BioSys-PhD
Biological Systems - Functional & Integrative Genomics

BioFIG
Center for Biodiversity, Functional and Integrative Genomics

BioSys/BioFIG Research Seminar

Exploratory data analysis: a crucial step towards transforming complexity of NGS data into biological evidence

Andreia J. Amaral
BioFIG / FCUL

Genomics is defined as the systematic study on a whole-genome scale for the identification of genetic contributions to a phenotype of interest. In recent years the development of next generation sequencing (NGS) has enabled a boom in the genomics field that has unraveled important biological insights with impact in biomedicine, agriculture and even in the research of alternative fuels. There is a wide range of experimental designs that can be supported by NGS, whole or partial genome studies, metagenome and transcriptome studies, protein-DNA interactions and nucleosome positioning. Although the analysis of NGS data of these different experimental designs involves specific steps, all involve a crucial step, exploratory data analysis for assessment of data quality and selection of good quality reads in order to efficiently identify genomic variants or differential gene expression. NGS can be affected by a range of artifacts that arise during the library preparation and sequencing processes, which can negatively impact the quality of the raw data for downstream analyses. These issues include platform specific error profiles, systematic variation in quality scores across the sequence read, biases in sequence generation driven by base composition, departure from optimal library fragment sizes and variation in the proportions of duplicate sequences introduced by PCR amplification. In this seminar I will show how assessment of NGS data quality was crucial to efficiently identify variants in the porcine genome for the development of porcine 60K SNP chip project and generate the first map of selective signatures of this species, as well as to support a robust differential gene expression analysis using RNA-seq libraries in an ongoing project in *D. melanogaster*.

Date: 16 May 2014 – 11h30

Faculdade de Ciências da UL, Lisboa (Portugal), Building C6 – room 6.2.53

PD+F FCT PhD PROGRAMMES

U LISBOA | UNIVERSIDADE DE LISBOA