

Megasequencing Data Management: from reads to candidate genes.

Day 1

10:30 – 10:30

Course presentation.

Joaquin Dopazo.

10:30 – 11:30

Introduction to NGS technologies.

Javier Santoyo

11:30 – 12:00

Computing infrastructure for NGS analyses.

Pablo Escobar.

12:00 – 12:30

Coffee break

12:30 – 14:00

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.

14:00 – 15:30

Lunch break

15:30 – 17:00

NGS data preprocessing practical session.

Jose Carbonell and Francisco Garcia.

Hands on session with FastQC and FastX-Toolkit programs.

17:00 – 17:30

Coffee break

17:30 – 19:00

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.

Day 2

9:00 – 10:30

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.

10:30 – 11:00

Coffee break

11:00 – 12:30

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.

12:30 – 13:00

Coffee break

13:00 – 14:30

Variant calling practical session.

Jorge Jimenez and Martina Marba.

Hands on with GATK, Annovar, Variant and VCFtools.

14:30 – 16:00

Lunch break

16:00 – 17:30

Results interpretation.

Jorge Jimenez and Enrique Vidal

A general view on how to interpret BIER analysis pipeline and results. Some remarks on experiment design.

17:30 – 18:00

Coffee break

18:00 – 19:00

Some CIPF tools for Gene prioritization.

Luz Garcia and Marta Bleda

SNOW, NetworkMiner, Variant, CellBase and Genome Maps