

Bioinformatics Basics

Course Leader: Dr. Paul C. Boutros (Ontario Institute of Cancer Research, U. Toronto)

Other Instructors: Fouad Yousif (OICR-U.Toronto), Julie Livingstone (OICR-U.Toronto)

Dates: Monday September 15th - Friday September 19th 2014

Structure: Each day will be divided into two halves. Each half will have a 90-120 minutes of lecture material followed by 90-120 minutes of practical, on-computer data-analysis. We will cover both the underlying characteristics of each type of data, as well as the key considerations in data-analysis. The software sessions will use freely-available open-source software, with an emphasis on the R statistical environment.

Course Outline:

Day One, Monday 15th: Introduction to Genomics, R and Linux

- Lecture #1: What is Genomics (~1 hour)
- Mixed #1: Introduction to linux/clusters (1-2 hours)
- Lecture #2: R intro (~2 hours)

Day Two, Tuesday 16th: Microarray

- Practical #1: Extended R example (~1-2 hours)
- Lecture #4: Microarrays 101 (~90 minutes)
- Practical #2: Using R to read microarray data (~2 hours)
- Lecture #5: Advanced Microarrays, including multiple-testing (~90-120 minutes)
- Practical #3: QC & Pre-Processing of Microarray Data

Day Three, Wednesday 17th: Next-Gen Sequencing

- Lecture: Intro to NGS - what is it, why we use it (~30 minutes)
- Lecture: What happens in lab - steps (what are adaptors, what is a flowcell) and where bias occurs (~1-2 hours)
- Mixed: what data we get of sequencer (~30 minutes)
- Mixed: Alignment - repeat regions, how alignment works, how PE helps, bwa index, running bwa, SAM/BAM format (~90 minutes)

- Mixed: converting file to BAM, how to view BAM files, coordinate sorting, collapsing, samtools, IGV (~2-3 hours)

Day Four, Thursday 18th: SNV Calling and Annotation

- Dataset: Human chr22 from 1k genomes
- Lecture 1: ~1 hr. Introduction to SNV calling
- Lab 1: ~1 hr. Demo of IGV. Manually call SNVs
- Lecture 2: ~1 hr Base calling algorithms and VCF format.
- Lab 2: ~2.5hrs. Generate calls using samtools, do some R visualization and filtering.
- Lecture 3: ~30 minutes. Annotation and Annovar
- Lab 3: ~1hr+ Generate annotations for our filtered VCF calls we generated earlier.

Day Five, Friday 19th: Machine Learning and Visualization

- Lecture #1: Why should you care (3 x 20-min vignettes = 1 hour)
- Lecture #2: Machine-Learning 101 (~90 minutes)
- Practical #1: Sample ML (~2 hours)
- Lecture #3: Data Visualization Basics (~60 minutes)
- Catch Up & Bring Your Own Problems (~2 hours)
- Bringing it All Together: The Heterogeneity Study: (~30-45 minutes)