ISCB Student Council

SYMPOSIUM 2023

19th ISCB STUDENT COUNCIL SYMPOSIUM

LYON, FRANCE | HYBRID

23RD / JULY / 2023

PROGRAM BOOKLET

#SCS2023 SCS2023_ ISCB_SCS ISCBSC

Contact:
symposium@iscbsc.org
www.scs2023.iscbsc.org

CONTENTS

WELCOME TO SCS 2023

Dear Participants,

Welcome to the 19th ISCB Student Council Symposium (SCS2023)!

We are thrilled to host this year's symposium, bringing together students and early career researchers in Computational Biology and Bioinformatics. SCS2023 offers a unique platform to present your work, connect with peers, and foster professional growth.

As the first-ever hybrid SCS, we have carefully crafted a program that combines captivating keynote speeches, engaging roundtable discussions, oral talks, flash talks, and poster presentations. This diverse range of sessions ensures a dynamic and inclusive environment for knowledge exchange and collaboration.

Beyond the scientific program, we have organized an exclusive networking session to facilitate meaningful connections among participants. This session presents an opportunity to expand your network and forge long-lasting relationships within the computational biology community. We thank Harvard Medical School for sponsoring this social event.

We extend our heartfelt gratitude to all participants, sponsors, and volunteers for their invaluable contributions in making SCS2023 a reality. We would also like to thank the International Society for Computational Biology (ISCB) for their unwavering support.

We hope that SCS2023 will inspire and empower you as you embark on this exciting journey. Let us embrace collaboration, innovation, and community as we collectively shape the future of computational biology.

Welcome to SCS2023, and we wish you a fruitful and memorable symposium!

Best regards,

Syed Muktadir Al Sium and Laura Veschetti
Chairs, Organizing Committee
ISCB Student Council Symposium 2023

ORGANIZING TEAM

Chairs

Syed Muktadir Al Sium
Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh

Laura Veschetti
IRCCS San Raffaele and University of Verona, Italy

Program Committee

Sanjana Fatema Chowdhury (Chair)
Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh

Gokce Senger
European Institute of Oncology, Milan, Italy

Rubaiat Ahmed
University of Dhaka, Bangladesh

Aakriti Jain
Department of Biophysics, University of Delhi, New Delhi, India

Mirko Treccani
CMLab, University of Verona, Italy

S. M. Salim Shadman
BRAC University, Bangladesh

Mallikarjuna Thippana
University of Hyderabad, India

Design Committee

Arsalan Riaz (Chair)
Precision Medicine Lab, Peshawar, Pakistan

Shahriar Hossain
Noakhali Science and Technology University, Naokhali, Bangladesh

Omar Faruk
Guest contributor

Finance & Fellowship Committee

Katherin Eliana Alejandra Otalora Acevedo (Chair)
Universidad de Barcelona, Spain

Md. Eashanul Karim Moon
Jashore University Science And Technology, Bangladesh

Jaime Tamayo
Universidad Nacional Mayor de San Marcos, Peru

Website Committee

Talukder Caleeb Shahriar (Chair)
University of Dhaka, Bangladesh

Estafania Torrejon
Nova Medical School, Portugal

Othmane Hayoun Mya
Barcelona Supercomputing Center (BSC-CNS), Spain

Social Media Committee

Tarhima Jahan Jerin (Chair)
Bangladesh Agricultural University, Bangladesh

Tanha Islam
Noakhali Science and Technology University, Bangladesh

Giulia Cesaro
University of Padova, Italy

Maharaf Hossain Rahat
BRAC University, Bangladesh

Shoaib Saikat
University of Barishal, Bangladesh

Miguel Fernandez
Barcelona Supercomputing Center (BSC), Spain

Juan Carlos Zavaleta Lopez
UPAO, Peru

Steering Committee

Pradeep Eranti
Université Paris Cité, France

Gabriel J. Olguin-Orellana
Universidad Talca, Talca, Chile

Wim Cuypers
Institute of Tropical Medicine Antwerp, Belgium

R. Gonzalo Parra
Barcelona Supercomputing Center, Barcelona, Spain

First version, July 2023

This booklet was designed and edited by Sanjana, Gokce, Rubaiat and Arsalan on behalf of #SCS2023. And this was supervised by Syed Muktadir Al Sium. For latest updates please visit the symposium page at https://scs2023.iscbsc.org/. This booklet was inspired by the program booklet of #ASCS2022 (https://zenodo.org/record/7934696) as a template.

Disclaimer
The ISCB Student Council has made all efforts to provide accurate information but does not guarantee the correctness of any information provided in this booklet. The ISCB Student Council is a committee of the International Society for Computational Biology (ISCB), which is incorporated as a 501(c)(3) non-profit corporation in the United States.

Supported By

The International Society for Computational Biology (ISCB) is a recognized 501(c)3 organization based in the United States of America. ISCB is a global society is aims for advocating and advancing scholarship, research, training, outreach, and inclusive community building in computational biology and its professions. Society membership reflects commitment toward the advancement of computational biology. The ISCB is an international non-profit organization whose members come from the global bioinformatics and computational biology communities. The ISCB serves its global membership by providing high-quality meetings, publications, and reports on methods and tools; by disseminating key information about bioinformatics resources and relevant news from related fields; and by actively facilitating training, education, employment, career development, and networking. We advocate and provide leadership for resources and policies in support of scientific endeavors and to benefit society at large.

ISCB Student Council, the student organization of the International Society for Computational Biology. The mission is to promote the development of the next generation of computational biologists. We achieve our goal through provision of scientific events, networking opportunities, soft-skills training, educational resources and career advice, while attempting to influence policy processes affecting science and education.

Harvard Medical School, one of the prestigious sponsors of SCS2023, is a world-renowned institution at the forefront of medical education, research, and patient care. Established in 1782, Harvard Medical School has consistently set the standard for excellence in medicine, producing exceptional physicians, scientists, and healthcare leaders. With its unwavering commitment to advancing knowledge and improving health worldwide, Harvard Medical School has made groundbreaking discoveries and pioneered innovative medical interventions that have transformed the field of medicine. The institution's renowned faculty members and researchers continue to push the boundaries of medical science, working across disciplines to unravel complex diseases, develop novel treatments, and improve healthcare delivery. As a sponsor of SCS 2023, Harvard Medical School demonstrates its dedication to fostering collaboration and promoting the next generation of scientists and computational biology professionals.

ACKNOWLEDGEMENTS

The success of Student Council Symposium 2023 (#SCS2023) relies on the dedication of numerous individuals. We express our gratitude to all those who participated in organizing this event throughout the year, regardless of whether their involvement was a brief task or months of devoted work. For some efforts, we are extraordinarily grateful, and we want to specifically acknowledge them, as they have gone above and beyond, and their efforts deserve special mention:

- Without the logistical support and invaluable advice of ISCB Chief Executive Officer Diane Kovats and Lead Technologist Seth Munholland the #SCS2023 would not have been possible. We sincerely appreciate their continued support.
- Special thanks to Steven Leard and the ISMB/ECCB 2023 organising committee.
- We also thank ISCB-SC Executive Team for their guidance and timely advice.

#SCS2023 would like to thank our

- Keynote speakers: Prof. Anais Baudot and Prof. Burkhard Rost,
- Round table discussion panellists: Dr. Alex Bateman, Dr. Lucia Peixoto, Dr. Dan DeBlasio and Dr. Farzana Rahman.
- Abstract reviewers for volunteering their time to contribute to this symposium's success and promote the next generation of Bioinformaticians and Computational Biologists.
- All the authors who shared their valuable works with the scientific community through SCS2023.

Furthermore, we would like to thank everyone on the organizing and steering committee for their time and effort in realizing this event.

BIOS

KEYNOTE SPEAKERS

PROF. ANAIS BAUDOT

Aix Marseille Univ, CNRS Marseille, France
Barcelona Supercomputing Center, Spain

Prof. Baudot is a CNRS director of research in Computational and Systems Biology, Marseille, France. Since 2018, she has been leading a research group dedicated to Systems Biomedicine in the Marseille Medical Genetics unit (https://www.marseille-medical-genetics.org/a-baudot/). Prof. Baudot's main objectives are to better understand the genotype-phenotype relationships of genetic diseases thanks to the analysis and integration of biology.

PROF. BURKHARD ROST

Department of Informatics, Bioinformatics, and Computational Biology, Technische Universität München, Munich, Germany

Prof. Burkhard Rost is the director of the Department for Computational Biology & Bioinformatics at the Faculty of Informatics of the Technical University of Munich (TUM). He is Professor of Computer Science & Computational Biology & Bioinformatics. He leads the Rostlab, whose goal is to predict important aspects of protein structure and function using sequence information, evolutionary information and results from other predictions, as well as improving the effectiveness and efficiency of structural genomics projects' ability to determine the structures of proteins on a large scale.

ROUND TABLE PANELISTS

Dr. Alex Bateman

Senior Team Leader and Head of Protein Sequence Resources, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)

Alex Bateman is a computational biologist and Head of Protein Sequence Resources at the European Bioinformatics Institute (EBI), part of the European Molecular Biology Laboratory (EMBL). He has led the development of the Pfam biological database and introduced the Rfam database of RNA families. His team also delivers the InterPro and RNAcentral resources. He is PI for the UniProt project at EMBL-EBI. He is currently Editor-in-Chief of Bioinformatics Advances.

Dr. Lucia Peixoto

Associate Professor, Department of Translational Medicine and Physiology, Washington State University, USA

Dr. Peixoto possesses a broad background in molecular biology, biochemistry, genomics and computational biology. Her research focuses on using genomic and computational biology approaches to study brain function. Peixoto completed her Bachelor's degree in Biochemistry, and later on obtained her Ph.D. She has received multiple awards, and completed her postdoctoral training at The University of Pennsylvania. Peixoto also trained at the Training Program in Neurodevelopmental Disabilities, and completed a clinical internship at the Center for Autism Research. Her research interest is in understanding the underlying molecular basis of Autism Spectrum Disorders and their comorbidity with Intellectual Disability and Sleep Impairments. Her lab uses mouse models for functional genomic studies of behavior and patient samples to study genetic contributions to disease. Besides these, her work includes developing bioinformatics approaches for genomic data analysis that yield precise and reproducible results.

Dr. Dan DeBlasio

Assistant Professor of Computer Science, University of Texas at El Paso, USA

Dr. DeBlasio's research interests are broadly in applied machine learning as well as algorithm design and analysis, and he takes inspiration from biological problems. Many times this not only leads to an interesting algorithmic result, but a useful biological tool. He also uses machine learning techniques in applications outside computational biology, most recently in identifying unresolved space objects using hyper- and multi-spectral telescope imaging. DeBlasio received a PhD in Computer Science from the University of Arizona as well as a MS and BS in CS from the University of Central Florida. He was previously a member of the Computational Biology Department at Carnegie Mellon University, where he will be rejoining from Fall 2023. He was a member of the ISCB Student Council, serving many leadership roles, until 2019 and has been a member of ISCB for more than 10 years.

Dr. Farzana Rahman

Assistant Professor, Faculty of Engineering, Computing and Environment, Kingston University London, UK

Farzana is an Assistant Professor in Computer Science at Kingston University London, UK. Her research focuses on evolutionary genomics, proteomics and natural crisis modelling using ML, deep learning, and cloud computing. She is actively involved in improving computational pedagogy utilising the Wikipedia knowledge base. She is an open-source science advocate and an experienced international STEM conference organiser. She is a co-chair of the International Society for Computational Biology's (ISCB) Wikipedia Committee. She is also a founding member of the ISCB publication committee. Farzana served as a Chair of ISCB-SC in 2016-18, followed by a 3-year term as an elected Board of Directors at the ISCB. Farzana has been recognised in 2023 by the EPSRC-WES UK as a promising Women in Engineering Ambassadors.

PROGRAM

Time (CET)	Sunday 23rd July	
9:00 - 9:10	**Introduction and Welcome words**	
9:10 - 9:55	**Keynote Lecture 1**	**Multimodal data integration for rare genetic diseases** Prof. Anais Baudot
9:55 - 10:15		**Coffee Break**
10:15 - 10:30	**Oral Talk 1**	**Computational resources for understanding the binding affinity of membrane protein-protein complexes and mutants (in person)** Fathima Ridha Karuvanthodikayil
10:30 - 10:45	**Oral Talk 2**	**Development and Application of the MultiSEp R Package to Identify Multiple Myeloma Achilles' Heels for Drug Discovery (in person)** Adeline McKie
10:45 - 10:50		**Technical Break**
10:50 - 11:05	**Oral Talk 3**	**Exploring Endogenous Peptides for Development of Safer Opioid Analgesics (virtual)** Md. Shahadat Hossain
11:05 - 11:20	**Oral Talk 4**	**Improved Prediction of Protein GO terms in the CAFA5 competition (virtual)** Zong Ming, Chua
11:20 - 11:35	**Oral Talk 5**	**ATLAS-AML: An Automated Bioinformatics Pipeline for Target Characterization in Acute Myeloid Leukemia (virtual)** Suraj Bansal
11:35 - 11:40		**Technical Break**
11:40 - 11:45	**Flash Talk 1**	**Impact of oral anti-diabetic drugs on gut-derived extracellular vesicles: Proteomic Signature (in person)** Estefania Torrejón
11:45 - 12:00		**Break**
12:00 - 12:15	**Oral Talk 6**	**Differential Production of Abnormal Isoforms in Brain Cells Contributes to Neuronal Death and Parkinson's Disease: Gender-Specific Differences (virtual)** Waqar Hanif
12:15 - 12:30	**Oral Talk 7**	**Decoding the functional roles of intronic microRNA Hsa-Mir-2355 and its host gene KLF7 in the immunobiology of cervical cancer (virtual)** Nure Sharaf Nower Samia
12:30 - 12:35	**Flash Talk 2**	**Changes on the Structure of Microbial Communities of the rhizosphere of peruvian fruit trees (Annona cherimola Mill. and Pouteria lucuma) across depth soil using PacBio HiFi sequencing (virtual)**

		Angie Tatiana Porras Valencia
12:35 - 12:40	**Flash Talk 3**	**Comprehensive annotation of miRNAs and lncRNAs in domesticated cotton species (virtual)** Vivek AT
12:40 - 12:45	**Flash Talk 4**	**mICKEY: Memory-Efficient Deep Learning for Personalized Biomarker Discovery and Cancer Origin Prediction from DNA Methylation Data (virtual)** Kasidech Aewsrisakul
12:45 - 12:50	**Flash Talk 5**	**Most frequently harboured missense variants of hACE2 across different populations exhibit varying patterns of binding interaction with spike glycoproteins of emerging SARS-CoV-2 of different lineages (virtual)** Rubaiat Ahmed
12:50 - 12:55	**Flash Talk 6**	**Gene subset signatures for complexity reduction of technical and functional comparisons of whole transcriptomes (in person)** Shruti Gupta
12:55 - 13:00	**Flash Talk 7**	**Virus-host interactions in a municipal landfill include non-specific viruses, hyper-targeting, and interviral conflicts (in person)** Nikhil George
13:00 - 14:10	**Lunch | Poster Session & Networking**	
14:10 - 14:55	**Keynote Lecture 2**	**Decoding the language of life** Prof. Burkhard Rost
14:55 - 15:10	**Oral Talk 8**	**Heterogeneous Domain Adaptation for Species-Agnostic Transfer Learning (in person)** Youngjun Park
15:10 - 15:25	**Oral Talk 9**	**CClens: a cellular communication workflow for large-scale single-cell RNA sequencing data (in person)** Giulia Cesaro
15:25 - 15:40	**Oral Talk 10**	**Low input capture Hi-C: a method to decipher the molecular mechanisms underlying non-coding alterations (in person)** Laureano Tomás-Daza
15:40 - 15:55	**Oral Talk 11**	**P53 orchestrates spatio-temporal epigenome rewiring to transcriptionally prevent malignant transformation (in person)** Monica Cabrera-Pasadas
15:55 - 16:25	**Exploring posters and Discussion with presenters**	
16:25 - 16:35	**Introducing ISCB Student Council activities**	
16:35 - 17:35	**Roundtable discussion**	**Exploring the Potential of AI in Revolutionizing Bioinformatics Research: Opportunities and Challenges.** (Dr. Alex Bateman, Dr. Lucia Peixoto, Dr. Dan DeBlasio, Dr. Farzana Rahman)
17:35 - 17:55	**Closing remarks**	
17:55 - 18:00	**All on stage for picture/photo of the event**	

ORAL TALKS (TITLE: A-Z)

ATLAS-AML: An Automated Bioinformatics Pipeline for Target Characterization in Acute Myeloid Leukemia

Suraj Bansal[1], *Andy Zeng*[1], *Amanda Mitchell*[1], *John Dick*[1]

[1]*Princess Margaret Cancer Centre, University Health Network*

Acute myeloid leukemia (AML) is an aggressively heterogeneous disease with poor survival outcomes. In AML, adverse genomic profiles and leukemia stem cell (LSC)-enriched cellular hierarchies are often linked to chemoresistance and frequently observed at relapse. Although single-cell and bulk transcriptomics have reshaped our understanding of hematopoiesis and AML, subsequent analyses pose technical barriers for tailoring interpretation for scientists' ongoing drug experiments.

To bridge computational and experimental drug development in AML, we introduce ATLAS-AML, an automated bioinformatics pipeline for transcriptomic meta-analysis of genes and gene signatures in AML. The ATLAS-AML pipeline is available as a containerized web application that experimental scientists can readily employ without bioinformatics expertise. ATLAS-AML integrates publicly available single-cell and bulk RNA-sequencing datasets with preconfigured pipelines to streamline five concomitant levels of visualization for queries: expression across normal and leukemic hematopoietic hierarchies, enrichment in functionally-validated LSC+ fractions, and correlation to disease relapse, clinical characteristics, and overall patient survival. For example, ATLAS-AML showed that DNMT3B, a well-documented transcriptomic hallmark of AML, was overexpressed in primitive AML cell types and associated with LSC+ fractions, disease relapse, FLT3-ITD mutations, GATA2-MECOM alterations, and worse patient outcomes.

Moreover, ATLAS-AML constitutes a powerful framework for accelerating target discovery. Reinterrogating our datasets using differential expression, we identified 282 novel target candidates. Querying each target in ATLAS-AML, we shortlisted candidates that were associated with disease-propagating LSCs and biologically distinct subsets of AML patients. For example, ATLAS-AML suggested that CNST, a trans-Golgi network receptor for targeting connexins to the plasma membrane, is enriched in malignant LSC populations and associated with adverse cytogenetic alterations, lending to CNST's therapeutic viability in AML.

Altogether, ATLAS-AML enables scientists to leverage insights from single-cell and bulk transcriptomics to inform preclinical studies towards risk-tailored treatments in AML.

CClens: a cellular communication workflow for large-scale single-cell RNA sequencing data

Giulia Cesaro[1], *Giacomo Baruzzo*[1], *Barbara Di Camillo*[1]

[1]*University of Padova*

BACKGROUND

Cellular communication plays a crucial role in controlling and regulating many biological processes, such cell development and tissue functionality, and diseases, as cancer progression. The advent of single-cell transcriptomics has enabled the study of cellular communication and several computational tools have been developed for inferring ligand-receptor interactions.

As single-cell transcriptomics has become cheaper, widespread and accessible, the availability of large-scale studies (i.e. cell atlases) of increasing complexity (different conditions, different subjects and time series studies) poses new challenges in cellular communication analysis, such as i) new biological questions to answer, e.g. identify changes in crosstalk across distinct contexts, ii) the increased computational demand and iii) the visualization and interpretation of results. Therefore, there is the need of a generalizable and scalable workflow to perform and support the interpretation of cellular communication analysis from large-scale single-cell RNA data in a user-friendly, efficient, and effective way.

DESCRIPTION

We propose CClens, a bioinformatic pipeline, that first comprises the quantification and characterization of cell-cell communication in each distinct context at both inter- and intra-cellular level. Then, it enables the identification of alterations in the communication patterns across distinct contexts, exploiting ad-hoc statistical methods to work with any multi-condition scenarios, including dataset where only information about experimental condition is available (i.e. multi-condition scenario) or coupled with patients' ID (i.e. multi-patient scenario). To handle the increased computational burden, we use advanced data structures from the bigmemory R package and the possibility to integrate C++ code from the Rcpp package, achieving both an efficient in-memory computation and exploiting shared-memory parallelism. Lastly, an R/shiny interface offers multiple functionalities (e.g. filtering options, advanced visualization tools) to inspect, summarize and interpret complex cell-cell communication data in user-friendly, accessible (no-code) and flexible way.

CONCLUSIONS

Single-cell transcriptomics is a quite young and fast evolving research area, having a huge impact in biological data analysis advancement, and cellular communication analysis represents an unprecedented opportunity to characterize biological systems. We believe that CClens will facilitate the analysis and interpretation of cell-cell communication, making it a valuable tool to gain new insights about biological processes that govern a multicellular system or different experimental conditions.

Computational resources for understanding the binding affinity of membrane protein-protein complexes and mutants

Fathima Ridha Karuvanthodikayil, *Department of Biotechnology, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India- 600036*

BACKGROUND

Membrane proteins (MPs) mostly function as complexes and the interaction between the proteins is are dictated by their strength of binding or binding affinity. Due to their intricate structure, however, the binding affinity of membrane proteins is less explored compared to globular proteins. Mutations in these complexes affect their binding affinity, as well as impair critical functions, and may lead to diseases. Despite an increase in experimental affinity data in the literature, they are dispersed, necessitating their compilation into a comprehensive database for further analysis. Also, experimentally determining the affinity of these complexes is expensive and time-consuming; making them infeasible for their large-scale applications. Therefore, there is high demand for accurate computational approaches to determine the affinity of these complexes.

RESULTS

We developed the first and specific database, MPAD (Membrane Protein complex binding Affinity Database), which contains experimental binding affinities of membrane protein-protein complexes and their mutants along with sequence, structure, and functional information, membrane-specific features, experimental conditions, as well as literature information. The current version of MPAD contains 5376 entries, which includes 1705 wild-type and 3671 mutant data. MPAD has an easy-to-use interface and options to build search queries, display, sort, download, and upload the data are among the other features available to users.

Using this database, we have developed the first ML-based method, MPA-Pred, for predicting the affinity of novel MP complexes. Our method showed a correlation and MAE of 0.83 and 0.91 kcal/mol, respectively, using the jackknife test on a set of 114 complexes. Classification of complexes based on membrane protein type and function improved the performance of the method.

CONCLUSION

MPAD is the first database for the binding affinity of membrane protein-protein complexes. The database can be used to understand the factors influencing the binding affinity in MPs as well as the impact of mutations on binding affinity. We have also developed a multiple regression-based method to predict the affinity of novel MP complexes. Thus, we anticipate that these resources can help contribute to an in-depth understanding of MP complexes which may have potential applications to drug design and also for further analysis in different directions.

Decoding the functional roles of intronic microRNA Hsa-Mir-2355 and its host gene KLF7 in the immunobiology of cervical cancer

Nure Sharaf Nower Samia[1], *Rafeed Rahman Turjya*[1]

[1]*BRAC University*

MicroRNAs (miRNAs) are short, noncoding RNAs involved in post-transcriptional gene regulation. Evidence of their roles in carcinogenesis and cancer progression is expanding gradually, implying their crucial involvement in diagnosis and therapy. How miRNA dysregulation may affect cervical cancer development, which is highly prevalent in women with poor prognosis, remains largely unclear. The purpose of this study was to reassess the studies that focused on miRNA expression in cervical cancer and determine whether intronic microRNAs have a role in the tumorigenic pathway by altering their host gene targets. Using available gene expression data from the GEO dataset GSE145372, we identified the dysregulated intronic miRs and corresponding host genes in cervical cancer patients. Among these, we selected the intronic miRNA Hsa-Mir-2355 and its transcription factor-encoding host gene KLF7, both differentially expressed in these patients. Our findings indicate that miRNAs play a significant role in the invasion and metastasis of cervical cancer by affecting specific signalling pathways. In silico modelling revealed that Hsa-Mir-2355 regulates KLF7 target genes ATG12, KRAS, PRKAR1A, and REL, which can otherwise influence cervical cancer prognosis through control of angiogenesis, apoptosis, and metastasis.Furthermore, we identified REL as a survival gene in cervical cancer, linked to patient survival and CD4+ recruitment, suggesting that it could be employed as a checkpoint for cervical cancer treatment.

Development and Application of the MultiSEp R Package to Identify Multiple Myeloma Achilles' Heels for Drug Discovery

Adeline McKie[1], *Mark Wappett*[2], *Benayu Priyanto*[1], *Lisa Crawford*[1], *Ian Overton*[1]

[1]*Queen's University Belfast,* [2]*Almac Discovery*

BACKGROUND

Almost all Multiple Myeloma patients relapse and ultimately succumb to therapy-resistant disease; there is urgent need for more effective treatment. Achilles' heel relationships arise when the status of one gene exposes a cell's vulnerability to perturbation of a second gene, such as chemical inhibition, providing opportunities for precision oncology. While there is a significant focus on genetic approaches, transcriptome data has advantages for the investigation of gene dependency relationships and remains relatively underexplored. Available multiomics resources for this purpose include the Cancer Dependency Map (DepMap) and the Multiple Myeloma Research Foundation (MMRF) CoMMpass study.

DESCRIPTION

We developed MultiSEp for integrative discovery of candidate gene dependency relationships in multiomics data. Clustering of samples by expression of one gene allows partitioning of another gene's CRISPR scores, mutations or gene expression to investigate signatures of synthetic lethality. MultiSEp performed well in benchmarking against other methods and we predicted multiple myeloma gene dependency relationships at genome-scale (27,288 genes, 372,303,828 candidate interactions) with CoMMpass data (n=859 patients). Following multiple filtering steps we derived a high-confidence predicted synthetic lethal network (8,695 edges, 5059 genes; SynLethNet), including characteristic mutual exclusive loss patterns (binomial $q<0.05$). We predicted the population coverage achieved by drugging SynLethNet genes, for example inhibiting a hub is predicted to achieve therapeutic cell killing if any neighbouring gene is mutated in the cancer cells. Our analysis only utilised deleterious mutations predicted by the variant effect prediction tools VARITY (score >=0.99), SNPeff and SNPsift (annotated 'high-impact' mutations). Of ten hubs with predicted therapeutic coverage >5%, two achieved 58% coverage and at least one is an attractive candidate drug target (coverage=14%). Functional annotation of SynLethNet revealed many genes involved in the ubiquitin-proteasome system, which is dysregulated in Multiple Myeloma and a target of current front-line therapy. Predictions were validated with the Cancer Dependency Map and the Cancer Therapeutics Response database. The MultiSEp R package will soon be available from CRAN and has passed 63-unit tests.

CONCLUSION

We present the MultiSEp R package, demonstrated with a case study in multiple myeloma where we predict candidate drug targets and provide mechanistic insights to advance precision oncology.

Differential Production of Abnormal Isoforms in Brain Cells Contributes to Neuronal Death and Parkinson's Disease: Gender-Specific Differences

Waqar Hanif, *NUST*

Parkinson's disease is the second most common neuropathological disorder having a considerable effect on public health with a wide range of complexities. It accounts for about 8.5 million cases worldwide, being highly prevalent among men with poor prognosis. Neuronal cell death and axonal injury are hallmarks of various neurodegenerative disorders including Parkinson's disease. This study aims to identify novel therapeutic targets against male-specific Parkinson's disease exploring isoform variants of dysregulated apoptotic neuron genes among men. RNA-seq pipeline was employed on sequencing reads (collected from NCBI GEO) to identify differentially expressed genes between Parkinson's disease men and women followed by functional enrichment analysis of dysregulated genes using enrichR. Subsequently isoform switching analysis of dysregulated neuronal genes was performed by IsoformSwitchAnalyzeR. Finally binding interactions between wild and variant structures of abnormal transcripts and Death Receptor 4 were analyzed using HDOCK.

Resultantly, 453 differentially expressed genes including 121 upregulated and 332 downregulated genes were obtained however aberrant isoform variants were identified for two under-expressed apoptosis-regulating neuron genes namely HSPB1 and CASP7, and single over-expressed apoptosis-modulating neuron gene TNFSF10 in men compared to women. Alternatively spliced isoforms elucidated the considerable role of post-transcriptional modification on altered structural and functional characteristics leading to loss of apoptotic activity that may play a neurodegenerative role. Furthermore, dysregulated neuronal genes have significantly afflicted neurological mechanisms such as tumor necrosis factor receptor binding and cysteine-type endopeptidase activity indicating neurodegeneration in PD males. Additionally molecular docking analysis revealed strong binding interactions between wild type TNFSF10 and Death Receptor 4 leading to increased apoptosis in neuronal cells. Whereas isoform variant TNFSF10 revealed loss of neuroprotection due to weak interaction between ENST00000420541 (generated protein) and Death Receptor 4. Based on the neurodegenerative role of wild and variant isoforms, TNFSF10 has been proposed as a potential diagnostic and prognostic biomarker that can be targeted therapeutically to ensure early neuroprotection against Parkinson's disease in men.

Exploring Endogenous Peptides for Development of Safer Opioid Analgesics

Md. Shahadat Hossain[1]*, Shafiqul Islam*[1]*, Mohammad Mehedi Hasan*[1]*, Md Ackas Ali*[2]*, Mohammad A. Halim*[2]

[1]*Division of Infectious Diseases and Division of Computer Aided Drug Design, The Red-Green Research Centre,* [2]*Department of Chemistry and Biochemistry, Kennesaw State University, 370 Paulding Avenue NW, Kennesaw, GA 30144, USA*

BACKGROUND

Opioid analgesics are widely used for pain management, but their efficacy is often accompanied by various adverse effects, such as tolerance, dependence, addiction, and overdose. The Mu opioid receptor (MOP) is the primary target for opioid analgesics, but its activation can also lead to unwanted effects through β-arrestin recruitment and receptor desensitization. Biased agonism of MOP presents a promising approach to developing safer and more effective opioids. Recently, peptide-based opioids have garnered attention, with over 25 endogenous peptides discovered that bind to MOP to regulate pain and other bodily activities.

DESCRIPTION

This study aimed to evaluate the potential efficacy of 80 G protein-coupled receptor (GPCR) endogenous peptides from human and animal sources as clinical therapeutics using a computational approach. The preliminary analysis involved screening the peptides through molecular docking, and the best five candidates, including Big Endothelin, Glucagon, Kisspeptin, Leptin, and Galanin, were selected for further investigation. Molecular dynamics simulations with POPC membrane-embedded conditions were conducted to evaluate their binding affinities and interaction patterns. The results suggested that Glucagon has the most promising potential as a clinical therapeutic. Glucagon exhibited proper binding with the binding pocket and showed significant interaction with active site residues, including TRP318, THR216, CYS 217, and other binding pocket residues such as TRP226, LYS209, ASP216, and GLU310. Moreover, the analysis of the therapeutic complex formation indicated the dominance of hydrogen (62%) and hydrophobic (28%) bonds, which promote peptide-protein complex formation. Furthermore, the structure-activity relationship (SAR) analysis revealed that positively charged residues played a critical role in peptide-protein interaction.

CONCLUSION

This computational study suggests that Glucagon has the potential to serve as a safer and more effective alternative to traditional opioids for pain management. The peptide exhibited strong binding with the MOP binding pocket and interactions with active site residues, indicating promising therapeutic efficacy. These findings lay the foundation for further investigations into Glucagon's molecular mechanisms and the development of novel analgesic drugs with reduced side effects.

Heterogeneous Domain Adaptation for Species-Agnostic Transfer Learning

Youngjun Park[1], *Nils Muttray*[2], *Anne-Christin Hauschild*[3]

[1]*Universitätsmedizin Göttingen,* [2]*University of Göttingen,* [3]*Institute for Medical Informatics at the University Medical Center Goettingen*

Model organisms such as mice and zebrafish play a crucial role in developing and validating new hypotheses in biomedical research, particularly in studying disease mechanisms and treatment responses. However, due to biological differences between species translating these findings into human applications remains challenging. Commonly used homologous gene information is often incomplete data, particularly for non-model organisms. Furthermore, it entails a significant information loss during gene-id conversion because of many-to-many mapping of orthologous genes. This current issue leads to limited availability in functional analysis of cross-species study. To address this issue, we present a novel methodology for species-agnostic transfer learning with heterogeneous domain adaptation.

The baseline model is a state-of-the-art heterogeneous domain adaptation method, Cross-Domain Structure Preserving Projection, which is the heterogeneous domain version of the Supervised Locality Preserving Projection. It utilizes known common labels in domains and aligns the heterogeneous latent space. This approach allows for knowledge integration and translation across various species' datasets without relying on external human-curated knowledge. It has been evaluated with related generalized zero-shot learning models and models with gene homology.

The evaluation is done using four different single-cell sequencing datasets, focusing on the out-of-sample cell type label prediction task. There are three results: (1) comparison with the Mutual Nearest Neighbors (MNNs) to clarify the SATL task, (2) SATL performance comparison with various related methods using three human-mouse pair single-cell datasets, (3) lps-stimulated four different species macrophage dataset analysis using SATL. As a result, it can predict unseen cell types based on other species' data. It is possible to utilize all genes on cross-species analysis. We observe similar gene ontologies amongst the most influential genes composing the primary latent space axis in both species.

This demonstrates that our novel approach allows knowledge transfer beyond species barriers without gene homologies, but utilizing all possible gene sets. However, this approach heavily relies on the performance of the feature extractor and requires auxiliary class information for all unseen classes to be available. Nevertheless, SATL can provide a method for predicting analogous genes having similar functions but dissimilar sequences.

Improved Prediction of Protein GO terms in the CAFA5 competition

Zong Ming Chua[1], *Adarsh Rajesh*[1]

[1]*Sanford Burnham Prebys*

Accurate prediction of protein Gene Ontology (GO) terms is crucial for understanding biological processes, protein function, and various disease mechanisms. Current methods to produce new GO annotations is both time intensive and often misses out on important aspects of protein function. Furthermore, many proteins from less studied organisms have no known function. Methods of GO prediction have recently seen exciting new developments based on large transformer-based language models that allow unprecedented accuracy in predicting the functions of previously little understood proteins. In this context, the fifth Critical Assessment of Functional Annotation (CAFA5) competition provides a platform for benchmarking such developments.

We propose an innovative ensemble learning approach that integrates pre-trained state of the art models to produce GO predictions superior to that of any individual model. We utilize and integrate pre-trained embeddings such as those from ProtT5-XL-BFD and ProtBert to produce GO models that significantly outperform most other models currently submitted on the live leaderboard of CAFA5. Our study exemplifies the significant improvement in protein GO term prediction using an ensemble of diverse embedding models. By incorporating both sequence and structural information, we achieved a more comprehensive and accurate protein function prediction. This improvement has potential implications in accelerating biological research of little studied proteins and enhancing our understanding of disease pathogenesis by pointing the way towards a fuller understanding of biological systems. Future directions include further optimization of the ensemble model to include more accurate predictors of diverse and novel protein sequences and the extension and complementation of such ensemble learning approaches to other open problems in biological modeling such as protein-protein interactions or protein structure folding.

Low input capture Hi-C: a method to decipher the molecular mechanisms underlying non-coding alterations

Laureano Tomás-Daza[1], *Llorenç Rovirosa*[1], *Alfonso Valencia*[2], *Biola M. Javierre*[1]

[1]*Josep Carreras Leukaemia Research Institute,* [2]*Barcelona Supercomputing Center*

BACKGROUND

Long-range interactions between regulatory elements and promoters are key in gene transcriptional control; however, their study requires large amounts of starting material, which is not compatible with clinical scenarios nor the study of rare cell populations.

DESCRIPTION

We have developed low input capture Hi-C (liCHi-C) as a cost-effective, flexible method to map and robustly compare promoter interactomes at high resolution. As proof of its broad applicability, we implement liCHi-C to study normal and malignant human hematopoietic hierarchy in clinical samples. We demonstrate that the dynamic promoter architecture identifies developmental trajectories and orchestrates transcriptional transitions during cell-state commitment. Moreover, liCHi-C enables the identification of disease-relevant cell types, genes and pathways potentially deregulated by non-coding alterations at distal regulatory elements. Finally, we show that liCHi-C can be harnessed to uncover genome-wide structural variants, resolve their breakpoints and infer their pathogenic effects.

CONCLUSION

Collectively, our optimized liCHi-C method expands the study of 3D chromatin organization to unique, low-abundance cell populations, and offers an opportunity to uncover factors and regulatory networks involved in disease pathogenesis.

P53 ORCHESTRATES SPATIO-TEMPORAL EPIGENOME REWIRING TO TRANSCRIPTIONALLY PREVENT MALIGNANT TRANSFORMATION

Monica Cabrera-Pasadas, *Barcelona Supercomputing Center, Barcelona and Josep Carreras Leukaemia Research Institute, Badalona*

BACKGROUND

Cell stress and DNA damage activate p53 transcription factor, triggering transcriptional activation of a myriad of target genes to ultimately facilitate distinct cellular outcomes, including cell cycle arrest, senescence, or apoptosis among others. However, the molecular mechanisms underlying p53-related gene transcription regulation are not completely understood. This gap of knowledge is critical since p53 is one of the most frequently mutated gene in blood malignancies leading the loss of one of the first barriers to prevent malignant transformation.

DESCRIPTION

In this project we addressed this gap of knowledge. p53 preferentially bind enhancers. Since enhancers control transcription of target genes through physical proximity with their promoters, proximity determined by the 3D genome folding within the nucleus, we aimed to study the role of spatio-temporal genome architecture in the p53 response. Specifically, we deciphered how p53 activation orchestrate the 3D epigenetic landscape to ultimately control gene transcription and block cancer development.

To study the dynamic crosstalk between spatiotemporal genome architecture, epigenetics and transcription triggered by p53 we have modelled p53 activation on time and performed a multiomics integration of Hi-C, Promoter Capture Hi-C (PCHi-C), ChIP-seq and RNA-seq data. Just to note, PCHi-C is the method that we previously developed to associated distal enhancers and target genes (Javierre et al., Cell, 2016).

We demonstrated that p53 drives dramatic changes in genome architecture, including A/B compartments, Topological Associated Domains (TADs) and DNA loops after minutes of its activation. These changes accompanied epigenetic landscape re-configuration to ultimately trigger p53-related transcriptional response. Then, we defined set of functional p53 binding sites, being most of these at enhancers as previously reported. p53 activation drove new-loop formations between p53-bound enhancers and gene promoters. However, in some cases the 3D chromatin topology was pre-established. In both cases, DNA loops allowed the propagation of the activating p53 effect from distal enhancers to promoter to ultimately lead gene transcription upregulation. Specifically, we associated these functional p53 binding sites at enhancers with a set of 331 distal target genes, which in most of the cases were not the closest gene in the linear genome (mean distance between p53-bound enhancers and target gene porters of 153Kb). Among these, we did not only identified examples of previously identified p53 target genes (e.g.TP53INP1, PLK2) but we also identified potentially new direct target genes and pathways distally controlled by p53.

CONCLUSION

Collectively, our results demonstrate that p53 activation dramatically reshapes the promoter-enhancer interactome landscape to ultimately control the transcriptional response. Besides, this study provides the first set of genes and pathways distally controlled by p53 and suggest candidate non-coding regions (i.e., p53-bound functional enhancers and linked promoters) that can be mutated or epimutated in blood malignancies to lead an aberrant p53 response in a wild p53 condition.

FLASH TALK POSTERS (TITLE: A-Z)

Changes on the Structure of Microbial Communities of the rhizosphere of peruvian fruit trees (Annona cherimola Mill. and Pouteria lucuma) across depth soil using PacBio HiFi sequencing

Richard Estrada[1] *and* ***Tatiana Porras***[1]

[1]*Instituto Nacional de Innovación Agraria (INIA)*

The characterization of soil microbiological structure at different depths is essential to understand the impact of microorganisms on nutrient availability, soil fertility, plant growth and stress tolerance, as well as to identify bacteria with bioremediation capacity. In this study, the microbiological structure was analyzed at three depths (3 cm, 12 cm and 30 cm) of two fruit trees native to the inter-Andean valleys of South America: Annona cherimola Mill. (Chirimoya) and Pouteria lucuma (Lucuma). These fruit trees not only have nutritional benefits, but also offer a combination of nutrients essential for a balanced diet, as lucuma is a source of vitamins and minerals, while cherimoya provides vitamin C and antioxidants. We used a high-throughput Pacbio HiFi long-read sequencing approach to explore the composition, diversity, and functions of bacterial communities of the rhizosphere soil of Annona cherimola Mill. and Pouteria lucuma native in different soil depth. Significant differences were observed in the alpha diversity indices, evaluated by Shannon's index (p=0.0114) and observed features (p=0.0105) between the soil depths. The family-level relative abundance analysis of Pouteria lucuma revealed that Acidobacteriaceae and Thermoanaerobaculaceae were predominant in the shallower soils, whereas the genera Thiobacter and Pirellula exhibited a significantly higher abundance in the soils at a depth of 30 cm. We found significant changes in beta diversity due to depth gradient and plant type. Also, we also found similar functional diversity profiles among microbial communities and the influences of edafic factor in the strcuture of microorganisms. This study will help for future research aimed at understanding the impact of microorganisms in different deep layers of the soil, as well as their influence on crop growth and quality.

Comprehensive annotation of miRNAs and lncRNAs in domesticated cotton species

Vivek At[1]*, Ajeet Singh*[1]*, Shailesh Kumar*[1]

National Institute of Plant Genome Research

Allotetraploid cotton plants, specifically Gossypium hirsutum and Gossypium barbadense, are widely grown for their natural and renewable textile fibers. Despite extensive research on non-coding RNAs in domesticated cotton species, systematic identification and annotation of long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) expressed in different tissues, developmental stages, and biological contexts remains limited. This limits our understanding of their functions and impedes future cotton research. To fill this void, we present a high-confidence set of lncRNAs and miRNAs from G. hirsutum and G. barbadense derived from large-scale RNA-seq and small RNA-seq datasets. This information is incorporated into CoNCRAtlas, a user-friendly database that provides comprehensive annotations of lncRNAs and miRNAs based on the systematic integration of extensive annotations. We anticipate that this comprehensive resource will accelerate evolutionary and functional studies of non-coding RNAs, providing critical insights for future cotton breeding programs. The CoNCRAtlas database is free and open to the public, and it can be accessed at http://www.nipgr.ac.in/CoNCRAtlas/.

Gene subset signatures for complexity reduction of technical and functional comparisons of whole transcriptomes

Shruti Gupta[1], Shandar Ahmad[1]

[1]*School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India*

BACKGROUND

Gene expression profiling has widely investigated cellular and disease contexts. While large-scale transcriptome data are often searched using phenotype metadata of each experiment, there is also an interest in the reverse inference of phenotype from a newly generated transcriptome profile called "content". Studies have been done on drug repositioning, identifying novel drug-disease connections or inferring drug-drug relationships from expression profiles, highlighting the significance of profile comparisons and content-based queries. In 2017, Subramanium et al. showed that much fewer landmark genes (L1000 genes) carry enough information about whole transcriptome profiles, and the latter can be predicted using a deep learning model. However, it was unclear if the selected L1000 set is the only combination that can do the job and if they have the same power to segregate biologically relevant expression profiles.

DESCRIPTION

In this work, we develop a database of global expression profiles of legacy microarray and RNASeq experiments, including those from single cells. Using these data sets, we assess the performance of L1000 and other similar-sized genesets by creating random and systematically selected samples. We evaluate the ability of these subsets to reproduce profile-profile comparisons that would result from the whole-transcriptome gene expression profile. We also investigated the pathways, gene ontologies and functional features which are critical for the selection of a good subset having the ability to reproduce original comparisons. Finally, profiles derived from selected subsets were also applied to extract biologically similar samples instead of simple expression profile signatures of each sample in the dataset. A framework for large-scale comparison of such data sets is also being developed.

CONCLUSION

Our results suggest that many different gene subsets are equally powerful to L1000 in reproducing profile-profile similarities between transcriptome data sets. Certain genes often form the part of best-performing gene subsets, highlighting their criticality for determining whole transcriptome patterns. The choice of these genes and subsets will aid in understanding the key regulatory factors in gene expression. These findings will also help in the large-scale imputation of gene expression profiles collected on smaller platforms and speed up the process of content comparison between them.

Impact of oral anti-diabetic drugs on gut-derived extracellular vesicles: Proteomic Signature

Estefania Torrejon[1], *Akiko Teshima*[1], *Ana Sofia Carvalho*[1], *Hans Christian Beck*[2], *Rune Matthiesen*[1], *Paula Macedo*[3], *Rita Machado de Oliveira*[1]

[1]*Nova Medical School,* [2]*Centre for Clinical Proteomics, Department of Clinical Biochemistry and Pharmacology, Odense University Hospital,* [3]*APDP*

BACKGROUND

Extracellular vesicles (EVs) mediate inter-organ communication in type 2 diabetes (T2D) pathogenesis. Gut derived EVs (GDE) protein content reflects metabolic state and administering prediabetic GDE induces a diabetogenic phenotype on healthy mice. Analysis of GDE proteomic profile showed an upregulation of acyl-CoA thioesterases and downregulation of rate-limiting enzymes for glycolysis. To unveil the relevance of oral antidiabetic drugs, we hypothesize that metformin and pioglitazone's metabolic actions are dependent on GDE's proteomic cargo.

DESCRIPTION

Two groups of mice were fed with either normal chow diet (NCD) or high-fat diet (HFD), then treated with metformin or pioglitazone. GDE were isolated and characterized by nanoparticle tracking analysis. Proteins were extracted and analyzed by nano-LC-MSMS. Statistical analysis was performed using the limma R package. After treatment, both drugs improved glucose intolerance and liver steatosis compared to prediabetic animals. We identified 159 proteins differentially expressed between HFD and HFD+metformin and 180 between HFD and HFD+pioglitazone and, together with the principal component analysis among groups, these results indicate both drugs alter the GDEs protein composition to resemble NCD.

CONCLUSION

Metformin and pioglitazone modify GDE-mediated interorgan crosstalk, which plays a role in the progression of dysmetabolism. Modulating this mechanism may have therapeutic implications.

mICKEY: Memory-Efficient Deep Learning for Personalized Biomarker Discovery and Cancer Origin Prediction from DNA Methylation Data.

*Pakanan Tussanapirom[1], **Kasidech Aewsrisakul**[1], Chanati Jantrachotechatchawan[2], Kobchai Duangrattanalert[3]*
[1]Triam Udom Suksa School, [2]Faculty of Medicine Siriraj Hospital Mahidol University, [3]Chulalongkorn University Technology Center (UTC)

Cancer claims over 10 million lives annually, and treatment depends on accurate identification of tissue origin. Traditional diagnostics have long wait times and invasive biopsies pose significant health risks. To tackle these challenges and pave the way for early detection, we developed mICKEY, a deep-learning pipeline for cancer prediction using CpG methylation data from solid and liquid biopsies. We use deep learning, instead of standard statistics, for its capacity to model complex non-linear biological data. Our model employs variational inference to boost consistency and masks to handle low-quality data. We use dense layers with regularization for CpG site selection. Fewer CpG sites allow healthcare facilities to measure and analyze with affordable multiplex techniques instead of high-throughput screening (>450k probes) that requires memory-consuming models. Lastly, mICKEY has a self-attention layer to encode interaction and pinpoint vital CpG sites in each sample. Our model, using fewer than 100 CpG sites, achieves robust sensitivity and specificity of over 95% on 18 cancer origins across demographic groups and sample types on solid biopsy DNA from holdout TCGA and independent GEO datasets. To make our model more suitable for early detection via non-invasive methods, mICKEY couples metric learning with domain adaptation to deal with data limitation and yields over 85% in sensitivity and specificity. Sample-specific attention maps reveal known biomarkers, validating its potential for future personalized treatment. Overall, mICKEY offers a practical solution for early detection of cancer together with a promising future use in personalized therapy.

Most frequently harboured missense variants of hACE2 across different populations exhibit varying patterns of binding interaction with spike glycoproteins of emerging SARS-CoV-2 of different lineages

Rubaiat Ahmed[1], *Anika Tahsin*[1], *Piyash Bhattacharjee*[1], *Maisha Adiba*[1], *Abdullah Al Saba*[1], *Tahirah Yasmin*[1], *Sajib Chakraborty*[1], *A. K. M, Mahbub Hasan*[1], *A. H. M. Nurun Nabi*[1]

[1]*University of Dhaka*

BACKGROUND

Since the emergence of SARS-CoV-2 in 2019, the virus accumulated various mutations, resulting in numerous variants. According to the mutations acquired, the variants are classified into lineages and differ greatly in infectivity and transmissibility. The world saw prominent surges in the rate of infection as newer variants emerged. However, not all populations suffered equally, which suggests a possible role of host genetic factors.

DESCRIPTION

We investigated the effect of the lineage-defining mutations of the SARS-CoV-2 variants: Mu, Delta, Delta Plus (AY.1), Omicron sub-variants BA.1, BA.2, BA.4, BA.5, and BA.2.12.1 on the strength of binding of the spike glycoprotein receptor-binding domain (RBD) with the human angiotensin-converting enzyme 2 (hACE2) missense variants prevalent in major populations (E37K in Africans, F40L in Latin Americans, D355N in non-Finnish Europeans, and P84T in South Asians) via molecular docking and molecular dynamics (MD) simulation. The results demonstrated variable strength of binding and showed altered interaction patterns in different hACE2-RBD complexes.

CONCLUSION

The missense variants of hACE2 and spike RBD mutation, both affect the binding energy and pattern of interaction between the two proteins. In vitro studies are warranted to confirm these findings which may enable early prediction regarding the risk of transmissibility of newly emerging variants across different populations in the future.

Virus-host interactions in a municipal landfill include non-specific viruses, hyper-targeting, and interviral conflicts

Nikhil George[1], *Laura Hug*[1]

[1]*University of Waterloo*

BACKGROUND

Viruses are the most abundant microbial guild on the planet and impact microbial community structure and ecosystem services. Viruses are underrepresented in databases and specifically understudied in engineered environments. In this study we aimed to expand the diversity and ecological roles of prokaryotic viruses in engineered environments.

DESCRIPTION

Using metagenomics, we examined host-associated viruses from multiple sites in a municipal landfill across two years via CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) spacer to viral protospacer mapping. Approximately 46,000 viral elements were identified using VirSorter. 1,858 Metagenome-Assembled Genomes were identified using three automated algorithms (CONCOCT, MetaBAT2, and Maxbin2), DAS Tool for consensus-based

binning, and CheckM for quality filtering. Viruses comprised ~5% of the assembled community DNA sequence. These viruses are diverse but show strong genomic similarity across space and time. A total of 506 unique virus-host connections capture hyper-targeted viral populations and host CRISPR array adaptation over time, as well as viruses predicted to infect across multiple phyla, suggesting that some viruses may be far less host-specific than is currently understood. Hyper-targeted viral elements showed a relative depletion in Chi sites which we predict influences host CRISPR-Cas systems' disproportionate recruitment of spacers from these genetic elements. We detected 190 viral elements that encode CRISPR arrays, including the longest virally-encoded CRISPR array described to date, with 187 spacers. CRISPR-encoding proviruses integrated into host chromosomes were observed, as examples of CRISPR-immunity-based superinfection exclusion.

CONCLUSION

Our networks highlight virus-host interactions that are rare or have not previously been described, but which influence the ecology of this dynamic engineered system. We intend to use these networks as a framework to understand how virus-induced host mortality may be influencing landfill processes such as nutrient cycling and contaminant degradation. The viruses detected in this study are not well-represented in reference databases nor are their host interactions frequently described in literature. This altogether suggests that municipal waste sites harbor distinct viromes and virus-host interactions that merit further study.

POSTERS (TITLE: A-Z)

Altered Temporal and Visual Resting State Network Biomarkers in Early Parkinson's Disease

Neel Kondapalli, *Dougherty Valley High School*

Parkinson's disease is a brain disorder that affects upwards of ten million people worldwide. It impairs motor and non motor skills, leading to symptoms such as tremor, stiffness, and cognitive dysfunction. Parkinson's disease drastically decreases quality of life and gradually reduces the ability of patients to function on their own. Most current diagnosis methods rely on observable physical characteristics of the disease in a clinical setting. The implications of this is that 1) Parkinson's disease detection and subsequent treatment administration are not as early as they could be, and 2) progress towards better medications for Parkinson's disease is slow since researchers are still trying to understand the disease's pathology. Neuroimaging biomarkers can help both with detecting the disease earlier through non-invasive means and deepening scientific comprehension on how it affects the brain. Here, we compared neuroimaging data from patients with Parkinson's and controls via graph theory and persistent homology analysis to see where diseased brain networks diverge from normal functioning at an early stage. Abnormalities in various metrics were found primarily in regions of the temporal and occipital lobes responsible for important cognitive and visual processing tasks that are associated with the early cognitive decline seen in Parkinson's disease. Regions such as the fusiform gyrus and the lateral occipital cortex were observed to exhibit abnormal network influence in patient brains. Divergence from normal functioning at an early stage in these temporo-occipital areas may be an indication of the development of certain Parkinson's-related symptoms.

Combining CNNs and RNNs for Automated Protein Function Prediction

Wasana Parackrama[1], *Dulmini Ekanayake*[1], *Kalani Prabodha*[1], *Damayanthi Herath, Indika Kahanda*[2]

[1]*Department of Computer Engineering, Faculty of Engineering, University of Peradeniya,* [2]*School of Computing, University of North Florida*

BACKGROUND

Gene Ontology (GO) is a widely used vocabulary for describing protein functions, covering cellular components, molecular functions, and biological processes. Despite how beneficial it is to know the GO annotations of the proteins, determining protein function through wet-lab experiments is laborious and time-consuming. Therefore, in order to keep up with the rate at which new proteins are sequenced, numerous computational methods for Automated Function Prediction (AFP) have emerged over the years; from the earliest statistical approaches to models which use the latest Deep Learning techniques.The primary aim of this project is to develop a more accurate AFP model that uses protein-protein interaction (PPI) data and protein sequence data by combining Convolutional Neural Network (CNNs) and Recurrent Neural Network (RNNs).

DESCRIPTION

We first experimented with multiple sub-models (sequence-CNN component, sequence-RNN component, PPI-RNN component) and tested the performance of CNNs vs RNNs with sequence data and PPI data separately, and the overall effect of the techniques used to combine the sub-models. We obtained sequence data for yeast from CAFA3 (The Critical Assessment of Functional Annotation) training data and PPI data from the STRING database. Subsequently, based on the preliminary observations from above experiments, we developed a novel RNN-CNN model for AFP that use protein sequences and PPI data.

CONCLUSION

Our experimental results suggest that there is a significant performance (Fmax) improvement in protein function prediction when CNN and RNN components are combined. In the future, we plan to further improve performance by training with a higher number of protein sequences while following additional data preprocessing steps. Another aspect worth exploring is the incorporation of other types of genomic data such as mutations as input. Testing on larger and more complex genomes such as human would be another future direction. Our findings from this study have implications for both the biologists and bioinformaticians working in AFP.

Common hallmarks between melanoma brain metastases and Alzheimer's disease

Irene Soler-Sáez[1], *Alcida Karz*[2], *Adolfo López-Cerdán*[3], *José F. Català-Senent*[3], *Gonzalo Antón-Bernat*[3], *María de la Iglesia-Vayá*[4], *Marta R. Hidalgo*[3], *Eva Hernando-Monge*[2], *Francisco García-García*[3]

[1]*Centro de Investigación Príncipe Felipe (CIPF),* [2]*Department of Pathology, New York University Grossman School of Medicine, New York (US),* [3]*Bioinformatics and Biostatistics Unit. Principe Felipe Research Center, Valencia (Spain),* [4]*Joint Unit CIPF-FISABIO Biomedical Imaging*

BACKGROUND

Melanoma is a skin cancer subtype caused by the uncontrolled growth of melanocytes. It has the highest mortality rate of skin cancers, mainly due to its high metastatic capacity. Among the most frequent and aggressive metastases, melanoma brain metastasis stands out since it is present in half of the patients and 75% of postmortem tissues with metastatic melanoma. Its biological insights are currently unknown, but recent studies suggest that melanoma brain metastasis mimics patterns of Alzheimer's disease, such as amyloid beta secretion. We performed this work to unveil the common molecular mechanisms of both diseases, which is key for understanding how melanoma colonises the brain, and which specific transcriptomic profile this metastasis possesses to survive in the brain microenvironment.

DESCRIPTION

To fully characterise the similarity between these diseases, we have compared the transcriptomic profiles of two melanoma brain metastasis studies with one Alzheimer's disease study from cortex and one Alzheimer's disease study from hippocampus. For each pair of melanoma brain metastasis versus Alzheimer's disease data, we identified common deregulated genes, and checked whether these genes do or do not intersect randomly by resampling. These genes were functionally characterised by the following five areas: biological processes, molecular functions, cellular components, signalling pathways and protein-protein interactions. As a result, in all comparisons we find common deregulated genes which provide non-random biological signals. In detail, 34 genes are altered commonly in melanoma brain

metastasis and the cortical region of Alzheimer's patients, being involved in glycosaminoglycan binding and the collagen-containing extracellular matrix. Moreover, 18 genes are commonly altered in melanoma brain metastasis and the hippocampus of Alzheimer's patients, being in this case related with axonogenesis. The consensus pattern of both regions exposes 8 genes as key deregulated genes: BEX1, PMP2, ST6GALNAC3, UCHL1, NFIB, PALLD, PBX3 and VCAN.

CONCLUSION

These results provide the basis for discovering novel potential melanoma brain metastasis biomarkers. In detail, this work has succeeded in obtaining a panel of genes that could lay the foundations for the definition of the molecular mechanisms by which melanoma metastasizes to the brain tissue so effectively.

Comparative Analysis of Computational Methods for Single-Cell RNA Sequencing Annotation in Mouse Motor Cortex

Olutoba Ojo, *UPMC Hillman Cancer Center*

BACKGROUND

Single-cell RNA sequencing (scRNAseq) is a newer method that enables scientists to quantify the RNAs in individual cells and infer more accurate identities for the cells in a sample. Computational methods are essential to infer cell types as accurately as possible.

METHODOLOGY

In this study, we annotated mouse motor cortex cell data using three different computational annotation tools: Azimuth, SingleR, and ScType. The scRNAseq mouse motor cortex data was retrieved from the PanglaoDB database and normalized prior to analysis. Azimuth used an annotated mouse motor cortex reference as a basis for mapping and produced two plots visualizing cell type prediction. SingleR used reference transcriptomic datasets of pure cells for its annotation and output a heatmap graph distinguishing the varying levels of expression among cell types, as well as a violin plot visualizing the distribution of cell type expression. ScType utilized a combination of its cell marker database and output a cluster graph of cell types and a mapping classification graph. Our study demonstrates that the use of various computational annotation methods can lead to differing interpretations of data. By utilizing multiple computational methods for cell type annotation, we gain a more comprehensive understanding of the cell types present in the mouse motor cortex cell data. The results show that the three methods produced different annotations for the mouse motor cortex cell data. SingleR and ScType require the normalization of data as a step prior to analysis. Azimuth, on the other hand, uses an annotated reference file as a basis for mapping. The utilization of multiple computational methods in cell type annotation can provide more reliable and accurate identification of cell types in scRNAseq datasets.

CONCLUSION

In conclusion, our study highlights the importance of utilizing multiple computational methods in single-cell RNA sequencing annotation. This approach can provide a more comprehensive understanding of cell types present in complex datasets, such as the mouse motor cortex cell data used in this study. These findings have indications for future studies aiming to understand the diversity of cell types in various biological systems.

Comparison of existing prediction tools and development of deep learning-based machine learning predictor for S-palmitoylation site in Arabidopsis Thaliana

Ayesha Sanjana Kawser Parsha, *University of Manchester*

S-palmitoylation is an irreversible bond where palmitate (74%) or stearate (22%) fatty acid bonds with cysteine residue either at COOH or NH2 terminal through a thioester linkage. There have been various experimental techniques over the years that can help to identify the S-palmitoylation site but since it is time-consuming the need for computational techniques has been increasing. However, there are few predictors that can identify S- palmitoylation sites in Arabidopsis Thaliana accurately enough.

In this research, prediction tools such as the GPS PALM 6.0, pCysMod, GPS LIPID 1.0, CSS PALM 4.0, and NBA PALM were compared to determine the type of machine learning algorithm that predicts this site more accurately for the experimental dataset. Based on the analysis, a AI-driven deep learning-based prediction tool was developed with three sequence-based input features such as the binary encoding profile, amino acid composition, and autocorrelation features. Different combinations of features were used to build the model and after the training and validation, the model performed better with the experimental dataset for 8 and 10- fold cross-validations when all features were present. The model was tested with unseen data and the results exhibited a higher area under curve score which is near to the score of 8 and 10-fold cross-validation with the experimental dataset. After comparing the area under curve score of 10-fold cross-validation of this new model with the established tools' area under curve score with their respective training set, it can be shown that this model is better than the previous tools in predicting the S- palmitoylation site in the experimental dataset.

The aim of this research is to build a better prediction tool for the Arabidopsis Thaliana and according to the area under curve score it has a higher prediction accuracy score than the existing tools. By using this tool to predict S- palmitoylation sites, food production can be controlled in plants and also it can be used to control immunological therapeutic targets.

Computational Analysis of Carbapenemics drug affinity against Acinetobacter baumannii CarO protein mutations

Lin Geraldin Zevallos Cuarite[1], Haruna Luz Barazorda Ccahuana[1]

[1]Universidad Católica de Santa María

Antibiotic resistance is a major threat to global health and could become one of the leading causes of death in the coming decades. Different studies show that the mechanisms of resistance to antibiotics allow the modification of targets. Acinetobacter baumannii is a multidrug-resistant pathogen; found in intensive care units where it mainly causes pneumonia, bacteremia, among other diseases. The CarO protein is located in the barrel-shaped outer membrane and is associated with carbapenem antibiotics. Genetic modifications, conformational changes, and insertion sequences affect its function. The main objective of this research is to analyze by computational approach the affinity of carbapenem drugs such as Imipenem, Meropenem, Doripenem, and Ertapenem against CarO mutations of Acinetobacter baumannii. We searched for the CarO protein sequences from five biological sources reported in the NCBI database and made multiple alignment of the sequences against the native protein using the Clustal Omega algorithm. For the 3D structure of proteins, homology modeling was performed on the SWISS-MODEL server. The PubChem server was used to search drugs, where the 3D conformation was analyzed with the Avogadro viewer. Finally, for molecular docking, each mutated CarO was docked against the drugs using the

Dockthor server. The results of the multiple alignment show that the sequences have a similarity percentage greater than 95.9% and the identity varies from 55.71% to 100% with the native protein. The molecular modeling of CarO showed the differences between the mutated sequences. This residue variation occurred in the channel gate of this protein. In the molecular docking of CarO with the four drugs, we observed that there are different affinity values, ranging from -6.928 of the Imipenem compound with the sputum sample (ARK08725) to -8.26 of the Ertapenem drug with the feces sample (APU28149). In conclusion, our results show that the sequence mutations could cause drug resistance since different variations of the affinity values between the protein and the ligands are observed. To continue our studies it is necessary a Molecular Dynamics analysis to validate the molecular interactions. Likewise, this study can be applied to molecular microbiological analysis and targeted therapy.

Detailed analysis of the aged brain transcriptome reveals differences in mRNA properties and impaired mRNA turnover

Nisha Hemandhar Kumar[1], *Eugenio F. Fornasiero*[1,2], *Verena Kluever*[1], *Mattia Pelizzola*[3], *Mattia Furlan*[3], *Emanuel Barth*[4], *Manja Marz*[4], *Sebastian Krautwurst*[4], *Cristina Cheroni*[5,6], *Giuseppe Testa*[5,6,7]

[1]*University Medical Center Goettingen,* [2]*University of Trieste,* [3]*Fondazione Istituto Italiano di Tecnologia,* [4]*Friedrich Schiller University Jena,* [5]*European Institute of Oncology,* [6]*University of Milan,* [7]*Human Technopole, Milan, Italy*

Brain aging is characterized by a progressive loss of tissue integrity and increased cellular heterogeneity, leading to impaired function, increased susceptibility to disease and death. Despite recent rapid technological advances in various omics measurements, the complexity and the coordination of the molecular processes associated with brain aging remain poorly understood, especially when combining and reconciling mRNA and protein datasets. In this work, we collected and analyzed two large datasets including mRNA levels from second and third generation sequencing technologies such as Illumina and Oxford Nanopore technologies in young adult and aged adult mice. We report the first transcriptome-wide differential transcript usage study of brain aging. In addition, we provide the community with a large resource of whole brain transcriptomes and comprehensive analyses that identify widespread diversity in RNAs during aging. Specifically, we observed that the mRNAs encoding for neuronal synaptic proteins are upregulated with age and this is conserved in the human context, although the difference is not observed at the protein level. We also observed that a subset of the genes that are upregulated in the aged brain is associated with neurodegenerative diseases. In addition, we report that the RNA molecules that are longer and have a shorter 3'UTR are abundant in the aged brain and a subset of these are alternatively spliced at the 3'UTR. Finally, we observed a difference in the turnover of mRNAs at different ages. Overall, based on these observations, we speculate that alterations at the 3'UTR may play an active role in the aging process, prompting future efforts to explain these observations from a molecular perspective and providing new ideas to understand the cellular and molecular mechanisms occurring in the aged mammalian brain.

Gene and functional signatures across time and severity in spinal cord injury. A meta-analysis of transcriptomic studies in rat

Rubén Grillo-Risco[1], *Marta Hidalgo-García*[1], *Beatriz Martínez-Rojas*[2], *Victoria Moreno-Manzano*[2], *Francisco García-García*[1]

[1]*Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF),* [2]*Neuronal and Tissue Regeneration Laboratory, Principe Felipe Research Center (CIPF)*

BACKGROUND

Spinal cord injury is a devastating condition characterized by the disruption of neural tissue within the spinal cord, resulting in partial or complete loss of motor, sensory, and autonomic functions. However, there are no effective therapeutic solutions for spinal cord injury nowadays, so it is essential to use multidisciplinary approaches that allow a better characterization of the molecular mechanisms underlying this condition.

DESCRIPTION

The objective of this study is the characterization of different phases and severities of spinal cord injury at the gene and functional level, as well as the identification of biomarkers. The employed strategy consists on the following steps: 1) systematic review and selection of transcriptomic studies; 2) individual analysis of the studies, including preprocessing, exploratory analysis, and differential expression analysis; 3) gene meta-analysis; 4) identification of biomarkers; 5) functional characterization including gene set analysis, inference of transcription factor activities, and construction of co-expression networks; 6) bioinformatic and experimental results validation. As a result, we have compiled a total of 14 studies and 273 samples, organized into 9 experimental groups. We successfully identified severity-specific gene markers, such as Srpx2, Hoxb8, Acap1, Snai1, or Aadat, useful for the prognosis prediction of the injury, as well as phase-specific genes like Il6, Fosl1, Cfp, C1qc, or Cp, enabling accurate identification and classification of spinal cord injury transcriptomic profiles.

CONCLUSION

We constructed a transcriptomic reference that provides a valuable framework for assessing spinal cord injury at the gene and functional levels, incorporating the dual perspective of severity and time. In addition, we have created an interactive web application for researchers interested in the field of spinal cord injury that offers a user-friendly interface to explore and visualize the results obtained in this work (https://rubbmetasci.shinyapps.io/metaSCI/). Our work has allowed us to assess how expression profiles change as a function of severity and time after injury, as well as to identify the biological functions, pathways and transcription factors that are deregulated.

High-performance annotation of biological sequence profiles

Martin Larralde[1], *Georg Zeller*[1]

[1]*European Molecular Biology Laboratory*

BACKGROUND

Sequence analysis has been a prevalent field of bioinformatics growing with each new advance in DNA sequencing. Methods for sequence-profile comparison which were developed in the 20st century are still in use today, processing an ever-expanding amount of sequencing data.

Position Specific Scoring Matrices (PSSMs) were first introduced to represent sequence motifs. PSSMs take into account the occurrence of individuals letters in a motif against background probabilities. The MEME suite is commonly used for discovery and search of motifs. Expanding on PSSMs, profile Hidden Markov Models (pHMMs) were proposed to better capture protein families by additionally modelling insertions and deletions. HMMER is a de-facto standard for pHMMs, powering workflows such as the NCBI Prokaryotic Genome Annotation Pipeline.

DESCRIPTION

To support our sequence learning model, which uses pHMM features, we developed PyHMMER, a Python library binding to HMMER. PyHMMER offers more flexibility than the HMMER command line, as it can process inputs directly from memory. The reworked threading model results in substantial time gains for smaller target databases. On a six-core machine, PyHMMER can annotate a bacterial proteome in roughly a quarter of the time needed by HMMER.

While HMMER offers high performance, thanks to a platform-accelerated compute kernel, there is no high-speed equivalent for PSSMs. Existing implementations motif scanning all rely on sequential scoring. Inspired by the matrix striping technique of HMMER, we developed Lightmotif, a Rust library implementing striped PSSM scanning. This technique allows for the vectorized computation of motif scores using modern CPU extensions. The AVX2 version can search a human chromosome for a binding site in a mere second, a 30x to 50x speed-up over sequential implementations. Lightmotif also provides a Python interface similar to Biopython to allow for drop-in replacement.

CONCLUSION

We developed two libraries for efficiently scanning sequence features using HMMs and PSSMs respectively. They leverage modern processor technologies like SIMD and multithreading to accelerate ubiquitous sequence analysis tasks. The efficiency improvements help scaling these methods to larger datasets, but also lead to a reduction in resource usage, which translates to reduced energy consumption and contributes to green computing.

MaGplotR: a software for the analysis and visualization of multiple MaGeCK screen datasets through aggregation

Alejandro Matía[1], *María M. Lorenzo*[1], *Duo Peng*[2]

[1]*Department of Biotechnology, INIA-CSIC,* [2]*Chan Zuckerberg Biohub*

In the recent years, genetic screens are gaining crucial importance for interrogating gene function in a high-throughput manner. Bioinformatic tools for analyzing and visualizing the results from these screens are being developed to help with these tasks. We present MaGplotR, a tool that analyzes multiple CRISPR screen datasets, identifies common hits, couples the hits to enrichment analysis (several GO categories) and network plots, and produces publication-level plots. Output plots give information on the quality control of the screen data (e.g., sgRNA distribution and PCA for quality control) and show the best hits by aggregation from multiple screen experiments. To maximize comparability, rank is used to identify common hits. MaGplotR can also be used to analyze experiments where a control condition is used for multiple treatment conditions. MaGplotR is executed from any command line and is fast and easy to use, with even just one argument. It is available at https://github.com/alematia/MaGplotR.

Melanoma biomarkers identification using novel computational strategies

Helena Gómez-Martínez[1], *Adolfo López-Cerdán*[1], *Gonzalo Antón-Bernat*[1], *Eva Hernando*[2], *Francisco García-García*[1]

[1]*Bioinformatics and Biostatistics Unit. Principe Felipe Research Center, Valencia (Spain),* [2]*Department of Pathology, New York University Grossman School of Medicine, New York (US)*

BACKGROUND

Melanoma is one of the most aggressive types of skin cancer, with a progressive increase in its incidence in recent years. The main factors influencing the high lethality of the disease are the late onset of symptoms, the lack of an effective treatment, and the difficulties in early disease detection. However, although clinical and epidemiological parameters show a differential pattern by sex in the diagnosis and progression of this disease, there is a lack of knowledge regarding efficient biomarkers that might improve the early detection and the progression of the disease. Consequently, the main objective of this work is the development and application of a computational strategy to improve the knowledge of the development of the disease and prediction of the diagnosis of the patients as well as their evolution.

DESCRIPTION

In this work, a systematic review of melanoma studies according to PRISMA statements, in different databases, was conducted. Only those that were microarrays or massive sequencing of miRNAs and met previous criteria, were selected. After data download, an exploratory analysis and data normalization were performed. Subsequently, an individual analysis of each study and a global meta-analysis was carried out by integrating all results obtained and considering the critical patient information. Results revealed significant biomarkers in a diagnosis and progression scenario.

CONCLUSION

The findings generated in this study have contributed to a better understanding of the molecular mechanisms and the developmental processes of melanoma, giving the possibility of developing novel resources for its detection and treatment in a precision medicine frame.

Multi-omic signature for the diagnosis in melanoma brain metastases

Gonzalo Antón-Bernat[1], *Helena Gómez-Martínez*[1], *Eva Hernando-Monge*[2], *Francisco García-García*[1]

[1]*Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), Valencia (Spain),* [2]*Department of Pathology, New York University Grossman School of Medicine, New York (US)*

BACKGROUND

Melanoma is the third most common source of brain metastases, surpassed only by lung and breast cancer, and is the solid tumor with the greatest propensity to target the brain. The incidence of clinically detected brain metastases in patients with advanced melanoma is approximately 40%. These patients have the highest morbidity and mortality of patients with melanoma. Although in recent years some clinical trials are including patients with brain metastases, most only accept patients with a single asymptomatic brain metastasis, which represents a small proportion of patients with metastatic brain melanoma (MMC). In summary, there is a great unmet medical need as these patients continue to have few clinical options. Furthermore, knowledge of the molecular and cellular mechanisms responsible for melanoma brain metastasis remains limited. The aim of this purpose is to develop a multi-omics signature to improve the diagnosis of these patients.

DESCRIPTION

The methodology is based on an in silico approach capable of interrogating and integrating all the information related to MMC, available in public omics data repositories. This strategy included: a) a systematic review that identified omics studies (genomics, transcriptomics and epigenomics) on this disease, b) a bioinformatics re-analysis of each study, and c) its integration to identify genes and functions associated to MMC. The initial systematic review results provided 413 studies of MMC, in Homo sapiens (340 transcriptomic, 34 genomic and 39 methylation studies). After their exhaustive review, 15% of these studies met the final inclusion requirements. In this first phase, we already have the results of the bioinformatics analysis for the transcriptomic datasets. This set of candidate genes constitutes a potential target that will be expanded when the next phases of the proposal (genomics and epigenomics) are completed.

CONCLUSION

The identification of diagnostic and progression biomarkers will generate relevant knowledge that will provide guidance on which molecular mechanisms are specific to each patient group. This will improve the diagnosis, prognosis and monitoring of cancer, benefiting both the patient and the viability of the healthcare system by facilitating medical decision-making. This is a line of cancer research with enormous potential for health impact.

Novel theoretical drug design strategy to discover highly potent urease inhibitors for H. pylori based on coumarin derivatives

Gabriel J. Olguín-Orellana[1], *Daniel Bustos*[1]

[1]*Laboratory of Bioinformatics and Computational Chemistry, Faculty of Medicine, Catholic University of Maule*

INTRODUCTION

Helicobacter pylori (H.p.) infection affects over 50% of the global population, causing various gastric diseases. Urease, a key pathogenic metalloenzyme of this microorganism, hydrolyzes urea to ammonia in the stomach, favoring the replication of H.p. Antibiotic therapies have become unfeasible mainly due to antibiotic resistance. Then, new strategies based on urease inhibitors are being proposed.

DESCRIPTION

Here, a protocol is proposed to discover urease inhibitors for H.p. from new 240 coumarin derivatives. To accomplish it, the entire set of coumarins inhibitors already reported (CIAR; 142 compounds) is being subjected to Induced-Fit Docking (IFD), MM/GBSA and QM/MM to characterize their binding energy with H.p. urease (PDB-ID: 6ZJA) and correlate it with experimental values such as the IC50 and other 49 physicochemical descriptors. Subsequently, this protocol will be applied to the new set.

CONCLUSION

Until now, we estimated the binding CIAR-urease free energy through IFD and MM/GBSA, but with a low correlation with the IC50 (R2=0.59). This value could be due to the higher polarization effects induced by the Ni2+ in the urease, which decreases the precision. Consequently, the reparametrization of the ligand charges with QM/MM should increase the correlation coefficient and improve the hit-rate of the protocol.

Phylogenetic reconstruction of the Campylobacter genus

Irene Ortega Sanz[1], *Beatriz Melero, Jordi Rovira*[1]

[1]*Department of Biotechnology and Food Science, University of Burgos*

BACKGROUND

The taxonomy of the Campylobacter genus has been a subject of ongoing research and refinement for many years due to the development of new techniques and technologies leading to a better understanding. When it was formed in 1963, the genus Campylobacter comprised just two species, Campylobacter fetus and 'Campylobacter bubulus' (now Campylobacter sputorum), that were formerly classified as Vibrio spp. Since then, several new species or subspecies have been identified, especially with the potential of the 16S rRNA and hsp60 gene sequencing, while others have been subsequently refined or reclassified into a new genus, Helicobacter. At present, a total of 42 species and 13 subspecies within the Campylobacter genus have been recognized under the List of Prokaryotic names with Standing in Nomenclature (LPSN). However, no phylogenetic study of the complete set of species has been performed to date.

DESCRIPTION

In this study, the phylogenetic tree of the Campylobacter genus to date was built based on 16S rRNA gene sequences. These sequences were obtained via GenBank, and a multiple alignment was generated with MEGA X v10.2.5 using neighbour-joining clustering (p-distance) with 1,000 bootstrap replications. The phylogeny allowed differentiation of three homology groups. Homology group I comprised the thermotolerant campylobacters as their optimal growth temperature is 42 °C, including species such as C. jejuni subsp. jejuni (commonly referred as C. jejuni nowadays), C. jejuni subsp. doylei, C. coli and C. lari, that are most frequently isolated from human and animal diarrhoea. Homology group II was formed among others by the hydrogen-requiring species, such as C. concisus, C. showae, C. curvus and C. rectus, mainly occurring in the human oral cavity. Other species like C. fetus subsp. fetus and C. fetus subsp. venerealis most associated with bovine reproductive disorders, and C. hyointestinalis subsp. hyointestinalis and C. hyointestinalis subsp. lawsonii, primarily found in the intestines of livestock animals, comprised the homology group III.

CONCLUSION

The Campylobacter species were clustered into groups with related characteristics that originated in similar sources. Hence, phylogeny provides a suitable approach to study the evolutionary relationships among species.

Prediction of bacterial interactomes based on genome-wide coevolutionary networks: an updated implementation of the ContextMirror approach

Miguel Fernández Martín[1], *Camila Pontes*[1], *Victoria Ruiz-Serra*[1], *Alfonso Valencia*[1]

[1]*Barcelona Supercomputing Center - Life Sciences*

The biological function of proteins is preserved through coevolution and can be quantified by computing the similarity between the phylogenetic trees of pairs of protein families. When the phylogenetic similarity is high, it indicates that proteins are likely to interact. However, this similarity is influenced by many factors, including background evolution. Current coevolution-based methods treat protein pairs independently, despite proteins interacting with multiple others. The ContextMirror methodology evaluates coevolution by integrating the influence of every interactor on a given protein pair (coevolutionary network), providing more accurate

protein-protein interaction predictions. In our study, we evaluate the ContextMirror pipeline, already shown to improve the prediction of protein-protein interactions, by predicting protein-protein interactions for the full proteome of Escherichia coli (4298 proteins). Preliminary predictions reveal the potential of this approach to improve our understanding of protein coevolution. The true positive rate of the top-500 predictions (≈ 50% accuracy) is approximate to other methods and compared to the STRING database, they map only to high-confident pairs (confident score > 0.8). In the next steps of our analysis, ContextMirror will be used to identify differences in bacterial interactomes, with potential implications in drug design and protein engineering.

Sex differences in Alzheimer's disease through meta-analysis of single cell studies

Borja Gómez Cabañes[1], *Francisco García García*[1], *Marta R. Hidalgo*[1], *Irene Soler Sáez*[1], *Fernando Gordillo Gonzalez*[1], *Héctor Carceller*[2]

[1]*Unidad de Bioinformática y Bioestadística, CIPF (Centro de Investigación Príncipe Felipe),* [2]*Institute of Biotechnology and Biomedicine (BIOTECMED), Universitat de València, Burjassot, Spain*

Alzheimer's disease is a neurodegenerative disease that causes memory loss, cognitive deficits, and behavioral changes, ultimately leading to the death of the patient, and is currently the most common type of dementia. Although Alzheimer's disease affects both sexes, females exhibit a higher incidence, prevalence, and a faster progression than males, with more prominent behavioral changes. All of this suggests that there are sex differences in the pattern of neurodegeneration caused by Alzheimer's disease. Despite the distinct epidemiological features between sexes, there are few studies focused on understanding and explaining such differences. Therefore, the objective of this work is to study and analyze the underlying mechanisms of these sex-related differences in the development of Alzheimer's disease, through the integration of various single-cell transcriptomic (scRNA-seq) studies on brain tissue from Alzheimer's disease patients.

Regarding this objective, a systematic review following the PRISMA criteria has been carried out. As a result, we selected 4 scRNA-seq prefrontal cortex studies including information on the sex of the patients. For each study, quality control and processing was performed to obtain the main characteristic cell types of brain tissue. Subsequently, we conducted a differential expression analysis and, finally, a meta-analysis was performed to integrate the data obtained from each of the studies. As preliminary results we have obtained a large number of differentially expressed genes, as well as several biological pathways related to stress and protein folding in oligodendrocytes, in addition to synaptic problems in the case of neurons, all of which are altered only in females.

Detecting the mechanisms involved in the sex differences of Alzheimer's disease provides valuable information, not only about the disease itself and how it's progression, but also opens the possibility for the development of drugs and treatments tailored to the patient's sex.

Sex-Based Variations in the Extracellular Vesicle Lipidome Induced by Alcohol Use Disorder

Carla Perpiñá-Clérigues[1], *Susana Mellado-Valero*[2], *Cristina Galiana-Roselló*[3], *María Pascual-Mora*[2], *Francisco García-García*[4]

[1]*Bioinformatics and Biostatistics Unit, CIPF - School of Medicine and Dentistry, University of Valencia,* [2]*School of Medicine and Dentistry, University of Valencia - Department of Mol. and Cellular Pathology of Alcohol (CIPF),* [3]*Department of Inorganic Chemistry, Institute of Molecular Science, University of Valencia,* [4]*Bioinformatics and Biostatistics Unit, Príncipe Felipe Research Center*

BACKGROUND

Alcohol Use Disorder is a chronic brain disease that hinders the ability to control drinking habits, leading to physical, mental, and social negative consequences. Moreover, exosome lipidomics is an emerging field for studying brain disorders such as Alcohol Use Disorder. This is because extracellular vesicles can cross the blood-brain barrier, making them non-invasive biomarkers that can help understand the molecular responses underlying brain diseases. In addition, lipid dysregulation has been linked to several inflammatory and neurological disorders, so lipid species could have biomedical applications. Likewise, extracellular vesicles isolated from adolescents exposed to alcohol intoxication demonstrated a sex-based difference in their microRNA and lipid profiles. Therefore, the aim of this study is to investigate how Alcohol Use Disorder affects the extracellular vesicles lipidome differently by sex in adults and the involvement of the immune response.

DESCRIPTION

To investigate the differences in the lipidome of female and male individuals with Alcohol Use Disorder compared to controls, we utilized a lipidomics computational strategy in R programming. Limma package was used to conduct differential abundance analysis. RefMet database was employed for lipid class annotation and a class enrichment were done using the mdgsa package to evaluate the biological function. We also used the web app LINEX2 to analyze lipid metabolic networks. Our findings revealed 32 and 39 lipids that were significantly different between Alcohol Use Disorder and control groups in females and males, respectively, with only 10 being shared. Additionally, significant sex-based differences were observed in the abundance of 15 lipids. The class enrichment analysis showed that Unsaturated FA, FAHFA, and Cer_NDS lipid subclasses were differently enriched between sexes. Moreover, the lipid metabolomic network suggested distinct enzymatic dysregulations between females and males.

CONCLUSION

The findings imply that Alcohol Use Disorder differentially alters the extracellular vesicles lipidome between sexes, highlighting the importance of sex-specific biomarkers to enhance the understanding and individualized treatment of Alcohol Use Disorder.

Uncovering sex differences in Parkinson's Disease through metaanalysis of single cell transcriptomic studies

Fernando Gordillo-González[1], *Zoraida Andreu*[2], *Irene Soler-Sáez*[1], *Héctor Carceller*[3], *Marta R. Hidalgo*[1], *Francisco García-García*[1]

[1]*Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), Valencia, Spain,* [2]*Foundation Valencian Institute of Oncology (FIVO), Valencia, Spain,* [3]*Institute of Biotechnology and Biomedicine (BIOTECMED), Universitat de València, Burjassot, Spain*

BACKGROUND

Parkinson's disease is the second most frequent neurodegenerative disorder in the population, as well as the most common movement disorder associated with aging. Its cause remains unknown, its diagnosis is mainly clinical and currently there is no treatment able to slow down the neurodegenerative process. Its variability between individuals makes Parkinson's disease an adequate prospect for precision medicine. In this subject, sex variable is a promising candidate for clustering Parkinson's disease patients because differences in prevalence, age of onset and symptoms have been described. Nevertheless, molecular mechanisms behind these sex differences remain unclear. This study aims to achieve further knowledge about these underlying mechanisms of sex differences in Parkinson's disease by performing a novel integration of single cell transcriptomic (scRNA-seq) studies accomplished in brain samples with Parkinson's disease.

DESCRIPTION

The systematic review carried out, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria, identified 4 scRNA-seq studies in post-mortem substantia nigra samples with sex information of the patients involved. In these studies, the processing, identification of cell types, and differential expression analysis of each identified cell type (neurons, astrocytes, microglia and oligodendrocytes) was carried out individually. Subsequently, a metaanalysis was performed as a strategy for integrating these individual outcomes, in order to obtain the genes that showed the same behavior throughout all the studies, thus providing great robustness to the results obtained. This novel computational strategy identified some genes and functional pathways that could be responsible for the observed differences between sex in Parkinson's disease, for instance, the higher representation of axon outgrowth in neurons in females.

CONCLUSION

We identified sex-differentially altered genes and functions in Parkinson's disease in every major cell type analyzed. These results may lead not only to a deeper understanding of the underlying mechanisms of disease progression in each sex, but also to an improvement in the sex-specific treatment of patients and disease monitoring strategies.

#SCS2023 SCS2023_ ISCB_SCS ISCBSC

Contact:
symposium@iscbsc.org
www.scs2023.iscbsc.org