

Bioinformatics for Integrative Genomics

Exploring and making sense of multidimensional data sets that result from merging heterogeneous genomic data resources

In the last few years, the study of the genetic basis of complex diseases has undergone a profound transformation. High-throughput techniques, such as microarrays and next generation sequencing technologies, allow scientists to identify alterations at genome-wide level in patient's genomes and transcriptomes. The application of these technologies yields complex, multidimensional genomic data sets that describe in detail the myriad changes that occur within individual cells and how these changes differ between patients, cells or conditions. Usually we want to use this data to answer some specific questions. For instance, which of the long list of mutations detected are likely to affect the function of the protein and which are probably neutral? Which of the genes affected by those mutations are already known to be involved in cancer or other diseases? Which pathways or biological processes are affected by the transcriptomic alterations detected in my experiment? Etc.

We need to apply bioinformatic approaches to answer these questions. The course will provide the students with theoretical and practical expertise in the field of bioinformatics for integrative genomics. Attendees will be updated on the state of the art in the field and they will become capable of handling and analyzing genomics data. Upon completion, the attendees will be able to perform integrative analysis and interpret the results, in order to answer specific biological questions.

To achieve these objectives we will use three tools developed in our lab:



Is a framework for analysis and visualization of genomic data. Data and results are represented as interactive heatmaps. Gitools allows importing data directly from IntOGen, Biomart, KEGG and Gene Ontology. It performs several analyses like enrichment analysis, correlations, overlaps, identification of driver alterations and combination of experiments.

Reference: Christian Perez-Llamas and Lopez-Bigas N. *Gitools: analysis and visualization of genomic data using interactive heat-maps. PLoS One*, 6,e19541 (2011).

<http://www.gitools.org>



Integrative OncoGenomics. A discovery tool for cancer researchers, is a resource that integrates multidimensional OncoGenomics Data for the identification of genes and groups of genes (biological modules) involved in cancer development.

Reference: Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, Furney S and Lopez-Bigas N. *IntOGen: Integration and data-mining of multidimensional oncogenomic data. Nature Methods*, 7, 92-93 (2010)

<http://www.intogen.org>



Is a method to integrate the output of various methods to identify deleterious non-synonymous variants, which we have shown to perform significantly better than any of the methods alone. Condel outputs a score for each SNVs that can be used as an indicator of the impact of the mutation on the functionality of the mutant protein.

Reference: Abel Gonzalez-Perez and Lopez-Bigas N. *Improving the assessment of the outcome of non-synonymous SNVs with a Consensus deleteriousness score (Condel). Am J Hum Genet*, 88(4):440-9 (2011).

<http://bg.upf.edu/condel>

➤ **Instructors:**

- **Nuria Lopez-Bigas** is the head of the Biomedical Genomics group at the Universitat Pompeu Fabra (UPF) in Barcelona. She completed her PhD degree on 2002 on the molecular basis of deafness at the Oncologic Research Institute in Barcelona. She then moved to the European Bioinformatics Institute in Hinxton (Cambridge, UK) to initiate a project on the computational study of disease and cancer genes. Since 2006 is a group leader at the UPF focusing on cancer genomics and bioinformatics. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
- **Sophia Derdak** obtained her PhD in immunology from the Medical University of Vienna, Austria, in 2003. After a postdoctoral project at the German Cancer Research Center (Heidelberg, Germany), she moved to Barcelona to work in cancer biomarker discovery at Oryzon Genomics before joining the Biomedical Genomics Group in 2010. She is interested in integrative Omics, personalized genomics and the analysis of Next Generation Sequencing data.
- **Alba Jené-Sanz** is a Biologist from the Universitat Pompeu Fabra. She has a Master in Bioinformatics from the same university and she is currently pursuing a PhD at the Biomedical Genomics group (<http://bg.upf.edu>) of the University Pompeu Fabra. Her main interests are the study of regulatory pathways in cancer from an epigenomics point of view, and the use of sequencing technologies for large-scale cancer genomics studies.

➤ **Target Audience:** The course is directed to researchers in biology and medicine.

➤ **Course Pre-requisites:** Basic Molecular Biology. Elementary computing skills. Background knowledge in bioinformatics is not necessary.

➤ **Course URL:** <http://bg.upf.edu/course-big>

➤ **Detailed Program**

Day 1 (June 21st)	
9:30 – 11:00	Introduction to integrative genomics
11:00 – 11:30	Coffee Break
11:30 – 12:30	Exploring IntOGen data We will answer questions like: Is my favourite gene altered in any type of cancer? Which? How? Which genes are more significantly up-regulated in breast cancer? Which pathways are affected in lung cancer?...
12:30 – 14:00	Lunch Break
14:00 – 16:00	IntOGen and Gitoools Exploring the similarities and differences between the alteration patterns of different tumour types at the level of genes and pathways. Pathway enrichment analysis.
16:00 – 16:30	Tea Break
16:30 – 18:00	Basic and intuitive analysis of microarrays using Gitoools Enrichment analysis of pathways and Gene Ontology terms.

Day 2 (June 22nd)	
9:30 – 11:00	Basic and intuitive analysis of microarrays using Gitoools Statistical combination of information from heterogeneous data sources. Identifying cancer driver alterations. Importing data from IntOGen to compare with our own data.
11:00 – 11:30	Coffee Break
11:30 – 12:30	Basic and intuitive analysis of microarrays using Gitoools Correlation and clustering of expression patterns among samples, cells types or conditions.
12:30 – 14:00	Lunch Break
14:00 – 16:00	Identifying functional single nucleotide variants Using Condel to identify from a long list of non-synonymous variants, which are the ones that are more likely to affect the function of the protein.
16:00 – 16:30	Tea Break
16:30 – 18:00	Summary, questions and wrap up