

equipment, and to Dr. G. E. R. Denton and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

¹ Young, F. B., Gerold, H., and Jeros, W., *Phil. Mag.*, **40**, 149 (1935).

² Longuet-Higgins, M. S., *Nat. Rep. Roy. Astro. Soc., Geophys. Suppl.*, **3**, 288 (1949).

³ Van Arx, W. S., Woods Hole Papers in Phys. Geodes., Notes, **11**, C3 (1950).

⁴ Ekstrand, V. W., *Arkiv. Mat. Astrofysik. (Stockholm)*, **3**(11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey¹. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate ester groups joining 2'-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's² model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There

This figure is purely diagrammatic. The two ribbons symbolize the two polynucleotide chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

We are much indebted to Dr. Jerry Donnison for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

is a residue on each chain every 3.4 Å. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs: a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-coordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally^{3,4} that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data^{5,6} on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donnison for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.

April 2.

¹ Pauling, L., and Corey, E. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **40**, 51 (1953).

² Furberg, B., *Acta Chem. Scand.*, **6**, 634 (1952).

³ Chargaff, E., for references see Kamenof, S., Braserman, G., and Chargaff, E., *Nucleic Acids et Biologie*, **10**, 603 (1948).

⁴ Wyatt, G. E., *J. Gen. Physiol.*, **26**, 293 (1952).

⁵ Astbury, J. T., *Proc. Roy. Soc. Exp. Biol. I. Nucleic Acid*, **66** (Camb. Univ. Press), 1947.

⁶ Wilkins, M. H. F., and Randall, J. T., *Biokim. et Biolog. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

WHILE the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury¹) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the poly-nucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline^{2,3}, semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure B' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3.4-Å. reflexion corresponded to the inter-nucleotide repeat along the fibre axis. The ~34 Å. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown⁵ (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of $J_{n\theta}$, the nth order Bessel function. A straight line may be drawn approximately through

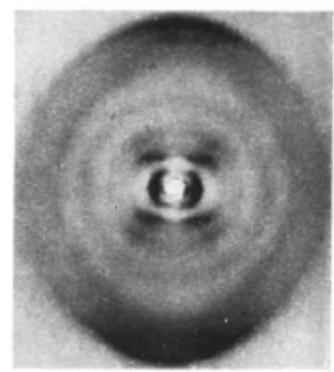


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *E. coli*. Fibre axis vertical.

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion ($J_{n\theta}^2$) on the nth layer line. The helical configuration produces side-bands on this fundamental reflection, the effect⁶ being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to C in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleoside. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

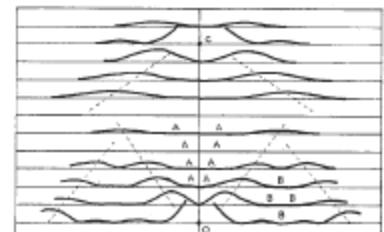
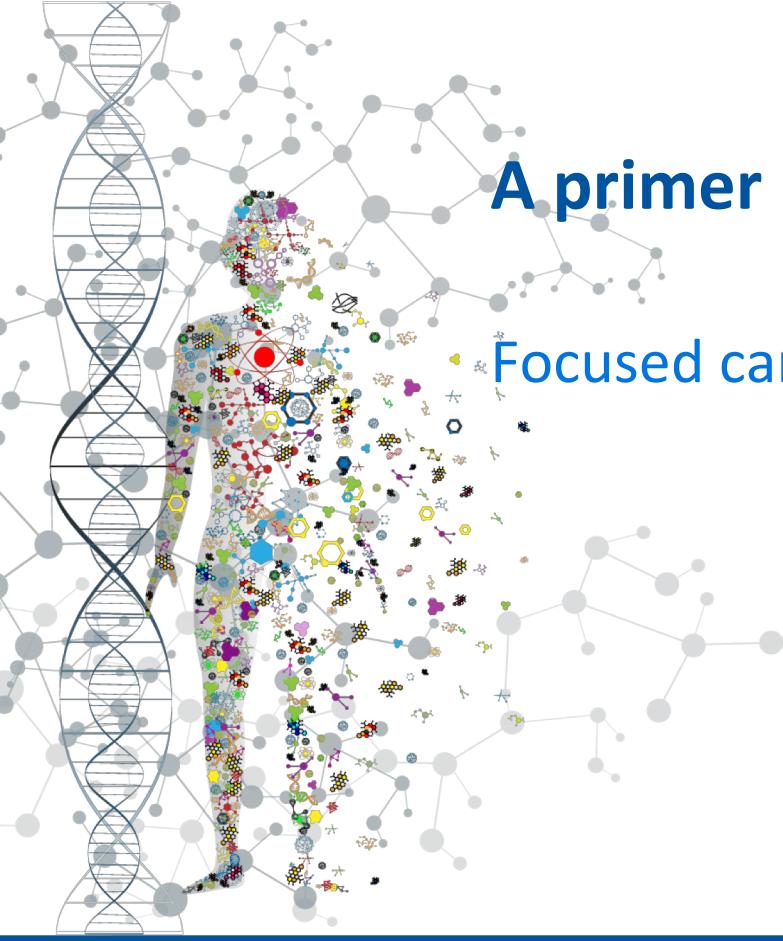


Fig. 2. Diffraction pattern of systems of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third, etc., layer lines. The radius of the helix is 20 Å. diameter and remains distributed along a radius, the mass at 20 Å. radius being proportional to the radius. About C on the tenth layer line similar functions are plotted for an outer diameter of 12 Å.



A primer in complex human genetics

Focused cardiovascular disease

Sander W. van der Laan, PhD

s.w.vanderlaan-2@umcutrecht.nl • [@swvanderlaan](https://twitter.com/swvanderlaan) • [swvanderlaan.github.io](https://github.com/swvanderlaan)



UMC Utrecht
Center for Circulatory Health

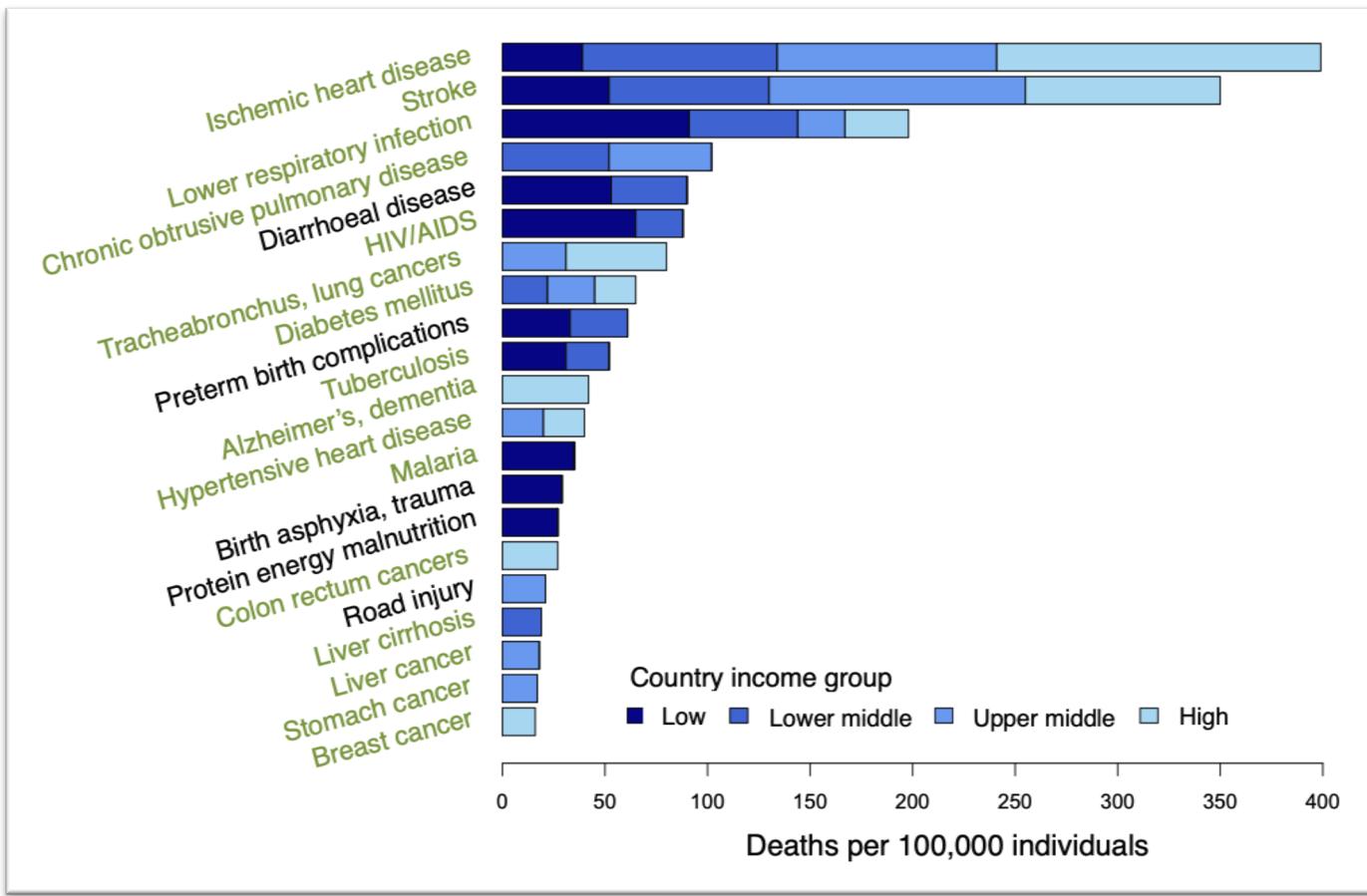
UMC Utrecht
Athero-Express Biobank Studies



Take home

- Basics of the human genome
- Analysis approaches
- Basic principles of modern-day genetic analyses

Human disease around the globe



The spectrum(s) of disease

Age of onset

Early onset

Late onset

Genetic architecture

One gene

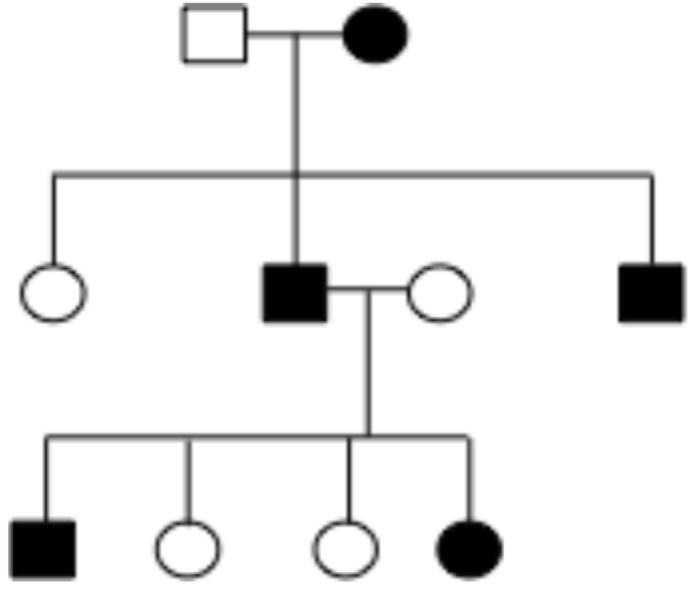
Many genes

Environmental architecture

No role

Major role

Rare (Mendelian) diseases



family-based studies

Genotype



Disease



Environment

Rare (Mendelian) diseases

1983

A polymorphic DNA marker genetically linked to Huntington's disease

James F. Gusella*, Nancy S. Wexler[†], P. Michael Conneally[†], Susan L. Naylor[§]
Mary Anne Anderson^{*}, Rudolph E. Tanzi^{*}, Paul C. Watkins^{*}, Kathleen
Margaret R. Wallace[‡], Alan Y. Sakaguchi[§], Anne B. Young^{||}, Ira S.
Ernesto Bonilla^{||} & Joseph B. Martin*

* Neurology Department and Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.

† Hereditary Disease Foundation, 9701 Wilshire Blvd, Beverly Hills, California 90212, USA.

‡ Department of Medical Genetics, Indiana University Medical Center, Indianapolis, Indiana 46225, USA.

§ Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263, USA.

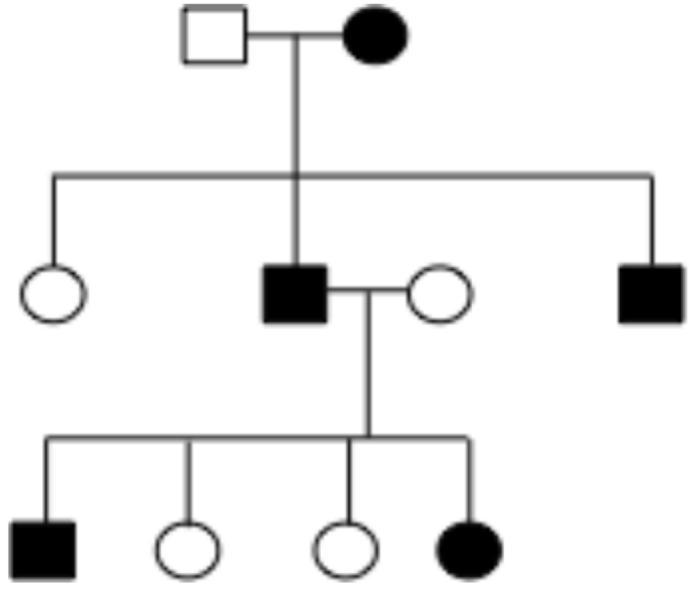
|| Venezuela Collaborative Huntington's Disease Project*

Family studies show that the Huntington's disease gene is linked to a polymorphic DNA marker on chromosome 4. The chromosomal localization of the Huntington's disease gene is the first step in using DNA technology to identify the primary genetic defect in this disorder.



1989

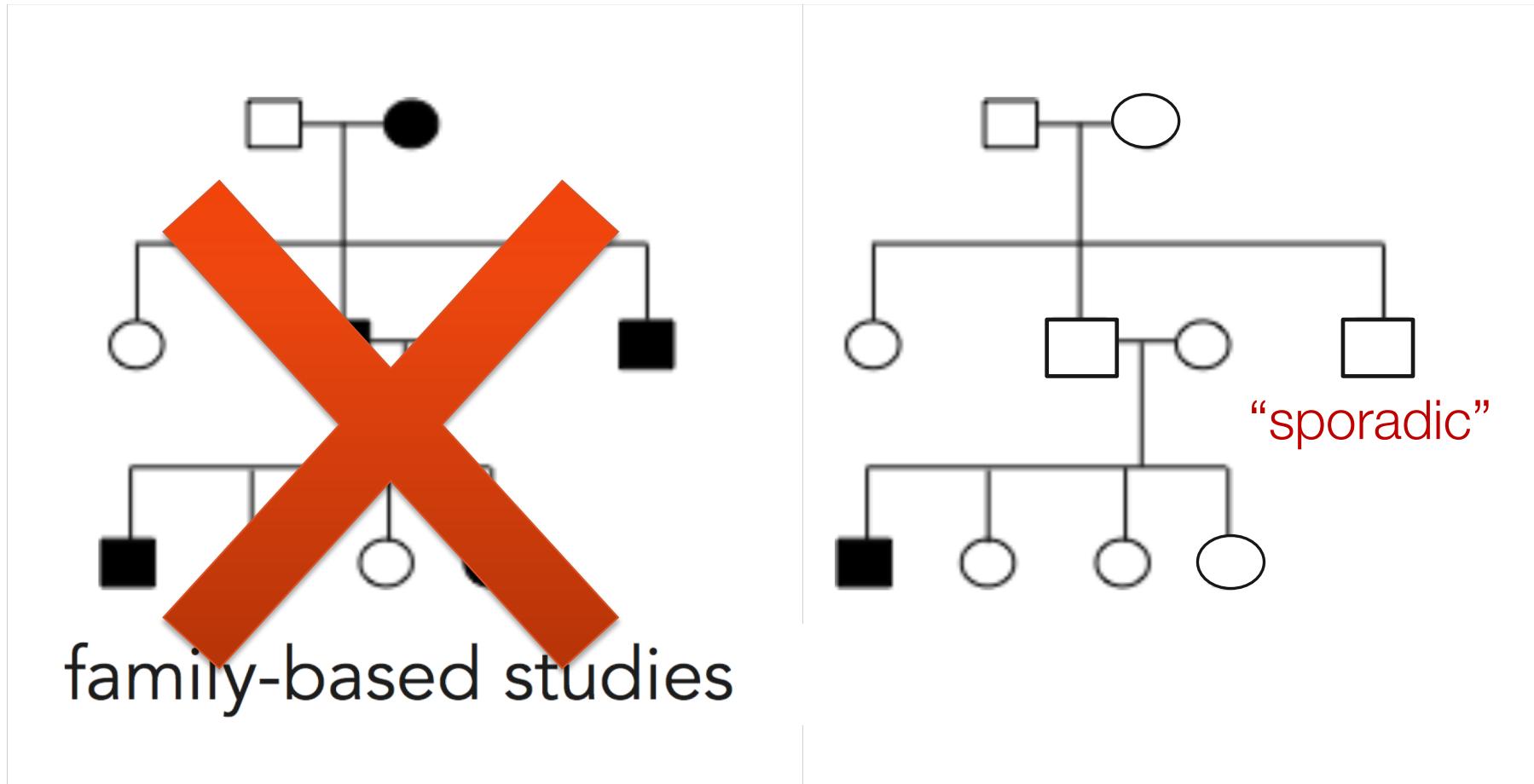
Common (complex) diseases



family-based studies



Common (complex) diseases



family-based studies

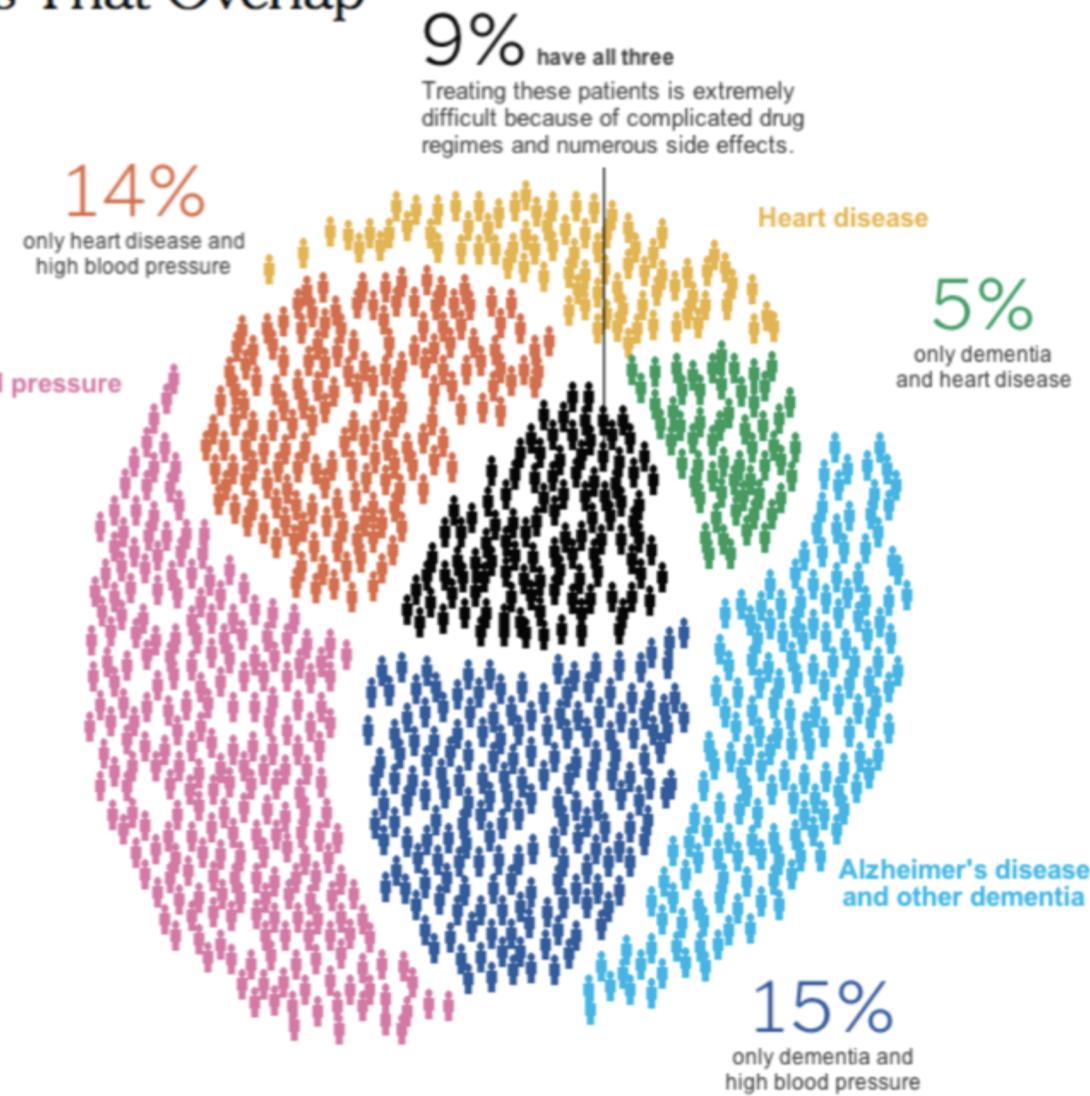
The challenges of common disease

- Heterogeneity
- Late (or broad age range for) onset
- Interaction of genes and environment (multifactorial)
- Overlap with other diseases

For the Elderly, Diseases That Overlap

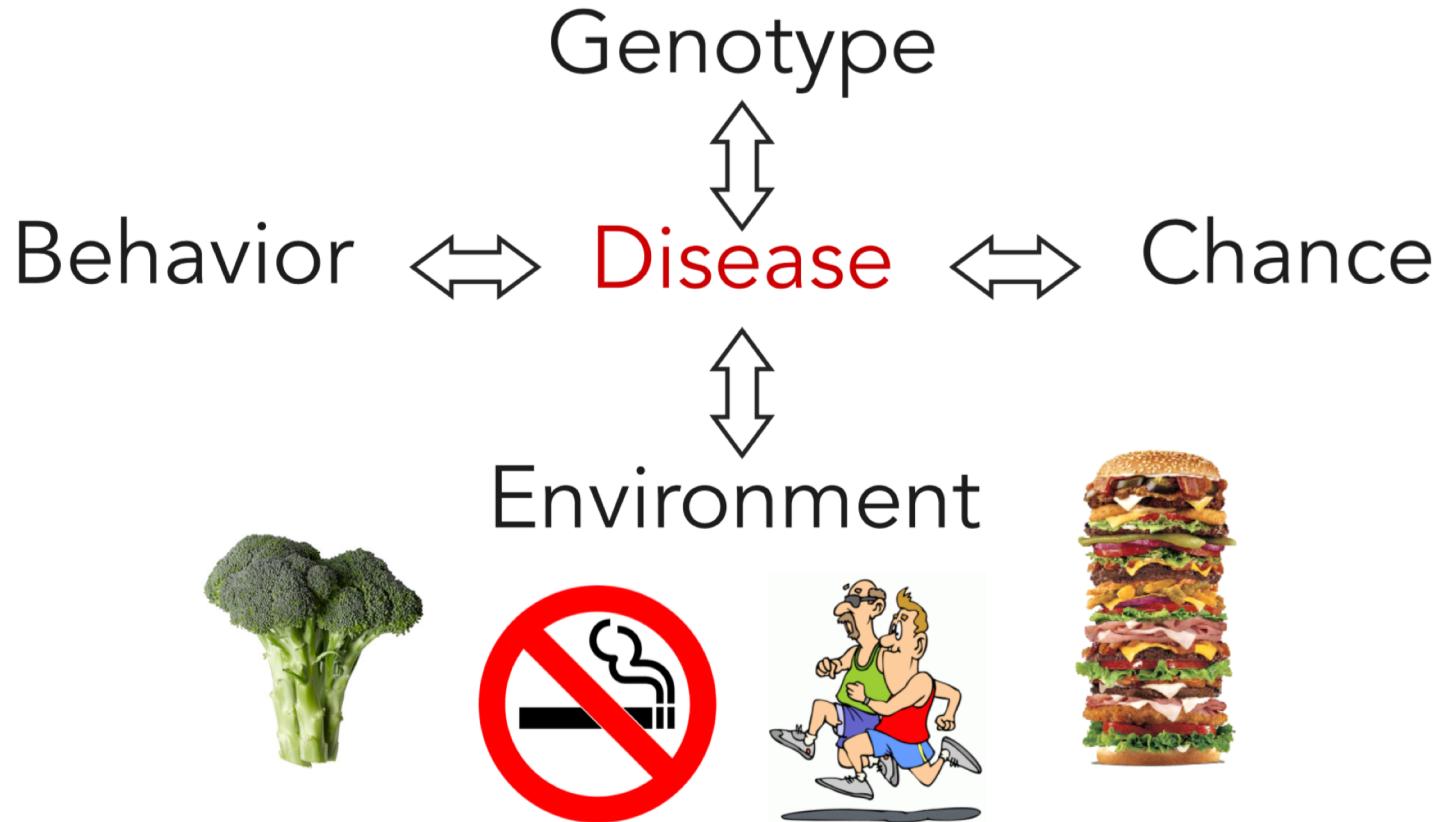
1 2 3 4 5 6 7 NEXT »

Researchers are beginning to focus more intently on the overlaps and possible interconnections, and some scientists argue that it may not be possible to treat dementia without treating vascular problems.

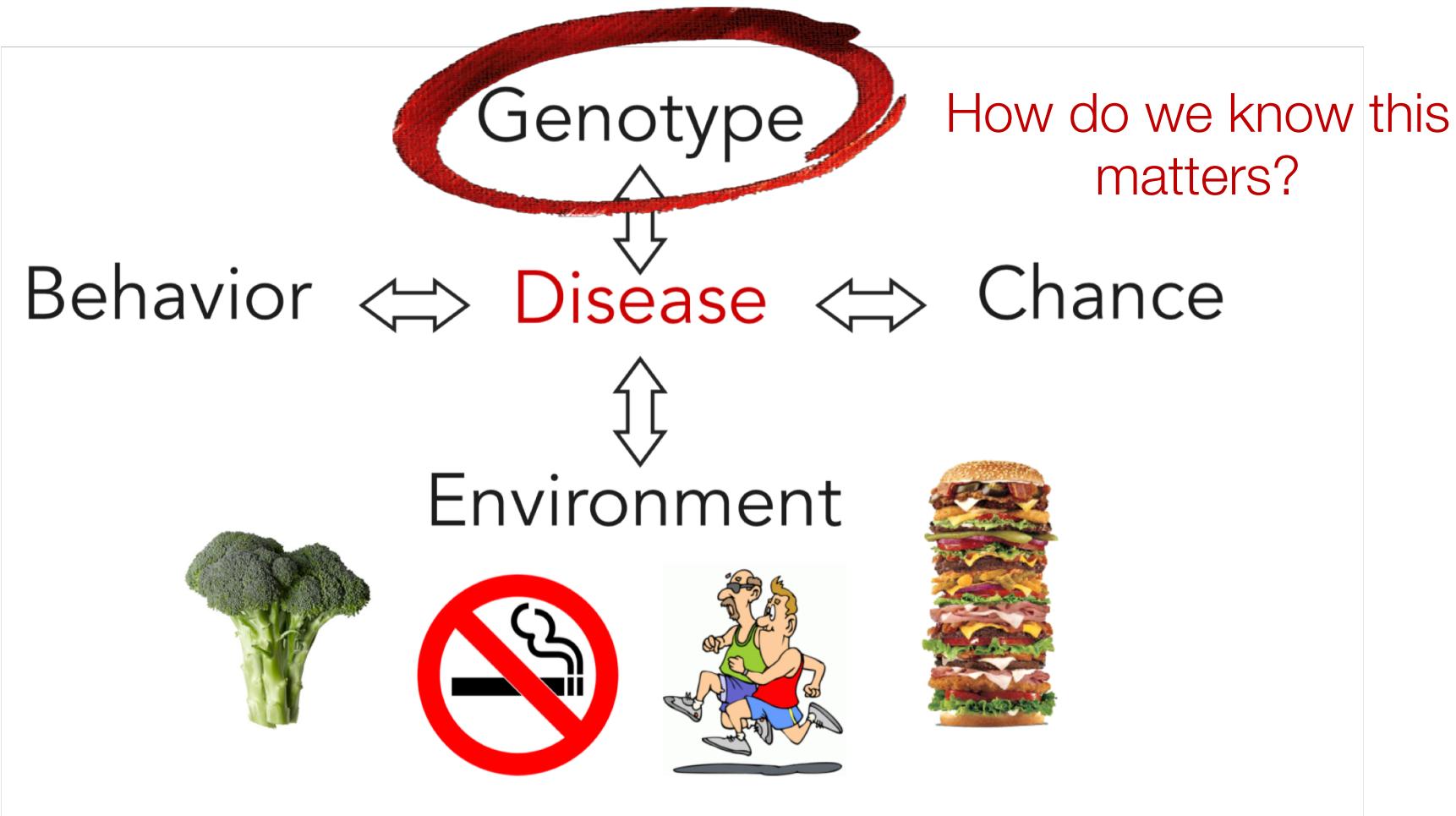


The New York Times, 15 April 2013

Multifactorial disease



Multifactorial disease



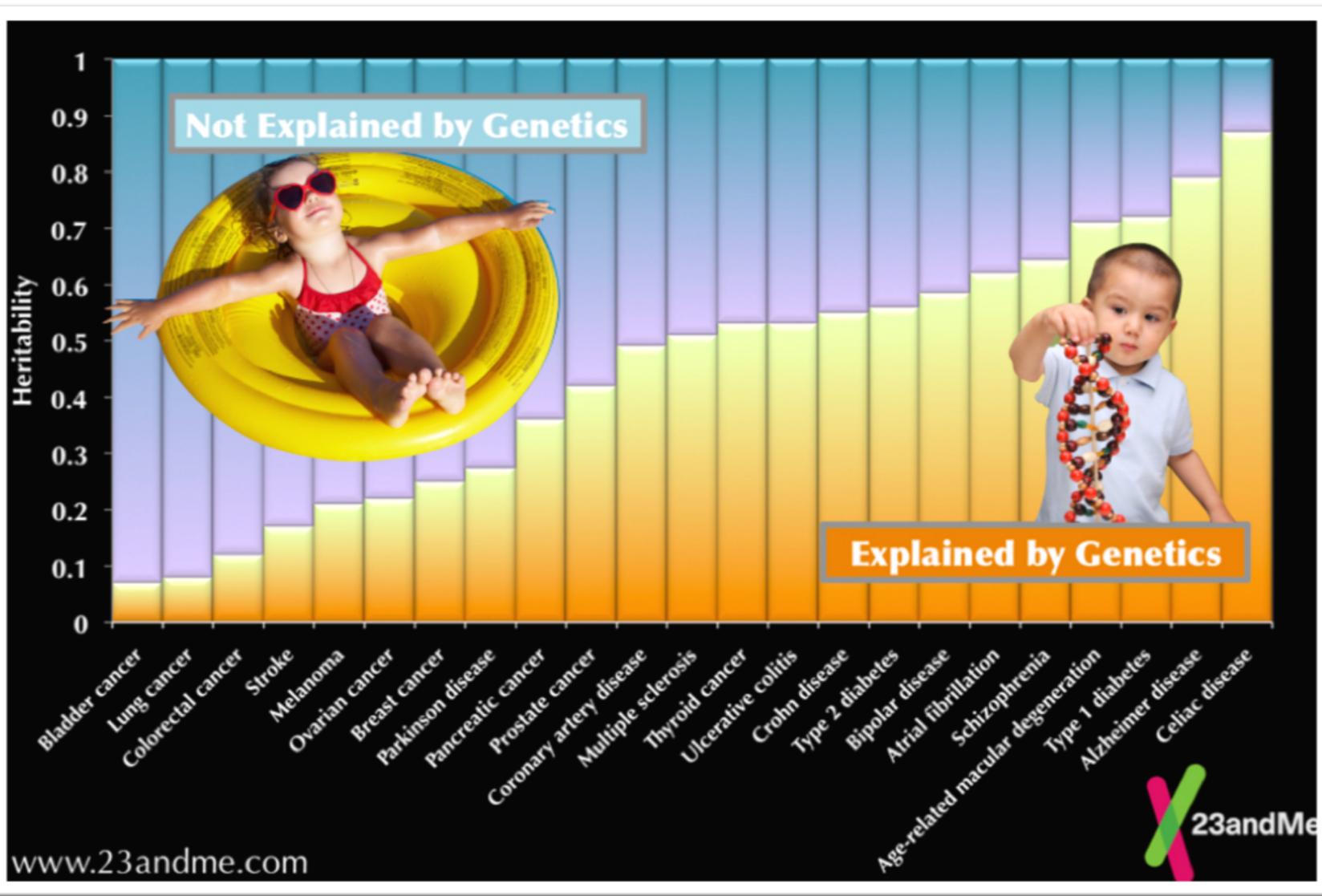


Heritability

Given I am a patient, what is risk of disease for...

	Type 1	Type 2
Your neighbor (unrelated)?	0.4%	5-10%
Your sibling?	6%	30%
Your identical twin?	30-50%	>80%

The range of heritability estimates



Family history

- Framingham Heart Study | www.framinghamheartstudy.org
 - A positive history of cardiovascular disease and associated risk factors tend to aggregate in families
 - Familial aggregation heritability of CVD estimated $\geq 90\%$ (before 46 years)
 - Family history is an independent risk factor (FHS)
 - Positive family history associated with pre-clinical atherosclerosis as measured by carotid IMT, $h^2 \approx 0.35$
- High concordance rate among monozygotic twins, compared to dizygotic twins
- Heritability of atherosclerosis (carotid IMT) $h^2 \approx 0.21-0.64$ and is increased by age and cardiovascular risk factors

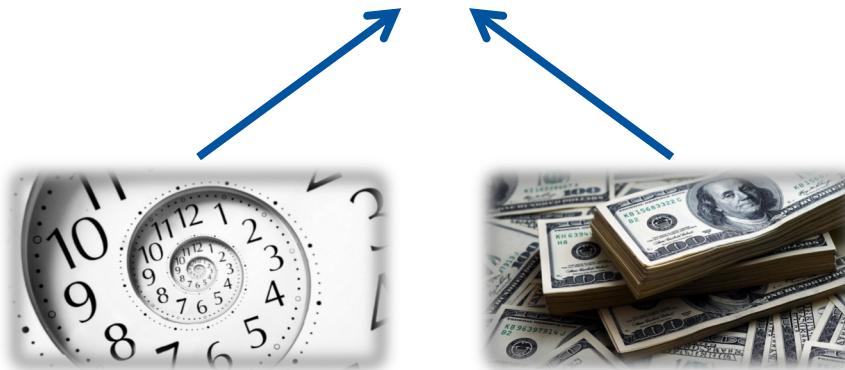
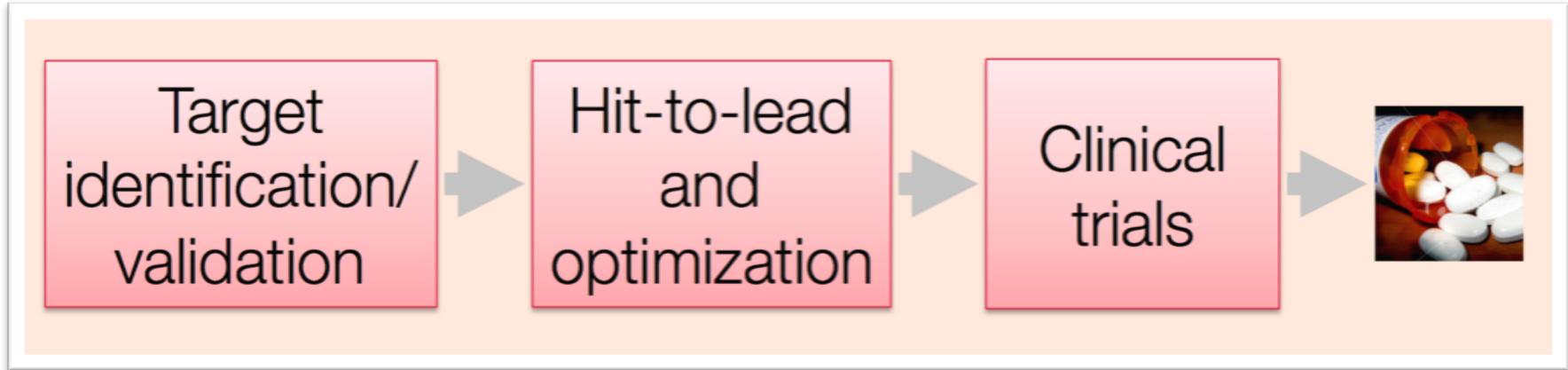
There is clearly a heritability factor for atherosclerotic and consequent cardiovascular disease



Why do some individuals have a higher risk for a disease than others?

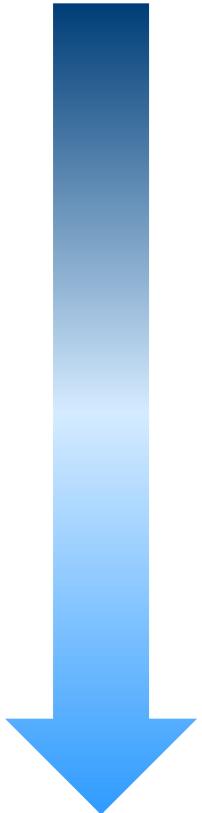
How can we alleviate disease burden in the human population?

Drug development



What's the goal of genetics?

- Understanding true causal disease pathways
 - Identify risk factors
 - Inform novel research directions
 - Enable rational and efficient drug development
- Precision medicine
 - Evaluate individual disease risk
 - Early disease identification or prevention
 - Understand patient's therapeutic response



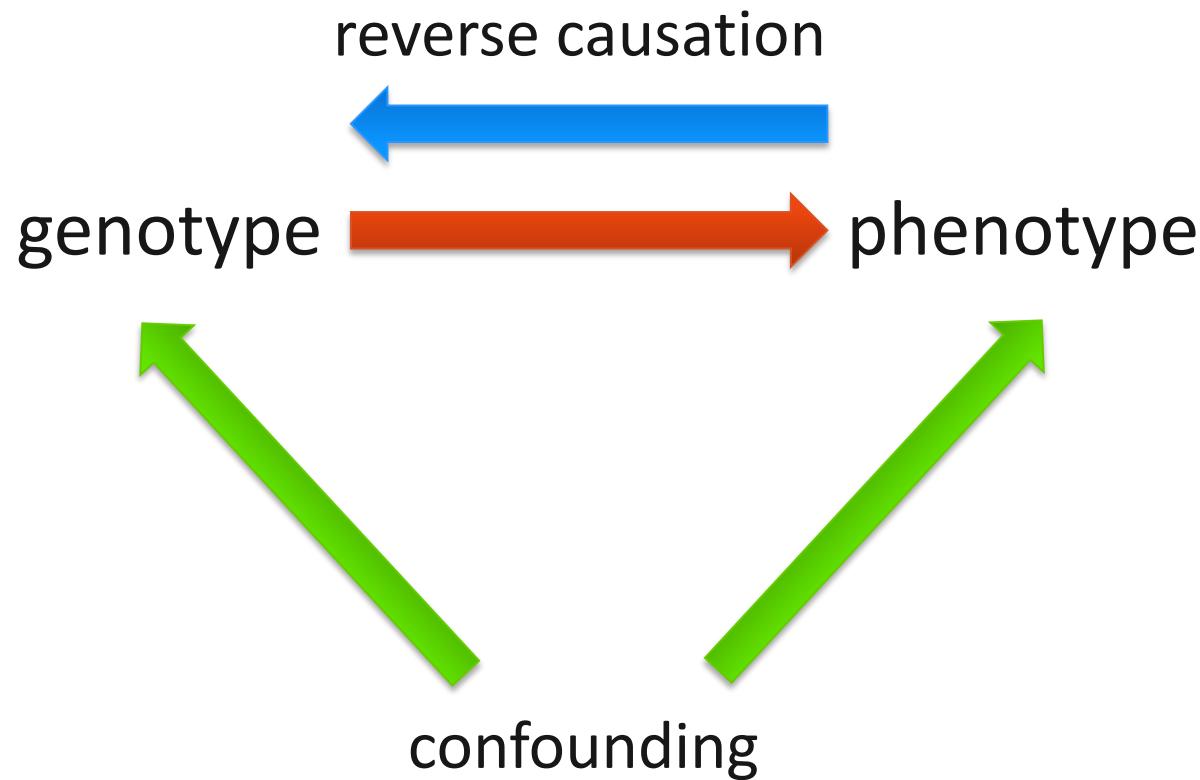
Why genetics at all?

- Genotypes are randomly assigned at meiosis
 - Nature's randomized clinical trial
- Genotypes are fixed and unaltered by the disease
 - Exception: somatic mutations in cancer
- We have become increasingly good at measuring genotypes
 - Lots and lots of data

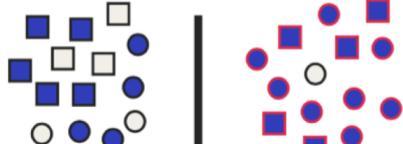
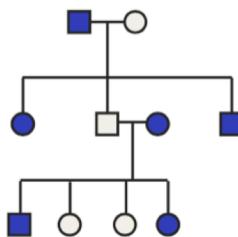
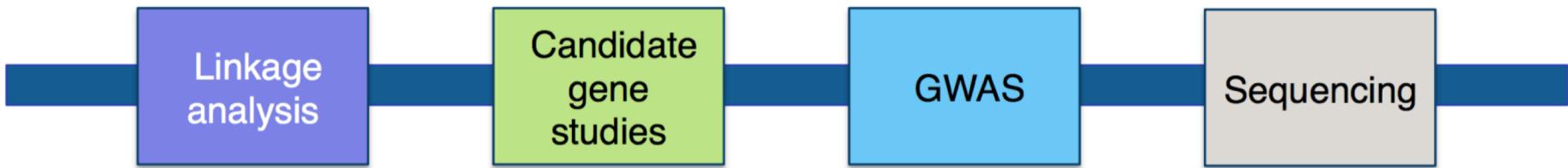
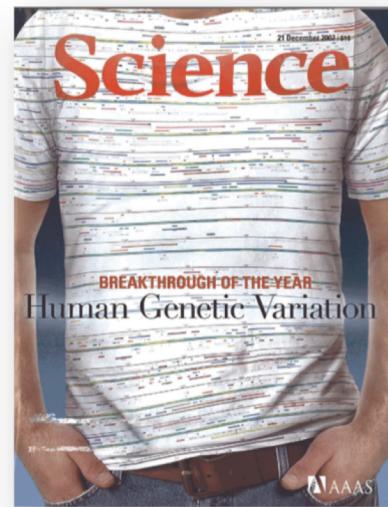
The limitations of genetics

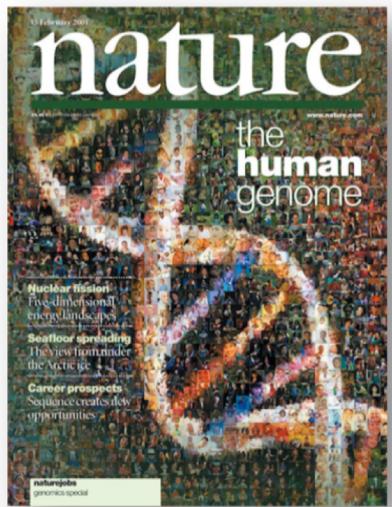
genotype  phenotype

The limitations of genetics

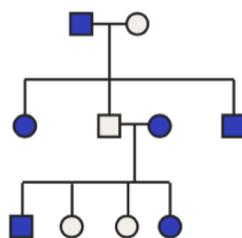


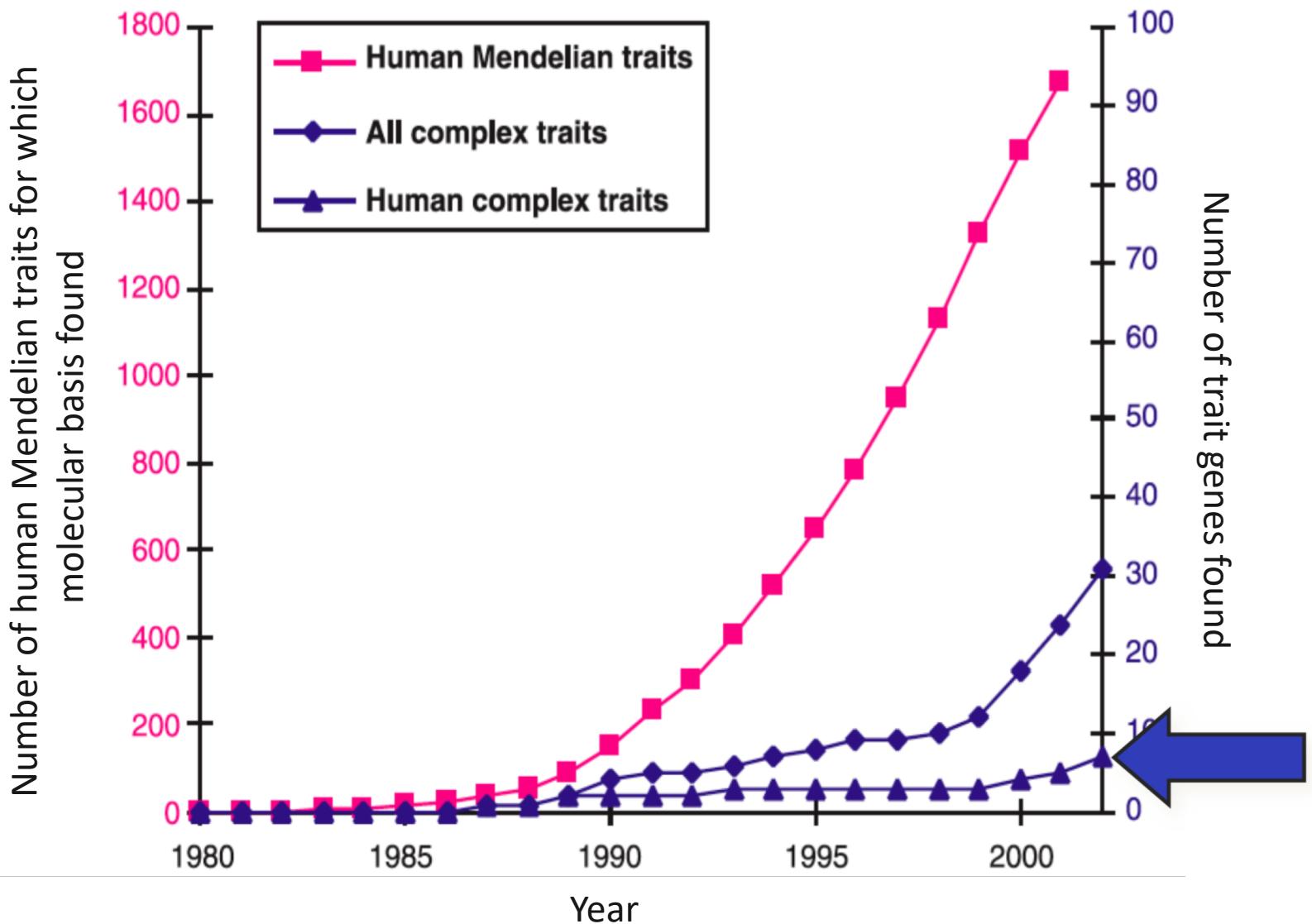
Where we've been and where we are

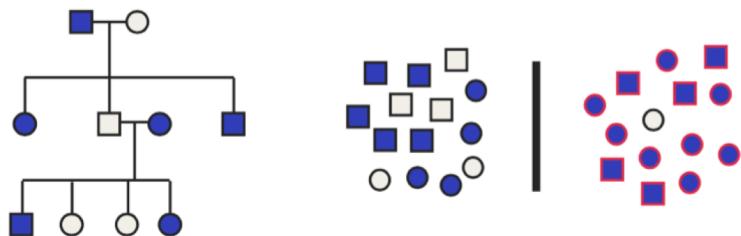
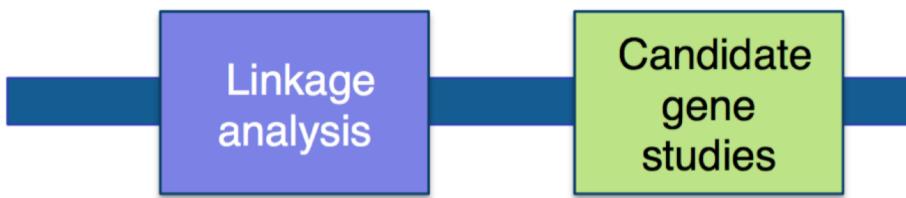




Linkage analysis







The candidate gene approach

- Pick a gene that might have a role in your disease
arbitrary
- Genotype individuals at a few sites around that gene
 - Typically 1,000 – 2,000 samples
no power
- Test genetic sites for association

A poor history of candidate gene studies

March/April 2002 · Vol. 4 · No. 2

review

A comprehensive review of genetic association studies

Joel N. Hirschhorn, MD, PhD^{1–3}, Kirk Lohmueller¹, Edward Byrne¹, and Kurt Hirschhorn, MD⁴

Most common diseases are complex genetic traits, with multiple genetic and environmental components contributing to susceptibility. It has been proposed that common genetic variants, including single nucleotide polymorphisms (SNPs), influence susceptibility to common disease. This proposal has begun to be tested in numerous studies of association between genetic variation at these common DNA polymorphisms and variation in disease susceptibility. We have performed an extensive review of such association studies. We find that over 600 positive associations between common gene variants and disease have been reported; these associations, if correct, would have tremendous importance for the prevention, prediction, and treatment of most common diseases. However, most reported associations are not robust: of the 166 putative associations which have been studied three or more times, only 6 have been consistently replicated. Interestingly, of the remaining 160 associations, well over half were observed again one or more times. We discuss the possible reasons for this irreproducibility and suggest guidelines for performing and interpreting genetic association studies. In particular, we emphasize the need for caution in drawing conclusions from a single report of an association between a genetic variant and disease susceptibility. **Genet Med 2002;4(2):45–61.**

Key Words: human genetics, association studies, common disease, polymorphisms

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

PloS Medicine, 2005

The candidate gene problem:

- Lack of statistical rigor
- Lack of large samples
- Lack of data quality control
- Lack of replication data

Need systematic, unbiased approach

Problems with the candidate gene approach

- Small sample sizes
- Weak effects
- No community-wide standards for QC, association claims
- Population stratification

Important side note: this still happens

OPEN  ACCESS Freely available online

PLOS GENETICS

AVPR1a and SLC6A4 Gene Polymorphisms Are Associated with Creative Dance Performance

Psychiatr Q (2014) 85:257–265
DOI 10.1007/s11126-013-9287-x

ORIGINAL PAPER

The 2-Repeat Allele of the MAOA Gene Confers an Increased Risk for Shooting and Stabbing Behaviors

SCIENTIFIC REPORTS



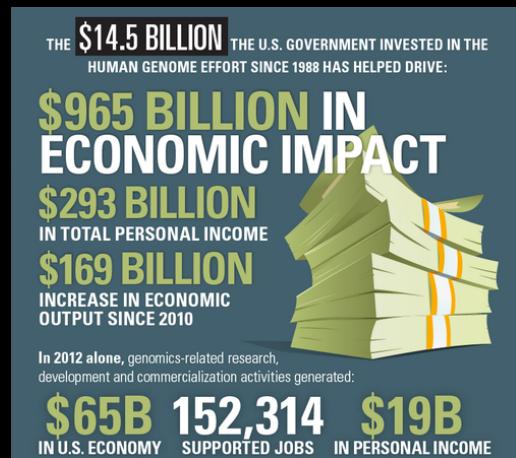
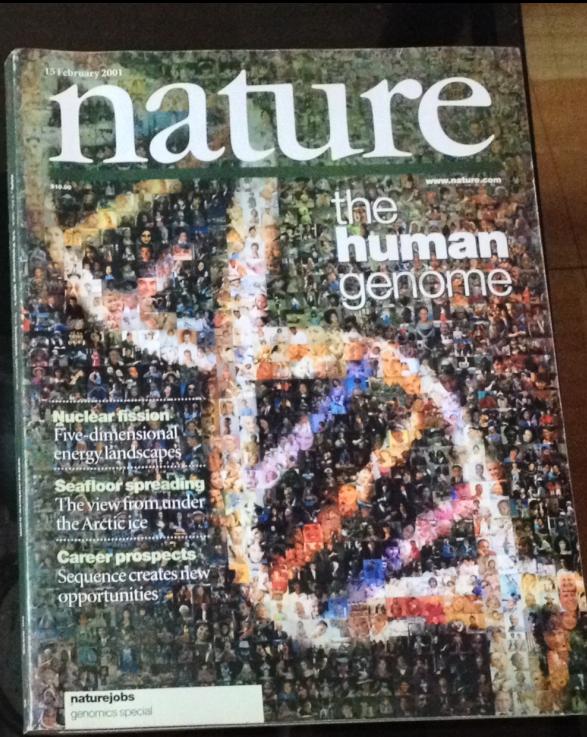
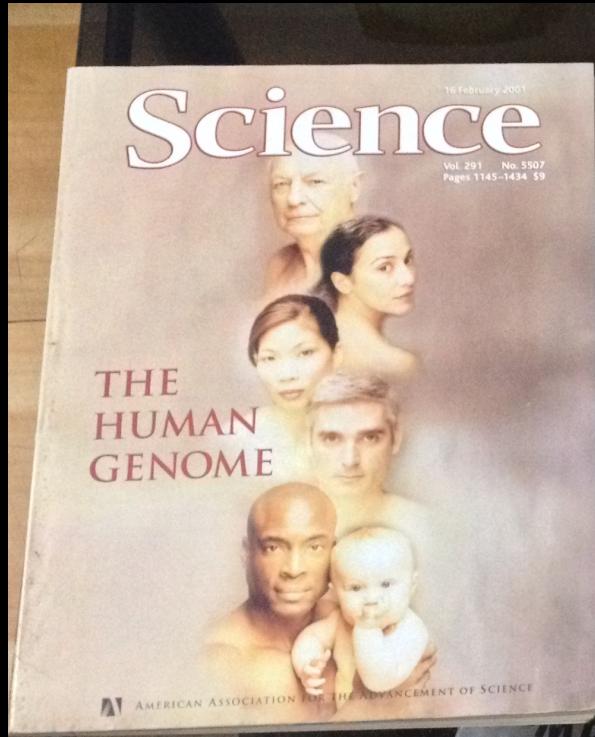
OPEN The association between romantic relationship status and 5-HT1A gene in young adults

SUBJECT AREAS:
HUMAN BEHAVIOUR
BEHAVIOURAL GENETICS

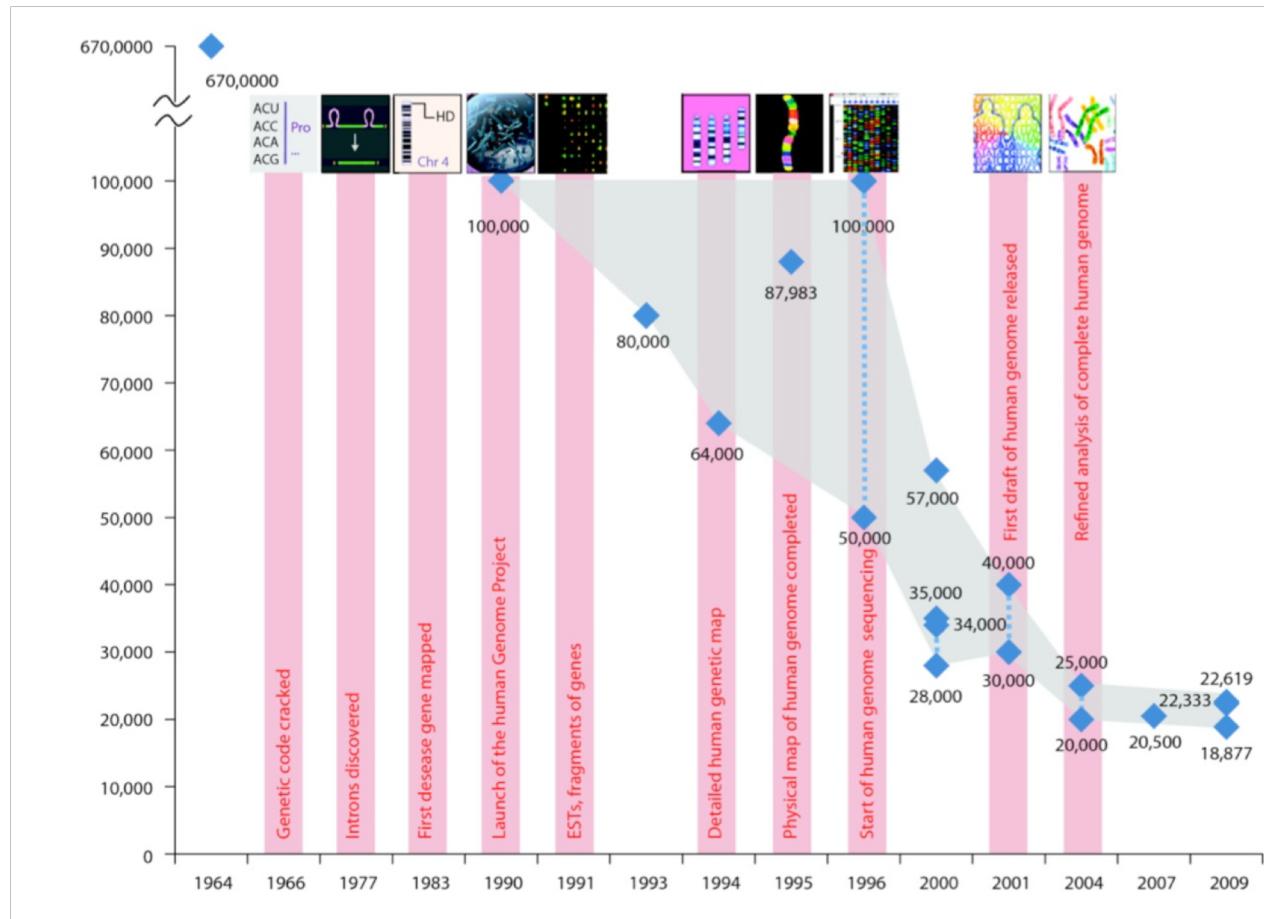




The Human Genome Project “paid forward” and paved the way for modern day genomics



(Finally) a complete map



Human Genome: *some statistics*

- 3.2 billion base pairs in the haploid genome
- \approx 18,000-25,000 genes
 - \approx 23,000 coding for proteins
 - Only 1.5% of the total genome
- Rest of the genome:
 - Non-coding RNA (rRNA, tRNA)
 - Regulatory sequences, e.g. promoter, enhancer regions
 - Repetitive elements and other variations
 - Transposable elements
- (So there's no such thing as "junk DNA"...)

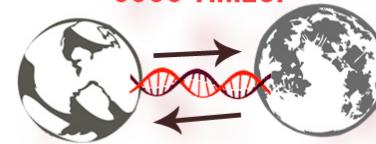
Our entire DNA sequence is called a genome... and there's an estimated **3,000,000,000** DNA bases in our genome.



A complete 3 billion base genome would take **3 GIGABYTES OF STORAGE SPACE.**



IF YOU UNWRAP ALL OF THE DNA YOU HAVE IN ALL YOUR CELLS, YOU COULD REACH THE MOON **6000 TIMES.**



99.9% OF OUR DNA SEQUENCE IS THE SAME AS OTHER HUMANS¹.



0.1% DNA DIFFERENCE
This 0.1% DNA difference between us may have to do with the number of nucleotides in a person's DNA.

When DNA is copied into a new life, the nucleotides are either gained or lost in the process. This gain or loss results in our differences.



= 50 YEARS

It would take a person typing 60 words per minute, 8 hours a day, around 50 years to type the human genome.

ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGCATCGATCTTGC
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTTACTTGC
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCAGCAAAG
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCATCGTACGACACATATCGTCATCGTACTGCCACGT
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGCACATATCGTCATAGCTCA
TCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTGGCA
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTATCCTATATTGA
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGCGGTC
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTAUTGACTGTCTAGTCTAAACACATCCCACATATGC
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGCATCGTACTGACTGTCTAGTCTAAACACATCCCACATCGC
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATATCGTCATCGTACTGACTGTCTAGTCTAAACACATATTG
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGCGCCT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTAUTGACTGTCTAGTCTAAACACATCCCACATATAATG
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAAACACATTTG
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATACATATCGTCATCGTACTGACTGTCTAGTCTAAACACACCCC
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGCGCTT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTAUTGACTGTCTAGTCTAAACACATCCCACATTTT
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCAC
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATAGCCGATCGTACGACACATATCGTCATCGTACTGCCCTGGCT
CTGTCAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGTACGACACATATCGTCATCGTACTGCCCTGGG
CTGTCAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGCACATATCGTCATACATAAAT
CGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCCCATGCGGC
ATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTATTCTAAACACATCCCAGCATGGCT
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTACTGCCCTAAAC
CTGTCAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGTACTGCACATATCGTCATACATACCCT
GTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTACCGGCT
ATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCAAT
CGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCCGG

ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGCATCGATCTTGC
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTTACTTGC
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCAGCAAAG
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCATCGTACGACACATATCGTCATCGTACTGCCACGT
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGCACATATCGTCATAGCTCA
TCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTGGCA
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTATCCTATATTGA
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGCGGTC
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTA

GACTGTCTAAACACATCCCAC
CGTACTGACTGTCTAGTCTAAACACATCCCAC
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATATCGTCATCGTACTGACTGTCTAGTCTAAACACATATTG
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGCGCCT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTA
CGTACTGACTGTCTAGTCTAAACACATCCCAC
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATACATATCGTCATCGTACTGACTGTCTAGTCTAAACACACCCCC
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGCGCTT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTA
CGTACTGACTGTCTAGTCTAAACACATCCCAC
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATAGCCGATCGTACGACACATATCGTCATCGTACTGCCCTGGCT
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGACACACATATCGTCATCGTACTGCCCTGGG
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCACATATCGTCATACATAAAT
CGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCCCATGCGGC
ATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATGGCT
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGACACACATATCGTCATCGTACTGCCCTAAAC
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGACTGCATCGTACTGACTGCACATATCGTCATACATAACCCT
GTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTACCGGCT
ATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCAAT
CGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCCGG

Most of genetic variation is due to *single nucleotide polymorphisms (SNPs)* -- single base changes that are common in the general population

Human genome: *individual variations*

- Human genome is ~99 % similar between individuals
- 0.5-1% different

articles

A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms

The International SNP Map Working Group*

* A full list of authors appears at the end of this paper.

We describe a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome, providing an average density on available sequence of one SNP every 1.9 kilobases. These SNPs were primarily discovered by two projects: The SNP Consortium and the analysis of clone overlaps by the International Human Genome Sequencing Consortium. The map integrates all publicly available SNPs with described genes and other genomic features. We estimate that 60,000 SNPs fall within exon (coding and untranslated regions), and 85% of exons are within 5 kb of the nearest SNP. Nucleotide diversity varies greatly across the genome, in a manner broadly consistent with a standard population genetic model of human history. This high-density SNP map provides a public resource for defining haplotype variation across the genome, and should help to identify biomedically important genes for diagnosis and therapy.

The International HapMap Project

Phase I

1.1 million SNPs

270 individuals from 4 populations



Phase II

3.1 million SNPs

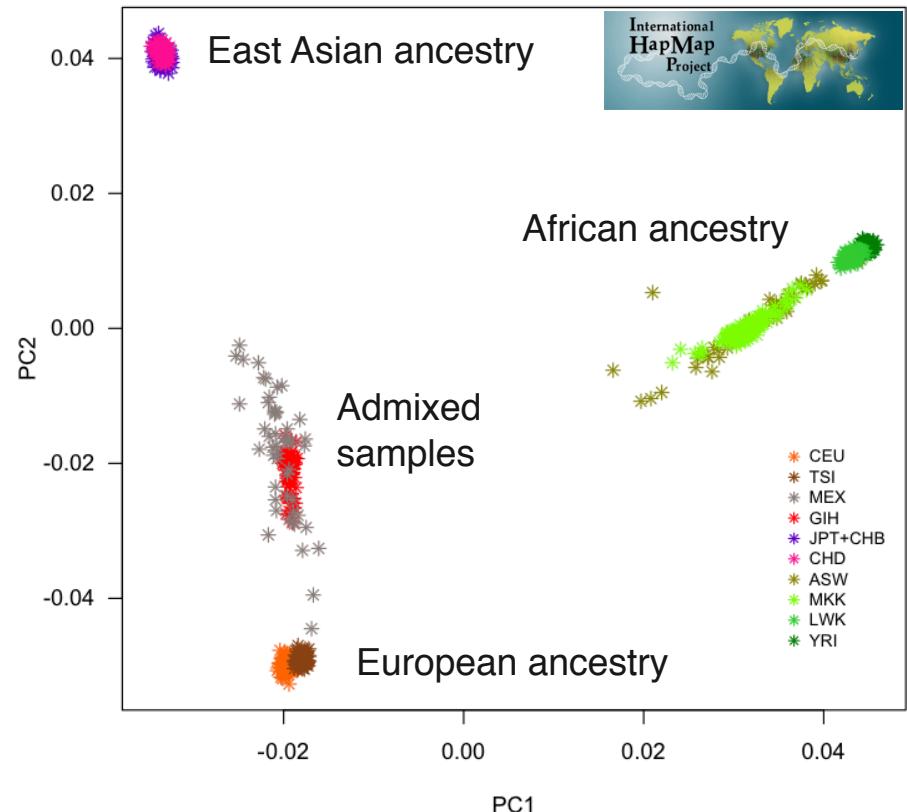
270 individuals from 4 populations



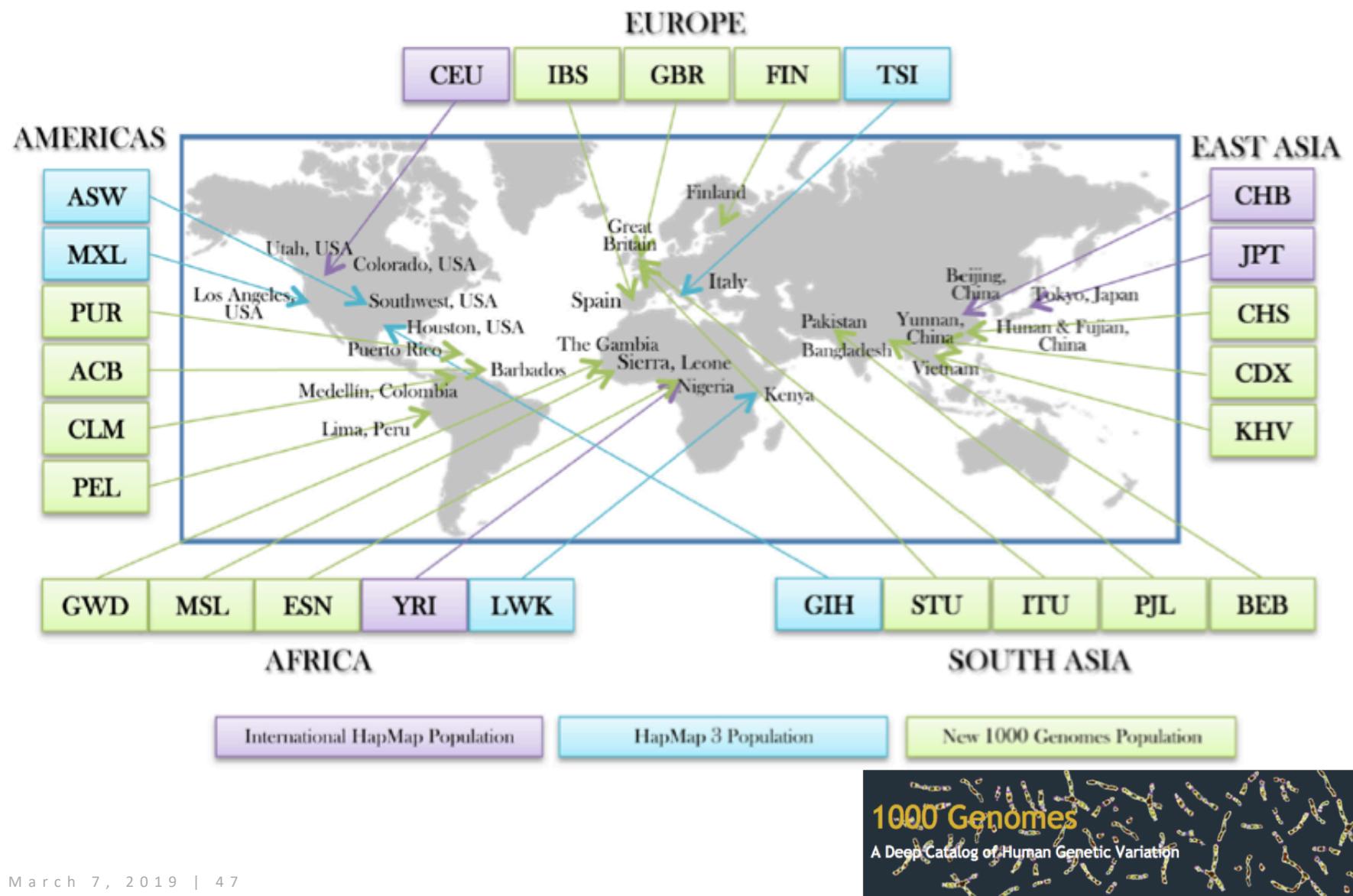
Phase III

1.6 million SNPs

1,184 individuals from 11 populations



The 1000 Genomes Project

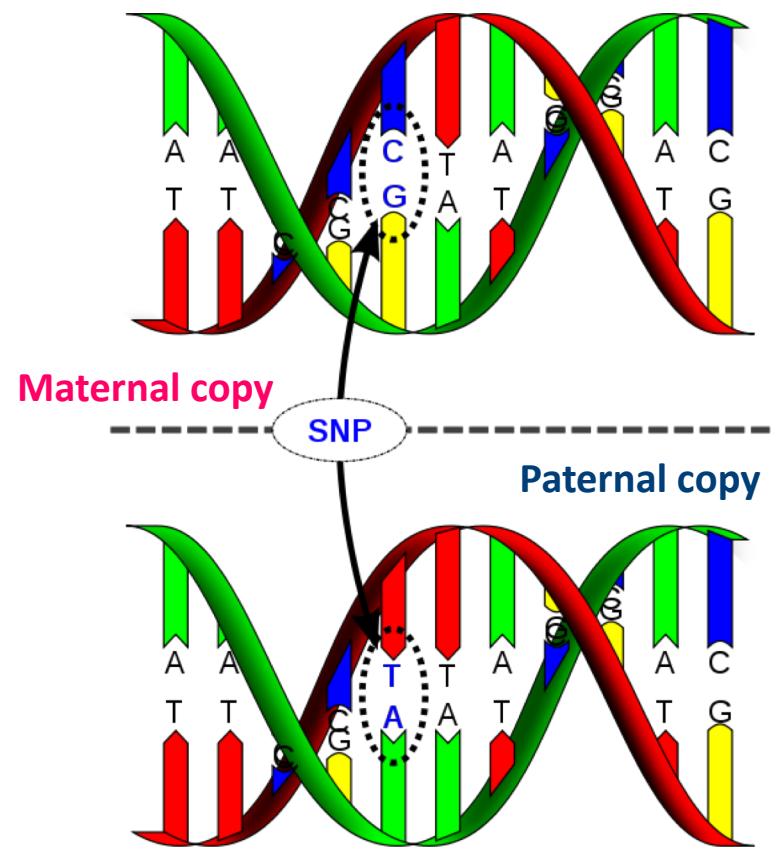


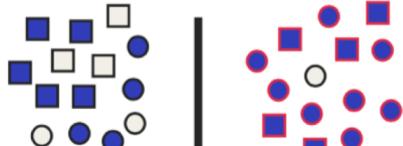
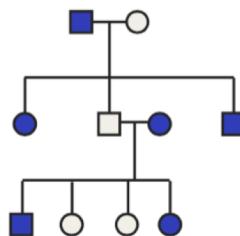
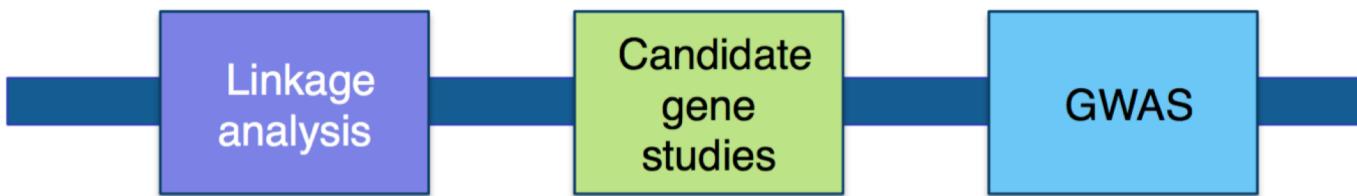
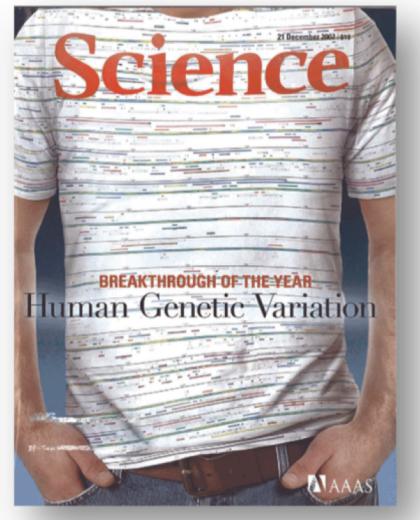
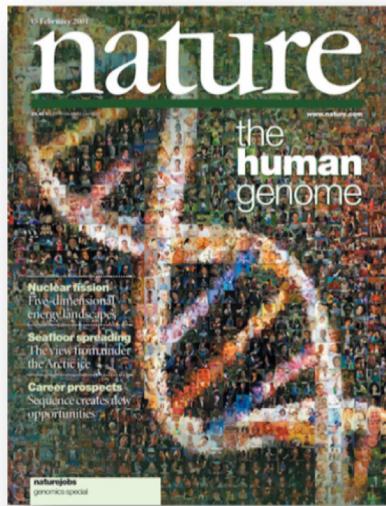
Single-Nucleotide Polymorphism

- “one base pair variation”
 - > 1% general population (common)
 - \approx 10 million SNPs (\approx 0.25% genome)
 - Makes you and me unique
 - Most common type of genetic variation



www.hapmap.org





Common variant, common disease hypothesis

- Most common diseases happen later in life
- If common variants are not selected against, they may associate to late-onset (after reproduction) disease
- Common variants are easier to find and characterize

The beginnings of GWAS

HapMap Phase I

HapMap Phase II

SNP arrays

WTCCC GWAS

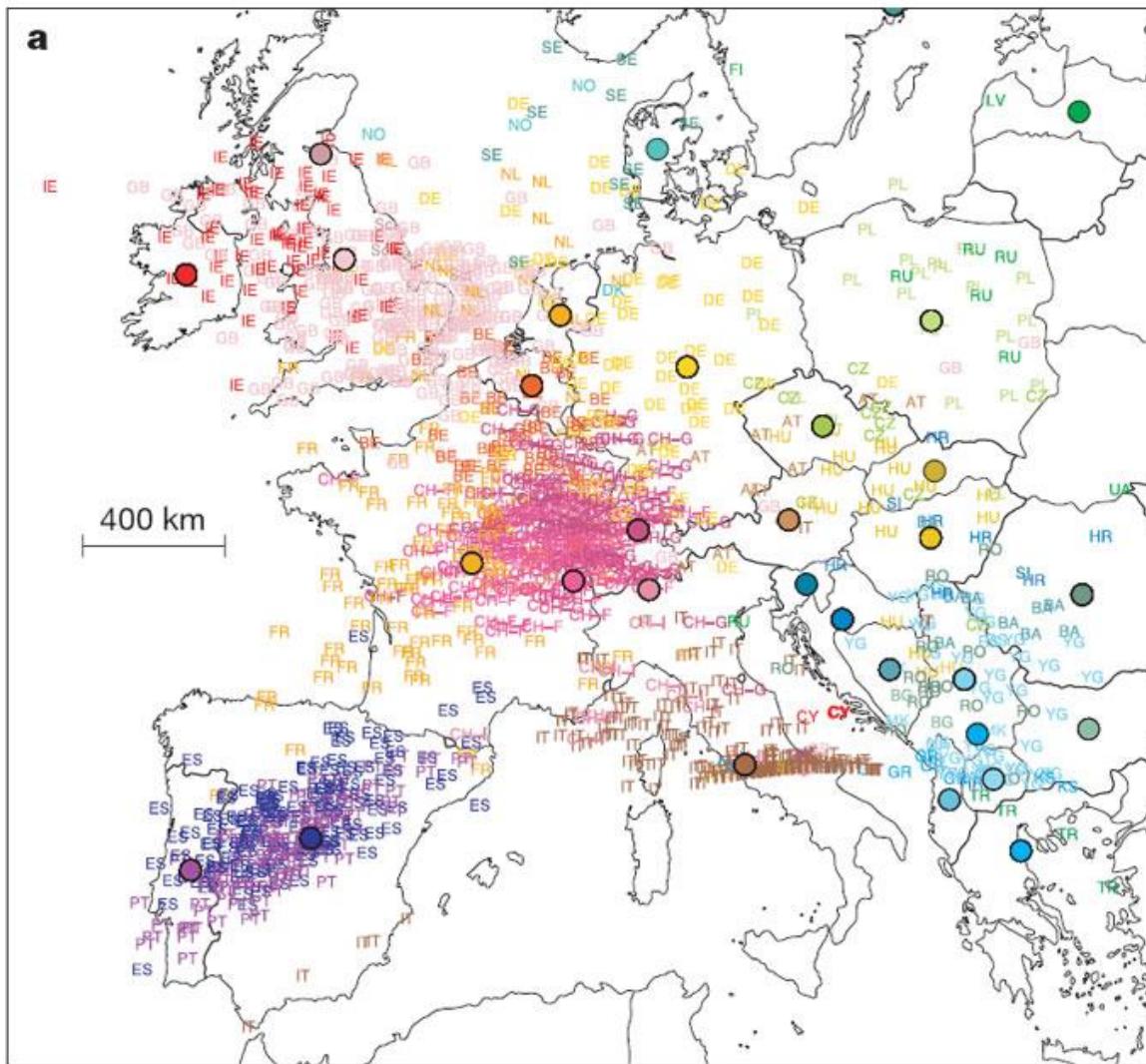
HapMap Phase III



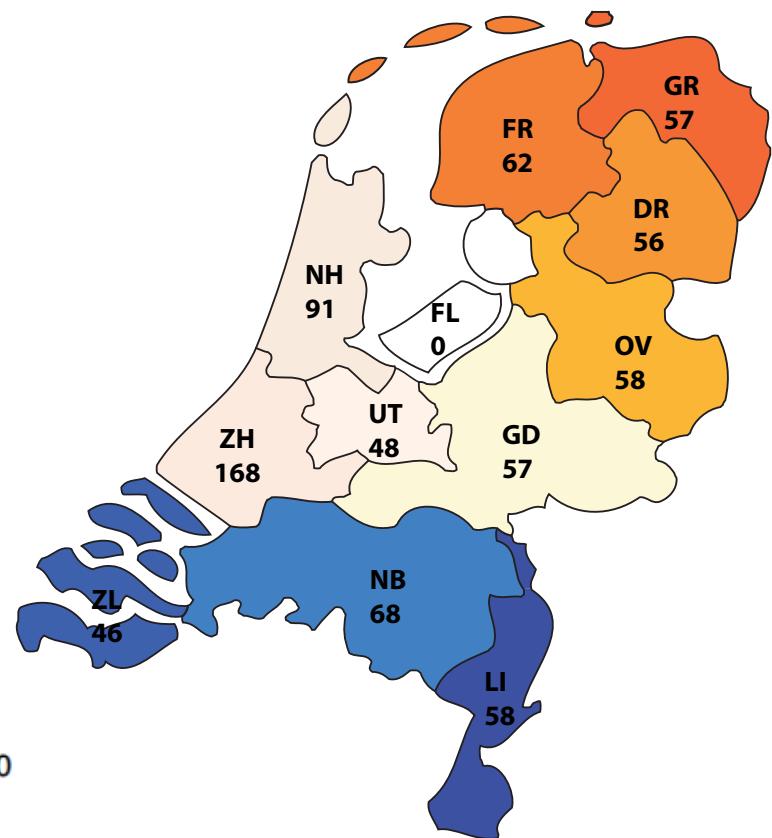
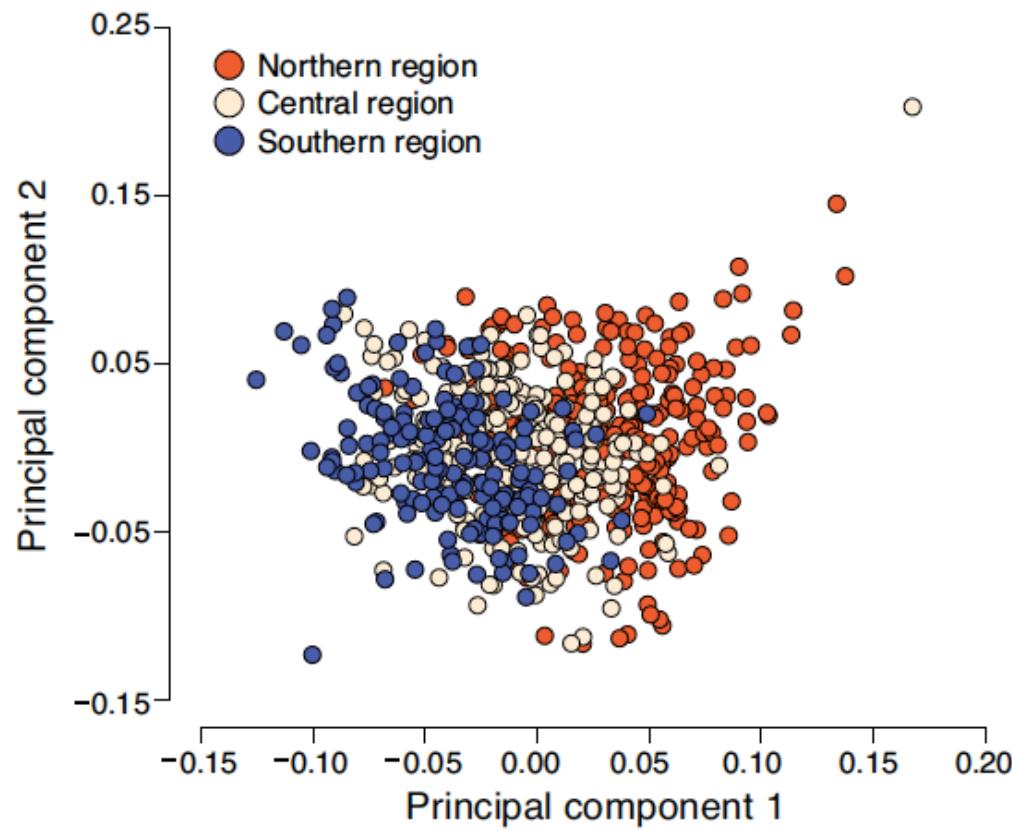
2003

2010

Differentiation of populations



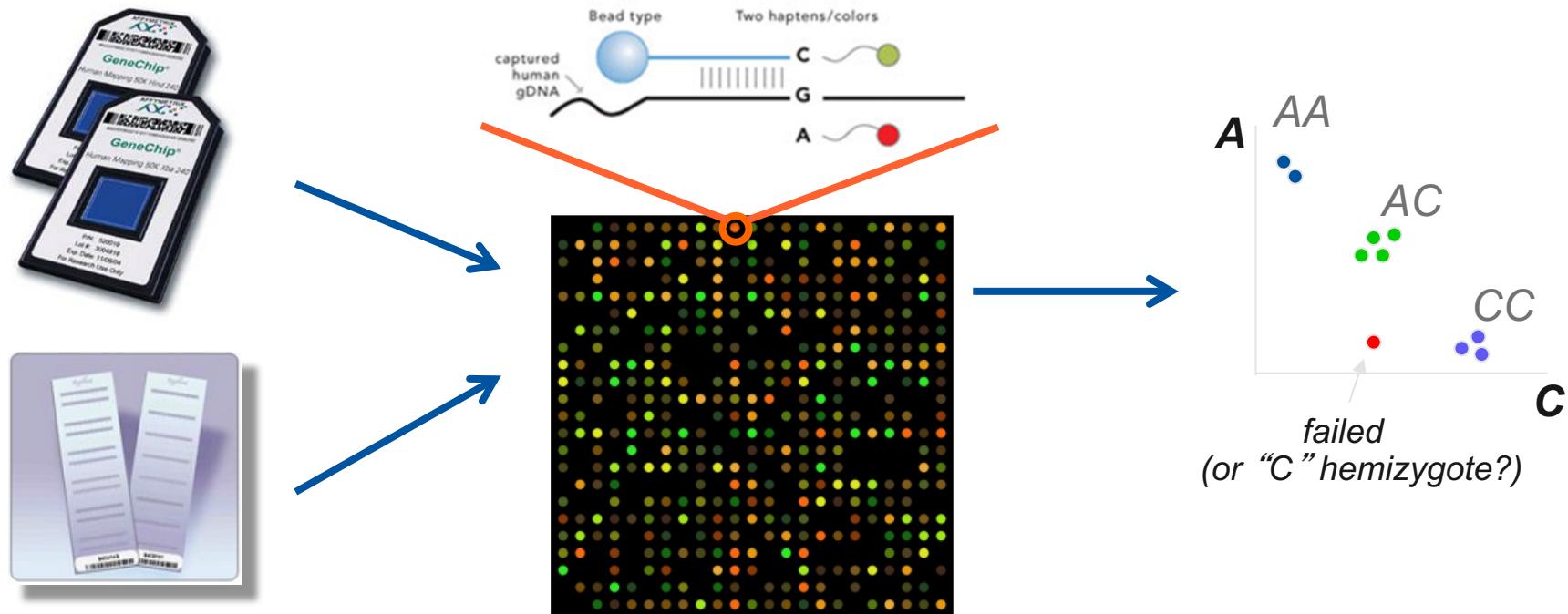
Within the Netherlands, North-South substructure



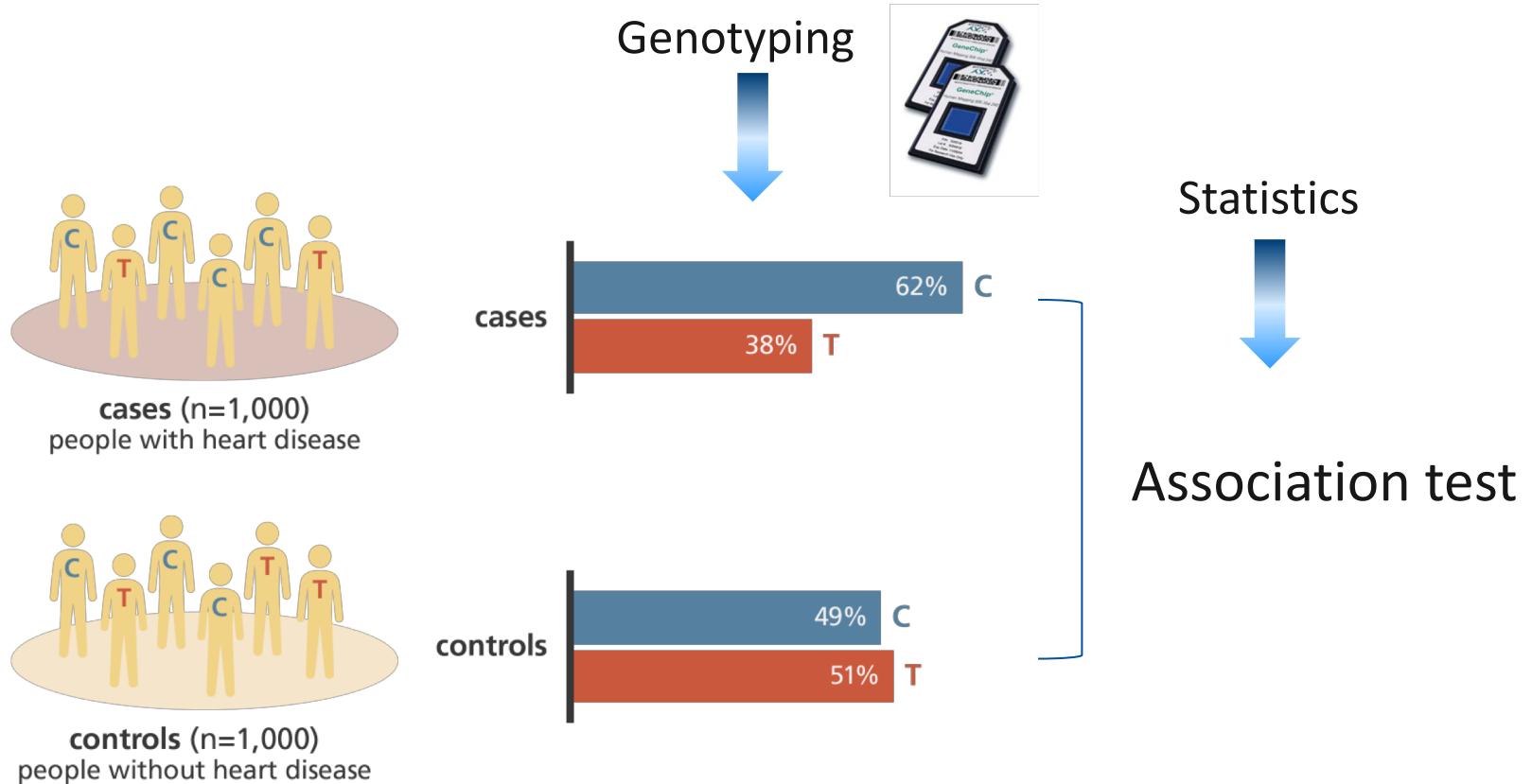
Genome of the Netherlands (250 families whole-genome sequenced)

Genotyping platforms

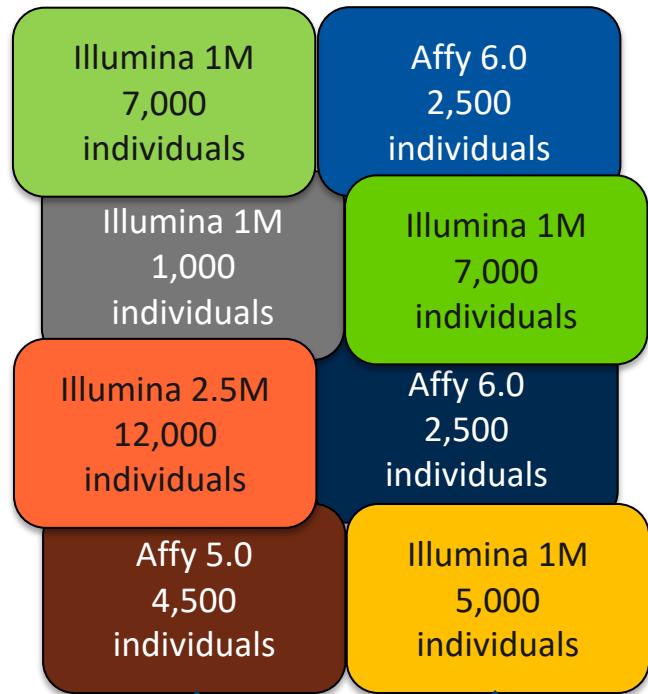
- Genome-wide SNP microarrays allow measurement of genotypes of 100,000's of SNPs in a single experiment
- Variety of microarrays (different SNP density, cost, etc) by Illumina and Affymetrix



GWAS (the big picture)

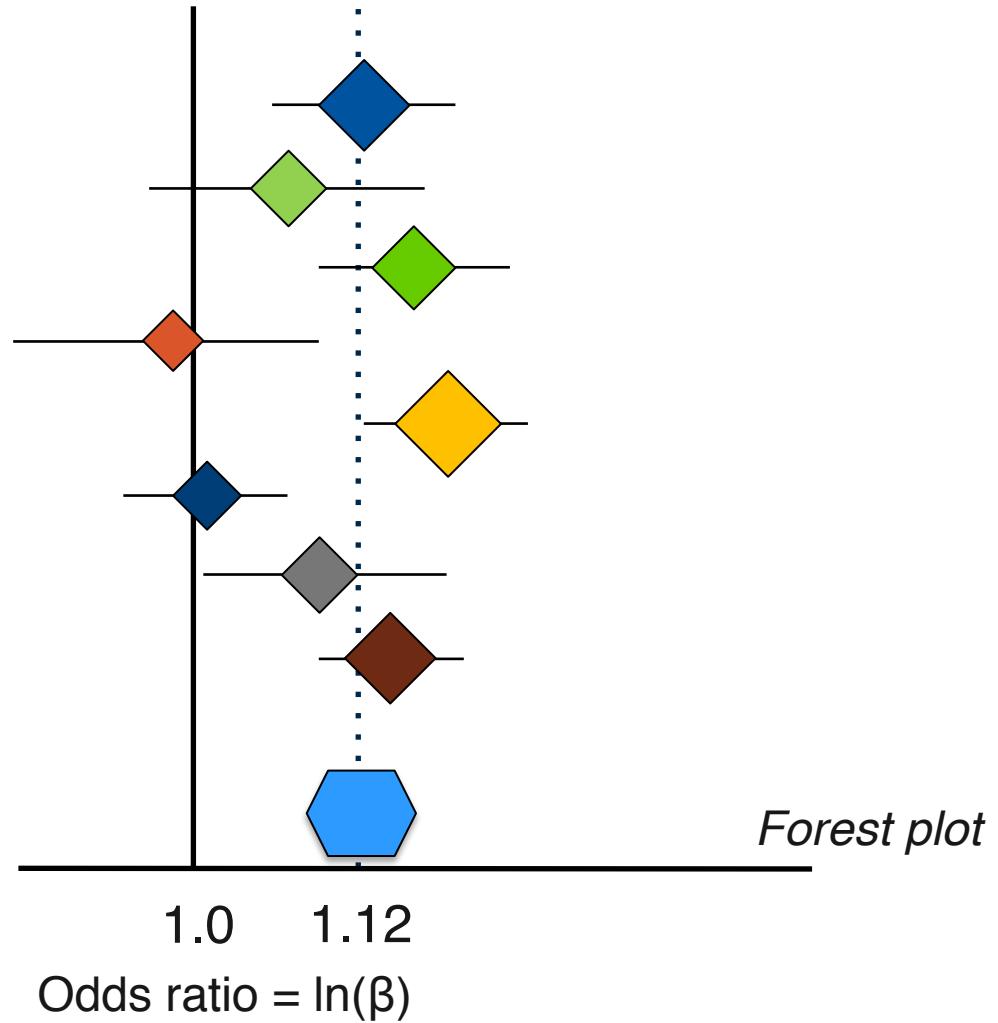


Combining GWAS datasets



Imputation
↓
Meta-analysis of GWAS

Results for one SNP



deCODE Genetics, Inc.

- >50% adult population of Iceland (>140,000) in biobank (blood)
- Pedigree information going back to the first settlements (\approx 1000 years ago)
- Extensive medical records & genotypic data
- Over 250 high-impact publications (Nature, Science, AJHG)
- 50 common diseases
 - Stroke (=CVA) association with *ALOX5AP*
 - MI association with *ALOX5AP*
 - Association of a variant on 9p21.1 with Abdominal aortic aneurysm (AAA), intracranial aneurysm, stroke and MI



The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke

Anna Helgadottir¹, Andrei Manolescu¹, Gudmar Thorleifsson¹, Solveig Gretarsdottir¹, Helga Jonsdottir¹, Unnur Thorsteinsdottir¹, Nilesh J Samani², Gudmundur Guðmundsson¹, Struan F A Grant¹, Gudmundur Thorgeirsson³, Sigurlaug Sveinbjornsdottir³, Einar M Valdimarsson³, Stefan E Matthiasson³, Halldor Johannsson³, Olof Guðmundsdottir¹, Mark E Gurney¹, Jesus Sainz¹, Margaret Thorhallsdottir¹, Margaret Andressdottir¹, Michael L Frigge¹, Eric J Topol⁴, Augustine Kong¹, Vilimundur Gudnason⁵, Hakon Hakonarson¹, Jeffrey R Gulcher¹ & Kari Stefansson¹

We mapped a gene predisposing to myocardial infarction to a locus on chromosome 13q12–13. A four-marker single-nucleotide polymorphism (SNP) haplotype in this locus spanning the gene *ALOX5AP* encoding 5-lipoxygenase activating protein (FLAP) is associated with a two times greater risk of myocardial infarction in Iceland. This haplotype also confers almost two times greater risk of stroke. Another *ALOX5AP* haplotype is associated with myocardial infarction in individuals from the UK. Stimulated neutrophils from individuals with myocardial infarction produce more leukotriene B4, a key product in the 5-lipoxygenase pathway, than do neutrophils from controls, and this difference is largely attributed to cells from males who carry the at-risk haplotype. We conclude that variants of *ALOX5AP* are involved in the pathogenesis of both myocardial infarction and stroke by increasing leukotriene production and inflammation in the arterial wall.

Helgadottir, A., et al. *Nature Genetics*; volume 36, 233; 2004

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir,^{1,*} Gudmar Thorleifsson,^{1,*} Andrei Manolescu,^{1*} Solveig Gretarsdottir,¹ Thorarinn Blöndal,¹ Aslaug Jonasdottir,¹ Adalbjorg Jonasdottir,¹ Asgeir Sigurdsson,¹ Adam Baker,¹ Amar Palsson,¹ Gisli Masson,¹ Daniel F. Gudbjartsson,¹ Kristinn P. Magnusson,¹ Karl Andersen,² Allan I. Levey,³ Valgerdur M. Backman,¹ Sigurborg Matthiasdottir,¹ Thorbjorg Jonsdottir,¹ Stefan Palsson,¹ Helga Einarsdottir,¹ Steinunn Gunnarsdottir,¹ Arnaldur Gylfason,¹ Viola Vaccarino,³ W. Craig Hooper,³ Muredach P. Reilly,⁴ Christopher B. Granger,⁵ Harland Austin,³ Daniel J. Rader,⁴ Svti H. Shah,⁵ Arshed A. Quyyumi,³ Jeffrey R. Gulcher,¹ Gudmundur Thorgeirsson,² Unnur Thorsteinsdottir,¹ Augustine Kong,^{1,†} Kari Stefansson^{1,‡}

Helgadottir, A., et al. *Science* volume 316, 1491; 2007

Wellcome Trust Case-Control Consortium

- 1,500 1958 Birth Cohort Controls (58BC)
- 1,500 UK Blood Services Controls (UKBS)
- 14,000 cases of seven common diseases
 - Bipolar disorder
 - **Coronary artery disease**
 - Crohn's disease
 - **Hypertension**
 - Rheumatoid arthritis
 - **Type 1 diabetes**
 - **Type 2 diabetes**

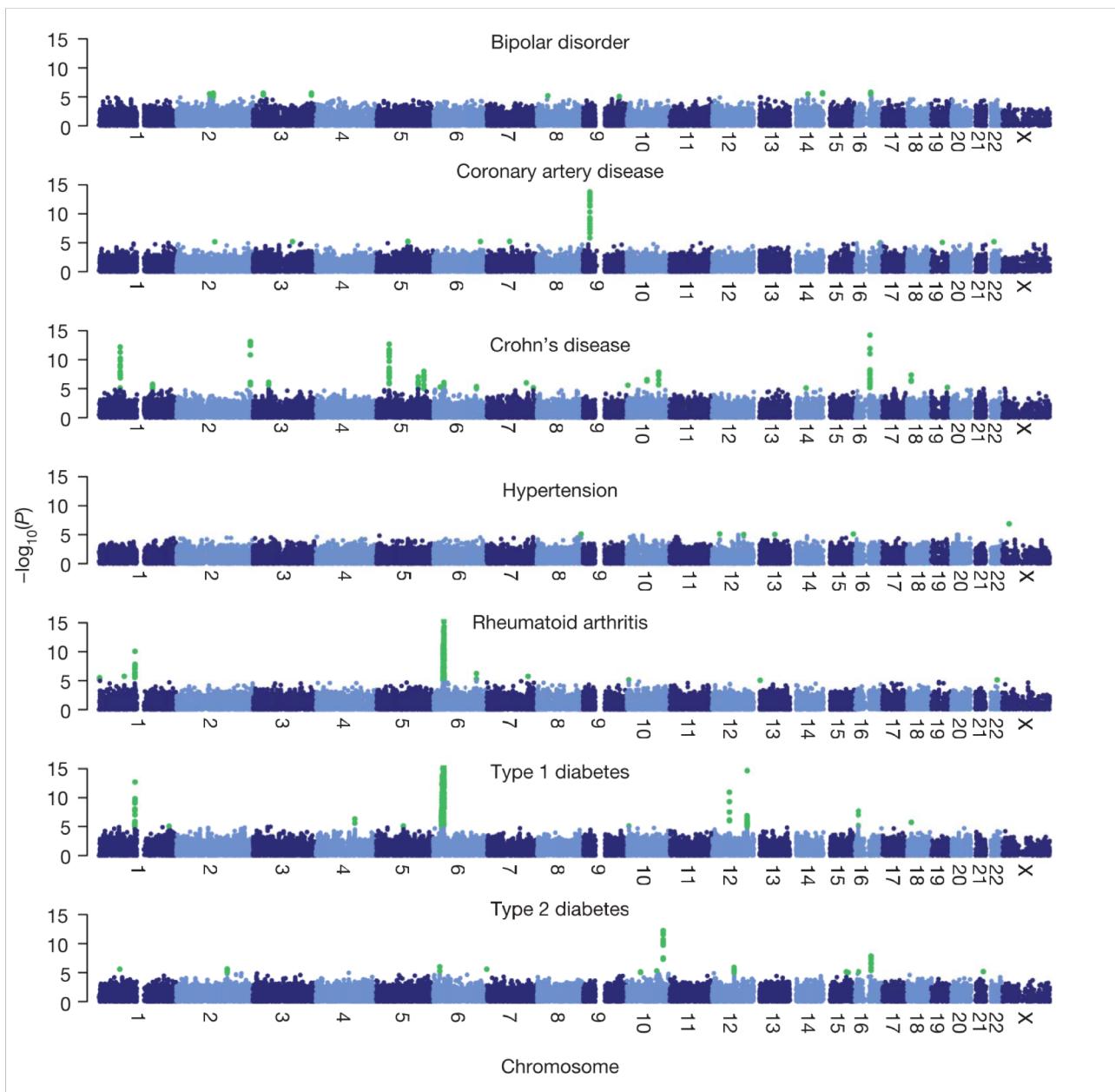
Vol 447 | 7 June 2007 | doi:10.1038/nature05911

nature

ARTICLES

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*



One famous example

- deCODE Genetics was the first to discover a SNP associated with myocardial infarction (MI) in 2007
- WTCCC, McPherson, and Samani were able to replicate the same finding in the same year, and many have reconfirmed it in different populations



A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Anna Helgadottir,^{1,*} Gudmar Thorleifsson,^{1,*} Andrei Manolescu,^{1,*} Solveig Gretarsdottir,¹ Thorarinn Blonadal,¹ Aslaug Jonasdottir,¹ Adalbjorg Jonasdottir,¹ Asgeir Sigurdsson,¹ Adam Baker,¹ Amar Palsson,¹ Gisli Masson,¹ Daniel F. Gudbjartsson,¹ Kristinn P. Magnusson,¹ Karl Andersen,² Allan I. Levey,³ Valgerdur M. Backman,¹ Sigurborg Matthiassdottir,¹ Thorbjorg Jonsdottir,¹ Stefan Palsson,¹ Helga Einarsdottir,¹ Steinunn Gunnarsdottir,¹ Amaldur Gylfason,¹ Viola Vaccarino,³ W. Craig Hooper,³ Muredach P. Reilly,⁴ Christopher B. Granger,⁵ Harland Austin,³ Daniel J. Rader,⁴ Svti H. Shah,⁵ Arshed A. Quyyumi,³ Jeffrey R. Gulcher,¹ Gudmundur Thorgeirsson,² Unnur Thorsteinsdottir,¹ Augustine Kong,^{1,†} Kari Stefansson¹

A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

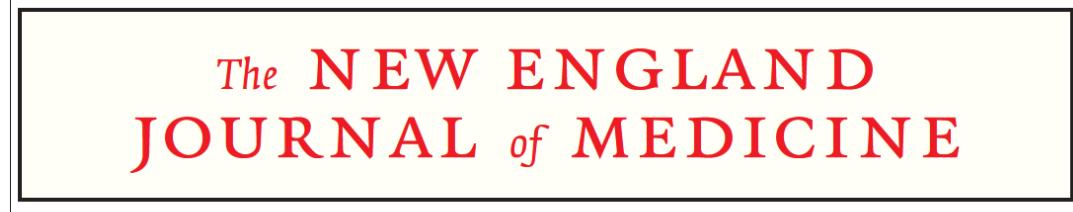
Ruth McPherson,^{1,*†} Alexander Pertsemlidis,^{2,*} Nihan Kavaslar,¹ Alexandre Stewart,¹ Robert Roberts,¹ David R. Cox,³ David A. Hinds,³ Len A. Pennacchio,^{4,5} Anne Tybjaerg-Hansen,⁶ Aaron R. Folsom,⁷ Eric Boerwinkle,⁸ Helen H. Hobbs,^{2,9} Jonathan C. Cohen^{2,10†}

Helgadottir, A., et al. *Science*; 316(5830):1491-1493, 2007

McPherson, R., et al. *Science*; 316(5830):1488-1491, 2007

Wellcome Trust Case Control Consortium. *Nature*; 447(7145):661-678, 2007

Samani, N.J., et al. *N Engl J Med*; 357(5):443-453, 2007



Genomewide Association Analysis of Coronary Artery Disease

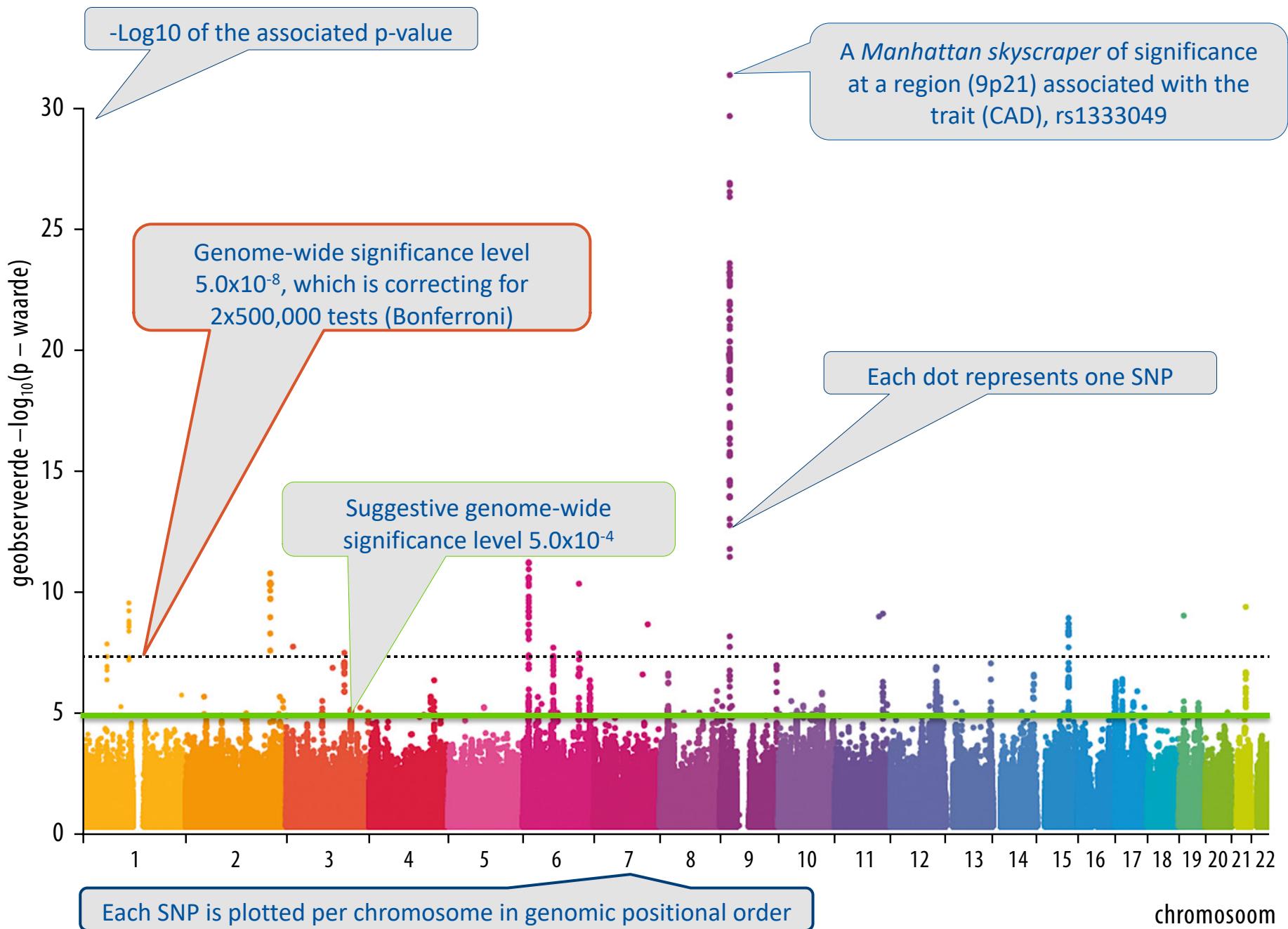
Vol 447 | 7 June 2007 | doi:10.1038/nature05911

nature

ARTICLES

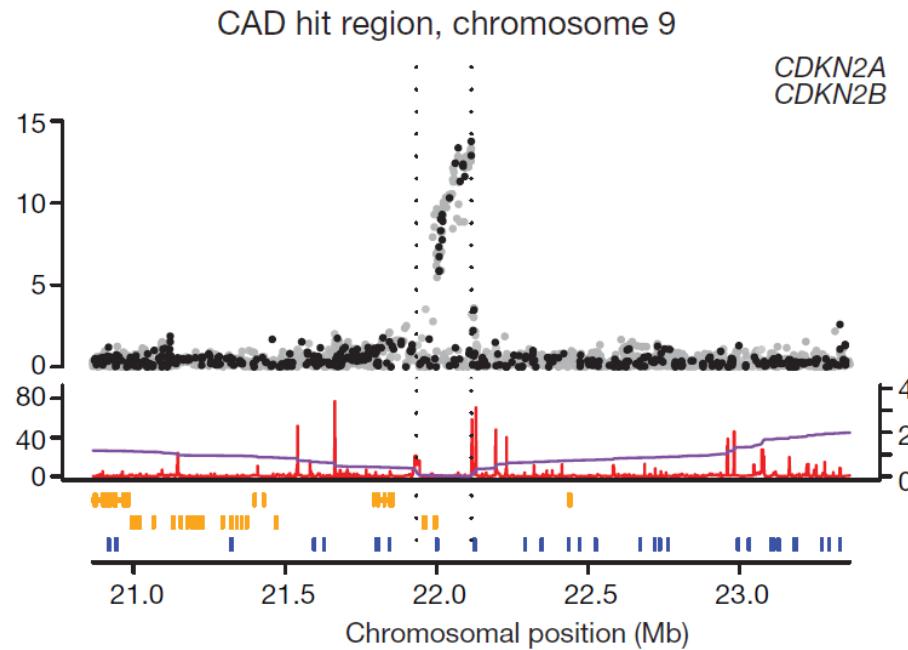
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*



9p21 and cardiovascular disease

- The SNPs associated with CAD on 9p21.1 are rs1333049, rs10757274, rs2383207, rs2891168, and rs10757278
- They are found in an *intergenic region*
- Genes nearby: *CDKN2A*, *CDKN2B*
 - also associated with *type 2 diabetes mellitus*
 - regulating cell proliferation, cell aging and the associated degeneration, and programmed cell death of many cell types



Wellcome Trust Case Control Consortium. *Nature*; 447(7145):661-678, 2007

A closer look at the results...

Table 3 | Regions of the genome showing the strongest association signals

Collection	Chromosome	Region (Mb)	SNP	Trend P value	Genotypic P value	$\log_{10}(BF)$, additive	$\log_{10}(BF)$, general	Risk allele	Minor allele	Heterozygote odds ratio	Homozygote odds ratio	Control MAF	Case MAF
CAD	9p21	21.93-22.12	rs1333049	1.79×10^{-14}	Standard analysis 1.16×10^{-13}	11.66	11.19	C	C	1.47 (1.27-1.70)	1.9 (1.61-2.24)	0.474	0.554

- CAD: coronary artery disease
- 9p21: chromosome 9, short arm (p)
- Region: 21.93-22.12 megabase pairs
- rs1333049: official dbSNP ID

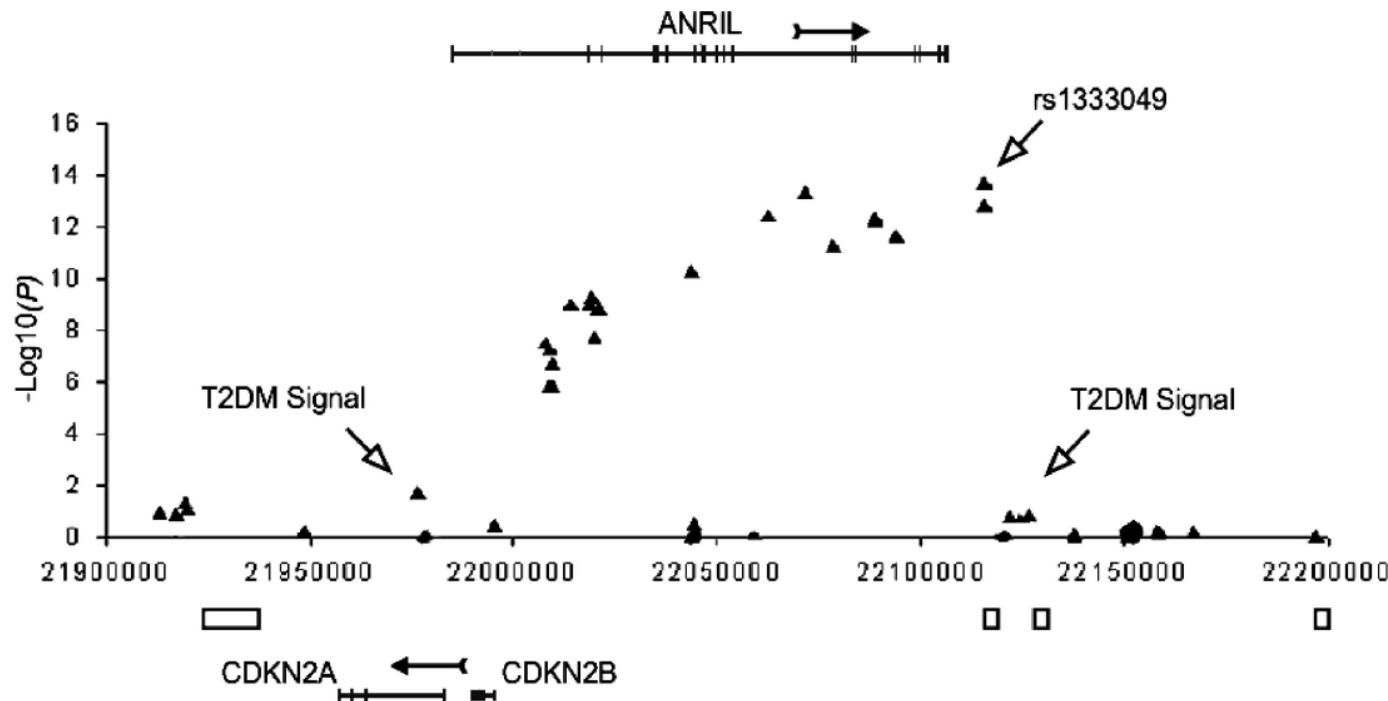
Risk allele: minor allele
 Odds ratio: the odds of exposure between cases and controls

P-value of association test: AA vs. AB vs. BB

Minor allele frequency: the frequency of the risk (minor) allele in the population

9p21 points to a RNA gene

- Resequencing unveiled a RNA gene, *ANRIL*
- Current efforts are aimed to elucidate the role of *ANRIL* in (A)MI
- Might be involved in *early-onset MI* (before age of 50 years)



Samani, NJ., et al. Circ Cardiovasc Genet; 1:81-84, 2008

CARDIoGRAMplusC4D Study

- Coronary Artery Disease Genome–Wide Replication And Meta–Analysis Study: CARDIoGRAM
- > 63,000 cases and > 130,000 controls
 - Myocardial infarction (MI), coronary artery disease (CAD) or both
 - CAD: MI, CABG, PTCA, AP
 - Age limit: 45–66
- Sample size greatly influences power and effect size to discover new variants
- CARDIoGRAMplusC4D sought to solves this issue
- 55 susceptibility loci for CAD were discovered



ARTICLES

Large-scale association analysis identifies new risk loci for coronary artery disease

The CARDIoGRAMplusC4D Consortium¹

Coronary artery disease (CAD) is the commonest cause of death. Here, we report an association analysis in 63,746 CAD cases and 130,681 controls identifying 15 loci reaching genome-wide significance, taking the number of susceptibility loci for CAD to 46, and a further 104 independent variants ($r^2 < 0.2$) strongly associated with CAD at a 5% false discovery rate (FDR). Together, these variants explain approximately 10.6% of CAD heritability. Of the 46 genome-wide significant lead SNPs, 12 show a significant association with a lipid trait, and 5 show a significant association with blood pressure, but none is significantly associated with diabetes. Network analysis with 233 candidate genes (loci at 10% FDR) generated 5 interaction networks comprising 85% of these putative genes involved in CAD. The four most significant pathways mapping these networks are linked to lipid metabolism and inflammation, underscoring the causal role of these activities in the genetic etiology of CAD. Our study provides insights into the genetic basis of CAD and identifies key biological pathways.

© 2012 Nature America, Inc. All rights reserved.

NPG

Coronary artery disease and its main complication, myocardial infarction, is the leading cause of death worldwide. Although, epidemiological studies have identified many risk factors for CAD, including plasma lipid concentrations, blood pressure, smoking, diabetes and markers of inflammation, a causal role has been proven only for some (for example, low-density lipoprotein (LDL) cholesterol and blood pressure), primarily through randomized clinical trials of drug therapy directed at the risk factor¹. Twin and family studies have documented that a significant proportion (40–50%) of susceptibility to CAD is heritable (for a review, see ref. 2). Because genotypes are not confounded by environmental exposures, genetic analysis has the potential to define which risk factors are indeed causal and to identify pathways and therapeutic targets^{3,4}. To date, genome-wide association studies (GWAS) have collectively reported a total of 31 loci, associated with CAD risk at genome-wide significance ($P < 5 \times 10^{-8}$)^{5–13}. However, variants at these loci explain less than 10% of the heritability of CAD. One likely reason for this is that, given the polygenic nature of complex traits and the relatively small observed effect sizes of the loci identified, many genuinely associated variants do not reach the stringent P -value threshold for genome-wide significance. Indeed, there is increasing evidence that the genetic architecture of common traits involves a large number of causative alleles with very small effects¹⁴. Addressing this will require the discovery of additional loci while leveraging large-scale genomic data to identify the molecular pathways underlying the pathogenesis of CAD. Such discovery is facilitated by building molecular networks, on the basis of DNA, RNA and protein interactions, which have nodes of known biological function that also show evidence of association with risk variants for CAD and related metabolic traits.

In the largest GWAS meta-analysis of CAD undertaken to date by the Coronary Artery Disease Genome-wide Replication and

Meta-analysis (CARDIoGRAM) Consortium⁵, which involved 22,233 cases and 64,762 controls, in addition to loci reported at genome-wide significance, a linkage disequilibrium (LD)-pruned set of 6,222 variants achieved a nominal association P value of less than 0.01. Here, we test these 6,222 SNPs in a meta-analysis of over 190,000 individuals, with the primary aim of identifying additional susceptibility loci for CAD. To this end, we used the Metabochip array¹⁵, which is a custom iSELECT chip (Illumina) containing 196,725 SNPs, designed to (i) follow-up putative associations in several cardiometabolic traits, including CAD, and (ii) fine map confirmed loci for these traits. All SNPs on the array with data in the CARDIoGRAM study were considered for analysis (79,138 SNPs, of which 6,222 were the replication SNPs and 20,876 were fine-mapping SNPs in the 22 CAD susceptibility loci identified at the time at which the array was designed; the remaining SNPs were submitted by the other consortia contributing to the Metabochip array¹⁵). In addition, we assess whether the genome-wide significant CAD risk alleles act through traditional risk factors by considering the available large GWAS for these traits^{16–20}. Finally, we identify a broader set of SNPs passing a conservative FDR threshold for association with CAD and use this set to undertake network analysis to find key biological pathways underlying the pathogenesis of CAD.

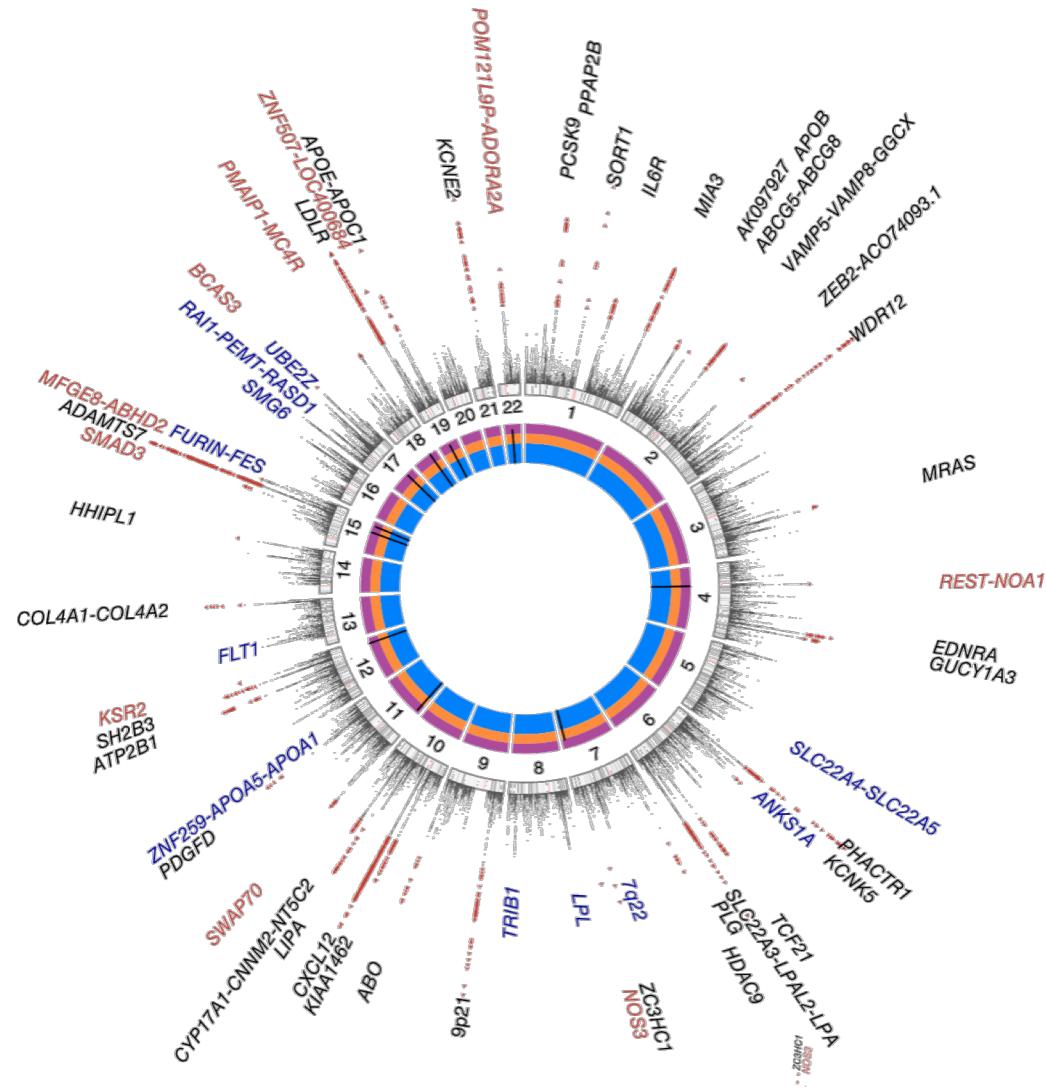
RESULTS Study design

We expanded the CARDIoGRAM discovery data set (22,233 cases and 64,762 controls⁵; stage 1) with 34 additional CAD sample collections (stage 2) of European or south Asian descent comprising 41,513 cases and 65,919 controls (study descriptions and sample characteristics are given in **Supplementary Tables 1a** and **2a**, respectively) and undertook a 2-stage meta-analysis to test SNPs on the Metabochip array

¹A full list of authors and affiliations appears at the end of the paper.

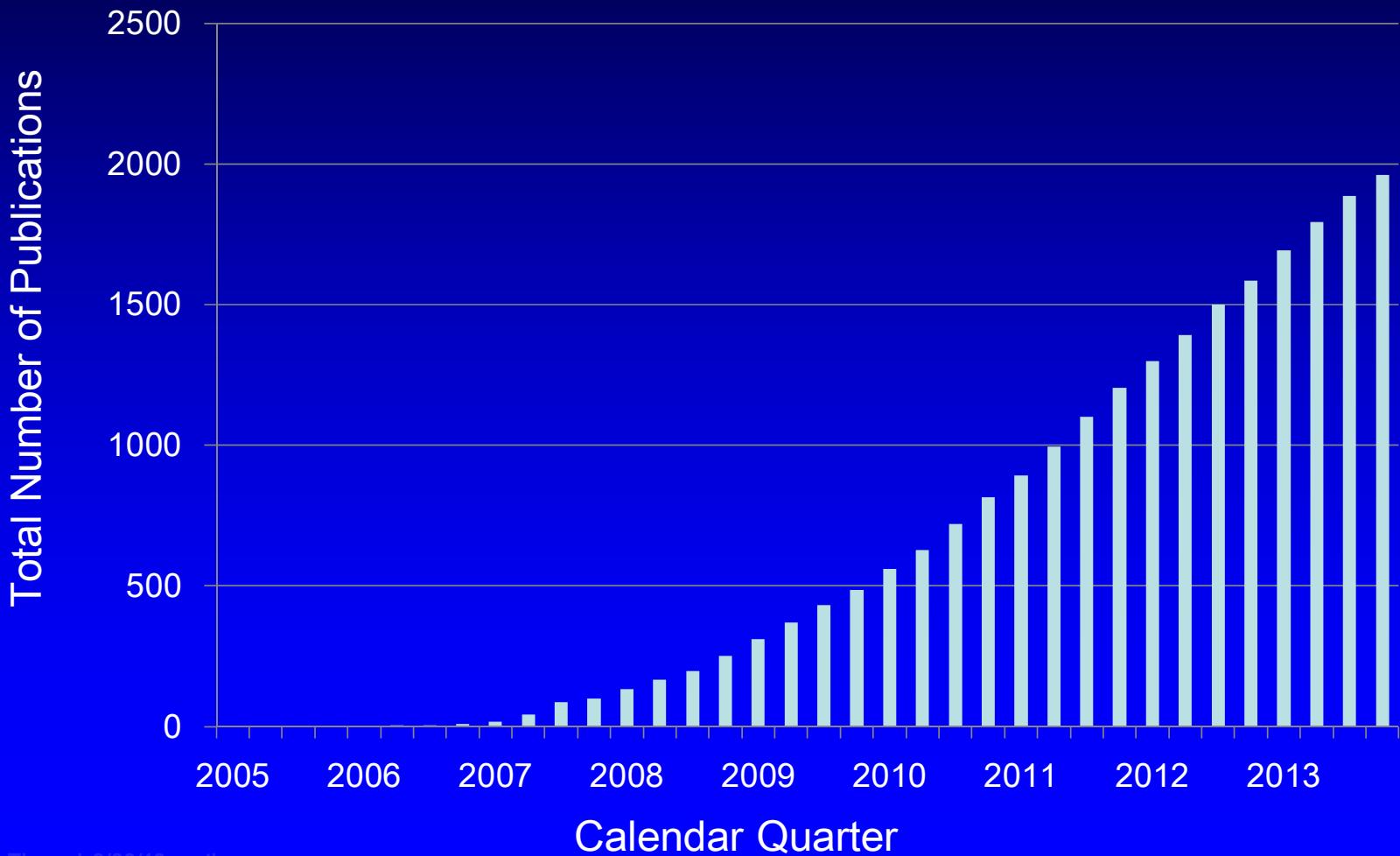
Received 24 April; accepted 2 November; published online 2 December 2012; doi:10.1038/ng.2480

And 8 years later (>15 times more samples)



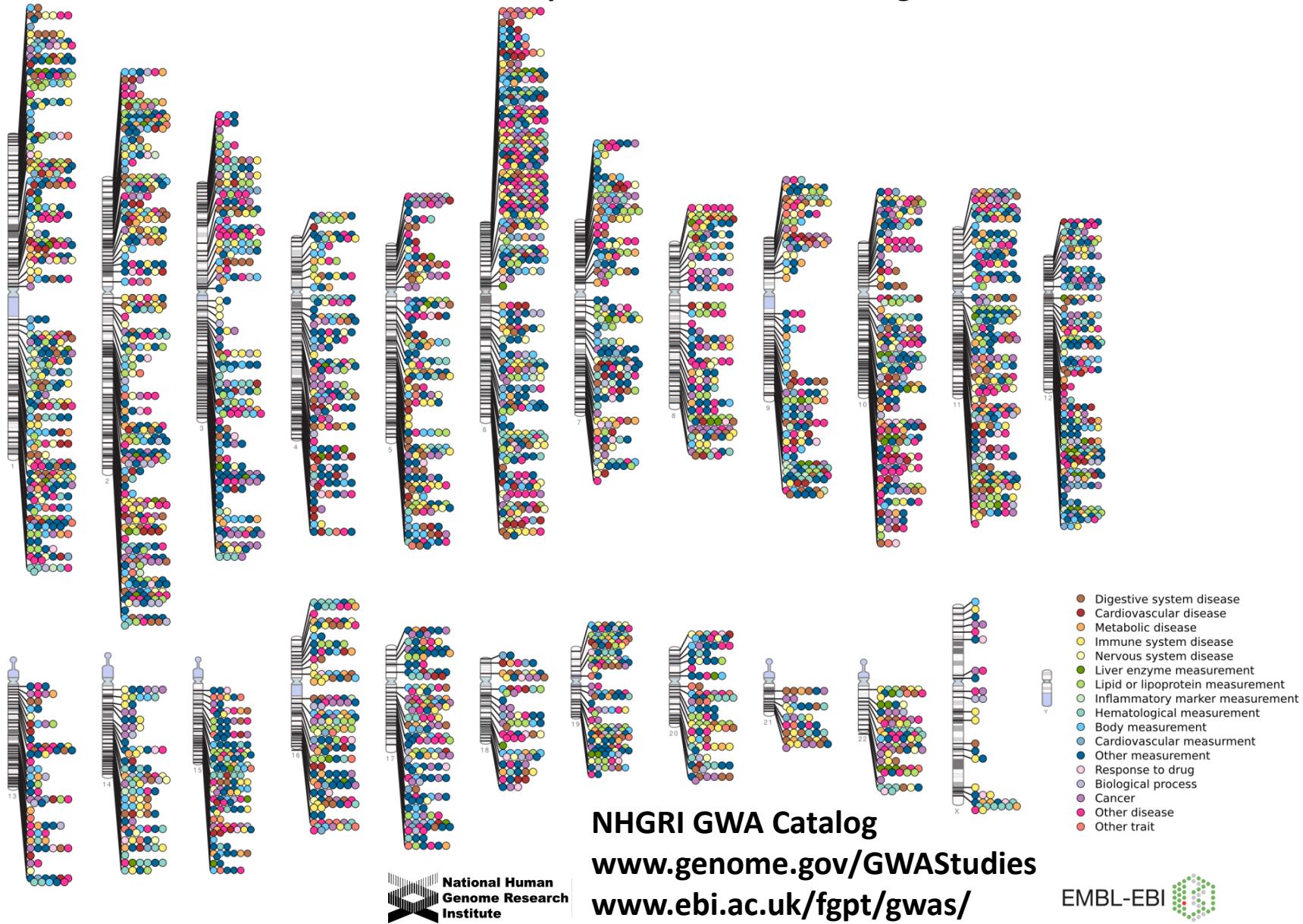
Published GWA Reports, 2005 – 2013

1960



Published Genome-Wide Associations through 12/2013

Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories



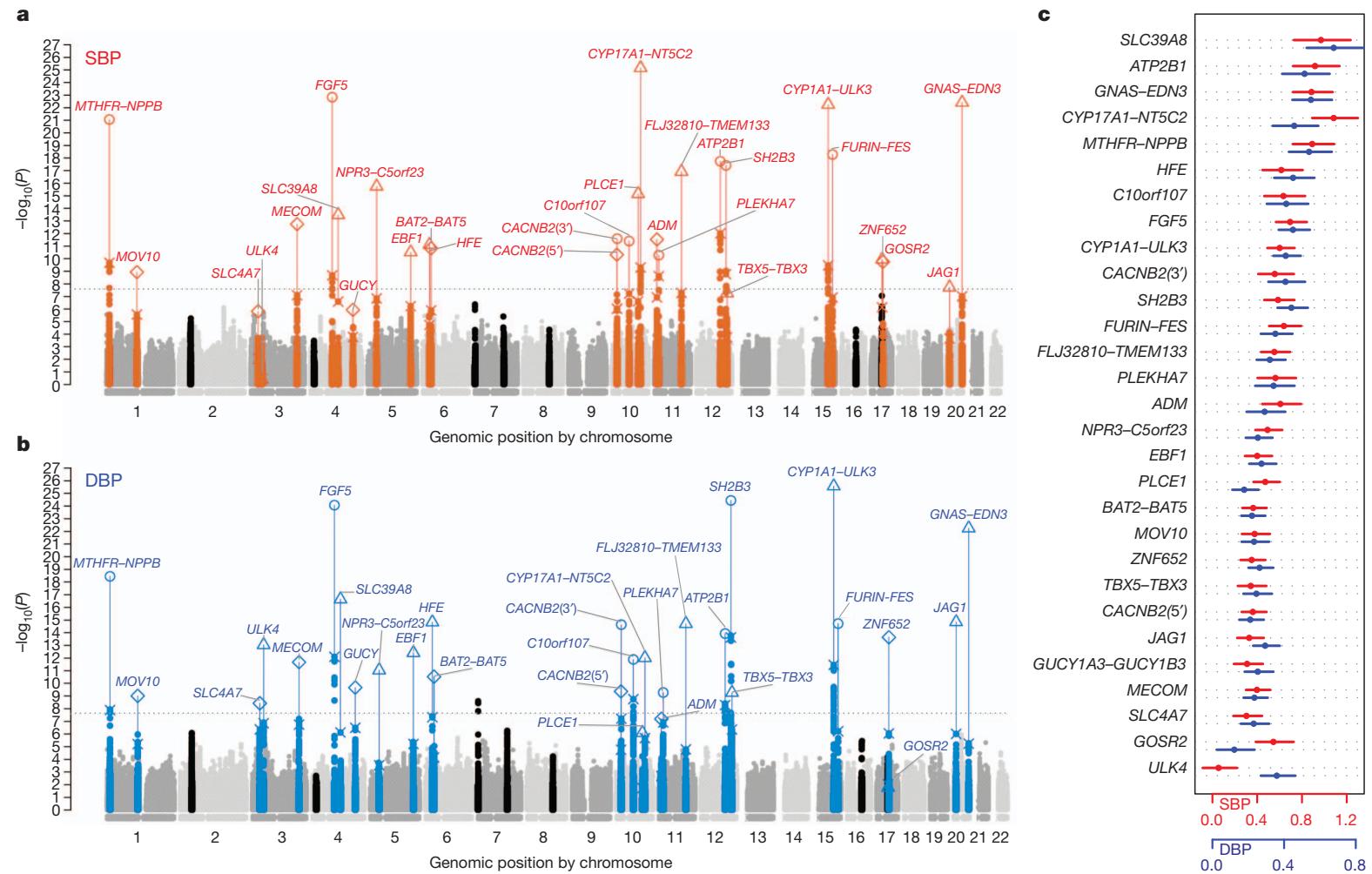


Figure 1 | Genome-wide $-\log_{10} P$ -value plots and effects for significant loci. **a, b**, Genome-wide $-\log_{10} P$ -value plots are shown for SBP (a) and DBP (b). SNPs within loci reaching genome-wide significance are labelled in red for SBP and blue for DBP (± 2.5 Mb of lowest P value) and lowest P values in the initial genome-wide analysis as well as the results of analysis including validation data are labelled separately. The lowest P values in the initial GWAS are denoted with a X. The range of different sample sizes in the final meta-

analysis including the validation data are indicated as: circle (96,000–140,000), triangle ($>140,000$ – $180,000$) and diamond ($>180,000$ – $220,000$). SNPs near unconfirmed loci are in black. The horizontal dotted line is $P = 2.5 \times 10^{-8}$. GUCY denotes GUCY1A3–GUCY1B3. **c**, Effect size estimates and 95% confidence bars per blood-pressure-increasing allele of the 29 significant variants for SBP (red) and DBP (blue). Effect sizes are expressed in mm Hg per allele.

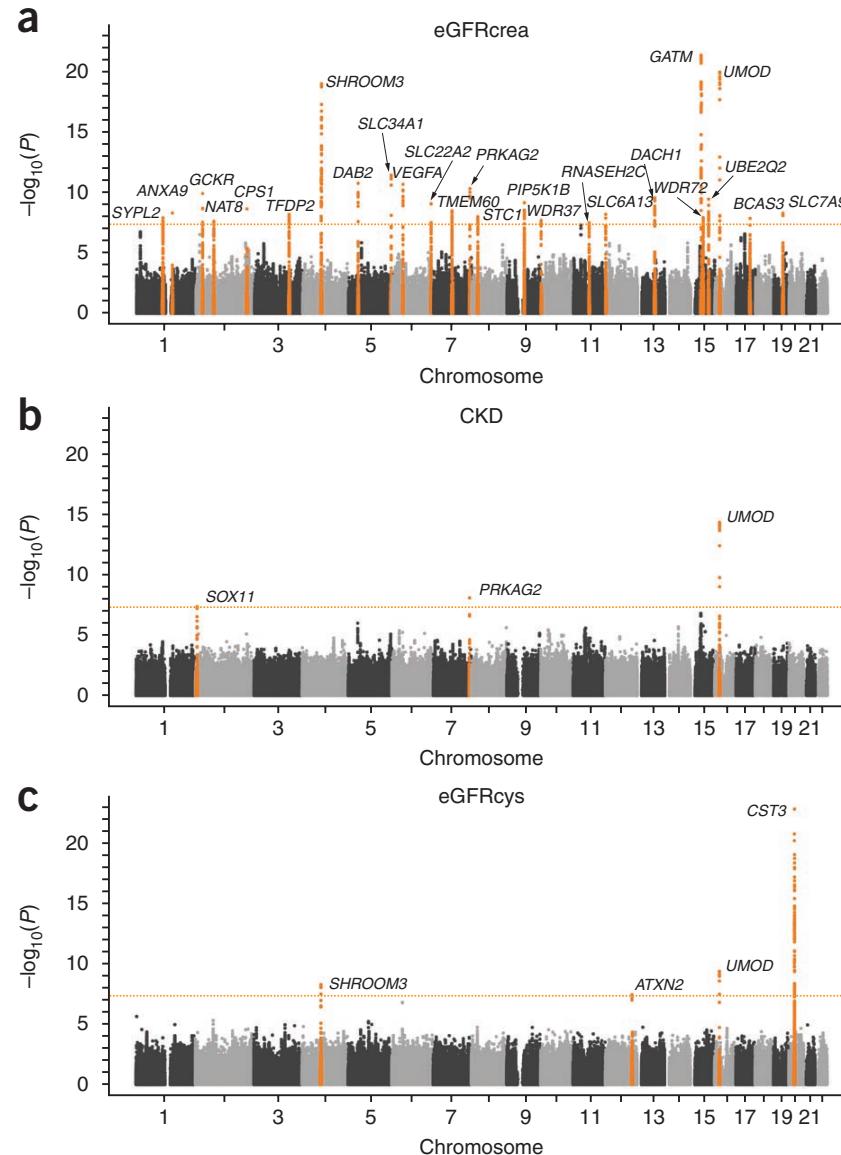


Figure 1 Genome-wide $-\log_{10} P$ value plot from stage 1. (a–c) Plots show discovery analysis of eGFRcrea (a), CKD (b) and eGFRcys (c). The dotted line indicates the genome-wide significance threshold at $P = 5 \times 10^{-8}$.

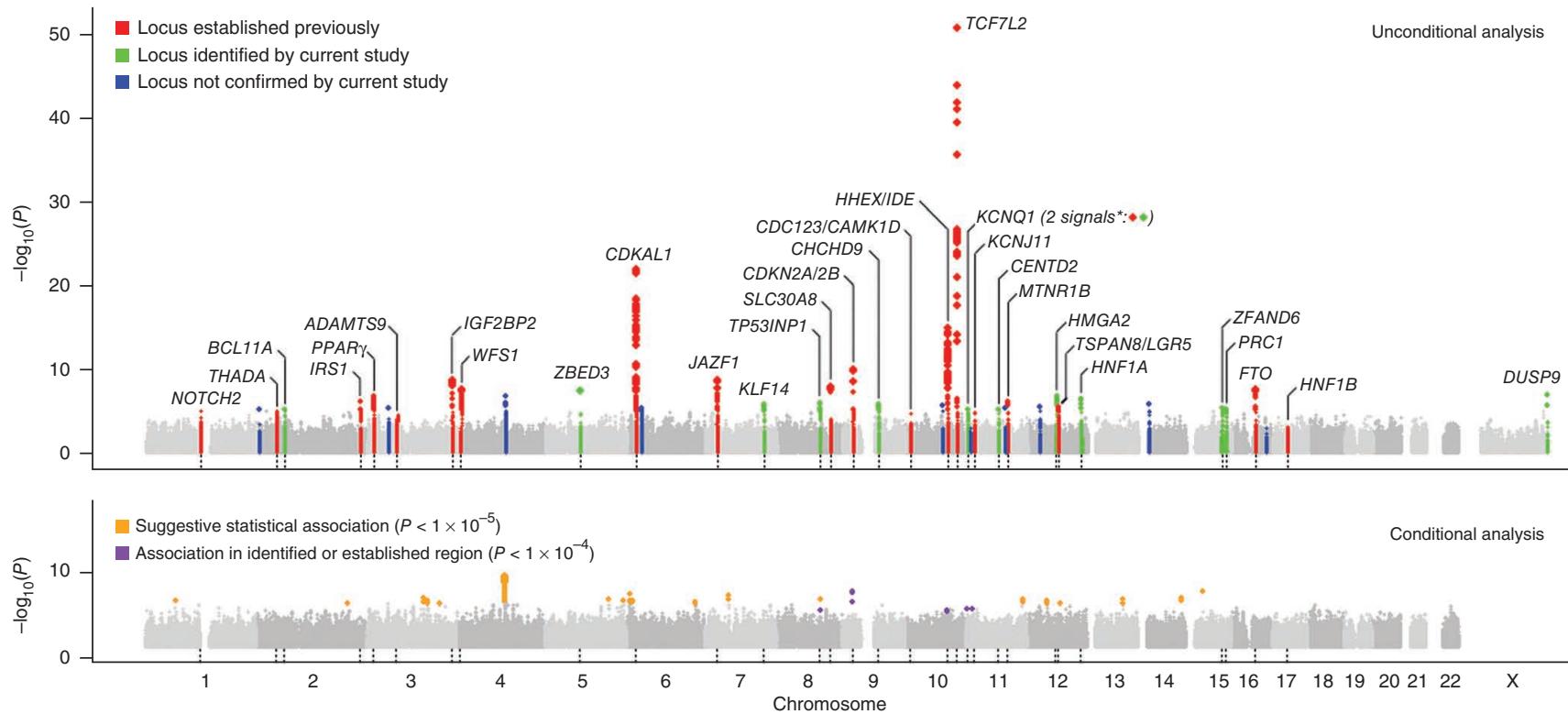
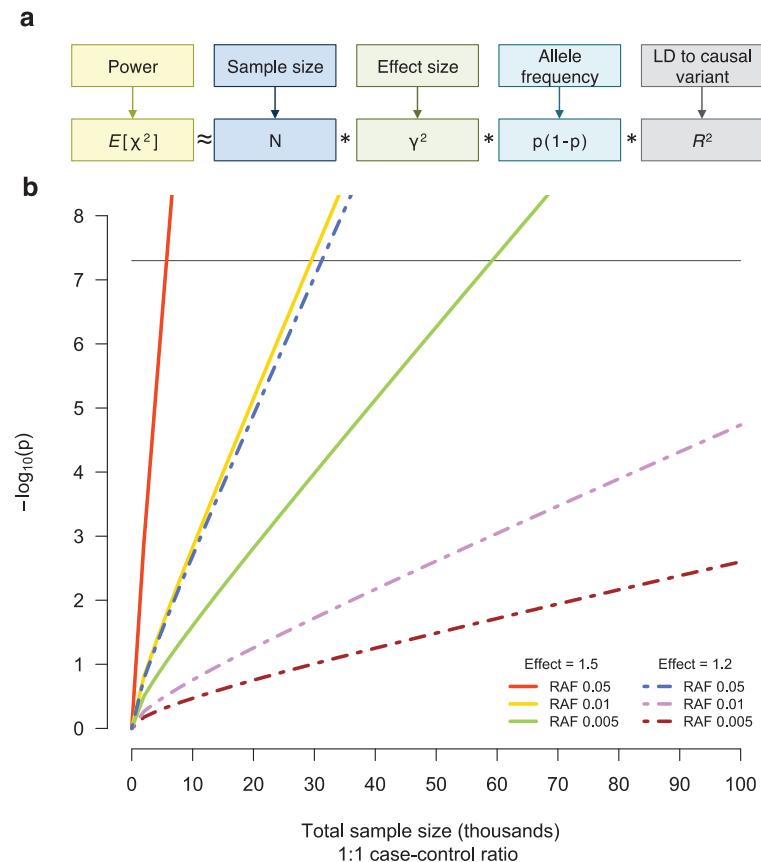
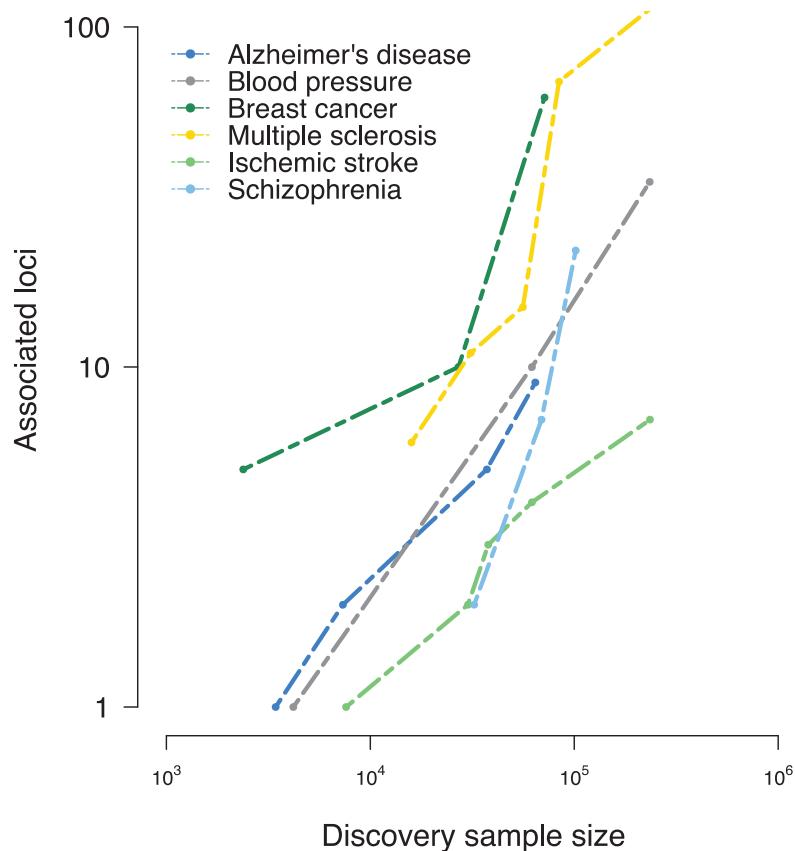
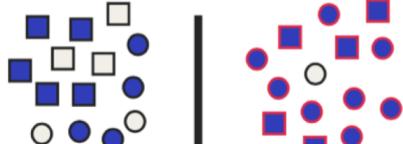
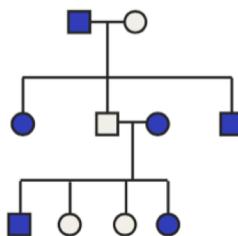
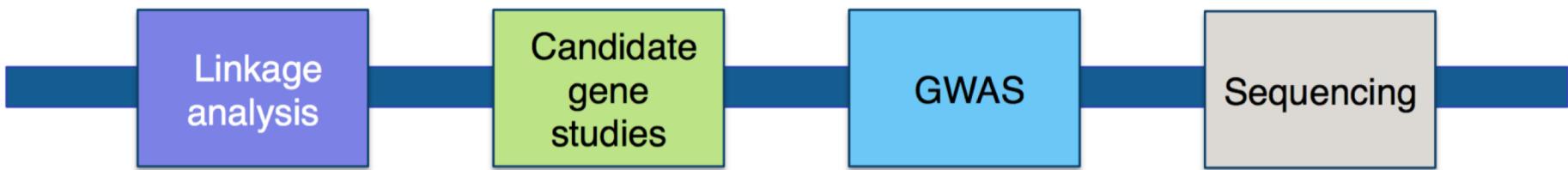
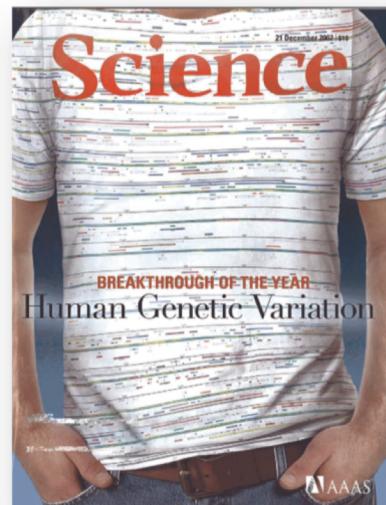


Figure 1 Genome-wide Manhattan plots for the DIAGRAM+ stage 1 meta-analysis. Top panel summarizes the results of the unconditional meta-analysis. Previously established loci are denoted in red and loci identified by the current study are denoted in green. The ten signals in blue are those taken forward but not confirmed in stage 2 analyses. The genes used to name signals have been chosen on the basis of proximity to the index SNP and should not be presumed to indicate causality. The lower panel summarizes the results of equivalent meta-analysis after conditioning on 30 previously established and newly identified autosomal T2D-associated SNPs (denoted by the dotted lines below these loci in the upper panel). Newly discovered conditional signals (outside established loci) are denoted with an orange dot if they show suggestive levels of significance ($P < 10^{-5}$), whereas secondary signals close to already confirmed T2D loci are shown in purple ($P < 10^{-4}$).

Power, Effect size, Sample size...

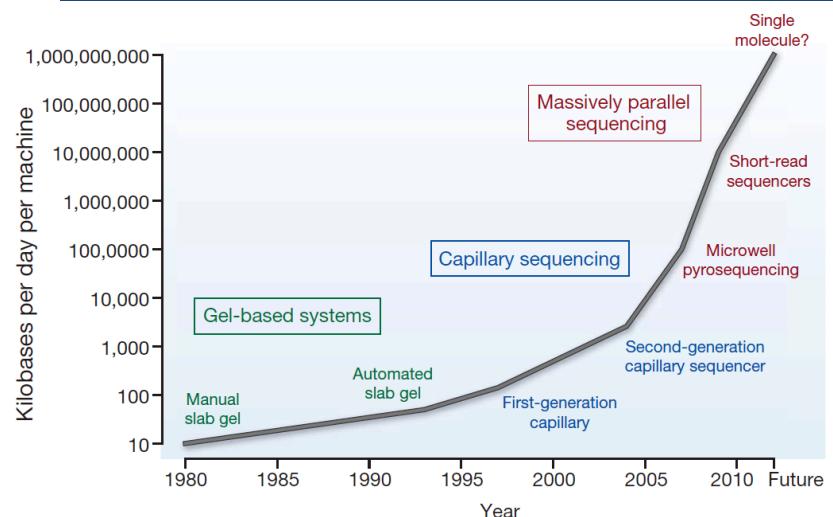
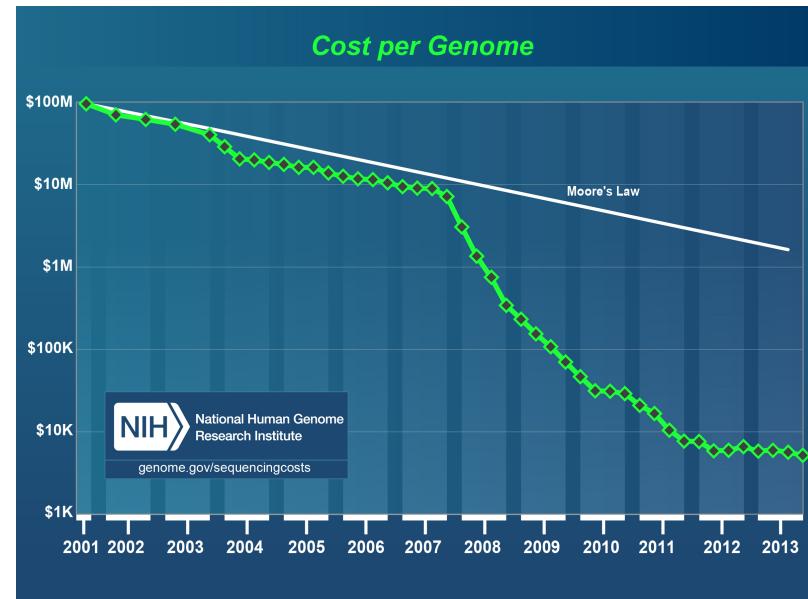




Next-generation sequencing

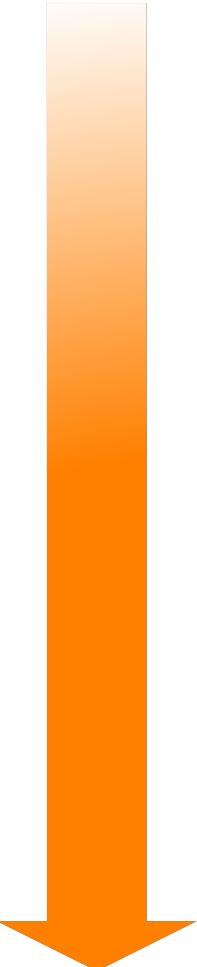
Milestone: \$1000 dollar genome
(2014, Illumina HiSeq X Ten Sequencer)

But how much money needs to be spent on annotation and (even more important) interpretation of the results?

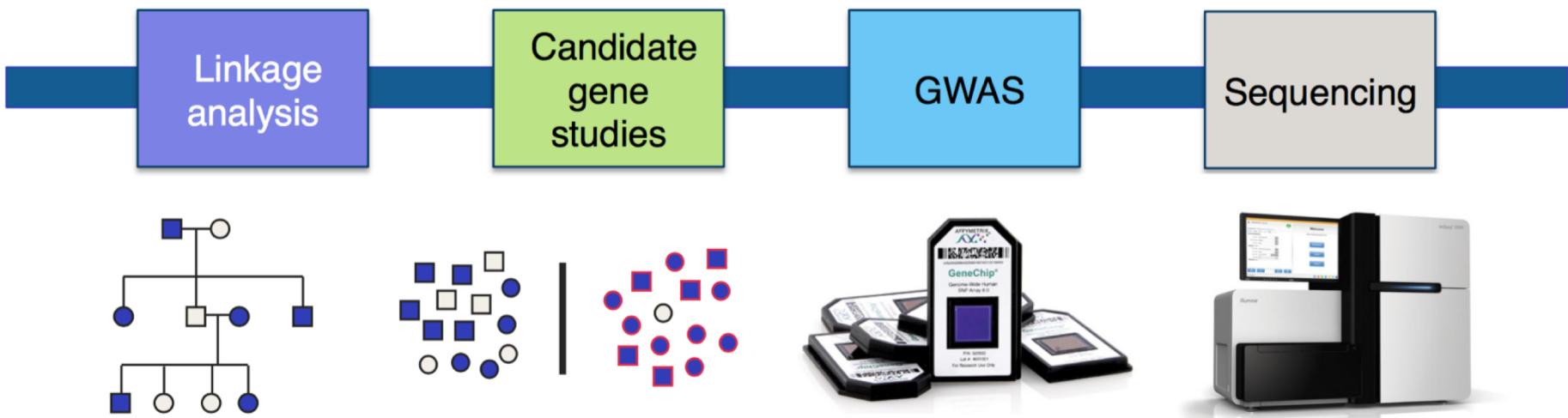


Summary: what's been (being) done?

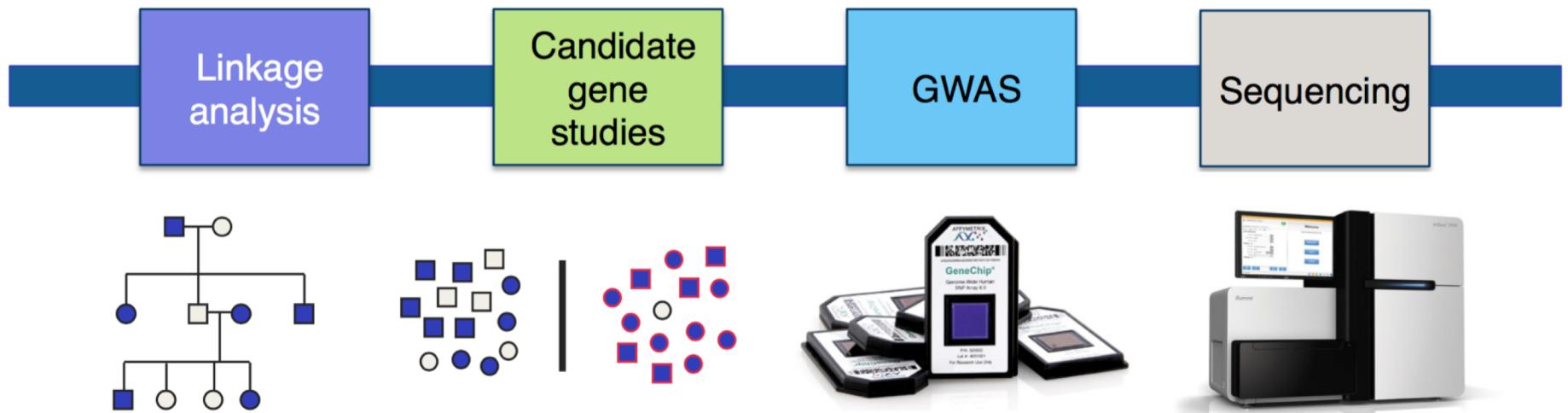
- Family-based linkage studies
 - Rare, Mendelian traits
- Candidate gene association studies
 - Many claims, few robust findings
 - Terrible track record in terms of reproducibility
- Genome-wide association studies (GWAS)
 - Complex traits and common diseases
- Whole-exome sequencing studies
 - Rare, Mendelian diseases (unsolved cases)
 - Complex traits and common diseases
- Whole-genome sequencing studies



What have we learned in the field of cardiovascular genetics?



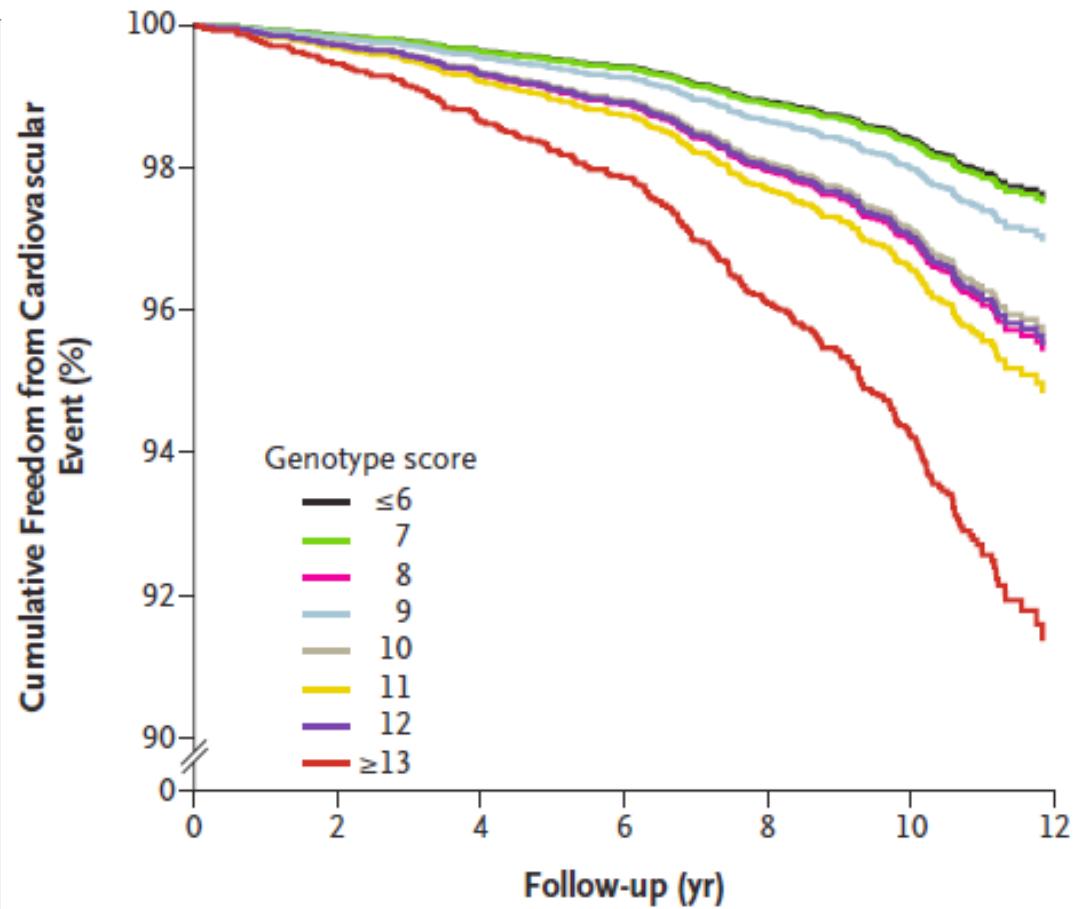
Prediction?



9 SNPs associations with CVD

Table 4. Multivariable Analysis of the Association between Genotype Score and the Time to First Cardiovascular Event.*

	Multivariable-Adjusted Hazard Ratio (95% CI)	P Value
Age, per SD	1.77 (1.52–2.07)	<0.001
Male sex	1.61 (1.20–2.17)	0.002
Parent or sibling with history of myocardial infarction	1.52 (1.17–1.97)	0.002
Cholesterol, per SD		
LDL	1.13 (0.99–1.29)	0.08
HDL	0.75 (0.61–0.91)	0.003
Log triglycerides, per SD	0.87 (0.73–1.04)	0.12
Blood pressure, per SD		
Systolic	1.29 (1.08–1.54)	0.005
Diastolic	1.16 (0.97–1.38)	0.11
Body-mass index, per SD	1.09 (0.94–1.25)	0.26
Diabetes mellitus	1.47 (1.02–2.13)	0.04
Status of cigarette smoking		<0.001†
Former versus never	1.17 (0.85–1.59)	
Current versus never	2.00 (1.41–2.83)	
Log C-reactive protein, per SD	1.14 (0.99–1.30)	0.06
Drug therapy		
Lipid-lowering	1.29 (0.63–2.64)	0.48
Antihypertensive	1.46 (1.08–1.97)	0.01
Genotype score, per single unfavorable allele	1.15 (1.07–1.24)	<0.001



But does not aid risk prediction

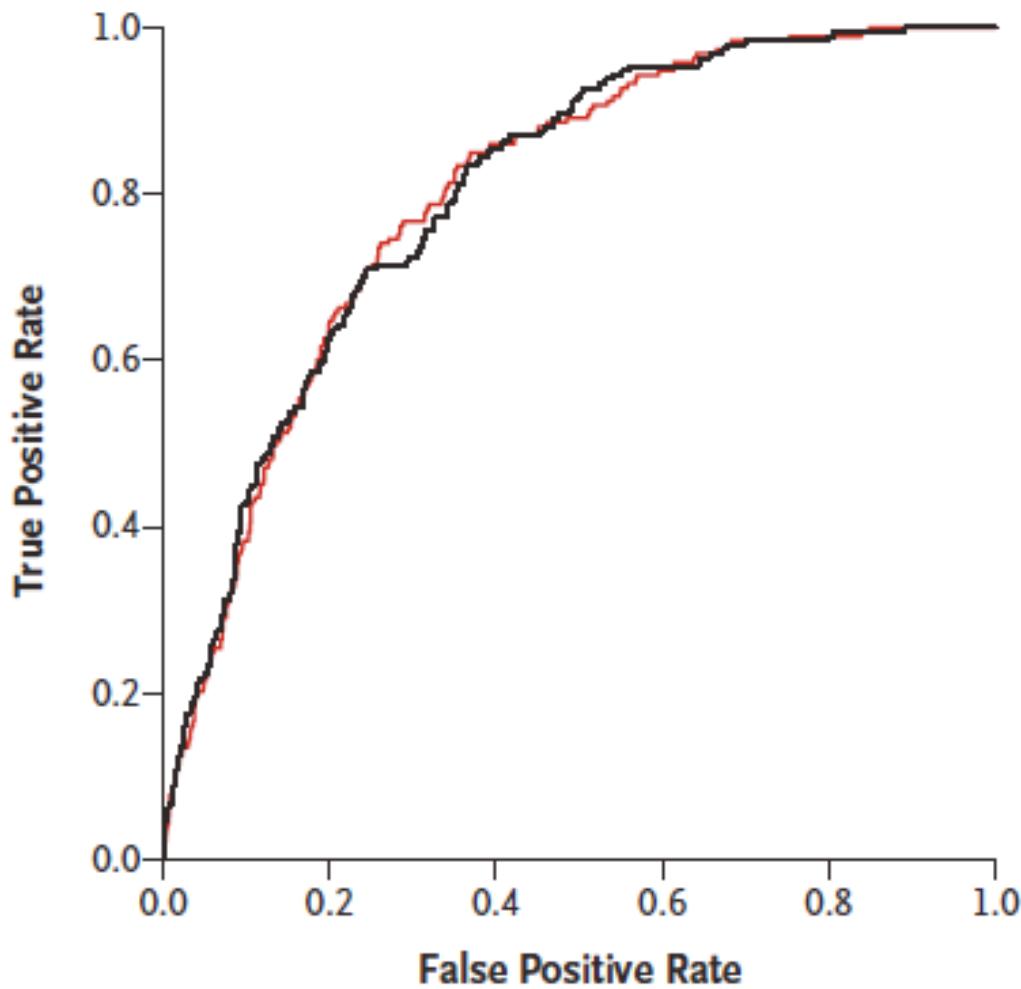


Figure 2. Receiver-Operating-Characteristic (ROC) Curves for Incident Myocardial Infarction, Ischemic Stroke, or Death from Coronary Heart Disease during 10-Year Follow-up.

The curves are based on risk-prediction models incorporating 14 clinical covariates that either included the genotype score (black line) or did not include the genotype score (red line). The C statistic (area under the ROC curve) for total cardiovascular events was the same (0.80) for both risk models.

MENU ▾

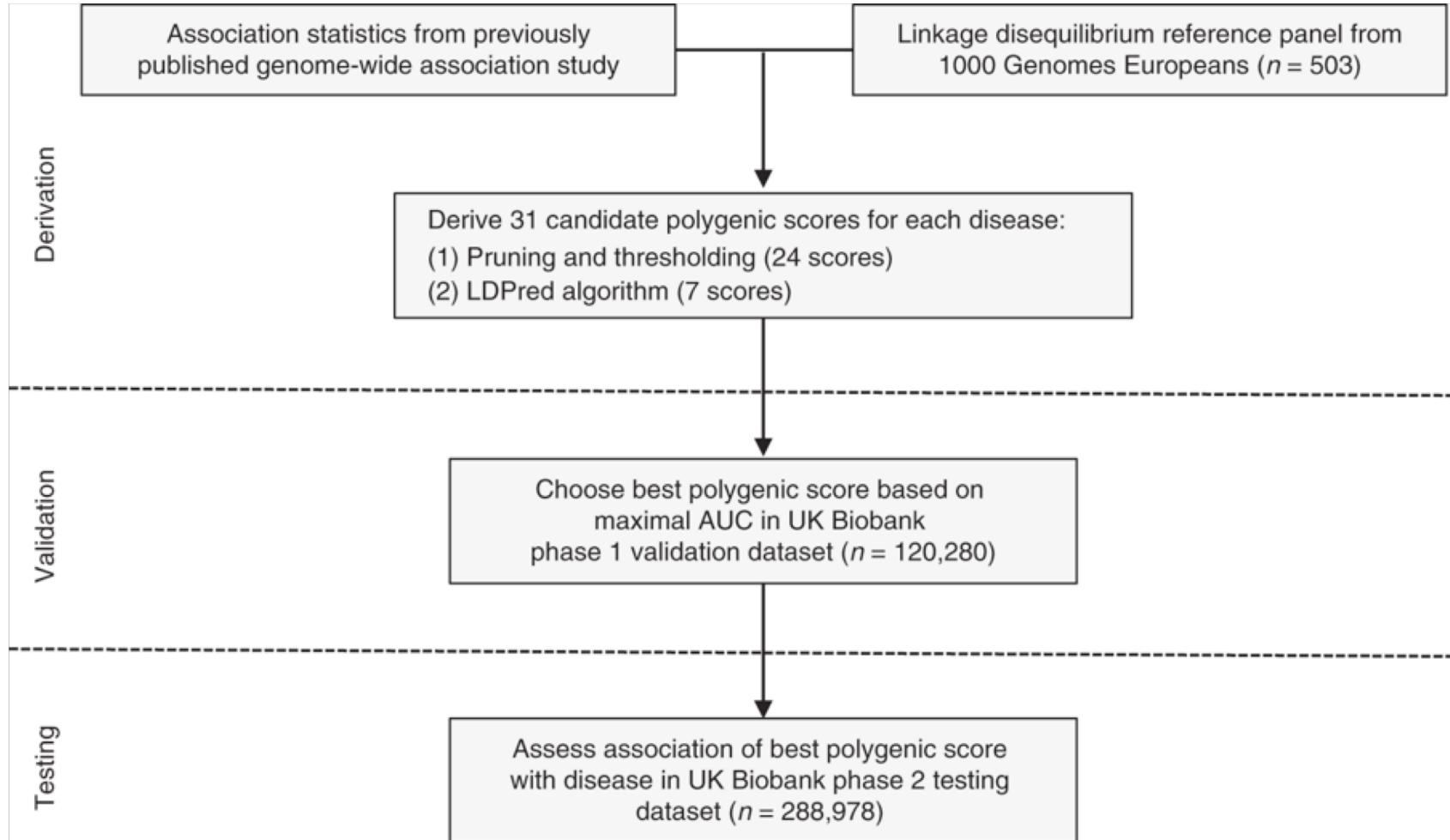
Letter | Published: 13 August 2018

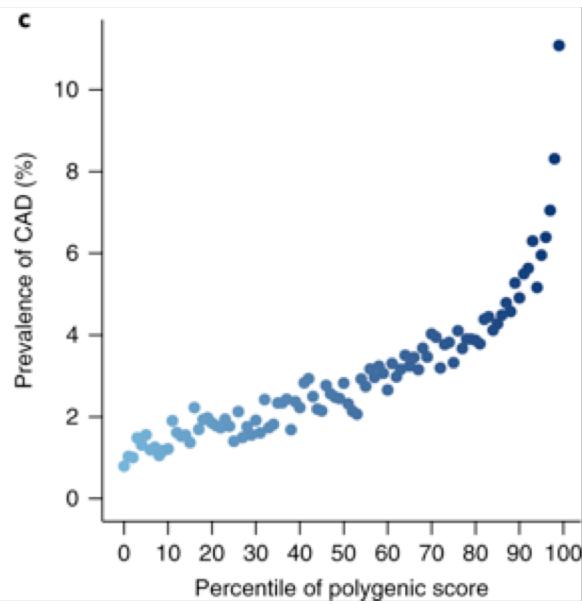
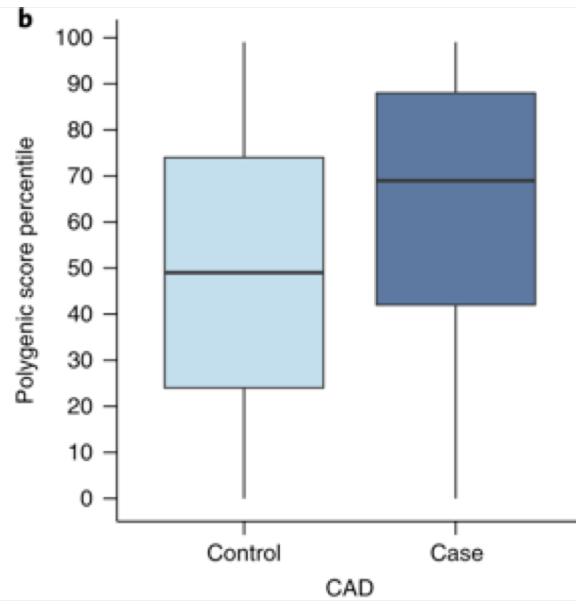
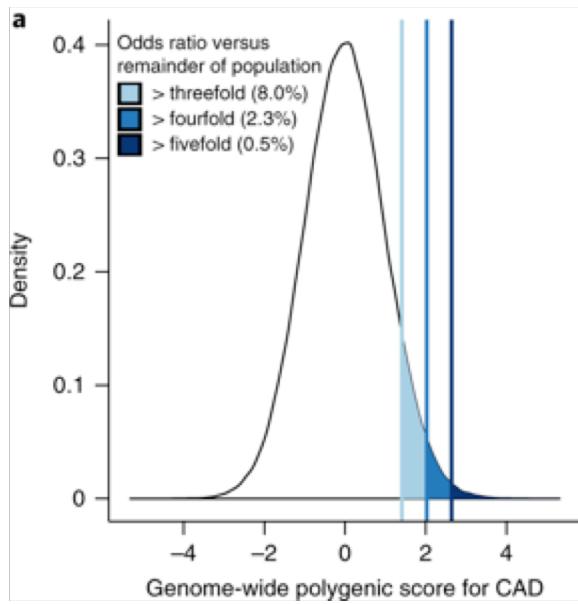
Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

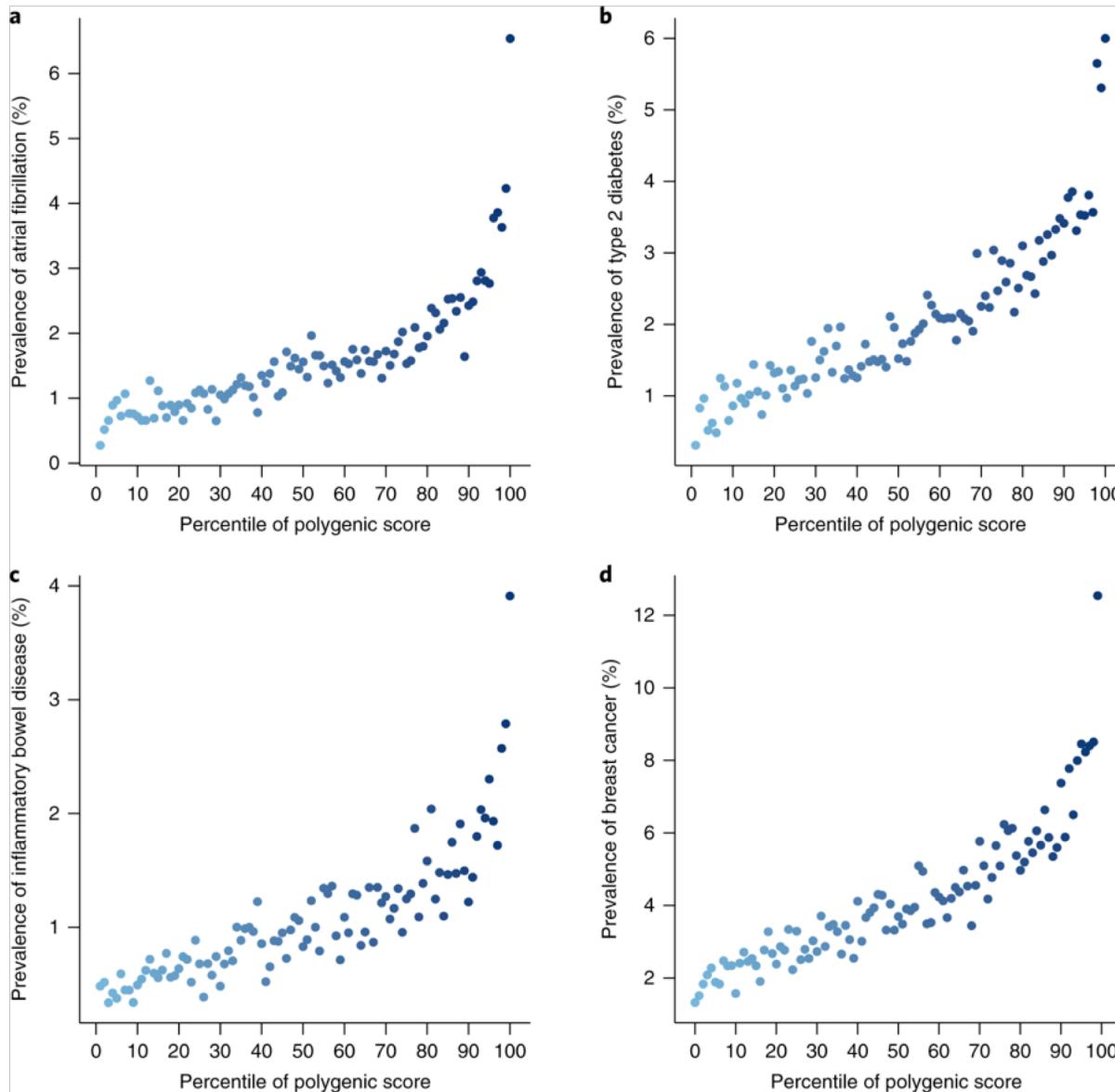
Amit V. Khera, Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S. Lander, Steven A. Lubitz, Patrick T. Ellinor & Sekar Kathiresan 

Nature Genetics **50**, 1219–1224 (2018) | Download Citation 

Include millions of variants of small effects







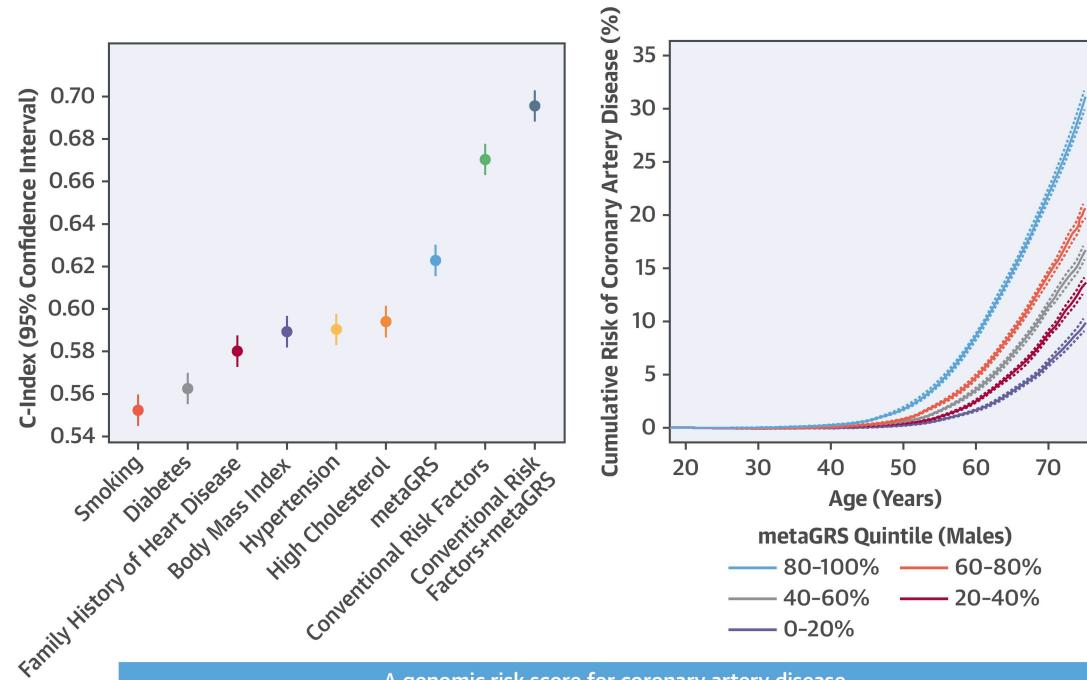
High GPS definition	Individuals in testing dataset (n)	% of individuals
Odds ratio ≥3.0		
CAD	23,119/288,978	8.0
Atrial fibrillation	17,627/288,978	6.1
Type 2 diabetes	10,099/288,978	3.5
Inflammatory bowel disease	9,209/288,978	3.2
Breast cancer	2,369/157,895	1.5
Any of the five diseases	57,115/288,978	19.8
Odds ratio ≥4.0		
CAD	6,631/288,978	2.3
Atrial fibrillation	4,335/288,978	1.5
Type 2 diabetes	578/288,978	0.2
Inflammatory bowel disease	2,297/288,978	0.8
Breast cancer	474/157,895	0.3
Any of the five diseases	14,029/288,978	4.9
Odds ratio ≥5.0		
CAD	1,443/288,978	0.5
Atrial fibrillation	2,020/288,978	0.7
Type 2 diabetes	144/288,978	0.05
Inflammatory bowel disease	571/288,978	0.2
Breast cancer	158/157,895	0.1
Any of the five diseases	4,305/288,978	1.5

Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults

Implications for Primary Prevention

Michael Inouye, Gad Abraham, Christopher P. Nelson, Angela M. Wood, Michael J. Sweeting, Frank Dudbridge, Florence Y. Lai, Stephen Kaptoge, Marta Brozynska, Tingting Wang, Shu Ye, Thomas R. Webb, Martin K. Rutter, Ioanna Tzoulaki, Riyaz S. Patel, Ruth J.F. Loos, Bernard Keavney, Harry Hemingway, John Thompson, Hugh Watkins, Panos Deloukas, Emanuele Di Angelantonio, Adam S. Butterworth, John Danesh, Nilesh J. Samani and for the UK Biobank CardioMetabolic Consortium CHD Working Group

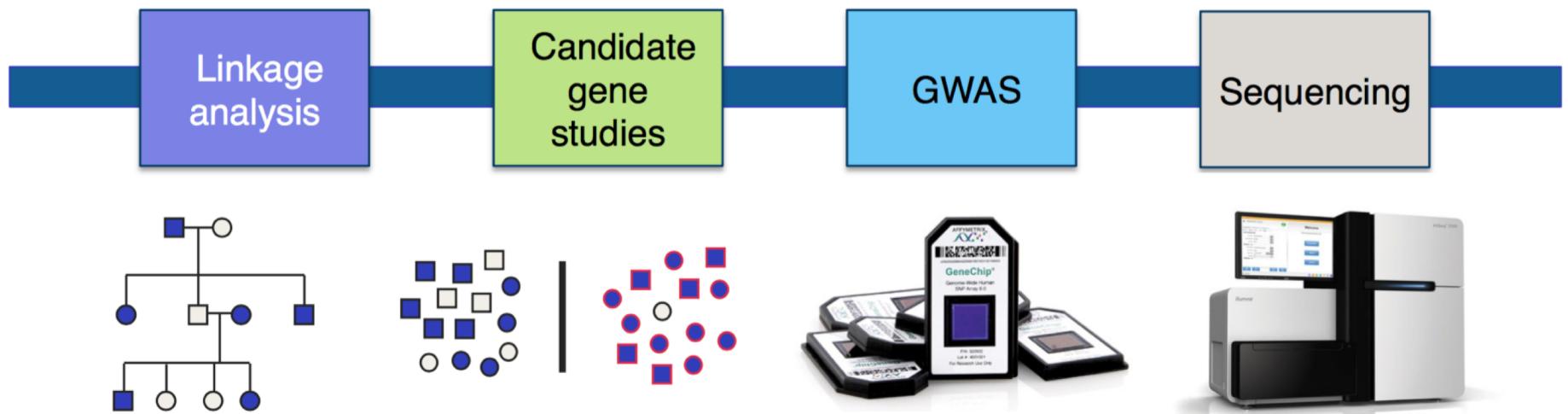
CENTRAL ILLUSTRATION: Genomic Risk Score for Coronary Artery Disease



A genomic risk score for coronary artery disease
Greater association with future coronary artery disease than any single conventional risk factor
Independent of yet complements conventional risk factors
Provides meaningful lifetime risk estimates of coronary artery disease
Quantifiable at or before birth and shows potential for risk screening in early life

Inouye, M. et al. J Am Coll Cardiol. 2018;72(16):1883-93.

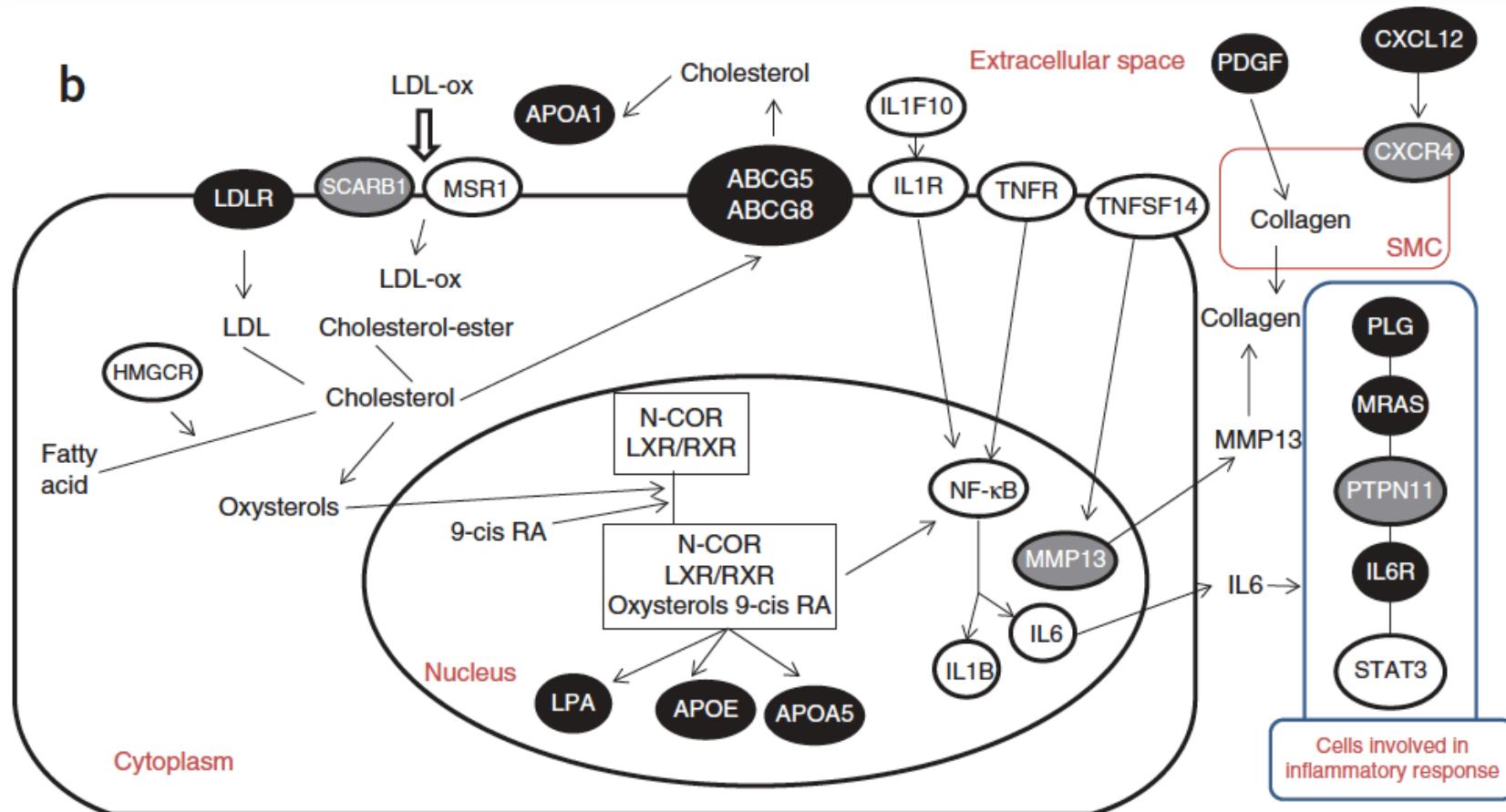
Biological mechanisms?



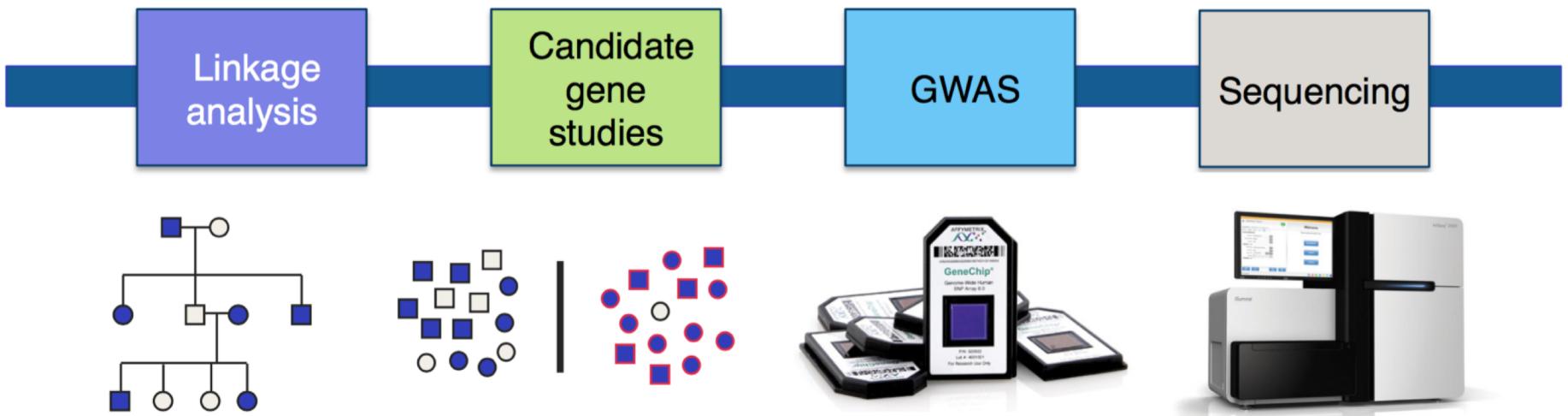
Large-scale association analysis identifies new risk loci for coronary artery disease

63,746 cases and 130,681 controls: 46 CAD loci

The CARDIoGRAMplusC4D Consortium¹



Causality: Mendelian Randomization?



APOLIPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR,—It is unclear whether the relation between low serum cholesterol levels and cancer¹ is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study cohorts with widely different diets and corresponding differences in mean cholesterol levels experienced similar mean cancer rates.^{2,3} On the other hand, within each region cancer incidence was higher in men with a serum cholesterol in the lowest part of the cholesterol distribution for that country.³ Thus, naturally low cholesterol levels are sometimes associated with increased cancer risk.^{1,3}

Differences in the aminoacid sequence of apolipoprotein E (apo E) are major determinants of differences in plasma cholesterol levels within a population. Apo E has a key role in the clearance of cholesterol from plasma.⁴ The synthesis of apo E is under the control of three independent alleles, located at a single gene locus, coding for the major isoforms E-2, E-3, and E-4 with respective population frequencies of about 8, 77, and 15%.⁵ The homozygous E-3/E-3 is the most common phenotype encountered and E-2/E-2 is the least common. From apo E-2 to apo E-3, one cysteine residue is replaced by arginine, and from apo E-3 to apo E-4 another cysteine residue is replaced. As a result the avidity of apo E containing lipoproteins for lipoprotein receptors increases from apo E-2 to apo E-3 to apo E-4. In several populations,⁶⁻⁸ including the Finns and the Japanese (Dr G. Utermann, personal communication), the gradient in serum cholesterol levels in the population is associated with a gradient in apo E phenotype, E-2 being associated with lower serum low-density lipoprotein and total cholesterol levels than E-3 and E-4. Thus, if a naturally low cholesterol favours tumour

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

Unlike most other indices of lipid metabolism, apolipoprotein aminoacid sequences are not disturbed by disease, and the apo E phenotype found in a patient will have been present since birth. A comparison of apo E phenotypes in cancer patients with those in matched controls might thus shed light on the relation between low cholesterol and cancer. If it is causal then the E-2 allele should be more common among patients and E-3 and E-4 more common among controls. On the other hand, equal distribution of apo E phenotypes among cases and controls would suggest that the association between low cholesterol and cancer is spurious. Measurement of apo E phenotype by isoelectric focusing of plasma is a routine determination in lipid laboratories; epidemiologists interested in cholesterol and cancer should include it in their studies.

Department of Human Nutrition,
Agricultural University,
6703 BC Wageningen, Netherlands

MARTIJN B. KATAN

1. McMichael AJ, Jensen OM, Parkin DM, Zaridze DG. Dietary and endogenous cholesterol and human cancer. *Epidemiol Rev* 1984; **6**: 192-216.
2. Keys A, Aravanis C, Blackburn H, et al. Serum cholesterol and cancer mortality in the seven countries study. *Am J Epidemiol* 1985; **121**: 870-83.
3. Katan MB. Effects of cholesterol-lowering diets on the risk for cancer and other non-cardiovascular diseases. In: Nestel PJ, et al, eds. Atherosclerosis VII: Proceedings of the Seventh International Atherosclerosis Symposium. Amsterdam: Elsevier, 1986.
4. Brown MS, Kovanen PT, Goldstein JL. Regulation of plasma cholesterol by lipoprotein receptors. *Science* 1981; **212**: 628-35.
5. Utermann G, Steinmetz A, Weber W. Genetic control of human apolipoprotein E polymorphism: comparison of one- and two-dimensional techniques of isoprotein analysis. *Hum Genet* 1982; **60**: 344-51.
6. Utermann G, Kindermann I, Kaffarnik H, Steinmetz A. Apolipoprotein E phenotypes and hyperlipidemia. *Hum Genet* 1984; **65**: 232-36.
7. Robertson FW, Cumming AM. Effects of Apoprotein E Polymorphism on Serum Lipoprotein Concentration. *Arteriosclerosis* 1985; **5**: 283-92.
8. Utermann G. Genetic polymorphism of apolipoprotein E: impact on plasma lipoprotein metabolism. In: Crepaldi G, et al, eds. Diabetes, obesity and hyperlipidemias III. Amsterdam: Elsevier, 1985: 1-28.

APOLIPOPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR.—It is unclear whether the relation between low serum cholesterol levels and cancer¹ is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study cohorts with widely different diets and corresponding differences in mean cholesterol levels experienced similar mean cancer rates.^{2,3} On the other hand, within each region cancer incidence is higher among patients with naturally low cholesterol levels than among those with high cholesterol levels.^{1,3}

Differences in the aminoacid sequence of apolipoprotein E (apo E) are major determinants of differences in plasma cholesterol levels within a population. Apo E has a key role in the clearance of cholesterol from

control of three genes coding for the protein. The apo E-2 allele is the least common. From apo E-2 to apo E-3, one cysteine residue is replaced by arginine, and from apo E-3 to apo E-4 another cysteine residue is replaced. As a result the avidity of apo E containing lipoproteins for lipoprotein receptors increases from apo E-2 to apo E-3 to apo E-4. In several populations,⁶⁻⁸ including the Finns and the Japanese (Dr G. Utermann, personal communication), the gradient in serum cholesterol levels in the population is associated with a gradient in apo E phenotype, E-2 being associated with lower serum low-density lipoprotein and total cholesterol levels than E-3 and E-4. Thus, if a naturally low cholesterol favours tumour

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

Unlike most other indices of lipid metabolism, apolipoprotein E aminoacid sequences are not disturbed by disease, and the apo E phenotype found in a patient will have been present since birth. A comparison of apo E phenotypes in cancer patients with those in matched controls might thus shed light on the relation between low cholesterol and cancer. If it is causal then the E-2 allele should be more common among patients and E-3 and E-4 more common among controls. On the other hand, equal distribution of apo E phenotypes among cases and controls would suggest that the association between low cholesterol and cancer is spurious.

Epidemiologists interested in cholesterol and cancer should include it in their studies.

Department of Human Nutrition,
Agricultural University,
6703 BC Wageningen, Netherlands

MARTIJN B. KATAN

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

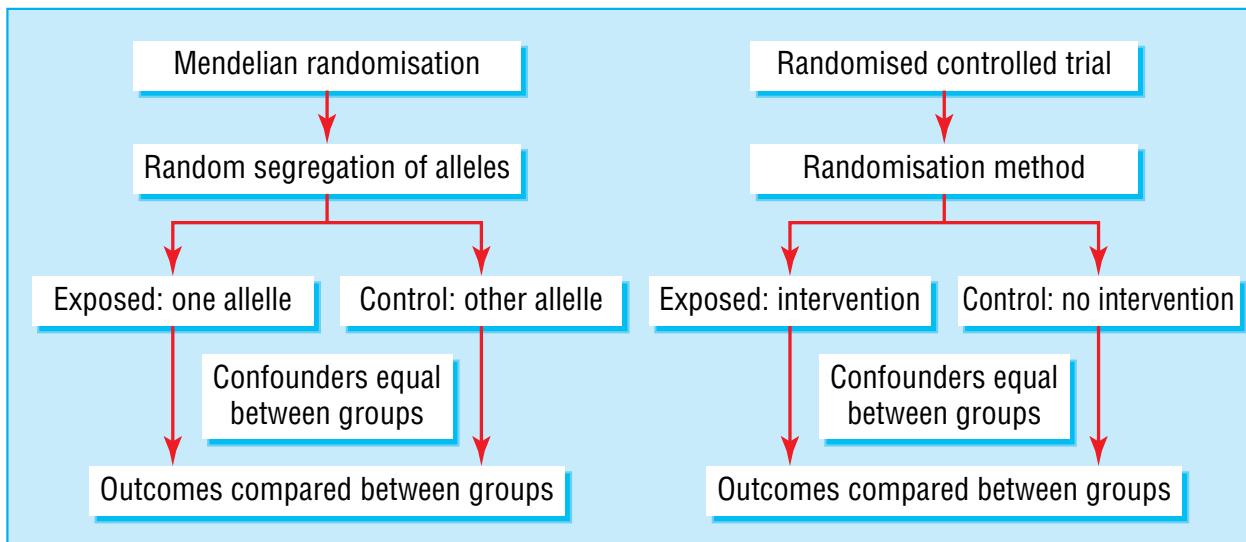
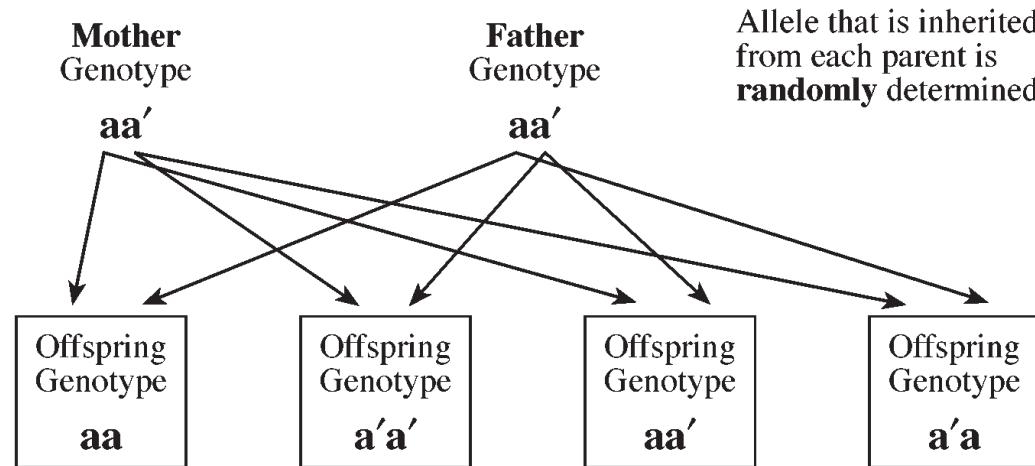
cardiovascular diseases. In: Nestel PJ, et al, eds. *Atherosclerosis VII: Proceedings of the Seventh International Atherosclerosis Symposium*. Amsterdam: Elsevier, 1986.

4. Brown MS, Kovanen PT, Goldstein JL. Regulation of plasma cholesterol by lipoprotein receptors. *Science* 1981; **212**: 628-35.
5. Utermann G, Steinmetz A, Weber W. Genetic control of human apolipoprotein E polymorphism: comparison of one- and two-dimensional techniques of isoprotein analysis. *Hum Genet* 1982; **60**: 344-51.
6. Utermann G, Kindermann I, Kaffarnik H, Steinmetz A. Apolipoprotein E phenotypes and hyperlipidemia. *Hum Genet* 1984; **65**: 232-36.
7. Robertson FW, Cumming AM. Effects of Apoprotein E Polymorphism on Serum Lipoprotein Concentration. *Arteriosclerosis* 1985; **5**: 283-92.
8. Utermann G. Genetic polymorphism of apolipoprotein E: impact on plasma lipoprotein metabolism. In: Crepaldi G, et al, eds. *Diabetes, obesity and hyperlipidemias III*. Amsterdam: Elsevier, 1985: 1-28.

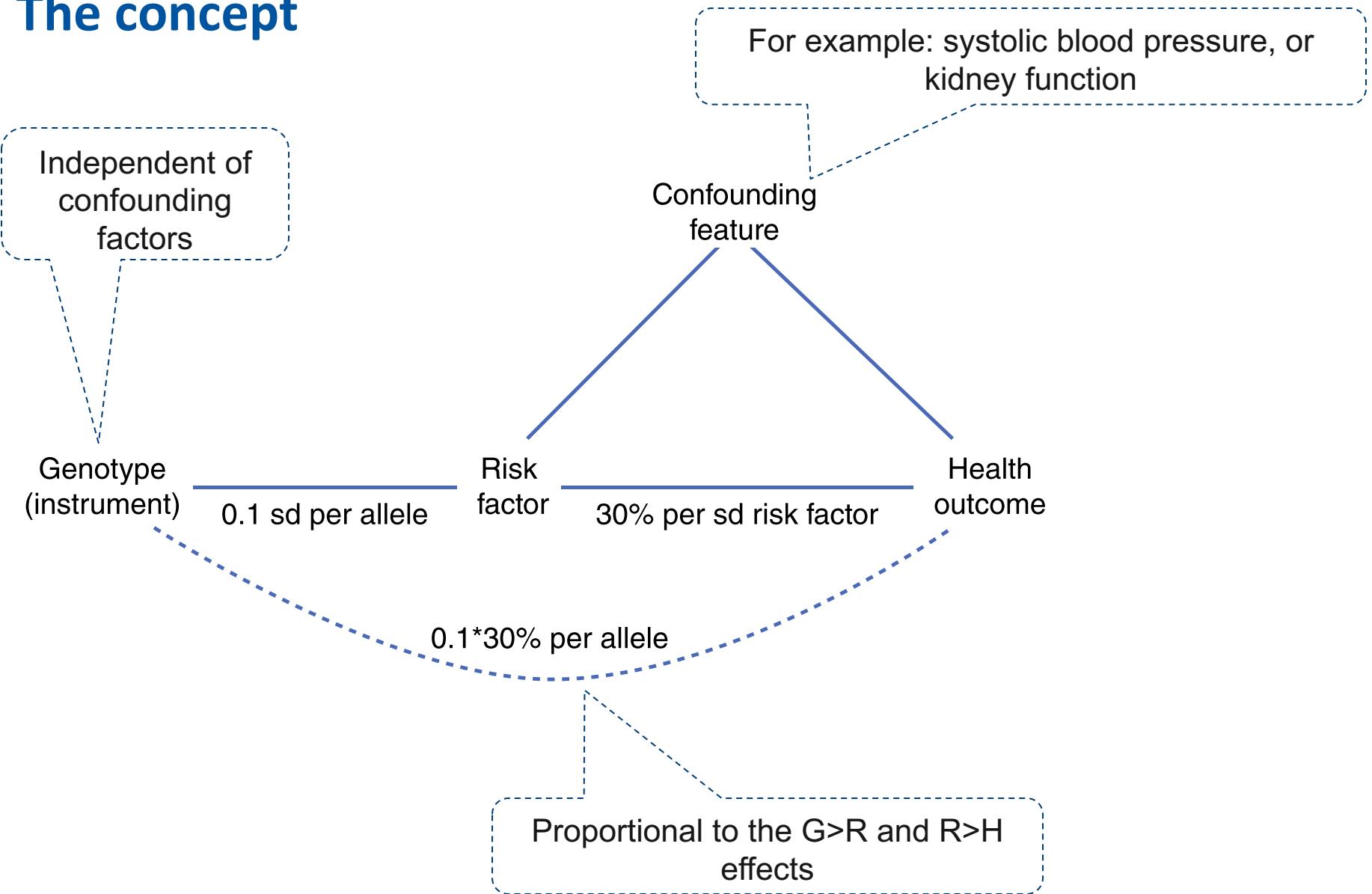


Gregor Johann Mendel - 20 July 1822 – 6 January 1884

Alleles are randomly assigned: a “natural” treatment in a “natural” randomized trial



The concept



Hypothesis: Inhibition of CETP might lead to raised HDL

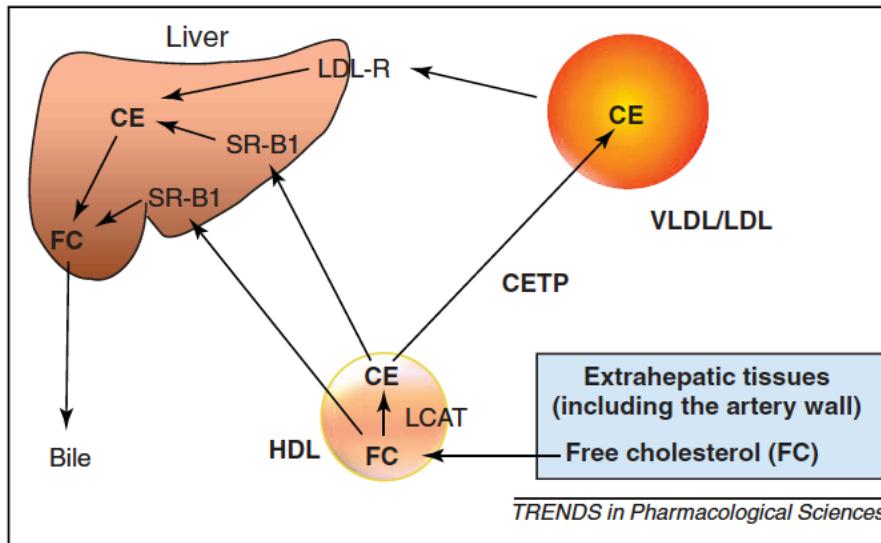


Figure 1. The role of CETP in plasma cholesterol transport. Cells in extra-hepatic tissues eliminate any cholesterol that is surplus to their needs by transferring it as free (unesterified) cholesterol (FC) to HDLs in the extracellular space. The FC in HDLs is then either delivered to the liver in a process dependent on hepatic scavenger receptor (SR)-B1 or converted into cholesteryl esters (CE) by lecithin cholesterol acyltransferase (LCAT). The CE formed in HDLs is subsequently transported to the liver by either of two pathways: a direct pathway mediated by SR-B1 and an indirect pathway in which HDL CE is first transferred to the VLDL/LDL fraction by CETP and then taken up by the liver following binding of LDL to hepatic LDL receptors (LDL-R).

CETP inhibitors

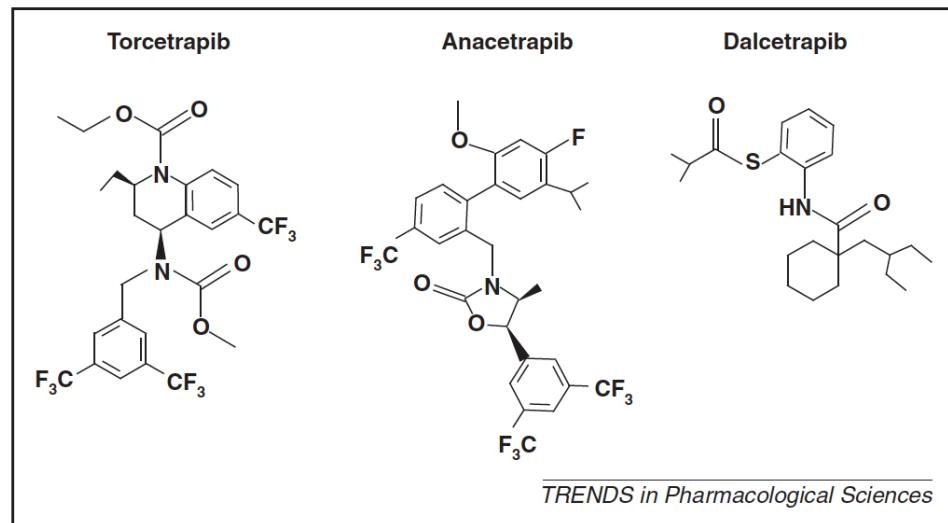


Figure 2. Structures of torcetrapib, anacetrapib and dalcetrapib.

- Torcetrapib – ILLUMINATE halted due to safety issues
- Dalcetrapib (raises HDL, does not lower LDL) stopped by Roche (no clinical benefit)
- Anacetrapib (raises HDL, lowers LDL) – DEFINE
- Evacetrapib (raises HDL, lowers LDL)

The NEW ENGLAND JOURNAL *of* MEDICINE

ESTABLISHED IN 1812

NOVEMBER 22, 2007

VOL. 357 NO. 21

Effects of Torcetrapib in Patients at High Risk for Coronary Events

Philip J. Barter, M.D., Ph.D., Mark Caulfield, M.D., M.B., B.S., Mats Eriksson, M.D., Ph.D.,
Scott M. Grundy, M.D., Ph.D., John J.P. Kastelein, M.D., Ph.D., Michel Komajda, M.D., Jose Lopez-Sendon, M.D., Ph.D.,
Lori Mosca, M.D., M.P.H., Ph.D., Jean-Claude Tardif, M.D., David D. Waters, M.D., Charles L. Shear, Dr.P.H.,
James H. Revkin, M.D., Kevin A. Buhr, Ph.D., Marian R. Fisher, Ph.D., Alan R. Tall, M.B., B.S.,
and Bryan Brewer, M.D., Ph.D., for the ILLUMINATE Investigators*

After 12 months, significant HDL increase (72%) and
LDL decrease (25%), but increased risk for CVD
(HR=1.3) and death (HR=1.6)

The New York Times

TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JOBS | REAL ESTATE | AUTOS

piwdebakker ▾ | Help

Health | All NYT

Search

Health

FITNESS & NUTRITION | HEALTH CARE POLICY | MENTAL HEALTH & BEHAVIOR

99 · 99 TRACKS ARTISTS THE 99%

CLICK HERE TO DOWNLOAD

Pfizer Ends Studies on Drug for Heart Disease

By ALEX BERENSON
Published: December 3, 2006

Pfizer announced last night that it had discontinued research on its most important experimental drug, a treatment for [heart disease](#). The decision is a stunning development that is likely to seriously damage the company's prospects through the next decades.

Preliminary research found that the drug, torcetrapib, appeared to be linked with deaths and heart problems in the patients who were taking it. For people with heart disease, Pfizer's decision to stop the trial represents the failure of a drug that many cardiologists had viewed as a potentially major advance in efforts to reduce heart attacks and strokes.

Torcetrapib is designed to raise levels of so-called good [cholesterol](#). It was to be used in combination with older drugs called statins, like Lipitor and Zocor, which reduce so-called bad cholesterol.

As recently as Thursday, Pfizer executives had hailed the drug at a meeting with investors and analysts at the company's research center in Groton, Conn. "This will be one of the most important compounds of our generation," said Jeffrey B. Kindler, Pfizer's chief executive.

Pfizer is the world's biggest drug company, with 106,000 employees and \$51 billion in sales in 2005.

In a news release issued yesterday, the company said that it would immediately halt clinical trials of the drug and end its development.

The decision was based on interim results from a 15,000-patient clinical trial. The trial, called Illuminate, was scheduled to be completed in 2009. Pfizer had hoped it would prove that the combination of the two drugs was significantly more likely to reduce heart attacks and strokes than Lipitor alone does.

Even before yesterday's announcement, some cardiologists had raised concerns about torcetrapib, noting that the drug raised [blood pressure](#) in many patients, a serious side

E-MAIL | PRINT | REPRINTS

LOLA VERSUS WATCH TRAILER

More Articles in Health »

TicketWatch: Theater Offers by E-Mail

Sign up for ticket offers from Broadway shows and other advertisers.
piwdebakker@mac.com | Sign Up | Change E-mail Address | Privacy Policy

THE VIRUS THAT CAUSES SHINGLES MAY ALREADY BE INSIDE YOU GET THE INSIDE STORY

MOST POPULAR - HEALTH

E-MAILED | BLOGGED | VIEWED

- Well: A Richer Life by Seeing the Glass Half Full
- Drugs Help Tailor Alcoholism Treatment
- Well: New Data on Harms of Prostate Cancer Screening
- Well: Really? Never Brush Your Teeth Immediately After a Meal
- Well: Sleep Apnea Tied to Increased Cancer Risk
- Recipes for Health: Collard Greens, Fast or Slow
- A Long View on Health Care: Think Like an Investor
- More Care Up Front For \$54 A Month
- Well: Some Fats May Harm the Brain More
- Well: The Doctor's Remedy: Biofeedback for Stress

Go to Complete List »

Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study

Benjamin F Voight*, Gina M Peloso*, Marju Orho-Melander, Ruth Frikke-Schmidt, Maja Barbalic, Majken K Jensen, George Hindy, Hilma Hålm, Eric L Ding, Toby Johnson, Heribert Schunkert, Nilesh J Samani, Robert Clarke, Jemma C Hopewell, John F Thompson, Mingyao Li, Gudmar Thorleifsson, Christopher Newton-Cheh, Kiran Musunuru, James P Pirruccello, Danish Saleheen, Li Chen, Alexandre F R Stewart, Arne Schillert, Unnur Thorsteinsdottir, Gudmundur Thorgeirsson, Sonia Anand, James C Engert, Thomas Morgan, John Spertus, Monika Stoll, Klaus Berger, Nicola Martinelli, Domenico Girelli, Pascal P McKeown, Christopher C Patterson, Stephen E Epstein, Joseph Devaney, Mary-Susan Burnett, Vincent Mooser, Samuli Ripatti, Ida Surakka, Markku S Nieminen, Juha Sinisalo, Marja-Liisa Lokki, Markus Perola, Aki Havulinna, Ulf de Faire, Bruna Gigante, Erik Ingelsson, Tanja Zeller, Philipp Wild, Paul IW de Bakker, Olaf H Klungel, Anke-Hilse Maitland-van der Zee, Bas J M Peters, Anthonius de Boer, Diederick E Grobbee, Pieter W Kamphuisen, Vera H M Deneer, Clara C Elbers, N Charlotte Onland-Moret, Marten H Hofker, Cisca Wijmenga, W M Monique Verschuren, Jolanda M A Boer, Yvonne T van der Schouw, Asif Rasheed, Philippe Frossard, Serkalem Demissie, Cristen Willer, Ron Do, Jose M Ordovas, Gonçalo R Abecasis, Michael Boehnke, Karen L Mohlke, Mark J Daly, Candace Guiducci, Noël P Burtt, Aarti Surti, Elena Gonzalez, Shaun Purcell, Stacey Gabriel, Jaume Marrugat, John Peden, Jeanette Erdmann, Patrick Diemert, Christina Willenborg, Inke R König, Marcus Fischer, Christian Hengstenberg, Andreas Ziegler, Ian Buyschaert, Diether Lambrechts, Frans Van de Werf, Keith A Fox, Nour Eddine El Mokhtari, Diana Rubin, Jürgen Schrezenmeir, Stefan Schreiber, Arne Schäfer, John Danesh, Stefan Blankenberg, Robert Roberts, Ruth McPherson, Hugh Watkins, Alistair S Hall, Kim Overvad, Eric Rimm, Eric Boerwinkle, Anne Tybjaerg-Hansen, L Adrienne Cupples, Muredach P Reilly, Olle Melander, Pier M Mannucci, Diego Ardissino, David Siscovick, Roberto Elosua, Kari Stefansson, Christopher J O'Donnell, Veikko Salomaa, Daniel J Rader, Leena Peltonen, Stephen M Schwartz, David Altshuler, Sekar Kathiresan

Asn396Ser in *LIPG* increases HDL-C, and does
not affect other relevant factors
(BP, T2D, BMI, CRP, LDL)

HDL increased in Ser carriers

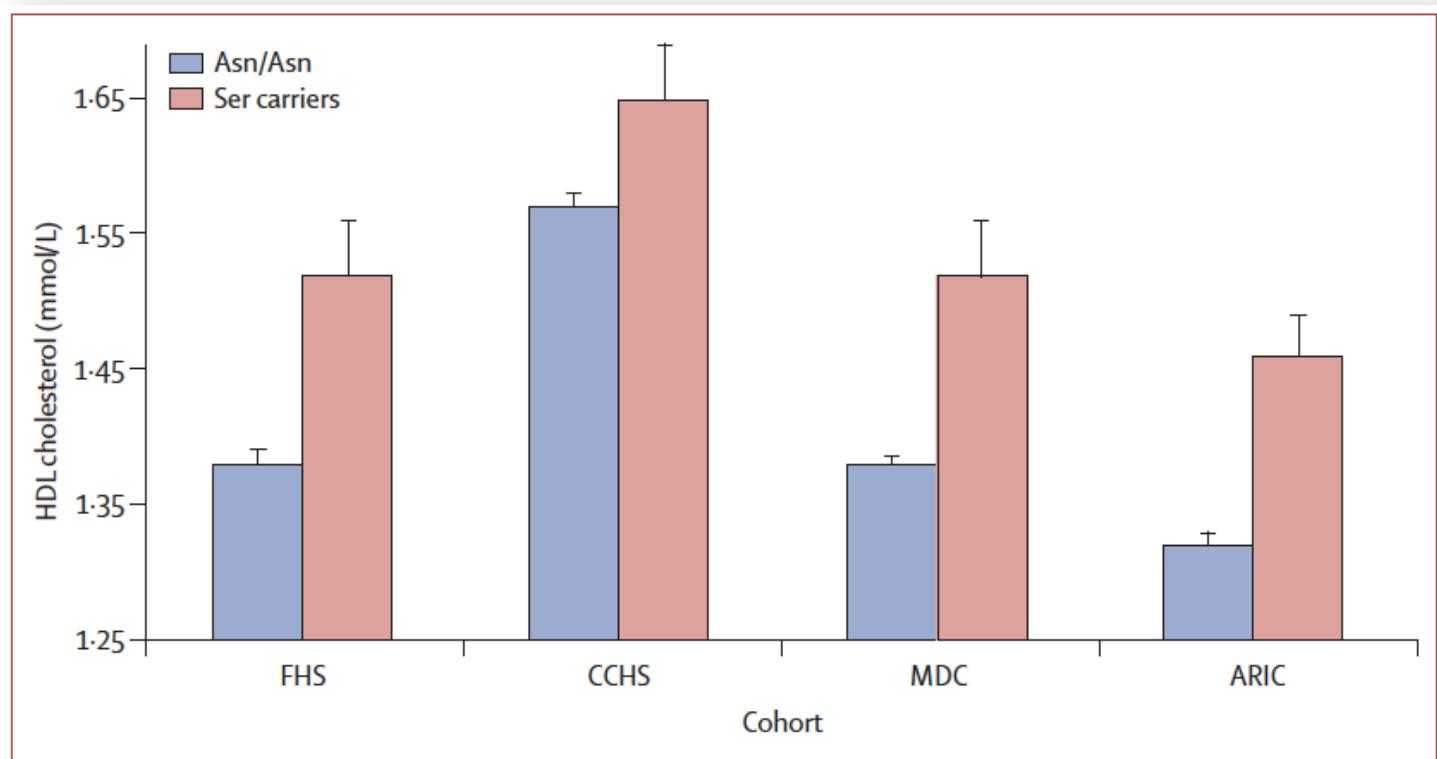


Figure 1: Plasma HDL cholesterol concentrations in carriers versus non-carriers of the Ser allele at the LIPG Asn396Ser polymorphism

Error bars show standard error. FHS=Framingham Heart Study. CCHS=Copenhagen City Heart Study. MDC=Malmo Diet and Cancer Study. ARIC=Atherosclerosis Risk in Communities Study.

But no protection against MI

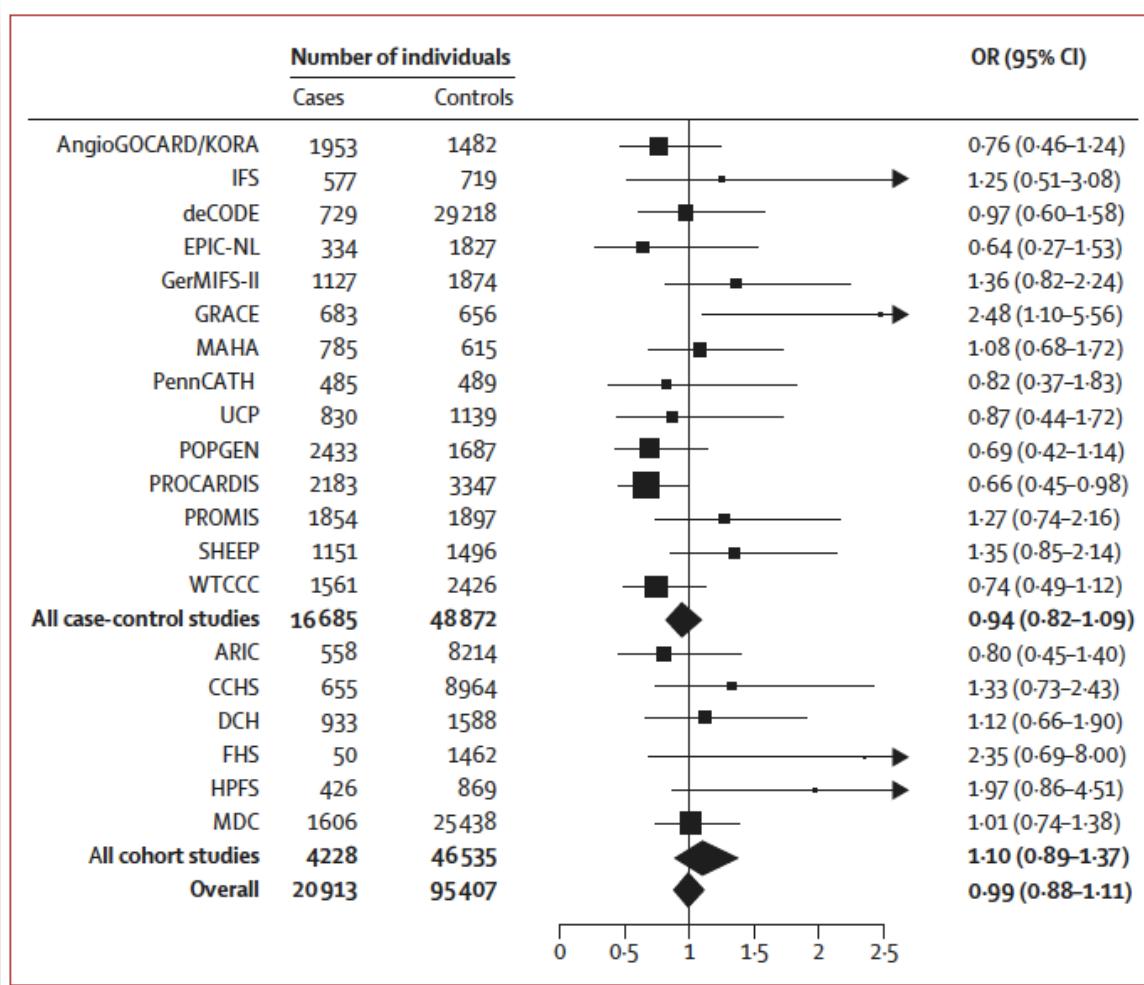


Figure 2: Association of LIPG Asn396Ser with myocardial infarction in 116 320 participants from 20 studies
In each study, the HDL-cholesterol-raising serine allele was modelled.

HDL has no impact on myocardial infarction risk

Odds ratio (95% CI) per SD increase in plasma lipid based on observational epidemiology*	Odds ratio (95% CI) per SD increase in plasma lipid conferred by genetic score†
LDL cholesterol	1·54 (1·45–1·63) 2·13 (1·69–2·69), $p=2\times10^{-10}$
HDL cholesterol	0·62 (0·58–0·66) 0·93 (0·68–1·26), $p=0·63$ 

*Observational epidemiology estimates derived from more than 25 000 individuals from prospective cohort studies as shown in the appendix p 22. †LDL genetic score consisting of 13 single nucleotide polymorphisms (SNPs) as shown in the appendix p 27; HDL genetic score consisting of 14 SNPs as shown in the appendix p 28.

Table 4: Estimate of the association of genetically raised LDL cholesterol or HDL cholesterol and risk of myocardial infarction using multiple genetic variants as instruments

Original Investigation

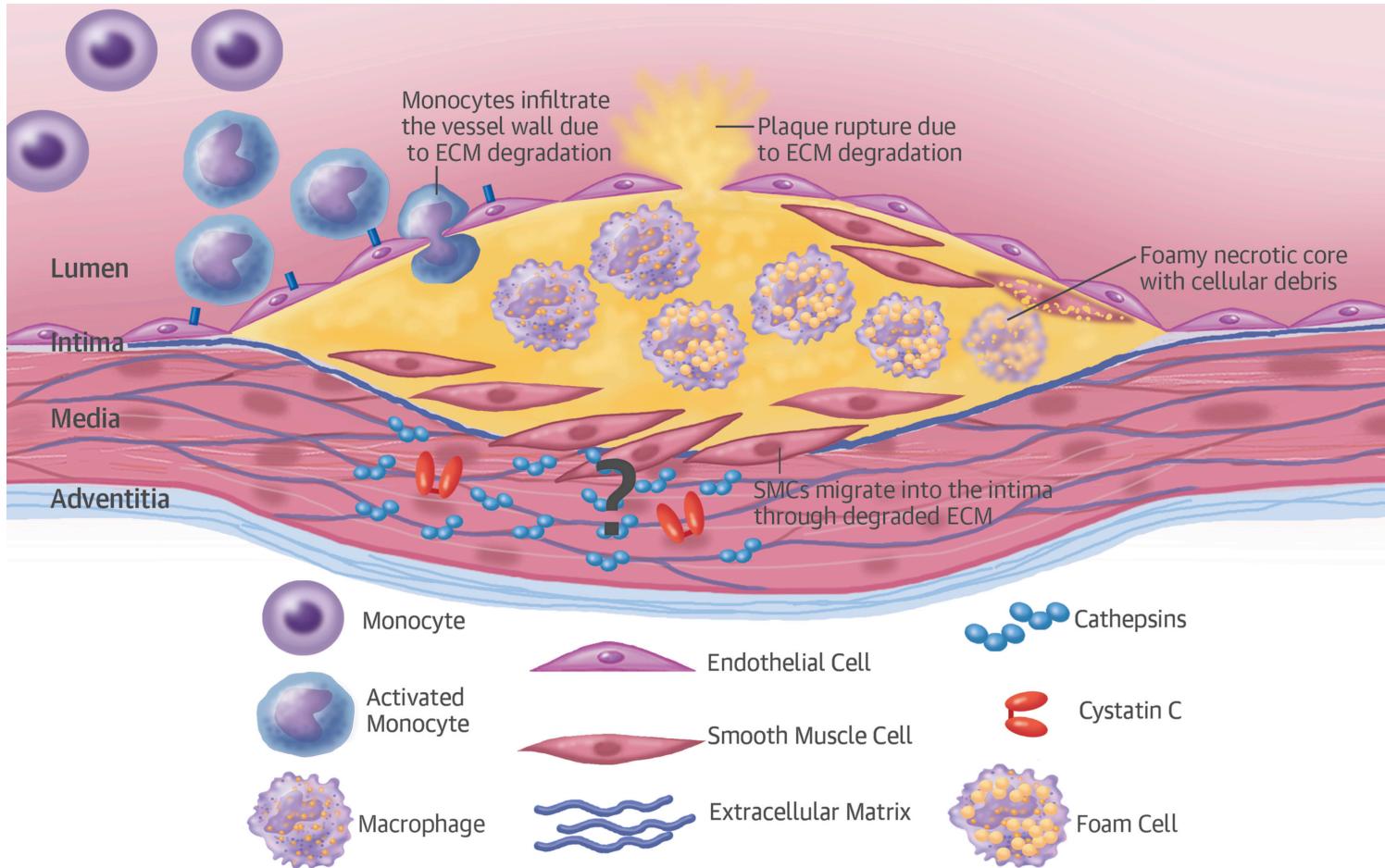
Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study

Sander W. van der Laan MSc ^{a,*}  Tove Fall PhD ^{b,*}  Aicha Soumaré PhD ^c, Alexander Teumer PhD ^d, Sanaz Sedaghat MSc ^f, Jens Baumert PhD ^d, Delilah Zabaneh PhD ^{b,f}, Jessica van Setten PhD ^d, Ivana Isgum PhD ⁱ, Tessel E. Galestro PhD ^k, Johannes Arpegård MD ^{i,n}, Philippe Amouyel MD, MPH ^{j,n}, Stella Trompet PhD ^{d,n}, Melanie Waldenberger PhD, MPH ^{j,f}, Marcus Dörr MD ^{d,n}, Patrik K. Magnusson PhD ⁱ, Vilmantas Giedraitis PhD ^d, Anders Larsson MD, PhD ^v, ... Folkert W. Asselbergs MD, PhD ^{f,k,xx,ooo,*} 

 Show more

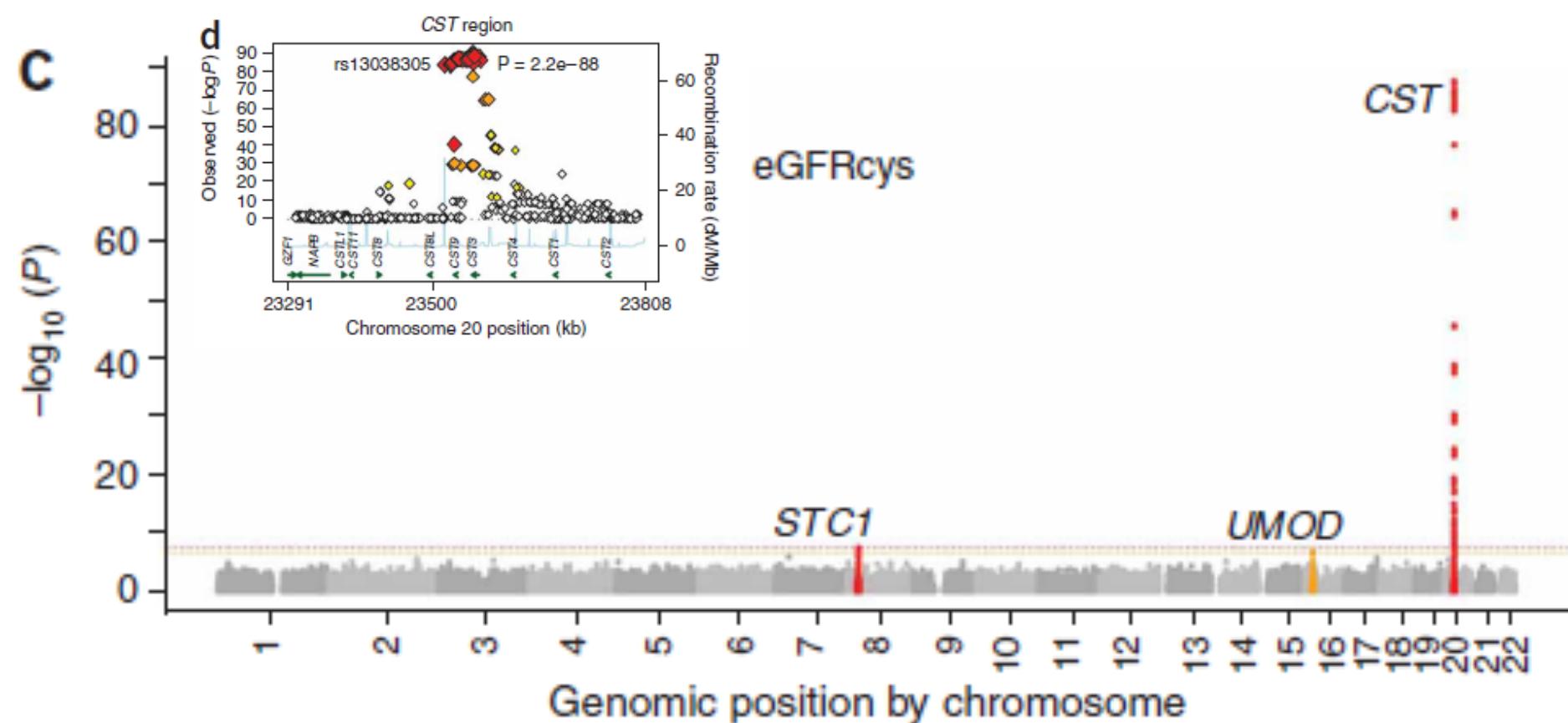
<https://doi.org/10.1016/j.jacc.2016.05.092>

Get rights and content



GWAS: locus with *CST3* pops up (naturally)

- Four Genome–Wide Association Study Meta-Analyses of eGFR and CKD revealed a region containing *CST3* (Cystatin C) as significantly associated
 - Identification of three loci associated with eGFRcys aka CystC expression: *STC1*, *UMOD*, *CST3*



Over 75,000 individuals included

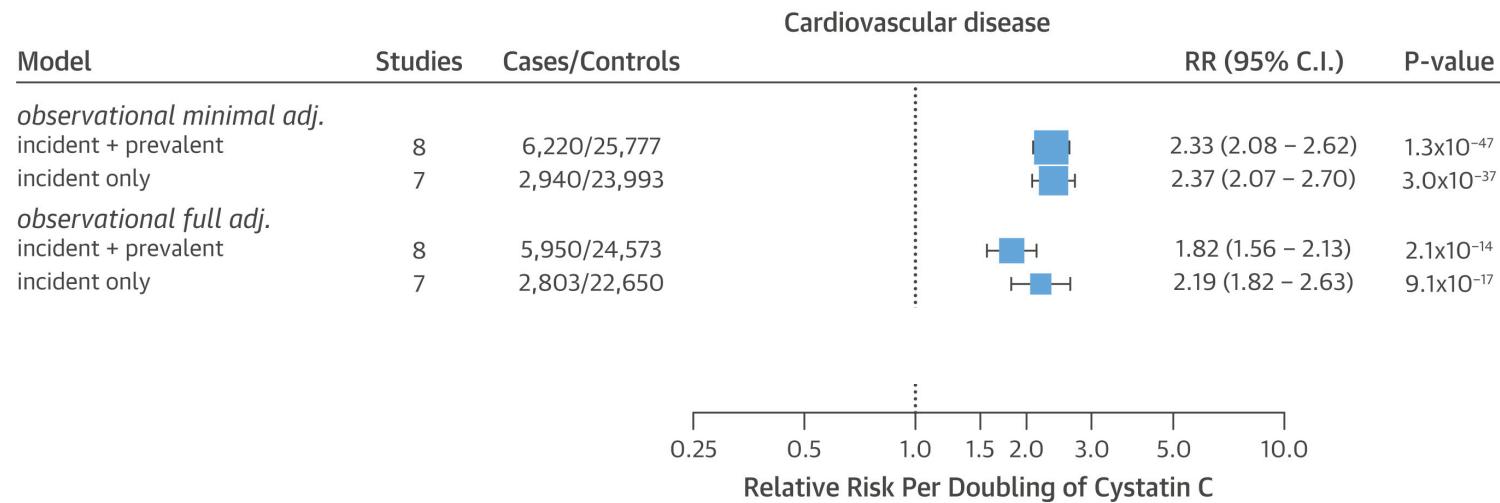
TABLE 1 Characteristics of Prospective Cohorts

Study	Total	SNP*	Cystatin C†	CVD‡	CHD‡	IS‡	HF‡	MI‡	Male	Age (yrs)	Cystatin C (mg/dl)
3C	6,440	6,435	1,244	1,717	1,235	459	439	486	39.19	74.30 ± 5.52	0.92 ± 0.24
EPIC-NL	6,265	5,192	—	1,967	1,430	537	—	1,430	22.39	53.80 ± 10.23	—
GOSH	1,478	1,479	—	493	111	235	233	—	42.08	51.08 ± 11.86	—
HRS	7,844	5,585	5,777	—	—	—	—	—	—	—	0.64 ± 0.34
KORA	4,856	1,867	4,676	540	341	255	—	341	49.53	49.75 ± 14.11	0.80 ± 0.21
NBS	1,819	1,297	—	66	—	66	—	170	49.48	61.05 ± 10.26	—
PIVUS	1,016	949	1,004	255	175	71	75	105	49.90	70.20 ± 0.17	0.90 ± 0.19
PREVEND	3,245	3,245	3,245	236	190	58	—	—	50.26	49.42 ± 12.25	0.87 ± 0.17
PROSPER§	5,244	5,150	—	2,561	2,034	779	211	762	48.13	75.34 ± 3.35	—
Rotterdam	7,983	5,974	3,906	3,579	1,934	1,328	1,625	1,176	38.90	73.06 ± 7.49	1.11 ± 0.28
SHIP	3,224	3,224	3,212	114	19	87	—	134	48.08	54.46 ± 15.26	0.88 ± 0.30
Tromsø	6,129	—	6,129	1,251	—	494	—	881	47.59	60.59 ± 10.25	0.86 ± 0.18
TWINGENE	6,902	6,902	6,740	932	610	287	206	—	47.23	64.83 ± 8.26	1.02 ± 0.30
ULSAM	1,221	1,107	1,193	503	285	175	220	—	100.00	71.00 ± 0.64	1.25 ± 0.27
WHI	7,854	7,844	—	4,831	2,934	2,115	—	2,934	0.00	67.97 ± 6.58	—
Whitehall II	4,961	5,011	—	349	254	111	—	254	74.58	49.19 ± 5.99	—
Overall	76,481	61,261	37,126	19,394	11,552	7,057	3,009	8,673	—	—	—

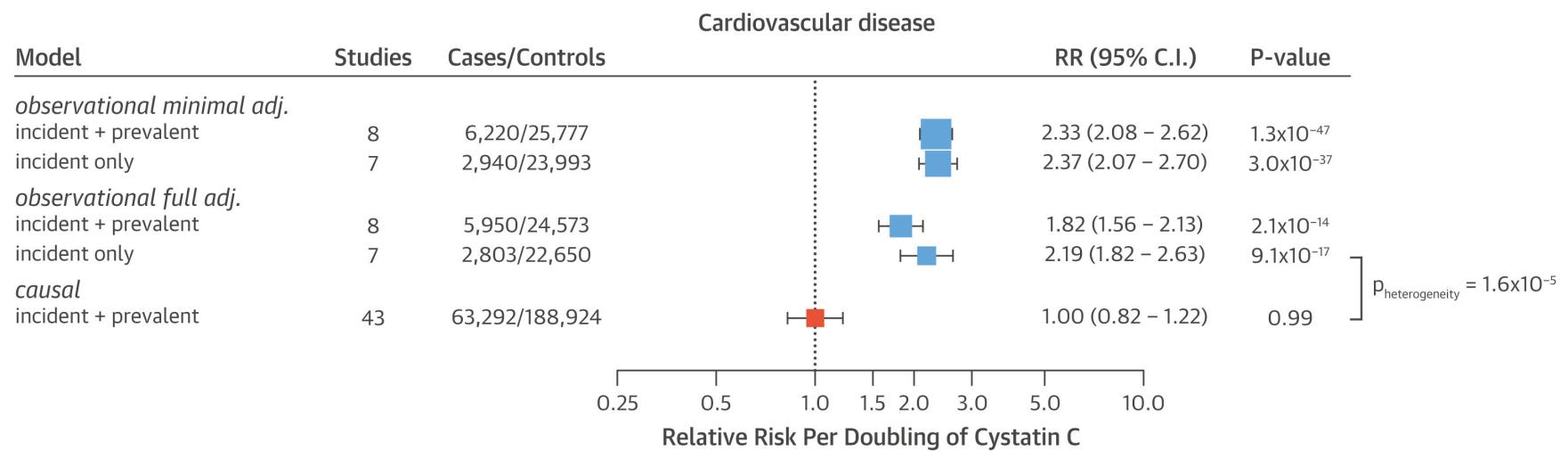
Values are n, %, or mean ± SD. *Total number of individuals with genotype data. †Genetic data were available in 29,805 of the 37,126 individuals that had values for cystatin C, which we used to associate rs911119 with circulating cystatin C. For the genetic analysis of CVD, CHD, IS, and HF, cohorts that contributed toward consortia were excluded. ‡Indicates total incident and prevalent cases of disease or composite diseases in the case of CVD. §PROSPER is a randomized clinical trial. ||For the association of SNP with cystatin C concentrations, 9,488 samples were available in TWINGENE.

CHD = coronary heart disease; CVD = cardiovascular disease; HF = heart failure; IS = ischemic stroke; MI = myocardial infarction; SNP = single-nucleotide polymorphism.

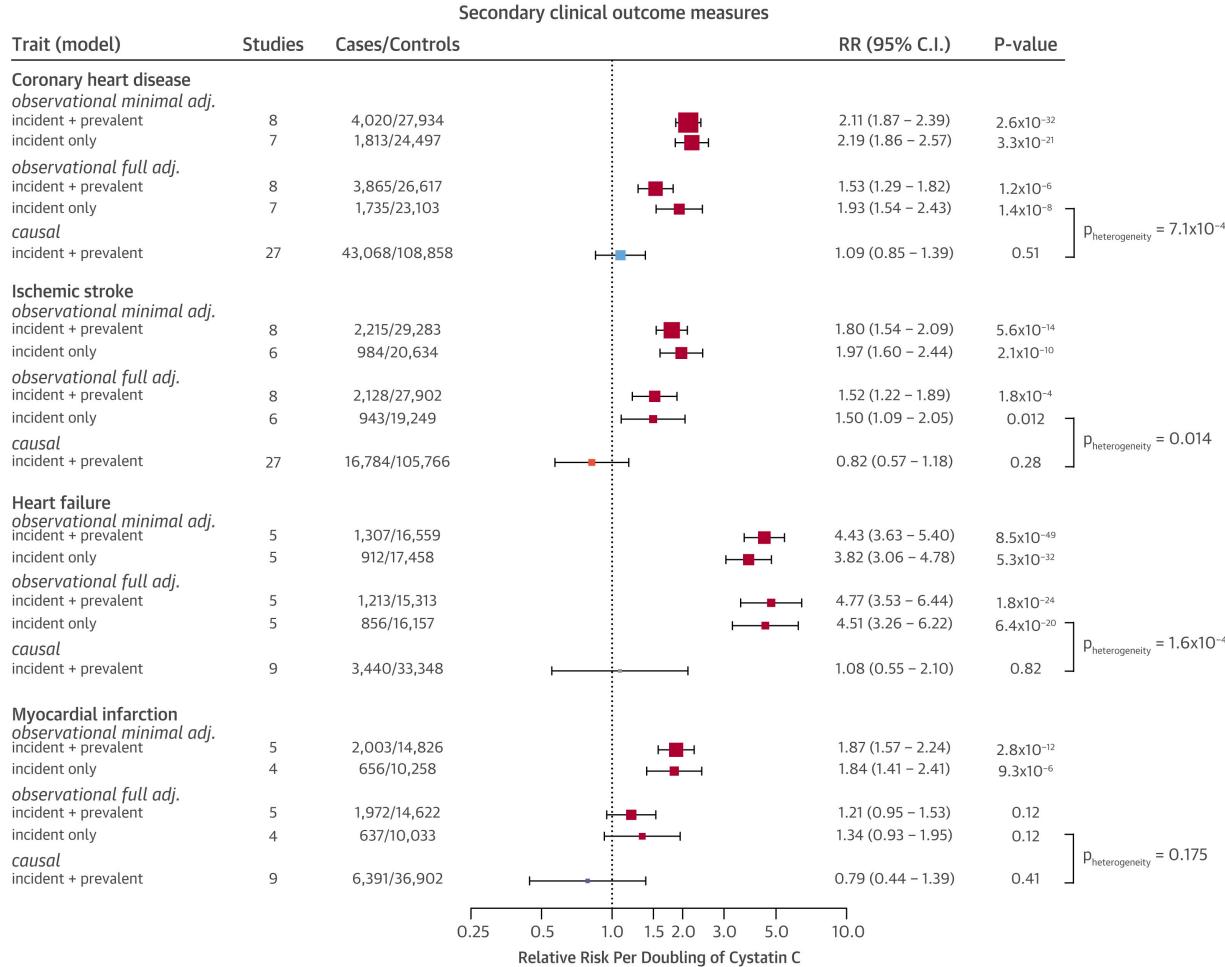
Cystatin C associates to CVD in observational studies



No causal effect of Cystatin C on CVD



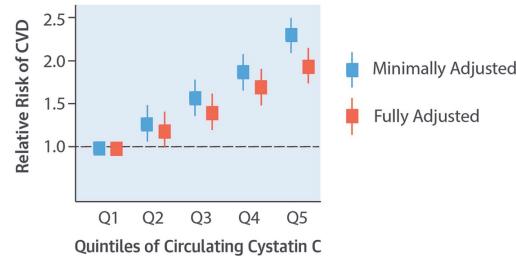
Secondary clinical endpoints: nada, nothing, zip



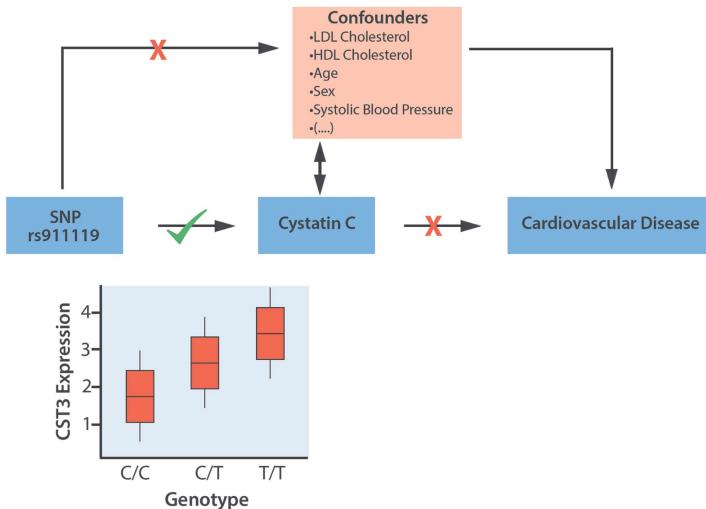
Cystatin C levels are not causal to CVD risk

CENTRAL ILLUSTRATION: Assessing Causality of Cystatin C in CVD

A. Observational Epidemiology

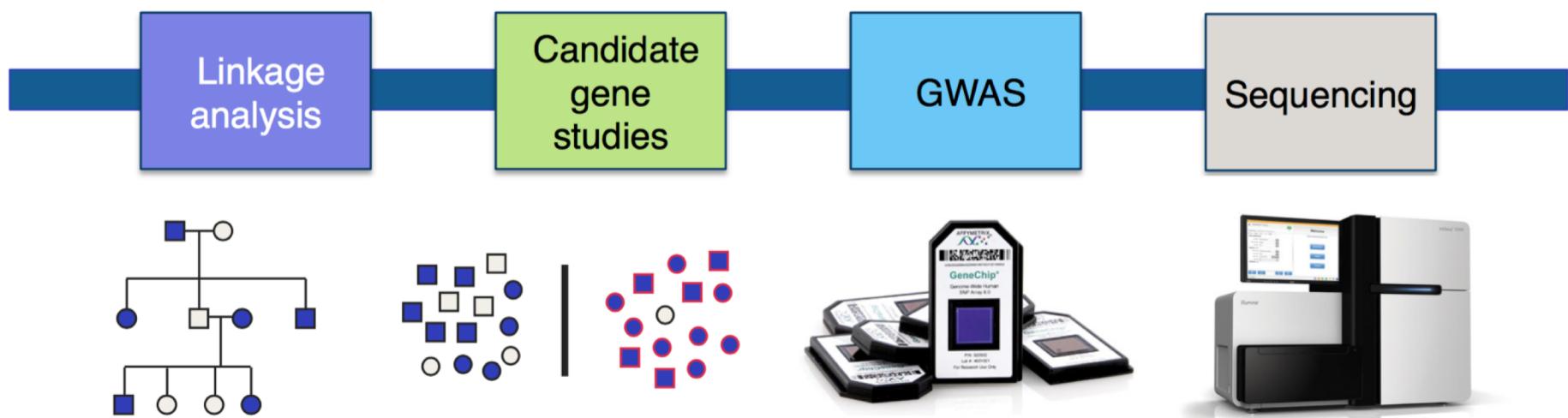


B. Mendelian Randomization



van der Laan, S.W. et al. J Am Coll Cardiol. 2016;68(9):934-45.

Drug discovery & validation



Drug discovery

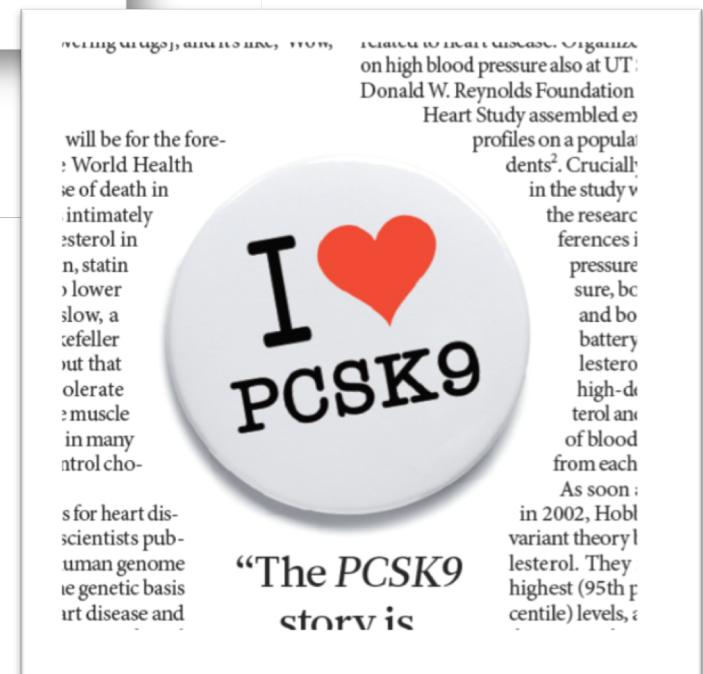
- Each locus likely harbors a disease-driving gene (or regulatory element)
- The magnitude of the odds ratio does not indicate
 - Potential biological value
 - Potential for therapy (“druggability”)
- Examples
 - *PPARG* in type 2 diabetes (thiazolidinediones)
 - *KCNJ11* and type 2 diabetes (sulfonylureas)
 - *PCSK9* and myocardial infarction (*PCSK9* inhibitors)

ORIGINAL ARTICLE

Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease

Jonathan C. Cohen, Ph.D., Eric Boerwinkle, Ph.D., Thomas H. Mosley, Jr., Ph.D., and Helen H. Hobbs, M.D.

- **PCSK9 first discovered in familial hypercholesterolemia**
- **Then discovered in a GWAS of EOMI**



LDL and PCSK9 in two populations

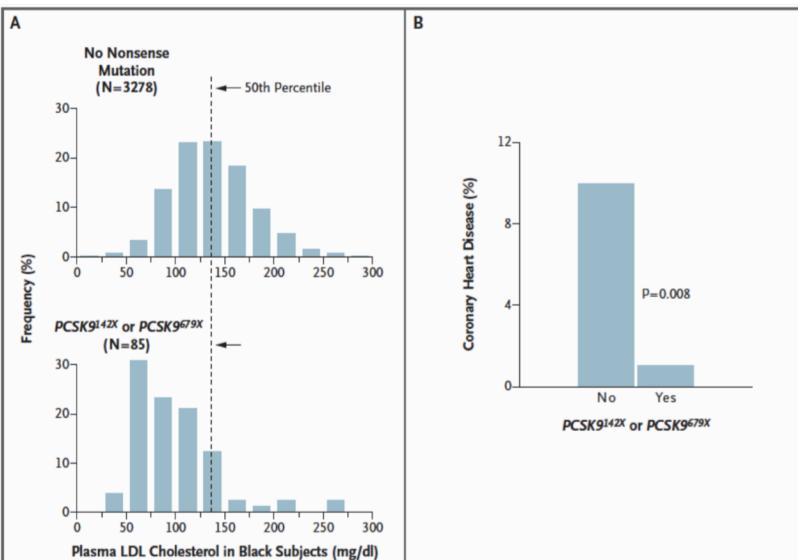


Figure 1. Distribution of Plasma LDL Cholesterol Levels (Panel A) and Incidence of Coronary Heart Disease (Panel B) among Black Subjects, According to the Presence or Absence of a *PCSK9^{142X}* or *PCSK9^{679X}* Allele.

In Panel A, the distribution of plasma LDL cholesterol levels at baseline among 3278 black subjects who did not have a *PCSK9^{142X}* or *PCSK9^{679X}* allele (top) is compared with the distribution of levels among the 85 black subjects who had one of these two alleles (bottom). Panel B shows the percentage of participants from these two groups who had no evidence of coronary heart disease at baseline and in whom coronary heart disease developed during the 15-year follow-up period. To convert values for LDL cholesterol to millimoles per liter, multiply by 0.02586.

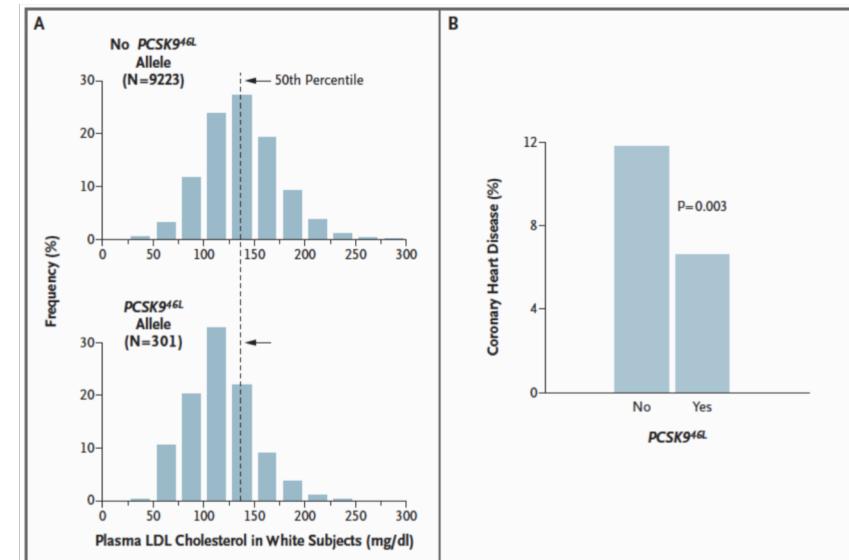
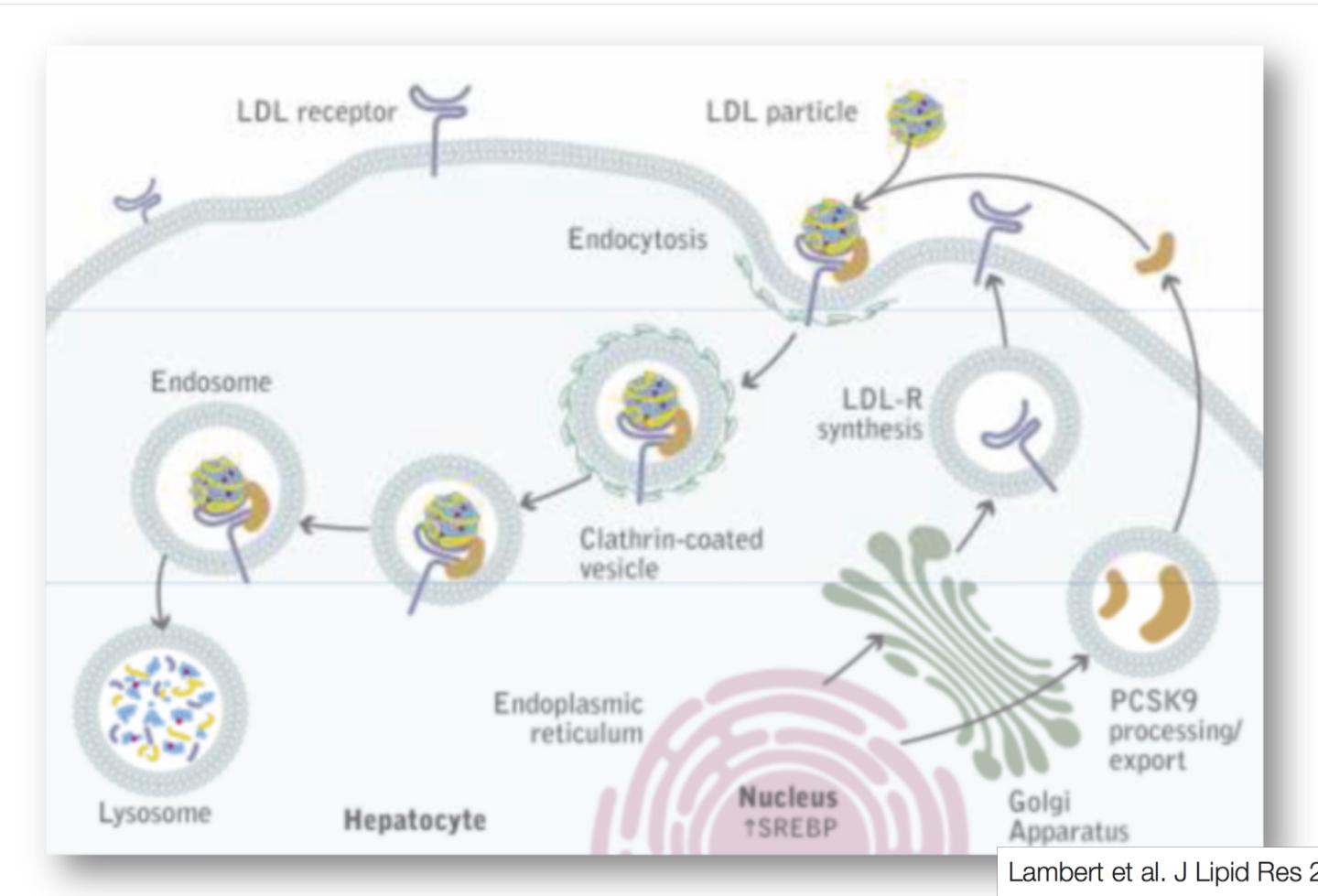


Figure 2. Distribution of Plasma LDL Cholesterol Levels (Panel A) and Incidence of Coronary Events (Panel B) among White Subjects, According to the Presence or Absence of a *PCSK9^{46L}* Allele.

In Panel A, the distribution of plasma LDL cholesterol levels at baseline among 9223 white subjects who did not have a *PCSK9^{46L}* allele (top) is compared with the distribution of levels among the 301 white subjects who were either heterozygous or homozygous for this allele (bottom). Panel B shows the percentage of participants from these two groups who had no evidence of coronary heart disease at baseline and in whom coronary heart disease developed during the 15-year follow-up period. To convert values for LDL cholesterol to millimoles per liter, multiply by 0.02586.

PCSK9 promotes *LDLR* degradation



Lambert et al. J Lipid Res 2012

Conclusions

- The genomic sequence is set at conception, no need to worry about confounding factors
- GWAS have been and will continue to be very successful
- Still need large sample sizes for sufficient power
- Complex genetics studies useful to gain information causality and drug discovery
- The jury is still out with respect to risk prediction
- Interpretation and translation will be the major challenge in the next decade

Cardiovascular Genetic Research

Laboratory of Clinical Chemistry and Hematology

Dr. Sander W. van der Laan

Research topics

Biomarker Discovery & Validation

Athero-Express | AAA-Express | CTMM | many more

Cardiovascular Genetics

Prof. Dr. Folkert W. Asselbergs

Prof. Dr. Gerard Pasterkamp

Dr. Jessica van Setten

Dr. Magdalena Harakalova

Dr. Floriaan Schmidt

Dr. Michal Mokry

Ischemic stroke

GWAS | 4C | MR | CRISPR-Cas9

Cardiovascular Genomics

*Next-Gen Sequencing | eQTL | pQTL |
Transcriptomics | Epigenomics | MR*

S.W. van der Laan

s.w.vanderlaan-2@umcutrecht.nl | @swvanderlaan



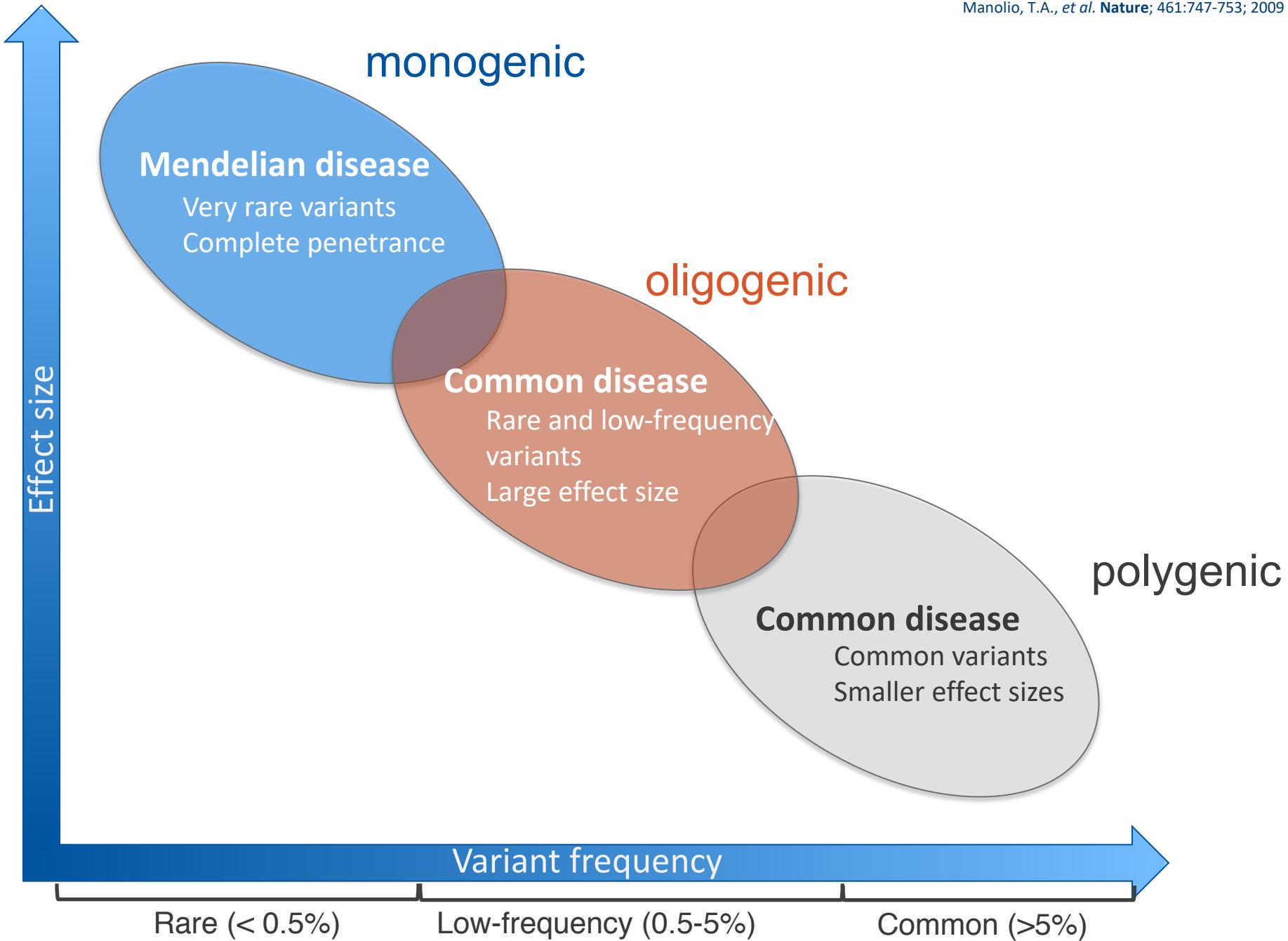
UMC Utrecht



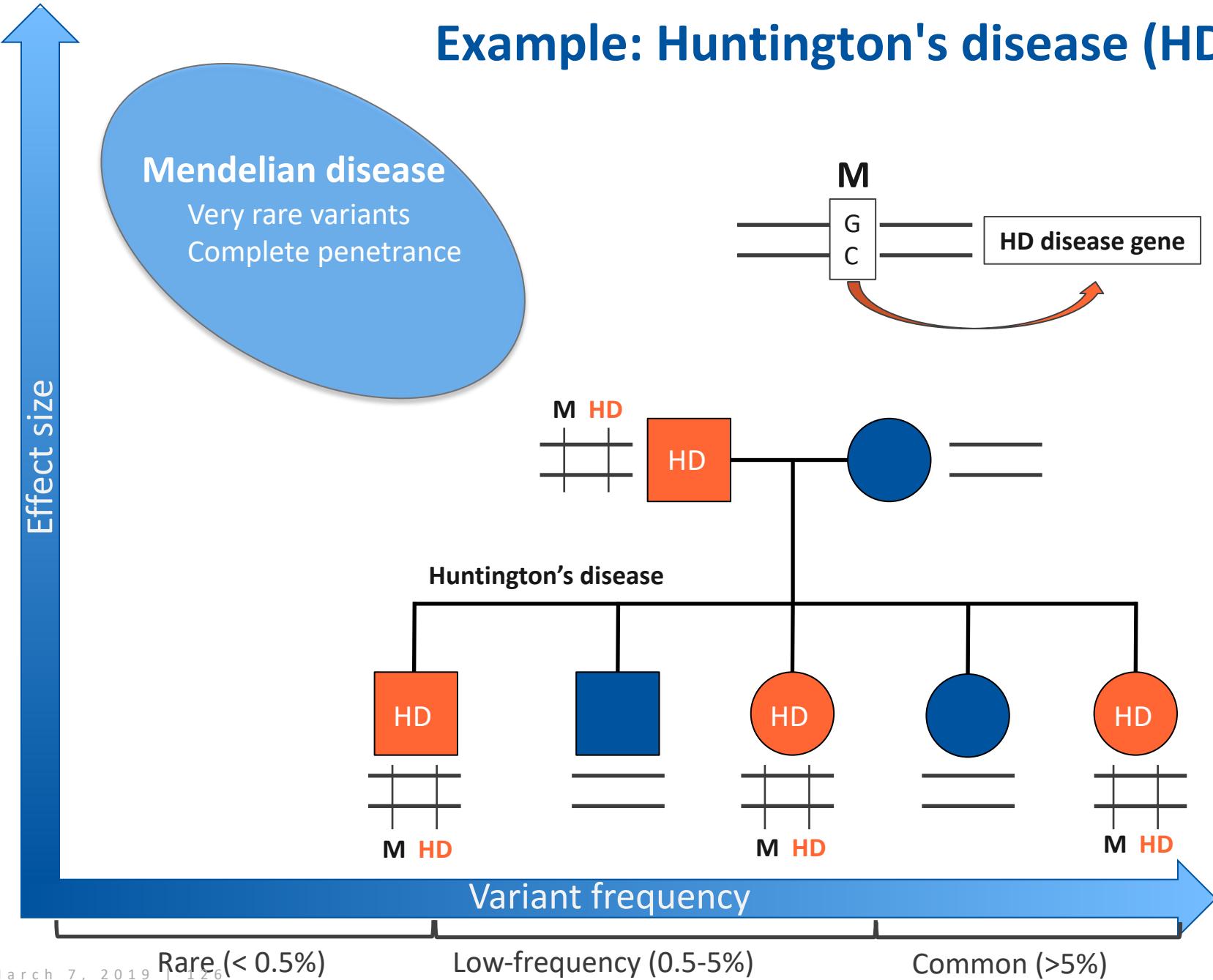
CVON

CardioVasculair Onderzoek Nederland
Een privaat-publiek initiatief





Example: Huntington's disease (HD)



A polymorphic DNA marker genetically linked to Huntington's disease

James F. Gusella*, Nancy S. Wexler^{†||}, P. Michael Conneally[†], Susan L. Naylor[§],
Mary Anne Anderson*, Rudolph E. Tanzi*, Paul C. Watkins*, Kathleen Ottina*,
Margaret R. Wallace[‡], Alan Y. Sakaguchi[§], Anne B. Young^{||}, Ira Shoulson^{||},
Ernesto Bonilla^{||} & Joseph B. Martin*

* Neurology Department and Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

† Hereditary Disease Foundation, 9701 Wilshire Blvd, Beverly Hills, California 90212, USA

‡ Department of Medical Genetics, Indiana University Medical Center, Indianapolis, Indiana 46223, USA

§ Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263, USA

|| Venezuela Collaborative Huntington's Disease Project*

Family studies show that the Huntington's disease gene is linked to a polymorphic DNA marker that maps to human chromosome 4. The chromosomal localization of the Huntington's disease gene is the first step in using recombinant DNA technology to identify the primary genetic defect in this disorder.

1983

A polymorphic DNA marker genetically linked to Huntington's disease

James F. Gusella*, Nancy S. Wexler^{†||}, P. Michael Conneally[†], Susan L. Naylor[§],

Mary Anne Anderson*, Rudolph E. Tanzi*

Margaret R. Wallace[‡], Alan Y. Sakaguchi[‡],
Ernesto Bonilla^{||} & James C. Ross^{||}

* Neurology Department and Genetics Unit, Massachusetts General Hospital

† Hereditary Disease Foundation, 9701 Wilshire Boulevard, Los Angeles, CA 90021

‡ Department of Medical Genetics, Indiana University Medical School, Indianapolis, IN 46285

§ Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, NY 14263
|| Venezuela Collaborative Huntington's Disease Study

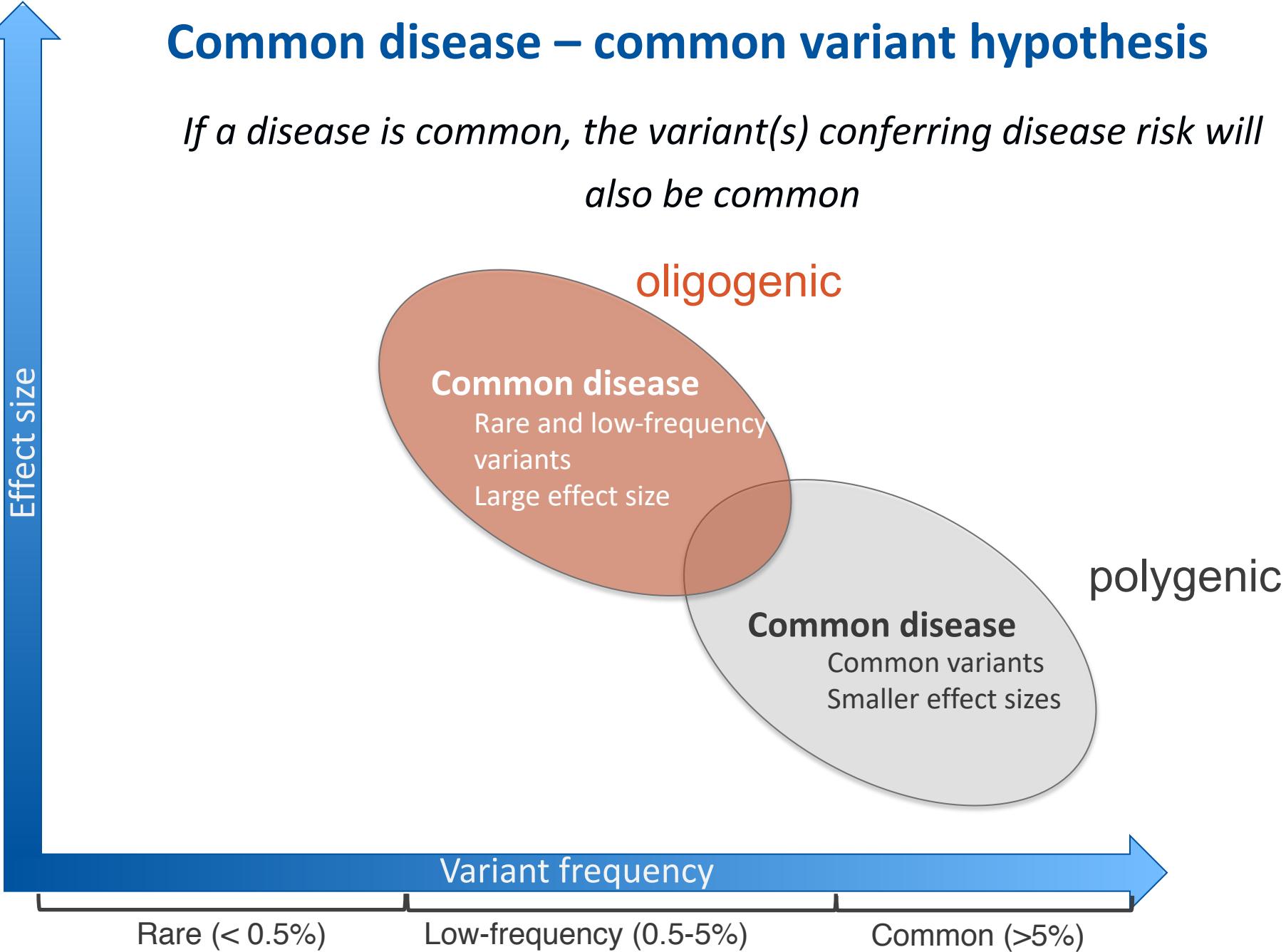
1983

1989



Common disease – common variant hypothesis

If a disease is common, the variant(s) conferring disease risk will also be common



What type of disease are we looking at?

• Complex diseases

- Polygenic, multifactorial diseases
 - Diabetes mellitus
 - Asthma
 - Cardiovascular disease
 - Hypertension

Each gene contributes a little to the disease

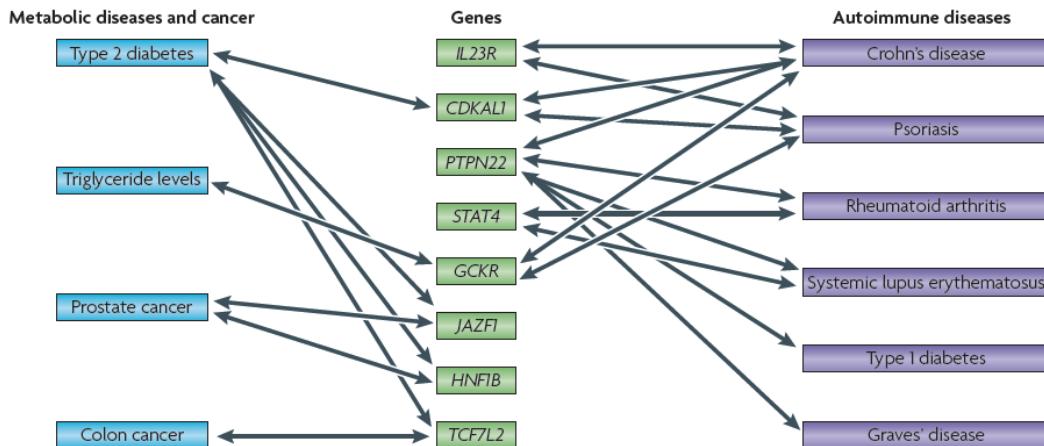
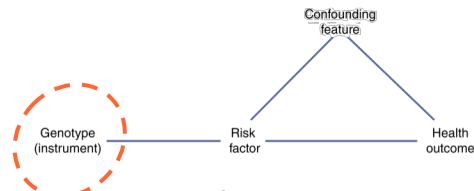


Table 3 Some Recent Genes/Loci Identified in Coronary Artery Disease

Gene/Locus	Functional Genomics	Independently Replicated	Reference(s)
CFH	Inflammation	Yes	49,50
LTA4H	Inflammation	No	12
FLAP	Inflammation	No	5
Lymphotxin α	Inflammation	No	69
Galectin 2	Inflammation	No	68
Stromelysin 1	Inflammation	No	61
MHC2TA	Inflammation	No	71
Kalirin	Inflammation	Yes	13
TSP 4	Endothelial integrity	Yes	57-61
Connexin 37	Endothelial integrity	No	61
MEF2A	Endothelial integrity	Yes	62,63
Apo E4	Lipoprotein handling	Yes	52,53,78
LRP6	Lipoprotein handling	No	56
PCSK9	Lipoprotein handling	No	42,43
VAMP8	Thrombosis	No	72
PAI-1	Thrombosis	No	61
Factor V (1691A)	Thrombosis	No	73
Prothrombin (20210A)	Thrombosis	No	73
9p21	Unknown	Yes	38-40

Apo E4 = apolipoprotein E4; FLAP = 5-lipoxygenase activating protein; LRP6 = low-density lipoprotein receptor-related protein 6; LTA4H = leukotriene A4 hydrolase; MEF2A = myocyte enhancer factor 2a; MHC2TA = major histocompatibility factor class 2 transactivator; PAI-1 = plasminogen activator inhibitor 1; PCSK9 = proprotein convertase subtilisin/kexin type 9; TSP 4 = thrombospondin 4; VAMP8 = vesicle-associated membrane protein 8; other abbreviations as in Table 2.



SNP selection

- One GWAS on CystC, rs1158167 in Framingham Heart Study
- Three (meta-analyses of) GWAS on eGFR_{CystC}
 - rs911119, rs13038305
 - $eGFR_{CystC} = 76.7 \times (\text{serum CystC})^{-1.19}$
 - Log linear relation between serum CystC and eGFR

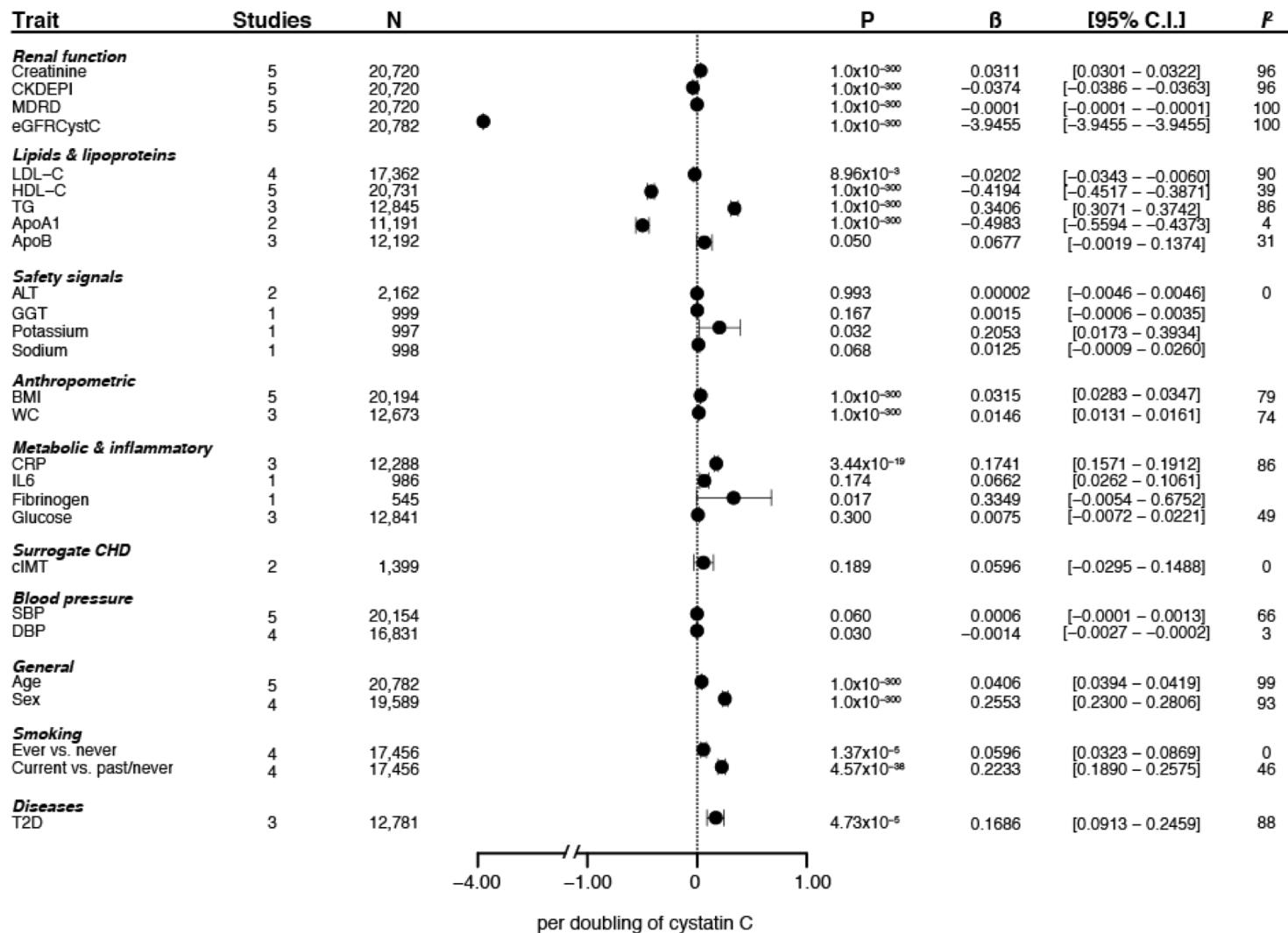
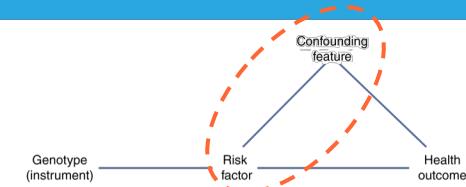
Proxy	Distance	r ²	Chr	Position	Minor	Major	MAF	Variant	Gene
<u>rs1158167</u>	34,548	0.913	20	23,526,189	G	A	0.25	downstream	n/a
<u>rs17751897</u>	20,023	0.955	20	23,540,714	C	T	0.242	intergenic	n/a
<u>rs12625716</u>	5,892	0.955	20	23,554,845	A	G	0.242	downstream	n/a
<u>rs6048952</u>	5,480	0.955	20	23,555,257	G	A	0.242	downstream	n/a
<u>rs13038305</u>	2,475	1.00	20	23,558,262	T	C	0.233	intronic	CST3
<u>rs911119</u>	0	1	20	23,560,737	C	T	0.233	intronic	CST3
<u>rs3827143</u>	6,880	1	20	23,567,617	G	A	0.233	upstream	n/a
<u>rs6114208</u>	8,997	1	20	23,569,734	G	C	0.233	upstream	n/a

Studies included

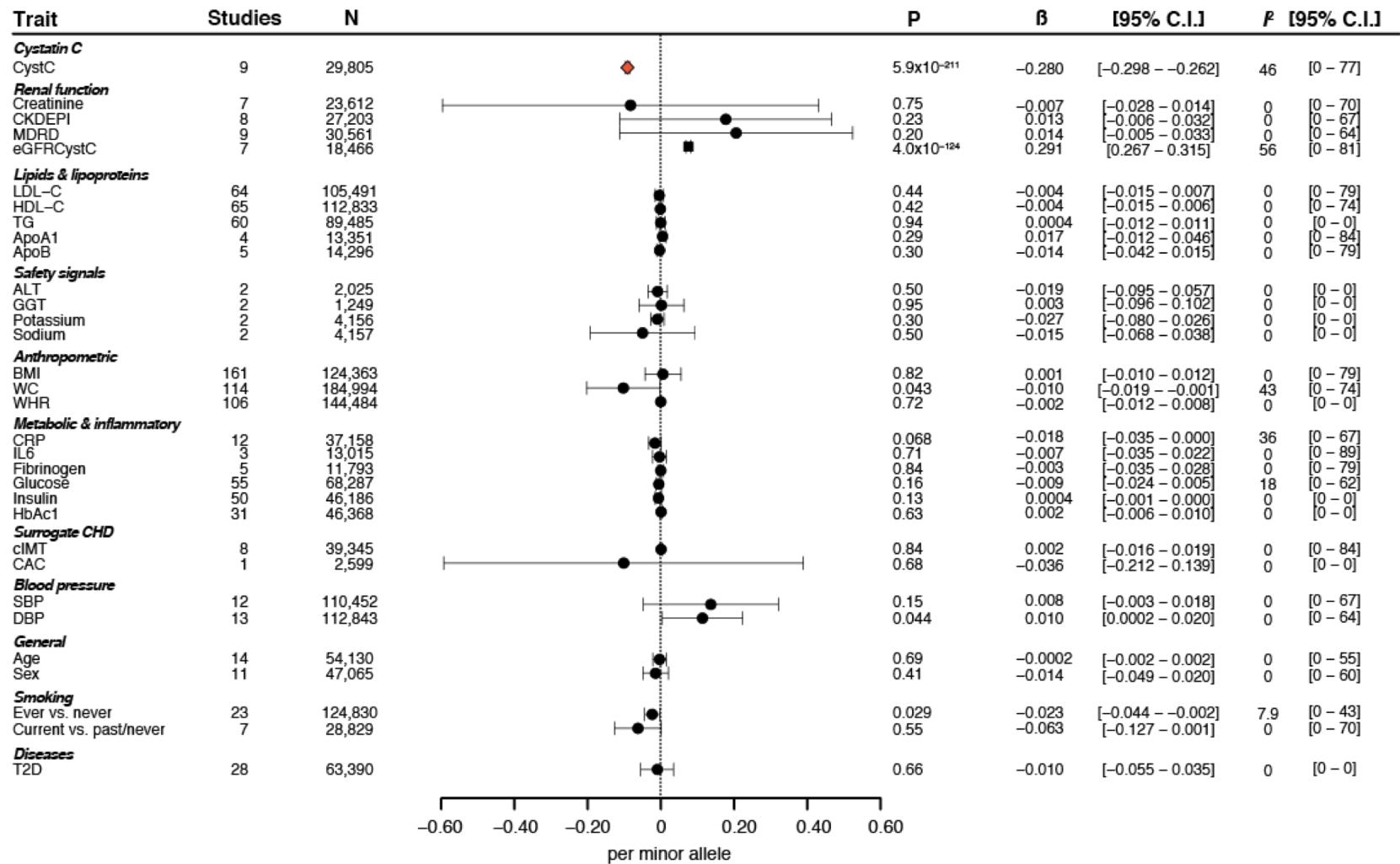
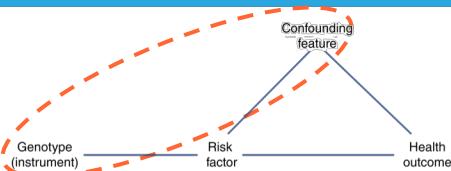
Study	N _{total}	N _{SNP}	N _{CystC}	N _{CVD}	N _{CHD}	N _{stroke}	N _{HF}	N _{MI}
3C	6,440	6,435	1,244	1,717	1,235	459	439	486
EPIC-NL	6,265	5,192	-	1,967	1,222	443	-	1,430
GOSH	1,478	1,479	-	493	111	235	233	-
HRS	7,844	5,585	5,777	-	-	-	-	-
KORA	4,856	1,867	4,676	540	341	86	-	341
NBS	1,819	1,297	-	66	113	49	-	170
PIVUS	1,016	949	1,004	255	175	63	66	105
PREVEND	3,245	3,245	3,245	236	190	58	-	-
PROSPER	5,244	5,150	-	2,561	1,994	766	210	762
Rotterdam	4,730	5,974	3,906	3,579	1,934	1,328	1,625	1,176
SHIP	3,224	3,224	3,212	114	6	27	-	134
Tromsø	6,129	-	6,129	1,251	-	494	-	881
TWINGENE	6,902	6,902	6,740	940	610	287	206	-
ULSAM	1,221	1,107	1,193	503	285	159	199	-
WHI	7,854	7,844	-	4,831	2,930	2,111	-	2,934
Whitehall II	4,961	5,011	-	349	253	111	-	254
CARDIoGRAM				20,251	20,251			
C4D				15,388	15,388			
METASTROKE				12,389		12,389		
CHARGE-HF				2,526			2,526	
Overall	73,228	61,261	37,126	63,292	47,038	19,065	5,504	8,673

Van der Laan *et al.* In revision @ JACC

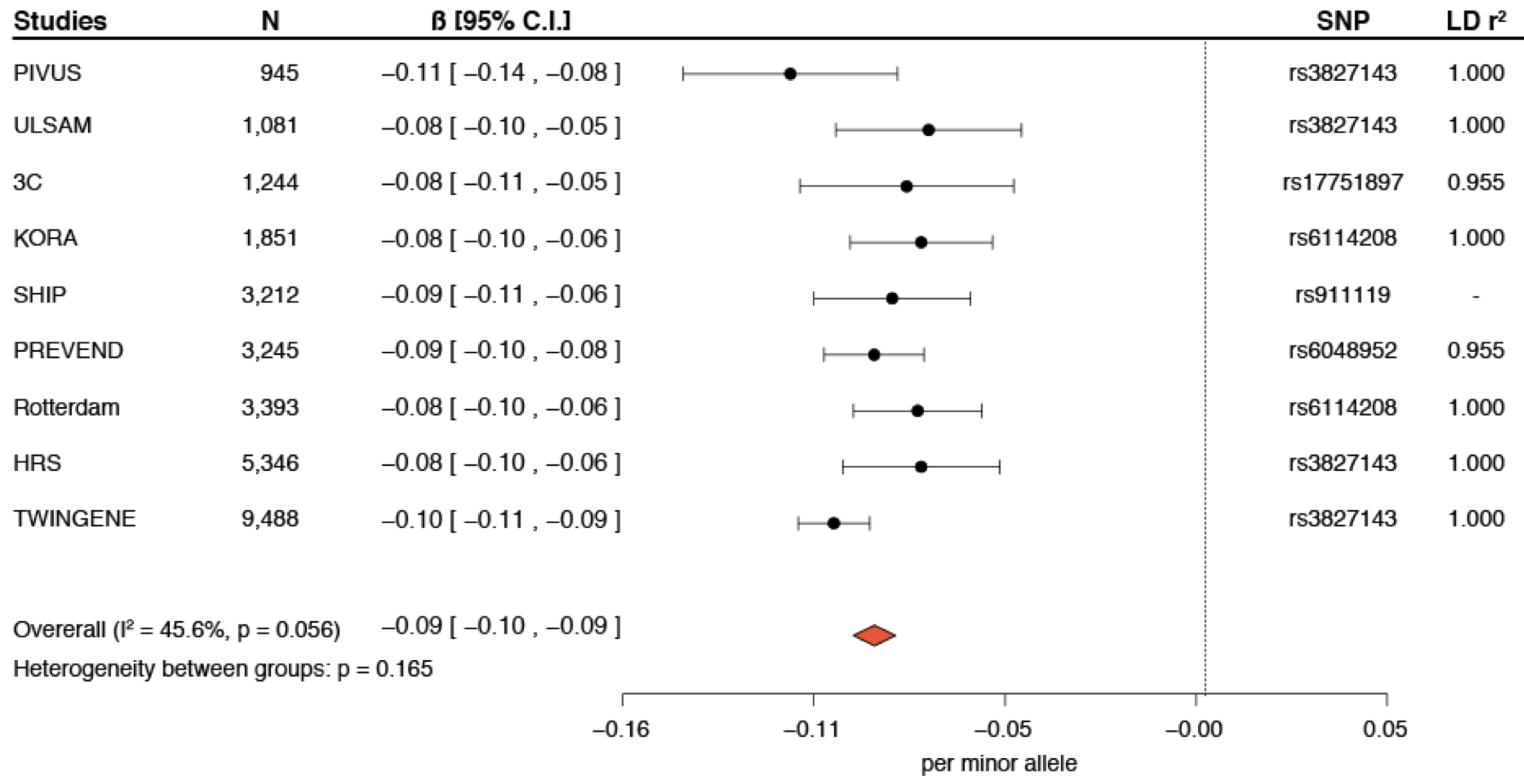
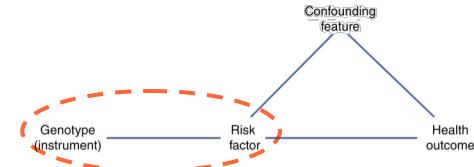
Serum cystatin C vs. risk factors



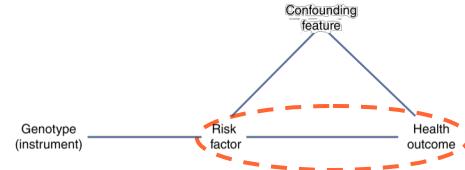
Cystatin C variant vs. risk factors



SNP vs. cystatin C

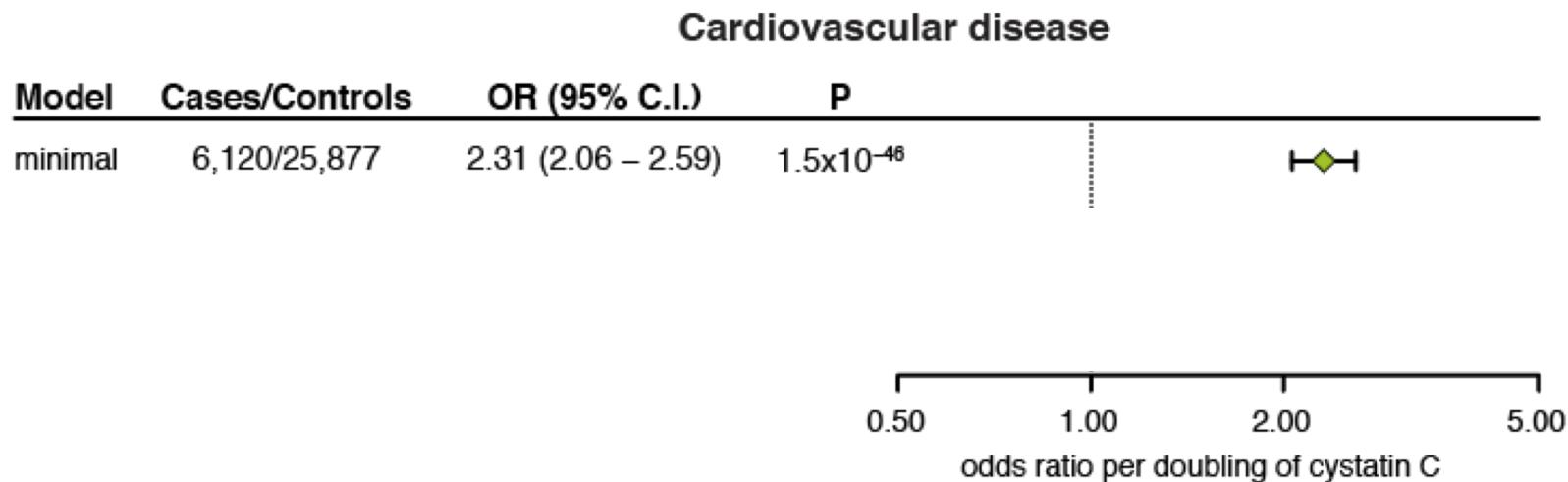


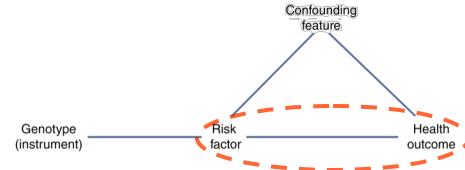
- Per minor allele there is 6.13% decrease in CystC [mg/L]
- $\beta = -0.09 [-0.10 - -0.09]$, $p = 5.95 \times 10^{-163}$, $N = 29,805$
- This explains $\approx 2.75\%$ of the phenotypic variation



Observational analysis

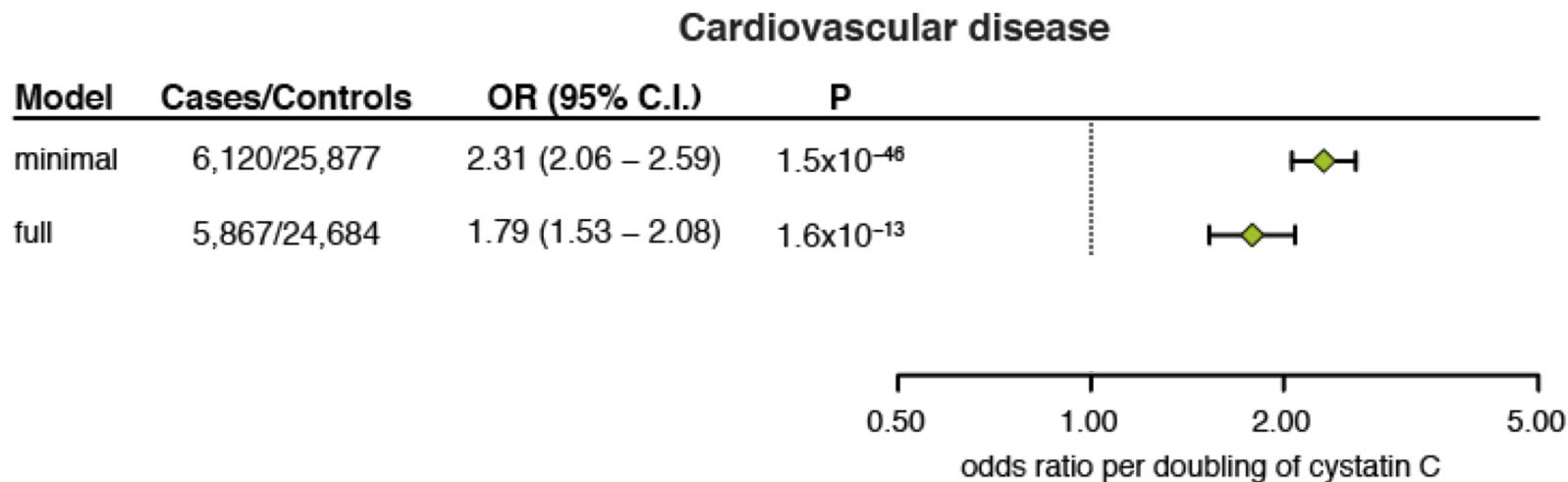
- Minimally adjusted model: age + sex



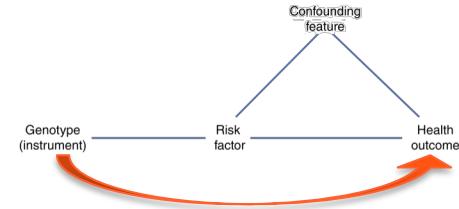


Observational analysis

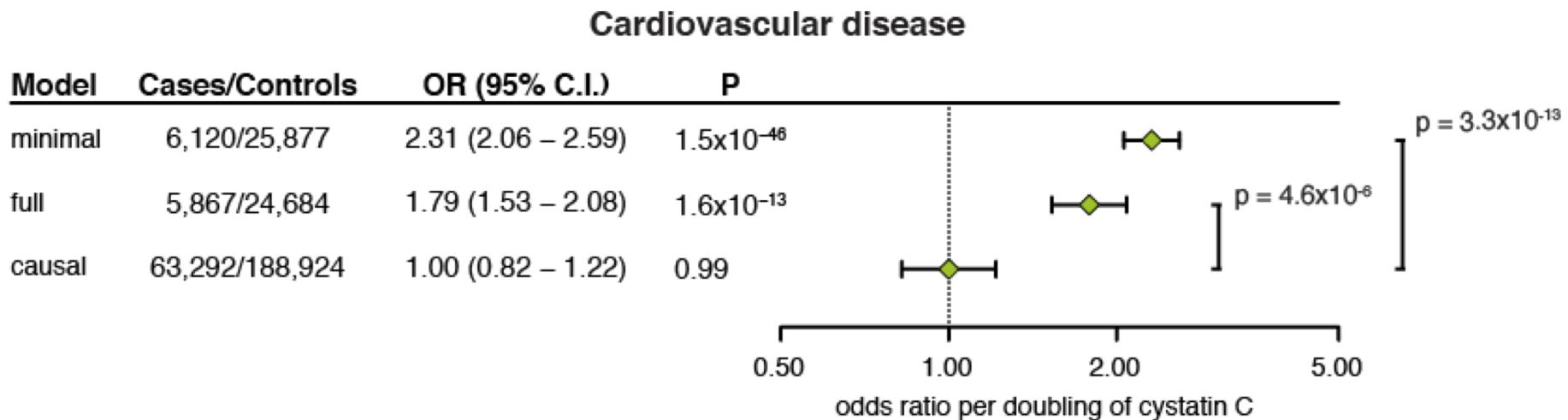
- Minimally adjusted model: age + sex
- Fully adjusted model: age + sex + HDL + systolic blood pressure + BMI + smoking status + CKDEPI



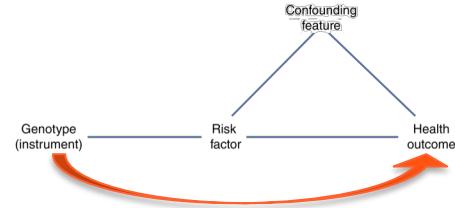
Mendelian randomization analysis



- Minimally adjusted model: age + sex
- Fully adjusted model: age + sex + HDL + systolic blood pressure + BMI + smoking status + CKDEPI
- Causal effect: *nada, nothing, niets, rien, nichts, ничего*
- *Interaction analysis*: significant – models are different



Other outcomes



- Causal effect: *nada, nothing, niets, rien, nichts, ничего*

