

Hands-on Biological Data Science with R

Introduction to Complex Genetics



Introduction

The Outline

- **Genetics Basics**
 - Introduction
 - Terminology
 - Simple vs Complex Traits
- **Complex Genetics**
 - Genetic Architecture of Complex Traits
 - Approaches on Identifying Genetic Determinants of Complex Traits
 - Genome Wide Association Studies (GWAS)

Genetics Basics

- Introduction
- Terminology
- Simple vs Complex Traits

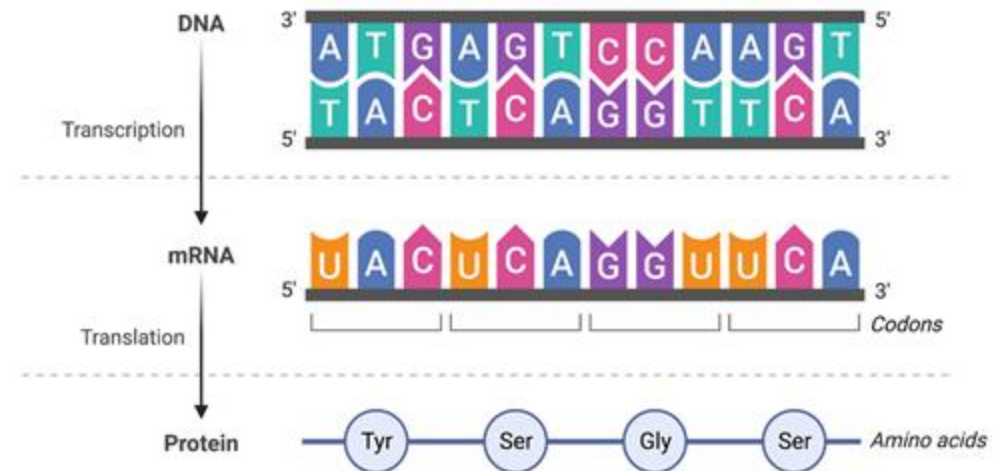
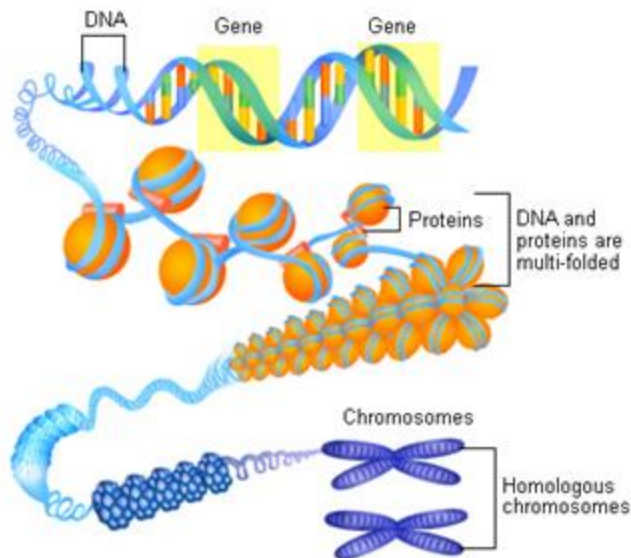
Genetics Basics

Introduction

What is Genetics?

- Genetics is the study of genes, genetic variation, and heredity in organisms. In medicine, the focus is on humans, aiming to understand our genetic structure and its influence on variations within a population.

The Central Dogma

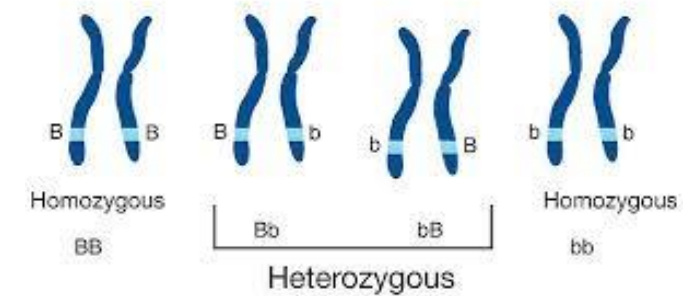
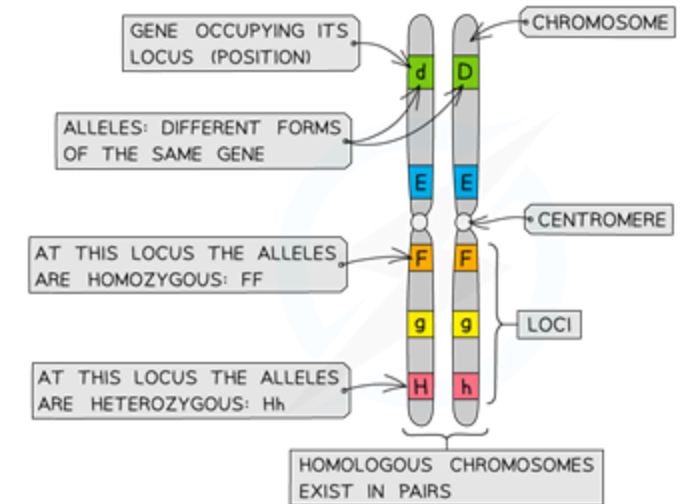


Genetics Basics

Terminology

Genetics Terminology

- **Locus (pl. Loci):** A unique physical position on the chromosome where a gene exists.
- **Gene:** A sequence of DNA that shapes a specific trait.
- **Allele:** An alternate version of a specific gene. Since humans have 2 paired sets of (homologous) chromosomes, we have 2 copies (alleles) of each gene.
- **Homozygous:** If both the member of allelic pair in homologous chromosome are identical.
- **Heterozygous:** If both the member of allelic pair in homologous chromosome are not identical.
- **Hemizygous:** when genes are present in only one copy. (e.g. genes of X chromosome and Y chromosome of male)
- **Genotype:** the genetic structure of the individual.
- **Phenotype:** the outward expression of the genotype, as affected by the environmental factors as well.



Genetics Basics

Simple vs Complex Traits

Traits

- Traits come in broadly two types: Simple (also called Mendelian) and complex.
- **Simple Traits:**
 - These are mostly either/or traits that are controlled by one gene, which has one effect. (eg. CF, CFTR)
 - They are strongly genetically determined, they cannot be changed and there is little to no influence from the environment. (eg. Blood type)
 - They have a finite number of genotypes, usually very small, since they caused by a single gene.
- **Complex Traits:**
 - These are mostly continuous traits that are influenced by more than one gene. (**polygenic**)
 - In addition to the genotype, they are influenced by environmental factors (eg. diabetes, obesity).
 - Because of their polygenicity, there can be many genotypes. The more genes that are involved, the finer the genotype categories until they converge into a continuous distribution.
 - (1 gene = 3 genotypes, 2 genes = 9 genotypes, 3 genes = 27 genotypes..)
 - It's often difficult to try to divide them into categories due to an infinite number of phenotypes

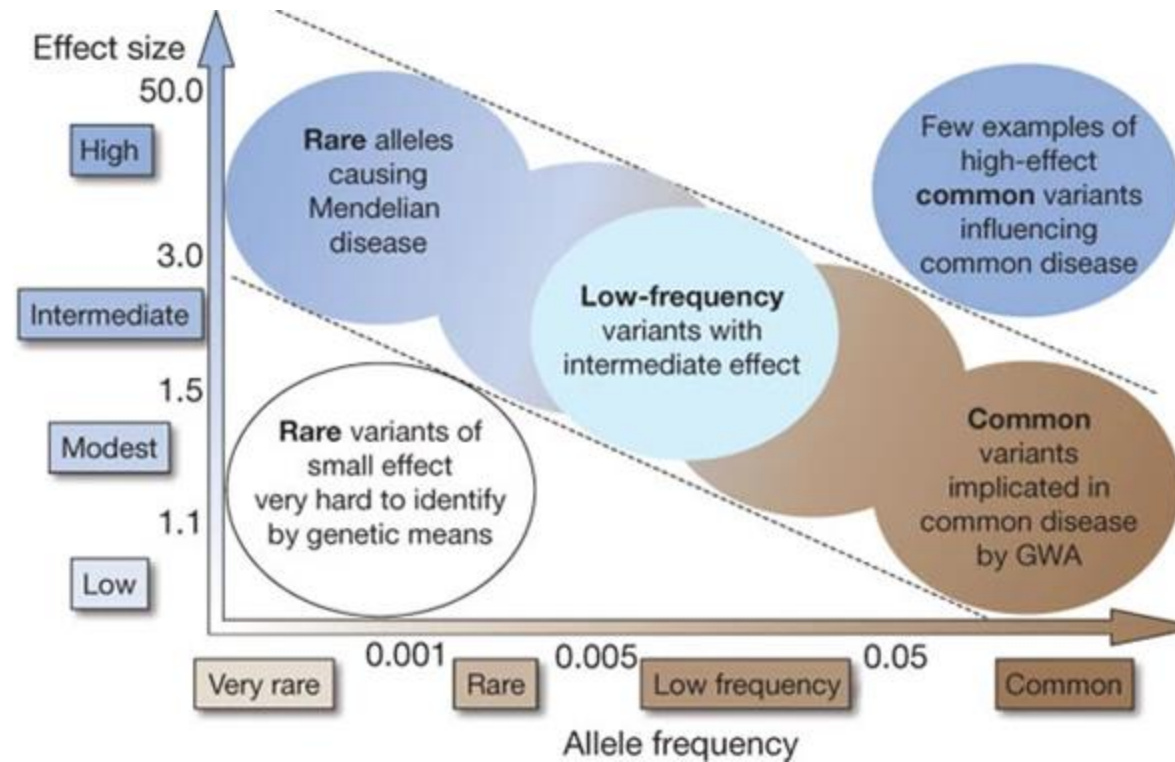
Complex Genetics

- Genetic Architecture of Complex Traits
- Approaches on Identifying Genetic Determinants of Complex Traits
- Genome Wide Association Studies (GWAS)

Complex Genetics

Genetic architecture of complex traits

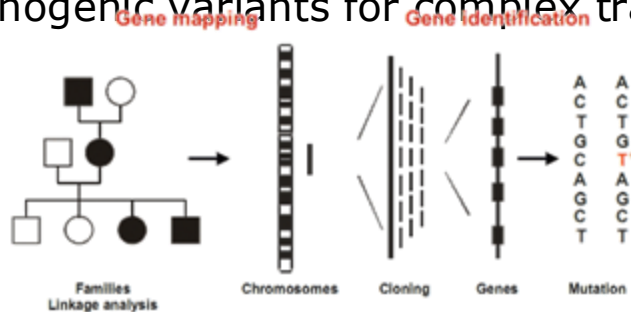
- The basic assumption regarding genetic determinants of complex disease is that **common variants in many genes** will each lead to a small change in the risk of disease, and that the overall risk of disease is determined by the combination of multiple variants and environmental exposures.



Complex Genetics

Approaches on Identifying Genetic Determinants of Complex Traits

- Until the mid-1990s, the primary methods for identifying susceptibility variants for complex traits were linkage analysis and candidate-gene association studies.
- **Linkage Analysis**, a statistical method that successfully identified causal genes for over 2,000 monogenic diseases, has had limited success in mapping genes associated with complex traits. However, identifying pathogenic variants for complex traits via this method remains highly challenging.

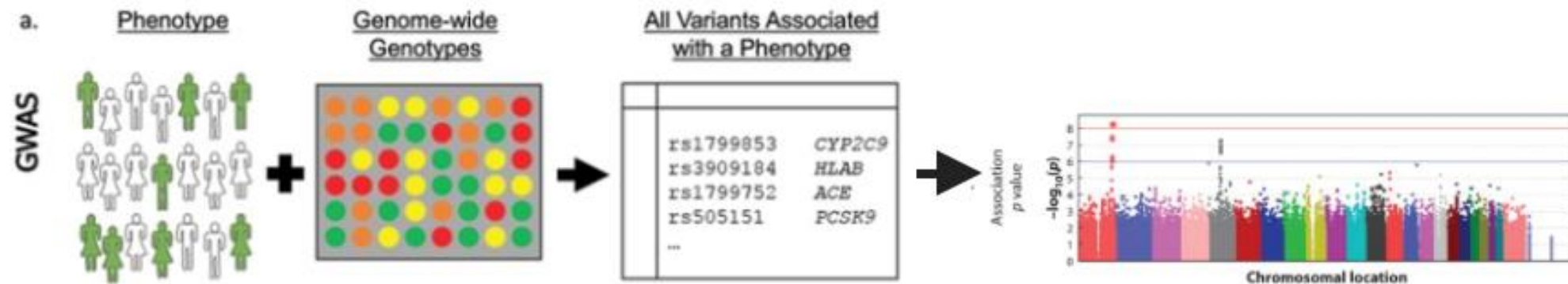


- **Candidate-gene association studies**, which involves hypothesizing the role of a gene in a disease based on prior knowledge of its function, has occasionally been successful—for example, in identifying APOE4 as a risk factor for Alzheimer's disease. However, replication of findings has often been inconsistent, and this method is limited to conditions where the underlying biology is already known.

Complex Genetics

Approaches on Identifying Genetic Determinants of Complex Traits

- **Genome-wide association studies (GWAS)** tests multiple genetic variants across the human genome simultaneously to identify associations with a specific disease or trait. Unlike candidate-gene approaches, GWAS does not rely on prior knowledge of gene function, providing an unbiased method to explore the entire genome.
- In its most basic form a GWAS is just a t-test or a chi-square test (for categorical variables) that aims to compare the effects of different genotype groups (AA, Ab, aa) on a selected phenotype, ran for from hundreds of thousands to millions of genetic variants. To account for any confounder affecting the phenotype, more complex methods of logistic and linear regression are being used



Complex Genetics

Genome Wide Association Studies (GWAS)

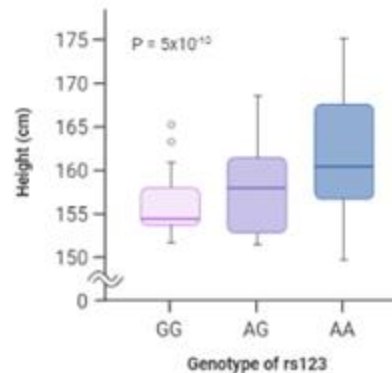
The Principle of a Genome-wide Association Study (GWAS)

- 1 Height and genetic data for individuals in study



- 2 Single-variant association test with a candidate variant, rs123

A alleles increase height on average*

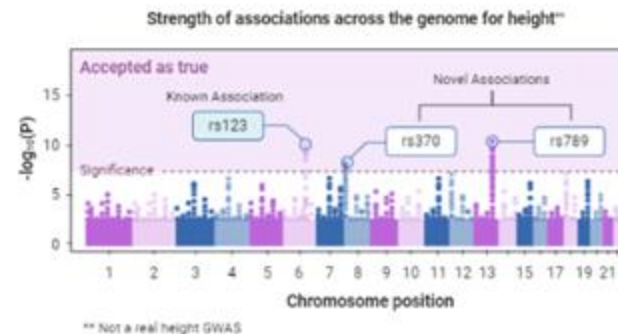


* Liberties were taken with the size of the allelic effect shown in this example

- 3 rs123 may not be causal, but a measured proxy for the causal variant, rs456

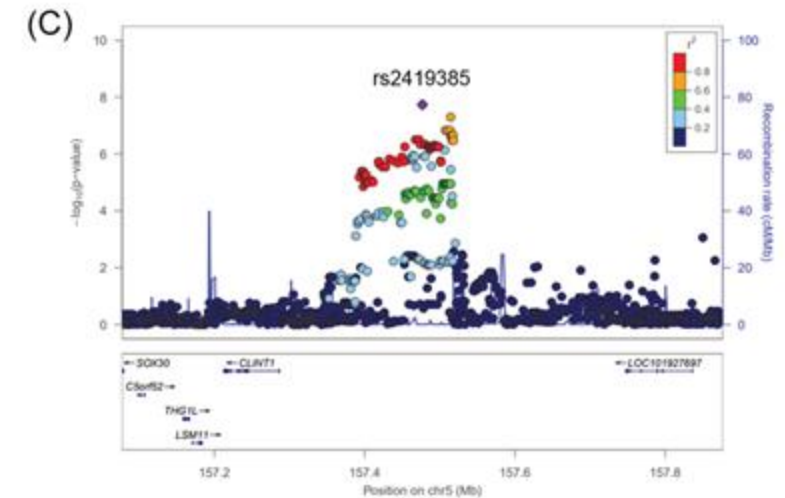
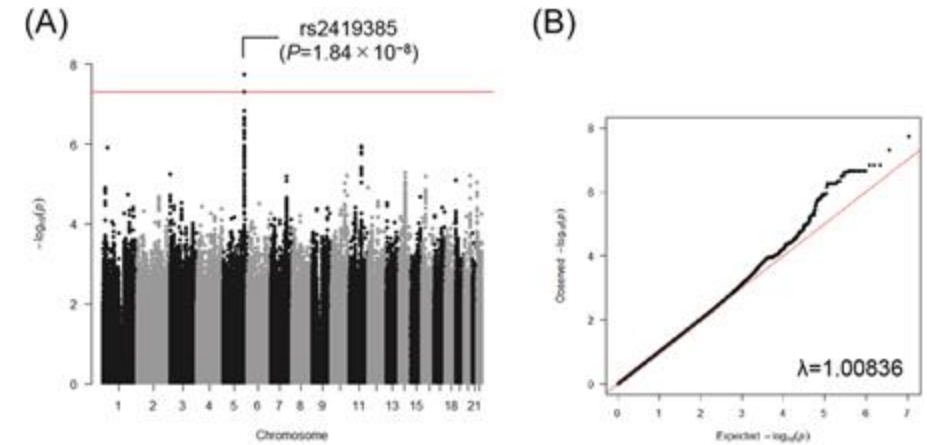


- 4 GWAS use evenly-spaced proxies for association tests across the genome; plots show if $-\log_{10}(P)$ passes the genome-wide significance threshold



- 5 Further analyses of the regions can identify causal variants and their function

Fine-mapping Functional work Meta-analysis



Thanks for your attention!

Contacts:

Ekin Yaman
Kim-Hellmuth

✉: Ekin.Yaman@med.uni-muenchen.de

Dr. med. Paula Rothämel

✉: Paula.Rothaemel@med.uni-muenchen.de

Dr. med. Sarah

<https://www.ccrc-hauner.de/kim-hellmuth-labor>