Tehran University of Medical Sciences

Institute of Biochemistry and Biophysics

ANNUAL BIOINFORMATICS CONFERENCE 2022

DISCUSSIONS AND INSIGHTS

MAJOR TOPICS

- System Biology
- Big Data in Biology
- Structural Bioinformatics
- Biological Sequences Analysis
- AI & Machine Learning in Biology
- Modeling in Computeational Biology
- Computational Drug Design & Discovery

1st International and 10th National Iranian Conference on Bioinformatics

22-24, February 2022

University of Tehran
Kish International Compus

Organizers

Events come in all shapes and sizes, but the principles of organization are always the same. The key to a successful event is making sure it's well planned and that it meets all of your targeted expectations.

University of Tehran
Kish International Campus

IBIS
Iranian Bioinformatics Society

Tehran University
Medical Science

Welcome Messages

01

Statement by the Chairman & Executive Secretary of the Conference:

Dr. Kaveh Kavousi,
University of Tehran

It is extremely fortunate that under the auspices of a fruitful collaboration with the Iranian Bioinformatics Society (IBIS), University of Tehran Kish International Campus, and Tehran University of Medical Sciences, the 1st international and the 10th national Iranian Conference on Bioinformatics (ICB) is being held virtually and in-person on February 22nd to 24th. The purpose of holding this conference is to provide a useful forum and to create an opportunity to exchange views and ideas on the latest scientific and research achievements, theoretical approaches and practical applications of bioinformatics by delivering keynote lectures, displaying posters and holding specialized roundtables. The conference also aims to establish effective communication between professors, students and researchers in the filed in order to exchange ideas, establish scientific collaborations and communication between the academia and the industry.

Favored by IBIS's 15 years of invaluable experience, ICB10 is being held this year as an international event under the supervision of the most prominent and internationally recognized great names and scholars in the field of bioinformatics.

It is such a great pleasure and honor that ICB10's keynote speakers are all among the most prominent bioinformatics experts. We intend to provide many learning opportunities to our dear participants by holding numerous specialized workshops and opening up new learning horizons ahead of our dear participants.

In the meantime, holding a conference on the beautiful Kish Island is an opportunity for tourism enthusiasts to visit and enjoy the attractions and wonders of this small, but lovely Persian Gulf island. It is hoped that holding this conference paves the way and contribute to the synergy of experience,

more constructive discourse and direct and face-to-face communication between researchers and students active in the field of bioinformatics.

As the Executive Secretary of the 1st international and the 10th national Iranian Conference on Bioinformatics (ICB), I cordially invite all professors, students and researchers to increase the academic richness of this event by participating and presenting the results of their original research and experience.

It is hoped that holding this conference will pave the way and contribute to the synergy of experience, more constructive discourse and direct and face-to-face communication between researchers and students active in the field of bioinformatics.

02

Statement by the Scientific Secretary of the Conference:

Dr. Sajjad Gharaghani, University of Tehran

By the hand of God almighty and in line with 15 years of invaluable experience of the Iranian Bioinformatics Society (IBIS) as well as the efforts of all our esteemed colleagues from University of Tehran Kish International Campus, and Tehran University of Medical Sciences, we were able to organize the 1st international and the 10th national Iranian Conference on Bioinformatics (ICB) to hold virtually and in-person in Kish Island. The beautiful Kish Island hosts the ICB10 international event from February 22nd to 24th 2022. ICB10 focuses on various aspects of bioinformatics, including artificial intelligence and machine learning in biology, modeling in computational biology, big data in biology, discovery and computational drug design, system biology, structural bioinformatics, and Biological Sequences Analysis. The Conference aims to create an opportunity for professors, students and researchers to contribute to the enhancement of useful scientific experiences by presenting their latest scientific-research achievements. It is also hoped that the exchange of ideas during various specialized meetings would lead to the formation of effective communication between the academia and the industry. A number of scientific lectures based on the latest original research are scheduled for each day. In addition, the eight Keynote Speakers of the Conference are among the world's leading bioinformatics experts who have presented numerous scientific works and valuable books on bioinformatics to the international community. More than 10 training workshops with retraining advantages are being held by professors and scholars in various fields of bioinformatics, including drug design, genomics and analysis of exome data, metagenomics, among others. It's a great opportunity for those interested not only to earn relevant knowledge, but increase their skills. As the Scientific Secretary of the Iranian Conference on Bioinformatics, I invite all professors, students and researchers to add to the scientific richness of this Conference by attending and presenting the results of their original research and valuable experiences so that we can share this interdisciplinary knowledge in our beloved homeland of Iran and look for chances to promote internationally.

Table of Contents

Main Topics

01

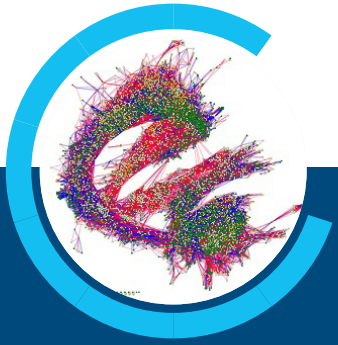
AI & Machine Learning in Biology

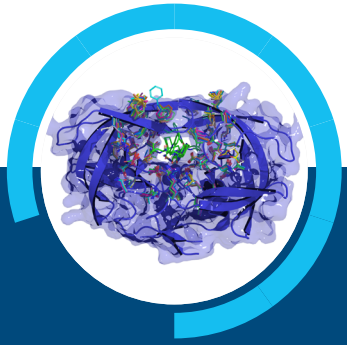
02

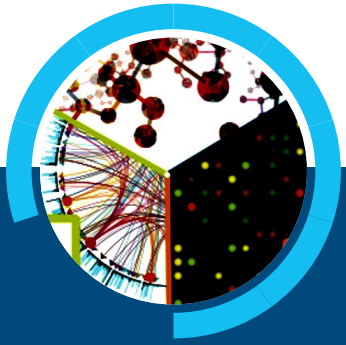
Big Data in Biology

03

Computational Drug Design and Discovery







04

Modeling in Computational Biology

05

Biological Sequence Analysis

Analysis of:

- Genomics Data
- Metagenomics Data
- Proteomics Data
- Transcriptomics Data
- Epigenomics Data
- Metabolomics Data

06

Structural Bio-informatics

07

Systems Biology

Agenda at a Glance

19 Feb

Hijri | 1400.11.30

Workshops

Tehran Time	Presenter	Title
5:00 PM	Dr. Seyed Shahriar Arab	Introduction of Molecular Dynamics Simulation and Its Application in Protein Engineering and Drug Design

20 Feb

Hijri | 1400.12.01

Workshops

Tehran Time	Presenter	Title
8:00 AM	Dr. Samaneh Maleknia	Microarray Data Analysis

25 Feb

Hijri | 1400.12.06

Workshops

Tehran Time	Presenter	Title
8:00 AM	Dr. Mohammad Hossein Norouzi-Beirami	Metagenomic Data Analysis
4:00 PM	Dr. Seyed Hamid Aghaee-Bakhtiari	Analysis of miRNA Data

24 Feb

Hijri | 1400.12.05

Workshops

Tehran Time	Presenter	Title
8:30 AM	Dr. Sajjad Gharaghani	Computational Drug Design
5:00 PM	Dr. Mohammadreza Modaresi Mr. Mohammad Sadra Modaresi	Evaluation of Computational Approaches in Gene Therapy, Focusing on Cystic Fibrosis

26 Feb

Hijri | 1400.12.07

Workshops

Tehran Time	Presenter	Title
8:00 AM	Dr. Mohammad Kera-matipour Mr. Naser Elmi Dr. Fahimeh Palizban	Exome Data Analysis for Germline Variant Calling
1:00 PM	Dr. Anna Meyfour Ms. Leila Ghanbari Maman	Analysis of Proteomics Data

27 Feb

Hijri | 1400.12.08

Workshops

Tehran Time	Presenter	Title
1:00 PM	Dr. Eskandar Omidinia Dr. Hamid Shahbazmo-hammadi Dr. Elham Rismani	Immunoinformatics In Silico Vaccine Design
4:00 PM	Dr. Esmaeil Shahsavand Ananloo	Application of Bioinformatics in the Normal and Abnormal Functioning of the Brain, Mind, and Behavior

21 Feb

Hijri | 1400.12.02

Workshops

Tehran Time	Presenter	Title
8:00 AM	Dr. Parvaneh Nikpour Ms. Maryam Sadat Hosseini Ms. Basireh Bahrami	Integrated Analysis of Cancer Transcriptome and Methylome Data
1:00 PM	Dr. Najmeh Salehi Ms. Fereshteh Fallah Atanaki	Single-Cell Transcriptome Analysis with R

22 Feb

Hijri | 1400.12.03

Keynote Speakers

Tehran Time	Presenter	Title
9:00 AM	Ali Shojaie, University of Washington	Differential Network Analysis
11:00 AM	Karoline Faust, Katholieke Universiteit Leuven	From Hairballs to Hypotheses: Microbial Network Analysis
1:30 PM	Wing-Kin Sung, National University of Singapore	Structural Variation Calling and Its Application in Understanding Large Indels in 1047 Arabidopsis
2:15 PM	Limsoon Wong, National University of Singapore	Conservative Batch-Effect Correction for Single-Cell RNA-Seq Data Enables Discovery of Rare Cell Populations
3:30 PM	Andreas Bender, University of Cambridge	Artificial Intelligence in Drug Design - Approaches, Solutions, and Illusions
4:15 PM	Alireza Fotuhi siahpirani, University of Wisconsin-Madison	Computational Methods for Integrative Inference of Genome-Scale Gene Regulatory Networks

23 Feb

Hijri | 1400.12.04

Keynote Speakers

Tehran Time	Presenter	Title
8:15 AM	Ehsaneddin Asgari on behalf of Mohammad Mofrad, University of California, Berkeley	Life Language Processing
11:00 AM	Yves Moreau, University of Leuven	Bayesian Matrix Factorization and Deep Learning for Drug-Target Activity Prediction
1:30 PM	Kwoh Chee Keong, National University of Singapore	Deep Learning for Long Non-coding RNA Functional Annotation
2:15 PM	Babak H. Khalaj, Sharif University of Technology	The Key Interaction to Watch in The Coming Decade: Bio, Data Science and Engineering
4:15 PM	Jonathan Pevsner, Kennedy Krieger Institute, USA	Bioinformatics and Genomics
5:00 PM	Iman Hajirasouliha, Weill Cornell Medicine of Cornell University	Deep Learning Applications in Digital Pathology and Embryology

28 Feb

Hijri | 1400.12.09

Workshops

Tehran Time	Presenter	Title
4:00 PM	Ms. Shiva Beigizadeh	Types of Epidemiological Studies

Detailed Schedule

Plan Table

22 Feb			
Tehran Time	**Plan**	**Speaker name**	**Title**
08:00 AM-09:00 AM	Conference Opening Ceremony		
09:00 AM-09:45 AM	1st Keynote Speaker	Ali Shojaie, University of Washington, USA (Washington)	Differential Network Analysis
09:45 AM-10:30 AM	Oral Presentation		
10:30 AM-11:00 AM	**Break**		
11:00 AM-11:45 AM	2nd Keynote Speaker	Karoline Faust, Katholieke Universiteit Leuven, Belgium	From Hairballs to Hypotheses: Microbial Network Analysis
11:45 AM-12:30 PM	Oral Presentation		
12:30 PM-01:30 PM	**Break**		
01:30 PM-02:15 PM	3rd Keynote Speaker	Wing-Kin Sung, National University of Singapore, Singapore	Structural Variation Calling and Its Application in Understanding Large Indels in 1047 rabidopsis
02:15 PM-03:00 PM	4th Keynote Speaker	Limsoon Wong, National University of Singapore, Singapore	Conservative Batch-Effect Correction for Single-Cell RNA-Seq Data Enables Discovery of Rare Cell Populations
03:00 PM-03:30 PM	**Break**		
03:30 PM-04:15 PM	5th Keynote Speaker	Andreas Bender, University of Cambridge, United Kingdom	Artificial Intelligence in Drug Design - Approaches, Solutions, and Illusions
04:15 PM-05:00 PM	6th Keynote Speaker	Alireza Fotuhi siahpirani, University of Wisconsin-Madison, USA (Wisconsin)	Computational Methods for Integrative Inference of Genome-Scale Gene Regulatory Networks
05:00 PM-06:00 PM	Oral Presentation		
06:0 AM-06:30 AM	**Break**		
06:30 PM- 07:30 PM	Scientific panel	Kaveh Kavousi, Hamed Sheydaeian, Aidin Parnia	Digital Transformation in Healthcare: Strategies, Trends and Issues

23 Feb			
Tehran Time	**Plan**	**Speaker name**	**Title**
08:15 AM-09:00 AM	7th Keynote Speaker	Ehsaneddin Asgari on behalf of Mohammad Mofrad, University of California, Berkeley, USA (California)	Life Language Processing
09:00 AM-09:30 AM	8th Keynote Speaker	Pooya Zakeri	An Explainable Matrix Factorization-based Data Fusion Model to Untangle the Connections between Different Alzheimer's Diseases Studies
09:30 AM-10:30 AM	Oral Presentation		
10:30 AM-11:00 AM	**Break**		
11:00 AM-11:45 AM	9th Keynote Speaker	Yves Moreau, University of Leuven, Belgium	Bayesian Matrix Factorization and Deep Learning for Drug-Target Activity Prediction
11:45 AM-12:30 PM	Oral Presentation		
12:30 PM-01:30 PM	**Break**		
01:30 PM-02:15 PM	10th Keynote Speaker	Kwoh Chee Keong, National University of Singapore, Singapore	Deep Learning for Long Non-Coding RNA Functional Annotation
02:15 PM-03:00 PM	11th Keynote Speaker	Babak H. Khalaj, Sharif University of Technology, Iran	The Key Interaction to Watch in the Coming Decade: Bio, Data Science and Engineering
03:00 PM-03:30 PM	**Break**		
03:30 PM-04:15 PM	Oral Presentation		
04:15 PM-05:00 PM	12th Keynote Speaker	Jonathan Pevsner, Kennedy Krieger Institute, USA	Bioinformatics and Genomics
05:00 PM-05:45 PM	13th Keynote Speaker	Iman Hajirasouliha, Weill Cornell Medicine of Cornell University, USA (New York)	Deep Learning Applications in Digital Pathology and Embryology
05:45 PM-06:30 PM	Conference Closing Ceremony		

Detailed Schedule

Oral Presentation

Read the abstract of these presentations at the end of the booklet.

22 Feb		
Tehran Time	**Speaker Name**	**Title**
09:45 AM-10:05 AM	Room1: Mohsen Hooshmand Room2: Mohadese Dousti	Room1: Indicator Regularized Non-negative Matrix Factorization vs. A Novel Combinatorial Heuristic Matrix Factorization: A Comparison of Matrix Factorization Methods as the Building Block of Drug Repurposing Room2: In Silico Transcriptome Analysis of Drought and Salt Involved Responsiveness Genes in Brassica Napus
10:05 AM-10:25 PM	Room1: Nazanin Hosseinkhan Room2: Sara Mohammdi	Room1: Investigating the Obesity Paradox in Patients With Hepatocellular Carcinomas Using Bioinformatics Approaches Room2: Comparison of Random Forest and Boosted Regression Tree in Improving Predicted Affinity
11:45 AM-12:05 PM	Room1: Naser Elmi Room2: Milad Nourozi	Room1: Inferring Microbial Communities Using Constrained Damped Lasso Regression Based on the Generalized Lotka-Volterra Model Room2: Casilico: A Versatile CRISPR Package for In silico CRISPR RNA Designing

22 Feb		
Tehran Time	**Speaker Name**	**Title**
12:05 PM-12:25 PM	Room1: Amirali Zandieh Room2: Donya Afsharjahanshahi	Room1: A Diffusion Kernel-Based Approach for Protein Domain Identification Room2: Exploration of Plastic Contaminated Soil Metagenome to Identify Novel Plastic Degradation Enzymes
05:00 PM-05:20 PM	Room1: Alexandre Rossi Paschoal Room2: Nasibeh Khayer	Room1: Non-coding and Transposable Elements Discovery Through Artificial Intelligence Approaches Room2: Using Liquid Association Analysis to Detect Controller Genes Involved in Pituitary Non-functioning Adenoma Invasiveness
05:20 PM-05:40 PM	Room1: Fateme Abbasi Room2: Mohammad Taheri-Ledari	Room1: A Novel Method for Predicting Drug Synergy Based on Matrix Factorization Room2: Simultaneous Inference of Cell-Line-Specific Gene Regulatory Networks and Mode-Of-Action of Drugs From Drug-Induced Gene Expression Measurements
05:40 PM-06:00 PM	Room1: Naser Faraji Room2: Zohre Toghraee	Room1: ApInAPDB: A Database of Apoptosis-inducing Anticancer Peptides Room2: Negative Binomial Mixed Models for Identifying Oncogenic Dependencies Through Analysis of RNAi Screening Data

Detailed Schedule

Oral Presentation

Read the abstract of these presentations at the end of the booklet.

23 Feb		
Tehran Time	**Speaker Name**	**Title**
09:00 AM-09:30 AM	Pooya Zakeri	Room1: An explainable matrix factorization-based data fusion model to untangle the connections between different Alzheimer's disease studies
09:30 AM-09:50 AM	Room1: Sara Fayazzadeh Room2: Maryam Hosseini	Room1: Identification of Peripheral Blood Mononuclear Cell Gene Signatures for Detection of Hepatocellular Carcinoma Room2: Integrative Analysis of DNA Methylation and Gene Expression to Identify Gastric Cancer Diagnostic Biomarkers via Machine Learning Approache
09:50 AM-10:10 AM	Room1: Zahra Salehi Room2: Faegheh Golabi	Room1: RNA-sequencing of CD4+ T cells in Relapsing-Remitting Multiple Sclerosis patients at relapse; deciphering the involvement of novel genes and pathways Room2: Development of a new oligonucleotide block location-based feature extraction (BLBFE) method for the classification of riboswitches
10:10 AM-10:30 AM	Room1: Leila Mirsadeghi Room2: Mehri Javid	Room1: EARN as a Precision Oncology Tool Leads Us to Propose the Targeted Genes Panel for Metastatic Breast Cancer Room2: How Does a Bacteriophage Enzybiotic Target Bacteria? Introducing a Structural Model of Bacteriophage PhaxI Lytic Enzyme

23 Feb		
Tehran Time	**Speaker Name**	**Title**
11:45 AM-12:05 PM	Room1: Shima Jamalirad Room2: Samira Shafiee	Room1: Transcriptome Profiling Analysis of Suaeda Salsa Shoots Under Salinity Stress Room2: Target Prediction for the Inhibitors of VEGFR2 as Anti-colorectal Cancer Compounds Using Similarity-Based Search Methods
12:05 PM-12:25 PM	Room1: Masoud Arabfard Room2: Saeideh Khodaee	Room1: Investigation of Common Genes in Different Stages of Non-alcoholic Fatty Liver Disease With Microarray Datasets Analysis Room2: A New Reconstruction of Mouse Metebolic Model Using Orthology-Based Approach
03:30 PM-03:50 PM	Room1: Somaieh Soltani Room2: Mojtaba Ghani-zadeh	Room1: Inflammatory Target Prediction for the FDA-Approved Anticancer Drugs Using Morgan Fingerprint Similarity-Based Methods Room2: Novel Insights in Targeted Therapy of Cancer by Mathematical Modeling of Cancer Immunoediting
03:50 PM-04:15 PM	Room1: Karim Rahimian Room2: Fatemeh Bayani	Room1: Classification of Autistic Patients and Control via Utilizing Dictionary of Functional Modes as Brain Atlas Room2: Molecular Dynamics Simulations Show Structural Insights Into the N-terminal Domain Mutations of the Spike Protein in the Omicron (B.1.1.529) Variant of SARS-CoV-2

Detailed Schedule

Workshops

Content	Date (AD)	Time	Presenter
Introduction of Molecular Dynamics Simulation and Its Application in Protein Engineering and Drug Design	2022.02.19	05:00 PM-07:00 PM	Dr. Seyed Shahriar Arab
Microarray Data Analysis	2022.02.20	08:00 AM-12:00 PM	Dr. Samaneh Maleknia
Integrated Analysis of Cancer Transcriptome and Methylome Data	2022.02.21	08:00 AM-12:00 PM	Dr. Parvaneh Nikpour Ms. Maryam Sadat Hosseini Ms. Basireh Bahrami
Single-Cell Transcriptome Analysis with R	2022.02.21	01:00 PM-03:00 PM	Dr. Najmeh Salehi Ms. Fereshteh Fallah Atanaki
Computational Drug Design	2022.02.24	08:30 AM-05:00 PM	Dr. Sajjad Gharaghani
Evaluation of Computational Approaches in Gene Therapy, Focusing on Cystic Fibrosis	2022.02.24	05:00 PM-09:00 PM	Dr. Mohammadreza Modaresi Mr. Mohammad Sadra Modaresi
Metagenomic Data Analysis	2022.02.25	08:00 AM-04:00 PM	Dr. Mohammad Hossein Norouzi-Beirami
Analysis of miRNA Data	2022.02.25	04:00 PM-08:00 PM	Dr. Seyed Hamid Aghaee-Bakhtiari
Exome Data Analysis for Germline Variant Calling	2022.02.26	08:00 AM-01:00 PM	Dr. Mohammad Keramatipour Mr. Naser Elmi Dr. Fahimeh Palizban
Analysis of Proteomics Data	2022.02.26	01:00 PM-05:00 PM	Dr. Anna Meyfour Ms. Leila Ghanbari Maman
Immunoinformatics In Silico Vaccine Design	2022.02.27	01:00 PM-04:00 PM	Dr. Eskandar Omidinia Dr. Hamid Shahbazmohammadi Dr. Elham Rismani
Application of Bioinformatics in the Normal and Abnormal Functioning of the Brain, Mind, and Behavior	2022.02.27	04:00 PM-06:00 PM	Dr. Esmaeil Shahsavand Ananloo
Types of Epidemiological Studies	2022.02.28	04:00 PM-08:00 PM	Ms. Shiva Beigizadeh

Keynote **Speakers**

Ali Shojaei | Ph.D.

Professional Positions	• Professor and Associate Chair for Strategic Research Affair at the Department of Biostatistics, University of Washington • Founding Director, Summer Institute for Statistics in Big Data (SISBID) • Affiliate Member, Center for Statistics in Social Sciences (CSSS), University of Washington
Main Research Areas	• Biostatistics • High-Dimensional Data and Statistical Learning • Longitudinal and Multilevel Data • Computational Biology • Statistical Network Analysis • Neuroimaging
Brief Research Description	Dr. Ali Shojaei is interested in developing statistical methods for analysis of large, complex systems, particularly biological and social systems. The main characteristic of these systems is that they are often comprised of a large number of interacting components, and the behavior of the system may not be evident from that of individual components. Networks offer a convenient framework for analyzing large, complex systems. His research thus focuses on statistical methods for high-dimensional networks. Dr. Shojaei also develops statistical machine learning methods for estimation and inference in high-dimensional problems (i.e., when there are more variables than observations), particularly, when variables and/or observations are correlated with each other. This intersection of statistical machine learning, network analysis is an emerging research area.

Andreas Bender | Ph.D.

Professional Positions	• Associate Director Computational ADME and Safety at AstraZeneca Cambridge • Head of Data-Driven Drug Discovery and Molecular Informatics group, Centre for Molecular Informatics, Department of Chemistry, Cambridge, United Kingdom • PhD in Cheminformatics • One of the Founders of Healx Ltd and PharmEnable Ltd
Main Research Areas	• AI/ML/data science and their application in drug discovery chemical biology, and in silico drug safety • Mode-of-action analysis and target prediction • Modelling mixtures of bioactive compounds and traditional medicines • Integrating chemical/biological data for characterizing compound action and disease • Toxicity prediction
Brief Research Description	Dr. Andreas Bender is a Reader for Molecular Informatics at the Centre for Molecular Informatics, part of the Department of Chemistry of the University of Cambridge. His research interests include molecular property - and toxicity - prediction using integrated chemical and biological information, utilizing gene expression data to characterize compound activities, and the modelling of bioactive mixtures of chemical compounds.

Iman Hajirasouliha | Ph.D.

Professional Positions	• Assistant Professor of Physiology and Biophysics, Weill Cornell Medicine, Cornell University, New York City • Assistant Professor of Computational Genomics in Computational Biomedicine in the Institute for Computational Biomedicine • Member of ICB and PBSB
Main Research Areas	• Machine Learning • Deep Learning • Genomics • Metagenomics • Cancer Research
Brief Research Description	Dr. Iman Hajirasouliha is an Assistant Professor of Computational Genomics in Computational Biomedicine and an Assistant Professor of Physiology and Biophysics, Physiology and Biophysics. Dr.Hajirasouliha and his lab team are passionate about developing new algorithms, machine learning and deep learning methods, and their applications to genomics, metagenomics and cancer research. Some of the current projects in the lab include characterizing human genomes and metagenomes sequenced by exciting new technologies, quantifying cancer evolution, a study of tumor heterogeneity using genomics and digital pathology images.

Jonathan Pevsner | Ph.D.

Professional Positions	• Professor and research scientist at the Kennedy Krieger Institute • Holder a primary faculty appointment in the Department of Psychiatry and Behavioral Sciences at the Johns Hopkins University School of Medicine • Author of Bioinformatics and Functional Genomics
Main Research Areas	• Cellular and Molecular Neuroscience • Neurobiology of Disease • Database Referencing of Array Genes ONline (DRAGON) • Standardization and Normalization of Microarray Data (SNOMAD) • Statistical Analysis and Visualization of Annotated Gene Expression data (SAVAGE)
Brief Research Description	Dr. Pevsner received his bachelor's degree in psychology from Haverford College and his doctoral degree in pharmacology and molecular sciences from the Johns Hopkins School of Medicine. He pursued post-doctoral training at the Stanford University School of Medicine, and joined the faculty of Kennedy Krieger Institute in 1995. The Pevsner Lab studies the molecular basis of childhood and adult brain disorders.

Keynote **Speakers**

Limsoon Wong | Ph.D.

Professional Positions	• Kwan-Im-Thong-Hood-Cho-Temple Professor in Computing at the National University of Singapore (NUS) • Professor (now honorary) of pathology in the Yong Loo Lin School of Medicine at NUS
Main Research Areas	• Bioinformatics • Data Mining • Databases • Computational Biology
Brief Research Description	Dr. Wong works mostly on knowledge discovery technologies and their application to biomedicine. Limsoon is a Fellow of the ACM, named in 2013 for his contributions to database theory and computational biology. Some of his other awards include the 2003 FEER Asian Innovation Gold Award for his work on treatment optimization of childhood leukemias, and the ICDT 2014 Test of Time Award for his work on naturally embedded query languages.

Mohammad R. K. Mofrad | Ph.D.

Professional Positions	• Professor of Bioengineering at the Department of Bioengineering, University of California, Berkeley • Fellow of the American Institute for Medical and Biological Engineering (AIMBE) • Invitational Fellow of Japan Society for Promotion of Science (JSPS)
Main Research Areas	• Computational Biology and Bioinformatics • Bacterial Communities and Microbiomes • Deep Learning for Biology and Medicine • Molecular Biophysics • Computational Biomechanics • Molecular Cell Biomechanics
Brief Research Description	Dr. Mohammad R. K. Mofrad is Professor of Bioengineering and Mechanical Engineering at the University of California, Berkeley. His research program is aimed to combine the state-of-the-art molecular and multiscale biomechanics, computational biology and bioinformatics, and statistical and machine learning approaches towards understanding the biological processes involved in human diseases. Mofrad Lab's multidisciplinary works have appeared in diverse scientific and biomedical engineering journals. He has co-edited several books, including Cytoskeletal Mechanics and Cellular Mechanotransduction published by Cambridge University Press.

Wing-Kin Sung | Ph.D.

Professional Positions	• Professor, School of Computing, National University of Singapore • Senior Group Leader, Genome Institute of Singapore • Ph.D. in Computer Science
Main Research Areas	• Bioinformatics • Computational Genomics • Health Informatics
Brief Research Description	Dr. Wing-Kin Sung received both the B.Sc. and the Ph.D. degree in the Department of Computer Science from the University of Hong Kong in 1993, 1998, respectively. He is a professor in the Department of Computer Science, School of Computing, NUS. Also, he is a senior group leader in Genome Institute of Singapore. He has over 20 years experience in Bioinformatics and Computational genomic research. He also teaches courses on bioinformatics for both undergraduate and postgraduate. He was conferred the 2003 FIT paper award (Japan), the 2006 National Science Award (Singapore), and the 2008 Young Researcher Award (NUS) for his research contribution in bioinformatics and algorithm.

Kwoh Chee-Keong | Ph.D.

Professional Positions	• Associate Professor School of Computer Science and Engineering, Nanyang Technological University, Singapore • Senior Member of IEEE • Senior Member of IES (Institution of Engineers), Singapore • Member of AMBIS (Association for Medical and Bioinformatics), Singapore • Deputy Executive Director of PaCE (Professional and Adult Continuing Education) Academy under Singapore Polytechnic Institute
Main Research Areas	• Data Analytics • Data Mining • Bioinformatics • Graph Theory • Biology and Medical Computing • Deep Learning • Pattern Clustering
Brief Research Description	Dr. Kwoh received Ph.D. degree in computing from the Imperial College of Science, Technology, and Medicine, University of London, U.K., in 1995. His main interests lie in the desire to making sense of big heterogeneous data for real application in engineering, life science, and medical. Also, his researches include data mining, soft computing, graph-based inference, and application areas include bioinformatics and engineering. He was the recipient of the best Faculty Mentor Award from Temasek Foundation in 2013 and 2014.

Keynote **Speakers**

Yves Moreau | Ph.D.

Professional Positions	• Full professor Faculty of Engineering Science at the University of Leuven, Belgium • Member of the Division STADIUS, Stadius Centre for Dynamical Systems, Signal Processing and Data Analytics • Member of Leuven.AI - KU Leuven Institute for Artificial Intelligence • Member of LISCO - KU Leuven Institute for Single Cell Omics
Main Research Areas	• Computational biology • Genomic medicine • Bioinformatics • Analyze and integrate large-scale, complex, heterogeneous biologica and clinical data
Brief Research Description	His team at the University of Leuven, Belgium focuses on AI algorithms and software platforms for the integration of complex data in clinical genomics and drug discovery: (1) federated analysis of real-world clinical and genomic data, (2) data fusion algorithms for the identification of pathogenic genetic variation in rare genetic disorders and liquid biopsies, and (3) data fusion for drug discovery and drug design. At the algorithmic level, he focuses on the development of novel AI methods, such as deep learning and Bayesian matrix factorization, for the fusion of heterogeneous sparsely-observed data.

Karolin Faust | Ph.D.

Professional Positions	• Associate Professor in Microbiological Bioinformatics at KU Leuven science 2021 • Principal Investigator (PI) in Laboratory of Molecular Bacteriology, Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven • Assistant Professor in Microbiological Bioinformatics at KU Leuven science 2016
Main Research Areas	• Metagenomics • Microbial ecology • Bioinformatics • Systems Biology
Brief Research Description	As a microbiologist and bioinformatician, Dr. Karoline Faust, who is on the faculty of Katholieke Universiteit Leuven (KU), has been moving into systems biology. The main interest of her lab is to explore microbial community structure and dynamics in silico and in vitro. She uses mathematical models to assess and model bustling microbial communities.

Babak H. Khalaj | Ph.D.

Professional Positions	• Professor of Electrical Engineering, Sharif University of Technology, School of Electrical Engineering • Director of Sharif Data Science Center • Senior Consultant in the areas of data communications and bio-signal processing, since 1999 • Ph.D. in Electrical Engineering, Stanford University, Stanford, CA, 1996
Main Research Areas	• Bio Big Data Analytics for Communication Networks • Algorithm Design and Analysis • Medical Networks for Cancer Diagnosis
Brief Research Description	Prof. Khalaj has been with the pioneering team at Stanford University. He was the recipient of the Alexander von Humboldt Fellowship from 2007 to 2008 and Nokia Visiting Professor Scholarship in 2018. He is currently working on different aspects of data science from both foundations side to application areas such as health and communications.

Introduction of **Main Topics**

01

AI & Machine Learning in Biology

Modern methods in biology, such as high-throughput sequencing and multi-omics approaches, generate enormous amounts of data. To discover knowledge about these data, Artificial Intelligence (AI) needs to be applied. AI technologies allow us not only to collect and analyze data at unprecedented scales but also to develop comprehensive predictive models that involve various sub-disciplines.

Meanwhile, machine learning algorithms as a branch of AI has ushered in a new era of high-throughput data analysis, where particularly it is difficult for researchers to make decisions based on laboratory evidence. On the other hand, applications of machine learning in biological phenomena have been increasing, especially with the growth in using deep-learning, transfer learning, and ensemble learning models.

Leveraged by both advancements in biological data generation and improvements in machine learning methods, researchers are promising for the analysis of complex biological data. Some of the advantages of using these methods include prognosis and diagnosis of complex diseases, such as cancers, studies that could have had an impact on COVID-19 and drug design and discovery.

02

Big Data in Biology

Recent technological advances allow us to generate high throughput data in a cost-efficient manner. Meanwhile, the low cost of data generation is leading us to the big data era. The availability of big data provides unprecedented opportunities; however, it also raises new challenges for data mining and analysis. In Omics era of the life sciences, data is presented in many forms, which represent the information at various levels of biological systems, including genome, transcriptome, epigenome, proteome, metabolome, molecular imaging, molecular pathways, different population of people and clinical or medical records.

The biological data is big, and its scale has already been well beyond petabyte (PB) even exabyte (EB). Nobody doubts that the biological data will create huge number of values, if scientists can overcome many challenges, e.g., how to handle the complexity of information, how to integrate the data

from very heterogeneous resources, what kind of principles or standards to be adopted when facing with the big data. Tools and techniques for analyzing big biological data enable us to analyze massive amount of information. It leads to a better understanding of the basic biomedical mechanisms, which can be further applied to personalized medicine.
In general, big data has four important features, so called four V's: Volume of data, Velocity of processing the data, Variability of data sources and Veracity of the data quality. Further, life sciences need more robust, accurate, and precise ways to handle the big data. Recent works in this area have already brought remarkable advantages and opportunities, which implies the central roles of bioinformatics and bioinformaticians in the future researches of the biological and biomedical fields.

03

Computational Drug Design and Discovery

Drug design aims to find a chemical compound that can fit a specific cavity on a protein target both geometrically and chemically. After passing the animal tests and human clinical trials, this compound becomes an available drug to patients. The process of drug discovery is very complex and requires an interdisciplinary effort to design effectively and commercially feasible drugs.
The conventional drug design methods include a random screening of chemicals found in nature or synthesized in laboratories. The problems with this method are the long design cycle and high cost. Computer-aided drug design (CADD) comprises a broad range of computational approaches that are part of modern drug discovery. CADD methods have played key role contributions in the development of drugs in clinical trials. The modern approach includes structural-based drug design with the help of computational methods that have speeded up the drug discovery process in an efficient manner. Also, an improved generation of software with easier operation and higher computational tools has been developed to generate worthy compounds with refinement capability. These tools can tap into chemical information to shorten the cycle of drug discovery, and thus make drug discovery more cost-effective.

04

Biological Sequence Analysis

Analysis of Genomics Data

Various NGS sequencing platforms have enabled the generation of genome-scale sequencing data in rapid time frames. Also, genome-wide sequencing technologies provide accumulated data to understand the crucial mechanisms of numerous biological functions in cells. Because of the huge density of genomic data, it is necessary to analyze them by computational tools.

To address this aim, the original files obtained from high throughput sequencing platforms are transformed to short reads by base calling. They are recorded in FASTQ format, which contains sequence information and corresponding sequencing quality data. Raw sequencing data may contain adapter contaminated and low-quality reads. Given that, these artifacts increase the complexity of downstream analyses and quality control is an essential step.

The workflow of the bioinformatics analysis pipeline for WGS/WES data contain raw data, clean data, alignment, SNP, InDel and annotation, respectively. Some of the applications of the genome-wide analysis encompass helping to enrich the genome databases, finding plausible driver genes to identify screening, prognostic, or diagnostic genomic biomarkers, development of targeted therapies based on personalized medicine and the emergence of precision medicine.

05

Biological Sequences Analysis

Analysis of Metagenomics Data

Since the beginning of the twentieth century, it has become clear that most micro-organisms cannot be cultured and their enzymes cannot be accessed using conventional methods. The introduction of metagenomics is one of the spin-offs from that challenge. Metagenomics is a culture-independent technique that makes it theoretically possible to study any type of sample and offers us the possibility of studying DNA from an entire organismal community.

Metagenomic analyses begin with the isolation of microbial DNA from an environmental sample and can be categorized as taxonomic profiling, functional profiling and comparative analysis. Additionally, pipeline of metagenomics analysis has been summarized to three major phases, e.g., extract high quality data, constructing gene catalogue and mapping, normalization, computing correlation between genes, and analyzing.

06

Modeling in Computational Biology

Simulation and In Silico modeling allow researchers to observe the dynamics of complex biological systems that are not perceptible in static diagrams. Indeed, computational models help to analyze and understand the nature of biological phenomena and complex systems, specifically when the system involves several subunits interacting in a time-dependent manner.

Models' production of biological networks is one of the prominent aims of simulation to provide insight into the basic biological organizations and predict more reliable medical treatments. Further, understanding of metabolic functions requires knowledge of the dynamic modeling and comprehension of the regulation of biological networks, e.g., the capture of intracellular metabolism has been applicable using metabolic modeling based on the systems biology approach.

Some applications of computational modeling include exploration of the cells and molecules' structure, such as protein, contribution to the discovery of intracellular biochemical mechanisms and metabolic processes understanding, study and prediction of the organisms' behavior, simulations of targeted drug delivery systems in cancers, reliable medical treatments developing.

07

Biological Sequences Analysis

Analysis of Proteomics Data

The proteome is defined as the totality of the expressed proteins by genome of an organism, tissue, or cell under a determined set of time, environmental or physiological conditions. Alongside this concept, proteomics involves studying post-translational changes and the interaction of proteins with other molecules. This branch of omics has generated a wealth of data and many useful insights into biological processes.

Furthermore, proteomics includes the suite of technologies and techniques used in particularly measurements of peptides and bioinformatics tools. Proteomics is crucial for prognosis, early detection, and effective monitoring of disease trend. On the other, it has a vital role in drug development.

Applications of proteomics data include new drug discovery, detection of various diagnostic bio-markers, finding candidates for vaccine production and understanding of pathogenicity mechanisms.

08

Biological Sequences Analysis

Analysis of Transcriptomics Data

The information content of an organism is recorded in the genome and expressed through transcription. Transcriptomics technologies are the techniques that used to study an organism's transcriptome; i.e., all of its RNA transcripts.

Measuring the expression of an organism's genes in different tissues, conditions, or time points gives information on how genes are regulated and reveals details of an organism's biology. It can also help to infer the functions of previously unannotated genes.

Transcriptomics analysis has enabled the study of how gene expression changes in different organisms and has been advanced in the understanding of human diseases. Some of the transcriptomics applications encompass development of immunological studies, offering achievements in personalized medicine, support of drug development and providing insight into the function of the genes.

09

Biological Sequences Analysis

Analysis of Epigenomics Data

Epigenomics is a branch of omics-data that studies changes other than those encoded in DNA sequences. Epigenetic mechanisms describe potential biological and molecular routes which could link the gap between genetic influences and environmental risk factors contributing to the occurrence of the biological phenomena.

Epigenetic changes are heritable. The most relevant epigenetic phenomenon occurs in cytosine methylation of DNA at CpG dinucleotides. Furthermore, epigenetic changes involve DNA–protein interactions, such as post-translational modifications of histone, chromatin remodeling, and micro-RNA interactions. Several experimental techniques have been developed to map genome-wide epigenetic information. They are ChIP-on-chip, ChIP-seq, and bisulfite sequencing, including RRBS and WGBS.

The importance of epigenetics is so great that has ushered the new horizon in precision medicine. Epigenome studies are critical for understanding the mechanisms underlying the development of cancer, e.g., aberrant DNA methylation is associated with the induction of a variety of tumors.

Therefore, epigenome analysis can represent a promising tool for prognosis, diagnosis, and a new scope in anticancer therapeutics. Some epigenomics applications are enhancement of precision medicine, explain the regulation of some defense mechanisms, understanding the mechanisms of diseases, particularly carcinogenesis and description of some aspects of the evolution of organisms.

10 Biological Sequences Analysis

Analysis of Metabolomics Data

Metabolomics is the comprehensive study of small molecule metabolites in biological systems. Analyzing thousands of metabolites in biological samples helps us to get accurate understanding of metabolic status and biochemical events happening within an organism. It has become an increasingly powerful tool in diseases research.

In metabolomics, it is common to deal with large amounts of data generated by nuclear magnetic resonance (NMR) and/or mass spectrometry (MS). Based on different aims, it is crucial to use a variety of data analysis methods or a combination of them to achieve an accurate result.

Indeed, metabolomics is the large-scale analysis of metabolites and requires bioinformatics tools for data analysis, visualization, and integration. Some processes, including noise filtering, peak selection and deconvolution, peak detection, peak alignment, and the creation of a final data matrix are applied to process metabolic data, statistically.

11 Structural Bioinformatics

Structural Bioinformatics is a branch of bioinformatics that also called computational structural biology. This field of science mainly deals with problems of biology, statistics, mathematics, bioinformatics, biophysics, computational chemistry, enzymology, pharmaceutical sciences, and much more other disciplines are making contributions to structural bioinformatics. It helps the researchers to analysis and predict the

three-dimensional structure of biological macromolecules such as proteins, RNA, and DNA.

Meanwhile, its applications are expanding in more fields, including comparisons of overall folds and local motifs of both primary, secondary and tertiary structures, structural and functional predictions, molecular mechanism of folding/unfolding of macromolecules, evolution, and bioengineering. Further, it has wide application in the researches of biological sciences and drug developments.

Some of the most important Structural Bioinformatics applications include helping to discover protein structure alignment, macromolecules dynamics, molecular docking and structure-based therapeutics design.

12
Systems Biology

Biological organisms are so complex, and their many parts interact in numerous ways. Systems biology helps us to study the behavior of complex biological organization and processes in terms of the molecular components. Indeed, systems biology is an approach to achieve high understanding concerning the organism, tissue, or cell. It tries to bridge the gap between quantitative sciences and experimental biology to extract global models from cellular processes.

One of the most exploited quantitative concepts used in systems biology are networks, e.g., in personalized cancer therapy, the network approach to cancer systems biology in many cases involves computational modeling that focusing on likely target nodes.

Sponsors

Active contributions and commitment from many great people went into creating this event.

We want to express our gratitude to our sponsors for their generous and unwavering time and resources during the event-setting, which helped us to make our vision of this scientific conference a reality.

IBIS

Iranian Bioinformatics Society

IBB

Institue of Biochemistry of Bioinformatics

Khatam

Khatam University

Zist Fanavari

Vice President for Science and Technology

FIBS

Federation of Iran Bioscience Societies

ISC

Islamic world science Citation Center

GLDR

Gastrointestinal and Liver Diseases Research Center

IBIS

Iranian Bioinformatics Society

ISMVIP

Iranian Society of Machine Vision and Image Processing

IPM

Institute for Research in fundamental Science

CIVILICA

Veles dolorestiis pligni cuptatibeari

ISO

Iranian Society of Optometry

ZFN

Zist Fanavari Novin

Committees List

Role	Names	Institution
Chairman & Executive Secretary	Dr. Kaveh Kavousi	University of Tehran
Scientific Secretary	Dr. Sajjad Gharaghani	University of Tehran
Scientific Committee Members	Dr. Mehdi Sadeghi	National Institute of Genetic Engineering and Biotechnology
	Dr. Bahram Goliaei	University of Tehran
	Dr. Hamid Pezeshk	University of Tehran
	Dr. Ali Masoudi-Nejad	University of Tehran
	Dr. Mohammad Hossein Karimi Jafari	University of Tehran
	Dr. Kaveh Kavousi	University of Tehran
	Dr. Hesam Montazeri	University of Tehran
	Dr. Abbas Nowzari-Dalini	University of Tehran
	Dr. Seyed-Amir Marashi	University of Tehran
	Dr. Changiz Eslahchi	Shahid Beheshti University
	Dr. Seyed Shahriar Arab	Tarbiat Modares University
	Dr. Yazdan Asgari	Gustave Roussy Institute
	Dr. Mehdi Sahihi	Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC)
	Dr. Ali Sharifi-Zarchi	Sharif University of Technology
	Dr. Mehdi Totonchi	Royan Institute
	Dr. Mehdi Mirzaie	Tarbiat Modares University
	Dr. Zarrin Minuchehr	National Institute of Genetic Engineering and Biotechnology
	Dr. Fatemeh Zare-Mirakabad	Amirkabir University of Technology
	Dr. Abbasali Emamjomeh	University of Zabol

Role	Names	Institution
Scientific Committee Members	Dr. Mohammad Reza Bakhtiarizadeh	University of Tehran
	Dr. Mohammad Reza Naghavi	University of Tehran
	Dr. Ghasem Hosseini Salekdeh	Macquarie University
	Dr. Esmaeil Shahsavand Ananloo	Tehran University of Medical Sciences
	Dr. Masoud Arabfard	Baqiyatallah University of Medical Sciences
	Dr. Ali Najafi	Baqiyatallah University of Medical Sciences
	Dr. Babak H. Khalaj	Sharif University of Technology
	Dr. Seyed Abolfazl Motahari	Sharif University of Technology
	Dr. Mohammad Reza Ghaffari	Agricultural Biotechnology Research Institute of Iran (ABRII)
	Dr. Zeinabsadat Mousavian hor	University of Tehran
	Dr. Javad Zahiri	University of California, San Diego
	Dr. Najmeh Salehi	Institute For Research In Fundamental Sciences
	Dr. Zahra Salehi	Tehran University of Medical Science
	Dr. Mohsen Shahlaei	Kermanshah University of Medical Sciences
	Dr. Alireza Fotuhi Siahpirani	University of Tehran
	Dr. Mohammad Ganjtabesh	University of Tehran
	Dr. Alexandre Rossi Paschoal	Federal University of Technology – Parana (UTFPR), Brazil
	Dr. Leila Mirsadeghi	University of Tehran

Committees List

Role	Names	Institution
Executive Committee Members	Dr. Najmeh Salehi	Institute For Research In Fundamental Sciences
	Dr. Leila Mirsadeghi	University of Tehran
	Dr. Zahra Salehi	Tehran University of Medical Sciences
	Dr. Mona Majd	University of Tehran Kish International Campus
	Dr. Masoud Arabfard	Baqiyatallah University of Medical Sciences
	Dr. Sayyed-Reza Behafarin	University of Tehran
	Kazhaleh Mohammadi	Department of Pharmacy, College of Pharmacy, Knowledge University, Erbil, Iraq
	Zahra Hossein Zadeh	Tehran University of Medical Sciences, Kish International Campus
	Mahdi Pursalim	University of Tehran, Kish International Campus
	Zeynab Piryaei	University of Tehran, Kish International Campus
	Ali Reza Dehghan	University of Tehran, Kish International Campus
	Maryam Mehrabani	University of Tehran, Kish International Campus
	Shohreh Yazdani	University of Tehran, Kish International Campus
	Romina Norouzi	University of Tehran, Kish International Campus
	Niloofar Agharezaee	University of Tehran, Kish International Campus
	Mohammad Amin Moragheb	University of Tehran, Kish International Campus
	Donya Afshar-Jahanshahi	University of Tehran, Kish International Campus
	Shiva Beigizadeh	University of Tehran, Kish International Campus
	Fatemeh Abbasi	University of Tehran
	Mozhgan Mozaffari Legha	University of Tehran
	Motahareh Hakiminejad	University of Tehran
	Amirali Zandieh	University of Tehran

Role	Names	Institution
Executive Committee Members	Maryam Hojati	Tarbiat Modares University (TMU)
	Sara Fayazzadeh	Tarbiat Modares University (TMU)
	Maryam Rafieipour	Tarbiat Modares University (TMU)
	Aref Tavangar	University of Tehran
	Sahar Borhani	Tarbiat Modares University (TMU)
	Amir Zarghami	Tarbiat Modares University (TMU)
	Sayeh Emadi	Tarbiat Modares University (TMU)
	Nafiseh Nafisi-Fard	University of Tehran
	Mohammad Amin Mahmanzar	University of Tehran, Kish International Campus
	Maryam A.Baghban	Islamic Azad University Of Share-Rey, Yadegar Emam

Ramzist

Have you ever felt passion for sharing your knowledge or skills with those who don't even know?
Have you heard of science communication?
Ramzist provides a friendly and energetic environment for those who want to express themselves.

@ramzist_mag

@ramzist

List of Abstracts

Oral

AI & Machine Learning in Biology

Integrative analysis of DNA methylation and gene expression to identify gastric cancer diagnostic biomarkers via machine learning approache

Maryam Sadat Hosseini[1], Maryam Lotfi Shahreza[2], Parvaneh Nikpour[1]*

[1] *Department of Genetics and Molecular Biology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran*
[2] *Department of Computer Engineering, Shahreza Campus, University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): pnikpour@med.mui.ac.ir

Abstract

Gastric cancer is the second cause of cancer-related deaths. Most patients are diagnosed in a late stage because there are not proper methods for early cancer detection. Therefore, there is an urgent need for finding such biomarkers for this cancer. In the current study, our purpose is to integrate methylation and gene expression data from The Cancer Genome Atlas (TCGA) to find potential diagnostic biomarkers for gastric cancer via bioinformatics and machine learning approaches. DNA Methylation and gene expression data for gastric cancer were downloaded from TCGA using TCGABiolinks R-package. At first, probes locating at sex chromosomes, containing single nucleotide polymorphisms (SNPs) or missing values were removed. Then, we used ChAMP R-package for finding differentially methylated CpGs (DMCs). CpGs with adjusted p-values<0.05 and |delta β| > 0.25 were considered DMCs. Gene expression dataset was normalized via DESeq2 R-package. Genes were considered differentially expressed (DEGs) if they satisfied the threshold of |log2 fold change| > 1 and adjusted p-values<0.05. Since promoter hyper-methylation of tumor suppressor genes is one of the most important observations in cancer, we only continued with hyper-methylated CpGs (38 probes) located in the promoter of downregulated genes. Recursive feature elimination with cross-validation (RFECV) method was used to find features with highest discriminative power between tumoral and normal samples resulting in 4 final probes including cg10604646, cg22083047, cg07730329 and cg12741420. These features where then used for constructing a logistic regression model. We validated these markers in an independent set from GEO database (GSE30601). The area under the curve (AUC) of model was 0.904 indicating that the four markers could achieve excellent performance in distinguishing tumoral and normal gastric samples. Overall, the four high-performance diagnostic signatures built through machine learning approaches can improve gastric cancer precision management upon prospective clinical validation.

Key Words: *Machine learning, DNA methylation, Gene expression, Diagnosis, Gastric cance*

Comparison of Random Forest and Boosted Regression Tree in improving predicted affinity

Sara Mohammadi[1], Zahra Narimani[2]*, Mitra Ashouri[1], Mohammad Hossein Karimi-Jafari[1]

[1] *Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[2] *Department of Computer Science and Information Technology, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran*

E-mail address (corresponding author): saramohammadi63@ut.ac.ir

Abstract

One of the challenges in predicting protein-ligand affinity is how target flexibility should be considered in the docking procedure. Recently, ensemble docking has gained increasing attention and is incorporated as a promising solution to this problem, however, there is still missing information on how an optimal set of conformations can be chosen in order to reduce computational costs and also the number of false positives in pose prediction. In order to generate an efficient ensemble of CDK2 X-ray structures, a robust graph-based selection algorithm is proposed, using which, 126 non-redundant CDK2 structures are selected in the ensemble dataset. A diverse set of ligands extracted from ChEMBL, and docked to the non-redundant receptor ensemble. A feature set of 512 features including feature energetics of docking results, beside other eight simple molecular features of ligands, are considered in the final dataset for the machine learning (ensemble-based) affinity prediction method. The use of machine learning eliminates the need of using classical scoring functions such as force-field, knowledge-based and empirical function, which are prone to limitations with increase in training data size. In this study, Random Forest (RF) and Boosted Regression Trees (BRT) ensemble learning algorithms are used for final affinity prediction. Finally, the impurity importance value of RF method is used in order to choose CDK2 structures which play a more important role in ensemble docking. Experiments show that docking to only those receptors selected by RF, reduces the error and also error skewness. Finally, using the mentioned methods, a $MSE_{RF} = 1.3, Rp_{RF} = 0.5$ for RF and $MSE_{BRT} = 1.37, Rp_{BRT} = 0.52$ for BRT is obtained (hyperparameters set to the default values and models iterated 50 times). By letting machine learning select important features, an accuracy of 1kcal/mol is achieved, which is significantly better than methods not based on machine learning.

Key Words: *Ensemble docking, Ensemble learning, Random forest, Boosted Regression Tree, CDK2*

Classification of Autistic Patients and Control Via Utilizing Dictionary of Functional Modes as Brain Atlas

Karim Rahimian[1], Mohsen Khodarahmi[2], Parimah Emadi Safavi[1], Seyed Shahriar Arab[1], Javad Zahiri[3]*

[1] *Biophysics Department, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*
[2] *Department of Radiology, Shahid Madani Hospital, Karaj, Iran*
[3] *Department of Neuroscience, University of California, San Diego, La Jolla, CA, United States.*

E-mail address (corresponding author): jzahiri@health.ucsd.edu

Abstract

Autism spectrum disorder is a set of neurodevelopmental disorders difficult to diagnose especially in childhood when the symptoms emerge. The early diagnosis of the disease can prevent many complications in patients in future years. Lacking suitable and applicable biomarkers and indices for the diagnosis, on the one hand, the similarity of these disorders to other neuropsychiatric diseases in another hand, made the early diagnosis almost overbearing. Functional Magnetic Resonance Imaging is one of the methods to diagnose neurological disorders by detecting changes associated with blood flow in the brain. Images from this method are high dimensional data which makes the training process of machine learning models challenging. One of the solutions is to extract essential features from images to ease the process. To this end, features from these images can be extracted using brain atlases for regions of interest. We used Dictionary of Functional Modes as an atlas for feature extraction followed by training a logistic regression model to classify images of autistic individuals and control cases. We leveraged datasets from the Autism Brain Imaging Data Exchange database containing images from autism and control individuals. The dataset has many variants due to variations in data collection from different imaging centers. Since these take-ups could obscure the accuracy of our machine learning model, we implicated the Combat method to remove these unwanted side-effects. Exploiting these methods and the atlas for feature extraction resulted in a significant increase in accuracy of our logistic regression classifier (71%) which is more optimized than previous methods such as asd-diagnet neural network. Our method can be applicable to classify other neurological disorders with regard to the brain region of interest. This is heartwarming for the future of precision medicine since it has the ability to investigate potential biomarkers of such complex disorders.

Key Words: *Machine Learning, Autism, Brain Atlas, DIFUMO, Classification*

Big Data in Biology

In silico transcriptome analysis of drought and salt involved responsiveness genes in Brassica napus

Mohadeseh Dousti[1]*, Leila Mazhari[1], Zarrin Minuchehr[2], Tahmineh Lohrasebi[1], Khadijeh Razavi[1]

[1] *Department of Plant Bioproducts, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran*
[2] *Department of Industrial and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran*

E-mail address (corresponding author): mohadese.doosti666@gmail.com

Abstract

Canola (Brassica sp.) as the most important oily seed product in the world is affected largely by salinity and drought stresses due to planting in arid and semi-arid regions. So, studying about impressive genes in salt/drought stresses in canola would be helpful to improve tolerant species. In this study genes that involved in salt and drought stresses in Brassica napus were investigated from integrating gene expression omnibus (GEO) and sequence read archive (SRA) databases in different time line since 2021. Then results were analyzed by web based software GEO2R and GALAXY for detect differentially expressed genes (DEGs). DEGs in short, mid and long-term of salinity and drought stresses were obtained from extensive meta-analysis (RRA method). Subsequently ontology of acquired robustDEGs were done by blast2go. By constructing PPI network with Cytoscape software, hub genes in each time line were identified. Among all 34 A. thaliana hub genes, HAI2 and DREB1B were selected to validate with real time qPCR in tolerant (Okapi) and sensitive (RGS) cultivars of canola. The hub known and novel genes that identified through our meta-analysis, provide insight to understanding the molecular response to salinity/drought stresses and engineering abiotic stress tolerance in canola.

Key Words: *Brassica napus, salinity, drought, meta analysis, candidate genes*

ApInAPDB: A database of apoptosis-inducing anticancer peptides

Naser Faraji[1´2], Alireza Doustmohammadi[3], Seyed Shahriar Arab[3]*, Norelle L. Daly[4], Ahmad Yari Khosroushahi[1´2]*

[1] *Department of Medical Nanotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz, Iran*
[3] *Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*
[4] *Centre for Molecular Therapeutics, Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Queensland [487]0, Australia*

E-mail address (corresponding author): sh.arab@modares.ac.ir

Abstract

ApInAPDB (http://bioinf.modares.ac.ir/software/ApInAPDB) is manually-collected database of 775 apoptosis-inducing anticancer peptides accumulated from published research articles and databases like DRAMP. ApInAPDB include 748 linear and 27 cyclic peptides having L-amino acids, D-amino acids or both of them. The current version of the database provides information related to a peptide including source, N-terminal modification, C-terminal modification, non-natural amino acid presence, binding target, function, IC50 of peptide, cell lines, and finally binding affinity and etc. Additionally, several physicochemical properties such as isoelectric point, net charge at pH 7, hydrophobicity and GRAVY among others and also 1559 descriptors for QSAR modeling are calculated for almost all peptides. Besides, secondary structure is predicted using Chou-Fasman, GOR, Neural Network algorithms and joint of them. Some user-friendly searching tools are designed to facilitate browsing and achieving the desired results including simple research, advanced research and top categories. Overall, ApInAPDB is the first database provides comprehensive and useful information of apoptosis-inducing anticancer peptides which can be helpful for researchers interested in cancer therapy and peptide design.

Key Words: *peptide; database; apoptosis; anticancer*

RNA-sequencing of CD4+ T cells in Relapsing-Remitting Multiple Sclerosis patients at relapse; deciphering the involvement of novel genes and pathways

Zahra Salehi[1], Saeed Talebi[2,3], Samaneh Maleknia[4], Fahimeh Palizban[5], Abdorreza Naser Moghadasi[6], Kaveh Kavousi[5], Mohammad Ali Sahraian[6]*, Maryam Izad[1,6]*

[1] Immunology Department, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[2] Department of Medical Genetics and Molecular Biology, Faculty of Medicine, Iran University of Medical Sciences, Tehran, Iran.
[3] Department of Pediatrics, Faculty of Medicine, Ali Asghar Children Hospital, Iran University of Medical Sciences, Tehran, Iran.
[4] Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[5] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[6] MS Research Center, Neuroscience Institute, Tehran University of Medical Sciences, Tehran, Iran.
E-mail address (corresponding author): izadm@sina.tums.ac.ir

Abstract

CD4+ T cells known as a noteworthy potential modulator of inflammation in multiple sclerosis (MS). In the current study, we investigated the transcriptome profile of CD4+ T cells in relapsing-remitting MS (RRMS) patients at relapse phase. We performed RNA sequencing of CD4+ T cells isolated from RRMS patients at relapse phase and age- and sex-matched healthy controls. The edgeR statistical method was employed to determine differential expression genes (DEGs). The gene set enrichment analysis was subsequently performed. Applying physical interaction network, genes with higher degrees were selected as hub genes. A total of 1278 genes and 1034 gens were defined at significantly higher or lower levels, respectively, in CD4+ T cells of RRMS patients at relapse phase as compared with healthy controls. The top up- and down-regulated gene were JAML and KDM3A. The detected DEGs were remarkably on chromosomes 1 and 2, respectively. The DEGs were mainly enriched in pathways such as 'regulation of transcription, DNA-templated', 'regulation of B cell receptor signaling pathway', 'protein phosphorylation', 'epidermal growth factor receptor signaling pathway', and 'positive regulation of neurogenesis'. Moreover, 16 KEGG pathways mostly associated with the immune system and viral infections were enriched. In the constructed physical interaction networks, UBA52 and TP53 were illustrated as the most highly ranked hub genes among up- and down-regulated genes, correspondingly.
Conclusions: By applying global transcriptome profiling of CD4+ T cells, we deciphers the involvement of several novel genes and pathways in MS pathogenesis. The present results need to be affirmed by in vivo and in vitro studies.

Key Words: *RRMS; CD4+ T cells; RNA-sequencing; Transcriptome; Functional modules; Chromosomal enrichment*

Biological Sequence Analysis

Exploration of Plastic Contaminated Soil Metagenome to Identify Novel Plastic Degradation Enzymes

Donya Afshar Jahanshahi[1´2], Kaveh Kavousi [2]*, Shohreh Ariaeenejad [3]

[1] *Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran*
[2] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[3] *Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research Education and Extension Organization (AREEO), Karaj, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Today, all kinds of plastics are widely used in the global economy for various purposes and 350-400 milion tons are produced every year. It has been shown that plastics cause adverse effects in all ecosystems and microplastics make particular concerns for the health of humans and other living things on the planet earth. Existing enzymes act mainly on high molecular weight polymers such as polyethylene (PE), polyethylene terephthalate (PET) and ester-based polyurethane (PUR). Various types of enzymes have been introduced as plastic digrader, including Laccase, Cutinase, PETase and Peroxidase. Laccases have been broadly applied as a multitasking biocatalyst in various industries. Also this enzymes are the most important enzymes that were introduced to degrade plastics. In this study, plastic contaminated soil samples are analyzed using metagenomics approaches including taxonomic profiling, assembly, binning, gene prediction, enzyme prediction, and 3D structure prediction to discover new microorganisms and enzymes in effective plastic degradation. Generated bins are analyzed using ECpred software and unbinned samples are analyzed using Metarenz software to predict laccase enzymes. Metarenz is a tool for identification target enzymes from assembled contigs. 3D structure of 86 Laccase candidates are predicted using Alphafold and TMalighn tools. The resulted metagenomic sequences are further investigated by NCBI CDD. Finally, sequences passed from all above stages will enter the cloning process and their degradability will be investigated in vitro.

Key Words: *Computational Metagenomics; Plastics; Degradation; Laccase; Soil;*

Identification of peripheral blood mononuclear cell gene signatures for detection of hepatocellular carcinoma

Sara Fayazzadeh[1], Anna Meyfour[2]*, Javad Zahiri[3]*

[1] *Bioinformatics and Computational Omics Lab(BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*
[2] *Basic and Molecular Epidemiology of Gastrointestinal Disorders Research Center, Research Center for Gastroenterology and Liver Diseases, Shahid Beheshti University, Tehran, Iran*
[3] *Department of Neurosciences, University of California, San Diego, CA, USA*

E-mail address (corresponding author): a.meyfour@sbmu.ac.ir

Abstract

Liver cancer is ranked sixth among the most frequently occurring cancers for both sexes worldwide in 2020 and the third leading cause of cancer deaths following lung and colorectal cancers. Hepatocellular carcinoma (HCC) is liver cancer's most common primary malignancy and accounts for ~90% of cases. Currently available diagnostic methods for HCC exhibit a relatively moderate sensitivity of 60%, therefore introducing gene signatures for the disease diagnosis seems necessary. The gene expression profile of peripheral blood mononuclear cells (PBMCs) has shown an alteration in the case of malignancies such as different types of cancer. In this study, we aim to detect the gene signatures of PBMCs that could distinguish HCC cases from healthy controls by analyzing multiple datasets retrieved from GEO database. Differentially expressed genes (DEGs) were determined using Limma package in R-4.1.0. A total of 11 genes, including 9 upregulated and 2 downregulated, were determined as DEGs based on the defined criteria. DAVID online tool was then utilized for enrichment analysis of DEGs. Several signaling pathways such as Toll-like receptor, NOD-like receptor, TNF signaling pathways, and pathways in cancer were enriched. These genes were also able to classify HCC from healthy controls with an accuracy of 93%. The results demonstrate PBMCs contain promising gene signatures that could shed light on the mechanism of the disease, leading to the introduction of a non-invasive diagnostic panel and/or therapeutic targets.

Key Words: *Hepatocellular carcinoma; Cancer; Microarray, Differential expression, DEG, Gene signature*

Development of a new oligonucleotide block location-based feature extraction (BLBFE) method for the classification of riboswitches

F. Golabi[1]*, M. Shamsi[2], M. H. Sedaaghi[3], A. Barzegar[4], M. S. Hejazi[5,6]

[1] *Department of Biomedical Engineering, Faculty of Advanced Biomedical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Genomic Signal Processing Laboratory, Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran*
[3] *Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran*
[4] *Department of Medical Biotechnology, Faculty of Advanced Biomedical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*
[5] *Molecular Medicine Research Center, Biomedicine Institute, Tabriz University of Medical Sciences, Tabriz, Iran*
[6] *Department of Pharmaceutical Biotechnology, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran*
E-mail address (corresponding author): faegheh.golabi@gmail.com

Abstract

As knowledge of genetics and genome elements increases, the demand for the development of bioinformatics tools for analyzing these data are raised. Riboswitches are genetic components, usually located in the untranslated regions of mRNAs, that regulate gene expression. Additionally, their interaction with antibiotics has been recently suggested, implying a role in antibiotic effects and resistance. Following a previously published sequential block finding algorithm, herein, we report the development of a new block location-based feature extraction strategy (BLBFE). This procedure utilizes the locations of family-specific sequential blocks on riboswitch sequences as features. Furthermore, the performance of other feature extraction strategies, including mono- and dinucleotide frequencies, k-mer, DAC, DCC, DACC, PC-PseDNC-General and SC-PseDNC-General methods, was investigated. KNN , LDA , naïve Bayes , PNN and decision tree classifiers accompanied by V-fold cross-validation were applied for all methods of feature extraction, and their performances based on the defined feature extraction strategies were compared. Performance measures of accuracy, sensitivity, specificity and F-score for each method of feature extraction were studied. The proposed feature extraction strategy resulted in a classification of riboswitches with an average correct classification rate (CCR) of 90.8%. Furthermore, the obtained data confirmed the performance of the developed feature extraction method with an average accuracy of 96.1%, an average sensitivity of 90.8%, an average specificity of 97.52% and an average F-score of 90.69%. Our results implied that the proposed feature extraction (BLBFE) method can classify and discriminate riboswitch families with high CCR, accuracy, sensitivity, specificity and F-score values.

Key Words: *Riboswitches; feature extraction; sequential blocks; block location-based feature extraction; classification; performance measures.*

Investigating the obesity paradox in patients with Hepatocellular carcinomas using bioinformatics approaches

Nazanin Hosseinkhan[1]*, Zahra Narimani[2], Laily Najafi[1], Mohammad Ebrahim Khamseh[1]

[1] *Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran*
[2] *Institute of advanced studies in Basic Sciences (IASBS), Zanjan, Iran*

E-mail address (corresponding author): hosseinkhan@ut.ac.ir

Abstract

Introduction: Obesity is a heritable multifactorial disease resulted from the interplay of environmental, genetics, epigenetics, and metagenomics factors. There is a common belief that higher body mass index (BMI) is associated with reduced cancer survival. However, several studies have challenged this by demonstrating better prognosis in overweight/obese patients, a phenomenon known as "obesity paradox". In this study, we aimed to compare the genetic composition of hepatocellular carcinoma (HCC) patients with/without overweight-obesity using whole genome sequencing data.
Methods: Mutation annotation files were retrieved from TCGA-LIHC project using TCGABiolinks library. In total, 158 normal weights and 161 over weights/obese HCC patients were selected for downstream analyses. Ensembl variant effect prediction (VEP) tool was used to further annotate the identified common and exclusive mutations in both BMI categories. In parallel, in order to find co-occurred mutations and their frequencies, we employed Apriori algorithm.
Result: Pathogenic cancer associated mutations in TP53, RYR2, CTNNB1, and FBN2 were detected in both BMI categories. Despite the presence of large number of exclusive mutations in the two BMI categories, they target common cancer associated pathways including TGF-beta, Notch, JAK-STAT, Wnt, Ras, MAPK, PI3K-Akt, and p53. In addition, co-occurred mutations were observed in only 4 overweight/obese. In addition, the number of deaths in normal weights carrying pathogenic mutations was higher than those with overweight/obesity (52 Vs. 45).
Discussion: Normal weight and obese patients with HCC represent exclusive genetic mutations. However, the signaling pathways are similar. Furthermore, co-occurred mutations were not common in overweight/obese patients. Considering high number of deaths in HCC patients with normal weight, we can conclude that overweight/obesity can protect HCC patients from worse outcomes. However, more investigations considering central obesity measures are required to have correct patient categories.

Key Words: *Hepatocellular carcinoma, Obesity paradox, Mutational analysis*

Transcriptome profiling analysis of Suaeda salsa shoots under salinity stress

Shima Jamalirad[1´2], Mohammad Reza Azimi[1], Nayer Azam Khoshkholgh Sima[3], Mehrshad Zeinalabedini[2], Laleh Karimi Farsad[2], Ghasem Hosseini Salekdeh[2´4], Mohammad Reza Ghaffari[2]*

[1] Department of Plant Breeding and Biotechnology (PBB), Faculty of Agriculture, University of Zanjan, Zanjan, Iran
[2] Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education, and Extension Organization (AREEO), Karaj, Iran
[3] Department of Molecular Plant Physiology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education, and Extension Organization (AREEO), Karaj, Iran
[4] Department of Molecular Sciences, Macquarie University, North Ryde, NSW, Australia

E-mail address (corresponding author): shimajamalirad@yahoo.com

Abstract

Salinity is one of the consequential abiotic stresses that significantly impact agricultural yield worldwide. euhalophyte Suaeda salsa is a model plant that may help identify salt tolerance molecular mechanisms to develop salt-resistant crops. In the present study, high-throughput RNA sequencing was conducted after the Suaeda salsa seedlings were exposed to 0 mM, 200 mM, 400 mM, and 800 mM NaCl for 30 days. The comparative analysis of transcripts identified 610, 1893, and 1691 genes significantly expressed at 200mM, 400mM, and 800 mM NaCl, compared to control, respectively. Gene Ontology (GO) and MapMan analyses revealed that the differentially expressed genes were related to solute transport and nutrient uptake, protein synthesis, modification, hemostasis, transcriptional regulation, and Phytohormones action. Moreover, we identified multiple genes associated with different transcription factor families, including 143 DEGs, which may regulate salt stress-response in S. salsa. Finally, the DEGs in the salt response pathway were verified by Quantitative real-time PCR (qRT-PCR).

Key Words: *Salinity, Euhalophyte, Suaeda salsa, High-throughput RNA sequencing*

A new reconstruction of mouse metebolic model using orthology- based approach

Saeideh Khodaee[1], Yazdan Asgari[2], Mehdi Totonchi[3,4], Mohammad Hossein Karimi-Jafari[1]*

[1] *Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[2] *Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran*
[3] *Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran*
[4] *Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran*

E-mail address (corresponding author): mhkarimijafari@ut.ac.ir

Abstract

A genome-scale metabolic (GSM) model is an in silco metabolic model including chemical reactions that take place inside a cell, tissue, or organism. In a GSM model, gene- protein reaction rules represent the mapping between gene encoding enzymes and the reaction they catalyze. These rules are a fundamental component in orthology-based model to reconstruct a new GSM model of a target organism through a GSM model of a reference organism.
In this study, we reconstructed a new mouse GSM model (iMM1865) from human Recon3D model. For this purpose, we used homologous gene mapping and extracting mouse specific reactions through literature and databases. iMM1865 has been validated using 431 metabolic objective functions which were created to verify Recon3 model. This model with no dead-end metabolites and blocked reactions is more comprehensive than previous reported mouse models. Also it has passed more metabolic objective functions tests. Furthermore, to evaluate of the predictive ability of the model, gene essentiality simulations was performed.

Key Words: *genome-scale metabolic model, orthology-based reconstruction, Mouse metabolic model, iMM1865, Recon3D*

EARN as a precision oncology tool leads us to propose the targeted genes panel for metastatic breast cancer

Leila Mirsadeghi[1]*, Kaveh Kavousi[1], and Ali Mohammad Banaei-Moghaddam[2]

[1] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[2] Laboratory of Genomics and Epigenomics (LGE), Department of Biochemistry, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

E-mail address (corresponding author): l.mirsadeghi@gmail.com

Abstract

Today, there are a lot of bio-markers on the prognosis and diagnosis of complex diseases such as primary breast cancer. However, our understanding of the drivers that influence cancer aggression is limited. Present investigation studies somatic mutation data consisting of 450 metastatic breast tumor samples from cBio Cancer Genomics Portal. We use four software tools to extract features from this data. Then, an ensemble classifier learning algorithm called EARN (Ensemble of Artificial Neural Network, Random Forest, and non-linear Support Vector Machine) is proposed to evaluate plausible driver genes for metastatic breast cancer (MBCA). It is an attempt to focus on the findings in four aspects of MBCA prognosis. First, drivers and passengers predicted by SVM, ANN, RF, and EARN are introduced. Second, the performance of four learning methods is evaluated using statistical criteria. Third, the outputs of the biological inference based on gene set enrichment analysis (GSEA) and pathway enrichment analysis (PEA) are discussed. Finally, the PEA using ReactomeFIVIz tool (FDR<0.03) for the top 100 predicted genes by EARN leads us to propose a new gene set panel for MBCA, including HDAC3, ABAT, GRIN1, PLCB1, and KPNA2 as well as NCOR1, TBL1XR1, SIRT4, KRAS, CACNA1E, PRKCG, GPS2, SIN3A, ACTB, KDM6B, and PRMT1. Furthermore, we compare results for MBCA to other outputs regarding 983 primary breast invasive carcinoma (BRCA) tumor samples obtained from The Cancer Genome Atlas (TCGA). Meanwhile, the 16-gene panel proposed by EARN has been surveyed in the whole-exome sequence of an archived FFPE sample obtained from breast tissue of an anonymous Iranian female patient with invasive breast carcinoma. This research leverages both computational and experimental approaches to assist precision oncologists to design compact targeted panels that eliminate the need for whole-genome/exome sequencing.

Key Words: *Metastasis breast tumor; Mutation data; Ensemble classifier; Plausible driver genes; Targeted gene panel*

Computational Drug Design and Discovery

A Novel Method for Predicting Drug Synergy Based on Matrix Factorization

Fatemeh Abbasi[1], Sadjad Gharaghani[1]*

[1] *Laboratory of Bioinformatics and Drug Design (LBD), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): s.gharaghani@ut.ac.ir

Abstract

Combination therapy has proven to be highly effective in the treatment of complex diseases such as cancer and infectious diseases. When compared to monotherapy, drug combination therapy can improve cancer treatment efficacy, reduce drug dose-dependent toxicity, and prevent drug resistance. In this study, we present a novel regression method based on matrix factorization. Using K-fold nested cross-validation, we compare the results of the presented method to the results of two novel regression methods, PRODeepSyn and DeepSynergy, on the DrugComb database. DrugComb collects data from four studies: I the O'Neil dataset, (ii) the Forcina dataset, (iii) the NCI Almanac dataset, and (iv) the CLOUD dataset. DeepSynergy is a feed forward neural network that converts sample input vectors into a single output value known as the synergy score. DeepSynergy used chemical information derived from drugs as well as genomic data pertaining to disease biology. PRODeepSyn is a deep neural network that predicts synergy scores based on cell line embeddings and drug features using Batch Normalization. PRODeepSyn constructs the feature vector for each drug using the molecular fingerprint and descriptors to represent the structural and physicochemical properties of drugs, and for cell line features they integrate three types of heterogeneous cell line features containing gene expression data, gene mutation data, and interactions between gene expression products to construct cell line embeddings. To compare these methods, the mean square error (MSE), root mean square error (RMSE), and Pearson correlation coefficient (PCC) between predictions and ground truth are used as primary evaluation metrics. The results show that the presented matrix factorization method outperformed the PRODeepSyn and DeepSynergy methods across the platform.

Key Words: *drug synergy; deep learning; graph convolutional network; feed forward neural network*

Core proteome mining of Salmonella enterica subsp. enterica to explore novel therapeutic targets and design a multi-epitope vaccine

Yamini Chand[1], Sachidanand Singh[2,3]

[1] *Faculty of Biotechnology, Institute of Biosciences and Technology, Shri Ramswaroop Memorial University, Lucknow-Deva Road, Barabanki,Uttar Pradesh, India.*
[2] *Faculty of Biotechnology, Institute of Biosciences and Technology, Shri Ramswaroop Memorial University, Lucknow-Deva Road, Barabanki, Uttar Pradesh, India*
[3] *Department of Biotechnology, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, 522213, Andhra Pradesh, India.*

E-mail address (corresponding author): yamini.chand4@gmail.com

Abstract

Salmonella enterica subsp. enterica (S. enterica) infects humans and animals, causing enteric and systemic salmonellosis. Antibiotics are no longer effective against S. enterica infection due to the advent of multidrug-resistant (MDR) strains. Additionally, typhoid vaccines are not currently advised for children under the age of six, and their long-term efficacy is questionable. In the present study, we employed a bioinformatics pipeline of subtractive proteomics and immunoinformatics to explore novel drug and vaccine targets and construct a multi-epitope vaccine against S. enterica. A total of 1964 protein-coding sequences were retrieved from the EDGAR database that comprised the core genome of 185 S. enterica strains. The core genome was screened for non-paralogous proteins employing CD-HIT tool which resulted in 1949 protein sequences. Subtractive proteomics prioritized twenty-seven cytoplasmic and six outer membrane proteins (OMPs) as potential drug and vaccine targets, respectively based on their non-homology to humans, presence of essential, virulent and resistant factors, interaction to host (Homo sapiens), and role in pathogen-specific metabolic pathways. The OMPs with low molecular weight, transmembrane helices, and high antigenicity were subjected to immunoinformatics analysis to identify cytotoxic T lymphocyte (CTL) epitopes, helper T lymphocyte (HTL) epitopes, and linear B lymphocyte (LBL) epitopes using various tools and servers. The highly antigenic, non-allergenic and non-toxic epitopes were selected to design four different vaccine constructs (VC1, VC2, VC3, and VC4) using linkers and immune-modulating adjuvants. The vaccine construct (VC4) was selected based on its physicochemical properties, allergenicity, antigenicity, and non-toxicity potential. The molecular docking and molecular dynamics (MD) simulation analyses ensured stable molecular interaction of the final prioritized vaccine construct with human immune cell receptors (TLR4). Finally, the vaccine construct (VC4) was cloned in silico to ensure its effectiveness. We conclude that the designed vaccine construct may provide immunological protection against S. enterica infections.

Key Words: *Cytotoxic T lymphocytes (CTL) epitopes; Helper T lymphocytes (HTL) epitopes; Immunoinformatics; In silico cloning; Linear B lymphocyte (LBL) epitopes; Molecular docking; Molecular dynamics (MD) simulation; Multidrug-resistance (MDR); Multiepitope-based subunit vaccine; Outer membrane proteins (OMPs); Subtractive proteomics.*

Indicator Regularized Non-Negative Matrix Factorization Vs. a Novel Combinatorial Heuristic Matrix Factorization: a Comparison of Matrix Factorization Methods as the Building Block of Drug Repurposing

Arash Zabihian[1], Mohsen Hooshmand[2]*, Sajjad Gharaghani[1]*

[1] *Laboratory of Bioinformatics and Drug Design (LBD), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[2] *Department of Computer Science and Information Technology,Institute for Advanced Studies in Basic Sciences, Zanjan, Iran*

E-mail address (corresponding author): mohsen.hooshmand@iasbs.ac.ir

Abstract

Drug repurposing is one of the ponderable computational methods suggested for rapid developing of a drug. This approach can be formulated as a matrix factorization (MF) method. First, a MF method is applied to different types of drug-target relation and similarity matrices. Then, the unseen data will be analysed to evalutae the predictive power of the model. Some types of data in bioinformatics are binary, e.g. binary representation of impacts of antiviral drugs on viruses. MF of such data is more complicated than the conventional numerical values. Most of conventional algorithms for binary matrix factorization use gradient-based methods. These methods utilize relaxation approach to solve continuous binary problems. The relaxation approach turns the binary constraint into a box constraint. Despite the widespread use, these methods do not perform well on sparse data sets, and also unable to return a proper approximation of problem. To avoid facing these limitations, we propose a binary matrix factorization method utilizing combinatorial optimization and modular arithmetic. In our prposal, binary values are used in each step and the results come from modular multiplication. We compare our results with Indicator Regularized Non-Negative Matrix Factorization (IRNMF) which is a gradient based method. Both methods are appled to the same antiviral-viruse interaction matrix and Five-fold cross-validation (CV) are used to report the performance. The results indicate that better performance and lower error value are the advantages of the suggested method in compare with IRNMF.

Key Words: *Matrix Factorization; Drug repurposing; Antiviral Drugs; Heuristic Factorization*

Target prediction for the inhibitors of VEGFR2 as anti-colorectal cancer compounds using similarity-based search methods

Samira Shafiee[1]*, Andreas Bender[2], Mostafa Zakariazadeh[3], David Schaller[4], Somaieh Soltani[1]

[1] *Pharmacy Faculty, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Cambridge, UK*
[3] *Department of Biology, Faculty of Sciences, Payame Noor University, Tehran, Iran*
[4] *Pharmaceutical and Medicinal Chemistry, Freie Universitat Berlin, Konigin-Luise-Str, Berlin, Germany*

E-mail address (corresponding author): basira6785@gmail.com

Abstract

Colorectal cancer is among the diseases that little is known about its etiology. Like other cancers, colorectal cancer is the end result of a multifactorial, multi genetic, and multistage process. VEGFR2 has been indicated as a key factor of angiogenesis in many cancers as well as colorectal cancer. Its inhibition is under investigation as anti-angiogenesis cancer therapy. To investigate the multi-targeting characteristic of VEGFR2 inhibitors, we utilized a rational approach combined with similarity-based search methods to the prediction of targets from a list of colorectal cancer-related drug targets. Data mining workflow was developed using KNIME. The data was extracted from the ChEMBL25 database. The Morgan fingerprints were calculated by RDKit and ChemAxon. Tanimoto similarity index was calculated and the clustering of the inhibitors was done using a procedure implemented by KNIME. Proper activity thresholds and Tanimoto index were optimized based on the obtained results and the method was validated using the available evidence for the probable inhibition of the predicted targets from literature. Compounds were considered similar if the Tanimoto score was > 0.7. Using the developed method, CDK2, HER1, and TGF-β2 were predicted for the investigated VEGFR2 inhibitors. CDK2 was predicted for AT-9283, as a multi-targeted VEGFR2 inhibitor using the developed method, there is published evidence for the inhibition of CDK2 by AT-9283 which approves the prediction capability of the developed method. In addition, we studied the interaction of AT-9283 with VGEFR2 and CDK2 using molecular docking and the obtained binding energy values were -9.20 Kcal/mol and -9.60 Kcal/mol, respectively. The results from the molecular docking study were in good agreement with reported experiments. According to the results, the developed method could be used for the target prediction with high reliability.

Key Words: *Colorectal cancer; Multi-target; KNIME platform; Molecular docking; VGEFR2*

Inflammatory Target prediction for the FDA-approved anticancer drugs using morgan fingerprint similarity-based methods

Somaieh Soltani[1]*, Gerhard Wolber, David Schaller, Andrease bender

[1] *Department of Medicinal Chemistry, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran*

E-mail address (corresponding author): somaieh.s@gmail.com

Abstract

As a result of complicated interactions between pharmacodynamic, pharmacokinetic, genetic, epigenetic, and environmental factors, most of the available drugs or multitarget therapies poses polypharmacological effects. Cancer as a disease with multiple reasons, has been one of the main mortalities causes and drug discovery for it has been one of the most interested subjects during the years. One of the most well-known causes for cancer is inflammation and anti-inflammatory effect of available anticancer drugs could be regarded as a mechanism for their efficacy and potency. In this study we developed predictive models based on the morgan fingerprint similarity search for anticancer drugs with focus on inflammatory target A list of FDA approved anticancer drugs were generated and their fingerprint were calculated using RDKit python module. The tanimoto index was calculated using the obtained fingerprints and the targets were predicted based on the Chembl25 target prediction mudule. The modified python code was utilized to predict the target profile for the investigated compounds, while the target list was limited to the inflammatory targets. The results indicated an interesting profile of anti-inflammatory predicted targets for the available FDA-approved anticancer drugs. The obtained results were clustered due to the synergistic, antagonistic and neutral target profiles for the investigated compounds. Litreture survey indicated the availabaility of experimental and clinical evidences for most of the predicted targets, while some targets were not reported previously. The results showed that the developed rational method combined with similarity search method could be used to the target prediction for FDA approved drugs.

Key Words: *Polypharmacology, Cancer, Morgan fingerprint, Chembl25, Target prediction*

Modeling in Computational Biology

Investigation of common genes in different stages of non-alcoholic fatty liver disease with microarray datasets analysis

Masoud Arabfard[1], Yazdan Hasani Nourian[1], Raheleh Halabian[2], B. Fatemeh Nobakht M. Gh.[1]*

[1] *Chemical Injuries Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran*
[2] *Applied Microbiology Research Center, Systems biology and poisonings institute, Baqiyatallah University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): fatemeh.nobakht@yahoo.com

Abstract

Background: Non-alcoholic fatty liver disease (NAFLD) is the most common type of chronic liver disease worldwide and is a risk factor for developing cirrhosis or hepatocellular carcinoma (HCC) if untreated. NAFLD affects around 90% and 25% of obese patients and people worldwide, respectively. Although lots of efforts have been performed by experts in order to find the underlying mechanisms of NAFLD, it remains a challenge to recognize them. The aim of this study is to distinguish common gene signatures and pathways in the human liver during NAFLD progression through the systems biology method.
Method: In this study, the three microarray datasets, GSE48452, GSE63067, and GSE89632, were selected from the NCBI GEO database to explore differentially expressed genes (DEGs) among healthy controls, simple steatosis, and non-alcoholic steatohepatitis (NASH) patients. Furthermore, protein-protein interaction (PPI) networks and pathway enrichment analysis were used to detect common genes and biological pathways in different stages of NAFLD.
Results: The current was included 47 healthy subjects, 36 patients with simple steatosis and, 46 NASH patients. Common high degree genes among all three sets were CHI3L1, GFBP2, NRG1, PEG10, and FADS2. The top five genes in the hepatic PPI networks of three datasets were STAT3, JUN, CANX, FN1, and MYC. Signal transduction, immune response, and anti-apoptosis were the most important biological pathways between healthy vs. NASH, while cell communication, signal transduction, and immune response were the three top biological pathways between healthy vs. simple steatosis. Also, the most eminent biological pathways between NASH vs. simple steatosis were metabolism and energy pathways.
Conclusion: The present study represented the unique and shared key genes and pathways between different stages of NAFLD, which may facilitate the understanding of NAFLD mechanism and identify potential therapeutic targets in this disease.

Key Words: *Microarray, Systems biology, Protein-protein interaction network, NAFLD, NASH*

Casilico: A versatile CRISPR package for in silico CRISPR RNA designing

Adnan Asadbeygi[1], Milad Norouzi[2], Mohammad Sadegh Vafaei Sadi[2], Saffari Mojtaba[3], Mohammad Reza Bakhtiarizadeh[1]*

[1] *Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran*
[2] *Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[3] *Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences (TUMS), Tehran, Iran*

E-mail address (corresponding author): mrbakhtiari@ut.ac.ir

Abstract

Background: The efficiency of CRISPR-Cas system is highly depends on well-designed CRISPR RNA (crRNA). To facilitate the use of various types of CRISPR-Cas systems, there is a need for development of computational tools to design crRNAs which cover different CRISPR-Cas systems with off-target analysis capability. Numerous crRNA design tools have been developed but nearly all of them are dedicated to design crRNA for genome editing. Hence, we developed a tool matching the needs of both beginners and experts, named Casilico, which was inspired by the limitations of the current gRNA design tools for designing crRNAs for Cas12, Cas13, and Cas14 CRISPR-Cas systems.
Materials and Methods: Using a list of important features such as mismatch tolerance rules, self-complementarity, GC content, frequency of cleaving base around the target site, target accessibility and protospacer flanking site or protospacer adjacent motif requirement, Casilico searches all potential crRNAs in a user-input sequence. Considering these features, help users to rank all crRNAs for a sequence and make an informed decision about whether a crRNA is suited for an experiment or not. Our tool is sufficiently flexible to tune some key parameters governing the design of crRNA and identification of off-targets, led to increases the chances of successful CRISPR-Cas experiments.
Results, and Conclusion: Casilico outperforms previous crRNA design tools in the following respects: 1) supporting any reference genome/transcriptome for which a FASTA file is available; 2) designing crRNAs that simultaneously target multiple sequences through conserved region detection among a set of sequences; 3) considering new CRISPR-Cas subtypes; 4) reporting a list of different features for each candidate crRNA, which can help the user to select the best one. Casilico was successfully applied to design crRNAs for different genes in SARS-CoV-2 genome, as some of the crRNAs have been experimentally tested in the previous studies.

Key Words: *Cas13, Cas12, Cas14, Guide RNA, CRISPR Cas*

Molecular dynamics simulations show structural insights into the N-terminal domain mutations of the spike protein in the Omicron (B.1.1.529) variant of SARS-CoV-2.

Fatemeh Bayani[1], Negin Safaei Hashkavaei[1], Yahya Sefidbakht[1]*

[1] *Protein Research Center at Shahid Beheshti University, Tehran, Iran*

E-mail address (corresponding author): y_sefidbakht@sbu.ac.ir

Abstract

Since the outbreak of SARS-CoV-2, many variants with different mutations in the structure of the virus have spread rapidly global. One of the new Variant of Concern (VOC) with high infectivity is Omicron (B.1.1.529), which was first reported in November 2021 from South Africa. Since Spike glycoprotein is a major driving force in virus infectivity, it is important to study the effects of mutations in understanding the structural changes of the new variants. Careful evaluation of the spike glycoprotein sequence in omicron reveals numerous point and deletion mutations as well as an additional insertion. The N-terminal domain (NTD) of spike glycoprotein is an effective factor in cell surface adhesion and plays an important role in the virus escaping the immune system. Therefore, the study of the effect of mutations through molecular dynamics simulations shows changes in the structure of NTD and its stability. these analyzes, helps researchers to achieve a relative understanding of the high infectivity and antigenicity of the new variant in the compared to the Wuhan-Hu-1 type.

Key Words: *SARS-CoV-2; Omicron (B.1.1.529); Spike glycoprotein; N-terminal domain (NTD); Molecular dynamics simulations.*

Inferring microbial communities using constrained damped lasso regression based on the generalized Lotka-Volterra model

Naser Elmi[1], Ahmad Kalhor[2], Kaveh Kavosui[1]*

[1] *Complex Biological systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[2] *Human and Robot Interaction Laboratory, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

In this study, we developed a new method to infer microbial communities. Synthetic and natural microbial communities play essential roles in the industry and our health. We used the generalized Lotka-Volterra (GLV) model to model the interaction between microbes. Due to the meaningfulness of the parameters in this model, it has been widely used for modeling microbial communities. To the best of our knowledge, most of the regression-based methods on the GLV equation did not consider the sparsity and constraints of the real problem. Hence, to solve the limitations of the available methods, we developed damped lasso regularization to solve this constrained-based convex optimization problem. We used CVX solver for this problem. We trained and tested our method on various simulated microbial communities with 3 to 5 interacting microbes with different dynamics, including stable fixed point, limit cycle, and chaotic dynamics. We used the cross-validation method to test our method's performance in inferring the magnitude and sign of the interactions. We calculated the correlation of the estimated abundance of interacting microbes based on the inferred model with actual data. Our results demonstrated that the developed method could accurately predict the parameters sign and magnitude. Furthermore, the correlation between the estimated abundance and real abundance was more than 0.9. We also evaluated the model performance in presence of noise.

Key Words: *microbial community, inference, lasso regression, GLV, Optimization*

Novel insights in targeted therapy of cancer by mathematical modeling of cancer immunoediting

Mojtaba Ghanizadeh[1]*, Seyed Peyman Shariatpanahi[1], Bahram goliaei[1]

[1] *Laboratory of Biophysics and Molecular Biology, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): moghaniz@ut.ac.ir

Abstract

Computational modeling has advanced in recent decades and provides new opportunities for a better understanding of immune system-cancer dynamics. Immunoediting is the process describing this dynamic and discusses immune system-cancer coevolution during tumorigenesis. It consists of three main phases; Elimination (elimination of cancer cells by the immune system), dormancy phase (balance between tumor growth and immune response), and escape (escape from immune response and growth of tumor). Here we modeled the immunoediting, using a system of two ordinary differential equations to study the fate of the tumor through three main phases of cancer immunoediting. Using the model, we evaluated the role of the tumor growth rate, tumor immunogenic properties, and immune cells intratumoral penetration on the tumor fate through coevolution with the immune system. The results show that, interestingly, immunogenic tumors can survive elimination and instead, become dormant by decreasing the growth rate and/or increasing accessibility of the tumor for the immune system; Dormant tumors may wake up later on, causing tumor escape and growth. The model also indicated that the immune response and tumor growth synchronization has a significant role for tumor eradication. In other series of simulations, we evaluated the fate of the residual tumor cells after surgical tumor resection. The results revealed the importance of timing in adjuvant targeted therapy of tumor residue: a too late or a too early start of targeted therapy may result in tumor escape or tumor dormancy, respectively, while starting at an intermediate timing results in the complete elimination of the tumor. Results of our model simulation are corroborated with experimental and clinical observations.

Key Words: *Immunoediting; Immune system-cancer synchronization; Mathematical modeling; Targeted therapy*

Negative Binomial Mixed Models for Identifying Oncogenic dependencies through analysis of RNAi Screening data

Zohreh Toghrayee[1]*, Sajjad Gharaghani[2], Hesam Montazeri[2]

[1] *Department of Bioinformatics, Kish International Campus, University of Tehran, Kish, Iran*
[2] *Department of Bioinformatics, Institute Biochemistry and Biophysics, University of Tehran, Iran*

E-mail address (corresponding author): zohreh_toghrayee@yahoo.com

Abstract

Background: Loss-of-function RNAi screening has been extensively used to identify cancer dependencies, including oncogene addiction. A few computational methods such as ATARiS, DEMETER, and RSA have been presented to compute gene-level scores by handling off-target effects in RNAi screening data. However, these methods often result in low statistical power in low sample-size settings. This paper presents a new statistical approach to tackle the off-target effects in order to provide higher statistical power compared to the mentioned methods.
Methods: We applied DRIVE project data and the CCLE data to detect gene drivers in pan-cancer and breast cancer, by thoroughly scrutinizing all shRNA data. In the proposed method, we first removed the effect of batches including pool and thermodynamic stability of shRNAs using an empirical Bayesian model available in the SVA package in R. Then negative binomial mixed effect models were performed on ranks of these logFC in each cell line.
Results: Among 6919 genes, known cancer genes such as KRAS, NRAS, BRAF, PIK3CA, CTNNB1, TP53, and CDK4 were reassuringly identified by the proposed method in Pan-cancer analysis. We demonstrated that the proposed approach outperformed ATARiS and DEMETER in terms of statistical power through sub-sampling approaches. In analyzing breast cancer data, we identified both putative oncogenes such as PAX5 and RASGRP2 and known oncogenes such as KRAS.
Conclusion: By using all information in RNAi screening data, the proposed method models on- and off-target effects and can identify oncogene addictions in cancer.

Key Words: *RNAi screening, Off-target effect, DRIVE Project, Batch effect removal*

Structural Bioinformatics

How does a bacteriophage enzybiotic target bacteria? Introducing a structural model of bacteriophage PhaxI lytic enzyme

Mehri Javid[1], Zargham Sepehrizadeh[1], Ahmad Reza Shahverdi[1], Azadeh Ebrahim-Habibi[2,3]*

[1] Department of Pharmaceutical Biotechnology and Pharmaceutical Biotechnology Research Center, School of Pharmacy, Tehran University of Medical Sciences, Tehran, Iran
[2] Biosensor Research Center, Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran
[3] Endocrinology and Metabolism Research Center, Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran

E-mail address (corresponding author): aehabibi@sina.tums.ac.ir, azadehabibi@yahoo.fr

Abstract

Spurred by the emergence and prevalence of multidrug resistance of human pathogens, there has been a global call for novel classes of antibacterials over the last decade. Phages as natural predators of bacteria respond to this call and have been considered a promising candidate since their discovery. Their lytic enzymes serve for cell wall degradation of the infected bacteria. The natural function of these enzymes, also called 'enzybiotics,' can be exploited as powerful enzyme-based antibiotics. These insights have paved the road towards the (pre)clinical development of enzybiotics in various trials phases. Due to poor structural and biochemical characterization, their significant potential as antimicrobials remains unexplored. Verifying structural features and enzymatic specificity is crucial for prospective enzyme modification and further application. Hence, this study aimed to shed light on the 3D structure, determination of essential amino acid residues, and assessment of interactions mode with the bacterial cell wall of a novel PhaxI bacteriophage enzybiotic. A computationally aided method determined that this enzybiotic has two domains: The structure was predicted using the YASARA program and dihedral angles were evaluated by Ramachandran plot with 96.7% residues in the most favored regions. After generating a tautomers library of NAG-NAM-NAG molecules as a piece of the bacterial cell wall, virtual screening was done with Autodock Vina to determine the best substrate-enzyme interaction. The stability and binding affinity of the complex were studied using AMBER 14 forcefield for molecular dynamics simulations for 150 ns (3 × 50 ns) at physiological pH (7.4). The predicted enzyme tunnel and essential residues were further confirmed using Dali server and evolutionary structural relationship studies. These results can subsequently lead to discovering new activities and modes of action of these enzymes to develop potent and more efficient broad-spectrum antibacterials.

Key Words: *Antibiotic resistant; phage lytic enzymes; enzybiotics; homology modeling; molecular dynamic simulation; evolutionary relationships*

A diffusion kernel-based approach for protein domain identification

Amirali Zandieh[1], Mohammad Taheri-Ledari[2], Seyed Peyman Shriatpanahi[1]*, Changiz Eslahchi[3,4]

[1] *Department of Biophysics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[2] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[3] *Department of Computer and Data Sciences, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran*
[4] *School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*

E-mail address (corresponding author): ch-eslahchi@sbu.ac.ir

Abstract

It is almost half a century since the concept of protein domain, as compact and recurring units that are able to fold and function independently, was introduced. Nevertheless, the inherent ambiguity of the definition besides the increasing number of newly solved structures keeps the accurate automated methods in high demand. Contrary to the majority of the state-of-the-art methods, we employed enhanced measures of proximity between amino acids rather than developing context-specific clustering algorithms. Here, the power of kernel functions to separate structural domains in their corresponding Hilbert spaces is investigated. For this purpose, utilizing four different diffusion kernels on protein graphs, a novel pipeline for protein domain assignment is developed. The result of the presented method on commonly used benchmark data sets shows a marginally better performance compared to the best available methods based on two different metrics. Moreover, by offering alternative partitionings, our method answers the problem of subjectivity in protein domain definition. The high prediction accuracy of the approach reveals the diffusion kernels' potential to split entangled structures of complex proteins. In addition to out-competing other methods by merely employing general (rather than context-specific) clustering algorithms, our pipeline provides the versatility to implement other graph node kernels that can potentially boost its performance.

Key Words: *Protein structure; Graph node kernel; Protein domain assignment; Clustering; Diffusion kernel*

Systems Biology

Using Liquid Association Analysis to detect controller genes involved in pituitary non-functioning adenoma invasiveness

Nasibeh khayer[1], Mehdi Mirzaie[2], Maryam Jalessi[1´3]

[1] Skull Base Research Center, The Five Senses Health Institute, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
[2] Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran
[3] ENT and Head & Neck Research Center and Department, Hazrat Rasoul Hospital, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

E-mail address (corresponding author): khayer.n@iums.ac.ir

Abstract

Nowadays, considerable disease-related high-throughput "omics" datasets are freely available. Such datasets contain valuable information about disease-related pathways and their corresponding gene interactions. Currently, knowledge of the non-functioning pituitary adenoma (NFPAs) invasion at the molecular level is not sufficient. The present study aimed to identify critical biomarkers and biological pathways associated with invasiveness in the NFPAs using a three-way interaction model. This model can detect the dynamic nature of the co-expression relationship of two genes ({X1, X2}) by introducing a third gene (X3), which is sometimes referred to as the controller gene. Indeed, the expression level of the controller gene modulates the correlation between X1 and X2. One of the statistical methods for this purpose is liquid association analysis.
This study used the Liquid association method to capture the statistically significant triplets involved in NFPAs invasiveness. Random Forest analysis was applied to select the most critical controller genes. Finally, gene set enrichment and gene regulatory network analyses were applied to detect the biological relevance of the statistically significant triplets. This study suggests Nkx3-1 and Fech as two controller genes that might be critical in the invasiveness behavior of NFAPs. Moreover, the "mRNA processing" and "spindle organization" pathways are suggested as two crucial pathways involved in the NFAPs' invasiveness.

Key Words: *Non-functioning pituitary adenomas, Invasiveness, fast liquid association, Random forest classification, Gene set enrichment analysis.*

Non-coding and Transposable elements discovery through artificial intelligence approaches

Alexandre Rossi Paschoal

E-mail address (corresponding author): paschoal@utfpr.edu.br

Abstract

We have learned that only a tiny portion of the eukaryotic genomes are translated into proteins. In the last years, significant efforts have been devoted to differentiating pervasive transcription from functional Non-coding RNAs (ncRNAs) transcripts. In the same field, the transposable element is one of the major mechanisms to promote variability in the genome. In this talk, we will discuss how we could use and apply Big Data, Artificial Intelligence, and Data Science to Non-coding and TE problems. In particular about: (i) large-scale data integration and analysis; (i) the relationship between ncRNAs and transposable elements; and (iii) a machine learning and deep learning tool for long RNAs and TE.

Key Words: *Non-coding RNA*

Simultaneous inference of cell-line-specific gene regulatory networks and mode-of-action of drugs from drug-induced gene expression measurements

Mohammad Taheri-Ledari[1], Amirali Zandieh[2], Kaveh Kavousi[1]*, Sayed-Amir Marashi[3]

[1] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[2] Department of Biophysics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[3] Department of Biotechnology, College of Science, University of Tehran, Tehran, Iran

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Personalized molecular networks help in elucidating drug resistance mechanisms, finding new therapeutic targets, and predicting the effectiveness of combination therapy. In this study a method for simultaneously inferring gene regulatory network and mode-of-action of drugs from gene expression measurements of drug-perturbed cancer cell-lines is proposed. An inferred model is a dynamic state-space representation whose variables are genes and inputs are drug dosages. Such a model includes gene-gene and drug-gene (mode-of-action) relationships. To implement the proposed method, drug-induced expression levels of nearly 1000 landmark genes for a set of cell-lines in the LINCS database were employed. Since no time-series gene expression data are available the cell-lines are assumed to be in steady-state. The central challenge of the present study is under-determinedness for which a set of solutions are examined. Finally, the performance of the proposed method is evaluated by performing a cross-validation on the set of drugs for a specific cell-line.

Key Words: *Dynamical system, State-space model, Network inference, Gene regulatory network, Drug mode-of-action, Under-determinedness, Cancer cell-line*

Poster

AI & Machine Learning in Biology

Machine Learning and Feature Selection to SEER Data to Novel Diagnosis Thyroid Cancer

Ali Abedini[1]*, Shirin Malehmir[2]

[1] *Department of Bioinformatics, Segal biotechnology, Tehran, Iran*
[2] *Department of Microbiology, Karaj Branch, Islamic azad university, Karaj, Iran*

E-mail address (corresponding author): ali.abedini.biotech.67@gmail.com

Abstract

In this research to create a machine learning prediction model that can be used to predict bone metastasis in thyroid cancer. Demographic and clinicopathologic variables of thyroid cancer patients in the Surveillance, Epidemiology, and End Results database from 2010 to 2016 were retrospectively analyzed. On this basis, we developed a random forest algorithm model based on machine-learning. The area under receiver operating characteristic curve (AUC), accuracy score, recall rate, and specificity are used to evaluate and compare the prediction performance of the random forestmodel and the other model. A total of 17,138 patients were included in the study, with 166 (0.97%) developed bone metastases. Grade, T stage, histology, race, sex, age, and N stage were the important prediction features of bone metastasis. The random forestmodel has better predictive performance than the other model (AUC: 0.917, accuracy: 0.904, recall rate: 0.833, and specificity: 0.905). The random forestmodel constructed in this study could accurately predict bone metastases in thyroid cancer patients, which may provide clinicians with more personalized clinical decision-making recommendations. In conclusion, here, we developed a random forest prediction model for bone metastases in thyroid cancer patients that outperformed traditional logistic regression models. This facilitates personalized diagnosis and refined clinical decision making for bone metastasis in thyroid cancer patients.

Key Words: *bone metastasis, machine learning, random forest, SEER.*

The Application of Feature weighting models for Identification of key genes associated with the Transcriptomic Response to Drought Stress in Populus Species

Sahar Akrami[1]*, Ahmad Tahmasebi[1], Ali Niazi[1]

[1] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*

E-mail address (corresponding author): akrami.sahar.19@gmail.com

Abstract

Introduction: Poplar varieties are planted in short rotation coppice, and are supposed to show high biomass production. Drought is a very important abiotic stressor. High throughput gene expression technologies provide valuable information about transcriptome. Feature weighting models are also known as attractive strategies to gain new biological insights. At the transcriptome level, the algorithms for identifying key signatures related to environmental stress have not been applied in Populus. In this study, we used the large transcriptome data to gain comprehensive view of drought stress response in Populus.
Method: The array expression datasets retrieved from GEO and ArrayExpress. RMA algorithm was used for background correction and normalization of gene expression data by Affy R package. Finally, an empirical Bayes method was performed to correct non-biological differences and remove batch effects from gene expression datasets using ComBat function in the SVA Rpackage. Feature selection algorithms were employed to reduce the dimensionality of expression dataset and identify the gene expression features. We implemented various attribute weighting algorithms include SVM, Chi Squared, Information Gain, Information Gain Ratio, Deviation, Gini Index, Uncertainty, Relief, and PCA to identify the most important genes using RapidMiner Studio software.
Result: In total 13 microarray datasets consisting of 324 arrays were considered. After pre-processing and removing the batch effect, the normalized datasets were obtained for further downstream analysis. In total, 648 genes were identified as the most important features by at least one of the models. Functional annotation showed that the feature genes were enriched in response to abiotic stimulus and MAPK signaling pathway. In addition, a lot of genes were related to secondary metabolic process. Interestingly, the seven methods selected auxin response factor 2-like and PYL4-like as important features.
Conclusion: Our analysis suggests that ARF2-like and PYL4-like genes can be potential candidates for screening and breeding purposes in Populus.

Key Words: *Populus Species; Feature weighting models; SVM; PCA*

Performance evaluation of different machine learning classification models on expression profiles of tumor educated platelets data

Sajedeh Bahonar[1]*, Fahimeh Palizban[1], Hesam Montazeri[1]

[1] *Department of Bioinformatics, Institute Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): sajedeh.bahonar@ut.ac.ir

Abstract

Since liquid biopsy is less invasive than tissue biopsy, studies on liquid biopsy biomarkers for the early detection of cancer are taken into consideration. Expression profiles of tumor-educated platelets (TEP) in liquid biopsy can be used as one of the biomarkers. Using classification machine learning models, given the feature space derived from the expression data of TEPs, has given us the ability to predict data categories. Here, the aim is performance evaluation of different classification models for diagnosis of cancer-based on expression profiles of platelets. First, expression profiles of TEPs in 230 patients with breast, liver, colorectal, brain, pancreatic, and lung cancers in addition to profiles of 55 healthy individuals were downloaded from the GEO database (GSE68086). Thereafter, the data were normalized using the edgeR package (R software version 4.1.0) and 2000 genes with the highest variance were selected. Then, different types of classification models namely SVM, LDA, logistic regression, boosting, classification tree, and random forest, were evaluated on the feature selected data in 10-fold cross-validation. In addition, the variable importance of selected genes was obtained using polynomial SVM. Then, pathway enrichment analysis was performed using H, C6, and C7 gene sets of MSigDB database using preranked GSEA method. The results showed that the polynomial SVM has the highest performance on the validation set (accuracy ~ 95%, mean AUC ~ 0.994, sd AUC ~ 0.0093). Also, the linear SVM model had the second-best performance on validation set (mean AUC ~ 0.9917). In pathway enrichment analysis 10 immunological pathways were enriched in cancer samples compared to healthy donors. Overall, the results showed that polynomial SVM can be a model with good performance for classifying TEP data. All in all, the results of this study indicate that the expression profile of TEPs can be considered as a candidate biomarker in liquid biopsy.

Key Words: *Tumor educated platelets (TEP), classification models, cancer, pathway enrichment analysis*

Prediction of immunogenic peptides derived from FVIII by machine learning approach

Mahsa Bahramimoghadam[1], Mohammadali Mazloumi[2], Gholam Ali Kardar[2], Ali Mohammad Banaei-Moghaddam[3], Kaveh Kavosui[1]*

[1] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[2] *Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran*
[3]*Laboratory of Genomics and Epigenomics (LGE), Department of Biochemistry, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Various factors are involved in the development of an immune response to factor eight (FVIII) in hemophilia patients, among which T cell epitopes (factor eight-derived peptides) play the most important role. Mutations in T cell epitopes while maintaining the structure of factor eight can be a good solution to this unwanted immune response. The need for this research is due to the fact that Treatment of patients in whose body exogenous factor eight exhibits an immune response (inhibitor patients) It is much more difficult than hemophiliacs who do not have this problem. Also, the treatment methods and strategies that exist today have a lot of costs and complications which are not very successful. In this study, we build a model using machine learning algorithms to predict the immunogenicity of immunogenic peptide sequences. We first used compositional features to predict the peptides that bind to class II molecules of the Major Histocompatibility Complex (MHCII). Including: AAC, APAAC, CKSAAGP, CTDC, CTDT, DPC, GDPC and PAAC. We then evaluated these features with some classifiers such as Random Forrest, Support Vector Machine, Decision Tree, Naive Bayes Classifier, XGBoost, and Perceptron, and the accuracy for each of these classifiers was, 0.62,0.51,0.35,0.54,0.58,0.66 respectively. In the next step some of the best features were selected. The accuracy of the classifiers including Random Forest (0.51), Support Vector Machine (0.59), Decision Tree (0.44), Naive Bayes (0.58), XGBoost (0.51) and Perceptron (0.46) were not good enough. The data that used in this method cover all types of human HLA-DR. Also, the features used were the most up-to-date features related to peptide-MHCII Binding. We hope to achieve higher accuracy by enhancing them.

Key Words: *T cell epitopes, immunogenic peptides, prediction, machine learning, FVIII*

Mutation prediction of Infectious viruses based on Different Machine learning approaches

Khashayar Ehteshami[1], Mohammad Hadi Azarabad[2], Paria Pashootan[2], Negar Khalili[2], Fereshteh Fallah Atanaki[2], Amirali ghafourian Ghahramani[1], Kaveh Kavousi*[2]

[1] *Department of Computer Engineering, School of Science and Engineering, Sharif University of Technology International campus-kish island, Kish, Iran*
[2] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

In general, the ability to predict the evolution of a pathogen enhances our ability to control, prevent, and treat diseases. Usually, only mutations that can escape the host immune system and affect the severity can sustain and spread throughout generations. Several different pandemics have happened through the years. For example, the 1918 influenza pandemic was one of the most severe in recent history, and the H1N1 virus caused it. By mutation prediction, pandemics can be recognized before they happen. Variational AutoEncoders (VAEs) and Generative Adversarial Networks (GANs) generate new samples from our data in machine learning. Sequence-to-Sequence (Seq2seq) networks are primarily used in translation tasks to generate a new sample from the previous one. The method that we use is a combination of GAN networks and sequence to sequence networks. We create a Seq2seq network as a Generator of our model with the help of Long Short Term Memories (LSTMs) and then use a discriminator to distinguish whether the sequences are fake or real. The most challenging task is that GANs are not good in sequential data, and Seq2seq has some problems in the long length of sequences. For the result, we find out which sequence is more possible for the mutant in the future, and we can use these results for preventing a future pandemic.

Key Words: *Machine Learning; GANs; Deep Learning; Sequence Generating*

Prediction of protein aggregation tendency based on the support vector machine algorithm

Fatemeh Eshari[1]*, Amirreza Farajnezhadi[1], Fahime Momeni[1], Mehran Habibi-Rezaei[1]

[1] *Protein Biotechnology Research Lab (PBRL), School of Biology, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): fatemeh.eshari@guest.ut.ac.ir

Abstract

Protein aggregation plays an important role in various diseases, for instance, type 2 diabetes (T2D), Alzheimer's disease (AD), Parkinson's disease (PD), prion encephalopathies, and Huntington's disease. Moreover, it has been recognized as a field with increasing importance in the biopharmaceutical industries because of its occurrence during bioprocessing steps to ensure the drug's effectiveness and decrease associated risks, such as increased immunogenicity. Therefore, aggregation prediction of proteins under different conditions has great importance for successful biopharmaceutics' development and theranostic approaches. In order to predict the aggregation propensity of proteins, a machine learning method was proposed to evaluate the aggregation propensity of hexapeptides of WALTZ-DB 2.0 databank, using the Support Vector Machine (SVM) algorithm based on the sequences of segments and beta-sheet formation propensity of residues as an intrinsic feature. To analyze the capability of the proposed method, two parameters were considered, which are F-measure and Matthews Correlation Coefficient (MCC), owing to their evaluative power. Finally, the applied approach was compared with the Pasta 2.0 server that uses similar inputs to make predictions. The mentioned parameters of the proposed method were resulted to be 0.830 and 0.633 for the proposed method, and 0.688 and 0.382 for the Pasta2 server, respectively. As a result, the new suggested strategy superiorly evaluates the aggregation propensity, which is essential for the basic and applied approaches.

Key Words: *Protein aggregation; Bioproducts engineering; Machine learning; SVM; Pasta 2.0*

Predicting the membrane proteins' classification using multi-dimensional wavelet and random forest classifier

Parham Hajishafiezahramini[1], Dr. Parviz Abdolmaleki[1]*

[1] *Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Jalal AleAhmad,Nasr, Tehran, Iran*

E-mail address (corresponding author): parviz@modares.ac.ir

Abstract

Concerning the difficulties and complexity of experimental methods to determine the functionality and structure of the proteins, the computational techniques have recently found their proper place in predicting protein function problems. While different techniques have been introduced based on the machine learning approach, there is no combination technique of exploiting Multidimensional discrete wavelet transform(DWT) analysis and machine learning. In this study, we have devised a handy, accurate, and time-efficient predictive model to classify the membrane proteins into five different classes, including single-pass type 1, single-pass type 2, Multi-Pass, Lipid-Chain, and GPI membrane proteins based on DWT analysis and machine learning approach.

We have applied our proposed method for Chou's membrane protein datasets, containing 2059 and 2625 membrane protein sequences from five different classes. The majority of the former studies used these datasets as the complete ones. In this technique, protein sequences were initially transformed into six-dimensional signals, including the hydropathy scale, polarity, secondary structure, molecular volume, codon diversity, and electrostatic charge indexes. These six-dimensional signals are then used as the multidimensional discrete wavelet transform input data to analyze the entire signals. Feature vectors were then generated regarding the proper criteria of approximate and detailed coefficients for every single protein. ventually, the feature vectors were used in a random forest classifier to avoid overfitting and take advantage of measuring variable importance.

As a result, we obtained an accuracy of 91.7% and 89.6% for the independent dataset and jackknife test, respectively. These results indicated that the proposed method yielded better results.

Key Words: *Membrane proteins; Predictive model; Discrete Wavelet transform; Hydropathy scale*

Functional annotation of Missense Mutations Based on Protein Features

Motahareh Hakiminejad[1]*, Hesam Montazeri[2], Bahram Goliaei[1]

[1] *Department of Biophysics, Institute Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[2] *Department of Bioinformatics, Institute Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): mot.hakimi.95@gmail.com

Abstract

Background: Among all genetic alterations in cancers, Single Nucleotide Variant (SNV) are the most common mutation. Still, identifying cancer-driving SNV (driver mutation) among the plenty of non-driving ones (passenger mutations) remains a challenge due to the innate bias in their population and the fact that most driver mutations are rare. We present a random forest tool that can annotate driver/passenger missense mutations based on protein information and helps find novel driving mutations.
Materials and Methods: Pan-cancer mutation data from the TCGA database (n = 600k) fetched and labeled passenger/driver (based on their prevalence). According to the mutations, a feature-set containing five categories was built: 1) Physio-chemical changes of the changed amino acid, 2) Changes in Pseudo-amino acid composition of the 21-mer sequence around the point of mutation 3) disorderness of the mutation region, 4) site of mutation reported region/functions in uniport 5) whether the gene is reported to be Oncogene/ Tumor Suppressor or none. A random forest model was trained on the feature set by the ranger package in R.
Results: The accuracy of the method on test data is 99% (sensitivity = 99%, specificity = 54%). The method was evaluated against other cancer missense annotations such as CHASMplus, CHASM, Mutation Assessor, Polyphen2, and VEST on experimentally-labeled cancer missense mutations. The receiver operating characteristic curve (auROC) of methods were 88%, 67%, 67%, 59%, 72%, respectively, and our method auROC was 83%. Also, it was tested against cancer SNV Golden standard based on extensive literature and database review, in which the accuracy was reported to be 72%, (sensitivity = 74%, specificity = 71%)
Conclusion: We developed a random forest method that discriminates drivers from passenger missense mutations. As the method is solely based on protein descriptors, it can give insight into the mutation mode of action.

Key Words: *Protein structure/function, cancer-type-specific driver, missense mutation, rare drivers*

Identification of molecular features and pharmacophores for selective inhibition of cyclin-dependent kinases: Application of counterpropagation artificial neural networks

Sara Kaveh[1], Marzieh Sadat Neiband[2], Ahmad Mani-Varnosfaderani[1]*

[1] *Department of Chemistry, Tarbiat Modares University, Tehran, Iran*
[2] *Department of Chemistry, Payame Noor University (PNU), Tehran, Iran*

E-mail address (corresponding author): a.mani@modares.ac.ir

Abstract

Protein kinases are one of the largest enzyme families, consisting of 2% of the translated human genome. Kinase inhibitors have enormous medical use for treating mortal diseases such as cancer and Parkinson. Due to the wide range of activities and functions for all kinases, there is a vital need for the development of selective kinase inhibitors for targeting only some specific types of them for effective treatments. The main aim of this project is to find isoform-selective pharmacophores and molecular features for different types of Kinase receptors. In order to achieve this goal, a total of 4201 drug-like molecules with recorded inhibition activity for the inhibition of CDK1, CDK2, CDK4, CDK5, and CDK9, were collected from Binding-Database. The variable importance in projection (VIP) method was used to select the molecular descriptors from the 3224 descriptor pool calculated via DRAGON 5.5 software. The dataset was divided into training (70%) and test (30%) sets, randomly. Counter propagation artificial neural network (CPANN) and supervised Kohonen networks (SKN), were used for the classification of the molecules. Some general parameters such as mean square distance index, number of Pyrazoles atoms, hydrophobicity, aromaticity index, and number of hydroxyl groups were found to be important parameters for describing the inhibition behavior of CDK's inhibitors. Generally, the performances of classification models were evaluated according to the statistical parameters derived from the confusion matrices. The classification rates range from 82 % to 79% for the training and validation procedure for the optimized CPANN models. The high accuracy values of the obtained classifiers for the training and test sets demonstrate that the information provided is reliable for describing and predicting the activity of CDK inhibitors. The reliable statistical values of the classified models can be applied by researchers in the pharmaceutical sciences whom aim to design selective kinase inhibitors.

Key Words: *Classification; Kinase; Selective drug design; structure-activity relationship; Virtual screening*

Prediction of Disease-Causing Genes in Breast Cancer by Graph Mining in Biological Networks

Alireza Meshkin[1]*, Alireza Molaei[1], Khosro Goudarzi[1]

[1] *Department of Computer Engineering, Islamic Azad University, Damavand Branch, Damavand, Iran*

E-mail address (corresponding author): meshkin@damavandiau.ac.ir

Abstract

Traditional studies in breast cancer, do not locate exactly the casual breast cancer genes in the genome and they often detect a region containing many candidates' genes. Gene prioritization problem tries to rank the candidate genes from most to least promising. It can lead to faster discovery of novel casual genes and can lead to better diagnostic accuracy and treatment in breast cancer.

Despite many advances in medical science and biology, many people die each year from breast cancer. This shows that science still has a long way to go to cure this cancer. Breast cancer is a very complex and deadly disease. Many of the genes associated with the disease are not yet known, for example, known genes in breast cancer, such as BRCA1 and BRCA2, account for only 5% of breast cancer cases. In this study, protein network data source and network-based algorithms were evaluated and then Network propagation algorithm applied to prioritize candidate genes for breast cancer.

The results showed that protein networks have a significant impact on the quality of the gene prioritization approach. This approach outperformed previously published algorithms (e.g., DIR, and ENDEAVOUR) in evaluation metrices such as AUC, average rank, and TOP 5% metrics.

Key Words: *Candidate Gene Prioritization, Biological Complex Network, Graph mining, Breast Cancer, Network Propagation.*

A greedy time-frequency analysis of electrodermal activity for cognitive state monitoring

Rezvan Mirzaeian[1], Peyvand Ghaderyan[2]*

[1] *Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran*
[2] *Computational Neuroscience Laboratory, Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran*

E-mail address (corresponding author): p_ghaderyan@sut.ac.ir

Abstract

Cognitive workload (CW) is defined as the required mental effort to perform a task, which is associated with capacity-limited cognitive system for processing information in the working memory. The development of a framework to not only minimize human errors, but also maximize performance, is an important application of CW estimation and management. It can also help to improve the diagnosis and treatment of neurological or cognitive disorders, and to facilitate human-computer interaction. In place of non-invasive nature, sensitivity to cognitive states, robustness against intentional conduct, and the possibility of online investigation, psychophysiological signal analysis have received special attention for CW estimation. Hence, a novel CW estimation method based on matching pursuit (MP) decomposition of electrodermal activity (EDA) and support vector machine classifier has been proposed. The MP, as adaptive and greedy time-frequency decomposition, has the advantages of reducing cross terms, improving time-frequency resolution and enhancing biomedical signal analysis performance [5, 6]. Applying two dictionaries, including RnIdent and daubechies wavelet (db5) at level 5, sparse coefficients have been extracted from the EDA signals. Then, several statistical and nonlinear features (mean, standard deviation, variance, covariance, and Shannon entropy) have been calculated from the MP coefficients. The proposed method evaluated on EDA of 30 healthy students performing an arithmetic task has achieved an average accuracy of 96.80% for two workload levels. Moreover, the experimental results have indicated that the combination of complementary information from the different extracted features has enhanced the estimation performance.

Key Words: *Matching Pursuit, Cognitive workload, Support vector machine, arithmetic task*

TM-Bench: A benchmark dataset for thermophilic-mesophilic proteins classification

Saber Mohammadi[1], Danial Khadivi[1], Javad Zahiri[2,3], Seyed Shahriar Arab[1]

[1] *Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*
[2] *Department of Neuroscience, University of California San Diego, California, USA*
[3] *Department of Pediatrics, University of California San Diego, La Jolla, CA, USA*

E-mail address (corresponding author): sabermohammdi.1994@gmail.com

Abstract

Recently, machine learning approaches have become conventional methods in order to solve biological problems. Thermal stability of thermophilic and hyper-thermophilic proteins has made them suitable candidates for medical and industrial applications. Thus, various machine learning methods have been introduced to predict the thermophilic proteins and discriminate them from their mesophilic counterparts based on the sequence information of these proteins. Most of these studies have reported accuracies of more than 90 percent, whereas it seems to be optimistic. Using an inappropriate dataset can be the main source of this overestimation. Hence, comparing the various approaches has become challenging due to the lack of a gold standard dataset. Here we introduce TM-Bench dataset. Zhang and Fang made an effort for the first time to discriminate thermophilic and mesophilic proteins via pattern recognition methods. Since then, a variety of approaches such as SVM, artificial neural network, decision tree, k-nearest neighbor, genetic algorithm, and Naive Bayes have been adopted for the classification of thermophilic and non-thermophilic proteins solely based on proteins sequence information. In this study, we used the BacDive database in order to extract a list of thermophilic and mesophilic organisms based on their optimum growth temperature. Next, after having extracted the corresponding protein sequences from Swiss-Prot database, redundancies in the primary dataset were removed by the CD-HIT tool. Subsequently, the balanced and imbalanced datasets were fed to the above-mentioned methods for re-evaluating their performance. Our results indicate that sensitivity, specificity, and accuracy were lower than previously reported measures for balanced data set, and with imbalanced data set, sensitivity drops dramatically. Overall, Multi-Layer Perceptron and Logit Boost showed better performance than other methods with the balanced dataset, with 81% and 78% accuracy, the sensitivity of 82% and 79%, and specificity of 80% and 77%, respectively.

Key Words: *Machine learning, thermophilic protein, mesophilic protein, neural network, SVM*

The efficiency of artificial neural network (ANN) for diagnosis of obesity and hypertension

Maryam Moradi[1]*, Anfal Shamsa[1]

[1] *Laboratory Science, Medipol University, Istanbul, Turky*

E-mail address (corresponding author): moradimaryam553@gmail.com

Abstract

Obesity and hypertension are health problems in any society. The aim of this study was to evaluate the sensitivity, specificity and accuracy of artificial neural network (ANN) for the diagnosis of obesity and hypertension. For this study, demographic information about 550 students aged 7-18 years was recorded in the ANN program. The recorded demographic information consisted of 11 input variables and 3 output variables. Input variables included age, sex, weight, height, waist circumference, body mass index, waist-to-height ratio, abdominal obesity, physical activity, genetics, and unhealthy eating behaviors, while output variables included obesity, systolic blood pressure, and diastolic blood pressure. In this study, Levenberg-Marquardt and Conjugate Gradient algorithms were used to training the network. The results showed that the selected neural network with Levenberg-Marquardt algorithm had 17 hidden neurons in the diagnosis of obesity and high diastolic blood pressure, while in the diagnosis of high systolic blood pressure it had 15 hidden neurons. Based on the results of the study, the sensitivity, specificity and accuracy of the network in the diagnosis of diastolic blood pressure were 0.8123, 0.9915 and 0.9713, respectively. While these values were 0.9672, 0.9962 and 0.9818 for obesity and 0.8559, 0.9912 and 0.9843 for systolic blood pressure, respectively. Based on the results of the present study, it can be concluded that ANN designed to diagnose obesity, systolic and diastolic blood pressure with equal accuracy of 96%, 85% and 81%, respectively. Therefore, it can be said that ANN program has high efficiency in diagnosing obesity and hypertension.

Key Words: *Artificial Neural Network; Health; Obesity; Hypertension, Efficiency; Iran.*

Similarity detection between modern human genome and their ancestors DNA sequences by Deep Learning

Keivan Naseri[1]*, Mahboobeh Golchinpour[2]

[1] *Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran*
[2] *Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): keivan.naseri2017@ut.ac.ir

Abstract

Neanderthals were a species of human that lived in Europe and parts of western Asia, Central Asia, and northern China (Altai). The first signs of early Neanderthals date back to about 350,000 years ago in Europe. There is ample genetic evidence that modern humans had sex with Neanderthals, Denisovans, and other ancient relatives.
In this study, we used in-depth learning to identify areas of Neanderthal intrusion in the modern human genome. Recent methods, such as the Markov latent model (HMM) to find the Neanderthal effect on the genome, are a memoryless model that does not consider the relationship between nucleotide distances along DNA sequences. Therefore, we used deep learning power to process crude genomic sequences and nucleotide long-term memory in genomes with short-term long-term memory (LSTM).
This model works better than linear models such as support vector machines (SVMs) or simple Bayesian classifiers, so we recommend the LSTM method for analyzing ancient biological data.
We first converted DNA sequences into k-mers with limited space. We then used the Bag Of Words model to compare k-mers frequencies between sequences inherited from Neanderthals and sequences from weak ancient ancestors. Finally, when classifying sequences, we learned Word Embeddings with a sequential model with the Keras Embeddings layer. The model achieved an accuracy of 87.6% in the data set that classifies the input Neanderthal sequences against the discharged source.
It should be noted that for the near future, our vision is to find similarities between modern humans and their ancestors in the genomic data of skin patients using the LSTM model.

Key Words: *Neanderthal genome; DNA-Sequencing; Deep Learning; SVM; LSTM*

Diagnosis of COVID-19 patients based on chest CT images using image processing algorithms

Ahmadreza Rasouli, Mohsen Khodarahmi, Javad Zahiri

E-mail address (corresponding author): ahmadrezarasouli7377@gmail.com

Abstract

The rapid diagnosis of the disease often makes the treatment procedure faster and less expensive. In the case of SARS-COV2, an infectious disease with a high rate if transmission, diagnosis without the need to see a doctor is of great importance.

In this study, we aim to introduce an algorithm with a high performamce based on which we can diagnose a patient based on the chest CT imaging features, and without any need for the suspected patient to see a doctor. Firstly, we enhanced the quality of CT images by using the classic algorithms and the most important filters for image processing. Then thr most important CT imaging feature were extracted using a convolutional neural network. We have used two imaging sets including 42 images from patients and 42 images from healthy persons. For each of CT features we assigned a certain weight.

Finally, we desgined an algorithm which gets a CT image from an individual as input, and determines whether this individual is healthy or patient, by enhancing the quality of the initial image, extracting the relevant imaging features by using a convolutional neural network, and adding the multiplication of each feature and its associated weight.

In this study, by examining 1288 photos of healthy people and 1343 photos of sick people, we reached about 90% accuracy in diagnosing the disease.

Key Words: *Corona-Image processing-Neural Network-Feature extraction-Preprocessing*

Evaluation of COVID-19 mutations and predicting the rate of disease transmission and pathogenicity based on the types of mutations.

Tina Sadeh

E-mail address (corresponding author): tinasadeh93@gmail.com

Abstract

Researchers have developed a new method that uses Artificial Intelligence to foresee the most likely mutations of pathogens like SARS-COV-2, the virus that causes COVID-19. SARS-CoV-2, a novel coronavirus mostly known as COVID-19 has created a global pandemic. The world is now immobilized by this infectious RNA virus. As of Jan 5, already more than 3.15M people have been infected and 5.73M people died RNA viruses are different than DNA-based viruses in the sense that they have higher mutation rates, and hence, they have higher adaptive capacity. This mutation causes continuous evolution that leads to host immunity and therefore, based on the type of mutation, it affects the rate of disease transmission and pathogenicity This RNA virus can do the mutation in the human body. Accurate determination of mutation rates is essential to comprehend the evolution of this virus and to determine the risk of emergent infectious disease. The collected dataset is processed to determine the mutation of different parts of the Covid-19 separately

We proposed a model for the Virus Mutations Prediction. The proposed approach consists of four main phases:

1. Sequences of datasets are preprocessed.
2. Once we have preprocessed sequences of data, they are transformed into a format that is suitable for training an LSTM network. In this case, a one-hot encoding of the integer values is used where each value is represented by a binary vector that is all “0” values except for the pointer to the word, which is set to 1
3. The input data are prepared to train on the LSTM encoder. After that, it is the role of the decoder to take the output from the encoder as integers and transform it into sequences
4. The obtained results are evaluated.

Key Words: *Artificial Intelligence, LSTM Algorithm, COVID-19, RNA viruses*

Identification of a three-miroRNA-based prognostic gene signature in hepatocellular carcinoma using the random survival forest method

Sadra Salehi-Mazandarani[1], Mohammad Hossein Donyavi[2], Parvaneh Nikpour[1]*

[1] Department of Genetics and Molecular Biology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran
[2] Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran

E-mail address (corresponding author): pnikpour@med.mui.ac.ir

Abstract

Background: Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related death worldwide. HCC has poor prognosis and monitoring of HCC patients with high risk is necessary. Many studies have revealed miroRNAs (miRNAs) as potential prognostic biomarkers in cancer. For this aim, we performed a bioinformatic-based analysis and random survival forest method to identify a prognostic gene signature in HCC.
Materials and Methods: miRNA-seq data and clinical information of HCC samples were downloaded from The Cancer Genome Atlas (TCGA) database using the TCGAbiolinks R package. Differentially-expressed miRNAs (DEMis) among cancerous and paracancerous samples were identified by the DESeq2 package based on $|logFC| > 2$ and adjusted p-value < 0.01. Univariate survival analysis was performed on the DEMis data of cancerous samples to identify prognosis-related miRNAs (hazard ratio (HR) $\neq 1$ and p-value < 0.01) by the survival package. Subsequently, randomForestSRC package was utilized to rank survival-related miRNAs. Then, multivariate cox regression analysis was performed to establish a risk scoring model based on the regression coefficient and gene expression. Survival and time dependent ROC (receiver operating characteristic) curve plots were generated by the survival and survivalROC packages, respectively.
Results: 131 DEMis (126 upregulated and 5 downregulated) were identified between 372 cancerous and 50 paracancerous HCC samples. Univariate survival analysis revealed 8 prognostic-associated miRNAs. Based on the random forest analysis, 3 miRNAs (hsa-miR-9-5p, hsa-miR-137-3p and hsa-miR-105-5p) which had the relative importance > 0.6, were selected to construct a prognostic gene signature. The HR and p-value of the model were 2.6 and 0.00003, respectively. Additionally, AUC of the model for 5, 3, and 1 year were 0.66, 0.61, and 0.64, respectively.
Conclusion: This study provides a reliable gene signature for prediction of
prognosis in HCC patients using miRNAs. Further studies are needed to evaluate the strength of this model in HCC.

Key Words: *Carcinoma, Hepatocellular; MicroRNAs; prognostic Biomarker*

Machine learning-driven set of peripheral blood microRNAs as diagnostic biomarkers for myocardial infarction

Mehrdad Samadishadlou[1], Zeynab Piryaei[2,3], Farhad Bani[1,4]*, Kaveh Kavousi[3]*

[1] *Department of Medical Nanotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran*
[3] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Iran*
[4] *Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Cardiovascular disease is the leading cause of mortality worldwide and myocardial Infarction (MI) is responsible for 85% of cardiovascular disease mortality. Since the survival rate in MI cases strongly depends on fast diagnosis and treatment, discovering novel biomarkers for rapid and accurate diagnosis is of great importance. MicroRNAs are significant regulators of adaptive and maladaptive responses in cardiovascular diseases. Hence, microRNAs have undergone intensive research as possible therapeutically and diagnostic targets. However, their role as novel biomarkers for diagnosing MI needs to be more investigated. The microarray GSE61741 dataset has been downloaded from the Gene Expression Omnibus (GEO) database including 863 microRNAs expression profile in peripheral blood. The selected samples included 94 healthy (as the control) and 62 samples with MI (as the case). At the first, differentially expressed microRNAs has been identified using the limma package with the adjusted P-value <0.05 and -1 > log2 FC > 1 criteria. Then, sequential forward and backward selection algorithms has been applied for feature selection. Finally, the support vector machine (SVM) algorithm has been performed on selected microRNAs to classify samples with 10-fold cross-validation. 100 differentially expressed microRNAs has been identified in samples with MI compared to healthy samples. 35 microRNAs with the greatest importance has been selected using feature selection algorithms. Among them, a unique signature of five microRNAs (including hsa-miR-1246, hsa-miR-1258, hsa-miR-1279, hsa-miR-132*, and hsa-miR-142-3p) have been chosen and an SVM model has been trained with their expression values. The trained model predictive values are 0.92, 0.84, and 0.91, for AUC, sensitivity, and specificity, respectively. Based on our findings, the multi-marker approach increases predictive values in comparison to single microRNAs. Therefore, microRNA signatures derived from peripheral blood could be valuable novel biomarkers for more accurate diagnosis of MI.

Key Words: *Myocardial infarction, MicroRNA, Machine Learning, MicroRNA Signature, Biomarker*

Machine Vision Based Conic Estimation: An Effective Tool for Reinforcement of Deep Learning in Measuring Fetal Head Circumference

Amir Saniyan[1], Seyed Vahab Shojaedini[2]*

[1] *Department of Computer Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran*
[2] *Department of Biomedical Engineering, Iranian Research Organization for Science and Technology, Tehran, Iran*

E-mail address (corresponding author): shojadini@irost.ir

Abstract

The artificial intelligence is considered as an effective way to estimate the border around the fetal head based on ultrasonic images. This approach has been remained as an open problem because of existing some challenges such as low contrast of the image under various conditions and its low signal to noise ratio.
Targets: In this article the machine vision based, conic estimation scheme is utilized as supplementary processing of a VNET deep network in order to promote the performance of artificial intelligenece in measuring fetal head parameters. Firstly, a cost function is constructed by using a family of curves, each of them may be a candidate for the fetal border. Then the objective function is minimized by using optimization framework in order to obtain the best fitted curve.
Methodology: The testbed for evaluation of the proposed idea was prepared by using the software Python with Keras on a computer with a NVIDIA 2080 TI GPU with 32 GB RAM. The tests were performed on HC18 challenge dataset provided by Heuvel and collaborators which includes 999 ultrasound images which have been divided into three categories of training, validation and test subsets, with ratios of 60%, 15% and 25% among the total images respectively.
The obtained results showed that proposed scheme may lead to significant promotion in measuring the parameters of fetal head. Therefore our method resulted in the accuracy, precision dice and jaccard parameters equal to 97.6%, 94.3%, 95.4% and 91.7% respectively. These results show at least 3.6% improvement over the same methods which did not use from geometric concept.
Conclusions: Improving the estimation parameters show effectiveness of our idea based on reinforcement of artificial intelligence tools in measuring fetal head circumference by considering the geometric nature of the problem.

Key Words: *Fetal head circumference; Deep learning; Conic fitting; Geometric properties.*

Identification of genes affecting in Non – small cell lung cancer using machine learning techniques and bioinformatics tools

Marzie Shadpirouz[1]*, Morteza Hadizadeh[2], Sadegh Raoufi[3]

[1] *Department of Applied Mathematics,Factulty of Mathematical Sciences,Shahrood University of Technology,Semnan*
[2] *Physiology Research Center, Institute of Neuropharmacology, Kerman University of Medical Sciences, Kerman, Iran*
[3] *Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*

E-mail address (corresponding author): m.shadpirouz@gmail.com

Abstract

Background & Objective: Lung cancer is the second most common cancer after breast cancer and the main trigger cause of death in women and men globally. Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancers. Because early detection of cancer plays a vital role in treatment, this study sought to identify genes that could potentially be effective in early Non – small cell lung cancer screening.
Material & Methods: Firstly, three micro-array datasets (GSE1987, GSE44077, and GSE74706) related to non-small cell lung cancer were downloaded from the Gene Expression Omnibus (GEO). After integrating and bath effect removal of these datasets, Lasso logistic regression was used to extract important genes. Processing of all data was performed using the R statistical programming language. Also, Gene Set Enrichment Analysis (GSEA) was performed by Metascape bioinformatics tool to identify KEGG pathways and Gene Ontology Enrichment.
Results: Finally, the introduced model selected 15 genes (ACVRL1, ANKRD1, C11orf80, CA4, EIF1B, FGF2, GRK5, KLHL18, LILRA1, MME, SDC1, STX11, TMOD1, TTN, WIF1). The accuracy level of the model was 100%. These genes are related to the Wnt signaling pathway, which plays a significant role in NSCLC. Until now, seven genes (47%) have been reported in biological studies as genes effective in NSCLC.
Conclusion: With the use of machine learning techniques and bioinformatics tools, this study has introduced new genes that can serve as the target of early diagnosis or treatment of NSCLC.

Key Words: *Non – small cell lung cancer; Gene expression; Gene selection; Machine learning; Lasso logistic regression.*

Prediction peptide activity and interaction in drug discovery by utilize machine-learning technique

Mozhgan Shavandi[1]

[1] *Materials and Energy reaserch center, Karaj, Iran*

E-mail address (corresponding author): mojhgan_19@yahoo.com

Abstract

These days, machine-learning-based predictions of anti cancer drug show many attentions due to accurately predicting. Peptide-based machine learning techniques, which is a rapid and perfect outcome prediction, play an important role in developing peptide drugs. Anti cancer peptides that usually contain 5 to 30 amino acid residues, can destroy cancer cells through apoptosis and necrosis that led to kill cells of cancer tumor. It actions selectively without damaging other normal cells and show less systemic toxicity. They possess high hydrophobicity and a positive net charge. We study and use a systematic review on the application of machine learning techniques and prediction of peptide drug activity. In this study computational tool construction based on machine learning algorithm were utilized to identify activity and interaction of peptide as an anti cancer drug. It has carried out by features calculated from the amino acid sequence and atomic composition. Using machine-learning approaches, we can develop prediction model for peptide drugs system.

Key Words: *Machine learning; Peptide drug; Anti cancer drug; drug discovery.*

Toxicoinfotrmatics: Recent Approaches in Integration of Big Data Biomath for the Benefit of healthcare policy

Parisa Shoae Hagh[1]*

[1] *Department of Pharmacodynamics and Toxicology, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran*

E-mail address (corresponding author): shoaeip941@gmail.com

Abstract

The era of in vivo, in vitro and ex vivo transmission to in silico experiments has been rapidly progressed in recent decades due to the benefits of computation in shortening the prediction time of results, reducing costs, and has provided significant advances in computational predictive toxicology, Including risk assessment of different organs toxicity at the encounter of high-risk molecules, estimating the co-effect of the time on life-long exposure, aggregating these effects with the consequences of multifactor underlying diseases and bringing forward system toxicology and ultimately using next generation sequencing results in designing individual-centered therapies for individual medicine. However, existing data and massive calculations are still not being used as efficiently as they should be in achieving health goals. In this article we are focused on introducing adverse outcome pathways coalition and outcome extrapolation strategies by artificial intelligence, integrated approaches of testing & prediction utility, some aspects of the most machine learning methods: support vector machines (SVMs), random forest (RF), decision trees (DTs), Naive Bayes, k-nearest neighbors (KNN) and neural networks and their current deficits in. finally propose practical solutions for current issues in big data analysis systems. We believe this is not a smooth road to endpoint application but a feasible prospect to beneficial decision-making regulators in favor of individuals one-by-one, in an example of what the world confrontes in the case of Covid-19, we summarize the unaccounted adverse outcomes of long-approved beneficial drugs because of the lack of massive calculations in the scale of whole population and how affected the general healthcare policies.

Key Words: *toxicoinformatics; big data; computational toxicology; system toxicology; machine learning*

Big Data in Biology

Evaluation overexpression of miR-451a in decreased CERK gene expression in breast cancer using bioinformatics analysis

Nada abadh[1], Majid Sadeghizadeh[1]*

[1] *Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*
[2] *Department of Immunology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): n.abadh1981@gmail.com

Abstract

MicroRNAs (miRNAs) are one of the essential components of gene regulatory networks today. MiRNAs can down-regulate gene expression by inhibiting translation or stimulating mRNA degradation. In some cases, they can also up-regulate the target gene. MiRNAs participate in regulating biological processes and thereby have the promising potential to act as valuable diagnostic biomarkers and even therapeutic molecules in various diseases.
Bioinformatics analysis is a pivotal initial step of microRNA research, starting from searching, targeting, interaction. In our study, we used bioinformatic analysis data to find out (1) the expression of miR-451a (2), the expression of miR-451a in breast cancer compared to other miRNAs (3), miR-451a target prediction (4), Interaction between miR-451a and cerk (5) the link between miR-451a and breast cancer. Finally, bioinformatics is continuing the development of new tools in the future and in the world, so it is more helpful for cancerous diseases and attractive in molecular therapy.

Key Words: *breast cancer, MicroRNAs, miR-451a, cerk*

Microarray meta-analysis focused on transcription factors involve in breast cancer cell response to soy isoflavones

Elham Ashrafi-Dehkordi[1]*, Ahmad Tahmasebi[1,2], Habil Zare[3,4], Seyed Mohammad Mazloomi[1]

[1] *Nutrition Research Center, Department of Food Hygiene and Quality Control, School of Nutrition and Food Sciences, Shiraz University of Medical Sciences, Shiraz, Iran*
[2] *Biotechnology Institute, College of Agriculture, Shiraz University, Shiraz, Iran*
[3] *Department of Computer Science, Texas State University, San Marcos, Texas, USA*
[4] *Department of Cell Systems & Anatomy, The University of Texas Health Science Center, San Antonio, Texas, USA*

E-mail address (corresponding author): elhamashrafi@gmail.com

Abstract

Breast cancer is the most common cancer among women and one of the leading causes of cancer death worldwide. Certain lifestyle behaviors and diet patterns are recognized as factors that influence the risk of breast cancer. Intake of phytoestrogens can affect the risk of breast cancer. For the first time, we performed large-scale comparative transcriptomic analysis via a meta-analysis of breast cancer cell response to soy isoflavones using microarray gene expression data implemented. By analyzing 283 microarray samples from 7 different experiments, we identified 3,890 genes differentially regulated in response to breast cancer cell response to soy isoflavones, with a false discovery rate (FDR) ≤ 0.001, of which 2,173 were up-regulated and 1,717 were down-regulated. Among the DEGs, 269 TF genes were identified, and these belong to 42 TF families. The C2H2 ZF, bZIP, and bHLH were the largest families with 85, 25, and 19 members, respectively. Recent studies show that zinc finger proteins are key TFs that contribute to cancer progression by regulating the transcription of downstream genes involved in proliferation, apoptosis, migration, and invasion. For example, ZNF217 has been reported to play a critical role in promoting breast cancer metastasis. In total, 55.4% of the TFs were up-regulated and 44.6% were down-regulated. Interestingly, the expression level of all TFs in the E2F family (including E2F1-6, and E2F8) were up-regulated in the isoflavones exposure condition. pRb–E2F complexes coordinate the expression of genes involved in the cell cycle and apoptosis. Previous studies have indicated that pRb binds to E2F1, E2F2, and E2F3 during much of the G1 phase of the cell cycle, and it silences gene expression by recruiting HDACs or HMTs. The tumor suppressor pRb exerts its effect on cell growth through interaction with other factors, such as activating E2F1-3.

Key Words: *Isoflavones, Breast cancer, Transcriptome data, Meta-analysis*

Gene ontology to explore functions of DEGs detected in breast cancer cell response to soy isoflavones

Elham Ashrafi-Dehkordi[1]*, Ahmad Tahmasebi[1,2], Habil Zare[3,4], Seyed Mohammad Mazloomi[1]

[1] *Nutrition Research Center, Department of Food Hygiene and Quality Control, School of Nutrition and Food Sciences, Shiraz University of Medical Sciences, Shiraz, Iran*
[2] *Biotechnology Institute, College of Agriculture, Shiraz University, Shiraz, Iran*
[3] *Department of Computer Science, Texas State University, San Marcos, Texas, USA*
[4] *Department of Cell Systems & Anatomy, The University of Texas Health Science Center, San Antonio, Texas, USA*

E-mail address (corresponding author): elhamashrafi@gmail.com

Abstract

Disparities in breast cancer incidence between Asian and Western countries have long been noted. A possible contributor to this difference may be the consumption of soy products because higher soy intake among Asian populations has been associated with a lower risk of breast cancer. However, the effect of isoflavones on breast cancer risk is controversial. Many studies have reported that a diet high in isoflavones decreases the risk of breast cancer, while other studies (e.g. in vitro and animal studies) suggest that isoflavones are not statistically significantly associated with breast cancer risk. The abundant available transcriptome data offer an excellent opportunity to investigate the mechanism-specific effects of isoflavones on cancer cells. A meta-analysis is a robust strategy that leverages multiple datasets and provides reliable results and information through larger sample sizes and thus more statistical power. For the first time, we performed a large-scale meta-analysis using 283 microarray samples from 7 different experiments. In addition, we used GO and pathway enrichment analyses, which are differentially expressed genes (DEGs) in breast cancer cell response to isoflavones. The result of biological process enrichment analysis of the up-regulated genes revealed that the terms associated with cell division, mitotic nuclear division, and viral processes were significantly overrepresented. These findings also suggest that isoflavones may influence cell growth through modulating signaling pathways. The terms telomere maintenance via recombination was also found to be significant. In addition, down-regulated genes revealed a number of terms related to protein transport and the negative regulation of cell proliferation. Moreover, the dominant categories of molecular functions were protein binding and ATP binding for the up-regulated genes. In the category of cellular components, the nucleus and cytoplasm were the most significantly represented groups for up-regulated and down-regulated genes, respectively.

Key Words: *Breast cancer, Isoflavones, Gene ontology, Meta-analysis*

Weighted gene coexpression network analysis of DEGs detected in breast cancer cell response to soy isoflavones

Elham Ashrafi-Dehkordi[1]*, Ahmad Tahmasebi[1,2], Habil Zare[3,4], Seyed Mohammad Mazloomi[1]

[1] Nutrition Research Center, Department of Food Hygiene and Quality Control, School of Nutrition and Food Sciences, Shiraz University of Medical Sciences, Shiraz, Iran
[2] Biotechnology Institute, College of Agriculture, Shiraz University, Shiraz, Iran
[3] Department of Computer Science, Texas State University, San Marcos, Texas, USA
[4] Department of Cell Systems & Anatomy, The University of Texas Health Science Center, San Antonio, Texas, USA

E-mail address (corresponding author): elhamashrafi@gmail.com

Abstract

Soy foods are rich in isoflavones, which belong to the phytoestrogen family that could reduce breast cancer risk. To define the biological networks affected by isoflavones in breast cancer cells, a WGCNA was performed to identify modules by analyzing 283 microarray samples from 7 different experiments. The results showed, genes were grouped into five discrete modules, each of which was coded by a unique color. The size of these modules ranged from 101 to 2,257 genes. The two major modules were named turquoise (2,257 genes), and blue (1,013genes). We performed an enrichment of GO to gain insight into the biological functions of the genes within these modules. The turquoise module was highly enriched for cell division and mitotic nuclear division. The blue module was associated with DNA replication. The majority of genes in blue module were up-regulated by isoflavones. The blue module also included 39 genes that were related to pathways in cancer (e.g., BAD, HSP90B1, and E2F3). BAD regulates cell growth. Role of Hsp90B1 in ovarian cell survival and cell apoptosis has been highlighted. E2F3 controls cell cycle progression. The brown module was highly enriched for genes involved in embryonic digit morphogenesis. Moreover, we observed this module was annotated in response to estrogen. This module was enriched for hippo signaling pathway. Hippo pathway plays an important role in cell proliferation and breast cancer metastasis. The GATA3 gene was found in the brown module. GATA3 requires normal development of the mammary gland and shows a relatively high sensitivity to breast carcinomas. The yellow module was observed to be enriched with terms such as cell-cell adhesion, translation, and protein folding. Finally, the green module was not enriched for any significant biological process term. In the blue module, DEGs were significantly enriched in the cell cycle and cancer pathways.

Key Words: *Breast cancer, Isoflavones, Weighted gene coexpression network analysis, Meta-analysis*

Integrated systems biology investigation (in-silico) of the hub regulatory RNA interactions (competitive endogenous RNA) in the lung cancer microarray samples

Parisa Rabiei Chamgordani, Mahdieh Bakhshayesh

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Background: Lung cancer is an insufficient treatment option, with only 20% of patients responding to systemic chemotherapy. miRNAs, discovered non-coding RNAs, regulate gene expression at the post-transcriptional level by base pairing with 3'UTR of their targeted mRNAs annotated as to their targetome. Analyzing the expression mRNAs microarray from lung cancer tissues and non-tumor tissues with GSE136043.

Result: According to the microarray data from comparing lung cancer tissue and normal tissue, we have examined five genes, including efcc1, LILRB2, JAML, SLC16A7, and DMRT2, had the most significant expression changes. We have found microarray data based on the highest score identified lncRNAs similar to the chromosomal number. We have examined interactions between lncRNAs and the genes. The low expression of our genes can lead to miR-1304-5p, miR-3173-3p, miR-148a-3p, miR-221-3p, miR-7-5, and miR-363-3p not attaching to the gene, binding to the THRB-IT1, LINC01002, LINC00665, and C9orf147 reducing the expression of the lncRNAs. We chose three genes from five genes that indicate the existence of CeRNA because the miRNA of genes including efcc1, LILRB2, and DMRT2 had similar lncRNAs, which indicates the existence of a complex network that can be used in gene therapy or drug. miR-7-5p targets a gene called CCND1 and a protein target called PA28gamma. PA28gamma emerges as a novel functional target of tumor suppressor miRNA-7 in non-small-cell lung cancer.

Methods: All miRNAs were regained from DIANA-TarBase v7.0 and miRwalk databases, followed by examining their expression in lung tumors using the GEO and DAVID database. They were finding interaction between lncRNAs and genes by LncRRIsearch, comparing the expression of genes in lung cancer by Gepia2 databases. The expression of lncRNAs in different tissues has been examined by the lnCAR databases. The study of metabolic pathways and expression of the gene was performed using KEGG and Reactome Pathway Database.

Key Words: *Lung cancer, miR-7-5P targetome, Complex network, Microarray*

Integrative gene expression analysis of peripheral blood from autism spectrum disorder cases

Parimah Emadi Safavi[1], Seyed Shahriar Arab[1], Javad Zahiri[2]*

[1] *Biophysics Department, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*
[2] *Department of Neuroscience, University of California, San Diego, La Jolla, CA, United States.*

E-mail address (corresponding author): jzahiri@health.ucsd.edu

Abstract

Autism spectrum disorders is an umbrella term used for the set of neurodevelopmental conditions with early childhood onset characterized by difficulties in social interactions and repetitive behaviors. Recent advances in high-throughput sequencing and public data sharing have provided us with new opportunities to investigate the underlying mechanisms of such complex diseases. nevertheless, identification of disorder heritability and gene regulatory signatures is still a challenge since each gene mutation involved in the disease only stands for a minor subset of autistic cases. Accordingly, transcriptomic analysis is a great approach since provides us the ability to identify pathways associated with disease-related genes. In this research, we used publicly available microarray datasets of peripheral blood of autistic people and typically developed cases in search of candidate risk genes. We incorporated disease genes from the Simon Foundation Autism Research Initiative database and gene sets of interest from prior literature for their association with our differentially expressed genes in blood datasets. Our pathway enrichment analyses results have shown critical pathways such as those in chromatin remodeling and cell adhesion which are crucial to neuron plasticity. These processes are an essential part of brain development in early childhood. It is remarkable since autism emerges by the time brain development happens which is evident that our other differentially expressed genes can introduce candidate risk genes in autism. These findings prove the applicability of data integration in the search for potential risk genes and pathways with almost high confidence. Our analysis would be a candidate reproducible method to illuminate pathways, heritability, and etiology of autism spectrum disorders.

Key Words: *autism spectrum disorder; data integration; microarray; peripheral blood; gene expression*

Cross talk between cancer immunity cycle and EMT; new light for study cancer progression

Amirali Hariri[1]*, Hamid Mir Mohammad Sadeghi[1], Mina Mirian[1]

[1] *Department of Pharmaceutical Biotechnology, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran*

E-mail address (corresponding author): mina6m@yahoo.com

Abstract

The immune system becomes involved in a complex set of interactions with different tumor cells. Several sequential actions must be initiated and allowed to proceed and repeatedly expand for an anticancer immune response to lead to the efficient destruction of cancer cells. These processes are known as the cancer Immunity cycle. On the other hand, epithelial to mesenchymal transmission (EMT) is another significant process involved in the development and invasion of cancer. Several similar pathways that regulate EMT are involved in tumor-immune interactions, but little is known about the mechanisms and consequences of interference between these regulatory processes. We performed comprehensive multi-omics analyses to determine the interference between EMT and cancer immunity cycle and their clinical relevance in breast cancer. We obtained relevant multi-omics data containing expression profiles, RNA sequencing data, and clinical data from 1119 breast cancer patients (BRCA, n = 1119) from The Cancer Genome Atlas (TCGA; tcga-data.nci.nih.gov/tcga). The EMT score of cancer cells was measured using the EMT signature, which compares the expression of epithelial marker genes and mesenchymal marker genes. We used two online servers (immune land escape and TIP) for the immune system status and merged these data. Our study demonstrated the existence of complex and dynamic interaction between the cancer immunity cycle and EMT and their effect on cancer prognosis and treatment. Our study highlights the potential for EMT-immune interference as a model for explaining the underlying molecular mechanisms of cancer progression and guiding more effective and generalized cancer.

Key Words: *Epithelial to mesenchymal transmission, Cancer immunity cycle, Omics*

Comprehensive RNA-Seq Meta-analysis identifies genes/pathways related to fat-tail development of sheep

Seyedeh Fatemeh Hosseini[1], Mohammad Reza Bakhtiarizadeh[1]*, Abdolreza Salehi[1]

[1]*Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): sf.hosseini75@ut.ac.ir

Abstract

Fat-tail content of sheep presents a valuable energy reserve that has historically facilitated adaptation to harsh environments. However, in modern intensive sheep industry systems, thin-tailed sheep are preferred, because fat deposition requires a greater energy cost than accretion of an equivalent amount of lean tissue. Comparative transcriptome analysis of thin-tailed against fat-tailed breeds can shed light on the genetic mechanisms controlling fat-tail development and improve breeding strategies to modulate fat deposition. Transcriptomic studies often suffer from issues with reproducibility, which can be improved by integrating multiple studies based on meta-analysis. Here, an RNA-Seq meta-analysis on the transcriptomes of six publicly available datasets was performed. To analyze each study, first, quality control and trimming of the raw reads were done using FastQC and Trimmomatic tools, repectively. Then, the clean reads were aligned against reference sheep genome (V1.0.104). Htseq was employed to quantify the expression levels of the genes and DESeq2 R package was used to identify the genes differentially expressed (DEGs) between fat- and thin-tailed sheep breeds. The output of the DEGs analysis from all studies was further employed for the meta-analysis based on two approaches (fishercomb and invnorm) from metaRNASeq R package. The genes that were identified to be DEGs by both approaches were considered as final DEGs. Totally, 650,605,259 reads related to 6 samples were analyzed. A total of 353 genes (134 upregulated, 219 down-regulated) were identified as DEGs. Some of these genes including SCD, FASN, ACACA, CPT2, ACLY, LPL and DGAT2 confirmed the previous reports of associations with fat-tail development. Functional analysis of DEGs showed 64 KEGG pathways and 2423 GO terms were signiciantly enriched including "fatty acid metabolism pathways", "carbohydrate metabolism" and "steroid biosynthesis" which may contribute to fat storage in sheep tails. Our results highlighted a core set of genes/pathways associated with fat-tail development in sheep.

Key Words: *Sheep, Fat tail, Fat deposition, RNA-seq, Meta-analysis*

Identification of oxidative stress responsive hub genes and their related miRNAs in Arabidopsis thaliana based on integrated bioinformatics meta-analysis

Samaneh Hosseinpour[1], Homa Hosseinpour[2], Mohammad Rezaei, Mansoureh Azadeh

[1] *Zist Fanavari Novin biotechnology institute, Strategic and Modern Technologies Skill training center*
[2] *Technical and Vocational Organization, Hezar Jerib street, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Oxidative stress is an integral component of various stress conditions in plants, and determines the substantial overlap in physiological and molecular responses to biotic and abiotic environmental challenges. To better understand plants how to adapt environmental stress, we retrieved the oxidative-responsive genes of Arabidopsis thaliana, the main model plant for the experimental analysis. The purpose of our study was the identification of oxidative-responsive genes and their related miRNAs and associated pathways in A. thaliana under stress conditions using a high-throughput in silico meta-analysis. In this study we integrate three microarray datasets (GSE20009, GSE40574 and GSE57286) in the Gene Expression Omnibus (GEO) database. Normalization and processing of raw data performed by (SVA) and (limma) packages of R (version 1.4.1). The DEGs were identified by the thresholds of p-value < 0.05, and |log2fold change (FC)| > 1. The analysis of gene ontology enrichment (GO), protein-protein interaction (PPI), pathway analysis, and potential miRNAs for hub genes, were performed using bioinformatics online tools DAVID, STRING, KEGG, psRNATarget and RNAhybrid. The result 90 DEGs, including 46 up-regulated and 44 down-regulated genes, were analyzed by integrated bioinformatics approaches enriched in glutathione metabolism, flavonoid biosynthesis, oxidation-reduction process, and response to stress. DFR, GSTU5, PGDH, GSTF6, GSTF12, PYD4, APL3, CYP81D1 and F3H was hub genes in oxidative stress. All hub genes except GSTF6, APL3 and F3H were target genes of 11 miRNAs. This research indicated that two hub genes DFR and GSTU5 and their related miRNAs: ath-miR418 and ath-miR391 have effect on flavonoid biosynthesis and glutathione metabolism pathways respectively in oxidative response adaptation. They also can be considered in breeding programs and genetic engineering for the production of tolerant plants.

Key Words: *Oxidative stress; Arabidopsis thaliana; miRNA; integrated bioinformatics meta-analysis*

Identification of novel genes in different regions of brains: Meta-analysis of schizophrenia transcriptome data

Arshia Jahanimoghadam[1], Masoud Shahsavari[2], Negin Ahmadi Jazi[3], Aria Jahanimoghadam[4], Javad Zahairi[5]

[1] *Department of Biology, Faculty of science, Islamic Azad University of Mashhad, Mashhad, Iran*
[2] *Department of Agronomy and Plant Breeding, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran*
[3] *Department of Zoology, Faculty of Biological Science, Shahid Beheshti University, Tehran, Iran*
[4] *Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany*
[5] *Department of Neuroscience, University of California San Diego, La Jolla, CA, USA*

E-mail address (corresponding author): arshia.jahanim@gmail.com

Abstract

Schizophrenia (SCZ) is a severe mental disorder caused by complex genetic and environmental interplay. Despite a large amount of research was carried out in SCZ, the underlying mechanism is remained to be fully understood. The available drug options for SCZ treatment showed lower than expected efficiency. With the advent of high-throughput sequencing techniques along with availability of sequencing data, an urgent need to data analysis has appeared. Of note, meta-analysis enables us to integrate several databases and therefore increase the reliability of the results. In the present study, we obtained data from different regions of brain, integrate them and introduce new genes associated with SCZ. Microarray gene expression data from 119 samples (56 normal and 63 SCZ) were retrieved from superior temporal gyrus (STG), hippocampus (HIP), Pre-frontal cortex (PFC) and associative striatum (ASR) following by analysis and integration by GEO2R and MetaDE package respectively. Initially, we identify 4026 genes with p-value < 0.05 using naive sum of the ranks method. We then narrowed our list down to 32 genes owing to removing genes having expression under |0.05| in either one of the regions. Furthermore, among the 32 mentioned genes, 27 genes were investigated experimentally in previous studies. Novel genes BOD1L1, CHM, EDN3, ENDOD1, and ANGPTL4 were found to be differentially expressed genes across all the regions. These candidate genes are of importance for further investigation as they may elucidating the SCZ etiology or being used as biomarkers in the clinical context.

Key Words: *Keywords: Schizophrenia; transcriptome; meta-analysis; biomarker*

Associated Gene Expression Profiles in Pathogenesis of Lung Adenocarcinoma

Sajad Jandaghian

E-mail address (corresponding author): jandaghian.78@gmail.com

Abstract

Lung cancer is the leading cause of cancer-related deaths in men and the second-leading cause in women worldwide and its occurrence and mortality are heavily influenced by cigarette smoking patterns. Based on histopathology assessment lung cancer can be divided into two general subtypes, non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) between which the former is the more common type. Lung adenocarcinoma (LAUD) is the most common among the three histological forms of NSCLC. It is speculated that subclasses of cancer may be related to different cells of origin and their related microenvironments NSCLS can be defined as an invasive and complex disease because it cannot be assessed by a single genetic alteration or a particular biological pathway.

Lung cancer is the leading cause of cancer-related deaths in men and the second-leading cause in women worldwide and its occurrence and mortality are heavily influenced by cigarette smoking patterns. Based on histopathology assessment lung cancer can be divided into two general subtypes, non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) between which the former is the more common type. Lung adenocarcinoma (LAUD) is the most common among the three histological forms of NSCLC. It is speculated that subclasses of cancer may be related to different cells of origin and their related microenvironments NSCLS can be defined as an invasive and complex disease because it cannot be assessed by a single genetic alteration or a particular biological pathway. This study was conducted through analysis of microarray data acquired from the GEO database. The analysis was conducted in an R language environment. The DEGs (Differentially Expressed Gene) related to LAUD cases were identified and gene profiles related to these set of DEGs was subsequently obtained.

Key Words: *Lung cancer, LAUD, Microarray, DEG, Gene profile*

Genome-wide profiling of DNA/RNA hybrids (R-loops) in psoriasis disease

Leila Kianmehr[1]*, Hassan Rajabi-Maham[1], Minoo Rassoulzadegan[2]

[1] *Animal Sciences and Biotechnology Department, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University,Tehran, Iran*
[2] *Université de Nice, INSERM-CNRS, France*

E-mail address (corresponding author): kianmehr96@gmail.com

Abstract

DNA/RNA hybrids (R-loops) are RNA-mediated epigenetic element involved in the molecular mechanism of human diseases. R-loops are three-stranded structure made up a DNA/RNA hybrid and the displaced single-stranded DNA. Recently proposed independent-antibody method of our group leads to detect these structures in different type of cells started from germ cells to currently somatic cells. One physiological instance of such structure present at our chromosomes ends (telomeres) as a lncRNA named TERRA (TElomeric Repeat-containing RNA; constitutive of hexameric repeats $(TTAGGG/CCCTAA)^n$. Using TERRA as a privilege model system to study R-loops helps us to understand DNA/RNA hybrids are accumulated at the telomeres of different chromosomes in psoriasis patient samples (lesional and non-lesioanl skin samples vs. healthy controls) as previously observed in germ cells. It is reported that short telomeres are associated with higher TERRA levels in aging research. TERRA/DNA hybrids opens new startegies to study psoriasis genome instability. Distribution of different non-coding RNA besides TERRA throughout the genome of psoriasis-derived skin cells revealed the unresolved DNA/RNA hybrids may facilitate skin lesions formation and may play an important role in epigenome shaping.

Key Words: *DNA/RNA hybrids, R-loops, epigenetics, RNA-seq, psoriasis, TERRA*

Computational prediction of MicroRNA-Cancer relations

Mahsa Morattab[1], Adel Eghbali[2], Shahab Bakhtiari[3]*, Narjes Takhshid[4], Shahrzad Zirak

[1] Department of biology , faculty of basic science,Central Tehran Branch Islamic Azad University,Tehran,Iran
[2] Department of Medical Laboratory Science, Faculty of paramedical science , Mazandaran university of medical science ,Sari, Iran
[3] Department of Biological Sciences, University of Kurdistan, Sanandaj, IranSchool of advanced sciences and technologies, Azad University of Tehran Medical Science,Tehran ,Iran
[4] Department of Life Science Engineering, Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran

E-mail address (corresponding author): shahabbakhtiari.gen@gmail.com

Abstract

Every year over a million people lose their lives because of cancer, which is why scientists are more motivated than ever to carry out numerous researches to understand the fundamental causes of cancer 1. MicroRNAs (miRNAs) are short RNA molecules that bind mRNA, altering their regulation 2. It is feasible to use them as potential diagnostic biomarkers 3 due to their stability and aberrant expression in the pathogenesis of cancer, cardiac, and other diseases 4. They modulate many processes that contribute to different stages of cancers 5. For example, Experimental evidence suggested that MIR21 could be involved in different cancers 6. Many databases provide detailed information regarding these small molecules 7. Link prediction is utilized to estimate hidden relations among existing links based on known topology 8, which is more efficient in a complex network to find missing links or even predict other ones 9. Furthermore, it promotes discovering the new relations among disease-related miRNAs that could broaden our horizons of the molecular mechanisms of human diseases 10, which could sometimes be challenging in bioinformatics research 11. The accuracy of new algorithms in link prediction has been proved by extensive experiments on both synthetic and real networks 7. In this article, we have created a miRNA-cancer network using four link prediction algorithms, including common neighbors (CN), Preferential attachment (PA), Jaccard (JC), and Adamic and Adar (AA), based on the Human microRNA Disease Database (HMDD v3.2) to predict new miRNA-cancer relations. According to our predictions, hsa-mir-146a-Glioma and hsa-mir-19a-Non-small-cell lung carcinoma (NSCLC), are some of the most probable Hsa-miRNA cancer associations that have not been published in any articles yet, and they could be the best probable candidates for additional laboratory and validation studies.

Key Words: *miRNA; Link prediction; Cancer; bipartite network; bioinformatics*

Investigation of the interaction between bta-miR2410 and PTEN gene in dairy cattle production

Zahra Roudbari[1], Saideh Eskandarynasab[1´2]

[1] Department of Animal Science, Faculty of Agriculture, University of Jiroft, Jiroft, Iran
[2] Department of Animal Science, Faculty of Agriculture, Zabol University, Zabol, Iranroudbari.zahra@ujiroft.ac.ir

E-mail address (corresponding author): rodbari.zahra@gmail.com

Abstract

Predicting microRNAs related to target genes using bioinformatics tools saves time and laboratory testing costs. One of the most important economic traits in the dairy cattle industry is milk production, which is controlled by many genes and microRNAs. In this study, we investigated the interaction between bta-miR2410 and the PTEN gene. For this purpose, initially the target genes of bta-miR2410 were extracted from miRWalk 2.0 database and according to the importance of PTEN gene in the milk production process, it was selected from the target genes. Then, the gene sequence was downloaded from the NCBI database and the miRBase database was used to find the bta-miR2410 sequence. The results of interaction of bta-miR2410 and PTEN gene with MFE = -29.9 kcal/mol show the high effect of this microRNA on PTEN gene. The effects of PTEN target gene on milk production have been studied in other studies, and It has been reported that this gene initiates lactation by playing a role in the AKT signaling pathway. The use of bioinformatics studies along with new techniques can be very helpful in increasing the speed and more accurate identification of biomarkers. Finally, the results of this study showed that bta-miR2410 by targeting the PTEN gene, which is one of the important genes involved in important biological processes of milk production, can alter milk production by altering the expression of PTEN gene in dairy cattle. Therefore, this gene can be further studied as a biomarker in future studies.

Key Words: *Dairy cattle, Genetic interaction, miRNA, Milk production*

Detection and evaluation of protein complexes affecting on the skeletal muscle of beef cattle

Zahra Roudbari[1], Saideh Eskandarynasab[2]

[1] *Department of Animal Science, Faculty of Agriculture, University of Jiroft, Jiroft, Iran*
[2] *Department of Animal Science, Faculty of Agriculture, Zabol University, Zabol, Iran*

E-mail address (corresponding author): rodbari.zahra@gmail.com

Abstract

Protein complexes are a collection of proteins that interact at a specific time and site to perform a specific biological process. proteins as single units cannot perform well, but what causes a particular phenotype is a set of protein interactions. one of the traits that has been considered in the livestock industry is skeletal muscle growth and meat quality. growth traits are quantity variables that are controlled by many genes. the propose of this study is the detection and study of protein complexes in protein-protein interaction networks related to beef skeletal muscle. gene expression data with GSE25554 access number were downloaded from GEO database. These data are related to the muscle tissue of beef cattle. filters for identifying genes with different and significant expression were (-2> fold Chang> 2) and (Adj P-Value <0.05). Which was done using the Limma package available in R software. Protein complexes were determined by the MCODE plugin in Cytoscape software. In these complexes, the proteins with the highest communication score were squared and introduced as seeds. DAVID database was used to identify the activity of seed genes. in the present study, 143 genes were examined, and three complexes were formed, in which the seed genes include FASN, FN1, and SDHC. The results of gene activity showed that FN1 gene has various roles including cell adhesion and cell growth. The FASN gene is involved in biological processes such as lipid transport, lipid metabolism and muscle growth, it also affects the quality of meat. The SDHC gene is involved in the mitochondrial electron transfer chain and is responsible for the transfer of electrons from succinate to coenzyme Q. Therefore, these genes can be introduced as the key genes that control the skeletal muscle growth of beef cattle.

Key Words: *Protein Complex ،Muscle Growth ،PPI network*

Investigating the differential expression of starch and transcription factor encoding genes and their relations in wheat cultivars with different bread quality at 5 days post anthesis

Zahra Sabili[1,2], Shabnam Hasrak[2], Tahmineh Lohrasebi[2]*, Sajad Rashidi-Monfared[1]

[1] *Agricultural Biotechnology Department, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran*
[2] *Department of Plant Bio-products, Institute of Agricultural Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran*

E-mail address (corresponding author): tlohrasebi@gmail.com

Abstract

Wheat starch is one of the most important sources of carbohydrates, since every study about the regulating factors of starch metabolism can improve starch quantity and quality. The aim of this research is studying the relation between differential expressed starch and transcription factor (TF) genes at 5-days post anthesis in wheat cultivars with high (Pishtaz) and low (Navid) bread quality.
The transcriptome of cultivars compared by RNA-Seq. and based on wheat reference genome. The result showed “Alpha-glucosidase”, “Endoglucanases” and “1,4-alpha-glucan branching enzyme” genes that involved in starch metabolism had differential expression. In addition, “General transcription factor IIH subunit 5 (Tfb5)”, “GTE9” and “MADS box” TFs showed differential expression. The 1200 upstream nucleotide of starch gene employed for searching cis-regulating motifs in Place and PlantCare databases. The DNA binding motif of differential TF identified based on literature search. We found CArG box DNA binding site for MADS box, HD1 and HD2 for Tfb5 and transcriptionally competent chromatin binding for GTE9. Alpha-glucosidase with lower and Endoglucanase with higher expressions in Pishtaz, have cis-acting elements motif for binding the Tfb5. This TF contributes in transcription re-initiation, early elongation steps and DNA repair. The reduced levels of Tfb5 and Alpha-glucosidase can guide us to assume a regulating role for this TF on Alpha-glucosidase gene. 1,4-alpha-glucan branching enzyme and Endoglucanase have lower and higher expression, respectively. We found a motif for binding the MADS box that actives in transcription activating , regulation of flower, leaf, and root development , growth negative regulators, positive regulators of stress tolerance. The decrease in 1,4-alpha-glucan gene can be contributed to lower expression of MADS box.It seems that MADS box and Tfb5 negatively controls Endoglucanase expression, has not affinity for binding to its upstream or this gene is not their target genes.

Key Words: *Triticum aestivum, Bromodomain, cis-regulating, Alpha-glucosidase, Endoglucanase.*

Studying the relation between differential expressed starch genes and bread quality in two wheat cultivars from 5 to 28 days post anthesis

Zahra Sabili[1´2], Shabnam Hasrak[2], Tahmineh Lohrasebi[2]*, Sajad Rashidi-Monfared[1]

[1] Agricultural Biotechnology Department, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran
[2] Department of Plant Bio-products, Institute of Agricultural Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

E-mail address (corresponding author): tlohrasebi@gmail.com

Abstract

Wheat endosperm, which is composed of starch, proteins and lipids, is the main flour source in bread industry. The amount and type of starch and its interactions with gluten proteins play key roles in the quality of bread. The aim of this research was to study genes with differential expression and their affected pathways on final bread quality in two wheat cultivars with high (Pishtaz) and low (Navid) bread quality at 5, 10, 14, 21 and 28 days post anthesis (DPA). The gene expression determined througth RNA-sequencing technique and diferentially expressed genes identified by downstream bioinformatic analysis. Starch encoding genes selected and their functions and related pathways identified through EMBL-EBI, Pfam and KEGG databases. Results showed that genes encodying Endoglucanase, Beta-fructofuranosidase1, Alpha-amylase and Enolase had higher expression in Pishtaz compared to Navid at 5, 10, 14, 21 and 28 DPA, respectively. These genes are involved in the cellulose hydrolysis, sucrose hydrolysis, starch degradation and glycolysis process, respectively. Alpha-glucosidase at 5 and 10 DPA, 1,4-alpha-glucan-branching enzyme at 5 DPA and Alpha-amylase at 28 DPA are genes with increased expression in Navid compared to Pishtaz and have roles in maltose hydrolysis , branching of starch polymer chains and starch degradation, respectively. Based on the results, it seems that processes linked with starch biosynthesis at 5 and 10 DPA and starch degradation at 14 and 21 DPA have higher activity and produce downstream products in Pishtaz. Also, starch degradation activity decreases and pyruvate accumulation significantly accured in Pishtaz at 28 DPA. In Pishtaz, contrary to Navid, the reactions are generally towards starch biosynthesis, branching its carbohydrate polymers especially crystalline structure and higher starch content in endosperm. This is consistent with previous results that Pishtaz has higher crystallinity in its starch structure and bread quality.

Key Words: *Triticum aestivum, Anthesis, Grain developement, RNA sequencing, Crystallinity*

correlation between BP1 and SPL1 and protein phosphorylation in spms(multiple sclerosis)

Armin Sharifi, mehran radak, Shahab Atefi

E-mail address (corresponding author): arminsharifi1222@gmail.com

Abstract

Introduction: MS, as a chronic autoimmune disease with unknown etiology, is the most common neurological disorder in young adults. It is characterized by the infiltration of auto-reactive immune cells into the central nervous system (CNS), which leads to inflammation, demyelination, and axonal degeneration. Several studies suggest that genetic, environmental, epigenetic and infectious factors may contribute to the etiology and pathogenesis of MS. This essay survey death ratio reason of SPMS patient.
Material and method: In order to start this study, we used GEO accession number GSE131282. In our work, we focus on patient with less disease duration, then by the use of GEO2R web tool, we analyzed the information. Eventually, for every gene list we removed differentially expressed genes (DEGs) which had p-value more than 0.05 and log 2 FC between ±0.6. protein-protein interaction (PPI) was used to find the proteins that is coded by DEGs, thereby string data base was utilized for this aim. The list of DEGs was deposited in this data base and output, PPI network, was received. To analyze PPI networks, Cytoscape 3.4.0 and Gephi software v 0.9.1 were used to visualization of networks.
Result: samples of patients and controls were compared to identify DEGs in each disease. The number of DEGs were 801 genes in SPMS which 356 were upregulated and 484 were downregulated. Analysis of modules and their functional annotation in SPMS show G-protein coupled receptor signaling pathway and extracellular matrix organization and etc.
Conclusion: This essay demonstrated that BP1 and SPL1 play crucial role in mortality of SPMS. positive regulation of gene expression and protein phosphorylation have limited in case of PPMS.

Key Words: *#protein phosphorylation#multiple sclerosis# BP1#SPL1*

New insights into the fusarium wilt resistance in chickpea using genomics and transcriptomics meta-analysis

Jahad Soorni[1], Fatemeh Loni[1], Parisa Daryani[1], Nazanin Amirbakhtiar[2], Zahra-Sadat Shobbar[1]*

[1] *Agricultural Biotechnology Research Institute of Iran (ABRII), Department of Systems Biology, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran*

[2] *National Plant Gene Bank of Iran, Seed and Plant Improvement Institute (SPII) Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran,*

E-mail address (corresponding author): shobbar@abrii.ac.ir

Abstract

Chickpea (Cicer arietinum L.) is one of the most important grain legumes providing a major source of dietry protein. Fusarium wilt (FW) caused by eight Fusarium oxysporum f. sp. ciceri pathogenic races affect chickpeas causing a great yield loss wordwide. In the present study, an integrated meta-analysis approach was used to find genomic regions and candidate genes involved in resistance to fusarium wilt. All the reported quantitative trait loci (QTLs) and transcroptomics data including RNA-seq and SAGE studies related to FW resistance were collected. A meta-analysis was performed using 29 initial QTLs associated with different races of FW resistance reported so far through nine independent experiments by Biomercator which led to indentification of seven meta-QTLs. The genes located in these hotspot regions were retrieved through Ensembl Plants and NCBI databases. Moreover, the differentially expressed genes in chickpea were found through analysis of the related transcroptomics data. Conclusively, 981 FW-resistance related genes were identified using meta-QTL analysis and compared with 291 and 5183 FW-responsive genes detected through RNAseq and SAGE techniques. The integrated genome and transcriptome analyses revealed a number of candidate genes underlying resistance to FW races, which included some genes from MAP kinase, serine threonine kinase, WRKY transcription factors, antioxidant enzymes and NBS-LRR gene families. These candidate genes/regions can be used for development of FW-resistant cultivars through genetics engineering or molecular breeding.

Key Words: *Meta-QTL; RNA-seq; Cicer arietinum; fusarium race; biotic stress*

Biological Sequence Analysis

Investigation of phylogenetic relationship of some isolates of Magnaporthe oryzae in different hosts

Kh. Abbasi[1]*, M. Mirab-balou[1], S. Afzalinia[2], T. Alizamani[3]

[1] *Department of Plant Protection, College of Agriculture, Ilam University, Ilam, Iran*
[2] *Department of Plant Protection, College of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz. Iran*
[3] *Department of Plant Protection, College of Agriculture, Lorestan University, Lorestan, Iran*

E-mail address (corresponding author): kh.abasi@ilam.ac.ir

Abstract

Rice is one of the most important agricultural products having more than 90% of its world production in Asia. However, a large amount of that is wasted by pests and diseases. Blast is an influential disease of rice, responsible for a significant yield loss of about 10% to 30% of total rice production every year. The fungal plant pathogen, Magnaporthe oryzae, is the causal agent of the devastating rice blast disease. It can infect millet and other grasses, but rice is the most economically significant host. In this study, the phylogenetic relationships of isolates obtained from rice and other wheatgrass are compared based on the ITS regions. The ITS nucleotide sequences were selected from the NCBI gene bank; Then, a strain of Gaeumannomyces graminis (GGG317 18S) was used as an out-group array for rooting deciduous trees. The obtained sequences were aligned using BioEdit v, 0,9 software and the Clustalw algorithm. Phylogenetic analysis was performed based on the sequence of ITS region using the maximum likelihood algorithm. Also, the validation test of 1000 replications was carried out by MEGA.6 software. ITS-based phylogenetic analysis of M. oryzae isolates, which were isolated from different grasses, were classified into four groups based on their host: group A (Rice), group B (Wheat), group C (Corn, Barley, and so on), and D (Rice). The results indicated that M. oryzae was identified for rice isolates in groups A and D, but the Blast agent of Iranian rice was not included in these categories. According to our studies, it can be found that the phylogenetic properties of M. oryzae in Iranian rice are similar to group C. Therefore, we conclude the Blast agent in Iran will probably be different from other regions.

Key Words: *Iran; ITS region; Phylogenetics; Rice Blast*

Bioinformatics analysis of regulatory miRNAs and target lncRNAs in cervical cancer

Sama Abbasi[1]*, Hassan Rajabi-Maham[1], Leila Kianmehr[1]

[1] *Animal Sciences and Marine Biology Department, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, G.C, Tehran[1983963411], Iran*

E-mail address (corresponding author): samaabbasi976@gmail.com

Abstract

Cervical cancer is one of the leading causes of mortality in women, worldwide. The involvement of non-coding RNAs in cancer addressed extensively in recent years. In the present study, the bioinformatics analysis of microarray data including two-gene dataset (GSE9750 and GSE52903) and one miRNA dataset (GSE30656) performed. Firstly, differentially expression analysis among cancer samples versus healthy controls showed 64 up-regulated and 106 down-regulated genes. PPI network analysis determined 28 genes as hub genes (degree about 48-54) which play a significant role in cell cycle pathways and cell division. Differential analysis of miRNAs dataset also identified 10 significant difference. Meta-analysis was performed between the three selected datasets and the relationship between hub-genes and miRNAs was examined. Interaction between 4 candidate miRNAs and 139 associated lncRNAs predicted in cervical tissues by miRNet online software. As a result, 19 of these lncRNAs with the highest degree of interaction selected. The present study has investigated the interaction of involved miRNAs and lncRNAs in cervical cancer. In conclusion, bioinformatics analysis could help us to suggest the different expressions and interactions of most important genes, regulatory miRNAs, and target lncRNAs. However, these findings should be further considered under experimental approaches although many reports confirm the involved pathways and genes in cervical cancer.

Key Words: *Cervical cancer, Microarray data, Bioinformatics, miRNA, LncRNA*

Evaluation of genetic mutation PLA2G6 gene in Parkinson's disease by WES method with using bioinformatics analysis

Shima Abbasnejad, Zahra Rezvani

E-mail address (corresponding author): shimaabbas535@gmail.com

Abstract

Introduction: In the last decade, (WES) has become a common tool for identifying genetic causes of human inherited disorders, and it has also recently been applied to Parkinson disease research. The recent research focus is to identify the function of genes and proteins involved in the genetic forms of Parkinson's disease that will help us better understand the genetic pathogenesis of the disease. This study provides a detailed analysis of mutations in the genes associated with Parkinson's disease that significantly advance our knowledge of clinical changes, brain imaging, and pathological features of the disease.

Materials and Methods: Using the WES method, a file related to the PLA2G6 gene was prepared. First we check it with 1000GENOME, ESP 6500 databases. Then using reported mutations in known genes according to the Human Genome Mutations Database are available from http://www.hgmd.org. For bioinformatics analysis, all genes and mutations of this disease were first extracted using the HGMD site. Then, the effect of mutations on disease was evaluated using POLY PHEN and SIFT site separately. Common mutations were extracted and analyzed statistically.

Results: Genes associate with Parkinson's disease was identified as follows: PLA2G6 autosomal recessive inheritance pattern are inherited from a patient's parent with muscle weakness and neur developmental regression

Conclusion: After analysis, the genes associated with Parkinson's disease with WES method and bioinformatics sites were identified as follows: PLA2G6 (C.C1715T/P.T572I) is common in many patients will determine role in the pathophysiology of Parkinson's disease Kalinderi K, Bostantjopoulou S, Fidani L. The genetic background of Parkinson's disease: current progress and future prospects. Acta Neurol Scand. (2016) 134:314–26. doi: 10.1111/ane.12563 Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain. 2002 Apr;125(Pt 4):861-70. doi: 10.1093/brain/awf080. PMID: 11912118.

Key Words: *Whole Exome Sequencing (WES), Parkinson's disease, bioinformatics analysis, PLA2G6,*

Bioinformatics analysis of changes FNDC5 gene expression in skeletal muscle of whey protein-treated mice

Alieh abdolrezaie[1]*, Hashem Nayeri[2], Ali Mohammad Ahadi[3]

[1] *Department of Biochemistry, Falavarjan Branch, Islamic Azad University, Isfahan, Iran*
[2] *Department of Biochemistry, Falavarjan Branch, Islamic Azad University, Isfahan, Iran*
[3] *Department of Genetic, School of Science, Shahr-e-kord University, Shahr-e-Kord, Iran*

E-mail address (corresponding author): a.abdolrezaie@yahoo.com

Abstract

Background: Irisin production is a metabolic pathway that results from the poroteolytic breakdown of the FNDC5 protein, which is secreted from skeletal muscle as an exercise hormone during exercise. Considering that food compounds can affect metabolic pathways, gene expression and protein production, the aim of this study was to investigate the results of the effect of whey protein on the expression of some genes, production of dependent proteins and the study of thermogenesis in skeletal muscle.
Materials and Methods: In this study, 20 male mice were divided into four groups: placebo, supplementation treatment, exercise and simultaneous supplementation and exercise treatment. Supplementation was performed with whey protein. Then muscle tissue samples were taken and the expression of FNDC5 gene in muscle was evaluated. Finally, we were investigated the correlation of FNDC5 gene expression with genes involved in skeletal muscle thermogenesis, protein interaction and metabolic pathways with string databases and cytoscape software.
Results: The results showed that whey protein supplementation had a positive effect on FNDC5 gene expression and bioinformatics analysis confirmed the correlation between irisin levels, gene expression and thermogenesis. Also found that UCP1 and PPARa are genes in the process of thermogenesis in skeletal muscle that are associated with FNDC5 / Irisin in this metabolic pathway.
Conclusion: Protein compounds with their structure and components have a great impact on the expression of genes and metabolism in the body. Therefore, it was found that whey protein increases the level of irisin by activating molecular signaling pathways and increasing gene expression. Today, the use of protein is a strategy to treat obesity and improve athletic performance, and protein supplements seem to be the most suitable combination for achieving high doses of Irisin and thermogenesis.

Key Words: *Key words: FNDC5, Bioinformatics, Gene expression, Irisin, Whey protein*

Bioinformatics evolution of an lncRNA signaling pathway and its related function in colorectal cancer

Irsa Alborzi[1], Maryam Modrek[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and Vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Genetic investigations have discovered long non-coding RNA (lncRNA) genes in the human genome that do not encode proteins. The non-coding genome, mutations, single nucleotide polymorphisms, copy-number variations, and epigenetic modifications enhance the possibility of altering lncRNA expression levels, which can lead to cell cycle dysregulation and possibly cancer. Our aim of this study was to investigate the role an LncRNA in colorectal cancer (CRC). Through this study, the NCBI obtain some information about colorectal cancer. LncRNAsnp2 database helped to find lncRNA associated with CRC. The interactions between LncRNAs and selected LncRNA with other diseases were found from LncRNAdisease2 database. Based on our research, MALAT-1 (ID :NONHSAT022126.2) was the potential lncRNA in CRC. Another lncRNAs such as SCYL1, LTBP3, SSSCA1, FAM89B, KCNK7, MAP3K11 are involved in the process of CRC as well. MALAT-1 plays role in other diseases such as gallbladder cancer, Pancreatic cancer, and hepatocellular carcinoma too. It is inferred that overexpression of MALAT-1 could promote cell proliferation and promote tumor growth and metastasis in CRC. The results reveal that one of the five pieces (6918 nt-8441 nt) at the 3' end of MALAT-1 plays an essential role in biological processes such as cell migration, invasion, and proliferation. The underlying mechanism was associated with SFPQ gene as a tumour suppressor and PTBP2 as a proto-oncogene. MALAT-1 should bind to SFPQ, freeing PTBP2 and growing cell proliferation and migration. It is concluded that lncRNAs manage numerous vital most cancers traits via their interplay with different mobile macromolecules and satisfactory law of lncRNA transcription would possibly offer indicators of malignant transformation. Recent advances in expertise in the molecular mechanics of lncRNAs have given the capacity to functionally annotate cancer-associated transcripts, making those molecules attractive healing goals withinside the combat in opposition to cancer.

Key Words: *Long Non-coding RNA; MALAT-1; Tumorigenesis; Databases*

Distribution of HLA alleles in malignant tissues of patients with hepatocellular carcinoma

Shahram Aliyari[1´2], Zahra Salehi[2], Mohammad Hossein Norouzi-Beirami[3], Kaveh Kavousi[2]*

[1] *Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran*
[2] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[3] *Department of Computer Engineering, Osku Branch, Islamic Azad University, Osku, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Hepatocellular carcinoma (HCC) is one of the most common malignancies and a leading cause of cancer-related deaths worldwide. Despite evident role of human leukocyte antigen (HLA) in immune detection and killing of cancer cells, patients. HLA genotypes has not been fully characterized in HCC. The present study aimed to investigate the distribution of HLA allele frequency in HCC patients by in Silico HLA typing using RNA-Sequencing (RNAseq) sequence reads. We inferred the HLA genotypes across 371 TCGA-LIHC primary tumor samples using arcasHLA tool from input RNA sequencing data. A total of 58 HLA-A, 87 HLA-B, 52 HLA-C, 41 HLA-DPB1, 27 HLA-DQA, 43 HLA-DQB1, and 43 HLA-DRB1 alleles were identified in the HCC population. The most frequent allele in the HLA-A, HLA-B, HLA-C, HLA-DPB1, HLA-DQA, HLA-DQB1, and HLA-DRB1 loci were (A*02:01:01, 14.5 %), (B*07:02:0, 8.7%), (C*07:02:01, 12.5%), (DPB1*04:01:01, 28.3%), (DQA1*01:02:01, 15%), (DQB1*03:01:01, 14.8%), and (DRB1*07:01:0, 10.2%), respectively. HLA-B (89%) and HLA-DQB1 (90%) showed the highest level of heterozygosity. Together, our findings showed high prevalence of some HLA alleles with probable clinical significance in HCC. To provide a better plan for disease prognosis and immunotherapy treatment further evaluations of the identified alleles, particularly in patients as well as healthy controls, is necessary.

Key Words: *Hepatocellular carcinoma; HLA; Allele; RNA-Sequencing*

Whole genome analysis of fungal plant pathogens to find pathogenicity genes

Adel Pordel[1], Amirreza Amirmijani[2], Khaled Miri[3]

[1] Plant Protection Research Department, Baluchestan Agriculture and Natural Resources Research and Education Centre, AREEO, Iranshahr, Iran
[2] Department of Plant Protection, University of Jiroft, Jiroft, Iran
[3] Research Department, Seed and Plant Improvement, Baluchestan Agricultural and Natural Resources Research and Education Center, AREEO, Iranshahr, Iran

E-mail address (corresponding author): ar.amirmijani@ujiroft.ac.ir

Abstract

Sequencing the fungal genomes reveals extensive variability in the structure and dynamics of the genome to better understand the mechanisms that have led to the increase in disease incidence related to pathogenicity genes. The genomes exhibit different numbers of chromosomes and other establishments of chromosomal, defining so-called accessory compartments that are crucial to pathogenicity in plant-infecting fungi. Pathogens produce small proteins called effectors that exert their activity in the host's extracellular environment or inside host cells and manipulate the plant immune system. A significant interest in sequencing the genomes has understood the localization of effector genes in pathogen genomes. Effectors trigger plant immune responses that are molecularly similar and often identical in pathogenic and beneficial microbes. The first layer of biochemical plant defense is achieved by detecting pathogen-associated molecular patterns (PAMPs) and microbe-associated molecular patterns (MAMPs). We are focusing on recent advances in genome sequencing technologies, genome alignment and assembly (MACSE and ABySS), gene annotation (AUGUSTUS), and effector identification methods (Proteinortho, OrthoFinder, Signal p, Go ontology, Ensemble, and Swissprot) that hold promise to disclose complete and correct effector repertoires. The whole-genome analysis allows exploiting entire effector repertoires and knowledge of their diversity within pathogen populations to develop durable and sustainable resistance breeding strategies, disease control, and management of plant pathogens.

Key Words: *genome sequencing, plant-microbe interaction, pathogenicity genes, era.*

The importance of DNA sequencing on taxonomy of the genus Cladosporium

Amirreza Amirmijani

Department of Plant Protection, Faculty of Agriculture, University of Jiroft, Jiroft, Iran

E-mail address (corresponding author): ar.amirmijani@ujiroft.ac.ir

Abstract

Cladosporium is one of the largest hyphomycetous genera that the species of its are distinguishable from allied genera by having coronate conidial hila. By introducing the new concept of Cladosporium, the taxonomy of this genus has changed and in many papers, various species have been re-examined and described. Finely, these studies have led to publish a monograph of this genus, viz "the genus Cladosporium" by Bensch et al. in 2012. Due to limitation in the use of morphological traits, identification of most species are difficult. In recent decays, the use of DNA-based methods such as DNA barcoding has played a great role in identification of species especially cryptic species in most fungal groups. According numerous studies, molecular data has led to the description of new species in the genus Cladosporium. Considering the new concept of Cladosporium, taxonomy of this genus has started using morphological and molecular data in Iran, and these studies will be continued until reaching a comprehensive monograph.

Key Words: *DNA barcoding, identification, molecular data, new species*

Demonstration of cellular dynamic behavior in response to DNA damage through gene regulatory network analysis of single-cell resolution DNA repair phenotype and gene expression multi-omics dataset

Abolfazl Arab[1]*, Sama Goliaei[1], Faramarz Mehrnejad[1]

[1] *Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): abarbiology@gmail.com

Abstract

A lot of genes may involve in the response of cells to DNA damage occurrence and regulation of these genes are complicated at different levels –. Also, it is known that their missing activity has important roles in cancer diseases and treatment. Droplet microfluidics and Next Generation Sequencing (NGS) are enabling technologies to evaluate genomic complexity in single-cell resolution. In general, these methods extract a large amount of data from different layers of biological information which opens computational challenges to analyze these big data toward biological interpretations. Richer et. al, 2020, used nano-biomimetic substrates containing arbitrary lesion DNA damage to quantitatively evaluate the relationship of gene expression and repair phenotype in single-cell resolution. Among their data, we have selected time-series experiments on the HAP1 cancer cell line and designed our analysis workflows to reanalyze the available multi-omics data. Basically, for each cell, we have both gene expression profile and DNA repair measurement simultaneously. We used statistical inference, mathematical modeling, and graph analysis of context-specific gene regulatory networks (GRN) to mine this data –. Also, we are connecting our data with publicly available datasets such as protein-protein interaction (PPI) and molecular signature database (MSigDB).

Key Words: *Single-cell; RNA expression; RNA-seq; Omics; Multi-omics; Enzyme activity; DNA Damage; DNA Repair; Gene Regulatory Network; Gene-set analysis; Network analysis*

the relation between aging process and contributor factors

Shahab Atefi, Armin Sharifi, Mehran Radak

E-mail address (corresponding author): atefishahab@gmail.com

Abstract

No one dares to deny the importance of aging in the different systems that balance our body and mind arterial system pressure affect the elasticity and the ability in a aged elderly people due to missing multi factors.

In our study, we use Meta annalistic vision to fond essential differentially transcription factor in aging process of cardiac mesenchymal stromal cells.

Methods: We used the GSE129656 dataset that information is available free for all in the NCBI database. The data set was analyzed through using a lima package embedded in the GEO2R NCBI instrument to determine significant differentially expressed genes (DEGs) with a P-value of less than 0.05 in old cardiac mesenchymal stromal cells (cMSCs) cells against young cMSCs. Next, select genes that have a 0.6 change in log2 Fold. After that, we estimate the transcription factors for the various genes expressed on the Enricher site with ChEA2016. Finally, we used Cytoscape 3.4.0 and Gephi 0.9.2 software to construct and analyze a network of differentially expressed transcription factors (DE-TF).

Results: The results showed that there is a transcription factors with different expression (DE-TF) In senescent cells. In addition, the DE-TF is JARID2 which regulation of this was less.

Conclusion: this essay suggests that JARID2 has vital function in aging process of cardiac mesenchymal stromal cells.

Key Words: *#aging process#JARID2#cardiac mesenchymal stromal cells.*

Bioinformatics study of level expression of miR-212 as a biomarker of Glioblastoma

Ali Attaripour Isfahani[1], Mohammad Sina Pirmoradian[1], Elnaz Nasiriyan[1], Zahra Rafiezadeh[1], Mansoureh Azadeh[2]*

[1] *Faculty of Sciences, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran*
[2] *Zist-fanavari Novin Biotechnology Department, State Technical and Vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): bionovin@gmail.com

Abstract

Introduction: MicroRNAs are small non-coding nucleic acids (around 20 nucleotides) that are able to post-transcriptionally regulate gene expression by binding to complementary sequences of target messenger RNA. In this study, miR-212 expression levels, target genes and genetic pathways in glioblastoma were investigated using a bioinformatics database.
Methods: This bioinformatic study was performed by several main databases, such as "miRdSNP, mirbase, GeneCards and DAVID". At first from mirbase, the chromosomal profile of this miRNA was obtained. The position of miR-212 on the human chromosome and complete information about that, was identified In GeneCards. Finally the DAVID database was used for genetic pathways evaluation.
Results and Discission: MiR-212 gene is located on chromosome 17p13.3. miR-212 is overexpressed in cancers of the central nervous system and is found in various parts of the cell, particularly in the extracellular space and plasma membrane. miR-212 plays a role in the post-transcriptional pathway of 2950 genes of cellular sections. In addition, of the 2562 identified miR-212 gene targets, which with influence on these genes could result in a variety of diseases, it can act as a tumor suppressor in glioblastoma by effecting on 53 genes. It was also found that miR-212 by Suppression of SGK3 gene as an important tumor-promoting gene is directly involved in the suppression of glioblastoma. Based on these results, miR-212 inhibits glioblastoma cell proliferation.
Conclusion: The current study shows that, increased expression of miR-212 in brain cells as a tumor suppressor, can indicate the presence of cerebral glioblastoma, which may also lead to a rapid diagnosis of this chronic disease. Especially, since miR-212 can also be detected in serum and plasma, which are much more readily obtainable than tissues, it attracts increased clinical attention as a biomarker in glioblastoma.

Key Words: *Key words: Glioblastoma, miR-212, Tumor suppressor, Tumor cell biomarker*

Bioinformatics and phylogeny analysis on phytoanticipine tomatinase

Akram bakhshi, mohhamad Ali Tajik Ghanbari

E-mail address (corresponding author): akrambakhshi100@gmail.com

Abstract

Plant pathology is the science that studies the causes of plant diseases, the mechanisms by which diseases develop in individual plants and in plant populations, and the ways and means by which plant diseases can be managed or controlled. Bioinformatics applies information technologies to the allied fields of agriculture, horticulture, forestry, biotechnology, microbiology, plant physiology and molecular biology. Bioinformatics devises strategies for data management, analysis and integration tools that enable rapid scientific discovery and informed decision making. In plant pathology, the 'contemporary' application stage of bioinformatics is typically after a pathogen has been identified as a causative agent for a given plant host and subjected to biotechnological studies. pathogenicity is defined as the capability of a pathogen to cause disease, and virulence is defined as "the degree of pathogenicity of a given pathogen. α-Tomatine is a saponin found in tomato plants, in high concentrations Many fungal pathogens of tomato produce extracellular enzymes(phytoanticipin), collectively known as tomatinase, that detoxify the preformed antifungal steroidal glycoalkaloid alpha- tomatine. In the present study, amino acid sequences related to proteins involved in tomatinase synthesis in several genera of fungi and bacteria plant pathogenic were extracted from databases and clustered by MEGA7 software. Comparison of sequences at different levels showed significant similarities and differences, including the placement of each gene in different clones and the similarity of sequences in the genera Clavibacter and Streptomyces, the similarity of sequences between species. Fusarium tricinctum, Colletotrichum truncatum as well as Clavibacter michiganensis. Given the importance of phytoanticipin, these differences and similarities may be candidates for further bioinformatics studies.

Key Words: *Tomatinase, MEGA7, Phylogeny, phytonaticipin*

Investigation of IL-17RA editing sites in patients with severe COVID-19

Javad Behroozi[1]*, Mehrdad Nasrollahzadeh Sabet[1], Mojtaba Yousefi zoshk[2], Ebrahim Hazrati[3], Peyman Aslani[4], Reza Mosaed[5]

[1] *Department of Genetics and Advanced Medical Technology, Faculty of Medicine, AJA University of Medical Sciences, Tehran, Iran*
[2] *Department of Pediatrics, Faculty of Medicine, AJA University of Medical Sciences, Tehran, Iran*
[3] *Department of Anesthesiology and Critical Care, Aja University of Medical Sciences, Tehran, Iran*
[4] *Department of Parasitology and Mycology, Faculty of Medicine, Aja University of Medical Sciences, Tehran, Iran*
[5] *Department of Clinical Pharmacy, Faculty of Medicine, Aja University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): javad_behroozi65@yahoo.com

Abstract

Interleukin 17 (IL-17) family is highly multi-purpose pro-inflammatory cytokines critical for a variety of processes, including host defense and pathogenesis of inflammatory disease. Interleukin 17 receptors (IL-17R) consists of 5 members and despite these functions, regulating immune and inflammatory responses, are another role of IL-17R. The aim of this study was to investigate IL-17RA editing sites in patients with severe COVID-19. RNA-seq raw data related to 40 COVID-19 hospitalized patients and 10 healthy donors were retrieved from the publicly available Gene Expression Omnibus (GEO) database. Quality control, alignment and variant calling were performed using FastQC, Hisat2 and Freebayes tools, respectively. RNA-DNA differences were filtered to remove known SNPs and low quality SNVs. Additionally, several quality-aware filtering steps were employed to increase the accuracy of identifying true RNA editing sites. Finaly, statistical significance for differences between patients and healthy donors editing ratios was assessed by the T test. Our results showed difference in 7 A to G editing sites on the 3'UTR of IL-17RA gene between patients and healthy donors ($P<0.05$). These editing sites could affect microRNA target recognition and subsequently affect the expression profile of IL-17RA. RNA editing of IL-17RA significantly differ between hospitalized COVID-19 patients and heathy peoples and anti-IL-17RA therapy in COVID-19 should be considered.

Key Words: *RNA-sequencing; COVID-19; RNA editing; IL-17RA.*

Identification of key transcription factors and microRNAs associated with drought stress in rice: a systems biology approach

Ali Asghar Dehghan[1]*, Ali Moghadam[2], Akbar Karegar[1], Yaser Biniaz[3], Fateme Amiri[1]

[1] *Department of Plant Protection, School of Agriculture, Shiraz University*
[2] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*
[3] *Plant Virology Research Center, Shiraz University, Shiraz, Iran*

E-mail address (corresponding author): aliasghardehghan96@gmail.com

Abstract

Drought is one of the major environmental risks that decreases the successful production of crops, especially rice, in the world, which can occur at any time during the growing season. Hence, one of the main challenges in agriculture is to produce more food with less water. Therefore, study of molecular mechanisms in order to characterization of key gene networks responsive to this stress is more considered. For this purpose, we retrieved the transcriptome RNA seq data from the European Bioinformatics Institute (EBI) database. Data quality analysis and final analysis were performed using CLC Genomics Workbench 20.0. The reference genome as well as the rice annotation were downloaded from the NCBI. Adjusted P-values (FDR < 0.001) were considered significant and, differentially expressed genes (DEGs) were used for further analysis. Analysis of GO enrichment for DEGs showed that transcripts were more involved in cellular processes such as cellular process and localization. In addition, In the category of cellular components, DEGs were enriched for cellular anatomical presence and intracellular anatomical structure. In the category of molecular function most genes were referred to ion membrane transmitter activity, transmembrane transporter activity of molecular entity. Based on the results of Kyoto Encyclopedia of Genes and Genomes analysis, it was concluded that endocytosis and mRNA surveillance pathways, these two had a greater role in the plant's response to drought stress. In microRNA analysis, miR156, miR160, miR171, and miR172 showed the highest expression in drought stress conditions. In the analysis of transcription factors, it was shown that the bHLH family had the highest expression in both down-regulated and up-regulated genes in response to drought stress. Also, transcription factors bZIP, WRKY, MYB, C2H2, which play a key role in drought resistance and stress, had the highest increase in expression.

Key Words: *Key words: Abiotic stress, microRNA, System biology, Drought*

A Fully Automatic Registration and Analysis Platform for Next Generation Sequencing Data

Toktam Dehghani[1]*, Reza Jafarzadeh Esfehani[2], Maliheh Alimardani[2,3,4], Maryam Gholizadeh[1], Ariane Sadr-Nabavi[3], Mohammad Reza Abbaszadegan[2,3], Saeid Eslami[1]*

[1] *Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
[2] *Medical Genetics Research Center, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
[3] *Department of Medical Genetics and Molecular Medicine, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
[4] *Student Research Committee, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*

E-mail address (corresponding author): edu.dehghani@gmail.com

Abstract

Background: Clinical genetics plays a crucial role in the healthcare systems by providing valuable information for the investigation of genetic variations, prevention and diagnosis of genetic disorders, and determination of the best treatments for patients. Next Generation Sequencing (NGS) is a popular technology for sequencing patients' whole exome. NGS is a rapid, high-throughput, and cost-effective approach for genetic testing. Despite the growth of NGS utilization in clinical genetics, major challenges occur in registration and analysis of the NGS data, such as dealing with the huge amount of data, the requirement of high-level computational systems and programming skills, the variety of bioinformatics tools, and the low quality of results visualization. Thus, further improvements in algorithms and pipelines are still essential.
Results: We provide an online and totally automatic framework for the registration and analysis of NGS data to make this procedure simpler and easier. In this platform, five major goals are pursued. (1) The presented platform is simple and fully automated; it reduces the need for human resources and the genetic database by developing a user-friendly graphical interface. (2) It achieves reliability and accuracy through performing standard analysis, such as minimal pipeline, gene panels, incidental finding, and carrier screening analysis. (3) Traceability and reproducibility are achieved by storing and reporting via a secure website, which could be practical for re-analyzing data. (4) All outputs and results are graphically displayed. (5) It makes shareable results; details of all steps could be shared with collaborators or publications.
Conclusion: Therefore, this automatic platform would enhance registration, analysis, visualization, and sharing of NGS data. It could improve the decision-making process of genetic disorder diagnosis and treatment and could reach a new era in clinical genetics research. The platform is available at https://www.SmartDXCloud.ir.

Key Words: *Automatic Platform, Pipeline, Genetic Variants, Next Generation Sequencing (NGS), Whole-Exome Sequencing (WES), Personalized Genomic Medicine, User-friendly Graphical Interface.*

EDC-Protein network formation analysis in genetic response of human epithelial cells to SteA

Nazila Eslami[1]

[1]*Department of Biology, Tabriz Branch, Islamic Azad University, Tabriz, Iran*

E-mail address (corresponding author): nazilaeslami2@gmail.com

Abstract

Salmonella enterica serovar Typhimurium is a gram-negative facultative anaerobic enteric pathogen in humans and animals, and a leading cause of gastroenteritis. The Salmonella pathogenicity island-1(SPI-1) type III secretion system (T3SS) is critical for invasion of host cells via the trigger mechanism by deploying a macropiocytosis-related process in enterocytes and the SPI-2 of the T3SS is responsible for the zipper mechanism and intracellular survival of Salmonella Typhimurium [2,3]. These systems translocate proteins called effectors into eukaryotic host cell. Effectors interfere with certain host signal transduction pathways to allow the internalization of pathogens and their survival and proliferation inside vacuoles. SteA is one of the few Salmonella effectors that are substrates of both T3SSs. Nothing is known about the function of this protein inside the host cells.

We scheduled a study to evaluate SteA gene expression in patients with Salmonella-induced gastroenteritis and natural specimens, and obtained effective endocrine disrupting chemicals (EDCs). Then, the protein-protein interaction network was constructed using the STRING database and analyzed using Cytoscape using differentially expressed genes (DEGs) with adjusted p value of less than 0.05. Subsequently, the network produced a PPI module. The genes in the module were then analyzed for GO and pathway enrichment. For module genes, EDC gene interactions were collected and reconstructed as a single EDC gene network. In Salmonellosis, 324 putative EDCs were discovered to influence gene regulation. The 3 genes TGFB1, CCND1 and LUM were genes that were affected by EDCs. However, these results need to be experimentally confirmed to suggest improved prevention.

Key Words: *Salmonella enterica; Type III secretion system; SteA; MicroarrayHuman epithelial cell; Salmonella pathogenicity island*

Investigation of the gene expression profile of podocytopathy in glomerulonephritis

Mohammadreza Esmaeili[1], Marjan Nejati[1], Reza Moghadasali[1]*, Sara Taleahmad[1]*

[1] *Department of Stem Cells and Developmental Biology, Cell Sciences Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran*

E-mail address (corresponding author): sara_taleahmad@yahoo.com

Abstract

Chronic kidney disease affects more than 10% of the world population. 90% of all Cases are attributed to glomerular disease. The most sever forms of kidney diseases are often associated with irreversible damage to glomerular podocytes -highly specialized epithelial cells that encase glomerular capillaries and regulate removing toxin and wasted components from the blood. As a consequence, finding signaling pathways related to the podocyte damage by bioinformatics approaches is essential. In this study we focused on gene expression profiling in damaged podocyte cells and investigated signaling pathways associated with this matter. At first, we selected suitable studies from GEO database (GSE151690). This dataset consists of three groups of glomerulonephritis with three difference times (W0, W1, W5) and analyzed with GEO2R tool. Then uploaded the up and down regulated genes to VENNY version 2.1.0 tool and selected processed in common genes between up/down regulated groups to examine the signaling pathways in the DAVID database and the KEGG library then loaded the genes involved in biological process, cell component and molecular function. Finally, we used STRING database to select protein interaction.
Results showed that 167 genes up regulated and 190 genes were down regulated. Among those up regulated genes most of them were expressed in focal adhesion, PI3K-Akt signaling pathway, cell adhesion molecules (CAMs), cytokine-cytokine receptor interaction, in other hand metabolic and TGF-β signaling pathway were observed in down regulated genes. JacK3, Col1a1, Fn1, Ptprc, Pdgfb, Pdgfbr genes were up regulated and the results also obtained that Acadsb, Acmsd, Acsm3, Bmp4, Dmgdh were down regulated. By this analysis and results we observed the genes that identified in podocyte cells damages that highly related to cytoskeleton structure denature, metabolic pathways.

Key Words: *Kidney, chronic kidney disease, Glomerulonephritis, podocytopathy, Microarray analysis, gene expression profile*

Bioinformatics study of miR-301a as a blood-borne biomarker for detection of breast cancer

Reza Ghanbari[1], Ghazal Ziarizi[1], Mohsen Hosseini[1], Mohammad Niazi[1], Mansoureh Azadeh[1]*

[1] *Zist-fanavari Novin Biotechnology Department, State Technical and Vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): bionovin@gmail.com

Abstract

Introduction: MicroRNAs are small non-coding molecules, which play a regulatory role in post-transcriptional gene expression. In this study, using bioinformatics databases, expression levels, target genes, cellular location and genetic pathways of miR-301a were investigated as a blood-borne biomarker for detection of breast cancer.
Methods: Focusing on bioinformatics aspects, miRbase and GEO databases were used to determine chromosomal profile, target genes and validation of miR-301a as a blood-borne biomarker. The cellular location of miR-301a and its expression levels in different organs, especially the breast in a healthy and tumorous state were evaluated in the GeneCards and GEPIA2 database. Finally, the DAVID database was used for genetic pathways analysis.
Results and Discussion: MiR-301a gene is located on chromosome 17q22. The miR-301a is overexpressed in breast cancer and plays a role in the post-transcriptional pathway of approximately 2760 genes which could lead to a variety of diseases by mutations in these genes. Among these genes, mutations in 151 of them can cause breast cancer. MiR-301a can be widely found in whole blood because it was identified to be transported out of the cells after transcription.
Conclusion: Current study reveals that miR-301a plays a role in the post-transcriptional pathway of a large number of genes. It is also highly expressed in breast cancer as a blood-borne biomarker. As a result, miR-301a has a great potential for clinical diagnosis of breast cancer.

Key Words: *Blood-borne biomarker, Breast cancer, MiR-301a*

Bioinformatic tools for microRNA dissection and identify thier target genes in sugarcane

Rahim Ghezzi[1], Leila Nejadsadeghi[2]*, Khosro Mehdi Khanloo[2], Daryoosh Nabati Ahmadi[3]

[1] *Faculty of Agriculture, Shahid Chamran University of Ahvaz*
[2] *Department of Plant Production Engineering and Genetics, Faculty of Agriculture, Shahid Chamran University of Ahvaz*
[3] *Department of Plant Production Engineering and Genetics, Faculty of Agriculture, Shahid Chamran University of Ahvaz*

E-mail address (corresponding author): l.nejadsadeghi@scu.ac.ir

Abstract

Recently, microRNAs (miRNAs) have emerged as important elements of gene regulatory networks. MiRNAs are endogenous single-stranded non-coding RNAs (~22-nt long) that regulate gene expression at the post-transcriptional level. Through pairing with mRNA, miRNAs can down-regulate gene expression by inhibiting translation or stimulating mRNA degradation. In some cases they can also up-regulate the expression of a target gene. Sugarcane is an important industrial crop accounting for nearly 85% of sugar produced worldwide and it is fast becoming an energy crop for the production of bio-fuel ethanol. Waterlogging is one of the serious environmental constraints for optimum growth and yield of sugarcane. Current research has shown that microRNAs (miRNAs) play vital roles in plant response to stresses caused abiotic stress such as waterlogging.

The expression of six candidate miRNAs and their targets were validated using quantitative real-time PCR (qRT-PCR) technology. To identify the mechanisms involved in miRNA-mediated tolerance, it was necessary to first identify the miRNA-modulated genes by bioinformatics mthodes. The mature miRNA sequences were obtained from the miRbase database, and miRNA-modulated genes were predicted using psRNATarget software. Target annotation (biological process, molecular function, and subcellular localization) was performed using Gene Ontology and Uniprot online tools. Co-expression networks were constructed using the GeneMANIA prediction server software, with default analysis parameters. Arabidopsis thaliana and sorghum bicolor were used as the reference genome for all analyses.

Our results showed that miRNAs and their respective target genes were differentially expressed in sugarcane seedling leaves exposed to waterlogging stress. Relative expression data revealed significant differences between miR160, miR166, miR171, miR393, miR396 and miR398 expression levels in sugarcane leaves under waterlogging stress. The predicted target genes regulated by the evaluated miRNAs also show significant difference in relative expression.

Key Words: *Bioinformatic, Gene ontology, miRNA, Target genes*

Genome analysis of Cardamine hirsuta using whole genome sequensing approach

Mahmood Ghorbani Marghashi[1], Hedayat Bagheri[2]

[1] *Department of Horticulture, Faculty of Agriculture, Malayer University, Malayer, Iran*
[2] *Department of Biotechnology, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran*

E-mail address (corresponding author): mgh724@gmail.com

Abstract

Recently, Cardamine hirsuta has emerged as an exclusively powerful genetic system for comparative studies of development. C. hirsuta is a small plant in the cruciferous family, which is a very close relative of Arabidopsis thaliana and has many desirable characteristics as a model species. The C. hirsuta genome is estimated to be 1.5 times that of A. thaliana, with eight chromosomes. Also the constructed high-quality reference genome of the C. hirsuta strain 'Oxford' provided a powerful platform for molecular studies. Genomic DNA of Mashhad ecotype was extracted using CTAB method. 30 microliters of high quality DNA at a concentration of 20 ng /µL were sent to Novogene for genomic sequencing. Sequencing was performed by HiSeq X Ten using the Pair end method. The obtained sequence was sorted using IGV software and reference genome and then analyzed for sequencing quality using Fastqc software. Using Smudgeplot and Genome Scope software, the polyploidy level and homozygosity of the sequenced genome of Mashhad were determined. Genomescope uses k-mer frequencies generated from raw read data to estimate the genome size, the abundance of repetitive elements and rate of heterozygosity, and Smudgeplot to visualize and estimate the ploidy and genome structure of a genome by analyzing heterozygous k-mer pairs.

Using the Genome Scope program and kmer16 size, the size of the Mashhad ecotype genome was estimated between 190 to 230 Mb, which was similar to the size of the Oxford genome as the reference genome. Also, the rate of homozygosity was 99.95% and the rate of heterozygosity was 0.05%. Smudgeplot showed that cardamine hirsuta is a diploid plant. Visualization and estimating the ploidy and genome structure of a genome by analyzing heterozygous k-mer pairs is an important assessment prior to genome assembly.

Key Words: *Cardamine hirsuta, GenomeScope, Smudgeplot*

Progressively Multiple Protein Sequences Alignment Using Intuitionistic Fuzzy Approach

Behzad Hajieghrari[1]*, Naser Farrokhi[2], Mojahed Kamalizadeh[1]

[1] *Department of Agricultural Biotechnology, College of Agriculture, Jahrom University, Jahrom, Iran*
[2] *Department of Cell & Molecular Biology, Faculty of Life Sciences & Biotechnology, Shahid Beheshti University G.C., Evin, Tehran, Iran*

E-mail address (corresponding author): bheghrari@jahromu.ac.ir

Abstract

The progressive alignment approach constitutes one of the most convenient and effective ways to align multiple sequences. Atanassov modifies the fuzzy set by proposing an intuitionistic concept. In this concept, a related degree and a non-relationship degree is detected. However, the sum of the relationship and non-relationship degrees is bare than or equal to one. As a result, the hesitancy degree equals one minus from the whole of the related degree and the non-relationship degree. Traditional hierarchical clustering algorithms are employed broadly to cluster numerical information. Some modifications need the formal hierarchical clustering algorithms to deal with the data expressed in an intuitionistic fuzzy set. In this study, we proposed a measure of the distance between pairs of protein sequences by intuitionistic fuzzy approach and construction merge the tree by a hierarchical grouping to improve the sensitivity of progressive multiple sequence alignments. Both unweighted paired groups with arithmetic mean (UPGMA)- and neighbor-joining (NJ)-based hierarchical clustering were employed to evaluate the algorithm performance. The merging continues until one group remains. Ultimately, the sequences have progressively aligned according to the branching order in the merge tree. Reference sequences from BALiBASE 4.0 (hand-aligned), PREFAB 4.0 (structurally supervised), and OXBench were employed to evaluate the method performance. We computed the quality of the alignments using the Friedman ranks test in the $P<0.05$ statistical significance level, not only in terms of SP-and C- but also TC-score. The UPGMA and NJ-groping of the proposed method perform well in improving the alignment sensitivity and accuracy. Comparatively, where the sequences are not close to each other, the NJ clustering model has more reliable performance. However, UPGMA clustering was the top performer in aligning all the BALiBASE reference sequence sets. The drawback of this approach is its higher time complexity with similar memory usage to the ClustalW.

Key Words: *Intuitionistic fuzzy approach; Multiple sequence alignment; Progressive alignment.*

Identification of effective miRNAs in breast cancer by bioinformatics analysis

Dorsa Hayati[1]*, Hassan Rajabi-Maham[1], Leila Kianmehr[1]

[1] *Animal Sciences and Marine Biology Department, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, G.C, Tehran, Iran*

E-mail address (corresponding author): dorsa.hayati@gmail.com

Abstract

Breast cancer is the greatest reason for deaths associated with cancer among women. In the latest decades, miRNAs (microRNAs) have been reported to partake in the regulation of tumorigenesis-associated various genes, tumor progress, and breast cancer metastasis. The Gene Expression Omnibus was used to download the main data set of GSE4566 and GSE26659. The package of LIMMA of R software was applied to identify the expressed microRNAs differentially among 178 patients with breast tumor and 32 controls. Cytoscape imagined the interaction networks of gene-miRNA and then analyzed using MCODE (Molecular Complex Detection). The candida miRNAs were chosen and investigated by the dataset of TCGA (The Cancer Genome Atlas) and the plotter of KM (Kaplan–Meier) was used to analyze the prognostic values and expression levels of the candida miRNAs. Moreover, KEGG analysis was applied to detect the common pathways that the candida miRNAs were contained. Generally, 35 down-regulated and 18 up-regulated DE-microRNAs were recognized. Additionally, eight candida miRNAs with four down-regulated miRNAs (hsa-miR-143, hsa-miR-125b, hsa-miR-497, hsa-miR-26a) and four up-regulated miRNAs (hsa-miR-155, hsa-miR-375, hsa-miR-203a, hsa-miR-21) were recognized in the network of microRNA target gene and KM found their OS (overall survival) in breast cancer significantly. Lastly, three recognized miRNAs possess the most ordinary in the majority of five pathways that detected using KEGG for instance p53 signaling pathway and cell cycle. This investigation demonstrated that these five pathways might be significant in the progress of breast cancer. The three miRNAs of hsa-miR-375, hsa-miR-203a, and has-miR-21 could be effective microRNAs in breast cancer and might be applied in breast cancer as biomarkers.

Key Words: *Breast Cancer, signaling pathway, networks of gene-miRNA.*

The importance of protein sequence control through Bioinformatics tools to achieve a large amount of desirable proteins in protein purification projects

Zahra Hemati[1], Maziyar veisi[2]

[1] *Department of Pathobiology, School of Veterinary Medicine, Shahrekord University, Shahrekord, Iran*
[2] *School of Veterinary Medicine, Shahrekord University, Shahrekord, Iran*

E-mail address (corresponding author): z.hemati@sku.ac.ir

Abstract

Despite the fact that many protein expression systems have been developed today, but making a pure protein remains a problematic duo to the presence of intensive cloning, expression and purification steps. Recent bioinformatic tools with In-silico analysis have allowed scientists to successfully identify immunogenic proteins.
In our study, the In-silico analysis of the MAP2191 gene sequence was carried out using Bioinformatics tools, along with in vitro experiments. The antigenicity and hydrophobicity analysis of the Mce protein encoded by this gene was investigated using CLC Genomics Workbench 7.5.1. In-silico prediction of the primary structure of Mce protein showed the presence of a highly hydrophobic loop in the N-terminal amino acid sequence of residues 10 to 32. Then a new Mce protein encoded with the MAP2191 gene was amplified and subcloned into E. coli. We have attempted to express Mce protein in a different state with positive expression control.
Western blot analysis of the Mce protein and the control protein only showed the presence of a control protein. Comparing the results of the prediction of the primary structure of the Mce protein with the experimental results confirmed that the expression of this protein was influenced by its hydrophobic nature. Our data support the hypothesis that the presence of a hydrophobic structure, in particular during the first initialization of amino acids in the protein chain, may influence the protein expression system. Our information provide valuable guidance for future researchers to scan the protein sequence for the hydrophobic region in the starting stages of protein purification to have a sense of the expected final protein yield and make better informed choices.

Key Words: *Bioinformatic, In-silico, Hydrophobicity, Protein purification*

Hsa-miR-24 is a novel and critical regulatory biomarker of LBL by regulation of cell growth signaling pathway: integrated microarray and bioinformatics analyses

Saba Hojjati[1], Shadi Omid Ghaemi[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Lymphoblastic leukemias lymphomas (B-All / LBL) account for 2% of lymphoid neoplasms of precursor T-Cells or lymphoblasts. The incidence appears to be rising in both children and adults. B-Cells acute lymphoproliferative disease manifest as pure leukemia (B-All) in 80% of cases, isolated extramedullary presentations frequently have marrow involvement at diagnosis that may be morphologically biomarker may require detection by high-resolution flow cytometry. In recent decades, miRNAs and lncRNAs have been studied and are considered impactful biomarkers in cancer. Therefore, in this bioinformatic approach, the goal was to spot and determine a biological network of genes, miRNAs, and lncRNAs, which have a notable influence on progression LBL. Microarray data analysis of the LBL gene expression profile was performed by GEO2R online software. GSE29986 was analyzed in this study. Using miRWalk and miRdSNP, microRNA – mRNA interaction analysis was performed. The common microRNAs were selected by venny 2.1 online software. Pathway enrichment and mRNA interaction analyses were performed by DAVID. Single nucleotide polymorphism analysis was performed by miRNASNP v3. Based on microarray analysis by GEO2R, PDGFRB has a significant dysregulation in the LBL samples compared to control (adj. p. value < 0.0001). DAVID database revealed that PDGFRB and the related mRNAs, EBF1, LPAR1, WNT6, and FN1, are the essential genes in the cell energy and growth signaling pathway. Also, hsa-miR-24 is the commonly interacted microRNAs with all of mentioned five genes, and its binding affinity is correlated to the rs246387. In conclusion, hsa-miR-24 regulates the cell growth pathway and LBL development by suppressing the expression of mentioned genes, especially in the rs246378 region of the PDGFRB gene.

Key Words: *Keywords: (B-ALL), LBL, PDGFRB, ra246387, has-miR-24*

Phylogenetic analysis of the human gut microbiota and pathogenic bacteria using 16S rRNA

Amir Jalali[1]*

[1] *Department of Biology, Faculty of Sciences, Arak University, Arak, Iran*

E-mail address (corresponding author): a-jalali@araku.ac.ir

Abstract

The human gut microbiota is a large and diverse microbial community that inhabits the gastrointestinal tract and provide multiple benefits for us including nutrient absorption, maturation of the immune system, and also plays a central role in protection of the host from pathogenic bacterial infection. On the other hand, many enteric pathogens have developed strategies in order to be able to outcompete the gut community, leading to infection and/or chronic diseases. Phylogenetic analysis can be effective in choosing the appropriate antibiotics that have the greatest effect on pathogenic bacteria and least impact on the normal human gut flora. The 16S rRNA gene sequences for 50 species of human gut microbiota and pathogenic bacteria were downloaded from GenBank and aligned using Clustal Omega. The tree was constructed using the Maximum-likelihood method in MEGA-X. Also, bacterial diversity in environmental samples was analyzed using RDP Classifier v.2.12. The Species of Collinsella aerofaciens, the most abundant actinobacterium in the gastrointestinal tract of healthy humans, are closely related to other normal gut bacterium Bifidobacterium animalis. Clostridium, includes several significant human pathogens, are closest strains to genus of Clostridioides which includes Clostridioides difficile, a human pathogen causing an infectious diarrhea. Bacillus cereus cause infections of the gastrointestinal tract and surprisingly, the closest strains to this pathogen are Streptococcus salivarius and Streptococcus viridans which are normal human oral bacteria. Vibrio cholerae share close relationship with the other causal agents of the infections of the gastrointestinal tract, Yersinia, Escherichia coli, Shigella and Salmonella. Campylobacter jejuni closely related to Helicobacter pylori. Based on the phylogenetic analysis there are close relationship between human gut microbiota but only some of them are related to gut pathogenic bacteria. In general, this phylogenetic relationship suggested that human gut microbiota and pathogenic bacteria were originated from the different ancestor.

Key Words: *16S rRNA gene; Normal human gut flora; Pathogenic bacteria; Phylogenetic analysis*

Molecular detection of Rickettsia spp. on ovine ticks collected northwest Iran: First report of Rickettsia sibirica

Vahab jamai[1], Jalal Shayagh[1]*, Mohammad Sadaghian[2]

[1] Department of Veterinary Medicine, Shabestar branch, Islamic Azad University, Shabestar, Iran
[2] Department of Pathobiology, School of Veterinary Medicine, Islamic Azad University of Shabestar Branch, Shabestar, Iran

E-mail address (corresponding author): jalalshayegh@gmail.com

Abstract

Background: This study aimed to investigate the species of Rickettsia spp. that were detected at ticks isolated from sheep in four provinces of northwestern Iran, in the Middle East.
Methods: Ticks collected in this study were isolated from four provinces in northwestern Iran. Various counties from East-Azerbaijan, West-Azerbaijan, Ardabil, and Kurdistan provinces were studied. A total of 614 ticks were collected from the studied provinces during this study (40 polled samples). A polymerase chain reaction (PCR) test was used to molecularly evaluate the presence of Rickettsiae in the samples. The gene used in this experiment was the gene encoding citrate synthase protein (gltA). Two of the confirmed samples of Rickettsia were sent to the laboratory for gene sequencing. The sequences were trimmed and edited using Sequencher© V5.4.6 software, and the dendrogram was aligned and drawn with MEGA5© software.
Results: Ticks that found during this study are included Hyalomma marginatum, Hyalomma anatolicum, Hyalomma asiaticum, Rhipicephalus turanicus, Rhipicephalus bursa, Rhipicephalus sanguineus, and Dermacentor marginatus. The results of sequencing of these two genes showed that they belong to two species of Spotted Fever Group Rickettsiae (SFG).
Conclusions: Co-morbidity of Rickettsiae in humans and animals, especially domestic livestock, has made this pathogen one of the most critical tick-borne pathogens worldwide. A total of 614 ticks were collected from the studied provinces during this study (40 polled samples). Ticks found during this study in West-Azerbaijan province included Hyalomma marginatum, Hyalomma anatolicum, Hyalomma asiaticum, Rhipicephalus turanicus, Rhipicephalus bursa, Rhipicephalus sanguineus, and Dermacentor marginatus. Our research and previous articles show that this is the first time that signs of Rickettsia sibirica have been reported in Iran.

Key Words: *Rickettsia, Iran, Vector Borne Diseases, Sequence Analysis.*

Evaluation of ZmNAC10 gene sequence and protein structure of this gene to evaluate salinity stress resistance in plants

Iman Yousefi Javan[1]*, Masoud Alipanah[1]

[1] *Department of Plant Production, Faculty of Agriculture, University of Torbat Heydarieh, Torbat Heydarieh, Iran*

E-mail address (corresponding author): iman.javan@yahoo.com

Abstract

Corn is one of the most important cereals of American origin after wheat and has the largest agricultural lands in the world.This plant contains nutrients and can be a good source of fiber. Due to the value of this plant, recognizing the genes involved in the path of resistance to stress is one of the most important issues of this plant. Irrigation method, environmental factors and medical stresses on it. To better understand and understand the molecular mechanism for resistance to stress, the gene and expression of ZmNAC10 were specifically studied. NACs are one of the largest family of plant-specific transcription factors (TFs) and play a key role in growth and stress responses. The sequence of this gene is highly similar to the sequence of receptor kinases (high-affinity cell surface receptors). In this study, ZmNAC10 gene was studied using bioinformatics sources and primers were designed to amplify this gene. Then, its cDNA was made by reverse transcriptase enzyme and used as a template for ZmNAC10 gene amplification in PCR reaction. ZmNAC10 gene expression was lowest in stems and roots and highest in leaves. During the PCR reaction, a 561 bp fragment of cDNA and a 1440 bp fragment of genomic DNA were amplified. This gene encodes a protein of 187 amino acids that has a region with a kinase sequence. By analyzing the molecular structure of the transcribed protein, the isoelectric point and the molecular weight were obtained, and also based on the homologous modeling, its spatial shape showed that this protein has a regular structure.

Key Words: *Keywords: Molecular mechanism, Resistance to stresses, Kinases Receptor, Homologous modeling*

Phylogenetic analysis of some isolates of Fusarium spp.

Behnaz Karimi[1], Amirreza Amirmijani[1], Nasrin Seyedi[1], Mousa Najafinia[2], Adel Pordel[3]

[1] *Department of Plant Protection, Faculty of Agriculture, University of Jiroft, Jiroft, Iran*
[2] *Iranian Research Institute of Plant Protection, Tehran, Iran*
[3] *Plant Protection Research Department, Baluchestan Agriculture and Natural Resources Research and Education Centre, AREEO, Iranshahr, Iran*

E-mail address (corresponding author): ar.amirmijani@ujiroft.ac.ir

Abstract

The genus Fusarium is highly diverse in morphology and have a worldwide distribution. There are several species complexes in Fusarium that are morphologically indistinguishable but phylogenetically clustered in different phylogenetic lineages. During recent years, molecular studies have led to introduction of new species and combinations in the genus Fusarium. This study was carried out for preliminary phylogenetic analysis of some isolates of this genus obtained from greenhouse crops in south of Kerman province. For this regards, 87 sequences of EF1-α region (including 67 sequences from different species of Fusarium and 20 sequences obtained from GeneBank) were used in phylogenetic analysis. The data analyzed using Minimum-Evolution (ME) and Maximum-parsimony (MP) methods in MEGA 7.0 software. According to the results, this genomic region showed suitable resolution for all of the species. However, there are still several isolates of the species complexes such as F. equiseti and F. incarnatum that are phylogenetically clustered in several subclades.

Key Words: *Greenhouse crops, maximum parsimony, sequencing, species complex*

Analysis of targeted base ITS sequences in the genus Astragalus using Hill Climbing Algorithm

Negar Karimi[1]*, Hojjatollah Saeidi[1]

[1] *Department of Animal and Plant Biology. Faculty of Biological Sciences and Technology, University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): negar.karimi.gh@gmail.com

Abstract

The genus Astragalus is an important taxon with recognized center/centers of diversification and origination in Iran. different species of the genus relating to old and new world occurs in Iran. So, this region has been of great attentions considering phylogenetic studies of the genus. In the present study 30 accessions of different species of the genus were sequenced targeting ITS1 region and 32 seqneced datasets were added from NCBI data base, making a complete representative of 10 phylogenetically close sections of the genus in Iran. Forward and reverse contigs were alighned and used for further analysis. The Hill Climbing algorithm was chosen as the best model representing phylogenetic relationships, testing by GTR+I+G method and, selected through AKAIKE INFORMATION CRITERTION test, by the generation of 2,000,000 generations. The validity of the number of generations was checked with Tracer software. The resulting standard deviation was estimated to be 0.022. The outgrouping taxa were defined and the resulting tree was plotted using FigTree software. The Consistency Index (CI) was calculated for the tree as; CI= 0.391. The results indicated that different species were grouped based on their similarity and dissimilarity and grouped mainly based on their evolutionary priorities defining two specific clades relating to old and new world respectively. Among them 5 species A. aegobromus, A. coluteocarpus, A. siliquosus, A. zerdanus and A. hamosus, all belonging to the old world, were grouped apart. The rest of the species belonging to 2 sections of Hymenostegis and Anthylloidei were grouped in a specific clade, being a representative of the new world taxa occurring in Iran. The results indicated that although the genus Astragalus has been experiencing different geographical and ecological conditions in Iran, the species occurring in Iran follow thier ancestral relationships and possess a potential germplasm to be used with conservation porposus.

Key Words: *Fabaceae, Phylogeny, Bayasian method, ITS, Iran*

Studying relationships of subfamily Panicoideae (Poaceae) in Iran with Bayesian and Maximum Parsmony approaches using ITS and MATK sequences

Negar Karimi[1]*, Hojjatollah Saeidi[1]

[1] *Department of Animal and Plant Biology. Faculty of Biological Sciences and Technology, University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): negar.karimi.gh@gmail.com

Abstract

Subfamily Panicoideae include agronomically important crops like rice and sorghum. Phylogenetic relationships of different members of subfamily Panicoideae occurring in Iran was analyzed using ITS and MATK genes. 62 sequences, representative of 26 species, were provided from NCBI data base. Sequences Alignment for each species was performed using Clustal W. To evaluate the compatibility of the sequences, the ILD test was performed, using Paup. Based on the heuristic search method, the P-value was calculated as 0.001 and defined as standard; P-values <0.05. Different phylogenetic trees were constructed using maximum parsimony and bayasian methods. Among which the tree based on combined data using maximum parsimony showed a more significant values as; CI = 0.3 and RI = 0.4 and was selected for further interpretations. According to the results the two species of Oryza sativa and Poa trivialis belonging to the tribes Oryzoideae and Pooideae, respectively, were grouped apart from other members of the subfamily. Aristida and Stipagrostis were sister genera belonging to the tribe Aristidoideae, which together with the members of the subfamily Panicoideae make the evolutionary clade of PACMAD and show closer kinship with members of this subfamily. Three tribes of Paspaleae, Andropogonaea and Paniceae each created specific clades, creating monophyletic groups. The species belonging to the phylogenetically known Centothecoid clade were also grouped apart from the rest taxa. However; the relationships among members showed polyphyletic situations which could be due to the low number of accessions used in the study. Finally, it is concluded that the ITS and MATK are good markers recognizing relationships in high taxonomic levels and the maximum parsimpony and bayasian methods are a good source of information for such studies.

Key Words: *Poaceae, Phylogeny, MATK, ITS, Iran*

In silico analysis of the deduced protein sequence of eugenol o-methyltransferase of Ocimum basilicum L.

Mohammad Rafiei[1], Fatemeh Khakdan[2]*

[1] *Department of Plant Biotechnology, Isfahan University of Technology, Isfahan, Iran*
[2]*Biology Department, Farzanegan Campus, Semnan University, Semnan, Iran*

E-mail address (corresponding author): f.khakdan@semnan.ac.ir

Abstract

Eugenol are the principal bioactive components of the defensive arsenal of the Ocimum species, which act as signal molecules among plants, humans, and microbes. The final stage in the biosynthesis of phenylpropene, methyl eugenol is catalysis by eugenol O-methyltransferase (EOMT) enzyme. The phylogenetic results indicated different EOMT form evolving from a single ancestral gene and resulting tree was classified into four clusters. ObEOMT was found associated with O. tenuiflorum. The ObEOMT protein molecular mass was predicted to be about 40236.52 and the isoelectric point was predicted to be at 5.59 calculated by the Online Computer pI/MW Tool (http://cn.expasy.org/tools). In addition, based on InterProScan tool and pfam database, the putative amino acids sequence revealed more homology with conserved active site consensus sequence at the N terminus of a variety of plant O-methyltransferases (MSLKCAIQLGIPDILHKHGRPMTLSQLLQSIPINKEKTQCFQRLMRALV) to mediate dimerisation and methyltransferase of these proteins. As PROSITE motif search show, ObEOMT have four major motifs, ASN_GLYCOSYLATION (consisting of 116 amino acids; N-{P}-[ST]-{P}), PKC_PHOSPHO_SITE (Protein kinase C phosphorylation site; [ST]-x-[RK]), CK2_PHOSPHO_SITE (Casein kinase II phosphorylation site; [ST]-x(2)-[DE]), and MYRISTYL (N-myristoylation site; Ala, Ser, Thr, Cys, Asn and Gly). According to conserved domain database, the ProDom and TrEMBL results showed ObEOMT protein consist of the essential domain, EQLLQAQVHVWNHMYAFANSMSLKCAIQLGIPDILHKHGRPMTLS has been specified in SAM-dependent O-methyltransferase class II-type profile. PSORT, signal P and Target P analyses showed the presence of a GxG motif in the N-terminal region of ObEOMT amino acid sequence, which was not found in the protein localization signal peptide at the C-terminal region. Results revealed that predicted ObEOMT protein was a predominantly α-helical protein, which mainly consisted of alpha helices (49.3%) and random coils (27.22%), extended strands (12.9%), and 37.8% beta turns.

Key Words: *eugenol O-methyltransferase (EOMT); phylogenetic analysis; conserved domains; catalytic domains; motif prediction*

Bioinformatics analysis of the key enzyme of important volatile phenylpropanoid pathway in Ocimum basilicum

Zeynab Zare[1], Fatemeh Khakdan[2]*

[1] *Department of clinical biochemistry, Faculty of Veterinary, Urmia University, Urmia, Iran*
[2]*Biology Department, Farzanegan Campus, Semnan University, Semnan, Iran*

E-mail address (corresponding author): f.khakdan@semnan.ac.ir

Abstract

Essential oils of O. basilicum are rich in phenylpropanes like chavicol, methylchavicol, eugenol, methyleugenol, and some terpenoid compounds, which are contributed to the particular properties of many spices and herbs. The final step in the biosynthesis of the methylchavicol, the conversion of chavicol to methylchavicol, is catalyzed by the enzyme chavicol O-methyltransferase (CVOMT). Here, we study the structural and functional important regions (motif prediction), conserved domains, protein domain families, catalytic domains were detected in the protein sequence via PROSITE, InterProScan, ProDom, TrEMBL, respectively. PSORT Prediction, predict Nuclear Localization Signals (NLS) tool and Subcellular location prediction (PLS) web tools. The ObCVOMT protein molecular mass was predicted to be about 39.95 kDa and the isoelectric point was predicted to be at 5.58 calculated. In addition, the putative amino acids sequence revealed more homology with conserved active site consensus sequence of all the PAL protein conserved motifs (MSLKCAIQLGIPDILHKHDHPMTLSQLLQAIPINKEKSQSFQRLMRALV), and belongs to the class I-like SAM-binding methyltransferase superfamily and Cation-independent O-methyltransferase family. MotifScan results revealed that the protein sequence of ObCVOMT had 11 catalytic domains consisting of eight different catalytic domains viz. ASN_GLYCOSYLATION (N-glycosylation site), CK2_PHOSPHO_SITE (Casein kinase II phosphorylation site), MYRISTYL (N-myristoylation site). Results from protein structure modeling program SOPMA revealed that predicted ObCVOMT protein was a predominantly α-helical protein, which mainly consisted of alpha helices (49.16%) and random coils (32.02%), extended strands (13.48%), and 5.34% beta turns.

Key Words: *ObCVOMT, Catalytic domains, Conserved domains, protein structure modeling, motif scan*

Identification of Potential key Genes and Pathways in Prostate Cancer Using Bioinformatic Analysis

Yasaman khamineh[1], Mahmood Talkhabi[1]*

[1] *Department of Animal Biology and Biotechnology, Faculty of Life Sciences and Biotechnology, University of Shahid Beheshti, Tehran, Iran*

E-mail address (corresponding author): m_talkhabi@sbu.ac.ir

Abstract

Prostate cancer (PCa) is the world's second most frequent malignancy that threatens men's health. There are still significant challenges in the treatment of prostate cancer. Its mortality accounts for around 10% of all tumor-related deaths, and it is increasing year by year. The precise molecular pathways are still unknown, prompting urgent study and experience. Therefore, we used bioinformatics analysis to identify potential biomarkers and efficient pathways for PCa early detection. The GSE103512 dataset, which has been downloaded from the Gene Expression Omnibus database (GEO), was normalized using the Transcriptome analysis console (TAC). The genes with adjusted p-value (FDR)< 0.05 and -2<|[log FC]|<2 were identified as differentially expressed genes (DEGs) between 7 normal prostate tissue and 50 PCa samples. Protein-protein interaction (PPI) and visualization were constructed using string, Cytoscape, and Gephi, respectively. We examined these through KEGG pathway enrichment analysis to determine which biological processes, molecular activities may be linked to the overlapping DEGs. According to the findings,822 DEGs (585 up-regulated and 237 down-regulated) were discovered. V-myc avian myelocytomatosis viral oncogene homolog (MYC), SRC proto-oncogene, non-receptor tyrosine kinase (SRC), and cadherin 1, type 1 (CDH1) are three overexpressed genes enriched in the KEGG pathway of peroxisome proliferator-activated receptors (PPARs) and bladder cancer, while caveolin 1 (CAV1), jun proto-oncogene (JUN), and elastin (ELN) are three low expressed genes enriched in the KEGG pathway of protein digestion and absorption and focal adhesion pathways. DEGs in prostate cancer were significantly enriched in the following GO terms (most significant) under the biological process's group: ‘extracellular matrix organization’(GO:0030198), ‘aortic valve morphogenesis’(GO:0003180), ‘fatty-acyl-CoA biosynthetic process’(GO:0046949), and ‘polyamine metabolic process’(GO:0006595). Taken together, we discovered new potential key genes in PCa that might function as robust biomarkers and provide important information for future molecularly targeted therapeutics.

Key Words: *Key words: prostate cancer; differentially expressed genes; pathway; bioinformatics; biomarker*

Transcriptomic analysis of walnut/ Ophiognomonia leptostyla interaction using RNA-seq

Fatemeh Khelghatibana[1]*, Mohammad Javan-Nikkhah[2], Naser Safaie[3], Ahmad Sobhani[4], Ehsan Sari[5]

[1] *Plant Pathology department, Iranian Research Institute of Plant Protection, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran*
[2] *Department of Plant Protection, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran* [3] *Department of Plant Pathology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran*
[4]*Agricultural Biotechnology Research Institute of Iran - Isfahan Branch, Agricultural Research, Education and Extension Organization (AREEO), Isfahan, Iran*
[5]*Wheat Genomics and Pathology, Durum Wheat Breeding and Genetic Team, Crop Development Centre, University of Saskatchewan, Saskatoon, Canada*

E-mail address (corresponding author): khelghati50@gmail.com

Abstract

Anthracnose disease caused by Ophiognomonia leptostyla, is one of the most important and widespread fungal disease on walnut species including Persian walnut. Knowledge of molecular defense responses, could help to develop pathogen-tolerant walnut varieties. In this study RNA-seq was used to profile the transcriptome of O. leptostylia -inoculated and mock-inoculated leaf samples of Persian walnut (cv. Chandler) at three time points: 48(T1), 96(T2) and 144(T3) hours post inoculation. Short reads were generated as 100 PE on Illumina Novaseq 6000 platform. HISAT2 and StringTie were used for mapping and transcript assembly respectively [2;3]. Transcripts were annotated by sequence alignment similarity search BLASTX to protein database UniProt, TAIR10 and NCBI nr (e-value<1e-5). Geneontology enrichment was carried out based on annotation results using Wego software. In sum 30961, 41283 and 41283 genes were characterized in T1, T2 and T3 samples respectively. In contrast to T2 and T3 samples, GO terms in T1 were rather of biological process than molecular function or cellular components. That means the plant host responses to infection increased over 96hpi and 144hpi. Differential gene expression identified 892 and 1255 and 3174 DEGs in T1, T2 and T3 samples respectively (FDR < 0.001 LogFC=2) compared with healthy control. Overall in T1 sample the number of DEGs with decrease in expression were more than DEGs with increase, in contrast to T2 and T3 samples. Further analysis revealed a long list of diverse plant defense related genes identified among DEGs including different transcription factors like WRKY, ERFs, NAK, PR proteins, Zinc finger and proteins and NBS-LRR proteins, serine threonine kinases, ABC transporter and a proteins involve in hormone signal transduction. These are involved in hormone signal transduction, MAPKs signaling and phenylpropanoid biosynthesis pathways and constitute a complex defense network in the response of walnut cv. Chandler against to O. leptostyla infection.

Key Words: *Juglans regia; Ophiognomonia leptostyla; plant pathogen interaction, RNA-seq; DEGs; plant defense responses; cv Chandler*

A reference transcriptome for walnut anthracnose pathogen, Ophiognomonia leptostyla

Fatemeh Khelghatibana[1]*, Mohammad Javan-Nikkhah[2], Naser Safaie[3], Ehsan Sari[4] Ahmad Sobhani[5], Somayeh Shams

[1] Plant Pathology department, Iranian Research Institute of Plant Protection, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran
[2] Department of Plant Protection, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran [3] Department of Plant Pathology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran
[4]Wheat Genomics and Pathology, Durum Wheat Breeding and Genetic Team, Crop Development Centre, University of Saskatchewan, Saskatoon, Canada
[5]Agricultural Biotechnology Research Institute of Iran - Isfahan Branch, Agricultural Research, Education and Extension Organization (AREEO), Isfahan, Iran

E-mail address (corresponding author): khelghati50@gmail.com

Abstract

Despite the economic losses due to Ophiognomonia leptostyla on walnut trees, no genome or transcriptome reference for the pathogen has yet been available. In the present study, the transcriptome of O. leptostyla isolate SA-SE was assembled using four assemblers i.e., Trinity, Oases, SOAPdenovo-Trans and Bridger. RNA sequencing of fungal mycelia grown on axenic culture and the leaf samples of Persian walnut (cv. Chandler) inoculated with the fungal conidia and sampled at three time points 48h, 96h and 144h post inoculation. The completeness and contiguity of assemblies were assessed by tools such as Transrate and BUSCO to identify the superior assembly [1,2]. In most of the assessment criteria such as N50, Transrate score, number of ORFs with known description in gene bank, the percentage reads mapped back to the transcript (RMBT), BUSCO features, SwissProt coverage bin and RESM-EVAL score, the Bridger assembly was the superior assembler. Next, the expression of transcript was profiled over sampling time point to understand how O. leptostyla deploys transcriptome for infection and colonization within host plants. K-means clustering of expressed transcripts resulted in transcripts grouped in four different clusters over three sampling time points. Candidate effectors and Cazyme were identified in silico and their co-expression pattern was examined to identify candidate key genes in infection and host colonization.

Key Words: *Keywords: Juglans. Regia (cv Chandler), Ophiognomonia leptostylaRNA-seq, de novo assembly, K-mean analysis, effector*

Tandem repeats ubiquitously flank and contribute to translation initiation sites

Ali M.A. Maddi[1], Kaveh Kavousi[1]*, Masoud Arabfard[2], Hamid Ohadi[3], Mina Ohadi[4]*

[1] *Laboratory of Complex Biological systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[2] *Chemical Injuries Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran*
[3] *School of Physics and Astronomy, University of St. Andrews, St. Andrews KY16 9SS, United Kingdom.*
[4] *Iranian Research Center on Aging, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Background: Recent findings in yeast and human suggest that evolutionary divergence in cis-regulatory sequences can impact translation initiation sites (TISs). Here we employed the TIS homology concept to study a possible link between every category of tandem repeats (TRs) and TIS selection.
Materials and Methods: We selected human as reference, and 83 other species, and extracted the entire protein-coding genes (n=1,611,368) and transcripts (n=2,730,515) annotated for those species from Ensembl 102. Two different weighing vectors were implemented to designate homologous vs. non-homologous TISs. The threshold for TIS homology was set based on three thousand simulations of random pair-wise comparisons of the initial five amino acids (excluding the initial methionine) of protein-coding genes in human. TISs that were flanked by TRs in human were BLASTed against the initial TISs in the orthologous genes across the 83 species, and the number of events in which human-specific and non-specific TRs occurred with homologous (≥ 50% homology) and non-homologous (<50% homology) TISs were subsequently calculated. On average, every transcript was flanked by 1.19 TRs of various categories in their 120 bp upstream RNA sequence.
Results and Conclusion: We detected statistically significant excess of non-homologous TISs co-occurring with human-specific TRs and vice versa (on average, significant p-values << 0.1 near the zero were calculated for all experiments). At the interspecies level, human proteins flanked by human-specific TRs were significantly less homologous to other species than those flanked by non-specific TRs. We conclude that TRs are abundant cis-elements in the upstream sequences of TISs across species. We also conclude a link between all categories of TRs and TIS selection based on the patterns of co-occurrence of TRs with TISs. Asymmetric and stem-loop structures formed as a result of TRs may function as genetic marks for TIS selection.

Key Words: *Translation initiation site; tandem repeat; genome-scale; TIS selection; homology*

Identification of genes with significant differential expression in breast cancer patients using transcriptome data analysis

Pegah Mahbobniya[1], Mostafa Rafiepour[1], Reza Mahdian[2], Farinaz Behfarjam[1]*

[1] *Department of Genetics, Faculty of Science, Danesh Alborz University, Qazvin, Iran*
[2] *Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran*

E-mail address (corresponding author): behfarjam.farinaz@gmail.com

Abstract

Background: Breast cancer is a common cancer that starts in the breast tissue. This disease is the most common cancer in women worldwide with a poor prognosis due to its aggressive biological behaviour and lack of therapeutic targets. Therefore, identifying the key genes involved in potential initiation and progression mechanisms of breast cancer can be useful in the diagnosis, prognosis and treatment of this disease. Accordingly, the aim of the present study was to identify some novel genes with significant differential expression affecting breast cancer patients through bioinformatics analysis.
Materials and Methods: Illumina-sequenced transcript data from three samples with breast cancer and four healthy individuals as control were collected from the NCBI database, and after controlling the quality of the readings with FastQC program, the reads alignment were performed with reference genomes by STAR software. Then, data features were evaluated using featureCounts program. DESeq2 software was used to identify the genes with differential expression between breast cancer samples and healthy individuals as control. Finally, Volcano graph, bar plot and point diagrams were traced to determine genes with significant differential expression.
Results: RNA sequencing data analysis between patients with breast cancer and healthy individuals showed valuable results. The results of these study were identified 11 downregulated differentially expressed genes (DEGs) and 20 upregulated DEGs. Among the key genes of this study can be noticed RNF223, ETV3L, CACNA1S, UPS34, NEURL3, CRYBG3 and FABP4 and other important genes such as CCL20, CXCL5, CEP295 and FLG2 that significantly associated with breast cancer pathogenesis.
Conclusion: In this study, we found some key genes with important roles in the pathways and molecular mechanisms involved in the pathogenesis of breast cancer that can be candidate targets for diagnostic, therapeutic and preventive purposes.

Key Words: *Keywords: Breast cancer; Differential gene expression; Transcriptome; Bioinformatics; NCBI.*

Towards accurate avian tree of life

Sina Majidian[1]

[1] *Department of Computational Biology, University of Lausanne*

E-mail address (corresponding author): sina.majidian@unil.ch

Abstract

Species tree reconstruction was long ago based on morphological characters. In 1990s, a few universal orthologous genes (gene markers) including the small and large subunit ribosomal RNAs were utilised for this purpose. This built the foundation of molecular phylogenetics. However, thanks to the advent of DNA sequencing technologies, sequence data are being widely used for species tree inference. Thus, nowadays hundreds of gene markers could be used in this regard. Nevertheless, inferring these marker loci is computationally demanding and requires complicated pipelines to extract the single copy ortholog groups (OGs). Recently, a new phylogeny for a dataset of 363 birds has been proposed. Here, we discuss the speed and accuracy of their approach compared to the NCBI taxonomy using the Robinson–Foulds metric. Besides, we present our fast method for tree inference for arbitrary species sets. In this method, the OMAmer software was used to place proteins of species of interest onto a database of hierarchical orthologous groups from OMA. Then, each protein is mapped to an OG if possible. Next, the most informative OGs are selected for which multiple sequence alignment (MSA) matrices are computed. Finally, the species tree is inferred using IQ-tree using the super-matrix. The super-matrix of orthologous characters was computed within 50 CPU hours and IQ-Tree needs 11h on 48 CPUs to infer the avian tree.

Key Words: *Comparative genomics, Species tree, NCBI taxonomy, Orthologous groups.*

Bioinformatics analysis of hsa-miR-135a on APC gene in patients with Familial Adenomatous Polyposis

Maryam Modrek[1], Irsa Alborzi[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and Vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Familial Adenomatous Polyposis (FAP) is an autosomal-dominant inherited disease caused by germline mutations in the Adenomatous Polyposis Coli (APC) gene. Locating on choronosome 5, the APC gene encodes a tumor suppressor protein that acts as an antagonist in Wingless-Related Integration Site (WNT) signaling pathway. Defects in this gene cause FAP that usually results to malignancy. The aim of our study was to find bioinformatics information about the role of APC gene in the process of FAP. For this purpose, GEO datasets was studied in order to achieve information about APC gene, miRdSNP was applied to choose microRNA involved in FAP disease. The miRBase database helped to detect data about selected microRNA. The pathways associated with APC were detected through sortment of genes in DAVID database. The information about APC gene in the carcinogene pathway was obtained from KEGG pathways. The results showed that hsa-miR-135a is predicted to target the APC gene. It targets the 3'UTR region of APC, suppresses its expression. Inactivation of the APC gene is a major initiating event in colorectal tumorigenesis. Based on the findings, it is inferred that β-catenin levels may increase due to the cessation of APC gene function. Since the WNT signaling pathway is a conserved pathway, tumorigenesis may begin because of incorrect signaling of the WNT pathway. APC-free β-catenin provokes the WNT signaling pathway as a protein, leading to active transcription of target genes. Also we saw SNP:rs6875894 (single nucleotide change c/t conversion) in nucleotide 141 on miR-135a sequence which had no effect on the emrgene of this disease.It is concluded that as APC is a tumor suppressor gene so it is not the cause of the disease and when its function is stopped by miR-135a , the cell progresses to tumorigenesis.

Key Words: *FAP; miRNA; SNP; KEGG, DAVID*

Investigation of CFTR: c.364T>A (p.Tyr122Asn) variant using in silico predictive tools

Keivan Moradi[1]*, Sahand Khamooshian[2]

[1] *Department of Biochemistry, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran*
[2] *Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran*

E-mail address (corresponding author): keivan.moradi@kums.ac.ir

Abstract

As a disorder with autosomal recessive (AR) inheritance, cystic fibrosis (CF) is caused by pathogenic variants in cystic fibrosis transmembrane conductance regulator (CFTR) gene. Therefore, it is necessary to classify the variants identified in this gene into pathogenic and non-pathogenic types. The aim of this study was to investigate the deleterious effect of CFTR: c.364T>A (p.Tyr122Asn) variant. To this end, ten in silico predictive tools were selected. Our results are shown in the table below; accordingly, CFTR: c.364T>A (p.Tyr122Asn) variant had deleterious effects on all in silico tools used by us. This variant was first reported by Iranome project (http://www.iranome.ir/) in a healthy individual with Lurish (Luri) ethnicity from Iran (in heterozygous form). Our search in the databases of LOVD, HGMD, dbSNP, ClinVar, gnomAD, 1KGP, CFTR1, and CFTR2 as well as in the google scholar did not reveal any other reports for CFTR: c.364T>A (p.Tyr122Asn) variant. In conclusion, with a threshold of deleterious effects in seven or more in silico predictive tools, this variant could be accepted as pathogenic. However, for its final classification, it is necessary to consider the other criteria provided by American College of Medical Genetics and Genomics (ACMG-AMP) guidelines.

PhD- SNPg	PANTHER PSEP	SNPs & GO	FATHMM-XF	I-Mutant	PolyPhen -2	PROVEAN	Mutation Taster	CADD	SIFT	# of tools with pathogenic prediction
Pathogenic	Probably damaging	Disease	Pathogenic	Disease	Probably damaging	Deleterious	Disease causing	25	Damaging	10:10

Key Words: *Cystic fibrosis (CF); CFTR gene; In silico analysis*

In silico assessment of CFTR: c.1118A>G (p.Asp373Gly), a novel variant detected among Iranian population

Keivan Moradi[1]*, Sahand Khamooshian[2]

[1] *Department of Biochemistry, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran*
[2] *Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran*

E-mail address (corresponding author): keivan.moradi@kums.ac.ir

Abstract

Recent advances in genome sequencing have led to the identification of large amounts of variants in human genes. As a national project, the Iranome project (http://www.iranome.ir/) was established recently in Iran and resulted in detecting many genomic variants not reported before. In this study, the effect of a variant detected in cystic fibrosis transmembrane conductance regulator (CFTR) gene, c.1118A>G (p.Asp373Gly), was investigated using ten in silico predictive tools including PhD- SNPg, PANTHER PSEP, SNPs & GO, FATHMM-XF, I-Mutant Disease, PolyPhen-2, PROVEAN, Mutation Taster, CADD, and SIFT. NM_000492.4 was used to determine the variant position and the position of the variant in protein was determined based on UniProtKB/SwissProt P13569-1. According to our results, except for PolyPhen-2 and SNPs & GO, the other eight in silico tools predicted deleterious effects for this variant. CFTR: c.1118A>G (p.Asp373Gly) variant had been detected in two healthy individuals with Persian ethnicity from Iran (both in heterozygous forms). Our efforts to find this variant in the literature were unsuccessful. In addition, it was not reported in LOVD, HGMD, dbSNP, ClinVar, gnomAD, 1KGP, CFTR1, and CFTR2 public databases. In conclusion, with a threshold of deleterious effects in seven or more in silico predictive tools, CFTR: c.1118A>G (p.Asp373Gly) variant could be accepted as a pathogenic variant. However, for its final classification, it is necessary to consider the other criteria provided by American College of Medical Genetics and Genomics (ACMG-AMP) guidelines.

Key Words: *Cystic fibrosis (CF); CFTR gene; In silico analysis*

In silico pathogenicity evaluation of QDPR: c.135T>A (p.Asn45Lys) variant

Keivan Moradi[1]*, Sahand Khamooshian[2]

[1] *Department of Biochemistry, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran*
[2]*Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran*

E-mail address (corresponding author): keivan.moradi@kums.ac.ir

Abstract

Genetic mutations occurred in the genes related to BH4 cofactor, including quinoid dihydropteridine reductase (QDPR), may lead to hyperphenylalaninemia (HPA). DHPR deficiency is caused by defects in the corresponding gene QDPR (1). In this study, we evaluated the pathogenicity of QDPR: c.135T>A (p.Asn45Lys) variant using ten in silico predictive tools including PANTHER PSEP, PhD-SNPg, PROVEAN, Mutation Taster, SNPs & GO, FATHMM-XF, I-Mutant Disease, PolyPhen-2, CADD, and SIFT. NM_000320.3 and P09417-1 were used as reference sequences. Except for PhD- SNPg, FATHMM-XF, and SNPs & GO, the other seven in silico tools predicted deleterious effects for this variant. QDPR: c.135T>A (p.Asn45Lys) variant had been reported recently for the first time in an Iranian patient with DHPR deficiency (in compound heterozygous form) (2). There were no reports of this variant in the literature as well as in the LOVD, HGMD, dbSNP, ClinVar, gnomAD, 1KGP, and BIOPKU public databases. In conclusion, with a threshold of deleterious effects in seven or more in silico predictive tools, QDPR: c.135T>A (p.Asn45Lys) variant could be accepted as a pathogenic variant. However, for its final classification, it is necessary to consider the other criteria provided by American College of Medical Genetics and Genomics (ACMG-AMP) guidelines (3).

Key Words: *DHPR deficiency; QDPR gene; In silico analysis*

Integrated Bioinformatics Analysis of Genes Associated with Lung Cancer Pathway

Zahra mosayebi Dorcheh[1], Pegah Javid.Mansoreh Azadeh[1]*

[1] *Zist fanavari Novin Biotechnology Department state Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Lung Cancer (LC) is a disorder that influences lung tissue cells. LC is generated while bizarre cells are divided in an out of control manner to shape a tumor withinside the lungs. The remedy technique relies upon at the form of most cancers in addition to the patients' trendy health. Mortality from lung most cancers may be drastically decreased with early detection of the disorder. In this study, the bioinformatics analysis was conducted to find more information about the involved markers and pathway through the disease. For this purpose, NCBI, miRNASNP, miRdSNP, miRBase, miRWalk, and DAVID databases were surveyed. In this regard, miRNASNP and miRdSNP databases were used to identify related microRNAs (miRs) and single Nucleotide Polymorphisms (SNPs) in LC. Studies at NCBI indicated that three of the most active genes involved in LC, are TPM3 gene (Located on chromosome 1), ITGB3 gene (Located on chromosome 17) and CYP1B1 gene (Located on chromosome 2). Studies in miRdsNP database showed that in the process of LC disease, the conversion of T to C base through dSNP rs1051370 in nucleotide 2179 could affect the binding of miRNA has-miR-1 to associated 3'UTR of TPM3 gene on chromosome 1 and change its function. In addition, the occurrence of has-miR-1 sequence on 3'UTR of TPM3 gene on chromosome 1 would change its function. Furthermore, has-miR-1, has-let-7a and has-miR-27b could be the most important microRNAs associated with these genes. The dSNP rs2317676 occurs on nucleotide 712 at the junction of has-let-7a to 3'UTR of ITGB3 gene and converts the alleles A to allele G. In the process of LC, the has-miR-27b binds to 3'UTR of CYP1B1 gene on chromosome 2 and the dSNP rs9341266 accrues in nucleotide 694 where C is converted to T. It is concluded that since changes in sequences of genes mediated by dSNP can change their function, it could be a potential marker through disease diagnosis.

Key Words: *miRNA; cancer; dSNP; miRdSNP; fatal diseas*

In silico detection of a new deleterious single nucleotide polymorphism and miRNA related to CHEK2 gene in breast cancer and their interactions

Sajede Naghiyan Fesharaki[1], Sajjad Sisakhtnezhad[1], Mansoureh Azadeh[2]

[1] *Department of Biology, Faculty of Science, Razi University, Kermanshah, Iran*
[2] *Zist Fanavari Novin biotechnology institute, Isfahan, Iran*

E-mail address (corresponding author): minota_2012@yahoo.com

Abstract

Introduction: Checkpoint Kinase 2 (CHEK2) is a serine/threonine protein kinase that acts as a tumor suppressor gene through the induction of cell cycle checkpoint, the inhibition of cellular proliferation, and the activation of DNA repair pathways and apoptosis. Although CHEK2 mutations are rare, it reported that the risk of developing breast cancer is higher in carriers of truncating mutations. We also propose that deleterious single nucleotide polymorphism (SNPs) in 3′-UTR of the CHEK2 gene may affect the interaction of microRNAs (miRNAs) that regulate the expression of this gene in breast cancer. Therefore, this in silico study was conducted to identify the possible deleterious single nucleotide polymorphisms (SNPs) in the 3′-UTR of the CHEK2 gene and also miRNAs that target it. Moreover, we aimed to investigate the influence of the identified SNPs on binding of the miRNAs in the 3′-UTR of CHEK2 gene.
Methods: The CHEK2 gene sequence was obtained from NCBI database. The dbSNP tool of NCBI was also used to identify the possible deleterious SNPs in the gene. Next, integrated bioinformatics data analysis by miRWalk, PhenomiR, miRBase, David, and miRNASNP-V3 were performed to identify miRNAs that target CHEK2 gene mRNA and also to evaluate the effect of SNPs on its binding site.
Results: Our results introduced rs540410451 as a deleterious SNP that result to the substitution of G to A allele in CHEK2 gene of breast cancer patients. It also found that hsa-miR-12136 may regulate CHEK2 gene expression. Moreover, in silico data analysis indicated that rs540410451 may affect the performance and binding of hsa-miR-12136 in the 3'-UTR of the CHEK2 mRNA.
Conclusion: This study for the first time, introduces rs540410451 and hsa-miR-12136 as important biomarkers that may be associated with CHEK2 gene and breast cancer. However, experimental studies are needed to confirm the results of this study

Key Words: *Single nucleotide polymorphism, rs540410451, CHEK2 gene, hsa-miR-12136, Breast cancer.*

New insights into the evolutionary history of Bromus pumilio and Bromus gracillimus based on nuclear and plastome molecular data

Akram Nasiri[1]*

[1] *Department of Plant Biology, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): anasiri89@yahoo.com

Abstract

Bromus L., classified in its own tribe Bromeae, is a large and complicated grass genus containing about 165 annual to perennial species that is widely distributed in temperate regions of the world. To reassess the phylogenetic relationships among B. pumilio and B. gracillimus, species whose respectively classification within Bromus sects. Boissiera and Nevskiella, and other sections of the Bromus genus, the phylogenetic analyses of 46 Bromus accessions and 4 outgroup specimens were conducted based on DNA sequences from four molecular markers (nrDNA ITS and ETS and plastid rpl32-trnLUAG and matK) in this study. Sequences were assembled, trimmed, and visually assessed using Geneious 11.1.5 software and edited sequences for each gene were aligned separately using the MAFFT v7.388 alignment algorithm. Then, sequences were concatenated as two combined nuclear and plastid DNA dataset matrices using SequenceMatrix. Phylogenetic relationships of Bromus sections were eventually reconstructed by using maximum parsimony, Maximum Likelihood, and Bayesian analyses. In the phylogenetic trees derived from both nrDNA and plastid sequences, Bromus s.l. is monophyletic with strong support and B. pumilio and B. gracillimus are maximally supported as sister taxa. The relationship of the B. pumilio–B. gracillimus lineage to other Bromus lineages differs between the plastid and nrDNA trees. In the nrDNA trees, B. densus and B. pumilio–B. gracillimus are sister groups, whereas in the plastid trees B. pumilio–B. gracillimus and B. sect. Ceratochloa form a strongly supported clade, and this clade and B. densus are successive sister groups to the rest of the genus.

Key Words: *Bromus sect. Boissiera; Bromus sect. Nevskiella; grasses; molecular phylogenetics; evolutionary reconstruction*

Investigation of renal tubular cells gene expression during acute kidney injury induced by Cisplatin

Marjan Nejati[1], Mohammadreza Esmaeili[1], Reza Moghadasali[1]*, Sara Taleahmad[1]*

[1] *Department of Stem Cells and Developmental Biology, Cell Sciences Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran*

E-mail address (corresponding author): sara_taleahmad@yahoo.com

Abstract

Acute kidney injury (AKI) is defined by a sudden loss of excretory kidney function. (1) AKI is part of a range of conditions summarized as acute kidney diseases and disorders (AKD), in which slow deterioration of kidney function or persistent kidney dysfunction is associated with an irreversible loss of kidney cells and nephrons, which can lead to chronic kidney disease (CKD). (2, 3, 4) As a consequence, finding signaling pathways related to the proximal tubule injury by bioinformatics approaches is essential. In this study we focused on gene expression profiling in damaged Tubular cells and investigated signaling pathways associated with this matter. At first, we selected suitable studies from GEO database (GSE85957). This dataset consists of three groups of injured PTCs with four difference times (D3, D5, D8 and D26) and analyzed with GEO2R tool. Then uploaded the up and down regulated genes to VENNY version 2.1.0 tool and selected processed in common genes between up/down regulated groups to examine the signaling pathways in the DAVID database and the KEGG library then loaded the genes involved in biological process, cell component and molecular function. Finally, we used STRING database to select protein interaction. Results showed that 127 genes up regulated and 128 genes were down regulated. Among those up regulated genes most of them were expressed in p53 signaling pathway, Ribosome, TNF signaling pathway, HTLV-I infection in other hand metabolic pathways, Citrate cycle (TCA cycle) and Carbon metabolism were observed in down regulated genes. Mdm2 ،Cdkn1a ،Egr1 ،Fos ،Rps7 ،Atf3 genes were up regulated and the results also obtained that Adk, Pdha1، Sdha ،Sdhd ،Uqcrc2 were down regulated. By this analysis and results we observed the genes that identified in tubular cells damages that highly related to ribosome, metabolic pathways.

Key Words: *Acute kidney disease, Proximal tubule cell, Microarray analysis, gene expression profile*

Integrative analysis of protein-coding genes, long non-coding genes, and microRNAs in breast cancer by GDC RNATOOLS and drug-gene interaction analysis

Sajjad Gharaghani[1], Romina Norouzi[2]

[1] Institute of Biochemistry and Biophysics (IBB), Department of Bioinformatics, Laboratory of Bioinformatics and Drug Design (LBD), University of Tehran
[2] University of Tehran, Kish International Campus, Institute of Biochemistry and Biophysics, Department of Bioinformatics, Laboratory of Bioinformatics and Drug Design (LBD)

E-mail address (corresponding author): s.gharaghani@ut.ac.ir

Abstract

Introduction: For this study, all mRNAs (Gene Expression Quantification (and microRNAs (BCGSC miRNA Profiling) count matrix (1204 samples and 60483genes for mRNAs and 1182 samples and 2588genes for microRNAs) data for breast cancer (the primary tumor) get from TCGA by GDC client after that normalized by TMM method and finding out differential express genes by Deseq2 package. Volcano plot and bar plot and Hierarchical clustering are used to visualize DEGS analysis results.
For ceRNAs network analysis of internally incorporated databases of miRNA-mRNA and miRNA-lncRNA interactions are used. lncRNA-miRNA-mRNA interactions reported and visualized in Cytoscape. For Functional enrichment analysis, perform Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Disease Ontology (DO) on the list of DEGs.
To find drug-gene interaction, by using DGIdb, selected (up and downregulated) gene in DEG file with criteria (logFC>2 and adjP.Val >0.005) for up-regulated and (logFC<-2 and adjP.Val >0.005) for the downregulated gene) and the results used as input for the DGIdb database. Then the interactions between drugs and differential express genes and the list of interactions are used as input to generate a network of genes and drugs.
Results: Based on analysis 1763 differential express genes find out in breast cancer which among them 1514 are protein-coding, 191 are lncRNAs and noncoding RNAs, and 47 pseudogenes and 4 TR and 7 TEC gens and 2 IG genes. Based on ceRNAs network analysis finding 404 miRNAs sponges by 404 lncRNAs, and 404 miRNAs directly interact with 404 mRNAs.
Drug -gene interaction prediction by DGIdb: based on the filter that is set on DEGs, find 606 downregulated genes that interact with the same number of drugs such as HRITONAVIREPARIN, CISPLATIN, RITONAVIR, etc. and 73 upregulated genes interact with the same numbers of drugs such as METHOTREXATE, TRASTUZUMAB, and so on.

Key Words: *Keywords: GDCRNA; lncRNAs, circulating RNA, breast cancer, drug-gene interaction, DGIdb*

Bioinformatics Evaluation of the KRT Gene's Effects on Epidermolysis Bullosa Disease

Fatemeh Rostamaei[1], Mitra Omidi[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

The Epidermolysis Bullosa (EB) disease may be caused by mutations in the KRT14 and KRT5 genes, that lead to abnormal polymerization of the middle fibers in the keratinocyte layer and result to weak epidermal [1,2]. The mutation in keratine (KRT) gene mediated by Single Nucleotide Polymorphisms (SNP) and micro-RNA (miRNA) change some alleles that result in EB. The SNPs residing within miRNA at the 3'UTR of genes may, impair miRNA biogenesis and alter target selection have potentially profound [4,5]. In the current study, we aimed to investigate the process of EB disease bioinformatically through SNPs and microRNA affecting the disease. For this purpose, miRNASNP and LncRNASNP2 databases were studied in order to find the effective SNPs in EB and profile changes of genome. The results indicated that the most important genes involve in EB were KRT10 and KRT5. The conversions of alleles A to G, T to C and A to C were observed in SNPs rs780277802, rs774862920 and rs773100474 respectively. The occurrence of these SNPs on miRNA sequence may have potentially profound biological effect and lead to changes in chromosomal structure. Bioinformatics analysis revealed that frequent mutations in chromosomes 17 and 12 lead to mutations in KRT14 and KRT5 genes due to interaction of SNPs and miRNAs. In detail, the occurrence of rs780277802 within hsa-miR-106a-5p, rs774862920 on hsa-miR-20a-5p, and rs773100474 on the sequences of hsa-miR-17-5p on 3'UTR of KRT5 gene may lead the cells to tumorigenesis. In this study and previous studies, it was observed that the main causes of EB is abnormal production of protein and keratin due to interaction of SNPs and mRNA in KRT genes. Overall, it is concluded that the EB phenotype harbor a spectrum of mRNA (hsa-miR-20a-5p) and SNP (rs774862920) mutations in the KRT5 gene.

Key Words: *Skin disease; EB; KRT genes; bioinformatics*

Agonist/ antagonist compounds' mechanism of action on estrogen receptor-positive breast cancer: a system-level investigation assisted by meta-analysis

Zeynab Piryaei[1,2], Zahra Salehi[3], Mohammad Reza Tahsili[4], Esmaeil Ebrahimie[5,6], Mansour Ebrahimi[4], Kaveh Kavousi[2]*

[1] *Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran*
[2] *Laboratory of Complex Biological Systems and Bio-informatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Iran*
[3] *Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran*
[4] *Department of Biology, Faculty of Science, University of Qom, Qom, Iran*
[5] *School of Animal and Veterinary Sciences, The University of Adelaide, South Australia, Adelaide, Australia*
[6] *Genomics Research Platform, School of Life Sciences, La Trobe University, Melbourne, Victoria, Australia.*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

The largest group of breast cancer patients are estrogen receptor-positive (ER+). There is a vast amount of studies focused on breast cancer. That vastness provides the requisites for the integration and meta-analysis of the related studies. Meta-analysis could lead to more accurate results than single investigations. In the present work, a specific layout for meta-analysis of multiple RNA-seq datasets is proposed in order to obtain a final accurate, least error-prone methodology. Meta-analysis was separately performed on two estrogen-treated MCF7 and T47D versus untreated cell lines to obtain meta-differentially expressed genes. Also, only shared significant genes between MCF7 and T47D cell lines were enriched to obtain more stringent results. The ER+ patients respond to both ER agonist (E2) and ER antagonists (Tamoxifen, Fulvestrant, and Brilanestrant). Hence, we compared the meta-analysis results with genes obtained from ER antagonists to understand the function of ER and its affected genes. Genes involved in human mitochondria and several keratin family members' genes were up-regulated in the meta-analysis. Still, they showed no alteration neither in individual datasets treated with E2 and ER antagonists. Our findings indicated that Tamoxifen does not block specific genes directly affected by ER and has no effect on their expression. Moreover, to the best of the authors' knowledge, pathways were identified that were not previously reported in BC. Meta-analysis of RNA-seq data with correct methodology could identify new genes and pathways that are essential in breast cancer. If there are suitable datasets, it is recommended that the methodology be used for other diseases to obtain more accurate results.

Key Words: *Breast cancer; Estrogen receptor positive; ER agonist/antagonist; RNA-sequencing; Meta-analysis; MCF7/T47D cell lines*

The gene expression patterns of stromal cells in prostate tissue during prostate cancer

Maryam Rahimi Tesiye[1], Mona Zaersabet[2], Amir Charkaneh[2], Elnaz Khorasanian[1], Zahra Abrishami Kia[3]

[1] *Faculty of Life Science and Biotechnology, Shahid Beheshti university, Tehran, Iran*
[2] *Department of Biology, Faculty of Sciences, University of Guilan, Rasht, Iran*
[3] *Department of Exercise Physiology, Faculty of Physical Education and Sport Sciences, University of Mazandaran, Babolsar, Iran*

E-mail address (corresponding author): maryamrahimi563@yahoo.com

Abstract

Introduction: The development of stromal cells in prostatic glands revealed that these cells not only nourish and support the secretory epithelium of the glands, but also contribute to the morphogenesis, its hormonal metabolism and secretory function. It has also been exhibited that they are involved in the onset and development of cancer by producing reactive stroma. Accordingly, this study aims to analyze the gene expression patterns of stromal cells in prostate gland during prostate cancer in comparison to normal tissue.
Methods: The information related to normal prostate tissue was obtained by GEO database with GSE 26910 and the stromal cells' data was provided by GSE 35373. Four sets of data were gained for each group and the platform used for both was [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Data analysis was performed by R software. After entering the commands related to directory setting, showing microarray images, drawing histogram, data were normalized. Afterwards, heatmap commands were typed in order to observe the gene expression patterns.
Results and discussion: The results of data analysis and heatmap illustration showed that there was a significant difference between the gene expression patterns of stroma and normal cells which indicates the importance of these cells in tumorigenesis. Given the substantial difference in the gene expression patterns of stromal cells and normal tissue as well as the complexity of prostate cancer treatment, it seems that more attention to these cells could be an important therapeutic goal.

Key Words: *Keywords: stromal cells, prostate cancer, gene expression, prostate gland, R software*

Bioinformatics analysis of profilin genes (PRF) in brachypodium

Mohammad Ramezani Kaporchali

E-mail address (corresponding author): kaporchaly@gmail.com

Abstract

Profilin proteins (PRFs) are small (12-15 kDa) actin-binding proteins that play an important role in cytoskeletal dynamics and plant growth by regulating actin polymerization. In this study, PRF genes were selected in Brachypodium. Phylogenetic tree and prediction of cis-elements as well as analysis of chromosomal distribution in Brachypodium species were analyzed to evaluate the performance of these genes. Our results showed that the phylogenetic tree of PRF genes of different species (beans, Arabidopsis, Brachypodium and barley) were divided into two groups. Based on chromosomal position analysis, PRF genes were distributed on chromosomes 1 and 2. Also, based on the analysis of PRF protein sequences using MEME database, this protein contained 3 conserved motifs. Our findings can be considered as a useful source for future studies of PRF genes in studies between different plant species. The aim of this study was to identify and describe PRF genes in Brachypodium species through genome analysis.

Key Words: *Profilin protein, conserved motif, phylogenetic tree, chromosomal location*

A computationally optimized broadly reactive antigen analysis of tick midgut surface protein Bm86 in vaccine development against Rhipicephalus microplus and Rhipicephalus annulatus

Mohammad Mehdi Ranjbar[1]*, Gholamreza Karimi[1], Naebali Ahmadi[2], Sajjad Yazdansetad[3]

[1] *Razi Vaccine and Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran*
[2] *Proteomics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
[3] *Infectious Diseases Research Center, Golestan University of Medical Sciences, Gorgan, Iran*

E-mail address (corresponding author): sajjad.yazdansetad@gmail.com

Abstract

Background: The tick-borne diseases cause significant economic losses in animals and humans. The tick midgut surface protein Bm86 of Ripsephalus spp. is considered as an immunoreactive antigen for vaccine candidate against cattle fever ticks. In this bioinformatics report, we focused on the computationally optimized broadly reactive antigen (COBRA) tool to decipher the tick midgut protein Bm86 for vaccine strategy in the fight against Rhipicephalus species.
Materials and Methods: In this bioinformatics analysis, the consensus sequences of R. microplus and R. annulatus were retrieved, aligned with the GenBank data using the BLASTN algorithm and Sequencher v4.8 program, and finally edited. The phylogenetic trees were built using the maximum likelihood method. Bm86-based vaccine using the COBRA method was designed according to the taxa grouping and the center-of-tree (COT) sequences. The protein superimposition algorithms and molecular modeling analysis were also carried out.
Results: The hypervariable regions were identified in the amino acid residues 177-181, 270-276, and 351-352, respectively in both R. microplus and R. annulatus sequences. Six sequences were selected to anticipate the evolutionary vaccine and 12 sequences were used to realign and obtain the consensus sequences to do COBRA. R. annulatus sequences were in sister branches with more similarity to each other compared to Bm86 protein sequences in R. microplus except ADQ19687. The sequences AJE29931, AJE29932, ATW75472, ATW5476, ADM86722, and ACZ55133 were selected for vaccine designing according to the ancestral center-of-tree (COT).
Conclusion: Anti-tick vaccines using the COBRA approach could be a more cost-effective alternative with a broader spectrum compared to the commonly used recombinant vaccines.

Key Words: *Rhipicephalus; Bm86; Phylogeny; Vaccine; COBRA*

Diabetes type 2 and pancreatic cancer share biomarkers

Fatemeh Razavi[1], Zahra Akhlaghi[1], Atena Zadeh Rokni[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Pancreatic cancer is the 7th cause of death in the world. Pancreatic cancer has no symptoms until advanced stages and TP53 gene effects in 50% of cancers. There are some diseases like diabetes type 2 in which TP53 gene is involved. The aim of this study was to find the shared biomarker between pancreatic cancer and diabetes type 2 through bioinformatics analysis which can be effective to diagnose the disease in earlier stages. For this purpose, the associated Single Nucleotide Polymorphism (SNP) and microRNA with TP53 were found in miRdSNP database. DIANA Tools and LncRNADisease v2.0 databases helped to analyze the data in order to find a related long non-coding RNA (lncRNA) that has interaction with the same miRNA. Venn was used to find the shared effective genes between pancreatic cancer and diabetes and their interaction as well. The results showed that rs17884306 as SNP and has-miR-22-3p can interact on TP53 leading to pancreatic cancer. The lncRNA: AC060780.1 in pancreas tissue has also interaction with has-miR-22-3p which results in diabetes type 2. In addition, the lnc MALAT was investigated as the mutual one in both diseases. Moreover, the mutual genes including ACASB, NT5E, TEME237, MYC, PHLDA1 were obtained through Venn diagrams in the process of pancreatic cancer and diabetes type 2. Based on the current study and results, it is concluded that rs17884306 could be a useful biomarker to diagnose pancreatic cancer. The has-miR-22-3p and MALAT can prove that diabetes type 2 can be an underlying disease for pancreatic cancer, and they can be useful to diagnose pancreatic cancer in patient who have diabetics type 2.

Key Words: *TP53; rs17884306; has-miR-22-3p; AC060780.1; MALAT*

Selection of blood biomarkers to diagnose therapeutic resistance in people with epilepsy

Foroozan fatahinia[1]* Saeid Latifi-Navid[1], Kamran Ghaedi[1]

[1] *Department of Biology, Faculty of Science, University of Mohaghegh Ardabili, Ardabil, Iran*

E-mail address (corresponding author): foroozanfatahinia@gmail.com

Abstract

Epilepsy is a common brain disease affecting 70 million people worldwide. Antiepileptic drugs dramatically prevent seizures in patients, but about one-third of patients are resistant to these drugs. Therefore, finding blood biomarkers helps to diagnose and treat patients. Research shows that mutations in genes encoding ion channels play a role in epilepsy.LncRNAs play an important role in the central nervous system, with dysfunction and expression changes associated with epilepsy.

Method: Initially, Array express and GEO databases were examined to identify and express gene changes in this disease. Finally, the GSE71058 study, which is a type of RNA sequencing technology, was selected. According to the type of data, R software and related packages were used for analysis. After evaluation and changes in gene expression, a list of genes and LncRNAs that had more than double and significant expression changes (pvalue <0.05) were selected among the study groups. Because these ion channels play an important role in epilepsy, the list of genes related to ion channels was extracted using the HGNGO database. Also, to identify lncRNAs related to ion channel genes, the correlation between them was evaluated using the Pearson correlation coefficient. At this stage, lncRNAs with a correlation coefficient greater than +0.7 or less - 0.7 and a significant level of pvalue <0.05 were selected.

Results: According to the analyzes performed on the GSE71058 data, the hcn1, hcn2, hvcn1 genes, which are related to ion channels, and the LncRNAs of FOXN3-AS1 and SLC26A4-AS1 have more than double the expression changes and have become.

Conclusion: Considering the results of data analysis and changes in the expression of ion channel genes and related LncRNAs and their importance in causing epilepsy, it is suggested that changes in gene expression be studied in a detailed laboratory study.

Key Words: *epilepsy, ions channel, biomarker*

Investigation of competing endogenous RNA regulation of BCAS4 and SHISA7 in tau pathology in Alzheimer's disease

Hani Sabaie[1]*, Mohammad Reza Asadi[1], Zoha Salkhordeh[2], Maryam Rezazadeh[1]

[1] *Department of Medical Genetics, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Department of Medical Genetics, Afzalipour Faculty of Medicine, Kerman University of Medical Sciences, Kerman, Iran*

E-mail address (corresponding author): sabaieh@tbzmed.ac.ir

Abstract

Alzheimer's disease (AD) is a heterogeneous neurodegenerative disorder (NDD) that is increasing in prevalence globally. Emerging data suggest that GABAergic signaling may experience pathogenic changes and contribute to the pathological hallmarks of AD, such as tau phosphorylation. Moreover, it is widely established that a disruption in the equilibrium of competing endogenous RNA (ceRNA) crosstalk plays a role in NDDs [3; 4]. In this study, we employed bioinformatic approaches to investigate the ceRNA regulation of BCAS4 and SHISA7 in tau pathology. Two microarray datasets were downloaded from the Gene Expression Omnibus database, including information on mRNAs (GSE106241) and miRNAs (GSE157239) from individuals with AD who had varying degrees of AD-associated neurofibrillary pathology in temporal cortex (TC) tissue specimens and matched controls. The R package limma was used to discover differentially expressed mRNAs and miRNAs associated with AD-related neurofibrillary pathology. We utilized miRWalk (version 3) to identify interactions between miRNAs linked with AD-related neurofibrillary pathology and BCAS4/SHISA7. The ceRNA regulatory axes were constructed utilizing co-expression and miRNA-mRNA interactions. Only Braak stage V exhibited dysregulation of the BCAS4 and SHISA7 genes. Hsa-miR-185-5p targeted both genes BCAS4 and SHISA7 and its levels of expression were statistically higher in TC samples of AD than in controls. Based on co-expression and miRNA-mRNA interactions, BCAS4 was proven to act as a ceRNA for SHISA7 by sponging hsa-miR-185-5p in TC tissue specimens. The current work is the first evidence to highlight the expression of the BCAS4/hsa-miR-185-5p/SHISA7 ceRNA axis in the brains of AD patients.

Key Words: *Alzheimer's disease; BCAS4; competing endogenous RNA; miR-185; SHISA7*

Improving the prediction of physical protein interaction by Balanced Random Forest inter-protein residue contact predictions using sequence covariation information

Sara Salmanian[1]*, Hamid Pezeshk[2,3], Mehdi Sadeghi[4]

[1] *Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[2] *School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran (currently visiting Department of Mathematics and Statistics, Concordia University, Montreal, Canada)*
[3] *School of Biological Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran*
[4] *National Institute of Genetic Engineering and Biotechnology, Tehran, Iran*

E-mail address (corresponding author): salmanian1983@gmail.com

Abstract

Protein-protein interactions are essential for most cellular processes. There are a lot of protein interactions and a large number of protein sequences with unknown interacting partners. Prediction of protein interaction from sequence information has always been a great challenge. Those predictions would be more challenging when someone is supposed to specifically detect physical but not functional protein interplays. Therefore, developing new approaches for the accurate prediction of sequence-based physical protein interactions could be an important advancement in computational biology. Inter-protein spatially interrelating residue positions exhibit correlated patterns of sequence evolution in multiple sequence alignments. Those co-evolutions are wisely exploited for the prediction of physical protein interactions.
It is shown that feeding norm values of whole covariation information of protein heterodimers into Support Vector Machines (SVM), could accurately predict the possibility of physical interaction of those dimers using sequence information. In the present study, Balanced Random Forest (BRF) models were trained with the covariations of inter-protein residues at different hypothetical interacting sites and then the models were employed for the prediction of possible inter-protein residue contacts. Instead of considering whole co-evolutionary information, those BRF predictions could take into account the covariation information of more probable physically interacting residues for further prediction of protein dimers at higher protein scales. BRF predicted those more probable contacting residues as positive class and other interacting pairs of amino acids as negative. After BRF predictions, previously computed covariation scores of negatively predicted residue partners were zeroized, thereby the role of those pairs in the final calculation of norm values were driven out. Results of the current study indicated that feeding the updated norm values of residue-residue covariation matrices, obtained after BRF predictions, into SVM models could significantly increase the accuracy of the final protein interaction predictions at the protein family level.

Key Words: *residue contacts, physical interaction, covariation, protein interaction prediction*

Comprehensive RNA, protein interaction and SNP analysis of CAX1 in the Arabidopsis thaliana: integrated systems biology approach

Kianoush Sanei Ataabadi[1], Fatemeh Ahmadi[1], Zohre Nasirian[1], Behrooz Salehi-Eskandari[1]*

[1] *Department of Biology, Payame Noor University (PNU), Tehran, Iran*

E-mail address (corresponding author): behsalehi@pnu.ac.ir

Abstract

Calcium exchanger 1 (CAX1) is the encoder of vacuolar Ca2+/3H+ antiporter. This protein involves regulating Ca2+ levels in the Arabidopsis thaliana. In this study, we performed a protein-protein network analysis to demonstrate the possible regulatory gene network of the transporter Ca2+/3H+ in the Arabidopsis Thaliana. After literature mining, we indicated the protein-protein interaction network of the CAX1 gene in the Arabidopsis by STRING online software. The miRNA-mRNA interactions analysis of the selected genes was performed by the miRWalk 2.0 online database. The analysis of single nucleotide polymorphism (SNP) was performed by miRNASNP-v3. Based on this network analysis, the CAX1 has significant protein interactions with CAX3, TPC1, CAX11, ACA11, CXIP4, SOS2, SSE1, NHX1, T22C5.23, and ECA. There is no possible Arabidopsis microRNA interaction with the selected genes. However, there is a significant microRNA interaction with hsa-miR-6894-3p and SOS2 in the 3UTR position (Score: 1.00). Among mentioned genes, TPC1 is a key regulatory network in the conductance of sodium and calcium ions and the intravascular pH. ACA11 is a ATPase in the mature vacuole membrane expressed in transgenic plants and visualized in root cells. Based on miRNA SNP, rs1314179105 is an innovative and significant SNP on the 3UTR region of the SOS2 gene (14 gains and 15 losses). The expression level and RNA interactions of the mentioned genes are highly recommended to be evaluated in the Arabidopsis thaliana, bioinformatically and experimentally.

Key Words: *Arabidopsis Thaliana, protein interaction, gene network, SNP, CAX1*

Liver tissue transcriptome data analysis for identification of key miRNAs in sheep (Ovis aries) with different dites

Sajed Sarabandi[1], Farinaz Behfarjam[2], Mostafa Rafiepour[2]*

[1] *Department of Veterinary, Faculty of Sience and Medecin, Islamic Azad University Karaj branch, Iran*
[2] *Department of Genetics, Faculty of Science, Danesh Alborz University, Qazvin, Iran*

E-mail address (corresponding author): mostafarafiepour@yahoo.com

Abstract

Background: Today, genetic engineering methods used to change the genome in a way to produce more products like milk. Understanding regulatory mechanisms and the identification of small RNAs involved in metabolism is one of the most challenging problems in genetic science. Therefore, the aim of the present study was to identify some regulator genes in Ovis aries with different diets using bioinformatics analysis.
Materials and Methods: In this research, 13 samples of Illumina-sequence transcriptome data for sheep (Ovis aries) livers with different diets (high energy group n=3, oil supplement group n=3, light grazing n=2, over-grazing n=2 and control samples n=3) were collected. Firstly, quality control of data were done with FastQC program. Then, the reads aligned to the reference genome (ARS-UI_Ramb_v2.0) with STAR software. Finally, to identify the differentially expressed miRNAs, data were analyzed using FeatureCounts and DESeq2 programs, respectively.
Result: Our results showed several differentially expressed miRNAs in each group (high energy group n=13, oil supplement group n=11, light grazing n=10 and overgrazing n=10). For example, in oil diet, miRNA221 and miRNA29B significantly upregulated ($p<0.05$) that related to glucose metabolism. Also, the result showed expression of miRNA30 increased in oil supplement groups which has a key role in lipid metabolism. In the high energy and overgrazing groups, miRNA143 was significantly downregulated ($p<0.05$). Studies showed that this microRNA is an important regulator in glucose metabolism. In addition to, the resulted showed that the expression of mir16b, mir30b, and mir199a significantly decreased ($p<0.05$) in all groups. Studies indicated that these miRNAs have the important roles in glucose methbolism and immune system response.
Conclusion: In this study, our results indicated that different diets are determinant in regulating of key pathways and mechanisms like lipid metabolism and miRNAs can be candidate targets for purposeful nutrition.

Key Words: *Metabolism, miRNA, Ovis Aries, Gens expression, Bioinformatic*

An immunoinformatics approach for the design of a multi-epitope vaccine targeting lukD of Staphylococcus aureus

Parisa Khodabandeh shahraki[1], Mokhtar Nosrati[2], Mansour siavash[1]

[1] *Isfahan Endocrine & metabolism Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*
[2] *Department of Biotechnology, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): shahrakimailbox@gmail.com

Abstract

Background: S. aureus is the pathogen most frequently isolated from diabetes foot ulcer. These days, antibiotic-resident S.aureus is an active debate in science and clinical world. Today we can harness immunoinformatics tools with reverse vaccinology principles to design a multi epitope vaccine. In this study we aim to design a vaccine against lukD and lukE, which are two of most frequent virulence factor of S. aureus in diabetes foot ulcer.
Methods: The present study used immuno-informatics and computational approaches to identify both T-cell and linear B-cell epitopes of the lukD virolance factor of S. aureus. The three-dimensional structure of the final vaccine protein and conformational B cell epitopes was predicted with online servers.
Results: Among linear B-cell epitopes, "NITPKREKKVDD", "GDINISNGLSGGLN", and "NYKQESYRTTIDRKTNHK" were identified as highly antigenic. Among T-cell epitopes, "FQILTFNFI" (MHC class-I peptide) and "TSTYEVDWQNILLKL", "LNIFQILTFNFIKDK", "KLNIFQILTFNFIKD", "NDKLNIFQILTFNFI", "NQNTVTFTSTYEVDW", "LGKSSVASSIALLLL" (MHC class-II peptide) displayed the highest population score, and was immensely antigenic and conserved. All selected epitopes was non-allengen. We put these epitopes together with KK linker. Human Beta-defensin 1 and PADRE sequence were used as adjuvant. In vaccine tree dimensional structure, four conformational B-cell epitope were detected.
Conclusion: The present study identified several epitopes of the lukD virulence factor of S. aureus that are highly antigenic and conserved.

Key Words: *S.aureus, Staphylococcus, T-cell epitopes, B-cell epitopes, bioinformatics, Computational approaches, Vaccine design*

effective factor to change spermatogenesis in multiple sclerosis

Armin Sharifi, Shahab Atefi, Hosein Mohammadi

E-mail address (corresponding author): *arminsharifi1222@gmail.com*

Abstract

Introduction: behavioral and cognitive disabilities are rooted in Neurodegenerative diseases are rooted in the central nervous system (CNS). abnormalities in the genome and/or metabolism in brains stems Neurodegenerative diseases such as Multiple sclerosis (MS) is a chronic disability disease which is one of common disease in the world.in this study we investigate the cause of death ratio in PPMS patient. **Material and method:** In order to start this study, we used GEO accession number GSE131282. In our work, we considered patients with disease duration 54 PPMS years in PPMS as control groups and disease duration 4 years as treatment groups and by the use the limma package in R software data was analyzed. differentially expressed genes (DEGs) list by utilizing string database was analyzed to build Protein-protein interaction (PPI) network. Cytoscape 3.4.0 have been software that construct modules and Gephi software visualized our network.
Result: In case of PPMS, the number of DEGs were 845 genes which 707 were upregulated and 138 were downregulated. Analysis of modules shows that translation and regulation of apoptotic process are two major function. in addition, top ten of nodes determine as hubs genes such as RPL11and PLEC. **Conclusion:** Our study illustrated that changing in expression of TP53 and MAPK14 and so on can affected violence of PPMS disease. Thus patient with type of intensive PPMS has drastically decrease in spermatogenesis and cell differentiation.

Key Words: *#multiple sclerosis #spematogenesis#TP53*

Genome-wide identification and characterization of the metal tolerance protein (MTP) gene family in Eucalyptus grandis

Zahra Shirazi[1]*, Fatemeh Khakdan[2]

[1] *Biotechnology Research Department, Research Institute of Forests and Rangelands, Agricultural Research Education and Extension Organization (AREEO)*
[2] *Department of Biology, Farzanegan Campus, Semnan University, Semnan, Iran*

E-mail address (corresponding author): z.shirazi@rifr-ac.ir

Abstract

Metal tolerance proteins (MTPs) play an important role in the transport of metals at the cellular, tissue and whole plant levels. Considering the limited information on the MTP gene family in Eucalyptus plants which have phytoremediation properties and their different roles, this study aimed to identify the members, protein characteristics, simple sequence repeats (SSR) markers, miRNAs targets and gene expression analysis of this gene family by bioinformatics tools. The twenty MTP gene members were identified in the genome of Eucalyptus. The phylogenetic tree displayed that MTPs were categorized into three groups (Mn-MTP, Zn/Fe-MTP, and Zn-MTP) and seven sub-groups and were predicted to be localized in vacuoles. EqMTP proteins were predicted to encode polypeptides of 315-884 amino acids with molecular weights ranging from 36.07 to 97.32 kDa. Fifty-seven SSRs were identified in 18 MTP sequences and most of the genes had a single SSR. Identification of the simple sequence repeats (SSRs) marker of this family was found to be helpful in selecting the associated marker. Prediction of MTPs-targeted miRNAs, illustratd 14 EqMTP genes was targeted for 36 miRNAs. All the genes were suppressed by at least more than one miRNA, except for EqMTP9.1. EqMTP12 and EqMTP8.2 were suppressed by eight and seven miRNAs, respectively. Many metabolic pathways and activities related to plant growth and development, signal transduction, and reaction to stresses (including heavy metals) can be arranged by miRNAs. EqMTP11 gene expression was increced in various developmental stages and biotic stresses in different Eucalyptus species. The EqMTP11 gene was proposed as a good candidate for genetic engineering and breeding programs in order to introduce a metal-tolerant cultivar.

Key Words: *Cation diffusion facilitator; Eucalyptus; Heavy metals; Phytoremediation; Metal tolerance protein*

Identification and characterization of raffinose biosynthesis pathway genes in Vitis vinifera

Zahra Shirazi[1]*

[1] *Biotechnology Research Department, Research Institute of Forests and Rangelands, Agricultural Research Education and Extension Organization (AREEO)*

E-mail address (corresponding author): z.shirazi@rifr-ac.ir

Abstract

Grape (Vitis vinifera L.) is one of the most important fruit crops that various adverse abiotic stresses negatively affect its growth and yield. Raffinose is the classe of soluble carbohydrates that biosynthesis of its starts with activity of galactinol synthase (GolS) enzyme. In the next step, raffinose is made by raffinose synthase (RfS) activity. More than one isoform of GolS and RfS are found in different plant species that are expressed in various organs and responds to diverse environmental conditions. In the present investigation, we carried out a whole genome association study to identify the RfS and GolS gene families, protein characteristics, conserved motives, cis- element in prometor, gene structure and chromosomal location and duplication in grape by bioinformatics tools. Based on protein sequences of GolS and RfS found in A. thaliana; 15 VvGolS and 10 VvRfS were identified in grapes genome which clustered in five and six groups, respectively. The peptide length of RfS and VvGolS members was predicted from 346 to 792 and 162 to 590 amino acid residues. Subcellular localization of all VvRfS members' family was predicted in cytoplasmic, and most of the VvGolS members were also cytoplasmic. VvRfS had at least three motives of raffinose synthase or seed imbibition protein Sip1domain. The VvGolS family had contained most similar gene structure compared to VvRfS family. The VvGolS and VvRfS genes families were divided into ten linkage groups and four segmental duplication and tandem groups. The existence of four main cis-acting elements (ABRE, LTR, HSE and ARE) indicated that raffinose gene family is stimulated in cold, heat and anaerobic conditions. Abiotic stress can regulate the expression of raffinose gene family through binding to these cis-acting elements in upstream of VvGolS and VvRfS genes

Key Words: *Abiotic stress; Carbohydrates; Galactinol synthase; Raffinose synthase*

The Effect of a Single Nucleotide Polymorphism on a miRNA Function in the Papillary Thyroid Carcinoma (PTC)

Fatemeh Skandari[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and Vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Papillary Thyroid carcinoma (PTC) is the most common form of well-differentiated thyroid cancer, and the most common form of thyroid cancer to result from radiation exposure. PTC has shown strong heritability and no predisposing mutations in the germ-line. In the current study, we have researched the effect of common T/G polymorphism (rs1357866392) as an SNP on the miR-34a-3p sequence and its binding to the mRNA sequence of MET proto-oncogene, receptor tyrosine kinase gene (MET) (ENSG00000105976.14) in PTC through bioinformatics. The MET gene encodes a member of the tyrosine kinase receptor family of proteins and the MET proto-oncogene product. For this purpose, GEO datasets showed the top 250 differentially expressed genes. The David database was used to cluster the gens and it revealed the involved genes in transcriptional misregulation pathway in cancer including MET gene. The miRNASNP-v3 database showed the relationship between studied miRNA and SNP. The results showed that binding to the 3'UTR region of MET, the miR-34a-3p blocks gene's expression, and acts as a tumour suppressor in PTC. The occurrence of rs1357866392 (T/G on chr1:9151709) in the seed match of miR-34a-3p (chr1:9151708-9151714) and MET (chr7:116796124-116798386), made the binding to be lost. It makes the gene not to be under the control of miRNA anymore and upregulation will occur. Thus, our data suggest that rs1357866392 is a disease-associated SNP since as a common polymorphism in miR-34a-3p affects the regulation of the MET gene and result in the genetic predisposition to PTC. Its role in the tumorigenesis through somatic mutation Preliminary evidence suggests that these effects are mediated through target genes which expressions are affected by the SNP status.

Key Words: *PTC, MET gene, miR-34a-3p, rs1357866392*

How to used Bioinformatics Tools to Disease Variant Detection?

Eskandar Taghizadeh[1], Gordon A. Ferns[2], Majid GhayourMobarhan[3]

[1] Department of Medical Genetic, Faculty of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
[2] Department of Medical Education, Brighton and Sussex Medical School, Falmer, Brighton, UK
[3]Metabolic Syndrome Research Centre, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

E-mail address (corresponding author): eskandar.taghizadeh@yahoo.com

Abstract

Whole exome sequencing (WES) is a powerful tool to identifying mutations in rare genetic disorders. The aim of present study was to identifying the genetic variants in a family with three affected members born from consanguineous marriages. The affected members had frequent attacks of apnea for 20–60 s with bradycardia and hypotonic; abnormal breathing patterns; fast or slow breathing (hyperpnoea), abnormal movement in eyes, distinctive facial features, ataxia and intellectual disability episodic hyperpnoea, polydactyly, syndactyly and Brain MRI showed the molar tooth sign, atrial septal defect (ASD) and ventricular septal defect (VSD).
After WES with using NextSeq 500 Illumina platform with 100 million reads (100 X) the sequence reads were mapped to the hg38 genome version. GATK pipeline were used for variant calling and in the next step we used the wANNOVAR tool for variant annotation. There were about 178,000 variants in raw data of proband and after this we used bioinformatics analysis for sequencing results by using standard bioinformatics software and international databases. We used filtration against benign variants such common single nucleotide polymorphisms that were previously reported in some genetic related databases as non-pathogenic or those variants with frequency more than 1%. Also we excluded variants that did not fit with the expected pattern of family pedigree. Finally, novel variants that had a deleterious effect on proteins and tissue expression were considered. We checked selected genes variants in VarElect, Phenolyzer. After this we used prediction software's such as PolyPhen 2, SIFT, MutationTaster, FATHMM, PROVEAN and international databases including, OMIM, GeneCards and MalaCard. Finally, we found a new variant in C5ORF42 gene (c.3080A>T: p. D1027V) which is a related gene with Joubert syndrome. Confirmation tests were carried out by PCR and Sanger sequencing for identified variant in C5ORF42 gene in proband and her family members.

Key Words: *Keywords: Bioinformatics, Whole exome sequencing, Joubert syndrome, C5ORF42 gene*

Differences between hg-19 and hg-38 in transcriptomic analysis of cumulus cells in PCOS patient

Mohammadhosein fakherbaheri[1], Kimia Monshizadeh[1], Mohammad Yahya Vahidi Mehrjardi[1]*

[1] *Department of Medical Genetic, Medical School, Shahid Sadoughi University of Medical Science, Yazd, Iran*

E-mail address (corresponding author): mmvahidi@gmail.com

Abstract

Understanding of oocyte maturation process is prominent in infertility malady's such as Polycystic ovary syndrome(PCOS). PCOS patients need to be assisted reproduction technologies (ARTs) like IVF for fertilization. So this is a prerequisite to identify oocyte with reproductive power. In this study we analyzed cumulus cells which are around oocytes, they are appropriate choice due to their directive relationship with oocyte and separating of them is a non-invasive method to check oocyte maturation. We analyzed 5 transcriptomic profile cumulus samples from PCOS patients (2 sample were for GV oocyte and 3 samples were for MII oocyte) with RNA-Seq metod with hg-19 and hg-38. This study showed us 1775 DEG with hg-19 and 1403 with hg-38 because of this were various significant pathways and gene ontologies. Based on the different gene expressions of hg19 and hg38, real time pcr should be applied to confirm expression differences

Key Words: *PCOS, cumulus cell, hg-19, hg-38, transcriptomic analysis*

Bioinformatics analysis of gene expression profiling for identification of key proteins associated with non_alcoholic fatty liver disease

Maryam Yaraghi Isfahani[1], Dr Najaf Allahyari Fard[2]*

[1] *Higher Education Institute, meymeh. Iran*
[2] *The National Research Center for Genetic Engineering and Biotechnology, Tehran, Iran*

E-mail address (corresponding author): maryamye20@yahoo.com

Abstract

Introduction: Non-alcoholic fatty liver disease (NAFLD) is a condition in which triglycerides accumulate in the liver cells of people who have no history of alcohol use. Damage to liver cells such as apoptosis, necrosis, inflammation and then fibrosis are features of this disease.. The importance of this disease is due to the destruction of liver cells, which is known as one of the most common and important causes of cirrhosis and liver failure in the world. Its prevalence in the general population in different countries is about 3-25%, and in the general population of iran is about 2.9-7.1%, and in diabeticsis is about 55.8%. this statistic is increasing due to lifestyle changes.
Methods: By visiting the Geo site, the raw micro data was downloaded with the access code GSE89632. Using GEO2R online software, genes with different expression (DEGs) were selected in two cases of fatty liver disease, in cluding: simple steotitis and non- alcoholic steatohepatitis due to their significance. Enrichment analysis of altered gene groups for cell pathway and processes associated with nonalcoholic fatty liver disease was performed using the metascape web server and the construction of a protein interaction network (PPI). Key genes and proteins involved in NAFLD were also identified using string, cytoscape and centiscape software.
Results: 14 important and key expressed mutation genes between healthy and diseased groups in two different cases of non-alcoholic fatty liver disease, using protein-protein interaction network (PPI) and name-related software known as: TYMS ،ASPM ،MYC ،IL6 ،VEGFA ،JUN ،IL1B ،FOS ،TLR2،IL10 ،CXCL8 ، CD44 ،MMP9 and also obtained similar cellular pathways and processes related to NAFLD is drawn below.
Conclusion: The present study shows that some important genes and pathways may be associated with the occurrence and progression of NAFLD disease. Important biomarkers for prevention, treatment and new therapeutic goals were identified in this study.

Key Words: *Non-alcoholic fatty liver disease, microarray technique, protein interaction network, gene expression pattern.*

Bioinformatics Study on Slc34a2 Gene in Lung Cancer Tumorigenesis

Atena Zadeh Rokni[1], Fatemeh Razavi[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

A novel paradigm in tumor biology suggests that Non-Small Cell Lung Cancer (NSCLC) growth is driven by lung cancer stem cell-like cells (LCSCs). SLC34A2, a member of the solute carrier gene family, encodes NaPi2b (the type II Na/Pi co-transporter). NaPi2b is a multi-transmembrane sodium-dependent phosphate transporter responsible for transcellular inorganic phosphate absorption, maintenance of phosphate homeostasis, and tumorigenesis. The biological function of SLC34A2 and its underlying mechanisms in NSCLC has remained unclear. In this study, the SLC34A2 gene was the target for bioinformatics findings. For this purpose, single nucleotide polymorphism (SNP) and microRNA (miRNA) associated with SLC34A2 were found in the GeneCards and NCBI databases. LncRNADisease (version 2.0) database helped to analyze the data to find a long non-coding RNA (lncRNA) related to same miRNA. The analysis showed that the occurance of the SNP:rs5996397 can lead the cells to lung cancer and affect the involved miRNA-650 and lncRNA-MEG3 in this process. The rs5996397 occures at the seed match of miRNA-650 on 3'UTR of SLC34A2 gene. According to findings from the GEPIA2 database, the SLC34A2 gene is highly expressed in lung cancer. Through the process of this disease, the disease-associated SNP: rs5996397 is located on the mRNA and causes the disease. When miRNA-650 binds to 3'UTR of SLC34A2 mRNA, it increases gene expression and intensifies the tumorigenesis of LCSCs. When lncRNA-MEG3, which is a tumor suppressor, binds to miRNA-650 and suppresses SLC34A2, it silences miRNA-650 and reduces pathogenicity. based on different aspects of this pathway, it is inferred that the interaction of lnc-MEG3 with rs5996397 could be useful biomarkers for the diagnosis of lung cancer. In conclution, the miRNA-650 prove that SLC34A2 is an effective gene in lung cancer outbreak and its tumorigenesis.

Key Words: *Key words: NSCLC; miRNA-650; rs5996397; lnc-MEG3; SLC34A2.*

Comparative genomics of prokaryotes

Shiva zarean[1], Samin Golchin[1], Negin Daneshmand[1]

[1] *Azad university of Falavarjan, Isfahan, Iran*

E-mail address (corresponding author): shivazarean1@gmail.com

Abstract

Comparative genomics methods enable the reconstruction of bacterial regulatory networks using available experimental data. The analyses of type III secretion and SOS response regulatory networks illustrate instances of convergent and divergent evolution of these regulatory systems, showcasing the power of formal ancestral state reconstruction at inferring the evolutionary history of regulatory networks.Gene order in prokaryotes is conserved to a much lesser extent than protein sequences.Only several operons, primarily those that code for physically interacting proteins, are conserved in all or most of the bacterial and archaeal genomes. Multiple genome alignments are a resource for studies on operon rearrangement and disruption, which is central to our understanding of the evolution of prokaryotic genomes.
The coverage of the archaeal genomes was only slightly lower than that of bacterial genomes. The majority of the conserved gene strings are known operons, with the ribosomal super operon being the top scoring string in most genome comparisons. Subsequent comparisons of the sequenced bacterial and archaeal genomes have shown that even most of the operons are extensively rearranged during evolution. Only a few operons, typically coding for physically interacting proteins, are conserved in all or most of the genomes.In this research We were interested in systematically and quantitatively exploring the conservation of gene order among the 25 currently available complete bacterial and archaeal genomes and its dependence on the evolutionary distance between these genomes, and in assessing the potential of such comparisons for identification of previously undetected operons and prediction of gene functions

Key Words: *Comparative genomics, Prokaryotes, Ribosome, Operon, Bacterial genomes*

Toward elucidation of shoot-root communications involved in the developmental transition of C3 to C4 photosynthesis by comparative transcriptome analysis of hypocotyls in Halimocnemis mollissima

Mahdis Zolfaghar[1], Mohammad Reza Ghaffari[2], Ali Mohammad Banaei-Moghaddam[1]*

[1] *Institute of Biochemistry and Biophysics, Laboratory of Genomics and Epigenomics (LGE), University of Tehran, Tehran, Iran*
[2] *Department of Systems Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Karaj, Iran*

E-mail address (corresponding author): m.zolfaghar@ut.ac.ir

Abstract

C4 photosynthesis increases carbon fixation efficiency by separating photosynthesis phases into the mesophyll and bundle sheath cells and concentrating CO2 around Rubisco, reducing photorespiration in arid and hot environments. Consequently, C4 plants have higher water and nitrogen use efficiency with increased photosynthesis rates than C3 species. As a striking instance of convergent evolution, evolving unique anatomical and biochemical features from C3 progenitors in more than 60 independent lineages ensures these characteristics. Nevertheless, exploring the evolutionary paths and identifying the global regulators of crucial components of C4 photosynthesis are constrained by the phylogenetic noises in comparative studies. H. mollissima switches C3 (in cotyledons) to C4 photosynthesis (in first leaves) during its life cycle. Such species provided an excellent model for studying the C3 to C4 transition regulations.
Here, we aimed to investigate probable long-distance communication pathways involved in this developmental transition through transcriptome analysis of hypocotyls. Thus, high-quality RNA was extracted from hypocotyls before and after the first leaves' formation and sequenced by Illumina Hi-seq 2000 sequencer. After de novo assembly, we identified differentially expressed genes between two developmentally different hypocotyls. They belonged to biosynthesis, transport, and signaling of phytohormones, various transcription factors (TFs), and plant signaling peptides. Two members of the GRAS family of TFs, SHORT-ROOT (SHR) and SCARECROW (SCR), and one member of the bZIP TFs, HY5, were more expressed in hypocotyls after the formation of first leaves. It has already been shown that SHR is a mobile transcription factor and SCR controls Kranz anatomy in maize leaves. Moreover, HY5 is a master regulator and a phloem-mediated shoot-to-root signal that adjusts the carbon-nitrogen balance in Arabidopsis. Our findings show the possible signaling pathways involved in the emergence of more advantageous features of C4 photosynthesis.

Key Words: *C4 photosynthesis, Halimocnemis mollissima, Long-distance signaling, RNA sequencing, Shoot-root coordination*

Computational Drug Design and Discovery

Molecular docking and ADMET study of bioactive compounds of Salvia officinalis against main protease of SARS-CoV2

Tooba Abdizadeha[1]*

[1] *Clinical Biochemistry Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran*

E-mail address (corresponding author): medicinalchemistry18@gmail.com

Abstract

COVID-19 is an infectious illness caused by SARS-CoV-2 has been declared as a global pandemic by WHO. There is an emergent need to search for possible medications and we investigated the potential of compounds of Salvia officinalis against main protease (Mpro) of the SARS-CoV-2. The molecular docking process was performed using Molecular Operation Environment (MOE) software to predict the mode of interaction between the best possible biological conformations of compounds in the active site of Mpro enzyme. The 2D structures of compounds of Salvia officinalis such as Carnosol, Jasmonic acid, Salvianolic acid A, Tanshinone I, Caffeic acid and Luteolin were prepared by Chem Draw ultra 8.0 software and converted into 3D format by Hyper Chem7 using AM1 semiempirical method. The compounds were docked into the active site of Mpro (PDB ID: 6LU7) by MOE software. The best pose of compounds with the higher score was selected for ligand-target interaction analysis by the LigX module in MOE software. The docking results showed a high potency of Carnosol and Luteolin as Mpro inhibitors with binding energies of − 12.40 and − 11.85 kcal/mol, respectively. Docking studies shows that Salvia officinalis compounds bind strongly with some of the amino acid residues in the active site of Mpro and these active compounds could form π-π stacking interactions with His41, Met49 and Met165 and hydrogen bonds with Cys145, Gly143, Ser144, Glu166 and electrostatic interactions with His163, Gln189 and Asn412. In silico ADMET properties prediction also shows that Salvia officinalis active compounds had good solubility, absorption, permeation, and non-toxic characteristics. The results of the study showed that two active compounds of Salvia officinalis have high binding affinity with Mpro of SARS-CoV2 and have good ADMET properties and could be considered as promising compounds for the development of COVID-19 potential inhibitors after further studies.

Key Words: *COVID-19, SARS-CoV-2, Main Protease, Molecular Docking, Salvia officinalis*

Hologram quantitative structure-activity relationship (HQSAR) study for anticancer activities of 2H-chromen-2-one derivatives

Tooba Abdizadeha[1]*

[1] *Clinical Biochemistry Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran*

E-mail address (corresponding author): medicinalchemistry18@gmail.com

Abstract

Cancer is one of the deadliest diseases and the second-leading cause of death in the world and remains a significant health problem worldwide. The numerous drugs have been used for the cancer treatment but have severe side effects. The current status in the fight against cancer suggests that there is a critical demand of more potent anticancer agents with the lowest toxicity and resistance profile. Consequently; increasing interest has been devoted to the design and discovery of more effective anticancer agents in current medicinal chemistry. In this work, Hologram quantitative structure-activity relationship (HQSAR) models using 2H-chromen-2-one derivatives were generated to discover the relationship between the different chemical structures and the anti-cancer activity of agents. In the proposed HQSAR model, three fragment parameters, fragment distinction, fragment size and fragment length, were set to "A, C and H", "4-7" and "53" respectively. Conventional validation techniques, internal and external validations such as, non-cross-validated correlation coefficient (r^2_{ncv}), cross-validated correlation coefficient (q^2) and predicted correlation coefficient (r^2_{pred}), were utilized to evaluate the forecasting accuracy of proposed model. The HQSAR model (q^2, 0.851; r^2_{ncv}, 0.974; r^2_{pred}, 0.841) for data set (training and test set) of anticancer agents yielded significant statistical results. The HQSAR contribution maps generated from these models illustrated that the yellow, blue, green-blue and green fields played key roles for improve the anti-proliferative activity of anticancer agents. The final QSAR models could be useful for rational design and development of novel potent anticancer agents in cancer treatment.

Key Words: *HQSAR, anti-Cancer, 2H-Chromen-2-one, Fragment Size, Fragment Distinction*

Comparison of the binding curcumin and Imatinib to some oncogenic tyrosine kinases

Hanie Abdollahzade[1]*, Yaghub Pazhang[1]

[1] *Department of biology, faculty of sciences, Urmia University, Urmia, Iran*

E-mail address (corresponding author): st_h.abdollahzadeh@urmia.ac.ir

Abstract

Background: Some tyrosine kinases, including c-SRC, c-ABL, and BCR-ABL, are essential in cancer promotion and progression. Some compounds bind to the kinases and inhibit their functions and reduce cancer development. This study chose two crucial compounds, Imatinib and curcumin, to compare their binding to the kinases and calculate their affinity.
Methods: Autodock-vina, ViewerLit, BIOVIA, LIGPLOT+, and OpenBable. To clarify the active site, we used BIOVIA; both Autodock and Autodock Vina have analyzed the interaction between drug and protein.
Results: The results of the Autodock-vina software (CPU 8 and RMSD: 0.000) showed that Imatinib had the most affinity to c-SRC protein with affinity energy -9.1 kcal/ mole. Curcumin on c-ABL protein with affinity energy -7.8 kcal/mole had the most significant impact.
Discussion: As the results showed, Imatinib and curcumin bind effectively to c-Src and c-ABL, respectively. Therefore the compounds may reduce the growth of the cancer cells overexpressing the kinases. These results are theoretical and require laboratory and experimental results.

Key Words: *Keywords: Molecular Docking, ligand, Autodock, BCR-ABL, c-Src, c-ABL*

14-3-3 sigma: new peptidic inhibitors by fragment-based drug design

Yasin Fahimi[1], Dara Dastan[1], Ahmad ebadi[2]*

[1] *Department of Pharmacognosy, School of Pharmacy, Medicinal Plants and Natural Products Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[2] *Department of Medicinal Chemistry, School of Pharmacy, Medicinal Plants and Natural Products Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*

E-mail address (corresponding author): a.ebadi@umsha.ac.ir

Abstract

Modulating disease-relevant protein-protein interactions (PPIs) is a reliable and evolving approach. Studies have revealed almost 700 PPIs of the 14-3-3 family. Regulating a wide variety of signaling pathways, 14-3-3s are potential targets for drug discovery. 14-3-3 Sigma is a unique isoform which is most associated with the occurrence and development of malignancies. Overexpression of SFN (14-3-3 sigma gen) is related to gallbladder, liver, and breast cancer. Inhibition of 14-3-3 sigma interactions can both lead malignant cells to apoptosis and prevent the patient from chemotherapy resistance. Due to the structure of 14-3-3sigma in protein interactions, formation and occupation, the U-shaped groove between the both monomers of 14-3-3 sigma, plays the most important role in establishing complexes that lead to cell cycle regulation. The design of anti-cancer peptides (ACPs) is a promising method for inhibiting protein interactions. So, in this study, we designed ACPs to inhibit the sigma isoform using a fragment-based drug design (FBDD) strategy. After binding the potential fragments, a peptide library consisting of 1059049 peptides were designed. We performed a supervised machine learning in such a way that dataset were divided into 70% to 30% for training and testing, respectively. Classification and logistic regression model were used for analysis in which 100 high-scored peptides were docked at the amphipathic groove of 14-3-3sigma. In the next step, 10 selected peptides were examined by molecular dynamic simulation. Finally, p11 (NKWRRF, Mw: 840.48 gr/mol) and p15 (NRWRRF, Mw: 937.54 gr/mol) hexapeptides with the least free energy were synthesized by the solid-phase peptide synthesis (SPPS) and then characterized by LC–MS and RP-HPLC.

Key Words: *PPIs, peptidic inhibitor, 14-3-3 sigma, FBDD, Machine Learning*

Microarray analysis for Glioblastoma drug-repositioning

Zahra Alaeddini[1]*

[1]*Department of Mathematics, Alzahra University, Tehran, Iran*

E-mail address (corresponding author): z.alaeddini@student.alzahra.ac.ir

Abstract

Background: Glioblastoma is an aggressive type of cancer that can occur in the brain or spinal cord. Glioblastoma forms from cells called astrocytes that support nerve cells. Glioblastoma can occur at any age but tends to occur more often in older adults. One of the most important things to treat this type of brain cancer is an appropriate drug, but drug discovery is a high-priced pipeline that may be taken a long time to achieve a good component that satisfies most of the pharmacokinetic assays for our interested disease. One way to reach this aim is the drug repositioning that we investigate in this research.
Materials and Methods: We downloaded GSE4290 from the GEO database for our purpose and did an analysis on it by R. After that we extracted genes that had high expression by setting LogFc >1 and adj.p.val < 0.05. These genes did input as queries on the Enrichr web tool for investigating diseases pathway that these genes have an effect on them. One of the diseases based on the KEGG 2021 Human, was small cell lung cancer. We search on ChEMBL for small cell lung cancer relative compounds.
Results: by filtering on the ChEMBL results based on the Lipinski and Ghose, drug-likeness roles, we compared the SMILE structures of TALAZOPARIB which is accessible by results filtering, and EVEROLIMUS that is used for Glioblastoma treating, on the Swiss ADME. The TALAZOPARIB satisfies all Drug-likeness assays than EVEROLIMUS and based on the ChEMBL references TALAZOPARIB is used for treating many types of cancers, so we can use instead of EVEROLIMUS.
Conclusion: When drug discovery has a huge financial load on drug companies, drug repositioning is a good and fast choice for them to provide the best drug component with satisfying most of the pharmacokinetic assays and drug-likeness roles.

Key Words: *Glioblastoma; Drug repositioning; drug discovery; cancer*

Bioinformatic screening of Lamiaceae family compounds as potential inhibitors of spike-mediated receptor-binding of SARS-CoV-2

Abbas Alibakhshi[1]*, Shima Gharibi[2], Shahrzad Ahangarzadeh[3]*, Fatemeh Yarian[4]

[1] *Molecular Medicine Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[2] *Core Research Facilities (CRF), Isfahan University of Medical Sciences, Isfahan, Iran*
[3] *Infectious Diseases and Tropical Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*
[4] *Department of Medical Biotechnology, School of Medicine, Fasa University of Medical Sciences, Fasa, Iran*

E-mail address (corresponding author): alibakhshi2630@yahoo.com

Abstract

COVID-19 (coronavirus disease 2019) still remains a major challenge worldwide. The coronavirus causes the disease by binding to ACE2 receptors on lung cells, infecting the cells and triggering the onset of symptoms. The prevention of such binding in which the virus is eventually unable to enter the cell could be a promising therapeutic approach. In this in silico study, 306 compounds of Lamiaceae family native in Iran were retrieved from ChEBI, PubChem, Chemspider and ZINC15 serevers as 3D structures in SDF files. Crystal structure of SARS-CoV-2 spike RBD was obtained as bound with ACE2 from Protein Data Bank (PDB) database (PDB ID: 6M0J). Then, the compounds with inhibitory potential were selected by molecular docking and virtual screening using Autodock vina tools embedded in PyRx software, in terms of free energy binding against the spike protein of the virus. The pharmacokinetic profile of selected compounds was evaluated using SwissADME website. Afterwards, by molecular dynamic simulation using GROMACS package and RMSD and RMSF analysis as well as MM/PBSA method, four compounds were further assessed for binding affinities against the receptor-binding domain of spike. The results showed that two compounds, Catechin gallate and Perovskone B derived from Stachys and Salvia genus generated stronger binding affinity against the target and thus could act as potential inhibitory compounds of RBD of the SARS-CoV-2 spike protein.

Key Words: *Lamiaceae family; SARS-CoV-2; RBD; Molecular docking; Molecular dynamic*

Evaluation of native plants of Hamedan province through their potential ability to treat colorectal cancer at the molecular level by systems biology, drug design and experimental methods

Abbas Alibakhshi[1]*, Ali Shojaeian[1], Atefeh Asgari[2]

[1] *Molecular Medicine Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[2] *Hamadan Core Facility, Hamadan University of Medical Sciences, Hamadan, Iran*

E-mail address (corresponding author): alibakhshi2630@yahoo.com

Abstract

Colorectal cancer is the second deadliest and third most common cancer in the world. The use of computers and bioinformatics methods in laboratory studies has increased and especially in the diagnosis of factors involved in the disease and the identification of molecules with high therapeutic potential have been widely used, which shows the importance of this method. Also, the importance of effective compounds of medicinal plants in human life has found a special place. In this study, the list of compounds from native plants of Hamedan province were obtained from literatures and their information and 3D structures in SDF files were retrieved from ChEBI and PubChem servers. The SDF files were entered into the PharmMapper server, an integrated pharmacophore matching platform for potential target identification. The 20 top potential targets of all compounds were shared and then compared with all factors involved in the colorectal cancer, obtained from NCBI gene database. Afterwards, between the potential targets of plant compounds and factors obtained from colorectal cancer as targets of this cancer, a drug-target (DT) network was constructed using Cytoscope software. Molecular docking was used for the best targets and plant compounds with the highest betweenness centrality and their interaction was investigated by the interaction energy using Autodock vina tools embedded in PyRx software and Discovery Studio software. The three potential compounds were then further analyzed using the GROMACS package and the analysis of RMSD and RMSF as well as interaction energies. the results showed that some compounds from Apiaceae, Hypericaceae and Lamiaceae families exhibited good potential inhibitory effects against several targets such as MAPK14, Insulin receptor and Angiogenin. In the following, we effect of these compounds on some of the targets will be examined in vitro.

Key Words: *Native plants; Colorectal cancer; Drug-target network; Molecular docking*

Computational design of peptide inhibitors targeting the interaction of ACE2 and the spike receptor-binding domain of the SARS-CoV-2 Omicron variant

Fatemeh Arabi Jeshvaghani[1]*

[1] *Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): f.arabi68@yahoo.com

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the seventh human coronavirus that causes the coronavirus disease (COVID-19) after attacking the upper respiratory system and may lead to respiratory, enteric, hepatic, and neurologic diseases. Omicron variant (B.1.1.529) is a recent SARS-CoV-2 variant of concern (VOC) with a large number of mutations in spike protein and with impacts on viral transmissibility, disease severity, and efficacy of vaccines and therapeutics. Here, we designed a set of peptide inhibitors to block the protein–protein interactions (PPIs) of the receptor binding domain (RBD) of the SARS-CoV-2 Omicron variant with human ACE2 using computational methods. Based on the key interacting residues involved in the interaction site of RBD and ACE2 that were analyzed by hotspot prediction tools, PIC and Ligplot^{+}, a set of inhibitory peptides was designed. CABS-dock was used for molecular docking of designed peptides with spike protein of Omicron variant. Prodigy and HawkDock webservers predicted the binding affinity of peptide-protein complexes and MM/GBSA free energy decompositions. Drug-likeness and ADME-Tox analysis were done by SwissADME and FAF-drugs4 to compute the physicochemical descriptors and properties of the designed peptides. Peptides are less immunogenic than recombinant proteins or monoclonal antibodies, with higher solubility and better biological efficiency than therapeutic proteins. Antibodies cannot penetrate to the cell membrane to target intracellular PPIs and small molecules are not promising candidates for targeting challenging desired large and flat PPIs binding sites. The designed interfering peptides with the most negative binding energies, the proper binding sites, and the acceptable properties are the potential candidates to develop as novel anticoronaviral therapeutics for COVID-19 or future related CoV's outbreaks.

Key Words: *SARS-CoV-2; Protein-protein interactions; Inhibitory peptide; Omicron variant; Receptor-binding domain*

Ligand Discovery for the human monocarboxylate transporter 1 (MCT1) in an open-outward conformation by virtual screening on ZINC's FDA-approved drugs

Mina Barhoon[1], Hamid Mahdiuni[1], Soraya Sajadimajd[1]

[1] *Bioinformatics Lab., Department of Biology, School of Sciences, Razi University, Kermanshah, Iran*

E-mail address (corresponding author): hmahdiuni@razi.ac.ir

Abstract

Monocarboxylate transporters (MCTs), encoded by the solute carrier 16 (SLC16) gene family, are a 14-member group of membrane transporters facilitating the displacement of monocarboxylates like pyruvate, lactate, ketone bodies, short-chain fatty acids, and thyroid hormones. The extracellular transport of lactate occurs mainly by MCT4 and intracellular uptake of lactate is mediated by MCT1, being a proton-dependent process involved in the regulation of intracellular pH. It has been shown that overexpression of MCT4 and MCT1 is associated with the development of a variety of malignancies including breast, stomach, lymphoma, brain, lung, skin, and soft tissue cancers. In the tumor microenvironment, differences in cancer cells' access to nutrients and oxygen modify cellular metabolism. Cancer cells in a hypoxic state turn to glycolytic metabolism to continue their survival and proliferation, producing large amounts of lactate inside the cell, which must be transported out of the cell via MCT4. Conversely, oxidative cancerous cells express MCT1 to enter excess lactate, using it as the preferred fuel instead of glucose. This process, called "metabolic symbiosis" can be targeted as a potential treatment for a variety of cancers. MCT1 inhibition in oxidative cells increases the rate of glucose consumption, causing glycolytic cell death due to glucose deficiency and acidification of their cytosol. In this study, a structure-based virtual screening was performed using the MCT1 atomic coordinates (in outward-open conformation) downloaded from protein data bank (PDB) with a code of 6LYY, and 1778 FDA-approved drugs downloaded from the ZINC database. "FAF-drugs4" webserver and "Open Babel" software were used to remove PAINS compounds and ligand preparation, respectively. Molecular docking calculations were done using "Autodock Vina", "Molegro Virtual Docker" and "DOCK6" programs. The ligands that showed high binding energy were visually analyzed and were introduced for further studies.

Key Words: *Monocarboxylate transporter1, Virtual screening, Chemical inhibitors, Cancer*

Application of activity cliffs to virtual screening for identify lactate dehydrogenase inhibitors using machine learning approaches

Sedigheh Damavandi[1], Abbasali Emamjomeh[1,2], Fereshte Shiri[1]

[1] *Laboratory of Computational Biotechnology and Bioinformatics (CBB), Department of Bioinformatics, Faculty of Science, University of Zabol, Iran*
[2] *Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Zabol, Iran*

E-mail address (corresponding author): sedigheh.damavandi@mazust.ac.ir

Abstract

Activity Cliffs (ACs) are groups of structurally identical active chemicals with significant potency variances. Since ACs are so common in SAR data, it's crucial for modern computer-aided drug design and discovery to be able to deal with them effectively. ACs, on the other hand, pose considerable challenges for bioactivity-supervised discovery approaches that presume smooth and continuous structure-activity connections. Many ligand-based approaches have high-performance predictions, indicating that the hypothesis is correct. However, there are certain limitations. They are unable to explain the activity cliff , and their results for new compounds are suboptimal. The ACs verified to yield pharmacophore and machine learning models that were equivalent to known ligand- and structure-based pharmacophores in terms of accuracy. We applied the ACs to investigate the protein Lactate dehydrogenase which has a high-resolution crystallographic structure and a number of known inhibitors. The number of ACs was determined in each protein's inhibitor population. The missing edges (e.g. unknown interactions) are predicted using various kinds of Machine Learning models. Then, we have searched a large number of machine learners (MLs) to see if we could link protein features to the existence or absence of ACs in the ligand population. Therefore, by identifying ACs that had not been previously considered, the presence of different atomic shares and the differential effects of power associated with ACs formation are determined, which indicated the occurrence of unknown interactions. Finally, using the information deriving from the activity cliff analysis to suggest how virtual screening protocols might be improved to favor the early identification of potent and selective lactate dehydrogenase inhibitors in molecular databases.

Key Words: *Key words: Activity Cliffs; Machine Learning; Lactate dehydrogenase; virtual screening.*

Repurposing of Latanoprost as an ADAM17 inhibitor: A ligand- and structure-based approach

Kourosh Daneshvarnejad[1], Homa Azizian[2]*

[1] *Department of Medicinal Chemistry, Faculty of Pharmacy, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran*
[2] *Department of Medicinal Chemistry, School of Pharmacy International Campus, Iran University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): kourosh.daneshvarnejad@outlook.com

Abstract

Background: A disintegrin and metalloproteinase 17 (ADAM17), with its role in shedding various intercellular mediators, including tumor necrosis factor α (TNFα), has been considered a prominent drug target, particularly in inflammatory disorders and malignancies. Despite efforts, no selective ADAM17 inhibitor has yet entered the pharmaceutical market due to systemic toxicities. Drug repurposing seems to be a promising strategy for developing new ADAM17-based therapies. The current study aimed to identify FDA-approved drugs with the potential to be repurposed as selective ADAM17 inhibitors via a ligand- and structure-based approach.
Method: In this study, we screened the e-Drug3D database of FDA-approved drugs over the generated pharmacophore hypothesis AAADHR. All compounds that matched the hypothesis were screened over the ADAM17 crystallographic structure via a three-step Glide docking protocol. To identify ADAM17 selective inhibitors against matrix metalloproteinases (MMPs), we also docked the virtual screening hit compounds over the structures of MMP-1, MMP-2, MMP-3, MMP-7, MMP-8, MMP-9, MMP-10, MMP-12, MMP-13, MMP-14, and MMP-16. Among potential selective inhibitors, we selected the compound that formed the most stable complex with ADAM17 based on MM-GBSA free binding energy and molecular dynamics simulation findings.
Results: With a pharmacophore fitness score of 1.336 and an XP GScore of -9.838 in the ligand- and structure-based screening process, Latanoprost was identified as a possible inhibitor of ADAM17. Based on docking results over the MMPs structures (XP GScore > -9.17), this compound can selectively inhibit ADAM17. MM-GBSA ΔGBind of -38.30 and the average protein Cα RMSD value of 1.52±0.18 during the 120 ns MD simulation period confirm the stability of the Latanoprost-ADAM17 complex.
Conclusion: According to the results of this study and the Latanoprost pharmacokinetic properties, this drug has the potential to be a repurposing drug candidate, especially for inflammatory skin conditions.

Key Words: *ADAM17; Metalloproteinases; TNFα; Drug repurposing; Inflammation*

Design, and molecular docking studies of novel benzosuberone derivatives as potential anticancer agents

Omid Ebrahimpour[1], Shohreh Mohebbi[1], Hafezeh Salehabadi[1]*

[1] *Department of Medicinal Chemistry. School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, Iran*

E-mail address (corresponding author): omidebrahimpour1991@gmail.com

Abstract

Introduction: Aminopeptidase N (APN) is one of the most critical metalloenzymes in the body belonging to the M family, containing zinc atoms (Zn2 +). This enzyme attaches to the n-terminus of amino acids and destroys amino acids and proteins. APN plays a vital role in angiogenesis, and tumor metastasis. The expression level of this enzyme is increased in most cancers of the stomach, pancreas, prostate, and kidney and causes cancer cells to multiply, so that APN can be a helpful marker in cancer diagnosis. As a result, inhibition of this enzyme can be very effective in inhibiting and controlling cancer. For this reason, its inhibitory synthesis has become attractive [2, 3]. Tetralones such as benzosuberon derivatives, as a potent APN non-peptide inhibitor, competitively and selectively inhibit the enzyme aminopeptidase N. The application of bioinformatics could help to rational drug design and identification of new potent lead compounds. In this project, APN inhibitory activity of 30 Benzosuberon derivatives was investigated by molecular docking studies, and the best compounds were selected to evaluate the enzymatic assay.
Methods: The Crystal structure of APN, with the PDB ID of 4FKK and resolution of 2.60 Aº was achieved from Protein Data Bank (www.rcsb.org). After validation, all derivatives were investigated by docking studies. Finally, compounds with best docking score have been selected for synthesis and evaluation by enzymatic assay.
Results and Discussion: Here molecular docking studies were used to identify new compounds with inhibitory effects on APN. The binding energy and main interactions between the benzosuberon derivatives and APN binding pocket were precisely investigated in detail. Compounds with appropriate docking score selected for enzymatic assay. Selected compounds can be considered as a proper candidate in order to develop new APN inhibitors.

Key Words: *Inhibitor, Molecular Docking, Aminopeptidase N, Benzosuberone, Cancer*

Computational Design of a Candidate Vaccine to Increase Immune Responses against Omicron SARS-CoV-2 variant

Shiva Mohammadi[1]*, Babak Elyasi Far[2], Nahid Rezaei[3], Motahare Mahi Birjand[4]

[1] *Department of Medical Biotechnology, School of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran*
[2] *Department of Physiology and Pharmacology, School of Medicine, Dezful University of Medical Sciences, Dezful, Iran*
[3] *Department of Medical Immunology, School of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran*
[4] *School of Pharmacy, Infectious Diseases Research Center, Birjand University of Medical Sciences, Birjand, Iran*

E-mail address (corresponding author): shivamohamadi66@yahoo.com

Abstract

Background: SARS-CoV-2 has been emerging in the form of different variants since its first emergence in 2019. A new Variant named the Omicron variant was reported. This variant has many mutations in the Spike, membrane, envelope, and nucleocapsid proteins. Immuno-informatics tools are cost-effective methods to accelerate the design of a suitable candidate vaccine against Omicron variant. The employment of vaccines has been demonstrated to be a promising immunization approach against viruses due to the induction of long-term protective immunity.
Methods: In the study, a computational approach was conducted to design a vaccine composed of cytotoxic T lymphocyte and helper T lymphocyte epitopes of Spike and Envelope proteins. The potential viral peptides as the candidate vaccine were screened regarding convenient features like hydrophilicity, flexibility, and antigenicity. The final assembled construct was fused with the assistance of suitable linkers and cloned in a pET28a expression vector for the production of the vaccine in a bacterial host.
Results: After validation of the final construct in terms of its efficacy, stability, and exposure ability, molecular docking analysis was carried out to reveal its interaction with toll-like receptor 4. The molecular simulations by iMODS software confirmed the stability of the binding interface. Additionally, the computational cloning of the assembled vaccine in pET28a plasmid showed the possibility of producing a vaccine construct in E. coli.
Conclusion: The computational analysis showed that this construct could be a potent vaccine candidate against Omicron SARS-CoV-2 variant once its effectiveness is verified by experimental studies.

Key Words: *Immuno-informatics, Vaccine, Envelope protein, Omicron SARS-CoV-2 variant, Spike protein.*

In silico Design of a Candidate Vaccine to Hasten Immune Responses against Group B Streptococcus Bacteria

Shiva Mohammadi[1]*, Babak Elyasi Far[2]

[1] *Department of Medical Biotechnology, School of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran*
[2] *Department of Physiology and Pharmacology, School of Medicine, Dezful University of Medical Sciences, Dezful, Iran*

E-mail address (corresponding author): shivamohamadi66@yahoo.com

Abstract

Background: Streptococcus agalactiae (Group B Streptococcus) is a pathogen, that asymptomatically colonizes the genitourinary tract of healthy adults. Nevertheless, Group B Streptococcus carriage in pregnant women can lead to several health issues in newborns causing life-threatening infections, for example, pneumonia, sepsis, or meningitis. Immuno-informatics servers are cost-effective methods to boost the design of a skilled candidate vaccine against Streptococcus agalactiae. The use of vaccines has been confirmed to be a hopeful immunization approach against bacteria due to the induction of long-term protective immunity.
Methods: In the study, a complete in silico approach was conducted to design a vaccine composed of cytotoxic T lymphocyte and helper T lymphocyte epitopes of Sip (surface immunogenic protein) and Alp proteins. The potential bacterial peptides as the candidate vaccine were screened regarding convenient features like hydrophilicity, flexibility, and antigenicity. The final construct was fused with the assistance of suitable linkers and cloned in a pET28a vector for the production of the vaccine in a bacterial host.
Results: After validation of the vaccine construct in terms of its efficacy, stability, and exposure ability, molecular docking analysis was carried out to show its interaction with toll-like receptor 4. The molecular simulations by iMODS server confirmed the stability of the binding interface. Additionally, the in silico cloning of the assembled vaccine in pET28a showed the possibility of producing a vaccine construct in E. coli.
Conclusion: The in silico analysis indicated that the construct could be a potent candidate vaccine against Group B Streptococcus once its effectiveness is verified by experimental studies.

Key Words: *Alp proteins, Candidate Vaccine, Group B Streptococcus, In silico, Sip protein*

Quantitative analysis for identifying important disease-related microRNAs

Alborz Esfandyari[1]*, Hossein Hajnoruzi[1]*, Zeinab Zali[1]

[1] *Isfahan University of Technology*

E-mail address (corresponding author): alborz.esfandyari@ec.iut.ac.ir

Abstract

Irregularities in gene expression, and therefore, overproduction or underproduction of essential proteins, are the key reasons for the development of various diseases. Lately, one of the treatment methods that researchers are exploring is to investigate the regulatory mechanisms of gene expression. MicroRNAs are one of these regulatory factors that bind to the genes in specific regions called the binding sites and prevent the expression of proteins. As a result, if microRNAs fail to execute their function, protein production will be out of the normal range.

Studying the microRNAs involved in a disease can potentially lead to finding a cure for that disease; however, finding the relevant microRNAs is very complex since one microRNA can target multiple genes, and on the other hand, one particular gene can be regulated by multiple microRNAs. Therefore, identifying disease-related microRNAs requires extensive experiments and takes a deal of time and resources. The purpose of this study is to facilitate the work of researchers by ranking the disease-related microRNAs based on their importance in disease. The approach of this study consists of 1) mining the well-established disease-gene and microRNA-target association databases; 2) counting the number of disease-associated genes that are targeted by each microRNA; and 3) calculating the proportion of disease-associated microRNA targets relative to all of the predicted targets of each microRNA.

Finally, according to the timeliness, the proposed approach of this study has been applied to three diseases, coronavirus, covid-19, and B.1.1.7. This approach resulted in the identification of hsa-miR-548c-3p and hsa-miR-551a as the most important microRNA associated with coronavirus, hsa-miR-652-3p and hsa-miR-548c-3p associated with covid-19 and hsa-miR-423-5p and hsa-miR-4446-3p associated with B.1.1.7.

Key Words: *MicroRNAs; Gene expression; Protein; Counting; Coronavirus.*

SARS-CoV-2 spike protein targeting by Saffron extracts: in silico molecular docking

Amirreza Farajnezhadi[1]*, Fatemeh Eshari[1], Mehran Habibi-Rezaei[1]

[1] *Protein Biotechnology Research Lab (PBRL), School of Biology, University Of Tehran, Tehran, Iran*

E-mail address (corresponding author): ar.farajnezhadi@ut.ac.ir

Abstract

The widespread pandemic of COVID-19 is one of the challenges that influence all humans from 2019. The viral agent of this infectious disease targets the angiotensin-converting enzyme (ACE2) receptor of the host cell membrane by its surface spike (S) glycoprotein. This protein is composed of two subunits, S1 and S2, which the first subunit comprises receptor-binding domain (RBD) that is responsible for binding to ACE2 receptor ,. Accordingly, the spike protein was targeted with molecules derived from saffron, to disable the binding of the RBD region to the ACE2 receptor. Saffron is the most valuable medicinal food product, also it is essential in Iran's agricultural economy. Dried stigmas of the plant Crocus sativus (Iridaceae) are processed to produce saffron which has some other importance in pharmaceutics, cosmetics, perfumery, and textile dye-producing industries. Recently, the importance of this plant increased in the world due to the various reports about its pharmacological activity, for instance, Antinociceptive and anti-inflammatory activities. In this study, 5 extracts of saffron including, Crocin, Crocetin, Safranal, Kaempferol, Picrocrocin were examined in silico for their binding affinity with RBD by AutoDock Vina software. The results revealed that Crocin had the highest potential to block the RBD (binding affinity = -32.64 kJ/mol). Also, some papers reported the protease blocking activity of Crocin. Immediately after Crocin, Kaempferol had a better capability to block the RBD with a binding affinity of -28 kJ/mol. Moreover, the lowest binding affinity was -20.92 kJ/mol for Safranal. In conclusion, these results suggested that Saffron and its extracts are potentially suitable candidates for controlling the COVID-19.

Key Words: *Covid-19; Receptor-binding domain; Saffron; Molecular Docking; Crocin*

Immunoinformatics approach for designing poly-epitope based vaccines

Elham Ghafouri[1]*, Dr. Hossein Khanahmad[1]

[1] *Department of Genetics and Molecular Biology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran*

E-mail address (corresponding author): elhamghafouri412@gmail.com

Abstract

The primary purpose within all the immunizations is the ability of the vaccine to induce a robust immune response in a faster mode than the pathogen itself. Designing the best vaccine in the fastest time is very important. Reverse vaccinology employs bioinformatics to find proteins and epitopes that indicate antigenicity. Some of the advantages of this approach are reduced cost and time of vaccine development and facilitating antigen identification selection. This method does not require the cultivation of risky microorganisms, allowing to study non-cultivate microorganisms and screening of all microorganism proteins from genome sequences, making them biologically safe. Additionally, their selectivity allows accurate activation of immune responses.

Methods: We will highlight the immune-informatics and computational approach technique known as reverse vaccinology to predict the potential epitope for designing poly epitope vaccines, including the T cell and B cell epitopes. Each epitope's antigenicity, toxicity, and allergenicity are checked; then, the selected epitopes are joined together via different linkers.

Various physical-chemical properties of the poly-epitopes vaccine, including molecular weight (MW), isoelectric point (IP), stability index, half-life in vitro and in vivo, aliphatic index, and grand average of hydropathicity (GRAVY), are estimated in silico. The threedimensional structure of the mentioned vaccine will be subjected to molecular docking studies with MHC-I and MHC-II molecules.

Results: In the final, we selected the non-toxic and non-allergic epitopes with a high antigenicity score. These epitopes are linked together and indicate suitable properties.

Conclusion: The proposed vaccine needs to be validated clinically to ensure its safety and immunogenic profile.

Key Words: *Keywords: Reverse vaccinology; Immuno-informatics; Poly epitopes; T-and B-cell epitopes; Vaccine design*

Designing a Bioinformatics Model to Select Antimicrobial Peptides against Helicobacter pylori Infection

Saba Ghasemi[1], Karim Heydari[2]*, Maryam Homayon[3]

[1] *Department of Biology, Zarghan Azad University, Zarghan, Iran*
[2] *Department of Computer Science, Zarghan Azad University, Zzarghan, Iran*
[3] *Department of Biology, Zand Institute of Higher Education, Shiraz, Iran*

E-mail address (corresponding author): homayoon.mar@gmail.com

Abstract

Helicobacter pylori is the most common chronic bacterial infection that colonizes the gastrogenic mucosa of half of the population worldwide. The treatment of H. pylori has been limited to the use of antibiotics, and due to increase in resistance to antibiotics used in the anti-H. pylori therapy, the development of an alternative treatment is one of the most common challenges for physicians. Therefore, a strategic, and effective antimicrobial peptides (AMPs) design against H. pylori is essential where the use of the most prevalent bioinformatics tools could help achieve it.

In this study, biotechnology software and AMP databases were used to design new AMPs against H. pylori. In fact, five AMP databases were used to extract the best AMP for gram-negative bacteria based on some criteria such as hydrophobicity, length, charge, and three-dimensional structure of peptides. In order to create specific AMPs to target Helicobacter pylori, a designed peptide was added to the end of AMPs. This peptide has been designed based on the tendency to bind to the predicted binding sites on one of the proteins in the outer membrane of this bacterium (HopB).

Twelve AMPs (CPF-St7, XT-4, XPF-SE3, Warnericin RK, Magainin-F3, Acipensin 6, Carnobacteriocin B2, Fallaxin, Dermaseptin-DI1, Pseudin-2, Hylin-a1, Piscidin-3) were selected considering some important factors for gram-negative antibacterial peptides. The designed peptide was added to the N or C-terminal of AMP considering the AMP properties. Bioinformatics methods can be the first step in achieving new and efficient AMPs.

The results of this study can be promising to introduce new treatment strategies against H. pylori. However, to validate the therapeutic effects of our AMPs design, in vitro and in vivo studies are required.

Key Words: *Antimicrobial peptide; Bioinformatics; Helicobacter pylori; AMP databases.*

Target prediction for the AKT inhibitors in the inflammatory pathway of colorectal cancer using similarity-based methods from Chembl25 database

Alireza Ghavami[1], Andrease Bender[2], Mostafa Zakariazadeh[3], Samira Shafiee[1], Somaieh Soltani[1]*

[1] Department of Medicinal Chemistry, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
[2] Department of Chemistry, Centre for Molecular Informatics, University of Cambridge, Lensfield Road, Cambridge, CB²¹EW, UK
[3] Department of Biology, Faculty of Sciences, Payame Noor University, Tehran, Iran

E-mail address (corresponding author): alireza.ghavami98@gmail.com

Abstract

The pathway of Interleukin-6 (IL6) as one of the inflammatory cytokines that its contribution to colorectal cancer pathology is known, was investigated to extract the related target list. AKT is a serine/threonine kinase protein which can stimulate a large different extracellular factor. Synergestic effect of simultaneous inhibition AKT and some targets in the pathway of IL6 have been reported. In this study we used a similarity-based method for the target prediction for AKT inhibitors from IL6 pathway. The target list was extracted from the KEGG database and Chembl25 was used to extract the inhibitors. The Morgan fingerprints was calculated using RDKit that was implemented in a Knime platform. Tanimoto similarity index was calculated and different thresholds were used to indicate similarity. The predicted targets for AKT inhibitors were GSK3β, PDKI1, PIM1, mTOR, JAK1, TIE-2 and STAT3. The literature and databases survey indicated the available evidences for the inhibition of the predicted targets. As the synergestic effect of AKT-STAT3 has been reported earlier. The pyrimidine compound with the predicted inhibition of STAT3 was further evaluated using molecular docking. Docking simulation was done by AutoDock software. The results indicated that the appropriate interaction between AKT and selected compound. Binding energy value is -8.34 Kcal/mol between them.

Key Words: *Interleukin-6; Colorectal Cancer; Inflammation; Database; Molecular Docking*

Target prediction for GSK3β inhibitors in colorectal cancer pathway based on Morgan fingerprints

Zahra Jannat Doust[1], Salva Golgoun[1]*, Gerhard Wolber[2], Mostafa Zakariazadeh[3], Samira Shafiee[1], Somaieh Soltani[1]

[1] *Department of Medicinal Chemistry, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Pharmaceutical and Medicinal Chemistry (Computer-Aided Drug Design), Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany*
[3] *Department of Biology, Faculty of Sciences, Payame Noor University, Tehran, Iran*

E-mail address (corresponding author): salva.golgoun@gmail.com

Abstract

Colorectal cancer is associated with inflammation. A systematic data-mining study was performed to investigate colorectal cancer related targets for inhibitors of GSK3β. Morgan fingerprints were calculated for all studied compounds. Tanimoto index was calculated using fingerprints and compounds with Tanimoto >0.7 was indicated as similar. The developed procedure for target prediction was run using knime platform. RDKit and Chemaxon were used to fingerprint calculation. The inhibitors were extracted from Chembl25. According to the results some GSK3β inhibitors can be an inhibitor for AKT, IKKA, PKB and CDK1 as well. The reliability of the results was evaluated using the available evidences for the inhibition of predicted targets from literature and the survey showed that there is published evidences for the inhibition of IKKα, PKB and CDK1. The mode of interaction of the compounds with the highest and lowest potency were evaluated using molecular docking methods and the results indicated good correlation with experimental results. In conclusion the developed method is reliable for the target prediction.

Key Words: *Colorectal cancer; inflammation; GSK3β; knime platform; molecular docking*

Meta-Analysis of the Microarray Expression Data and Molecular Docking of the Secondary Metabolites of the Medicinal Plants (licorice, Echinacea Angustifolia, Ginseng,) to Determine Their Active Ingredients Against SARS-CoV-2

Afifeh Hajipour, Masoud Tohidfar*, Yousef Saeedi Honar

E-mail address (corresponding author): hajipourafifeh77@gmail.com

Abstract

Coronaviruses are viruses that often cause cold-like symptoms in the respiratory system. A number of them, such as SARS and MERS in recent years and COVID-19 since January 2019, have killed thousands of people and caused many pandemics. Medicinal plants are used to treat a variety of ailments, including infectious diseases. The aim of this study was to determine the genes of effective antiviral compounds in the medicinal plants of licorice, Echinacea angustifolia, ginseng, and chicory coffee against such symptoms as pulmonary embolism caused by coronavirus. In this research, first, meta-analysis of the secondary metabolites of the mentioned plants on the human species was done and their key proteins were identified. Then, the interaction of the proteins was assessed in the STRING database and the Cytoscape software and PANTHER database were applied for obtaining the candidate genes and studying their ontology, respectively. The expressions of the genes were measured by drawing HeatMap. The meta-analysis resulted in identifying 14 genes related to the immune system and blood coagulation factors that would cause blood clotting by converting fibrinogen to fibrin and activating platelets. Then, the extracts of the 4 mentioned plants, whose antiviral and antimicrobial effects had been proven, were extracted from the Pubchem database for performing docking and their interaction was evaluated by using the STRING database. Afterwards, the results obtained from the STRING server were assessed in the Cytoscape software and (F10) and (F2) as receptors were extracted from the Protein Data Bank (PDB) database. Finally, the receptor-ligand docking was performed and its validation was confirmed by the ligand-protein server. According to the results, the antiviral extracts of the mentioned plants can be used as inhibitors of the two mentioned coagulation factors, which convert fibrinogen to fibrin and activate platelets, and prevent the danger of Pulmonary Embolism (PE).

Key Words: *SARS-CoV 2, microarray, molecular docking, drug design, medicinal plant*

Interaction effect of Pennyroyal essential oil compounds on Streptococcus pyogenes tyrosine phosphatase enzyme by molecular docking

Zeynab Hejazi[1], Marzieh Etehadpour[1]*, Mahboobe Bagheri[1]

[1] *Agricultural Faculty of Bardsir, Shahid Bahonar University of Kerman, Kerman, Iran*

E-mail address (corresponding author): etehadpour@uk.ac.ir

Abstract

Streptococcus pyogenes is one of the most important Gram-positive bacteria. Streptococcus pyogenes type a (GAS) is the most important group of streptococci due to the activity of pharyngitis, which annually causes about 600 million cases of streptococcal pharyngitis worldwide. Streptococcus pyogenes encodes a hydrolyzing enzyme called tyrosine phosphatase (SP - PTP), which has dual phosphatase properties for both phosphorylated serine / threonine tyrosine amino acids. SP-PTP controls the growth, division, adhesion, and invasion of bacterial cells into host cells through its phosphatase activity. In this study, first, the three-dimensional structure of Pennyroyal and protein compounds was obtained from PubChem and PDB databases, respectively. ViewerLite, Chimera 1.14, Discovery Studio, AutoDockTools-1.5.6 and AutoDock Vina softwares were used to perform molecular docking and interaction between plant compounds and enzyme. Based on the results of docking studies, all 64 studied compounds were able to occupy the active site of the enzyme, among which, the best docking results are related to the Carvol compound. In fact, this compound has the most negative binding energy level (-7.1 Kcal / mol), the highest affinity for binding to key amino acids and the active site of the enzyme.

Key Words: *Key words: AutoDock Vina; Molecular docking; Screening; Tyrosine phosphatase.*

Molecular docking of sixteen common phenolic compounds of medicinal plants in inhibition of Streptococcus pyogenes tyrosine phosphatase

Zeynab Hejazi[1], Marzieh Etehadpour[1]*, Mahboobe Bagheri[1]

[1] *Agricultural Faculty of Bardsir, Shahid Bahonar University of Kerman, Kerman, Iran*

E-mail address (corresponding author): etehadpour@uk.ac.ir

Abstract

Streptococcus pyogenes is the most common bacterial cause of pharyngitis. Patients with sore throat should be treated promptly with appropriate antibiotics. Analyzing the effectiveness of a wide range of plant compounds in order to obtain safe and new antibiotics through the laboratory is time consuming and costly. Therefore, in recent years, research based on bioinformatics methods due to time and cost savings have been considered as complementary methods in the design of new drugs, including antibiotic compounds. In this regard, the aim of this study was in silico investigation of the inhibitory effect of common phenolic compounds in medicinal plants on the Streptococcus pyogenes tyrosine phosphatase. First, the structure of plant compounds and proteins were obtained from PubChem and PDB databases, respectively. Finally, the ability of plant compounds to inhibit the enzyme in ViewerLite, Chimera 1.14, Discovery Studio, AutoDockTools-1.5.6 and AutoDock Vina were investigated. The results showed that most of the studied compounds have good interactions with the studied protein. The results also showed that Quercetin and Umbelliferon have the strongest interactions with the studied protein. In fact, these compounds with the most negative binding energy levels (-7.6 and -7.4 Kcal / mol, respectively) have a greater tendency to bind to the key amino acids of the active site of the studied protein. Therefore, these compounds can be considered as good candidates for laboratory and in vivo study of anti-Streptococcus pyogenes activity.

Key Words: *Keywords: Antibacterial, Molecular docking, Bioinformatics, Medicinal plants*

IN-SILICO STUDY OF HERBAL COMPOUNDS AS NOVEL MAO-B INHIBITOR FOR PARKINSON'S DISEASE TREATMENT

Afsaneh Hoseini[1], Mohammad Reza Zamani[2], A.m Mansoreyan[1], Fatemeh hedareyan[1], Ali shamsi[3]

[1] *Nor Danesh Institute, Isfahan, Iran*
[2]*National Institute for Genetic Engineering and Biotechnology Tehran Iran*
[3]*Fan & Danesh Institute Saveh Iran*

E-mail address (corresponding author): hoseniaf@chmail.ir

Abstract

Parkinson disease (PD) is the second most common neurodegenerative disorder.Monoamine oxidase B (MAOB) is expressed in the mitochondrial membrane and has a key role in degrading various neurologically active amines such as benzylamine, phenethylamine and dopamine with the help of Flavin adenine dinucleotide (FAD) cofactor and have important functions in the metabolism of neuroactive and vasoactive amines in the central nervous system. 3D structure of MAOB were used from PDB database and are available with PDBID (2V5Z). Some novel herbal bioactive compounds have been reported to be useful in neurodegerative disorders, depression and have antioxidative property. Docking was successfully performed for MAO B protein with 35 herbal compounds and active site coordination was found to be similar as the template residues involved in binding with experimentally used inhibitors. The binding affinities computed using autodock-vina are shown to be non-reliable. To be effective as a drug, a potent molecule must reach its target in the body in sufficient concentration, and stay there in a bioactive form long enough for the expected biologic events to occur. Drug development involves assessment of absorption, distribution, metabolism and excretion (ADME) increasingly earlier in the discovery process, at a stage when considered compounds are numerous but access to the physical samples is limited. According to the above, it was found that herbal compounds (coumarin, Vanilic acid, Cinnamic acid, Isoferulic acid) can have a significant effect in the treatment of Parkinson's disease.

Key Words: *Parkinson disease, modeling, Monoamine oxidase B inhibitor, docking, Herbal compounds*

Structure-based virtual screening of FDA-approved compounds to find potential inhibitors for human monocarboxylate transporter 1 in an open-inward conformation

Fatemeh Jabari[1], Hamid Mahdiuni[1], Soraya Sajadimajd[1]

[1] *Bioinformatics Lab., Department of Biology, School of Sciences, Razi University, Kermanshah, Iran*

E-mail address (corresponding author): hmahdiuni@razi.ac.ir

Abstract

Owing to the high rate of glycolysis, increased access to glucose is necessary for cancer cells to survive and proliferation. To adapt to the changes in the tumor microenvironment, two groups of cells develop: glycolytic and oxidative cancer cells. Altered metabolism and increased glucose consumption in the glycolytic cells lead to increased intracellular lactate levels, which have to be removed from those cells and absorbed by the oxidative ones. Lactate is transported by a number of monocarboxylate transporters (MCT), which has been identified as a family of 14 members, ranging from MCT1 to MCT14. The monocarboxylate transporter 1 (MCT1) catalyzes the movement of monocarboxylates such as lactate and pyruvate across the plasma membrane of mammalian cells. The overexpression of MCT1 has been reported in several cancers including lung, colon, brain, skin and lymphoma cancers. Therefore, inhibition of MCT1 holds great promise for cancer treatment. In the present study, we screened the FDA-approved compounds of the Zinc Database using structure-based virtual screening methods to find effective compounds for inhibition of the MCT1 transporter in an open-inward conformation. To this end, the resolved structure of MCT1 was downloaded from RCSB (PDB ID: 7CKO). To prepare the chemical library, FAF-drugs4 was applied on FDA-approved subset of ZINC to remove compounds having more than 10 rotatable bonds and PAINS. Next, the clean library was energy minimized using Mmff94 force field. In follow, we docked the resultant library against MCT1 using a funnel-like pattern by Autodock Vina, Molegro Virtual Docker, and DOCK6. The compounds obtained from this step were studied for further investigation.

Key Words: *Monocarboxylate Transporter1, Virtual screening, inhibitors, Cancer*

Investigating the protection effects of pharmaceutical polysaccharides on Interferon beta-1a against heat stress: A molecular dynamic simulation

Kasra Karami[1], Pegah Yaghoubi[1], Mohsen Shahlaei[2]*, Sajad Moradi[2]*, Mohamadjavad Naseri[1]

[1] *Pharmaceutical Sciences Research Center, Faculty of Pharmacy, Kermanshah University of Medical Sciences, Kermanshah, Iran*
[2] *Nano Drug Delivery Research Center, Kermanshah University of Medical Sciences, Iran*

E-mail address (corresponding author): kasra1998karami@gmail.com

Abstract

Many proteins, including therapeutic proteins, are structurally unstable in the face of multiple stresses, such as rising temperature. The drug delivery system through polymers increases the therapeutic effect and reduces side effects along with reaching the drug to targeted site at the specified time and the drug is also protected. In this study, the effect of five carbohydrate polymers as alginate, chitosan, heparin, hyaluronic acid and pectin on the stability of interferon beta-1a protein at 300 and 363 K was investigated using molecular dynamics simulation methods.

The two-dimensional structure of polymers was mapped and optimized with ChemOffice software, the contents of the resulting PDB file were given to the PRODRG server in order to preparing the topology parameters of the polymers. The protein coordinate file was also obtained from the Brookhaven Protein Database (http://www.rcsb.org/pdb) in PDB format. The all simulations were performed using Gromacs version 5.1.1 software and up to 100 ns at two temperatures of 300 and 363 K. Then the results were analyzed using the different computational tools.

Using molecular dynamics simulation, we found that increasing temperature causes protein instability and among the five polymers investigated, hyaluronic acid and heparin polymers were able to protect the protein at high temperature.

Finally, this study suggests that the preparation of interferon beta-1a formulations from hyaluronic acid and heparin polymers can be effective in preventing high temperature structure degradation.

Key Words: *molecular dynamics simulation, Gromacs, interferon beta-1a, polysaccharide polymers, protein-polymer interaction*

Spike/ membrane multi-epitope based protein as a future candidate vaccine against sars-cov-2

Armity Kiumarsy[1], Mohammad Reza Shafaati[2]*, Masoud Ghorbani[3], Shaghayegh Yazdani[4], Vahid Zarghami[5], Farzaneh Kazemianfar[6]

[1] Department of microbiology, faculty of advanced science & technology. Tehran medical sciences, Islamic Azad University, Tehran, Iran
[2] Department of cellular & molecular biology, Faculty of Basic Sciences, Hamedan Branch, Islamic Azad University, Hamadan, Iran
[3] Pasteur Institute of Iran, Research and production Complex, Department of Research and Development, Karaj, Iran
[4] Department of microbiology, faculty of advanced science & technology. Tehran medical sciences, Islamic Azad University, Tehran, Iran
[5] Institute of Nanoscience and Nanotechnology, Sharif University of Technology, Tehran, Iran
[6] Department of Infectious Diseases and Tropical Medicine, Shahid Fayaz-Bakhsh Hospital, Iran University of Medical Sciences, Tehran, Iran

E-mail address (corresponding author): shafaati@iauh.ac.ir

Abstract

Since humans are not immune to SARS-COV-2, it is urgent to produce vaccines to reduce epidemics and prevent the reappearance of SARS-CoV-2 and possible variants of this virus. This study aimed to design a multivalent recombinant protein based on SARS-CoV-2 virus spike and Membrane epitopes associated with HSP70 of Mycobacterium tuberculosis bacteria to evaluate the induced immune response to enhance the immune response induced by the recombinant SARS-CoV-2 vaccine. Initially, to select the appropriate strain for extraction of membrane and viral spike, phylogenetic analysis was performed. Then, based on the frequency of HLA in the population of Iran and the Middle East, three HLA-A*02:01 and HLA-C*04:01 and HLA-DRB1*01:01 were selected to predict T cell epitopes. These epitopes were divided into two groups of GP100 and GP200 and were connected to HLA-A and HLA-C, and from each group, protein candidates were selected. Then, two selected proteins, and whole protein HSP70 tuberculosis were attached by suitable linkers and analyzed immunogenicity, antigenicity, the allergenicity of these recombinant proteins, and one protein was selected as a final protein candidate. In our laboratory, this triple cassette was cloned into a clone donor and transected into the expression host of Bacillus subtilis. Results: The increase in antibody titer against SARS-COV-2 in mice immunized with the recombinant vaccine was significant. Discussion and Conclusion: This achievement shows that an increase in humoral and cellular immune response from the recombinant vaccine of SARS-COV-2 based on the epitopes can be used as an effective candidate vaccine against the possible epidemics caused by this Virus and possible variants of this virus.

Key Words: *SARS-CoV-2; Immunogenicity; Antigenicity; Allergenicity; Recombinant protein design; Candidate Vaccine;*

Identification of effective natural compounds for reducing α-Synuclein aggregation involved in Parkinson's disease using Molecular docking on the ZINC database

Marzyeh golestanfard[1], Milad Lagzian[1]*, Maliheh Mohammadi[1], Arezou Ghahghaei[1]

[1] *Department of Biology, Faculty of Science, University of Sistan and Baluchestan, Zahedan, Iran*

E-mail address (corresponding author): m.lagzian@science.usb.ac.ir

Abstract

Parkinson's disease is one of the most common neurodegenerative disorders in the elderly age. One of the mechanisms involved in neurodegeneration is related to the aggregation of the presynaptic protein α-synuclein. Despite numerous studies and efforts, discovering compounds that can serve as promising inhibitors and delay α-synuclein aggregation remains an important challenge. This study aims to identify potential compounds that inhibit α-synuclein accumulation using structure-based virtual screening in ZINC's natural compounds Database. Initially, the structure of human α-synuclein covering the full sequence of the protein (PDB ID:1XQ8) was retrieved and prepared from the Protein Data Bank. The natural compound library from the ZINC database that contains more than 340,000 compounds was prepared. The potential α-synuclein binding pocket contained was identified by the literature review. It contained four key residues, including L38, V40, K43, and K45, used to create docking constrain. Molecular docking studies at three different precision levels, including HTVS (High throughput virtual screening), SP (standard precision), and XP (extra precision), were performed by Schrodinger 2021-2. Subsequently, the MM-GBSA method estimated the binding free energy between top-ranked ligands and the protein. The analysis of the molecular docking calculation showed that the compounds ZINC31169817 ()1S(-1,5-Anhydro-1-)2,6-dihydroxy-4-methylphenyl(-D-glucitol), ZINC12405202()1R,2S,3R,4S,5R,6S(-6-Methoxy-1,2,3,4,5-cyclohexanepentol), Compound825 (}3-)Hexopyranosyloxy(-2-[(27)-2-penten-1-yl]cyclopentyl}acetic acid) with docking score -7.998, -7.711, -7.225 Kcal/mol and the binding free energy -29.69, -26.57, -28.56 have the highest affinity for interaction to α-synuclein monomer. The identified candidates need to verify by molecular dynamics simulation before any in-invitro study.

Key Words: *Parkinson's disease, α-synuclein, aggregation, Molecular docking, ZINC Natural Library*

Applying virtual screening approach to find out effective agonist against RXR (retinoid X receptor) as promising candidates to treat Alzheimer disease

Alireza rahimpour[1], Behzad Shahbazi[2], Ladan Mafakher[3]

[1] *Islamic Azad University of science and research branch Tehran, Iran*
[2] *Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran*
[3] *Thalassemia & Hemoglobinopathy Research center, Health research institute, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran*

E-mail address (corresponding author): ladan.mafakher@gmail.com

Abstract

Introduction: Alzheimer's is neurodegenerative disease that causes accumulation of extensive amyloid-beta in the brain. The prevalence Alzheimer of patients increased exponentially so that in United states estimates 13.7 million people will be affected by Alzheimer by 2080. Alzheimer's patients can be classified into two different types including Sporadic or late-onset AD (LOAD) with prevalence of 95% and early-onset AD (EOAD). In the LOAD, one of the predominate risk factor gene is APOE which dysregulation of this gene increase accumulation of amyloid-beta. So regulation the expression of APOE by RXR receptor as transcriptional regulator of APOE by designing Agonist against RXR receptor could be control clearance of amyloid-beta and decrease symptoms of Alzheimer.
Material and Methods: Two herbal Database (TCM and IBscreen) were selected as ligand libraries. Pharmit webserver was chosen to check properties of small molecules based on Lipinski's rules. The selected small molecules which pass Lipinski's rules were applied to dock by SMINA package against binding site of RXR receptor. The small molecules that effectively interact with RXR binding site nominated for cytotoxic and blood barrier properties analysis by OSIRIS Data Warrior and BBB predictor. Small molecules that could pass blood-brain barrier and have not mutagenesis, tumorigenic, reproductive effect properties were entered in Molecular Dynamic simulation and efficacy of their interaction were checked by MM/PBSA package.
Results and Discussion: Among total compounds of TCM and IBscreen, around thirty compounds pass Lipinski's rules and effectively interact with binding site of RXR receptor. Cytotoxic and blood brain barrier analysis revealed that all these compounds have no mutagenesis, tumorigenic, reproductive effect and able to pass blood brain barrier. Molecular dynamic simulation discovered that five compounds able to interact efficiently during simulation. MM/PBSA analysis figured out that these compounds effectively bind to binding site of RXR receptor.

Key Words: *Alzheimer disease, virtual screening, retinoid X receptor, Molecular Dynamics simulation, Molecular docking*

The development of natural dietary supplements and medicines utilizing bioinformatics as a tool to identify the biosynthesis processes of bioactive compounds of edible and medicinal mushrooms

Mahnaz Mehrabi[1´2], Nasser Poursaeid[2´3]

[1] *PharmD, University Pécs Medical, Pécs, Hungary.*
[2] *Department of Pharmaceutical Biotechnology, MahPharmed Pharmaceutical Company, Isfahan, Iran*
[3] *Technology Incubator Center of Medicinal Plants, Islamic Azad University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): drmahnazmehrabi@yahoo.fr

Abstract

There are a wide range of bioactive compounds in edible and medicinal mushrooms including polysaccharides (especially β-glucans), glycoproteins, peptides, terpenoids, sterols, and phenolic compounds. These compounds help prevent, treat, and recover from illness. Identification of factors affecting biosynthesis is very important for the production of bioactive compounds on industrial scales and pharmaceutical systems. Therefore, bioinformatics methods can be used to analysis of the cellular and molecular structure of enzymes effective in biosynthesis processes and increase the production of bioactive compounds. The results of studies show that mineral ions such as zinc, copper, calcium, iron, and selenium can enhance the expression of genes that lead to the production of more bioactive compounds of edible and medicinal mushrooms, such as polysaccharides, triterpenoids and ganodric acids. The use of these findings could improve the biosynthesis of bioactive compounds and the performance of pharmaceutical biotechnology systems for the production of natural superfoods, dietary supplements and medicines.

Key Words: *Edible-medicinal mushrooms, Bioactive compounds, Biosynthesis, Dietary supplements, Natural medicines*

Bioinformatic Selection and Evaluation of Novel Drug Candidates to Inhibit Notch Signaling Pathway Based on Drug Repositioning Approach

Mina Mirian[1]*, Abbas Jafarian Dehkordi[2], Fatemeh Saffari[1], Amirali Hariri[1]

[1] Department of Pharmaceutical Biotechnology, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran
[2] Department of Pharmacology, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran

E-mail address (corresponding author): mina6m@yahoo.com

Abstract

The notch signaling pathway has a vital role in cell division, proliferation and cell survival among other pathways. There are some evidences that show inhibition of this pathway could overcome drug resistance in chemotherapy. Several attempts have been performed to block this pathway independently and concurrently with chemotherapy in cancer therapies. Drug repositioning -using approved drugs for new purposes- has been interested as a new strategy for drug development. Here, some approved drugs were selected with respect to Notch pathway and their cytotoxic effects were evaluated with drug repositioning approach. KEGG and Reactome databases were used to identify the critical proteins in Notch signaling pathway. Then, molecular docking has performed on the target protein using Auto dock software. To evaluate the accuracy of docking results, A549 cells were treated with selected drugs, and cell survival was determined using MTT assay and flowcytometry analysis. Moreover, A549 cells were simultaneously treated with doxycycline and ibuprofen to evaluate synergism effects.

ADAM17 protein was selected as a critical protein for blocking the Notch pathway, based on intracellular network findings. Doxycycline, tetracycline, ibuprofen, diclofenac and meloxicam were selected by docking. For each drug, IC50 was determined against the A549 cell line. Doxycycline and ibuprofen showed a synergistic cytotoxic effect on A549 cells when they were simultaneously used. The flowcytometry results showed doxycycline and ibuprofen can induce cell death using apoptosis mechanisms, especially in the co-treatment conditions.

Taken together, the determination of ADAM17 relation with other proteins and transformation in cancers has illustrated the importance of this protein (4). These findings showed the cytotoxic effects of doxycycline and ibuprofen on A549 cell lines, especially when they were simultaneously applied. These findings suggest that selected drugs could be utilized for lung cancer chemotherapy, after more investigation.

Key Words: *Drug Repositioning, Lung Cancer, Notch Signaling Pathway*

Molecular docking and spectroscopic studies of sulfonamide-based imines with human serum albumin

Baran Mohamadi Moghadam[1], Aria Tajalli[1], Ahmad Amiri[1]*

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): baran2181@gmail.com

Abstract

Human serum albumin contains about 60% of the total plasma protein and commonly used to study drug-protein interactions. HSA shows high affinity to different drugs, nutrients, metal ions, and their metabolites. Based on materials' affinity to HSA, their absorption, distribution, metabolism, and toxicity could be changed in vivo and then affect their pharmacokinetics, pharmacodynamics, and toxicity. Sulfonamides are an important class of antibiotics, which have been widely used as feed additives in agriculture for decades.We report here, the synthesis of two sulfonamide ligands, SulfHB and SulfTP. The chemical structure of the ligands were dtetermined by different spectroscopic techniques. The binding modes of SulfHB and SulfTP ligands with HSA was explored experimentally and computationally. The fluorescence titration experiment revealed a strong complex formation between the ligands and HSA, with a significant quenching by a blue shift of ~18 nm at the λ max of 343.The molecular docking approaches revealed that SulfHB or SulfTP could spontaneously enter into the binding sites of HSA through H-bond interactions and van der Waals forces, and that SulfHB exhibited much stronger binding affinity toward HSA than SulfTP at different temperatures (ΔG =-11.29 for SulfHB, and ΔG = -9.52 for SulfTP). The binding constants for SulfTP-HSA were determined to be 39.95 × 10 5 L.mol −1 at 298 K., and SulfHB had a greater effect on the α-helix content of HSA. Observations from molecular docking studies revealed that the hydrogen bonds might be a key factor contributing to the binding affinity of sulfa drugs and HSA. The aminoacids such as Arg117, Asn130, Asp129, Lys199, Ala 297 and Arg222, play key roles in the sulfonamide-HSA binding process via their interaction with the sulfone (O=S=O) and hydroxyl groups of the ligands. These findings might be helpful to understand the biological effects of sulfonamides in humans.

Key Words: *HSA; Sulfa Drugs; Molecular Docking; DFT-BPV86; GAUSSIAN; Sulfadiazine*

Synthesis, Enzymatic Assay and Molecular Docking Studies of Biuret Derivatives for Discovery of New Urease Inhibitors as Potential Agents Against Helicobacter Pylori Infection

Atefeh Mohammadian Berneti[1], Hafezeh Salehabadi[1]*, Neda Adibpour[1], Massoud Amanlou[2,3]

[1] *Department of Medicinal Chemistry, School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, Iran*
[2] *Department of Medicinal Chemistry, Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran, Iran*
[3] *Drug Design and Development Research Center, The Institute of Pharmaceutical Sciences (TIPS), Tehran University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): atefeh.mohammadian75@gmail.com

Abstract

Helicobacter pylori is a gram-negative spiral bacterium that causes infections in the human stomach and is able to survive in the acidic environment of the stomach with the help of the enzyme urease. By converting urea to ammonia and carbon dioxide, this enzyme modulates the pH and facilitates survive of H. pylori in the stomach because of providing a neutral environment in acidic conditions. Therefore, the virulence of H. pylori could be controlled using substances that inhibit urease activity. Due to the structure of biurets , which consists of two ureases and is similar to the urease substrate, they can be considered as potential urease inhibitors. In this project, urease inhibitory activity of 18 biuret derivatives that have already been synthesized was investigated by molecular docking studies and the best compounds were selected to evaluate the enzymatic assay.
The crystal structure of Jack bean urease (PDB code:3la4) was obtained from Protein Data Bank (www.rcsb.org). After validation, biuret derivatives were investigated by docking studies. The selected compounds are the ones that have the lowest docking score. Finally, selected compounds were evaluated by enzymatic assay.

Key Words: *Synthesis, Molecular Docking, Biuret Derivatives, Inhibitor, Urease*

Discovery of new natural inhibitors for aldosterone synthase (CYP11B2) for treating cardiovascular disease

Sina Mohammad Mehri[1], Mahshad Shahriari[1], Samad Nejad Ebrahimi[1]*

[1] *Department of phytochemistry, Medicinal Plants and Drug Research Institute, Shahid Beheshti University, Evin, Tehran, Iran*

E-mail address (corresponding author): sinamohamadmehri98@gmail.com

Abstract

Background: Aldosterone is a main mineralocorticoid steroid hormone responsible for the body's balance of sodium and potassium ions and blood pressure regulation. The aldosterone biosynthesis pathway is mediated by stimulation of Mineralocorticoids receptors and activation of the aldosterone synthase enzyme (CYP11B2). Abnormal increases in aldosterone levels cause kidney and heart diseases such as increased blood volume and stimulation of cardiac fibroblasts and cardiac hypertrophy diseases of myocardial fibrosis, and ventricular arrhythmias. More than 1000 natural compounds were screened to discover a potent natural inhibitor for aldosterone synthase enzyme (CYP11B2) which is responsible for cardiovascular disease prevention.
Methods: Virtual screening based on docking studies was performed using the Glide and Induced fit docking (IFD) program in Maestro 12.8 (Schrodinger, LLC). More than 1000 natural compounds were downloaded from the Zinc database and prepared with Ligprep software. Protein was obtained from the protein database (PDB ID: 4FDH) and prepared using protein preparation software. QSAR and Qikprop studies also were performed to evaluate the IC50 and Lipinski data for the selected compounds.
Results: The results of molecular binding in XP and IFD studies determined the Nu.1 ligand (Zinc ID 14690026) with a docking score -12.469 kcal/mol, IFD score -1030.47 kcal/mol and calculated IC50 8.223 µM as a potent inhibitory compound for CYP11B2 enzyme. The HB60 with CID: 10263082 was considered as a native ligand with a docking score of -6.59 Kcal/mol and an IFD score of -1011.44 Kcal/mol and Calculated an IC50 7.92 µM with the enzyme.
Conclusion: According to the mentioned data Ligand with Zinc ID 14690026 was selected as a potent natural inhibitor for the CYP11B2 receptor compared to the native ligand. This natural ligand helps balance blood pressure by inhibiting the aldosterone biosynthesis pathway and preventing cardiovascular diseases.

Key Words: *Aldosterone synthase, CYP11B2 enzyme, Cardiac hypertrophy, Blood pressure diseases*

Discovery of new natural inhibitors for Human epidermal growth factor receptor-2 (HER2) for treating Breast cancer disease

Sina Mohammad Mehri[1], Samad Nejad Ebrahimi[1]*

[1] *Department of phytochemistry, Medicinal Plants and Drug Research Institute, Shahid Beheshti University, Evin, Tehran, Iran*

E-mail address (corresponding author): sinamohamadmehri98@gmail.com

Abstract

Background and Aim: Breast cancer is one of the most common diseases among women that is directly related to epidermal growth factor (EGF). Human epidermal growth factor receptor-2 (HER2) is a component of the tyrosine kinase family, which is overexpressed in this disease and its exacerbation and overexpression cause the proliferation and metabolism of breast cancer cells. In recent years, attempts have been made to use small molecules such as HER2 inhibitor molecules instead of chemotherapy to inhibit the overgrowth of cancer cells and to treat breast cancer.
Methods: Virtual screening based on docking studies was performed using the Glide and Induced fit docking (IFD) program in Maestro 12.8 (Schrodinger, LLC). More than 700 natural compounds were downloaded from the Zinc database and prepared with Ligprep software. Protein was obtained from the protein database (PDB ID: 3PP0) and prepared using protein preparation software. QSAR and Qikprop studies also were performed to evaluate the IC50 and Lipinski data for the selected compounds.
Results: The results of molecular binding in XP and IFD studies determined the Nu.1 ligand (Zinc ID 253502387) with a docking score -10.207 kcal/mol, IFD score -1281.20 kcal/mol and calculated IC50 5.255 µM as a potent inhibitory compound for HER2. The NL2 with CID: 16736274 was considered as a native ligand with a docking score of -4.561 Kcal/mol and an IFD score of -1271.87 Kcal/mol and Calculated an IC50 5.273 µM with the enzyme. The obtained results show a high inhibitory activity of the selected compound compared to the native ligand.
Conclusion: According to the mentioned data Ligand with Zinc ID 253502387 was selected as a potent natural inhibitor for the HER2 compared to the native ligand. This Natural ligand helps to better inhibit the HER2 and stop cell growth and proliferation

Key Words: *Epidermal growth factor (EGF), HER2 receptor, Breast cancer, small molecules*

Novel QSAR Model for Cyclic Sulfonamide Derivatives as Potent COVID-19 Inhibitors

Nathalie Moussa[1]*, and Hoda Mando[2]

[1] *Department of Pharmaceutical chemistry and quality control of medicaments, Faculty of Pharmacy, Al-andalus University, Tartus, Syria*
[2] *Department of Pharmaceutical chemistry and quality control of medicaments, Faculty of Pharmacy, Damascus University, Damascus, Syria*

E-mail address (corresponding author): dr.nathalooo@hotmail.com

Abstract

The recent outbreak of the deadly coronavirus disease 19 (COVID-19) pandemic poses serious health concerns around the world. The lack of approved drugs continues to be a challenge and further necessitates the discovery of new therapeutic molecules. Computational methods such as ligand-based drug design are promising approaches to discover novel inhibitors for coronavirus disease.
In this study, novel quantitative structure−activity relationship QSAR model for 28 cyclic sulfonamide derivatives that inhibit SARS-CoV-2 was built by multiple linear regression (MLR). To validate the proposed model, the studied compounds were divided into 23 compounds (training set) and 5 compounds (test set). The developed model was valid, robust, and predictive with correlation coefficient (R2) of 0.77 and 0.95 for training and test groups, respectively.
The model obtained six descriptors which best describe the activity. The six descriptors encode barysz matrix, atom count, and autocorrelation. The descriptors nCl that related to the atom count play more significant role in SARS-CoV-2 inhibitory activity as sensitivity analysis has shown.
The model is expected to be useful in virtual screening, providing important tools in the field of drug design, and orienting the direction of designing new SARS-CoV-2 inhibitors with better activity.

Key Words: *SARS-CoV-2; QSAR; descriptors; inhibitors; drug design.*

Pharmacoinformatics Study and High-Throughput Virtual Screening to Discover Effective Natural Compounds in Treatment of Erectile Dysfunction

Farshad Qalekhani[1´2], Fatemeh Nabizadeh[1´3*]

[1] Pharmaceutical Sciences Research Center, Health Technologies Institute, Kermanshah University of Medical Sciences (KUMS), Kermanshah ,Iran
[2] Medical Biology Research Center, Health Technologies Institute, Kermanshah University of Medical Sciences (KUMS), Kermanshah , Iran
[3] Department of Human Genetics, Faculty of Medicine, Hormozgan University of Medical Sciences (HUMS), Bandar Abbas , Iran

E-mail address (corresponding author): fatima.nbz71@gmail.com

Abstract

Erectile dysfunction (ED) is the consistent or recurrent inability to achieve and/or maintain penile erection sufficient for a satisfying sexual performance. It is expected that ≈322 million men worldwide will suffer from ED by 2025. Phosphodiesterase 5 (PDE5) is a well-studied enzyme that specifically targets cGMP produced by nitric oxide-mediated activation of the guanylyl cyclase. Considering the crucial function of cGMP generated through the activation of this signaling pathway in several physiological processes, pharmacological inhibition of PDE5 has been revealed to be beneficial in ED treatment. Sildenafil was the first PDE5 inhibitor that showed a wonderful efficacy in ED, but it can still cause several side effects. Here, we used a library of 1228 purchasable natural products to investigate their inhibitory effect on PDE5 through bioinformatics tools and databases. Primarily, the three-dimensional structure of PDE5 was downloaded from the "Protein Data Bank" database, containing the crystal structure of the enzyme together with its inhibitor, Sildenafil. Structure preparation for docking was performed using the UCSF Chimera program, and ADT software. Water molecules and Sildenafil were removed, and hydrogens were added to the structure. Then, docking between Sildenafil and PDE5 was performed using AutoDock Vina software, to validate the grid box and determine the positive control binding energy. Docking between PDE5 and the library of natural compounds was carried out by AutoDock Vina software on the MTiOpenScreen server. The top 20 compounds were later evaluated by the Lipinski rule, and five compounds that followed the rule were assessed for toxicity. Finally, Alpha-Spinasterol glucoside (Affinity = -11.1 (kcal/mol)) and Yuccagenin (Affinity = -11 (kcal/mol)) were chosen as the leading compounds, and the ADME properties of them were evaluated through the pkCSM server. After pharmacokinetics studies, we concluded that Alpha-Spinasterol glucoside and Yuccagenin as PDE5 inhibitors can be candidates for the treatment of ED.

Key Words: *Alpha-Spinasterol glucoside; Erectile dysfunction; Phosphodiesterase 5; Sildenafil; Yuccagenin.*

An In Silico Approach Towards Identification of Novel Chalcone Derivatives as Potential Monoamine Oxidases Inhibitors

Reza Naderi[1], Neda Adibpour[1], Shohreh Mohebbi[1], Hafezeh Salehabadi[1]*

[1] *Department of Medicinal Chemistry, School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, Iran*

E-mail address (corresponding author): r3naderi@zums.ac.ir

Abstract

Parkinson's disease is a chronic and progressive neurological disorder of the central nervous system that affects and disrupts the functioning of the body's motor system. In Parkinson's disease, Abnormal mitochondrial expression of the enzyme monoamine oxidases B (MAOB) causes the oxidation of compounds such as dopamine, serotonin, norepinephrine, 2-phenylethylamine, etc. Oxidation of these compounds leads to the production of reactive oxygen species, which by products ultimately lead to neurotoxic effects such as disruption of neurotransmitters. Chalcones are unsaturated carbonyl α and β compounds that exhibit a variety of biological effects. Recent studies show that chalcone derivatives also have good inhibitory effects on the MAOB enzyme. Today, various bioinformatics techniques are used in the manufacture and design of medicine to save time and money. Due to the need to develop inhibitory compounds that can selectively and reversibly have an inhibitory effect on the MAOB enzyme, in this study, we decided to design new chalcone derivatives and study molecular docking and how they interact with the active site of the enzyme.

Key Words: *Monoamine Oxidases B (MAOB), Molecular Docking, Chalcone Derivatives, Inhibitor*

Antiviral effect assessment of some venom gland peptides from Iranian yellow scorpion, "Odontobuthus doriae" against SARS-CoV-2 with the aim of drug development

Maryam Naderi Soorki[1]*

[1] *Department of Biology, School of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran*

E-mail address (corresponding author): maryam_naderi_soorki@yahoo.com

Abstract

SARS-CoV-2 is from enveloped virus family responsible for the COVID-19 pandemic. No efficient drugs currently available for treatment of infection caused specifically by this virus. Therefore, searching for effective therapeutic treatments for severe illness caused SARS-CoV-2 is crucial. Scorpion venoms are significant sources of peptides with pharmaceutical potential including antivirals. Although, some studies determined the antiviral effects of some scorpion peptides on the other members of Coronaviridae family, but no anti-SARS-CoV-2 effects of these peptides have been reported until now. In this study antiviral effects of three predicted antimicrobial peptides (ODAMP1, ODAMP2 and ODAMP3) from Iranian yellow scorpion "Odontobuthus doriae" were assessed by Computational methods.

Methods: 3D model of selected potent antimicrobial peptides was designed with I-TASSER. The best models based on Z-score and Ramachandran plot were. Themodels were refined by a 200 ns Molecular Dynamics (MD) simulation by using Gromacs 2021.2 software. Refined models were Docked with RBD domain of SARS-CoV-2 spike protein by using HADDOCK software. Docking of human ACE2 peptide with RBD domain also assessed at the same time. The docked complexes (RBD-peptide and RBD-ACE2) were refined again by a 100 ns MD simulation and then analyzed.

Result & discussion: the results of 3D-model designing and models refinement showed one cluster for ODAMP1 and 2 cluster for each one of the ODAMP2 and ODAMP3. The results from molecular docking based on HADDOCK Score and Z-score showed that ODAMP1 peptide has a high affinity for RBD domain compared to other peptides but the results of molecular dynamics simulation showed ODAMP2 has a high stability and affinity to the RBD domain of covid-19 spike protein among the three peptides. In fact, this peptide can be a good candidate for used as a factor to inhibit the RBD domain of Sars-COV2 virus in in clinical studies with pharmacological purposes.

Key Words: *COVID-19, Antimicrobial peptide, Iranian yellow scorpion, Molecular Dynamics simulation, Molecular Docking*

Molecular docking of HMG-CoA reductase and inhibitory effect of new derivatives of atorvastatin on this enzyme

Shiva najafi[1]*, Hajar Sirous[2]

[1] *Department of Biochemistry, Faculty of Basic Sciences, Falavarjan Branch, Islamic Azad University, Isfahan, Iran*
[2] *Bioinformatics Research Center, School of Pharmacy and pharmaceutical Sciences, University of Medical Science, Isfahan, Iran*

E-mail address (corresponding author): shivanajafi7@gmail.com

Abstract

Introduction: In the intracellular mevalonate cycle, HMG-CoA reductase catalyzes the conversion of HMG-CoA to mevalonate. Mevalonate's metabolic pathway is a very important metabolic pathway in eukaryotes and many bacteria that eventually produce many important compounds such as cholesterol, Isoprenoids and terpenes such as lycopene. All animal cells produce cholesterol at a rate of production that varies in cell type and organ function. About 20-25% of daily cholesterol production occurs in the stomach, other organs with a higher production rate include the intestines, adrenal glands and reproductive organs. The intracellular synthesis begins after a acetyl CoA molecule and a aceto acetyl CoA-1 molecule, which is hydrated to form 3-hydrogen-3methylglutaryl CoA1 (HMG-CoA). This molecule is then reduced to mevalonate by the HMG-CoA enzyme. This reaction has a slow rate, is a recoverable step in cholesterol synthesis and is the site of action for statins.
Description: As we know HMG-CoA reductase is inhibited by the statins family drugs. This enzyme, encoded by PDB ID: 1HWK, via homo saponins taxonomy is inhibited by atorvastatin. This drug inhibits this enzyme by blocking the active site of the enzyme and establishing hydrogen and ionic bonds with the amino acids in the active site, eventually reducing cholesterol. The aim of this project was to investigate the novel derivatives of atorvastatin and the HMG-CoA reductase inhibition.
Discussion and conclusion: After drawing the structure of the new atorvastatin derivatives by Marvin Sketch software and examining the amount of energy and its binding to the HMG-CoA reductase enzyme by Auto Dock Tools software, it was concluded that the new atorvastatin derivatives also inhibited the enzyme with high efficiency and inhibiting the active site of the enzyme, it inhibits it.

Key Words: *HMG-CoA, Atorvastatin, Auto Dock Tools, Cholesterol*

Bioinformatics analysis of Epitope-based candidate vaccine against the novel (SARS-CoV-2)

Sajjad Gharaghani[1], Romina Norouzi[2]

[1] Assistant Professor, Institute of Biochemistry and Biophysics (IBB), Department of Bioinformatics, Laboratory of Bioinformatics and Drug Design (LBD), University of Tehran
[2] University of Tehran, Kish International Campus, Institute of Biochemistry and Biophysics, Department of Bioinformatics, Laboratory of Bioinformatics and Drug Design (LBD)

E-mail address (corresponding author): s.gharaghani@ut.ac.ir

Abstract

The complete sequence of SARS-CoV-2 genome that get from NCBI (ACCESSION number: NC_045512.2) use in this study, and find ORF in the genome and to find signal peptide, use signalp database. for assessment of allergenicity of these proteins, use ALGpred database and filter protein based on amino acid composition and Blast search on allergen representative peptides (ARPs), and choose a protein that both algorithms confirmed that proteins are not Allergen. To investigate immunogenicity of proteins, use the vaxijen server and choose proteins with a threshold Overall Prediction for the Protective Antigen score up to 0.9.for toxicity analysis of proteins uses the Toxinpred database. To remove inside cell membrane proteins use (TMHMM) database. To Epitope prediction use IEDB, for linear epitope prediction and CBTOPE database for conformational epitope prediction. Epitopes that pass the criteria of toxicity, allergenicity, antigenicity are chose and sort based on scores, then get-togethers with linker (KK, GGGGS), and adjuvant (OmpA) (GenBank: AFS89615.1) add. (6xHis-tag) adds at the C terminal end of the vaccine construct. Use cluspro server for molecular docking study to identify binding affinity characterizing of candidate vaccines 1 and 2 with MHC Class I and MHC Class II.

Key Words: *Keyword: SARS-CoV-2: Epitope-based, vaccine design*

Drug-Target Interaction Prediction with Deep Learning and Recommender Systems

Seyed Amirhossein Nosrati[1]*, Seyed Amir Ali Ghafourian Ghahramani[1], Kaveh Kavousi[2]

[1] *Department of Computer Engineering, Sharif University of Technology, International Campus, Kish Island, Iran*
[2] *Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*

E-mail address (corresponding author): nosrati75@gmail.com

Abstract

Drug designing is a complex, costly, and time-consuming process with high failure chance. Therefore, drug repurposing is gaining importance. To this aim, one method is to identify new interactions between chemical compounds and protein targets called predicting Drug-Target Interactions (DTIs). DTIs are usually represented by bi-partite networks where predicting drug-target interaction can be formulated as the link prediction. Recently, Graph Neural Networks (GNNs) have achieved tremendous success on machine learning tasks defined over graph-structured data such as node classification and link prediction. They can extract informative features of the input graph and improve the performance of down-stream tasks. We propose a GNN-based framework to predict drug-target interactions. To this aim, we represent the drug-target interactions as a bi-partite network and we construct a protein-similarity network between proteins based on their structural similarities. The proposed framework learns informative representations for both drugs and target proteins in an end-to-end fashion. The learned representations are used to predict interactions. We prepared two data sets. The first one extracted from DrugBank containing relevant information about drugs, target proteins, and their interactions. As the second data set, we used the Yamanishi benchmark dataset containing interactions between drugs and different groups of enzymes, ion channels, GPCRs, and nuclear receptors.
The results indicate that the proposed framework exhibits acceptable performance and can get better results compared to some proposed methods in the literature. On the first data set, our method achieved 92% and 85% of accuracy over training and test sets. For the second data set, accuracies are 89%, 86%, 82%, and 80%, respectively, on four classes of targets.
The proposed GNN-based framework starts with random representations for drug and proteins and learns highly informative embeddings to predict the possible interactions. Results, indicate the high abilities of the method in predicting DTIs.

Key Words: *Graph Neural Network; Drug-Target Interaction; Drug Repurposing; DTI Classification;*

The new inhibition strategy of the p38 MAP Kinase protein by using mutagenesis and COXH11 inhibitor

Jalil Parchekani[1], Abdollah Allahverdi[1], and Hossein Naderi-Manesh[1´2]*

[1] Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran,Iran
[2] Department of Nanobiotechnology, Faculty of Biological Sciences, Tarbiat Modares University, Tehran,Iran

E-mail address (corresponding author): parchekanijalil@gmail.com

Abstract

Mitogen-activated kinase proteins family is involved in a variety of cellular functions, such as apoptosis, differentiation, transcription, and proliferation. Besides, p38 MAP kinases are associated with regulating cellular responses to various stimulus activities, like pro-inflammatory cytokines, heat shock, mitogens, and osmotic stress. In this study, we mutated the three amino acids of the p38 protein active site. Then inhibitor of COXH11 was docked to the binding pocket of p38 protein and subsequently, molecular dynamics (MD) simulations were carried out. The COXH11 inhibitor is a pyrazolo benzothiazine-based compound that has four rings and has an inhibitory effect on the P38 MAP kinase protein. Gromacs software version 5.0.1 was used to investigate mutation effects. The MD simulation time was set at 250 nanoseconds and the p38 protein molecule dissolved in the water of the TIP3P model. The results showed mutation caused p38 protein to form more non-bonded interactions with COXH11 inhibitor. The significant effects of COXH11 inhibitor on the p38 protein can be the potential novel strategy for the treatment of related diseases

Key Words: *Keywords: P38 MAP kinases, Molecular Dynamic Simulation, Protein Mutation*

Design of Potential Aminopeptidase N (APN) inhibitors via Pharmacophore and Docking-Based Virtual Screening

Majid Pourzamani Molan[1], Shohreh Mohebbi[1], Hafezeh Salehabadi[1]*

[1] *Department of Medicinal Chemistry, School of Pharmacy, Zanjan University of Medical Sciences, Zanjan, IranEmail address*

E-mail address (corresponding author): mpourzamani@zums.ac.ir

Abstract

Introduction: M1 family of metalloamineopeptidases. The enzyme aminopeptidase N regulates various of cellular Aminopeptidase N (APN), also known as CD13, is a metal-dependent membrane protease belonging to the functions by different mechanisms, such as peptide breakdown . Aminopeptidase N has a significant effect on various cellular pathways, including migration, invasion, angiogenesis, and tumor cell metastasis. Overexpression of APN with many diseases such as inflammation, viral infection and especially cancer. Since the enzyme aminopeptidase N plays a crucial role in cancer cell metastasis, it is now considered an attractive target for designing of anticancer drugs. Drug design using computers today is dramatically increasing due to its advantages such as low cost, high speed, and good accuracy. By using informatics, it is possible to identify new effective and robust compounds. In this study, a hybrid strategy including docking and virtual screening was used to identify new APN inhibitors.
Method: The crystal structure of porcine aminopeptidase-N complexed with bestatin, with the PDB ID of 4fkk, was obtained from Protein Data Bank (www.rcsb.org). A proper pharmacophore model was generated using Ligand Scout 3.12 on the most critical area on the APN active site. Then ZINC libraries (over 35 million compounds) were applied for virtual screening.
Results and Discussion: Collected compounds from virtual screening were followed by molecular docking studies. Seven compounds were selected based on docking score and complying with Lipinski's "rule of five". Selected compounds can be considered as a proper candidate in order to develop new APN inhibitors.

Key Words: *Virtual Screening, Molecular Docking, Aminopeptidase N, Inhibitor*

Synthesis and spectroscopic characterization of carboxamide ligands: anti-cancer potential validation by in vitro interaction studies with HSA, DFT, molecular docking

Parisa Rezaei[1], Sudabeh Shokrolahi[1], Aria Tajali[1], Ahmad Amiri[1]*

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): prezaii123@gmail.com

Abstract

Human serum albumin (HSA) is one of the basic components of blood plasma and it serves as a storage and carrier protein. Understanding and characterizing the interaction of drugs with HSA has attracted great research interests from decades. The nature and importance of these bindings have direct consequence on drug delivery, pharmacokinetics, pharmacodynamics, therapeutic efficacy and drug designing. Herein, two carboxamide ligands have been synthesised by the reaction of adenine with picolinic acid (HL1) and adenine with pyrazine-2-carboxylic acid (HL2) and characterised by UV–Vis, FT-IR, 1H-NMR, and Mass spectroscopy. Also, the optimized structures of these ligands have been investigated using the DFT/B3LYP method with the 6–311++G(d,p) basis set. These show the results of the calculations to be in accordance with the experimental ones.

In the current work, we studied the mechanism of interaction between the anticancer drug and carrier protein human serum albumin (HSA) by using a variety of spectroscopic techniques (fluorescence spectroscopy, and circular dichroism (CD) spectroscopy) and computational methods (molecular docking and molecular dynamic simulation). Fluorescence data indicated that interaction of drug with HSA changed the microenvironment around the tryptophan residue with excellent binding constant ($38.27\ mM^{-1}$ for HL1). Also considering that the K_q values for these ligands are $>2.0 \times 10^{10}\ M^{-1}\ s^{-1}$, indicated that the fluorescence of HSA was quenched by ligands with a static quenching mechanism. Computational molecular docking was carried out to investigate the HSA-binding pose of the compounds. the study of molecular docking also indicated that our ligands could strongly bind to the site I (subdomain IA) of HSA. According to the CD results these ligands can bind to the main blood carrier protein (HSA) and change the secondary structure of the protein.

Key Words: *carboxamide; HSA Binding; Molecular Docking; DFT; Circular Dichroism; MTT assay.*

Spectroscopic and Molecular Docking Studies of Co(III) Carboxamide Complexes with HSA

Parisa Rezvaninia[1], Sudabeh Shokrollahi[1], Aria Tajally[1], Ahmad Amiri[1]*

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): r.parisa72@gmail.com

Abstract

The binding of small molecules to Human Serum Albumin (HSA) has been investigated for many years through different spectroscopic techniques to illustrate details of the protein structure and binding mechanism. The carboxamide [–C(O)NH–] group, ubiquitous throughout nature in the primary structure of proteins, is an important ligand construction unit for coordination chemists. Metal complexes of carboxamide ligands have paid special attention in biological systems and DNA cleavage. We report here the synthesis, spectroscopic characterization and molecular docking studies of [Co(bpb)(Br)(H_2O)], (1) and [Co(bpb)(N_3)(H_2O)], (2) complexes where bpb^{2-}= N,N'-bis(pyridine-2-carboxamido) benzene dianion. The interaction of 1 and 2 with HSA was investigated under physiological conditions, using fluorescence and circular dichroism techniques. The interaction between the complexes and HSA resulted to fluorescence quenching with a blue shift at λ_{max} of 343 nm. The magnitude of the Kq value for both 1 and 2 complexes (~ 10^{11} $M^{-1}s^{-1}$) indicates the static mecnanism for the interaction type. Furthermore, molecular docking analysis by using Vina was done to obtain more information regarding the binding sites between the HSA and complexes. The results demonstrated the presence of strong HSA-complex interactions with the binding affinities of -9.7 kcal/mol and -9.3 kcal/mol for 1 and 2, respectively. However, the H-bonding interactions between H atoms of the axially coordinated water molecule and the oxygen atoms of proline (Pro-447B) and cysteine (Cys-448B) residues, led to the changes in the secondary structure of the protein.

Key Words: *carboxamide; HSA Binding; Molecular Docking; DFT; Circular Dichroism; MTT assay.*

Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against SARS-CoV-2 Omicron variant

Esmaeil Roohparvar Basmenj[1], Seyed Shahriar Arab[1]

[1] *Biophysics Department, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, IRAN*

E-mail address (corresponding author): e.roohparvar@modares.ac.ir

Abstract

SARS-CoV-2 is a single-stranded RNA virus which belongs to the own coronavirus circle of relatives and reasons Covid-19. Omicron variant is a SARS-CoV-2 variant and was first reported on November 9, 2021. Moreover, World Health Organization (WHO) declared it a cause for concern on November 26, 2021. In this project, we designed a multi-epitope vaccine candidate targeting the spike glycoprotein of the Omicron variant. Spike glycoprotein is the largest protein of four structural proteins discovered in coronaviruses. This protein is incredibly immunogenic, and spike anti-protein antibodies are discovered in improved SARS and Covid-19 instances, and most COVID-19 vaccine development efforts in the reaction to the COVID-19 pandemic aim to set off the immune response against the spike protein. In addition to being effective against the Omicron variant, the bioinformatics vaccine presented in the current project is also effective against other variants of the former coronavirus. Furthermore, the vaccine is expected to be effective against other variants that we will encounter in the future; because the epitopes used in this present research are found via the existing coronavirus sequences, which were sequenced in different countries since the beginning of the pandemic. It means that these sequences are somehow conserved in the coronavirus genomic sequence. The proposed vaccine candidate was evaluated for various characteristics such as antigenicity, allergenicity, toxicity, physical and chemical properties, secondary and tertiary structure, and its interaction with the immune system using the docking technique. Finally, the suggested vaccine candidate has passed all the computational tests and is now acknowledged as a viable coronavirus vaccine candidate.

Key Words: *COVID-19, Coronavirus, Reverse vaccinology, Spike Glycoprotein, Multiepitopes Vaccine*

Synthesis and in silico docking studies of curcumin-polyhydroxy derivatives as novel dual inhibitors of α-amylase and α-glucosidase

Pedram Routabi[1], Maryam Mehrabi[1]*, Reza Khodarahmi[2]*, Hadi Adibi[3], Masoumeh Mehrabi[2]

[1] Department of Biology, Faculty of Sciences, Razi University, Kermanshah, Iran
[2] Medical Biology Research Center, Health Technology Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran
[3] Pharmaceutical Sciences Research Center, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

E-mail address (corresponding author): m5_mehrab@yahoo.com

Abstract

Over the past twenty years, the prevalence of diabetes as one of the most common metabolic diseases has become a public health problem worldwide. Blood glucose control is an important factor in delaying the onset and progression of diabetes-related complications. Therefore, in this study, curcumin- polyhydroxy derivatives were synthesized and used as effective agents in the treatment of diabetes with inhibitory properties against two carbohydrate-hydrolyzing enzymes α-glucosidase (α-Glu) and α-amylase (α-Amy) which are known to be significant therapeutic targets for the reduction of postprandial hyperglycemia. In search of potent dual inhibitors of α-Amy and α-Glu, we have synthesized curcumin-polyhydroxy derivatives, characterized by FTIR and MS then the binding interaction details of compounds to α-Amy and α-Glu were determined using a molecular docking study. Molecular docking was performed using AutoDock Vina and docking results were analyzed by PyMOL v.2.3.2.1 and LigPlot+ v.1.4.2 software. Molecular docking indicated that curcumin derivatives mainly interacted with amino acid residues located in the active site of α-Amy and α-Glu. The binding energies obtained from the docking of α-Amy with curcumin (C1), C2, C3, and C4 derivatives were -8.4, -8.2, -8.5 and -8.7, kcal/mol, respectively. Also, C1, C2, C3, and C4 derivatives showed binding energies of -7.6, -7.8, -8.0 and -8.0 kcal/mol in communication with the active site of α-Glu, respectively. This study indicated that curcumin-polyhydroxy derivatives could occupy the active catalytic site of α-Amy and α-Glu, resulting in the inhibitory effect due to steric hindrance, but more preclinical and clinical studies should be carried out in the future.

Key Words: *Curcumin-based derivatives, Anti-diabetic, α-Amylase, α-Glucosidase, Molecular docking*

Discovery of novel small molecular ligands for TLR8 by computational methods

Farideh Sadeghkhani[1], Zahra Hajihassan[1]*, Sajjad Gharaghani[2]*

[1] Department of Life Science Engineering, Faculty of New Sciences & Technologies, University of Tehran, Tehran, Iran
[2] Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

E-mail address (corresponding author): s.gharaghani@ut.ac.ir

Abstract

The TOLL-like receptor 8 (TLR8) is an endosomal protein expressed in some of the innate immune cells and involved in viral infections. however, many recent studies are available, to prove their role in forming anticancer responses. This study aimed at the virtual screening of a great number of compounds to detect ligands with appropriate pharmacological properties. with this purpose, PDB codes of TLR8 in complex with its agonists were studied and a pharmacophore model was designed for each of them by using the Pharmit webserver. In the next step, all compounds available in two great databases, ZINC and Pubchem were screened by these models. Afterward, several filters were applied to detect the ADME-Tox features by use of the FAF-Drugs4 server. Then all the output compounds were docked to the TLR8 receptor. Subsequently, the highest docking scores were prepared for Molecular Dynamics (MD) simulation with GROMOS43a1 energy field by GROMACS software in 20ns. The result demonstrated two compounds are great candidates for further pharmaceutical studies for discovering either antiviral or anticancer medicines in the future.

Key Words: *Toll-like receptor 8; Virtual screening, Docking; Molecular Dynamics Simulation*

Molecular docking studies of novel drug derivative bound to main protease enzyme (Mpro) in Covid-19

Reza Sahebi[1]*

[1] *Metabolic Syndrome Research Center, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*

E-mail address (corresponding author): reza_sahebi2001@yahoo.com

Abstract

Background: COVID-19 antiviral drugs have recently been discovered and introduced in light of their reported high virulence in recent months. The 3C-like protease (Mpro/7L8I) is a promising target for anti-CoV drugs. The goal of this study is to inhibit COVID-19 with FDA-approved viral antiprotease drugs and their derivative.
Materials and Methods: This is a descriptive-analytic research project. PubChem and the Protein Data Bank (PDB) were used to obtain the tertiary structure of COVID-19 Mpro as well as the drug's compounds. Molecular docking was screened using MVD (molegro virtual docker), version 6, with a grid resolution of 0.30 Å. The calculated ligand receptor (protein) interaction energy is represented by docking scores (DOS). As a result, more negative scores indicate a stronger binding tendency.
Results: The docking of COVID-19 protease (7L8I) investigated with three selected FDA-approved viral antiprotease drugs (Lopinavir, Cobicistat, and Ritonavir) followed by a cobicistat derivative. Cobicistat and cobicistat derivative had the best DOS (-188.89 and -215.12 respectively) and Lopinavir and Ritonavir had the worst DOS (7252 and 3199 respectively). Mpro could be targeted with 11 different conformations at its binding site by the best-screened ligand, a cobicistat derivative (InChIKey=ZHYPAEQWZKWGPZUHFFFAOYNA-N), which was identified as the most effective.
Conclusion: The findings revealed that one of the three FDA-approved drugs selected for the study can be a potent inhibitor of COVID-19. Among them, a cobicistat derivative may be the most effective for the treatment of the disease. On the basis of the findings, it is recommended that in-vitro and in-vivo studies be carried out to determine the efficacy of this drug against the COVID-19 infection virus.

Key Words: *Cobicistat; Ritonavir; Lopinavir; Molecular Docking; Main Protease; Anti Viral Drugs*

Looking for a AKT inhibitor; From similarity search to molecular docking

Mozhgan Samiepur[1]*

[1] *Bioinformatic department, Zist Pardazesh institute, Isfahan, Iran*

E-mail address (corresponding author): siva.85374@gmail.com

Abstract

Akt is identified as a protein kinase B (PKB) member of the Serin/Threonine kinase family. This protein is an oncogene that regulates growth, proliferation, cell survival, apoptosis, and glycogen metabolism. Overexpressed AKT is observed in many human cancers such as ovarian, lung, and pancreatic cancers. Thus, AKT can be considered a therapeutic target for cancer.Recent studies have shown that chemical inhibitors of AKT can successfully overcome chemotherapy resistance in cancer patients.

In this study, we used molecular docking for virtual screening of compounds similar to WFE (an inhibitor of AKT) to find a compound with high-affinity binding to akt1. At first, we performed a similarity search for WFE by enforcing the Lipinski rule of five filters. Then, we conducted molecular docking by "PyRx" software and ultimately used "lig plot" for analysis of dockings results.

The results indicate that the compound with PubChem CID=147779364 has the best energy binding from the compound library (ΔG=-12.1). Therefore, this compound can be utilized as a potential inhibitor for more research on cancer treatment.

Key Words: *Key words: In-silico Drug design; AKT inhibitor; Virtual screening; Cancer*

Virtual Screening of EGFR Inhibitors for Cancer Treatment

Samane Samiepur[1]*

[1] *Bioinformatic department, Zist Pardazesh institute, Isfahan, iran*

E-mail address (corresponding author): samanehsamiepour@gmail.com

Abstract

EGFRs are a big family of tyrosin kinase receptors involved in many cancers such as breast, lung, esophageal, head, and neck. EGFR and its family contribute to a signaling cascade regulating cancer cells' growth, differentiation, adhesion, migration, and survival. The Binding of ligands such as TGF and EGF to the extracellular domain of EGFR induces tyrosin kinase domain dimerization and initiates tumor formation. EGFR is overexpressed in some human cancers and is considered a promising target for new anticancer drugs. Erlotinib is a common inhibitor of EGFR that blocks its function by binding to the intracellular tyrosine kinase domain and restricts downstream signaling pathway leading to cancer.
In this project, we used virtual screening method by using molecular docking. For this purpose, we used erlotinib as a query for similarity search to make a library of compounds. We filtred them by Lipinski's role of five filters. Then, high-throughput docking software "PyRx" was used for molecular docking of these compounds with EGFR. Finally, the results of docking were analyzed by "Lig Plot" software.
Results show that of the 135 compounds obtained from similarity search, the compound with pubchem CID=1040492277 has the most significant binding energy(ΔG=-12.7). This compound can be considered a potential inhibitor of EGFR that has more significant energy binding than erlotinib and can be used in further investigation.

Key Words: *Key words: cancer; EGFR; molecular docking; virtual screening; signaling pathway*

In Silico Design of a novel peptide inhibitor derivative of azurin peptide affecting gastric cancer

Shahrzad Shahbazi[1], Somayeh Reiisi[1]*

[1] *Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran*

E-mail address (corresponding author): shahrzad.shahbazi1997@gmail.com

Abstract

Background: Antimicrobial peptides are short amino acid sequences, involved in diverse functions. They have anti-cancer properties, probably due to controlling mutagenic compounds by direct binding to carcinogens or inhibition of microbes that produce these agents. Azurin is a copper-comprising redox protein secreted by Pseudomonas aeruginosa. Azurin and its derived peptide can selectively enter into different types of cancer cells and induce cell cycle arrest or apoptosis. The use of computational methods for drug designing has been considered a powerful technique in recent decades. Thus, the aim of the current study is to improve the approach for the identification of possible peptide drug molecules for gastric cancer targeting through bioinformatics analysis.
Materials and Methods: Azurin peptide sequence submits to AntiCP server for predict and design of newly anticancer peptides. The 2D and 3D structures of these peptides were drawn with Protean (DNAstar software) and Phyre2 respectively. The 3D structures of peptides were docked with the VEGFR using docking server swissdock and the peptide with the maximum binding energy value was identified. As well, anticancer activity for each peptide was estimated with iACP software.
Results: AntiCP server predicts >200 peptides with the anticancer property. Anticancer activity of newly designed peptides showed 98-99% anticancer specificity. Results of this study indicated that among 200 peptides that were virtually screened, 18 peptides showed the high anticancer property and among these peptides, DSRVIEHTKLIGSGEKDSVTFDVSK was identified as a potential sequence for VEGFR ligand.
Conclusion: Our results revealed the possibility of a successful design of a new anticancer peptide derived from the Azurin peptide. Therefore, the newly designed peptide can be proposed as a potential inhibitor agent against gastric cancer.

Key Words: *Anti-cancer peptides; Azurin peptide; Gastric cancer; In Silico*

Modeling anticancer drugs on the pathway JAK-STAT in cancer stem cells

Zahra Shakibay Senobari[1]*, Mohsen Masoumian Hosseini[2], Ziba Veisi Malekshahi[3], Azadeh Hekmat[1]

[1] *Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran,*
[2] *Department of E-learning in medical science, Virtual University of Medical Sciences, Tehran, Iran*
[3] *Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): zahra.shakibaysenobari@yahoo.com

Abstract

In mammals, the JAK / STAT pathway is the major signaling mechanism for a wide range of cytokines and growth factors [1,2]. Many recent studies show that inhibition of STAT3 in various tumor cells has anticancer effects. In this paper, the functional regions of STAT3 were analyzed and structures with PDB codes 4E68-A, 1BG1-A, 3CWG-A, 1YVL-A, 1BF5-A, 4ZIA-A and 1Y1U-A were selected to generate homology models. Water, ligands, and additional chains of these structures were removed using Chimera 1.8 software for modeling with Modeler 9. ProSA-website and Rampage were used to confirm the accuracy of the modeling. STAT Inhibitors FLLL32, HJC0123, STX-0119, LLL12, INS3-54, and Piperlongumin were created using MarvinSketch software and minimized using Hyperchem software. The compounds Stattic and Sta-21 were obtained from PubChem. Molecular docking was performed using AutoDock tools and evaluated after 100 runs. This study showed that despite the explanation of some articles regarding the inhibitory effect of compounds HJC0123, Ins354, Sta-21, and Statiic on the DNA-binding domain, Ins354 appears to bind more strongly to DNA by binding to arg432 and lys574 than the other compounds (-0.26; -8.29). Among these compounds, Statiic appears to bind to both the DNA-binding domain and the SH2 domain with greater stability, inhibiting dermirization and DNA binding. By inhibiting these processes, transcription of the gene is inhibited and the cancer cell can no longer replicate and continue to grow. Since the drugs Ins354 and Sta-21 have the same inhibitory effect, it is expected that taking two drugs at the same time will lead to a competitive situation and the inhibitory effect will not be different from the indicated amount, but may be worse due to the side effects of the two drugs. However, taking one of these two drugs in combination with Statiic may increase the inhibitory effect.

Key Words: *STAT3; SH2 domain; DNA binding domain; Anti cancer drugs; Molecular Docking*

The new approach based on MD studies to design novel anticancer drugs based on amino acids and carbohydrates to inhibit DHFR

Amin Mohammad Shapouri, Alireza Fattahi, Sepideh kalhor

E-mail address (corresponding author): aminmohammad.shapouri@sharif.edu

Abstract

DHFR is one of the essential enzymes in the Thymidylate synthesis cycle. The expression of this enzyme in tumor cells and some related autoimmune diseases such as Psoriasis, which is a derma-related illness, has been observed. The inhibition of this enzyme will lead to suppressing the production of 5,10-Methylenetetrahydrofolate in cells. 5,10-Methylenetetrahydrofolate is produced from tetrahydrofolate with the aid of SHMT. As a result, the mentioned inhibition will cause cell death. Methotrexate is one of the drugs for treating Psoriasis, which has many side effects such as intestinal problems, respiratory disorders, blood disorders, central nervous system disorders, and elevated liver function.

This project aims to replace Methotrexate with drugs based on carbohydrates and amino acid blocks with fewer side effects.

To achieve our goals in designing new drugs, we employ quantum mechanics and molecular dynamic methods. Firstly, concerning the main drug structure (Methotrexate), some new structures based on amino acids and carbohydrates will be suggested. Then, the newly designed structures will be optimized at the DFT/rB3LYP/6-31G+(d) level. It should be noted that the primary conformer search of the structures will be carried out using Spartan software. Following that, Autodock and Autodock/vina are the software used to select the best replacements. Outputs of the docking procedures will be used for molecular dynamics simulations. Furthermore, the energy analyses for this study will be performed by MM/PBSA method. We hope that the results of this study will lead to the design of more effective and less toxic anticancer and anti-autoimmune drugs.

Key Words: *Inhibitor; Drug design; Computational design; Anti-cancer compounds; Sugar and amino acid; Methotrexate*

Investigation and molecular dynamic study of polymeric encapsulation of anticancer drug doxorubicin as drug delivery systems

Mozhgan Shavandi[1]*

[1] *Materials and Energy reaserch center, Karaj, Iran*

E-mail address (corresponding author): mojhgan_19@yahoo.com

Abstract

In this study, Molecular dynamics (MD) simulations were used to investigate the dynamics of doxorubicin anticancer drug poly (ethylene glycol)/ Polysaccharide system. Process conditions were performed in an aqueous environment at human body temperature (310 K) and physiological pH level of 7.4. All simulations were performed by GROMACS 5.0.7, and the time-step was 2 fs. In all system, number of polymeric system and DOX molecules with different charge fractions were randomly solvated in a box include TIP3P water molecules. The particle mesh Ewald (PME) method was utalized to calculate the long-range electrostatic interaction. The cut-off was 1.2 nm and in all system, the MD simullation was carried out in the NPT ensemble. We investigated different polymer chain length (3, 5 and 7) in diffrent concentration. The result showed that the polymer chain length impact on DOX loading. The PEG-DOX system illustrated that interactions increased as the PEG chain length increase.

Key Words: *Anticancer drug; Doxorubicin; Molecular dynamics simulations, MD, Polymer*

In-Silico Study of Fibroblast Activation Protein Binding of Recent FDA Approved Inhibitors

Sudabeh Shokrollahi[1], Ahmad Amiri[1]*, Aria Tajalli[1]

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): s.shokrollahi1991@ut.ac.ir

Abstract

Fibroblast activation protein (FAP) is a well-defined marker, expressed at high levels on the cell surface of cancer associated fibroblasts (CAFs). FAP, a constitutively active serine peptidase with both dipeptidyl peptidase IV (DPP IV) and collagenase/gelatinase activity, promotes malignant and invasive behavior of epithelial cancers. High stromal expression levels of FAP correlate with poor prognosis. FAP is difficult to detect in non-diseased adult tissue, but it is generally expressed at sites of tissue remodeling. Several potential FAP-targeted approaches consisting of vaccines, antibodies, prodrugs, and inhibitors have been exploited in preclinical studies. Among them, a class of FAP inhibitors (FAPi) with a N-(4-quinolinoyl)-Gly-(2-cyanopyrrolidine) scaffold displayed nanomolar affinity and high selectivity against other interfering dipeptidyl peptidases and prolyl oligopeptidase. In this study, several positron emission tomography (PET) tracers including FAPI-02, FAPI-04, FAPI-46, and FAP-2286 were investigated. Quantum chemical calculations of these compounds have been carried out by DFT at the B3LYP/6–311++G(d,p) level. An analysis of the calculated vibrational frequencies was performed and significant bands were specified. Furthermore, the binding affinity between the above-mentioned inhibitors and FAP was studied under simulated physiological conditions, using molecular docking (MD). The results show that the FAP-2286 compound has more affinity for binding to FAP.

Key Words: *FAP; DFT; Molecular Docking; CAFs; PET.*

In-Vitro and In-Silico Study of Naphtalene-Based Schiff-Bases; Synthesis, Charactrization, HSA Binding, and MTT Assay

Sudabeh Shokrollahi[1], Ahmad Amiri[1]*

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): s.shokrollahi1991@ut.ac.ir

Abstract

Drug–albumin complexes can be regarded as fundamental model for the drug–protein-binding, owing to the stability, availability, and extraordinary binding capacity of albumin. In summary, the investigation of the interactions between anticancer agents and HSA is bound to be revealing [1,2]. Herein, four Schiff-bases have been synthesised by the reaction of 1,5-Naphthalenediamine (p-ND) with the four aldehyde drivatives and characterised by UV–Vis, FT-IR, 1H-NMR, and mass spectroscopy. Quantum chemical calculations of synthesized compounds have been carried out by DFT at the B3LYP/6–311++G(d,p) level; these show the results of the calculations to be in accordance with the experimental ones. A comparative analysis of the experimental and calculated vibrational frequencies was performed and significant bands were specified. The binding affinity between our Schiff-bases and human serum albumin (HSA) was studied under simulated physiological conditions, using absorbance titration experiments, fluorescence spectroscopy, circular dichroism (CD), and molecular docking (MD). Interaction results revealed one molecule of synthesised Schiff-bases to bind to protein. In-vitro anticancer activity of the synthesised compounds was evaluated against the human hepatocellular carcinoma (HepG2) and human breast (MCF-7) cancer cells using MTT assay. Among the compounds, L3 (containing a metoxy substituents) exhibited the highest anticancer activity.

Key Words: *Schiff Base; HSA Binding; Molecular Docking; DFT; Circular Dichroism; MTT assay.*

Identification of a novel human dihydrofolate reductase inhibitor by virtual screening, docking, ADMET prediction and molecular dynamics simulations

Asma Soofi[1], Nahid Safari-Alighiarloo[2]*

[1] *Department of Physical Chemistry, School of chemistry, College of Sciences, University of Tehran, Tehran, Iran*
[2] *Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): safari.nhd@gmail.com

Abstract

Dihydrofolate reductase (DHFR) is a critical enzyme that catalyzes the reduction of dihydrofolate to tetrahydrofolate (THF), responsible of the synthesis of raw material for cell growth and proliferation. Inhibition of DHFR resulted in a deficiency of THF and cell death. This feature makes human DHFR (hDHFR) as an attractive target for cancer therapy. Accordingly, we aimed to identify a novel effective hDHFR inhibitor using series of in silico approaches. In this study, we performed structure-based virtual screening of the PubChem database to identify potent hDHFR inhibitors. First, we retrieved all compounds from the PubChem database having at least 90% structural similarity with the known natural hDHFR inhibitors. Afterwards, the compounds were screened through molecular docking methods against hDHFR to investigate protein–ligand interaction and estimate their binding affinities. The sixteen compounds with higher binding affinity compared to methotrexate (MTX), as a reference compound, were selected. These compounds displayed essential molecular orientation and interactions with key residues of the hDHFR active site. They further investigated by Lipinski and ADMET prediction. At this stage, nondrug-like compounds were excluded based on computational approaches. Then, potential inhibitor (PubChem CID: 46886812) were identified and docked into the actives sites of hDHFR to investigate the binding modes. Finally, molecular dynamics (MD) simulation carried out to evaluate the stability of ligand-target complexes. We found that binding of CID: 46886812 stabilized the structure of hDHFR with the lowest fluctuations. In conclusion, this compound (CID: 46886812) can be suggested as a potential anti-cancer drug due to its non-toxicity, high binding affinity, and specificity towards the inhibition of hDHFR.

Key Words: *Dihydrofolate reductase (DHFR), Virtual screening, Molecular docking, Molecular dynamics simulation, ADMET*

Carotenoids as potential inhibitors of TNFα in COVID-19 treatment

Farzaneh Taghipour[1], Nasrin Motamed[1]*, Mohammad Ali Amoozegar[2], Maryam Shahhoseini[1,2,3], Soodeh Mahdian[3,4]*

[1] *Department of Cellular and Molecular Biology, School of Biology, College of Science, University of Tehran, Iran*
[2] *Department of Microbiology, School of Biology, College of Science, University of Tehran, Iran*
[3] *Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran*
[4] *Reproductive Epidemiology Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran*

E-mail address (corresponding author): motamed2@khayam.ut.ac.ir

Abstract

Tumor necrosis factor-alpha (TNF-α) is a multifunctional pro-inflammatory cytokine, responsible for autoimmune and inflammatory disorders. In COVID-19 patients, increased TNF-α concentration may provoke inflammatory cascade and induce the initiation of cytokine storm that may result in fatal pneumonia and acute respiratory distress syndrome (ADRS). Hence, TNFα is assumed to be a promising drug target against cytokine storm in COVID-19 patients. In the present study, we focused on finding novel small molecules that can directly block TNF-α-hTNFR1 (human TNF receptor 1) interaction. In this regards, TNF-α-inhibiting capacity of natural carotenoids was investigated in terms of blocking TNF-α-hTNFR1 interaction in COVID-19 patients with the help of a combination of in silico approaches, based on virtual screening, molecular docking, and molecular dynamics (MD) simulation. A total of 125 carotenoids were selected out of 1204 natural molecules, based on their pharmacokinetics properties and they all met Lipinski's rule of five. Among them, Sorgomol, Strigol and Orobanchol had the most favorable ΔG with the best ADME (absorption, distribution, metabolism, excretion) properties, and were selected for MD simulation studies, which explored the complex stability and the impact of ligands on protein conformation. Our results showed that Sorgomol formed the most hydrogen bonds, resulting in the highest binding energy with lowest RMSD and RMSF, which made it the most appropriate candidate as TNF-α inhibitor. In conclusion, the present study could serve to expand possibilities to develop new therapeutic small molecules against TNF-α.

Key Words: *TNF-α, TNFR1, Carotenoid, COVID-19, Virtual Screening, Molecular Dynamics*

Experimental and Computational Studies of ND-Drived Schiff base ligands with HSA; Molecular Dynamics Simulation, QSAR Modeling and Molecular Docking Studies

Aria Tajally[1], Sudabeh Shokrollahi[1], Ahmad Amiri[1]*

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): aryatajally@yahoo.com

Abstract

Human Serum Albumin (HSA) is a negatively charged, is present in high concentration in plasma. Existence of several binding sites and an extraordinary binding capacity with a high degree of conformational flexibility makes HSA a versatile carrier to transport a variety of endogenous and exogenous ligands including drug molecule. Herein, twelve Schiff-base ligands (L1-L12) have been synthesised by the reaction of 1,5-naphthalenediamine (ND) with four aldehyde drivatives and characterised by different spectroscopic techniques. The interaction of these ligands with HSA was investigated under pseudo-physiological conditions by fluorescence and circular dichroism (CD). The fluorescence quenching of HSA at 343 nm upon addition of the L1-L12, reveals the formation of complexes between Schiff-bases and HSA. The magnitude of the Kq values for L1-L12 ligands (greater than 2.0 × 1010 M−1 s−1) confirm the static mechanism for quenching. Furthermore, the CD spectra show that the random coil and antiparallel parts of the secondary structure have trends inverse to the helix part in the presence of Schiff base ligands. The optimization of the structures was performed by using DFT/B3LYP method with the 6–311++G(d,p) basis set. Finally, the molecular docking studies with AutoDocVina software and molecular dynamics simulations using Gromacs package were applied to estimate the binding affinity between HSA and L1-L12 ligands. The results show the strong intreraction between the protein and the ligands with the binding affinities in the range of -7.2 ─ -11.1 kcal/mol. The diffrences of the binding constants for the interaction of L1-L12 with HSA was realized by QSAR toolbox.

Key Words: *Schiff Base; HSA Binding; Molecular Docking; DFT; Circular Dichroism; MTT assay, Gromacs package, Qaussian,Q-SAR toolbox.*

Structure based pharmacophore modeling, virtual screening and molecular docking approaches for identification of natural anti-cancer metastasis agents targeting Wnt/β-catenin pathway

Neda vaghefinezhad[1], Majid tafrihi[1]*

[1]*Department of Molecular and Cell Biology, Faculty of Sciences, University of Mazandaran, Babolsar, Mazandaran, Iran*

E-mail address (corresponding author): m.tafrihi@umz.ac.ir

Abstract

Cancer metastasis is a multi-stage process in which a cancer cell spreads from the primary site of the lesion, passes through the circulatory system, and establishes a secondary tumor at a new nonadjacent organ. Dietary phytochemicals (DPs) modulate numerous biological events including epigenetic changes and signaling transduction pathways such as Wnt/β-catenin. Some of these compounds which stabilize cell-cell adhesions are controversial topics that change the expression of a variety of Wnt target genes. So they may induce cell-cycle arrest, apoptosis, and/or inhibition of Epithelial-Mesenchymal Transition (EMT) and metastasis. Natural molecules that target the Wnt/β-catenin pathway include flavonoids, polyphenols, terpenes and terpenoids, secosteroids, and alkaloids.

In this study, we aimed to discover novel potent WNT/β-catenin pathway inhibitors through β-catenin (PDB ID: 1JDH) structure-based virtual screening and pharmacophore modeling. 28 bioactive molecules were selected from different plants, after which we performed analyzes such as molecular docking, pharmacophore modeling, and Lipinski's Rule of Five (RO5) filter. The Camptothecin molecule had the best ligand-based pharmacophore model and docking energy. This model was applied to screen Pubchem molecular library with more than 9.3 million compounds for the novel β-catenin inhibitor. The hits were subsequently subjected to molecular docking after being filtered by Lipinski's rules. After screening the molecular library through molecular docking, pharmacophore modeling, and Lipinski's Rule of Five (RO5) filter, we proposed eight compounds out of 539 structurally representative top hits as the most potent inhibitors (Compound CID: 11317647; 102335601; 137313625; 129670516; 57244149; 90680097; 90680098; 57016104). Finally, the novel inhibitors proposed in this study need further consideration to uncovering cancer treatment and with the generated pharmacophore model, more potent β-catenin inhibitors can be easily screened.

Key Words: *β-catenin inhibitors; Cancer metastasis; molecular docking; pharmacophore modeling.*

Investigation of some natural extract compounds against COVID-19 by Molecular Docking study

Akbar Vaseghi[1], Sanaz Soltangheis[2], Reza Ashrfi Parchin[3], Kosar Rezaee chamanie[4], Majid Sadeghizadeh[1]*

[1] Demarteman of Nanobiotechnology, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran
[2] Department of Chemical engineering, University of Mohaghegh, Ardabil, Iran
[3] Excir Faravaran Sabalan company, Ardabil Science and Technology Park, Ardabil, Iran
[4] Department of Biology, Faculty of Science, University of Guilan, Rasht, Iran

E-mail address (corresponding author): vaseghi.akbar@gmail.com

Abstract

We used the molecular docking technique on the ACE2, Heat Shock Protein A5 substrate-binding domain b (HSPA5 SBDb), proteins in the human body, and the main protease (PDB6LU7) protein of SARS-CoV-2. We describe from silico studies on the host-cell receptor recognized by the viral spike protein that leads to an essential foundation about SARS-CoV-2 resistance of individual compounds. In this study, 11 natural compounds, which have antiviral properties according to previous studies, have been selected as small molecules candidates in the molecular docking study of spike and PDB6LU7 proteins of SARS-CoV2 and also ACE2, TMPRSS2, and HSPA5 proteins in human cells. Binding constants of CAPE, Apigenin, Acacetin, Rutin, Chrysin, Galangin, Kaempferol, Quercetin, Artepillin c, Cinnamic acid, Prenyl caffeate, and three drugs as conventional antivirus include Oseltamivir, Heparan sulfate, and Acyclovir were measured using the AutoDock 4.2 molecular docking program. The results showed a high binding affinity for the Rutin, Galangin, and Quercetin to the ACE2, HSPA5, TMPRSS2, and 6LU7 protein from -8.1 to -10.7 kcal/mol. Also, Chrysin had the best inhibition potentials among the studied molecules with high binding energy -9.4 kcal/mol from S protein. Our studies showed that rutin had the best inhibition potentials among the studied molecules with high binding energy again ACE2, HSPA5, TMPRSS2, 6LU7, and S protein. Among these compounds, Rutin might compete with Covid-19 for ACE2, HSPA5, TMPRSS2, 6LU7, and S proteins and might prevent or delay the entry of Covid-19 into the cell. It is followed by myricetin, caffeic acid phenethyl ester, hesperetin, and pinocembrin. In conclusion, the high potential of polyphenolic agents and flavonoids in propolis to bind to human and viral proteins associated with the SARS-CoV-2 pandemic indicates that has high potential in the treatment of Covid-19.

Key Words: *SARS-CoV-2, natural compounds, Molecular Docking*

Exploring the interactions between artemisinin with serum paraoxonase -1 using molecular docking technique

Maziyar Veisi[1]*, Behnaz Karimi[2], Marzie Shadpirouz[3], Zahra Hemati[4], Shaahin Veisi[5]

E-mail address (corresponding author): maziyar.veisi@gmail.com

Abstract

Background & Objective: Artemisia annua is an herb native to Asia that is employed in traditional medicine for the therapy and prevention of fever and trembling. Artemisinin is one of the most important derivatives of Artemisinin annua with anti-inflammatory, Antimicrobial, and Antioxidant properties. On the other hand, Paraoxonase -1 is a vital enzyme with Anti-inflammatory and antioxidant properties, produced in the liver and linked to HDL. So, studies about their impact on each other can have a significant position in pharmacology and drug design.
Material & Methods: This is a descriptive-analytical study. In the present study, the structures of the drug (artemisinin) and serum Paraoxonase -1 enzyme were respectively obtained from the PubChem and Protein Data Bank (PDB) databases. Then to investigate how the compound is attached to the active site of the enzyme, a docking study was performed by AutoDockTools-1.5.6 software.
Results: Findings from molecular docking show that ligand artemisinin had the most negative ΔGbind (-6.45 Kcal/mol) than Paraoxonase -1 (-2.58 kcal/mol), which indicated favorable interactions with the critical amino acid residues at the active site of the enzyme.
Conclusion: The interaction studies indicated that Artemisinin with the serum Paraoxonase1 possesses a high binding affinity.

Key Words: *Artemisinin; serum Paraoxonase-1; Molecular docking; drug design*

2D QSAR model development for nitroimidazoles radiosensitizers DNA binding activity based on 2D and 3D descriptors

Mostafa Zakariazadeh[1], Somaieh Soltani[2]*

[1] *Department of Biology, Faculty of Sciences, Payame Noor University, Tehran, Iran*
[2] *Department of Medicinal Chemistry, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran*

E-mail address (corresponding author): somaieh.s@gmail.com

Abstract

Nitroimidazoles compounds have been used in cancer therapy as radiosensitizers. They interact with DNA duplex structure via intercalating mode and change its conformer and interrupts its replication cycle. To the prediction of their radiosensitization effectiveness (pC1.6), a linear model was developed based on 2D and 3D descriptors generated by PaDEL software. 84 nitroimidazole compounds were divided into train (62 compounds) and test set (22 compounds). The 3D structures of compounds were prepared by Avogadro software and geometry optimized by MMFF94 force field. Descriptor selection was done using stepwise regression. Developed linear models were validated based on the internal and external validations. 9 descriptors were selected from 1875 calculated descriptors. Selected descriptors were AlogP, BCUTc1h, ASP4, MLFER_A, MLFER_E, TDB6s, TDB10s, MOMI-X, and RDF30m. The intercorrelatuon between selected descriptors were < 0.65. The correlatin coefficient between experimental and predicted pC1.6 were 0.77 for the external test set. Obtained results indicated a good prediction potency for the developed model in comparision to previous models. The selected descriptors represented the importance of both 2-D and 3-D characteristics of nitroimidazoles in their effect on DNA duplex conformation change. The obtained data can be used for improve the effectiveness of studied compounds against cancer due to their binding to DNA.

Key Words: *QSAR; DNA-binding; nitroimidazole compounds, descriptor*

Theoretical study of the inhibitory effect of bee venom peptides on coronavirus spike protein

Mahtab Zarean[1], Karim Mahnam[1]*

[1] *Department of Biology, Faculty of Basic Sciences, Shahrekord University, Iran*

E-mail address (corresponding author): mahnam.karim@sku.ac.ir

Abstract

The spike protein of coronavirus binds to the ACE2 receptor protein at the surface of human cells and causes the entrance virus to the cell. Bee venom peptides have anti-inflammatory effects and have been suggested for the treatment of coronary heart disease. To test their ability for binding to the spike protein, the three-dimensional structure of the spike protein and peptides were downloaded from the protein bank or prepared by the Hypercom software and subjected to molecular dynamics simulation for 100 nanoseconds by the Gromacs software. The peptides were then docked to the RBD domain of the spike protein by the Haddock server and the complexes were simulated again for 100 ns. The binding energy of the peptides to the RBD domain was calculated by MM/PBSA method. The results showed that the peptides MISMLRCIYLFLSVILITSY and MKFLVNVALVFMVVYISFIY had better binding energy and could inhibit it better by binding to the spike protein and preventing the virus from entering the cells. Also, several mutations in these peptides were designed with the help of Beatmusic server, and the same procedure was repeated for them. The results showed that the MISMLRCIYLFLSVILITSY mutation had better binding energy than the others and this peptide suggests as a potential drug for coronavirus disease

Key Words: *Coronavirus, Spike protein, Molecular dynamics simulation, Bee venom peptides, Mutation*

Modeling in Computational Biology

Investigation of Novel Regulators of the COPD Development Based on Systems Biology Approaches: A WGCNA Study

Hojat Borna[1], B. Fatemeh Nobakht M. Gh[1], Yazdan Hasani Nourian[1], Masoud Arabfard[1]*

[1] *Chemical Injuries Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): arabfard@gmail.com

Abstract

Background: Chronic Obstructive Pulmonary Disease (COPD) is a well-known chronic disease, which is the third worldwide disease-related leading cause of death. A complex combination of factors over a long time prepares conditions for emergence of this progressing disease. Understanding genetic basics and behavior of the major actors of the disease including immune systems Polymorphonuclear Cells (PMNCs), sheds light on turning points of COPD development.
Method: To achieve the goal, different transcriptomics containing databases were surveyed to find proper COPD containing microarray data. Chosen dataset (GSE42057) was submitted to analyze with WGCNA and Limma, in order to find co-expressed genes as well as relationship between modules and clinical traits and possible differentially expressed genes, respectively.
Results: Data analysis with Limma was unable to present any possible Differentially Expressed Gene (DEG). However, expression network analysis revealed prominent hub genes as well as strong relationships between some identified modules and Eigengenes with clinical variables. It is revealed that two highly altered modules containing mostly inflammatory and remodeling pathways are highly related with BMI and golden stages of COPD.
Conclusion: It is interesting that cellular integrity control genes such as tubulin proteins(TUBA1A), aquaporins (AQP1), peroxisome proliferator-activated receptor (PPAR), and Nitric Oxide Synthase 2(NOS2) are also discovered to be significantly involved in the pathogenesis of the process.

Key Words: *COPD, WGCNA, Module, PMNC*

Molecular docking and molecular dynamic simulation of Curcumin on three pre-apoptotic factors of Bad, Bak, and Bim

Fateme Davarani Asl[1], Pegah Khosravian[2], Mehdi Rezaee[1], Javad Saffari-Chaleshtori[3]*

[1] Department of Medical Biotechnology, School of Advanced Technologies, Shahrekord University of Medical Sciences, Shahrekord, Iran
[2]Medical plants Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran
[3]Clinical Biochemistry Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran

E-mail address (corresponding author): j_saffari@yahoo.com

Abstract

Curcumin, a polyphenol compound, is derived from the plant Curcuma longa. This antioxidant with apoptotic effects can lead cells to death especially in cancer cells. This molecular dynamic and molecular docking simulation study investigated the effects of curcumin on three pre-apoptotic factors of Bad, Bak, and Bim. The PDB files of these three factors were obtained from www.rcsb.org and the 3D structure of curcumin was obtained from PubChem and converted to a PDB file by Avogadro v.1.2 software. Via GROMACS 2018, studies on molecular dynamic simulation were conducted in water and ion environment. Furthermore, AutoDock v.4.2.6 software performed the docking of curcumin as a ligand to these pre-apoptotic proteins. Also, LigPlot^{+} v.4.5.3 was used to determine the hydrogen and hydrophobic bonds at the binding sites. Curcumin with the highest tendency could bind to Bad by -6.58 kcal/mol of binding energy. In addition, the binding of curcumin to the other proteins induced some changes in molecular dynamic factors such as radius of gyration (Rg), root-mean-square deviation (RMSD), and root-mean-square fluctuation (RMSF). A significant increase in Rg and RMSD was seen in Bad and Bak after docking with curcumin. Moreover, curcumin induced some variations in RMSF and secondary structure in these factors. According to this study, curcumin could bind to Bad, Bak, and Bim directly, induce conformational changes, and increase their likelihood of dimerization that can lead to activating apoptotic pathways. These results confirmed the apoptotic effects of curcumin on cancer cells.

Key Words: *Bad, Bak, Bim, Curcumin, Molecular dynamics*

Generalized dissimilarity models as a tool to predict Odonata assemblages species composition

Zohreh Eslami Barzoki[1]*, Mehregan Ebrahimi[1´2], Saber Sadeghi[1]

[1] *Department of Biology, Faculty of Sciences, Shiraz University, Shiraz, Iran*
[2] *College of Science and Engineering, Flinders University, Adelaide, Australia*

E-mail address (corresponding author): eslami.zohre86@gmail.com

Abstract

Central Iran is considered one of the most water-stressed regions of the Central Plateau of Iran. In recent years a combination of climate change, population growth, and economic development is placing enormous stress on freshwater ecosystems of this area. Thus the quality and quantity of freshwater habitats are declining and freshwater biodiversity is threatened. However, information about the diversity patterns of freshwater species and drivers of these patterns is scarce. Therefore assessing the effects of anthropogenic environmental alterations would be challenging. One way to overcome this challenge is to use statistical and modelling techniques. By using these techniques we can estimate species assemblages across a less explored region by extrapolating the results of a few surveys conducted in that region. Results of such analysis will have the best chance of being effectively used in conservation decision-making if it is done for flagship taxa like Odonata. Generalized dissimilarity modelling (GDM) is a statistical approach that uses the pairwise dissimilarity of surveyed localities to fit a model predicting the biological distance between any given pair of localities as a multivariate non-linear function of differences in the environmental characteristics of those localities. In this study, we used data of 41 water bodies surveyed regarding Odonata species in Central Iran (Qom and Esfahan provinces). We perform GDM using the R package gdm to extrapolate species assemblage composition in areas where the species composition data was not available. GDM provided a map that clusters the predicted communities into a discrete set of communities with a common profile. It shows how Odonata communities are structured along the altitudinal gradient of Central Iran in terms of their species composition. We can use this map for priority conservation planning and to suggest the best distribution of potential locations for conducting future field surveys.

Key Words: *Arid and semi-arid; Dragonflies; GDM, Extrapolation; Ecological Modelling*

Comparison of the Salmonella enterica subsp. LT2 metabolic model in biochemical databases

Sajjad Gharaghani[1], Shiva Beigizadeh[2]

[1]Assistant Professor, Institute of Biochemistry and Biophysics (IBB), Department of Bioinformatics, Laboratory of Bioinformatics and Drug Design (LBD), University of Tehran
[2] Dept. of Biostatistics, Jahrom University of Medical Sciences, Jahrom, Iran

E-mail address (corresponding author): s.gharaghani@ut.ac.ir

Abstract

Metabolic models include the description of all biochemical reactions, metabolites, and metabolism genes for a particular organism, biochemical, genetic, and genomic knowledge. We analyzed five biochemical databases (Bigg, KEGG, SEED, PATRIC, and BIOCYC) commonly used for metabolic modeling and performed pairwise comparisons to examine compatibility, incompatibility, and ambiguity of reactions, metabolites, and genes between databases. Salmonella enterica subsp. LT2 that there's a metabolic model for it in the Bigg database. (ID BIGG: STM_v1_0) The SEED model is also the primary source of genomic-scale metabolic models based on microbial or plant annotation genomes. We created the metabolic model of our reference genome using the Model SEED web server and reconstructed it in the PATRIC metabolic database. (PATRIC Database ID: 99287.1) We compared the Bigg and Seed metabolic models for the number of reactions, metabolites, and genes. The Seed model had 1646 reactions, 1726 metabolites, and 958 genes, and the Bigg model had 2545 reactions, 1802 metabolites, and 1271 genes. There were 1935 reactions and 1547 metabolites in the Seed model that were not presented in the Bigg model, and conversely, there were 1073 reactions and 951 metabolites in the Bigg model that were not presented in the Seed model. In the SEED database, our organism had 566 subsystems and in the PATRIC database, it had 344 subsystems. To find functional proteins that are in the metabolic pathway but are not in the PATRIC database, we searched for organisms in the KEGG database.
Conclusion: Each database uses its own naming space; when different tools are used, the same metabolites and reactions may have different naming conventions. It is now recommended to use unique identifiers, independent of the specific databases used or referred to different namespaces. Also, manual verification of maps is a useful solution to eliminate inconsistencies while combining the models.

Key Words: *databases, incompatibility, name ambiguity, chemical nomenclature*

A New Analytical Model Based on Cellular Automata to Study the Influence of Various Drugs on Alzheimer's disease Progress

Niloofar Jafari[1], Faegheh Golabi[1]*, Yashar Sarbaz[2], Abbas Ebrahimi-kalan[3]

[1] *Department of Biomedical Engineering, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*
[2] *Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran*
[3] *Department of Neuroscience and Cognition, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*

E-mail address (corresponding author): faegheh.golabi@gmail.com

Abstract

In recent years, extensive research has been done to predict, treat, and recognition of Alzheimer's disease (AD). Among these scientific works, mathematical modeling of AD is an efficient way to study the influence of various parameters such as drugs on AD progression. This article aims to introduce and propose a novel mathematical model for the study of Alzheimer's disease (AD) progress. The presented model is based on Cellular Automata for better representation of AD progression. The differential equations of the Puri-Li model are utilized in this paper to calculate the number of Amyloid-β molecules. A new definition for AD rate is presented in this study. Moreover, other useful factors such as Critical Rate (CR) and Warning Rate (WR) are defined to determine the status of AD progression. To get exact insight into the neuron-to-neuron communications, the model is obtained for a 3×3 neuron system to investigate the influence of drug injection on the reduction of AR, CR, and WR factors. It will be shown that the usage of the strong and suitable drug can decrease AR and CR factors and also enhance the WR. The presented study can be utilized for the investigation of various factors in the control and treatment of AD progression.

Key Words: *Alzheimer disease; cellular automata; modelling; disease progress; Amyloid Beta*

Computational prediction of microRNAs' target genes in rat brain involved in morphine addiction

Maryam Koraei[1]*, Aso Mafakheri[2], Shamseddin Ahmadi[1], Sadegh Sulaimany[2]

[1] *Department of Biological Science, Faculty of Science, University of Kurdistan, Sanandaj, Iran*
[2] *Department of Computer Engineering, Faculty of Engineering, University of Kurdistan, Sanandaj, Iran*

E-mail address (corresponding author): maryamkoraei2016@gmail.com

Abstract

MicroRNAs are small non-coding RNAs with a pivotal role in morphine addiction and tolerance. However, little is known about specific microRNAs involved in regulating molecular mechanisms of addiction. Therefore, the identification of microRNAs targeting specific genes involved in addiction through computational methods could be very helpful in reducing the time and cost of laboratory methods. The aim of this study is to investigate the possible connections between microRNAs and genes involved in morphine addiction in rats by using bioinformatics tools. In this way, high probable communications that will be useful for purposeful future laboratory studies will be identified and proposed.
First, a list of microRNAs and genes involved in morphine addiction was collected by searching in NCBI database and 30 microRNAs and 144 related genes were selected. Then, links between the genes and the microRNAs involved in morphine addiction in rats were found using MicroRNA Target Prediction Database (www.mirdb.org). Thereafter, the desired connections were modeled in the form of a bipartite network, the relationships between microRNA and genes was graphed, and then a customized version of link prediction algorithms was set to find the most probable relations between the related microRNAs and target genes. Further, the accuracy of the results was calculated and confirmed by measuring the AUC. Finally, the 20 top relations predicted were reported in order of priority. new relation between rno-miR-1-3p, rno-miR-1b, rno-miR-9a-5b, rno-miR-16-5p, rno-miR-27a-3p, rno-miR-132-3p, rno-miR-19b-3p and rno-miR-212-3p recpectively with Ms4a7, Pi4ka, Creb1, Avpr1a, Cxcr5, Slain2, Creb1 and Tdrd7 are predicted. These computational findings are the most promising connections that are likely to exist but have not yet been reported in the online databases. They are the best choices to be validated using in vitro studies. Therefore, a proposed way is to perform related laboratory tests to confirm or disprove such predictions.

Key Words: *Morphine, Addiction, microRNA, Target gene, Link prediction, Rat*

Extraction of temperature distribution of high intensity focused ultrasound: Nonlinear propagation

Sare Mortazavi[1], Manijhe Mokhtari-Dizaji[1]*

[1] *Department of Medical Physics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): manijhem@yahoo.com

Abstract

Background: The aim of the study is the extraction of acoustic pressure distribution in the target tissue layers based on the nonlinear behavior of waves. The nonlinear behavior effect of high intensity focused ultrasound (HIFU) on the temperature distribution of the tissue was extracted and compared with the linear behavior. Materials and Methods: The acoustic pressure field was calculated using the Westervelt equation and was coupled with Pennes thermal transfer equation. The simulations were performed for three layers of skin, fat, and muscle using Comsol software. The disagreement between two linear and nonlinear models was analyzed with Kolmogorov–Smirnov test. The pressure and temperature distributions were calculated in nonlinear model by changing the acoustical parameters of the transducer including; intensity, effective radiation area, focal length and sonication time.

Results: Model results were validated with experimental results with 98% correlation coefficient ($p<0.05$). There are not significantly different between the pressure amplitude and temperature distribution in linear and nonlinear models at low intensity ($p>0.05$), but with increasing intensity to 10 W/cm2, in nonlinear model, maximum pressure and maximum temperature increased 40% and 20% compared with linear model. For input intensities of 1.5, 2, 8 and 10 W/cm2, the maximum pressure (at focal point) increased 10, 12, 22, 40% and maximum temperature increased 1, 2,12,20% in nonlinear model compared to linear model. At 0.8 and 1.5 cm2 effective radiation area, the maximum temperature in nonlinear model increased from 43 to 79 °C. By decreasing the focal lengths from 10 to 7.5 mm, the maximum temperature increased from 45 to 87 °C.

Conclusion: It is concluded a change in the input parameters of the transducer, it can be very effective in treating. The results emphasize the effects of nonlinear propagation and acoustical radiation parameters to improve the HIFU treatment.

Key Words: *Simulation; Ultrasound; Nonlinear; Thermo-viscous Effect*

The potential of antimicrobial peptides as Covid-19 therapeutics using computational methods

Mortezaali Sara[1], Mahboobeh Zarrabi[1]

[1] *Department of Biotechnology, Biological Faculty, Alzahra University, Tehran, Iran*

E-mail address (corresponding author): sarah.mortezaali@gmail.com

Abstract

Background: A new type of coronavirus was first detected in December 2019 at Wuhan city, China. The pneumonia-like disease caused by the virus is globally known as COVID-19. Apart from China, COVID-19 has spread to 213 countries and killed over 5,735,179 people in total as of today (7 February 2022). With the case count and death toll rising each day, there is an urgent need for antiviral drugs against SARS-CoV-2. Antimicrobial peptides (AMPs) have been widely regarded as a promising solution to combat harmful microorganisms. They are biologically active molecules produced by different organisms as an essential component of their innate immune response against invading pathogens.

Materials and Methods: 110 antimicrobial peptides with a length of 5 to 50 amino acids have been chosen from DBAASP and the third structure of proteins have modeled by pepfold3 server. Then peptides have been refined by MODELLER. The structure of SARS_CoV_2 spike protein has been obtained from protein data bank (PDBID:6lzg). Spike's ligand has been deleted by Chimera and non-polar hydrogens have been merged by autodock tools.

Finally, peptide files and receptor files have been loaded on pyrx, and docking was carried out for the spike protein with the grid box set to center x-, y-, z-axis values of −40.34, 30.85, 6.61 with a dimension of 35 Å x 50 Å x 30 Å.

Result: For binding affinity, the best result was the 15288-spike complex with -11.5 Kj/mol. Then this complex was subsequently analyzed by MD simulation. RMSD, RMSF, and potential energy analysis were performed by gromacs.

Conclusion: Our results and studies show that a number of peptides in this library have the ability to bind to the spike protein. This interaction has the potential to disturb the function of the virus and these peptides can be used as a therapeutic agent.

Key Words: *Molecular Dynamic; Docking; Antimicrobial peptide; Coronavirus; Spike protein*

The effect of Phenol Hydroxylase enzyme on some phenolic compounds by molecular docking and simulation study

Mohamadhossein Najafi-Chaleshtori[1], Hadi Zarei Mahmoud Abadi[2], Javad Saffari-Chaleshtori[3]*

[1] Faculty of Agriculture and Environment, Islamic Azad University, Meybod, Yazd,Iran
[2] Department of Environment, College of Agriculture and Natural Resources, Islamic Azad University of
[3] Clinical Biochemistry Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran

E-mail address (corresponding author): j_saffari@yahoo.com

Abstract

Pseudomonas phenol hydroxylase is a bacterial enzyme with hydroxylase activity that can add the hydroxyl group to chemical compounds and increases their water solubility. This in-silico study investigated the effect of pseudomonas phenol hydroxylase enzyme on some phenolic compounds by simulation and molecular docking study. In this study, after obtaining the FASTA file sequencing of phenol hydroxylase from Uniprot server, the production of its PDB file was performed by i-tasser server. Then the confirmation of 3D structure was performed by Ramachandran Plot Server. Also, the structure of phenolic compounds were obtained from pubchem server and converted to PDB files by Avogadro V1.2 software. The molecular dynamic of phenol hydroxylase was done by Gromacs 2021 software. Then the molecular docking of phenolic compounds on phenol hydroxylase enzyme was done by Autodock v.4.2 software after stabilization of RMSD parameter. The ramachandran plot showed that more than 98% of total residues are in preferred and highly preferred regions. According to RMSD results, the simulation system was stabilized after 50 nanoseconds of simulation time. The results of docking study showd that some phenolic compounds such as Bisphenol(-7.17 kj/mol), Dinoseb(-6.20 kj/mol), 4-Nonylphenol(-6.03 kj/mol), 4-Octylphenol(-5.93 kj/mol), 4-Chlorophenol(-4.92 kj/mol), Catechol(-4.23 kj/mol), Phenole(-4.18 kj/mol) have the high tendency to interact with phenol hydroxylase enzyme. These interactions occurred via some hydrogen and hydrophobic bounds at the binding site of phenol hydroxylase. This simulation and molecular docking study showed that phenolic compounds have the high tendency to interact with Pseudomonas phenol hydroxylase enzyme. These results confirm that this enzyme can be effective on phenolic compounds.

Key Words: *Docking, Phenol Hydroxylase, Phenolic compounds, Simulation*

Presentation of a mathematical model for describing and predicting the release of methotrexate and sodium valproate from polymer nanogels using finite difference method

Pedram Porbaha[1], Amir Azadi[1], Ali Akbar Roosta[2]

[1] *Department of Pharmaceutics, School of Pharmacy, Shiraz University of Medical Sciences*
[2] *Department of Chemical Engineering, Faculty of Chemical Engineering, Shiraz University of Technology, Shiraz, Iran*

E-mail address (corresponding author): p.porbaha@gmail.com

Abstract

In recent years, nanotechnology have gained a great deal of connection with drug delivery systems. One of the biggest challenges facing scientists in this field is the development and optimization of sustained-release systems. One of the biggest challenges facing scientists in this field is the development model for predicting the Release of the Matrix sustained-released systems. a differential mechanistic model for the drug loaded in the chitosan nanogels using MATLAB software has been presented and an attempt has been made to avoid statistical methods that require costly experiments as much as possible. Then this model was tested on two series of experimental data of nanogels containing sodium valproate and methotrexate and compared with other models in terms of R2 parameters and mean absolute error.It was observed that the proposed model for nanogels containing methotrexate has R2 = 0.958 and the mean absolute error is equal to 10.03% and for nanogels containing sodium valproate has R2 = 0.936 and the mean absolute error is equal to 10.53% and compared to other existing models, was among the models with higher accuracy.In this study, a mechanistic model based on differential relations was obtained and this model was solved in both 1- minimizing liquid concentration and 2- varying liquid concentration and then comparing the results with eight other experimental models. The result showed that the model in the variable state considering the concentration of liquid compared to other models was among the best models in terms of R2 and absolute error percentage.

Key Words: *Mathematical Modeling; Sustained-Released Drug Delivery Systems; Hydrogel; Finite Difference Method; Nanotechnology*

Structural Variation Detection from Paired-end NGS data using Hidden Markov Model

Muhammad Amin Rahimi[1]*

[1] *Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): m.aminrahimi91@gmail.com

Abstract

The objective of the present study was to identify the areas variated in the samples' genomes, which was achieved by the Hidden Markov Model. The single-read data of IIlumina technology displays the exact same correlation with Array-CGH and almost the same algorithms can be applied. In this study, Hidden Markov Models were used for the more precise and abundant paired-end data, which is quite unusual for this type of algorithm. For this purpose, two methods of identification were used [2, 3]: 1) in the first method, using a specific threshold, the ratios compared with the normal samples were extracted and after the labelling of variated areas, the Hidden Markov Model was applied, 2) the second method utilized the ground truth data and SVM machine leaning technique to label variated areas. The Hidden Markov Model was then applied for re-labelling of variated areas. Finally, for evaluation of the model, artificial data were acquired using simulation techniques. After the identification of variated areas by Hidden Markov Model, the percentage of the found duplication, deletion and translocations were calculated. The novelty of this study lies in the identification of structural variations in paired-end data by Hidden Markov Model whereas, in previous studies single-read data were used. Furthermore, this study identifies translocations using Hidden Markov Model for the first time.

Key Words: *Structural Variation; Hidden Markov Model; Support Vector Machine; Next Generation Sequencing.*

Comparative molecular dynamics study of receptor-binding domains in wild type and Omicron (B.1.1.529) variant of SARS-CoV-2

Negin Safaei Hashkavaei[1] and Fatemeh Bayan[1] and Yahya Sefidbakht[1]*

[1] *Protein Research Center at Shahid Beheshti University, Tehran, Iran*

E-mail address (corresponding author): y_sefidbakht@sbu.ac.ir

Abstract

The new variant of SARS-CoV-2, Omicron (B.1.1.529), first reported in November 2021 is currently infecting people around the world. Since the spike glycoprotein plays a key role in the early events in viral replication through a receptor binding domain (RBD) which bindes to angiotensin-converting enzyme 2 (ACE2), it is important to study the effects of mutations on the structure of receptor binding domain (RBD). Based on Structural analysis, 15 substitution mutations were identified in RBD of this new variant. In this paper, we used computational comparison of receptor-binding domain (RBD) in wild-type (WT) SARS-CoV-2 and Omicron variant of concern (VOC) in free forms and structural models were built using the I-TASSER server. Based on our molecular dynamics simulation results, mutations can affect the RBD structure and in this new variant of concern, RBD structure is more open than the wild-type (WT) RBD. The results of this study could help to achieve a better understanding for researchers, and be effective in the development of novel therapies.

Key Words: *Receptor-binding domain (RBD); SARS-CoV-2; Omicron (B.1.1.529); Molecular dynamics simulations.*

Molecular docking and molecular dynamic simulation of the effect of Pseudomonas phenol hydroxylase enzyme on polycyclic aromatic hydrocarbons (PAHs)

Javad Saffari-Chaleshtori[1]*, Gashtasb Mardani[2], Majid Validi[2]

[1] *Clinical Biochemistry Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran*
[2] *Medical Plants Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran*

E-mail address (corresponding author): j_saffari@yahoo.com

Abstract

Polycyclic aromatic hydrocarbons (PAHs) are hydrocarbon compounds that are composed of multiple aromatic rings. PAHs are found in coal and oil deposits and are also produced by the combustion of organic matter for example, in engines and incinerators or when biomass burns in forest fires. PAHs are one of the most significant toxins in carcinogenesis and it is important to find the best way to remove these compounds from the environment. This study investigated the effect of pseudomonas phenol hydroxylase enzyme on PAHs by molecular dynamic and simulation study. The FASTA file sequencing of phenol hydroxylase was obtained from the Uniprot server and its PDB file was performed from the i-tasser server. Then Ramachandran plot was performed from Ramachandran Plot Server. Also, the structure of PAH compounds were obtained from the PubChem server and converted to PDB files by Avogadro V1.2 software. The molecular docking of PAH compounds on phenol hydroxylase was done by Autodock v.4.2 software and the Gromacs 2021 software was used for molecular dynamics studies. The results of this study showed that the PAH compounds have a high affinity to interact with phenol hydroxylase by hydrogen and hydrophobic bonds. Among them, the Dibenzo Anthracene, Benzo Fluoranthene, and Indeno Pyrene were more potent to interact with phenol hydroxylase. Binding of Dibenzo Anthracene, Benzo Fluoranthene, and Indeno Pyrene to phenol hydroxylase binding side can induce the conformational changes on enzyme structure and induce the change on RMSD, RMSF, RG, and other molecular dynamic parameters. According to this simulation study, phenol hydroxylase enzyme can affect PAH compounds so the bacteria with enzymes such as Pseudomonas can be used in contaminated environments by PAH compounds.

Key Words: *Docking, Phenol Hydroxylase, Phenolic compounds, Simulation*

Reaction-diffusion modeling suggests a novel amplification mechanism for the extremely-low frequency magnetic field bioeffects

Amirali Zandieh[1], Abbas Ali Ravassipour[1], Seyed Peyman Shriatpanahi[1]*, Bahran Goliaei[1]

[1] *Department of Biophysics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): pshariatpanahi@ut.ac.ir

Abstract

There have been a growing number of researches indicating the role of reactive oxygen species (ROS) as a mediating agent for the observed effect of extremely low-frequency electromagnetic field (ELF-EMF) on living organisms. Here we propose a mechanism to explain how interactions with energy dozens of magnitudes below kBT are not being masked by thermal noise.

Grounded on the previously discovered phenomenon of "Radical Pair Mechanism", first, we introduce a scheme that suggests how applying a low-intensity magnetic field of the order of a few tens of milliTeslas can alter the superoxide production rate at Qo site of mitochondrial cytochrome bc1. Our proposal is well-backed by a recent experimental finding that indicates changes in mitochondrial electron transport chain ROS production under the effect of magnetic fields.

Next, using a reaction-diffusion model previously developed for the simulation of the observed oscillation of ROS level in myocytes, we show how the mentioned superoxide production change can raise the whole cellular ROS level via amplification. Mitochondrial ROS production level is generally escalated in cancerous cells through mutations which in turn can contribute to the transformation of healthy cells into tumors. The results of the model reveal how an alternating magnetic field can amplify the change in superoxide production in such cells via two modes: Either by inducing a whole cellular ROS oscillation in the case of the cells with borderline mitochondrial ROS level or by synchronizing the mitochondria in the cells with asynchronous mitochondrial ROS oscillation through a resonance effect.

Finally, the proposed model is in good agreement with our recent experimental fluorescence microscopy study in which we observed frequency-dependent changes in both mitochondrial membrane potential and ROS level under the effect of alternating magnetic fields. Also, the observed bimodal responses of the different cell lines provide further support for our novel mechanism.

Key Words: *Reaction-diffusion Modeling; Amplification; Reactive Oxygen Species; ELF Magnetic Field; Radical Pair Mechanism*

Structural Bioinformatics

A Computational Approach to Modeling a Thermophilic Psychrophilic Protease Fusion Protein for Laundry Detergent

Atena Abed[1], Ladan Mafakher[2], Elham rismani[3], Hamed Mirzaei[4]

[1] *Cellular and Molecular Research Center, Shahrekord University of Medical Sciences, Shahrekord, Iran*
[2] *Thalassemia and Hemoglobinopathy Research Center, Health Research Institute, Ahvaz Jundishapur University of Medical Science, Ahvaz, Iran*
[3] *Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran*
[4] *Research Center for Biochemistry and Nutrition in Metabolic Diseases, Institute for Basic Sciences, Kashan University of Medical Sciences, Kashan, Iran*

E-mail address (corresponding author): abed.atena@yahoo.com

Abstract

Protease is one of the main enzyme which is used in laundry industry. As the thermal condition in lanundry industry is varied from high to low temperature, having protease enzyme which could activate in broad thermal range has great potential. One of the approach in this field is fusion two proteins with different thermal activity. For doing this in current work by computational method, Thermolysin protease an extracellular metalloendopeptidase from a gram-positive thermophilic bacterium Bacillus thermoproteolyticus which activate in thermal range from 25 to 88°C, was joined by a four repeat of G4S as a flexible linker with a psychrophilic protease from a marine bacterium Flavobacterium YS-80 with thermal range from 0°C to 30°C, to produce protease enzyme which activate in broad thermal range from 0 to 88° C.
Material and methods:
As the whole structure of thermolysin did not exist in PDB bank, homology modeling by Lomet webserver was performed and Modeller package was applied to construct whole protein fusion structure. The quality of model structure was verified by Procheck, Verify3D, ERRAT, ProSA webserver. The secondary structure analysis was done by PSIPRED. The solubility and physiochemical properties of fusion protein was determined by Protein-sol and Protparam webserver. Molecular Dynamics simulation was perfomed by Gromacs package to assess stability of fusion protein structure in simulated condition.
Result and Discussion:
The quality structure assessment of fusion protein showed that this protein modelled with high quality compared to non-fusion form. Also, secondary structure analysis revealed that no secondary structure changes was observed in fusion form compared to non-fusion form. The physicochemical properties of fusion protein figured out this structure had stable structure and good solubility in water. the Molecular dynamic simulation result discovered that no drastic structure changes was happened during simulation time.

Key Words: *protease, protein engineering, homology modeling, molecular dynamics simulation*

Harnessing Bioinformatic Approaches to Designing a Multi-epitope Peptide Vaccine Against COVID-19

Tooba Ahmadi[2], Vajihe Ghorbanzade[1], Hassan Dariushnejad[1,2]*

[1] *Cardiovascular Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran*
[2] *Department of Medical Biotechnology, Faculty of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran*

E-mail address (corresponding author): toobaahmadi98@yahoo.com

Abstract

Introduction: SARS-CoV-2 is the causative agent of Coronavirus 2019 or COVID-19 in the world. Novel coronavirus disease is a respiratory disease. To date, there have been challenges in the treatment for COVID-19 and emerged new variants. Accordingly, an effective prevention regime is needed for this infection, which covers most variants. The purpose of this research was to predict the conserved epitopes of Spike and Nucleocapsid proteins from SARS-CoV-2 for the design of a novel coronavirus 2019 multi-epitope vaccine using in silico tools.

Methods: Computational analysis and immunoinformatics approaches include identification of potential conserve epitopes and selection of epitopes based on allergenicity, toxicity, antigenicity, and molecular docking were used for epitope prediction and screening. In the next step, selected segments of the epitopes were attached by the suitable linkers. Finally, Maltese-bound protein (MBP) as an adjuvant was added to the novel vaccine structure. The secondary and third structures of the designed multi-epitope vaccine were predicted via immunoinformatics algorithms. Predicted structure refined and validated for attaining best stability. In the end, immunoinformatics evaluation, molecular docking, and molecular dynamics were performed to confirm vaccine efficiency. Codon optimization and in silico cloning were done to ensure the expression yield of the novel multi-epitope vaccine in the target host.

Results: This study showed that our data support the suggestion that the designed vaccine could induce immune responses against SARS-CoV-2 variants.

Conclusion: The Structure designed had acceptable quality with software reviews. Further in vitro and in vivo experiments.

Key Words: *COVID-19, Immunoinformatics, SARS-CoV-2, multi-epitope vaccine, bioinformatics*

Investigation of SDCCAG8 gene in COLON ADENOCARCINOMA and the possibility of rs769698229 in hsa-mir-450b-5p

Aliasghar Akbarinia[1], Zahra Maravandi[1]

[1]*Department of Biology, Faculty of Science, Naghshejahan Higher Education Institute, Isfahan, Iran*

E-mail address (corresponding author): aliasghar.akbarinia@naghshejahan.ac.ir

Abstract

Introduction: Colon cancer is a type of cancer that starts in the colon or rectum.The large intestine forms the organs of the lower gastrointestinal tract.There is more than one type of colon cancer.There are cells in different parts of the gastrointestinal tract that can lead to colon cancer.The most common type of colon cancer starts with adenocarcinoma.Adenocarcinomas form inside the cells that make up the lining of the large intestine or rectum.Cancer may be caused by genetic mutations that can be inherited or acquired.These mutations are not a guarantee of colon cancer, but they do increase the risk.The SDCCAG8 gene encodes a centrosome-related protein.This protein may be involved in the organization of the centrosome during the interphase phase,and mitotic mutations in this gene are associated with renal retinal ciliopathy.
Materials and methods: At the NCBI site, the SDCCAG8 gene (involved in interphase and mitosis) was selected(2).At the GEPIA2 site,the expression of this gene was examined,and at the MIRWALK site, the microRNAs that acted on this gene were studied,and at LNCRNASNP2, this microRNA was examined for expression in COLON ADENOCARCINOMA and finally the SNPs affecting this microRNA were examined at the MIRNASNP site.
Discussion: SDCCAG8 gene is more expressed in tumor patients.MIRNA (has-mir-450b-5p) that was selected had a score: 0.85 and np:16, which has a great effect on the SDCCAG8 gene.This MIRNA had an expression of 7.18 in COLON ADENOCARCINOMA.And many SNPs affect this MIRNA.
Conclusion: SNPs are high-density natural sequence variations in genomes and are considered as a major genetic source of phenotypic variation within a species and are considered important genetic markers. Based on previous research, it is predicted that SNP rs769698229 will be created between a gene or MIRNA and cause a functional disorder.

Key Words: *Colon, cancer, lnc, snp, gene*

Co-Expression Analysis for Identification of Critical Genes in Pancreatic Cancer

Sahar Akrami*[1], Ali Niazi[1], Ahmad Tahmasebi[1], Amin Ramezani[2,3], Abbas Alemzadeh[4], Ali Moghadam[1]

[1] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*
[2] *Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran*
[3] *Shiraz Institute for Cancer Research, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran*
[4] *Department of Crop Production and Plant Breeding, School of Agriculture, Shiraz University, Shiraz, Iran*

E-mail address (corresponding author): akrami.sahar.19@gmail.com

Abstract

Introduction: Pancreatic ductal adenocarcinoma (PDAC) is known as king of carcinoma. PDAC is a distinctly aggressive cancer, with a 5-year survival rate of <1.0%. Considerable efforts have been made to identify potential PDAC biomarkers that may be used to develop anti-metastatic treatments and improve prognostic evaluation. Novel biomarkers for PDAC are urgently needed because of its poor prognosis. Recently, computational analyses using high-throughput expression data have helped to recognize putative molecular mechanisms involved in various cancers. Despite the importance of differentially expressed genes (DEGs) identification, this strategy mostly focuses on the discovery of gene contents and suffers from exploring relationships among genes. Co-expression network analyses allow us to apply a system-level view of gene-gene connections. In this study, we carried out both methods of finding the DEGs and also network analysis by co-expression method, to find the possible biomarker candidates for PDAC. Method: Transcriptomics data of normal pancreatic tissue and pancreatic cancer, of pancreatic cancer, were retrieved from TCGA database. To normalize data and identify the differentially expressed genes (DEGs), the edgeR package was used with an FDR of ≤0.01 and a |fold change| ≥1. Co-expression method was applied with Hmisc package in R based on Pearson correlation. The network was visualized with Cytoscape, and finally with the help of CytoNCA plug-in, hub genes were topologically identified. Result: A total of 798 DEGs which were differentially expressed between pancreatic cancer and normal tissues were found. The network constructed by the co-expression analysis showed 890 nodes and 108742 edges. PCNA, CD49B, CEP250-AS1, MTOR_Ps2448 and PI3KP110ALPHA with top node degrees were selected as the hub genes. Generally these genes are involved in mismatch repair, base excision repair, DNA replication and cell cycle. Conclusion: The analysis suggests that these genes may be potential diagnostic biomarkers and/or therapeutic molecular targets in patients with PDAC.

Key Words: *Co-expression analysis; Pancreatic Cancer; Biomarker; Network Analysis*

Identification of Key long noncoding RNAs-related to Pancreatic Cancer Using Network Analysis

Sahar Akrami[1]*, Ali Niazi[1], Ahmad Tahmasebi[1], Amin Ramezani[2,3], Abbas Alemzadeh[4], Ali Moghadam[1]

[1] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*
[2] *Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran*
[3] *Shiraz Institute for Cancer Research, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran*
[4] *Department of Crop Production and Plant Breeding, School of Agriculture, Shiraz University, Shiraz, Iran*

E-mail address (corresponding author): akrami.sahar.19@gmail.com

Abstract

Introduction: Pancreatic ductal adenocarcinoma (PDAC) is one of the most malignant tumors that threaten human health. The molecular mechanisms underlying PDAC still remain unclear. Unreasonable excessive mortality and low survival rates for this disease, mainly result from the delay in diagnosis and treatment. Long non-coding RNAs (lncRNAs), a class of transcripts ≥200 nucleotides, have been proved to regulate various biological processes including apoptosis, invasion, metastasis and angiogenesis through interactions with miRNAs or mRNAs in different cancer types. In this work, we find important lncRNAs involved in PDAC, which are identified by mining The Cancer Genome Atlas (TCGA) PDAC RNA-sequencing differentially expressed data between cancer and normal state and visualization of the network by co-expression method.
Method: The RNA-seq expression data of PDAC for cancerous and normal condition retrieved from TCGA database. The lncRNA data was extracted via Biomart tool in Ensemble. To identify the differential expression of lncRNA, the edgeR package was used, with the standard thresholds of |fold change| ≥1 and FDR of ≤0.01. The network of lncRNA-gene was constructed based on the Pearson correlation. Finally, CytoNCA plug-in was used to screen hubs of the network. GO and KEGG pathway analyses were performed, by gProfiler database, to determine the significantly enriched functions and pathways of these lncRNAs in PDAC.
Result: We detected 592 mRNAs and 206 lncRNAs that were differentially expressed. After constructing the co-expression network of the lncRNA-mRNA, a total of 5 lncRNAs were found, which includes MIR600HG, C9orf139, LINC01410, IRF1-AS1, and BTG1-DT. GO showed the crucial roles of them in immune system process, immune response, leukocyte activation, cell activation. Also KEGG analysis demonstrated enrichment in natural killer cell mediated cytotoxicity and B cell receptor signaling pathway.
Conclusion: Our findings uncovered that these lncRNAs may be used as diagnostic indicators and prognostic factors in PC patients.

Key Words: *Pancreatic Cancer; lncRNA; Network analysis; TCGA data*

In silico design of a multi-epitope vaccine candidate against SARS-CoV-2

Abbas Alibakhshi[1]*, Armina Alagheband Bahrami[2], Elmira Mohammadi[3,4], Shahrzad Ahangarzadeh[5], Meysam Mobasheri[6]

[1] *Molecular Medicine Research Center, Hamadan University of Medical Sciences, Hamadan, Iran*
[2] *Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Science, Tehran, Iran*
[3] *Core Research Facilities, Isfahan University of Medical Sciences, Isfahan, Iran*
[4] *Applied Physiology Research Center, Cardiovascular Research Institute, Isfahan University of Medical Sciences, Isfahan, Iran*
[5] *Infectious Diseases and Tropical Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*
[6] *Department of Biotechnology, Faculty of Advanced Sciences and Technology, Tehran Islamic Azad University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): alibakhshi2630@yahoo.com

Abstract

Frequent and sometimes more dangerous mutations in the SARS-CoV-2 indicate that a stronger strategy is needed to produce an effective vaccine. A vaccine that contains a wider range of virus factors and remains effective if one or more mutations is happen in a part of the genome. In this study, four important virus proteins, S, E, M and Mpro were used to make a multi-epitope protein vaccine. For this purpose, the sequences were retrieved from NCBI gene database and then antigenic determinant of 4 proteins using IEDB, MHCpred, ProPred, RANKPEP, ABCpred, BCpreds, BepiPred, Ellipro servers were selected and a protein structure was designed using 4 domains containing epitopes. After examining its antigenic and allergenicity potential usingVaxiJen v2.0 server AllerTOP v2.0 server respectively, its three-dimensional structure was designed by two servers, Robetta and I-TASSER and then it was docked with immune system receptors using Autodock vina program embedded in the PyRx software. Finally, using the dynamic molecular method by Gromacs package 5.1.1, complexes and interactions were investigated and their interaction energies were measured. The results of the study showed that the designed structure has good relative stability and interacts well with its receptors and can be used as a vaccine candidate for further studies.

Key Words: *SARS-CoV-2; Covid-19; Multi-epitope; Vaccine; Molecular docking; Molecular dynamics*

Computational saturation mutagenesis to predict the effects of systematic mutations on the stability and binding affinity of nonstructural proteins 7 and 8 in SARS-CoV-2

Fatemeh Arabi Jeshvaghani[1], Fatemeh Javadi-Zarnaghi[1]*

[1] *Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran*

E-mail address (corresponding author): f.arabi68@yahoo.com

Abstract

The replication of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is mediated by protein-protein interactions of nonstructural proteins (nsps) and nsp-viral RNAs. The central component of replication machinery in SARS-CoV-2 is nsp12 with RNA-dependent RNA polymerase (RdRp) activity that catalyzes the synthesis of viral RNAs. Nsp7 and nsp8 as cofactors of RdRp play vital roles in stimulating the polymerase activity of RdRp and promoting its processivity. In this study, we analyzed the effects of all possible mutations (a total of 3724 mutations) generated by the computational saturation mutagenesis of all residues in nsp7 and nsp8 proteins of SARS-CoV-2 to all other 19 residues on the stability of proteins and protein-protein binding affinity. mCSM-PPI2 server was used to predict the effects of missense mutations on nsp7-nsp8 binding affinity. The impacts of the generated systematic mutations on stability and flexibility of investigated proteins were predicted by structure-based prediction tools including DynaMut, mCSM, SDM, and CUPSAT servers as well as PROVEAN and I-Mutant 2.0 as sequence-based prediction tools. ConSurf was used for analyzing the evolutionary conservation of the residues. A significant majority of mutations have the potential to destabilize the interaction of nsp7 and nsp8. K2P, M52D, and L56G mutations in nsp7 and L108D, I111D, and L96D mutations in nsp8 were predicted to significantly decrease the nsp7-nsp8 binding affinity. Mutations in highly conserved interface residues of nsp7-nsp8 were predicted to considerably decrease the stability and flexibility of proteins. This study provides comprehensive insights into the consequences of mutations in nsp7 and nsp8 proteins with importance in antiviral design and development for COVID-19 or future possible outbreaks related to coronaviruses.

Key Words: *SARS-CoV-2; nsp7-nsp8; Computational Saturation mutagenesis; COVID-19; Missense mutation*

A new mechanism in potassium channel blockage identified for a scorpion venom peptide

Masoumeh Baradaran[1]*

[1] *Toxicology Research Center, Medical Basic Sciences Research Institute, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran*

E-mail address (corresponding author): mb.baradaran@gmail.com

Abstract

Background: Scorpion venom is a source of ion channel modifier peptides with interesting pharmacological properties. Here, in addition to reporting the structure of a peptide(meuK2-2), identified in the Mesobuthus eupeus venom gland, its interaction with Kv1.3 channel was also interpreted.
Methods: Three-dimensional structure of meuK2-2 was generated using MODELLER and I-TASSER, PHYRE2, Robetta servers. The best model was selected according to Z-score;and considered for further optimization using MD simulation.Interaction of meuK2-2 with the Kv1.3 channel was also evaluated using peptide-protein docking experiments with HADDOCK software, subsequently pose clustering and also, 100 ns MD simulations using a protein-water system in the NVT ensemble by Gromacs to evaluate the binding interaction between meuK2-2 and Kv1.3.The final docked complexes were then subject to minimization with CHARMM force field and investigated key interacting residues, electrostatic interactions, binding free energies, folding pattern, hydrogen bond formation, hydrophobic contacts.
Results: 3-D structure of the meuK2-2 is composed of a cysteine-stabilized α-helical and β-sheet(CSα/β) folding.Two key residues and H-bonds are involved in the binding of meuK2-2 to Kv1.3.In a new mechanism meuK2-2 binds to both turret and pore loop of the channel.The binding of meuK2-2 induces some conformational changes to Kv1.3. This is followed by occupation of the pore of the channel with the side chain of a His9 residue.Altogether blocks the ion permeation pathway.
Conclusions: A new mechanism was predicted for channel blocking with meuK2-2.Since Kv1.3 plays a significant role in human T cell activation, meuK2-2 have a potential for further investigations to develop as a pharmacological tool in treatment of autoimmune diseases.

Key Words: *Mesobuthus eupeus; Venominvformatics; Potassium channel blocker; Homology modeling; Natural peptide*

Polymorphisms effects on structure and function of OeACP2 and OeACP3 proteins

M.Beizaei nejad[1], T.Jadidi[1], V.Shariati[1], Z.Minoochehr[1], SM.Houseini Mazinani[1], K.Razavi[1]*

[1] *National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran*

E-mail address (corresponding author): m720906@gmail.com

Abstract

Olive oil has fundamental positive influences on human health and prevent cardiovascular diseases. Fatty acid biosynthesis (FAS) pathways produce and store lipids in mesocarp of olive (Olea europaea L.). An important cofactor of fatty acid biosynthesis is acyl carrier protein (ACP) that acts as a shuttle. It makes covalent binds with fatty acyl intermediates and moves them from previous enzyme to next enzyme of the pathway. So mutations in ACPs may affect and disrupt fatty acid biosynthesis and oil producing. ACP has three loci in plants that each of them has two alleles. OeACP2 and OeACP3 have more crucial role and largely found in olive fruit and leave. Hence, we extracted DNA from two Iranian olive cultivars, Shenge and Mari, with low and high quality oil respectively. Then whole genome next generation sequencing (NGS) was done and mutations in these genes were found. We identified single nucleotide polymorphisms (SNPs), insertions and deletions (InDels) in DNA sequences and predicted tertiary structure of proteins with online servers such as I-TASSER and Robetta by homology modeling and ab initio methods. Finally we found some effective polymorphisms that seem to impact on structure and function of acyl carrier proteins.

Key Words: *olive oil; FAS; ACP; polymorphism*

The Binding Assessment with Human Serum Albumin of Binuclear Ni(II) and Cu(II) Complexes Containing Bidentate 8-Hydroxyquinoline Ligand

Mohammad Reza Binaeizadeh[1], Sudabeh Shokrollahi[1], Ahmad Amiri[1]*

[1] *Department of Chemistry, College of Science, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): ahmadamiri@ut.ac.ir

Abstract

The study of human serum albumin (HSA) binding in biophysical and biochemical studies is important since it has a well-known primary structure and it has been associated with the binding of many different categories of small molecules. Metal ions bind to proteins via coordinate bond formation with amino acid side chains. The use of metal complexes allows the formation of targeted protein-bound adducts that show binding via coordination or noncovalent interactions. Herein, two Ni(II) and Cu(II) complexes of the general type [Ni2(Q)6] and [Cu2(Q)4] (Q = 8-hydroxyquinoline) were synthesised and characterised by FT-IR, UV–Vis and 1H-NMR. The optimized structures and frontier orbitals properties were computed using DFT/B3LYP method and LANL2DZ and 6–31G** as basis sets. The electronic structure and spectral characterization of the synthesized complexes have been investigated using the density functional theory (DFT). The interaction of these complexes with protein (HSA) was followed via fluorescence spectroscopy, circular dichroism (CD) and molecular docking simulation. The observation of remarkable quenching in the emission spectrum of HSA upon titration with the complexes with an excellent binding constant (Kb= 93.88 mM-1 for [Ni2(Q)6]) reveals the strong interaction between the complexes and HSA. Furthermore, the Kq values of the fluorescence quenching are greater than 2.0 × 1010 M−1 s−1, establishing the fact that the fluorescence quenching is static. Molecular docking results revealed these complexes bind with HSA, and the binding site was positioned in Sudlow Site II of HSA (subdomain IIB).

Key Words: *8-hydroxyQuinoline; HSA Binding; Molecular Docking; DFT; Circular Dichroism.*

Identifying Thermostability Characteristics of Family A DNA Polymerases

Seddigheh Borhani[1], Seyed Shahriar Arab[1]*

[1] *Biophysics Department, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): sahar_borhani@hotmail.com

Abstract

DNA polymerases create complimentary DNA strand in living cells and are crucial to genome transmission and maintenance. All of these enzymes possess similar human right-handed fold which contain of thumb, fingers and palm subdomains and contribute to polymerization activities. These enzymes classified to 7 evolutionary families A, B, C, D, X, Y and RT based on amino acid sequences analysis and biochemical characteristics. Family A DNA polymerases exist on extended range of organisms included mesophilic, thermophilic and hyper thermophilic bacteria, participate in DNA replication and repair and have broad application in molecular biology and biotechnology. In this study, we attempt to detect factors play a role in thermostability properties of this family member despite their remarkable similarities in structure and function. For this purpose, similarities and differences in amino acid sequences, structure and dynamics of these enzymes have inspected. Our results demonstrated that thermophilic and hyper thermophilic enzymes have more charged, aromatic and polar residues than mesophilic ones, and consequently show further electrostatic and cation-pi interactions. In addition, in thermophilic enzymes, Tryptophan, Histidine and aliphatic residues tend to position in buried states more than mesophilic enzymes. These residues within their aliphatic parts increase hydrophobic core packing and therefore enhance thermostability of these enzymes. Moreover, molecular dynamic simulation results reveal that increasing of temperature impact mesophilic enzymes further than thermophilic ones, showing as rise in structural fluctuations and flexibilities.

Key Words: *Family A DNA Polymerases; Thermostability; Thermophilic Enzymes; Molecular Dynamic Simulations.*

Bioinformatics Analysis of the Expression of Key Genes Involved in Biosynthesis Pathway of Carvacrol of White Savory Essential Oil Under Drought Stress

Fatemeh Davari Shalamzari[1]*, Leyla Nejadsadeghi[1], Khosro Mehdikhanlou[1], Qing-Yan Shu[2]

[1] Department of Plant Production and Genetics, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[2] Key Laboratory of Plant Resources/Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing, China

E-mail address (corresponding author): f.davari1995@gmail.com

Abstract

White savory (Satureja mutica Fisch. & CAMey) is a member of the mint family that is extended in the eastern and western of Iran. This plant has antioxidant and antimicrobial properties. In this study, White Savory was planted in two conditions of drought control and treatment at 30% of field capacity in the Faculty of Agriculture, Shahid Chamran University. Using the KEGG database, the biosynthetic pathway and genes involved in the production of this pathway were determined. The active ingredient in White Savory is Carvacrol, in which three key genes including: Geranyl geranyl de phosphate synthase, Limonene synthase, Limonene hydroxylase are involved in the biosynthesis pathway of Carvacrol. Utilizing bioinformatics and real-time PCR, the expression of genes effective in biosynthesis of active substance and biochemical pathways was largely determined. Due to the lack of white savory genomic information, the gene sequences of plants of this family were examined. After predicting the sequence of these three genes, suitable primers were designed. In this study, the intracellular location of proteins was determined to be all three cytoplasmic genes. The properties and characteristics of each gene were also examined. Investigation of physical and chemical properties of proteins that are unstable proteins in all three genes, signal prediction of potential peptides in proteins and prediction of spatial structure of proteins, which includes structural evaluation and estimation of structural model quality through bioinformatics detection took. Bioinformatics study revealed that all three genes do not have peptide signal sequences. The quality of the structural model was examined through QMEANDisCo Local and QMEAN Z-Scores. All genes were found to be highly similar to the mentioned structural model. According to the authors of this study, bioinformatics study of genes at the protein level will create more awareness and facilitate future research.

Key Words: *White savory, Carvacrol, Biosynthetic pathway*

Investigating the effects of carboxyl terminal elimination from 604 retinal isoform of IMPDH1: an experimental and molecular dynamics simulation approach

Parisa Elyasi Ebli[1], Razieh Yazdanparast[1]*, Sajjad Gharaghani[1]

[1] *Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): parisaeliasi@ut.ac.ir

Abstract

Mutations in inosine monophosphate dehydrogenase 1 (IMPDH1), are known to be a root cause of Retinitis Pigmentosa (RP), a common hereditary blindness. Regulation of the activity of IMPDH1 is dependent on the occupation of nucleotide binding sites with GDP/GTP, performing an inhibitory effect. The retinal isoforms of IMPDH1, identified with distinct catalytic activity, contain additional terminal peptides which their possible roles in regulation of enzymatic activity relies undiscovered. The current study investigated the probable interactions of N-ter peptide extension of 604-isoform in absence of the C-ter peptide. MD simulations on dimer structures of wild-type and engineered proteins were performed using GROMACS 2019.1 simulation Package and CHARMM36 force field. The energy of the system was minimized and the system was equilibrated using Velocity-rescaling and Parrinello-Rahman algorithms for NVT and NPT equilibrations, respectively. The final production step of the system continued up to 20 ns. Further, RMSD and RMSF analyses were performed. The proteolytic digestion indicated a rapid digestion of the recombinant protein in contrast to the wild type 604-isoform, recommending a higher accessibility of α-Chymotrypsin to digestion sites due to the removal of C-ter peptide. Our computational data, also, revealed the formation of a novel helix of N-ter peptide in GTP2 binding site. This helix formation could affect the regulation of enzyme's catalytic activity, either by masking the nucleotide binding site or acting as an GTP-independent internal inhibitory element.

Andashti, B. et al. (2020). The functional impact of the C/N-terminal extensions of the mouse retinal IMPDH1 isoforms: a kinetic evaluation. Molecular and cellular biochemistry, 465(1-2), 155–164.
Buey, R. M. et al. (2015). Guanine nucleotide binding to the Bateman domain mediates the allosteric inhibition of eukaryotic IMP dehydrogenases. Nature communications, 6, 8923

Key Words: *Retinitis Pigmentosa; IMPDH1; Moleular Dynamics Simulation; retinal isoforms.*

Protein Engineering of Bacillus α-Amylase to Improve Thermostability and Low Water Activity: A bioinformatics approach

Babak Elyasi Far[1]*, Meysam Mard-Soltani[2], Ladan Mafakher[3], Shiva Mohammadi[4]

[1] Department of Physiology and Pharmacology, School of Medicine, Dezful University of Medical Sciences, Dezful, Iran*
[2] Department of Clinical Biochemistry, Faculty of Medical Sciences, Dezful University of Medical Sciences, Dezful, Iran
[3] Thalassemia & amp
Hemoglobinopathy Research center, Health research institute, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
[4] Department of Medical Biotechnology School of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran

E-mail address (corresponding author): b.elyasifar@gmail.com

Abstract

Background: α-Amylase produced by Bacillus licheniformis (BLA) is widely used in starch industries. Water is so crucial for α-Amylase activity; due to the low amount of available water in the starch industry, it can lead to denaturation of the enzyme. In this study, we attempt to reduce BLA aggregation and improve low water activity by protein engineering without reducing its activity by computational methods. **Material and Methods:** As surface residues are responsible in aggregation and water activity, the surface amino acids were defined by PyMol software (PDB ID: 1VJS). Aggrescan webserver was used to define hotspot aggregation residues. The surface amino acids of the α-Amylase which did not locate in the active site of the enzyme and were hotspot aggregation residues, replaced with glutamic and aspartic acid as negative amino acids by Chimera software. The stability and solubility of engineered α-Amylase compared to native form were done by Protparam and Protein-sol, respectively. To analyze secondary protein Psipred webserver was applied. Moreover, molecular dynamic simulation was done by Gromacs package to assess the stability of the tertiary structure of mutant form compared to native BLA during the time. **Result and Discussion:** According to the active site analysis of BLA, surface Amino acid Q340E, S356D, G474E, L318E, and S310E have neutral charges and are recognized as hot spots of aggregation that was changed to negative charge amino acids. Physicochemical properties analysis, protein solubility, and aggregation potential discovered that these properties of engineered enzyme improved from engineered mutant compared to native form. Also, secondary structure evaluation indicates that these mutations caused no change in protein structure which was proved by molecular dynamics simulation during time. So, these mutations could reduce BLA aggregation and improve low water activity although experimental analysis needs to prove it.

Key Words: *α-Amylase, protein engineering, low water activity, In Silico*

Diagnosis of breast cancer using K-Means and fuzzy C-Means clustering

Hamid Reza Erfanian[1]*, Zahra Ghafari[1]

[1] *Department of Bioinformatics, University of Science and Culture, Tehran, Iran*

E-mail address (corresponding author): erfanian@usc.ac.ir

Abstract

In recent years, breast cancer has been one of the most common causes of death among women. Thermography is one of the fastest, cheapest, risk-free, radiation-free, and painless diagnostic methods available for this cancer. The use of new methods in image processing and machine learning has led to the use of thermographic images to successfully conduct studies to establish breast cancer diagnostic systems. In this study, the diagnosis of breast cancer has been studied using K-Mean and fuzzy C-Mean clustering and an intelligent method has been used to separate healthy from unhealthy tissue and to separate mass in unhealthy tissue. In this method, clustering was performed using two methods: K-Mean and fuzzy C-Mean, then the cluster with the highest center intensity as the input of the area growth algorithm and the brightest pixel as the grain point of the area growth method were selected. The suspected area was determined according to the growth algorithm of the area and then the specificity of the suspected area was extracted based on the coefficient matrix. At this stage, the threshold was set for the four properties of the co-occurrence matrix and based on it, a decision was made about the suspicious area. Using this method, an intelligent system was designed that reduces the amount of human error in the diagnosis of cancer and will be able to detect the mass in the early stages of breast cancer with 91.67% classification and 89.65% sensitivity.

Key Words: *Breast cancer; Image processing; K-Means clustering Method; Fuzzy C- Means clustering Method.*

Binder design for targeting SARS-CoV-2 spike protein: An in silico perspective

Ali Etemadia[1], Hamid Reza Moradi[1], Farideh Mohammadian[2], Mohammad Hossein Karimi-Jafari[3], Babak Negahdari, Yazdan Asgari, Mohammadali Mazloomi[1]*

[1] *Medical Biotechnology Department, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences (TUMS), Tehran, Iran*
[2] *Department of Medical Biotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran*
[3] *Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): moradi-h@razi.tums.ac.ir

Abstract

The COVID-19 pandemic is affecting all aspects of people lives around the world. Coronaviruses are single-stranded RNA viruses and SARS-CoV2 belongs to the second group of this family. Spike glycoprotein (S protein) of the virus is the key molecule employed for cell penetration by binding to the receptor-binding domain (RBD) of angiotensin converting enzyme 2(ACE2). It is the only homing receptor recognized for the virus till date. Blocking the Interaction of S protein and ACE2 can inhibit the virus from entering the cell and replication. Small Proteins (2 – 20 kDa) as antiviral medications, may hold great promises as the future of therapeutics. Owing to the advances in the field of structural biology, it is now possible to design, highly selective small proteins. Here, we aimed at designing some novel binders to block the S protein of the SARS-CoV2.

The design process was based on collecting a list of natural proteins from PDB database. Collected proteins were monomeric, expressed in E.Coli and had no RNA, DNA, ligand or mutation in their structure. Then an Initial directed docking was performed by Patchdock against SARS-CoV2 S protein. Best performing protein was selected (3HGL) and used for interface designing by Rosettas FastDesign. Next the new designed models were filtered and Cluspro and PatchDock was used for blind docking to select the best binders. Also their structural parameters and characteristics was assessed with different web based tools. In the end, according to the data gathered from previous step three best performing binders have been chosen (BIN32,BIN78,BIN91). In order to investigate the conformational behavior of the binder's models and native scaffold, MD simulation was used. The results showed that BIN32,BIN78,BIN91 have a high binding energy towards S proteins of the SARS-CoV2. (-22.43, -20.91 and -17.01 kcal/mol respectively)

Key Words: *Protein design, COVID-19, Rosetta, MD simulations, Small Proteins*

Engineering carbonic anhydrase with improved solubility for green CO2 capture

Amirreza Farajnezhadi[1]*, Mona Masoumparast[1], Mehran Habibi-Rezaei[1]

[1] *Protein Biotechnology Research Lab (PBRL), School of Biology, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): ar.farajnezhadi@ut.ac.ir

Abstract

Carbonic anhydrase (CA) catalyzes the reversible conversion of CO2 to bicarbonate in living creatures. It is recommended for green CO2 capture technologies to enhance the low capturing rates of bioseparation processes. Different strategies were considered for the bioengineering of CA to stabilize the enzyme in extreme temperature and pH conditions, making it suitable for industrial operations. This study aims to improve enzyme solubility by recognizing aggregation-prone regions of CA and eliminating them. The 3D structure of the α-carbonic anhydrase (α-CA) from Thermovibrio ammonificans (TaCA; PDB ID: 4C3T), a thermostable tetrameric enzyme, was analyzed by Aggrescan3D 2.0 to identify aggregation hotspots on its surface. The average aggregation score of TaCA was calculated to be −0.9431, which indicates its low tendency for aggregation. However, seven aggregation-prone regions were discovered, centered on isoleucine 49, tyrosine 74, valine 237, cysteine 67, valine 73, leucine 45, and leucine 197, respectively by their severance. Cysteine 67 is necessary for the tetramerization of the enzyme that makes it thermostable, and leucine 197 is located in TaCA active site; therefore, these two remained untouched, and five other residues were mutated to alanine due to its low aggregation tendency. These mutations decreased the average aggregation score of TaCA by 10%, and aggregation hotspots disappeared, while the overall energy of protein structure experienced a 20.4 kJ/mol increase. Therefore, the mutated protein structure needs further dynamic examinations to assess enzyme stability. In conclusion, this study revealed that a small number of mutations could improve the overall solubility of the CA enzyme to prevent aggregation through CO2 bioseparation processes.

Key Words: *Protein aggregation; Carbonic Anhydrase; CO2 capture; Enzyme engineering.*

P23 protein is a candidate for vaccination against Cryptosporidiosis

Roohollah Fattahi[1]

[1] *Faculty of Para Veterinary Medicine, Ilam University, Ilam, Iran*

E-mail address (corresponding author): r.fattahi@ilam.ac.ir

Abstract

Cryptosporidiosis is an important parasite causing severe diseases in the immunodeficient people especially AIDS patients. Cryptosporidiosis has been also reported as a common serious primary cause of outbreaks of diarrhea in newborn calves. The aim of this study was to analyze different bioinformatics' characterization of sporozoite surface antigen P 23 of Cryptosporidium parvum.
Our result showed that this protein has 111 amino acids, MW 11232.39 Da and PI 5.18. the number of negatively and positive charged residues was 16 and 14 residues, respectively. The total number of atoms was 1564. The estimated half-life for P 23 protein was as follows: 30 h (mammalian reticulocytes, in vitro), > 20 h (yeast, in vivo) and > 10 h (Escherichia coli, in vivo). The instability index (II) is computed to be 70.11 and Aliphatic index was 44.59.
Also, our survey showed that P 23 has high antigenicity and it is not an allergen. This protein has four Serin phosphorylation sites and one acylation site. Transmembrane domains and subcellular localization of P 23 were 62.2 %: nuclear, 8.7 %: cytoskeletal, 4.3 %: cytoplasmic and 4.3 % plasma membrane. Secondary structure prediction in the P 23 sequence showed to 51 (45.95%) alpha helix, 58 (52.25%) random coil and 2 (1.8%) extended strands. moreover, 3D structure of this protein was predicted and refined. For this protein, three high-ranked linear epitopes were determined using BCPREDS server. Also, at least five linear epitopes with high score were confirmed by ABCpred Prediction Server.
this survey confirm that P23 could be a favorable candidate for vaccination against C. parvum infection.

Key Words: *Cryptosporidiosis, P 23, bioinformatic.*

Molecular docking study on the LMNA variant (R427C) effect in hypertrophic cardiomyopathy patient

Serwa Ghasemi[1], Iman Salahshourifar[1], Mohammad Mahdavi[1], Majid Maleki[1], and Samira Kalayinia[1]*

[1] *Cardiogenetic Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): samira.kalayi@yahoo.com

Abstract

We studied the effect of the LMNA variant (R427C) in the TGF-β pathway in an Iranian family with hypertrophic cardiomyopathy (HCM) by Molecular docking.
LMNA gene encodes lamin A and C (lamin A/C), which are intermediate filament proteins implicating in different cellular processes. Mutant lamin A/C may change TGF-β1 activity in the regulation of cell cycle progression, negatively. Therefore, cell proliferation will be stimulated in the presence of mutant lamins A/C. Molecular docking was performed using HADDOCK Web Server to investigate the binding mode between the normal and mutant forms of lamin A/C and the MAN1. Data visualization and the interactions were checked by PyMOL and LigPlus+, respectively.
In homology modeling, the structural alteration in lamin A/C R427C mutant was revealed. Furthermore, molecular docking identified the affinity of lamin A/C and MAN1 interaction has been reduced in presence of R427C, which followed by TGF-β signaling inhibition and Smad pathway reduction through sequestering at the nuclear envelope (NE).
Molecular docking analysis provided useful information of decreased binding affinity between mutant lamin A/C (R427C) and MAN1 in the HCM disease.

Key Words: *Hypertrophic cardiomyopathy; TGF-β signaling pathway; LMNA; Molecular docking*

Protein structure assessment using knowledge-based statistical potential function based on Ramachandran plots

Maryam Hojati[1], Seyed Shahriar Arab[1]*

[1] *Biophysics Department, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): maryhojati@gmail.com

Abstract

Background: Protein structural data is of prime importance for researchers in various biological fields, and the accuracy of the data can impact the significance of analysis results. However, native protein structures are not always available, so predicted protein models are widely used in many cases. Generally, in predicting protein structures, knowledge-based scoring methods based on statistical potential are used to select the most stable and native-like structure from the predicted models. For instance, the Modeller simulation tool uses "dope-score" to rank protein structures. According to a study by Anfinsen et al., the native structures have the least amount of free energy in a set of simulated models.

Materials & Methods: Methods based on statistical potential do not have the certainty and accuracy for calculating molecular mechanics force field; in contrast, statistical potential functions are computationally possible. For example, out of 55 decoy sets created by QUARK and I-TASSER algorithms for CASP11 targets, dope-score allocates only 15 minimum energies to native structures. Therefore, the need for research to achieve better evaluation criteria remains essential. This project intends to design a new knowledge-based scoring function based on PDB database information. Ramachandran diagrams are created for amino acid windows with 2 to 7 residues. The probability of any angle occurring in each window using statistical potential and the Boltzmann function is converted into energy. The protein with the least energy is considered as the best model.

Results & Conclusion: Our method could successfully rank more native protein structures as the top 5% among all modeled proteins in CASP11 than dope-score function.

Key Words: *Validation of structure, protein structure, Ramachandran, statistical potential, protein quality*

Bioinformatics Evaluation of gene and microRNA through HER2-positive breast cancer

Fatemeh Jalali[1], Fatemeh Hamidifar[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Breast cancer is the most common cancer in women. Due to the limitations of common cancer diagnostic tests, the introduction of biomarkers with higher specificity for the diagnosis of breast cancer is important [reference]. The aim of this study was to evaluate the bioinformatics of HER2-positive breast cancer and the resistance mechanisms in this disease caused by the mir-8 family expression. The study was done through genomics databases. The NCBI database was used to search for HER family genes in intracellular signaling pathways. Drug targets for HER2 were then assessed using the GeneCards database. Finally, the role of the mir-8 family in the pathogenesis of HER2 was determined using the MirBase database. According to the data obtained from the NCBI database, the HER family including EGFR (HER1/ErbB1), HER2 (ErbB2), HER3 (ErbB3), HER4 (ErbB4), and IGF-IR can activate several oncogenic signaling pathways to stimulate growth. According to GeneCards database, HER2 inhibitor drugs such as trastuzumab, pertuzumab, margotosumab or T-DM1 can be used for suppression of cell growth in the treatment of this cancer. These drugs block the Ras and PI3K/mTOR pathway and prevent cell division. Trastuzumab, as a HER2 antagonist, is one of the approved drugs for this cancer. The miRBase database showed that downregulation of mir141 in cancer cells has led to resistance to the drug. In addition, the miR-200c /141 cluster, both of which belong to the mir-8 family, plays an important role in the epithelial to mesenchymal transition (EMT) process. The expression of these two miRNAs is inversely correlated with HER2-positive breast cancer. Mir141 plays a prominent role as a metastasis suppressor gene. It is concluded that members of the miR-8 family by targeting HER2 involved in the growth and processes involved in breast cancer could possibly be investigated as diagnostic biomarkers in future studies.

Key Words: *Key words: HER family genes; Mir141; miR-200c; suppressive drugs*

Expression Analysis of the Key Genes Involved in Biosynthesis Pathway of Major Component of Ziziphora persica Bunge Essential Oil under Drought Stress

Mozhdeh Jalilifar[1]*, Khosro Mehdikhanlou[1], Lila Nejadsadeghi[1], Qing-Yan Shu[2]

[1] Department of Plant Production and Genetics, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[2] Key Laboratory of Plant Resources/Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing, China

E-mail address (corresponding author): jalilifarmozhdeh@yahoo.com

Abstract

Iranian Kakoti (Zizphora persica) is a plant belonging to the mint family, with a wide range of distribution throughout the country, that has edible and medicinal properties with antioxidant and antimicrobial properties. In order to conduct this research, Iranian Kakoti was cultivated in two conditions of control and drought stress treatment at 30% of field capacity (30% FC) in the Faculty of Agriculture at Shahid Chamran University of Ahvaz. Using the KEGG database, the metabolic pathway and genes involved in the production of this pathway were identified. The active ingredient of this plant is Polgen, with four key genes including: Isopiperitenol dehydrogenase, Limonene synthase, limonene hydroxylase and Isopiperitenone reductase involved in the path of biogen synthesis. Due to the lack of genomic information of the studied plant, these four genes were studied by examining their sequences in families close to this genus and species, namely Mentha Logifolia and Mentha Piperita. After predicting the sequence of these four genes, the corresponding primers were designed. In this study, the properties and the intracellular position of proteins showed that the protein belongs to all four cytoplasmic genes. Investigation of physical and chemical properties of unstable proteins in all four genes signal prediction of potential peptides in proteins and prediction of spatial structure of proteins, which includes structural evaluation and estimation of structural model quality through bioinformatics identification. The results of bioinformatics study showed that there was no peptide signal sequence for any of the studied proteins. The quality of the structural model was evaluated through QMEANDisCo Local and QMEAN Z-Scores and showed that all four genes have a high similarity with the model. The authors of this article believe that the bioinformatics study of genes at the protein level is a cheap way to gain more knowledge for future studies.

Key Words: *Bioinformatics, KEGG, Sequencing, Genomic information.*

The investigation of horse myostatin with human follistatin by molecular docking method

Elnaz Karbaschian[1], Ali Javadmanesh[1]*

[1] *Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran*

E-mail address (corresponding author): javadmanesh@um.ac.ir

Abstract

Myostatin is a transforming growth factor-beta (TGF-beta) family member that plays roles as a negative regulator of muscle mass development. Follistatin is a myostatin-binding protein that can inhibit myostatin activity to promote muscle growth. This is an important way for increasing growth especially in demo horses. Here, we aimed to prediction inhibition activity of equine myostatin protein with human follistatin, through bioinformatics tools. The Swiss-model server was applied to predict the third (three-dimensional) structure of horse myostatin protein and studies with SAVES 6.0 online server. After that, the interactions of myostatin with human follistatin were evaluated using variety of tools such as Verify3D, ERRAT and ClusPro2.0 online software [2, 3]. The results showed that the Verify3D for this protein was at least 88% of the amino acid residues have an average score of 3D-1D> = 0.2, which is acceptable for our protein. strategy evaluates proteins using a three-dimensional structure. This score varies from -1 (poor score) to +1 (good score). ERRAT is an online server that confirms the structure of a protein on the assumption of nuclear fusion between different types of atoms, with total quality index in this study was 83.52 which is acceptable. According to the Ramachandarn plot, for horse myostatin 88.7% of the amino acids in this structure were in the desired region, which is acceptable for predicting the third structure. The docking result showed that the N-terminal of follistatin was in contact with TGF-beta region in myostation that related to myostatin activity according to crystallography structure, derived from the docking of human follistatin with mouse myostatin. Finally, our research indicated that the binding energy in our predicted model was -903 which is close to the result of binding energy of the docking human follistatin with mouse myostatin which is reported –1323.3.

Key Words: *Follistatin, Equus caballus, Myostatin, Bioinformatics, Molecular Docking*

Identification of potential diagnostic microRNAs and lncRNAs in breast carcinoma: Integrated high-throughput bioinformatics investigation

Soudabeh Madhkhan Isfahani, Mohammad Rezaei, Ziba Rezvani Sichani, Samaneh Hosseinpour, Mansoureh Azadeh

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Breast cancer is the second foremost cause of death among women worldwide. MicroRNAs and lncRNAs are involved in regulating the expression of certain genes, as well as some biological processes in breast carcinoma. In the present study, we focused on identifying potential diagnostic microRNAs and lncRNAs in breast cancer using bioinformatics analysis. Breast cancer Microarray dataset GSE29431 was selected from Gene Expression Omnibus (GEO), then differentially expressed genes (DEGs) were screened. DEG analysis was performed using limma and GEOquery packages in R studio (v.4.0.2). The most significant genes (logFC > 3 and adjusted p-value < 0.05) were selected, and taken to miRWalk, to find relevant microRNAs. LncBase v.2 was employed to find decent lncRNAs (score = 1.000). Finally, we used STRING to find PPI network. After careful analysis of GSE29431 dataset, a total number of 286 up-regulated genes were detected in breast carcinoma. AGR2 (logFC = 3.84, adjusted p-value = 2.48E – 05), Muc1 (logFC = 3.45 and adjusted p-value = 3.76E - 06) and PLPP4 (logFC = 3.41 and adjusted p-value = 1.20E – 07) had remarkable expressions. All of genes were taken to miRWalk. We identified hsa-miR-197-3p, hsa-miR-145-5p as well as hsa-miR-148a-3p relating to AGR2, Muc1 and PLPP4, respectively. According to lncBase v.2, various lncRNAs including KCNQ1OT1, GLIDR and OIP5-AS1 was detected as potential lncRNAs, showing score = 1.000. Except for AGR2-Muc1 interaction network (gene co-expression score = 0.213), there was no gene interaction among other genes. Based on results, three microRNAs including has-miR-197-3p, has-miR-145-5p and has-miR-148a-3p, up-regulated in breast cancer, and have oncogenic role in this kind of carcinoma. Relating to these microRNAs, the expressions of KCNQ1OT1, GLIDR and OIP5-AS1 rise in breast cancer, so, they could be considered as potential diagnosis factors in breast carcinoma.

Key Words: *Breast cancer, microRNAs, lncRNAs, GEO.*

Rs78293998 promotes the hepatocellular carcinoma by disturbing the folding and interaction of CDKN3 protein: integrated gene expression and proteomics analyses

Benyamin Mashhadi[1], Mohammad Rezaei[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Hepatocellular carcinoma (HCC) is now one of the most common cancers and the top cause of cancer-related death worldwide. This investigation aimed to find a novel differentially expressed gene (DEG) in the HCC patients compared to control samples. Expression analysis of GSE121248 achieved from GEO2R online software and validation of expression analyses performed by GEPIA2 database. Single nucleotide polymorphisms (SNPs) of CDKN3 extracted from dbSNP and identification of deleterious SNPs Brought out from SIFT and PROVEAN databases. Biophysical validation of deleterious SNPs realized from HOPE software. Based on microarray analysis, CDKN3 have a significant up-regulation in the HCC samples, compared to control (logFC: 2.237, p-value < 0.0001). From all of the extracted SNPs on the coding region, SIFT and PROVEAN online software revealed that rs78293998 is the most significant deleterious SNP in the protein-coding region CDKN3. This SNP is the arginine mutation into Isoleucine at the 72nd position of CDKN3 protein based on HOPE software. This mutation is located in the Tyrosine-protein phosphatase domain and can disturb this function of CDKN3 protein. Based on the biophysical validation of HOPE, the mutant residue has less charge and smaller size, which can lead the protein to the loss of main interactions. Furthermore, the mutant residue is more hydrophobic than the wild one, which can result in loss of hydrogen bonds and correct protein folding of CDKN3. In conclusion, rs78293998 can promote the HCC development by changing the expression level of CDKN3 while changing the correct folding and protein interactions of CDKN3 and miss-regulation of Tyrosine-protein phosphatase activity.

Key Words: *Microarray; Data analysis; Protein structure; Single nucleotide variations*

Function of lectins in the efficiency of ferritin nanoparticle vaccines

Seyedeh Zeinab Masoomi Nomandan[1], Seyed Masoud Hosseini[1], Maryam Azimzadeh Irani[1]*

[1] *Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran*

E-mail address (corresponding author): m_azimzadeh@sbu.ac.ir

Abstract

Introduction: Viral infections constitute a serious public health and social problem worldwide, especially during the recent COVID-19 pandemic caused by the betacoronavirus SARS-CoV-2. Display of antigens on ferritin generally elicits a more robust neutralizing antibody response against the target pathogen as compared to immunization with the antigen alone. The ferritin-RBD nanoparticle vaccine induced not only a persistent RBD specific antibody response but also long-term protective memory. Because the ferritin nanoparticle vaccine uses glycosylated RBD (receptor binding domain) in its structure, it may be recognized and gets inefficient by human lectins. Lectins are glycan-binding proteins, present ubiquitously in all forms of life. In this study, molecular docking to examine the lectin binding site on the ferritin nanoparticle vaccine protein used.

Material & Methods: The lectin protein structure was obtained from the protein data bank (PDB id: 5b1w). haddock 2.4 server was used to dock the lectin and ferritin nanoparticle vaccine complex together and biomolecular visualization of the complex was carried out by pymol.

Results: The structure of the full length human lectin was obtained in complex with RBD-ferritin nanoparticle. Lectin (residues:51-64) binds close to the RBD-ferritin interface (residues: 41-49) and Haddock score (3.7 +/- 12.1) and RMSD value (10.9 Å) of the molecular docking showed that this connection has low stability and it does not interfere with ferritin and RBD binding. The results suggesting that efficiency of ferritin nanoparticle vaccine may Not be affected by human lectins.

Conclusions: Binding of human lectin to the RBD-ferritin nanoparticle shows that although the binding site of the lectin is near to the RBD and ferritin interface it does not inhibit the ferritin-RBD binding.

Key Words: *Lectin, RBD-ferritin, Nanoparticle vaccine, Molecular docking.*

Investigation of the mechanism of phenol inhibitory activity in the function of fructosyl peptide oxidase

Mona Masoumparast[1]*, Amirreza Farajnezhadi[1], Mehran Habibi-Rezaei[1]

[1] *Protein Biotechnology Research Lab (PBRL), School of Biology, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): m.masoumparast@ut.ac.ir

Abstract

Fructosyl peptide oxidase (FPOX, EC: 1.5.3) that belongs to the oxidoreductase family can identify fructosyl valine (FV) and fructosyl lysine (FK) as substrate. In this reaction, fructosyl amino acids are broken into glucosone, amino acids, and hydrogen peroxide (H_2O_2). The amount of produced H_2O_2 measures reacted substrate that could be monitored with colorimetric methods. As a measurement method, 4-aminoantipyrine and phenol act as electron scavengers to adsorb free radicals produced by H_2O_2 in the presence of peroxidase. It was observed in laboratory studies that a high phenol concentration could interfere with FPOX activity. Therefore, we conducted an in silico study to discover the phenol inhibition activity. Firstly, FPOX (PDB ID: 4RSL) active sites and pockets were identified by the Computed Atlas of Surface Topography of proteins (CASTp). Then Auto Dock Vina was employed to investigate the binding of FK and phenol to the FPOX. Accordingly, the binding affinities of the FK and phenol to the FPOX active site were calculated as -23.87 and -21.35 kJ/mol, respectively. These results suggested that phenol could not interfere with the enzyme's function, especially in low concentrations. Although, when phenol concentration increases, it can compete with FK for binding FPOX active site to inhibit its activity possibly in a competitive manner of action.

Key Words: *Fructosyl peptide oxidase; Enzyme Activity; Molecular docking; Phenol.*

Homology modeling of FimH protein of Klebsiella pneumoniae :a key protein in bacterial pathogenesis

Fahimeh Mazroe[1], Mohamad Reza Dayer[2]*, Fatemeh Khademimoghadam[2]

[1] *Department of Biology, Faculty of Science, Kherad Institute of Higher Education*
[2] *Department of Biology, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran*

E-mail address (corresponding author): fahimehmazro766@gmail.com

Abstract

Klebsiella pneumoniae, as an important gram-negative pathogen causes urinary tract and respiratory infections which leads to large numbers of cases each year. One of the main factors involved in the pathogenicity of this bacterium is the FimH protein. This protein binds the bacterium to the host mucosal tissue, which leads to colonizing and multiplying the pathogen in that site and eventually inhibiting the host immune system.
In this study, the tertiary structure of FimH protein of Klebsiella pneumoniae was predicted by homology modeling carried out by MODELLER software (Version 9.15). Two models (M1 and M2) were selected as output models for further analysis. After energy minimization of the models, their validation was checked by the analysis of the second structure (Ramachandran diagram), z-score, energy level, protein binding site, and docking experiments using mannose and FimH antagonist (alpha-D-mannopyranoside) as ligands. Ramachandran plot showed that more than 92% of amino acids of M1 and M2 are in acceptable regions and their energy level analysis confirmed that they had appropriate structures. otherwise, the values for M1 and M2 in the Z-score plot were, -2.65 and -2.08 respectively. The values matched the experimentally determined values for the X-ray structures. It is noted that 21 possible active sites in FimH protein were identified by predicting protein binding sites. Finally, the docking results showed that binding energies of mannose to M1 and M2 were -195.45, -201.81 kJ/mol respectively while the binding energies of alpha-D-mannopyranoside to M1 and M2 were 310.45, -282.23 kJ/mol respectively.
In conclusion, our findings indicate that M1 model contrast to M2 seems to be a better model for further studies as FimH structure to control the pathogenicity of Klebsiella pneumoniae via drug design.

Key Words: *: Homology Modelling, K. pneumoniae, FimH, Structure Analysis*

Bioinformatics analysis of gene expression profiling for identification of key proteins associated with Hepatocellular Carcinoma

Elnaz Mehvari[1], Dr Najaf Allahyari Fard[2]*

[1] *Higher Education Institute, meymeh. Iran*
[2] *The National Research Center for Genetic Engineering and Biotechnology, Tehran, Iran*

E-mail address (corresponding author): allahyar@nigeb.ac.ir

Abstract

Hepatocellular carcinoma is a type of primary liver cancer.HCC is the sixth most common cancer in the world and the third leading cause of cancer deaths. The most important cause of death in patients with HCC is more than 80% of cronic hepatitis B or hepatitis C in fection and long _term use of aflatoxin. One of the characteristics of this malignancy is the high rate of progression of this disease HCC is a malignant tumor with a poor diagnosis that surgery and liver transplantation have been the two main treatments that can be preformed only in the early stages.
Objective: The aim of this study was to identify key genes, biological pathways and new drug targets using appropriate bioinformatics toola to combat hepatocellular carcinoma.
Methods: in this study microarray expressive data downloaded from GEO database was used, which used GEO2R online server to analyze the data.protein interaction was also potted using a network analysis criteria were preformed using cytoscape software using the centiscape plugin analysis of gene categories was also obtained using a metascape. finally, 9 key genes were identified, including CCNA2, CCNB1, TYMS, EZH2, PCNA, CDK1, CAT, ESR1, ALB.
Also different key pathways have been identifided that are associated with hepatocellular carcinoma.identification of these genes and key pathways involved in hcc discovering the disease mechanism and appropriate drug targets.

Key Words: *Hepatocellular Carcinoma, Hepatitis B, Hepatitis C, microarray Technique.*

Evaluation of site-directed mutagenesis on stability and affinity of Bevacizumab antibody

Shima Mirhoseini[1], Mahsa Mirzaei[1], Mohammad Mehdi Heidari[1]*, Mehri Khatami[1]

[1] *Department of Biology, Faculty of Sciences, Yazd University, Yazd, Iran*

E-mail address (corresponding author): ma_mirzaei76@yahoo.com

Abstract

Bevacizumab is an anti-angiogenesis monoclonal antibody which prevents the interaction of VEGF-A with VEGFR and thereby inhibits the activation of VEGF signaling pathways that promote neovascularization. The type of amino acids at the site of antigen and CDR interaction is important and also the amino acid diversity is accumulated in the three complementarity determining regions (CDRs). Cysteine, proline and methionine are avoided for their chemical reactivity or constraints in folding. Isoleucine and leucine are underrepresented to decrease the surface hydrophobicity. Furthermore, arginine, lysine, serine, threonine, tyrosine and asparagine are preferred as these amino acids occur frequently in the paratope, where they are involved in antigen interaction. As a result, these substitutions enable us to optimize antibodies' affinity and stability. VEGF-A amino acid sequence and Bevacizumab amino acid sequence was also extracted from https://www.rcsb.org with 6BFT PDB code. This web server was also used for CDR prediction. Initialy we used pymol software to visualize antibody-epitope interaction. Some amino acids were selected as significant residues in Bevacizumab structure by employing the results of different software. These residues located in one of the three CDR regions. Pymol software can predict amino acid substitution influence on antibodies' affinity by measuring bond length and strength. In order to explore the effect of amino acid substitution on antibodies' stability mCSM web server also was used. Several characteristics of monoclonal antibody should be developed in order to use them as therapeutic agents, including biding affinity, folding stability and pharmacokinetics. Optimization strategies for monoclonal antibody provide the possibility of modification and improvement of an antibody molecule. By analysis with mCSM some of mutants of antibody improved stability, especially conversion of tyrosine 102 to glutamic acid. Evaluation of optimization for antibody affinity by Pymol software indicated that the conversion of tryptophan 108 to arginine improved affinity.

Key Words: *Site-directed mutagenesis; Stability; Affinity; Bevacizumab antibody*

Genetic Analysis of 40 MLPA-Negative Duchenne Muscular Dystrophy Patients by Whole-Exome Sequencing

Gholam Reza Zamani[1] Mohammad Farid Mohammadi[2], Ali Reza Tavasoli[1], Mahmoud Reza Ashrafi[1], Sareh Hosseinpour [3], Homa Ghabeli[1], Elham Pourbakhtyaran[1], Roya Haghighi[1], Seyed Mohammad Mahdi Hosseini[1], Pouria Mohammadi [4]*, Morteza Heidari[1]*

[1] *Department of Pediatric Neurology, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran*
[2] *Department of Cell and Molecular Sciences, Faculty of Biological Sciences, Kharazmi University, Tehran, Iran*
[3] *Department of Pediatric Neurology, Vali-e-Asr Hospital, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran*
[4] *Department of Medical Genetics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): m.mahamadi@khu.ac.ir

Abstract

This manuscript aimed to determine the underlying point mutations causing DMD in a heterogeneous group of Iranian patients, who are clinically suspected. Whole-Exome sequencing was utilized to detect disease-causing variants in 40 MLPA-negative DMD patients. Disease-causing variants were detected in the DMD gene in 36/40 of the patients (90%), and 4/40 of them (10%) remained undiagnosed. WES analysis revealed that nonsense mutation was the most common type in our study (23/36 of the cases). Besides, 12/36 of the cases had frameshift mutation, and one of the patients had a likely pathogenic splice variant in the DMD gene. Carrier testing revealed that 21/40 of the mothers had the identified variant. Therefore, most mutations were inherited (58.3%), while 19/40 were de novo (41. 7%). The present study has demonstrated the importance of performing WES to detect disease-causing point mutations in MLPA-negative DMD patients and to identify carrier females. Due to regulatory challenges, the clinical development of therapeutic approaches is time-consuming and may not be available to all patients shortly. Therefore, it appears that the techniques used to accurately detect mutations in carrier mothers are a more efficient solution to prevent the increased prevalence of DMD.

Key Words: *Duchenne muscular dystrophy. Multiplex ligation-dependent probe amplification. Whole-exome sequencing, Dystrophin*

In Silico Identification of key microRNAs in Cannabis sativa L.

Motahhareh Mohammadi[1], Sedigheh Fabriki-Ourang[1], Jafar Ahmadi[1]

[1] *Genetics and Plant Breeding, Imam Khomeini International University, Qazvin, Iran*

E-mail address (corresponding author): mohammadi.m314@gmail.com

Abstract

Cannabis (Cannabis sativa L.) is an annual, herbaceous, and dicotyledonous plant of the Cannabinaceae family. Cannabis has not been considered as important food, medicinal and industrial source and little research has been done on its agronomic, physiological, and biochemical properties. MiRNAs are small, non-coding RNAs that regulate gene expression in plants and animals. These small RNAs play an important role in regulating post-transcriptional gene expression and act as negative regulators of gene expression in eukaryotes. In recent years, the use of bioinformatics tools to identify miRNAs and their target genes has been widely considered and used due to the availability of genomic sequences in databases due to their efficiency and low cost. The main purpose of this study was to identify conserved miRNAs in the genome of the Cannabis sativa L. through the computational genomics homology search approach. Therefore, to identify cannabis microRNAs, a homology-based search was performed between Cannabis sativa L. genome data in the NCBI database and known microRNAs in the miRBase database, using BLASTn, then to better identify the secondary structure, With the help of the BEDFile tool, 200 nucleotides were added to the beginning and end of each candidate target miRNA and in the next step, BLASTx was performed to remove the coding sequences of the proteins. The secondary structures of candidate microRNAs were predicted using MFOLD software. Finally, eight miRNAs were identified, which belonged to six different families. Csa-miR169e 5p and csa-miR169k belonged to the miR169 family and csa-miR319b along with csa-miR319a3p belonged to the miR159 family. Other identified microRNAs include csa-miR8014 5p, csa-miR8005c, csa-miR2111b and csa-miR1533. The current results provide molecular evidence for understanding the possible involvement of miRNAs in the regulation of Growth and development by miR159 family and Stresses response by miR169 family in this plant species.

Key Words: *Bioinformatic, Cannabis sativa, Comparative genomics, Genome, MicroRNA.*

α-Glucosidase Un-Competitive Inhibition investigation via Residual Interaction Network, and Free Energy Landscapes: 0.5 μs Molecular Dynamics Simulations

Zahra Moosavi-Movahedi[1], Najmeh Salehi[2], Mohammad Hossein Karimi-Jafari[1]*

[1] *Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*
[2] *School of Biological Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*

E-mail address (corresponding author): mhkarimijafari@ut.ac.ir

Abstract

Residual Interaction Network (RIN) analyses combined with molecular dynamics (MD) simulations have shown to be a powerful tool for the analysis of protein conformational dynamics, that plays a key role in the allosteric mechanism of enzymes. Here, with the coherence between RIN analysis and Free Energy Landscapes (FEL), two 250 ns trajectories are compared, which were created by MD simulations on free and xanthene derivatives bound of α-glucosidase. The use of two-dimensional principal component analysis (PCA) on cartesian coordinates of MD trajectories, leads to the discovery four basins in the conformational landscape of α-glucosidase and its complex to identify the representative of their sub-states. By integrating the networks of each basin conformations, and creating the mean network for each one, four intermediate networks of α-glucosidase were created. We also utilize the Girvan-Newman algorithm to detect communities and its interactions as a structural organization in α-glucosidase. The role of residues with large variations in the centrality values were examined on the pathways between the active site and distant allosteric site. Also, the hotspot residues distal from the active site, and residues with significant dynamical changes upon ligand-binding, and their communities are identified. So, the structure of α-glucosidase is well organized with different communities acting different roles in the ligand binding and allosteric un-competitive inhibition mechanism. This study demonstrates that the combination of the RINs with FEL on MD trajectories could be an effective method which can be extended to investigate allosteric communications for other macromolecular interaction systems.

Key Words: *Residue Interaction Networks, α-glucosidase Uncompetitive Inhibition, Allostery, Molecular Dynamics Simulation, Hotspot Residues.*

Molecular docking of Enterocin-P peptide with DNA: an in silico study

Ali Javadmanesh[1], Zahra Mousavi[1]

[1] *Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran*

E-mail address (corresponding author): javadmanesh@um.ac.ir

Abstract

Antibiotic resistance is a looming public health crisis especially in animal industry. Hence, researchers are looking for natural alternatives such as Antimicrobial peptides (AMPs). Nowadays, AMPs, also known as host defense peptides, are short and generally positively charged peptides found in a wide variety of life forms from microorganisms to humans. One of the modes of action of some AMPs were the binding with DNA and can directly affect the expression of related genes and inhibit the synthesis of macromolecules or destroy. Among AMPs Entrocin-P peptide has attracted a lot of attention from researchers. The aim of this study was to evaluated the possibility of binding DNA to Entrocin-P peptide. The structure of Entrocin-p peptide was obtained from previous study. A general strategy for docking and modelling DNA–protein complexes has been developed with HADDOCK server. HADDOCK uses non-structural experimental data to drive the docking during a rigid-body energy minimization, and semi-flexible and water refinement stages. The latter allow for flexibility of all DNA nucleotides and the residues of the protein at the predicted interface. Using this method, we have reproduced the structure of Entrocin-P peptide bound to DNA. The results have shown that the model of the Entrocin-P -DNA complex successfully has predicted. Furthermore, the results indicated that interactions were occurred by the major groove of DNA with α-helix part of this peptide. The best haddock score was -74.3 +/- 3.5. In addition, the value of Z-Score was -2.5. Its Z-score indicates how many standard deviations from the average this cluster is located in terms of score that the more negative the better.

Key Words: *Entrocin-P, Peptide, Bioinformatics, DNA, Molecular docking*

Bioinformatics evaluation of hsa-miR-4496 related to a single nucleotide polymorphism (rs1031305330) of GSTP1 gene in gastric cancer

Sajede Naghiyan Fesharaki[1], Sajjad Sisakhtnezhad[1], Mansoureh Azadeh[2]

[1] *Department of Biology, Faculty of Science, Razi University, Kermanshah, Iran*
[2] *Zist Fanavari Novin biotechnology institute, Isfahan, Iran*

E-mail address (corresponding author): minota_2012@yahoo.com

Abstract

Introduction: Gastric cancer remains one of the most common and deadly cancers worldwide. Helicobacter pylori infection is the major risk factor associated with the development of gastric cancer (GC). The transition from normal mucosa to non-atrophic gastritis, triggered primarily by H. pylori infection, initiates precancerous lesions which may then progress to atrophic gastritis and intestinal metaplasia. GSTP1 gene is located position on choromosome11q13.2. Different studies indicated that this gene is associated with gastric cancer. Based on accomplished studies, expression of this gene may be a prognostic indicator for several type of cancer. Single nucleotide polymorphisms (SNPs) are consisting genetic marker related to many of the genetic disease. microRNAs are counted as bio- markers with regulation and control of genes expression at level of mRNA. Given the importance of SNPs and microRNAs as cancer biomarkers that influence the expression of genes, this study aimed to investigate association between hsa-miR-4496 and a single nucleotide (rs1031305330) of GSTP1 gene in patients suffering from gastric cancer.
Methods: In the present study, databases such as NCBI, miRNASNP-V3, miRWalk, PhenomiR, miRBase and David were used for in silico data analyses.
Results: Our bioinformatics analysis indicated that rs1031305330 is associated with the conversion of C to T allele in GSTP1 gene, which may affect the performance and binding of miRNAs in this region. In this regard, results of this study also revealed that hsa-miR-4496 can bind to the 3'-UTR transcript of the GSTP1 gene.
Conclusion: Altogether, this study identifies rs1031305330 and hsa-miR-4496 as important biomarkers that are associated with GSTP1 gene and gastric cancer. Also, it suggests that the formation of the T allele in rs1031305330 affects the binding of hsa-miR-4496 to the GSTP1 gene.

Key Words: *Single nucleotide polymorphism, rs1031305330, GSTP1 gene, hsa-miR-4496, Gastric cancer.*

Relative expression analysis of COL1A1 mRNA in oral tongue squamous cell carcinoma: a study based on gene expression omnibus (GEO) and bioinformatics analysis

Naeimeh Parsapour[1], Mohammad Rezaei[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Oral Tongue Squamous Cell Carcinomas (OTSCC) is one of the most aggressive and common malignant tumor of the head and neck distinct. Our study aimed to define a competitive endogenous RNA (CeRNA) network that provides valuable biomarkers and boosts the treatment process of OTSCC. Therefore, bioinformatics analysis targets novel biomarkers for diagnosis and curing OTSCC. From the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/), we obtained the gene expression data of 10 OTSCC tumor and control samples (GSE160042). GEO2R analyzed the data sets to find out differentially expressed genes in patients (DEGs). A threshold value of |logFC| > 2 and a p-value< 0.05 were chosen, and one of the most significant genes was taken to miRWalk 2.0 to find target miRNA. Target miRNA was searched in LncBase v.3, and appropriate lncRNAs were found. At last, Cytoscape software (3.9.0) was used to show the interaction between the components of the CeRNA network. Pathway enrichment was performed by KEGG. COL1A1 had a significant high expression in the tumor samples (logFC > 2, adj. p. value < 0.05). Moreover, mir-615-3P is a novel suppressor for COL1A1 (score: 1.00). miR-516-3p have RNA interactions with, NORAD, XIST, and MIAT lncRNAs. COL1A1 is an invasion-related gene observed in the 'extracellular matrix' (ECM) and cell adhesion signaling pathway. It has been reported to be involved in multiple biological behaviors, including cell proliferation, invasion, metastasis, and angiogenesis. miR-615-3p have a significant suppressor effect on an over-expressed gene (COL1A1) in OTSCC patients. Also, it has the significant RNA interaction with NORAD, XIST, and MIAT lncRNAs. The mentioned lncRNA can regulate the OTSCC development and metastasis by affecting the expression level of COL1A1 by regulating cell proliferation.

Key Words: *Oral Tongue Squamous Cell Carcinomas (OTSCC), mir-615-3P, COL1A1, Biomarker*

Genome-Wide Identification of Conserved MiRNAs in Andrographis paniculate Nees

Heydar Rabbaniha[1], Jafar Ahmadi[2], Sedigheh Fabriki-Ourang[2], Daryuosh Talei[3]

[1] *Genetics and Plant Breeding, Imam Khomeini International University, Qazvin, Iran*
[2] *Genetics and Plant Breeding, Faculty of Agriculture & Natural Resources, Imam Khomeini International University, Qazvin, Iran*
[3] *Medicinal Plant Research Center, Shahed University, Tehran, Iran*

E-mail address (corresponding author): heydar.rb@gmail.com

Abstract

No study has been performed on identifying microRNAs in the medicinal plant, Andrographis paniculata. Nain-e Havandi (Andrographis paniculata Nees) is a medicinal plant of the Acanthaceae family. Nain-e Havandi has numerous medicinal properties and is widely used in Chinese traditional medicine. Medicinal properties of this plant include anti-cancer, anti-diabetes and anti-hepatitis. MicroRNAs are a class of short (18 to 24 nt), non-coding regulatory RNAs that play a major role in post-transcriptional gene silencing by targeting mRNAs for degradation or inhibiting translation. MiRNAs are the main regulators of a wide range of biological processes, including evolution, cell differentiation, signal transduction, secondary metabolic pathways, and response to a variety of biotic and abiotic stresses. The purpose of this study was to identify conserved miRNAs in the genome of the Andrographis paniculata through the computational genomics homology search approach. The required genomic sequences of the Andrographis paniculata and the mature sequences of previously identified plant-conserved miRNAs were obtained from the NCBI and miRBase databases, respectively. The Blastn was performed to find identical sequences with conserved miRNAs in the genome of Andrographis paniculata. The protein-coding sequences were then deleted using the Blastx algorithm. The mfold web server software was used to study the secondary structure of the remaining sequences. Also, the analysis of the minimal folding free energy index, base content, and percentage of base pairs formed in secondary structures was performed. Finally, a total of 46 conserved miRNAs were identified in the genome of Andrographis paniculata, of which 26 miRNAs belonged to 17 independent families. The miR156 family with 4 members was the largest identified family. The miR399 family had three members and the miR167_1, miR160, miR169_2 and miR5067 families each had two members. Only one member was identified for each of the 11 remaining miRNA families.

Key Words: *Andrographis paniculata, Bioinformatics, Comparative genomics, Genome, MicroRNAs.*

Protein Structure Preidction of Glutamate receptor 3

Fatemeh Rahimi[1]*, Sayed Mojtaba Ghavami[2]

[1] *Department of Cell and Molecular Biology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran*
[2] *Department of Horticulture, Faculty of Agriculture, University of Ilam, Ilam, Iran*

E-mail address (corresponding author): rozitarahimi1374@gmail.com

Abstract

Metabotropic Glutamate receptor 3 (mGlu3) is a Receptor for glutamate that functions as ligand-gated ion channel in the central nervous system and plays an important role in excitatory synaptic transmission. This receptor is a type of multi-pass membrane protein. mGlu3 is one of the therapeutic targets due to its genetic association with psychiatric disorders. It seems that the use of protein structure modeling and prediction methods on Glutamate receptor 3 protein with the aim of developing more effective and efficient agonists for the treatment, can have a significant effect on promoting Drug Discovery studies. For this purpose, the structure of the protein predicted by I-TASSER, Robetta and SWISS-MODEL server. The best model selected and subjected to ProSA and VADAR servers in order to determine the quality of the predicted structure. The Z-score of the model was -9.31 and the percentage of amino acids in most favored regions was 88.161. Then the protein structure was placed into a membrane model by CHARM-GUI server. Overally the predicted model could be an ideal structure for further studies such as docking. In this regard the result of molecular docking studies using the predicted model as receptor and L-glutamate as ligand shows that the best binding pose has -3.58 kCal/mol.

Key Words: *I-TASSER; SWISS-MODEL; Structure Prediction; CHARM-GUI; Modeling*

Molecular Docking Studies of Vitamin K Epoxide Reductase Subunit1

Fatemeh Rahimi[1]*, Sayed Mojtaba Ghavami[2]

[1] *Department of Cell and Molecular Biology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran*
[2] *Department of Horticulture, Faculty of Agriculture, University of Ilam, Ilam, Iran*

E-mail address (corresponding author): rozitarahimi1374@gmail.com

Abstract

Vitamin K Epoxide Reductase (VKOR) is a membrane protein that regenerates Vitamin K 2,3-epoxide and converts vitamin K to vitamin K hydroquinone by enzymatic catalysis. Considering the importance of VKORC1 protein in several diseases and its importance as the target of many protein complex inhibitory antagonists in the prevention and treatment of thromboembolic diseases, cardiovascular diseases and myocardial infarction, docking studies and evaluation of the binding and efficacy of existing drugs affecting this protein are of particular importance. Common drugs for this protein include Warfarin, Brodifacoum, Phenindione, and Cholorophacinone. It seems that docking and the study of protein-ligand binding on VKORC1 protein with the aim of evaluating available drugs in the market, can have a significant impact on the promotion of drug discovery studies. For this purpose docking of VKORC1 with each drug done by AutoDock software. By comparing the energy of the best conformation from each cluster for various VKORC1 protein antagonists and Vitamin K Epoxide, it was found that the best drugs in inhibiting this receptor and subsequently preventing related diseases, are Brodifacoum, Chlorophacinone, Warfarin and Phenindione, respectively. Also, all of these drugs bind with more binding energy than the main protein ligand (Vitamin K Epoxide), which indicates the effectiveness of all these drugs. This could be due to an increase in the number of amino acids involved in binding at the active site of the enzyme based on information from comparisons in Ligplus software.

Key Words: *Molecular Docking; VKORC1; AutoDock; Brodifacoum; Chlorophacinone*

The Role of CD81 Gene in the Expression of Tumor-Suppressive miR-let-7 in HCV-Infected Patients with Hepatocellular Carcinoma

Niloofar Rahimzadeh[1], Golnaz Shirazi[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Hepatocellular carcinoma is the most frequent type of cancer in the world, and it is intimately associated to advanced liver disease. Given the importance of hepatocellular carcinoma, there is a growing interest in studying CD81 function in hepatocytes. Several members of tetraspanin superfamily have been described as managing different virus infectious phases at different stages, including viral entry, viral replication, virion exit, and infectivity which is associated with immunodeficiency disease, including hepatitis C (HCV) which is a serious human pathogen with a high chronicity rate [3,4]. Chronic hepatitis C (CHC) is a virus that causes hepatocellular cancer (HCC) in people. The family of miR-let-7, a kind of short non-coding RNA that regulates the expression of certain genes, contains 12 members of miRNAs. Some of them have demonstrated dysregulation results in the less differentiated cellular phase and development of cell-based disorders like cancer. In this study, the roles of CD81 gene and miR-let-7 were discussed in promotion of hepatocellular carcinoma in HCV-infected patients to validate the related prior reports. The function of CD81 gene was investigated through GeneCards and KEGG databases. MiRWalk and miRBase databases showed the involved members of miR-let-7 family in HCC and related to CD81 gene. According to the bioinformatics analysis, CD81 plays a role in HCV infection, which may lead to the development of HCC by interaction with miR-let-7 downregulation. It is concluded that considering level of miR-let-7 expression may suppress cell line proliferation in HCC. Overall, it seems reasonable to assume that CD81 is associated with HCV and HCC.

Key Words: *HCC; HCV; CHC; miRNA; Bioinformatics*

A multi-epitope vaccine design to target SARS-CoV-2 employing an in silico approach

Hamid Reza Rasouli Jazi[1]*

[1] *Faculty of Chemistry and Chemical Engineering, Malek Ashtar University of Technology, Tehran, Iran*

E-mail address (corresponding author): hamid.rasouli1993@gmail.com

Abstract

The Covid-19 pandemic started in 2019 and has become a threat to human health with its rapid spread worldwide. Despite advances in the field of SARS-CoV-2 vaccines, the effectiveness of existing vaccines against new variants of the virus is in doubt, and repeated booster doses are needed to maintain antibody levels at an adequate level. Here, we used structural proteins S, M, E, and N as appropriate candidate targets to develop a vaccine against SARS-CoV-2. We analyzed these proteins to find analogs and conserve domains in different SARS-CoV-2 variants. We used various immunoinformatics tools to predict different T and B lymphocyte epitopes in conserved regions of target proteins. We performed analyses to select the best epitopes based on allergenicity, antigenicity, physicochemical properties, and toxicity. Six RNA fragments were built and linked in a specific order with linkers. We predict the tertiary structure of epitopes and implied dockings with corresponding MCH to evaluate binding affinities. Our results suggest a multiepitope vaccine against different variants' converse regions of SARS-CoV-2 proteins.

Key Words: *Covid-19, SARS-CoV-2, multiepitope vaccine, in silico vaccine design, immunoinformatics*

Integrated Bioinformatics Analysis of Hub Genes and Pathways in Thyroid Cancer

Ziba Rezvani Sichani[1], Adel Rezvani Sichani[1], Soudabeh Madhkhan Isfahani[1], Mohammad Rezaei[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Thyroid cancer (TC) is the sixth most common type of cancer among women worldwide.The aim ofthe present study was to identify hub genes and pathways in Thyroid cancer by microarray expression profiling Furthermore,a protein-protein interaction(PPI) network was constructed, and 8 hub genes were identified based on this network.Microarray data selected for bioinformatic analysis is Gene Expression data stored with GSE35570, GSE3678, GSE33630 code in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. Thyroid cancer and healthy control groups were compared,That including 100 thyroid tumors and 74 normal tissues to obtain the intersection of differentially expressed genes, and a protein-protein interaction network was constructed to obtain the HUB gene and analyzed by Search Tool for the Retrieval of Interacting Genes online tool and R software Based on the whole network, we identified 8 hub genes that included CCNA2,FEN1,CCNB1,CDK1,RRM2,CDC45,CHRDL1,UBE2C. which were highly expressed in TC tissues. Bioanformatic data suggested that the expression levels of CDK1, UBE2C and CCNA2 genes were also upregulated in other histological subtypes of thyroid carcinoma. High expression of FEN1,CDC45,CCNB1,CHRDL1 and RRM2 gene significantly decreased disease-free survival of patients with other thyroid carcinomas ,and Enrichment analysis showed that these hub genes were primarily accumulated in ‘cell cycle’ and ‘p53 signaling pathway’, ‘viral carcinogenesis’These findings suggest that may several hub genes(CCNA2, FEN1,CCNB1,CDK1,RRM2,CDC45,CHRDL1,UBE2C) and pathways, which will contribute to elucidating the pathogenesis of TC and providing therapeutic targets for TC.

Key Words: *Thyroid cancer, Hub Genes, Bioinformatics analysis*

Integrated Bioinformatics Analysis of Key Genes in Hepatitis B Virus-Associated Hepatocellular Carcinoma

Adel Rezvani Sichani[1], Ziba Rezvani Sichani[1], Elham Naderi Semiromi[1], Mohammad Rezaei[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

The e majority of primary liver cancers in adults worldwide are hepatocellular carcinomas (HCCs, or hepatomas).Hepatocellular carcinoma (HCC) is one of the most lethal cancers globally.The purpose of this study is to find the key genes in Hepatitis B Virus-Associated Hepatocellular Carcinoma Gene expression profiles of GSE55092, GSE94660, GSE121248 from the GEO database were obtained and analyzed.There were 120 HCC samples and 138 nontumour tissue samples Functional assessment of DEGs was performed by Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG).enrichment analyses showed that these DEGs were mainly enriched in cell division and DNA replication biological processes, nucleoplasm and microtubule cellular components, protein binding molecular functions, and cell cycle and DNA replication pathways.Key genes were selected by the protein-protein interaction (PPI) network and further validated by GSE25599 clinical data. Through protein– protein interaction analysis,14 Key gene DEGs were indicated,including CDKN3,CCNA2,PBK,CCNB2,CCNB1,MAD2L1,FEN1,CDC20,TPX2,BUB1B,TOP2A,RRM2,ECT2,CDK1. that closely related to the of hepatitis B. In addition, in comparison with normal tissues, the expression levels of TOP2A, MAD2L1, CDC20, and CCNB2 were higher in HBVrelated HCC. In conclusion, our results suggest that CDC20 and TOP2A might serve as a key gene for prognosis and as a therapeutic target for HBVassociated HCC

Key Words: *Hepatocellular Carcinoma, Hepatitis B, Key gene, Bioinformatics Analysis*

Investigation of structural properties of lunasin with computational tools

Fatemeh Safakhah[1], Mahboobeh Zarrabi[1]*, Fakhri Sadat Hoseyni[1]

[1] *Biotechnology Group, Al-Zahra University, Tehran, Iran*

E-mail address (corresponding author): fatemehsafakhah75@gmail.com

Abstract

Lunasin is a soy-derived chemical that slows the proliferation of newly identified cancer cells. This substance's anti-cancer properties have gotten increased attention .In the treatment of disorders, lunasin has been found to have cholesterol-lowering and hypertensive effects. Lunasin is a 43-amino-acid peptide having a SKWQHQQDSCRKQLQGVNLTPCEKHIMEKIQGRGDDDDDDDDD sequence and a molecular weight of 5.5 KDa .Tumor cells can adhere to the extracellular matrix due to the RGD motif found in lunasin. In fact, by fighting for adhesion, the peptides in this motif can suppress cancer.

The properties of lunacin were investigated using computational tools using bioinformatics tools. The second structure of lunasin was derived using the database http://cib.cf.ocha.ac.jp/bitool/MIX, then its third structure was analyzed utilizing the ITASSER database due to laboratory restrictions and the tiny size of this sequence. Its structure is seen using the Spdb viewer software. The Co - factor server was used to find the peptide's functional area. The hotspot wizard server looked at possible locations for examining mutations in the structure of lunasin. It was tested utilizing a prosper site and a peptide cutter to make greater utilization of enzymes. The anticp database was used to investigate cancer qualities, as well as other ACPHP properties. The amino acids S1, R1, R3, T3, K2, K24, K29, K12, and D18 from the region that plays a role in stability can be altered, however the functional region cannot. The results revealed that this peptide possesses anti-cancer capabilities, with KHIMG receiving the highest score of 87%. QGRGDDDDDD has also been discovered to have antihypertensive characteristics, with an 81 percent score. CNBr enzymes were found to be capable of selective cleavage of the lunasin peptide Chymotripsine, Clostripain.

Key Words: *Computing tools, cancer, soybeans, active peptides, components*

The rational design of stabilizing mutation in IMPDH1 by molecular dynamics simulations: a structural study

Samira Sattari[1], Razieh Yazdanparast[1]

[1] *Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*

E-mail address (corresponding author): sattari_samira@alumni.ut.ac.ir

Abstract

Regulation of purine concentrations is essential for cell growth and proliferation and any imbalance will lead to diseases such as Retinitis Pigmentosa (RP). Inosine monophosphate dehydrogenase (IMPDH) catalyzes the rate-limiting step in the GTP biosynthesis pathway. Regulation of IMPDH activity is by the CBS domain with a role in the self-assembly of the enzyme in the form of filaments. ATP increases filament formation and enzyme activity, while GTP inhibits the enzyme via compression and stabilization. Thus, the main goal in this study was to inhibit the enzyme via creating stabilizing mutations in the ATP binding site. Based on the computational approach, an I157V mutation was expected to lead to the enzyme stabilization and inhibition due to both CBS intrinsically disordered conformation and the impaired ATP binding site. The effect of rational design of point mutation was validated by CD spectroscopy, intrinsic and extrinsic fluorescence, and thermal denaturation analysis. Overall, we suggest that I157 mutation to V157 is an important target in drug design for RP disease by increasing enzyme stability, reducing dynamics required for activity, and thus the enzyme inhibition.

Key Words: *IMPDH, Retinitis pigmentosa, Stabilizing mutations, Molecular dynamics*

In silico Molecular dynamics of lead nitrate interaction with replication compounds of E. coli

Najmeh Shadfar[1], Mahboobeh Zarrabi[1]*

[1] *Department of Biotechnology, Faculty of Biological Sciences, Alzahra University, Tehran, Iran*

E-mail address (corresponding author): shadfarnajmeh75@gmail.com

Abstract

The spread of pollutants, including heavy metals such as pb^{2+}, in water resources, has negatively affected many marine and freshwater ecosystems and, consequently, their habitats, especially fish. Lead is a major used extensively in industry. Most heavy metals can be soluble in water and this is a big threat. The main mechanism of toxicity involves the production of free radicals. Which causes damage to biological molecules. Escherichia coli is a Gram-negative, facultative anaerobic bacterium. It is bacteria normally live in the intestines of healthy people and animals. Bacterial replication enzymes have vital functions, disruption of each of which can cause serious damage to the bacterial genome. The replication process of DNA is carried out with high efficiency and precision by DNA polymerases. The replicative enzyme in E. coli is DNA Pol III, which is a complex of 10 subunits that coordinates replication strands. Computational simulations have been applied to investigate various facets of DNA polymerase structure and function. The identified in the DNA Pol III, the two divalent Mg^{2+} ions that are essential for catalysis. In this work, the binding properties of pb^{2+} on DNA Pol III holoenzyme complex have been studied based on molecular dynamics. GROMACS a molecular dynamics simulation software, version 5.0.1 has been used to derive dynamic and thermodynamic properties.
The results obtained in the study revealed the possible attachment sites of lead ions and their interactions with domains that have metal bonding properties and examining the obtained parameters, we found that the binding of lead to enzymes is very stable and causes a change the configuration second structure.

Key Words: *Heavy Metal Poisoning; Simulations; Molecular Dynamics; Docking; Lead*

Bioinformatics approach for potential inhibitory of pyrogallol in ovarian cancer by Cdc25A targeting

Azin Shirani[1], Somayeh Reiisi[1]*

[1] *Department of Genetics, Faculty of Basic Sciences, Shahrekord University*

E-mail address (corresponding author): azinnsh@gmail.com

Abstract

Introduction: Pyrogallol, one of the natural polyphenols, was known to have anti-inflammatory and antitumor effects in some cancers. However, the underlying antitumor mechanisms of pyrogallol, still remain unclear so far. Cell division cycle 25 A (Cdc25A), is one of the most vital cell cycle regulators, and positively controls the functions of CDKs that lead to cell cycle progression. Overexpression of Cdc25A promotes tumorigenesis, and is observed in ovarian cancer. Therefore, the present study aimed to determine the potential therapeutic effect of pyrogallol for Cdc25A inhibition.
Methods: Pyrogallol structure was drawn in the HyperChem software. Cdc25A protein structure was retrieved from the RCSB PDB database. For molecular docking and preparation of ligand and target protein, Autodock 4.2 was used. Cdc25A protein structure docking studies were performed with this software. Nonpolar hydrogen atoms were assembled and fixed in ligand.
Result: According to the molecular docking studies, pyrogallol showed high binding energies with Cdc25A protein, with maximum values of -34.22 kJ/mol. pyrogallol mostly interacts with H, R, K, S,G, I amino acids.
Conclusion: This evaluation concludes that pyrogallol may be used as an anticancer drug for ovarian cancer as it can suppress Cdc25a proteins and may stop the cell proliferation in cancers. Pyrogallol is recommend as an effective, safe, and commercial drug to inhibit tumor progression in patients with ovarian cancer.

Key Words: *Key words: Pyrogallol; Cdc25A; ovarian cancer; in silico*

Evaluation of BRCA1 transcription in Glioblastoma Disease through Bioinformatics analysis

Sara Soltani Tehrani[1], Melika Golshayan[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): sarahsoltani13777731@gmail.com

Abstract

Glioblastoma (GBM) is a malignant (cancerous) brain tumor that develops from a specific type of brain cell called an astrocyte. Glioblastomas are often very aggressive and grow into surrounding brain tissue In most cases, the exact underlying cause is unknown. There is currently no cure for glioblastoma. Treatment is palliative and may include surgery, radiation therapy, and/or chemotherapy. Therefore, this study aimed to identify the genetic biomarkers for GBM with the concentration on the expression of BRCA1. This study is a theoretical type of bioinformatics, by referring to related databases including GEPIA2, genecards, mirwalk, mirdsnp, mirnasnp. Probably, the overexpression of BRCA1, has-miR-143 and has-miR-504 have an effective role in case of tumor survival. Increaseing the expression of BRCA1 increases the risk of tumorigenesis and metastasis while decreasing its expression reduces the number of cells involved in glioblastoma. It can possibly be used as a diagnostic confirmatory marker and predictor of cancer progression by increasing metastasis, reducing intercellular adhesion, and inhibiting cell death.

Key Words: *GBM; miRNA; Biomaekers*

Comparison of spatial distribution patterns of SNPs throughout the evolution

Forough Taheri[1]

[1] *Department of Genetics, Faculty of Basic Sciences, Islamic Azad University Shahrekord Branch, Shahrekord, Iran*

E-mail address (corresponding author): brightness_ta@yahoo.com

Abstract

Single Nucleotide Polymorphism (SNP) is the most common type of genetic variations which include only substitution of one basepair within the DNA sequence. Associations have been found between SNPs and many illnesses in human; while the severity of illness and how the body responds to the treatments are also associated with variations caused by the SNPs. Around 4 to 5 million SNP incidents have been found in human genome which are detected using bioinformatics tools and various methods such as SNP microarrays. Previous Studies have shown that SNPs are not monotonously distributed along the DNA or do not follow a normal distribution pattern. SNPs are in fact clustered, while the density of those clusters is higher in X chromosome than the Autosomal chromosomes. around 25.4% of SNPs have at least a neighboring SNP within 25 bp. The proportion of adjacent SNPs increase sharply as we move towards the shorter mutual distances. do the mutation hotspots or the SNP clusters change their location on the genome through the evolution process? and is there pattern for this movement?
Advanced and high throughput methods of genome sequencing have made much more data available regarding SNPs on various species. The present study aims to investigate the spatial distribution of SNPs in genomes of few selected species and tries to explore the SNP hotspot movement patterns and the differences of their distribution throughout the evolution process. As we find the spatial distribution of SNPs which somehow shows the mutational hotspots, this possibility may arise that significant patterns for SNP cluster movements through chromosomes of various parts of the genome emerge. This could in turn lead to exploration of new evolutionary trends and its potential association between the spatial concentration of SNPs within various genotypes and their physical phenotypes observed through the evolution process of species.

Key Words: *Single Nucleotide Polymorphism, spatial distribution, mutation hotspot, evolution pattern*

Bioinformatic analysis on regulation of GGH gene in paranoid schizophrenia patients

Mehrnoosh Tavakoli[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

The Gamma-Glutamyl Hydrolase (GGH) gene spans 24 Kb on chromosome 8q12 in human and contain 9 exons. This gene produces γ-glutamyl hydrolase which is a lysomal enzyme involved in the metabolism of folates and anti-folates. An association between folate deficiency and schizophrenia has been reported. In this study, we analyzed the distinctive transcriptome alterations of GGH gene in paranoid schizophrenia bioinformatically, which can uncover new genetic associations. The data related to this gene was obtained from GEO datasets based on GSE93987 microarray gene expression study. The gene ontology and genome pathway analysis were performed by DAVID and KEGG databases respectively. The interaction among proteins was constructed through STRING database. Based on DAVID disease-class, GGH gene is involved in immune, neurological diseases. The volcano plot resulted from GEO analyses showed differentially expression of GGH gene between paranoid schizophrenia samples and unaffected comparisons. The log2fold change was -1.887E-1 for GGH gene that means its down-regulation in the patient samples. The hypofunctional 677C>T variant of the MTHFR gene has been associated with symptom risk in schizophrenia, as has Low serum folate levels. According to the STRING pathway, the MTHFR and GGH have correlation link in metabolic pathway, carbon metabolism and one carbon pool by folate. The interactions of 20 genes associated with GGH retrieved from the GeneMania database and the network analysis showed their most connectivity in physical interactions. Our findings match with the evidence for 5 genes (TRIP6, GDI1, OS-9, PARK7 and UGDH) related to schizophrenia disorder. It is concluded that the GGH gene and correlated pathway can play important roles in human neurodegenerative diseases caused by certain abnormalities of protein synthesis, metabolic pathway, DNA synthesis, and gene expression.

Key Words: *Gamma-Glutamyl Hydrolase; Paranoid Disorder; Folate Pathway*

Molecular docking study of the quenching mechanisms of luminol through interaction with ssDNA aptamer

Raheleh Torabi[1], Hedayatollah Ghourchian[1]*

[1] *Laboratory of Bioanalysis, Institute of Biochemistry & Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): r_torabi@ut.ac.ir

Abstract

Generally, luminol as a signal producing element in chemiluminescent aptasensor is used for recognition of biomarkers. However, in some cases it was observed that luminol emission intensity is quenched when it came close to single stranded DNA (ssDNA) aptamer. It was hypothesized that this phenomenon could be happened due to the interaction between luminol and nucleobases in ssDNA. To realize this idea, luminol was superimposed with nucleobases by LS-align and then, luminol structure was docked with ssDNA structure by HADDOCK. The obtained models were analyzed by Discovery Studio Visualizer and Ligplot for detection of intermolecular interactions. The results revealed that due to the high structural similarity between luminol and double-ring nucleobases, luminol tends to form hydrogen bonds with pyrimidines (cytosine and thymine) rather than purines (adenine and guanine). Formation of hydrogen bonding prevents the excitation of luminol by inhibiting the reversible generation of $HOO^{\bullet}$ and $O_2^{\bullet-}$ radicals, and finally, less $L^{\bullet-}$ were oxidized to $3\text{-}AP^{2-*}$, which led to the static quenching. Moreover, docking results showed that at excited state π-π stacking are formed between π-orbitals of excited luminol and nucleotides. π-π stacking beside hydrogen-bonding could lead to Förster resonance energy transfer in which nucleotides as acceptor of energy absorb the emitted fluorescence of the excited luminol and consequently the fluorescence intensity of excited-state is quenched. Therefore, it seems that ssDNA is able to quench the luminol emission via static and/or dynamic quenching mechanism.

Key Words: *Aptamer, Docking, Luminol, Quenching*

Design, In silico analysis, expression and evaluation of hybrid and bivalent forms of scFvs against Neisseria meningitides factor H binding protein (fHbp)

Fatemeh Yarian[1]*, Mona Salahshoor Dahr[2]*, Abbas Alibakhshi[3], Shahrazad Ahangarzadeh[4], Maryam Ghane[2], Marjan Khajavi[5]

[1] Department of Medical Biotechnology,School of Advanced Technologies in Medicine, Fasa University of Medical Sciences, Fasa,Iran
[2] Department of Biology, Islamshahr Branch, Islamic Azad University, Islamshahr, Iran
[3] Molecular medicine Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
[4] Infectious Diseases and Tropical Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[5] Department of Biology, School of Basic Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran

E-mail address (corresponding author): f.yarian@gmail.com

Abstract

Meningococcal disease is an acute disease caused by the bacterium Neisseria meningitidis. Humans are the only natural hosts of this bacterium. It has several pathogenic factors, among which the factor H binding protein (fHbp) plays a vital role in the survival of the pathogen in the host. The single-chain variable fragment antibody (scFv) have been shown to potentially be well used in therapeutic and diagnostic applications against this factor. With the aim of stronger effect, in the present study, two scFvs with anti-fHbp potency were constructed in hybrid format and docked to fHbp antigen. An anti-fHbp bivalent format of the scFvs was also constructed. Both formats were obtained by expression into pET28a vector and purified by affinity chromatography. For determination of affinity of the hybrid and bivalent formats, ELISA method was used. According to bioinformatics results, one of the hybrid formats with a suitable structure was selected. The affinity of selected hybrid format was 7.6 × 10-9 M and the affinity for bivalent format was 3.77 × 10-9 M. The higher affinity of hybrid format indicates that development of scFv hybrid formats seems to be a better solution than bivalent format.

Key Words: *In silico analysis, Hybrid scFv; bivalent scFv; Neisseria meningitides; factor H binding protein*

Gold Nanoparticles as Efficient Scaffolds for Aptamer Adsorption: An In-silico Study

Hamed Zahraee[1], Zahra Khoshbin[2,3]*, Mohammad Reza Bozorgmehr[4]

[1] *Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran*
[2] *Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran*
[3] *Department of Medicinal Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran*
[4] *Department of Chemistry, Mashhad Branch, Islamic Azad University, Mashhad, Iran*

E-mail address (corresponding author): khoshbinz991@mums.ac.ir

Abstract

Aptasensors apply aptamers as recognition segments. Optical aptasensors are popular assays, due to high sensitivity. Gold nanoparticles (AuNPs) are efficient in developing aptasensors, due to easy synthesis and chemical stability. AuNPs are applied as fluorescence quencher or colorimetric indicator. Here, we theoretically study the adsorption behavior of the tobramycin-specific aptamer on the AuNP surface by molecular dynamics (MD) simulation to provide a confirmation on the ability of AuNP as an anchorage scaffold for aptamers, suitable for designing AuNPs-based aptasensors. The aptamer structure was obtained from PDB (ID: 1LC4). To have its optimum conformation, 60 ns MD simulation was done using GROMACS 2018.4. The simulated output was added to a simulation box containing a spherical AuNP with a radius of 2.5 nm. 110 ns MD simulation was performed considering AMBER-99SB-IDLN force field for AuNP. The RMSD graph reached a constant state about 100 ns MD simulation, proved that 110 ns was suitable for the study. The gyration radius (Rg) was approximately constant about 2.3 nm, very close to the AuNP radius, can be a sign for opening of the aptamer structure due to the adsorption on the AuNP surface. The RMSF proved the higher flexibility of the terminal nucleotides of the aptamer instead of the internal ones, in accordance with the distances between the center of mass (COM) of each aptamer nucleotide and COM of AuNP. Mean squared displacement (MSD) reflected the adsorption behavior of the aptamer, a potent anchoring of the aptamer on the AuNP surface after the simulation. The free energy landscape (FEL) analysis proved that the aptamer-AuNP complex is energetically favorable. Finally, the output observation by PyMOL software illustrated the aptamer adsorption on AuNP with the effective role of the non-terminal nucleotides.

Key Words: *Aptamer; Gold nanoparticle; Molecular dynamics simulation; Free energy landscape.*

Molecular Insight into the Adsorption of Tobramycin Specific Aptamer on Graphene Sheet

Hamed Zahraee[1], Zahra Khoshbin[2,3]*, Mohammad Reza Bozorgmehr[4]

[1] *Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran*
[2] *Pharmaceutical Research Center, Pharmaceutical Technology Institute, Mashhad University of Medical Sciences, Mashhad, Iran*
[3] *Department of Medicinal Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran*
[4] *Department of Chemistry, Mashhad Branch, Islamic Azad University, Mashhad, Iran*

E-mail address (corresponding author): khoshbinz991@mums.ac.ir

Abstract

Aptasensors are biosensors based on aptamers as the bio-recognition. Graphene is a two-dimensional carbon sheet, attractive for the design of fluorescence and electrochemical aptasensors. Graphene provides some anchorage sites for aptamers; hence, we theoretically study the adsorption behavior of the tobramycin-specific aptamer on graphene by molecular dynamics (MD) simulation to provide a molecular confirmation on the aptamer adsorption by graphene, suitable for the design of graphene-based aptasensors. To have an initial aptamer (PDB ID: 1LC4) configuration, 25 ns MD simulation was done using GROMACS 2018.4 [5,6] with CHARMM36-jul2021 force field. The output was added to a simulation box containing a graphene sheet, and 190 ns MD simulation was performed. The decrease in the RMSD fluctuations proved the steady adsorption of the aptamer on the sheet. The RMSF analysis proved the higher flexibility of the nucleotides at the 3´-region in comparison with that at the 5´-region. Indeed, the aptamer adsorption was first happened through its 5´-region, and then, its 3´-end, in accordance with the distance between the center of mass (COM) of each nucleotide and graphene along the Z-axis. Mean squared displacement (MSD) parameter reflected the adsorption behavior of the aptamer, a potent anchoring of the aptamer on the graphene surface after the simulation. The free energy landscape (FEL) proved that the aptamer-graphene complex was energetically favorable. The reduced density gradient (RDG) analysis clarified that the hydrogen bonding and van der Waals interactions played an important role in the aptamer adsorption on graphene. Finally, the output observation by PyMOL software illustrated that the aptamer was first adsorbed through its 5´-region, and then, its 3´-region, and finally, its structure was completely located on the graphene surface.

Key Words: *Adsorption; Aptamer; Graphene; Molecular dynamics simulation; Tobramycin*

Identification of key genes and signaling pathways in comparative transcriptome analysis of bovine mammary epithelial cells challenged with Escherichia coli

Masoumeh Naserkheil[1], Farzad Ghafouri[2], Ali Kadivar[3], Sonia Zakizadeh[4]*

[1] *Animal Breeding and Genetics Division, National Institute of Animal Science, Cheonan-si, Korea*
[2] *Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran*
[3] *Department of Clinical Sciences, Faculty of Veterinary Medicine, Shahrekord University, Shahrekord, Iran*
[4] *Animal Breeding and Genetics Department, Animal Science Research Institute of Iran, Agriculture Research, Education, and Extension Organization (AREEO), Karaj, Iran*

E-mail address (corresponding author): sonia_zaki@yahoo.com

Abstract

Substantial evidence demonstrates that the identification of pathogens is of major importance in order to correct actions, prevent transmission to other cows, reduce the risk of appearance of chronic infections, and aid to reduce inappropriate use of antibiotics, antimicrobial resistance, and cost of treatment. Escherichia coli (E. coli) is one of the most prevalent Gram-negative pathogens known that causes mammary gland infections in dairy cattle and may greatly influence animal welfare and milk composition/quality negatively. Mammary epithelial cells (MEC) act as sentinels to bacterial intrusion by producing mediators of inflammation and local defense. In this study, a public microarray dataset GSE47599 was used to explore the differentially expressed genes (DEGs) from 16 mammary epithelial cell samples containing 8 samples infected with E. coli and 8 samples as control. Differential gene expression analysis between infected and control samples was performed identifying 147 DE genes based on fold change > ±0.05 and false discovery rate <0.05, of which 13 were underexpressed and 134 genes were overexpressed in the infected samples compared to control samples. Among these genes, five genes—ISG15, CCL2, TLR2, MX1, and JAK2—were identified as the hub genes (highly connected genes), which were subjected to protein-protein interaction (PPI) network. Moreover, functional annotation and enrichment analysis identified 27, 9, and 3 GO terms related to inflammatory and immune responses in the biological process, molecular function, and cellular component categories, respectively. KEGG pathway significant enrichment analysis revealed that the DEGs participated in 44 pathways which are known to be enriched in TNF signaling pathway, IL-17 signaling pathway, Toll-like receptor signaling pathway, and chemokine signaling pathway. Therefore, the identification of disease-causing genes and genetic mechanisms related to immune or inflammatory responses helps to improve diagnosis, prognosis, and monitoring of responses to therapy.

Key Words: *Differentially expressed genes; Mammary epithelial cells; Pathway analysis; Bovine*

RNAseq analysis of Bovine infected with Mycobacterium avium subsp. Paratuberculosis in three time point

MohadeseDousti[1], Fazileh Esmaeili[1], Masoumeh Naserkheil[2], AzarRashedi Dehsahraei[3], Sonia Zakizadeh[4]*

[1] *Department of Plant Bioproducts, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran*
[2] *Animal Breeding and Genetics Division, National Institute of Animal Science, Cheonan-si, Korea*
[3] *Expert of Animal Breeding Center and Improving Livestock Production, Karaj, Iran*
[4]*Animal Breeding and Genetics Division, Animal Science Research Institute of Iran, Agriculture Research, Education, and Extension Organization, Karaj, Iran*

E-mail address (corresponding author): sonia_zaki@yahoo.com

Abstract

Johne's disease, a chronic intestinal disease of ruminants which caused by pathogen Mycobacterium avium subsp. Paratuberculosis (MAP), leads to an infection and serious economic losses for cattle production in the world. Consequently, the identification of MAP diseaseresponsive genes and their interaction can help us to improve diagnosis, prognosis, and monitoring of therapy. The present study aimed to explore the key differentially expressed genes (DEGs) and draw the protein-protein interaction networks using 8 RNAseq experiments in short (1h and 2h), mid (4h, 6h, and 8h) and long terms (24h and 72h) of infection, as well as provide the potential transcriptional markers for identifying the status of MAP infection. In total, we observed 1777 (670 up, 1107 down), 1834 (685 up, 1149 down) and 1480 (951 up, 529 down) DEGs in response to short, mid and long terms of MAP infection, respectively. DAVID Gene ontology annotation and enrichment analysis revealed these DEGs were mainly involved in immune response, inflammation response, and cell survival pathways. By constructing and analyzing three PPI networks, we found some hub genes in each term of MAP infection: TNF, NDUFB7, NDUFS8 and MRPS2 were upregulated during 1h and 2h of infection while FN1, EGFR, MET and CDK1 were downregulated. TNF gene plays an important role in regulating the innate immune response to mycobacterial infection. Among all candidate genes, the expression of EGFR and FGF1 increased during the progression of the disease. Thus, the candidate genes identified in our study can be suggested as biomarkers for the diagnosis of Johne's disease in dairy cattle.

Key Words: *Dairy cattle; John's disease; RNAseq analysis; Enrichment analasis*

Systems Biology

Bioinformatics Evaluation of Interactions amonge Gene, microRNA, and Potential Drugs through HER2-Positive Breast Cancer

Fatemeh Jalali[1], Fatemeh Hamidifar[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Breast cancer is the most common cancer in women. Due to the limitations of common cancer diagnostic tests, the introduction of biomarkers with higher specificity for the diagnosis of breast cancer is important. The aim of this study was to evaluate the bioinformatics of HER2-positive breast cancer and the resistance mechanisms in this disease caused by the mir-8 family expression. The study was done through genomics databases. The NCBI database was used to search for HER family genes in intracellular signaling pathways. Drug targets for HER2 were then assessed using the GeneCards database. Finally, the role of the mir-8 family in the pathogenesis of HER2 was determined using the MirBase database. According to the data obtained from the NCBI database, the HER family including EGFR (HER1/ErbB1), HER2 (ErbB2), HER3 (ErbB3), HER4 (ErbB4), and IGF-IR can activate several oncogenic signaling pathways to stimulate growth. According to GeneCards database, HER2 inhibitor drugs such as trastuzumab, pertuzumab, margotosumab or T-DM1 can be used for suppression of cell growth in the treatment of this cancer. These drugs block the Ras and PI3K/mTOR pathway and prevent cell division. Trastuzumab, as a HER2 antagonist, is one of the approved drugs for this cancer. The miRBase database showed that downregulation of mir141 in cancer cells has led to resistance to the drug. In addition, the miR-200c /141 cluster, both of which belong to the mir-8 family, plays an important role in the epithelial to mesenchymal transition (EMT) process. The expression of these two miRNAs is inversely correlated with HER2-positive breast cancer. Mir141 plays a prominent role as a metastasis suppressor gene. It is concluded that members of the miR-8 family by targeting HER2 involved in the growth and processes involved in breast cancer could possibly be investigated as diagnostic biomarkers in future studies.

Key Words: *Key words: HER family genes; Mir141; miR-200c; suppressive drugs*

Bioinformatics analysis of gene expression profiles to identify key genes associated with Non-small cell lung cancer (NSCLC)

Fatemeh Afkhaminejad[1], Kamran Rafiei[1], Najaf Allahyari Fard[2]*

[1] *Nourdanesh Institute of Higher Education*
[2] *National institute of Genetic engineering and biotechnology (NIGEB)*

E-mail address (corresponding author): allahyar@nigeb.ac.ir

Abstract

Background: Lung cancer is one of the most common malignancies in the world, as well as the leading cause of cancer death, with over one million people dying from it each year. Non-small cell lung cancer (NSCLC) accounts for roughly 85 percent of all lung malignancies and has a five-year survival rate of fewer than 20%. An accurate understanding of the molecular mechanisms underpinning lung cancer progression, as well as the genes (proteins) implicated in the disease's pathogenesis, is critical for the development of diagnostic and treatment techniques.
Materials and Methods: In this study, four microbial gene expression datasets GSE27262, GSE31210, GSE19188, and GSE19804 from Gene Expression Omnibus (GEO) obtained and different genes expressed) (DEGs) were analyzed between SCLC samples and healthy control samples using GEO2R analytical toolkit. In addition to the STRING database and Cytoscape software to create and analyze the protein-protein interaction network (PPI) genes of expression and reduction commonly expressed between these Four data sets and determination of hub and key genes were used. Confirmation of key gene expression and associated survival was also confirmed using the GEPIA and HPA databases. Functional enrichment analysis and functional pathway enrichment analysis for DEGs were performed by the Enrichr web server.
Results and conclusion: After gene integration, a total of 850 DEGs (including 264 overexpressed genes and 586 overexpressed genes) were identified. Findings related to the pathogenesis of NSCLC were identified and after checking the susceptibility of genes (proteins) (CCNB2, TTK, PBK, RRM2, CCNB1, CCNA2, HMMR, KIF11, and TYMS as potential targets in the pathogenesis of NSCLC disease and design of target targets. Drugs were identified against the disease, and functional enrichment showed that the cell cycle is the most important pathway involved in the pathogenesis of NSCLC with the highest significant value.

Key Words: *Lung cancer, NSCLC, GEO, Cytoscape, DEGs*

Homotopy analysis of homeostatic mechanisms in a feedforward network

Zohreh Dadi[1]*, Fatemeh Mohammadi[1]

[1] *Faculty of Basic Science, University of Bojnord, Bojnord, Iran*

E-mail address (corresponding author): dadizohreh@gmail.com

Abstract

One of the significant phenomena in the regulation of biological systems is homeostasis. Homeostasis occurs when an output variable is roughly constant while there is a variation of an input parameter on an interval. In some networks such as feedforward networks, some nodes will inherit homeostasis from other nodes. This phenomenon is studied from the point of view dynamical system by Golubitsky and his colleagues. In this study, we consider a feedforward network. By using dynamical system theory, we study the behavior of nodes near homeostasis points. It should be pointed out that if a node of the network has a bifurcation point coinciding with the homeostasis point then some switch-like response or stable clock rhythms reveals in networks. We use the homotopy analysis method to give an approximate solution of this network. This method allows us to calculate the solution of the system in the form of an infinite series.

Key Words: *Homeostatisis; Network; Homotopy analysis method.*

hsa-miR-8081 is a novel key regulatory RNA in the thyroid cancer development by regulation of HMGA2 in the transcriptional misregulation in cancer signaling pathway: high-throughput data analysis of microarray experiment

Mina Dehghani- Samani[1]*, Khatereh Firouzi-Farsani[1]

[1] *Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran*

E-mail address (corresponding author): minadehghani1372@gmail.com

Abstract

Gene interaction networks, in which genes activate and repress the transcription of other genes, are responsible for much of cellular life's complexity, and their failure can be fatal to an organism. As a result, systems biology has long sought to comprehend these networks. This research performed an integrated computational study to demonstrate a novel gene regulatory network and select a significant microRNA-mRNA interaction axis in the thyroid cancer samples.

Gene expression analysis of thyroid cancer samples was performed by microarray data analysis of the GSE33630 microarray dataset. Data analysis of microarray samples was performed by affy and limma statistical packages in R Studio (4.0.2) environment. microRNA – mRNA interaction analysis was performed by miRWalk. Pathway enrichment and Gene Ontology analysis were performed by Enrichr online software. Microarray data analysis revealed that HMGA2 has a significantly high expression (logFC: 5.225, adjusted. P. Value < 0.0001) in patients with thyroid cancer. Also, miRWalk analysis revealed that hsa-miR-8081 regulates the expression level of HMGA2 in the 3UTR region (Score: 1). The HMGA2 is a key regulatory gene in the transcriptional misregulation in the cancer signaling pathway. This protein has a significant role in the histone-serine phosphorylation (GO:0035404) in the nuclear chromosome (GO:0000228) and nuclear lumen (GO:0031981).

Based on this integrated high-throughput and RNA interaction analysis, for the first time, we demonstrated that hsa-miR-8081 could be a critical regulatory RNA in the thyroid cancer patients, by regulation of HMGA2 – a potential biomarker and oncogene of thyroid cancer - mRNA level and affecting the histone-serine phosphorylation in the nuclear lumen. It is highly recommended that a luciferase assay experiment validates the mentioned RNA interaction to achieve more accurate information about the precise role of mentioned microRNA in thyroid cancer development.

Key Words: *Systems biology; Big biological data; Microarray; HMGA2; microRNA*

Integrative gene network analysis of rice transcriptome to identification of key microRNAs involved in salinity stress

Zohreh Ebrahimi[1], Yaser Biniaz[2], Fatemeh Amiri[3], Ali Moghadam[3]*, S. Mohsen Taghavi[1]

[1] *Department of Plant Protection, Shiraz University, Shiraz, Iran*
[2] *Plant Virology Research Center, Shiraz University, Shiraz, Iran*
[3] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*

E-mail address (corresponding author): z.ebrahimi6326@gmail.com

Abstract

Salt stress is a harsh abiotic stress that decreases the crop yield dramatically. Rice (Oryza sativa), a major cereal, the main food source for many countries, is sensitive to this stress. Understanding the molecular mechanisms of rice under salt stress is a pivotal factor for developing stress-tolerant genotypes. Therefore, an integrative transcriptome data analysis was performed. For this purpose, some RNA-seq data were retrieved from European Bioinformatics Institute (EBI) database. CLC Genomics Workbench v.12 software used for data analysis. After quality control, reads were mapped to rice reference genome. The adjusted P-values (FDR < 0.01) were considered the significant and, thus, differentially expressed genes (DEGs) were used for further analysis. Gene ontology analysis of DEGs showed that the genes were enriched for response to cellular process, metabolic process, organic substance metabolic process in biological process category. Furthermore, the DEGs in molecular function category were enriched for catalytic activity, organic cyclic compound binding and oxidoreductase activity. In cellular component category, most genes were referred to cellular anatomical entity and intracellular anatomical structure. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed that pathways that involve most of DEGs are metabolic pathways, biosynthesis of secondary metabolites, and biosynthesis of amino acids respectively. The microRNA analysis showed that miR156, miR160, and miR396 were the candidates that involved the most families. Genetic engineering has been proved to be an efficient approach to the development of salinity-tolerant genotypes, and this approach will become more powerful as more regulatory elements like key microRNAs associated with salinity tolerance are identified.

Key Words: *Key words: RNA-seq, microRNA, Gene ontology, Rice, Salinity*

Interconnectedness of Metastasis and Invasion at single-cell resolution

Delaram Jebreili-Rizi[1], Modjtaba Emadi-Baygi[1]*

[1] *Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran*

E-mail address (corresponding author): emadi-m@sku.ac.ir

Abstract

Single-cell sequencing-based methods are changing whole-organism sciences by opening up new opportunities for investigating complex ecosystems, particularly cancers. It gives researchers an excellent chance to analyze the functional states of cancer cells at single-cell resolution, allowing them to investigate the functional heterogeneity of cancer cells and gain a better understanding of cancer cells as functional units performing specific biological functions in the initiation and progression of cancer. Because metastasis and invasion are mostly responsible for cancer mortality and morbidity, understanding how they interact cooperatively or competitively throughout the tumor's lifecycle is crucial. Therefore, by identifying specific patterns of gene expression in single cells, we will clarify the mechanisms behind tumor invasion and metastasis, which are critical for preventing cancer from spreading and simplifying its treatment. For this purpose, we established a computational analysis based on CancerSEA and Omnipath database by employing the Omnipath and GOfuncR packages in R software to look for genes that are shared by these two functional states and result in shared phenotypes. We found that among, 41900 cancer single cells in 25 different human cancers, there are only 5 protein-coding genes (PCGs) (including ACTB, ANXA2, CFL1, STMMN1, and YWHAZ) that are associated with 178 significant phenotypes (GO terms). According to our findings, these five PCGs may be critical genes in metastasis and invasion since they affect cell membrane, cell adhesion, cell movement, cell polarity, vesicle budding and fusion, cytoskeleton organization, and signal transduction. To be concluded, our results may pave the way for considering these five PCGs as plausible promising targets for targeted therapy.

Key Words: *Cancer; Phenotype; Single cell; Invasion; Metastasis; Functional states*

A systems biology approach shows that RUNX1 can be considered as a potential target to inhibit glioblastoma progression

Sharareh Mahmoudian-Hamedani[1], Parvaneh Nikpour[1]*

[1] *Department of Genetics and Molecular Biology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran*

E-mail address (corresponding author): pnikpour@med.mui.ac.ir

Abstract

Background: Glioblastoma is the most common brain tumor with a 5 year-survival rate of about 5% which puts it among the most life-threatening cancers. One of the notable genes that has been the center of focus in many cancer studies, specially in those related to aggressive solid tumors, is RUNX1. It has been found that RUNX1 has a dual role in cancer progression, meaning that in some cancers promotes the proliferation and progression while in some others acts as a tumor suppressor. In the following study, using gene expression profiles from Gene Expression Omnibus (GEO) database and a systems biology approach, we tried to examine the role of RUNX1 in the glioblastoma A172 cell line.
Materials and Methods: Gene expression data from GEO (GSE174634) including six samples of A172 glioblastoma cell line, in three of which RUNX1 expression is suppressed, was firstly downloaded. Differentially expressed genes were then extracted using GEO2R tool with adjusted p-value <0.05 and a network was constructed using Cytoscape. Functional enrichment analysis was performed using CluePedia plug-in. MCODE was used to screen the modules of the network.
Results: 220 genes were shown to be differentially-expressed between RUNX1- and RUNX1+ samples. A network was then constructed with 166 nodes and 568 edges. By using MCODE plug-in, four clusters were identified in the network. Reactome pathway analysis for 166 nodes showed that highly enriched pathways in the network are related to condensation of pro-metaphase chromosomes, G1/S transition, cell cycle and transcriptional regulation of TP53. The results of KEGG functional enrichment analysis demonstrated that selected genes are highly involve in p53 signaling pathway, cell cycle and FOXO signaling pathway.
Conclusion: This study showed that RUNX1 knock-down results in upregulation of pathways related to tumor suppression, and RUNX1 can be considered as a potential target for glioblastoma treatment.

Key Words: *Key words: Glioblastoma, RUNX1, Systems biology, Cytoscape, Gene expression network*

Dysregulation of LINC01985 lncRNA leads to the abnormal exocytosis by low expression of MMRN1 in the thyroid cancer patients: an integrated computational biology approach

Khatereh Firouzi-Farsani[1]*, Mina Dehghani-Samani[1]

[1] *Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran*

E-mail address (corresponding author): khatereh.firouzi1106@gmail.com

Abstract

Genomic characterization and evolutionary histories of both primary and metastatic cancer development have been made possible by high-throughput and multi-omics technologies. These tools, which give insight into cancer cells' genome, transcriptome, metabolome, and proteome, confirm our grasp of systems biology-level approaches to cancer. This study performed a high-throughput microarray data analysis to find novel regulatory genes in the thyroid cancer samples, using a systems biology approach. Furthermore, the RNA-interaction and functional enrichment analysis was performed to demonstrate the precise regulatory mechanism of the coding and non-coding RNAs found in the experiment. The GSE33630 microarray dataset was analyzed by R Studio (4.0.2). Normalization of raw data was performed by affy package. Differential expression (DE) and statistical analyses were performed by limma package. RNA interaction and functional enrichment analyses (Pathway and Gene Ontology (GO)) was performed by Enrichr online software. MMRN1 has a significantly low expression in the thyroid cancer patients, compared to control (logFC: -4.205, adj. P. Value < 0.0001). LINC00477 and LINC01985 lncRNAs have a significant co-expression with the NMRN1 mRNA (adjusted p-value: 0.005). Based on literature mining, LINC00477 isoforms can have a significant role in developing gastric cancer (by down-regulation). In this study, we demonstrate the possible regulatory role of the two mentioned lncRNAs in developing thyroid cancer for the first time by disturbance of the regulation of exocytosis (GO:0045055) in the secretory granule lumen (GO:0034774). LINC00477 and LINC01985 lncRNAs are the two novel non-coding RNAs that can regulate the expression of MMRN1. Dysregulation of the two mentioned RNAs can lead to the abnormality in the expression of MMRN1 and disturb the regular exocytosis mechanism in the secretory granule lumen in thyroid cancer patients.

Key Words: *Systems biology; High-throughput data; Microarray; MMRN1; RNA interaction*

SERPINB2 Overexpression As a Potential Contributor to Development of Multiple Myeloma And Its Predictive CeRNA Network

Ali Barani[1], Mohammad Rezaei[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Multiple myeloma (MM) is a type of bone marrow cancer in which monoclonal plasma cells undergo rapid growth, resulting in overexpression of immunoglobins. The exceptional growth of information and advances in technology in biomedical fields has been overwhelming for labratorary scientists over the last decade. Therefore, research has been moving from being strictly conducted in a real life lab environment to a 'virtual lab' environment where management and analysis of data are done.

Initially, gene expression data of MM patients (GSE146649) was obtained from the NCBI Gene Expression Omnibus (GEO) and then analyzed by GEO2R to find differentially expressed genes (DEGs). GENT2 database was then used to reinforce the possibility of correlation between the selected gene and bone marrow cancer. Through UniProt and GeneCards, gene ontology information and biological pathway involvement were understood. Furthermore, miRWalk, TargetScan and miRDB were utilized to find significant miRNA-mRNA interactions in the 3'UTR region. Additionally, the selected miRNA was searched in LncBase v.2 to find strong interactions with lncRNAs and construct a predictive ceRNA network.

Through analysis of the GEO dataset, a gene named SERPINB2 was found to be considerably upregulated (|logFC| = 4.246 , adj. P value =4.10e-04) in MM samples. The product of this gene is a protein called plasminogen activator inhibitor 2 (PAI-2) which mainly inhibits urokinase-type plasminogen activator in fibrinolysis pathway but it can also protect the cell from apoptosis by interacting with Proteasome subunit beta type-1 (PSMβ1). Analysis of possible miRNA-mRNA interactions revealed hsa-miR-182-5p as a significant interactor to SERPINB2 mRNA. This miRNA was then searched in LncBase v.2 (Bone marrow) and MIR663AHG, FAM106A and LINC00963 had the strongest interactions.

In conclusion, SERPINB2 is overexpressed in MM and forms a possible ceRNA network among hsa-miR-182-5p, MIR663AHG, FAM106A and LINC00963.

Key Words: *Multiple Myeloma; SERPINB2; Cancer; CeRNA*

A bioinformatics analysis of exosomal microRNAs related to colorectal cancer: An approach for identifying diagnostic biomarkers

Ali Khalafizadeh[1], Seyedeh Donya Hashemizadegan[1], Majid Sadeghizadeh[1], Sadegh Babashah[1]*

[1] *Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): alikh21175@gmail.com

Abstract

Background: Colorectal cancer (CRC) has shown a remarkable increase in the annual global incidence rate among prevalent kinds of gastrointestinal malignancies. Exosomes have recently been recognized as an important player in CRC pathogenesis. They are secreted nanoparticles that are present in both normal and pathophysiological conditions in all body fluids and are also effective in cell-cell communication. microRNAs (miRNAs), as the most important cargo in exosomes, are potential cancer biomarkers that can be useful in diagnosing and prognosis. To find a potent diagnostic biomarker, we checked the EVmiRNAs, a database of miRNA profiling in extracellular vesicles, and re-analyzed the expression profiles of specific exosomal miRNAs in CRC that were present in 11 samples.
Methods: The expression of exomiRs was studied using the EVmiRNA database. This database provided 535 small RNA sequencing samples of EVs from 11 sources / diseases and constructed the EVmiRNA database (http://bioinfo.life.hust.edu.cn/EVmiRNA) to show the miRNA expression profiles. We selected 10 miRNAs (hsa-miR-10a-5p, hsa-miR-192-5p, hsa-miR-191-5p, hsa-miR-92a-3p, hsa-miR-21-5p, hsa-miR-148a-3p, hsa-miR-182-5p, hsa-let-7a-5p, hsa-miR-215, hsa-miR-27b-3p) that have most expression in CRC from EVmiRNA database and used Diana miRpath V.2 algorithm (http://diana.imis.athena-innovation.gr/DianaTools) to make a heatmap.
Results: We found that among these selected exosomal miRNAs that have highest expression in CRC, hsa-miR-21-5p and hsa-let-7a-5p have high expression in pathways in cancer and P53 signaling pathway by Diana miRpath V.2 algorithm. We demonstrated that these miRNAs have specific roles in these pathways in CRC and we found that our selected miRNAs have high expression in these pathways in CRC.
Conclusion: We established that hsa-miR-21-5p and hsa-let-7a-5p as exosomal miRNAs have the potential of targeting crucial pathways in CRC that can be chosen as a target for CRC.

Key Words: *Keywords: colorectal cancer, microRNA, EVmiRNA database, Diana miRpath V.2*

Investigation of neuropeptide databases to identify and predict important Neuropeptides for pest control management

Mehrbanoo Kazemi Alamuti[1], Mohammad Majdi[1], Mohammad Reza Ghaffari[2]*, Ghasem Hossini Salekdeh[3]

[1] Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran
[2] Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran, (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran
[3] Department of Molecular Sciences, Macquarie University, North Ryde, NSW

E-mail address (corresponding author): ghaffari@abrii.ac.ir

Abstract

Pest insects can have unfavorable and negative impacts on agricultural production and the food supply. The damage caused by agricultural pests is estimated at an 18-20% loss in annual and $ 470 billion on a global scale. Although insecticides have helped minimize the impact of insect pests, chemical control entails economic, health, and environmental costs. Neuropeptides and G-protein coupled receptors are essential signaling molecules in multicellular organisms, and they regulate various physiological processes, such as reproduction, osmoregulation, growth, and development. Besides, these small peptides and their target receptors have been potent and promising targets for pest control and new environmentally friendly insecticidal agents. This study examines the features and applications of NeuroPep, NeuroPID, DINeR, and NeuroPIpred databases to identify and predict new neuropeptides in pests. Likewise, the prediction models will develop using input features like amino acids and dipeptide compositions, binary profiles, and implementing different machine learning techniques. We believe that research on neuropeptides and G-protein coupled receptors will provide more information for pest insect management.

Key Words: *Neuropeptides, Insecticide, Database, Machine Learning*

Discovering potential hub genes in gastric cancer using in silico analysis

Mohammad Hossein Habibi[1]*, Atena Sadat Hosseini[2], Sara Razi[3]

[1] *Amol University of Special Modern Technologies, Amol, Iran*
[2] *Islamic Azad University of Shahr Rey (Yadegar -e- Imam Khomeini), Shahr Rey, Iran*
[3] *Vira pioneers of modern science (VIPOMS), Tehran, Iran*

E-mail address (corresponding author): mohammadhosseinhabibi1999@gmail.com

Abstract

Gastric cancer was responsible for 7.7% of total cancer death worldwide in 2020. This research aims to discover hub genes and potential drug targets in Gastric cancer using systems biology approaches. Three sets of microarray profiles, including GSE79973, GSE13911, GSE54129, were downloaded from Gene Expression Omnibus (GEO) and analyzed through R programming (Bioconductor package including limma, affy, and oligo libraries). Differentially expressed genes (DEGs) were filtered with P_val < 0.05, Log Fc > 1, and Log Fc <-1, then a total of 217 overlapped genes (143 downregulated and 74 upregulated) were extracted, and a PPI network was constructed using Cytoscape software. Firstly, three subnetworks were identified with the MCODE plugin based on connectivity. After merging three subnetworks, 35 resulting genes were transferred to the Enricher database for enrichment analysis. The analysis was performed in four groups: the KEGG pathway, GO cell components, GO molecular functions and GO Biological process. The first result for each group was protein digestion and absorption, collagen-containing extracellular matrix, platelet-derived growth factor binding and, extracellular matrix organization, respectively. In the next step, centrality analysis with a focus on degree and betweenness was performed on overlapped genes through centiscape 2.2 plugin. The five top genes with the highest betweenness were FN1, MUC5AC, PTGS2, ATP4A, and CXCL8 and the five top genes with the highest degree were FN1, COL1A1, COL1A2, THBS2, and TIMP1.

Key Words: *Gastric cancer; systems biology; microarray profiles; differentially expressed genes; enrichment analysis*

Systems biology analysis of rice transcriptome: key genes responsive to salinity stress

Mozhde Hamidizade[1], Yaser Biniaz[2], Fateme Amiri[3], S. Mohsen Taghavi[1], Ali Moghadam[3]*

[1] *Department of Plant Protection, School of Agriculture, Shiraz University*
[2] *Plant Virology Research Center, Shiraz University, Shiraz, Iran*
[3] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*

E-mail address (corresponding author): m.hamidizade71@gmail.com

Abstract

Salt stress is one of the main abiotic stresses limiting rice production. Better understanding of the tolerance mechanisms is a significant subject in agribiotechnology in rice. In this study, transcriptome data were obtained from a European Bioinformatics Institute (EBI) database. CLC Genomics Workbench software was used for data processing. Reference genome and also rice annotation were retrieved from NCBI. To plot the protein-protein interactions and network construction, the String database was used and Cytoscape software was applied for visualizing the interaction networks. The iTAK database was used for transcription factor analysis. Hub genes were identified using Cytoscape software, with the highest number of interactions belonging to the OsJ_02628 gene with an unknown role and a score of 628. Due to the importance of hub genes, their function was evaluated. The most important of these were nucleotide linkers (Os08t0117100-01), similar to ATP-dependent RNA helicase (Os03t0827700-01), helix domains such as tetratricopeptide containing the protein (Os07t0590600-00), and the precursor (Os0000-02). In the results of ontological analysis, the highest number in biological process was related to ribonucleoprotein complex subunit organization and localization, in molecular function to transporter activity and membrane and also in cellular component to plant-type cell wall and membrane. Thus, transcription, kinase activity along with transmitters were introduced as the most important enriched groups in this study. In the TFs analysis, it was defined that the number and type of families at different time points, are almost the same. Hence, the highest numbers belonged to bHLH and WRKY families. Plants overexpressing OsWRKY45-2 have significantly lower survival rates under salinity due to the suppression of SNAC1, DREB1B, NCED4 and Rab16D genes, indicating that OsWRKY45-2 might be a transcriptional repressor of these genes. These results provided valuable information for future research of crops improvement exposed to abiotic stresses.

Key Words: *Transcriptome, Systems Biology, Gene Ontology, Transcription factor, Abiotic stress*

CAMTA gene family analysis in tomato and expression of genes responsive to developmental stages

Zohreh Hajibarat[1]*, Abbas Saidi[1], Mohammad Reza Ghaffari[2], Mehrshad Zienalabedini[2]

[1] Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[2] Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

E-mail address (corresponding author): zohreh.hajbarat@yahoo.com

Abstract

Calmodulin-binding transcription factors (CAMTAs) are recognized as one of the stress-responsive proteins. In this study, six SlCAMTA genes were selected in tomato. Chromosomal distribution, gene structure, conserved motifs, synteny analysis, phylogenetic tree of SlCAMTA genes in tomato were analyzed to further evaluate their performance. To detect expression levels in developmental stages, mRNA analysis of SlCAMTA genes were performed using publicly available expression data in the genvestigator. The aim of study was to identify and characterize SlCAMTA genes, via insilico genome-wide analysis approach. Chromosomal position indicated that SlCAMTA genes were distributed on chromosomes 1, 4, 5, and 12. Our findings showed that all genes were increased expression except SlCAMTA3 gene. The second group of SlCAMTA genes had 2 introns, while the first group contained one intron. Gene structure was similar in most proteins in each group, confirming the phylogenetic classification of SlCAMTA. Prediction of cis-elements in the promoter region of genes showed that DOF and AP2 / ERF, GATA and Homodomain had the highest cis-elements in the promoter region of SlCAMTA genes. The conserved motifs and gene structure in most proteins in each group were similar, validating the CAMTA phylogenetic classification. The analysis of synteny showed that SlCAMTA4 with SlCAMTA4.1 genes were orthologous.This study could be considered as a useful source for future CAMTA comparative studies in different plant species. The maximum number of cis-elements was belonged to SlCAMTA4 gene which this indicates that this gene is highly resistant to different stresses. This study could be considered as a useful resource for future comparative studies of CAMTA in different plant species and provide useful information for finding candidate genes in response to stress.

Key Words: *Calmodulin-binding transcription factors, gene expression, Developmental stages, gene structure*

Genome-wide transcriptome analysis reveals many alternative splicing events in peanut roots (Arachis hypogaea L.) under drought stress conditions

Zahra Ghorbanzadeh[1], Feba Jacob[2], Mohammad Reza Ghaffari[1], and Rasmieh Hamid[3]*

[1] Department of Systems Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research Education and Extension Organization (AREO)
[2] Centre for Plant Biotechnology and Molecular Biology, Kerala Agricultural University, Thrissur, India
[3] Department of Plant Breeding, Cotton Research Institutes of Iran (CRII), Agricultural Researches, Education and Extension Organization, Gorgan, Iran

E-mail address (corresponding author): rasihamid@gmail.com

Abstract

Alternative splicing (AS) is a posttranscriptional regulatory mechanism that promotes proteome diversity. AS is often observed in the context of changing environmental conditions. However, it is unclear whether or not AS plays a role in plant response to drought stress. In this work, we conducted a genome-wide survey of AS events in peanut (Arachis hypogaea) roots grown under normal and drought conditions. Under normal conditions, drought stress, and post-drought recovery, we found 234, 456, and 415 AS events, respectively, that were substantially differentially spliced. Alternative skipped exons and splice sites were the most common types of AS. These drought-treated root samples produced 856 genes with significant AS change. Enrichment analysis revealed that AS modulation of binding activity is critical for peanut root response to drought. Specifically, genes encoding the splicing regulatory components of the spliceosome pathway and the mRNA surveillance pathway were enriched in the Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis. Genes related to the splicing regulatory factor of peanut root also responded to drought stress and were alternatively spliced. In summary, our results suggest that genes with specific AS patterns could be used to improve plant adaptation to drought stress. These discoveries would open new avenues for enhancing plant stress resistance as the role and mechanism of AS in the process of abiotic stress are further explored.

Key Words: *Alternative splicing, Drought stress, Peanut, Root*

Genome wide detection and Evolution of the Myosin Gene Family in Arabidopsis (Arabidopsis thahliana)

Zahra Hajibarat[1]*, Abbas Saidi[1], Mehrshad Zienalabedini[2], Mohammad Reza Ghaffari[2]

[1] Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[2] Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

E-mail address (corresponding author): zahrahajibarat@yahoo.com

Abstract

The actin-based myosin system is essential for the organization and dynamics of the endomembrane system and transport network in plant cells. Plants harbour two unique myosin groups, class VIII and class XI. There is a little information about myosin in Arabidopsis. Here, we detected 17 myosin genes from the genome of Arabidopsis. Phylogenetic analysis indicated that Arabidopsis myosin genes could be grouped into class VIII and class XI, with three and 11 members, respectively. Analysis of transcription factor binding sites (TFBS) in the promoter region of myosin genes revealed the maximum and minimum number of TFBS in NAC and CSD, respectively. Moreover, variable frequencies of TFSB in myosin genes could indicate that these genes control different developmental stages and are involved in complex regulatory mechanisms. Gene structure analysis showed that the number of exons in myosin genes ranged from 21 to 42 and most of these genes in the same subfamily had similar exon–intron patterns. According to the results of genome distribution, myosin genes were located unevenly on the five A.thahliana chromosomes. Arabidopsis myosins genes were generated from segmental and tandem duplication. Five putative motifs in Arabidopsis myosins were identified using the Pfam and SMART database. Overall, these data provide significant understandings into the evolutionary and functional characterization of Arabidopsis myosin genes that could transfer to the identification and application of homologous myosins of other species.

Key Words: *TFBS, Arabidopsis, exon, actin, Phylogenetic, NAC*

Bioinformatic analysis of long non-coding RNA ZFAS1 that is involved in colorectal cancer via indirect correlation

Seyedeh Donya Hashemizadegan[1], Ali khalafizadeh[1], Majid Sadeghizadeh[1], Sadegh Babashah[1]*

[1] *Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

E-mail address (corresponding author): dhmanshadi@gmail.com

Abstract

Background and aim: Colorectal cancer (CRC) is the second and third most frequently diagnosed cancer in women and men, respectively, with an increasing global incidence rate. Many researchers have recently focused on long non-coding RNAs (lncRNAs) because of their abnormal expression patterns in CRC and their potential function in carcinogenesis as oncogenes or tumor suppressors. LncRNAs are non-protein coding RNA molecules with a length of more than 200 nucleotides.
Methods: To find a potent diagnostic biomarker, we determined RNA-seq expression profiles of 50 lncRNAs, selected between tumor and non-tumor colorectal tissues, and identified a signature of lncRNAs differentially expressed in CRC patients. Based on the fold-change (p<0/05), we selected 10 lncRNAs and then used research articles that focused on them, finally selecting Zinc Finger Antisense 1(ZFAS1). Furthermore, we used the non-code database (http://www.noncode.org/), and LNCipedia (http://www.lncipedia.org) and finally, we found an indirect correlation between ZFAS1 and the p53 signaling pathway.
Results: ZFAS1 has been increased in CRC and plays a role in proliferation, invasion, and EMT. ZFAS1 operates as an oncogene in CRC by deregulating p53 indirectly, causing cell cycle advancement and apoptosis inhibition. In CRC, overexpression of ZFAS1 resulted in decreased p53 protein expression, increased proliferation and cell cycle arrest, and inhibited apoptosis.
Conclusion: We established that ZFAS1 plays a role in the p53 network via deregulation of p53 and inhibition of apoptosis with indirect correlation, indicating that ZFAS1 participates in the p53-mediated apoptosis pathway in CRC.

Key Words: *Keywords: colorectal cancer, lncRNA, ZFAS1, p53 signaling pathway*

Investigation of metabolic pathways of genes related to the QTL of reproduction traits in sheep genome using gene network and gene ontology

Yasaman Zeraatpisheh[1], Saeed Zerehdaran[1] and Ali Javadmanesh[1]*

[1] *Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran*

E-mail address (corresponding author): javadmanesh@um.ac.ir

Abstract

Naturally, Fertility characteristics such as ovulation rate, litter size, total number of lambs born, stillbirth and age at first puberty are regulated by different genes with different effects in sheep and all have a significant impact on the economy of the sheep industry. The development of DNA-based marker technology and genomic tools opened an opportunity to identify and annotate functional genes associated with economic traits. In this study, QTLs related to fertility in sheep were prepared through AnimalQTL database. The genes for each QTL were then obtained from the sheep reference genome in the NCBI database. Then, in order to understand the relationship between the obtained genes, gene networks for each trait were drawn using Cytoscape_v3.8.0 software, and finally, Cytoscape software was used to interpret gene networks and study gene ontology. The results of this study showed that there are a total of 124 QTLs for the parasite resistance trait that are controlled by 388 genes. Most of these markers were mapped using methods such as the Genome-Wide Association Study (GWAS), Regional Heritability Mapping (RHM), or Single Nucleotide Polymorphisms (SNPs). Ontological analysis showed 32 biological pathways related to this trait in sheep and some pathways contributed more than others were including: positive regulation of organelle assembly, positive regulation of canonical Wnt signaling pathway, smoothened signaling pathway, neuron fate commitment, NF-KappaB binding, ovulation cycle process, regulation of viral life cycle. In this study, the ontology of genes was investigated and the metabolic pathways associated with the reproduction trait in sheep were obtained through experimentally validated QTL reports. To our knowledge, this was the first study pointing out the related pathways from known and valid QTLs. This method could be applied for other traits as well.

Key Words: *Ontology, QTL, Marker, Fertility*

The role of miR-429 in regulation the activity of colorectal carcinoma cells via autophagy using system biology approaches

Seyed Taleb Houseini[1]*, Mansoureh Azadeh[2], Arash Sattari [3], Farkhondeh Nemati[1], Roya Bishehkolaei[1]

[1] Department of Biology, Faculty of Basic Sciences, Islamic Azad University, QaemshahrBranch, Mazandaran, Iran
[2] ZistFanavari Novin Biotechnology Institute, Isfahan, Iran
[3] Department of Medical laboratory sciences, faculty of medical sciences, Gorgan Branch, Islamic Azad University, Gorgan , Iran

E-mail address (corresponding author): seyedtalebhouseini@gmail.com

Abstract

Background: Colorectal cancer (CRC) is among the most frequent cancers (3rd) and is one of the major causes of cancer-related deaths (4th) worldwide. Autophagy is a highly evolutionary conserved housekeeping process that relies on the degradation of cellular components (e.g., organelles, protein aggregates) in autophagosomes. Altered activity in this pathway (compared to normal conditions) is associated with numerous human diseases, including cancer. In this study, we investigated the role of miR-429 on the CRC cells in autophagy.
Material and Methods: In this study, For extracting miRNAs target genes, miRTareBase, miRWalk, miRDB, DIANA databases were selected for predicted purposes. Genes related to CRC cell activation were extracted from Coremine database and then the common genes were selected. Cytoscape software were used to graph the protein-protein interaction network. Finally, the signaling pathways and functional analysis of the final genes were examined using the Enrichr database.
Results: The results of miR-429 targets showed that out of 183 genes obtained in Coremine database and 1252 genes in the mentioned databases, 6 genes including ZEB1, TFA2PA, FBXW7, WASF3, CRKL and RND3 in the process of CRC cell activation are common. Finally the protein-protein interaction network was drawn and examined.
Conclusion: By examining the target genes of miR-429 in the process of activating CRC cells and mapping the protein-protein interaction network of these genes and analyzing the signaling and functional pathways, it was found that the mechanism of autophagy plays an effective role in this process. The results of this study can help to clarify the mechanism of activation of CRC cells.

Key Words: *Colorectal Carcinoma, miR-429, autophagy, system biology*

Computational Analysis of ARID1A and its correlated miRNAs Involved in Colon cancer

Seyed Taleb Houseini[1]*, Mansoureh Azadeh[2], Arash Sattari [3], Farkhondeh Nemati[1], Roya Bishehkolaei[1]

[1] *Department of Biology, Faculty of Basic Sciences, Islamic Azad University, QaemshahrBranch, Mazandaran, Iran*
[2] *ZistFanavari Novin Biotechnology Institute, Isfahan, Iran*
[3] *Department of Medical laboratory sciences, faculty of medical sciences, Gorgan Branch, Islamic Azad University, Gorgan , Iran*

E-mail address (corresponding author): seyedtalebhouseini@gmail.com

Abstract

Background: Colorectal cancer (CRC) is one of the most common gastrointestinal malignant cancers and is a leading cause of cancer-related mortality across the globe. miRNAs are a new family of endogenous non-coding RNAs, almost 20-24 nucleotides in long that arrange the expression of various genes involved in normal development as well as human disease including cancer. ARID1A is frequently deleted in multiple human tumors. It is located on chromosome 1p36.11, a region that is commonly deleted in various cancer types and suspected to contain tumor suppressor genes. In this study, we aim to focus on discovering the crosslink between ARID1A and its related miRNAs as a key oncogene in colon cancer.
Methods: The interaction between ARID1A and its target miRNA was found by implementing miRTargetLink database. Overall, 98 miRNA was found to be highly associated with ARID1A expression. Among the predicted miRNAs, three of them were nominated for their strong correlation with ARID1A in colon cancer. The candidate miRNAs were further examined for KEGG enrichment pathway analysis by using Diana miRpath V.2 algorithm.
Results: The analyzed data obtained from miRTargetLink databases showed that hsa-miR-31-5p, hsa-miR-101-3p and hsa-miR-1-3p are significantly correlated to ARID1A. Furthermore, by using Diana miRpath algorithm, we demonstrated that these miRNAs play important roles in the cell cycle.
Conclusion: Taking into account the low level of expression of hsa-miR-31-5p, hsa-miR-101-3p and hsa-miR-1-3p in colon cancer and their direct interaction with ARID1A, it can be observed that these miRNAs can severs as potent regulators of signaling pathways involved in colon cancer progression.

Key Words: *Computational Analysis, ARID1A, microRNAs, Colon cancer*

Bioinformatic analysis of differentially expressed genes in Ulcerative Colitis

Mahsa Ekhtiar[1], Pegah Javid[1], Mansoureh Azadeh[1]*

[1] *Zist Fanavari Novin Biotechnology Department, State Technical and Vocational Training Organization, Isfahan, Iran*

E-mail address (corresponding author): mazadeh@phd.iaurasht.ac.ir

Abstract

Ulcerative colitis (UC) is an idiopathic inflammatory colon disease characterized by widespread friability and superficial erosions of the colonic wall, as well as bleeding. It is the most prevalent type of inflammatory bowel disease in the world. The aim of our study was to determine the potential biomarkers as targets for treatment of UC. Thus, the gene expression profile was analyzed through GEO datasets based on GSE11223 to search out the top differentially expressed genes where analyzed in DAVID database. MicroRNAs and SNPs involved in UC were studied through miRdSNP. The LncRNADisease database revealed the long noncoding RNAs. According to GEO analysis, IL1b, FCGR3, IRAK3, and TIMP1 were considered as Up-regulated genes. The miRNAs and SNP involved in UC were hsa-miR-578, hsa-miR-122, hsa-miR-224, and rs315951 respectively. In addition, the lncRNAs H19 and BC012900 were implicated in UC as well. The Genes' biological mechanisms and pathways were enriched in interleukins and cytokines, linked to different pathways such as inflammatory responses and TNF signaling pathway. The results reveal that the occurance of rs315951 at the binding site of hsa-miR-578, at 965 nucleotides upstream of hsa-miR-224, or at 625 nucleotides upstream of hsa-miR-122 on 3'UTR of IL1RN gene, downregulate the expression of IL1RN and may cause UC disease. It means that the expression of IL1RN gene inhibits the activity of interleukin-1 by binding to IL1R1 receptor. H19 overexpression reduced Vitamin D receptor (VDR) expression significantly. VDR signaling has been considered to play an important role in the regulation of inflammation and has a vital role in intestinal epithelial barrier. Inflammatory cytokines promote BC012900, and it cause apoptosis in intestinal epithelial cells. It is concluded the mentioned genes, miRs, and SNP could raise our understanding of UC pathogenesis and IL1RN may be used as a potential therapeutic purpose for UC patients.

Key Words: *UC; differentially expressed genes; miRNA, SNP, lncRNA; Inflammation*

Nonlinear dynamics of birdsongs

Zohreh Dadi[1]*

[1] *Faculty of Basic Science, University of Bojnord, Bojnord, Iran*

E-mail address (corresponding author): dadizohreh@gmail.com

Abstract

Interaction between a nervous system and a highly nonlinear vocal device of birds, the syrinx, reveals a complex vocal behavior which is known as birdsong. Birdsong, as an animal model for vocal learning, is a research area of neuroscience. It should be noted that this is not exclusive to birdsong. The work of Zhang and Ghazanfar showed that the vocalizations of marmosets during their development are due to nonlinear interaction between the nervous system and the biomechanics involved in respiration. In this study, we study low dimensional dynamical system, a biomechanical model, which is introduced by Amador and Mindlin. As they stated, this model is important to allow designing precise stimuli for exploring the sensorimotor integration of acoustic signals. By using bifurcation theory, we study some codimension one bifurcations in this model. Investigating bifurcation in this model is important because it is related to to spectrally different sounds.

Key Words: *Vocal behavior; Biomechanical model; Dynamical system; Codimension one bifurcations.*

Prediction of Plants lncRNAs with machine Learning based approaches

Maryam Azizi[1], Ehsan Mohseni Fard[1], Kaveh Kavousi[2], Mohammad Reza Ghaffari[3]*

[1] *Department of Plant Production and Genetics, Faculty of Agriculture, University of Zanjan, Zanjan, Iran*
[2] *Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran*
[3] *Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran, Agricultural Research, Education and Extention Organization (AREEO), Karaj, Iran*

E-mail address (corresponding author): ghaffari@abrii.ac.ir

Abstract

Long noncoding RNAs(lncRNAs) play a significant role in molecular mechanisms, including transcription, post-transcription, and epigenetic regulation. Coding and non-coding sequences can be detected using available high-throughput sequencing technologies and prediction tools. Since lncRNA sequences are less conserved than mRNA sequences, homology cannot identify lncRNA transcripts by itself. Thus, detecting lncRNAs is typically done by identifying known genes based on several manual curation and removal steps of coding RNAs, but these predictions may be incorrect due to incomplete information. The presence of ORFs in lncRNA, which encode a small peptide and play an important regulatory role, adds to the detection's complexity. Machine learning is used in many of the lncRNA prediction software programs currently available to researchers. Some of these software's are used to predict protein-coding potential and are not specific to lncRNAs, and some of them were developed for particular species. Machine learning can help automate lncRNA detection of big data more accurately and quickly. A general overview of the studies on lncRNAs prediction and validation and their advantages are presented in detail in the current review.

Key Words: *lncRNAs; Machine learning; Prediction*

Investigation of FGF7 gene expression profile in 20 common cancers by R and WGCNA softwares

Reza Ghezelbash*, Zahra Rezvani

E-mail address (corresponding author): orazgan.reza@gmail.com

Abstract

Cancer is a group of diseases that involve the abnormal growth of cells that can spread to other parts of the body. There are related genetic pathways that may play an important role in the formation, spread, and metastatic processes of cancer. The fibroblast growth factor gene family, in close association with signaling pathways including PI3K / AKT, MAPK / ERK1, is an influential target for many cancers. Various studies have shown that incorrect expression of FGF can induce deformity of epithelial cells and fibroblasts, and these deformed cells cause tumors. Considering the role of FGF family, with emphasis on FGF7 gene and considering that the study and evaluation of molecular and regulatory pathways involved in cancer patients is important, so in the present study we intend to identify FGF7 gene expression in 20 common cases Cancers by R and WGCNA software.

WGCNA is a biological system method for describing patterns of correlation between genes in microarray samples. In this study, 20 cancers were examined using mathematical methods as well as R and WGCNA programming. Preliminary studies for each cancer were performed online with Geo2R. The results showed that FGF7 gene was significantly expressed in 9 cancers. WGCNA was applied to each cancer expression matrix individually, and FGF7 co-expression genes were obtained. For each cancer, the information of the modules was classified using R software and its diagrams were drawn. KEGG analysis and gene ontology were performed on each cancer for FGF7-related genes, and a network of FGF7-related regulatory pathways was plotted by Cytoscape software.

In breast cancer, the FGF7 gene module was enriched in fatty acid metabolism and in bladder, thyroid, melanoma, and prostate cancers in focal adhesion. FGF7 co-expression genes in colorectal, esophageal, and cervical cancers were enriched in the immune response pathways.

Key Words: *Gene expression profiles, Cancer, Bioinformatic, System Biology, FGF7*

Novel Insight into Pancreatic Adenocarcinoma Pathogenesis Using Three-way Interaction Model

Nasibeh Khayer[1]

[1] Skull Base Research Center, The Five Senses Health Institute, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

E-mail address (corresponding author): khayer.n@iums.ac.ir

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a lethal malignancy associated with a poor prognosis. High-throughput disease-related-gene expression data provide valuable information on gene interaction, leading to more profound insight into pathogenesis. The co-expression analysis is a common approach used to investigate gene-gene interaction. However, such an approach is too simplistic for explaining complex molecular interactions. In this study, we applied a three-way interaction model to capture switch genes whose expression levels modulate the dynamic nature of the co-expression gene relationship between two other genes. Subsequently, we determine the biological relevancy of statistically significant triplets in the gene regulatory network and biological pathways. Moreover, we validate in silico the outcome of this research in other related and unrelated gene expression datasets. The results of the current study suggest two critical biological processes which might be involved in PDAC triggering, and in addition, six crucial switch genes may pave the way for identifying drug targets.

Key Words: *Pancreatic ductal adenocarcinoma, Three-way Gene Interaction, Gene Set Enrichment Analysis, Therapeutic Targets.*

Nominal p-value: An Inaccurate Predictor of the Gene Significance

Maryam Maghsoudi[1]*, Rosa Aghdam[1], Changiz Eslahchi[1´2]

[1] *School of Biological Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran*
[2] *Department of Computer and Data Sciences, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran*

E-mail address (corresponding author): maryammagy@gmail.com

Abstract

A recent discovery in cancer research revealed that many of the randomly selected genes in some cancer types are significantly associated with patients' survival time. Studies show that this phenomenon in breast cancer is influenced by the activity of the proliferation signature and by eliminating the effect of this signature from the expression data, the association of a random gene set with survival time is dramatically reduced. In another study, it has been demonstrated that using a proliferation signature in other breast cancer datasets does not remove this association. We argue that the nominal p-value is not a good estimator for determining the significance of a random gene sets' association with survival time. As a result, we define a function that can distinguish the difference between a random gene set and published signatures. To do this, we used a random gene set denoted by *X*, to divide samples into two different groups using principal component analysis (PCA), and the groups' survival time were compared to each other using a log-rank test to obtain a nominal p-value (*pvalue(X)).* If the nominal p-value was significant, we used significance analysis of microarray (SAM) to find genes, denoted by *Y*, that are differentially expressed between these two groups. Using a method similar for random gene set *X*, we obtain p-value for set Y *(pvalue(Y))* and define the function as $f(X) = \frac{|X \cup Y|+1}{|X \cap Y|^2+1} * \max(pvalue(X), pvalue(Y))$.We consider the gene set *X* to be significantly associated with survival time if $f(x)$ is less than 0.05. By utilizing this function on 34 different cancer types, the result show that randomly selected genes that are biologically meaningless and unrelated to cancer progression and metastasis are no longer considered to be significantly associated with patient survival time.

Key Words: *breast cancer; nominal p-value; principal component analysis; survival time*

Identification of hub genes and pathways connect gut microbiome to hepatocellular carcinoma using bioinformatics analysis in mus musculus

Mohammad Amin Malekraeisi[1]*

[1] *School of Medicine, Iran University of Medical Sciences, Tehran, Iran*

E-mail address (corresponding author): malekraeisi.m@iums.ac.ir

Abstract

Background: Liver cancer was the sixth common cancer and the third fatal cancer in 2020. Hepatocellular carcinoma (HCC) included 75%-85% of liver cancer cases. Gut microbiome-induced inflammation and elevated deoxycholic acid, secondary bile acid, were associated with HCC. In a study on mice, deoxycholic acid increased development to HCC in mice receiving STHD-01. The current study used bioinformatics approaches to reveal pathways and hub genes related to HCC and the gut microbiome.
Materials and methods: GSE114400 and GSE71628 for liver transcriptome related to gut microbiome changes, and GSE23680 and GSE104627 for HCC in mus musculus were chosen from the GEO. GEO2R, edgeR, and limma packages of R software were used to identify differentially expressed genes (DEGs) with Adjusted.P.value<0.05 and |logFC|>0. Shared genes of the up-regulated gut microbiome and down-regulated genes of HCC (up microbiome&down HCC) and the intersection of down-regulated genes of the gut microbiome and up-regulated genes of HCC (down microbiome&up HCC) were determined distinctly. Genes entered Enrichr, and BioPlanet identified enriched pathways. STRING and Cytoscape were used for network construction and analysis of protein-protein interaction (PPI) networks.
Results: Up microbiome&down HCC had 44 genes. The number of genes was 51 for down microbiome&up HCC. Pathway analysis using BioPlanet detected amino acid metabolism, pyrimidine nucleotides metabolism, urea cycle, and drug metabolism as significant pathways which involved up microbiome& down HCC genes. The unwinding of DNA, DNA strand elongation, cell cycle, DNA replication, and CDK regulation of DNA replication were significant pathways for the other intersection. Analysis of PPI network suggested Csp1, Fah, Upb1, Kynu for up microbiome&down HCC genes. Aurka, Incenp, Trip13 were hub genes for the other network.
Conclusion: Understanding pathways and hub genes related to the gut microbiome and HCC interaction would help provide more efficient treatments targeting HCC based on the gut microbiome.

Key Words: *bioinformatics; gut microbiome; hepatocellular carcinoma; hub gene; pathway*

Identification of Differentially Expressed Genes in AML and ALL patients by system biology approaches

Sara Parizan[1]

E-mail address (corresponding author): saraparizan01@gmail.com

Abstract

Leukemia is a hematologic malignancy caused by increasing numbers of abnormal white blood cells in the blood-forming organs. Leukemia is classified into four main types according to cell line of origin and the clinical course of the disease. AML (acute myeloid leukemia), ALL (acute lymphoblastic leukemia), CML (chronic myeloid leukemia), CLL (chronic lymphoblastic leukemia), The initial treatment of leukemia is chemotherapy but it can cause some side effects such as immune suppression, or dysfunction of the central nervous system. Therefore according to the side effects of chemotherapy and the increasing incidence of leukemia in some countries, it is urgent to discover targeted therapeutic strategies for improving leukemia patients' health.

Nowadays, the microarray is one of the most efficient and accurate techniques used to determine potential biomarkers and molecular factors involved in the diagnosis and treatment of cancer.

In the current, we downloaded two microarray datasets from the GEO database to evaluate differentially expressed genes between AML/ALL patients using the R language package. Then functional enrichment analysis and the protein-protein interaction network analysis were performed for further investigation of screened DEGs.

In this study potential target genes and molecular mechanisms were involved in AML/ALL patients were successfully identified. These findings may represent a new approach to enhance effective strategies for patients with leukemia; however, further research is required to confirm the results of our study.

Key Words: *Leukemia, gene Expression, bioinformatics analysis, biomarker*

In silico analysis of identification miRNA regulatory network in gallic acid treatment of breast cancer cell line

Afsane Masoudi chelegahi[1], Somayeh Reiisi[1]*

[1] *Department of Genetics, Faculty of Basic Sciences, Shahrekord University*

E-mail address (corresponding author): masoudi.ir74@gmail.com

Abstract

Introduction: Gallic acid, a poly-hydroxylphenolic compound plentifully found in plants, fruits, and foods, has been described to have numerous biological functions including an anticancer effect. Gallic acid can be effect on genes and miRNA expression. miRNAs are a class of non-coding RNA that involved in various biological processes including cancers. Therefore, many studies explaining the role of miRNAs in breast cancer. Therefore this study aims to investigate putative target genes and networks where they are involved in gallic acid treatment.
Methods: The original data set GSE85871 was selected from the GEO dataset (NCBI), and then the differentially expressed miRNAs in MCF7 cell line treated by gallic acid were identified using the GEO2R. Their target genes were predicted from four (Tagetscan, miRWalk, miRDB, miRmap) miRNA target prediction databases. Then, functional analysis was accomplished for the target genes using by construction of a miRNAs target gene network.
Results: In current study, described 4 miRs (miR-127-5p, hsa-miR-511-3p, hsa-miR-500a-5p, and hsa-miR-3652) with up-regulation and 3 miRs (hsa-miR-492, hsa-miR-155, and hsa-miR-3685) with down-regulation in treatment with gallic acid. The miRNAs were exposed to the most used predictions software and >250 overlap target genes predicted. Then, enrichment analysis was performed revealing KEGG pathway, comprising cell cycle, cGMP-PKG signaling pathway, and apoptosis. A network construction was generated and links between the selected miRNAs and the predicted targets.
Cnclusions: In this study, we merged miRNA expression analysis with a bioinformatics-based workflow. Some genes (GAD2, CAPN6, TNIP3 and FOXJ1), pathways and interactions, putatively involved in breast cancer inhibition, were identified.

Key Words: *miRNA, breast cancer, miRNA network, Gallic acid treatment*

Determination of biomarkers and hub genes of psoriasis disease; a bioinformatics datasets analysis

Eqbal Mohamadi[1]*, Zarrin Minuchehr[2]

[1] *Department of Bioprocess Engineering, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran*
[2] *Department of Systems Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran*

E-mail address (corresponding author): eqbalmohamadi@gmail.com

Abstract

Psoriasis, a chronic self-inflammatory disease that affects about 3% of adults worldwide, shows diverse phenotypical and pathological symptoms such as skin plaques and joint lesions with adverse physical and psychological effects in patients. The disease has a complex immunogenetic mechanism. So novel studies thanks to high-throughput technologies and tools in the diverse layers of omics led to identification of the main biomarkers of the disease. This outcomes promising application of the results in the clinical field aiming early diagnosis and precise treatment of the disease. In this study, we analyzed several GEO datasets from the NCBI website, searching for psoriasis skin disease keywords. Firstly, we set the effective DEGs that had a higher relative expression in the disease, furthermore, the major hub genes related to the disease, including ccl20, IL-17, IL-8 were determined. Afterwards, using the STRING enrichment analyzing tool, the predicted protein-protein interactions (PPI) were identified. Furthermore, using the KEGG database, the molecular pathways of the biomarkers and their functional role, determined. The results of this study can be used in the design of ligands and drugs for drug delivery purposes alongside the therapeutic studies of inflammation in psoriasis.

Key Words: *Keywords: Psoriasis, Biomarkers, DEGs, STRING, Hub genes, Inflammation*

Drug design against influenza pH1N1 polymerase through computational biology and molecular modeling

Atousa Moradzadegan[1]

[1] *Department of Experimental Sciences, Dezful Branch, Islamic Azad University, Dezful, Iran*

E-mail address (corresponding author): atoosa803@yahoo.com

Abstract

Influenza H1N1 virus is the main cause of worldwide epidemics and annual influenza outbreaks in humans. But few drugs are available for its treatment. Consequently, researchers have been engaged in efforts to discover new antiviral mechanisms that can lay the foundation for novel anti-influenza drugs. The antiviral drugs that selectively act on RNA polymerase are less prone to resistance and possess fewer side effects on the patient. Therefore, there is increased interest in screening compounds that can inhibit influenza virus RNA polymerase. The viral RNA polymerase is comprised of subunits PA, PB1 and PB2. PA has endonuclease activity and is a common target for drug design. In this study, we employed molecular docking, molecular dynamics (MD), MMPBSA and ADME studies to detect and propose an inhibitor among 11873 structures against PA. Our molecular docking, MD and MMPBSA studies showed that ZINC15340668 has ideal characteristics as a potent PA inhibitor and can be used in experimental phase. Also, ADME prediction showed that all physic-chemical parameters are within the acceptable range defined for human use. Molecular mechanism based study revealed that upon inhibitor binding; the flexibility of PA backbone is enhanced. This observation demonstrates the plasticity of PA active site, and it should be noted in drug design against PA Influenza A viruses.

Key Words: *Influenza H1N1, Endonuclease, Molecular docking, Molecular dynamics, ADME prediction*

miRDisNet: A Meta-Database of miRNA, lncRNA and circRNA-Disease Association Resources With Graph Presentation

Mohammad Hasan Jafari Najaf Abadi[1], Seyed Hamid Aghaee-Bakhtiari[1´2]*

[1] *Department of Medical Biotechnology and Nanotechnology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*
[2] *Bioinformatics Research Group, Mashhad University of Medical Sciences, Mashhad, Iran*

E-mail address (corresponding author): aghaeibh@mums.ac.ir

Abstract

Nowadays, the role of ncRNAs, which include miRNAs, lncRNAs and circRNAs in the beginning, and progression of various diseases has been fully proven. Today, these molecules are used for prognosis, diagnosis, drug response and therapy in various diseases. Because of the importance of these molecules, many databases have been created to store ncRNA-related communications. Most of these databases are limited to relations of a single type of disease like immune disease or cancer or single category of ncRNAs. Therefore, researchers have to go to many databases and combine the data manually to achieve desired result. Moreover, due to several names for a unique disease and differences in the naming of the same disease in various databases and the diverse naming of miRNAs in different versions of miRBase, this is more difficult and the data loss is very probable. We combined the information of 17 existing databases including miRNA (BioM2MetDisease, Cardio_ncRNA, DMPred, EpimiR, HMDD, ImmunemiR, miR2Disease, miRandola, miRCancer, miREnvironment, OncomiR and PhenomiR), lncRNA (Cardio_ncRNA, Lnc2Cancer and LncRNADisease) and circRNA (circRNAdb) and minimize the loss of data by uniforming the names of diseases and miRNAs. In this research, we provide for the first time a simultaneous search for multiple ncRNAs and diseases for users, which can see the results as an intersection or union. In addition, we analyze the upregulation or downregulation of a specific ncRNA in a disease. Due to the complexity of the relationship in ncRNA-disease, tables cannot present the desired results but graphs are more suitable items that broadly used in the systems biology to display and analyze the complex networks. In this study, results will be displayed in graphs and help to provide researchers with a variety of filters to achieve results that are more favorable. miRDisNet is accessible via http://mirdisnet.ir/.

Key Words: *miRNA; lncRNAs; circRNA; ncRNA-Disease Association; miRDisNet; Graph presentation*

Analysis of floral development related gene co-expression networks in Arabidopsis thaliana

Pouya Motienoparvar[1]*, Mokhtar Jalali Javaran[1], Kaveh Kavousi[2]

[1] *Department of Agricultural Biotechnology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran*
[2] *Institute of Biochemistry and Biophysics (IBB), University of Tehran*

E-mail address (corresponding author): p.motienoparvar@modares.ac.ir

Abstract

The study of flower development is of great scientific, practical and applicable importance; because in this way, flowering control can be provided in plants. For example, it can be used in the production of male sterile plants to produce hybrid seeds. It should be noted that many regulatory genes are involved in the flowering process that have been identified previously. The interactions of genes involved with the formation of flower components are complex. In general, identifying the function of individual genes involved in flower development has not led to a complete and accurate understanding of the developmental process. In order to overcome this challenge, nine co-expression genes network of Arabidopsis taliana for two wild and three mutant ecotypes were reconstructed and analysed using WGCNA R package and cytoscape software platform. In the continuation of the research, the modules involved in trait control were identified and compared. After performing costumized RandomWalk algorithm on the co-expression networks, 10 key genes in addition to the deed genes used in this study were determined that has have important role in phase transition. Topological anf functional analysis of the networks indicated deeper changes during the phase transition phenomena in biochemical pathways (alternative passways were activated). In conclusion, focusing on the hormone alternative pathways could provide novel approaches to control flowering by genetic engineering.

Key Words: *systems biology; Arabidopsis thaliana; floral development; network analysis; community detection; flower differentiation.*

The importance of circular RNAs in hepatocellular carcinoma based on molecular pathways analysis of their microRNA sponges

Mohammad Hossein Donyavi[1], Sadra Salehi-Mazandarani[2], Parvaneh Nikpour[2]*

[1] *Department of Cell and Molecular Biology and Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran*
[2] *Department of Genetics and Molecular Biology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran*

E-mail address (corresponding author): pnikpour@med.mui.ac.ir

Abstract

Background: Hepatocellular carcinoma (HCC) is the most frequent kind of liver cancer and the fourth leading cause of cancer-related death globally. Circular RNAs (circRNAs), as a type of non-coding RNAs, have been noticed in recent years due to their important features. CircRNAs can function as microRNA sponges and transcription modulators, and can interact with RNA-binding proteins. According to studies, circRNAs can play a significant role in cellular metabolism, which can help us to identify new cancer biomarkers and therapeutic targets. In this study, we determined differentially expressed circRNAs in HCC and recognized their potentially related molecular pathways.
Materials and Methods: Two HCC datasets, including GSE94508 (5 cancerous and 5 paracancerous samples) and GSE97332 (7 cancerous and 7 paracancerous samples) were selected from the Gene Expression Omnibus (GEO) database to identify differentially expressed circRNAs (DECs). These two datasets were combined, and the batch effect was eliminated using the sva R package. The limma R package was then used to identify DECs with threshold of $|\log2FC| > 2$ and adjusted p-value < 0.05. Subsequently, the CircInteractome database was utilized to detect microRNAs interacting with the DECs based on the context+ score percentile ≥ 90. Furthermore, DIANA-miRPath v.3 was used to reveal molecular pathways relating to these microRNAs (p-value < 0.01).
Results: Nine DECs (including five upregulated and four downregulated) were identified between cancerous and paracancerous HCC samples. 57 unique microRNAs interacting with the DECs were revealed. Finally, the mirPath database demonstrated 44 significant signaling pathways such as proteoglycans in cancer, Fatty acid biosynthesis, hippo signaling pathway, and adherens junction, which are cancer-associated.
Conclusion: This study provides an overview of differentially expressed circRNAs in HCC that are linked to cancer-related pathways as microRNA sponges. Further research can uncover the precise processes by which circRNAs function in HCC.

Key Words: *Carcinoma, Hepatocellular; RNA, Circular; MicroRNAs; Bioinformatic*

A Meta-Analysis of Comparative Transcriptomic Data Reveals a Set of Genes Involved in the lignin synthesis in Nicotiana tabacum

Maryam Rashidifar[1], Hossein Askari[2]*, Ali Moghadam[3]*

[1] Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[2] Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[3] Institute of Biotechnology, Shiraz University, Shiraz, Iran

E-mail address (corresponding author): marayam.rashidifar@gmail.com

Abstract

Lignin is the second component of plant biomass and provides mechanical strength to tree trunks and confers impermeability to vascular tissues. It is mainly involved in defense mechanisms against biotic and abiotic stresses. This study was designed to explore the gene expression regulatory networks of this pathway in tobacco under drought stress conditions. We retrieved four datasets from different gene expression studies on tobacco in drought stress conditions from Gene Expression Omnibus. Preprocessed reads were aligned to the reference genome with Hisat2. HTSeq was used to count the number of reads mapped to each gene. The meta-analysis approach evaluated differentially expressed genes by the combined data (P-value ≤ 0.05). In addition, based on the R Package Weighted Gene Co-Expression Network Analysis (WGCNA), we identified some modules related to the lignin biosynthesis pathway. The gene network analysis also identified several hub genes such as FAS1and PPC2.51 which may play crucial roles in the lignin biosynthesis pathway. The previous study represents that FLAs are cell wall structural glycoproteins that mediate cellulose deposition and cell wall development and they are abundant in the xylem. It is also shown PP2C signaling cascade provides land plants with a hormone-modulated, resulting in a tolerance strategy allowing them to support tissues built of cells with thicker cell walls. In the present study, we identified several hub genes. The results showed that these hub genes may have vital roles in regulation of lignin biosynthesis. The current findings provide an overall insight into lignin biosynthesis and can expand the potential for engineering genome-scale pathways and systems metabolic engineering to alter the production of lignin by plants.

Key Words: *Phenylpropanoid pathway; RNA-seq data; Meta-analysis; co-expression; WGCNA*

Bioinformatics-Based Prediction of Recurrence in Tamoxifen-Treated Patients with Estrogen Receptor-Positive Breast Cancer

Fatemeh Nabizadeh[1´2], Farshad Qalekhani[1´3]*, Shabnaz koochakkhani[2]

[1] Pharmaceutical Sciences Research Center, Health Technologies Institute, Kermanshah University of Medical Sciences(KUMS), Kermanshah, Iran
[2] Department of Human Genetics, Faculty of Medicine, Hormozgan University of Medical Sciences(HUMS), Bandar Abbas, Iran
[3] Medical Biology Research Center, Health Technologies Institute, Kermanshah University of Medical Sciences(KUMS), Kermanshah, Iran

E-mail address (corresponding author): ghalekhani73@outlook.com

Abstract

Estrogen receptor (ER)-positive breast cancer is the most common subtype of all invasive breast cancer types. After binding with the ligand, the activated ER can promote cell proliferation while inhibiting cell apoptosis. Tamoxifen (TAM) acts as a selective ER modulator in adjuvant therapy for ER-positive breast cancer, inhibiting the proliferation of cancer cells and activating apoptosis. Although TAM treatment can drastically decrease mortality rates among breast cancer patients, about half of the patients still suffer from the recurrence of therapy resistance tumors. Therefore, identification of a molecular signature that predicts the relapse of TAM-treated patients could help the therapeutical management of ER-positive breast cancer. Here, we used network analysis to compare the responses of ER-positive breast cancer patients to TAM, as some patients developed metastasis while others showed partial or complete remission after treatment. To do this, two microarray-based gene expression profiling datasets, GSE9893 and GSE82172 were downloaded from GEO. After merging the datasets and batch effect removal, differentially expressed genes (DEGs) were obtained by comparing the expression values between recurrent and nonrecurrent breast cancer samples. Limma package was used to calculate fold changes of the DEGs. Next, we selected the genes with adjusted p-value <0.05 and Log2 fold change <-1 and >1. Functional and pathway enrichment analyses of DEGs were performed using the Enrichr web server. Following network analysis within Cytoscape software, the hub genes were identified by selecting the top 10% of nodes harboring the highest degree of connectivity, using the cytoHubba plugin. In conclusion, we found that FN1, ACTB, COL1A2, COL3A1, VIM, CTGF, YWHAZ, ACTA2, LUM, and COL4A1 act as hub genes in the TAM-responsive regulatory network. The results can be helpful in predicting the response of ER-positive breast cancer patients to TAM and identifying patients who do not benefit from Tamoxifen treatment.

Key Words: *Breast cancer; ER-positive, Microarray; Recurrence; Tamoxifen.*

In silico systems biology approach for analyzing of Genetic network and pathways in major depressive disorder

Mitra Farahmandnejad[1], Seyed Mahdi Sadati[1´4], Pouria Mosaddeghi[2], Amirhossein Sakhteman[2´3], Negar Firouzabadi[1]*

[1] Department of Pharmacology & Toxicology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
[2] Department of Medicinal Chemistry, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
[3]Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
[4] Department of Neuroscience, School of Sciences and Advanced Technology in Medicine, Hamadan University of Medical Sciences, Hamadan, Iran

E-mail address (corresponding author): mahdisms2020@gmail.com

Abstract

Major depressive disorder is a long-term relapsing condition associated with high levels of disability, mortality and reduce the quality of life. According to the World Health Organization, about 350 million people in the world are affected by this condition. An estimated 50% of depressed patients are inadequately treated by available interventions. Therefore, the purpose of this study was to systematically investigate and reconstruct the genetic network of major depressive disorder (MDD) for finding most important genes and biological pathways to assay with experimental condition in next studies and finally hypothesize new treatment.

In this study, through an extensive review of existing published studies and popular databases, all genes associated with MDD were found. Then, in order to integrate the results, all the interactions between these genes were explored and the achievement was represented as an interactive genetic network with GeneMANIA and other plugins. Furthermore, the reconstructed network was analyzed with Cytoscape. After finding most important genes, we found most important pathways with Gene Ontology (GO) and Reactome databases. Finally, we investigated diseases with the most number of shared genes with MDD by DisGeNET. It was found that GRIN2A, APP, JUN, PPP3CA and EGFR are the most central nodes in the genetic network of MDD. Functional and pathway enrichment analysis showed that GPCR signaling, cytokine and immune system signaling are the main systems in patients with MDD. By studying genes shared between MDD and other diseases, it was cleared that MDD, Schizophrenia, Alzheimer's disease and some types of cancer have the most number of shared genes.

The results of this study, in addition to reviewing the available results as a comprehensive and integrated manner, provide new hypotheses for future studies. It shows that we must use systems biology approach to manage and treat disorders. These outcomes must be done in experimental study.

Key Words: *Major depressive disorder, Systems biology, Cytoscape, Genetic network, pathway*

Bioinformatics investigation of human genes expression to bacterial meningitis pathogenesis

Hossein Sahragard[1], Mahlagha Cheraghi[1], Amirreza Hooshmand[2], Behin Omidi[1]

[1] *Department of Biology, Central Tehran Branch, Islamic Azad University, Tehran, Iran*
[2] *Department of Cell and Molecular Sciences, Faculty of Advanced Science & Technology, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran*

E-mail address (corresponding author): hossin.sahragard85@gmail.com

Abstract

Introduction: The most common bacteria that cause bacterial meningitis are Neisseria meningitis, Streptococcus pneumonia, Listeria monocytogenes, Haemophilus influenzae, and Escherichia coli. Meningitis, pneumonia, and sepsis are all life-threatening symptoms of Streptococcus pneumonia infection, which is one of the most common bacterial causes of morbidity and mortality worldwide. Bacteria enter the CSF in the subarachnoid space via crossing the blood–CNS barrier (either the blood-brain barrier through the brain parenchymal microvasculature or the blood–CSF barrier through the choroid plexus or the pial or arachnoidal microvasculature) or from a nearby infection site. The goal of this study was to figure out which genomic pathways and hub genes were turned on during bacterial meningitis.
Method and material: GSE40586, a microarray dataset, was obtained from the Gene Expression Omnibus (GEO) There are 22 bacterial meningitis samples and 18 control samples in this collection. The transcriptome analysis console (TAC) was used then utilized to normalize and assess the genes with differential expression (DEGs). DEGs between normal and BM samples were chosen based on adjusted p-value (FDR) -< 0.05 and |log2 FC|>-2. Protein-protein interaction (PPI) and visualization were created using String, Cytoscape, and Gephi, respectively.
Results: DEGs were obtained for 624 genes (316 upregulated, 308 downregulate). Our research identified five hub genes, namely, IL CD4 (T-cell surface glycoprotein cd4), CD8A (T-cell surface glycoprotein cd8 alpha chain), KCNA3 (potassium voltage-gated channel subfamily A member), ITGAM (integrin alpha-M), and LCK (Tyrosine-protein kinase Lck). Furthermore, the results of the KEGG pathway analysis revealed that these genes were enriched in significant pathways Th1 and Th2 cell differentiation, T cell receptor signaling pathway, and Hematopoietic cell lineage.
Conclusions: In conclusion, the findings of this study may aid in the development of new targets for medication discovery and therapy of bacterial effects

Key Words: *Keywords*

Systems biology, Bioinformatics, Gene network analysis, biomarker

Deep neural network prediction of interplay between lncRNAs and salinity response in rice

Raheleh Mirdar Mansuri[1], Amir-Hossein Azizi[1], Zahra-Sadat Shobbar[1]*

[1] *Agricultural Biotechnology Research Institute of Iran (ABRII), Department of Systems Biology, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran*

E-mail address (corresponding author): shobbar@abrii.ac.ir

Abstract

Salt stress seriously constrains growth and fertility of rice worldwide. Long noncoding RNAs (lncRNAs) play crucial roles in plant abiotic stress response, however, no systematic screening of lncRNAs under salt stress in rice has been reported. Herein, a transcriptional study using RNA-seq was conducted which resulted in the identification of differentially expressed lncRNAs (DE-lncRNAs) in FL478 as salt tolerant rice genotypes compared to its susceptible parent (IR29). Initially, a total of 15131 and 16256 potential noncoding RNAs were respectively identified in FL478 and IR29, among which 8724 (FL478) and 9235 (IR29) transcripts with length of > 200 bp were nominated as lncRNAs. Applying a strict pipeline, four and nine DE-lncRNAs were respectively detected in FL478 (2 up- and 2 down-regulated) and IR29 (6 up- and 3 down-regulated) under salt conditions, while only 2 DE-lncRNAs were in common among both genotypes. Using ATAC-seq data, the genomic regions of all four lncRNAs in FL478 and 6/9 in IR29 were confirmed to be significantly accessible for transcription. To identify the potential functions of salt-responsive DE-lncRNAs in cis-regulation, protein-coding genes were initially searched 100 kb upstream and downstream of these lncRNAs; then DE-lncRNAs and their neighboring mRNA were subjected to co-expression network analysis. We further explored potential functions of salt-responsive DE-lncRNAs to identify trans-regulatory networks of lncRNAs in each genotype. The crosstalk between DE-lncRNAs and miRNAs were subsequently investigated by exploring the DE-lncRNAs acting as target mimic of known miRNAs in Oryza sativa. WGCNA analyses identified 49 modules in the two genotypes. Interestingly, lncRNA.2-FL in FL478 was highly correlated with 173 mRNAs in the transcriptional module M39, whereas this module was not significant in IR29. Furthermore, we used a deep neural network trained on the rice genome to predict the effect of nearby mutations on the expression of mRNAs. We identified many important regulatory elements and showed the causal effect of lncRNA.2-FL on the expression level of a differentially expressed gene in the tolerant genotype. The identified DE-lncRNAs might be involved in rice response to salt stress, and lncRNA.2-FL may play a role as a regulatory hub in the salt stress tolerance of FL478.

Key Words: *Oryza Sativa; Salinity; LncRNA; co-expression; weighted correlation network analysis*

In-silico analysis of a missense SNP (rs1805323 C>T) in human PMS2 gene

Sahar Yaqubi[1]

[1]*Department of Molecular and Cell Biology, University of Mazandaran, Babolsar, Mazandaran, Iran*

E-mail address (corresponding author): saharyaqubi2000@gmail.com

Abstract

Research has shown that the endometrial cancer is the most common type of uterine cancer, affecting most often women over 55. The incidence rate of endometrial cancer is increasing rapidly, so much so that the cancer is estimated to increase by more than 50% worldwide by 2040. About 5% of endometrial cancers are linked to hereditary factors. One of the genes playing a significant role in endometrial cancer is PMS2. PMS2 (PMS1 homolog 2), which is located on chromosome 7p. It encodes a protein that functions to correct DNA mismatches and deletions that can occur during DNA replication and homologous recombination. There is, however, little experimental research into the relationship between SNPs occurring within the PMS2 and endometrial cancer. To address this gap, missense SNP and its FASTA sequence were first retrieved from NCBI. It was subsequently evaluated through the following seven software: SIFT, POLYPHEN-2, PORVERN, PhD-SNP, SNPs&GO, I-Mutant, and HOPE. The purpose was to determine whether the retrieved SNP affects the stability and functions of the protein encoded from the PMS2 gene. According to I-Mutant report, a decrease in the stability of the encoded protein happens during the single-nuclotid polymerization. HOPE confirms I-Mutant's prediction that the wild-type residue is more hydrophobic than the mutant residue. While the mutant residue charge is positive, the wild-type residue charge was neutral. As to the functions of the encoded protein, SIFT predicts that this SNP deleteriously affects the protein function. HOPE confirms SIFT prediction: the mutated residue is located in a domain that is important for the main activity of the protein. Hence, the mutation of the residue might disturb this function. According to the results of this study, with the decrease in the stability and disturbance in the function of the protein in the uterine cells, the probability of endometrial cancer increases.

Key Words: *endometrial cancer; PMS2; single-nucleotide polymorphism (SNP); mutation effect prediction*

Finding the cross-tissue biomarkers gastric cancer diagnosis by transcriptome high throughput data analysis

Kiarash Zare[1´2], Hossein Pourmontaseri[1´3], Pardis Mohammadi Pour[4], Mohammad Mehdi Naghizadeh Jahromi[1]

[1] Student Research Committee, Fasa University of Medical Sciences, Fasa, Iran
[2] Computational systems biology circle, Fasa University of Medical Sciences, Fasa, Iran
[3] Bitab knowledge enterprise, Fasa University of Medical Sciences, Fasa, Iran
[4] Phytochemistry Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

E-mail address (corresponding author): kiarashzare1998@yahoo.com

Abstract

Gastric cancer (GC) is one of the most common gastrointestinal malignancies. Although several approaches have been developed to diagnose GC in recent years, still, there is a long distance to provide rapid, non-invasive, and inexpensive detections. The present study aimed to detect cross-tissue biomarkers in tissue, blood, and salivary samples of GC patients. The GC-related terms were searched in array express and Gene Expression omnibus to collect and analyze appropriate datasets. Each dataset was analyzed with GEO2R tools to calculate differentially expressed genes. Then, a defined formula [-log(p.value)×|log(fold change)|] was applied to select the 2000 top significant probs. In the end, we selected the intersection genes which were repeated at least in four out of eight tissue datasets. At last, we integrated two GC blood and salivary datasets and found the common genes in all datasets. The analysis showed 190 and 144 common genes in GC tissue and blood datasets. The evaluation of VENN diagram results demonstrated that common gene expression between tissue and blood, salivary and blood, tissue and salivary, was 6, 17, and 17. Also, the common genes between tissue, salivary, and blood were 4 genes, including AKR1C1, SORBS2, CKM, and PDLIM3. Our finding indicated that explored genes would manipulate oxidoreductase activity, production of structural proteins, and transferase activity in favor of GC pathogenesis. We concluded that four candidate common genes which have been explored in the present study between salivary, tissue, and blood samples of GC patients would develop promising diagnostic approaches in the future. If laboratory investigations confirm the results of the present study in the future, it is possible to obtain rapid and inexpensive tests to accurately diagnose GC patients.

Key Words: *gastric cancer, biomarkers, transcriptome, data analysis, diagnostic approach*

The Role of Tumor Microenvironment Communication in Cancer Progression Using Single-Cell RNA Sequencing: A System Level Study

Fereshteh Fallah Atanaki[1], Hesam Montazeri[1], Kaveh Kavousi[1]*

[1] *Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): kkavousi@ut.ac.ir

Abstract

Tumor microenvironments are complex ecosystems comprised of cancer cells, infiltrating immune cells, and other cell types. Diverse cell types interact with each other and collectively create complex networks of intercellular and intracellular signals. In fact, these networks determine cancer progression as well as response to therapy. Recent advances in single-cell transcriptomics have laid the foundations for studying tumors at the microscopic resolution. This technology provides a compelling strategy to fill gaps in knowledge of human tumors and uncover complex networks between diverse cell types in a tumor ecosystem. In this study, we proposed a framework to explore the molecular mechanisms in different stage of cancer by integrating single-cell RNA-sequencing (scRNA-seq) data and a multi-layer network that integrate information from different levels. The heterogeneous multi-layered network provides a great way to represent the biological system's hierarchy: from gene to cellular function and final phenotype. It allows to uncover novel relations between biological entities on a microscopic view. The pipeline was applied to glioma scRNA-seq data and complex network analysis (topology and functional analysis) of the different networks from different cancer stages revealed important signaling module, important ligand-receptor connection, and most significant ligand – receptor-TF axis. Differential network analysis between stage III and IV glioma discovered the most important nodes and edges in the rewiring of interactions. In addition, hidden hierarchies in data have been revealed by heterogeneous multi-layered network models.

Key Words: *Key words: Tumor microenvironment; complex networks; heterogeneous multi-layered network; scRNA-seq*

Integrating bioprocessing engineering and metagenomics approach to optimize fast production of enriched biocomppost and humic acid from rice straw

Reza Sharafi, Gholamreza Salehi Jouzani, Ebrahim Karimi, Hossein Ghanavati, Mojegan Kowsari

Microbial Biotechnology Department, Agricultural Biotechnology Research Institute (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

E-mail address (corresponding author): gsalehi@abrii.ac.ir

Abstract

The objective of the present study was to optimize fast production of enriched biocompost and humic acid from rice straw at lab and pilot level using bioprocess engineering and metagenomics approach. The effective bacteria and fungi with high hydrolytic activities were isolated and characterized from simulated composting process based on their cellulase, xylanase and amylase activities. The 16SrDNA and ITS sequencing showed that the selected strains were belonged to Bacillus licheniformis, Nocardiopsis alba, B. subtilis and Thermoascus aurantiacus. The effects of different materials, including rice straw (RS), chicken manure (CM), urea, olive waste (OW), zeolite, biochar and two groups of native microbial boosters were assayed using 8 treatments and 3 replicates at lab level (9 kg) during 60 days. Carbon/nitrogen ratio, pH, EC, temperature, and macro/micro element contents and humic acid production were measured during the process. Then, two selected treatments, including K and M (containing RW, CM, OW, zeolite, biochar and microbial booster 1 and 2 (B. licheniformis, N. alba, B. subtilis and T. aurantiacus or Trichoderma sp, respectively) and the control (RW) were used for the pilot level (500 kg). The treatments K and M showed maximum temperature increase (up to 67 °C), NPK contents, C/N ratio reduction (72.11 and 65.07¬%, respectively) and maximum humic acid production (310 and 264 g/kg) compared to the control (36 g /kg). Moreover, the humic acid extracted from the treatment K showed maximum positive effects on wheat growth indexes. Metagenome analysis showed that two phyla of Proteobacteria and Firmicutes (with a frequency of 67% and 27%, respectively) were dominant in the treatments K and M during three composting phases, whereas genera Sphingobacterium and Pseudomonas were predominant in the control (A). The genus Bacillus was dominant in K and M. Alpha analysis showed that the richness and evenness in the maturation and mesophilic phases were greater than in the thermophilic phase. Moreover, the beta analysis revealed that maturation phase in the treatment M had a similar variety to mesophilic phase in the K. Interestingly, microbiome of the control was not significantly changed during the three temperature phases. Based on the functional analysis, the maximum enzymatic activities were observed in the thermophilic phases in K and M. Among the strains added to the composting process, T. aurantiacus and B. subtilis had significant effects on the decomposition of lignin and cellulose during the process.

Key Words: *Rice straw, Biocompost, Humic acid, Metagenomics*

ANNUAL BIOINFORMATICS CONFERENCE 2022

1st International and 10th National Iranian Conference on Bioinformatics

22-24, February 2022

University of Tehran
Kish International Compus