

# 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



# Introduction

- Basic Genetics
- Genetics & Family history
- Types of Studies
- WTCCC

**So what about genetics in atherosclerosis and consequent cardiovascular disease?**

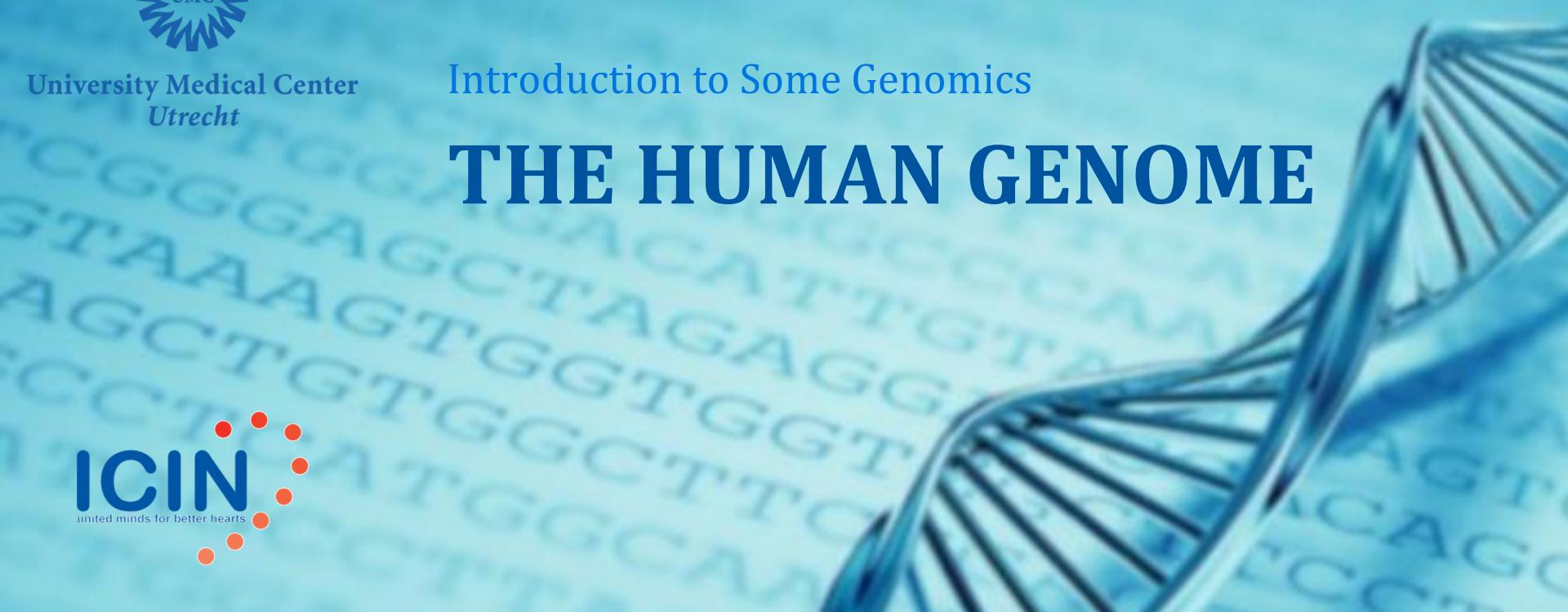


University Medical Center  
Utrecht



Introduction to Some Genomics

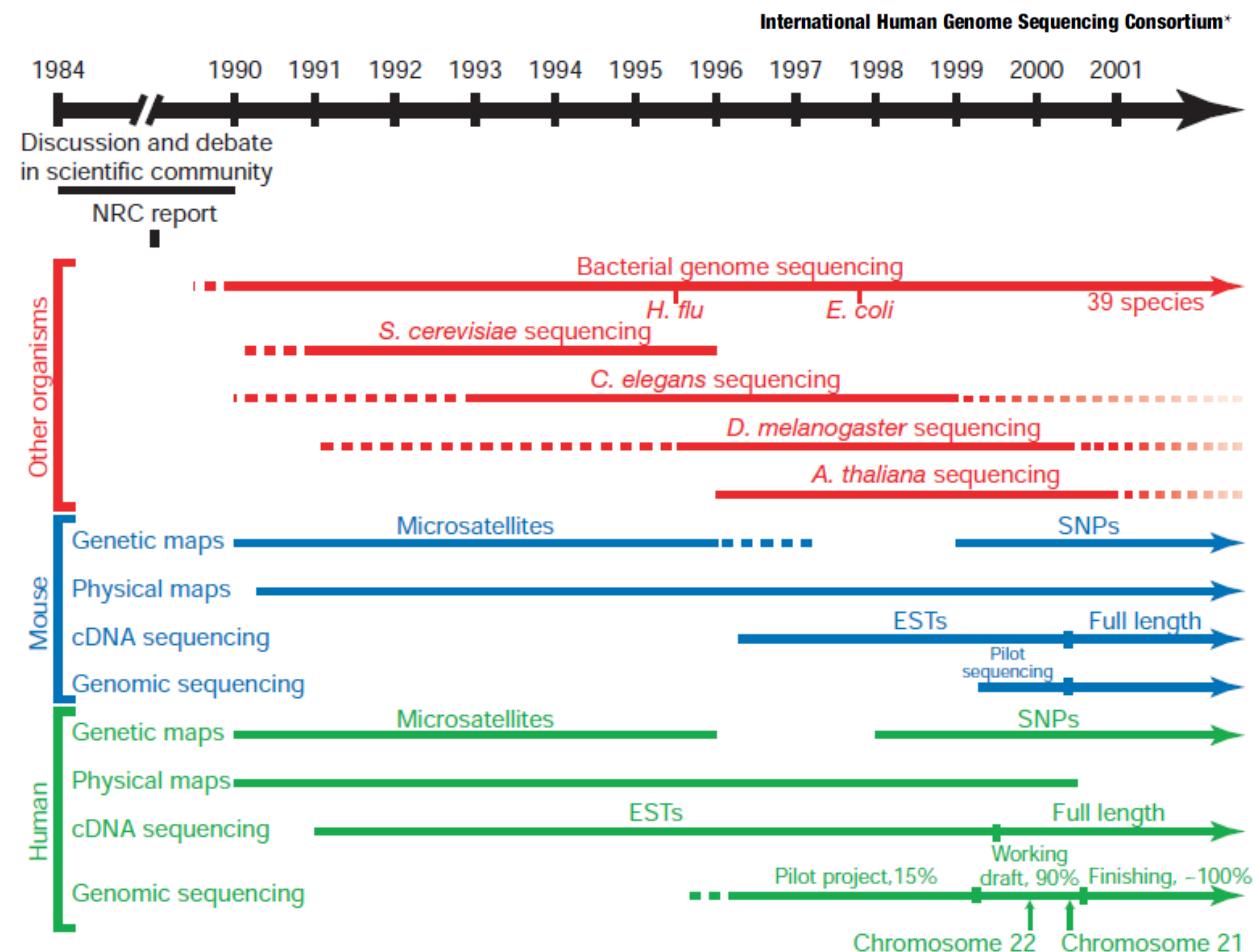
# THE HUMAN GENOME



# Human Genome Project: 2001

## articles

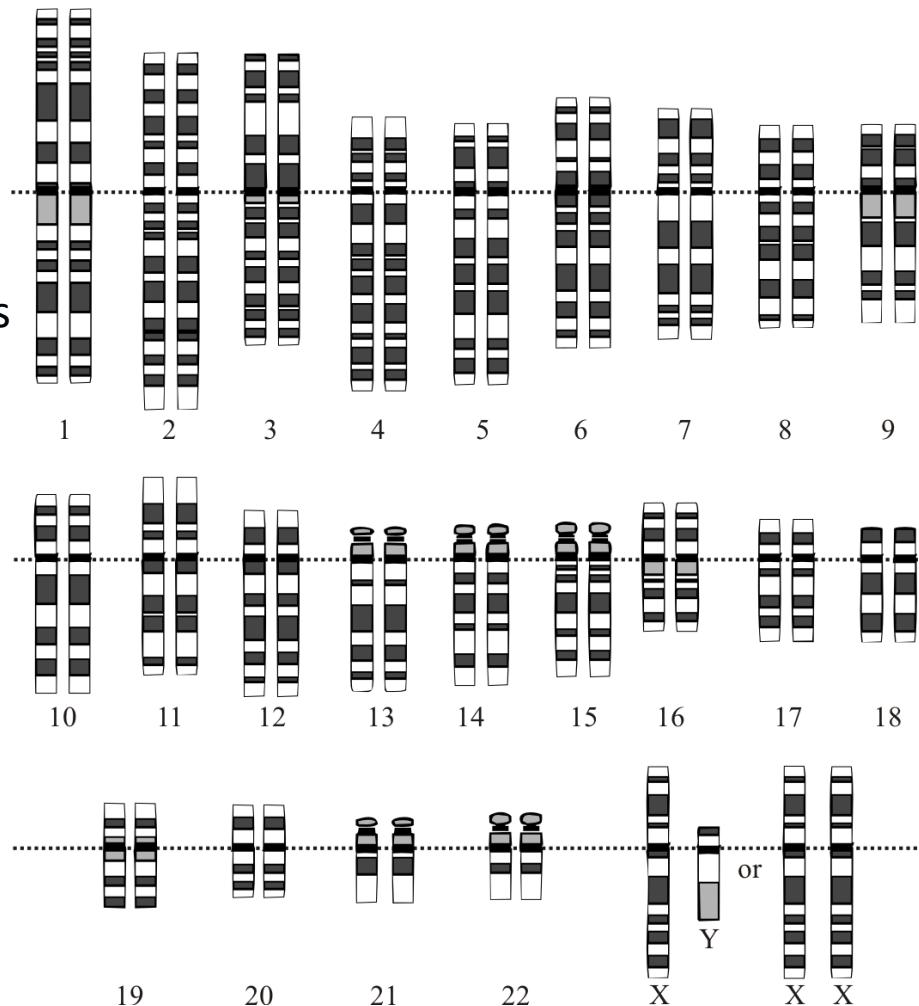
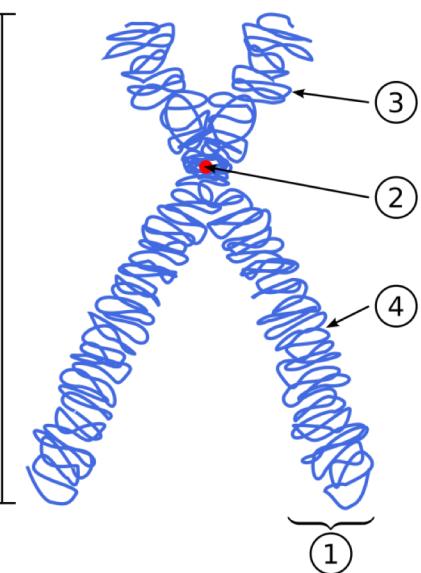
# Initial sequencing and analysis of the human genome



- 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
  - One copy of one (part) of the chromosome(s) of each of your parents
    - 1) chromatid
    - 2) centromere
    - 3) short arm, p
    - 4) long arm, q



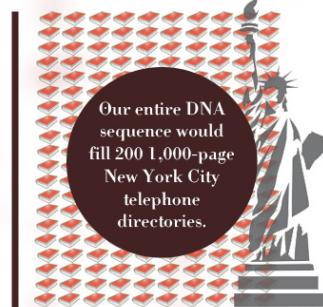
# Human Genome: *some statistics*

- 3.2 billion base pairs in the haploid genome
- $\approx$  20,000-25,000 genes
  - $\approx$  23,000 coding for proteins
  - Only 1.5% of the total genome
- Rest of the genome:
  - Non-coding RNA (rRNA, tRNA)
  - Regulatory sequences, e.g. promoter, enhancer regions
  - Repetitive elements
    - 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"...)

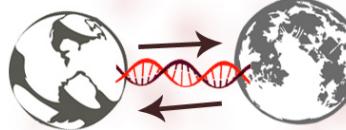
Our entire DNA sequence is called a genome... and there's an estimated **3,000,000,000** DNA bases in our genome.



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.**

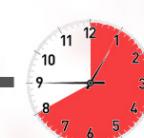


**99.9%** OF OUR DNA SEQUENCE IS THE SAME AS OTHER HUMANS'.



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.

ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGCATCGATCCATT  
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTTACCCAT  
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCA  
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGG  
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGCACATATCGTCATACATAGA  
TCGTAUTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTACCCAT  
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTAT  
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCG  
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTAUTGACTGTCTAGTCTAAACACATCCACATATCGTC  
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGCATCGTACTGACTGTCTAGTCTAAACACATCCACATATCG  
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTAT  
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCG  
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTAUTGACTGTCTAGTCTAAACACATCCACATATCGTC  
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCA  
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTA  
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCG  
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTAUTGACTGTCTAGTCTAAACACATCCACATATCGTC  
CGTACTGACTGTCTAGTCTAAACACATCCCACTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCA  
TATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATAGCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGG  
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGG  
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGACTGCACATATCGTCATACATAGAC  
CGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTACCCAT  
ATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCAT  
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGG  
CTGCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACGCCGATCGTACTGACTGCACATATCGTCATACATAGACT  
GTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCACTTACCCATG  
ATCGTACTCGTACTGACTGTCTAGTCTAAACACATCCCACACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTAC  
CGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTAC

**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.

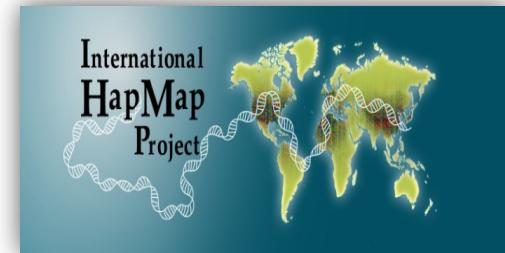
---

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.

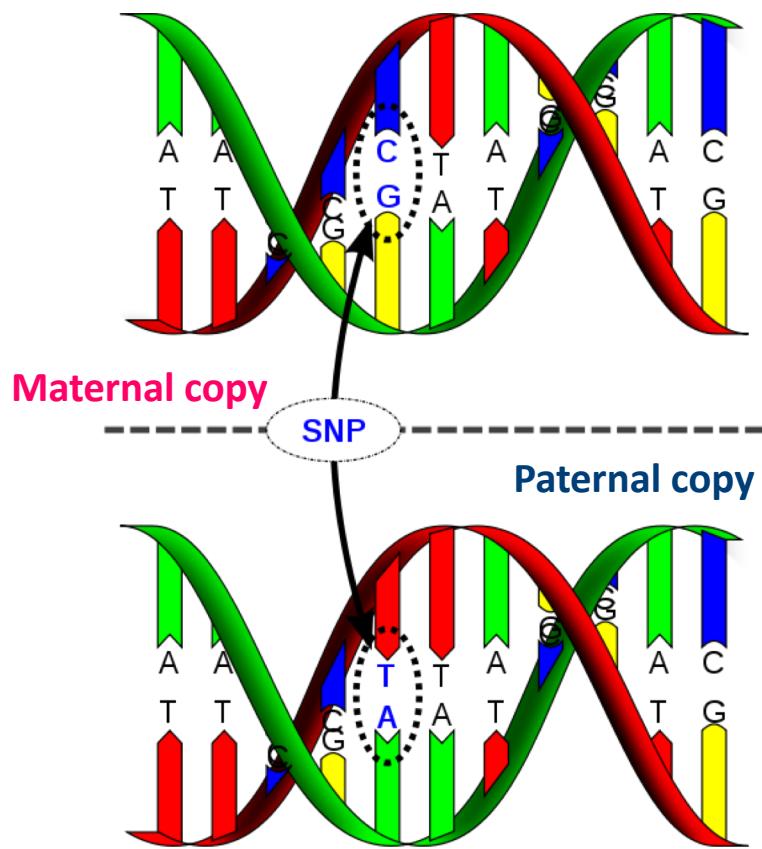
# Single-Nucleotide Polymorphism

*a genetic variation as proxy*

- Single-nucleotide polymorphism (SNP)
- “one base pair variation”
  - > 1% general population
  - ≈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

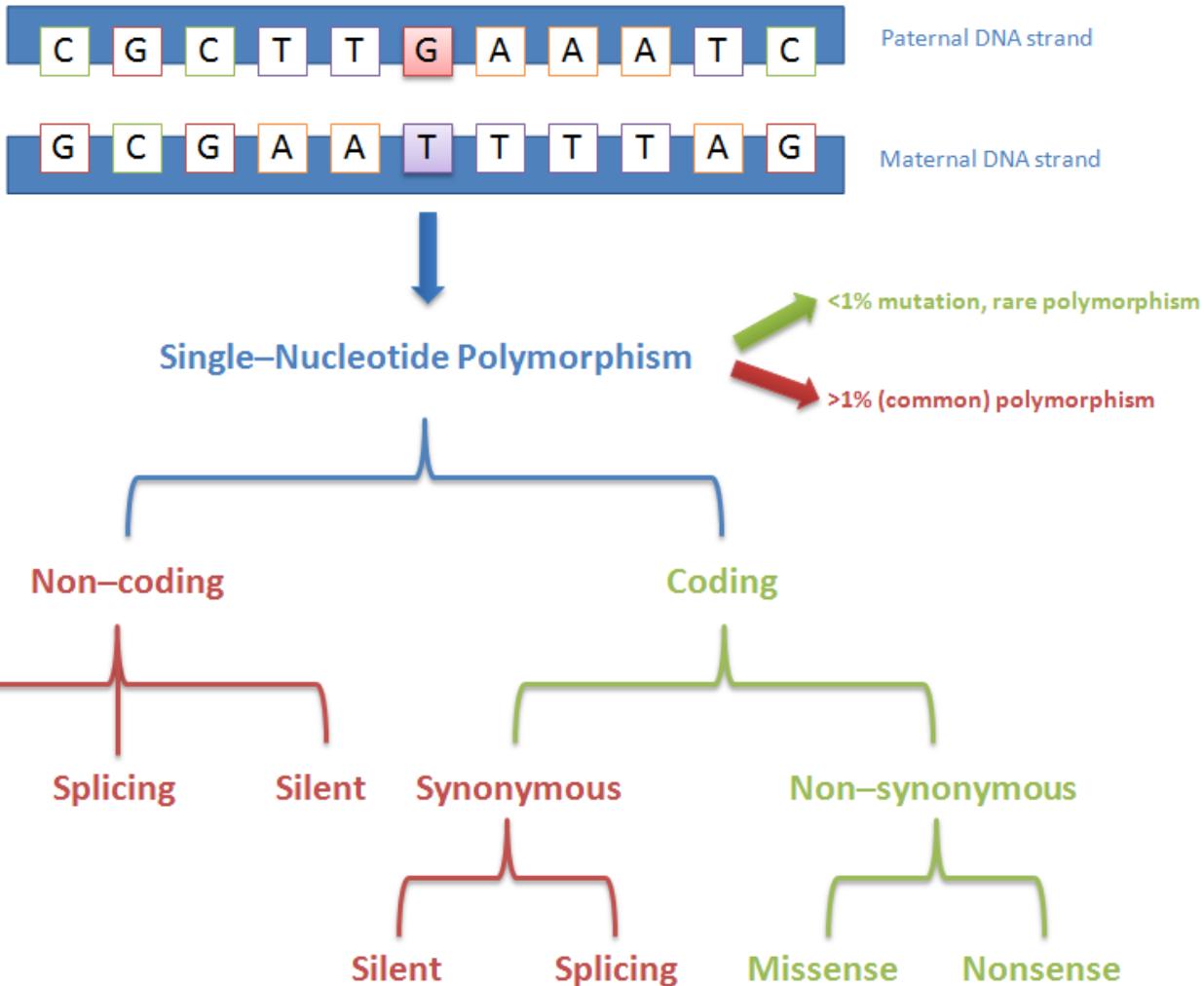


[www.hapmap.org](http://www.hapmap.org)

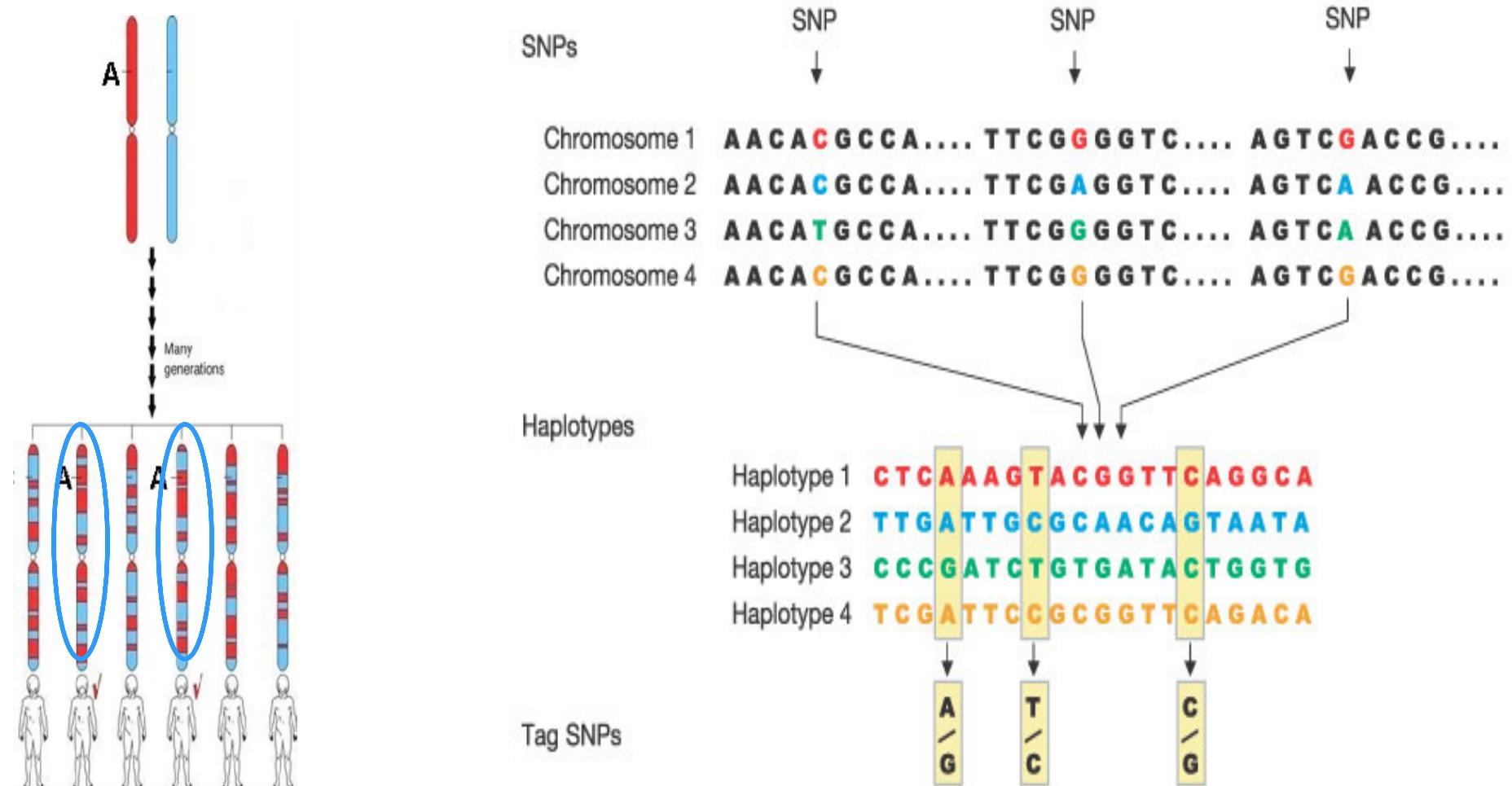


# Types of SNPs

## Chromosome

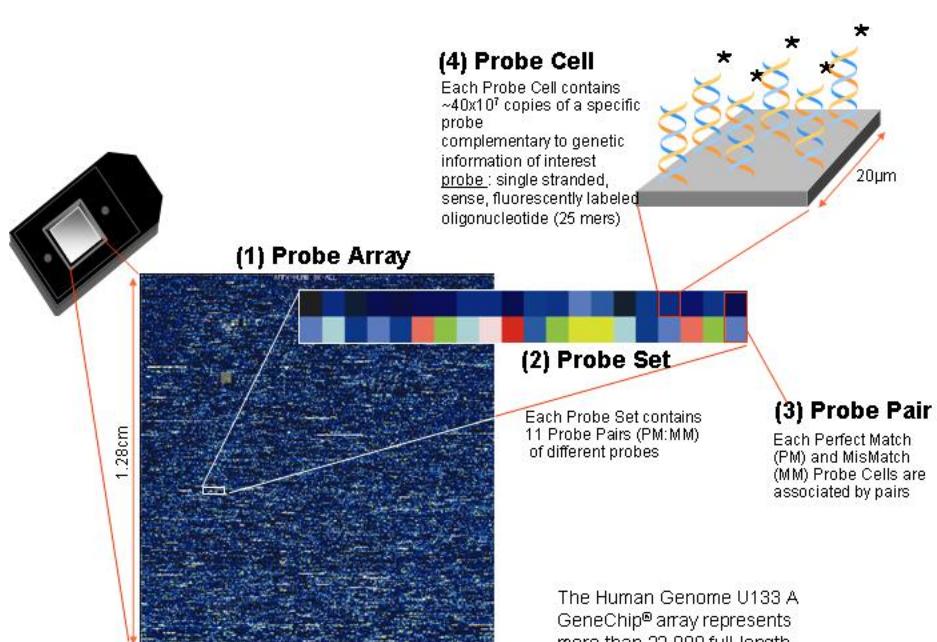
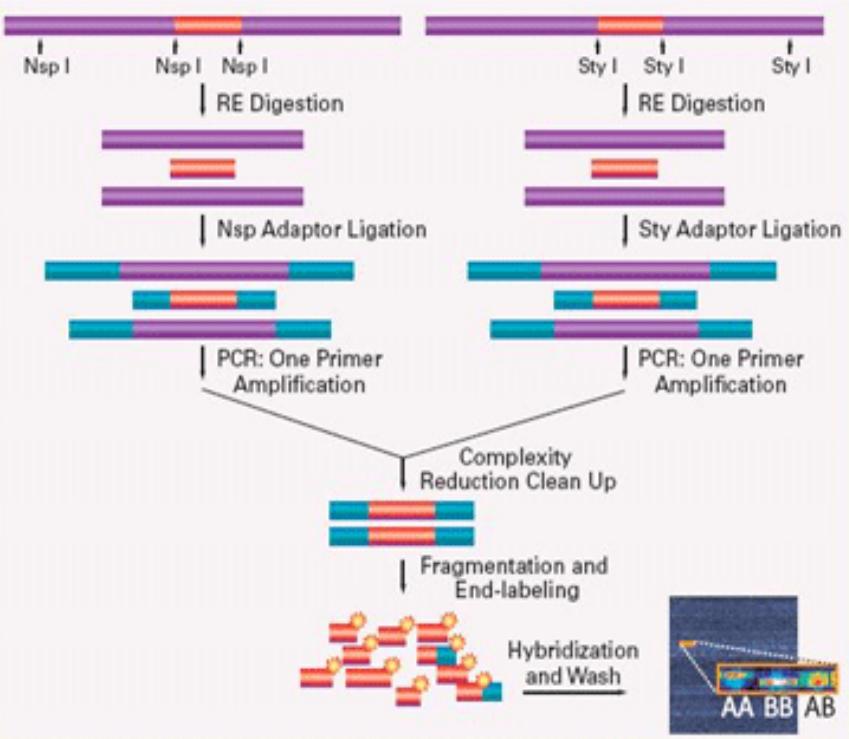


# SNPs → Haplotype → tagSNP



# SNP “genotyping”

The fifth-generation Whole-genome Sampling Assay.



# Genotyping Platforms



# Genotyping Platforms: *examples*

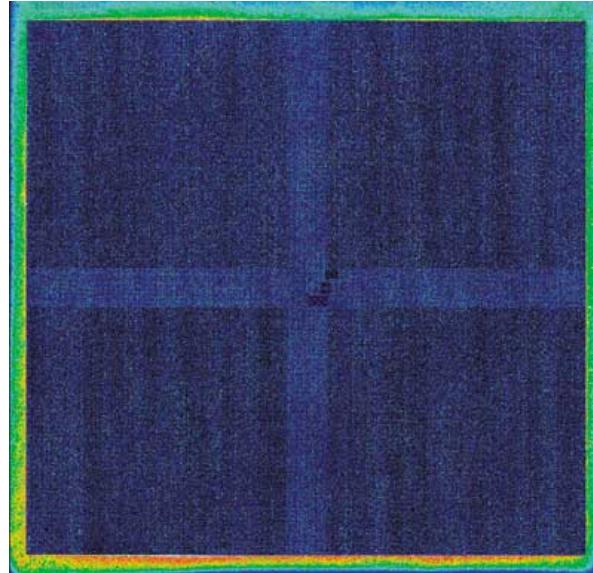
- Affymetrix Genome-Wide Human SNP Array 5.0
  - 500,568 SNPs
  - Chromosomal coverage
    - Good representation of autosomal chromosomes
    - X-chromosome poor representation
    - No Y-chromosome representation
    - No mitochondrial chromosome representation
- Illumina Human660W-Quad v1
  - 657,366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

# Genotyping Platforms: *genotype calling*

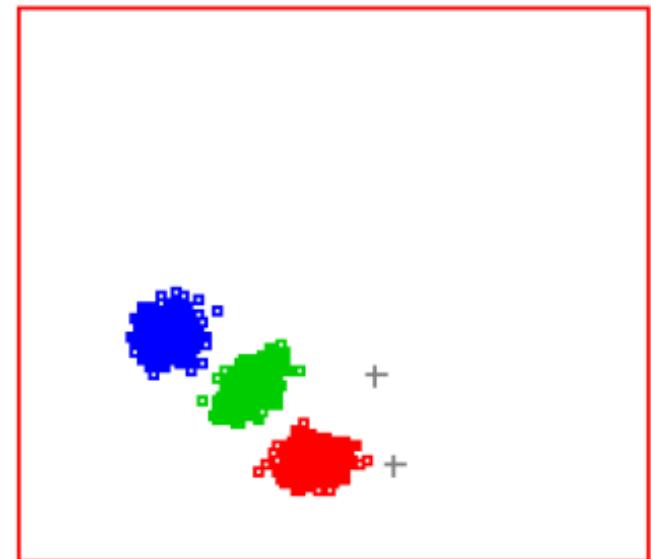
- Several *mathematical algorithms* exist
- Several software programs use different algorithms
- All are platform independent (Illumina, Affmetrix)
  - BEAGLECALL
  - Birdsuite
  - BRLMM-P



One picture per chip/patient



Clustering for each SNP!!!





University Medical Center  
Utrecht

# CARDIOVASCULAR GENOMICS



# 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

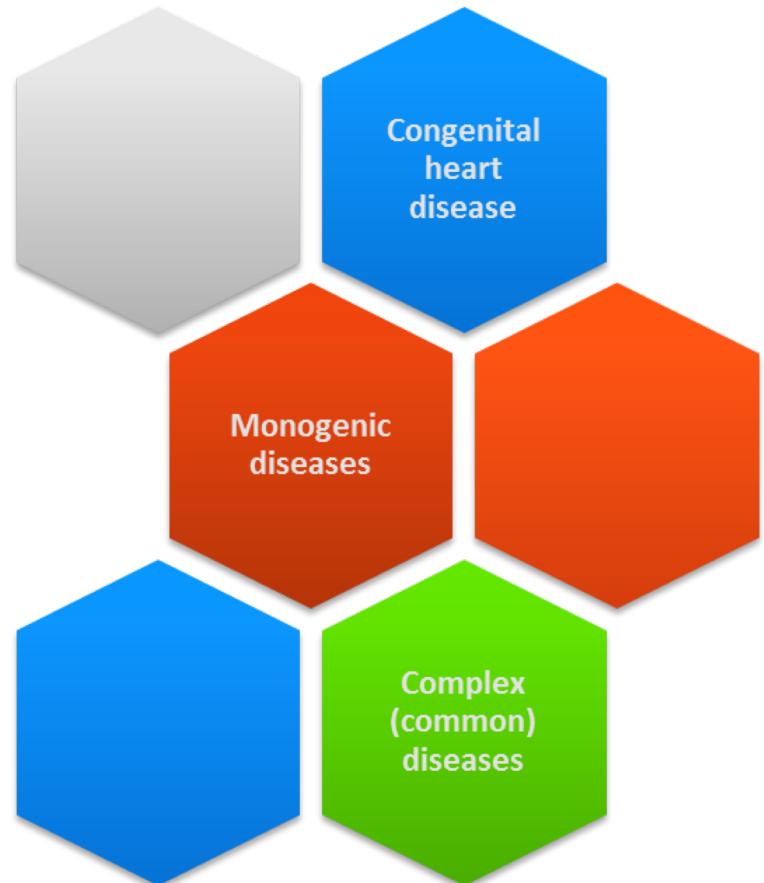
*There is clearly a heritability factor for atherosclerotic and consequent cardiovascular disease*

# What type of disease are we looking at?

## *monogenic diseases*

- **Monogenic diseases**

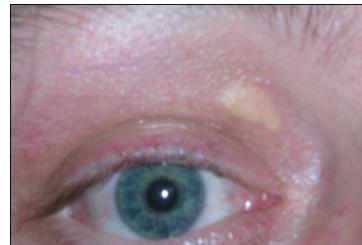
- Mendelian pattern
  - Autosomal dominant, e.g.:
    - Marfan Syndrome
    - **Familiar hypercholesterolemia**
  - Autosomal recessive , e.g.:
    - Sickle cell anemia
    - Cystic fibrosis
  - X-linked , e.g.:
    - Duchene muscular dystrophy
  - Y-linked/Mitochondrial



# Monogenic disease

## Familial Hypercholesterolemia (FH)

- Characterized by high cholesterol levels (LDL)
  - Genetic variations in *cholesterol metabolism genes*
    - One copy of mutant *LDLR* (heterozygote) CVD before 30-40 years
    - Two copies of mutant *LDLR* (homozygote) severe cardiovascular problems in childhood
      - Heterozygous FH: 1:500
      - Homozygous FH: 1:1,000,000
- Genes involved:
  - *LDLR*, LDL-receptor (prevalence 1 in 500)
    - Class I: *LDLR* is not synthesized
    - Class II: No proper transportation of *LDLR* endoplasmic reticulum to the Golgi apparatus
    - Class III: *LDLR* does not properly bind LDL on the cell surface (*apoB/LDLR*)
    - Class IV: *LDLR* bound to LDL does not properly cluster in clathrin-coated pits for receptor-mediated endocytosis
    - Class V: No recycling to the cell surface of *LDLR*
  - *ApoB*, apoprotein B100 (prevalence 1 in 1,000)
    - Mutation of one amino acid (arginine to glutamine) no proper binding of *LDLR*
  - Other genes: *PCSK9*, *LDLRAP1* (prevalence less than 1 in 2,500)



Xanthelasma palpebrarum

# 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*

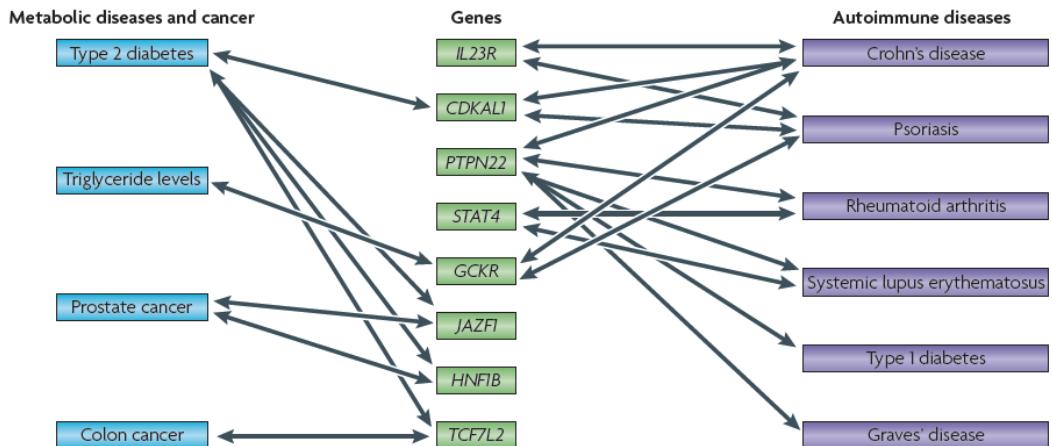


Table 3

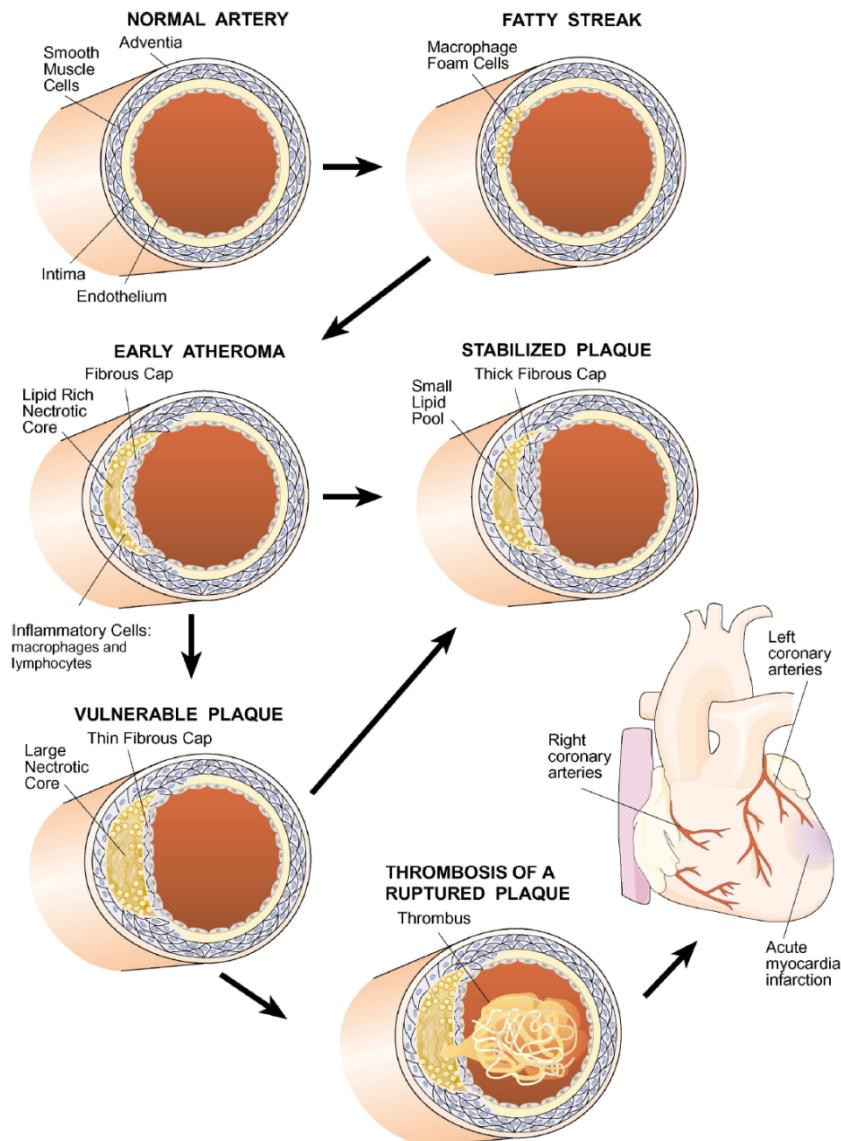
Some Recent Genes/Loci Identified in Coronary Artery Disease

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.

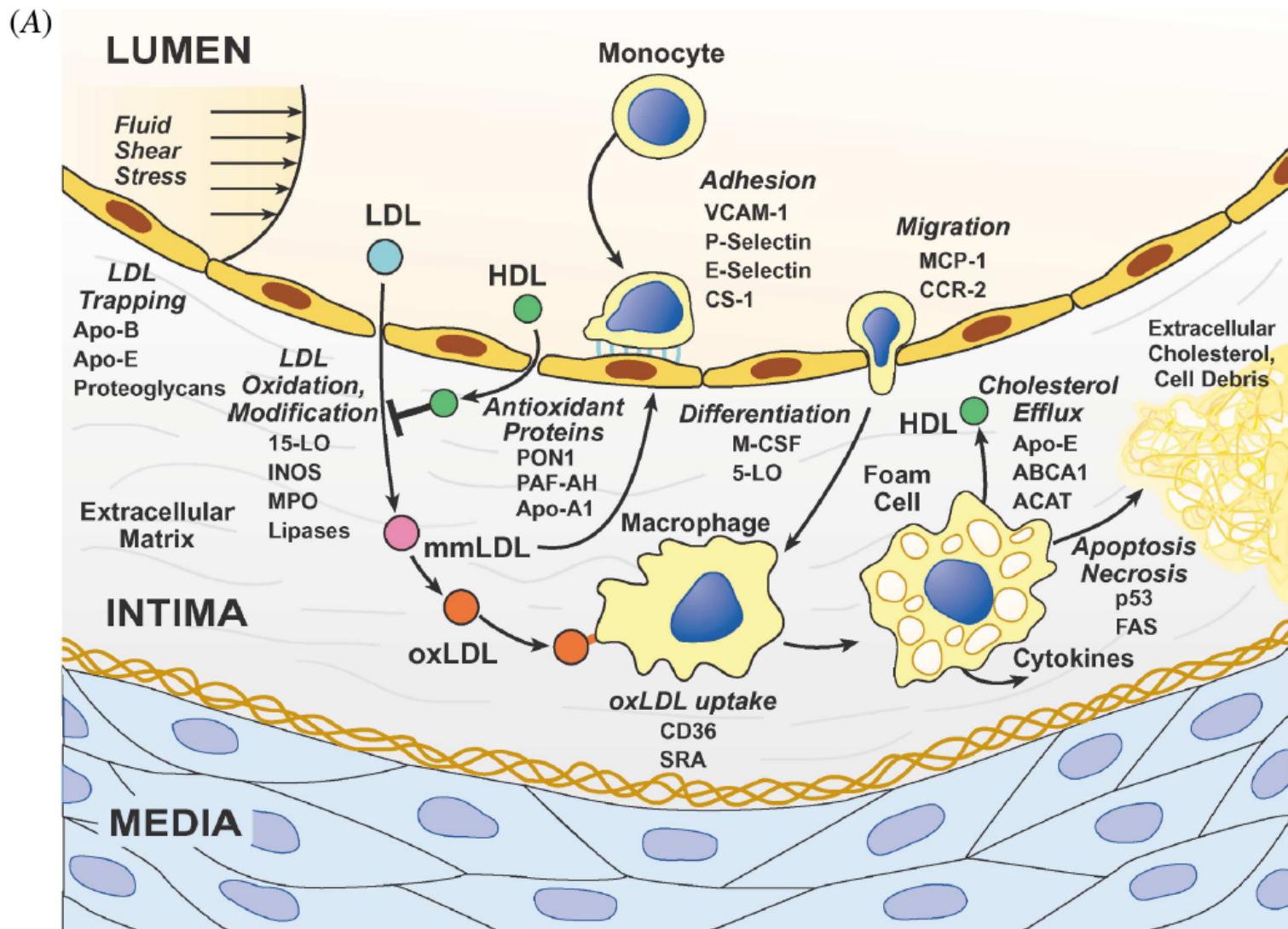
# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



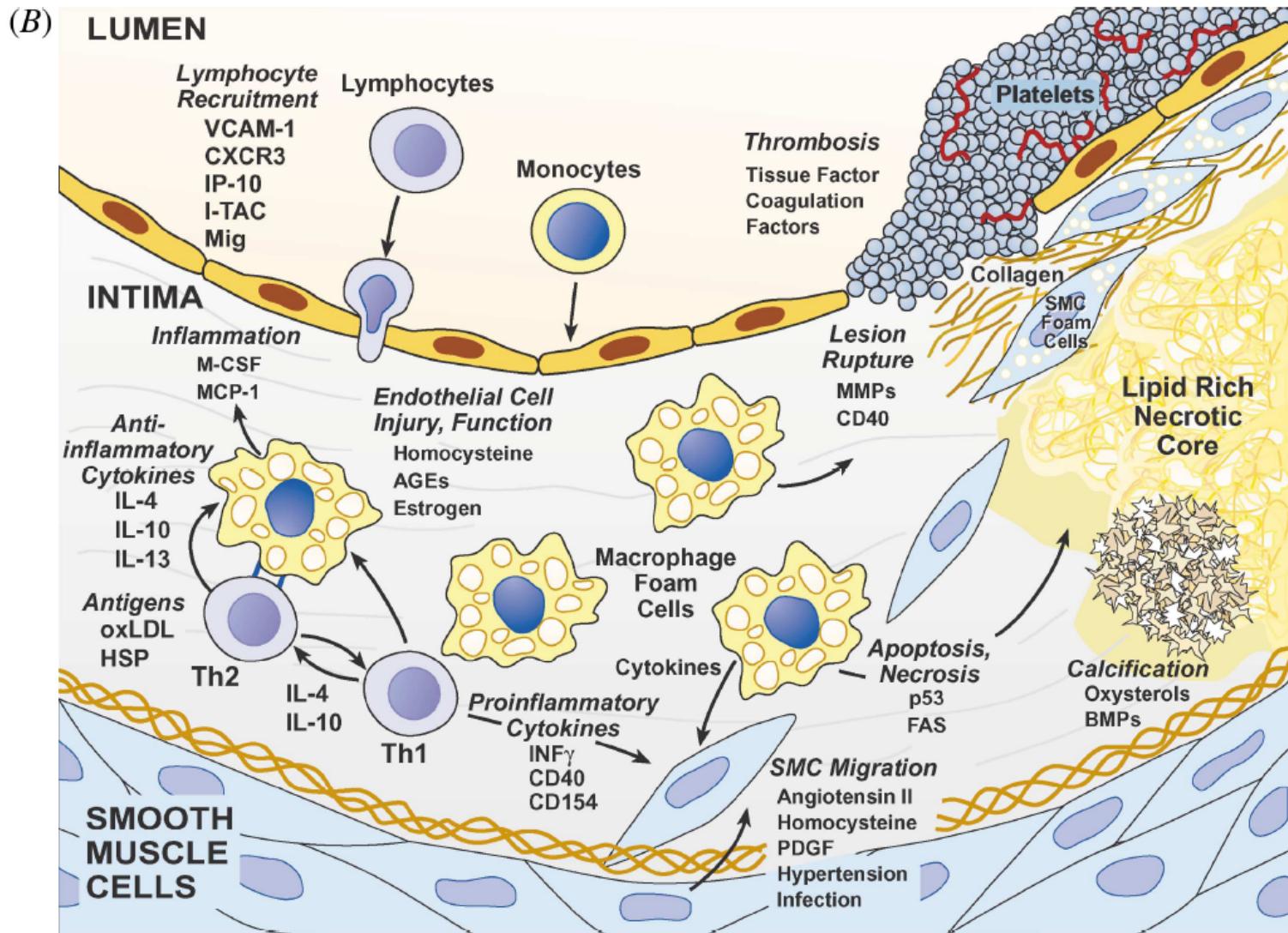
# Atherosclerosis & Cardiovascular Disease

*a very complex process and disease*



# Atherosclerosis & Cardiovascular Disease

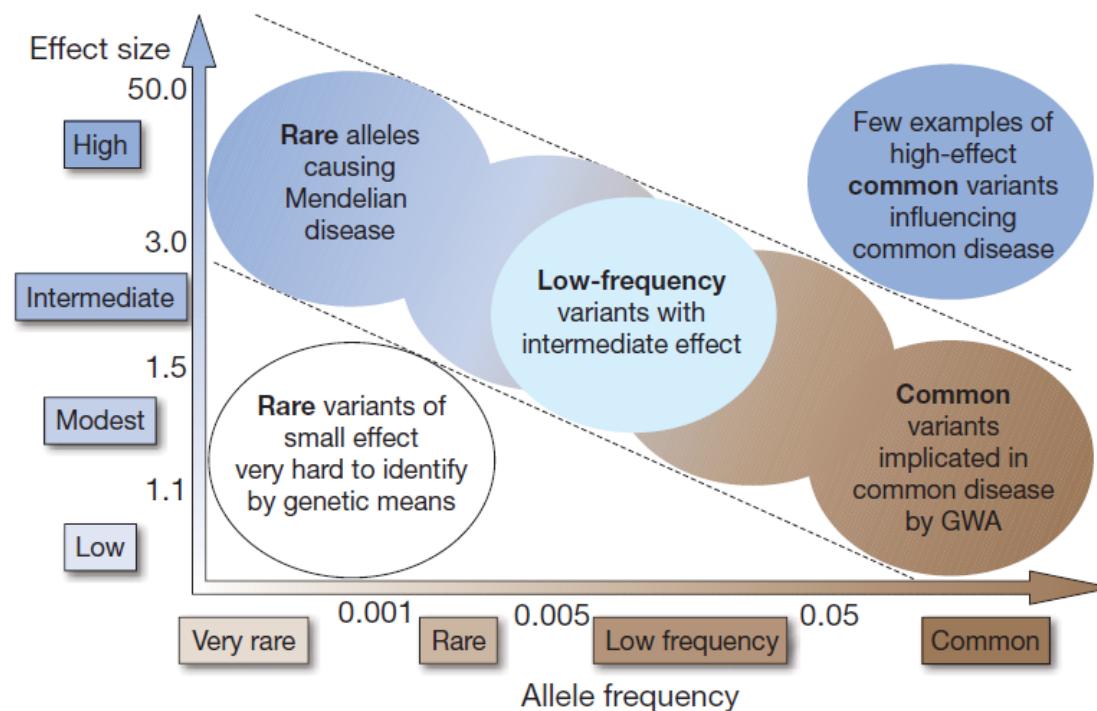
*a very complex process and disease*



# 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

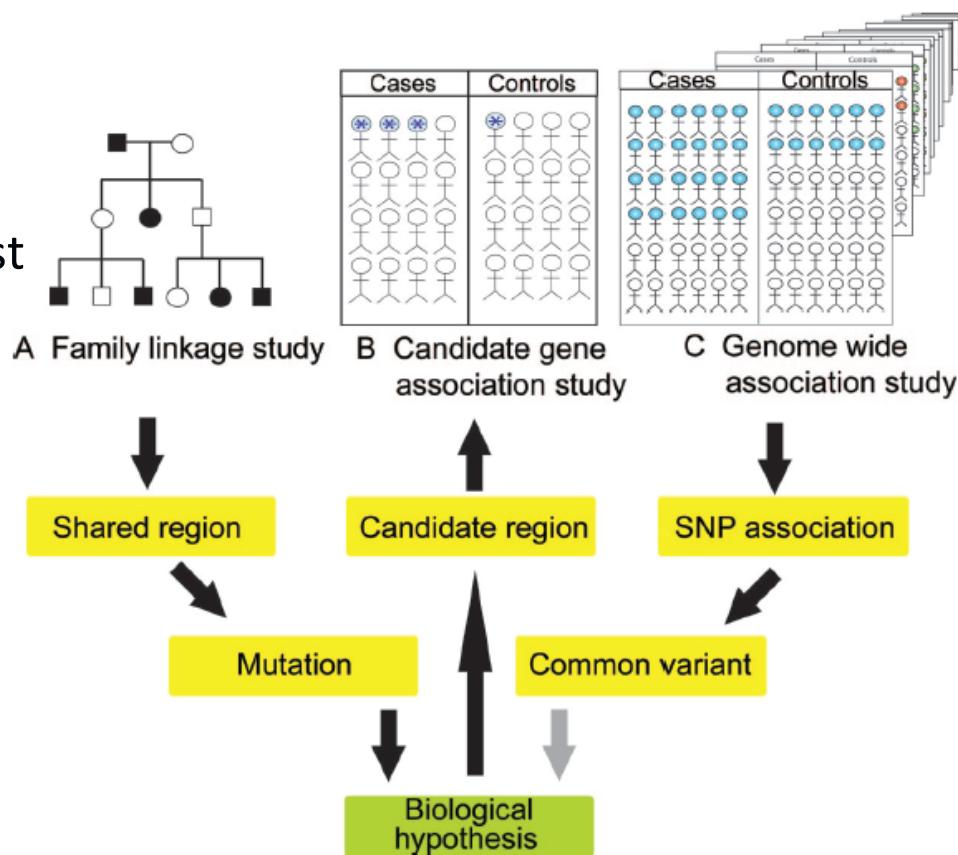
**Table 1.** Study Designs Used in Genome-wide Association Studies

	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



## Genetic and Genomic Insights into the Molecular Basis of Atherosclerosis

Yaoyu Chen,<sup>1</sup> Jarod Rollins,<sup>1</sup> Beverly Paigen,<sup>1</sup> and Xiaosong Wang<sup>2,\*</sup>

<sup>1</sup>The Jackson Laboratory, Bar Harbor, ME 04609, USA

<sup>2</sup>Novartis Institutes for BioMedical Research, Cambridge, MA 02139, USA

\*Correspondence: [xiaosong.wang@novartis.com](mailto:xiaosong.wang@novartis.com)

DOI 10.1016/j.cmet.2007.07.001

Atherosclerosis is a complex disease involving genetic and environmental risk factors, acting on their own or in synergy. Within the general population, polymorphisms within genes in lipid metabolism, inflammation, and thrombogenesis are probably responsible for the wide range of susceptibility to myocardial infarction, a fatal consequence of atherosclerosis. Genetic linkage studies have been carried out in both humans and mouse models to identify these polymorphisms. Approximately 40 quantitative trait loci for atherosclerotic disease have been found in humans, and approximately 30 in mice. Recently, genome-wide association studies have been used to identify atherosclerosis-susceptibility polymorphisms. Although discovering new atherosclerosis genes through these approaches remains challenging, the pace at which these polymorphisms are being found is accelerating due to rapidly improving bioinformatics resources and biotechnologies. The outcome of these efforts will not only unveil the molecular basis of atherosclerosis but also facilitate the discovery of drug targets and individualized medication against the disease.

- 19 studies in 9 populations identified 40 disease-regulating loci
  - 4x study Coronary Artery Disease (CAD): 9
  - 6x study Myocardial Infarction (MI): 12
  - 1x study CAD & MI: 1
  - 1x study Acute Coronary Syndrome (ACS): 3
  - 2x study Carotid Intimal-Medial Thickness (CIMT): 9
  - 2x study Coronary Artery Calcification (CAC): 4
  - 2x study Cerebral Vascular Accident (CVA=stroke): 1
  - 1x study Peripheral Artery Disease (PAD): 1

# 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 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>4</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

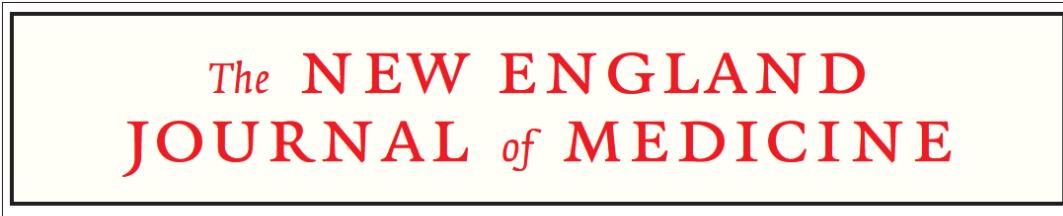
Ruth McPherson,<sup>1,\*†</sup> Alexander Pertsemlidis,<sup>2,\*</sup> Nihan Kavaslar,<sup>1</sup> Alexandre Stewart,<sup>1</sup> Robert Roberts,<sup>1</sup> David R. Cox,<sup>3</sup> David A. Hinds,<sup>3</sup> Len A. Pennacchio,<sup>4,5</sup> Anne Tybjaerg-Hansen,<sup>6</sup> Aaron R. Folsom,<sup>7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

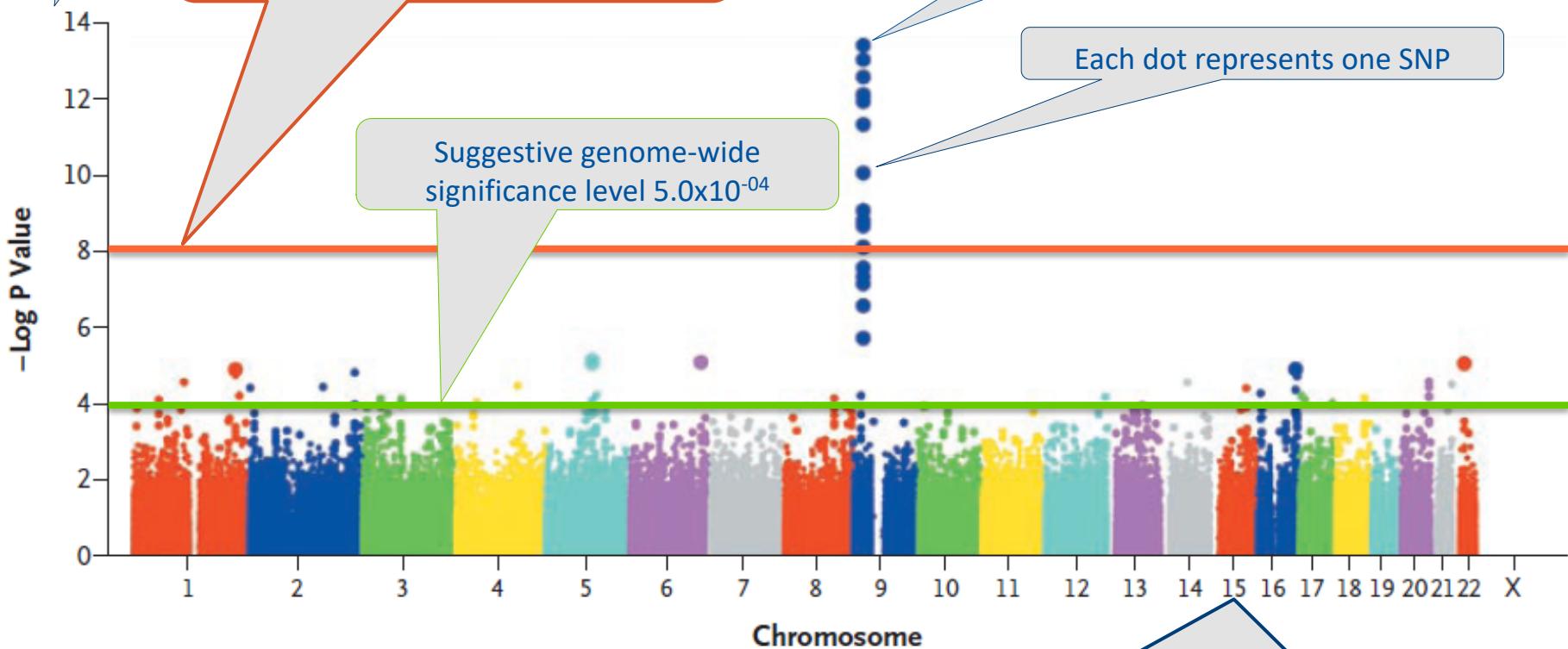
-Log10 of the associated p-value

Genome-wide significance level  
 $5.0 \times 10^{-8}$ , which is correcting for  
2x500,000 tests (Bonferroni)

Suggestive genome-wide  
significance level  $5.0 \times 10^{-4}$

A Manhattan skyscraper of significance  
at a region (9p21) associated with the  
trait (CAD), rs1333049

Each dot represents one SNP



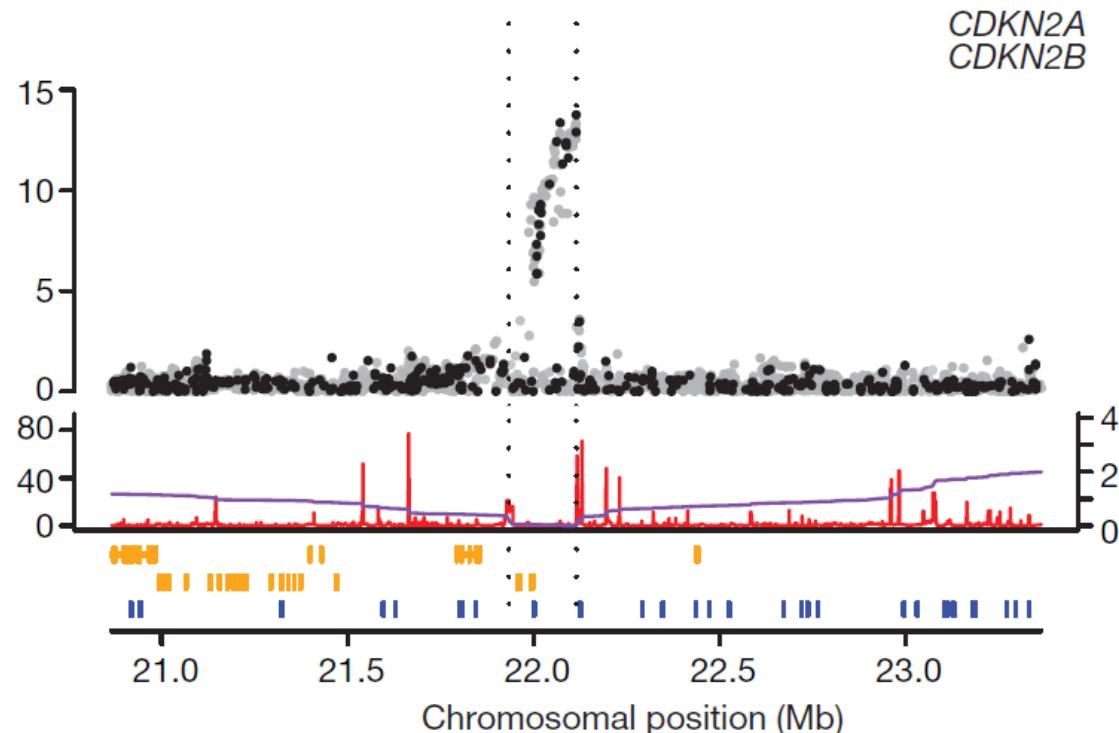
Each SNP is plotted per chromosome in genomic positional order

# One famous example

## 9p21 a closer look

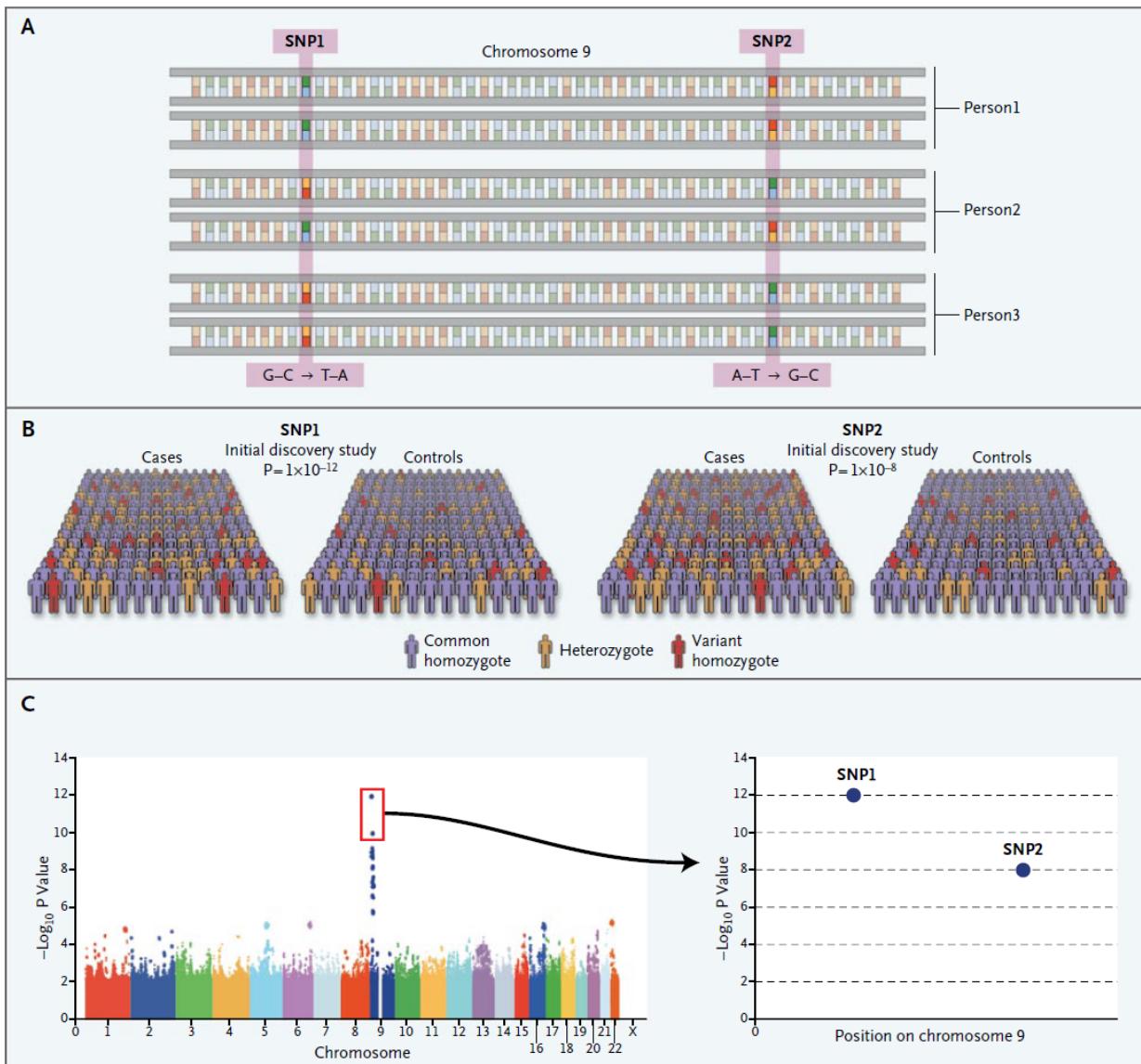
- The SNPs on 9p21.1 are rs1333049, rs10757274, rs2383207, rs2891168, and rs10757278
- They are found in an *intergenic region*
- Genes nearby: *CDKN2A*, *CDKN2B*
  - also associated with *type 2 diabetes mellitus*
  - regulating cell proliferation, cell aging and the associated degeneration, and programmed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

CAD hit region, chromosome 9



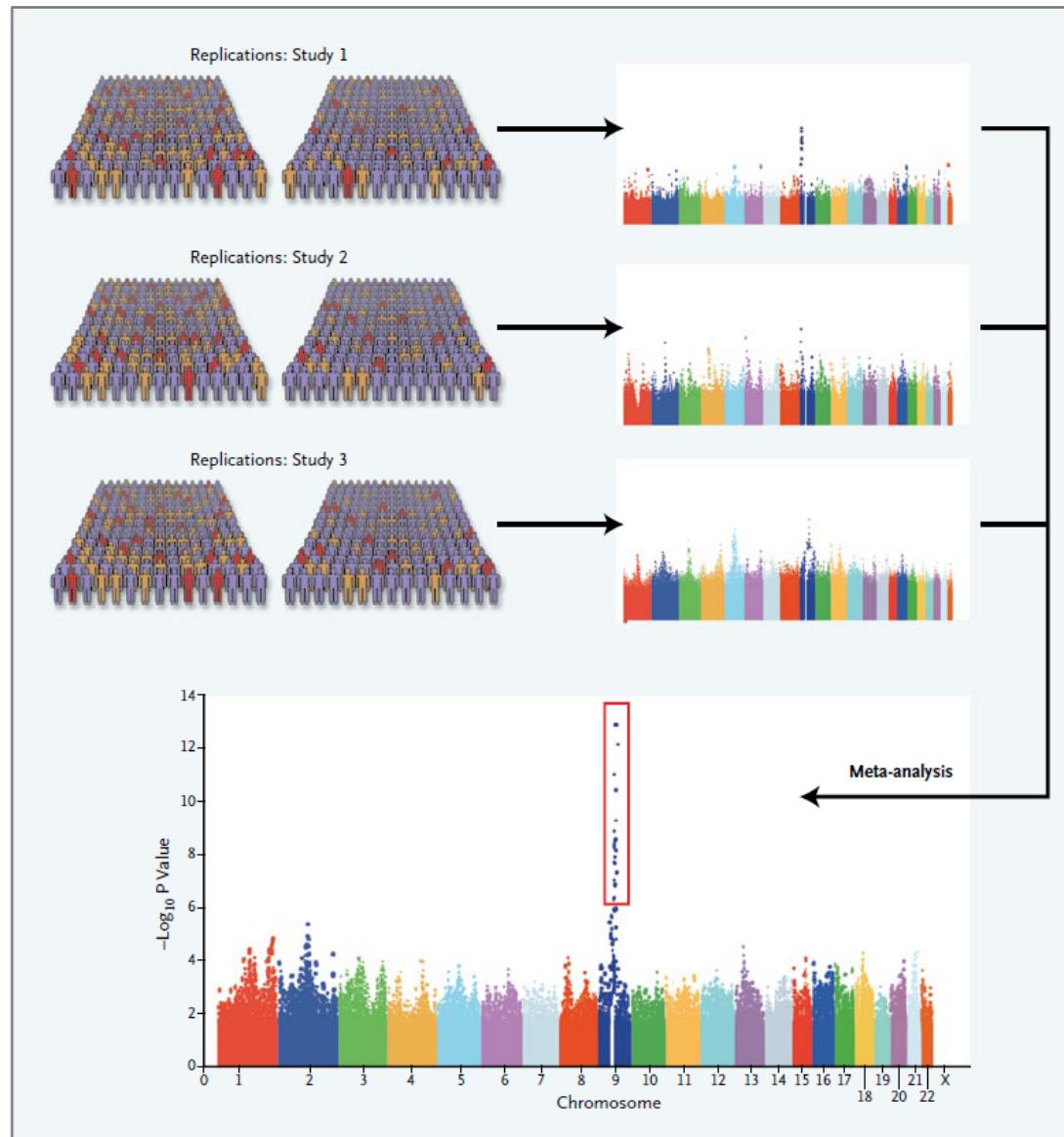
# Replication is the key because “one study is no study”

- First discover SNPs associated with the disease...



# Replication is the key *because “one study is no study”*

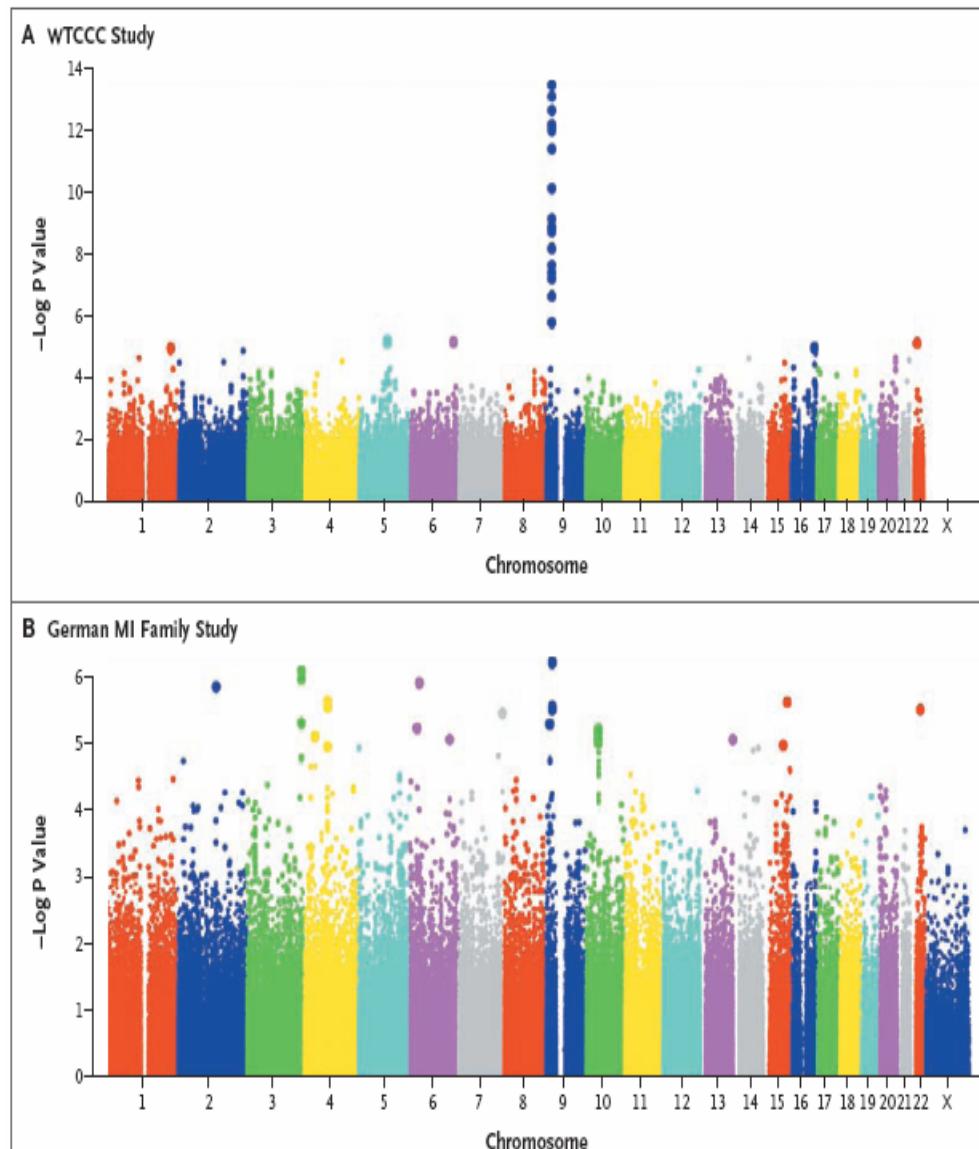
- Then try to replicate...
  - In many other populations
  - With the same *phenotype*
- Meta-analysis



# One famous example

## 9p21 in the Germans

- In GWAS: “*One study is no study*”
- Replication in *German MI Family Study*
  - 875 MI cases
    - <60 years of age
    - ≥1 family member affected (CAD)
    - 1644 controls



# 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
Standard analysis													
BD	16p12	23.3-23.62	rs420259	$2.19 \times 10^{-04}$	$6.29 \times 10^{-08}$	1.96	4.79	A	C	2.08 (1.60-2.71)	2.07 (1.6-2.69)	0.282	0.248
CAD	9p21	21.93-22.12	rs1333049	$1.79 \times 10^{-14}$	$1.16 \times 10^{-13}$	11.66	11.19	C	C	1.47 (1.27-1.70)	1.9 (1.61-2.24)	0.474	0.554
CD	1p31	87.3-87.48	rs11003303	$8.45 \times 10^{-13}$	$5.83 \times 10^{-12}$	10.87	7.41	-	-	1.59 (1.22-1.88)	1.88 (1.54-2.24)	0.337	0.331
CD	2q37	233.92-234	rs10210302	$7.10 \times 10^{-14}$	$5.26 \times 10^{-14}$	11.11	11.28	T	C	1.19 (1.01-1.41)	1.85 (1.56-2.21)	0.481	0.402
CD	3p21	49.3-49.87	rs9858542	$7.71 \times 10^{-07}$	$3.58 \times 10^{-08}$	4.24	5.22	A	A	1.09 (0.96-1.24)	1.84 (1.49-2.26)	0.282	0.331
CD	5p13	40.32-40.66	rs17234657	$2.13 \times 10^{-13}$	$1.99 \times 10^{-12}$	10.41	9.89	G	G	1.54 (1.34-1.76)	2.32 (1.59-3.39)	0.125	0.181
CD	5q33	150.15-150.31	rs1000113	$5.10 \times 10^{-08}$	$3.15 \times 10^{-07}$	5.36	5.01	T	T	1.54 (1.31-1.82)	1.92 (0.92-4.00)	0.067	0.098
CD	10q21	64.06-64.31	rs10761659	$2.68 \times 10^{-07}$	$1.75 \times 10^{-06}$	4.69	4.13	G	A	1.23 (1.05-1.45)	1.55 (1.3-1.84)	0.461	0.406
CD	10q24	101.26-101.32	rs10883365	$1.41 \times 10^{-08}$	$5.82 \times 10^{-08}$	5.91	5.48	G	G	1.2 (1.03-1.39)	1.62 (1.37-1.92)	0.477	0.537
CD	16q12	49.02-49.4	rs17221417	$9.36 \times 10^{-12}$	$3.98 \times 10^{-11}$	8.93	8.47	G	G	1.29 (1.13-1.46)	1.92 (1.58-2.34)	0.287	0.356
CD	18p11	12.76-12.91	rs2542151	$4.56 \times 10^{-08}$	$2.03 \times 10^{-07}$	5.42	5.00	G	G	1.3 (1.14-1.48)	2.01 (1.46-2.76)	0.163	0.208
RA	1p13	113.54-114.16	rs6679677	$4.90 \times 10^{-26}$	$5.55 \times 10^{-25}$	22.36	21.99	A	A	1.98 (1.72-2.27)	3.32 (1.93-5.69)	0.096	0.168
RA	6	MHC	rs6457617*	$3.44 \times 10^{-76}$	$5.18 \times 10^{-75}$	74.84	73.18	T	T	2.36 (1.97-2.84)	5.21 (4.31-6.30)	0.489	0.685
T1D	1p13	113.54-114.16	rs6679677	$1.17 \times 10^{-26}$	$5.43 \times 10^{-26}$	23.07	22.83	A	A	1.82 (1.59-2.09)	5.19 (3.15-8.55)	0.096	0.169
T1D	6	MHC	rs9272346*	$2.42 \times 10^{-134}$	$5.47 \times 10^{-134}$	141.9	142.2	A	G	5.49 (4.83-6.24)	18.52 (27.03-12.69)	0.387	0.150
T1D	12q13	54.64-55.09	rs11171739	$1.14 \times 10^{-11}$	$9.71 \times 10^{-11}$	8.89	8.24	C	C	1.34 (1.17-1.54)	1.75 (1.48-2.06)	0.423	0.493
T1D	12q24	109.82-111.49	rs17696736	$2.17 \times 10^{-15}$	$1.51 \times 10^{-14}$	12.53	11.88	G	G	1.34 (1.16-1.53)	1.94 (1.65-2.29)	0.424	0.506
T1D	16p13	10.93-11.37	rs12708716	$9.24 \times 10^{-08}$	$4.92 \times 10^{-07}$	5.15	4.70	A	G	1.19 (0.97-1.45)	1.55 (1.27-1.89)	0.350	0.297
T2D	6p22	20.63-20.84	rs9465871	$1.02 \times 10^{-06}$	$3.34 \times 10^{-07}$	4.15	3.98	C	C	1.18 (1.04-1.34)	2.17 (1.6-2.95)	0.178	0.218
T2D	10q25	114.71-114.81	rs4506565	$5.68 \times 10^{-13}$	$5.05 \times 10^{-12}$	10.14	9.43	T	T	1.36 (1.2-1.54)	1.88 (1.56-2.27)	0.324	0.395
T2D	16q12	52.36-52.41	rs9939609	$5.24 \times 10^{-08}$	$1.91 \times 10^{-07}$	5.35	5.05	A	A	1.34 (1.17-1.52)	1.55 (1.3-1.84)	0.398	0.453
Multi-locus analysis													
T1D	4q27	123.26-123.92	rs6534347	$4.48 \times 10^{-07}$	$1.83 \times 10^{-06}$	5.15	4.69	A	A	1.30 (1.10-1.55)	1.49 (1.25-1.78)	0.351	0.402
T1D	12p13	9.71-9.86	rs3764021	$7.19 \times 10^{-05}$	$5.08 \times 10^{-08}$	2.12	4.55	C	T	1.57 (1.38-1.79)	1.48 (1.25-1.75)	0.467	0.426
Sex differentiated analysis													
RA	7q32	130.80-130.84	rs11761231	$3.91 \times 10^{-07}$	$1.37 \times 10^{-06}$	-	-	G	A	1.44 (1.19-1.75)	1.64 (1.35-1.99)	0.375	0.327
Combined cases													
RA+T1D	10p15	6.07-6.17	rs2104286	$5.92 \times 10^{-08}$	$2.52 \times 10^{-07}$	5.26	4.45	T	C	1.35 (1.11-1.65)	1.62 (1.34-1.97)	0.286	0.245

Regions with at least one SNP with a P value of less than  $5 \times 10^{-7}$  for our primary analyses. The  $\log_{10}$  value of the Bayes factor (BF) for the bayesian analysis corresponding to the trend and genotypic tests is also given. Region marks the boundaries of signal defined by recombination and return of test statistics to background levels. The minor allele is defined in the controls and its frequency in that group as well as the case sample is reported. MAF, minor allele frequency. Cluster plots for each SNP have been inspected visually, and are shown in Supplementary Fig. 10. Positions are in NCBI build-35 coordinates \*Multiple SNPs in the MHC region are significant, we report the most extreme.

**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

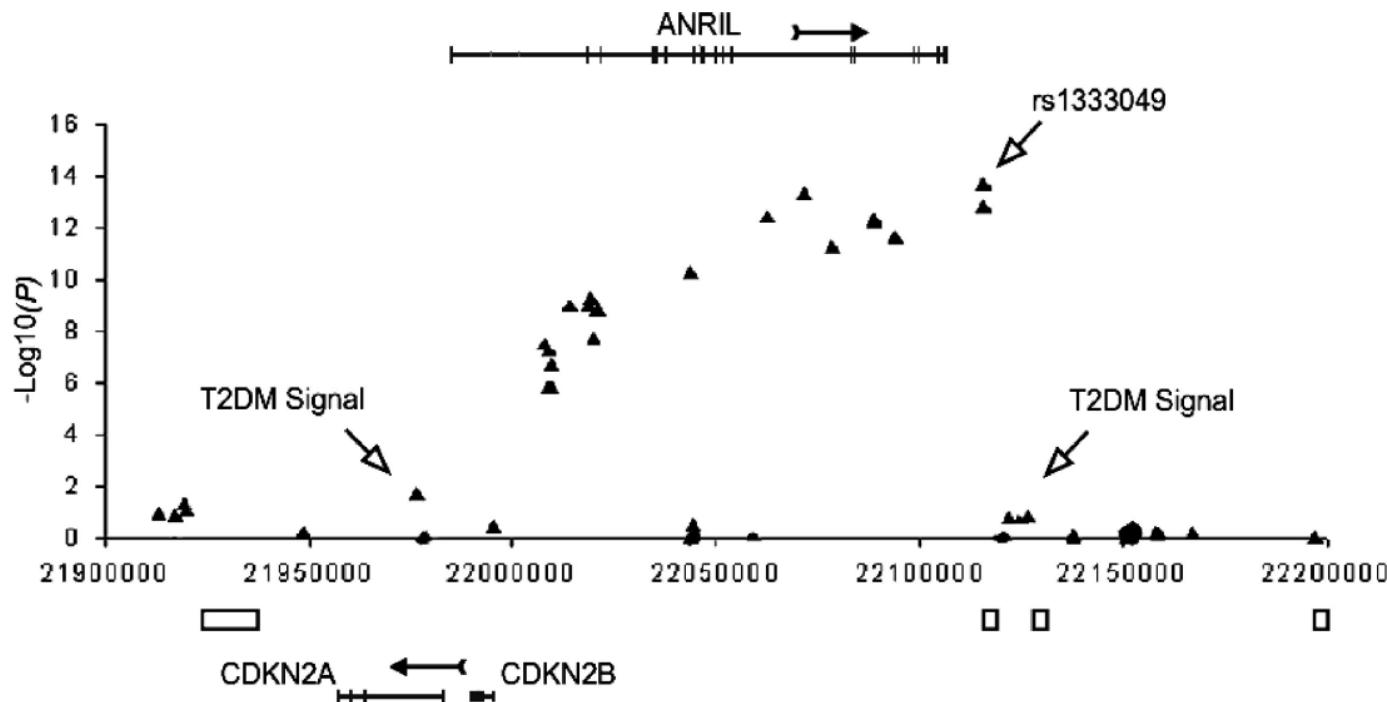
**Risk allele:** minor allele  
**Odds ratio:** the odds of exposure between cases and controls

**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		
(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	

Key	Adding up the points
Color	Risk
Green	Very low
White	Low
Yellow	Moderate
Rose	High
Red	Very high

Point total:

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	= 25.9 %
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 Vista Customer Service: 15700 W. 103rd St. Suite 200, Lemoore, CA 93643 – Phone: (833) 222-8510 Fax: (833) 783-0998 – www.decodediagnostics.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 = 25.9 %
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

# GWAS in March 2010

$\approx 545$  studies,  $\approx 800$  SNPs &  $p \leq 5 \times 10^{-8}$



University Medical Center  
Utrecht



# Interested in Cardiovascular Research?

University Medical Center  
Utrecht

## Experimental Cardiology Laboratory

Prof. Dr. G. Pasterkamp

Prof. Dr. D.P.V. de Kleijn

Dr. J.P.G. Sluijter

Dr. I. Höfer

Dr. J.K. van Keulen

Dr. M. Smeets

## 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



[www.umcutrecht.nl/experimentalcardiology](http://www.umcutrecht.nl/experimentalcardiology)

&

[www.atheroexpress.nl](http://www.atheroexpress.nl)



University Medical Center  
Utrecht



# Further reading

- Pearson, T.A. et al. *How to interpret a GWAS* - **JAMA** 2008 (PMID: 18349094)
  - Excellent to grasp the concepts of GWAS'
- McCarthy, M.I. et al. *GWAS for complex traits, consensus, uncertainty and challenges* - **Nature Rev Genet** 2008 (PMID: 18398418)
  - Another excellent start to grasp the concepts of GWAS'
- WTCCC - *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls* - **Nature** 2007 (PMID: 17554300)
  - Pivotal article on a GWAS for CAD, only the CAD-section (including pictures/tables) is interesting)
- **Robbins and Cotran – Pathologic Basis of Disease, 7<sup>th</sup> Edition**
  - Page 149- 152: “Mendelian Disorders” on autosomal dominant, autosomal recessive and X-linked disorders
  - Page 154: “Marfan Syndrome”
  - page 156-158: “Familial Hypercholesterolemia”