Medhanie A. Mulaw (Jun.-Prof. Dr.)
medhanie.mulaw@uni-ulm.de

Basics in bioinformatics and functional genomics using R and Bioconductor

Scientific data analysis and presentation play a pivotal role in biological sciences and molecular medicine. Clinical decisions, including patient diagnosis and prognosis, are based on proper quantitative and qualitative analysis of various molecular and histological data. Integrated analysis of Large-scale whole-genome, transcriptome, and epigenome datasets are providing novel insights into disease initiation, progression, and treatment in clinical settings. Additionally, functional studies using animal models and *in vitro* experiments have seen a paradigm shift with the advent of next-generation sequencing.

This course will introduce participants to methods of data retrieval and preprocessing using R and Bioconductor. Participants will use published and in-house datasets and familiarize themselves with biostatistical methods in bioinformatics and functional genomics.

Target: Be able to understand the basics of acquisition, analysis, and presentation of small- and large-scale high throughput sequencing data. Hypothesis-specific data analysis design; a quick overview of data analysis considerations depending on the biological source (DNA, RNA, histone …) and platform (bulk, single-cell, sequencing method …) will be covered.

Desired background: Basic knowledge of experimental design in scientific research; theoretical background in biological data analysis and bioinformatics; basic understanding of computer operating systems (esp. Linux/Unix); some hands-on experience in R or related data analysis packages is a plus. Participants are strongly encouraged to read reviews and introductory textbooks related to biostatistics, bioinformatics, and functional genomics.

Methodology: R and Bioconductor; public and in-house -omics datasets. The lecture will be held once per week. Participants will have practical questions to work on until the next session. At the end of the lecture series, participants (in a group of 3 to 4) will work on a dataset related to a selected scientific question. They will work as a team to come up with a concise and clear plan on how to tackle the problem and undertake the data analysis. Each group will finally give a presentation (~15 minutes) followed by a discussion with the attendees.

References:

Gentleman et al. (Eds.) (2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor / R.

Singer et al. (2017). Bioinformatics for precision oncology *Briefings in Bioinformatics*, bbx143, https://doi.org/10.1093/bib/bbx143.

Michael F. Berger & Elaine R. Mardis (2018). The emerging clinical relevance of genomics in cancer medicine. *Nature Reviews Clinical Oncology,* volume 15, pages353–365 (2018).

Michael Agostino (2012). Practical Bioinformatics. Garland Science.