Genetic Data and Personalized Medicine

Panel 5:

Moderator: Maryam Daneshpour
Members: Azita Zadehvakili, Ashraf Mohammad-Khani, Abdolazim Nejatizadeh, Mehdi Mirzayii, Bhareh Sedaghati Khayat, Kamran Guity, Alireza Haghighi

Outline

- Define the role of genetic study in cohorts
- How to establish a standard genomic bank?
- How to collect and develop family relationships in a cohort
- Share the genetic data and define the genomic map
- Create a database of phenotypes in existing cohort
- How to control data quality and integration of genomic data?
- Draw a road map for personalized medicine in the context of cohort studies

Personalized medicine

- Personalized medicine is a medical procedure that separates patients into different groups with:
  - Medical decisions
  - Practices
  - Interventions
  - and/or products
- Being tailored to the individual patient based on their predicted response or risk of disease.
- The terms personalized medicine, precision medicine, stratified medicine and P4 medicine are used interchangeably to describe this concept

President Obama's Precision Medicine Initiative

State of the Union Address, 20 January 2015, USA — Budget : **215 M$**

70 M$

Focus CANCER

- With NCI
- Building a *Cancer Knowledge Network*

130 M$

Creation COHORT

- Voluntary National Research Cohort
- 1 million volunteers

5 M$

PRIVACY

- Input from Experts
- Patients groups, bioethicists, advocates…

10 M$

REGULATORY

Modernization

- FDA thinks of new evaluation of NGS

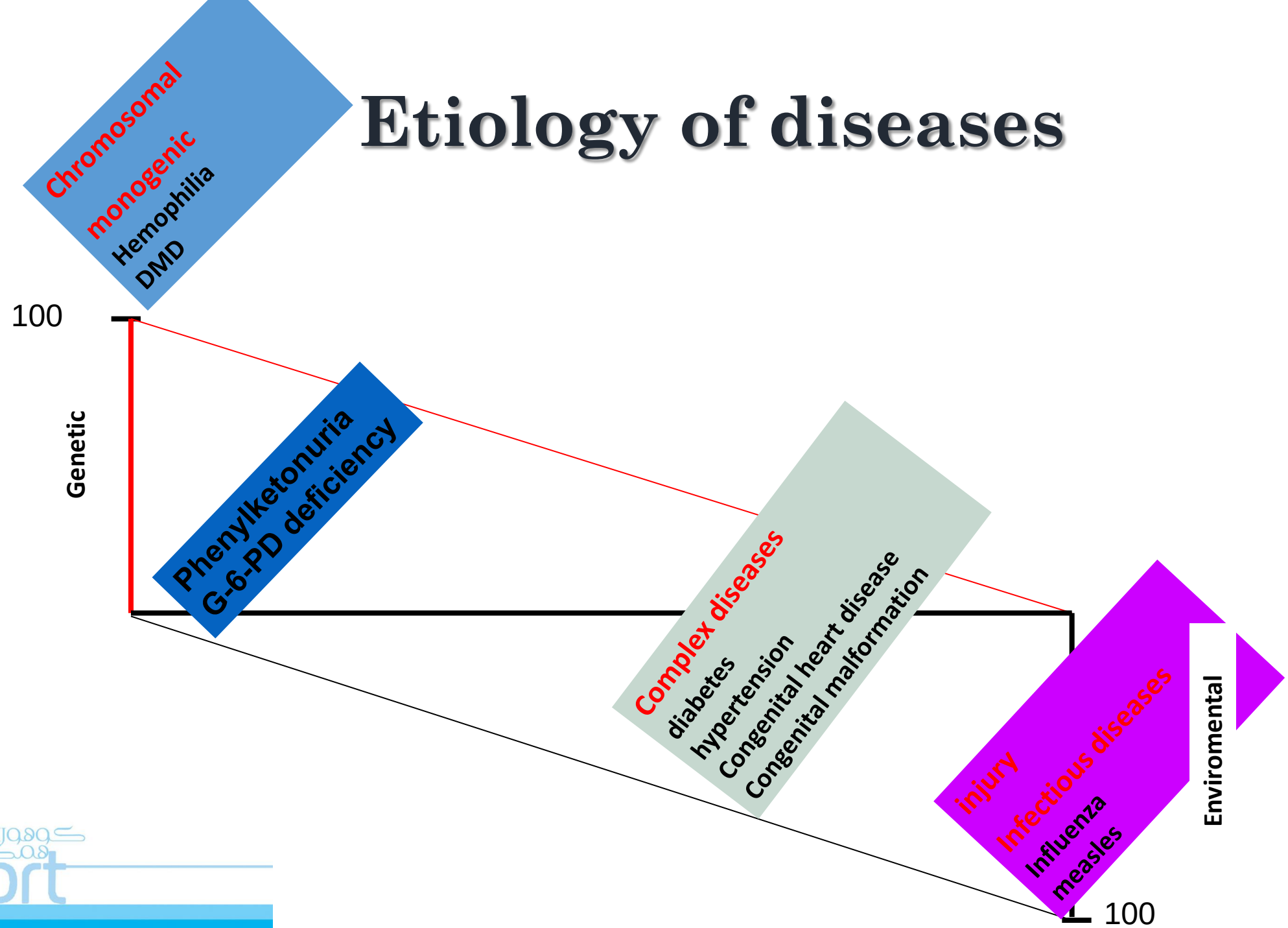
PARTNERSHIP

Public / Private partnerships

https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

Role of genetic study in cohorts

Etiology of diseases

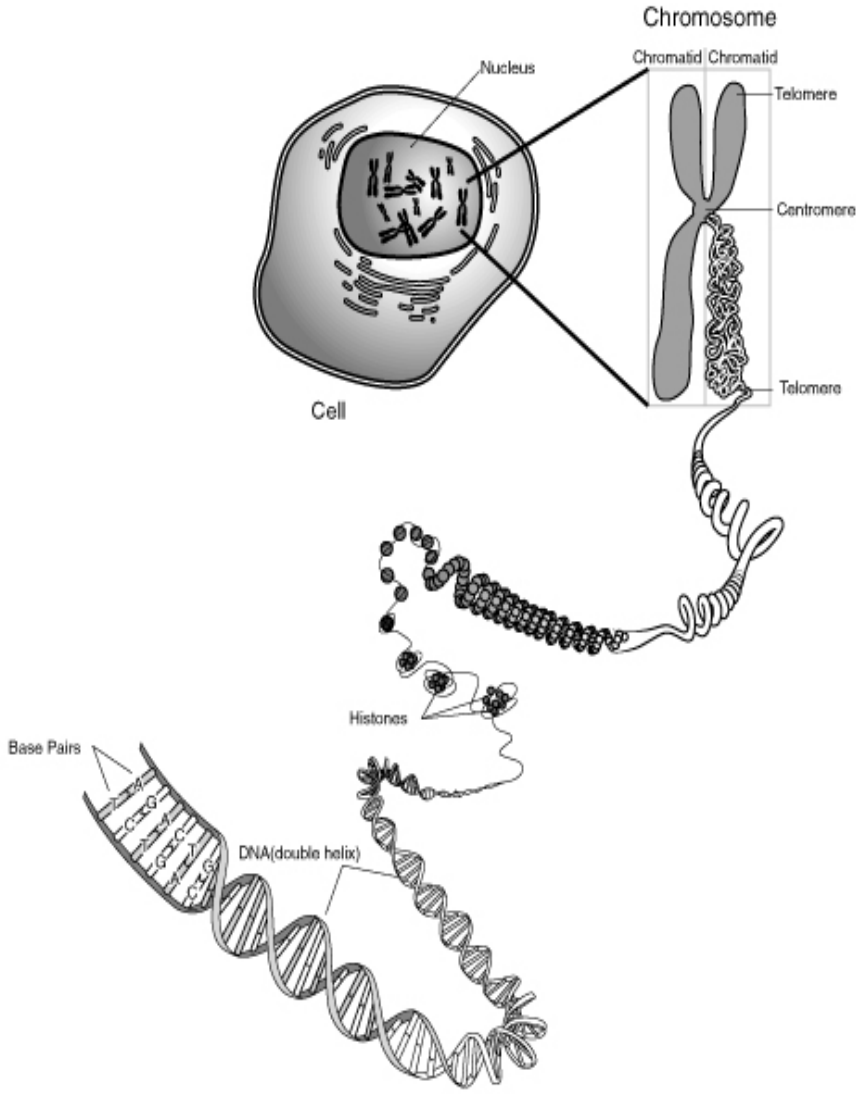


Penetrance and Environmental Factors

- Highly penetrant Mendelian single gene diseases
  - Huntington's Disease caused by excess CAG repeats in huntingtin's protein gene
  - Autosomal dominant, 100% penetrant, invariably lethal
- Reduced penetrance, some genes lead to a predisposition to a disease
  - BRCA1 & BRCA2 genes can lead to a familial breast or ovarian cancer
  - Disease alleles lead to 80% overall lifetime chance of a cancer, but 20% of patients with the rare defective genes show no cancers
- Complex diseases requiring alleles in multiple genes
  - Many cancers (solid tumors) require somatic mutations that induce cell proliferation, mutations that inhibit apoptosis, mutations that induce angiogenesis, and mutations that cause metastasis
  - Cancers are also influenced by environment (smoking, carcinogens, exposure to UV)
  - Atherosclerosis (obesity, genetic and nutritional cholesterol)
- Some complex diseases have multiple causes
  - Genetic vs. spontaneous vs. environment vs. behavior
- Some complex diseases can be caused by multiple pathways
  - Type 2 Diabetes can be caused by reduced beta-cells in pancreas, reduced production of insulin, reduced sensitivity to insulin (insulin resistance) as well as environmental conditions (obesity, sedentary lifestyle, smoking etc.).

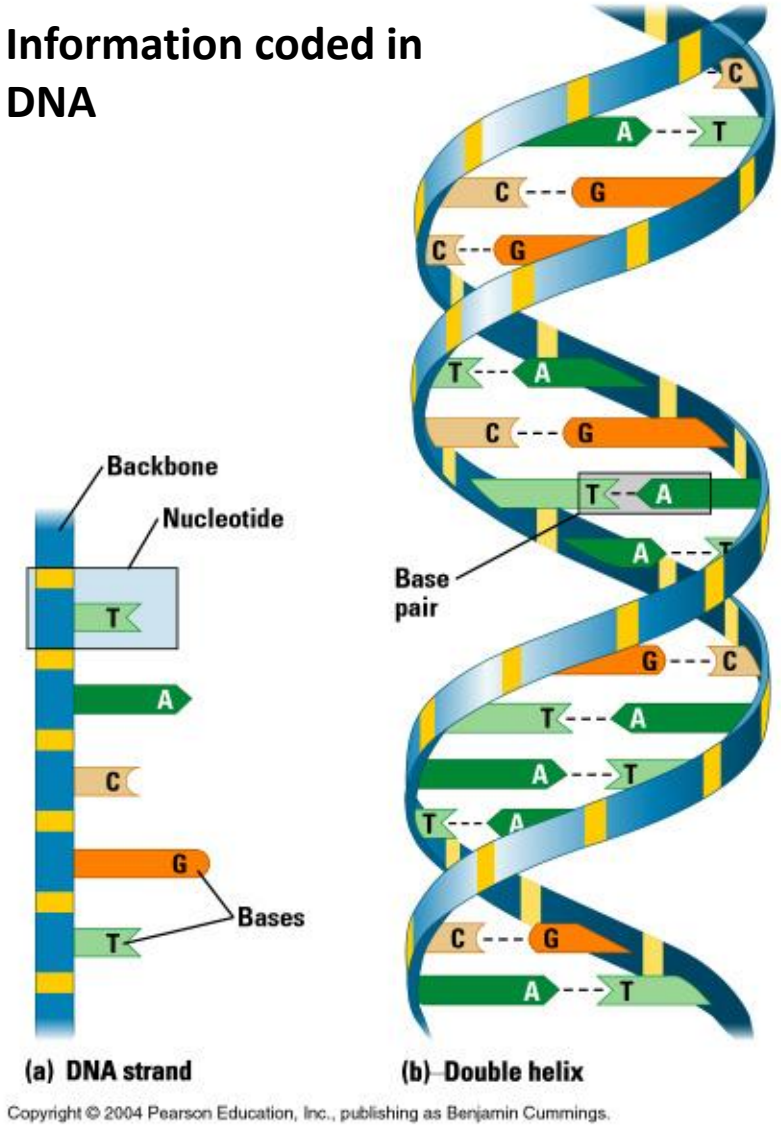
The genome is our Genetic Blueprint

- Nearly every human cell contains 23 pairs of chromosomes
  - 1 - 22 and XY or XX
    - XY = Male
    - XX = Female
- Length of chr 1-22, X, Y together is ~3.2 billion bases (about 2 meters diploid)



The Genome is Who We Are on the inside!

- Chromosomes consist of DNA
  - molecular strings of A, C, G, & T
  - base pairs, A-T, C-G
- Genes
  - DNA sequences that encode proteins
  - less than 3% of human genome

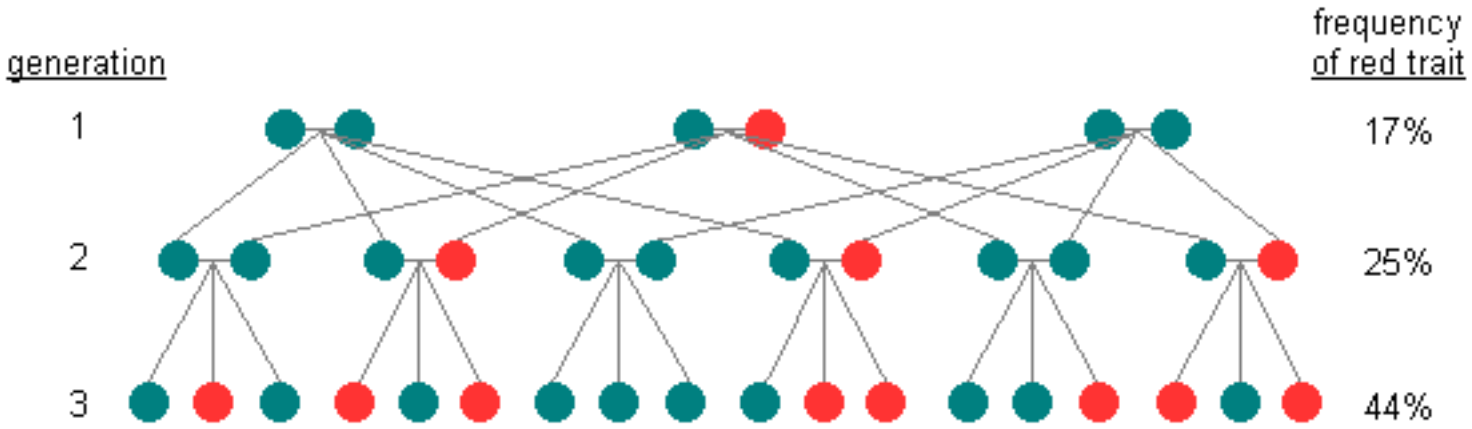


Issues in genetic association studies

- Many genes
  - ~20,000-25,000 genes, many can be candidates
- Many SNPs
  - ~12,000,000 SNPs, ability to predict functional SNPs is limited
- Methods to select SNPs:
  - Only functional SNPs in a candidate gene
  - Systematic screen of SNPs in a candidate gene
  - Systematic screen of SNPs in an entire pathway
  - Genomewide screen
  - Systematic screen for all coding changes

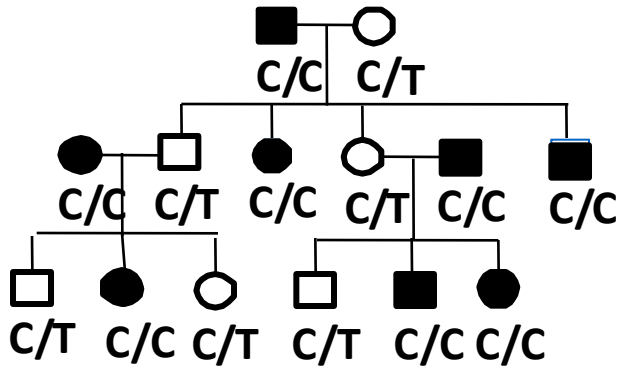
Population Genetics

- Chromosome pairs segregate and recombine in every generation
- Every allele of every gene has its own independent evolutionary history (and future!)
- Frequencies of various alleles differ in different sub-populations of people.



Human Genetic Analysis

Families
Linkage Studies



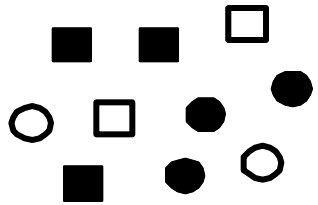
Simple Inheritance (Segregate)

Single Gene with Major Effect

Variant Rare in the Population

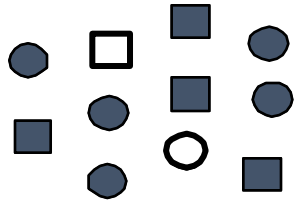
~600 Short Tandem Repeat Markers

Populations
Association Studies



40% T, 60% C

Cases



15% T, 85% C

Controls

Complex Inheritance (Aggregate)

Multiple Genes with Small Contributions and Environmental Contexts

Variant(s) Common in the Population

Polymorphic Markers > 500,000 -1,000,000 Single Nucleotide Polymorphisms (SNPs)

Genomic landscape of variation

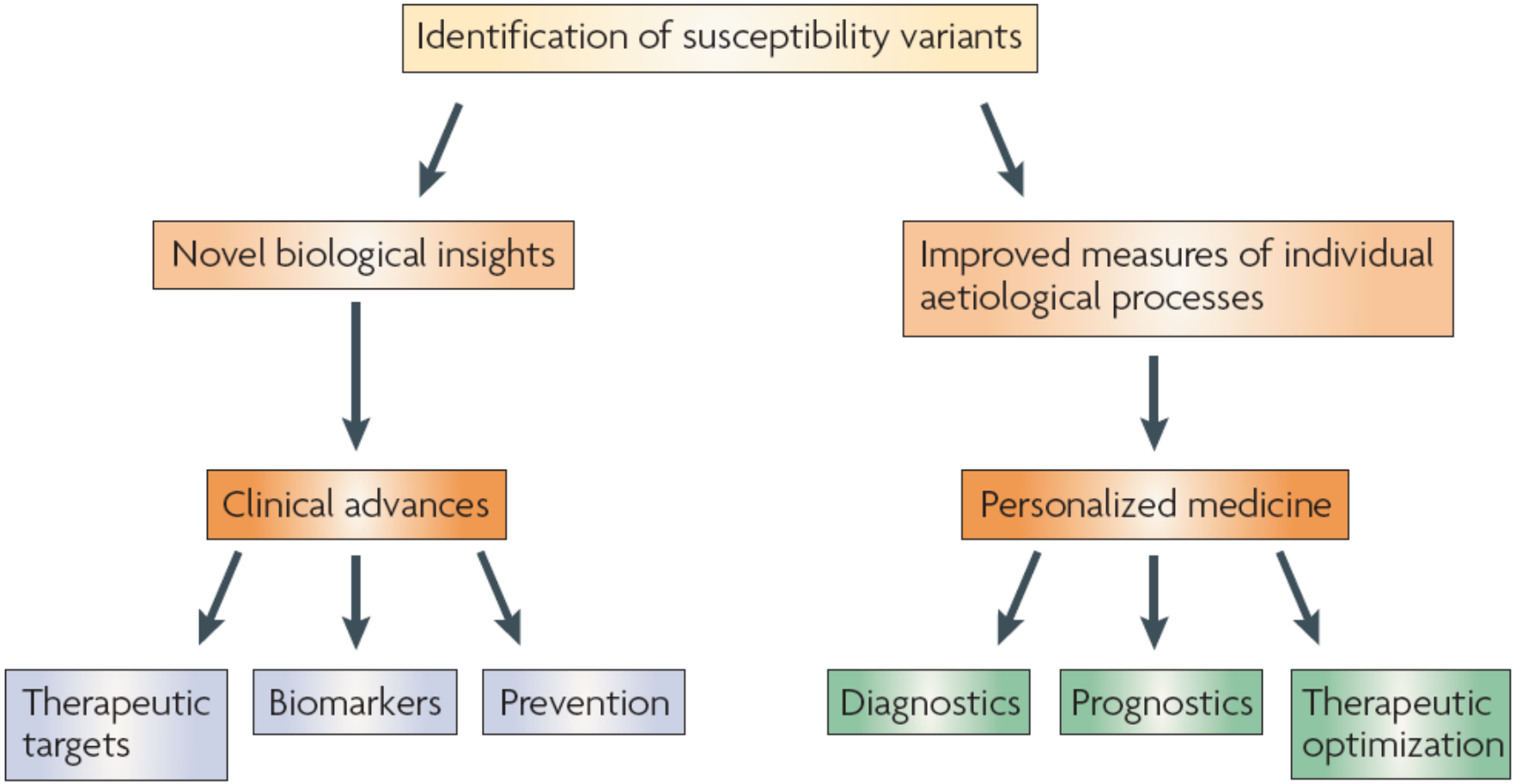
- The average genome (2x 3 billion bases) contains:
  - Base substitutions, compared to the human reference genome
    - ~ 10 million
  - Small insertions and deletions, 'indels' (1-100bp)
    - ~0.5 million
  - Larger deletions, duplications and insertions (>100bp)
    - ~5,000
- Variation in (~20,000) genes - the 'exome'
  - Variants
    - ~18,000
  - Variants might be expected to influence gene function
    - ~8-9,000
  - ~95% of variants are common in the population, only 5% are 'novel'
  - ~500-1000 genes contain novel, potentially functional variants
  - ~100-200 genes contain variants that unambiguously 'knock-out' the gene
  - ~20-40 genes contain novel 'knock-out' variants

GWAS

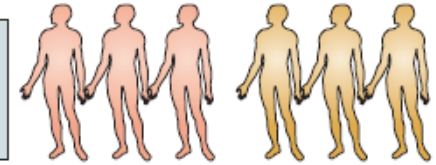
- A genome-wide association study is an approach that involves rapidly scanning markers across the complete sets of DNA, or genomes, of many people to find genetic variations associated with a particular disease.
- Once new genetic associations are identified, researchers can use the information to develop better strategies to detect, treat and prevent the disease. Such studies are particularly useful in finding genetic variations that contribute to common, complex diseases, such as asthma, cancer, diabetes, heart disease and mental illnesses.

http://www.genome.gov/20019523

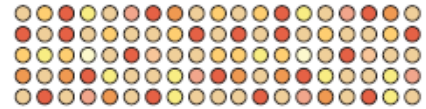
Potential of GWAS



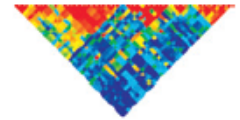
Large cohort of cases and controls (n >1,000)
- Matched for confounding variables, such as race, ethnicity and sex
- Stratified in order to maximize signals



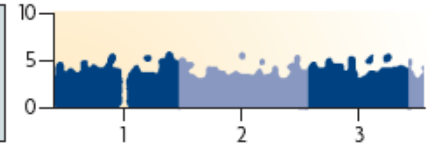
Microarray-based SNP genotyping
- ~1 million random marker SNPs or ~25,000 risk-enhancing SNPs (for example, nsSNPs)



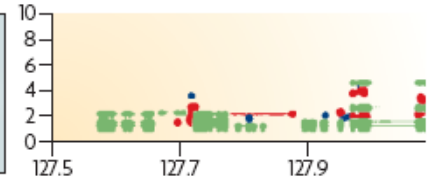
Derivation of haplotypes
- Predicated on International HapMap



Detection of association signals
- χ^2 or similar test
- Uncorrected P <10^{-7} or false discovery rate-like correction



Fine mapping of association signal (see FIG. 2)
- Directed genotyping of additional SNPs in region
- Fine mapping of LD in region of association
- Empirical derivation of haplotypes
- Examination of effect of stratification, if available



Replication of association
- Large independent cohort of cases and controls (n >1,000)
- Genotyping of nominated candidate SNPs (<20)
- χ^2 or similar test; replication of initial signal

Genotypes	CC	AA	CA	Total
Cases observed	59	27	98	184
Controls observed	60	89	36	185
Total	119	116	134	369

Biological validation of association
- Identification of risk-enhancing variant
- Examination of functional consequence of variant
- Determination of mechanism of risk-enhancement

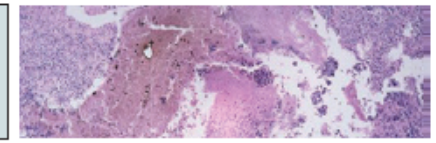


Figure 1 | **Overview of the general design and workflow of a genome-wide association (GWA) study.** The discovery phase entails genotyping many case and control DNA samples and evaluation for significant associations. The replication phase involves fine mapping of association signals and independent confirmation in a second cohort. Biological validation is important for translation of GWA findings into diagnostic or therapeutic discoveries.

GWAS ➡Improved Health?

1. Use of genetic information regarding common disease to individualize providers' approach to patients and change patients' behaviors in ways that lead to improved health ("**Personalized Medicine**").
2. Use of genetic information regarding common disease to understand the biology of human disease to lead to improved **diagnostic, therapeutic, and preventive** approaches.

Selection of SNPs
(Genome-wide association studies)

- Molecular
  - Higher requirements: Affymetrix and Illumina
- Analytical
  - Highest requirements: Data management, automation
- Advantages
  - No biological assumptions and can identify novel genes/pathways
  - Excellent chance to identify risk alleles
  - Utility in individual risk assessment
- Disadvantages
  - High costs
  - Concern of multiple tests

SNP Selection

a Direct:
catalogue and test all functional variants for association

b Indirect:
use a dense SNP map and test for linkage disequilibrium

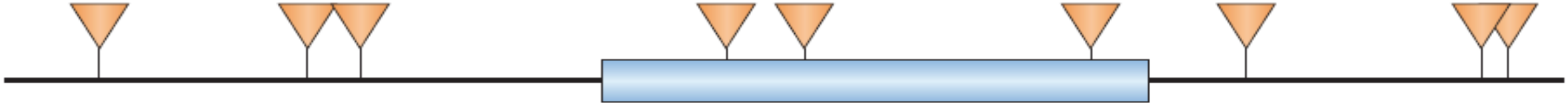


Figure 2 | **Alternative designs for genome-wide association studies. a** | Direct approach of testing a catalogue of all common functional variants in the genome. **b** | Indirect approach of testing a dense map of SNPs and relying on linkage disequilibrium to detect associations that are due to untested functional variants.

SNP Selection

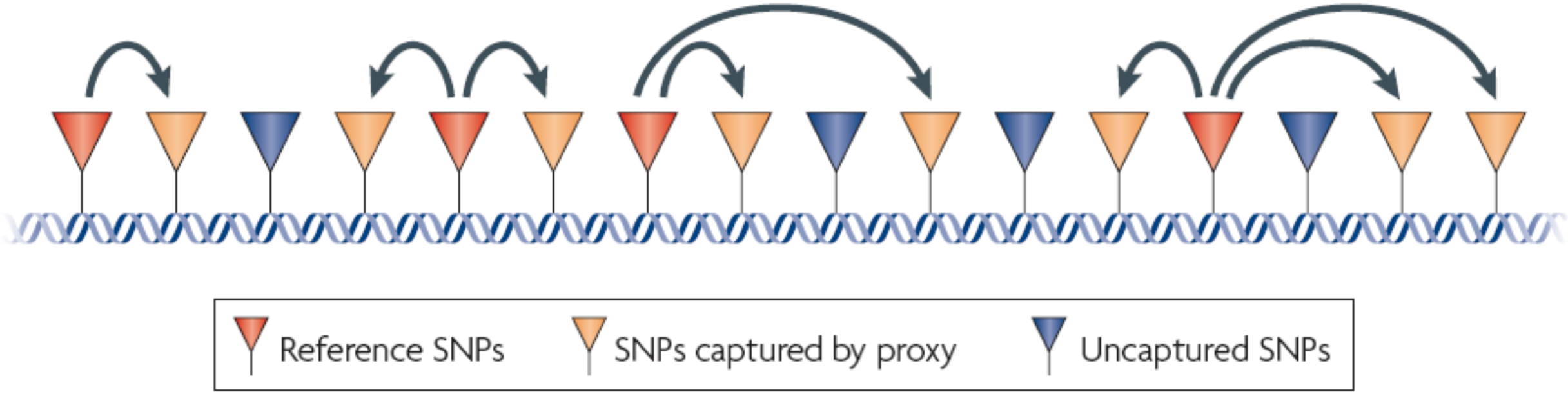


Figure 3 | **Schematic of a genomic region to be tested for association with a phenotype.** The four reference SNPs in the mapping panel are indicated by red triangles; these are genotyped directly. The eight SNPs indicated by yellow triangles are captured through linkage disequilibrium (by proxy) with 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.

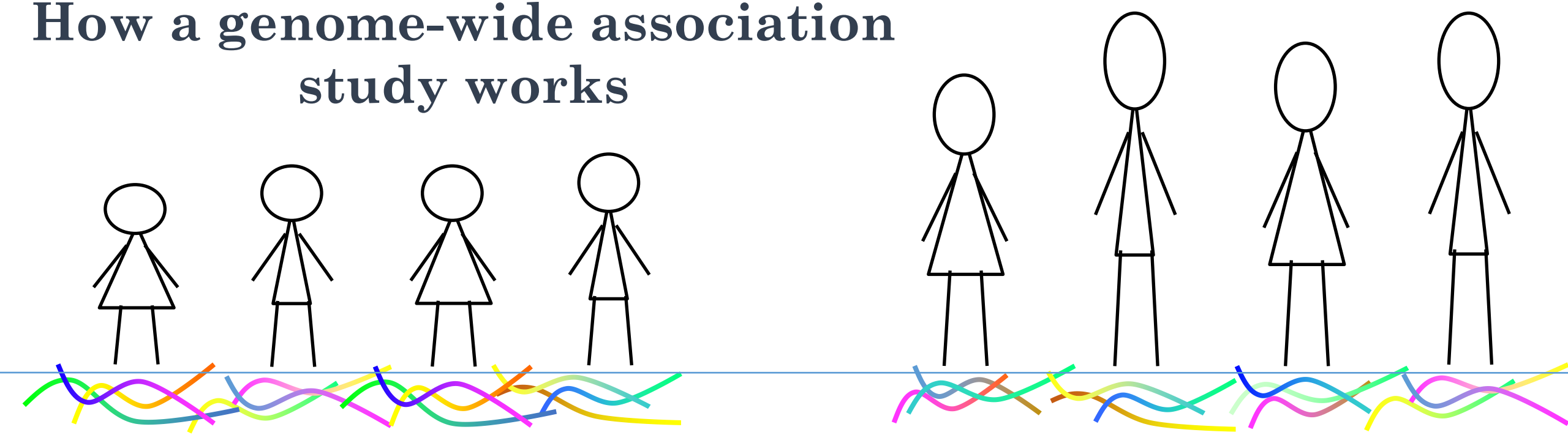
Genome-wide Association Studies (GWAS)

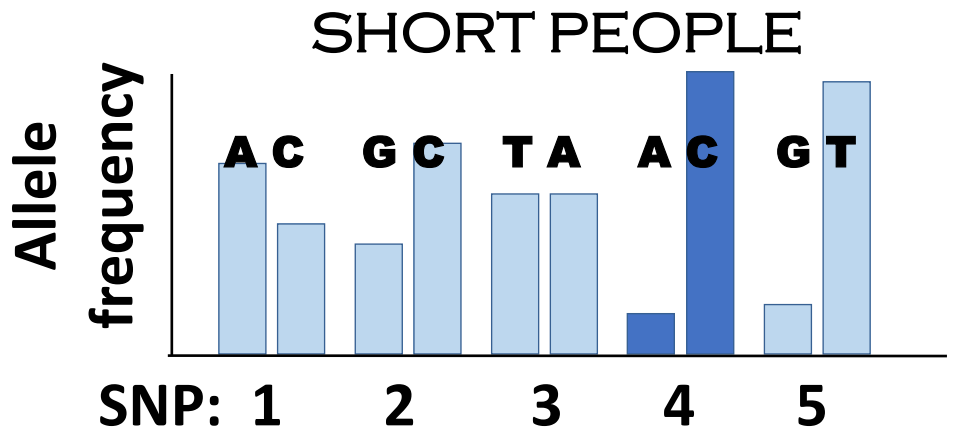
Use analysis of SNPs to find places in genome 'associated' with differences in the trait of interest

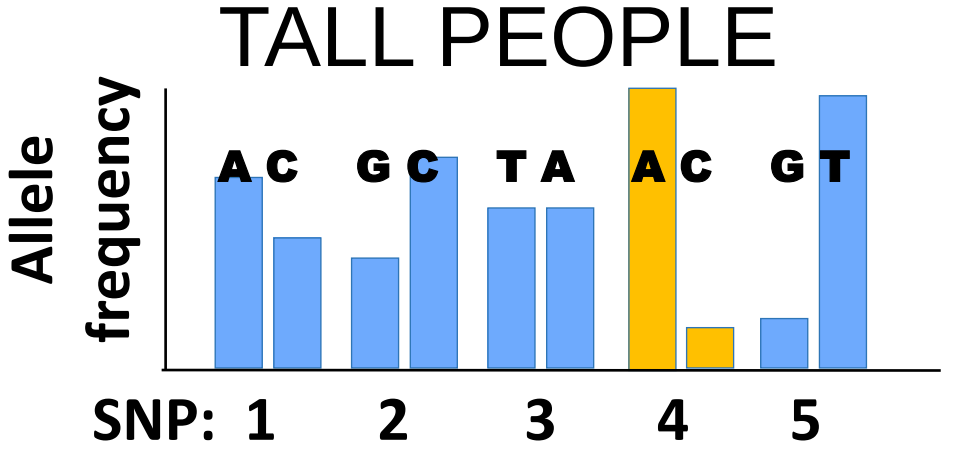
SNP = single-nucleotide polymorphism
= common simple DNA variant

1. Find many thousand people who differ for the trait of interest
2. Use a 'DNA chip' to identify (genotype) their alleles at each of ~10^6 SNP genome positions
3. Look for the SNPs where the two populations have different allele frequencies

How a genome-wide association study works







Genome-wide Association Studies (GWAS)

SNP#	Tall vs short?
1	same
2	same
3	same
4	same
5	same
6	same
7	same
8	same
9	same
10	same
.....	

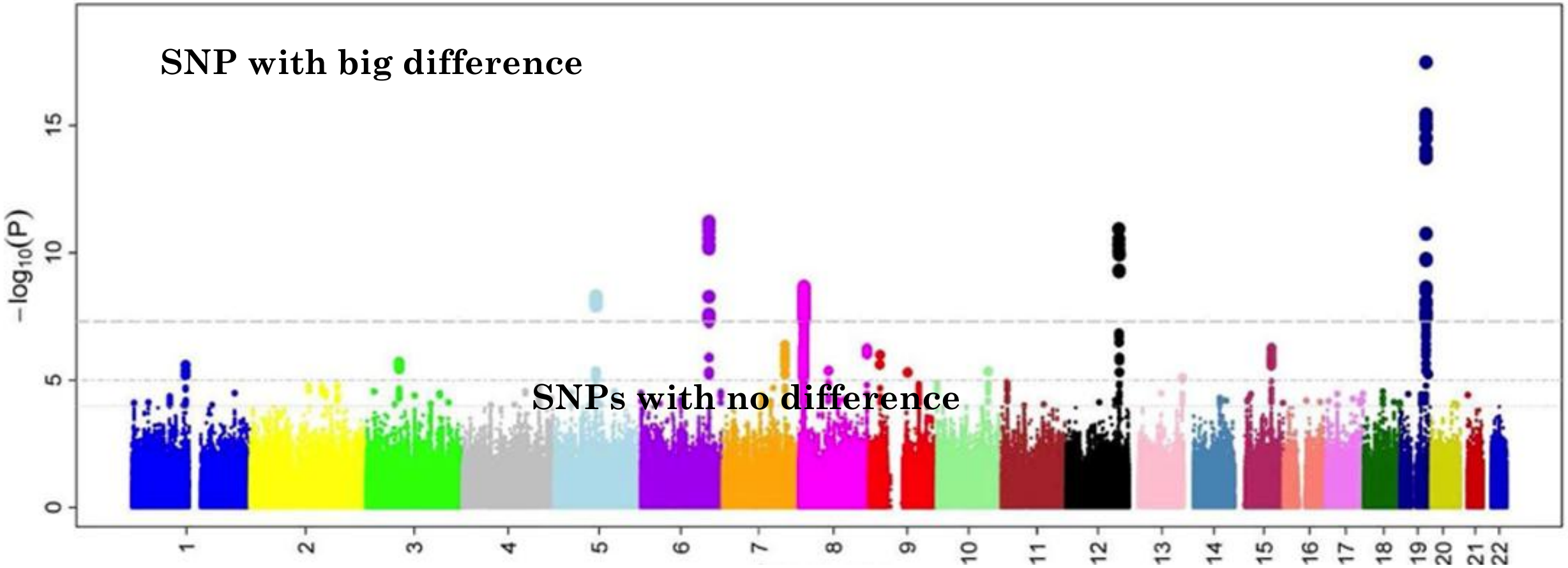
SNP#	Tall vs short?
101	same
102	same
103	**different**
104	same
105	same
106	same
107	same
108	same
109	same
1010	same
.....	

SNP#	Tall vs short?
10001	same
10002	same
10003	same
10004	same
10005	same
10006	**different**
10007	same
10008	same
10009	same
100010	same
.....	

SNP#	Tall vs short?
999991	same
999992	same
999993	same
999994	same
999995	same
999996	same
999997	same
999998	same
999999	same
1000000	same

≈1,000,000 SNPs

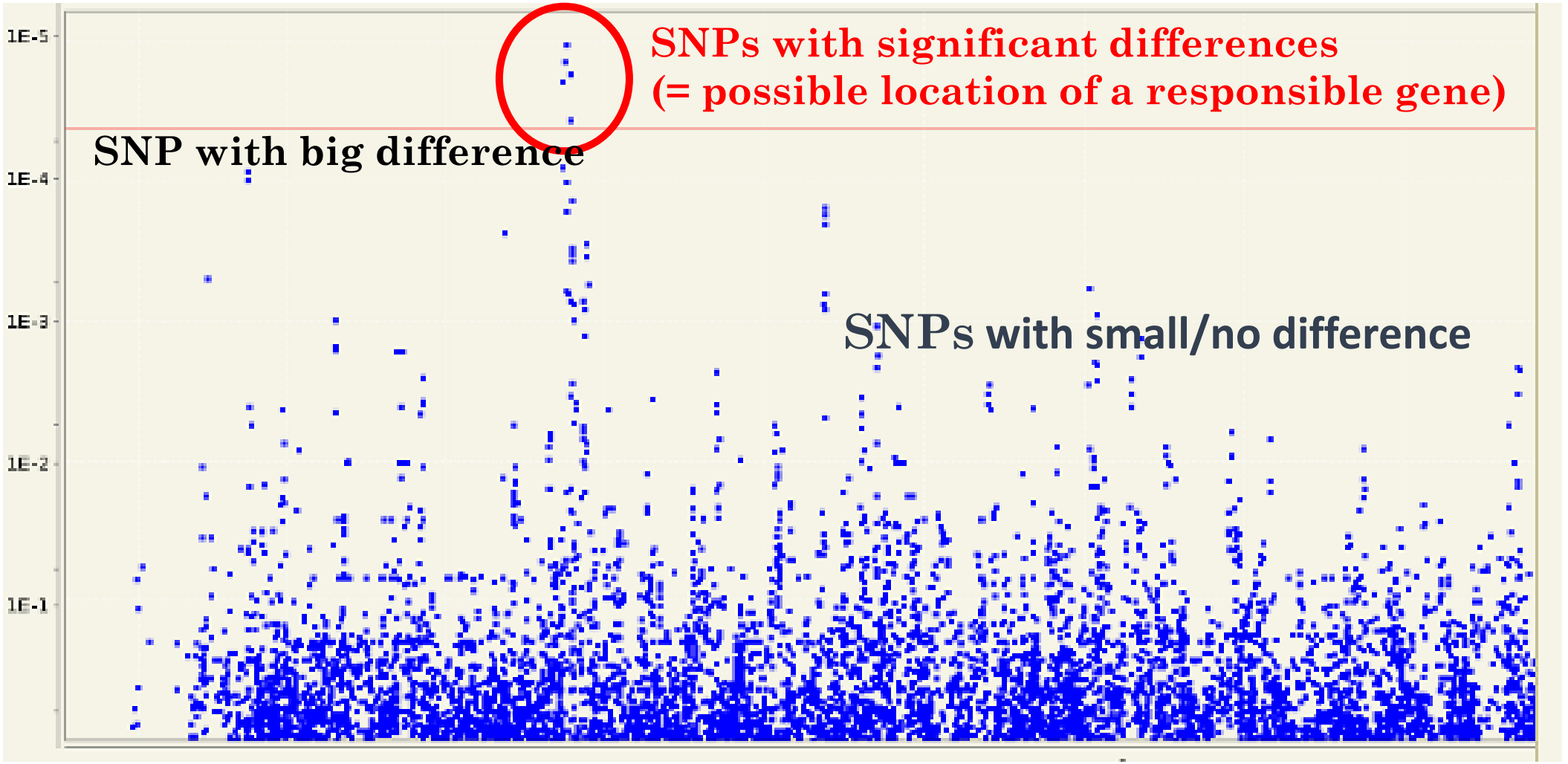
‘Manhattan plot’ of GWAS results



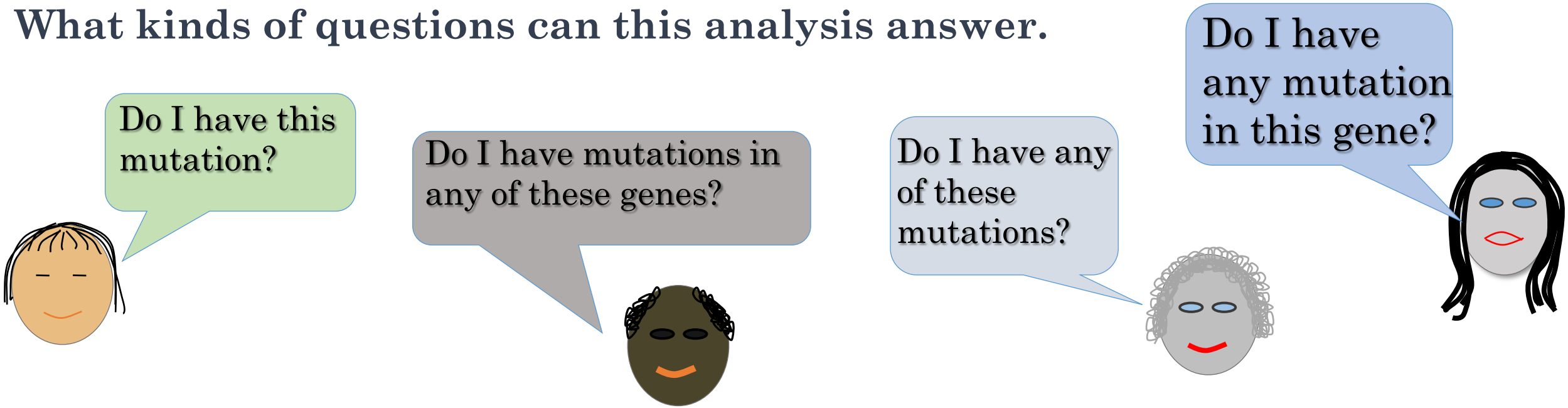
Position of SNPs along each of the chromosomes

Ikram MK et al (2010) Four novel loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation *in vivo*. PLoS Genet. 2010 Oct 28;6(10):e1001184. doi:10.1371/journal.pgen.1001184.g001

Zoom in on GWAS results



What kinds of questions can this analysis answer.



When/why would this information be needed?

When the person has a condition ('syndrome') that has a well characterized genetic cause.

When the person's relative is known to have a specific genetic condition

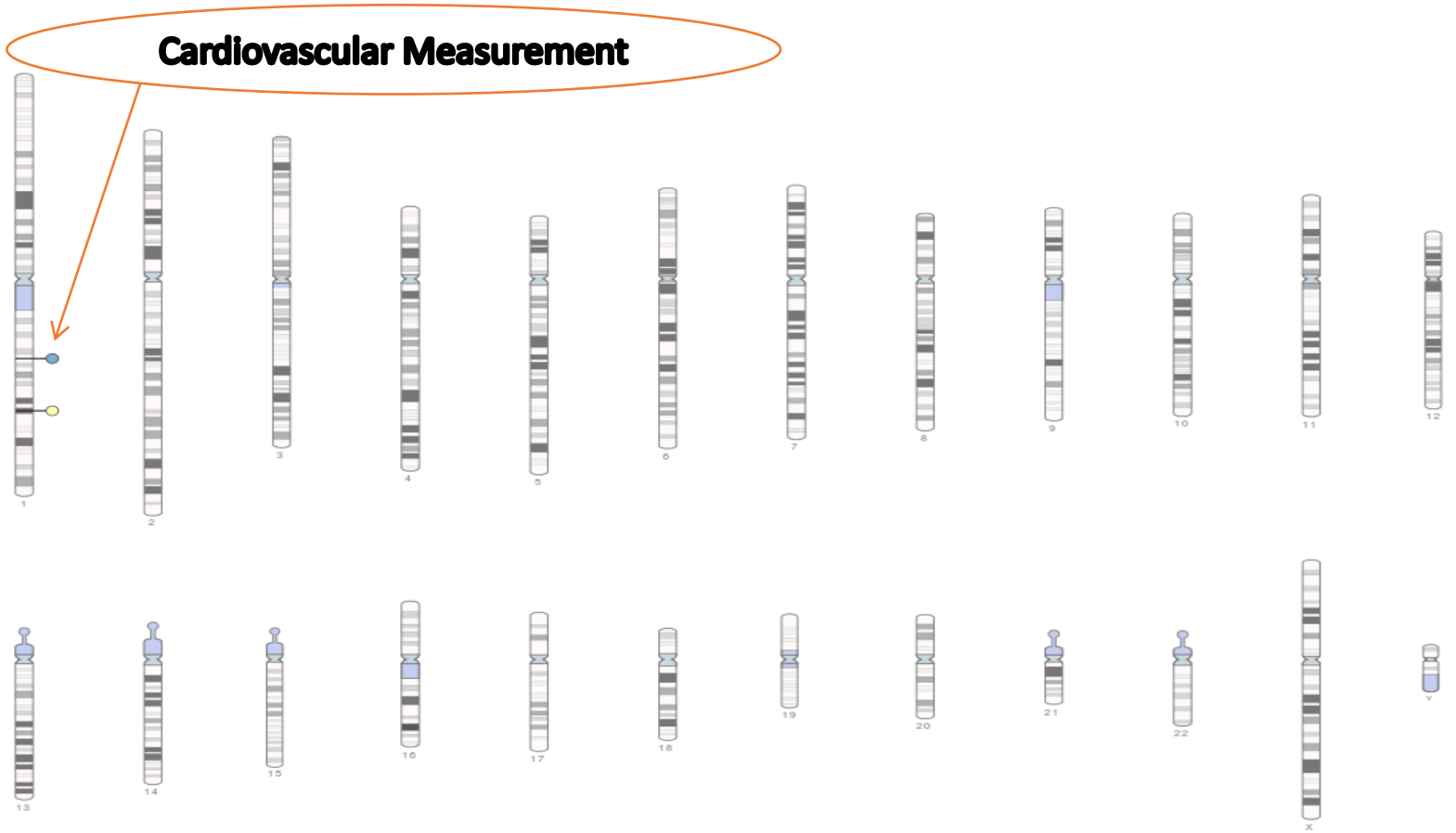
When choice of treatment depends on the genetic cause

Issues of GWAS

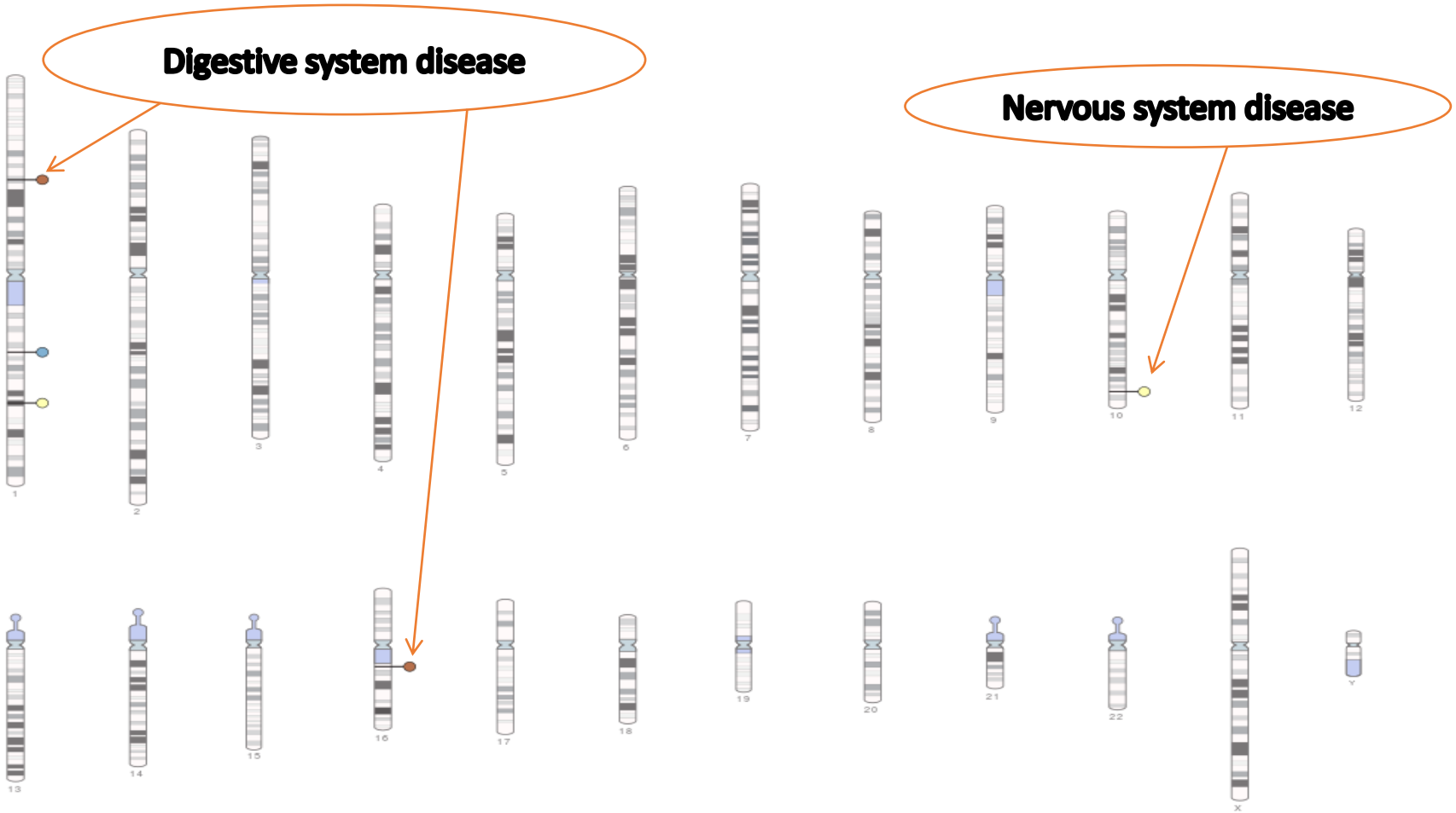
- Population stratification
- Multiple Testing: False Positives
- Gene-Environmental Interaction
- High Costs

مطالعات GWAS (2005)

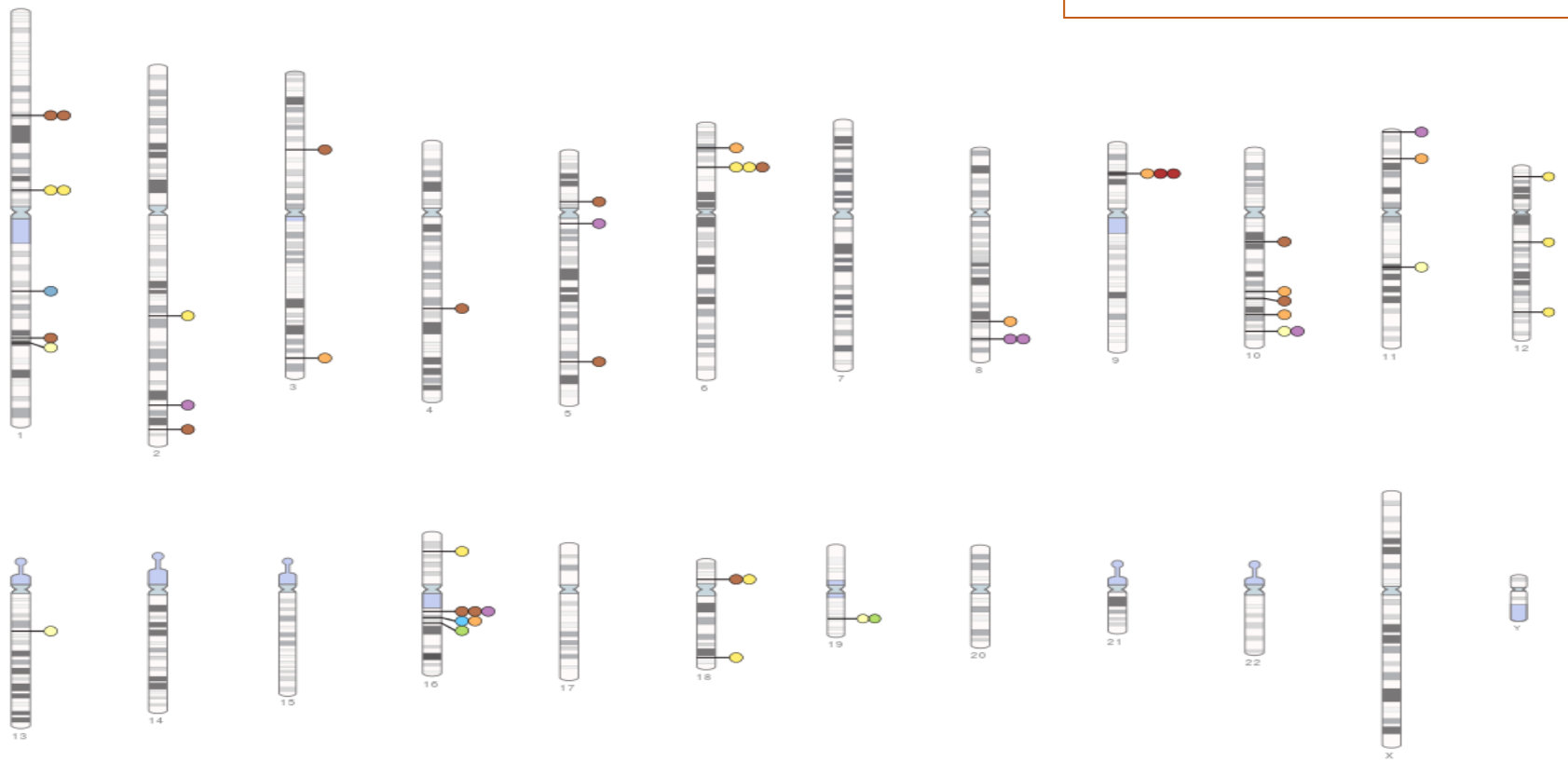
مطالعات GWAS (2006)



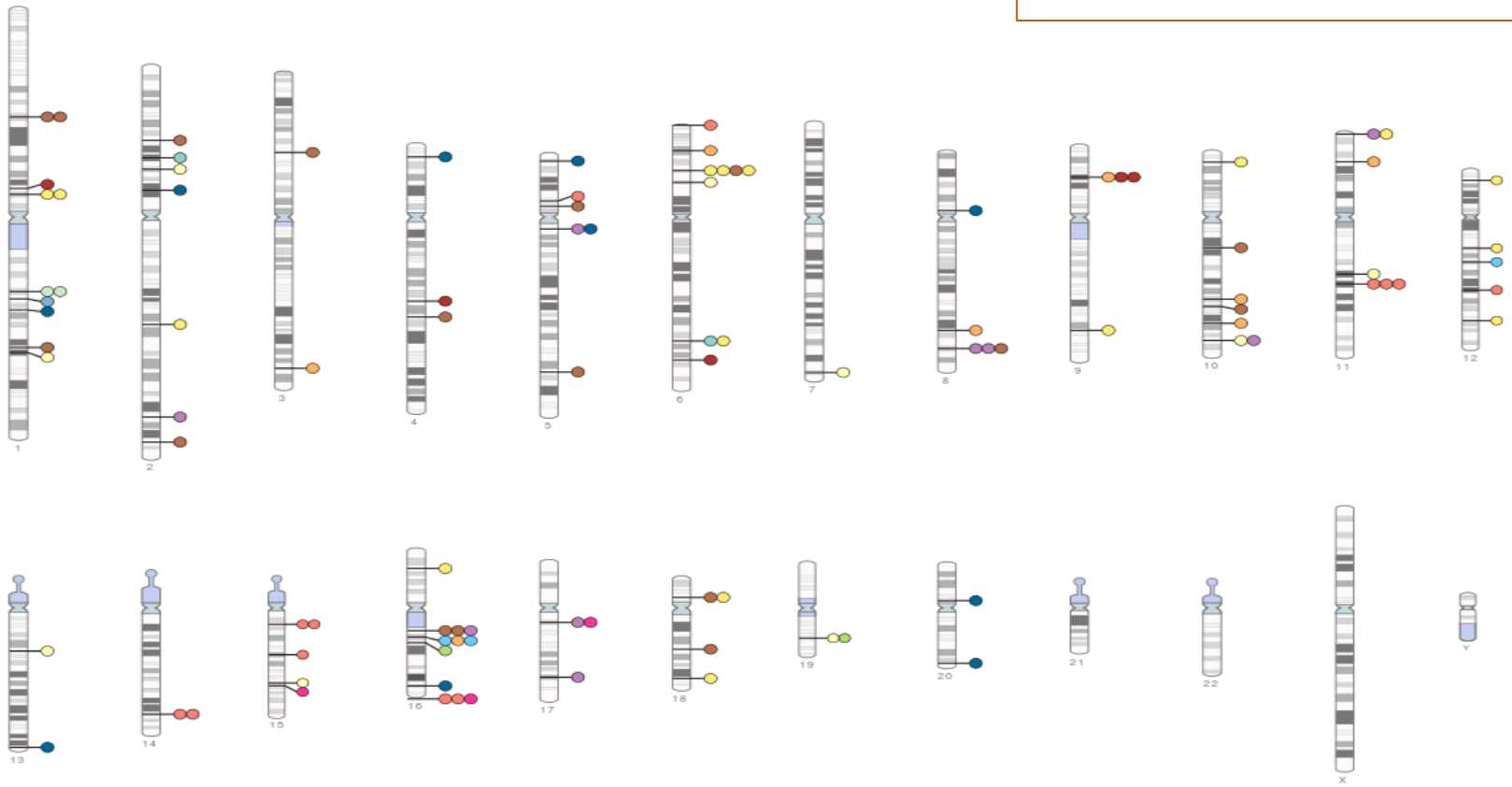
مطالعات GWAS (2006)



مطالعات GWAS (2007)



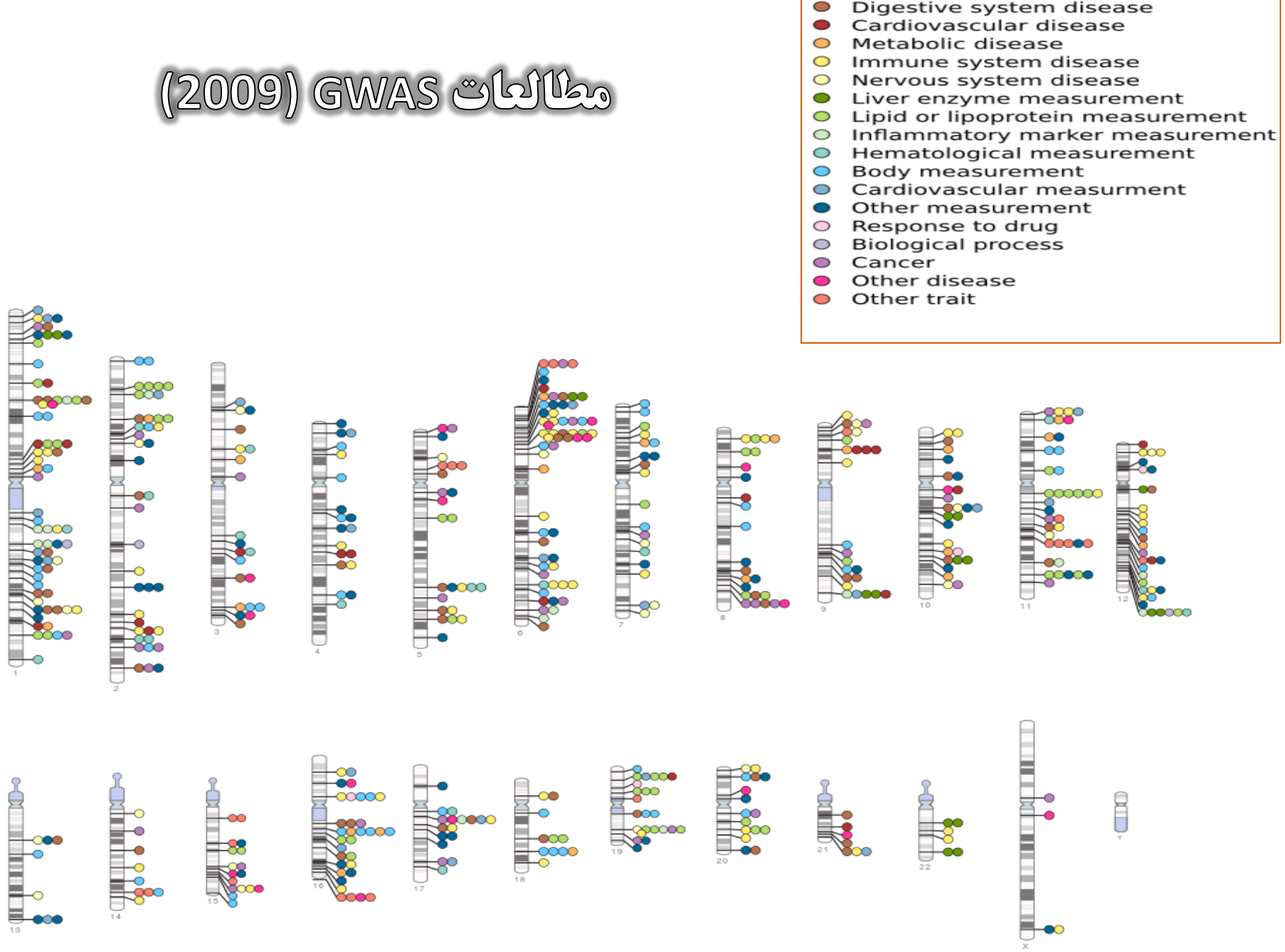
مطالعات GWAS (2007)



مطالعات GWAS (2008)

مطالعات GWAS (2008)

مطالعات GWAS (2009)

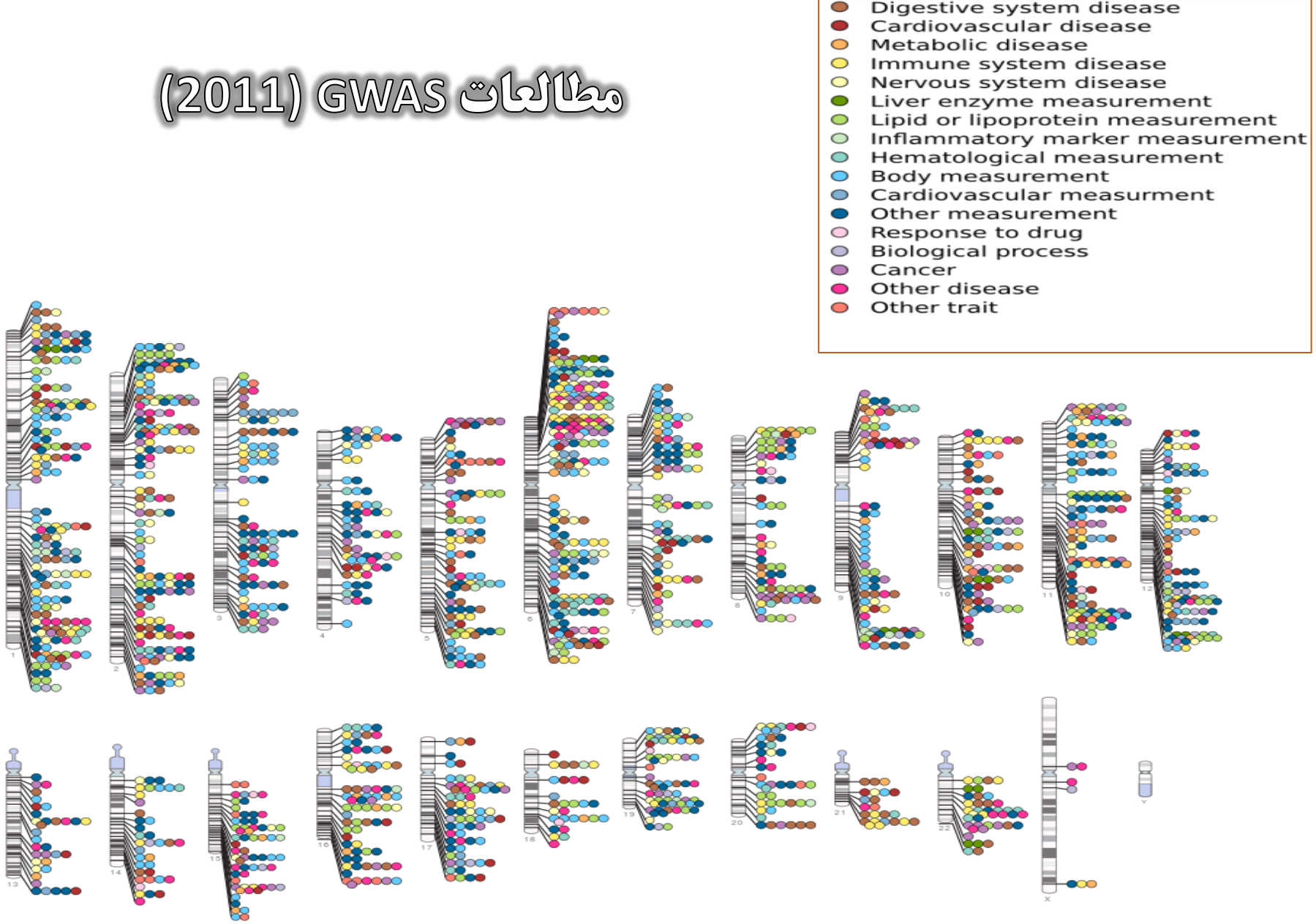


مطالعات GWAS (2009)

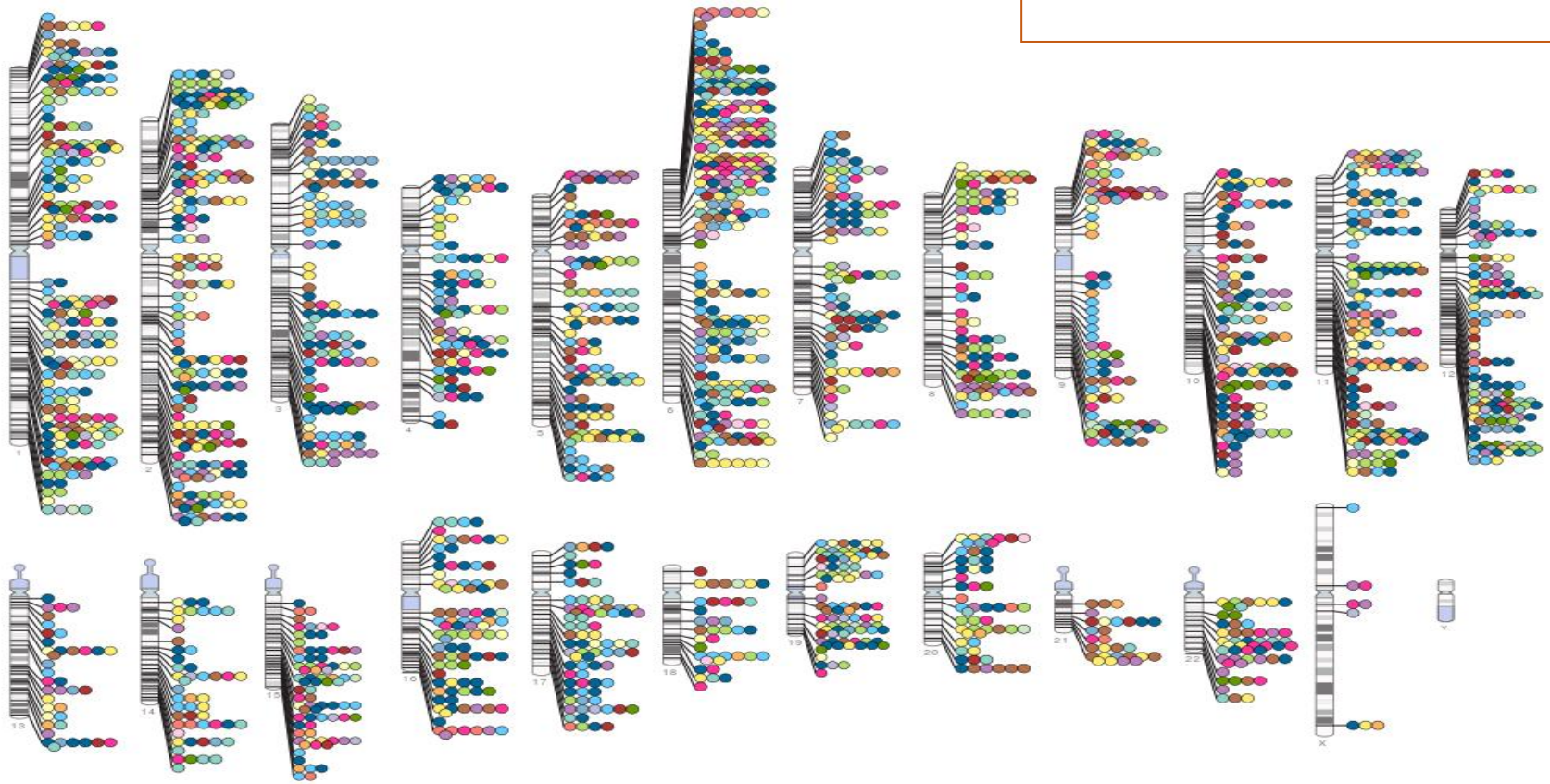
مطالعات GWAS (2010)

مطالعات GWAS (2010)

مطالعات GWAS (2011)

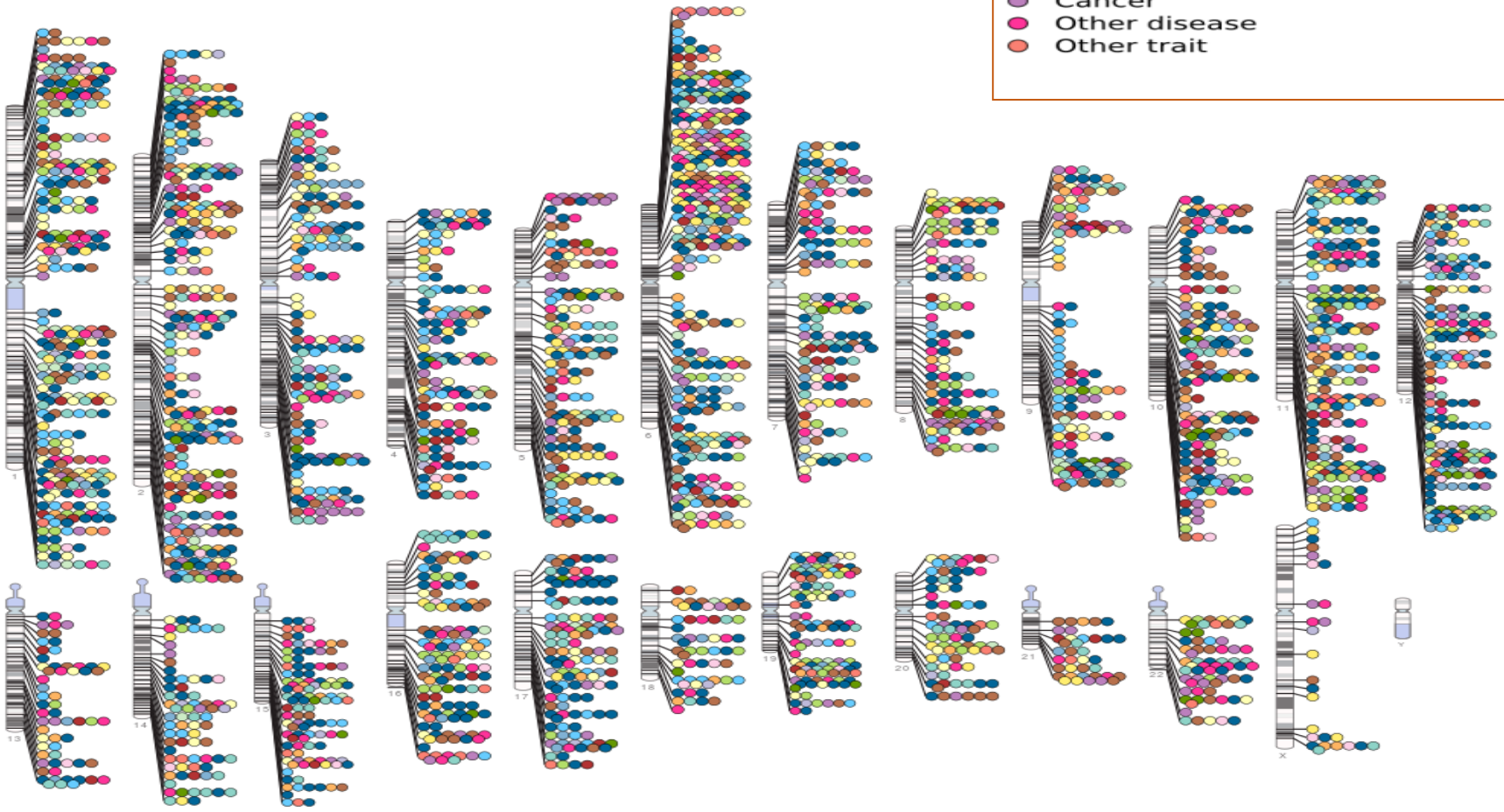


مطالعات GWAS (2011)



مطالعات GWAS (2012)

مطالعات GWAS (2012)



مطالعات GWAS (2013)

مطالعات GWAS (2013)

مطالعات GWAS (2014)

Ten Basic Questions to Ask About a Genome-wide Association Study Report

1. Are the cases defined clearly and reliably so that they can be compared with patients typically seen in clinical practice?

2. Are case and control participants demonstrated to be comparable to each other on important characteristics that might also be related to genetic variation and to the disease?

3. Was the study of sufficient size to detect modest odds ratios or relative risks (1.3-1.5)?

4. Was the genotyping platform of sufficient density to capture a large proportion of the variation in the population studied?

5. Were appropriate quality control measures applied to genotyping assays, including visual inspection of cluster plots and replication on an independent genotyping platform?

Pearson, T. A. et al. JAMA 2008;299:1335-1344

Ten Basic Questions to Ask About a Genome-wide Association Study Report

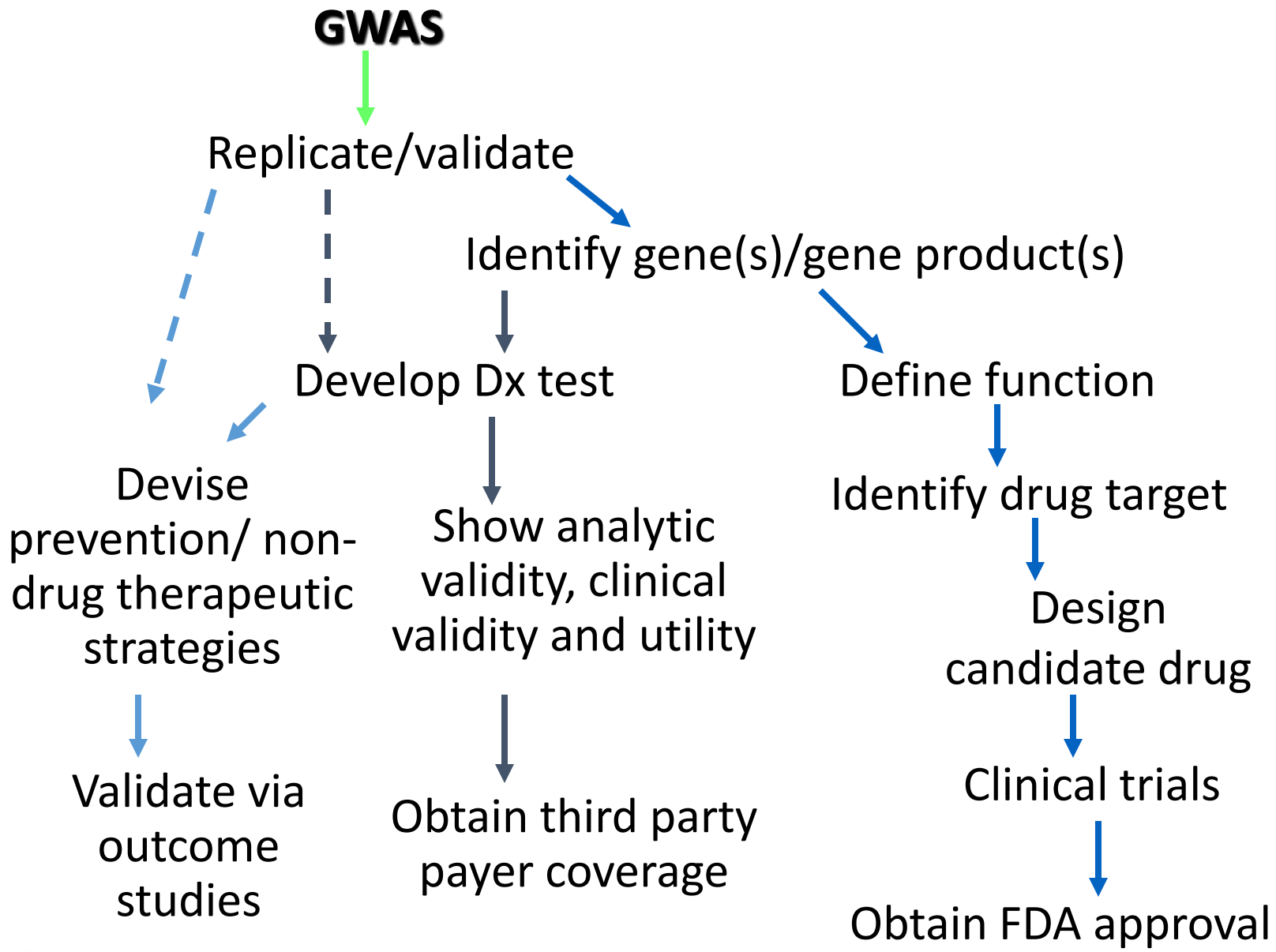
6. Did the study reliably detect associations with previously reported and replicated variants (known positives)?

7. Were stringent corrections applied for the many thousands of statistical tests performed in defining the P value for significant associations?

8. Were the results replicated in independent population samples?

9. Were the replication samples comparable in geographic origin and phenotype definition, and if not, did the differences extend the applicability of the findings?

10. Was evidence provided for a functional role for the gene polymorphism identified?

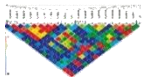

GWAS
Replicate/validate
Identify gene(s)/gene product(s)
Develop Dx test
Define function
Devise prevention/ non-drug therapeutic strategies
Show analytic validity, clinical validity and utility
Identify drug target
Design candidate drug
Validate via outcome studies
Obtain third party payer coverage
Clinical trials
Obtain FDA approval

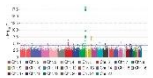
https://www.ebi.ac.uk/**gwas**

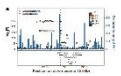
© Francis Collins, 2008

روند انجام مطالعه گسترده ژنومی

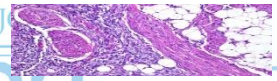
Genome Wide Association Study







Genotypes	CC	CA	AA	Total
Case Observed	59	27	98	184
Control Observed	60	89	36	185
Total	119	116	134	369



مطالعه قند و لیپید تهران
Tehran lipid and Glucose Study

فاز اول
- ۱۳۷۸-۱۳۸۰
- پانزده هزار نفر شرکت کننده

فاز دوم
- ۱۳۸۱-۱۳۸۳
- تشکیل بانک ژنومی

فاز سوم
- ۱۳۸۴-۱۳۸۷
- رسم ۵۴۰۰ شجره خانوادگی

فاز چهارم
- ۱۳۸۸-۱۳۹۰
- عقد تفاهم نامه جهت انجام مطالعه گسترده ژنومی

فاز پنجم
- ۱۳۹۱-۱۳۹۳
- ارسال نمونه و انجام ژنوتایپینگ

فاز ششم
- ۱۳۹۴-....
- کنترل کیفی نتایج و آزمون های آماری

مطالعه گسترده ژنومی در مطالعه قند و لیپید تهران

Tehran Cardio-metabolic Genetic Study