Megasequencing Data Analysis: from reads to candidate genes

Dates: 10, 11 and 12 December 2012

Day 1

11:00 – 11:30	**Course presentation.** Joaquin Dopazo.
11:30 – 13:00	**Introduction to NGS technologies.** Javier Santoyo
13:00 – 13:30	**Computing infrastructure for NGS analyses.** Pablo Escobar.
13:30 – 15:00	*Lunch break*
15:00 – 16:30	**NGS data preprocessing concepts.** Jose Carbonell. Handling sequence files. Fasta and fastq file formats overview. Encoding quality metrics. Quality control tools. Bias detection. Sequence filtering.
16:30 – 17:00	*Coffee break*
17:00 – 18:30	**NGS data preprocessing practical session.** Jose Carbonell and Francisco Garcia. Hands on session with FastQC and FastX-Toolkit programs.

Day 2

10:00 – 11:30

Sequence alignment concepts.
Enrique Vidal.

Reference genome as a concept.
NGS alignment programs and algorithms.
SAM/BAM and BED
file formats overview.
Mapped reads visualization.
Quality control of the mapping.
Local realignment.
Computational needs.

11:30 – 12:00 *Coffee break*

12:00 – 13:30 **Sequence alignment practical session.**
Enrique Vidal and David Montaner.

Hands on session with alignment programs: bowtie, bwa and bfast
.
Overview of quality control and visualization tools: bamQC, GATK and IGV
.
Data handling with SAMtools.

13:30 – 15:00 *Lunch break*

15:00 – 16:30 **Variant calling in NGS experiments; concepts.**
Jorge Jimenez.

Variant types: SNPs and InDels.

Calling procedures and algorithms.
VCF file formats overview.
Introduction to the programs GATK, Annovar and Variant.
Variant filtering.
The problem of the missing data.
Variant annotation.
Data bases and data repositories: dbSNP and the 1000 genomes project.

16:30 – 17:00 *Coffee break*

17:00 – 18:30 **Variant calling practical session.**
Jorge Jimenez and Martina Marba.
Hands on with GATK, Annovar, Variant and VCFtools.

Day 3

9:30 – 11:30	**Results interpretation.** Jorge Jimenez and Enrique Vidal A general view on how to interpret BIER analysis pipeline and results. Some remarks on experiment design.
11:30 – 12:00	*Coffee break*
12:00 – 14:00	**Some CIPF tools for Gene prioritization.** Luz Garcia and Marta Bleda SNOW, NetworkMiner, Variant, CellBase and Genome Maps