Genetic Epidemiology

Evangelos Evangelou, PhD

vangelis@cc.uoi.gr

eevangelou

Εντοπίστηκαν γονίδια που σχετίζονται με πρώιμη εφηβεία και εμμηνόρροια στα κορίτσια

Ανακαλύφθηκαν 30 γονίδια που σχετίζονται με την εμφάνιση πρώιμης εφηβείας και εμμηνόρροιας στα κορίτσια.

Ανακαλύφθηκαν γονίδια που σχετίζονται με το άσθ

Βρέθηκαν γονίδια που σχετίζονται με την μυωπία

Ελπίδες για μελλοντική πρόληψη μόνο με σταγόνες ή χάπια.

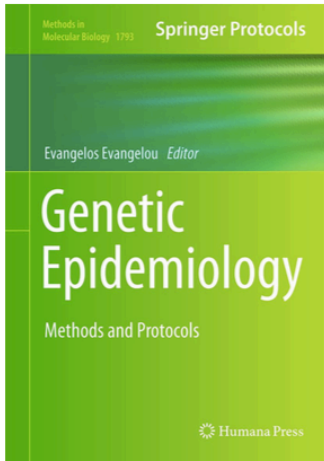## Ανακαλύφθηκαν γονίδια που συνδέονται με την ανορεξία

ΕΠΙΣΤΗΜΗ

Βρέθηκαν γονίδια που καθορίζουν το ύψος του ανθρώπου

Course contents (week I)

- Introduction to Genetic Epi

- Study designs in Genetic Epi

- Genetic association studies

Readings

Chapters:
2. Key concepts in genetic epidemiology
3. Quality control of common and rare variants
4. Genome-wide association studies
5. Assesing rare variation in complex traits
6. Meta-analysis of common and rare variants

Additional Readings

- Warren HR, Evangelou E, Cabrera et. Al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. 2017;49(3):403-415

- Evangelou E, Warren HR, Mosen-Ansorena D et al. Genetic analysis of over one million people identifies 535 novel loci associated with blood pressure traits. Nature Genet. 2018; in press
  - https://www.biorxiv.org/content/early/2017/10/11/198234

- Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013;14(6):379-89

- Visscher PM et al. 10 years of GWAS discovery: Biology, Function, and Translation. Amer J Human Genet. 2017;101(1):5-22

Introduction to Genetic Epidemiology

Learning outcomes

- Describe genome structure and human genetic variation

- Provide an account of key concepts of population genetics and genetic epidemiology (i.e., heritability, linkage disequilibrium)

- Understand the goals and principles of family- and population-based designs in genetic epidemiology

Outline

- Genetic Epidemiology
- Genome structure and genetic variation
- Genetic and epidemiological study designs
  - Estimation of genetic effects
- Key concepts
  - Heritability
  - Linkage disequilibrium

Genetic Epidemiology

A hybrid science focusing on *complex* diseases (where both genetic & environmental factors contribute to etiology of disease)
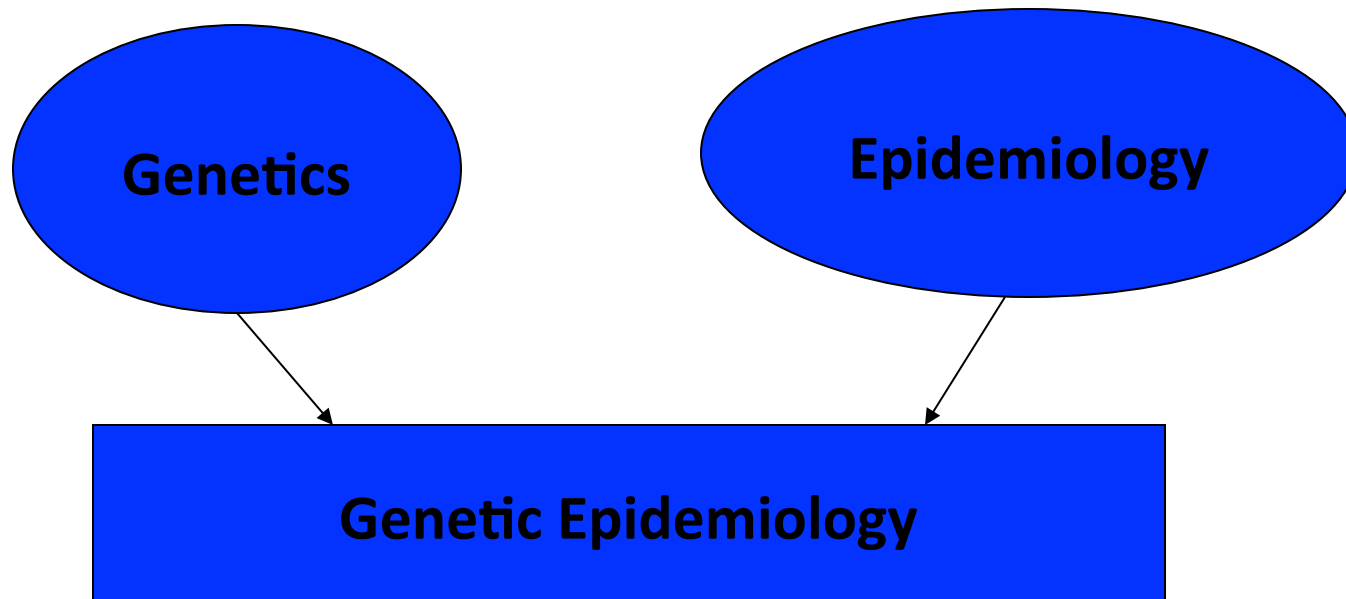
Parent sciences (genetics & epidemiology) share common goals but they differ in their histories & perspectives.

Genetic Epidemiology

- "A field of science that focuses on the role of genetic factors and their interaction with environmental factors in the occurrence of disease in human populations"

Fundamentals of Genetic Epidemiology. Khoury M, et al.

Genetic Epidemiology

- Is based on principles of population genetics

- Utilizes statistical approaches to detect the genetic effects on susceptibility to chronic diseases and quantitative traits
  - Type 2 Diabetes
  - Prostate cancer
  - Obesity or quantitative trait, e.g. BMI

Genetic Epidemiology
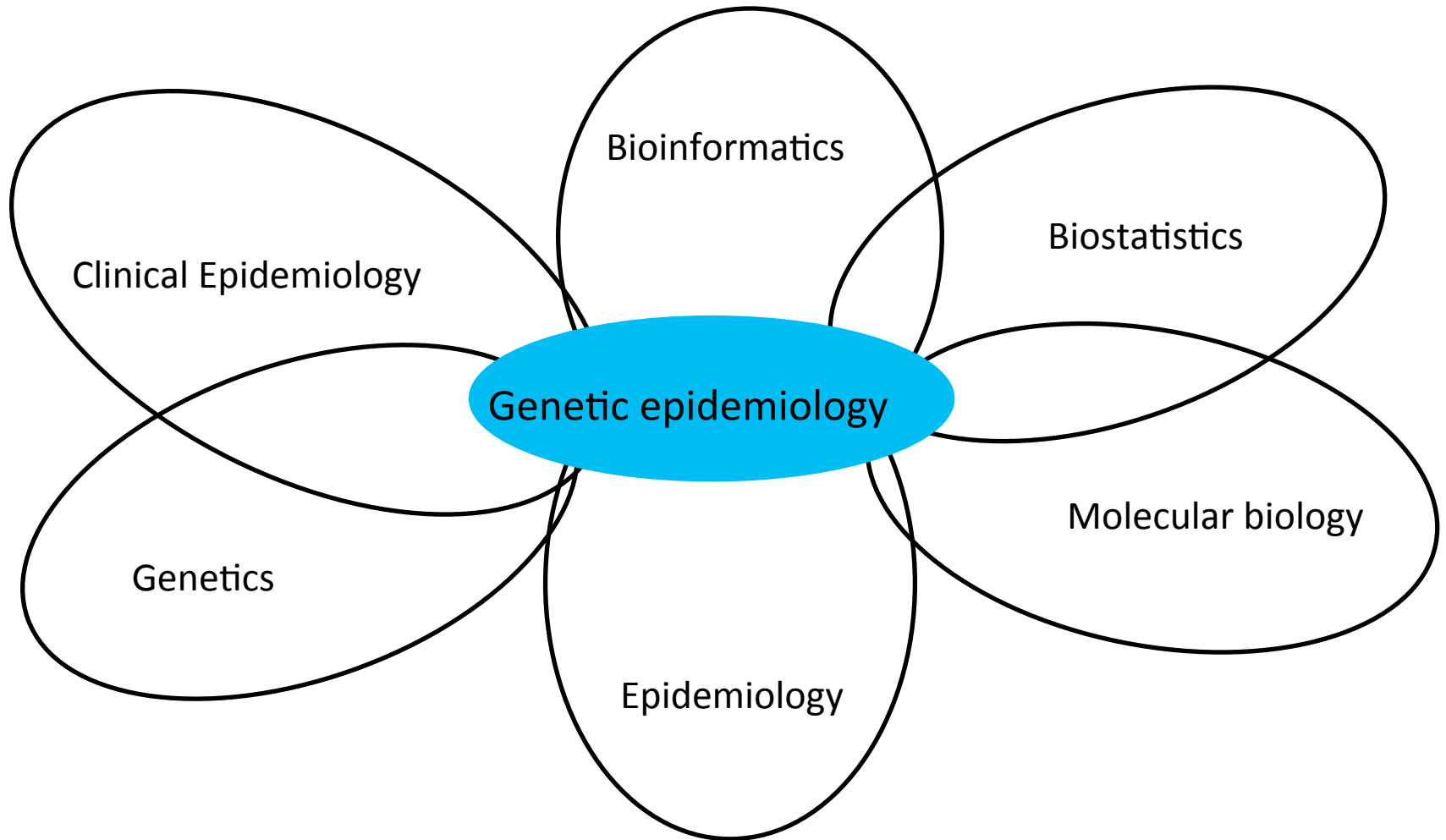
Landmarks in Genetics

Year	Event
1865	Gregor Mendel publishes work on peas describing fundamentals of inheritance
1871	DNA is isolated from the cell nucleus
1900	3 people independently "re-discover" Mendel's work (Correns, DeVries & vonTschermak)
1901-02	Garrod discovers human example of Mendelian disease (alkaptonuria) & Landsteiner discovers 1st genetic marker (ABO)
1908	Hardy & Weinberg lay the foundation for modeling genes in populations

Landmarks in Genetics (cont'd)

1930's	Biometrical school of genetics develops statistical models for genes in families & populations
1953	Double helix structure of DNA identified by Watson & Crick (& R. Franklin)
1966	Genetic code established (3 nucleotides per codon)
1972	Recombinant DNA techniques developed
1987	Human Genome Project proposed
2001	Draft sequence of human genome available
2008	1000 Genomes Project commences

Central questions in Genetic Epidemiology

1. Does the trait cluster in families?
2. Can familial clustering be explained by genes or shared environment?
3. What is the best model of inheritance?
4. Can we locate genes for complex diseases/traits?
5. How does the gene control risk of disease?

Use of genetic terms over time
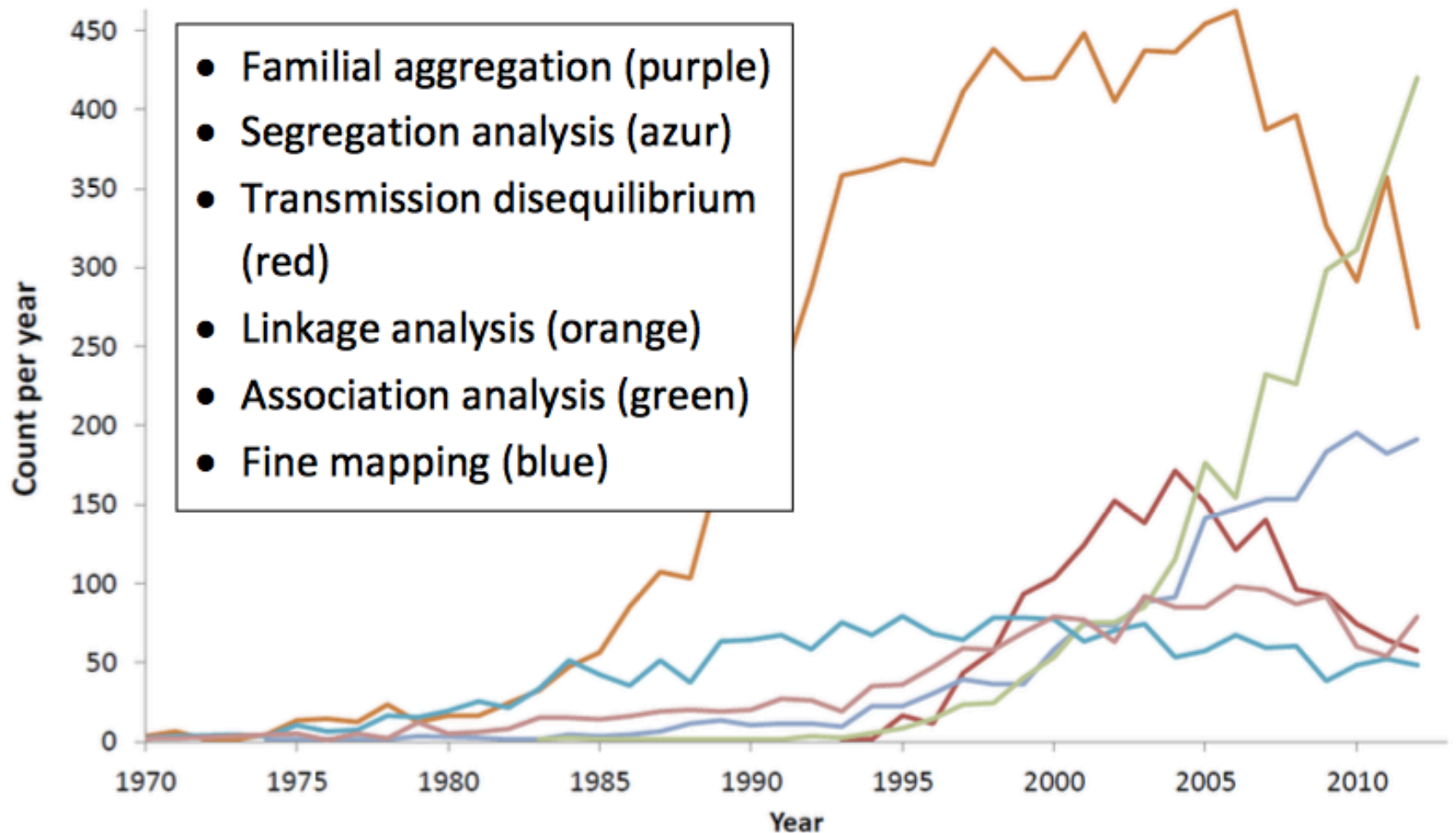
(adapted from IGES presidential address A Ziegler, Chicago 2013)

Figure 1. ROR and 95% CIs for Each Comparison of an Unrelated Case-Control Study versus Family-Based Study

PLOS | GENETICS
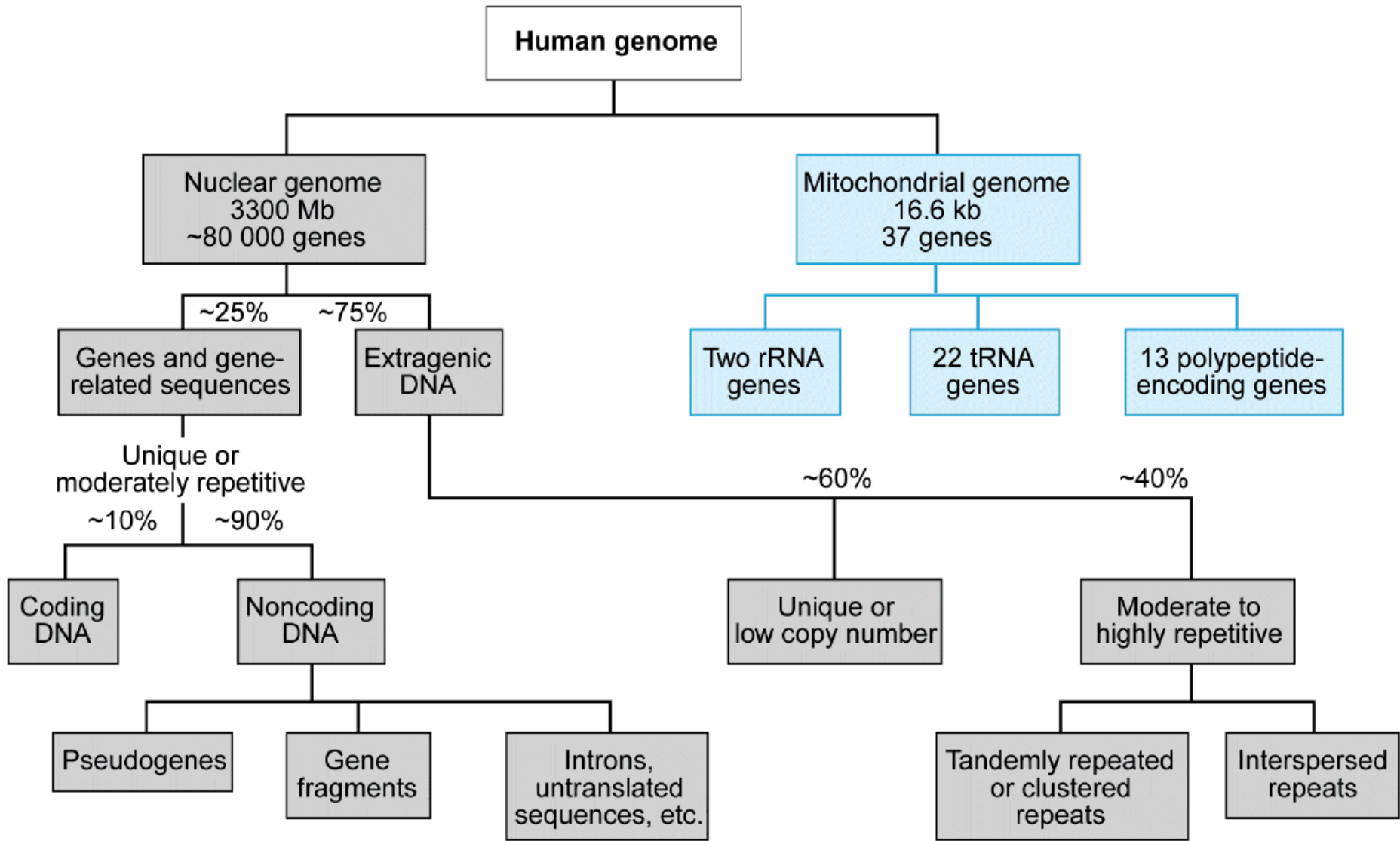
Organization of the human genome

- Nuclear genome
  - 3200 Mb
  - 23 (XX) or 24 (XY) linear chromosomes
  - ~20,000 protein-coding genes
  - 1 gene/30-60kb
  - Only 10% is coding sequence

  - Introns
  - 3% coding
  - Repetitive DNA sequences (45%)
  - Recombination
  - Mendelian inheritance (X + auto, paternal Y)
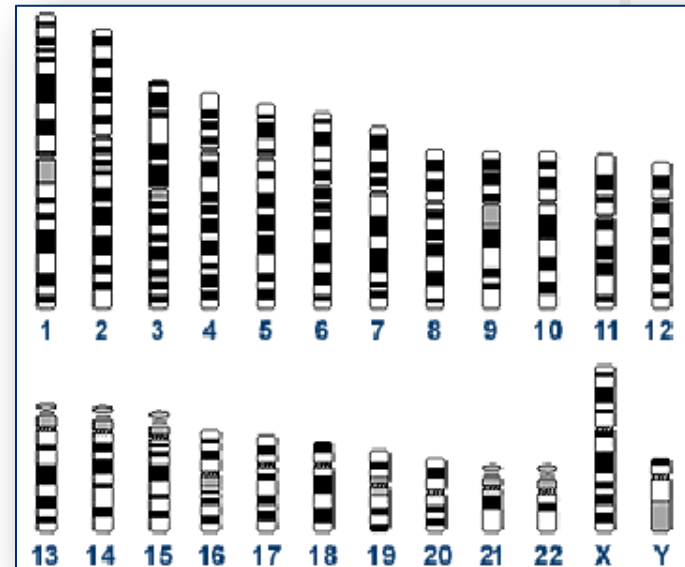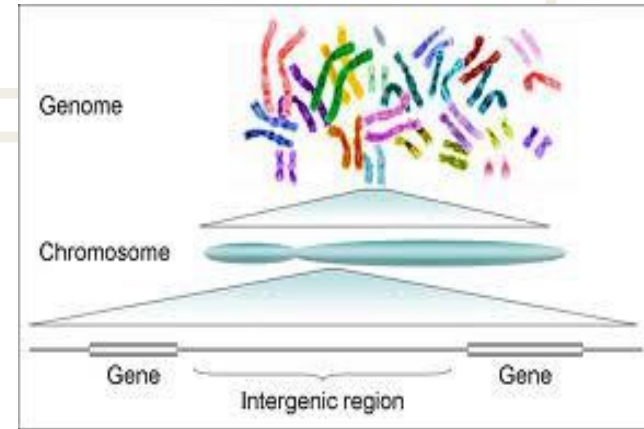
Organization of the human genome

- Genes vary in size and exon content
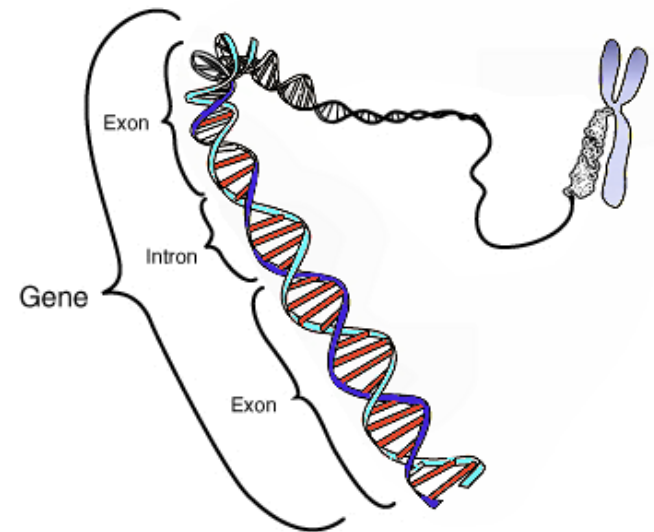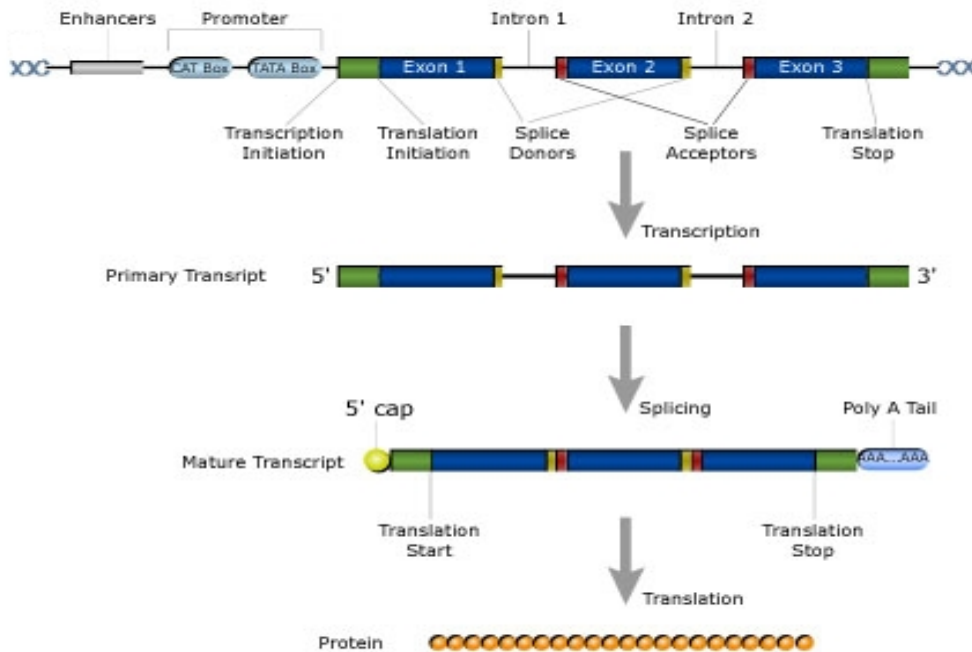
Human Genetic variation and disease

- What is a SNP?

- Types of SNP

- SNPs as genetic markers

- GWAS

Watson and Crick
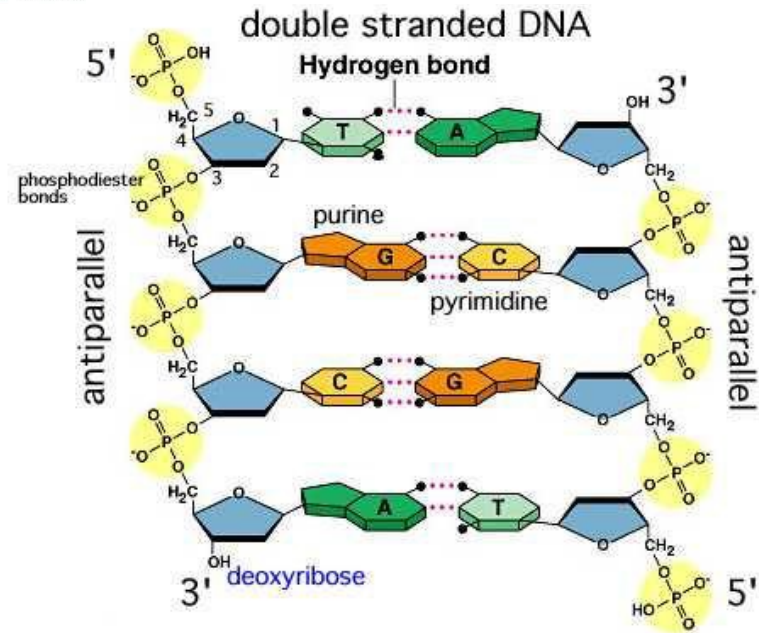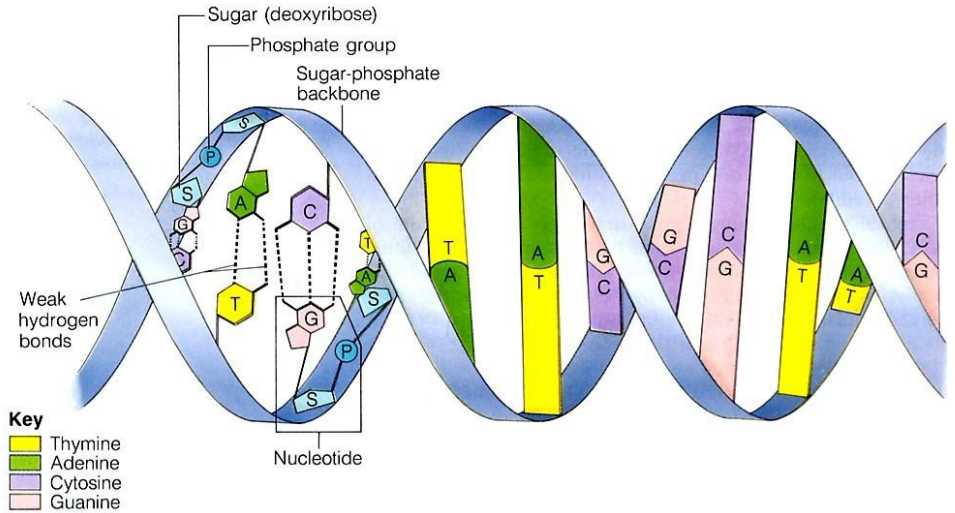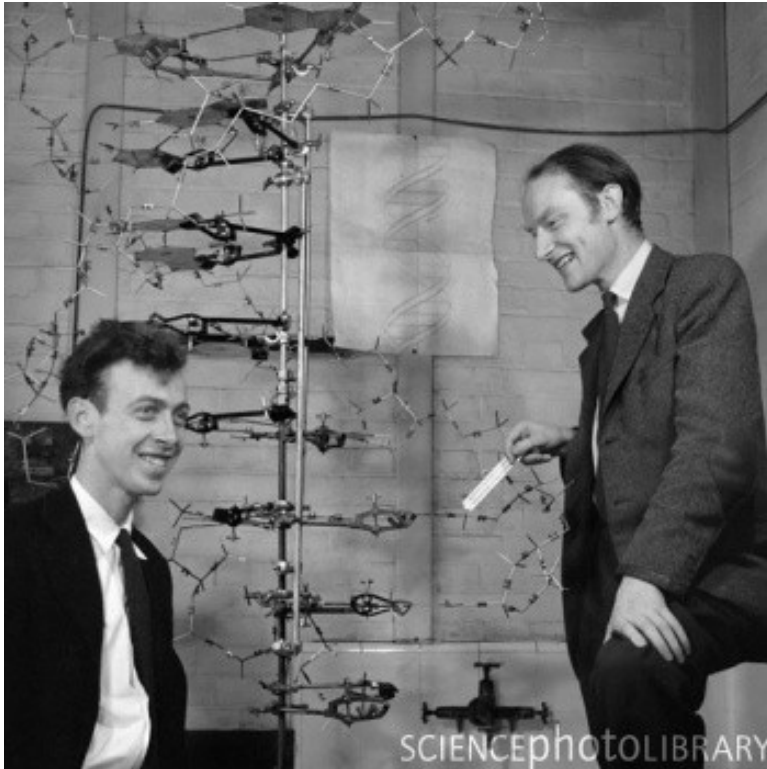
SNPs

- Single/simple nucleotide polymorphism – SNP
- A single nucleotide variant in the DNA
- SNP is a DNA sequence variation within a single nucleotide— A, T, C or G — in the genome
- E.g. Adenine to Guanine, Thymine to Cytosine
- Mostly biallelic (two alleles) polymophism,
  - AAGGTTA σε ATGGTTA
  - but large number of tri- and quadri-allelic SNPs is now described
- Could also be a 1bp indel, duplication, etc

SNP or mutation?

- Typically not considered to have a functional effect, hence polymorphism and not mutation

- However, SNP is often used as a term for all single-base changes, functional or not.

- The difference is that a mutation has a functional effect and a polymorphism does not necessarily

Genetic terms

- Each variant of a gene (at a given locus) is an **allele**
  - e.g. *A* or *a*
- For a genetic marker, the two parentally-inherited variants combined are called a **genotype**
  - e.g. *A,a*
- The site or location on a chromosome occupied by a gene is called a **locus**

Minor Allele Frequency (MAF)

- This is how often the less frequent allele of a biallelic variant occurs in a group (often a percentage)
- As the total allele frequency is 1 (100%), a MAF must always be less than 0.5 (50%), otherwise it would be a major allele
- E.g. if we genotype a variant (A/G) in 1000 people
  - 550 are (A,A), 400 are (A,G) and 50 are (G,G)
  - There are 2000 alleles in total
  - The G allele is less common, accounting for 500 alleles
  - Therefore, the MAF is 500/2000 = 0.25 or 25%

Types of SNPs based on MAF

- Common SNPs
  - MAF > 5%
  - ~7 M in human genome

- Low frequency SNPs
  - 1% < MAF < 5%.
  - ~11 M in human genome

- Rare SNPs
  - MAF < 1%
  - >100 M in human genome

Types of SNP

- Non-Synonymous
- Synonymous
- Promoter
- Terminator
- Splicing
- Neutral

SNPs as Genetic Markers

- Bi-allelic
- Very common across the genome
- Good evidence that they can directly cause disease
- Easily genotyped using high-throughput technologies
- Widely used for association and linkage studies

(db)SNP
http://www.ncbi.nlm.nih.gov/snp

➢ More information on genetic variation

- genotype

- allele frequencies

- chromosome position

- sequence

- population diversity

- visual displays

PubMed Gene
http://www.ncbi.nlm.nih.gov/pubmed/gene/

➢ Database of genes in humans (n=43,828) and other organisms

- gene name

- alias (e.g. *TCF7L2*)

- function

- lineage in other organisms

- biological pathways

Step back: Key concepts of population genetics

- Heritability

- Hardy-Weinberg equilibrium

- Linkage disequilibrium

Heritability (of a trait) definitions

- **Fraction of phenotypic variability that is attributable to genetic variation**

- IS NOT: how much genetics influences trait in one person

- is relative to specific population in a particular environment (since contribution of genetic factors is relative to contribution of other factors such as environment)

Heritability

- Phenotype P
- Genotype G
- Environment E
- Var(P) = Var(G)+Var(E)+2Cov(G,E)

Broad Sense Heritability:
(includes additive, epistatic,
dominant genetic effects)

$$H^2 = \frac{Var(G)}{Var(P)}$$

Narrow Sense Heritability
(includes only additive genetic
effects)

$$h^2 = \frac{Var(A)}{Var(P)}$$

Heritability

- Classically, in twins, heritability of a trait is twice the difference in the correlation between identical (MZ) and non-identical twins (DZ)

- $H^2=2(r(MZ)-r(DZ))$

Heritability of Traits
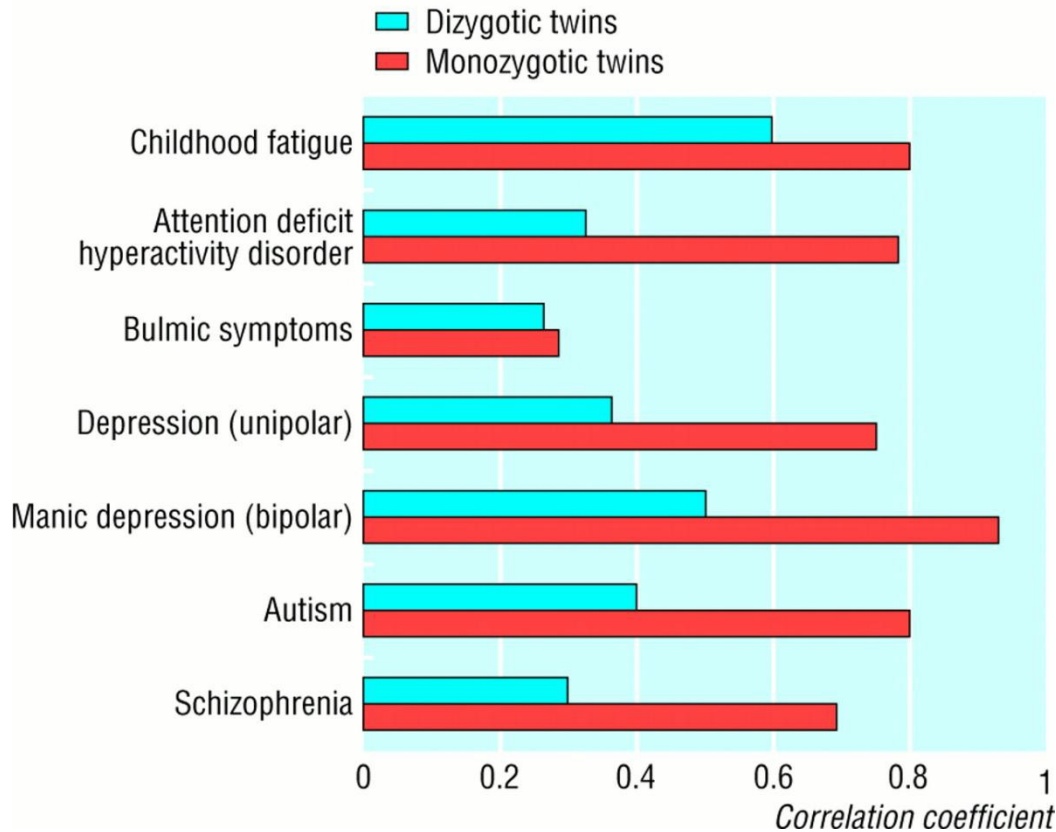
Schizophrenia

r(MZ) = 0.7
r(DZ) = 0.3

$H^2 = 2(0.7-0.3)$

Heritability = 0.8

Examples of estimated heritability

- Alcoholism 50-60%
- Alzheimer's 58-79%
- Asthma 30%
- Bipolar Disorder 70%
- Depression 50%
- Hair Curliness 85-95%
- Lung Cancer 8%
- Height 81%
- Obesity 70%
- Longetivity 26%
- Sexual Orientation 60%
- Schizophrenia 81%
- Type 1 diabetes 88%
- Type 2 diabetes 26%

Hardy-Weinberg Equilibrium (HWE)

➤ Mathematical model of expected genotype frequencies in a population

➤ Allele and genotype frequencies will remain constant from generation to generation in the absence of other evolutionary influences

Hardy-Weinberg Equilibrium (HWE)

➢ Violations of HWE could be due to:
- Non-random mating (i.e., inbreeding)
- Natural selection
- Mutation
- Migration
- Chance (in small populations)

- Genotypic Errors
- Association?

Hardy-Weinberg Equilibrium (HWE)

➤ Let's imagine a genetic locus with two alleles (A and a)

➤ p: frequency of A

➤ q=1-p: frequency of a

➤ p^2: frequency of AA

➤ q^2: frequency of aa

➤ 2pq: frequency of Aa

➤ $p^2 + 2pq + q^2 = 1$

		Paternal gametes	
		A (p)	α (q)
Maternal gametes	A (p)	AA (p^2)	Aa (pq)
	α (q)	Aa (pq)	Aa (q^2)
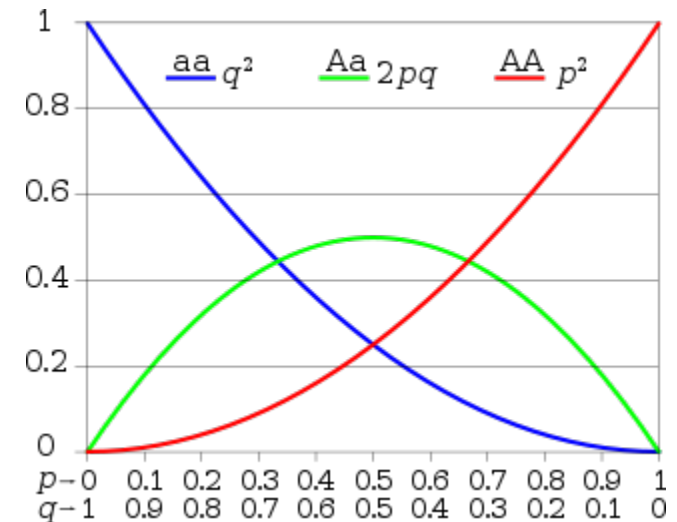
Example of HWE

- p=(2*AA+1*Aα)/2N
- q=1-p
- N population
- 2N alleles

Genotype	Number
AA	136
Aa	209
aa	80
Total (N)	425
p=(2*136+1*209)/2*425=0,57	
q=0,43	

Example of HWE

- $\chi^2 = \Sigma(O_i - E_i)^2/E_i$
- H_0: HWE
- $\chi^2 = 0{,}003$ with 1 df
- H_0 cannot be rejected
- If $\chi^2 \geq 3{,}84$ with 1 df then $P < 0{,}05$

Genotype	Observed number (O)	Expected number (E)
AA	136	$p^2*N = 136{,}1$
Aa	209	$2pq*N = 208.8$
aa	80	$q^2*N = 80.1$

Linkage disequilibrium (LD)

➢ Non-random association of alleles at different loci

➢ Presence of statistical associations between alleles at different loci that are different from what would be expected if alleles were independently transmitted from generation to generation

Linkage disequilibrium (LD)

➤ Measures of LD
  • D'
  • r^2
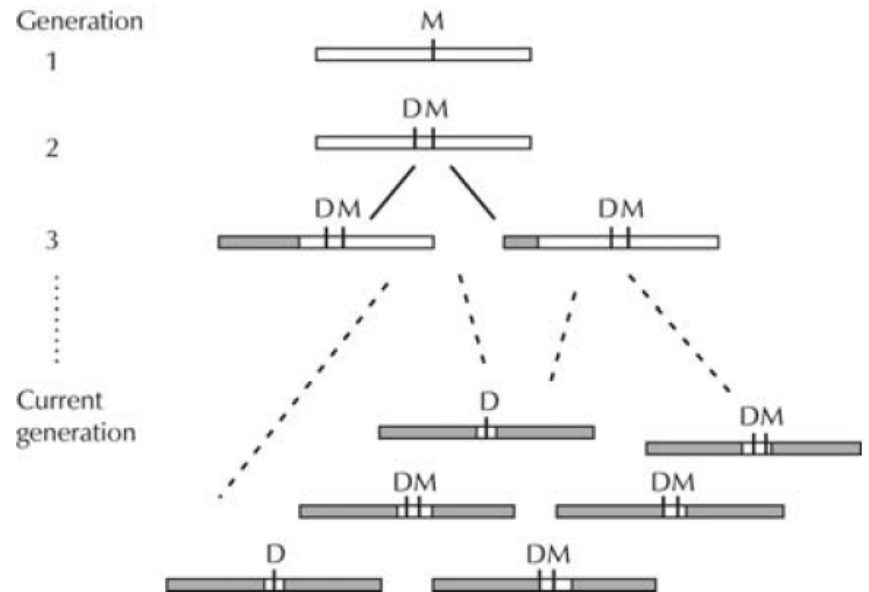
LD is diminished with time and increased recombination rate
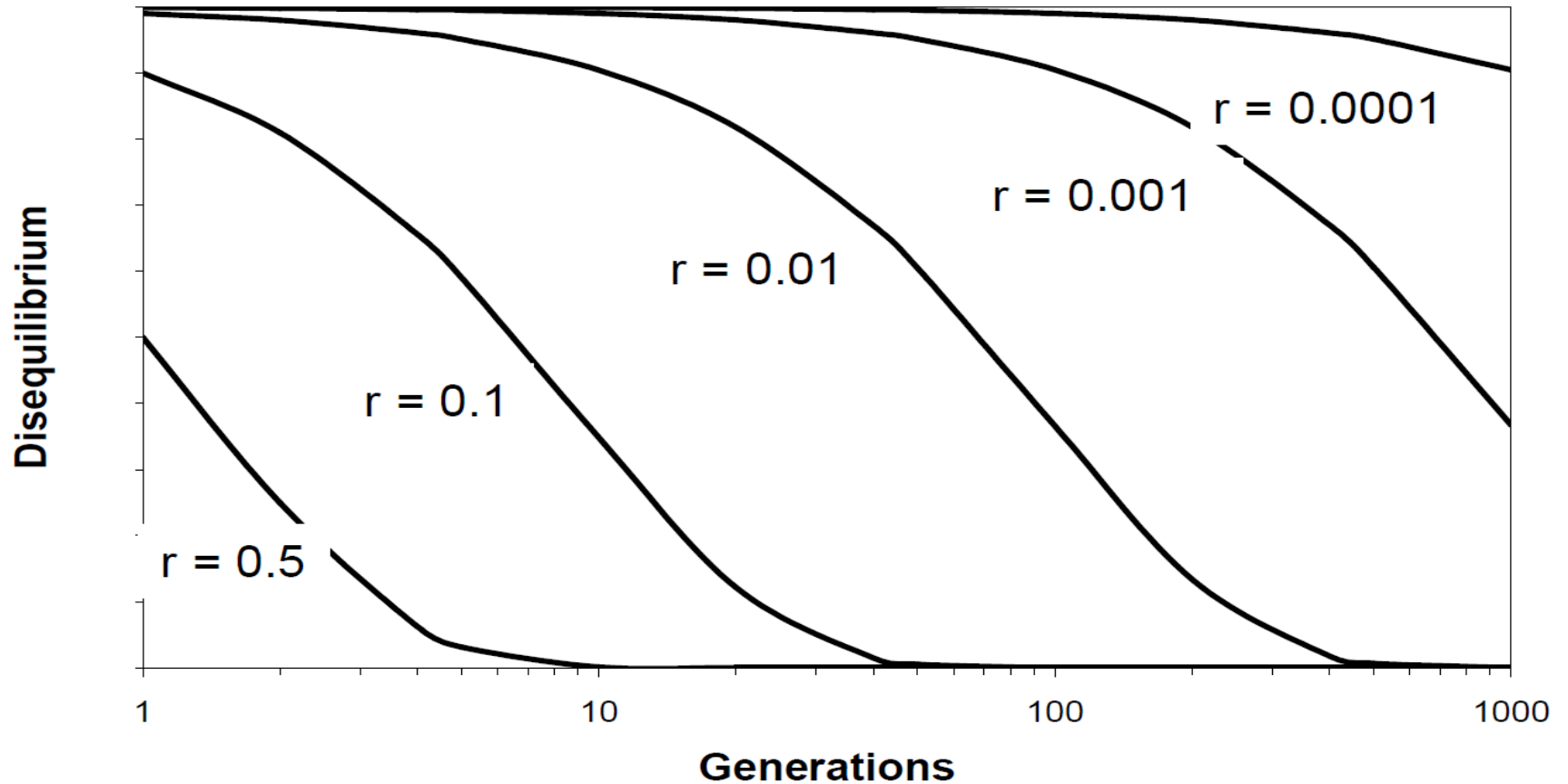
Linkage disequilibrium map
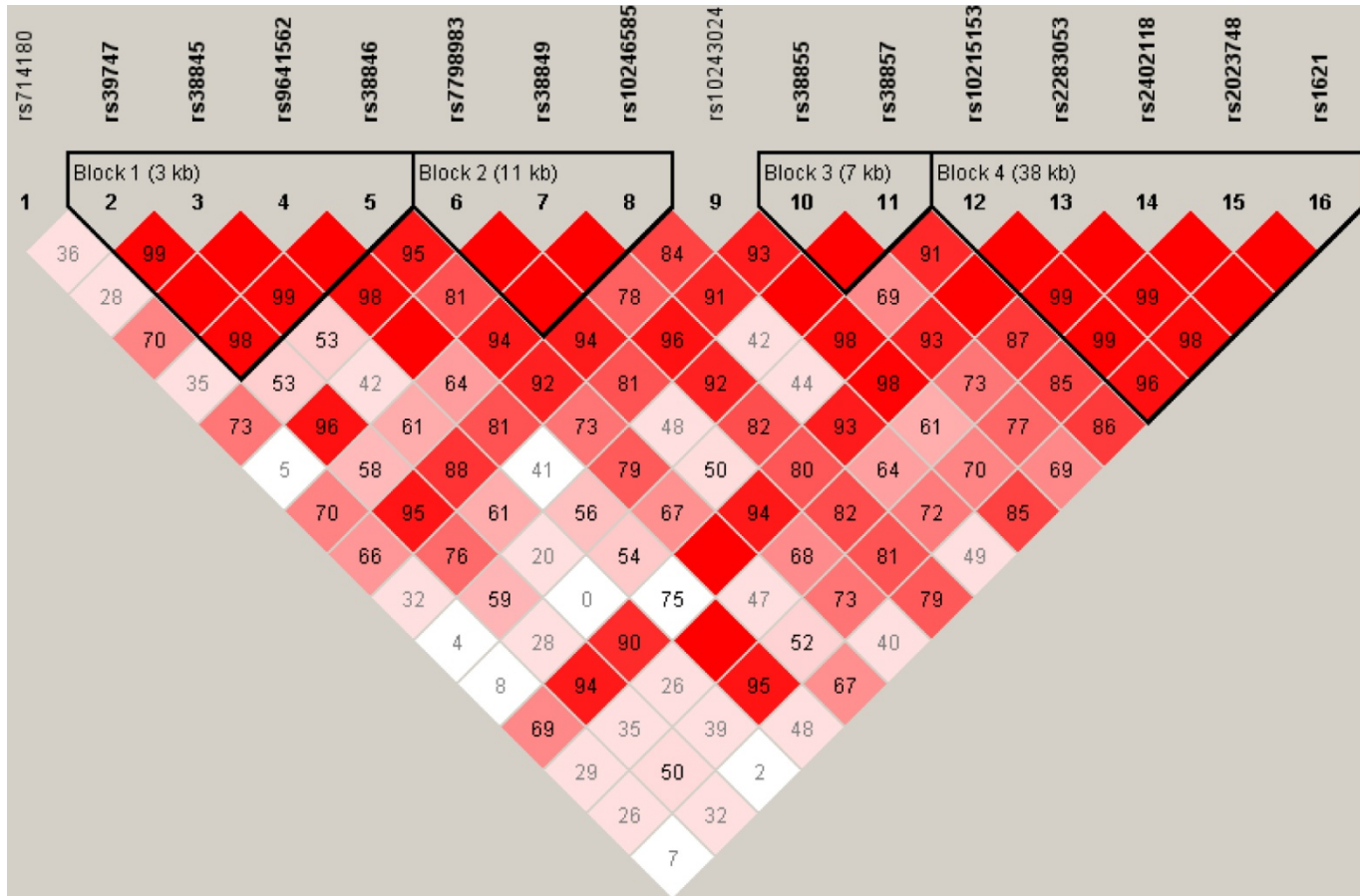
Study designs in Genetic Epidemiology

Study designs of genetic associations

Case-control studies

Prospective cohorts

Retrospective cohorts

Cross-sectional studies

Nested case-control studies

Case-control study design in genetics

- Design: identify participants based on their disease/outcome status, compare presence of genetic variant
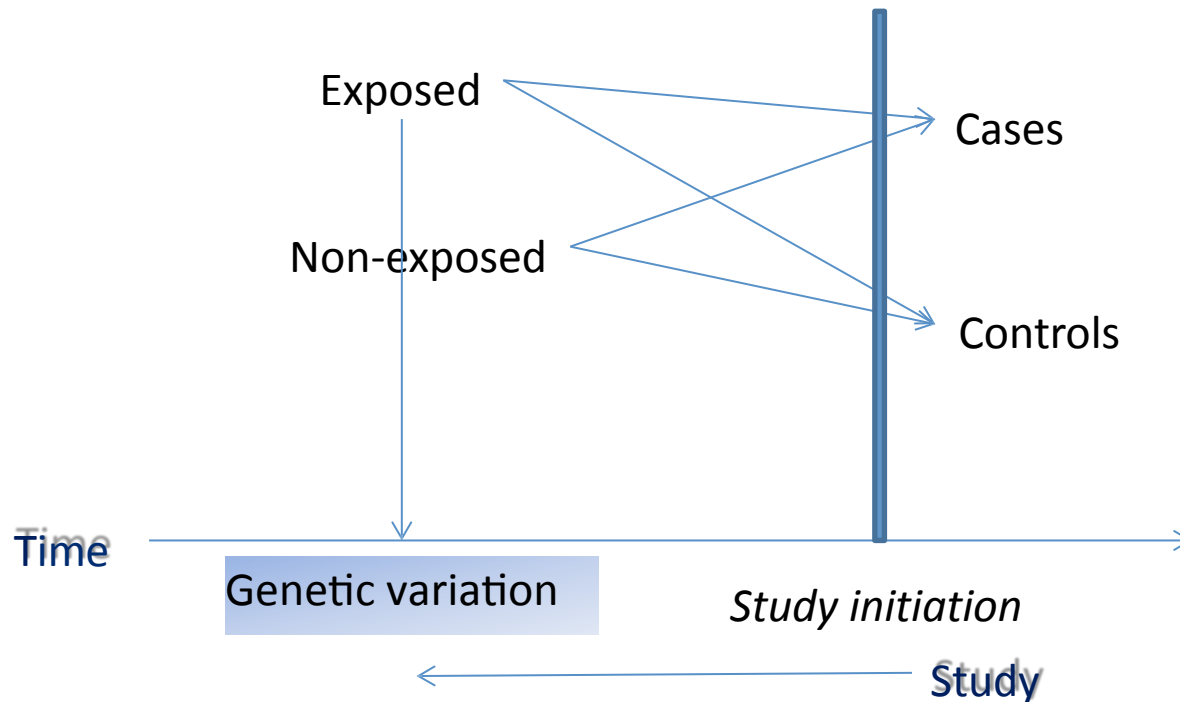
Assumptions

- Cases representative of all cases of disease

- Controls drawn from the same population as cases (and at risk for the outcome)

- Exposure data (genetic information) collected similarly in cases and controls
  - Genetics: T2D cases DNA is extracted from whole blood, controls DNA is from cell lines

Advantages of a case-control study

- Suitable for rare outcomes
- Suitable for outcomes with long induction period
- Cheaper
- Need fewer people in some cases
- Readily evaluate multiple exposures
- Convenient
- If assumptions are met, valid estimates of relative risk

Disadvantages of a case-control study in genetics

- Retrospective (not so much of a problem in genetic epi)
- Difficult to study rare exposures
- Genetic confounding (population stratification)
- Problematic when investigating G*E interactions
- Special considerations (more later)
  - Exposure-related
    - Recall bias: Disease status may influence reporting (not so much of a problem in genetic epidemiology as genetic variation is determined at the time of gamete formation)
  - Outcome-related
    - is studying survivors of the disease

Subtypes of case-control studies

- Nested case-control
  - Within a cohort study, compares all cases to a subset of persons who did not develop disease
- Case-cohort
  - Within a cohort study, compares all cases to a random subsample of the cohort
  - Sub-cohort can be used for multiple case groups
- Super-cases and super-controls
  - Extremes of the phenotypes
  - Maximizes opportunity to detect signal

Cohort studies

- Identify individuals based on their exposure status, follow-up to ascertain disease/ outcome status

Exposed

Diseased

Non-exposed

Non-diseased

Time

Exposure assessment

Study initiation
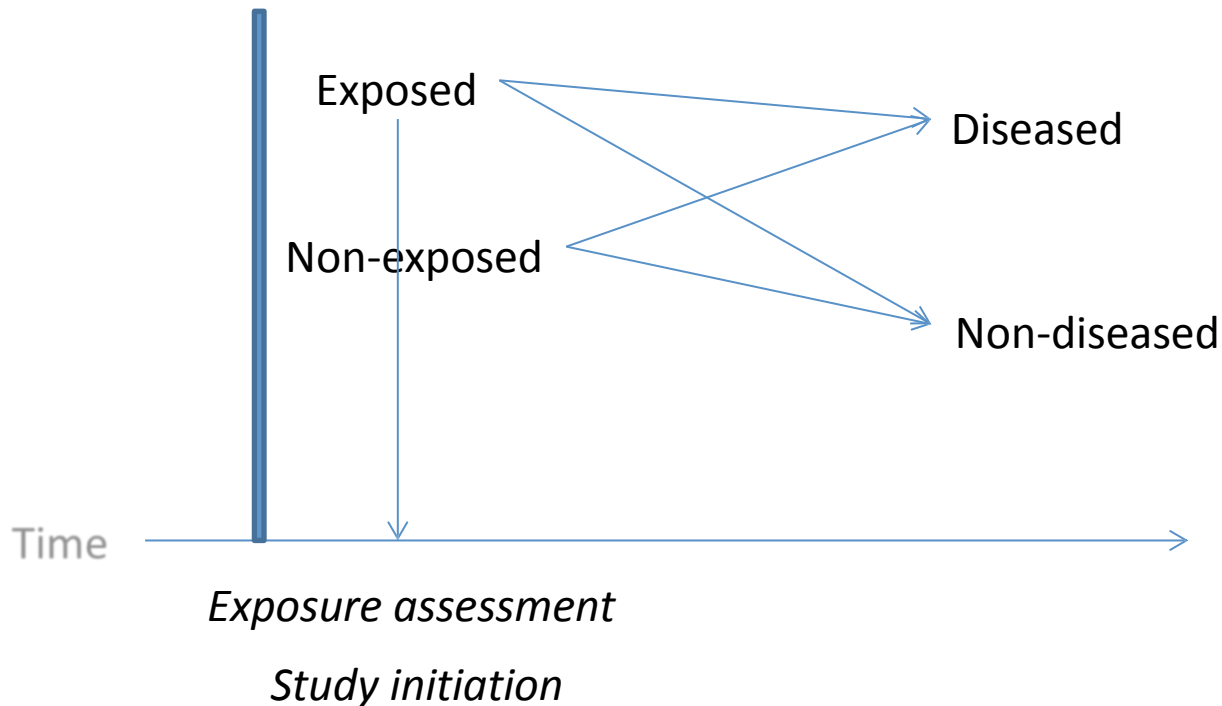
Assumptions

- Exposed and non-exposed groups are representative of a well-defined general population

- Outcome assessment comparable between exposed and non-exposed

Measure of genetic effects

- Cohort studies are often used for quantitative outcomes in genetic studies
  - BMI, eye colour, blood pressure
- Genetic model assumes additive genetic effects to test for association
  - r-fold increase in phenotype values for each risk allele
  - Uses linear regression with number of risk alleles as predictor and trait value as outcome
  - Trend test, 1df

Advantages of a cohort study in genetics

- Able to directly estimate disease incidence
- Optimal for short induction periods
    - Induction period = time from exposure to manifest disease
- Can look at multiple outcomes
- Potential to investigate natural history of disease
- Amenable to both quantitative and binary outcomes
- Risk factors ascertained prior to disease
- Ideal for gene*environment interaction analyses

Disadvantages of a cohort study

- Not suitable for rare exposures or rare outcomes

- Requires large populations

- May be more expensive, time consuming

Genetic associations

- True association

- Indirect association due to linkage disequilibrium

- Association due to random errors

- Association due to systematic errors

- Association due to random and systematic errors

Indirect tests of association using "tag SNP" genetic markers

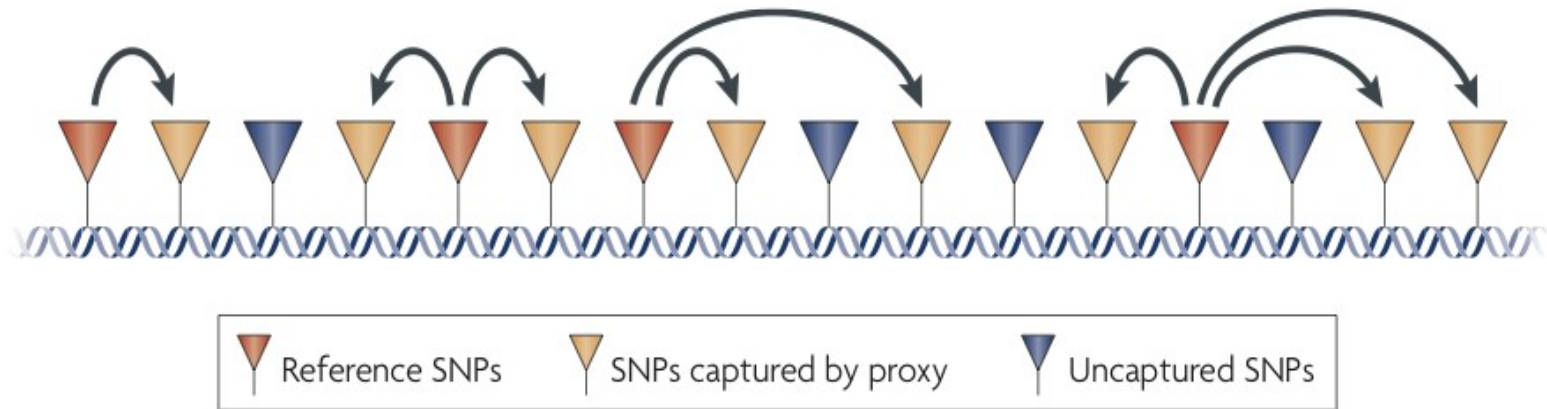Reference SNPs   SNPs captured by proxy   Uncaptured SNPs

Figure 3 | **Schematic of a genomic region to be tested for association with a phenotype.** The four reference SNPs in the mapping panel are indicated by red triangles; these are genotyped directly. The eight SNPs indicated by yellow triangles are captured through linkage disequilibrium (by proxy) with the reference SNPs denoted by arrows. The four SNPs indicated by blue triangles are neither genotyped nor in linkage disequilibrium with the reference SNPs; phenotypic association that is due to one of these would be missed.

Old and new problems

Small sample sizes

Small effect sizes

Large numbers of genetic variants

Absence of replication

Two different approaches

Genome-wide association studies

- High-throughput genotyping technologies to essay hundred of thousand of SNLs
- Hypothesis-free agnostic approach
- Millions of associations tested simultaneously
- Adjust for multiple comparisons
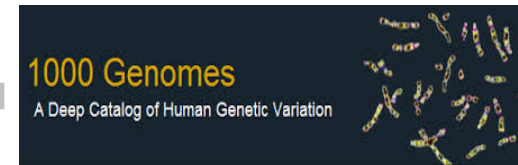- Two-stage or one-stage designs
- Replication

Candidate gene studies

- Research based on previous hypothesis
- Biological-functional background
- Ad hoc analysis of published results
- Replication

QC

MACH-IMPUTE

International HapMap Project

1000 Genomes
A Deep Catalog of Human Genetic Variation

~2.5M SNPs (HapMap)-~90M 1KG Project/HRC
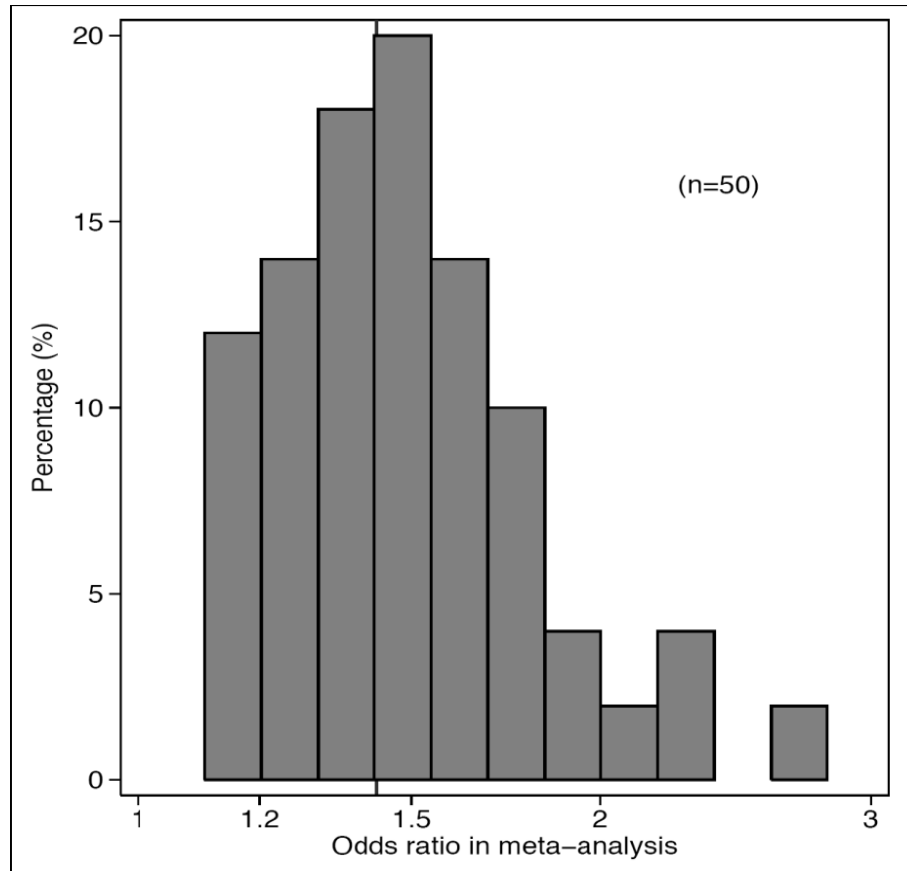
Association testing

Coverage and efficiency in current SNP chips

Table 1 Chip size, the lowest MAF covered by the chip, the number of non-synonymous SNPs, and design notes of recent Illumina and Affymetrix chips according to their datasheets provided by the companies

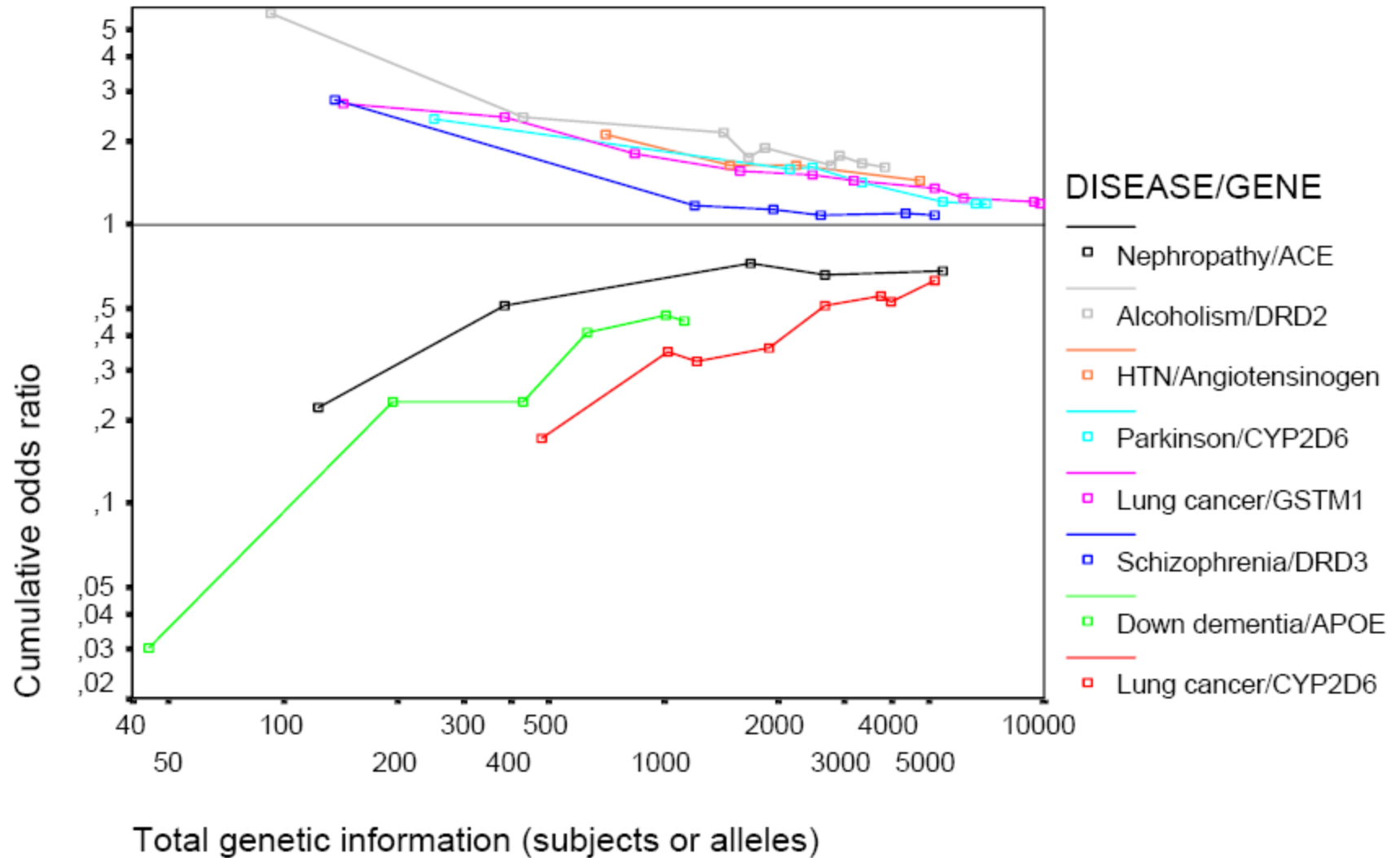|	Chip size in number (SNPs)	Lowest MAF captured	Number (non-synonymous SNPs)	Based on	Note
Affymetrix					
Axiom Genome-Wide Human EU (Axiom GW EU)	~600000	1%	10648	HapMap, Single Nucleotide Polymorphism database (dbSNP), 1000 GP	Targeting European population
Axiom Genome-Wide Human ASI (Axiom GW ASI)	~600000	1%	10346	HapMap, dbSNP, 1000 GP	Targeting Asian population
Axiom Genome-Wide Human CHB (Axiom GW CHB)	~1200000	2%	10560	HapMap, dbSNP, 1000 GP	Targeting CHB subpopulation
Axiom Genome-Wide Human PanAFR (Axiom GW PanAFR)	~2200000	2%	12250	HapMap, dbSNP, 1000 GP, Southern African Genomes Project	Targeting African population
Illumina					
Human OmniExpress	~700000	5%	15062	HapMap	Optimized tag SNP
Human Omni1S-8	~1000000	5%	5641	1000GP	Optimized tag SNP
Human Omni2.5-8	~2500000	2.5%	41900	1000GP	Targeting common and rare variants
Human Omni2.5S-8	~2500000	1%	57360	1000GP	Targeting rare variants

Estimated effect sizes in genetic epi are small

GENE	Polymorphism	Fixed effects OR (95% CI)
---	rs9300039[a]	1.25 (1.15-1.37)
FTO	rs8050136	1.17 (1.12-1.22)
PPARG	rs1801282	1.14 (1.08-1.20)
CDKAL1	rs10946398[b]	1.12 (1.08-1.16)
SLC30A8	rs13266634	1.12 (1.07-1.16)
CDKN2B	rs564398	1.12 (1.07-1.17)
HHEX	rs5015480- rs1111875	1.13 (1.08-1.17)
KCNJ11	rs5215[c]	1.14 (1.10-1.19)
IGF2BP2	rs4402960	1.14 (1.10-1.18)
CDKN2B	rs10811661	1.20 (1.14-1.25)
TCF7L2	rs7901695[d]	1.37 (1.31-1.43)

Absence of replication

Absence of replication

Nature 1994
TNFA associates with
cerebral malaria
>1000 εώς σήμερα

Legend:
- Data on different diseases and genes
- Data on same disease only
- Data on same gene only
- Methods
- Replications
- Review, editorial, letter, news
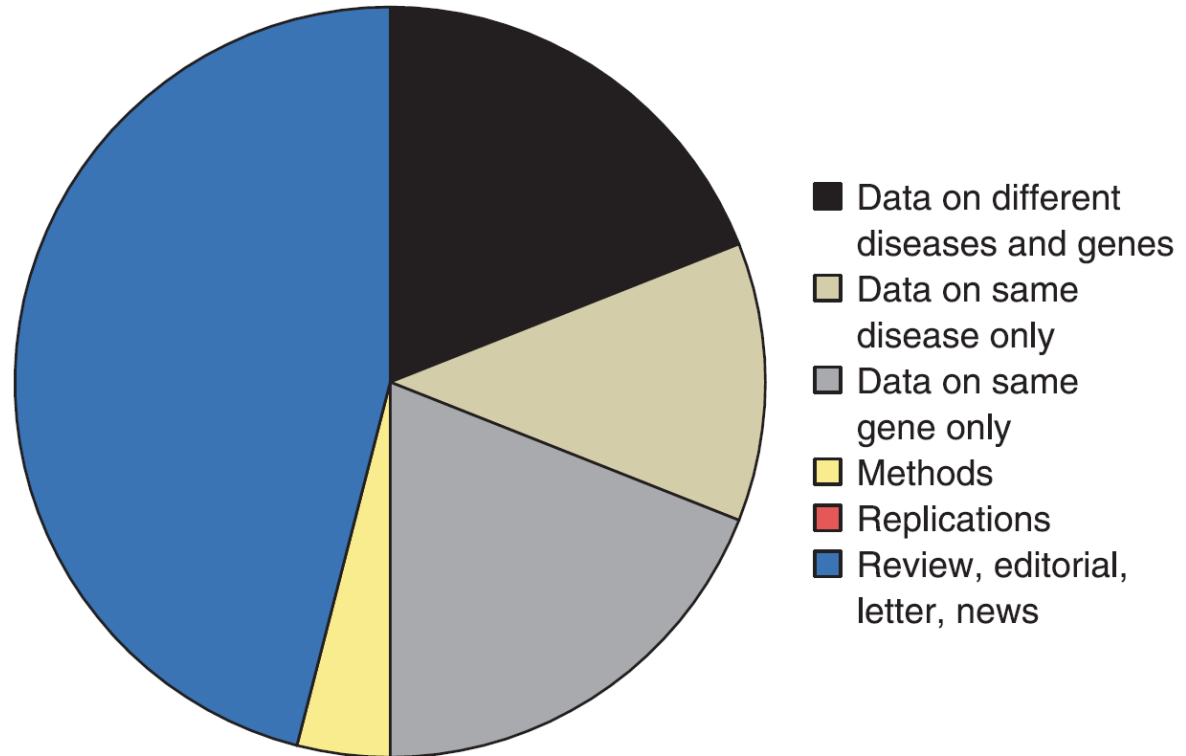
Replication efforts

(b) Genome-wide association findings for Parkinson disease

Legend:
- Data on different diseases and genes
- Data on same disease only
- Data on same gene only
- Methods
- Replications
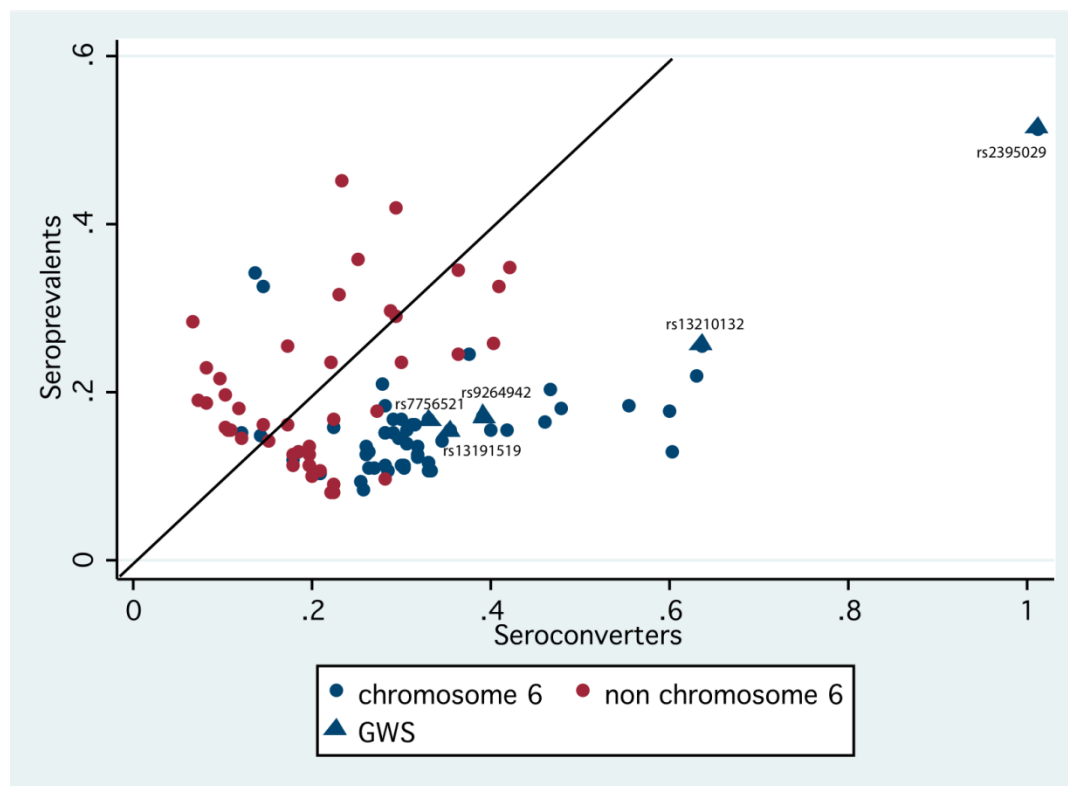- Review, editorial, letter, news
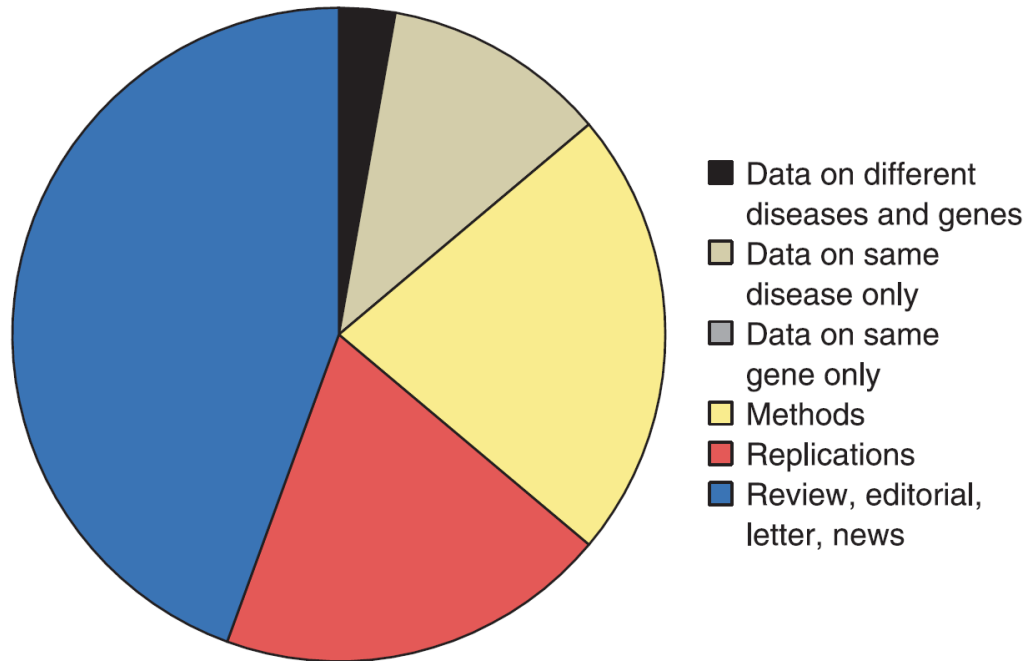
Definition of the phenotype

HIV GWAs

Differencies in the definition of the phenotype, can cause differences in the effect sizes of the associations

The estimate of the genetic effect was larger in seroconverters

Could be considered in optimizing power for discovering new associations

Evangelou et al. AJE, 2011

International consortia

Consortium	Disease/Trait	Teams	Participants
GEFOS	Osteoporosis	40	133000
TREATOA	Osteoarthritis	20	30000
GEOPD	Parkinson's	20	12000
DIAGRAM	Σ. Διαβήτης	30	100000
GIANT	Ύψος, βάρος	80	250000

Mega-Analysis

ARTICLE

New genetic loci link adipose and insulin biology to body fat distribution

⊖ Collaborators (2208)

Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, Henneman P, Heid IM, Kizer JR, Lyytikainen LP, Fuchsberger C, Tanaka T, Morris AP, Small K, Isaacs A, Beekman M, Coassin S, Lohman K, Qi L, Kanoni S, Pankow JS, Uh HW, Wu Y, Bidulescu A, Rasmussen-Torvik LJ, Greenwood CM, Ladouceur M, Grimsby J, Manning AK, Liu CT, Kooner J, Mooser VE, Vollenweider P, Kapur KA, Chambers J, Wareham NJ, Langenberg C, Frants R, Willemsvan-vanDijk K, Oostra BA, Willems SM, Lamina C, Winkler T, Psaty BM, Tracy RP, Brody J, Chen I, Viikari J, Kähönen M, Pramstaller PP, Evans DM, St Pourcain B, Sattar N, Wood A, Bandinelli S, Carlson OD, Egan JM, Böhringer S, van Heemst D, Kedenko L, Kristiansson K, Nuotio ML, Loo BM, Harris T, Garcia M, Kanaya A, Haun M, Klopp N, Wichmann HE, Deloukas P, Katsareli E, Couper DJ, Duncan BB, Kloppenburg M, Adair LS, Borja JB, Wilson JG, Musani S, Guo X, Johnson T, Semple R, Teslovich TM, Allison MA, Redline S, Buxbaum SG, Mohlke KL, Meulenbelt I, Ballantyne CM, Dedoussis GV, Hu FB, Liu Y, Paulweber B, Spector TD, Slagboom P, Ferrucci L, Jula A, Perola M, Raitakari O, Florez JC, Salomaa V, Eriksson JG, Frayling TM, Hicks AA, Lehtimäki T, Smith GD, Siscovick DS, Kronenberg F, van Duijn C, Loos RJ, Waterworth DM, Meigs JB, Dupuis J, Richards JB, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier JB, Johansson Å, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikäinen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney AS, El Mokhtari N, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Müller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schäfer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, van der Schoot CE, Wagner PJ, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle

Era of Biobanks

- Population-based
  - UK Biobank, Japanese biobank
- Deep phenotyping
  - Cartagene

Large sample sizes

Genetic analysis of over one million people identifies 535 novel loci for blood pressure.

Evangelos Evangelou, Helen R. Warren, David Mosen-Ansorena, Borbala Mifsud, Raha Pazoki, He Gao, Georgios Ntritsos, Niki Dimou, Claudia P. Cabrera, Ibrahim Karaman, Fu Liang Ng, Marina Evangelou, Katarzyna Witkowska, Evan Tzanis, Jacklyn N. Hellwege, Ayush Giri, Digna R. Velez Edwards, Yan V. Sun, Kelly Cho, J. Michael Gaziano, Peter W. F. Wilson, Philip S. Tsao, Csaba P. Kovesdy, Tonu Esko, Reedik Magi, Lili Milani, Peter Almgren, Thibaud Boutin, Stephanie Debette, Jun Ding, Franco Giulianini, Elizabeth G. Holliday, Anne U. Jackson, Ruifang Li-Gao, Wei-Yu Lin, Jian'an Luan, Massimo Mangino, Christopher Oldmeadow, Bram Prins, Yong Qian, Muralidharan Sargurupremraj, Nabi Shah, Praveen Surendran,
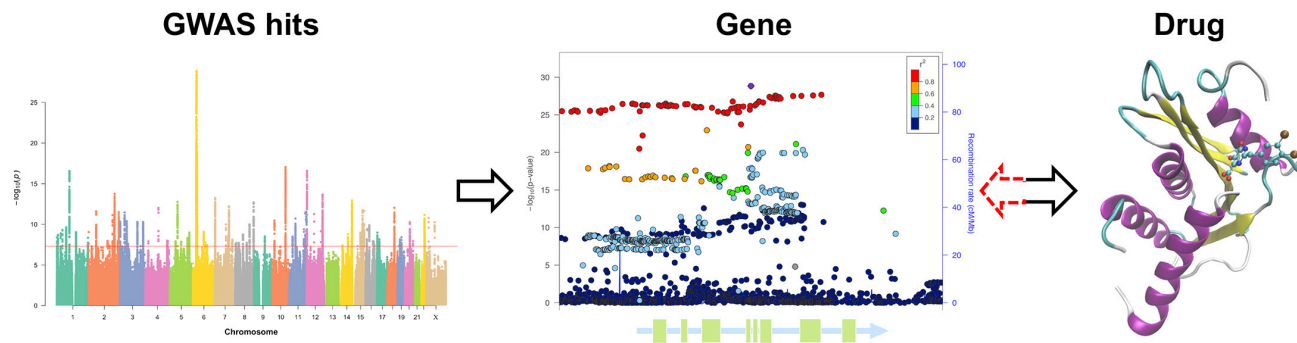
More problems

- Large number of identified genes
- Inheritance models
- Gene-gene interactions
- Gene-environment interactions
- Errors in genotyping and phenotyping
- Systematic errors

GWAS SNP-Trait Discovery Timeline
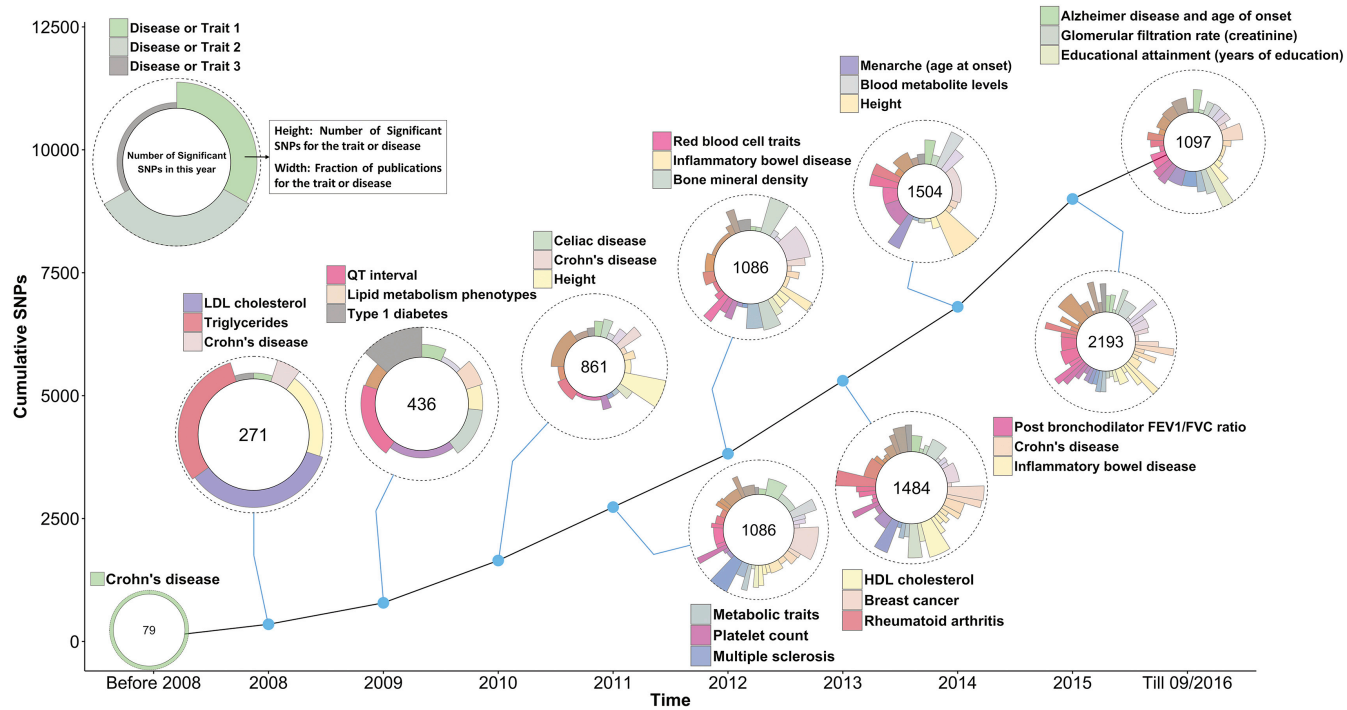
Examples of links between GWAS discoveries and drugs

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

Gene-Gene interactions
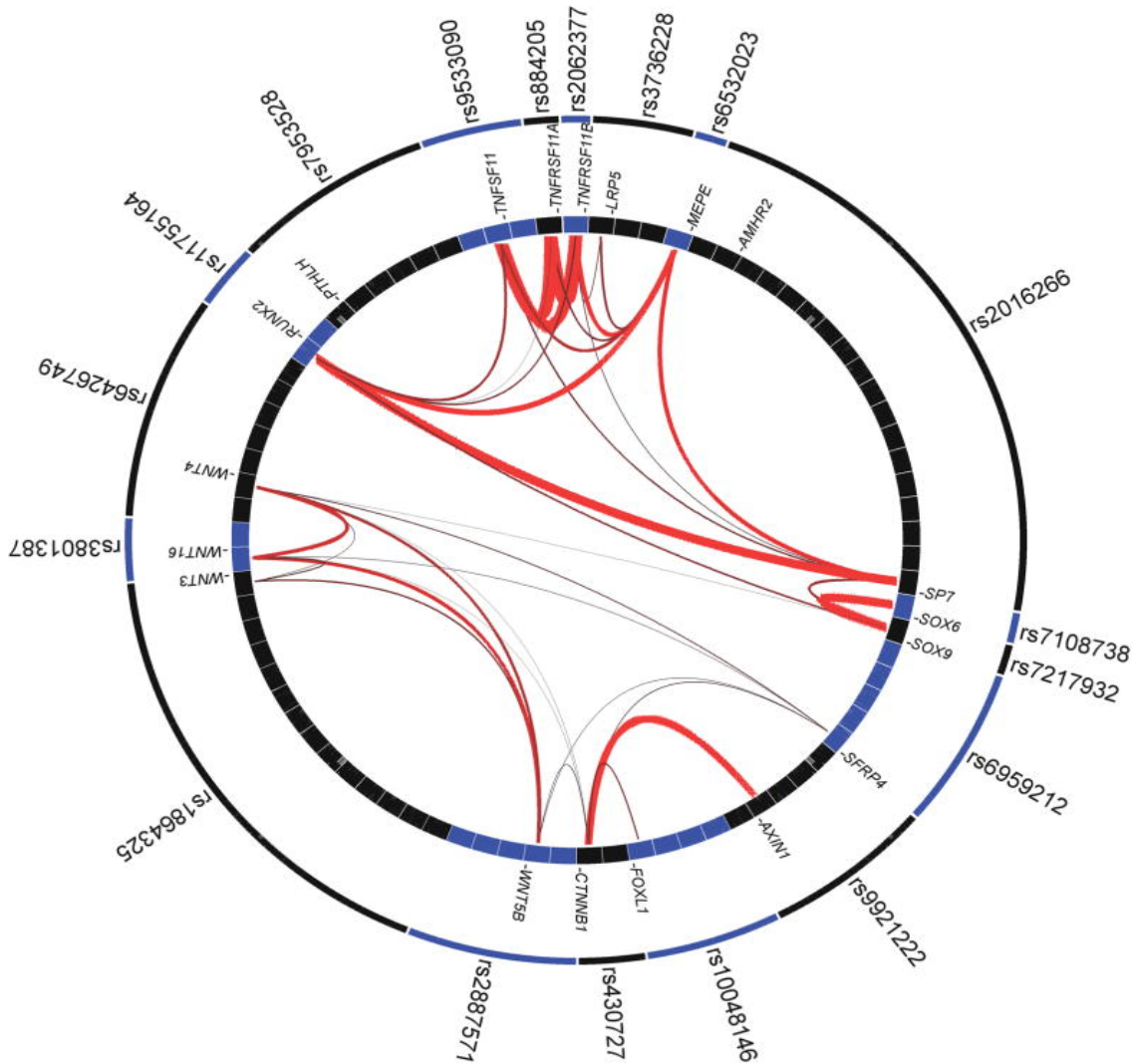
Large sample sizes are required to support evidence of gene-gene interactions

Gene-environment interactions
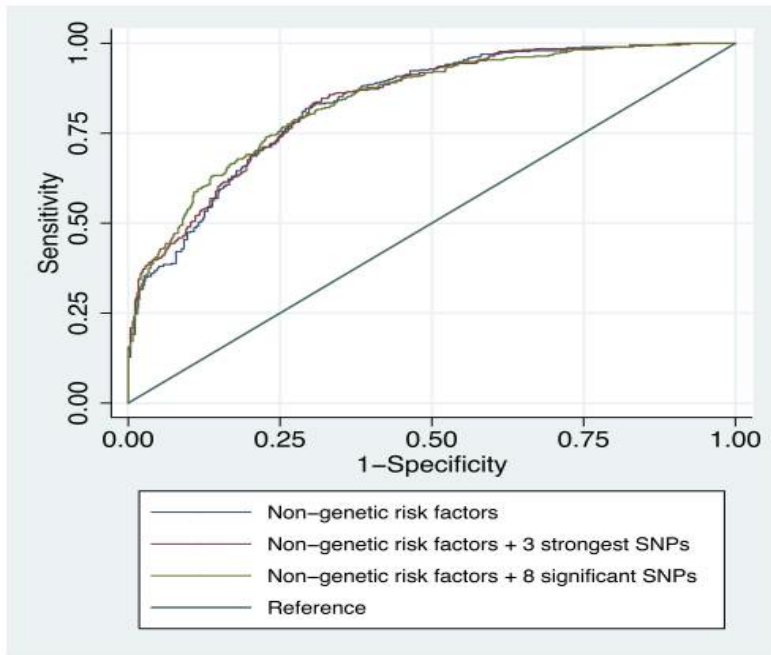
Stefanaki I et al. PLoS One; 2013

Kypreou KP et al. J Invest Dermatol; 2016

Table 2. Risk prediction performance for the four different models of predictors in the Greek dataset

	AUC	95% CI
Phenotypic risk factors only[1]	0.764	0.741−0.787
Phenotypic risk factors + GRS$_{GWS}$	0.775	0.752−0.797
Phenotypic risk factors + GRS$_{ALL}$	0.775	0.752−0.798

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; GRS, genetic risk score; GWS, genome-wide significant.

[1]Risk factors are sex, age, eye color, hair color, skin color, phototype, and tanning ability.

Whole exome and whole genome sequencing

Copyright © 2012 University of Washington
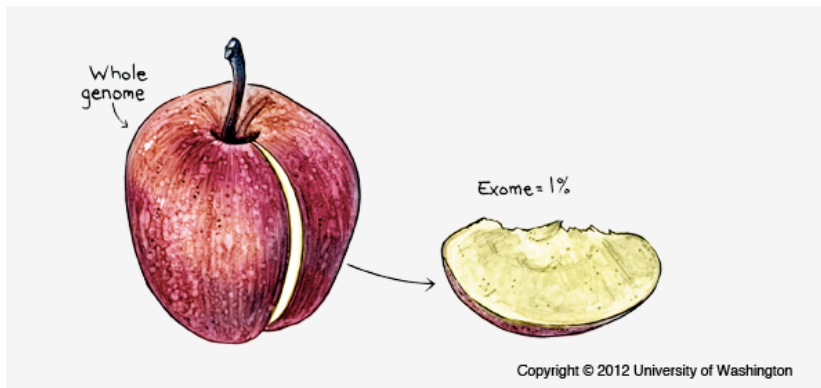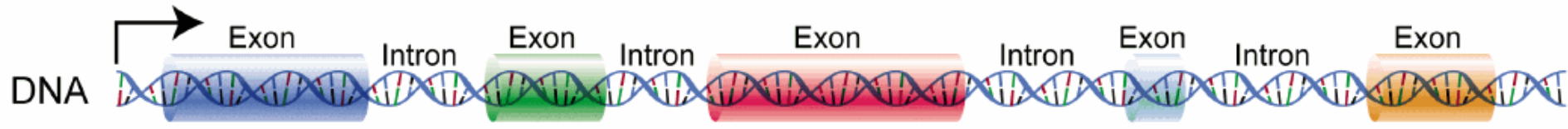
Human Genome Epidemiology (HuGE) Review

Genome-wide Significant Associations for Variants With Minor Allele Frequency of 5% or Less—An Overview: A HuGE Review

Orestis A. Panagiotou, Evangelos Evangelou, and John P. A. Ioannidis*

In the near future
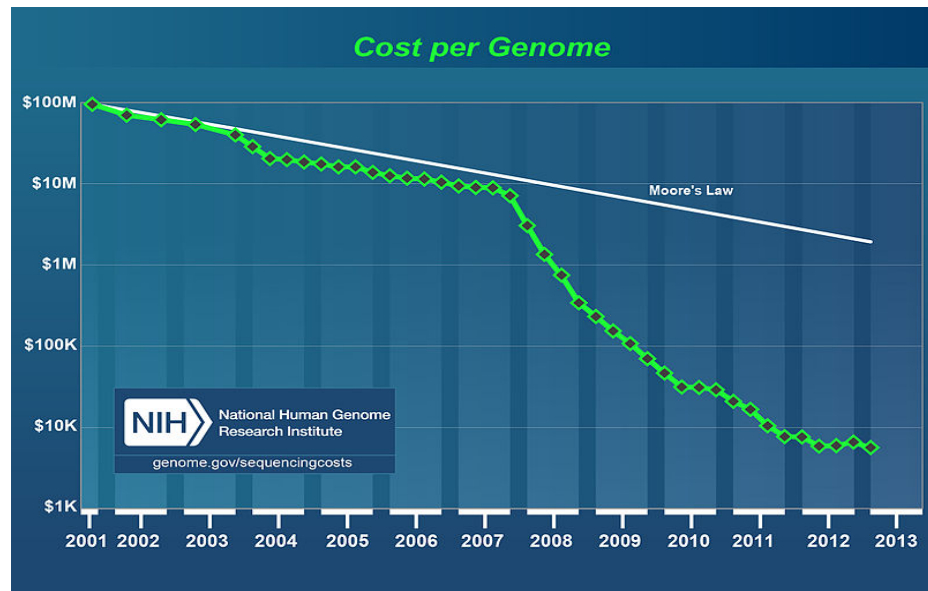
- Exome sequencing-Whole genome sequencing

Cost reductions

⬇

Personal Genome

⬇

Precise Medicine

Cost per Genome

Moore's Law

NIH National Human Genome Research Institute

genome.gov/sequencingcosts

Clinical assessment incorporating a personal genome

Euan A Ashley, Atul J Butte, Matthew T Wheeler, Rong Chen, Teri E Klein, Frederick E Dewey, Joel T Dudley, Kelly E Ormond, Aleksandra Pavlovic, Alexander A Morgan, Dmitry Pushkarev, Norma F Neff, Louanne Hudgins, Li Gong, Laura M Hodges, Dorit S Berlin, Caroline F Thorn, Katrin Sangkuhl, Joan M Hebert, Mark Woon, Hersh Sagreiya, Ryan Whaley, Joshua W Knowles, Michael F Chou, Joseph V Thakuria, Abraham M Rosenbaum, Alexander Wait Zaranek, George M Church, Henry T Greely, Stephen R Quake, Russ B Altman
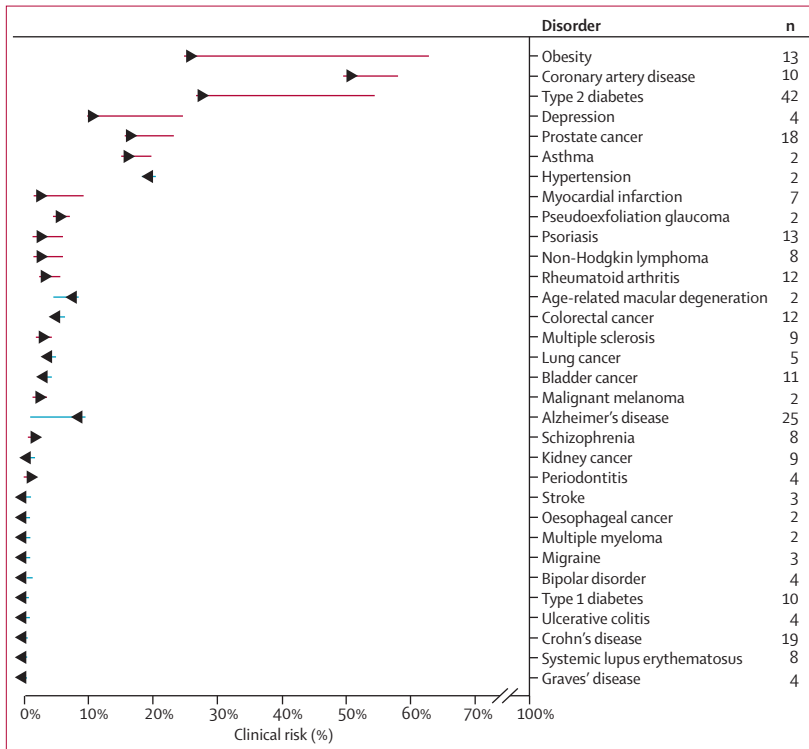
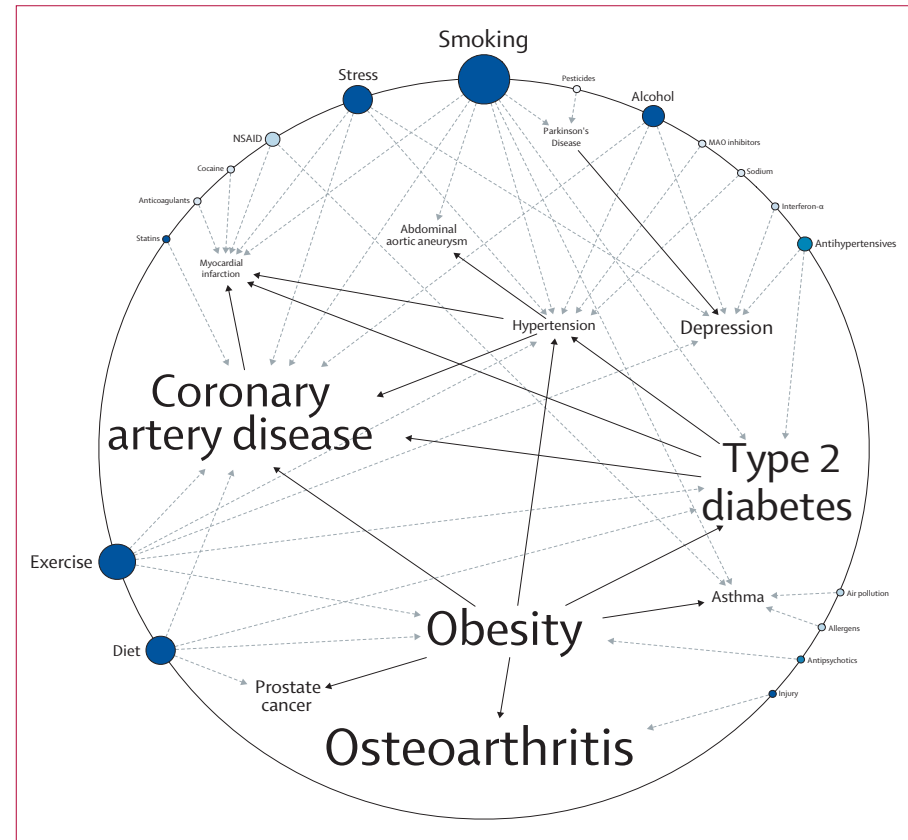*Figure 3:* **Clinical risk incorporating genetic-risk estimates for major diseases**

Figure 5: **Gene-environment interaction**

HEALTH RISKS

✳ 23andMe Research Discoveries were made possible by 23andMe members who to

SHOW RESULTS FOR [Cyrus Farivar ⇕]   ED REPORTS »

Cyrus:

Average:

This is the estimated lifetime incidence
of Gout for someone with Cyrus's genotype
compared to average.

Read more »

NOTE: This result applies to people
of European ancestry. We cannot yet estimate
risk for those with Multiple ancestries
ancestry. (more)

Elevated Risk ⌾

NAME	CONFIDENCE	YOUR			AVERAGE
Gout	★★★★	35.			
Alzheimer's Disease	★★★★	12.6%	7.2%	1.75x	▬
Chronic Kidney Disease	★★★★	5.0%	3.4%	1.45x	▪
Restless Legs Syndrome	★★★★	2.5%	2.0%	1.25x	▪
Exfoliation Glaucoma	★★★★	2.2%	0.7%	2.90x	▪
Celiac Disease	★★★★	0.59%	0.12%	4.98x	▪
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★	0.43%	0.36%	1.21x	▪
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★	0.28%	0.23%	1.22x	▪

Rare variants with large effects

MAF=0.2%, OR=50.6