Introduction to GWAS and Polygenic Risk Scores

Workshop Lead: Yuan Ding

Registration link: https://involvement.mcgill.ca/event/297006

Approximate duration: 4 hours

Prerequisites:

1. Basic understanding of molecular biology and genetics
2. Basic understanding of statistical concepts (regression, p-value)
3. Basic understanding of Python

Summary: This workshop introduces participants to the fundamental concepts and practical applications of Genome-Wide Association Studies (GWAS) and Polygenic Risk Scores (PRS). Attendees will learn how GWAS identify genetic variants associated with complex traits, understand the construction and interpretation of PRS, and recognize common pitfalls and limitations of these methods. The workshop includes hands-on activities using real and simulated genomic data to calculate PRS and explore their predictive utility.

Learning Objectives: (List 2-5 learning objectives participants will learn upon completion of this workshop)

1. Explain the principles and methodology behind Genome-Wide Association Studies.
2. Describe how Polygenic Risk Scores are constructed and applied in disease risk prediction.
3. Identify common pitfalls and limitations inherent in GWAS and PRS analyses, including issues related to linkage disequilibrium and ancestry-specific effects.
4. Perform basic genomic data manipulation and quality control using PLINK.
5. Calculate and interpret Polygenic Risk Scores using provided datasets.

1. **Module 1 (GWAS and Polygenic Risk Score)**
    a. Overview of genetic variants (SNPs, Indels, CNVs)
    b. Distinction between complex traits and Mendelian traits
    c. GWAS
        i. Statistical model to identify risk variants
        ii. Pitfalls (LD, ancestry specific effects)
        iii. Hands on activity: browsing GWAS summary statistics on GWAS catalogue
    d. Polygenic risk score
        i. Different ways to adjust for weights of SNPs when scoring individuals
        ii. Necessary quality control steps for PRS calculation
2. **Module 2 (Hands-on activity: calculate Polygenic Risk Score)**

a. Using PRS to predict cardiovascular disease with simulated data
b. Pitfalls of PRS (Poor accuracy of PRS when applied across ancestries)