Statistical Methods for High Dimensional Biology
STAT/BIOF/GSAT 540

Lecture 22 – Polygenic Risk Scores and Phenome-Wide Association Studies (PheWAS)

March 31, 2021

Jessica Dennis, MSc, PhD

Assistant Professor of Medical Genetics

Dennis Lab in Complex Trait Genomics at UBC and BC Children's Hospital Research Institute

Priority

Genomic determinants of brain-related traits and disorders

Projects

- Prenatal origins of neurodevelopmental disorders and neurodiversity
- Perinatal (maternal) mental health
- Loneliness and depression
- Traumatic brain injury recovery

Themes

Sex differences
Ancestral diversity
Gene-environment interactions
Multi-omics data integration
Longitudinal / trajectory analyses

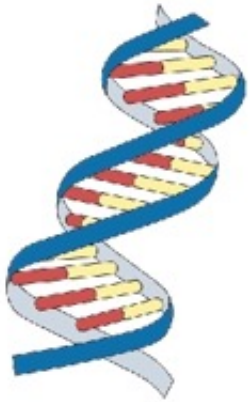
Platforms

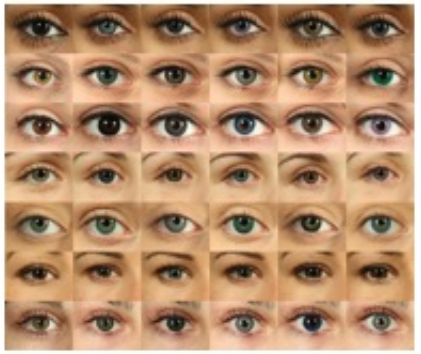
Genomics and 'omics
Health administrative records (e.g., PopData BC, EHRs, biobanks)
Population-based cohort studies (e.g., CHILD, CLSA)
Bioinformatics and data science
Genetic epidemiology and statistical genetics

PhD in Epidemiology from University of Toronto

Postdoctoral Fellow in Genetic Medicine at Vanderbilt University (Nashville, TN)

Genotype vs. Phenotype



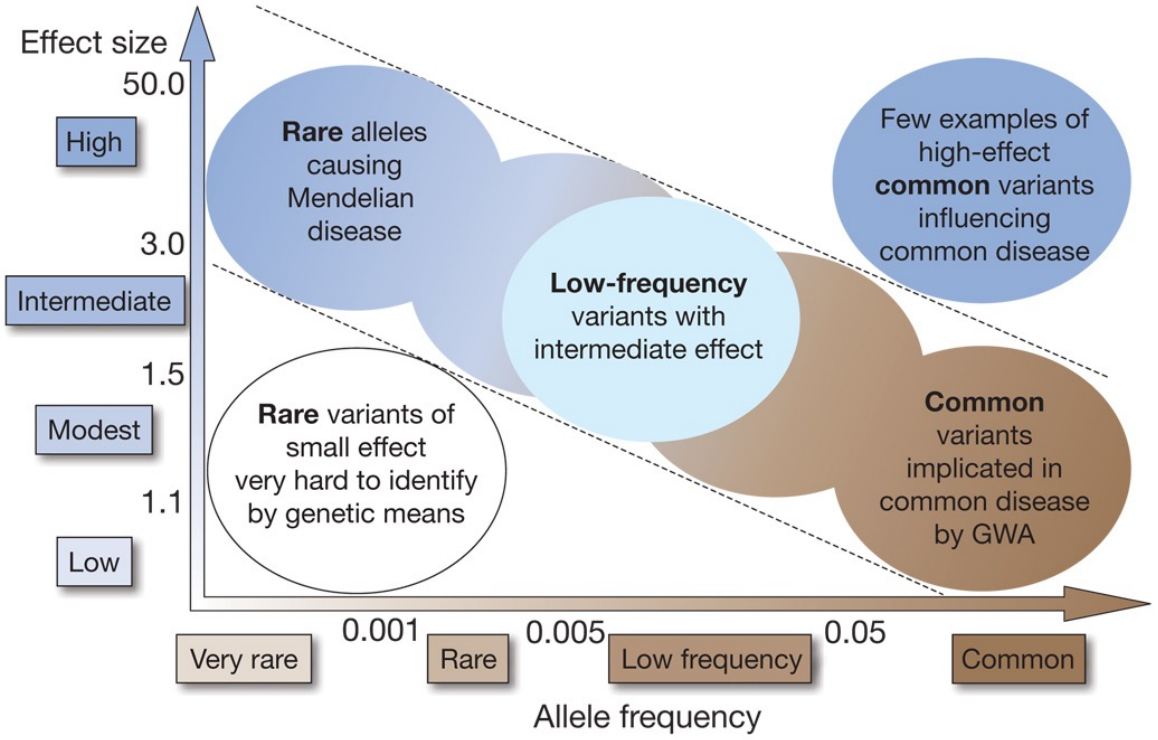


Genotypes are the genetic make-up of an individual.

Phenotypes are the physical traits and characteristics of an individual and can be influenced by their genotype and/or by the environment.
Complex traits are multifactorial phenotypes.

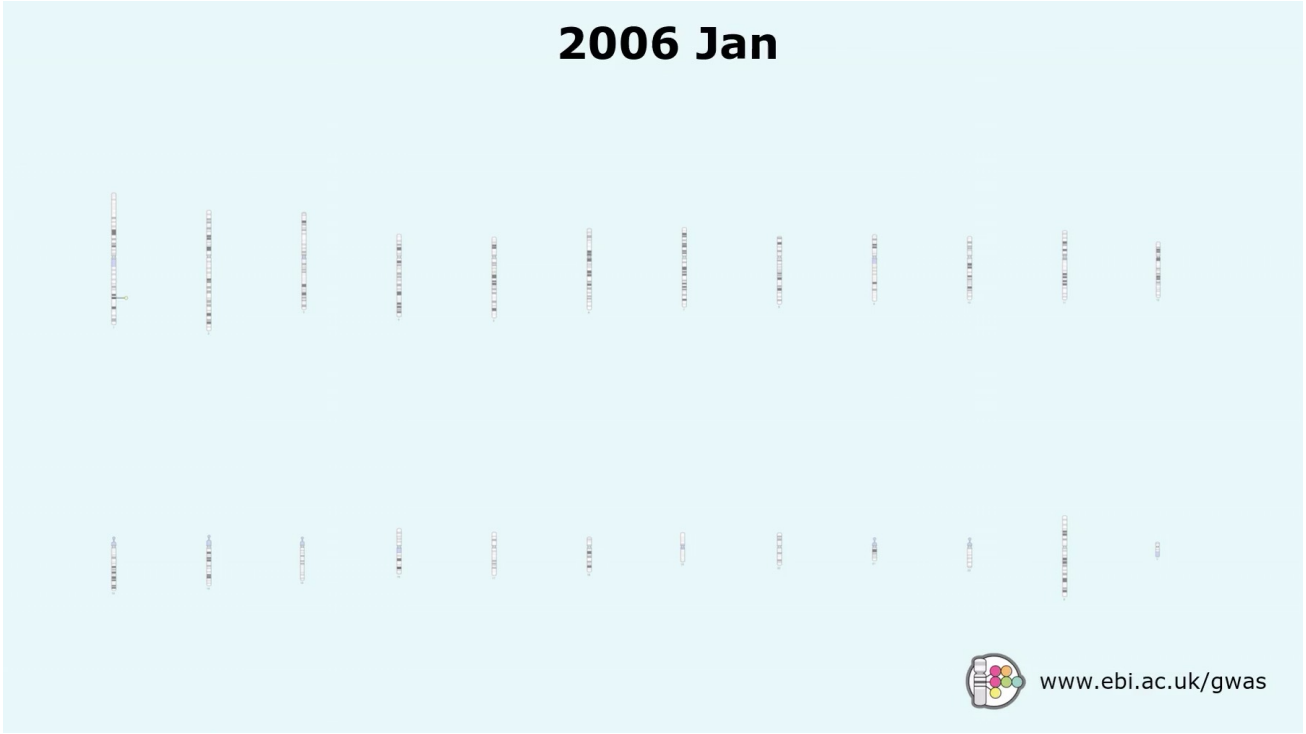
Source: https://www.ebi.ac.uk/training/online/course/human-genetic-variation-i-introduction-2019/what-genetic-variation/genotype-or-phenotype

Common disease common variant model motivating GWAS



Manolio TA et al. Nature. 2009 Oct 8;461(7265):747-53.

15 Years of GWAS Discovery: Now What?



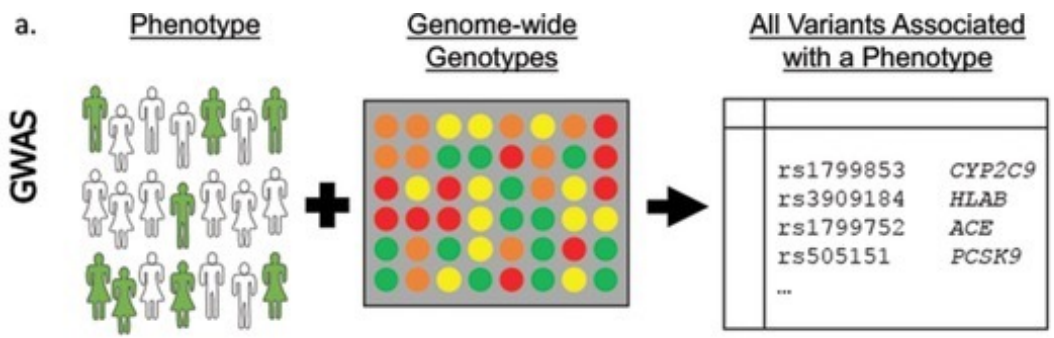
Learning Objectives

1) Explain what GWAS can and cannot tell us about the contribution of genetics to different human traits and diseases.

2) Calculate a polygenic risk score and describe potential applications.

3) Describe data elements contained within a biobank and the types of investigations made possible by these types of resources.

GWAS Study Design



Phenotype could be dichotomous, continuous, categorical, count, etc.

Array genotyping (typically >2 million SNP genotypes) followed by imputation to ~9 million SNP genotypes

Stringent multiple testing correction: $p<5x10^{-8}$

Out of scope: study design, phenotype and genotype QC

Robinson JR et al. Clin Transl Sci. 2018 Mar; 11(2): 112–122.

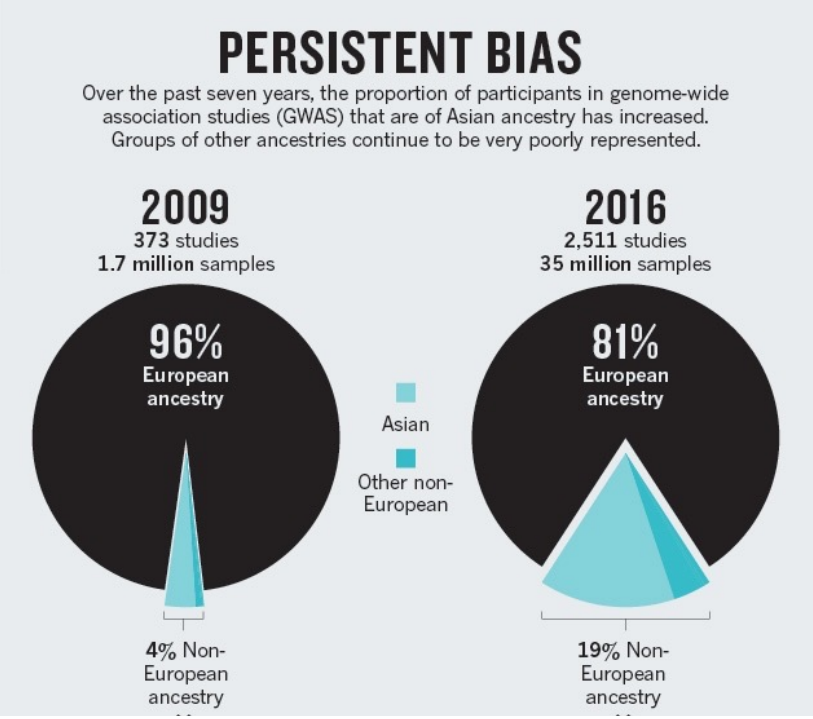
Activity: Explore GWAS catalogue (5 min.)

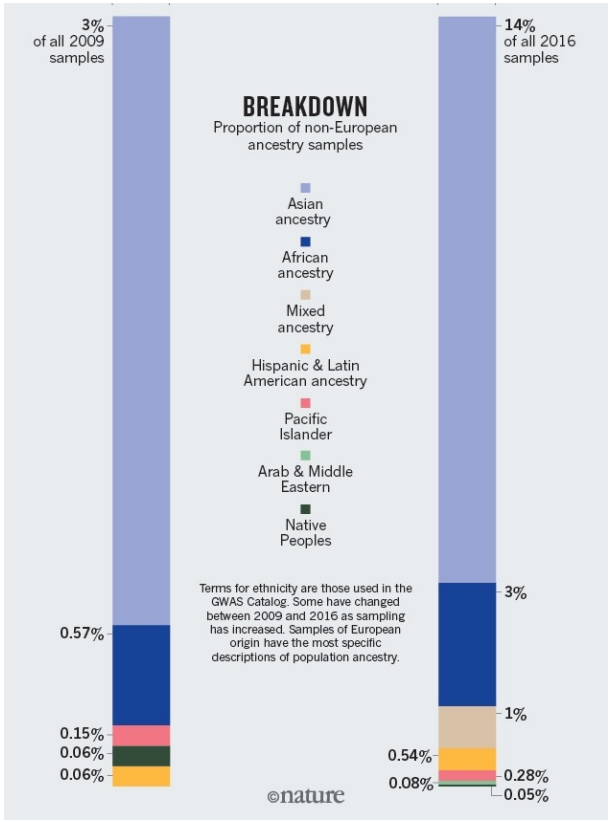
- Diagram: https://www.ebi.ac.uk/gwas/diagram
  - Hover over different dots to see the phenotypes studied
- Search Catalogue: https://www.ebi.ac.uk/gwas/
  - E.g., coronary artery disease
  - E.g., loneliness

Discussion:

1) What kinds of phenotypes have been studied in a GWAS?
2) What is the typical strength of association between genotype and phenotype (OR or Beta)?
3) Who is typically included in GWAS? What are typical sample sizes?
4) What insights do these results provide?

Genomics is failing on diversity





Self-reflection:

How diverse are the data you use in your research? How might this affect interpretation?

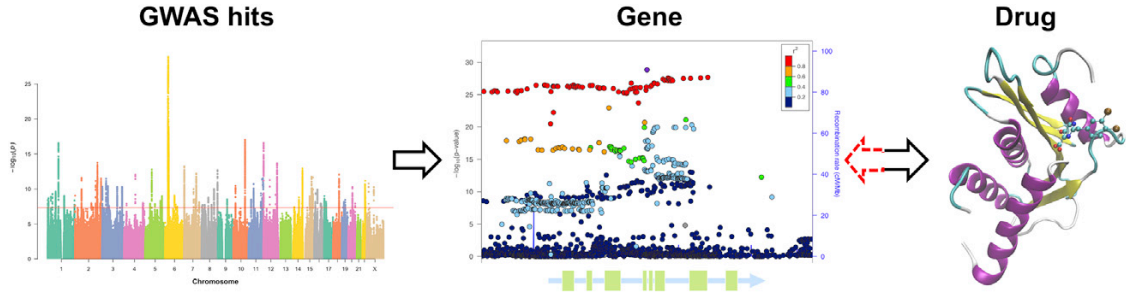
Popejoy AB, Fullerton SM. Nature. 2016 Oct 13;538(7624):161-164.

ASIDE: Race, ethnicity, and genetic ancestry

- ‘‘Race,’’ ‘‘ethnicity,’’ and ‘‘ancestry’’ are often used interchangeably, yet they have no universal definitions.
- **Race**: A culturally and politically charged term, for which definitions and meaning are context-specific. Race is related to individual and/or group identity and is often linked to stereotypes of visible physical attributes such as skin and hair pigmentation.
- **Ethnicity**: Describes people as belonging to cultural groups, usually on the basis of shared language, traditions, foods, etc.
- **(Genetic) ancestry**: a description of the population(s) from which an individual’s recent biological ancestors originated, as reflected in the DNA inherited from those ancestors. Genetic ancestry can be estimated via comparison of participants’ genotypes to global reference populations

Peterson R et al. Cell. 2019 Oct 17;179(3):589-603.

GWAS discoveries can improve drug development



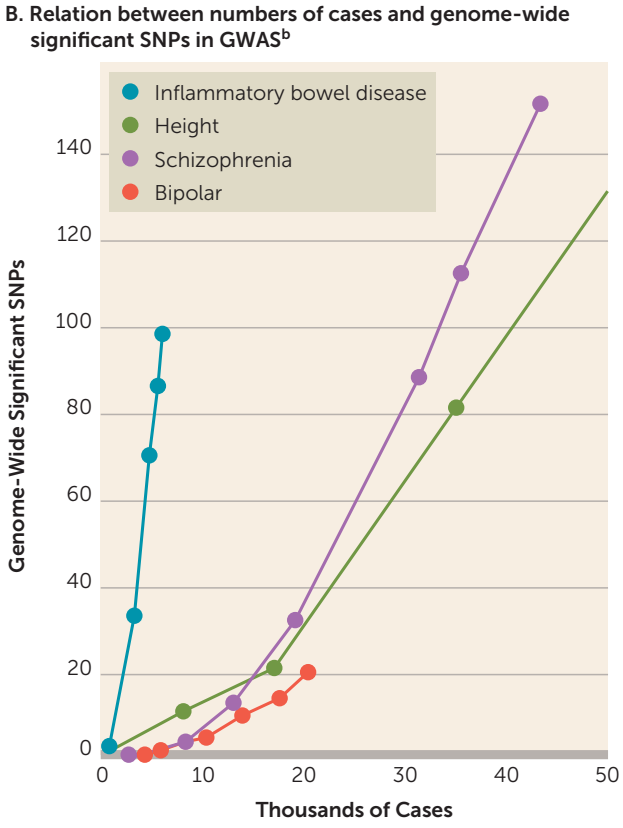
Trait	Gene with GWAS hits	Known or candidate drug
Type 2 Diabetes	*SLC30A8/KCNJ11*	ZnT-8 antagonists/Glyburide
Rheumatoid Arthritis	*PADI4/IL6R*	BB-Cl-amidine/Tocilizumab
Ankylosing Spondylitis(AS)	*TNFR1/PTGER4/TYK2*	TNF-inhibitors/NSAIDs/fostamatinib
Psoriasis(Ps)	*IL23A*	Risankizumab
Osteoporosis	*RANKL/ESR1*	Denosumab/Raloxifene and HRT
Schizophrenia	*DRD2*	Anti-psychotics
LDL cholesterol	*HMGCR*	Pravastatin
AS, Ps, Psoriatic Arthritis	*IL12B*	Ustekinumab

Visscher P et al. Am J Hum Genet. 2017 Jul 6;101(1):5-22.

GWAS can provide insights on genetic architecture

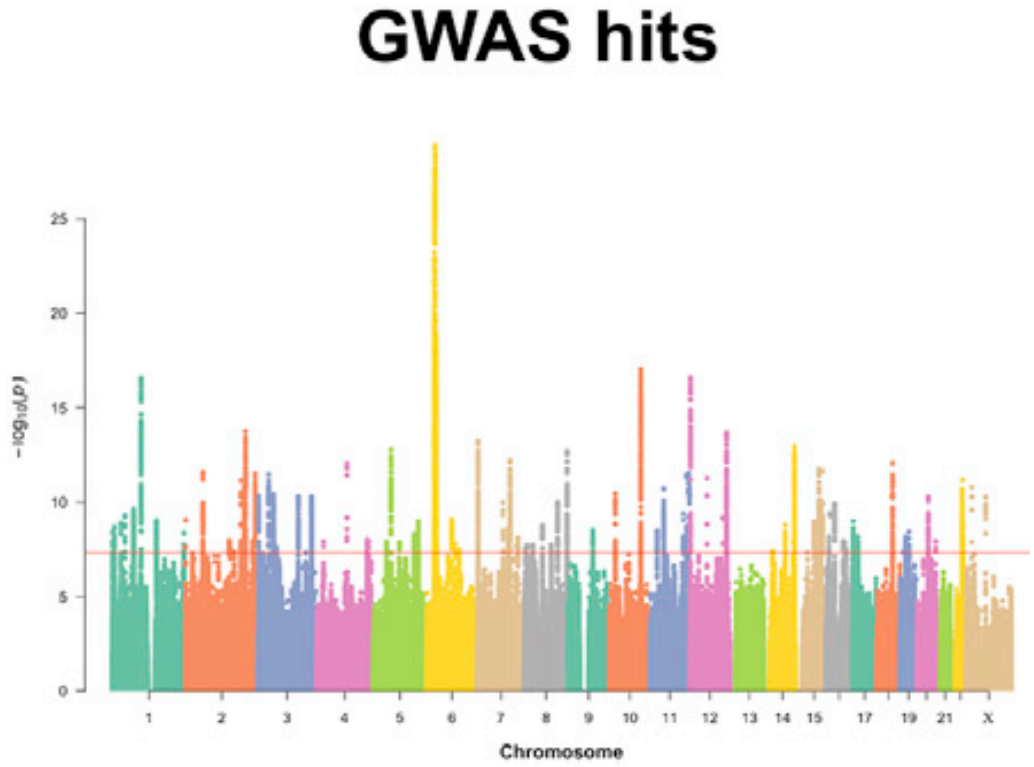
Genetic architecture of a trait refers to the total number of associated variants and, for each, the allele frequencies and the degree of risk conferred.

The concept of genetic architecture is applicable to any trait (e.g., Huntington's disease is caused by a rare, deterministic variant).



Sullivan PF et al. Am J Psychiatry. 2018 Jan 1;175(1):15-27.

Visualizing GWAS results: Manhattan plot



Visscher P et al. Am J Hum Genet. 2017 Jul 6;101(1):5-22.

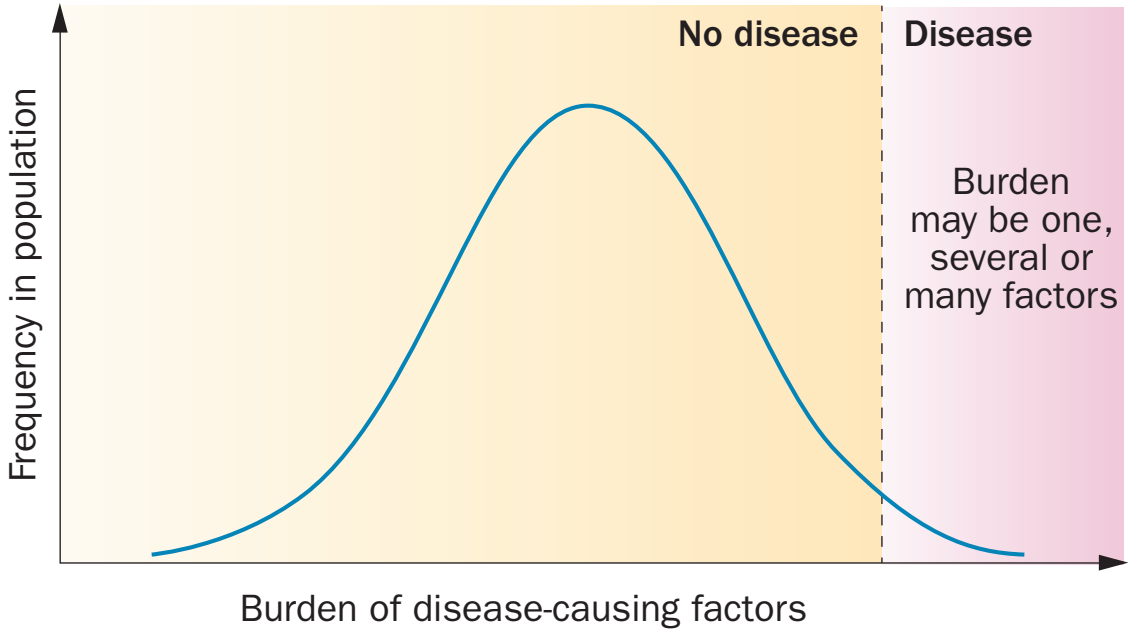
Learning Objectives

1) Explain what GWAS can and cannot tell us about the contribution of genetics to different human traits and diseases.

2) Calculate a polygenic risk score and describe potential applications.

3) Describe data elements contained within a biobank and the types of investigations made possible by these types of resources.

Most complex traits are polygenic



Burden of disease-causing factors can be modeled as a continuous liability in the population.

Disease is manifest in people whose burden exceeds some threshold.

We all carry some genetic burden for disease, which can be quantified by a polygenic score.

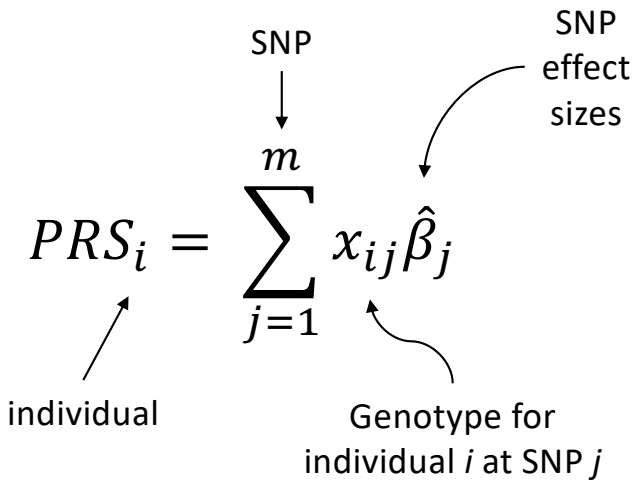
Al-Chalabi A , Hardiman O. Nat Rev Neurol. 2013 Nov;9(11):617-28.

Polygenic Risk Scores (PRS)

Also known as: genetic risk scores; PGS

Method to predict an individual's genetic predisposition for disease

Simplest form:



Polygenic Scores (PRS): Example

$$PRS_i = \sum_{j=1}^{m} x_{ij}\hat{\beta}_j$$

PRS are allelic sums weighted by effect estimates from a discovery GWAS.

Discovery GWAS Summary Statistics

SNP	Allele Effect ($\widehat{\boldsymbol{\beta}}$)
rs1	1.002
rs2	1.03
rs3	1.10
SNP_j	$\widehat{\boldsymbol{\beta}}_j$

Target Sample (Genotyped patients)

ID	PRS
R2348	PRS= 0*1.002 + 2*1.03 + 0*1.10 …
R9842	PRS= 1*1.002 + 1*1.03 + 0*1.10 …

The number of SNPs included in the PRS calculation can range from a small number of genome-wide significant SNPs to several million

Activity: Calculate a PRS (3 min.)

$$PRS_i = \sum_{j=1}^{m} x_{ij}\hat{\beta}_j$$

Schizophrenia GWAS Summary Statistics:

SNP	Allele Effect ($\widehat{\boldsymbol{\beta}}$)	P-value
rs1	0.52	7.2×10^{-12}
rs2	0.44	6.3×10^{-12}
rs3	0.35	9.3×10^{-10}
rs4	0.29	3.6×10^{-7}
rs5	0.27	7.2×10^{-6}
rs6	0.24	7.2×10^{-6}
rs7	0.17	7.2×10^{-4}

Sasha's genotype data:

SNP	Allele copies (x)
rs1	2
rs2	0
rs3	1
rs4	1
rs5	2
rs6	1
rs7	0

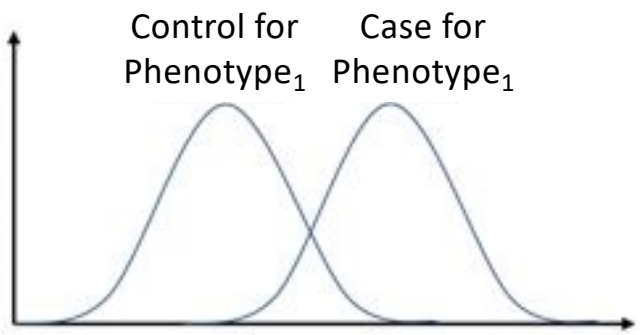
Calculate a PRS for schizophrenia for Sasha, only including genome-wide significant SNPs

Discussion:

1) Now that you have calculated a PRS for schizophrenia in Sasha, what could you do with the score?

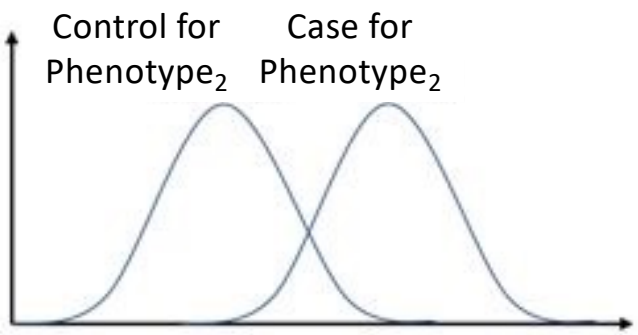
Applications of PRS

Within-Phenotype Association



Polygenic score for Phenotype$_1$

Cross-Phenotype Association

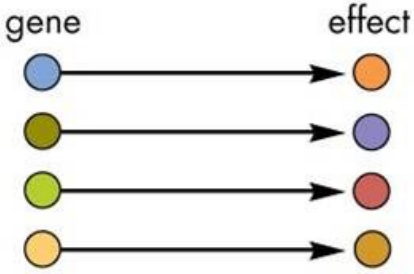


Polygenic score for Phenotype$_1$

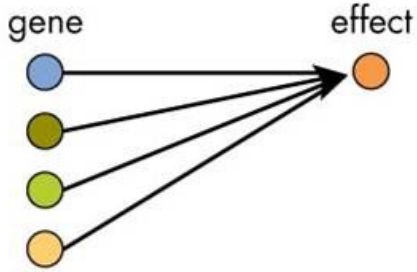
What would this tell us?

Smoller JW et al. Mol Psychiatry. 2018 Jan 9.

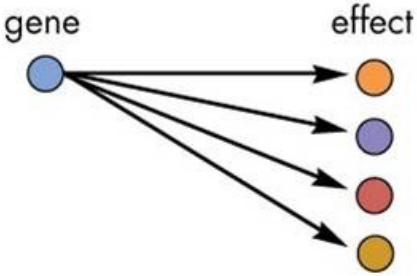
Polygenicity vs. pleiotropy



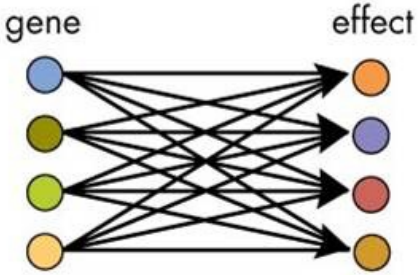
Each gene has a distinct biological effect.



Polygenic trait: Many genes contribute to a single effect.



Pleiotropy: A gene has multiple effects.

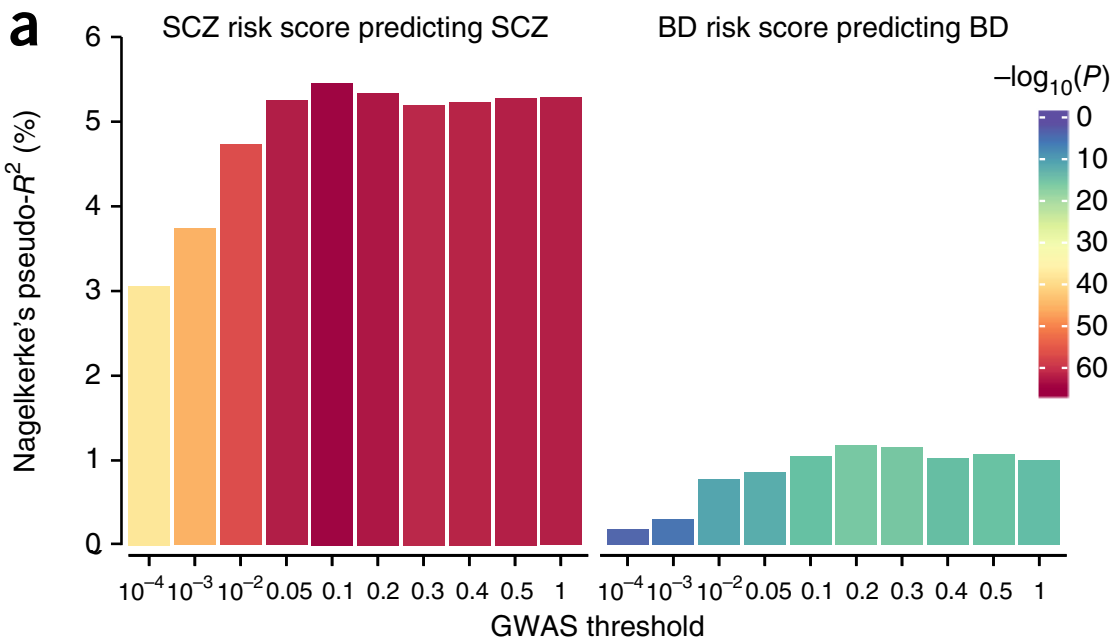


Polygenic traits and pleiotropy

Pleiotropy describes the **genetic** effect of a single **gene** on multiple phenotypic traits. The underlying mechanism is **genes** that code for a product that is either used by various cells or has a cascade-like signaling function that affects various targets.

E.g., Mutations in the *CFTR* gene cause a multitude of symptoms in cystic fibrosis patients

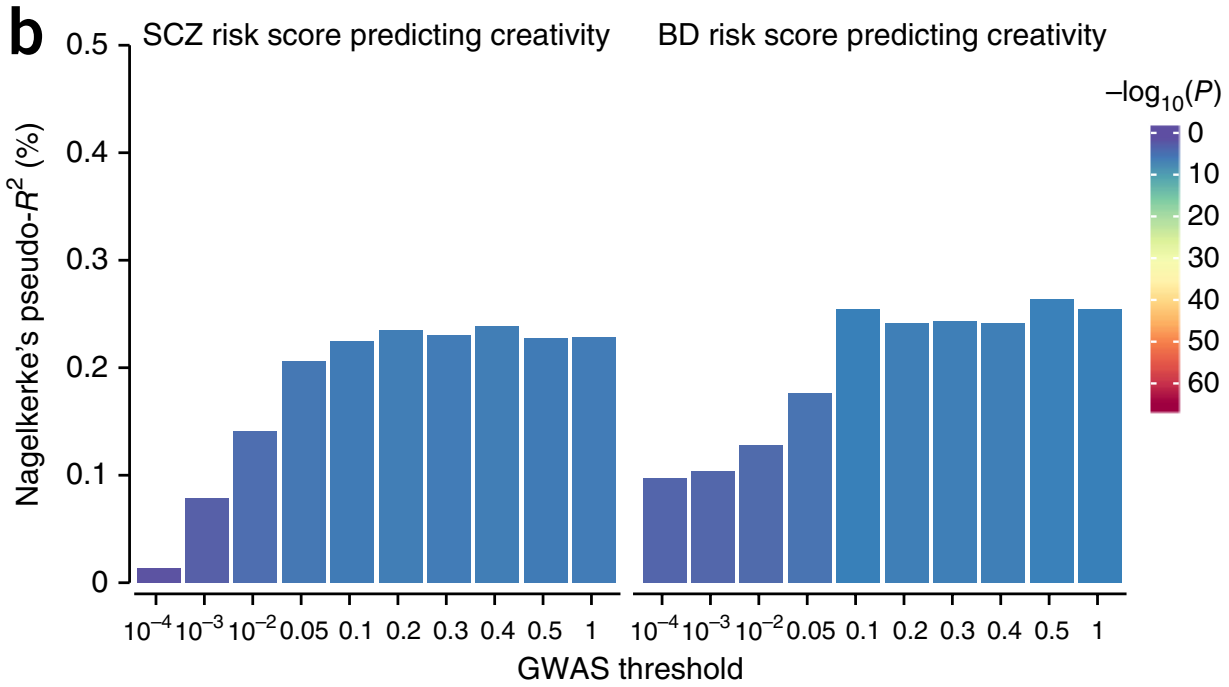
Application of PRS: Within-phenotype associations



R^2: Proportion of the variance for the dependent variable that is explained by the independent variable

Power RA et al. Nat Neurosci. 2015 Jul;18(7):953-5.

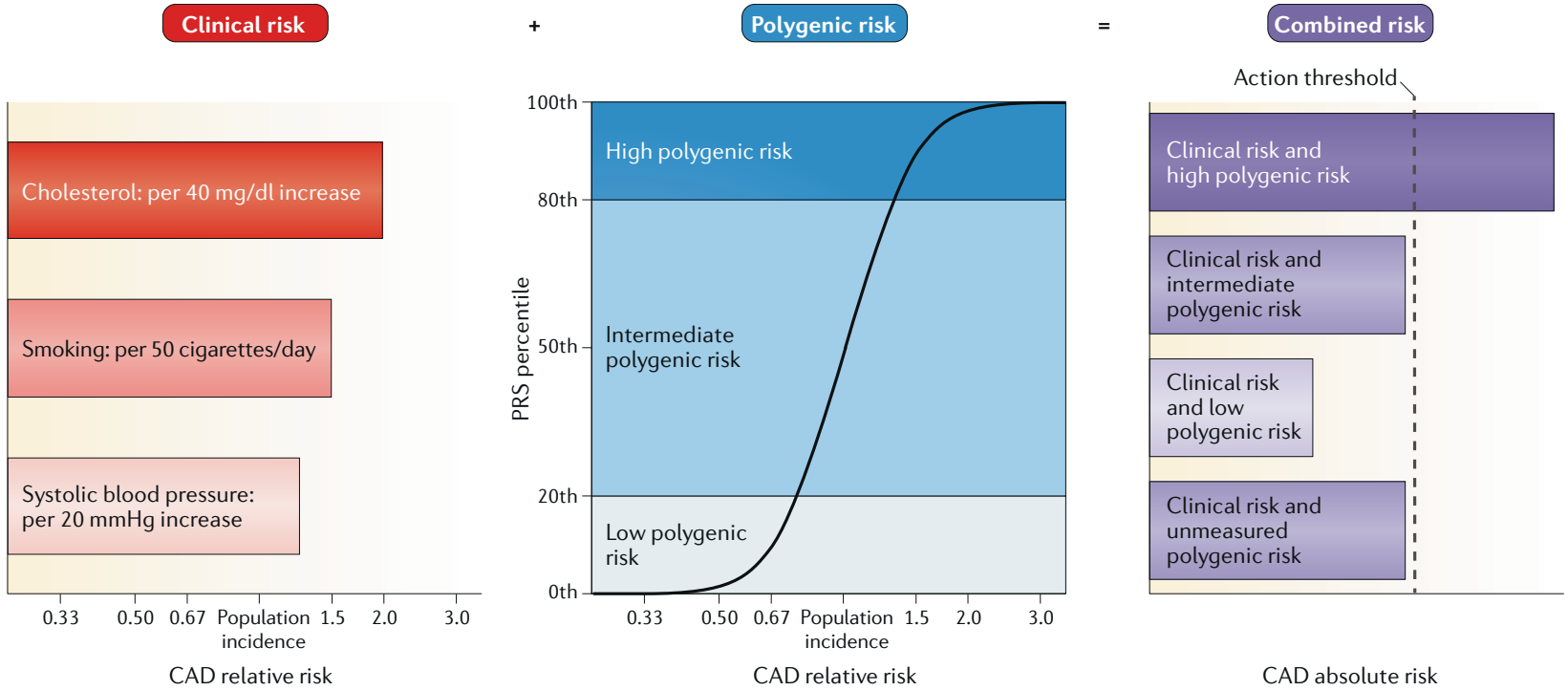
Application of PRS: Across-phenotype associations



R^2: Proportion of the variance for the dependent variable that is explained by the independent variable

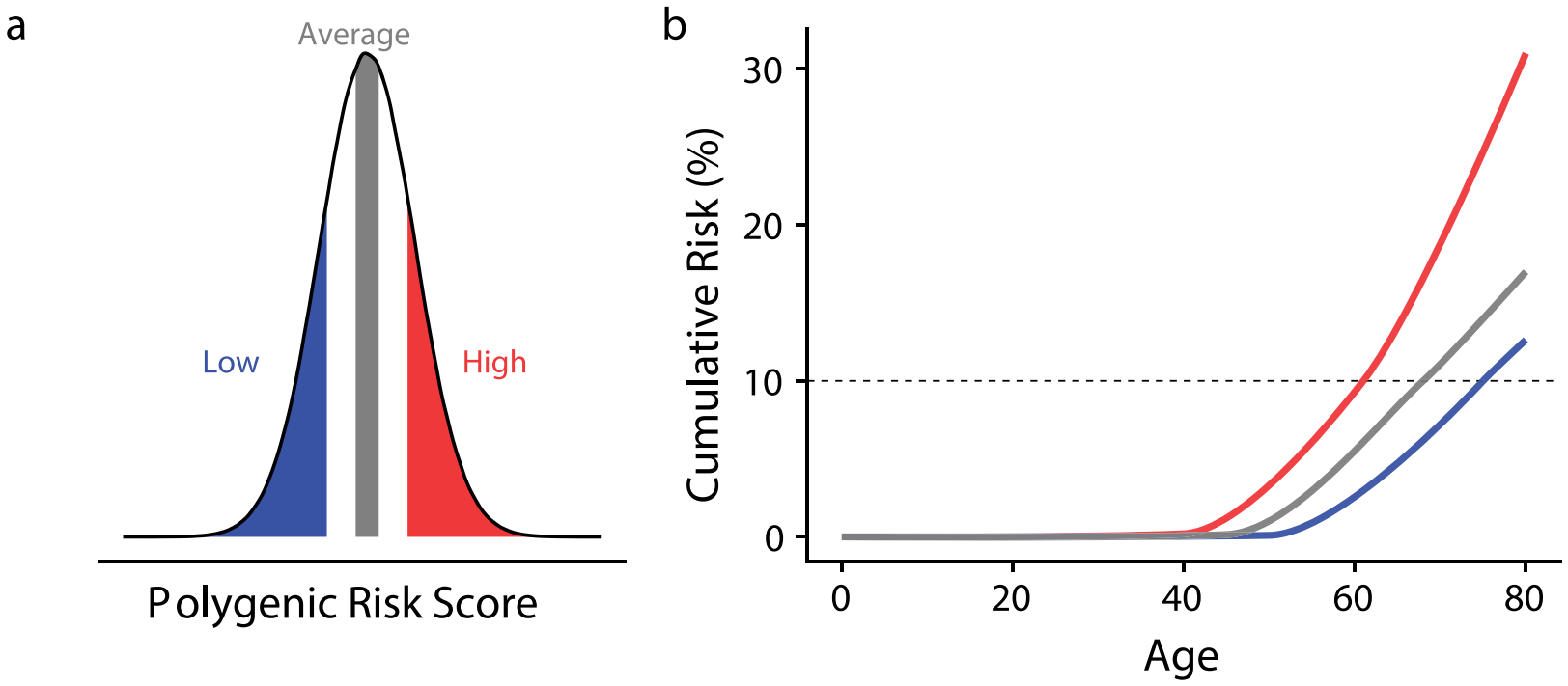
Power RA et al. Nat Neurosci. 2015 Jul;18(7):953-5.

Clinical Applications of PRS



Torkamani A et al. Nat Rev Genet. 2018 Sep;19(9):581-590.

Clinical Applications of PRS



Lambert SA et al. Hum Mol Genet. 2019 Nov 21;28(R2):R133-R142.

Learning Objectives

1) Explain what GWAS can and cannot tell us about the contribution of genetics to different human traits and diseases.

2) Calculate a polygenic risk score and describe potential applications.

3) Describe data elements contained within a biobank and the types of investigations made possible by these types of resources.

Biobanks

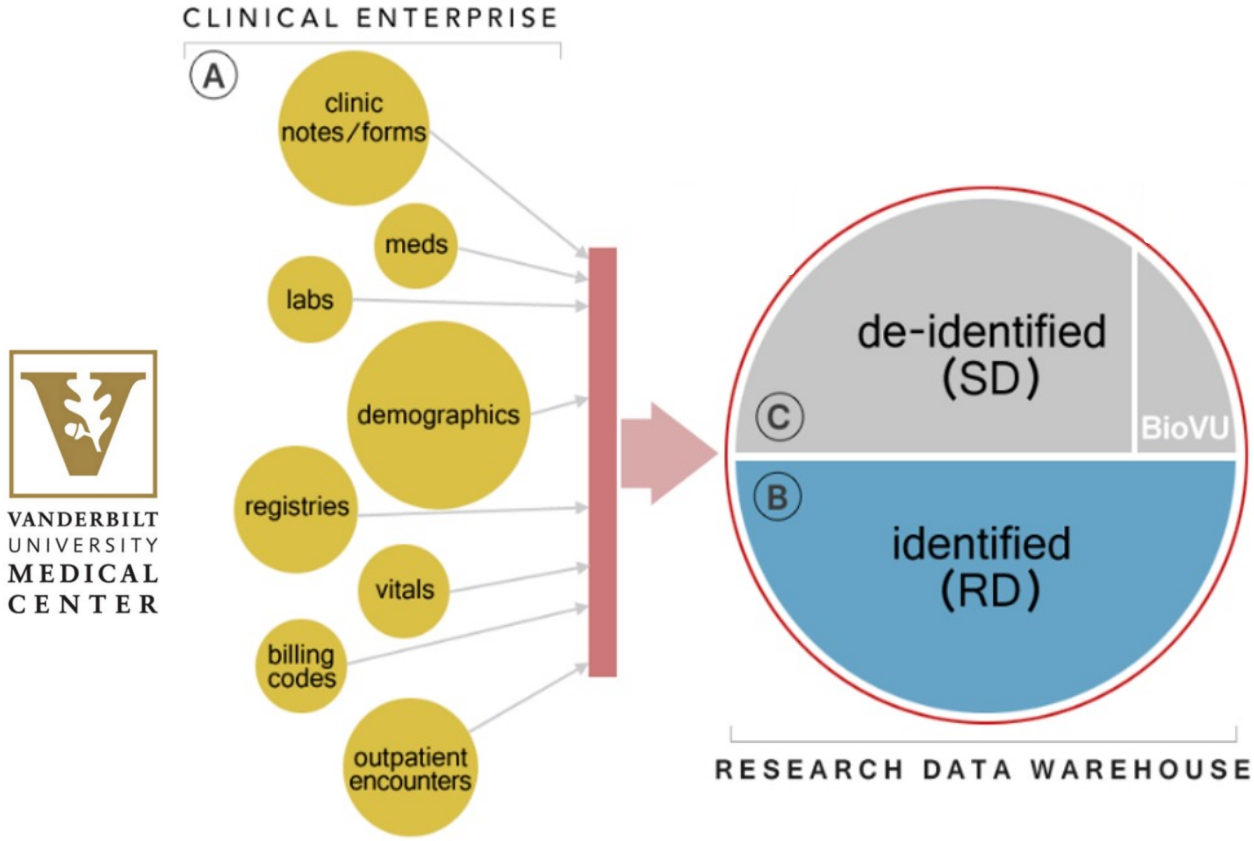
Collections of biological specimens linked to phenotype data.

Used for health research.

What biobanks have you heard of?

Million Veteran Program

Biobanks: BioVU



SD: 2.8 million patient EHRs

BioVU: 285,000 patients EHRs linked to DNA samples

Average EHR length of BioVU patients is 10 years

Danciu I et al. J Biomed Inform. 2014 Dec;52:28-35.

UK Biobank – explore YouTube videos, website

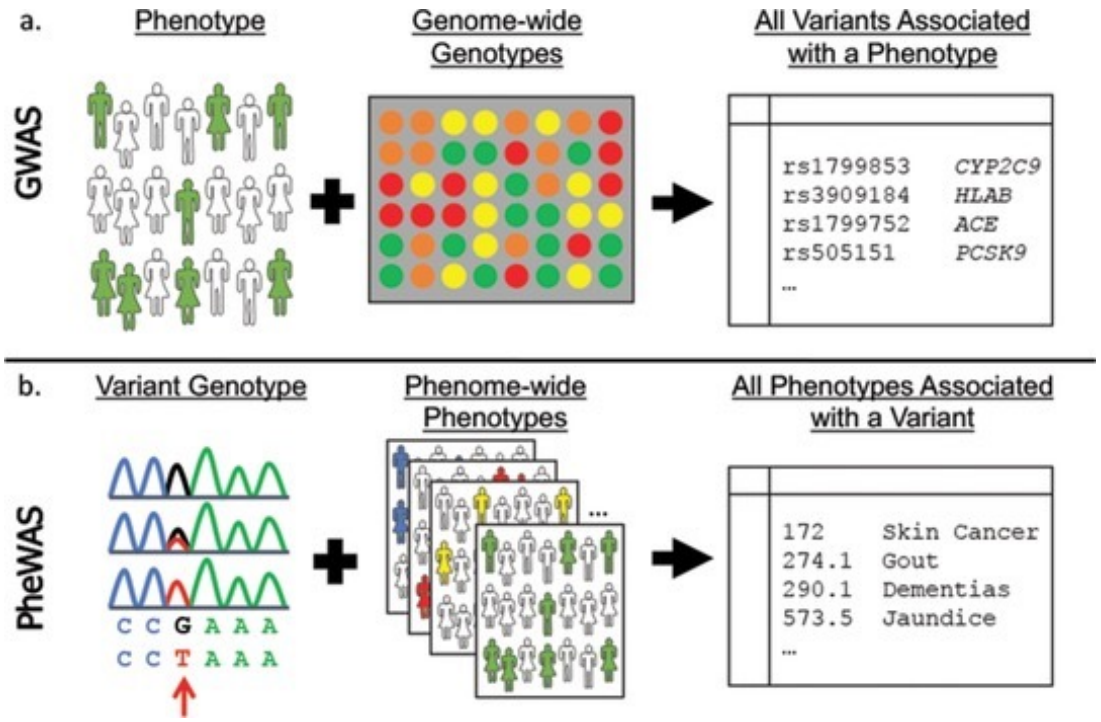
Activity: Explore the UK Biobank Data Showcase (5 min.)

http://biobank.ndph.ox.ac.uk/showcase/

Discussion:

1) Who is included in these studies?
2) How generalizable are the results?
3) What would you do with these data?

Phenome-wide association study (PheWAS)



Robinson JR et al. Clin Transl Sci. 2018 Mar; 11(2): 112–122.

Activity: Explore polygenic burden associations across the human phenome (5 min.)

- http://mrcieu.mrsoftware.org/PRS_atlas/

  - E.g., PRS for “Neuroticism” in a PheWAS

Discussion:

1) What are the strengths and weaknesses of PheWAS?
2) What are the strengths and weaknesses of PRS-PheWAS?

Learning Objectives and Wrap-Up (5 min.)

1) Explain what GWAS can and cannot tell us about the contribution of genetics to different human traits and diseases.

2) Calculate a polygenic risk score and describe potential applications.

3) Describe data elements contained within a biobank and the types of investigations made possible by these types of resources.