



## STAT 540, Statistical Methods for High Dimensional Biology 2016/17, Term 2

**Instructors: Sara Mostafavi & Paul Pavlidis**

**Time and Place:** M/W 9:30-11:00 AM, ESB 2012 (lecture)  
W 11:00-12:00 PM ~OR~ 12:00-1:00 PM, ESB 1042 (lab/seminar)

**Seminar:** *officially* runs Wed 12pm - 1pm; *unofficially* students are welcome to come after class around 11am and begin a ~1 hour guided analysis with TA support; TA will remain in the lab until 1pm to help those who start at 12pm and for general office hours. Seminars will take place in ESB 1042. Registration in a seminar is **REQUIRED**.

**Prerequisites:** Officially none BUT in reality...

- Statistics: you should have already taken university level introductory statistics course. Prior to starting the course, we expect you to review notes from an introductory statistics course, to be familiar with common probability distribution, random variables, and hypothesis testing. We will review these topics, but you should have already heard/seen these before.
- Biology: No requirements, but you are expected to learn things like the difference between a DNA and RNA and a gene and a genome.
- R: no experience required but be prepared to do *a lot* of self-guided learning. Go ahead and start now by [installing R](#) and the HIGHLY RECOMMENDED “integrated development environment” (IDE) [RStudio](#)! Students are expected to run R on their own computer or a computer they have plenty of access to and control over. The best set-up, if possible, is to bring your own laptop to the computing seminars.

### Assessment:

- Homework.
- Group project. Groups formed and projects conceived during January/February. Primary deliverable is a poster, presented in last class meeting. Each student also produces a short report and the group submits a digital supplement.
- Paper critiques: You will review and critique two Bioinformatics/Computational Biology papers.
- “Other”, e.g. preparedness, participation, handing in credible effort on *small* weekly computing lab mini-exercises.

### Topics/timeline (tentative):

#### Week 1

seminar 00 | R, RStudio Set Up & Basics, students complete set-up and exploration on their own in advance

lecture 01 | Introduction to high dimensional biology and the course |

seminar 01 | Basic Data Analysis in R |

seminar 01 | Basics of Molecular biology/genetics 101 |

## **Week 2**

lecture 02 | Overview / review of probability and statistical inference, part1  
lecture 03 | Overview/review of probability and statistical inference, part2  
seminar 02 | Cont'd introduction to R and Git/GitHub |

## **Week 3**

lecture 04 | Exploratory analysis |  
lecture 05 | Data QC and preprocessing |  
seminar 03 | Introduction to R graphics |  
Project groups should be formed around now

## **Week 4**

lecture 06 | Statistical inference: hypothesis testing & two group comparisons |  
lecture 07 | Statistical inference: linear models 1 |  
seminar 04 | Data aggregation and two group testing |  
Initial project proposals due around now

## **Week 5**

lecture 08 | Statistical inference: linear models 2 |  
lecture 09 | Statistical inference: large scale inference; fitting many linear models; limma|  
seminar 05 | Fitting and interpreting linear models (low volume) |

## **Week 6**

Homework #1 assigned around now  
lecture 10 | Large scale inference: multiple testing and permutation testing ||  
seminar 06 | Fitting and interpreting linear models (high volume), limma package | Feedback to groups re: initial project proposals around now.

## **Week 7**

(UBC closed for mid-term break)

## **Week 8**

lecture 11 | Analysis of RNA-Seq data - Part 1 |  
lecture 12 | Analysis of RNA-Seq data – Part 2 |  
seminar 07 | RNA-Seq data part 1 |  
Homework #1 and final project proposals due around now.

## **Week 9**

lecture 13 | Analysis of RNA-Seq Data - Part 2 |  
lecture 14 | Guest lecture: Microbiome data and/or Methylation data analysis |  
seminar 08 | RNA-Seq data part 2  
Homework #2 assigned around now

## **Week 10**

lecture 15 | Clustering algorithms |  
lecture 16 | Dimensionality reduction – non-linear approach |

seminar 09 | Clustering and PCA |

### **Week 11**

lecture 17 | Classification and variable selection |

lecture 18 | Variable selection |

seminar 10 | Supervised learning, cross validation, variable selection |

### **Week 12**

Homework #2 due around now

lecture 19 | Enrichment analysis, gene networks |

lecture 20 | Enrichment analysis, gene networks, continued |

seminar 11 | TA office hours during seminar time ... group project work |

### **Week 13**

lecture 21 | Multi-omic analysis 1 |

lecture 22 | Guest lecture TBD|

seminar 12 | TA office hours during seminar time ... group project work |

### **Week 14**

lecture 23 | Poster session 9:30am - 1:00pm |

lecture 22 | Guest lecture TBD

seminar 12 | TA office hours during seminar time ... group project work |

### **Week 14**

lecture 23 | Poster session 9:30am - 1:00pm