POLYGENIC RISK SCORE

From clinic to disease risk assessment

PRESS RELEASE · 5 MINUTE READ · MAR | 2 | 2020

Massachusetts General Hospital launches new Preventive Genomics Clinic

Julie Cunningham · 617-724-6433 · julie.cunningham@partners.org

Key Takeaways

- New clinic will empower patients to better understand, predict and preve genetic information.
- The clinic will be embedded within primary care practices at MGH and a 'eConsult' program

> *We believe DNA testing will be a key piece of routine care in the future.*

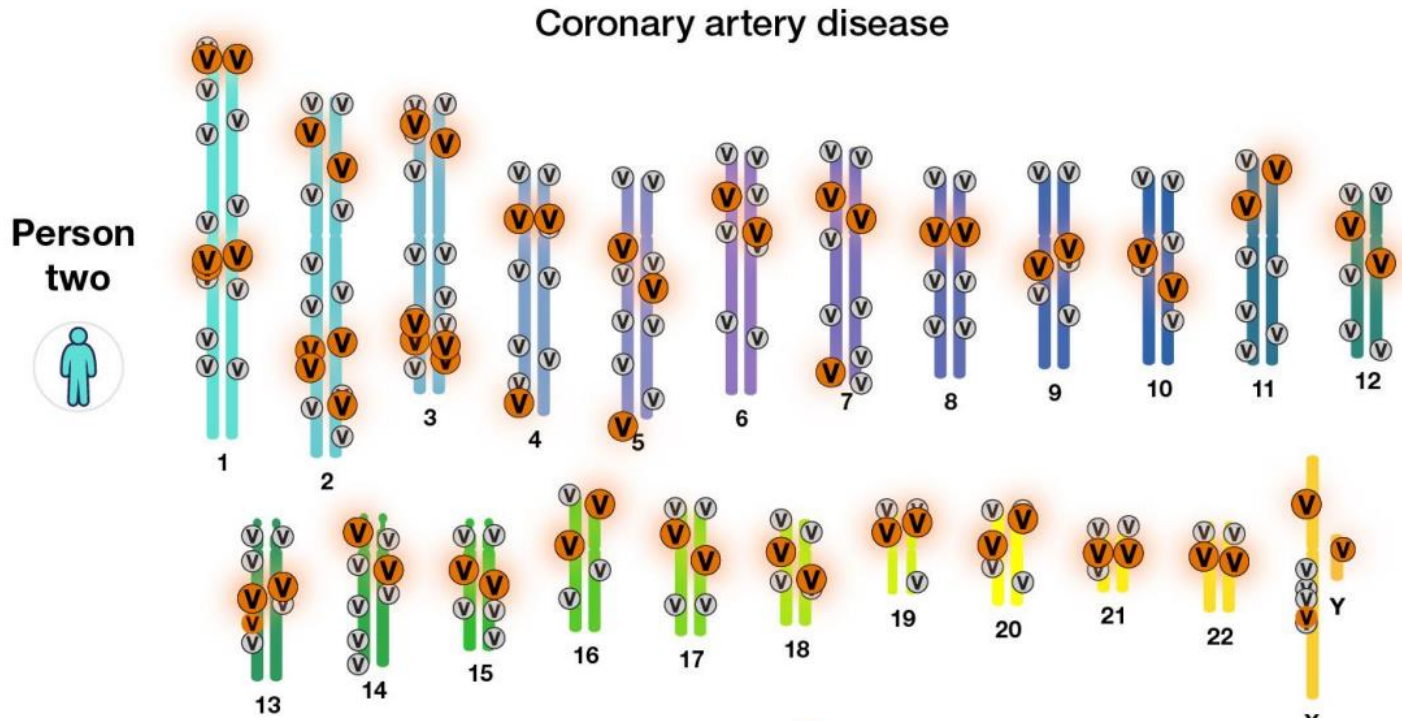
Amit V. Khera, MD
Co-founder, Mass General Preventive Genomics Clinic

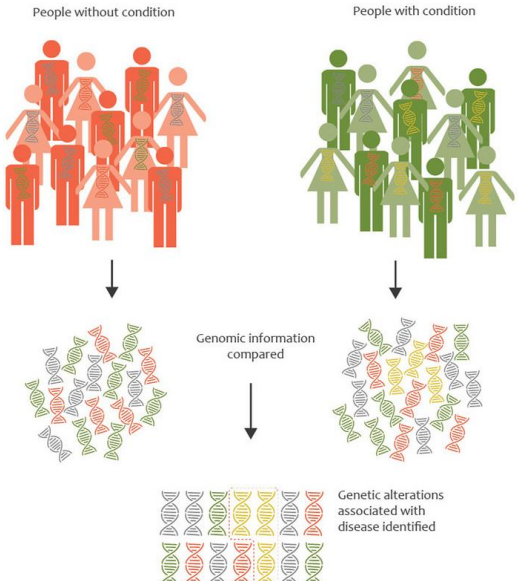
Table 1 | Prevalence and disease associations of high-risk PRS for six diseases in MGBB overall and by reported race

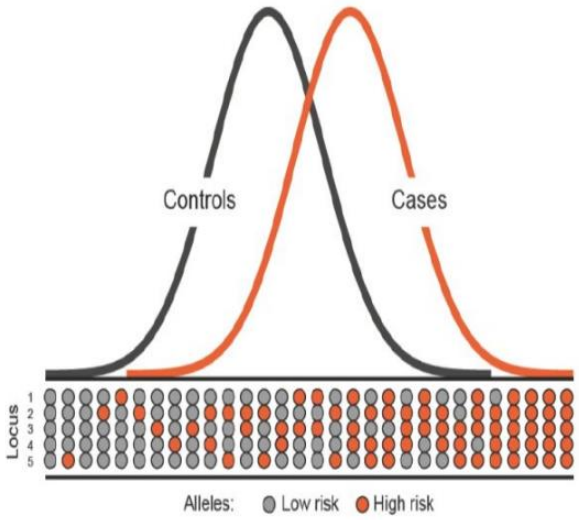
Disease	High risk (%)[a]	OR overall	OR white	OR Black	OR Asian	OR Other/Unknown
		OR (95% CI)[b] (*n*/*n*, *n*/*n*)[c]	OR (95% CI)[b] (*n*/*n*, *n*/*n*)[c]	OR (95% CI)[b] (*n*/*n*, *n*/*n*)[c]	OR (95% CI)[b] (*n*/*n*, *n*/*n*)[c]	OR (95% CI)[b] (*n*/*n*, *n*/*n*)[c]
BrCa	8.6	2.38 (2.07–2.73) (286/1,400, 1,427/16,606)	2.39 (2.07–2.76) (270/1,156, 1,318/13,495)	2.24 (0.97–5.15) (7/73, 43/1004)	0.51 (0.07–3.9) (1/33, 24/405)	2.35 (1.08–5.1) (8/138, 42/1,702)
CRCa	5.4	2.37 (1.74–3.24) (46/1,913, 346/34,117)	2.29 (1.65–3.19) (41/1,646, 312/28,717)	4.11 (1.17–14.48) (3/83, 15/1706)	0 (0–NaN) (0/35, 7/744)	3.30 (0.73–14.88) (2/149, 12/2,950)
PrCa	13.1	2.22 (1.98–2.48) (498/1,698, 1,693/12,813)	2.31 (2.05–2.59) (468/1,448, 1,544/11,017)	1.39 (0.74–2.59) (14/71, 74/521)	2.58 (0.5–13.28) (2/36, 6/279)	1.41 (0.78–2.58) (14/143, 69/996)
AFib	8.3	2.37 (2.12–2.64) (450/2,589, 2,282/31,101)	2.40 (2.14–2.69) (422/2,179, 2,101/26,014)	1.47 (0.72–3.01) (9/137, 71/1590)	2.00 (0.57–7.03) (3/62, 17/704)	2.28 (1.32–3.94) (16/211, 93/2,793)
CAD	9.8	1.86 (1.69–2.05) (562/3,018, 2,991/29,851)	1.91 (1.73–2.12) (503/2,459, 2,680/25,074)	1.41 (0.86–2.29) (21/177, 125/1484)	3.96 (1.79–8.76) (9/51, 31/695)	1.47 (0.97–2.22) (29/331, 155/2,598)
T2D	8.4	1.75 (1.57–1.95) (439/2,612, 2,924/30,447)	1.93 (1.71–2.17) (367/2,284, 2,159/25,906)	1.21 (0.7–2.09) (18/57, 358/1374)	1.07 (0.37–3.08) (4/49, 52/681)	1.58 (1.14–2.19) (50/222, 355/2,486)

Complex diseases (Polygenic disease)

ZoteroToExcalidraw.md

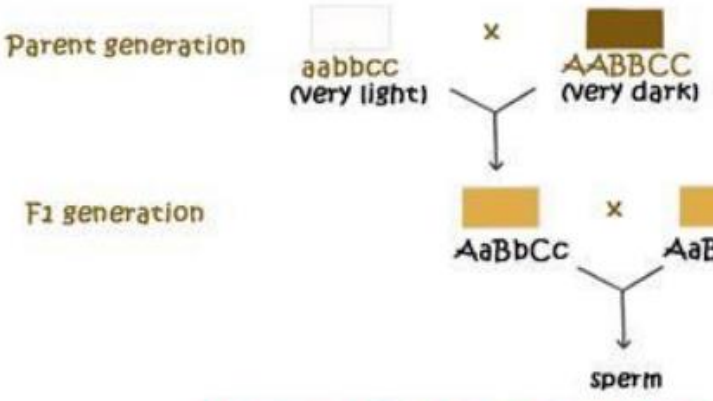






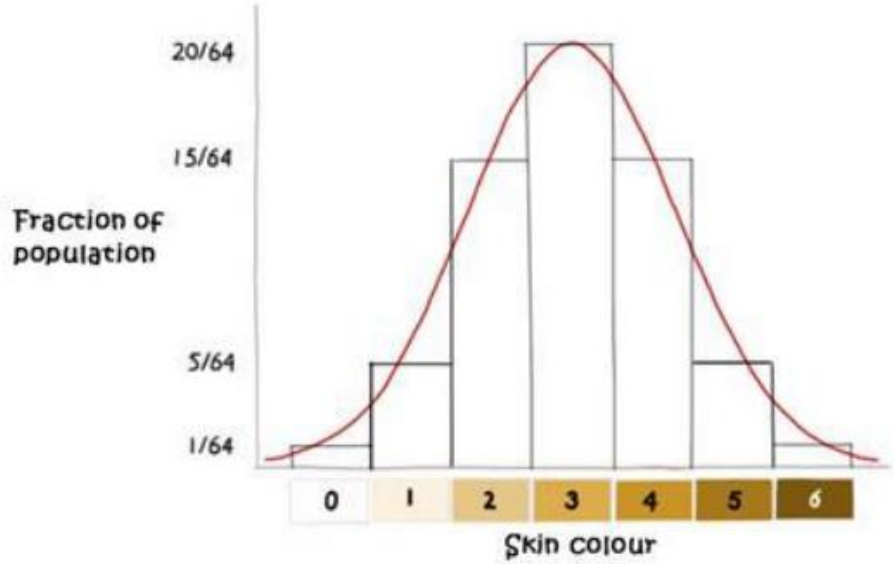
The distribution of risk alleles in both cases and controls follows a normal distribution. However, cases have a shift towards a higher number of high risk alleles.

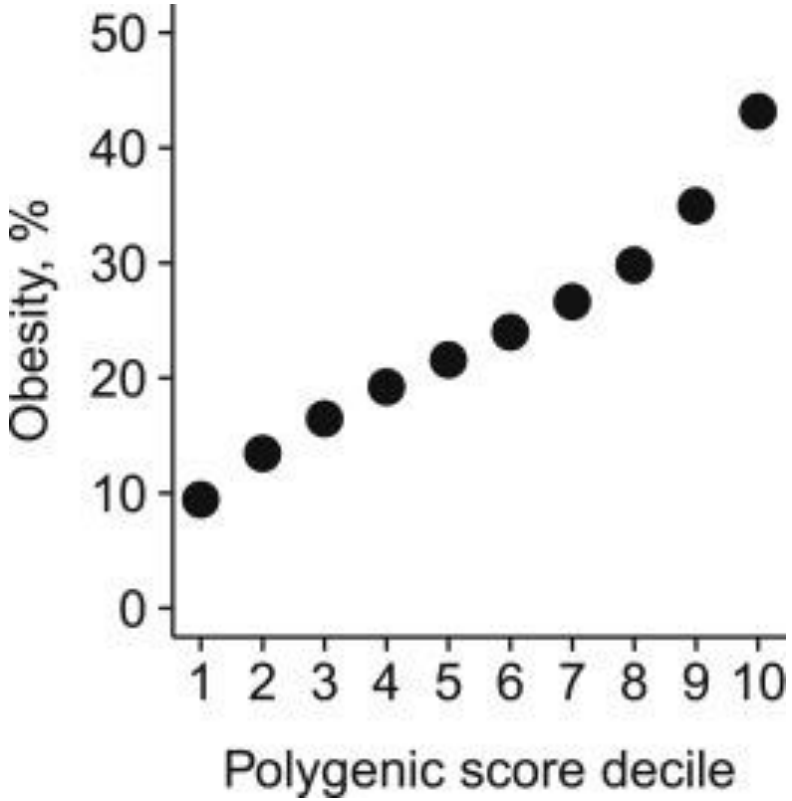
Whiffin and Houlston, Genes 2014, 5, 270-284

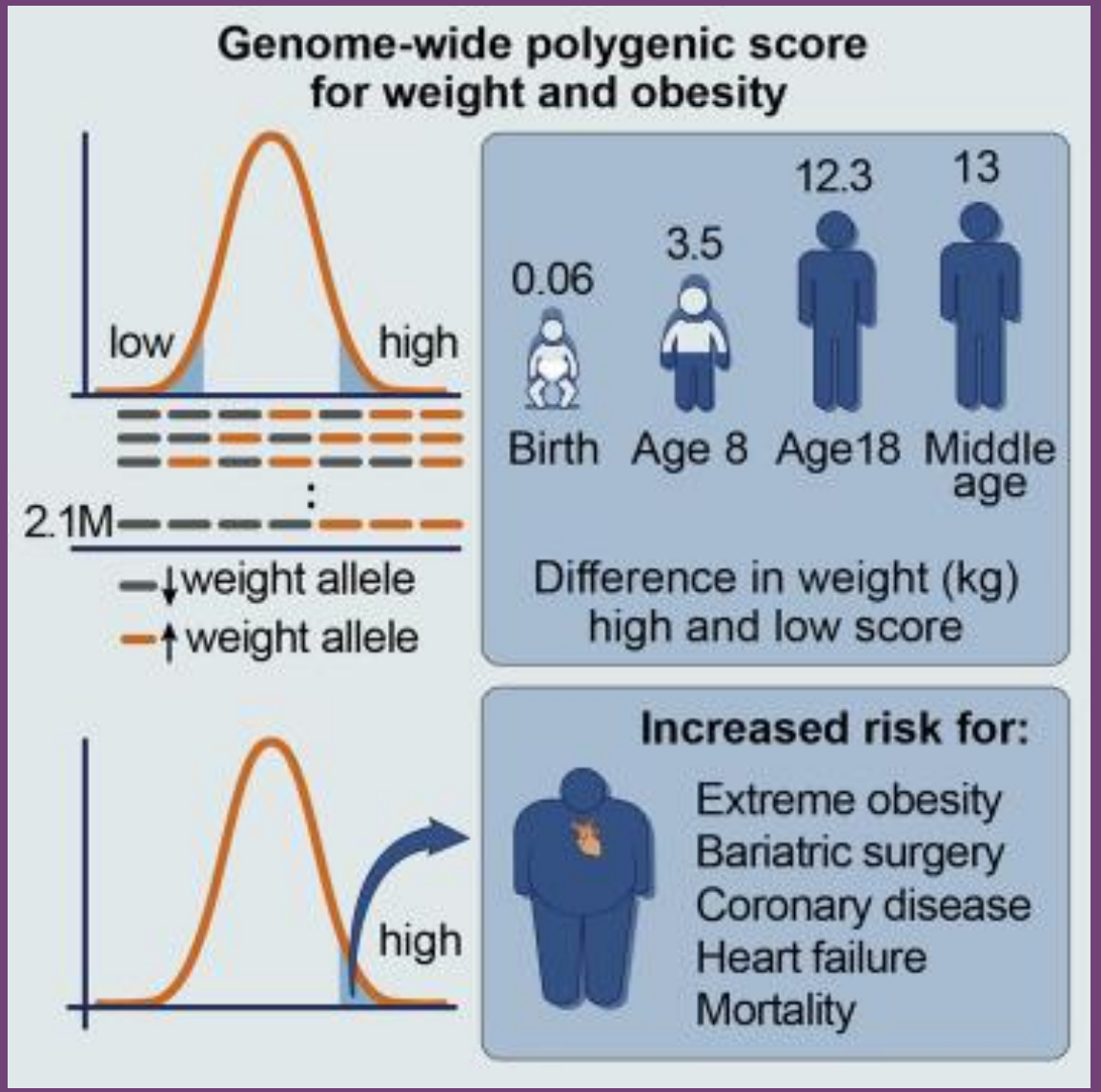


F2 generation

eggs \ Gametes	ABC	ABC	AbC	Abc	aBC	aBc	abC	abc
ABC	6	5	5	4	5	4	4	3
ABC	5	4	4	3	4	3	3	2
AbC	5	4	4	3	4	3	3	2
Abc	4	3	3	2	3	2	2	1
aBC	5	4	4	3	4	3	3	2
aBc	4	3	3	2	3	2	2	1
abC	4	3	3	2	3	2	2	1
abc	3	2	2	1	2	1	1	0







Cell. 2019 Apr 18;177(3):587-596.e9.

MULTIPLE LINEAR REGRESSION FRAMEWORK

***p* vector of snp effect size**

***n* 個人的 phenotype 向量** → $y = X\beta + \varepsilon$ ← ***n*-vector of riidual error**
each element follows an independent normal distribution

***n* 個人 x *p* snp 的 genotype 矩陣**
Genotype 常以 reference allele 的數量 (0, 1, 2) 表示，
也可以是 impute 後的連續數值

$$GPS = \sum_{snp=i}^{n} Beta_i * Genotype\_Score_i$$

- *Effect size estimated from independent samples*
- *SNPs were pruned to be independent with each other.*

- 遺傳分佈是 sparse or polygenic 通常未知，故有靈活的建模假設會有較好的表現
- 超參數的推理策略會影響準確性 (如 LD window size …)
- 有個人級的資料則無須LD推論，故通常表現更好

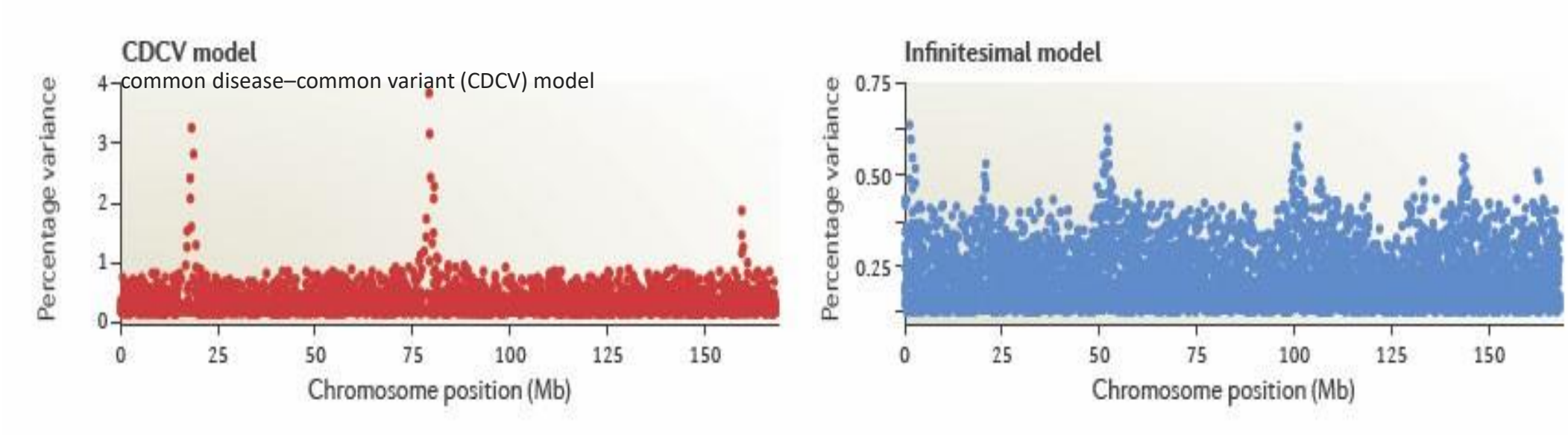
THEORETICAL PRS ACCURACY

$$R^2 \approx \frac{h_m^2}{1 + \frac{m}{Nh_m^2}}$$

h_m^2 : True variance explained by the predictor depends on the m SNP set.
N: GWAS sample size

- PRS have a **theoretical** upper limit dependent on the **broad sense trait heritability**
- PRS have a **technical** upper limit associated with the proportion of **variance tagged** by the DNA variants measured (or **SNP-based heritability** as we are using GWAS data).
- PRS have a **practical** upper limit dependent on the **sample size of the discovery dataset** used to estimate effect sizes of risk alleles, and the **quality** of the discovery data.
- PRS can be pushed closer to the technical upper limit by the **statistical methodology** used to generate the optimal weighting given to the risk alleles, and new methods integrate new biological data (e.g., functional annotations).

DIFFERENT EXPECTED SIGNATURES FROM GENOME-WIDE ASSOCIATION STUDIES FOR FOUR MODELS OF DISEASE.



Sparse modeling

假設只有一小部分的SNP具有非零效應

effect size 為 0

$$\beta_j \sim \pi\, N\left(0, \sigma_\beta^2\right) + (1-\pi)\delta_0$$

Snp effect size 為 normal distribution 的比例

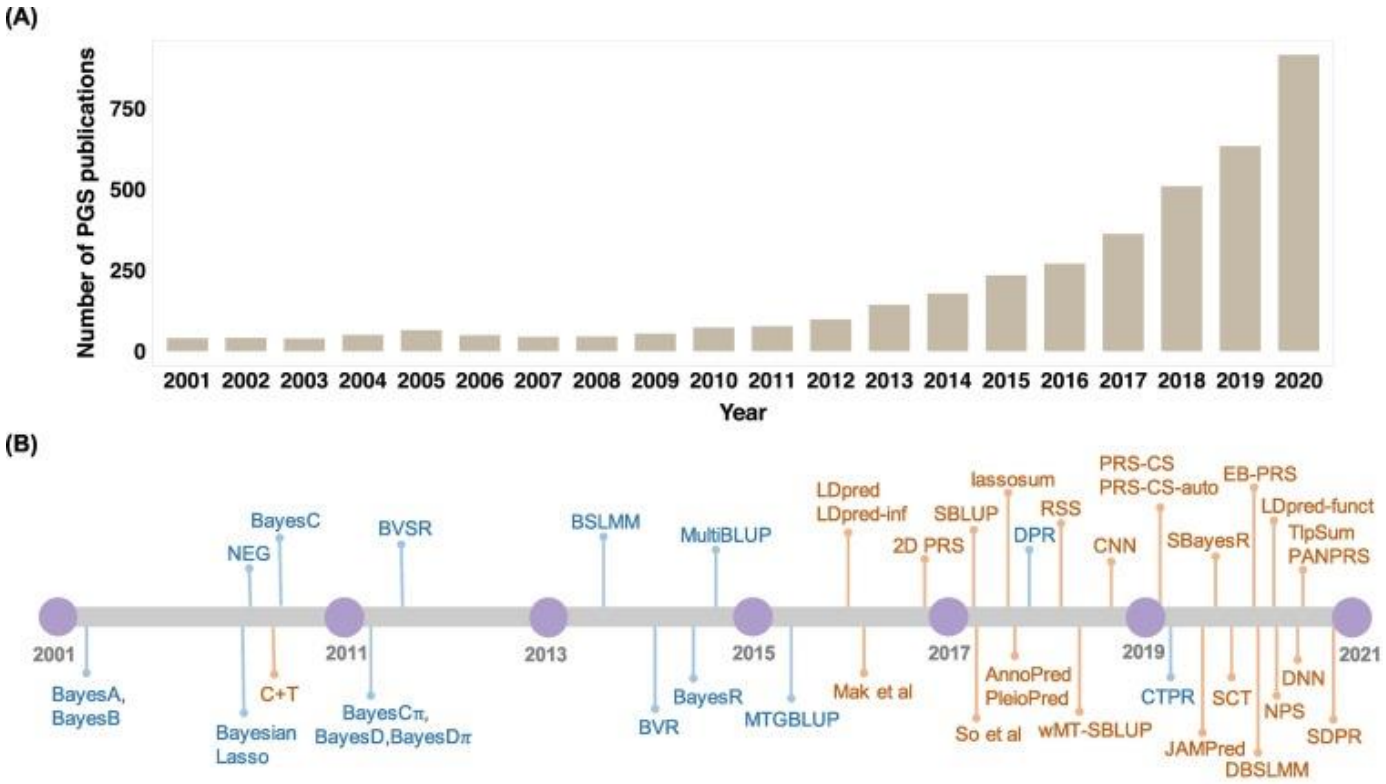
Polygenic modeling

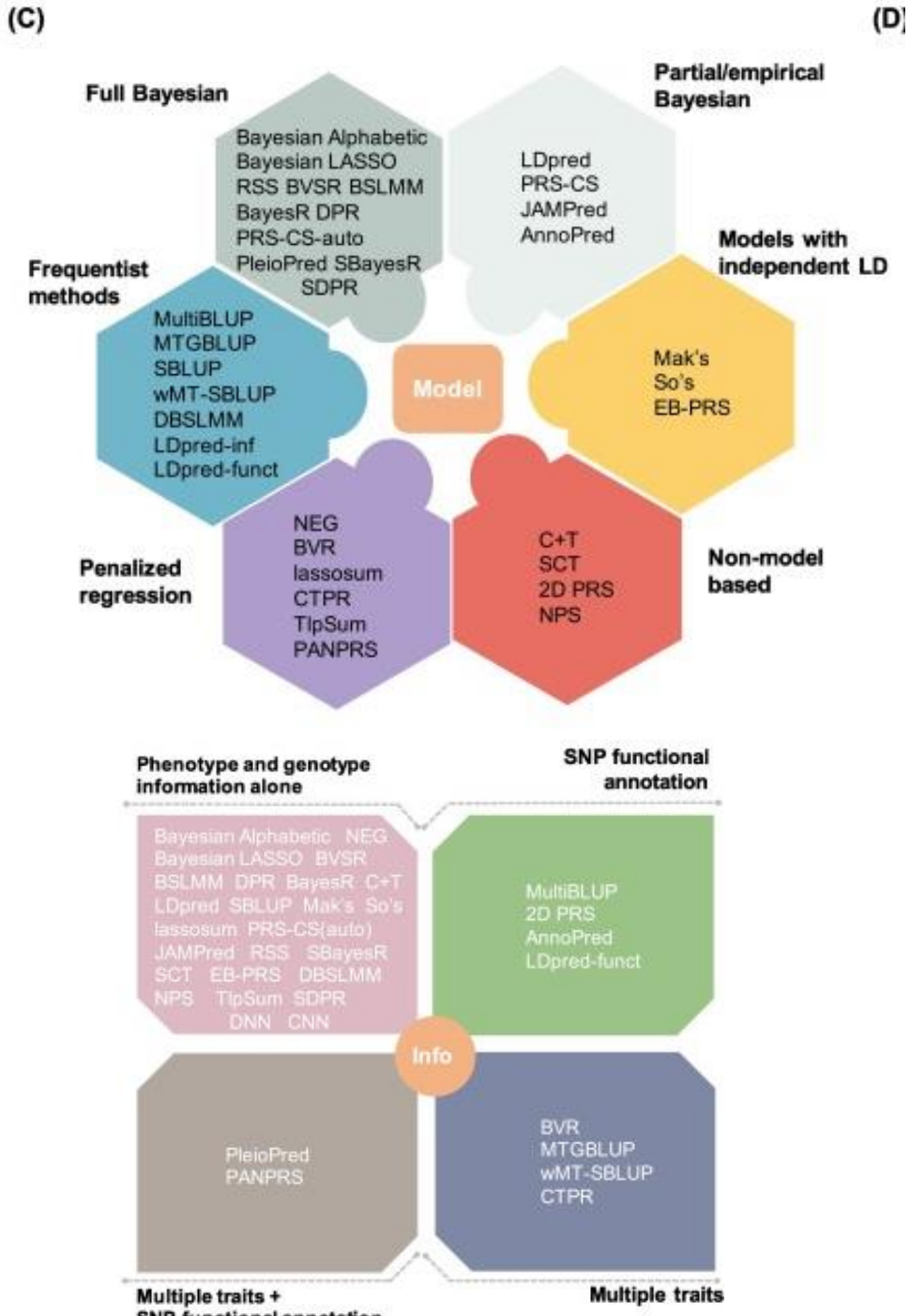
所有SNP效應皆非零

$$\beta_j \sim N\left(0, \sigma_\beta^2\right)$$

https://www.nature.com/articles/nrg3118

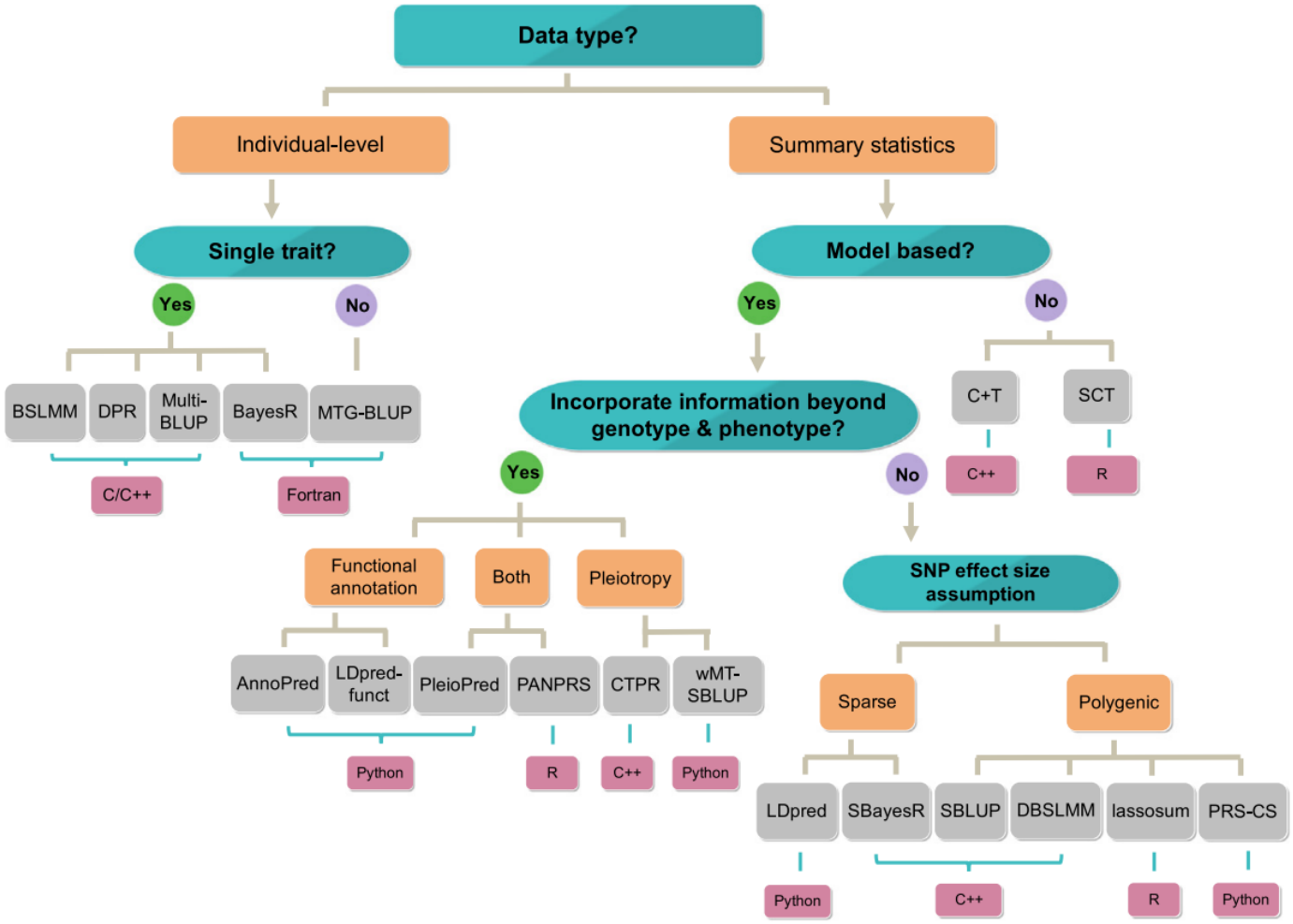
GENETIC PREDICTION OF COMPLEX TRAITS WITH POLYGENIC SCORES: A STATISTICAL REVIE



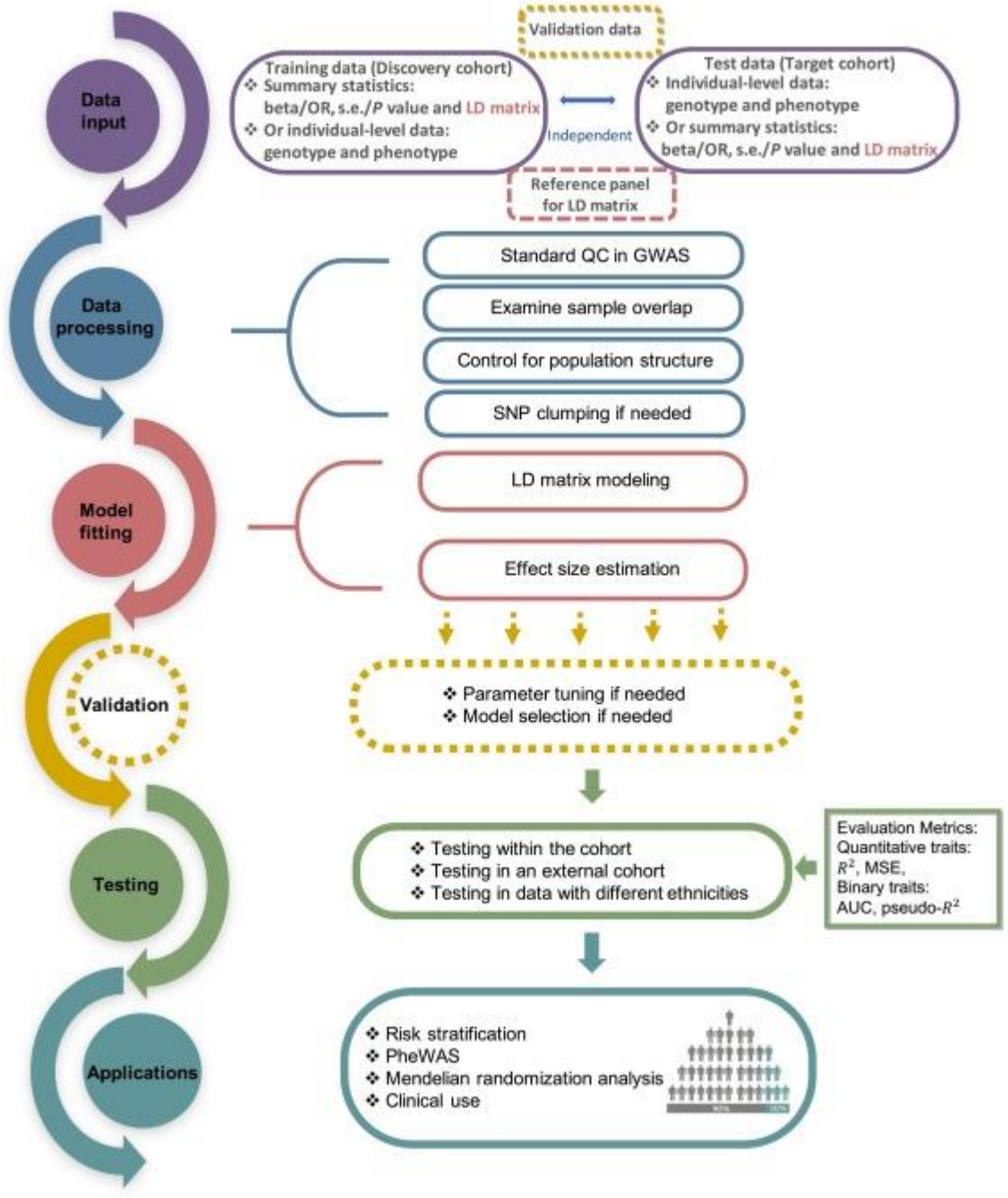


https://doi.org/10.1016/j.tig.2021.06.004

DECISION FOR PRS

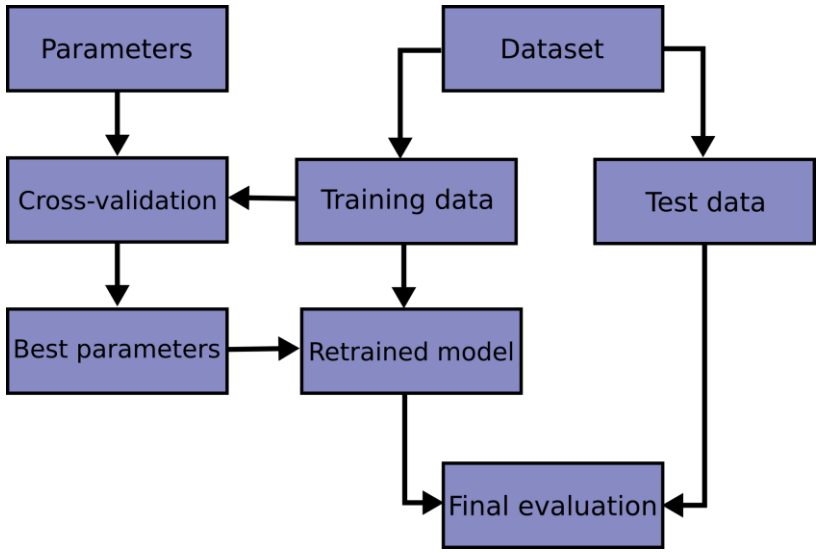


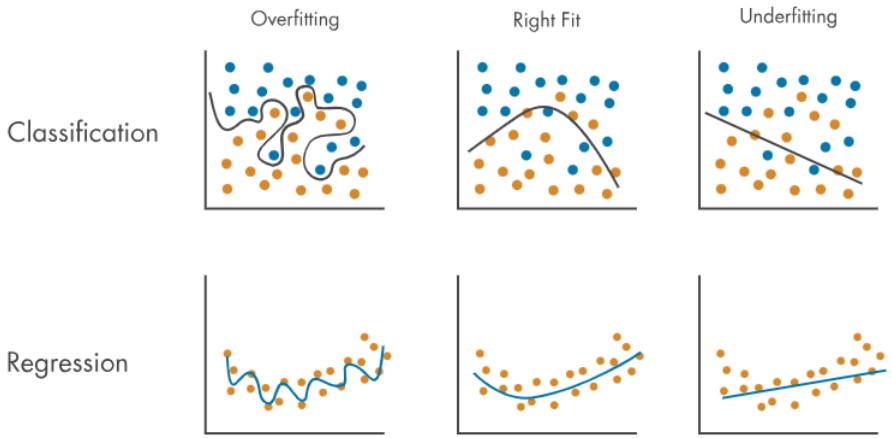
Trends in Genetics

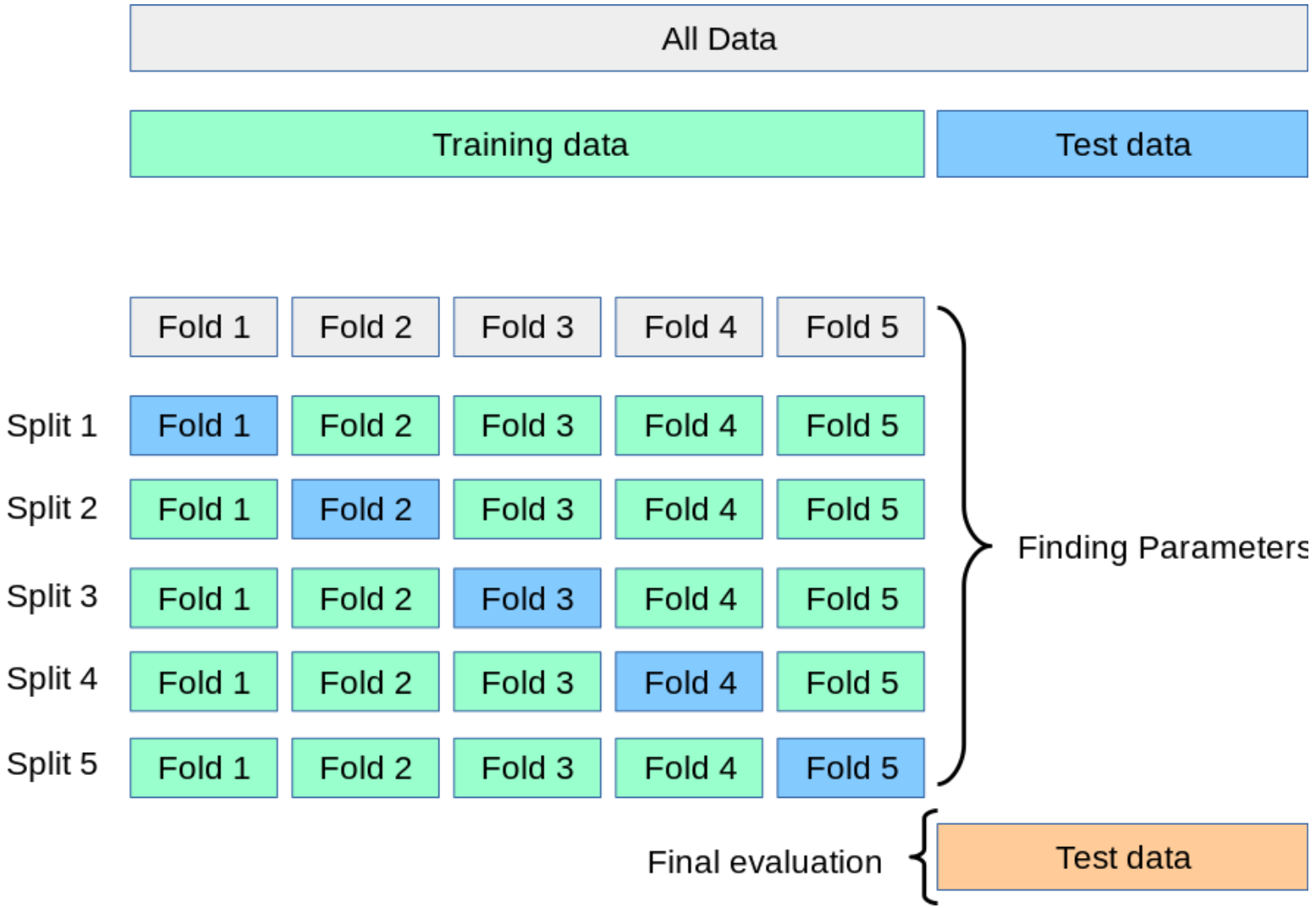


Trends in Genetics

EVALUATING MODEL PERFORMANCE
TRAINING, TESTING AND CROSS-VALIDATION

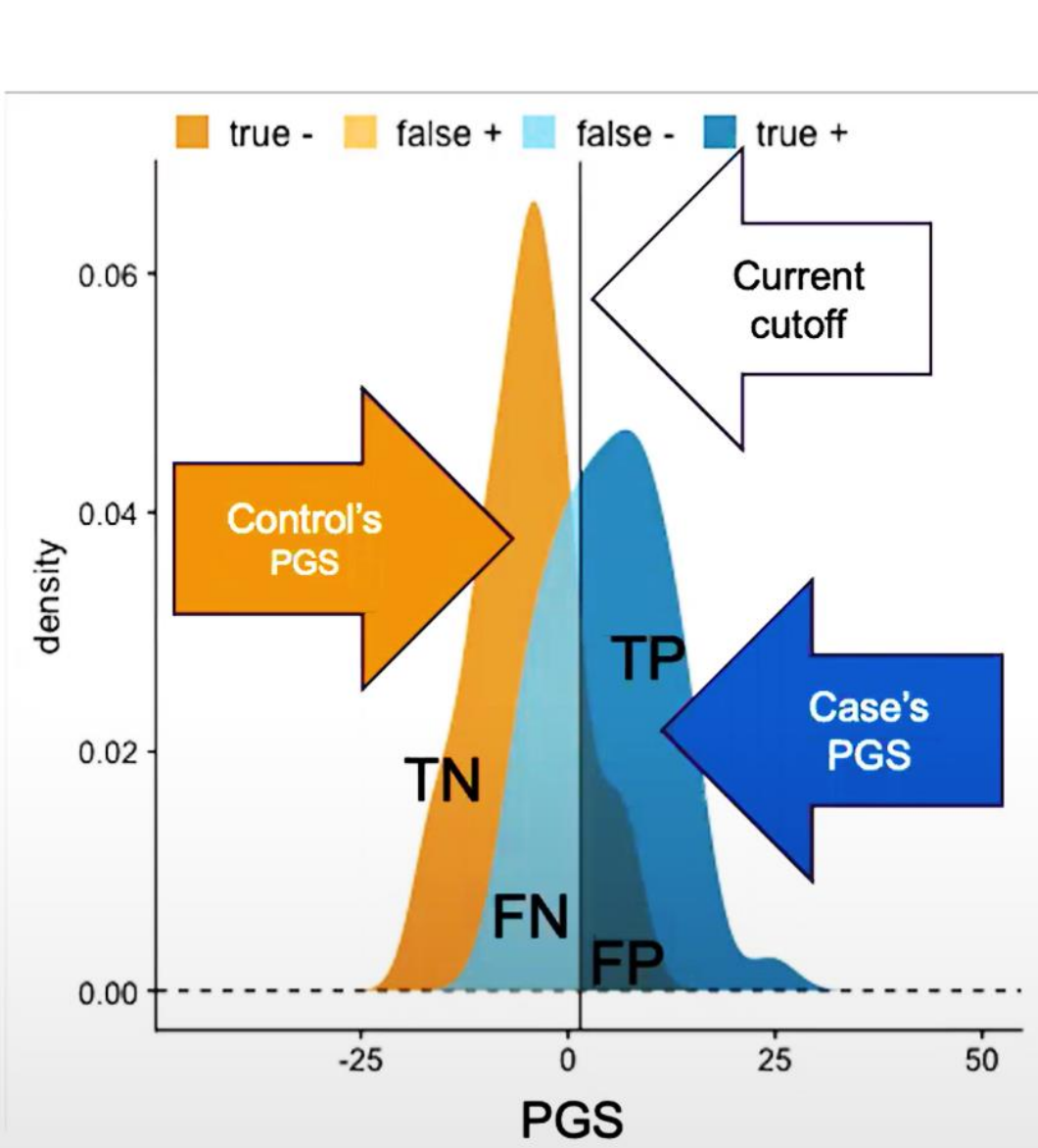


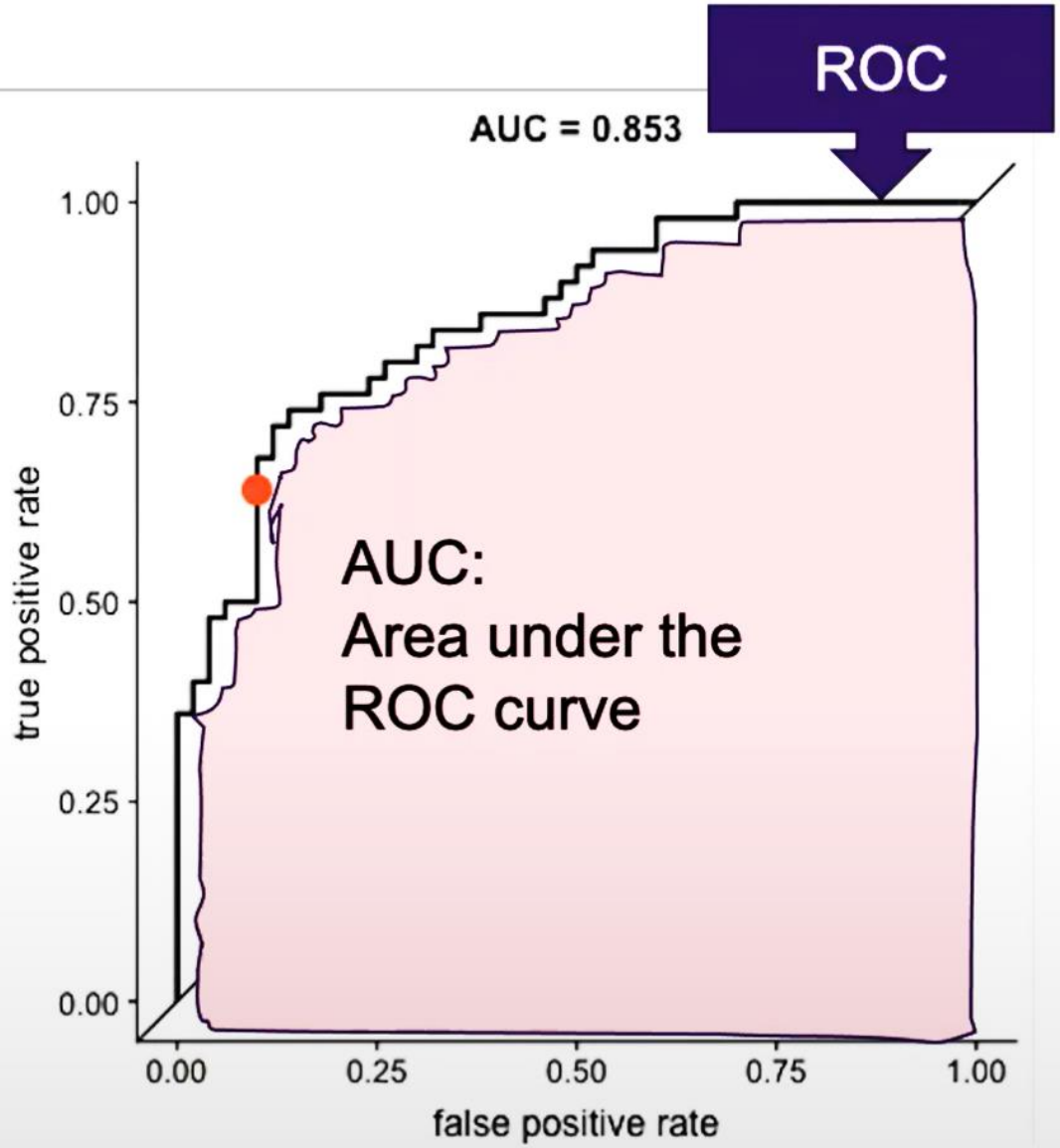




https://scikit-learn.org/stable/modules/cross_validation.html
https://www.mathworks.com/discovery/overfitting.html

AREA UNDER CURVE (AUC)





Receiver
Operating
Characteristic
curve

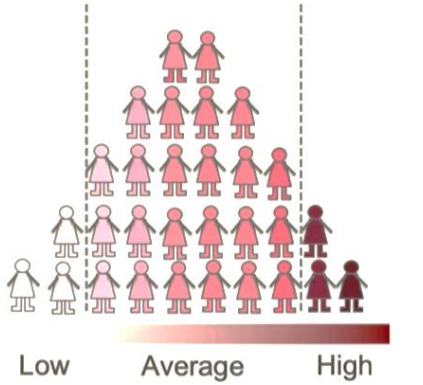
GENETIC TEST NEED TO BE VALID AND USEFUL

Analytical Validity

Analytical Validity | Clinical Validity | Clinical Utility | Implement

Clinical Validity

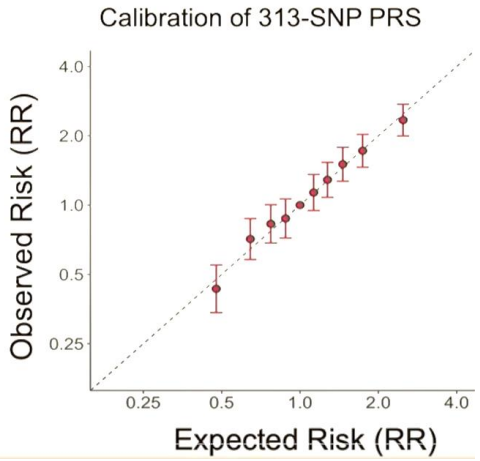
Calibration: how close is the predicted risk to actual risk?



15 Prospective cohorts

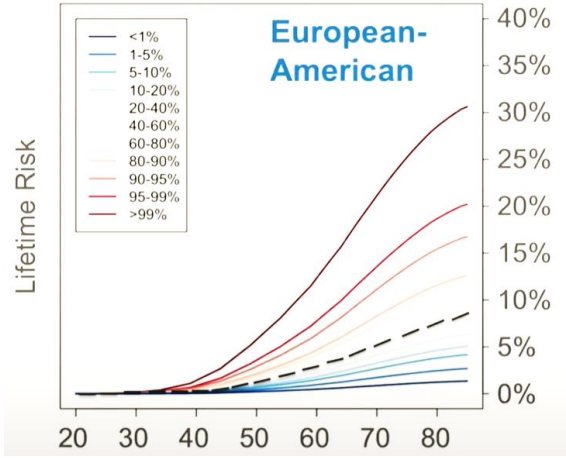
230,000 women

6,000 breast cancers



Clinical Validity

Risk discrimination: how well can we separate people by risk?



Clinical Utility

- Acceptability, uptake
- Delivery in health care system
- Cost-effectiveness
- Ethical, legal, social issues

- Behavioral change
- Risk-based mammography
- Anti-estrogen medication
- Prophylactic mastectomy

Use-Case Example - UK Breast Cancer

- **Background** - Routine screening is not offered to women in their 40s. This is because the disease is less common in this group so mass screening becomes less cost-effective.
- **Problem** - Circa 8,000 new breast cancer cases diagnosed in this group every year in the UK. Symptomatic presentation has a worse outcome
  - 15% of women have actionable levels of breast cancer risk without being aware of it*
- **Solution** - Risk based screening invitations to women in their 40s to support a decision about screening
  - Offering screening in the 40-49 age group could identify circa 30% of cases in this currently unscreened group.
  - Genomic risk tool to stratify screening identifies more breast cancers than using rare variants

Population stratification of Breast Cancer Screening using either PRS or rare variants

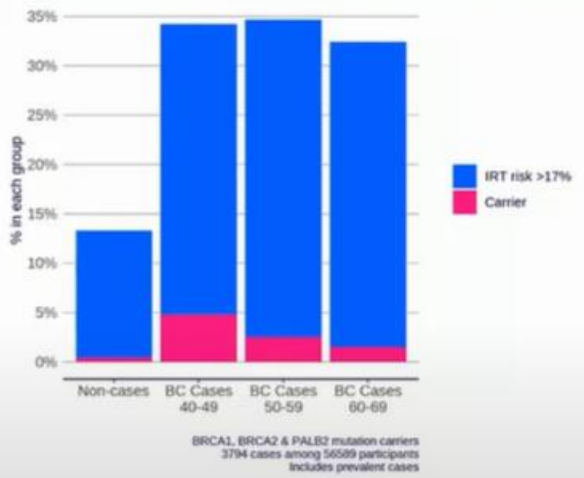


Figure: Genetic factors are the most powerful predictors of breast cancer. If used to stratify screening in the population, PRS (using routine clinical thresholds) will identify more cancers than highly penetrant rare-mutations.

GENOMICS

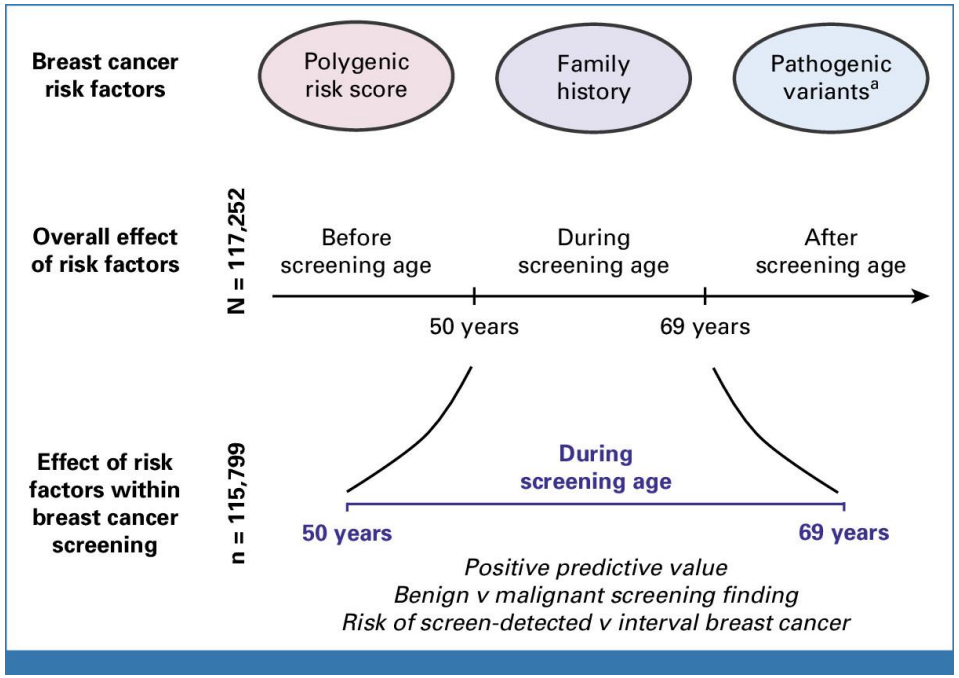
*Lifetime risk > 17% to age 80
Source: Genomics plc

資料來源：TBCA社團法人乳癌病友協會提供，詳見國健署及美國國家癌症協會2018，
https://www.setn.com/News.aspx?NewsID=1130455，

https://tw.news.yahoo.com/%E7%AF%A9%E6%AA%A2%E7%8E%87%E5%83%853%E6%88%903-%E5%8F%B0%E4%B9%B3%E7%99%8C%E5%A5%BD%E7%99%BC%E5%B9%B4%E9%BD%A1%E6%AF%94%E6%AD%90%E7%BE%8E%E6%97%A910%E6%AD%B2-%E9%86%AB%E5%BB%BA%E8%AD%B0%E8%A3%9C%E5%8A%A9%E9%AB%98%E9%A2%A8%E9%9A%AA%E6%97%8F%E7%BE%A4%E6%8F%90%E6%97%A9%E6%AA%A2%E6%9F%A5-160000554.html

Journal of Clinical Oncology
An American Society of Clinical Oncology Journal

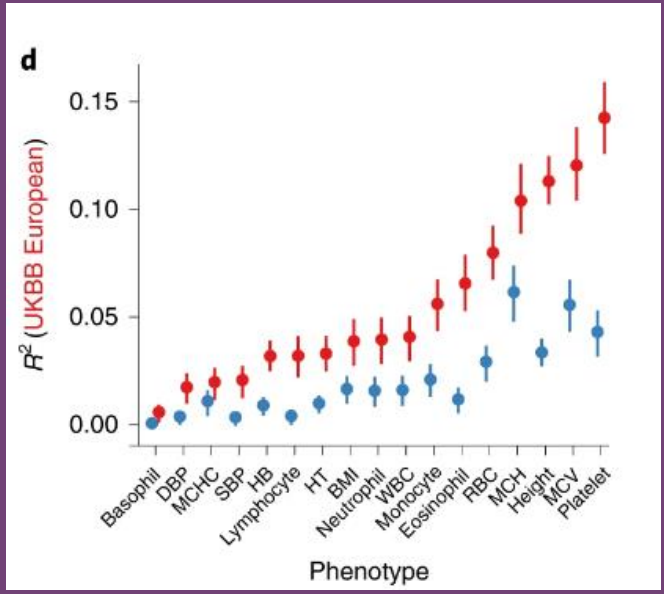
COMPREHENSIVE INHERITED RISK ESTIMATION FOR RISK-BASED BREAST CANCER SCREENING IN WOMEN



- 11,556 breast cancer
- 2,437 pathogenic variants (PVs) carriers, 2.1%; *CHEK2* c.1100delC 1.6%, *CHEK2* c.319+2T>A 0.2%, and *PALB2* c.1592delT 0.3%
- Family History (FH) of breast cancer, parental causes of death, first-degree relatives diagnosed with breast cancer, or an ICD-10 diagnosis for FH.

Category	Before Screening Age	During Screening Age	After Screening Age
Any breast cancer, No.	1,453	7,905	2,198
Invasive breast cancer, No.	1,377	7,145	2,058
In situ breast cancer, No.	74	760	140
Bilateral breast cancer, No.	20	96	36
Age at disease onset, years, median (IQR)	45.9 (42.9-47.9)	59.1 (54.2-64.1)	73.5 (70.1-76.7)
PRS >90% in cases, No. (%)	341 (23.5)	1,663 (21.0)	404 (18.4)
PRS >90% in controls, No. (%)	11,210 (9.9)	9,547 (8.8)	3,652 (8.5)
PV carriers in cases, No. (%)	94 (6.5)	345 (4.4)	73 (3.3)
PV carriers in controls, No. (%)	2,343 (2.0)	1,998 (1.9)	740 (1.7)
Positive FH in cases, No. (%)	107 (7.4)	489 (6.2)	53 (2.4)
Positive FH in controls, No. (%)	3,605 (31)	3,116 (2.9)	865 (2.0)

THE CHALLENGE OF PRS PREDICTION IN NON-EUROPEAN ANCESTRIES



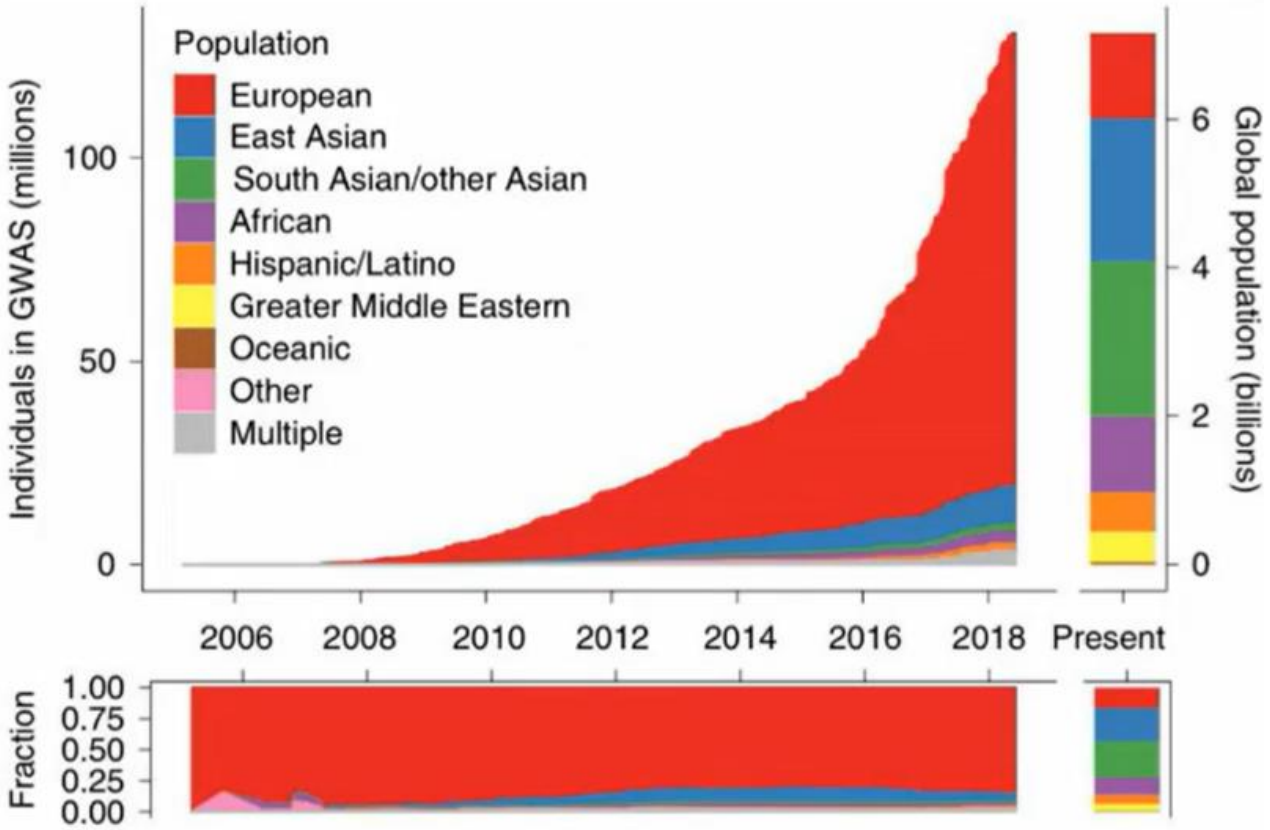


Fig. 1 | Ancestry of GWAS participants over time, as compared with the global population. Cumulative data, as reported by the GWAS catalog[76]. Individuals whose ancestry is 'not reported' are not shown.

Martin AR, Kanai M, et al. Nat Genet. 2019;51:584–591.

REFERENCE