

GENOME-WIDE ASSOCIATION STUDIES IN CANCER

Presented By :

Onnapa Kongphan

Master's degree student of Biomedical science program,
Khon Kaen university

Contents

1

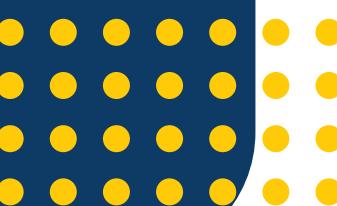
Genome-Wide Association studies (GWAS):
an overview

2

The GWAS workflow

3

Application of GWAS in Cancer



GWAS overview

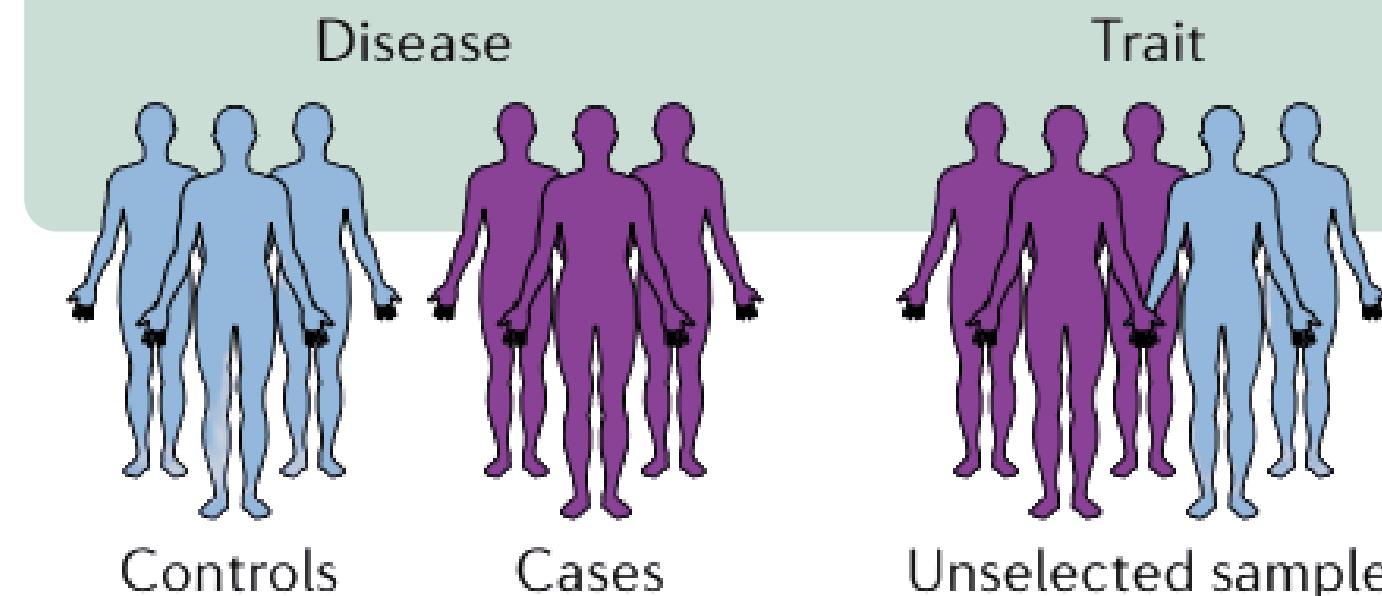
GWAS are research methods used to identify genetic variations (e.g., SNPs) across the entire genome that are associated with specific diseases or traits

Genotyping data

- Single nucleotide polymorphism (SNPs) are commonly used as marker
- Copy number variations (CNVs)
- Insertions and deletions (INDELS)

Phenotyping data

- Disease status (cases vs control)
- Quantitative traits (e.g., drug levels)
- Environmental traits (e.g., smoking, exercises)



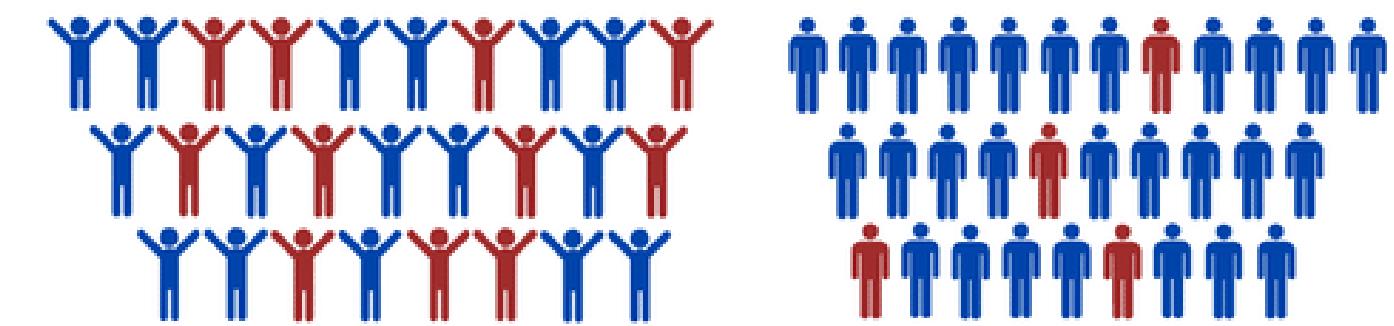
Purpose

To uncover genetic factors that contribute to disease risk by identifying SNPs that are linked to increased susceptibility.

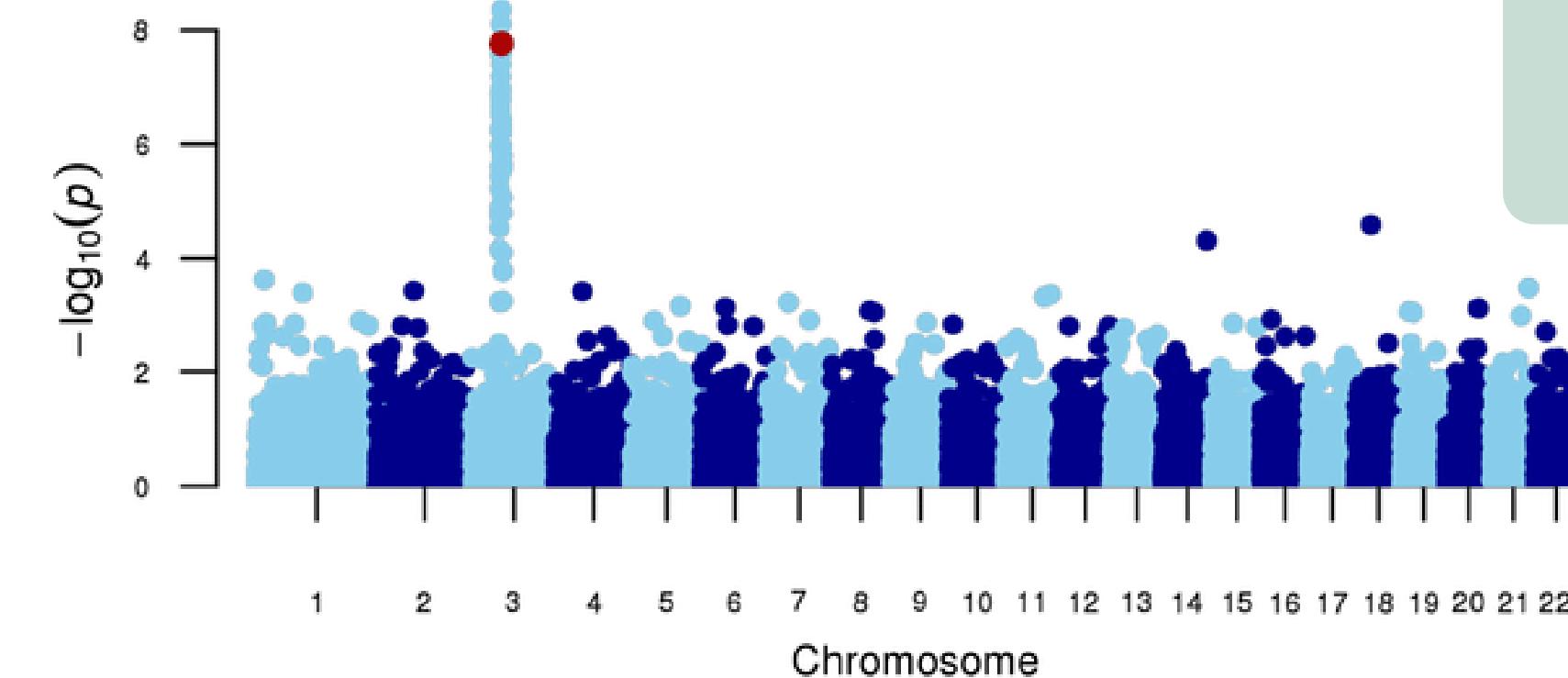
Sequence Variation

ATGCCAGTGTTCAGATGCTGGCCAGCTGGACGAGGGCGATGAC
ATGCCAGTGTTCAGATG**T**GGCCAGCTGGACGAGGGCGATGAC

Disease



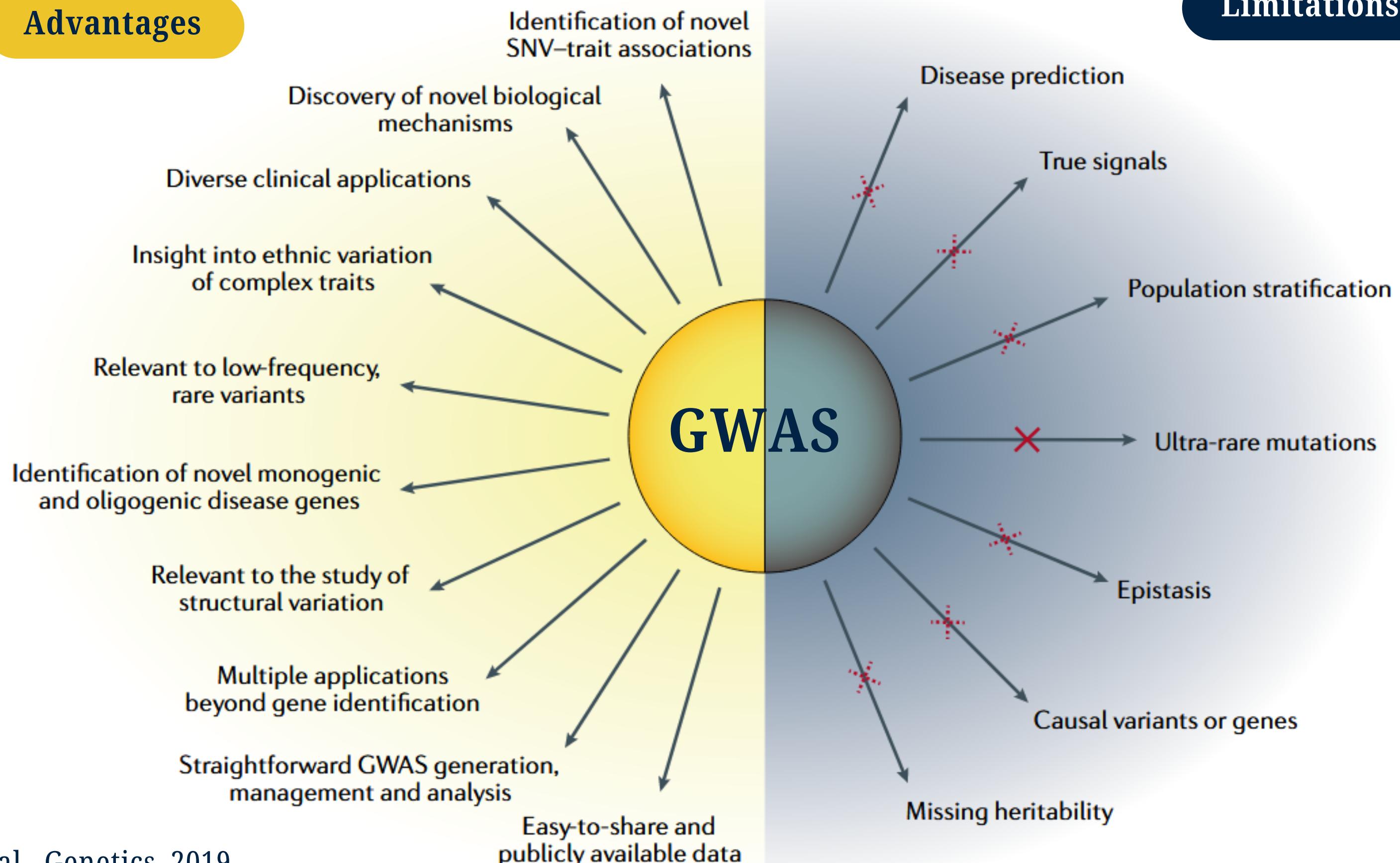
GWAS



- To use genetic risk factors to predict who is at risk
- Uncover genetic factors underlying disease susceptibility to enable new prevention and treatment strategies

Advantages and Limitations of GWAS

Advantages



Genotyping

Genetic variants can be genotyped using numerous technologies, including SNP arrays and WGS



SNP arrays

VS

Whole-genome sequencing (WGS)



Costs

Genomic coverage

GWAS analysis

Relatively inexpensive (~US\$40 per sample)

Biased towards variants discovered in wellstudied or sequenced populations
(Common and low-frequency variants)

Well-established analytical pipeline and tools for data analysis

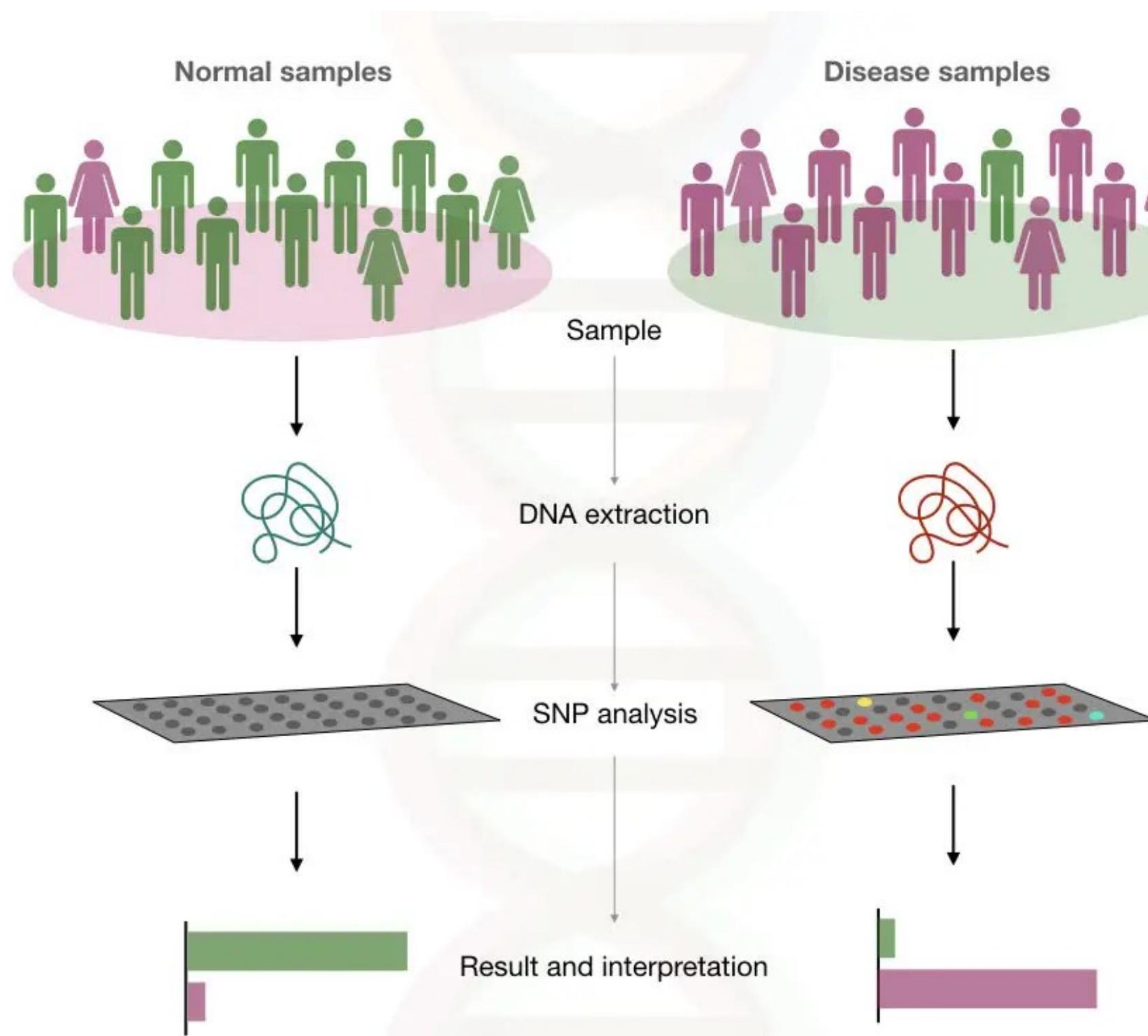
Expensive (>US\$1,000 per sample)

From low-frequency, common variants to nearly all genetic variation in the genome

Higher computational costs and greater analytical complexity

Key Features of GWAS

GWAS often require very large sample sizes to identify reproducible associations and depending on the genetic architecture of the traits



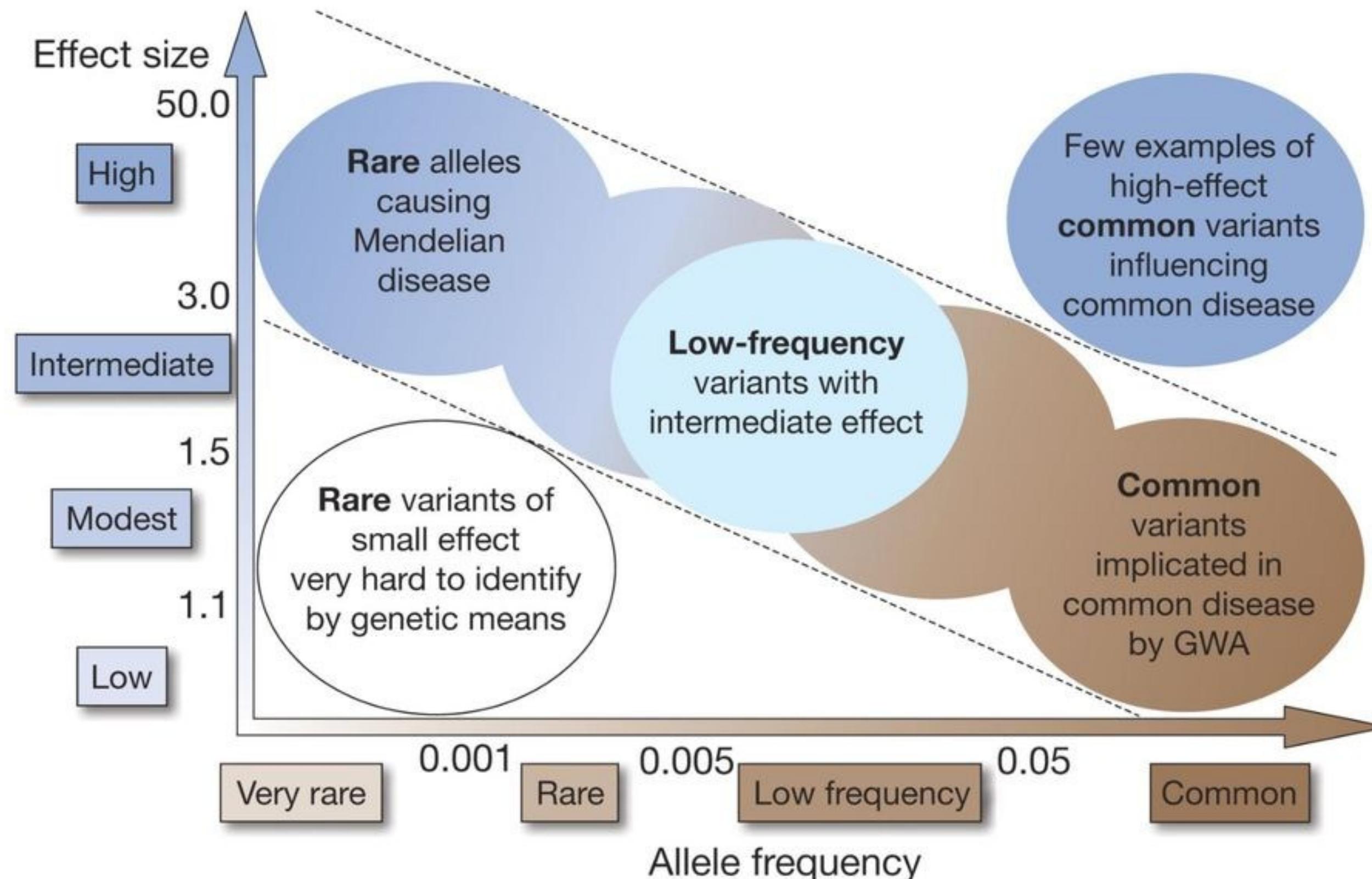
Large-Scale Studies

Case-Control Design

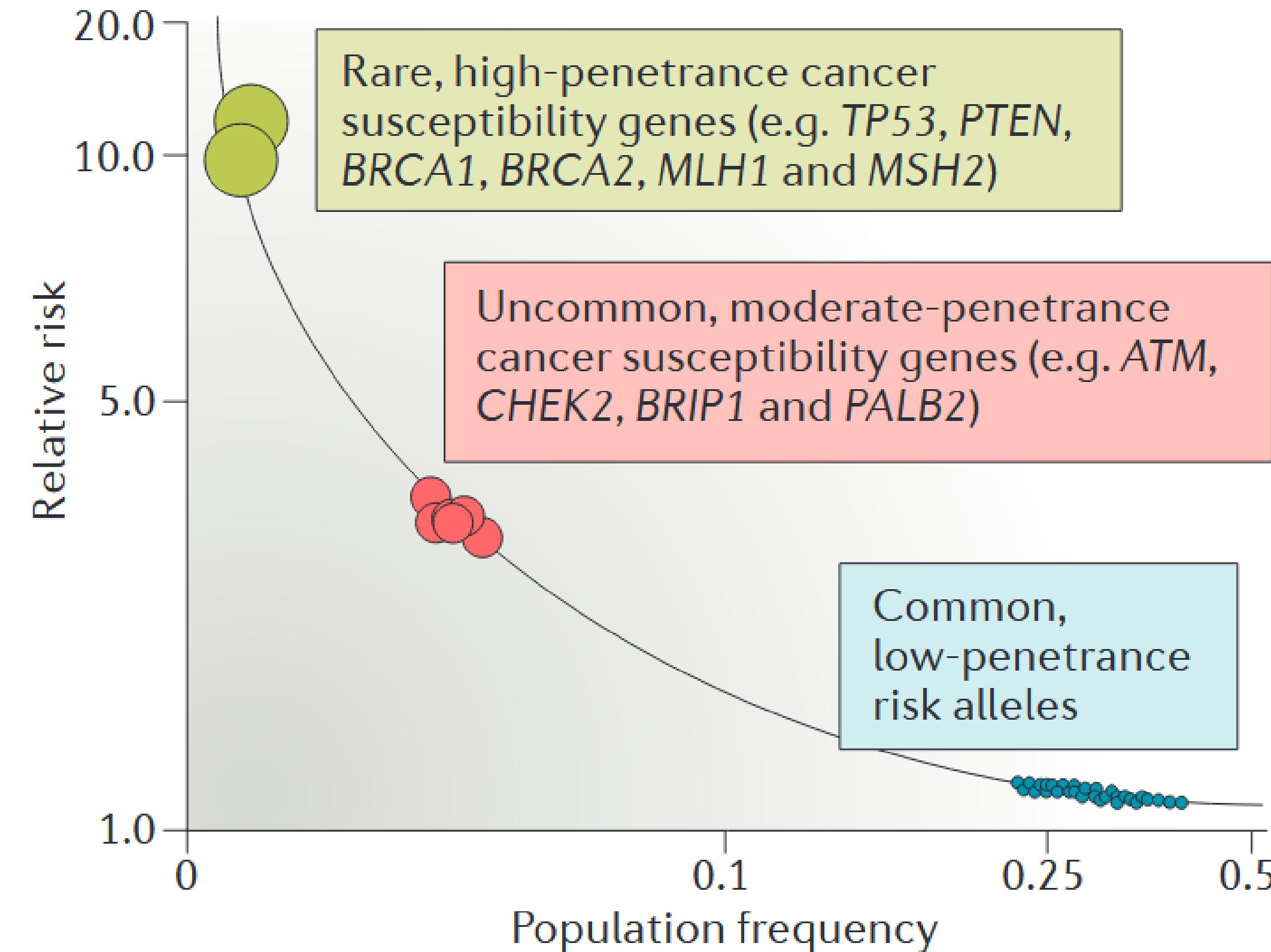
Phenotype-Genotype Association

Focus on Common Variants

Genetic variation and disease susceptibility



Genetic architecture of cancer risk



Contents

1

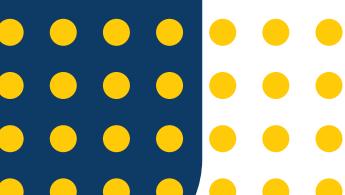
Genome-Wide Association studies (GWAS):
an overview

2

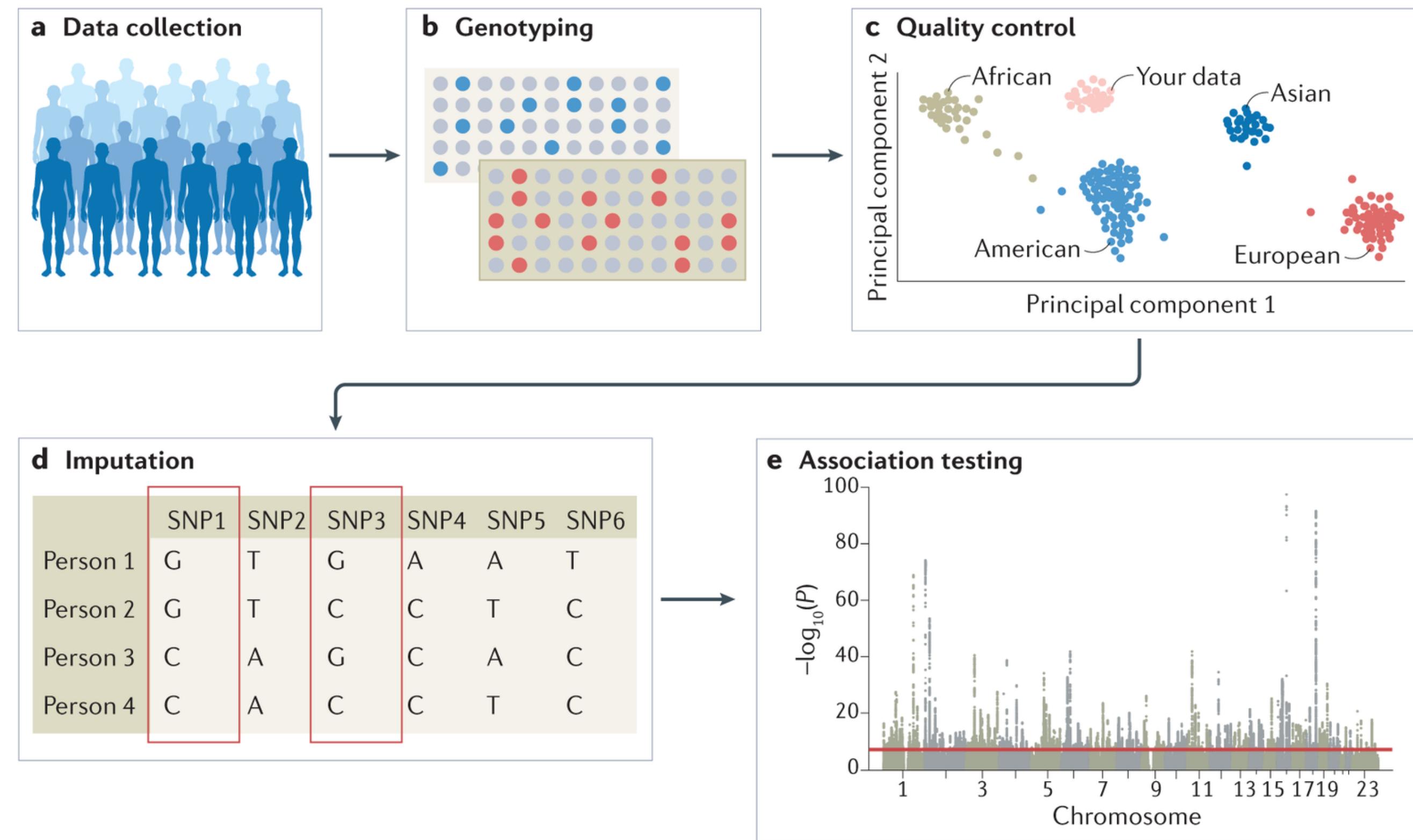
The GWAS workflow

3

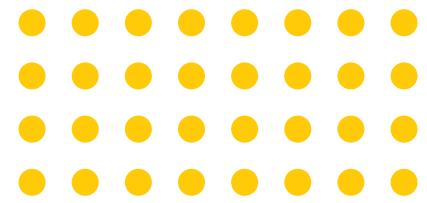
Application of GWAS in Cancer



The GWAS workflow



Quality control



1

Missingness of SNPs and individuals

2

Sex discrepancy

3

Minor allele frequency (MAF)

4

Hardy-Weinberg equilibrium (HWE)

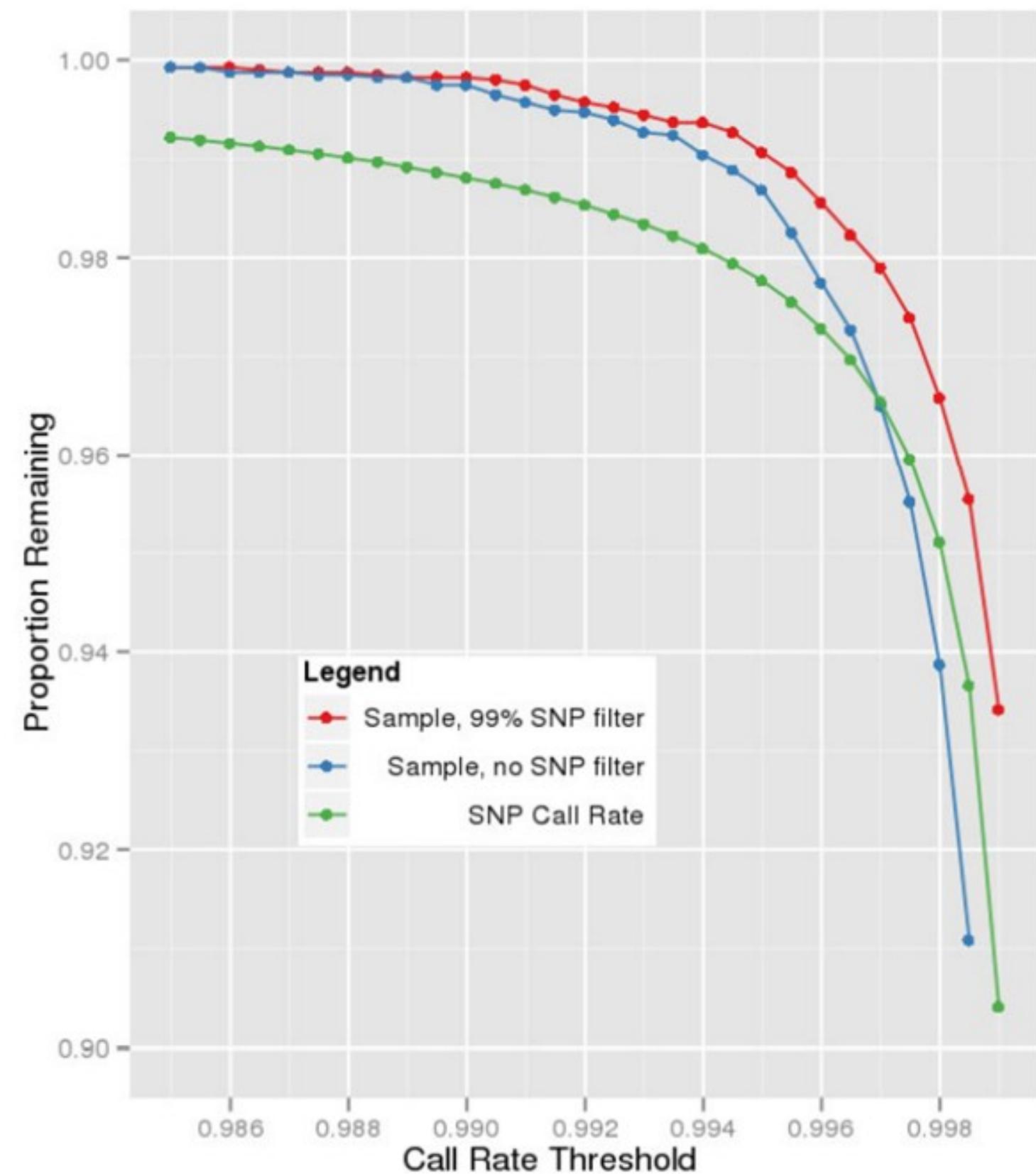
5

Relatedness

6

Population stratification

Missingness of SNPs and individuals



A recommended threshold is 98-99% efficiency

Samples missing rate < 98-99%

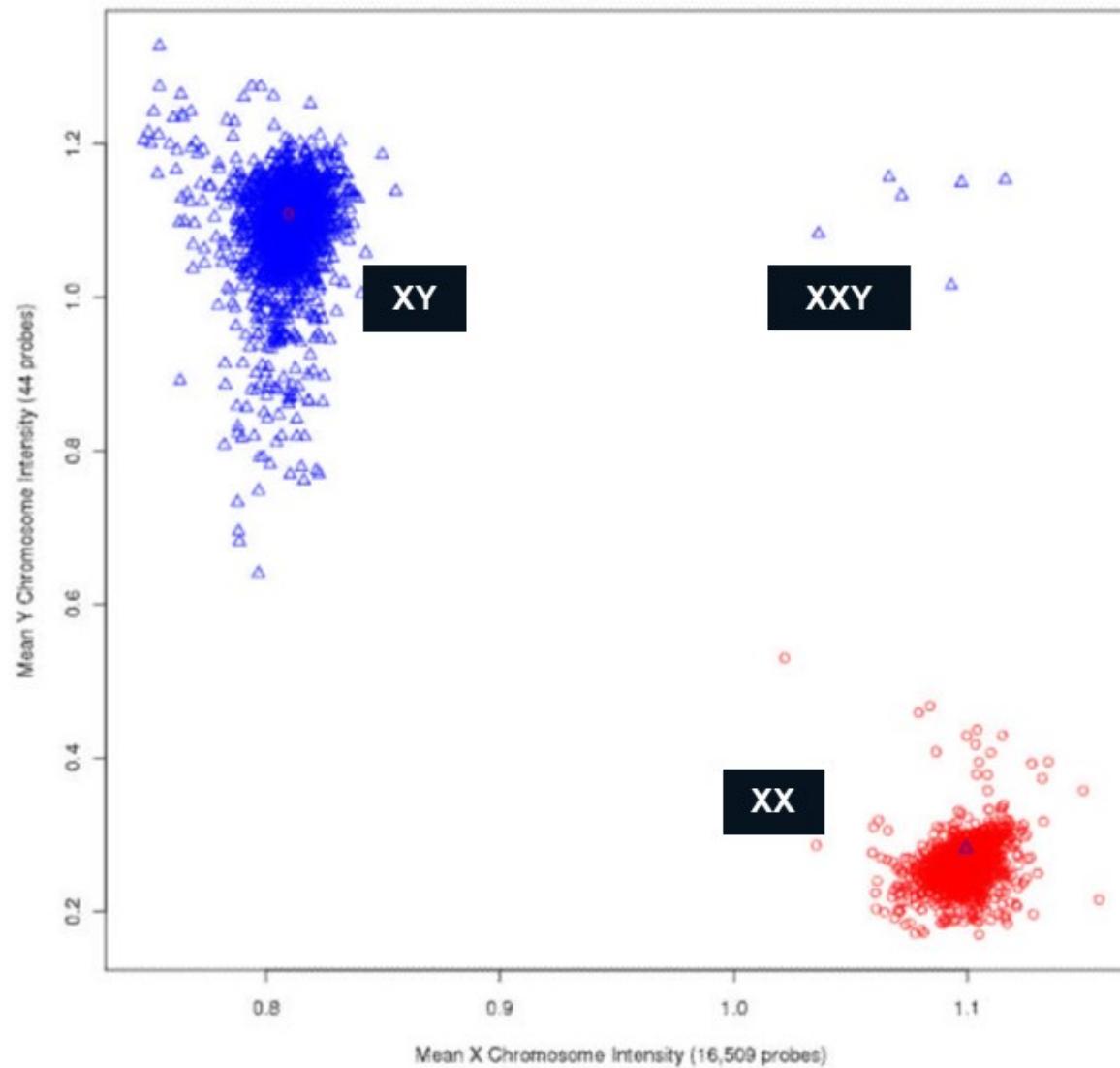
- Poor quality DNA samples
- Aberrant genotype calling

SNP missing rate < 98-99%

- Bad indicator of marker quality

Sex discrepancy

Checks for discrepancies between sex of the individuals recorded in the dataset [PEDSEX] and their sex based on X chromosome heterozygosity/ homozygosity rates [SNPSEX]

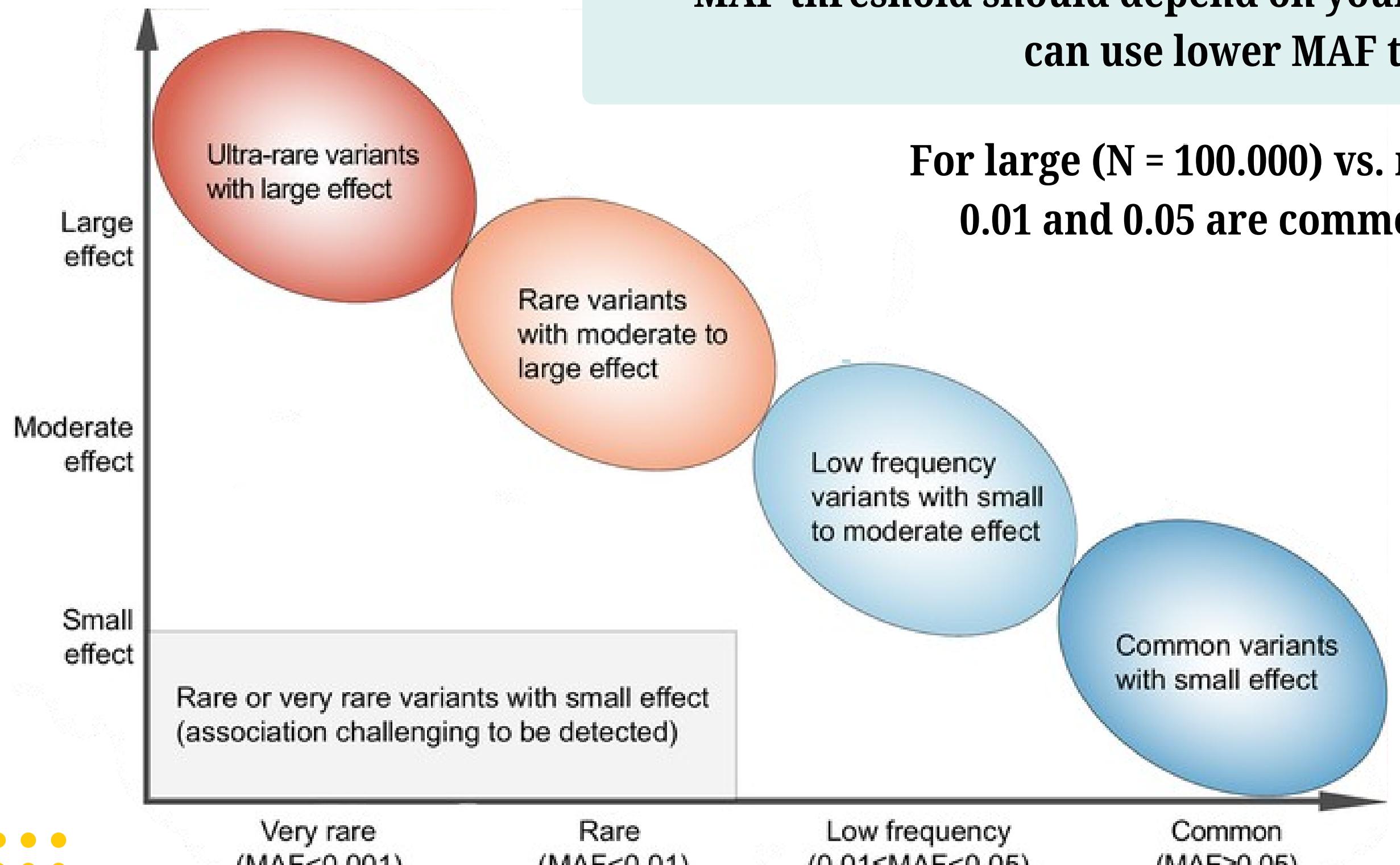


IID	PEDSEX	SNPSEX	STATUS	F	Explanation
1	1	1	OK	0.98	Male
2	2	2	OK	0.03	Female
3	2	1	PROBLEM	0.99	Recorded female, genetically male
4	1	2	PROBLEM	0.02	Recorded male, genetically female
5	2	0	PROBLEM	0.28	Likely a female with sex chromosome anomaly (e.g. XX/XO mosaic, loss-of-heterozygosity on X)
6	1	0	PROBLEM	0.35	Likely a male with sex chromosome anomaly (e.g. XXY or XX/XY mosaic)

1 = male, 2 = female, 0 = unknown

Males should have an X chromosome homozygosity estimate >0.8 and females should have a value <0.2

Minor allele frequency (MAF)



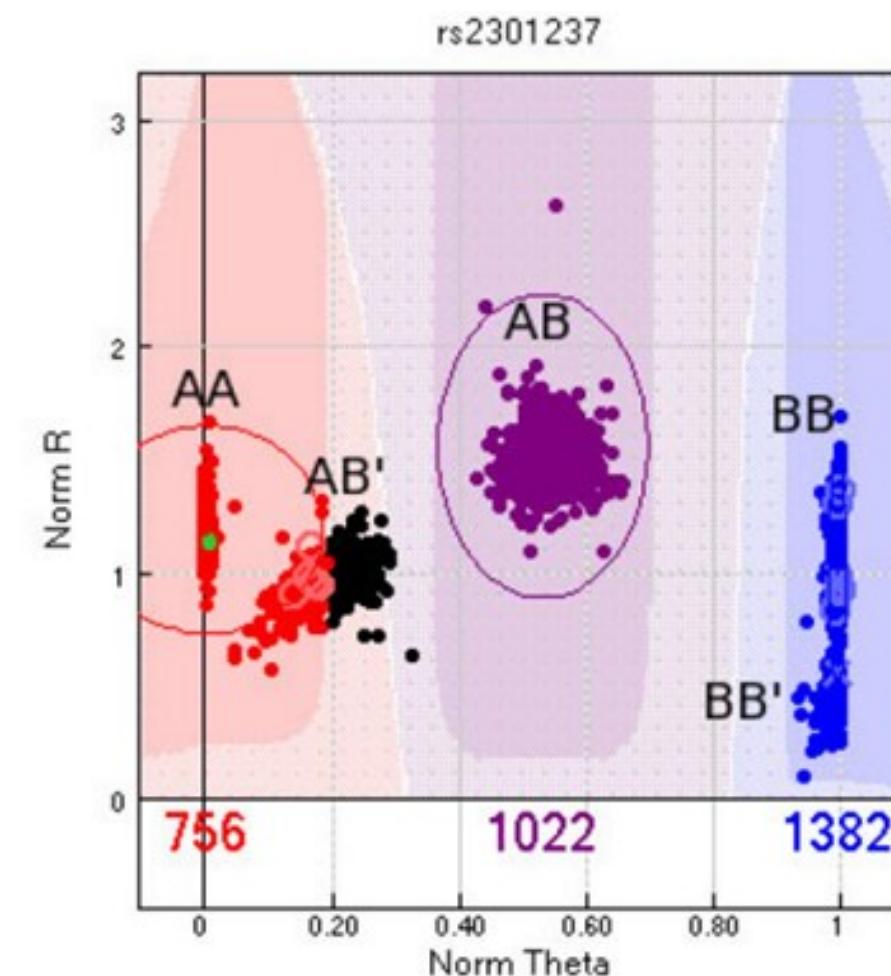
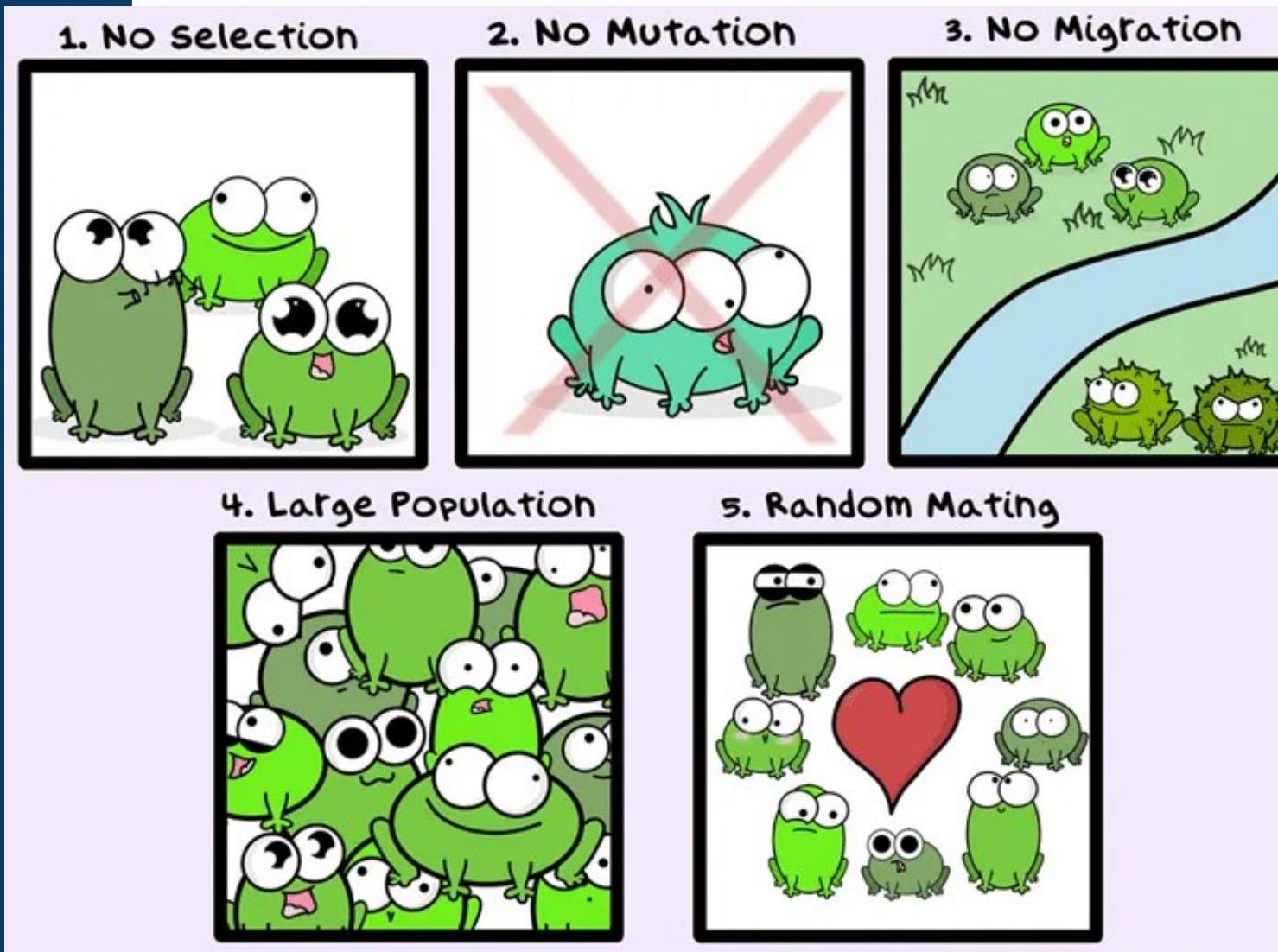
Hardy-Weinberg equilibrium (HWE)

It assumes the population with no selection, mutation, or migration

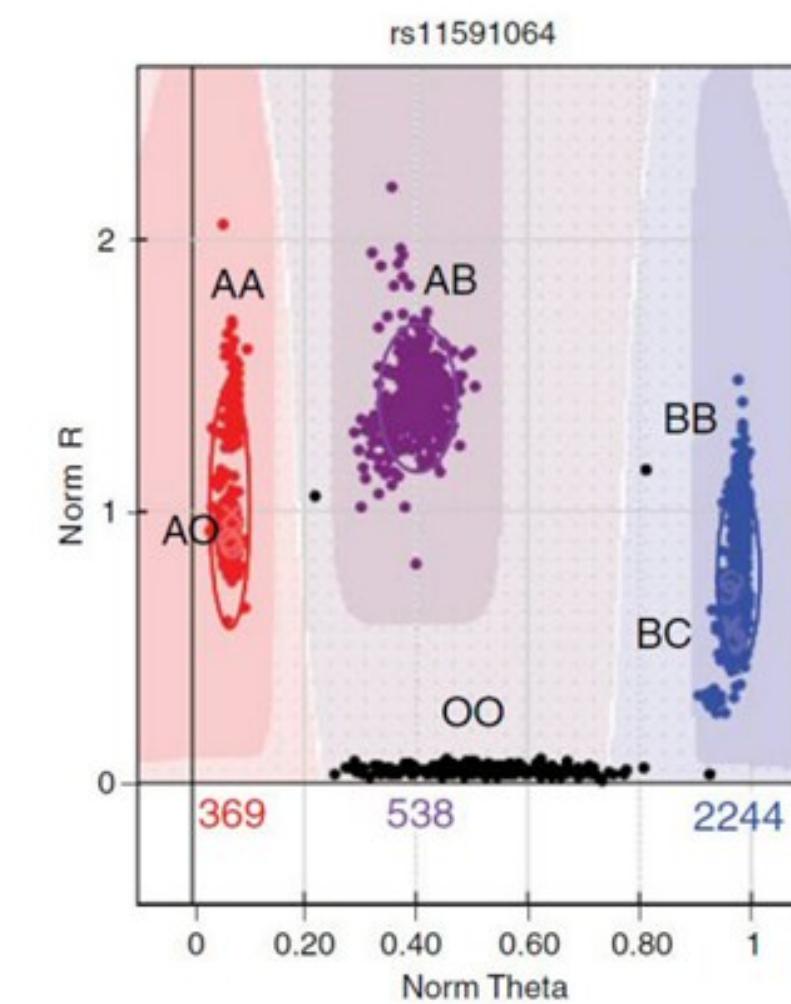


Variants with low P value

- Genotyping errors
- Indicate evolutionary selection



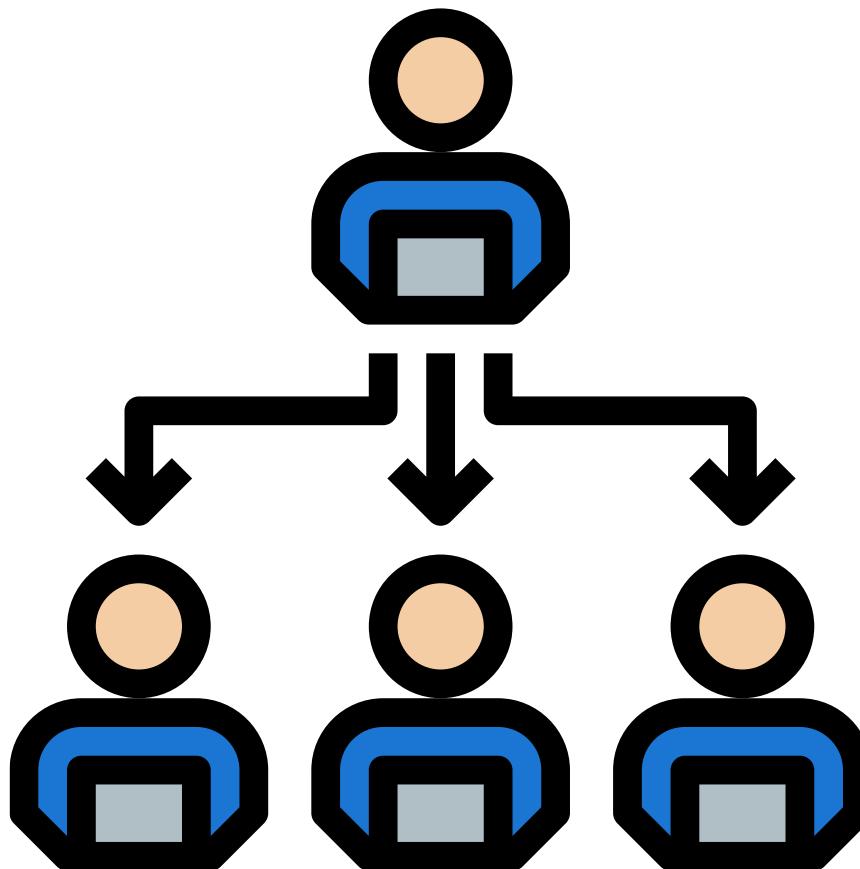
deviation from HWE due to excess homozygosity



deviation from HWE due to copy loss

Relatedness

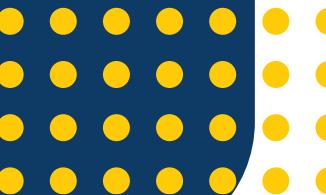
Relatedness refers to the degree of genetic similarity or kinship between individuals in a study population.



related individuals share a higher proportion of their genome

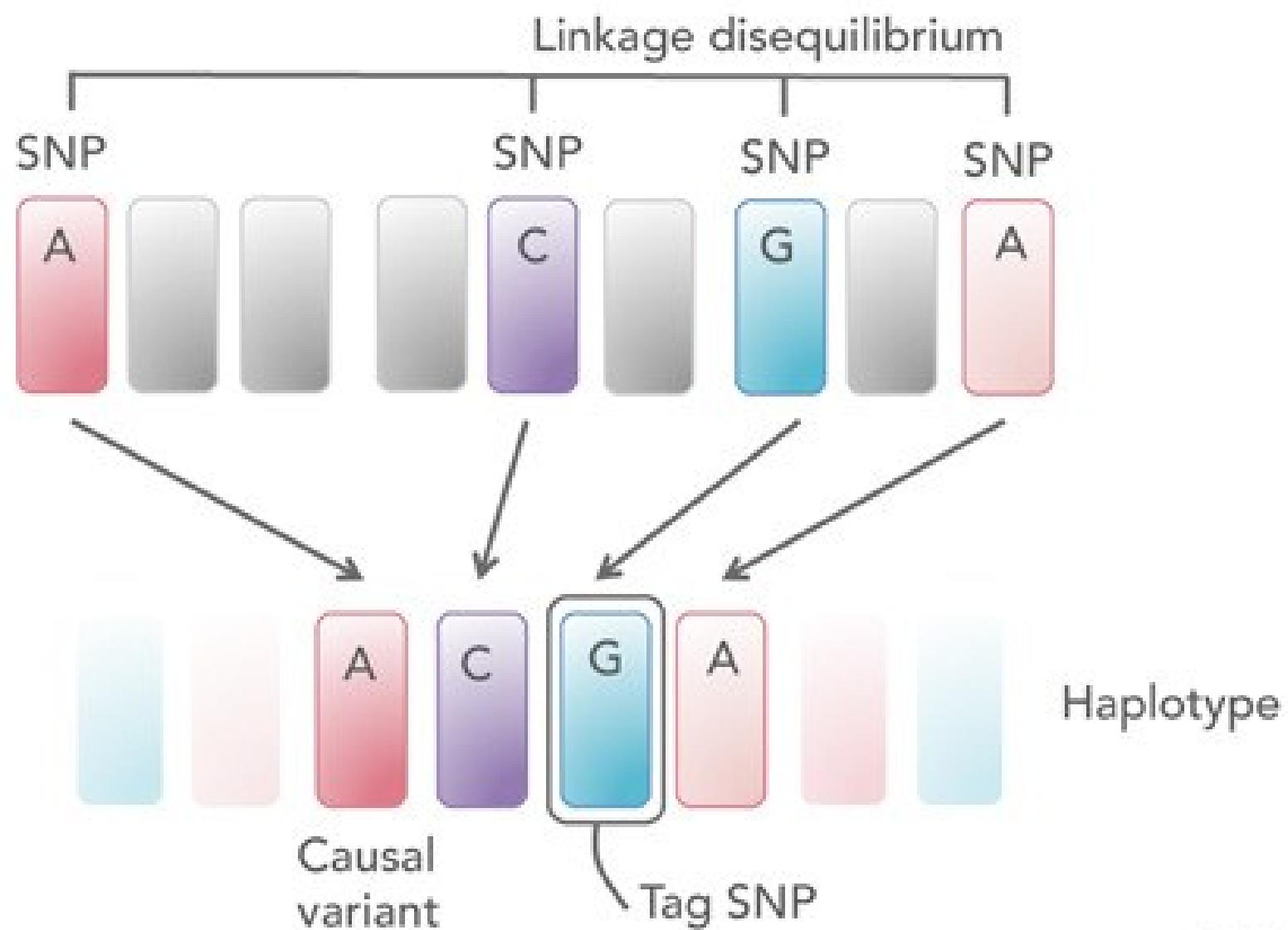


introduce bias or confounding effects in the association analysis



Linkage disequilibrium (LD)

Property of one allele in an SNPs being correlated with an allele in another SNPs along a contiguous stretch of the genome



Importance of LD in Genetics and GWAS

- Tagging SNPs
- Identifying Causal Variants
- Haplotype Blocks
- Population History

CHOI

Contents

1

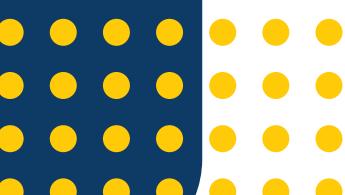
Genome-Wide Association studies (GWAS):
an overview

2

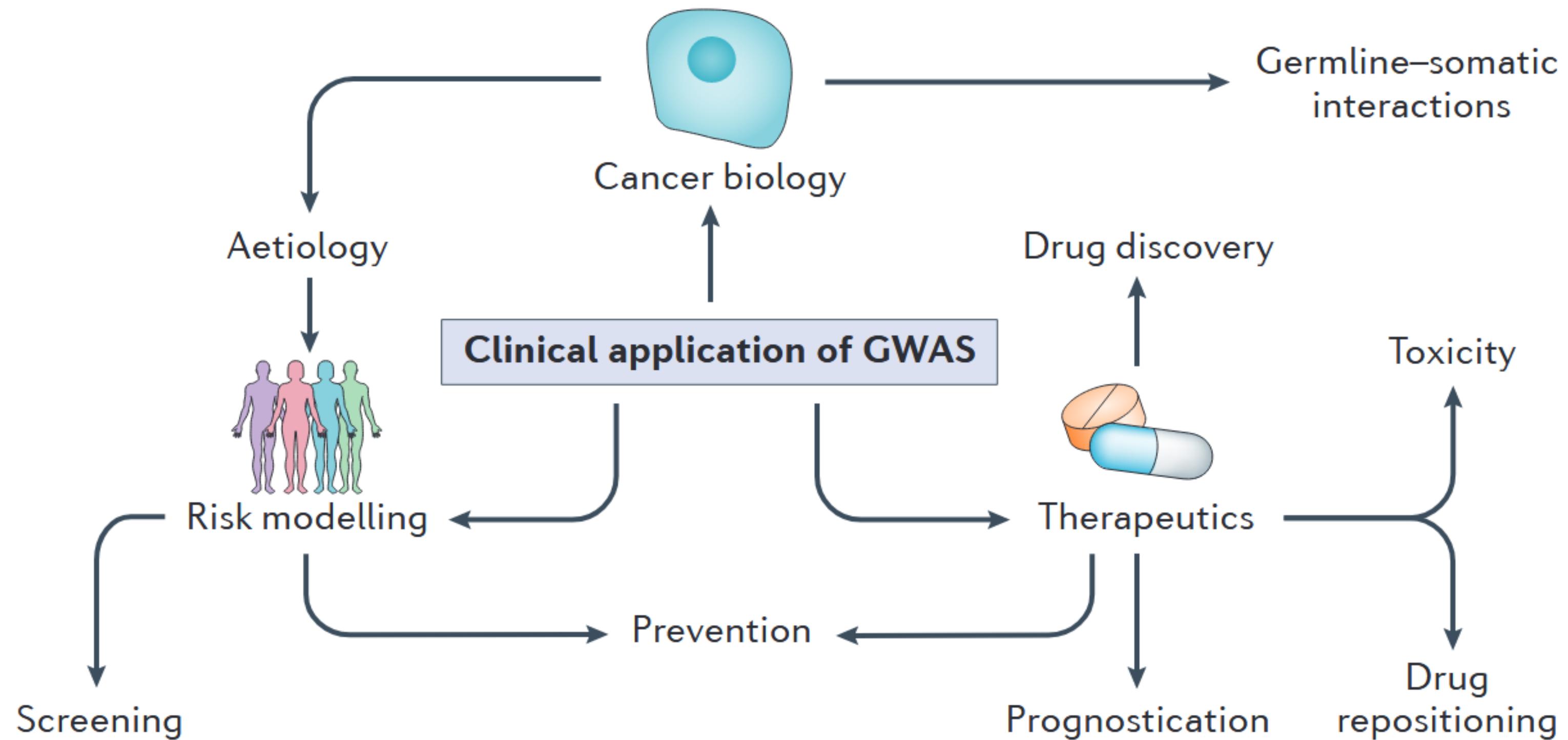
The GWAS workflow

3

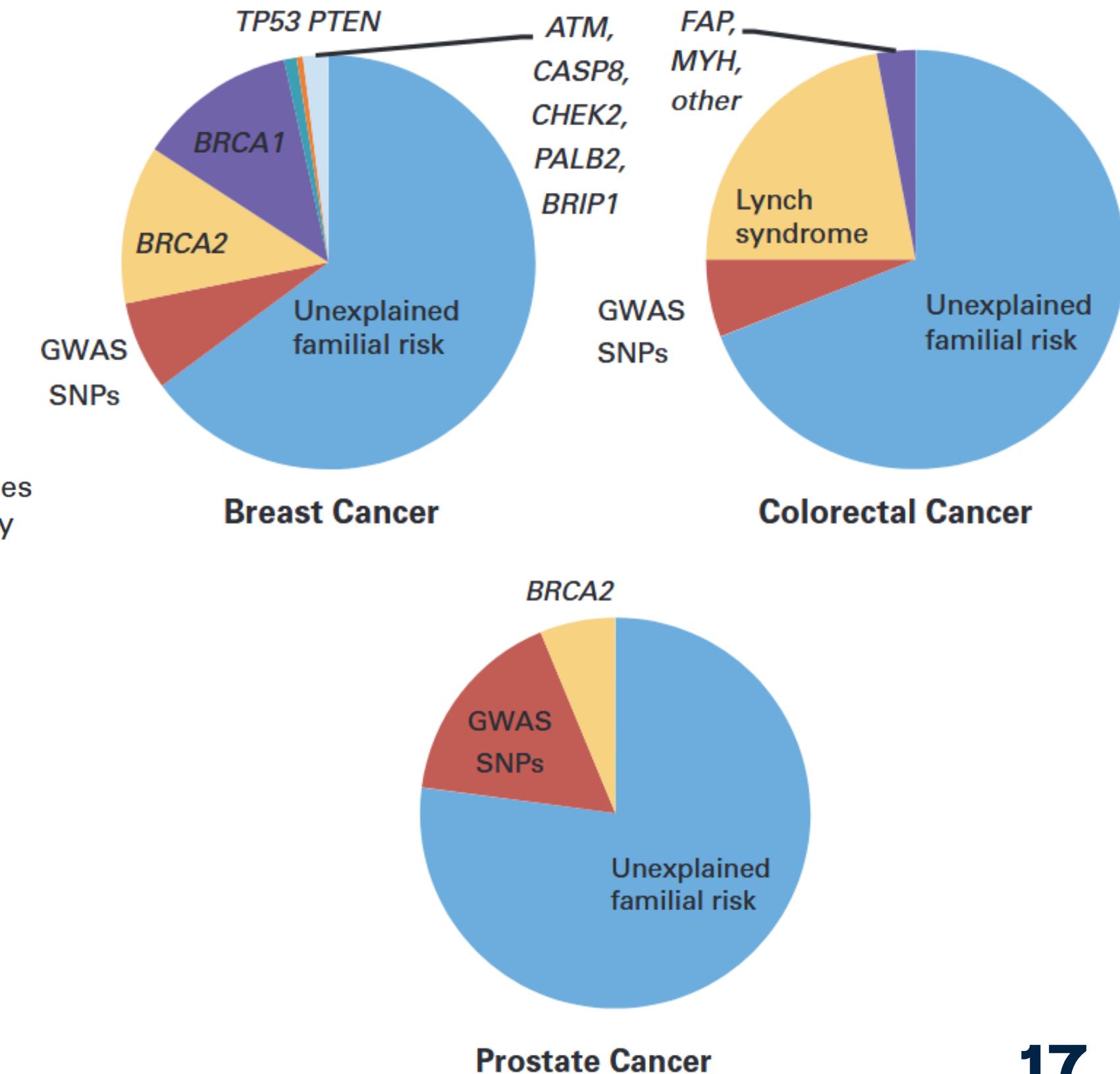
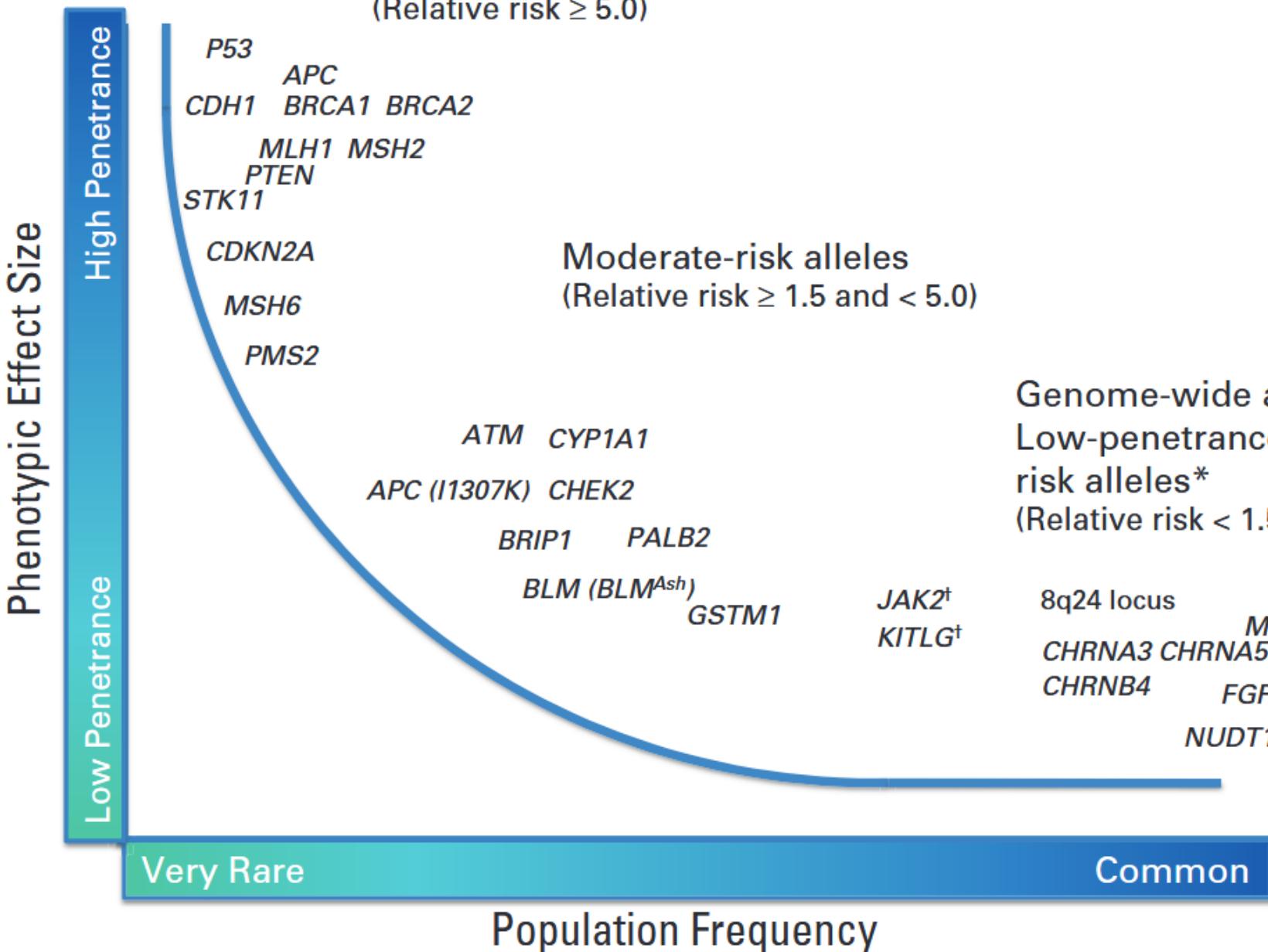
Application of GWAS in Cancer



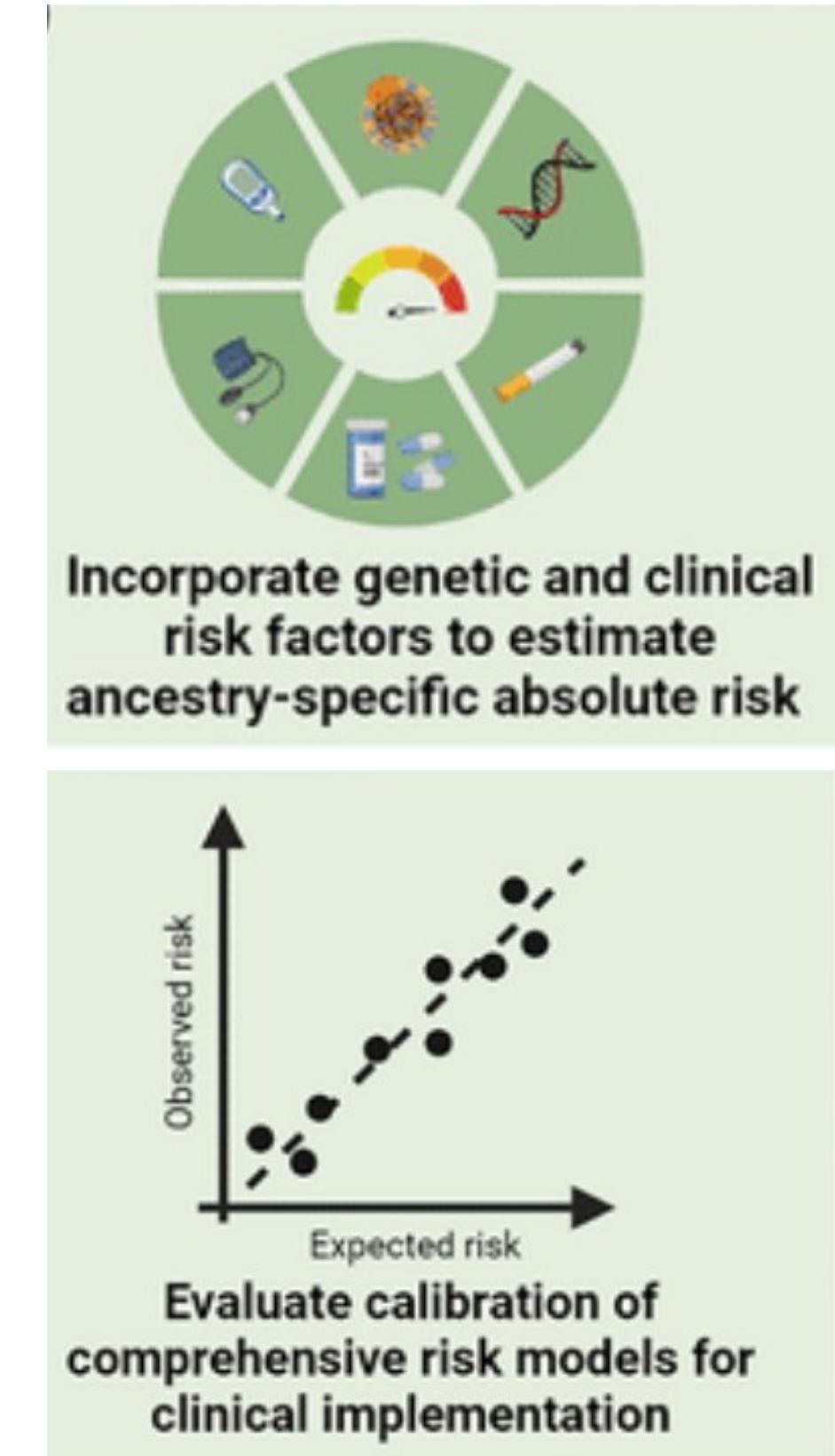
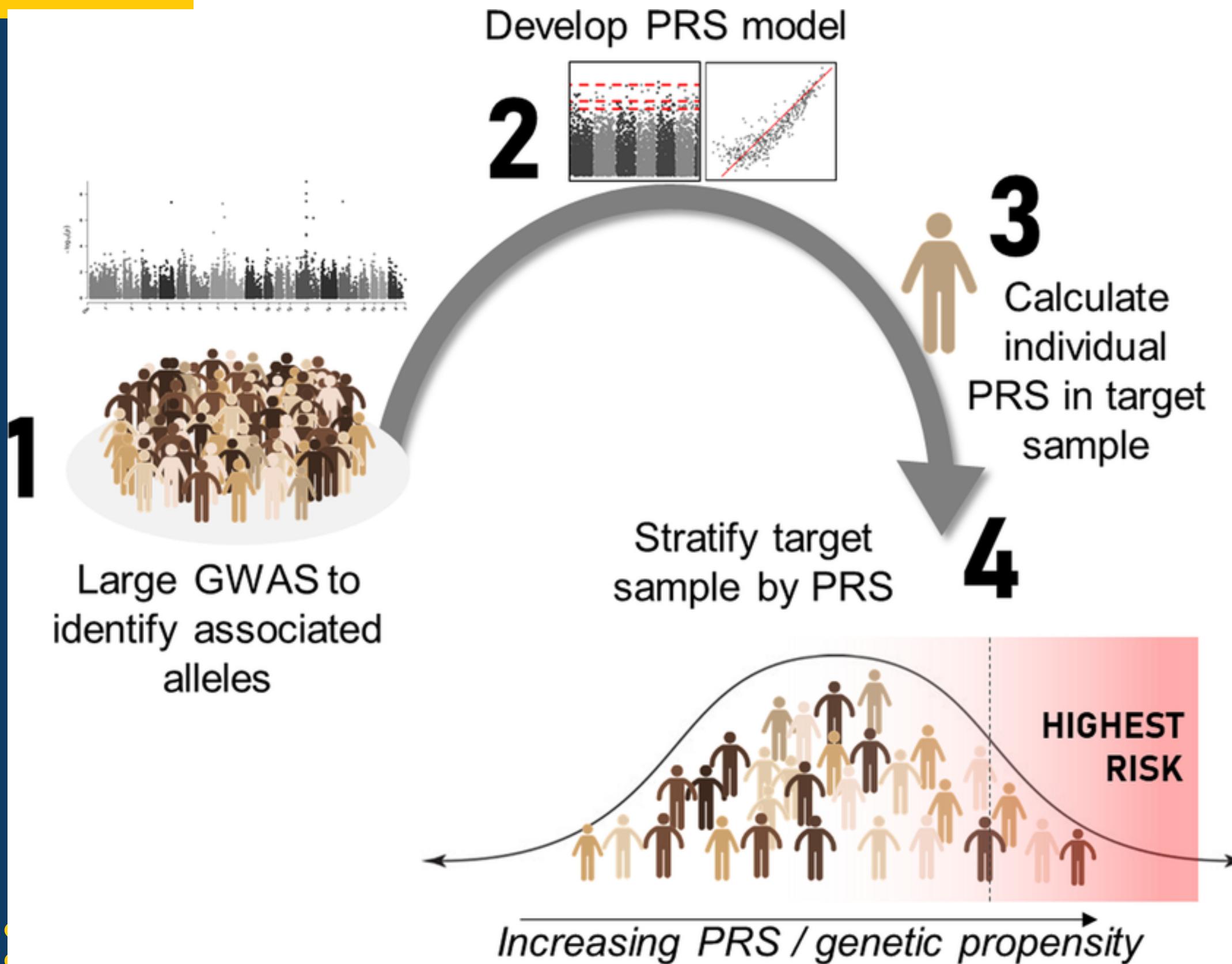
The clinical application of GWAS



GWAS have identified many low-penetrance susceptibility loci



Develop Polygenic risk score (PRS)



GWAS in Somatic VS Germline mutation

Somatic DNA changes

Acquired over a persons lifetime in single cells

Can lead to cancer

Can NOT be inherited

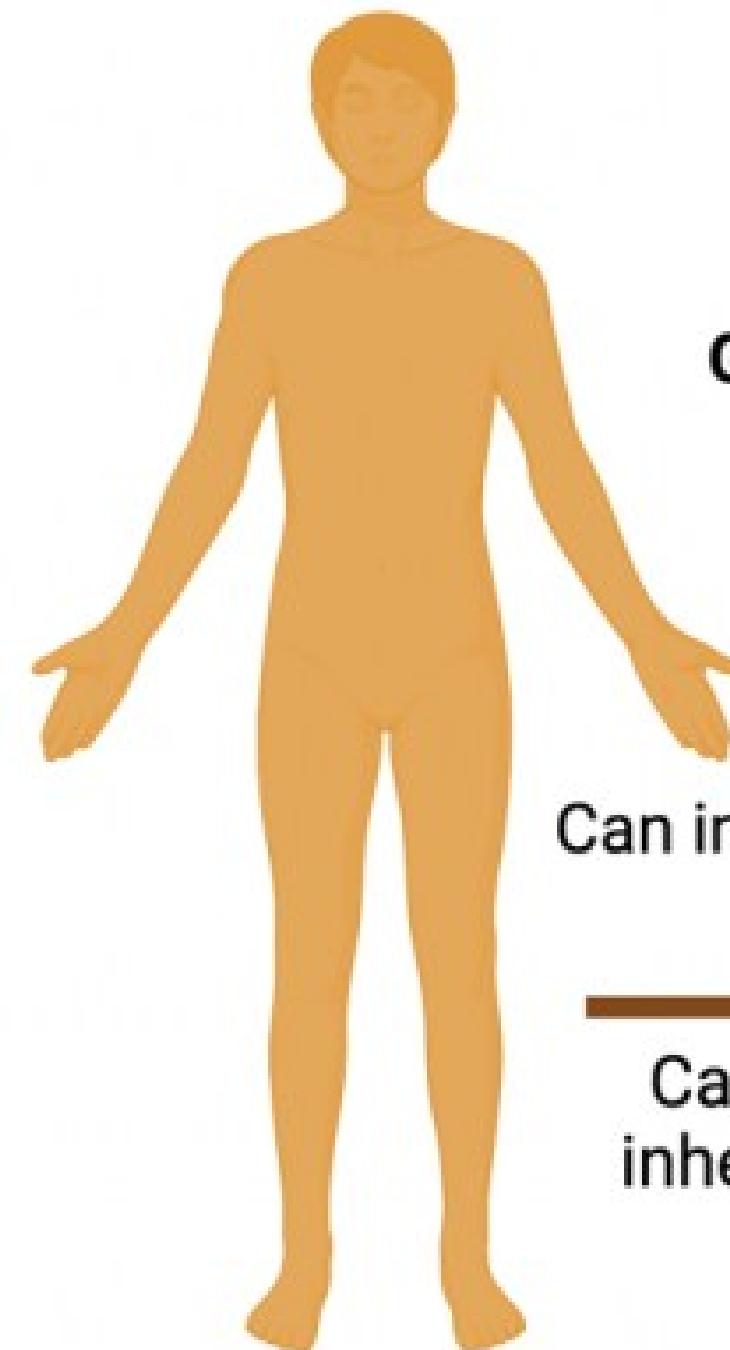


Germline DNA changes

Present in every cell of the body including egg and sperm

Can increase cancer susceptibility

Can be inherited



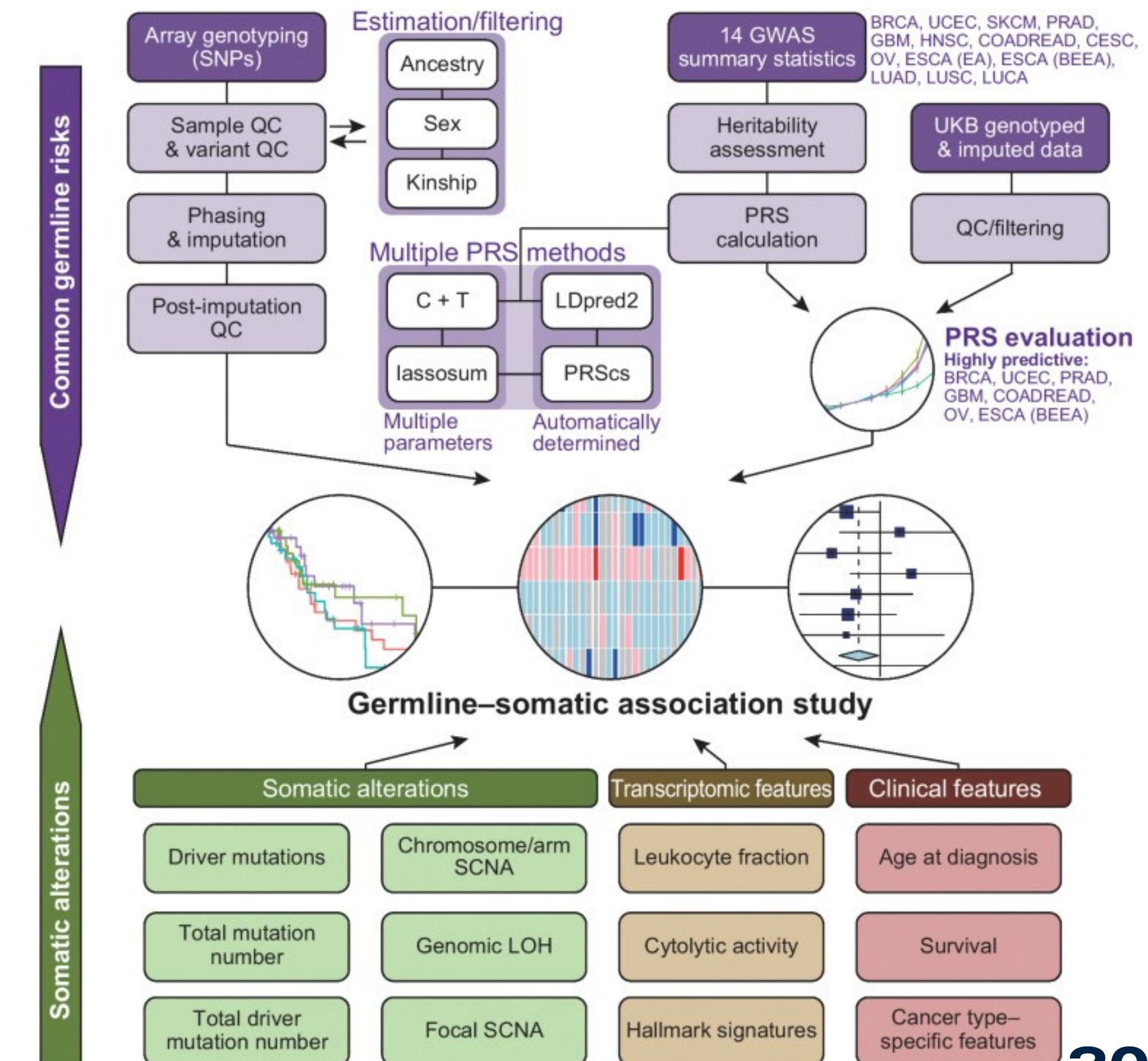
GWAS in Somatic VS Germline mutation

Germline GWAS

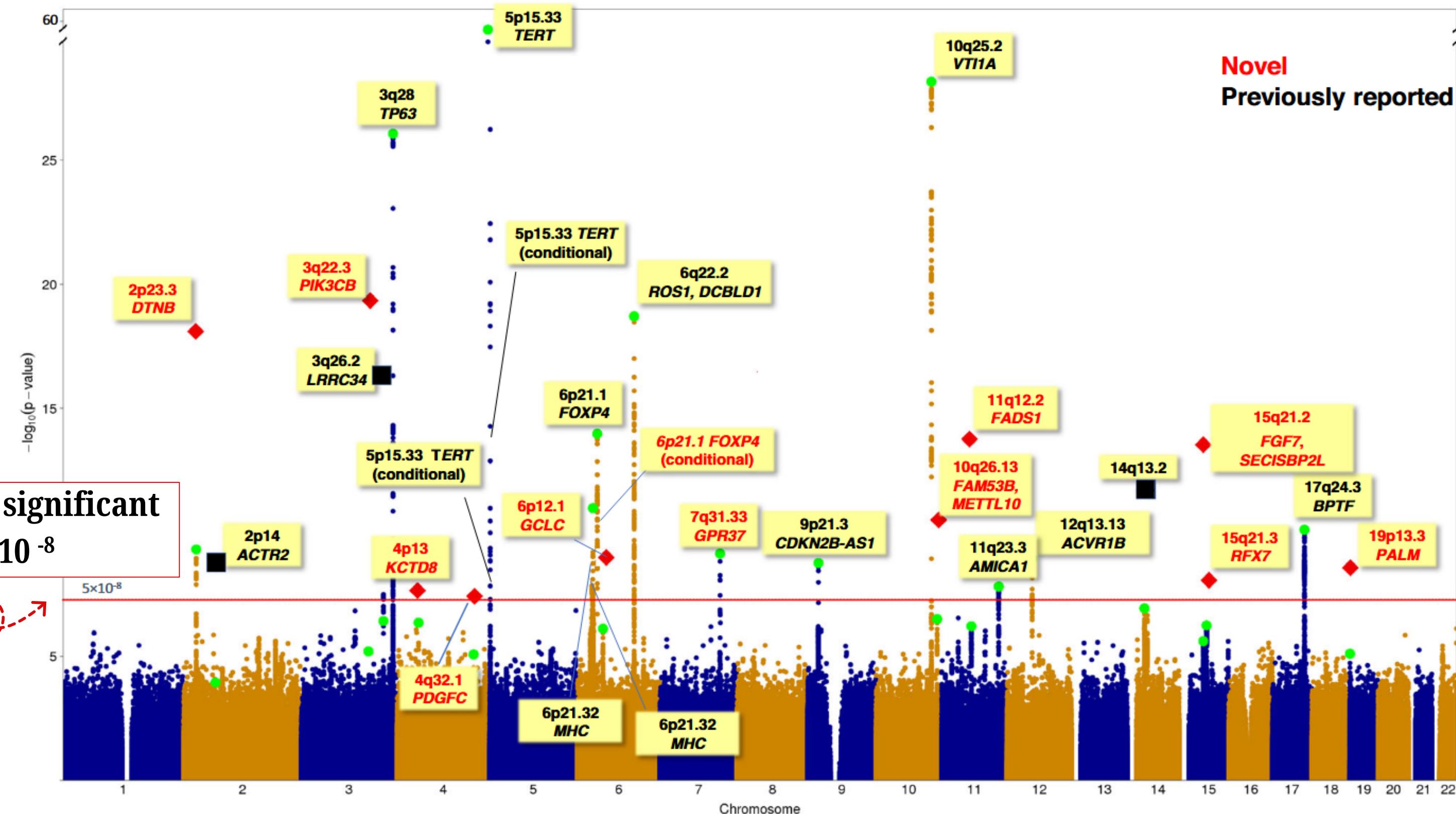
- Identify inherited variants associated with cancer risk or susceptibility
- Predict disease susceptibility or risk before onset (PRS)

Somatic GWAS

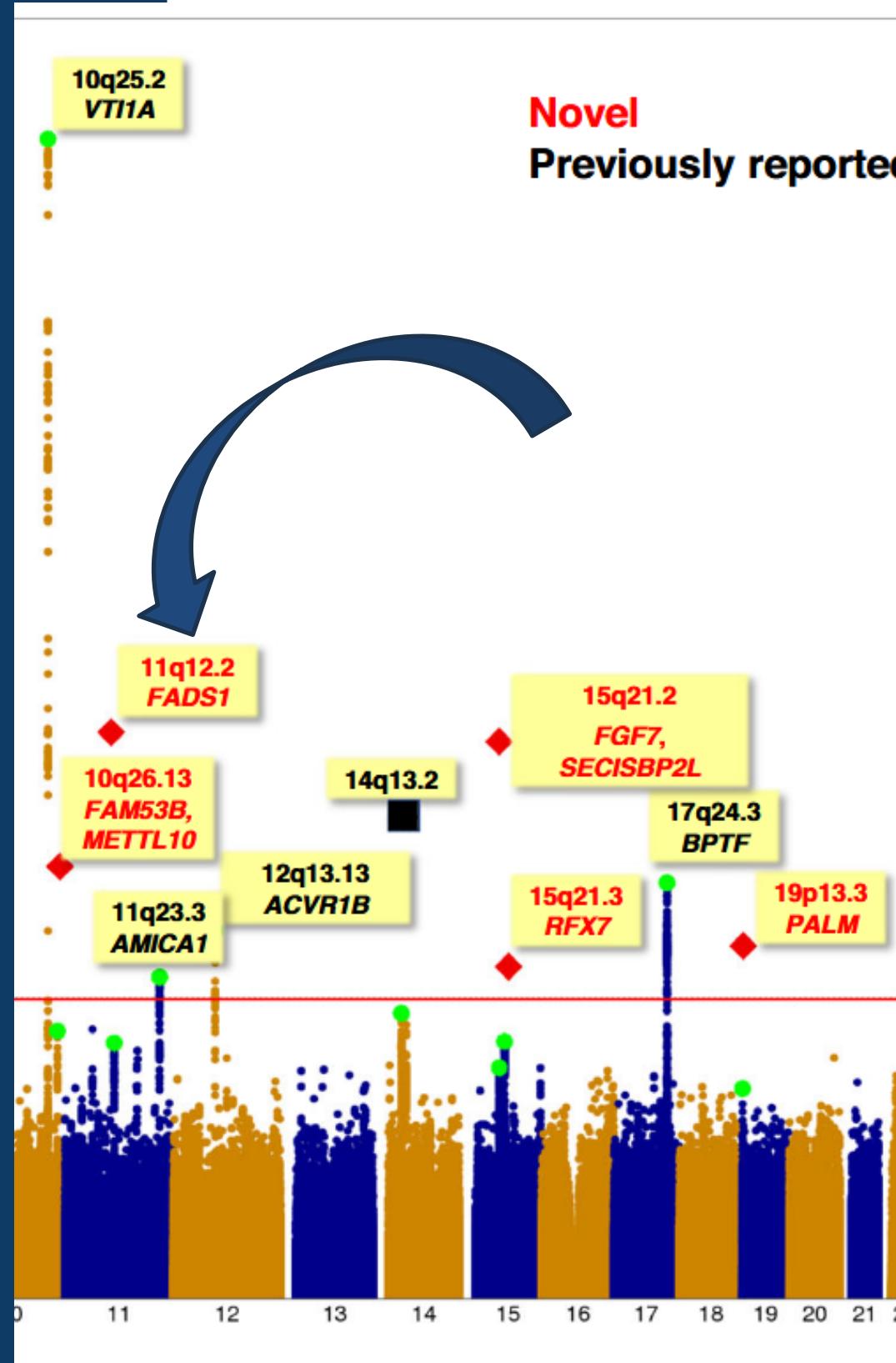
- Identify somatic mutations associated with tumor development, progression, or heterogeneity



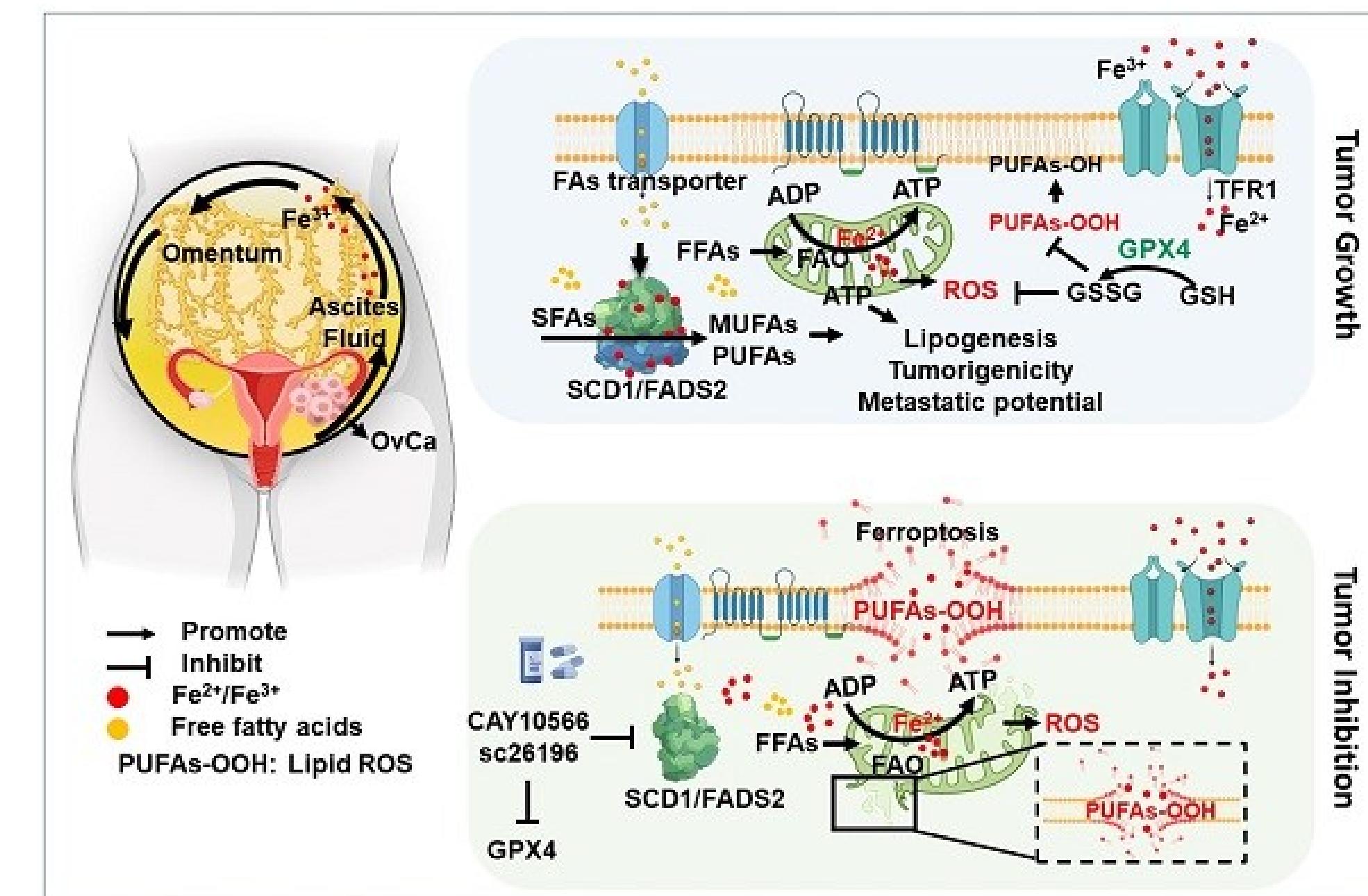
GWAS of Lung cancer in East Asian compare European population



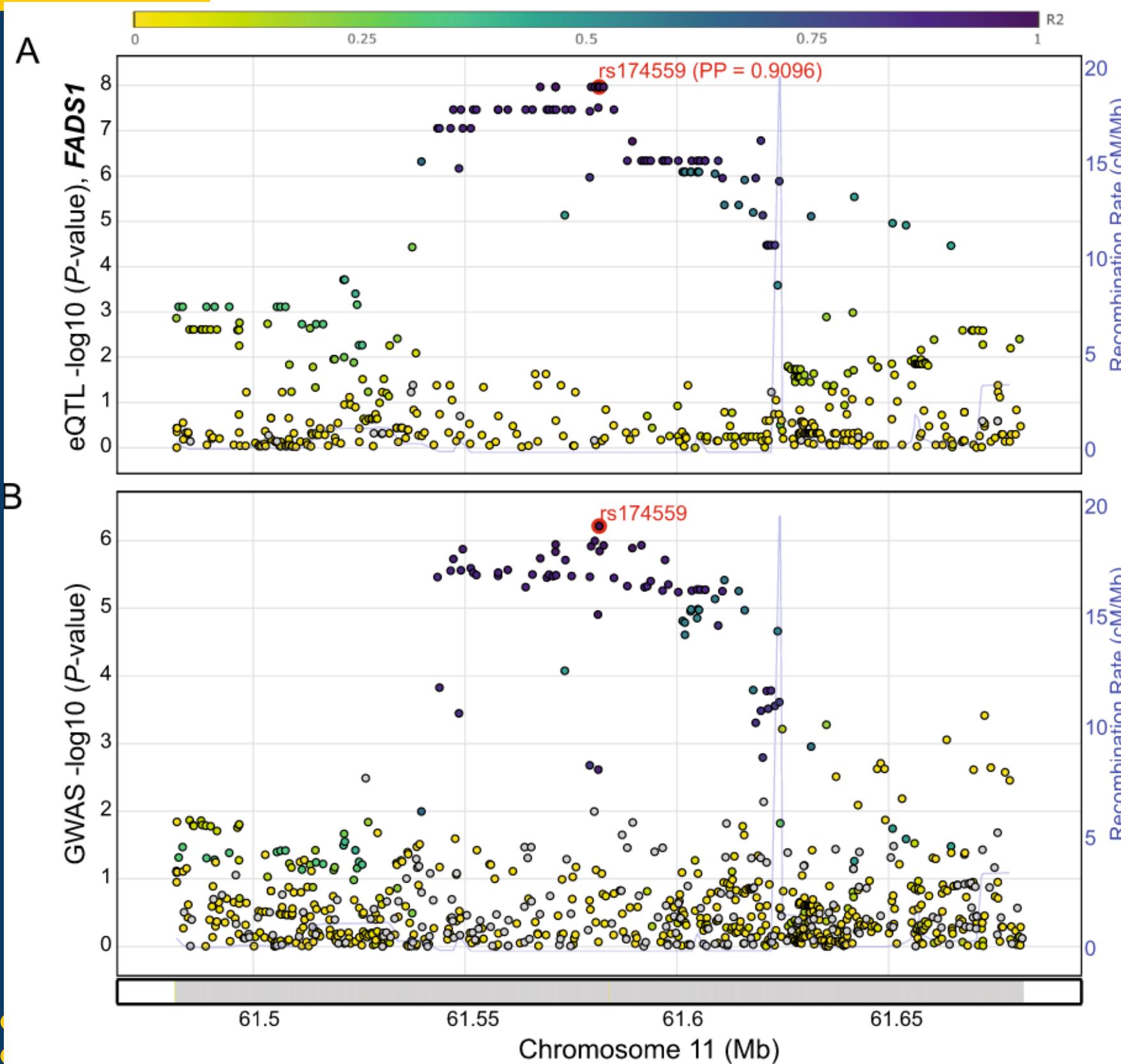
GWAS of Lung cancer in East Asian compare European population



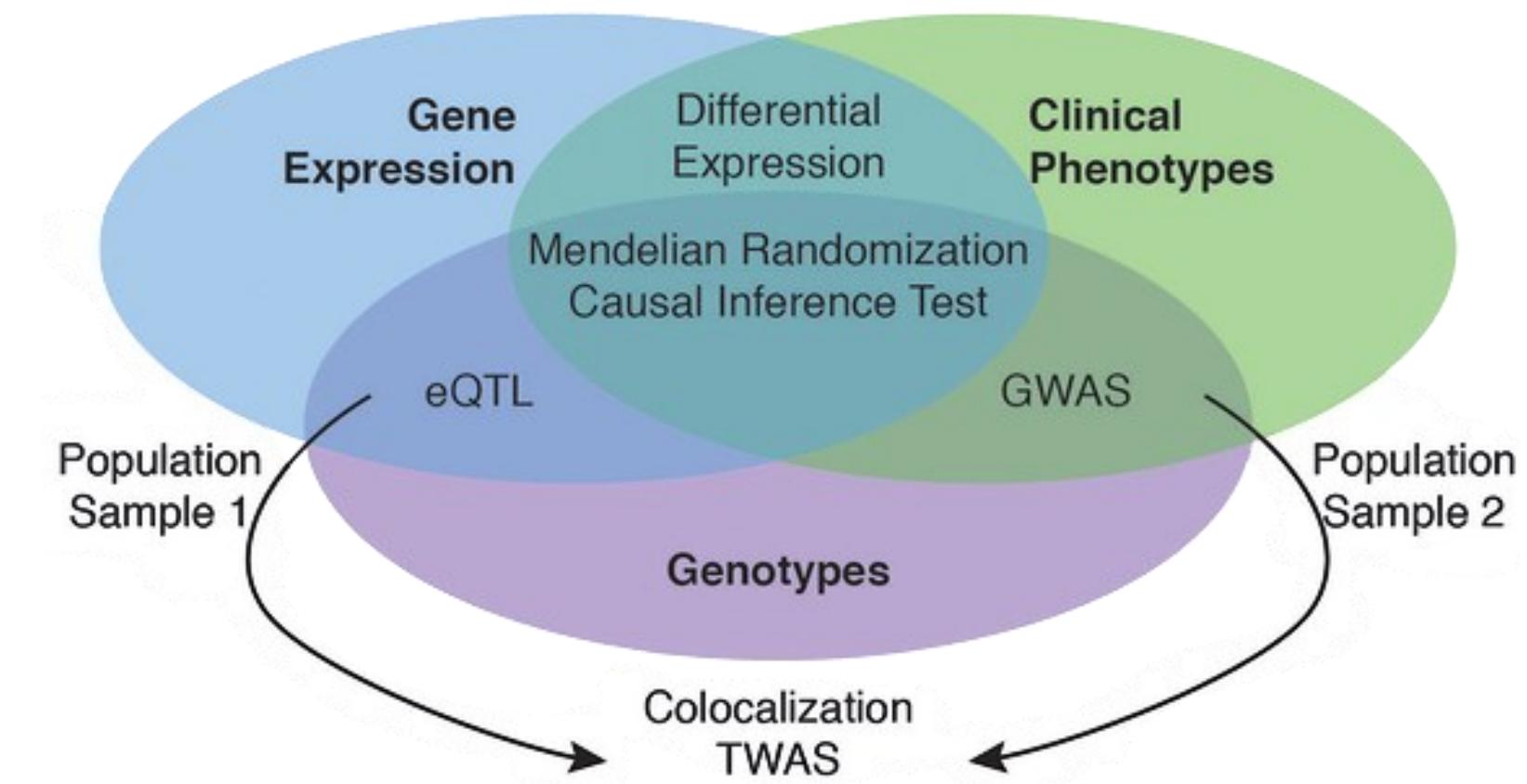
FADS1 (Fatty Acid Desaturase 1) an enzyme involved in the biosynthesis of long-chain polyunsaturated fatty acids (LC-PUFAs)



GWAS of Lung cancer in East Asian compare European population



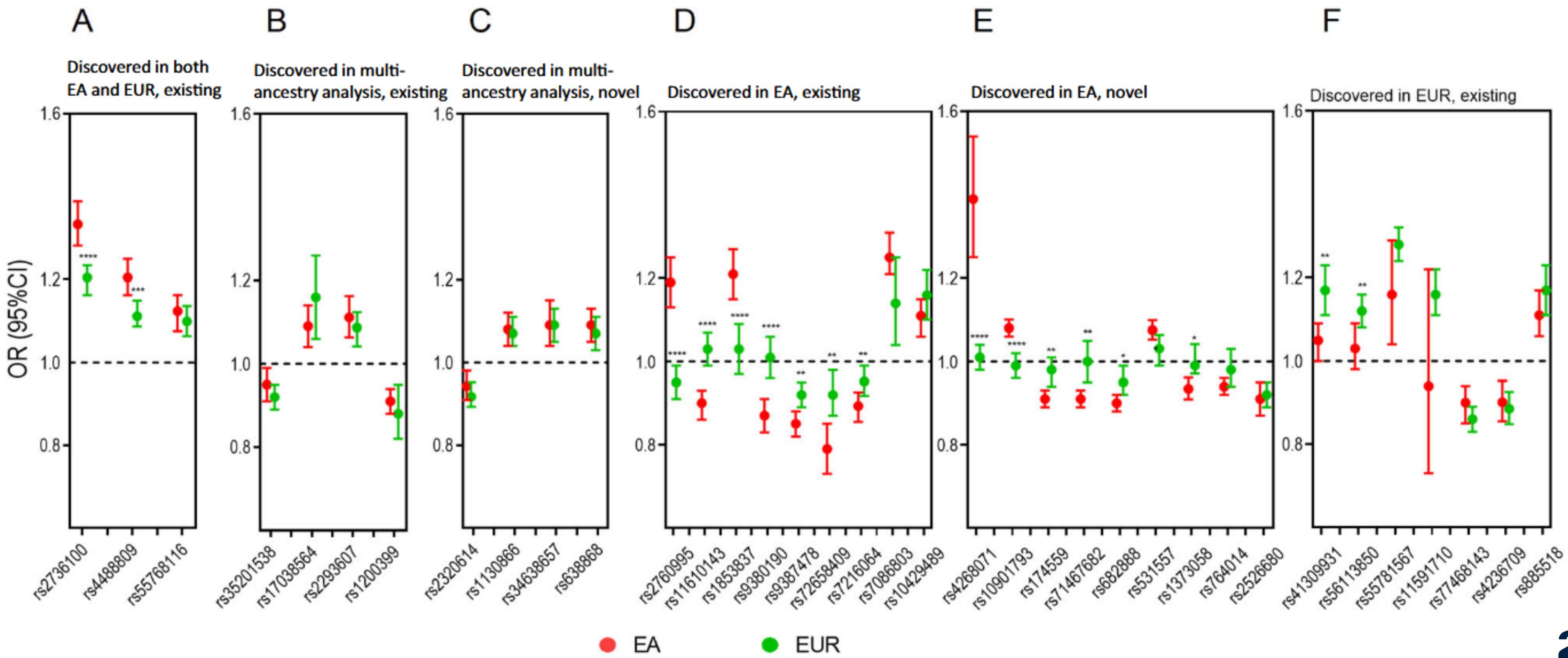
Colocalization of lung adenocarcinoma
GWAS signal from the new locus on Chr11
with *FADS1* eQTL signal.



Posterior probability (PP) > 0.8
= Strong colocalization

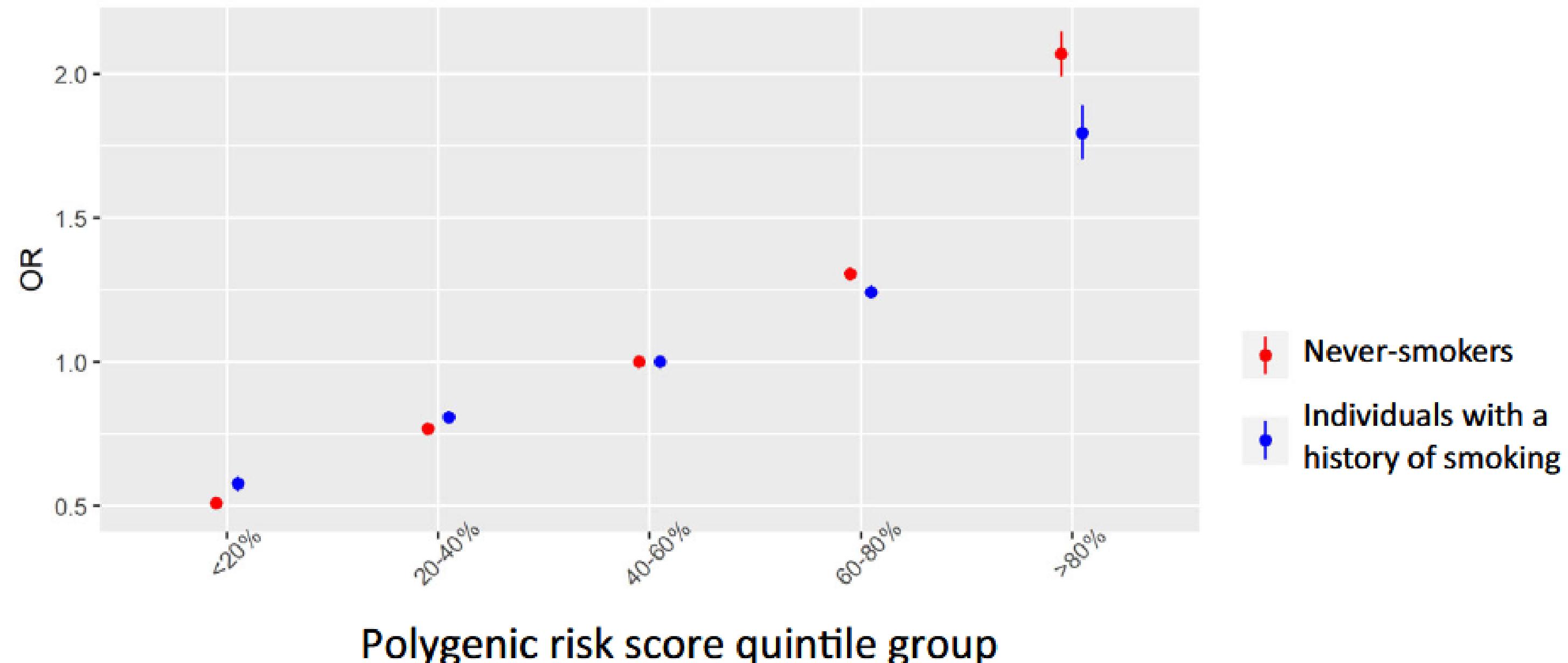
GWAS of Lung cancer in East Asian compare European population

Comparing odds ratios of lung adenocarcinoma susceptibility variants between East Asian and European populations.

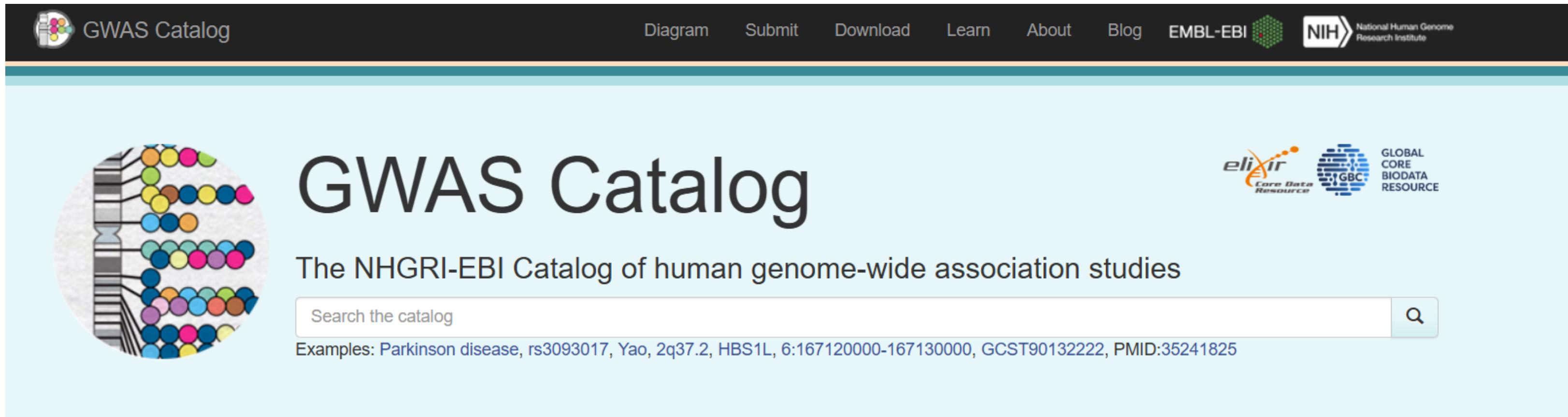


GWAS of Lung cancer in East Asian compare European population

A polygenic risk score (PRS) is more strongly associated with risk of lung adenocarcinoma in never-smokers than in individuals with a history of smoking ($P = 0.0058$)



GWAS database (<https://www.ebi.ac.uk/gwas/>)



The screenshot shows the main landing page of the GWAS Catalog. At the top, there's a navigation bar with links for "Diagram", "Submit", "Download", "Learn", "About", "Blog", "EMBL-EBI", and "NIH". Below the navigation is a large, stylized circular graphic composed of colored dots and lines, representing genetic data. To the right of this graphic, the title "GWAS Catalog" is displayed in a large, bold, black font. Underneath the title, the subtitle "The NHGRI-EBI Catalog of human genome-wide association studies" is shown. A search bar with the placeholder "Search the catalog" and a magnifying glass icon is positioned below the subtitle. Below the search bar, a line of text provides examples of search terms: "Examples: Parkinson disease, rs3093017, Yao, 2q37.2, HBS1L, 6:167120000-167130000, GCST90132222, PMID:35241825". In the top right corner of the main content area, there are logos for "elixir Core Data Resource" and "GCB Global Core Biodata Resource".

Are you accessing the GWAS catalog programmatically? The newly designed REST API v2.0 is here! Find out more in our [blog post](#).

Download

Download a full copy of the GWAS Catalog in spreadsheet format as well as current and older versions of the GWAS diagram in SVG format.

Summary statistics

Documentation and access to full summary statistics for GWAS Catalog studies where available.

Submit

Submit summary statistics to GWAS Catalog.



Thank you for
your attention

