Rafael Irizarry

1) Statistical Issues in the analysis of Microarray Data (3 horas)

Microarrays are an example of powerful high through-put genomics tools that are revolutionizing the measurement of biological systems. In this and other technologies, a number of critical steps are required to convert the raw measures into the results relied upon by biologists and clinicians. These data manipulation have enormous influence on the quality of the ultimate measurements and studies that rely upon them.

Microarrays are now being used to measure diverse high genomic endpoints including genotype, chromosomal abnormalities including deletions/insertions, protein binding sites, methylation, and alternative splicing. In each case, the genomic units of measurement are short oligonucleotides referred to as probes. Without appropriate understanding of the bias and variance of these measurements, biological inferences based upon probe analysis will be compromised.

In these lectures we will review some basic concepts of molecular biology and genetics, describe the technology, and review some of the more influential statistical approaches.

2) Statistical and Computational Issues in the analysis of Second Generation Sequencing Data

Second-generation sequencing (sec-gen) technology can sequence millions of short fragments of DNA in parallel, making it capable of assembling complex genomes for a small fraction of the price and time of previous technologies. In fact, a recently formed international consortium, the 1000 Genomes Project, plans to fully sequence the genomes of approximately 1200 people. The prospect of comparative analysis at the sequence level of a large number of samples across multiple populations may be achieved within the next five years. These data present unprecedented challenges in statistical analysis. In this lecture we will review some of these issues