Epigenomics: ChIP-seq data analysis

Instructor: Ariel Madrigal Aguirre

Registration link: https://forms.gle/5ftowfB2q6LFZkXW9

Hours of instruction: 4

Prerequisites:

- Access to a terminal in your computer (MacOS and Linux users will have it already)
- Proficiency with command line interface (e.g., know how to navigate between directories, execute programs)
- Knowledge of basic programming
- Install:
    - R (>=4.0.0) and Rstudio
    - R dependencies:
        - ChIPQC (https://bioconductor.org/packages/release/bioc/html/ChIPQC.html)
        - TxDb.Hsapiens.UCSC.hg38.knownGene (https://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg38.knownGene.html)
        - rGREAT(https://bioconductor.org/packages/release/bioc/html/rGREAT.html)
    - Python >=3.6
    - Bowtie2 (https://bowtie-bio.sourceforge.net/bowtie2/index.shtml)
    - Samtools (http://www.htslib.org/)
    - MACS2 (https://pypi.org/project/MACS2/)
    - Deeptools (https://deeptools.readthedocs.io/en/develop/#)
    - Bedtools (https://bedtools.readthedocs.io/en/latest/)

*Students that register will receive a guide to help set up requirements (through Docker or manual installation) for the workshop beforehand.

*There will be a pre-workshop session online to solve software installation issues (Date/time to be confirmed).

Summary:

ChIP-seq is a method to identify genome-wide DNA binding sites for a protein of interest. This worskhop will guide you through the basics of ChIP-seq analysis with hands-on exercises. Worskhop participants will learn how to process ChIP-seq data: perform read alignment, peak calling, visualization through the genome browser, motif finding and gene set enrichment analysis.

Contents:

I. **Module 1: Introduction (20 min)**
   a. Outline (2 min)
   b. Intro to ChIP-seq (18min)
II. **Module 2: Alignment and identification of binding sites (1h 10 min)**
   a. Alignment (5 min)
   b. Peak calling (10 min)
   c. Peak overlap (5 min)
   d. Hands-on (50 min)

10 min BREAK

III. **Module 3: Quality control (15 min)**
   a. Common metrics for QC (15 min)
IV. **Module 4: Visualization (45 min)**
   a. Generation of bigwig tracks (5 min)
   b. UCSC genome browser (5 min)
   c. Deeptools (5 min)
   d. Hands-on (30 min)

10 min BREAK

V. **Module 5: Motif finding and gene set enrichment analysis (45 min)**
   a. Motif enrichment (10 min)
   b. Gene set enrichment analysis (10 min)
   c. Hands-on (25 min)
VI. **Concluding remarks (10 min)**