Societat Catalana de BIOLOGIA

BIOINFORMATICS BARCELONA

II Jornada de Bioinformàtica i Biologia Computacional

Organitzada per la
Secció de Bioinformàtica i Biologia Computacional de la SCB i
Bioinformatics Barcelona

PROGRAMA I RESUMS DE LES COMUNICACIONS

INSTITUT D´ESTUDIS CATALANS

Sala Prat de la Riba, Institut d'Estudis Catalans
Carrer del Carme, 47. Barcelona

Carrer del Carme 47

Barcelona

12 de desembre de 2014

PROGRAM

08:30 - 9:00	Registration
09:00 - 9:15	Welcome

Dr. Ana Ripoll, President of the Bioinformatics Barcelona (BIB)
Dr. Roderic Guigó, SCB Computational Biology Section Coordinator

09:15 - 11:00 Session 1. Structural Bioinformatics (Chair: Marc A. Marti-Renom)

9:15 - 10:00 *Structural Pharmacology - The scope and role of structural approaches to understanding drug action.*
John Overington (EMBL-EBI, Cambridge).

10:00 - 10:30 *IntSide: a web server for the chemical and biological examination of drug side effects.*
Teresa Juan-Blanco (IRB)

10:30 - 11:00 *Conformational selection mechanism in protein-protein association: insights from docking.*
Chiara Pallara (BSC)

11:00 - 11:30	Coffee break (POSTERS)

11:30 - 1:15 Session 2. Evolutionary Bioinformatics (Chair: Jaume Bertranpetit)

11:30 - 12:15 Phylogenetic inference from genome-wide data.
David Posada (University of Vigo, Vigo).

12:15 - 12:45 *The genomics salad: applying phylogenomics in the green kingdom.*
Salvador Capella (CRG).

12:45 - 13:15 *A Machine-Learning Framework to Detect and Classify Hard Selective Sweeps in Human Populations.*
Marc Pybus (IBE/UPF)

13:15 - 14:15	LUNCH (POSTERS)

14:15 - 16:00 Session 3. Computational Genomics (Chair: Roderic Guigó)

14:15-15:00 *A comprehensive model of the ESC's epigenetic network.*
Alfonso Valencia (CNIO, Madrid).

15:00-15:30 *Kernel based variable importance for gene-set analysis in genetic association studies.*
Vicente Gallego (UVIC).

15:30-16:00 *The genes of the deep. A global ocean metagenomic survey from the Malaspina expedition.*
Pablo Sánchez (ICM-CSIC)

16:00 - 16:30	Coffee break (POSTERS)

16:30 - 18:15 Session 4. Biomedical Bioinformatics (Chair: Ferran Sanz)

16:30 - 17:15 *The VISC project: big data for health related research.*
Ramon Maspons (AQuAS, Barcelona)

17:15 - 17:45 *Identification of low-frequency and rare variants in type 2 diabetes…*
Silvia Bonas (BSC)

17:45 - 18:15 *In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals novel targeting opportunities.*
Carlota Rubio-Perez (IMIM/UPF)

18:15 - 18:30 Closing Remarks & JdB2014 Award
Marc A. Marti-Renom, Chair of the JdB2014

Oral presentations

STRUCTURAL BIOINFORMATICS

John Overington, EBI-EMBL, Cambridge.

Structural Pharmacology - The scope and role of structural approaches to understanding drug action

We know now the structure and bioactive conformations of a large number of drugs complexed with their molecular targets - both those targets responsible for the therapeutic efficacy, and also those for metabolism and transport. These ability to interpret drug action at a 3-D structural level has been transformed recently by progress in the determination of structures for membrane proteins, in particular for various GPCRs, and ion-channels. Our group maintains a large-scale database of drug-like ligand bioactivity data - ChEMBL https://www.ebi.ac.uk/chembl and we integrate this against ligand structures and targets proteins found in PDBe, part of the wwPB consortium. In the presentation we will overview drugs and drug targets, discuss general principles of drug action that are important to factor in to analyses, and then review the current set of drug targets from a familial and structural perspective. Future directions of our own activities and community research challenges will presented.

INTSIDE: A WEB SERVER FOR THE CHEMICAL AND BIOLOGICAL EXAMINATION OF DRUG SIDE EFFECTS

Teresa Juan-Blanco [1], Miquel Duran-Frigola[1] and Patrick Aloy [1,2,*]

[1]Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), c/ Baldiri Reixac 10-12, 08028 Barcelona, Spain.

[2]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluis Companys 23, 08010 Barcelona, Spain.

Drug side effects (SEs) are one of the main health threats worldwide, and an important obstacle in drug development. Understanding how adverse reactions occur requires knowledge on drug mechanisms at the molecular level. Despite recent advances, the need for tools and methods that facilitate side effect identification still remains.

Very recently, we presented a top-down approach to identify chemical and biological drug features that may be involved in the development of adverse drug reactions (Duran- Frigola & Aloy, 2013). We delimited the chemical and biological space for each compound by gathering molecular properties from major biomedical resources and carried out an enrichment analysis, associating more than 1,000 SEs with molecular features. On the biological side, we considered drug targets and off-targets, pathways, molecular functions and biological processes. From a chemical viewpoint, we included molecular fingerprints, scaffolds and chemical entities.

Here, we introduce a web server, named IntSide, which automates this analysis and enables the quick and easy access to our findings. Moreover, we further extend the method by integrating additional biological information, like protein interactions and disease-related genes, to facilitate mechanistic interpretations. IntSide is available at http://intside.irbbarcelona.org/.

Reference:

Duran-Frigola M, Aloy P. Analysis of chemical and biological features yields mechanistic insights into drug side effects. Chemistry & biology. 2013;20(4):594-603.

CONFORMATIONAL SELECTION MECHANISM IN PROTEIN-PROTEIN ASSOCIATION: INSIGHTS FROM DOCKING

Chiara Pallara, Manuel Rueda and Juan Fernández-Recio Joint BSC-IRB Research Programme in Computational Biology, Barcelona Supercomputing Center Center,
Barcelona, Spain. E-Mail: chiara.pallara@bsc.es

To understand cellular processes at molecular level we need to improve our knowledge of protein-protein interactions, but determining the atomic structure of many protein complexes is still challenging. Thus, structural prediction of protein-protein association is one of the major goals of computational biophysics. Despite methodological advances in docking protocols, dealing with molecular flexibility is a major bottle-neck, as the experiment CAPRI (Critical Assessment of PRediction of Interactions) [1] has shown. Indeed, state-of-the-art rigid-body docking approaches like pyDock [2] show excellent success rates [3], but have difficulties in cases with large conformational changes upon binding [4]. For complexes that form via conformational selection mechanism, in which the unbound state can sample bound conformers, a largely unexplored strategy to include flexibility in docking predictions would consist on the use of precomputed conformational ensembles generated from unbound protein structures [5]. Recently we applied this approach to a series of ubiquitin complexes, in which the use of RDCderived ensembles significantly improved docking predictions [6]. Here, we have extended this strategy to the set of 124 cases in Protein-Protein Docking Benchmark 3.0 [7]. Conformational ensembles for the unbound docking partners were automatically generated by using three different computational approaches, modeling minimization (MM), molecular dynamics (MD) and normal mode analysis (NMA). To establish the limits of the approach in optimal conditions, we first used for docking only those conformers that would be expected to give best results based on their similarity to the bound structure. Then, for a small sub-set of cases we devised a more realistic protocol by using all conformers for the docking simulations. The results show that the use of small conformational ensembles can significantly improve docking predictions in high-affinity, medium-flexibility complexes. In addition to the relevance for methodology development, his work shows that the definition of the conformational selection mechanism should focus on the sampling of bound conformations of key interface residues.

References

[1] J. Janin, "Protein-protein docking tested in blind predictions: the CAPRI experiment", Mol Biosyst, 6, 12, 2351-62, 2010.
[2] T. M. Cheng, T. L. Blundell and J. Fernandez-Recio, "pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking", Proteins, 68, 2, 503-15, 2007.
[3] C. Pallara, B. Jimenez-Garcia, L. Perez-Cano, M. Romero-Durana, A. Solernou, S. Grosdidier, C. Pons, I. H. Moal, and J. Fernandez-Recio, "Expanding the frontiers of protein-protein modeling: from docking and scoring to binding affinity predictions and other challenges", Proteins, 81, 12, 2192-200, 2013.
[4] C. Pons, S. Grosdidier, A. Solernou, L. Perez-Cano and J. Fernandez-Recio, "Present and future challenges and limitations in protein-protein docking", Proteins, 78, 1, 95-108, 2010.
[5] N. Andrusier, E. Mashiach, R. Nussinov and H. J. Wolfson, "Principles of flexible protein-protein docking", Proteins, 73, 2, 271-89, 2008.
[6] C. Pons, R. B. Fenwick, S. Esteban-Martín, X. Salvatella and J. Fernandez-Recio, “Validated Conformational Ensembles Are Key for the Successful Prediction of Protein Complexes”, Journal of Chemical Theory and Computation 9, 3 1830-1837, 2013.
[7] H. Hwang, B. Pierce, J. Mintseris, J. Janin and Z. Weng “Protein-protein docking benchmark version 3.0.”, Proteins, 73, 705-9, 2008.

EVOLUTIONARYBIOINFORMATICS

David Posada, Universidad de Vigo, Vigo.
Phylogenetic inference from genome-wide data

The unprecedented amount of data resulting from next-generation sequencing has opened a new era in phylogenomics. However, although large datasets should in theory increase resolution, multilocus data has also uncovered a great deal of phylogenetic incongruence among different genomic regions, due both to stochastic error and to the action of different evolutionary process like incomplete lineage sorting, gene duplication and loss and horizontal gene transfer. In this talk I will explain some of the most important challenges we will have to face to reconstruct the history of species. I will also describe different strategies for the phylogenetic analysis of genome-wide data that we have recently developed, including a new species tree approach that offers a very good compromise between model complexity and computational feasibility.

THE GENOMICS SALAD: APPLYING PHYLOGENOMICS IN THE GREEN KINGDOM

Salvador Capella-Gutierrez 1,2,3 & Toni Gabaldón 1,2,4.

1) Bioinformatics and Genomics Programme. Centre for Genomic Regulation (CRG). Dr. Aiguader, 88. 08003 Barcelona, Spain
2) Universitat Pompeu Fabra (UPF). 08003 Barcelona, Spain.
3) Yeast and Basidiomycete Research Group. CBS Fungal Biodiversity Centre. Uppsalalaan 8. 3584 LT Utrecht, The Netherlands.
4) Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

With a rough estimation of 300,000 different species, plants constitute one of the most diverse group of organisms. Despite their major role for oxygenation and as food source for animals - including humans - there were few complete genomes sequenced until very recently. In recent years, the publication of plant genomes has grown rapidly. From their analyses it is clear that many of them are complex in terms of genome size, number of genes, presence of alternative transcripts and a high percentage of repetitive elements. With up to 80% of repetitive regions, partially due to multiple events of polyploidization and the activity of transposable elements, the identification and comparison of gene sets across species is not easy. Phylogenomics, the intersection of evolutionary studies and genomics, can provide an appropriate framework to identify and classify differences in gene content across species.

Here, we will present the results of using a highly accurate phylogenomics pipeline in the context of a number of genome sequencing projects. We have contributed to understand different aspects of plant evolution across the whole kingdom, from red algae (*Chondrus crispus*), to sugar beet (*Beta vulgaris)* to melon (*Cucumis melo)* among others. We will describe how large-collections of single-gene phylogenies (i.e. phylomes) can help defining a stable gene-set and identifying recently-expanded transposable elements.

A MACHINE-LEARNING FRAMEWORK TO DETECT AND CLASSIFY HARD SELECTIVE SWEEPS IN HUMAN POPULATIONS

Marc Pybus[1,¶] , Pierre Luisi[1,¶] , Giovanni Dall'Olio[1,¶] , Manu Uzkudun[1] , Hafid Laayouni[1] , Jaume Bertranpetit[1], Johannes Engelken[1]

Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
C/ Doctor Aiguader 88, 08003, Barcelona, Spain

E-mail: marc.pybus@upf.edu

Detecting adaptive (or positive) selection in human genomic regions is a recurrent topic in human population genetic studies. Over the years, many positive selection tests have been implemented to highlight specific genomic patterns left by a selective event when compared to neutral expectations. However, there is little consistency among the regions detected in several genome-wide scans using different tests: population-specific demographic dynamics, local genomic features or different types of selection acting along the genome at different times and strength might explain such discrepancies. We have implemented a machine-learning classification framework based on the boosting algorithm that exploits the combined ability of some positive selection tests to uncover different features of a given selective sweep (such as completeness and oldness). Our simulation-calibrated framework estimates composite scores of several positive selection tests while controlling for population-specific demography within a hard sweep model context. As a result, we increase the sensitivity toward hard selective sweeps while adding insights about the completeness and oldness of the sweep. Our method also allows to determine the relevance of a given positive selection test under specific selection scenarios. We calibrated and applied the method to three reference populations from The 1000 Genome Project to generate a genome-wide classification map of hard selective sweeps, which identifies putative regions under positive selection in the human lineage. This study is aimed at improving the way a selective sweep is inferred by taking into account all genomic features left by selective sweeps of different intensities and time scales. We found very few signals of hard selective sweeps in the African population analyzed and compared to Asians or Europeans.

COMPUTATIONAL GENOMICS

Alfonso Valencia, CNIO, Madrid
A comprehensive model of the ESC's epigenetic network

My group is interested in the reconstruction of protein networks using computational methods to disentangle direct and indirect interactions and in particular those classified under the general category of co-evolution based approaches (Juan et al, 2013).

In this case we have collected the heterogeneous high-throughput epigenomic datasets available in public repositories for mouse embryonic stem cells (mESCs) including 139 experiments from 30 datasets with a total of 77 epigenomic features, various cytosine modifications (5mC, 5hmC and 5fC), histone marks and Chromatin related Proteins (CrPs).

We applied a set of newly developed statistical analysis methods (see Lasserre et al., 2013) with the goal of understanding the associations between chromatin states, detecting significant co-occurrence between CrPs and epigenetic modifications in different chromatin regions.

The resulting networks reveal the complex relations between cytosine modifications, protein complexes and their dependence on ESC chromatin contexts.
Furthermore, we have applied a newly developed method based on (Juan et al, 2008) to evaluate the co-evolution between families of "Chromatin related Proteins". The co-evolutionary approach brings orthogonal information that completes the epigenetic network and makes possible a new level of biological interpretation.

I will present the initial network model together with the methodology developed for this study.

This work corresponds to the paper in preparation by Carrillo de Santa Pau, Perner, Juan et al., (2014) and it was developed in collaboration with Martin Vingron's lab (MPIMG, Berlin) in the context Blueprint EU consortium (www.blueprintepigenome.eu)

References:

- Juan, D., Pazos, F., and Valencia, A. (2008). High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. Proc. Natl. Acad. Sci. U.S.a. 105, 934-939.
- Juan, D., Pazos, F., and Valencia, A. (2013). Emerging methods in protein co-evolution. Nat. Rev. Genet. 14, 249-261.
- Lasserre, J., Chung, H.-R., and Vingron, M. (2013). Finding associations among histone modifications using sparse partial correlation networks. PLoS Comp. Biol. 9, e1003168.

KERNEL BASED VARIABLE IMPORTANCE FOR GENE-SET ANALYSIS IN GENETIC ASSOCIATION STUDIES

Vicente Gallego[1], M. Luz Calle[2], Ramon Oller[3]
[1]Department of Systems Biology, University of Vic (Catalonia),
e-mail:vicente.gallego@uvic.cat.
[2]Department of Systems Biology, University of Vic (Catalonia),
e-mail:malu.calle@uvic.cat.
[3] Departament d'Economia i Empresa, University of Vic (Catalonia),
e-mail:ramon.oller@uvic.cat.
Postal Address: Sagrada Familia,7 (08500)

The identification of genetic variants that are associated with disease risk is an important goal of genetic association studies. Standard approaches perform univariate analysis where each genetic variant, usually Single Nucleotide Polimorphisms (SNPs), is tested for association with disease status. However, the marginal approach suffers from many limitations, due to small marginal effects. An alternative is testing the joint effect of a set of genetic variants within a gene or a genetic pathway. In this work we consider the Kernel logistic model for gene-set analysis where the genetic contribution is included through the kernel matrix that measures the similarities between the individuals on the basis of their SNP genotypes. However, the power of this approach may be limited if only a small fraction of the considered genetic variants are related to the disease. In this context, we propose a kernel based variable importance measure (KVI) that can be used to identify the most relevant genetic variants in association with the disease.

We proved through simulation studies that the proposed variable importance improves the rankings provided by other techniques, like the Random Forest, and also that selection of variables based on those rankings increases the power of Kernel Logistic Regression.

Acknowledgments:
This research was partially supported by grant MTM2012-38067-C02-02
from the Ministerio de Economía e Innovación (Spain) and grant 2009SGR-581 from Generalitat de Catalunya (Spain)

THE GENES OF THE DEEP. A GLOBAL OCEAN METAGENOMIC SURVEY FROM THE MALASPINA EXPEDITION

Pablo Sánchez[1*], Marta Sebastián[1], Ramiro Logare-Haurié[1], Guilem Salazar[1], Francisco M. Cornejo-Castillo[1], Carlos Duarte[2], Josep M. Gasol[1], Silvia G Acinas[1].
[1]Departament de Biologia Marina i Oceanografia. Institut de Ciències del Mar, ICM-.CSIC, Pg Marítim de la Barceloneta 37-49, 08003 Barcelona; [2]Departamento de Investigación del Cambio Global. Instituto Mediterráneo de Estudios Avanzados, CSIC, Esporles.
*Email Pablo Sánchez: pablosanchez@icm.csic.es

The deep ocean is the largest habitat on Earth. It is a vast repository of microbial genetic and metabolic diversity (prokaryotes, protists and viruses) which is largely unknown. It differs from the surface ocean on the high pressure, temperature and the absence of light. Recently it has been estimated that one third of the oceanic biological production of CO_2 takes place in the dark ocean, carried out mainly by microorganisms, stressing its importance in the biogeochemical cycles.
The Malaspina circunnavigation expedition 2010 has produced a large and exhaustive multidisciplinary dataset that includes more than 2.5 terabases of metagenomic information of the global ocean (over $12 \cdot 10^9$ sequencing reads), particularly from the deep layers of poorly explored geographic locations.
Here we present the preliminary data of the metagenomic analysis of 150 environmental samples from 30 Malaspina stations around the globe, sampled on 7 depths from surface to 4000 m deep, with emphasis on taxonomic and functional diversity. We also explore how the sequencing depth influences the amount of information that can be obtained from a marine metagenomic dataset, particularly related to the diversity and functionality of the rare biosphere.
This unprecedented dataset will allow to get an insight on the most abundant uncultured prokaryotic organisms inhabiting the dark ocean at a global level, the discovery of new genes and the exploration of their biogeography and the metabolisms driving the biogeochemical cycles in the deep.

IDENTIFICATION OF LOW-FREQUENCY AND RARE VARIANTS IN TYPE 2 DIABETES: THE LARGEST GENOME-WIDE ASSOCIATION META-ANALYSIS BASED ON IMPUTATION WITH 1000 GENOMES AND UK10K REFERENCE PANELS IN 13.201 CASES AND 59.656 CONTROLS

Bonas,S.[1], Sanchez,F.[1,2],Guindo, M.[1], Mercader,JM.[1], Torrents,D.[1,3]

[1]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center Center, Barcelona, Spain
[2]Computer Sciences Department, Barcelona Supercomputing Center Center (BSC-CNS), Barcelona, Spain
[3]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Despite the undeniable contribution of Genome-Wide Association Studies (GWAS) in the understanding of the genetics of Type 2 Diabetes (T2D), known associated loci published to date only explain a small fraction of the estimated T2D heritability. Genotype imputation consists in predicting unobserved genotypes using a dense reference panel of haplotypes to increase the statistical power of GWAS. The emergence of novel and large catalogues of genetic variation such as the *1000 Genomes* Project (1KG) reference panel and UK10K, based on low whole genome and exome sequencing data, constitute a unique opportunity to perform a more accurate fine-mapping of known associated regions and to identify new markers associated to T2D by re-analysing existing GWAS datasets available in public repositories.
In this work we have exploited the features of GWImp-COMPSs, an application developed in our group to perform genome-wide genotype imputation and association testing, requiring minimal user intervention. We re-analysed using genotype imputation with 1KG and UK10K (http://www.uk10k.org/) reference panels, 6 GWAS datasets available in dbGaP repositories comprising 72,857 (13,201 cases and 59,656 controls) effective subjects. The meta-analysis using 1KG panel resulted in 9,824,136 variants with minor allele frequency higher than 0.1%. This approach allowed us discovering two novel *loci* and replicating 11 previous known regions mainly found by large consortia meta-analysis. These novel *loci* include 3 imputed missense genome-wide significant variants, only available through 1KG imputation. These results demonstrate that re-analysing public available data using novel and denser reference panels results in the identification of novel candidate genes for complex diseases.

IN SILICO PRESCRIPTION OF ANTICANCER DRUGS TO COHORTS OF 28 TUMOR TYPES REVEALS NOVEL TARGETING OPPORTUNITIES

Carlota Rubio-Perez[1], David Tamborero[1], Michael P. Schroeder[1], Albert A. Antolín[2], Jordi Deu-Pons[1], Christian Perez-Llamas[1], Jordi Mestres[2], Abel Gonzalez-Perez[1], Nuria Lopez-Bigas[1,3]
[1]Biomedical Genomics lab, Research Program on Biomedical Informatics, IMIM Hospital del Mar Medical Research Institute and Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Catalonia, Spain, Barcelona, ES,
[2]Systems Pharmacology, Research Program on Biomedical Informatics, IMIM Hospital del Mar Medical Research Institute and Universitat Pompeu Fabra, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain, Barcelona, ES
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), E08010 Barcelona, Spain.

The development of targeted therapies against altered driver proteins holds the promise of selectively and efficiently eliminating cancer cells. Here, we present the first large-scale therapeutic landscape of cancer in a 6.792 sample cohort covering 28 tumor types.

To discover actionable driver events, we first identified mutational, copy number alteration and fusion cancer driver genes by detecting signals of positive selection across tumor cohorts. Next, we gathered information on FDA approved and on clinical or pre-clinical therapeutic agents targeting driver genes (Drivers Actionability Database). By combining data of Drivers and Drivers Actionability Databases, we developed the novel *in silico* drug prescription approach, which determines which drugs could benefit each of the cancer patients based on the genomic alterations of the tumor.

In all, we identified 460 mutational cancer driver genes acting in one or more of the tumor types along with 39 driver genes acting via CNAs or gene fusions. Fifty of these cancer driver genes are targeted by FDA approved agents, 63 by molecules currently in clinical trials and 74 are bound by pre-clinical ligands. We also identified 81 therapeutically unexploited targetable cancer genes. Lastly, by applying *in silico* drug prescription we found that only 6.7% of the patients could be treated following clinical guidelines, while up to 40% could benefit from different types of repurposing opportunities of approve drugs, and up to 78% considering treatments currently under investigation. This result highlights the current scope of targeted anti-cancer therapies and its prospects for growth.

Posters

Number	Author
1	Abril, Josep
2	Agullo, Lluis
3	Bau, Davide
4	Caceres, Alejandro
5	Campanera, JosepMaria
6	Capellades, Jordi
7	Carbonell, Pablo
8	Casserras, Teresa
9	Carreté, Laia
10	Castelo, Robert
11	Córdova, Aldo
12	Corrales, Marc
13	Duran Frigola, Miquel
14	Engelken, Johannes
15	Garcia, Gerardo
16	Giner, Carla
17	Guillén, Yolanda
18	Guindo, Marta
19	Isus, Laura
20	Julca Chávez, Irene
21	Malinverni, Roberto
22	Mandage, Ragendra
23	Martin-Lopez, Manuel
24	Martínez, Alexander
25	Martinez, Francisco
26	Matsoukas, Minos
27	Munar, Marta
28	Obiol_Pardo
29	Piñero, Janet
30	Rambla, Jordi
31	Rodriguez-Fos, Elias
32	Romero-Garcia, Javier
33	Sanchez, Alex
34	Sanchez, Friman
35	Timoneda, Natalia
36	Ventura, Salvador

DISTILLING A NETWORK OF INTERACTIONS TO UNCOVER GENES INVOLVED IN RETINITIS PIGMENTOSA DISEASE

Sergio Castillo-Lara[1], Daniel Boloc[1], Gemma Marfany[1,2,3],
Roser Gonzàlez-Duarte[1,2,3], Josep F Abril[1,2,*].
1 Departament de Genètica, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain;
2 Institut de Biomedicina (IBUB), Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain;
3 CIBERER, Instituto de Salud Carlos III, Barcelona, Catalonia, Spain.
* Corresponding autor: jabril@ub.edu

Retinitis Pigmentosa (RP) is a highly heterogeneous genetic disorder with more than 60 known causative genes. The identification of those genes, which are also closely related to non-syndromic retinal dystrophies, has increased steadily during the last decade. However, about 30% of the cases described for RP remain unassigned. Some of the known RP genes, such as CERKL, are still poorly understood at molecular level to properly describe their function and how this relates to the observed phenotype in retinal cells. Although a considerable amount of genetic and functional data on single RD genes and mutations has been gathered, a comprehensive view is still missing. In contrast to the analysis of individual isolated genes, finding the networks linking disease genes provides powerful ethiopathological insights. Our approach relies on the connectivity of the network generated by merging data from different sources: high-throughput data from BioGRID and STRING databases; manually curated data for interactions retrieved from iHOP; as well as interactions filtered out by syntactical parsing from up-to-date abstracts and full-text papers related to RP research field. We analyzed the paths emerging when known RP genes are used as bait over the whole interactome. In order to simplify the search space we kept the minimal number of connections among those genes and their closer neighbors. We are integrating tissue-specific expression levels and phenotypic data on top of this simplified network, centered on the RP causative genes, while providing an interactive interface for the molecular biologist to explore that network.

MODULATION OF THE IMMUNE RECEPTORS CD300F AND CD200R AS A NOVEL THERAPEUTIC STRATEGY FOR ACUTE CNS DAMAGE

Luis Agulló and Jordi Villà-Freixa
Research Group in Bioinformatics and Medical Statistics (BEM), Polytechnic School, University of Vic – Central University of Catalonia, 08500 Vic

The existence of different types of immune receptors with the capability to regulate microglia/macrophage function opens a new window for the development of new neuroprotective strategies in acute CNS damage. This project specifically focuses on the therapeutical potential of two pivotal immune receptors in this pathology, CD300f and CD200R. Our aim is, using structural bioinformatics approaches, to characterize the binding site of their endogenous ligands (Ca2+-sphingomyelin and CD200, respectively), and to obtain putative drugs able to modulate their function. Models of CD300f have been submitted to molecular dynamics at different Ca2+ concentrations and different structural frames resulting from these simulations used to perform docking studies with sphingomyelin. On the other hand, CD200R alone or forming complexes with its ligand CD200 has also been submitted to molecular dynamics and the interaction energy between receptor and ligand residues analyzed. Results obtained with CD300f suggest that, in the presence of Ca2+, the loop CDReq2 undergoes a conformational change uncovering a previously hidden binding site for Ca2+-sphingomyelin. The binding site extents from this loop to the cavity formed by β-hairpin C-C′ and loop CDReq3. Otherwise, matching of the calculated interaction energy (CD200R-CD200) with reported experimental data of the binding of peptides representing segments of the sequence of the proteic ligand CD200 and with agonistic and antagonistic behavior, has allowed us to define the minimum peptide with potential activating properties.

DISTINCT STRUCTURAL TRANSITIONS OF CHROMATIN TOPOLOGICAL DOMAINS CORRELATE WITH COORDINATED HORMONE-INDUCED GENE REGULATION

Francois Le Dily,1,2,3 Davide Bau_,1,3 Andy Pohl,1,2 Guillermo P. Vicent,1,2 Francois Serra,1,3 Daniel Soronellas,1,2 Giancarlo Castellano,1,2,4 Roni H.G. Wright,1,2 Cecilia Ballare,1,2 Guillaume Filion,1,2 Marc A. Marti-Renom,1,3,5 and Miguel Beato1,2

1Gene Regulacion, Stem Cells, and Cancer Program, Centre de Regulacio_ Geno_mica (CRG), 08003 Barcelona, Spain; 2Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain; 3Genome Biology Group, Centre Nacional d'An_alisi Geno_mica (CNAG), 08028 Barcelona, Spain; 4Hospital Cl_ınic, Universitat de Barcelona, 08036 Barcelona, Spain; 5Institucio_ Catalana de Recerca i Estudis Avanc ¸ats (ICREA), 08010 Barcelona, Spain

The human genome is segmented into topologically associating domains (TADs), but the role of this conserved organization during transient changes in gene expression is not known. Here we describe the distribution of progestin-induced chromatin modifications and changes in transcriptional activity over TADs in T47D breast cancer cells. Using ChIP-seq (chromatin immunoprecipitation combined with high-throughput sequencing), Hi-C (chromosome capture followed by high-throughput sequencing), and three-dimensional (3D) modeling techniques, we found that the borders of the ~2000 TADs in these cells are largely maintained after hormone treatment and that up to 20% of the TADs could be considered as discrete regulatory units where the majority of the genes are either transcriptionally activated or repressed in a coordinated fashion. The epigenetic signatures of the TADs are homogeneously modified by hormones in correlation with the transcriptional changes. Hormone-induced changes in gene activity and chromatin remodeling are accompanied by differential structural changes for activated and repressed TADs, as reflected by specific and opposite changes in the strength of intra-TAD interactions within responsive TADs. Indeed, 3D modeling of the Hi-C data suggested that the structure of TADs was modified upon treatment. The differential responses of TADs to progestins and estrogens suggest that TADs could function as ''regulons'' to enable spatially proximal genes to be coordinately transcribed in response to hormones.

INFERENCE OF POLYMORPHIC INVERSIONS ON SNP DATA REVEALS ASSOCIATION OF INVERSION AT 15Q24.2 WITH CHILDREN'S INTELLIGENCE

Alejandro Cáceres & Juan R Gonzalez, CREAL, PRBB, Barcelona, 08003

Inversion polymorphisms have important phenotypic and evolutionary consequences in humans. We have developed two different methodologies to infer inversions from SNP dense data, enabling the use of large cohorts for their study. We detected the convergence of the two methods at 15q24.2 indicating the existence of a novel inversion, in a deletion prone region associated with mental retardation and Autism. We characterized three putative inversion-induced haplotypes (called NI, Ia and Ib) that showed Mendelian inheritance in trio analysis and strongly correlated with expression levels of local genes, especially *MAN2C1* and *SNUPN*, in blood and brain. Worldwide population analysis revealed an African origin of the three haplotypes with significant population stratification. We found association with verbal and non-verbal intelligence quotient (IQ) in 2,735 children of European ancestry from three independent population cohorts. Homozygosity for the NI was associated with lower verbal IQ (2.5-point loss, p-value=0.008), while homozygosity for the Ia and Ib was associated with non-verbal IQ: 1.5-point gain for Ia (p-value=0.0004) and 2-point loss (p-value=0.001) for Ib. Our data indicate that common polymorphic inversion-related haplotypes at 15q24.2 influence human intelligence most likely by regulating gene expression in brain. We show that systematic analyses, such as convergence of different methods can reveal the contribution of inversions to the ancestral composition of populations and to the heritability of human disease.

SCIENCENODES: COLLABORATIVE PLATFORM FOR BIOINFORMATICS WEB APPLICATIONS. THE CASE OF THE CRUCIAL RESIDUES THAT MAY FAVOR THE OLIGOMERISATION OF AMYLOID-ß PEPTIDES

Albert Vilà,[a] Gerard Cegarra,[a] Lluís Belanche,[a] Ramon Pouplana,[b] and Josep M. Campanera[b]*
[a] Departament de Llenguatges i Sistemes Informàtics, Facultat d'Informàtica de Barcelona, Universitat Politècnica de Catalunya, C/Jordi Girona, 1-3, 08034, Barcelona
[b] Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII, s/n, Diagonal Sud, 08028 Barcelona, campanera@ub.edu

The understanding of structural bases of the binding (or stability) free energy between two biological moieties (or in a single structure) represents one of the procedures for unravelling the intrigues of biological functions performed by those moieties. MMGBSA (Molecular Mechanics-Generalized Born Surface Area) approach can decompose the binding and stability energies along a molecular dynamics trajectory or set of docking conformations, for instance, into residues or residue pairwise contributions. However, the amount of data generated for this decomposition is vast and thus impedes univariate exploration. Alternatively, multivariate data analysis techniques such as Partial Least Squares (PLS) or Principal Component Analysis (PCA) allows an in-depth exploration of the computed energy matrices in order to find significant residues or residue pairwise contributions that govern the binding or stability energy. Furthermore, an interactive analysis is also a requirement to really understand the complexity of the data.

To deal with challenges like this, the authors designed and developed ScienceNodes,[1] a collaborative platform for bioinformatics web applications that helps to a) create web applications easily from scripts, b) work collaboratively with your colleagues, c) analyse interactively your results and d) share your applications with your audience. Here we present one of the case studies in which this methodology and this platform have been applied satisfactorily: the unravelling of the crucial residues that may favor the oligomerisation of Amyloid-ß peptides in the Alzheimer disease. [2]

[1] http://science.lsi.upc.edu
[2] Campanera, Josep M. and Pouplana, R. *Energetic contributions of residues to the formation of early amyloid-β oligomers*, Physical Chemistry Chemical Physics, 2014, on-line, DOI: 10.1039/C4CP04544K

GENERALIZABLE STATISTICALLY-ORIENTED R-WORKFLOW FOR ISOTOPIC LABEL-TRACKING EXPERIMENTS

Jordi Capellades

Metabolite concentration is deeply regulated through homeostasis, whose fluctuations are modified by lower level processes, such as gene regulation or protein modification. The evolution of these fluxes can be traced using stable isotope-labelling. This approach makes use of isotopically-labeled substrates (i.e., uniformly labeled glucose [U-^{13}C]-Glc or uniformly labeled glutamine [U-^{13}C]-Gln) to trace back the cellular fate of labeled atoms into the structures of transformed metabolites. Additionally, the advent of comprehensive metabolic profiling technologies has lately broadened the coverage of stable isotope tracing studies, allowing an unbiased mapping of fluxes through multiple metabolic pathways. In spite of the potential benefits of unbiased stable isotope-labelling approaches, there are not yet computational solutions to deal with their derived-data. Here we present GeoRge, an R-based workflow capable of analyzing non-targeted stable isotope-labelling metabolomics data based on mass spectrometry. GeoRge is based on the recursive comparison of all detected ions across samples obtained from isotope-labeled experiments and their corresponding non-labeled controls. GeoRge queries Human Metabolome DataBase (HMDB) in order to obtain a first overview of those putative metabolite identinties resulting from transformation of labeled substrates. The global overview of the molecular fate of atoms in labeled precursors allows to determine the metabolic flux through both a priori known and unanticipated pathways.

SYSTEM-LEVEL ASSESSMENT OF TISSUE-SPECIFIC DRUG TOXICITY

Pablo Carbonell
Research Programme on Biomedical Informatics (GRIB), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, IMIM (Hospital del Mar Medical Research Institute), Dr. Aiguader, 88. Barcelona, Spain.

One critical step in drug development is toxicity assessment. Notably, about 45% of drug failures in investigational drug development can be attributed to toxicity, with adverse drug reactions being the 4th leading cause of death worldwide. However, *in vivo* testing of drug toxicity is resource intensive both in terms of time and cost. The development of *in silico* predictive models of drug toxicity, therefore, should facilitate the improvement in quality of drug candidates ensuring, thus, lower attrition rates. In this study, starting with high-quality preclinical drug toxicity data (eTOX project), we put the focus on a systems biology approach for modeling drug toxicity with the aim of providing a mechanistic understanding of the toxic effect of compounds at different levels (pathway, cell, tissue, organ). To that end, we have developed a protocol that integrates toxicity endpoint information with gene expression perturbation data in cell lines in response to toxic agents (LINCS project). Such toxicity information was then used to simulate the associated tissue-specific metabolic phenotype as predicted by a genome-scale model of human cells (Recon 2). By using flux variability analysis, we determined the set of reactions whose perturbation could critically disrupt the optimal state of the cell. The analysis of tissue-specific gene up/downregulation associated with toxicity responses provided clues about which critical pathways were altered as well as their relationship with toxicity mechanisms; helping us in that way to ultimately define a drug and/or particular dose as toxic or non-toxic.

IDENTIFYING TRANSCRIPTOMIC NETWORK DIFFERENCES IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE MAJOR PHENOTYPES

Teresa Casserras[1], Rosa Faner[2,3], Ignacio Coca[1], Tamara Cruz[2,3], Alejandra Lopez-Giraldo[2,3], Alvar Agustí[2,3,4] and Susana Kalko[1]

[1:] Bioinformatics Platform Institut d'investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
[2]Fundació Clínic per a la Recerca Biomèdica, Barcelona, Spain
[3]Ciber Enfermedades Respiratorias (CIBERES), Madrid, Spain
[4]Thorax Institute, Pulmonary Service, Hospital Clinic, IDIBAPS, University of Barcelona, Spain

Chronic Obstructive Pulmonary Disease (COPD) is characterized by persistent airflow limitation associated with an enhanced inflammatory response to inhaled particles and gases, mostly, tobacco smoking. Airflow limitation in COPD is due to a mixture of airways disease (bronchiolitis) and parenchymal destruction (emphysema), the relative contributions of which vary from person to person. Given the very distinct nature of these two pathological processes, we hypothesized that the inflammatory response and their structuration associated to bronchiolitis and emphysema may also be different. The characterization of these differences can contribute to identifying novel potential molecular targets for drug development.
Network analysis has the potential to integrate complex data sets and to reveal previously unrecognized mechanisms of disease. In this paper we used network analysis to: *(1)* test the hypothesis that the inflammatory response associated to airway (bronchiolitis) and parenchymal (emphysema) disease in COPD is different; and, *(2)* describe extensively by the first time their respective topologies main characteristics.
Lung transcriptomic profiling (Affymetrix mRNA arrays) was performed in 70 former smokers COPD. Bronchiolitic and emphysematous transcriptomic signatures were compared using statistical differential expression *RankProd* methodology. Immune correlation networks were constructed in every group, using the *Cytoscape* tool to derive main topological parameters. As a complementary result, this dataset will also allow us to evaluate other very innovative methodologies, focusing on differential networks analysis, as is the *Cytoscape* plug-in *de-MAP*.

Preliminary Results: Patients with emphysema present an up-regulation in B-lymphocyte activation and lymphoid follicle formation genes that is detected even in mild DLCO impairment (60-80%), and in the emphysematous immune network these genes are hubs. In contrast, in patients with bronchiolitis activation of Wnt signaling and collagen remodeling was observed.
We have explored and identified principal hubs and other critical elements in the networks that can function as checkpoints between sub-networks (clusters) and can facilitate/limit crosstalk between different nodes. Clustering coefficients and densities of the networks show large differences between both phenotypes.

GENOME VARIATION ACROSS CLINICAL AND COMMENSAL ISOLATES IN THE EMERGING FUNGAL PATHOGEN *CANDIDA GLABRATA*

Laia Carreté[1,2], Cécile Fairhead[3], Toni Gabaldón[1,2,4]
[1]Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Doctor Aiguader, 88, Barcelona, 08003, 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)
e-mail: Laia.Carrete@crg.eu, tgabaldon@crg.es

Infections caused by pathogenic yeasts are becoming of increasing medical importance. *Candida glabrata* is the second most common pathogenic fungi in humans. Phylogenetically, *Candida glabrata* group is a part of the Nakaseomyces, a group more closely related to Saccharomyces cerevisiae than to the model pathogen *Candida albicans* (Gabaldón et al., 2013). This indicates that virulence towards humans has independently and recently emerged within this clade. Also, virulence and antifungal resistance properties can vary across strains. In addition, some recent studies indicate that the genomic structure across clinical isolates is highly dynamic and may be associated to the emergence of drug-resistance traits (Piškur *et al.* 2014). Considering that virulence properties can vary significantly among strains of the same species, it is important to study the detailed genetic background of pathogenic and environmental isolates.

Here, we use a genome re-sequencing approach to investigate the variability among 32 different genomes from clinical and commensal *C. glabrata* samples. We did a computational analysis for detecting single-nucleotide polymorphism, ploidy, copy number variation and genomic re-arrangements between these genomes. Here, we show that differences at the nucleotide level is low (0.5%-0.7%), and that most differences consist of gene losses and gains, often involving cell-wall proteins. The sequences strains are structured in seven differentiated clades, which do not cluster by geographical origin or body-site. Clades are biased towards one of the two mating types *a* or *alpha*, indicating that mating type switching occurs at very low frequencies.

Gabaldón *et al.*: Comparative genomics of emerging pathogens in the Candida glabrata clade (2013). BMC Genomics 14:623.

Piškur *et al.*: Genome structure and dynamics of the yeast pathogen *Candida glabrata* (2014). FEMS Yeast Research 10.1111/1567-1364.12145

MAPPING EQTL NETWORKS WITH MIXED GRAPHICAL MARKOV MODELS

Inma Tur[1,2], Alberto Roverato[3] and Robert Castelo[1,2,*]

[1]Dept. Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, E-08003 Barcelona, Spain
[2]Research Program on Biomedical Informatics, Institut Hospital del Mar d'Investigacions Mèdiques, Dr. Aiguader 88, E-08003 Barcelona, Spain
[3]Dept. of Statistical Sciences, Università di Bologna, Via Belle Arti 41, I-40126 Bologna, Italy
[*]Corresponding author: robert.castelo@upf.edu

Expression quantitative trait loci (eQTL) mapping constitutes a challenging problem due to, among other reasons, the high-dimensional multivariate nature of gene expression traits. Next to the expression heterogeneity produced by confounding factors and other sources of unwanted variation, indirect effects spread throughout genes as a result of genetic, molecular and environmental perturbations. From a multivariate perspective one would like to adjust for the effect of all of these factors to end up with a network of direct associations connecting the path from genotype to phenotype. In this paper we approach this challenge with mixed graphical Markov models, higher order conditional independences and q-order correlation graphs. These models show that additive genetic effects propagate through the network as function of gene-gene correlations. Our estimation of the eQTL network underlying a well-studied yeast data set leads to a sparse structure with more direct genetic and regulatory associations that enable a straightforward comparison of the genetic control of gene expression across chromosomes. Interestingly, it also reveals that eQTLs explain most of the expression variability of network hub genes.

References

I. Tur, A. Roverato and R. Castelo. Mapping eQTL networks with mixed graphical Markov models. Genetics, in press.
http://dx.doi.org/10.1534/genetics.114.169573

GENOME-WIDE DNA METHYLATION DATA IN DEPRESSIVE PSYCHOPATHOLOGY: ADJUSTING THE EPIGENETIC VARIABLE OUTLIERS FOR RISK PREDICTION ANALYSIS (EVORA) ALGORITHM

Aldo Córdova-Palomera[1,2,*], Helena Palma-Gudiel[1], Rafael Tabarés-Seisdedos[2,3] and Lourdes Fañanás[1,2]
[1]Unitat d'Antropologia, Departament de Biologia Animal, Facultat de Biologia and Institut de Biomedicina (IBUB), Universitat de Barcelona; Av. Diagonal, 645, Barcelona, Spain, 08028.
[2]Centro de Investigaciones Biomédicas en Red de Salud Mental (CIBERSAM); C/Doctor Esquerdo, 46, Madrid, Spain, 28007.
[3]Department of Medicine, University of Valencia and Instituto de Investigación Sanitaria INCLIVA; Blasco-Ibáñez 17, Valencia, Spain, 46010.

An interesting manuscript (*Bioinformatics*, 28, 1487-1494) introduced a novel adaptive index prediction algorithm called EVORA (Epigenetic Variable Outliers for Risk prediction Analysis), which may have enormous importance in the biomedical literature in the coming years. Of note, assessment of software performance based on real datasets —to complement simulated data— can improve the quality and applicability of novel tools.
The authors of the present report have run the EVORA algorithm in a real DNA methylation dataset from the Illumina Infinium HumanMethylation450 Beadchip, which covers >450,000 CpG sites across the human genome. Data for this pilot evaluation come from a training set of 6 monozygotic adult twin pairs (12 individuals) discordant in their liability for anxious-depressive psychopathology.
After evaluating its performance, the EVORA code was modified to improve biological significance the results it generates. Pathway analysis using the information from the amended algorithm allowed identifying enrichment across biological processes relevant for depressive psychopathology.
In summary, running an adjusted version of the EVORA algorithm in a real dataset has led us to confirm its conceptual feasibility, despite the use of moderate sample sizes.

[*]To whom correspondence should be addressed. **Contact:** a.cordova@ub.edu.

Keywords: DNA methylation, genome-wide, Illumina Infinium HumanMethylation450, EVORA, depression

Marc Corrales

It is widely accepted that transcription factors (TFs) control gene expression by binding regulatory sequences upstream of genes. However correct, this model does not account for position effects, which is a long standing observation that the *locus* of a gene influences its expression. We have developed a high throughput method called TRIP (Thousands of Reporters Inserted in Prallel) to integrate, map and measure the expression of reporter constructs. We have inserted 10 different reporters in *Drosophila* cells and we have collected a dataset of about 200,000 integrations in total, giving us unprecedented insight into how the organization of the genome shapes transcriptional landscapes. The integrated reporters define large domains of high and low activity. These domains are the same for all the reporters, but some promoters are more responsive to the context than others. Surprisingly, centromeric heterochromatin has a very distinct action depending on the promoter. Whereas some promoters are repressed upon landing in heterochromatin, others are activated. This difference of behavior corresponds to different architectures of the promoter, which suggests that promoters interpret the chromatin context in different ways to determine the level of transcription of a gene. Our results uncover key information for understanding transcription regulation in eukaryotes.

A CHEMO-CENTRIC VIEW OF HUMAN HEALTH AND DISEASE

Miquel Duran-Frigola[a], David Rossell[b,c] and Patrick Aloy[a,d]

a. Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB) Barcelona, c/ Baldiri Reixac 10-12, 08028 Barcelona, Spain
b. Biostatistics and Bioinformatics Unit, Institute for Research in Biomedicine (IRB) Barcelona, Spain
c. Department of Statistics, University of Warwick, Coventry, United Kingdom
d. Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain

Efforts to compile the phenotypic effects of drugs and environmental chemicals offer the opportunity to adopt a chemo-centric view of human health that does not require detailed mechanistic information. Here, we consider thousands of chemicals and analyze the relationship of their structures with adverse and therapeutic responses. Our study includes molecules related to the etiology of 934 health threatening conditions and used to treat 835 diseases. We first identify chemical moieties that could be independently associated with each phenotypic effect. Using these fragments, we build accurate predictors for approximately 400 clinical phenotypes, finding many privileged and liable structures. Finally, we connect two diseases if they relate to similar chemical structures. The resulting networks of human conditions are able to predict disease comorbidities, as well as identifying potential drug side effects and opportunities for drug repositioning, and show a remarkable coincidence with clinical observations.

THE ANEUPLOIDY METAL TRANSPORTER CANCER HYPOTHESIS IS COMPATIBLE WITH TCGA DATA

Johannes Engelken[1], Matthias Altmeyer[2], Renty B Franklin[3]
[1] Institute of Immunology, RWTH Aachen, Germany [2] University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, Denmark, [3] University of Maryland, Department of Oncology and Diagnostic Sciences, Baltimore, MD

Here we present a new model of cancer genomics that incorporates aneuploidy (chromosome chaos), a ubiquitous phenomenon of cancer cells. By combining results from the fields of zinc biology, cancer genomics and evolutionary biology, we have built a testable link between aneuploidy, zinc concentrations and carcinogenesis. Our analysis of genomic data from The Cancer Genome Atlas (TCGA) project suggests that gains and losses of certain metal transporter genes occur frequently and correlate well with transporter gene expression levels (gene dosage). Contrasted with established cancer genes which mainly act through point mutations and focal amplification/loss, these novel chromosomal candidate driver mutations in metal transporters score better. We propose a number of ways to further test this aneuploidy metal transporter cancer (AMTC) hypothesis *in silico* and *in vitro*. If confirmed, it may help to explain a number of long-standing observations in cancer research:
- Aneuploidy has been known for at least 100 years and affects essentially all malignant cancer cells, nevertheless there is no convincing explanation for its causes and consequences.
- Many known cancer driver mutations affect zinc binding sites (e.g. the three most recurrent mutations in TP53) and many cancer genes interact with zinc or copper (e.g. TP53, RB1, PTEN, MDM4, PIK3CA, MEK1, MEK2).
- Concentrations of copper are markedly increased in cancer tissue and the blood plasma of cancer patients, while zinc levels are typically decreased
- Platinum based chemotherapeutic drugs (e.g. cisplatin) are transported through the cellular copper transport system and cisplatin resistance is mediated through gene dosage of effects of certain metal transporters, possibly caused by chromosomal mutations.
- Yeast cells adapt to nutrient deficiencies through aneuploidy (a plausible analogy to cancer cells)
- Zinc and copper have a systemic effect on the “hallmarks of cancer” including DNA repair, inflammation and apoptosis, and hereby may explain part of a cancer cell´s phenotype, which is insufficiently explained by established cancer genes.
- Overexpression of several of the identified metal transporter genes has been shown to lead to malignant cellular behavior in vitro.

Reference: http://biorxiv.org/content/early/2014/03/14/002105

FROM NEURAL PROGENITORS TO NEURONAL DIFFERENTIATION: LASER MICRODISSECTION AND GLOBAL TRANSCRIPTOME ANALYSES TO ELUCIDATE STRIATAL DEVELOPMENT

M. Pardo[1], G Garcia-Diaz-Barriga[1], A Rosser, R Martin-Ibañez[1], JM Canals[1]
1 *Departament de Biologia Cel.lular, Immunologia i Neurociencies, Universitat de Barcelona - IDIBAPS*
2 Cardiff University, UK

Brain development is a process requiring precise gene regulation, but our knowledge of the regional dynamics is incomplete. Outlining gene changes during striatal development has proven challenging due to factors such as its heterogeneous structure, and intermixed neuronal subpopulations. To overcome this, we used laser microdissection to extract mRNA from mouse striatal Germinal and Mantle Zones (GZ, MZ) during different developmental stages (E12.5, E14.5, E16.5 and E18.5) obtaining gene expression profiles of striatal progenitor and postmitotic cells.
Using a linear fit algorithm, 3,636 DEGs (Differentially expressed genes) were obtained, with Hierarchical Clustering showing a high similitude between region and biologically sound GOenrichment terms. K-means (k=15) clustering was applied to obtain representative gene profiles; using Bayesian Information Criterion, clusters were further classified into 6 biologically relevant groups. Relating the patterns to a Weighted Gene-Correlation Analysis, we constructed highly relevant modules, validated with external data via Metacore ® platform; WGCNA demonstrated a different network topology for GZ and MZ samples. Alternative Splicing Events were assessed using MiDAs..
Finally, Striatal-related human microarray samples (9 wpc to 40 yo.) from the Allen BrainAtlas Institute, were tested for DEG's (5737 genes), annotated with Xspecies identifiers and compared to mouse-DEGs. Remarkably, the mouse-human set included all classic striatal markers validating biologically our results. In conclusion, our approach uses a set of bioinformatics tools for a complete transcriptome analysis to discern biologically important gene patterns, find new biologically confirmed dynamics, and provide a powerful tool to find new target genes involved in striatal development.

EVOLUTIONARY TRAJECTORIES OF HUMAN POLYMORPHIC INVERSIONS

Carla Giner-Delgado[1,2]; David Castellano[1]; Magdalena Gayà-Vidal[1]; Sergi Villatoro[1]; Mario Cáceres[1,3];

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain;
[2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain;
[3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Chromosomal inversions have been a paradigm for evolutionary biology for decades. A key effect of inversions is that they suppress recombination as heterozygotes. Because of this property, they have been proposed as key factors in processes like local adaptation, evolution of sex chromosomes, and speciation. However, little is known about the evolutionary dynamics of these mutations, especially in humans. Here, we explore evolutionary trajectories of human polymorphic inversions to better understand the forces acting on them. This study is framed within an exhaustive project towards the characterization of inversions in the human genome (INVFEST). In particular, we took advantage of the large-scale genotyping effort of 41 inversions in ~550 individuals from 7 HapMap populations, and the information provided about the global frequencies and differences among populations. Additionally, by combining inversion genotypes with 1000 Genomes Project data, we have also been able to explore the nucleotide variation patterns associated to the inversions. This analysis revealed that a high proportion of those mediated by inverted repeats are recurrent, while those with clean breakpoints seem to have a unique origin. Focusing in the inversions derived from a unique inversion event, we have applied and adapted methods to estimate their age from both nucleotide variation and frequency data. Taking into account the uncertainties of each type of information, we have explored the different demographic and selective scenarios that could lead to the observed results. Our findings suggest that some candidate inversions may have been favored by selection and deserve further molecular and phenotypic characterization.

FUNCTIONAL PROFILING OF THE GUT MICROBIOME IN HIV INFECTION

Yolanda Guillén[1,2], Marc Noguera-Julian[1,2,3], Muntsa Rocafort[1,2], Mariona Parera[1,2], Maria Casadellà[1,2], Isabel Bravo[4], Josep Coll[1,4], Julià Blanco[1,2,3], Bonaventura Clotet[1,2,3,4], Roger Paredes[1,2,3,4]

[1]irsiCaixa AIDS Res. Inst., Badalona, Catalonia, Spain
[2]Univ. Autònoma de Barcelona, Bellaterra, Catalonia, Spain
[3]Univ. de Vic-Univ. Central de Catalunya, Vic, Catalonia, Spain
[4]Unitat VIH, Hosp. Univ. Germans Trias i Pujol, Badalona, Catalonia, Spain

Background. The gut microbiome plays an essential role in human physiology. Recently, it has been reported that compositional shifts in the microbiota inhabiting gastrointestinal tract are linked to health disorders. Here we explored how HIV infection might influence its functions.

Methods. We used PICRUSt to infer the functional profile from 16S rRNA Miseq™ sequencing data. Sequence data was obtained in a cross-sectional study comparing the intestinal microbiome of HIV-negative (HIVneg) with that of HIV-1-infected subjects clustered according different phenotypes, i.e. late presenters (LP: no ART, CD4≤200 c/mm^3), elite controllers (EC: no ART, HIV-1 RNA (VL)<50 c/mL for 1 year), viremic controllers (VC: no ART, VL 50-2000 c/mL for 1 year), ART-naïve (AN: no ART, CD4≥500 c/mm^3, VL>2000 c/mL), early treated (ET: ART started ≤6 months from HIV-1 infection, VL<50 c/mL), immune concordant (IC: on ART≥2 years, CD4≥500 c/mm^3, VL<50c/mL), and immune discordant (ID on ART≥2 years, CD4≤300 c/mm^3, VL<50c/mL). Univariate correlations involving >2 groups were done applying ANOVA plus Benjamini-Hochberg correction and Tukey-Kramer post-hoc tests; 2-group comparisons were performed with the Welch's t-test (STAMP package).

Results. The parent study included 80 subjects: 58 (73%) men, 21 (26%) women and 1 (1%) transgender woman. Microbial genes infered by PICRUST in each sample were associated to one or multiple KEGG pathways. Relative to men, women showed a significant enrichment in genes involved in amino acid metabolism. Considering only men, HIV+ subjects (5 LP, 1 EC, 3 VC, 6 AN, 5 ET, 17 IC, 8 ID) showed higher proportions of genes related to Ala, Asp and Glu metabolism; and secondary bile acid biosynthesis relative to HIVneg individuals (n=13). Moreover, most differences among HIV+ men phenotypes were dominant in ID group, whose microbiome was enriched for genes related to RNA degradation; and biotin metabolism. We also observed functional differences in MSM (n=46) relative to MSW (n=9) and PWID (n=3) men. Contrary to previous studies we did not find differences in tryptophan pathway representation among VIH phenotypes.

Conclusions. The functional profile of the gut microbiome of HIV-infected subjects varies according to gender, sexual behavior and immune status, and differs among HIV phenotypes.

Glossary. ART (antiretroviral therapy), VL (viral load), CD4 (cluster of differentiation 4), MSM (Men sex men or homosexuals), MSW (men sex women or heterosexuals), PWID (people who injected drugs).

COMBINING LATEST 1000 GENOMES PHASE 3 AND UK10K REFERENCE PANELS SIGNIFICANTLY IMPROVES GENOTYPE IMPUTATION OF LOW FREQUENCY AND RARE VARIANTS

Guindo, M.[1], Bonàs, S.[1], Sanchez, F.[1,2], Torrents, D.[1,3], Mercader, JM.[1]

[1]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center Center, Barcelona, Spain
[2]Computer Sciences Department, Barcelona Supercomputing Center Center (BSC-CNS), Barcelona, Spain
[3]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Although GWAS have allowed the identification of thousands of trait-disease associations, the current methodology is far from being able to explain the estimated heritability of each trait, even when analyzing hundreds of thousands of samples. Inferring missing genotypes through genotype imputation is essential to maximize the statistical power of GWAS. Imputation takes advantage of the denser reference panels that are steadily coming out (e.g., 1000 Genomes phase 1, 2010, and phase 3, 2014 and UK10K, 2013). We hypothesized that merging results after undergoing imputation separately improves the amount of non-genotyped variants and imputation accuracy. We used GWAS data of 4.672 subjects from the Wellcome Trust Case Control Consortium (WTCCC, 2007) and performed imputation with IMPUTE2. Starting with 881.337 SNPs, and after applying stringent quality filtering criteria (Bonàs et al., in preparation), we were able to impute and test for association ~8.7, ~8.8 and ~9.6 million of SNPs, with 1KG, UK10K, and 1KGphase3, respectively. However, we were able to accurately impute ~9.9 million of SNPs when combining 1KG+UK10K and ~10.5 million of SNPs with UK10K+1KGphase3. Most of the improvement of combining reference panels was in the low-frequency (0.05 >MAF>0.01) and rare variants (0.01>MAF>0.001). In summary, we determine that a combination of the UK10K and 1KGphase3 imputation represents a promising and cost-effective strategy to gain resolution and statistical power of existing GWAS.

A NETWORK APPROACH TO SPINAL CORD INJURY

Laura Isús[1], Jofre Tenorio-Laranga[1], Mireia Herrando-Grabulosa[2], Caty Casas[2] and Patrick Aloy[1,3]

[1]Institute for Research in Biomedicine (IRB Barcelona), Joint IRB-BSC Program in Computational Biology.
[2]Group of Neuroplasticity and Regeneration Department of Cell Biology, Physiology and Immunology Institute of Neurosciences, Universitat
Autònoma de Barcelona, and Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED) Bellaterra, Spain.
[3]Institució Catalana de Recerca I Estudis Avançats (ICREA)

Spinal Cord Injury (SCI) represents a severe health problem associated with lifetime disabilities. Immediate cell death occurring after SCI is followed by a progressive death of neurons and degeneration. Depending on the severity and the proximity to the soma of the axonal injury, spinal motoneurons (MNs) may evolve to a retrograde degeneration reaction or to a regenerative process [1]. Combining proteomic data with physical and functional interaction information can provide further insights into the dynamic behavior and mechanisms involved in the degeneration and regeneration of spinal motoneurons. We propose a directed integrative approach to decipher the distinct molecular and cellular changes that contribute to each type of process and particularly in the death mechanisms and the characteristic neuropatic pain associated with the degenerative process. Combining proteomic data with protein-protein interaction networks specifically containing disease-associated genes and their direct interactors, can help us to razionalize our findings and may provide new candidate and interesting proteins for further analyses.

Comparing GSEA analysis of our lists of candidates and networks with classical functional enrichment analysis (DAVID) we have been able to identify distinct enriched pathways (motives) between the degenerative and the regenerative process instead of general GO terms and KEGG pathways common to both models and many other disorders. In conclusion, some motives become siginificant only when direct interactors were included in the GSEA (e.g., Anoikis and Autophagosome fusion events) showing that by mapping our candidates to an interaction network we are increasing the statictical power of our analysis.

References

1. Penas, C. and Font-Nieves, M. (2011) Autophagy, and BiP level decrease are early key events in retrograde degeneration of motoneurons., Cell Death and Differentiation 18, 1617–1627.
2. Subramanian, A. and Tamayo, P. (2005) Gene set enrichment analysis: A knowledge based approach for interpreting genome-wide expression profiles., Bioinformatics, 102, 25545-50.

COMPARATIVE GENOME ANALYSIS OF *PENICILLIUM DIGITATUM* AND *PENICILLIUM EXPANSUM*

[1,2,3]Irene Julca-Chávez, [2,3]Marina Marcet-Houben, [2,3,4]Toni Gabaldón
1) Universidad Autónoma de Barcelona (UAB), Spain;
2) Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain;
3) Universitat Pompeu Fabra (UPF), Barcelona, Spain;
4) Institució Catalana d'Estudis Avançats (ICREA), Barcelona, Spain.

Penicillium expansum and *P. digitatum*, causing blue and green mould, respectively, are the most important postharvest necrotrophic fungi. *P. expansum* can be found in large range of host such as apples, pears, peaches, grapes and other fruits while *P. digitaum* specifically infects citrus fruits. The genome analysis of *P. digitatum* showed that there were few differences between two sequenced isolates from Spain. In contrast the same analysis on three strains of *P. expansum* revealed a much larger amount of differences. The objective of this study was to assess genomic variability in these two species using a larger set of genomes, and accounting for other types of genomic variations such as Copy Number Variations and Recombination events.
Our results show that the strains of *P. expansum* have more SNPs than the strains of *P. digitatum*. These SNPs were not equally distributed along the genome of *P. expansum*, but rather they clusterd together. The larger number of recombination predicted in *P. expansum* indicates that the large number of SNPs may be a result of recombination. For the prediction of CNV, we are testing different methodologies.
In summary, although the two Penicillium species are phylogenetically close, they present a highly contrasting behavior at the population level. This may be, associated with their specificity to the host and their ability to recombine.

REGIONER: AN R PACKAGE FOR THE MANAGEMENT AND STATISTICAL COMPARISON OF GENOMIC REGIONS

Anna Díez-Villanueva, Bernat Gel, Marcus Buschbeck, Miguel A. Peinado, Eduard Serra, Roberto Malinverni

Management and analysis of regional genomic information is increasingly important in biological studies, either as the main outcome of an analysis or as an additional layer in a dataset. Statistically assessing the spatial relations between region sets is a fundamental part of their analysis, but so far, the available options are lacking or limited in scope.

Here we present regioneR, an R package built on top of the Bioconductor's genomic regions functions with two main aims: (1) to offer a basic set of region manipulating functions with a simple interface and (2) to create a statistical framework based on customizable permutation tests to assess the relations between genomic region sets.

The core part of the package is a permutation test specifically designed to evaluate the relations between sets of genomic regions. All functions are prepared to work with a genome and a mask, either custom or automatically loaded from *BSGenome*, and custom masks can be used to deal with complex analysis. The randomization and evaluation functions are fully customizable and users can define their own functions. For example, in addition to the included evaluation functions dealing with overlaps, distances and base-level values, it is possible to evaluate other relevant information as GC content, methylation levels or position within the chromosomes. It is even possible to change the randomization process to take into account the structural complexity of the genome using alternative randomization strategies. In addition, the included plotting functionality creates publication-ready graphics representing the results of permutation tests.

Besides its easy-to-use design, regioneR is a customizable and powerful tool to manage and analyze sets of regions, and a useful addition to the NGS and genome wide analysis toolbox.

GENETIC FACTORS AFFECTING EBV LOAD IN TRANSFORMED LCLS FROM THE 1000 GENOME PROJECT: A GWAS ON TRANSFORMATION

Rajendra Mandage, Gabriel Santpere and Arcadi Navarro

Institute of Evolutionary Biology (Universitat Pompeu Fabra-CSIC), PRBB, Doctor Aiguader 88, 08003, Barcelona, Catalonia, Spain

Worldwide 99% of the adult population is infected by the Epstein-Barr virus (EBV), which persists for long time, can cause infectious mononucleosis and is associated with multiple sclerosis and different types of tumors. Although EBV has been the focus of extensive research work, much still remains unexplored related what makes some individuals more sensitive to EBV infection and to adverse outcomes as a result of infection.

EBV is used to transform B-cells into lymphoblastoid cell lines (LCLs). We hypothesized that differences among individual LCLs in the EBV load resulting from EBV transformation may reflect different genetic susceptibility to EBV infection. To test this hypothesis, we retrieved whole-genome sequenced LCLs from the 1000G Project derived from individuals of 26 different populations worldwide. In total, 1549 individuals were subjected to viral load estimation *in silico*. This *in sillico* approach consisted of gathering all reads not mapping into the human genome reference and mapping them against the EBV reference genome; the accuracy of our viral load estimates was validated by RT-PCR.

Our results showed considerable differences in viral load among populations. The proper estimation of EBV load has made it possible to perform a genome wide association analysis (GWAS) between estimated EBV loads and genetic variants determined within the 1000G project samples. GWAS yielded many putative candidates genes could be associated with higher EBV load in host.

These candidate genes necessitate further evaluation to reveal the biological mechanisms underlying higher EBV load and EBV associated diseases.

UNDERSTANDING PROTEIN RECOGNITION USING STRUCTURAL FEATURES

Manuel A. Marín-López*, Joan Planas-Iglesias, Jaume Bonet and Baldo Oliva
Structural Bioinformatics Laboratory, Universitat Pompeu Fabra, C/Dr. Aiguader 88, 08003 Barcelona, Spain
**Tlf. 647456388, manuel.marin@upf.edu*

Protein-Protein interactions (PPIs) play a crucial role in virtually all cell processes. Thus, understanding the molecular mechanism of protein recognition is a critical challenge in molecular biology. Previous works in this field show that not only the binding region but also the rest of the protein is involved in the interaction, suggesting a funnel-like recognition model as responsible of facilitating the interacting process. Further more, we have previously shown that three-dimensional local structural features (groups of protein loops) define characteristic patterns (interaction signatures) that can be used to predict whether two proteins will interact or not. A notable trait of this prediction system is that interaction signatures can be denoted as favouring or disfavouring depending on their role on the promotion of the molecular binding. Here, we use such features in order to determine differences between the binding interface and the rest of the protein surface in known PPIs. Particularly, we study three different groups of protein-protein interfaces: i) native interfaces (the actual binding patches of the interacting pairs), ii) partial interfaces (the docking between a binding patch and a non-interacting patch), and iii) back-to-back interfaces (the docking between non-interacting patches for both of the interacting proteins). Our results show that the interaction signatures in partial interfaces are much less favoured than the ones observed in native and back-to-back interfaces. We hypothesise that this phenomenon is related to the dynamics of the molecular association process. Back-to-back interfaces preserve the exposure of the real interacting patches (thus, allowing the formation of a native interface), while in a partial interface one interacting patch is sequestered and becomes unavailable to form a native interaction.

INVFEST, A DATABASE INTEGRATING INFORMATION OF POLYMORPHIC INVERSIONS IN THE HUMAN GENOME

Alexander Martínez-Fundichely[1], Sonia Casillas[1], Raquel Egea[1], Miquel Ràmia[1], Antonio Barbadilla[1], Lorena Pantano[1], Marta Puig[1], and Mario Cáceres[1, 2]

[1] Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Barcelona, Spain. [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Newest genome sequencing technologies have uncovered an unprecedented degree of structural variation in the human genome, and have provided new insights on the genetic basis of phenotypic and disease-susceptibility differences between individuals. Most of these variants are currently catalogued in the Database of Genomic Variants (DGV). However, inversions have been relatively overlooked compared to CNVs due to their difficulty of study. Therefore, we have created "InvFEST", a data-warehouse implementation that integrates several data of interest related to inversions with an online analytical processing engine (OLAP) to gather information and compute a report of each inversion. InvFEST merges inversion predictions from healthy individuals into a non-redundant dataset taking into account the resolution (error) in breakpoint location. Moreover, it stores information from validations and genotyping assays, the association with genes and segmental duplications, and the evolutionary history of the inversions. The initial results show a low overlap between the inversion predictions of the different studies, with more than 70% of inversions predicted only by one study. Nevertheless, after filtering unreliable locations, the total number of independent inversions is reduced by half, to less than 600. This suggests that there may be diverse biases in each inversion prediction method and that our knowledge of human inversions is still incomplete. The InvFEST database aims to fill the void in inversion information by becoming a central data repository to share results and collaborate towards the complete characterization of human polymorphic inversions.

LIGAND-TARGET PREDICTION BY STRUCTURAL NETWORK BIOLOGY USING NANNOLYZE

1. Martínez-Jiménez, F., Papadatos, G., … Marti-Renom, M. a. (2013). Target Prediction for an Open Access Set of Compounds Active against Mycobacterium tuberculosis. *PLoS Computational Biology*, *9*(10), e1003253. doi:10.1371/journal.pcbi.1003253

Target identification is essential for lead optimization, drug-drug interaction prediction, dosage adjustment and side effect anticipation. Specifically, the knowledge of structural details is essential for understanding the mode of action of a compound on a target protein. Here, we present nAnnoLyze, a method for target identification that relies on the hypothesis that structurally similar binding-sites bind similar ligands. nAnnoLyze integrates structural information into a bipartite network of interactions and similarities to predict structurally detailed compound-protein interactions at proteome scale. The method was benchmarked on a dataset of 6,282 pairs of known interacting ligand-target pairs reaching a 0.96 of area under the Receiver Operating Characteristic (AUC) curve when using the drug names as an input feature for the classifier, and a 0.70 of AUC for "anonymous" compounds or compounds not present in the training set. nAnnolyze has been previously applied in an Open Source Drug Discovery initiative against Mycobacterium tuberculosis [1]. Moreover, we have performed a screening in order to predict interactions for all the compounds in the DrugBank database with each human protein structure and providing examples of target identification for known drugs against human diseases. The accuracy and applicability of our method to any compound indicate that a comparative docking approach such as nAnnoLyze enables large-scale annotation and analysis of compound–protein interaction and thus may benefit drug development.

STRUCTURE BASED PHARMACOPHORE VIRTUAL SCREENING AGAINST LYSOPHOSPATIDIC ACID RECEPTORS AND CALCINEURIN

Minos-Timotheos Matsoukas, Gianluigi Caltabiano, Leonardo Pardo

minos.matsoukas@uab.es

Laboratori de Medicina Computacional, Unitat de Bioestadistica, Facultat de Medicina, Universitat Autonoma de Barcelona, E-08193, Bellaterra, Barcelona, Spain

The use of the fast growing number of protein crystal structures that could serve as potential drug targets is been extensively used for virtual screening methods to identify novel small molecules as potential ligands. Pharmacophore modeling is one of these methods that permit screening of large databases for potential lead molecules. In this study we wave build structure and homology model based pharmacophores and identified small molecules that bind to the Lysophosphatidic acid receptors and Calcineurin.

The Lysophosphatidic acid receptors types 1-3 ($LPA_{1-3}R$) are highly homologous receptors and belong to the G protein-coupled receptor family (GPCR). They are involved in a wide range of cellular responses, including calcium mobilization, cell proliferation, cell transformation, and chemotaxis. No crystal structure has been yet reported for the LPARs. However, using a homology model of LPA_1R, we developed a pharmacophore and performed a virtual screening of the ZINC database. The hit compounds were tested *in vitro* and five of them were found to be selective against LPA_1R or LPA_2R.

Calcineurin (CN) is a serine-threonine phosphatase involved in T cell signaling. CN dephosphorylates multiple phosphoserines on nuclear factor of activated T cells (NFAT), a transcription factor, leading to its nuclear translocation and activation. Using the crystal structure of CN bound to a peptide, the same method was used to identify potential inhibitors of CN. The selected molecules were pharmacologicaly tested and four of theme were found to have immunosuppressant activity through induction of NFATc-dependent gene expression.

CHARACTERIZATION OF COMPLEX CHROMOSOMAL REARRANGEMENTS IN CHRONIC LYMPHOCYTIC LEUKEMIA

Marta Munar[1], Santiago Gonzalez[1], Sílvia Beà[2], Itziar Salaverria[2], Montserrat Puiggròs[1], Romina Royo[1,3], Josep Ll. Gelpí[1,3,4], Carlos López-Otín[5], Xose S. Puente[5], Elías Campo[2,6], & David Torrents[1,6]

[1] Joint IRB-BSC Program in Computational Biology, Barcelona Supercomputing Center Center, Barcelona; [2] Department of Pathology, Hematopathology Unit, Hospital Clínic, August Pi I Sunyer Biomedical Research Institute (IDIBAPS), Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain; [3] Computational Bioinformatics, National Institute of Bioinformatics, Spain; [4] Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Spain; [5] Dpt. Bioquímica y Biología Molecular, Universidad de Oviedo – IUOPA, Oviedo, Spain; [6] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona.

Within the aim of characterizing the landscape of complex rearrangements and in the context of the Chronic Lymphocytic Leukemia (CLL) consortium, we have analysed more than 150 tumor-control pairs of whole genomes of the disease. For this, we have used a variant calling approach (SMuFin) in combination with data from SNP array and chromosome painting. This analysis uncovered different patterns: from no reorganizations to severe rearrangements of chromosomes that result from the massive rupture and subsequent joining of the chromatin. These rearranged chromosomes contain deletions of DNA fragments that carry tumor suppressor genes, translocations that cause a change of the genomic context of genes, and fusion genes, which in some cases have been shown to generate RNA transcripts. This study helps us understanding the role of chromosomal reorganizations in tumors formation and progression.

DEVELOPMENT OF NEW DRUGS AGAINST TUBERCULOSIS. COMPUTER-AIDED IDENTIFICATION OF INHIBITORS OF THE ENZYME CDP-METHYLERYTHRITOL SYNTHASE

Obiol-Pardo, C.[1]; Garcia-Sabaté, C[1,2]; Schreiber, E.[1]; Rubio-Martínez, J[1].; Imperial, S.[2]

[1]Departament de Quimica Fisica, Facultat de Química, Univerisdad de Barcelona; 2. Departament de Bioquímica i Biología Molecular, Facultat de Biologia, Univesitat de Barcelona

Tuberculosis is one of the leading infectious diseases in humans. Discovering new treatments for this disease is urgently required, especially in view of the emergence of multiple drug resistant organisms and to reduce the total duration of current treatments. The synthesis of isoprenoids in *Mycobacterium tuberculosis* has been reported as an interesting pathway to target, and particular attention has been focused on the methylerythritol phosphate (MEP) pathway comprising the early steps of isoprenoid biosynthesis. In this context we have studied the enzyme CDP-methylerythritol synthase (CMS), the third enzyme in the MEP pathway. In the initial part of this study a homology modeling of *M tuberculosis* CMS was performed. After evaluating the quality of the model, the most important sites useful for drug discovery were suggested. To validate this model *M. tuberculosis* CMS gen was cloned and the role of key residues was analysed by site-directed mutagenesis. Once crystallographic structures of the enzyme were available (2XWN (CTP), 3OKR and 3Q80 (CDP-ME)) they were used for the computer-aided identification of inhibitors. Both classical and accelerated molecular dynamics were performed. The enzyme active site and the interphase between the two enzyme subunits were used as pharmacophores. Compounds were docked into each site and were ranked by an ensemble-best scoring scheme and high-scoring compounds were selected to be tested experimentally.

NETWORK MEDICINE ANALYSIS OF COPD MULTIMORBIDITIES

Solène Grosdidier[1], Antoni Ferrer[2], Rosa Faner[3], Janet Piñero[1], Josep Roca[4], Borja Cosío[5], Alvar Agustí[3,4], Joaquim Gea2, Ferran Sanz[1] and Laura I Furlong[1*]

[*] Corresponding author: Laura I Furlong lfurlong@imim.es

Author Affiliations

[1] Integrative Biomedical Informatics Group, Research Program on Biomedical Informatics (GRIB), IMIM (Hospital del Mar Research Institute), Barcelona, DCEXS, Universitat Pompeu Fabra, Barcelona, Spain

[2] Servei de Pneumologia, IMIM (Hospital del Mar Research Institute), Barcelona, DCEXS, Universitat Pompeu Fabra, CIBERES, Barcelona, Spain

[3] FISIB, CIBERES, Palma de Mallorca, Spain

[4] Thorax Institute, Hospital Clínic, IDIBAPS, Universitat Barcelona, Barcelona, Spain

[5] Hospital Universitari Son Espases-IdISPa, Ciberes, Palma de Mallorca, Spain

The pathogenic mechanisms underlying COPD multimorbidities are not completely understood, thus the exploration of potential molecular and biological linkages between COPD and their associated diseases is of great interest. We developed a novel, unbiased, integrative network medicine approach for the analysis of the diseasome, interactome, the biological pathways and tobacco smoke exposome, which has been applied to the study of 16 prevalent COPD multimorbidities identified by clinical experts. The results show that all COPD multimorbidities studied here are related at the molecular and biological level. In addition, we identified known biological pathways involved in COPD, such as inflammation, endothelial dysfunction or apoptosis, serving as a proof of concept of the methodology. More interestingly, we found previously overlooked biological pathways that might contribute to explain COPD multimorbidities, such as hemostasis in COPD multimorbidities other than cardiovascular disorders, and cell cycle pathway in the association of COPD with depression. Moreover, we also observed similarities between COPD multimorbidities at the pathway level, suggesting common biological mechanisms for different COPD multimorbidities. Finally, chemicals contained in the tobacco smoke target an average of 69% of the identified proteins participating in COPD multimorbidities. The network medicine approach presented here allowed the identification of plausible molecular links between COPD and comorbid diseases, and showed that many of them are targets of the tobacco exposome, proposing new areas of research for understanding the molecular underpinning of COPD multimorbidities.

THE EGA AS AN EXPERIENCE OF SHARING CONTROLLED ACCESS HUMAN

Jordi Rambla[1], A. Carreño[1], M. Alberich[1], O. Martinez Llobet[1], Marc Sitges[1], S.de la Torre[1], Ilkka Lappalainen[4], Jeff Almeida-King[4], Cristina Yenyxe[4], Jag Kandasamy[4], Vasudev Kumanduri[4], Ignacio Medina[4],Gary Saunders[4], Alexander Senf[4], Dylan Spalding[4], Saif Ur-Rehman[4] Justin Paschall[4], Paul Flicek[1] and A. Navarro[1,2,3]
[1] Centre for Genomic Regulation (CRG), Dr. Aiguader88, 08003 Barcelona, Spain and Universitat Pompeu Fabra (UPF), Barcelona, Spain
[2] Institute of Evolutionary Biology, Barcelona, Spain
[3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
[4] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), WellcomeTrust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

Nowadays, there is an increasing focus on how to share human genomic data while keeping the required privacy. The European Genome-phenome Archive (EGA) has been dealing with this subject for more than 5 years, thus, it is in a good position to help in this community effort. In this communication we will explain the lessons learned during this period, will point out some current initiatives, and will highlight some open issues.

FINDING NEW DISEASE GENES THROUGH A NOVEL GENE-BASED GWAS APPROACH

Elias Rodriguez-Fos[1], Sílvia Bonàs[1], Michele Donato[2], Calin Voichita[2], Sorin Draghici[3,4,5], David Torrents[1,2#], Josep M. Mercader[1#]

[1] Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center Center, Barcelona, Spain; [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain [2]Department of Computer Science, Wayne State University, Detroit, MI, USA; [3]Perinatology Research Branch, NICHD/NIH/DHHS, Bethesda, Maryland, and Detroit, MI, USA [4]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, USA [5]Department of Clinical and Translational Science, Wayne State University, Detroit, MI, USA; [6]Institució Catalana de Recerca i Estudis Avançats (ICREA) Pg. Lluís Companys 23, 08010 Barcelona, Spain; [#]Both authors jointly directed this work.

Type II Diabetes, which affects nearly 400 million people in the world, is a complex metabolic disease caused by a combination of genetic and lifestyle factors. Family studies estimate that genetic factors account for 26% of the total phenotypic variance. To identify which genetic variants are responsible for the heritability of T2D, Genome Wide Association Studies (GWAS) have been performed and have found more than 80 different genomic regions, mapping to 162 genes associated with susceptibility with T2D. However, these associated variants only explain around 10% known genetic heritability, which suggests that there are still many genetic variants that contribute to the disease that are still to be discovered.
In order to gain more information from the existing GWAS datasets, we have developed a novel gene-based test that combines several variants in a given gene allowing for different models of inheritance. Using a combination filtering, logistic regression and permutation approaches, the method provides a single risk score that can also be subsequently used for systems biology or network biology approaches.
We applied this method to five T2D GWAS datasets, comprising a total of 5.200 cases and 6.600 controls. Besides confirming the association of some known genes, we identified a novel genome-wide significant gene associated with T2D. Replication of these findings is being performed in 11.000 cases and 56.700 controls from publicly available datasets.
In conclusion, we present a novel method for global gene-based association analysis and prove its utility to find novel genes involved in the susceptibility of complex diseases.

DISCOVERY OF AN AMPHIPATHIC HELIX IN MG517 PROTEIN AS RESPONSIBLE OF ITS MEMBRANE ADHESION

Javier Romero-García, Xevi Biarnés, Antoni Planas
e-mail: antoni.planas@iqs.edu
Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain

MG517 is an essential enzyme of the pathogenic organism Mycoplasma genitalium that regulates the membrane fluidity[1]. The N-terminal region of this protein was modeled by means of homology modeling and Molecular Dynamics[2] but the structure of its C-terminal extension remains elusive due the lack of templates. A number of empirical experiments demonstrated that this protein needs to be attached to the membrane in order to be active and suggests that the C-terminus has an important role in this adhesion[1]. We report here a computational strategy to study this region and to find structural elements of MG517 possibly associated to membranes. The combination of predictor servers and modeling has allowed us to identify a 23 residues long amphipathic helix in the apical extreme of the C-terminal extension. The membrane association process of this helix has been simulated by Molecular Dynamics and Metadynamics.

[1] Andrés E, Martínez N, Planas A. J Biol Chem. 2011 Oct 14;286(41):35367-79.
[2] Romero-García J, Francisco C, Biarnés X, Planas A. PLoS One. 2013 Dec 3;8(12):e81990.

SCATTERPLOT CLUSTERING FOR THE INTEGRATIVE ANALYSIS OF EXPRESSION AND METHYLATION DATA

Sánchez, Alex[1,4], Carmona, Francesc[1], Bazzoco, Sarah[2,3], Arango, Diego[2,3], Ruiz de Villa, M. Carme[1]

[1] Statistics Department, Universitat de Barcelona. Avda Diagonal 643, 08028 Barcelona, Spain.
[2,] Group of Molecular Oncology, CIBBIM-Nanomedicine, Vall d'Hebron Research Institute (VHIR), Passeig Vall d'Hebron, 119-129, 08035 Barcelona, Spain.
[3] CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN); Spain.
[4,] Statistics and Bioinformatics Unit (UEB), Vall d'Hebron Research Institute (VHIR), Passeig Vall d'Hebron, 119-129, 08035 Barcelona, Spain.

Methylation of CpG dinucleotides in the promoter of genes involved in the oncogenic process has been shown to be a key process contributing to tumor initiation and/or progression. The identification of genes regulated by methylation can be done looking for patterns in the relation between gene expression and methylation that suggest the known form of regulation. In practice this is done mainly by looking for genes with an L-shaped form in a scatterplot between gene expression and the degree of methylation, although other regulation patterns may be of interest. The main objectives of this work are: (1) to select an appropriate method for scatterplot clustering that can be used to mine a multiple high-throughput dataset formed by expression and methylation data and extract the desired patterns, (2) to test the methods selected on a colon cancer dataset formed by a panel of 30 cell lines derived from colorectal tumors and validate the findings based on their biological relevance.
We considered two methods to select patterns of regulation that could be attributable to methylation: (i) Conditional Mutual Information (Liu and Qiu, 2012) and (ii) Clustering based on Splines Regression (Hastie and Tibshirani, 2009).
These methods were tested on a multiple expression-methylation dataset where genes were previously filtered in order to keep only genes showing some degree of correlation with methylation. The combined application of both methods allowed to detect two main groups of genes associated with methylation. In order to test the validity of the selection, a functional analysis of the genes derived was performed and some of the results such as an enrichment of zinc finger proteins suggests that the selection was not only statistical but also biological meaningful.

GWIMP-COMPSS: AN INTEGRATED FRAMEWORK FOR GENOME-WIDE IMPUTATION AND ASSOCIATION STUDIES ON PARALLEL COMPUTING PLATFORMS

Friman Sanchez[1,2], Silvia Bonas[1], Marta Guindo[1], Carlos Díaz[2], Enric Tejedor[2], Rosa M. Badia[2,4], Josep M. Mercader[1], and David Torrents[1,3].

[1]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center Center, Barcelona, Spain.
[2]Computer Sciences Department, Barcelona Supercomputing Center Center (BSC-CNS), Barcelona, Spain.
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain;
[4]Artificial Intelligence Research Institute (IIIA), Spanish Council for Scientific Research (CSIC), Barcelona, Spain

Genome-wide association studies (GWAS) have been a successful methodology for identifying hundreds of associations between common genetic variants and human complex traits and diseases. Additionally, genotype imputation, the process of inferring non-genotyped genetic variants based on a denser reference panel of haplotypes, has become a key approach for improving the power of GWAS, fine mapping of known association regions and for allowing large GWA meta-analyses from data sets using different genotyping platforms.
However, whole genome imputation and association testing with increasingly larger reference panels represent a high computational burden and still face important limitations. First, they require the combination of several tools working coordinately in a workflow style, each tool having different requirements of performance, parallelism, memory usage, configuration parameters, etc. Second, imputation and GWAS workflows are not easily deployable across the variety of current distributed computing infrastructures (e.g., high-performance computing clusters, grids, clouds, etc.) and usually involve complex setup work and code adaptation.
We developed GWImp-COMPSs, a research tool to phase, impute genotypes and perform association testing that requires minimal configuration and delivers optimal and robust results. GWImp-COMPSs works on top of the COMPSs framework, which steers the parallelization of the application and makes it portable between different computing infrastructures ranging from simple desktop computers to distributed computing infrastructures, such as HPC clusters, grids and cloud, without software modifications. GWImp-COMPSs allows merging the results from different reference panels such as UK10K and 1000 Genomes, and generates graphical and summary outputs providing to the non-expert user amenable information for the biological interpretation of the results. With GWImp-COMPSs we were able to perform whole-genome imputation in a total of 6000 cases and controls with the UK10K reference panel and association testing in less than 15 hours and without any user intervention using 23 nodes on the Marenostrum III supercomputer. GWImp-COMPSs also represents an ideal tool for large multi-center GWAS consortia, allowing whole genome imputation and association testing in a standardized way across different institutions, without the need of sharing the individual-level data.

COMPUTATIONAL METHODS TO EXPLORER VIRAL SPECIES FROM WASTEWATER METAGENOMICS DATA

Natàlia Timoneda[123*], Joel Sabaniego[12], Xavier Fernandez-Cassi[3], Rosina Girones[3,] Josep F. Abril[12].

1 Departament de genètica, Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Catalunya, Spain.
2 Institut de Biomedicina (IBUB), Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Catalunya, Spain.
3 Departament de microbiologia, Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Catalunya, Spain.

*natalia.timoneda@gmail.com

Treated wastewater is increasingly recognized as a resource of water nutrients and is reused in industry, for landscape irrigation, aquifer recharge and, especially in Mediterranean Europe, also for irrigation of fresh produce for human consumption. Wastewater contains many potential well known pathogenic bacteria and viruses; but also other potential emerging bacterial and viral pathogens largely unknown.
Although ICTV recognized 2243 viral species to date, the most recent estimates determine that we have identified and characterized less than 0.1% of the viruses present on this planet. However this number itself is likely a gross underestimate. For instance, approximately 40% of all diarrhea, being the third leading infectious cause of death worldwide, cases are unknown etiology.
Our main objective is expand the list of known viruses present in urban sewage using metagenomics analyses. For that, we have generated a customized protocol to analyses viral wastewater data from NGS (MiSeq Illumina); and have had to adjust the parameters for the distinct bioinformatics tools used. Roughly; main steps of this protocol include standard cleaning of the raw metagenomics sequences, filtering most informative ones and performing search in order to detect known species. After that, unmapped sequences, at both levels raw reads and assembled contigs, may correspond to novel variants or species; so the sequences candidates are classified to facilitate the posterior experimental validation and markers selection.

WHAT MAKES A PROTEIN SEQUENCE A PRION?

Raimon Sabate[1], Frederic Rousseau[2,3], Joost Schymkowitz[2,3] and Salvador Ventura[4]

[1]Departament de Fisicoquímica, Facultat de Farmàcia, and Institut de Nanociència i Nanotecnologia (IN^{2}UB), Universitat de Barcelona, Avda. Joan XXIII s/n, E- 08028-Barcelona, Spain

[2]VIB Switch Laboratory, VIB, Leuven, Belgium

[3]Departement for Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium

[4]Institut de Biotecnologia i de Biomedicina and Dept. de Bioquímica i Biologia Molecular. Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Typical amyloid diseases such as Alzheimer's and Parkinson's were thought to exclusively result from *de novo* aggregation, but recently it was shown that amyloids formed in one cell can cross-seed aggregation in other cells or even individuals, following a prion-like mechanism. Despite the large experimental effort devoted to understanding the phenomenon of prion transmissibility, it is still poorly understood how this property is encoded in the primary sequence. In many cases, prion structural conversion is driven by the presence of relatively large glutamine/asparagine (Q/N) enriched segments. Several studies suggest that it is the amino acid composition of these regions rather than their specific sequence that accounts for their priogenicity. However, our analysis indicates that it is instead the presence and potency of specific short amyloid-prone sequences that occur within intrinsically disordered Q/N-rich regions that determine their prion behaviour, modulated by the structural and compositional context. This provides a basis for the accurate identification and evaluation of prion candidate sequences in proteomes in the context of a unified framework for amyloid formation and prion propagation.

PARTICIPANTS

Nom	Cognoms	Institució
Josep F	Abril	UB
Luis	Agulló	UVIC
Konstantinos	Alexiou	IRTA
Patrick	Aloy	IRB
Amaya	Amador	UB
Maite Garazi	Barrón Aduriz	IBE
Davide	Bau	CNAG
Jaume	Bertranpetit	UPF/IBE
Xevi	Biarnés Fontal	IQS
Sílvia	Bonàs	Barcelona Supercomputing Center Center
Marta	Bosch Rodriguez	IDIBAPS
Arnau	Bria	CRG
Alejandro	Caceres	CREAL
Mario	Cáceres	ICREA / UAB
M.luz	Calle	UVIC
Juanma	Calvo	UB
Josep Maria	Campanera Alsina	UB
Eva	Capdevila Busquets	IRB
Salvador	Capella	CRG
Jordi	Capellades	Yanes Lab
Maria Nerea	Capon Lamelas	UAB
Pablo	Carbonell	GRIB, IMIM/UPF
Laia	Carreté	CRG
Elvira	Carrio Gaspar	IMPPC
Teresa	Casserres	IDIBAPS
Robert	Castelo	UPF
Sergio	Castillo	UB
Emilio	Centeno	CRAG
Aldo	Cordova-Palomera	UB
Marta	Coronado Zamora	UAB
Marc	Corrales Berjano	CRG
Florencia	Correa Fiz	CReSA
Lenore	Cowen	Tufts University/CRG
Fernando	Cruz	CNAG
Yasmina	Cuartero	CNAG
Xavier	Daura	ICREA/UAB
Karel	De Pourcq	UB/CRAG
Bernat	Del Olmo Cabestré	Centre de Genètica Cardiovascular
Marco	Di Stefano	CNAG
Dani	Diaz de Quijano	UB
Anna	Díez Villanueva	IMPPC
Miquel	Duran-frigola	IRB

Jordi	Durban Sanchez	IBV -CSIC
Johannes	Engelken	RWTH Aachen
Antonio	Espinosa	UAB
Juan	Fernandez-Recio	Barcelona Supercomputing Center
Guillaume	Filion	CRG
Oscar	Flores Guri	IRB Barcelona - BSC Joint
Leonor	Frias Moya	CNAG
Vicente	Gallego	UVIC
Carolina	Garcia	
Sònia	Garcia	UB
Alfonso Gerardo	García Díaz Barriga	UB
Pere	Gelabert Xirinachs	UAB
Radina	Georgieva	CRG
Carla	Giner Delgado	UAB
Marisa	Girós	Hospital Clínic
Jessica	Gomez	CNAG
Alex	Gómez Garrido	Esteve
Juan Carlos	Gomez Verjan	UNAM
Santiago	González	Barcelona Supercomputing Center Center
Gloria	González Gacio	BITAC MAP S.L
Mike	Goodstadt	CNAG
Christian	Griñan Ferre	UB
Roderic	Guigo	CRG
Amy Lauren	Guillaumet	CNAG
Yolanda	Guillén Montalbán	IrsiCaixa
Marta	Guindo	Barcelona Supercomputing Center
Toni	Hermoso Pulido	CRG
Sergi	Hervás Fernández	UAB
Santiago	Imperial	UB
Laura	Isús Díaz	IRB Barcelona
Andrea	Izquierdo Bouldstridge	IBMB-CSIC
Rory	Johnson	CRG
Teresa	Juan Blanco	IRB Barcelona
Irene Consuelo	Julca	CRG
Susana	Kalko	IDIBAPS
Jordi	Leno Colorado	UB
Pablo	Librado	UB
Tomàs	Llorens Marès	CEAB-CSIC
Ruben	Lopez	Aladid
Pierre	Luisi	IBE (UPF-CSIC)
Roberto	Malinverni	IMPPC
Rajendra	Mandage	UPF-CSIC/PRBB
Marina	Marcet Houben	CRG
Manuel Alejandro	Marín López	UPF
Francisco	Martinez	CNAG/CRG

Alexander	Martinez Fundichely	IBB
Marc A.	Marti-Renom	CNAG/CRG
Gemma	Mas de Xaxars Giner	UB
Ramon	Maspons	
Jesús	Matés	GENCARDIO
Minos-timotheos	Matsoukas	UAB
Eduardo	Mayol	UAB
Josep M	Mercader	Barcelona Supercomputing Center
Ana	Milovanovic	Barcelona Supercomputing Center Center
Antoni	Moreno	UB
Loris	Mularoni	UPF
Marta	Munar Ortiz	Barcelona Supercomputing Center Center
Miguel Ángel	Naranjo	CRG
Arcadi	Navarro	UPF
Isaac	Noguera	IBB
Marc	Noguera-Julian	IrsiCaixa
Baldo	Oliva	UPF
Dorcas J.	Orengo Ferriz	UB
Rüdiger	Ortiz Álvarez	CEAB-CSIC
John	Overington	
Chiara	Pallara	Barcelona Supercomputing Center Center
Lorenzo	Pasquali	IGTP
Cinta	Pegueroles	CRG
Janet	Piñero	IMIM
Josep	Piñol	CREAF-UAB
Antoni	Planas Sauter	IQS
Mercè	Planas-Fèlix	Barcelona Supercomputing Center
David	Posada	
Eva M.	Puerma	UB
Roser	Pujol	UAB
Marc	Pybus	CSIC/UPF
Núria	Radó Trilla	Academia Europaea
Jordi	Rambla	CRG
Oscar	Reina Garcia	IRB Barcelona
Juliana	Ribera Catarina	MAG
Javier	Rivera	IrsiCaixa
Muntsa	Rocafort Juncà	IrsiCaixa
Gustavo	Rodriguez	CNAG
M.cristina	Rodríguez	IRSICAIXA
Elias	Rodriguez Fos	Barcelona Supercomputing Center
Bernardo	Rodríguez Martín	Barcelona Supercomputing Center
Lukasz	Roguski	CNAG
Javier	Romero-García	IQS
Julio	Rozas	UB

Carlota	Rubio-Perez	UPF(DCEXS)
Mateo	Ruiz Conca	UB
Pablo	Sánchez	ICM-CSIC
Friman	Sánchez Castaño	BSC, Life Science
Jose Francisco	Sanchez Herrero	UB
Alex	Sánchez Pla	UB
Walter	Sanseverino	CEO de Sequentia Biotech
François	Serra	CNAG
Gerard	Serra Vidal	IBE- UPF
Marc	Sitges	CRG
Carolina	Soriano Tarraga	IMIM
Daniel	Soronellas	CRG
Yannick	Spill	CNAG
Jean-françois	Taly	CRG
Javier	Tapial	CRG
Natàlia	Timoneda Solé	UB
David	Torrents	ICREA-BSC
Javier	Urchueguía	UPV
Alfonso	Valencia	
Salvador	Ventura	IBB-UAB
Gonzalo	Vera	CRAG
Maria	Vila Casadesús	IDIBAPS
José Luis	Villanueva-Cañas	FIMIM
Joel	Vizueta	UB