CING BIOINFORMATICS WORKSHOP #1

On

Next Generation Sequencing and Human Variation Data Analysis

Venue: The Cyprus Institute of Neurology and Genetics, Room 104

Date: April 3, Monday, 9:00-13:00

Potential Attendees: CING Scientists and CSMM Students

Registration: Free

Certificate of Attendance: It will be given to the registered attendees after the end of the Workshop**.**

Lecturers: Anastasis Oulas (CING Bioinformatics Group), Athina Theodosiou (CING Cytogenetics Department), Kyriaki Michailidou (CING Electron Microscopy and Molecular Pathology Department), Elena Loizidou (CING Bioinformatics Group)

Program

9:00-9:30 Welcome Session and Introduction. Next Generation Sequencing and Human Variation (Anastasis Oulas)

In this workshop we will elaborate on the current state-of-the-art practices, tools and methodologies for analysing data from large scale, high throughput experiments such as whole exome or whole genome sequencing (WES, WGS). It is directed to students and researcher who are interested in such methodologies and would like to gain the most from their data.

9.30-10:15 Lecture 1. NGS Data Pre-processing: From FASTQ to BAM files (Athina Theodosiou)

This lecture will focus on the initial steps performed in NGS pipelines for genomics analysis. Firstly, we will introduce the file formats that are generated i) from next generation sequencing (fastq files) and ii) from the alignment (sam/bam files). Next, we will present the quality control that is necessary to be done on the raw data and move on to describe the alignment procedure including any necessary preprocessing step. Moreover, we will present the statistical measurements that can be estimated in order to validate a sequencing run from the bam files. Last, we will demonstrate the use of IGV tool in order to visualize mapped reads on the genome.

10:15-11:00 Lecture 2. Variant (SNP & Indel) Discovery stage of the NGS pipeline analysis (Kyriaki Michailidou)

This is the stage where the study samples are compared to the human reference genome in order to identify sites that display variation. This is a critical part of the pipeline since we need to ensure that actual genetic variation is distinguished from sequencing artefacts. We will focus on the best practices from the GATK (broad institute) for germline SNP & Indel discovery in whole genome and exome sequencing.

COFFEE BREAK (11:00-11:30)

11:30-12:15 Lecture 3. Copy Number Variation (CNV) Discovery using NGS data (Elena Loizidou)

This lecture will emphasize specifically on the tools and methodologies available for detecting copy number variations (duplications and deletions) from WGS and WES data. This methodology relies on comparing samples with a reference panel in order to discover copy number variations that play an important role in generating necessary variation in the population as well as disease phenotype.

12:15-13:00 Lecture 4. Annotation, Variant Refinement and Filtering using Current State-of-the-art tools and Databases (Anastasis Oulas)

This talk will address issues of variant annotation using tools like Variant effect predictor (VEP) and GEnome MINIng (Gemini). These are flexible frameworks for exploring genetic variation in the context of the wealth of genome annotations available for the human genome in online databases (i.e. ENCODE tracks, UCSC tracks, OMIM, dbSNP, KEGG, HPRD, ExAC, ESP and 1000GP, ClinVar, COSMIC) They provide a simple, flexible, and powerful system for exploring genetic variation for disease and population genetics.