**NEXT-GENERATION SEQUENCING FOR ANTIBODY DISCOVERY AND ENGINEERING**

**Sai Reddy**, Ph.D., Department of Biosystems Science and Engineering, ETH Zurich, Switzerland

**Simon Friedensohn**, Department of Biosystems Science and Engineering, ETH Zürich

In this training seminar, participants will learn about Next-Generation Sequencing (NGS) of antibody repertoires. Part 1 will provide an introduction to the antibody repertoires, consisting of genetic background, generation of diversity, sequencing technologies and a hand-on session on the computational tools available for the analysis antibody repertoire NGS data. Part 2 will focus on the pre-processing and analysis of data. Each step of the preprocessing will be elucidated using the programming language R along with existing bioinformatics pipelines available. Repertoire analysis content will provide statistical quantification and visualization of high-dimensional data. The course will be fully interactive with case studies, participants will be able to download data and example scripts. Please bring your computer.

**Topics Include:**

**PART I: Introduction to NGS of Antibody Repertoires and Bioinformatics Tools**

**Antibody repertoires**

* Antibody responses: cells, biological function and applications
* Genetic background and basis of diversity

**Introduction to NGS of antibody repertoires**

* Sequencing platforms: advantages and disadvantages
* Single-cell and paired variable region sequencing

**Experimental design and considerations**

* Cell isolation and selection
* Sampling depth
* Library preparation methods
* Error and bias correction strategies

**Data analysis overview**

* Data formats and preprocessing
* Antibody sequence annotation

**Hands-on session: Introduction of common computational tools**

* Introduction to programming with the language R
* Data preprocessing
  + Errors and Quality control
  + Assembly of paired-end reads
* Bioinformatics tools for antibody repertoire annotation
  + IMGT
  + MiXCR
  + In-house aligner (MAF)
  + Single-cell sequencing software
* R packages for data analysis

**PART II: Preprocessing and Analysis of Antibody Repertoire NGS Data**

**Hands-on session: Preprocessing data in R**

* Additional data sources
  + Common databases and repositories
  + Guide to data retrieval and download
* Explore annotated reads
  + IMGT – V(D)J germline segment assignment, CDR1/2/3, FW1/2/3/4
* Filtering and cleaning
  + Filtering CDR3 singletons, short CDR3s, remove unproductive sequences

**Hands-on session: Immune repertoire analysis**

* Determine clonal variants
* Global characterization of immune repertoires
  + Diversity, somatic hypermutations (SHM) distribution, CDR3 length distribution, V(D)J gene and allele frequency
* Extraction of monoclonal antibody (mAb) candidates and somatic variants
  + Based on CDR3s, SHM, Germlines, repertoire frequencies
* Data visualization
* Saving output and reproducibility

**Advanced topics & case studies**

* Identification of biologically relevant sequences from NGS data using machine learning
* Phylogenetic trees of clonal lineages
* In-silico modelling of novel antibody sequences based on antibody repertoires
* Single-cell sequencing