Teaching how to do statistical analysis to prioritize genes or mutations for diseases using web tools

Francisco García-García[1], Alejandro Alemán[2] y Joaquín Dopazo[3]

1fgarcia@cipf.es [2]aaleman@cipf.es [3]jdopazo@cipf.es. Centro de Investigación Príncipe Felipe

Abstract

Recent advances in genome sequencing technologies have enabled a breakthrough in the investigation of the genetic basis of disease. However, the resolving power of these technologies has shifted the bottleneck of the discovery process from the production to the data analysis phase.

In order to help to circumvent this bottleneck, we have generated several web tools for the study of variations in genetic diseases. This is an appropiate resource to teach Statistics for this specific group of professionals:

1. BiERapp (http://bierapp.babelomics.org/)

Whole-exome sequencing has become a fundamental tool for the discovery of disease-related genes of familial diseases and the identification of somatic driver variants in cancer. However, finding the causal mutation among the enormous background of individual variability in a small number of samples is still a big challenge. BiERapp is a web-based interactive framework designed to assist in the prioritization of disease candidate genes in whole exome sequencing studies.

2. Spanish Population Variant Server (http://bioinfo.cipf.es/apps-beta/spvs/beta/)

This tool was created with the aim of providing the scientific and medical community with information about the variability of the Spanish population. It is useful for filtering polymorphisms and local variants.

3. TEAM (http://team.babelomics.org/)

A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications.

Key words: Genomics, statistical analysis, web tools, Biotechnology

En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), Didáctica de la Estadística, Probabilidad y Combinatoria, 2 (pp. 471-472). Granada, 2015.

Web tools to analyze and prioritize genes or mutations for diseases

Francisco García-García[1], Alejandro Alemán[1,2], Joaquín Dopazo[1,2,3]

1 Systems Genomics Lab, Computational Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain.
2 BIER, CIBER de Enfermedades Raras (CIBERER). 3 Functional Genomics Node (INB)

Aim

- Recent advances in genome sequencing technologies have enabled a breakthrough in the investigation of the genetic basis of disease. However, the resolving power of these technologies has shifted the bottleneck of the discovery process from the production to the **statistical data analysis** phase.
- Our team (**BIER**: Bioinformatics Platform for Rare Diseases; http://www.ciberer.es/bier) has developed several we tools aimed at different kind of statistical analysis: **discovery of new variants, disease diagnosis and visualization of genomics results**.

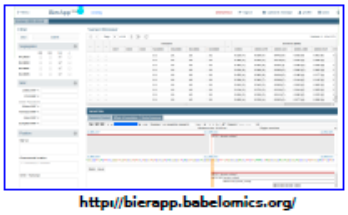
Methods

- BiER has designed pipelines for Genomics and Transcriptomics sequencing data analysis.
- These web applications makes an intensive use of new web technologies and standards like HTML5. R (free software environment for statistical computing) was used for analysis modules.
- Several training activities were carried out to facilitate the understanding and management of data.

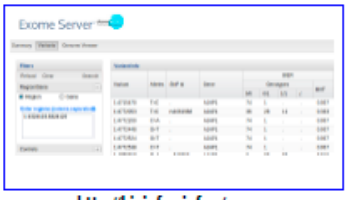
Results

- Scientific collaborations took place among **19 CIBERER research CIBERER.**
- **Recent publications** include the discovery of two new mutations in the BCKDK gene, responsible of a neurobehavioral deficit in pediatric patients (1), new mutations in different genes causing inherited retinal dystrophies (2) and metabolic diseases (3).
- These **web tools** were generated to analyze and improve the management of results:

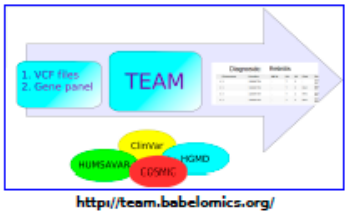
- **BiERapp** (4). A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome secuencing studies.
- **ExomeServer**. Created with the intention to provide the scientific and medical community, information about the variability in the Spanish population. It is useful for filtering polymorphisms and local variants.
- **TEAM** (5). A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications.
- **BABELOMICS** (6). Gene expression, genome variation and functional profiling analysis suite (http://babelomics.bioinfo.cipf.es/).



http://bierapp.babelomics.org/



http://bioinfo.cipf.es/apps-beta/exome-server/beta/



http://team.babelomics.org/

Conclusions

- Interaction between research groups and BIER platform has been an important factor in web design and adjustment tools for analyzing sequencing data and its interpretation.
- The results obtained from the analyzes have provided a better understanding of the genomic data of these diseases, as well as the detection of biomarkers that can be used in the prevention, diagnosis and clinical therapy design.
- The use of web tools improved the skills of researchers in the statistical analysis of genomic data.

References

(1) García-Cazorla A., Oyarzabal A., Fort J., Robles C., Castejon E., Ruiz-Sala P., Bodoy S., Merinero B., Lopez-Sala A., Dopazo J., et al. Two Novel Mutations in the BCKDK Gene (Branched-Chain Keto-Acid Dehydrogenase Kinase) are Responsible of a Neurobehavioral Deficit in two Pediatric Unrelated Patients. Hum. Mutat 2014.
(2) de Castro-Miro M., Pomares E., Lorso-Motta L., Tonda R., Dopazo J., Marfany G., Gonzalez-Duarte R. Combined genetic and high-throughput strategies for molecular diagnosis of inherited retinal dystrophies. PLoS ONE 2014;9:e88410.
(3) Ter F., García-Silva M.T., Ferrer-Cortes X., Navarro-Sastre A., García-Villoria J., Coll M.J., Vidal E., Jimenez-Almazan J., Dopazo J., Briones P., et al. Exome sequencing identifies a new mutation in SERAC1 in a patient with 3-methylglutaconic aciduria. Mol. Genet. Metab. 2013;110:73-77.
(4) Aleman, A; García-García, F; Medina, I; Dopazo, J. A web tool for the design and management of panels of genes for targeted enrichment and massive sequencing for clinical applications. Nucleic Acids Res. 2014 May 26. pii: gku347
(5) Aleman, A; García-García, F; Salavert, F; Medina, I; Dopazo, J. A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies. Nucleic Acids Research. 2014 May 6. PMID: 24803668.
(6) Medina et all. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. NAR-00481-Web-B-2010.R1 .