

Μαριάνθη Γεωργίτση, PhD

«Γενετική και Μοριακή Βάση Ασθενειών»

Εργαστήριο Γονιδιωματικής Ποικιλότητας και Γενετικής Επιδημιολογίας



mgeorgit@mbg.duth.gr

GWAS Genome-wide Association Studies

Μελέτες συσχέτισης
γονιδιωματικής κλίμακας

ΠΕΡΙΕΧΟΜΕΝΑ

- Εισαγωγή – Complex traits (Πολυπαραγοντικά γνωρίσματα)
- Προς την ανάπτυξη τεχνικών γονιδιωματικής κλίμακας
- GWAS basics (concept, workflow, analyses)
- GWAS και Πολυγονιδιακοί Δείκτες Κινδύνου [Polygenic Risk Scores (PRSs)] – Πιθανοί βιοδείκτες?

Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond

Niels Grarup • Camilla H. Sandholt • Torben Hansen •
Oluf Pedersen

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

Missing heritability: the discrepancy between h^2 estimated in twin studies and the summed variation explained by genome-wide significant variants. h^2_{SNP} , which includes the influence of variants not reaching genome-wide significance, accounts for some missing heritability. While the cause of ‘still missing’ heritability ($h^2 - h^2_{\text{SNP}}$) remains unclear, it is likely attributable to (mostly rare) variants that are poorly tagged by common SNPs on arrays.

Glossary
<i>1000 Genomes Project</i> The 1000 Genomes Project, launched in January 2008, is an international research effort to establish a detailed catalogue of human genetic variation. Scientists planned to sequence the genomes of 2,500 participants from a number of different ethnic groups
<i>Allele</i> One of a number of alternative forms of the same gene or same genetic locus
<i>De novo mutation</i> An alteration in a gene that is present for the first time in one family member as a result of a mutation in a germ cell (egg or sperm) of one of the parents or in the fertilised egg itself
<i>Epistasis</i> When the effect of one gene depends on the presence of one or more ‘modifier genes’ (genetic background). Also referred to as gene–gene interaction
<i>Exome</i> The protein coding part of the human genome. The exome of the human genome consists of roughly 180,000 exons, constituting about 1% of the total genome, or about 30 megabases of DNA
<i>Heritability</i> The proportion of phenotypic variation of a trait that is due to underlying genetic variation
<i>Imputation</i> In genetics, imputation refers to the statistical inference of unobserved genotypes. It is achieved by using known haplotypes in a reference population, such as the 1000 Genomes Project, thereby allowing non-genotyped genetic variants to be tested for association with a trait of interest
<i>Linkage disequilibrium</i> A non-random association between alleles at different loci
<i>Minor allele frequency</i> Ranging from 0% to 50%, this is the proportion of alleles at a locus that contain the less frequent allele.
<i>Private variants</i> Variants restricted to probands and immediate relatives
<i>Sequencing depth</i> In DNA sequencing, depth refers to the number of times a nucleotide is read during the sequencing process. Deep sequencing indicates that the depth of the process is many times larger than the length of the sequence under study

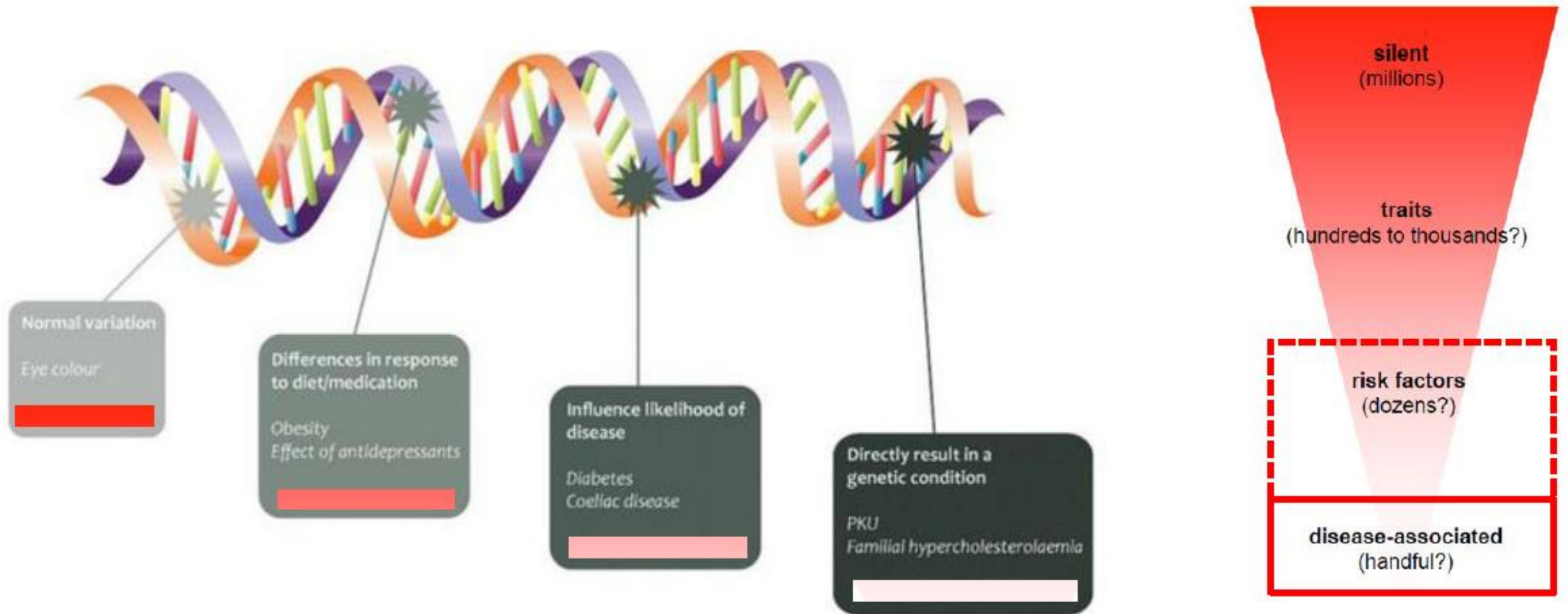
GWAS Genome-wide Association Studies

Μελέτες συσχέτισης
γονιδιωματικής κλίμακας

ΠΕΡΙΕΧΟΜΕΝΑ

- Εισαγωγή – Complex traits (Πολυπαραγοντικά γνωρίσματα)
- Προς την ανάπτυξη τεχνικών γονιδιωματικής κλίμακας
- GWAS basics (concept, workflow, analyses)
- GWAS και Πολυγονιδιακοί Δείκτες Κινδύνου [Polygenic Risk Scores (PRSs)] – Πιθανοί βιοδείκτες?

Η ποικιλομορφία του ανθρώπινου γονιδιώματος



Adapted and modified from:

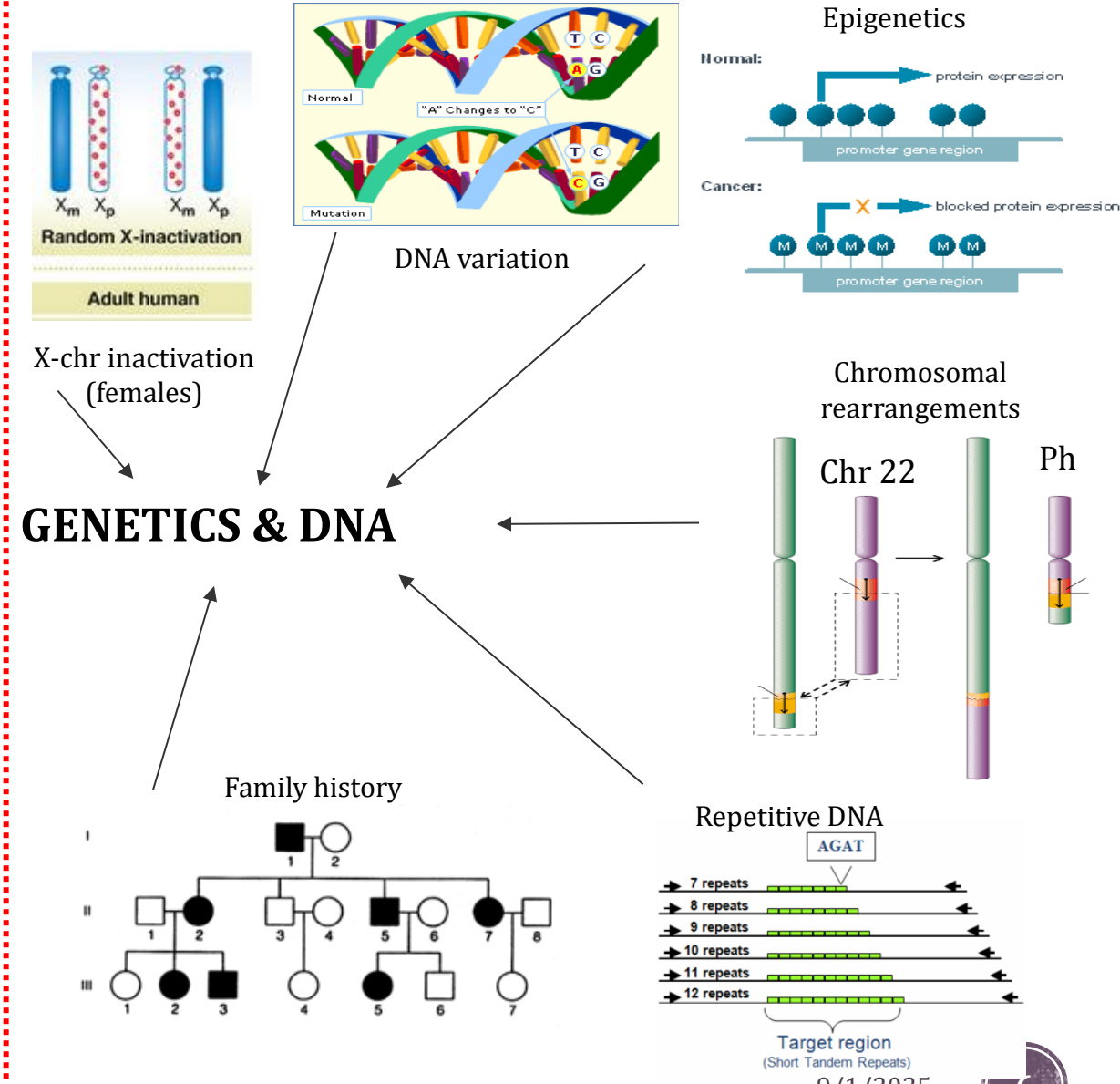
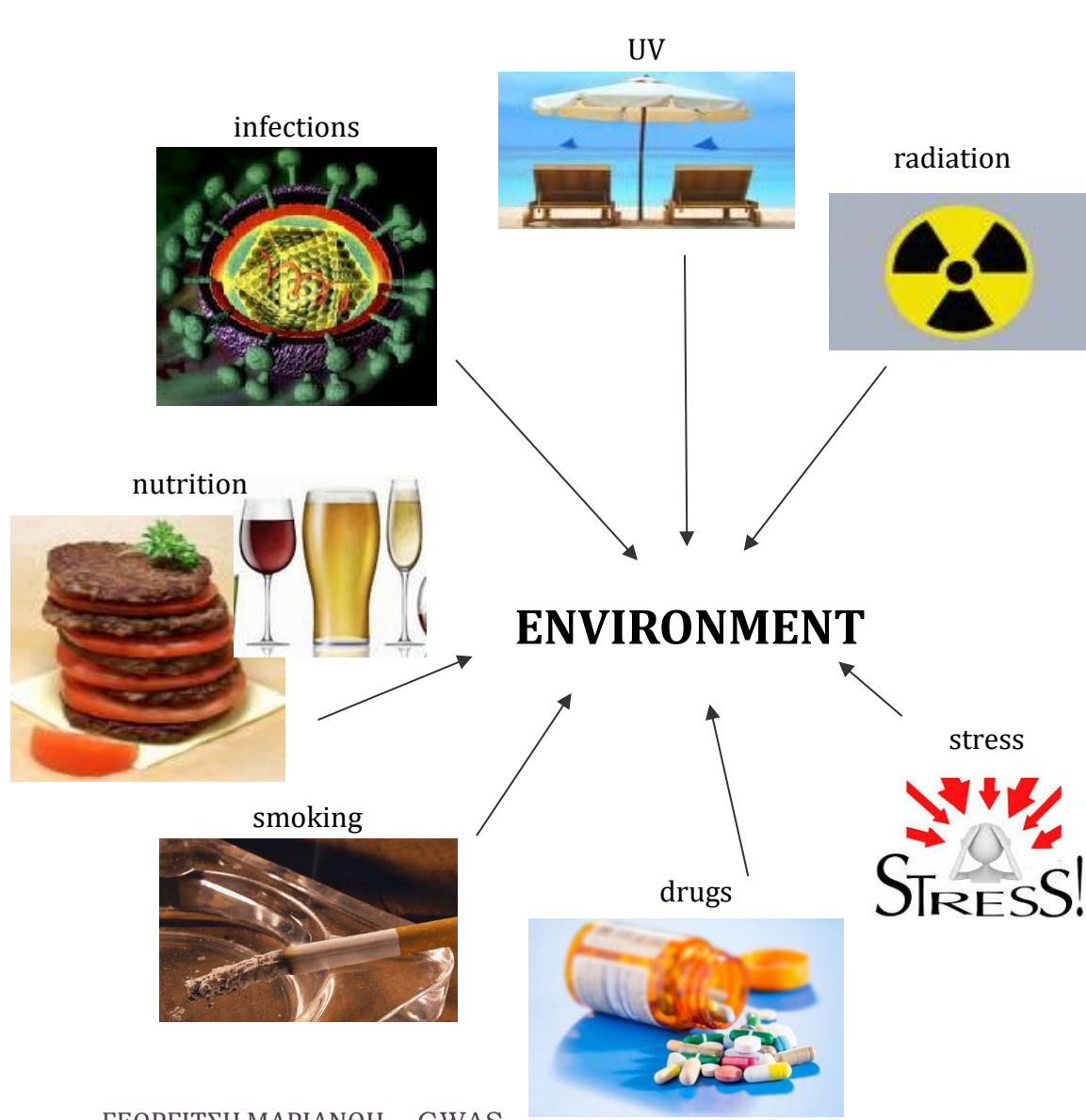
@genomicsedu

www.genomicseducation.hee.nhs.uk

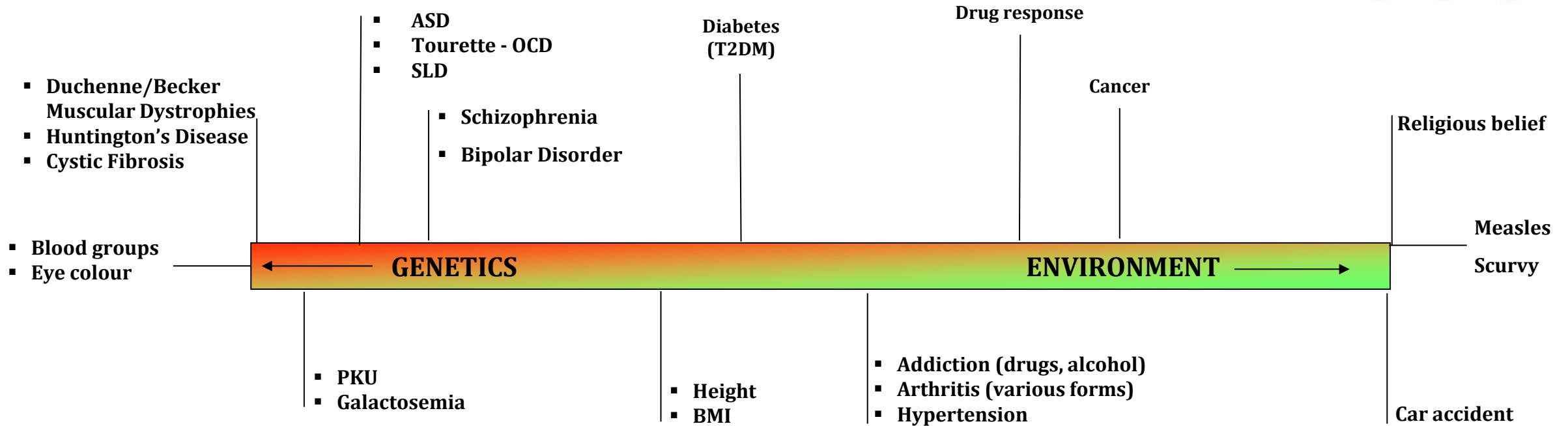
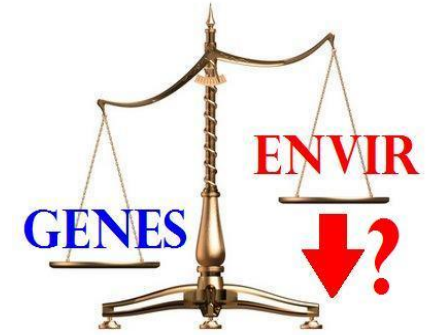
/genomicsedu

Ποιες παραλλαγές είναι σημαντικές για την υγεία του ατόμου?

Πολυπαραγοντικά γνωρίσματα – Complex traits



Γενετική προδιάθεση & φαινοτυπικοί χαρακτήρες



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

Κληρονομισιμότητα πολυπαραγοντικών γνωρισμάτων

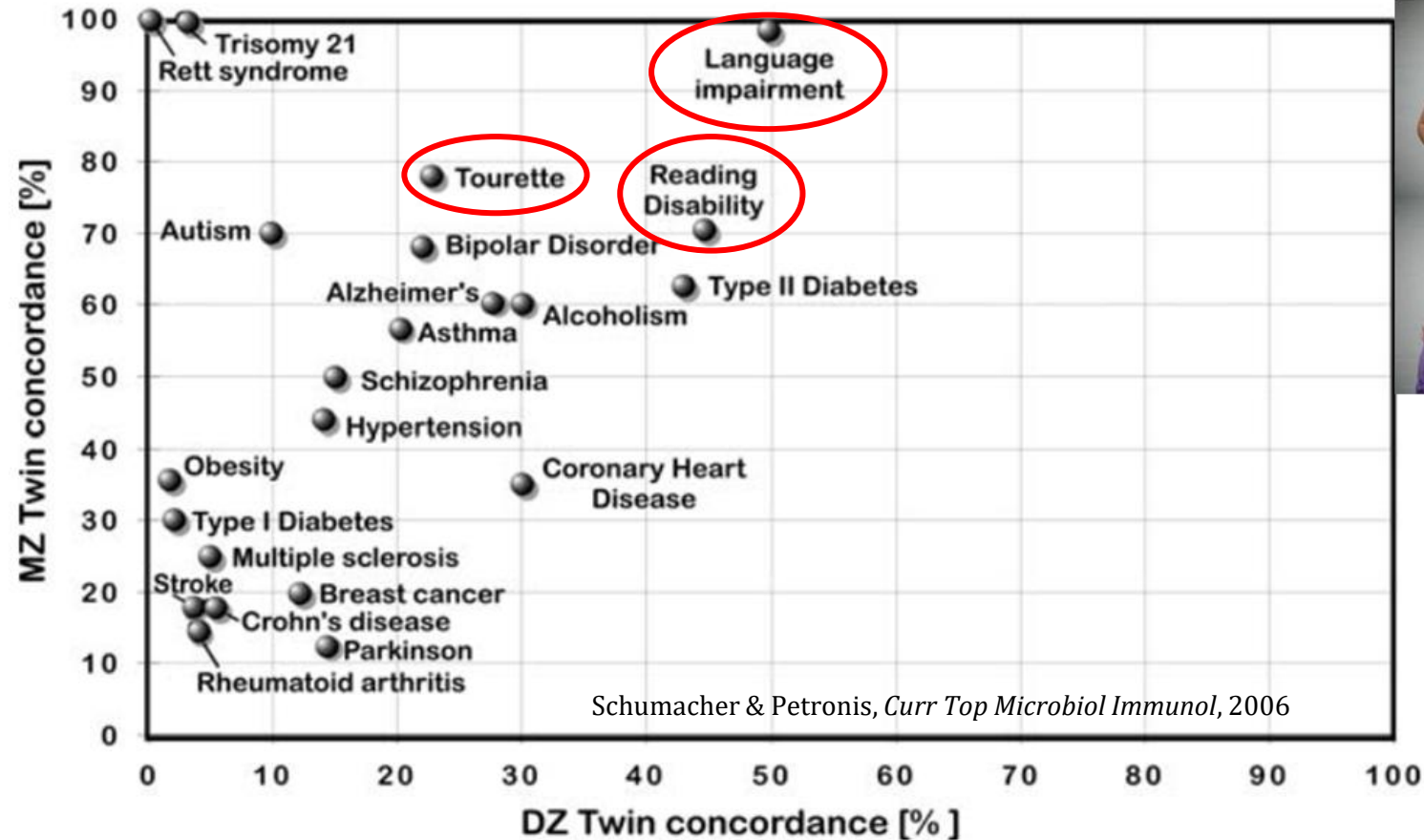


Fig. 1 Concordance of MZ and DZ twins for different disorders. As a rule, the degree of concordance in MZ twins is lower than 100% for nearly all complex diseases but substantially higher in comparison to the concordance rate in DZ twins

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)

$$P = G + E$$

$$V_P = V_G + V_E + V_{G \times E}$$

$$V_P = V_A + \underbrace{V_D + V_I}_{\text{Non-Additive Genetic Variance}} + V_E$$

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

$$V_G = V_A + V_D + V_I$$

$$H^2 = V_G / V_P$$

Broad-sense
heritability

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

Narrow-sense
heritability

- 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

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

Κληρονομησιμότητα πολυπαραγοντικών γνωρισμάτων

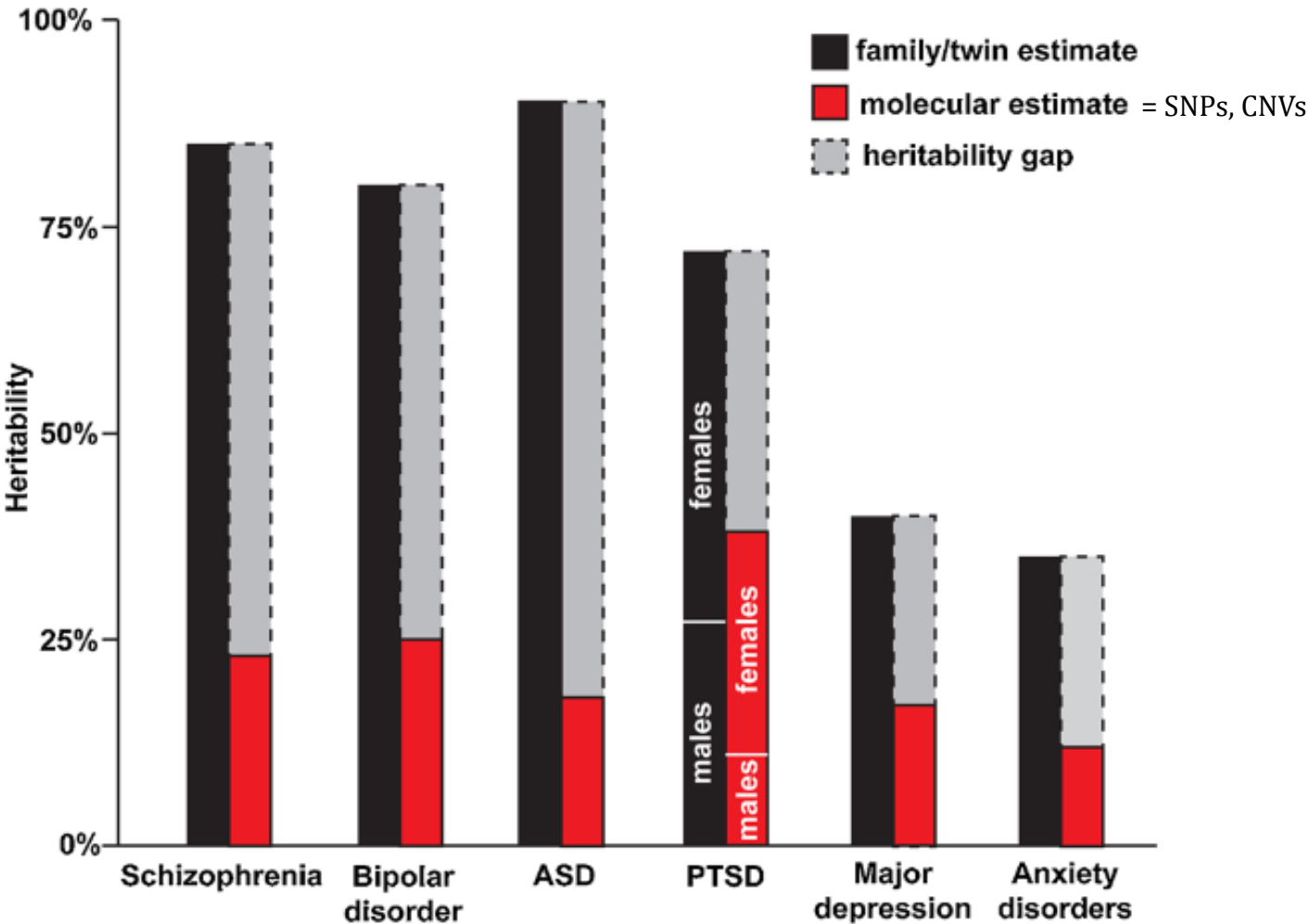
$$h^2 = \frac{V_A}{V_P}$$
 Narrow-sense heritability

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

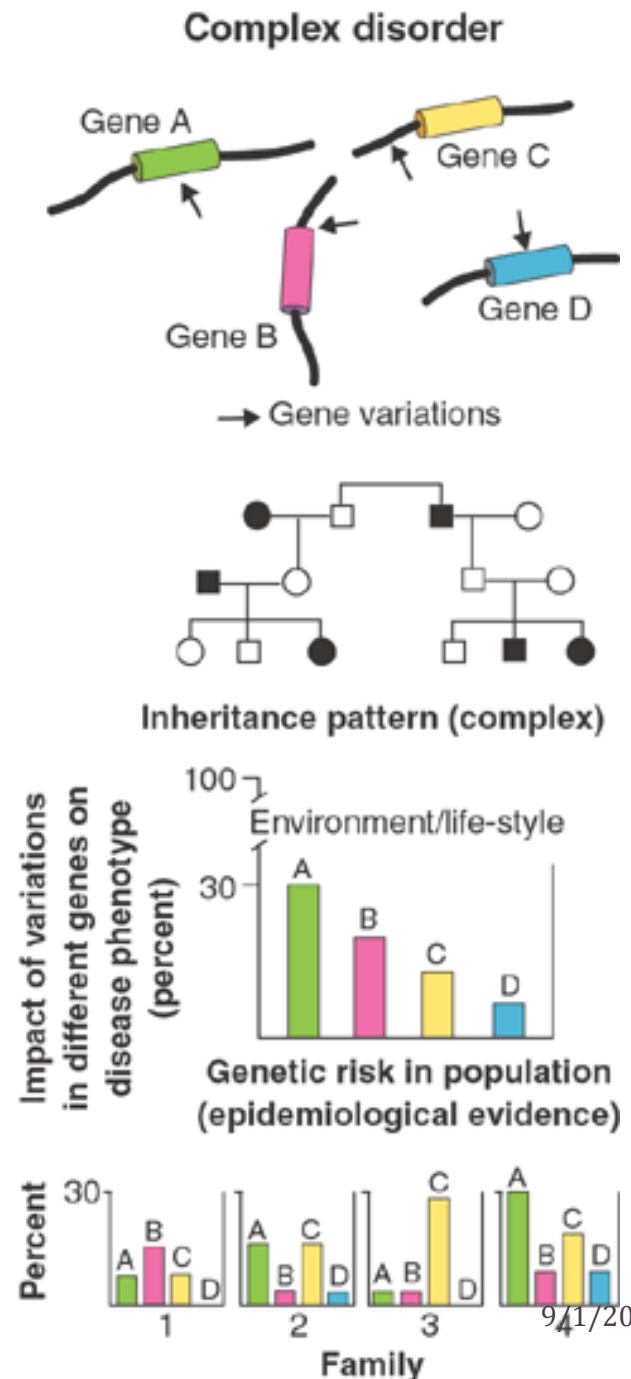
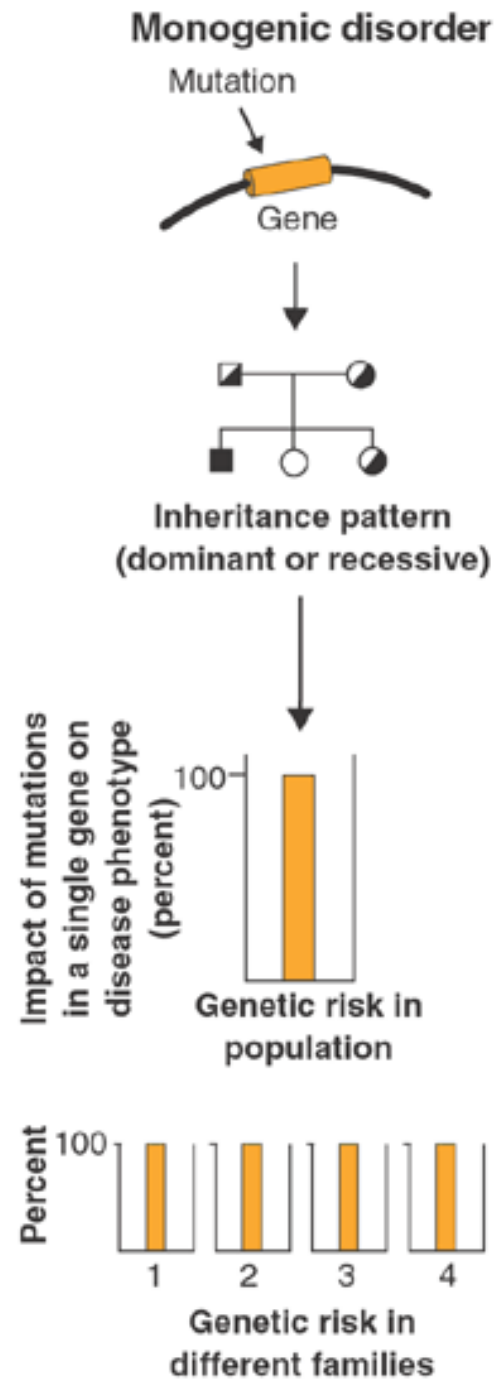


Monogenic

versus

Complex

traits/disorders



Single-gene disorders

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

Multifactorial disorders (Complex traits)

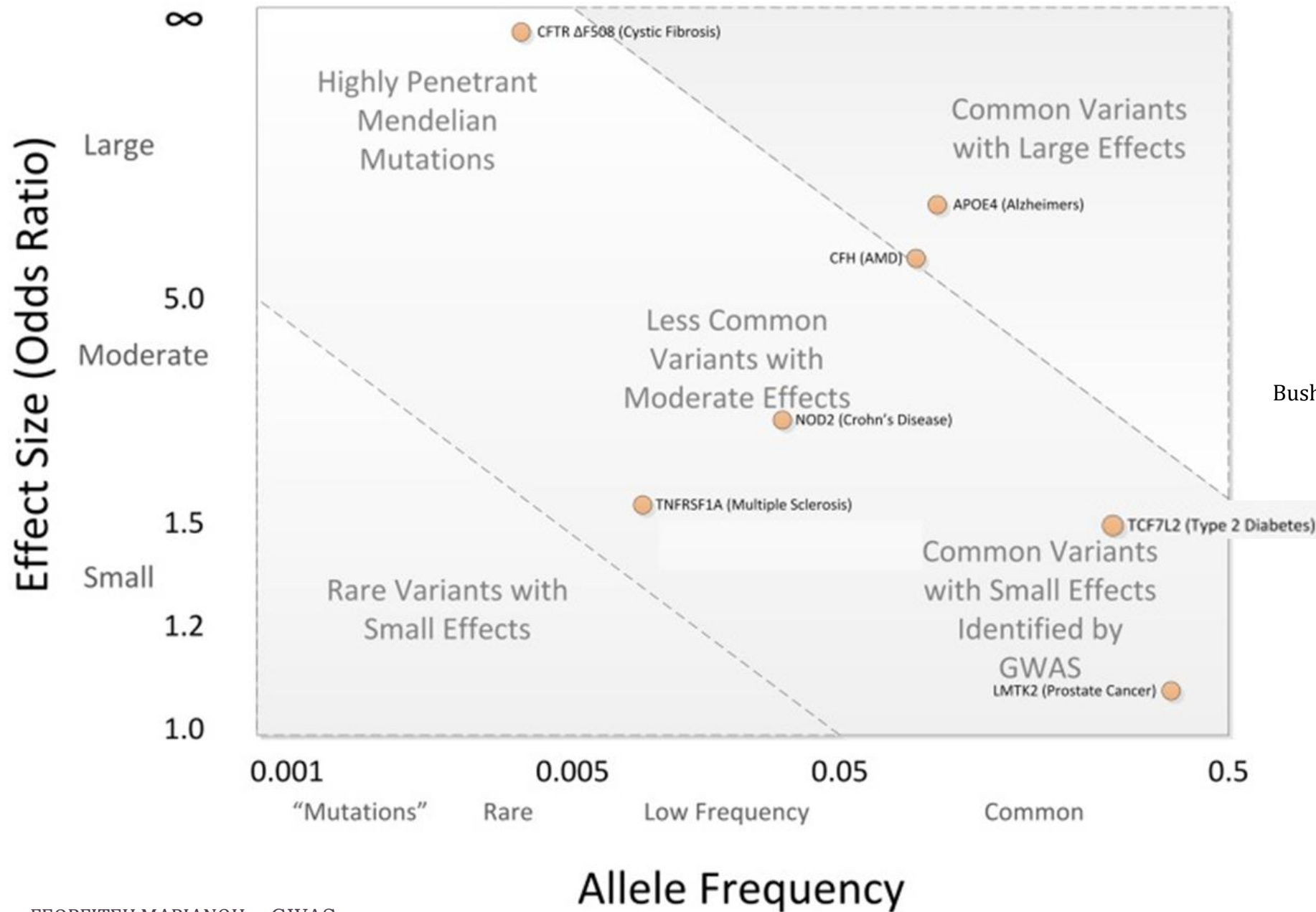
- Serious impact on public health cost
- Multiple genes and loci
- Complex pattern of inheritance (additive)
- Variable heritability (h^2)
- Common and rare genetic variants
- Whole-genome scans → new technologies/analytical tools
- Examples:
 - Stroke/CVD
 - Diabetes (Type 2)
 - Schizophrenia/Bipolar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

GWAS Genome-wide Association Studies

Μελέτες συσχέτισης
γονιδιωματικής κλίμακας

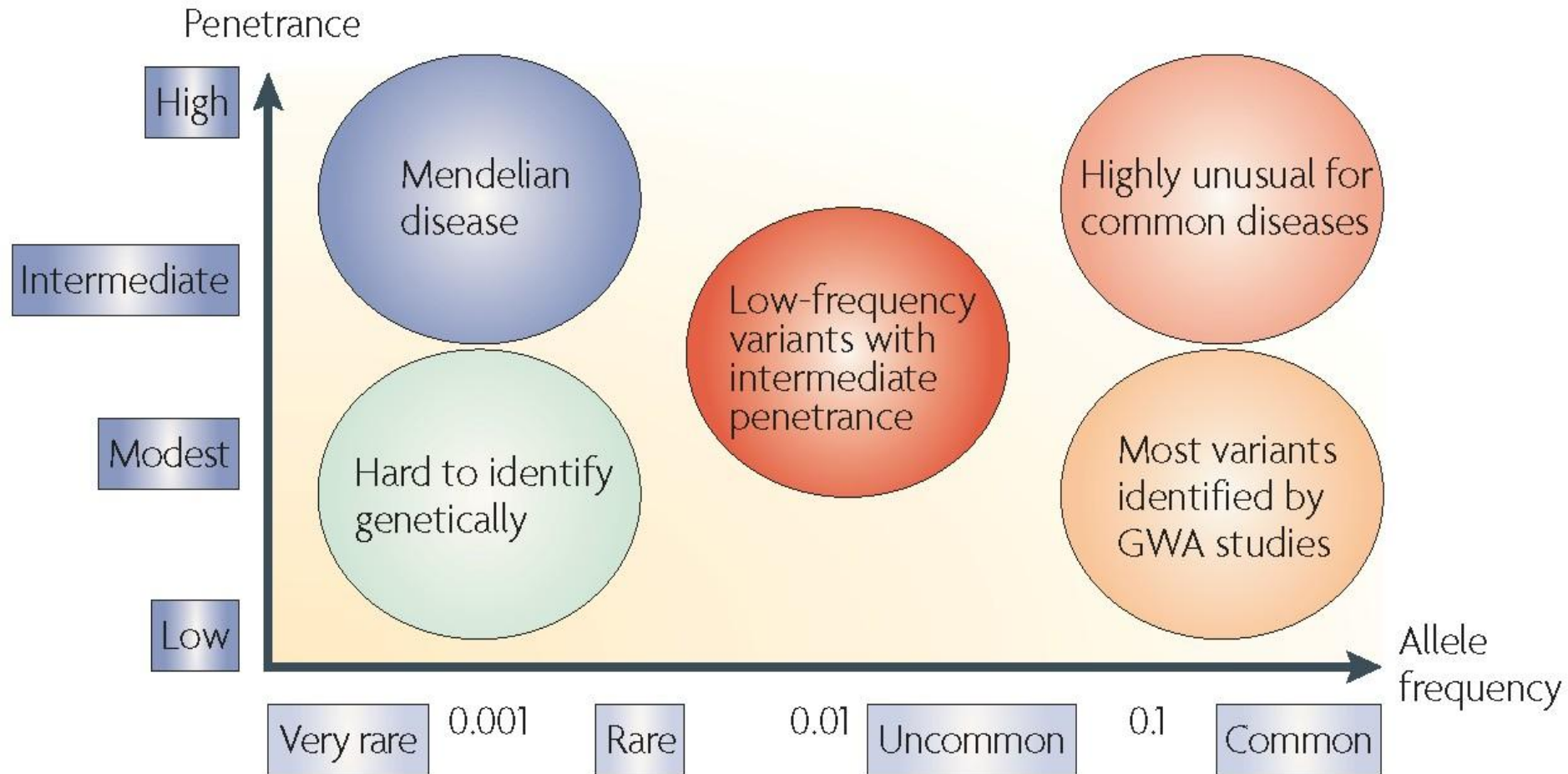
ΠΕΡΙΕΧΟΜΕΝΑ

- Εισαγωγή – Complex traits (Πολυπαραγοντικά γνωρίσματα)
- Προς την ανάπτυξη τεχνικών γονιδιωματικής κλίμακας
- GWAS basics (concept, workflow, analyses)
- GWAS και Πολυγονιδιακοί Δείκτες Κινδύνου [Polygenic Risk Scores (PRSs)] – Πιθανοί βιοδείκτες?



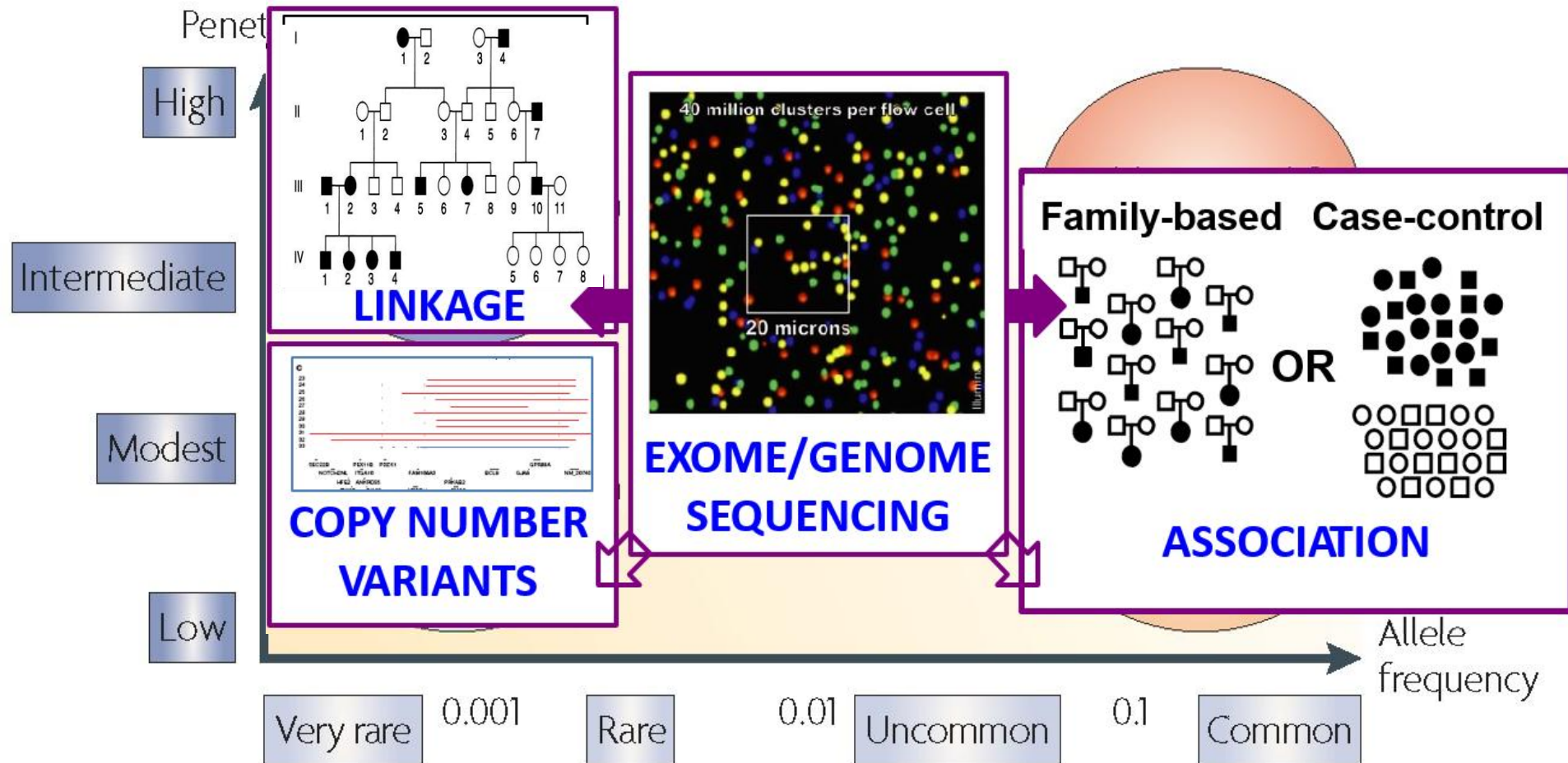
Bush & Moore, *PLoS Comput Biol*, 2012

Η ποικιλομορφία του ανθρώπινου γονιδιώματος



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

Η ποικιλομορφία του ανθρώπινου γονιδιώματος

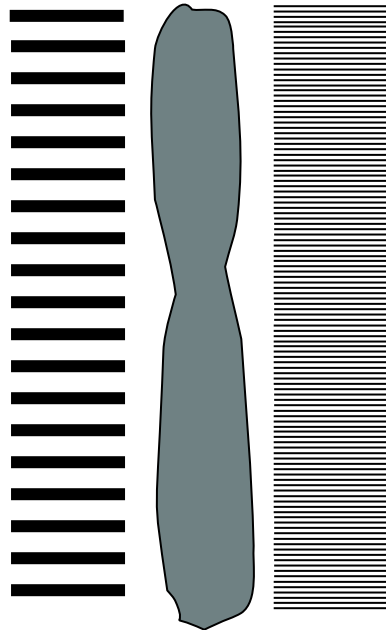


McCarthy *et al.*, *Nat Rev Genet*, 2008

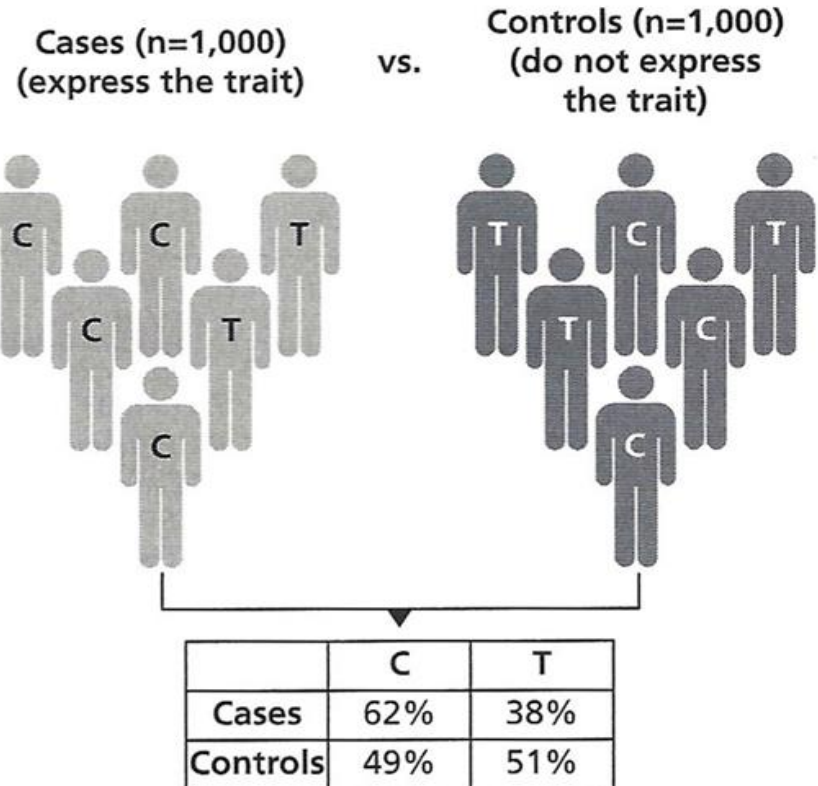
Μελέτες Συσχέτισης (1990 – σήμερα ?)

Σκοπός: Η ανίχνευση συσχετίσεων μεταξύ **γενετικών δεδομένων** (δηλ. αλληλομόρφων ή γονοτύπων) συχνών παραλλαγών του γονιδιώματος με **παραμέτρους** που αφορούν ένα γνώρισμα ή μία κλινική κατάσταση (πχ νόσημα) υπό μελέτη. Αυτό επιτυγχάνεται με χρήση **στατιστικών αναλύσεων** σε ένα ικανοποιητικά **μεγάλο πληθυσμιακό δείγμα** (συνήθης σχεδιασμός «περίπτωση-μάρτυρα»/case-control), προκειμένου να υποστηριχθεί ή να καταρριφθεί ο ισχυρισμός ότι η(οι) συγκεκριμένη(ες) παραλλαγή(ες) συνεισφέρουν στον καθορισμό του γνωρίσματος/νοσήματος.

Microsatellites
(παλαιότερα)



SNPs
(σήμερα)



Genome-wide association studies (2005 - σήμερα)

Περιπτώσεις/Ασθενείς



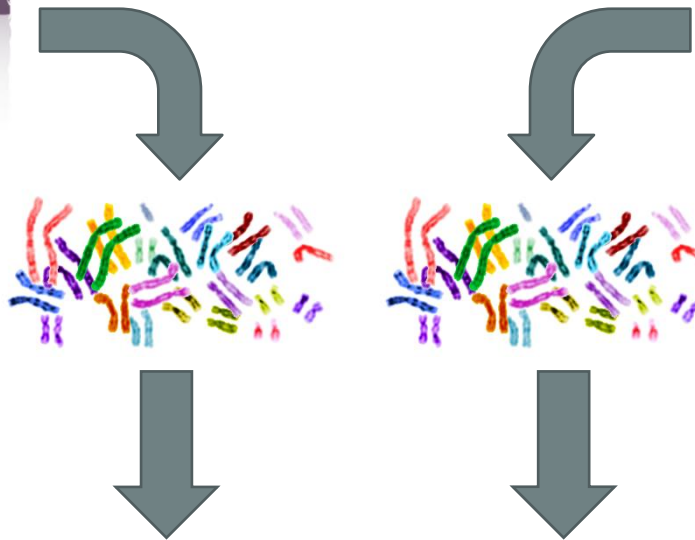
Μη συγγενείς!

Χιλιάδες δειγμάτων –
Όσο περισσότερα τόσο
καλύτερα!

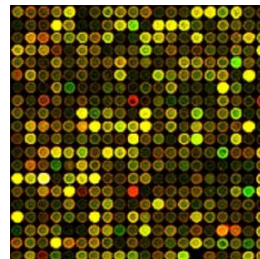
Μάρτυρες



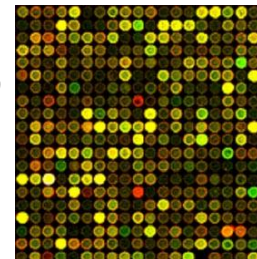
Μη συγγενείς!



- Γονοτύπηση:
100k -> 500k -> 6M SNPs



Σύγκριση συχνοτήτων
αλληλομόρφων ή/και
γονοτύπων –
Στατιστική ανάλυση



- Ερμηνεία συσχετίσεων
- Μετανάλυση

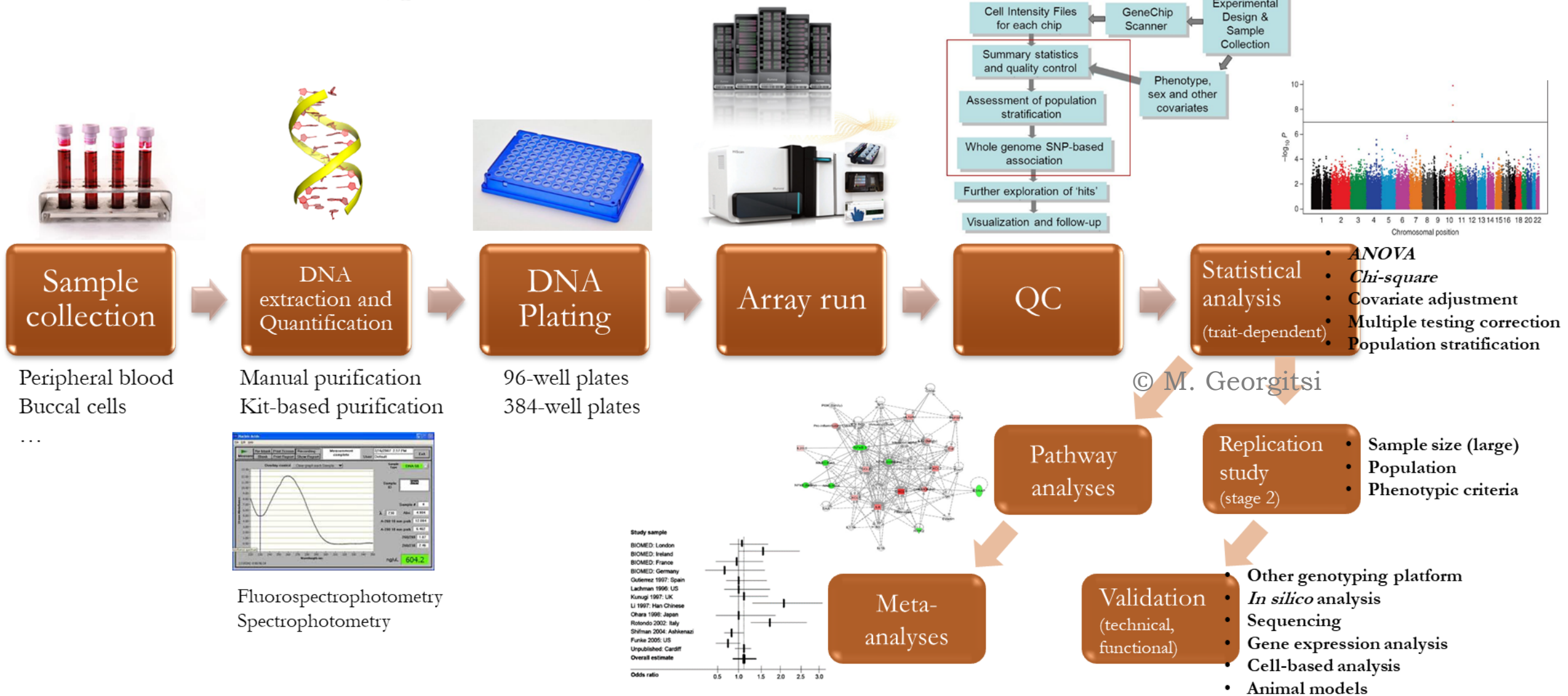
GWAS Genome-wide Association Studies

Μελέτες συσχέτισης
γονιδιωματικής κλίμακας

ΠΕΡΙΕΧΟΜΕΝΑ

- Εισαγωγή – Complex traits (Πολυπαραγοντικά γνωρίσματα)
- Προς την ανάπτυξη τεχνικών γονιδιωματικής κλίμακας
- GWAS basics (concept, workflow, analyses)
- GWAS και Πολυγονιδιακοί Δείκτες Κινδύνου [Polygenic Risk Scores (PRSs)] – Πιθανοί βιοδείκτες?

Workflow of a typical GWAS study



Ανάλυση δεδομένων

Δείκτες που
γονοτυπήθηκαν:

0=AA
1=Aa
2=aa

Συμμετέχοντες
στη μελέτη:

1=cases
0=controls

Y	g ₁	g ₂	g ₃	g ₄	...	g _m
1	2	0	1	0		2
0	0	1	0	0		1
1	0	1	1	0		2
1	1	2	0	0		2
0	0	1	1	0		1
...	...	...	...	...		...
0	1	0	2	0		0

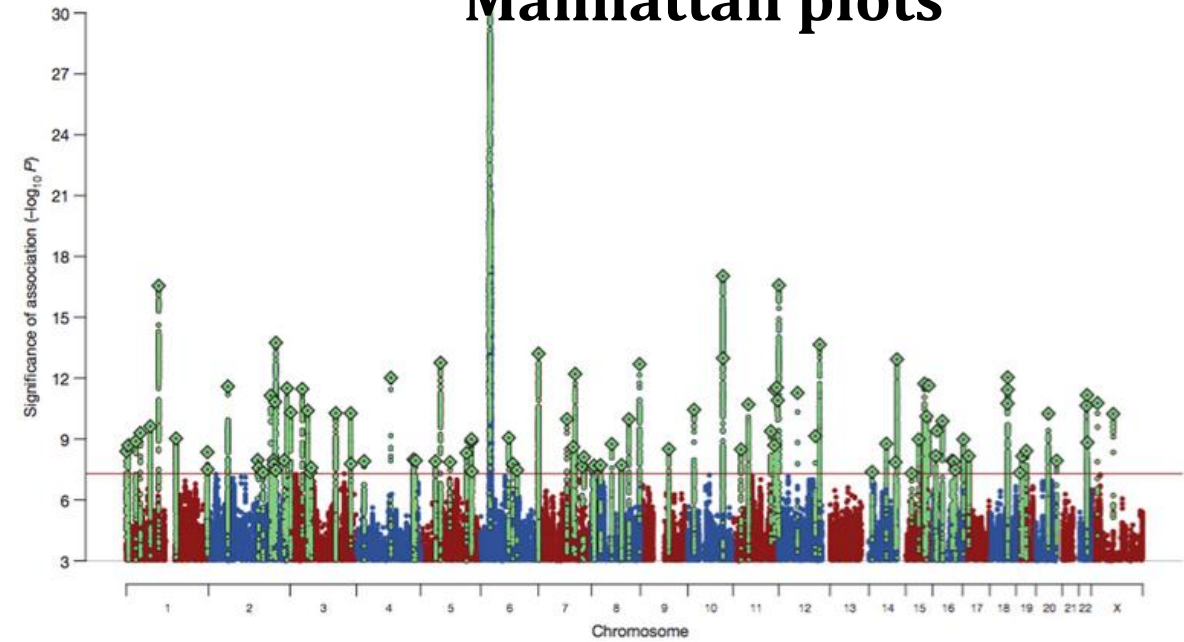
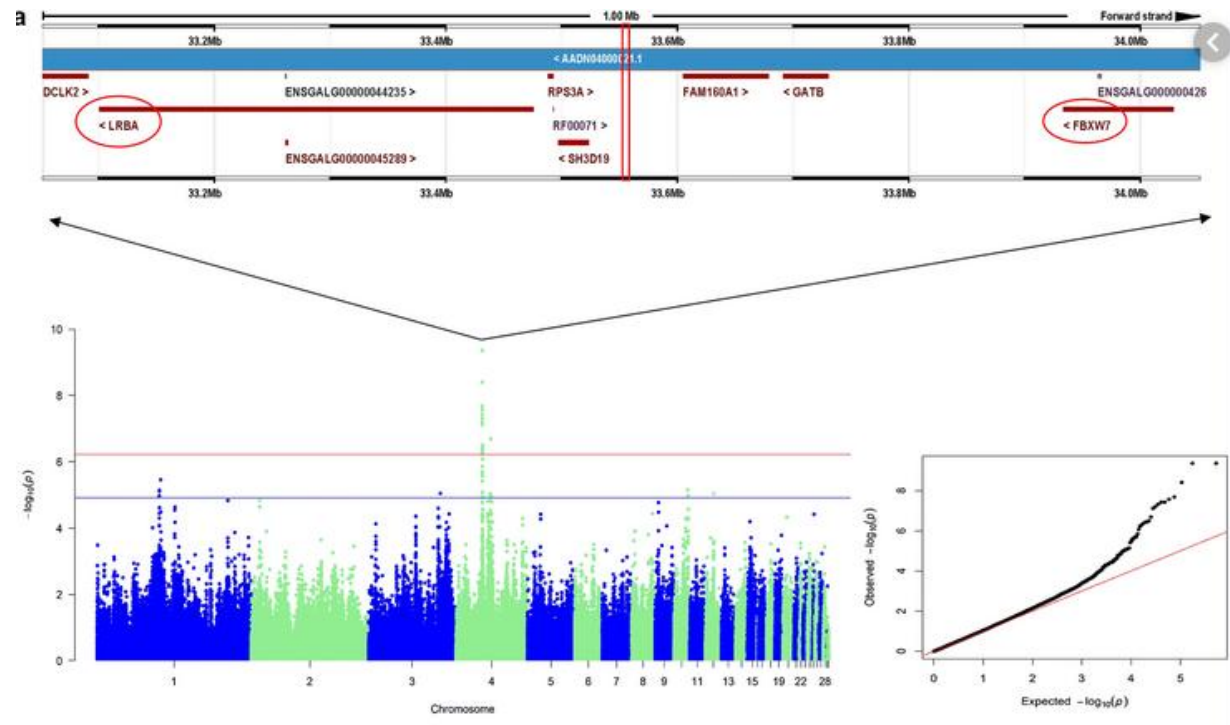
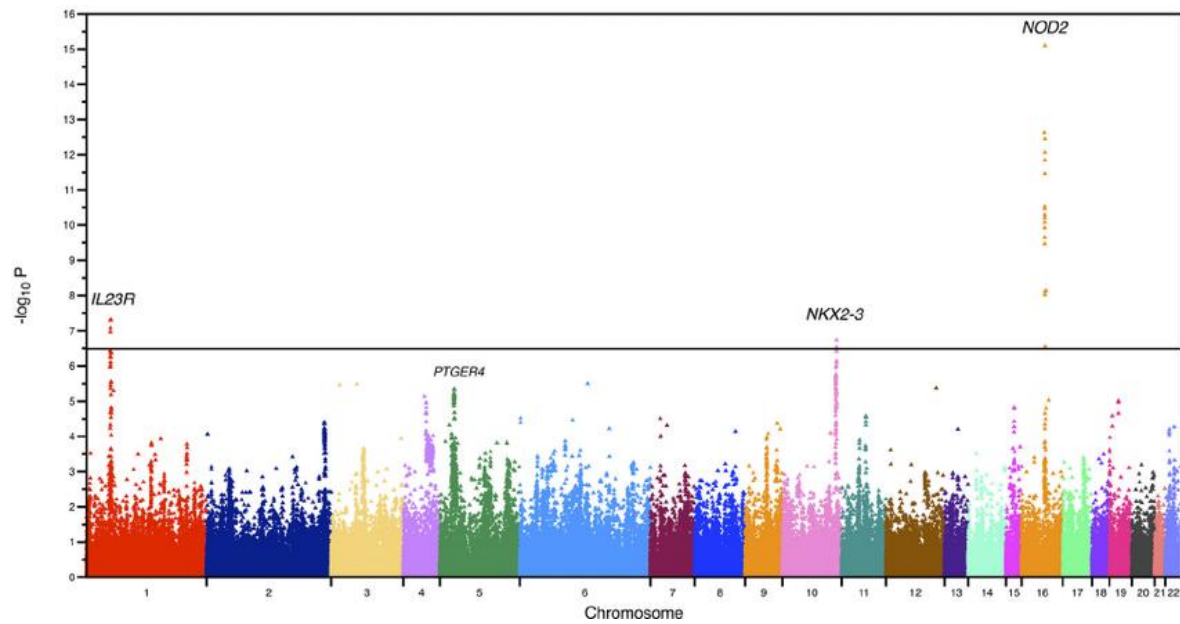
Συμμεταβλητές

z ₁	z ₂	...	z _c
Sex	Age		BMI
1	54		24.5
0	36		23.7
1	72		30.2
0	66		26.0
0	65		27.0
...	...		...
0	55		22.7

- A) Γνωρίσματα ποιοτικά
(διχότομα, διακριτά δεδομένα)
- B) Γνωρίσματα ποσοτικά (συνεχή)

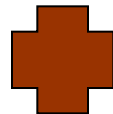
Μεταβλητή 1/ Μεταβλητή 2	Συνεχής	Διακριτή
Συνεχής	Συντελεστής συσχέτισης	t-test Z-test Wilcoxon
Διακριτή	t-test Z-test Wilcoxon	χ ² test Fisher's exact test McNeamar test

Γραφική αναπαράσταση αποτελεσμάτων Manhattan plots



Γενετική Αρχιτεκτονική

Πληθυσμός



Γενετικές παραλλαγές

SNP																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Είδος (ποιες)

Αριθμός (πόσες)

Συχνότητα
(πόσο συχνές)

Αποτέλεσμα
(πόσο σημαντικές)

Linkage
Disequilibrium
(με ποιες άλλες)

Πληθυσμιακή διαστρωμάτωση

Είναι ένας παράγοντας σύγχυσης που μπορεί να οδηγήσει σε **ψευδώς θετικές συσχετίσεις** και λανθασμένα συμπεράσματα, λόγω α) διαφορετικών συχνοτήτων αλληλομόρφων & β) ελαφρώς διαφοροποιημένων φαινοτύπων

Πληθυσμός 1

Population 1

	Allele 1	Allele 2
Affected	50 ($f_{1,Aff}=0.2$)	200
Unaffected	25 ($f_{1,Unaff}=0.2$)	100

$\chi^2 = 0.00$ p-value = 1.0

 **AA**
 **Aa**
 **aa**

Μάρτυρες



Πληθυσμός 2

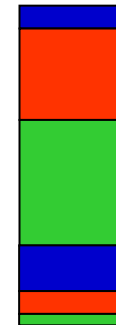


Population 2

	Allele 1	Allele 2
Affected	100 ($f_{1,Aff}=0.8$)	25
Unaffected	200 ($f_{1,Unaff}=0.8$)	50

$\chi^2 = 0.00$ p-value = 1.0

Περιπτώσεις



Combined

	Allele 1	Allele 2
Affected	150 ($f_{1,Aff}=0.4$)	225
Unaffected	225 ($f_{1,Unaff}=0.6$)	150

$\chi^2 = 29.2$ p-value = 6.5×10^{-8}

Πιθανές λύσεις: α) προσεκτικότερος σχεδιασμός (επιλογή πληθυσμών, μαρτύρων),
β) Καλύτερη φαινοτύπιση ασθενών, γ) PCA

Σύνοψη – Περιπτώσεις συσχέτισης

1. Υπαίτια αλληλόμορφα (ιδανικό σενάριο, αλλά εξαιρετικά σπάνιο!) – Bingo!

= Τα αλληλόμορφα που ανιχνεύονται ως στατιστικά σημαντικά συσχετιζόμενα δημιουργούν προδιάθεση ή προσφέρουν προστασία έναντι της εκδήλωσης κάποιου φαινοτύπου (αυξάνουν ή μειώνουν άμεσα τον κίνδυνο)

2. Αλληλόμορφα σε LD με υπαίτια αλληλόμορφα (τυπικό σενάριο, οι περισσότερες περιπτώσεις) – Χρήσιμο για περαιτέρω μελέτες

= Τα αλληλόμορφα που ανιχνεύονται ως στατιστικά σημαντικά συσχετιζόμενα στην πραγματικότητα είναι συνδεδεμένα, δηλαδή συγκληρονομούνται με τα πραγματικά υπαίτια αλληλόμορφα (που μάλλον παραμένουν άγνωστα και χρήζουν περαιτέρω ανίχνευσης)

3. Πληθυσμιακή διαστρωμάτωση

= Τα αλληλόμορφα που ανιχνεύονται ως στατιστικά σημαντικά είναι ψευδώς θετικά συσχετιζόμενα με τον υπό μελέτη φαινότυπο (γνώρισμα, νόσημα, κλπ). → Τα αποτελέσματα είναι παραπλανητικά

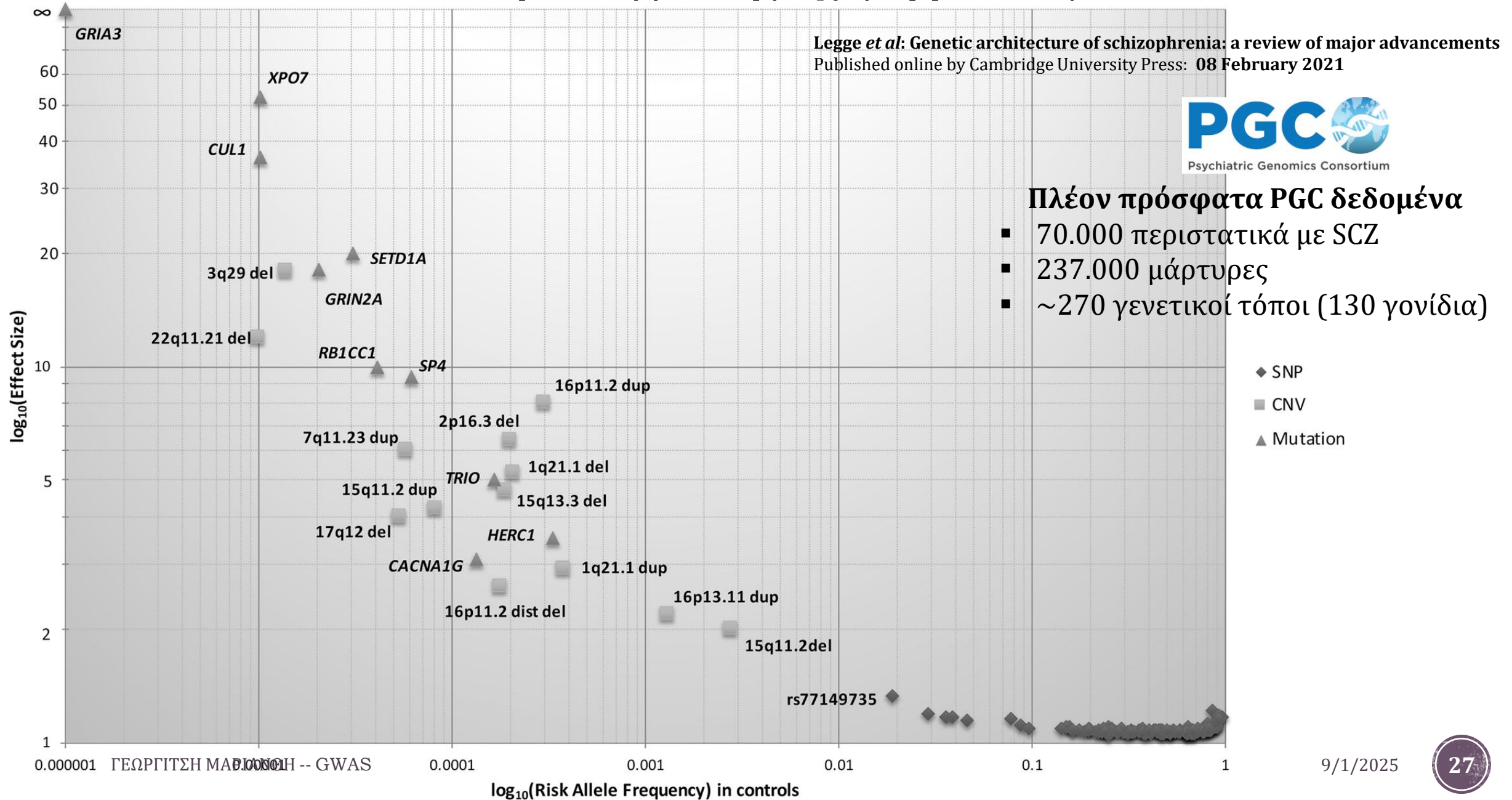
Box 2. GWAS methods

A GWAS consists of a series of independent regressions of a phenotype on genetic variants (specifically, SNPs) across the genome, one regression per SNP (see [Figure 2A](#) in main text). A modern SNP array includes fewer than 1 million ‘common’ SNPs [those whose **minor allele frequency (MAF)** > 1%], even though there are ~15 million common SNPs in the human genome [[105](#)]. Despite this discrepancy, the arrays capture the influence of almost all common SNPs as well as most other types of common polymorphisms (e.g., deletions/insertions), because genetic variants that are nearby on the genome tend to be correlated (in LD). Thus, each measured SNP ‘tags’ other nearby variants. Most modern GWASs impute nearly all common SNPs and use these imputed scores directly in the analysis.

Due to LD, the influence of a single causal variant will typically manifest itself in nearby SNPs. Thus, GWAS hits implicate a region where a causal variant is likely to exist rather than pinpointing the actual causal variant, making biological interpretation of individual variants challenging. Nevertheless, in aggregate, GWAS hits are not random. They occur more often in or near genes and regulatory regions and they often cluster in biologically meaningful genetic pathways (e.g., SNPs associated with bipolar disorder cluster in synaptic signaling pathways [[106](#)]), giving insight into the pathophysiology of disease [[107](#)].

The field has agreed upon standards and practices designed to minimize false positives. Because systematic differences in allele frequencies exist between subpopulations, and because environmental influences can differ across these subpopulations, it is routine to control for such **population stratification** by using mixed linear models and/or by including genomic principal components as covariates. Regardless of the number of SNPs examined, the maximum effective number of independent tests conducted on common SNPs in the genome is ~1 million [[108](#)], so the standard in the field is to require a Bonferroni-corrected threshold of $0.05/10^6 = 5 \times 10^{-8}$ for a SNP association to be considered ‘genome-wide significant’. Modern GWASs are performed on tens of thousands to millions of individuals, with sample size being the most important factor in determining the number of significant associations [[19](#)]. To achieve large samples, the field has adopted a highly collaborative approach, including forming large GWAS consortia, open access to large biobank datasets, and sharing summary statistics for all SNP tests following a GWAS. These practices reduce barriers of entry to the field, maximize data usage, increase transparency, and encourage crossfertilization of ideas.

Το παράδειγμα της σχιζοφρένειας



SUMMARY OF SCHIZOPHRENIA GENETICS

HERITABILITY

Schizophrenia has a strong genetic component. Twin and family studies estimate heritability at ~80%. SNP based estimates currently explain ~24% of the heritability.

Common variants (>1%)

SNPs

The largest GWAS to date by the PGC (69,369 cases, 236,642 controls) identified 329 independent genome-wide significant SNPs mapping to 270 distinct loci. Fine-mapping refined association to 130 genes most likely to explain these associations.

CNVs

The largest study (21,094 cases, 20,227 controls) in schizophrenia identified 8 CNV loci to be associated, collectively explaining 0.85% of variance in liability. 1.4% of cases carried risk CNVs. The most strongly associated CNV for schizophrenia is 22q11.2 deletion which confers a 20-fold risk.

SNVs

The latest SCHEMA Consortium analysis (24,248 cases, 97,322 controls) identified 10 genes with an excess of ultra-rare variants in schizophrenia cases. Other studies have found an increased burden of disruptive and damaging ultra-rare variants, and loss-of-function variants enriched in loss-of-function intolerant genes.

DNVs

In an analysis of 3444 trios, loss-of-function DNVs were found to be significantly enriched in loss-of-function intolerant genes in schizophrenia cases and this burden was higher for genes previously associated with neurodevelopmental disorders.

Rare variants (<1%)

POLYGENIC RISK SCORES

PRS is a single measure of common genetic liability. It can explain ~7.7% of the variance in schizophrenia case/control status with an AUC ~0.71. PRS calculated in one population explains less variance when applied to another, highlighting need for greater diversity in GWAS.

PLEIOTROPY

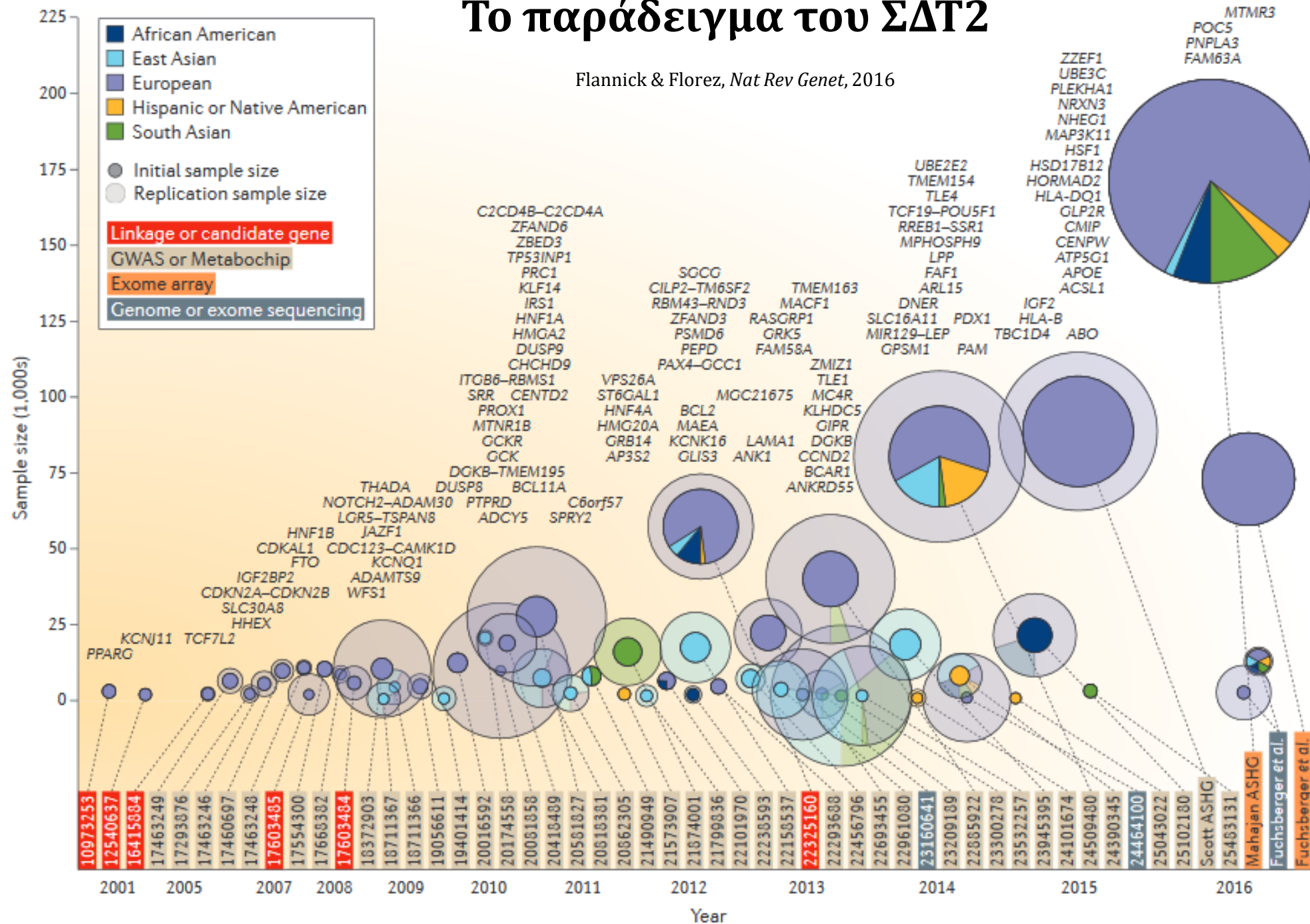
Risk variants for schizophrenia are also associated with other psychiatric and somatic disorders. Genetic correlations of common variants show significant overlap with most psychiatric conditions, cognitive ability, and BMI. Rare variants associated with schizophrenia are all also associated with intellectual disability and developmental disorders.

BIOLOGICAL INSIGHTS

Gene set analyses provide evidence that risk variants for schizophrenia act by disrupting the functioning of the synapse through multiple inter-related mechanisms. GWAS and rare variant studies have identified associations with targets of FMRP, voltage-gated calcium ion channel complexes, synaptic plasticity, and abnormal long-term potentiation. Rare variant studies have also found associations with ARC and NMDAR gene sets.

Το παράδειγμα του ΣΔΤ2

Flannick & Florez, *Nat Rev Genet*, 2016



GWAS Genome-wide Association Studies

Μελέτες συσχέτισης
γονιδιωματικής κλίμακας

ΠΕΡΙΕΧΟΜΕΝΑ

- Εισαγωγή – Complex traits (Πολυπαραγοντικά γνωρίσματα)
- Προς την ανάπτυξη τεχνικών γονιδιωματικής κλίμακας
- GWAS basics (concept, workflow, analyses)
- GWAS και Πολυγονιδιακοί Δείκτες Κινδύνου [Polygenic Risk Scores (PRSs)] – Πιθανοί βιοδείκτες?

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)

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

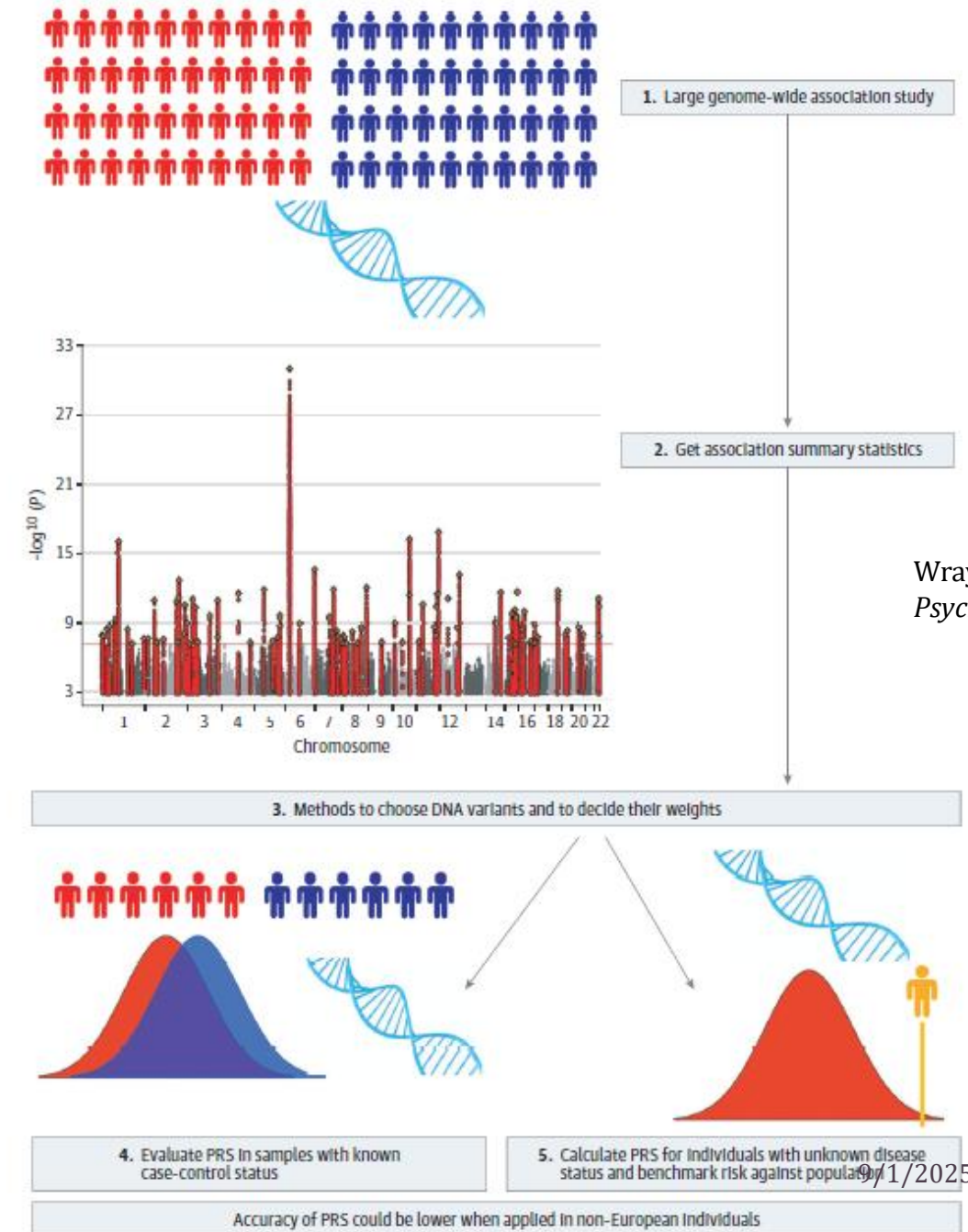
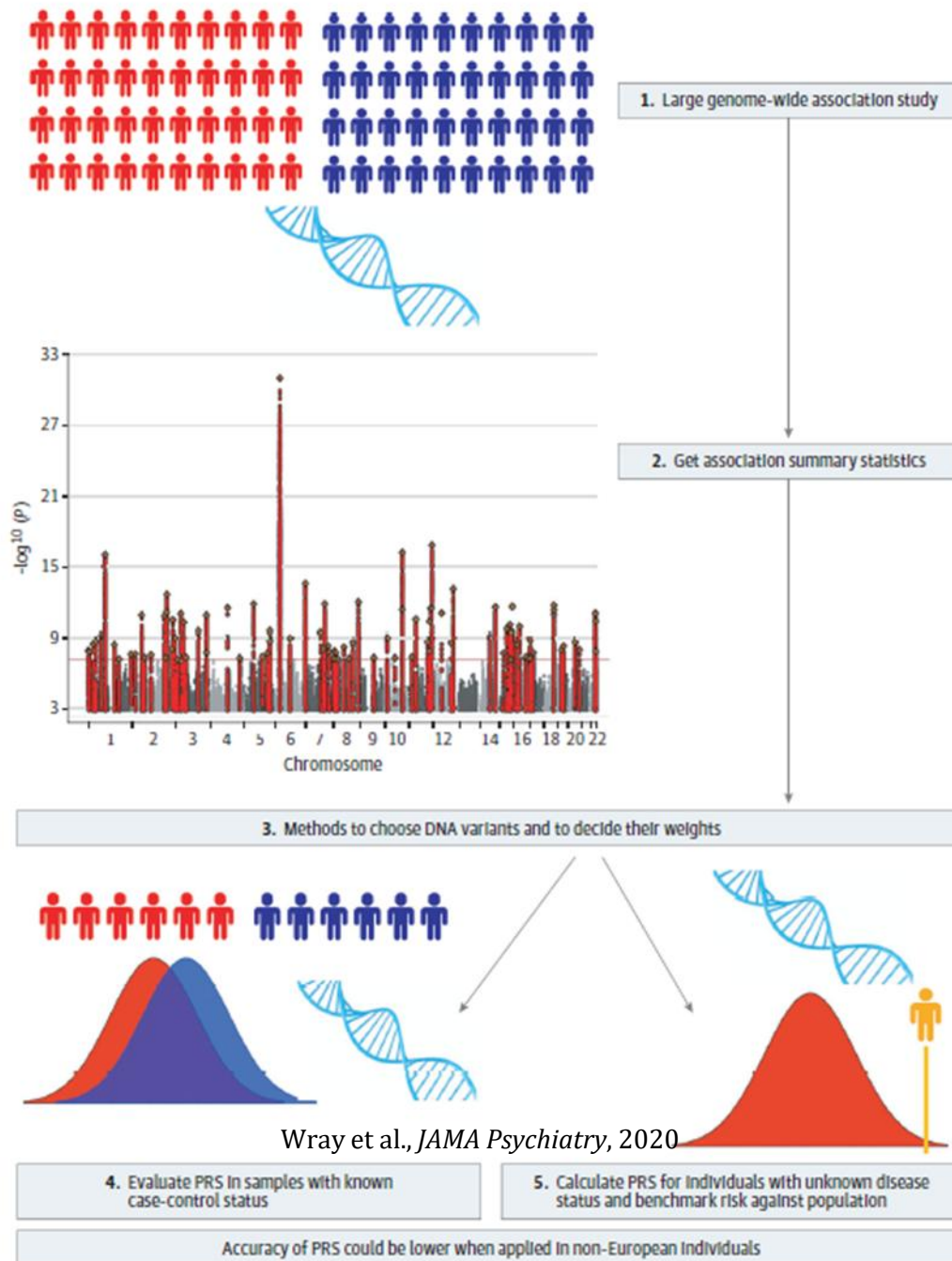
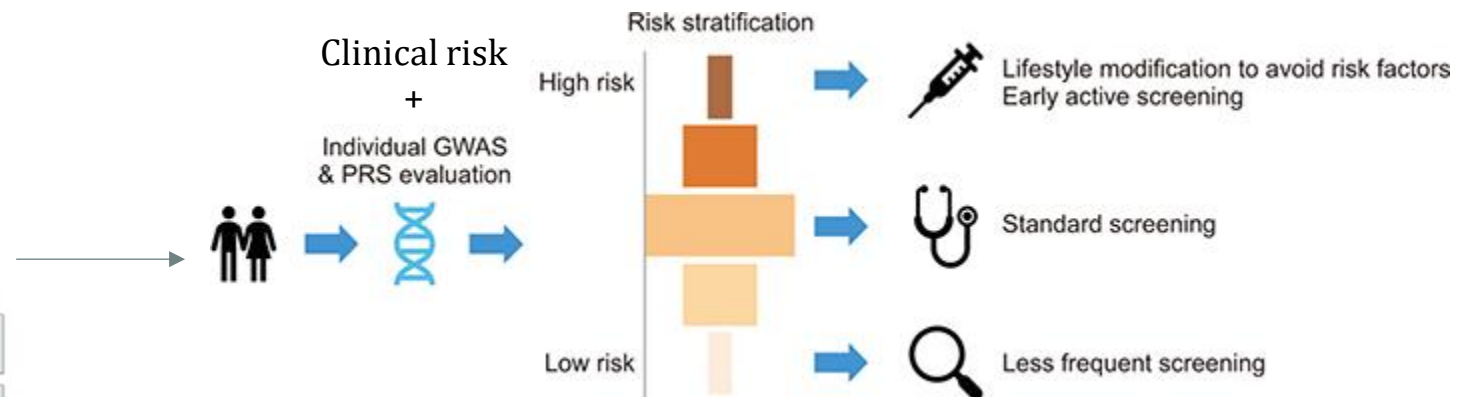


Figure 2. Schematic of the Steps Needed to Generate and Validate Polygenic Risk Scores (PRS)

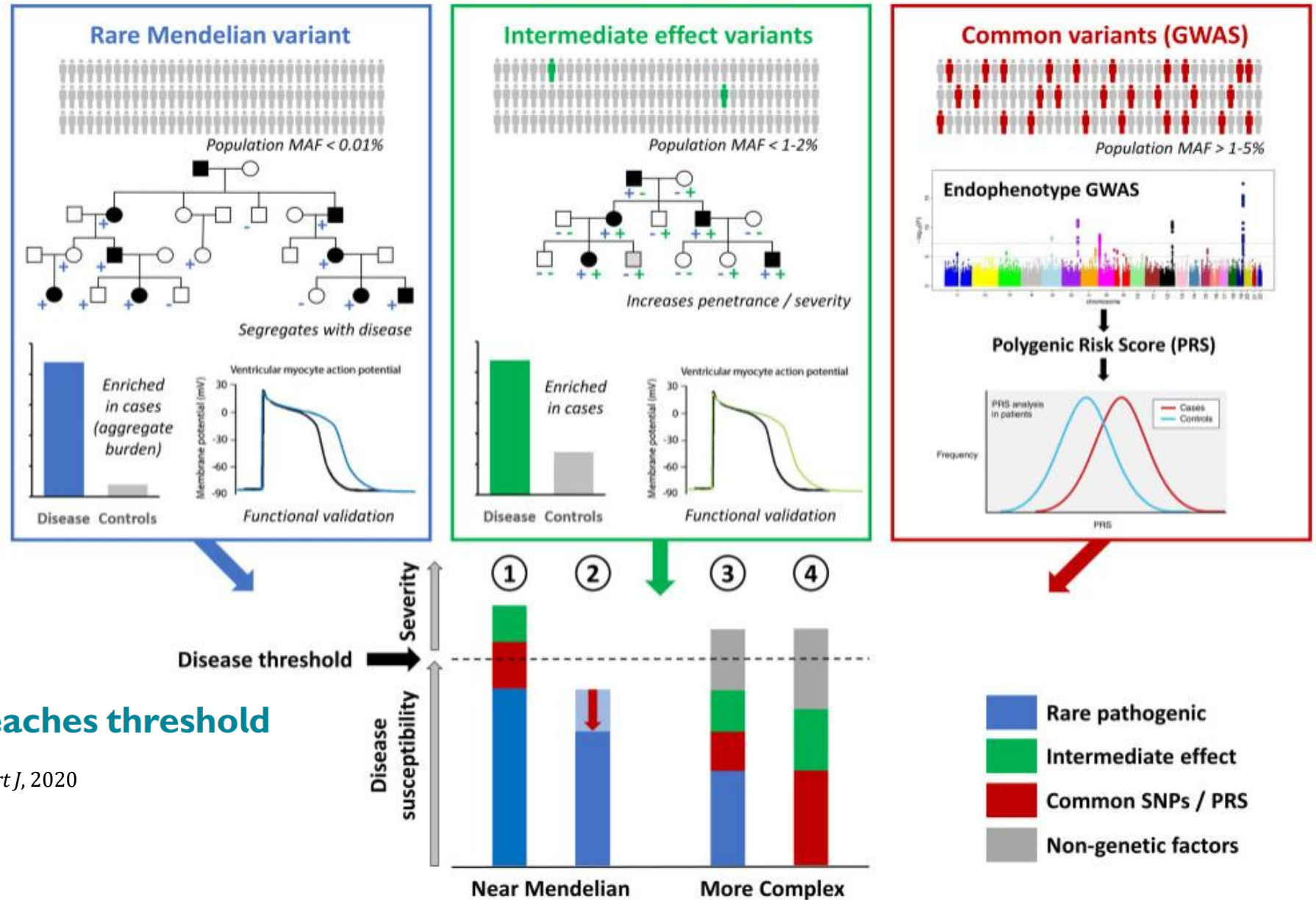


Polygenic risk scores

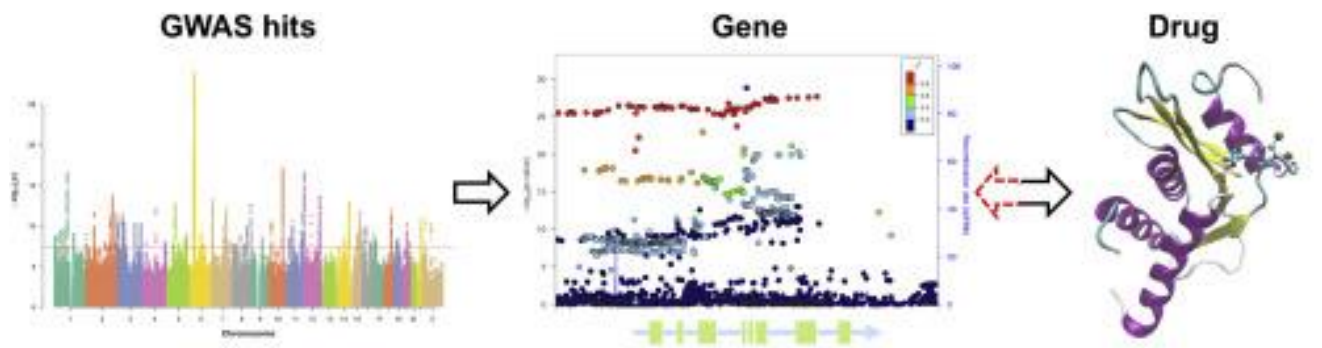
- ✓ PGSs are a **quantitative measure** of the additive genetic burden (polygenic contribution) for a particular disease or trait that can be used to assess **individual genetic load**.
- ✓ PGSs typically aggregate hundreds or thousands of variants with small individual effect sizes and follow a normal distribution on a population level.
 - 6–8 % for SCZ (Purcell et al., *Nature*, 2009)
 - 3-4 % for BD (Purcell et al., *Nature*, 2009)
 - 2–3 % for MDD
 - ~3% for OCD (Yu et al., *Am J Psychiatry*, 2015)
 - 0.6-0.8% for GTS (Yu et al., *Am J Psychiatry*, 2015; 2019)



Σύνοψη



Επιτυχίες του πεδίου



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

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

Η ανίχνευση γενετικών δεικτών με προοπτική για τη **μεταφραστική έρευνα** και **διάγνωση** ή **πρόληψη** είναι εξαιρετικής σημασίας για πεδία της ιατρικής επιστήμης στα οποία:

A) Τα **διαγνωστικά κριτήρια** δεν **basίζονται σε βιοδείκτες** (για παράδειγμα στην ψυχιατρική, τις νευροαναπτυξιακές διαταραχές και τη μελέτη συμπεριφορικών χαρακτήρων)

B) Υπάρχει **έλλειψη αποτελεσματικής θεραπείας**

Γ) Είναι εξαιρετικά μεγάλη η **ετερογένεια στο φαινότυπο**

Δ) Είναι νοσήματα με **όψιμη έναρξη**, άρα προκύπτει μεγάλο όφελος από την έγκαιρη γνώση για την πρόληψη

Εφαρμογές στην έρευνα και την κλινική πράξη

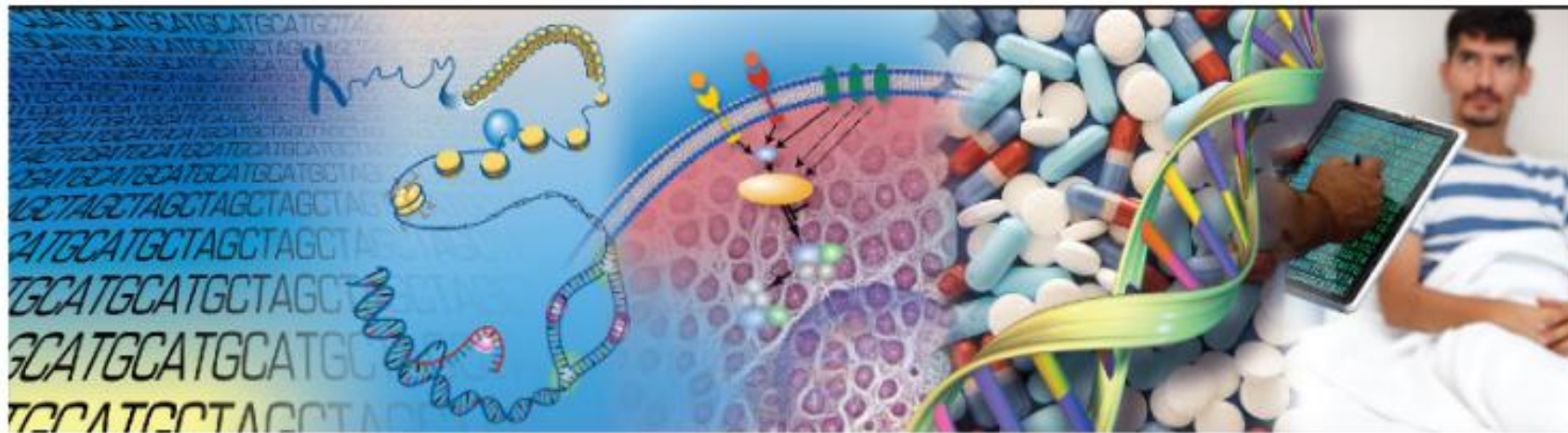
Understanding
the structure of
genomes

Understanding
the biology of
genomes

Understanding
the biology of
disease

Advancing
the science of
medicine

Improving the
effectiveness of
healthcare



<https://www.genome.gov/Pages/About/Planning/2011NHGRIStrategicPlan.pdf>

BOX 3

Bioinformatics and computational biology



The major bottleneck in genome sequencing is no longer data generation—the computational challenges around data analysis, display and integration are now rate limiting. New approaches and methods are required to meet these challenges.

Data analysis. Computational tools are quickly becoming inadequate for analysing the amount of genomic data

that can now be generated, and this mismatch will worsen. Innovative approaches to analysis, involving close coupling with data production, are essential.

Data integration. Genomics projects increasingly produce disparate data types (for example, molecular, phenotypic, environmental and clinical), so computational approaches must not only keep pace with the volume of genomic data, but also their complexity. New integrative methods for analysis and for building predictive models are needed.

Visualization. In the past, visualizing genomic data involved indexing to the one-dimensional representation of a genome. New visualization tools will need to accommodate the multidimensional data from studies of molecular phenotypes in different cells and tissues, physiological states and developmental time. Such tools must also incorporate non-molecular data, such as phenotypes and environmental exposures. The new tools will need to accommodate the scale of the data to deliver information rapidly and efficiently.

Computational tools and infrastructure. Generally applicable tools are needed in the form of robust, well-engineered software that meets the distinct needs of genomic and non-genomic scientists. Adequate computational infrastructure is also needed, including sufficient storage and processing capacity to accommodate and analyse large, complex data sets (including metadata) deposited in stable and accessible repositories, and to provide consolidated views of many data types, all within a framework that addresses privacy concerns. Ideally, multiple solutions should be developed¹⁰⁵.

Training. Meeting the computational challenges for genomics requires scientists with expertise in biology as well as in informatics, computer science, mathematics, statistics and/or engineering. A new generation of investigators who are proficient in two or more of these fields must be trained and supported.