XII EDITION BIOINFORMATICS OPEN DAYS

CONFERENCE BOOK

UNIVERSITY OF MINHO
GUALTAR CAMPUS

MARCH 16, 17 AND 18TH, 2023

Bioinformatics Open Days 2023

/bioinformaticsopendays

Bioinformatics Open Days 2023

WELCOME MESSAGE

Greetings,

It is with great pleasure that we welcome you all to the XII Edition of the Bioinformatics Open Days! This student-led initiative has been promoting knowledge exchange among students, teachers, and researchers in the Bioinformatics and Computational Biology fields since 2012, and this year's event promises to be as exciting as ever. The scope of the main scientific sessions and keynote lectures cover a range of topics that include genomic data infrastructures, the role of bioinformatics in the study of viruses, and a fresh perspective focused on sustainability and the impact of our daily tasks on the current climate state.

In addition to these stimulating talks, we have planned three workshops with different applications and backgrounds, as well as innovative oral and poster communications. Furthermore, we are also including a discussion on the reality of the corporate world, which will provide participants with valuable insights and guidance on starting their professional life in Bioinformatics.

We are thrilled to be a platform for knowledge sharing and community building within the Bioinformatics field. We hope that this year's event will provide a great experience for everyone and help advance the future of Bioinformatics, both nationally and internationally.

Thank you for your participation and enthusiasm, and we look forward to seeing you at the XII Edition of the Bioinformatics Open Days.

Kind regards,

The Organizing Committee of BOD 2023.

Bioinformatics Open Days 2023

THE ORGANIZING COMMITTEE

MIGUEL ROCHA
General Chair

CAMILA BABO
President
Scientific Program

ALEXANDRA COELHO
Vice-President
Logistics

ROBERTO BULLITTA
Finances

ANA RAFAELA PEREIRA
Scientific Program

VÂNIA MIGUEL
Scientific Program

DANIELA LEMOS
Logistics

ALEXANDRE CASTRO
Logistics

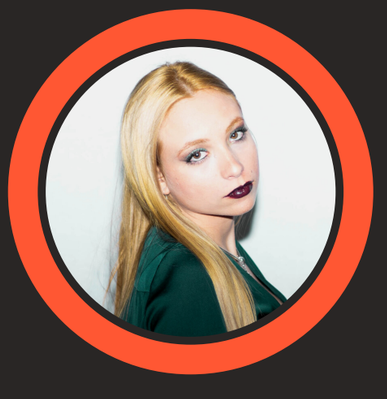
CATARINA FERREIRA
Logistics

ANDREIA GOMES
Logistics

SÉRGIO MENDES
Logistics

RODRIGO RIBEIRO
Logistics

MÓNICA LEIRAS
Event Promotion



RUTE CASTRO
Event Promotion

ANA LISBOA
Event Promotion

Bioinformatics Open Days 2023

/bioinformaticsopendays

Bioinformatics Open Days 2023

CONTENTS

Bioinformatics Open Days 2023

/bioinformaticsopendays

Bioinformatics Open Days 2023

PROGRAM

MARCH 16TH - THURSDAY

09:30 H	**Opening session**
10:00 H	**Scientific Session [Sponsored by BioData.pt]** • Genomic Data Infrastructure for Health Research Panel Chaired by Ana Melo Kjell Peterson, Ana Teresa Freitas, Sérgio Sousa and Maria Salazar
11:30 H	**Coffee Break**
11:45 H	**Oral Communications [Session 1]** • "Structural Bioinformatics", chaired by Vítor Pereira A bioinformatics approach to study the permeation of solutes through PfAQP for the development of new antimalarial therapies - Marta Batista Computational study of promising smart metallodrug delivery systems - Inês Pires Molecular Dynamics as an active resource on the development of new formulations for commercial use: The case of natural silicone alternatives - Tiago Ferreira The impact of the SARS-CoV-2 Omicron variant on the receptor binding domain conformational dynamics and interaction with human ACE2 - Rita I. Teixeira
12:45 H	**Posters Highlights**
13:00 H	**Lunch**
14:30 H	**Keynote Lecture** • Multi-country outbreak of monkeypox virus: genomics as a tool to identify genetic clustering, transmission dynamics and the role of Portugal in the international virus spread João Paulo Gomes

Bioinformatics Open Days 2023

PROGRAM

MARCH 16TH - THURSDAY

15:15 H	**Oral Communications [Session 2]** • "Genomics and Systems Biology", chaired by Miguel Rocha Energy demand and enzyme budget trade-offs modulate the accuracy of dynamic Flux Balance Analysis - David Henriques MUG: a mutation overview of GPCR subfamily A17 receptors - Ana B. Caniceiro A comparative genomics approach to assess interspecific variability associated with cork development - Filippo Bergeretti Characterization of grapevine (Vitis vinifera) intra and inter-varietal diversity using whole genome resequencing - Sara Freitas
16:15 H	**Poster Communications [Session 1] + Coffee Break** • "Methods and applications", chaired by Andreia Salvador Posters n°. 5, 8, 14, 19, 20, 21, 25, 29, 30, 31, 33, 35 and 41. *The respective submissions can be checked in the Annex "Scientific Submissions".
17:15 H	**Bioinformatics Portuguese League Final Session**
20:00 H	Formal Dinner @ Taberna do Migaitas, Braga
22:30 H	Pub&Quiz

Bioinformatics Open Days 2023

PROGRAM

MARCH 17TH - FRIDAY

09:30 H	**Keynote Lecture** • The environmental impacts of bioinformatics: how bad is it and what can we do about it? Loïc Lannelongue
10:15 H	**Poster Highlight**
10:30 H	**Poster Communications [Session 2] + Coffee Break** • "Software and applications", chaired by Andreia Salvador Posters no. 2, 3, 6, 9, 13, 16, 17, 24, 26, 36, 37, 39 and 40. *The respective submissions can be checked in the Annex "Scientific Submissions".
11:30 H	**Oral Communications [Session 3]** • "Knowledge representation and machine learning", chaired by Óscar Dias Identification of gene regulatory modules acting in the interaction between cork development and environmental variable - Hugo Rodrigues The ImmunoPeptidomics Ontology: design and evaluation - Patrícia Eugénio Exploring self-attention mechanisms and deep reinforcement learning for the de novo drug design - Tiago Pereira ngest: A Scalable Snakemake Pipeline for Customized Knowledge Graph construction - Ana Santos-Pereira
13:00 H	**Lunch**
14:30 H	**Round Table + Network Session** Discussing the reality of the corporate world Silico Life, OmniumAI, Accenture
16:30 H	**Ending Session**
19:00 H	Informal BBQ Dinner @ Industrial, Braga

Bioinformatics Open Days 2023

PROGRAM

MARCH 18TH - SATURDAY

09:30 H **Workshops**

Introduction to Chemoinformatics using DeepMol
- João Capela and João Correia

The importance of machine learning in structural biology
- Irina Moreira, Luana Afonso and Catarina Marques-Pereira

Ontology Matching in the Biomedical Domain
- Marta Silva e Patrícia Eugénio

Genomic Data Infrastructure for Health Research Panel

Participants: Ana Teresa Freitas, Sérgio Sousa, and Maria Manuel Salazar

Moderation: Ana Portugal Melo

This panel will begin with a presentation by Kjell Petersen on the Genomic Data Infrastructure in Norway. The presentation will cover various topics, including technical, legal, and ethical issues, as well as data sharing. Ana Teresa Freitas will follow with the Portuguese counterpart. Finally, Sérgio Sousa and Maria Manuel Salazar will share their perspective on the need and major challenges to use the National and European Infrastructure for depositing and cross border sharing of genomic information towards advancing research and personalized medicine in Portugal and Europe. A discussion among the panelists with the participation of the audience will close this session.

Sponsored by BioData.pt, in the frame of Elixir Convergence and European Genomic Data Infrastructure projects. The project European Genomic Data Infrastructure in Portugal is being developed by INSA, BioData.pt, Instituto Superior Técnico, University of Aveiro and INESC-ID. It is built to support the One Million Genomes initiative that in Portugal is coordinated by INSA.

Kjell Petersen

Kjell Petersen is a distinguished researcher and the Group Leader of the Computational Biology Unit (CBU) at the University of Bergen in Norway. His primary research interest lies in gene set and network-based analysis approaches for omics datasets, and he has made significant contributions to gene expression profiling, especially in the context of Endometrial Cancer. Kjell also leads the CBU Service Group, which provides essential bioinformatics services to Norwegian researchers in molecular biology, biochemistry, biomedicine, and microbiology. Furthermore, he is heavily involved in e-infrastructure development and operation for life science research.

SCIENTIFIC SESSIONS AND KEYNOTE LECTURES

Multi-country outbreak of monkeypox virus: genomics as a tool to identify genetic clustering, transmission dynamics and the role of Portugal in the international virus spread

Mpox is a viral zoonosis caused by the mpox viruus, a member of the genus Orthopoxvirus. Mpox is endemic in West and Central Africa and has often been caused by spill-over events from small rodents and non-human primates to humans. However, it can also be transmitted from person-to-person through direct contact with lesions, body fluids and respiratory droplets. A large multi-country mpox outbreak has been ongoing worldwide since May 2022, with 85,922 cases and 96 deaths being reported in 110 countries by February 24th, 2023. On July 23rd, the WHO declared this outbreak a public health emergency of international concern. The dissemination has mostly affected men who have sex with men, who frequently displayed skin lesions in the anogenital area, suggesting transmission dissemination essentially through sexual networks. Portugal was one of the first countries to report mpox cases and the first to sequence the virus genome of the multicounty 2022 outbreak. By taking advantage of the Portuguese unprecedented sequencing rate (we sequenced the virus genome of nearly 60% of mpox cases) and the availability of patient-associated epidemiological data, we were able characterize the transmission dynamics in Portugal, as well as to understand the role our country in the mpox international dissemination.

João Paulo Gomes

Dr. João Paulo Gomes is a highly accomplished researcher who leads the Genomics and Bioinformatics Unit at the Department of Infectious Diseases of the Portuguese National Institute of Health, known as the Instituto Ricardo Jorge. With a PhD in Microbial Genomics, he has contributed significantly to the field of infectious disease research through his specialization in Microbial Genomics. Dr. Gomes is widely recognized for his pioneering work in using genomics to monitor important outbreaks, which has been instrumental in the fight against infectious diseases. As the PI of the Genomics and Bioinformatics Unit, Dr. Gomes continues to make important contributions to the field of Microbial Genomics.

Bioinformatics Open Days 2023

SCIENTIFIC SESSIONS AND KEYNOTE LECTURES

The environmental impacts of bioinformatics: how bad is it and what can we do about it?

The environmental impact of (scientific) computing, computational biology in particular, is a growing concern in light of the urgency of the climate crisis, and there is widespread interest in the research community; so what can we all do about it? Tackling this issue and making it easier for scientists to engage with sustainable computing is what motivated the Green Algorithms project. We will discuss what we learned along the way, how to estimate the impact of our work and dive into the carbon footprint of common bioinformatic analyses. We will also highlight the levers scientists and institutions have to make their research more sustainable, debate the ethical implications of these environmental costs and examine what is still needed moving forward.

Loïc Lannelongue

Loïc Lannelongue is a highly accomplished Research Associate in Biomedical Data Science at the University of Cambridge. He is a Software Sustainability Institute Fellow and holds several academic positions, including College Post-Doctoral Associate at Jesus College and Associate of the Senior Common Room at King's College, Cambridge. Loïc's research interests focus on the use of machine learning algorithms for patient care, particularly in the context of cardiovascular disease. He combines genetic information with medical imaging to better understand disease mechanisms and improve individual diagnoses. In addition, Loïc is leading the Green Algorithms project, which promotes more sustainable computational science.

ORAL COMMUNICATIONS

Session 1 - Structural Bioinformatics

A bioinformatics approach to study the permeation of solutes through PfAQP for the development of new antimalarial therapies

Marta S. P. Batista, Paula J. Costa, and Bruno L. Victor

BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal

Malaria is one of the largest public health problems. Although most variants are successfully treated with the existing antimalarial drugs, this disease is still responsible for a large number of global deaths. Severe malaria in humans is mostly caused by infection with Plasmodium (P.) falciparum whose complications include severe anemia, end-organ damage, pulmonary complications, and hypoglycemia1. The development of hybrid antimalarial agents has been pursued as a promising strategy to tackle resistant parasite strains, eliminating the actively-infecting P. falciparum organisms in human red blood cells, and also the replicative and dormant liver forms of the parasite. The aquaporin of P. falciparum (PfAQP) is a water and glycerol membrane protein channel, allowing the permeation of these molecules from the host to the parasite. The fast reproduction of P. falciparum in the host red blood cells requires massive biogenesis, in which glycerol is incorporated into the lipids of newly synthesized parasite membranes. Therefore, PfAQP is seen as a promising therapeutic target for the development of new antimalarial therapies. In this communication, we will present a recently developed bioinformatics approach focused on the identification of PfAQP structural features that regulate the permeation of molecules through its pores, to boost the development of a hybrid therapeutical agent that either blocks or transports currently available antimalarial drugs to P. falciparum medium. By using methods such as Molecular Dynamics, Umbrella Sampling, and Potential of Mean Force calculations we gathered relevant structural information regarding the permeation of water, glycerol, erythritol, and xylitol through PfAQP pores, which will allow us to identify and develop in the future a new antimalarial hybrid drug.

Acknowledgments: We acknowledge Fundação para a Ciência e Tecnologia (FCT) for funding through projects UIDB/04046/2020, UIDP/04046/2020, 2021.09731.CPCA, and BioISI Junior Program. FCT is also acknowledged for 2021.00381.CEECIND contract (PJC).

ORAL COMMUNICATIONS

Session 1 - Structural Bioinformatics

Computational study of promising smart metallodrug delivery systems

Inês D. S. Pires[1] , Tânia S. Morais[2] , and Miguel Machuqueiro[1]

1. BioISI - Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Ed. C8, Lisboa, Portugal.
2. Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal.

Cancer, especially breast cancer, has been rising to the top of the most prevalent and deadly diseases. The triple-negative (TN) subtype of breast cancer (BC) is associated with high aggressiveness and poor prognosis. Other subtypes currently employ targeted therapies, taking advantage of receptors like hormone (estrogen and progesterone) and human epidermal growth factor receptor 2. The TN subtype lacks their expression, thus, its treatment is still heavily reliant on chemotherapy, especially with cisplatin-like drugs which are known for lack of selectivity and tendency to develop multidrug resistance.

TM34 is a ruthenium-based compound that has been suggested to be a more efficient and selective therapy than cisplatin. TM34 derivatives have been under study in the last years, in an attempt to increase their selectivity but preserve their activity, by adding a pH-sensitive linker and a peptide sequence that is recognized by receptor proteins from the FGFR family (overexpressed in TNBC cancers). Once in the presence of the altered pH of the tumor micro-environment, the linker hydrolyses and releases the active species.

This work aimed to study the impact of different substituent groups on the active specie's biophysical profile. This included examining the interaction of several TM34 derivative compounds with a membrane model (POPC) and calculating their membrane crossing energy profiles that can be used to estimate the membrane permeability coefficients. We used Molecular Dynamics simulations coupled with an Umbrella-sampling scheme to obtain the potential of mean force profiles, which allowed the calculation of the membrane permeability using the inhomogeneous solubility-diffusion model.

Acknowledgments: We acknowledge Fundação para a Ciência e Tecnologia (FCT) for funding through projects UIDB/00100/2020 (CQE), UIDB/04046/2020 & UIDP/04046/2020 (BioISI), and PTDC/QUI-QIN/0146/2020. T.S. Morais and M. Machuqueiro thank the CEECIND 2017 Initiative for projects CEECIND/00630/2017 and CEECIND/02300/2017, respectively (acknowledging FCT, as well as POPH and FSE-European Social Fund).

ORAL COMMUNICATIONS

Session 1 - Structural Bioinformatics

Molecular Dynamics as an active resource on the development of new formulations for commercial use: The case of natural silicone alternatives

Tiago Ferreira[1,2,3], Ana Loureiro[1,2,3], Jennifer Noro[1,2,3], Artur Cavaco-Paulo[1,2,3] and Tarsila Castro[1,2]

1. CEB- Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal
2. LABBELS –Associate Laboratory, Braga/Guimarães, Portugal.
3. SOLFARCOS—Pharmaceutical and Cosmetic Solutions, 4710-053 Braga, Portugal

The world of cosmetics is an always-evolving field with constant developments on its formulation components. The current reality asks for an ever-increasing need for natural and sustainable replacements for synthetic compounds in all fields of modern consumer products. However, the research and development stages to find these alternatives can be an expensive, time-consuming, and often wasteful process that turns this task into a laborious process. These factors all support the need for a better way of reaching results through a more time and cost-effective standpoint. For this reason, Molecular Dynamics simulations present a suitable set of resources for the assessment of the molecular superstructure of natural silicone alternatives. The work at hand has not only developed a computational methodology to research formulation components' behaviour in solution but also showed how modern silicone alternatives' distribution and densities can behave accordingly to their counterparts using Molecular Dynamics simulations. The work was built on two systems, A and B, where the former is composed of one ester (Dipentaerythrityl Hexa C5 Acid Ester) and the latter by an ester combined with an alkane (Triheptanoin and C13-Isoalkane); all three molecules are commercial options widely used. All systems under study were submitted to a 3-step thermal regulation strategy. The systems have gone through initial simulation at 25°C and with heating to 70°C, then a temperature switch (25°C↔70°C), then a shock to 200°C and finally a Simulated Annealing protocol onto 250°C. Ultimately, all systems converged towards micelle-like structures. These results come to further ascertain the position of computational chemistry and Molecular Dynamics Simulations as an important part of R&D processes in modern science and investigation.

ORAL COMMUNICATIONS

Session 1 - Structural Bioinformatics

The impact of theSARS-CoV-2 Omicron variant on thereceptor binding domain conformational dynamics andinteractionwith human ACE2

Rita I. Teixeira[1], Mariana Valério[1], Luís Borges-Araújo[1,*], Manuel N. Melo[1,2,3], Cláudio M. Soares[1], João B. Vicente[1,*], Diana Lousa[1,*]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
2. iBB - Institute for Bioengineering and Biosciences, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
3. Molecular Microbiology and Structural Biochemistry, UMR5086 CNRS & University of Lyon

The COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus2(SARS-CoV-2), has led to over 6.6 million deaths worldwide, as of 9th January 2023. TheSARS-CoV-2 mechanism of transmission and infection involves the binding of the virus to the angiotensin-converting enzyme 2 (ACE2) host receptor through the Spike (S) protein receptor-binding domain (RBD). The RBD is a privileged target of our immunesystem and antiviral therapies. Thus, understanding the molecular details of the binding mode isv ery relevant. Additionally, the unbound RBD presents conformational changes in the receptor binding motif (RBM) region that are hypothesized to also affect receptor recognition. Multiple vaccines and new therapeutics against SARS-CoV-2 have been developed over the last few years. However, the emergence of variants of concern (VOC) poses a great challenge due to the loss of natural and vaccine immunity. The rise of the Omicron VOC raised considerable global concern due to the amount of S protein substitutions, fifteen of which are located in the RBD. Later on, the Omicron variant has been divided into seven lineages, of which BA.1, BA.2, and BA.5 became the most concerning ones due to their increased transmissibility. Here, we investigated the impact of the Omicron subvariants on the binding to the human ACE2 (hACE2), by performing microsecond atomistic molecular dynamics (MD) simulations of the Omicron RBDs bound to hACE2. Our analysis of the interface and structural dynamics of the Omicron RBD substitutions provided a detailed characterization of the binding mode and the identification of specific substitutions that may affect binding affinity via the establishment of new interprotein interactions. A complementary study of the effect of the Omicron subvariantson the RBD conformational dynamics was also performed by simulating the unbound Omicron RBDs. We observed that the Omicron subvariants impact the RBD conformational dynamics towards an efficient binding to hACE2 providing them fitness advantage.

ORAL COMMUNICATIONS

Session 2 - Genomics and Systems Biology

Energy demand and enzyme budget trade-offs modulate the accuracy of dynamic Flux Balance Analysis

David Henriques[1], Vítor Pereira[2], and Eva Balsa-Canto[1]

1. Biosystems and Bioprocess Engineering Group, IIM-CSIC, Vigo, Spain.
2. Centre of Biological Engineering, University of Minho, Braga, Portugal.

Genome-scale metabolic models (GEM) and flux balance analysis (FBA) are widely used to predict intracellular fluxes in yeast fermentation of wild-type phenotypes and knockout or under-expression mutants. GEMs collect the list of known metabolic reactions (typically associated with gene products) in a matrix format (the stoichiometric matrix). When the rates of production and consumption of extracellular metabolites are known, it is possible to formulate linear systems of equations which encode a space of feasible solutions of the internal fluxes. Given that the number of fluxes (variables) to estimate is generally greater than the number of metabolites (equations), the problem is undetermined.

One possibility to deal with underdetermination is to use a flux balance analysis (FBA) approach, that is, assume that cells behave optimally in some respects and solve an optimization problem. Unfortunately, the optimal of an FBA problem can (usually) be achieved by many different combinations of fluxes. A common approach to this is to use prior knowledge or biological intuitions in order to bias the fluxes towards more plausible solutions. Often, the underlying objective is the maximization of biomass and the minimization of total flux in the so-called parsimonious FBA (pFBA). Interestingly, despite the fact that S. cerevisiae is a workhorse of the biotechnology industry and arguably the best studied eukaryote model organism, the pFBA approach generally does not lead to intracellular solutions compatible with experimental measurements of intracellular fluxes with radiolabelled substrates.

In this work, we identify some of the issues that interfere with the reliable estimation of intracellular fluxes in nitrogen-limited fermentations of S. cerevisiae. To achieve this, we combined a dynamic model that details the growth, biomass composition, and production of the main extracellular metabolites with an enzyme-constrained metabolic network. Our results indicate that considering the proton pump activity associated with the uptake of ammonium ions addresses the previous limitations associated with central nitrogen metabolism. Furthermore, our results indicate that, while using pFBA or minimizing enzyme use, the choice of an excessively high GAM parameter in the biomass equation can bypass other plausible metabolic sinks of ATP, such as plasma membrane ATPase.

ORAL COMMUNICATIONS

Session 2 - Genomics and Systems Biology

MUG: a mutation overview of GPCR subfamily A17 receptors

Ana B. Caniceiro[1,2], Beatriz Bueschbell[1,3], António J. Preto[1,3], Carlos A. V.Barreto[1,3], and Irina S. Moreira[1,4]

1. CNC-Center for Neuroscience and Cell Biology, Center for Innovative Biomedicineand Biotechnology, University of Coimbra, Coimbra, Portugal
2. PhD in Bioscience, Department of Life Sciences, University of Coimbra, CalçadaMartim de Freitas, 3000-456 Coimbra, Portugal
3. PhD Programme in Experimental Biology and Biomedicine, Institute forInterdisciplinary Research (IIIUC), University of Coimbra, Casa Costa Alemão,3030-789 Coimbra, Portugal
4. Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, 3000-456 Coimbra, Portugal

More than 800 genes encode G-protein-coupled receptors (GPCRs), making them the largest family of membrane proteins. GPCRs mediate several signalling pathways through a general mechanism that involves the activation of a chain of events, leading to the release of molecules responsible for cytoplasmic action and further regulation. These physiological functions can be severely altered by mutations in GPCR genes. Besides many physiological functions, GPCRs have also been reported to be associated with various pathophysiological states and diseases. The root cause of such pathologies involving GPCRs is genetic errors, which alter the normal function of the receptor. A detailed classification of all known non-synonymous mutations in GPCRs would help to understand deregulation and guide appropriate therapy.

The GPCRs subfamily A17 (dopamine, serotonin, adrenergic, and trace amine receptors) is directly related to neurodegenerative diseases, and it is crucial to explore the known mutations in these systems and their impact on structure and function. A comprehensive and detailed computational framework-MUG (Mutations Understanding GPCRs (MUG), is presented here to illustrate the key reported mutations and their effect on receptors of the subfamily A17 of GPCRs. The types of mutations occurring overall and in the different families of subfamily A17 were explored, as well as their localization within the receptor and potential effects on receptor functionality. The mutated residues were further analyzed by considering their pathogenicity.

Based on this analysis, the MUG database, which is an interactive web application, is available for the management and visualization of this mutational dataset. This reveals a high diversity of mutations in the GPCR subfamily A17 structures, drawing attention to the considerable number of mutations in conserved residues and domains. Furthermore, pathogenic mutations are mostly found in regions critical to the structure and function of receptors. This interactive database will help to explore GPCR mutations, their influence, and their family wise and receptor-specific effects, constituting the first step in elucidating their structures and molecules at the atomic level.

ORAL COMMUNICATIONS

Session 2 - Genomics and Systems Biology

A comparative genomics approach to assess interspecific variability associated with cork development

Filippo Bergeretti, Pedro M. Barros , M. Margarida Oliveira

Genomics of Plant Stress Lab, Instituto De Tecnologia Química E Biológica António Xavier (ITQB NOVA), Av. da República, 2780-157 Oeiras, Portugal

Corresponding author: fbergeretti@itqb.unl.pt

Cork oak (Quercus suber) is a unique woody plant, member of Fagaceae family that is distributed in the west Mediterranean region. It is an emblematic resource due to cork production, having high economical, ecological and social significance in Portugal. Cork is a protective layer developed in the trunk of the tree derived from the activity of a meristematic layer named phellogen (cork cambium). While in most woody species the phellogen's life span is limited, in cork oak it remains active throughout the tree life cycle, being fully regenerated after harvest. Because the specific regulation of phellogen activity in Q. suber remains mostly unknown, the main goal of this project is to elucidate the evolution of the cork oak genome and the genetic elements regulating cork development. For that, we performed a phylogenetic orthology inference of all the protein sequences predicted in the genomes of multiple Quercus spp. (Q. suber, Q. robur, Q. lobata, Q. mongolica and Q. rubra), another Fagaceae (Castanea mollissima) and using Populus trichocarpa as an outgroup. A rooted species tree was obtained, representing the evolution process predicted for the Quercus genus, highlighting different gene duplication events, and variation in evolutionary rates among gene families. Finally, we merged the evolutionary information with functional annotation data. We found 1428 well-supported gene duplication events for the common ancestor of the Quercus genus, obtaining evolutionary evidence of gene expansion/contraction events for predicted gene families containing genes putatively involved in cork development. Among 1,786 transcription factors predicted in the Q. suber genome, 919 were determined as actively transcribed in cork tissues using available RNA-seq data, and 45 gene duplication sand 3 gene loss events were found among this group. As the biological relevance of orthogroups with specific variability found in Q. suber will be further explored, this comparative approach will be the basis for future gene regulatory network data integration and synteny analysis, to further look for inter and intraspecific features (gene content and order) in cork oak and related species genomes.

Characterization of grapevine (Vitis vinifera) intra and inter-varietal diversity using whole genome resequencing

Sara Freitas[1,2,3], **Antonio J. Muñoz-Pajares**[1,3,4], **David Azevedo-Silva**[1,2,3], **Mariana Sottomayor**[1,2,3], **Pedro Humberto Castro**[1,3], **João Tereso**[1,3,5,6], **Antero Martins**[7,8], **Elsa Gonçalves**[7,8], **Miguel Carneiro**[1,2,3], **Herlander Azevedo**[1,2,3]

1. CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal-
2. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto,4099-002 Porto, Portugal
3. BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
4. Departamento de Genética, Facultad de Ciencias, Universidad de Granada. Campus Fuentenueva, 18071, Granada, Spain.
5. MHNC-UP-Museum of Natural History and Science of the University of Porto-PO Herbarium, University of Porto, Praça Gomes Teixeira, 4099-002, Porto, Portugal
6. Centre for Archaeology, UNIARQ, School of Arts and Humanities, University of Lisbon, Portugal
7. LEAF-Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal
8. Portuguese Association for Grapevine Diversity-PORVID, Tapada da Ajuda, 1349-017 Lisboa, Portugal

Corresponding author: hazevedo@cibio.up.pt

Human activity led to the creation of thousands of grapevine (Vitisvinifera L.) varieties with extensive phenotypic diversity. Grapevine is a widely cultivated and economically significant crop. Unfortunately, it has been experiencing extensive genetic erosion, driven by favoring of specific varieties/clones, and the globalization-driven exposure to pathogens. Fighting this genetic erosion whilst addressing issues of resilience to climate change, yield and other traits, requires a crucial understanding of the genetic basis of grapevine phenotypic variation. Byt aking advantage of the extensive Portuguese reservoirs of grapevine varietal diversity, we are using Next Generation Sequencing (NGS) strategies to unlock the biodiversity and evolutionary relationships associated with various of those grapevine varieties. Here, we report on the use ofbioinformatic/NGS-enabled whole genome resequencing strategies, geared towards the characterization of extant biodiversity associated with grapevine germplasm.

Funding: Fundação para a Ciência e Tecnologia (FCT/MCTES) for project GrapeVision (PTDC/BIA-FBT/2389/2020) and support to H.A.(CEECIND/00399/2017/CP1423/CT0004); FCT/MCTES and POCH/NORTE2020/FSE for support to S.F. (SFRH/BD/120020/2016); FCT/MCTES and POPH-QREN/FSE for support to M.C. (CEECINST/00014/2018/CP1512/CT0002).

ORAL COMMUNICATIONS

Session 3 - Knowledge representation and machine learning

The ImmunoPeptidomics Ontology: design and evaluation

Patrícia Eugénio[1], Daniel Faria[2], and Catia Pesquita[1]

1. LASIGE, Faculdade de Ciências da Universidade de Lisboa, Portugal
2. INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

There is significant hardship in manipulating the complex and variable data generated by biomedical subdomains such as immunopeptidomics. Therefore, ontologies are often used to express knowledge in a domain to help establish a standard nomenclature to help handle and integrate medical data.

As a new field, immunopeptidomics lacks standardization: there is no recognized vocabulary or explicitly defined meanings. Hundreds of ontologies cover the biomedical domain, including neighboring subdomains like proteomics and immunology, but none adequately cover immunopeptidomics. This shortcoming needed to be addressed so cancer research in this domain could start to affect clinical practice.

This work details the development of the ImmunoPeptidomics Ontology, ImPO, to allow later integration of data produced by immunopeptidomics in personalized oncology. For this purpose, ImPO was designed following a process that comprised: capturing domain specialist knowledge in immunopeptidomics; iterative conceptual modeling of the domain through an Entity-Relationship model, semantic modeling by OWL formalization of the ER, cross-referencing ImPO with 28 external ontologies, and evaluation with competency questions and construction pitfalls.

Unlike most biomedical ontologies currently accessible, ImPO was created to be populated with data and function as the semantic backbone of a knowledge network (KG). ImPO was created as part of the KATY project, which aims to apply “AI-empowered knowledge” to clinical practice in Clear Cell Renal Cell Carcinoma. ImPO is one of the KATY KG components that will facilitate data integration in the project and give explainability to the AI techniques created in the project. Nonetheless, the ImPO ontology was created to be utilized independently of the KATY KG as a stand-alone knowledge model to aid in data integration and knowledge discovery in immunopeptidomics.

ORAL COMMUNICATIONS

Session 3 - Knowledge representation and machine learning

Exploring self-attention mechanisms and deep reinforcement learning for the de novo drug design

Tiago O. Pereira, Maryam Abbasi and Joel P. Arrais

1. Centre for Informatics and Systems, Department of Informatics Engineering, University of Coimbra

Over the last few years, deep learning (DL) methods have contributed to significant advances in drug development, namely for the targeted generation of novel compounds. In this work, we apply DL capabilities to streamline the generation of new compounds with optimized properties using Reinforcement Learning (RL) and self- attention mechanisms. The practical case addressed was the generation of a set of putative hit compounds against the ubiquitin-specific protein 7 (USP7) due to the importance of this enzyme for the proliferation of different types of tumours.

Typically, the application of RL implies training a molecular Generator and an Evaluator model to predict biological affinity against the desired target. The Evaluator assigns a reward to each molecule, and based on this value, the Generator updates its parameters to improve the optimization in subsequent iterations. Hence, the feedback that the Generator receives is not directly related to the individual choices of the atoms that comprise the molecule but rather to the molecule as a whole. This is a limitation of the Generator's learning process since it prevents the model from directly perceiving which choices it should or should not make when constructing molecules to better optimize the drug-like properties.

Herein, we propose a novel RL setting in which the Generator learns directly through its individual actions. In other words, we intend to move the evaluation of the actions from the level of molecules to the level of the tokens that constitute them. The main idea is to provide different rewards for different molecule regions for the Generator to learn how to generate compounds with the active sites typically involved in the interaction with the target. The framework comprises an RNN-based Generator and a pIC50 Predictor connected by a Transformer-encoder. The latter model will apply multi-head self-attention to the sampled molecules to extract the attention scores and the informative embedding vectors that characterize the molecules. By distributing the attention scores along the molecule, it will be possible to indicate to the Generator the different levels of importance of each molecular region. The results demonstrate that the Generator's learning process becomes more efficient by using the Transformer-encoder to assign individual rewards to tokens based on their attention scores. Hence, highlighting the tokens that promote higher or lower rewards makes the RL process more well-oriented and straightforward.

ORAL COMMUNICATIONS

Session 3 - Knowledge representation and machine learning

ngest: A Scalable Snakemake Pipeline for Customized Knowledge Graph construction

Ana Santos-Pereira[1,2], Joana Vilela[1,2], Ana Rita Marques[1,2], João Xavier Santos[1,2], Célia Rasga[1,2], Astrid Vicente[1,2], Hugo Martiniano[1,2]

1. Instituto Nacional de Saúde Doutor Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, Portugal
2. Faculdade de Ciências, BioISI - Biosystems & Integrative Sciences Institute, Universidade de Lisboa, Lisboa, Portugal

In the last few decades, massive amounts of biological data have become universally available, creating several opportunities for its analysis, which has been shown to be crucial for obtaining novel and valuable clinical insights. A growing trend for using data integration approaches as means for understanding complex and heterogeneous data, such as Knowledge Graphs (KGs), has been observed due to their ability to exploit the vast amount of information on interactions between chemical and biological entities. However, the dissemination of this information across distinctly organized databases and within the literature represents the biggest challenge when analyzing the available datasets. Thus, in this project, we aimed to build a software tool, the ngest, to automate the process of building biomedical KGs. Ngest integrates data from high quality curated databases such as Ensembl, STRINGDB, Gene Ontology and DisGenet taking advantage of the biolink model to standardize the categorization of different entities and relationships. Moreover, we include information regarding non-coding RNA (ncRNA) interactions with genes, proteins and other RNAs from NPInter and MirTarBase databases, which constitute a differentiation factor from other approaches that commonly target protein coding regions of the genome. Ngest extracts information from 14 datas ources, which results in 367.760 entities and 20.492.625 relationships. This scalable and flexible pipeline for automated download, data processing and production of the final KG was developed using Snakemake and includes a script for uploading the resulting KG toa neo4j graph database for visualization and querying. We envision the use of the KGs produced with ngest as a basis for the application of KG embedding approaches, aiming at contributing to better diagnosis or personalized treatment based on genetic, epigenomic and clinical information.

ImmspaCy: Extracting Gene-disease Associations for Systems Immunology Discoveries

Hayden So[1], Steven H Kleinstein[2], and Kei-Hoi Cheung [3,4]

1. UWC Red Cross Nordic, Flekke, Vestland, Norway.
2. Yale University School of Medicine, New Haven, CT.
3. Yale University School of Public Health, New Haven, CT.
4. Yale Center for Medical Informatics.

A human's adaptive immune responses to diseases result from the diversity of antibodies and rearrangements of immunoglobulin (IG) genes. We present immspaCy, a custom NLP pipeline in spaCy's python framework, for extracting orthographic variations of IG genes and disease entities from the growing quantity of immunology literature. As a result, we create a novel IG-specialized gene-disease association network, visualized in Cytoscape, allowing for downstream profiling of prominent genes and usage in gene set enrichment context. This project addresses the largest problem of ambiguity and diversity in biomedical naming conventions.

Leveraging spaCy v3's usability and components, a comprehensive list of 639 IG gene terms and 10862 human disease ontology terms were retrieved. Literature mining and extraction were done on relevant abstracts from PubMed. With a custom-built spaCy pipeline with enhanced tokenization, sentence segmentation, context awareness, and negation, complications in scientific biomedical text handling were addressed more appropriately. As the CNN NER works on generalization and having unseen data, the corpus of abstracts was split into training and evaluation data with an 80:20 split with cross-validation. It was trained by feeding the RegEx-annotated data which most closely represented gold standard data. The extracted entities are linked to a Cytoscape network for visualization.

While rule-based matching provided suitable results, the CNN NER model was able to identify edge cases comparable to manual annotation (such as HV1-69-sBnAbs). Most prominent variations included randomized orders and prefixes/suffixes. However, the further integration of custom contextual components led to a decrease in entities extracted. For example, a majority of IGHD cases represented isolated growth hormone deficiency instead of the IG-heavy genes and having negation cleared the false positives, making the model more precise. The downside is that while the rule-based extraction would pick up rare ruled variations, without enough annotations, the CNN-trained model would forget this (catastrophic forgetting).

When expanded to other immunological terminology and accessed by researchers, they now can trace V-gene usages and connect gene profiles to previously unknown conditions. The next step is to include AUROC performance, association scores, and relative weighting depending on the closeness of the terms.

Holistic Biomedical Knowledge Graph Integration

Marta Silva[1], Daniel Faria[2], and Catia Pesquita[1]

1 LASIGE, Faculdade de Ciências da Universidade de Lisboa, Portugal.
2 INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Data-centric approaches have taken the forefront in biomedical research, driven by the increasing availability of multi-domain biomedical data. There has also been an increased interest in adhering to the FAIR principles, and in order for data to be Interoperable and Reusable, it needs to be described objectively and in a structured manner within its context. This can be achieved by using ontologies, as they are structured representations of concepts and relationships that provide the scientific context onto which data is embedded. However, the necessary knowledge is often fragmented across hundreds of biomedical ontologies, and it is necessary to ensure interoperability between them to enable holistic research approaches (such as precision medicine or systems biology). Constructing a comprehensive and complete view of the biomedical domain requires multiple ontology matching.

The problem of performing ontology matching in multiple ontologies to create a single final alignment in a manner that is scalable, complete, and correct has not been properly addressed by the state-of-the-art, which has focused mainly on simple pairwise equivalencies.

This work explores holistic ontology matching and how can ontologies be considered collectively, complex ontology matching and how large language models and lexical approaches can find complex correspondences in the absence of shared instances, and interactive ontology matching and how to support complex interactions and human-in-the-loop strategies.

POSTER COMMUNICATIONS

Session 1 - Methods and Applications

Genome-wide transcriptomic analysis reveals novel genes involved in cynaropicrin synthesis in Cynara cardunculus

A. Paulino[1,2], R.C. Pires[1], I. Fernandes[2], A. Faustino[1,3], J. Santos[1,3], T. Brás[1,3], D. Rosa[1,3,4], O.S. Paulo[2], M.F. Duarte[1,3], and L. Marum[1,3]

1. Alentejo Biotechnology Center for Agriculture and Agro-food (CEBAL)/Instituto Politécnico de Beja (IPBeja), 7801-908 Beja, Portugal.
2. cE3c – Centre for Ecology, Evolution and Environmental Changes & CHANGE – Global Change and Sustainability Institute, Computational Biology and Population Genomics Group (CoBiG2), Campo Grande, 1749-016 Lisboa, Portugal
3. MED – Mediterranean Institute for Agriculture, Environment and Development & CHANGE – Global Change and Sustainability Institute, CEBAL, 7801-908, Beja, Portugal.
4. Allelopathy Group, Department of Organic Chemistry, INBIO Institute of Biomolecules, Campus de Excelencia Internacional Agroalimentario (ceiA3), University of Cádiz, 11510 Puerto Real, Cádiz, Spain.

Corresponding author: ana.paulino@cebal.pt

Cynara cardunculus L. (Cc), commonly named cardoon, is a Mediterranean plant from the Asteraceae family, which has gained growing interest for its natural source of sesquiterpene lactones (STLs), namely cynaropicrin (Cyn), the major SLT presented in Cc leaves. Portugal has a huge natural variability of Cc at morphological, genetic, and biochemical levels.

To identify the molecular mechanisms essential for STLs biosynthesis, transcriptomes of Cc genotypes were analyzed by comparing different biochemical profiles observed over 4 months (March – June) in regard to Cyn, collecting all samples when contrasting amounts of Cyn were achieved (May): (HP) samples with a peak of high production, (LP) samples with a peak of low production, (AH) samples with constant high levels and AL samples with constant low levels.

Total RNA was extracted from the Cc leaves for cDNA libraries synthesis. Stranded paired-end sequencing was performed by the DNBseq platform. The raw reads were pre-processed to remove low-quality reads and adaptors contamination yielding a set of high-quality reads which were then mapped with STAR v.2.7.19a against the assembly version 2.0 of the Cc genome from The Global Artichoke Genome Database. Unique mapped reads were used for differential gene expression analysis, which was performed using DESeq2. Differentially expressed genes were defined as genes with a log2 fold change (logFC) $\geq |2|$ and a False Discovery Rate (FDR) ≤ 0.05. The differential expression analyses yield a total of 36 and 212 DEGs when comparing HP vs LP and HP+AH vs LP+AL, respectively. In both comparisons, most of the DEGs were more expressed in samples with a high amount of Cyn (HPvsLP: 78%; HP+AH vs LP+AL 57%).

Several DEGs involved in stress and related to photosynthesis and cellular respiration were identified suggesting that environmental factors have an essential role in the regulation of Cyn production. Considering the potential biotechnological, agronomic, and pharmaceutic potential use of Cc plants, the relevance of this study is substantially high-level.

Genome-wide transcriptomic analysis reveals novel genes involved in cynaropicrin synthesis in Cynara cardunculus

A. Paulino[1,2], R.C. Pires[1], I. Fernandes[2], A. Faustino[1,3], J. Santos[1,3], T. Brás[1,3], D. Rosa[1,3,4], O.S. Paulo[2], M.F. Duarte[1,3], and L. Marum[1,3]

1. Alentejo Biotechnology Center for Agriculture and Agro-food (CEBAL)/Instituto Politécnico de Beja (IPBeja), 7801-908 Beja, Portugal.
2. cE3c – Centre for Ecology, Evolution and Environmental Changes & CHANGE – Global Change and Sustainability Institute, Computational Biology and Population Genomics Group (CoBiG2), Campo Grande, 1749-016 Lisboa, Portugal
3. MED – Mediterranean Institute for Agriculture, Environment and Development & CHANGE – Global Change and Sustainability Institute, CEBAL, 7801-908, Beja, Portugal.
4. Allelopathy Group, Department of Organic Chemistry, INBIO Institute of Biomolecules, Campus de Excelencia Internacional Agroalimentario (ceiA3), University of Cádiz, 11510 Puerto Real, Cádiz, Spain.

Corresponding author: ana.paulino@cebal.pt

Acknowledgments: This work is supported by Program Alentejo 2020, through the European Fund for Regional Development (FEDER) under the scope of MedCynaraBioTec – Selection of Cynara cardunculus genotypes for new biotechnological applications: the value chain improvement of cardoon, a well-adapted Mediterranean crop (ALT20-03-0145-FEDER-039495). Authors also acknowledge FCT for Contrato – Programa to L. Marum (CEECINST/00131/2018), PhD grant to A. Paulino (SFRH/BD/145383/2019) and D. Rosa (SFRH/BD/143845/2019), and Project UIDB/05183/2020 to Mediterranean Institute for Agriculture, Environment and Development (MED), and Project LA/P/0121/2020 to CHANGE – Global Change and Sustainability Institute.

POSTER COMMUNICATIONS

Session 1 - Methods and Applications

Metagenomic approach to identify genes encoding forglycosidehydrolases in composting samples

Joana Sousa[1], Cátia Santos-Pereira[1], Ângela M. A. Costa[1], Andréia O. Santos[1], Ricardo Franco-Duarte[2,3], Lígia R. Rodrigues[1] and Sara S. Silvério[1]

1. CEB - Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal
2. CBMA - Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.
3. IB-S - Institute of Science and Innovation for Bio-Sustainability, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.

Metagenomics involves the study of the genomic DNA from a set of microorganisms present in a particular environmental sample. This approach has emerged as a promising culture-independent technique to explore the diversity and function of microbiomes, allowing the discovery of novel biochemical compounds, namely enzymes with high potential for industrial applications. Composting habitats are characterized by a high microbial diversity, and represent a suitable source of robust enzymes able to convert the recalcitrant structure of lignocellulose, such as cellulases, endo-hemicellulases, oligosaccharide-degrading enzymes, and debranching enzymes. In fact, several lignocellulose-degrading enzymes have been successfully identified in composting samples following metagenomic approaches. The efficient handling, processing, and analysis of the large metagenomic datasets generated by next-generation sequencing platforms can be achieved using advanced bioinformatics pipelines.

In this work, composting samples were collected from three Portuguese composting units, which handle different types of wastes. The metagenomic DNA was extracted from the composting samples, the three composting metagenomes were analyzed by shotgun sequencing and a comparative analysis was performed between our samples and composting samples selected from the literature to evaluate the potential of these environments for lignocellulosic biomass conversion. The metagenomic sequencing data from all samples were processed using appropriate bioinformatics tools and the functional annotation of genes encoding glycoside hydrolases was carried out using the CAZy database. Our bioinformatics pipeline revealed that all samples were enriched in cellulases, endoglucanases, and β-glucosidases, which confirms the richness of composting habitats, regardless of waste compositions, in lignocellulose-degrading enzymes. As these compost samples were collected in the thermophilic phase, the identified enzymes may harbor interesting features for industrial purposes, including catalytic activity under high temperatures.

POSTER COMMUNICATIONS

Session 1 - Methods and Applications

Exploring Aveiro salterns to discover new and robust biosurfactant producers

Cátia Santos-Pereira[1], Joana Gomes[1], Joana Sousa[1], Marta Simões[2], André Antunes[2], Ricardo Franco-Duarte[3,4], Sara S. Silvério[1], Lígia R. Rodrigues[1]

1. Centre of Biological Engineering (CEB), University of Minho, Braga, Portugal
2. State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau SAR, China, Taipa, China
3. CBMA (Centre of Molecular and Environmental Biology), Department of Biology,University of Minho, Braga, Portugal
4. IB-S (Institute of Science and Innovation for Bio-Sustainability), University of Minho, Braga, Portugal

Nowadays, the awareness of the concepts of sustainability ,environmental impacts and green biotechnology has reached more and more importance. In this line, biotechnology techniques emerged as a powerful tool to find new and bio-sustainable molecules, such as biosurfactants, with interesting properties. Biosurfactants are versatile molecules produced by microorganism that can be applied in several fields, namely in cosmetics, pharmaceutical, detergents,food, textile or petroleum industries. Worldwide production of surfactants is expected to reach 65,570 million USD by 2029, however chemical surfactants exhibit toxicity and are not biodegradable. Thus, finding new and robust bio-based surfactants is of utmost importance. Saline and hypersaline environments, like those found inA veiro salterns, are potentially rich in unique microbial diversity that can be the source of interesting molecules, including biosurfactants.

In this study, samples from Aveiro salterns were collected and chemically characterized revealing interesting values of salinity, pH and heavy metal content. Culture-dependent techniques allowed the identification of several halophilic microorganisms with remarkable surfactant-like properties, which can have promising industrial applications. Genomic DNA was extracted from saltern water andampliconsequencingof16S rRNA gene was performed to proceed with taxonomic annotation and identification of the isolated microorganisms.

Integration of Multi-Omics Data for the Classification of Glioma Subtypes and Identification of Novel Biomarkers

Francisca G. Vieira[1], Regina Bispo[2], Marta B. Lopes[2,3]

1. NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal
2. NOVAMATH Center for Mathematics and Applications, Department of Mathematics, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal
3. NOVA Laboratory for Computer Science and Informatics (NOVA LINCS), NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal

Corresponding author: fmg.vieira@campus.fct.unl.pt

Glioma is currently one of the most prevalent types of primary brain cancer with increasing incidence worldwide and ascendent mortality rates. Given its high level of heterogeneity along with the complex biological molecular markers, many efforts have been made to classify the subtype of glioma in each patient, which, in turn, is critical to improve early diagnosis and survival. The fast-growing technological advances in high-throughput sequencing have enabled researchers to collect multiple omics data for the same individuals, typically leading to high-dimensional datasets. In addition, the integration of the different omics levels provides a window to build a more comprehensive landscape of biological processes. Several computational and statistical approaches, such as joint dimensionality reduction (JDR) and network-based techniques, have been proposed to tackle high-dimensional data and extract relevant features from multiview data. Within the JDR approach, supervised methods based on sparse canonical correlation analysis are explored in this study, to discriminate between glioma subtypes (glioblastoma, astrocytoma, and oligodendroglioma) while seeking for common information across different datatypes. Hence, using data from The Cancer Genome Atlas (TCGA), this research aims at (1) finding consistent patterns across datasets that vary between the different phenotypes, and (2) identifying molecular biomarkers signatures to each disease group.

Transcriptional characterization of cTfh cells in a viral infection at single-cell resolution

Diogo Fonseca[1,a], **Ruido Amaral Vieiraa**[1,a], **Saumya Kumarb**[1,b], **Luis Graca**[1,b]

1. Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Centro Académico de Medicina de Lisboa, Universidade de Lisboa, Lisboa, Portugal

a. Joint first authors

b. Joint senior authors

Respiratory viral infections are the world's most deadly communicable disease and the fourth most common cause of death at a global level. The emergence of COVID-19 has rushed the scientific community to understand the immune response to viral infections better. With this objective, we characterized in COVID-19 patients the transcriptional signature of circulating Tfh cells (cTfh)–the CD4+T cell subpopulation that drives B cell-affinity maturation leading to the production of the high-affinity antibodies that neutralize the virus.

T-cell subsets were sorted from fresh peripheral blood samples collected from COVID-19 patients and processed for single-cell RNA sequencing (10x Genomics). The datasets from the different patients were integrated, and the transcriptome was analyzed using the R package Seurat.

By analyzing the transcriptomeof~18,000 cTfh cells, we observed the segregation of the clusters essentially along two axes: the well-defined cTfh subtypes (cTfh1, cTfh2, and cTfh17 subsets) and cell-state specific phenotype, i.e., active or quiescent. We found that the cells in the active state (ICOS+andPD1+) are predominantly cTfh1 (CXCR3+), cells reported to be associated with immune responses to viral infections. Within the CXCR3+cells, we observed two distinct phenotypes: effector cTfh1 cells, and CD8-like (cytotoxic) cTfh cells. The significance of a cytotoxic population of cTfh1 cells in viral infections is a novel finding that requires further dissection.

In conclusion, using a single-cell transcriptomics approach, we demonstrated the emergence of cTfh1 cells in COVID-19 patients, with the specialization of some of those cells in cytotoxicity.

POSTER COMMUNICATIONS

Session 1 - Methods and Applications

A new GIMME-based compartmentalised algorithm for transcriptomics data integration

Diego Troitiño-Jordedo[1,3], Lucas Carvalho[2], David Henriques[1], Vítor Pereira[2], Miguel Rocha[2], Eva Balsa-Canto[1]

1. Biosystems and Bioprocess Engineering Group, IIM-CSIC, Vigo, Spain.
2. Centre of Biological Engineering, University of Minho, Braga, Portugal.
3. Applied Mathematics Department, University of Santiago de Compostela, Spain.

The integration of multi-omics data into genome-scale models (GEMS) has the potential to improve our knowledge of the relationships between genotypes, environment and phenotypes of microbial species.

In this context, the integration of transcriptomics data has received substantial attention. Most methods are based on the activation or deactivation of particular reactions attending to a given transcriptomics reference value. Gene Inactivity Moderated by Metabolism and Expression (GIMME), a software tool that systemizes the integration of transcriptomics data, uses a transcriptomic threshold value for the entire cell. However, the selection of this threshold comes with risks. This is the case, for example, when deactivating reactions that might condition the viability of the cell due to associated low transcript values; or reactions that are required to meet the production of a certain external metabolite.

In this work we address the issue. To do so, we designed a new heuristic within GIMME that enables the selection of different transcriptomics thresholds for different cellular compartments. These compartments might correspond to cellular organelles (for example, cytosol, mitochondria, etc.) or can be manually designed by the user.

We have tested this new approach to describe the metabolism of Saccharomyces cerevisiae under batch fermentation conditions with very satisfactory results. Specifically, we used the iMM904 metabolic reconstruction of the species, and compared the results achieved with the standard GIMME using a fixed transcriptomics threshold with the novel heuristic with compartmentalised transcriptomics thresholds. Models were constrained with glucose uptake data at an industrial setup. The standard approach results in unfeasibility while the novel GIMME implementation, proposed in this work, successfully recovered flux distributions.

The heuristic is implemented in MEWpy and facilitates the integration of transcriptomics thresholds by compartments, pathways, etc. It can be applied to other metabolic studies, and opens the opportunity to obtain more refined and realistic flux distributions.

Design, production and characterization of antiviral proteins targeting SARS-CoV-2

Susana Parreiras[1], Carlos H. Cruz[1], Mariana Valério[1], Pedro M. F. Sousa[2], Cláudio M. Soares[1], João B. Vicente[1], Diana Lousa[1]

1. ITQB NOVA, Instituto de Tecnologia Química e Biológica António Xavier, www.itqb.unl.pt
2. iBET, Instituto de Biologia Experimental e Tecnológica, www.ibet.pt

SARS-CoV-2 is the virus responsible for the COVID-19 pandemic, which has caused >600 million confirmed cases of infections and >6 million deaths (as of January 2023). Despite the vaccination efforts, with 13 billion vaccine doses administered (as of December 2022), there remains an urgent need to develop strategies to control infection and treat patients.

One of the proteins attached to the viral membrane is the spike (S) protein, that is primarily responsible for the virus' ability to enter host cells. It consists of two subunits: S1, containing a receptor-binding domain (RBD) responsible for binding to the host cell receptor, and S2, that facilitates the membrane fusion. This makes it one of the most promising therapeutic targets.

The aim of this work was to design and produce antiviral proteins that can prevent the interaction between the S protein and the host receptor, angiotensin converting enzyme- 2 (ACE2) protein, to block infection.

First, several antiviral proteins were computationally designed using the Rosetta program based on the interactions between ACE2 and the RBD. Next, molecular dynamics simulations (MD) of 1 μs x 5 replicates of three candidates, free in solution and in complex with the RBD, were performed to test their interaction with the RBD. This was followed by experimental validation that began with the expression and purification of the three candidates. After obtaining pure fractions, the secondary structure and thermostability of these proteins were tested by far-UV circular dichroism spectropolarimetry. Surface plasmon resonance was used to evaluate the affinity of each candidate for the RBD. Neutralization assays were performed to investigate the neutralization ability of the proteins.

The experimental results showed that one of the developed proteins is a promising therapeutic approach that will be further improved in the future.

POSTER COMMUNICATIONS

Session 1 - Methods and Applications

Comparative analysis of Human and Mouse oral fibroblasts by scRNA-seq

Alexandre Fernandes[1], Diana Pereira[2], Inês Sequeira[2], Márcia Barros[1]

1. LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
2. Institute of Dentistry, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, United Kingdom

Different tissues respond differently to damage, such as scar formation, pain level, and healing time. Oral mucosa is a tissue with a high capacity to heal scarless and faster when compared with other tissues like the skin. Differences in the healing process result from a conjunction of factors, such as structure, environment, keratinocyte proliferation, and subpopulations of the fibroblasts present in the tissue. Fibroblasts are responsible for maintaining the extracellular matrix. Thus, different gene expression levels between fibroblast subpopulations can reveal new treatments for the wound healing process. Next-generation sequencing (NGS) technologies such as Single Cell RNA sequencing allow us to study the heterogeneity of the tissues, helping to detect subpopulations of fibroblasts. This study will compare differences in oral fibroblasts between two species, Homo sapiens and Mus musculus, using bioinformatics tools to reveal the gene markers responsible for scarless and faster healing processes in this tissue. We will perform a scRNA-seq datasets analysis of the two species to achieve this aim using Scanpy, a machine learning algorithm for the Python environment. The datasets initially needed to be standardized by applying the same methodologies to allow the wanted comparison between species. The preliminary results of the analysis have revealed the existence of 5 initial clusters of fibroblast subpopulations in each dataset. Beyond this evidence, we noticed differences with the analysis performed by the mouse dataset authors.

This work was supported by FCT through the LASIGE Research Unit, ref. UIDB/00408/2020 and ref. UIDP/00408/2020.

RNA-Seq analysis of laser capture microdissected tissues of the cork oak

Rita Costa Pires[1], Ana Usié[1,2], Ana Ferro[1,2,a], Tiago Capote[1,2,a], Liliana Marum[1,2]

1. Alentejo Biotechnology Center for Agriculture and Agro-food (CEBAL) / Polytechnic Institute of Beja (IPBeja), 7801-908 Beja, Portugal
2. MED – Mediterranean Institute for Agriculture, Environment and Development & CHANGE – Global Change and Sustainability Institute, CEBAL – Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo, 7801-908, Beja, Portugal
a. Current Address: Center for Genomics and Systems Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

Cork oak is an evergreen broad-leaved tree widely distributed throughout the Mediterranean basin. It is the main cork-production species worldwide playing a significant economic, ecological and social role in particular in Portugal and Spain. Cork offers exceptional physical and chemical properties for industrial applications. The presence of lenticular channels in the cork, contributes to the cork porosity, one of the main parameters for the industrial quality evaluation of cork. The entire biological and molecular pathway of development of cork tissue and the origin of lenticels is still poorly understood.

The main goal was to study the transcriptome profiling of phellogen, xylem and lenticels isolated by Laser capture microdissection (LM), for a better comprehension of cork formation and quality. Thus, samples of different tissues were collected through LM according to Pires et al. (2022), for RNA extraction. Total RNA was sequenced on the Illumina HiSeq 2500 platform to produce stranded paired-end reads of 125 bp in length. Differential expression analyses between tissues was conducted with EdgeR (R v.4.2.0) yielding a universe of 1,248 differentially expressed genes (log fold Change (logFC) ≥|2| and False Discovery Rate (FDR) ≤ 0.05). These genes were found differentially expressed in at least one pairwise tissue comparison. Genes involved in cell wall formation, defense and stress-related mechanisms, and in activation of plant growth regulators signaling pathways (brassinosteroids, ethylene and jasmonic acid) were identified as differentially expressed between tissues. This study helps to understand the molecular mechanisms associated with the development of phellem and xylem, and also to identify specific candidate genes linked to lenticels formation.

Acknowledgments: This work was supported by Program Alentejo 2020 under the scope of Lentidev—A molecular approach to cork porosity (ALT20-03-0145-FEDER-000020), and by Program PORTUGAL 2020 Partnership Agreement, under the scope of Biodata.pt– Infraestrutura Portuguesa de Dados Biológicos (22231/01/SAICT/2016), through the European Regional Development Fund (ERDF). Authors also acknowledge FCT for funding researchers through Contrato–Programa: L. Marum (CEECINST/00131/2018) and A. Usié (CEECINST/00100/2021), and for the financial support to Research Unit UIDB/05183/2020 (MED - Mediterranean Institute for Agriculture, Environment and Development).

A user-friendly bioinformatics pipeline for metaproteogenomics data analysis

João C. Sequeira[1,2], Vítor Pereira[1,2], M. Madalena Alves[1,2], Miguel Rocha[1,2], and Andreia F. Salvador[1,2]

1. CEB-Centre of Biological Engineering, University of Minho, Braga, Portugal
2. LABBELS – Associate Laboratory, Braga/Guimarães, Portugal

Metaproteogenomics integrates the analysis of genomics and proteomics data obtained from microbial communities. Meta-Omics Software for Community Analysis (MOSCA) is a command-line pipeline developed to perform the combined analysis of Metagenomics (MG) and Metatranscriptomics (MT) data. MOSCA was built as a Snakemake workflow to ensure reproducibility of its results, connecting many state-of-the art tools for omics-analysis. These tools are either directly run through Snakemake or integrated in Python and R scripts. In this work, we present the incorporation of metaproteogenomics analysis in MOSCA, and the novel web graphical user interface (GUI) of MOSCA, called MOSca's GUI TO perform meta-omics analyses (MOSGUITO), among other upgrades to MOSCA first version.

Major upgrades include: 1) the addition of a binning step, useful for obtaining Metagenome Assembled-Genomes (MAGs); 2) the option of performing the assembly with MT sequencing reads by using Trinity, which is essential if MG data is not available; 3) the incorporation of complementary functional annotation tools, UPIMAPI and reCOGnizer, which employ different algorithms for protein annotation, based on whole-sequence and domain-based homology, respectively; 4) the inclusion of KEGGCharter, a tool that represents MOSCA results in KEGG metabolic maps, showing either the genomic potential or the differential gene/protein expression of the entire microbial community.
Regarding the metaproteogenomics workflow, MOSCA builds a reference protein database from the genes identified in the MG data for the Peptide-to-Spectrum matching step. Three different search engines available through SearchCLI are used, namely X! Tandem, MyriMatch, and MS-GF+, to match tandem mass spectra with peptide sequences. Protein inference, identification and quantification is performed with PeptideShaker.

The intuitive web interface MOSGUITO was created to facilitate the utilization of MOSCA, by not requiring a personal device to run the pipeline. MOSGUITO also promotes the interactive navigation of MOSCA's outputs, which can be accessed from the same interface where the configuration and execution of MOSCA happen.

MOSCA can be installed from bioconda, and MOSGUITO can be accessed at (https://iquasere.github.io/MOSGUITO/).

Development of a bioinformatic workflow to evaluate tyrosine kinase inhibitor derivatives with improved membrane permeabilities

Rita F. C. C. Guerra, Nuno F. B. Oliveira, Pedro M. S. Suzano, Bruno L. Victor, and Miguel Machuqueiro

BioISI - Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Ed. C8, Lisboa, Portugal.

Despite tumor multi-drug resistance (MDR) being multi-factorial in nature, it has been proposed that the acidity in the lumen of lysosomes (pH~4.5–5) and tumor microenvironment (pH~6.2–6.8) play a significant role hindering the anti-cancer activities of hydrophobic weak base drugs (Lewis bases; pKa~7.5–9.5), by efficiently entrapping/excluding them, via protonation events. Some of these compounds, the tyrosine kinase inhibitors (TKI), exhibit high and complementary clinical relevance by being vital mediators of signal transduction and cancer cell proliferation, angiogenesis, and apoptosis. We have developed a computational strategy to chemically modify this class of TKI molecules and exchange the cationic amino groups with anionic ones. The rationale is that the anionic group should also have good solubility in the aqueous media and, in contrast to the weak base, have its membrane permeability increased with acidity. These acidic derivatives should selectively target cancer cells over normal tissues and effectively evade lysosomal sequestration, circumventing several crucial factors related to MDR.

In this study, we are optimizing a molecular docking protocol based on different search methods and scoring functions to study systematically the impact of replacing cationic groups found on TKIs with negative chemical building blocks on the binding to RTKs. This chemical modification strategy will allow us to simultaneously improve the druggability of such compounds, and evaluate the impact on the affinity to their therapeutical targets. In the future, we will use a consensus docking approach where the score and/or rank of various freely available docking suits, including Autodock 4.2, Autodock-GPU, Autodock Vina 1.2, and Dock 6.9, will be combined to achieve the best results.

Acknowledgments: We acknowledge Fundação para a Ciência e Tecnologia (FCT) for funding through projects UIDB/04046/2020, UIDP/04046/2020, 2021.09731.CPCA, and grants CEECIND/02300/2017 and 2021.06409.BD

Machine learning-based assessment of Multi-omics Integration Tools

Mohamed Emam[1,2,3], Ahmed Tarek[2,3], Mohamed El hadidi[2], Mohamed Hamed[3], and Agostinho Antunes[3]

1. CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208 Porto, Portugal.
2. Bioinformatics group, Center for Informatics Sciences (CIS), Nile University, Giza, Egypt
3. Institute for Biostatistics and Informatics in Medicine and Ageing Research (IBIMA), Rostock University Medical Center, Rostock, German

Multi-omics data integration enhances our understanding of biological systems and their underlying mechanisms. The advent of high-throughput technology and the increasing availability of multi-omics data has led to the development of several statistics-based integration methods. However, the performance of these methods is variable and there is a need to assess the performance of the selected features output of the tools in an unbiased manner. Here, we have performed a rigorous assessment of three representative multi-omics integration tools MOFA+, GFA, and ICluster using two complementary benchmarks. First, we assessed how well the features selected by each tool could discriminate between patient and control samples using both linear and non- linear classification models. Secondly, we quantified how much each type of omics data selected features contributed to the total variance. Through such detailed comparisons, we observed that the features selected by MOFA+ and GFA gave the best F1 score (0.7) in the nonlinear classification model which discriminates between patient and control classes. Hence, we recommend these two tools as unsupervised integration tools for feature selection purposes. Our analyses were conducted on a real biological dataset to further study prediabetes patients. We also take advantage of the multi-omics data to detect subtypes of prediabetes and provided several clinical insights which will open a new gate toward the era of personalized medicine for diabetic disease. The next step in this research is to develop our in-house deep learning-based multi-omics integration , which can capture non-linear signals, discreet data variance, and mixed factors.

BIT by BIT: Associating Transcription to Ageing

Ian Teixeira[1,2], Ana Rita Grosso[2], and Sérgio F. de Almeida[1]

1. Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina da Universidade de Lisboa, 1649-028 Lisboa, Portugal
2. UCIBIO, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

A common and longstanding view of the ageing process poses that the onset of ageing hallmarks is driven by the gain of somatic mutations deriving from the inaccurate repair of DNA lesions, the most catastrophic of which are DNA double-strand breaks (DSBs). However, compelling evidence accumulated during the last decade demonstrated that, in addition to causing mutations, DSBs might initiate gene expression events. Using live-cell imaging, it was recently provided direct evidence of transcription initiation at DSB sites, a process that we dubbed DNA Break-Induced Transcription (BIT). A trademark of the ageing cell is the widespread rewiring of gene expression, leading us propose the disruptive hypothesis that alternative gene expression events established through BIT at DSBs contribute to the establishment of ageing hallmarks. By using data from genome wide detection techniques of DSBs (BLESS, BLISS DSBCapture) and high-throughput sequencing (RNA-seq), whose cells underwent no treatment, it was possible to identify natural BIT events at the whole genome level.

Development of comprehensive methodologies for bacteriophage-host studies

Maria Vieira[1,2], José Duarte[1,2], Oscar Dias[1,2], Miguel Rocha[1,2], Joana Azeredo[1,2] and Hugo Oliveira[1,2]

1. Centre of Biological Engineering, School of Engineering, University of Minho, Braga 4704-553, Portugal
2. LABBELS – Associate Laboratory, Braga/Guimarães, Portugal

The emergence of antibiotic resistance has become a problem in the medical field, driving the search for new therapies that can fight bacterial infections. Bacteriophages, also known as phages, play an important role in this research due to their ability to infect and kill bacteria. These organisms can encode enzymes capable of degrading polysaccharides present on the bacterial cell surface. These enzymes are usually called depolymerases, and they may have specificity to recognize and degrade some capsular types of bacteria.

As the study of these proteins is still scarce, a machine learning tool, called PhageDPO, was developed and is capable of predicting depolymerases based on bacteriophage genomes using machine learning methods. This tool was integrated into Galaxy framework available online at: bit.ly/phagedpo.

In the scope of FITTED project, a comprehensive database of bacteriophage-host interactions, called PhageKDB, was also developed. The database includes information about species, phages (name, accession number) their respective depolymerases (topology, accession number) and the consequent capsular type that depolymerase can degrade. This database was created using a combination of manual curation and automated data mining techniques. The user-friendly interface allows for easy searching and browsing of the data. The database will be a valuable resource for researchers studying bacteriophages and their interactions with hosts, as well as for those interested in potential applications of bacteriophages in biotechnology and medicine. This database was integrated into the FITTED project website available at fitted.ceb.uminho.pt.

Acknowledgments: This study was supported by the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding of UIDB/04469/2020 unit and “la Caixa” Foundation and FCT under the grant agreement HR21-FCT-00533.

Prediction of plastic-degrading enzymes in omics data from marine environments

José P. Freitas[1,2], João C. Sequeira[1,2], Diogo Cachetas[1,2], Miguel Rocha[1,2], Andreia F. Salvador[1,2]

1 Centre of Biological Engineering, School of Engineering, University of Minho, Braga 4704-553, Portugal

2 LABBELS – Associate Laboratory, Braga/Guimarães, Portugal

Plastic pollution is an environmental problem with inestimable consequences to public health. Due to the recalcitrant nature of synthetic plastics, they are the most difficult to suffer biodegradation, thus accumulating and disrupting ecosystems, having a massive impact on marine life. However, there are microorganisms which synthesize enzymes with the ability to biodegrade plastics, including the most recalcitrant, like polyethylene terephthalate (PET) and polyethylene (PE). Knowing these enzymes, it is possible to search for other proteins with similar metabolic activity in ecosystems contaminated with plastics, as it is the case of marine environments. Identifying the microorganisms and respective enzymes with such capability, strategies of in-situ or ex-situ biodegradation can be developed, to help fighting plastic pollution. Mining metagenomes from environmental samples have been shown to be an efficient strategy to find novel plastic-degrading enzymes. However, there are still no reported automated bioinformatics tools designed for that purpose.

In this work, a bioinformatics tool was developed to predict plastic-degrading enzymes in omics data, with the aim of finding new plastic-degrading enzymes. This tool receives two different inputs: 1) reference protein sequences corresponding to enzymes with plastic-degrading activity, there are used to create Hidden Markov Models (HMMs), and 2) omics datasets (protein FASTA files) as sink of proteins similar to those used for the HMMs construction. The tools output the proteins in the omics datasets that match the reference proteins.

This tool uses as default an HMM database for PET-degrading enzymes. The constructed HMMs are validated using the “leave-one-out” cross validation method. Alternatively, this tool can be applied to case studies different from plastics biodegradation, once the HMMs may be constructed using other reference enzymes.

The bioinformatics tool is available at bioconda (https://anaconda.org/bioconda/m-party).

This tool will contribute to unveil putative enzymes in environmental metagenomes that might be involved in plastics biodegradation. Those enzymes can then be tested in laboratory in order to confirm their activity towards plastic polymers.

POSTER COMMUNICATIONS

Session 2 - Software

Improving phenylalanine hydroxylase catalytic activity for tryptophan using an automated platform of structure-based enzyme engineering

Caio S. Souza[1], João Correia[1], Miglė A. Bartels[2], Sofia Ferreira[1], Diana Lousa[1], Isabel Rocha[1] and Cláudio M. Soares[1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal
2. Universidade Livre de Berlim, Kaiserswerther Str. 16-18, 14195 Berlin, Alemanha

Poor expression, low catalytic levels, low concentrations of co-factors or substrates, toxicity towards the final product and other factors, all contribute to the low performance of microorganisms using natural or novel pathways. Protein engineering is an effective strategy for improving metabolic pathways and overcoming these problems by redesigning an enzyme's catalytic properties in favour of a certain reaction.

To address this problem, we developed an automated platform for enzyme engineering, which aims to enhance the enzymes that catalyse the limiting steps in the targeted pathways. The objectives ranged from increasing an enzyme's efficiency to enabling it to catalyse a different transformation than the one catalysed by the wild-type sequence. The steps of enzyme engineering are divided into three major modules: a random mutation generator, an atomistic homology modeler and a binding energy evaluator. This method may therefore build a set of mutant enzymes and filter which sequences are putative candidates for experimental expression and enzymatic testing.

Phenylalanine hydroxylase (PAH) is an enzyme that catalyses the hydroxylation of the aromatic side-chain of phenylalanine to generate tyrosine. With the aim of increasing PAH affinity to tryptophan (Trp) (instead of phenylalanine) to produce 5-hydroxytryptophan (5HTP), an enzyme PAH was engineered with the automated platform. After a manual inspection of platform's best results, mutants E219S, F197Y and the triple mutant F197Y_S215N_E219Q, were the ones chosen for experimental testing. In vivo results indicate that all the mutants were able to produce 5HTP, with mutant F197Y_S215N_E219Q producing the maximum yield, reaching a maximum concentration of 71.5 mg/L after 144 hours. This shows that these mutations had an impact on the enzyme's affinity for Trp as a substrate and is able to convert it into 5HTP, showcasing the potential of enzyme engineering approaches.

POSTER COMMUNICATIONS

Session 2 - Software

Workflow for identification of InDel markers for rice (Oryza sativa) varietal discrimination

M. Beatriz Vieira[1], Pedro M. Barros[1], Tiago F. Lourenço[1] and M. Margarida Oliveira[1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal

Rice is the staple food for over half of the world's population. It has rich genetic diversity and a vast number of certified varieties, ranging in economic value and quality. As such, rice is highly prone to adulteration, especially fraudulent variety claims. Therefore, developing methods for fraud detection is of utmost importance. Molecular markers-based methods, particularly centred on polymerase chain reactions (PCR), are considered efficient and relatively inexpensive. Thus, we are developing a DNA-based method for the discrimination of varieties that circulate in the European market and Egypt, and for the identification of undesirable mixtures. We started by sequencing and mapping the genome of 20 varieties and used two additional ones previously sequenced. Using these 22 genomes, we followed the proposed GATK workflow for short variants calling and obtained an array of single nucleotide polymorphisms (SNPs), and insertions and deletions (InDels). By focusing on the InDel markers, we applied additional filtering to select those bigger than 15 bps (52,174 InDels). To further optimize the selection of the markers, we implemented a method previously described for SNPs' selection, the Conditional-Random-Selecting (CRS) method. The CRS method randomly selects a group of InDels and, by eliminating redundant ones, results in combinations of the least amount of InDels necessary to discriminate all the selected varieties. We obtained groups of six and seven markers that we are experimentally validating, by testing primers and optimizing PCR. Moreover, we aim to select InDels that may be associated with traits of interest, namely grain quality. Therefore, we also used CRS to select InDels only present in quantitative trait loci (QTLs) associated with seed and eating quality traits and we are increasing these QTLs of interest. Work is ongoing to validate and optimize the method of InDels' selection and PCR amplification, for further development of multiplex-PCR and capillary electrophoresis. Ultimately, the method should also be extended and tested to discriminate and identify more varieties.

Acknowledgements: H2020 TRACE-RICE Grant n.1934 (call-2019, Agrofood) PRIMA Programme, and FCT-Portugal through CEECIND/03641/2017 (TFL) ,and DL57/2016/CP1369/CT0029 (PMB).

Bioinformatics characterization of grapevine inter-varietal diversity using WGS data

João Nunes[1,2,3], João Monteiro [1,3,4], João Tereso [1,3,5,6], Miguel Carneiro[1,2,3], Sara Freitas[1,2,3] and Herlander Azevedo[1,2,3,*]

1. CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.
2. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto, Portugal.
3. BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal.
4. Departamento de Informática, Escola de Engenharia, Universidade do Minho, 4710-057 Braga, Portugal.
5. MHNC-UP - Museum of Natural History and Science of the University of Porto - PO Herbarium, University of Porto, Praça Gomes Teixeira, 4099-002, Porto, Portugal.
6. Centre for Archaeology, UNIARQ, School of Arts and Humanities, University of Lisbon, Portugal.

*__Corresponding author:__ hazevedo@cibio.up.pt

Grapevine (Vitis vinifera) is a fruit crop belonging to the Vitaceae family. In Portugal, the grapevine is among the three most grown crops. Its fruit (grape) is mostly used for winemaking but it can also be used for table consumption. There are thousands of grape varieties in Europe, with distinct levels of inter- and intra-varietal diversity. Such diversity results from a complex domestication history over multiple historical periods. To elucidate different properties of its recent evolutionary history, we resorted to whole genome resequencing of individual clones using Illumina sequencing data. Our results supported a model in which a central domestication event in grapevine was followed by post-domestication hybridization with local wild genotypes, leading to the presence of an introgression signature in modern wine varieties across Western Europe. To further characterize and distinguish the varieties suffering introgression with local wild varieties we are extending our genomics and bioinformatics characterization. First, we carried out a survey for publicly available whole genome sequencing (WGS) efforts, compiled all the metadata available regarding these sequences, eliminated redundancies and merged the information from similar categories. Then we developed and optimized the pipeline for trimming and mapping against the newest grapevine reference genome (PN40024v.4), implementing checkpoint statistics between steps to ensure the reliability of the data. At this stage we plan to continuously increase the dataset and initiate the population structure analysis using probabilistic methods, geared towards PCA, phylogeny and structure analysis.

Funding: Fundação para a Ciência e Tecnologia (FCT/MCTES) for project GrapeVision (PTDC/BIA-FBT/2389/2020) and support to H.A.(CEECIND/00399/2017/CP1423/CT0004); FCT/MCTES and POCH/NORTE2020/FSE for support to S.F. (SFRH/BD/120020/2016); FCT/MCTES and POPH-QREN/FSE for support to M.C. (CEECINST/00014/2018/CP1512/CT0002).

Bioinformatic characterization of genetic variation in grapevine using PoolSeq data

João Monteiro[1,2,3,4], João Nunes[1,2,5], Antero Martins[6,7], Elsa Gonçalves[6,7], Miguel Carneiro[1,2,5], Oscar Dias[3,4], Sara Freitas[1,2,5] and Herlander Azevedo[1,2,5,*]

1. CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.
2. BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal.
3. Centro de Engenharia Biológica, Universidade do Minho, 4710-057 Braga, Portugal.
4. LABBELS Laboratório Associado, Guimarães, Braga, Portugal.
5. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto, Portugal.
6. LEAF- Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal.
7. Portuguese Association for Grapevine Diversity-PORVID, Tapada da Ajuda, 1349-017 Lisboa, Portugal.

*__Corresponding author:__ hazevedo@cibio.up.pt

Vitis vinifera L. (the common European grapevine) is one of the most cultivated fruit plants and an economically important crop. It is essential to have an understanding of the genetic and molecular mechanisms underpinning crop traits such as yield. Nowadays, most cultivated grapevines are clonally propagated. Therefore, the expected genetic variability is low within a given variety. However, as genomic research advances, more refined technologies are being used to assess the relative contribution of genome-wide polymorphisms within a variety of evolutionary forces. One of these approaches is the genomic sequencing of pools of individuals using short-read sequencing technology (PoolSeq). With PoolSeq it is possible to study information regarding the population's polymorphic sites and corresponding allele frequencies of variants. In this project, various datasets of PoolSeq resequencing of important Portuguese grapevine clonal diversity will be studied. The main goal is to use these datasets to interrogate for genetic variation of different grapevine cultivars, quantify diversity and genetic differentiation, and detect selection signatures. Given the large potential, yet also the specificity of the PoolSeq strategy, a range of dedicated bioinformatics tools will be used for the comparison of results and summarization of their overall performance.

Funding: Fundação para a Ciência e Tecnologia (FCT/MCTES) for project GrapeVision (PTDC/BIA-FBT/2389/2020) and support to H.A.(CEECIND/00399/2017/CP1423/CT0004); FCT/MCTES and POCH/NORTE2020/FSE for support to S.F. (SFRH/BD/120020/2016); FCT/MCTES and POPH-QREN/FSE for support to M.C. (CEECINST/00014/2018/CP1512/CT0002).

POSTER COMMUNICATIONS

Session 2 - Software

DeepMol: a python-based machine and deep learning framework for drug discovery

João Correia[1,2], João Capela[1,2], Vítor Pereira[1,2] and Miguel Rocha[1,2]

1. Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710–057 Braga, Portugal.
2. LABBELS Laboratório Associado, Guimarães, Braga, Portugal.

Artificial intelligence has demonstrated outstanding results in many research areas, such as computer vision and natural language processing. In recent decades, there has been a growing interest in using machine and deep learning techniques in other research fields, including in chemoinformatics for molecular property prediction. This has led to the development of numerous algorithmic solutions proposed for various steps in the molecular property prediction pipeline. However, the wide range of approaches and methods available can be challenging for those without a strong computational background to navigate. Additionally, the lack of domain expertise, established data processing guidelines, and standard benchmarks can make it challenging to effectively apply artificial intelligence to drug discovery. To address these issues, a python-based framework called DeepMol was created to simplify the creation of machine and deep learning pipelines applied to chemoinformatics and make it more accessible to a wider audience.

DeepMol offers an intuitive interface and a comprehensive collection of methods for molecular property prediction, making it an useful tool for researchers and practitioners in the field of drug discovery. It covers a wide range of important steps, including standardization of molecules, dealing with unbalanced data, feature generation, feature selection, unsupervised learning, data splitting based on molecular information, model construction, hyperparameter optimization, and feature explainability. DeepMol relies on Tensorflow (https://www.tensorflow.org/), Keras (https://keras.io/), Scikit-Learn (https://scikit-learn.org/), and DeepChem (https://deepchem.io/) for the model construction allowing for the implementation of custom models or the use of pre-built ones. For operations on molecular data, DeepMol uses the RDKit framework (https://www.rdkit.org/).

DeepMol has already been used in many publications, including for evaluating molecular representations in machine learning models for drug response prediction and interpretability, and for predicting relationships between chemical structures and sweetness. DeepMol was developed under the BSD-2-Clause License and is available at https://github.com/BioSystemsUM/DeepMol.

POSTER COMMUNICATIONS

Session 2 - Software

Bioinformatic approaches to address new perspectives on genotype-phenotype associations for complex diseases

Daniel Martins [1,2], **Conceição Egas** [2] **and Joel Arrais** [1]

1. CISUC - Centre for Informatics and Systems of the University of Coimbra. PoloII, Pinhal de Marrocos, 3030-290 Coimbra, Portugal.
2. CIBB - Centre for Innovative Biomedicine and Biotechnology (UC- Biotech). ParqueTecnológico de Cantanhede, Núcleo 04, Lote 8, 3060-197 Cantanhede, Portugal.

Complex diseases are a major matter of interest on biomedicine research. Althoughfor the majority of those diseases its pathophysiology is largely understood andtheenvironmental risk factors are well reported, their genetic causes are usually unknownto a large extent.

Schizophrenia (SCZ) is a complex diseases with severe long-term implications and treated with a combination of medicine and individually-tailored therapy. It is estimatedto affect 1 in 300 people worldwide. The ability to predict more accurately the riskof manifesting this condition would largely improve its prevention.

Although several loci have been associated to the disease, they are usually foundonGWAS studies which assume independent effects of variants on the manifestationof aphenotype. However, with the crescent application of Machine Learning methods onthebiomedical scope and the uprising of Deep Learning, it is now possible to studycombined effects of several genes.

GenNet is a Deep Learning framework that replaces Fully-Connected Neural Networks with structures that reflect genetic and biological connections provided bytheuser, as variant-gene and gene-pathway annotations. We were able to access a Swedishpopulation-based case-control WES dataset on Schizophrenia, stored on dbGaPunder the accession phs000473.v2.p2. We filtered the dataset, keeping the variants witha significant direct association to the disease after a chi-square test without correction. This new dataset was provided as input to the GenNet framework.

With this approach, we reached similar results as reported on the literature andour results reveal several genes of interest. Both genes not previously linked with SCZ, asCLIP2 and MAIP1, and genes with recent associations to the disease as EGFR.

Funding: FCT - Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUCR&D Unit - UIDB/00326/2020 and the PhD Scholarship SFRH/BD/146094/2019.

Bioinformatics Open Days 2023

BIOINFORMATICS PORTUGUESE LEAGUE

Final Session

The "Bioinformatics Portuguese League" (Liga Portuguesa de Bioinformática - LPB) is a national contest on the scope of Bioinformatics and Computational Biology. The LPB has been created for higher education students with interest on those areas of knowledge and the national Bioinformatics community. It promotes and benefits the integration of future researchers and professionals on Bioinformatics and encourages active learning and improvement of skills on that field. The LPB was inspired on its Brazilian counterpart (LBB) that, at the time, had already completed two editions.

The contest was divided into 3 phases. The first phase consisted on a set of 30 questions on the subjects of Biology, Computer Science and Bioinformatics. The second phase consisted on 5 Computational Biology challenges. The three highest scoring teams were qualified for the third, and last, phase. Which consisted on the development of an Investigation Project. The final results will be presented by each team during the Bioinformatics Open Days event.

Bioinformatics Open Days 2023

ROUND TABLE + NETWORK SESSION

Discussing the reality of the corporate world

This year's edition will integrate a Round Table initiative, where invited colaborators will be present to represent their company and discuss the corporate world, whilst answering questions from the public. This discussion will then be followed by a Networking Session. In this component participants will have the opportunity to interact with each company representative, incentivizing follow-up questions and general discussion about starting their professional career. The aim is to expose successful and growing technological-based companies, whilst closing the gap between company representatives and all those interested.

SilicoLife

SilicoLife designs optimized microorganisms and novel pathways for industrial biotechnology applications. The team includes specialists in several areas, namely biotechnology, computational biology, metabolic engineering, molecular biology, systems biology, bioinformatics, and text mining. SilicoLife builds computational models of microbial cells and develops proprietary state-of-the-art algorithms to find the most efficient pathways between raw materials and end-products, streamlining the strain design process and explore non-intuitive pathway modifications.

Accenture

Accenture is amongst the leading consulting companies worldwide, with offices spread throughout 49 countries and over 200 cities. Assisting leading businesses accelerate revenue growth and optimize processes (from economic to technological), their dedicated team of 738,000 employees, with ranging areas of specialization, allows for the shared application of knowledge and expertise to solve a myriad of problems. This allows for innovative and client-specific solutions that leverage modern technological advancements and help build lasting relationships of trust between service-providers and clients.

Bioinformatics Open Days 2023

ROUND TABLE + NETWORK SESSION

Discussing the reality of the corporate world

OmniumAI

Focusing on leveraging artificial intelligence to solve current issues within the field of biomedicine, OmniumAI aims to automate machine and deep learning methodologies to infer knowledge based on relevant biological data, making it accessible for experts and beginners in the field alike. Their dedicated team is composed of both scientists and engineers with broad know-how in artificial intelligence and data science. This spin-off's ambitions are to offer tailored solutions for current, significant issues, deploying high-performance models to deal with different types of biological data, including proteins, compounds and DNA.

Bioinformatics Open Days 2023

WORKSHOPS

Introduction to Chemoinformatics using DeepMol

The aim of this workshop is to introduce the audience to the field of Chemoinformatics, especially in the area of drug discovery. The workshop will focus on key topics such as chemical compound representations, standardization, and feature generation. Participants will also learn about techniques for handling imbalanced data, training and evaluating ML/DL models, and understanding model explainability.

In this workshop, we will use DeepMol (https://github.com/BioSystemsUM/DeepMol), a Python-based machine learning and deep-learning framework for drug discovery. To fully participate, attendees must have a basic understanding of Python programming.

Join us for a hands-on exploration of how chemoinformatics can be used to tackle real-world drug discovery problems.

João Correia

PhD Candidate at University of Minho
Developer of the DeepMol package

João Capela

PhD Candidate at University of Minho
Developer of the DeepMol package

Bioinformatics Open Days 2023

WORKSHOPS

The importance of machine learning in structural biology

In this workshop we will perform a binary classification task to predict protein residues that interact with a ligand. We will train three different machine learning models (Gaussian Naive Bayes, Random Forest, and Multi-layer Perceptron) and evaluated their performances using appropriate metrics. We will cover the theoretical aspects of a typical machine-learning project and engage in hands-on activities with actual proteins using a Google Colaboratory notebook. After this workshop, you will be able to apply the basic steps of a machine-learning project with structural biology data.

Irina Moreira

Leader and Researcher of the Center for Neuroscience and Cell Biology of the University of Coimbra (CNC UC)

Catarina Marques-Pereira

PhD Candidate at CNC UC

Luana Afonso

PhD Candidate at CNC UC

Bioinformatics Open Days 2023

WORKSHOPS

Ontology Matching in the Biomedical Domain

Ontology matching is the process of defining correspondences between two or more related ontologies, which can be used to either map or integrate them. This is critical to ensure data findability and interoperability when datasets are described using different ontologies, a problem that is increasingly more common in the biomedical domain due to the prolific development of ontologies therein.

Biomedical ontologies pose unique challenges to ontology matching due to their distinct profile. In this hands-on tutorial, we overview these challenges, the state-of-the-art solutions to address them, the ontology matching tools that implement such solutions, and their performance in independent evaluation. Furthermore, we discuss the role of the user in validating ontology alignments and/or performing interactive matching. Finally, we review current infrastructures, initiatives and applications involving ontology matching.

Marta Silva

PhD Candidate at LASIGE

Patrícia Eugénio

PhD Candidate at LASIGE

Bioinformatics Open Days 2023

SPONSORS AND COLLABORATORS

Thank you note

Dear Sponsors, Collaborators, Volunteers, and Members of the Organizing Committee,

On behalf of the XII Edition of the Bioinformatics Open Days, we would like to express our most sincere appreciation for your support and contributions to the event. Your hard work and dedication are key in making the conference a success.

We extend our deepest thanks to our sponsors and collaborators, whose generosity and partnership made it possible to create a program of keynote talks, workshops, and networking opportunities that will be both informative and enjoyable. Your contributions ensured that our event remains an accurate representation of the present and future of the Bioinformatics field, and we are grateful for your commitment to our cause.

We also want to express our gratitude to the volunteers and members of the organizing committee, whose efforts were essential in the planning and execution of the event. Your dedication and commitment to the conference are critical to its success, and we appreciate the time and effort you put into organizing the various aspects of the event.

We would like to express a special thank you to the Student Association for the Bioinformatics students of the University of Minho (NEBIUM, Núcleo de Estudantes de Bioinformática da Universidade do Minho). We are grateful for the time, effort, and dedication you put into organizing various aspects of the conference. Your expertise and advice were essential in shaping the program and ensuring that it met the needs of all attendees.

Finally, we would also like to extend our appreciation to the general chair, Professor Miguel Rocha, for his insightful opinions and experience. His guidance and leadership were instrumental in ensuring the success of the event.

In conclusion, the success of the XII Edition of the Bioinformatics Open Days was a result of the collective efforts of all of you. We are deeply grateful for your support and look forward to continuing our partnership in future editions.

Sincerely,

Camila Babo
President of the Bioinformatics Open Days 2023

Bioinformatics Open Days 2023

SPONSORS AND COLLABORATORS

thank you!

Bioinformatics Open Days 2023

ANNEX I: SCIENTIFIC SUBMISSIONS

N°	Authors	Title	Accept as
1	Hugo Rodrigues et al	Identification of gene regulatory modules acting in the interaction between cork development and environmental variable	Oral
2	João C. Sequeira et al	A user-friendly bioinformatics pipeline for metaproteogenomics data analysis	Poster
3	Rita Guerra et al	Development of a bioinformatic workflow to evaluate tyrosine kinase inhibitor derivatives with improved membrane permeabilities	Poster
4	Patrícia Eugénio et al	The ImmunoPeptidomics Ontology: design and evaluation	Oral
5	Hayden So et al	ImmspaCy: Extracting Gene-disease Associations for Systems Immunology Discoveries	Poster
6	Mohamed Emam et al	Rigorous Assessment of Multi-omics Integration Tools for the Study of Pre-diabetes	Poster
7	Marta Batista et al	A bioinformatics approach to study the permeation of solutes through PfAQP for the development of new antimalarial therapies	Oral
8	Marta Silva et al	Holistic Biomedical Knowledge Graph Integration	Poster
10	Inês Pires et al	Computational study of promising smart metallodrug delivery systems	Oral
12	David Henriques et al	Energy demand and enzyme budget trade-offs modulate the accuracy of dynamic Flux Balance Analysis	Oral
13	Ian Teixeira et al	BIT by BIT: Associating Transcription to Ageing	Poster
14	A. Paulino et al	Genome-wide transcriptomic analysis reveals novel genes involved in cynaropicrin synthesis in Cynara cardunculus	Poster
15	Tiago Ferreira et al	Molecular Dynamics as an active resource on the development of new formulations for commercial use: The case of natural silicone alternatives	Oral
16	Maria Fernanda Vieira et al	Development of comprehensive methodologies for bacteriophage- host studies	Poster
17	José Freitas et al	Prediction of plastic-degrading enzymes in omics data from marine environments	Poster
18	Tiago Pereira et al	Exploring self-attention mechanisms and deep reinforcement learning for the de novo drug design	Oral
19	Joana Sousa et al	Metagenomic approach to identify genes encoding for glycoside hydrolases in composting samples	Poster
20	Cátia Santos-Pereira et al	Exploring Aveiro salterns to discover new and robust biosurfactant producers	Poster
21	Francisca G. Vieira et al	Integration of Multi-Omics Data for the Classification of Glioma Subtypes and Identification of Novel Biomarkers	Poster
23	Rita I. Teixeira et al	The impact of the SARS-CoV-2 Omicron variant on the receptor binding domain conformational dynamics and interaction with human ACE2	Oral
24	Caio S. Souza et al	Improving phenylalanine hydroxylase catalytic activity for tryptophan using an automated platform of structure-based enzyme engineering	Poster
25	Diogo Fonseca et al	Transcriptional characterization of cTfh cells in a viral infection at single-cell resolution	Poster
26	Beatriz Vieira et al	Workflow for identification of InDel markers for rice (Oryza sativa) varietal discrimination	Poster

ANNEX I: SCIENTIFIC SUBMISSIONS

27	Ana B.Caniceiro et al	MUG: a mutation overview of GPCR subfamily A17 receptors	Oral
28	Ana Santos-Pereira et al	ngest: A Scalable Snakemake Pipeline for Customized Knowledge Graph construction	Oral
30	Diego Troitiño-Jordedo et al	A new GIMME-based compartmentalised algorithm for transcriptomics data integration	Poster
31	Susana Parreiras et al	Design, production and characterization of antiviral proteins targeting SARS-CoV-2	Poster
32	Filippo Bergeretti et al	A comparative genomics approach to assess interspecific variability associated with cork development	Oral
33	Alexandre Fernandes et al	Comparative analysis of Human and Macaca oral fibroblasts by scRNA-seq	Poster
34	Sara Freitas et al	Characterization of grapevine (Vitis vinifera) intra and inter- varietal diversity using whole genome resequencing	Oral
35	Rita Costa Pires	RNA-Seq analysis of laser capture microdissected tissues of the cork oak	Poster
36	João Nunes et al	Bioinformatics characterization of grapevine inter-varietal diversity using WGS data	Poster
37	João Monteiro	Bioinformatic characterization of genetic variation in grapevine using PoolSeq data	Poster
39	João Correia et al	DeepMol: a python-based machine and deep learning framework for drug discovery	Poster
40	Daniel Martins	Bioinformatic approaches to address new perspectives on genotype-phenotype associations for complex diseases	Poster

XII EDITION BIOINFORMATICS OPEN DAYS

Thank you for joining us at XII BOD!
Your participation and support are greatly appreciated. We look forward to seeing you at future events and continuing to grow our Bioinformatics community.

Bioinformatics Open Days 2023

/bioinformaticsopendays