

BIOSTATISTICS SEMINAR



Michael Love,
Dr. rer. nat
Assistant Professor,
Gillings School of
Global Public Health,
Department of
Biostatistics,
UNC

Avoiding two pitfalls in data analysis of gene expression

I will discuss two basic errors that can occur in data analysis of gene expression, and discuss solutions in the form of software packages developed by my lab. One is more of a statistical nature, while the other concerns computational reproducibility. In the first part, I will discuss quantification challenges associated with measuring the abundance of RNA transcripts with similar sequence via short sequenced reads, and a non-parametric method called "Swish" to address this problem, developed primarily by Anqi Zhu (graduated from PhD program in 2019). In the second part, I will discuss the widespread problem of missing annotation metadata for RNA-seq datasets (e.g. which set of genes were used, different release versions), and how this can lead to problems with computational reproducibility and data analysis. We have a newly published solution, "tximeta", which can eliminate these issues by the use of cryptographic hashing followed by subsequent lookup in a hash table of commonly used reference transcripts.

Thursday April 23, 2020

3:30 pm - 4:30 pm

Zoom meeting: Please also find a link in the email invite.

https://uncsph.zoom.us/j/893328432?pwd=MnRUWitYaEZva3ZpYzhGMkJOVktBQW09&_ga=2.141183730.1187571104.1585657287-1769805331.1548686859



UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH