Lecture 2

GWAS Unveiled: The Historical Journey and Fundamental Concepts

by Dr. Mustafa İsmail Özkaraca

Contents

What is Genome-wide Association Studies (GWAS)?

Statistical inference procedure based on **association testing** applied **genome-wide**.

Goal:

Identify genetic variants that are associated with traits/diseases.

Motivation:

Finding an address (before the invention of smartphones!): the One O'Clock Gun at Edinburgh Castle:

- Ask for directions to Edinburgh Castle. (GWAS)
- Once there, ask a staff member where to find the One O'Clock Gun.

Figure is from: https://www.edinburghcastle.scot/see-and-do/highlights/one-oclock-gun

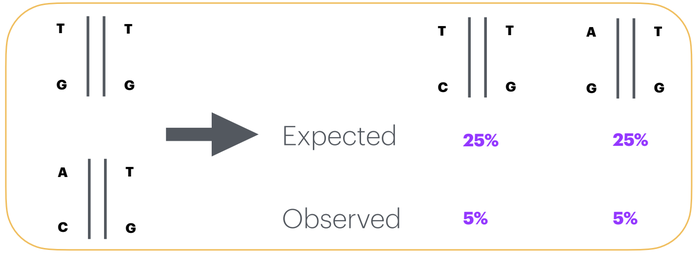
Why GWAS?

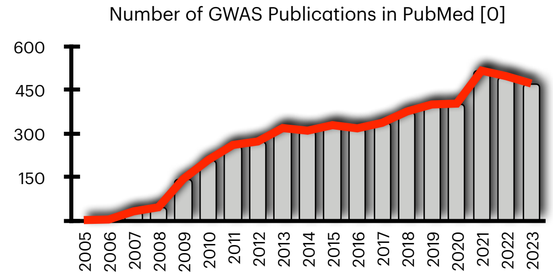
- Identifies genetic variants associated with traits **without prior hypotheses**

- Highlights candidate genes and pathways

- Disease risk predictions

Historical Evolution of Quantitative Genetic Studies





Advances in Linkage Analysis

Late 1800s
Early 1900s

1937 1950s **1963** 1970s 1980s **1996** **2003** **2005**

Family Studies **Genetic Screening** **Candidate Gene Studies** **GWAS**

Francis Galton (1869, 1886)
Sir Archibald Garrod (1902)

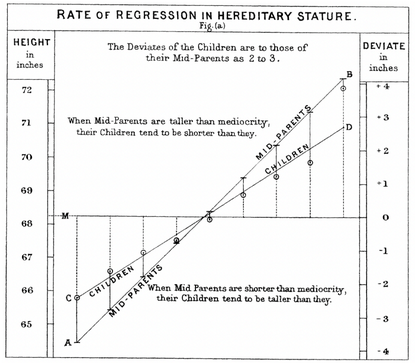


Figure from Galton F.
"Towards Mediocrity in Heredity Stature"
(1886)

First Human Linkage Analysis
Bell and Haldane (1937)

Newborn Screening for Phenylketonuria (PKU)
Guthrie, Robert, and Ada Susi.
"A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants."
Pediatrics 32.3 (1963): 338-343.

Linkage vs Association [1]

Human Genome Project

First GWAS [2]

[0] https://pubmed.ncbi.nlm.nih.gov/?term="genome-wide+association+study"[Title]+AND+("2005"[Date+-+Publication]+:+"2024"[Date+-+Publication])
[1] Risch, Neil, and Kathleen Merikangas. "The future of genetic studies of complex human diseases." *Science* 273.5281 (1996): 1516-1517.
[2] Klein, Robert J., et al. "Complement factor H polymorphism in age-related macular degeneration." *Science* 308.5720 (2005): 385-389.

The GWAS Catalog

Published Genome-Wide Associations as of July 2019

p≤5X10-8 for 17 trait categories

National Human Genome Research Institute

EMBL-EBI

NHGRI-EBI GWAS Catalog
www.ebi.ac.uk/gwas

http://ftp.ebi.ac.uk/pub/databases/gwas/timeseries/current/GWAS_Catalog_annotated_diagram.pdf

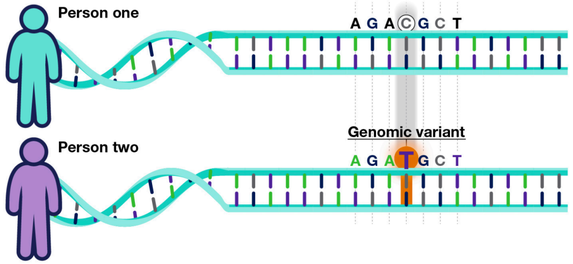
"The Catalog **was founded** by the **NHGRI** in **2008**, in response to the rapid increase in the number of published genome-wide association studies (GWAS)." [3]

"As of **2024-11-20**, the GWAS Catalog contains **7083** publications, **692444** top associations and **96947** full summary statistics." [4]

[3] https://www.ebi.ac.uk/gwas/docs/about
[4] https://www.ebi.ac.uk/gwas/home

Fundamental Concepts in GWAS

Genomic variant (g)



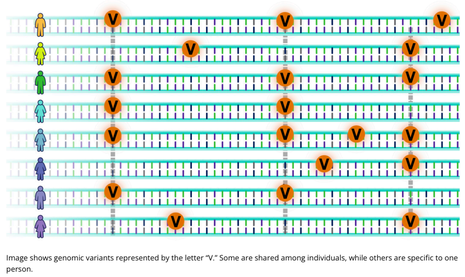


Image shows genomic variants represented by the letter "V." Some are shared among individuals, while others are specific to one person.

Figures are from: https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores

Two Vital Questions for a given variant g, and a trait T:

1. Is g **associated with** T ?
2. If associated, **how much impact** does g have on T ?

Manhattan plot

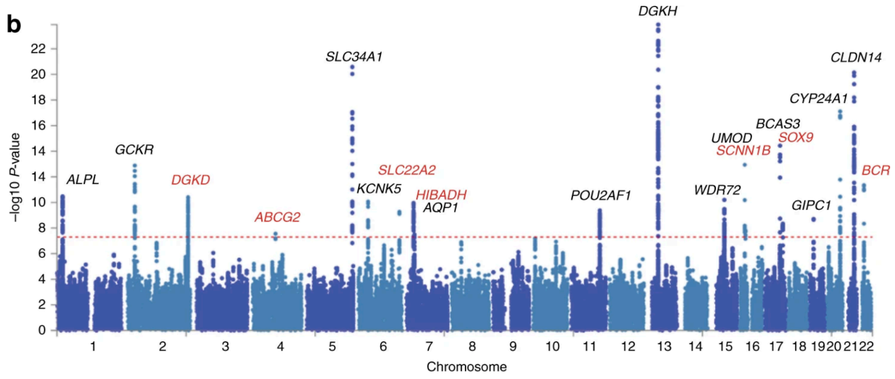
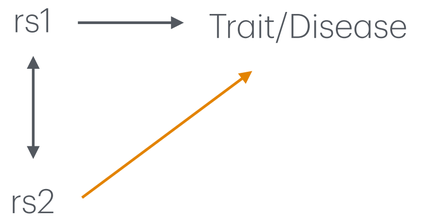


Figure is from:
Howles, S.A., Wiberg, A., Goldsworthy, M. *et al.* Genetic variants of calcium and vitamin D metabolism in kidney stone disease. *Nat Commun* **10**, 5175 (2019).

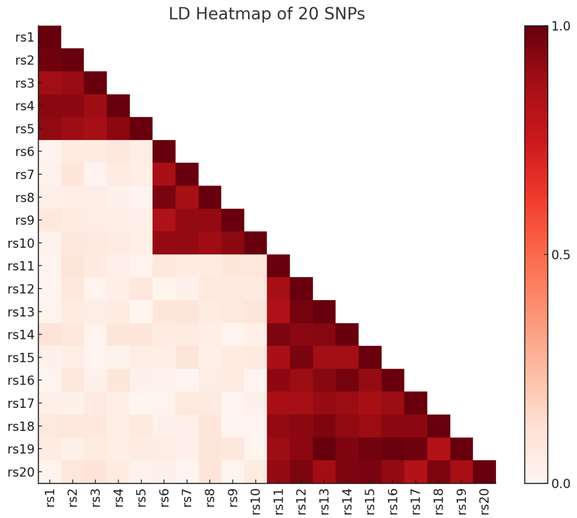


Effect size Measures the impact of a genetic variant on a particular trait or phenotype of interest.

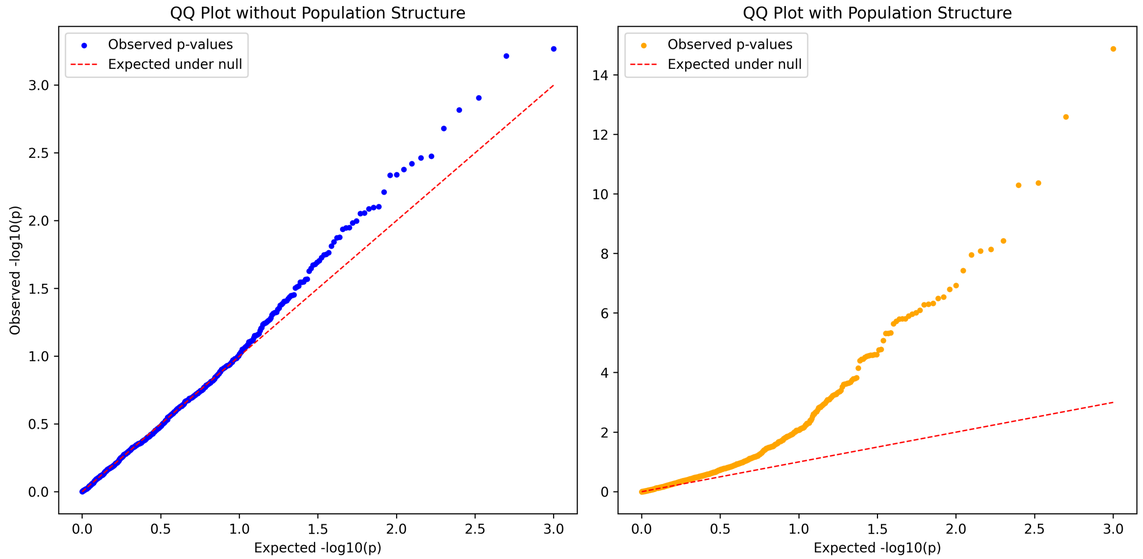
Fundamental Concepts in GWAS

Linkage Disequilibrium (LD)

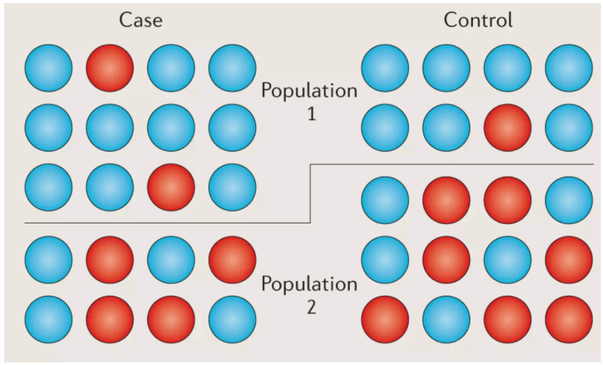
non-random association of alleles (at different loci).



QQ plot



Population structure



Genomic Inflation Factor (λ)

$$\lambda = \frac{Median(Observed\ Test\ Statistics)}{Median(Expected\ Test\ Statistics\ under\ the\ null)}$$

.

$\lambda > 1$ might indicate the presence of population structure

.

Figure is from: Balding, D. A tutorial on statistical methods for population association studies. *Nat Rev Genet* **7**, 781–791 (2006).

Fundamental Concepts in GWAS

Identity by State (IBS)

Identical genetic segments between individuals.

Identity by Descent (IBD)

Identical genetic segments between individuals due to inheritance.

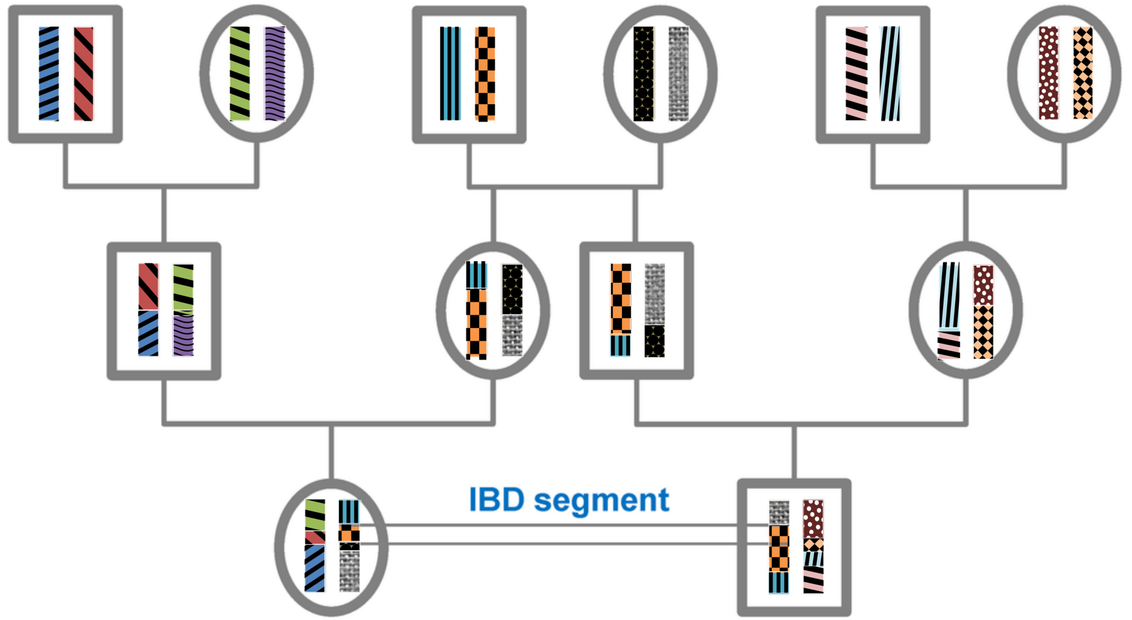


Figure is from: https://en.wikipedia.org/wiki/Identity_by_descent

Coefficient of relationship (r)

A measure of biological relationship between individuals. Examples:

$$r(self, self) = 1,\ r(self, son) = 0.5,\ r(self, daughter) = 0.5,\ r(self, father) = 0.5,\ r(self, mother) = 0.5$$

Fundamental Concepts in GWAS

Minor Allele Frequency (MAF)

The frequency of the less common allele in a given population.

Hardy-Weinberg Equilibrium (HWE)

HWE describes a stable state where allele and genotype frequencies remain constant over time in a given population, provided random mating occurs and there is no influence of evolutionary forces such as selection, mutation, migration, or genetic drift..

Genetic Relationship Matrix

A matrix $\pi = \left[\pi_{i,j}\right]$ that estimates the genetic relatedness between individuals based on their genotypes.

$$\pi_{i,j} = \frac{1}{M}\sum_{k=1}^{M}\frac{(g_{ki} - 2f_i)(g_{kj} - 2f_i)}{2f_i(1 - f_i)}.$$

What’s Next

1. Association Analysis by Linear Regression

2. Why Linear Regression does not work?

3. Association Analysis by Linear Mixed Models

4. Advantages and (further) Challenges