Interuniversity PhD program in Bioinformatics

Annual Internal Workshop

Aula Magna

Universitat de Vic – Universitat Central de Catalunya

Vic, February 3rd 2023

Book of abstracts

UAB Universitat Autònoma de Barcelona	Universitat de Barcelona	Universitat de Girona	Universitat de Lleida
UNIVERSITAT DE VIC UNIVERSITAT CENTRAL DE CATALUNYA	Universitat Oberta de Catalunya	UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH	UNIVERSITAT ROVIRA i VIRGILI

In collaboration with

BIOINFORMATICS BARCELONA

The organizing committee is composed by the members of the academic comission of the PhD program:

- Xavier Daura, coordinator of the PhD program, UAB
- Alexandre Sànchez, UB
- Beatriz López, UdG
- Rui Alves, UdL
- Jordi Villà-Freixa, UVic-UCC
- Ferran Prados, UOC
- Alexandre Perera, UPC
- Sergio Gómez, URV

The secretariat is composed by:

- Jordi Villà-Freixa, UVic-UCC
- Àlex Sánchez, UB
- Nayanika Das, UVic-UCC
- Meritxell Pujolassos, UVic-UCC
- Jing Yang, UVic-UCC
- Yolanda Tristancho, Uvic-UCC

SUPPORTED BY
The event is supported by

Vicerectorat de Recerca i Transferència de Coneixement UVIC·UCC
UAB Universitat Autònoma de Barcelona
ESCOLA DE DOCTORAT UVIC·UCC
FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES UVIC | UVIC·UCC

with the collaboration of

U Divulga Ciència i Societat UVIC·UCC
Bi-squared Research Group Bioinformatics and Bioimaging UVIC·UCC

Program

When	Who	What
9:15	Opening: Xavier Daura (UAB, Coordinator of the Interuniversity PhD program) and Antoni Tort (Director of the PhD school, UVic-UCC), Jordi Villà-Freixa (UVic-UCC)	
9:30	**Mar Albà (IMIM/ICREA)**	**Invited talk: "Uncovering the small proteome"**
10:15	Miquel Estévez Gay (UdG)	Computational Exploration and Design of new Halohydrin Dehalogenase variants
10:35	Guillem Vila Julià (UPC)	Allosterism in Bak: Sheding the Light in the Regulation of Apoptosis
10:55	*Coffee break and poster session 1(posters 10-33)*	
11:50	Gülnur Ungan (UAB)	MRSI-detected pattern in glioblastoma patients one month after concomitant chemoradiotherapy
12:10	Oriol Basallo Clariana (UdL)	Changing terpenoid biosynthesis in rice through synthetic biology
12:30	Marcos Lacasa Cazcarra (UOC)	The use of oxygen consumption in the CPET test as a biomarker in chronic fatigue syndrome patients
12:50	Ariadna Acedo (UVic-UCC)	Role of RNA-seq Preprocessing Steps In Molecular Subtype Classification In Muscle-invasive Bladder Cancer
13:10	*Lunch time*	
14:30	Mónica Cabrera Pasadas (UPC)	p53 as a master regulator of the chromatin organization
14:50	Rubén G. Barriada (UOC)	Deep-learning-based Methods for Cardiovascular Risk Assessment with Retinal Images
15:10	*Coffee break and poster session 2 (posters 34-58)*	
16:10	Laura Tiesler (UAB)	Detection of heme binding sites for the design of artificial hemoenzymes
16:30	**María Ortiz (BMS)**	**Invited talk: Disease Strategy for Translational Medicine (case study on Multiple Myeloma)**
17:15	Best-poster announcement and closing	

1. KEYNOTE: Uncovering the small proteome

Mar Albà

1 ICREA
2 Evolutionary Genomics Group, Research Program on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Barcelona 08003, Spain

Abstract

Emerging sequencing techniques have revealed that there is a large number of small non-canonical proteins that have remained unannotated. They are translated from alternative open reading frames or from regions previously believed to be non-coding, such as long non-coding RNAs (lncRNAs) and untranslated regions (UTRs). Some of these small ORFs show strong phylogenetic conservation but other are species- or lineage-specific. Using studies in yeast, we provide evidence that these proteins are important for short evolutionary time scale adaptations. We also show that they can provide plenty of raw material for de novo gene birth, a process by which proteins with completely new sequences emerge from previously non-coding parts of the genome.

2. KEYNOTE: Disease Strategy for Translational Medicine (case study on Multiple Myeloma)

María Ortíz

1 Bristol Myers Squibb

Abstract

The treatment landscape in many cancers has changed dramatically within the past years, providing more and more opportunities for patients. This increase in treatment options has also impacted the way we think and approach drug development, having now a much more complex environment to deal with. The field is going towards precision medicine, but how we get there is key. Here I will be presenting how improving our knowledge of disease heterogeneity can help us improve future treatments for patients in unmet need.

1. ORAL: Computational Exploration and Design of new Halohydrin Dehalogenase variants

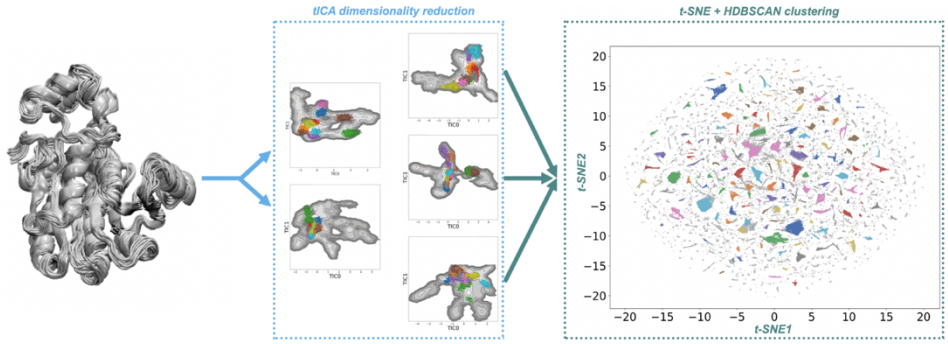
Miquel Estévez-Gay[1,*], Javier Iglesias-Fernández[1,3], Sílvia Osuna[1,2]

1. CompBioLab Group, Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, c/Maria Aurèlia Capmany 69, 17003 Girona, Catalonia, Spain,
2. ICREA, Passeig Lluís Companys 23, 08010 Barcelona, Catalonia, Spain
3. Nostrum Biodiscovery, Carrer de Baldiri Reixac, 10–12, 08028 Barcelona, Catalonia, Spain.

Abstract

Many enzymes present in Nature could be potentially used to synthesize chemically relevant chiral intermediates for drugs. One example is the enzyme family halohydrin dehalogenases (HHDH)[1], which catalyse the enzymatic conversion of ethyl (S)-4-chloro-3-hydroxybutyrate (ECHB) into the corresponding epoxide. This enzyme in the presence of a chloride ion can also catalyse the epoxide ring opening reaction yielding an ethyl (R)-4-cyano-3-hydroxybutyrate (HN), which is a precursor of Lipitor, a drug used for lowering the levels of cholesterol in blood. Within this enzyme family, different subclasses of HHDH have been identified showing a big spectrum of beneficial properties like thermal and organic solvent stability, activity, enantioselectivity and expanded substrate scope[2].

The computational exploration of the conformational landscape of HHDH can reveal the key residues and conformations that grant these enzymes the properties mentioned. In this talk, Molecular Dynamics (MD) simulations coupled with machine learning and correlation-based tools are used to explore and analyze the conformations of different HHDH subclasses[3,4] and mutants. These results, coupled with Cluster Model calculations for the WT system and variants, provide critical information on the protein k_{cat} and thus explain the variants that showed no difference in the landscape. With this knowledge, a rational in silico design protocol of some HHDHs was conducted to design new HHDHs being tested in the lab.



References

[1] RM, de Jong. *EMBO J*. **2003**, 22 (19), 4933–4944.

[2] A, Schallmey; M, Schallmey. *Appl. Microbiol. Biotechnol*, **2016**, 100, 7827–39.

[3] M. Estévez-Gay, *Catalysts* **2020**, *10*(12), 1403.

[4] J, Wessel; G, Petrillo; M, Estevez-Gay *FEBS J.*, **2021.** DOI: 10.1111/febs.15777

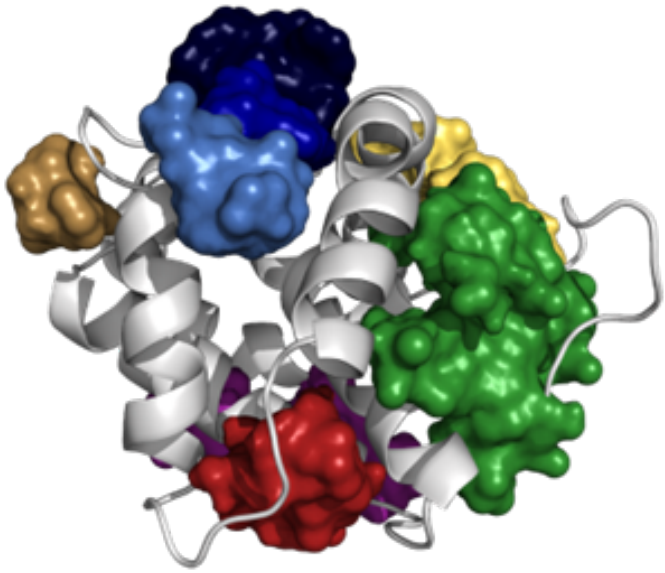
2. ORAL: Allosterism in Bak: Sheding the Light in the Regulation of Apoptosis

Guillem Vila-Julià[1,2], Juan J. Pérez[1] and Jaime Rubio-Martinez[2]

1 Department of Chemical Engineering. Universitat Politècnica de Catalunya-Barcelona Tech. Av. Diagonal, 647. 08028, Barcelona, Spain
2 Department of Materials Science and Physical Chemistry, University of Barcelona and the Institute de Recerca en Química Teòrica i Computacional (IQTCUB). C. Martí I Franqués, 1. 08028, Barcelona, Spain

Abstract

Apoptosis is a programmed cell death mechanism involved in many essential functions in multicellular organisms. (Taylor *et al.*, 2008) Thus, its deregulation has a major impact in disease: an excessive response can lead to ischemic conditions or drive neurodegeneration, whereas an inefficient response has a major impact in tumour processes. (Favaloro *et al.*, 2012)



Bak proapoptotic protein is directly involved in the permeabilization of the mitochondrial outer membrane, a crucial step during apoptosis. This protein, member of the Bcl-2 family of proteins, share a common structure with the other members of this family, which means a selectivity complexity when trying to identify new molecules capable of alter Bak activation. (Czabotar *et al.*, 2014; Chipuk *et al.*, 2010) As any drug capable of activating Bak has been identified yet, we have used different Bak structures to study the potential binding sites using FTMap and D3Pockets. Moreover, we have used our semi-automatic and systematic procedure called fragment-dissolved Molecular Dynamics to better decipher new allosteric binding sites and use them as starting points for docking studies.

References

Chipuk,J.E. et al. (2010) The BCL-2 Family Reunion. 37, 299–310.

Czabotar,P.E. et al. (2014) Control of apoptosis by the BCL-2 protein family: Implications for physiology and therapy. Nat. Rev. Mol. Cell Biol., 15, 49–63.

Favaloro,B. et al. (2012) Role of Apoptosis in Disease. Aging (Albany. NY)., 4, 330–349.

Taylor,R.C. et al. (2008) Apoptosis: Controlled demolition at the cellular level. Nat. Rev. Mol. Cell Biol., 9, 231–241.

3. ORAL: MRSI-detected pattern in glioblastoma patients one month after concomitant chemoradiotherapy

Gulnur Semahat Ungan[1,2] ; Albert Pons Escoda[3] ; Daniel Ulinic[2] ; Carles Arús[2,1] ; Alfredo Vellido[1,4] ; Carles Majós[1,3] ; Margarida Julià-Sapé[2,1]

[1]Centro de Investigación Biomédica en Red, Spain ;
[2]Dept de Bioquímica i Biología Molecular, Biociències, Universitat Autònoma de Barcelona, Spain ;
[3]Institut d'Investigació Biomèdica de Bellvitge (Idibell), Hospital Universitari de Bellvitge, Spain ;
[4]IDEAI-UPC research center, UPC BarcelonaTech, Spain

Abstract

The standard treatment for Glioblastoma (GB) patients includes surgical removal of the tumour followed by concomitant treatment with chemoradiotherapy with temozolomide (TMZ)[1]. The first treatment response evaluation is performed one month after the end of concomitant treatment (P1M). Patient follow-up is necessary because an abnormal contrast-enhancing region (CER) at this time point can be due to real tumor growth (true progression) or to pseudoprogression, a phenotype caused by treatment-associated changes. Progression status can only be determined in the next exploration, two months after P1M (P3M). We evaluated whether the MRS metabolic patterns of the CER regions at the P1M exam are predictive of the progression status two months later. We retrospectively analysed 53 consecutive patients' magnetic resonance spectroscopic imaging (MRSI) data, explored at the "Hospital de Bellvitge" between Jul. 2016-Jan. 2019. Pooled spectra were used to extract characteristic spectral patterns by convex-Non-Negative Matrix Factorisation (cNMF)[2,] which factorizes the data matrix (***X***) into two non-negative matrices: one including the so-called sources (***F***) with dimensions ***d×k*** where k is the number of sources, and d is the data dimension, and another (***H***) with the source-mixing coefficients, with dimensions ***k×n***, where n is the number of spectra. The values of H were used to calculate a "winning source" (predominant metabolic pattern for each spectrum) that was then used to create nosologic maps as overlays onto the reference MRI displaying the CER regions of each patient. H values were also used as input features for supervised Machine Learning-based classification with logistic regression (LR), support vector machines (SVM), random forests (RF) and linear discriminant analysis (LDA). Association was measured using Cramer's V[3]. We found three distinct patterns or sources, the first contributed by mobile lipids at 0.9 and 1.28 ppm, as well as some choline and importantly, a 2.8 ppm polyunsaturated fatty acids (PUFA) signal. The second one has the main contributions from mobile lipids at 0.9 and 1.28 ppm. The third pattern is mainly of proliferative type, with a choline-containing compounds being the tallest peak, also showing mobile lipids at 0.9, 1.28 ppm. Cramer's V value was 0.38, indicating a moderate association of the sources with the outcome. The best performing classifier was LDA with a balanced accuracy and area under the curve (AUC) of 80%. Therefore, at P1M after the Stupp treatment protocol, patients with GB have distinctive metabolic patterns, that are associated with response, but a perfect classification is not possible yet due to tumour heterogeneity.

References

1. Stupp,R. et al. (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. New England Journal of Medicine, 352, 987–996.

2. Ding,C.H.Q. et al. (2010) Convex and semi-nonnegative matrix factorizations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32, 45–55.

3. Cohen,J. (1988) Statistical Power Analysis for the behavioral sciences L. Erlbaum Associates, Hillsdale, NJ.

4. ORAL: Changing terpenoid biosynthesis in rice through synthetic biology

Oriol Basallo[1,2], Abel Lucido[1,2], Rui Alves[1,2], Lucia Perez [3,4], Paul Christou[3,4,5], Teresa Capell[3,4], Ester Vilaprinyo[1,2]

[1]Faculty of Medicine, Universitat de Lleida.
[2]IRBLleida.
[3]ETSEA, Universitat de Lleida.
[4]Agrotecnio CERCA Center.
[5]ICREA.

Abstract

Many high valued chemicals in the pharmaceutical, biotechnological, cosmetic, and biomedical industries belong to the terpenoid family. Biosynthesis of these chemicals relies on polymerization of Isopentenyl di-phosphate (IPP) and/or dimethylallyl diphosphate (DMAPP) monomers, which plants synthesize using two alternative pathways: a cytosolic mevalonic acid (MVA) pathway and a plastidic methyleritritol-4-phosphate (MEP) pathway. As such, developing plants for use as a platform to use IPP/DMAPP and produce high value terpenoids is an important biotechnological goal. Still, IPP/DMAPP are the precursors to many plant developmental hormones. This creates severe challenges in redirecting IPP/DMAPP towards production of non-cognate plant metabolites. A potential solution to this problem is increasing the IPP/DMAPP production flux *in planta*. Here, we aimed at discovering, understanding, and predicting the effects of increasing IPP/DMAPP production in plants through modelling.

We used synthetic biology to create rice lines containing an additional ectopic MVA biosynthetic pathway for producing IPP/DMAPP (Pérez et al., 2022). The rice lines express three alternative versions of the additional MVA pathway in the plastid, in addition to the normal endogenous pathways. We collected data for changes in macroscopic and molecular phenotypes, gene expression, isoprenoid content, and hormone abundance in those lines. To integrate the molecular and macroscopic data and develop a more in depth understanding of the effects of engineering the exogenous pathway in the mutant rice lines, we developed and analyzed data-centric, line-specific, multilevel mathematical models. These models connect the effects of variations in hormones and gene expression to changes in macroscopic plant phenotype and metabolite concentrations within the MVA and MEP pathways of WT and mutant rice lines.

Our models allow us to understand how an exogenous IPP/DMAPP biosynthetic pathway affects the flux of terpenoid precursors. They also allow us to quantify the effect of hormonal regulation on the alternative IPP/DMAPP biosynthetic pathways, enabling the prediction of macroscopic plant characteristics from molecular data. In addition, the line-specific models are a tool that can help in prioritizing the most optimal re-engineering strategy for the exogenous pathway in order to achieve specific metabolic goals.

References

Pérez,L. *et al.* (2022) Multilevel interactions between native and ectopic isoprenoid pathways affect global metabolism in rice. *Transgenic Res.*, **31**, 249–268.

5. ORAL: The use of oxygen consumption in the CPET test as a biomarker in chronic fatigue syndrome patients

Marcos Lacasa[1], Patricia Launois[5], Ferran Prados[1,2,3,4], José Alegre[5], Jordi Casas-Roma[1]

[1] e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain
[2] Center for Medical Image Computing, University College London, London, United Kingdom
[3] National Institute for Health Research Biomedical Research Centre at UCL and UCLH, London, United Kingdom
[4] Queen Square MS Center, Department of Neuroinflammation, UCL Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom
[5] Myalgic Encephalomyelitis / Chronic Fatigue Syndrome Unit, Division of Rheumatology, Vall d'Hebron Hospital Research Institute Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

Background

Chronic fatigue syndrome (CFS) is a disabling chronic disease. The search for a biomarker to objectively determine the health status of CFS patients has been ongoing in recent years. The aim of the study is to demonstrate that oxygen consumption, by means of a cardiopulmonary exercise stress test (CPET) is a biomarker for grading CFS patients.

Method and Results.

This study included 92 patients who fulfilled the 1994 CDC/Fukuda definition and 2003 Canadian Consensus Criteria for CFS. Short form 36-item health survey (SF-36) questionnaire and both the response matrix and subscales and peak oxygen consumption during CPET and Weber classification were analysed. It is shown that for clustering by unsupervised machine learning it is better to use the decoded answers matrix. 239 results of the CPET stress test have been analysed, and the peak oxygen consumption is considered to classify according to Weber's classification. A contingency table of 92 validated records of patients who have performed the CPET test and answered the SF-36 questionnaire is constructed and labeled according to the cluster and the assigned Weber classification. The result shows that a worse Weber classification infers a worse result on the SF-36.

Conclusion

The use of oxygen consumption in the CPET test should be considered for use as a biomarker for the status of the diagnosed CFS patient. Clustering of records from health questionnaires such as the SF-36 should be performed with the decoded response data as it maintains the initial information and improves the quality of the model.

6. ORAL: Role of RNA-Seq Preprocessing Steps In Molecular Subtype Classification In Muscle-Invasive Bladder Cancer

Acedo-Terrades, Ariadna,[1] Perera-Bel, Júlia,[1] Bellmunt, Joaquim,[1,2] Nonell, Lara[3]

[1] Hospital del Mar Medical Research Institute (IMIM), Research Programme of Biomedical Informatics (GRIB): Barcelona, Spain
[2] Division of Hematology and Oncology, Beth Israel Deaconess Medical Center; Boston, USA
[3] Bioinformatics Unit, Vall d'Hebron Institute of Oncology (VHIO) ; Barcelona, Spain

Abstract

Muscle-invasive bladder cancer (MIBC) is an heterogeneous disease that is characterized by genomic instability and a high mutation rate. For that reason, transcriptome profiling has been used to classify MIBC into six molecular subtypes , for a more precise patient stratification according to prognosis, clinical characteristics and therapeutic options (Kamoun, 2020). One of the most common approaches to obtain transcriptome profiling is RNA-Seq technology, which consists of several preprocessing steps (ie. alignment, quantification) in which a great number of methods can be used to obtain the final table of counts. Our hypothesis is these steps affect the resulting table of counts and, hence, might also influence the classification into molecular subtypes. Assess the role of RNASeq preprocessing steps in the molecular subtype classification of bladder cancer samples A study of different preprocessing methodologies was conducted by comparing: STAR and Hisat2 tools for the alignment step, and featureCounts, HTSeq, StringTie and RSEM for the quantification step. Regarding the normalization step, the methodologies that were used are: TPM, log2TPM, TMM, rawData and log2rawData. We applied the pipelines to bladder cancer samples from 3 datasets from the GEO database. Accuracy was evaluated using the classification label obtained from the consensusMIBC classifier (Kamoun, 2020). The most shared subtype across the methods was selected as the gold standard for each sample. Our preliminary results demonstrate that STAR was the best aligner, producing the highest accuracy and being the most stable classification results across the different quantification and normalization methods. Regarding quantifiers, almost all of them have high accuracy. Finally, TMM and log2TPM were the normalization methods that showed a high accuracy across almost all the methods used for alignment and quantification steps. Conclusions: According to our results, we propose STAR+featureCounts+TMM as the most accurate pipeline to generate a counts table to use for downstream molecular subtype classification.

References

1. Kamoun,A. Et al (2020) A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. European Urology, 77, 420-43

7. ORAL: p53 as a master regulator of the chromatin organization

Cabrera-Pasadas M,[1,2] Javierre BM,[2] Valencia[1]

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

Abstract

ATP53 is a tumor suppressor gene that codes for p53, a sequence-specific DNA binding protein that regulates transcription of hundreds of genes upon cellular damage to promote cell arrest and/or apoptosis. In normal unstressed cells, MDM2 promotes p53 degradation maintaining its cellular levels low. However, cellular damage, or MDM2 pharmacological inhibition, leads to p53 accumulation and activation. TP53 is one of the most frequently mutated genes in cancer and p53 response is deficient in more than half of the tumors. For this reason, the complete understanding ofthe molecular mechanism by which p53 regulates gene expression and induces tumor-suppressive and anti-proliferative effect as a consequence of cellular damage is a clinical need. Spatial-temporal genome architecture, which is cell type specific, plays a key role in gene regulation, and it could be dynamically rewired by p53 as a mechanism of transcriptional regulation. p53 could support the formation of DNA loops that physically connect regulatory elements and gene promoters to allow gene transcription. To test this hypothesis, we propose to decipher the potential effect of p53 activation via MDM2 inhibition on genome architecture based on a "omics" approach that integrates Promoter Capture Hi-C, RNA-seq and ChIP-seq data. Preliminary analysis demonstrates that, in contrast to cells lacking activated p53, p53 activated cells present higher number of significant gene promoter-anchored interactions, especially at the range between 300 and 1000 kilobases of distances. These p53-specific interactions mainly connect gene promoters with other genomic region that could contain regulatory elements. Therefore, these new interactions could imply a mechanism of p53-mediated gene expression control. Contrarily, no difference in co-localization of gene promoters are observed. In addition, network analysis shows that p53 activated cells are characterized by higher number of connected hubs in comparison with cells without p53 activated reinforcing the role of p53 in chromatin organization. Finally, promoter-centred interactions specific of p53 activated cells are enriched in p53 binding at interacting regions, suggesting that part of changes on genome architecture are a direct effect of p53. Therefore, these results lead us to postulate that activated p53 may act as a master regulator in the organization of thechromatin.

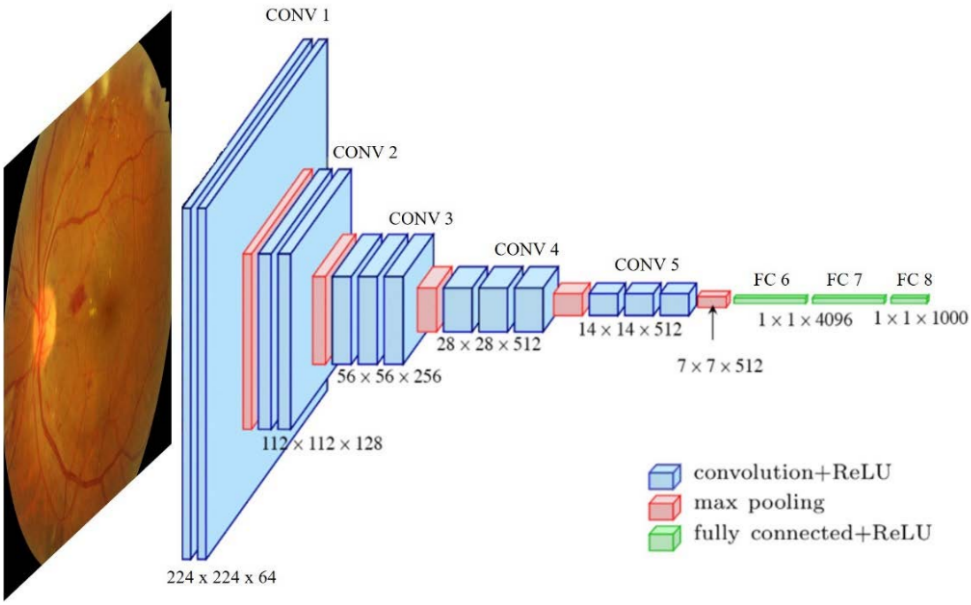
8. ORAL: Deep-learning-based Methods for Cardiovascular Risk Assessment with Retinal Images

Rubén G. Barriada and David Masip

AIWell Research Group, Faculty of Computer Science, Multimedia and Telecommunications, Universitat Oberta de Catalunya, 08018 Barcelona, Spain

Abstract

The progress of computing systems during the last years has allowed deep learning (DL), a sub-field of AI, to become a feasible methodology to analyze complex sources of data, such as medical images. Within medical imaging techniques, retinal photography analysis has gained popularity due to its noninvasive and cost-effective nature. Different biomarkers and eye structures can be identified from a RFI, playing an important role in identifying retinal abnormalities and diseases. In recent years, DL applied to oculomics has aroused great interest in the scientific community. Studies on the identification and prediction of ocular biomarkers of systemic diseases are becoming increasingly interesting for researchers in the field . DL techniques are providing insights about eye–body associations through retinal morphology analysis to enhance the understanding of complex disorders, such as musculoskeletal diseases, traumatic brain injury, cardiovascular disease, renal impairment, Alzheimer's disease or anemia detection. The motivation of this thesis is to study the feasibility of predicting cardiovascular events using deep learning models for reliable and robust application in medical environments. In an initial phase, we prepared the methodology (DL) and performed the necessary training to train networks on existing problems with public retinal image databases. In a central phase, the problem was focused on oculomics, focusing the thesis on the prediction of cardiovascular problems, and starting to work on a specific predictor such as the CAC score, with a paper already published in a journal. In the final phase, we hope to complete the thesis by extending the experiments to specific cardiovascular events, applying self-supervision, multi-modality and improvements in the architectures for infarct prediction using retinal imaging.



References

1. Barriada, R. G., Simó-Servat, O., Planas, A., Hernández, C., Simó, R., & Masip, D. (2022). Deep Learning of Retinal Imaging: A Useful Tool for Coronary Artery Calcium Score Prediction in Diabetic Patients. Applied Sciences, 12(3). https://doi.org/10.3390/app12031401

2. Cheung, C. Y., Ran, A. R., Wang, S., Chan, V. T. T., Sham, K., Hilal, S., Venketasubramanian, N., Cheng, C.-Y., Sabanayagam, C., Tham, Y. C., Schmetterer, L., McKay, G. J., Williams, M. A., Wong, A., Au, L. W. C., Lu, Z., Yam, J. C., Tham, C. C., Chen, J. J., … Wong, T. Y. (2022). A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. The Lancet Digital Health. https://doi.org/10.1016/s2589-7500(22)00169-8

3. Goutam, B., Hashmi, M. F., Geem, Z. W., & Bokde, N. D. (2022). A Comprehensive Review of Deep Learning Strategies in Retinal Disease Diagnosis Using Fundus Images. IEEE Access.

4. Harris, G., Rickard, J. J. S., Butt, G., Kelleher, L., Blanch, R., Cooper, J. M., & Oppenheimer, P. G. (2022). Review: Emerging Oculomics based diagnostic technologies for traumatic brain injury. IEEE Rev. Biomed. Eng., PP.

5. Kim, B. R., Yoo, T. K., Kim, H. K., Ryu, I. H., Kim, J. K., Lee, I. S., Kim, J. S., Shin, D.-H., Kim, Y.-S., & Kim, B. T. (2022). Oculomics for sarcopenia prediction: a machine learning approach toward predictive, preventive, and personalized medicine. EPMA J., 13(3), 367–382.

6. Maldonado García, C., Bonazzola, R., Ravikumar, N., & Frangi, A. F. (2022). Predicting Myocardial Infarction Using Retinal OCT Imaging. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13413 LNCS, 787–797. https://doi.org/10.1007/978-3-031-12053-4_58

7. Mitani, A., Huang, A., Venugopalan, S., Corrado, G. S., Peng, L., Webster, D. R., Hammel, N., Liu, Y., & Varadarajan, A. V. (2020). Detection of anaemia from retinal fundus images via deep learning. Nature Biomedical Engineering, 4(1), 18–27.

8. Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering, 2(3), 158–164. https://doi.org/10.1038/s41551-018-0195-0

9. Rim, T. H., Lee, G., Kim, Y., Tham, Y. C., Lee, C. J., Baik, S. J., Kim, Y. A., Yu, M., Deshmukh, M., Lee, B. K., Park, S., Kim, H. C., Sabayanagam, C., Ting, D. S. W., Wang, Y. X., Jonas, J. B., Kim, S. S., Wong, T. Y., & Cheng, C. Y. (2020). Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. The Lancet Digital Health, 2(10), e526–e536. https://doi.org/10.1016/S2589-7500(20)30216-8

10. Sabanayagam, C., Xu, D., Ting, D. S. W., Nusinovici, S., Banu, R., Hamzah, H., Lim, C., Tham, Y. C., Cheung, C. Y., Tai, E. S., Wang, Y. X., Jonas, J. B., Cheng, C. Y., Lee, M. L., Hsu, W., & Wong, T. Y. (2020). A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. The Lancet Digital Health, 2(6), e295–e302. https://doi.org/10.1016/S2589-7500(20)30063-7

11. Simó, R., Bañeras, J., Hernández, C., Rodŕiguez-Palomares, J., Valente, F., Gutierrez, L., González-Alujas, T., Ferreira, I., Aguadé-Bruix, S., Montaner, J., & others. (2019). Diabetic retinopathy as an independent predictor of subclinical cardiovascular disease: baseline results of the PRECISED study. BMJ Open Diabetes Research and Care, 7(1), e000845.

12. Wagner, S. K., Fu, D. J., Faes, L., Liu, X., Huemer, J., Khalid, H., Ferraz, D., Korot, E., Kelly, C., Balaskas, K., & others. (2020). Insights into systemic disease through retinal imaging-based oculomics. Translational Vision Science & Technology, 9(2), 6–6.

9. ORAL: Detection of heme binding sites for the design of artificial hemoenzymes

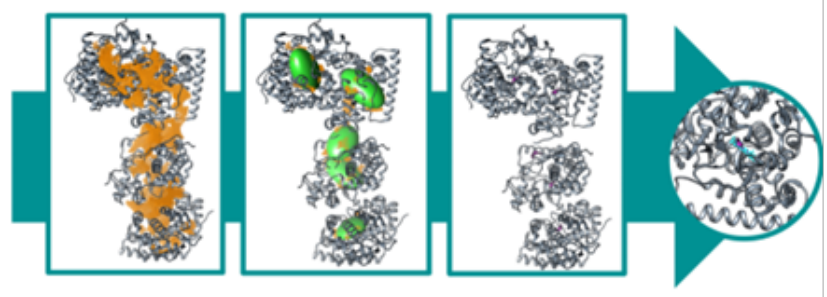
Laura Tiessler-Sala[1], Raúl Fernández-Díaz[1], Jean-Didier Maréchal[1]

[1]Universitat Autònoma de Barcelona, Departament de Química, Bellaterra, Espanya.

Abstract

More than 40% of proteins bind metals or metal complexes in order to carry out their functions, both catalytic and structural (Andreini *et al.*, 2008). **Heme** is one of the most common metal complexes in nature and it is involved in several crucial biological processes, such as oxygen transport, redox catalysis or transcription regulation (Poulos, 2014). Over the last decade, advances in the field of designing artificial enzymes have led to heme-based artificial metalloenzymes that are able to catalyse cyclopropanations and N-H or S-H insertions (Villarino *et al.*, 2018; Wang *et al.*, 2014; Kan *et al.*, 2016).

Despite the importance of heme, the binding of heme to proteins and its prediction has not been widely studied. So far, most programs are based on a combination of sequence and structure predictions. In this work we present, **HemeFinder**, a new program that allows detecting heme binding sites and designing new hemoenzymes based only on the protein structure and the geometrical predisposition of heme binding sites.



HemeFinder workflow is based on first detecting cavities within the protein and then storing them as grid points using a pyKVFinder module (Guerra *et al.*, 2021). The volume of these cavities is calculated and defined as ellipsoids to assess if a heme molecule would fit. Then, considering only geometric criteria, for each of the points in the cavity it is calculated if any surrounding amino acids would be able to coordinate or if any could be mutated in order to coordinate with heme.

References

Andreini,C. *et al.* (2008) Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem*, **13**, 1205–1218.

Guerra,J.V. da S. *et al.* (2021) pyKVFinder: an efficient and integrable Python package for biomolecular cavity detection and characterization in data science. *BMC Bioinformatics*, **22**, 607.

Kan,S.B.J. *et al.* (2016) Directed Evolution of Cytochrome c for Carbon–Silicon Bond Formation: Bringing Silicon to Life. *Science*, **354**, 1048–1051.

Poulos,T.L. (2014) Heme Enzyme Structure and Function. *Chem. Rev.*, **114**, 3919–3962.

Villarino,L. *et al.* (2018) An Artificial Heme Enzyme for Cyclopropanation Reactions. *Angewandte Chemie International Edition*, **57**, 7785–7789.

Wang,Z.J. *et al.* (2014) Cytochrome P450-Catalyzed Insertion of Carbenoids into N-H Bonds. *Chem Sci*, **5**, 598–601.

1. POSTER S1: Whole-genome assessment of runs of homozygosity in children from 6 European populations: Demographic insights, selective pressure and health effects

Laura Balagué-Dobón[1], Alejandro Cáceres[1,2.3] and Juan R. González[1,2.4]

[1] Bioinformatics Research Group in Epidemiology, ISGlobal, 08003 Barcelona, Spain
[2] Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública (CIBERESP), 08003 Barcelona, Spain
[3] Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya, 08019 Barcelona, Spain
[4] Department of Mathematics, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain

Abstract

Approximately the 2% of the average human genome is covered by genomic regions with identical haplotypes inherited from a common ancestor, namely Runs of Homozygosity (ROH). The study of the burden and distribution of ROH reveals demographic histories of populations and contributes to the understanding of inbreeding depression (Ceballos *et al.*, 2018; Pemberton *et al.*, 2012). While the causes of Mendelian diseases become apparent in ROH regions of inbred individuals (Botstein and Risch, 2003), less is known about the early effects that high-frequency ROH regions may have on the general population. We aimed to characterize ROH regions along the whole genome of 1304 children from 6 European birth cohorts and study their effect in 19 children's traits but also in their mothers' health during the pregnancy period. We observed an average of 45Mb for the total length ROH (SROH) and of 52 for total number of ROH regions (NROH), that highly varied between cohorts. We observed 139 ROH consensus regions across the cohorts, 7 of them showing divergent frequencies between cohorts, and therefore denoting a putative intra-European selective pressure. We found strong associations of ROH in 17q21.32 with ponderal index, revealing likely causal associations previously reported at a SNP level; and in 2q21.3 with gestational diabetes, supporting the idea of an active role of the placenta metabolism in the outcome of pregnancy diseases (Burton *et al.*, 2019; Lu *et al.*, 2018; Traglia *et al.*, 2018).

References

Botstein,D. and Risch,N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.

Burton,G.J. *et al.* (2019) Pre-eclampsia: pathophysiology and clinical implications. *BMJ*, **366**.

Ceballos,F.C. *et al.* (2018) Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet. 2018 194*, **19**, 220–234.

Lu,Y.P. *et al.* (2018) Fetal Serum Metabolites Are Independently Associated with Gestational Diabetes Mellitus. *Cell. Physiol. Biochem.*, **45**, 625–638.

Pemberton,T.J. *et al.* (2012) Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.*, **91**, 275–292.

Traglia,M. *et al.* (2018) Cross-genetic determination of maternal and neonatal immune mediators during pregnancy. *Genome Med.*, **10**, 1–17.

2. POSTER S2: Operational integration of data analysis models in production environments

Jan Borràs Ros

Eheus SL

Abstract

Nowadays we generate such an amount of data that is hardly managed without specialized algorithms. The amount of biological molecular data generated from any omics analysis is huge, and there is a need for standard methods to deal with such complexity when a successful data analysis research project reaches its final goal.

First we must build and deploy our model and make it operational for production though an application, which will operate in a productive environment along with other services to accomplish the final goal, which it can go from giving a user some specific insight, to operate an ecosystem of data models and give a personalized report with a lot of valuable information. The standardization methods must combine aspects from different fields, from developing operations (DevOps) with containerization and its continuous delivery/continuous integration (CD/CI), to data analysis with data modelling, or data engineering with Extract-Transform-Load (ETL) pipelines. And the final goal of this research is to highlight fundamental aspects of how all the components and their lifecycle are managed while we provide a service using a real case example in a real data analysis company, Exheus.

In Exheus it has been created an AI algorithm to classify the normality in a population using transcriptomic analysis of whole blood samples, then all this data is transformed though an structured pipeline to finally give valuable information to clients of their health status using a personalized report. This is a good example to elucidate possible problems and solutions for the operational integration of data analysis models to real world cases.

Materials and methods

To fully deploy the data modelling applications the operations pipeline has been divided into small microservices, microservices work as functional units with their own dependencies and independent lifecycle. To containerize each service Docker Engine and Kubernetes have been used to operate those containers, both are the most used tools for managing the operational units of a production environment. To manage code delivery and implementation inside each container Gitlab community edition has been used which is a custom implementation of Git software in a web platform build by a community of software developers. And data modelling and classification was performed using Python programming language.

Results and conclusions

The first phase of the deployment into production environment was the coding part of the microservices which ended with different functional units including preprocessing of data, data quality assessment, data modelling, data postprocessing, data transformation and data visualization methods each maintained independently. In parallel each microservice has been put in an independent Gitlab project to keep track of all the changes in the code structure, and to be able to implement CI/CD pipelines and a registry with a docker container for each project. Finally when all the elements were correctly deployed to a server, Exheus operations finally were ready to attend the needs of their clients.

3. POSTER S3: Methylation biomarkers for colorectal cancer early detection and survival prognostics impact gene expression and link to cancer-related biological pathways

P. Canal-Noguer[1,3,4,5], A. Requena-Bermejo[1], F. Mattia Mancuso[1], JC. Higareda-Almaraz[1], M. Manrique-López[1], M. Chersicola2, P. Pérez-Martínez[2], P.A. Camino[1], P. Knap[2], V. Erklavec Zajec[2], K. Kruusma[2]

1 Universal Diagnostics S.A., Seville, Spain
2 Universal Diagnostics d.o.o., Ljubljana, Slovenia
3 B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Spain
4 Networking Biomedical Research Centre in the subject area of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain
5 Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain

Abstract

Background

DNA methylation has been previously shown to have diagnostic and predictive potential for colorectal cancer (CRC). Aim of this study was to evaluate putative methylation markers in the context of early cancer development and diagnostics as well as further investigate the biological significance of these regions.

Methods

Biomarker discovery was done by whole genome bisulfite sequencing (WGBS) of 88 CRC, 48 advanced adenoma (AA) and corresponding adjacent normal tissue (NAT) samples. Short-list of significantly hypermethylated regions (DMRs) was correlated to transcriptomics data from 512 CRC patients in The Cancer Genome Atlas (TCGA) cohort. Pathway enrichment for biological pathway analysis of the DMRs was done by using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. Survival analysis was performed using Kaplan–Meier method on sub-groups of patients divided by the methylation status of individual markers. Finally, individual marker significance of selected regions was evaluated by analyzing 26 plasma samples from early stage (stage I-IIA) CRC samples and 42 colonoscopy verified controls (CNT) with targeted methylation sequencing assay.

Results

4167 putative marker regions were identified from biomarker discovery with WGBS. Differential signal could be observed both between AA and NAT and CRC and NAT, while several of these regions were differentially methylated also between AA and CRC samples, indicating biological signal change with adenoma progression to cancer. 84 hypermethylated DMRs from several verification studies were further evaluated against transcriptome data from TCGA, where overlap for 69 genes was found. 19 of these genes showed a significant down-regulation ($p< 0.05$), indicating a link between hypermethylation and gene expression. 2 genes showed significant up-regulation ($p< 0.05$), which could indicate other epigenic processes to be in place. KEGG pathway analysis revealed that the top pathways involved were axonal guidance, ephrin receptor signaling, epithelial-mesenchymal transition and FGF signaling, which all play significant role in the context of cancer development and progression. Kaplan-Meier analysis showed significant correlation to patients 5- year survival prediction linked to 3 genes: FGF14 ($p=0.025$, HR = 1.75) , DPY19L2P1 ($p=0.012$, HR = 1.86), PTPRO ($p=0.046$, HR = 1.63). Targeted sequencing analysis on plasma samples of patients with early stage (I-IIA) colorectal cancer and age and gender matching colonoscopy-verified controls, showed

high individual marker accuracy with AUC= 0.78 for FGF14, AUC= 0.81 for DPY19L2P1 and AUC= 0.73 for PTPRO.

Conclusions

Methylation markers have distinct signals in early development of CRC, with high individual accuracy for separating early-stage cancers from matching controls. These regions have impact on gene expression and can be linked to relevant biological pathways. Extending early detection potential of the markers to further prognostics and stratification, could lead to better outcomes and improved survival of the patients.

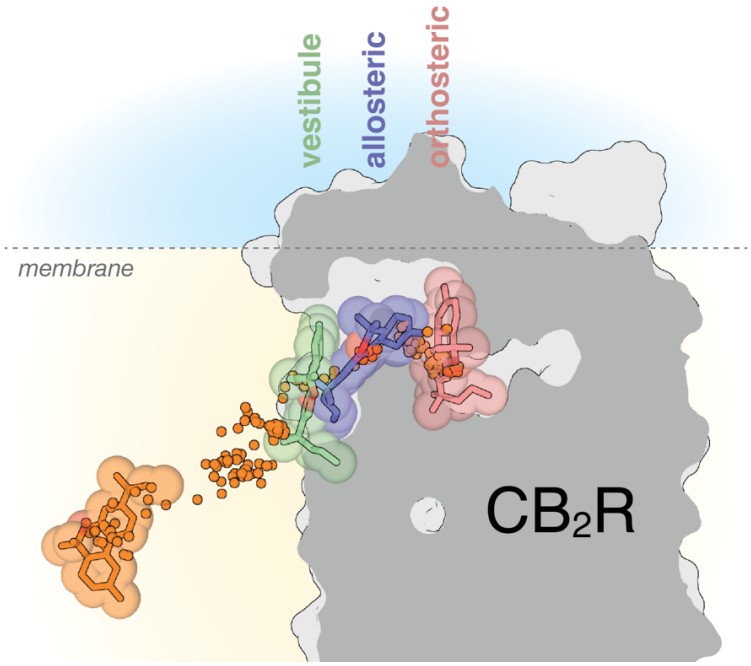
4. POSTER S4: A single point mutation blocks the entrance of ligands to the cannabinoid CB_2 receptor via the lipid bilayer

Nil Casajuana-Martin[1], Gemma Navarro[2,3], Angel Gonzalez[1], Claudia Llinas del Torrent[1], Marc Gomez Autet[1], Aleix Quintana-García[1], Rafael Franco[3,4], Leonardo Pardo[1]

[1] Universitat Autonoma de Barcelona, Laboratory of Computational Medicine, Bellaterra, Spain
[2] Universitat de Barcelona, Department of Biochemistry and Physiology, Faculty of Pharmacy and Food Sciences, Barcelona, Spain
[3] Instituto de Salud Carlos III, CIBERNED, Madrid, Spain
[4] Universitat de Barcelona, Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Barcelona, Spain

Abstract

Molecular dynamic (MD) simulations have become a common tool to study the pathway of ligand entry to the orthosteric binding site of G protein-coupled receptors. Here, we have combined MD simulations and site-directed mutagenesis to study the binding process of the potent JWH-133 agonist to the cannabinoid CB_2 receptor (CB_2R). In CB_2R, the N-terminus and extracellular loop 2 fold over the ligand binding pocket, blocking the access to the binding cavity from the extracellular environment. We, thus, hypothesized that the binding pathway is a multi-stage process consisting of the hydrophobic ligand diffusing in the lipid bilayer to contact a lipid-facing vestibule, from which the ligand enters to an allosteric site inside the transmembrane bundle, and finally moves to the orthosteric binding cavity. This pathway was experimentally validated by a single point mutation that blocked the entrance of the ligand, as JWH-133 was not able to decrease the forskolin-induced cAMP levels in the mutant receptor. This proposed ligand entry pathway defines transient binding sites that are potential cavities for the design of synthetic modulators.



5. POSTER S1: Dynorphin A as a Cell Penetrating Peptide

Èric Catalina-Hernández[1], David Masnou-Sanchez[1], Mario Lopez-Martin[1], Aguilella-Arzo, Marcel[2], Alex Peralvarez-Marin[1]

1 Unitat de Biofísica, Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona
2 Departament de Física, Escola Superior de Tecnologia i Ciències Experimentals, Universitat Jaume I (Castelló)

Abstract

Membrane proteins are crucial in connecting the cell with its environment. Most current drug therapeutic approaches are directed to membrane protein targets, mainly because of the involvement of these proteins in signalling cascade events, as well as in the transport between the exterior and the interior of the cell and among cellular compartments. Understanding how membrane proteins regulate all these processes will lead to a better knowledge of how to tackle diseases at the molecular level. Dynorphins are endogenous peptides and the canonical kappa-opioid receptor (KOR) substrate. Dynorphins are prohormones derived from prodynorphin (PDYN), which is cleaved into big dynorphin (BigDyn, 32 residues), and further processed into dynorphin A (DynA, 17 residues) and B (DynB, 13 residues)[1,2]. Furthermore, 3 clinical variants have been described for DynA: L5S, R6W, R9C. Pathophysiological implications for dynorphins have been described owing to their cell penetrating peptide (CPP)-like behaviour, as molecules capable of inducing membrane translocation via membrane destabilization and/or pore formation. Besides, they are related to Alzheimer's and Parkinson's diseases (AD, PD). Hence, we have studied DynA -and its clinical variants- CPP activity. Thus, we have simulated DynA in 3 types of membrane (DPPC, DPPC:DOPC:CHOL, and DPPC:DOPC:DPPS:DOPS:CHOL), for the 4 types of DynA and a control (R9 peptide) and have induced membrane translocation through adaptatively Steered Molecular Dynamics (aSMD) simulations[3]. Afterwards, we run a conventional MD (cMD) of 100 ns. In some cases, we observed a proteolipidic pore after the simulations, since dynorphins, as positively charged peptides, interact with the negatively charged polar head groups of phospholipids, whereas, in other cases, the pore is rapidly closed. We have computed the potential of mean force (PMF) as a way of comparing the CPP activity to peptides with similar characteristics. We are currently analysing and undergoing more simulations to extract concluding information.

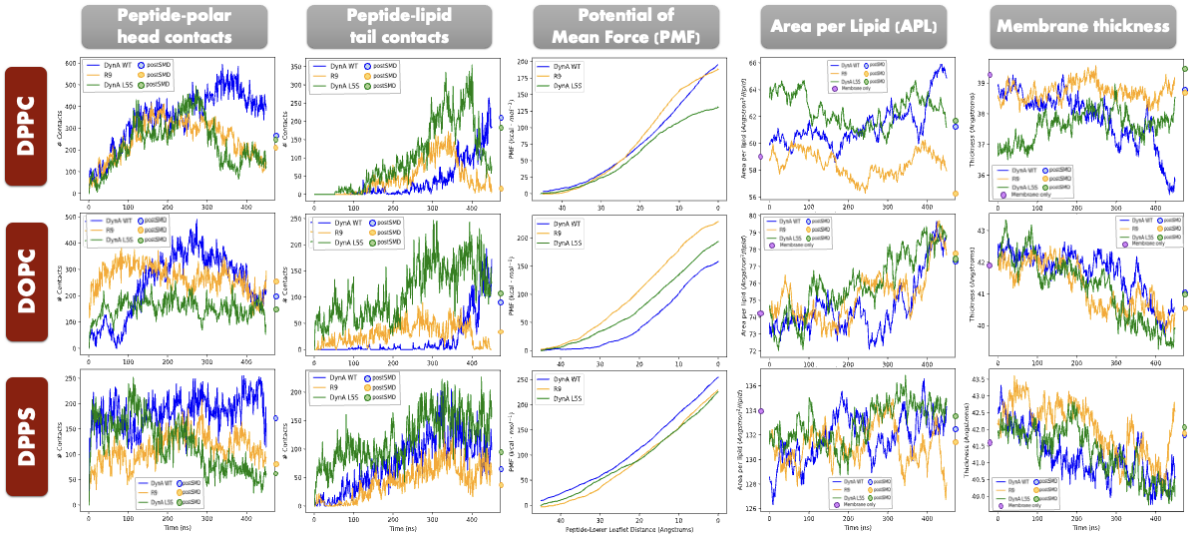


Figure 1. Analysis of peptide-lipid contacts, potential of mean force, area per lipid, and membrane thickness for each of the membranes and the 3 peptides whose simulations have finished.

References

(1) Gallego-Villarejo, L.; Wallin, C.; Król, S.; Enrich-Bengoa, J.; Suades, A.; Aguilella-Arzo, M.; Gomara, M. J.; Haro, I.; Wärmlander, S.; Muñoz, F. J.; Gräslund, A.; Perálvarez-Marín, A. Big Dynorphin Is a Neuroprotector

Scaffold against Amyloid β-Peptide Aggregation and Cell Toxicity. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 5672–5679. https://doi.org/10.1016/j.csbj.2022.10.014.

(2) Perini, D. A.; Aguilella-Arzo, M.; Alcaraz, A.; Perálvarez-Marín, A.; Queralt-Martín, M. Dynorphin A Induces Membrane Permeabilization by Formation of Proteolipidic Pores. Insights from Electrophysiology and Computational Simulations. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 230–240. https://doi.org/10.1016/j.csbj.2021.12.021.

(3) Gimenez-Dejoz, J.; Numata, K. Molecular Dynamics Study of the Internalization of Cell-Penetrating Peptides Containing Unnatural Amino Acids across Membranes. *Nanoscale Adv.* **2022**, *4* (2), 397–407. https://doi.org/10.1039/D1NA00674F.

6. POSTER S2: Machine learning computational tools to assist the performanceof systematic reviews: A mapping review.

Ramon Cierco Jimenez1,2*, Teresa Lee3, Nicolás Rosillo4, Reynalda Cordova5,6, Ian A Cree1, Angel Gonzalez2 and Blanca Iciar Indave Ruiz1

1 International Agency for Research on Cancer (IARC/WHO), Evidence Synthesis and Classification Branch, Lyon, France
2. Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain
3. International Agency for Research on Cancer (IARC/WHO), Services to Science and Research Branch, Lyon, France
4. Servicio de Medicina Preventiva, Hospital Universitario 12 de Octubre, Madrid, Spain
5. International Agency for Research on Cancer (IARC/WHO), Nutrition and Metabolism Branch, Lyon, France
6. Department of Nutritional Sciences, University of Vienna, Vienna, Austria

Abstract

Background: Within evidence-based practice (EBP), systematic reviews (SR) are considered the highest level of evidence in that they summarize the best available research and describe the progress in a determined field. Due its methodology, SR require significant time and resources to be performed; they also require repetitive steps that may introduce biases and human errors. Machine learning (ML) algorithms therefore present a promising alternative and a potential game changer to speed up and automate the SR process. This review aims to map the current availability of computational tools that use ML techniques to assist in the performance of SR, and to support authors in the selection of the right software for the performance of evidence synthesis.

Methods: The mapping review was based on comprehensive searches in electronic databases and software repositories to obtain relevant literature and records, followed by screening for eligibility based on titles, abstracts, and full text by two reviewers. The data extraction consisted of listing and extracting the name and basic characteristics of the included tools, for example a tool's applicability to the various SR stages, pricing options, open-source availability, and type of software. These tools were classified and graphically represented to facilitate the description of our findings.

Results: A total of 9653 studies and 585 records were obtained from the structured searches performed on selected bibliometric databases and software repositories respectively. After screening, a total of 119 descriptions from publications and records allowed us to identify 63 tools that assist the SR process using ML techniques.

Conclusions: This review provides a high-quality map of currently available ML software to assist the performance of SR. ML algorithms are arguably one of the best techniques at present for the automation of SR. The most promising tools were easily accessible and includea high number of user-friendly features permitting the automation of SR and other kinds of evidence synthesis reviews.

Keywords: Systematic reviews, Mapping review, Evidence-based practice, Software development, Machine learning, Automatization.

Published article accessible at BMC Medical Research Methodology
https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01805-4#author-information

7. POSTER S3: Image-based survival model for multiple sclerosis: a preliminary study

Llucia Coll[1], Deborah Pareto[2], Xavier Lladó[3], Carmen Tur[1]

[1] Multiple Sclerosis Centre of Catalonia (Cemcat), Department of Neurology/Neuroimmunology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
[2] Magnetic Resonance Unit, Dept of Radiology, Vall d'Hebron University Hospital, Barcelona, Spain
[3] Research Institute of Computer Vision and Robotics, University of Girona, Girona, Spain

Abstract

Multiple sclerosis (MS) is the most common disabling (non-traumatic) condition in young adults. Progression independent of relapse activity (PIRA) is considered the main mechanism through which patients with MS accumulate irreversible disability. Thus, early and accurate predictions of long-term risk of presenting a first PIRA event after symptom onset are crucial for patient management (Montalban *et al.*, 2018) . Combining deep learning techniques with magnetic resonance imaging (MRI), we aim to extract the features that better define the survival model curve to predict time-to a first PIRA event. To do that, we will use the baseline MRI from 383 MS patients (58 with a PIRA event). From them, we will explore different convolutional neural networks (CNNs) strategies, such as dimensionality (3D, 2D and 2.5D), transfer learning or addition of clinical data (e.g., age), as well as the time intervals for the conditional survival probabilities (Vale-Silva and Rohr, 2021). Afterwards , our model performance will be compared with traditional statistical models obtained on the same dataset (Tur *et al.*, 2022) .

References

Montalban,X. *et al.* (2018) ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. *Multiple Sclerosis Journal*, **24**, 96–120.

Tur,C. *et al.* (2022) Association of Early Progression Independent of Relapse Activity With Long-term Disability After a First Demyelinating Event in Multiple Sclerosis. *JAMA Neurol*.

Vale-Silva,L.A. and Rohr,K. (2021) Long-term cancer survival prediction using multimodal deep learning. *Sci Rep*, **11**, 1–12.

8. POSTER S4: Fast inference of selection coefficients from genealogies

Aina Colomer-Vilaplana[1], Sònia Casillas[1], Leo Speidel[2,3]

1 Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, Barcelona
2 Genetics Institute, University College London, London, UK,
3 Francis Crick Institute, London, UK

Abstract

Our species evolved to survive in most environments encountered while spreading through Africa and later in the departure for the rest of the planet. A fundamental question left to resolve is to what extent selection has played a role in shaping our genomes to achieve so. Particularly over the past tens of thousands of years as a response to the constraints faced during the spread, such as different climates, inbreeding with other archaic species, exposure to new pathogens, and changes in diet.

Historically different sets of metrics have been used to identify genomic regions that have undergone selection. However, these metrics are limited by the different types of confounders like structure, admixture, background selection, recombination, or the high amount of weak selection present in the genome due to traits being highly polygenic.

With Relate[1], it has recently become possible to reconstruct genealogies from the variation data of thousands of modern-day people. This method describes how individuals are genetically related back in time, providing a robust framework for how genetic variation has evolved.

Here we propose a new statistic for fast inference of selection coefficients from genealogies data based on allele trajectory inferences. We benchmark this method using discoal [3] and SLiM [3] simulations looking at different modes of selection (including standing variation) and investigating if we can infer the timing of selection onset and strength. Finally, we plan to apply these to infer selection on 1000G data and compare their power and performance across different approaches, including genealogy-based and more classical approaches.

References

Speidel,L. *et al.* (2019) A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* **51**, 1321–1329.

Kern,A.D. and Schrider,D.R. (2016) Discoal: flexible coalescent simulations with selection, *Bioinformatics* **32**, 3839–3841.

Haller,B.C. and Messer,P.W. (2019) SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Molecular biology and evolution*, **36**, 632-637.

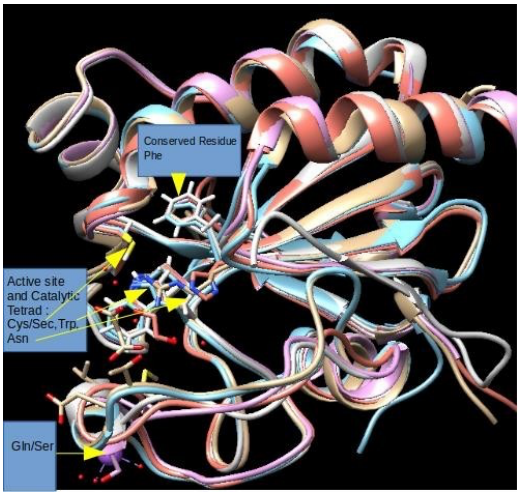
9. POSTER S1: Understanding the peroxidase activity of Mammalian Redox Regulating Glutathione Peroxidase using QM and QM/MM calculations

Nayanika Das, Jordi Villà-Freixa

Research Group on Bioinformatics and Bioimaging (BI2); Facultat de Ciències, Tecnologia i Enginyeries; Universitat de Vic - Universitat Central de Catalunya

Abstract

The biological effects of selenium are largely mediated by selenium-containing proteins (selenoproteins) [1]. Among them, different isoforms of Glutathione Peroxidase (GPX) have been labeled as potential targets to control or cause the oxidative stress during cancer evolution. [2] [3] In particular, eight different cysteine and selenocysteine containing isoforms of Glutathione Peroxidase (GPX1-8) isoforms have been identified in humans.[4] Mammalian GPX1, GPX2, GPX3, and GPX4 have shown to be selenium containing enzymes, whereas GPX6 is a selenoprotein in humans with cysteine containing homologues in rodents. GPX5, GPX7 and GPX8 contain cysteine. [4] It is known that glutathione peroxidase (GPX) oxidizes thiols to disulfides with an active site that contains a Sec/Cys residue [5].There exists strong structural homology among the different isoforms (see figure for details). The relationships between the mechanism and the structure is not very well known, which has led us to use a combination of studies at an atomic level by QM/MM methods and structural bioinformatics tools to unravel sequence-structure-function relationships. Our calculations target the understanding of the overall chemical mechanism of GPX enzymatic action within the context of the different homologs in humans, as well as the effect of Sec or Cys presence in the active site. We have used the GAMESS program gas phase and implicit solvent QM calculations to explore the non-enzymatic mechanism of the peroxidase activity in the first step of the reaction. Preliminary hybrid QM/MM calculations of the enzymatic mechanism with pDynamo are also shown.



References

Labunskyy, V. M., Hatfield, D. L., & Gladyshev, V. N. (2014). Selenoproteins: Molecular pathways and physiological roles. In Physiological Reviews (Vol. 94, Issue 3, pp. 739–777). American Physiological Society. https://doi.org/10.1152/physrev.00039.2013

Lubos E, Loscalzo J, Handy DE. Glutathione peroxidase-1 in health and disease: from molecular mechanisms to therapeutic opportunities. Antioxid Redox Signal. 2011 Oct 1;15(7):1957-97. doi: 10.1089/ars.2010.3586. Epub 2011 Apr10. PMID: 21087145; PMCID: PMC3159114.

Liu, H., Forouhar, F., Seibt, T. et al. Characterization of a patient-derived variant of GPX4 for precision therapy. NatChem Biol 18, 91–100 (2022). https://doi.org/10.1038/s41589-021-00915-2

Scheerer, P., Borchert, A., Krauss, N., Wessner, H., Gerth, C., Höhne, W., & Kuhn, H. (2007). Structural basis for catalytic activity and enzyme polymerization of phospholipid hydroperoxide glutathione peroxidase-4 (GPX4). https://doi.org/10.1021/bi700840d

Nauser, T., Steinmann, D., & Koppenol, W. H. (2012). Why do proteins use selenocysteine instead of cysteine? Amino Acids, 42(1), 39–44. https://doi.org/10.1007/s00726-010-0602-7

10. POSTER S2: Longitudinal Segmentation Of Multiple Sclerosis Lesions Using 3d U-Net And Attention

M.Díaz-Hurtado[1] J.Casas Roma[1], F.Prados Carrasco[3,4,5]

[1] e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain
[3] Centre for Medical Image Computing, University College London, London, United Kingdom
[4] National Institute for Health Research Biomedical Research Centre at UCL and UCLH, London, United Kingdom
[5] Queen Square MS Centre, Department of Neuroinflammation, UCL Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom

Abstract

Multiple sclerosis (MS) is the most common non-traumatic cause of neurological disability in young people. Magnetic resonance imaging is very important in diagnosis, follow-up of severity, measurement of disease burden and a biomarker of response to treatment. Measuring lesions changing in number and volume along time is cumbersome, time-consuming and prone to errors with poor inter-rater reliability. Many methods of automatic measurement have been developed but with variable success. One of the most innovative techniques currently is called U-Net(Ronneberger et al., 2015), a deep learning architecture based on convolutional neural networks (CNN). Use of U-Nets in MS MRI image segmentation is still scarce and under development. My ongoing thesis studies its implementation adding innovative attention mechanisms and a longitudinal approach by using baseline image information to predict follow-up image segmentation.

After having developed a literature review on longitudinal image segmentation (Diaz-Hurtado et al., 2022) we have implemented a U-Net with 3D CNN, inception convolutions and attention gates based on previous models developed for pancreas image segmentation by Schlemper and Oktay (Schlemper et al., 2018, 2019) Currently we are tuning hyperparameters and training the model on a patch based training system. Model uses PyTorch and library Torchio for preprocessing, patch sampling and reconstruction. For model training Keras/Tensorflow is used. We are also exploring adding new features to the architecture and increasing calculation power with professional GPU architectures. In a future stage we aim to tune up the model in other datasets and put it to the test in a clinical environment in collaboration with MS clinical researchers.

References

Diaz-Hurtado,M. et al. (2022) Recent advances in the longitudinal segmentation of multiple sclerosis lesions on magnetic resonance imaging: a review. Neuroradiology, 64, 2103–2117.

Ronneberger,O. et al. (2015) U-net: Convolutional networks for biomedical image segmentation. In, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 234–241.

Schlemper,J. et al. (2019) Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis, 53, 197–207.

Schlemper,M.B. et al. (2018) Cultivating Student Citizens. Spatial Citizenship Education, 88–116.

11. POSTER S3: Expanding the Knowledge of the Coagulation Contact Pathway Genomic Regulation

Laia Diez, on behalf of the CHARGE Consortium Hemostasis Working Group

1 Genomics of Complex Disease Unit, Sant Pau Biomedical Research Institute. IIB-Sant Pau, Barcelona, Spain.

Abstract

Background: Coagulation is an important process that reduces bleeding when there is an injure. Coagulation factors are the blood's proteins that help form blood clots to stop bleeding when a damage occurs. Coagulation factors IX (FIX), XI (FXI) and XII (FXII) are key components of the contact pathway of coagulation that participate in the propagation and stabilization of a thrombus.

Previous studies have suggested FIX and FXI to be related with the increased risk of having a thrombotic event (Van Hylckama Vlieg *et al.*, 2000; Gill *et al.*, 2018; Bertina, 2003; Yang *et al.*, 2006; Yuan *et al.*, 2021; Cushman *et al.*, 2009; Meijers *et al.*, 2000). Additionally, some studies related to activated partial thromboplastin time (aPTT), a test that characterizes coagulation of the blood, suggest that a short aPTT is a stable predictor of thrombosis risk (Tripodi *et al.*, 2004).

Genome-wide association studies (GWAS) are used to identify genomic variants that are statistically associated with a risk of a disease or a particular trait. Previous GWAS have been performed for FIX, FXI and aPTT using European participants. FXI GWAS (N= 16,567) revealed associations at *F11, KNG1*, and GCKR loci while aPTT GWAS (Houlihan *et al.*, 2010) (N=9,240) identified associations at *F11, F12, KNG1, ABO, HRG, F5* and *AGBL1* loci, accounting for 29% of the variance in aPTT (Tang *et al.*, 2012). However, no GWAS have been published for FIX and FXII levels before.

Aims and Methods: We aim to perform GWAS of the coagulation factors FIX, FXI, FXII and aPPT in larger sample sizes, and including non-European participants and sex-specific analyses. We will also perform a multi-phenotype analysis of these coagulation factors. Finally, we will conduct mendelian randomization analyses between FIX, FXI, FXII and the most frequent cardiovascular outcomes, such as venous thromboembolism, ischemic stroke peripheral artery disease or coronary artery disease, to elucidate the causal role of these proteins on cardiovascular disease.

Expected Results: We expect to unmask novel genetic factors that regulate the risk of thrombosis as well as elucidate the causal effect of these proteins on cardiovascular disease.

Overall, we anticipate that our work will identify additional loci associated to these factors, therefore helping elucidate the genetic regulatory mechanisms of the intrinsic coagulation pathway. This can shed light on identifying additional thrombosis mechanisms.

References

Bertina,R.M. (2003) Elevated clotting factor levels and venous thrombosis. *Pathophysiol. Haemost. Thromb.*, **33**, 395–400.

Cushman,M. *et al.* (2009) Coagulation factors IX through XIII and the risk of future venous thrombosis: the Longitudinal Investigation of Thromboembolism Etiology. *Blood*, **114**, 2878–2883.

Gill,D. *et al.* (2018) Genetically Determined FXI (Factor XI) Levels and Risk of Stroke. *Stroke*, **49**, 2761–2763.

Houlihan,L.M. *et al.* (2010) Common variants of large effect in F12, KNG1, and HRG are associated with activated partial thromboplastin time. *Am. J. Hum. Genet.*, **86**, 626–631.

Van Hylckama Vlieg,A. *et al.* (2000) High levels of factor IX increase the risk of venous thrombosis. *Blood*, **95**, 3678–3682.

Meijers,J.C.M. *et al.* (2000) High levels of coagulation factor XI as a risk factor for venous thrombosis. *N. Engl. J. Med.*, **342**, 696–701.

Tang,W. *et al.* (2012) Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am. J. Hum. Genet.*, **91**, 152–162.

Tripodi,A. *et al.* (2004) A shortened activated partial thromboplastin time is associated with the risk of venous thromboembolism. *Blood*, **104**, 3631–3634.

Yang,D.T. *et al.* (2006) Elevated factor XI activity levels are associated with an increased odds ratio for cerebrovascular events. *Am. J. Clin. Pathol.*, **126**, 411–415.

Yuan,S. *et al.* (2021) Genetically Proxied Inhibition of Coagulation Factors and Risk of Cardiovascular Disease: A Mendelian Randomization Study. *J. Am. Heart Assoc.*, **10**.

12. POSTER S4: Optimizing computation through parallelization: Advances in differential expression analysis

Xavier Escribà Montagut[1,2], Juan R. Gonzalez[1,2]

1 Universitat Autónoma de Barcelona, Bellaterra, Spain
2 Institut de Salut Global de Barcelona (ISGlobal), Barcelona, Spain

Abstract

As researchers in the field of bioinformatics, it is essential to understand the advances in computing, including improvements in core counts, memory bandwidth, and other quantitative measures (Shalf, 2020). These advances have significant implications for our field and can be leveraged to our advantage. In the rapidly evolving world of computing, it is important to keep up with the latest developments in order to effectively utilize the latest technologies.

Particularly, parallel computing (Asanovic *et al.*, 2009), which involves the use of multiple processors or computers to perform tasks simultaneously, resulting in faster completion times and increased efficiency. Parallel computing is particularly useful for tasks that can be broken down into smaller, independent parts, such as scientific simulations, data analysis, and machine learning. While the technical terms used to describe these systems can be intimidating to those unfamiliar with them, the benefits of parallel computing are undeniable and have made it an important tool in a variety of fields.

The author demonstrates the practical use of parallel computing by analyzing an open-source algorithm (differential expression analysis using the BioConductor *limma* package (Ritchie *et al.*, 2015)) and identifying opportunities to parallelize it using R language tools. The implementation of these changes results in a significant reduction in task completion time. More precisely, a reduction of 30+ minutes compute time to 3 minutes when running on an advanced research computer facility.

In conclusion, parallel computing is a valuable tool for researchers in the field of bioinformatics and it is important for researchers to understand how to effectively utilize these technologies to their advantage.

References

Asanovic,K. *et al.* (2009) A view of the parallel computing landscape. *Commun ACM*, **52**, 56–67.

Ritchie,M.E. *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, e47–e47.

Shalf,J. (2020) The future of computing beyond Moores Law. *Philosophical Transactions of the Royal Society A*, **378**.

13. POSTER S1: Stem cell like states suppress use of alternative end joining DNA damage repair

Roderic Espín[1], Sònia Farran[1], Francesca Mateo[1], Lidia Franco[1], Adrián Martínez[1], Xieng C Wang[1], Alexandra Baiges[1], Rana El Bizri[1], Ines Guix[2], Nadia Garcia[1], Luis Palomero[1], Mary Helen Barcellos-Hoff[2], Alvaro Aytes[1] & Miquel Angel Pujana[1]

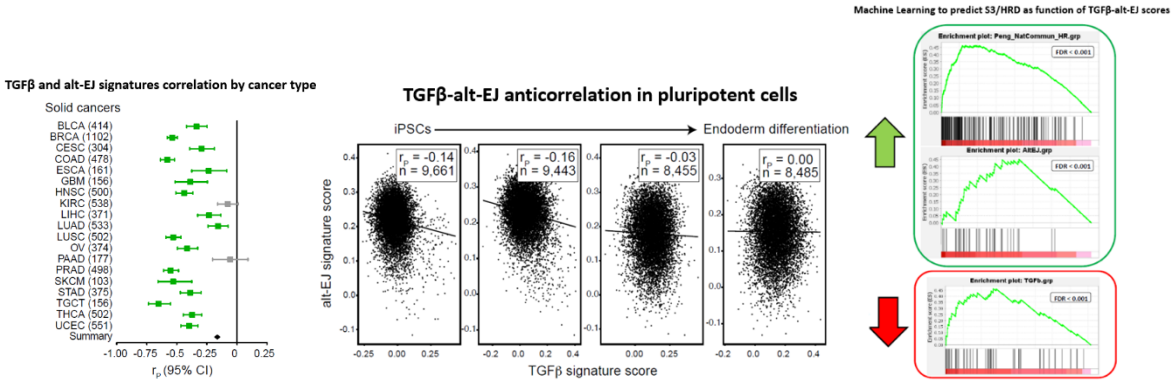
1 ProCURE, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain
2 Department of Radiation Oncology and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, United States

Abstract

Precise maintenance of genetic information is fundamental for organismal life and viability of progeny. Failure to repair DNA double-strand breaks (DSBs) can lead to cell death or mutations causing carcinogenesis, among other diseases(Jackson and Bartek, 2009). DSBs are repaired by two major pathways: homology-directed repair (HDR) and canonical non-homologous end joining (NHEJ)(Scully *et al.*, 2019). When these pathways are compromised,

a microhomology-mediated alternative end joining (alt-EJ) pathway is activated(Ahrabi *et al.*, 2016). Alt-EJ is frequently used in cancer cells linked to HDR-deficiency(Alexandrov *et al.*, 2013). However, the context and mechanisms by which normal and/or cancer cells chose either pathway remain poorly understood.

Functionally-validated gene expression signatures of transforming growth factor β (TGFβ) and alt-EJ signaling are found to be anticorrelated in most cancer types(Liu *et al.*, 2021). The signatures anti-correlation is linked to a cancer cell stem status and, in turn, is characteristic of normal cells or tissue with embryonic-undifferentiated or pluripotent features. The TGFβ-alt-EJ anticorrelation is confirmed in the analysis of single-cell transcriptome profiles.



References

Ahrabi,S. *et al.* (2016) A role for human homologous recombination factors in suppressing microhomology-mediated end joining. *Nucleic Acids Res.*, **44**, 5743–5757.

Alexandrov,L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.

Jackson,S.P. and Bartek,J. (2009) The DNA-damage response in human biology and disease. *Nature*, **461**, 1071–1078.

Liu,Q. *et al.* (2021) Loss of TGFβ signaling increases alternative end-joining DNA repair that sensitizes to genotoxic therapies across cancer types. *Sci. Transl. Med.*, **13**.

Scully,R. *et al.* (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.*, **20**, 698–714.

14. POSTER S2: Targeting epigenetic synthetic lethality for cancer treatment

Maria Farina-Morillas[1,2], Jose A. Seoane[1]

[1] Cancer Computational Biology Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona
[2] Universitat Autònoma de Barcelona (UAB), Barcelona

Abstract

Synthetic lethality (SL) is a specific interaction between genes, where the perturbation of either gene individually does not alter cell viability however the perturbation of both genes simultaneously leads to loss of viability (1). This allows for selective cancer therapies design, as inhibiting a synthetic lethal gene partner to a specifically altered gene in tumour cells will promote cell death just in cancer cells and spare the normal ones (2,3). Furthermore, SL can be expanded to epigenetic SL, as epigenetic alterations can lead to chromatin deregulation and oncogenesis beyond the effect of somatic mutations, and affect very different transcriptional programs, thus highlighting the relevance of epigenetic SL and its potential therapeutic value (4).

Computational methods are being used to predict SL partners to overcome the experimental screening methods limitations, and of especial relevance are Machine Learning algorithms (5). We are expanding the current algorithms to identify epigenetic synthetic lethal gene pairs using gene expression data and later integrating other multi-omics data, to determine novel biomarkers and therapy approaches. We are analysing molecular cancer cohorts and we will establish relationships between genes based on chromatin accessibility, exploring both cis- and trans-regulation by different Chromatin Regulatory Genes.

References

1. O'Neil, N. J *et al*. (2017). Synthetic lethality and cancer. Nature Reviews Genetics 2017 18:10, 18(10), 613–623.

2. Huang, A. *et al*. (2019). Synthetic lethality as an engine for cancer drug target discovery. Nature Reviews Drug Discovery 2019 19:1, 19(1), 23–38.

3. Magen, A. *et al*. (2019). Beyond Synthetic Lethality: Charting the Landscape of Pairwise Gene Expression States Associated with Survival in Cancer. Cell Reports, 28(4), 938-948.e6.

4. Yang, H. *et al*. (2019). Epigenetic synthetic lethality approaches in cancer therapy. Clinical Epigenetics, 11(1), 1–7.

5. Wang, J. *et al*. (2022). Computational methods, databases and tools for synthetic lethality prediction. Briefings in Bioinformatics, 23(3), 1–22.

15. POSTER S3: goSorensen: an R package and app to compare feature lists based on the Sorensen-Dice index and enrichment of GO terms

Pablo Flores[1], Jordi Ocaña[2], Álex Sánchez[2], Miquel Salicrú[2].

[1] Department of Statistics and Operational Research, Faculty of Mathematics and Statistics, Universitat Politècnica de Catalunya, Barcelona, Spain
[2] Department of Genetics, Microbiology and Statistics, Statistics Section., Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

Abstract

Given n GO terms (Consortium, 2004) for a specific GO ontology and GO level, the number of enriched and non-enriched terms in two feature lists $L1,L2$ can be summarized as a 2×2 contingency table $nij,\ i,j=1,0$. The expression $dS\hat{}= 1-2n11 / (2n11+n10+n01)$ based on Sorensen-Dice index (Sørensen, 1948) is a good measure of enrichment degree coincidence between $L1$ and $L2$ which are taken as similar if they share a great proportion of common enriched GO terms. If $dS\hat{}$ approaches zero, it means a positive dependence of the enrichment degree of both lists and it implies biological similarity. Using the delta method (Doob, 1935) , it is possible to show that $dS\hat{}$ is asymptotical normal, however for poor levels of enrichment, bootstrap is a more conservative but preferable approach, with better type I error control. Rejecting the null hypotheses in $H0: dS \geq d0$ Vs $H1: dS < d0$ can be taken as evidence of equivalence (Wellek, 2010) between $L1$ and $L2$, up to a previously given irrelevance limit $d0$.

This work presents computer tools allowing to apply this theory above explained (Flores *et al.*, 2022) . The **goSorensen R-Package**[1] available in Bioconductor is built using the R programming language (R Core Team, 2021) under a S3 object-oriented approach. Functions in this package recognize two types of data inputs: i) Feature lists containing the ENTREZ identifiers and ii) enrichment frequencies. When more than two feature lists are compared, the Bonferroni-Holm criterion (Holm, 1979) is applied. The **goSorensen app**[2] is implemented using the R-Shiny framework (Chang *et al.*, 2022) . All computations in this app can be obtained directly using the R-package, but this app is intended to be a more user-friendly tool for users who are not familiar with R code. App shows a graph allowing to decide which of the approximations (normal or bootstrap) provides more reliable results.

References

Chang,W. et al. (2022) shiny: Web Application Framework for R.

Consortium,G.O. (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res, 32, D258–D261.

Doob,J.L. (1935) The limiting distributions of certain statistics. The Annals of Mathematical Statistics, 6, 160–169.

Flores,P. et al. (2022) An equivalence test between features lists, based on the Sorensen–Dice index and the joint frequencies of GO term enrichment. BMC Bioinformatics, 23, 207.

Holm,S. (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 65–70.

R Core Team (2021) R: A Language and Environment for Statistical Computing.

Sørensen,T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons.

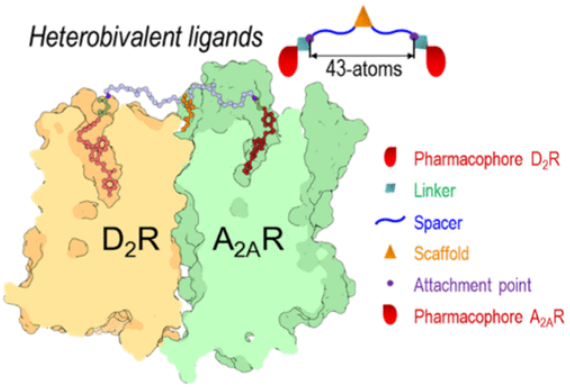
Wellek,S. (2010) Testing statistical hypotheses of equivalence and noninferiority Calver,R. (ed) Chapman and Hall/CRC.

16. POSTER S4: Heterobivalent Ligand for the Adenosine A_{2A} – Dopamine D_2 Receptor Heteromer

Daniel Pulido,[#1,2] Verónica Casadó-Anguera,[#3] Marc Gómez-Autet,[#4] Natàlia Llopart,[3] Estefanía Moreno,[3] Nil Casajuana-Martin,[4] Sergi Ferrer,[5] Leonardo Pardo,[*4] Vicent Casadó,[*3] and Miriam Royo[*1,2]

[1] Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), 08034 Barcelona, Spain
[2] Department of Surfactants and Nanobiotechnology, Institute for Advanced Chemistry of Catalonia (IQAC-CSIC), 08034 Barcelona, Spain
[3] Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Institute of Biomedicine of the University of Barcelona (IBUB), University of Barcelona, 08028 Barcelona, Spain
[4] Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[5] Integrative Neurobiology Section, National Institute on Drug Abuse, Intramural Research Program, National Institutes of Health, Baltimore, Maryland 21224, United States

Abstract



Bitopic and bivalent ligands are single chemical entities composed of two pharmacophore units covalently linked by an appropriate linker. In particular, bivalent ligands for G protein-coupled receptors (GPCRs) have emerged to interrogate receptor dimer function by simultaneously binding both orthosteric sites of the (homo/hetero) dimer (Hiller *et al.*, 2013) . Homobivalent ligands contain two copies of the same pharmacophore, whereas heterobivalent ligands link two different pharmacophores.

A G protein-coupled receptor heteromer that fulfills the established criteria for its existence *in vivo* is the complex between adenosine A_{2A} ($A_{2A}R$) and dopamine D_2 (D_2R) receptors. These heteromers have been found in transfected cells (Canals *et al.*, 2003) , in native tissue (Trifilieff *et al.*, 2011) , and in human postmortem brains in which the amount of heteromer differs from healthy and diseased, increasing in Parkinson's disease (Fernández-Dueñas *et al.*, 2019) and decreasing in schizophrenia (Valle-León *et al.*, 2021) . Here, we have designed and synthesized heterobivalent ligands for the $A_{2A}R$-D_2R heteromer. A key factor for the design of these type of compounds is the spacer length of the bivalent ligand, which depends on the dimer interface, the surface of the extracellular domains of the dimer, and the position and the orientation of the attachment points of the pharmacophore-linker moieties (Pérez-Benito *et al.*, 2018) . The interface of the $A_{2A}R$-D_2R heteromer was previously determined using biomolecular fluorescence complementation (BiFC) experiments where disruption of fluorescence complementation has been only significantly observed in the presence of peptides with the amino acid sequence of TM4 and TM5 of both $A_{A2A}R$ and D_2R, suggesting a TM4/5 interface for the $A_{2A}R$-D_2R heteromer (Navarro *et al.*, 2018) . This information has been used to computationally model the heteromer and, after docking the two pharmacophore units into the orthosteric binding cavities of each of the two receptors, calculate the preferred spacer length by adjusting different spacer moieties to the van der Waals surface of the heteromer.

The indispensable simultaneous binding of these ligands to the two different orthosteric sites of the heteromer has been evaluated by radioligand competition-binding assays in the absence

and presence of specific peptides that disrupt the formation of the heteromer, label-free dynamic mass redistribution assays in living cells, and molecular dynamic simulations. This combination of techniques has permitted us to identify a compound with a spacer length of 43-atoms, as a true bivalent ligand that simultaneously binds to the two different orthosteric sites. Moreover, bioluminescence resonance energy transfer experiments indicate that this compound also favors the stabilization of the $A_{2A}R$-D_2R heteromer.

References

Canals,M. *et al.* (2003) Adenosine A2A-dopamine D2 receptor-receptor heteromerization: qualitative and quantitative assessment by fluorescence and bioluminescence energy transfer. *J Biol Chem*, **278**, 46741–46749.

Fernández-Dueñas,V. *et al.* (2019) Revealing Adenosine A2A-Dopamine D2 Receptor Heteromers in Parkinson's Disease Post-Mortem Brain through a New AlphaScreen-Based Assay. *Int J Mol Sci*, **20**.

Hiller,C. *et al.* (2013) Class A G-protein-coupled receptor (GPCR) dimers and bivalent ligands. *J Med Chem*, **56**, 6542–6559.

Navarro,G. *et al.* (2018) Evidence for functional pre-coupled complexes of receptor heteromers and adenylyl cyclase. *Nat Commun*, **9**, 1–12.

Pérez-Benito,L. *et al.* (2018) The size matters? A computational tool to design bivalent ligands. *Bioinformatics*, **34**, 3857–3863.

Trifilieff,P. *et al.* (2011) Detection of antigen interactions ex vivo by proximity ligation assay: endogenous dopamine D2-adenosine A2A receptor complexes in the striatum. *Biotechniques*, **51**, 111–118.

Valle-León,M. *et al.* (2021) Decreased striatal adenosine A2A-dopamine D2 receptor heteromerization in schizophrenia. *Neuropsychopharmacology*, **46**, 665–672.

17. POSTER S1: The landscape of gastrointestinal stromal tumour(GIST) progression uncovers a mutually-exclusive chromosomal instability (CIN)-dependent and CIN-independent tumour evolution

David Gómez-Peregrina1, Alfonso García-Valverde1, Sebastian Bauer2, ChiaraColombo3, Piotr Rutkowski4, Armelle Dufresne5, Patrick Schöffski6, Kjetil Boye7,Pablo Marín-García8, Marta Gut9,10, Anna Esteve9,10, Genís Parra9,10, ClaudiaValverde11, Iván Olivares-Rivas1, Daniel Pilco-Janeta1, Jordi Rosell1, GemmaMur1,12, César Serrano1,11

1. Sarcoma Translational Research Laboratory, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain.
2. Department of Medical Oncology, Sarcoma Center, West German Cancer Center, University Duisburg-Essen, Medical School, Essen, Germany.
3. Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.
4. Department of Soft Tissue-Bone Sarcoma and Melanoma, Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland.
5. Medical Oncology Department, Centre Léon Bérard, Lyon, France.
6. Department of General Medicine Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium.
7. Department of Oncology, Oslo University Hospital, The Norwegian Radium Hospital, Oslo, Norway.
8. Zetta Genomics, Cambridge, UK.
9. CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.
10. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
11. Department of Medical Oncology, Vall d'Hebron University Hospital, Barcelona, Spain.
12. Clinical trials office, Vall d'Hebron Institute Oncology (VHIO), Barcelona, Spain.

Keywords: Sarcoma, Chromosome instability, TKI resistance.

Abstract

GISTs are malignant mesenchymal neoplasms arising from the interstitial cells of Cajal. They are generally classified as genomically simple sarcomas, bearing a low number of karyotype abnormalities (Heinrich *et al.*, 2002). The most common initiating event is the oncogenic activation of KIT (80%) or PDGFRA (10%) tyrosine kinase receptors by gain-of-function mutations, remaining present throughout the course of the disease (Hirota *et al.*, 1998; Corless *et al.*, 2011). Less commonly, GISTs have KIT/PDGFRA-independent drivers (known as GIST WT) such as loss-of-function alterations in NF1 and SDH genes (Corless *et al.*, 2011). GIST therapeutic strategies are based on tyrosine kinase inhibitors (TKIs) with KIT/PDGFRA inhibitory activity (Serrano *et al.*, 2017). TKI resistance results from the polyclonal expansion of KIT/PDGFRA secondary mutations in most patients (Liegl *et al.*, 2008; Serrano and Fletcher, 2019). Given the low frequency of this neoplasm (10-15 cases/million/year), little is known about other aspects of GIST biology and tumour progression.

In order to unravel the biological processes underlying GIST evolution, a unique collection of GIST samples with clinical information (n=68) was gathered from European sarcoma-expert institutions and sequenced by WES and RNA-seq. This collection aims to recapitulate the evolutionary history of GIST, from pre-treated localised tumours to multi-TKI refractory metastases.

Low tumour mutational burden (4.06 muts/Mb) was consistent with prior sarcoma data. KIT/PDGFRA secondary mutations were the predominant mechanism of TKI resistance and only n=6 cases displayed oncogenic activation of KIT-downstream signalling intermediates.

Unexpectedly, >70% of the cohort showed high chromosome instability (CIN) properties with at least one whole-genome doubling (WGD) event, when assessing the fractions of the exome affected by copy number alterations (CNA), LOH fractions and tumour ploidies. NGS data from patients identified alterations in cell cycle processes and checkpoints, increased proliferative properties, disrupted p53-network activity (with nearly no mutations in p53 or related genes) and impaired DNA damage sensing and response, ultimately providing a CIN-permissive context. No significant associations were found between CIN and previous TKI treatments nor with localized and metastatic disease.

Collectively, we show that CIN is an unprecedented novel driver mechanism of tumour evolution in a meaningful subset of GIST patients. Ongoing work attempts to define the driver mechanisms of this CIN-dependent fate in GIST evolution.

References

Corless,C.L. et al. (2011) Gastrointestinal stromal tumours: origin andmolecular oncology. Nat Rev Cancer, 11, 865–878.

Heinrich,M.C. et al. (2002) Biology and genetic aspects of gastrointestinalstromal tumors: KIT activation and cytogenetic alterations. Hum Pathol, 33,484–495.

Hirota,S. et al. (1998) Gain-of-function mutations of c-kit in humangastrointestinal stromal tumors. Science, 279, 577–580.

Liegl,B. et al. (2008) Heterogeneity of kinase inhibitor resistance mechanismsin GIST. The Journal of Pathology, 216, 64–74.

Serrano,C. et al. (2017) Novel Insights into the Treatment of Imatinib-ResistantGastrointestinal Stromal Tumors. Targeted Oncology, 12, 277–288.

Serrano,C. and Fletcher,J.A. (2019) Overcoming heterogenity in imatinib-resistant gastrointestinal stromal tumor. Oncotarget, 10, 6286–6287.

18. POSTER S2: *In-silico* simulation and identification of predictive biomarkers of first-line single-agent regorafenib on a metastatic colorectal cancer virtual population

Juan Manuel García-Illarramendi[1,2], Pedro Matos-Filipe[1,3], Guillem Jorba[1,3], Xavier Daura[2], Judith Farrés[1], José Manuel Mas[1]

[1]Anaxomics Biotech SL, Barcelona, Spain
[2]Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Barcelona, Spain
[3]Structural Bioinformatics Group, Research Programme on Biomedical Informatics, Department of Experimental and Health Science, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Metastatic colorectal cancer (met-CRC) remains as one of the most aggressive cancer subtypes in Western civilization, with a poor prognosis specially in elderly people. Despite the large number of approved treatments for colorectal cancer, there is still a lack of predictive biomarkers for met-CRC. In this study, an *in-silico* clinical trial of first-line single-agent regorafenib on elderly met-CRC patients was done, where Therapeutic Performance Mapping System (TPMS) was used to build patient models and simulate drug treatment.

Transcriptomic data from 408 elderly met-CRC patients were retrieved from Gene Expression Omnibus and used to build a cohort of virtual patients based on their individually differentially expressed proteins (IDE). Target-based simulation of regorafenib effect on a knowledge-based met-CRC protein set allowed defining *in-silico* good and bad responders. 213 mechanistical biomarkers were identified by comparing the regorafenib TPMS models of the good and bad responder patients (Welch's t-test, adjusted p-value by Benjamini-Hochberg < 0.05; absolute mean difference > 0.1). Similarly, 178 dysregulated proteins of the IDEs were found as potential biomarkers by comparing the IDEs of the good and bad responders (Two-sided Fisher's exact test, p-value < 0.05). 4 proteins; MARK3, HSF1, RBCK1 and LHCGR, were found as potential mechanistic and predictive biomarkers of regorafenib.

19. POSTER S3: Exploring brain features variation associated with a higher genetic predisposition to Alzheimer's disease in the Alzheimer's disease spectrum

Patricia Genius Serra[1], Malu Calle, Carles Falcón, Carolina Minguillón, Manel Esteller, Arcadi Navarro, Juan Domingo Gispert, Natalia Vilor-Tejedor

1.Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain

Abstract

Imaging genetics (IG) studies analyze how genetic information influences brain features by combining neuroimaging-based brain features and genetics data. Common multivariate methods do not take into account the joint modulation of brain features. The objective of this project consisted of exploring which cortical and subcortical brain regions vary together as a function of the genetic predisposition to Alzheimer's disease (AD) at different stages of the disease.

The sample of the study was defined by 351 (233 amyloid-beta negative (Aβ-) and 118 Aβ+) cognitively unimpaired (CU) middle-age participants from the ALzheimer's and FAmilies (ALFA) project and 330 individuals (230 mild cognitive impaired (MCI) subjects and 100 AD patients) from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Participants had available information on genetics, neuroimaging data and cerebrospinal fluid AD core biomarkers. Four groups were defined covering the AD disease-stage continuum. Genetic predisposition to AD (PRS-AD) was estimated by calculating a polygenic risk score. A binary variable was created (Low vs High risk of AD) taking as cut-off point the quantile 0.8. Compositional data analysis was performed using *coda4microbiome* [1]. The algorithm is based on two main steps. At first, variable selection is performed through an elastic net penalized regression procedure. The joint volumetric change of the selected variables is summarized in a log-contrast function (brain signature). In a second step, a logistic regression is defined to determine the association between the brain signatures with the PRS-AD. Models were stratified by disease stage. All the models were adjusted for age and sex.

Individuals at different stages of the AD continuum displayed different brain signatures associated with higher genetic predisposition to AD. Remarkably, hippocampal volume was positively associated with higher risk of AD in Aβ- individuals. On the contrary, it was negatively associated with an increased risk of AD in MCI Aβ+ [Figure 1].

Results suggested that after an initial increase in the volume of the hippocampus in healthy individuals at higher risk of AD, damage in this region might occur in later stages of the disease but before AD clinical diagnosis. This work provides an innovative modelling perspective for IG studies, providing an approximation that may be closer to the volumetric changes that individuals along the AD can display.

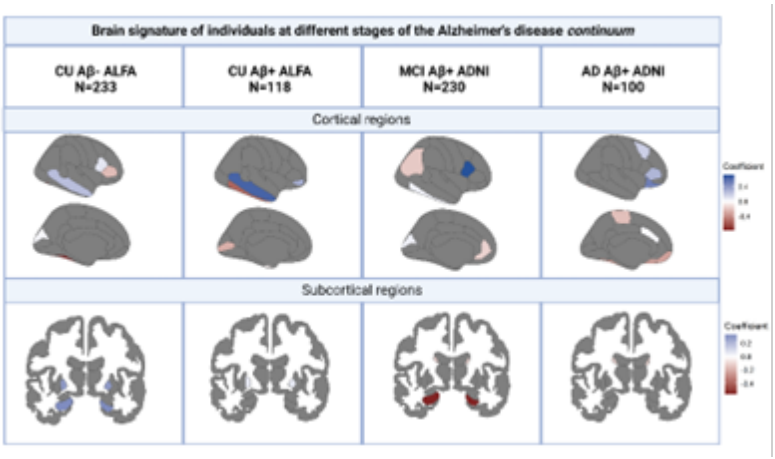


Figure 1. *Displaying the brain signature that is most closely associated with a higher genetic predisposition to AD. In blue, we show the brain regions whose volume is positively associated with a higher genetic predisposition to AD. In red, we show the brain regions whose volume is negatively associated with an increased risk of AD. Legend: Aβ- (amyloid-beta negative), Aβ+ (amyloid-beta positive), MCI Aβ+ (mild cognitive impaired amyloid-beta positive), AD Aβ+ (Alzheimer's disease amyloid-beta positive).*

References

Calle, M.L. and Susin, A. (2022) coda4microbiome: compositional data analysis for microbiome studies. *bioRxiv*

20. POSTER S4: Enhancing metabolome characterization for LC-MS and LC-MS/MS using HERMES

Giné, Roger[1,2], Capellades, Jordi[1,2], Vinaixa, Maria[1,2], Yanes, Oscar[1,2,3]

[1] Metabolomics Platform, Institut d'Investigació Sanitària Pere Virgili (IISPV)
[2] Metabolomics Interdisciplinary Laboratory (MIL@B). Universitat Rovira i Virgili (URV)
[3] Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM)

Abstract

Despite the large number of peak signals (10^4 to 10^5) detected in a single LC-MS metabolomics experiment, in practice, the number of consistently identified metabolites is still very limited (10^2 to 10^3) (Mahieu and Patti, 2017; Sindelar and Patti, 2020). This gap in identifications is mainly caused by the lack of structural information gathered in LC-MS/MS experiments: current LC-MS/MS data acquisition strategies fragment precursor ions based on either on a mass range or intensity criteria, while not taking into account the biological origin of the ions. Thus, most of the time the instrument is fragmenting redundant or non-biological signals (Duan *et al.*, 2016), hampering identification rates.

In a previous publication (Giné *et al.*, 2021), we presented a new strategy, HERMES, that greatly improves the characterization of metabolomics samples compared to current state-of-the-art methods. By querying the ions against a database of potential ionic formulas, HERMES generates a very sample-specific list of target ions for LC-MS/MS fragmentation, which in turn improves the number of identified metabolites in a sample by up to 2-3X.

The objective of my thesis project is to improve upon HERMES and current metabolite annotation tools to characterize the metabolome, extend the annotation process to multiple samples and create an expanded annotation framework for ion mobility data (LC-IM-MS). Lastly, all created software tools will be migrated to a cloud-based computational environment to improve their performance and usability for the general scientific community.

References

Duan,L. *et al.* (2016) Discrimination and Quantification of True Biological Signals in Metabolomics Analysis Based on Liquid Chromatography-Mass Spectrometry. *Molecular Plant*, 9, 1217–1220.

Giné,R. *et al.* (2021) HERMES: a molecular-formula-oriented method to target the metabolome. *Nat Methods*, 18, 1370–1376.

Mahieu,N.G. and Patti,G.J. (2017) Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal Chem*, 89, 10397–10406.

Sindelar,M. and Patti,G.J. (2020) Chemical Discovery in the Era of Metabolomics. *J Am Chem Soc*, 142, 9097–9105.

21. POSTER S1: Analyzing relationships among nutrition, health, and sustainability in Catalonia by data science models

Giner MP[1], Bach-Faig A[2]

[1] Doctorate Program in Bioinformatics (UAB, UB, UdG, UdL, UOC, UPC, URV, UVic-UCC).
[2] FoodLab Research Group (2017 SGR 83), Faculty of Health Sciences, Universitat Oberta de Catalunya (UOC), 08018, Barcelona, Spain

Abstract

Dietary habits have a direct impact not only on people's health, but also on the health of the planet [1]. The increasing demand of food for the growing global population, together with the environmental impact of the food system, make it necessary to employ strategies to promote health and sustainable diets. Therefore, it is important to analyze the different individual and contextual factors related to health and dietary habits to promote and reinforce a sustainable lifestyle that guarantees the well-being of both the population and for the planet [2,3,4]. One of the tools to retrieve such information is through surveys [5,6], which can be linked to the clinical data of the participants [7]. In addition, the use of current omics technologies, such as metabolomics, can be applied to link potential biomarkers, not just to health but also to sustainability patterns. For that, the use of computational strategies is key to handle and dig into these different sources of information. Different approaches can be applied to optimize diets with environmental constraints by mathematical programming and metaheuristics [8], or to build predictive models by machine learning to recognize patterns in favor of health and sustainable habits [9].

References

1. Bach-Faig, A. *et al* (2022) Consensus-building around the conceptualisation and implementation of sustainable healthy diets: a foundation for policymakers. BMC Public Health, **22**(1)
2. Duchin F. (2005) Sustainable consumption of food: A framework for analyzing scenarios about changes in diets. J Ind Ecol **9**, 99-114.
3. Tichenor B.N., *et al.* (2018) Linking sustainability to the healthy eating patterns of the Dietary Guidelines for Americans: a modelling study. The Lancet Planetary Health,**2**(8); e344-52.
4. European Public Health Association - EUPHA (2017) Healthy and Sustainable Diets for European Countries - annual report.
5. Sivertsen B. *et al* (2019) Cohort profile: the SHoT-study, a national health and well-being survey of Norwegian university students. BMJ Open, **9**(1).
6. Innstrand S. and Christensen M. (2020) Healthy Universities. The development and implementation of a holistic health promotion intervention programme especially adapted for staff working in the higher educational sector: the ARK study. Global Health Promotion, 68-76, **27**(1).
7. Burcin M. *et al.* (2019) Optimizing college health promotion in the digital age: Comparing perceived well-being, health behaviors, health education needs and preferences between college students enrolled in fully online versus campus-based programs. Health Promotion Perspective, 270-278, **9**(4).
8. Ileri Y. and Hacibeyoglu M. (2019) Advancing competitive position in healthcare: a hybrid metaheuristic nutrition decision support system. International Journal of Machine Learning and Cybernetics, 1385-1398, **10**(6).
9. Chaudhary A. and Krishna V. (2019) Country-specific sustainable diets using optimization algorithm. Environmental Science and Technology, 7694-7703, **53**(13).

22. POSTER S2: Uncovering the mechanisms of resistant B-ALL through integration of multi-omics data

Kathleen Imbach[1], Manel Esteller[1], Eduard Porta-Pardo[1]

1 Josep Carreras Leukaemia Research Institute (IJC), Badalona 08916, Spain

Abstract

As biological sequencing has become more accessible and computational capabilities have exponentially increased, so has the possibility to integrate multiple layers of biological data to better understand the molecular roots of disease. One example of such an effort has been undertaken by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), which has analyzed integrated proteogenomic data from patients spanning 10 different solid tumor types to unravel the nuances of cancer progression and therapeutic resistance[1]. In this study, we aim to apply similar proteogenomic analysis approaches to better understand resistant B-cell acute lymphoblastic leukemia (B-ALL). Although many patients diagnosed with B-ALL respond to initial therapeutic interventions, the population of individuals who experience relapse have much poorer prognosis and overall survival rates[2]. Studying the biological signatures unique to resistant B-ALL is essential to predict which patients might not respond to standard therapies and to develop novel approaches to target unresponsive tumors. In this study, we aim to integrate multi-omics data taken at diagnosis from around 100 Spanish B-ALL patients. Harmonizing relevant clinical information with individuals' genetic, epigenetic, transcriptomic, and proteomic profiles will permit molecular insights into disease manifestations and resistance in leukemia. Utilizing the capabilities of biologically informed machine learning approaches, as was recently implemented to assess prostate cancer[3], will allow us to discern patterns hidden in the multiple biological layers of B-ALL tumors. Mimicking known biological networks in the algorithm will permit interpretability of the results, unlike the opacity of traditional neural network algorithms[4]. Specifically, we hope to investigate the different mutation profiles in leukemia, understanding how somatic alterations and epigenetic signatures influence protein expression, and broadly associate clinical attributes with biological patterns. This undertaking will ultimately result in a rich resource to be used by other scientists working to understand leukemia and further the effort to develop more effective treatment regimens.

References

Rodriguez, H.*, et al.* The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. *Cell* 2021;184(7):1661-1670.

Hefazi, M. and Litzow, M.R. Recent advances in the biology and treatment of B-cell acute lymphoblastic leukemia. *Blood and Lymphatic Cancer: Targets and Therapy* 2018;8:47-61.

Elmarakeby, H.A.*, et al.* Biologically informed deep neural network for prostate cancer discovery. *Nature* 2021;598(7880):348-352.

Murdoch, W.J.*, et al.* Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci U S A* 2019;116(44):22071-22080.

23. POSTER S3: Spectra to molecule prediction using generative deep learning models

Muhammad Faizan Khan, Óscar Yanes, Roger Guimerà and Marta Sales-Pardo

Departament D'enginyeria electrònica Elèctrica i Automàtica, Universitat Rovira i Virgili

Abstract

Predicting the structure of a small molecule based on its tandem MS/MS spectrum is a challenging task in metabolomics. In this study, we are investigating state-of-the-art generative deep learning model to predict molecules from MS/MS spectra. Figure 1 presents the proposed workflow for predicting molecules from spectra. Here we use 4 metabolite databases: NIST20, MSDIAL, GNPS and Metlin (1,2), containing approximately 600k experimental spectra from over 42k compounds as ground truth to train and test generative deep learning models.

We divide our work in two blocks: (i) a Junction Tree Variational autoencoder (3) is used to generate valid molecules from SMILES; and (ii) deep learning models to predict molecules from spectra.

Junction Tree Variational autoencoder (JTVAE) is a tool that generates valid molecular graph representations. We trained a JTVAE using a metabolite dataset of approximately 390k SMILES and generated valid latent representation for each compound, which are then used as a representation to label MS/MS spectra for the second block of the workflow. Table 1 shows JTVAE results on original dataset (Zinc) and on our metabolite data.

Table 1 JTVAE RESULT

Model	SMILE	Reconstruction score (Tanimoto==1)
Original JTVAE	300K	58%
Retrained JTVAE	390K	68%

Currently, we are exploring different deep learning models such as convolutional neural networks, deep neural networks, LSTM and RNN for predicting molecules from MS/MS spectra.

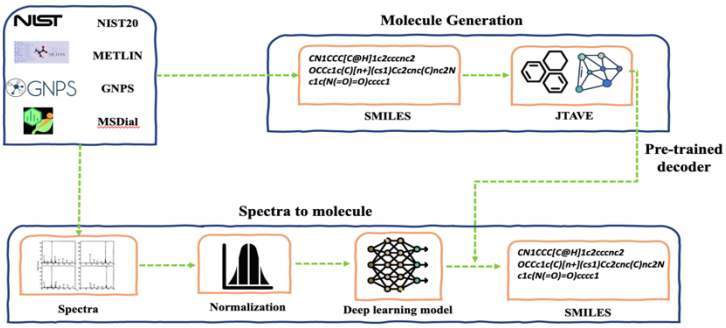


Figure 1 Proposed Workflow

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945413 and from the Universitat Rovira i Virgili (URV).

References

1. Guijas C, *et al.* (2018) METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal Chem, 3156-3164.

2. Jin, *et al*. (2018) Junction tree variational autoencoder for molecular graph generation. *International conference on machine learning*, PMLR.

24. POSTER S4: Identification of potential autoepitopes triggering Diabetes Type I in the contest of Sars-Cov2 infenction

Filippo Guerri[1,2], Xavier Daura[2]

1 Anaxomics SL
2 Universidad Autonoma de Barcelona

Abstract

Most infectious disease, such as viruses, bacteria, and parasites can trigger autoimmunity disease producing antibodies that recognize endogenous protein of the host [1]. This work provides a bioinformatic pipeline to identify potential autoepitope in the host's proteome by means of sequence similarity with the infectious organism (BLAST) [2] and epitope prediction algorithm (CNNPEPPRED and Netmhciipan_4.1) [3,4]. In particular, we are investgating the relation between Sars-cov2 infections and the insurgence of Diabetes type[5]. Our results provides a set of endogenoues human protein that could be further evaluated as trigger for autoimmunity.

References

1. Getts,D.R. *et al.* (2013) Virus infection, antiviral immunity, and autoimmunity. *Immunological Reviews*, **255,**197-209

2. Altschul,S.F. *et al.* (1990) Basic local alignemnt search tool. J Mol Bio, **215**, 403-410.

3. Junet,V. and Daura,X. (2021) CNN-PepPred: an open-source tool to create convolutional NN models for the discovery of patterns in peptide sets--application to peptide–MHC class II binding prediction. *Bioinformatics*, **37-23**, 4567-4568.

4. Reynisson,B. *et al.* (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, **48**, 449-454

5. Boddu,S.K. *et al.* (2020) New Onset diabetes, type1 diabetes and COVID-19. *Diabetes Metabolic Syndrome,* **14**, 2211-2217

25. POSTER S1: Identification of epigenetic biomarkers for molecular subgrouping of ependymoma

Joshua Llano-Viles[1,2,3,4], Soledad Gómez-González[3,4], Cinzia Lavarino[3,4], Alexandre Perera-Lluna[1,2]

1 B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Spain
2 Networking Biomedical Research Centre in the subject area of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain
3 Institut de Recerca Pediàtrica Hospital Sant Joan de Dèu, Esplugues de Llobregat, Barcelona, Spain
4 Pediatric Cancer Center Barcelona (PCCB), Hospital Sant Joan de Déu, Barcelona, Spain.

Abstract

Ependymomas are rare tumors of the central nervous system (CNS) that may arise in the supratentorial region or posterior fossa of the brain, or in the spinal cord. Its incidence is slightly lower in children (3/1,000,000) than in adult men and women (5/1,000,000 and 4/1,000,000 respectively), with pediatric patients having the worst prognosis (Saleh et al., 2022).

The clinical prognosis of ependymomas varies depending on the clinical, histopathological and molecular features (Pajtler et al., 2015). In 2016 the World Health Organization (WHO) established a classification of ependymoma subtypes based on the combination of histopathologic and molecular parameters (Louis et al., 2016). However, several studies have shown that risk stratification using molecular subgroups obtained through methylation profiling is superior to histological grading. For this reason, the 2021 revision of the WHO classification of CNS tumors presents a significant change from the previous histomorphologic classification of ependymal tumors to a classification of ten types of ependymomas based on anatomic location and molecular features (Louis et al., 2021; Saleh et al., 2022). Figure 1 shows the different molecular subgroups grouped by anatomical location.

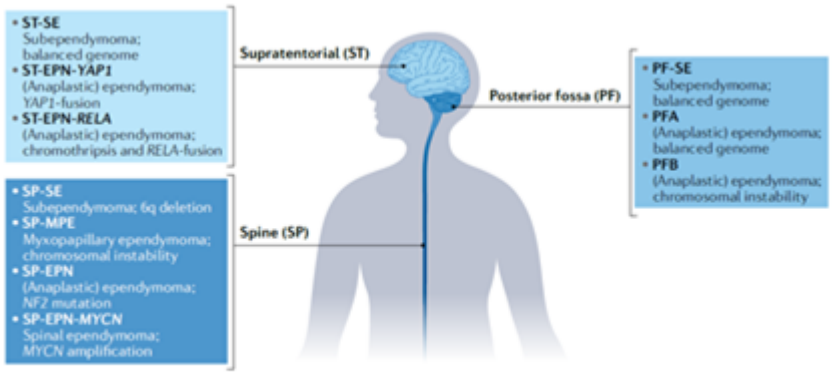


Fig 1. **Molecular subgroups of ependymoma grouped by anatomic location.**
Figure extracted from Saleh et al., 2022.

However, the application of array-based technology in a routine diagnostic setting can be time-consuming, costly and sometimes inaccessible for many centers worldwide treating patients with brain tumors. Consequently, a significant number of patients cannot benefit from the clinical advances associated with methylation-based ependymoma classification.

The aim of this study is to obtain epigenetic biomarkers from the methylation profiles of each molecular subgroup of ependymoma. In addition, one of the main characteristics of these biomarkers is that they should be implemented into clinical routine. For this purpose, we collected DNA methylation microarray data (Illumina Infinium HumanMethylation 450

BeadChip (HM450K) and Illumina methylation EPIC BeadChip array (EPIC)) from 8 different studies (n= 1748) (Mack et al., 2014; Pajtler et al., 2018; Fukuoka et al., 2018; Brabetz et al., 2018; Rogers et al., 2018; Capper et al., 2018; Cavalli et al., 2018; Michealraj et al., 2020). The proposed workflow is based on the previous work of our group (Gómez et al., 2018), first we will classify the samples based on the 2021 WHO classification. Next, we will preprocess the tumoral data and perform differential methylation analysis between the 10 ependymoma subgroups. In addition, we will compare the methylation patterns of ependymoma subtypes with methylation patterns of other pediatric tumors and healthy tissues in order to identify and filter DNA methylation biomarkers capable of exclusively discriminating ependymoma subtypes. Finally, we will reduce and select a group of biomarkers that meet the selection criteria to be included in the routine diagnostic.

References

1. Saleh, A. H. et al (2022). The biology of ependymomas and emerging novel therapies. *Nature Reviews Cancer*, *22*(4), 208–222. https://doi.org/10.1038/s41568-021-00433-2

2. Pajtler, K. W. et al (2015). Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell*, *27*(5), 728–743. https://doi.org/10.1016/j.ccell.2015.04.002

3. Louis, D. N. et al (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica*, *131*(6), 803–820. https://doi.org/10.1007/s00401-016-1545-1

4. Louis, D. N. et al (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology*, *23*(8), 1231–1251. https://doi.org/10.1093/neuonc/noab106

5. Mack, S. C. et al (2014). Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature 2014 506:7489*, *506*(7489), 445–450. https://doi.org/10.1038/nature13108

6. Pajtler, K. W. et al (2018). Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. *Acta Neuropathologica*, *136*(2), 211–226. https://doi.org/10.1007/S00401-018-1877-0/METRICS

7. Fukuoka, K. et al (2018). Significance of molecular classification of ependymomas: C11orf95-RELA fusion-negative supratentorial ependymomas are a heterogeneous group of tumors. *Acta Neuropathologica Communications*, *6*(1), 134. https://doi.org/10.1186/S40478-018-0630-1/FIGURES/5

8. Brabetz, S. et al (2018). A biobank of patient-derived pediatric brain tumor models. *Nature Medicine 2018 24:11*, *24*(11), 1752–1761. https://doi.org/10.1038/s41591-018-0207-3

9. Rogers, H. A. et al (2018). Limitations of current in vitro models for testing the clinical potential of epigenetic inhibitors for treatment of pediatric ependymoma. *Oncotarget*, *9*(92), 36530–36541. https://doi.org/10.18632/ONCOTARGET.26370

10. Capper, D. et al (2018). DNA methylation-based classification of central nervous system tumours. *Nature 2018 555:7697*, *555*(7697), 469–474. https://doi.org/10.1038/nature26000

11. Cavalli, F. M. G. et al (2018). Heterogeneity within the PF-EPN-B ependymoma subgroup. *Acta Neuropathologica*, *136*(2), 227–237. https://doi.org/10.1007/S00401-018-1888-X/METRICS

12. Michealraj, K. A. et al (2020). Metabolic Regulation of the Epigenome Drives Lethal Infantile Ependymoma. *Cell*, *181*(6), 1329-1345.e24. https://doi.org/10.1016/j.cell.2020.04.047

13. Gomez, S. et al (2018). A novel method for rapid molecular subgrouping of medulloblastoma. *Clinical Cancer Research*, *24*(6), 1355–1363. https://doi.org/10.1158/1078-0432.CCR-17-2243

26. POSTER S2: Allosterism in the adenosine A_{2A} and cannabinoid CB_2 heteromer

Claudia Llinas del Torrent[1], Iu Raïch[2,3,4], Angel González[1], Nil Casajuana-Martin[1], Rafael Franco[3], Leonardo Pardo[1], Gemma Navarro[2,3,4]

[1] Laboratory of Computational Medicine, Biostatistics Unit, Faculty of Medicine, Universitat Autònoma Barcelona, 08193 Bellaterra (Barcelona)
[2] Department of Biochemistry and Physiology, Faculty of Pharmacy and Food Sciences, Universitat de Barcelona, 08028 Barcelona
[3] Centro de Investigación en Red, Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, 28031 Madrid
[4] Institute of Neuroscience, University of Barcelona (NeuroUB), Av Joan XXIII 27-31, 08028, Barcelona, Spain.

Abstract

Allosterism is a common regulatory mechanism for G protein-coupled receptors (GPCRs) that can be attained by ligand-binding or protein-protein interactions with a different GPCR or other proteins. GPCRs can engage protein-protein allosteric interactions among them, by forming complexes constituted by the same (homo) or different (hetero) receptor protomers[1]. GPCRs are dynamic proteins that permit rapid small-scale structural fluctuations[2]. Thus, one of the protomers in the heteromer can influence the conformation of the interacting receptor and allosterically modulate its functional properties (affinity and/or efficacy and/or functional selectivity)[3].

Here, we have studied the influence of the dimer interface on the allosteric properties of the adenosine A2A (A2AR) and cannabinoid CB2 (CB2R) heteromer. Our results show that A2AR blocks CB2R induced signalling in a ligand-independent manner. Notably, this effect is reverted by the action of disruptive synthetic peptides or the action of a selective adenosine A2AR antagonist. We have named this new described phenomenon as cross-agonism. Interestingly, going deep inside in the analysis of the cross-agonism using computational and experimental results, a change in the heteromeric interface of the A2AR-CB2R heteromer induced by antagonists binding is suggested.

These novel results shed light on a different type of allosteric mechanism and extend the repertoire of GPCR heteromer signaling.

References

1. Ferre, S. *et al. (*2009) Building a new conceptual framework for receptor heteromers. *Nat. Chem. Biol.*, **5**, 131-4.

2. Nygaard, R. *et al*. (2013) The Dynamic Process of beta(2)-Adrenergic Receptor Activation. *Cell*, **152**, 532-42.

3. Rozenfeld, R. and Devi, L. A. (2010) Receptor heteromerization and drug discovery. *Trends Pharmacol. Sci*., **31**, 124-30.

Integrative multi-ome analysis of the CEBPa-induced B-cell to macrophage cell fate conversionAnna Vanessa Lopez Rubio, GemmaValcàrcel Ximenis, Jose Luis Sardina OrtegaAbstractCell fate decisions are typically initiated by the binding to DNA of sequence specific transcription factors (TFs) to regulate gene expression. This occurs in concert with epigenetic modifications, which include DNA and histone modifications that together create transcriptionally permissive or repressive chromatin environments (Apostolou and Hochedlinger, 2013). Recent studies have shown thatDNA modifications modulate TF binding to DNA (Yin et al., 2017)(Kaluscha et al. 2022 PMID:36471082) and influence the recruitment of chromatin remodeling complexes (Zhang et al., 2015). However, whether the dynamic interplay between DNA modifications and chromatin-associated proteins is a driving force of cell-fate decisions is unknown. This is largely due to the low proportion of cells that can be experimentally induced to change fate (Takahashi and Yamanaka, 2006).This obstacle has been overcome by the development of highly efficient cellular conversion protocols including the highly-efficient C/EBPaconversion of leukemic B-lymphocytes into non-tumorigenic macrophages (Rapino et al, 2013. PMID:23545498). Using this unique cellular system we have now collected a precious genome wide dataset containing information about the chromatin (accessibility -by ATAC-seq-, activity -by H3K27ac and H3K4me levels-and 3D configuration -by HiC-seq-); enhancer (by TT-seq and BRD4 ChIP-seq); DNA methylation (by WGBS-seq); and the transcriptomic status (by RNA-seq) of the cells at 4 different time points (0 -uninduced-, 24, 96 and 168 hours). We will integrate all these different layers of epigenome information using amulti-omics approach to gain comprehensive knowledge about the chromatin status at relevant gene regulatory regions during the C/EBPa induced loss of tumorigenicity.ReferencesApostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature502, 462-471.Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al.(2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science356.Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., Zhao, D., Liu, Y., Wang, C., Zhang, X., et al.(2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature525, 389-393.Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell126, 663-676.

28. POSTER S4: A mathematical model for strigolactone biosynthesis in plants

Abel Lucido[1,2], Oriol Basallo[1,2], Albert Sorribas[1,2], Ester Vilaprinyo[1,2], Rui Alves[1,2]

1 Systems Biology Group, Department Ciències Mèdiques Bàsiques, Faculty of Medicine, Universitat de Lleida, Lleida, Spain
2 Institut Recerca Biomedica de Lleida (IRBLleida), Lleida, Spain

Abstract

Strigolactones mediate plant development, trigger symbiosis with arbuscular mycorrhizal fungi, are abundant in 80% of the plant kingdom and help plants gain resistance to environmental stressors. They also induce germination of parasitic plant seeds that are endemic to various continents, such as *Orobanche* in Europe or Asia and *Striga* in Africa. While strigolactones are crucial hormones, their biosynthesis is less well understood. The genes involved in the early stages of strigolactones biosynthesis are well known in several plants. The regulatory structure and the latter parts of the pathway, where flux branching occurs to produce alternative strigolactones, are less well understood. Here we present a computational study that collects the available experimental evidence and proposes alternative biosynthetic pathways that are consistent with that evidence. Then, we test the alternative pathways through *in silico* simulation experiments and compare those experiments to experimental information. Our results predict the differences in dynamic behavior between alternative pathway designs. Independent of design, the analysis suggests that feedback regulation is unlikely to exist in strigolactone biosynthesis. In addition, our experiments suggest that engineering the pathway to modulate the production of strigolactones could be most easily achieved by increasing the flux of β-carotenes going into the biosynthetic pathway. Finally, we find that changing the ratio of alternative strigolactones produced by the pathway can be done by changing the activity of the enzymes after the flux branching points.

References

Alder,A. *et al.* (2012) The path from β-carotene to carlactone, a strigolactone-like plant hormone. *Science*, **335**, 1348–1351.

Mashiguchi,K. *et al.* (2021) Strigolactone biosynthesis, transport and perception. *The Plant Journal*, **105**, 335–350.

Waters,M.T. *et al.* (2012) The Arabidopsis Ortholog of Rice DWARF27 Acts Upstream of MAX1 in the Control of Plant Development by Strigolactones. *Plant Physiology*, **159**, 1073–1085.

Yoneyama,Kaori *et al.* (2018) Conversion of carlactone to carlactonoic acid is a conserved function of MAX1 homologs in strigolactone biosynthesis. *New Phytologist*, **218**, 1522–1533.

Zhang,Y. *et al.* (2014) Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis. *Nature Chemical Biology*, **10**, 1028–1033.

Zhu,C. *et al.* (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *PNAS*, **105**, 18232–18237.

29. POSTER S1: A BERT base model for the analysis of Electronic Health Records

Enrico Manzini[1,2,3], Alexandre Perera Lluna[1,2,3]

[1] B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Spain.
[2] Networking Biomedical Research Centre in the subject area of Bioengineering, Biomaterials and Nanomedicine, Madrid, Spain.
[3] Institut de Recerca Sant Joan de Deu, Barcelona, Spain

Abstract

In recent years the digitization of health data and the increasing availability of Electronic Health Records (EHRs) is opening new opportunities -as well as new challenges- for improving the health care process through precision medicine (Noura 2019), i.e. that process of creating better diagnostic and treatment response models tailored on patients data. Deep learning (DL), a subfield of Machine Learning, is having a big impact in the way EHRs are analyzed and used to create personalized prediction models. In particular, DL has been shown to get better results compared to traditional approaches in different tasks: from disease detection to sequential prediction of clinical events; from data augmentation to concept embedding (Xiao 2018). One of the biggest limitations in training DL models for specific tasks is the availability of labeled data for training and validation. For this reason the concept of transfer learning, i.e. training a model on a generic domain to transfer this knowledge on a different, more specific, domain (Weiss 2016), is gaining more and more strength in DL research. This is especially true in the field of Natural Language Processing (NLP), where the BERT model achieved state of the art performances in different tasks. In this work we proposed a BERT based model designed to work with diagnosis and medicament codes and different continuous variables. Moreover we introduced: a state vector describing static information about the patient that helps the model to better learn the sequence of EHRs; and a mechanism of relative time attention based on the Relative Position Representation (RPR) (Shaw 2018), in order to better learn the irregularity of the data. Results of this model outperformed classical supervised learning techniques such as recurrent neural networks and random forest algorithms.

References

1. Noura S. Abul-Husn and Eimear E. Kenny (2019), Personalized medicine and the power of electronic health records, Cell, 177, 58–69

2. Cao Xiao, et al (2018), Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review, Journal of the American Medical Informatics Association, 25, 1419–1428

3. Karl Weiss, et al. (2016)A survey of transfer learning. Journal of Big Data, 3:9, 12

4. Peter Shaw, et al (2018) Self-attention with relative position representations. NAACL HLT 2018-Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2, 464–468

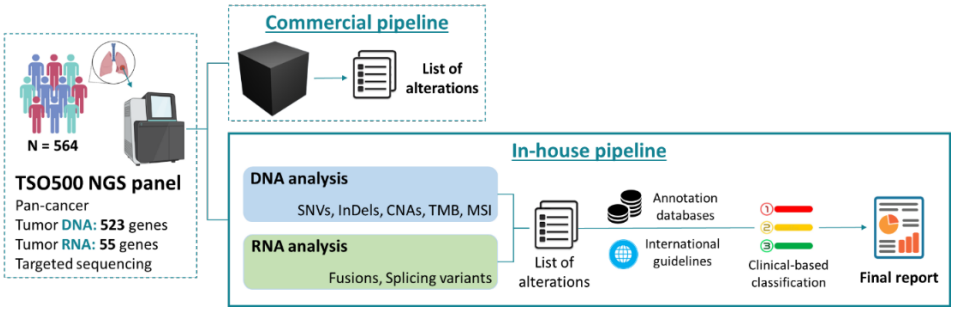
30. POSTER S2: An integrative bioinformatics pipeline for the analysis of a comprehensive NGS pan-cancer panel

Marín R,[1,2] Alay A,[1,2] Hijazo-Pechero S,[3,4] Moreno V,[1,2] Nadal E,[1,2] Solé X[4]

[1] Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, Spain.
[2] Institut Català d'Oncologia (ICO), L'Hospitalet de Llobregat, Spain.
[3] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.
[4] Hospital Clínic de Barcelona, Barcelona, Spain.

Abstract

Gene sequencing panels have emerged as a key diagnostic tool for the application of precision oncology in the clinical practice (Li *et al.*, 2017; Mateo *et al.*, 2018) . However, commercial panels often provide proprietary and non-customizable tools which generate numerous output metrics without visual reports, thus hindering the subsequent interpretation of results. By using tumor DNA and RNA sequencing data from the Illumina® TruSight™ Oncology 500 (TSO500) panel as input, here we present an open and customizable bioinformatics pipeline to (i) identify tumor small variants, copy number alterations (CNAs), cancer-related splicing variants, and gene fusions, (ii) measure the tumor mutational burden (TMB) and the microsatellite instability (MSI), and (iii) provide interpretable and visual results. Up to now, tumor alterations from 564 patients have been identified using both our pipeline and the commercial one (TSO500). Regarding small variants, our pipeline implements an integrative approach of multiple callers and external databases to generate a consensus and more robust list of annotated variants. Concerning CNAs, the TSO500 pipeline only reports amplifications of 59 genes, while we provide both amplifications and deletions of 518 genes. Moreover, the TSO500 only analyzes three genes (*AR*, *EGFR*, and MET) for detecting splicing variants, while our pipeline can identify cancer-specific isoforms for all the 55 RNA genes. Therefore, our project shows how the implementation of a customized pipeline provides more insight about the genomic alterations, graphical reporting of results, and it can be updated on a continuous to incorporate other analyses and new bioinformatics advances. All this makes our pipeline a robust option for analyzing this type of data.



References

Li,M.M. *et al.* (2017) Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *Journal of Molecular Diagnostics*, **19**, 4–23.

Mateo,J. *et al.* (2018) A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). *Annals of Oncology*, **29**, 1895.

31. POSTER S3: Fast evolution of piRNA cluster expression across three mouse species

Adrià Mitjavila Ventura[1,2,3], Tanya Vavouri[2]

1. Germans Trias i Pujol Research Institute (IGTP), Can Ruti Campus, 08916 Badalona, Spain
2. Regulatory Genomics group, Josep Carreras Leukaemia Research Institute (IJC), Campus ICO-Germans Trias i Pujol, 08916 Badalona, Spain.
3. PhD Program in Bioinformatics, Universitat Autònoma de Barcelona (UAB), 08193 Cerdanyola del Vallès, Spain.

Abstract

Piwi-interacting RNAs (piRNAs) are small non-coding RNAs that arise from long, single stranded piRNA precursors and are responsible for the silencing of transposable elements (TEs) in the germline of most animals[1]. Despite piRNA machinery being highly conserved across animals[2], piRNAs and piRNA-producing loci (piRNA clusters) are highly diverse across and within species[3,4]. In the current work, we studied the piRNA expression and its variation in the male germ line of three closely related mouse species, providing the first small RNA datasets in two of these species. In addition, we evaluated changes in sequence that could influence the emergence of piRNA clusters, their expression and their diversity across species. In summary, we found that piRNA clusters are highly species-specific, and most conserved clusters likely produce more piRNAs. We also found significant differences between 3'UTRs of genic piRNA clusters and other genes, but not compared to their non-piRNA producing orthologs. Finally, we found significant associations between species-specific TEs and differential piRNA production. Our results suggest that presence of transposon insertions may be the origin of a subset of new piRNA clusters. Thus far, this in one of the first piRNA studies comparing closely related species within the mammalian clade and it is a first step towards unravelling the mechanisms by which piRNA clusters emerge and evolve.

References

1. Ozata, D.M. *et al*. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics,* **20**, 89-108. doi: 10.1038/s41576-018-0073-3.

2. Gainetdinov, I. *et al.* (2018) A single mechanism of biogenesis, initiated and directed by PIWI proteins, explains piRNA production in most animals. *Molecular Cell*, **71**, 775-790. doi: 10.1016/j.molcel.2018.08.007.

3. Chirn, G.W. *et al* (2015) Conserved piRNA expression from a distinct set of piRNA cluster loci in eutherian mammals. *PLoS Genetics,* **11**, e1005652. doi: 10.1371/journal.pgen.1005652.

4. Ozata, D.M. *et al*. (2020) Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. *Nature Ecology Evolution,* **4**, 156-168. doi: 10.1038/s41559-019-1065-1.

32. POSTER S4: Characterization of post-COVID-19 condition subsyndromes through hierarchical clustering

Francisco Muñoz-López[1], Maria Nevot[1], Cristian Tebé[2], Gemma Lladós[3,4], Sergio España[3,4], José Ramón Santos[3,4], Cristina López[3,4], Cora Loste[3,4], Ruth Toledo[4], Marta Font[4], Anna Chamorro[4], Roger Paredes[1,3,4], Bonaventura Clotet[1,3,4], Lourdes Mateu[3,4,6], Marta Massanella[1,5,6]

[1]IrsiCaixa-AIDS Research Institute and Germans Trias i Pujol Health Research Institute (IGTP), Badalona, Spain
[2]The Bellvitge Biomedical Research Institute (IDIBELL), Hospitalet de Llobregat, Spain
[3]Infectious Diseases Department, Germans Trias i Pujol Hospital, Badalona, Spain
[4]Fight against AIDS Foundation (FLS), Germans Trias i Pujol Hospital, Badalona, Spain
[5]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas, CIBERINFEC, Madrid, Spain
[6]University of Vic–Central University of Catalonia (UVic-UCC), Vic, Spain

Abstract

During the past three years, over 600 million of confirmed COVID-19 cases have been reported. A remarkable portion (8-15%) of these patients present persistent symptoms at least three months after infection, some of them even lasting over a year. This heterogeneous set of symptoms, which have an impact on the everyday life of the patients, has been named as post-COVID-19 condition (PCC) or long COVID[1]. Several hypotheses have been proposed to understand the cellular and molecular processes underlying this disease, being the ones involving the immune system the most promising. The aim of this thesis is to fully characterize a set of PCC patients coming from KING Cohort extension, at both clinical and immunological level.

This particular work will be focused on the clinical and demographic aspects of the patients. For that matter, 358 well-characterized PCC patients were evaluated for 59 different persistent symptoms. Gower distance between subjects was calculated using symptomatology and several methods were applied to determine the most suitable clustering algorithm. According to the previous step, hierarchical clustering with Ward's minimum variance method was performed using the distance matrix. Average silhouette method was used to determine the optimal number of clusters, resulting in a global maximum of two clusters, and a local maximum of five clusters. Those five clusters were selected in order to increase the resolution of the subsyndromes. All analysis was performed using R 4.2.2.

69% of PCC were female and the median of age was 48 years (IQR [40-56]). Only 37% of PCC required hospitalization during acute SARS-CoV-2 infection. PCC showed a median of 10 symptoms (IQR [6-15]) and few comorbidities (median 2, IQR [1-3]). Symptomatology-based clustering revealed five differentiated groups of PCC with several degrees of severity. Most severe subsyndromes of PCC (cluster 3 and 5) experienced median of 20 [17-23] and 17 [15-19] symptoms respectively, being neurological, respiratory, gastrointestinal, and musculoskeletal most common symptoms. Both groups were predominantly mid age (40-50 years) women (89% in both cases) and <25% required hospitalization during acute infection. On the other hand, men tend to present a milder PCC, characterized by fatigue and dyspnoea.

References

1. Soriano JB *et al*. (2022). A clinical case definition of post-COVID-19 condition by a Delphi consensus. Lancet Infect Dis. 22(4):e102-e107.

33. POSTER S1: Characterization of sewage water samples in environmental proteomics using the ROIMCR-PLSDA methodology

C. Pérez-Lopez[1], A. Ginebreda[1], M. Carrascal[2], J. Abian[2], D. Barcelo[1,3] and R. Tauler[1]

[1] Institute of Environmental Assessment and Water Studies (IDAEA-CSIC), Department of Environmental Chemistry, Jordi Girona 18-26, 08034 Barcelona, Spain
[2] Biological and Environmental Proteomics, Institute of Biomedical Research of Barcelona, Spanish National Research Council (IIBB-CSIC/IDIBAPS), Rosellón 161, E-08036 Barcelona, Spain
[3] Catalan Institute for Water Research (ICRA-CERCA), Emili Grahit 101, Parc Científic i Tecnològic de la Universitat de Girona, Edifici H2O, 17003 Girona, Spain

Abstract

The Regions of Interest-Multivariate Curve Resolution (ROIMCR) methodology, recently proposed as a proteomic tool (Perez-Lopez *et al.,* 2021) , combined with Partial Least Squares-Discriminant Analysis (PLS-DA) is shown for the analysis of environmental proteomics (Carrascal *et al.,* 2020) samples.

Polymeric devices were placed for 11 days in the influent water of the Gavà-Viladecans (Barcelona, Spain) wastewater treatment plant at 3 different times, between April and May 2020, during the quarantine period of the SARS-COV 2 pandemic. Slices from these polymeric devices were cut, purified, and digested with trypsin. Tryptic peptides (a triplicated at each sampling time) were analyzed by LC-HRMS/MS and the data obtained were exported to MATLAB computational environment and analyzed by the ROIMCR procedure (Gorrochategui *et al.,* 2019) . A first step of filtering and compression of datasets was performed by the Regions of Interest (ROI) procedure, without losing mass accuracy. Multivariate Curve Resolution-Alternating Least Square (MCR-ALS) was applied to the ROI compressed data matrix resolving 181 'pure' components, explaining most of the experimental data variance (96.86%). Most of these components can be associated with peptide signals. The peak heights of the elution profiles of these components were then analyzed by Partial Least Squares-Discriminant Analysis (PLS-DA) (Wold *et al.,* 2001) , resulting in 41 MCR components (potential peptides) as being the main responsible of the observed differences among samples collected at different times. A final identification step of reanalyzing the target signals chosen by the ROIMCR-PLSDA was performed to set those peptides, and their protein inference, responsible for the main variability among samples.

As a final result, 39 proteins, represented by 333 peptides, have been identified among all the signals associated with the variability among samples resulting from ROIMCR-PLSDA. These proteins, such as immunoglobin domains, chaperonins, elongation factors or ribosomal proteins, belong to both eukaryotic organisms and bacterial species (e.g. human, mouse or bull). Finally, 27 of the 41 resolved MCR components were associated with a peptide and its corresponding protein providing a better characterization of the signals responsible for the main variability among samples.

References

Carrascal,M. *et al.* (2020) Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. *Science of the Total Environment*, **747**.

Gorrochategui,E. *et al.* (2019) ROIMCR: A powerful analysis strategy for LC-MS metabolomic datasets. *BMC Bioinformatics*, **20**, 1–17.

Perez-Lopez,C. *et al.* (2021) Non-target protein analysis of samples from wastewater treatment plants using the regions of interest-multivariate curve resolution (ROIMCR) chemometrics method. *J Environ Chem Eng*, **9**.

Wold,S. *et al.* (2001) PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **58**, 109–130.

34. POSTER S2: METHYLCLOCK: A Bioconductor Package To Estimate Dna Methylation Age

Dolors Pelegrí-Sisó,[1] Paula de Prado,[1] Justiina Ronkainen,[2] Mariona Bustamante,[1,3] Juan R González[1,3]

1. Barcelona Research Institute for Global Health (ISGlobal), Doctor Aiguader 88, 08003 Barcelona, Spain.
2. Center for Life Course Health Research, University of Oulu, Pentti Kaiteran katu 1, Linnanmaa, Oulu, Finland.
3. CIBER in Epidemiology (CIBERESP), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0 28029 Madrid, Spain

Abstract

Ageing is a biological and psychosocial process related to diseases and mortality. It correlates with changes in DNA methylation (DNAm) in all human tissues. Therefore, epigenetic markers can be used to estimate biological age using DNAm profiling across tissues. Alterations of the correlation between biological and chronological age (i.e. age acceleration or deceleration) have been related to several age-related diseases, pathological stages and can occur before clinical manifestation of diseases become overt. Therefore, epigenetic clocks are being considered as a tool in prevention, diagnostic and even in forensic applications. We developed methylclock[1], a Bioconductor package to compute epigenetic age. Since that moment, the package is being routinely updated with new epigenetic clocks. So far, it includes 15 DNAm adult/childhood and gestational clocks. They range from the classic Hovarth's[2] clock to more recent ones like EPIC predictor of gestational age. Methylclock has been used, among many others projects, to investigate the association between the early life exposome and epigenetic age acceleration in children[3] using the Pediatric-Buccal-Epigenetics' clock. It has also been used to characterize the ethnic differences in DNA methylation between UK-residents, South Asians and Europeans[4]. Another interesting study using methyclock showed that age acceleration is associated with alcohol use disorder[5], providing the first evidence for a recovery of this effect upon abstinence from alcohol. methylclock package is in continuous development and new clocks are added, some of them requested by researchers. In summary, methylclock enables quick and user-friendly computation of a large set of epigenetic clocks that allows the investigation of this biomarkers in different research areas aiming at improving personalized medicine and public health.

References

Pelegí-Sisó, Dolors, et al. "methylclock: a Bioconductor package to estimate DNA methylation age." Bioinformatics 37.12 (2021): 1759-1760.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome biology, 14(10), 1-20.

de Prado-Bert, Paula, et al. "The early-life exposome and epigenetic age acceleration in children." Environment international 155 (2021): 106683.

ELLIOTT, Hannah R., et al. Characterisation of ethnic differences in DNA methylation between UK resident South Asians and Europeans. medRxiv, 2022.

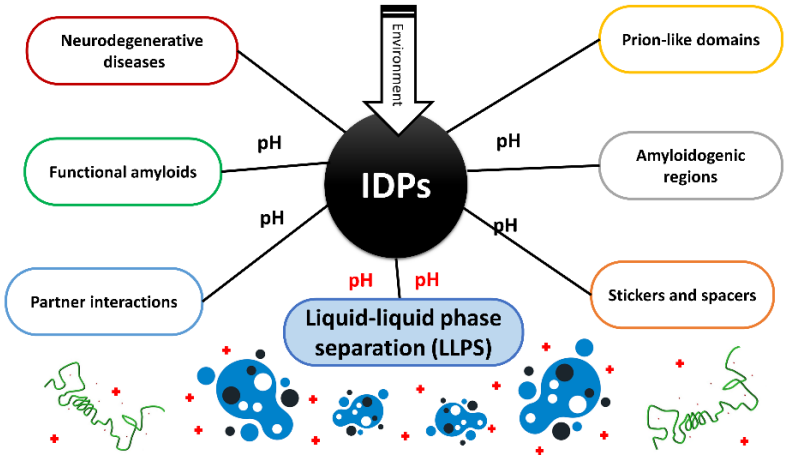
Zindler, Tristan, et al. "How alcohol makes the epigenetic clock tick faster and the clock reversing effect of abstinence." Addiction Biology 27.5 (2022): e13198.

35. POSTER S3: Large-scale computational study of pH-dependent solubility and disorder in liquid-liquid phase separating disordered regions

Carlos Pintado-Grima[1], Oriol Bárcenas[1] and Salvador Ventura[1]

[1]Institut de Biotecnologia i Biomedicina, Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain.

Abstract



Intrinsically disordered proteins (IDPs) are essential players in the assembly of biomolecular condensates during liquid-liquid phase separation (LLPS) (Nott, et al., 2015). Disordered regions (IDRs) are significantly exposed to the solvent and, therefore, highly influenced by fluctuations in the microenvironment (Uversky, 2009). Extrinsic factors, such as pH, modify the solubility and disorder state of IDPs, which in turn may impact the formation of liquid condensates (Santos, et al., 2020; Santos, et al., 2020). However, little attention has been paid to how the solution pH influences LLPS, despite knowing that this process is context-dependent (Alberti, et al., 2019). In this work (Pintado-Grima, et al., 2022), we have conducted a large-scale in-silico analysis of pH-dependent solubility and disorder in IDRs known to be involved in LLPS (LLPS-DRs). We found that LLPS-DRs present maximum solubility around physiological pH, where LLPS often occurs, and identified significant differences in solubility and disorder between proteins that can phase-separate by themselves or those that require a partner. We also analyzed the effect of mutations in the resulting solubility profiles of LLPS-DRs and discussed how, as a general trend, LLPS-DRs display physicochemical properties that permit their LLPS at physiologically relevant pHs.

References

Alberti, S., Gladfelter, A. and Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* 2019;176(3):419-434.

Nott, T.J.*, et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* 2015;57(5):936-947.

Pintado-Grima, C., Barcenas, O. and Ventura, S. In-Silico Analysis of pH-Dependent Liquid-Liquid Phase Separation in Intrinsically Disordered Proteins. *Biomolecules* 2022;12(7).

Santos, J.*, et al.* DispHred: A Server to Predict pH-Dependent Order-Disorder Transitions in Intrinsically Disordered Proteins. *International journal of molecular sciences* 2020;21(16).

Santos, J.*, et al.* pH-Dependent Aggregation in Intrinsically Disordered Proteins Is Determined by Charge and Lipophilicity. *Cells* 2020;9(1).

Uversky, V.N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The protein journal* 2009;28(7-8):305-325.

36. POSTER S4: Computational investigation of the developmental bases of morphological plasticity in *Drosophila* wings

Aleksa Ratarac[1], Carlos Mora Martinez[2], Isaac Salazar Ciudad[1,3]

1 Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Spain
2 Institute of Biotechnology, University of Helsinki, Finland
3 Centre de Recerca Matemàtica, Barcelona, Spain

Abstract

Although the wing of the fruit fly (*Drosophila melanogaster*) is one of the best understood animal organs in terms of development, there still remain significant gaps in our understanding of this process. The pupal stage is of particular interest, as the mechanics and the succession of its phases of isotropic growth, elongation and asymmetrical contraction are still largely elusive.

Building upon the framework of vertex models (Farhadifar *et al.*, 2007), and resting on the assumption that all morphogenetic movements are explainable by a finite set of cell behaviors and properties, the Salazar group has previously constructed a computational model of the asymmetrical contraction phase (Ray *et al.*, 2015), which reliably reproduced the wild-type wing morphology as well as some distinct mutant phenotypes solely by varying the system- or tissue-wide parameters. Based on previous successful attempts in other models (Salazar-Ciudad and Jernvall, 2010), this project aims to take this approach further by expanding and improving the model in an attempt to reproduce much finer variation (e. g. populational variation) and simulate all the experimentally observable phases of the pupal stage within a single model. In addition, a phenotypic plasticity experiment is being run in order to provide the morphological data on populational variation in wing morphology of flies reared at different temperatures and population densities.

The primary challenges of the project are producing a satisfactory wild-type morphology, reproducing the extent of experimentally observed variation by applying random perturbations to systematically chosen parameters of the model, and interpreting these results in the light of environment-phenotype and environment-genotype-phenotype interactions. The results are expected to provide insight and suggest hypotheses on which aspects of wing development can be influenced by environmental factors and to which extent, thus offering mechanistical explanations on how the environmental changes lead to different morphologies that are currently lacking.

References

Farhadifar,R. *et al.* (2007) The Influence of Cell Mechanics, Cell-Cell Interactions, and Proliferation on Epithelial Packing. *Current Biology*, **17**, 2095–2104.

Ray,R.P. *et al.* (2015) Patterned Anchorage to the Apical Extracellular Matrix Defines Tissue Shape in the Developing Appendages of Drosophila. *Developmental Cell*, **34**, 310–322.

Salazar-Ciudad,I. and Jernvall,J. (2010) A computational model of teeth and the developmental origins of morphological variation. *Nature*, **464**, 583–586.

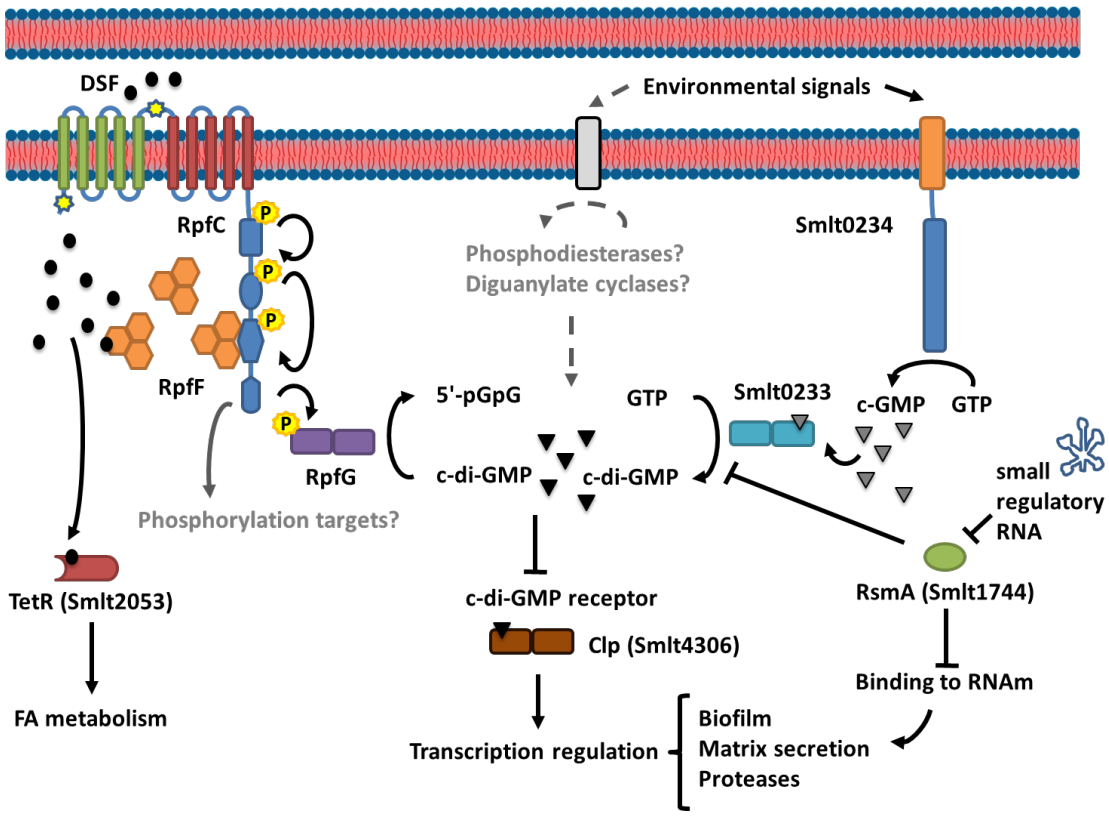
37. POSTER S1: Identification of potential inhibitors of *Stenotrophomonas maltophilia* Quorum Sensing

Enrique Rodríguez Bote[1], Xavier Daura Ribera[2]

1 Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès (Barcelona), Spain
2 Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès (Barcelona), Spain

Abstract

Stenotrophomonas maltophilia is an opportunistic Gram-negative pathogen with increasing incidence in clinical settings (Brooke, 2014). The most critical aspect of *S. maltophilia* is its frequent resistance to most of the antibiotics of clinical use (Kim *et al.*, 2019). Quorum Sensing (QS) systems coordinate bacterial populations and act as major regulatory mechanisms of pathogenesis in both pure cultures and poly-microbial communities. Disruption of QS systems, a phenomenon known as Quorum Quenching (QQ), represents a new promising paradigm for the design of novel antimicrobial strategies (Martínez *et al.*, 2019). In this context, we aim to target of the quorum sensing systems of *S. maltophilia* which can be used as an effective antimicrobial strategy, either alone, disarming the microorganism, or in combination with current antibiotics, providing a potentiator effect, paying special attention to Diffusible Signal Factor (DSF) signaling.



Here we aim to identify the allosteric sites in QS targets and perform a virtual screening of our compound libraries against active and allosteric sites for the targets selected. We selected two different proteins as druggable targets, the TETR family member PsrA and Clp transcriptional regulator, both known to be part of the DSF signaling system. The selection of the active site is performed using a Molecular Dynamics (MD) approach, we performed an MD simulation using AMBER (Case *et al.*, 2022) and selecting a putative allosteric site performing a principal component analysis of the residue pairwise interaction covariance matrix. Using the Enamine

database we have performed a Molecular Docking with MOE (Chemical Computing Group UL, 2022), selecting the best hits.

References

Brooke,J.S. (2014) New strategies against Stenotrophomonas maltophilia: A serious worldwide intrinsically drug-resistant opportunistic pathogen. *Expert Rev. Anti. Infect. Ther.*, **12**, 1–4.

Case,D.A. *et al.* (2022) Amber.

Kim,E.J. *et al.* (2019) Risk factors for mortality in patients with Stenotrophomonas maltophilia bacteremia and clinical impact of quinolone-resistant strains. *BMC Infect. Dis.*, **19**, 1–8.

Martínez,O.F. *et al.* (2019) Recent advances in anti-virulence therapeutic strategies with a focus on dismantling bacterial membrane microdomains, toxin neutralization, quorum-sensing interference and biofilm inhibition. *Front. Cell. Infect. Microbiol.*, **9**, 1–24.

Chemical Computing Group ULC. (2022) Molecular Operating Environment (MOE).

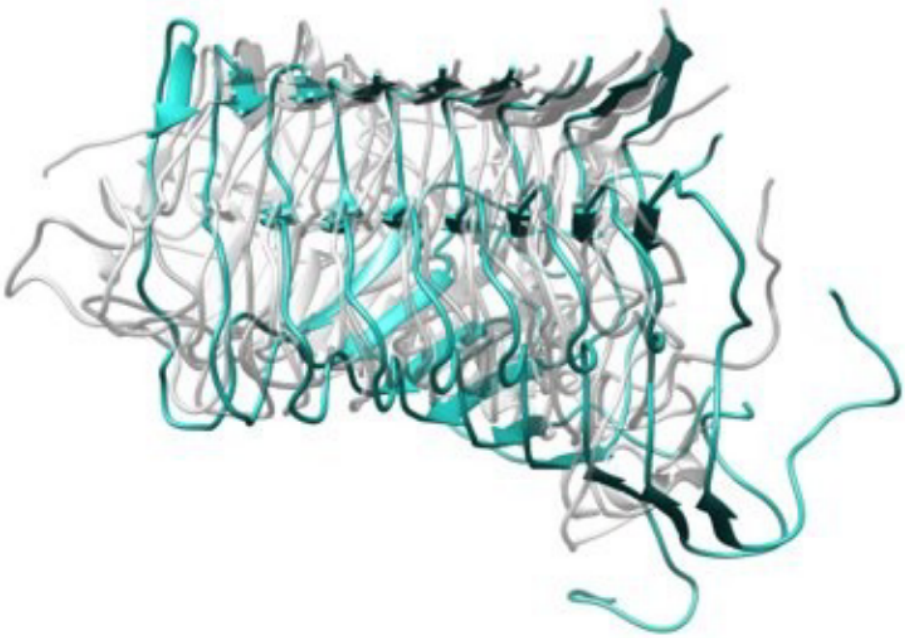
38. POSTER S2: Computational Assessment of theImpact of Cu(II) and Al(III) on b-Amyloid$_{42}$ Fibrils: Binding Sites, Structural Stability and PossiblePhysiological Implications

Lorena Roldán-Martín[1], Mariona Sodupe[1] and Jean-Didier Maréchal[1]

[1]Departament de Química, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

Abstract

One of the Alzheimer's Disease major hallmarks is the aggregation of b-amyloidpeptide, a process in which metal ions play an important role. In the presentwork, an integrative computational study has been performed. First, to identifymetal-binding regions, the in-house developed program BioMetAll and Golddocking software have been applied. Then, through classical and Gaussianaccelerated molecular dynamics, the conformational impact of Cu(II)and Al(III)ion binding to the b-amyloid (Ab42) fibrillar structure has been determined. It hasbeen observed that the metal-free fiber shows a hinge fan-like motion of the S-shape structure, maintaining the general conformation. Upon metal coordination,distinctive patterns are observed. On one side, Cu(II) binds to the flexible N-terminal region, in residues 13 and 15, similarly to the monomeric structure. Itsbinding induces structural changes that could ultimately disrupt the fibrillarstructure. On the contrary, Al(III) binding takes place on residues Glu22 andAsp23, whose binding reinforces the core stability of the system. These resultsgive clues on the molecular impact of the interaction of metal ions with theaggregates and sustain their non-innocent roles in the evolution of the illness.



39. POSTER S3: Integrating Artificial Intelligence Methods in Pharmacokinetics & Pharmacodynamics Processes

Sergio Sánchez-Herrero[1], Laura Calvet[2], Angel A. Juan[3]

1. Computer Science, Multimedia and Telecommunication Department, Universitat Oberta de Catalunya, 08018, Barcelona, Spain
2. Dept. of Telecomm. and Systems Engineering, Autonomous University of Barcelona, 08193, Bellaterra, Spain
3. Dept. of Statistics and Operations Research, Universitat Politècnica de València, 03801, Alcoy, Spain

Abstract

Characterization of the pharmacokinetics (PK) and pharmacodynamics (PD) of a drug product through population-based approaches and mathematical modelling of the relationships between exposure, safety, and efficacy is vital for pharmacological experts or pharmaceutical companies. In addition, PK and PD descriptions are essential for the approval process at regulatory agencies, such as the United States Food and Drug Administration (U.S. Food) or the European Medicines Agency (European Medicines Agency) [1, 2, 3]. There are many pharmacokinetics (PK) and pharmacodynamics (PD) software developed for these aims. However, common software use for developing Artificial Intelligence methods and techniques like R and Python could improve PK and PD processes. R or Python open-source software has developed a large number of complex and complete packages to perform data analysis, scientific computing, estimation and optimization algorithms, application development, backend web development or machine learning. Subsequently, R or Python packages are potential and general-purpose programming languages, which are embracing PK and PD modules too [4].

For this reason, the first aim of this research is embed R and Python inside PhysPK® PK/PD bio simulation software to improve pharmacokinetics processes [5]. PhysPK® is a powerful biosimulation and modeling software for population pharmacokinetic and pharmacodynamic systems analysis. Despite PhysPK provides well enough pre- and post-analysis diagnostic plots to characterize PK and PD to population-based approaches and mathematical modelling of the relationships between exposure, safety, and efficacy, R and Python could complete PK and PD analysis performing better data analysis, stats, graphs, reports, and so on to improve PhysPK outputs.

Second, PK/PD analysis provides mechanistic insight into biological processes but is time- and labor-intensive. In contrast, machine learning (ML) models are much quicker trained, but offer less mechanistic insights. The opportunity of using ML predictions of drug PK as input for a PK/PD model could strongly accelerate analysis efforts. It could be explore the ability of different ML algorithms to predict drug PK. Based on simulated data, we trained different ML algorithms to predict the plasma concentration-time inside PK/PD population model [6].

Third, during and after PK/PD model development, the likelihood of the predictions to fit the data, i.e., the probability of the model parameters being able to describe the data, is estimated in a maximum likelihood parameter estimation, often using differential equation systems. PhysPK is able to export PK/PD models to Python. As consequence, it could be analyses other estimation and optimization method to predict plasma concentration-time with estimation algorithms in Python [7].

PhysPK´s new connection to cutting-edge technologies, such as R and Python, allows a more comprehensive analysis and results with high-quality graphs, statistics, summary reports and tables, faster estimation and optimization methods, and any machine learning advantages.

With this new approach, greater efficiency, cost-effectiveness and timeliness in the drug development process are achieved (Figure 1).

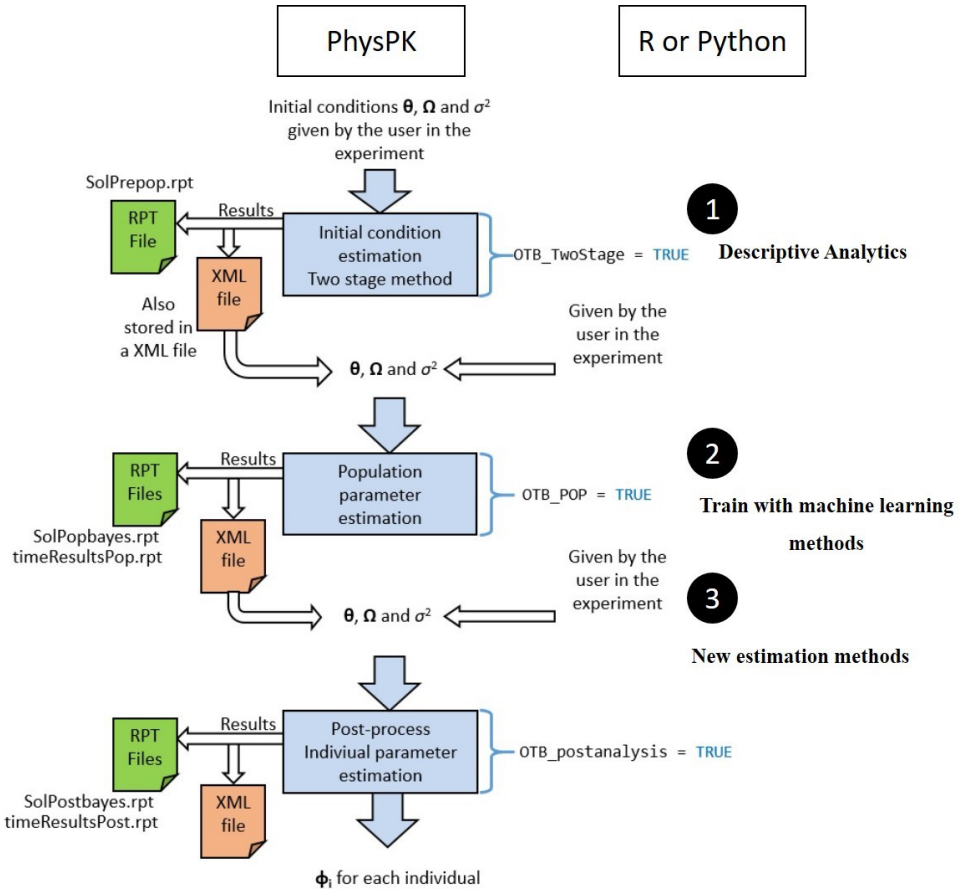


Figure 1. PhysPK provides the process for PK/PD population analysis based on three phases (Figure 1): (i) Initial condition estimation. It is applied the Two Stage (TS) method [8] to obtain an initial solution of mean population parameters and their distribution. (ii) Population parameter estimation. The population parameters are estimated using FO, FOCE or FOCE-I estimation methods, starting from values from TS stage or user's manual, as desired. (iii) Post-process. The specific parameters for each individual are estimated through the population parameters distribution computed in the second stage or provided by the user.

References

1. Rajman, I. (2008). PK/PD modelling and simulations: utility in drug development. Drug discovery today, **13(7-8)**, 341-346.

2. Guiastrennec, B., *et al*., (2013). AMGET, an R-based post processing tool for ADAPT 5. CPT: pharmacometrics & systems pharmacology, **2(7)**, 1-10.

3. Sun, D., *et al*., (2022). Why 90% of clinical drug development fails and how to improve it?. Acta Pharmaceutica Sinica B.

4. Acharya, C., *et al*., (2016). A diagnostic tool for population models using non-compartmental analysis: the ncappc package for R. Computer Methods and Programs in Biomedicine, **127**, 83-93.

5. Reig-Lopez, J., *et al*., (2020). A multilevel object-oriented modelling methodology for physiologically-based pharmacokinetics (pbpk): Evaluation with a semi-mechanistic pharmacokinetic model. Computer Methods and Programs in Biomedicine, **189**, 105322.

6. Keutzer, L., *et al*., (2022). Machine learning and pharmacometrics for prediction of pharmacokinetic data: differences, similarities and challenges illustrated with rifampicin. Pharmaceutics, 14(8), 1530.

7. Mao, J., *et al.,* (2022). Applying machine learning to the pharmacokinetic modeling of cyclosporine in adult renal transplant recipients: a multi-method comparison. Frontiers in Pharmacology, **13**.

8. Davidian, M. (2010). "Nonlinear mixed effects models" in International Encyclopedia of Statistical Science.Editor M. Lovric (New York: Springer).

40. POSTER S4: Radiogenomics for personalized medicine in chronic obstructive pulmonary disease (RadGen4COPD)

David Sarrat González[1,2]

1 Universitat Autónoma de Barcelona, Bellaterra, Spain
2 Institut de Salut Global de Barcelona (ISGlobal), Barcelona, Spain

Abstract

Chronic obstructive pulmonary disease (COPD) is a major global health issue, and current methods for predicting clinical outcomes and stratifying patients are inadequate[1,2]. The RadGen4COPD project aims to address this challenge by using a combination of medical images and genomic data to improve prediction and stratification in COPD[3].

To achieve this goal, we propose the development of a platform called RadioSHIELD, which will be a federated radiogenomics platform that allows for the analysis of data from multiple sources in a way that complies with the General Data Protection Regulation (GDPR)[4]. This platform will use transfer learning algorithms to improve prediction in COPD[5] by reusing existing models, and will also integrate imaging and genomic data to discover new COPD biomarkers and understand the biological mechanisms underlying the disease[6].

In addition to its predictive capabilities, the RadioSHIELD platform will be designed to be open-source, user-friendly, and GDPR-compliant. It will also use a Medical Imaging Data Structure (MIDS) to harmonize and organize data from different sources[7]. We will also employ non-disclosive advanced methods to integrate images and genomes through transfer deep learning and scalable univariate massive genome-wide association studies (GWAS).

Overall, the RadGen4COPD project has the potential to make pioneering progress in the field of radiogenomics and COPD research. By demonstrating the utility of this approach in clinical translation and research, we hope to significantly improve the prediction of clinical outcomes and stratification of patients with COPD.

References

Agustí, A., Celli, B. & Faner, R. (2017) What does endotyping mean for treatment in chronic obstructive pulmonary disease? Lancet (London, England) 390, 980–987.

Celli, B. R. & Agustí, A. (2018) COPD: time to improve its taxonomy? ERJ open Res. 4.

Fan, C. C. et al. (2018) Beyond heritability: Improving discoverability in imaging genetics. Human Molecular Genetics 27, R22–R28.

F, S. et al. (2018) To share or not to share? Expected pros and cons of data sharing in radiological research. Eur. Radiol. 28, 2328–2335.

Martinez, A. R. (2020) Classification of COVID-19 in CT Scans using Multi-Source Transfer Learning.

Shui, L. et al. (2021) The Era of Radiogenomics in Precision Medicine: An Emerging Approach to Support Diagnosis, Treatment Decisions, and Prognostication in Oncology. Front. Oncol. 10, 3195.

Saborit-Torres, J. M. et al. (2020) Medical imaging data structure extended to multiple modalities and anatomical regions.

41. POSTER S1: Integration of metagenomics, metatranscriptomics and metabolomics in IBD

Gerard Serrano Gómez[1,3], Chaysavanh Manichanh[1,2]

1 Gut Microbiome Group, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
[2] Medicine Department, Autonomous University of Barcelona (UAB), 08193 Cerdanyola del Vallès, Spain
[3] Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193, Spain.

Abstract

Inflammatory Bowel Disease (IBD) is the term used to describe two of the most common chronic inflammatory diseases of the gastrointestinal (GI) tract: Crohn's Disease (CD) and Ulcerative Colitis (UC). Both IBD subtypes are characterised by the alternation of periods of clinical remission and relapse. Discrimination between the two subtypes proves to be challenging, with up to 10% of the patients being incorrectly diagnosed[1].

Although the factors that trigger CD and UC development are still unknown, several studies demonstrated that gut microbiota alterations are associated with IBD, with an increase of Proteobacteria and depletion of Firmicutes in CD-affected individuals[2–4] and a decrease of butyrate-producing bacteria in UC[5,6]. However, environmental factors such as diet, smoking habits, antibiotic usage, stress or sleeping schedule have also been linked to the development of IBD[7]. It is believed that the increase of pathogenic bacteria in the GI tract alters gut permeability, causing a disbalance in the microbial community known as dysbiosis, and an alteration of the metabolite composition in the GI tract[8–11], which ultimately leads to gut inflammation.

Previous studies used metagenomics to identify taxonomic changes in IBD[12,13], metatranscriptomics to characterize the metabolic pathways involved in the disease[14] and metabolomics to describe metabolites linked to CD and UC[15]. Although there have been efforts to integrate the layers of information from different omics techniques[14,16–18], to the best of our knowledge there are no studies that integrate in a completely untargeted manner information from shotgun metagenomics, metatranscriptomics and metabolomics in stool samples to provide a comprehensive overview of the IBD mechanisms in human patients.

Here, we present the research plan of our project, which aims to integrate information from various omics technologies aiming to identify the functional role of the microbiome in IBD subtypes and the mechanisms that lead to the development of the disease.

References

1. Monteiro, S. et al. Essential role of small bowel capsule endoscopy in reclassification of colonic inflammatory bowel disease type unclassified. World J. Gastrointest. Endosc. 9, 34–40 (2017).

2. Halfvarson, J. et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat. Microbiol. 2, 17004 (2017).

3. Baumgart, M. et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. ISME J. 1, 403–418 (2007).

4. Manichanh, C. et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut 55, 205–211 (2006).

5. Kumari, R., Ahuja, V. & Paul, J. Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India. World J. Gastroenterol. 19, 3404–3414 (2013).

6. Machiels, K. et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. Gut 63, 1275–1283 (2014).

7. Gomaa, E. Z. Human gut microbiota/microbiome in health and diseases: a review. Antonie Van Leeuwenhoek 113, 2019–2040 (2020).

8. Strugala, V., Dettmar, P. W. & Pearson, J. P. Thickness and continuity of the adherent colonic mucus barrier in active and quiescent ulcerative colitis and Crohn's disease. Int. J. Clin. Pract. 62, 762–769 (2008).

9. Schmitz, H. et al. Altered tight junction structure contributes to the impaired epithelial barrier function in ulcerative colitis. Gastroenterology 116, 301–309 (1999).

10. Söderholm, J. D. et al. Epithelial permeability to proteins in the noninflamed ileum of Crohn's disease? Gastroenterology 117, 65–72 (1999).

11. Nishida, A. et al. Gut microbiota in the pathogenesis of inflammatory bowel disease. Clin. J. Gastroenterol. 11, 1–10 (2018).

12. Pascal, V. et al. A microbial signature for Crohn's disease. Gut 66, 813–822 (2017).

13. Franzosa, E. A. et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. 4, 293–305 (2019).

14. Zhang, Y. et al. Metatranscriptomics for the Human Microbiome and Microbial Community Functional Profiling. Annu. Rev. Biomed. Data Sci. 4, 279–311 (2021).

15. Marchesi, J. R. et al. Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. J. Proteome Res. 6, 546–551 (2007).

16. Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662 (2019).

17. Jwj, L. et al. Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. Cell Host Microbe 29, (2021).

18. N, M. et al. Longitudinal multi-omics analysis identifies early blood-based predictors of anti-TNF therapy response in inflammatory bowel disease. Genome Med. 14, (2022).

42. POSTER S2: Development and application of tools for automated integration and analysis of big data in forestry management

Eva Tejada

Alter Software, Colombia

Abstract

There is a huge amount of forest-related data from around the world that can be analyzed. Researchers also need to know how is the status of this data. This Ph.D. work aims to integrate forest data, analyze what its quality is and apply some artificial intelligence techniques to find patterns over time that contribute to decision making about the land management, sustainability and prevent the effects of climate changes. The objective of this presentation is to socialize the context, objectives, methodology and the first results of the research work carried out in the Ph.D. period.

43. POSTER S3: Reanalysis of next generation sequencing data from patients with Fanconi anaemia-like and cardiac diseases

Eudald Tejero[1], Massimo Bogliolo[1-3], Roser Pujol[1-3], Benjamín Rodríguez-Santiago[3,4], Jordi Surrallés[1,4]

[1]Genome Instability and DNA Repair Syndromes Group. IR-Sant Pau - Sant Pau Institute of Biomedical Research (IIB)
[2]Joint Unit in Genomic Medicine UAB-IR Sant Pau
[3]Center for Biomedical Network Research on Rare Diseases (CIBERER)
[4]Genetics Department, Hospital de la Santa Creu i Sant Pau, Barcelona

Abstract

The utilization of exome sequencing technology in genetics diagnosis of rare diseases patients is becoming routinary in the majority of laboratories (Vinkšel *et al.*, 2021). Depending on patients' clinical phenotypes the percentage of successful diagnosis is between 25 and 58% (Fung *et al.*, 2020). Recent literature recommends reanalysing negative cases after a period of time to reach a genetic diagnosis (Dai *et al.*, 2022), taking advantage of variant databases updates and improved pipelines. Here we propose next generation sequencing data reanalysis of two different patients groups: 1) Fanconi anaemia (FA)-like patients, and 2) patients with different cardiac diseases, including dilated, hypertrophic and arrhythmogenic cardiopathies, and aortic pathology. WES data from six FA-like patients (group 1) was reanalysed by using a custom pipeline. To find candidate variants, filters considering variant frequency expected disease incidence, inheritance model and mutation type were employed. Likely disease-causing variants were found in two patients. Copy number variants of patients negative for the single nucleotide variants and small indels reanalysis are being analysed. Regarding group 2, data from 383 patients will be reanalysed to find candidate intronic variants affecting splicing using AI prediction tools such as SpliceAI (Jaganathan *et al.*, 2019).

References

Dai,P. et al. (2022) Recommendations for next generation sequencing data reanalysis of unsolved cases with suspected Mendelian disorders: A systematic review and meta-analysis. Genetics in Medicine, 24, 1618–1629.

Fung,J.L.F. et al. (2020) A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. npj Genom. Med., 5, 37.

Jaganathan,K. et al. (2019) Predicting Splicing from Primary Sequence with Deep Learning. Cell, 176, 535-548.e24.

Vinkšel,M. et al. (2021) Improving diagnostics of rare genetic diseases with NGS approaches. J Community Genet, 12, 247–256.

44. POSTER S4: Study of differential expressed genes between abdominal aortic cases and controls

Gerard Temprano-Sagrera[1], Mercedes Camacho[1], Maria Sabater-Lleal[1,2], Ana Viñuela[3]

[1]Unit of genomics of Complex Disease, Institut d'Investigació Biomedica Sant Pau, Barcelona, Spain.
[2]Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden
[3]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, UK.

Abstract

The abdominal aortic aneurysm (AAA) is a complex vascular disease characterized by a local dilatation of the aorta on its abdominal segment, and a deterioration of the vascular tissue internal structure. (Sakalihasan *et al.*, 2018). AAA is difficult to diagnose as it usually cause no symptoms until the moment it breaks, when it causes mortality rates around 80% (Aggarwal *et al.*, 2011). Our objectives are to I) characterize gene expression profiles of aortic tissue, ii) identify AAA specific gene expression profiles, and iii) identify molecular drives of aortic rupture. For this, we are using RNAseq data from abdominal aortic control samples from deceased organ donors (N=47) and aortic AAA samples from alive donors (N=97). Given that samples from healthy donors were obtained from cadavers, the detection of differentially expressed genes (DEG) between cases and controls would be virtually indistinguishable from the influence of ischemic time. To overcome this limitation, we are currently evaluating the influence of ischemic time in gene expression using available suitable samples (Artery Aorta (N=387), Artery Coronary (N=213), Artery Tibial (N=584), Whole Blood (N=670)) from the GTEx consortium. We performed a preliminary analysis identifying 12,134 DEG between cases and controls (FDR < 0.05). Gene set enrichment analysis among the most significantly DEG (Bonferroni < 0.05) demonstrated an enrichment with immune response processes**.** To evaluate the reliability of our results, we compared our list of DEG with a previously published similar study that used microarray expression data (Lindquist Liljeqvist *et al.*, 2020). Among the 8,590 identified DEG, 5,288 were also identified in our study, showing that even with a suboptimal experimental design we would detect genes modulated by AAA. Current work is further evaluating the molecular differences across AAA individuals using protein-protein interaction network analysis, co-expression networks with weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) and the role of alternative splicing in AAA (Li *et al.*, 2018). Finally, we are performing an allelic-specific expression analysis to identify genetic drivers of aortic rupture (van de Geijn *et al.*, 2015). Overall, our initial approach to identify AAA gene expression profiles has been able to replicate previously published findings, proving the value of RNAseq data to investigate complex diseases.

References

Aggarwal,S. *et al.* (2011) Abdominal aortic aneurysm: A comprehensive review. *Exp Clin Cardiol*, **16**, 11–15.

van de Geijn,B. *et al.* (2015) WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*, **12**, 1061–1063.

Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Li,Y.I. *et al.* (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet*, **50**, 151–158.

Lindquist Liljeqvist,M. *et al.* (2020) Tunica-Specific Transcriptome of Abdominal Aortic Aneurysm and the Effect of Intraluminal Thrombus, Smoking, and Diameter Growth Rate. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **40**, 2700–2713.

Sakalihasan,N. *et al.* (2018) Abdominal aortic aneurysms. *Nat Rev Dis Primers*, **4**, 1–22.

45. POSTER S1: Low input promoter capture Hi-C: a method to decipher the molecular mechanisms underlying non-coding alterations

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

1. Josep Carreras Leukaemia Research Institute (IJC)
2. Barcelona Supercomputing Center (BSC-CNS)

Abstract

Long-range interactions between regulatory elements and promoters are key in gene transcriptional control but challenging to study due to the lack of low input methods. Here we introduce low input capture Hi-C (liCHi-C), a cost-effective, customizable method to map and robustly compare promoter interactomes at high resolution in rare cell populations. As a proof of it broaden applicability, we use liCHi-C to in vivo study normal and malignant human hematopoietic hierarchy. We demonstrate that the dynamic promoter architecture foreshadows developmental trajectories meanwhile orchestrates transcriptional decisions along cell commitment. Simultaneously, liCHi-C enables the identification of new disease-relevant cell types, genes and gene pathways potentially deregulated by non-coding alterations at distal regulatory elements. Besides, we show liCHi-C ability to genome-wide uncover structural variants, resolved their breakpoints and propose pathogenic effects, including the formation of new regulatory landscape. Therefore, liCHi-C allows the study of previously unmeasurable cell types to ultimately illuminate disease etiopathogenesis.

46. POSTER S2: Applications of Data Science, Artificial Intelligence, Data Analytics, Machine Learning, and Federated Learning in Healthcare

Abtin Tondar[1,4*], Laura Calvet Liñan[2], Angel A. Juan1[1,3], Amir Bahmani[4,5]

[1]Faculty of Computer Science and Telecommunications, Universitat Oberta de Catalunya, Barcelona, Spain
[2]Autonomous University of Barcelona, 2 Calle Emprius, Sabadell, Spain
[3]Department of Applied Statistics & Operations Research at the Universitat Politècnica de València, Valencia, Spain
[4]Stanford Deep Data Research Computing Center, Stanford University School of Medicine, Palo Alto, CA, USA
[5]Stanford Healthcare Innovation Lab, Stanford University School of Medicine, Palo Alto, CA, USA
Corresponding Author: Abtin Tondar

Abstract

This poster presents a summary of some of our recent research findings as well as our present study related to the applications of data science, artificial intelligence, data analytics, machine learning (ML), and federated learning (FL) in healthcare. Using ML in healthcare to extract information that can enhance healthcare decision-making is becoming increasingly common. For example, when scheduling multi-period medical treatments for patients with cancer, medical committees must consider a large amount of data, variables, sanitary and budget constraints, as well as probabilistic elements. Hence, decisions must be made based on many factors, such as the priority of each patient as well as available treatments and their expected effects. To support this complex decision-making process, we introduce a novel methodology that combines a biased-randomized heuristic with simulation, to return[1].

More than ever, healthcare systems worldwide can use data, predictive models, and intelligent algorithms to optimize their operations and the service they provide to citizens. This paper reviews the existing literature regarding the use of data science/analytics methods and artificial intelligence algorithms in healthcare. This paper also discusses how healthcare organizations can benefit from these tools to efficiently deal with a myriad of new possibilities and strategies they can use to fulfill the citizens' needs and generate better healthcare services[2].

In our ongoing study, we work on FL to address one of the biggest challenges for training ML algorithms for healthcare decision-making, which is protecting patient privacy. Due to the sensitivity of patient medical data, there is an increasing demand for developing high-quality ML models based on different and extensive datasets. We are developing FL approaches that allow decentralized machine-learning model training with cloud operating datasets without compiling and damaging data[3].

Keywords: Data Science, Artificial Intelligence, Data Analytics, Machin-Learning, Healthcare

References

1-Leandro, CM., Juliana C, Angel A.J, Abtin T, Laura C, Barry B, Jose LS. (2021) Supporting Efficient Assignment of Medical Resources in Cancer Treatments with Simulation-Optimization. *2021 Winter Simulation Conference (WSC)*

2- Sergio B, Juliana C, Abtin T, Laura C, Angel A.J. (2022) Data Science, Analytics and Artificial Intelligence in e-Health: Trends, Applications, and Challenges. *Intl. Trans. in Op. Res* (In-press)

3- Amir B,Kyle F,Vandhana K,Arash A,Amir A,Philip S. T, Michael P.S ,Cuiping P (2021). Swarm: A federated cloud framework for large-scale variant analysis. *PLOS Comput. Biol,* **17**, 1-27.

47. POSTER S3: Benchmarking and improving imputation approaches for recurrent inversions in the human genome

Illya Yakymenko[1], Jon Lerga-Jaso[1], Mario Cáceres[1,2]

Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain
ICREA, Barcelona, Spain

Abstract

Inversions are a type of structural variant that are involved in phenotypic differences among individuals. Due to certain features, such as the fact of usually implying no loss or gain of DNA or the presence of large inverted repeats at their breakpoints, the characterization of inversions is quite difficult. It has been recently found that in humans many inversions are recurrent and they are not linked to other genomic variants. For this reason, the effect of recurrent inversions has been largely missed in current GWAS, and it is necessary to develop new methods to predict inversion genotypes accurately in the datasets of interest. Here, we have done a benchmarking analysis of the genotype predictions among different imputation tools, comparing IMPUTE2[1] with other softwares, such as IMPUTE5[2], BEAGLE[3] or/and scoreInvHap[4]. The accuracy was calculated as the r2 between experimental and imputed genotypes. From our set of 130 experimentally genotyped inversions 55 are recurrent. We found out that 23 and 18 out of 55 were imputable (r2 > 0.8) in European and African populations, respectively, but it varies among softwares. Nevertheless, this ratio increases to 26/55 for both populations when we filter out samples with a post-imputation genotype probability lower than 0.8. Finally, we are also testing a tool based on deep learning which could avoid the HMM-based algorithm limitations, as the position or linkage dependence and region complexity, to increase our catalogue of imputable inversions.

References

1. Howey, B.N. et al. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics, 5, 6.
2. Rubinacci, S. et al. (2020) Genotype imputation using the Positional Burrows Wheeler Transform. PLoS genetics, 16, 11.
3. Browning, B. L. et al. (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. American journal of human genetics, 103, 338-348.
4. Ruiz-Arenas, C. et al. (2019) scoreInvHap: Inversion genotyping for genome-wide association studies. PLoS genetics, 15, 7.

48. POSTER S4: A systems view of glioblastoma cell state transitions at single-cell resolution

Jing Yang[1], Jordi Villà-Freixa[1] and Adrián López García de Lomana[2]

[1] BI-SQUARED Research Group, Department of Biosciences, Faculty of Sciences, Technology and Engineering, Universitat de Vic - Universitat Central de Catalunya, 08500 Vic, Spain.
[2] Center for Systems Biology, University of Iceland, 101 Reykjavík, Iceland.

Abstract

Glioblastoma (GBM) is the most common and lethal primary brain tumour, which remains incurable due to its predominant characteristics, including intratumor heterogeneity, high filtration and rapid progression. GBM patients undergoing standard surgical resection of tumour mass followed by radiotherapy and chemotherapy still have a median overall survival of 8 months, and a five-year survival rate of 6.8% [1]. Coexisting treatment resistance and lack of response to targeted therapies contribute to recurrence. A concurrent proneural-to-mesenchymal cell state shift has been observed, where the proneural subtype alone is drug-susceptible and estimated to drive significant prolonged survival, up to ten years [2]. In this project, we propose to elucidate the transcriptional mechanisms and potential metabolic vulnerabilities that underlie drug-induced state transitions in GBM at whole-genome scale and single-cell resolution. Specifically, we plan to use publicly available single-cell transcriptome profiles, chromatin accessibility and phenotype data from primary-recurrent patient-matched paired GBM tumour samples [3] to build a gene regulatory network model using available computational tools like SCENIC [4]. Furthermore, we plan to develop expression-constrained genome-scale metabolic models from single-cell transcriptome data using available methods like rFASTCORMICS [5] to further understand the metabolic consequences of these cell state transitions. Ultimately, we plan to identify FDA-approved drugs against critical transcriptional regulators and metabolic enzymes as potential modulators of particular microstates across the cell state transition trajectory. The outcomes of this project will include (i) a systems-scale regulatory network from single-cell transcriptional profiles across a phenotypic transition, (ii) a catalogue of metabolic vulnerabilities across a drug-induced state transition, and (iii) a rational strategy to discover and repurpose FDA-approved drugs in treating GBM.

References

1. About glioblastoma. National Brain Tumor Society. (2022, September 15). Retrieved December 24, 2022, from https://braintumor.org/events/glioblastoma-awareness-day/about-glioblastoma/
2. Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., Mesirov, J. P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B. A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., Feiler, H. S., … Cancer Genome Atlas Research Network (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer cell, 17(1), 98–110. https://doi.org/10.1016/j.ccr.2009.12.020
3. Wang, L., Jung, J., Babikir, H., Shamardani, K., Jain, S., Feng, X., Gupta, N., Rosi, S., Chang, S., Raleigh, D., Solomon, D., Phillips, J. J., & Diaz, A. A. (2022). A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. *Nature cancer*, *3*(12), 1534–1552. https://doi.org/10.1038/s43018-022-00475-x
4. Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J. C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods*, *14*(11), 1083–1086. https://doi.org/10.1038/nmeth.4463
5. Pacheco, M. P., Bintener, T., Ternes, D., Kulms, D., Haan, S., Letellier, E., & Sauter, T. (2019). Identifying and targeting cancer-specific metabolism with network-based drug target prediction. EBioMedicine, 43, 98–106. https://doi.org/10.1016/j.ebiom.2019.04.046