

# 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?
  - Mendelian Diseases
  - Complex Diseases
- Genetics of Advanced Atherosclerotic Disease
- Genetic Burden for Disease (Risk)
- Genetics, Biomarkers & Disease



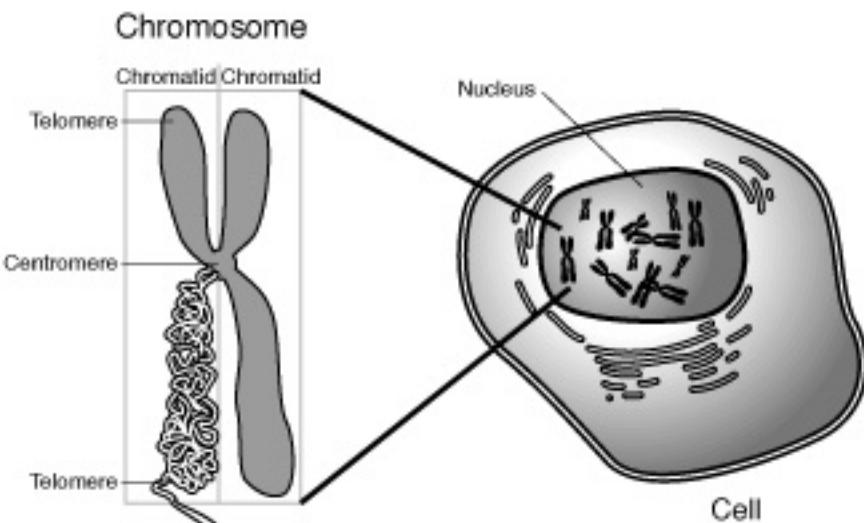
University Medical Center  
Utrecht



Recapturing Some Basic Genetics

# THE HUMAN GENOME

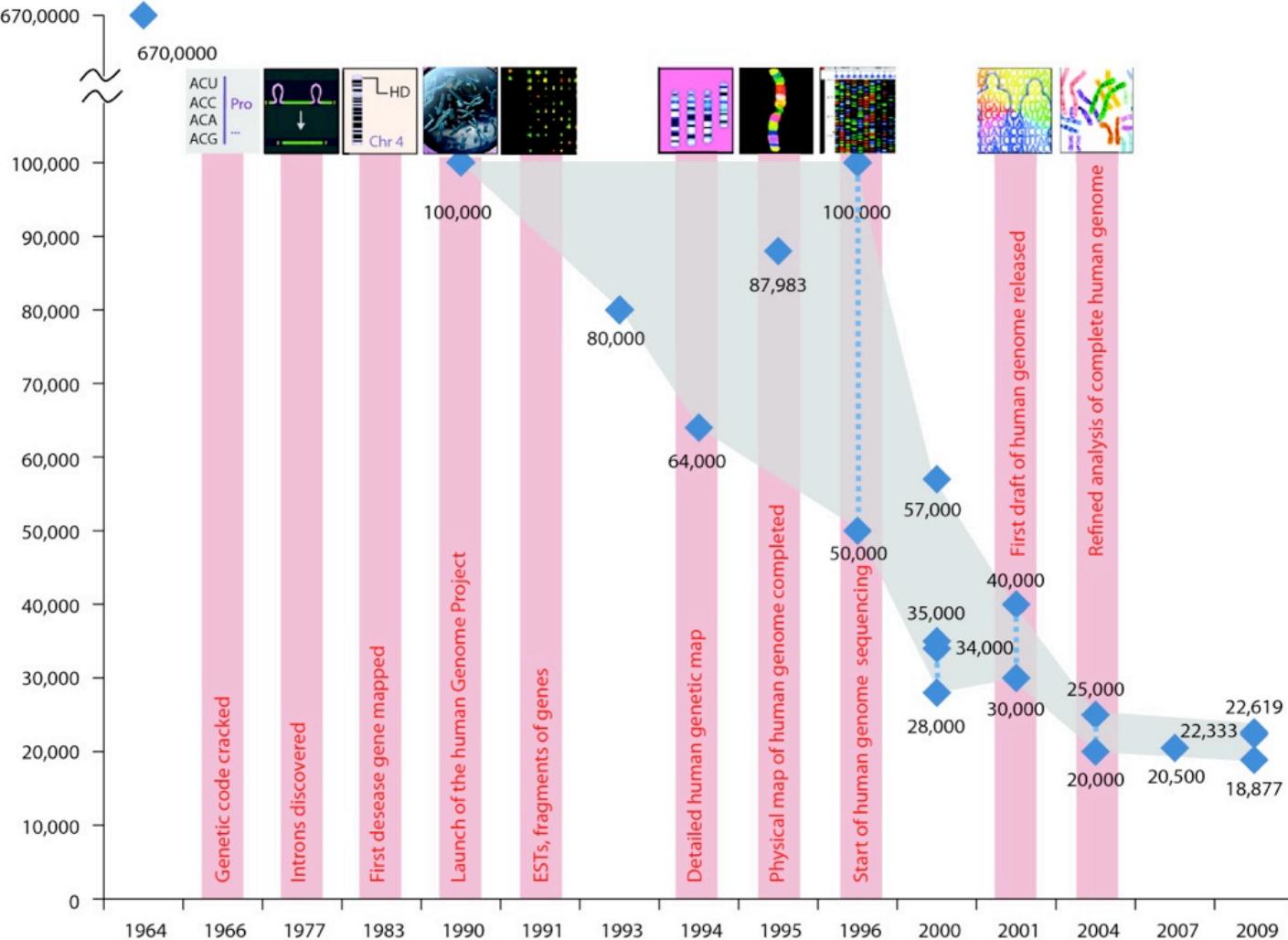




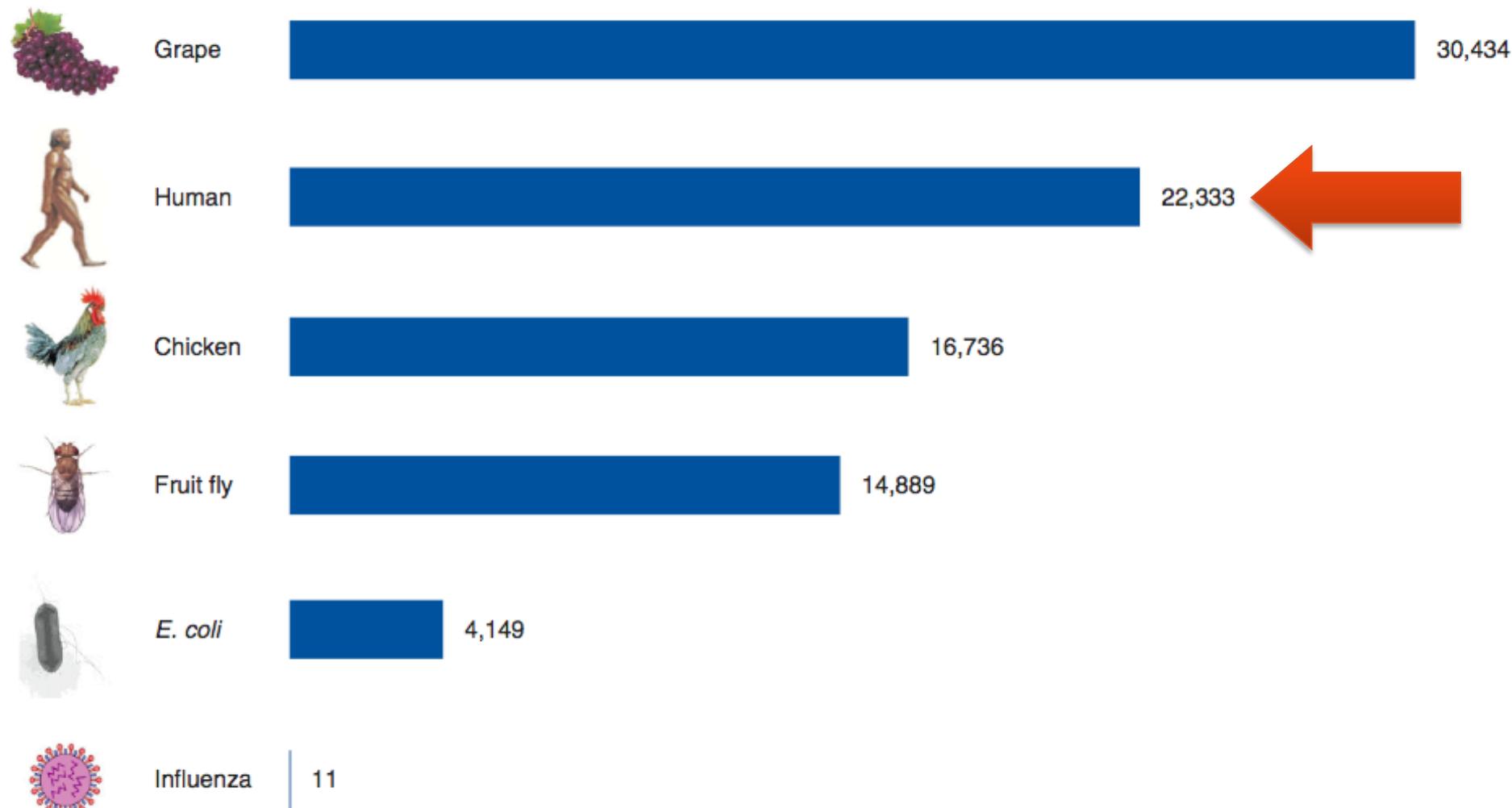
Within each cell

2 genome copies  
23 chromosomes  
~20,000 genes  
3 billion bases

# Somewhere around 22,000 genes?

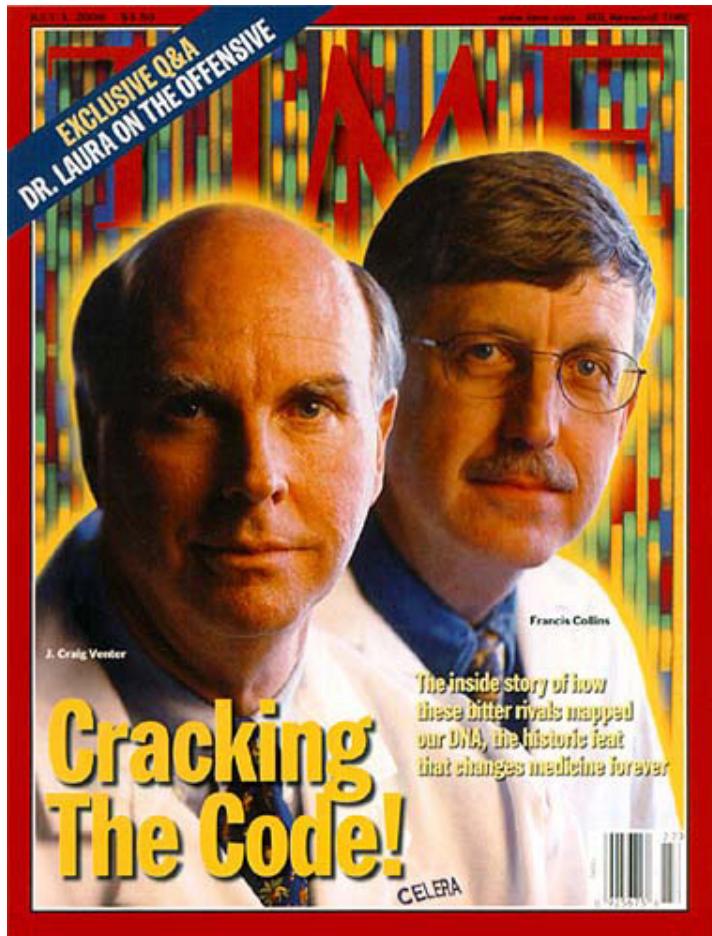
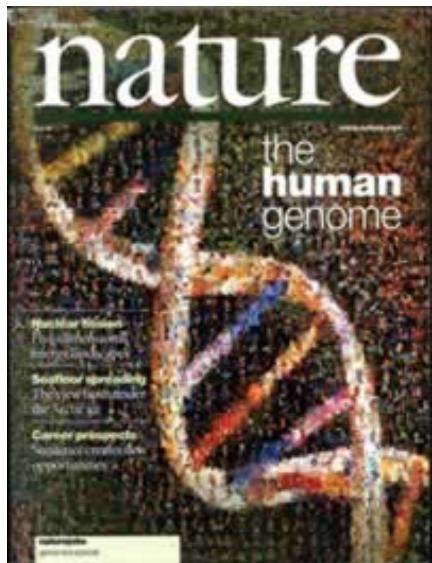


# The numbers compared



# The Human Genome Project

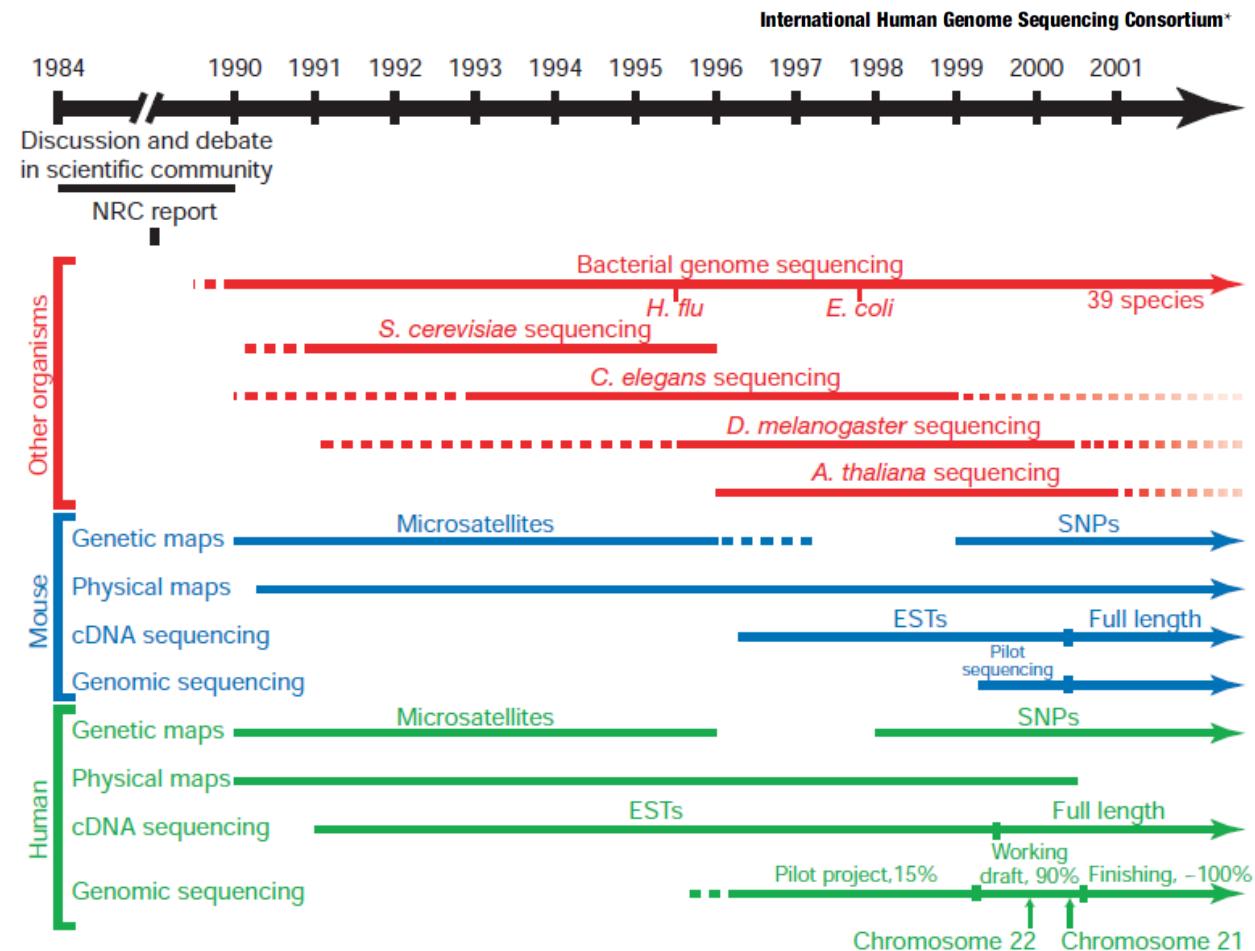
University Medical Center  
Utrecht



# Human Genome Project: 2001

## articles

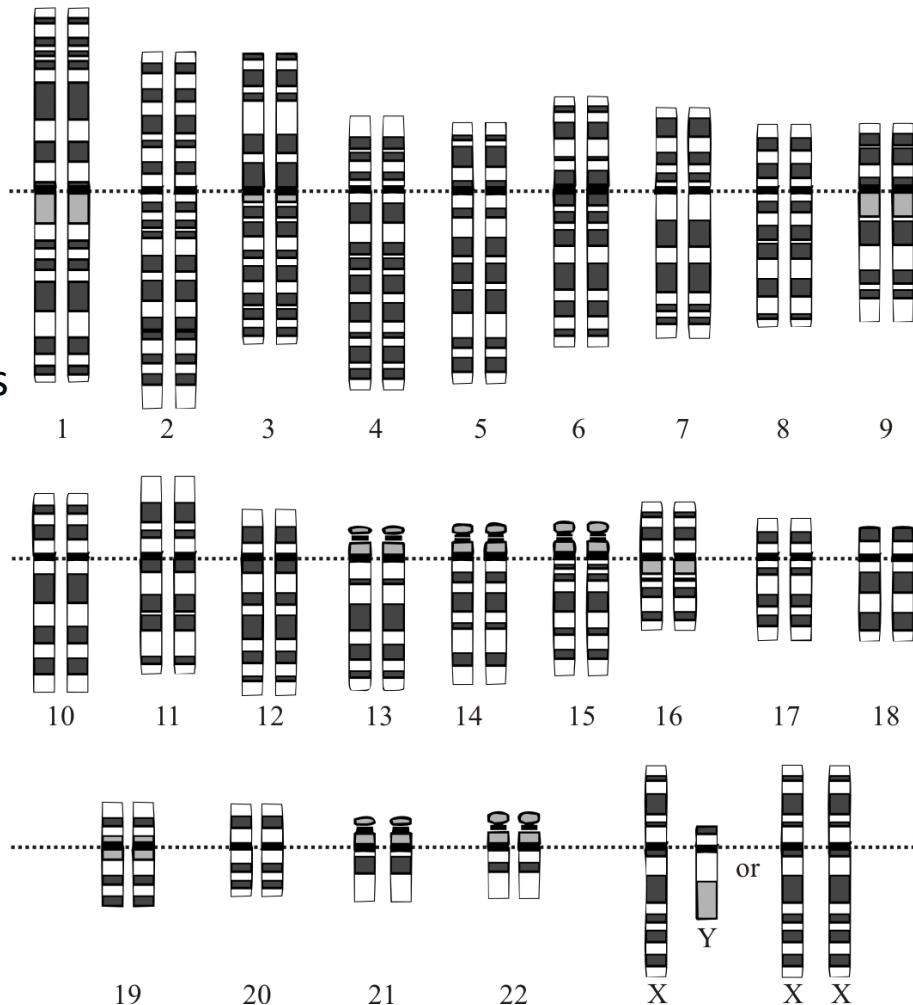
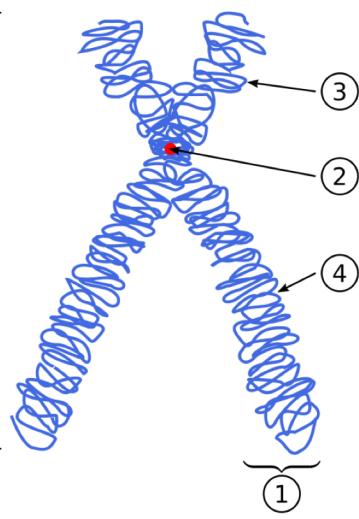
# Initial sequencing and analysis of the human genome



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

# 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



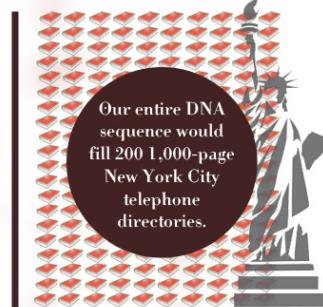
# Human Genome: *some statistics*

- 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.

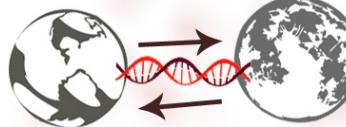


A complete 3 billion base genome would take **3 GIGABYTES OF STORAGE SPACE.**



**99.9%** OF OUR DNA SEQUENCE IS THE SAME AS OTHER HUMANS'.

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



99%

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 99.9% similar between individuals
- 0.1% 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.

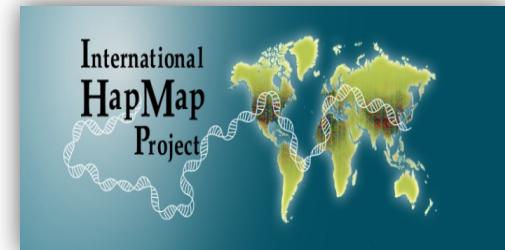
---

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.

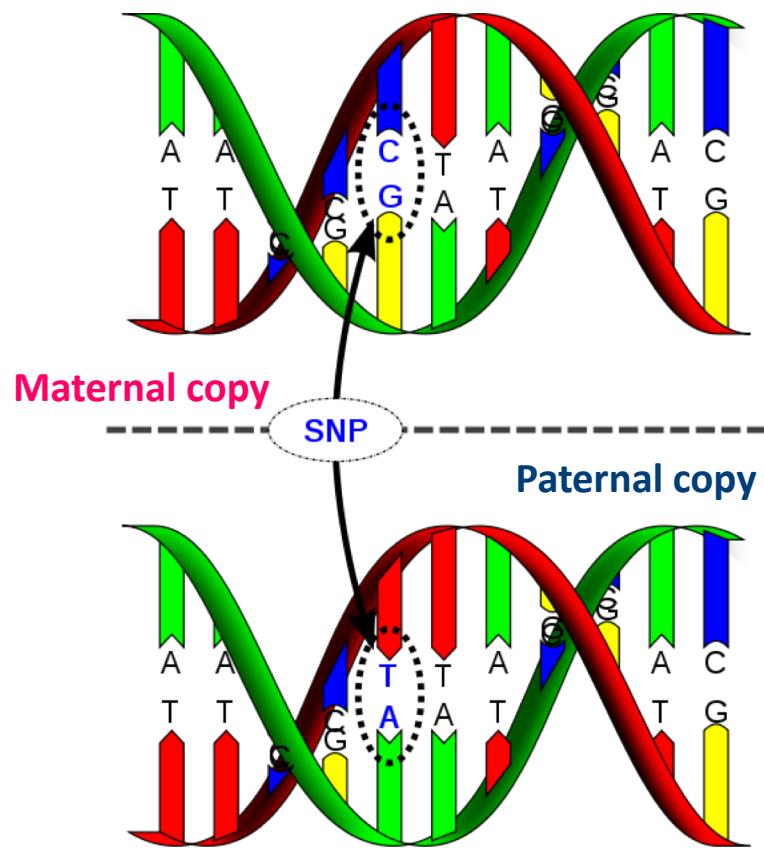
# Single-Nucleotide Polymorphism

*a genetic variation as proxy*

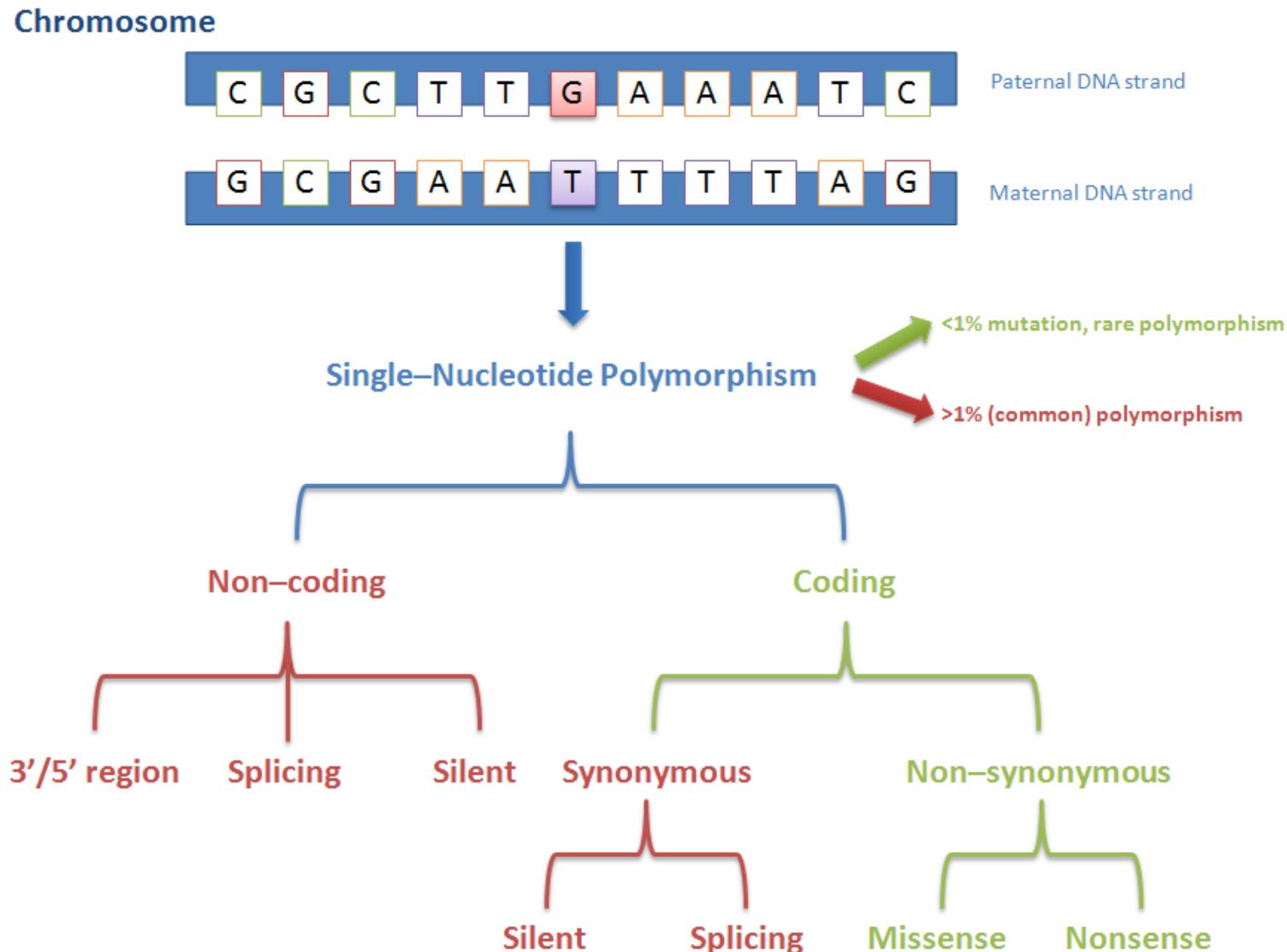
- Single-nucleotide polymorphism (SNP)
- “one base pair variation”
  - > 1% general population (common)
  - ≈10 million SNPs ( $\approx 0.25\%$  genome)
  - Makes you and me unique
- SNPs are common variants which are used as proxies of the actual genomic variation



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



# Types of SNPs



# 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					

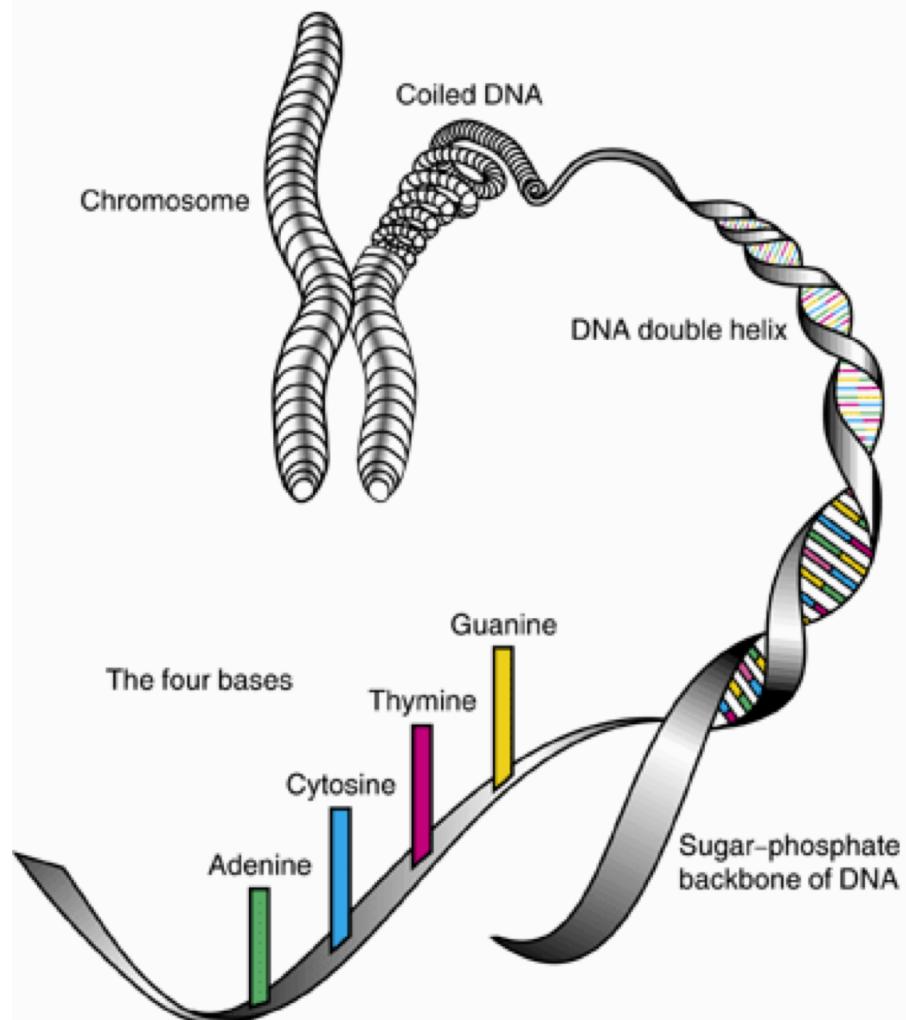
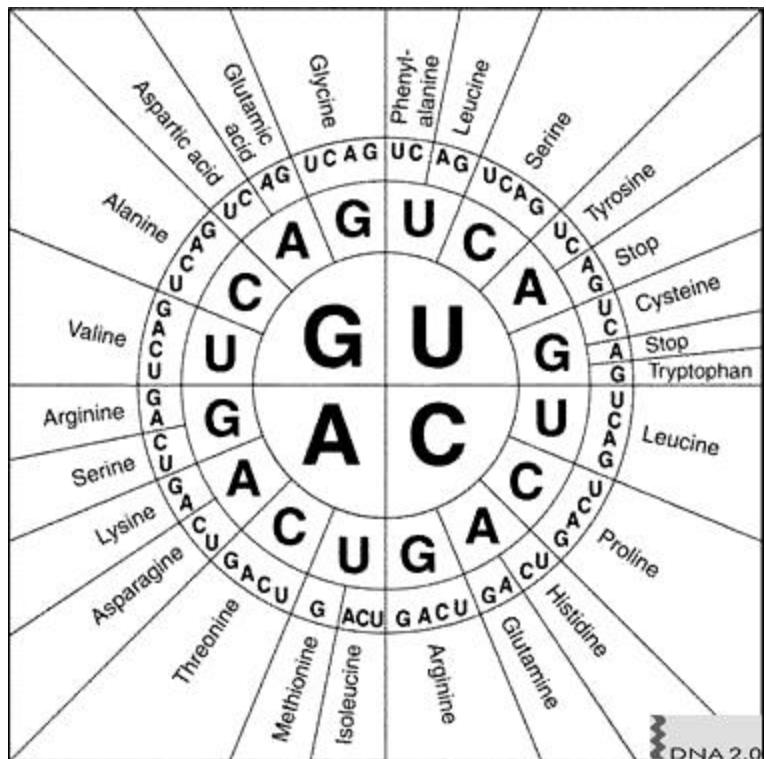
Diagram illustrating single nucleotide substitutions and their effects on protein structure:

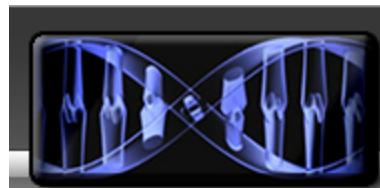
- Normal:** ATG CCG GAC (Methionine) → TCG TTT CTC GGG (Serine).
- A Silent:** ATG CCG GAO (Methionine) → TCA (Proline). (G>A, Ser to Ser)
- B Missense:** ATG CCG GAC (Methionine) → CCG (Proline). (T>C, Ser to Pro)
- C Nonsense:** ATG CCG GAC (Methionine) → TAG (Termination). (C>A, Ser to Ter)

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





# PolyPhen-2 prediction of functional effects of human nsSNPs

[Home](#)[About](#)[Help](#)[Downloads](#)[Batch query](#)[dbSNP query](#)

**PolyPhen-2 (Polymorphism Phenotyping v2)** is a tool which predicts possible impact of an amino acid substitution on the structure and comparative considerations. Please, use the form below to submit your query.

**dbSNP query** provides quick access to precomputed **PolyPhen-2** annotations for a missense human SNPs subset of **NCBI dbSNP** bu

**08-May-2011:** This dataset was obtained using **PolyPhen-2** v2.0.22. Updated predictions for **dbSNP build 132 / PolyPhen-2 v2.1.0 wi**

## Search for reference SNP

rsID

Software & web support: ivan adzhubey

# Imputing Untyped Variants

C.C.T.CC.A  
A.C.A.C..A  
C.T.A.C..G



1000 Genomes  
data

C....T....A  
A....A....A  
C....A....G

A.......  
C....T....G

...

C.C.T.CC...A  
A.C.A.C...A  
C.T.A.C...G

A.C.A.C...A  
C.C.T.CC/C.G

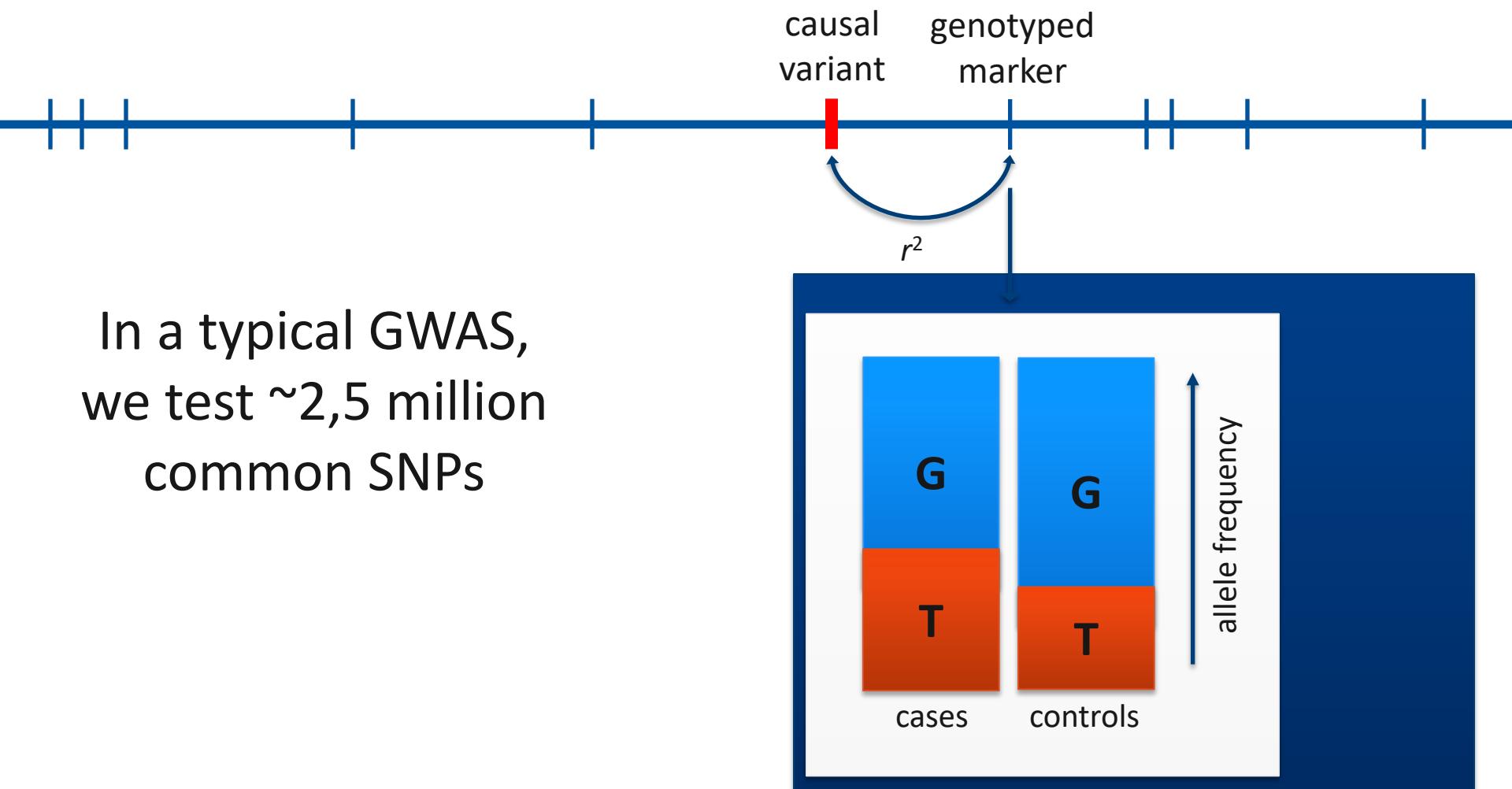


GWAS genotype  
data

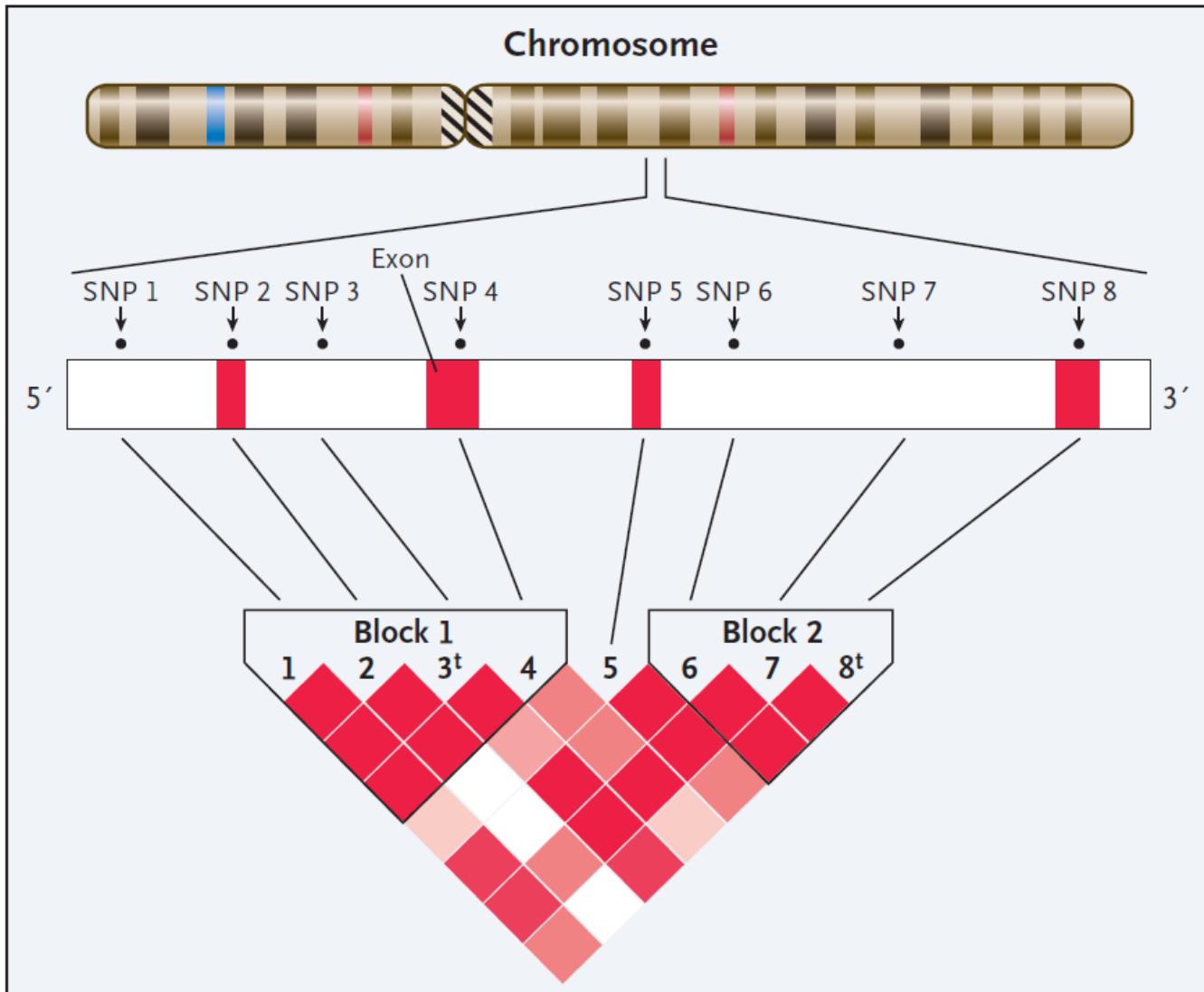
GWAS with  
imputed data  
**Free!**

# Linkage disequilibrium

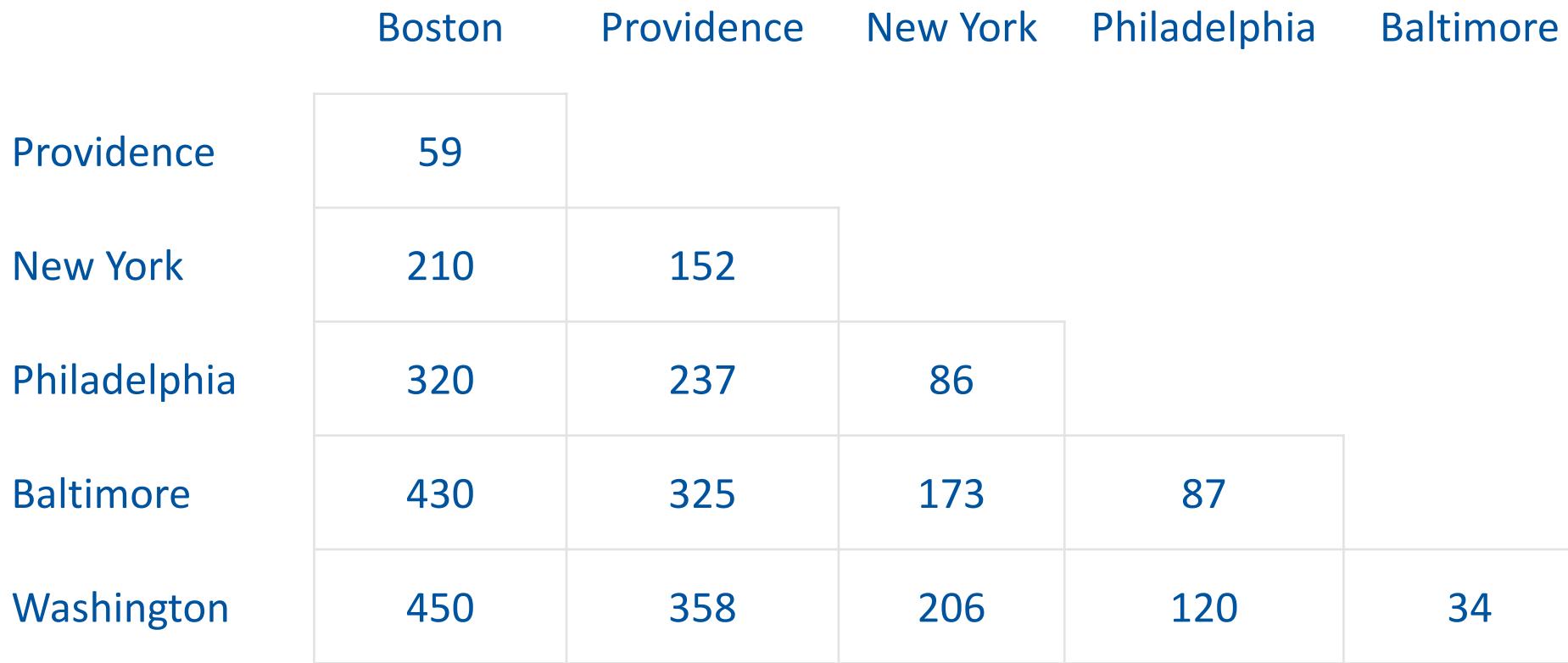
Non-random association of alleles at two or more loci



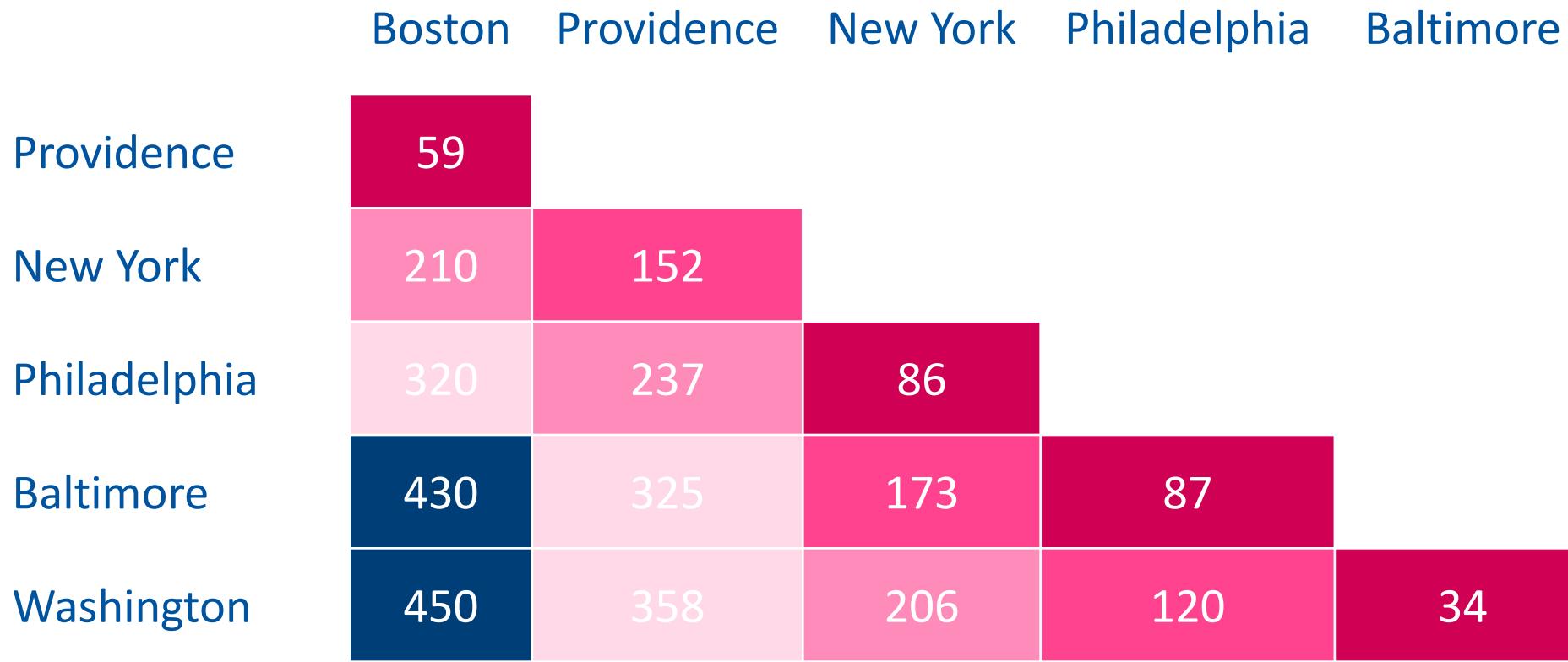
## LD: blocks of correlated SNPs



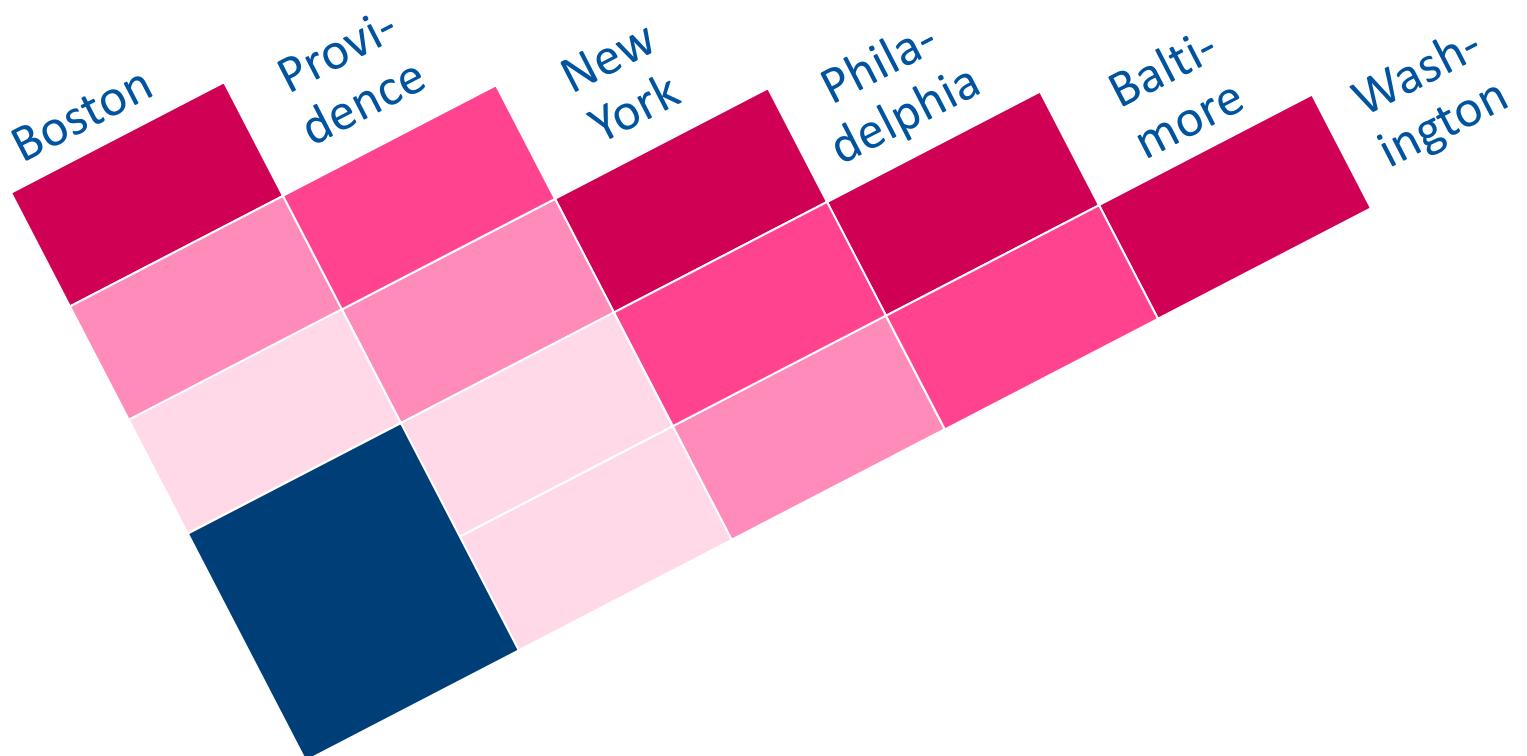
# Distances among East coast cities



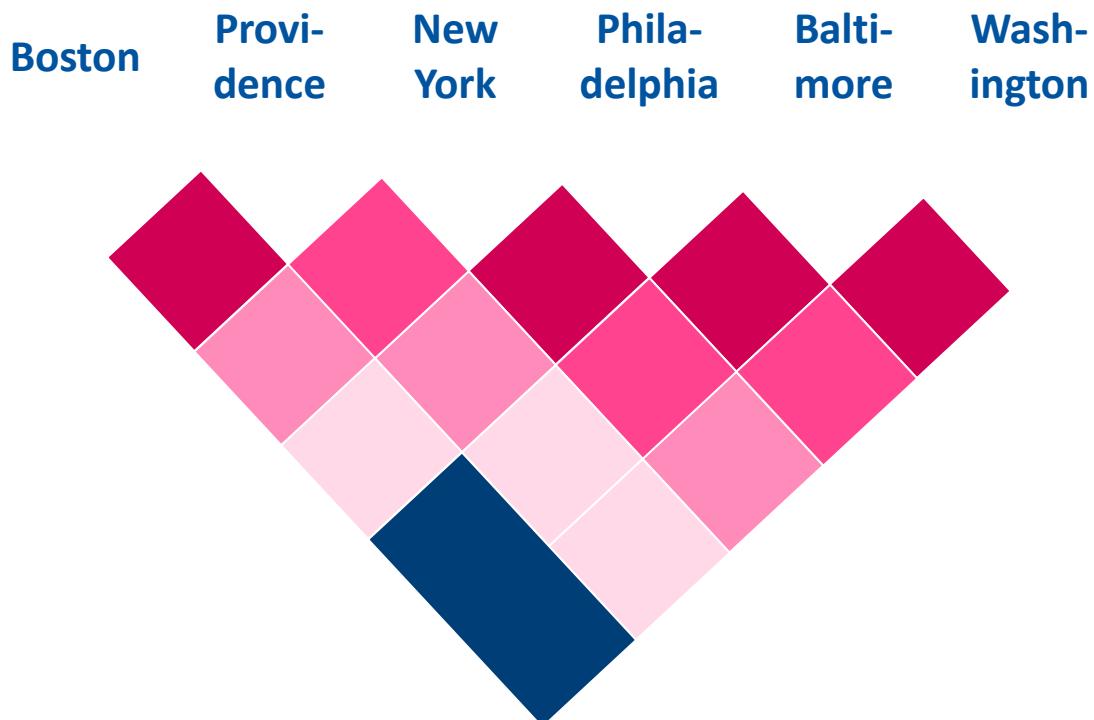
# Distances among East coast cities



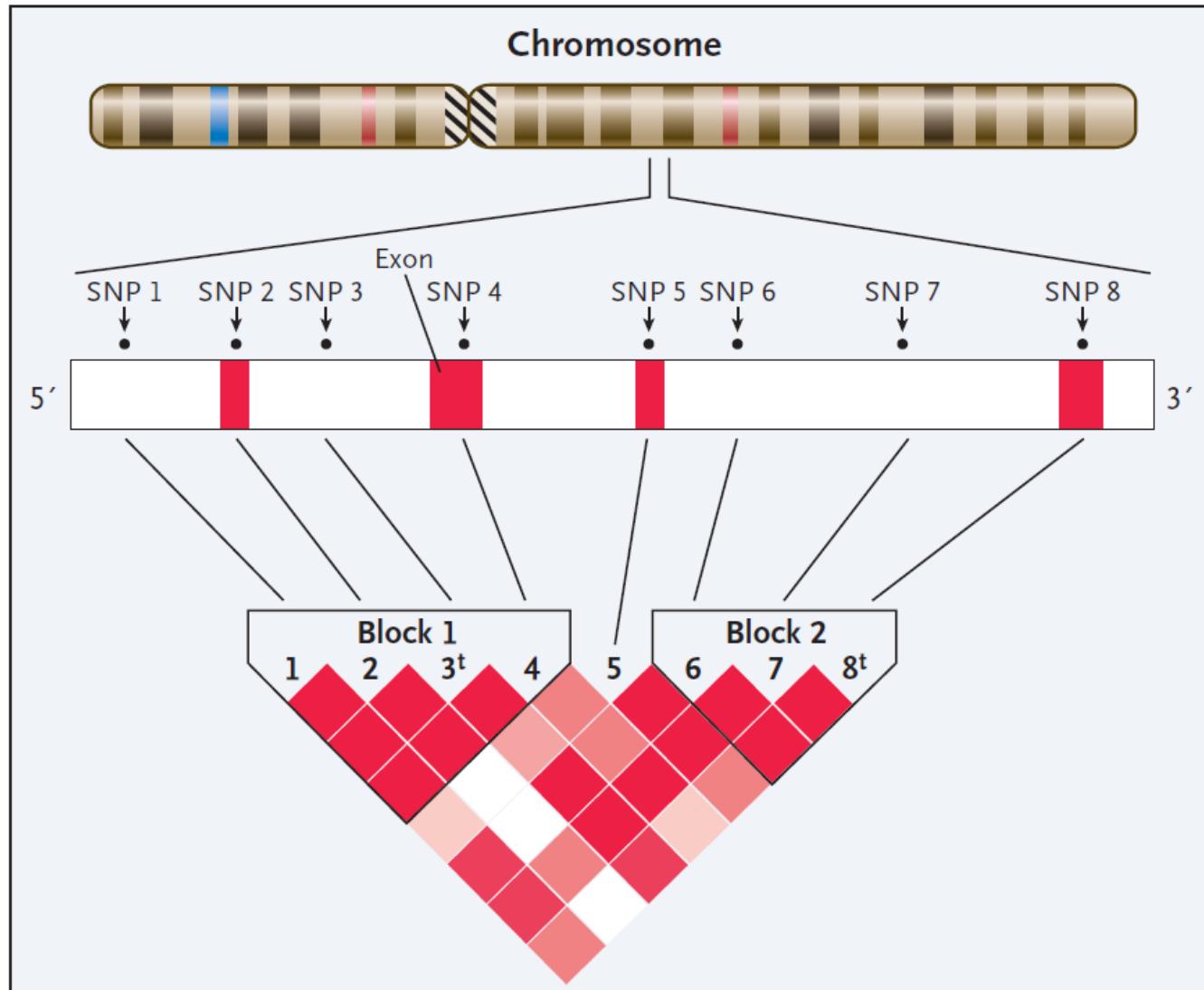
# Distances among East coast cities



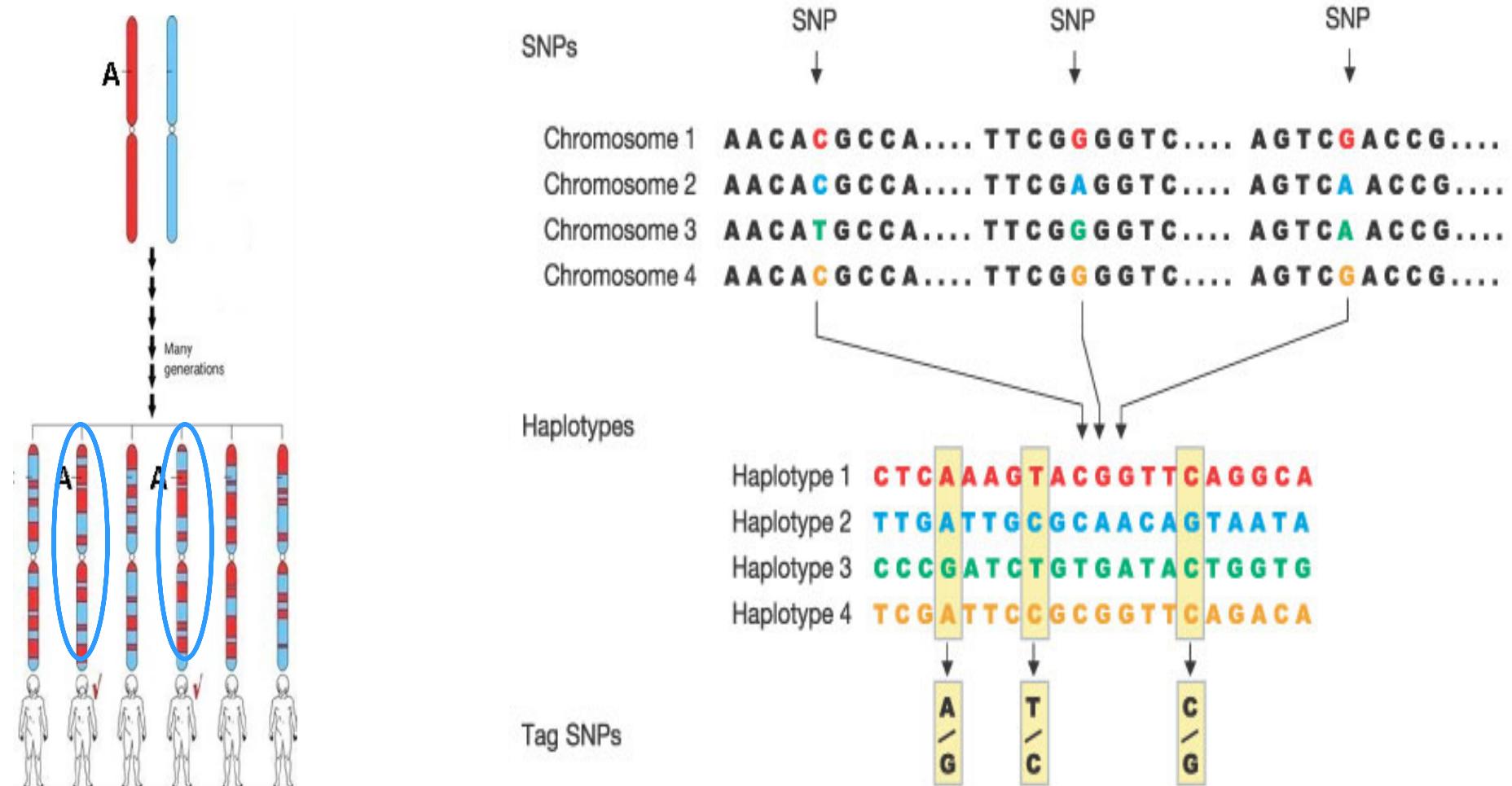
# Distances among East coast cities



# LD: blocks of correlated SNPs



# SNPs → Haplotype → tagSNP

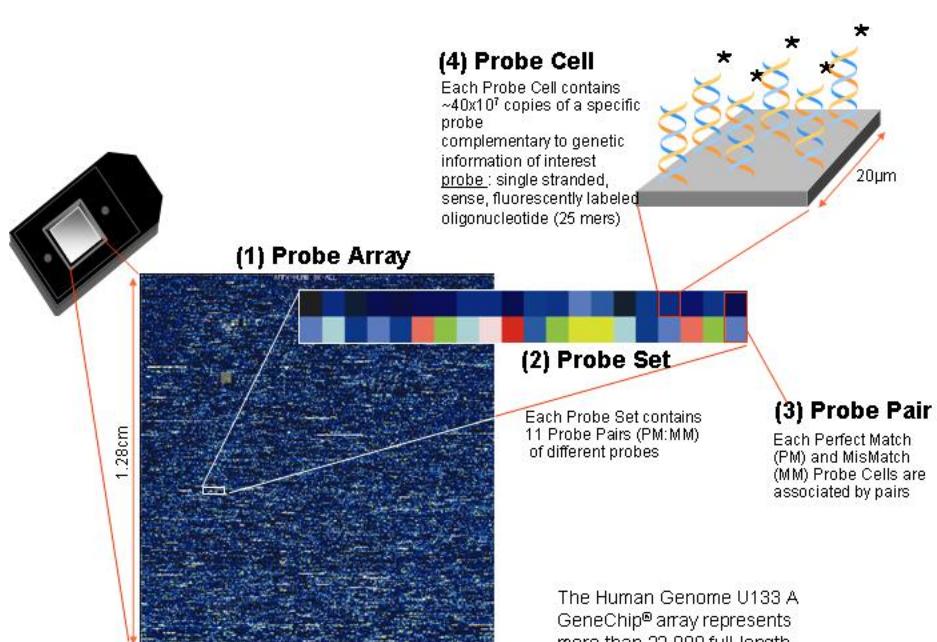
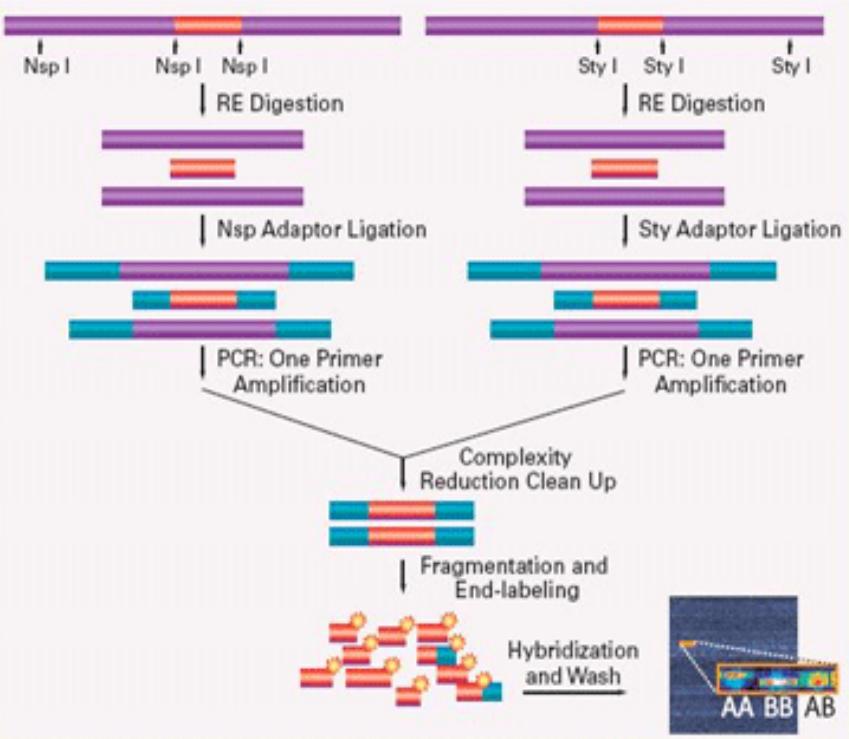


# Genotyping Platforms



# SNP “genotyping”

The fifth-generation Whole-genome Sampling Assay.



# Genotyping Platforms: *examples*

- Affymetrix Genome-Wide Human SNP Array 5.0
  - 500,568 SNPs
  - Chromosomal coverage
    - Good representation of autosomal chromosomes
    - X-chromosome poor representation
    - No Y-chromosome representation
    - No mitochondrial chromosome representation
- Illumina Human660W-Quad v1
  - 657,366 SNPs
  - Chromosomal coverage
    - Good representation of autosomal chromosomes
    - X-chromosome reasonable representation
    - Y-chromosome poor representation
    - Mitochondrial chromosome poor representation



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?

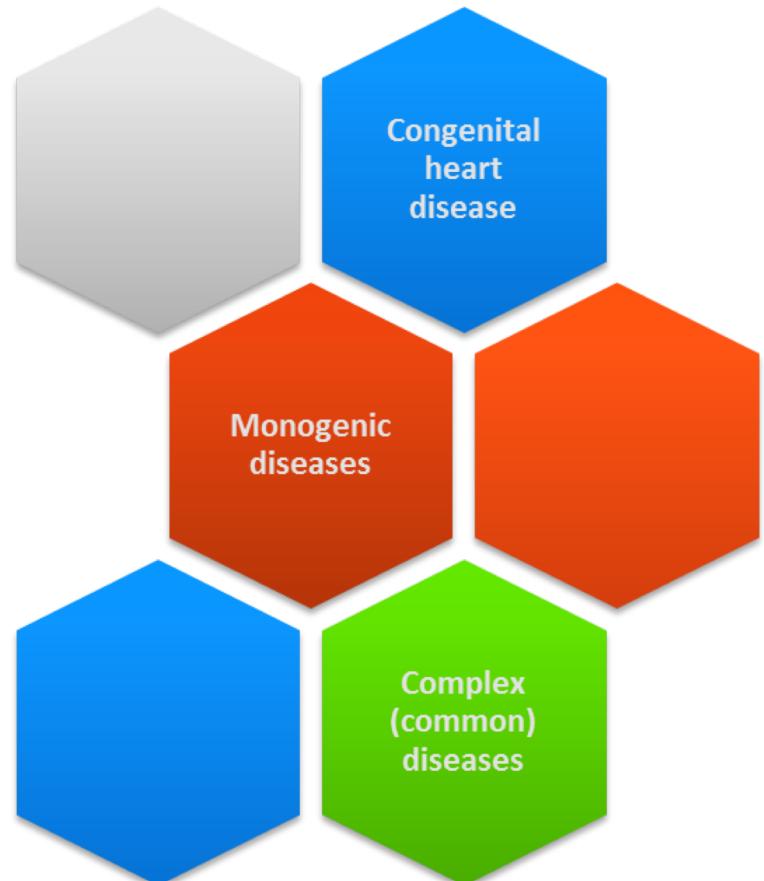
## *monogenic diseases*

- **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

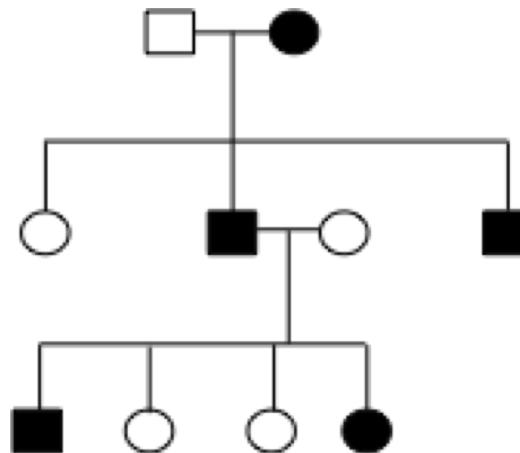
## Genotype



## 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**

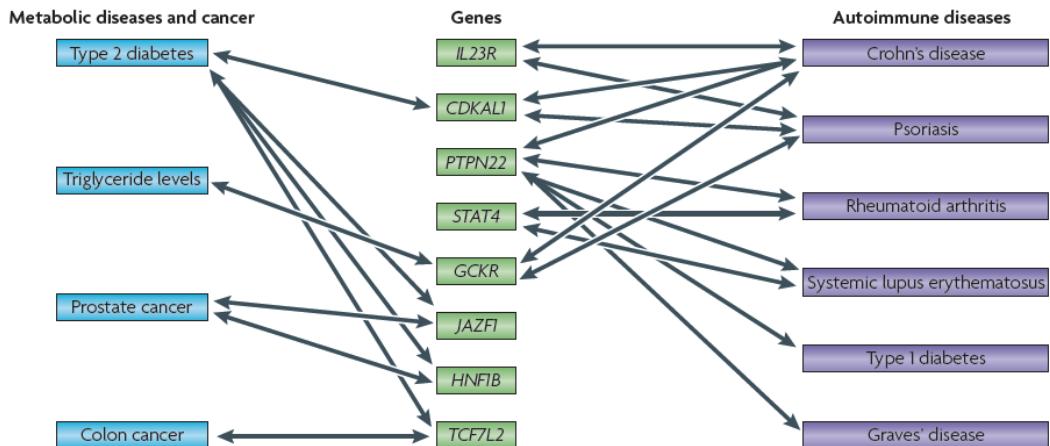


Table 3

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

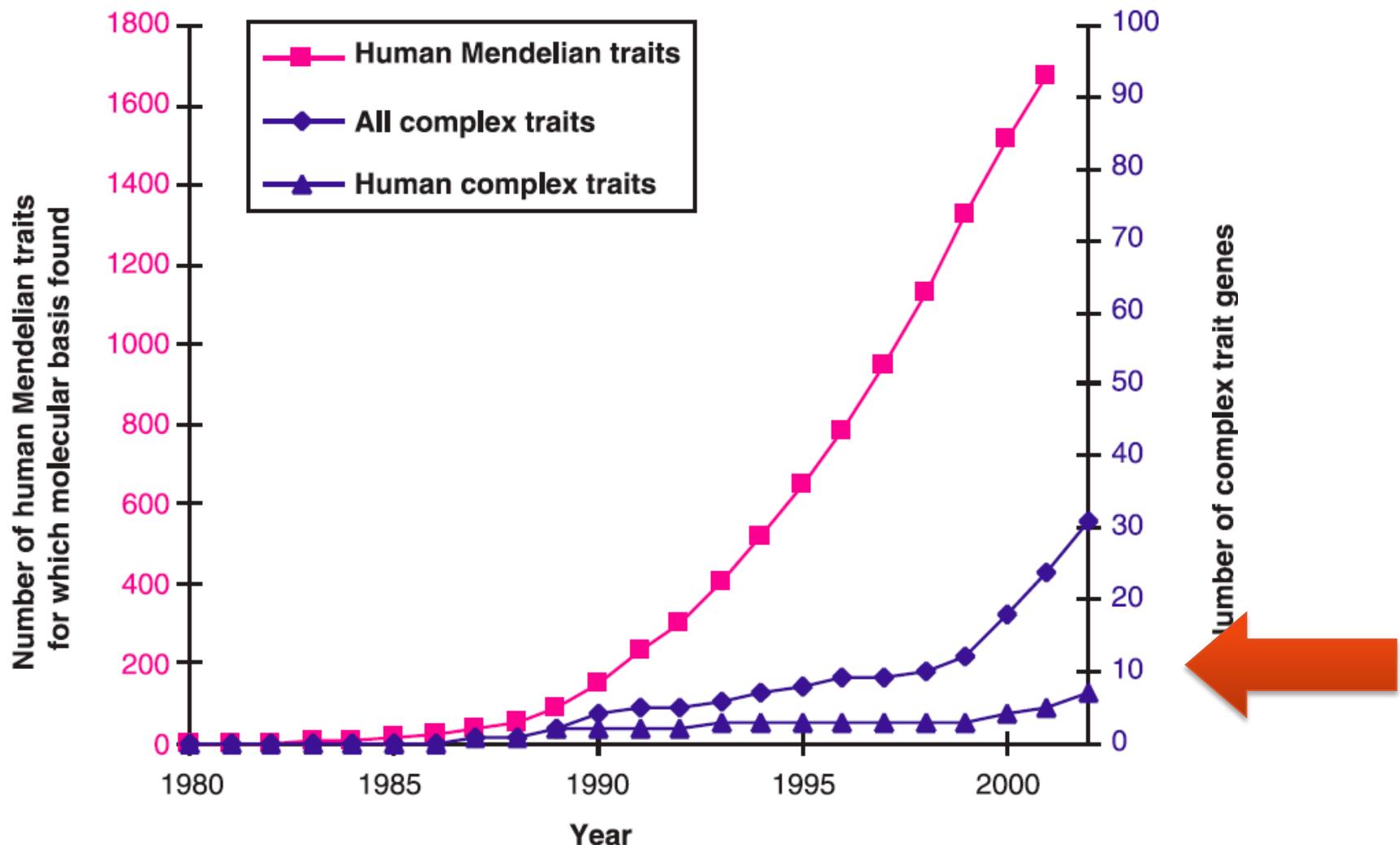
Polygenic

Genotypes → Disease

Environment



# Early successes genomic research of monogenic disease



# Challenges of common disease research

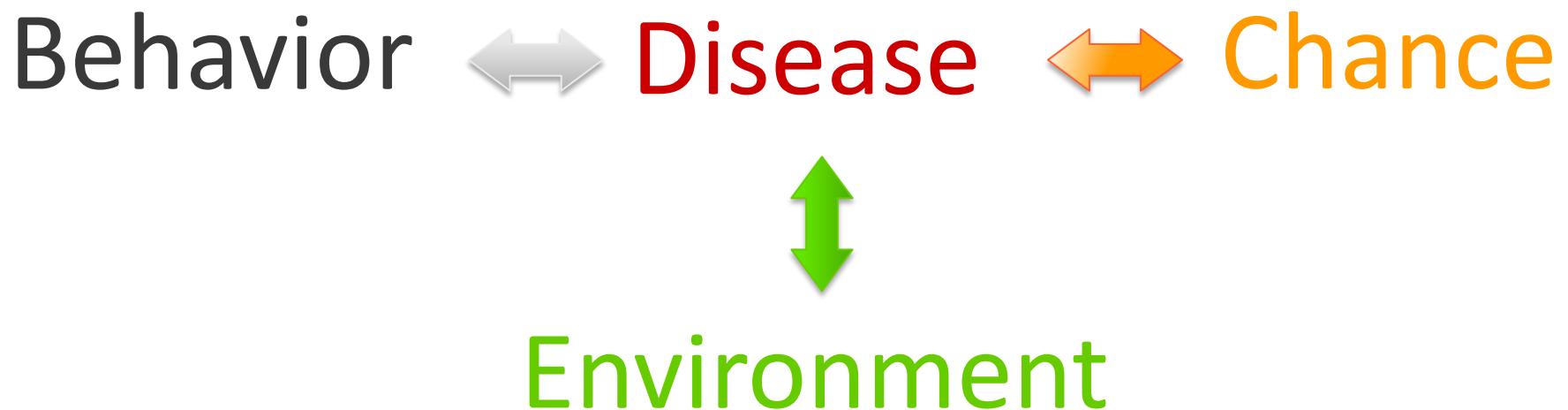
- Involvement of many genes and environment
  - Each plays only a small role (consequence - traditional approaches designed to find major genetic effects are ineffective – no such effects exist!)
  - Mutations are neither necessary nor sufficient on their own to cause disease
  - Environmental triggers 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

# Catalogs of inherited DNA variation

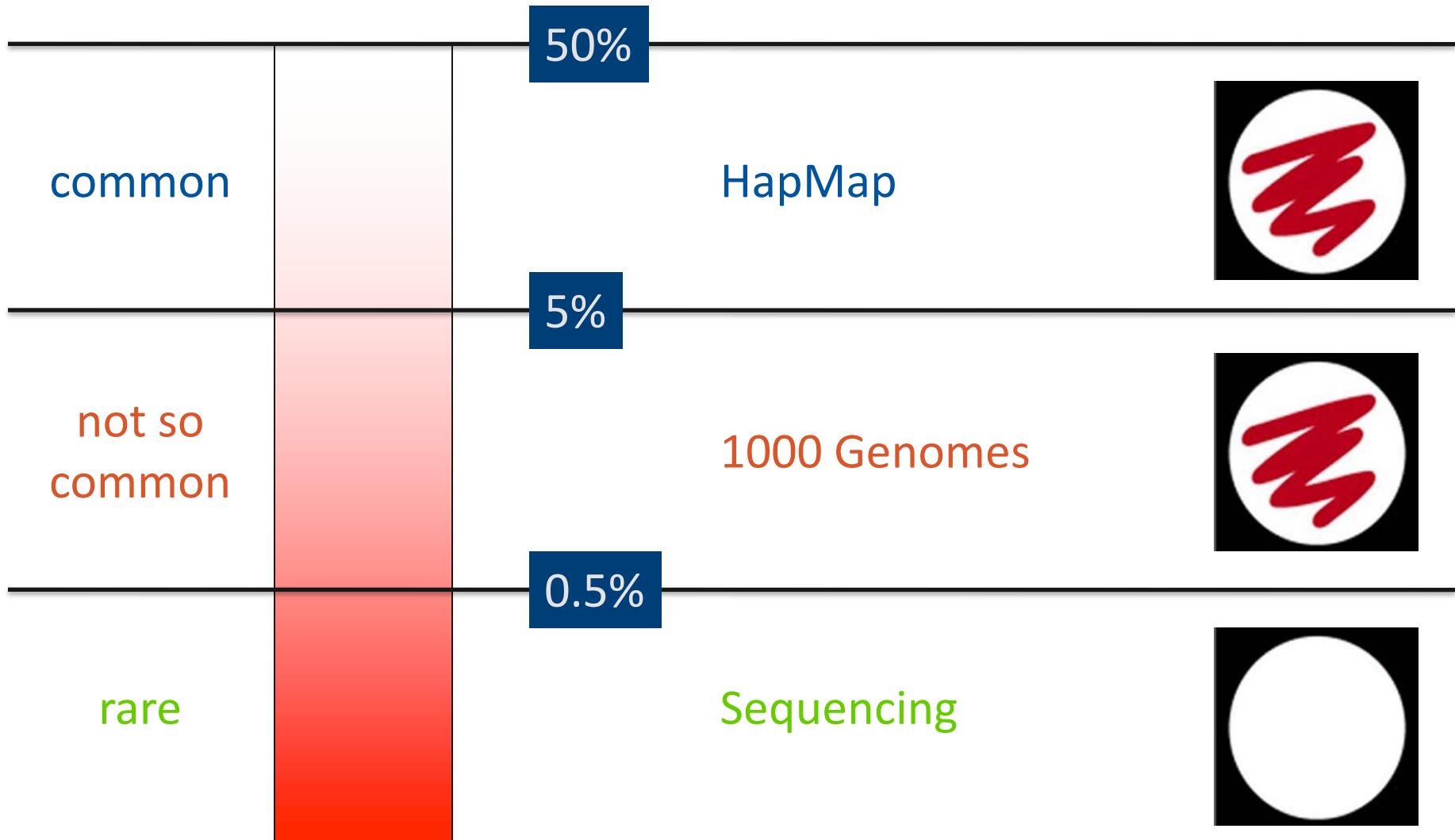
University Medical Center  
Utrecht

- 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?

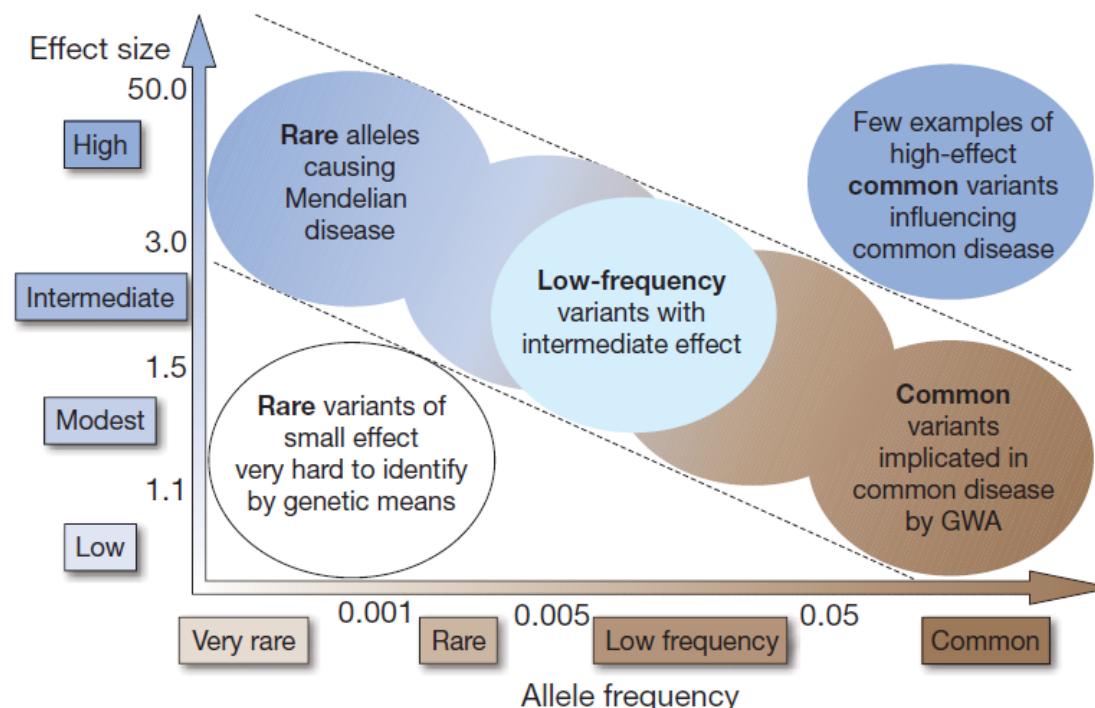
University Medical Center  
Utrecht



# 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!)



# What type of study design can you choose?

## *genome-wide association study*



University Medical Center  
Utrecht

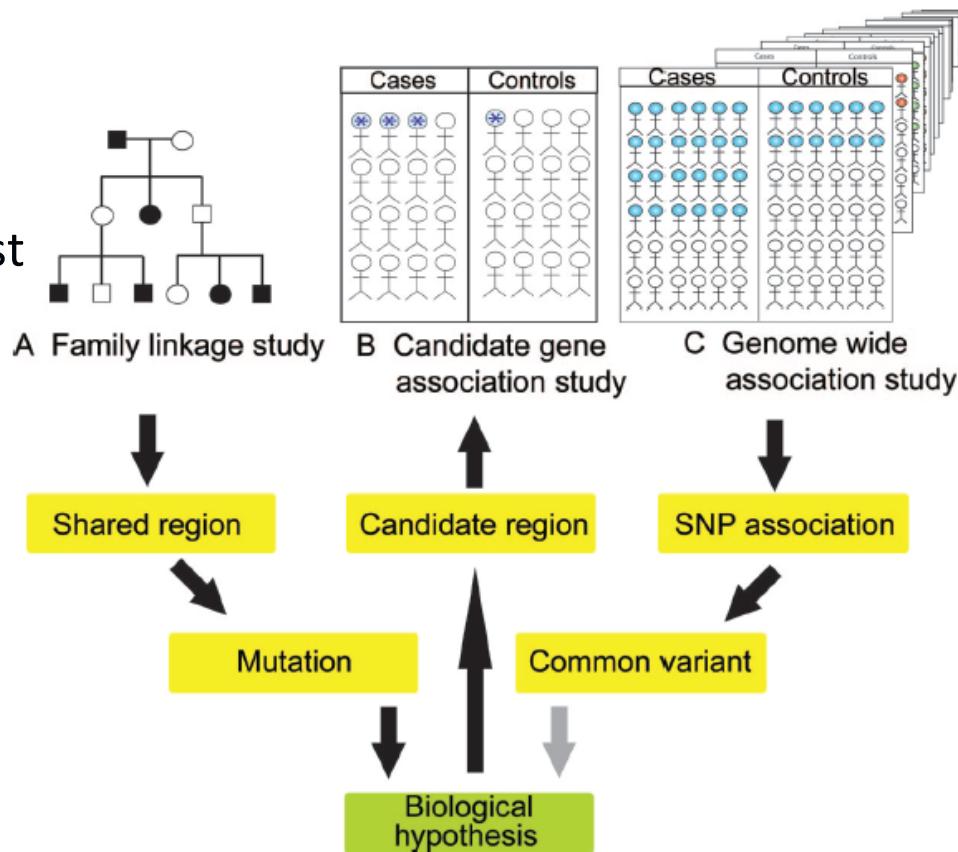
**Table 1.** Study Designs Used in Genome-wide Association Studies

	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



- >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 Gudmundsson<sup>1</sup>, Struan F A Grant<sup>1</sup>, Gudmundur Thorgeirsson<sup>3</sup>, Sigurlaug Sveinbjornsdottir<sup>3</sup>, Einar M Valdimarsson<sup>3</sup>, Stefan E Matthiasson<sup>3</sup>, Halldor Johannsson<sup>3</sup>, Olof Guðmundsdóttir<sup>1</sup>, Mark E Gurney<sup>1</sup>, Jesus Sainz<sup>1</sup>, Margaret 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 Blonadal,<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 Matthiassdottir,<sup>1</sup> Thorbjorg Jonsdottir,<sup>1</sup> Stefan Palsson,<sup>1</sup> Helga Einarsdottir,<sup>1</sup> Steinunn Gunnarsdottir,<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 Thorsteinsdottir,<sup>1</sup> Augustine Kong,<sup>1,†</sup> Kari Stefansson<sup>1</sup>

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

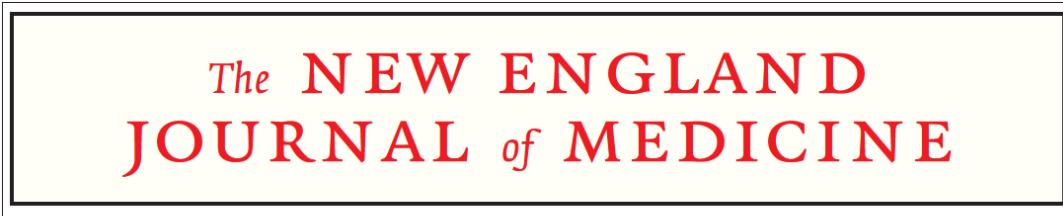
Ruth McPherson,<sup>1,\*†</sup> Alexander Pertsemlidis,<sup>2,\*</sup> Nihan Kavaslar,<sup>1</sup> Alexandre Stewart,<sup>1</sup> Robert Roberts,<sup>1</sup> David R. Cox,<sup>3</sup> David A. Hinds,<sup>3</sup> Len A. Pennacchio,<sup>4,5</sup> Anne Tybjaerg-Hansen,<sup>6</sup> Aaron R. Folsom,<sup>7</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

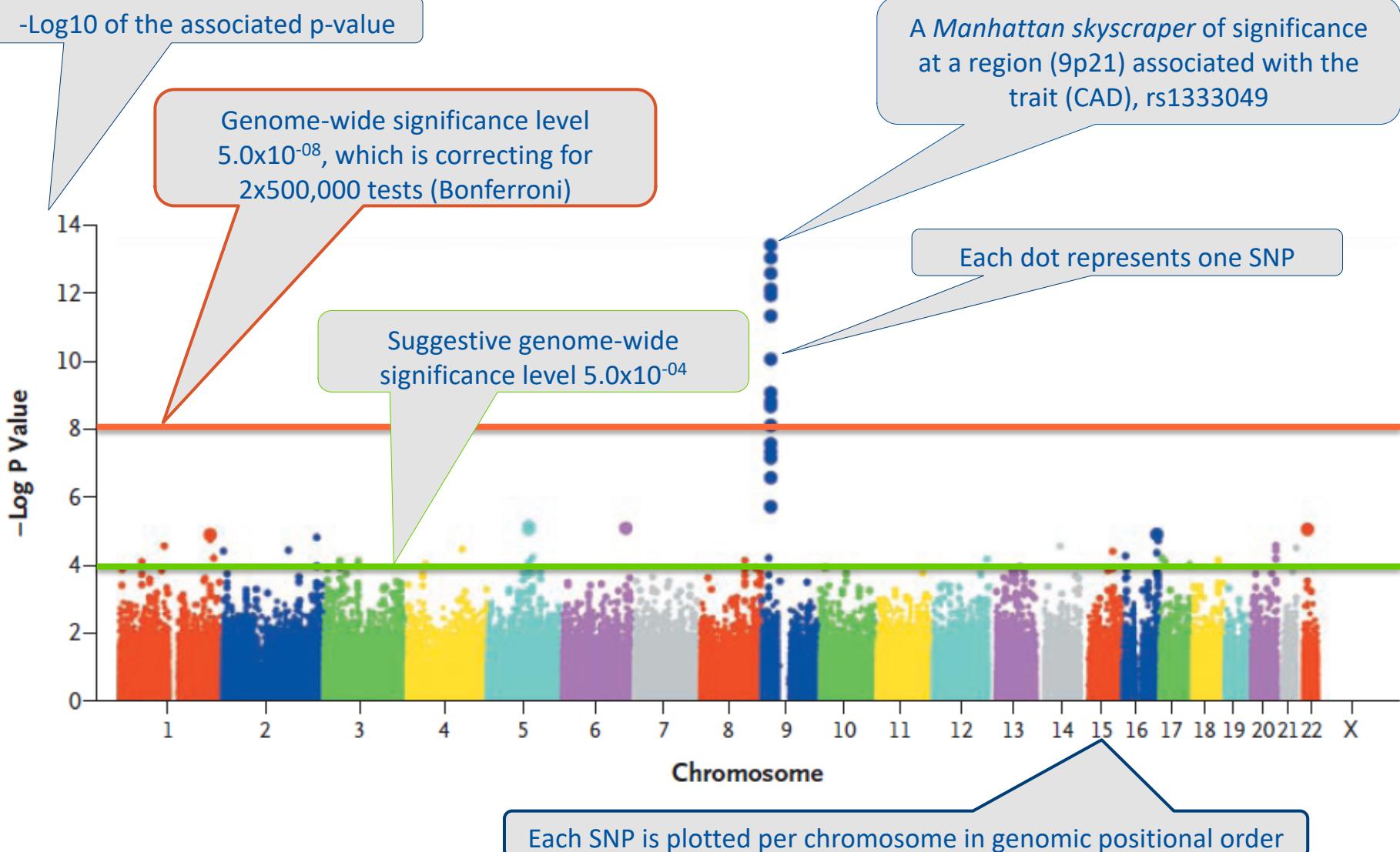
ARTICLES

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

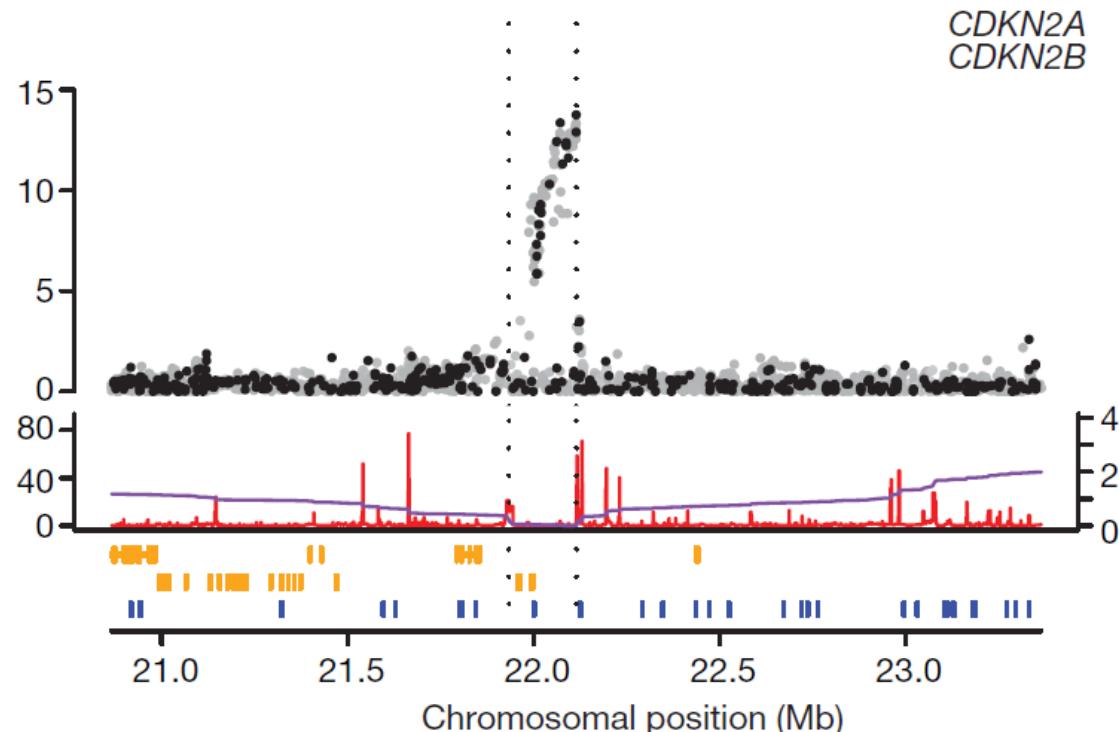


# One famous example

## 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

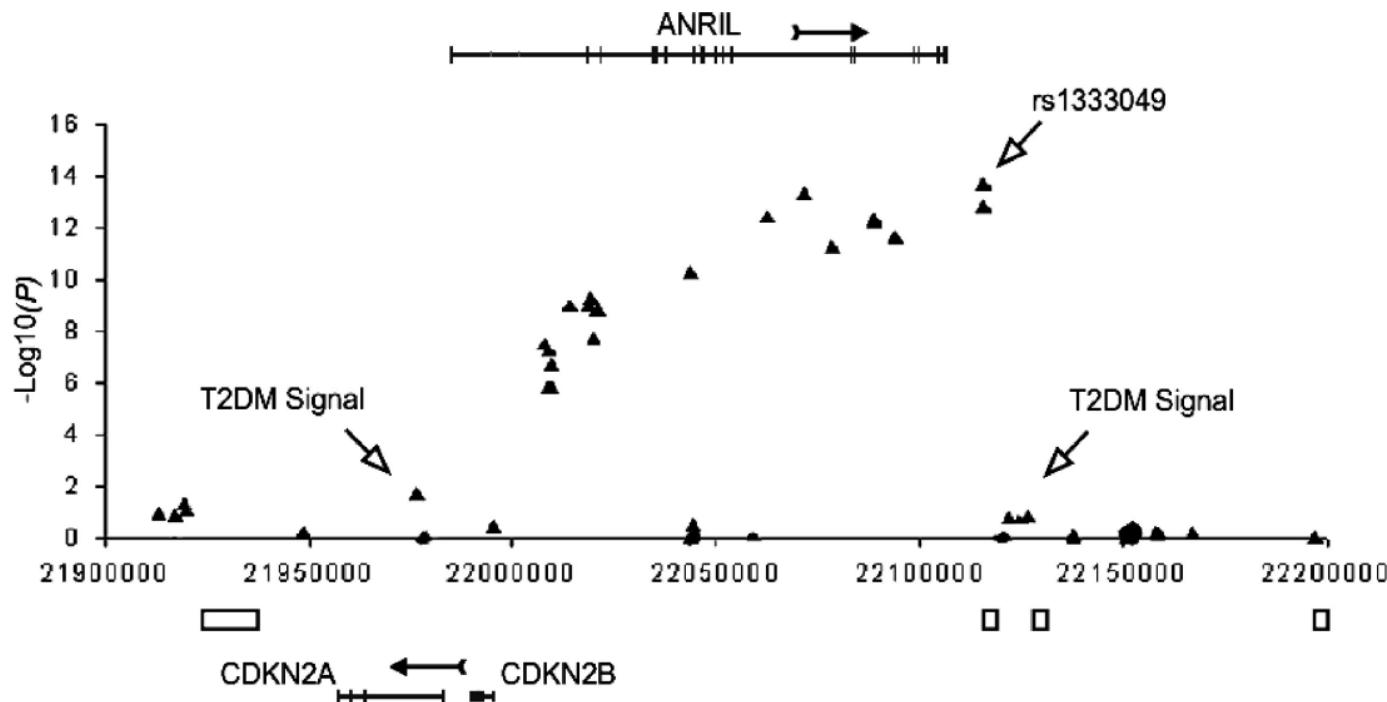
Risk allele: minor allele  
 Odds ratio: the odds of exposure between cases and controls

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

# deCODE MI™

*provides a modified 10 year CHD risk*



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

Blood Pressure		
Systolic Diastolic (mmHg)		
HDL - Cholesterol	Age	
(mg/dL) (mmol/L)	Years	Points
< 35 ≤ 0.90	30-34	-9
35-44 0.91-1.16	35-39	-4
45-49 1.17-1.29	40-44	0
50-59 1.30-1.55	45-49	3
≥ 60 ≥ 1.56	50-54	6
	55-59	7
	60-64	8
	65-69	8
	70-74	8

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

LDL - Cholesterol		
Key		
(mg/dL) (mmol/L)	Color	
< 100 ≤ 2.59	Risk	
100-129 2.60-3.36	Green Very low	
130-159 3.37-4.14	White Low	
160-189 4.15-4.91	Yellow Moderate	
≥ 190 ≥ 4.92	Rose High	
	Red Verh high	

## 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.

This test was developed and its performance characteristics determined by the deCODE genetics Diagnostic Laboratory. It has not been cleared or approved by the U. S. Food and Drug Administration (FDA). deCODE Diagnostics Laboratory – Testing Site: St. Lucie & 101 Raynfield, Isla

Customer Service: 15700 W. 103rd St. Suite 200, Lemont IL 60439 – Phone: (877) 222-8510 Fax: (830) 785-0998 – [www.decodediagnostic.com](http://www.decodediagnostic.com)

This document contains private and confidential health information protected by state and federal law. ©2009 deCODE genetics Diagnostic Laboratory

All rights reserved – document version 2.1

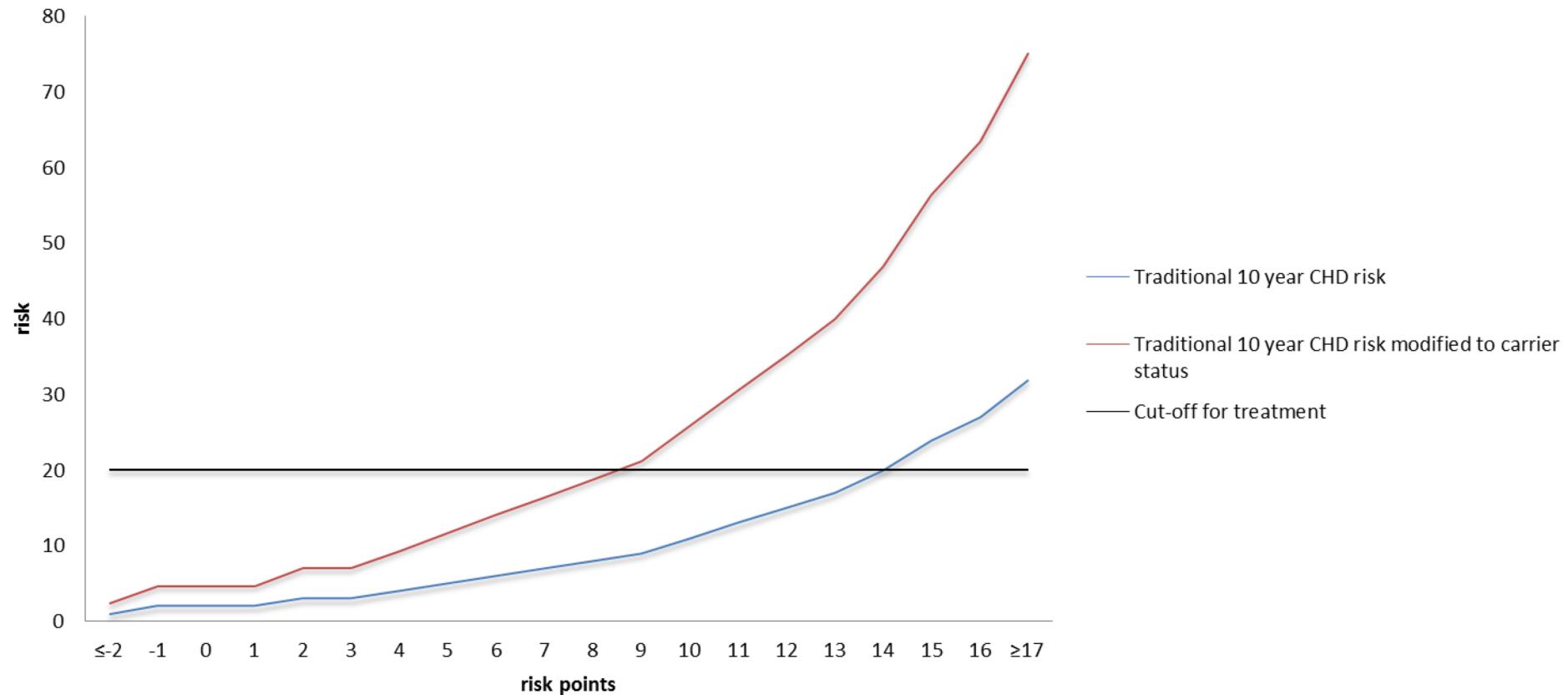
## CHD Traditional Risk & Your Modified Risk

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





University Medical Center  
Utrecht

Genetic Burden of Disease (Risk)

# GENETIC BURDEN SCORES



# CARDIoGRAM Study

- Coronary Artery Disease Genome–Wide Replication And Meta–Analysis Study: CARDIoGRAM
- > 20,000 cases and > 60,000 controls
  - Myocardial infarction (MI), coronary artery disease (CAD) or both
  - CAD: MI, CABG, PTCA, AP
  - Age limit: 45–66
- Recent studies uncovered around 13 variants associated with MI/CAD
- Sample size greatly influences power and effect size to discover new variants
- CARDIoGRAM sought to solves this issue
- 13 novel susceptibility loci for CAD were discovered

nature  
genetics

## Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease

We performed a meta-analysis of 14 genome-wide association studies of coronary artery disease (CAD) comprising 22,233 individuals with CAD (cases) and 64,762 controls of European descent followed by genotyping of top association signals in 56,682 additional individuals. This analysis identified 13 loci newly associated with CAD at  $P < 5 \times 10^{-8}$  and confirmed the association of 10 of 12 previously reported CAD loci. The 13 new loci showed risk allele frequencies ranging from 0.13 to 0.91 and were associated with a 6% to 17% increase in the risk of CAD per allele. Notably, only three of the new loci showed significant association with traditional CAD risk factors and the majority lie in gene regions not previously implicated in the pathogenesis of CAD. Finally, five of the new CAD risk loci appear to have pleiotropic effects, showing strong association with various other human diseases or traits.

It has been estimated that heritable factors account for 30%–60% of the inter-individual variation in the risk of coronary artery disease (CAD)<sup>1</sup>. Recently, genome-wide association studies (GWAS) have identified several common variants that associate with risk of CAD<sup>2</sup>. However, in aggregate, these variants explain only a small fraction of the heritability of CAD, probably partly due to the limited power of previous studies to discover effects of modest size. Recognizing the need for larger studies, we formed the transatlantic Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) consortium<sup>3</sup>. We performed a meta-analysis of 14 GWAS of CAD comprising 22,233 cases and 64,762 controls, all of European ancestry (Supplementary Table 1a–c and Supplementary Fig. 1). We then genotyped the lead SNPs within the most promising previously unidentified loci as well as a subset of previously reported CAD loci in up to 56,682 additional subjects (approximately half cases and half controls) (Supplementary Table 2a,b). Lastly, we explored potential mechanisms and intermediate pathways by which previously unidentified loci may mediate risk.

Nine of the twelve loci previously associated with CAD through individual GWAS achieved genome-wide significance ( $P < 5 \times 10^{-8}$ ) in our initial meta-analysis (Table 1 and Supplementary Table 3). We were, however, unable to test the previously reported association with a haplotype and a rare SNP in *LPA* in our GWAS data<sup>4,5</sup>, but we observed robust association with the rare *LPA* variant in our replication samples through direct genotyping (Table 1).

Thus, 10 of the 12 loci previously associated with CAD at a genome-wide significance level surpassed the same threshold of significance in CARDIoGRAM.

We selected 23 new loci with a significance level of  $P < 5 \times 10^{-6}$  in the meta-analysis for follow up (Online Methods and Supplementary Note). Taking the number of loci into consideration, our replication study had >90% power to detect effect sizes observed in the GWAS meta-analysis. Of the 23 loci, 13 replicated using our *a priori* definition of a validated locus, that is, showing independent replication after Bonferroni correction and also achieving  $P < 5 \times 10^{-8}$  in the combined discovery and replication data (Table 2, Fig. 1 and Supplementary Figs. 2 and 3). Results for all loci from the replication phase are shown in Supplementary Tables 4 and 5.

The 13 new loci had risk allele frequencies ranging from 0.13 to 0.91 and were associated with a 6% to 17% increase in the risk of CAD per allele (Table 2). Out of the 13 new loci, the additive model appeared most appropriate for 6 whereas the recessive model performed best at 5 and the dominant model at 2 loci (Supplementary Table 6).

In sub-group analyses, 20 out of 22 loci with  $P < 5 \times 10^{-8}$  (known and new loci combined; for one locus, age subgroups were not available) had higher odds ratios for early onset than for late onset CAD ( $P = 1.2 \times 10^{-4}$  for observed versus expected; Supplementary Table 7). The CAD loci showed consistent associations irrespective of case definition, although the odds ratios for most individual SNPs tended to be slightly greater for cases with angiographically proven CAD than for cases with unknown angiographic status ( $P = 0.019$  for observed versus expected) (Supplementary Table 8). In contrast, subgroup analyses in males and females revealed no sex-specific effects for any risk alleles (Supplementary Table 7) or for their observed versus expected pattern of association ( $P = 0.4$ ).

Among 7,637 CAD cases and 7,523 controls for whom we had individual level genotype data, the minimum and maximum number of risk alleles observed per individual was 15 and 37, respectively, when considering 23 CAD susceptibility loci. The mean weighted risk score was significantly higher for cases than for controls ( $P < 10^{-20}$ ). Furthermore, being in the top tenth percentile or lowest tenth percentile of the weighted score was associated with an odds ratio for CAD of 1.88 (95% CI 1.67–2.11) and 0.55 (95% CI 0.48–0.64), respectively, compared to the fiftieth percentile. The change in odds ratio for CAD across a broader spectrum of categories of the weighted score is shown in Supplementary Figure 4.

<sup>1</sup>A full list of authors and affiliations appears at the end of the paper.

Received 10 August 2010; accepted 10 February 2011; published online 6 March 2011; doi:10.1038/ng.784

# Replication & Discovery

- 10 out of 12 previously associated loci could be replicated
- 13 novel loci were uncovered

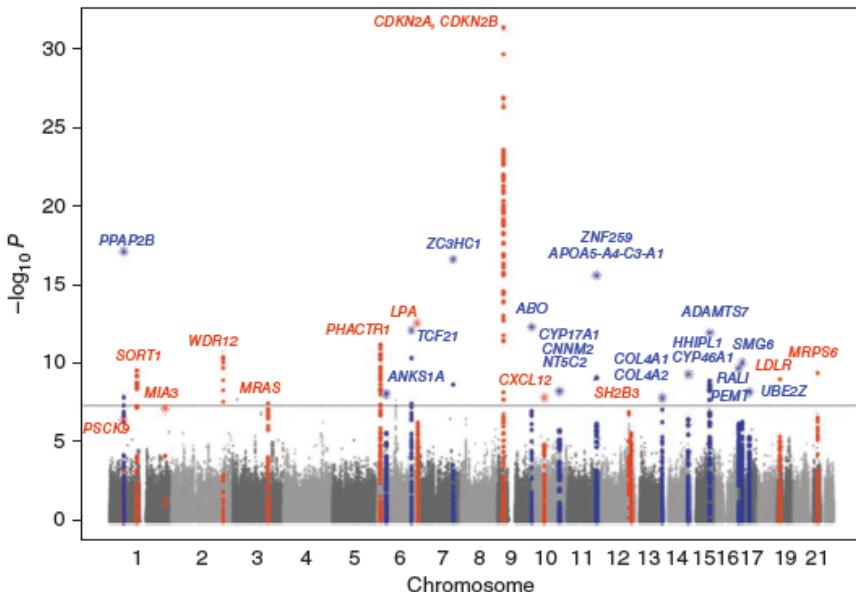
**Table 1** Association evidence in CARDIoGRAM for previously published loci for coronary disease (previously reported with genome-wide significance,  $P < 5 \times 10^{-8}$ )

Band	SNP	Gene(s) in region	n	Risk allele frequency (risk allele)	CARDIoGRAM		Reference
					OR (95% CI)	P	
1p32.3	rs11206510 <sup>a</sup>	PCSK9	102,352	0.82 (T)	1.08 (1.05–1.11)	$9.10 \times 10^{-8}$	1.15 (1.10–1.21) <sup>26</sup>
1p13.3	rs599839 <sup>b</sup>	SORT1	83,873	0.78 (A)	1.11 (1.08–1.15)	$2.89 \times 10^{-10}$	1.29 (1.18–1.40) <sup>21</sup>
1q41	rs17465637 <sup>c</sup>	MIA3	25,197	0.74 (C)	1.14 (1.09–1.20)	$1.36 \times 10^{-8}$	1.20 (1.12–1.30) <sup>21</sup>
2q33.1	rs6725887 <sup>b</sup>	WDR12	77,954	0.15 (C)	1.14 (1.09–1.19)	$1.12 \times 10^{-9}$	1.16 (1.10–1.22) <sup>26</sup>
3q22.3	rs2306374 <sup>b</sup>	MRAS	77,843	0.18 (C)	1.12 (1.07–1.16)	$3.34 \times 10^{-8}$	1.15 (1.11–1.19) <sup>23</sup>
6p24.1	rs12526453 <sup>b</sup>	PHACTR1	83,050	0.67 (C)	1.10 (1.06–1.13)	$1.15 \times 10^{-9}$	1.13 (1.09–1.17) <sup>26</sup>
6q25.3	rs3798220 <sup>d</sup>	LPA	32,584	0.02 (C)	1.51 (1.33–1.70)	$3.00 \times 10^{-11}$	1.92 (1.48–2.49) <sup>5</sup>
9p21.3	rs4977574 <sup>b</sup>	CDKN2A, CDKN2B	84,256	0.46 (G)	1.29 (1.23–1.36)	$1.35 \times 10^{-22}$	1.25 (1.18–1.31), 1.37 (1.26–1.48) <sup>20,21,27,28</sup>
10q11.21	rs1746048 <sup>a</sup>	CXCL12	136,416	0.87 (C)	1.09 (1.07–1.13)	$2.93 \times 10^{-10}$	1.33 (1.20–1.48) <sup>21</sup>
12q24.12	rs3184504 <sup>b</sup>	SH2B3	67,746	0.44 (T)	1.07 (1.04–1.10)	$6.35 \times 10^{-6}$	1.13 (1.08–1.18) <sup>38</sup>
19p13.2	rs1122608 <sup>b</sup>	LDLR	49,693	0.77 (G)	1.14 (1.09–1.18)	$9.73 \times 10^{-10}$	1.14 (1.09–1.19) <sup>26</sup>
21q22.11	rs9982601 <sup>b</sup>	MRPS6	46,230	0.15 (T)	1.18 (1.12–1.24)	$4.22 \times 10^{-10}$	1.19 (1.13–1.27) <sup>26</sup>

Data taken from <sup>a</sup>the combined analysis, <sup>b</sup>the meta-analysis, <sup>c</sup>only genotyped data from a subset of studies and <sup>d</sup>the replication.

**Table 2** New loci for coronary disease

Band	SNP	Gene(s) in region	Risk allele frequency (risk allele)	Meta-analysis		Replication		Combined analysis	
				P	n	P	n	OR (95% CI)	P
1p32.2	rs17114036	PPAP2B	0.91 (A)	$1.43 \times 10^{-8}$	80,870	$3.18 \times 10^{-12}$	52,356	1.17 (1.13–1.22)	$3.81 \times 10^{-19}$
6p21.31	rs17609940	ANKS1A	0.75 (G)	$2.21 \times 10^{-6}$	83,997	$1.18 \times 10^{-3}$	53,415	1.07 (1.05–1.10)	$1.36 \times 10^{-8}$
6q23.2	rs12190287	TCF21	0.62 (C)	$4.64 \times 10^{-11}$	78,290	$3.25 \times 10^{-4}$	52,598	1.08 (1.06–1.10)	$1.07 \times 10^{-12}$
7q32.2	rs15156924	ZC3HC1	0.62 (C)	$2.22 \times 10^{-9}$	80,011	$7.37 \times 10^{-10}$	54,189	1.09 (1.07–1.12)	$9.18 \times 10^{-18}$
9q34.2	rs579459	ABO	0.21 (C)	$1.16 \times 10^{-7}$	77,138	$7.02 \times 10^{-8}$	46,840	1.10 (1.07–1.13)	$4.08 \times 10^{-14}$
10q24.32	rs12413409	CYP17A1, CNNM2, NT5C2	0.89 (G)	$1.47 \times 10^{-6}$	80,940	$1.38 \times 10^{-4}$	48,801	1.12 (1.08–1.16)	$1.03 \times 10^{-9}$
11q23.3	rs964184	ZNF259, APOA5- A4-C3-A1	0.13 (G)	$8.02 \times 10^{-10}$	82,562	$2.20 \times 10^{-9}$	52,930	1.13 (1.10–1.16)	$1.02 \times 10^{-17}$
13q34	rs4773144	COL4A1, COL4A2	0.44 (G)	$4.15 \times 10^{-7}$	77,113	$1.31 \times 10^{-3}$	37,618	1.07 (1.05–1.09)	$3.84 \times 10^{-9}$
14q32.2	rs2895811	HHIP1	0.43 (C)	$2.67 \times 10^{-7}$	63,184	$4.59 \times 10^{-5}$	51,054	1.07 (1.05–1.10)	$1.14 \times 10^{-10}$
15q25.1	rs3825807	ADAMTS7	0.57 (A)	$9.63 \times 10^{-6}$	80,849	$1.39 \times 10^{-8}$	48,803	1.08 (1.06–1.10)	$1.07 \times 10^{-12}$
17p13.3	rs216172	SMG6, SRR	0.37 (C)	$6.22 \times 10^{-7}$	57,238	$2.11 \times 10^{-4}$	54,303	1.07 (1.05–1.09)	$1.15 \times 10^{-9}$
17p11.2	rs12936587	RASD1, SMCR3, PEMT	0.56 (G)	$4.89 \times 10^{-7}$	76,952	$1.35 \times 10^{-4}$	52,648	1.07 (1.05–1.09)	$4.45 \times 10^{-10}$
17q21.32	rs46522	UBE2Z, GIP, ATP5G1, SNF8	0.53 (T)	$3.57 \times 10^{-6}$	83,867	$8.88 \times 10^{-4}$	53,766	1.06 (1.04–1.08)	$1.81 \times 10^{-8}$



**Figure 1** Graphical summary (Manhattan plot) of genome-wide association results. The x axis represents the genome in physical order; the y axis shows  $-\log_{10} P$  for all SNPs. Data from the discovery phase are shown in circles, and data from the combined discovery and replication phases are shown in stars. Genes at the significant loci are listed above the signals. Known loci are shown in red and newly discovered loci are shown in blue.

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

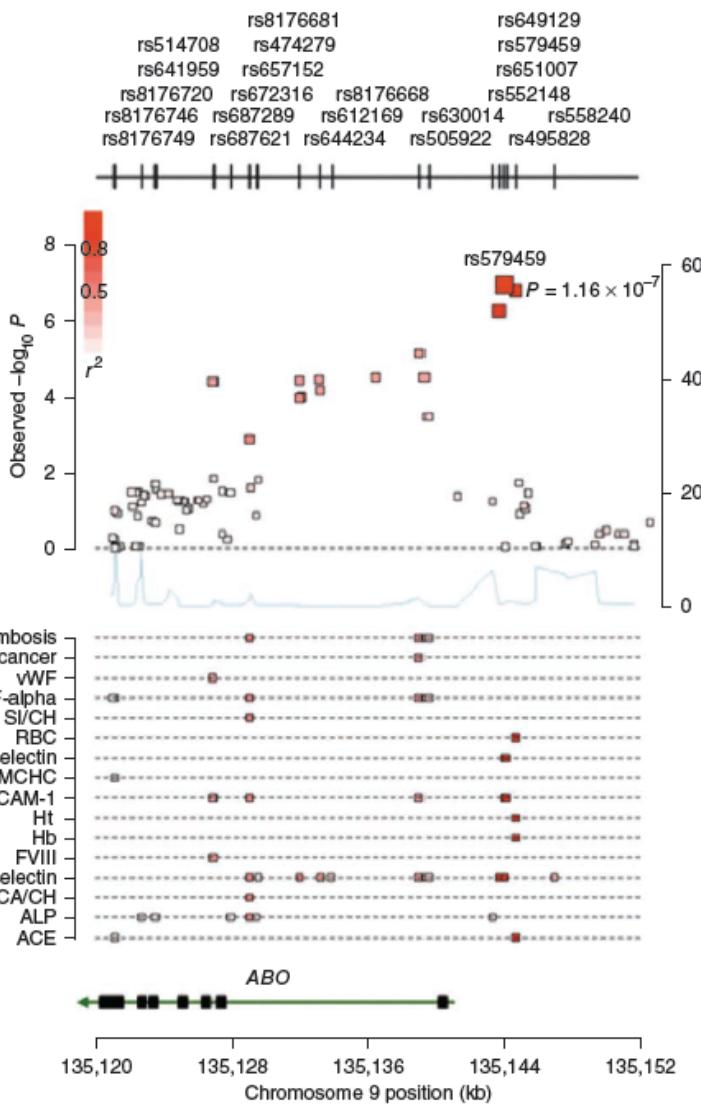
**Table 3** Effects of new CAD loci on traditional risk factors in combined analysis of ARIC and KORA F3 and F4 ( $n = 13,171$ )

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.



# Weighted Polygenic Burden Score

- TraitX is a risk factor and/or (status) marker and associated with atherosclerotic diseases and/or phenotypes
- Hypothesis:** A weighted polygenic burden score based on SNPs associated with TraitX (GBS TraitX ) is associated with atherosclerotic disease
- Validation:** A GBS TraitX is associated with natural log-transformed TraitX
- Question:** Is a GBS TraitX associated with:
  - Plaque phenotype
  - Plaque protein expression
  - (secondary) clinical events
  - Symptoms prior to surgery
  - Or any other relevant (clinical) feature?

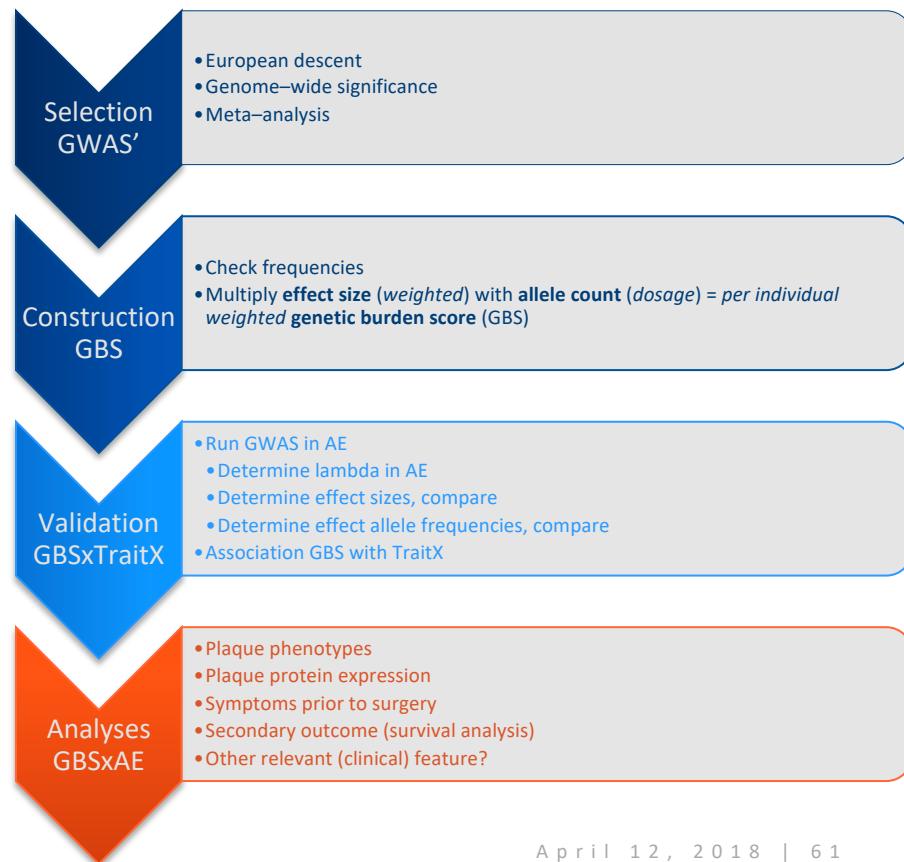
## Weighted Polygenic Burden Score

$$\sum_n^i \beta_i \times SNP_n$$

$i$ =effect size ( $\beta$ ) at locus

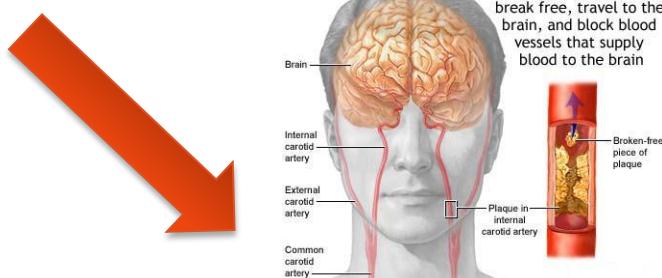
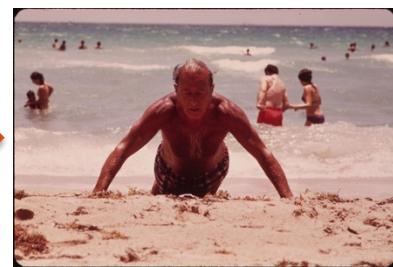
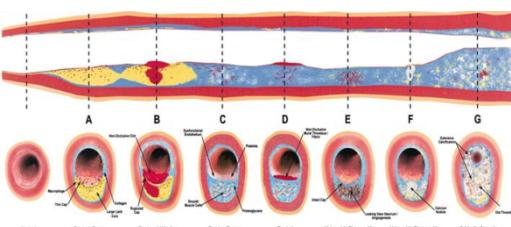
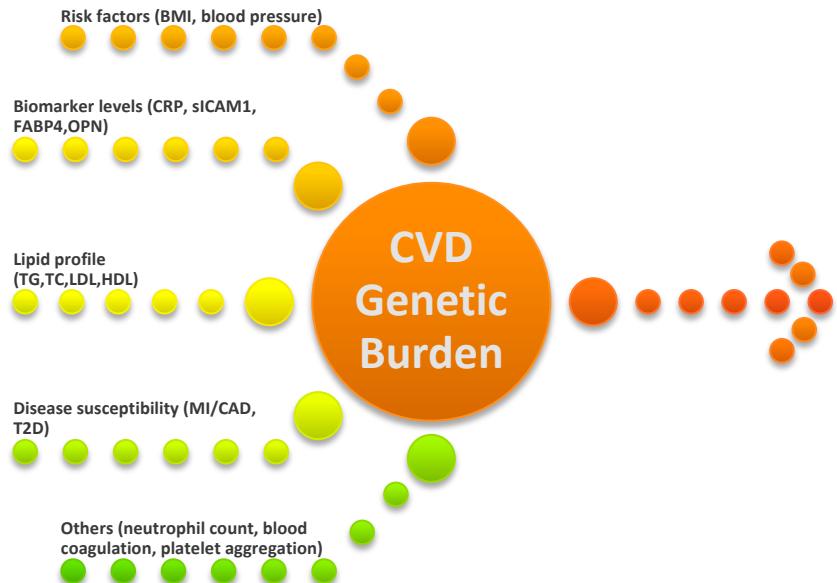
$n$ =number alleles at locus (range: 0 to 2), i.e. dosage

$SNP$ =locus proxy



# Genetic Burden to Atherosclerotic disease

Some *disease susceptibility* (T2D) leads to diseases (T2D, CVD), which on them selves are also *risk factors* (T2D, hypertension) and *vice versa* (hypertension)



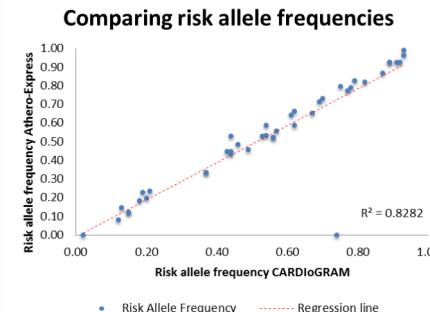
# MI & CAD in the Athero-Express

- Similar to CARDIoGRAM
- CAD cases
  - (fatal) MI, AP, PTCA, CABG prior to surgery (self-reported) or during follow-up
- Controls: all CAD-free patients
- Age limits: <45, <50, <60, <66 years → focus on CAD <60 years (similar to CARDIoGRAM)

	CAD <45	CAD <50	CAD <60	CAD <66
case	32	52	99	128
control	437	437	437	437
N	469	489	536	565

# Variants included

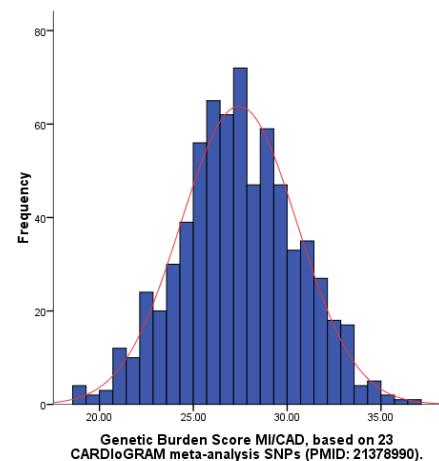
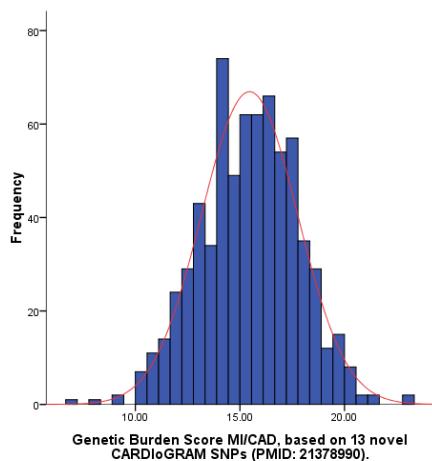
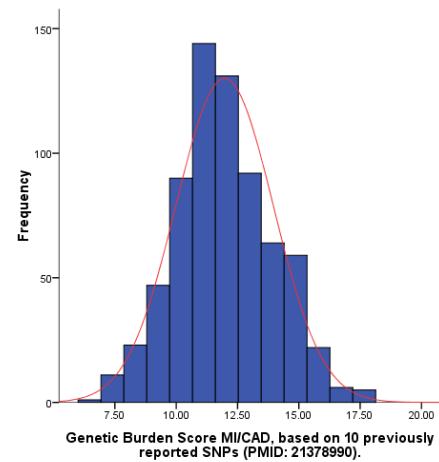
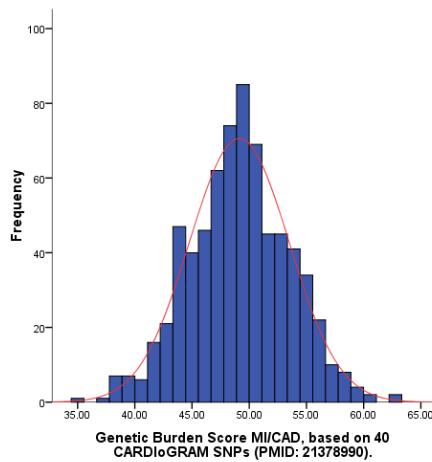
- 42 variants are “discovered”, of which 23 met the final meta-analysis criteria
- 2 variants are not in HM2r22, and therefore cannot be imputed
- 13 variants are genotyped, 27 are imputed
- 4 different *genetic burden scores* can be calculated



dbSNP rs#	Chr.	Position (bp)	Gene	Risk allele	Risk allele frequency (CARDIoGRAM)	Risk allele frequency (Athero-Express)	Odds ratio	95% CI lower upper	P-value	Analysis source	Type	Athero-Express Information			Genetic Burden Score Type				
												Source	A-allele	B-allele	Allele discovery	previous	novel	meta	
rs599839	1	109623689	SORT1	A	0.78	0.79	1.11	1.08 - 1.15	<b>2.89E-10</b>	meta-analysis	previously reported	genotyped	A	G	A	Y	Y	.	Y
rs11206510	1	55268627	PCKS9	T	0.82	0.82	1.08	1.05 - 1.11	<b>9.10E-08</b>	combined analysis	previously reported	genotyped	C	T	B	Y	Y	.	Y
rs17465637	1	n/a	MIA3	C	0.74	n/a	1.14	1.09 - 1.20	<b>1.36E-08</b>	subset studies	previously reported	n/a	n/a	n/a	n/a	n/a	n/a	n/a	
rs6725887	2	20345130	VDR12	C	0.15	0.12	1.14	1.09 - 1.19	<b>1.12E-09</b>	meta-analysis	previously reported	genotyped	C	T	A	Y	Y	.	Y
rs2306374	3	139602642	MRSAS	C	0.18	0.18	1.12	1.07 - 1.16	<b>3.34E-09</b>	meta-analysis	previously reported	genotyped	C	T	A	Y	Y	.	Y
rs12526453	6	13035530	PHACTR1	C	0.67	0.65	1.10	1.06 - 1.13	<b>1.15E-09</b>	meta-analysis	previously reported	genotyped	C	G	A	Y	Y	.	Y
rs3798220	6	n/a	LPA	C	0.02	n/a	1.54	1.36 - 1.74	<b>9.62E-12</b>	replication	previously reported	n/a	n/a	n/a	n/a	n/a	n/a	n/a	
rs4977574	9	22088574	CDKN2A/B,ANRIL	G	0.46	0.48	1.29	1.23 - 1.36	<b>1.35E-22</b>	meta-analysis	previously reported	genotyped	A	G	B	Y	Y	.	Y
rs1746048	10	44095830	CXL12	C	0.87	0.87	1.09	1.07 - 1.13	<b>2.12E-10</b>	combined analysis	previously reported	genotyped	C	T	A	Y	Y	.	Y
rs3184504	12	110368991	SH2B3	T	0.44	0.53	1.07	1.04 - 1.10	<b>6.35E-06</b>	meta-analysis	previously reported	imputed	C	T	B	Y	Y	.	Y
rs1122608	19	11024601	LDLR	G	0.77	0.77	1.14	1.09 - 1.18	<b>9.73E-10</b>	meta-analysis	previously reported	genotyped	G	T	A	Y	Y	.	Y
rs9982601	21	34520998	MRPS6	T	0.15	0.12	1.18	1.12 - 1.24	<b>4.22E-10</b>	meta-analysis	previously reported	imputed	C	T	B	Y	Y	.	Y
rs17114036	1	56735409	PPAP2B	A	0.91	0.93	1.17	1.13 - 1.22	<b>3.81E-19</b>	combined analysis	new locus	imputed	A	G	A	Y	.	Y	Y
rs17609940	6	35142778	ANKS1A	G	0.75	0.80	1.07	1.05 - 1.10	<b>1.07E-08</b>	combined analysis	new locus	genotyped	C	G	B	Y	.	Y	Y
rs12190287	6	134256218	TCF21	C	0.62	0.66	1.08	1.06 - 1.10	<b>1.07E-12</b>	combined analysis	new locus	imputed	C	G	A	Y	.	Y	Y
rs11556924	7	129450732	ZC3H1C1	C	0.62	0.59	1.10	1.07 - 1.12	<b>5.68E-18</b>	combined analysis	new locus	genotyped	C	T	A	Y	.	Y	Y
rs579459	9	135143989	ABO	C	0.21	0.23	1.10	1.07 - 1.13	<b>4.08E-14</b>	combined analysis	new locus	imputed	C	T	A	Y	.	Y	Y
rs12413409	10	104709086	CYP17A1,CNNM2,NT5C2	G	0.89	0.92	1.12	1.08 - 1.16	<b>9.51E-10</b>	combined analysis	new locus	genotyped	A	G	B	Y	.	Y	Y
rs964184	11	116154127	ZNF259,APOA5-A4-C3-A1	G	0.13	0.15	1.13	1.10 - 1.17	<b>3.14E-18</b>	combined analysis	new locus	imputed	C	G	B	Y	.	Y	Y
rs4773144	13	109758713	COL4A1, COL4A2	G	0.44	0.43	1.07	1.05 - 1.10	<b>3.83E-09</b>	combined analysis	new locus	imputed	A	G	B	Y	.	Y	Y
rs2895811	14	99203695	HH1PL1HH1PL1	C	0.43	0.45	1.07	1.05 - 1.10	<b>4.03E-10</b>	combined analysis	new locus	imputed	C	T	A	Y	.	Y	Y
rs3825807	15	76876166	ADAMTS7	A	0.57	0.56	1.08	1.06 - 1.10	<b>1.07E-12</b>	combined analysis	new locus	imputed	A	G	A	Y	.	Y	Y
rs12936587	17	17484447	RASD1,SMCR3,PEMT	G	0.56	0.52	1.07	1.05 - 1.09	<b>4.45E-10</b>	combined analysis	new locus	imputed	A	G	B	Y	.	Y	Y
rs216172	17	2073254	SMG6,SRR	C	0.37	0.33	1.07	1.05 - 1.09	<b>1.18E-09</b>	combined analysis	new locus	imputed	C	G	A	Y	.	Y	Y
rs46522	17	44343596	UBE2Z,GIP,ATP5G1,SNF8	T	0.53	0.53	1.06	1.04 - 1.08	<b>9.03E-09</b>	combined analysis	new locus	imputed	C	T	B	Y	.	Y	Y
rs2404715	1	56781366	-	C	0.92	0.93	1.16	1.12 - 1.20	<b>3.75E-15</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	C	T	A	Y	.	.	.
rs10933436	2	233707625	-	A	0.49	0.46	1.06	1.04 - 1.09	<b>7.06E-06</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	C	A	Y	.	.	.
rs7651039	3	15623008	-	C	0.54	0.53	1.06	1.04 - 1.09	<b>1.64E-06</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	C	T	A	Y	.	.	.
rs9838412	3	86202991	-	C	0.79	0.83	1.04	1.01 - 1.08	<b>7.50E-03</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	C	B	Y	.	.	.
rs12190423	6	72259432	-	G	0.61	0.64	1.05	1.02 - 1.07	<b>5.42E-05</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	C	G	B	Y	.	.	.
rs12524865	6	134238367	-	C	0.70	0.73	1.07	1.04 - 1.09	<b>2.29E-08</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	C	B	Y	.	.	.
rs7808424	7	116855058	-	G	0.12	0.08	1.10	1.06 - 1.14	<b>1.17E-06</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	G	T	A	Y	.	.	.
rs12682131	8	137311147	-	G	0.93	0.99	1.10	1.02 - 1.19	<b>1.05E-02</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	genotyped	A	G	B	Y	.	.	.
rs651007	9	135143696	-	T	0.19	0.23	1.10	1.07 - 1.13	<b>2.36E-13</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	genotyped	C	T	B	Y	.	.	.
rs7920682	10	30357832	-	A	0.54	0.59	1.05	1.02 - 1.08	<b>1.44E-04</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	G	A	Y	.	.	.
rs16911227		59369844	-	G	0.93	0.96	1.09	1.03 - 1.15	<b>2.31E-03</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	G	B	Y	.	.	.
rs12411886	10	104675289	-	C	0.89	0.92	1.11	1.07 - 1.15	<b>5.59E-09</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	C	B	Y	.	.	.
rs4937126	11	125787107	-	G	0.69	0.71	1.06	1.04 - 1.09	<b>4.73E-06</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	G	B	Y	.	.	.
rs4624107	14	99197193	-	C	0.44	0.45	1.07	1.05 - 1.09	<b>6.75E-09</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	C	G	A	Y	.	.	.
rs12924776	16	88114093	-	T	0.20	0.20	1.06	1.03 - 1.08	<b>9.93E-06</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	T	B	Y	.	.	.
rs1231206	17	2072355	-	A	0.37	0.33	1.07	1.05 - 1.09	<b>8.52E-10</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	A	G	A	Y	.	.	.
rs12449964	17	17485429	-	C	0.56	0.52	1.06	1.04 - 1.09	<b>8.43E-09</b>	combined analysis	new locus, not meeting CARDIoGRAM criteria	imputed	C	T	A	Y	.	.	.

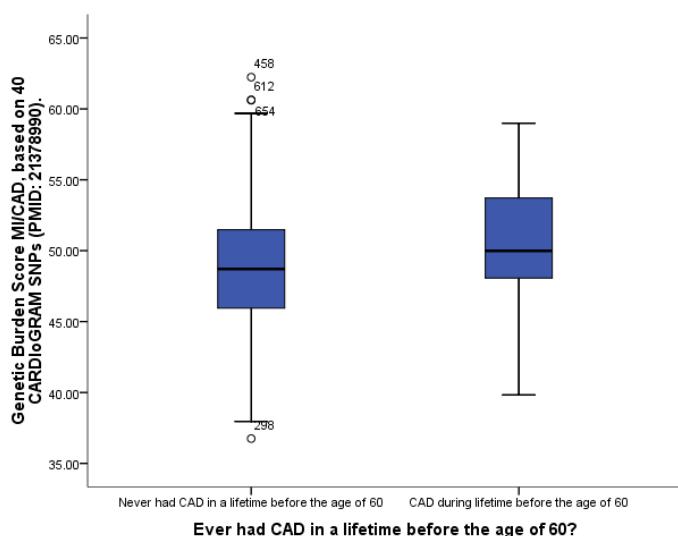
# Genetic Burden Scores of CAD variants

- 4 GBS' are constructed, including:
  - 40 SNPs associated with CAD in *discovery* phase
  - 10 *previously reported* SNPs associated with CAD
  - 13 *novel* reported SNPs associated with CAD
  - 23 SNPs *combined genome-wide significantly associated* with CAD



# CAD <60 years per GBS category

## *discovery phase & previously associated*



Ranks		
Genetic Burden Score MI/CAD, based on 40 CARDIoGRAM SNPs (PMID: 21378990).	N	Mean Rank
Ever had CAD in a lifetime before the age of 60	437	257.13
Never had CAD in a lifetime before the age of 60	99	318.70
Total	536	

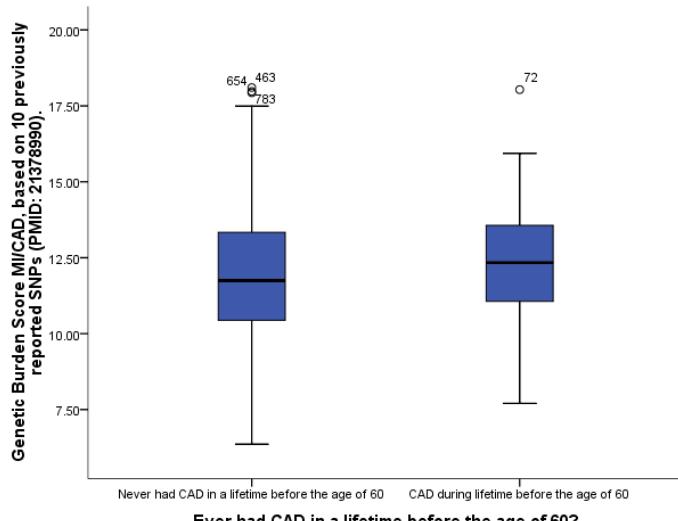
**Test Statistics<sup>b,c</sup>**

Genetic Burden Score MI/CAD, based on 40 CARDIoGRAM SNPs (PMID: 21378990).	
Chi-Square	12.756
df	1
Asymp. Sig.	.355E-4
Monte Carlo Sig.	.001 <sup>a</sup>
99% Confidence Interval	
Lower Bound	.000
Upper Bound	.001

a. Based on 10000 sampled tables with starting seed 1502173562.

b. Kruskal Wallis Test

c. Grouping Variable: Ever had CAD in a lifetime before the age of 60?



Ranks		
Genetic Burden Score MI/CAD, based on 10 previously reported SNPs (PMID: 21378990).	N	Mean Rank
Ever had CAD in a lifetime before the age of 60	437	261.40
Never had CAD in a lifetime before the age of 60	99	299.83
Total	536	

**Test Statistics<sup>b,c</sup>**

Genetic Burden Score MI/CAD, based on 10 previously reported SNPs (PMID: 21378990).	
Chi-Square	4.969
df	1
Asymp. Sig.	.026
Monte Carlo Sig.	.026 <sup>a</sup>
99% Confidence Interval	
Lower Bound	.022
Upper Bound	.030

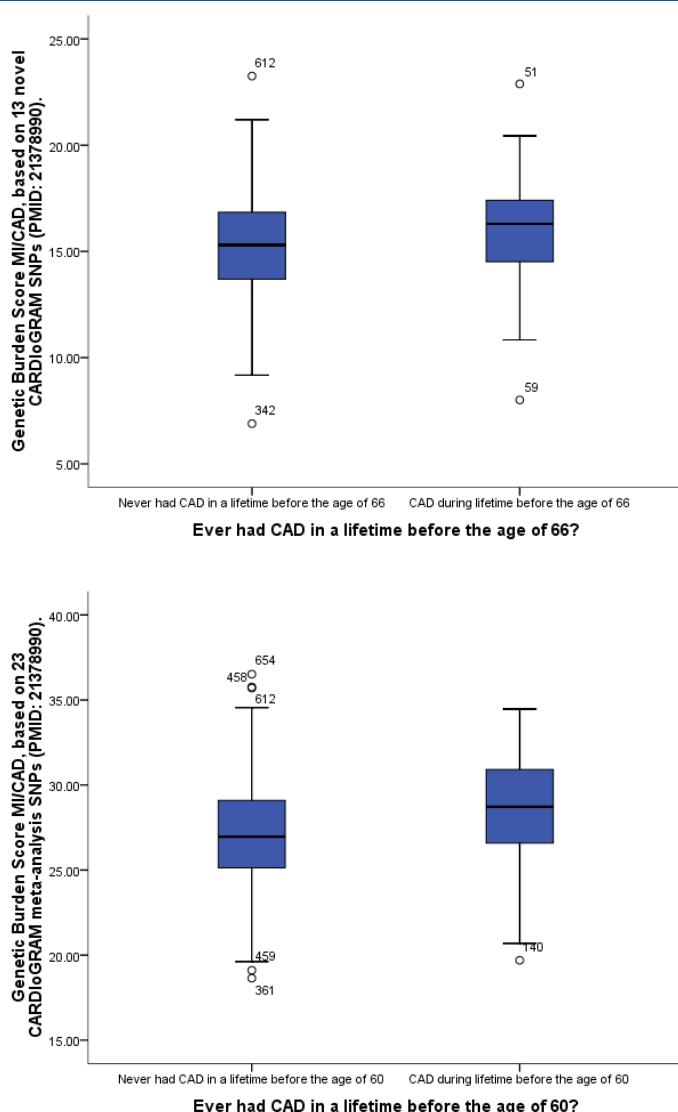
a. Based on 10000 sampled tables with starting seed 562334227.

b. Kruskal Wallis Test

c. Grouping Variable: Ever had CAD in a lifetime before the age of 60?

# CAD <60 years per GBS category

## *novel variants & final meta-analysis*



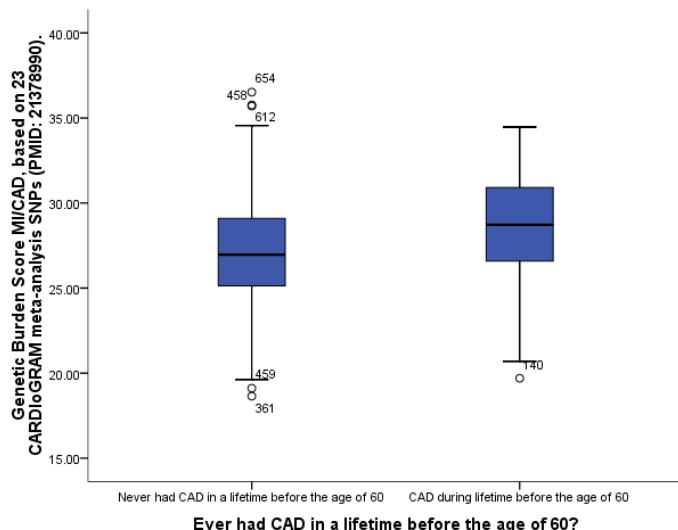
Ranks

Ever had CAD in a lifetime before the age of 60?	N	Mean Rank
Never had CAD in a lifetime before the age of 60	437	254.98
CAD during lifetime before the age of 60	99	328.16
Total	536	261.57

Test Statistics<sup>b,c</sup>

	Genetic Burden Score MiCAD, based on 13 novel CARDioGRAM SNPs (PMID: 21378990)
Chi-Square	18.020
df	1
Asymp. Sig.	2.19E-5
Monte Carlo Sig.	.000 <sup>a</sup>
Sig.	.000
99% Confidence Interval	Lower Bound .000 Upper Bound .000

a. Based on 10000 sampled tables with starting seed 1585587178.  
b. Kruskal Wallis Test  
c. Grouping Variable: Ever had CAD in a lifetime before the age of 60?



Ranks

Ever had CAD in a lifetime before the age of 60?	N	Mean Rank
Never had CAD in a lifetime before the age of 60	437	253.46
CAD during lifetime before the age of 60	99	334.88
Total	536	294.17

Test Statistics<sup>b,c</sup>

	Genetic Burden Score MiCAD, based on 23 CARDioGRAM meta-analysis SNPs (PMID: 21378990)
Chi-Square	22.306
df	1
Asymp. Sig.	2.32E-6
Monte Carlo Sig.	.000 <sup>a</sup>
Sig.	.000
99% Confidence Interval	Lower Bound .000 Upper Bound .000

a. Based on 10000 sampled tables with starting seed 1122541128.  
b. Kruskal Wallis Test  
c. Grouping Variable: Ever had CAD in a lifetime before the age of 60?

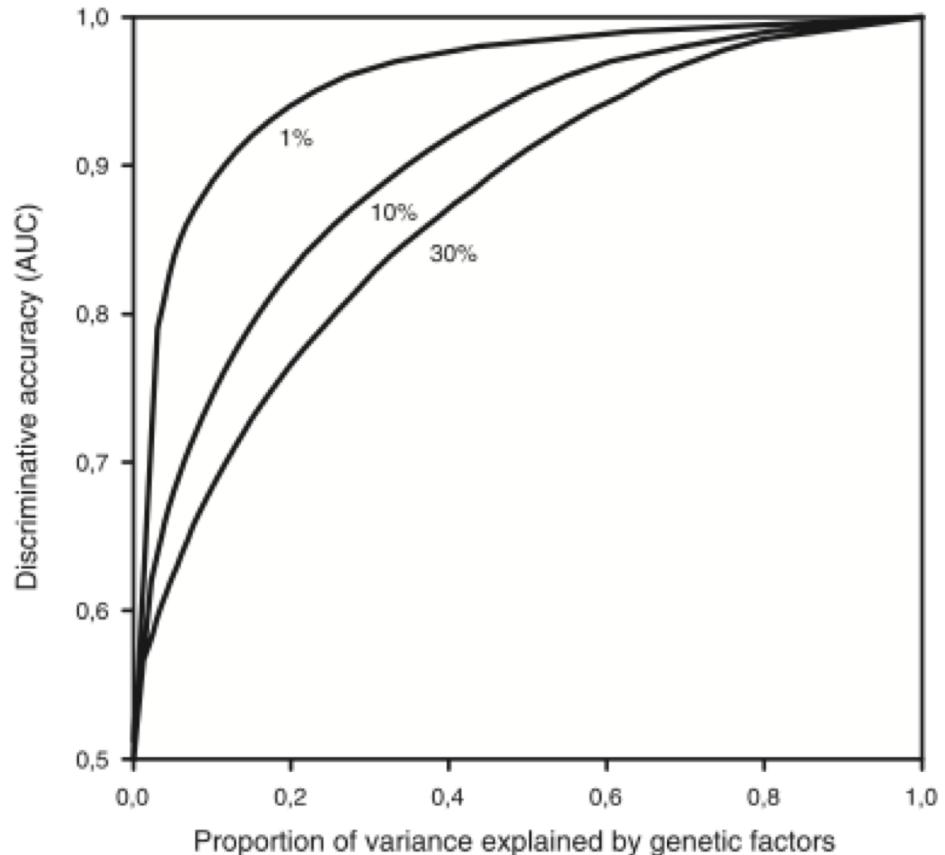
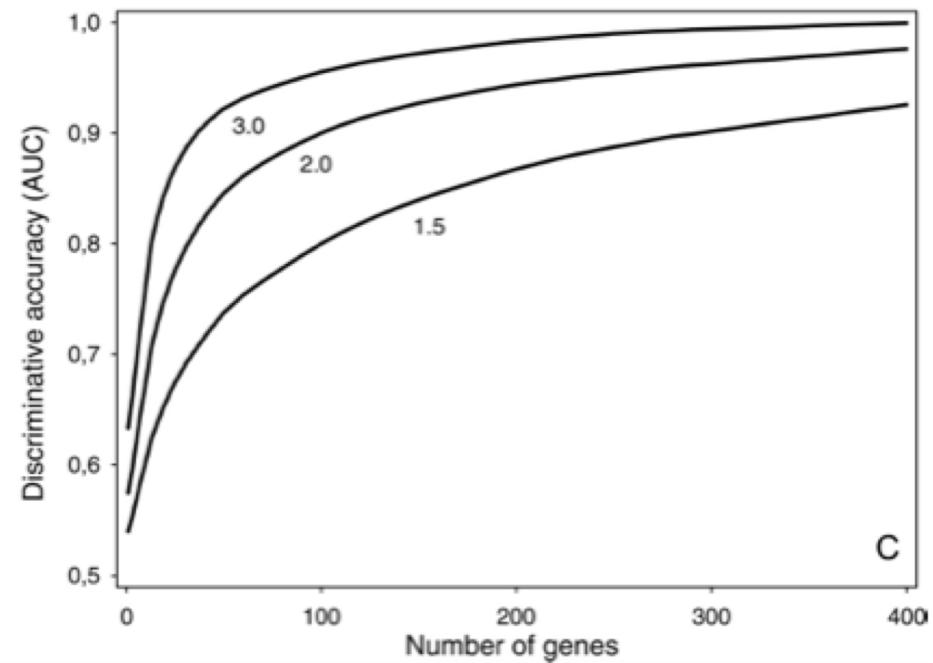
# Logistic regression for CAD age groups

- Covariates: age at onset for cases, age at inclusion for controls, gender
  - The novel variants and the final meta-analysis variants are best in predicting CAD < 60-66 years
  - It appears that these variants do not predict CAD before 50 years...

GBS type	Age group	Effect size ( $\beta$ )	SE $\beta$	Odds ratio	95% CI		P - value
					lower	upper	
discovery	< 45 years	-0.100	0.083	0.905	0.768	1.065	0.230
	< 50 years	-0.053	0.071	0.948	0.826	1.089	0.449
	< 60 years	0.047	0.038	1.048	0.973	1.130	0.216
	< 66 years	0.034	0.030	1.034	0.976	1.097	0.257
previous	< 45 years	0.022	0.171	1.022	0.732	1.428	0.896
	< 50 years	-0.128	0.137	0.880	0.673	1.151	0.351
	< 60 years	0.077	0.079	1.080	0.926	1.260	0.327
	< 66 years	0.058	0.064	1.059	0.934	1.201	0.368
novel	< 45 years	-0.213	0.180	0.808	0.567	1.151	0.238
	< 50 years	0.123	0.134	0.880	0.673	1.151	0.360
	< 60 years	0.166	0.075	1.181	1.020	1.368	<b>0.026</b>
	< 66 years	0.126	0.058	1.134	1.013	1.270	<b>0.029</b>
meta-analysis	< 45 years	-0.070	0.110	0.932	0.752	1.156	0.523
	< 50 years	-0.002	0.092	0.998	0.833	1.196	0.981
	< 60 years	0.126	0.054	1.134	1.019	1.262	<b>0.021</b>
	< 66 years	0.096	0.043	1.100	1.012	1.197	<b>0.026</b>

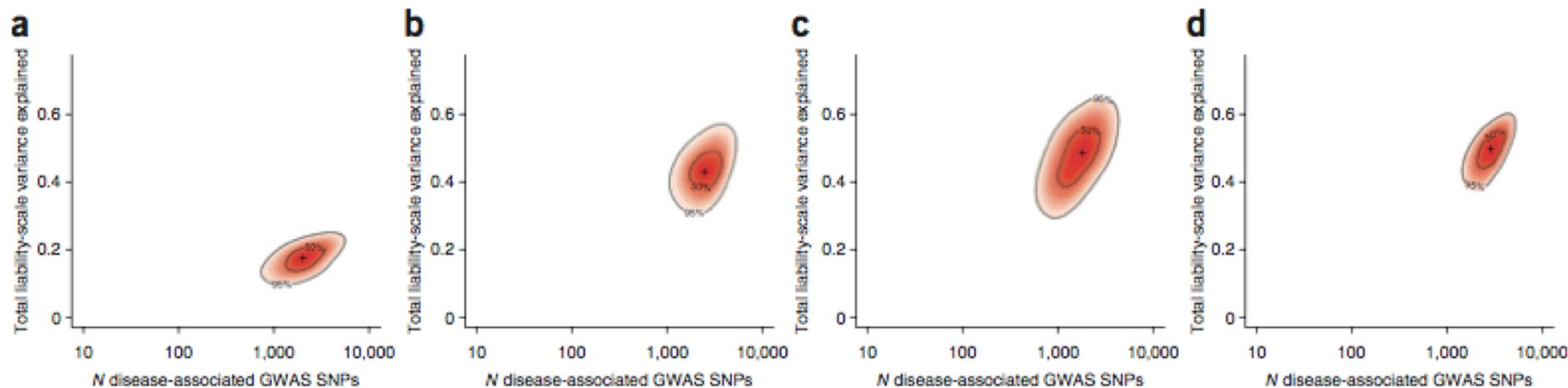
# So can we even predict?

- 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.



University Medical Center  
Utrecht

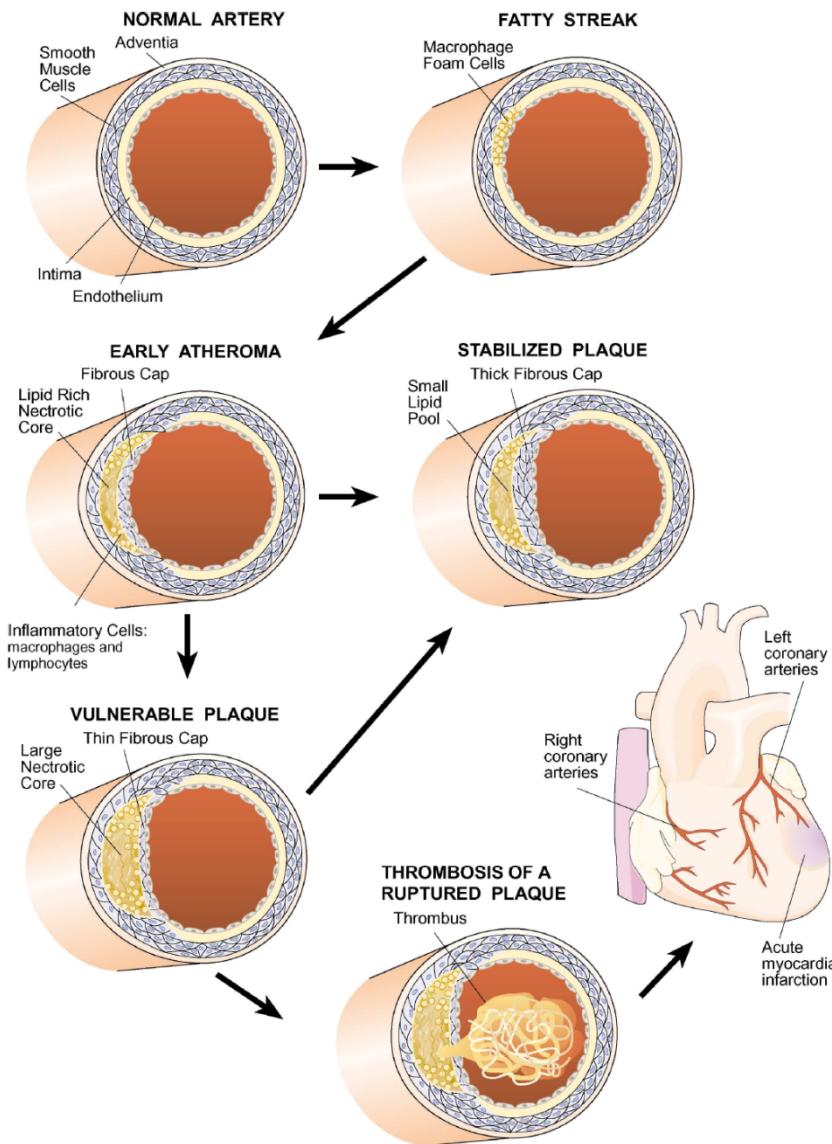
Genetics of Advanced Atherosclerotic Disease

# GWAS OF PLAQUE PHENOTYPES



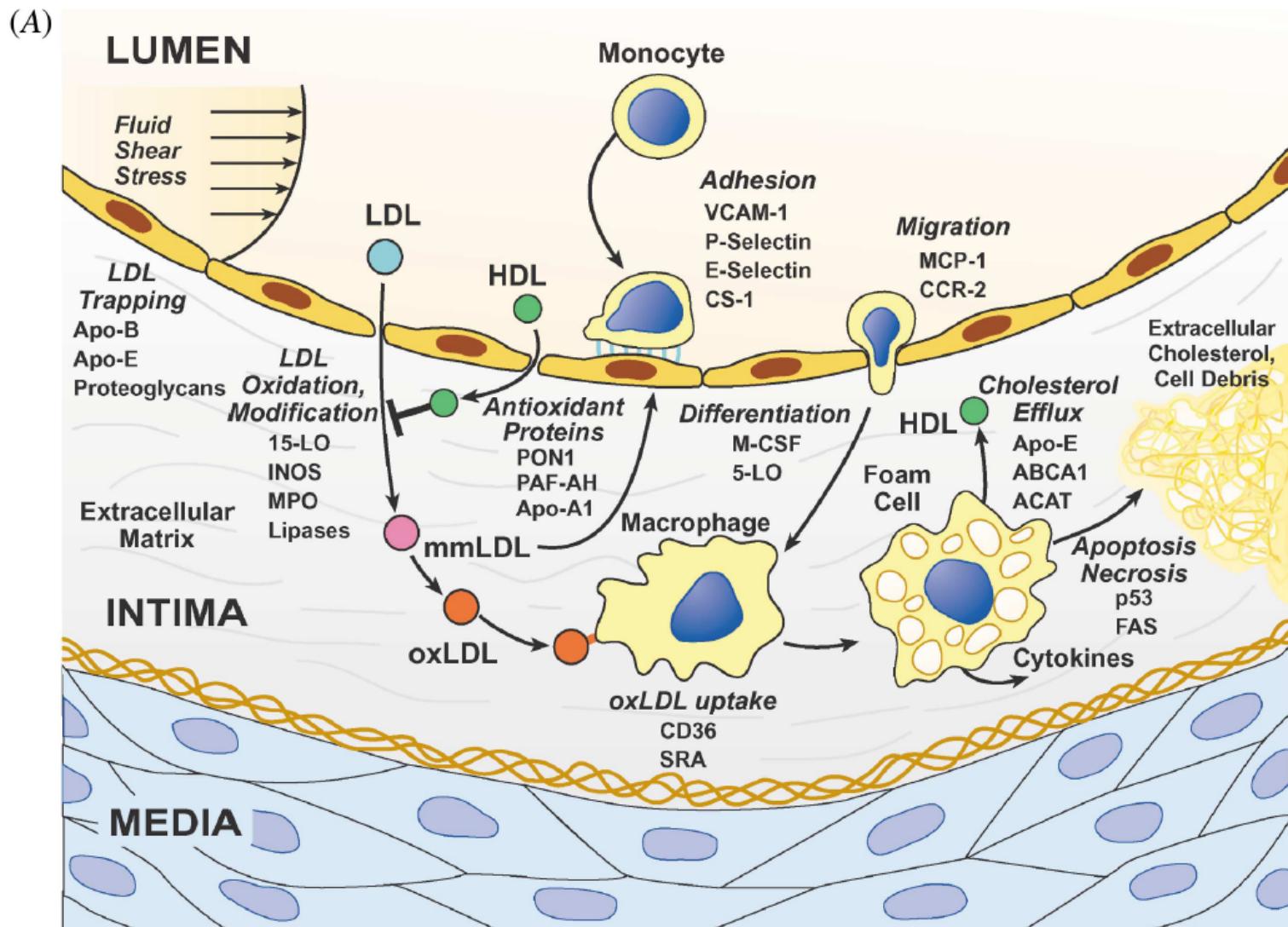
# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



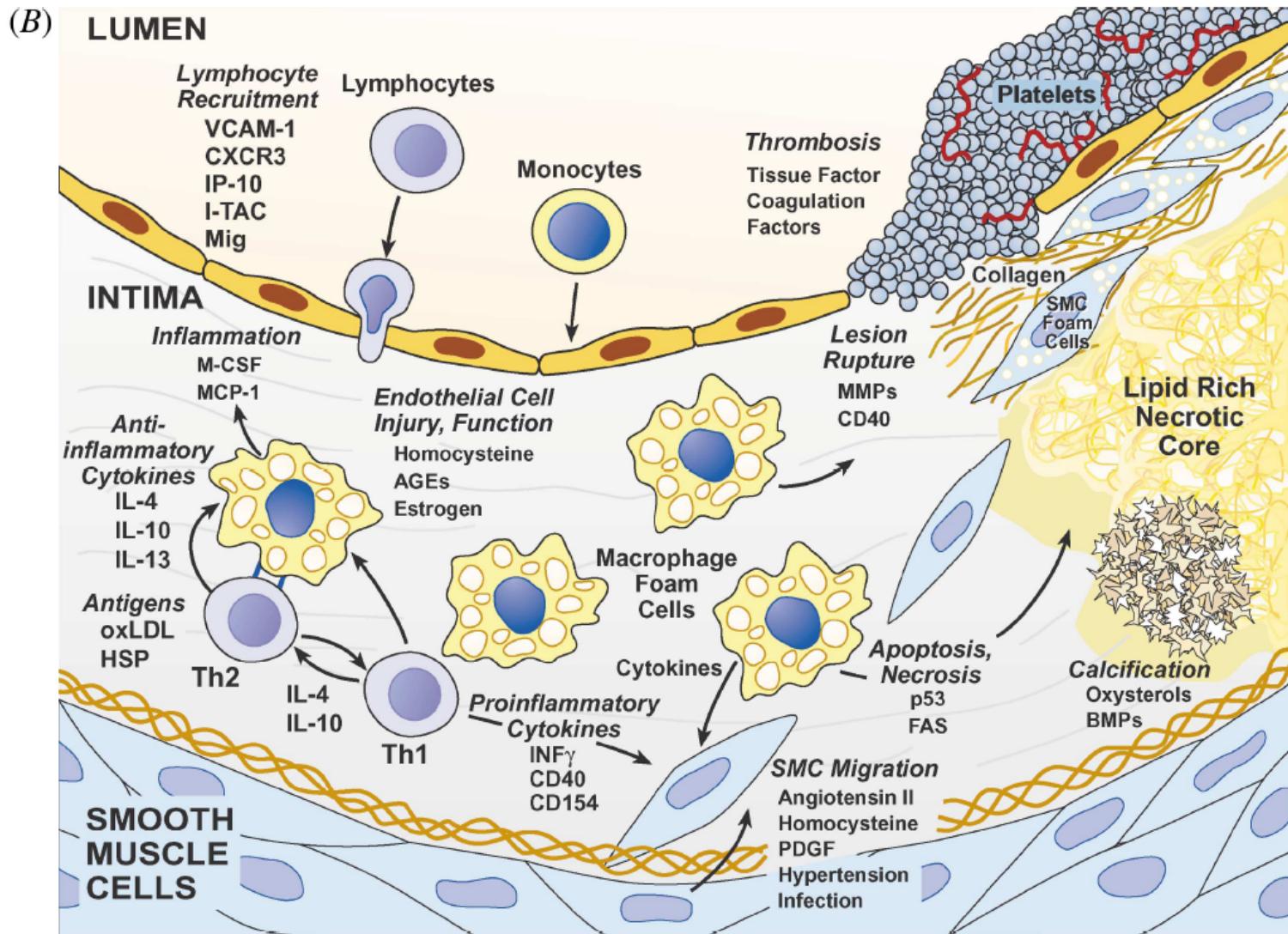
# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



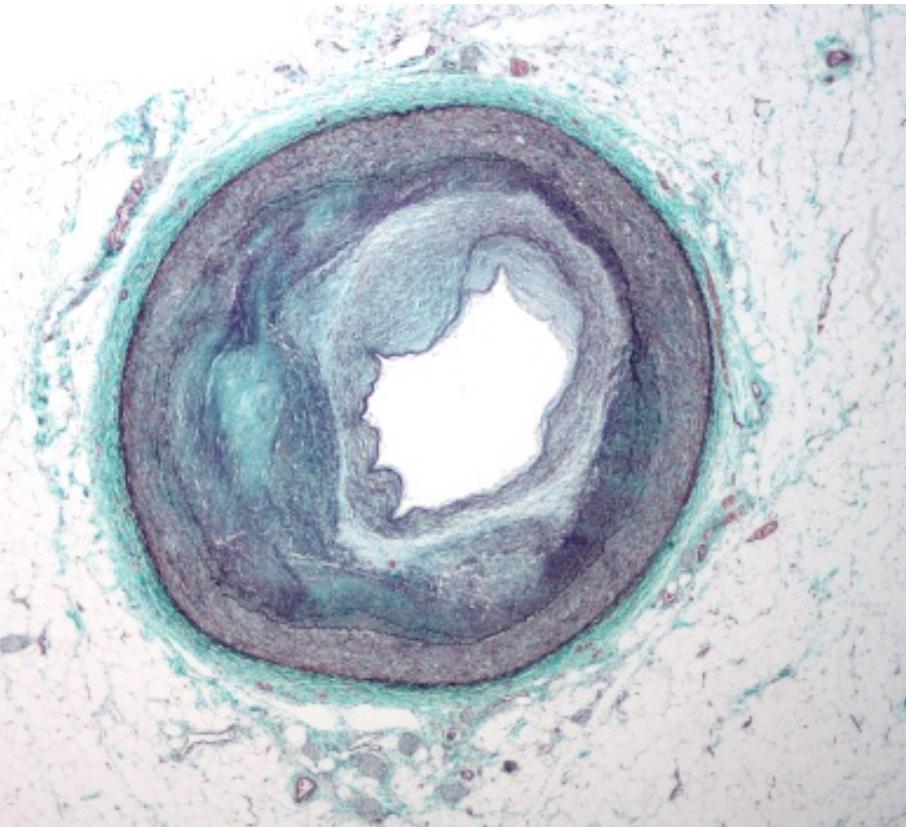
# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



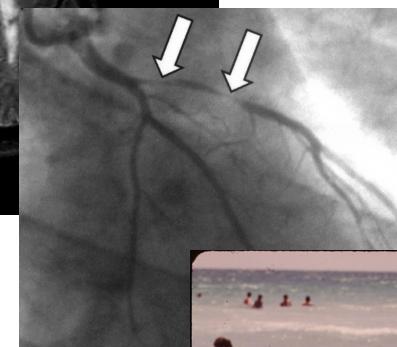
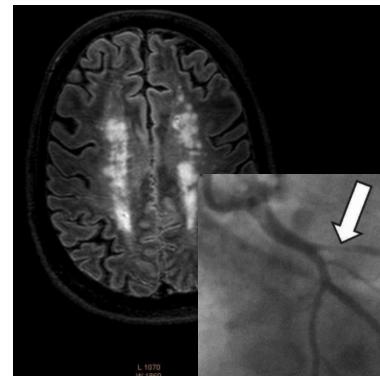
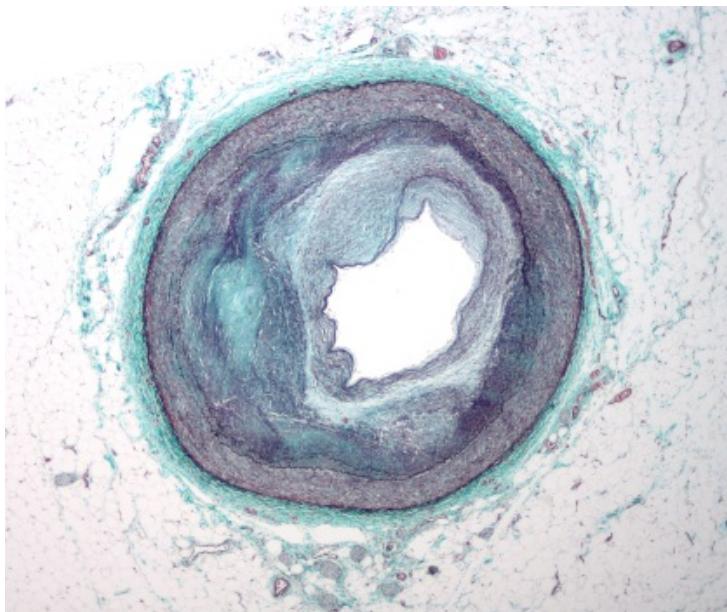
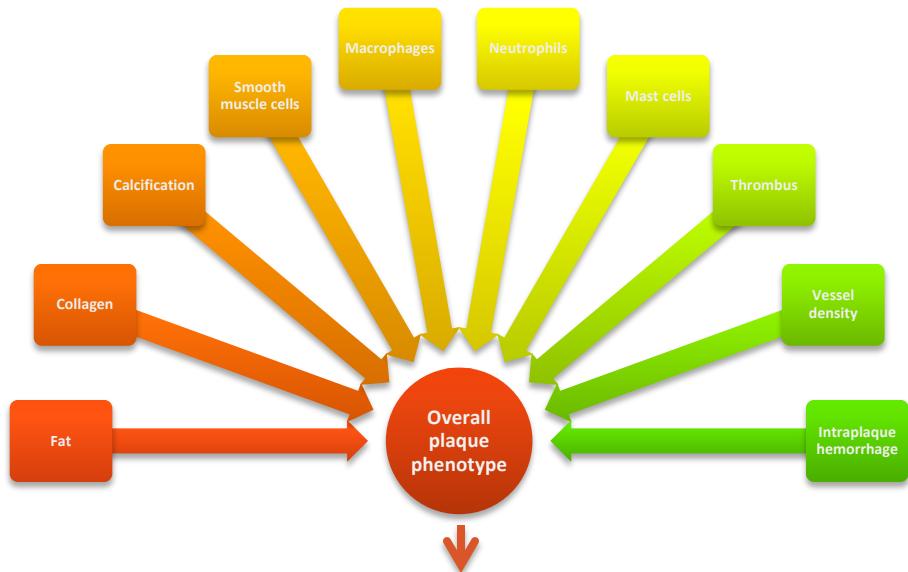
# Atherosclerosis causes cardiovascular disease

- *Atherosclerosis* is a chronic inflammation of the arteries and the underlying cause of *cardiovascular disease*
- Histologically *atherosclerosis* can be defined by *plaque phenotypes*



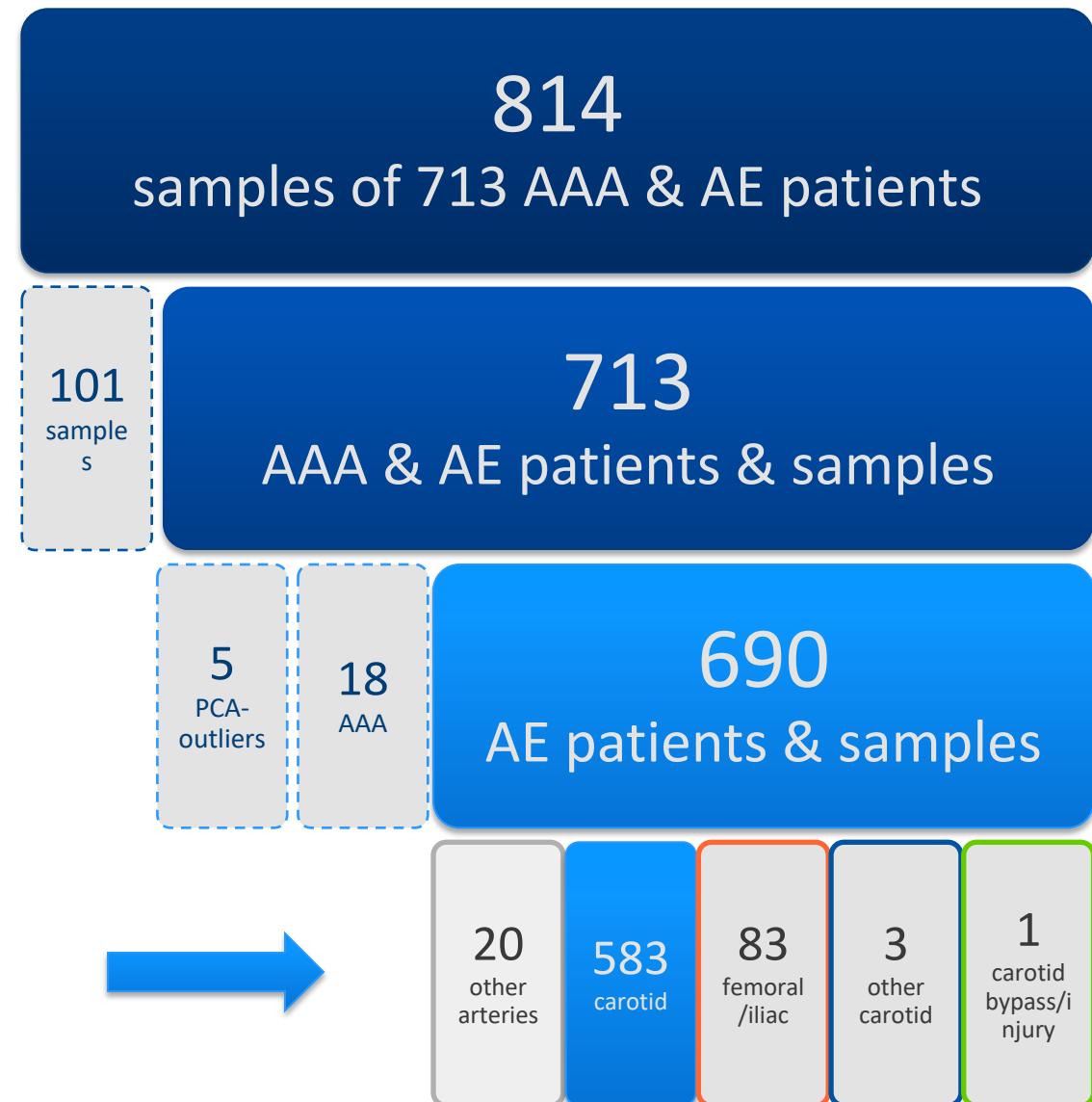
- Fat
- Collagen
- Calcification
- Smooth muscle cell (SMC) content
- Macrophage content
- Neutrophil content
- Mast cell content
- Thrombus presence
- Intraplaque hemorrhage
- Vessel density
- Overall plaque phenotype

# The plaque is the sum of its parts



# Breakdown of samples & patients

- 814 samples of 713 patients
  - Exclude 101 duplicate samples
  - Exclude 18 AAA
  - Exclude 5 PCA-outliers
  - 690 patients can be used (AE)
    - 583 carotid arteries (CEA)
    - 83 femoral arteries (FEA)
    - 4 other carotid artery surgeries
    - 20 other artery surgeries

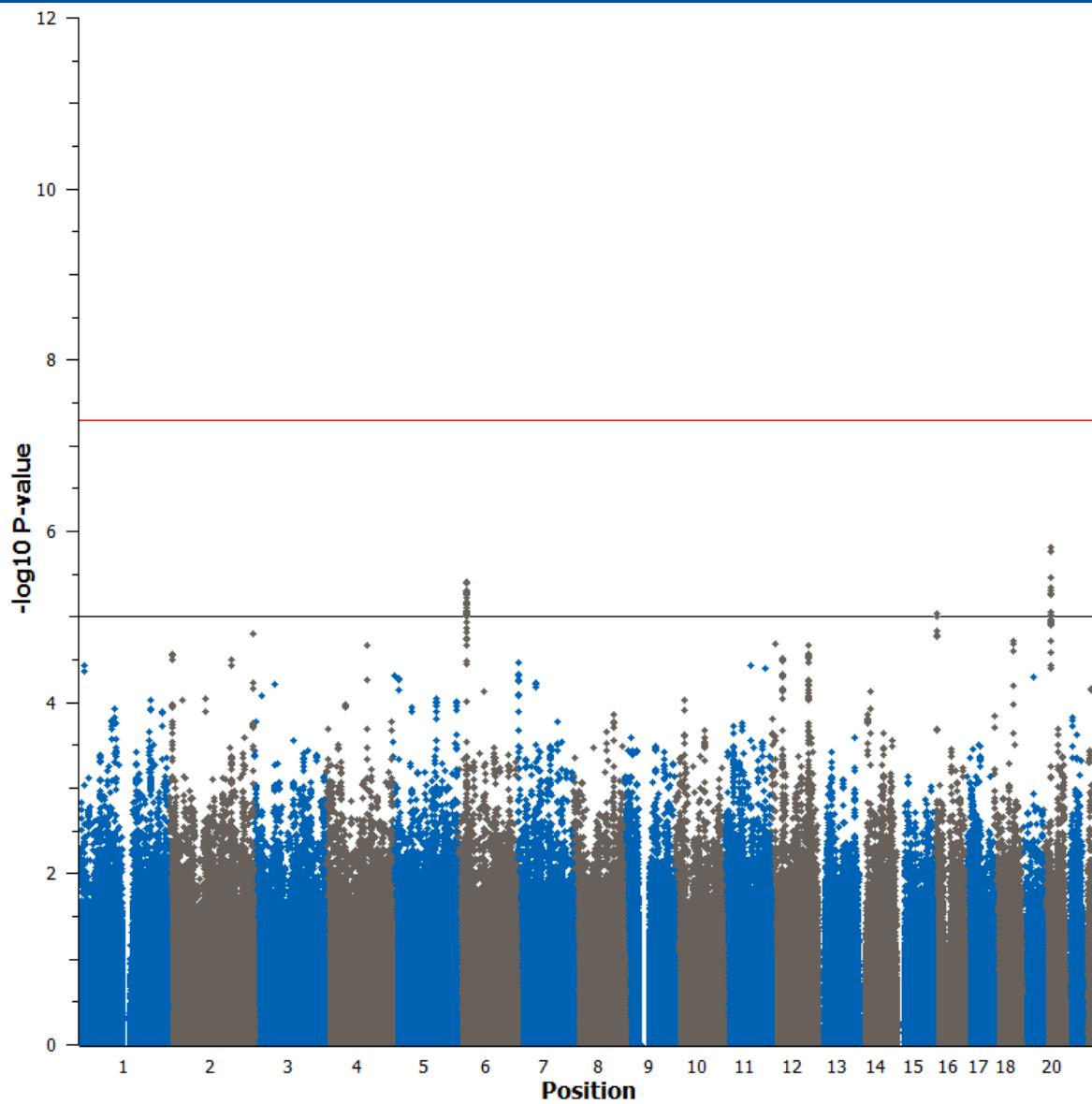


Analyses: Focus on CEA

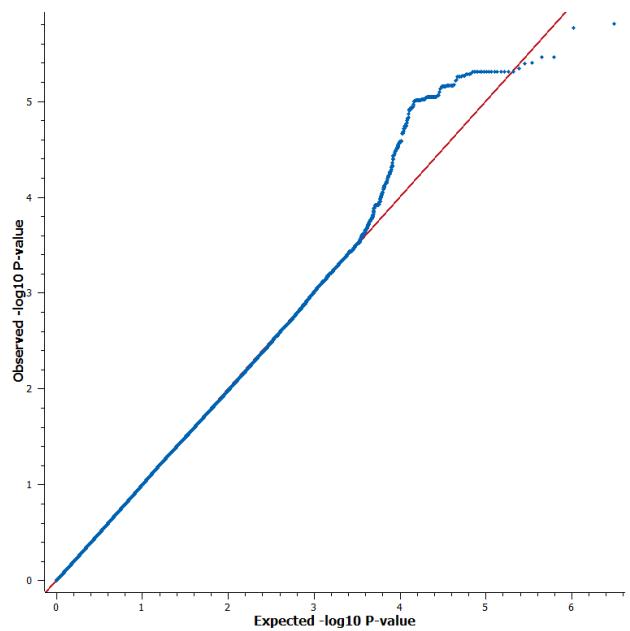
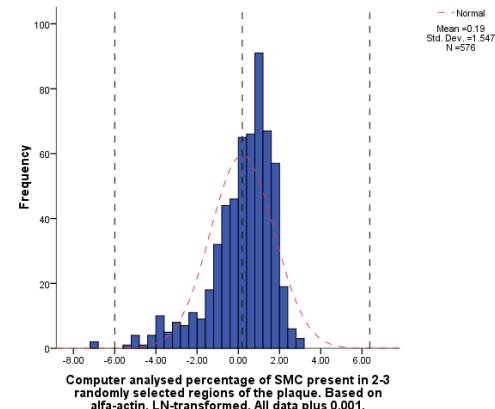


# Percentage smooth muscle cell per area

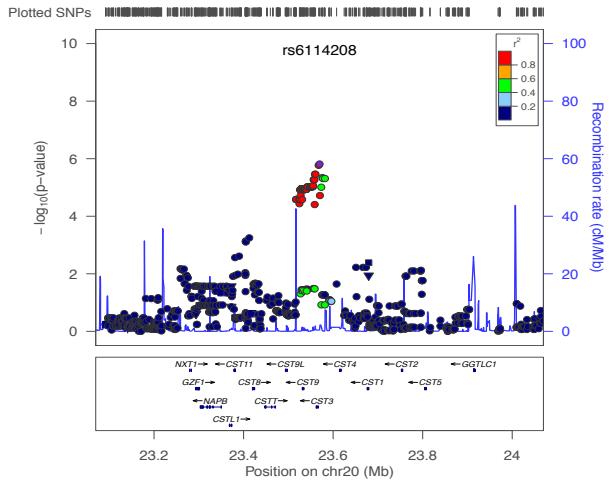
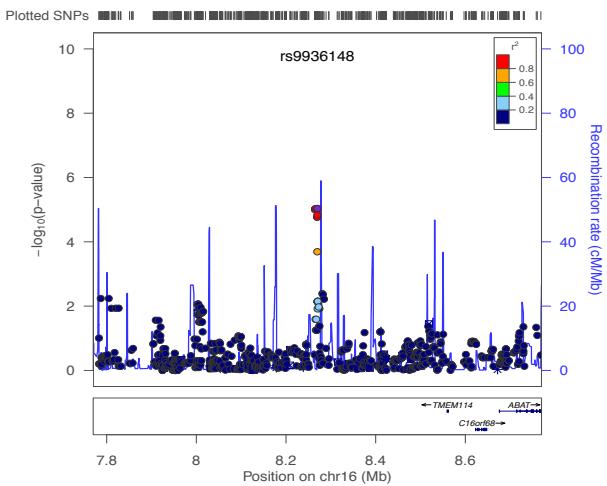
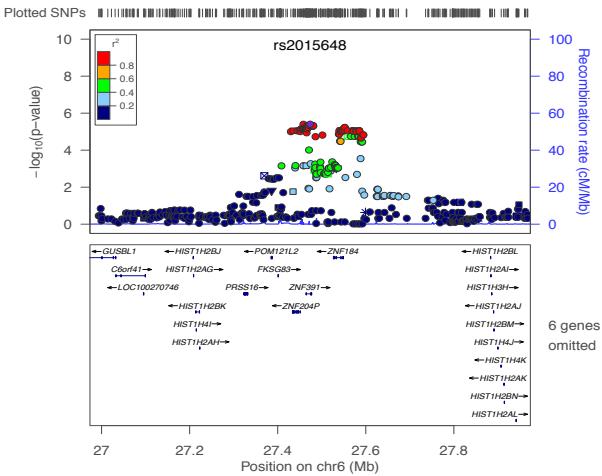
*covariates: age, gender, 10 PCs; BEAGLE  $R^2 \geq 0.90$ ; MAF > 5%*



Smooth muscle cells; N=576;  $\lambda=0.992$

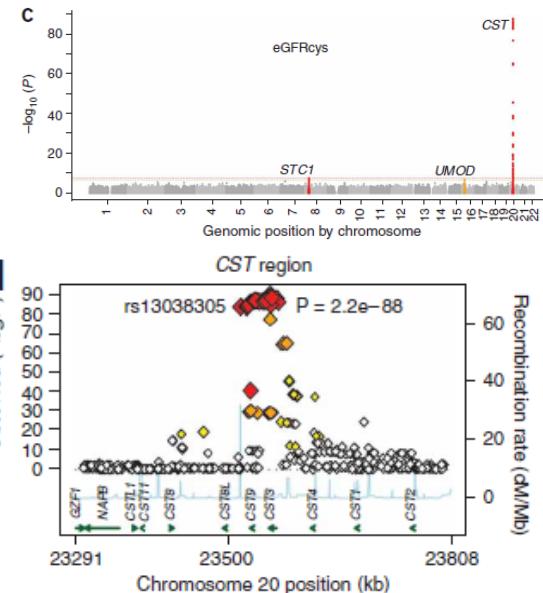


# Loci regions



# Kidney Function & Disease

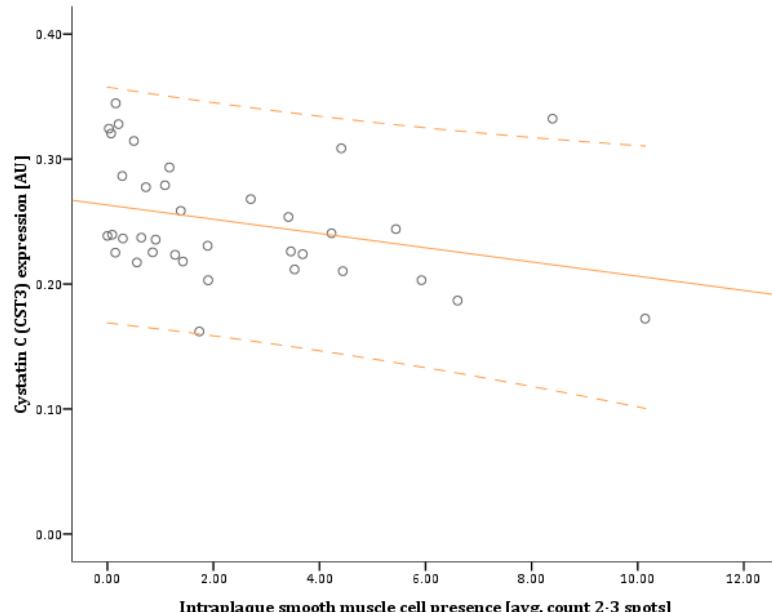
- Three Genome–Wide Association Study Meta-Analyses of eGFR and CKD revealed a region containing *CST3* (Cystatin C) as significantly associated
- We find the same region (*CST3*) significantly associated with smooth muscle cells in the atherosclerotic plaque
  - SNP association study
  - Versatile gene-based association study (VEGAS)



Chromosome	Gene(s)	SNP <i>originally reported</i>	SNP <i>Athero-Express</i>	R <sup>2</sup> <i>between SNPs</i>	P-value <i>SNP</i>	Gene	P-value <i>VEGAS</i>
20	<i>CST3, CST4, CST9</i>	rs911119	rs6114208	0.9113	1.53x10 <sup>-6</sup>	<i>CST3</i>	0.000009
					1.53x10 <sup>-6</sup>	<i>CST4</i>	0.000252
					1.53x10 <sup>-6</sup>	<i>CST9</i>	0.000018

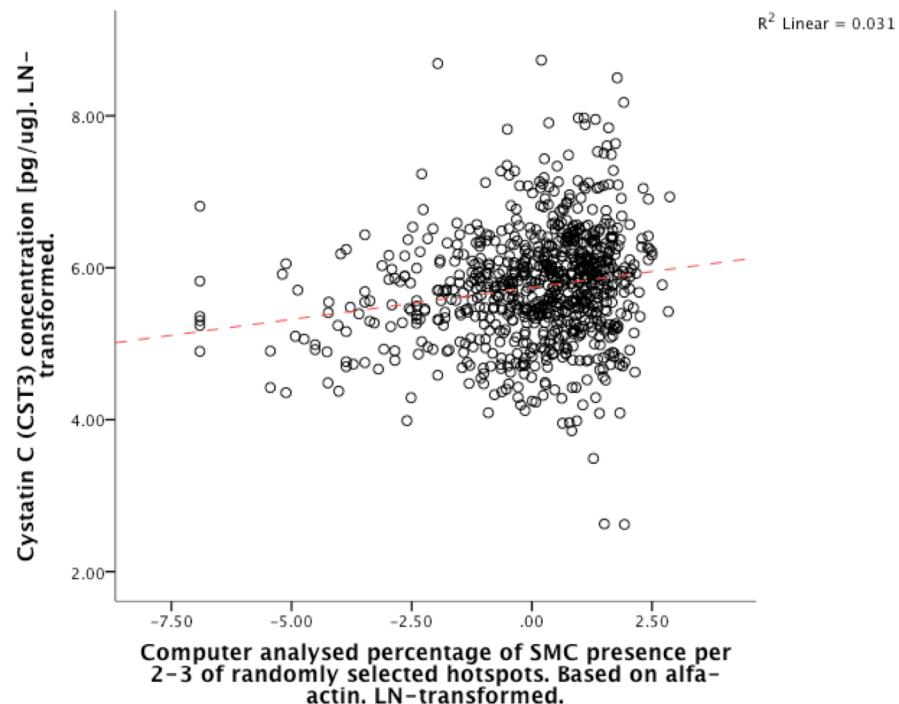
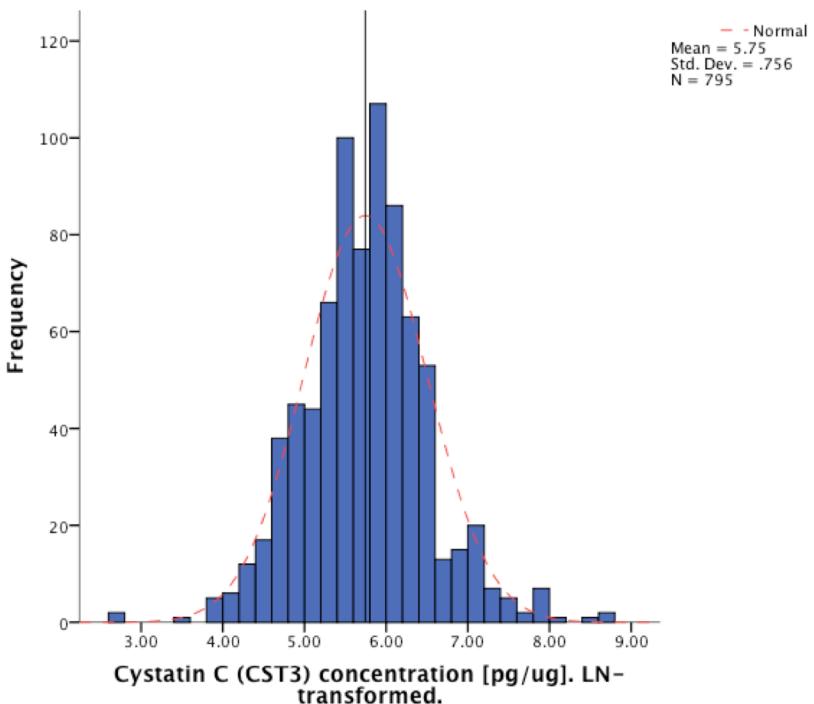
# Quantitative Proteomics substantiates Genomic Analyses

- 40 Athero-Express patients selected for plaque quantitative proteomics
- Cystatin C protein expression associated with intraplaque smooth muscle cells
  - Rho=-0.415, P=0.012
  - Not normally distributed, but untransformed!
  - Literature: Cystatin C is expressed by SMCs, and markedly decreased in atherosclerotic tissue [G.P. Shi, P.M. Ridker, P. Libby, J Clin Invest 1999] → our results do not make sense
- Validation of these results in the remaining Athero-Express CEA cohort:
  - N=1,1711 patients
    - Males: 1,174
    - Females: 537
  - Average age: 68.03 [years]
    - Males: 67.90
    - Females: 68.33
  - Average eGFR: 70.77 [mL/min/1.73m<sup>2</sup>]
    - Males: 71.95
    - Females: 68.04



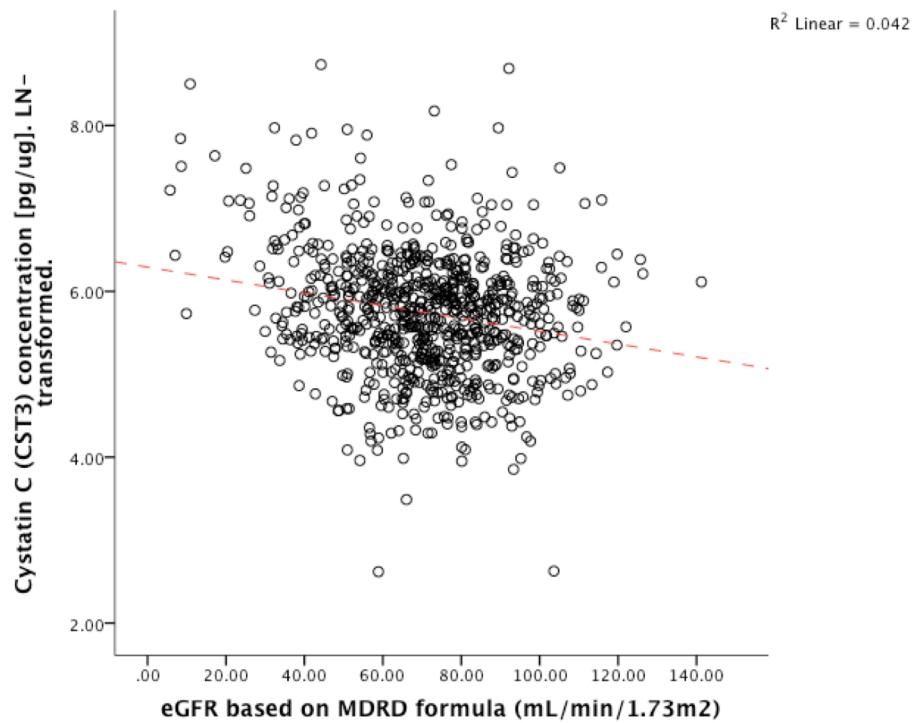
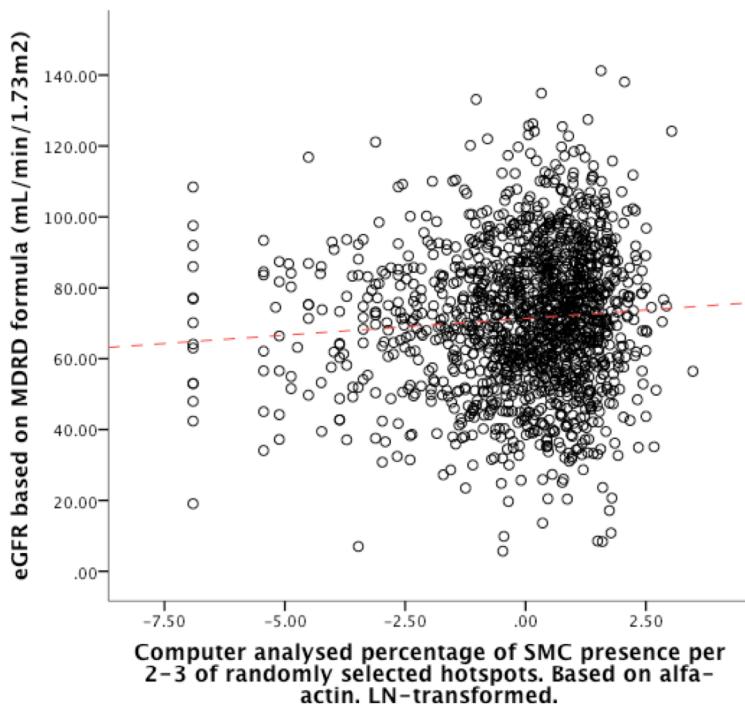
# Luminex Cystatin C plaque levels

- Rho=0.185 ( $R^2=0.176$ ),  $P=1.10\times 10^{-7}$



# Associations with eGFR

- SMC vs. eGFR: Rho=0.084 ( $R^2=0.076$ ),  $P=1.35\times 10^{-3}$
- eGFR vs. CST3: Rho=-0.161 ( $R^2=-0.026$ ),  $P=4.92\times 10^{-3}$

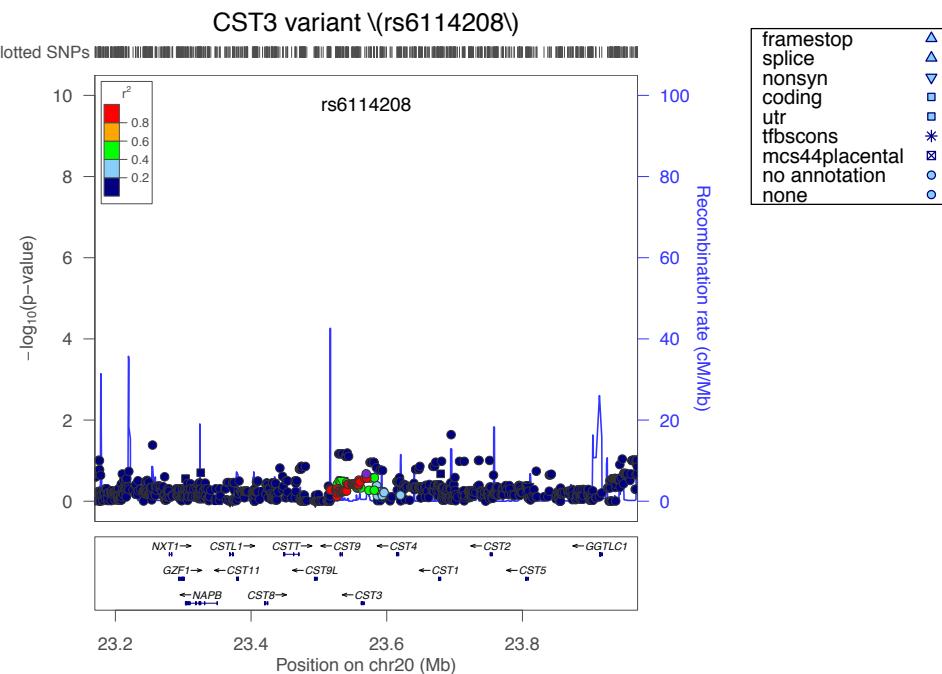
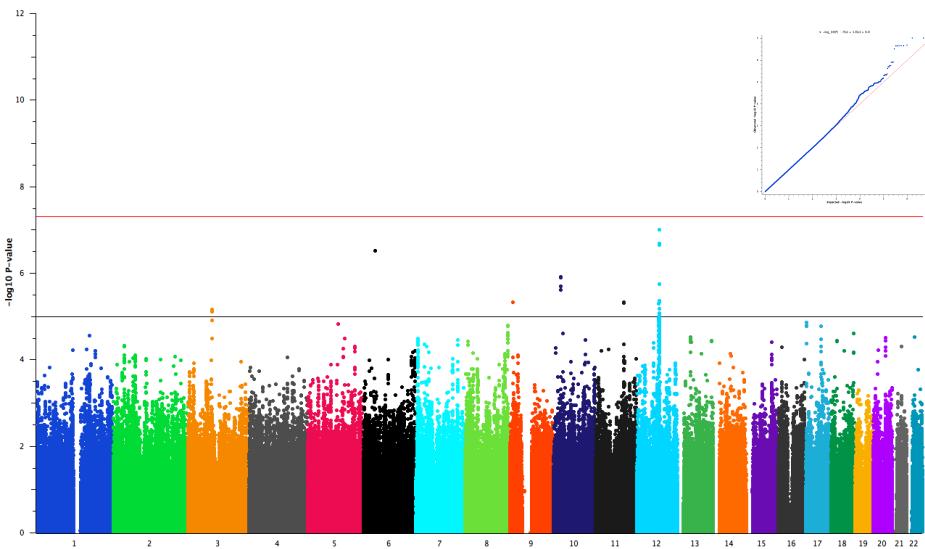


# Preliminary conclusions

- A variant (rs6114208) in *CST3* is associated with SMCs in the plaque
- *CST3* protein levels in the plaque are associated with SMCs
- *CST3* protein levels in the plaque are negatively correlated with  $eGFR_{creat}$
- $eGFR_{creat}$  is positively correlated with SMCs in the plaque
- The variant rs6114208 is not associated with *CST3* plaque levels

# Preliminary GWAS on plaque CST3 levels

- CST3 variant associated with CST3 protein levels in the plaque:
  - rs6114208, effect allele G, EAF=0.2351, INFO-metric=0.8284,  $\beta = -0.0847 \pm 0.0676$ s.e.m.,  $P=0.2111$ ,  $N=432$





University Medical Center  
Utrecht

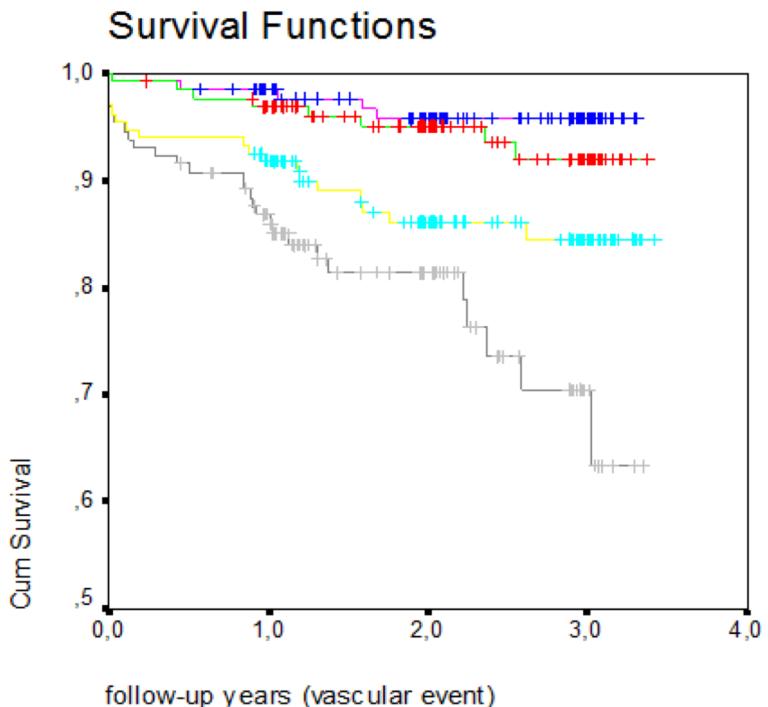
Genetics, Biomarkers & Disease

# OPN AS A BIOMARKER



# Some background information

- Osteopontin (OPN) expression in plaques is predictive for future vascular events
- Is OPN a good biomarker?



## Local Atherosclerotic Plaques Are a Source of Prognostic Biomarkers for Adverse Cardiovascular Events

Dominique P.V. de Kleijn, Frans L. Moll, Willem E. Hellings, Gonen Ozsarlak-Sozer, Peter de Bruin, Pieter A. Doevedans, Aryan Vink, Louise M. Catanzariti, Arjan H. Schoneveld, Ale Algra, Mat J. Daemen, E.A. Biessen, W. de Jager, Huoming Zhang, Jean-Paul de Vries, Erling Falk, Sai K. Lim, Peter J. van der Spek, Siu Kwan Sze, Gerard Pasterkamp

**Objective**—Atherosclerotic cardiovascular disease is a major burden to health care. Because atherosclerosis is considered a systemic disease, we hypothesized that one single atherosclerotic plaque contains ample molecular information that predicts future cardiovascular events in all vascular territories.

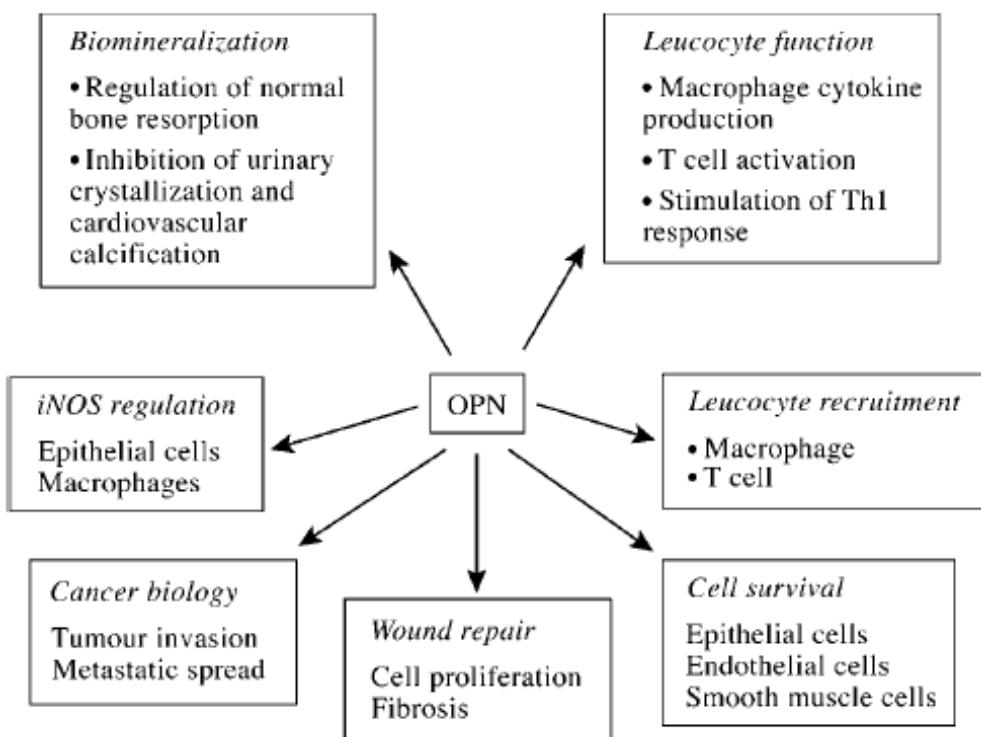
**Methods and Results**—AtheroExpress is a biobank collecting atherosclerotic lesions during surgery, with a 3-year follow-up. The composite primary outcome encompasses all cardiovascular events and interventions, eg, cardiovascular death, myocardial infarction, stroke, and endovascular interventions. A proteomics search identified osteopontin as a potential plaque biomarker. Patients undergoing carotid surgery ( $n=574$ ) served as the cohort in which plaque osteopontin levels were examined in relation to their outcome during follow-up and was validated in a cohort of patients undergoing femoral endarterectomy ( $n=151$ ). Comparing the highest quartile of carotid plaque osteopontin levels with quartile 1 showed a hazard ratio for the primary outcome of 3.8 (95% confidence interval, 2.6–5.9). The outcome did not change after adjustment for plaque characteristics and traditional risk factors (hazard ratio, 3.5; 95% confidence interval, 2.0–5.9). The femoral validation cohort showed a hazard ratio of 3.8 (95% confidence interval 2.0 to 7.4) comparing osteopontin levels in quartile 4 with quartile 1.

**Conclusion**—Plaque osteopontin levels in single lesions are predictive for cardiovascular events in other vascular territories. Local atherosclerotic plaques are a source of prognostic biomarkers with a high predictive value for secondary manifestations of atherosclerotic disease. (*Arterioscler Thromb Vasc Biol*. 2010;30:612–619).

Key Words: arterectomy ■ atherosclerosis ■ biomarker ■ plaque

# More on OPN

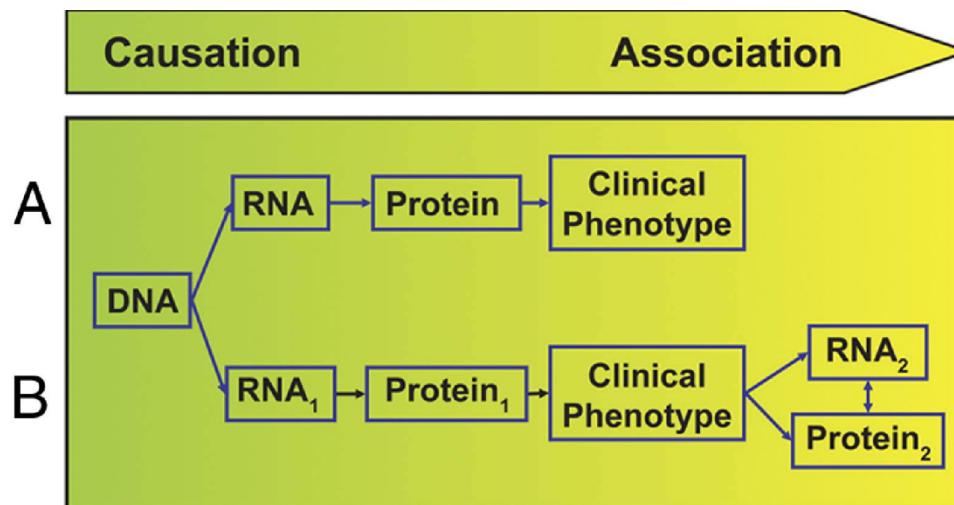
- What constitutes a good biomarker?
  - Mechanistically involved
  - Preferably causal
  - Measurable
  - Treatable
    - Biomarker of drug efficacy



# Biomarkers & Disease

## *chicken or the egg*

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



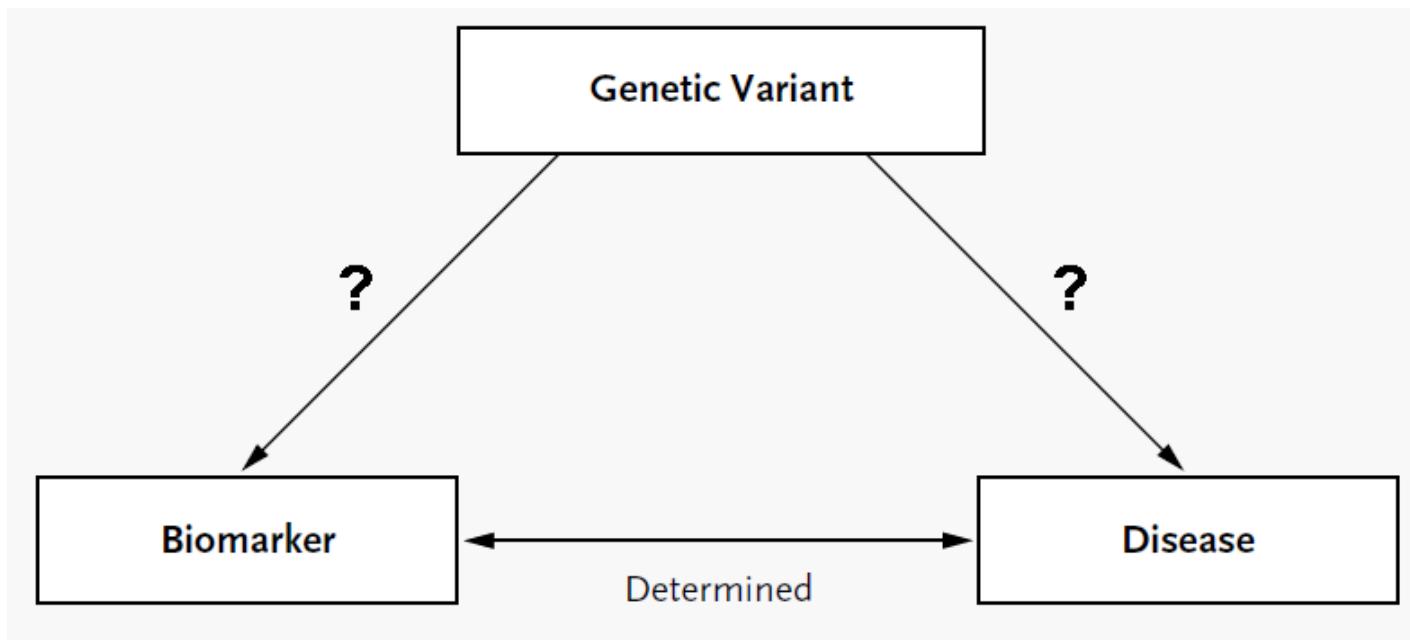
# SNPs, Biomarkers & Disease

*can genetic variation tell us something?*



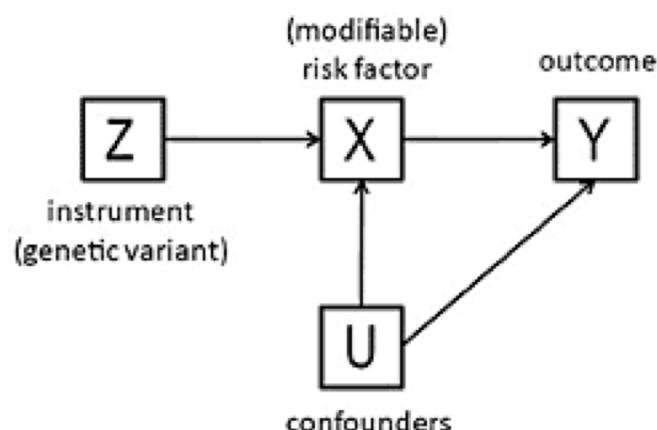
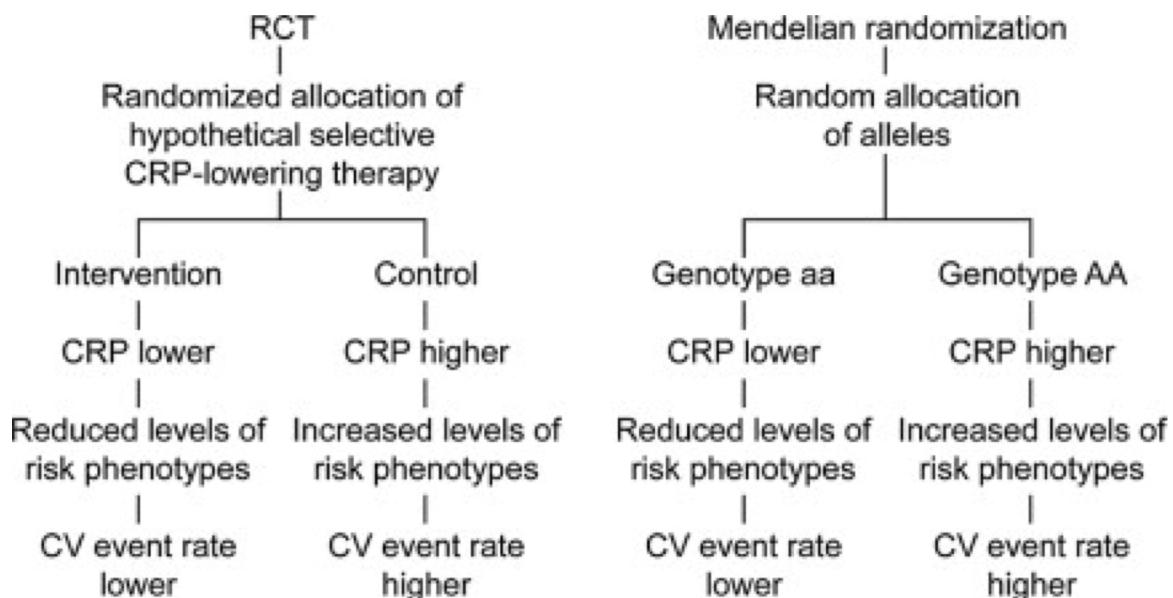
University Medical Center  
Utrecht

- 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



# Mendelian Randomization of OPN variation, expression and CVD

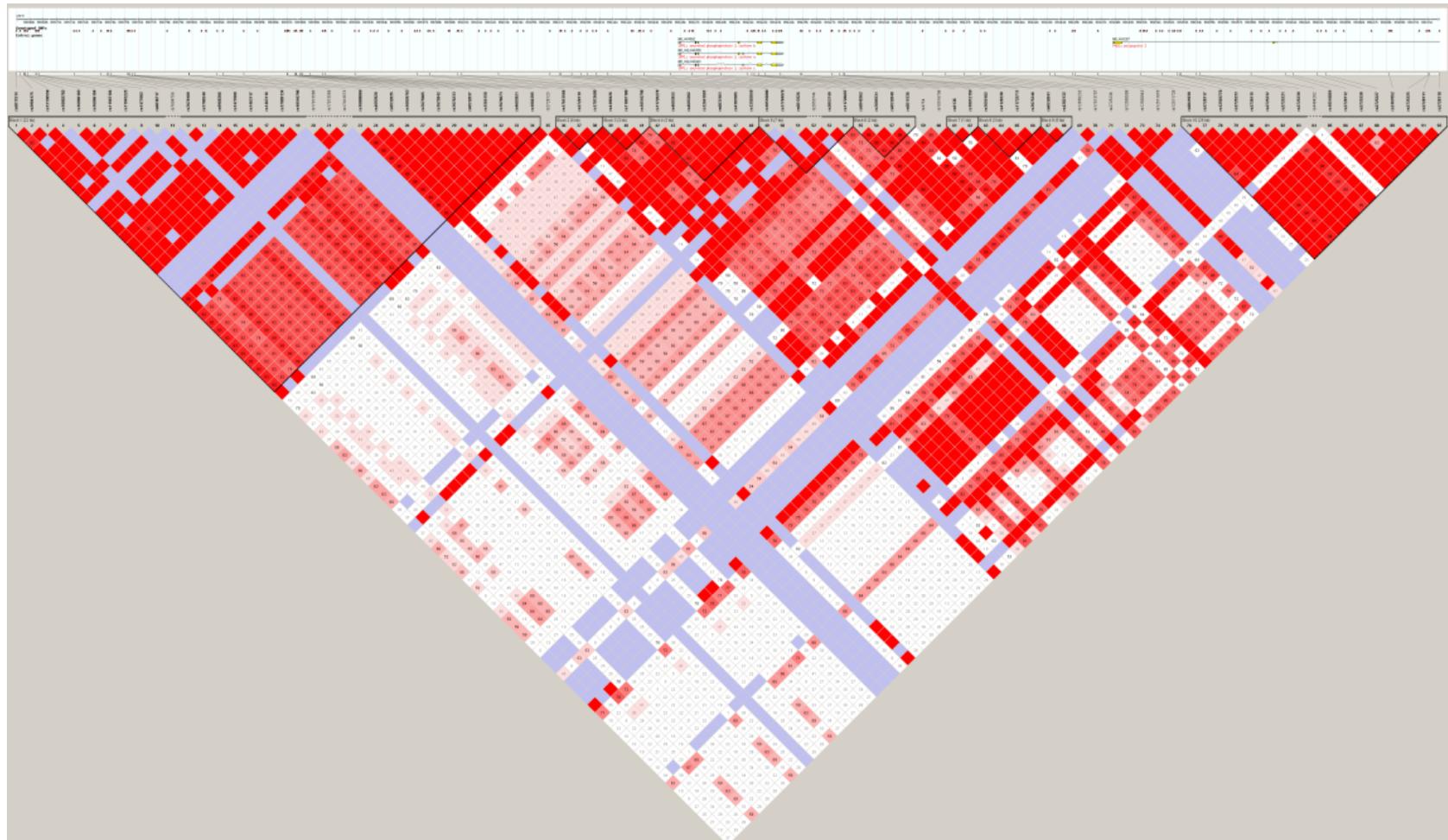


University Medical Center  
Utrecht

- OPN variant vs. CVD outcome
- OPN variant vs. OPN serum levels
- OPN serum levels vs. CVD outcome
- If all have the same observed and expected significant effect, OPN is likely to be causal to CVD

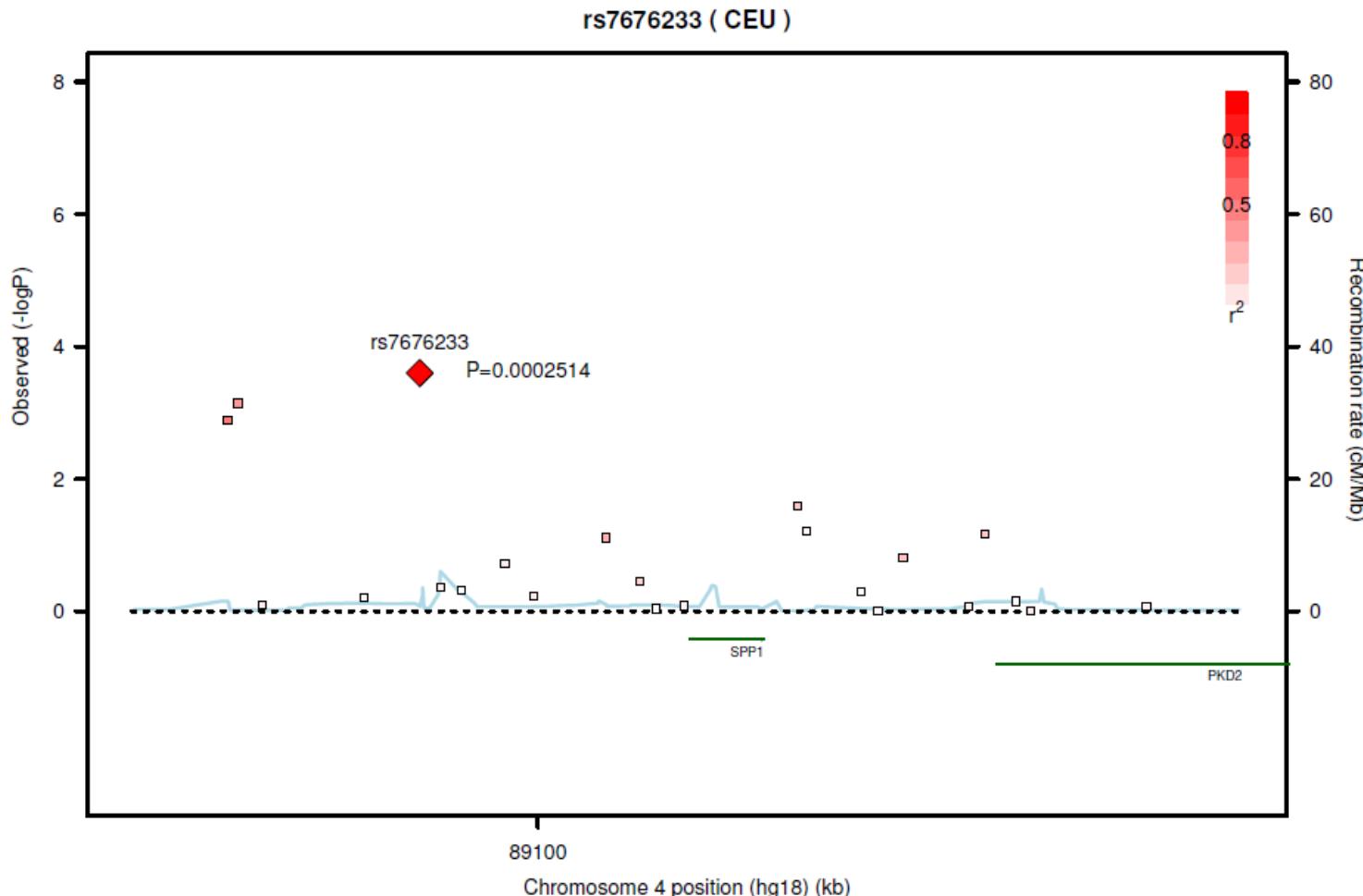
# SNPs tagging OPN

- Look at LD-plot



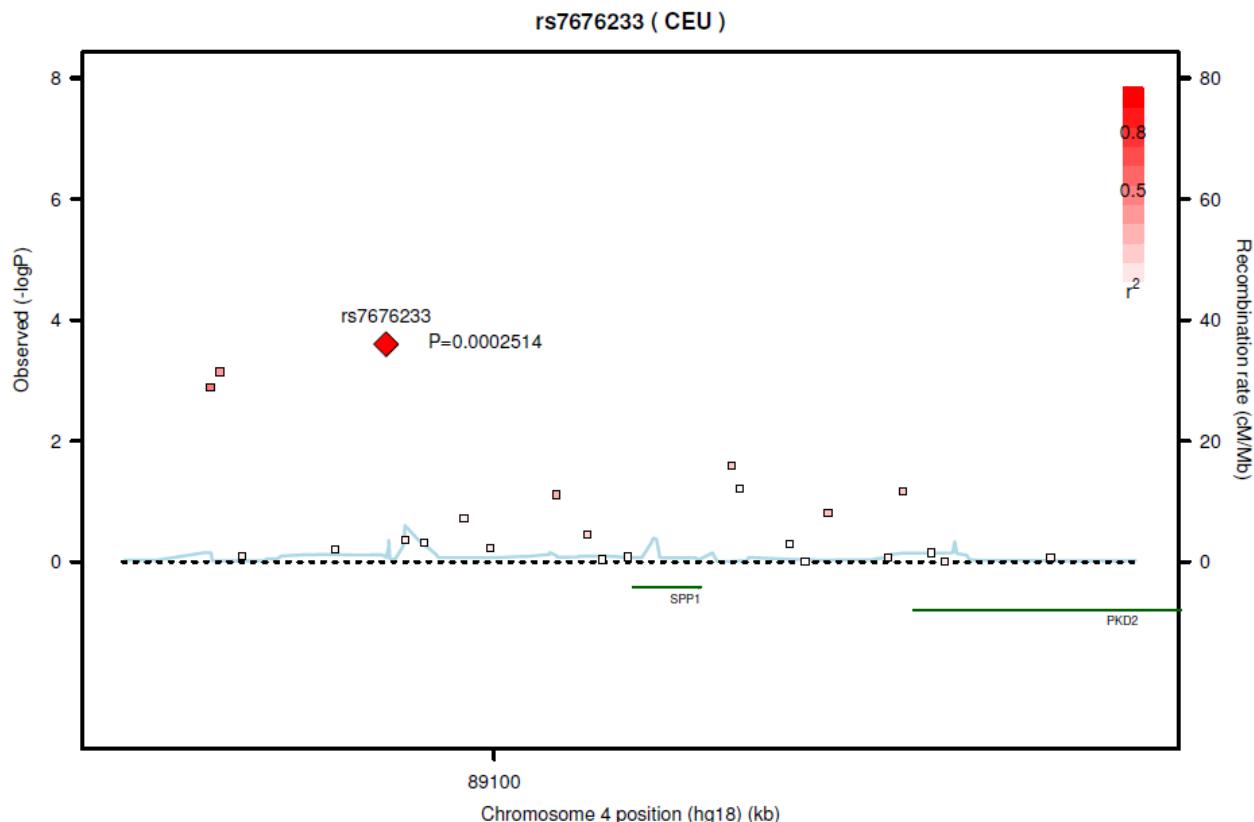
# SNPs associated with OPN plasma expression

- Associate each tagSNP with the expression of OPN in the plasma



# OPN & SNPs: what's next?

- Associate significant SNPs with disease
- Associate plaque expression with disease



# Conclusions #1

- Genetics is successful in identifying mutations causing mono-genetic diseases
- Next-generation sequencing will further elucidate Mendelian traits
- A near-complete catalog of common and rare variation is available to investigate complex traits

# Conclusions #2

- GWAS identified many genomic loci of modest effect that influence complex traits
- Collectively, these associated SNPs explain little of heritability. Explaining the missing heritability will be the challenge of the next years
- For most loci, causal (functional) variants are still unknown

# Conclusions

- A GWAS of plaque phenotypes
  - Might identify genes involved in atherosclerotic plaque formation, development and progression
  - Explain association of vessel density with disease
- A genetic burden score
  - Could stratify patients in high and low risk groups
- Mendelian Randomization can be used
  - To understand the association of biomarkers with disease
  - Assess the causality of a biomarker

PHASE : TWO : INTERPRETATION

I THINK I  
FOUND  
A CORNER  
PIECE.



# 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