Stanford MEDICINE | Department of Biomedical Data Science

CuGen: A GPU-accelerated framework for large-scale genomics

Tuomo Tapio Johannes Kiiskinen
Postdoctoral Scholar
Department of Biomedical Data Science

November 5, 2026
1:30PM-2:50PM
Li Ka Shing Center, Room:120

Reading list:

1. Great material as an intro and deeper diver for large scale genomics and GWAS genome-wide association studies (GWAS):
    a. Jonathan K. Pritchard, intro to population genetics in general: https://web.stanford.edu/group/pritchardlab/HGbook.html
    b. Matti Pirinen(university of Helsinki), more specifically on GWAS methodology: https://www.mv.helsinki.fi/home/mjxpirin/GWAS_course/
2. Summarizing previous methods for confounding control in GWAS
    a. SAIGE https://www.nature.com/articles/s41588-018-0184-y
    b. REGENIE https://www.nature.com/articles/s41588-021-00870-7
3. Some prev GPU methods
    a. torchGWAS: https://arxiv.org/abs/2604.21095
    b. SAIGE on GPUs: https://academic.oup.com/bioinformatics/article/42/3/btag032/8438945?login=false
4. Our preprint: https://www.medrxiv.org/content/10.64898/2026.07.15.26358178v1

Going through these quickly walks a student through the idea on:

1. how is genetic variation used to study human traits and diseases?
2. What is a genome-wide association study?
3. What methods have been applied to confounding control previously in GWAS?
4. What makes a GWAS a good candidate for GPUs and what obstacles there are?
5. How does CuGen solve these?

A good strategy can be just to for example focus on the intro items on list item 1 (Pritchard + Pirinen) and then just summarize the rest on the context of how they are relevant in regards to CuGen.