Μαριάνθη Γεωργίτση, PhD

Επίκουρη Καθηγήτρια Γενετικής και Μοριακής βάσης ασθενειών

Εργαστήριο Γονιδιωματικής Ποικιλότητας και Γενετικής Επιδημιολογίας

Τμήμα Μοριακής Βιολογίας και Γενετικής - Σχολή Επιστημών Υγείας

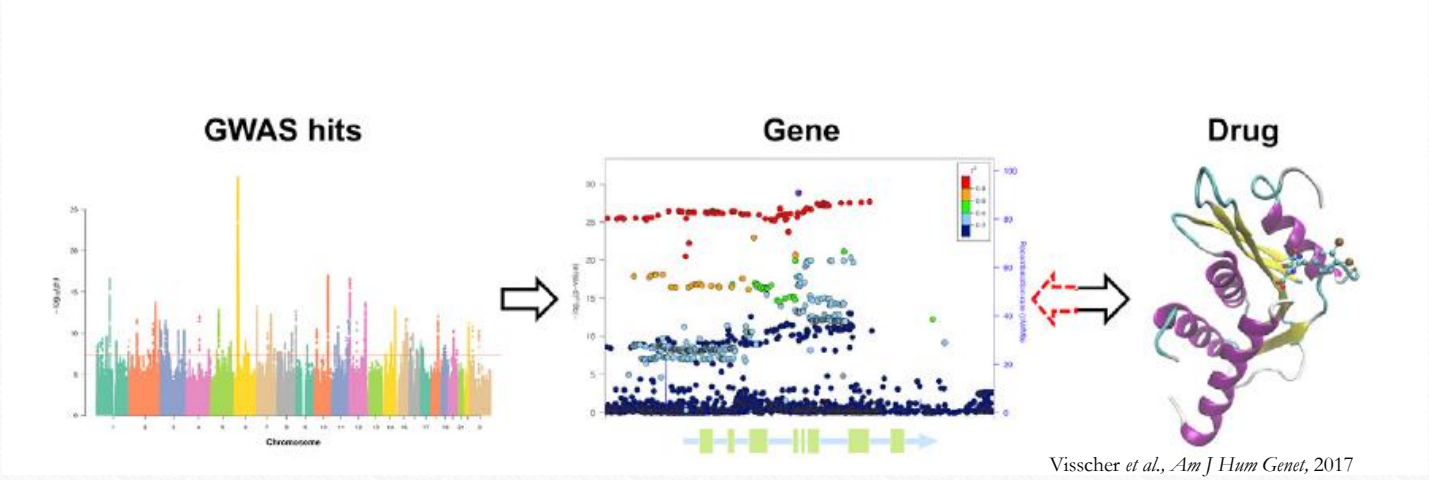
Δημοκρίτειο Πανεπιστήμιο Θράκης

E-mail: mgeorgit@mbg.duth.gr

Prospects and limitations in the use of genetic markers in translational research of complex diseases

«ΜΕΤΑΦΡΑΣΤΙΚΗ ΕΡΕΥΝΑ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ»

Μοριακή Διαγνωστική, Βιοδείκτες και Στοχευμένες Θεραπείες

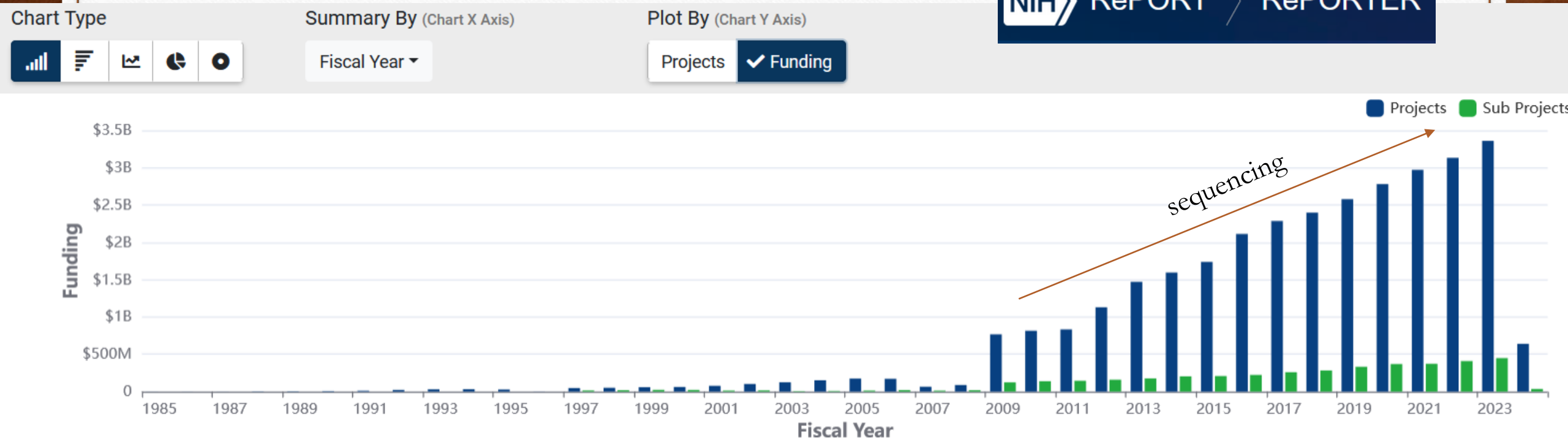


Visscher *et al., Am J Hum Genet,* 2017

Alexandroupoli, 20.04.2024

https://reporter.nih.gov/

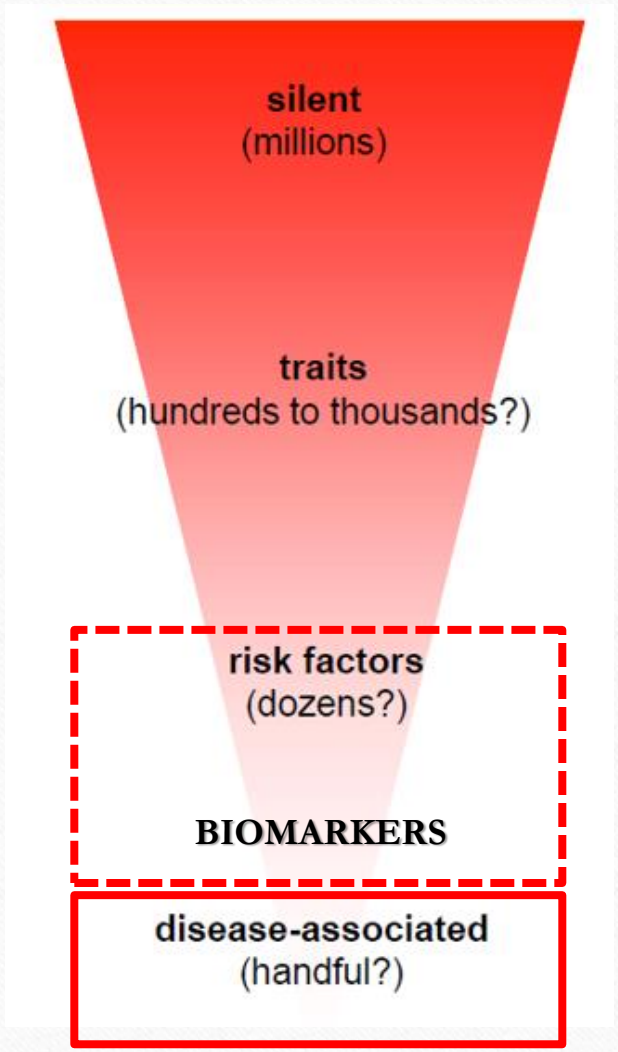
Search terms: "genetic biomarker"
Parameters: (agency) NIH, (summary by) fiscal year, (plot by) funding



Prospects and limitations in the use of genetic markers in translational research of complex diseases

O U T L I N E

- Introduction (basic concepts, tools, methodological approaches)
- Limitations & prospects
- Type 2 Diabetes Mellitus
- Examples of translational success in T2D



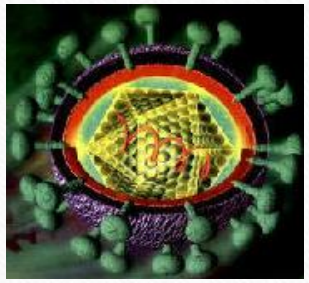
Which are medically relevant?

Genetic Biomarkers

! Genetic risk estimation as the earliest measurable contributor to any disease !

ENVIRONMENT

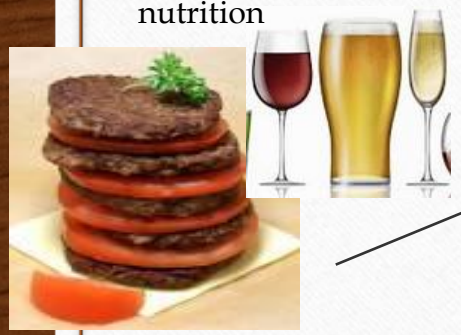
infections



UV

radiation

nutrition

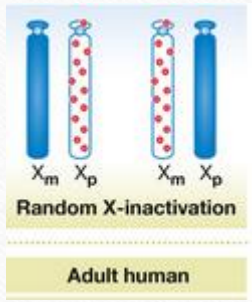


smoking

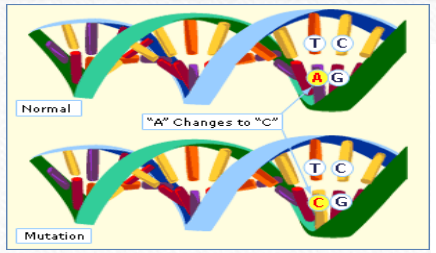
drugs

stress

GENETICS & DNA

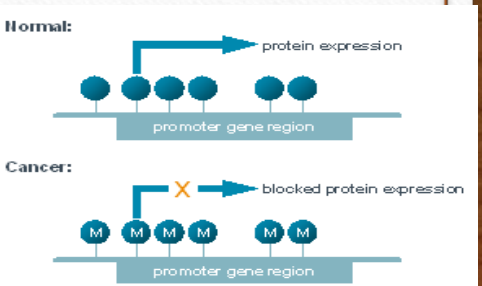


X-chr inactivation (females)

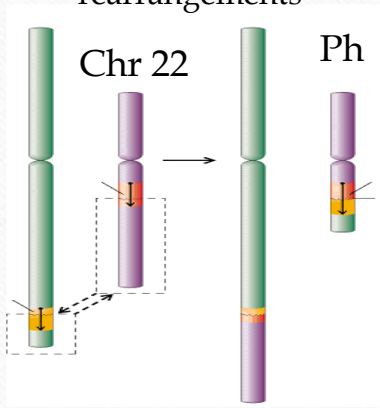


DNA variation

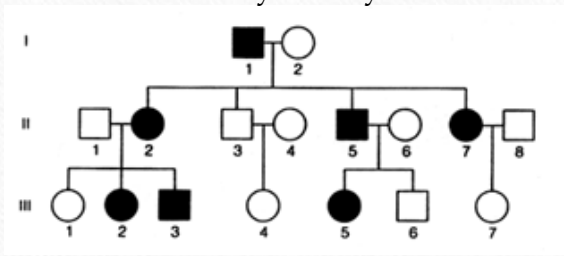
Epigenetics



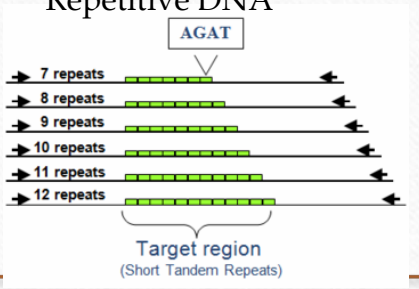
Chromosomal rearrangements

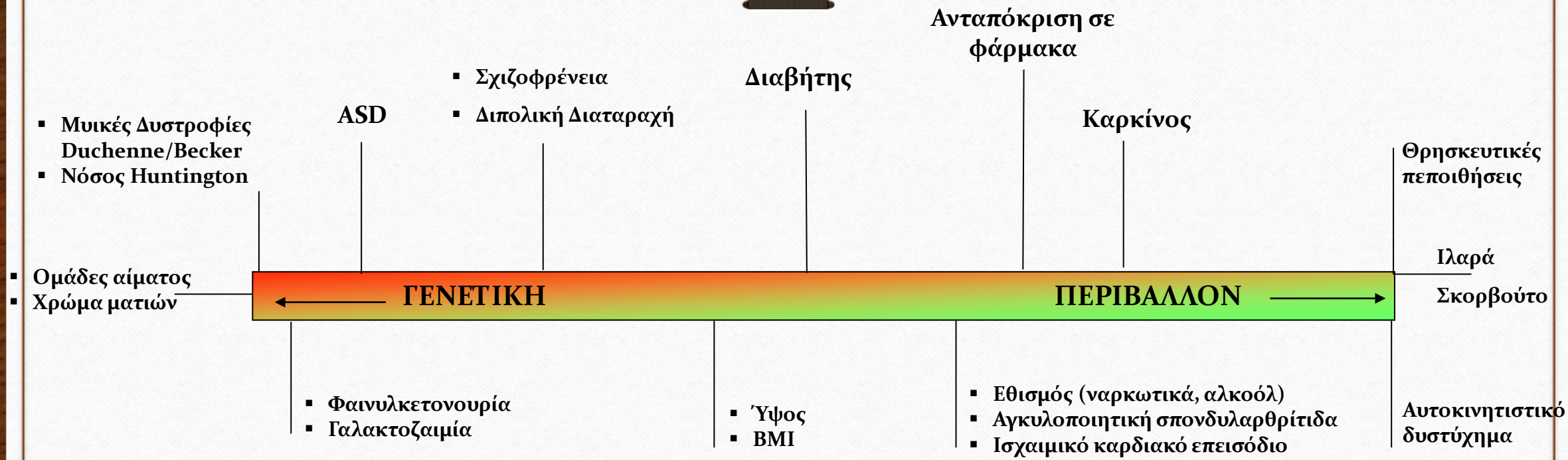


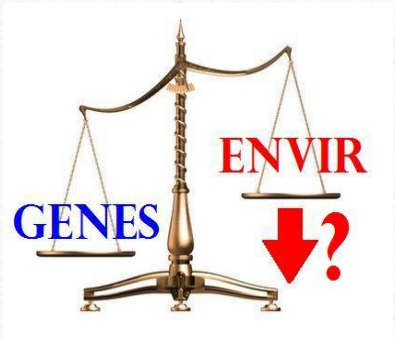
Family history



Repetitive DNA







Rare diseases
Mendelian genetic basis
Mono-/Oligo-genic
"Rare disease-rare variant" hypothesis

Common diseases
Complex genetic basis
Polygenic
"Common disease-common variant" hypothesis

Adapted and modified from Emery's Elements of Medical Genetics, 15th ed., 2017

Single-gene disorders

- Low impact on public health cost
- One or a few gene(s)
- Mendelian inheritance (dominant/recessive)
- Rare variants of large effect → **BIOMARKERS**
- Classical genetics approaches
- Examples:
  - Huntington's disease/Myotonic dystrophies
  - Cystic fibrosis
  - Muscular dystrophies (Duchenne/Becker)
  - Rett Syndrome
  - Fragile X
  - Osteogenesis Imperfecta
  - Hereditary cancer syndromes

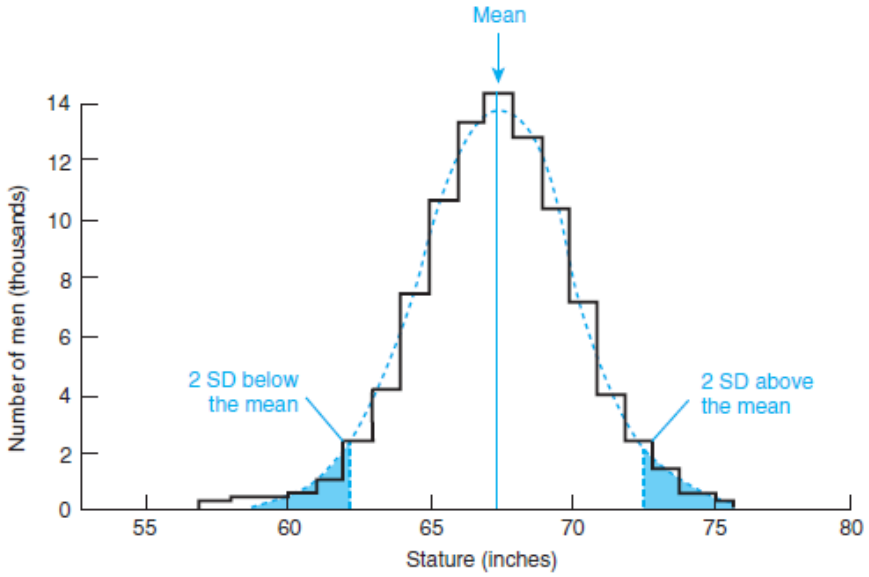
Multifactorial disorders (Complex traits)

- Serious impact on public health cost
- Multiple genes and loci
- Complex pattern of inheritance (additive)
- Variable heritability (h^2)
- Common and rare genetic variants → **LACK OF BIOMARKERS**
- Whole-genome scans → new technologies/analytical tools
- Examples:
  - Stroke/CVD
  - Diabetes (Type 2)
  - Schizophrenia/Bipolar Disorder/OCD
  - Autism Spectrum Disorder (ASD)/ADHD/Language & Learning Disorders
  - Osteoarthritis
  - Alzheimer's/Dementia

Quantitative traits and Quantitative Trait Loci (QTLs)

"All-or-none" vs "Shades of grey"

Figure 7.3 The inheritance of height. Genetics students at the University of Notre Dame lined up by height in inches, revealing the continuously varying nature of height. *David Hyde/Wayne Falda/McGraw-Hill Education*



Genetics in Medicine, Thompson & Thompson, 7th ed., 2007

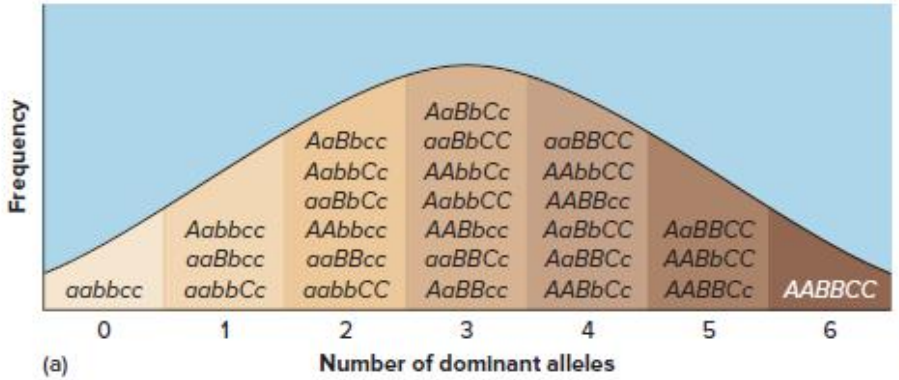
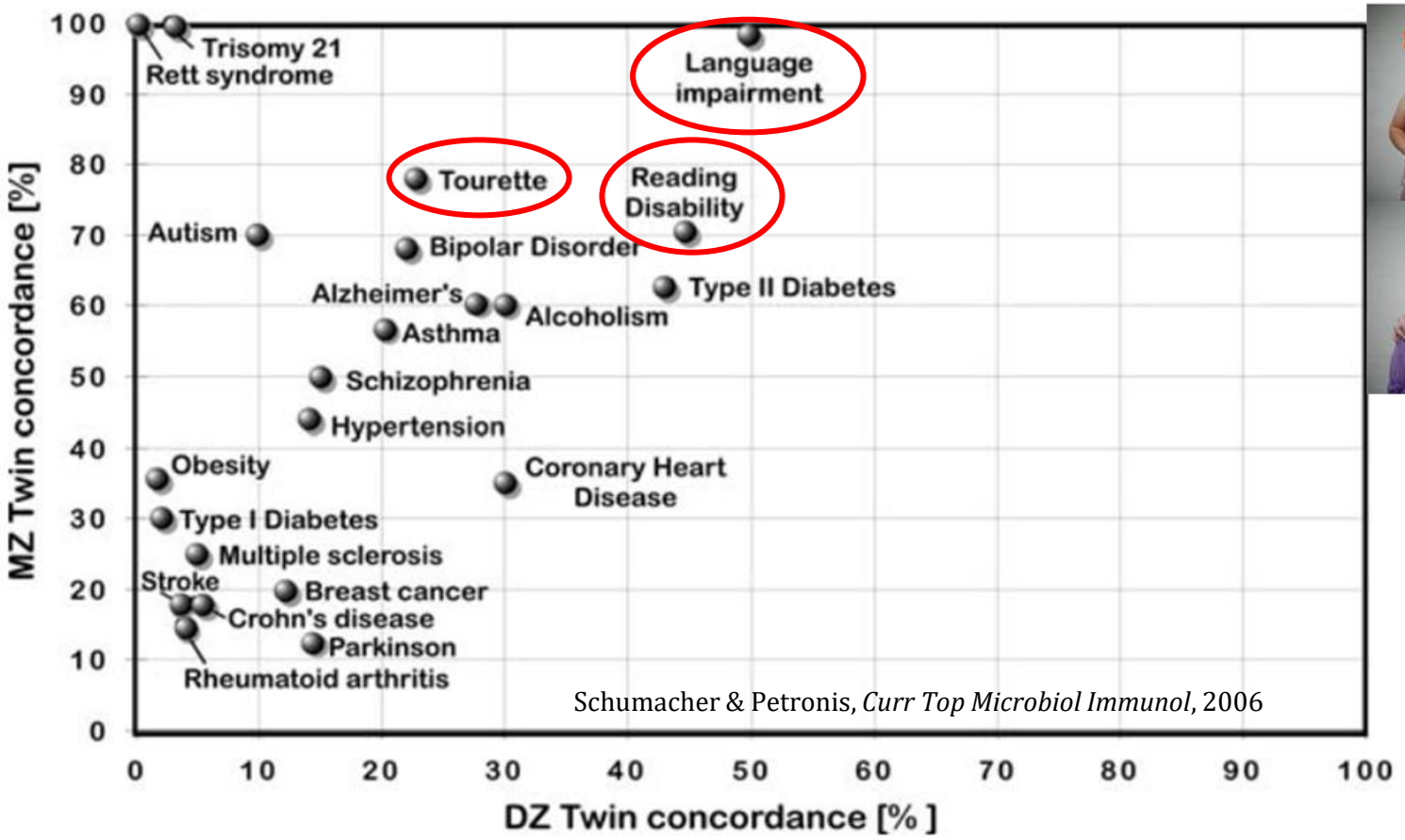


Figure 7.4 Variations in skin color. (a) A model of three genes, with two alleles each, can explain broad hues of human skin. In actuality, this trait likely involves many more than three genes. **(b)** Humans come in a great variety of skin colors. Skin color genes can assort in interesting ways. These beautiful young ladies, Marcia and Millie, are twins! Their father is Jamaican with dark skin and tight dark curls and their mother is European with fair skin and golden-brown hair. *(b): SWNS/South West News Service Ltd.*

Human Genetics: Concepts and Applications, McGraw Hill, 13th ed., 2021

Complex trait heritability



Schumacher & Petronis, *Curr Top Microbiol Immunol*, 2006

Fig. 1 Concordance of MZ and DZ twins for different disorders. As a rule, the degree of concordance in MZ twins is lower than 100% for nearly all complex diseases but substantially higher in comparison to the concordance rate in DZ twins

Photos by Jason Reed/Reuters

Human Genetics: Concepts and Applications, McGraw Hill, 13th ed., 2021

$$h^2 = \frac{\text{Variance in DZ pairs} - \text{Variance in MZ pairs}}{\text{Variance in DZ pairs}}$$

Heritability

(Κληρονομησιμότητα ή Κληρονομική Ικανότητα)

$$P = G + E$$

$$V_P = V_G + V_E + V_{GxE}$$

$$V_P = V_A + V_D + V_I + V_E$$

Phenotypic Variance | Additive Genetic Variance | Non-Additive Genetic Variance | Environmental Variance

$$H^2 = V_G/V_P$$

Broad-sense heritability

$$h^2 = \frac{V_A}{V_P}$$

Narrow-sense heritability

A heritability close to 1 indicates a large portion of the phenotypic variation is due to genetic factors

$$V_G = V_A + V_D + V_I$$

- The total *genetic* variance for a character (V_G) is a function of:
- **Additive genetic variance** (V_A) – variation due to the additive effects of alleles
- **Dominance genetic variation** (V_D) – variation due to dominance relationships among alleles
- **Epistatic genetic variation** (V_I) – variation due to interactions among loci

Complex trait heritability

$$h^2 = \frac{V_A}{V_P}$$

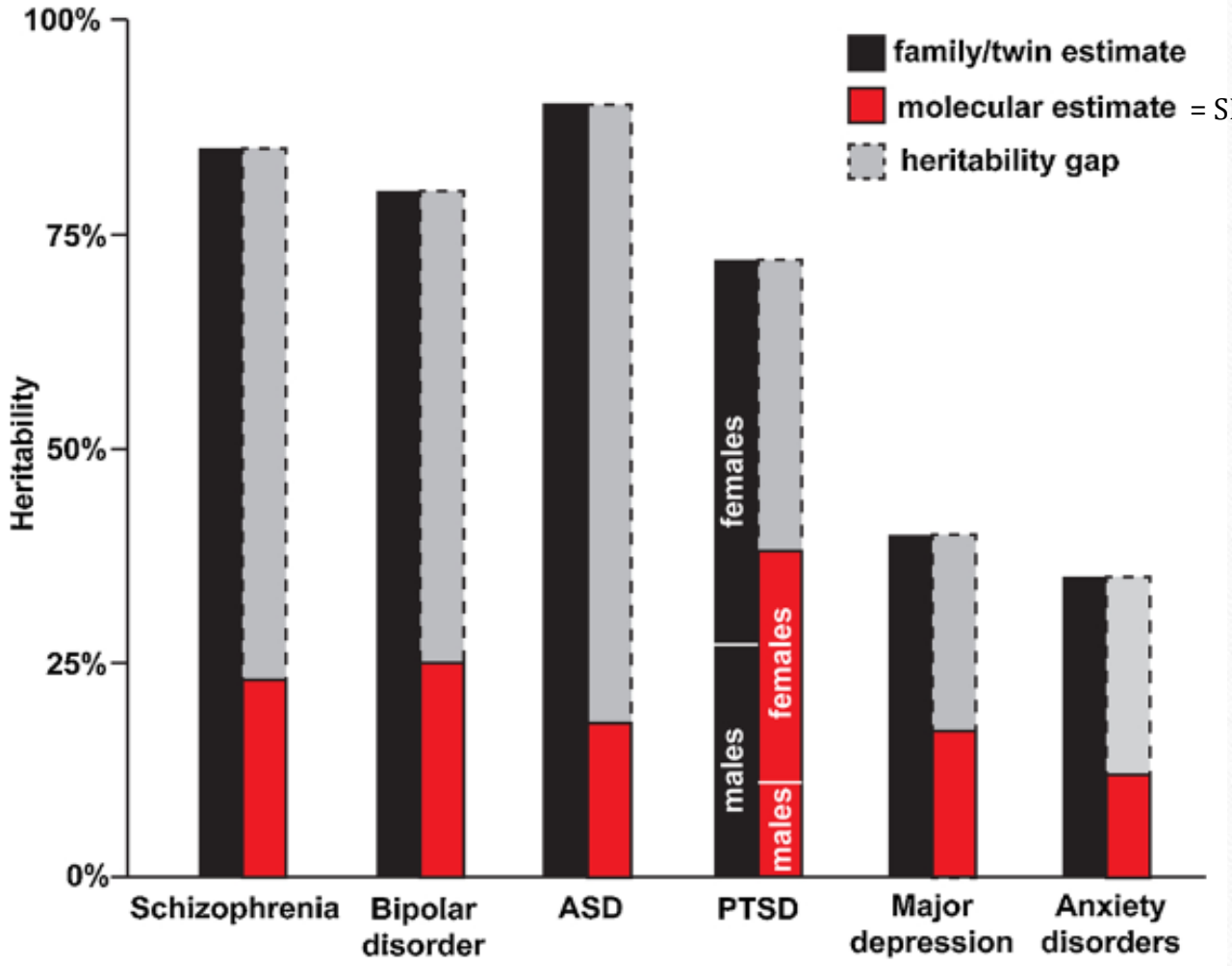
Narrow-sense heritability

Table 7.2 Heritabilities for Some Human Traits

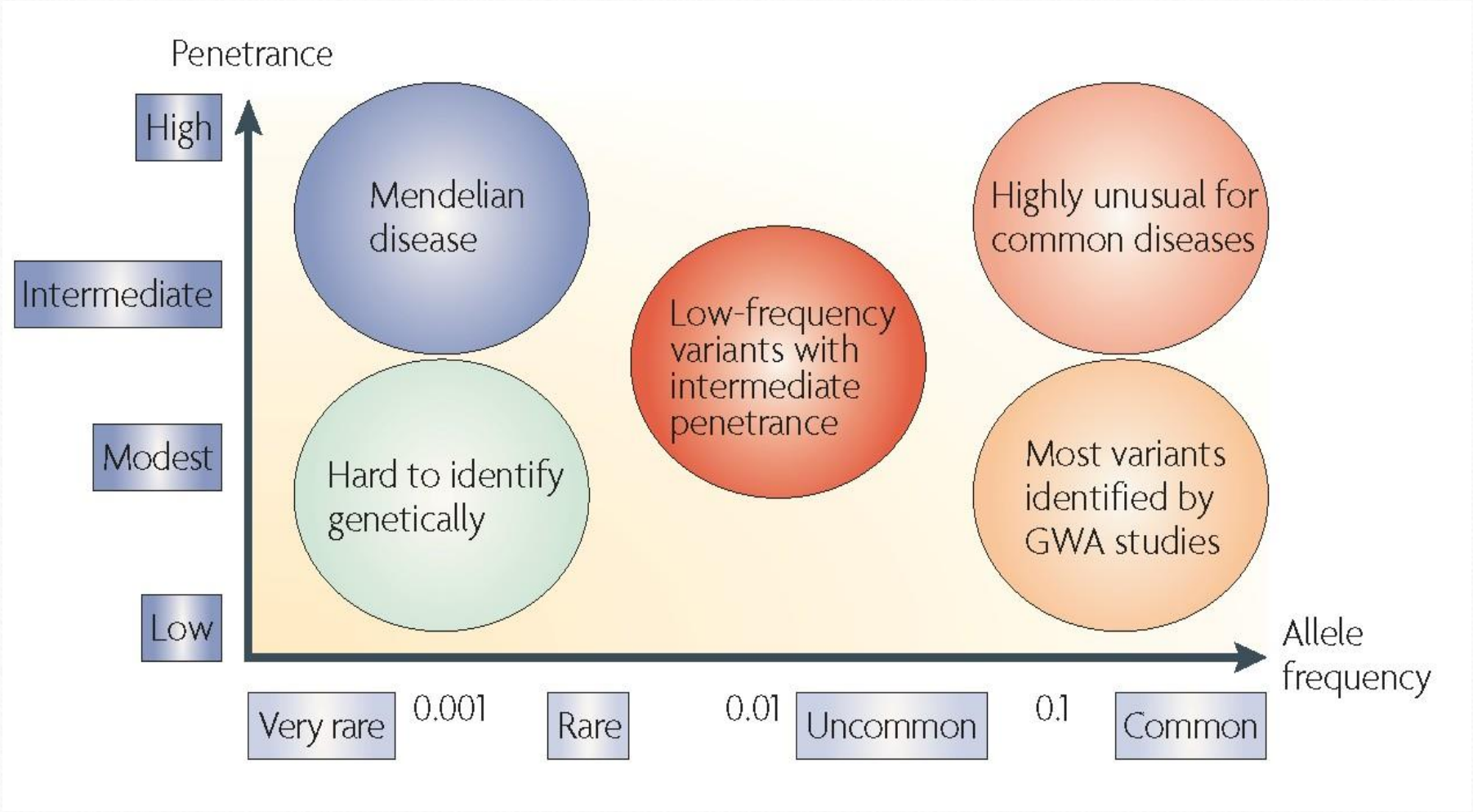
Trait	Heritability
Clubfoot	0.8
Height	0.8
Blood pressure	0.6
Body mass index	0.4–0.7
Verbal aptitude	0.7
Mathematical aptitude	0.3
Spelling aptitude	0.5
Total fingerprint ridge count	0.9
Intelligence	0.5–0.8
Total serum cholesterol	0.6

Human Genetics: Concepts and Applications, McGraw Hill, 13th ed., 2021

D. van Calker and T. Serchov *Neuroscience and Biobehavioral Reviews 126 (2021) 23–42*

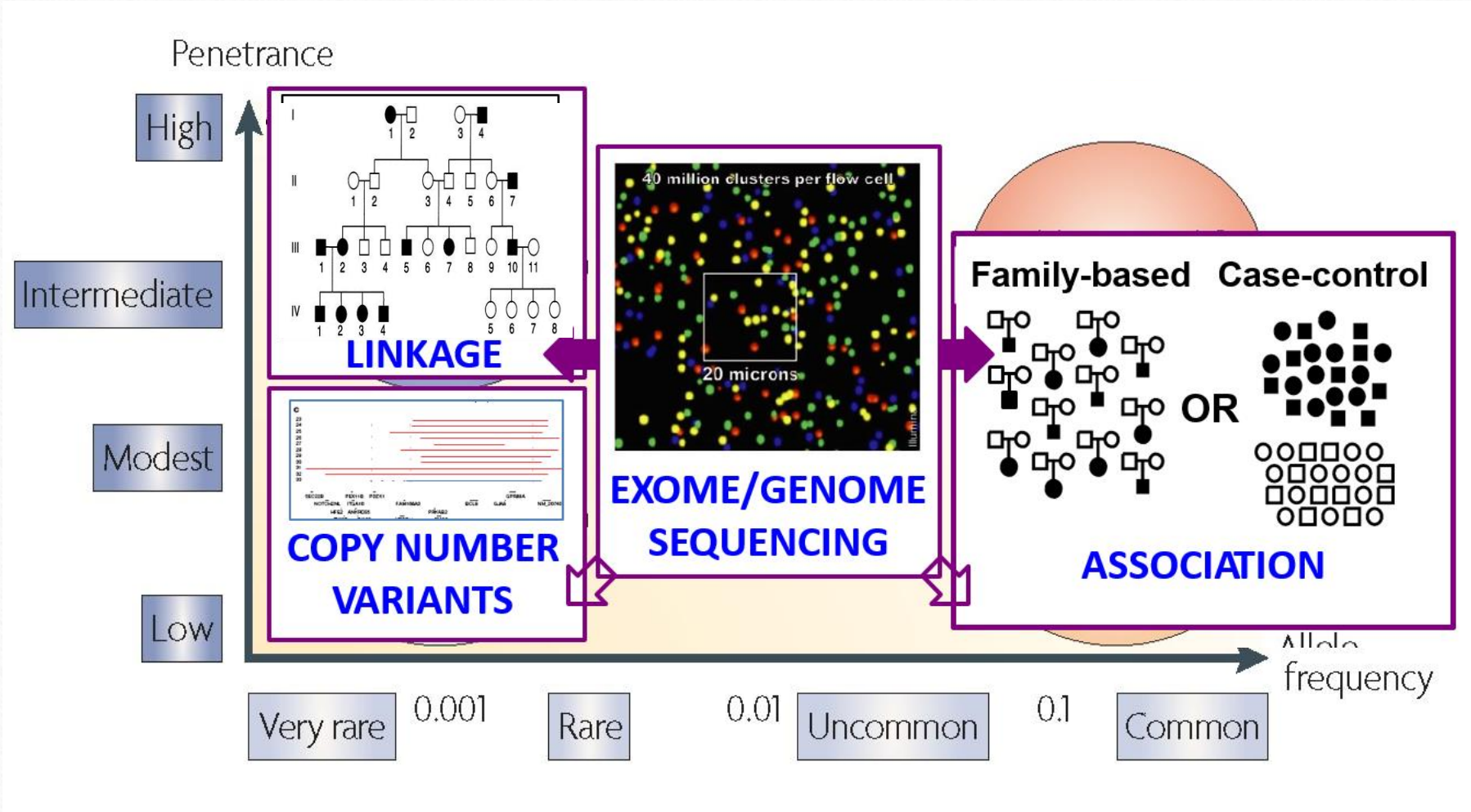


The landscape of human genome variation



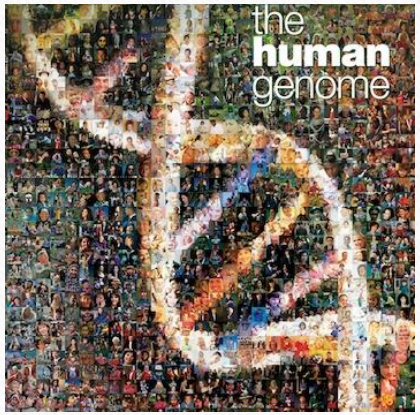
McCarthy *et al.*, *Nat Rev Genet*, 2008

The landscape of human genome variation



McCarthy *et al.*, *Nat Rev Genet*, 2008

Bush & Moore, *PLoS Comput Biol*, 2012



https://www.genome.gov/10001772

NOTE: The first genetic association studies focused on **candidate gene analysis** and, therefore, were not suited for novel genetic risk loci identification.

The completion of the **Human Genome Project**, the **HapMap Project** and **1000Genomes Project**, along with many technological and conceptual advances, have paved the way to array (chip)-based GWAS and Next-Generation Sequencing

→ **High-throughput targeted genotyping and NGS**

https://www.genome.gov/10001688/international-hapmap-project

GENOME 10K

Autism Genome 10K

Genomics england

The 100,000 genomes project

THE PRECISION MEDICINE INITIATIVE

biobank uk

Enabling scientific discoveries that improve human health

500k UK people data

1000 Genomes

A Deep Catalog of Human Genetic Variation

https://www.internationalgenome.org/

Genome-wide association studies (GWAS)

Patients

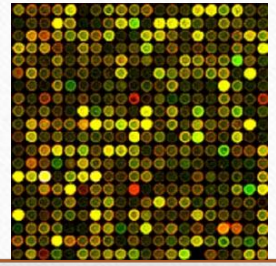
Controls

Thousands of samples
– The more the better !

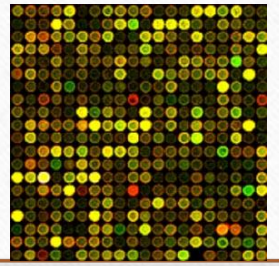
- ✓ Better phenotyping of cases – Inclusion/Exclusion criteria
- ✓ More careful selection of controls
- ✓ Investigators joining forces – Data sharing
- ✓ Funding of large-scale projects (multicentered – multiethnic)
- ✓ International data repositories – Free access

- ✓ Catalogued SNPs (Minor Allele Frequencies – MAFs) and tagSNPs
- ✓ Patterns of linkage disequilibrium (LD) per population
- ✓ Population-specific variation (AIMs)
- ✓ Frameworks to analyze enormous datasets – Development of bioinformatics tools
- ✓ Pathway analyses
- ✓ Meta-analyses

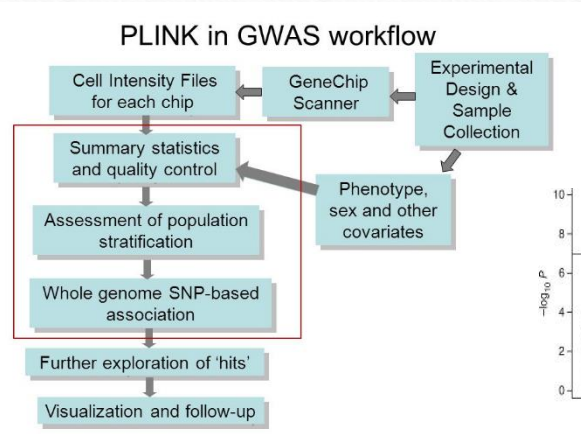
Genotyping of eg 100k-6M SNPs

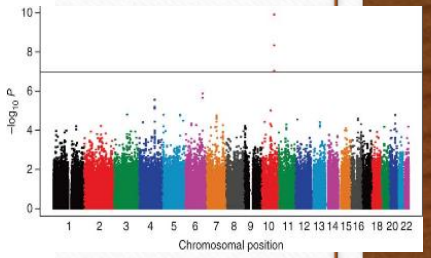


Comparison of alleles
–
Statistical analysis



Workflow of a typical GWAS study



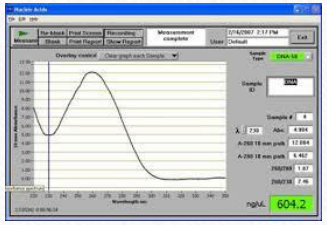


Sample collection

Peripheral blood
Buccal cells
…

DNA extraction and Quantification

Manual purification
Kit-based purification



Fluorospectrophotometry
Spectrophotometry

DNA Plating

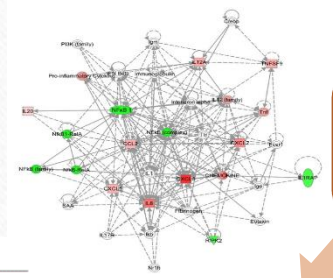
96-well plates
384-well plates

Array run

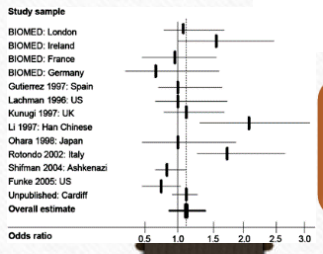
QC

Statistical analysis (trait-dependent)

- ***ANOVA***
- ***Chi-square***
- **Covariate adjustment**
- **Multiple testing correction**
- **Population stratification**



Pathway analyses



Meta-analyses

Replication study (stage 2)

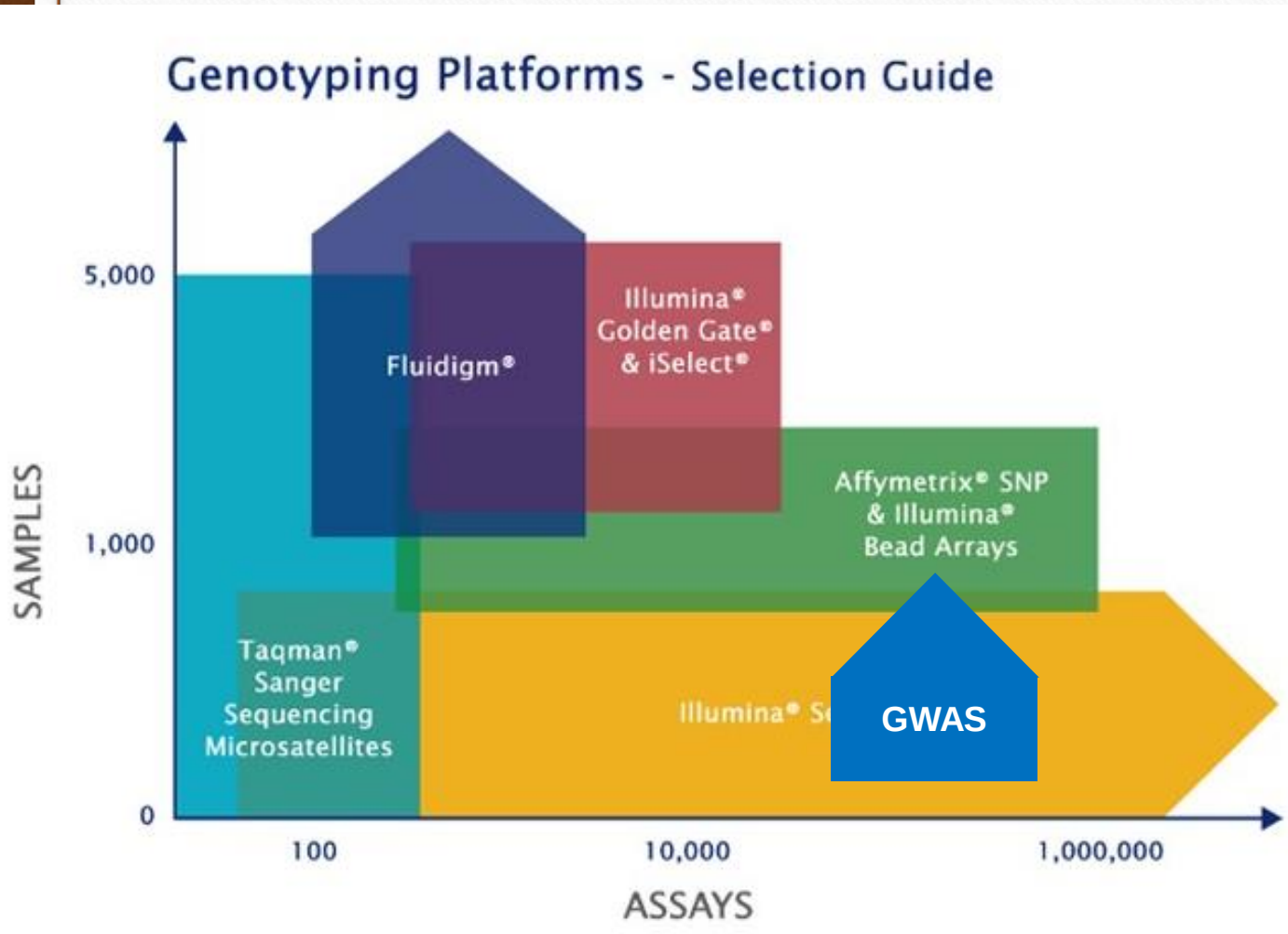
- **Sample size (large)**
- **Population**
- **Phenotypic criteria**

Validation (technical, functional)

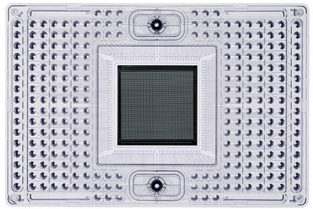
- **Other genotyping platform**
- ***In silico* analysis**
- **Sequencing**
- **Gene expression analysis**
- **Cell-based analysis**
- **Animal models**

Προς την ανάπτυξη τεχνικών γονιδιωματικής κλίμακας

1. Κόστος
2. Στόχος
3. Εργαλεία ανάλυσης



http://www.lifesciences.sourcebioscience.com/media/426305/genotyping%20image.png



Juno 96.96 Genotyping IF

WES
WGS

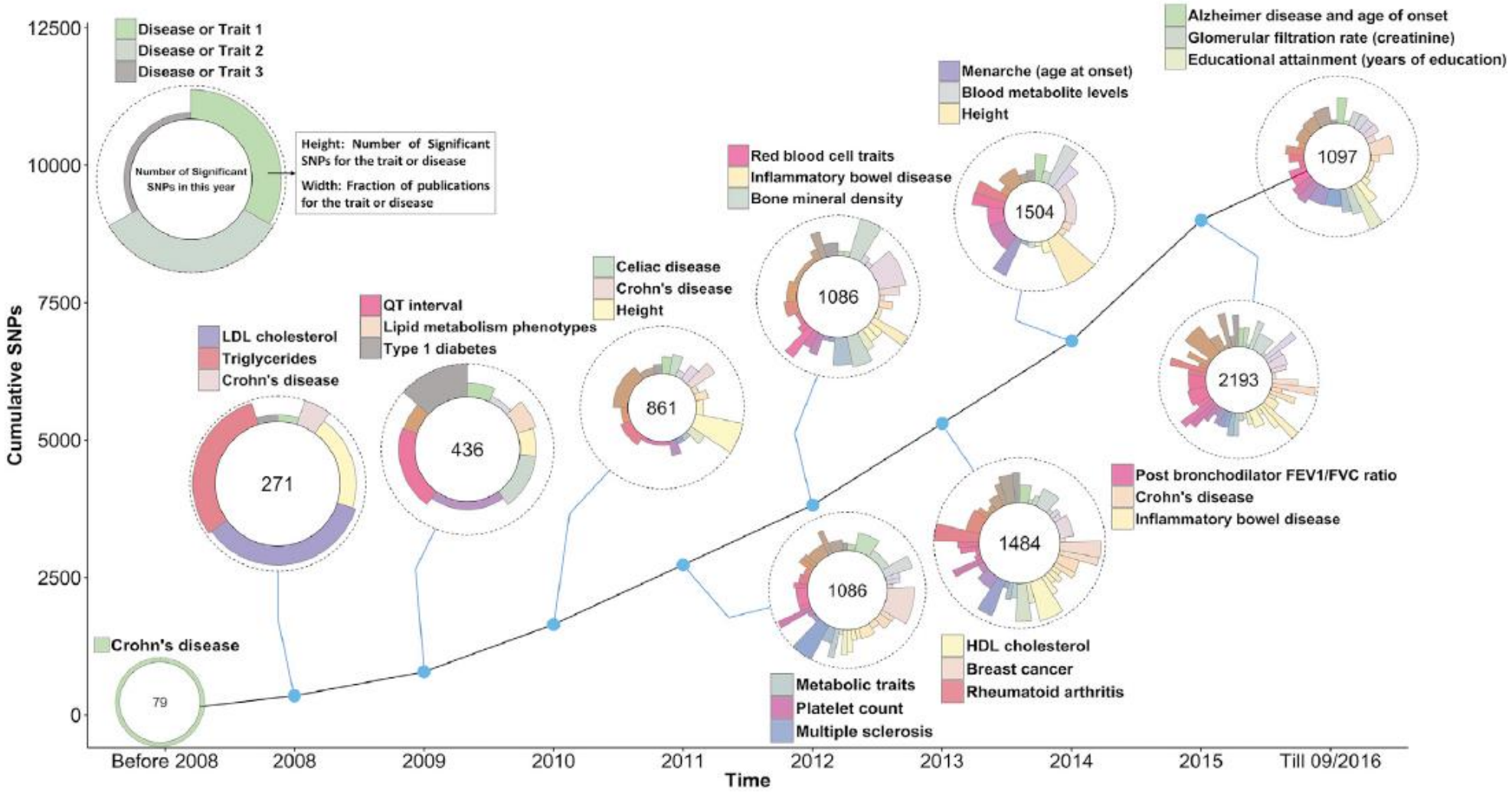


Figure 2. GWAS SNP-Trait Discovery Timeline

Data used for generating the graph were taken from the GWAS Catalogue.[10] SNPs and traits were selected according to the following filters. SNPs were selected with a p value $< 5 \times 10^{-8}$. For each trait with two or more selected SNPs, SNPs were removed if they had an LD $r^2 > 0.5$ (calculated from 1000 Genomes phase 3 data) with another selected SNPs and their p value was larger. For each year of discovery, only the top three traits and diseases with the largest number of SNPs are labeled in the circle.

Visscher *et al., Am J Hum Genet,* 2017

Genome-wide association studies (2005 – present day)

- ✓ Καλύτερος φαινοτυπικός χαρακτηρισμός των περιστατικών – Κριτήρια ένταξης/Αποκλεισμού
- ✓ Προσεκτικότερη επιλογή μαρτύρων
- ✓ Σύμπραξη ερευνητών – Διαμοιρασμός δεδομένων
- ✓ Χρηματοδότηση ερευνητικών έργων μεγάλης κλίμακας (πολυκεντρικές – πολυεθνικές μελέτες)
- ✓ Θέσπιση και λειτουργία διεθνών αποθετηρίων δεδομένων – Ελεύθερη (?) πρόσβαση ερευνητών

- **Συλλογή δειγμάτων**
- **Θέσπιση και δραστηριότητα ερευνητικών κοινοπραξιών**

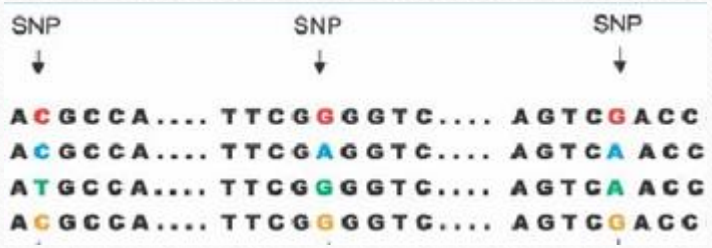
- ✓ Καταλογογράφηση SNPs
- ✓ Πρότυπα ανισορροπίας σύνδεσης (Linkage Disequilibrium – LD)
- ✓ Ποικιλομορφία γονιδιώματος σε πληθυσμιακό επίπεδο (Ancestry-Informative Markers – AIMs)
- ✓ Πλαίσια (pipelines) ανάλυσης δεδομένων μεγάλου όγκου – Ανάπτυξη εργαλείων βιοπληροφορικής ανάλυσης
- ✓ Αναλύσεις μονοπατιών
- ✓ Μεταναλύσεις - PRSs

- **Γενετική αρχιτεκτονική**
- **Αναλύσεις**
- **Βιοπληροφορική**

Γενετική Αρχιτεκτονική

Πληθυσμός

Γενετικές παραλλαγές



Είδος (ποιες)

Αριθμός (πόσες)

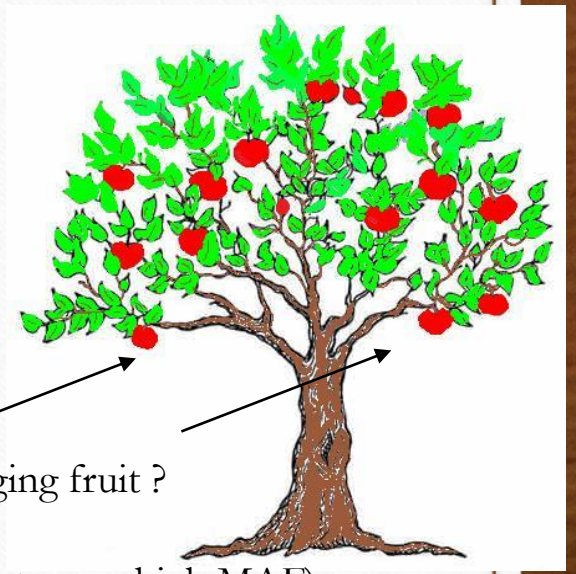
Συχνότητα
(πόσο συχνές)

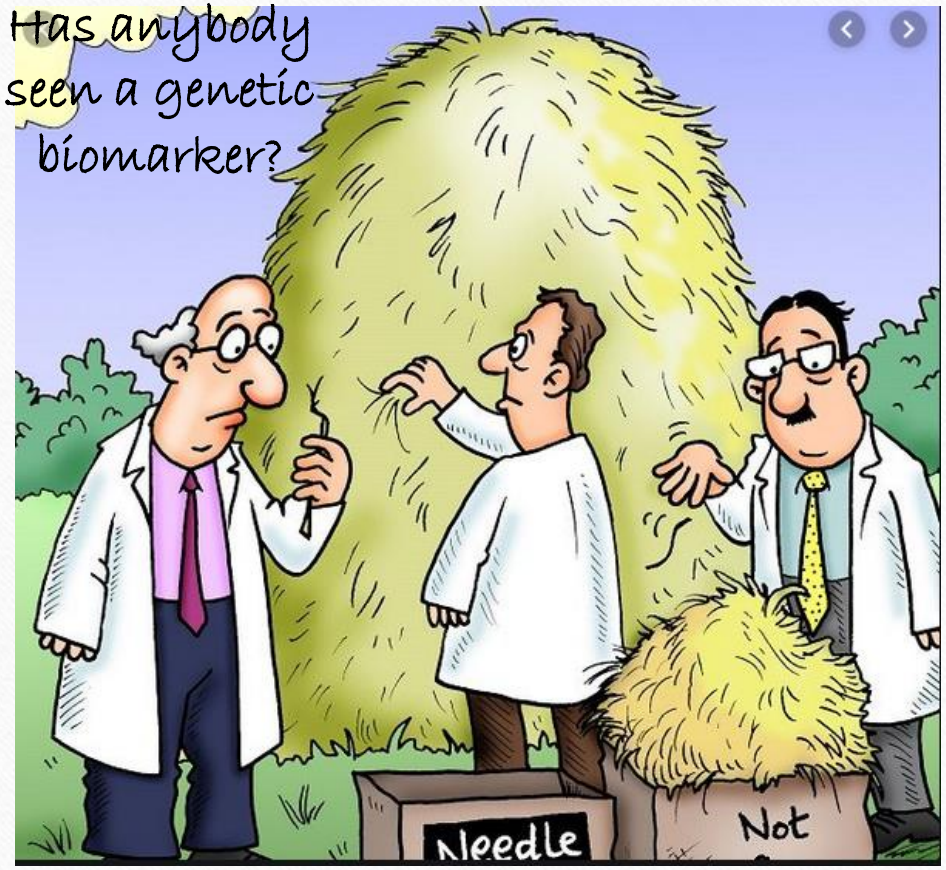
Αποτέλεσμα
(πόσο σημαντικές)

Linkage
Disequilibrium
(με ποιες άλλες)

GWAS limitations

- Only associations revealed – Not causation (Positive or negative correlation may imply underlying mechanisms, but is not proof that A actually causes B)
- Sex chromosomes (X, Y) omitted from analysis (until recently)
- Patient selection (population-based vs selected group) and controls selection (are they truly free or any disease? → genetic correlation between disorders/cross-disorder analyses)
- Can results from one ancestry (eg Caucasians) be extrapolated to another (eg Asians)?
- Distinguish true cases from phenocopies (=a trait or illness that resembles an inherited one but is attributed to an environmental cause)
- Unable to uncover tissue-specific *gene x gene* interactions





We should not expect common variants to have large effects, for evolutionary reasons.

Large effects are bad, as most variants with large effects are consequently selected against, thus never become fixed and common.

Missing heritability

(ελλείπουσα κληρονομησιμότητα)

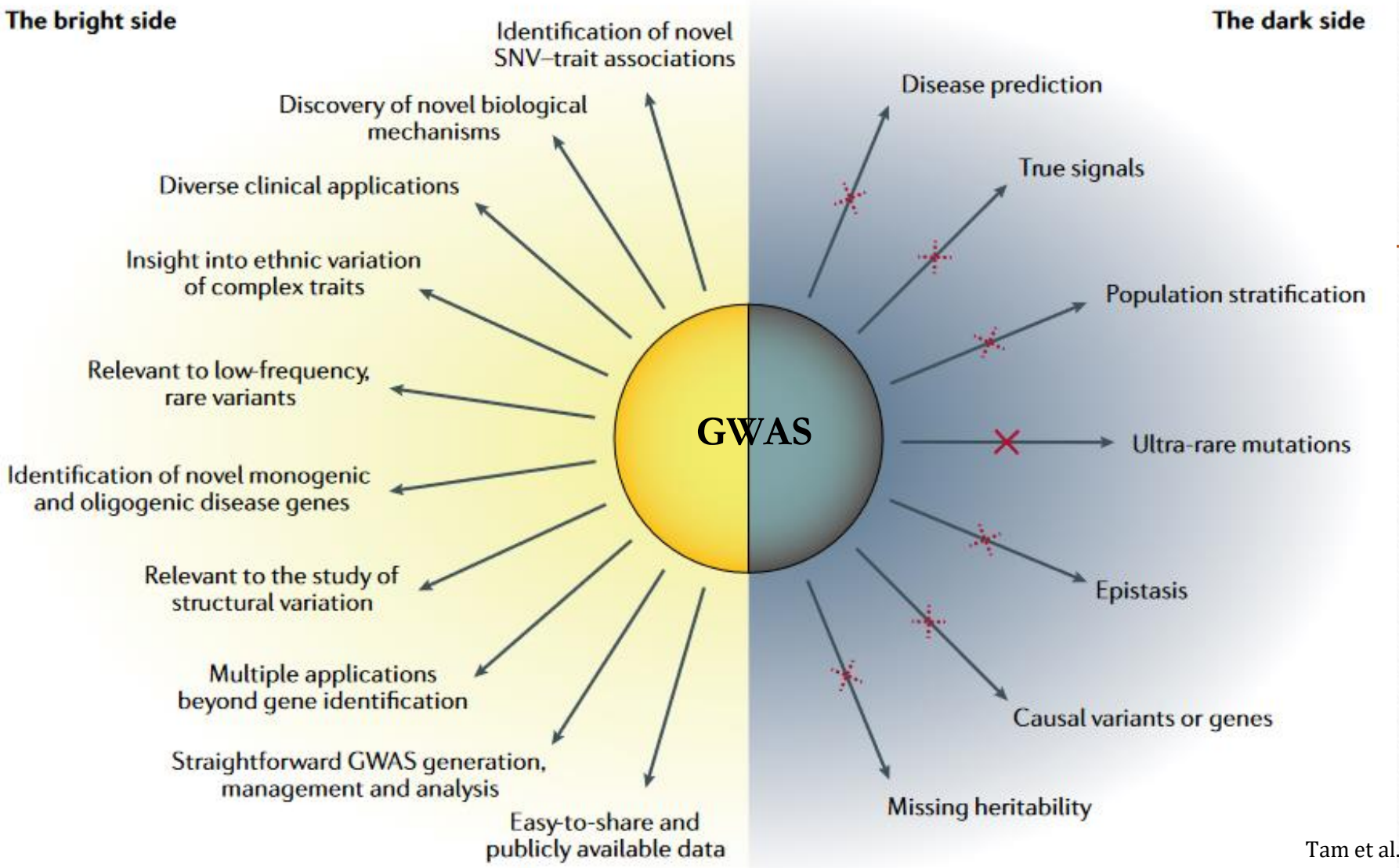
Rare variants (MAF<1%)

- Structural variants (CNVs)
- *Gene x gene* interactions (epistasis)
- Epigenetics
- *Gene x environment* interactions
- Something unknown?

There could be scarier and more intractable reasons for unaccounted-for heritability that are not even being discussed. "It's a possibility that there's something we just don't fundamentally understand," Kruglyak says. "That it's so different from what we're thinking about that we're not thinking about it yet."

Still the mystery continues to draw its sleuths, for Kruglyak as for many other basic-research scientists. "You have this clear, tangible phenomenon in which children resemble their parents," he says. "Despite what students get told in elementary-school science, we just don't know how that works." ■

https://www.nature.com/news/2008/081105/pdf/456018a.pdf - freely available

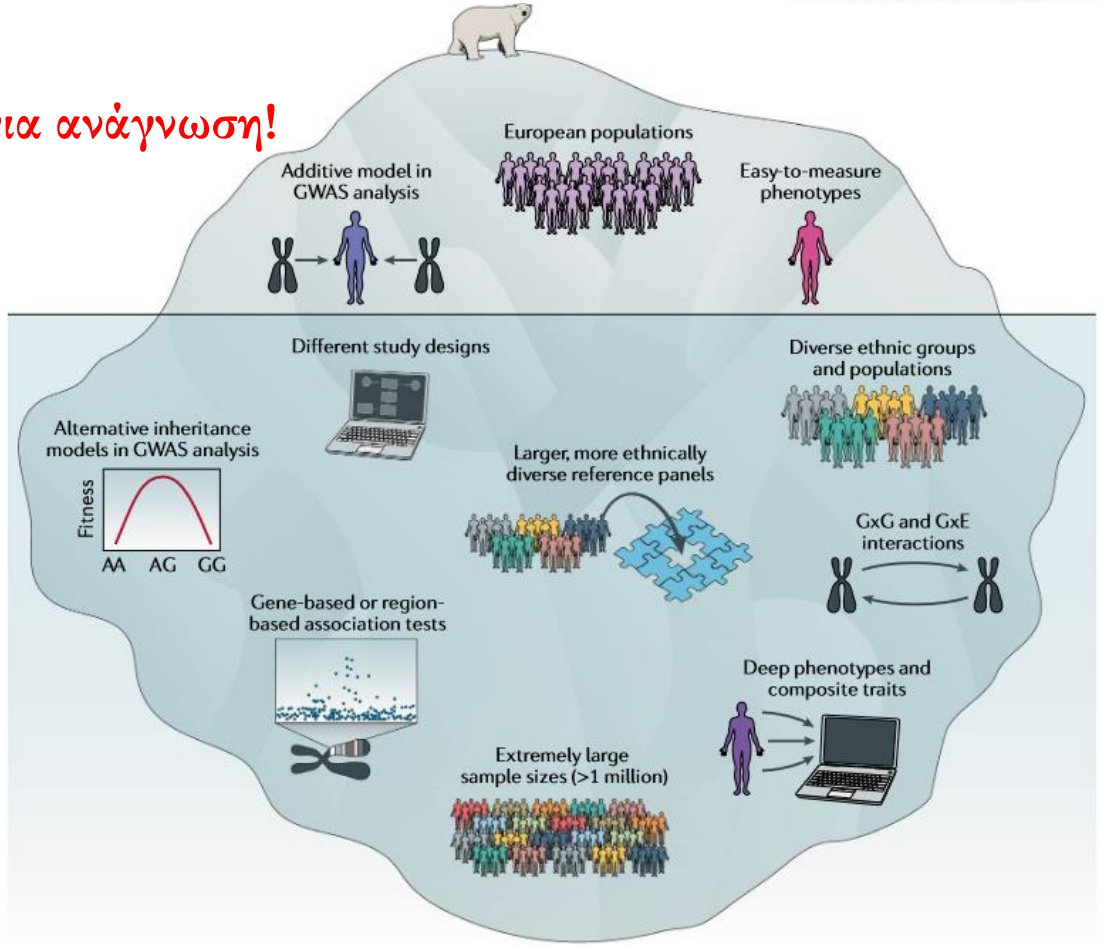


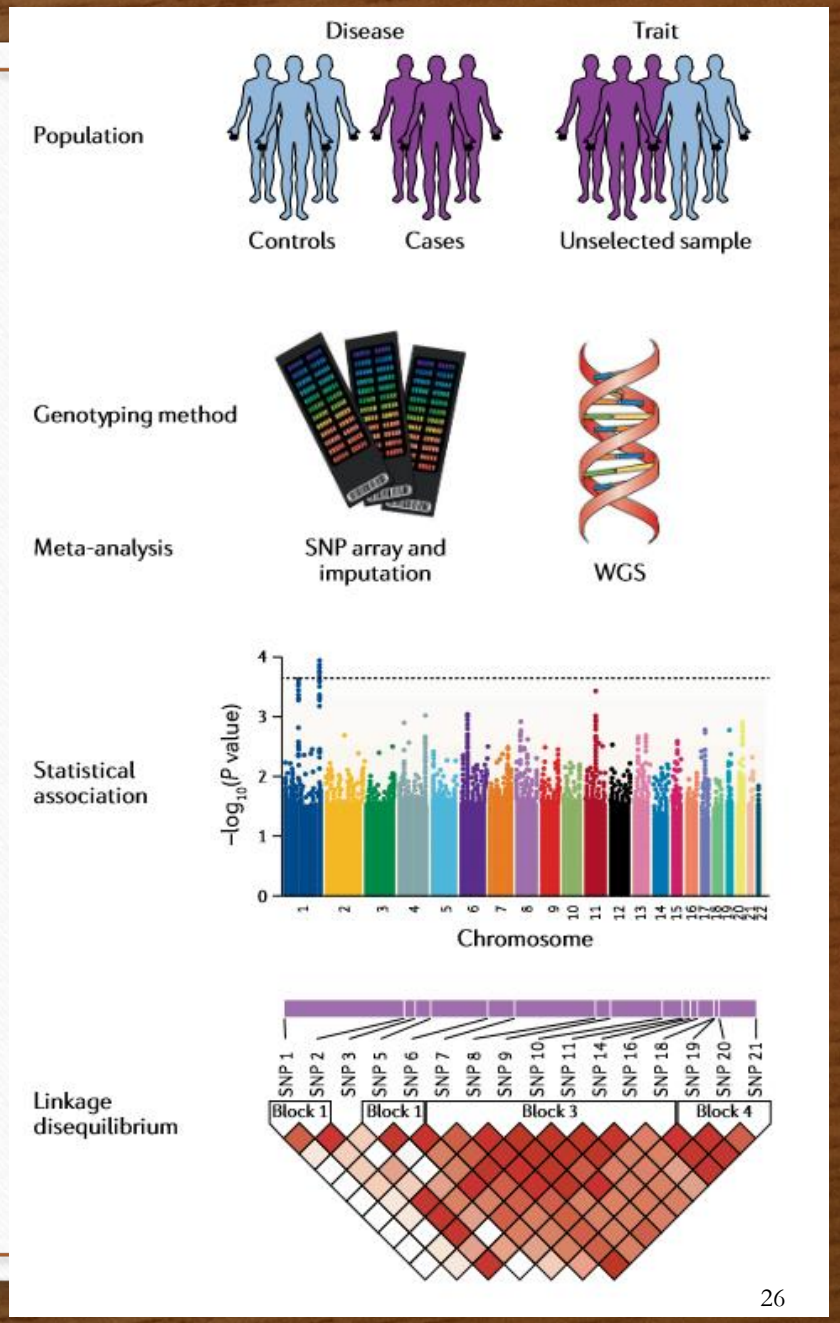
Tam et al., *Nat Rev Genet*, 2019

Benefits and limitations of genome-wide association studies

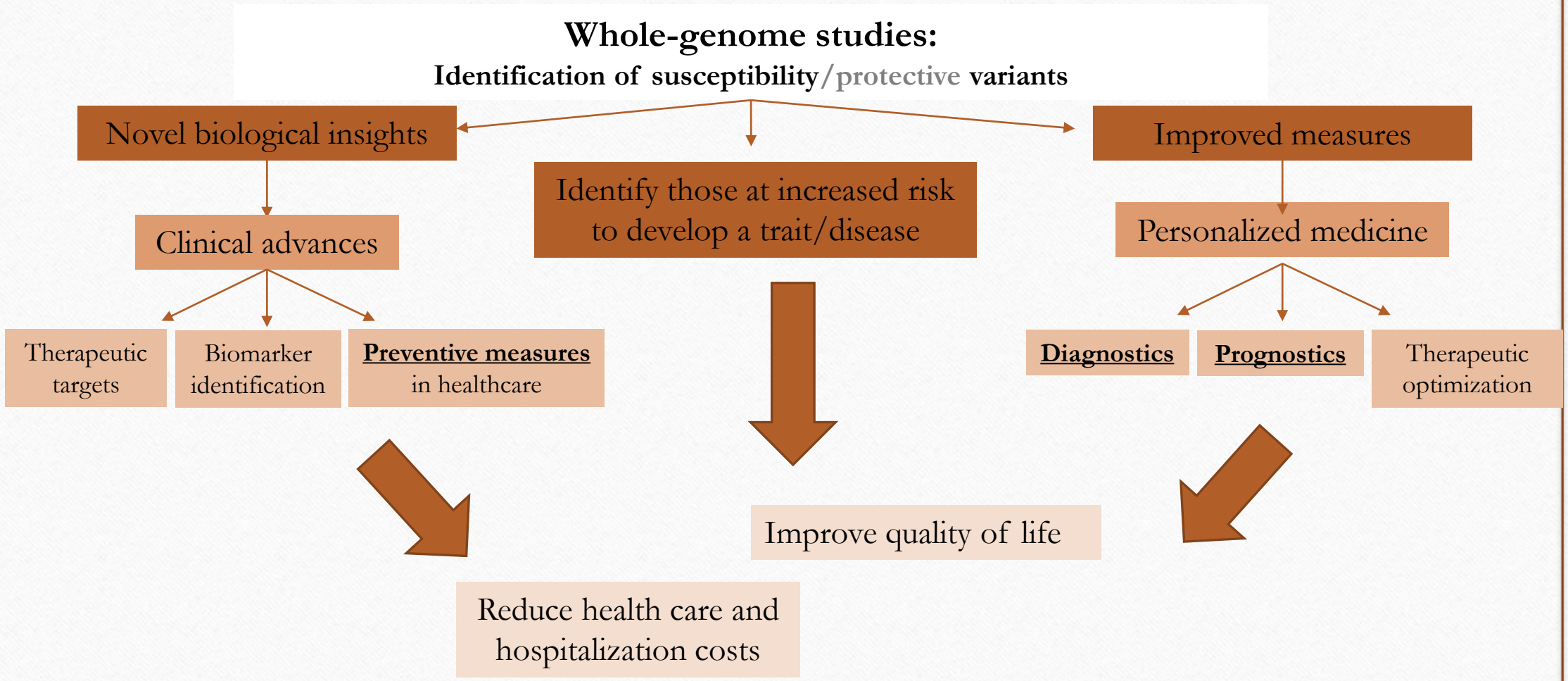
*Vivian Tam[1], Nikunj Patel[1], Michelle Turcotte[1], Yohan Bossé[2,3], Guillaume Paré[1,4] and David Meyre[1,4,5]**

Ισχυρή σύσταση για ανάγνωση!





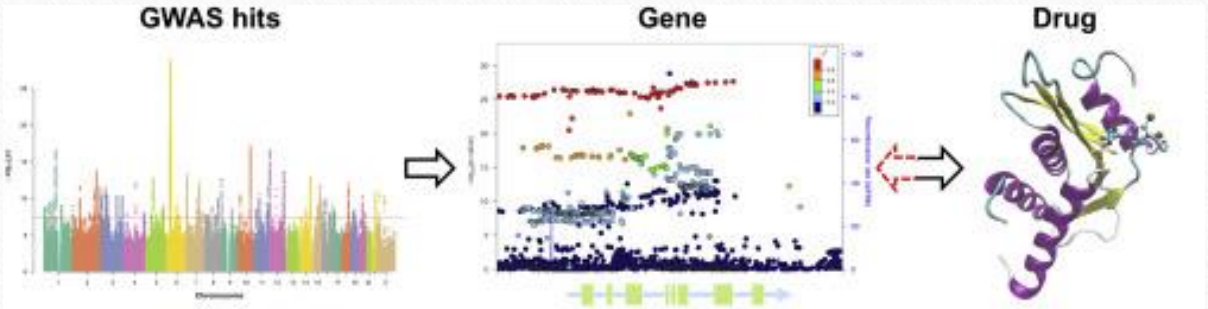
What whole-genome studies have to offer – Translational research



Adapted from McCarthy *et al.*, *Nat Rev Genet*, 9:357-369, 2008

Examples of links between GWAS discoveries and drugs:

Genetically-informed translational research/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

Identifying genetic markers with translational potential is crucial for medical **fields where:**

a) Diagnostic criteria are not based on biological markers, such as psychiatry and behavioural traits, and

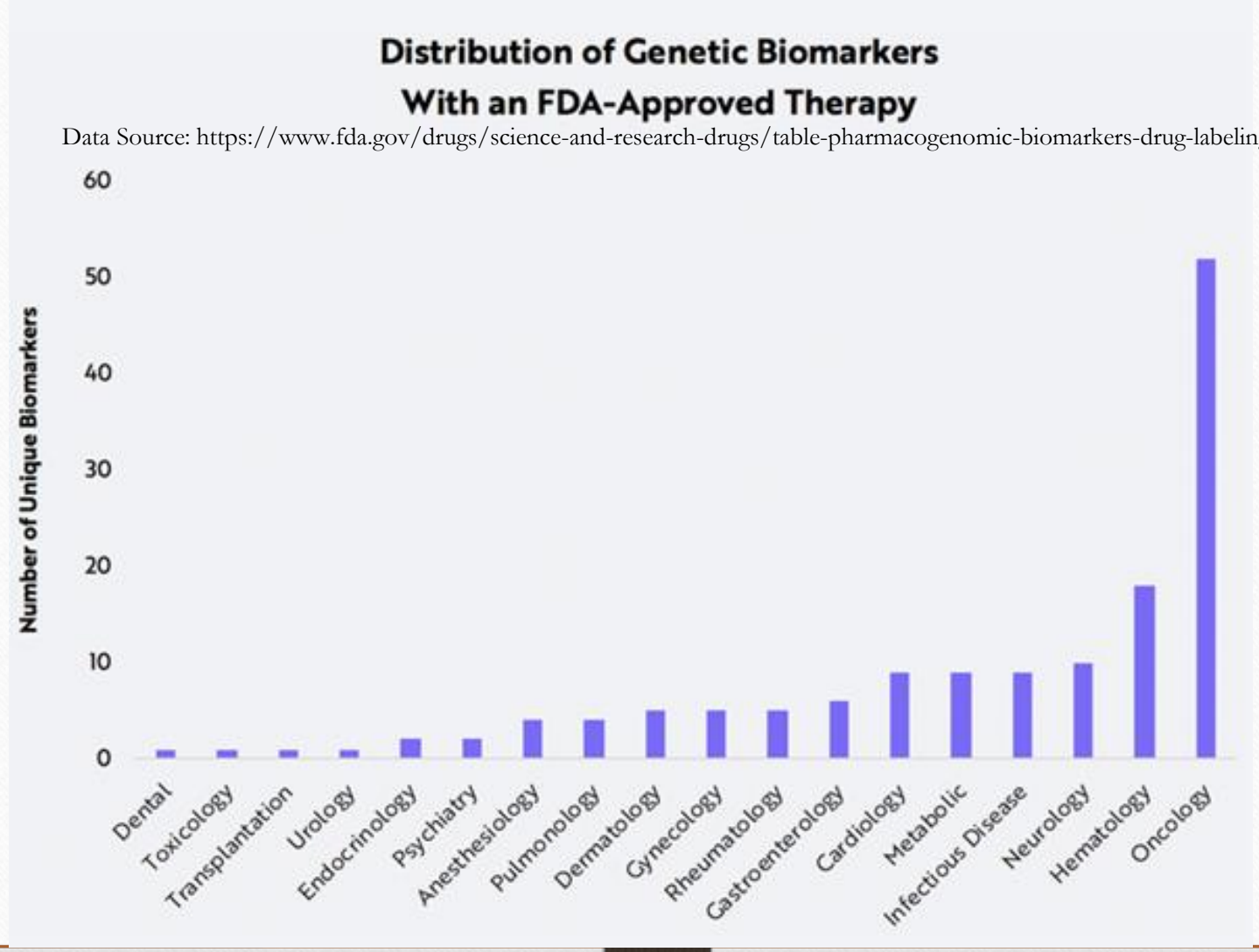
b) There is **lack of effective treatment**

c) The **phenotype** is quite **heterogeneous**

d) Complex diseases are often **late-onset**, thus predisposed individuals could benefit from early detection and preventive testing/medical follow-up

Visscher *et al., Am J Hum Genet,* 2017

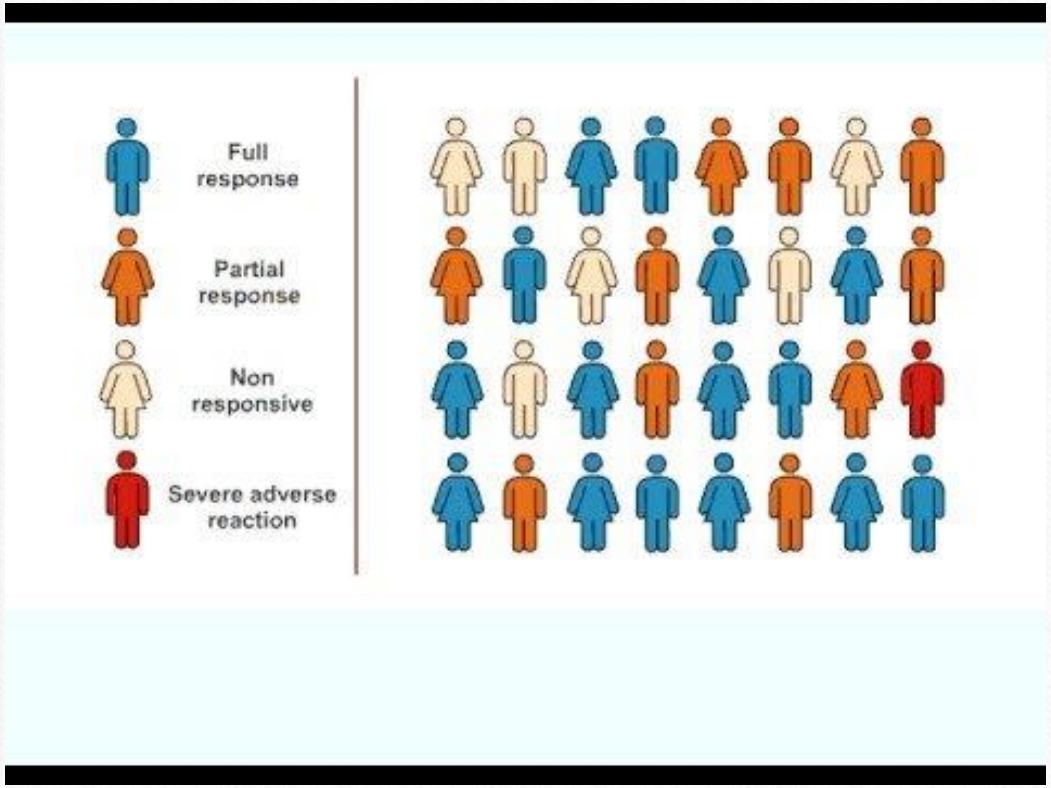
Genetically-informed translational research/Drug development

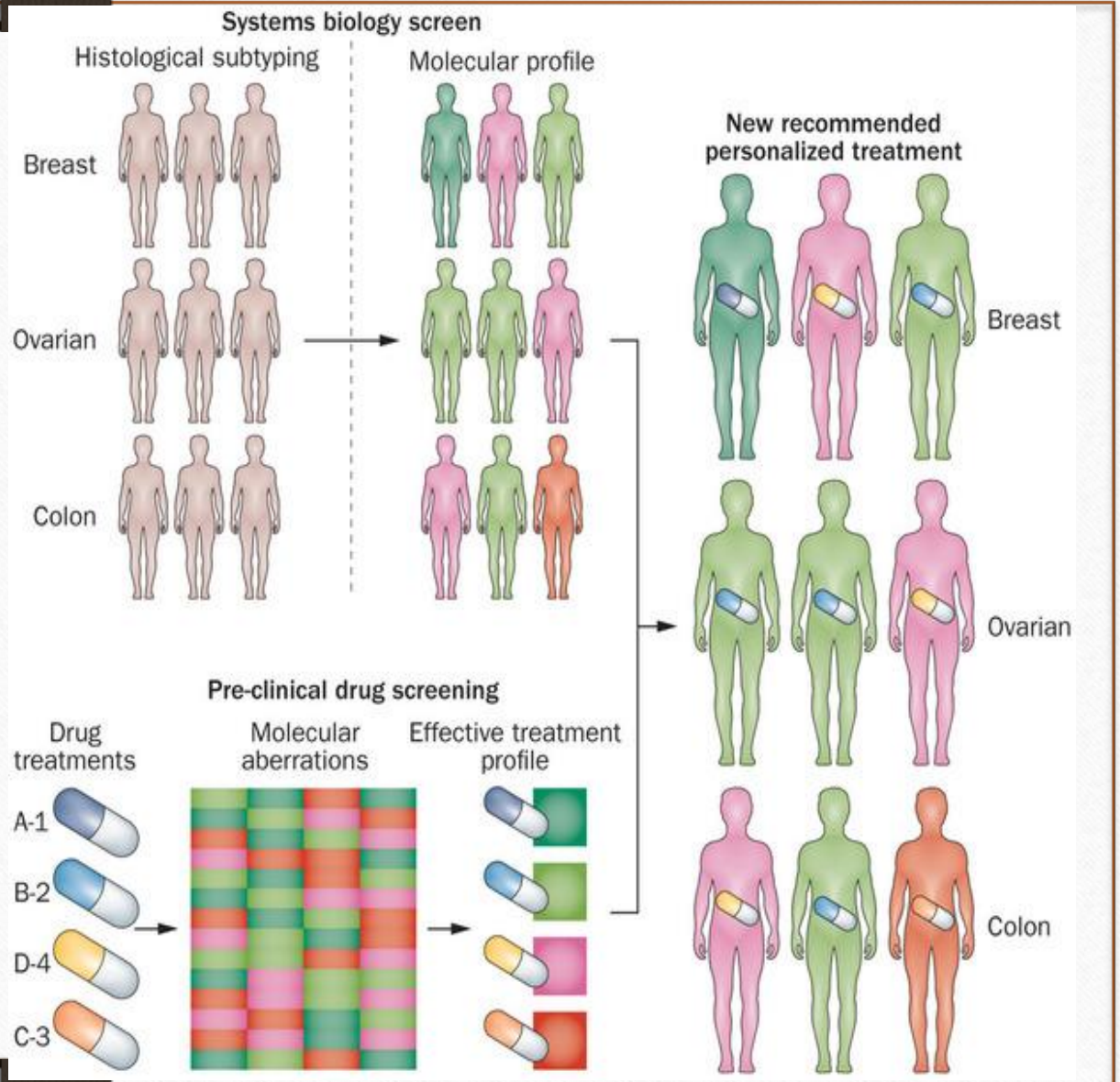


Graph source: ARK Investment Management LLC, 2020

Genetically-informed translational research/ Drug development

Pharmacogenomics





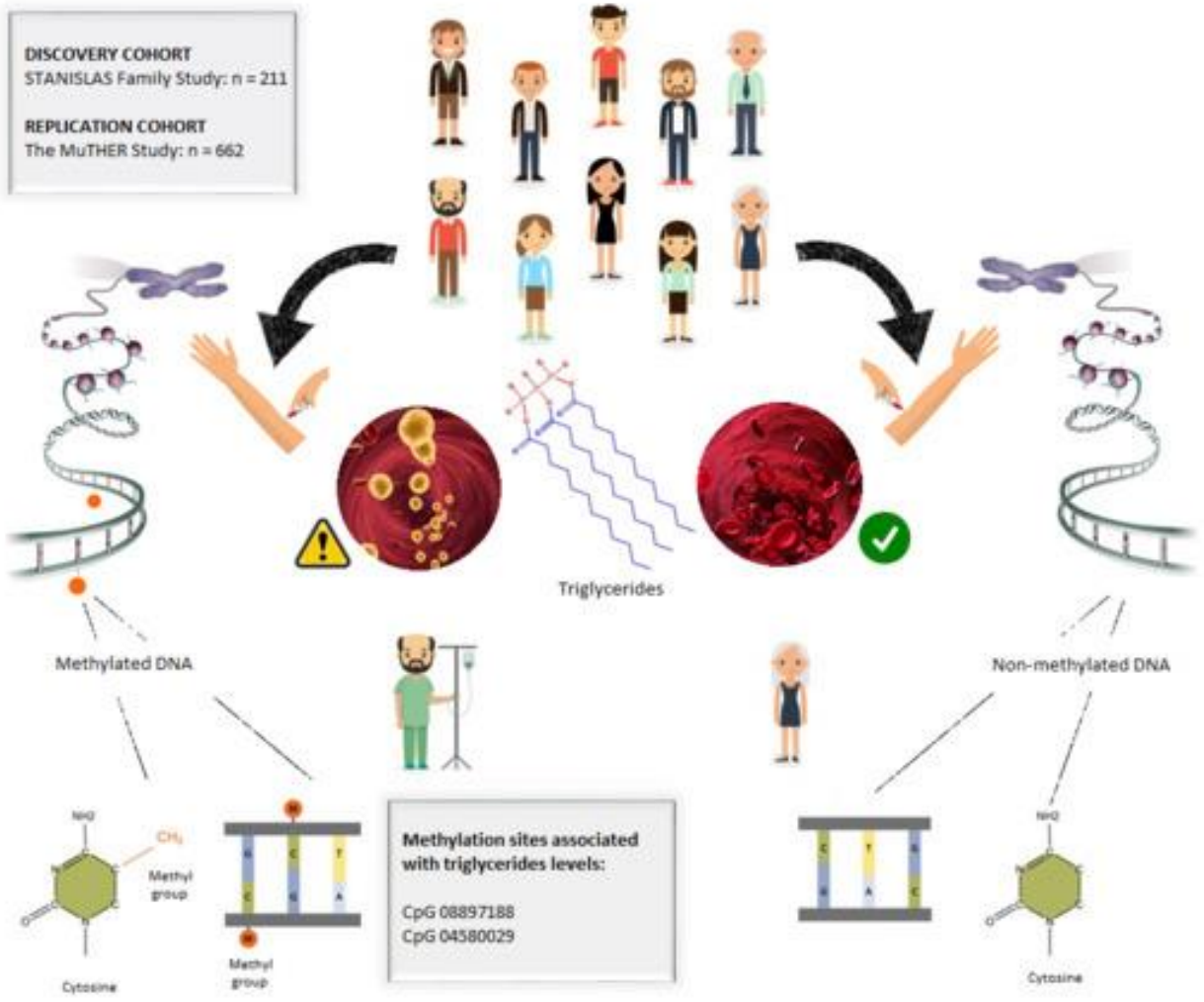
Werner, Mills & Ram, *Nat Rev Clin Oncol*, 2014

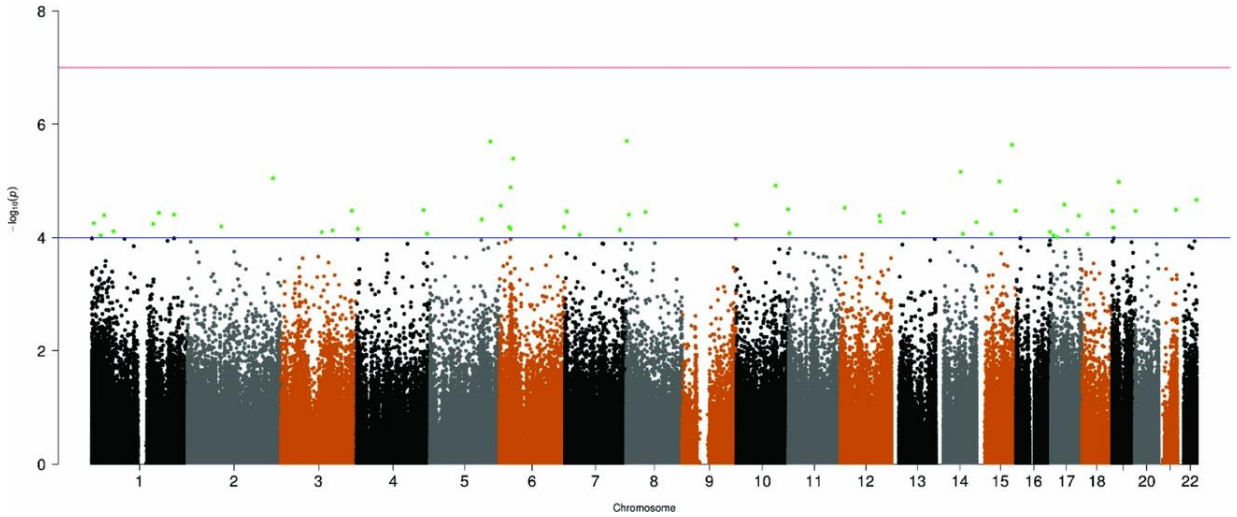
EWAS

(Epigenome-wide Association Study)

EWAS Catalog β

http://www.ewascatalog.org/





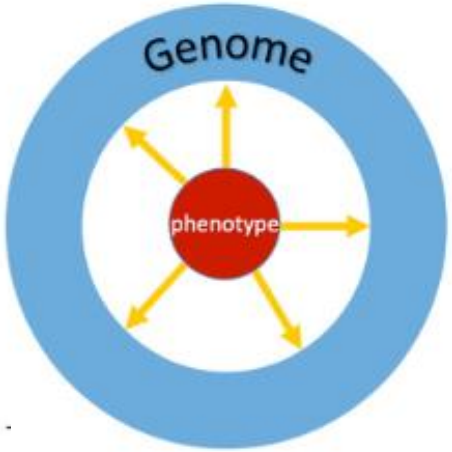
CpG site	CHR	CpG Location (bp)	*p*-value	Δβ	[Nearest] gene	Gene location
cg15583738	8	2176944	$1.98e^{-06}$	0.00408	—	—
cg06026425	5	157284650	$2.03e^{-06}$	0.008678	*CLINT1*	5q33.3
cg01321816	15	91358514	$2.32e^{-06}$	0.006536	*BLM*	15q26.1
cg03573179	6	36165382	$4.07e^{-06}$	0.003791	*BRPF3*	6p21.31
cg20519670	14	65172006	$6.91e^{-06}$	-0.00561	*PLEKHG3*	14q23.3
cg16463697	2	223886480	$8.97e^{-06}$	0.006082	*KCNE4*	2q36.1
cg00785856	15	59041883	$1.03e^{-05}$	0.005749	*ADAM10*	15q21.3
cg22033061	19	17531746	$1.04e^{-05}$	0.003424	*FAM125A/MVB12A*	19p13.11
cg19830950	10	102729375	$1.21e^{-05}$	0.002377	*SEMA4G*	10q24.31
cg08093277	6	29595299	$1.31e^{-05}$	0.006787	*GABBR1*	6p22.1
cg12961733	22	50165244	$2.17e^{-05}$	0.009437	*BRD1*	22q13.33
cg22430950	17	35166190	$2.62e^{-05}$	-0.0146	—	—
cg16208491	6	4021748	$2.76e^{-05}$	0.003249	*PRPF4B*	6p25.2
cg14830166	12	11821908	$3.01e^{-05}$	-0.01192	*ETV6*	12p13.2

Xie et al., *Int J Mol Sci*, 2019

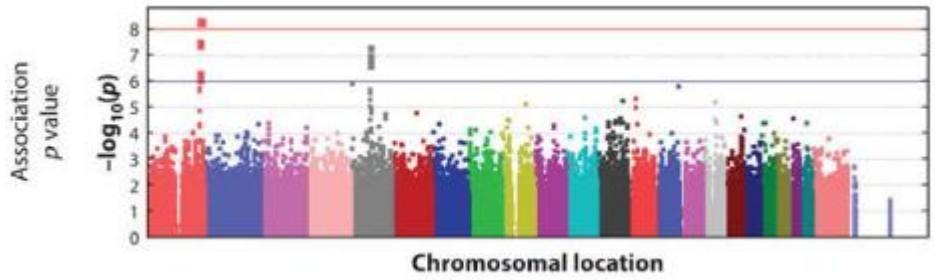
Zilhao et al., *Twin Res Hum Genet*, 2015

PheWAS (Phenome-wide Association Study)

GWAS: examines associations between specific phenotypes and genetic variants across the genome

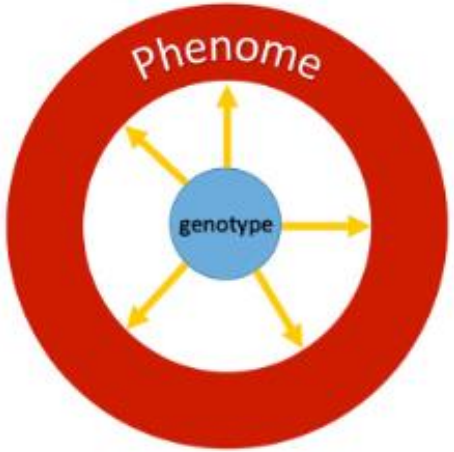


GWAS: Target phenotype -



- Ένα γνώρισμα/νόσημα
- Όλο το γονιδίωμα
- Συγκεκριμένες συσχετίσεις

PheWAS: examines associations between specific genetic variants and a large number of different phenotypes (phenome)



- Ένα υποσύνολο των παραλλαγών του γονιδιώματος
- Μεγάλος αριθμός φαινοτύπων
- Πολλαπλές συσχετίσεις

PheWAS: Target genotype –
(or other input variable, e.g., a specific disease, trait, or exposure)

Electronic Health Records (EHR)

Polygenic Risk Scores (PRSs)

(Πολυγονιδιακοί Δείκτες Κινδύνου)

- A count of the number of the risk variants across multiple genomic loci present in the person's DNA, weighted so that the presence of some risk variants is considered more important than others.

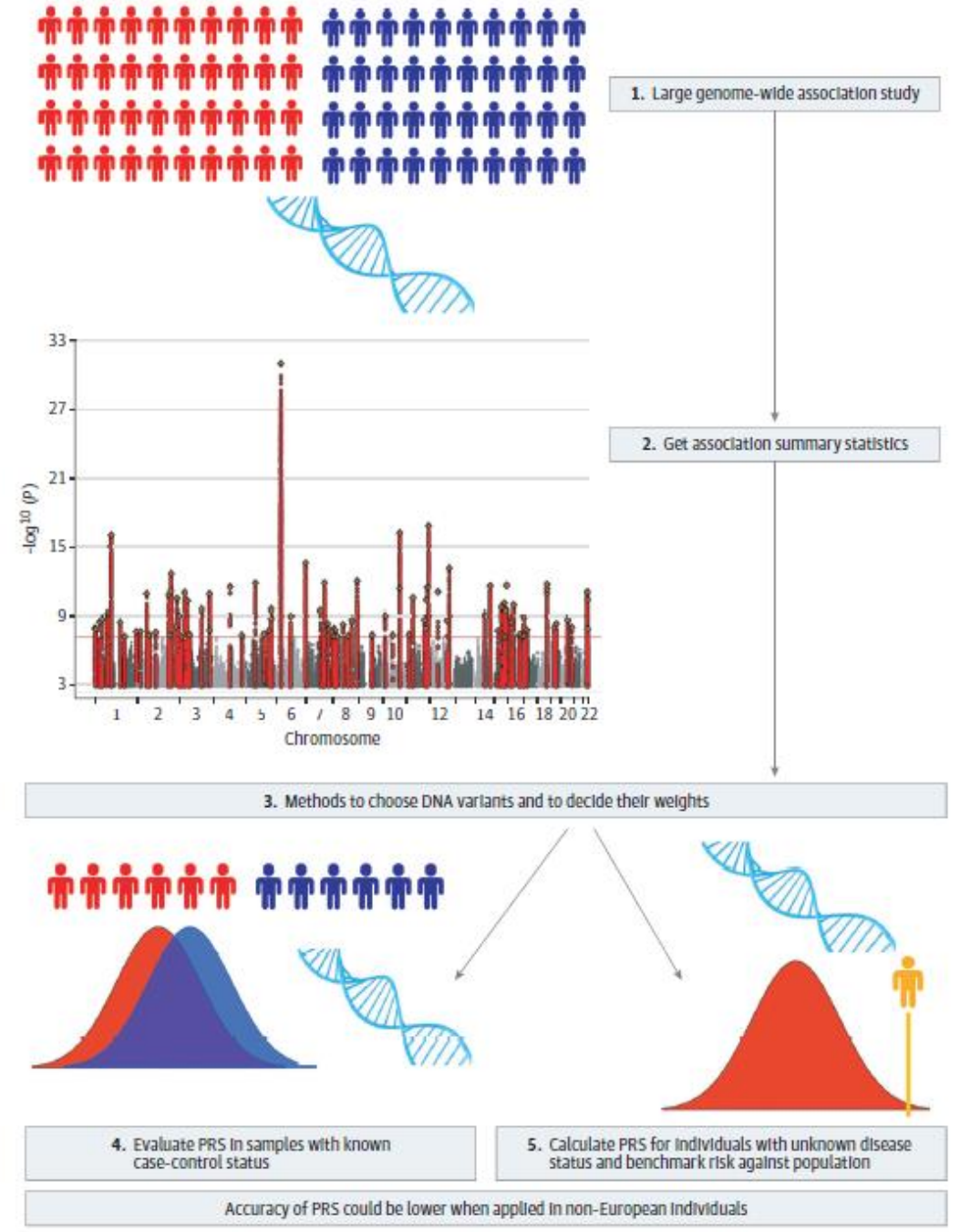
  - ✓ IBD (Crohn's, UC)
  - ✓ T1DM and autoimmune diseases
  - ✓ T2DM and obesity
  - ✓ CVD and hypertension
  - ✓ AD and neurodegenerative disorders
  - ✓ ADHD and ASD
  - ✓ Schizophrenia

  - ✓ Preterm delivery
  - ✓ Drug response (PGx)

  - ✓ Ancestry testing

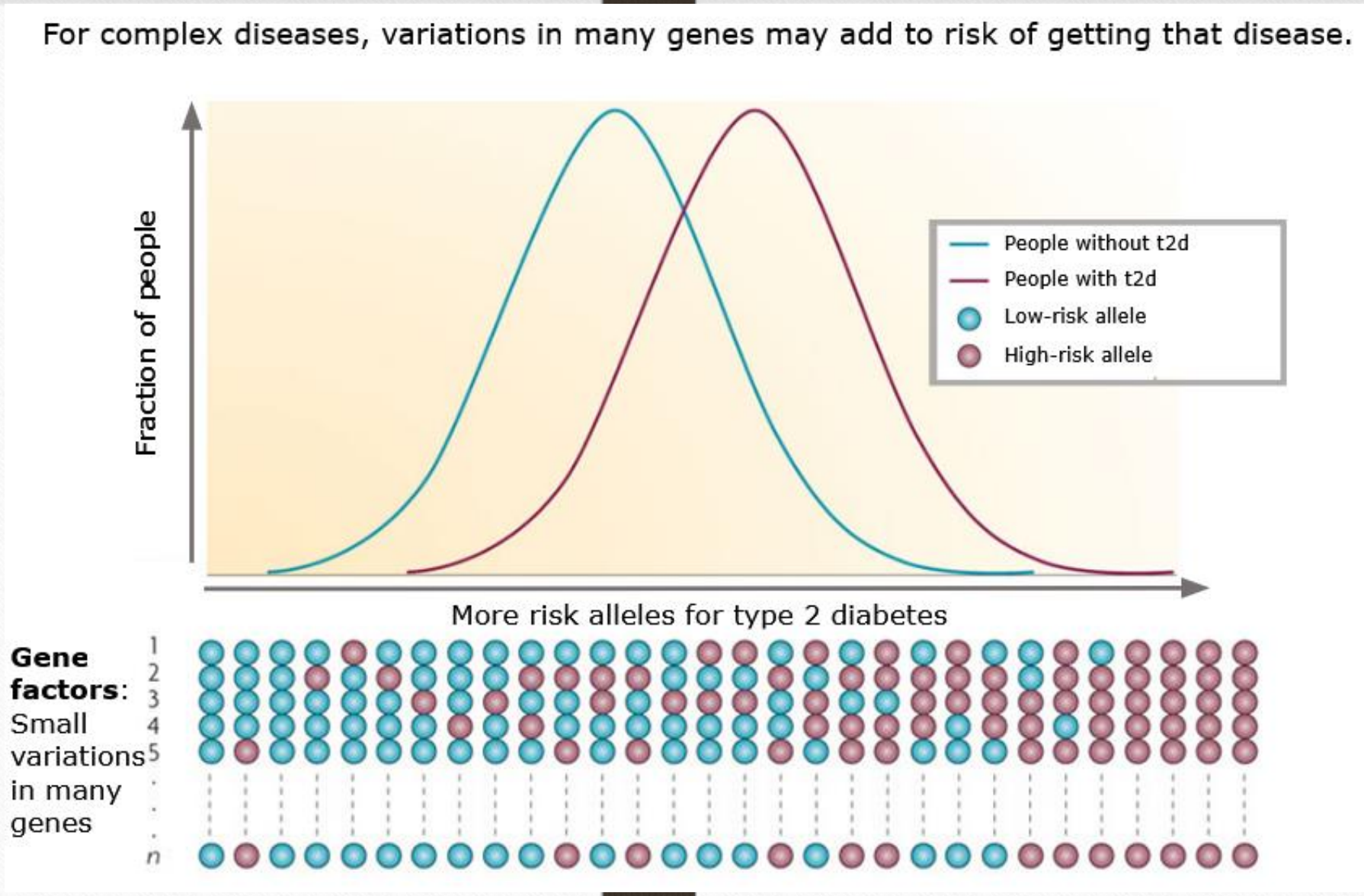
PGS Catalog

Figure 2. Schematic of the Steps Needed to Generate and Validate Polygenic Risk Scores (PRS)



Wray et al., *JAMA Psychiatry*, 2020

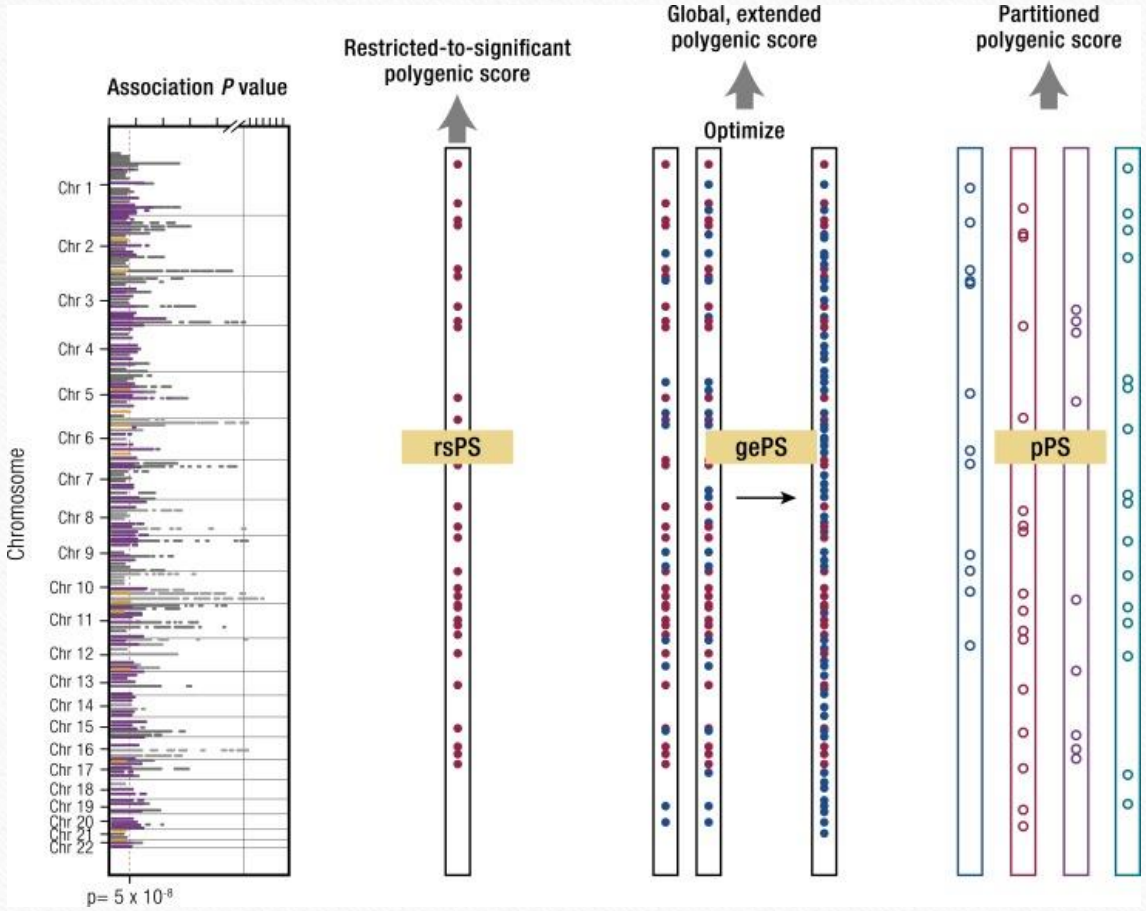
Polygenic Model of Disease



Fletcher & Houlston, *Nat Rev Cancer*, 2010 – Modified from: https://sites.google.com/uw.edu/gemnet-bio-t2d-curriculum/home

Polygenic Risk Scores (PRSs)

(Πολυγονιδιακοί Δείκτες Κινδύνου)



Restricted-to-significant polygenic scores (rsPSs): scores composed of variants at the extreme of a statistical distribution, most usually those that pass the genome-wide significant threshold ($p<5x10^{-8}$) for the trait concerned.

Global extended polygenic scores (gePSs): scores generated from a deeper set of variants generated from genome-wide analyses, typically involving large numbers of sub-threshold significant variants ($p>5x10^{-8}$).

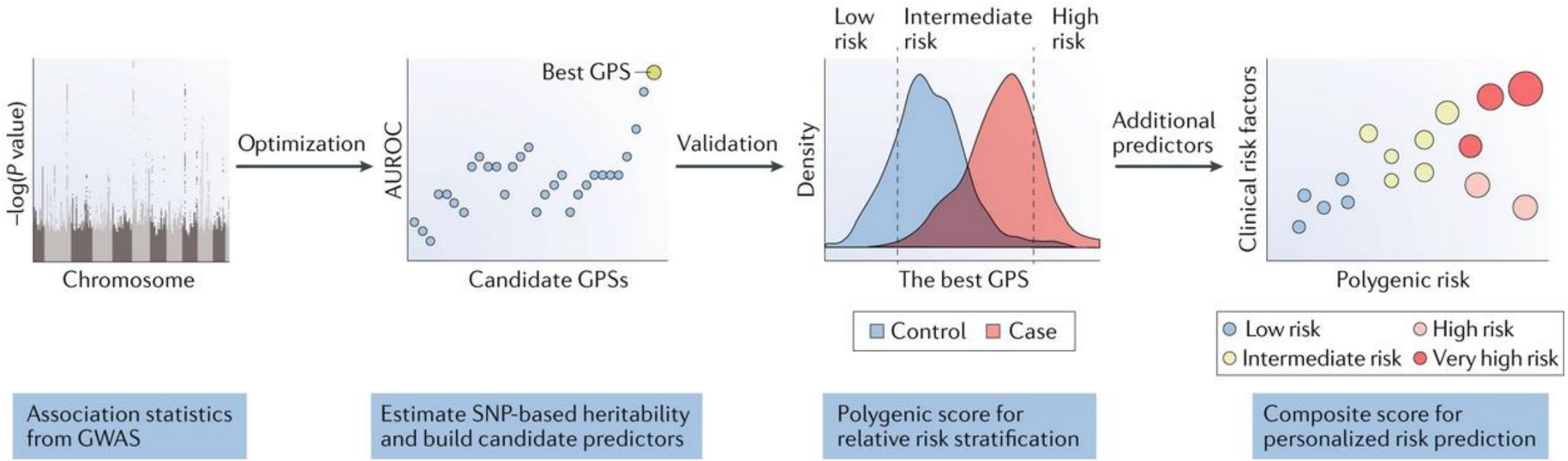
Partitioned (or process-specific) polygenic scores (pPSs): scores composed of variants grouped according to some common biological process (*e.g.*, association with a related endophenotype, tissue expression of related genes, chromatin state).

Alzheimer Disease (AD):

- AD risk, age-at-onset
- Progression (normal to MCI to AD)
- Cognitive function and memory
- Brain structure and function (MRI, PET)
- Biochemical changes in brain and periphery (CSF, blood, post-mortem)

Udler *et al.*, *Endocr Rev*, 2019

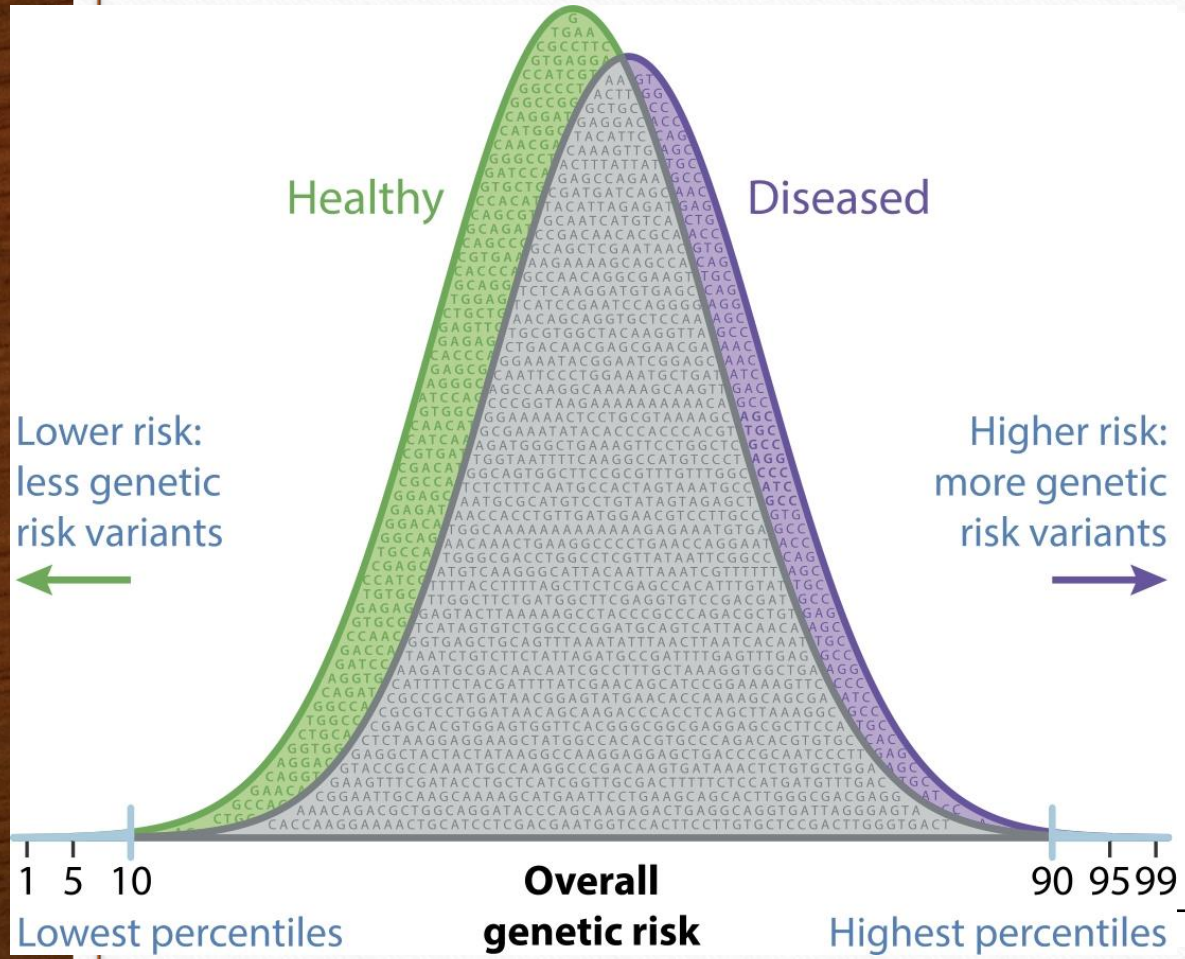
Polygenic Risk Scores (PRSs)

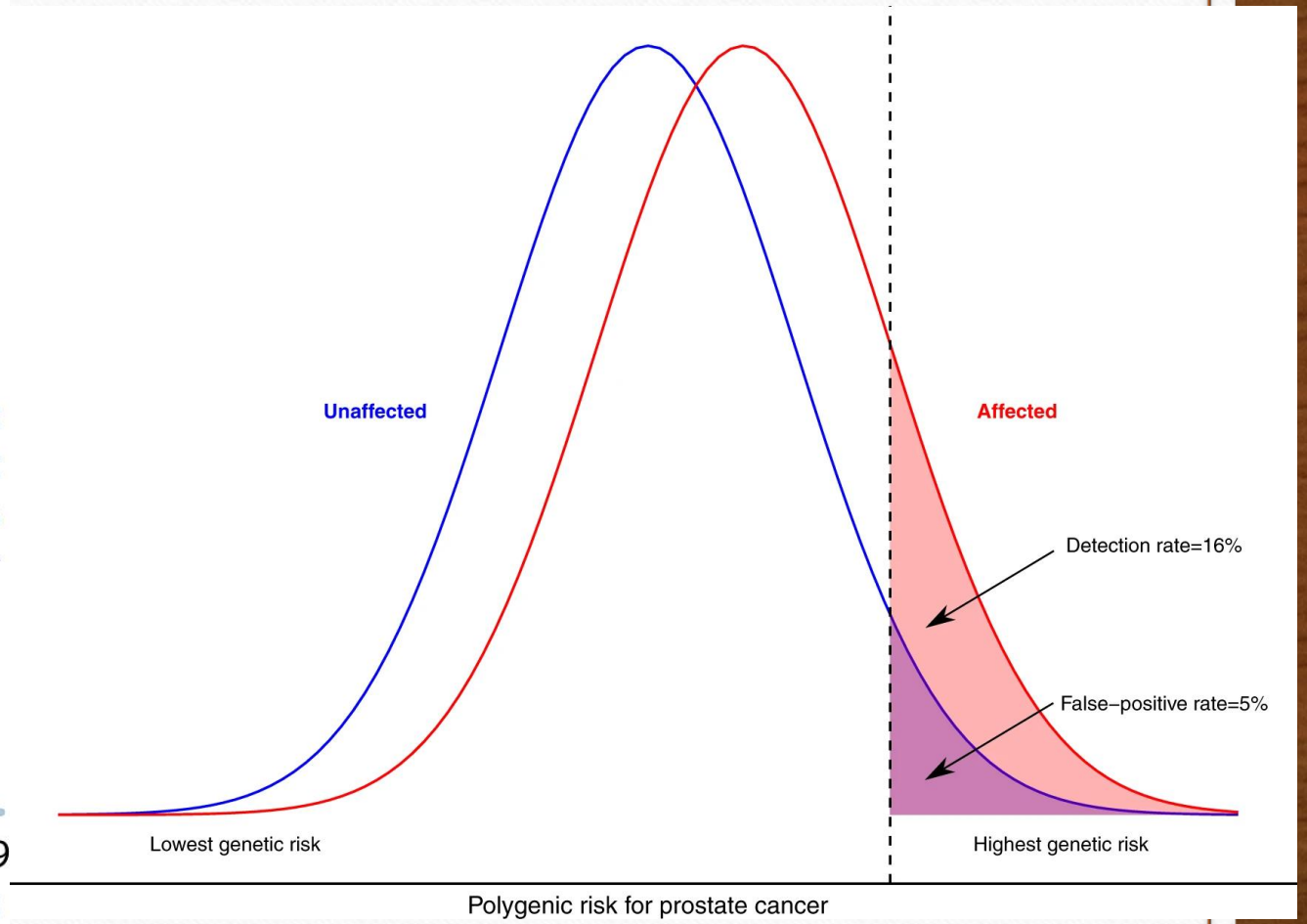


A genome-wide polygenic risk score (GPS) is based on genome-wide association study (GWAS) summary statistics. The optimization step enables selection of the best method according to the genetic architecture of a disease under study. The validation step requires an external cohort and is critical to obtaining reliable metrics of performance. Clinical predictors of absolute risk will require incorporation of additional demographic, clinical or lifestyle factors into composite risk models. AUROC, area under receiver operating characteristic. SNP, single-nucleotide polymorphism.

Liu & Kiryluk, *Nat Rev Nephrol*, 2018

Polygenic Risk Scores (PRSs)

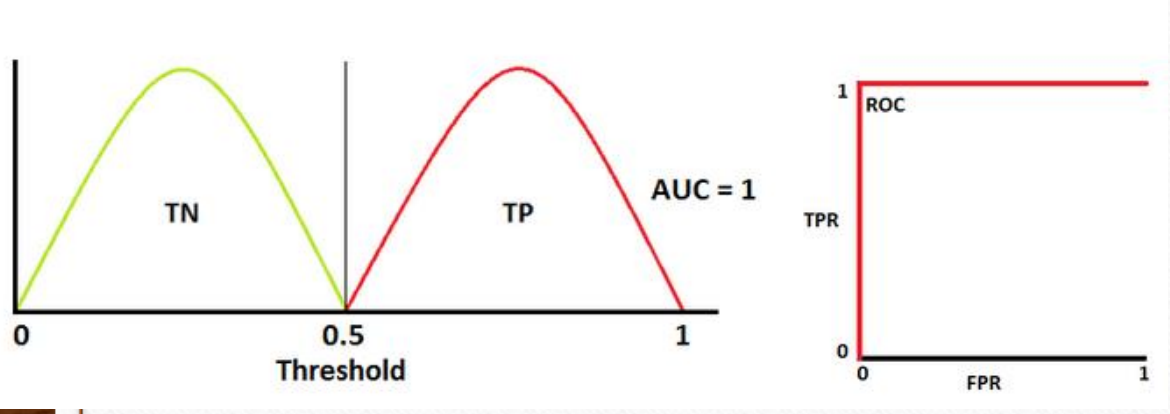




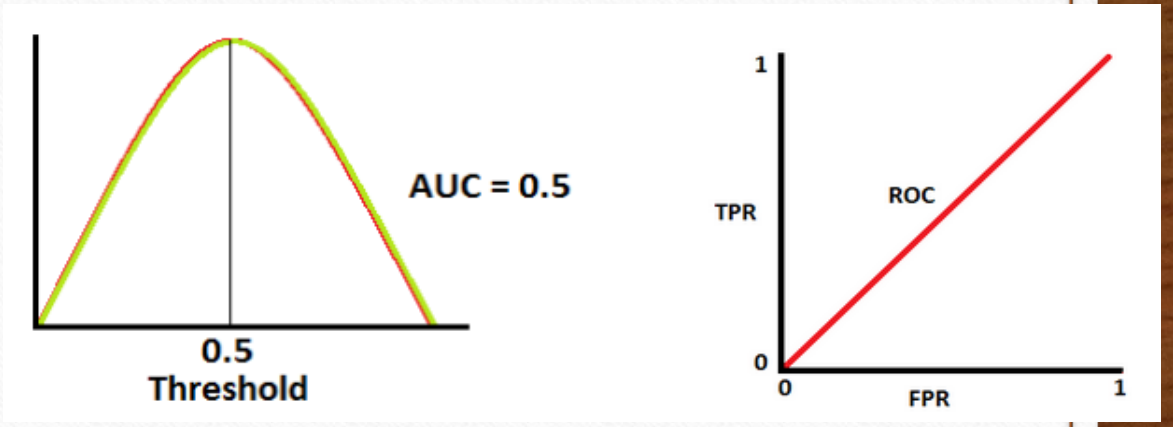
https://www.degruyter.com/document/doi/10.1515/medgen-2020-2006/html

Sud *et al, njp Precision Oncology*, 2021

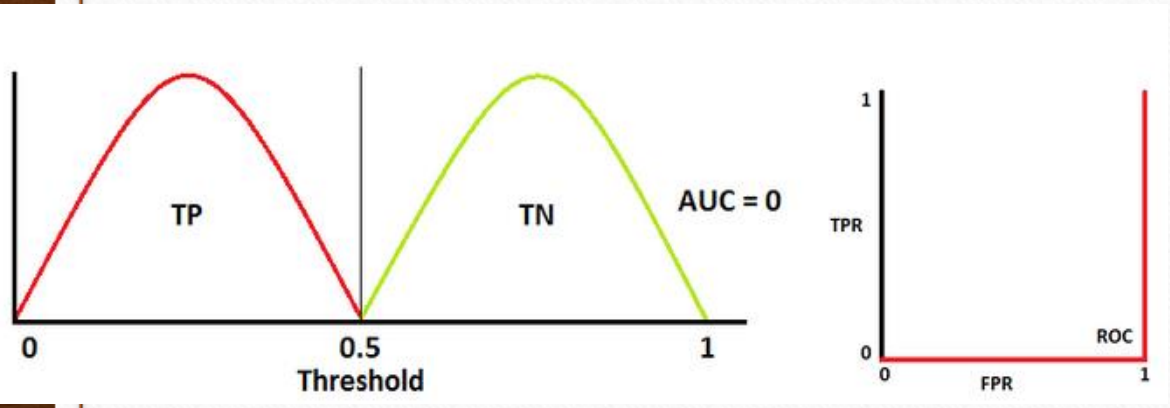
The perfect model



The non-informative model



The worst model



The current status in PRS research

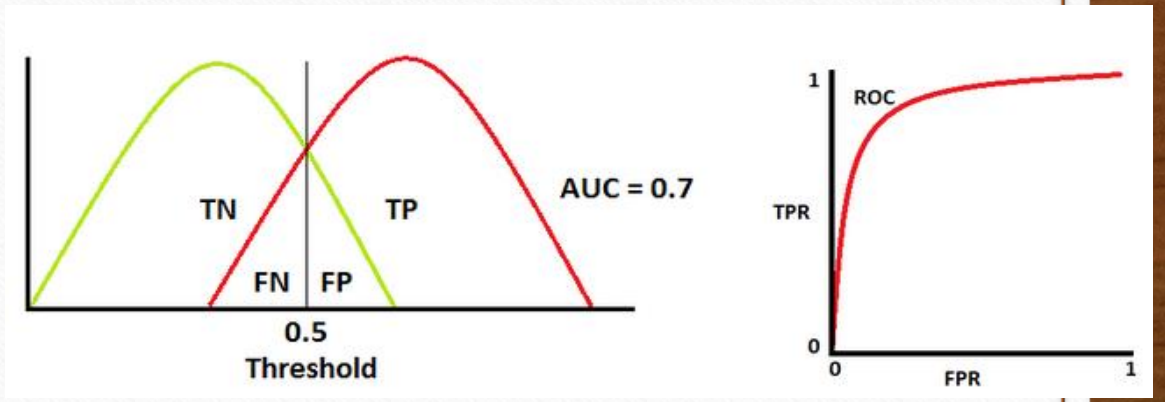
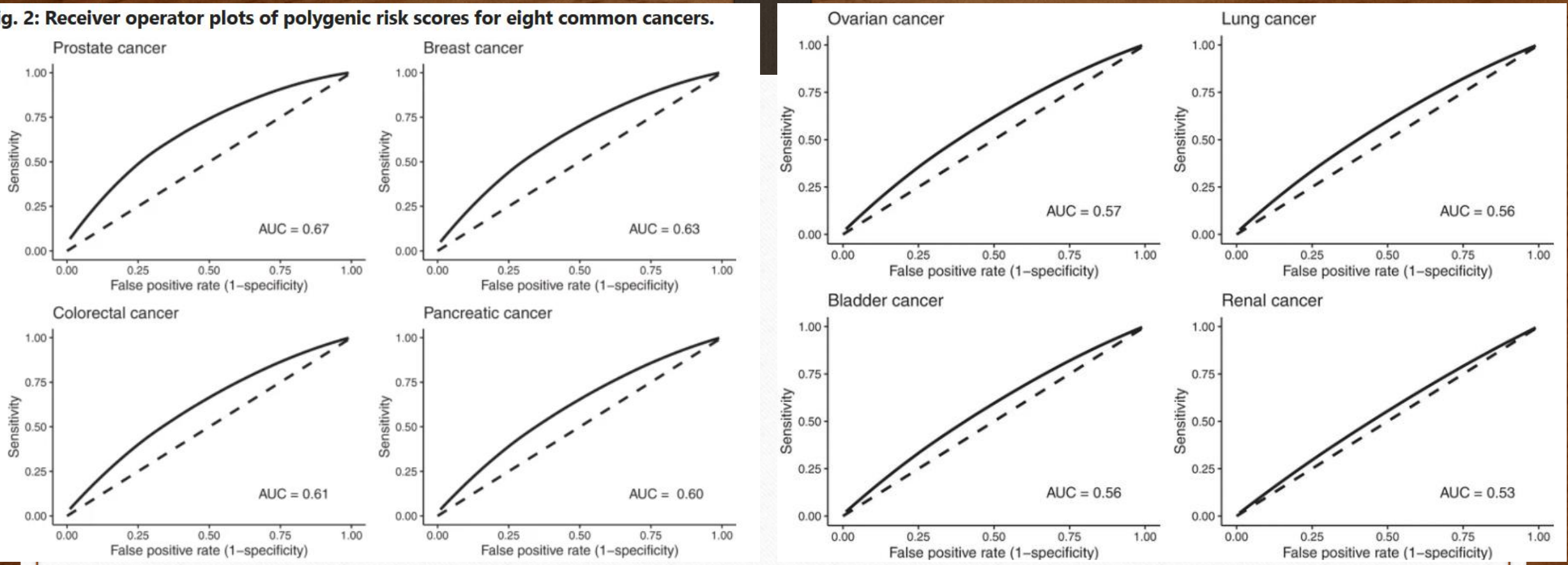
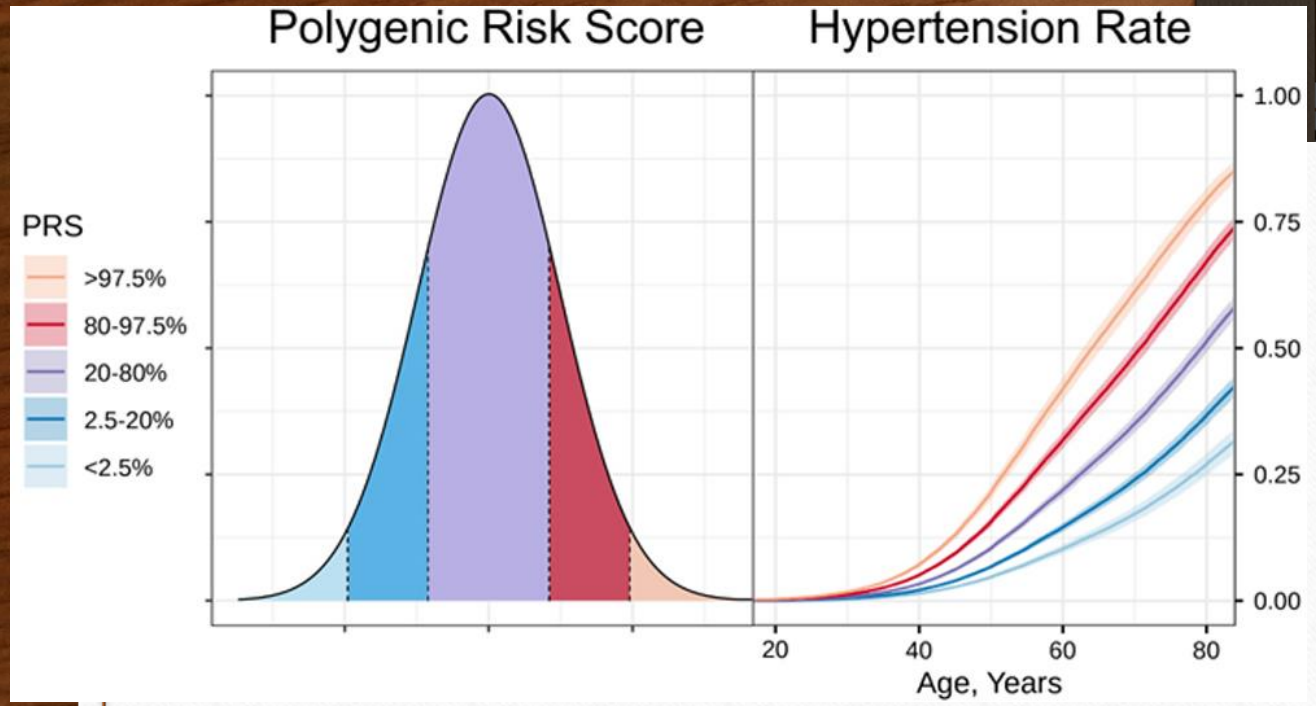


Fig. 2: Receiver operator plots of polygenic risk scores for eight common cancers.

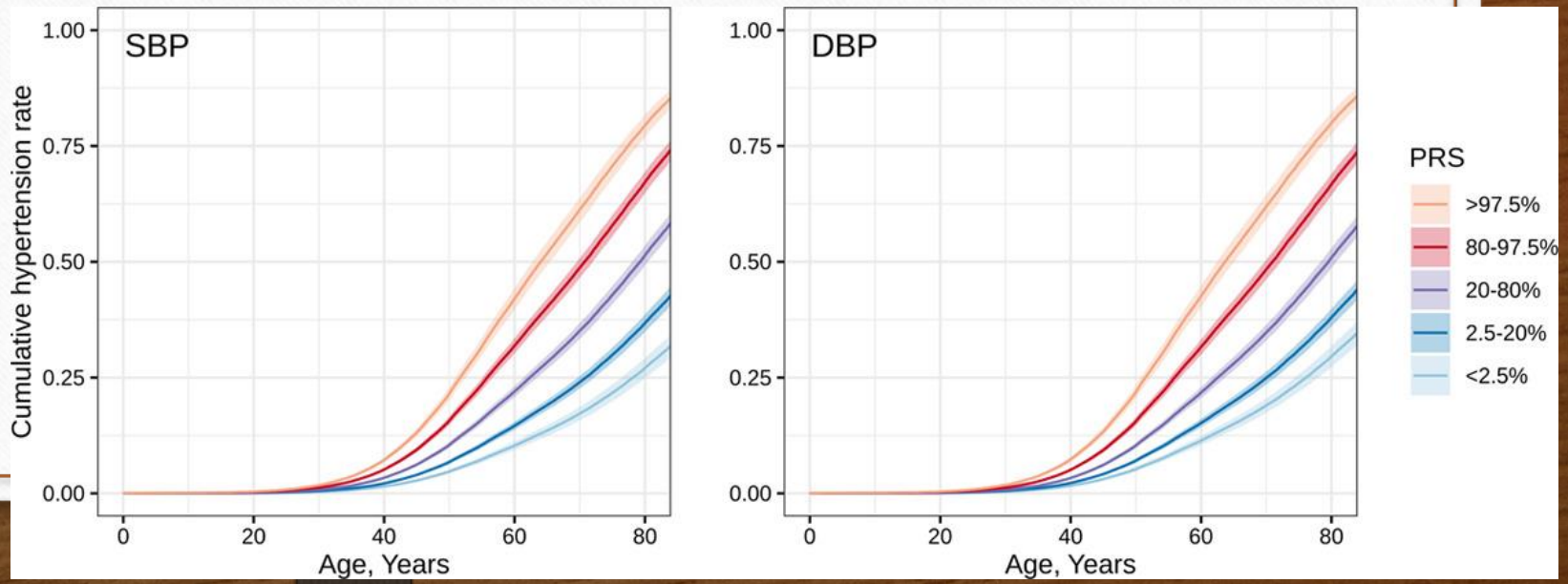


- ✓ **AUC values** of 0.7–0.8 are considered as acceptable and >0.8 as affording good discrimination
- ✓ Digital **mammography** for breast cancer screening has an AUC of 0.78!
- ✓ Breast cancer has the best characterized set of **non-genetic risk factors**: the AUC for these risk factors is 0.637 while the AUC of the current PRS is 0.631; and <u>in combination the AUC is 0.683</u>

Sud *et al, njp Precision Oncology*, 2021

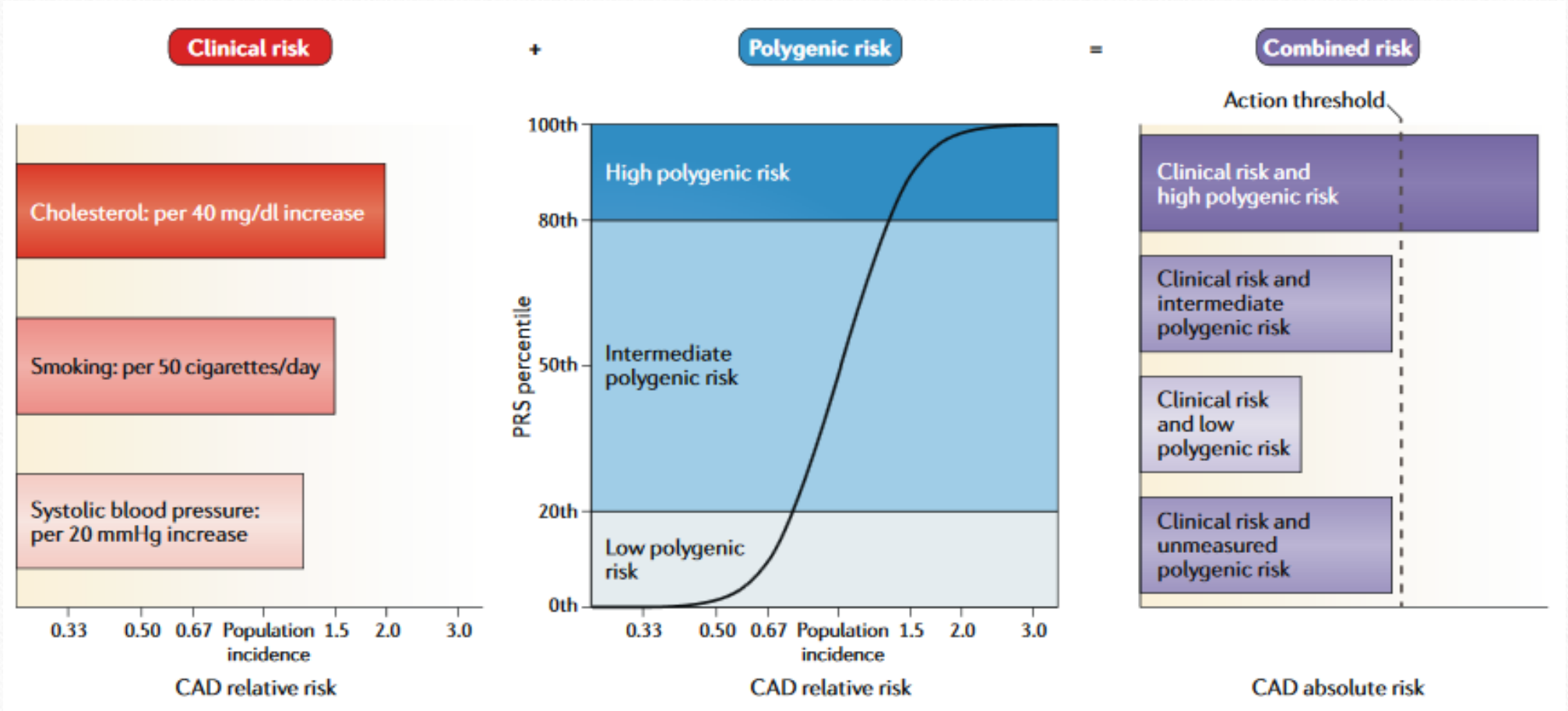


BP PRSs together with traditional risk factors could improve prediction of hypertension and particularly early-onset hypertension, which confers substantial CVD risk



Vaura *et al, Hypertension,* 2021

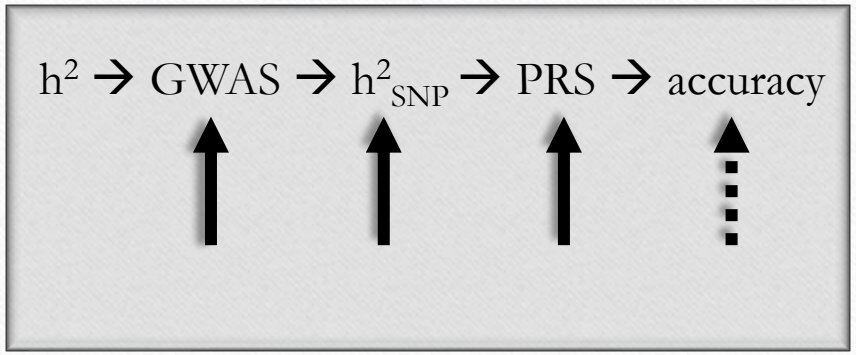
Polygenic Risk Scores (PRSs) – The future prospects



Torkamani *et al., Nat Rev Genet,* 2018

PRSs pros

- DNA collection is performed once
- The costs of generating PRSs is very low
- Can be recalculated from the same genetic data if new information to improve PRS becomes available
- From a single DNA test, many PRSs can be generated
- Important for population-based screening and prevention programs (ex. CRC, BrCa, glaucoma)
- Specific diagnosis in early phase of illness when symptoms may be general and non-specific (ex. CKD)
- Contribution to treatment choices, response to treatment, adverse outcomes (PGx)

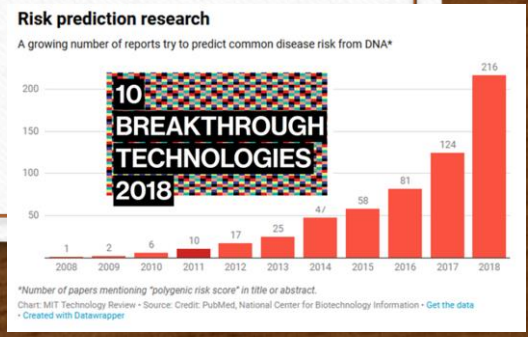


PRSs cons

- Prerequisite: Large GWAS sample sizes
- Still limited predictive value (accuracy):
  - Genetic factors are not the only risk factors in common diseases (age, sex, environment, etc)
  - PRSs explain only part of the genetic contribution (based on common variants only, each with a small effect)
  - Prediction is not equally valid when using the same PRSs across ethnic ancestries → Need for ancestry-specific PRSs
- The utility of the PRSs varies between conditions – dependent on their calculated heritability

---

- How should PRS-associated risk be communicated to the general public? (patient, family members, etc)
- D-t-C companies already provide PRS for a number of diseases and traits – Are clinicians ready to/should they interpret "online PRS calculator" data?
- Implications for health insurance?



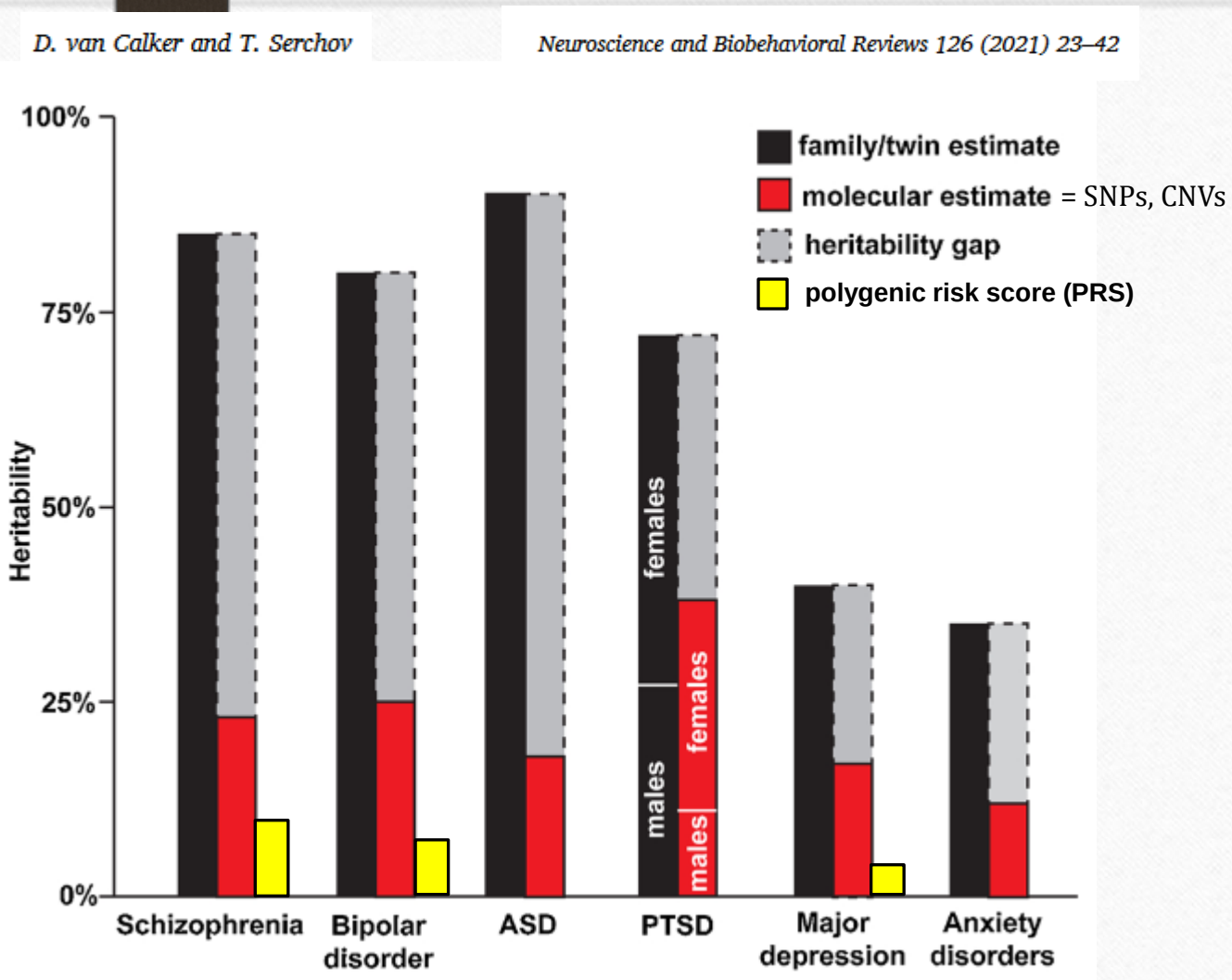
Heritability of neuropsychiatric phenotypes

$$P = G + E$$

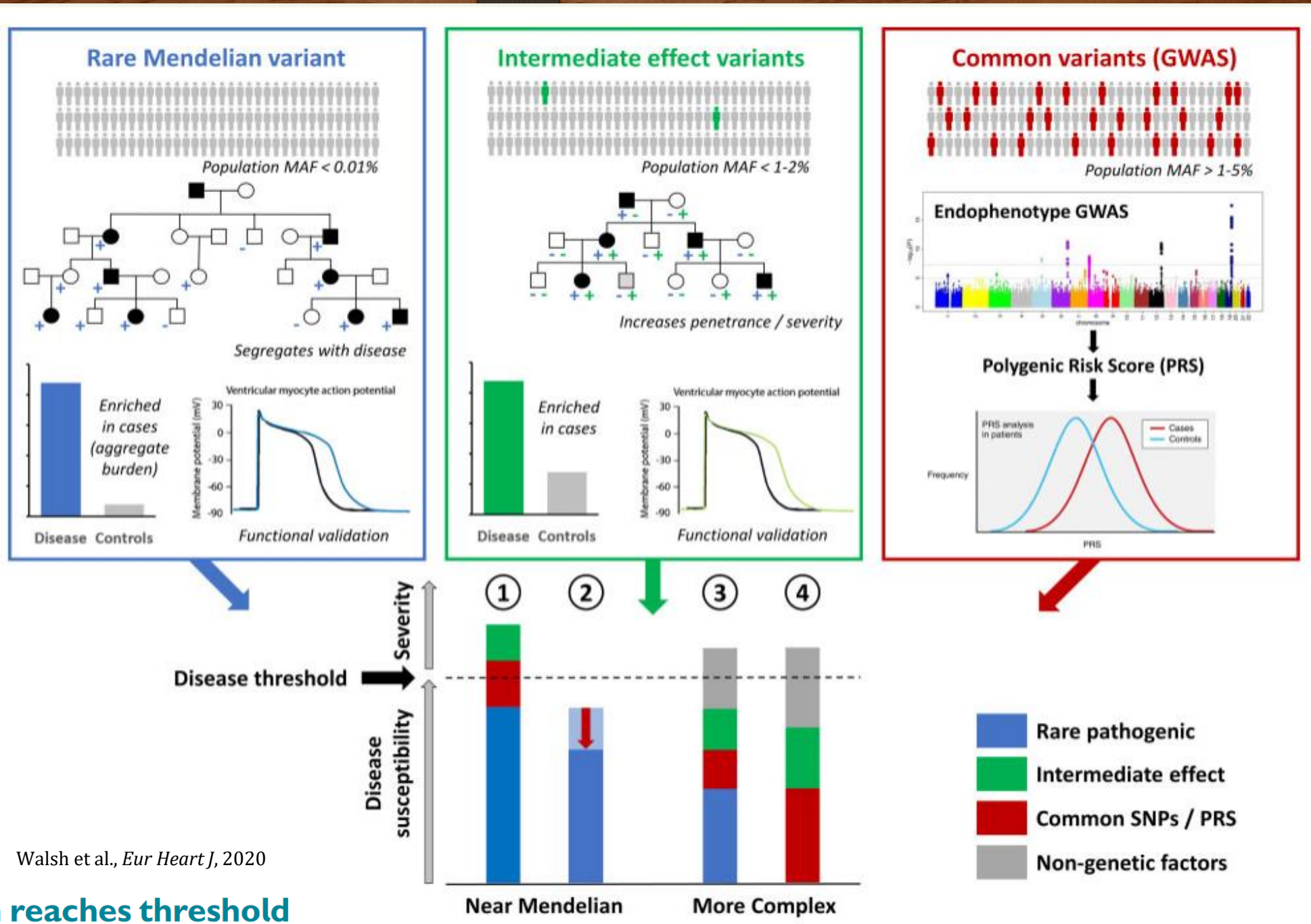
$$V_P = V_G + V_E + V_{GxE}$$

$$V_P = V_A + V_D + V_I + V_E$$

Phenotypic Variance | Additive Genetic Variance | Non-Additive Genetic Variance | Environmental Variance



Σύνοψη

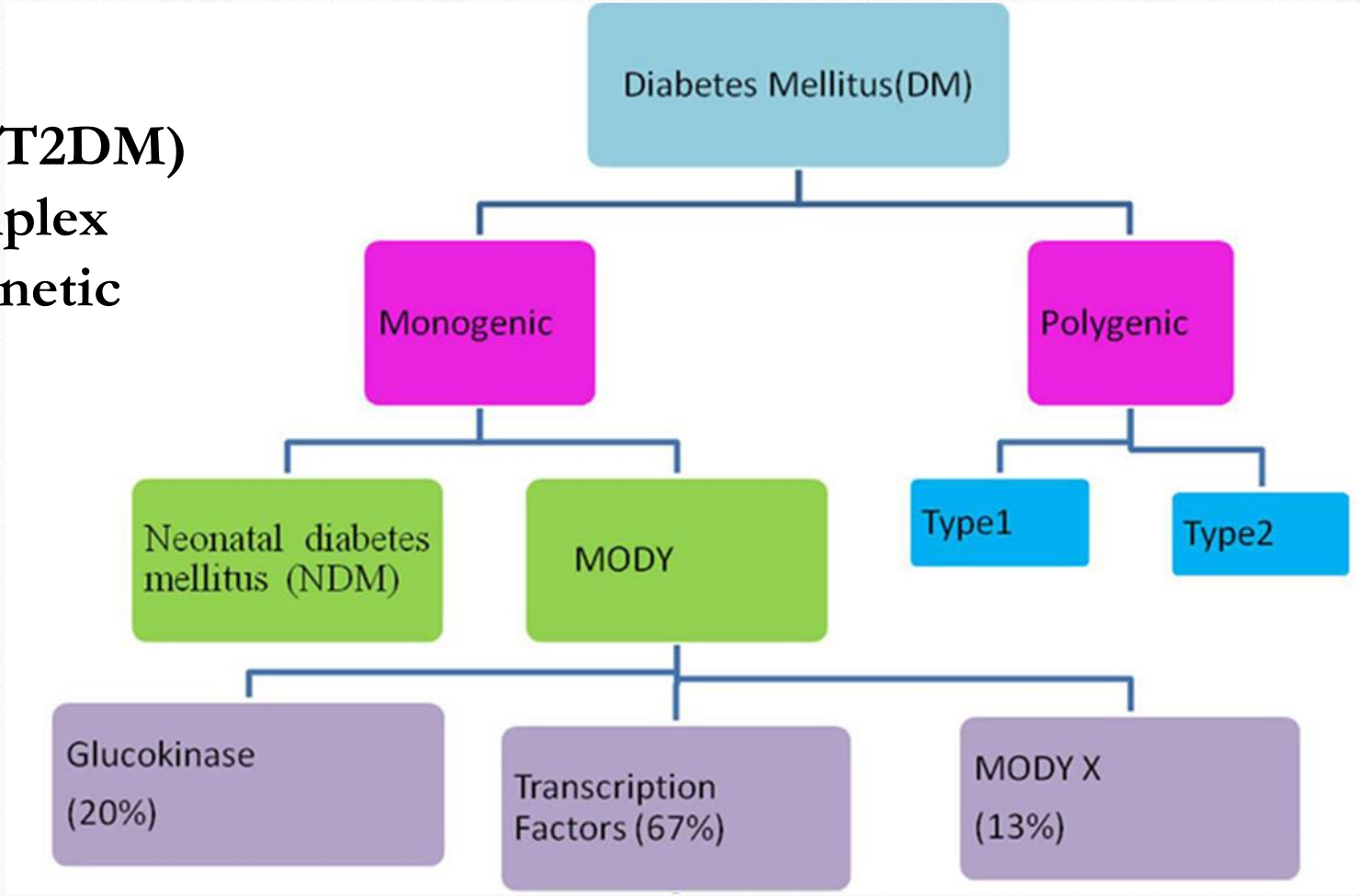


Walsh et al., *Eur Heart J*, 2020

When genetic burden reaches threshold

Part II

Type 2 Diabetes Mellitus (T2DM) as a paradigm of a complex disease with a strong genetic component



Type 2 Diabetes Mellitus (T2DM)

- 80-90% of all diabetes cases
- NIDDM/T2DM (70-85%)
- 1-5% prevalence (6-8% in USA, 7% GR, 5-10% Malta/Finland, 15% Turkey, 25% N. Zealand aborigines, 39% Nauru-Pacific Ocean, 50% Pima Indians)
- ~500M cases, ~700M by 2050 (International Diabetes Federation)
- 1st degree relatives: ~3X higher risk of developing T2DM
- Macrovascular and microvascular complications
- Genes, family history, sex, ancestry, age → non-modifiable T2DM predisposing factors

COMPARISON OF TYPE 1 AND TYPE 2 DIABETES MELLITUS

Characteristic	Type 1 (IDDM)	Type 2 (NIDDM)
Sex	Female = male	Female > male
Age at onset	Childhood and adolescence	Adolescence through adulthood
Ethnic predominance	Whites	African Americans, Mexican Americans, Native Americans
Concordance		
Monozygotic twins	33%-50%	69%-90%
Dizygotic twins	1%-14%	24%-40%
Family history	Uncommon	Common
Autoimmunity	Common	Uncommon
Body habitus	Normal to wasted	Obese
Acanthosis nigricans	Uncommon	Common
Plasma insulin	Low to absent	Normal to high
Plasma glucagon	High, suppressible	High, resistant
Acute complication	Ketoacidosis	Hyperosmolar coma
Insulin therapy	Responsive	Resistant or responsive
Oral hypoglycemic therapy	Unresponsive	Responsive

Genetics in Medicine, Thompson & Thompson, 7th ed., 2007

1 in 3 US adults by 2050

Admixed populations

The Staggering Costs of Diabetes

GROWING EPIDEMIC

Diabetes affects **30 million** children and adults in the U.S.

That's **1 in 11** Americans.

84 million Americans have prediabetes and are at risk for developing type 2 diabetes.

90% of them don't know they have it.

Every **21 seconds** someone in the U.S. is diagnosed with diabetes.

HUMAN COSTS

African Americans and Hispanics are over **50%** more likely to have diabetes than non-Hispanic whites.

People with diabetes are at higher risk of serious health complications:

STROKE | BLINDNESS | KIDNEY DISEASE | HEART DISEASE | LOSS OF TOES, FEET, OR LEGS

ECONOMIC COSTS

The total cost of diabetes and prediabetes in the U.S. is **$322 billion.**

The average price of insulin increased nearly **3x** between 2002 and 2013.

People with diabetes have health care costs **2.3x greater** than those without diabetes.

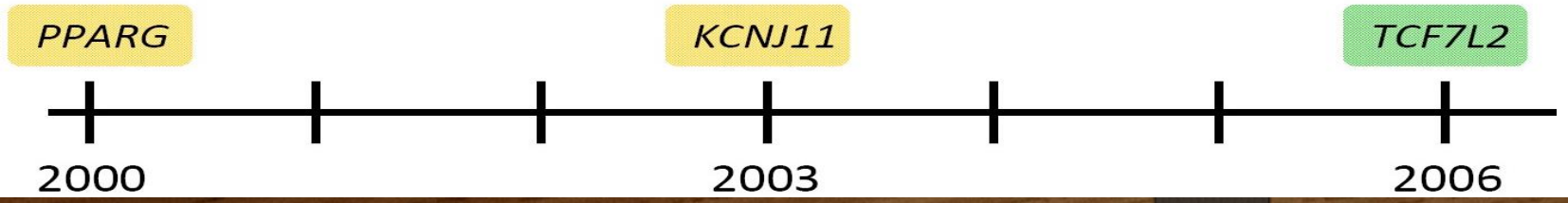
Learn more at **diabetes.org**

American Diabetes Association.

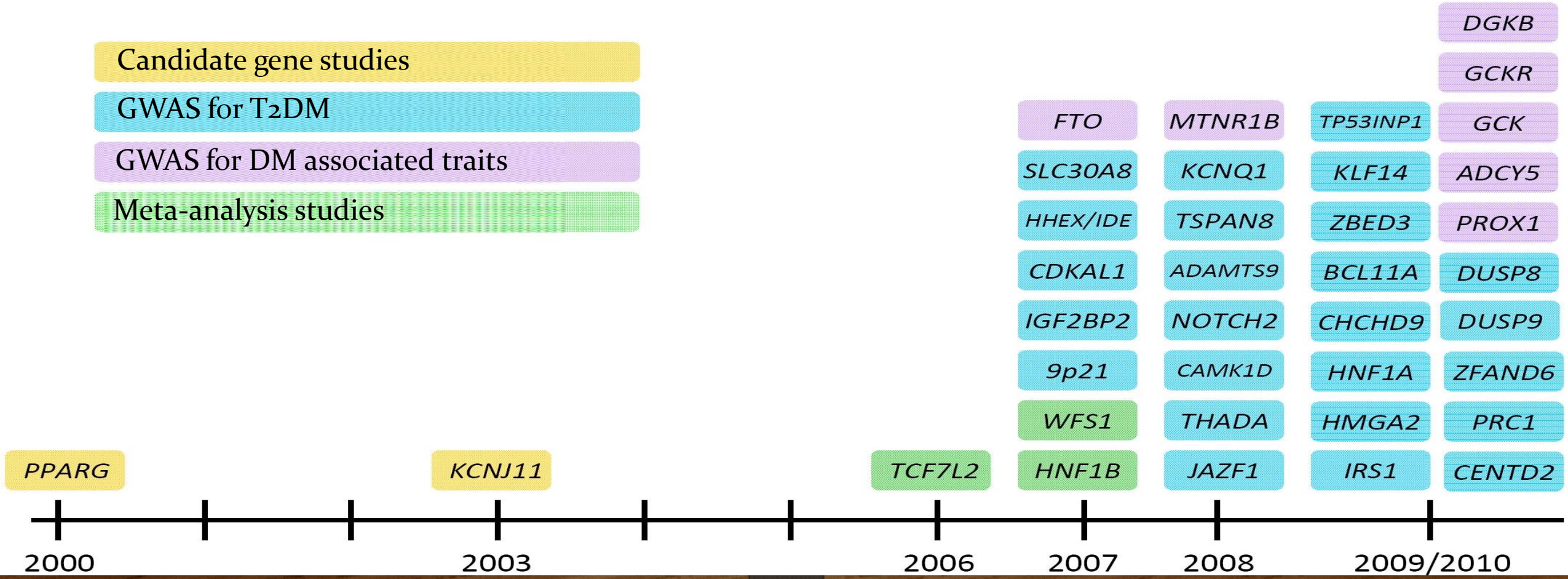
Prevention!!!

T2DM associated genes before 2007 – The pre-GWAS era

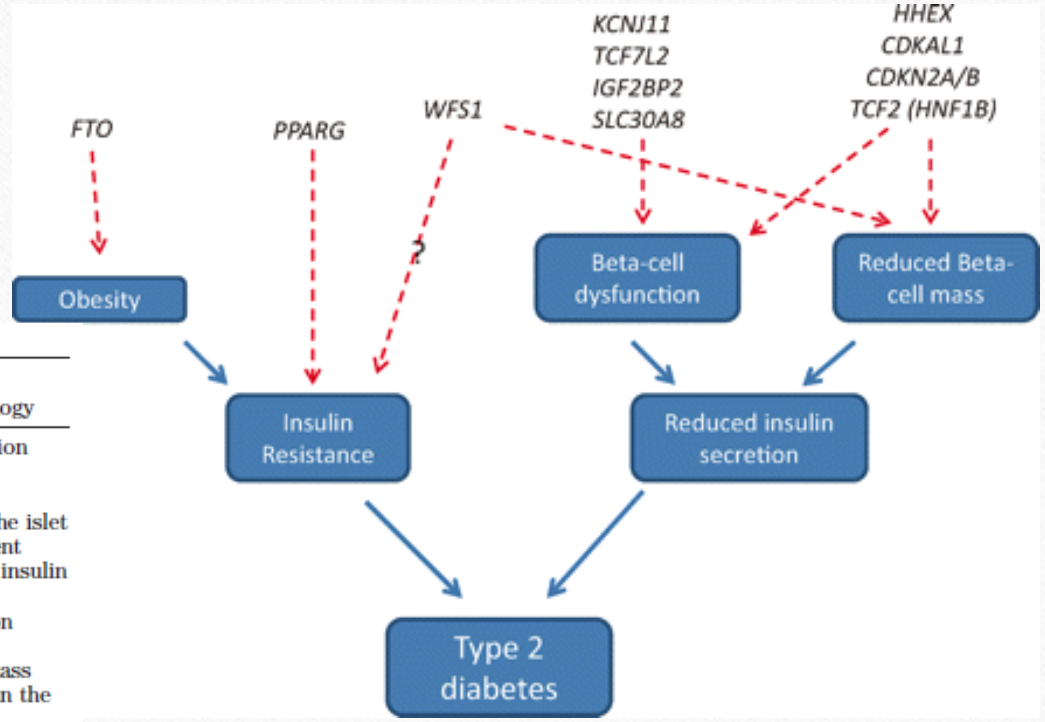
- Biological (functional) candidates (ignores less-obviously implicated genes)
- Positional candidates (small numbers of genes assessed)



T2DM associated genes after 2007 – The GWAS era



T2DM associated genes and molecular functions



Summary details of the first 17 loci with a proven role in type 2 diabetes susceptibility

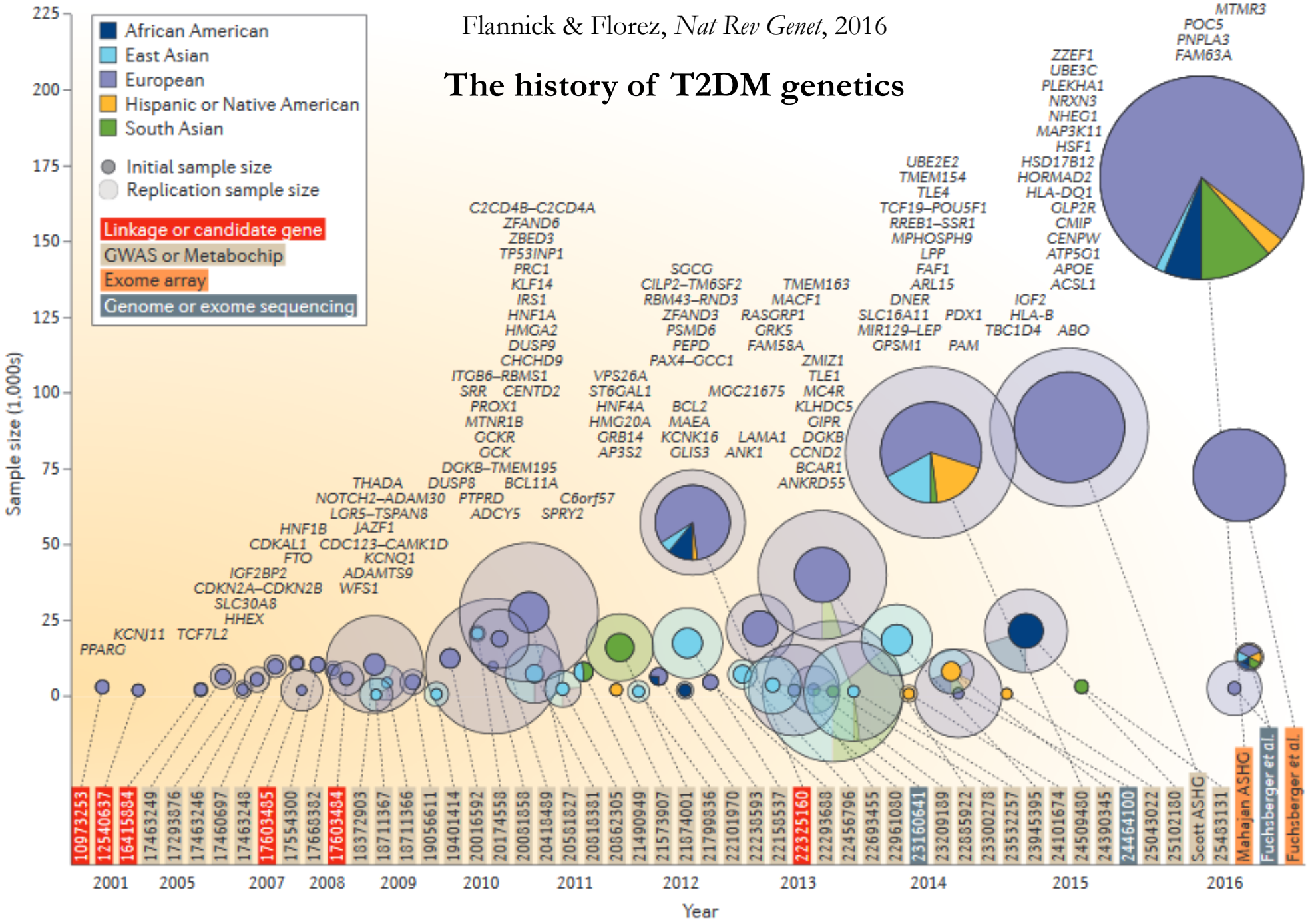
Signal	Chromosome	Representative SNP	Risk allele frequency	Effect size	How found	Hypothecated biology
PPARG	3	rs1801282	0.85	1.23	Candidate	Adipocyte differentiation and function
KCNJ11	11	rs5219	0.40	1.15	Candidate	β-Cell K_{ATP} channel
TCF7L2	10	rs7901695	0.40	1.37	Large-scale association	Incretin signaling in the islet
HHEX	10	rs5015480	0.63	1.13	GWA	Pancreatic development
SLC30A8	8	rs13266634	0.72	1.12	GWA	Zn transport in β-cell insulin granules
FTO	16	rs8050136	0.45	1.23	GWA	Hypothalamic effect on weight regulation
CDKAL1	6	rs10946398	0.36	1.16	GWA	β-Cell function and mass
CDKN2A/B	9	rs10811661	0.86	1.19	GWA	Cell cycle regulation in the β-cell
IGF2BP2	3	rs4402960	0.35	1.11	GWA	mRNA processing in the β-cell
WFS1	4	rs10010131	0.60	1.11	Large-scale association	Endoplasmic reticulum stress
TCF2/HNF1B	17	rs757210	0.43	1.08	Large-scale association	β-Cell development and function
JAZF1	7	rs864745	0.50	1.10	GWA	Transcriptional repression in the islet
CDC123/CAMK1D	10	rs12779790	0.18	1.09	GWA	Cell cycle regulation (*CDC123*)
TSPAN8	12	rs7961581	0.27	1.09	GWA	Cell surface glycoprotein
THADA	2	rs7578597	0.90	1.12	GWA	Apoptosis
ADAMTS9	3	rs4607103	0.76	1.06	GWA	Metalloprotease
NOTCH2	1	rs10923931	0.11	1.11	GWA	Pancreatic development

All loci have been shown to attain significance levels consistent with genome-wide significance in European populations. Note that in most cases the single nucleotide polymorphisms (SNPs) denoted are unlikely to be causal. Effect size is given as the estimated OR per copy of the risk allele. The biological processes listed are based on best available knowledge, but empirical data confirming these are not yet available

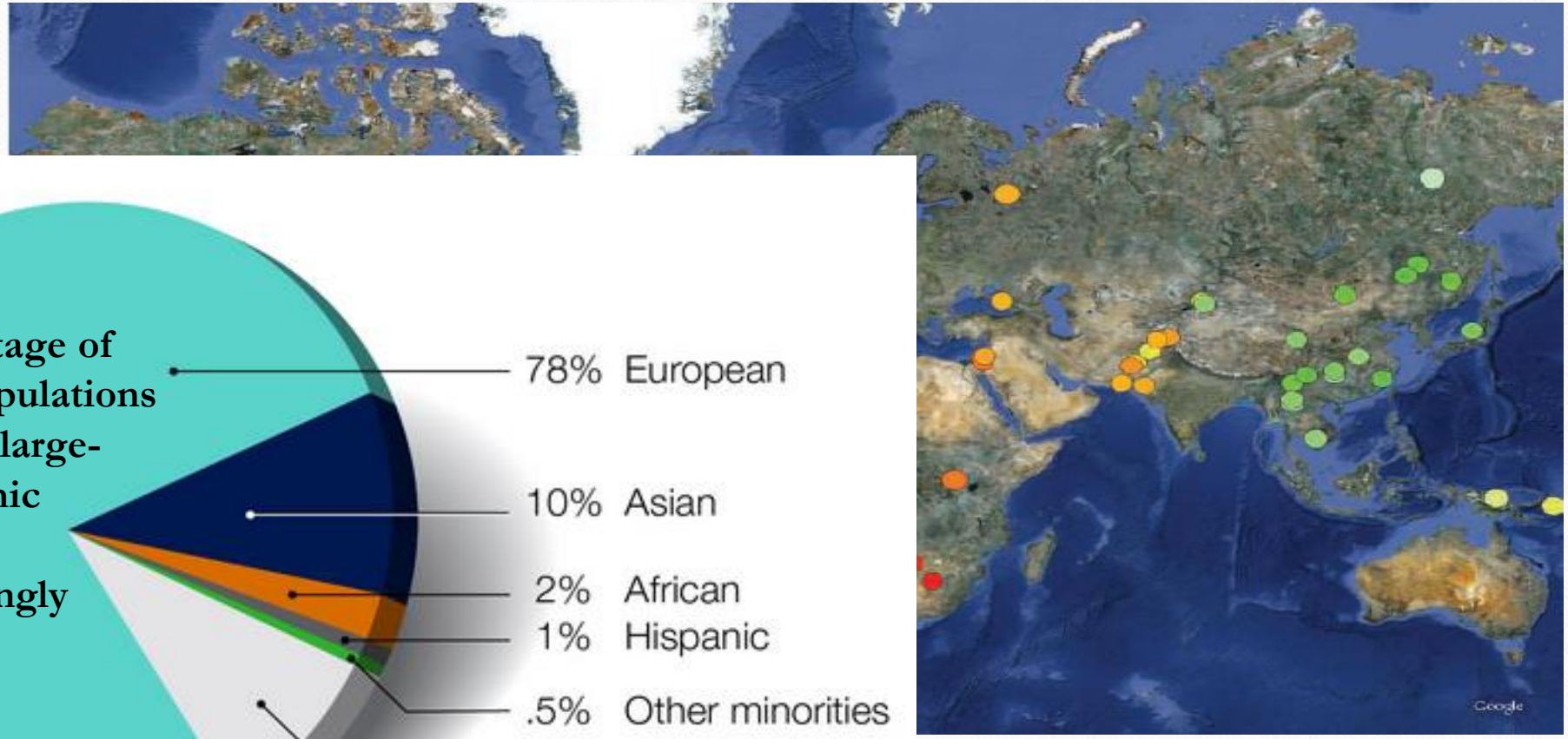
McCarthy & Hattersley, *Diabetes*, 2008

Flannick & Florez, *Nat Rev Genet*, 2016

The history of T2DM genetics



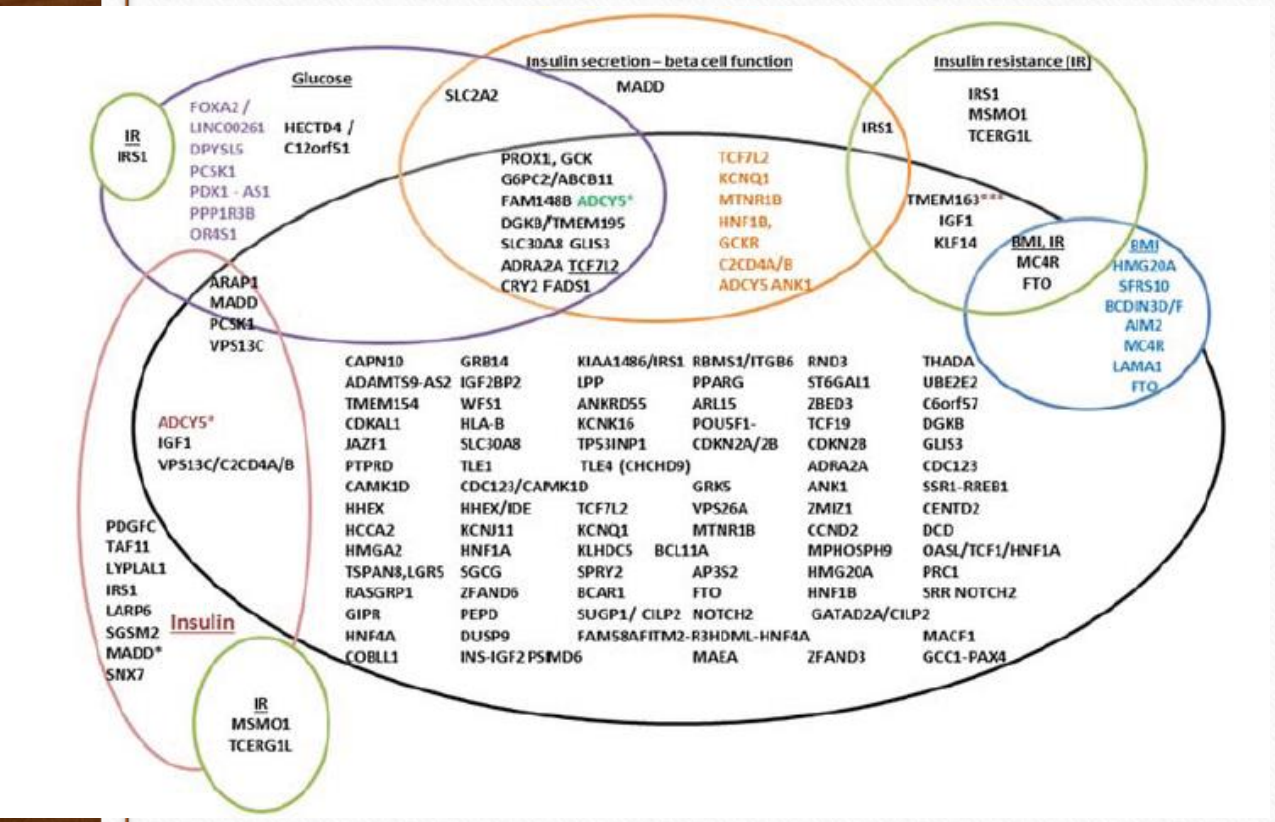
Alleles associated with complex disorders differ in frequency around the world

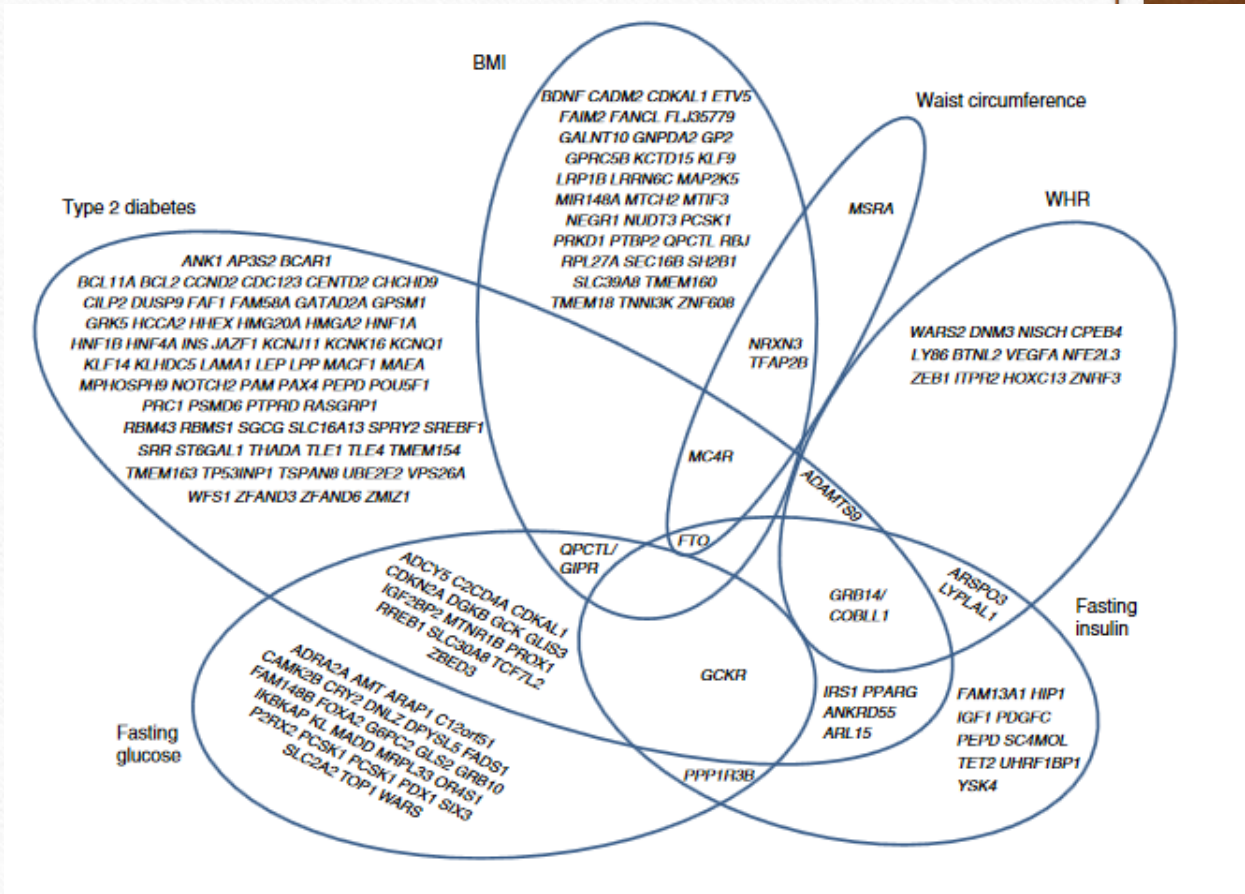


https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores#four

Corona *et al.*, 12th International Congress on Human Genetics, 2011

T2DM and glycemic traits – The GWAS era





Prasad & Groop, *Genes*, 2015

Grarup *et al.*, *Diabetologia*, 2014

From GWAS hits to novel therapeutic targets

Translational success story following up on much controversy

- A common missense variant in *SLC30A8* (p.W325R) associates with T2D risk and glucose levels with great statistical significance ($p<5\times10^{-8}$) (Sladek *et al., Nature*, 2007).
- *SLC30A8* encodes a pancreatic islet Zn^{2+} transporter (ZnT8), highly expressed in the membrane of insulin granules within pancreatic β-cells, where it transports zinc ions for crystallization and storage of insulin.
- Flannick *et al.* reported 12 rare, loss-of-function(LoF) *SLC30A8* alleles (haploinsufficiency) with a protective role against T2DM in humans, which collectively explain a 65% reduction in diabetes risk (Flannick *et al., Nat Genet*, 2014).
- *SLC30A8* LoF variants and T2D risk, across multiple ethnic backgrounds → a 50% reduction in gene dosage protects against T2D in humans!

Table 1 Association of *SLC30A8* variants with T2D

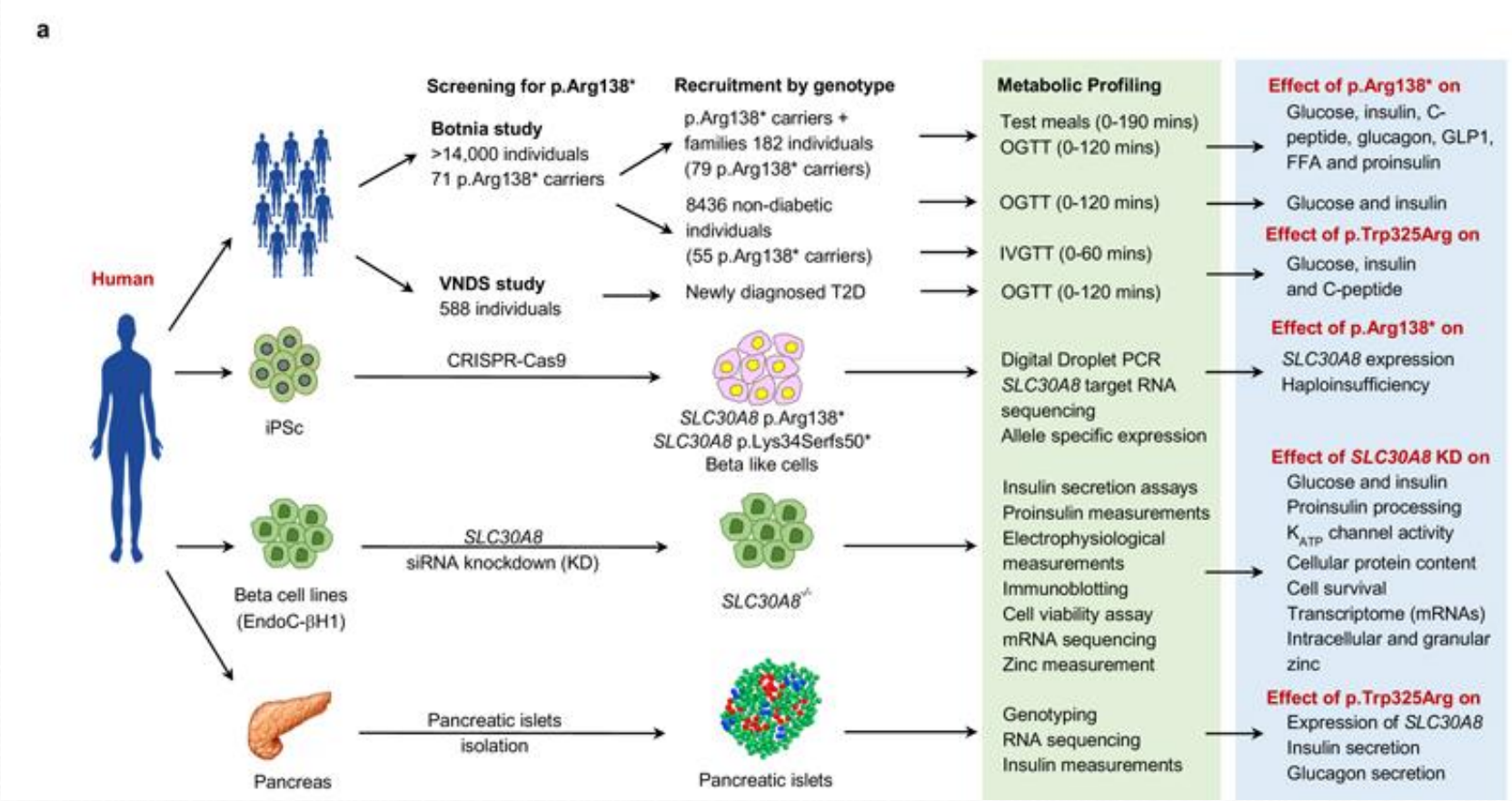
				N		Carriers		Allele frequency		OR	
Variant	Ancestry	Country	Cohort	Cases	Controls	Cases	Controls	Cases (%)	Controls (%)	(95% CI)	*P*
p.Arg138*	European	Finland	Botnia	3,727	5,440	9	39	0.12	0.36	0.47 (0.27–0.81)	0.0067
	European	Sweden	Malmo	6,960	5,480	2	3	0.014	0.027		
	European	Sweden	PIVUS/ULSAM	270	1,734	1	3	0.19	0.087		
	European	Denmark	Danish	3,889	7,869	0	9	0.0	0.057		
	European	Finland	Finnish	4,050	8,696	1	2	0.012	0.011		
	South Asian	Singapore	Singapore Indians	562	585	1	1	0.089	0.085		
	European	UK	UKT2D	321	319	0	1	0.0	0.16		
p.Lys34Serfs*50	European	Iceland	deCODE	2,953	67,919	2	248	0.034	0.18	0.17 (0.05–0.52)	0.0019
	European	Norway	HUNT2	1,645	4,069	0	3	0.0	0.037		
c.71+2T>A	African American	United States	WFS	501	527	1	0	0.1	0.0	0.30 (0.14–0.64)	0.0021
	African American	United States	JHS	530	533	0	1	0.0	0.094		
p.Met50Ile	European	Germany	KORA	97	91	0	1	0.0	0.55		
c.271+G>A	East Asian	Korea	KARE	520	551	0	1	0.0	0.091		
	South Asian	Singapore	Singapore Indians	562	585	0	1	0.0	0.085		
c.419–1G>C	South Asian	UK	LOLIPOP	530	537	1	0	0.094	0.0		
p.Trp152*	European	Finland	Botnia	134	180	0	1	0.0	0.28		
p.Gln174*	South Asian	UK	LOLIPOP	530	537	1	5	0.094	0.47		
c.572+1G>A	African American	United States	JHS	530	533	0	1	0.0	0.094		
p.Tyr284*	South Asian	UK	LOLIPOP	530	537	0	2	0.0	0.19		
	South Asian	Singapore	Singapore Indians	562	585	0	1	0.0	0.085		
p.Ile291Phefs*2	African American	United States	JHS	530	533	0	1	0.0	0.094		
p.Ser327Thrs*55	African American	United States	WFS	501	527	0	2	0.0	0.19		
Combined	–	–	–	30,433	118,701	19	326	–	–	0.34 (0.21–0.53)	1.7×10^{-6}

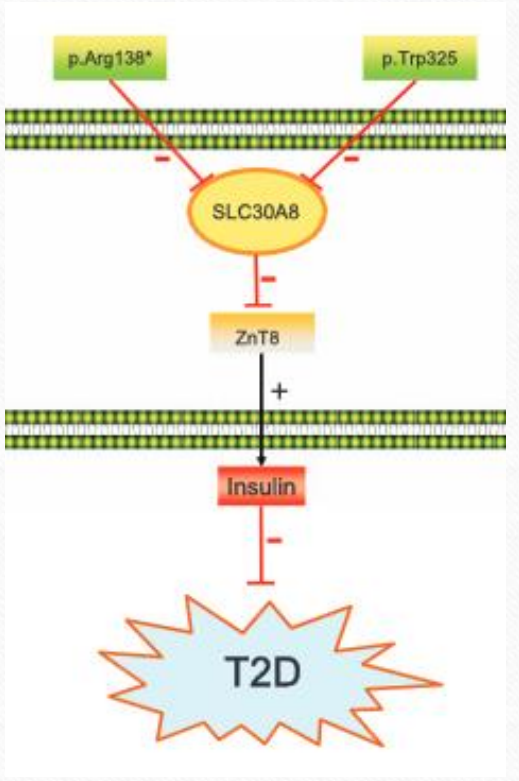
- The data suggested ZnT8 inhibition as a therapeutic strategy in T2D prevention.

Flannick *et al., Nat Genet*, 2014

From GWAS hits to novel therapeutic targets

Translational success story following up on much controversy



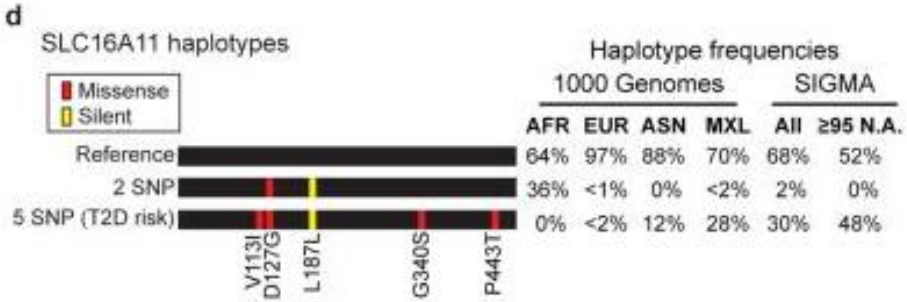


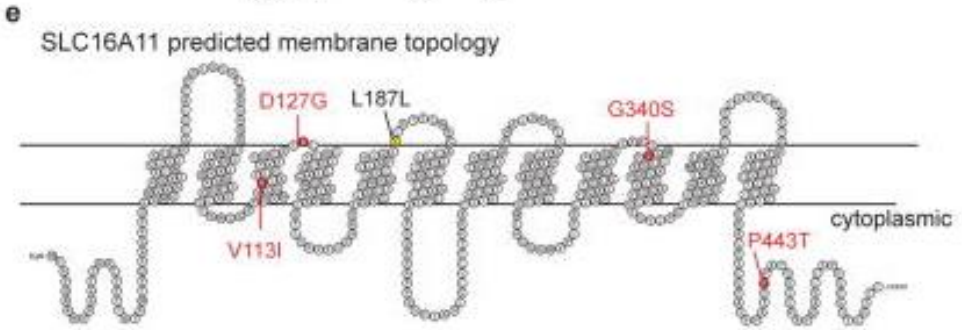
- Heterozygosity for the LoF allele p.Arg138* and homozygosity for the common allele p.W325W of *SLC30A8* are associated with enhanced glucose-stimulated insulin secretion, combined with enhanced proinsulin conversion, as a potential explanation for T2D protection.

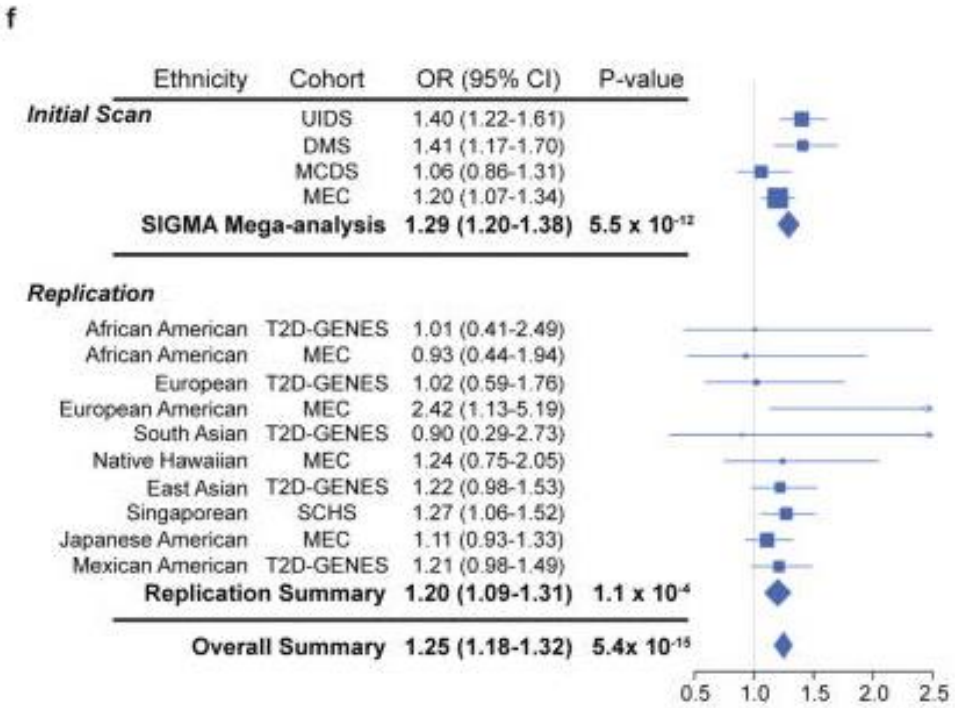
Dwivedi *et al., Nat Genet*, 2019

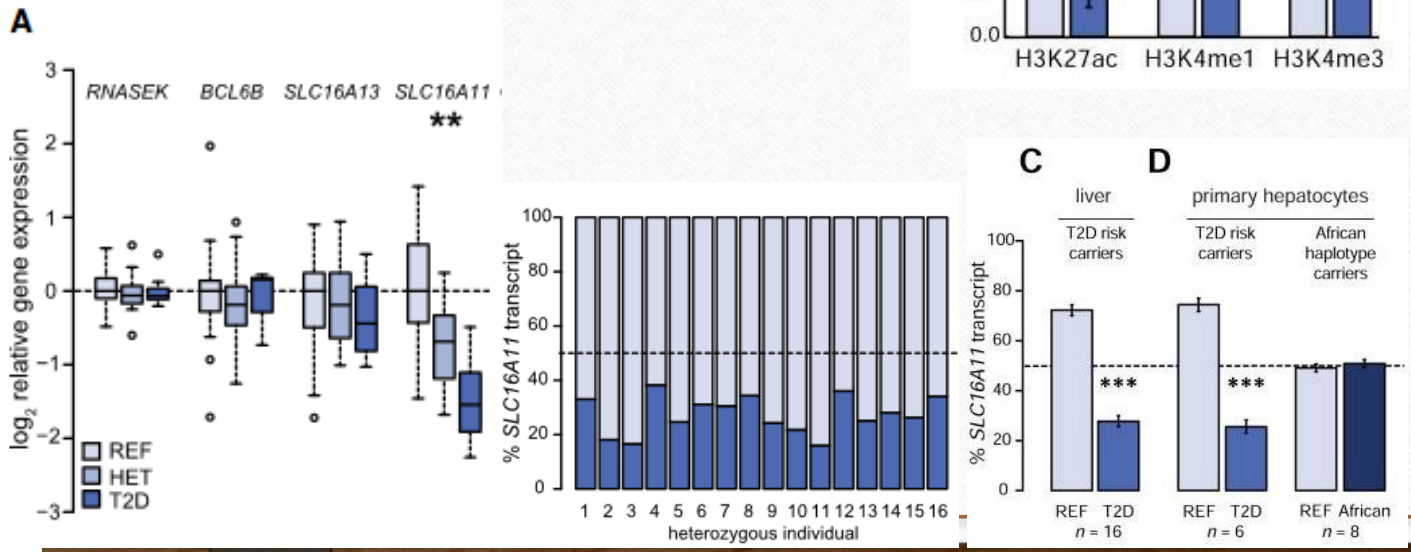
From a GWAS hit to a novel therapeutic target

- 9.2M SNPs in 8,214 Mexicans and other Latin Americans: 3,848 with T2D and 4,366 non-diabetic controls
- Each haplotype copy is associated with a ~20% increased risk of T2D, expected to contribute to the higher burden of T2D in Mexican and Latin American populations
- The T2D-risk haplotype contains a cis-eQTL for lower *SLC16A11* expression in liver









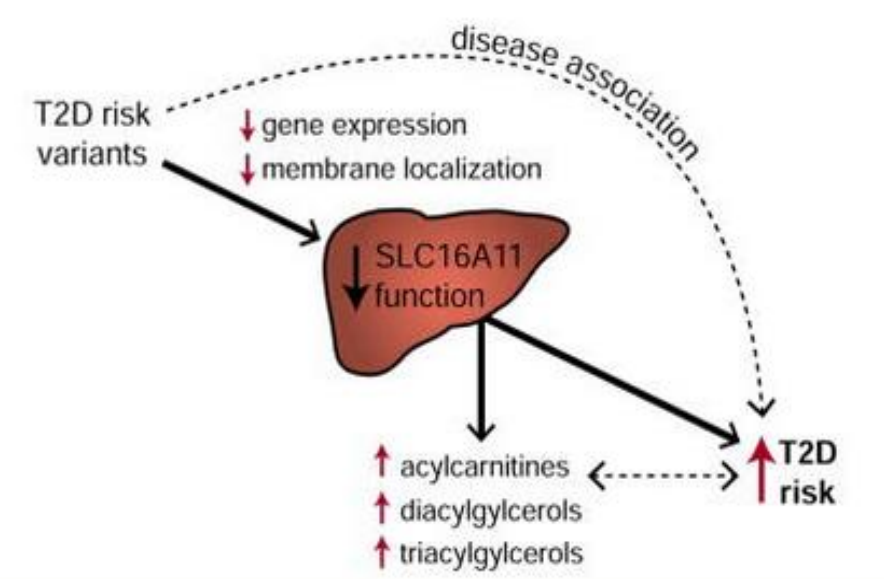
The SIGMA Type 2 Diabetes Consortium; Williams *et al., Nature*, 2014

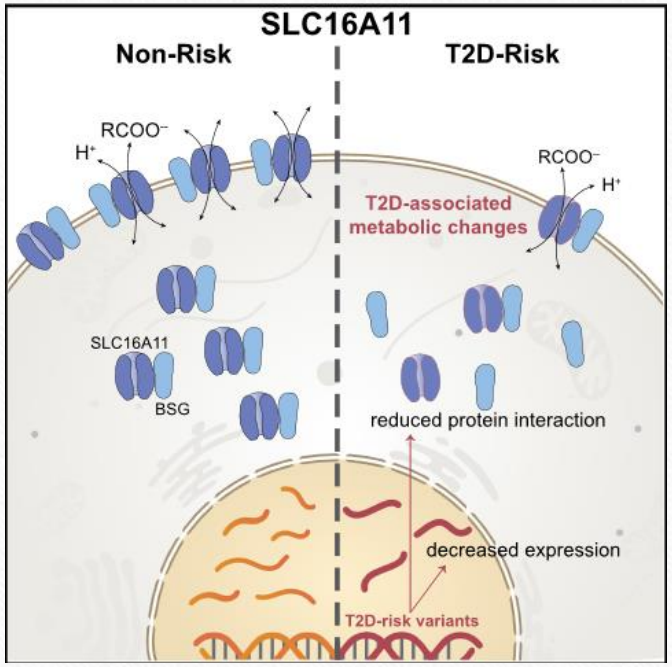
Rusu, Hoch *et al., Cell*, 2017

From a GWAS hit to a novel therapeutic target

Table 1. Categorization of SLC16 family members

Category	Family member	Primary substrates	Mechanism	Ancillary proteins
I	SLC16A1 SLC16A3 SLC16A7	Pyruvate, Lactate, Ketone bodies	H^+-coupled	Basigin (BSG) Embigin (EMB)
	SLC16A8	Lactate		
II	SLC16A2	T3, T4 hormones	Facilitated diffusion	No interaction
	SLC16A10	Aromatic amino acids		-
Uncategorized	SLC16A6	β-hydroxybutyrate	-	-
	SLC16A9	Carnitine	Not H^+-coupled	
	SLC16A4 SLC16A5 **SLC16A11** SLC16A12 SLC16A13 SLC16A14	-	-	





SLC16A11 may influence diabetes risk through effects on lipid metabolism in the liver

T2D-risk variants disrupt a SLC16A11-BSG interaction and cell-surface localization

Reduced SLC16A11 induces metabolic changes associated with increased T2D risk

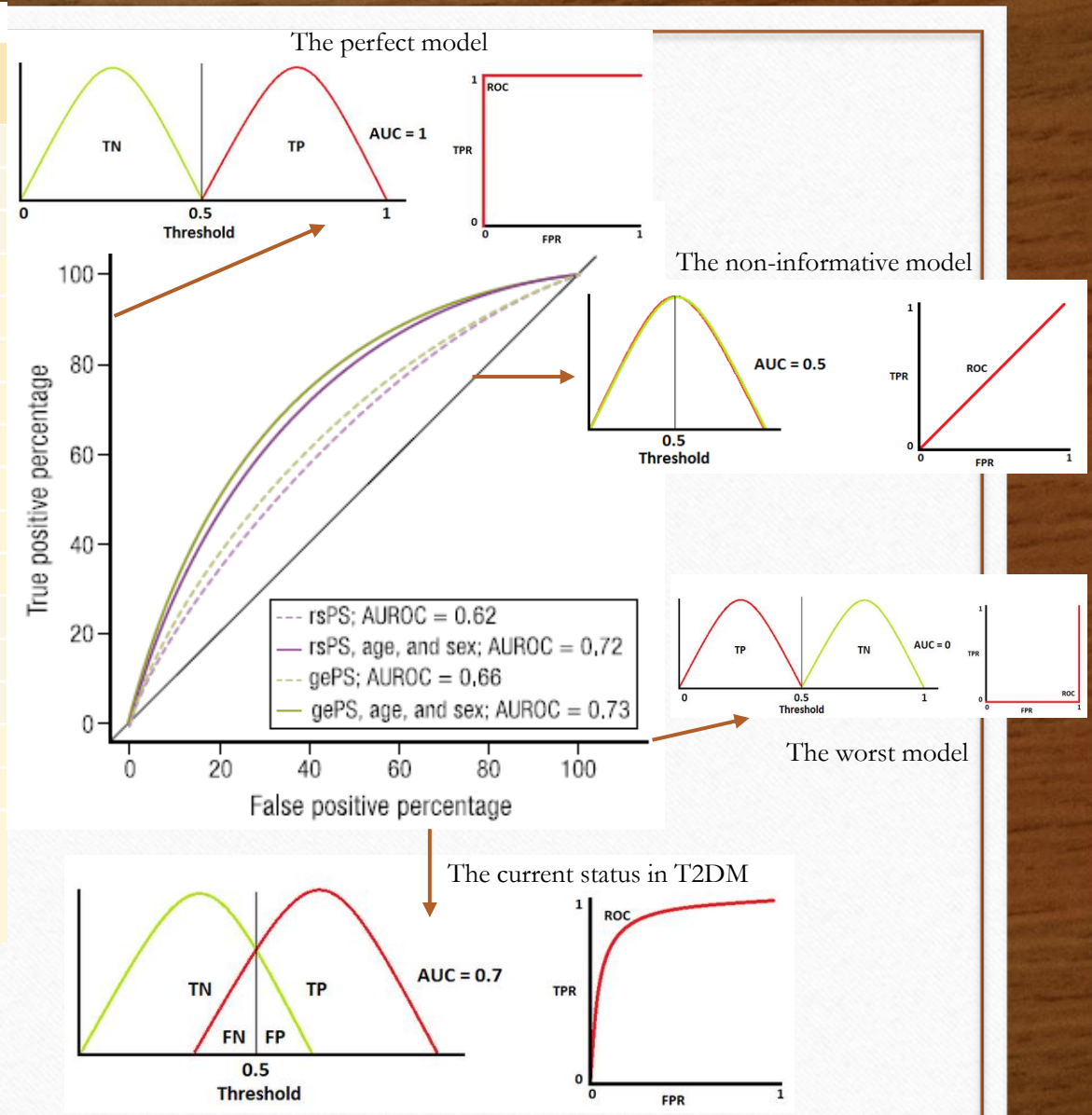
Therapeutics that enhance SLC16A11 levels or activity may be beneficial for T2D

Rusu, Hoch *et al., Cell*, 2017

PRSs and diabetes

Table 1. Comparison of Three Published Global, Extended Polygenic Scores for T2D

		Study		
		Khera *et al.*, 2018 (13)	Mahajan *et al.*, 2018 (9)	23andMe (43)
Discovery GWAS	Number of cases	26,676	55,005	80,792
	Number of controls	132,532	400,308	1,479,116
	Reference	Scott *et al.*, 2017 (44)	Mahajan *et al.*, 2018 (9)[a]	Multhaup *et al.*, 2019 (43)
Optimization data set	Methods	LDpred	Pruning and thresholding	Predetermined cutoffs
	Number of cases	2785	5639	48,028
	Number of controls	120,280	112,307	893,692
	P value threshold	—	0.1	1×10^{-5}
	LD pruning threshold	—	$r^2 > 0.6$	50-kb window
	Tuning parameter	$\rho = 0.01$	—	—
	Polymorphisms in risk score	6,917,436	171,249	1244
	Reference	UK Biobank	UK Biobank	23andMe[b]
Testing data set	Number of cases	5853	13,480	9008
	Number of controls	288,978	311,390	167,622
	Reference	UK Biobank	UK Biobank	23andMe
AUROC in testing data set (Europeans)	Not adjusted for age and sex[c]	0.64[d]	0.66	0.65
	Adjusted for age and sex	0.73	0.73	—
	OR of top 5% bin vs remainder population	2.75	2.75 without age and sex adjustment	2.76[d]
			4.52 with age and sex adjustment	

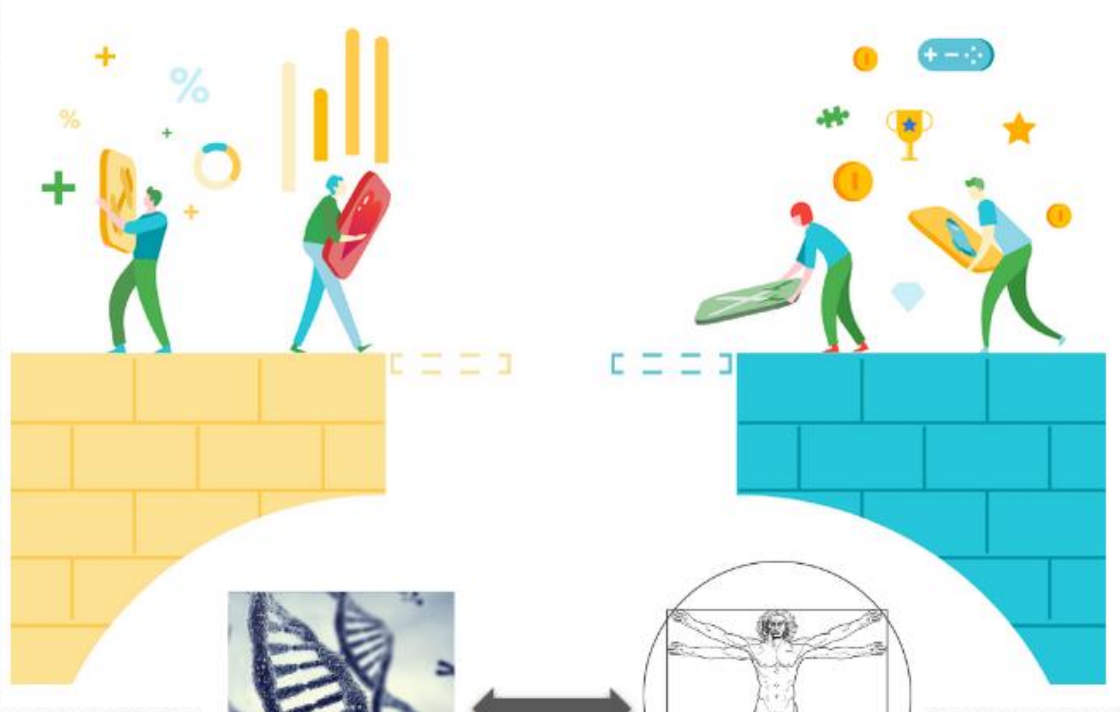


If the estimates of relative risk seen in UK Biobank participants in recent studies generalize to the population level, then there are likely to be >1M individuals in the UK, who, on the basis of their PRS alone, have a lifetime risk of T2D!

Udler *et al.*, *Endocr Rev*, 2019

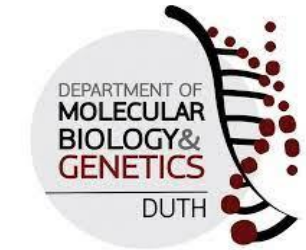
https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

We need to bridge the knowledge gap from sequence to consequence



Genotype to Phenotype

Prospects and limitations in the use of genetic markers in translational research of complex diseases



«ΜΕΤΑΦΡΑΣΤΙΚΗ ΕΡΕΥΝΑ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ»

Μοριακή Διαγνωστική, Βιοδείκτες και Στοχευμένες Θεραπείες

ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ | DEMOCRITUS UNIVERSITY OF THRACE

Μαριάνθη Γεωργίτση

Επίκουρη Καθηγήτρια Γενετικής και Μοριακής Βάσης Ασθενειών

E-mail: mgeorgit@mbg.duth.gr

Alexandroupoli, 20.04.2024