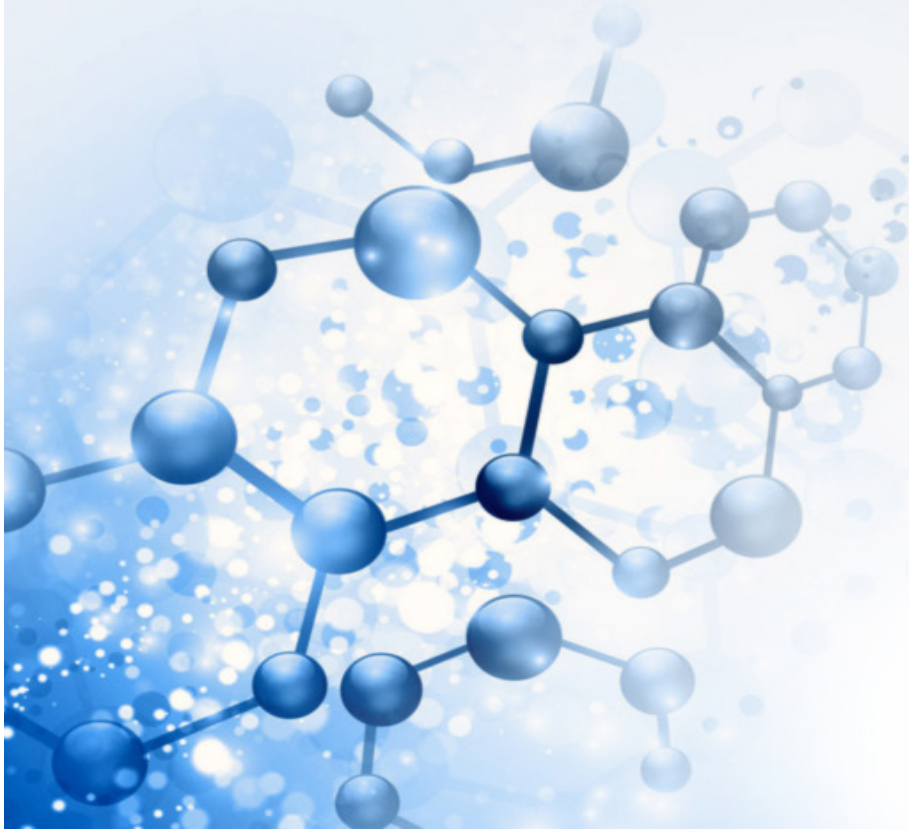
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

JBI2016

XIII Symposium on Bioinformatics

May, 10th -13th 2016



Valencia - Spain

INB Spanish National Bioinformatics Institute

Main sponsor: FUJITSU intel

Sponsors:

FEDER Fondo Europeo de Desarrollo Regional "Una manera de hacer Europa" UNIÓN EUROPEA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

PRB2

etsinf Escola Tècnica Superior d'Enginyeria Informàtica

ESCUELA TÉCNICA SUPERIOR INGENIEROS INDUSTRIALES VALENCIA

ITACA

IUMPA Instituto Universitario de Matemática Pura y Aplicada

biobam BIOINFORMATICS SOLUTIONS

SEQUENTIA

elixir

Affiliated Conference:

ISCB International Society for Computational Biology

Collaborators:

acm

BIB BIOINFORMATICS

#047

Functional Meta-Analysis for Genomic Studies

Francisco García-García[1], Iván Ansari Toledano[2], Cristina Escribano[1], Joaquin Dopazo[3], and David Montaner[4]

[1]*Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain*
[2]*Departamento de Informática, ETSE, Universidad de Valencia, Valencia, Spain*
[3]*Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF). Bioinformatics of Rare Diseases (BIER), Ciber de Enfermedades Raras (CIBERER). Functional Genomics Node, (INB) at CIPF, Valencia, Spain.*
[4]*Genomics England, London, United Kingdom.*

Introduction. Computational methods play a key role in the resolution of clinical and biological problems. Generation of large amounts of data from high-throughput technologies and increasing information accessible in biological databases has boosted the demand for new methodologies able to link both. Functional enrichment analysis of genomic data provides outcomes that are nowadays an integral part of the results of the experiment. However, the small sample size of most of the experiments and their connection to a specific scenario, represent limiting factors when evaluating such studies. Therefore, to improve the integration of various experiments in the functional context and provide clarity in the interpretation, we present a meta-analysis method to detect functional results of global interest, reducing experiment-specific context effects.

Data. We used two different genomic datasets to evaluate this methodology. From Gene Expression Omnibus (GEO) we selected 26 microarrays studies related to psoriasis and dermatitis. In the initial screening, a case-control experimental design for human was required, where cases showed skin lesion and controls were free of injury. On other hand, we downloaded a second dataset with 20 microRNA data from The Cancer Genome Atlas project, which contain both tumoral and healthy samples.

Methods. After preprocessing these data, differential expression analysis from Babelomics [1] and enrichment analysis using logistic models [2][3] were carried out for each study. Biological information from Reactome, KEGG pathways and Gene Ontology databases was used. Odds ratios for each functional term were combined to detect a global association between experimental groups and studies. Variability for each study was estimated from several methods for fixed and random effects (DerSimonian-Laird, Hedges, Hunter-Schmidt, Sidik-Jonkman, ...)

#046

PanelMaps: a web tool for detection and visualization of altered regions for targeted sequencing

José M. Juanes[1], Francisco García-García[2], Joaquin Dopazo[3], and Vicente Arnau[1]

[1]*Departamento de Informática, ETSE, Universidad de Valencia, Valencia, Spain*
[2]*Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia Spain*
[3]*Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF). Bioinformatics of Rare Diseases (BIER), Ciber de Enfermedades Raras (CIBERER). Functional Genomics Node, (INB) at CIPF, Valencia, Spain*

Introduction. Gene panel sequencing allow us to detect variants associated with different diseases. Sometimes, the possible cause of the disease is not due to variations in SNPs (Single Nucleotide Polymorphism) or INDELs (small insertions-deletions) and can be motivated by the presence of a larger variation: deletion or insertion.
The aim of this work is the design and development of a web tool for detection and visualization of altered regions from targeted sequencing data. How does PanelMaps work?

Inputs. One or several BAM files (one file for each individual) and a BED file including all regions for gene panel. After loading data, the coverage of each sample is calculated and these values are normalized between all samples considering the total number of reads of each sample.

Methods. The tool includes two modules: visualization of genes or regions from coverage data and a module analysis to detect regions of interest. for detection of altered regions, PanelMaps uses a control sample selected by the user or combines all samples to get a common reference. Users can also specify comparisons between subgroups of samples of interest. The analysis incorporates a sliding window algorithm with various parameters adjustable by the user and related to the precision and characteristics of the region to be detected.

Outputs. PanelMaps shows a graphical description of coverage levels for genes and samples to confirm that all regions are covered. This web tool visualizes all regions included in the panel and shows a selection of altered regions between samples.

Conclusions. Panelmaps is a useful tool for detection and visualization regions of genes altered in panels that improves the knowledge of the genetic basis of diseases and produces useful information for diagnosis in clinical contexts.
This web tool is an alternative to the use of molecular biology techniques such as MLPA (Multiplex Ligation-dependent Probe Amplification), which are very costly and sometimes have some technical problems such as failure to detect variants or micro-deletions in positions where primers are incorporated.

#066

Spanish Population-Specific Differences in Disease-Related Genetic Variation

Alicia Amadoz[1], Marta Bleda[1], Luz Garcia-Alonso[1], Alejandro Alemán[1], Francisco García-García[1], Juan A. Rodríguez[2], Josephine T. Daub[2], Gerard Muntané[2], Antonio Rueda[3], Alicia Vela-Boza[3], Francisco J. López-Domingo[3], Javier P. Florido[3], Pablo Arce[3], Arcadi Navarro[2], Salud Borrego[4], Javier Santoyo-López[5], Guillermo Antiñolo[3], and Joaquín Dopazo[1]

[1]*Centro de Investigación Príncipe Felipe (CIPF). Spain*
[2]*Universitat Pompeu Fabra. Spain*
[3]*Genomics and Bioinformatics Platform of Andalusia (GBPA). Spain*
[4]*Universidad de Sevilla. Spain*
[5]*University of Edinburgh. United Kingdom*

Recent results from large-scale genomic projects suggest that allele frequencies, which are highly relevant for medical purposes, differ considerably across different populations. The need for a detailed catalog of local variability motivated the whole-exome sequencing of 267 unrelated individuals [1], representative of the healthy Spanish population. Like in other studies, a considerable number of rare variants were found (almost one-third of the described variants). There were also relevant differences in allelic frequencies in polymorphic variants, including ~10,000 polymorphisms private to the Spanish population. The allelic frequencies of variants conferring susceptibility to complex diseases (including cancer, schizophrenia, Alzheimer disease, type 2 diabetes, and other pathologies) were overall similar to those of other populations. However, the trend is the opposite for variants linked to Mendelian and rare diseases (including several retinal degenerative dystrophies and cardiomyopathies) that show marked frequency differences between populations. This observation agrees with the fact that, while high-frequency variants and variants underlying complex diseases tend to be shared across populations [2], low-frequency alleles tend to be private [3]. Moreover, we observed a total of 121 variants affecting the binding sites of different drugs, suggesting an important role for local variability in population-specific drug resistances or adverse effects. In addition, our findings highlight the relevance of local variability to distinguish real disease associations from population-specific polymorphisms [1]. We have made available the Spanish population variant server that contains population frequency information for the complete list of 199,669 variant positions we found in the 267 healthy individuals (http://csvs.babelomics.org/).

[1] Dopazo et al. Molecular Biology and Evolution In press (2016)
doi:10.1093/molbev/msw005
[2] Marigorta et al. PLoS Genetics 9(6) (2013) e1003566.
doi:10.1371/journal.pgen.1003566
[3] Casals et al. PLoS Genetics 9(9) (2013) e1003815.
doi:10.1371/journal.pgen.1003815

#070

BiER collaborative projects

Francisco García-García[1], Alejandro Alemán[1], Francisco Salavert Torres[1], Mercedes Medina García[1], Julen Mendieta Esteban[2], and Joaquin Dopazo[1]

[1]*Computational Genomics. Príncipe Felipe Research Center. Spain*
[2]*Departamento de Informática, ETSE, Universidad de Valencia, Valencia, Spain*

BIER (Bioinformatics Platform for Rare Diseases) is a transversal working group whose mission is to provide bioinformatics and technological support to experimental CIBERER units.

During the last years the BiER platform has maintained an intense collaborative relationship with more than 25 CIBERER groups, mainly within the context of intramural sequencing projects, addressing transcriptomic and genomic studies (exomes and panels of genes). BiER has provided advice as well as technological and bioinformatics support in 70 projects of groups belonging to different CIBERER programs: Medical Genetics, Hereditary Metabolic Medicine, Endocrine Medicine, Pathology Neurosensorial, Neuromuscular and Mitochondrial Medicine and Related Syndromes Hereditary Cancer. We have also worked on the development of new methods of transcriptome analysis in the context of signaling pathways, functional meta-analysis and functional enrichment analysis for microRNAs.

We actively participate in collaboration intra-groups receiving more than 50 researchers in our unit and organizing the training activity "NGS course: from reads to candidate genes" which has been held during the last 4 years with an average of 25 attendants from different groups CIBERER per edition.

BiER has provided support and development of new versions of web tools for processing and analysis of genomic data: Babelomics (http://www.babelomics.org/) BiERapp (http://bierapp.babelomics.org/), TEAM (http://team.babelomics.org/), CIBERER Spanish Variant Server (http://csvs.babelomics.org/), among others.

The results of these analysis and bioinformatics developments have contributed to the discovery of 13 new disease genes in which 27 new mutations were identified and the identification of 36 new causal mutations in known disease genes, and also generated 56 collaborative scientific publications in the last three years
(http://bioinfo.cipf.es/publications) [1] [2] [3].

[1] Global Transcriptome Analysis of Primary Cerebrocortical Cells: Identification of Genes Regulated by Triiodothyronine in Specific Cell Types. Gil-Ibañez P et al. Cereb. Cortex. 2015.

#084

Web tools for the analysis of genomic data and the discovery new disease genes.

Alejandro Alemán[1], Fransico Salavert-Torres[1], Mercedes Medina[1], Francisco García-García[1], Jose Carbonell[1], Marta Hidalgo[1], Alicia Amadoz[1], Cankut Çubuk[1], Asunción Gallego[1], and Joaquín Dopazo[1]

[1]*Centro de Investigación Príncipe Felipe (CIPF). Spain*

The continuously increasing data production capability of sequencing technologies has shifted the bottleneck of the discovery process from the production to the data analysis phase. Our contribution to bridge the gap between genomic data production and its biological interpretation consists on the generation of a set of web-based tools used in different large-scale projects (MGP, CIBERER etc.). These include tools for gene prioritization, as BiERapp [1] (http://bierapp.babelomics.org/), which only during the last year was used for the analysis of more than 1000 exomes of patients of more than 70 different inherited pathologies, or TEAM [2] (http://team.babelomics.org/), designed for the efficient management of NGS targeted sequencing data for diagnostic. The CIBERER Spanish Variant Server (http://csvs.babelomics.org) is a public resource that provides information about the variability of the Spanish. We are in our fifth version of Babelomics [3] (http://babelomics.org), a general purpose platform for the analysis of Transcriptomics, Proteomics and Genomics data with advanced functional profiling, with more than 2000 registered users and about 2000 analysis carried out per month. and we are proud to present our new generation of precision medicine web tools that allow exploring disease mechanisms in the context of signaling pathways, hiPathia (http://hipathia.babelomics.org/), and the PathAct (http://pathact.babelomics.org/), an interactive framework to study of the consequences that KOs or over-expressions. More than 40,000 analyses were carried out in our tools during 2015.

[1] Alemán et al. Nucleic acids research, (2014), p. gku407.
[2] Alemán et al. Nucleic acids research, (2014), vol. 42, no W1, p. W83-W87.
[3] Alonso et al. Nucleic acids research, (2015), vol. 43, no W1, p. W117-W121.