III Jornada de Bioinformàtica i Biologia Computacional

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

Esponsoritzada per

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

18 de desembre de 2015

PROGRAM

9:00 – 9:30 Registration
9:30 – 9:45 Wellcome (Roderic Guigó & Ana Ripoll)

9:45 - 10:15 **Arcadi Navarro (ICREA, UPF, CRG)** "Big - Omics Data and cool things one can do with it."

10:15 - 10:30 Radhakrishnan Sabarinathan (IMIM, UPF), "Nucleotide excision repair is impaired by binding of transcription factors to DNA."
10:30 - 10:45 Elena Álvarez de la Campa (VHIR), "Prediction of pathogenic variants in mitochondrial disease: state of the art".
10:45 - 11:00 Bernardo Rodríguez-Martín (BSC), "Novel recurrent *ZFP36L1-IGH* fusion gene in mature b-cell neoplasms."
11:00 - 11:15 Babita Singh (UPF), "Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks."

11:15 - 11:45 Coffee Break

11:45 - 12:00 Brian Jimenez-Garcia (BSC) "LIGHTDOCK: a novel protein-protein docking framework for the new challenges in the interactomics era".
12:00 - 12:15 Francisco Martínez-Jiménez (CNAG-CRG), "Rationally designed drug blending as a mechanism to overcome drug resistance in cancer."
12:15 - 12:30 Jordi Serra-Musach (ICO-IDIBELL), "Network activity predicts cancer therapeutic response and synergism."
12:30 - 12:45 Pablo D. Dans (IRB) "T-RNASAURUS REX and the freezing of the genetic code"
12:45 - 13:00 Atos as a global IT partner in life science projects, AtoS

13:00 - 13:30 **Tim Hubbard (King's College, London)** "The 100.000 Genome project".

13:30 - 15:15 Lunch in "El Fresc Co" (Carrer del Carme 16)

15:15 - 15:45 **David Carrera (BSC, UPC)** "Technology Trends in Data-Centric Computing for Bioinformatics."
15:45 - 16:00 Cinta Pegueroles (CRG, UPF) "Selection on long non-coding RNAs is acting at both sequence and structural level."
16:00 - 16:15 Lopez-Sanchez M (CREAL), "MADSEQLOY: a bioconductor package to analyse loss of chromosome Y (loy) in age- and smoke-related diseases."
16:15 - 16:30 Narcís Fernandez-Fuentes (GRIB-IMIM) "*DE NOVO* design of peptides to modulate protein-protein interactions."

16:30 - 17:00 Coffee Break

17:00 - 17:30 **Modesto Orozco (IRB, UB)** “The complex relationship between computers and biology”

17:30 - 17:45 Fran Supek (CRG, UPF, RBI) “The functional importance of synonymous mutations in cancer and microbes analyzed in massive genomic data sets.”

17:45 - 18:00 Adrià Aterido (VHIR) “Genome-wide pathway analysis identifies new genetic pathways associated with psoriasis.”

18:00 - 19:30 Current status of the **Galaxy Project** and its growing community
Toni Espinosa (UAB) & Gonzalo Vera (CRAG)

David Torrents, Chair of the JdB2015

Oral presentations

NUCLEOTIDE EXCISION REPAIR IS IMPAIRED BY BINDING OF TRANSCRIPTION FACTORS TO DNA

Radhakrishnan Sabarinathan[1], Loris Mularoni[1], Jordi Deu-Pons[1], Abel Gonzalez-Perez[1], Núria López-Bigas[1,2,†]

1. Research Program on Biomedical Informatics, IMIM Hospital del Mar Medical Research Institute and Universitat Pompeu Fabra, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain

2. Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

†Corresponding author. E-mail: nuria.lopez@upf.edu

Presenter Email: sabari.radha@upf.edu

Somatic mutations are one of the major genetic alterations that contribute to the transformation of normal cell into cancer cell. The rate of somatic mutations appear in great variability across the genome due to chromatin organization, DNA accessibility and replication timing. However, the impact of DNA-binding regulatory proteins, like transcription factors, on mutation rate variability in the local regions of the genome has not been studied in detail. Here, by analysing the whole genome somatic mutations of three different tumor types (melanoma, lung adenocarcinomas and lung squamous cell carcinomas) sequenced by TCGA, we demonstrate that the rate of somatic mutations is highly increased at active Transcription Factor binding sites (TFBS) compared to their flanking regions. Using the recently available excision-repair sequencing (XR-seq) data (Hu et al., 2015), we show that the higher mutation rate at these sites is caused by a decrease of the levels of nucleotide excision repair (NER) activity. Therefore, our work demonstrates that DNA-bound proteins interfere with the NER machinery, which results in an increased rate of mutations at their binding sites. This finding has important implications in our understanding of mutational and DNA repair processes and in the identification of cancer driver mutations.

PREDICTION OF PATHOGENIC VARIANTS IN MITOCHONDRIAL DISEASE: STATE OF THE ART

Elena Álvarez de la Campa, Sergi Lois, Xavier de la Cruz, Tomàs Pinòs, Elena García

In this work, we analyse the predictive power of bioinformatics tools in relation to the functional impact of missense mutations in mitochondrial protein encoding genes. Specifically, we have focused on the following predictors: Polyphen-2, MAPP, MutAssessor, SIFT, Condel and PhD-SNP. In addition, we have developed a novel pathological predictive tool used specially for mitochondrial proteins: Biotrans.

For benchmarking purposes, we employed a set of 186 variants, of which 124 were neutral and 62 pathological. The latter include variants tagged as disease mutations in UniProtKB/Swiss-Prot and variants tagged as Confirmed in MitoMap.

Overall comparison of the predictors shows that three of them have a moderately better performance: Polyphen2 (MCC = 0.67), MutAssessor (MCC = 0.77) and Biotrans (MCC = 0.81). Condel and SIFT did not classify mutations in mitochondrial proteins. PhDSNP (MCC = 0.45) and MAPP (MCC = 0.11) displayed a poorer performance on mitochondrial mutations than with nuclear mutations. Interestingly, a protein-by-protein analysis shows that the performance of all these methods fluctuates, making them complementary in many cases.

On the basis of these results, we are defining an optimal protocol of usage for these predictors when applied to mutations in mitochondrial proteins.

NOVEL RECURRENT ZFP36L1-IGH FUSION GENE IN MATURE IN B-CELL NEOPLASMS

I Nagel*[1], I Salaverria*[2], S Gonzalez*[3], B Rodríguez-Martín[3], G Clot[2], J Pinteno[3], E Campo[2,4], D Torrents[#3,4], S Beà[#2] and R Siebert[#1]

[1] Institute of Human Genetics, Christian-Albrechts-University Kiel & University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany
[2] Hematopathology Unit, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Barcelona, Spain
[3] Joint IRB-BSC Program in Computational Biology, Barcelona Supercomputing Center (BSC), Barcelona, Spain.
[4] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
* contributed equally, # corresponding authors

Chronic Lymphocytic Leukemia (CLL) is the most common type of leukemia in adults. The ICGC-CLL Spanish consortium was launched with the aim of better understand the genetic bases of this heterogeneous disease and translating this knowledge into the clinic.

In this context, we have identified a recurrent 37 Mb deletion of the chromosome 14 in 52 patients with mature B-cell malignancies, mainly CLL. Interestingly, patients with the deletion displayed a more aggressive form of CLL than the ones with a normal chromosome. The fine-mapping of the deletion breakpoints with SMuFin and sanger sequencing revealed that 14q deletions fuse the Immunoglobulin Heavy Chain (IGH) locus with the ZFP36L1 gene, which is a post-transcriptional regulator involved in cellular growth and apoptosis pathways in B-cells. The RNA-seq analysis and RT-PCR in eight patients confirmed that IGH-ZFP36L1 fusion is transcribed into a chimeric mRNA with the potential to encode for a truncated ZFP36L1 protein. Additionally, the comparison of gene expression profiles revealed a pronounced down-expression of genes involved in the post-transcriptional regulation of gene expression for those patients with IGH-ZFP36L1 fusion. This finding suggests that the expression of an aberrant ZFP36L1 protein may interfere in those regulatory processes where it is involved. Finally, 14q deletions were associated with the motif recognized by the Translin protein, which is known to mediate the generation of IGH oncogenic fusions.

In summary, the integrated analysis of clinical, genomic and transcriptional data suggests that this novel and recurrent IGH-ZFP36L1 fusion is a driver event in CLL.

LARGE-SCALE ANALYSIS OF GENOME AND TRANSCRIPTOME ALTERATIONS IN MULTIPLE TUMORS UNVEILS NOVEL CANCER-RELEVANT SPLICING NETWORKS

Babita Singh1,*,Endre Sebestyén1,*, Belén Miñana1,2, Amadís Pagès1, Francesca Mateo3, Miguel Angel Pujana3, Juan Valcárcel1,2,4, Eduardo Eyras1,4,5

1Universitat Pompeu Fabra, Dr. Aiguader 88, E08003 Barcelona, Spain
2Centre for Genomic Regulation, Dr. Aiguader 88, E08003 Barcelona, Spain
3Program Against Cancer Therapeutic Resistance (ProCURE),Catalan Institute of Oncology (ICO), Bellvitge Institute for Biomedical Research (IDIBELL), E08908 L'Hospitalet del Llobregat, Spain.
4Catalan Institution for Research and Advanced Studies, Passeig Lluís Companys 23, E08010 Barcelona, Spain

*Equal contribution
5Correspondence to: eduardo.eyras@upf.edu

Change in alternative splicing is known to play crucial role in many diseases including cancer. Many hallmarks of cancer such as cell proliferation, angiogenesis and metastasis involve aberrant splicing patterns. However role of alternative splicing in tumor development and factors that regulate these splicing has not been studied yet in an exhaustive way. To understand splicing in cancer, we analyzed RNA and DNA sequencing data provided by The Cancer Genome Atlas (TCGA) for 11 solid tumor types from more than 4000 patients. We also studied 1348 RNA-binding proteins to understand the mechanisms behind these changes. We found several differential splicing patterns in genes between normal and corresponding tumor samples. We found novel mutations and copy number variations on RNA-binding proteins associated with splicing changes in cancer drivers and oncogenic pathways. We observed differential splicing of NUMA1, a cancer driver gene involved in mitotic process, in breast and kidney tumors. This was mainly controlled by MBNL1, a splicing factor known to be involved in cell differentiation. Knock down experiments of MBNL1 or using antisense oligonucleotides to modify NUMA1 splicing pattern lead to enhanced cell proliferation, mimicking tumor-like behavior in a normal breast epithelial cell model. Our study therefore unveils novel splicing patterns in multiple tumors and propose detailed splicing regulatory network that potentially contribute to tumor development and progression.

LIGHTDOCK: A NOVEL PROTEIN-PROTEIN DOCKING FRAMEWORK FOR THE NEW CHALLENGES IN THE INTERACTOMICS ERA

Brian Jimenez-Garcia (speaker) <brian.jimenez@bsc.es>
Juan Fernandez-Recio <juanf@bsc.es>

Structural prediction of protein-protein interactions can contribute to understand essential cell processes at molecular level. The field is rapidly developing and many computational docking methods have been reported in the past years [1]. While current docking algorithms, including our pyDock approach, can successfully model a significant number of protein-protein cases [2, 3], it is necessary to explore new approaches that can overcome current challenges such as conformational flexibility during association. Here we present LightDock, a novel protein-protein docking framework, which has been designed and implemented with the goal of being adaptable to current and future challenges and easily extensible by its users.
LightDock is based in the Glowworm Swarm Optimization [4] (GSO) strategy, an algorithm from the swarm intelligence family of the artificial intelligence algorithms. Each agent of the GSO algorithm, a glowworm, represents a protein docking conformation formed by two partner proteins. The glowworms emit light depending on the *luciferin*, a variable encoding the performance of the energetic scoring function, and attract other glowworms from the surrounding area. This process of emitting light and attracting other glowworms describes the sampling of the different docking conformations. The protein-protein docking scoring function can be easily defined by the user and makes the algorithm independent from any force-field.
Application of this new approach to rigid-body partners has improved the sampling of the docking poses in terms of ligand RMSD and enrichment in the number of near-native solutions. The method is easily adaptable to use precomputed conformational ensembles in order to include flexibility in the model. Different full-atom and coarse-grained functions have been included for multi-scale scoring, such as pyDock energetic terms, and the algorithm has been extended to include the anisotropic network model (ANM) for normal mode analysis (NMA) and local minimization.

REFERENCES

[1] Janin, J., Wodak, S. J., Lensink, M. F., & Velankar, S. (2015). Assessing Structural Predictions of Protein–Protein Recognition: The CAPRI Experiment. *Reviews in Computational Chemistry Volume 28*, 137-173.
[2] Cheng, T. M. K., Blundell, T. L., & Fernandez-Recio, J. (2007). pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, *68*(2), 503-515.
[3] Pallara, C., Jiménez-García, B., Pérez-Cano, L., Romero-Durana, M., Solernou, A., Grosdidier, S., ... & Fernandez-Recio, J. (2013). Expanding the frontiers of protein–protein modeling: From docking and scoring to binding affinity predictions and other challenges. *Proteins: Structure, Function, and Bioinformatics*, *81*(12), 2192-2200.
[4] Krishnanand, K. N., & Ghose, D. (2009). Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions. *Swarm intelligence*, *3*(2), 87-124.

RATIONALLY DESIGNED DRUG BLENDING AS A MECHANISM TO OVERCOME DRUG RESISTANCE IN CANCER

Francisco Martínez-Jiménez[1,2], John P. Overington[3] , Bissan Al-Lazikani[4], and Marc A. Marti-Renom[1,2,5]

1. Structural Genomics Group, CNAG-CRG, , Centre for Genomic Regulation Barcelona, Spain.
2. Gene Regulation Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain.
3. Stratified Medical, 40 Churchway, London NW1 1LW.
4. Institute of Cancer Research, Cotswold Road, Sutton, Surrey, United Kingdom
5. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Drug resistance is one of the major problems in cancer treatment. Rapid mutation and selective pressure can efficiently select drug-resistant mutants. Although there are many mechanisms for drug resistance, a classic mechanism is due to coding mutations in the drug-target binding site. Numerous efforts have been made to individually understand and overcome resistance to imatinib in chronic myelogenous leukemia (CML) treatment caused by the T315I mutation of ABL1 or resistance to gefitinib in non-small-cell lung cancers (NSCLC) due to point mutations in epidermal growth factor receptor (EGFR). However, there is a lack systematic analysis of the mutational landscape that can potentially cause resistance to targeted therapies.

To address this issue, we have developed a framework that predicts mutations in a drug target with the highest likelihood-resistance ratio in a specific cancer class. Subsequently, our model finds the least sensitive compounds for these resistance mutations. Finally, it defines a blend of molecules potentially overcoming resistance caused by generation of spontaneous mutations in the target in a particular cancer type. We exemplified the applicability of the framework using the ERK1/2, MEK1 and EGFR kinases. Our results show overlapping with previously reported resistant mutations in these proteins. We also provide a set of candidate molecules potentially insensitive to these mutations. In summary, our work aims to reduce the difficulties in the choice of the optimal treatment and thus, it is a step further in the development of the personalized medicine for the treatment of cancer.

NETWORK ACTIVITY PREDICTS CANCER THERAPEUTIC RESPONSE AND SYNERGISM

Jordi Serra-Musach[1], Francesca Mateo[1], Eva Capdevila-Busquets[2], Gorka Ruiz de Garibay[1], Alberto Villanueva[1], Mika Gustafsson[3], Mikael Benson[3], Samira Jaeger[2], Holger Heyn[4], Miguel Vizoso[4], Hector Pérez[4], Alex Cordero[4], Eva Gonzalez-Suarez[4], Manel Esteller[4-6], Xuexi Yang[7], Joaquín Arribas[6,8,9], Gema Moreno-Bueno[10,11], Conxi Lázaro[12], Marc Ferrer[13], Patrick Aloy[2,6] & Miguel Angel Pujana[1]*

1 Program Against Cancer Therapeutic Resistance (ProCURE), Breast Cancer and Systems Biology Unit, Catalan Institute of Oncology (ICO), Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona 08908, Catalonia, Spain.
2 Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Barcelona 08028, Catalonia, Spain.
3 The Centre for Individualized Medicine, Department of Clinical and Experimental Medicine, Linköping University, Linköping 58185, Sweden.
4 Cancer Epigenetics and Biology Program (PEBC), IDIBELL, L'Hospitalet del Llobregat, Barcelona 08908, Catalonia, Spain.
5 Department of Physiological Sciences II, School of Medicine, University of Barcelona, L'Hospitalet del Llobregat, Barcelona 08908, Catalonia, Spain.
6 Catalan Institution for Research and Advanced Studies (ICREA), Barcelona 08010, Catalonia, Spain.
7 School of Biotechnology, Southern Medical University, Guangzhou 510515, China.
8 Preclinical Research Program, Vall d'Hebron Institute of Oncology (VHIO), Barcelona 08035, Catalonia, Spain.
9 Department of Biochemistry and Molecular Biology, Autonomous University of Barcelona, Bellaterra 08193, Catalonia, Spain.
10 Department of Biochemistry, Autonomous University of Madrid (UAM), Biomedical Research Institute "Alberto Sols" (Spanish National Research Council (CSIC)-UAM), Hospital La Paz Institute for Health Research (IdiPAZ), Madrid 28029, Spain.
11 MD Anderson International Foundation, Madrid 28033, Spain.
12 Hereditary Cancer Program, ICO, IDIBELL, L'Hospitalet del Llobregat, Barcelona 08908, Catalonia, Spain.
13 Division of Preclinical Innovation, National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Bethesda, MD 20892-4874, USA.

Cancer patients often show no or only modest benefit from a given therapy. This major problem in oncology is generally attributed to the lack of specific predictive biomarkers, yet a global measure of cancer cell activity could provide a comprehensive understanding of therapy efficacy. Here, we have developed a weighted interactome network score and shown that it predicts therapeutic compound sensitivity and synergism. The integration of protein-protein interactions with basal gene expression in hundreds of cancer cell lines defines a weighted score that differentiates compounds classes, target families or types of therapies, and effector pathways. Cancer-associated biological processes and transcriptional regulators critically contribute to network activity. The analysis, centered on the major cancer drivers, used uncorrelated compound effects to reveal novel synergistic compound combinations. We validate the computational strategy on metformin, which in combination with AZD-8055, AZD-2281 (olaparib) or SL-0101 shows a pronounced synergistic effect in breast cancer cells dependent on PI3K/AKT signaling. Overall, this study defines a novel network activity function that is valuable for improving cancer therapy.

T-RNASAURUS REX AND THE FREEZING OF THE GENETIC CODE

Adélaïde Saint-Léger[1], Carla Bello-Cabrera[2,3], **Pablo D. Dans**[1,4], Adrian Gabriel Torres[1], Noelia Camacho[1], Modesto Orozco[1,4-6], Fyodor A. Kondrashov[2,3,6], and Lluís Ribas de Pouplana[1,6]

1. Institute for Research in Biomedicine (IRB), The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Catalonia, Spain.
2. Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Catalonia, Spain.
3. Universitat Pompeu Fabra (UPF), 08003 Barcelona, Catalonia, Spain.
4. Joint BSC-CRG-IRB Research Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10, 08028 Barcelona, Catalonia, Spain.
5. Departament de Bioquimica, Facultat de Biologia, Avgda Diagonal 647, 08028 Barcelona, Catalonia, Spain.
6. Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Catalonia, Spain.

The extant genetic code uses sixty-four codons that translate into twenty amino acids. How the code emerged and reached its current complexity is a question central to the origin of life, and important for the design of proteins with artificial amino acids. But it is still unknown why the standard genetic code is limited to twenty residues. The expansion of the genetic code necessitates the evolution of new transfer RNA identities, and here we show that this can be impaired by preceding constraints that block the emergence of new identity elements.
We have investigated the emergence of recognition signals required by an eukaryotic tRNA modification enzyme. We show that, in certain tRNAs, the incorporation of essential nucleotides for the activity of this enzyme was prevented by pre-existing structural features of the anticodon loop. Thus, the evolution of new tRNA identities can be blocked by the accumulation of preceding and incompatible identity elements. We extend this conclusion to complete genomic tRNA populations through the comparison of tRNA gene complements between species. We show that the rate of nucleotide substitutions in tRNAs is higher in species whose genomes contain smaller number of tRNAs, thus supporting the idea that signal saturation blocks the evolution of tRNA identities.
Thus, the evolution of tRNAs can be blocked by preexisting functional constrains, and slows down as the complexity of tRNA populations grows. We propose that the repertoire of amino acids used by the extant genetic code did not grow beyond its current size because an operational limit was reached in the number of identity elements that can be effectively encoded in tRNAs.

SELECTION ON LONG NON-CODING RNAS IS ACTING AT BOTH SEQUENCE AND STRUCTURAL LEVEL.

Cinta Pegueroles[1,2] and Toni Gabaldón[1,2,3]

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) Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

Metazoan genomes are known to encode many long non-coding RNAs (lncRNAs). Previous studies showed that lncRNAs are poorly conserved across species, and suggested that in those species having an small effective population size they may accumulate mutations mainly by drift. Here we evaluate signatures of selection in lncRNAs annotated in human, *Drosophila melanogaster* and *Caenorhabditis elegans* using three types of analysis: conservation across species, patterns of polymorphism within a species, and relationships between sequence constraints and secondary structure. In all analyses, we included a curated dataset of experimentally characterized human lncRNAs, as a reference for truly functional lncRNAs. We found evidence of purifying selection acting on lncRNAs in all species, including those with small effective population sizes. We describe that RNA secondary structure constrains sequence variation in lncRNAs, so that single nucleotide polymorphisms accumulate in highly accessible, unpaired regions and tend to be neutral with respect to structural stability. Importantly, this relation is independent of the GC content and the presence of splice-related motifs. We conclude that numerous predicted lncRNAs may be functional and may play key roles that remain to be discovered.

MADSEQLOY: A BIOCONDUCTOR PACKAGE TO ANALYSE LOSS OF CHROMOSOME Y (LOY) IN AGE- AND SMOKE-RELATED DISEASES

Lopez-Sanchez M, Gonzalez JR
Bioinformatics Research Group in Epidemiology (BRGE), Centre for Research in Environmental Epidemiology (CREAL)

Mortality and cancer incidence is higher among men than women. Genetic mosaicisms, including mosaic loss of chromosome Y (LOY) have been described as a risk factor for both haematological and solid tumours [1]. Smoking habits have a transient and dose-dependent mutagenic effect on LOY status [2]. Therefore, studying LOY as a genetic marker of susceptibility for complex diseases where ageing and smoking are known risk factor could be of interest. Novel technologies such as Next Generation Sequencing (NGS) allow the possibility of detecting LOY. So far, there are no bioinformatic tools to properly detect, visualize and analyse these events.
In this work, we introduce a Bioconductor package to detect mosaic LOY events for both NGS and SNP array data. The method relies in properly analysing probes in the male-specific region of chromosome Y in the 56-Mb region between pseudoautosomal regions1 and 2 (PAR1 and PAR2) on chromosome Y. Our functions take advantage of parallel computing capabilities and optimized algorithms. Several functions and classes have been defined to detect and visualize LOY alterations. Statistical methods to compare groups of individuals have also been created. The application and usability of the package is demonstrated by analysing a large number of samples from TCGA belonging to prostate (PRAD) and bladder (BLCA) cancer.
[1] Jacobs KB, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012, 44(6):651-8
[2] Dumansky JP et al. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. Science. 2015, 347(6217):81-3.

DE NOVO DESIGN OF PEPTIDES TO MODULATE PROTEIN-PROTEIN INTERACTIONS

Narcis Fernandez-Fuentes, Baldo Oliva

Structural Bioinformatics Group, GRIB-IMIM.

University Pompeu Fabra

narciso.fernandez@upf.edu

Protein-protein interactions (PPIs) underpin virtually all cellular processes both in health and disease. Modulating the interaction between proteins by means of small (chemical) agents is therefore a promising route for future novel therapeutic interventions. In this context, peptides are gaining momentum as emerging agents for the modulation of PPIs. In this talk, I will present an original computational approach: PiPreD, developed to model and design orthosteric peptides to block PPIs. PiPreD relies on a precompiled and bespoken library of structural motifs, iMotifs, extracted from protein complexes and a fast structural modeling algorithm driven by the location of native chemical groups on the interface of the protein target named anchor residues. PiPreD comprehensive and systematically samples the entire interface deriving peptide conformations best suited for the given region on the protein interface. Besides the technical aspects of the development of PiPreD, I will also present the experimental validation of PiPreD predictions in the development of peptides to modulate RAS-mediated pathways – and important drug target- and as complement to peptides microarrays.

THE FUNCTIONAL IMPORTANCE OF SYNONYMOUS MUTATIONS IN CANCER AND MICROBES ANALYZED IN MASSIVE GENOMIC DATA SETS

Fran Supek (fran.supek@crg.eu) [1,2,5], Belén Miñana [1,2], Juan Valcárcel [1,2,3], Toni Gabaldón [1,2,3,], Anita Kriško [4], Tea Copić [4], Ben Lehner [1,2,3]

1 Centre for Genomic Regulation (CRG), 08003 Barcelona, Catalonia, Spain.
2 Universitat Pompeu Fabra (UPF), 08003 Barcelona, Catalonia, Spain.
3 Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain.
4 Mediterranean Institute for Life Sciences (MedILS), 21000 Split, Croatia.
5 Rudjer Boskovic Institute (RBI), 10000 Zagreb, Croatia.

The availability of massive genomics data sets is rapidly increasing. Large samples allow for more statistical power to detect genetic changes with subtle but important effects. A prime example are synonymous mutations, which don't affect the amino acid sequence but may alter splicing motives (in eukaryotes) or protein translation efficiency (most evident in microbes).

We have examined somatic single-nucleotide variants across ~3800 cancer genomes for putative consequences of such synonymous mutations on gene function, and found that cancer genes tend to accumulate more synonymous mutations than expected: at least 1/5 synonymous mutations in oncogenes appear to be selected due to their effects on splicing [1]. Around 6-8% of all cancer drivers are the synonymous mutations, implicating the mis-spliced oncogenes as a potential therapeutic target.

In microbes, highly expressed genes exhibit characteristic codon usage patterns, boosting the efficiency of translation; examples are known where this may change with the environmental niche of the microbe [2]. We introduce a general statistical framework for searching for such associations in 910 prokaryotes, while controlling for confounders such as phylogenetic inertia. We have discovered 200 gene-phenotype links, and extensively experimentally validated 35 (of 44 tested) genes whose optimal codon content was associated with thermophilic, aerobic and halophilic lifestyles [3]. This is a novel 'genome context' method whose usefulness will grow with increased availability of genomes and of their phenotypic annotations. We illustrate this by extending our analyses by text-mined microbial phenotypes [4], resulting in more accurate gene function inferences.

[1] Supek F, Miñana B, Valcarcel J, Gabaldón T, Lehner B (2014) Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. *Cell* 156: 1324-1335.
[2] Supek F. (2015) The Code of Silence: Widespread Associations between Synonymous Codon Biases and Gene Function. *J Mol Evol*, advance online publication.
[3] Kriško A, Copić T, Gabaldon T, Lehner B, Supek F (2014) Inferring gene function from evolutionary change in signatures of translation efficiency. *Genome Biology* 15: R44.
[4] Brbić M, Piškorec M, Vidulin V, Kriško A, Supek F. ProTraits: a compendium of prokaryotic phenotypes and lifestyles extracted from biological literature and validated in genomic data. (unpublished; available at http://protraits.irb.hr)

GENOME-WIDE PATHWAY ANALYSIS IDENTIFIES NEW GENETIC PATHWAYS ASSOCIATED WITH PSORIASIS

Adrià Aterido[1*], Antonio Julià[1], Carlos Ferrándiz[2], Lluís Puig[3], Eduardo Fonseca[4], Emilia Fernández-López[5], Esteban Dauden[6], José Luís Sánchez-Carazo[7], José Luís López-Estebaranz[8], David Moreno-Ramírez[9], Rubén Francisco Vanaclocha[10], Enrique Herrera[11], Pablo de la Cueva[12], Nick Dand[17], Núria Palau[1], Arnald Alonso[1], María López-Lasanta[1], Raül Tortosa[1], Andrés García-Montero[13], Laia Codó[14], Josep Lluís Gelpí[14], Jaume Bertranpetit[15], Devin Absher[16], Francesca Capon[17], Richard M Myers[16], Jonathan N. Barker[17,18], Sara Marsal[1].

[1]Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain; [2]Hospital Universitari Germans Trias i Pujol, Badalona, Barcelona, Spain; [3]Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; [4]Complejo Hospitalario Universitario de A Coruña, A Coruña, Spain; [5]Hospital Universitario de Salamanca, Salamanca, Spain; [6]Hospital Universitario La Princesa, Madrid, Spain; [7]Hospital General Universitario de Valencia, Valencia, Spain; [8]Hospital Universitario Fundación Alcorcón, Madrid, Spain; [9]Hospital Universitario Virgen Macarena, Sevilla, Spain; [10]Hospital Universitario 12 de Octubre, Madrid, Spain; [11]Hospital Universitario Virgen de la Victoria, Málaga, Spain; [12]Hospital Universitario Infanta Leonor, Madrid, Spain; [13]Spanish National DNA Bank, Universidad de Salamanca, Salamanca, Spain; [14]Life Sciences, Barcelona Supercomputing Centre, Barcelona, Spain; [15]Spanish National Genotyping Centre (CeGen), Universitat Pompeu Fabra, Barcelona, Spain; [16]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA; [17]Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Hospital, London, UK; [18]St John's Institute of Dermatology, King's College London, London, UK.

* Correspondence: Adrià Aterido, Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, 08035, Spain. Tel: +34934029082, Fax: +34934034510, e-mail: adria.aterido@vhir.org

Psoriasis is the most prevalent autoimmune disease in the world and is characterized by the immune cell infiltration into the skin and the hyperproliferation of keratinocytes. Psoriasis negatively impacts the patients' quality of life and has a high socioeconomic burden. Psoriasis is a genetically complex disease and, to date, its heritability is only partially explained. There is increasing evidence that the missing heritability could be explained by multiple genetic variants from common biological pathways. The objective of this study was to identify new genetic variation associated with psoriasis using a genome-wide pathway analysis (GWPA). We genotyped 598,258 SNPs in a discovery cohort of 2,281 case-control individuals from Spain. We performed a GWPA using 1,053 reference biological pathways. A total of 14 pathways (P_{FDR}≤2.55e-2) were significantly associated with psoriasis. Using an independent cohort of 7,353 individuals from the UK, we significantly replicated 6 pathways (P_{FDR}≤3.46e-2). We found, for the first time, that psoriasis is genetically associated with retinol metabolism (P=1.84e-4), the transport of inorganic ions and amino acids (P=1.57e-7) and post-translational protein modification (P=1.57e-7). In the latter pathway, *MGAT5* showed a strong network centrality, and its association with psoriasis was further validated in an additional case-control cohort of 3,429 individuals (P<0.05). Additionally, using T cell *in vitro* cultures, we found a significant association between *MGAT5* variation and the activation levels of this key cell type in psoriasis. These findings provide new biological mechanisms in psoriasis and demonstrate that GWPA is a powerful approach to characterize the missing heritability of complex diseases.

Participants

Nom	Cognom	Institució
Luis	Agulló	UVIC
Riccardo	Aiese Cigliano	SEQUENTIA BIOTECH SL
Mar	Albà	ICREA - FUNDACIO IMIM
Joan	Albiol Sala	UAB
Eva M.	Alloza	BIB
Elena	Álvarez de La Campa Crespo	VHIR
Pau	Andrio Balado	BARCELONA SUPERCOMPUTING CENTER
Irantzu	Anzar	SEQUENTIA BIOTECH SL
Adrià	Aterido Ballonga	VHIR
Miquel	Barrabia Villarroya	UB
Didier	Barradas Bautista	BARCELONA SUPERCOMPUTING CENTER
Federica	Battistini	IRB
Vicky	Berez	BITAC
Silvia	Bonas Guarch	BARCELONA SUPERCOMPUTING CENTER
Ferran	Briansó Castilla	VHIR
Raquel	Buj Gómez	IMPPC
Mónica	Cabrera Pasadas	UPF
Juan Manuel	Calvo	UB
David	Carrera	BSC, UPC
Elvira	Carrio Gaspar	IMPPC
Josema	Cavanillas	ATOS
Laia	Codo Tarraubella	BARCELONA SUPERCOMPUTING CENTER
Paula	Cortes	BARCELONA SUPERCOMPUTING CENTER
Alfred	Cortés Closas	ISGLOBAL
Marta	Crespo Petit	UVIC
Pablo	Dans Puiggròs	IRB
Bernat	Del Olmo	GENCARDIO
Sophia	Derdak	CNAG
Jeannine	Diesch	IJC
Jordi	Durban	IBV-CSIC
Marc	Elosua Bayés	UDG
Ivan	Erill	UNIV. OF MARILAND
Paula	Escuer Pifarré	UB
Toni	Espinosa	UAB
Irene	Farabella	CNAG-CRG
Domènec	Farré Marimon	UB
Narcis	Fernandez-fuentes	UPF - GRIB-IMIM
Joan	Ferrer Obiol	UB
Oscar	Flores	GENOMCORE SL
Clàudia	Fontserè Alemany	IBE-UPF
Cristina	Frías López	UB
Leonor	Frias Moya	GENOMCORE SL

Silvia	Galan	CNAG-CRG
Diego	Gallego Perez	UAB
Sònia	Garcia	INSTITUT BOTÀNIC DE BARCELONA - CSIC
Francisco	Garcia	CIPF
Laura	García Gerique	HOSPITAL - FUNDACIÓ SANT JOAN DE DEU
Alicia	Garrido Garcia	HOSPITAL - FUNDACIÓ SANT JOAN DE DEU
Josep Ll.	Gelpi	BARCELONA SUPERCOMPUTING CENTER
Carla	Giner Delgado	UAB
Jèssica	Gómez Garrido	CRG-CNAG
Soledad	Gómez González	HOSPITAL - FUNDACIÓ SANT JOAN DE DEU
Jose Carlos	Gómez Tamayo	UAB
Juan Ramón	Gonzalez	CENTRE DE RECERCA EN EPIDEMIOLOGIA AMBIENTAL
Gloria	Gonzalez Gacio	BITAC MAP SL
Santiago	Gonzalez Rosado	BARCELONA SUPERCOMPUTING CENTER
Mike	Goodstadt	CNAG-CRG
Josep	Gregori i Font	ROCHE DIAGNOSTICS, SANT CUGAT
Roderic	Guigó	CRG
Marta	Guindo Martínez	BARCELONA SUPERCOMPUTING CENTER
Alba	Gutiérrez Sacristán	GRIB
Txema	Heredia Genestar	IBE- UPF
Carles	Hernandez-Ferrer	CENTRE DE RECERCA EN EPIDEMIOLOGIA AMBIENTAL
Tim	Hubbard	KING'S COLLEGE, LONDON
Valentín	Iglesias Mas	UAB
Andrea	Izquierdo Bouldstridge	IBMB-CSIC
Natalia	Jiménez Lozano	ATOS
Brian	Jiménez-García	BARCELONA SUPERCOMPUTING CENTER
Antonio	Julià Cano	VHIR
Hafid	Laayouni	UPF
Steven	Laurie	CNAG-CRG
Cristina	Leal Rodríguez	IMIM-UPF (GRIB)
Carles	Llorca Longares	IBE
Victor	Lopez Ferrando	BARCELONA SUPERCOMPUTING CENTER
Juana	López Redondo	UNIVERSIDAD DE ALMERÍA
Marcos	López Sánchez	CENTRE DE RECERCA EN EPIDEMIOLOGIA AMBIENTAL
Damaris	Lorenzo	STUDENT
Roberto	Malinverni	IJC
María	Maqueda González	UPC
Òscar	Marín Sala	VHIR
Gracia Ester	Martín Garzón	UNIVERSIDAD DE ALMERÍA
Francisco	Martinez Jimenez	CNAG-CRG
Pilar	Martínez Ortigosa	UNIVERSITY OF ALMERIA
Xavier	Martínez Serrano	VALL HEBRON INSTITUT RECERCA
Marc A.	Marti-Renom	CNAG-CRG
Gemma	Mas de Xaxars	UB
Jesús	Matés Ramírez	GENCARDIO
Josep Maria	Mercader Bigas	BARCELONA SUPERCOMPUTING CENTER

Rafael	Montiel	LANGEBIO, CINVESTAV-IPN, MEXICO
Antoni	Moreno	UB
Miguel Ángel	Naranjo Ortiz	CRG
Arcadi	Navarro Cuartiellas	UPF
Anel	Nurtay	CENTRE DE RESERCA MATEMATICA
Sergio	Oller Moreno	INSTITUTE FOR BIOENGINEERING OF CATALONIA
Dorcas J.	Orengo	UB
Modesto	Orozco	IRB, UB
Natàlia	Padilla Sirera	VHIR
Victòria	Pascal	VHIR
Andreu	Paytuví	SEQUENTIA BIOTECH SL
Cinta	Pegueroles Queralt	CRG
Alexandre	Perera Lluna	UPC
Merce	Planas Felix	BARCELONA SUPERCOMPUTING CENTER
Jordà	Polo Bardés	BARCELONA SUPERCOMPUTING CENTER
Marta	Pozuelo Del Río	VHIR
Eva M.	Puerma Rodríguez	UB
Pau	Puigdevall Costa	UPF
Montserrat	Puiggros Maldonado	BARCELONA SUPERCOMPUTING CENTER
Miquel Angel	Pujana	PROCURE, ICO-IDIBELL
Jordi	Pujols Pujol	UAB
Jordi	Quintana	PCB
Sabarinathan	Radhakrishnan	UPF
Dr. Miquel	Ràmia Jesús	MADE OF GENES
Mireia	Ramos	IGTP
Oscar	Reina Garcia	IRB BARCELONA
Dmitry	Repchevsky	BARCELONA SUPERCOMPUTING CENTER
Casandra	Riera	VHIR
Ana	Ripoll	BIB
Elias	Rodriguez Fos	BARCELONA SUPERCOMPUTING CENTER
Bernardo	Rodriguez Martin	BARCELONA SUPERCOMPUTING CENTER
Raquel	Rodríguez-pérez	INSTITUTE FOR BIOENGINEERING OF CATALONIA
Romina	Royo Garrido	BARCELONA SUPERCOMPUTING CENTER
Julio	Rozas	UNIVERSITAT DE BARCELONA
Lara	Rubio Araúna	IBE (CSIC-UPF)
Valentino	Ruggieri	SEQUENTIA BIOTECH SL
Carlos	Ruiz Arenas	CENTRE DE RECERCA EN EPIDEMIOLOGIA AMBIENTAL
Josep	Ruiz Sanchez	IRTA
Friman	Sanchez Castaño	BARCELONA SUPERCOMPUTING CENTER
Jose Francisco	Sanchez Herrero	UB
Alejandro	Sánchez-Gracia	UB
Alex	Sanchez-Pla	UB I VHIR
Walter	Sanseverino	SEQUENTIA BIOTECH SL
Eduardo	Santamaria-Navarro	MIND THE BYTE S.L.
Irene	Sanz	GRIB, IMIM
Gerard	Serra-vidal	UPF-CSIC

Babita	Singh	UPF-PRBB
Marc	Sitges Puy	GENOMCORE S.L.
Paula	Soler Vila	CNAG-CRG
Carolina	Soriano Tarrga	IMIM
Yannick	Spill	CNAG-CRG
Juan Antonio	Subirana	UPC-IMIM
Fran	Supek	CENTRE FOR GENOMIC REGULATION
Raul	Tonda Hernandez	CNAG
Mariona	Torrens Fontanals	UPF
David	Torrents	BARCELONA SUPERCOMPUTING CENTER
Maria	Touceda Suarez	CRG
Albert	Trill	ATOS
Covadonga	Vara González	UAB
Gonzalo	Vera	CRAG
Maria	Vila Casadesús	IDIBAPS
Joel	Vizueta	UB
Guillem	Ylla Bou	CSIC-UPF
Sanja	Zivanovic	IRB BARCELONA