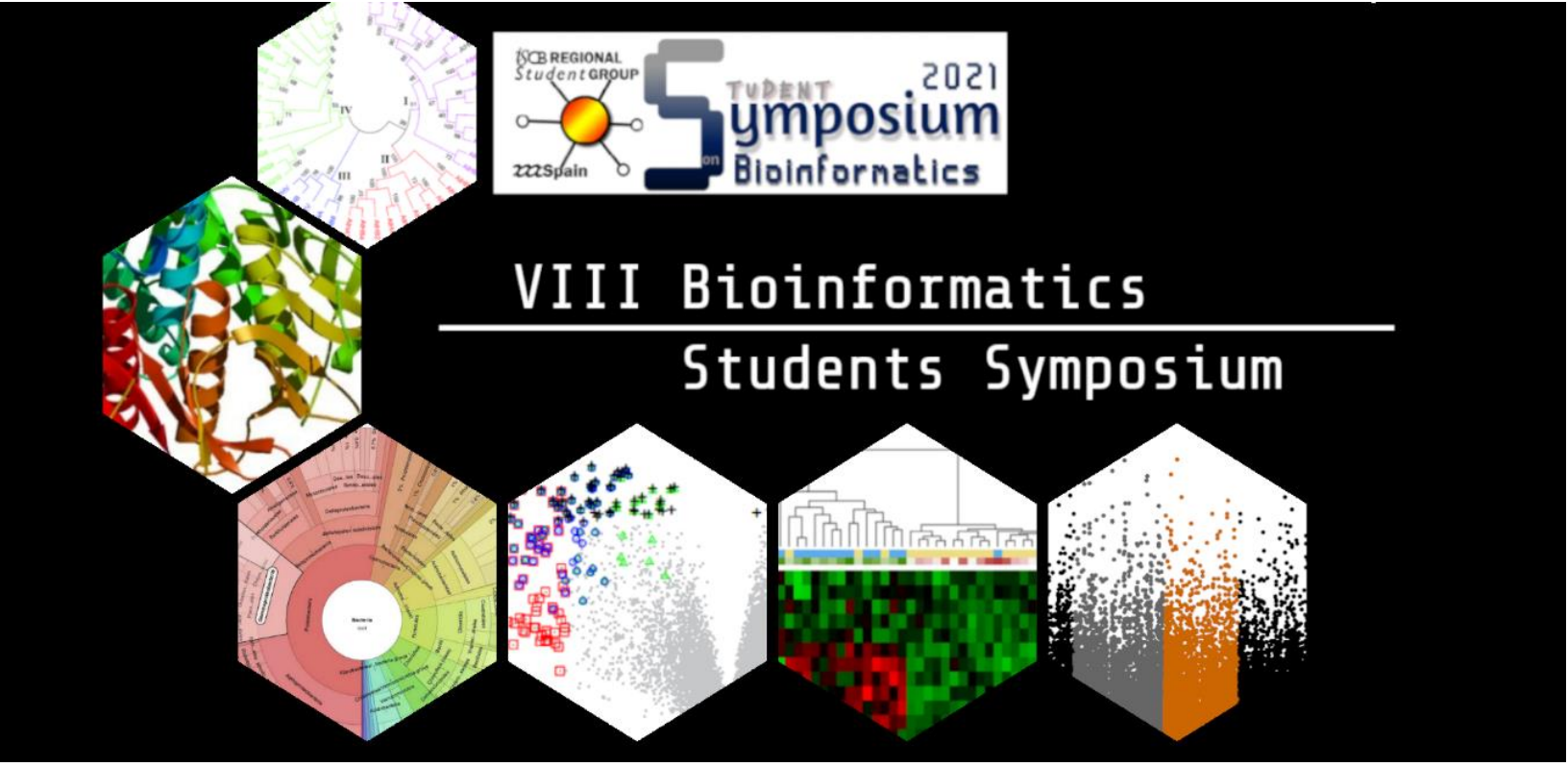


ONLINE 18-19th OCTOBER 2021

CONFERENCE ABSTRACT BOOK

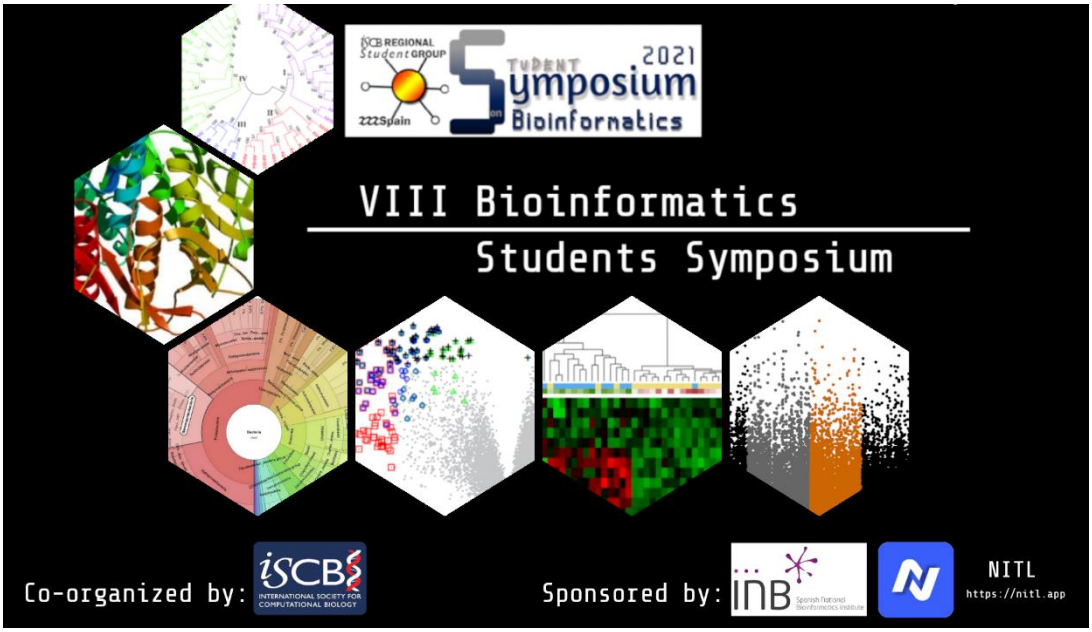
ISCB REGIONAL Student GROUP Spain

ISCB Student COUNCIL
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

Index

Organizers and collaborators

The **VIII Student Symposium in Bioinformatics** is organized by **RSG-Spain**, an association of junior scientists in Bioinformatics based in Spain and satellite group of the International Society for Computational Biology (ISCB), the largest scholar society in bioinformatics.



Gold sponsor:

Silver sponsors:

Organizing committee

- Symposium chair: **Sara Monzón** (Institute of Health Carlos III, BU-ISCIII).

- Speakers committee chair: **Inés Rivero** (Genetic Control of Organ Development and Regeneration Lab and Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares (CNIC).

- Evaluation committee chair: **Tamara Hernández** (Hospital Nuestra Señora de Candelaria).

  Team:

  Luis A. Rubio Rodríguez (Instituto Tecnológico y de Energías Renovables).

  Virginia Diez (University of Barcelona)

  Ana Díaz-de Usera (Instituto Tecnológico y de Energías Renovables).

  Eva Suárez (Hospital Nuestra Señora de Candelaria).

- Networking activities chair: **Adrián García Moreno** (Centro de Genómica e Investigación Oncológica (GENyO), BioInfoGRX).

  Team:

  Ana María Gómez Martínez (Instituto de Genética Humana de Ulm, BioInfoGRX)

  Álvaro Torres Martos (Centro de Investigación Biomédica de la Universidad de Granada, BioInfoGRX)

  Alberto Manuel Parra Pérez (Centro de Genómica e Investigación Oncológica (GENyO), BioInfoGRX)

  Raul López Domiguez (Universidad de Granada, Centro de Genómica e Investigación Oncológica (GENyO), BioInfoGRX)

  Ivan Ellson Lancho (Centro de Genómica e Investigación Oncológica (GENyO), BioInfoGRX)

  María Soledad Benítez Cantos (Centro de Genómica e Investigación Oncológica (GENyO), BioInfoGRX)

  Erica Padial Fuillerat (Universidad de Utrecht, BioInfoGRX)

Ane Martinez Larrinaga (Tecnun Escuela de Ingenieros de la Universidad de Navarra, BioInfoGRX)

Roberto Fornelino González (Imuno Therapeutics, BioInfoGRX)

- Outreach and web committee:

  Mario Rodríguez (Universidad Autónoma de Madrid).

  Mónica Cabrera (Barcelona Supercomputing Center).

  Lluís Revilla (CIBER, IDIBAPS).

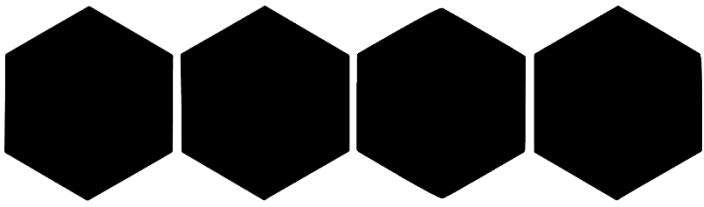
- Secretary and design committee:

  Javier Lanillos (Centro Nacional de Investigaciones Oncológicas).

  Miguel Juliá (Institute of Health Carlos III, BU-ISCIII).

  Eva Tosco (Hospital Nuestra Señora de Candelaria).

  Guillermo Gorines (Rey Juan Carlos University).



Symposium program

18th October: Meet&Greet event

10:00: Presentation of RSG and RSG-Nodes. Encouraging the creation of new nodes.

11:00: Interactive Survey: Surveying the young Spanish BioInfo topics and practices.

12:00: Round table with RSG-Spain representatives (bioinfo trends discussion and Q&A).

13:30: Abroad RSG Leader / Representative intervention - Argentina.

14:30: Social gathering on AirMeet or Remo rooms. Tables divided by topics.

19th October: VIII Student Symposium

10:00: Registration and online network gathering.

10:10: Keynote speaker. **Arcadi Navarro**. **Moderador:** Sara Monzón

"For more and Better Science. Big and Small Data Governance, Management & Sharing"

11:10: Women in Science. **Nataly Buslón**. **Moderador:** Mónica Cabrera

"Sex and gender perspective in research"

11:30: Break, online network gathering.

11:45: Lightning talks for selected abstract as a poster. **Moderador:** Eva Tosco

Tamara Hernández-Beeftink: *SAMD9 and genetic determinants of 28-day sepsis survival*

Eva Tosco-Herrera: *Software assessment for variant prioritization of free and open-source software*

Eva Suarez-Pajes: *A fine mapping association study of the HLA in sepsis susceptibility*

Almudena Neva-Alejo: *An integrated single-nucleus transcriptomics approach to identify sex molecular differences in Alzheimer's disease*

Irene Soler-Sáez: *A single-cell and single-nuclei transcriptomics approach reveals novel sex differences in multiple sclerosis*

Marina Murillo Recio: *tRNAstudio: analysis of tRNA-seq datasets without the need of programming skills*

Inés Rivero-García: *Assessment of the generalizability of convolutional neural network coexpression analysis for the inference of gene regulatory networks from scRNA seq expression data*

Cristina Zamora-Ballesteros: *Genome-wide identification, characterization and functional analysis of defense-related long non-coding RNAs in non-model plant Pinus radiata*

Siddhant Sharma: In silico characterization of prion-like domains in human protein families

Miquel Anglada-Girotto: *Robustica: customizable robust independent component analysis*

12:00: Selected talks from students on bioinformatics. **Moderador:** Tamara Hernández

Ana Díaz-de Usera: *Whole exome sequencing of the Canary Islands population to unravel natural genetic variation*

Patricia Sánchez-Jiménez: *Transcriptomic exploration of the epileptogenic zone of drug-resistant temporal lobe epilepsy patients*

Javier Lanillos Manchón: *Clinically actionable pharmacogenetic landscape for pre-emptive use of diagnostic Whole Exome Sequencing*

Blanca Rodríguez-Fernández: *Genetically predicted Telomere Length and its relationship with major Neurological Diseases and Life Expectancy: A Mendelian Randomization Study*

13:00: Lunch break.

14:00: Selected talks from students on bioinformatics. **Moderador:** Javier Lanillos

Olfat Khannous-Lleiffe: *Microbiome profiling from Fecal Immunochemical Test reveals microbial signatures with potential for Colorectal Cancer screening*

Edurne Urrutia-Lafuentel: *A systematic review and meta-analysis of transcriptomic studies unveil sex differences in rheumatoid arthritis*

Héctor Rodríguez-Pérez: *NanoRTax, a real-time pipeline for taxonomic and diversity analysis of nanopore 16S rRNA amplicon sequencing data*

Miquel Àngel Schikora-Tamarit: *perSVade: personalized Structural Variation detection in your species of interest*

Roc Farriol-Duran: Neutralizing antibodies as a therapeutic avenue for COVID-19

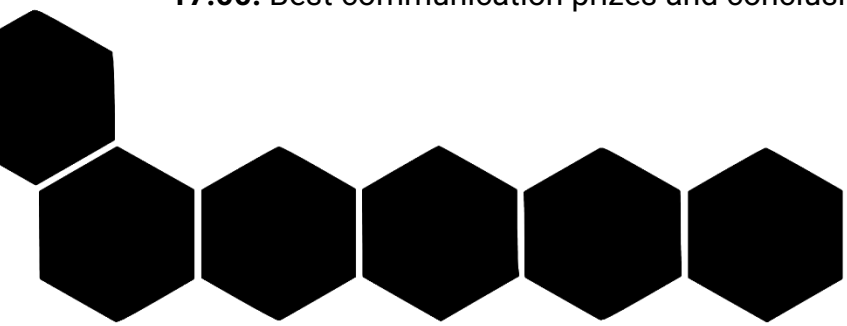
David Martínez: *Dynamic A-to-I RNA Editing Landscapes during Aging Process*

15:30: Break.

16:00: Keynote speaker: **Laura Furlong**. Career advice talk. **Moderador:** Inés Rivero

"From bench to in silico, and from academia to entrepreneurship"

17:00: Best communication prizes and conclusions. **Moderador:** Tamara Hernández



Keynote speaker: Arcadi Navarro

ICREA Research Professor at Institut de Biologia Evolutiva (CSIC-UPF)

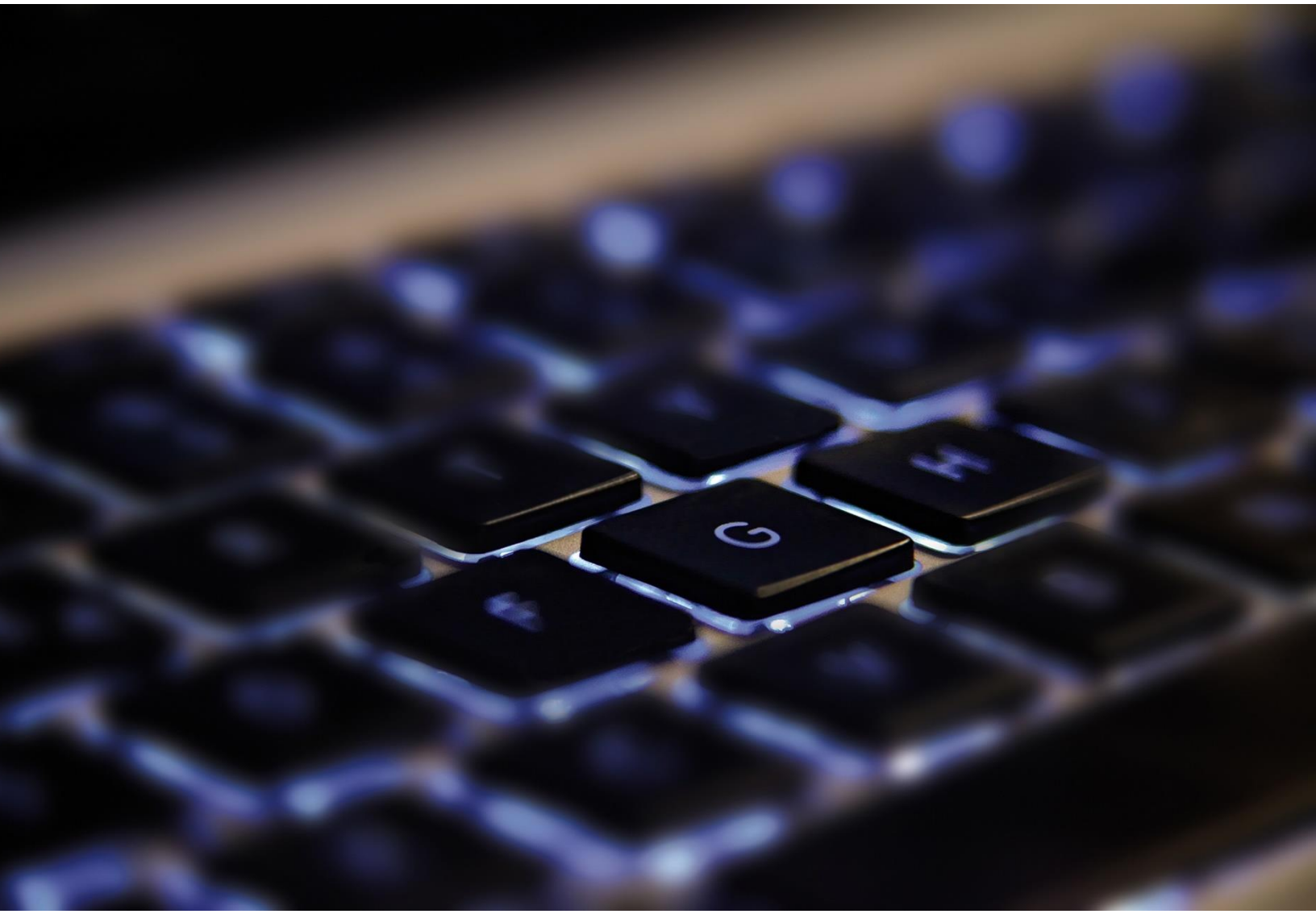
Prof. Arcadi Navarro received a PhD. in Genetics at Universidad Autónoma de Barcelona in 1998. After a postdoctoral stay in Edinburgh University, he set base at the Institute for Evolutionary Biology of UPF-CSIC, being appointed Research Professor in 2006, Professor of Genetics in 2010, and Director of the Department of Experimental and Health Sciences in 2013. In 2016, his career took a twist to become the Secretary of Universities and Research of La Generalitat, returning to UPF in 2019 and becoming the Director of the Pasqual Maragall Foundation in 2020. Intrigued by the genomic footprints of evolutionary processes, Prof. Navarro interrogates genomic variability to understand how evolutionary forces shape our genomes and what are their consequences on biodiversity, ageing and disease susceptibility. Prof. Navarro has provided answers to these very relevant questions in over 140 papers, books and over 20 conferences. Moreover, he is an active mentor, teaching at undergraduate and graduate level and having supervised 10 PhD. theses. Lastly, Prof. Navarro is also committed to communicating science to the public, and proof of this are his awards on the 11th European Scientific Divulgation Award (University of Valencia, 2005), and the Casa de las Ciencias Scientific Divulgation Award (A Coruña City Hall, 2007).

Keynote speaker: Laura Furlong

Scientific Director of MedBioInformatics Solutions & Senior Researcher at the Integrative Biomedical Informatics at GRIB (IMIM-UPF)

Dr. Laura Furlong obtained a PhD. in Biology from Universidad de Buenos Aires in 2002. She then joined Universidad Autonoma de Barcelona as a postdoctoral researcher and, after graduating from the UPF MSc in Bioinformatics in 2007, she joined the Integrative Biomedical Informatics Research group. Additionally, Dr. Furlong is an UPF associated lecturer since 2007 and Scientific Director of MedBioinformatics Solutions. In her research, Dr. Furlong unravels the pathological mechanisms of human disease from a network's perspective. To do so, her group develops bioinformatic methods for network analysis and text mining, and applies them to characterize disease and adverse drug effects. Fruit of her work are over 80 scientific papers and 4 PhD. theses, as well as the knowledge platform on disease genomics DisGeNET. DISGENET plus is the flagship product of MedBioinformatics Solutions, the spin-off company that she co-founded in 2020 to provide genomics tools and know-how for precision medicine and pharmaceutical R&D. Together with her academic and entrepreneurial successes, Dr. Furlong has been very active in sharing her scientific knowledge by teaching courses on network analysis and visualization, biocomputing and health informatics in prestigious research centers in Spain and abroad. She is also a member of the advisory boards of the Spanish Society of Health Informatics and the Gather Foundation.

Abstracts



A coarse-grained approach to model the dynamics of the actomyosin cortex

Miguel Hernández-del-Valle[1,2,3,4], Andrea Valencia-Expósito[6], Antonio López-Izquierdo [1,2,3,4], Pau Casanova-Ferrer[1,2,3,4], Pedro Tarazona[2,3,5], María D. Martín-Bermudo[6], David G. Míguez[1,2,3,4].

[1]Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid, 28049 Madrid; [2]IFIMAC, Fac. de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid; [3]Instituto Nicolás Cabrera, Fac. de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid; [4]Física de la Materia Condensada, Fac. de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid; [5]Física Teórica de la Materia Condensada, Fac. de Ciencias, Universidad Autónoma de Madrid,28049 Madrid; [6]Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide/CSIC/JA, Carretera de Utrera km 1, Sevilla 41013.

The dynamics of the actomyosin machinery is at the core of many important biological processes. Several relevant cellular responses such as the rhythmic compression of the cell cortex are governed, at a mesoscopic level, by the nonlinear interaction between actin monomers, actin crosslinkers and myosin motors. Coarse grained models are an optimal tool to study actomyosin systems, since they can include processes that occur at long time and space scales, while maintaining the most relevant features of the molecular interactions. Here, we present a coarse-grained model of a two-dimensional actomyos in cortex, adjacent to a three-dimensional cytoplasm. Our simplified model incorporates only well characterized interactions between actin monomers, actin crosslinkers and myosin, and it is able to reproduce many of the most important aspects of actin filament and actomyosin network formation, such as dynamics of polymerization and depolymerization, treadmilling, network formation and the autonomous oscillatory dynamics of actomyosin. Furthermore, the model can be used to predict the in vivo response of actomyosin networks to changes in key parameters of the system, such as alterations in the anchor of actin filaments to the cell cortex.

Area: Systems Biology and Networks

SAMD9 and genetic determinants of 28-day sepsis survival

Tamara Hernandez-Beeftink[1,2], Beatriz Guillen-Guio[1], Jose M. Lorenzo-Salazar[3], Almudena Corrales[1,4], M Isabel García-Laorden[2,4], Miryam Prieto González[5], Aurelio Rodríguez-Pérez[6,7], Demetrio Carriedo[8], Jesús Blanco[4,9], Alfonso Ambrós[10], Elena González Higueras[11], Elena Espinosa[12], Arturo Muriel[9], David Domínguez[12], Abelardo García de Lorenzo[13], José M. Añón[13], Javier Belda[14], Jesús Villar[2,4], Carlos Flores[1,3,4,5] and the Genetics of Sepsis (GEN-SEP) Network.

[1]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; [2]Research Unit, Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain; [3]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain; [4]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; [5]Servicio de Medicina Intensiva, Complejo Asistencial Universitario de Palencia, Palencia, Spain; [6]Department of Anesthesiology, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain; [7]Department of Medical and Surgical Sciences, University of Las Palmas de Gran Canaria, Gran Canaria, Spain; [8]Intensive Care Unit, Complejo Hospitalario Universitario de León, León, Spain; [9]Intensive Care Unit, Hospital Universitario Rio Hortega, Valladolid, Spain; [10]Intensive Care Unit, Hospital General de Ciudad Real, Ciudad Real, Spain; [11]Intensive Care Unit, Hospital Virgen de la Luz, Cuenca, Spain; [12]Department of Anesthesiology, Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain; [13]Intensive Care Unit, Hospital Universitario La Paz, Madrid, Spain; [14]Anesthesiology and Critical Care Department, Hospital Clinico Universitario of Valencia, Valencia, Spain.

Introduction. Sepsis is a severe systemic inflammatory response to infections that has a 20-30% mortality rate. To date, most genetic studies have focused on particular biological candidates. We performed the first genome-wide association study (GWAS) of 28-day survival in sepsis.

Methods. A GWAS was performed in 475 Europeans from the GEN-SEP cohort and 7.5 million imputed variants. Association analyses were conducted using Cox regression models, adjusting by gender, age, and the main two principal components of genetic variation. A replication was performed in 212 independent patients from the same cohort. Statistical significance was established at p<5.0E-8 in the meta-analysis. Whole-blood transcriptomics from septic patients were assessed by focusing on genes linked to significant variants.

Results and conclusions. The GWAS identified three independent common variants associated with reduced 28-day survival, including an exonic variant in *SAMD9* (HR=4.75, 95%CI=2.86-7.89, p=1.77E-9), which encodes the sterile alpha motif domain-containing protein 9, related with the inflammatory response to tissue injury. An upregulation of *SAMD9* expression in non-surviving septic patients (p=1.67E-03) was observed. In conclusion, we completed the first GWAS of 28-day survival in sepsis and identified novel candidate genes associated with reduced survival.

Funding: Instituto de Salud Carlos III (PI16/00049; PI17/00610; FI17/00177; PI19/00141; PI20/00876) and co-financed by the European Regional Development Funds, "A way of making Europe" from the European Union; and by the agreement OA17/008 with Instituto Tecnológico y de Energías Renovables (ITER) to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology; ECIT CGIEU0000219140.

Caracterization of sex differences in pancreatic ductal adenocarcinoma through an integrated transcriptomic approach

Lucía Fantin, Príncipe Felipe Research Center, UBB-CIPF, Spain
Zoraida Andreu Martínez, IVO-CIPF Joint Research Unit of Cancer, Spain
Carla Perpiñá-Clérigues, Príncipe Felipe Research Center, UBB-CIPF, Spain
Irene Pérez-Díez, FISABIO-CIPF Joint Research Unit in Biomedical Imaging, Spain
María de la Iglesia-Vayá, FISABIO-CIPF Joint Research Unit in Biomedical Imaging, Spain
Jose A. López-Guerrero, IVO-CIPF Joint Research Unit of Cancer, Spain
Francisco García-García, IVO-CIPF Joint Research Unit of Cancer, Spain. UBB-CIPF;

Corresponding author e-mail address: fgarcia@cipf.es.

Pancreatic cancer represents a major global health problem because it is one of the neoplasms with the highest rates of mortality worldwide. The most common subtype is pancreatic ductal adenocarcinoma (PDAC). Epidemiologically, it has been described that this disease has higher incidence rates in men than in women, but the reasons that generate this difference are not already clear. This study aims to identify the molecular mechanisms that explain these differences to develop personalized medicine for more effective and specific therapies that modify the prognosis of this disease. To achieve this, we have analyzed the transcriptomic data of PDAC studies from public repositories that includes sex information, obtained through a systematic review. For each of them we performed a differential expression analysis and then we integrated these data through a gene meta-analysis. Results were characterized through a functional enrichment that allowed identifying molecular functions differentially affected for each sex.

The meta-analysis results did not shown genes differentially expressed between men and women, however we detected functions related to the immune system and cell cycle regulation more expressed in men and functions related to the development of the tumor microenvironment and inflammatory diseases more expressed in women.

These results reveal molecular differences between men and women for PDAC, but further research is necessary to expand this knowledge.

Key words: pancreatic cancer, PDAC, sex, meta-analysis, transcriptomics.

An integrated single-nucleus transcriptomics approach to identify sex molecular differences in Alzheimer's disease

Almudena Neva-Alejo[1], Zoraida Andreu[2], Irene Soler-Sáez[1], José F. Català Senent[1], Adolfo López Cerdán[3], María de la Iglesia-Vayá[3], Francisco García-García[1].

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012 Valencia, Spain.
[2]Foundation Valencian Institute of Oncology (FIVO), 46009, Valencia, Spain
[3]Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012 Valencia, Spain

Corresponding author e-mail address: fgarcia@cipf.es.

Alzheimer's disease (AD) is the leading cause of dementia. The clinical symptoms of AD begin with short-term memory impairment and gradually important cognitive and conductual functions are impaired, finally driving to death. By 2050, the number of AD patients is expected to grow to more than 50 millions which, in addition to the lack of disease-modifying treatments and precise and early diagnostic methods, highlights the interest of the investigation in this field. As sex differences in prevalence, severity and progression of AD and many other diseases have been reported, sex perspectives in biomedical research may lead for a better and faster understanding of diseases, being a fundamental step towards personalized medicine. We have explored the sex-related molecular differences in AD by integrative analysis of single cell transcriptomics studies from postmortem brain tissues. Single nucleus RNA sequencing is a powerful approach to reveal the cellular heterogeneity masked in population-averaged detected gene expression in bulk RNA-seq. We have focused in sex differences in astrocytes and microglia for its important role in neuroinflammation in AD. For this purpose, we have performed a systematic review following the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) that afforded the selection of three studies. We conducted a differential gene expression analysis by cell type and a differential cell abundance analysis across sex/disease groups of subjects for each study individually and then, we integrated all the results. The results show significant differences in gene expression between women and men in AD, some of them were previously associated with AD which gives robustness to our analysis. On the other hand, no differences in cell abundance have been detected. Although further studies must be done, the found differentially expressed genes represent a potential key to understand how sex may influence in AD.

Keywords: Alzheimer's disease, Single-nucleus RNA sequencing, Sex perspective, Precision medicine

Microbiome profiling from Fecal Immunochemical Test reveals microbial signatures with potential for Colorectal Cancer screening

Olfat Khannous-Lleiffe[1,2], Jesse R. Willis[1,2], Ester Saus[1,2], CRI-PREV Consortium, Toni Gabaldón[1,2,3*].

[1]Barcelona Supercomputing Center (BSC-CNS). Carrer de Jordi Girona, 29, 31, 08034 Barcelona
[2]Institute for Research in Biomedicine (IRB). Carrer de Baldiri Reixac, 10, 08028 Barcelona
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA). Pg. Lluís Companys 23, 08010 Barcelona, Spain.

The understanding of the differential properties of the microbial communities living within each individual (microbiome), and its relationships with host genetics, diet and environment, can reveal links between the microbiota and disease or health. In the context of personalized medicine, the detection of microbiome differences between groups of individuals (e.g controls and cases) can be a valuable tool, not only to find risk biomarkers for early diagnosis or develop personalized treatments by modulating the microbiome, but also for discovering novel processes involved in disease origin or progression. Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer deaths worldwide. Early diagnosis of CRC, which saves lives and enables better outcomes, is implemented in many regions through a two-step population screening approach based on the use of Fecal Immunochemical Test (FIT) followed by colonoscopy if the test is positive. However, this intermediate assay has a high false positive rate in the first step, and there is a need to identify new predictive biomarkers to prioritize clinically relevant cases who should undergo a colonoscopy. We explored the diagnostic potential of fecal microbiota by applying a 16S metabarcoding strategy directly on FIT buffer. We uncovered microbial taxa and metabolic features in FIT positive samples that were significantly associated with different colonoscopy outcomes, underscoring a predictive potential and revealing changes along the path from healthy tissue to carcinoma. Using diagnostic evaluations from colonoscopy, we reconstructed changes in the composition, taxon co-occurrence and metabolic features of microbial communities associated to clinically relevant traits such as the presence of distinct precancerous lesions, hinting to potential microbial roles in the origin and progression of CRC. Finally, we used machine learning to develop a two-phase classifier that combines information from bacterial signatures, sex, age and FIT value. This classifier achieved close to 100% sensitivity for CRC, while significantly reducing the current false positive rate.

Whole exome sequencing of the Canary Islands population to unravel natural genetic variation

Ana Díaz-de Usera[1], Luis A. Rubio-Rodríguez[1], Adrián Muñoz-Barrera[1], Beatriz Guillen-Guio[2], Almudena Corrales[2,3], Antonio Íñigo-Campos[1], David Jáspez1, Itahisa Marcelino-Rodriguez[2,4], Antonio Cabrera de León[2], Rafaela González-Montelongo[1], Jose M. Lorenzo-Salazar[1], Carlos Flores[1,2,3].

[1]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain;
[2]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain;
[3]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain;
[4]Instituto de Tecnologías Biomédicas, Universidad de La Laguna, Santa Cruz de Tenerife, Spain.

Introduction: In order to provide the basis for an unbiased understanding of traits and diseases, many countries are developing their own catalogs of genetic variation. The Canary Islands (Spain) population exhibits the largest genetic proportion of North African ancestry among the Southwestern European populations. Here we provide, for the first time, an exome-wide analysis of their genetic variation.

Materials and Methods: DNA from 629 unrelated control donors were sequenced by commercial whole exome enrichment capture kits and a HiSeq 4000 (Illumina) using 75 bp paired-end reads. BWA-GATK v3.8 Best Practices were followed for germline variant calling using the GRCh37/hg19 human genome as the reference. ANNOVAR, Ensembl VEP, and InterVar tools were used for functional annotation.

Results: A total of 340,913 variants were identified by whole exome sequencing in the target regions, of which 70.1% were exonic. A quarter of them (i.e., 59,883) were absent from the 1000 Genomes Project, gnomAD, and TopMed, and 97.4% of those presented an allele frequency below 0.5%. Out of these, 3,276 were classified as pathogenic or likely pathogenic according to InterVar. Nearly 90% of those had high impact predictions or were loss-of-function variants, being frameshift the most common variation.

Conclusions: The high percentage of novel exome-wide genetic variation in Canary Islanders highlights the need for a specific catalog to improve the pathogenicity classification of variants in this population.

Funding: Ministerio de Ciencia e Innovación (RTC-2017-6471-1; AEI/FEDER,UE), agreement OA17/008 with ITER, Spanish Ministry of Education and Vocational Training (FPU16/01435); Instituto de Salud Carlos III (CD19/00231); Cabildo Insular de Tenerife CGIEU0000219140.

Novel computational strategies for the identification of a microRNA signature for diagnosis and prognosis in melanoma

Helena Gómez-Martínez[1], Adolfo López Cerdán[2], Francisco García-García[1].

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012 Valencia, Spain.
[2]Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012 Valencia, Spain.

Corresponding author e-mail address: fgarcia@cipf.es

Melanoma is one of the most aggressive types of skin cancer, with a progressive increase in its incidence in recent years. The lethality of the disease is mainly influenced by the late onset of symptoms, the lack of an effective treatment and the difficulties in early disease detection. However, although clinical and epidemiological parameters show a differential pattern by sex in the diagnosis and prognosis of this disease, there is a lack of knowledge regarding efficient biomarkers that favour the early detection and prediction of the evolution of the disease in patients. Consequently, a computational approach integrating all information could improve the understanding of the development of melanoma and the prediction of the patient diagnosis.

In this work, a systematic review of melanoma studies in different databases was conducted and based on previous criteria, only those that were microarrays or massive sequencing of miRNAs were selected. After data download, an exploratory analysis and data normalization were performed. Subsequently, an individual analysis of each study and a global meta-analysis was carried out by integrating all the data obtained and considering the gender-specific information of the participants.

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

Keywords: melanoma, microrna expression, sex-differences, meta-analysis, precision medicine.

Clinically actionable pharmacogenetic landscape for pre-emptive use of diagnostic Whole Exome Sequencing

Javier Lanillos Manchón[1].

[1]CNIO

Introduction: Whole exome sequencing (WES) is utilized in routine clinical genetic diagnosis. The technical robustness of repurposing large-scale next generation sequencing data for pharmacogenetics has been demonstrated, supporting the implementation of preemptive pharmacogenetics. However, few studies with limited sample size or limited to specific pharmacogenes have explored the clinical utility of adding clinical pharmacogenetics interpretation to diagnostic WES.

Aim: We performed a systematic analysis of a large cohort of individuals with diagnostic WES, to provide with global and gene-specific clinical pharmacogenetic utility data, population specific differences and rare loss-of-function variation.

Methods: 809 pharmacogenetic alleles, distributed through 19 genes, defined in a Clinical Pharmacogenetics Implementation Consortium guideline were interrogated in 5,001 individuals with a standard diagnostic WES testing (57% Spain; 27% Colombia; 11% Brazil; 5% other). Analysis included variant retrieval, quality data analysis, genotype to diplotype conversion and pharmacogenetic phenotype classification.

Results: We established that 302 alleles in 11 genes could be used to inform of pharmacogenetic phenotypes that changed drug prescription. Each individual carried in average 2.2 alleles and 93% of the cohort could be informed of at least one actionable pharmacogenetic phenotype. Differences in variant allele frequency were observed among the populations studied and the corresponding gnomAD population for 9.4% of the variants. Regarding novel rare variants, we uncovered 453 in 34 PharmGKB Very Important Pharmacogenes.

Conclusion: We provide with the landscape of pre-emptive actionable pharmacogenetic information using diagnostic WES data, together with population-specific allele variations.

Automatic information search for countering COVID-19 misinformation through semantic similarity

Álvaro Huertas García[1].

[1]Universidad Politécnica de Madrid.

Information quality in social media is an increasingly important issue and misinformation problem has become even more critical in the current COVID-19 pandemic, leading people exposed to false and potentially harmful claims and rumours. Civil society organizations, such as the World Health Organization, have demanded a global call for action to promote access to health information and mitigate harm from health misinformation. Consequently, this project pursues countering the spread of COVID-19 infodemic and its potential health hazards.

In this work, we give an overall view of models and methods that have been employed in the NLP field from its foundations to the latest state-of-the-art approaches. Focusing on deep learning methods, we propose applying multilingual Transformer models based on siamese networks, also called bi-encoders, combined with ensemble and PCA dimensionality reduction techniques. The goal is to counter COVID-19 misinformation by analyzing the semantic similarity between a claim and tweets from a collection gathered from official fact-checkers verified by the International Fact-Checking Network of the Poynter Institute. It is factual that the number of Internet users increases every year and the language spoken determines access to information online. For this reason, we give a special effort in the application of multilingual models to tackle misinformation across the globe. Regarding semantic similarity, we firstly evaluate these multilingual ensemble models and improve the result in the STS-Benchmark compared to monolingual and single models. Secondly, we enhance the interpretability of the models' performance through the SentEval toolkit. Lastly, we compare these models' performance against biomedical models in TREC-COVID task round 1 using the BM25 Okapi ranking method as the baseline. Moreover, we are interested in understanding the ins and outs of misinformation. For that purpose, we extend interpretability using machine learning and deep learning approaches for sentiment analysis and topic modelling. Finally, we developed a dashboard to ease visualization of the results.

In our view, the results obtained in this project constitute an excellent initial step toward incorporating multilingualism and will assist researchers and people in countering COVID-19 misinformation.

Keywords: Natural Language Processing, Machine Learning, Deep Learning, Transformers, Topic Modeling. Sentiment Analysis, Semantic search, COVID-19, Infodemic, Misinformation, Twitter, Factchecking

A systematic review and meta-analysis of transcriptomic studies unveil sex differences in rheumatoid arthritis

Edurne Urrutia-Lafuente[1], Irene Pérez-Díez[1, 2], Francisco García-García[1].

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012 Valencia, Spain.
[2]Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012 Valencia, Spain.

Corresponding author e-mail address: fgarcia@cipf.es

Autoimmune diseases, such as rheumatoid arthritis (RA), show sex-based differences in epidemiological and clinical features. According to the literature, there are disparities in the immune system of men and women. Although sex hormones are known to be involved in the process, the differences in the molecular mechanisms are not entirely understood. A greater knowledge of these mechanisms could lead to more accurate diagnostics and sex-specific treatments. Therefore, the aim of this study was to determine and understand the molecular mechanisms underlying sex differences in rheumatoid arthritis, to find prospective biomarkers that could foster personalised medicine.

This objective was accomplished in three steps. First, we conducted a systematic review of rheumatoid arthritis studies with gene expression microarray or RNA-sequencing data from patients and healthy individuals. Second, the selected studies were analysed individually. This step included an exploratory analysis, a differential expression analysis and a functional enrichment analysis of biological processes from the Gene Ontology (GO) and KEGG pathways. Finally, we performed a meta-analysis of genes using a random effects model, followed by functional enrichment of the overall results.

As a result, five datasets with 570 samples were retrieved from the Gene Expression Omnibus (GEO) repository. After the meta-analysis, 10 genes, 90 biological processes and 11 KEGG pathways were identified as significantly sex-dependent. Furthermore, the results were synthesised using the REVIGO tool and GO slim terms. The enrichment results suggest a higher immune system activity and a larger presence of energy precursors in female RA patients, among others. Male RA patients, in contrast, showed an overrepresentation of terms related to protein biosynthesis and negative regulation of osteoblasts.

Our findings may explain some of the sex differences in rheumatoid arthritis. Thus, they could lead to the discovery of new therapeutic targets and be the basis for future research devoted to sex-specific precision medicine.

Keywords: Autoimmune diseases, Sex characteristics, Gene expression, Meta-analysis, Precision medicine

A single-cell and single-nuclei transcriptomics approach reveals novel sex differences in multiple sclerosis

Irene Soler-Sáez[1], Zoraida Andreu[2], Almudena Neva-Alejo[1], José F. Català-Senent[1], Adolfo López-Cerdán[3], María de la Iglesia-Vayá[3], Francisco García-García[1].

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012 Valencia, Spain.
[2]Foundation Valencian Institute of Oncology (FIVO), 46009, Valencia, Spain
[3]Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012 Valencia, Spain.

Corresponding author e-mail address: fgarcia@cipf.es

Multiple sclerosis (MS) is a chronic autoimmune disease that corresponds to the first cause of non-traumatic disability among young adults. It is characterized by the damage of myelin (the lipidic layer that surrounds neuronal axons) induced by defective immune responses. As a result, patients suffer axonal injury and neuronal loss, which lead to the neurodegeneration of the central nervous system.

It has been reported sex bias in several epidemiological and clinical features of the disease. To reveal the molecular mechanisms underneath those differences, we propose an in silico analysis for processing transcriptomic data coming from single cells and single nuclei. This novel approach to find sex-based variability in MS enables the description of prospective biomarkers for every cell type evaluated, enhancing personalized and precision medicine from a sex perspective.

The approach can be divided into three steps. Firstly, we have performed a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement to select the input datasets. Secondly, we have developed pipelines to find statistical differences in cell type abundance and gene expression levels for each dataset independently. Finally, we have integrated the gene expression results to unveil a compromise among datasets and cell types. Those genes have allowed us to perform overrepresentation analysis to describe the biological processes that exhibit alterations by sex in MS.

Each of the datasets evaluated have provided samples from the nervous system or the immune system. In both of them, the analysis has revealed significant sex differences in cell type abundance as well as in several gene expression patterns. The integration of the latter results have shown commonly upregulated genes for each sex among cell types and for each cell type among datasets. Moreover, some of those genes represented biological functions that could be affected in a sex dependent manner.

Keywords: Multiple sclerosis, sex bias, gene expression, single cell, precision medicine.

Software assessment for prioritization of germline causal variants of disease on whole-exome sequencing data

Eva Tosco-Herrera[1], Alejandro Mendoza-Alvarez[1], Hector Rodriguez-Perez[1], Adrian Muñoz-Barrera[2], Antonio Iñigo-Campos[2], Almudena Corrales[1,3], Francisco Martinez Bugallo[4], Carol Prieto Morin[4], Rafaela González-Montelongo[2], Jose Miguel Lorenzo-Salazar[2], Carlos Flores[1,2,3], Laura Ciuffreda[1], Itahisa Marcelino-Rodriguez[1,5].

[1]Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife; [2]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife; [3]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid; [4]Clinical Analysis Service, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife; [5]Instituto de Tecnologías Biomédicas (ITB), Universidad de La Laguna, Santa Cruz de Tenerife, Spain.

Introduction: Whole-Exome Sequencing (WES) experiments analyze DNA sequences from protein-coding regions of the genome, where more than 80% of pathogenic and causal variants of Mendelian diseases are located. Frequently, WES provides a large number of variants per sample (15,000-25,000 variants). Despite causal variant prioritization is necessary for clinical diagnosis, bioinformatic standards are lacking. Here we benchmarked free software tools for variant prioritization of causal variants from empirical WES data.

Materials and methods: A total of 62 WES data from patients with a known genetic diagnosis were obtained using a HiSeq 4000 Illumina® system, using 75 bp paired-end reads and a 100 bp padding to capture the exon-flanking regions. Consistent quality controls and BWA-GATK v3.8 Best Practices were followed for germline variant identification using the GRCh37/hg19 human genome as reference. Whenever necessary, the Human Phenotype Ontology terms were manually inspected according to the declared phenotypes. Different parameters were considered for the performance assessment of nine software tools.

Results and conclusions: Five of the tools were fully evaluated, because of technical limitations and installation issues in the others. Based on conservative first-gene rankings, the highest diagnostic yield was found for Exomiser (69.3%), followed by AMELIE (46.7%), and Tapes (19.3%). In conclusion, the tools for variant prioritization provided largely different diagnostic yields, with Exomiser being the best performing tool.

Funding: Ministerio de Ciencia e Innovación (RTC-2017-6471-1; AEI/FEDER, UE); ITER agreement OA17/008; FIISC (FIIS19/48); ACIISI (TESIS2020010002); Instituto de Salud Carlos III (CD19/00231); ECIT CGIEU0000219140.

perSVade: personalized Structural Variation detection in your species of interest

Miquel Àngel Schikora-Tamarit[1,2] and Toni Gabaldón[1,2,3].

[1]Barcelona Supercomputing Centre (BSC-CNS). Jordi Girona, 29. 08034. Barcelona, Spain.
[2]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain.
[3]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain.

Structural variants (SV) such as translocations, inversions, deletions, and other genomic rearrangements can contribute significantly to genetic and phenotypic variability across many species. The role of SV has been traditionally overlooked due to the technical limitations of SV detection and interpretation from short-read sequencing datasets. Most available algorithms yield low recall when tested on humans, but few studies have investigated the performance in non-human genomes. Similarly, there are no specific indications about what parameters should be used for SV calling for most species. It is unclear whether the accuracy of each algorithm and running parameters validated on model species work equivalently for other species. In order to fill this gap we have developed perSVade (personalized Structural Variation Detection), a pipeline that identifies and annotates SVs in a way that is optimised for any input sample. Starting from a set of paired-end whole-genome sequencing reads, perSVade uses simulations on the reference genome to choose the best SV calling parameters. The output includes the optimally-called SVs, a report of the accuracy and a friendly graphical interface that shows the SVs on a genome browser. In addition, perSVade allows the calling of small variants and copy-number variation. In summary, this pipeline is useful to identify several types of genomic variation in short reads using a single bash command. We validated perSVade using simulations and real SVs for six diverse eukaryotic organisms. We consider that this tool will help to understand how SVs generate phenotypes across eukaryotes.

Keywords: structural variants, genomics, parameter optimization

tRNAstudio: analysis of tRNA-seq datasets without the need of programming skills.

Marina Murillo Recio[1], Ignacio Miguel Martínez de Lejarza Samper, Lluís Ribas de Pouplana and Adrian Gabriel Torres.

[1]Institute for Research in Biomedicine (IRB)

Transfer RNAs (tRNAs) are non-coding RNAs that bring amino acids to the ribosome and are thus essential for protein synthesis. tRNAs must undergo a maturation process in order to become functionally active. tRNA genes are transcribed as longer tRNA precursors (pre-tRNAs) that are processed and chemically modified post-transcriptionally in order to obtain mature tRNA sequences. These chemical modifications not only have an impact at structural and functional levels but also contribute to translation regulation. Mature and precursor tRNAs can also be further processed into tRNA-derived fragments (tRFs) that perform a wide range of roles beyond protein synthesis.

Alterations in tRNAs and in the enzymes that are responsible for tRNA biogenesis, modification and processing are related to complex diseases such as cancer, type 2 diabetes, neurological dysfunctions and rare pathologies. However, their molecular role in disease is currently poorly understood. In this sense, studying the changes in tRNA expression that can occur in different tissues, and further understanding the multiple functions of tRNAs and tRFs, could contribute to characterizing how alterations in tRNA biology affect the development of diseases.

It is possible to study tRNA biology by means of high throughput sequencing of tRNAs (tRNA-Seq). However, the presence of tRNA modifications, the sequence similarity between different tRNAs, and the large number of tRNAs encoded in the genome impair the interpretation and downstream analysis of tRNA-Seq datasets. Therefore, processing of tRNA-seq datasets require strong bioinformatics skills that are frequently not available in experimental laboratories.

Here we present tRNAstudio, a fully automated bioinformatics pipeline to analyse tRNA-Seq datasets that has been packaged into a user-friendly graphical user interface (GUI) that users can implement in macOS and linux-based platforms, without the need of programing skills. This tool detects, classifies and characterizes tRNA sequencing reads and can be used to extract information on tRNA expression, processing and post-transcriptional modification patterns. tRNAstudio provides the user with an interactive html summary report and extensive graphical representations of the data, including tRNA expression, relative proportion of reads derived from precursor and processed tRNAs, detection and quantification of tRNA modifications, and tRNA gene coverages that can aid in the identification of bona fide tRFs or of modified tRNA residues.

In addition, results are generated as spreadsheet tables that can be used for additional customized analyses.

This work brings bioinformatics closer to experimental laboratories and will help in expanding the knowledge on the role of tRNA biology in physiological and pathological scenarios.

In-silico characterization of prion-like domains in human protein families

Siddhant Sharma[1]*, Anjali Sharma1, Ayush Panda[1], Hameed Sanusi1, Jaskeerat Singh[1], Joy Olu-Benson[1], Maurya Sharma[1], Mohammed Emon[1], Mosadoluwa Agbonyin[1], Naomi Ndifon[1], Shruti Poojary[1], Sophie Fang[1], U.Favour Ugochi[1], Umasri Sankarlal[1], Valerii Pokrytiuk[1].

[1]The HackBio, Lagos, Nigeria
*siddhaantsharma.ss@gmail.com

Keywords: Prion-like Domains, Proteomics

The propagation of known prions relies on a protein fragment called the prion domain (PrD). These domains are often enriched in glutamine and asparagine residues and are usually found in the disordered fragments of proteins. Leveraging these principles, bioinformatic algorithms can be used to predict novel PrDs. The aim of this study is to uncover the most important PrLD candidates in classes of relevant human proteins with a greater likelihood of exhibiting prion-like behaviour. We collated 3342 reviewed human protein fasta files as a dataset from UniprotKB. The majority of prion proteins are marked by glutamine/asparagine-(Q/N-) enriched prion-forming domains, which are essential as well as sufficient for propagation. The collected protein fasta files were processed through bioinformatic algorithms such as prion-like amino acid composition - PLAAC on default background probability settings. The results were compiled together after placing a threshold filter of CORE_Score >15 from PLAAC. We identified 65 suspected candidates showing prion-like behaviour. The data was further passed through the PrionW tool to get precise locations for present prion-like domains (PrLD). Further verification of PLAAC positive candidates was done by using the MobiDB database for confirmed protein misfolding disorder predictions. Final ground-truthing was done using the pRANK algorithm for selected PrLDs in our study. Exploratory analysis of the data generated was done by mapping the PrLDs found in our study to known human gene-disease associations by the DisGeneNetR package and linking our candidate proteins having a CORE_Score ≥30 to a database of known protein-protein interactions through the ComPPI tool. Out of 65 candidates, we found 13 candidates with a CORE_Score ≥30 and a protein-disorder match on the MobiDB database confirming them as PrLD candidates. The pRANK analysis confirmed our PrLD candidates as lower the pRANK score, the more prion-like protein it resembles. The phylogenetic trees generated by iTOL for PLAAC and PrionW result-candidate proteins aid in the elucidation of the evolutionary closeness of human protein families and the CORE scores. Protein localization through ComPPI elucidates the differential concentration of the confirmed 13 PrLD's across their subcellular environments. The localization information can significantly improve protein targeting during drug development.

Unraveling microbial community assembly from 16S data

Silvia Talavera Marcos[1].

[1]Department of Biology, Area of Genetics, Universidad Autónoma de Madrid

Keywords: Microbial communities, Bioinformatics, Microbiome, Community assembly, Computational biology, Functional Annotation, Metabolic Modeling

In every ecosystem, be it macroscopic or microscopic, a series of niches exist that are occupied by different species and are responsible for selection. If a given species does not have the ability of surviving in any of the available roles when it arrives at the community, it will disappear because of negative selection. If more than one species has that ability, then they are said to “compete” for that niche – for instance, if the niche is glucose consumption only the species capable of consuming it at the highest rate will prevail.

In my master's thesis I focused my attention on some experimental data by Goldford et al. (2018) where my laboratory had detected two predominant PCGs belonging to the Pseudomonadaceae and Enterobacteriaceae families. I generated SBML metabolic models of members from those PCGs and analyzed them with bioinformatic tools like Smetana and ReFramed to observe the metabolites they were able to exchange.

As part of my doctoral thesis, I continued my studies on these data by trying to simulate the growth of each PCG on M9 medium and on the spent medium of the other. Growing Pseudomonadaceae on Enterobacteriaceae spent medium yielded the same growth rates ratio as seen experimentally.

Assuming the validity of these models, I examined their metabolic properties by using the EggNOG-mapper gene annotation tool and custom Python scripts. I obtained lists of enzyme descriptors that I parsed with my own pipeline, thus discovering some key differences between the metabolism of these PCGs that showed why Enterobacteriaceae initially had an advantage over Pseudomonadaceae in the proposed glucose medium.

In addition to this, to further understand why this relationship between PCGs had been established I generated Python metabolic graphs of some Pseudomonadaceae-Enterobacteriaceae pairs. By highlighting those metabolic functions unique to each of the PCGs and those that had an opposite sign flux, I clearly observed the relationship between these two PCGs - Enterobacteriaceae’s role is to consume glucose and Pseudomonadaceae’s is to consume the acetate produced by Enterobacteriaceae. This relationship has already been observed experimentally, which demonstrates that the study of metabolic models in the dry-lab, framed in our theoretical framework, can substitute to some extent wet-lab assays to understand real dynamics in microbial communities.

FLYCOP web, an user-friendly server for Synthetic Microbial Consortia optimization

David San León Granado and **Juan Nogales Enrique**.

Systems Biology, National Center of Biotechnology (CNB CSIC), Madrid, Spain

During the last century, a single strain of microbial species has been designed and engineered to produce multiple compounds or increase the efficiency of reproducible natural procedures. When the goal complexity is too high, this approach starts to be limited and Synthetic Microbial Consortia (SMC) are replacing engineered single strains to solve the limitations. This new paradigm generates a major challenge, due to the large amount of data that needs to be integrated – microbial metabolisms and their interactions-. There are few tools trying to approach these complex problems; one of them is Flexible sYnthetic Consortium Optimization (FLYCOP) (García-Jiménez at al, 2018), a computational framework able to simulate multiple scenarios in silico to detect the best configuration for a consortium when a purpose is defined. While FLYCOP is a quite flexible and powerful tool, it is very complex to configure. This work shows a simplified and user-friendly version of FLYCOP to reduce the configuration complexity adding a web interface and a toolbox of automatic functions. This new computational layer includes new functionality to FLYCOP and allows less experienced users the identification of important parameters and optimizations of their own consortia.

robustica: customizable robust independent component analysis

Miquel Anglada-Girotto[1], Luis Serrano[1,2,3]*, Sarah A. Head[1*].

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain.
[2]Universitat Pompeu Fabra (UPF), Barcelona, Spain.
[3]ICREA, Pg. LLuís Companys 23, Barcelona 08010, Spain.
*Corresponding authors: luis.serrano@crg.eu and sarah.dibartolo@crg.eu

Independent Component Analysis (ICA) allows the dissection of omic datasets into modules that help to interpret global molecular signatures. The inherent randomness of this algorithm can be overcome by clustering many iterations of ICA together to obtain robust components. Existing algorithms for robust ICA rely on the choice of clustering method and on computing a potentially biased and large Pearson distance matrix. We present robustica, a Python-based package to compute robust independent components with a fully customizable clustering algorithm and distance metric. We exploited its customizability to evaluate the performance of 6 popular clustering algorithms and to show how our subroutine to infer the components' signs across ICA iterations enables us to perform robust ICA using Euclidean distance while simultaneously increasing the precision, speed, and memory efficiency of the algorithm. We show its applicability by dissecting over 500 tumor samples from low-grade glioma (LGG) patients, where we define a new gene expression module with the key modulators of tumor aggressiveness downregulated upon IDH1 mutation.

Neutralizing antibodies as a therapeutic avenue for COVID-19

Farriol-Duran R[1], López-López R[2], Fernandez-Barat L[*2], Porta-Pardo E[*1,3].

[1]Barcelona Supercomputing Center
[2]Applied Research in infectious Respiratory Diseases and critically ill patients, Cirberes-IDIBAPS, Hospital Clinic, School of Medicine (University of Barcelona).
[3]Cancer Immunogenomics, Institute Josep Carreras for Leukaemia Research
[*]To whom correspondence should be addressed: lfernan1@clinic.cat, eporta@carrerasresearch.org
Funding: La Caixa Junior Leader Fellowship (LCF/BQ/PI18/11630003) BIOCAT XXX

COVID-19, the multi systemic disease caused by the virus SARS-CoV-2, has been responsible for an unprecedented pandemia during 2020 and 2021. Due to the novelty of the virus itself and the multi-organic affectations of the disease, COVID-19 treatment has been a challenge. In that scenario, both the scientific and political communities have focused their attention on this disease. As a result, the first vaccination strategies have been available in a record time. However, it is likely that COVID-19 will continue in our societies for the coming years, so it is not only important to develop prophylactic vaccines but to also design therapeutic strategies for severe patients.

In this project, we have developed a computational pipeline to predict the viral epitopes most likely to be recognised by neutralising antibodies, hence the targets of viral neutralization. Our pipeline includes both predictions of linear and structural epitopes, which are then prioritised according to multiple variables that influence the interaction between neutralizing antibodies and their epitopes: glycosylation status, localization on the viral membrane and accessibility in the 3D conformation of their parental protein. Additionally, we have developed a set of bioinformatic tools to manage and filter the predicted epitopes while increasing the pipeline's throughput to make a full-SARS-CoV-2-proteome epitope prediction feasible. We have then synthesised peptides spanning multiple epitopes and used these peptides to identify monoclonal antibodies in the serum of COVID-19-recovered patients. The final goal is to use these monoclonal antibodies to treat severe COVID-19 patients.

As a final layer, due to the importance of the variants of concern (VOCs), we have analyzed how the mutations these variants carry may affect the regions recognized by neutralizing antibodies.

In summary, we present a prioritised list of epitope candidates predicted from the whole SARS-CoV-2 proteome which will sequentially be tested against serum from COVID-19 recovered patients. Our in vitro data shows that the candidate epitopes predicted for the viral protein Spike have notable recognition capacities, suggesting that our pipeline is able to yield immunogenic epitopes.

Transcriptomic exploration of the epileptogenic zone of drug-resistant temporal lobe epilepsy patients

Patricia Sánchez-Jiménez[1], Marcos Elizalde-Horcada[2], Gemma Benito[1], Javier Fraga[3], Inmaculada Granero-Cremades[4], Cristina Virginia Torres[5], María de Toledo[6], Paloma Pulido[5], Marta Navas[5], Francisco Abad-Santos[7,8], Antonio Gómez-Martín[9,10], Paolo Maietta[1], María del Carmen Ovejero-Benito[7,11].

[1]NIMGenetics Genómica y Medicina S.L., Madrid.
[2]Unidad de Bioinformática, Plataforma de apoyo a la investigación, Instituto de Investigaciones Sanitarias Pere Virgili (ISS-PV), Reus, Tarragona.
[3]Departamento de Anatomía Patológica, Hospital Universitario de La Princesa, Madrid.
[4]Departamento de Análisis Clínico, Hospital Universitario de La Princesa, Madrid.
[5]Departamento de Neurocirugía, Hospital Universitario de La Princesa, Madrid.
[6]Departamento de Neurología, Hospital Universitario de La Princesa, Madrid.
[7]Departamento de Farmacología Clínica, Hospital Universitario de La Princesa, Instituto de Investigaciones Sanitarias la Princesa (IIS-IP), Madrid.
[8]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid.
[9]Área de Epidemiología y Salud Pública, Instituto de Investigaciones Biosanitarias (ibs-GRANADA), Granada.
[10]Grupo de Investigación en Salud Ambiental e Infancia, Escuela Andaluza de Salud Pública (EASP), Granada
[11]Departamento de Farmacología y Ciencias de la Salud, Escuela de Farmacia, Universidad San Pablo CEU, Alcorcón, Madrid.

INTRODUCTION: Epilepsy is a chronic brain disease that affects 0,5% of the world population. One-third of epilepsy patients do not respond well to pharmacological treatment, presenting drug-resistant epilepsy (DRE). It is an acquired disease and, once it appears, it is independent of the antiepileptic drug treatment. Moreover, DRE prevalence is higher in temporal lobe epilepsy patients (TLE). The neurosurgery resection of the epileptogenic zone is an efficient therapeutic approach for DRE patients. Thus, it is crucial to understand the molecular mechanisms underlying drug resistance in TLE patients. For that purpose, we perform a transcriptomic study of the epileptogenic zone of drug-resistant temporal lobe epilepsy patients.

MATERIALS AND METHODS: RNA-seq study was accomplished with 82 hippocampus RNAs extracted from formalin-fixed paraffin-embedded (FFPE) samples: 46 drug-resistant TLE patients neurosurgical resected from Hospital Universitario de La Princesa and 36 controls from post-mortem brain donors of four biobanks of the National Biobank Network (Navarrabiomed, IMIB, IDIBAPS, HUPHM). Quality control and quantification were performed before and after library preparations. Libraries were sequenced to an average depth of 100 million total reads (100 bp pair-end) on the Illumina NovaSeq™ 6000 platform. Raw reads were aligned to the GRCh37/hg19 reference genome. Differential gene expression (DEGs) was generated comparing the drug-resistant TLE patient expression patterns to control expression patterns. The genes with the highest differential expression were validated by RT-qPCR.

RESULTS: The significant differential expression genes applying the false discovery rate (FDR) were 8,773 (padj>0.05): 3,959 genes were upregulated, such as *TNFRSF10C, CPLX3, CRHR2;* and 4,774 genes were downregulated, such

as *TUBB3, FOXN1, PITX2*. Moreover, we made an enrichment in gene ontology (GO), and we found 24 GO-IDs, for example, synaptic plasticity regulation and neurotransmitter secretion. We also enriched DEGs in pathways, and we found 6 KEGG pathways, for instance: shigellosis and axon guidance.

CONCLUSIONS: The transcriptomic approach of the epileptogenic zone of drug-resistant temporal lobe epilepsy patients produced 8,773 differential expression genes. These results may help us to understand the neurological mechanisms involved in pathological processes. Furthermore, we contribute to improving the knowledge on drug resistance and the identification of therapeutic targets.

KEYWORDS: Drug-Resistant Epilepsy (DRE), Temporal Lobe Epilepsy (TLE), Transcriptomics, RNA-seq, Formalin-Fixed Paraffin-Embedded (FFPE).

Regulation of Transposable elements in aging and during cellular rejuvenation

Fernando Jose Galvez Sanchez[1,2], Diana Guallar Artal[2].

[1]Institut of Health Carlos III; [2]Center for research in molecular medicine and chronic diseases.

Ageing is the greatest risk factor for most human pathologies, resulting in a lower quality of life of elderly people, and a great burden for health care systems. Therefore, it is critical to dissect the molecular underpinnings behind the aging process in order to develope new therapies that could achieve or restore a healthy ageing. Transposable elements (TE) are repetitive sequences composing 45% of the human genome. In spite of their potential detrimental effects in the host genome if they are mobilized, a wealth of studies has demonstrated that TEs have been also pervasively coopted as regulatory elements for host genes, thus playing an important role in transcriptome regulation. During ageing, several families of TEs have been shown to be transcriptionally deregulated, although the specific impact on the host is yet to be elucidated. Due to the difficulties in identifying these sequences at the bioinformatic level according to their repetitive nature, few studies have studied the deregulation of these elements at the locus level in the aging process. Here, using an in silico software named SQuIRE, we will analyze the transcriptional deregulation of TEs during ageing. In particular, we will analyze the RNA-seq datasets of young and old fibroblasts (from 2 or 28 months-old mice, respectively) generated in our laboratory, to detect transcriptional changes of TEs at the class, family and element level. Furthermore, we will analyze the transcriptional dynamics of TEs during cellular rejuvenation of fibroblasts through somatic cell reprogramming with the Yamanaka factors, using RNA-seq samples deposited in public databases such as Gene Ontology (GEO) and European Nucleotide Archive (ENA). Understanding the contribution of TE deregulation to loss of cellular fitness during ageing, as well as its restoration by cellular rejuvenation will be key to identify novel molecular pathways which could be attractive druggable targets to achieve a healthier ageing.

Dynamic A-to-I RNA Editing Landscapes during Aging Process

David Martínez[1,2], Alejandro Fuentes-Iglesias[1], Vera Garcia-Outeiral[1], Miguel Fidalgo[1].

[1]Stem Cells and Human Diseases Group, Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela (USC), Santiago de Compostela, Spain; [2]Escuela Nacional de Sanidad (ENS), Instituto de Salud Carlos III, Madrid, Spain

dvidmd97@usal.es

Due to the continuous increase in life expectancy in advanced societies, aging represents the major risk factor for most human diseases including cancer, cardiovascular disorders, metabolic disorders, and neurodegenerative processes. Although the molecular causes of aging remain elusive, several evidences from yeast to human suggest that deregulation of chemical modifications within DNA and histones seems to be a major driver during the aging process. In contrast, little is known about the potential role of RNA chemical modifications in aging. RNA editing of adenosine to inosine (A-to-I), catalyzed by ADAR enzymes, is among the most abundant RNA modifications in mammalian cells. While A-to-I RNA modification can greatly affect the cellular transcriptome and has been involved in several human disorders, its functional roles in cellular aging are largely unexplored. Here, we have taken advantage of the fact that inosine is a biological mimic of guanosine to develop an optimized in silico approach to address the distribution and dynamics of this RNA modification during cellular aging. In sum, we have performed the analysis employing different bioinformatic software like STAR, to map and align the genome of the samples to the reference, and REDItools for RNA editing detection. By using both our own RNA-seq datasets of fibroblasts derived from mice at different ages obtained through high-throughput sequencing along with available datasets related to the aging process (i.e., skin), we were able to identify that A-to-I RNA editing is dynamically altered during organismal and cellular aging, with a greater number of RNA edited sites in older individuals. Notably, we found that A-to-I RNA editing profiles are sufficient to segregate samples according to their age. Together, our results yield important information about how aging occurs, and the important role of A-to-I editing in this process and open the possibility for new therapeutic approaches based on RNA modifications interventions to ameliorate aging associated disorders and improve human healthspan.

ACKNOWLEDGEMENTS:

Research from M.F. laboratory was supported by the Spanish Agencia Estatal de Investigación (PID2019- 105739GB-I00). M.F. is recipient of a Ramón y Cajal award (RYC-2014-16779) from the MINECO of Spain. A.F.-I. (MINECO, BES-2017-082007), and V.G.-O. (Ministerio de Ciencia, Innovación y Universidades, FPU17/01131) are recipients of fellowships.

Investigating epitranscriptomics as novel regulatory mechanisms involved in healthy aging in mice and humans

Alba Coego[1,2] & Diana Guallar[1,2].

[1]Center for Research in Molecular Medicine and Chronic Diseases (CiMUS); [2]Department of Biochemistry and Molecular Biology, USC, Santiago de Compostela 15782. Spain

CONTACT: alba.cortes.coego@usc.es

Aging is one of the greatest risk factors for most of human diseases and represents a social and economic burden both in Spain and the European Union. Aging is an unavoidable consequence of life and is characterized by a progressive loss of physiological integrity leading to a loss of function of cells and tissues, leading to disease and death. Despite the great amount of knowledge generated about how aging manifests at the cell and tissue levels, a comprehensive understanding of its molecular underpinnings is missing, thus limiting efficient rejuvenating or alleviating anti-aging therapies. Accumulating evidence supports a model where the transcriptome becomes less tightly regulated throughout the aging process. Indeed, a progressive degradation of transcriptional networks robustness and integrity has been observed during aging in several model organisms, including human and mouse tissues. Given that chemical RNA modifications (a.k.a. Epitranscriptomics) can regulate all stages of transcript metabolism, we propose to undertake an *in silico* approach to identify epitranscriptomic regulators (i.e. readers, writers and erasers) whose expression is altered during healthy aging. In particular, we have focused on the four most relevant RNA modifications in mammals: methylation of adenosines (m^6A) and cytidines (m^5C), hydroxymethylation of cytidines (hm^5C) and adenosine-to-inosine (A-to-I) editing. We have explored next generation sequencing RNA seq datasets of human and mouse fibroblasts of different ages, to identify the epitranscriptomic pathways related to m^6A / m^5C / hm^5C / A-to-I most altered during aging. Furthermore, we have carried out gene ontology analysis to explore conserved pathways altered with aging that could be regulated through epitranscriptomic modulation. The identification of conserved epitranscriptomic pathways associated to aging could represent exciting druggable pathways for extending a healthy lifespan. Understanding the regulatory mechanisms leading to transcriptional defects during aging will be key to identify novel regulatory targets whose modulation could ultimately reverse or delay the onset of age-related diseases.

ACKNOWLEDGEMENTS

This research was funded by the Spanish Agencia Estatal de Investigación, co-funded by the FEDER Program of the EU (RTI2018-096708-J-I00/AEI/FEDER,UE to D.G.), and the 2020 Leonardo Grant for Researchers and Cultural Creators BBVA Foundation (2020-PO041 to D.G.). D.G. is a recipient of a Ramón y Cajal Award from the MICI (RYC2019-027305-I). C.P. (Xunta de Galicia, ED481A-2020/187), AC (Xunta de Galicia, ED481A-2021/053)

Lipidomic profile studiy of extracellular vesicles to identify differences in adolescents exposed to alcohol intoxication with sex perspective

Carla Perpiñá-Clérigues[1,2], José F. Català-Senent[1], Francesc Ibáñez[3], Consuelo Guerri[3], María Pascual-Mora[2], Francisco García-García[1].

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012, Valencia, Spain.
[2]Department of Physiology, School of Medicine and Dentistry, University of Valencia (UV), 46003, Valencia, Spain.
[3]Cellular and Molecular Pathology of Alcohol, Principe Felipe Research Center (CIPF), 46012, Valencia, Spain.

Corresponding author e-mail address: fgarcia@cipf.es

Extracellular vesicles (EVs) have recently been discovered to have a crucial function in pathological processes including inflammation. Lipids in these vesicles could participate in their biogenesis, release and invagination in the target cells. Considering the ability of EVs to cross the blood-brain barrier, being good non-invasive biomarkers, the aim of this study is to evaluate if alcohol intoxication plays a relevant role at the lipidomic level in the extracellular plasma vesicles of adolescents and its possible differential profile by sex. To answer this question, we use plasma EVs vesicles from human adolescent males and females after an acute ethanol intoxication and their respective control individuals. Furthermore, a novel bioinformatic approach has been carried out through differential *expression* analysis and functional characterization of lipidomic data. The results show that there are significant differences in the lipidomic profile between plasma EVs from adolescent females and males, as well as significant differences in the lipid content of plasma EVs from intoxicated and control groups. Furthermore, the functional characterization shows lipid classes (mainclass *Fatty Acids* and its corresponding superclass *Fatty Acyls*) upregulated in adolescent females after an alcohol intoxication, compared to non-alcohol females. These classes are related to the inflammatory immune response supporting previous results from our group, demonstrating that adolescent females are more vulnerable to the effects of alcohol (e.g., higher expression of proinflammatory molecules than males). Therefore, we can conclude that the lipids of the plasma EVs could be good candidates as biomarkers and can help us to explain the mechanisms underlying the neuroinflammatory response after acute alcohol intoxication.

Keywords: extracellular vesicles, alcohol, lipidomics, differential expression, LSEA, sex differences

Metabolomic study in rats with chronic hyperammonemia. Implications in minimal hepatic encephalopathy

Rodriguez-Hernandez, Z[1,2,5]; Sonia Tarazona[3,6]. Marta Llansola[4,7].

[1]Department of Chronic Diseases Epidemiology, National Center for Epidemiology, Carlos III Health Institute, Madrid, Spain.
[2]Department of Biotechnology, Universitat Politècnica de València, Valencia, Spain.
[3]Department of Applied Statistics, Operations Research and Quality, Universitat Politècnica de València, Valencia, Spain.
[4]Laboratory of Neurobiology, Centro Investigación Príncipe Felipe, Valencia, Spain.
[5]zulema.rodriguez@isciii.es
[6]sotacam@eio.upv.es
[7]mllansola@cipf.es

Introduction: Hyperammonemia produces neurotoxicity, neurocognitive disorders and impaired motor function, contributing to the development of Minimal Hepatic Encephalopathy. However, it is not well understood the mechanisms underlying these effects of hyperammonemia. The aim of this study was to use a metabolomic aproach to study the effects of chronic hyperammonemia on metabolite levels, both in plasma and in different brain areas. **Material and methods:** Plasma samples from control and hyperammonemic rats at two and four weeks after the diet was administered, as well as different brain tissues (cerebellum, hippocampus and cerebral cortex) at 4 weeks after the diet, were obtained. A targeted metabolomic analysis was performed with these samples. **Results and discussion:** Differential analysis identified metabolites with significant changes due to hyperammonemia in each tissue: glutamine, glutamate, glycine, serine, histamine, asymmetrical dimethyl-arginine and serotonin, which are very important in neurotransmission. However, these changes were uneven between different areas of the brain. The temporal study of the plasma revealed that there are more alterations in the level of the metabolites at four weeks after having administered the diet than at two weeks, which shows the progression of chronic hyperammonemia. Finally, multivariate regression analysis between plasma and cerebellum data revealed that plasma serotonin and carnosine could serve as biomarkers to detect changes in the cerebellum due to hyperammonemia. **Conclusion:** Hyperammonemia produces specific metabolite alterations in each tissue. This study suggests new biomarkers and new pathways involved in hyperammonemia and Minimal Hepatic Encephalopathy, which should be further studied in future researches.

Keywords: Hyperammonemia; Minimal Hepatic Encephalopathy; Metabolomics; neurotransmitters; Partial Least Squares Regression; nitric oxide-cGMP pathway

Mouse Skin Cell Atlas

Aina Rill[1].

[1]Josep Carreras Leukaemia Research Institute

The skin is the primary barrier of our body against pathogens and physical injuries. It consists of three main layers with distinct cellular compositions and functions: the epidermis, mainly composed of keratinocytes; the dermis, a thick multilayered fibrous structure; and the hypodermis, that consists predominantly of fat. Skin homeostasis is carried out by a rich variety of cell types that together orchestrate several biological processes, such as directing the stem cell renewal, lineage commitment

and differentiation. Understanding the mechanisms underlying skin homeostasis and repair as well as aging and other inflammatory diseases, requires an in-depth knowledge of the molecular and functional properties of all the skin cell types.

The use of single-cell technologies has the advantage of detecting cellular heterogeneity of complex tissues, being even able to distinguish smaller subtypes or states, by which high resolution maps of dermal and epidermal populations have already been delineated. However, while a series of studies aimed to identify skin populations at steady or diseased states, a comprehensive molecular characterization of all skin cell types at single-cell resolution has not yet been performed.

Here, we have created a single cell transcriptomic atlas of *Mus musculus* skin by the integration of 8 different studies, including 17 biological samples and over 250k cells, in which we have identified more than 30 cell subpopulations located in the diverse layers of the mouse skin. We discovered cell-specific changes occurring across fibroblasts, which allowed us to classify them in different subtypes that show differential functional annotations. Additionally, we are also developing a versatile online toolbox to explore and download the skin cell atlas, including all characterized cells, to expedite further discoveries in skin biology.

Our mouse skin cell atlas represents a reference dataset that can be used as a basis for understanding fundamental skin biological processes. Jointly with the Human Skin Atlas consortium, we will perform a comparative analysis of the human and mouse skin, that will be fundamental to evaluate the mice as animal model for several biological processes.

Assessment of the generalizability of convolutional neural network co expression analysis for the inference of gene regulatory networks from scRNA seq expression data

Inés Rivero-García[1,2], Miguel Torres[2], Fátima Sánchez Cabo[1,2].

[1]Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. [2]Group of Genetic Control of Organ Development and Regeneration, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain.

Gene regulatory networks (GRN) govern all biological processes, and their inference is fundamental for understanding life and fighting against disease. With big data driving biomedical research, single-cell RNA sequencing (scRNA-seq) and Deep Learning are becoming two key methodologies for GRN inference. Particularly, convolutional neural network co-expression (CNNC) analysis is one Deep Learning-based approach that uses transcriptional co-occurrence images to predict regulatory interactions between pairs of genes. This method leverages the large amount of expression data coming from scRNA seq experiments in order to train and test the model, and it has been found to work well on cell-type specific contexts. Although scRNA-seq protocols are a common practice, it is not always possible to obtain data sets that are large enough for training a cell type specific and deep learning-based GRN inference model. This need calls for an evaluation of the generalisability of the CNNC framework, and whether it is able to predict regulatory interactions in more than one cellular type or context. By using ChIP-seq-annotated images of scRNA-seq expression co-occurrence from differentiated and undifferentiated cell types, we assessed how generalisable is the CNNC algorithm for the inference of regulatory interactions. Our findings suggest that the CNNC method extends well when using data from related cell types, supporting the possibility of training and predicting using different cell populations. Additionally, the CNNC framework is able to detect the immature regulatory status of undifferentiated cells, which show a more continuous read out of interaction probabilities. More importantly, models trained with data from more than one cell type have an equal or better performance than cell type-specific models, pointing towards the potential benefit of pooling different cell types during training. Taken together, these results suggest that the CNNC method may be generalizable across different cell types and conditions, which could be attributed to the conservation of GRN transcriptional co-occurrence signature patterns across different cell types or contexts.

Genetically predicted Telomere Length and its relationship with major Neurological Diseases and Life Expectancy: A Mendelian Randomization Study

Blanca Rodríguez-Fernández[1], Arcadi Navarro[1-5], Juan Domingo Gispert[1,6,7,8], Natalia Vilor-Tejedor[*1-3,9], Marta Crous-Bou[*1, 10, 11].

[1]Barcelonaβeta Brain Research Center - Pasqual Maragall Foundation. Barcelona, Spain.
[2]Centre for Genomic Regulation (CRG). The Barcelona Institute for Science and Technology, Barcelona, Spain.
[3]Universitat Pompeu Fabra (UPF). Barcelona, Spain.
[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.
[5]Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra.
[6]IMIM - Hospital del Mar Medical Research Institute. Barcelona, Spain.
[7]Centro de Investigación Biomédica en Red Bioingeniería, Biomateriales y Nanomedicina, Madrid, Spain.
[8]Centro Nacional de Investigaciones Cardiovasculares (CNIC). Madrid, Spain.
[9]Erasmus University Medical Center, Department of Clinical Genetics, Rotterdam, the Netherlands.
[10]Department of Epidemiology, Harvard T.H. Chan School of Public Health. Boston, MA, USA.
[11]Unit of Nutrition and Cancer, Cancer Epidemiology Research Program. Catalan Institute of Oncology (ICO)-Bellvitge Biomedical Research Center (IDIBELL), Hospitalet del Llobregat, Spain.

Introduction: Telomere length (TL) is a putative biomarker of biological aging. Shorter telomeres have been associated with mortality and increased rates of age-related diseases. However, observational studies are limited to conclude whether TL is causally associated with those outcomes or with related underlying pathological processes. Mendelian randomization (MR) was developed for assessing causality using genetic variants in epidemiological research.

Objective: The objective of this study was to test the potential causal role of TL in neurodegenerative disorders and life expectancy through MR analysis.

Methods: Summary level data were extracted from the most recent genome-wide association studies for TL (N = 78,592), Alzheimer's disease (AD) (N = 455,258), Parkinson's disease (N = 482,730), Frontotemporal dementia (N = 4,131), Amyotrophic Lateral Sclerosis (N = 36,052), Semantic Dementia (N = 924), Progressive Supranuclear Palsy (N = 12,308) and life expectancy (N = 75,244). Causal effects of TL were estimated using the Two-Sample MR

inverse-variance weighted (IVW) method. Effect sizes were reported in Odds Ratio (OR) for all the pathologies and the estimated coefficient (β_{IVW}) for life expectancy. Sensitivity analyses using the MR-Egger method were conducted to test for directional pleiotropic effects. False Discovery Rate corrections for multiple comparisons were performed at significant level 0.05.

Results: MR estimates revealed that longer TL inferred a protective effect on risk of AD (OR = 0.964; adjusted p-value = 0.045). Moreover, longer TL was significantly associated with increased life expectancy (β_{IVW} = 0.011; adjusted p-value = 0.045). Sensitivity analyses suggested evidence for directional pleiotropy in AD analyses.

Conclusions: Preliminary results showed that genetically predicted longer TL may increase life expectancy and play a protective causal effect on AD. The presence of pleiotropy may reflect the complex interplay between TL

homeostasis and AD pathophysiology. Further observational studies are needed to confirm these results.

This work was supported and awarded by the Alzheimer's Association Research Grant (AARG 19-618265). NV-T is funded by a postdoctoral grant, Juan de la Cierva Programme (FJC2018- 038085-I), Ministerio de Ciencia, Innovación y Universidades – Spanish State Research Agency.

Keywords: Life expectancy, Mendelian Randomization, Neurodegenerative diseases, Telomere length

Optimization of molecular dynamic simulations using parallel programming

Miguel Carmena-Bargueño[1].

[1]Bioinformatics and High Performance Computing Research Group (BIO-HPC), Universidad Católica San Antonio de Murcia (UCAM) , Murcia, Spain

Molecular Dynamics Simulations (MDS) is one of the most power tools of the *in silico* techniques. This gadget allows to predict the evolution of the biological system along the time. Using it in conjunction with high performance computing (HPC), we can evaluate the possible interactions between a drugs library against one protein. This method was used by our group with the rCUDA software [1] and the results of this study shown how these types of tools were faster than the classical MDS. However, the main problem of this study was the use of GROMACS; which is a power software but it could be used only by users with high knowledge of programming. Whereas, Desmond [2], which is a software used to perform MDS, applied with Maestro [3] allows to run MDS with a graphical interface, allowing to be used by non-expert users. The aim of our work is to create a method to run MDS using parallel programming with Maestro-Desmond. This method is called TOLEDO (Throughput Optimization of Ligand-Protein systems Exploration through Dynamics simulation in Optimized HPC systems). First, we process the protein-ligand complex with the tools of Maestro, the next step is to run MDS in a supercomputer such as Turgalium of CIEMAT (https://www.ceta-ciemat.es/) using parallel programming and finally, show the results with graphics that are very simple to interpret. In summary, TOLEDO is a method that researches can run MDS with few parameters, very fast and with results very clear. Also, we can run multiple MDS; what is impossible to do with the graphical interface of Maestro.

References

[1] Prades, J., Imbernón, B., Reaño, C., Peña-Garcia, J., Cerón-Carrasco, JP., Illa, F., Pérez- Sanchéz, H. (2020) Maximizing resource usage in multifold molecular dynamics with rCUDA, *The International Journal of High Performance Computing Applications, 34(1), 5–19. https://doi.org/10.1177/1094342019857131*

[2] Schrödinger Release 2021-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.

[3] Schrödinger Release 2021-1: Maestro, Schrödinger, LLC, New York, NY, 2021.

Keywords: Dynamic Molecular Simulations, Parallel Computing, Maestro-Desmond, Discovery of Drugs

Genome-wide identification, characterization and functional analysis of defense-related long non-coding RNAs in non-model plant *Pinus radiata*

Zamora-Ballesteros, C.[1,2]; Martín-García, J.[1,2]; Suárez-Vega, A.[3]; Diez, J.J.[1,2].

[1]Sustainable Forest Management Research Institute, University of Valladolid—INIA, Palencia, 34004, Spain
[2]Department of Vegetal Production and Forest Resources, University of Valladolid, Palencia, 34004, Spain
[3]Department of Animal Production, Faculty of Veterinary Medicine, University of León, Campus de Vegazana s/n, León, 24071, Spain

Long non-coding RNAs (lncRNAs) are emerging as important transcriptional regulators under biotic stresses in plants. However, to date, characterization of lncRNAs in conifer trees has not been reported. The identification and the functional prediction of lncRNAs requires the use of different bioinformatics approaches that result especially challenging for non model organisms such as *Pinus radiata*. In this study, transcriptomic identification of lncRNAs was carried out using strand-specific paired-end RNA sequencing, from *Pinus radiata* samples inoculated with the pathogenic fungus *F. circinatum*. Overall, 13,312 lncRNAs were predicted after the computational analysis, including long intergenic non coding RNAs (92.3%), antisense lncRNAs (3.3%) and intronic lncRNAs (2.9%). Compared with protein-coding RNAs, pine lncRNAs are shorter, have lower expression, lower GC content and harbour fewer and shorter exons. A total of 164 differentially expressed (DE) lncRNAs were identified in response to *F. circinatum* infection in the inoculated versus mock-inoculated *P. radiata* seedlings. The predicted *cis*-regulated target genes of these pathogen-responsive lncRNAs were related to defence mechanisms such as kinase activity, phytohormone regulation, and cell wall reinforcement. Co-expression network analysis of DE lncRNAs, DE protein-coding RNAs and lncRNA target genes also indicated a potential network regulating pectinesterase activity and cell wall remodelling. This study presents the first comprehensive genome-wide analysis of *P. radiata* lncRNAs and provides the basis for future functional characterizations of lncRNAs in the *Pinus-F. circinatum* interaction.

SARS-CoV-2-specific T cell receptors after disease and vaccination

Maria S. Benitez-Cantos[1], Juana M. Vivo-Molina[2], Marta Cuadros1, Carlos Cano[3].

[1]Department of Biochemistry, Molecular Biology III and Immunology, Universidad de Granada, Granada, 18010 Spain.
[2]Department of Statistics and Operational Research, Universidad de Murcia, CEIR Campus Mare Nostrum, IMIB-Arrixaca, 30100 Murcia, Spain.
[3]Department of Computer Science and Artificial Intelligence, Universidad de Granada, Granada, 18010 Spain.

Immunization against SARS-CoV-2 following infection or vaccination has been extensively studied from the perspective of antibody responses. However, T cells also play an important role in protection against COVID-19, conferring in some cases a more durable immunization. Signals of past and present infections are encoded in the set of up to 10^{10} different T cell receptors (TCRs) that bind to antigens to trigger an immune response. A tremendous effort has been made for *in vitro* identification of SARS-CoV-2-specific TCR sequences, and in 2021 the first TCR repertoire studies in disease and vaccination have been published. Nevertheless, these analyses are often shallow and centered only in SARS-CoV-2 TCRs, disregarding the information from the rest of the TCR repertoire. In the present study, published TCR repertoires of 19 COVID-19 convalescent, 19 Ad26.COV2.S (developed by Janssen Pharmaceutica) vaccinated and 5 placebo recipient individuals have been re-analyzed with a different SARS-CoV-2-specific TCRs dataset and from a sequence similarity perspective. Convalescent and vaccinated cohorts exhibited a similar TCR response to the viral spike protein –the single antigen in the vaccine– in terms of breadth, depth and epitope location, whereas the response to non-spike antigens was significantly higher in convalescent subjects compared with vaccinated and placebo cohorts. TCRs against the N protein in convalescent subjects represent the main non-spike response, with several epitopes being targeted as in the spike protein. Furthermore, sequence similarity network analysis revealed that putatively unspecific TCRs (i.e. any SARS-CoV-2 TCRs in placebos and non-spike TCRs in vaccinated individuals) are less connected in the repertoire network, suggesting that they may be cross-reactive to other antigens and do not represent a true immune response. These findings aim to broaden the understanding of how SARS-CoV-2 exposure shapes the TCR repertoire in disease and vaccination to establish accurate and durable biomarkers of immunization.

Keywords: SARS-CoV-2, vaccines, immunogenicity, TCR repertoire, TCR biomarkers.

Bioinformatic tools for proteomic data analysis

Luisa Chocarro.

Quantitative proteomics research has demonstrated to be essential to understand how biological systems work at a molecular level. However, the bioinformatic analysis and processing of this enormous amount of data for meaningful biological and functional interpretation remains to be challenging. Here, we will discuss various current standardized methods and strategies for bioinformatic analysis, and we will highlight some available software programs, computational tools, and meta-analysis resources for proteomic analysis.

Keywords: proteomics, bioinformatics, molecular biology.

Characterization of sex differences in stroke determining a consensus transcriptional profiling through meta-analysis

Yassir El Hamdani El Amri[1*], Irene Cánovas Cervera[1*], Sergio Romera-Giner[1†], Pablo Malmierca-Merlo[1†], Susana Barceló-Cerdá[2], Francisco García-García[1].

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012 Valencia, Spain.
[2]Universitat Politècnica de València. Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, 46022 Valencia, Spain.
[3]Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012 Valencia, Spain.
[*]Both authors contributed equally to the present work
[†]Both authors contributed equally to the present work.

Corresponding author e-mail address: fgarcia@cipf.es

The cerebrovascular accident, also known as stroke, is the consequence of the interruption of cerebral blood flow, triggering cerebral hypoxia and, consequently, neuronal death. It is both one of the leading causes of reduction of years due to disability and one of the diseases with the highest mortality and morbidity rates in the world.

However, its impact on the population is disproportional. Although it is the second leading cause of death nationally, it is the first in women. This divergence in epidemiological figures is the consequence of differences in molecular mechanisms. Their analysis would allow a better diagnosis and treatment of these types of accidents. Unfortunately, in most biomedical studies on this pathology, sex is not taken into account as a key factor in the research and thus, not much information is readily available.

This systematic review and complementary meta-analysis of transcriptomic studies aim to elucidate these differences due to sex, determining new biomarkers that improve diagnosis, clinical prognosis, and most important of all, better precision and efficacy in the treatment of the disease.

We have identified 7 specific biomarkers of this pathology: 5 genes expressed differentially in women and 2 in men. A subsequent analysis of the 25 most expressed genes for each sex has revealed to have implications with anti-stroke genes. This would suggest that women are not more prone to suffer a stroke, but rather men are more protected. Additionally, specific functions were identified in women, related to the integrity of the ribosome and mitochondria, and oxidative phosphorylation. In the case of men, specific functions related to blood-brain barrier homeostasis, immune response, and axonal guidance were found.

However, a greater number of studies is required to delve into these findings, as well as promoting the implementation of sex as a key factor in the experimental design of new studies, to better understand the existing discrepancies between the sexes regarding this disease.

Keywords: Stroke; Cerebrovascular accident; Meta-analysis; Sex differences; Transcriptomics

A fine mapping association study of the HLA in sepsis susceptibility

Tamara Hernandez-Beeftink[1,2], Itahisa Marcelino-Rodriguez[1], **Eva Suárez-Pajes**[1], Megan L Paynton[3], Luis A. Rubio-Rodríguez[4], Beatriz Guillen-Guio[1], Jose M. Lorenzo-Salazar[4], Almudena Corrales[1,5], M. Isabel García-Laorden[2,5], Miryam Prieto González[6], Aurelio Rodríguez-Pérez[7,8], Demetrio Carriedo[9], Jesús Blanco[5,10], Alfonso Ambrós[11], Elena González Higueras[12], Elena Espinosa[13], Arturo Muriel[10], David Domínguez[13], Louise V Wain[14], Abelardo García de Lorenzo[15], José M. Añón[15], Javier Belda[16], Jesús Villar[2,5], Carlos Flores[1,4,5] for the Genetics of Sepsis (GEN-SEP) Network.

[1]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; [2]Research Unit, Hospital Universitario Dr. Negrin, Las Palmas de Gran Canaria, Spain; [3]Department of Health Sciences, University of Leicester, Leicester, United Kingdom; [4]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain; [5]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; [6]Servicio de Medicina Intensiva, Complejo Asistencial Universitario de Palencia, Palencia, Spain; [7]Department of Anesthesiology, Hospital Universitario de Gran Canaria Doctor Negrín, Las Palmas de Gran Canaria, Spain; [8]Department of Medical and Surgical Sciences, University of Las Palmas de Gran Canaria, Gran Canaria, Spain; [9]Intensive Care Unit, Complejo Hospitalario Universitario de León, León, Spain; [10]Intensive Care Unit, Hospital Universitario Rio Hortega, Valladolid, Spain; [11]Intensive Care Unit, Hospital General de Ciudad Real, Ciudad Real, Spain; [12]Intensive Care Unit, Hospital Virgen de la Luz, Cuenca, Spain; [13]Department of Anesthesiology, Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain; [14]National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom; [15]Intensive Care Unit, Hospital Universitario La Paz, Madrid, Spain; [16]Anesthesiology and Critical Care Department, Hospital Clinico Universitario of Valencia, Valencia, Spain.

Background: Sepsis is a severe inflammatory response to infections with a high death rate. We previously conducted the first GWAS of copy number variations in 839 sepsis cases from the Gen-Sep Network and 1,453 controls, highlighting 6p22.1 as one of the significant loci linked to sepsis susceptibility. Due its importance in inflammatory and immunological diseases, here we performed a fine mapping of the Human Leukocyte Antigen (HLA) region contained in that locus.

Methods: We used SHAPEIT v2.837 for phasing the haplotypes and Impute2 to impute the classic HLA alleles, amino acids, and single nucleotide polymorphisms (SNPs). Association analyses were performed by logistic regressions using EPACTs v3.2.6. A Bonferroni correction was applied to identify significant classic HLA alleles ($p<2.58\text{-}4$) and amino acids ($p<4.91E\text{-}5$). For SNPs, a significance threshold for significance was established at $p<1.50E\text{-}5$ based on the number of independent variants.

Results and conclusions: We analyzed a total of 194 classic HLA alleles, 1,019 amino acids and 10,919 SNPs. None of the classic HLA alleles (p lowest =0.01), amino acids (p lowest =0.01), or SNPs (p lowest =9.84E-4) were significantly associated with sepsis. Given the complexity of this phenotype, these results suggest that the HLA genetic variation is not a major driver of sepsis susceptibility or has a modest effect size.

Funding: Instituto de Salud Carlos III (CD19/00231, FI17/00177, PI17/00610, PI20/00876), Ministerio de Ciencia e Innovación (RTC-2017- 6471-1; AEI/FEDER, UE), and agreement OA17/008 with ITER; ECIT CGIEU0000219140. The genotyping

service was performed at CEGEN-PRB3- ISCIII, supported by grant PT17/0019, of the PE I+D+i 2013-2016, funded by ISCIII and FEDER.

NanoRTax, a real-time pipeline for taxonomic and diversity analysis of nanopore 16S rRNA amplicon sequencing data

Héctor Rodríguez-Pérez[1], Laura Ciuffreda[1], Carlos Flores[1,2,3].

[1]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, 38010, Santa Cruz de Tenerife, Spain; [2]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, 28029, Madrid, Spain; [3]Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), 38600 Granadilla, Santa Cruz de Tenerife, Spain

Since the adoption of NGS technologies, the continuous development sequencing techniques and cost reduction have revolutionized the study of microbial communities. Nanopore long read sequencing platform allows for going beyond genus-level when performing pathogen identification or diversity analysis due to the increase in read length and features the unique ability to access and therefore analyze sequences of an ongoing experiment at the time they are generated.

We present NanoRTax *(https://github.com/genomicsiter/nanortax)*, a nextflow-based long-read metagenomics pipeline for taxonomic and sample diversity analysis. The pipeline features the integration of state-of-art read classification methods, downstream analysis steps and real-time capability to enable early analysis and benchmarking of full-length 16S rRNA classification workflows. Additionally, it can be paired with an independent Dash web application which provides immediate access to an interactive dashboard with taxonomic information, diversity statistics and visualizations from in-progress or finished pipeline runs.

We applied the NanoRTax workflow to the evaluation of early data analysis from full-length 16S rRNA sequencing in the context of predicting mortality in sepsis patients admitted in ICU. We previously observed that a reduction in bacterial lung diversity within 8 h of sepsis diagnosis is associated with ICU mortality, providing a potentially novel and early prognostic biomarker of non-pulmonary sepsis. Simulating a real-time environment, the previously observed dysbiosis in deceased patients is accountable as early as at 0-2h from experiment start to 24-48h. Our results derived from the analysis of early sequencing data support the development and implementation of real-time sequencing approaches in clinical context.

Deciphering the cellular, molecular, clinical, and therapeutic implications of lung cancers lacking targeted therapies: from gene expression to tailored therapies

Ane Martínez Larrinaga.

UNIVERSITY OF NAVARRA

Background: Lung cancer remains the most common cancer worldwide, both in terms of incidence and mortality. In 2040 > 3.1 million new cases are expected. The survival rate of patients with lung cancer is around 5-year and it strongly depends on the stage of the tumor. The main cause factor of lung cancer is smoking and the use of tobacco products. Lung cancer is classified in two broad histologic classes, which grow and spread differently: small-cell lung carcinomas (SCLC) and non-small cell lung carcinomas (NSCLC). In this project we focused our analysis in the two predominant subtypes of NSCLC which are: lung adenocarcinoma (LUAD) and lung squamous cancer (LUSC) because the account to the 85% of total lung cancers. LUAD and LUSC have a high mutational rate, however, around one third of LUAD patients lack druggable genomic mutations and can't benefit from target therapies. Regarding LUSC, no targeted therapies have been approved so far. Thus, there is a need to identify subgroups of patients amenable to new pharmacological treatments

Hypothesis: The main hypothesis of this project was that from gene expression data we could identify different subgroups in LUAD and LUSC. Furthermore, we hypothesized that if we were able to understand the molecular and cellular mechanism that differentiated these groups, we would predict specific drugs that may induce an antitumor effect on each group genotype.

Methods: RNA-Seq data of LUAD and LUSC was download from The Cancer Genome Atlas (TCGA). Frist a clustering analysis, using non negative matrix factorization (NMF) and K-means as clustering analysis, was performed to identify the possible subgroups of each pathology. Second, different analysis were made to obtain the biological description of the identify groups: **1)** the differentially expressed genes (DEGs) were identified, **2)** a functional enrichment analysis was done to identify the enriched GO Terms in each groups signature, **3)** the tumor microenvironment (TM) and the tumor infiltrating immune cells (TIICs) were analyzed, **4)** the differentially expressed (DE) transcription factors (TFs) were identify and **5)** with the clinical data from the TCGA a survival analysis was performed to obtain the survival rate of each group. Third, a more extense study of the TFs was made, with a gene regulatory network (GRN) to find the most biologically relevant TF, which were rewired. Lastly, a drug repositioning analysis was performed to predict the drugs that could reverse the signature of each group.

Results: The clustering analysis identified 3 groups which had relevant difference at a biological and clinical level. In both pathologies the groups presented

differences enriched terms and different DE TFs and rewired TFs. Furthermore, the proposed drugs were specific for each group signature which evidence the idea that we had 3 groups in both LUAD and LUSC.

Conclusions: In both LUAD and LUSC based on their transcriptome different groups could be identify. This groups could be characterized at a molecular and cell level, also, specific drugs based on each groups signature were proposed as possible treatment option.