Week 6

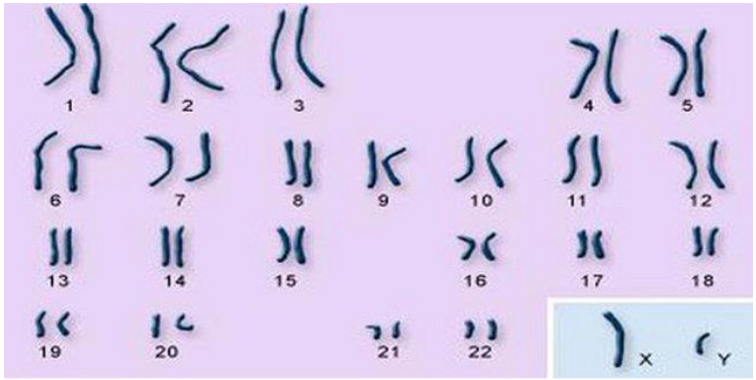
2025.03.25

Genome-Wide Association Study (GWAS) (全基因體關聯性分析)

- a study of a genome-wide (全基因體) set of genetic variants(基因變量) in different individuals to see if any variant(變量) is associated with a trait (性狀).

https://img.jagranjosh.com/images/2022/July/1172022/Compress_world_Population.webp

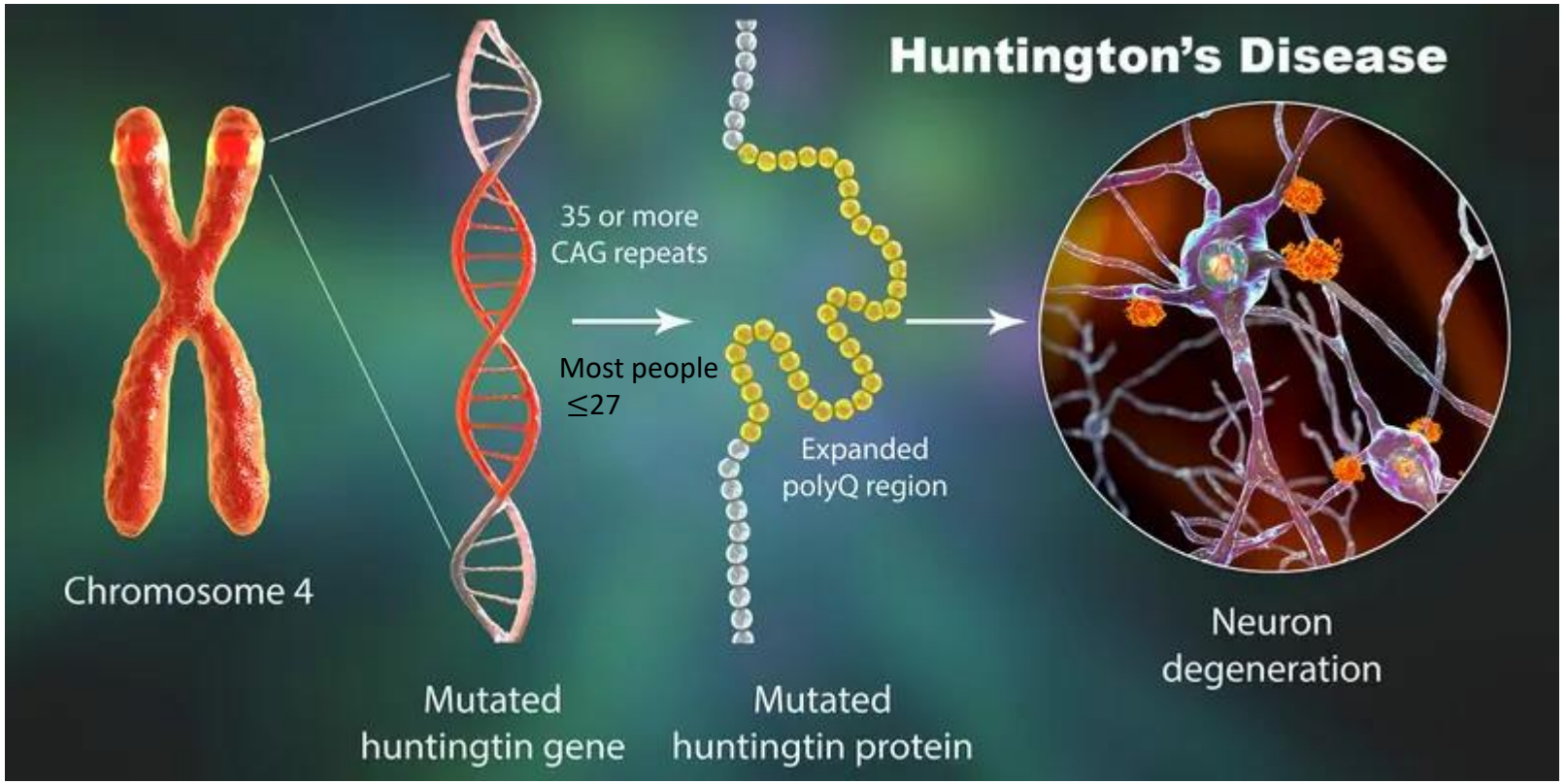
GWAS



GWAS versus Mendelian disorders

- Mendelian disorder: ex. Huntington's disease (亨丁頓舞蹈症)

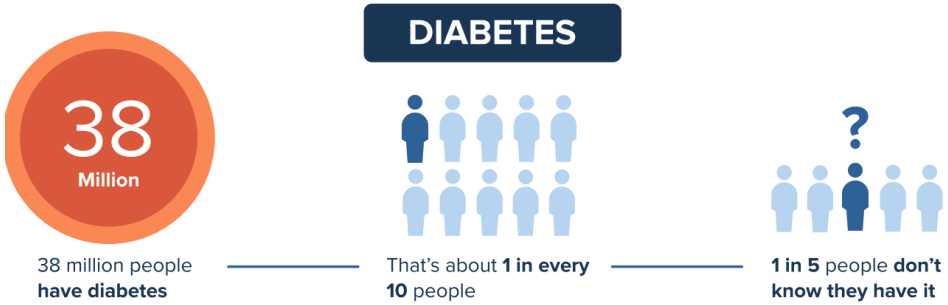
https://my.clevelandclinic.org/-/scassets/images/org/health/articles/14369-huntingtons-disease



https://cdn.drugdiscoverynews.com/assets/articleNo/15759/aImg/40650/61422-m.webp

GWAS versus Mendelian disorders-2

- GWAS: common variants common disease
  - BMI, type 2 diabetes

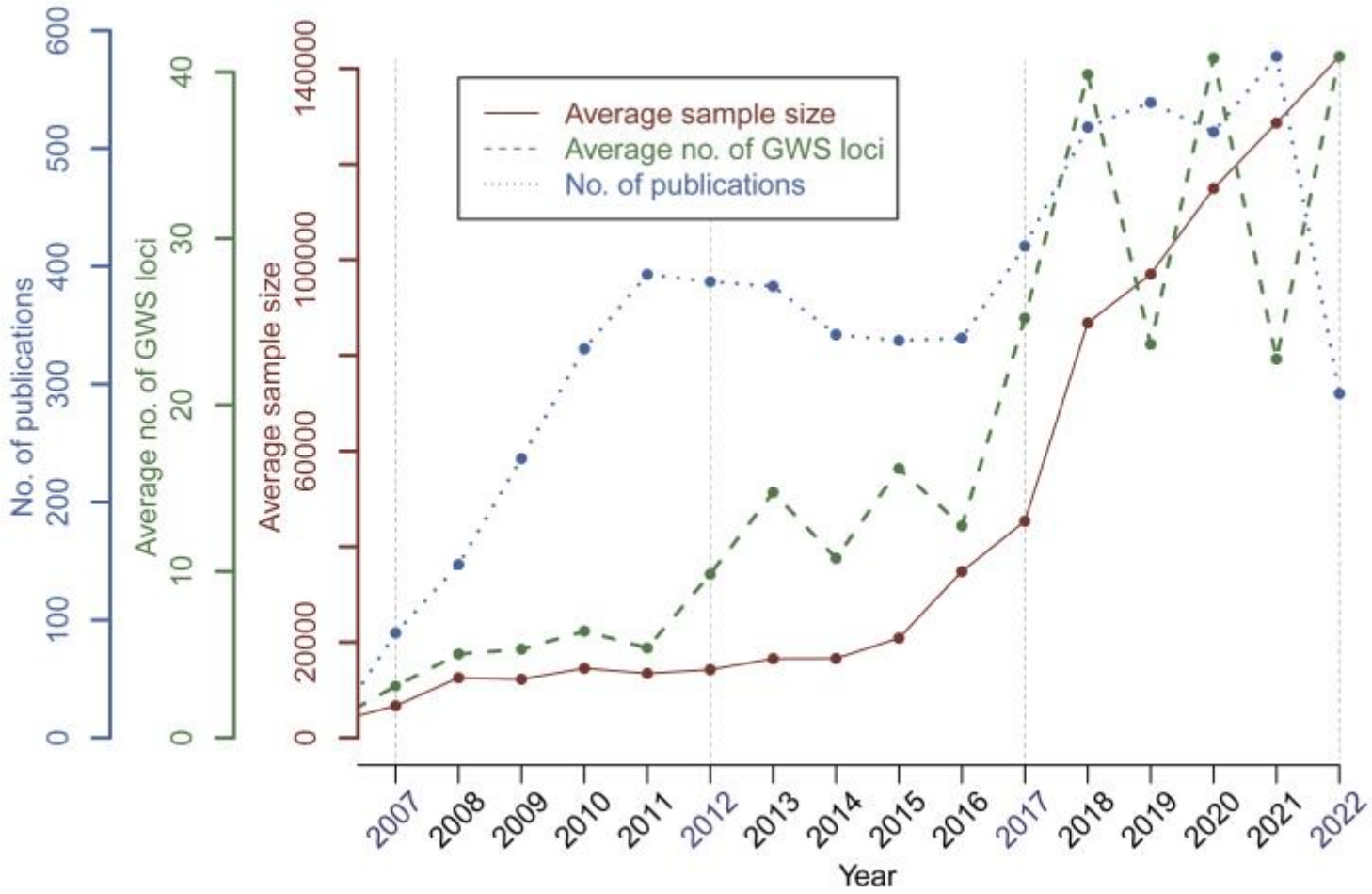


https://www.cdc.gov/diabetes/images/library/socialmedia/diabetesintheus_print.pdf
https://images.app.goo.gl/4CPEMJGXewUafA8x6

- Compare with Mendelian disorders
  - Less concentrated within families
    - large pedigrees (Mendelian), unrelated individuals (GWAS)
  - Larger numbers of variants, genome-wide, contribute to the phenotype
  - Phenotype may be modified to non-genetic risk factors

GWAS

- 2003: Human Genome Project
- 2005: First success stories (Klein et al., 2005)
  - age-related macular degeneration (AMD)
  - 100,000 SNPs
  - 96 cases, 50 controls
- 2007: landmark GWAS (Wellcome Trust Case Control Consortium,2007)
  - 7 complex disease (type 1 diabetes, type 2 diabetes, coronary heart disease, hypertension, bipolar disorder, rheumatoid arthritis, inflammatory bowel disease)
  - 500,000 SNPs
  - Total 2000 cases, 3000 shared controls



. 2023 Jan 11;110(2):179–194. doi: 10.1016/j.ajhg.2022.12.011

Statistical models

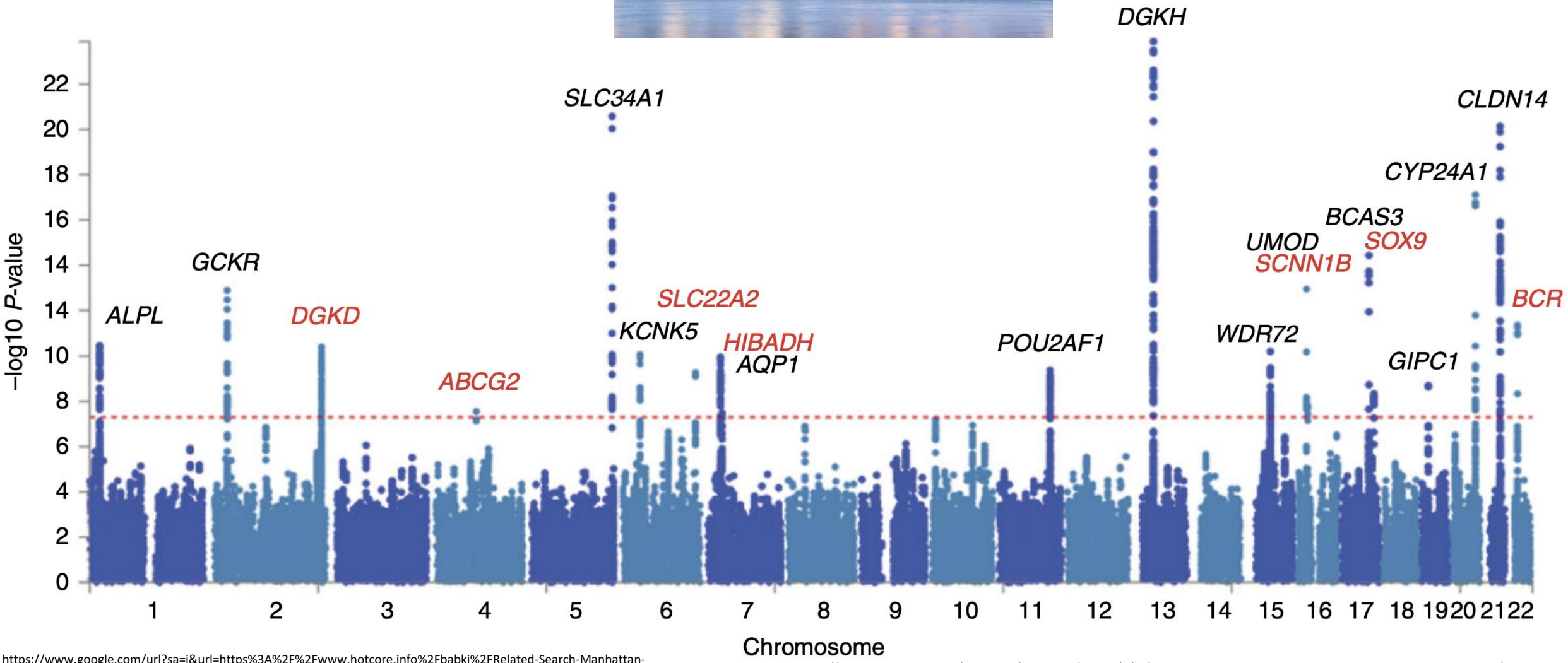
- Continuous trait: linear regression model
  - BMI, height, triglycerides
- Binary trait: logistic regression model
  - Type 1 diabetes, coronary artery disease, lung cancer
  - Case-control studies

Multiple testing problem

- One SNP at a time
- Familywise error rate (FWER), Bonferroni correction
- False discovery rate (FDR), q-values

Manhattan plot

Manhattan skyline
曼哈頓天際線



https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.hotcore.info%2Fbabki%2FRelated-Search-Manhattan-Skyline.htm&psig=AOvVaw3E5IjVcW3WM1s7P6E4o9I2&ust=1695544484837000&source=images&cd=vfe&opi=89978449&ved=0CBAQjRxqFwoTCPDTzMepwIEDFQAAAAAdAAAAABA0

https://upload.wikimedia.org/wikipedia/commons/thumb/4/4f/Manhattan_plot_from_a_GWAS_of_kidney_stone_disease.png/1200px-Manhattan_plot_from_a_GWAS_of_kidney_stone_disease.png

Reproducibility in GWAS

- Reproducibility:

  the ability to confirm genetic associations with traits or diseases by replicating results across independent cohorts, populations, or studies.

- Data Quality control
  - SNPs, 3 allele type
  - Call rate, HWE, sex check, relativeness,…

- Replication: confirm findings in independent cohorts
- P-value threshold (Bonferroni correction, p-value < 5×10^{-8})