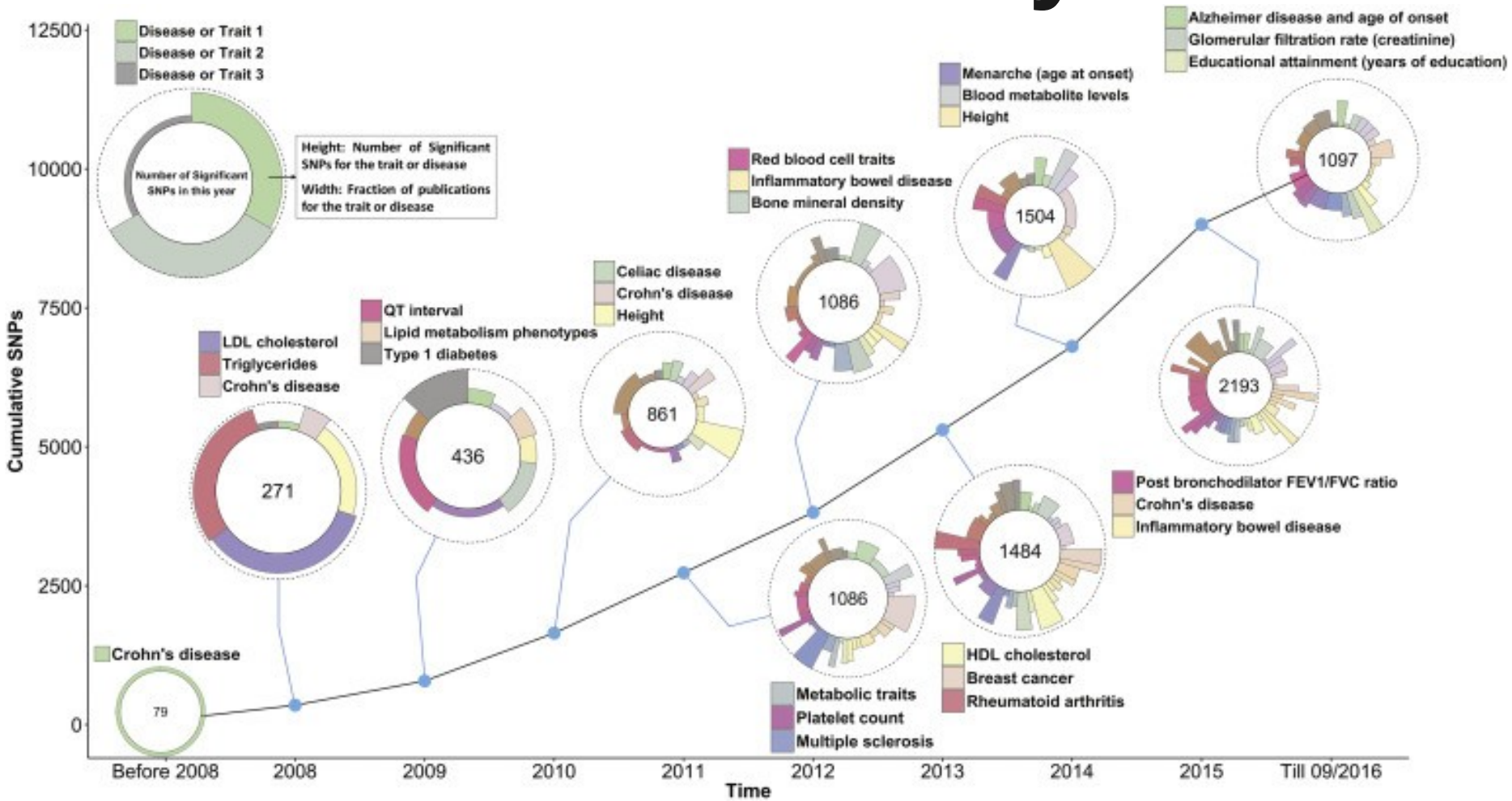
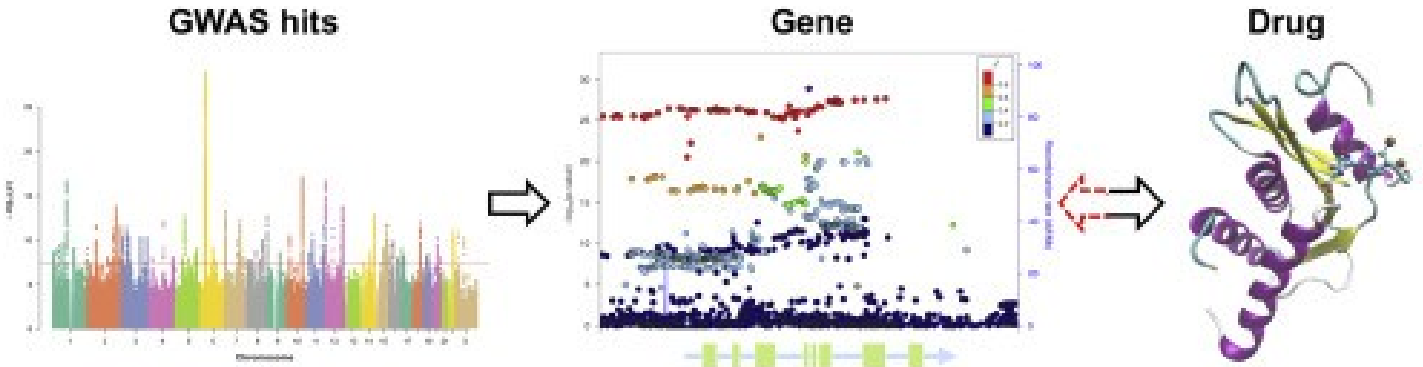
Genome-wide Association Studies (GWAS) and disease

GWAS SNP Discovery Timeline



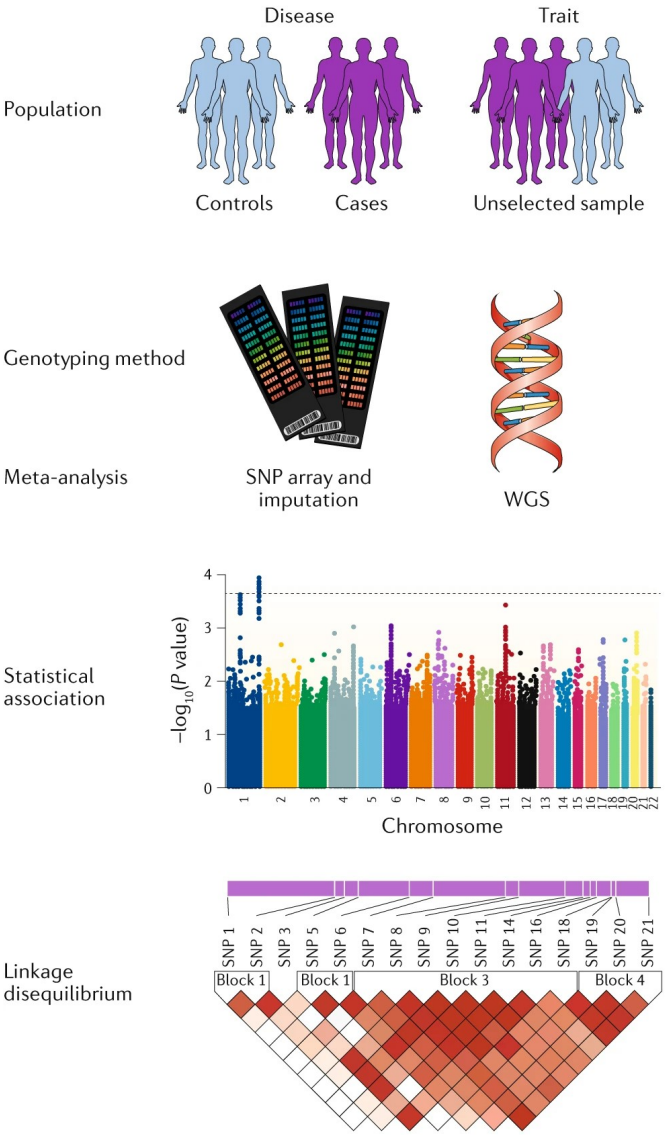
Visscher, Peter M., et al. "10 years of GWAS discovery: biology, function, and translation." The American Journal of Human Genetics 101.1 (2017): 5-22. https://doi.org/10.1016/j.ajhg.2017.06.005

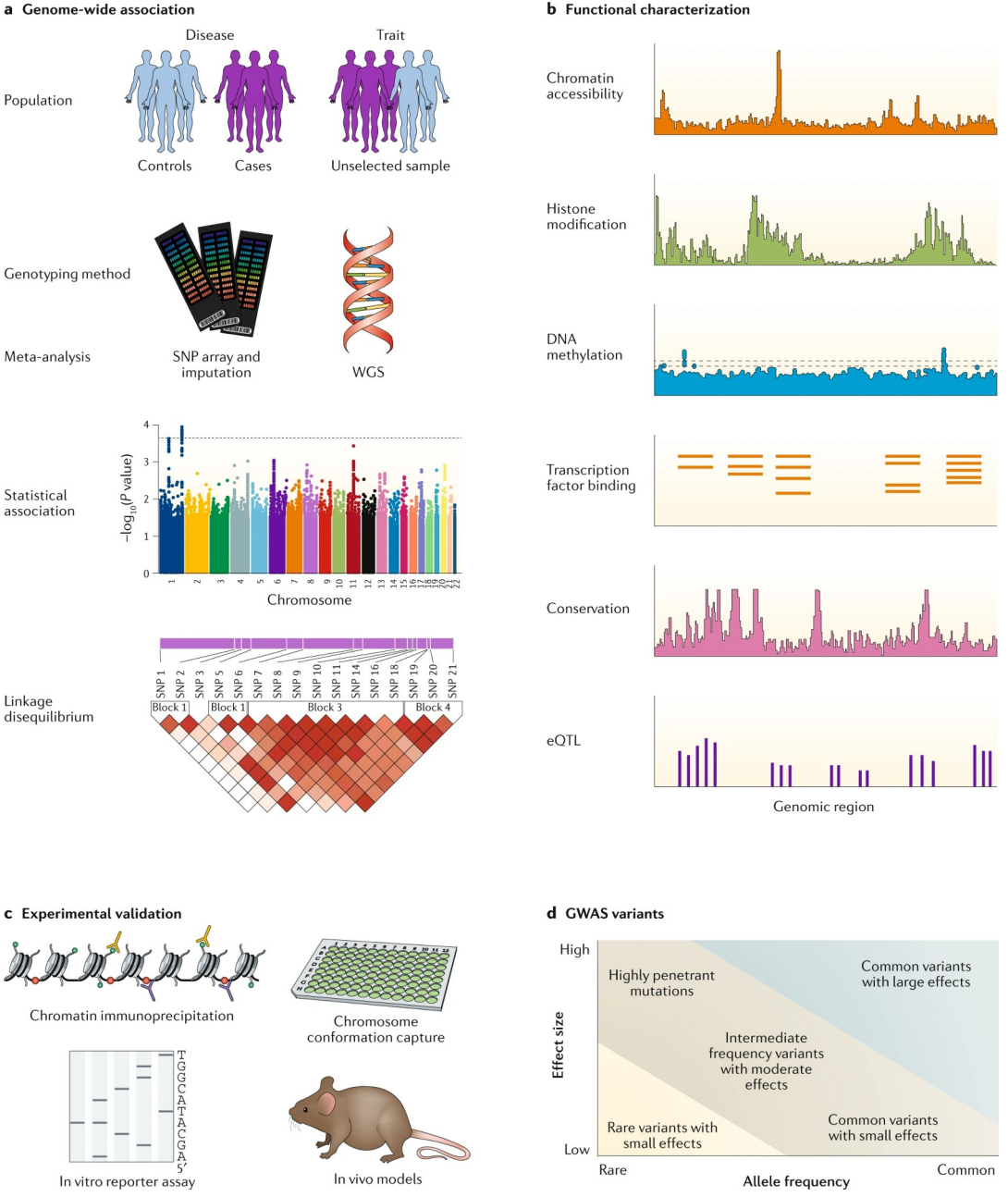
GWAS discoveries and drugs



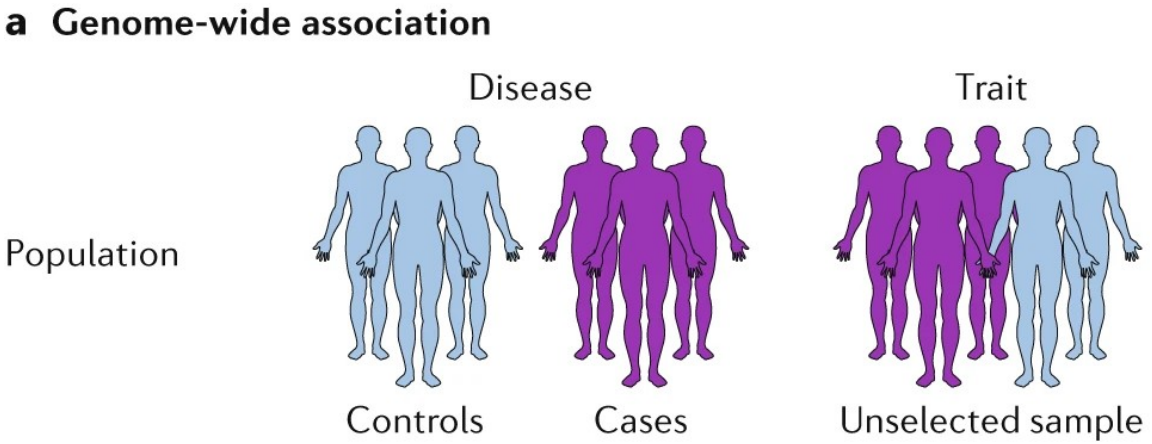
Trait	Gene with GWAS hits	Known or candidate drug
Type 2 Diabetes	*SLC30A8/KCNJ11*	ZnT-8 antagonists/Glyburide
Rheumatoid Arthritis	*PADI4/IL6R*	BB-Cl-amidine/Tocilizumab
Ankylosing Spondylitis(AS)	*TNFR1/PTGER4/TYK2*	TNF-inhibitors/NSAIDs/fostamatinib
Psoriasis(Ps)	*IL23A*	Risankizumab
Osteoporosis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

Visscher, Peter M., et al. "10 years of GWAS discovery: biology, function, and translation." The American Journal of Human Genetics 101.1 (2017): 5-22. https://doi.org/10.1016/j.ajhg.2017.06.005


a Genome-wide association
Disease
Trait
Population
Controls
Cases
Unselected sample
Genotyping method
Meta-analysis
SNP array and imputation
WGS
Statistical association
−log10(P value)
Chromosome
Linkage disequilibrium
SNP 1
SNP 2
SNP 3
SNP 5
SNP 6
SNP 7
SNP 8
SNP 9
SNP 10
SNP 11
SNP 14
SNP 16
SNP 18
SNP 19
SNP 20
SNP 21
Block 1
Block 1
Block 3
Block 4


a Genome-wide association
Disease
Trait
Population
Controls
Cases
Unselected sample
Genotyping method
Meta-analysis
SNP array and imputation
WGS
Statistical association
−log10(P value)
Chromosome
Linkage disequilibrium
SNP 1
SNP 2
SNP 3
SNP 5
SNP 6
SNP 7
SNP 8
SNP 9
SNP 10
SNP 11
SNP 14
SNP 16
SNP 18
SNP 19
SNP 20
SNP 21
Block 1
Block 1
Block 3
Block 4
b Functional characterization
Chromatin accessibility
Histone modification
DNA methylation
Transcription factor binding
Conservation
eQTL
Genomic region
c Experimental validation
Chromatin immunoprecipitation
Chromosome conformation capture
In vitro reporter assay
In vivo models
d GWAS variants
High
Low
Effect size
Highly penetrant mutations
Common variants with large effects
Intermediate frequency variants with moderate effects
Rare variants with small effects
Common variants with small effects
Rare
Common
Allele frequency

Study population



- could be cases and controls for a disease
- or a population sample of a trait

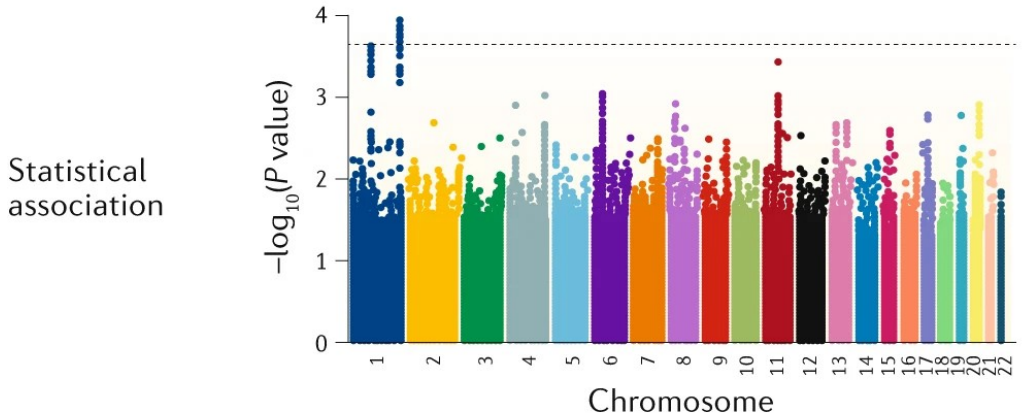
Tam, V., Patel, N., Turcotte, M. et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet 20, 467–484 (2019). https://doi.org/10.1038/s41576-019-0127-1

Genotyping method

- SNP array (cheap)
- whole exome sequencing (all exons i.e. coding)
- whole genome sequencing (expensive)

Tam, V., Patel, N., Turcotte, M. et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet 20, 467–484 (2019). https://doi.org/10.1038/s41576-019-0127-1

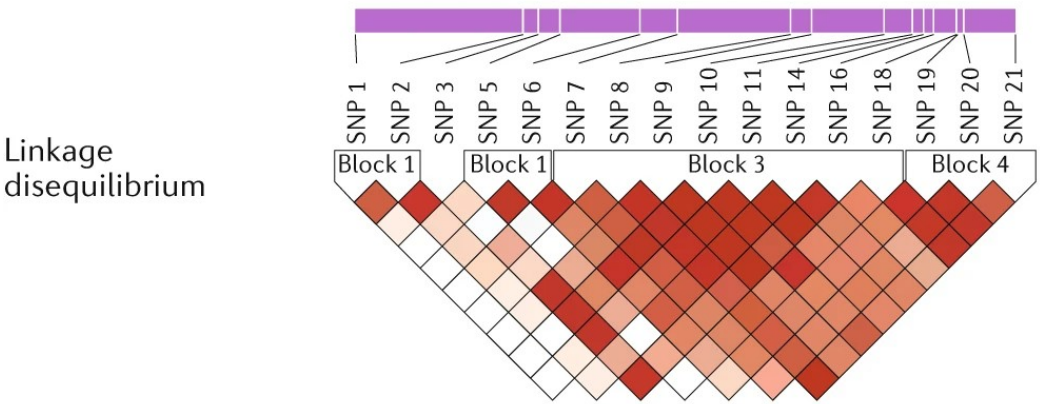
Statistical association



- e.g. logistic regression for binary (disease vs. healthy)
- e.g. linear regression for quantitative traits
- many other possible analyses

Tam, V., Patel, N., Turcotte, M. et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet 20, 467–484 (2019). https://doi.org/10.1038/s41576-019-0127-1

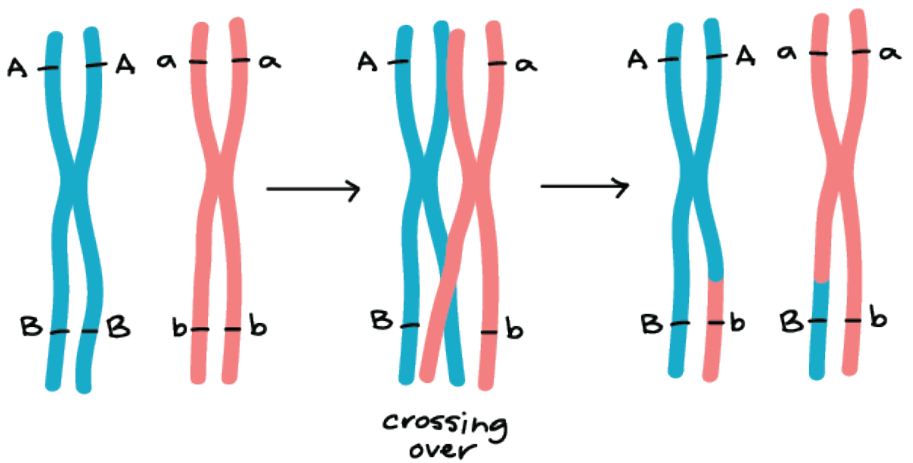
Linkage disequilibrium (LD)



- linkage equilibrium = random association of SNPs/alleles at different loci
- in reality many non-random associations, or linkage disequilibrium

Tam, V., Patel, N., Turcotte, M. et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet 20, 467–484 (2019). https://doi.org/10.1038/s41576-019-0127-1

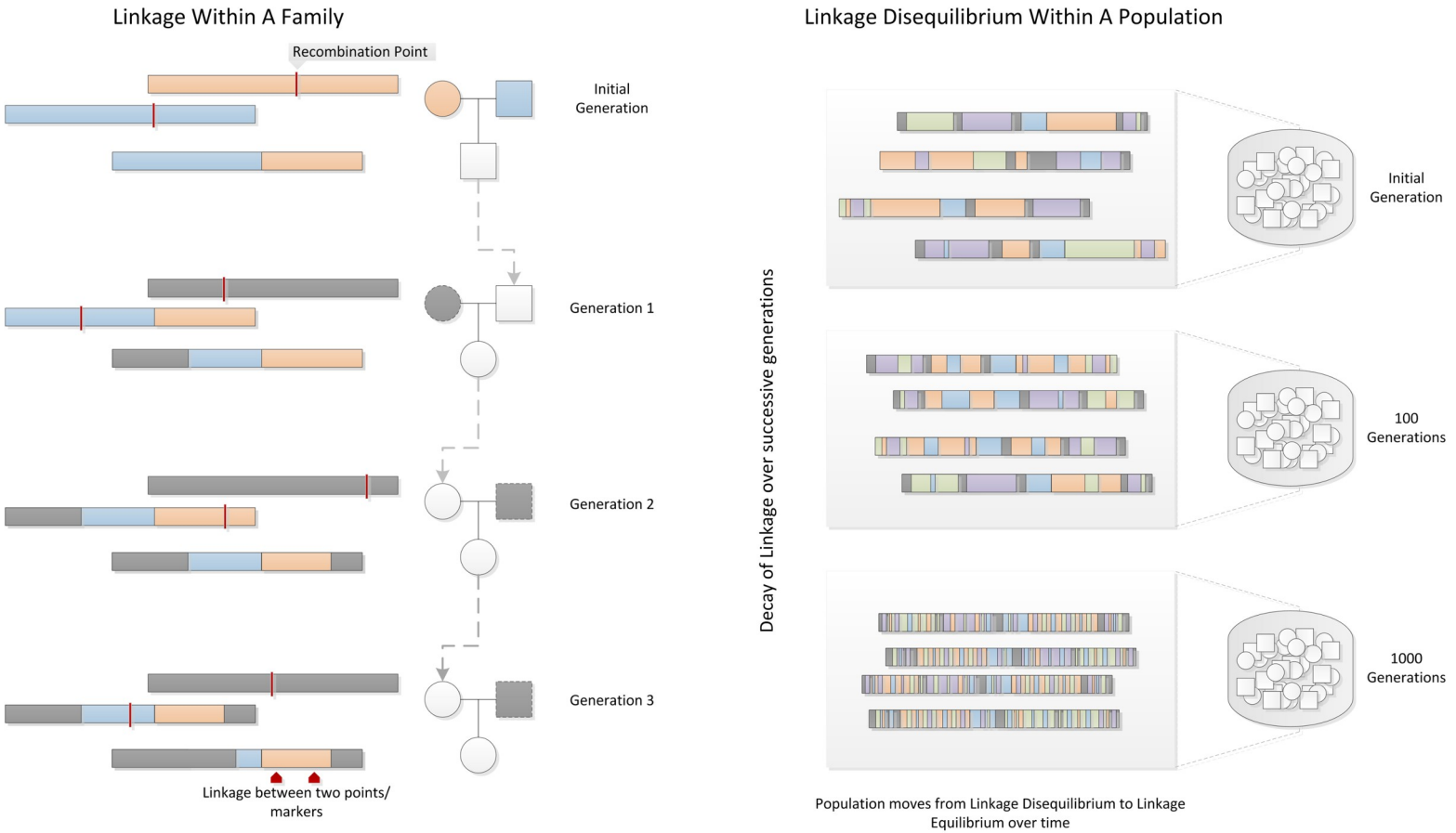
Crossing over



- Crossing over (homologous recombination) occurs during meiosis
- New allele combinations can occur in gametes

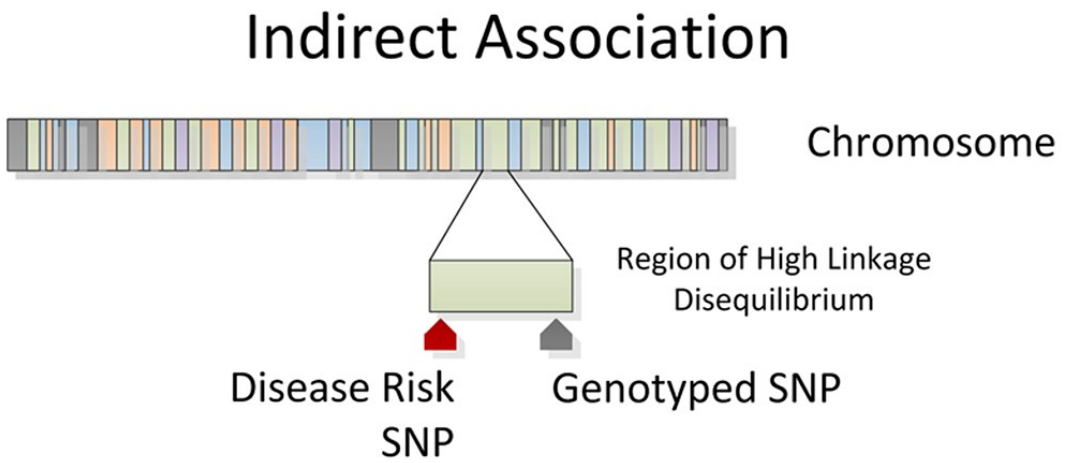
https://www.khanacademy.org/science/ap-biology/heredity/non-mendelian-genetics/a/linkage-mapping

Linkage disequilibrium (LD)



Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8(12): e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Linkage disequilibrium (LD)



- Indirect association – genotyped SNP in high LD with (un-genotyped) disease SNP

Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8(12): e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Missing heritability

Table 1 | Estimates of heritability and number of loci for several complex traits

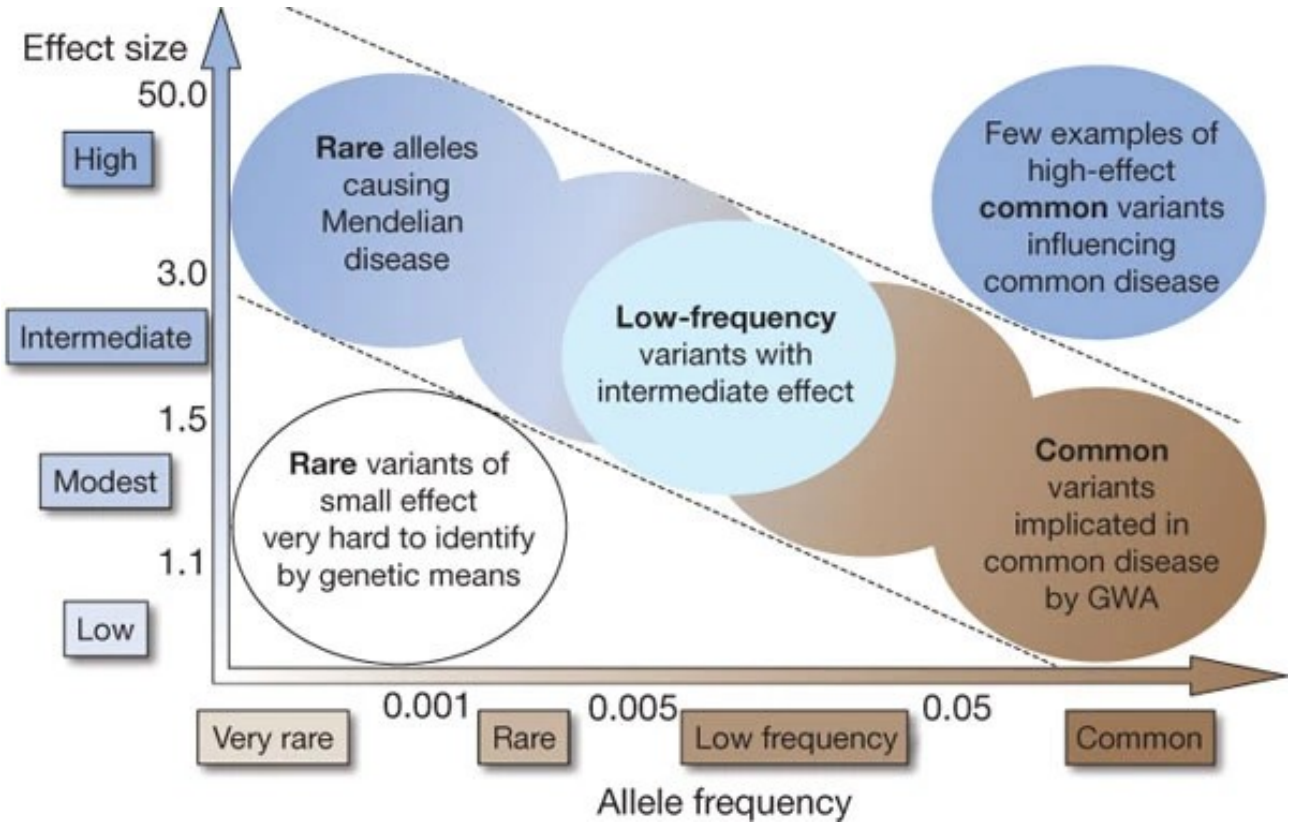
Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration[72]	5	50%	Sibling recurrence risk
Crohn's disease[21]	32	20%	Genetic risk (liability)
Systemic lupus erythematosus[73]	6	15%	Sibling recurrence risk
Type 2 diabetes[74]	18	6%	Sibling recurrence risk
HDL cholesterol[75]	7	5.2%	Residual* phenotypic variance
Height[15]	40	5%	Phenotypic variance
Early onset myocardial infarction[76]	9	2.8%	Phenotypic variance
Fasting glucose[77]	4	1.5%	Phenotypic variance

* Residual is after adjustment for age, gender, diabetes.

- GWAS variants often have small effect
- May explain small proportion of heritability

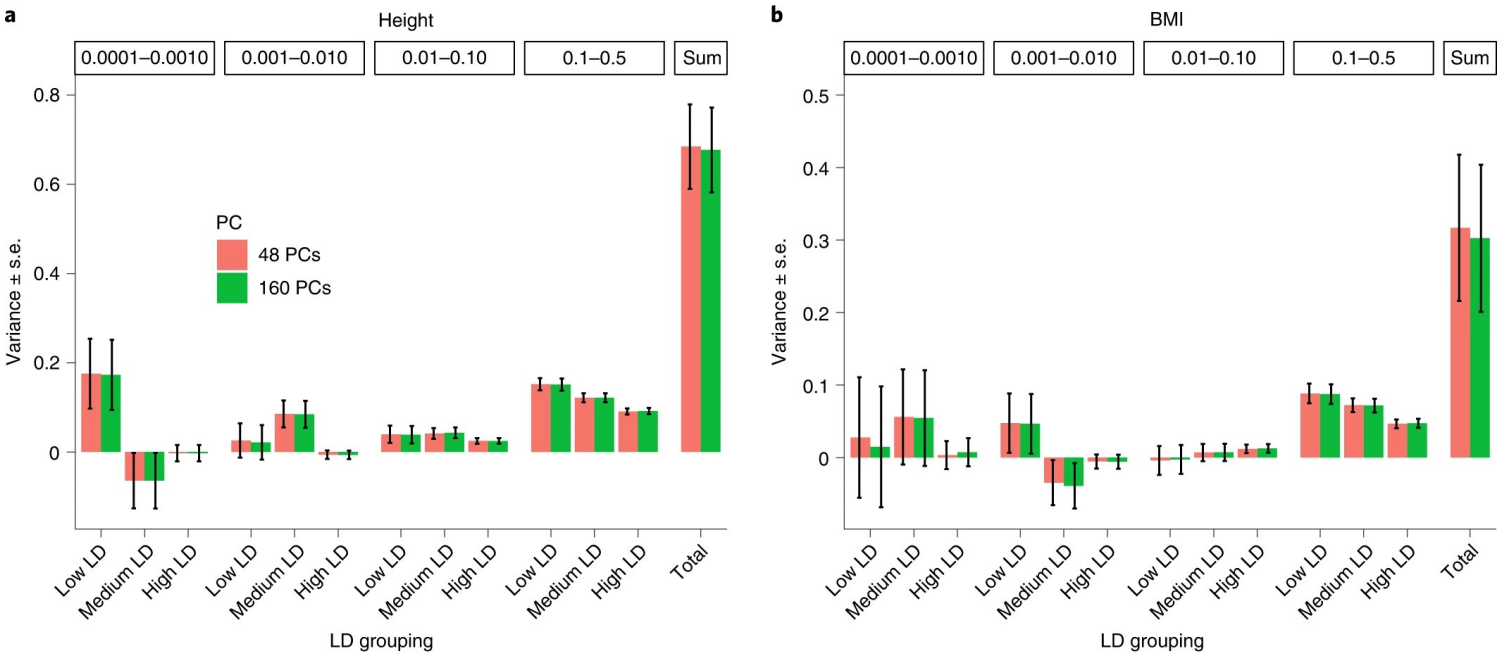
Manolio, T., Collins, F., Cox, N. et al. Finding the missing heritability of complex diseases. Nature 461, 747–753 (2009). https://doi.org/10.1038/nature08494

Missing heritability



Manolio, T., Collins, F., Cox, N. et al. Finding the missing heritability of complex diseases. Nature 461, 747–753 (2009). https://doi.org/10.1038/nature08494

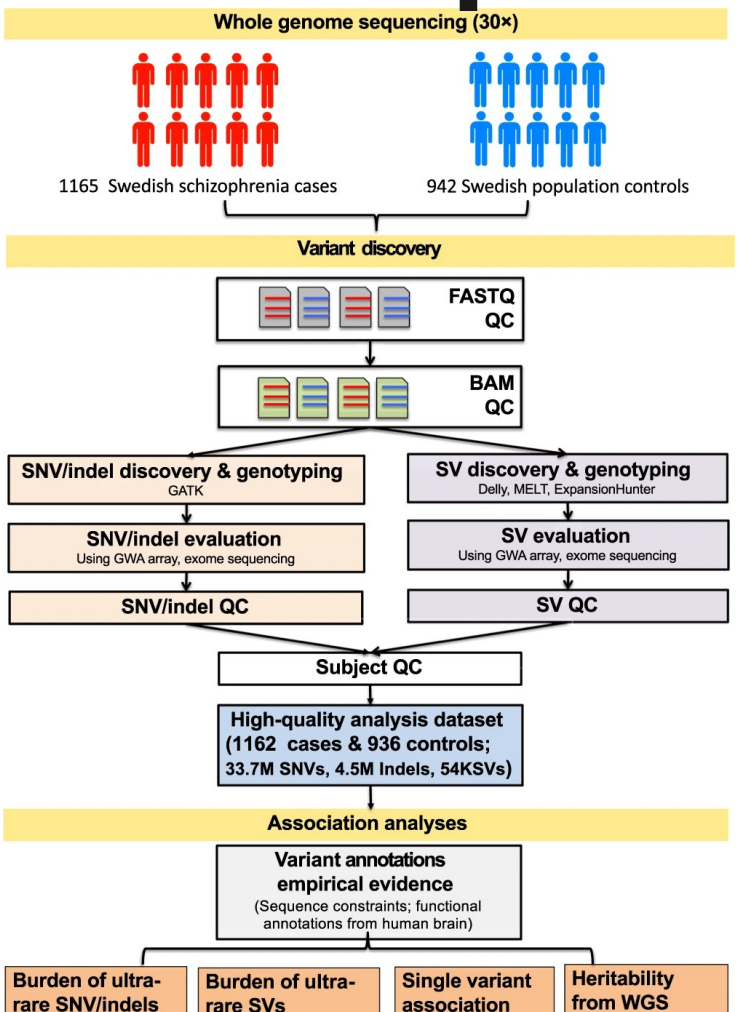
Missing heritability



- WGS can help capture additional heritability (in some cases)

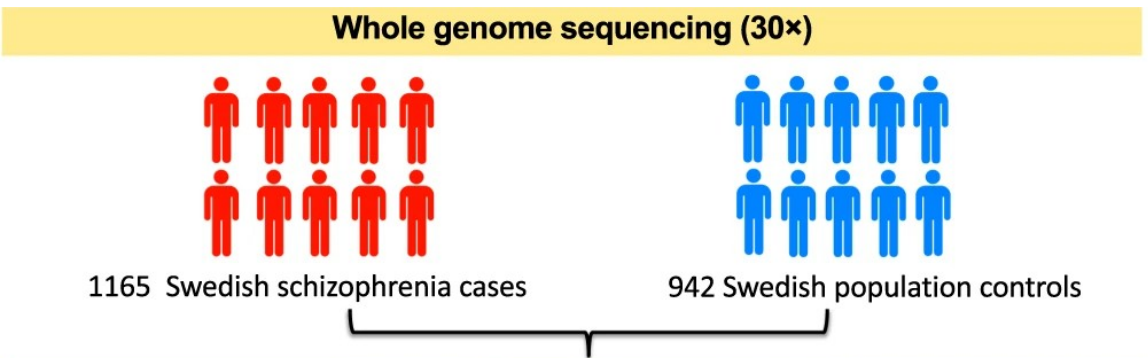
Wainschtein, P., Jain, D., Zheng, Z. et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet 54, 263–273 (2022). https://doi.org/10.1038/s41588-021-00997-7

Swedish Schizophrenia Study



Halvorsen, M., Huh, R., Oskolkov, N. et al. Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia.Nat Commun 11, 1842 (2020). https://doi.org/10.1038/s41467-020-15707-w

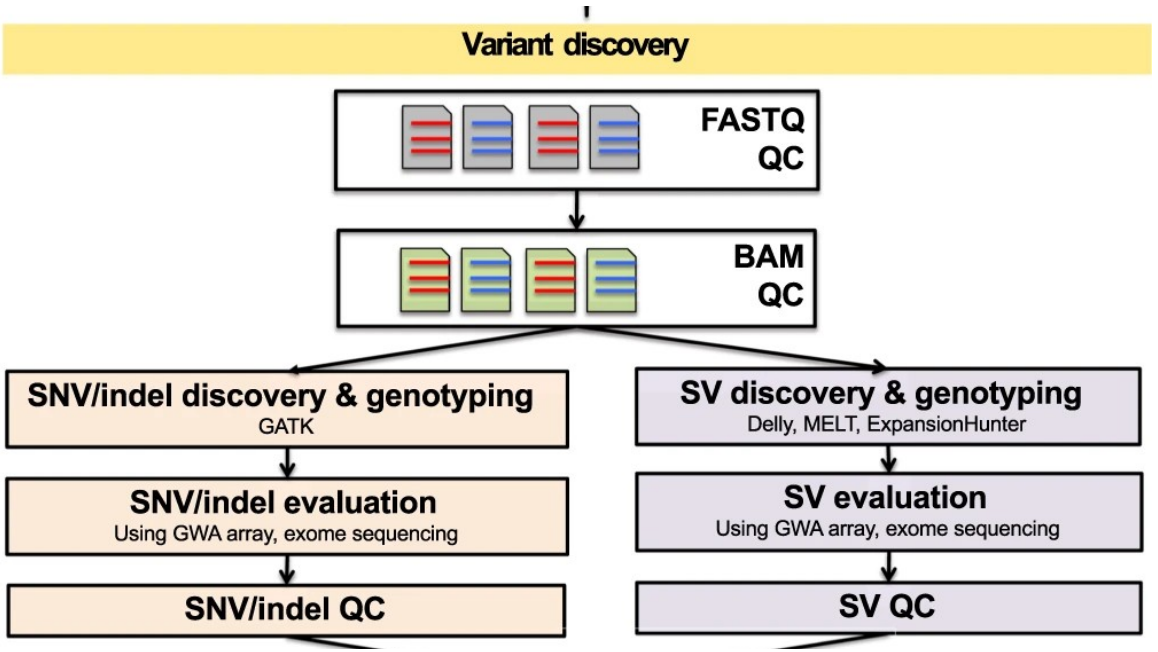
Swedish Schizophrenia Study



- Ancestry matched population controls
- Whole genome sequencing (WGS)

Halvorsen, M., Huh, R., Oskolkov, N. et al. Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia.Nat Commun 11, 1842 (2020). https://doi.org/10.1038/s41467-020-15707-w

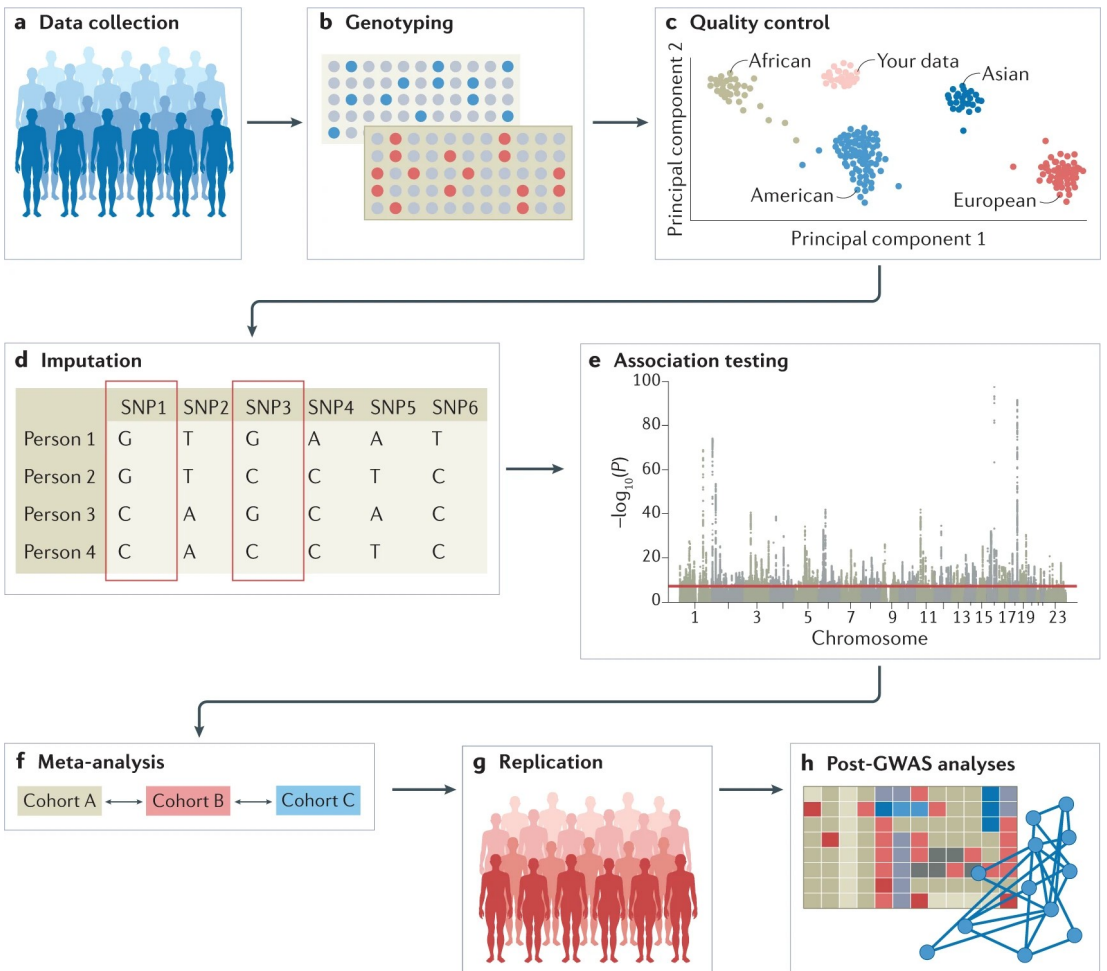
Swedish Schizophrenia Study



- WGS enables detection of structural variants (SV)
- Higher analytical burden

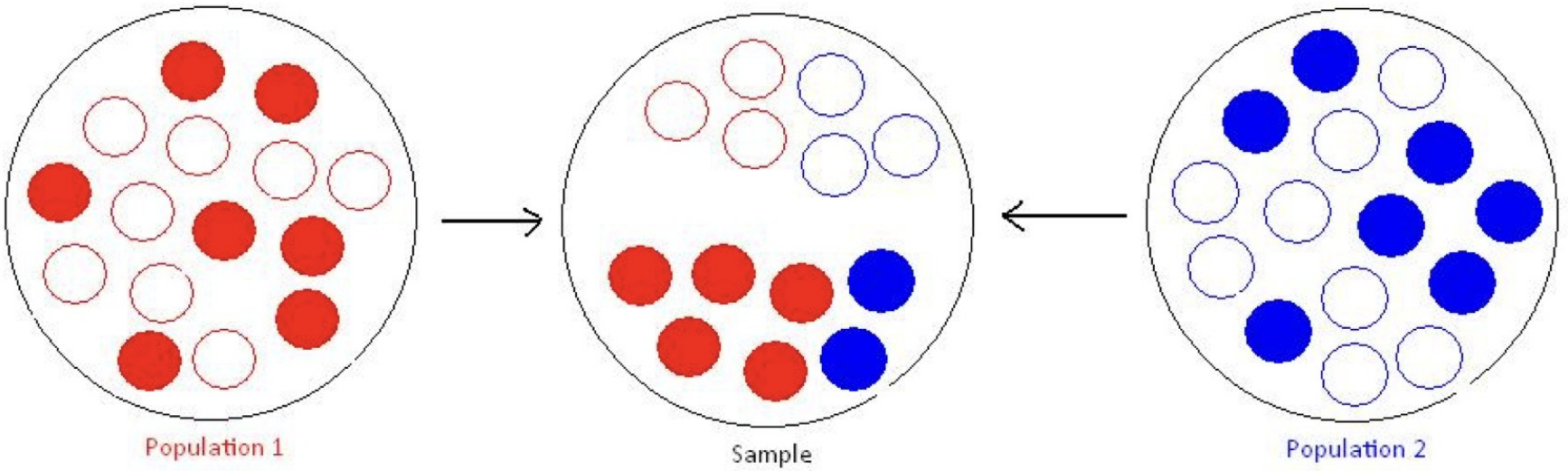
Halvorsen, M., Huh, R., Oskolkov, N. et al. Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia.Nat Commun 11, 1842 (2020). https://doi.org/10.1038/s41467-020-15707-w

GWAS from biobanks



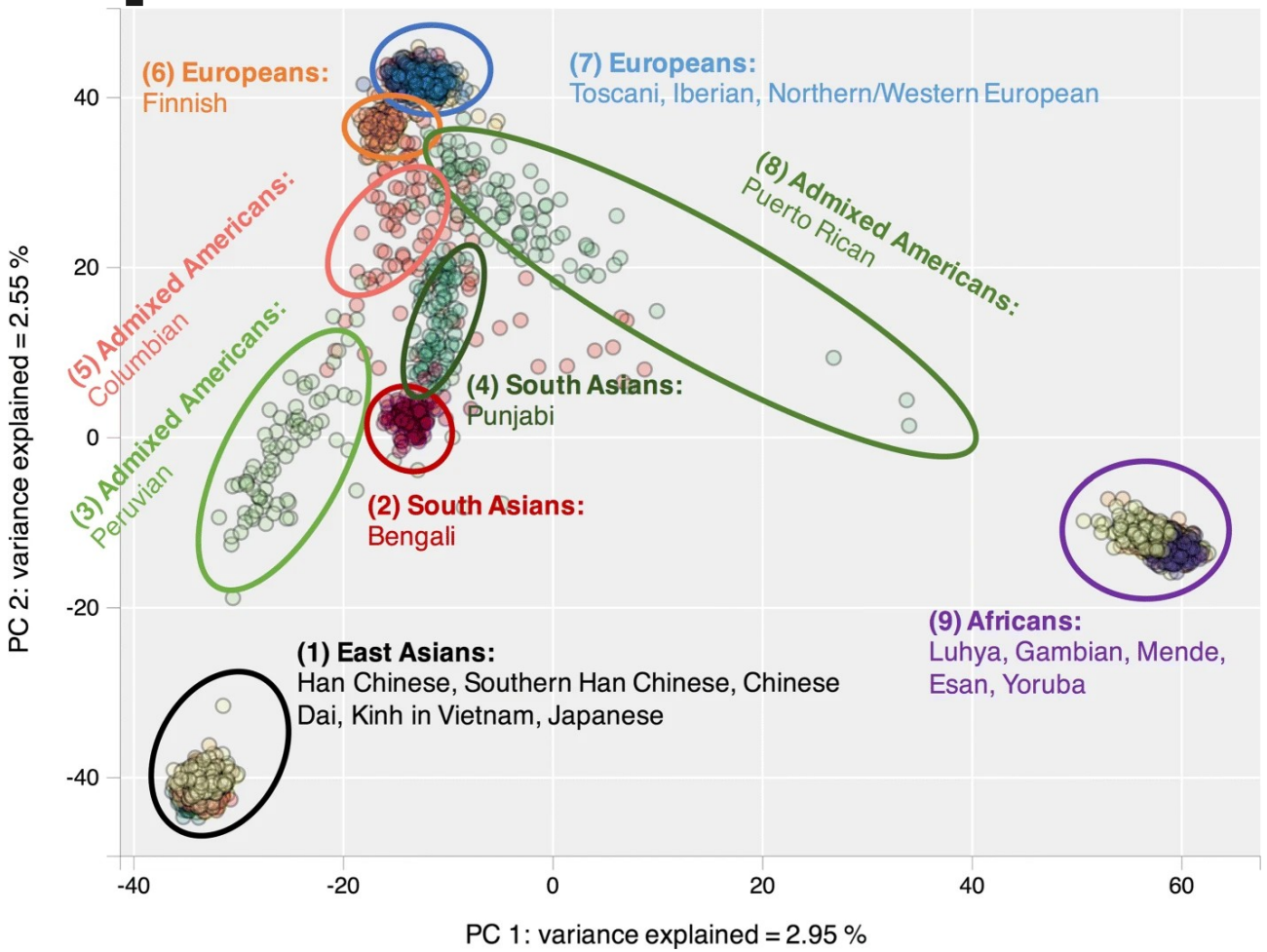
Uffelmann, E., Huang, Q.Q., Munung, N.S. et al. Genome-wide association studies. Nat Rev Methods Primers 1, 59 (2021). https://doi.org/10.1038/s43586-021-00056-9

Population stratification



- Closed circles are cases
- Open circles are controls
- Red ancestry overrepresented in cases
- i.e. ancestry could drive association

Population stratification



Gaspar, H.A., Breen, G. Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics. BMC Bioinformatics 20, 116 (2019). https://doi.org/10.1186/s12859-019-2680-1

Population stratification

PROGRESS

Figure 1 | **P–P plots for the visualization of stratification or other confounders.** The figure shows simulated P–P plots under three scenarios for genome-wide scans with no causal markers. **a** | No stratification: *p*-values fit the expected distribution. **b** | Stratification without unusually differentiated markers: *p*-values exhibit modest genome-wide inflation. **c** | Stratification with unusually differentiated markers: *p*-values exhibit modest genome-wide inflation and severe inflation at a small number of markers.