IV Jornada de Bioinformàtica i Genòmica

Organitzada per:
Secció de Bioinformàtica i Biologia Computacional de la SCB
Secció de Genòmica i Proteòmica de la SCB
Associació Bioinformatics Barcelona – BIB

Patrocinada per:

Atos Damm

PROGRAMA I RESUMS DE LES COMUNICACIONS

PARC DE RECERCA BIOMÈDICA DE BARCELONA (PRBB)

Auditori PRBB

Carrer Aiguader 88

Barcelona

20 de desembre de 2016

COMITÈ ORGANITZADOR:

Núria López-Bigas (ICREA, IRB)
Mario Cáceres (ICREA, UAB)
Roderic Guigó (CRG-UPF)
Ana Ripoll (UAB, BIB)

SUPORT:

Mariàngels Gallego (SCB)
Maite Sánchez (SCB)
Eva Alloza (BIB)

8:30 – 9:15 Registration

9:15 – 9:30 Wellcome and opening of the symposium
Dr. Arcadi Navarro (Secretari d'Universitats i Recerca, Generalitat de Catalunya
Dr. Marc Martí-Renom (Board of Directors of the Societat Catalana de Biologia)
Dra. Ana Ripoll (President of the Bioinformatics Barcelona Association - BIB)

SESSION I. Chair: Núria López-Bigas (ICREA, IRB)

9:30 - 10:15 **Morning's Keynote Lecture: Peter Campbell** (Wellcome Trust Sanger Institute, UK). Interrogating the architecture of cancer genomes.
10:15 - 10:30 **Ferran Nadeu** (IDIBAPS). Clinical impact of the quantitative subclonal architecture in chronic lymphocytic leukemia.
10:30 - 10:45 **Marcos Díaz-Gay** (IDIBAPS). Integrated analysis of germline and tumor DNA for the identification of new genes involved in familial colorectal cancer.
10:45 - 11:00 **Natàlia Padilla** (VHIR). Identification of causative mutations in breast and ovarian inherited cancers.

11:00 - 11:30 Coffee Break (PRBB terrace)

SESSION II. Chair: Cedric Notredame (CRG)

11:30 - 11:45 **Lidia Mateo** (IRB). A PanorOmic view of personal cancer genomes.
11:45 - 12:00 **Eduardo Eyras** (ICREA, UPF). Alternative splicing remodels the protein interaction network of cancer gene drivers.
12:00 - 12:15 **Fran Supek** (CRG). The rules and impact of nonsense-mediated mRNA decay in human cancers.
12:15 - 12:30 **Marta Guindo** (BSC-CNS). GUIDANCE: An Integrated framework for large scale genome and phenome-wide association studies on parallel computing platforms.
12:30 - 12:45 **Jordi Pujols** (UAB). AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures.
12:45 - 13:00 **Oriol Senan** (URV). AddClique: A network based model for adduct identification in LC/MS metabolomics.
13:00 - 13:15 **Albert Trill** (AtoS). Adapting HPC architectures to the bioinformatics specific needs.

13:15 - 14:30 Lunch (PRBB terrace) and free poster viewing

SESSION III. Chair: Jaume Bertranpetit (IBE-UPF)

14:30 - 14:45 **Sanja Zivanovic** (IRB). Multi-level strategy for analysis of bioactive drug conformations.

14:45 - 15:00 **Julien Lagarde** (CRG). High-throughput manual-quality annotation of full-length long noncoding RNAs with capture long-read sequencing (CLS).

15:00 - 15:15 **Juan M. Vaquerizas** (Max Planck Institute for Molecular Biomedicine, Germany). Transcriptional nucleation of topological domains during early embryogenesis.

15:15 - 15:30 **Jordi Moreno-Romero** (Swedish University of Agricultural Sciences, Sweden). Parental-specific epigenomics in Arabidopsis.

15:30 - 15:45 **Cristina Frías-lópez** (UB). Evolutionary genomics and transcriptomics of arthropod chemosensory systems: software development and experimental approaches.

15:45 - 16:00 **Mayukh Mondal** (IBE-UPF). Unraveling extinct genomes in present day human genomes.

16:00 - 16:30 Coffee Break (PRBB terrace)

SESSION IV. Chair: Mario Cáceres (ICREA, UAB)

16:30 - 16:45 **Roger Pique-Regi** (Wayne State University, USA). Functional Genomic analysis of Gene-by-environment interactions across 250 environments.

16:45 - 17:00 **Maria Maqueda** (UPC). Independent components of gene expression in endurance runners.

17:00 - 17:15 **Ricard Argelaguet** (European Bioinformatics Institute (EMBL-EBI), UK). Disentangling the common and specific sources of variation between different biological layers in multi-view single-cell sequencing.

17:15 - 18:00 **Afternoon's Keynote Lecture: John Marioni** (European Bioinformatics Institute (EMBL-EBI), UK). Using single-cell approaches to understand cell fate decisions in early mammalian development.

18:00 - 19:00 Poster viewing with authors and beer session organized jointly with the bioinformatics student group RSG-Spain

19:00 BIB award to the best oral communication and poster and end of the symposium.

Oral presentations

CLINICAL IMPACT OF THE QUANTITATIVE SUBCLONAL ARCHITECTURE IN CHRONIC LYMPHOCYTIC LEUKEMIA

Ferran Nadeu[1], Julio Delgado[1,2], Guillem Clot[1], David Martín-García[1], Tycho Baumann[2], Magda Pinyol[3], Pedro Jares[1,2], Sílvia Beà[1], Itziar Salaverria[1], Alba Navarro[1], Helena Suárez-Cisneros[3], Marta Aymerich[1,2], Maria Rozman[1,2], Neus Villamor [1,2], Dolors Colomer[1,2], Marcos González[4], Miguel Alcoceba[4], Maria José Terol[5], Enrique Colado[6], Armando López-Guillermo[1,2], Xose S Puente[7], Carlos López-Otín[7], Anna Enjuanes[3], Elías Campo[1,2,8]

1. Lymphoid Neoplasm Program, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Rosselló, 149, 08036 Barcelona, Spain
2. Departament d'Hematologia, Hospital Clínic, Villarroel, 170, 08036 Barcelona, Spain
3. Unitat de Genòmica, IDIBAPS, Rosselló, 149, 08036 Barcelona, Spain
4. Biología Molecular e Histocompatibilidad, Hospital Universitario, Paseo de San Vicente, 88-182, 37007 Salamanca, Spain
5. Unidad de Hematología, Hospital Clínico Universitario, Av. de Blasco Ibáñez, 17, 46010 Valencia, Spain
6. Servicio de Hematología y Hemoterapia, Hospital Universitario Central de Asturias, Av. Roma, s/n, 33011 Oviedo, Spain
7. Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología, Universidad de Oviedo, Fernando Bongera s/n, Edificio Santiago Gascón, 33006 Oviedo, Spain
8. Departament d'Anatomia Patològica, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain
Presenter e-mail: nadeu@clinic.ub.es

Purpose: Recent studies have revealed the presence and prognostic impact of small mutated subclones in chronic lymphocytic leukemia (CLL). However, these studies focused only on 5 genes. We set out to define the deep mutational architecture of 28 CLL driver genes, and determine the relevance of the quantitative subclonal architecture of the tumors in the progression of the disease.
Patients and methods: We assessed the mutational status of 406 untreated CLL patients by deep next-generation sequencing. Mutations present at less than 1% of tumor cells were identified and verified. Copy number alterations (CNA) were analyzed by SNP-arrays.
Results: Small subclonal mutations were the sole alteration in 22% of the mutated cases, and were frequently detected in nearly all genes analyzed. We identified three gene-specific patterns that linked the magnitude of the mutated clones (or mutated cancer cell fraction, CCF) with the prognosis of the patients: *CCF-independent pattern*: mutations at any CCF conferred an adverse prognosis; *CCF-clonal pattern*: only mutations with a CCF above a certain threshold impacted the outcome; and *CCF-gradual pattern*: the poor prognostic impact was a continuous variable directly related to the size of the clone. Finally, by combining mutations and CNA we analyzed how the global tumor architecture influences the clinical outcome. In this regard, the accumulation of driver alterations shortened the overall survival of the patients only when at least one alteration was present in a subclone. Conversely, tumors with only clonal alterations (non-heterogeneous) were slightly more aggressive than tumors carrying subclonal alterations as revealed in time to first treatment analyses.

Conclusions: The emergence of subclonal mutations is a dynamic phenomenon in CLL involving most driver genes. The clinical impact of the mutations is related to their CCFs, and the integration of the subclonal architecture in clinical analyses is desirable for a correct assessment of CLL outcome.

INTEGRATED ANALYSIS OF GERMLINE AND TUMOR DNA FOR THE IDENTIFICATION OF NEW GENES INVOLVED IN FAMILIAL COLORECTAL CANCER

Marcos Díaz-Gay[1], Sebastià Franch-Expósito[1], Clara Esteban-Jurado[1], Jenifer Muñoz[1], Anna Gratacós-Mulleras[1], Teresa Ocaña[1], Sabela Carballal[1], Miriam Cuatrecasas[2], Maria Vila-Casadesús[3], Juan José Lozano[3], Sophia Derdak[4], Steven Laurie[4], Sergi Beltran[4], Antoni Castells[1], Luis Bujanda[5], Joaquín Cubiella[6], Francesc Balaguer[1], Sergi Castellví-Bel[1]

1. Gastroenterology Department, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Hospital Clínic, Rosselló 149-153, 08036 Barcelona, Spain
2. Pathology Department, Hospital Clínic, Villarroel 170, 08036 Barcelona, Spain
3. Bioinformatics Platform, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Rosselló 149-153, 08036 Barcelona, Spain
4. Centre Nacional d'Anàlisi Genòmica-Centre de Regulació Genòmica (CNAG-CRG), Parc Científic de Barcelona, Baldiri Reixac 4, 08028 Barcelona, Spain
5. Gastroenterology Department, Hospital Donostia-Instituto Biodonostia, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Basque Country University, Begiristain Doktorea Pasealekua, s/n, 20014 San Sebastián, Spain
6. Gastroenterology Department, Complexo Hospitalario Universitario de Ourense, Instituto de Investigación Sanitaria Galicia Sur, Ramon Puga Noguerol 54, 32005 Ourense, Spain.
Presenter e-mail: diaz2@clinic.cat

Colorectal cancer (CRC) is one of the most common and lethal neoplasms worldwide. Genetic factors account for 35% of its susceptibility. Hereditary forms are mainly due to highly penetrant variants in genes such as APC, MUTYH and the DNA mismatch repair family. Familial CRC shows also familial aggregation, it does not present alterations in the hereditary CRC genes and part of its heritability remains unexplained. Our study presents a new methodology to identify novel variants in genes linked to familial CRC germline predisposition. An integrated germline-tumor analysis based on Knudson's two-hit hypothesis, which allows assessing the potential role as tumor suppressor genes of the selected candidates. Whole exome sequencing was performed both in germline and tumor DNA of five unrelated patients with strong CRC aggregation, compatible with an autosomal dominant pattern of inheritance and without alterations in known hereditary CRC genes. Deleterious single nucleotide (SNVs) and copy number variants (CNVs) were evaluated as candidates for first germline or second somatic hits by a multi-step filtration process. An analysis pipeline programmed in R language was used in the case of SNVs and four different algorithms were implemented to CNV inference: CoNIFER, ExomeDepth, VarScan2 and CNVkit. Sequencing quality, population frequency, family segregation, affected protein function and expression, pathogenicity and copy number status were taken into account as main variant features. Somatic mutational signatures characterization was also performed. Thirteen candidate germline-somatic variant pairs were prioritized. Three corresponded to pairs of SNVs and ten were composed by a germline SNV and a somatic CNV. SNV validation was performed with Integrative Genomics Viewer and Sanger sequencing, whereas CNV

verification and further functional studies will be required. A hypermutator phenotype was suggested in one of the analyzed tumors, according to the large number of variants detected in the somatic mutational profile analysis.

IDENTIFICATION OF CAUSATIVE MUTATIONS IN BREAST AND OVARIAN INHERITED CANCERS

Natàlia Padilla[1], Casandra Riera[1], Gemma Montalbán[2], Sandra Bonache[2], Orland Díez[2,3], Sara Gutiérrez-Enríquez[2], Xavier de la Cruz[1,4]
1. Translational Bioinformatics, Vall d'Hebron Institute of Research (VHIR)
2. Oncogenetics Group, Vall d'Hebron Institute of Oncology (VHIO)
3. Àrea de Genètica Clínica i Molecular, Hospital Universitari Vall d'Hebron (HUVH)
4. ICREA, Barcelona, Spain
Presenter e-mail: natalia.padilla@vhir.org

Diagnosis of inherited breast and ovarian cancers is a challenging problem, where sequence information plays an important role. Indeed, identification of potentially causative mutations in *BRCA1* and *BRCA2* genes can help physicians to choose the best treatment for their patients. However, interpretation of sequencing evidence is far from trivial. For several years, the program AlignGVGD (Tavtigian et al., Human Mutation 2008), a bioinformatics tool based on the use of evolutionary information, has been a reference in the field. But recent experimental studies (Starita et al., Genetics 2015) suggest that for *BRCA1* it may be preferable to use functional assays rather than evolutionary information. In this work, we revisit the value of this information using a set of 227 (77 pathogenic and 150 neutral) and 141 (36 pathogenic and 105 neutral) manually curated variants for *BRCA1* and *BRCA2*, respectively. We first analyze their distribution relative to a set of molecular properties, such as protein domain structure, conserved regions and structure location. Our results indicate that pathogenic and neutral mutations are differentially distributed relative to some of these properties, to a degree supporting their predictive value. We have confirmed that this is indeed the case, building a simple predictor for each gene with a success rate clearly above that of a random predictor. This confirms that indeed, in the case of inherited breast and ovarian cancers, evolutionary information combined other molecular-level features have the ability to identify causative mutations. We have confirmed this result with other predictors, including AlignGVGD, PON-P2, SIFT and PolyPhen-2. Interestingly, while AlignGVGD surpasses the other predictors, its performance drops to random when using another alignment as main input. This phenomenon, absent in other tools, indicates that it is preferable to use more than one program simultaneously for the scoring of variants in *BRCA1* and *BRCA2* genes.

A PANOROMIC VIEW OF PERSONAL CANCER GENOMES

Lídia Mateo[1], Oriol Guitart[1], Carles Pons[1], Patrick Aloy[1,2]

1. Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology. c/ Baldiri i Reixac 10-12. 08028 Barcelona, Catalonia, Spain
2. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain
Presenter e-mail: lidia.mateo@irbbarcelona.org

In order to turn precision oncology promise into reality we need to complement histopathologic analyses with molecular diagnostics. Currently, we can generate a comprehensive and accurate molecular portrait of an individual's tumor by putting together multiple layers of molecular profiles, such as somatic mutations, DNA copy-number alterations or mRNA expression. Computational solutions to large-scale data integration and visualization are needed to facilitate the interpretation of such complex molecular profiles. Cancer PanorOmics is a web-based service that looks at the genomic alterations found in an individual tumor in the light of the body of scientific evidence already available for 28 tumor type cohorts. Additionally, it also provides information about the molecular context of the mutations, by mapping mutations on the human structural interactome. This “PanorOmic” view of personal cancer genomes should enable a better understanding and interpretation of the results and increase the diagnostic utility of high throughput methods. Together with an appropriate genetic counseling and/or the medical advice from molecular tumor boards at the clinical onset, the visualization of cancer alterations in Cancer PanorOmics could contribute to the identification of actionable alterations guiding the clinical decision-making process. The results are shown in an interactive graphical interface that can be used to give back the genomic test results to the patient, whose information and empowerment should play a central role in personalized cancer management.

ALTERNATIVE SPLICING REMODELS THE PROTEIN INTERACTION NETWORK OF CANCER GENE DRIVERS

H. Climente-González[1], M. Reixachs[1], E Porta-Pardo[2], A. Godzik[2], E. Eyras[1,3]

1. Pompeu Fabra University (UPF). PRBB, E08003, Barcelona, Spain
2. Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, 92037, USA
3. Catalan Institution of Research and Advanced Studies (ICREA), E08010, Barcelona, Spain
Presenter e-mail: eduardo.eyras@upf.edu

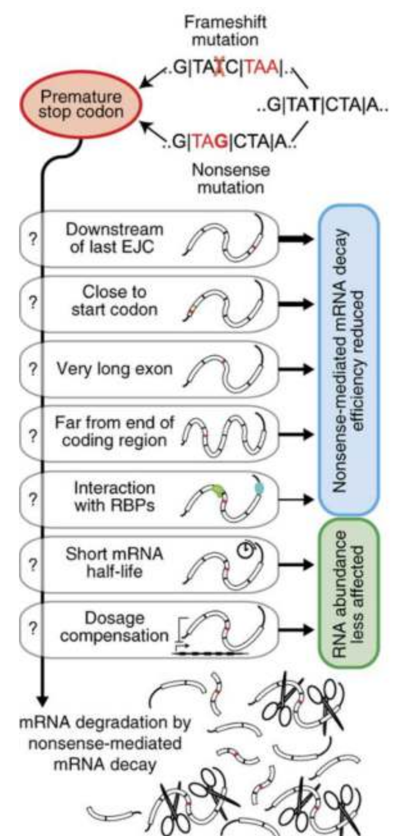
Alternative splicing changes are frequently observed during physiological processes like cell differentiation, as well as during disease states, including cancer. However, the functional relevance of these splicing changes remains mostly unknown. We have carried out a systematic analysis to characterize the potential functional consequences of alternative splicing changes in thousands of tumor samples. This analysis reveals that a subset of alternative splicing changes affect protein domain families that are frequently mutated in tumors, disrupt the protein-protein interaction network of cancer gene drivers, and show mutual exclusion with mutations in cancer drivers. We further show that tumor samples present a negative correlation between the number of these splicing changes and the number mutated cancer drivers. We extend this analysis to show that splicing changes affect protein production using sequencing of RNAs occupied by Ribosomes (Ribo-seq) and that the remodeling of the protein interaction network of cancer drivers is conserved between human and mouse. We propose that a subset of the alternative splicing changes observed in tumors represents independent oncogenic processes and could potentially be considered alternative splicing drivers (AS-drivers).

THE RULES AND IMPACT OF NONSENSE-MEDIATED mRNA DECAY IN HUMAN CANCERS

Rik G.H. Lindeboom[1,2], Fran Supek[1,2,3] and Ben Lehner[1,2,4]

1. EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Catalonia, Spain.
2. Universitat Pompeu Fabra (UPF), 08003 Barcelona, Catalonia, Spain.
3. Division of Electronics, Rudjer Boskovic Institute, 10000 Zagreb, Croatia.
4. Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain.
Presenter e-mail: fran.supek@crg.eu

Premature termination codons (PTCs) cause cancer as well as a large proportion of inherited human genetic diseases. PTC-containing transcripts can be degraded by an mRNA surveillance pathway termed nonsense-mediated decay (NMD). However, the efficiency of NMD varies, for example being inefficient when a PTC is located downstream of the last splice site in the mRNA (known as the exon junction complex /EJC/ model). We used matched exome and transcriptome data from 9,769 human tumors to systematically elucidate the rules of NMD targeting in human cells.



An integrated model incorporating multiple rules beyond the canonical EJC model explains 74% of the non-random variance in NMD efficiency across thousands of PTCs. For instance, we find that start codon-proximal PTCs commonly evade NMD *via* downstream re-initiation of translation. Moreover, NMD is less efficiently triggered by PTCs in very long exons and by PTCs that are far upstream of the wild-type stop codon. Sequence motifs corresponding to known RNA-binding proteins may modulate NMD activity in particular instances. We also show that rapid mRNA turnover and gene dosage compensation mask the effects of NMD in many genes.

Applying the NMD model reveals signatures of both positive and negative selection on NMD-triggering somatic mutations in human tumors and provides a novel classification of tumor suppressor genes. Taken together therefore, this study [1] provides important mechanistic insight into NMD and into tumor evolution, as well as a broader framework for predicting the effects of nonsense variants in human disease.

References:

1. RGH Lindeboom, F Supek and B Lehner. ***Nature Genetics*** **48**, 1112–1118 (2016) doi:10.1038/ng.3664

GUIDANCE: AN INTEGRATED FRAMEWORK FOR LARGE-SCALE GENOME AND PHENOME-WIDE ASSOCIATION STUDIES ON PARALLEL COMPUTING PLATFORMS

Marta Guindo-Martínez[1,*], Friman Sánchez[1,2,*], Silvia Bonàs-Guarch[1], Montserrat Puiggròs[1], Jorge Ejarque[2], Carlos Díaz[2], Enric Tejedor[2], Rosa M. Badia[2,3], Josep M. Mercader[1,#], David Torrents[1,4,#]

1. Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain
2. Computer Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
3. Artificial Intelligence Research Institute (IIIA), Spanish Council for Scientific Research (CSIC), Barcelona, Spain
4. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
* Both authors contributed equally to this work.
# These authors jointly directed this work.
Presenter e-mail: marta.guindomartinez@bsc.es

Current Genome-Wide Association Studies (GWAS) are dealing with continuously increasing large sample sizes available from large-scale biobanks and meta-analysis consortia. To allow an efficient analysis of the current and upcoming GWAS datasets, we developed GUIDANCE, an integrated modular framework that performs genotype imputation using multiple reference panels and allows both single trait and phenome-wide association testing with optional user intervention. We applied GUIDANCE to The Resource for Genetic Epidemiology Research on Aging (GERA) Cohort, comprising 56,637 subjects and 22 phenotypes, using 1000G phase 3, UK10K and GoNL for genotype imputation. As a result, we were able to replicate 54 *loci* at the genome-wide significance level (p<5x10-8) that have been previously described for 11 phenotypes. Moreover, for 14 phenotypes, 19 not previously reported *loci* were found. Interestingly, we identified a new variant in the X chromosome for dyslipidemia (OR males=0.83(0.78-0.87), p=5.19x10-11; OR females=0.91(0.86-0.97), p=0.006). Furthermore, through cross-phenotype analysis of GERA results, 18 genome-wide associated *loci* were found significant (p<3.39x10-5) for at least one additional disease. Some of the cross-phenotype associations that we identified had been previously reported, for instance, the 5q22 region associated with both asthma and allergic rhinitis. Additionally, novel cross-phenotype associations were also found, but further analysis should be performed in order to know whether these cross-phenotype associations are driven by comorbidity among diseases. All these novel findings are of interest for future validations and follow-up analysis and highlight the value of GUIDANCE for easy, efficient, standardized and complete genome and phenome-wide association studies.

AGGRESCAN3D (A3D): SERVER FOR PREDICTION OF AGGREGATION PROPERTIES OF PROTEIN STRUCTURES

Jordi Pujols[1], Rafael Zambrano[1], Michal Jamroz[2], Agata Szczasiuk[2], Sebastian Kmiecik[2] and Salvador Ventura[2]

1. Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain
2. Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw, Pasteura 1, Warsaw
Presenter e-mail: jordi.pujols@uab.cat

Protein aggregation has moved beyond being a mostly ignored area of protein chemistry to become a key topic in biomedicine and biotechnology. It underlies more than 40 human disorders, including neurodegenerative diseases -such Alzheimer's and Parkinson's- and non-neuronal related diseases such diabetes type II or some type of cancers. In addition, it is a pivotal factor to take into account while manufacturing protein-based therapeutics like monoclonal antibodies, growth factors or replacement enzymes. With the aim of anticipate this phenomenon, the present understanding on the molecular determinants of protein aggregation has crystalized in a series of predictive algorithms to identify the aggregation-prone sites of proteins. The vast majority rely on the aminoacidic sequence. Therefore they find difficulties to predict the aggregation properties of folded globular proteins, where aggregation-prone sites are often not contiguous in sequence or buried inside the native structure. The Aggrescan 3D server overcomes these limitations by projecting onto the protein structure the experimental aggregation propensity scale from the well-established AGGRESCAN method. In this way, the native neighbouring tendencies modulate the aggregation propensity score for each amino acid to obtain high confident predictions. Using the A3D server, the identified aggregation-prone residues can be virtually mutated to design variants with increased solubility, or to test the impact of pathogenic mutations. Additionally, A3D server enables to take into account the dynamic fluctuations of protein structure in solution, which may influence aggregation propensity, by using the fast simulations of CABS-flex approach. The A3D server can be accessed at http://biocomp.chem.uw.edu.pl/A3D/.

ADDCLIQUE: A NETWORK BASED MODEL FOR ADDUCT IDENTIFICATION IN LC/MS METABOLOMICS

Oriol Senan[1], Antoni Aguilar-Mogas[1], Miriam Navarro[2], Óscar Yanes[2], Marta Sales-Pardo[1] i Roger Guimerà[1,3].

1. Universitat Rovira i Virgili, departament d'Enginyeria Química, Tarragona
2. Center for Omics Sciences, Universitat Rovira i Virgili, Reus
3. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona

Presenter e-mail: oriol.senan@urv.cat

Liquid chromatography coupled to mass spectrometry (LC/MS) is the most common experimental setup for untargeted metabolomics. In a typical LC/MS experiment, compounds are first separated in the chromatography and then ionized at the spectrometer, where they are analyzed. In this process, molecules from the same metabolite can undergo different transformations, usually incorporating or losing precise molecular moieties. Each of these newly formed molecules or adducts of the same metabolite produces a different signal in the spectrometer, increasing the complexity of the analysis. Correct interpretation of these signals is crucial for a rigorous and accurate metabolomics experiment. To aid in this interpretation we have developed AddClique, a network based algorithm that is able to systematically discriminate adducts from the same metabolite from those from another metabolite, and then identify them. AddClique first computes correlations between peaks of the chromatogram to obtain the probability that those peaks correspond to adducts of the same molecule. With these probabilities AddClique builds a network, with adducts as nodes and probabilities as links. Then, AddClique identifies groups of fully connected components (cliques), as these are the most probable adducts of the same metabolite. Finally AddClique identifies the adducts with their masses as well as the metabolite. We tested our algorithm with datasets of increasing complexity: from single molecule experiments to a complex biological sample. We show that AddClique is a valid tool for the identification of adducts and can be incorporated to the regular workflow in the analysis of metabolomics data.

MULTI-LEVEL STRATEGY FOR ANALYSIS OF BIOACTIVE DRUG CONFORMATIONS

Sanja Zivanovic[1,2], Adam Hospital[1,2], Modesto Orozco[1,2,3]

1. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain
2. Joint BSC-IRB Research, Barcelona, Spain
3. Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Spain
Presenter e-mail: sanja.zivanovic@irbbarcelona.org

The procedure of drug discovery is very time and resources consuming process. Computer-aided drug design (CADD) is one of the powerful tools which can be used to increase the efficiency of the drug discovery. Estimating the relative free energy of a ligand in its target-bound state (i.e. the bioactive conformation) is necessary to optimize the potency of bioactive molecules and to improve the accuracy of SDBB methods. Our aim is to develop an efficient framework for finding the bioactive conformation of the flexible ligands. Since the bioactive conformation of the ligand may differ from the global minimum of the free ligand in the physiological environment, one has to evaluate the energetic cost required for adopting the bioactive conformation. A set of 100 crystal structures of pharmaceutically relevant drug-like molecules was tested using multi-level approach. We combined low-level method (LL) for sampling the conformational minima and high-level (HL) ab-initio calculations for estimating their relative stability.The method was automated and tested on various ligands with different numbers of atoms, charge and rotatable bonds. The analysis show that is necessary to perform Hamiltonian Replica Exchange simulations in order to explore all possible states of energy landscape of given dihedrals. Our findings suggest that the method is an effective way to improve analysis of the bioactive conformations of drug-like molecules. It is worth noting that present framework for multilevel strategy is a complex and long-term task, which requires a lot of rehearsals and implementations. Taking into account the flexible nature of molecules, protonation state and tautomeric forms, make our task even more challenging. The proposed strategy may represent an efficient tool for predicting the conformational landscape of drugs while keeping a reasonable balance between chemical accuracy and computational cost.

Keywords: drug design; computational chemistry; pharmacology; development;

HIGH-THROUGHPUT MANUAL-QUALITY ANNOTATION OF FULL-LENGTH LONG NONCODING RNAS WITH CAPTURE LONG-READ SEQUENCING (CLS)

Julien Lagarde[1], Barbara Uszczynska[1], Silvia Carbonell[2], Carrie Davis[3], Thomas R. Gingeras[3], Adam Frankish[4], Jennifer Harrow[4], Roderic Guigó[1], Rory Johnson[1] and GENCODE

1. CRG,
2. QGenomics, Barcelona, Spain
3. Cold Spring Harbor Laboratory, New York, United States
4. Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

Presenter e-mail: julien.lagarde@crg.eu

Efforts to decipher the functions of long noncoding RNAs depend on availability of full-length and comprehensive transcript annotations. The gold-standard manually-curated GENCODE collection remains far from complete: many genes models are fragmentary, while thousands more need to be catalogued. To accelerate lncRNA annotation, we have developed RNA Capture Long Seq (CLS), combining targeted RNA capture with 3rd generation long-read sequencing (PacBio). We present a survey of the entire GENCODE intergenic lncRNA populations in matched human and mouse tissues at a depth of 2 million reads each. Mapping and merging these data more than doubles the annotation complexity of these loci. Novel splice junctions are supported by short-read sequencing and are enriched in known splicing motifs. Combining transcription initiation and termination data, we catalogue a non-redundant set of 8,500/3,200 full-length novel lncRNA transcript structures, respectively. This unique full-length dataset reveals fundamental similarities and differences of lncRNA and protein-coding genes for the first time. CLS removes a longstanding bottleneck of transcriptome annotation, generating manual-quality full-length transcript models at high-throughput scales.

TRANSCRIPTIONAL NUCLEATION OF TOPOLOGICAL DOMAINS DURING EARLY EMBRYOGENESIS

Clemens B. Hug[1], Alexis Grimaldi[1], Kai Kruse[1] and Juan M. Vaquerizas[1]

1. Max Planck Institute for Molecular Biomedicine, Roentgenstrasse 20, 48149 Muenster, Germany
Presenter e-mail: jmv@mpi-muenster.mpg.de

The three dimensional organisation of the genome plays a fundamental role in the regulation of gene expression. Recent examinations of chromatin conformation have revealed the presence of hundreds of self-associating domains and thousands of regulatory loops between enhancers and target genes that ensure a correct deployment of developmental programmes. Mutations affecting these highly conserved and relatively tissue-unspecific regulatory features often result in striking developmental defects and disease. Despite the characterisation of such regulatory features, it is currently unknown when chromatin conformation is established and the cellular mechanisms that drive this process. Here, using *Drosophila* as a model system, we investigated chromatin conformation in tightly staged, hand-sorted embryos at five different developmental time points. *In situ* Hi-C maps for these embryos reveal that chromatin structure is significantly remodelled at hundreds of loci during embryo development. We further demonstrate that specific expressed loci serve as nucleation sites for early topologically associating domain (TAD) boundaries. Strikingly, inhibition of RNA polymerase II processivity does not preclude the formation of TADs, suggesting that transcriptional elongation is not necessary for the establishment of such domains. However, reduction of RNA polymerase II activity results in a marked decrease of insulation between topological domains. Our results have important implications for our understanding of how the three-dimensional structure of the genome is established and the mechanisms that trigger this organisation.

PARENTAL-SPECIFIC EPIGENOMICS IN ARABIDOPSIS

Jordi Moreno-Romero, Juan Santos-González, Claudia Köhler

Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, Uppsala, Sweden
Presenter e-mail: jordi.moreno@slu.se

Genomic imprinting is an epigenetic phenomenon by which certain genes are expressed in a parent-of-origin-specific manner. This process has been demonstrated to play an important role in animal and plant development. In plants, genomic imprinting occurs in the endosperm, an ephemeral tissue supporting embryo growth, similar to the nourishing role of the mammalian placenta. As the placenta, the endosperm develops after fertilization. Imprinted genes are marked by parental-specific epigenetic marks (the imprint) that are established in the gametes and maintained after fertilization. Although it is known that histone and DNA methylation marks are required to establish imprinted expression, the precise mechanism remains to be established. In order to understand how chromatin marks trigger imprinting regulation in plants, we optimized a procedure allowing to isolate endosperm-specific nuclei at early stages of Arabidopsis seed development followed by chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) or bisulfite sequencing (bs-seq). By crossing two Arabidopsis accessions that differ in large numbers of sequence polymorphisms we were able to differentiate the maternal or paternal origin of the epigenetics marks. Using that experimental setup, we were able to generate high-quality parental-specific endosperm epigenome profiles. Our study revealed that Polycomb-mediated H3 lysine 27 trimethylation (H3K27me3) is localized to DNA hypomethylated regions on the maternal genome, linking DNA demethylation and H3K27me3 to imprinted gene expression. We furthermore show that H3K27me3 marked regions are located at paternal centromeric and pericentromeric regions, unlike the euchromatic localization of this mark in vegetative tissues, suggesting a silencing role of Polycomb group proteins for paternal heterochromatic regions in the endosperm.

EVOLUTIONARY GENOMICS AND TRANSCRIPTOMICS OF ARTHROPOD CHEMOSENSORY SYSTEMS: SOFTWARE DEVELOPMENT AND EXPERIMENTAL APPROACHES

Cristina Frías-López[1,2], Sara Guirao-Rico[1,3], Francisca Cunha Almeida[4], Alejandro Sánchez-Gracia[1] and Julio Rozas[1]

1. Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain
2. Departament de Biologia Animal i Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain
3. Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Barcelona, Spain
4. Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), San Miguel de Tucumán, Argentina
Presenter e-mail: cristinafriaslopez@ub.edu

Unlike insects and vertebrates, the knowledge of the specific molecules involved in chemoreception in the major arthropod lineages other than hexapods comes exclusively from the comparative analysis of genome sequences. Indeed, the myriapods and chelicerates genomes contain genes encoding homologs of some insect membrane receptors and small soluble chemosensory proteins. The advent of RNAseq techniques can delimit much better (i.e., using organ specific transcriptomes) the specific genes and gene families involved in chemical communication in these organisms. The application of a *de novo* assembly or a mapping/alignments strategy will depend on the availability of a well annotated reference genome. Recently, we have developed TRUFA, a free web server designed to handle and analyze transcriptomics data in the absence of a reference genome. The pipeline integrates popular NGS tools for a full *de novo* transcriptome assembling, transcripts annotation, and differential expression analyses. Moreover, TRUFA has access to high-performance computing resources and incorporates the possibility of parallel computing. The performance of this new bioinformatics tool has been evaluated using four different published transcriptomic data sets. In order to advance in our understanding of the specific members of chemosensory families involved in the detection of chemical cues in myriapods, we sequenced the specific transcriptomes of the antenna (olfactory structures), the head (taste organs) and the remaining parts of the body of the centipede *Strigamia maritima* (*Chilopoda*) using Illumina paired-end libraries. Since, we have access to a completely sequenced and well annotated genome, all our analysis are based on the alignment of the reads from each different tissue to this reference. The integrative analysis of the results from this organ specific RNAseq and the comparative genomics of chemosensory families across arthropod species is allowing us to obtain a set of candidates to be the specific proteins involved in olfaction and taste in arthropods.

UNRAVELING EXTINCT GENOMES IN PRESENT DAY HUMAN GENOMES

Mayukh Mondal[1], Ferran Casals[2], Partha P. Majumder[3] and Jaume Bertranpetit[1]

1. Institut de BiologiaEvolutiva (UPF-CSIC), UniversitatPompeuFabra, Barcelona, Catalonia, Spain
2. Servei de Genòmica, UniversitatPompeuFabra, Barcelona, Catalonia, Spain
3. National Institute of BioMedical Genomics, Kalyani, West Bengal 741251, India
Presenter e-mail: mayukh.mondal@upf.edu

We have recently published a paper (Mondal et al. 2016) showing that an unknown archaic population (an extinct Eurasian hominin which is either Neanderthal nor Denisova) introgressed in the Andamanese and Indian populations. In this work we try to delimit which are the populations having had introgression from this unknown hominin. By using whole-genome sequences available and comparing to simulated data we hypothesized that Out of Africa (OOA) Eurasian populations of modern human have had introgression at least three times from extinct hominins. First from Neanderthal to all OOA populations, second all Asian and Pacific populations have had introgression from this unknown population and third, all Pacific populations having introgression from Denisova. In this follow up study we try to delimitate the geographic extent of the introgression with the unknown hominin population, which could be wider than initially proposed and encompass all Asian and Pacific populations.

FUNCTIONAL GENOMIC ANALYSIS OF GENE-BY-ENVIRONMENT INTERACTIONS ACROSS 250 ENVIRONMENTS

G. Moyerbrailean[1], A.L. Richards[1], C. Kalita[1], D. Kurtz[1], O. Davis[1], C. Harvey[1], A. Alazizi[1], D. Watza[1], Y. Sorokin[3], N. Hauff[3,4], X. Zhou[2], X. Wen[2], R. Pique-Regi[1,3], F. Luca[1,3]

1. Center for Molecular Medicine and Genetics, Wayne State University
2. Department of Biostatistics, University of Michigan
3. Department of Obstetrics and Gynecology, Wayne State University
4. College of Nursing, Wayne State University

Presenter e-mail: rpique@wayne.edu

Gene-by-environment interactions (GxE) determine common disease risk factors and biomedically relevant complex traits. However, quantifying how the environment modulates genetic effects on human quantitative phenotypes presents unique challenges. Environmental covariates are complex and difficult to measure and control at the organismal level, as found in GWAS and epidemiological studies. An alternative approach focuses on the cellular environment using in vitro treatments as a proxy for the organismal environment. These cellular environments simplify the organism-level environmental exposures to provide a tractable influence on sub-cellular phenotypes, such as gene expression. Expression quantitative trait loci (eQTL) mapping studies identified GxE in response to drug treatment and pathogen exposure. However, eQTL mapping approaches are infeasible for large scale analysis of multiple cellular environments. Recently, allele-specific expression (ASE) analysis emerged as a powerful tool to identify GxE in gene expression patterns by exploiting naturally-occurring environmental exposures. Here we characterized genetic effects on the transcriptional response to 50 treatments in 5 cell types. We discovered 1,455 genes with allele-specific expression (ASE) (FDR<10%) and 215 genes with GxE. We demonstrated a major role for GxE in complex traits. Genes with a transcriptional response to environmental perturbations showed a 7-fold higher odds of being found in GWAS. Additionally, 105 genes that indicated GxE (49%) were identified by GWAS as associated with complex traits. Examples include GIPR-caffeine interaction and obesity, and LAMP3-selenium interaction and Parkinson disease. Our results demonstrate that comprehensive catalogs of GxE interactions are indispensable to thoroughly annotate genes and bridge epidemiological and genome-wide association studies.

INDEPENDENT COMPONENTS OF GENE EXPRESSION IN ENDURANCE RUNNERS

Maria Maqueda[1,2], Emma Roca[3], Jose Manuel Soria[4], Alexandre Perera[1,2]

1. B2SLab, Dep. ESAII, Centre for Biomedical Engineering Research, Technical University of Catalonia (UPC), Pau Gargallo 5, 08028, Barcelona, Spain
2. Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Av. Monforte de Lemos 3-5, 28029,Madrid, Spain
3. Dep. Electronic Engineering, Center for Biomedical Engineering Research, Universitat Politècnica de Catalunya (UPC), Av. d' Eduard Maristany 10-14, 08019, Barcelona, Spain
4. Unitat de Genomica de Malalties Complexes, Institut de Recerca de l'Hospital de la Santa Creu i Sant Pau, Sant Antoni Maria Claret 167, 08025, Barcelona, Spain
Presenter e-mail: maria.maqueda@upc.edu

This study assessed the gene expression (GE) response from a group of 16 endurance runners after participating in an 82km competition. We used microarray technology (HuGene2.0st from Affymetrix, Inc., California) to analyse the genome-wide GE profile of runners' blood samples that were collected before and after the race. A linear regression model was fit to each transcript cluster (TC) expression value considering gender and completed distance as covariates. Statistically significant differentially expressed TCs were obtained from a moderated t-statistics performed for each coefficient in the model (adjusted p-value 5%, FDR). An over-representation analysis was applied to the differential TCs querying KEGG PATHWAY and The Gene Ontology (GO) databases. An independent component analysis (ICA) was computed to extract the independent expression modes among the differential TCs. 5084 distinct genes were prioritized being 37% down-regulated and 63% up-regulated. Over-represented biological pathways were mostly associated with: genetic information processing, infectious disease and immune system. ICA identified seven independent regulation expression modes as response to exercise. The intervention impacted heavily on the gene regulation processes, so that up to one fourth of the total variance related to coding RNA activity was captured as response to the exercise. This response produced extensive alterations in several pathways in the human biology, which we decomposed through a statistical method targeting independent expression modes.

DISENTANGLING THE COMMON AND SPECIFIC SOURCES OF VARIATION BETWEEN DIFFERENT BIOLOGICAL LAYERS IN MULTI-VIEW SINGLE-CELL SEQUENCING

Ricard Argelaguet[1], Florian Buettner[1], Oliver Stegle[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambrigde, UK
Presenter e-mail: ricard.argelaguet@ebi.ac.uk

Single-cell RNA sequencing is becoming a well-established routine that is revolutionising our understanding of cellular phenotypes. Interestingly, other data modalities are also starting to be assayed at the single-cell level, including epigenetics, proteomics and metabolomics, raising the question of how to jointly analyse these set of complex high-dimensional data sets using a statistically rigorous framework. Here we present single-cell Bayesian Group Factor Analysis (scBGFA), a multi-view generalisation of principal component analysis suited to the analysis of noisy single-cell sequencing data. scBGFA performs a joint dimensionality reduction for all views, yielding a low dimensional representation of the data which hopefully captures an inherent structure that might be masked by the noisy high-dimensional representation. Furthermore, scBGFA disentangles the variation unique to a single view and the variation shared between two or more views, thereby revealing hidden sources of covariation between different data modalities. We applied scBGFA to a recent data set of 61 embryonic stem cells generated by a technology called scMT-seq, a method which uses single-cell genome-wide bisulfite sequencing and RNA sequencing to perform a parallel profiling of the DNA methylation and the gene expression in single cells. Our results show the existence of several axes of variation related to known biological processes and suggest the existence of three subpopulations that are associated with different pluripotency potential and genome-wide methylation rate.

Posters

CHARACTERIZING THE IMPACT OF THE INCIDENTALOME IN CLINICAL GENE SEQUENCING PANELS

Josu Aguirre[1], Natàlia Padilla[1], Xavier De la Cruz[1,2]

1. Vall d'Hebron Research Institute (VHIR), Passeig de la Vall d'Hebron 119-129, 08035 Barcelona, Catalonia, Spain
2. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain
Presenter e-mail: josu.aguirre@vhir.org

The Incidentalome (the identification of unexpected pathogenic variants in genome-scale sequencing tests) has been an important source of concern in the clinical setting since it was first described (Kohane et al., 2006). For this reason, the presence of incidental findings in whole-genome and whole-exome sequencing experiments has been studied and documented. Surprisingly, this is not the case for gene sequencing panels, in spite of the fact that that they are broadly used with diagnostics purposes. Our main objective is to address this issue, exhaustively characterizing the Incidentalome in four representative gene panels (Illumina's TruSight One, Inherited, Cardio and Autism). To this end, we mapped a set of 81271745 natural variants from 2504 healthy unrelated individuals retrieved from the 1000 Genomes Project, focusing on those that happened in the coding region (201451 cases). We find that about 30% of the coding variants were of a pathogenic nature (potential cases of Incidentalome), with 566 amongst them belonging to the 56 ACMG reportable genes. We also show that there is a relationship between the Incidentalome and gene size, relevant for the design of diagnostics panels. These data give us an overall view of the relevance of the Incidentalome in gene panels. Next, to refine our study we repeated the previous analyses on a per individual basis, finding a substantial amount of pathogenic variants per healthy individual: between one and ten variants. In summary, the Incidentalome has a non-negligible presence in gene panels of clinical use, thus constituting a serious confusing factor.

Poster 2

FINDING AN OPTIMAL COMBINATION OF PATHOGENICITY PREDICTORS FOR THE CLINICAL SETTING

Elena Álvarez de la Campa[1,2], Xavier de la Cruz[1,3]

1. Vall d'Hebrón Research Institute (VHIR), Passeig Vall d'Hebron, 119-129, 08035 Barcelona, Catalonia, Spain
2. High Council for Scientific Research (CSIC), Barcelona Science Park, Helix Building, Room 02-19, C/ Baldiri Reixac 15-21, 08028 Barcelona, Spain
3. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain
Presenter e-mail: elena.alvarezdelacampa@gmail.com

In recent years, over 40 mutation prediction methods have been described, each based on a specific set of properties. These tools have relatively good success rates, in the 70-85% range. However, their results do not always coincide, and deciding which method (or methods) is preferable in each case is a difficult task, since different benchmark studies pointing to different predictors. Strict guidelines delimit the use of computational information in the clinical setting, due to the still moderate accuracy of in silico tools. These guidelines indicate that several tools should always be used and that full coincidence between them is required if we want to consider their results in medical decision processes. Application of this simple rule indeed decreases the error rate of in silico pathogenicity assignments. However, it is unclear how we shall proceed when predictors give contradictory results, a situation that happens with non-negligible frequency and this rule results in the rejection of potentially valuable information for a set of variants. In this work, we focus on these variants and develop specific predictors, for all the possible combinations of five known methods (SIFT, PolyPhen-2, PON-P2, CADD and MutationTaster2), to help improve the success rate of their annotation. We find that these specific predictors can discriminate between neutral and pathogenic variants, with a success rate different from random. They tend to outperform the constitutive methods, a trend that decreases as the performance of the constitutive predictor improves (e.g. with PON-P2 and PolyPhen-2). We also find that specific methods outperform standard consensus methods (Condel and CAROL).

GENOME-WIDE PATHWAY ANALYSIS OF CLINICAL PHENOTYPES IN SYSTEMIC LUPUS ERYTHEMATOSUS

Adrià Aterido[1], Antonio Julià[1], Patricia Carreira[2], Ricardo Blanco[3], Francisco Javier López-Longo[4], José Javier Pérez Venegas[5], Àlex Olivé[6], José Luís Andreu[7], Mª Ángeles Aguirre-Zamorano[8], Paloma Vela[9], Joan M. Nolla[10], José Luís Marenco de la Fuente[11], Antonio Zea[12], José María Pego[13], Mercedes Freire[14], Elvira Díez[15], María López-Lasanta[1], Mireia López Corbeto[1], Núria Palau[1], Raül Tortosa[1], Josep Lluís Gelpí[16], Devin Absher[17], Richard M Myers[17], Antonio Fernández-Nebro[18], Sara Marsal[1]

1. Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, 08035, Spain
2. Rheumatology Department, Hospital Universitario 12 de Octubre, Madrid, 28041, Spain
3. Rheumatology Department, Hospital Marqués de Valdecilla, Santander, 39008, Spain
4. Rheumatology Department, Hospital Universitario Gregorio Marañón, Madrid, 28007, Spain
5. Rheumatology Department, Hospital del SAS de Jerez de la Frontera, Cádiz, 11407, Spain
6. Rheumatology Department, Hospital Germans Trias i Pujol, Badalona, 08916, Spain
7. Rheumatology Department, Hospital Universitario Puerta de Hierro, Madrid, 28222, Spain
8. Rheumatology Department, Hospital Universitario Reina Sofía, Córdoba, 14004, Spain
9. Rheumatology Department, Hospital General de Alicante, Alicante, 03010, Spain
10. Rheumatology Department, Hospital Universitari de Bellvitge, Barcelona, 08907, Spain
11. Rheumatology Department, Hospital de Valme, Sevilla, 41014, Spain
12. Rheumatology Department, Hospital Universitario Ramón y Cajal, 28034, Madrid, Spain
13. Instituto de Investigación Biomédica de Vigo, Ourense y Pontevedra, 36204, Spain
14. Rheumatology Department, Hospital Universitario A Coruña, A Coruña, 15006. Spain
15. Rheumatology Department, Hospital Complejo Asistencial de León, León, 24001, Spain
16. Life Sciences, Barcelona Supercomputing Centre, Barcelona, 08034, Spain
17. HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, 35806, USA
18. Rheumatology Department, Hospital Regional de Málaga, Instituto de Investigación Biomédica de Málaga, Universidad de Málaga, 29010, Spain

Presenter e-mail: adria.aterido@vhir.org

Systemic lupus erythematosus (SLE) is a genetically complex disease characterized by the dysregulation of the immune system and a high phenotypic diversity. From a clinical point of view, this phenotypical heterogeneity is a major obstacle for the development of efficient therapies. Understanding the genetic basis of SLE phenotype heterogeneity would therefore be of high relevance to develop more efficient therapies. The objective of the present study was to identify the genetic variation associated with the most relevant phenotypes in SLE using a genome-wide pathway analysis (GWPA) approach. A total of 598,258 single-nucleotide polymorphisms were genotyped in a discovery cohort of 482 SLE patients collected from the rheumatology departments of 15 Spanish University Hospitals belonging to the Immune-Mediated

Inflammatory Diseases Consortium. Using a GWPA approach, we tested the association between 798 biological pathways and the main clinical phenotypes in SLE. The significant genetic pathways were subsequently tested for validation in an independent cohort of 425 SLE patients from the same ancestry. Finally, using an *in silico* functional analysis, we investigated whether the validated biological pathways are modulated by common SLE therapies. After multiple testing correction, we found two pathways significantly associated with the presence of oral ulcers and antinuclear antibodies in the discovery stage ($P_{FDR}<0.05$). In the replication stage, we validated the association between oral ulcers and the vascular endothelial growth factor (VEGF) pathway. Analyzing the functional effect of immunotherapies commonly used to treat cutaneous/mucocutaneous manifestations in SLE, we found highly significant perturbations of the gene expression of VEGF pathway ($P<5e-4$). Using a GWPA we have identified the genetic basis of a common phenotype in SLE. Our results suggest that drugs that target the VEGF pathway will have a higher probability of being effective in the treatment of oral ulcers in SLE.

SNAPSHOT: SINGLE NUCLEOTIDE POLYMORPHISMS SELECTION AND PROMOTER PROFILING

Daniel Boloc[1], Susana García[1], Patricia Gassó[1,2,3], Anna Gortat[1], Amalia Lafuente[1,2,3], Sergi Mas[1,2,3], Natalia Rodríguez[1], Teresa Torres[1], Miquel Bernardo Arroyo[2,3,4,5]

1. Department of Pathological Anatomy, Pharmacology and Microbiology, University of Barcelona, Barcelona, Spain.
2. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.
3. Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain.
4. Barcelona Clinic Schizophrenia Unit (BCSU), Psychiatry Service, Hospital Clínic de Barcelona, Barcelona, Spain.
5. Department Psychiatry and Clinical Psychobiology, University of Barcelona, Barcelona, Spain.

Presenter e-mail: danielboloc@gmail.com

The selection of a Single Nucleotide Polymorphism (SNP) can be very time consuming done by bibliographic methods. Moreover, the SNP is not always put in its genomic context in a visual manner. Thus, we propose a new automated and visual method of selecting SNP based on their position in the genome and the possible regulatory structures it could disrupt. In this manner, it allows scientists to prioritize SNPs in genotyping and disease studies. SNaPShOt is bundled into a web portal. It maps the different structures affecting the expression of a gene, especially in the promoter. Such structures are: transcription factors, histones, enhancers, promoters, promoter flanking regions and miRNA sites. Additionally it provides eQTL and linkage disequilibrium (LD) properties for the SNPs giving more clues about other indirectly associated SNPs. Possible disruptions of the aforementioned structures affecting gene transcription are reported using multiple resource databases. It has a user-friendly interface, allowing single or list queries of genes and coordinates. Therefore, SNaPShOt is the only portal providing useful promoter characterization in both report format and visual images (through *circos*) for SNP selection.

A DIGENIC MODEL BETWEEN *PTH1R AND ATP4A GENES*, EXPLAIN A FAMILY WITH GASTRIC NEUROENDOCRINE TUMORS, HYPOTHYROIDISM AND ARTHRITIS

Oriol Calvete[1, 2], Maite Herraiz[3], Jose Reyes[4], Ana Patiño[5] and Javier Benitez[1, 2]

1. Human Genetics Group, Spanish National Cancer Research Center (CNIO). 28029 Madrid, Spain.
2. Network of Research on Rare Diseases (CIBERER). 28029 Madrid, Spain.
3. Department of Gastroenterology. Clínica Universidad de Navarra. Pamplona Spain.
4. Department of Gastroenterology, Hospital INCA. 07300 Majorca, Spain.
5. Department of Pediatrics and Clinical Genetics Unit, University Clinic of Navarra, Pamplona, Spain.
Presenter e-mail: ocalvete@cnio.es

Type I gastric neuroendocrine tumors (gNETs) classically arise from hypergastrinemia in patients who present autoimmune atrophic gastritis that can cause hyperplasia of gastric enterochromaffin-like (ECL) cells. gNETs cause parietal cells (PC) destruction, which are responsible of gastric acid secretion through ATP4A proton pump and intrinsic factor production. Therefore, lack of PC cause achlorhydria and megaloblastic anemia due to the vitamin B12 malabsorption. Gastrin directly upregulates gastric acid production through cholecystokinin receptor-2 (CCKR2). By NGS, we recently identified an homozygous deleterious mutation in the *ATP4A* gene that explained an aggressive familial form of gNETs. The achlorhydria was described as the causative parameter of hypergastrinemia that develops gNET rather than classic tumorogenesis process. By feeding the knockin mice for the human mutation with HCl-water, we restored gastric homeostasis and prevented the premalignant conditions that lead gNETs. A second family with eleven siblings and three affected members with classic clinic of gNETs plus hypothyroidism and rheumatoid arthritis was also studied. By whole exome sequencing, we identified another missense mutation in heterozygosis in the same gene (*ATP4A*). Carriers of this variant had low ferritin and vitamin B12 levels but without gNETs development. A second heterozygous mutation in a novel gene (*PTH1R*$^{p.E546K}$) was also uncovered. Gastrin/CCKR2 also activates PTHLH/PTH1R regulation factor, which is involved in PC development and gastric homeostasis. Activation of PTH/PTH1R, which is upregulated by thyrotropin in thyroid, is also involved in *RANKL* expression to regulate bone homeostasis. Thyrotropin and *RANKL* expression were found deregulated in those members carrying *PTH1R* mutation plus hypothyroidism and rheumatoid arthritis suggesting a link between PTH1R gene and these pathologies and gastric disease. Both mutations in *ATP4A* and *PTH1R* genes suggest a digenic model for this family and contribute to the function and viability of PC and lead to the achlorhydria that drives hypergastrinemia and gNETs.

IDENTIFICATION OF NEW SINGLE NUCLEOTIDE POLYMORPHISMS INVOLVED IN ADHERENT-INVASIVE *E. COLI* PATHOTYPE BY COMPARATIVE GENOMICS

Carla Camprubí-Font[1], Mireia Lopez-Siles[1], Meritxell Ferrer-Guixeras[1], Laura Niubo-Carulla[1], Carles Abellà-Ametller[1], Jesús Garcia-Gil[1], Margarita Martinez-Medina[1]

1. Biology Department, University of Girona, Girona, Spain
Presenter e-mail: carla.camprubi@udg.edu

To date, adherent-invasive *Escherichia coli* (AIEC) can only be identified phenotypically by cell culture infection assays, which is a highly time-consuming technique. Due to its implication in Crohn's disease pathogenesis, having new molecular tools for AIEC identification would be of great significance. The aim of this study was to identify signature sequences specific of the AIEC phenotype by comparative genomics. The genome of three *E.coli* strain pairs from different phylogroups, which consisted of one AIEC and one non-AIEC clonally identical, were sequenced by Illumina and PacBio platforms. SPAdes was used for genome assembly and Harvest for comparative genomics between strains using a reference genome (AIEC UM146). PCR and Sanger sequencing were performed to verify the presence of Single Nucleotide Polymorphisms (SNPs) causing non-synonymous amino acid change in coding regions. Finally, SNPs distribution was evaluated in a collection of 22 AIEC and 29 non-AIEC isolates. Pearson's x^2 test or Mann-Whitney test were employed, as required, to analyse the nucleotides present in each SNP according to AIEC phenotype, adhesion and invasion indices. Comparative genomics identified a total of 18 SNPs between the strains of the D-phylogroup pair, 17 in the B2-pair and 30 in B1-pair accomplishing the selection criteria. From those, 24 SNPs (found in 13 genes) were confirmed by Sanger sequencing and further analysed in the strain collection. Three of the SNPs-encompassing genes were related with adhesion/invasion and two with stress tolerance. Three SNPs obtained differential nucleotide distribution between AIEC and non-AIEC strains ($p<0.016$), showing particular nucleotides being more associated with one phenotype or another. Of interest, these SNPs also presented association with adhesion and invasion strain capacities ($p<0.006$). To conclude, our study corroborates that there is not any AIEC-specific genetic marker that is widely distributed across all AIEC strains. Nonetheless, three SNPs putatively involved with AIEC phenotype have been described.

Poster 7

SEX IN AN ASEXUAL YEAST: GENOMIC EVIDENCE OF A SEXUAL CYCLE AND ERROR-PRONE MATING TYPE SWITCHING IN THE OPPORTUNISTIC PATHOGEN CANDIDA GLABRATA

Laia Carreté[1,2], Cinta Pegueroles[1,2], Cécile Fairhead[3], Toni Gabaldón[1,2,4]

1. Bioinformatics and Genomics Programme. Centre for Genomic Regulation (CRG), the Barcelona Institute of Science and Technology (BIST). Dr. Aiguader, 88. 08003 Barcelona, Spain
2. Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra (UPF).
3. Institut de Génétique et Microbiologie, UMR8621 CNRS-Université Paris Sud, Bât 400, UFR des Sciences, Orsay Cedex, F 91405, France.
4. Institució Catalana d'Estudis Avançats (ICREA)

Presenter e-mail: Laia.Carrete@crg.eu, tgabaldon@crg.es

Candida glabrata is considered an asexual species despite it maintains genes homologous to those involved in mating in *Saccharomyces cerevisiae*. A previous study found distinct mating types in *C. glabrata* natural populations, suggesting that this yeast may be able to mate in certain undiscovered conditions, similar to its distant relative *Candida albicans* (Muller *et al.,* 2008). Importantly, a later study based on clinical samples found that MLST-based clades showed a strong bias in the presence of only one of the two mating types, suggesting a low incidence of mating type switching (Brisse *et al.,* 2009) and fueling the debate about the hidden sexuality of *C. glabrata*. Here, we set out to address whether *C. glabrata* is able to mate and switch mating types using whole genome data and to perform for the first time a genome-wide characterization of the population structure in this species. Here, we analysed the genomes of 33 strains of *C. glabrata,* using comparative genomics with *C. albicans* and *S. cerevisiae*. First, we refined the population structure of those species using SNP data and assessed levels of genetic variation in *C. glabrata* genes, which were compared with those of *C. albicans* and *S. cerevisiae*. Second, we investigated different genomic variants, computed recombination rates and estimated the strength of purifying selection to elucidate whether *C. glabrata* is able to mate. We determined that *C. glabrata* is highly genetically structured , having high genetic differentiation between clades and low genetic differentiation within them. The existence of some form of sexual cycle is strongly supported by similar patterns of evolutionary constraints in reproduction related genes in *C. glabrata, C. albicans,* and *S. cerevisiae*. Importantly, our results are consistent with previous reports of successful mating-type switching in MTL1 from **a** to alpha, but also reveal frequent illegitimate recombination at the other MTL loci. In conclusion, we provide for the first time a genome-wide perspective of the genetic diversity of the species which reinforces the idea that *C. glabrata* is able to mate at a very low frequency.

Muller H., et al.: The asexual yeast *Candida glabrata* maintains distinct a and alpha haploid mating types (2008). Eukaryotic Cell. 7(5):848-58

Brisse S., et al.: Uneven distribution of mating types among genotypes of *Candida glabrata* isolates from clinical samples (2009). Eukaryotic Cell. 8(3):287-95

IN DEPTH EVALUATION OF THE FECAL MICROBIOTA: A FLEXIBLE AND OPTIMIZED ION TORRENT 16S rRNA GENE-BASED ANALYSIS WORKFLOW

Adrià Cereto-Massagué[1], Helena Torrell[1], Maria Guirro[1], Lorena García[1], Adrià Hernàndez[1], Lluís Arola[1], Núria Canela[1]

1. Centre for Omic Sciences (COS), Universitat Rovira i Virgili - Eurecat, Avinguda Universitat 1, 43204 Reus, Spain
Presenter e-mail: adrian.cereto@urv.cat

Microbes found in human intestinal tract enable nutrient absorption from otherwise non-digestible food, as they perform functions not present in their hosts, including essential vitamin and aminoacid synthesis. Those present in skin and mucosae protect their hosts against pathogenic elements such as parasites or other microbes. Likewise, foods harbour complex microbial communities that are of technological importance and invest them with desirable sensorial characteristics Eg., food microbial ecosystems condition final product properties in dairy, pickles, bread, wine, cheese, beer and other fermented foods. Identification of microbial species is often carried by PCR amplification and further 16S ribosomal RNA region sequencing. This is composed of conservative regions and nine hypervariable regions that give information for bacterial identification enabling individual read classification, but it is not clear which of these regions is the best choice for sequencing in metagenomics analysis. This approach can be used even when only traces or poor quality DNA are available and it does not require complete reference genome sequences. Relative performance of two distinct library preparation methods for the PGM Ion Torrent Next Generation Sequencing platform were compared: custom fusion-tag primers and a commercial kit, as Ion 16S Metagenomics kit from Life Technologies. A known community of 11 representative bacterial strains individually cultured was examined through both methods of library preparation and further next-generation sequencing. The sequencing results were analyzed using two different workflows: Ion Torrent's Ion Reporter and QIIME, the latter with several different configurations regarding OTU picking method, taxonomy assignment method and taxonomy database. Finally, a flexible metagenomics workflow with the most appropriate configurations for accurate data abundances and identifications were described.

A COMPARATIVE GENOMIC STRATEGY FOR TARGET DISCOVERY AND EVOLUTIONARY INFERENCES IN PATHOGENIC BACTERIA: *Stenotrophomonas sp*. AS A CASE STUDY.

Oscar Conchillo-Solé[1], Daniel Yero[1], Aniol Valera[1], Xavi Coves[1], Sonia Martinez-Servat[1], Isidre Gibert[1], Xavier Daura[1,2]

1. Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain.
2. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain
Presenter e-mail: ocs@bioinf.uab.es

The genus *Stenotrophomonas* comprises organisms with clinical and environmental importance, and it is genetically and phenotypically heterogeneous. The most predominant species, *S. maltophilia*, besides its biotechnological importance, is a leading multi-drug resistant (MDR) pathogen in hospitals worldwide and has been associated with serious infections in pediatric and immunocompromised patients. Here we present a comparative genomic strategy aimed to find new targets for virulence and resistance in *S. maltophilia* while also to unravel the complex inter- and intraspecies phylogenetic relationships in this genus. We have sequenced four new relevant strains of *S. maltophilia* form clinical origin with the Illuminia Miseq system. Clean reads were assembled and contigs submitted to NCBI for their annotations. In addition, all available genome sequences at NCBI for the genus *Stenotrophomonas* that passed the quality criteria (60 in total) were used in this study. Our comparative approach was based on all-versus-all theoretical proteome comparisons and ortholog identification was performed using a full reciprocal reverse hit algorithm identifying, this way, those proteins unique for each organism and those shared by all of them (core-proteome) or a subgroup. With this data the algorithm grouped organisms by two different methods (both represented as a Neighbor Network): by how many ortholog proteins they share or by their genetic distances based on the concatenated core proteins. Both clustering methods agreed to arrange the taxa in the same clades. Interestingly, some undetermined *Stenotrophomonas* and the only available strain of *S. pavanii* clustered into the clade including all *S. maltophilia* strains. All members of this clade share 41 unique proteins which include several antibiotic resistance related genes, membrane proteins, ABC transporters and regulatory proteins. Moreover, the clustering strategy highlighted a defined group of phenotypically relevant *S. maltophilia* MDR strains from clinical origins also sharing a set of unique proteins.

Acknowledgments:
This work has been supported by Spanish MICINN/FEDER (BIO2015-66674-R) and Catalan AGAUR (2009SGR-00108).

CHARACTERISATION OF THE ZEBRAFISH 3D GENOME

Noelia Díaz[1], Benjamín Hernández-Rodríguez[1], Kai Kruse[1], Juan M. Vaquerizas[1]

1. Max Planck Institute for Molecular Biomedicine, Röntgenstraße 20, 48149 Münster, Germany.
Presenter e-mail: noelia.diaz@mpi-muenster.mpg.de

One of the key players in regulating the transcriptional program of cells is the three-dimensional organization of chromatin in the interphase nucleus. In species ranging from human to worm, the genome adapts a specific topology into compartments and topologically associating domains. This domain-based organization of the genome seems to be strongly conserved across species. In Zebrafish, the 3D structure and conformational changes of its genome have remained elusive - in part due to its genome duplication event and the presence of ohnologous genes scattered throughout its genome. Here, we present the first set of genome-wide chromatin conformation capture (Hi-C) studies of the zebrafish genome at a 10kb resolution. We were able to characterize the structural features of chromatin topology using interaction maps obtained at 48 hours post fertilization. We show how chromatin in fish embryos has a similar organization as other higher eukaryotes with topological domains readily apparent at the megabase scale. Our results provide a first insight into the three dimensional organization of chromatin in zebrafish and open the door to further studies on conformational aspects of zebrafish chromatin.

Poster 11

NETWORK TOPOLOGY AFFECTS THE EVOLUTION OF ENZYME-CODING GENES

Begoña Dobón, Ludovica Montanucci, Hafid Laayouni, and Jaume Bertranpetit

IBE Evolutionary Biology Institute (CSIC-UPF), Universitat Pompeu Fabra, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain
Presenter e-mail: begona.dobon@upf.edu

To determine the role of topology on shaping the evolution of enzyme-coding genes, we performed a topological and evolutionary analysis of the human metabolic network at different scales: from the whole metabolic network, to metabolic subsystems, to metabolic pathways. Pathways were described as directed reaction graphs, where nodes represent enzymatic reactions and by extension, the genes that encode those enzymes. For each node we measured its connectivity, centrality and position within the pathway. For each enzyme-coding gene we studied signals of positive selection at: i) population level, by a machine-learning algorithm (Pybus et al, 2015), and at ii) inter-specific level during the divergence of rodents and primates. Strength of purifying selection was estimated by dN/dS ratio. At inter-specific level, genes under positive selection code for enzymes whose products are substrates of many reactions (high out-degree) or are located in a central position within the network (high closeness). At intra-specific level, the frequency of the selected allele is affected by the gene's connectivity: the selected allele in genes with high out-degree will not reach fixation. From a functional point of view, genes encoding enzymes from the inner core of metabolism show stronger purifying selection than those belonging to pathways of the outer layer. The opposite trend is found at population level: enzymes from the inner and intermediate metabolism show a more relaxed evolution. This study shows the importance of taking into account metabolic network structure when analyzing positive and purifying selection in enzyme-coding genes.

DECONSTRUCTING OBSSESIVE-COMPULSIVE DISORDER BY WHOLE EXOME SEQUENCING

Laura Domenech[1], Kelly Rabionet[1], Georgia Escaramís[1], Eva Real[2], Daniel Trujillano[1], Hana Susak[1], Mattia Bosio[1], Stephan Ossowski[1], Angel Carracedo[3], Pino Alonso[2], Xavier Estivill[1].

1. Centre for Genomic Regulation (CRG), The Barcelona Institue of Science and Technology, Barcelona, Spain.
2. Hospital Universitari de Bellvitge, IDIBELL, L'Hospitalet de Llobregat, Spain.
3. Universidad de Santiago de Compostela, Spain.
Presenter e-mail: laura.domenech@crg.eu

Obsessive-compulsive disorder (OCD) is a neuropsychiatric condition that affects 1-3% of the population worldwide and that is listed by the World Health Organization as the tenth most disabling illnesses of any kind. OCD is characterized by intrusive unwanted thoughts, fears or images (obsessions) and/or ritualized behaviours or mental acts (compulsions), which are performed to relieve the anxiety and/or distress caused by the obsessions. Nevertheless, current available treatments for this disorder only provide partial relief of the symptoms. Therefore, improving our understanding of the biology of OCD is an essential step towards an effective and personalized treatment. Family and twin studies demonstrated that OCD involves both environmental and polygenic risk factors. However, despite an abundance of candidate genes, linkage studies and GWAS, very little progress has been made to elucidate the genetic causes of this disorder. For this reason, this project aims to identify genetic profiles associated to OCD that will contribute to explain the aetiology of OCD. We performed Whole Exome Sequencing in 306 OCD cases and 611 controls followed up with Rare Variant Association Study and GWAS-like association study, identifying more than 350 genes harbouring a higher number of mutations in cases compared to controls. Stronger evidence of association for these significant genes will be obtained through replication in larger cohorts of OCD cases and controls by a targeted resequencing study. Moreover, we are validating some interesting mutations found in the OCD cases through functional studies in neuronal cell lines and zebrafish models.

Poster 13

REMOTE HOMOLOGY DETECTION BY PSEUDO-DOMAIN PREDICITION

Ramon Duran[1], Oscar Conchillo-Solé[1], Xavier Daura[1,2]

1. Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona
2. Institució Catalana de Recerca i Estudis Avançats (ICREA)
Presenter e-mail: ramondur@gmail.com

Identifying proteins that are functionally related in different bacterial species is one of the most challenging problems of the post-genomic era. The functional regions of a protein are called domains and tend to be conserved throughout evolution. By predicting them we can identify proteins with equivalent domains and conclude that they will have a similar function. In this context, we present a program for remote protein-homology detection by pseudo-domain prediction (REHDBP^2). The program starts by doing an iterative Jackhmmer search for the query protein against a given database. The result is used to build a multiple alignment, from which the conservation of every position in the alignment is calculated and domain boundaries are predicted. The multiple alignment of each resulting domain is then used to build a Hidden Markov Model (HMM), with which a second database (or the same one) is then searched. The result of this pipeline contains the proteins with the most significant matches for each HMM profile, as well as additional information extracted from the UniProt database. In addition, it displays all genera for which each possible combination of domains of the query protein is found, providing insight into the phylogeny of the protein. This method is still in development and is currently being used to identify putative bacterial virulence factors and to detect their presence along the bacterial tree of life.

SYSTEMS ANALYSIS OF INTRACELLULAR pH VULNERABILITIES FOR CANCER THERAPY

Erez Persi[1, *, #], Miquel Duran-Frigola[2, *], Mehdi Damaghi[3], John L. Cleveland[4], William R. Roush[5], Patrick Aloy[2, 6], Robert J. Gillies[3], Eytan Ruppin[1, #]

1. Center for Bioinformatics and Computational Biology, Institute of Advanced Computer Studies, Department of Computer Science, University of Maryland College Park, Maryland 20742, USA
2. Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain
3. Department of Cancer Imaging and Metabolism, Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA
4. Department of Tumor Biology, Moffitt Cancer Center & Research Institute, Tampa, Florida 33612, USA
5. Department of Chemistry, The Scripps Research Institute, 110 Scripps Way, Jupiter, USA
6. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain
Presenter e-mail: miquel.duran@irbbarcelona.org

A hallmark of cancer metabolism is a reverse pH-gradient manifest as acidification of the extracellular milieu and alkalization of the cytoplasm. While consequences of extracellular acidosis are known, the role of intracellular alkalization is not understood. Here we develop a computational methodology that explores how intracellular pH (pHi) is coupled to network-wide metabolic alterations, by reconstructing and integrating enzymatic pH-dependent activity profiles into genome-scale metabolic models. We show that at alkaline pHi cancer proliferation rate is optimal and fundamentally coupled to increased glycolysis and adaptation to hypoxia. Moreover, lowering pHi is revealed as a vulnerability for cancer cells that disables these metabolic adaptations, where inhibition of systems-informed metabolic targets selectively kills acid-adapted malignant breast cancer cells. This systems approach thus reveals essential roles of pHi in cancer metabolism, establishes a new therapeutic strategy, and provides a conceptual and computational framework for exploring roles of pHi in other fields in biology.

HOST RESPONSE PROCESSES AND HOST-PARASITE INTERACTIONS IN *TRYPANOSOMA BRUCEI* INFECTIONS

Pau Erola[1], Siddharth Jayaraman[1], Sofie Demeyer[1], Christiane Hertz-Fowler[2], Michael P. Barrett[3], Liam J. Morrison[1] & Tom Michoel[1]

1. Roslin Institute, University of Edinburgh, Bush Estate, Midlothian EH25 9RG
2. Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB
3. Wellcome Trust Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow G12 8TA

Presenter e-mail: pau.erola@roslin.ed.ac.uk

African trypanosomes are unicellular protozoan parasites that infect humans, livestock and wild animals. In Africa, trypanosomiasis has historically affected agricultural production and animal husbandry, having a high economic and social impact. The use of RNA sequencing technology can help us to better understand the biology of host response processes and how they vary between different host and parasite genotypes, to enable better understanding of the relative contribution of host susceptibility and pathogen virulence to disease severity. To this aim two genetically distinct strains of *Trypanosoma brucei*, TREU927 (the genome reference strain) and STIB247, were inoculated into 33 susceptible experimental BALB/c mice and 40 tolerant C57BL/6 mice, and the fold-change expression levels with respect to an uninfected control group (8 replicates for each mice strain) were measured at several time points after infection. We developed a novel computational method to reconstruct modules of co-expressed genes, which allows the tracking of the conservation and divergence of modules across the different host-parasite specific sub-datasets. Gene set enrichment analysis of these modules revealed significant genotype-independent as well as host-parasite genotype-specific changes in expression of genes involved in various immune response pathways. These results elucidate the presence of different pathogenic mechanisms underlying these parasitic infections and have the potential to inform the design of tools to control animal trypanosomiases.

Poster 16

PERGOLA: BOOSTING VISUALIZATION AND ANALYSIS OF BEHAVIORAL LONGITUDINAL DATA BY UNLOCKING GENOMIC ANALYSIS TOOLS

J. Espinosa-Carrasco[1,2], I. Erb[1,2], Marta Fructuoso-Castellar[1,2], M. Dierssen[1,2], C. Notredame[1,2]

1. Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain.
2. Universitat Pompeu Fabra (UPF), Barcelona, Spain
Presenter e-mail: joseantonio.espinosa@crg.cat

The analysis of complex temporal sequences of behavior is an important tool for understanding the nervous system. Thanks to the rapid developments of technologies in the field of behavioral neuroscience it has become possible to collect large amounts of longitudinal data or big behavioral data. In order to understand these data whose nature is primarily sequential, it is essential to have good tools for its visualization and analysis. Dealing with such a huge amount of sequential data is not a new problem in biology. Genomics, for instance has provided us with widely used tools for genome visualization such as genome browsers as well as very efficient tools for sequence analysis such as BEDtools, among others. So why not to apply these tools for the understanding of longitudinal behavioral data? Here we present the Python bEhavioRal GenOmebrowser LibrAry (Pergola), a python library that makes longitudinal behavioral data compatible with popular desktop genome browsers and genomic analysis tools. Pergola accomplishes data format conversion using a simple correspondence between behavioral and genomic data. Behavioral longitudinal data can be then displayed in genome browsers, simplifying both data aggregation and the integration with other sources of data such as environmental data or downstream analyses. Data conversion also allows the user to analyze the data using genomic tools. We show some applied examples of behavioral data visualization and analysis on different model organisms.

THE UNIQUE OLEATE DIOL-SYNTHASE ACTIVITY OF *PSEUDOMONAS AERUGINOSA*: DISCOVERY OF A NEW BACTERIAL DI-HEME CYTOCHROME C PEROXIDASE SUBFAMILY

Mónica Estupiñán[1,2], Daniel Álvarez-García[3,4], Xavier Barril[3,4,5], Pilar Diaz[1]*, M. Ángeles Manresa[2]

1. Department of Genetics, Microbiology and Statistics, Faculty of Biology.
2. Department of Biology, Sanity and Environment, Faculty of Pharmacy and Food Science. University of Barcelona.
3. Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain.
4. Institut de Biomedicina de la Universitat de Barcelona (IBUB), PCB-Edifici Hèlix Baldiri Reixac, Barcelona, Spain.
5. Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys, Barcelona, Spain.
Presenter e-mail: mestupinan@ub.edu

Pseudomonas aeruginosa displays the ability to perform bioconversion of oleic acid into a class of hydroxylated fatty acids known as oxylipins. The biotechnological applications of oxylipins have extensively been studied, constituting important emulsifying agents in food and cosmetics industries, acting as antibacterial or antifungal substances, or being used as intermediate compounds for fine chemistry, with an important role in the pharmaceutical area (1–4). A diol synthase activity is responsible for such a conversion, which proceeds through the dioxygenation of oleic acid to release hydroperoxide 10-H(P)OME ((10*S*)-hydroxy-(8*E*)-octadecenoic acid), followed by conversion of the hydroperoxide intermediate into 7,10-DiHOME ((7*S*,10*S*)-dihydroxy-(8*E*)-octadecenoic acid), both of which accumulate in the culture supernatant by means of the identified transporter ExFadlO (5). Moreover, hydroperoxide 10-H(P)OME is spontaneously converted into the monohydroxylated form of oleic acid (10*S*)-hydroxy-(8*E*)-octadecenoic acid (10-HOME), which also accumulates in the supernatant (5,6). Through *P. aeruginosa* genome mutant screening, we identified the genes responsible for the two enzymatic activities: a 10*S*-dioxygenase encoded by *PA2077* catalyses the first step of the reaction, whereas the diol-synthase encoded by PA2078 converts the hydroperoxide into 7,10-DiHOME by a 7,10-diol synthase activity (6). Analysis of the amino acid sequence of both enzymes revealed the presence of two heme-binding motifs (CXXCH) on each protein. Phylogenetic analysis showed the relation of both proteins to bacterial di-heme cytochrome c peroxidases (Ccps), similar to *Xanthomonas sp.* 35Y rubber oxidase RoxA. Structural homology modelling of PA2077 and PA2078 was achieved using RoxA (pdb 4b2n) as a template. From the 3D model obtained, presence of significant amino acid variations in the predicted heme-environment was found. Taking together the *in silico* and *in vivo* results obtained we conclude that enzymes PA2077 and PA2078 are the first described members of a new subfamily of bacterial peroxidases, designated as Fatty acid-di-heme Cytochrome c peroxidases (FadCcp)(7).

References:

1. Martín-Arjol et al., (2010). *Chem Phys Lipids* 163: 341–346
2. Hou CT (2009). *N Biotechnol* 26: 2–10
3. Bajpai VK et al., (2009). *N Biotechnol* 26: 122-130

4. Martínez et al., (2010). *J Biol Chem* 285: 9339–9345
5. Martínez et al., (2013). *Biochimie* 95: 290–298
6. Estupiñán et al., (2014). *Biochim Biophys Acta Mol Cel Biol Lipids* 1841: 1360–1371
7. Estupiñán et al., (2015). *PLoS One* 10: e0131462

Acknowledgements:
MINECO CTQ2010-21183-C02-02/CTQ2014-59632-R and Red Temática AC2015-00008-00-00; 2009-2013 SGR 00327/2014 SGR 534 and Biotechnology R+D+I Reference Network, Generalitat de Catalunya

3'UTR-EMBEDDED ALUS AS FACILITATORS OF THE GENESIS OF PROCESSED PSEUDOGENES AND THE CAPTURE OF HOST GENES BY HERPESVIRUSES

Domènec Farré[1,2], Pablo Engel[1,2], Ana Angulo[1,2]

1. Departament de Biomedicina, Facultat de Medicina, Universitat de Barcelona, Casanova 143, 08036 Barcelona, Spain
2. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Rosselló 149-153, 08036 Barcelona, Spain
Presenter e-mail: domenecfarre@ub.edu

The functional significance of Alu elements, which are highly abundant in the human genome, is still a subject of debate. It has been demonstrated that primate Alu and rodent Alu-like elements are retrotransposed by a mechanism driven by the proteins encoded by LINE1 (L1) elements, the same machinery that generates the processed pseudogenes (PPs). In this study we hypothesize that Alu elements embedded inside 3' untranslated regions (UTRs) could be involved in the genesis of PPs. We discovered that the existence of 3'UTR-embedded Alu elements, but not of other retroelements, is overrepresented in genes source of PPs. We found a similar overrepresentation for 3'UTR-embedded B1 (Alu-like) elements in PP parent genes of mouse and rat. Interestingly, we also demonstrated that the overrepresentation of 3'UTR Alus is particularly significant in PP parent genes with low germline gene expression level. Altogether our results suggest a novel role for Alu or Alu-like elements inside 3'UTRs as facilitators of the genesis of PPs, particularly in lowly expressed genes. Finally, we provide data that support the hypothesis that the L1 machinery, aided by the presence of 3'UTR-embedded Alus, is also the system that herpesviruses use to incorporate host genes to their genomes.

REPRODUCIBLE IN-SILICO OMICS ANALYSES ACROSS CLOUDS AND CLUSTERS WITH NEXTFLOW

Evan W. Floden[1,2], Paolo Di Tommaso[1], Maria Chatzou[1,2], Pablo Prieto Barja[1,2], Emilio Palumbo[1], Cedric Notredame[1]

1. Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain
2 Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
Presenter e-mail: evan.floden@crg.eu

Reproducibility has become one of biology's most pressing issues. This impasse has been fueled by the combined reliance on increasingly complex data analysis methods and the exponential growth of biological datasets. When considering the installation, deployment and maintenance of bioinformatic pipelines, an even more challenging picture emerges due to the lack of community standards. We introduce Nextflow, a pipeline orchestration tool that has been designed to ease deployment and guarantee reproducibility across platforms. It is a computational environment, which provides a domain specific language (DSL) to simplify the writing of complex distributed computational workflows in a portable and replicable manner. It allows the seamless parallelization and deployment of any existing application with minimal development and maintenance overhead, irrespective of the original programming language.

EMPIRICAL VALENCE BOND SIMULATIONS OF THE AZ28 CATALYTIC ANTIBODY FOR THE OXY-COPE REARRANGEMENT

Martin Floor[1], Luis Agulló[1], Gerald Zapata-Torres[2], and Jordi Villà-Freixa[1]

1. Computational Biochemistry and Biophysics lab, Research Group on Bioinformatics and Medical Statistics, Department of Biosciences, Universitat de Vic - Central University of Catalonia
2. Molecular graphical unit, Department of Inorganic and Analytical Chemistry, Universidad de Chile.
Presenter e-mail: martin.floor@uvic.cat

The oxy-Cope rearrangement is a [3.3] sigmatropic reaction that has been intensively studied from both experimental and theoretical perspectives. The antibody AZ28m was developed against a transition state analog (TSA) for this reaction and has been reported to catalyze the reaction. The germline-encoded version of this antibody (AZ28g) is a better catalyst (k_{cat} = 0.80 min^{-1}) than its mature form (k_{cat} = 0.023 min^{-1}), however the former binds poorer the TSA (k_d = 670 nM) than the latter (k_d = 17 nM). Such apparent inverse correlation of activity has been a subject of discussion in theoretical studies. These studies concluded that there exist conformational differences between the TSA and the TS that lead to differential binding events.
Here we report Empirical Valence Bond (EVB) simulations of the AZ28 system that account for their experimental activation free energies. Linear Interaction Energies (LIE) calculations were used to assess the differences in binding between substrate, product and transition state conformations. This allowed to explain the stabilizing effect of the antibody over different regions of the reaction coordinate. Additionally, mutantions were evaluated to account for the experimental differences in catalysis, between the AZ28m and AZ28g antibodies, in order to understand the observed inverse correlation. This work is aimed at testing the predictive power of the EVB methodology within the context of an artificial system, which could be useful in *de novo* enzyme design applications for reactions that do not have natural enzymatic counterparts.

UNVEILING THE ROLE OF TKTL1 IN CANCER METABOLISM THROUGH GSMMS

Effrosyni Karakitsou[1,2], Carles Foguet[1,2], Ines Baptista[1,2], Pedro de Atauri[1,2], Marta Cascante[1,2]

1 Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Av Diagonal 643, 08028, Barcelona, Spain
2 Institute of Biomedicine of University of Barcelona (IBUB) and Associated unit to CSIC, Av Diagonal 643, 08028, Barcelona, Spain
Presenter e-mails: e.karakitsou@ub.edu , cfoguet@ub.edu, martacascante@ub.edu

Development of malignancy in cancer is accompanied by a complete metabolic reprograming that provides advantages to cancer cells in terms of energy production and synthesis of biomolecules and is essential for tumor progression. Understanding metabolic reprograming is key to identifying metabolic vulnerabilities in cancer that can be exploited to selectively target cancer cells [1]. Genome-Scale Metabolic Models (GSMMs), which are mathematical representations of the entire metabolic reaction complement encoded by an organism's genome, serve as platforms where a wide range of omics data can be integrated to predict metabolic phenotype. Therefore, GSMMs have emerged as possible solutions to decipher the metabolic reprograming underlying tumour progression and systematically identify vulnerabilities that can be exploited in therapy [2]. We have previously identified TKTL-1 as a key player in cancer cell proliferation and metabolic reprograming in Leukemia [3]. The goal of this work is to generate specific genome scale reconstructions of Acute Myeloid Leukemia (AML) and Chronic Myeloid Leukemia (CML) cell lines and identify synthetic lethal pairs containing TKTL-1. The most promising candidates will be validated in vitro using gene silencing.

References:

1. Tarrado-Castellarnau, M., P. De Atauri, and M. Cascante, *Oncogenic regulation of tumor metabolic reprogramming*. Oncotarget, 2016. **7**(38): p. 62726-53.
2. Marin de Mas, I., et al., *Cancer cell metabolism as new targets for novel designed therapies*. Future Medicinal Chemistry, 2014. **6**(16): p. 1791-810.
3. Diaz-Moralli, S., et al., *A key role for transketolase-like 1 in tumor metabolic reprogramming*. Oncotarget, 2016.

EXOME SEQUENCING DATA ANALYSIS TO CHARACTERIZE RARE GERMLINE AND SOMATIC COPY NUMBER VARIANTS INVOLVED IN COLORECTAL CANCER

Sebastià Franch-Expósito[1], Clara Esteban-Jurado[1], Pilar Garre[2], Isabel Quintanilla[1], Saray Duran[1], Miriam Cuatrecasas[3], Esther Samper[1], Jenifer Muñoz[1], Marcos Díaz-Gay[1], Teresa Ocaña[1], Sabela Carballal[1], María López-Cerón[1], Antoni Castells[1], EPICOLON consortium, Maria Vila-Casadesús[4], Enric Serra[5], Sergi Beltran[5], Jaime Carvajal[6], Luis Bujanda[7], Clara Ruiz-Ponte[8], Jordi Camps[1], Meritxell Gironella[1], Juan José Lozano[4], Francesc Balaguer[1], Joaquín Cubiella[9], Trinidad Caldés[2], Sergi Castellví-Bel[1]

1. Gastroenterology, Hospital Clínic, IDIBAPS, CIBERehd, Barcelona, Spain
2. Molecular Oncology Laboratory, Hospital Clínico San Carlos, Madrid, Spain
3. Pathology, Hospital Clínic, Barcelona, Spain
4. Bioinformatics, CIBERehd, Barcelona, Spain
5. CNAG-CRG, Barcelona, Spain
6. Centro Andaluz de Biología del Desarrollo, Sevilla, Spain
7. Gastroenterology, Hospital Donostia, CIBERehd, San Sebastián, Spain
8. FPGMX, CIBERER, Santiago de Compostela, Spain
9. Gastroenterology, Complexo Hospitalario Universitario de Ourense, Ourense, Spain

Presenter e-mail: sebasfranch@gmail.com

Colorectal cancer (CRC) represents one of the most common cancers worldwide when taking into account both genders together. Availability of exome next generation sequencing data has increased tremendously and, besides point variants, copy number variants (CNVs) can be inferred using calling algorithms. CNV are common somatic events in cancer but they can be also the germline mutational event involved in CRC predisposition.

Methods: We analyzed germline DNA whole exome sequencing (WES) data from 38 families with strong CRC aggregation without alterations in known hereditary genes. In 5 index cases, paraffin embedded tumor DNA WES data was also studied. Rare germline CNVs were called using ExomeDepth and CoNIFER, while tumoral CNVs were inferred from data using CNVkit and VarScan2. Only CNVs detected by each pair of tools were considered. Germline variants shared between family members were compared to Database of Genomic Variants catalog and our in-house database, whereas somatic CNV were assessed with the COSMIC database. For one germline CNV, validation and segregation were performed by comparative genome hybridization, and gene expression and protein studies were also conducted. OncoScan array data was available for comparison for one CRC tumor sample.

Results: Twenty-one germline rare CNVs (16 duplications and 5 deletions) were detected. We identified 419 somatic variants (363 duplications and 56 deletions). Regarding germline variants, a duplication in chromosome 1 stood out as interesting and was validated. This duplication included TTF2, TRIM45, VTCN1 and miR942. Gene expression and protein studies pointed out to an overexpression of TTF2 and miR942 and a downregulation of TMEM158, previously involved in colorectal carcinogenesis.

Conclusions: WES data can be used as a first approach to identify CNVs in germline and somatic data. Chromosome 1 duplication may correspond to the mutational event involved in CRC predisposition in the carrier family.

POPFLY: THE DROSOPHILA POPULATION GENOMICS BROWSER

Sergi Hervás, Esteve Sanz, Sònia Casillas, Antonio Barbadilla.

Universitat Autònoma de Barcelona, Institut de Biotecnologia i Biomedicina / Departament de Genètica i Microbiologia, 08193 Bellaterra (Barcelona), Spain
Presenter e-mail: shervas91@gmail.com

The recent compilation of over 1000 worldwide wild-derived *Drosophila melanogaster* genome sequences reassembled using a standardized pipeline provides a unique resource for population genomic studies (Drosophila Genome Nexus, DGN). Thus, a visual display of the metrics describing the genome-wide variation and selection patterns would allow gaining a global view and understanding of the evolutionary forces shaping genome variation. Here we present PopFly, a population genomics oriented genome browser, based on JBrowse software, that contains a complete inventory of population genomic parameters estimated from the DGN data. This browser is designed for the automatic analysis and display of genetic variation data within and between populations along the *Drosophila melanogaster* genome. PopFly allows the visualization and retrieval of functional annotations, nucleotide diversity metrics, linkage disequilibrium statistics, recombination rate estimates and a battery of neutrality tests at different window sizes through the euchromatic chromosomes. PopFly contains the broadest catalogue of population genomic estimates in *Drosophila melanogaster* developed so far. It stores more than 3500 precomputed tracks with information about 30 populations from 17 countries and 5 continents, which can be filtered based on different features. The automated data processing pipeline, starting from raw sequence data and ending up at the browser visual elements, allows the incorporation of the growing number of population genomic sequences from this and other *Drosophila* species (http://flybook-mpg.uab.cat/flybook-mpg/). Overall, PopFly aims to become the reference *Drosophila* population genomics browser.
PopFly is open and freely available at site http://popfly.uab.cat.

Poster 24

HIGH RESOLUTION MAPPING IN DIFFERENT OCTOPLOID STRAWBERRY POPULATIONS FOR MQTLS DETECTION.

Jose Manuel Hidalgo López[1], Pol Rey Serra[1] and Amparo Monfort[1]

1. IRTA, Center for Research in Agricultural Genomics CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Bellaterra (Cerdanyola del vallés),08193 Barcelona, Spain
Presenter e-mail: josemanuel.hidalgo@irta.cat

The cultivated strawberry, Fragaria x ananassa, is the most important berry specie of the Rosaceae family, which includes species like peach, apple, pear, prune and berries. Its octoploid nature (2n = 8x = 56) presents a challenge to the development of molecular breeding tools, although its closely related to diploid Fragaria species (2n = 2x = 14) provides great knowledge and advanced genomic tools. After sequencing the diploid genome , the Axiom IStraw90™ SNP array (90.000 SNPs) was developed. This array was used to genotype a cultivated F2 population (135 individuals) that segregates for polyphenol compounds content (nutritional profile interest). Recently, a low-coverage resequencing was performed in order to genotype more than 8.000 SNPs on the same F2 population. Originally, we developed a genetic map using the IStraw90™ SNP array with more than 13.000 segregating markers. Lately, this map was improved with last SNPs genotyped. In addition, another two populations were genotyped using the same resequencing method, a F1 (67 individuals) and a F2 (54 individuals) segregating for volatile compounds (aroma). Using the same SNPs for different populations will allow us to improve our genetic maps, obtain stronger QTLs and improve knowledge on octoploid strawberry genome. Nutritional profiling has been done using HPLC-MS and aroma profiling is currently being done using GC-IMS.

CONTEXT-SPECIFIC REGULATORY VARIANT ANNOTATION OF LARGE-SCALE CANCER GENOME SEQUENCING INITIATIVES

Aliaksei Holik, Shalu Jhanwar, Stephan Ossowski

Bioinformatics and Genomics Programme. Centre for Genomic Regulation (CRG), the Barcelona Institute of Science and Technology (BIST). Dr. Aiguader, 88. 08003 Barcelona, Spain
Presenter e-mail: Aliaksei.Holik@crg.eu

Advances in genomics technologies have led to an emergence of large-scale cancer genome sequencing initiatives, such as TCGA and ICGC. These studies to date have mainly focused on understanding the role of mutations in the coding regions, primarily due to high interpretability of coding mutations. While this approach has provided some valuable insights into the molecular mechanisms of carcinogenesis, it largely ignores a vast number of potentially functional mutations in regulatory regions, such as promoters and enhancers. A number of recent studies have proposed resources and methods for estimating the effect of non-coding regulatory variants, broadly based on epigenetic marks and evolutionalry conservation (e.g RegulomeDB, CADD, GWAVA, etc). A common drawback of these methods is the lack of tissue-specificity in relation to the potential role of the variants analysed, which limits the extent to which the variant annotation can be generalised to different cancer types. In order to overcome this limitation we develop approaches that enable us to estimate the damage potential of the regulatory variants in the context of a specific cancer by integrating epigenome information from the relavant tissue types. We take advantage of the large-scale sequencing projects generating context-specific epigenome and transcriptome data (e.g. ENCODE and Roadmap Epigenomics) and apply this information in a tissue-specific manner to the 2700 whole cancer genomes spanning 22 cancer types from the Pan-Cancer Analysis of Whole Genomes (PCAWG) project. First, we use tissue-specific epigenome information to train a support vector machine based classifier to identify genomic regions consistent with promoter and enhancer activity. In order to estimate the damage potential of the variants overlapping these regions we further train a random forest based classifier using a set of commonly exploited features (histone marks, DNAse hyper-sensitivity, evolutionary conservation, etc), as well as pre-computed scores from existing methods (e.g. CADD). Additionally, we compute features based on the variants' potential to affect binding of the common transcription factors according to the motif position-frequency matrix. We train this classifier using tissue-specific expression quantitative trait loci (eQTLs) from GTEx project, as a positive training set. Finally, we apply the classifiers trained on the tissue-specific training sets to the somatic and germline variants from the relevant cancer types from the PCAWG project.

MECHANISMS OF ACTION OF SACUBITRIL/VALSARTAN ON CARDIAC REMODELING: A SYSTEMS BIOLOGY APPROACH

Oriol Iborra-Egea[1], Carolina Gálvez-Montón[1,2], Santiago Roura[1,2,3], Isaac Perea-Gil[1], Cristina Prat-Vidal[1,2] Carolina Soler-Botija[1,2], Antoni Bayes-Genis[1,2,4,5]

1. ICREC Research Program, Instituto de Investigación Germans Trias i Pujol, Badalona, Spain.
2. CIBERCV, Instituto de Salud Carlos III, Madrid, Spain.
3. Center of Regenerative Medicine in Barcelona, Barcelona, Spain.
4. Servicio de Cardiología, Hospital Universitario Germans Trias i Pujol, Badalona, Spain.
5. Departmento de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain.
Presenter e-mail: oriol.iborra@gmail.com

Sacubitril/valsartan, proved superiority over other conventional heart failure (HF) management treatments, but its mechanisms of action remains obscure. In this study, we sought to explore the mechanistic details for sacubitril/valsartan in HF and post-myocardial infarction (MI) remodeling, using an *in silico*, systems biology approach. Myocardial transcriptome obtained in response to MI in swine was analyzed to address post-infarction ventricular remodeling. Swine transcriptome hits were mapped to their human equivalents using Reciprocal Best (blast) Hits, Gene Name Correspondence, and InParanoid database. HF remodeling was studied using public data available in GEO (accession GSE57345, subseries GSE57338), processed using the GEO2R tool. Using the Therapeutic Performance Mapping System (*TPMS*) technology (Anaxomics Biotech SL.), dedicated mathematical models trained to fit a set of molecular criteria were generated. All relationships incorporated into the biological network were drawn from public resources (including KEGG, REACTOME, INTACT, BIOGRID, and MINT). An artificial neural network analysis revealed that sacubitril/valsartan acts synergistically against cardiomyocyte cell death and left ventricular extracellular matrix remodeling via 8 principal synergistic nodes. When studying each pathway independently, valsartan was found to improve cardiac remodeling by inhibiting members of the guanine nucleotide-binding protein family, while sacubitril attenuated cardiomyocyte cell death, hypertrophy, and impaired myocyte contractility by inhibiting PTEN. The complex molecular mechanisms of action of sacubitril/valsartan post-MI and HF were delineated using a systems biology approach. Further, this dataset provides pathophysiological rationale for the use of sacubitril/valsartan to prevent post-infarct remodeling.

PRIONW: A SERVER FOR THE PREDICTION OF PRION-LIKE DOMAINS AND THEIR AMYLOID CORES

Valentin Iglesias[1], Rafael Zambrano[1], Oscar Conchillo-Sole[1], Frederic Rousseau[2], Joost Schymkowitz[2], Raimon Sabate[3], Xavier Daura[4] and Salvador Ventura[1]

1. Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain
2. VIB Switch Laboratory and Department for Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium.
3. Institut de Nanociència i Nanotecnologia (INUB) and Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain.
4. Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
Presenter e-mail: valentin.iglesias.mas@gmail.com

Prions are a particular type of amyloids with the ability to self-perpetuate and propagate *in vivo*. Prion-like conversion underlies crucial biological processes in a growing number of species but is also connected to several human diseases including Creutzfeldt–Jakob disease and kuru. Yeast prions are the best understood transmissible amyloids. In these proteins, prion formation from an initially soluble state involves a structural conversion, often driven, by specific domains enriched in glutamine/asparagine (Q/N) residues. Importantly, domains sharing this compositional bias are also present in the proteomes of higher organisms, thus suggesting that prion-like conversion might be an evolutionary conserved mechanism. We have recently shown that the identification and evaluation of the potency of amyloid nucleating sequences in putative prion domains allows discrimination of *bona fide* prions. PrionW is a web application that exploits this principle to scan sequences in order to identify proteins containing Q/N enriched prion-like domains (PrLDs) in large datasets. A scan of the complete yeast proteome with PrionW identifies previously experimentally validated prions with a high success rate: a sensitivity of 0.917, a specificity of 0.949, a precision rate of 0.846, an accuracy of 0.941 and a false discovery rate of only 0.154. This website is free and open to all users who can analyze up to 10000 sequences at a time, PrLD-containing proteins are identified and their putative PrLDs and amyloid nucleating cores visualized and scored. The output files can be downloaded for further analysis. PrionW server can be accessed at http://bioinf.uab.cat/prionw/.

RATIONALIZING DRUG RESPONSE IN CANCER CELL LINES

Teresa Juan-Blanco[1], Patrick Aloy[1,2]

1. Institute for Research in Biomedicine. Joint IRB-BSC-CRG Program in Computational Biology
2. Institució Catalana Recerca i Estudis Avançats (ICREA)
Presenter e-mail: teresa.juanblanco@irbbarcelona.org

Cancer cell lines (CCLs) play an important role in the initial stages of drug discovery allowing, among others, for the high-throughput screening of drug candidates. However, their clinical relevance remains controversial. Among other limitations, CCLs often do not represent primary tumors which may result in the loss of tissue-specific drug response (1, 2). Many polymorphisms in genes encoding drug-metabolizing enzymes, transporters and drug targets, as well as disease-related genes have been linked to altered drug sensitivity. Yet, identifying the correlation between this variability and pharmacological responses remains challenging (3). One of the main complications in understanding drug responses is that differential pharmacological profiles are caused by the interplay of multiple genes. For this reason, system biology approaches that integrate genomic data with protein-protein interaction (PPI) networks may be more efficient to connect genetic variations to drug effects. Here, we propose a system biology method to identify the mechanisms that may affect drug sensitivity in CCLs. Specifically, we exploited somatic mutations, gene expression and drug response data provided by the Cancer Cell Line Encyclopedia (CCLE). We integrated these molecular profiles with PPI data to detect groups of CCLs with similarly perturbed network regions and that present a similar drug response. We can use this data to identify molecular biomarkers responsible for the differential drug response and propose complex gene expression patterns that may predict sensitivity in CCLs.

References:

1.Wilding, J.L. and Bodmer, W.F. (2014) Cancer cell lines for drug discovery and development. Cancer Research 74:2377-84.
2. Jaeger, S., Duran-Frigola, M. & Aloy, P (2015) Drug sensitivity in cancer cell lines is not tissue-specific. Molecular Cancer 14:40.
3. Evans, W.E. and Relling M.V. (2004) Moving towards individualized medicine with pharmacogenomics. Nature 429(6990), 464-8.

COMPARATIVE GENOMIC ANALYSIS OF THE *OLEA EUROPAEA* COMPLEX

Irene Julca[1,2,3], Marina Marcet-Houben[1,2], Damian Loska[1,2], Fernando Cruz[2,4], Jèssica Gómez-Garrido[2,4], Emilio Cano[5], Beatriz Galán[6], Leonor Frias[2,4], Paolo Ribeca[2,4], Marta Gut[2,4], Manuel Sánchez-Fernández[7], Jose Luis García[6], Ivo G. Gut[2,4], Pablo Vargas[5,*], Tyler S. Alioto[2,4], and Toni Gabaldón[1,2,8]

1. Bioinformatics and Genomics Department, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain
2. Universitat Pompeu Fabra (UPF). 08003 Barcelona, Spain.
3. Universitat Autònoma de Barcelona
4. CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain
5. Royal Botanical Garden of Madrid. Consejo Superior de Investigaciones Científicas (CSIC). Madrid, Spain
6. Centro de Investigaciones Biológicas (CIB). Consejo Superior de Investigaciones Científicas (CSIC). Madrid, Spain
7. Dirección de Paisajismo. Área Corporativa de Inmuebles. Grupo Santander. Boadilla del Monte, Madrid (Spain).
8. Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

Presenter e-mail: irene.julca@crg.eu

Olea europaea L. (olive) is one of the oldest cultivated fruit trees in the Mediterranean basin. It belongs to the *Oleaceae* family and is composed of six subspecies. The subspecies *europaea* has two varieties: the wild form (var. *sylvestris*) and the cultivars (var. *europaea*). To gain insight into the biology and evolution of this emblematic tree, we sequenced, assembled and annotated a reference genome obtained from a single individual (*Olea europaea* L. subsp. *europaea* var. *europaea* cv. 'Farga'). We also re-sequenced additional subspecies (*laperrinei*, *cuspidata*, *guanchica*, *maroccana* and *cerasiformis*), one var. *sylvestris* (*O. europea* subsp. *europaea* var. *sylvestris*) and four var. *europaea* ('Arbequina', 'Beladi', 'Picual', 'Sorani'). The final reference genome has a length of 1.31 Gb and 56,349 coding genes. We also assembled the chloroplast and mitochondrial genome of 'Farga'. Using all these genomes we analyzed genome variation through SNP calling. At a genomic level the cultivars had fewer homozygous SNPs than the wild form, and the two polyploid subspecies (*maroccana* and *cerasiformis*) had more heterozygous SNPs than the other subspecies. However, at the chloroplastic and mitochondrial level, the wild type had fewer SNPs than the cultivars. We used the prediction of genomic and chloroplast SNPs to reconstruct phylogenetic relationships between the *O. europaea* complex and used the subsp. *cuspidata* as outgroup. The genomic tree showed that all cultivars group together, but the var. *sylvestris* does not cluster with them, as expected. However the chloroplast tree showed that the subsp. *europaea* is divided in two groups, one composed by the other cultivars, and the other by 'Farga' and the var. *sylvestris*. In summary, the subspecies *europaea* is not monophyletic and the evolutionary signal of 'Farga' may differ from that of the other cultivars. Further analysis is needed to fully clarify the origins of 'Farga'.

PATHWAY-CENTERED ANALYSIS OF PIG DOMESTICATION

J. Leno-Colorado[1], N.J. Hudson[2], A. Reverter[3], M. Pérez-Enciso[1,4]

1. Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, 08193 Bellaterra, Spain
2. School of Agriculture and Food Science, University of Queensland, Queensland 4343, Australia
3. CSIRO Agriculture and Food, Queensland Bioscience Precinct, 306 Carmody Rd., St. Lucia, Brisbane, Queensland 4067, Australia
4. ICREA, Carrer de Lluís Companys 23, Barcelona 08010, Spain
Presenter e-mail: jordi.leno@cragenomica.es

The pig (*Sus scrofa*) is an excellent model to study domestication, which occurred independently in East Asia and Europe ca. 9,000 years ago. Due to the wide distribution of the wild ancestor, the wild boar, this species is an excellent model to study domestication. Analyzing genome variability patterns in terms of metabolic pathways is attractive since it considers the interrelated functions of genes, in contrast to genome-wide scans that treat genes or genome-windows in isolation. To that end, here we studied pig domestication using 40 wild boars and 123 domestic pig genomes from Asia and Europe. We computed statistical significance for differentiation (Fst) and linkage disequilibrium (nSL) statistics at the pathway level. In terms of Fst, we found 21 and 12 pathways significantly differentiated at a q-value < 0.05 in Asia and Europe, respectively; five were shared across continents. In Asia, we found six significant pathways related to behavior, which involved essential neurotransmitters like dopamine and serotonin. Several of the significant pathways were interrelated, and shared a variable percentage of genes. There were 12 genes present in more than 10 significant pathways (in terms of Fst); comprising genes involved in the transduction of a large number of signals, like phospholipase PCLB1, which is expressed in the brain, or ITPR3, which has an important role in taste transduction. In terms of nSL, significant pathways were mainly related to reproductive performance (ovarian steroidogenesis), an important target trait as well during domestication and modern animal breeding. Different levels of recombination cannot explain these results, since there was no correlation between Fst and recombination rate. We conclude that pathway analyses facilitate the biological interpretation of genomic studies. In the case of pig domestication, behavior played an important role, among other physiological and developmental processes.

QUANTITATIVE ASSESSMENT OF ESSENTIALITY IN A MINIMAL BACTERIUM

Samuel Miravet-Verde[1,2] , Maria Lluch-Senar[1,2] and Luis Serrano[1,2,3]

1. EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain,
2. Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain and
3. Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluis Companys 23, 08010 Barcelona, Spain
Presenter e-mail: samuel.miravet@crg.eu

Identifying the minimal number of genes required to sustain life is one of the major challenges in synthetic biology. Thus, classifying genomic regions as: essential (E; non-disruptible) and non-essential (NE; disruptible); is the first step for doing rational engineering of genomes. Recently, by combining high-throughput insertion tracking by deep sequencing (HITS), a new category of essentiality, fitness (F), has been found. This category comprises genomic regions that cannot be classified as E or NE, suggesting that their essentiality depends on the condition. However, the study of essentiality in different conditions is difficult. A statistical assessment of essentiality that allows the assignment of a quantitative value would facilitate their characterization. Herein, we describe the first quantitative approach used to study the genome essentiality in the minimal bacterium *Mycoplasma pneumoniae*. The implemented pipeline can be applied to other genomes and we envision that it will facilitate the characterization of F genes.

EVALUATION OF SPLICING PREDICTION TOOLS BY *IN VITRO* ANALYSIS IN BREAST/OVARIAN CANCER PREDISPOSITION GENES

Alejandro Moles[1,2], Laura Duran-Lozano[1,2], Gemma Montalban[1,2], Sandra Bonache[1], Irene Esteban[3], Neus Gadea[3,4], Estela Carrasco[3], Judith Balmaña[3,4], Sara Gutiérrez-Enríquez[1], Orland Diez[1,5]

1. Oncogenetics Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain
2. Universitat de Barcelona (UB), Barcelona, Spain
3. High Risk and Cancer Prevention Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain
4. Medical Oncology Department, Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Spain
5. Àrea de Genètica Clínica i Molecular, Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Spain
Presenters e-mail: amoles@vhio.net, lduran@vhio.net

A significant number of genetic variants identified for clinical diagnosis using massive sequencing have a potential splicing effect giving the *in silico* predictors a key role to assess the clinical significance of each variant. Our aim was to evaluate the performance of a set of commonly used splicing *in silico* tools compared against RNA *in vitro* results. This was done for natural splice sites of clinically relevant genes in hereditary breast/ovarian cancer (HBOC). *BRCA1* and *BRCA2* potentially spliceogenic variants were selected from HBOC patients analysed routinely for diagnostic purposes. We included variants in *ATM, BRIP1, CDH1, CHEK2, NF1, PALB2, PTEN, RAD51C, RAD51D, STK11* and *TP53*, obtained in a research series of *BRCA1/2* negative HBOC patients. Following an in-house pipeline based on quality control and minor allele frequencies (≤ 0.005), we identified 59 variants localized within the 10 exonic and 20 intronic nucleotides adjacent to exon-intron boundaries. These variants were analysed *in silico* by: SPANR, SpliceSiteFinder-like (SSF-like), MaxEntScan (MES), Splice Site Prediction by Neural Network (NNPLICE), GeneSplicer (GS) and Human Splicing Finder (HSF). For each predictor we adopted thresholds pre-established in literature. For the analysis *in vitro* RNA was isolated from variant carriers and controls, and splicing products were characterized by RT-PCR assays and Sanger sequencing. To date, we have analysed 44 variants, and most of the *in silico* tools predicted correctly their effect on splicing, being the most accurate SSF (86,36% of cases), HSF (84,09%) and SPANR (82,93%). Data analysed with the six *in silico* tools suggest MaxEnt as the most sensitive predictor and SPANR as the most specific. *In silico* tools are useful to predict splicing alterations. Combinations of several of them will allow the detection of higher number of variants with splicing effects.

KARYON, AN AUTOMATED PIPELINE FOR THE STUDY OF UNORTHODOX GENOMES

Miguel Ángel Naranjo-Ortiz, Toni Gabaldón

Center for Genomic Regulation, Barcelona
Presenter e-mail: miguelangelnaranjoortiz@gmail.com

Due to technical and practical limitations, genome sequencing projects have traditionally restricted themselves to organisms with small and highly homozygous genomes, creating a false impression of genomic homogeneity in the living world. However, with the exponential drop of sequencing prizes, genome sequencing has arrived to a myriad of new organisms and has exposed the field to a new array of genomic oddities. Polyploids, aneuploids, hybrid and chimeric genomes are known for many eukaryotic organisms and may in fact be much more common than previously thought. Here we present our later advancements in an automated pipeline that aims to help researchers to assembly problematic genomes and understand their structure.

COMPARATIVE TRANSCRIPTOMICS OF IMMUNE CELL REPROGRAMMING IN HUMAN AND MOUSE

Ramil Nurtdinov[1,2], Alexandre Esteban[1,2], Amaya Abad[1,2], Maria Sanz[1,2], Marina Ruiz[1,2], Dmitri Pervouchine[1,2], Sebastian Ullrich[1,2], Cecilia Klein[1,2], Alessandra Breschi[1,2], Silvia Perez[1,2], Rory Johnson[1,2], Roderic Guigo[1,2]

1. Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Bioinformatics and Genomics, Barcelona, Spain
2. Universitat Pompeu Fabra (UPF), Bioinformatics, Barcelona, Spain
Presenter e-mail: ramil.nurtdinov@crg.eu

Transcriptomes reflect cellular and organismic phenotypes. Therefore transcriptome comparisons across species uncover the molecular basis of both conserved and species-specific phenotypes. Since the laboratory mouse is top choice organism for human biology, transcriptome comparisons across multiple tissues between human and mouse have been extensively carried out. However, because phenotypic differences between human and mouse tissue are difficult to quantify, it is difficult to relate transcriptional differences to phenotyipic differences. Here we have compared human and mouse transcriptional patterns in a very well controlled time course cellular differentiation process for which the time of differentiation is quantifiable phenotype that can be precisely defined in human and mouse. Specifically, we employed the BlaER cell line, a powerful *in vitro* model of haematopoietic transdifferentiation. These engineered lymphoblastic leukaemia B-cells can be reproducibly and efficiently converted to a macrophage-like identity by the activation of a single CEBPα transgene. This process is accompanied by loss of the B-cell-specific surface marker CD19, and a concomitant gain of the Mac-1 marker, allowing cellular phenotype to be monitored and quantified by flow cytometry. The process takes seven days in human, but only three in mouse. We collected and sequenced cytoplasmic RNA at several consecutive time-points (12 in human and 10 in mouse). We aligned the human and mouse time points according to transcriptional similarity, and identified both the conserved and the species-specific transcriptional programs that participated in the process. Thus we identified 2689 orthologous protein coding gene with twofold expression change and concordant profiles during the transdiferentiation process, as well as 2149 variable ones that behaved differently in human and mouse. We also identified splicing events with conserved human-mouse patterns. Based on the species specific transcriptional pattern, we build models to predict transdifferentiation time depending on gene expression, and we are currently experimentally validating these models.

SCREENING OF NOVEL ENZYMES FROM AN ISOLATED STRAIN FROM SOIL

Pau Soteras[1], Susana Valenzuela[1,2], F.I. Javier Pastor[1,2]

1. Department of Genetics, Microbiology and Statistics, University of Barcelona.
2. IN2UB. Institute of Nanoscience and Nanotechnology, University of Barcelona.
Presenter e-mail: fpastor@ub.edu

The strain BP-23 isolated from soil presents a high depolymerizing activity of vegetal biomass due to its capability to degrade polysaccharides. One of the most relevant activity of the strain is xylanase. This activity allows the strain to grow in minimum medium with xylan, the second most abundant polysaccharide in the earth, as a unique source of carbon. With the aim of characterize all the xylanases involved in the xylanolytic system of the strain and to identify all the proteins involved in the degradation of this substrate, the sequencing of the genome of the strain has been performed. As a first stage to identify new enzymes, a multiple alignment of sequences encoding known enzymes with the genome of the strain allowed the identification of several DNA segments showing homology to known genes. DNA segments with the best score were chosen and analysed their upstream and downstream sequences to identify putative *orfs* in the genome of strain BP-23. As a third stage, a BLASTx focusing the search on microorganisms taxonomically related to the studied strain was performed, in order to select those *orfs* with homology to genes encoding relevant enzymes. Then, specific primers were designed to amplify these *orfs* by PCR, which were cloned in expression vectors and transformed in *E. coli*. The recombinant strains obtained were cultured to produce the new enzymes, which were purified and characterized. Multiple new *orfs* for xylanases of the families GH10, GH30 and GH8, expansins of the family GH63 and for several other new enzymes (xylosidases and arabinofuranosidases) of the family GH43 have been identified. Experimental evidence has confirmed that most of this *orfs* codify for active proteins. Among them, several new xylanases, expansins, celulases, esterases and pectatolyases from the strain have been characterized by the research group. It doesn't seem to be a unique fact that different enzymes of different families act in the degradation of a single polysaccharide source.

FINDING PATHWAY-SPECIFIC RESPONSES IN GENE EXPRESSION FAMILY STUDIES THROUGH VARIANCE COMPONENT ANALYSIS AND APPLICATION TO GAIT2 THROMBOPHILIA PROJECT

Sergio Picart-Armada[1], Andrey Ziyatdinov[2], Angel Martinez-Perez[2], Ana Viñuela[3], Juan Carlos Souto[2], Emmanouil Dermitzakis[3], Alexandre Perera[1], José Manuel Soria[2], Alfonso Buil[4]

1. Universitat Politècnica de Catalunya, Barcelona, Spain.
2. Institut d'Investigació Biomèdica, Sant Pau, Barcelona, Spain.
3. Department of Genetics Medicine and Development, University of Geneva, Geneva, Switzerland.
4. Mental Health Center Sct. Hans, Roskilde, Denmark.
Presenter e-mail: sergi.picart@upc.edu

Family studies offer an opportunity to assess the inherited and acquired nature of the quantified traits. Here, we analyse phenotypes related to thrombophilia by partitioning the explained variance into a genetic component, a transcriptomic component and the residual effect. Our data corresponds to the GAIT2 project on idiopathic thrombophilia, whose cohort includes 935 individuals stemming from 35 pedigrees. 481 quantitative phenotypes were measured at time of recruitment, including anthropometric measurements, hemogram, hemostasis traits, as well as phenotypes related with platelet activity, homocysteine metabolism, inflammation and flow cytometry. We also sequenced the mRNA transcriptome from whole blood. We perform variance component analysis on linear mixed effect models that explain the phenotypes in terms of the kinship structure and the gene expression. Additionally, we extend these models conceiving a pathway enrichment technique that tests whether including the gene expression similarity based on a particular pathway improves the explained variance in the phenotype. On one hand, models involving kinship and expression recover previously reported heritability estimates for known phenotypes. Furthermore, their error term decreases by introducing expression in the models, therefore explaining former residual variance in terms of ambient factors. On the other hand, relevant pathways improve specific phenotype models, such as platelet activation explaining platelet abundance in blood.

Funding:
Ministerio de Economía y Competitividad TEC2014-60337-R. Grup de Recerca Consolidats de la Generalitat de Catalunya 2014-SGR1063. CIBER-BBN is an initiative of the Spanish ISCIII. Local grant FI-AGAUR 2015.

Poster 37

GENOME INSTABILITY AND GENETIC DIVERSITY - HOW TO ADAPT IN THE ABSENCE OF SEX IN THE PROTOZOAN PATHOGEN LEISHMANIA DONOVANI

P. Prieto Barja[1,6], P. Pescher[2], G. Bussotti[3], F. Dumetz[4], H. Imamura[4], D. Kedra[1], V. Chaumeau[5], H. Himmelbauer[1,7], M. Pages[5], Y. Sterkers[5], J.C. Dujardin[4], C. Notredame[1,6], and G. F. Späth[2]

1. Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain
2. Institut Pasteur, INSERM U1201, Unité de Parasitologie moléculaire et Signalisation, Paris, France
3. HUB de Bioinformatique et Biostatistiques, Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), Institut Pasteur, Paris, France
4. Institute of Tropical Medicine, Molecular Parasitology Unit, Antwerpen, Belgium
5. University Montpellier, Faculty of Medicine, Laboratory of Parasitology-Mycology, CNRS 5290, IRD 224, University Montpellier (UMR "MiVEGEC") and Centre Hospitalier Universitaire, Montpellier, France
6. Universitat Pompeu Fabra (UPF), Barcelona, Spain; 7Institute of Biotechnology Muthgasse 18 1190 Vienna
Presenter e-mail: pablo.prieto.barja@crg.eu

Leishmania donovani causes visceral leishmaniasis, a fatal disease when left untreated. The process through which the parasite adapts to environmental change remains largely unknown. Here we show that aneuploidy is an integral part of parasites adaptation and that karyotypic fluctuations allow for selection of beneficial haplotypes, with important impact on parasites transcriptomic output that correlates with phenotypic variations including, proliferation and infectivity. To avoid loss of diversity resulting from karyotype and haplotype selection, L. donovani takes advantage of two mechanisms: (i) polyclonal selection of beneficial haplotypes resulting in co-existing subpopulations that preserve the original diversity, and (ii) generation of new diversity as a result of higher mutation rates tolerated by aneuploidy-prone chromosomes. Our results uncover high aneuploidy turnover and haplotype selection as a new mechanism of L. donovani evolutionary adaptation that preserves genetic diversity under strong selection. This process may be of broad significance to other human diseases, including fungal infection and cancer.

Poster 38

MADE OF GENES: A GENOMIC DATA MANAGEMENT FRAMEWORK FOR PERSONALIZED MEDICINE.

Miquel Ràmia-Jesús[1], Marc Sitges Puy[1], Leonor Frias Moya[1], Oscar Flores Guri[1]

1. Made of Genes, C/Virgen de Guadalupe 18, 08950 Esplugues de Llobregat, Catalonia, Spain
Presenter e-mail: we.are@madeofgenes.com

We're entering a new era: genome democratization. The cost of sequencing a single human genome has dropped from billions to thousands. Genome analysis is migrating from hypotheticals and research into healthcare institutions, where professionals are revolutionizing the way we treat disease: personalized medicine. One of the main challenges of this change of paradigm in medicine is transferring knowledge and results from remote labs to the common public. Many research and healthcare institutions are unprepared to handle the bureaucratic and computational demands of analyzing, storing and sharing genomic data. Also, when physicians finally approach the fast-growing, but young genetic testing market, they often lack familiarity with the tests's availability, clinical validity and results interpretation. Occasionally, they receive the results of genome tests that patients have accessed without appropriate supervision. Health benefits are difficult for professionals to identify, and patients develop significant concerns about how much of their genomic data really remains private and protected. 'Made of Genes' believes that the best way to improve our interactions with genomics is to connect all the players, while empowering the genome data owner, via an accessible, high-performance compliant computing platform. A genome only needs to be sequenced once. It is processed, encrypted and stored in our servers, available to be shared with researchers and test-providers at the owner's discretion and convenience. All genome accesses are overseen by a novel electronic informed consent system based on the bitcoin blockchain. And always under continuous supervision by a healthcare professional, who facilitates the test-provider's entire data interpretation journey. Made of Genes is a place where researchers share their knowledge, physicians apply it, and everyone accesses the best healthcare.

THE ROLE OF TRANSPOSABLE ELEMENTS IN ALTERNATIVE SPLICING REGULATION DURING EARLY DEVELOPMENT

Quirze Rovira[1,2], Chris D.R. Wyatt[1,2], Barbara Pernaute[1,2] and Manuel Irimia[1,2]

1. Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain
2. Universitat Pompeu Fabra (UPF), Barcelona, Spain
Presenter e-mail: quirze92@gmail.com

Through the process of alternative splicing (AS) eukaryotes can increase the diversity of their transcriptome and proteome by differentially processing the introns and exons of the pre-mRNA from a single gene. This mechanism is highly relevant in vertebrates where it affects around 95% of human multiexon genes. This plasticity of the transcriptome can be detected during the early stages of development, where many important cell fate decisions (e.g. trophoectoderm and inner cell mass differentiation) take place. At the very beginning, the cell processes are driven by the maternal mRNAs and proteins until embryo genome activation (EGA) and maternal degradation remodel the transcriptome. An intriguing feature of EGA is the activation of a large fraction of transposable elements (TEs). TEs are active components of the genome and constitute a significant fraction of the vertebrates' genomes. In addition to contributing to gene expression, TEs can be co-opted as new alternative exons or AS regulatory elements.
In this study we investigated the role that TEs have in the regulation of AS during early development in mammals (with data from human, mouse and cow), using RNA-seq data from the oocyte to blastocyst developmental stages. We used vast-tools to identify and quantify the inclusion of cassette exon events. Exons that share a similar pattern during development were clustered together. Some clusters showed a significant enrichment or depletion for TE-containing exons. Interestingly, the clusters depleted in TE-containing exons show a clear drop in exon inclusion during stages close to EGA while the enriched cluster show up-regulation of exon inclusion at the same stages. Moreover, exons from enriched and depleted clusters show high percentage of disrupting the protein open reading frame by opposite strategies: inclusion or exclusion of the exon. Finally, the clusters of study are enriched in gene ontology terms related to nucleus cellular components and biological processes in all three species.

INTERPRETING CANCER GENOMIC ALTERATIONS OF SINGLE PATIENTS

David Tamborero[1], Carlota Rubio-Perez[1], Jordi Deu-Pons[1], Michael Schroeder[1,2], Ana Vivancos[3], Ana Rovira[4], Ignasi Tusquets[4,5], Joan Albanell[4,6], Jordi Rodon[3], Josep Tabernero[3], Carmen de Torres[7], Rodrigo Dienstmann[3], Abel Gonzalez-Perez[1] and Nuria Lopez-Bigas[1,8]

1. Biomedical Genomics Lab, Research Program on Biomedical Informatics (GRIB), IMIM Hospital del Mar Medical Research Institute and Pompeu Fabra University, Barcelona, Spain
2. Charité – Universitätsmedizin Berlin, Germany
3. Vall d´Hebron Institute of Oncology, Vall d'Hebron University Hospital, Autonomous University of Barcelona, Barcelona, Spain
4. Medical Oncology Service, Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain
5. Autonomous University of Barcelona, Barcelona, Spain
6. Pompeu Fabra University, Barcelona, Spain
7. Developmental Tumor Biology Laboratory, Institut de Recerca Pediàtrica- Hospital Sant Joan de Déu Barcelona, Spain
8. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
Presenter email: carlota.rubio@irbbarcelona.org

The genomic alterations in tumors are increasingly being used to effectively tailor anti-cancer therapies for individual patients. However, the difficulty to interpret the somatic variants constitutes a major hurdle to the implementation of precision cancer medicine. In particular, most of the alterations observed in tumors, including those in well known cancer genes, are mutations of uncertain significance. Moreover, the information on tumor alterations that influence the response to anti-cancer therapies is fragmented across the literature and several specialized resources. Here we present an approach that combines extensive expert curation and computational analyses to assess the biological and clinical significance of alterations in tumor genomes. With the aim to support a broad range of applications in both the pre-clinical and the translational onocology settings, we have implemented the approach in an automatic platform, which is freely available at http://www.cancergenomeinterpreter.org.

INTEGRATING MULTIDISCIPLINARY DATA TO STUDY AGRONOMICAL IMPORTANT TRAITS IN CUCURBITACEAE SPECIES

Valentino Ruggieri[1], Kostantinos Alexiou[1], Lara Pereira[1], Jason Argyris[1], Marta Pujol[1], Jordi Garcia-Mas[1]

1. IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Bellaterra, Barcelona, Spain
Presenter e-mail: valentino.ruggieri@irta.cat

The need of increasing food supply in the world, scarcity of land and climate change, impose a sustainable intensification of the agricultural production and a concrete effort of plant science to improve traits having a great potential for yield, fruit quality and resilience towards biotic and abiotic stresses. Melon (Cucumis melo L.) is one of the most important horticultural crops (Diaz et al., 2015) and huge amounts of diverse data were until now generated for this species, including high-throughput genomic, transcriptomic and phenotypic data as well as wide collections of genetic resources (Garcia-Mas et al., 2012, Argyris et al., 2015). The aim of this work points to develop a framework to integrate and concatenate all the available multidisciplinary information by using affordable and innovative bioinformatic strategies. At this purpose the genetic resources (Introgression lines, RILs), the genomic data (reference genome sequence and functional annotation of the genes) and phenotypic characterization of the traits (fruit quality traits, fruit metabolome and resistance to biotic stress) were used to map the Quantitative Trait Loci (QTL) underlying these traits. Then the transcriptomic data coming from extensive RNA-Seq analysis (>100 experiments) contributed to provide a spatio-temporal overview for the regulation of the genes of interest. Finally, the re-sequencing of 21 melon genotypes allowed surveying the number and the type of polymorphisms putatively associated with the phenotypic variation observed. The interconnection of all these layers will provide the necessary information to achieve a preliminary full picture of the biological systems underlying agronomical important traits (as, for example, climateric/non-climateric fruit ripening) and may support the development of new strategies for improving sustainable production for species belonging to the Cucurbitaceae.

COMPUTATIONAL METHODS FOR PREDICTIONS OF SELENOPROTEIN GENES

Didac Santesmasses, Marco Mariotti and Roderic Guigó

Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain
Universitat Pompeu Fabra (UPF), Barcelona, Spain
Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), 08003 Barcelona, Spain
Presenter e-mail: didac.santesmasses@crg.eu

Selenoproteins contain the 21st amino acid selenocysteine (Sec), a selenium-containing cysteine analogue. Sec is co-translationally inserted in response to specific in-frame UGA codons, normally a stop, through a dedicated machinery. The main signal for UGA recoding is a RNA hairpin loop, known as the Sec insertion sequence (SECIS), present in all selenoprotein mRNAs. Although they constitute a very small fraction of the proteome, selenoproteins cover important roles in antioxidant defense, redox regulation, thyroid hormone activation and several others. Since the UGA codon can serve as both stop or Sec signal, standard gene annotation tools do not correctly predict selenoprotein genes, and hence they are usually misannotated in genome projects and protein databases. For this reason, we developed different computational methods to identify UGA-Sec codons in genomic sequences. Our toolkit consists of: Selenoprofiles, a profile-based gene predictor for the annotation of known selenoproteins; Seblastian, a pipeline based on the search of SECIS elements as first step, that can predict new selenoproteins; and Secmarker, for the identification of selenocysteine tRNA gene (tRNA-Sec), a marker for the Sec utilization trait.
Selenoproteins are present in the three domains of life, but not in all species. We characterized the set of selenoproteins present across sequenced genomes, revealing a detailed map of the use of Sec across the tree of life. Using genome sequencing, we traced with precision the path of genomic events that lead to recent independent selenoprotein extinctions in several Drosophila species.

A NEW APPROACH TO STUDY MUTATIONS IN NON-CODING REGIONS UNCOVERS NEW SOMATIC ALTERATIONS IN CANCER RELATED TO ALTERNATIVE SPLICING.

Babita Singh[1], Juan Luis Trincado[1], Eduardo Eyras[1,2]

1. Pompeu Fabra University (UPF), E08003, Barcelona, Spain
2. Catalan Institution of Research and Advanced Studies (ICREA), E08010, Barcelona, Spain

Presenter e-mail: babita.singh@upf.edu

Alternative splicing is highly regulated process that occurs in 94% of genes and is controlled by a complex network involving multiple non-coding RNAs and hundreds of RNA-binding proteins (RBPs). Alternative splicing is related to many essential cell processes as well as diseases (Chabot & Skhreta 2016). We have recently shown that trans-acting splicing regulatory factors are frequently altered in cancer and have an impact in alternative splicing and function in cancer cells (Sebestyen et al. 2016). Additionally, it has been shown that mutations on splice-sites and on synonymous sites potentially affect splicing in tumor patients (Jung et al. 2015, Supek et al. 2014). However, splicing-altering mutations can occur anywhere on exons and introns (Sterne-Weiler, T. & Sanford 2014), and it is not yet known the extent to which mutations anywhere along the gene, including non-synonymous sites, as well as deep intronic mutations and mutations in non-coding exons may affect splicing in cancer. We have performed a comprehensive study of the mutations in multiple RNA splicing regulatory motifs present anywhere in coding and non-coding regions along genes using data from whole genome sequencing from multiple tumor types. Using ~17 million mutations from whole genome sequencing data from TCGA, about 8.6M mutations were mapped to 32,400 genes. Using a sliding window approach, we found 140,764 significantly mutated regions (SMRs), which have more mutations than expected by the local mutation rate of the gene and by the local sequence biases. The majority of these SMRs occur in introns (134,682), followed by 3'UTRs (2915), non-coding RNAs (2183), CDS regions (628) and 5'UTRs (356). These SMRs are enriched for RNA binding protein motifs corresponding to various factors, including others ELAVL2, TRA2B, SFPQ, KHSRP, MBNL1, ZFP36, as well as novel motifs. These motifs show an enrichment of mutations on specific positions compared to the mutation background and compared to those observed in the germline, indicating a cancer specific disruption of splicing regulation processes. To further identify whether these specific position of mutations have possible downstream effect on splicing, we analyzed the RNA-seq reads from the same samples. This analysis yielded a number of genes whose splicing is affected in samples with mutations in RBP motifs, including CAMLG, FAM104A, SRSF11, SRSF2, METTL23, TMEM175, and TPD52 among others. Our study describes the first genome-wide map of somatic mutations that impact RBP binding sites in cancer and their effect on splicing. Our extensive shows that many mutations in non-coding regions, as well as in coding regions, potentially affect the regulation of RNA processing and stability, and in particular, alternative splicing. Our analysis has also uncovered new somatic alterations with relevance in cancer as well as a new approach to investigate the mutations in non-coding regulatory regions of the genome.

References:

Chabot, B. & Shkreta, L. Defective control of pre-messenger RNA splicing in human disease. *J. Cell Biol.* **212,** 13–27 (2016).

Sebestyén, E., Singh, B., et al.. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res. 2016 Jun;26(6):732-44.

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156,** 1324–35 (2014).
Jung, H. *et al.* Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47,** 1242–1248 (2015).

Sterne-Weiler, T. & Sanford, J. R. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol.* **15,** 201 (2014).

NETWORK MODULES UNCOVERS MECHANISMS OF SKELETAL MUSCLE DYSFUNCTION IN COPD PATIENTS

Ákos Tényi[1,2], Isaac Cano[1,2], Francesco Marabita[3,4], Narsis Kiani[3,4],Susana Kalko[1,2], Esther Barreiro[2,5], Marta Cascante[6] and David Gomez-Cabrero[3,4,7*], Josep Roca[1,2,*]

1. Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona. Barcelona.Spain.
2. Center for Biomedical Network Research in Respiratory Diseases (CIBERES). Spain.
3. Unit of Computational Medicine, Department of Medicine, Karolinska Institutet, Stockholm, 171 77, Sweden
4. Center for Molecular Medicine, Karolinska Institutet, Stockholm, 171 77, Sweden
5. Pulmonology Dept, Muscle and Respiratory System Research Unit, IMIM-Hospital del Mar. Universitat Pompeu Fabra, PRBB, Barcelona
6. Departament de Bioquimica i BiologiaMolecular, Facultat de Biologia-IBUB, Universitat de Barcelona, 08028Barcelona, Spain
7. Mucosal and Salivary Biology Division, King's College London Dental Institute, London, SE1 9RT, United Kingdom.
Presenter e-mail: tenyi@clinic.cat

Rationale: Chronic Obstructive Pulmonary Disease (COPD) patients can show significant non-pulmonary effects, among which skeletal muscle dysfunction has a prominent impact on prognosis and disease burden.
Objective: The research explores underlying mechanisms of the phenomenon, both at rest and after endurance training.
Methods: We identified and characterised the transcriptionally active network modules of interacting proteins in the vastus lateralis in COPD patients (FEV1 46 ± 12 %pred, age 68 ± 7 yrs.) and in healthy sedentary controls (age 65 ±9 yrs.) at rest and after an 8-week endurance training program. Results were compared with various independent measurements derived from the same patients.
Main Results: At baseline, we identified four COPD specific network modules showing abnormalities in contractility, creatinine metabolism, TGF-beta signalling pathway and interferon response. The two groups presented significant physiological training effects (Δ peak oxygen uptake and Δ blood lactates peak) ($p<0.05$ each). Importantly, training-induced adaptive mechanisms displayed striking differences. Briefly, healthy subjects showed four network modules associated to training-induced changes of cell bioenergetics, not seen in the patients; whereas, training-induced adaptive changes in COPD were only seen in muscle remodelling and inflammatory responses. The comparison with independent measurements supported the above findings.
Conclusions: The study characterizes network modules identifying both disease mechanisms and abnormal training-induced adaptations in skeletal muscle of COPD. Abnormal muscle bioenergetics constitutes the most striking finding in these patients, potentially driving other altered muscle responses. The study may contribute to the design of innovative therapeutic strategies for these patients.

FAST AND ACCURATE DIFFERENTIAL SPLICING ANALYSIS ACROSS MULTIPLE CONDITIONS WITH REPLICATES

Juan C. Entizne[1,*], Juan L. Trincado[1,*], Gerald Hysenaj[2], Miha Skalic[1], Babita Singh[1], David J. Elliott[2], Eduardo Eyras[1,3]

1. Pompeu Fabra University, E08003, Barcelona, Spain
2. Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle NE1 3BZ, UK.
3. ICREA, E08010, Barcelona, Spain
* equal contribution
Presenter e-mail: juanluis.trincado@upf.edu

Multiple approaches have been proposed to study differential splicing from RNA sequencing (RNA-seq) data[1], including the analysis of transcript isoforms[2,3], clusters of splice-junctions[4,5], alternative splicing events[6-8] and exonic regions[9]. However, many challenges remain unsolved, including the limitation in speed, the computing capacity and storage requirements, the constraints in the number of reads needed to achieve sufficient accuracy, and the lack of robust methods to account for variability between replicates and for the analysis across multiple conditions. We present here a significant extension of SUPPA[8] to enable streamlined analysis of differential splicing across multiple conditions, taking into account biological variability. We show that SUPPA differential splicing achieves high accuracy using extensive experimental and simulated data compared to other methods; and shows higher accuracy at low sequencing depth, with short read lengths, and using replicas with unbalanced depth, which has important implications for the cost-effective use of RNA-seq data for splicing analysis. We also validate the analysis of multiple conditions with SUPPA by studying the differential splicing during iPS-cell to neuron differentiation and during erythroblast differentiation, providing support for the applicability of SUPPA for the robust analysis of differential splicing beyond binary comparisons.

DECIPHERING MIRNA-MRNA INTERACTIONS IN PANCREATIC CANCER USING MIRCOMB R-PACKAGE

Maria Vila-Casadesús[1,2], Elena Vila-Navarro[1], Antoni Castells[1], Juan José Lozano[2], Meritxell Gironella[1]

1. Gastrointestinal & Pancreatic Oncology Group, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD)/Hospital Clínic of Barcelona/ Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), C/Rosselló 153, 08036, Barcelona, Spain
2. Bioinformatics Platform, CIBEREHD. C/Rosselló 153, 08036, Barcelona, Spain
Presenter e-mail: maria.vila@ciberehd.org

Introduction

MiRNAs are small, stable, non-coding RNA molecules from 18-25nt long that modulate mRNA expression. One miRNA can modulate up to hundreds of genes with different functions, including oncogenes and tumor repressor genes, and one gene may be regulated by more than one miRNA. Pancreatic ductal adenocarcinoma (PDAC) has a very bad prognosis and presents an altered miRNA expression profile. Here, we try to find new deregulated mRNA targets based on miRNA-mRNA interactions, by using a combination of bioinformatics and wet lab methods.

Methods

Genome-wide miRNA profiling was done in a set of 12 fresh-frozen macrodissected surgical samples (3 Healthy, 9 PDAC) by next-generation sequencing technology. Matched mRNA expression profiling was done using affymetrix microarrays. MiRNA-mRNA interactions were analyzed using miRComb R package. This package selects differentially expressed miRNAs and mRNAs, computes miRNA-mRNA correlations and annotates them with previously described databases. The final selected miRNA-mRNA interactions are those that have a significant negative correlation and are predicted in at least one database (TargetScan, miRDB and miRSVR used).

Results

We found 201 and 1613 upregulated miRNAs and mRNAs respectively, and 342 and 2030 downregulated miRNAs and mRNAs in PDAC samples. MiRComb detected 17401 miRNA-mRNA interactions. 75% of the deregulated mRNAs are targeted by at least one miRNA, and only 37 miRNAs (~7%) do not have any mRNA target. Hsa-miR-106b, previously described as upregulated in pancreatic cancer, is on the top 10 miRNAs with more miRNA-mRNA interactions.

Conclusions

MiRNAs may be an important player in pancreatic cancer, regulating the expression of thousands of mRNAs. Our method helps us to reduce the list of mRNA targets obtained from theoretical databases and focus on the most promising ones. These results helped us to select 10 biologically relevant miRNA-mRNA interactions that will be tested experimentally and if confirmed, could be evaluated as potential therapeutic targets.

INDEPENDENT MULTIFACTORIAL ANALYSIS TO ANALYZE MULTIBLOCK DATA: APPLICATION TO IMAGING GENETIC STUDIES

Natalia Vilor-Tejedor[1-3], Silvia Alemany[1-3], Alejandro Cáceres[1-3], Mariona Bustamante[1-4], Jordi Sunyer[1-3,5], Juan R. González[1-3]

1. ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain.
2. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
3. CIBER Epidemiologia y Salud Pública (CIBERESP), Barcelona, Spain.
4. Center for Genomic Regulation (CRG), Barcelona, Spain.
5. IMIM (Hospital Del Mar Medical Research Institute), Barcelona, Spain.
Presenter e-mail: natalia.vilor@isglobal.org

Current research is increasingly based on the assessment of more than one high-throughput data from different biological modalities to help explain subject differences in complex phenotypes. As such, it is common to have measurements such as genomic, transcriptomic, neuroimaging or exposomic data on the same individuals. A common statistical strategy to assess potential associations between a phenotype and a pair of modalities is based on the massive marginal linear method in which extensive paired-wise correlations are performed between both data sets. However, the method has important limitations such as the inability to exploit the multidimensionality of data and the requirement of a large number of subjects for well-powered inferences. To address these problems, we present a statistical methodology based on an extension of Multifactorial Analysis (MFA), referred as Independent Multifactorial Analysis (IMFA). This approach is designed to evaluate potential relationships between two sources of data based on Independent Component Analysis (ICA). IMFA methodology comprises five steps: i) an ICA of each data set is performed in order to search linear combinations of variables that optimize statistically independence, ii) data sets are normalized by the square root of the first ICA component, iii) data sets are concatenated, iv) an ICA is computed on the resulting data set and v) regression analyses for the phenotype are computed to determine relevant predictors. By analyzing a real data set from BREATHE project including 135 individuals with genomic and neuroimaging information, we concluded that the use of well-established methodologies can help to detect significant causal factors in IG studies, improving the results obtained from a massive marginal analysis.

Keywords: ICA, Imaging Genetics, Multifactorial analysis, genetics, neuroimaging

Acknowledgments:
Natalia Vilor-Tejedor is funded by a pre-doctoral grant from the Agència de Gestió d’Ajuts Universitaris i de Recerca (2015 FI_B 00636), Generalitat de Catalunya- Fons Social Europeu. The research leading to these results has received funding from the European Research Council under the ERC Grant Agreement number 268479. This research was also supported by grant MTM2015-68140-R from the Ministerio de Economía e Innovación (Spain).

Poster 48

DECIPHERING THE REGULATORY TRANSCRIPTIONAL NETWORK CONTROLING *DROSOPHILA* WING DISC REGENERATION

Elena Vizcaya[1,2], Cecilia Klein[3,4], Florenci Serras[1,2], Roderic Guigó[3,4] and Montserrat Corominas[1,2]

1. Departament de Genètica, Universitat de Barcelona, Barcelona, Spain
2. Institut de Biomedicina (IBUB), Universitat de Barcelona, Barcelona, Spain
3. Centre for Genomic Regulation (CRG), Barcelona, Spain
4. Universitat Pompeu Fabra, Barcelona, Spain

Presenter e-mail: elenavizcayamolina@gmail.com

Most organisms possess some ability to repair and regenerate damaged tissues, however while some can generate whole body parts or entire limbs others can only superficially seal wounds or restore small patches of tissue. Successful regeneration processes demand a hierarchal and well- controlled balance between proliferation, differentiation and metabolic functions, which are mostly orchestrated by signaling molecules and transcriptional regulation. Although similar gene networks participate in development and regeneration, there are differences in the intensity of the signals or the levels of transcription. The ultimate goal of our research group is to understand how transcription is regulated during development and regeneration using *Drosophila* wing imaginal discs, epithelia that develop adult structures and are able to regenerate upon an injury or cell death. We have used a genetic approach to study regeneration, which consists in genetic activation of apoptosis, followed by RNA-Seq and ATAC-Seq analyses at different times after induction of damage. Then, we have correlated expression data with chromatin accessibility to identify regulatory regions such as specific enhancers for regeneration. We have identified genes that are induced or repressed upon cell- death induction and we have in vivo characterized several putative enhancer regions that may mediate the regulation of genes following damage. We also have applied motif discovery tools to find putative transcription that will allow enhancer and gene activation. Finally, we are currently trying to correlate genes to enhancers in order to have the complete chromatin landscape of the regenerative process.

Participants

Name	Surname	Institution
Josep F	Abril	UB
Josu	Aguirre Gómez	VHIR
Luis	Agulló	UVIC
Montserrat	Alavedra Farell	BIB
Mar	Albà	ICREA-IMIM-UPF
Moles Fernández	Alejandro	VHIR
Konstaninos	Alexiou	CRAG
Tyler	Alioto	CNAG-CRG
Eva	Alloza	BIB
Lorena	Alonso Parrilla	BSC
Patrick	Aloy	IRB Barcelona
Raquel	Alvarez	IBMB-CSIC
Elena	Álvarez de La Campa	CSIC - VHIR
David	Aranda Garcia	UB
Alex	Arenas	URV
Ricard	Argelaguet	EMBL-EBI
Magdalena	Arnal Segura	HOSPITAL DEL MAR
Adrià	Aterido Ballonga	VHIR
Goutham	Atla	IDIBAPS
Adrià	Auladell Martín	ICM
Laura	Avinyó	ESCI-UPF
Alejandro	Barberá Mourelle	BSC
Maria	Bargués i Ribera	Institute of Systems and Synthetic Biology, France
Davide	Bau	CNAG-CRG
Silvia	Bea Bobet	IDIBAPS
Renée	Beekman	IDIBAPS
Xavier	Belles	IBE-UPF
Vicky	Bérez	Bitac MAP, S.L.
Marti	Bernardo Faura	CRAG -
Jaume	Bertranpetit	UPF
Erica	Bianco	IBE-UPF
Juan	Blanco-heredia	Barcelona Supercomputing Center
Daniel	Boloc	UB
Sandra	Bonache Real	VHIO
Silvia	Bonàs Guarch	Barcelona Supercomputing Center
Beatrice	Borsari	BSC
Mattia	Bosio	CRG
Mario	Cáceres	ICREA / UAB
Oriol	Calvete Torres	CNIO
Juan Manuel	Calvo Martín	UB
Elias	Campo	IDIBAPS
Carla	Camprubí Font	UDG

Isaac	Cano	IDIBAPS
Laia	Carreté Muñoz	CRG
Ferran	Casals	UPF
Estefania	Casaña Lorente	UAB
Marta	Cascante	UAB
Sònia	Casillas	UAB
Pol	Castellano	Fundació Sant Joan de Déu
Sergio	Castillo	UB
David	Castillo Andreo	CNAG - CRG
Adrià	Cereto i Massagué	Centre for Omics Sciences
Vicente	Chapaprieta Martínez	UB
Heng-chang	Chen	CRG
Cecilia	Coimbra Klein	CRG
Moisès	Coll Macià	Bioinformatic Research Center (BiRC), Aarhus Universiy
Carlos	Company	CRG
Oscar	Conchillo-solé	IBB - UAB
Montserrat	Corominas Guiu	UB
Marta	Coronado Zamora	UAB
Paula	Cortés Sánchez	BSC
Lourdes	Criado Mesas	CRAG
Fernando	Cruz	CNAG - CRG
Marc	Dabad Castella	CNAG - CRG
Sara	Darriba Jiménez	UAB
Xavier	De La Cruz	VHIR
Marc	De Manuel Montero	IBE- UPF
Verónica	De Pinho Mixao	CRG
Karel W. F.	De Pourcq	UB
Bernat	Del Olmo Cabestré	IDIBGI
Santiago	Demajo Meseguer	IDIBAPS - Hospital Clínic
Sophia	Derdak	CNAG-CRG
Paolo	Di Tommaso	CRG
Noelia	Díaz	Max Planck for Molecular Biomedicine
Marcos	Díaz-gay	IDIBAPS
Orland	Diez Gibert	VHIO
Begoña	Dobon Berenguer	IBE- UPF
Laura	Domenech	CRG
Ana	Dueso Barroso	BSC
Ramon	Duran	IBB - UAB
Laura	Duran Lozano	VHIO
Martí	Duran-ferrer	IDIBAPS
Miquel	Duran-frigola	IRB Barcelona
Anna	Enjuanes	Hospital Clínic
Pau	Erola	The Roslin Institute, The University of Edinburgh
Paula	Escuer	UB
Jose	Espinosa-Carrasco	CRG
Anna	Esteve Codina	CNAG-CRG
Mónica	Estupiñán Romero	UB

Eduardo	Eyras	UPF
Irene	Farabella	CNAG-CRG
Domènec	Farré	UB
Marta	Federico	Clarivate Analytics
Evan	Floden	CRG
Martin	Floor	UVIC
Carles	Foguet Coll	UB
Clàudia	Fontserè Alemany	IBE (UPF-CSIC)
Sonia-vanina	Forcales	IGTP
Sebastià	Franch Expósito	IDIBAPS
Leonor	Frias Moya	Made of Genes
Cristina	Frías-lópez	UB
Joan	Frigola	IRB
Silvia	Galan	CNAG-CRG
Sonia	Garcia	INSTITUT BOTÀNIC DE BARCELONA
Miquel	Garcia	UAB
Sandra	Garcia	UVIC
Carla	García Fernández	UPF
Raquel	Garcia Pérez	IBE- UPF
Diego	Garrido	MARTÍN
Pere	Gelabert Xirinachs	IBE-UPF
Alejandro	Gella	UAB
Pol	Gimenez Xavier	IDIBAPS
Carla	Giner-delgado	UAB
Jèssica	Gómez Garrido	CNAG-CRG
Alexandre	Gomila	IBEC
Gloria	González Gacio	BITAC
Juan Ramón	González Ruiz	ISGLOBAL-Campus Mar
Jordi	Grau	UOC
Josep	Gregori	Roche Diagnostics S.L.
Roderic	Guigó	CRG
Marta	Guindo Martínez	BSC
Maria	Guirro Castellnous	COS-URV
Oriol	Guitart Pla	IRB Barcelona
Emre	Guney	IRB Barcelona
Sara	Gutierrez-enriquez	VHIO
So Jung	Han	IBE- UPF
Toni	Hermoso Pulido	CRG
Eva	Hernández Illán	IDIBAPS
Adrià	Hernàndez Marsal	URV
Jessica	Hernandez Rodriguez	IBE- UPF
Benjamín	Hernández Rodríguez	Max Planck Institute for Molecular Biomedicine
Tanausú-marc	Herrero Montalvillo	UB
Sergi	Hervás	
Holger	Heyn	CNAG
Jose Manuel	Hidalgo López	CRAG-IRTA
Silvia	Hinojosa	UB
Aliaksei	Holik	CRG

Hrant	Hovhannisyan	CRG
Oriol	Iborra Egea	IGTP
Valentín	Iglesias Mas	UAB
Manuel	Irimia	CRG
Laura	Isus Diaz	IRB Barcelona
Andrea	Izquierdo Bouldstridge	IBMB-CSIC
Vanessa Cristina	Jacovas	UPF
Samira	Jaeger	IRB Barcelona
Claudia	Jambrina Pallarés	UAB
Eduard	Jose Cunilleras	UAB
Ricky	Joshi	IBMB-CSIC
Teresa	Juan-blanco	IRB Barcelona
Irene Consuelo	Julca Chávez	CRG
Antonio	Julià Cano	VHIO
Effrosyni	Karakitsou	UB
Ahmed	Khafaga	UPF
Hafid	Laayouni	IBE- UPF
Julien	Lagarde	CRG
Marina	Laplana	IBB - UAB
Steven	Laurie	CNAG-CRG
Jordi	Leno Colorado	CRAG
Ion	Lerga Jaso	IBB - UAB
Audald	Lloret-villas	CRG
Irene	Lobon Garcia	IBE- UPF
Ángela	López Del Río	UPC
Núria	López-Vigas	ICREA, IRB
Damian	Loska	CNAG
Roberto	Malinverni	IJC
Maria	Maqueda Gonzalez	UPC
Joan	Martí Carreras	ICM-CSIC
David	Martin Garcia	IDIBAPS
Francisco	Martinez	CNAG-CRG
Marc A.	Marti-renom	CNAG-CRG
David	Mas Ponte	CRG
Lidia	Mateo Ramos	IRB Barcelona
Jesús	Matés	IDIBGI
Ana	Melero Caujapé	UAB
Sergio	Mendoza Funes	Clarivate analytics
Monica	Millan	ISglobal
Samuel	Miravet Verde	crg
Manuel	Molina Marin	IES Provençana
Mayukh	Mondal	IBE- UPF
Gemma	Montalban	VHIO
Antoni	Moreno	Universitat de Barcelona
Jordi	Moreno-romero	UB
Jose Luis	Mosquera	IRB Barcelona
Richa	Mudgal	IRB Barcelona
Ferran	Muiños	IRB Barcelona

Loris	Mularoni	IRB
Sergio Antonio	Muñoz	UAB
Francesc	Muyas	CRG
Ferran	Nadeu	IDIBAPS
Miguel Ángel	Naranjo Ortiz	CRG
Isaac	Noguera	IBB - UAB
Cedric	Notredame	CRG
Ramil	Nurtdinov	CRG
Jessica	Nye	IBE- UPF
Dorcas J.	Orengo	UB
Bruna	Oriol Tordera	Institut de Recerca de la Sida IrsiCaixa
Modesto	Orozco	IRB Barcelona
Natàlia	Padilla	VHIO
Sara	Pagans	UDG
Luis	Palomero	IDIBELL
Lorenzo	Pasquali	IGTP
Cinta	Pegueroles	CRG
Simon	Perera Del Rosario	Anaxomics Biotech
Sergio	Picart-armada	UPC
Mel·lina	Pinsach	UDG
Roger	Pique Regi	Wayne State University
Francisco	Pisciottano	IBYME - CONICET (Argentina) / CRG
Maria-dolors	Piulachs	IBE-UPF
Albert	Pla Planas	University of Oslo
Mercè	Planas Félix	BSC
Cristina	Prat Ferrer	IJC
Pablo	Prieto	CRG
Eva M.	Puerma	UB
Montserrat	Puiggròs Maldonado	BSC
Miquel Angel	Pujana	IDIBELL
Roser	Pujol Calvet	UAB
Jordi	Pujols Pujol	UAB
Júlia	Quer Riera	UB
Francesc	Queralt	Oncoheroes
Javier	Quilez Oliete	CRG
Emanuele	Raineri	CNAG-CRG
Jordi	Rambla	CRG
Miquel	Ràmia-jesús	Made of Genes
Mireia	Ramos Rodríguez	IGTP
Georgina	Raventós Izard	UAB
Marina	Reixachs Solé	GRIB-IMIM/UPF
Pol	Rey Serra	CRAG
Meritxell	Riera Belles	IBE- UPF
Maria Angels	Rigola Tor	UAB
Ana	Ripoll	BIB
Aida	Ripoll Cladellas	UPF-CRG
Jordi	Rodó	UAB
Maria Mercè	Rodríguez Rosàs	

Elias	Rodriguez-fos	BSC
Quirze	Rovira	CRG
Meritxell	Rovira	IRB
Julio	Rozas	UB
Carlota	Rubio	UPF-IRB
Lara	Rubio Araúna	IBE
Valentino	Ruggieri	IRTA - CRAG
Jorge	Ruiz Orera	IMIM
Marina	Ruiz-romero	CRG
Beatriz	Ruz	GRIB-IMIM/UPF
Víctor	Sacristán Fraile	UAB
Núria	Sala Serra	ICO-IDIBELL
Ahmed	Salama	UAB
Jose Francisco	Sanchez Herrero	UB
Alex	Sánchez Pla	UB-VHIO
Marta	Sánchez Quiñonero	Jesuïtes Educació
Pablo	Sánchez	ICM-CSIC
Didac	Santesmasses	CRG
Ferran	Sanz	GRIB (IMIM/UPF)
Oriol	Senan Campos	URV
Inés	Sentís	IRB Barcelona
Aitor	Serres Armero	IBE- UPF
Daniel	Silberschmidt	UAB
Babita	Singh	UPF
Marc	Sitges Puy	Made of Genes
Pau	Soteras Vidiella	UB
Juan Antonio	Subirana	UPC-IMIM
Lauro	Sumoy	IGTP
Fran	Supek	CRG
Hana	Susak	CRG, Barcelona
Ryoji	Takahashi	CNAG
Akos	Tenyi	IDIBAPS - Hospital Clinic
Natàlia	Timoneda	UB
Raul	Tonda Hernandez	CNAG
Apostolia	Topaloudi	IBE- UPF
David	Torrents Arenales	ICREA-BSC
Antonio	Torres Mendez	CRG
Albert	Trill	Atos Iberia
Juan Luis	Trincado Alonso	UPF
Gabriel	Valiente	UPC
Jordi	Valls Margarit	BSC
Juanma	Vaquerizas	Max Planck Institute for Molecular Biomedicine
Aitor	Vega Sanchez	UVic
Jordi	Vélez	UVic
Gonzalo	Vera	CRAG
Carlos	Vicient	CRAG
Maria	Vila Casadesús	IDIBAPS
Roser	Vilarrasa-blasi	UB

Jesus	Villagrasa	COSTAISA
Pablo	Villegas Mirón	IBE- UPF
Natàlia	Vilor-tejedor	ISGlobal
Elena	Vizcaya	UB
Joel	Vizueta	UB
Sandra	Walsh Capdevila	IBE- UPF
Chris	Wyatt	CRG
Lus	Zapata	CRG
Jaime	Zaragoza Riumallo	bsc
Sanja	Zivanovic	IRB Barcelona
Eduard	Zorita	CRG