Syllabus:

Block 1: Advanced programming in R (2.5 credits)
This block provides the students a deep knowledge of programming in R, which is crucial for working with large data sets. It also teaches the students how to write maintainable code giving reproducible scientific results.

Teachers: Carl Nettelblad, Behrang Mahjani, Silvelyn Zwanzig
Total number of lectures in this block: 4 lectures + 2 labs

1. **Version control, GIT, Libraries, CRAN**
2. **Language Foundations (**Data structures, Subsetting, Vocabulary, Style, Functions, OO field guide, Environments, Exceptions and debugging)
3. **Functional programming (**Functionals, Function operators, Metaprogramming)
4. **Non-standard evaluation (**Expressions, Domain specific languages)

Block 2: High performance programing in R (2.5 credits)
The goal of this block is to teach the students how to write and analyze a high performance code in R, which is essential in handling computationally intensive algorithms.

Teachers: Carl Nettelblad, Salman Toor, Behrang Mahjani (lab)
Total number of lectures in this block: 4 lectures + 2 labs

Part 1: Performant code in R (Performance, Profiling, Memory, Rcpp, R's C interface)
(Carl Nettelblad) (1 lecture + 1 lab)

Part 2: Parallelization in R

1. **Parallel Computing (**Explicit parallelism, implicit parallelism, GPUs)
   (Carl Nettelblad) (1 lecture)
2. **Big data on Cloud Computing (**Hadoop, Spark) *(Salman Toor) (2 Lectures+ 1 lab)*

Block 3: Statistical and numerical methods for analysis of large data sets, with focus on bioinformatics applications (2.5 Credits)
This block showcases how to move your R usage from individual machines and modest-size datasets to the extremely large data sets that are becoming increasingly common. We cover both aspects of code and software architecture, and the statistical treatment, including some common mistakes.

Teachers: Guest lecturer
Total number of lectures in this block: 3 lectures + 1 labs

1. **Numerical precision in extremely large data sets**
2. **Packages for handling distributed and very large data in R**
3. **Statistical challenges when analyzing billions of data points**