

# Cardiovascular Genomics

An introduction to the analysis of genetic variants  
in cardiovascular disease

Sander W. van der Laan

[s.w.vanderlaan-2@umcutrecht.nl](mailto:s.w.vanderlaan-2@umcutrecht.nl)



University Medical Center  
Utrecht



# What we'll discuss today...

- Recapture Some Basic Genetics
  - Human Genome & Genetic Variation
  - How do we measure it?
- Discoveries in Mendelian Diseases & Complex Diseases
  - Coronary artery disease
  - Risk factors
- Clinical utilities?
  - Genetic Burden for Disease (Risk)
  - PCSK9
  - Mendelian Randomization of HDL & MI



University Medical Center  
Utrecht



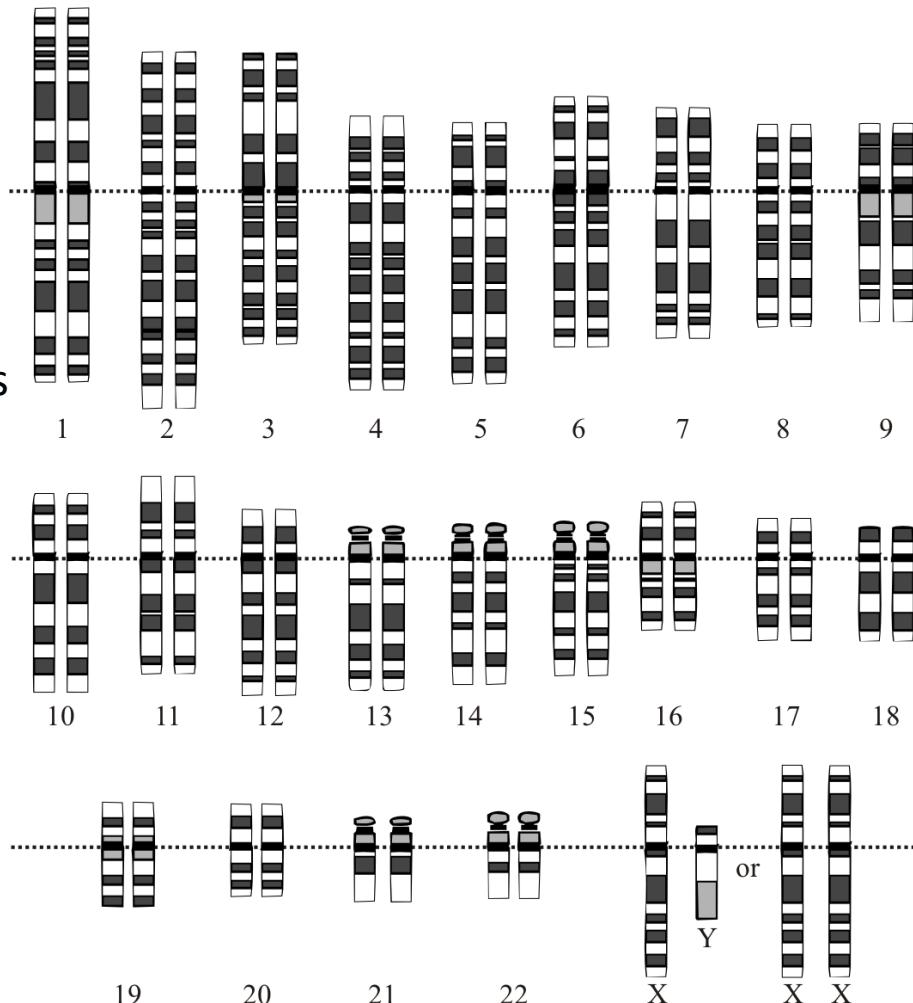
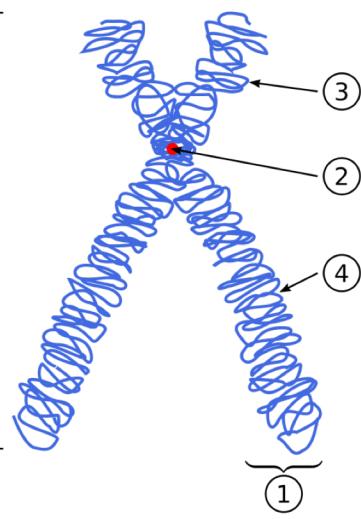
Recapturing Some Basic Genetics

# THE HUMAN GENOME

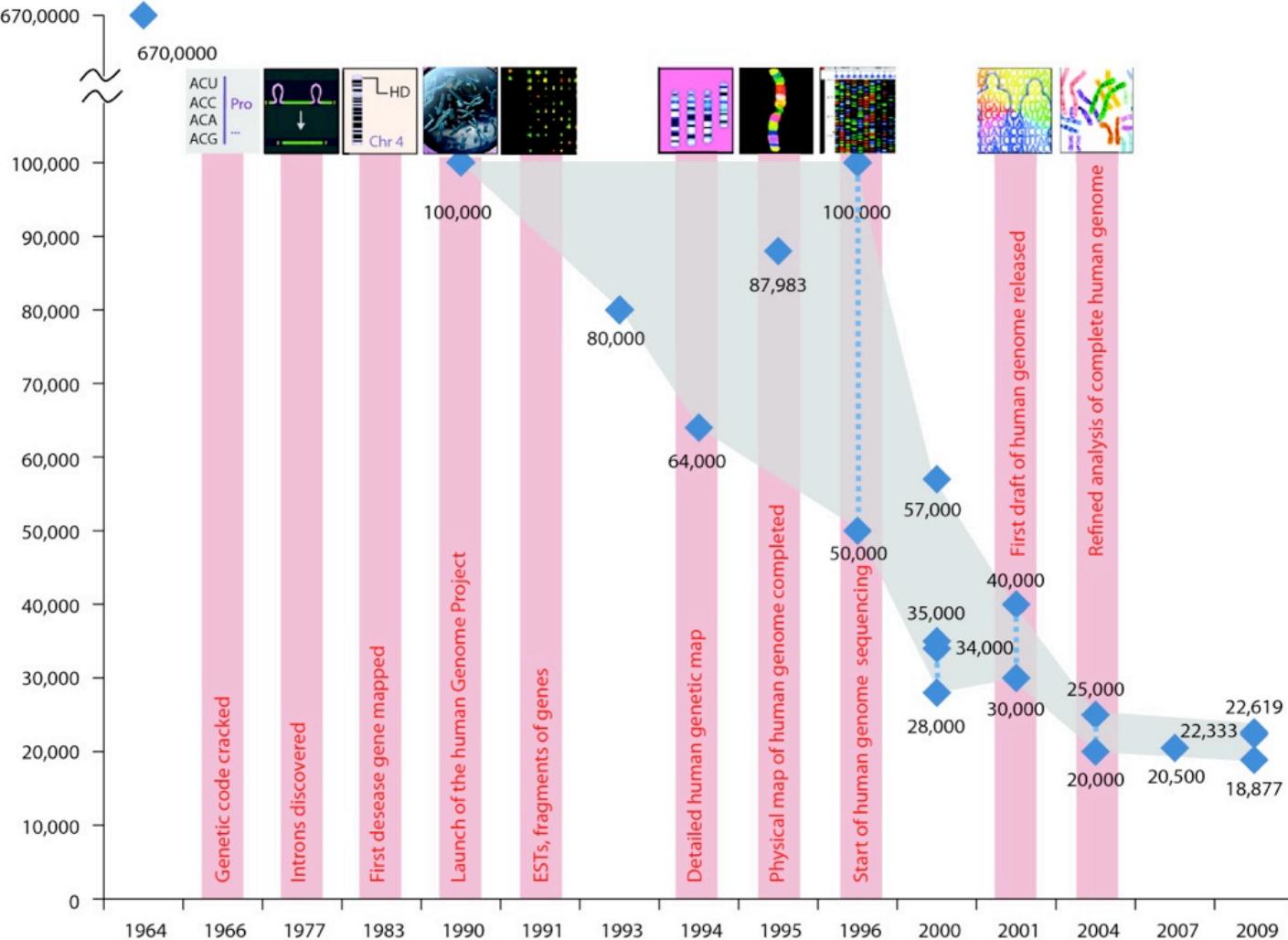


# Human Genome: *the chromosomes*

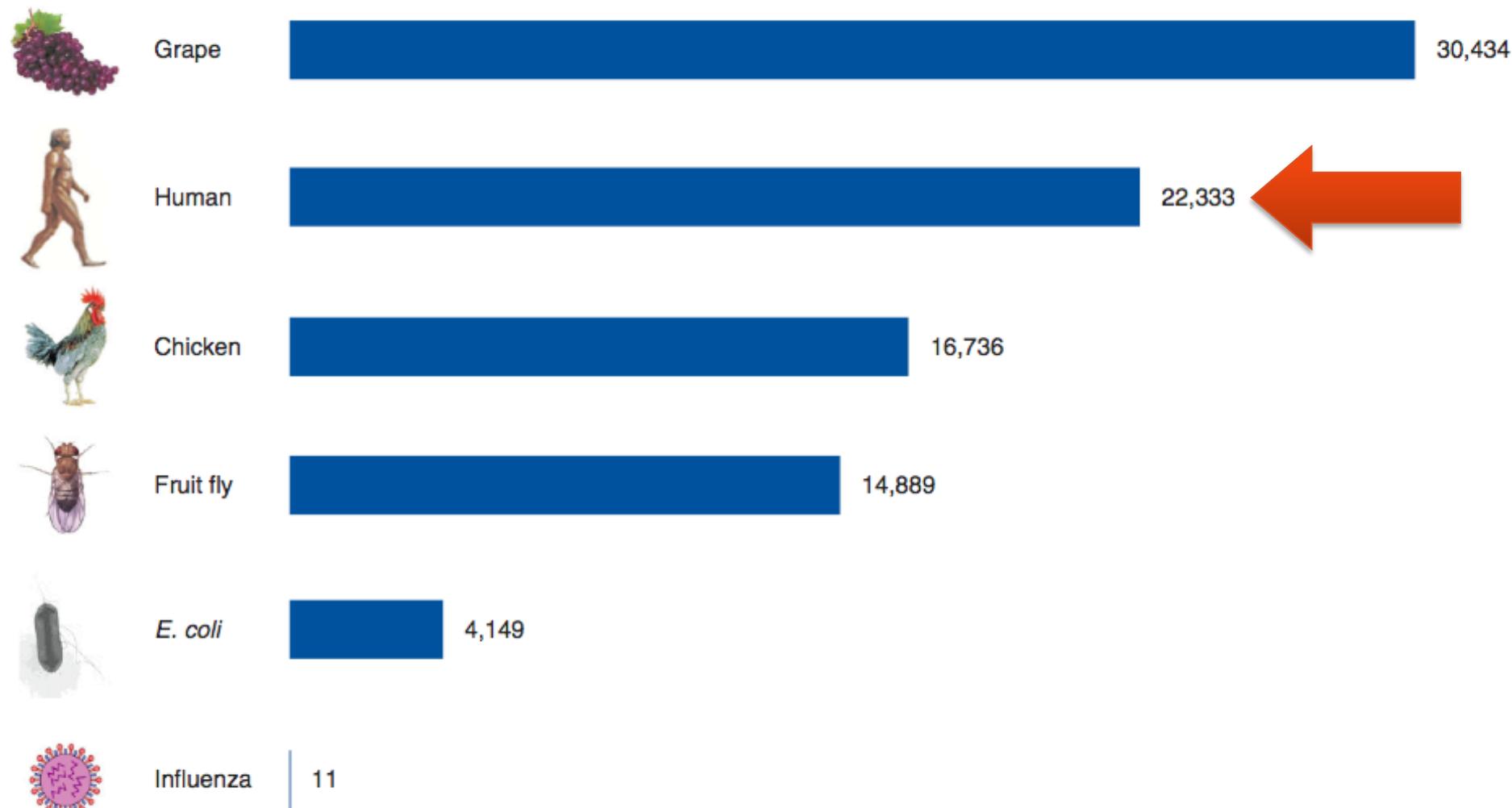
- Chromosomes:
  - Autosomes, 22 pairs
  - Sex-chromosomes, X and Y
  - Mitochondrial chromosome
  - One copy of one (part) of the chromosome(s) of each of your parents
    - 1) chromatid
    - 2) centromere
    - 3) short arm, p(etit)
    - 4) long arm, q



# Somewhere around 22,000 genes?



# The numbers compared



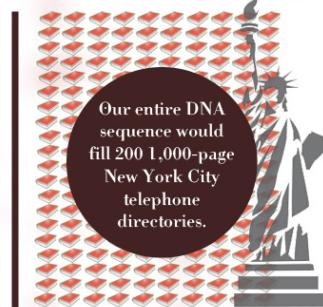
# Human Genome: *some statistics*

- 3.2 billion base pairs in the haploid genome
- $\approx 18,000\text{-}25,000$  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
    - Variable number of tandem repeats (VNTR)
    - Copy-number variations (CNV)
  - Transposable elements
    - Viral or bacterial origins
- (So there's no such thing as "junk DNA"...)

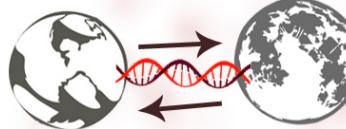
Our entire DNA sequence is called a genome... and there's an estimated **3,000,000,000** DNA bases in our genome.



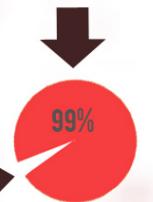
A complete 3 billion base genome would take **3 GIGABYTES OF STORAGE SPACE.**



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 HUMANS'.



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 in to 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.

ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGATCC  
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTTAC  
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCAGCA  
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTACTGCC  
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGACTGCACATATCGTCATACA  
TCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACTTA  
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTATCCTATA  
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGC  
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCACATAT  
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGCATCGTACTGACTGTCTAGTCTAAACACATCCACAT  
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATATCGTCATCGTACTGACTGTCTAGTCTAAACACA  
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGC  
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCACATAT  
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAAACACAT  
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATACATATCGTCATCGTACTGACTGTCTAGTCTAAACACA  
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGC  
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCACATAT  
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAAACACAT  
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATAGCCGATCGTACGACACATATCGTCATCGTACTGCCCT  
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGTACGACACATATCGTCATCGTACTGCCCT  
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGACTGCACATATCGTCATACAT  
CGTACTGACTGTCTAGTCTAAACACATCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCCCATGC  
ATCGTACTGACTGTCTAGTCTAAACACATCCACATATCGTCATCGTACTGACTGTCTATTCTAAACACATCCCAGCAT  
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTACTGCCCT  
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGTACTGACTGCACATATCGTCATACATA  
GTACTGACTGTCTAGTCTAAACACATCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACTTACC  
ATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATC  
CGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCAT

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 99.5% similar between individuals
- 0.5% different → individual point variations
- Single-nucleotide polymorphisms

## 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.

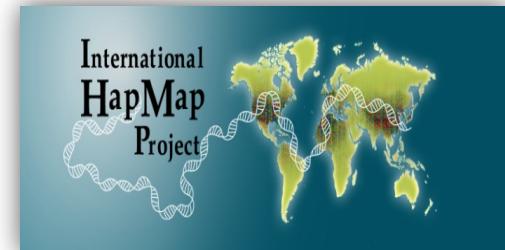
---

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.

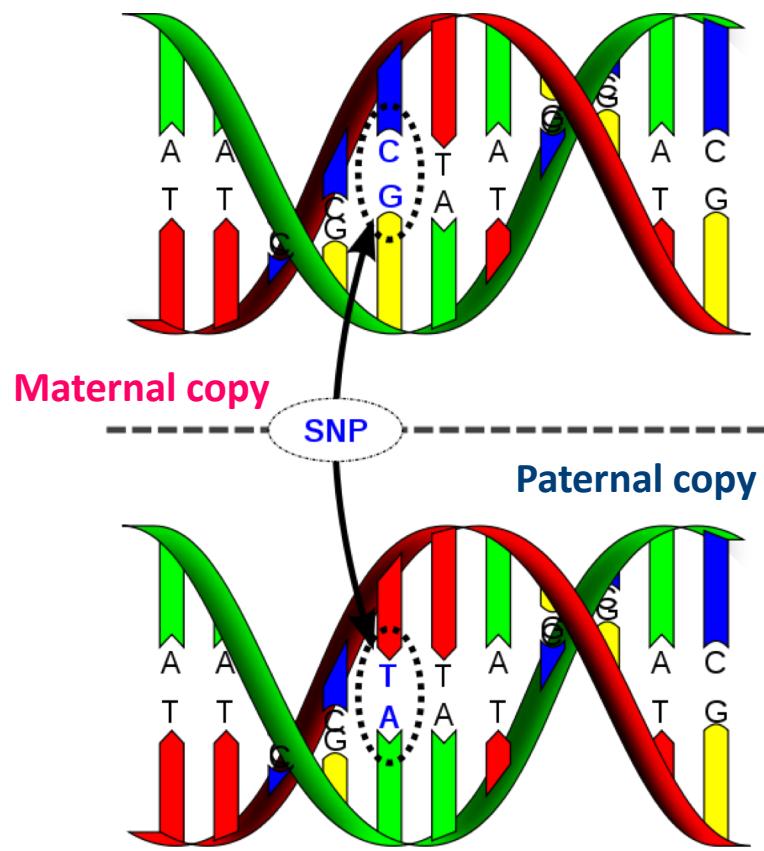
# Single-Nucleotide Polymorphism

*a genetic variation as proxy*

- Single-nucleotide polymorphism (SNP)
- “one base pair variation”
  - > 1% general population (common)
  - ≈10 million SNPs ( $\approx 0.25\%$  genome)
  - Makes you and me unique
- SNPs are common variants which are used as proxies of the actual genomic variation

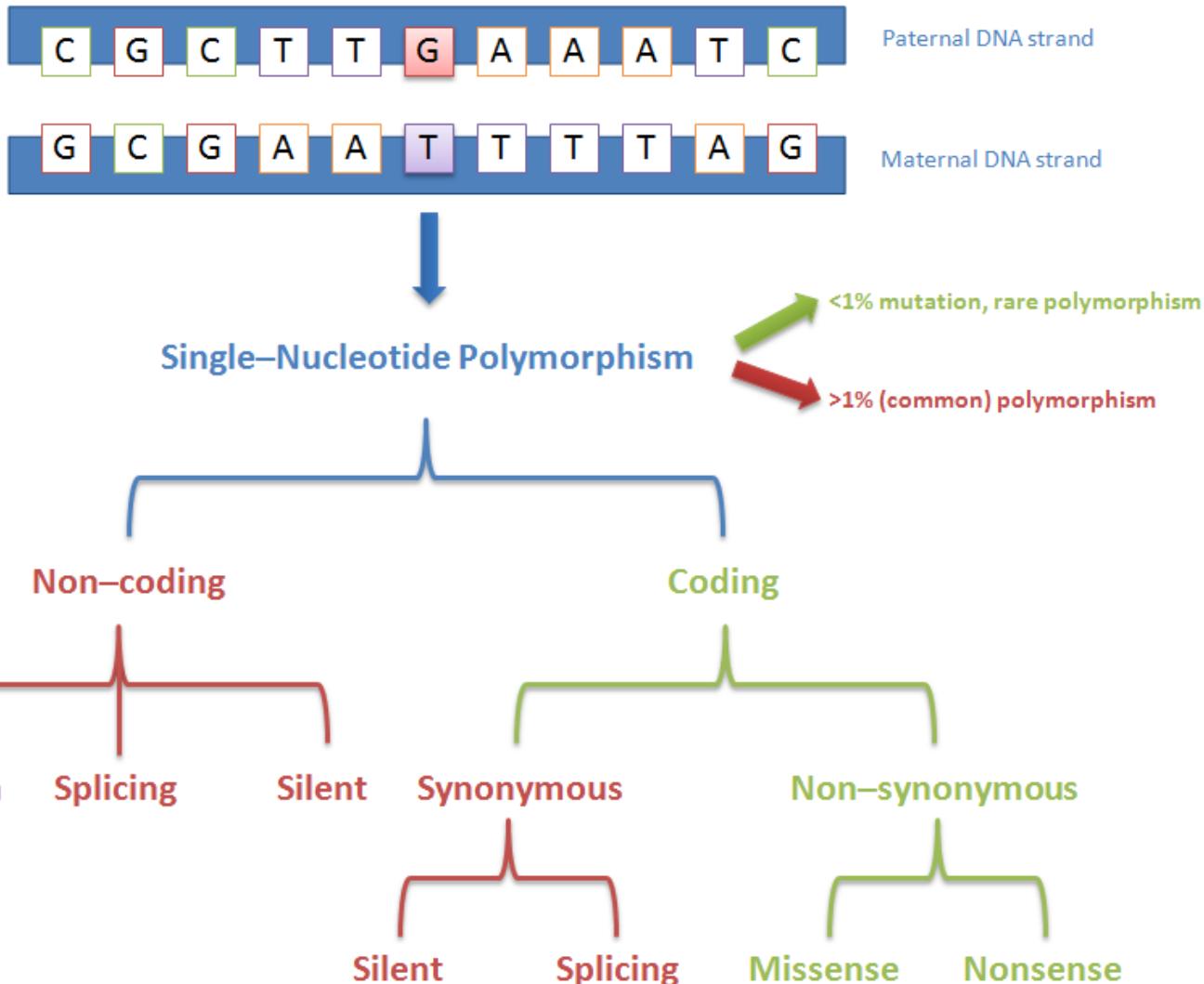


[www.hapmap.org](http://www.hapmap.org)



# Types of SNPs

## Chromosome



# Single base pair changes

## Single Nucleotide Substitutions

		Normal											
		ATG CCG GAC			TCG TTT CTC GGG								
		M	P	D	S	F	L	G					
<b>A</b>	<b>Silent</b> (G>A, Ser to Ser)												
		ATG CCG GAO	TCA		TTT CTC GGG								
		M	P	D	S	F	L	G					
<b>B</b>	<b>Missense</b> (T>C, Ser to Pro)												
		ATG CCG GAC	CCG		TTT CTC GGG								
		M	P	D	P	F	L	G					
<b>C</b>	<b>Nonsense</b> (C>A, Ser to Ter)												
		ATG CCG GAC	TAG		TTT CTC GGG								
		M	P	D	X	F	L	G					

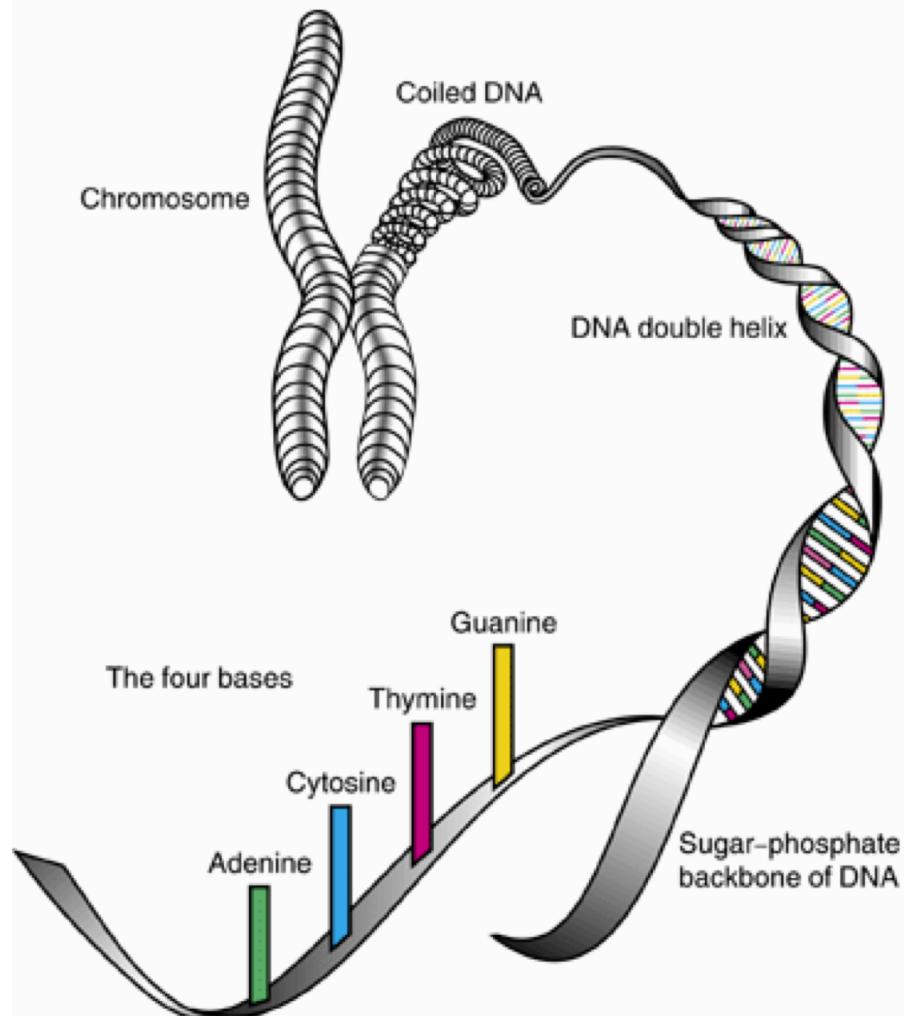
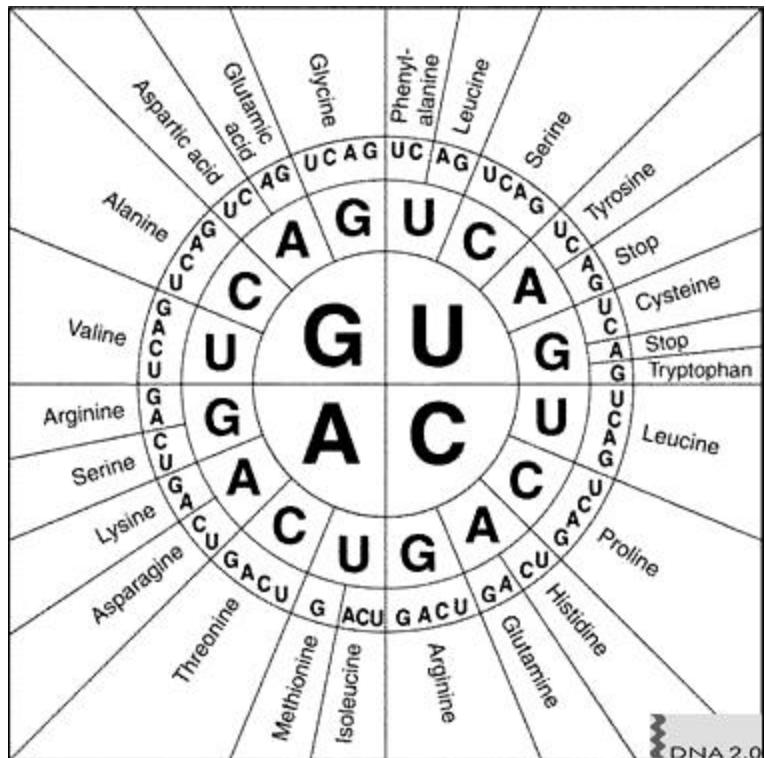
Diagram illustrating single nucleotide substitutions and their effects on protein structure:

- Normal:** ATG CCG GAC (Methionine) → TCG TTT CTC GGG (Serine).
- A Silent:** ATG CCG GAO (Methionine) → TCA (Proline). (G>A, Ser to Ser)
- B Missense:** ATG CCG GAC (Methionine) → CCG (Proline). (T>C, Ser to Pro)
- C Nonsense:** ATG CCG GAC (Methionine) → TAG (Termination). (C>A, Ser to Ter)

Chemical structures of amino acids:

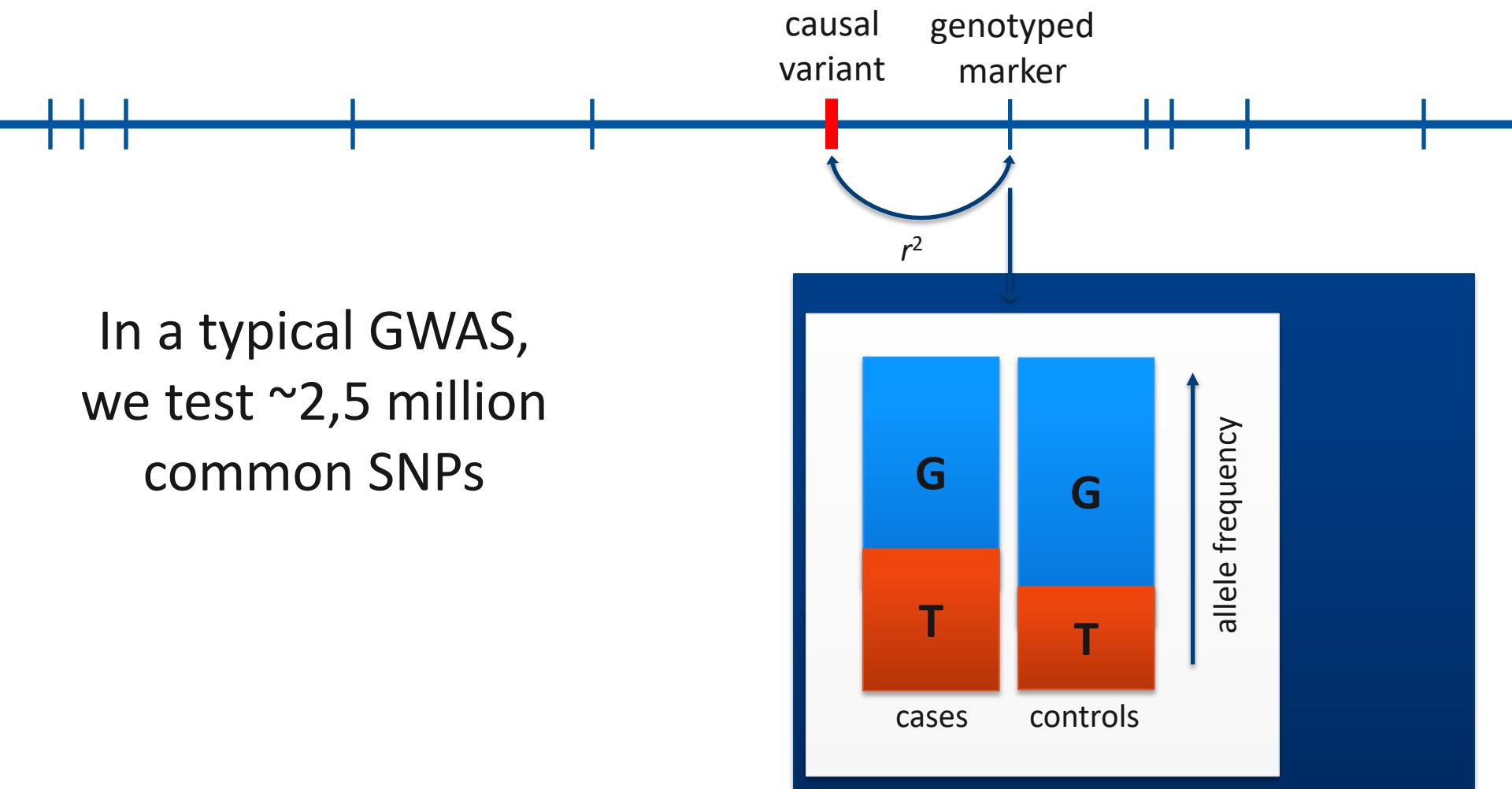
- Methionine (Met, M): CC(C)SSC(=O)N
- Proline (Pro, P): CC1=CSC1
- Aspartic Acid (Asp, D): CC(=O)OCC(=O)N
- Serine (Ser, S): CC(O)C(=O)N
- Phenylalanine (Phe, F): CC(=O)Oc1ccccc1
- Leucine (Leu, L): CC(C)C(=O)N
- Glycine (Gly, G): CH3C(=O)N

# The DNA code

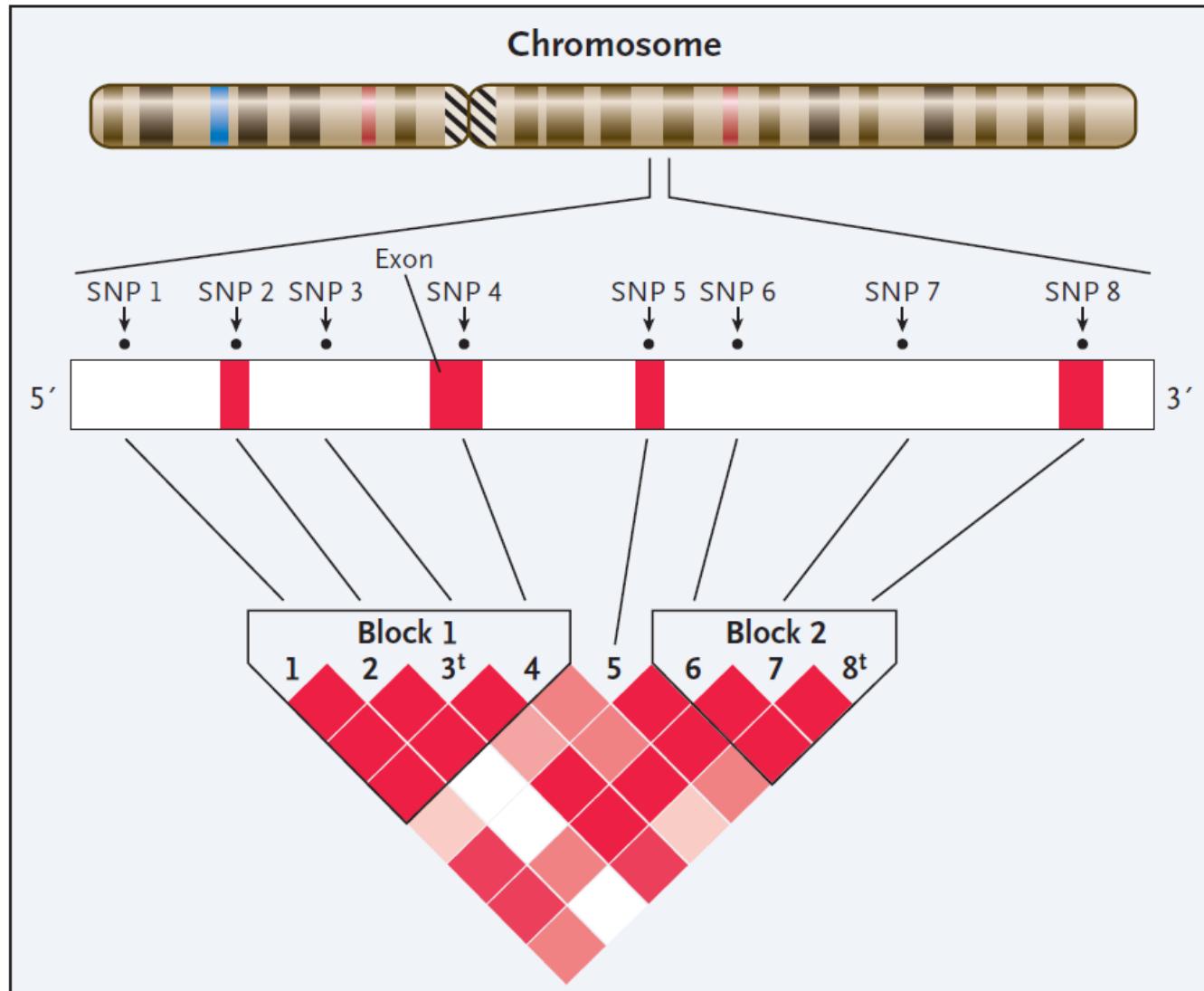


# Linkage disequilibrium

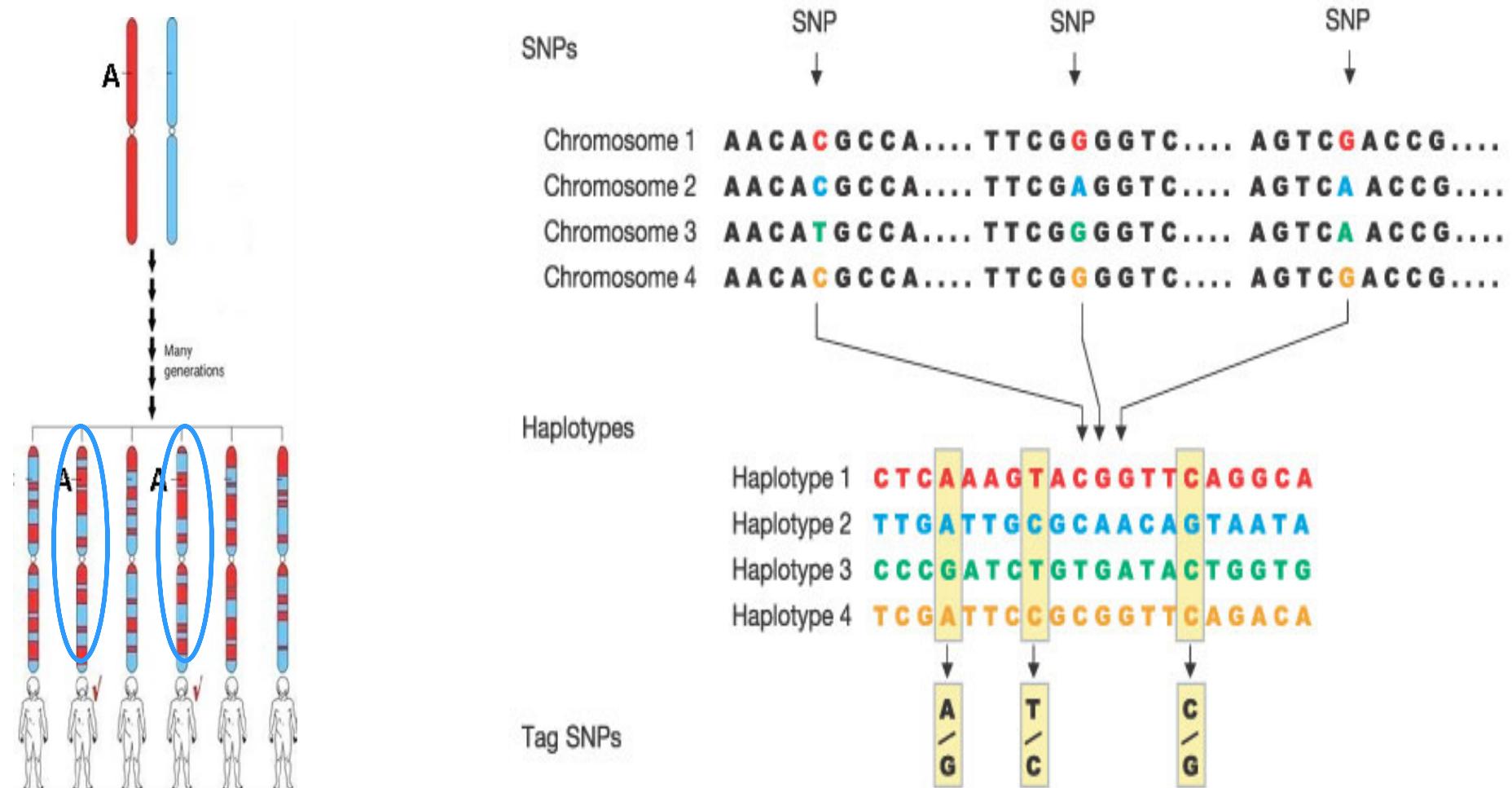
Non-random association of alleles at two or more loci



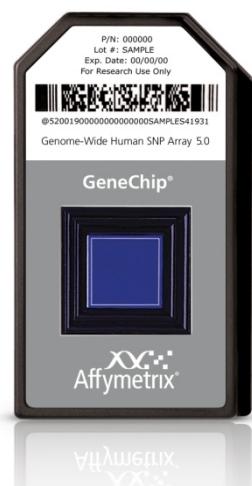
# LD: blocks of correlated SNPs



# SNPs → Haplotype → tagSNP



# Genotyping Platforms



# Catalogs of inherited DNA variation

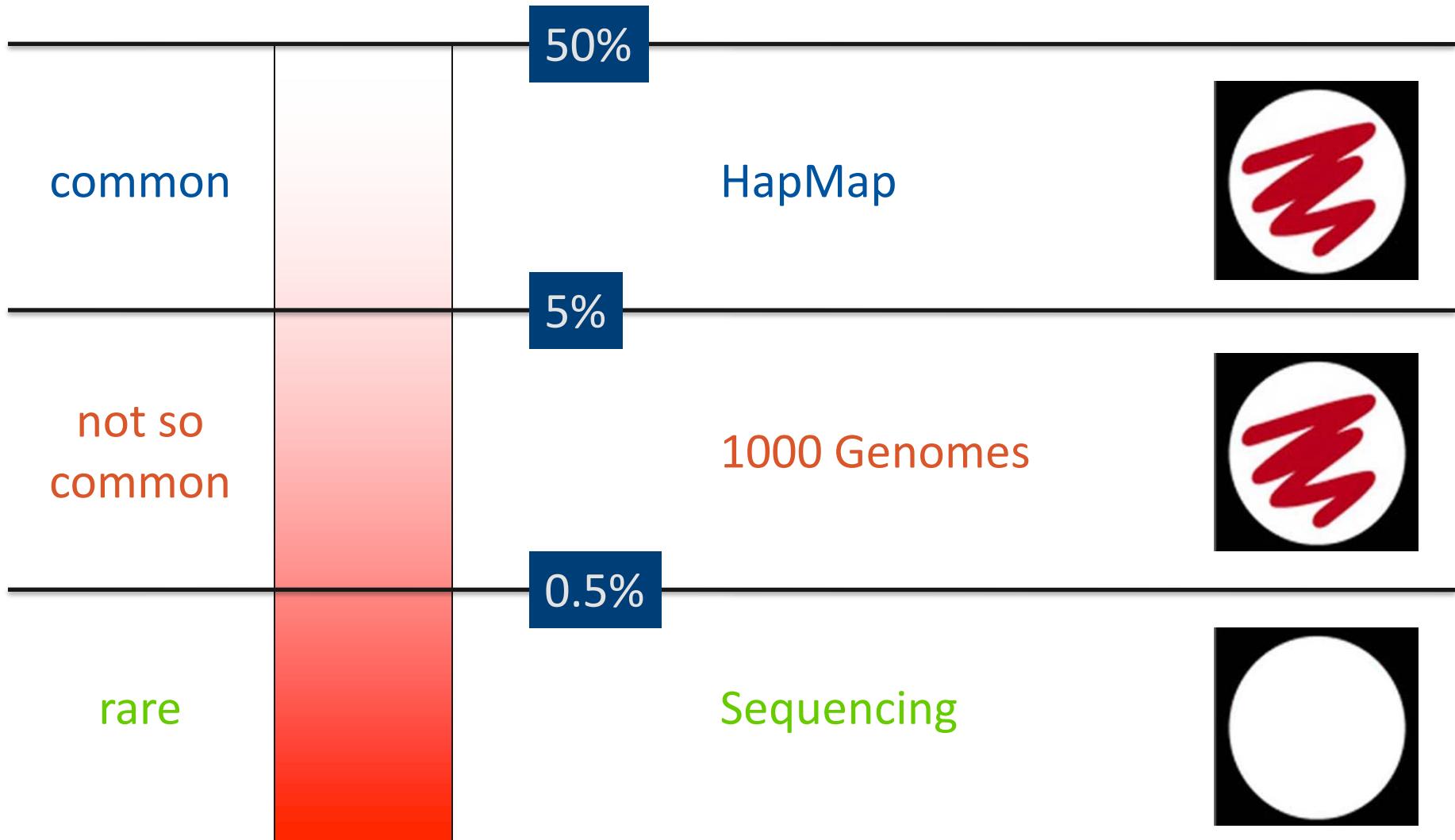
University Medical Center  
Utrecht

- The HapMap project built a genome-wide inventory of 3 million SNPs in 270 individuals from 4 populations
- 1000 Genomes project uses sequencing to identify nearly all variants in 2500 individuals of 15 populations



# So what's been done?

University Medical Center  
Utrecht





University Medical Center  
Utrecht



Recapturing Some Basic Genetics

# MENDELIAN AND COMPLEX DISEASES

# Family history

- Framingham Heart Study | [www.framinghamheartstudy.org](http://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*

# What type of disease are we looking at?

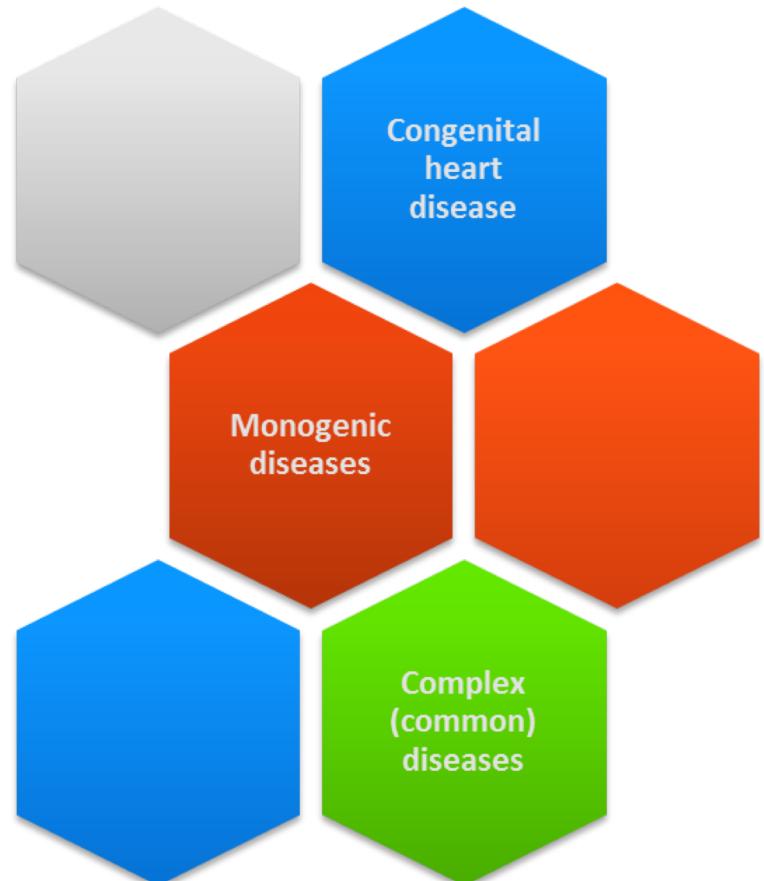
## *monogenic diseases*

- **Congenital heart disease**

- Atrial septal defects
- Ventricular septal defects
- Electrical septal defects

- **Monogenic diseases**

- Mendelian pattern
  - Autosomal dominant, e.g.:
    - Marfan Syndrome
    - Familial hypercholesterolemia
  - Autosomal recessive , e.g.:
    - Sickle cell anemia
    - Cystic fibrosis
  - X-linked , e.g.:
    - Duchene muscular dystrophy
  - Y-linked/Mitochondrial



# Monogenic disease

- Highly *penetrant* alleles are associated with monogenic, Mendelian diseases

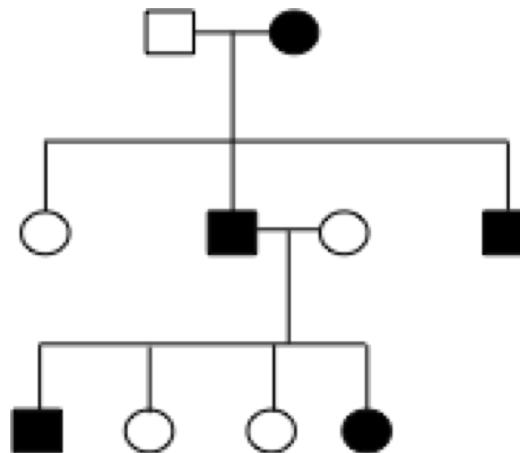
## Genotype



## Disease



Environment



family-based studies

# Family-based linkage analyses

- More than >2,000 single-gene disorders identified:
  - sickle cell anemia
  - cystic fibrosis
  - Huntington's disease
  - muscular dystrophy
  - rare forms of many common diseases, including breast cancer (BRCA genes), diabetes (MODY), hypertension
- Online Mendelian Inheritance in Man (OMIM)
  - <http://www.ncbi.nlm.nih.gov/omim>

# What type of disease are we looking at?

## *complex diseases*

- Complex diseases

  - Polygenic, multifactorial diseases

    - Diabetes mellitus
    - Asthma
    - Cardiovascular disease
    - Hypertension

*Each gene contributes a little to the disease*

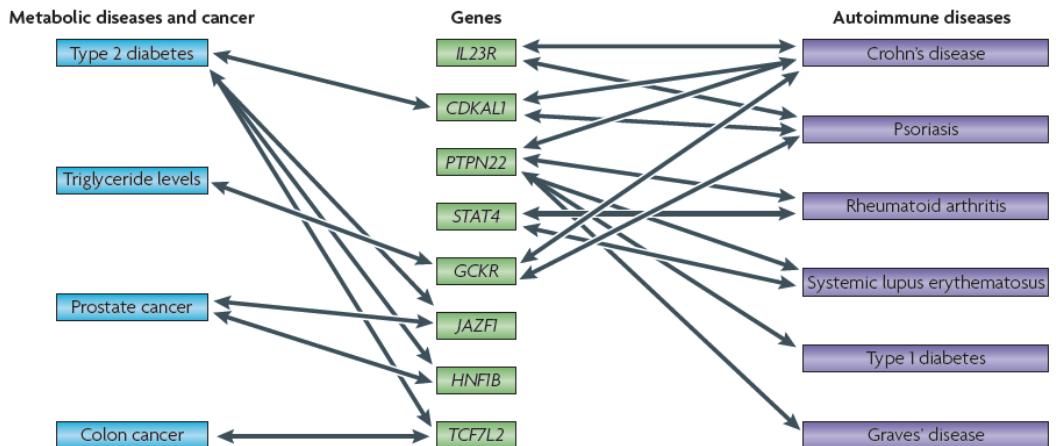


Table 3

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
Lymphotoxin $\alpha$	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.

Monogenic

Genotype → Disease

Polygenic

Genotypes → Disease

Environment



# Challenges of common disease research

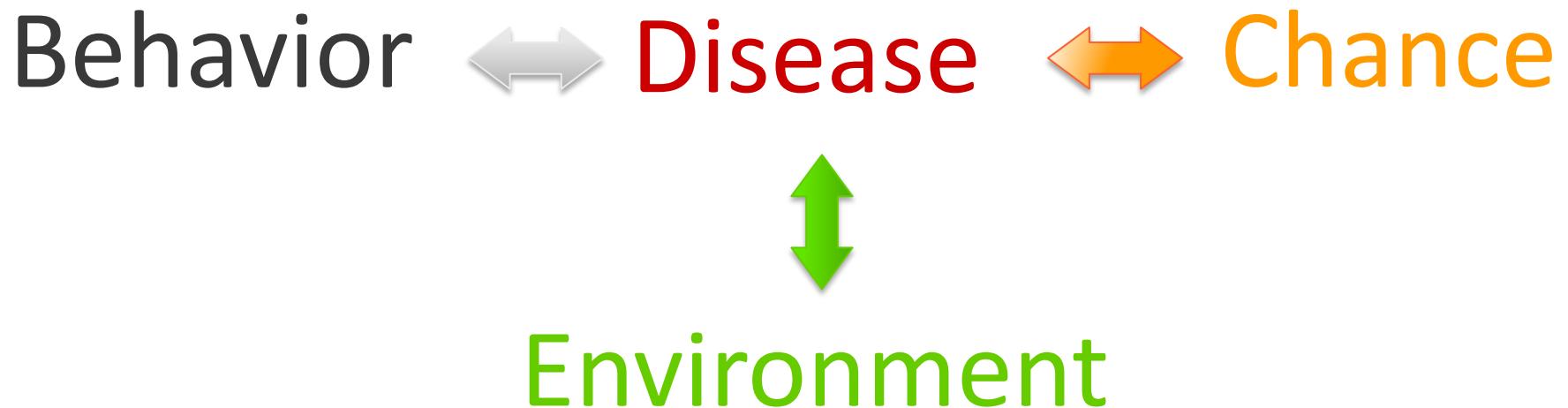
- Involvement of many genes and environment
  - Each plays only a small role (consequence - traditional approaches designed to find major genetic effects are ineffective – no such effects exist!)
  - Mutations are neither necessary nor sufficient on their own to cause disease
  - Environmental triggers are common
- Inability to recognize DNA changes that have subtle functional impact from majority of variation (likely neutral)

# Many factors influence complex traits and common disease



University Medical Center  
Utrecht

Genotype



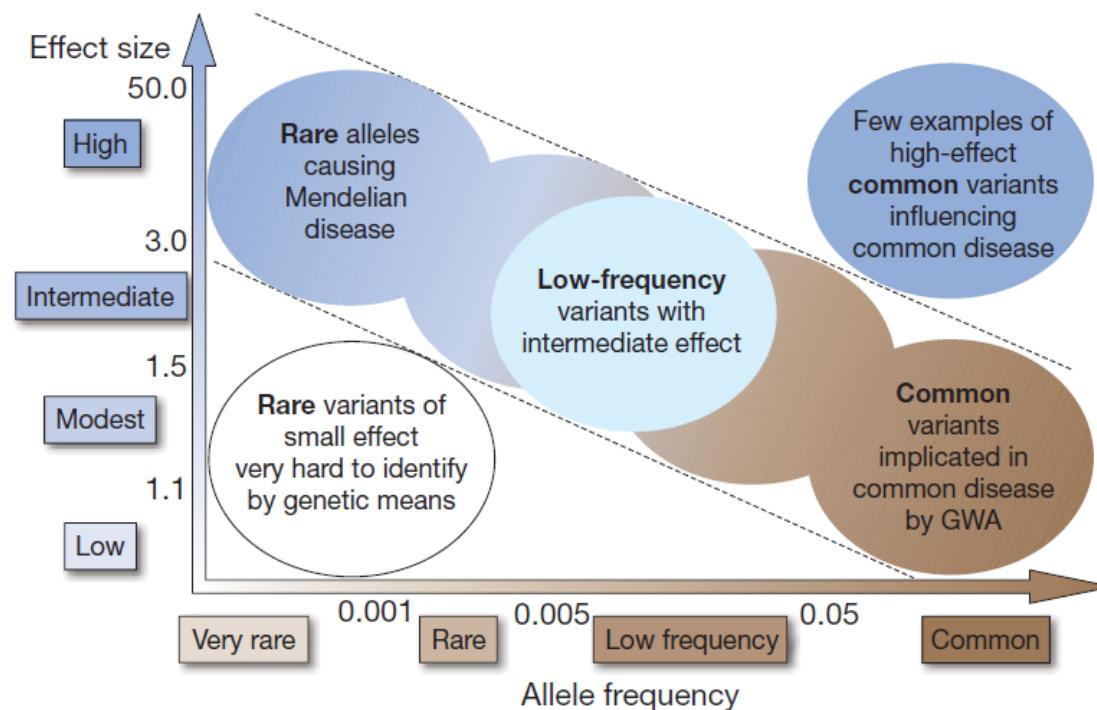
# Can we study every gene and variation in the human genome?

- Genome-wide association studies (GWAS)
  - Specific genotyping of known SNPs (HapMap)
  - Current platforms (Illumina, Affymetrix) allow genotyping of <2.5 million SNPs per DNA sample
  - Many associations identified to date for range of complex traits and diseases
- Complete sequencing is truly comprehensive
  - >10 million variants with frequency above 1%
  - Many, many more with frequency below 1%
  - But still expensive
- Whole-exome is being done in the interim period

# Common Disease, Common Variant

## *CDCV, Effect size, and Allele frequency*

- “Common Disease, Common Variant” (CDCV) hypothesis:
  - Common variants (SNPs) underlie common diseases/traits (atherosclerotic disease)
  - Why? Evolution: natural selection, fitness & genetic drift
- Effect size vs. Allele frequency
  - Low to intermediate penetrance
  - Low to intermediate **odds ratio (OR) 1.1-1.5**
  - Higher penetrance results in decreased reproductive fitness (unlikely in common diseases!)



# 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 Helgadottir<sup>1</sup>, Andrei Manolescu<sup>1</sup>, Gudmar Thorleifsson<sup>1</sup>, Solveig Gretarsdottir<sup>1</sup>, Helga Jonsdottir<sup>1</sup>, Unnur Thorsteinsdottir<sup>1</sup>, Nilesh J Samani<sup>2</sup>, Gudmundur Guðmundsson<sup>1</sup>, Struan F A Grant<sup>1</sup>, Gudmundur Thorgeirsson<sup>3</sup>, Sigurlaug Sveinbjornsdottir<sup>3</sup>, Einar M Valdimarsson<sup>4</sup>, Stefan E Matthiasson<sup>3</sup>, Halldor Johannsson<sup>3</sup>, Olof Guðmundsdottir<sup>1</sup>, Mark E Gurney<sup>1</sup>, Jesus Sainz<sup>1</sup>, Margret Thorhallsdottir<sup>1</sup>, Margaret Andressdottir<sup>1</sup>, Michael L Frigge<sup>1</sup>, Eric J Topol<sup>4</sup>, Augustine Kong<sup>1</sup>, Vilimundur Gudnason<sup>5</sup>, Hakon Hakonarson<sup>1</sup>, Jeffrey R Gulcher<sup>1</sup> & Kari Stefansson<sup>1</sup>

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

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

# Wellcome Trust Case-Control Consortium

University Medical Center  
Utrecht

- 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

**9p21**

- 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,<sup>1,\*</sup> Gudmar Thorleifsson,<sup>1,\*</sup> Andrei Manolescu,<sup>1,\*</sup> Solveig Gretarsdottir,<sup>1</sup> Thorarinn Blöndal,<sup>1</sup> Áslaug Jónasdóttir,<sup>1</sup> Adalbjørg Jónasdóttir,<sup>1</sup> Asgeir Sigurdsson,<sup>1</sup> Adam Baker,<sup>1</sup> Amar Palsson,<sup>1</sup> Gisli Masson,<sup>1</sup> Daniel F. Gudbjartsson,<sup>1</sup> Kristinn P. Magnusson,<sup>1</sup> Karl Andersen,<sup>2</sup> Allan I. Levey,<sup>3</sup> Valgerður M. Backman,<sup>1</sup> Sigurborg Matthiassdóttir,<sup>1</sup> Thorbjørg Jónsdóttir,<sup>1</sup> Stefan Palsson,<sup>1</sup> Helga Einarsdóttir,<sup>1</sup> Steinunn Gunnarsdóttir,<sup>1</sup> Amaldur Gylfason,<sup>1</sup> Viola Vaccarino,<sup>3</sup> W. Craig Hooper,<sup>3</sup> Muredach P. Reilly,<sup>4</sup> Christopher B. Granger,<sup>5</sup> Harland Austin,<sup>3</sup> Daniel J. Rader,<sup>4</sup> Svti H. Shah,<sup>5</sup> Arshed A. Quyyumi,<sup>3</sup> Jeffrey R. Gulcher,<sup>1</sup> Guðmundur Þorgerðsson,<sup>2</sup> Unnur Thorsteinsdóttir,<sup>1</sup> Augustine Kong,<sup>1,†</sup> Kari Stefansson<sup>1</sup>

## A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

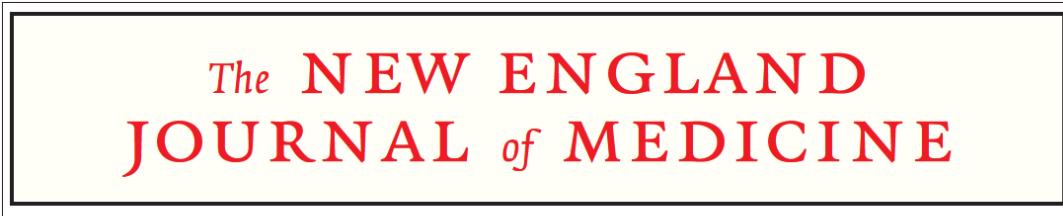
Ruth McPherson,<sup>1,\*†</sup> Alexander Pertsemlidis,<sup>2,\*</sup> Nihan Kavaslar,<sup>1</sup> Alexandre Stewart,<sup>1</sup> Robert Roberts,<sup>1</sup> David R. Cox,<sup>3</sup> David A. Hinds,<sup>3</sup> Len A. Pennacchio,<sup>4,5</sup> Anne Tybjaerg-Hansen,<sup>6</sup> Aaron R. Folsom,<sup>7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

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

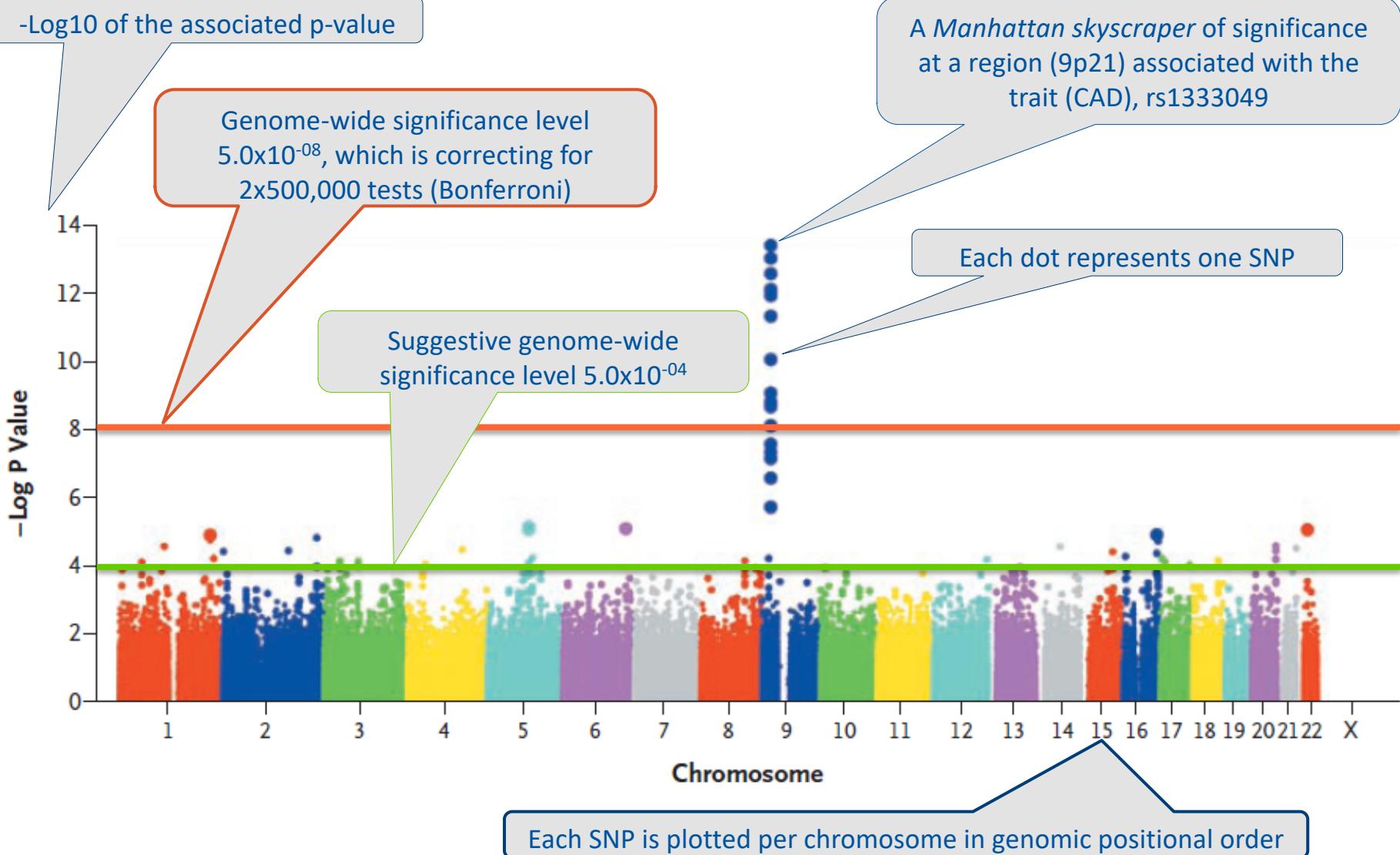
ARTICLES

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

The Wellcome Trust Case Control Consortium\*

# One famous example

## 9p21 in the WTCCC

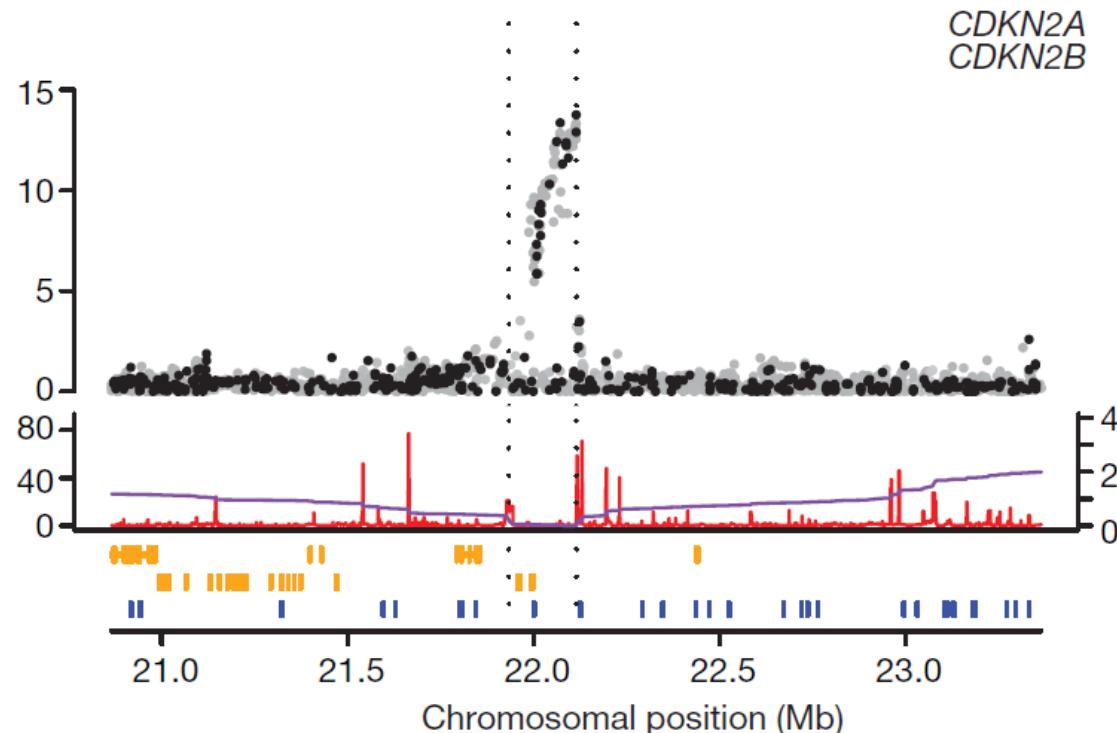


# One famous example

## 9p21 a closer look

- The SNPs 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

CAD hit region, chromosome 9



# 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	$1.79 \times 10^{-14}$	Standard analysis	$1.16 \times 10^{-13}$	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

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

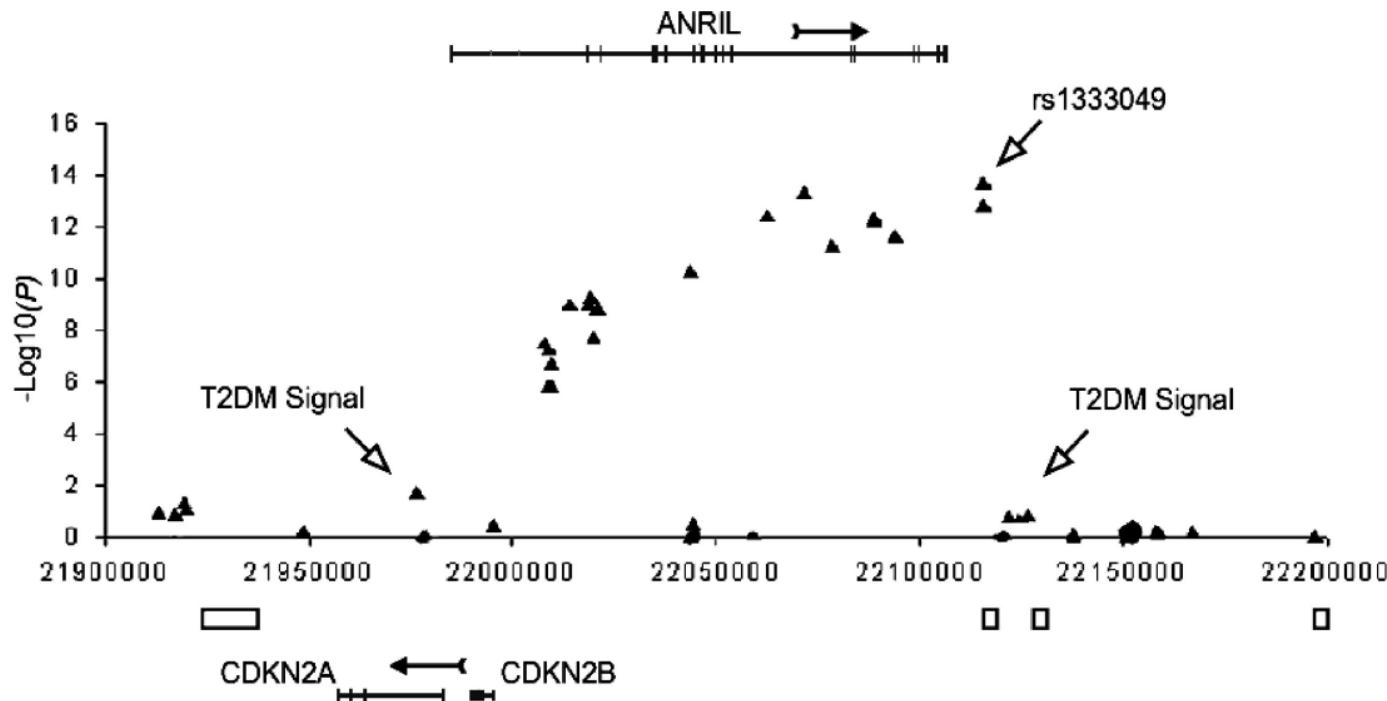
Risk allele: minor allele  
 Odds ratio: the odds of exposure between cases and controls

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

# 9p21 points to a RNA gene

*how does this explain an acute phenomenon like MI?*

- 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)



# 9p21 is used in a laboratory DNA test

- deCODE Genetics' deCODE MI™
- Assessment of the risk for (early-onset) myocardial infarction
  - SNPs rs133049 and rs10757278 located in vicinity to *CDKN2A* and *CDKN2B*
  - Risk allele is *independent* and *additive* to traditional risk factors (Framingham Heart Score, Reynold's score, AIRIC score)
  - 20-22% of the general population carry the risk allele
  - ≥40% in patients suffering early MI (male < 50 years, female < 60 years)
  - Carrying two copies of the risk allele correspond to an approximate *1.6 fold increase over the general population* of early onset MI and a *1.3 fold risk of MI in general*
- Tested according to CLIA, but not FDA approved
- Collaboration with clinicians/clinical geneticists
- Clear and concise report

# 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
- Recent studies uncovered around 30 variants associated with MI/CAD
- Sample size greatly influences power and effect size to discover new variants
- CARDIoGRAMplusC4D sought to solves this issue
- 15 novel susceptibility loci for CAD were discovered

© 2012 Nature America, Inc. All rights reserved.



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

The CARDIoGRAMplusC4D Consortium<sup>1</sup>

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 to 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.

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 factor<sup>1</sup>. 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<sup>3,4</sup>. To date, genome-wide association studies (GWAS) have collectively reported a total of 31 loci, associated with CAD risk at genome-wide significance ( $P < 5 \times 10^{-8}$ )<sup>5–13</sup>. 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 effects<sup>14</sup>. 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) Consortium<sup>5</sup>, 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 array<sup>15</sup>, 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 array<sup>15</sup>). 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<sup>16–20</sup>. 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

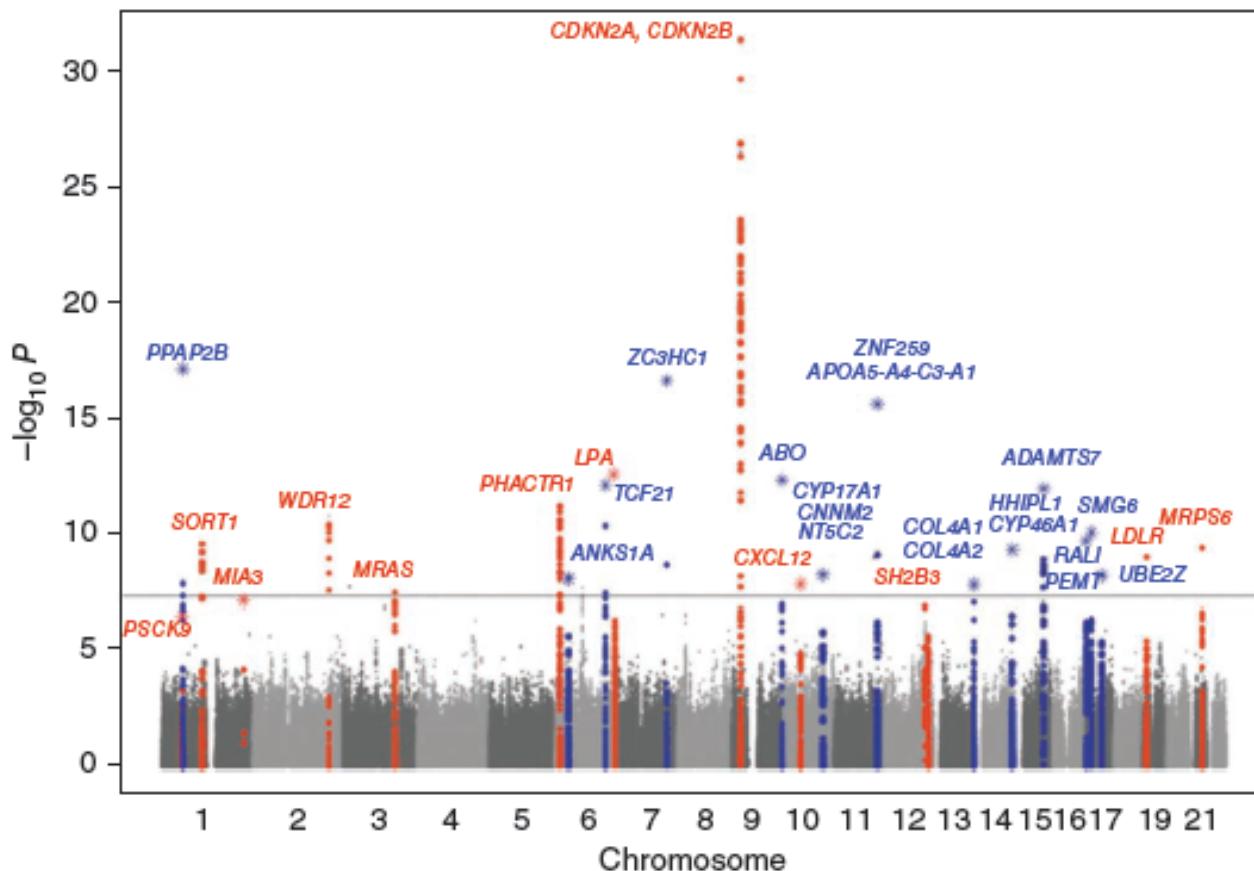
We expanded the CARDIoGRAM discovery data set (22,233 cases and 64,762 controls<sup>5</sup>; stage 1) with 34 additional CAD sample collections (stage 2) of European or south Asian descent comprising 41,513 cases and 65,919 controls (study descriptions and sample characteristics are given in **Supplementary Tables 1a** and **2a**, respectively) and undertook a 2-stage meta-analysis to test SNPs on the Metabochip array

<sup>1</sup>A 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

# Replication & Discovery

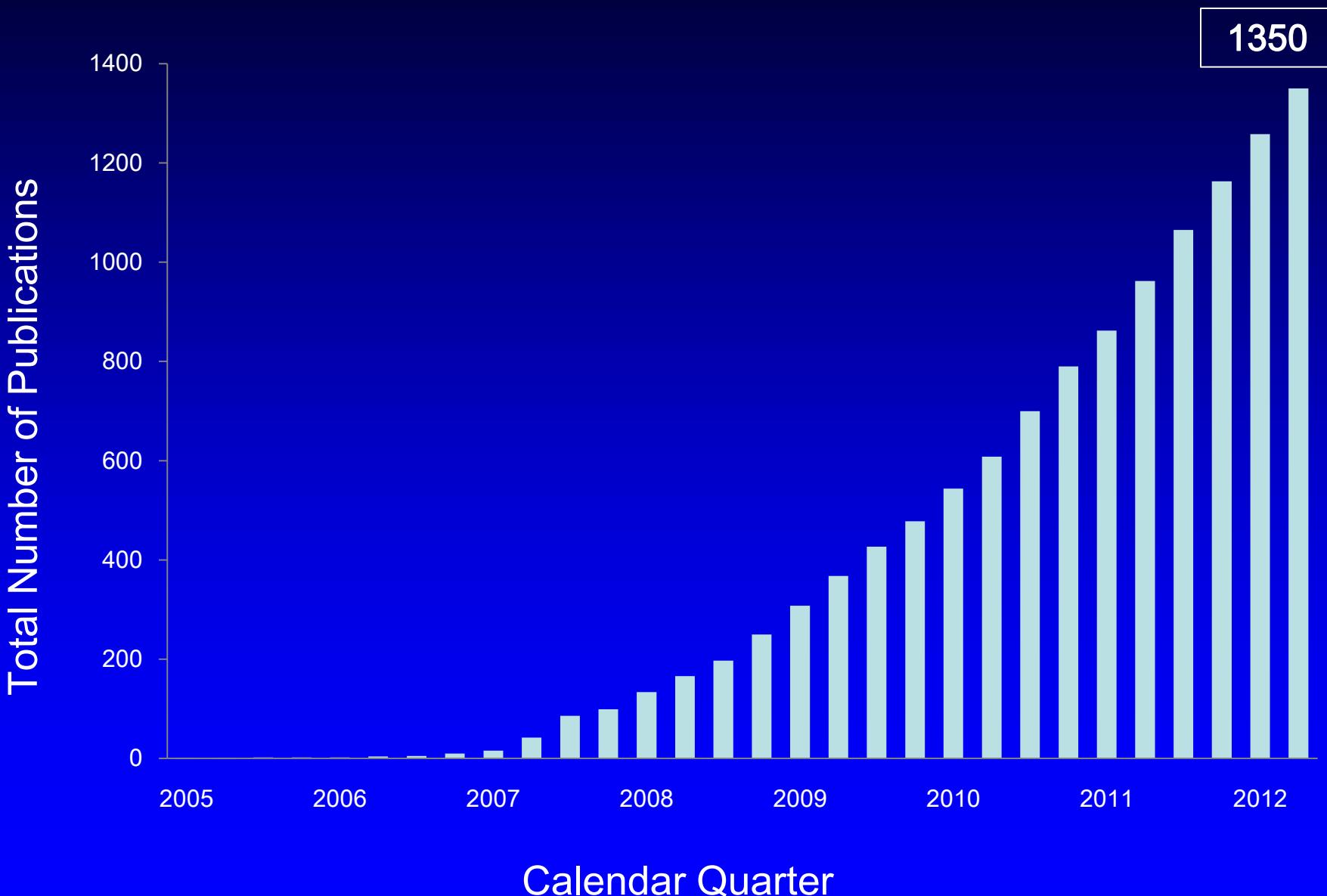
- 30 previously associated loci could be replicated
- 15 novel loci were uncovered, some of which are also associated with
  - Carotid plaque presence
  - Ischemic large artery (=carotid) stroke



# Success of GWAS

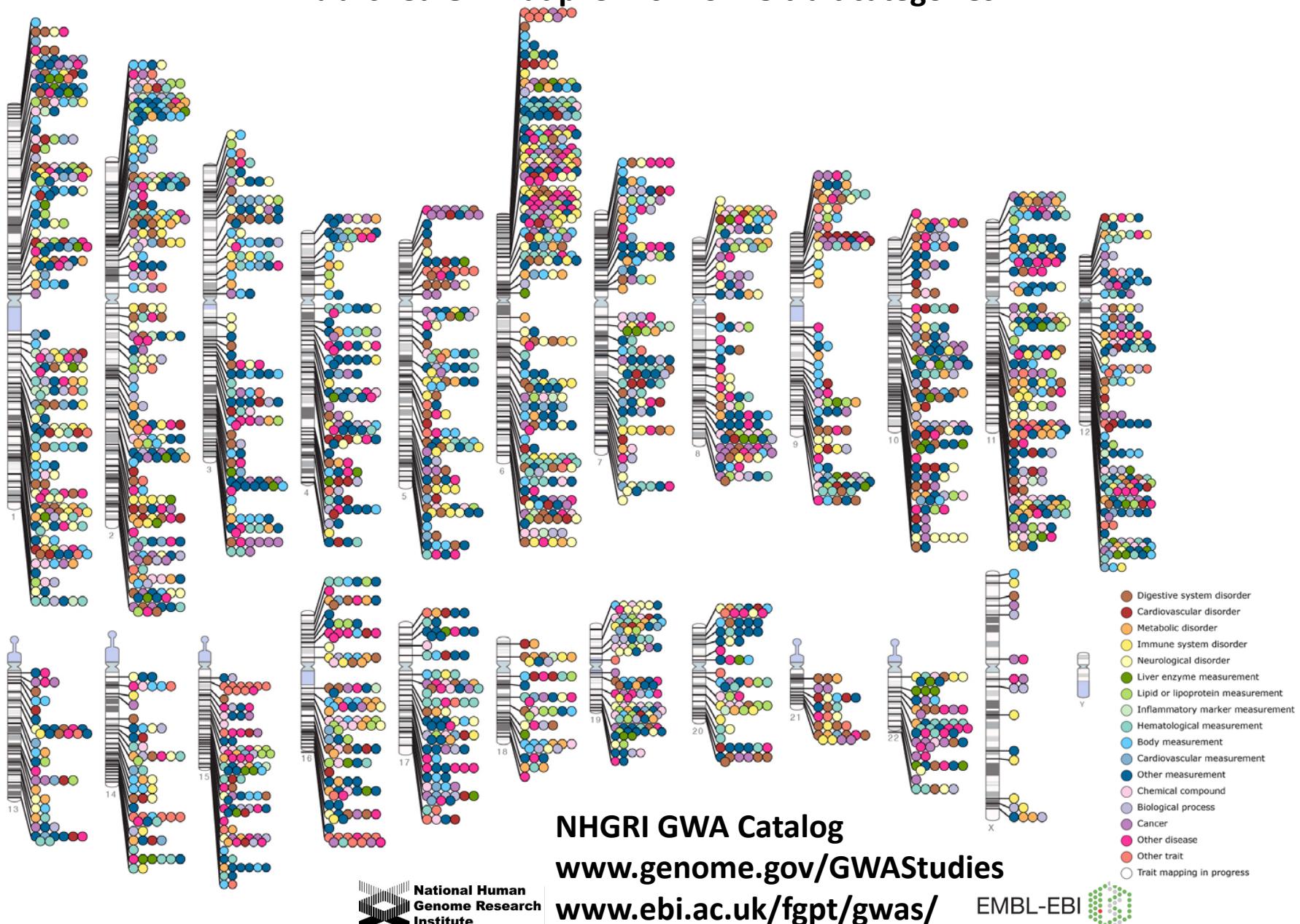
- Hundreds of variants associated with hundreds of traits and diseases
  - Since 2005 (when the “final” version of the human genome was made public)
- GWAS of the two major CVD:
  - Coronary artery disease (including MI)
  - Ischemic stroke
- GWAS of essentially all traditional risk factors of CVD have uncovered hundreds of variants associated with these risk factors
  - Type 2 diabetes
  - Hypertension & blood pressures (SBP, DBP, MAP, PP)
  - Glucose levels
  - CRP levels
  - Kidney function (eGFR) and chronic kidney disease
  - Smoking (!) → interestingly no overlap with any of the other risk factors or CVD
  - ...

# Published GWA Reports, 2005 – 6/2012



# Published Genome-Wide Associations through 07/2012

Published GWA at  $p \leq 5 \times 10^{-8}$  for 18 trait categories





University Medical Center  
Utrecht

Genetics, Biomarkers & Disease

# CLINICAL UTILITIES



# deCODE MI™

*provides a modified 10 year CHD risk*

<b>deCODE MI™</b> Genetic risk test for Heart Attack	NAME Jane Doe	DOB 10-13-1968	GENDER Female	Patient ID n/a	deCODE ID DGMIW#8570	5/5
---	------------------	-------------------	------------------	-------------------	-------------------------	-----

---

**FRAMINGHAM RISK SCORING ALGORITHMS FEMALE SPECIFIC**

HDL - Cholesterol		
(mg/dL)	(mmol/L)	Points
< 35	≤ 0.90	5
35-44	0.91-1.16	2
45-49	1.17-1.29	1
50-59	1.30-1.55	0
≥ 60	≥ 1.56	-2

LDL - Cholesterol		
(mg/dL)	(mmol/L)	Points
< 100	≤ 2.59	-2
100-129	2.60-3.36	0
130-159	3.37-4.14	0
160-189	4.15-4.91	2
≥ 190	≥ 4.92	-2

Blood Pressure						
		Diastolic (mmHg)				
		< 80	80 - 84	85 - 89	90 - 99	≥100
(mmHg)		< 120	-3	0	1	2
		120-129	0			
		130-139		1		
		140-159			2	
		≥ 160				3

Note: When systolic and diastolic pressure provide different estimates for point scores, use the higher number.

Key		Adding up the points	
Color	Risk	Age	
Green	Very low	LDL Cholesterol	
White	Low	HDL Cholesterol	
Yellow	Moderate	Blood Pressure	
Rose	High	Diabetes	
Red	Verv high	Smoker	

CHD Traditional Risk & Your Modified Risk				
Points Total	10 Year CHD Risk Traditional	Reclassified MI Risk Factor	10 Year CHD Risk Modified	
-2	=	≤1 %	x	2.35 = ≤2.4 %
-1	=	2 %	x	2.35 = 4.7 %
0	=	2 %	x	2.35 = 4.7 %
1	=	2 %	x	2.35 = 4.7 %
2	=	3 %	x	2.35 = 7.1 %
3	=	3 %	x	2.35 = 7.1 %
4	=	4 %	x	2.35 = 9.4 %
5	=	5 %	x	2.35 = 11.8 %
6	=	6 %	x	2.35 = 14.1 %
7	=	7 %	x	2.35 = 16.4 %
8	=	8 %	x	2.35 = 18.8 %
9	=	9 %	x	2.35 = 21.2 %
10	=	11 %	x	2.35 = 25.0 %
11	=	13 %	x	2.35 = 30.6 %
12	=	15 %	x	2.35 = 35.3 %
13	=	17 %	x	2.35 = 40.0 %
14	=	20 %	x	2.35 = 47.0 %
15	=	24 %	x	2.35 = 56.4 %
16	=	27 %	x	2.35 = 63.5 %
217	=	≥32 %	x	2.35 = ≥75.2 %

In the line with your Points Total, you will find your 10 CHD Risk in the column to the right: 10 Year CHD Risk Modified. Enter YOUR 10 Year MODIFIED CHD Risk on the appropriate age group line in the Comparative Risk table to see how your risk compares to the average and low 10 year CHD risk.

"Low risk was calculated for a woman the same age, normal blood pressure, LDL cholesterol 100-129 mg/dL, HDL cholesterol 45 mg/dL, non-smoker and no diabetes.

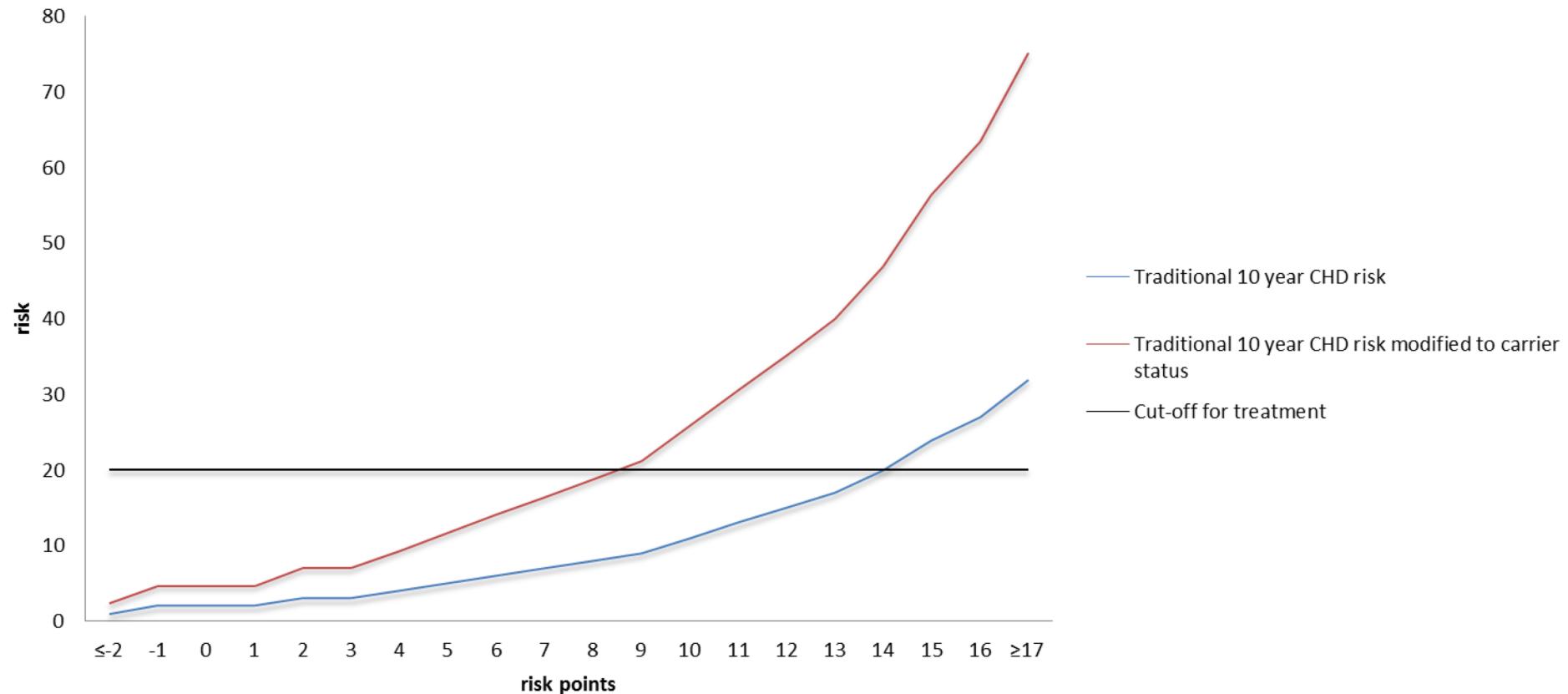
## CHD Traditional Risk & Your Modified Risk

Points Total	10 Year CHD Risk Traditional	Reclassified MI Risk Factor	10 Year CHD Risk Modified	
≤-2	=	≤1 %	x	2.35 = ≤2.4 %
-1	=	2 %	x	2.35 = 4.7 %
0	=	2 %	x	2.35 = 4.7 %
1	=	2 %	x	2.35 = 4.7 %
2	=	3 %	x	2.35 = 7.1 %
3	=	3 %	x	2.35 = 7.1 %
4	=	4 %	x	2.35 = 9.4 %
5	=	5 %	x	2.35 = 11.8 %
6	=	6 %	x	2.35 = 14.1 %
7	=	7 %	x	2.35 = 16.4 %
8	=	8 %	x	2.35 = 18.8 %
9	=	9 %	x	2.35 = 21.2 %
10	=	11 %	x	2.35 = 25.0 %
11	=	13 %	x	2.35 = 30.6 %
12	=	15 %	x	2.35 = 35.3 %
13	=	17 %	x	2.35 = 40.0 %
14	=	20 %	x	2.35 = 47.0 %
15	=	24 %	x	2.35 = 56.4 %
16	=	27 %	x	2.35 = 63.5 %
≥17	=	≥32 %	x	2.35 = ≥75.2 %

Cut-off, at which point some form of (drug) therapy starts

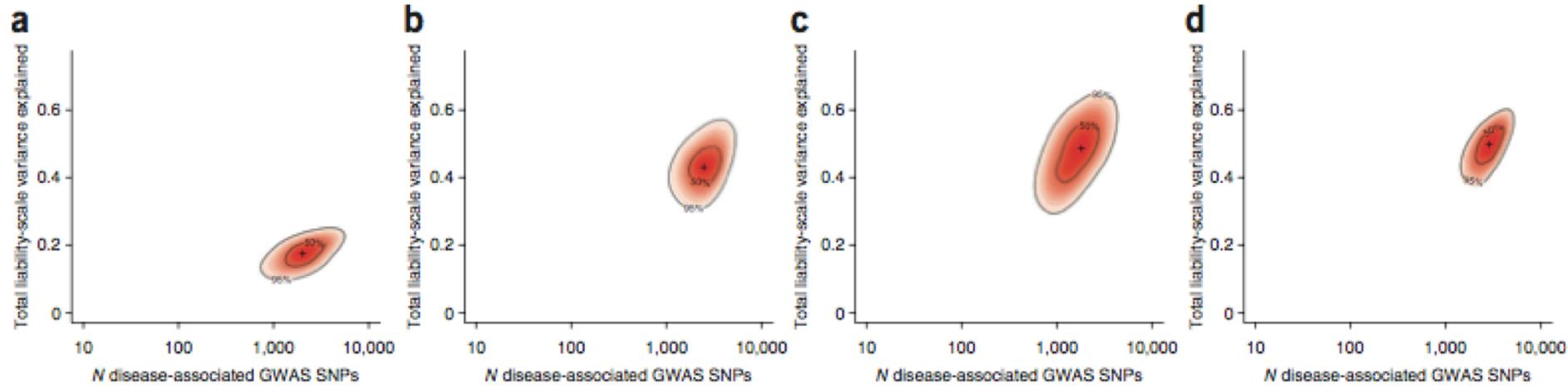
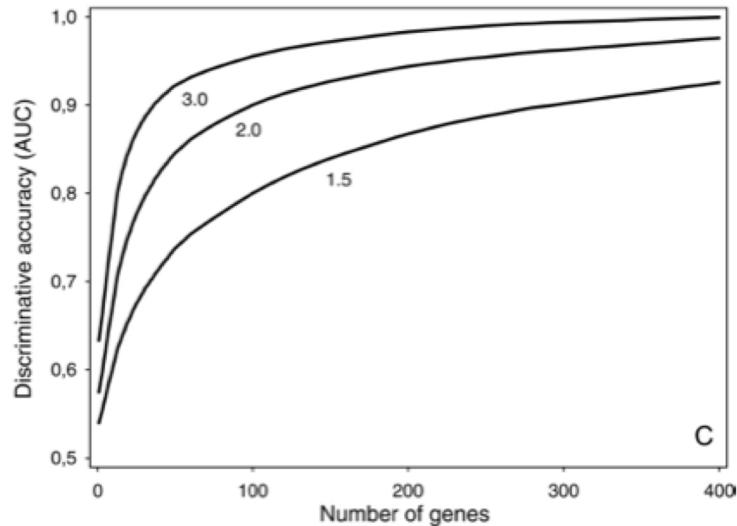
# From 9 to more than 20% risk

## 10 year CHD risk modified by carrier status



# The added value of Genetic Burden Scores?

- Genetic Burden Scores in addition to existing risk factor models
- Hundreds of variants are needed for any meaningful addition
- Upside: many variants remain to be discovered



# Proprotein convertase subtilisin/kexin type 9 PCSK9

- Variants in PCSK9 associated with low LDL and lower risk for CHD
  - Cure?

## 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.

### ABSTRACT

#### BACKGROUND

From the Donald W. Reynolds Cardiovascular Clinical Research Center (J.C.C., H.H.H.), the Center for Human Nutrition (J.C.C.), the Departments of Internal Medicine (J.C.C., H.H.H.) and Molecular Genetics (H.H.H.), and the Howard Hughes Medical Institute (H.H.H.), University of Texas Southwestern Medical Center, Dallas; the Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston (E.B.); and the Department of Medicine, University of Mississippi Medical Center, Jackson (T.H.M.). Address reprint requests to Dr. Hobbs at the Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas TX 75390-9046, or at helen.hobbs@utsouthwestern.edu.

#### METHODS

We compared the incidence of CHD (myocardial infarction, fatal CHD, or coronary revascularization) over a 15-year interval in the Atherosclerosis Risk in Communities study according to the presence or absence of sequence variants in the proprotein convertase subtilisin/kexin type 9 serine protease gene (*PCSK9*) that are associated with reduced plasma levels of LDL cholesterol.

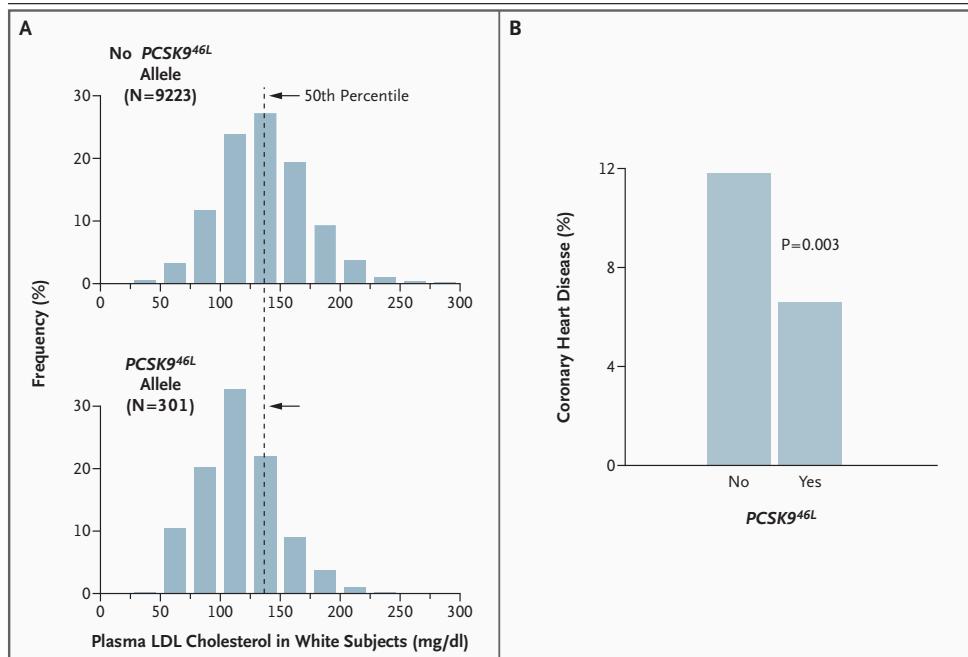
#### RESULTS

Of the 3363 black subjects examined, 2.6 percent had nonsense mutations in *PCSK9*; these mutations were associated with a 28 percent reduction in mean LDL cholesterol and an 88 percent reduction in the risk of CHD ( $P=0.008$  for the reduction; hazard ratio, 0.11; 95 percent confidence interval, 0.02 to 0.81;  $P=0.03$ ). Of the 9524 white subjects examined, 3.2 percent had a sequence variation in *PCSK9* that was associated with a 15 percent reduction in LDL cholesterol and a 47 percent reduction in the risk of CHD (hazard ratio, 0.50; 95 percent confidence interval, 0.32 to 0.79;  $P=0.003$ ).

#### CONCLUSIONS

These data indicate that moderate lifelong reduction in the plasma level of LDL cholesterol is associated with a substantial reduction in the incidence of coronary events, even in populations with a high prevalence of non-lipid-related cardiovascular risk factors.

N Engl J Med 2006;354:1264-72.  
Copyright © 2006 Massachusetts Medical Society.



**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<sup>46L</sup>* Allele.**

In Panel A, the distribution of plasma LDL cholesterol levels at baseline among 9223 white subjects who did not have a *PCSK9<sup>46L</sup>* 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.

# Clinical trials

Phase I	Phase II	Phase III	Phase IV
<b>20-80 participants</b>	<b>100-300 participants</b>	<b>1,000-3,000 participants</b>	<b>Thousands of participants</b>
Up to several months	Up to (2) years	One (1) - Four (4) years	One (1) year +
Studies the safety of medication/treatment	Studies the efficacy	Studies the safety, efficacy and dosing	Studies the long-term effectiveness; cost effectiveness
70% success rate	33% success rate	25-30% success rate	70-90% success rate

# Clinical trials for PCSK9 antibodies

## Effect of a Monoclonal Antibody to PCSK9 on LDL Cholesterol

Evan A. Stein, M.D., Ph.D., Scott Mellis, M.D., Ph.D.,  
George D. Yancopoulos, M.D., Ph.D., Neil Stahl, Ph.D., Douglas Logan, M.D.,  
William B. Smith, M.D., Eleanor Lisbon, M.D., M.P.H., Maria Gutierrez, M.D.,  
Cheryle Webb, M.D., Richard Wu, Ph.D., Yunling Du, Ph.D.,  
Therese Kranz, R.N., M.B.A., Evelyn Gasparino, B.S.,  
and Gary D. Swerdlow, M.D., Ph.D.

### ABSTRACT

#### BACKGROUND

Proprotein convertase subtilisin/kexin 9 (PCSK9), one of the serine proteases, binds to low-density lipoprotein (LDL) receptors, leading to their accelerated degradation and to increased LDL cholesterol levels. We report three phase 1 studies of a monoclonal antibody to PCSK9 designated as REGN727/SAR236553 (REGN727).

#### METHODS

In healthy volunteers, we performed two randomized, single ascending-dose studies of REGN727 administered either intravenously (40 subjects) or subcutaneously (32 subjects), as compared with placebo. These studies were followed by a randomized, placebo-controlled, multiple-dose trial in adults with heterozygous familial hypercholesterolemia who were receiving atorvastatin (21 subjects) and those with nonfamilial hypercholesterolemia who were receiving treatment with atorvastatin (30 subjects) (baseline LDL cholesterol, >100 mg per deciliter [2.6 mmol per liter]) or a modified diet alone (10 subjects) (baseline LDL cholesterol, >130 mg per deciliter [3.4 mmol per liter]). REGN727 doses of 50, 100, or 150 mg were administered subcutaneously on days 1, 29, and 43. The primary outcome for all studies was the occurrence of adverse events. The principal secondary outcome was the effect of REGN727 on the lipid profile.

#### RESULTS

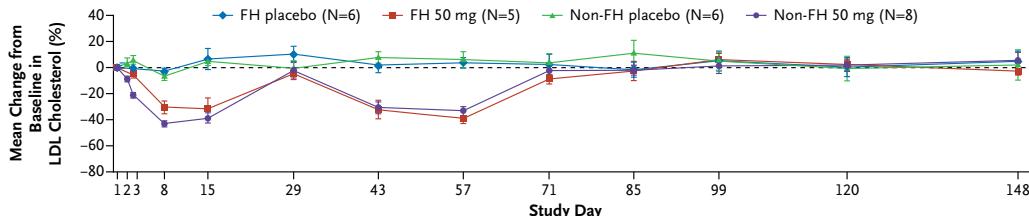
Among subjects receiving REGN727, there were no discontinuations because of adverse events. REGN727 significantly lowered LDL cholesterol levels in all the studies. In the multiple-dose study, REGN727 doses of 50, 100, and 150 mg reduced measured LDL cholesterol levels in the combined atorvastatin-treated populations to 77.5 mg per deciliter (2.00 mmol per liter), 61.3 mg per deciliter (1.59 mmol per liter), and 53.8 mg per deciliter (1.39 mmol per liter), for a difference in the change from baseline of -39.2, -53.7, and -61.0 percentage points, respectively, as compared with placebo ( $P<0.001$  for all comparisons).

#### CONCLUSIONS

In three phase 1 trials, a monoclonal antibody to PCSK9 significantly reduced LDL cholesterol levels in healthy volunteers and in subjects with familial or nonfamilial hypercholesterolemia. (Funded by Regeneron Pharmaceuticals and Sanofi; ClinicalTrials.gov numbers, NCT01026597, NCT01074372, and NCT01161082.)

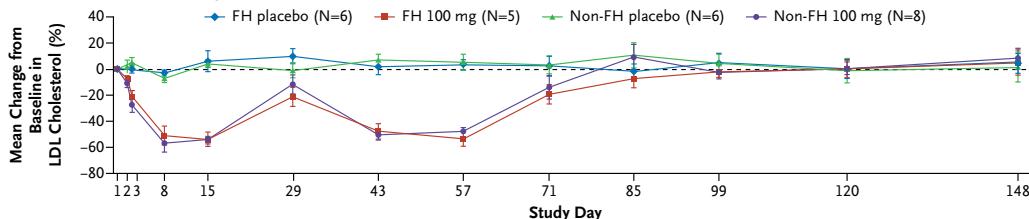
From the Metabolic and Atherosclerosis Research Center (E.A.S., C.W., T.K.) and the Medpace Clinical Pharmacology Unit (D.L.) — both in Cincinnati; Regeneron Pharmaceuticals, Tarrytown, NY (S.M., G.D.Y., N.S., R.W., Y.D., E.G., G.D.S.); New Orleans Center for Clinical Research, University of Tennessee Medical Center, Knoxville (W.B.S.); Quintiles, Overland Park, KS (E.L.); and Comprehensive Phase One, Miramar, FL (M.G.). Address reprint requests to Dr. Stein at the Metabolic and Atherosclerosis Research Center, 4685 Forest Ave., Cincinnati, OH 45212, or at esteinmrl@aol.com.

N Engl J Med 2012;366:1108-18.  
Copyright © 2012 Massachusetts Medical Society.

**A Subcutaneous Dose of 50 mg with Atorvastatin**

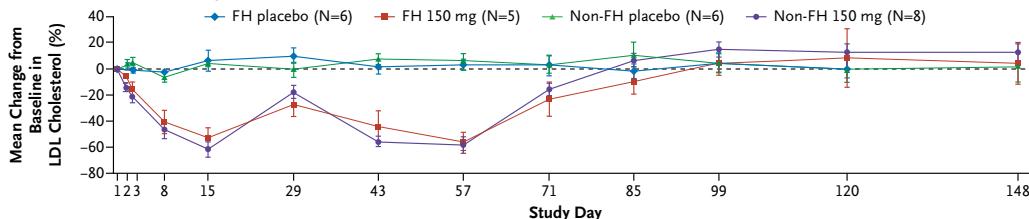
No. at Risk

FH placebo	666	6	6	6	6	6	6	6	6	6
FH 50 mg	555	5	5	5	5	5	5	5	5	5
Non-FH placebo	666	6	6	6	6	6	6	6	6	6
Non-FH 50 mg	888	8	8	8	8	8	8	8	8	8

**B Subcutaneous Dose of 100 mg with Atorvastatin**

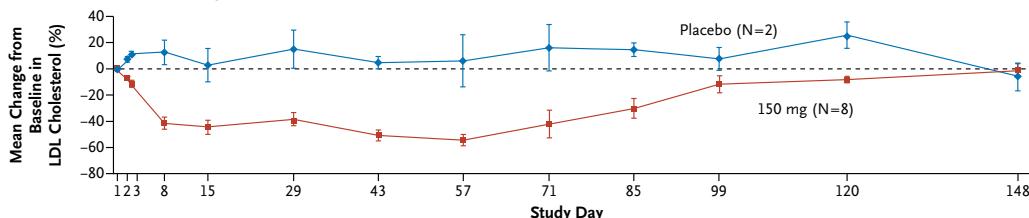
No. at Risk

FH placebo	666	6	6	6	6	6	6	6	6	6
FH 100 mg	555	5	5	5	5	5	5	5	5	5
Non-FH placebo	666	6	6	6	6	6	6	6	6	6
Non-FH 100 mg	888	8	8	8	8	8	8	8	8	8

**C Subcutaneous Dose of 150 mg with Atorvastatin**

No. at Risk

FH placebo	666	6	6	6	6	6	6	6	6	6
FH 150 mg	555	5	5	5	5	5	5	5	5	5
Non-FH placebo	666	6	6	6	6	6	6	6	6	6
Non-FH 150 mg	888	8	8	8	8	8	8	8	8	8

**D Subcutaneous Dose of 150 mg without Atorvastatin**

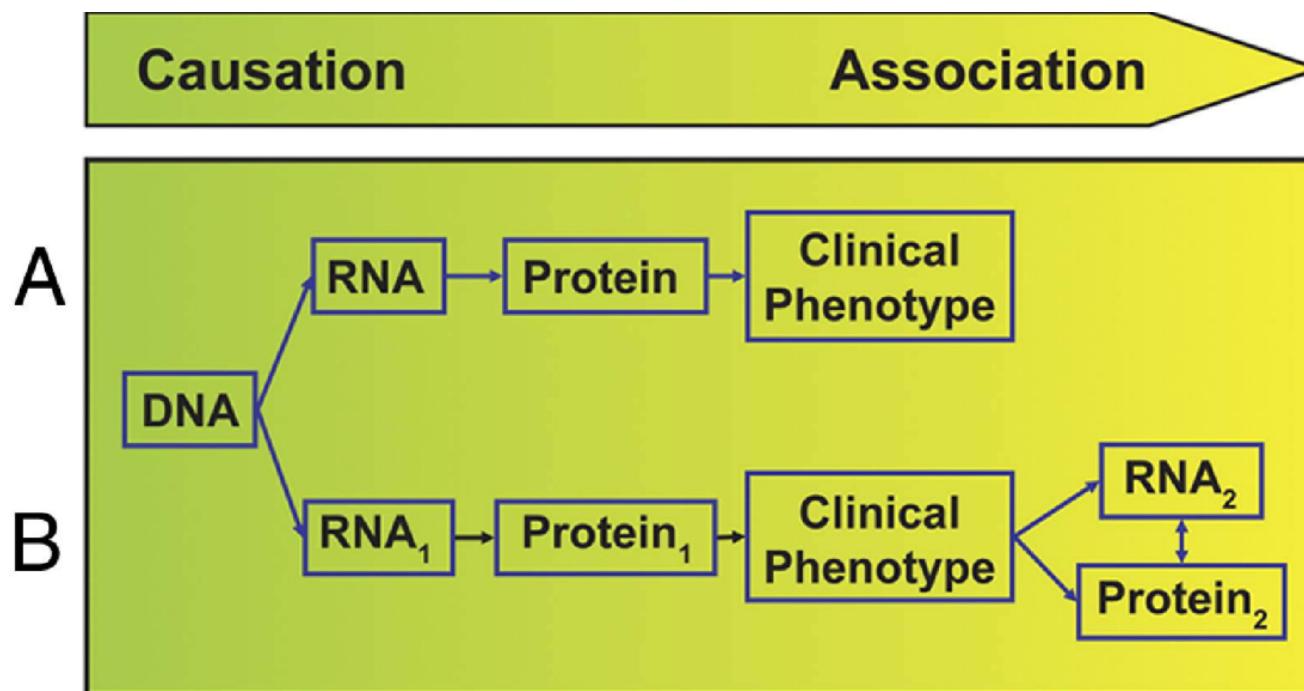
No. at Risk

Placebo	222	2	2	2	2	2	2	2	2	2
150 mg	888	8	8	8	8	8	8	8	8	8

# Biomarkers & Disease

## *chicken or the egg*

- A biomarker can be a protein causing the disease
- A biomarker can be a protein associated with disease



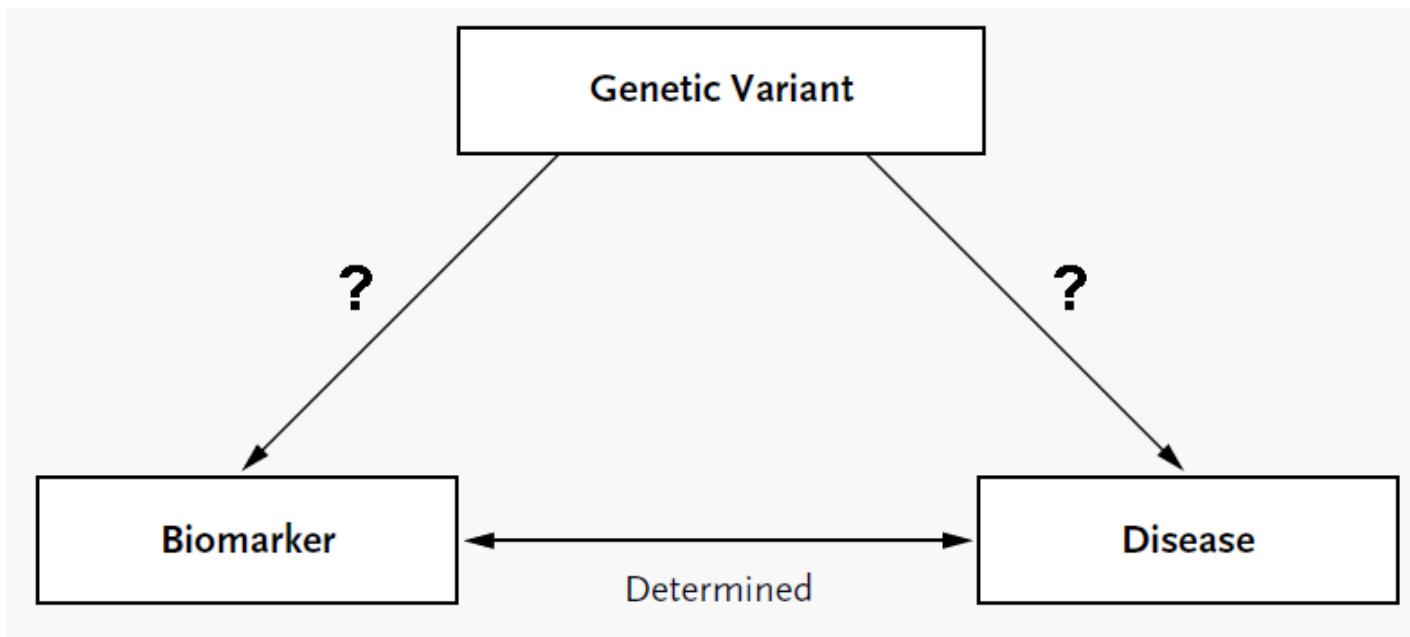
# SNPs, Biomarkers & Disease

*can genetic variation tell us something?*



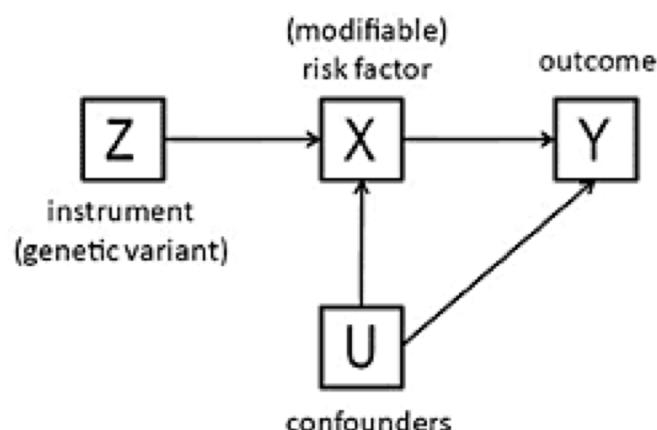
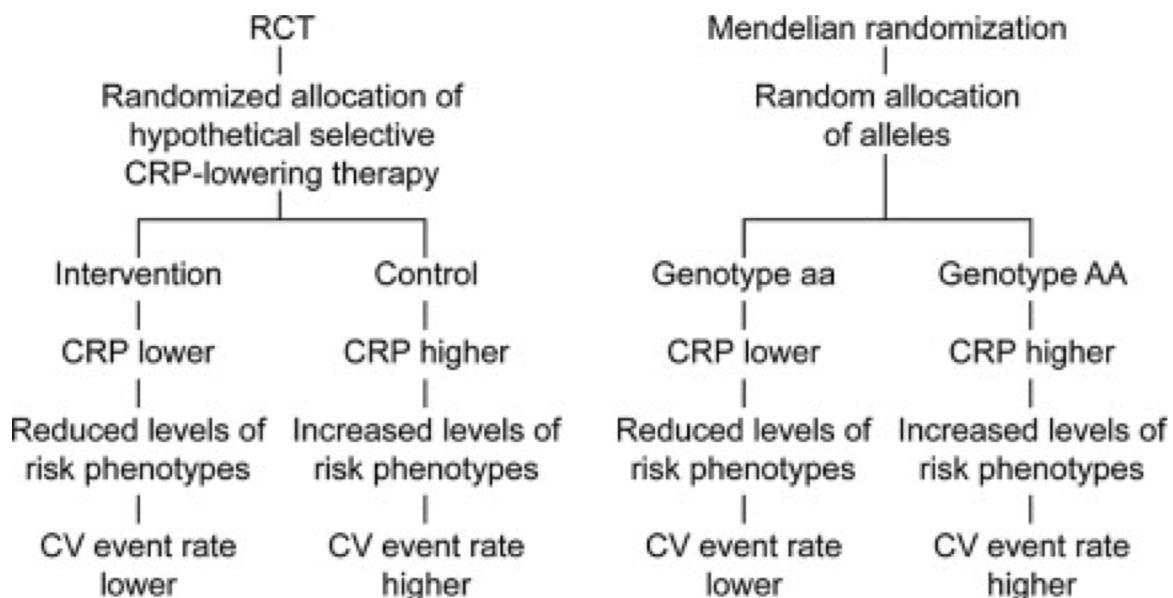
University Medical Center  
Utrecht

- A biomarker is associated with disease
- A genetic variant is associated with the expression of the biomarker
- A genetic variants is associated with disease



# Mendelian Randomization

- Genotypes are randomly assigned during meiosis and transferred from parent to child
- This is a “natural randomized controlled trial” based on genetic variation
- Use genetic variation as an instrument to infer causality



# HDL protective of MI?

- Various variants known to influence HDL levels
- HDL levels are associated with lower risk of MI (or CAD)
- Some also associated with the risk of MI (or CAD)
- Is a lower HDL level causal to a lower risk for MI?



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

Benjamin F Voight\*, Gina M Peloso\*, Marju Orho-Melander, Ruth Fririke-Schmidt, Maja Borbalic, Majken Kjeldsen, George H Cindy, Hilma H ölm, Eric L Ding, Toby Johnson, Heribert Schunkert, Niles 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 Thorsteinsdóttir, Guðmundur Þorgerðsson, Sónia Arand, James C Engert, Thomas Morgan, John Spertus, Monika Stoll, Klaus Berger, Nicola Martinielli, 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 Sirola, Marja-Liisa Lokki, Markus Perola, Aki Havulinna, Ulf de Faire, Bruna Gigante, Erik Ingelsson, Tanja Zeller, Philipp Wild, Paul I W de Bakker, Olaf H Klüngel, Anke-Hilse Maitland-van der Zee, Bas J M Peters, Anthoniou de Boer, Diederick E Grobbee, Pieter W Kamphuisen, Vera H M Deneer, Clara C Elbers, N Charlotte Onland-Moret, Marten H Hofker, Cisca Wijmenga, WM 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 Mihalek, Mark J Daly, Candace Guiducci, Noël P Butt, 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 Buyssechaert, Diether Lambrechts, Frans Van de Werf, Keith A Fox, Nour Eddine El Mokhtar, Diana Rubin, Jürgen Schrezenmeier, Stefan Schreiber, Anne Schäfer, John Danesh, Stefan Blanckenberg, Robert Roberts, Ruth McPherson, Hugh Watkins, Alastair S Hall, Kim Overvad, Eric Rimm, Eric Boerwinkle, Anne Tybjærg-Hansen, L Adrienne Cupples, Muredach P Reilly, Olle Melander, Pier M Mannucci, Diego Ardissino, David Siscovick, Roberto Eliasua, Kari Stefansson, Christopher J O'Donnell, Veikko Salomaa, Daniel J Rader, Leena Peltonen, Stephen M Schwartz, David Altschuler, Sekar Kathiresan

### Summary

**Background** High plasma HDL cholesterol is associated with reduced risk of myocardial infarction, but whether this association is causal is unclear. Exploiting the fact that genotypes are randomly assigned at meiosis, are independent of non-genetic confounding, and are unmodified by disease processes, mendelian randomisation can be used to test the hypothesis that the association of a plasma biomarker with disease is causal.

**Methods** We performed two mendelian randomisation analyses. First, we used as an instrument a single nucleotide polymorphism (SNP) in the endothelial lipase gene (*LIPG* Asn396Ser) and tested this SNP in 20 studies (20 913 myocardial infarction cases, 95 407 controls). Second, we used as an instrument a genetic score consisting of 14 common SNPs that exclusively associate with HDL cholesterol and tested this score in up to 12 482 cases of myocardial infarction and 41 331 controls. As a positive control, we also tested a genetic score of 13 common SNPs exclusively associated with LDL cholesterol.

**Findings** Carriers of the *LIPG* 396Ser allele (2·6% frequency) had higher HDL cholesterol (0·14 mmol/L higher,  $p=8\times10^{-13}$ ) but similar levels of other lipid and non-lipid risk factors for myocardial infarction compared with non-carriers. This difference in HDL cholesterol is expected to decrease risk of myocardial infarction by 13% (odds ratio [OR] 0·87, 95% CI 0·84–0·91). However, we noted that the 396Ser allele was not associated with risk of myocardial infarction (OR 0·99, 95% CI 0·88–1·11,  $p=0·85$ ). From observational epidemiology, an increase of 1 SD in HDL cholesterol was associated with reduced risk of myocardial infarction (OR 0·62, 95% CI 0·58–0·66). However, a 1 SD increase in HDL cholesterol due to genetic score was not associated with risk of myocardial infarction (OR 0·93, 95% CI 0·68–1·26,  $p=0·63$ ). For LDL cholesterol, the estimate from observational epidemiology (a 1 SD increase in LDL cholesterol associated with OR 1·54, 95% CI 1·45–1·63) was concordant with that from genetic score (OR 2·13, 95% CI 1·69–2·69,  $p=2\times10^{-10}$ ).

**Interpretation** Some genetic mechanisms that raise plasma HDL cholesterol do not seem to lower risk of myocardial infarction. These data challenge the concept that raising of plasma HDL cholesterol will uniformly translate into reductions in risk of myocardial infarction.

**Funding** US National Institutes of Health, The Wellcome Trust, European Union, British Heart Foundation, and the German Federal Ministry of Education and Research.

### Introduction

Cholesterol fractions such as LDL and HDL cholesterol are among the most commonly measured biomarkers in clinical medicine.<sup>1</sup> Observational studies have shown that LDL and HDL cholesterol have opposing associations

with risk of myocardial infarction, with LDL cholesterol being positively associated and HDL cholesterol being inversely associated.<sup>2,3</sup> However, observational studies cannot distinguish between a causal role in the pathological process and a marker of the underlying

# The answer in three steps... #1

- Various variants associated with HDL levels

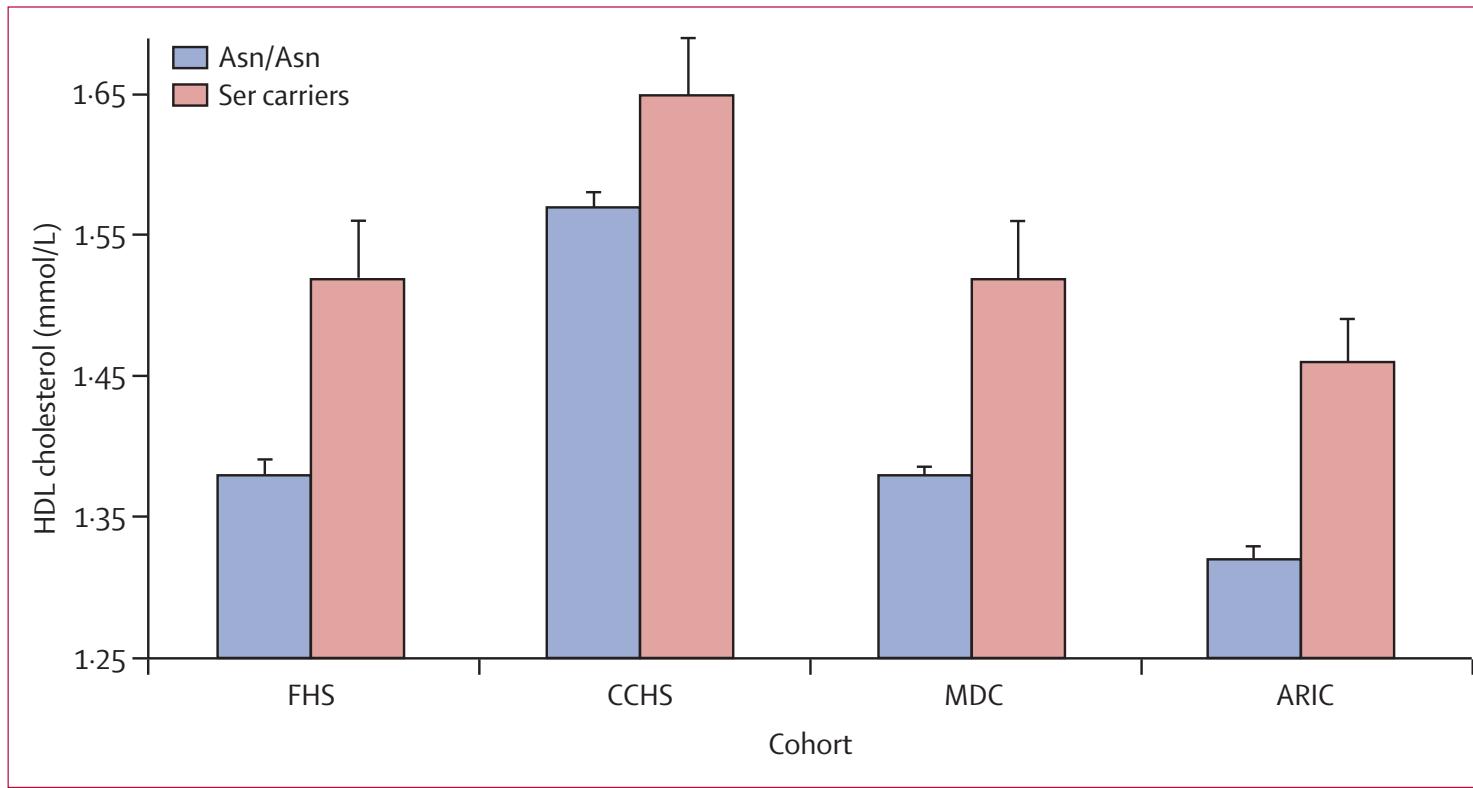
	Chromosome	Gene(s) of interest within or near associated interval	Major allele, minor allele (minor allele frequency)*	Modelled allele	Effect of modelled allele on plasma HDL cholesterol (mmol/L)*	Effect of modelled allele on plasma triglycerides (mmol/L)*	Effect of modelled allele on plasma LDL cholesterol (mmol/L)*	Sample size (MI cases/MI-free controls)	For modelled allele, observed change in MI risk (%; 95% CI)	For modelled allele, p value for association with MI
rs17482753	8p21	LPL†	G, T (0.10)	T	0.08	-0.24	..	19 139/50 812	-12% (-16 to -7)	4×10 <sup>-7</sup> †
rs17321515	8q24	TRIB1†	A, G (0.45)	G	0.02	-0.11	-0.05	19 139/50 812	-7% (-9 to -4)	2×10 <sup>-6</sup> †
rs6589566	11q23	APOA1-APOC3-APOA4-APOA5†	A, G (0.07)	A	0.05	-0.27	-0.09	18 310/49 897	-10% (-15 to -5)	8×10 <sup>-5</sup> †
rs4846914	1q42	GALNT2†	A, G (0.40)	A	0.02	-0.03	..	19 139/50 812	-3% (-6 to -1)	0.02†
rs2967605	19p13	ANGPTL4†	C, T (0.16)	C	0.05	-0.07	..	13 595/16 423	-5% (-10 to -1)	0.03†
rs3764261	16q13	CETP†	C, A (0.32)	A	0.10	..	-0.03	16 503/46 576	-4% (-7 to 0)	0.04†
rs61755018 (Asn396Ser)	18q21	LIPG	A, G (0.015)	G	0.14‡	..	..	17 165/49 077	-6% (-18 to 9)	0.41
rs17145738	7q11	MLXIPL	C, T (0.11)	T	0.03	-0.15	..	19 139/50 812	-1% (-4 to 3)	0.61
rs3890182	9q31	ABCA1	G, A (0.14)	G	0.03	..	0.05	19 139/50 812	-1% (-5 to 4)	0.76
rs2338104	12q24	MMAB, MVK	G, C (0.46)	G	0.03	..	..	19 139/50 812	0% (-3 to 3)	0.85
rs471364	9p22	TTC39B	T, C (0.12)	T	0.03	..	..	15 693/47 098	0% (-5 to 5)	0.97
rs2271293	16q22	LCAT	G, A (0.11)	A	0.03	..	..	19 139/50 812	4% (-1 to 8)	0.10
rs174547	11q12	FADS1-FADS2-FADS3	T, C (0.33)	T	0.03	-0.06	..	19 139/50 812	3% (-1 to 6)	0.11
rs1800588	15q22	LIPC	C, T (0.22)	T	0.05	0.07	..	17 917/49 514	4% (0 to 7)	0.04
rs16988929	20q13	HNF4A	C, T (0.01)	T	0.01	..	..	17 041/20 137	31% (12 to 54)	9×10 <sup>-4</sup>

\*Data presented from a meta-analysis of seven cohorts (n up to 19 840) as presented in reference 16; the effect of each SNP on a lipid trait was modelled if the association of the SNP with a plasma lipid trait exceeded nominal significance ( $p < 0.05$ ). †Loci and SNPs that exceeded nominal significance ( $p < 0.05$ ) for association of modelled allele with MI; all modelled alleles increased HDL cholesterol. ‡Effect size presented is from the Atherosclerosis Risk in Communities Study.

Table 2: Association of myocardial infarction (MI) with single nucleotide polymorphisms (SNPs) previously found to relate to plasma HDL cholesterol

# The answer in three steps... #2

- *LIPG* variant associates with HDL levels



**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.

# The answer in three steps... #3

- No association with MI...

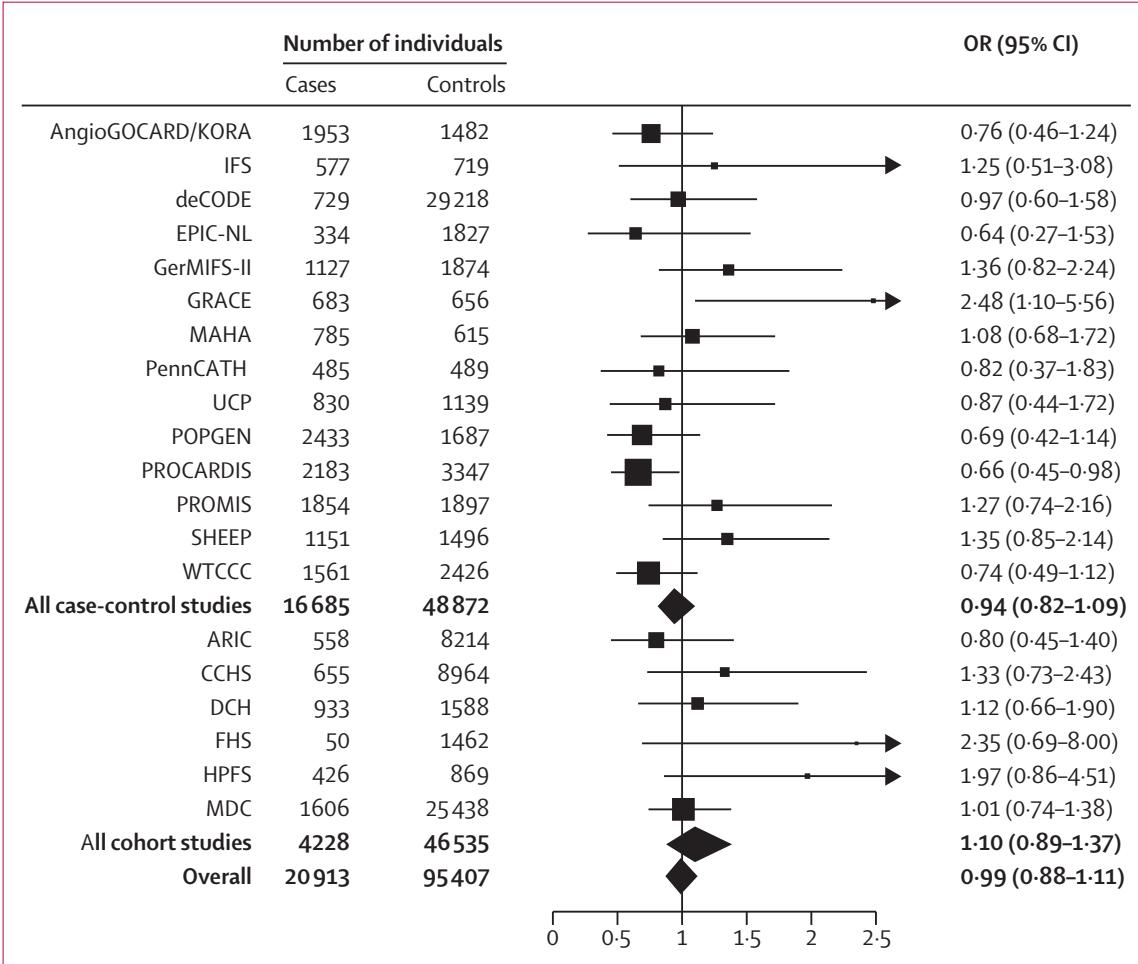


Figure 2: Association of LIPG Asn396Ser with myocardial infarction in 116 320 participants from 20 studies  
In each study, the HDL-cholesterol-raising serine allele was modelled.

# Cardiovascular Genetic Research

## Experimental Cardiology Laboratory

Prof. Dr. G. Pasterkamp

Prof. Dr. D.P.V. de Kleijn (Singapore)

Dr. J.P.G. Sluijter

Dr. I. Höfer

## Medical Genetics

Prof. Dr. P.I.W. de Bakker

Dr. F.W. Asselbergs

## Research topics

Biomarker Discovery & Validation

*Athero-Express | CTMM: Circulating Cells | Toll Express*

Regenerative Medicine in Ischaemic Heart Disease

*Stem Cells | Progenitor Cells*

Mechanisms of Arterial Occlusive Disease

*Toll-like receptors | Regenerative Medicine |  
Arteriogenesis*

Cardiovascular Genomics

*Next-Generation Sequencing | GWAS |  
Pharmacogenomics*



S.W. van der Laan – [s.w.vanderlaan-2@umcutrecht.nl](mailto:s.w.vanderlaan-2@umcutrecht.nl)



University Medical Center  
Utrecht



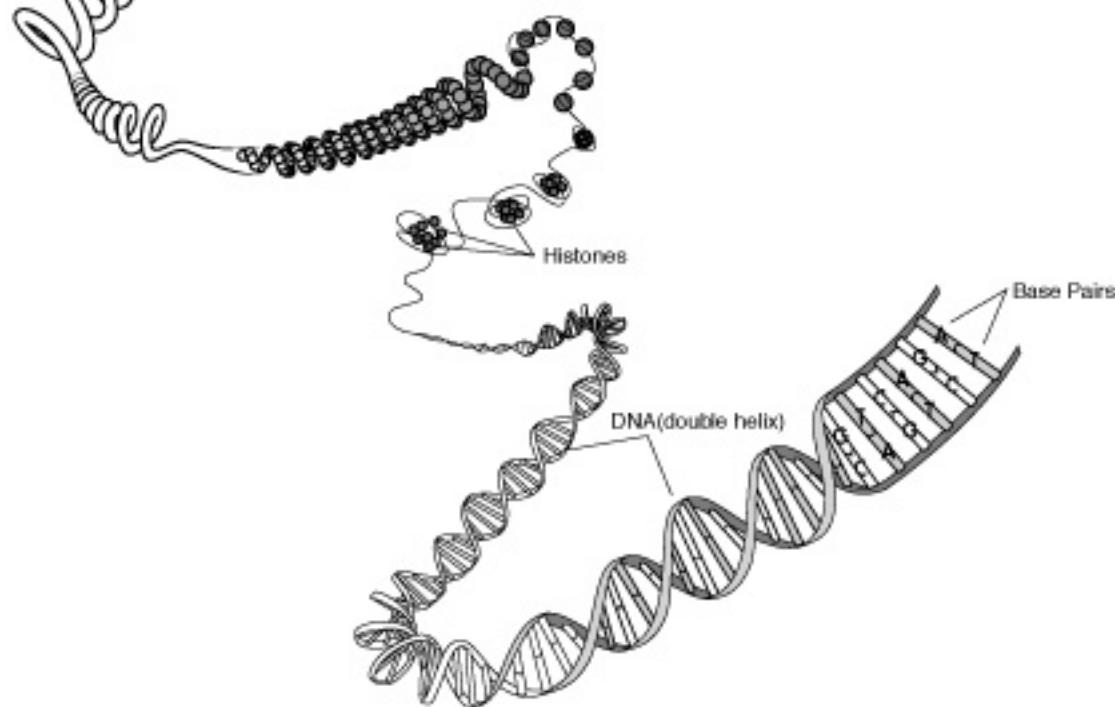
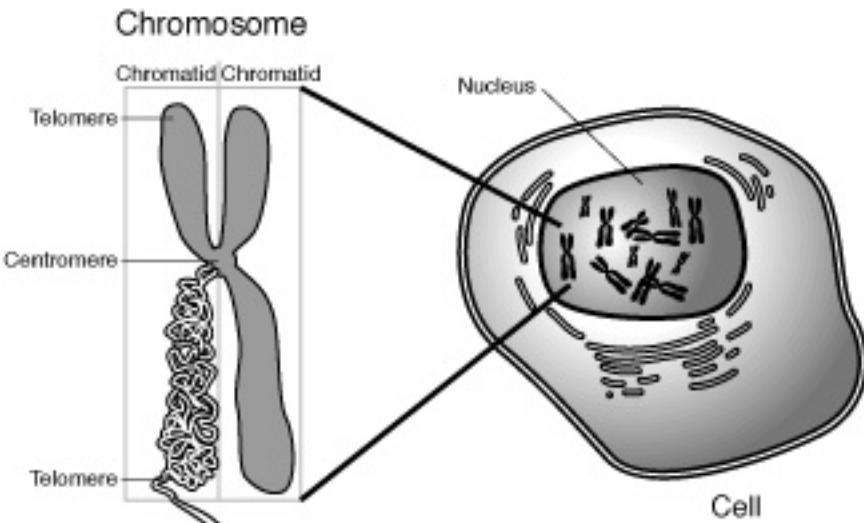
**CVON**  
CardioVascular Onderzoek Nederland  
Een privaat-publiek initiatief



**CAVADIS**

# Further reading

- McCarthy MI, Abecasis GR, Cardon LR, et al. **Genome-wide association studies for complex traits: Consensus, uncertainty and challenges.** *Nat Rev Genet.* 2008;9:356-369
- Feero WG, Guttmacher AE, Collins FS. **Genomic medicine--an updated primer.** *N Engl J Med.* 2010;362:2001-2011
- Damani SB, Topol EJ. **Future use of genomics in coronary artery disease.** *J Am Coll Cardiol.* 2007;50:1933-1940
- Pearson TA, Manolio TA. **How to interpret a genome-wide association study.** *JAMA : the journal of the American Medical Association.* 2008;299:1335-1344
- Dietz HC. **New therapeutic approaches to mendelian disorders.** *N Engl J Med.* 2010;363:852-863
- Manolio TA. **Genome wide association studies and assessment of the risk of disease.** *N Engl J Med.* 2010;363:166-176
- Schork NJ, Murray SS, Frazer KA, Topol EJ. **Common vs. Rare allele hypotheses for complex diseases.** *Curr Opin Genet Dev.* 2009;19:212-219
- Manolio TA, Collins FS, Cox NJ, et al. **Finding the missing heritability of complex diseases.** *Nature.* 2009;461:747-753
- Schunkert H, Konig IR, Kathiresan S, et al. **Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.** *Nature genetics.* 2011;43:333-338
- Pollex RL, Hegele RA. **Copy number variation in the human genome and its implications for cardiovascular disease.** *Circulation.* 2007;115:3130-3138



Within each cell

2 genome copies

23 chromosomes

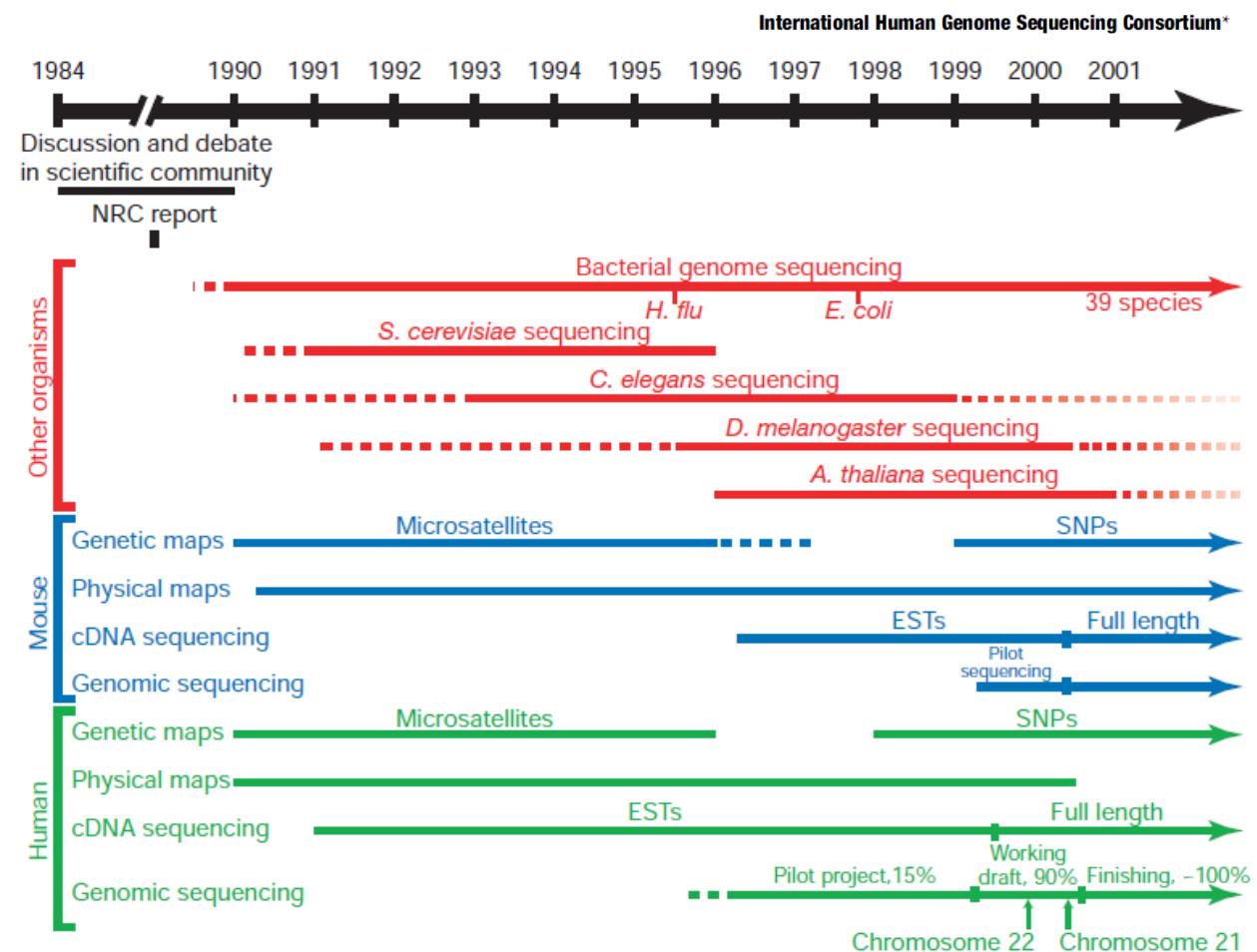
~20,000 genes

3 billion bases

# Human Genome Project: 2001

## articles

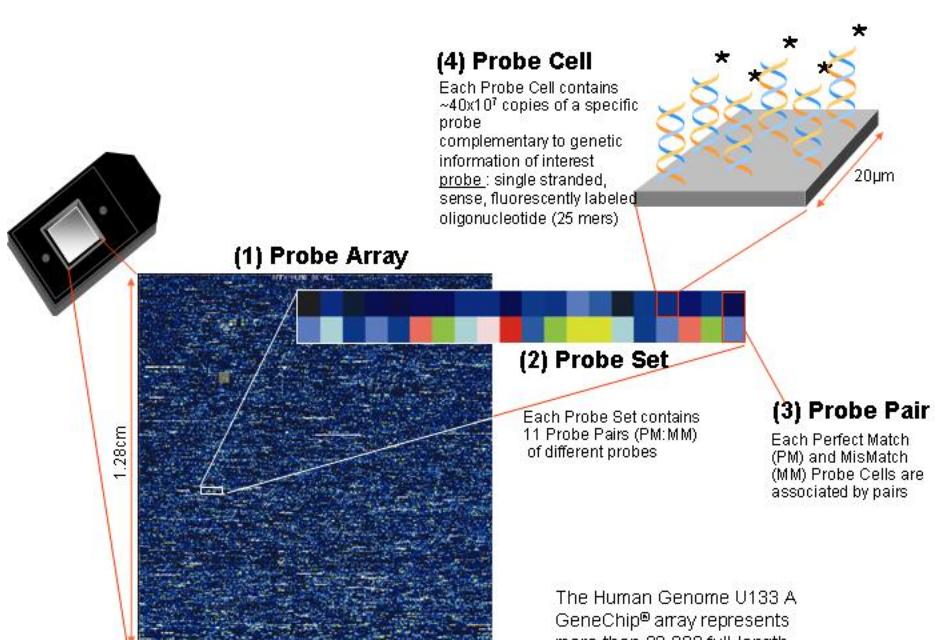
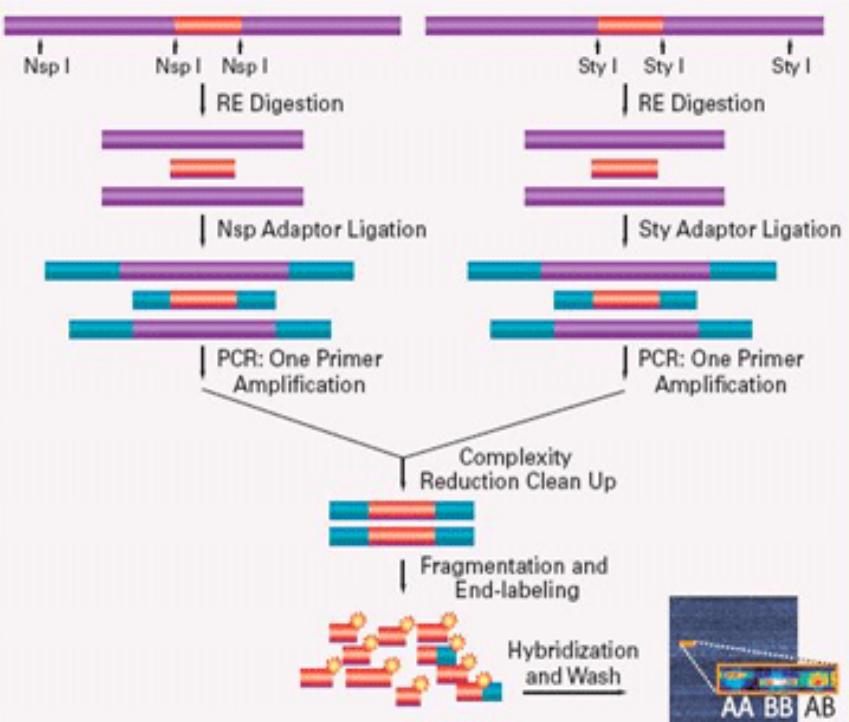
# Initial sequencing and analysis of the human genome



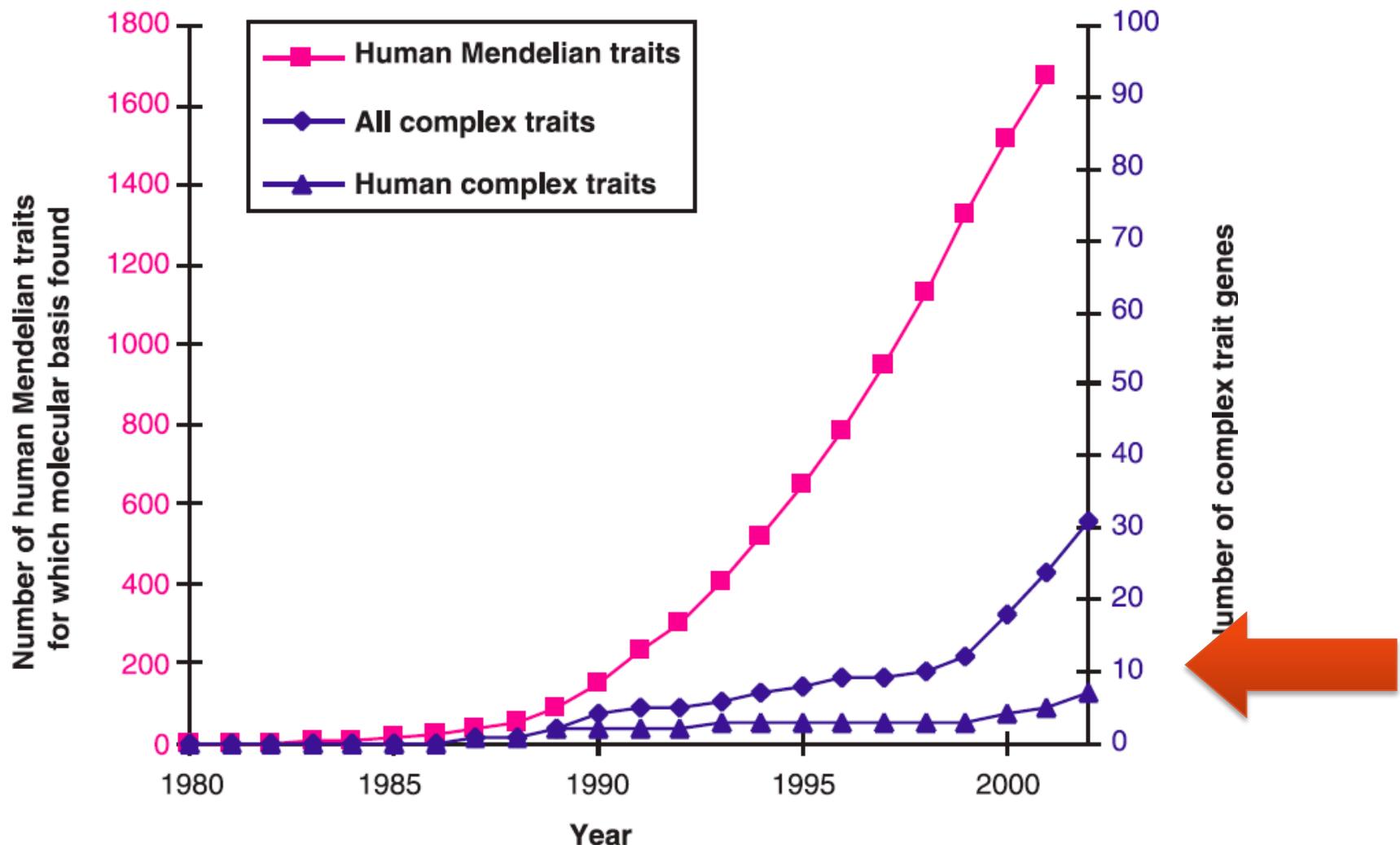
- Largest publication ever in *Nature*, February 2001: 62 pages
- “Back-to-back with a publication by Craig Venter, *et al.* in *Science*

# SNP “genotyping”

The fifth-generation Whole-genome Sampling Assay.

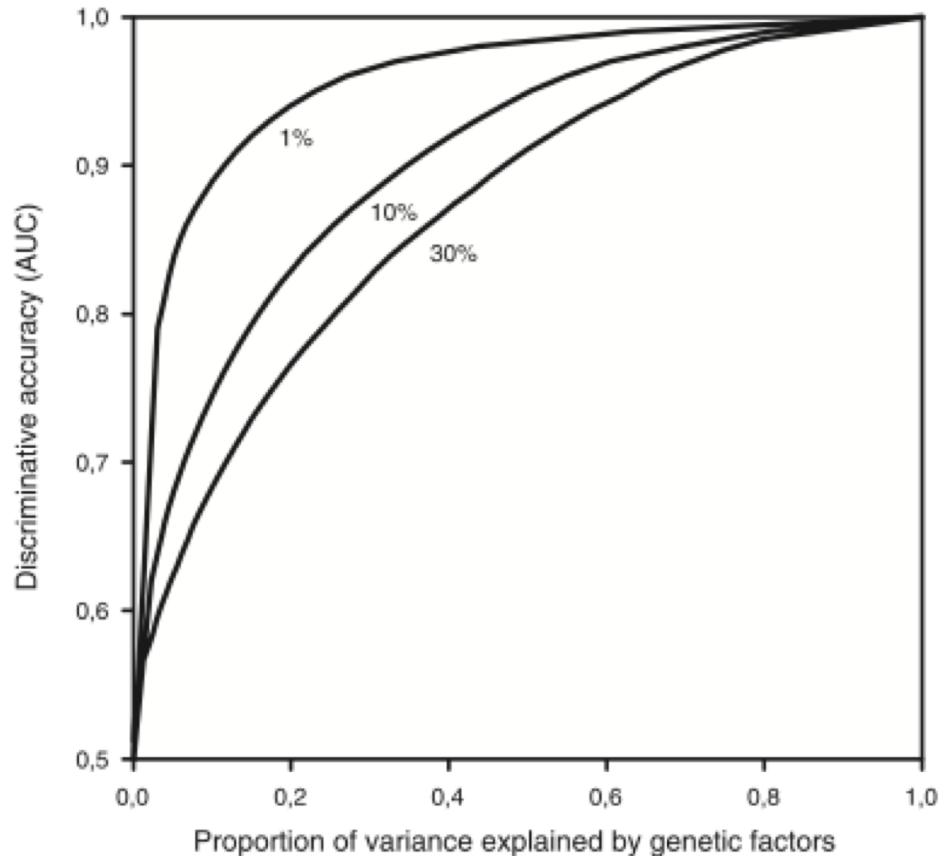
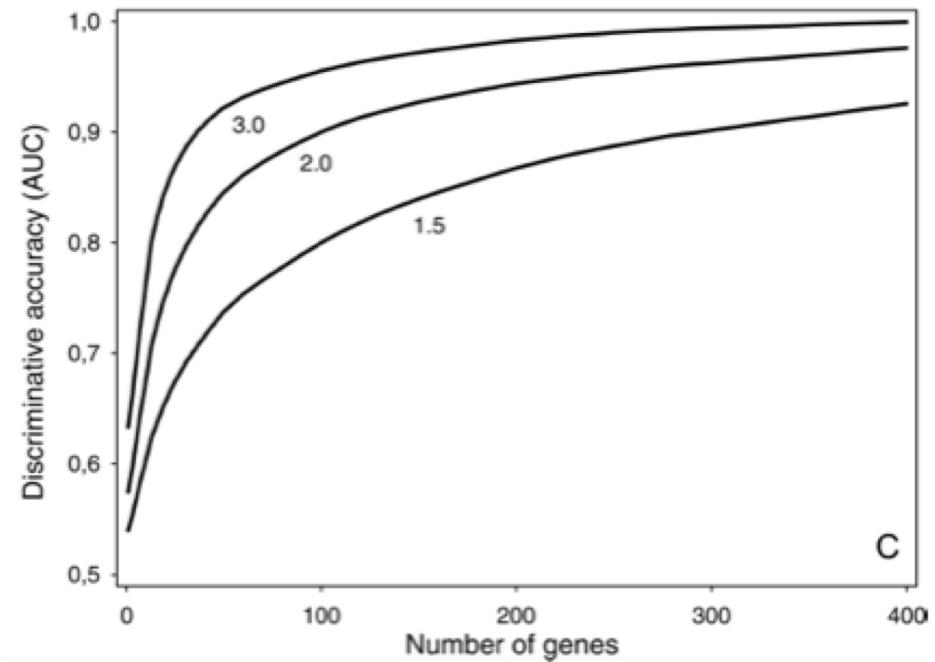


# Early successes genomic research of monogenic disease



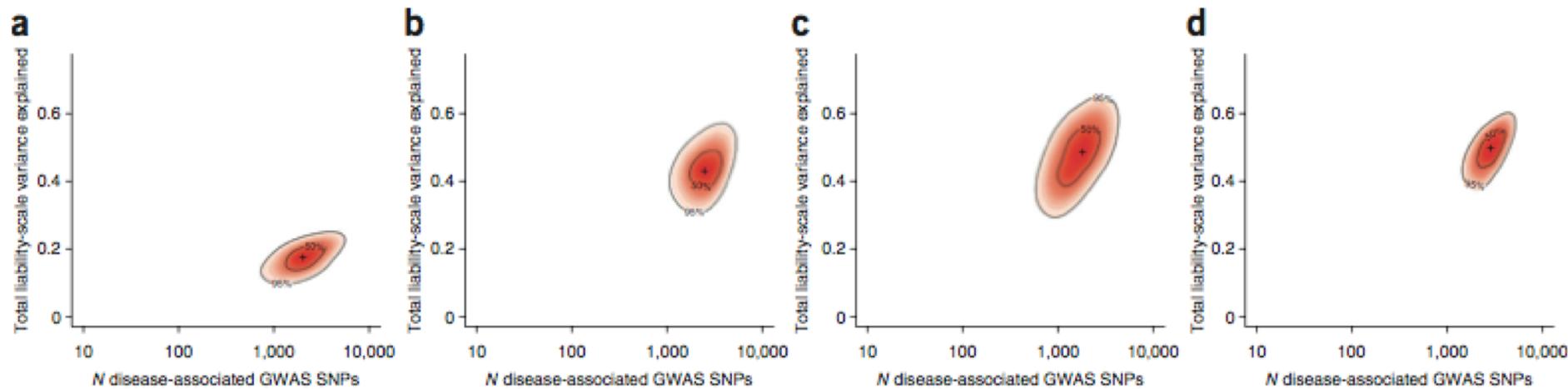
# So can we even predict?

- For common disease hundreds of genetic variants are needed for a clinically useful prediction



# Is there even hope?

- Research indicates hundreds of variants are awaiting discovery



**Figure 2** Posterior probability densities of the number of associated SNPs and the total liability-scale variance explained for the Bayesian analysis of the polygenic analysis results.  $N_{SNPs}$  are shown on the  $\log_{10}$  scale on the x axis, and  $V_{tot}$  values are shown on the y axis. The heat map colors represent the probability density height, with darker colors indicating higher density. Contour lines show the highest posterior density and the 50%, 90% and 95% credible regions. (a) Rheumatoid arthritis (with all known risk loci removed). (b) Celiac disease (with the extended MHC region removed). (c) MI/CAD. (d) T2D.

# What type of study design can you choose?

## *genome-wide association study*



University Medical Center  
Utrecht

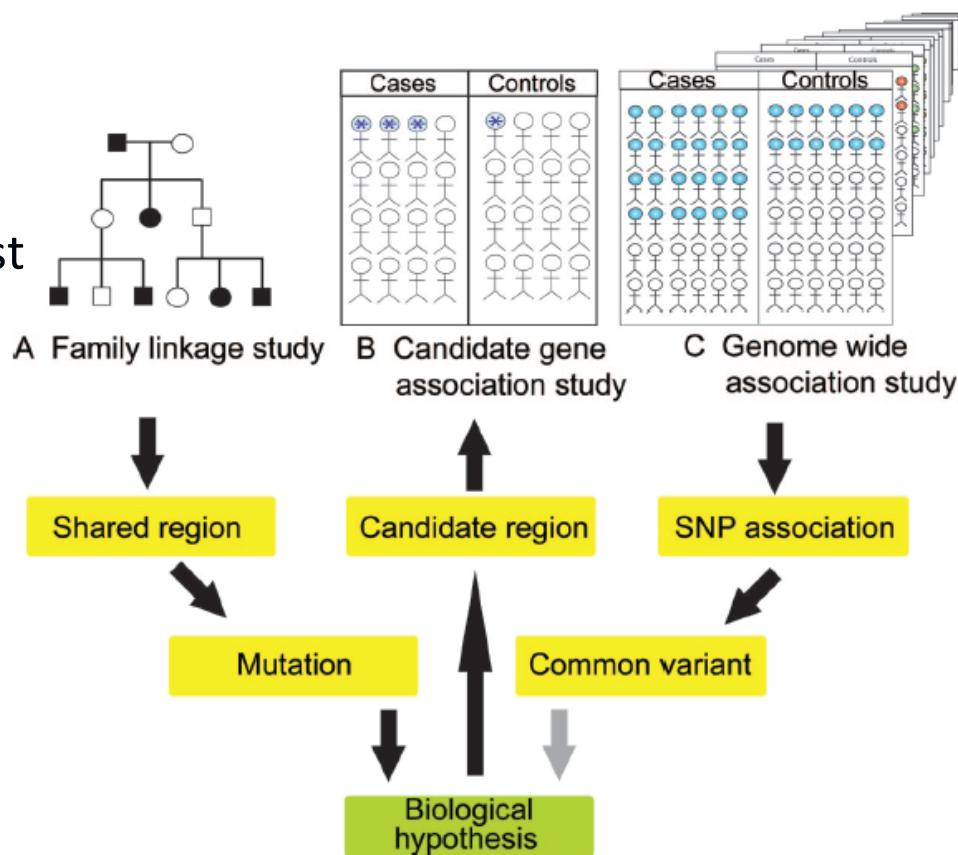
**Table 1.** Study Designs Used in Genome-wide Association Studies

	Case-Control	Cohort	Trio
Assumptions	Case and control participants are drawn from the same population Case participants are representative of all cases of the disease, or limitations on diagnostic specificity and representativeness are clearly specified Genomic and epidemiologic data are collected similarly in cases and controls Differences in allele frequencies relate to the outcome of interest rather than differences in background population between cases and controls	Participants under study are more representative of the population from which they are drawn Diseases and traits are ascertained similarly in individuals with and without the gene variant	Disease-related alleles are transmitted in excess of 50% to affected offspring from heterozygous parents
Advantages	Short time frame Large numbers of case and control participants can be assembled Optimal epidemiologic design for studying rare diseases	Cases are incident (developing during observation) and free of survival bias Direct measure of risk Fewer biases than case-control studies Continuum of health-related measures available in population samples not selected for presence of disease	Controls for population structure; immune to population stratification Allows checks for Mendelian inheritance patterns in genotyping quality control Logistically simpler for studies of children's conditions Does not require phenotyping of parents
Disadvantages	Prone to a number of biases including population stratification Cases are usually prevalent cases, may exclude fatal or short episodes, or mild or silent cases Overestimate relative risk for common diseases	Large sample size needed for genotyping if incidence is low Expensive and lengthy follow-up Existing consent may be insufficient for GWA genotyping or data sharing Requires variation in trait being studied Poorly suited for studying rare diseases	May be difficult to assemble both parents and offspring, especially in disorders with older ages of onset Highly sensitive to genotyping error

- GWAS is hypothesis-free: no *a priori* ideas on which variant is associated

# To hypothesize or not to hypothesize...

- **Family linkage study**
  - Trio-design: parents plus child
- **Candidate Gene Association Study**
  - *A priori* hypothesis
- **Genome-Wide Association Study**
  - No *a priori* hypothesis
  - Cases: some phenotype of interest
  - Controls: random population sample



# Pleiotropic effects of discovered loci association with risk factors and other diseases

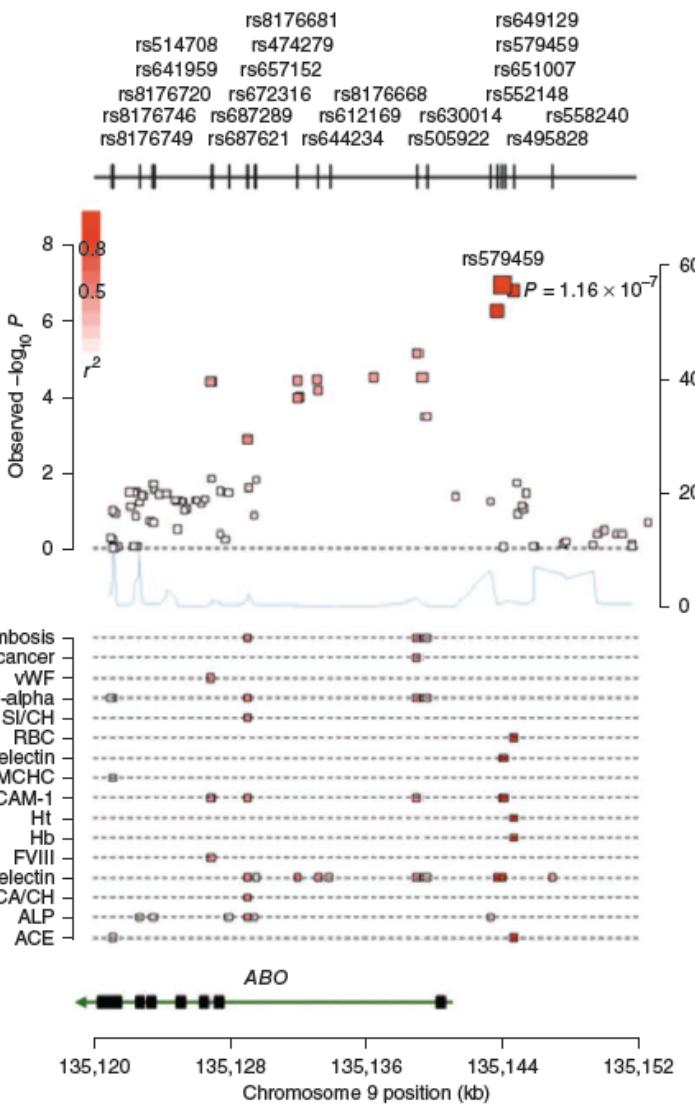
**Table 3** Effects of new CAD loci on traditional risk factors in combined analysis of ARIC and KORA F3 and F4 ( $n = 13,171$ )

SNP	Band	Gene(s) in region	Phenotype	$\beta$ (95% CI) <sup>a</sup>	$P$
rs579459	9q34.2	<i>ABO</i>	Total cholesterol	1.720 (0.554–2.885)	0.0038
			LDL cholesterol	1.538 (0.468–2.608)	0.0049
rs12413409	10q24.32	<i>CYP17A1, CNNM2, NT5C2</i>	Hypertension	0.141 (0.044–0.238)	0.0043
rs964184 <sup>b</sup>	11q23.3	<i>ZNF259, APOA5-A4-C3-A1</i>	HDL cholesterol	-1.926 (-2.441 to -1.411)	$2.28 \times 10^{-13}$
			Total cholesterol	4.578 (3.191–5.964)	$9.84 \times 10^{-11}$
			LDL cholesterol	1.699 (0.417–2.980)	0.0094

Results from fixed-effects meta-analysis based on  $\beta$  coefficients and standard errors from linear (for total, LDL and HDL cholesterol) and logistic (for hypertension) regression analysis of the single studies for which meta-analytic  $P < 0.01$ . LDL, low-density lipoprotein; HDL, high-density lipoprotein.

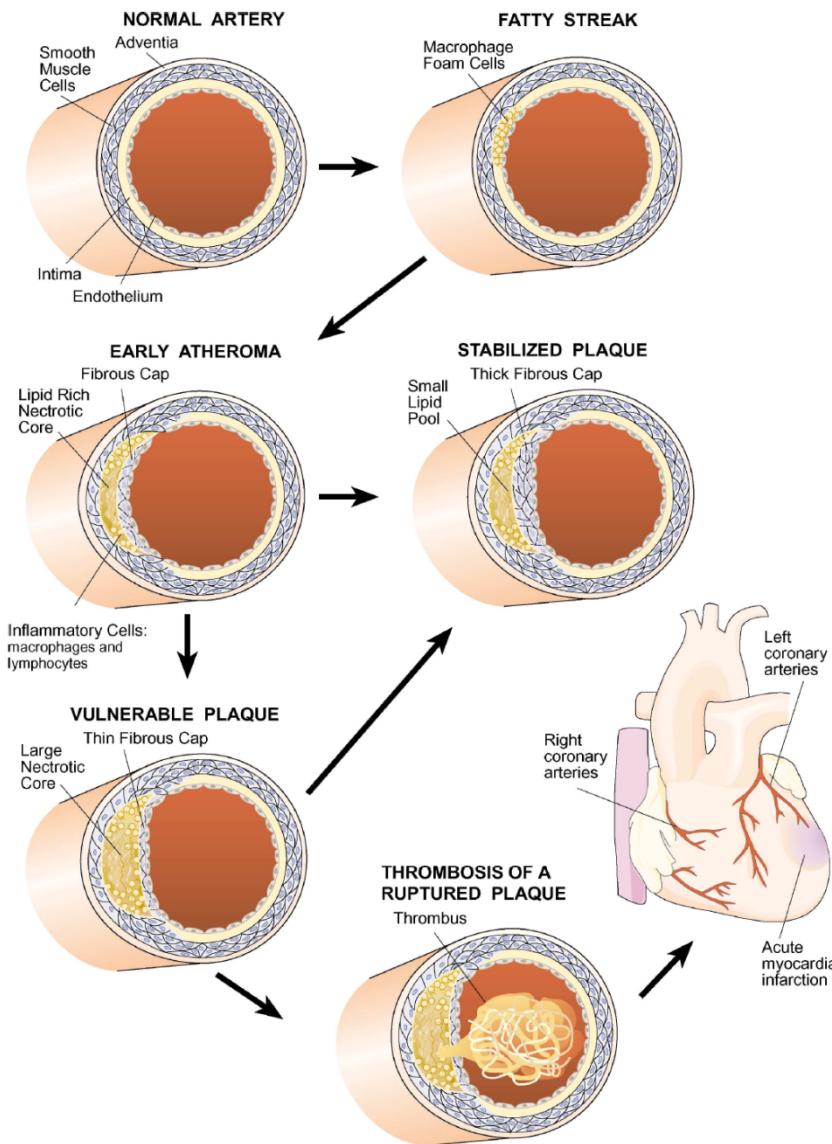
<sup>a</sup>Estimated pooled regression coefficients with 95% confidence intervals. Cholesterol levels are in mg/dl. <sup>b</sup>Previous genome-wide studies have demonstrated strong association of rs964184 with triglycerides<sup>39</sup>.

**Figure 2** Example of overlapping association signals for multiple traits at the *ABO* gene region on chromosome 9q34. In the upper panel, the association signal for coronary disease at the *ABO* gene region in CARDIoGRAM and the positions and rs numbers of SNPs in this region are shown. The size of the boxes illustrates the number of individuals available for this respective SNP. In the lower panel, all SNPs with  $P$  values at the genome-wide significance level of  $P < 5 \times 10^{-8}$  based on the National Human Genome Research Institute GWAS catalog (accessed on 28 June 2010) for all diseases and traits are shown. The degree of linkage disequilibrium ( $r^2$ ) between the lead SNPs for coronary disease and the other traits is reflected by the color of the squares (upper panel) and the small bars (lower panel), from dark red (high LD) to faint red (low LD). SI/CH, sitosterol normalized to cholesterol; CA/CH, campesterol normalized to cholesterol; ALP, alkaline phosphatase; ACE, angiotensin converting enzyme; FVIII, coagulation factor VIII; vWF, von Willebrand factor.



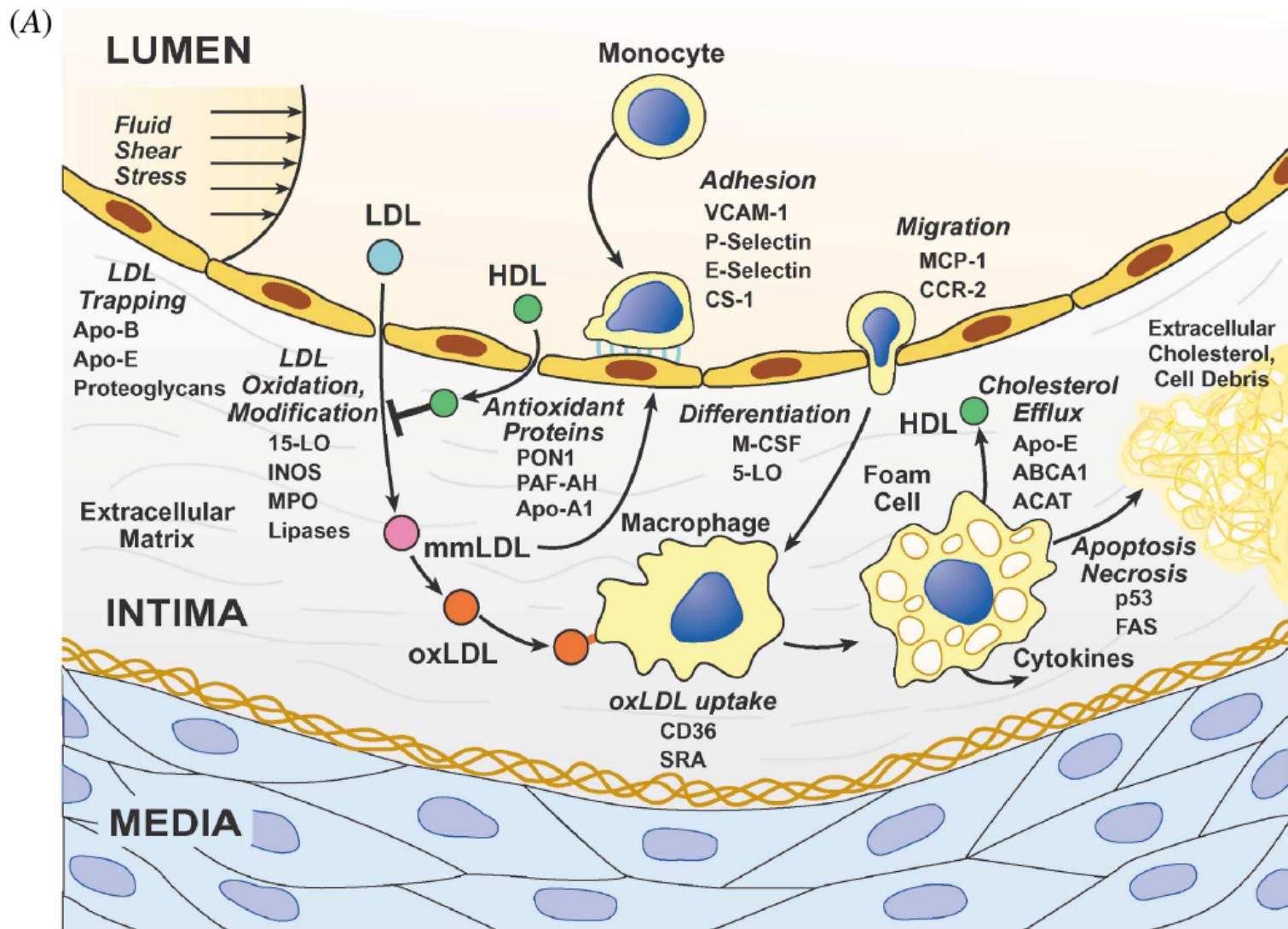
# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*

