

## Introduction and Materials

Welcome to Statistics for Genomic Data Science. I'm really excited for you to be a part of this class and to be a part of the Genomic Data Science Specialization from Johns Hopkins. Statistics is a key part of genomic data science and an area I'm incredibly excited about. I have been working on statistical genomics for about the last 10 years now (gulp!) and find that it is always teaching me something new. Doing statistics well is a challenge for data as high-throughput and variable as you often find in genomics. But it is well worth it! The best and most reproducible genomic results have all been discovered using sound experimental design, solid statistical analysis, and careful code. As you might imagine it is tough to cover all of the nuances of a topic as complicated as statistics for genomics in just four weeks. So I have structured this class to hit the key conceptual ideas of normalization, exploratory analysis, linear modeling, testing, and multiple testing that arise over and over in genomic studies. Then I will point you to further resources that can help you understand your specific data type and move your project along. I hope that if you complete this course you will have a grounding in the basics and a good handle on where you can go for more help. You can find all the course lecture materials in the course pages or from the companion site.

Course materials: [http://jtleek.com/genstats\\_site/](http://jtleek.com/genstats_site/)

The course R package that contains all the tutorials for the R coding examples during the course.

Course R package: <https://github.com/jtleek/genstats>

This class assumes a basic knowledge of genomics and computing that you can find in the courses:

[Introduction to Genomic Technologies](#)

[R programming](#)

[Bioconductor for Genomic Data Science](#)

In particular, the assignments will require some analysis in the R programming language. If you haven't had experience in R and Bioconductor that might make some of the assignments challenging, but I've done my best to include example R and Bioconductor workflows that will give you some idea where to start. All of the lectures are open from day 1, so you can work at your own pace. However, we've set deadlines for the assignments to help keep you on task and motivated. We will be releasing the assignments over the course of this first week, and the first quiz is open now. Please make sure to read the syllabus and check out the deadlines. Your first quiz is due by the end of this week. Awesome to have you in class and good luck!Jeff