





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



【 EVERY \$1 INVESTED IN THE HUMAN GENOME PROJECT 】
HAS TRIGGERED **\$178** IN
U.S. ECONOMIC ACTIVITY

THE **\$14.5 BILLION** THE U.S. GOVERNMENT INVESTED IN THE
HUMAN GENOME EFFORT SINCE 1988 HAS HELPED DRIVE:

\$965 BILLION IN
ECONOMIC IMPACT

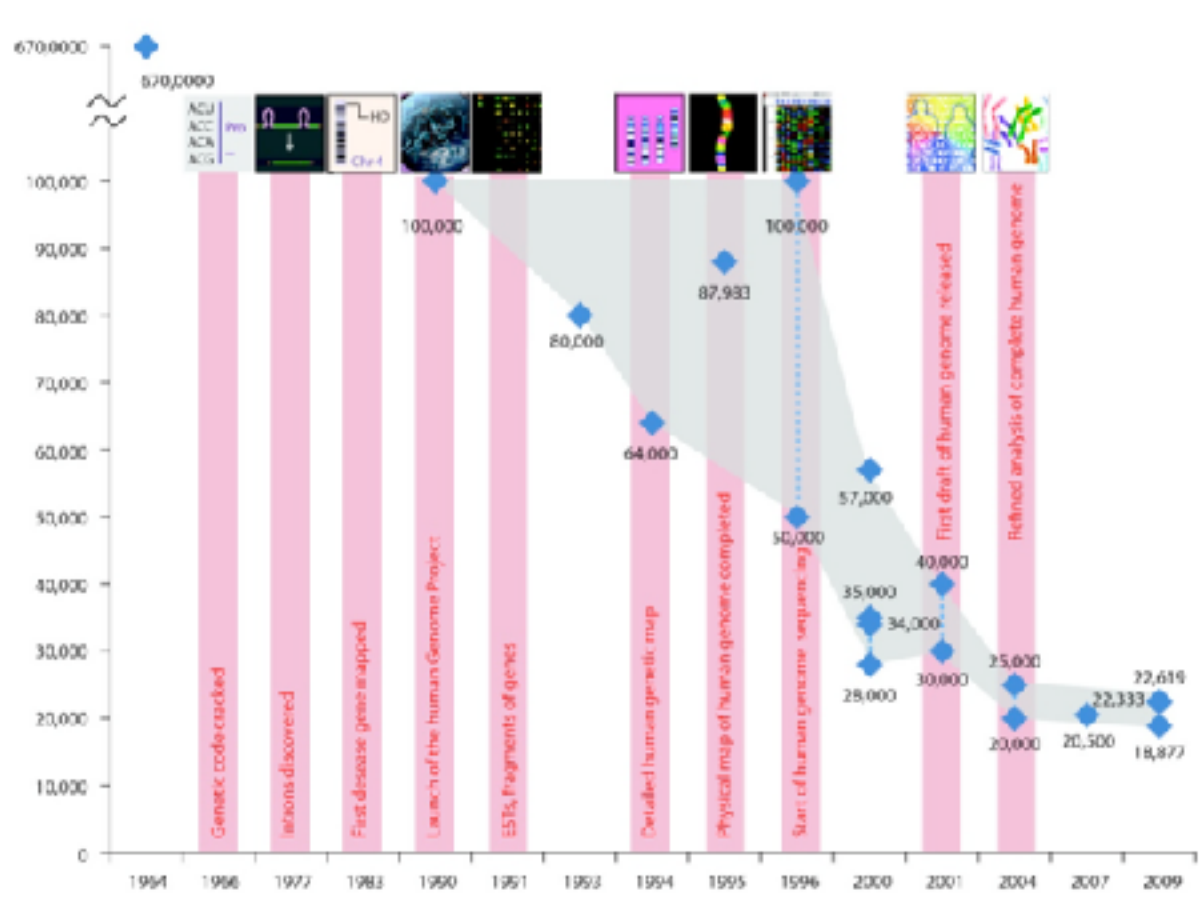
\$293 BILLION
IN TOTAL PERSONAL INCOME

\$169 BILLION
INCREASE IN ECONOMIC
OUTPUT SINCE 2010

In 2012 alone, genomics-related research,
development and commercialization activities generated:

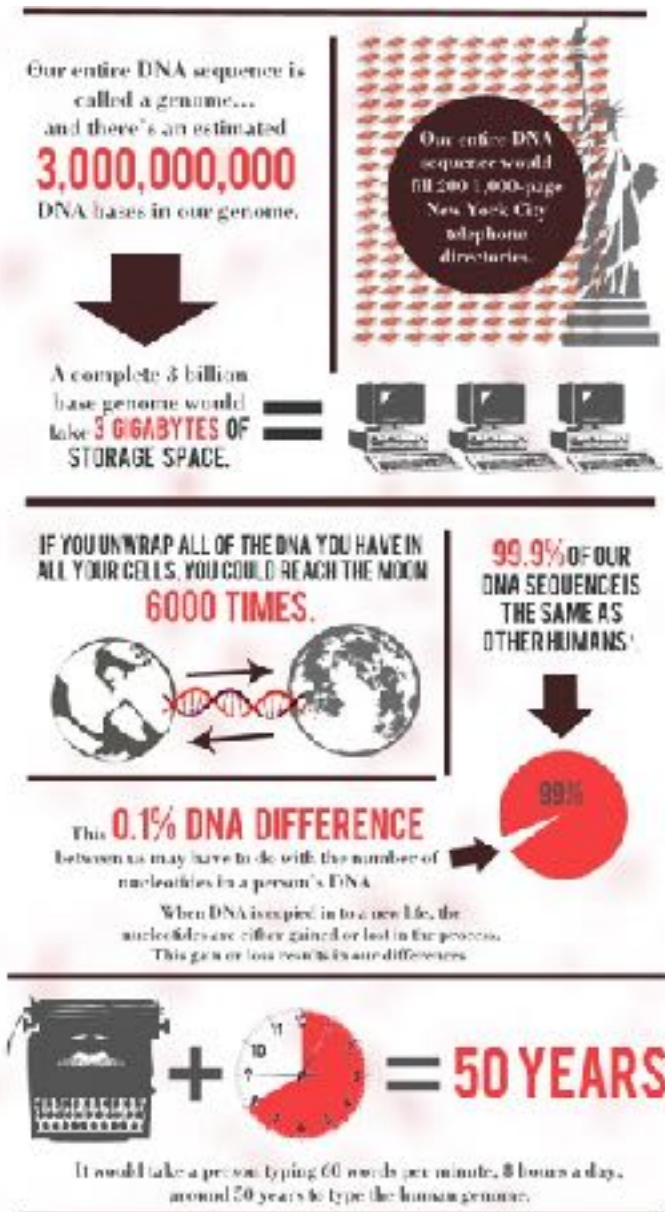
\$65B **152,314** **\$19B**
IN U.S. ECONOMY SUPPORTED JOBS IN PERSONAL INCOME

(Finally) a complete map



Human Genome: *some statistics*

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



ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCAT
CGATCTTGC
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCC
TTTACTTGC
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCC
CAGCAAAG
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTAC
TGCCCACGT
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGACTGCACATATCGT
CATAGCTCA
TCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCC
ACTTTGGCA
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTATC
CTATATTGCA
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTG
ACTGCGGTC
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCC
ACATATATGC
CGTACTGACTGTCTAGACTAAACACATCCCACCTTTACCCATGCATCGTACTGACTGTCTAGTCTAAACACATC
CCACATCGC
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATCCATATCGTCATCGTACTGACTGTCTAGTCTAA
ACACATTG
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTG
ACTGCGCCT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCC
ACATATAATG
CGTACTGACTGTCTAGTCTAAACACATCCCACCTTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAA
CACATTTCG

ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCAT
CGATCTTGC
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCC
TTTACTTGC
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCC
CAGCAAAG
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTAC
TGCCACGT
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGACTGCACATATCGT
CATAGCTCA
TCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCC
ACTTTGGCA
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTATC
CTATATTGCA
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTG
ACTGCGGTC
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCC
ACATATATGC
CGTACTGACTGTCTAGGCTAAACACATCCCACCTTTACCCATGCATCGTACTGACTGTCTAGTCTAAACACATC
CCACATCGC
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATCCATATCGTCATCGTACTGACTGTCTAGTCTAA
ACACATTG
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTG
ACTGCGCCT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCC
ACATATAATG
CGTACTGACTGTCTAGTCTAAACACATCCCACCTTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAA
CACATTTCG

ATGCCGATCGTACGACACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCATCGTACTGACTGCAT
CGATCTTGC
TACTGACTGCATCGTACTGACTGCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCC
TTTACTTGC
CATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCC
CAGCAAAG
CATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCTATGCCGATCGTACGACACATATCGTCATCGTAC
TGCCCACGT
ACTGTCTAGTCTAAACACATCCATCGTACTGACTGCATCGTACTGACTGCATCGTACTGACTGCACATATCGT
CATAGCTCA
TCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCC
ACTTTGGCA
ATATCGTCATCGTACTGACTGTCTAGTCTAAACACATCCCACATATCGTCATCGTACTGACTGTCTAGTCTATC
CTATTGGA
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTG
ACTG
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCC
ACATATATGC
CGTACTGACTGTCTAGGCTAAACACATCCCACCTTACCCATGCATCGTACTGACTGTCTAGTCTAAACACATC
CCACATCGC
ATCGTACTGACTGTCTAGTCTAAACACATCCCAGCATCCATCCATATCGTCATCGTACTGACTGTCTAGTCTAA
ACACATTTCG
GCCGATCGTACGACACATATCGTCATCGTACTGCCCTACGGGACTGTCTAGTCTAAACACATCCATCGTACTG
ACTGCGCCT
TGACTGCATCGTACTGACTGCACATATCGTCATACATAGACTTCGTACTGACTGTCTAGTCTAAACACATCCC
ACATATAATG
CGTACTGACTGTCTAGTCTAAACACATCCCACCTTACCCATGATATCGTCATCGTACTGACTGTCTAGTCTAA
CACATTTCG



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



Human genome: *individual variations*

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

articles

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

The International SNP Map Working Group*

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

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

The International HapMap Project

Phase I

1.1 million SNPs

270 individuals from 4 populations



Phase II

3.1 million SNPs

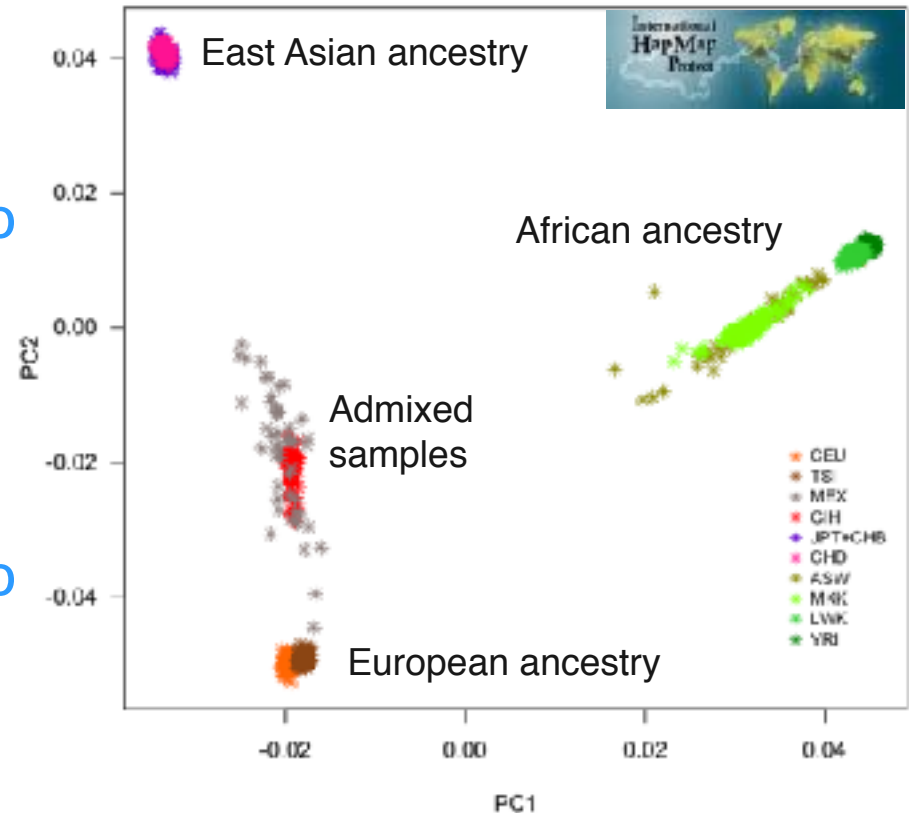
270 individuals from 4 populations



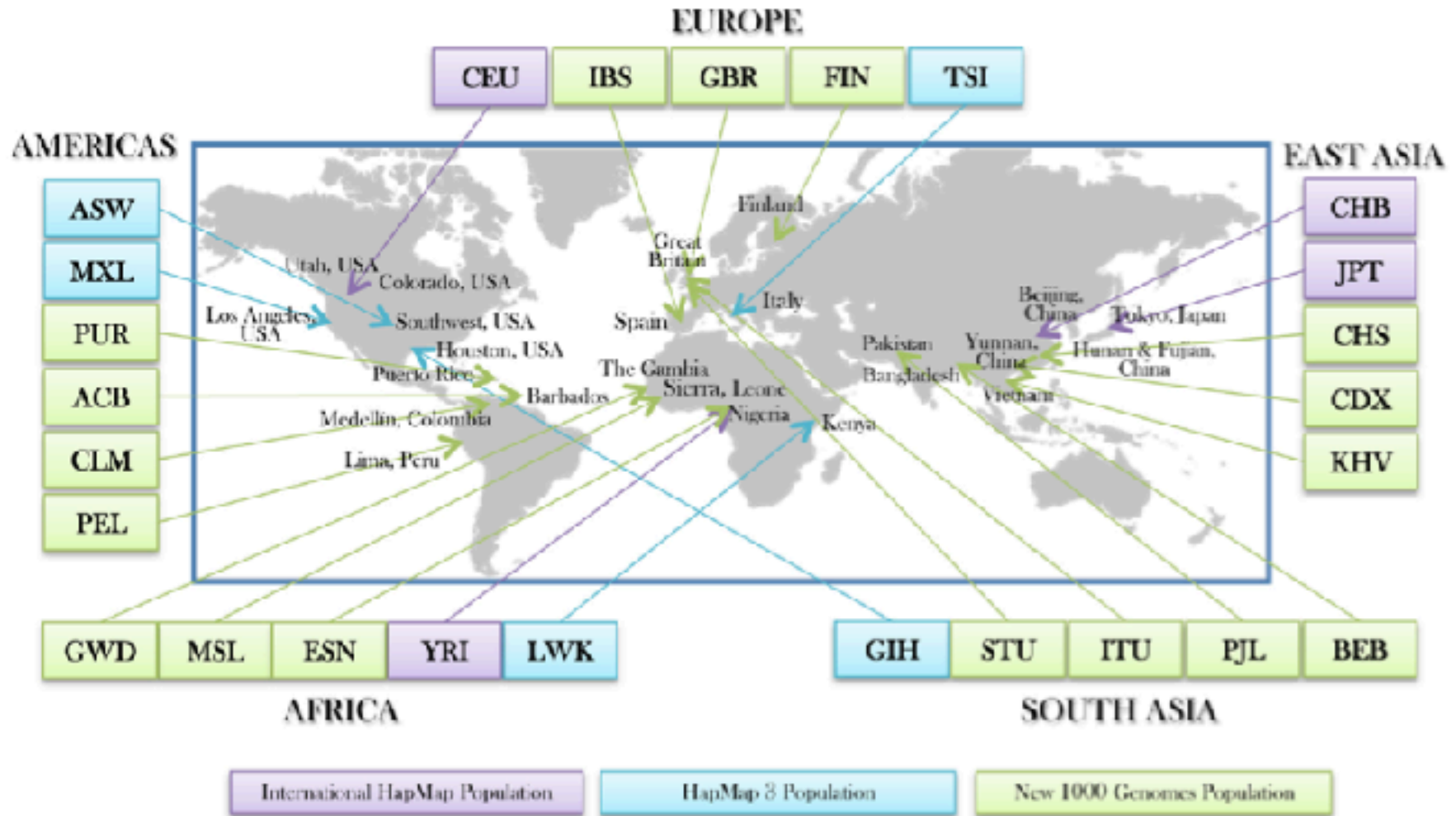
Phase III

1.6 million SNPs

1,184 individuals from 11 populations



The 1000 Genomes Project

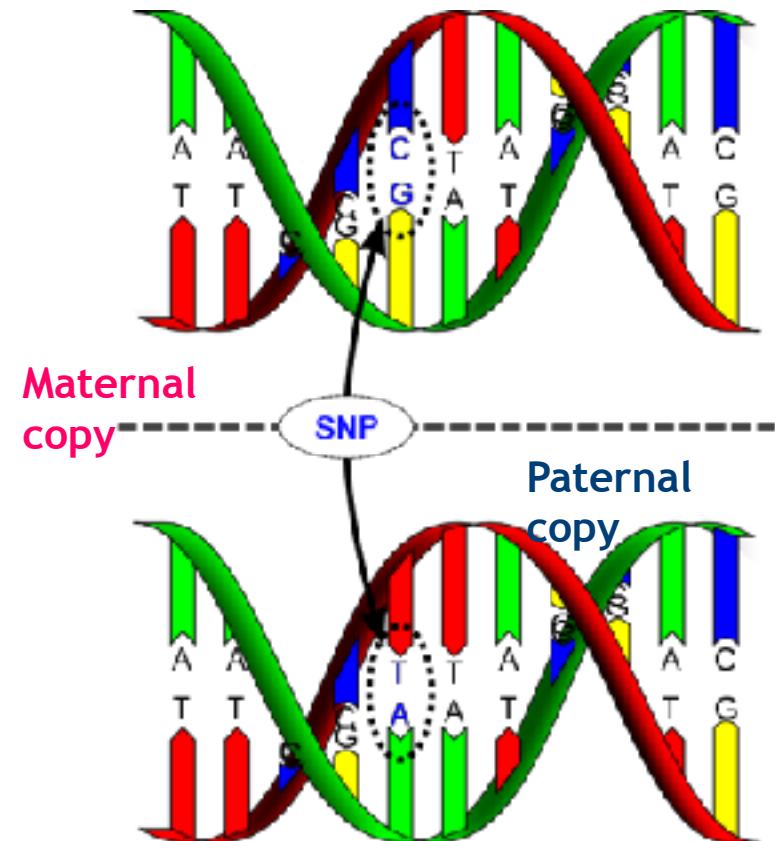


Single-Nucleotide Polymorphism

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



www.hapmap.org

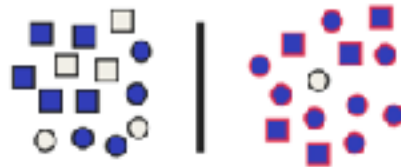
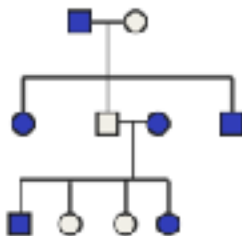




Linkage
analysis

Candidate
gene
studies

GWAS



Common variant, common disease hypothesis

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

The beginnings of GWAS

HapMap Phase I

HapMap Phase II

SNP arrays

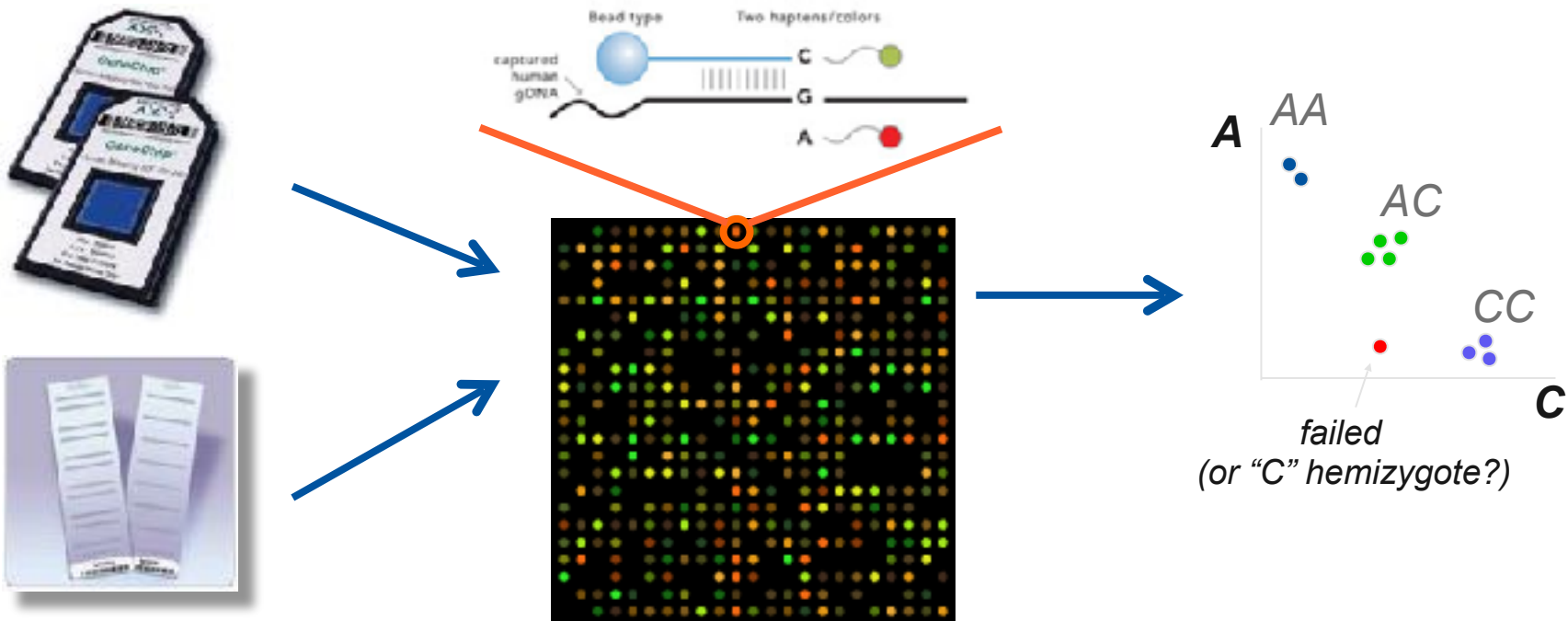
WTCCC GWAS

HapMap Phase III

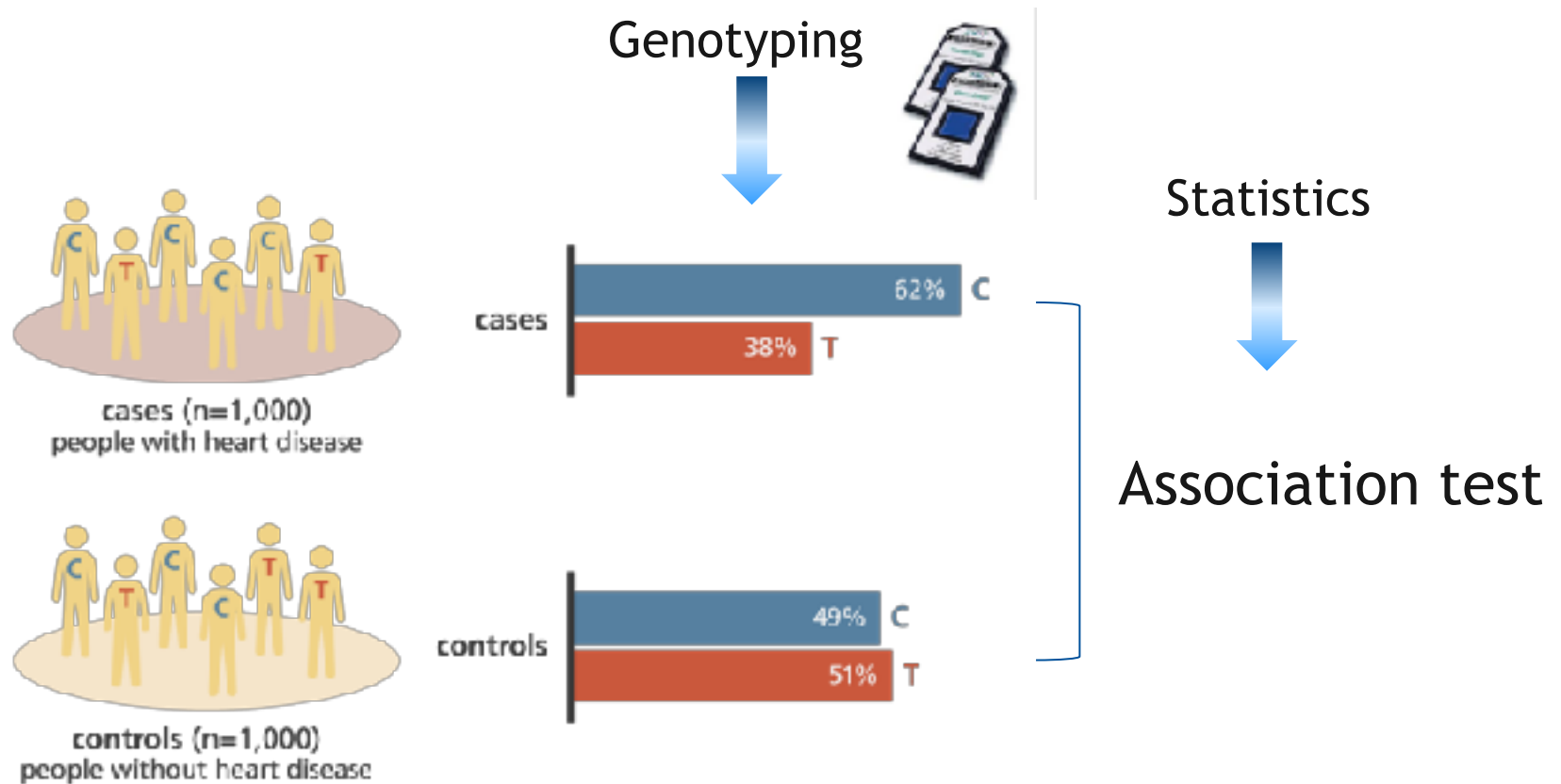


Genotyping platforms

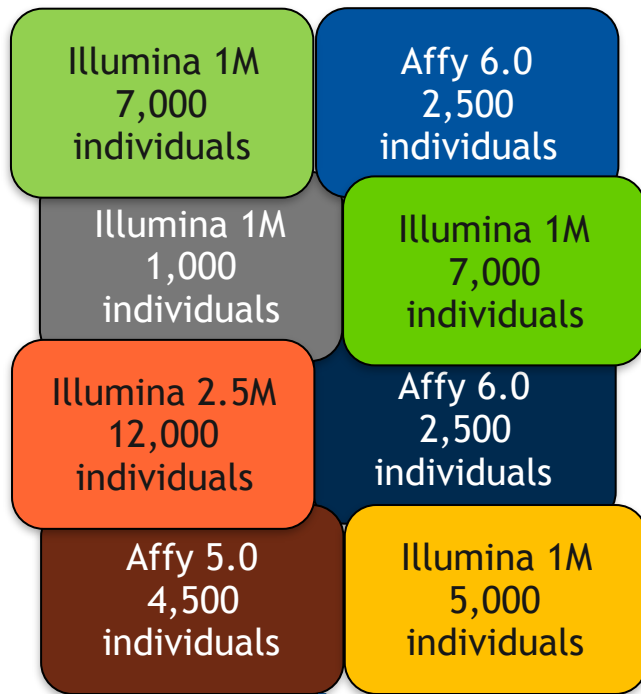
- Genome-wide SNP microarrays allow measurement of genotypes of 100,000's of SNPs in a single experiment
- Variety of microarrays (different SNP density, cost, etc) by Illumina and Affymetrix



GWAS (the big picture)



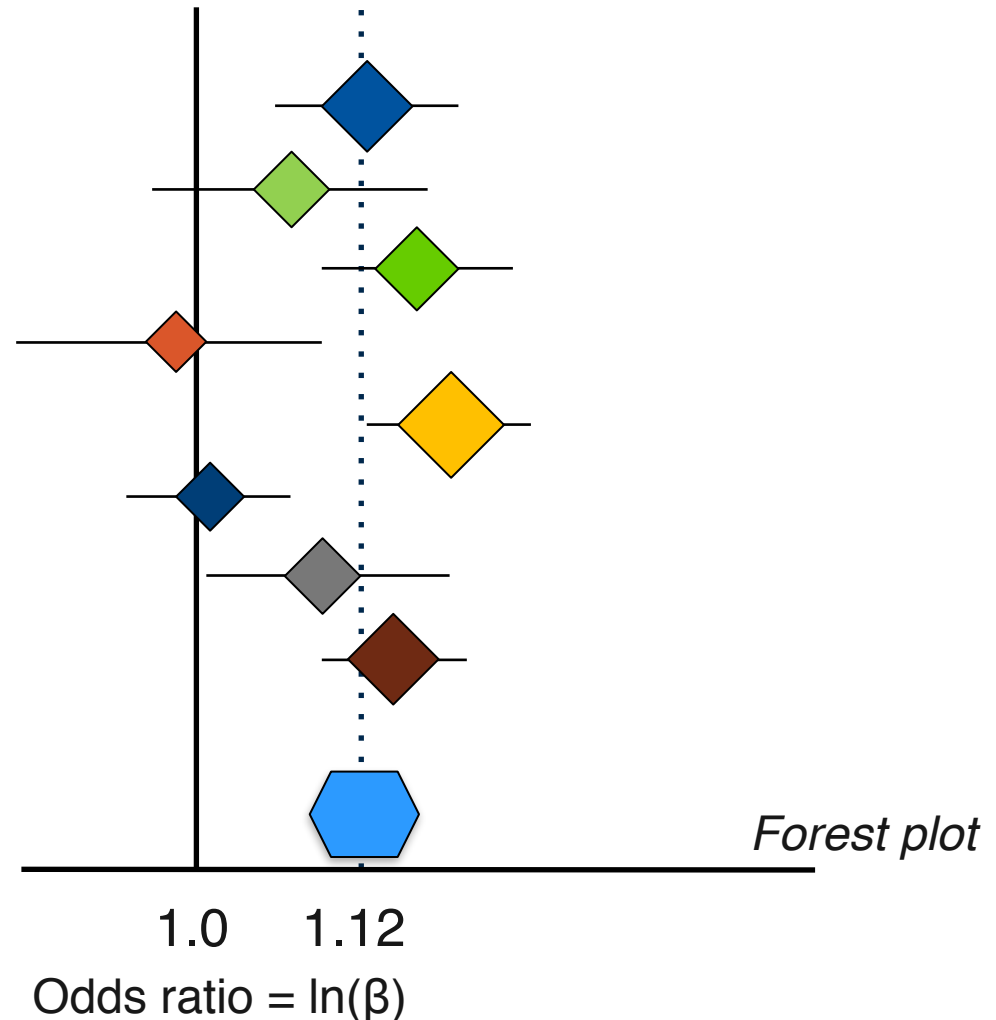
Combining GWAS datasets



Imputation

Meta-analysis of GWAS

Results for one SNP



deCODE Genetics, Inc.

- >50% adult population of Iceland (>140,000) in biobank (blood)
- Pedigree information going back to the first settlements (~1000 years ago)
- Extensive medical records & genotypic data
- Over 250 high-impact publications (Nature, Science,
- 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 Helgadóttir,¹ Andrei Mianaleva,¹ Gudmar Thorleifsson,¹ Solveig Gretarsdóttir,¹ Hilga Jonsdóttir,¹ Unnur Thordisdothir,¹ Nishil J. Sarnaik,¹ Gudmundur Gudmundsson,¹ Struan F. A. Grant,¹ Gudmundur Thorgeirsson,¹ Sigrúnur Jóhannsson,¹ Kinnar M. Valdimarsson,¹ Stefan B. Mathiasen,¹ Halldor Johannsson,¹ Olafur Gudmundsson,¹ Mark R. Gensler,¹ Jónas Sævi,¹ Margrét Thelma Jónsdóttir,¹ Margrét Andriadsdóttir,¹ Michael L. Frigge,¹ Elin T. Topol,¹ Augustine Kong,¹ Vilhjálmur Gudnason,¹ Hakon Hakonarson,¹ Jeffrey R. Goldner,¹ & Kari Stefansson¹

We mapped a gene predisposing to myocardial infarction to a locus on chromosome 14q32.1. A four-marker single nucleotide polymorphism (SNP) haplotype in this locus spanning the gene *ALOX5AP* encoding 5-lipoxygenase activating protein (LAP) is associated with a two-fold increase in risk of myocardial infarction in Iceland. This haplotype also confers a higher risk of stroke. Another *ALOX5AP* haplotype was linked with myocardial infarction in individuals from the UK. Simulated recombination from individuals with myocardial infarction produced more variation in the *ALOX5AP* haplotype than in individuals from controls, and this difference is especially evident in cells from individuals with the stroke haplotype. We conclude that variants of *ALOX5AP* are involved in the pathogenesis of both myocardial infarction and stroke by increasing leukocyte production and inflammation in the arterial wall.

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

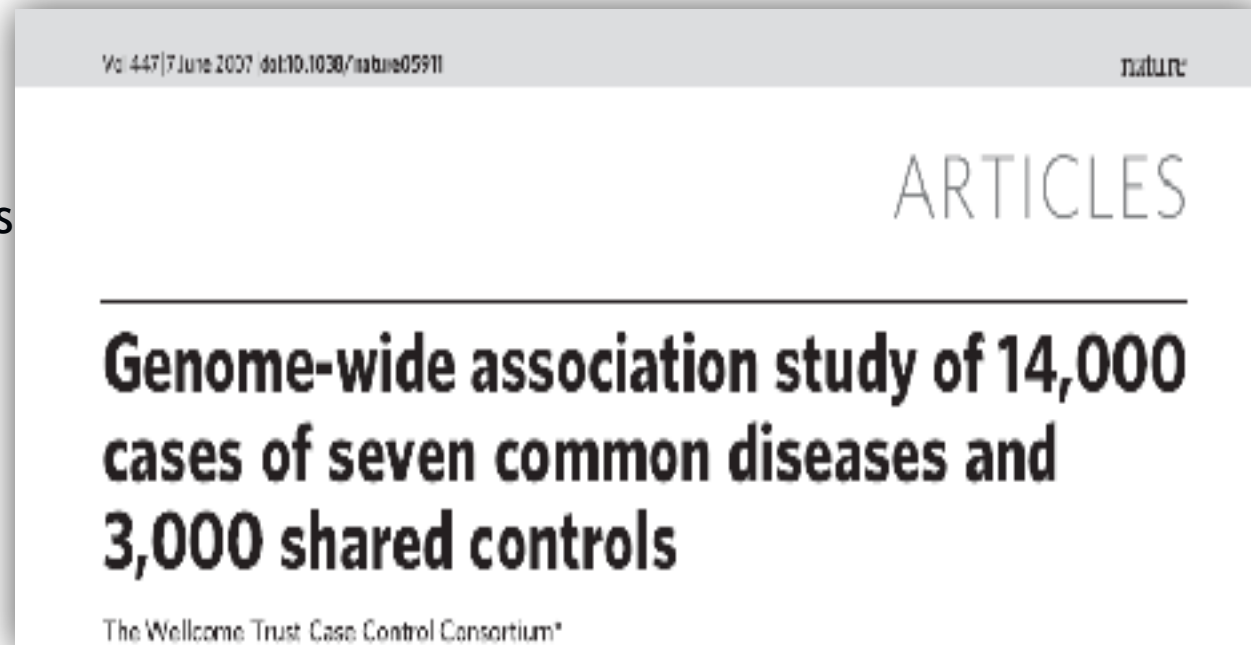
A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

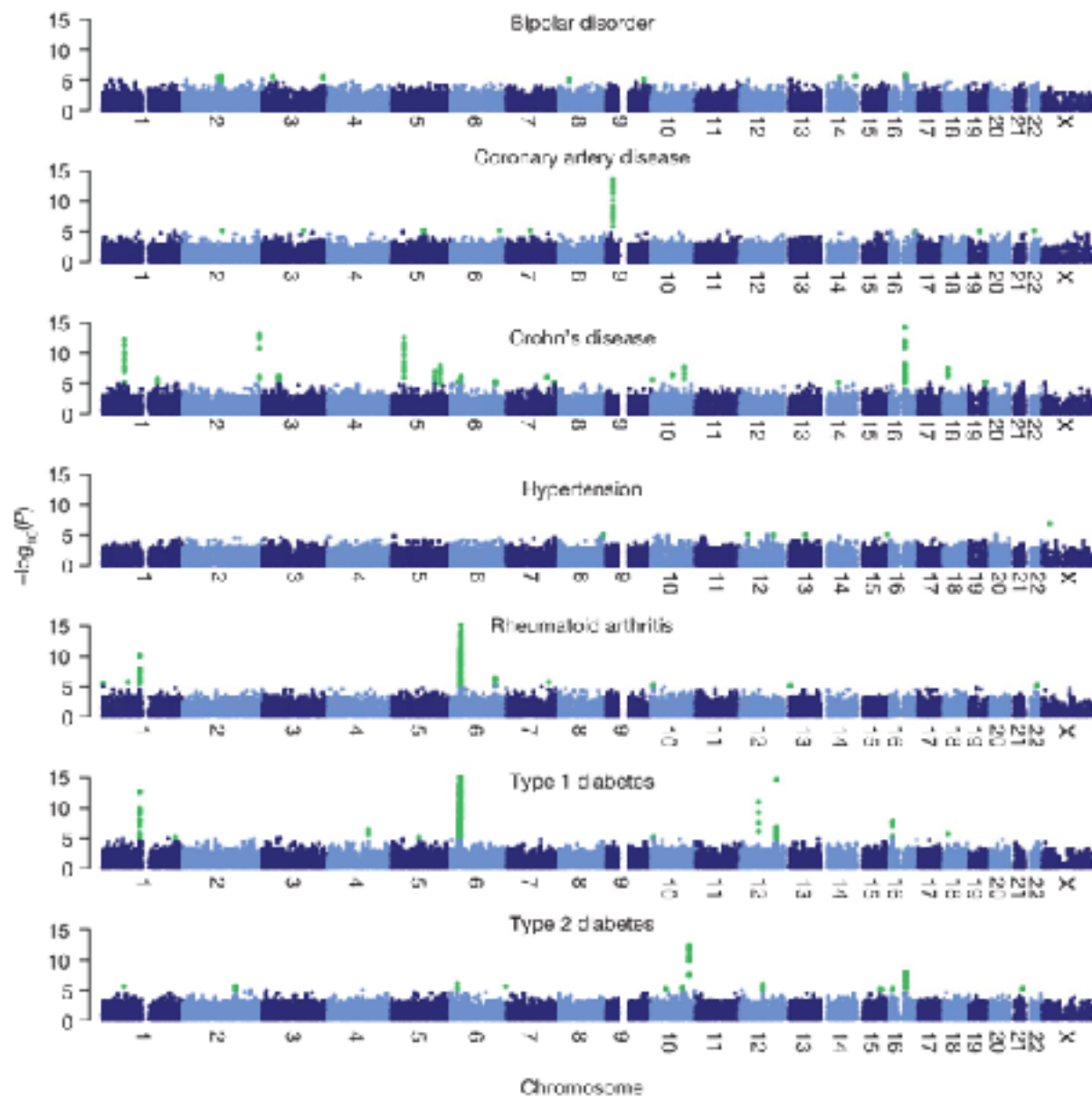
Anna Helgadóttir,^{1,2} Gudmar Thorleifsson,^{1,2} Andrei Mianaleva,^{1,2} Solveig Gretarsdóttir,¹ Thorarinn Blondal,¹ Asleaug Jonsdóttir,¹ Adalbjörg Jonsdóttir,¹ Asgeir Sigurdsson,¹ Adam Sævi,¹ Annar Pálsson,¹ Jóni Músson,¹ Gunnar P. Gudbjartsson,¹ Arnttinn P. Magnusson,¹ Karl Andersen,¹ Allan I. Levey,³ Valgerdur M. Baldursson,¹ Sigrúnur Jóhannsson,¹ Thorbjörg Jonsdóttir,¹ Stefan Pálsson,¹ Jóni Músson,¹ Sveinna Gunnarsdóttir,¹ Arnaldur Gylfason,¹ Viola Vercarone,³ W. Craig Hoppel,³ Mardret P. Kelly,⁴ Christopher B. Gieger,¹ Farand Austin,¹ Gail J. Rode,¹ Jóni H. Stolt,¹ Asgeir A. Gunnarsson,¹ Jeffrey R. Goldner,¹ Gudmundur Thorgeirsson,¹ Unnur Thordisdothir,¹ Augustine Kong,¹ & Kari Stefansson¹

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

Wellcome Trust Case-Control Consortium

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





One famous example

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



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

Auna Helgadóttir,^{1,4} Gudmar Thorolfsson,^{1,4} Andrei Manolescu,^{1,4} Jónleif Gröndal,¹ Thorarinn Benediktsson,¹ Asleik Jónsdóttir,¹ Adalbjörg Jónsdóttir,¹ Angel Thorgeirsson,¹ Adam Haxel,¹ Umar Farooq,¹ Jóni Músson,¹ Daniel F. Gudbjartsson,¹ Kristján P. Magnússon,¹ Karl Andersen,¹ Allan I. Levey,² Valgerður M. Baldursson,¹ Sigurðinn Atladóttir,¹ Thorbjörg Jónsdóttir,¹ Stefan Palsson,¹ Jóna Björnsdóttir,¹ Sveinna Gunnadóttir,¹ Arnaldur Gylfason,¹ Viola Vaccarino,³ W. Craig Hoeser,³ Mardis E. Daly,⁴ Christopher B. Gieger,¹ Farand Austin,¹ Gavin J. Ross,¹ Iván H. Shah,⁵ Anders A. Gylfason,¹ Jeffrey R. Gulcher,¹ Guðmundur Thorleifsson,¹ Umar Thorgeirsdóttir,¹ Augustine Borg,¹† Karl Stefánsson¹

A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

Ruth McPherson,^{1,†} Alexander Pertsemlidis,^{2,†} Nihan Kavazlar,¹ Alexandre Stewart,¹ Robert Roberts,¹ David R. Cox,² David A. Hinds,³ Len A. Pennacchio,^{4,5} Anne Tybjærg-Hansen,⁶ Aaron R. Folsom,⁷ Eric Boerwinkle,⁸ Helen H. Hobbs,^{2,9} Jonathan C. Cohen^{2,30,†}

Helgadóttir, 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

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

AUGUST 2, 2007

VOL. 357 NO. 5

Genomewide Association Analysis of Coronary Artery Disease

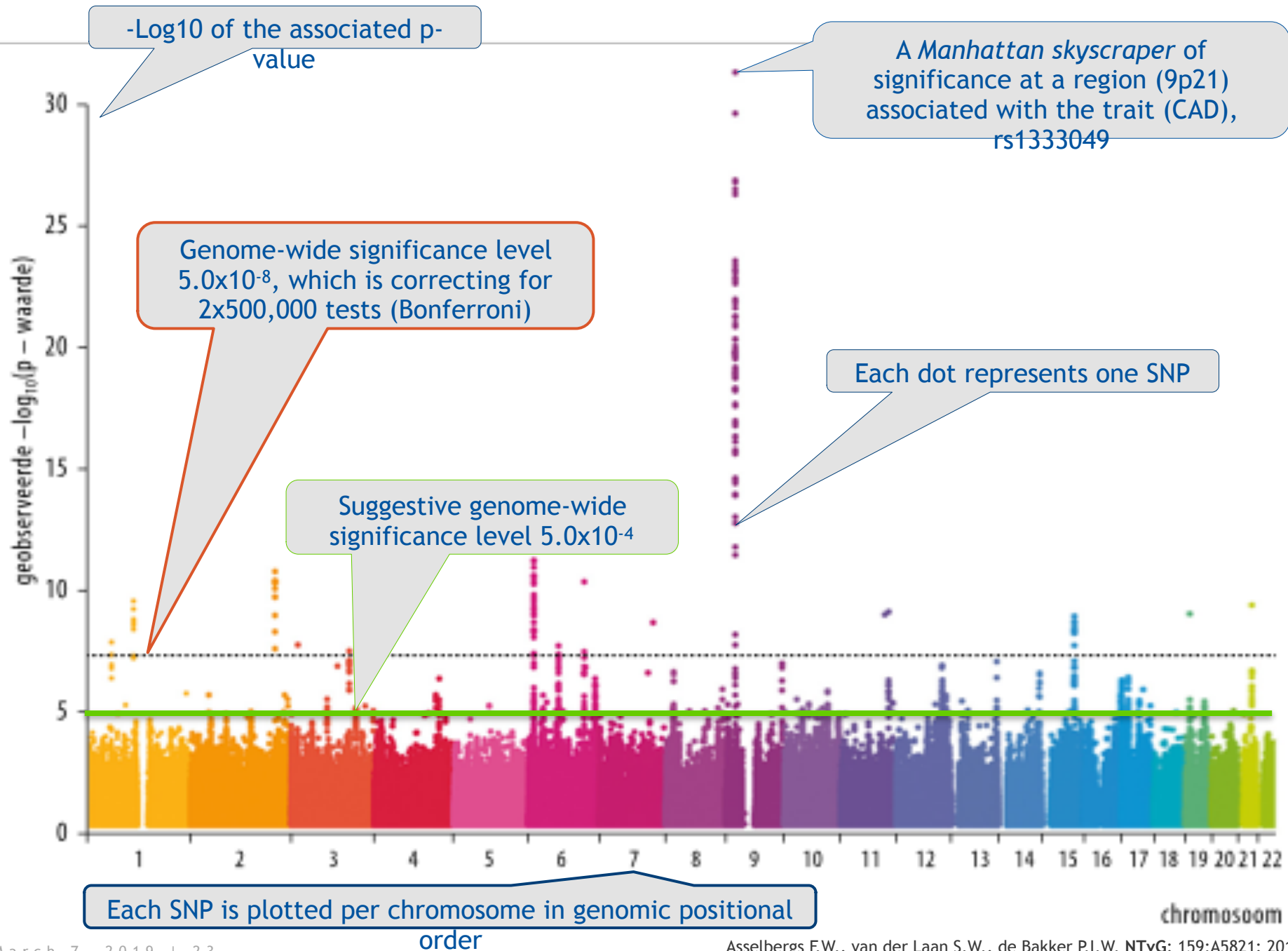
DOI:10.1056/NEJMoa0708483

1481-1489

ARTICLES

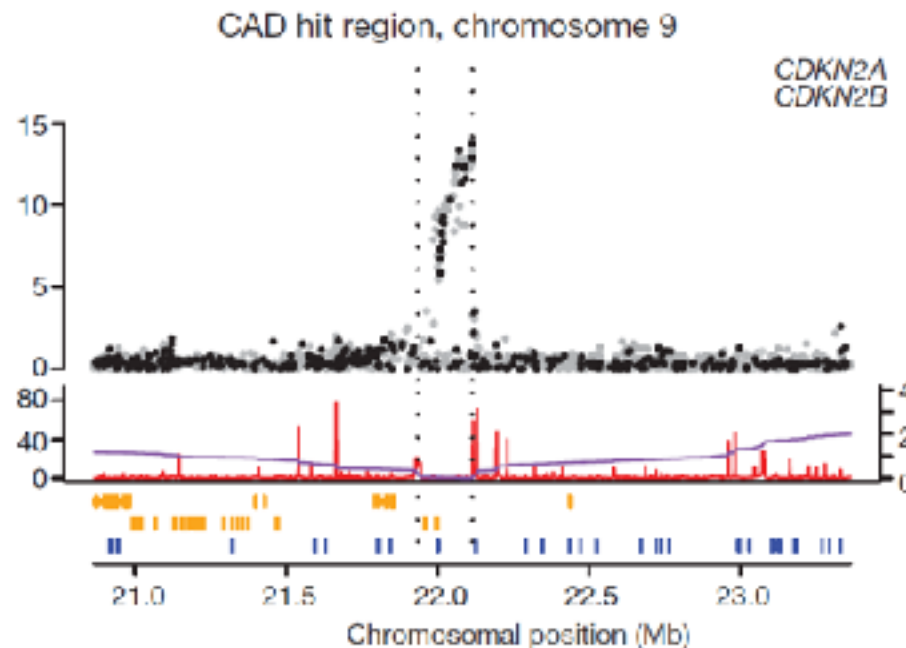
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium[†]



9p21 and cardiovascular disease

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



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

A closer look at the results...

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

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

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

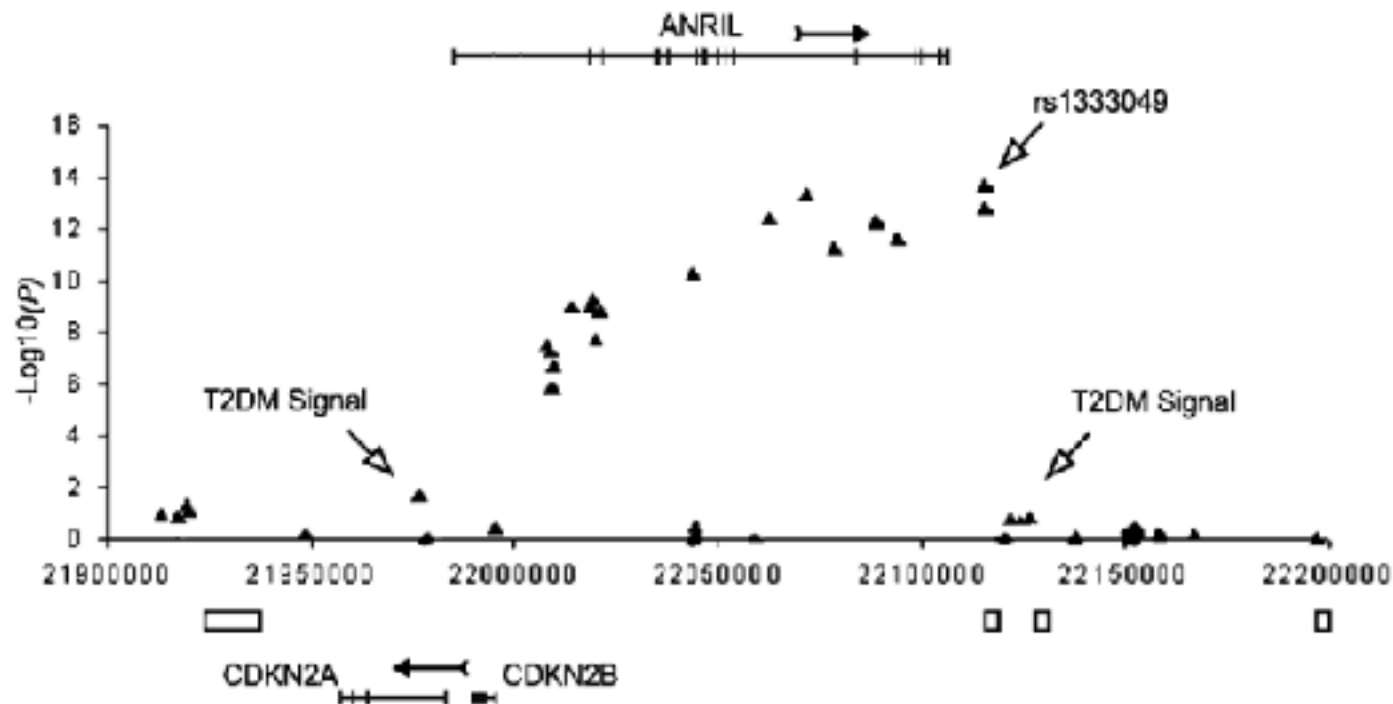
Risk allele: minor allele
Odds ratio: the odds of exposure between cases and controls

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

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

9p21 points to a *RNA* gene

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



CARDIoGRAMplusC4D Study

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

ARTICLES

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

The CADtoGRAMplusCAD Consortium¹

Concomitant artery disease (CAD): In the commonness: cause of death. Here, we report an association analysis meta-746 CAD cases and 133,671 controls identifying 13 loci revealing genome-wide significance, tailoring the number of susceptibility loci for CAD to 44, and a further 164 independent variants ($p < 3.2 \times 10^{-6}$, strongly associated with CAD) at 15 loci false discovery rate (FDR). Together, these variants explain approximately 10.6% of CAD heritability. Of the 46 genome-wide significant lead SNPs, 12 show a significant association with a light trait, and 1 show a significant association with blood pressure, but none is significantly associated with stroke. Network analysis with 281 confounding genes (found at 16% FDR) generated 5 distinct networks, with 25% of the SNPs in the network of genes involved in the regulation of the inflammatory response. These genes are linked to lipid metabolism and inflammation, underscoring the causal role of these pathways in the genetic biology of CAD. Our study provides insights into the genetic basis of CAD and identifies key biological pathways.

Environmental factors diverse include stress, nutrition, epigenetics, infections, etc. in the rising cause of death worldwide. Although, epidemiological studies have identified major risk factors for CVD, including plasma lipid concentrations, blood pressure, smoking, diabetes and markers of inflammation, a causal role has been shown only for some (for example, low-density lipoprotein (LDL) cholesterol and blood pressure), remotely through randomised double-blind drug therapy directed at the risk factor¹. Fats and family studies have suggested that a significant proportion (40–50%) of susceptibility to CVD lies in the family (in a certain sense, 20–30% of genetic variation are not understood by conventional Mendelian genetics; genetic analysis has the potential to define key risk factors) and suggest a causative role for diet, physical activity and therapeutic agents^{2–4}. In fact, genome-wide association studies (GWAS) have identified several loci for CVD, including LDL cholesterol, HDL cholesterol, triglycerides, blood pressure, and smoking^{5–10}. The

meta-analysis of E2159 of adult 17q21.31 carriers (22,23). Our analysis and 14,742 controls, imputation to loci representing genome-wide significance, a linkage disequilibrium (LD) by panel of 6,121 variants (see methods) and a meta-analysis (association *P* value of 1.6×10^{-10}). Here we integrate 6,121 SNPs in a meta-analysis of over 190,000 individuals with the primary aim of identifying additional susceptibility loci to GAE. To this end we used the Metabochip array[†] which has common SNPs (150,000) from 14,742 SNPs, designed to tag SNPs with $r^2 \geq 0.8$ in a population of several European ethnic groups including GAE, and 100 rare SNPs representing loci for three genes, APOB, SNRNP, and the *CD38* gene. The GAE study was conducted for analysis of 17q21.31 SNPs, of which 6,121 were the top custom SNPs and 14,742 were fine-mapping SNPs in the *CD38* gene. An explicitly loci identifier at the time at which the array was designed, the meta-analysis SNPs were submitted by the other consortia to control for the Metabochip array[†]. In addition, we assess whether the genome-wide association study (GWAS) identifies loci through traditional methods by considering the available data (25,26) for these two loci. In this study, we use a meta-analysis of the two studies to identify SNPs that are highly correlated with the GWAS threshold for association with GAE and use this set to undertake network analysis as first described biologically pathway underlying the pathogenesis of GAE.

RESULTS

Study design

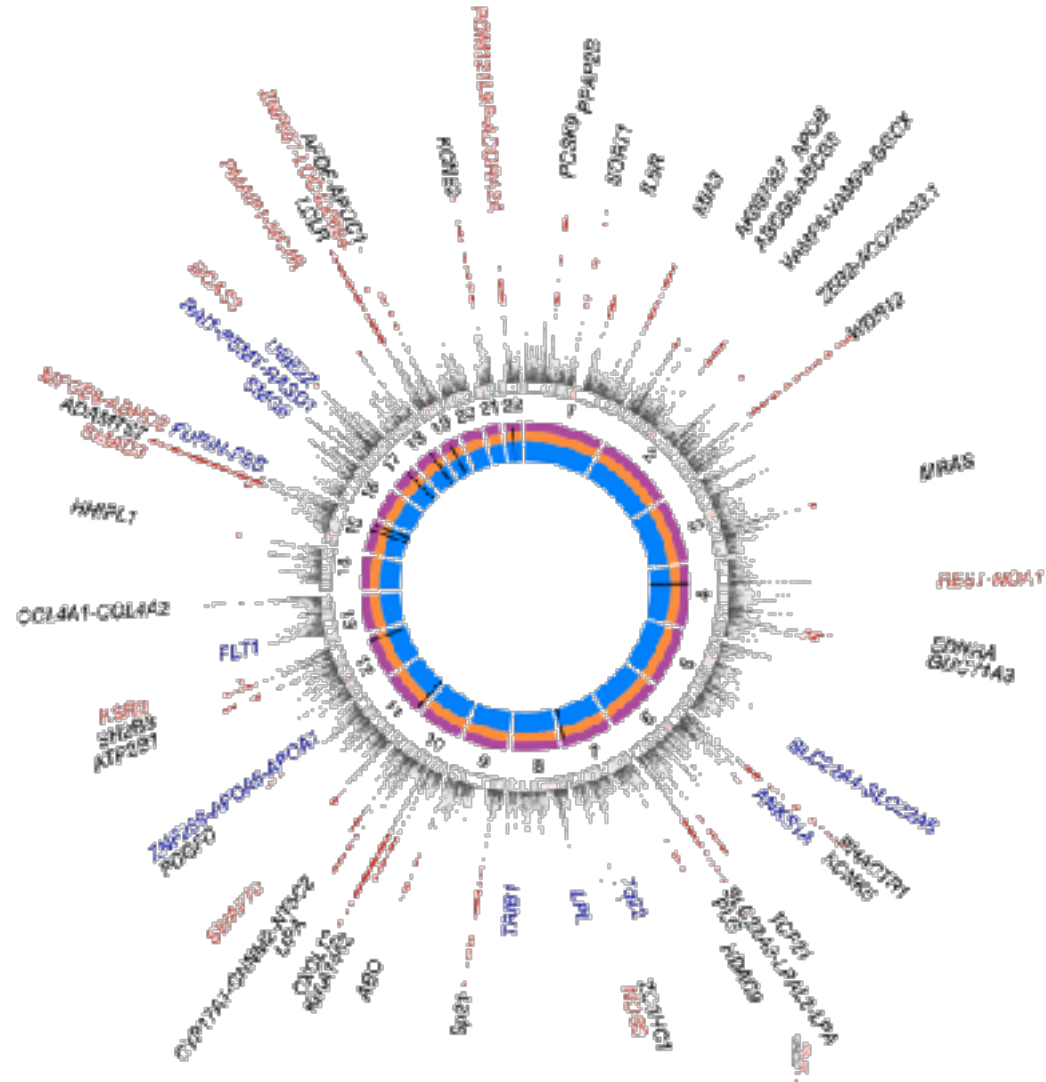
Measures of the CACNA1G discovery data set (21,130 cases and 64,782 controls), stage 1, with 34 additional CADD sample collections (stage 2) of European or south Asian descent comprising 61,313 cases and 15,905 controls (study descriptions and sample characteristics are given in **Supplementary Tables 1 and 2**, [www.nature.com/naturegenetics](#)).

¹For full list of authors, see Jettisoned Acknowledgements page at the end of the book.

Received 24 April 2012/ Accepted 17 November 2012/ Published online 2 December 2012/ doi:10.1017/S0022278X12001340

科学出版社出版 北京 100070

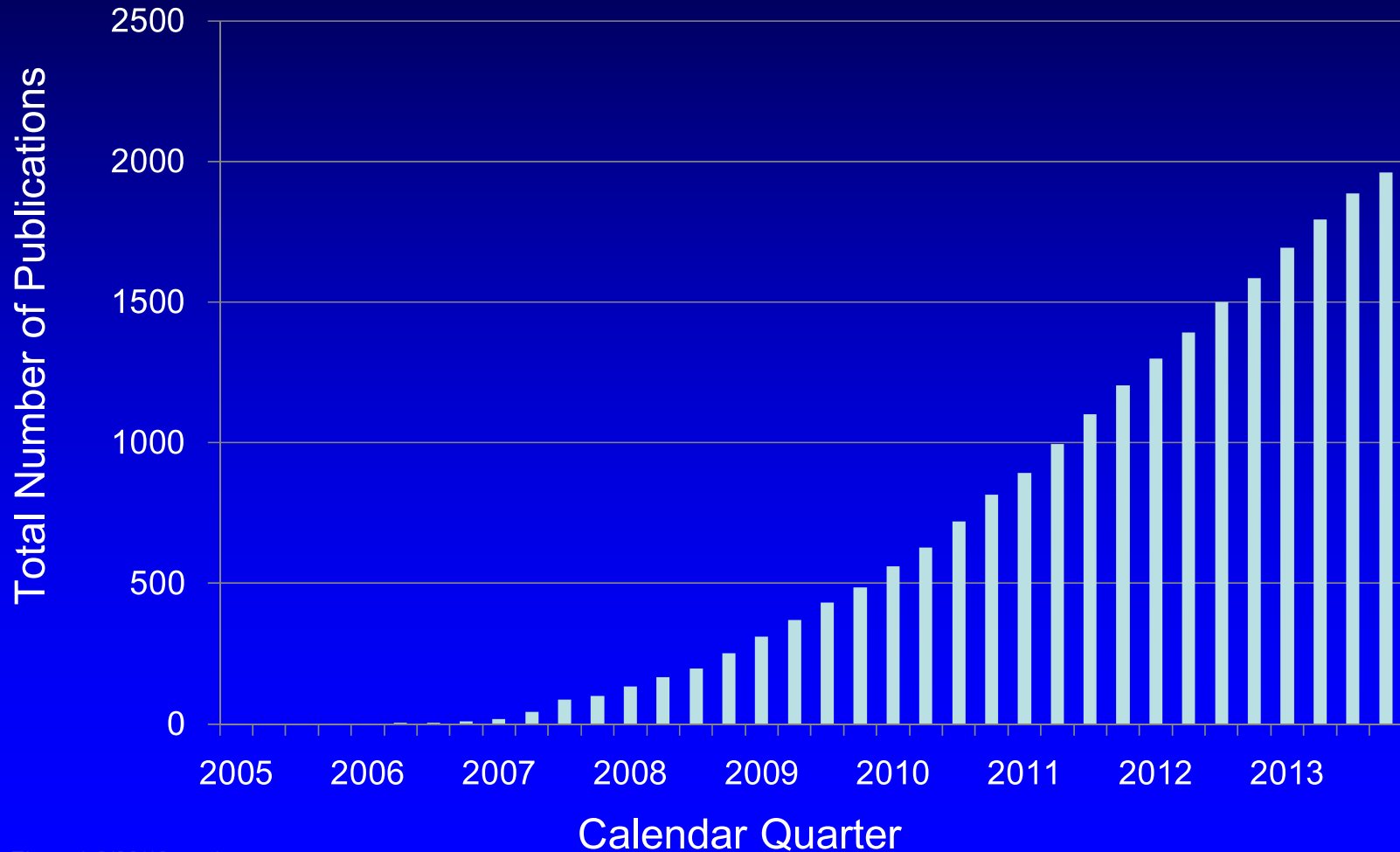
And 8 years later (>15 times more samples)



9p21 plus an additional 47 loci (!)

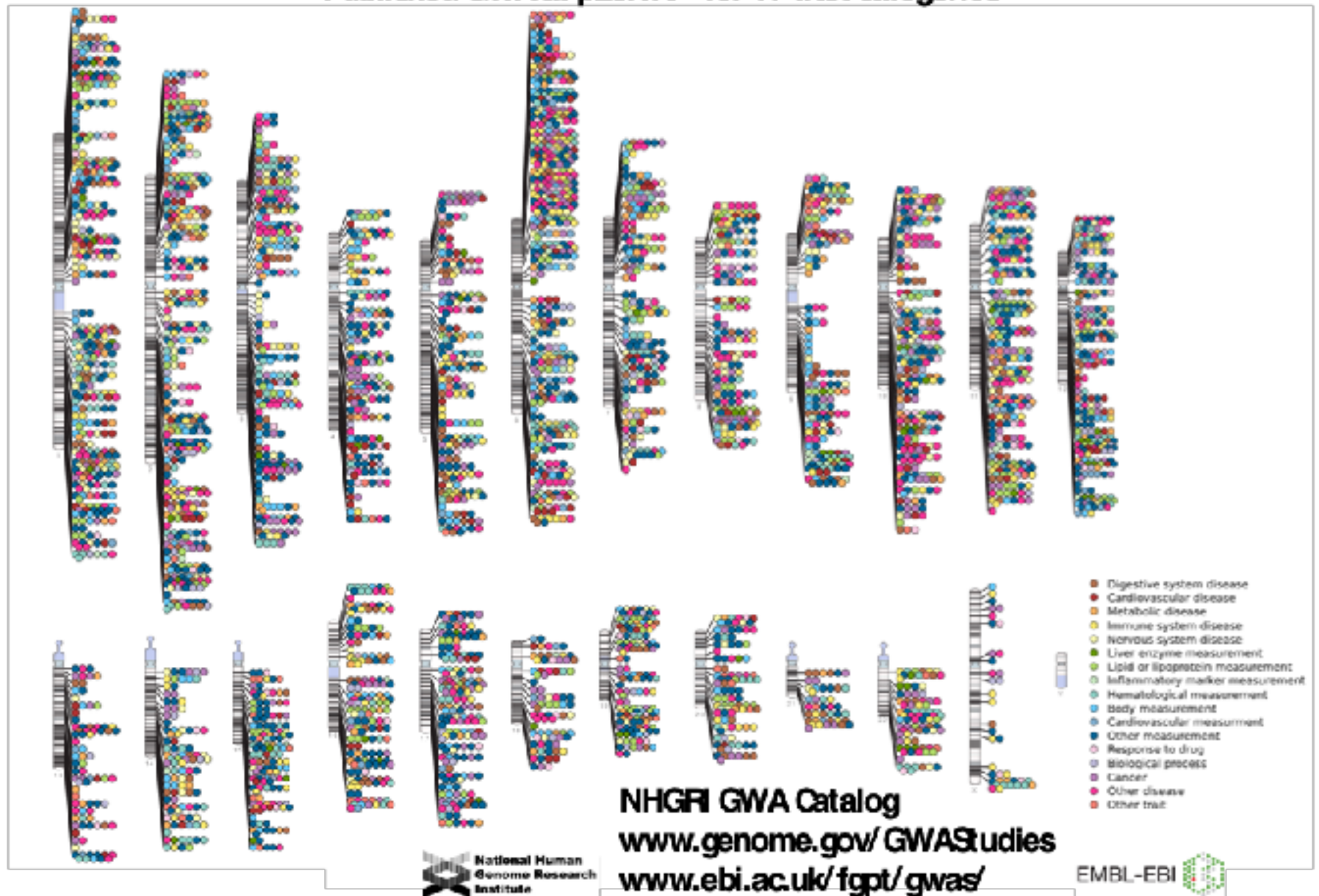
Published GWA Reports, 2005 – 2013

1960



Published Genome-Wide Associations through 12/2013

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



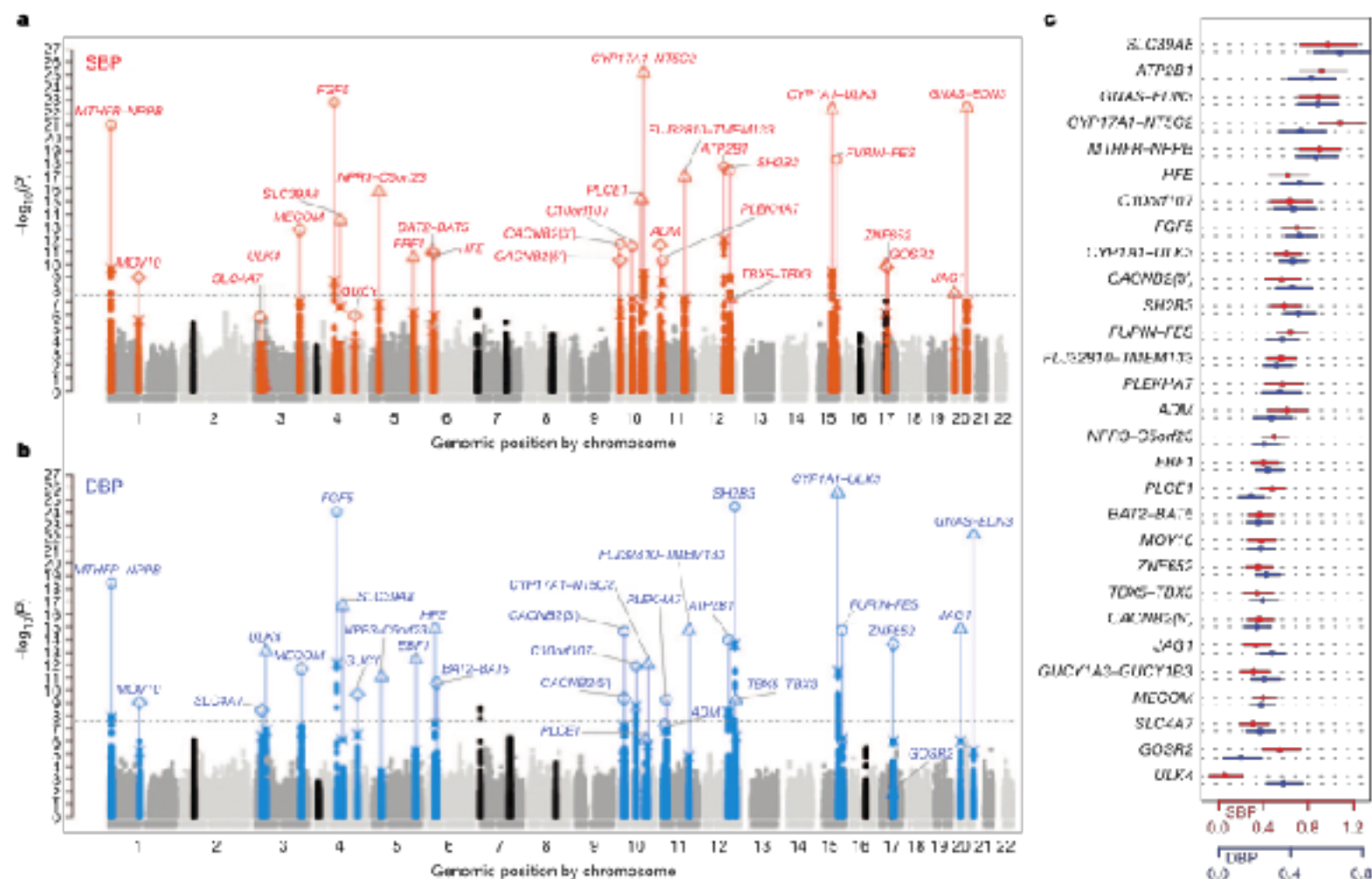


Figure 1 Genome-wide $-\log_{10} P$ -value plots and effects for significant loci. **a, b**, Genome-wide $-\log_{10} P$ -value plots are shown for SBP (**a**) and DBP (**b**). SNPs within loci reaching genome-wide significance are labeled in red for SBP and blue for DBP (± 2.5 Mb of lowest P value) and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

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

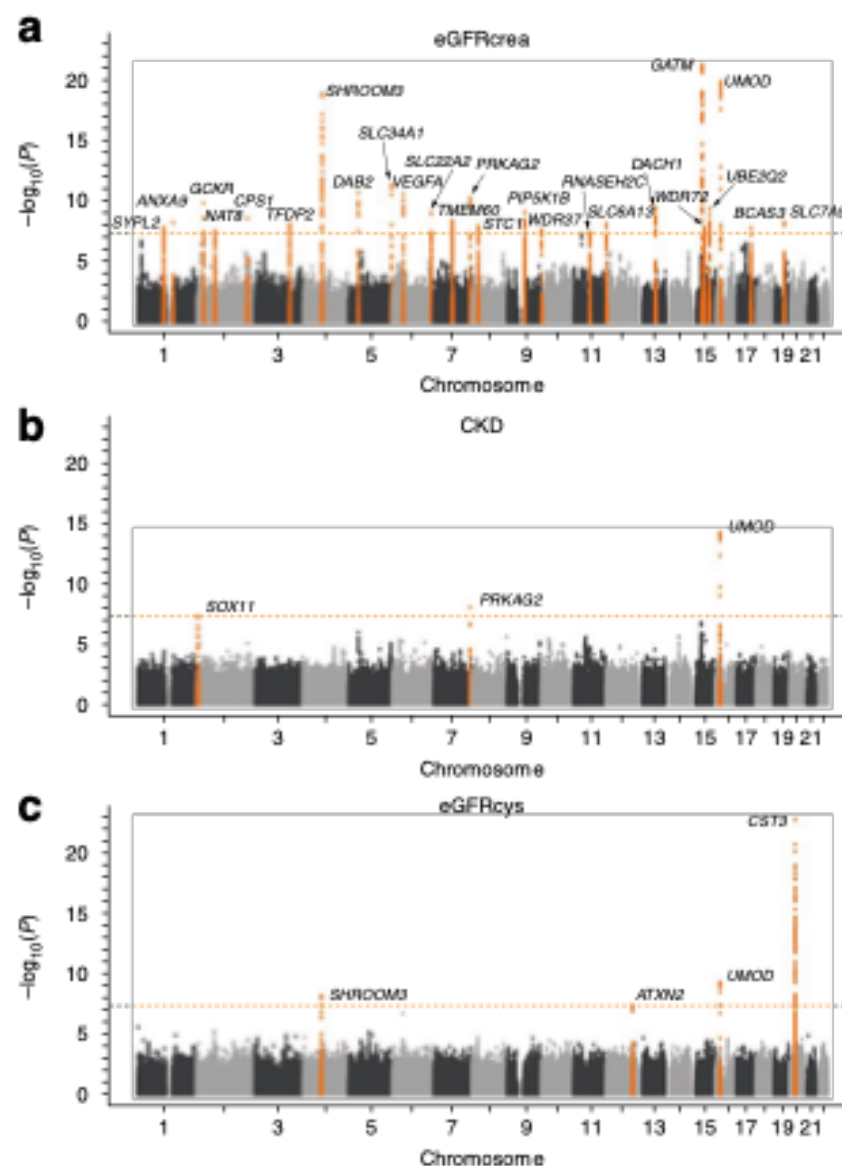


Figure 1 Genome-wide $-\log_{10} P$ value plot from stage 1. (a–c) Plots show discovery analysis of eGFRcrea (a), CKD (b) and eGFRcys (c). The dotted line indicates the genome-wide significance threshold at $P = 5 \times 10^{-8}$.

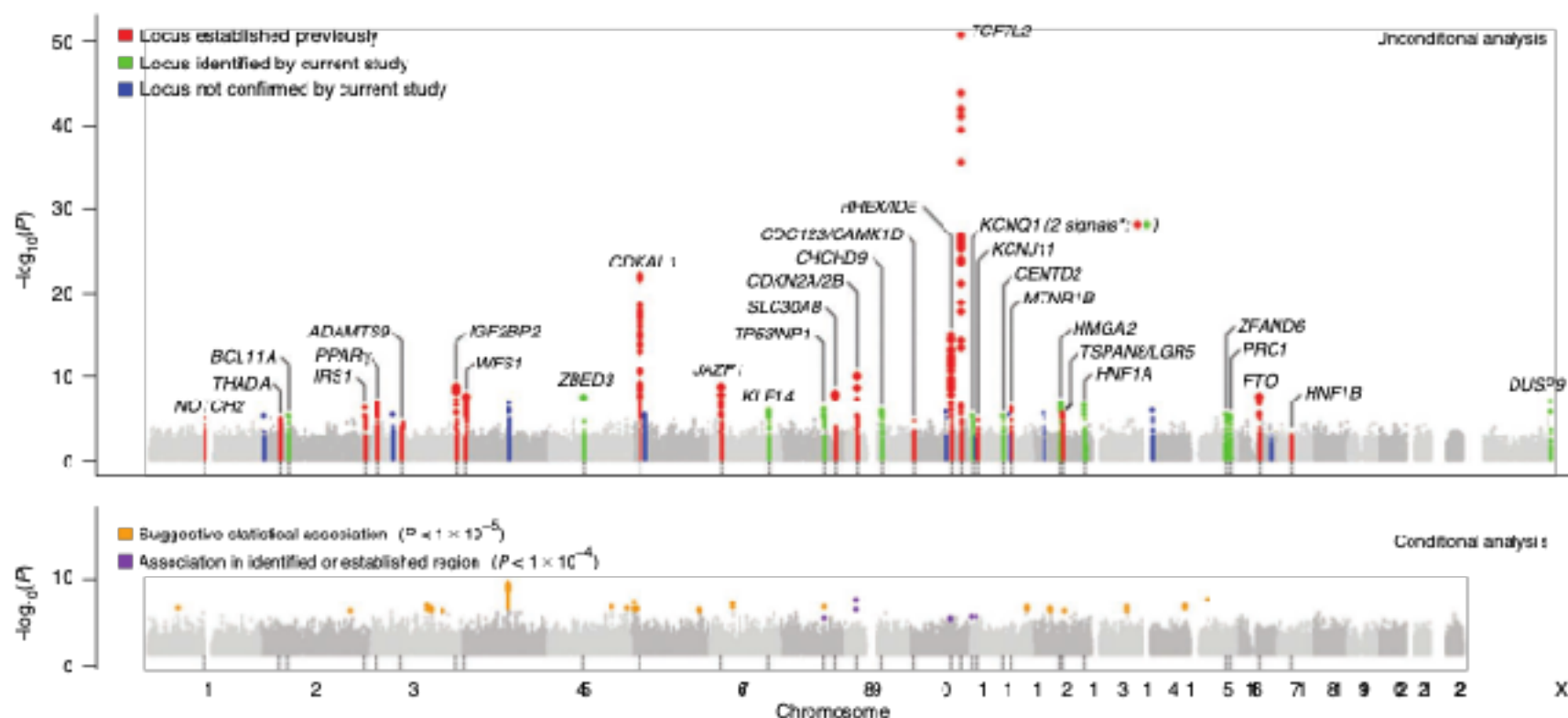
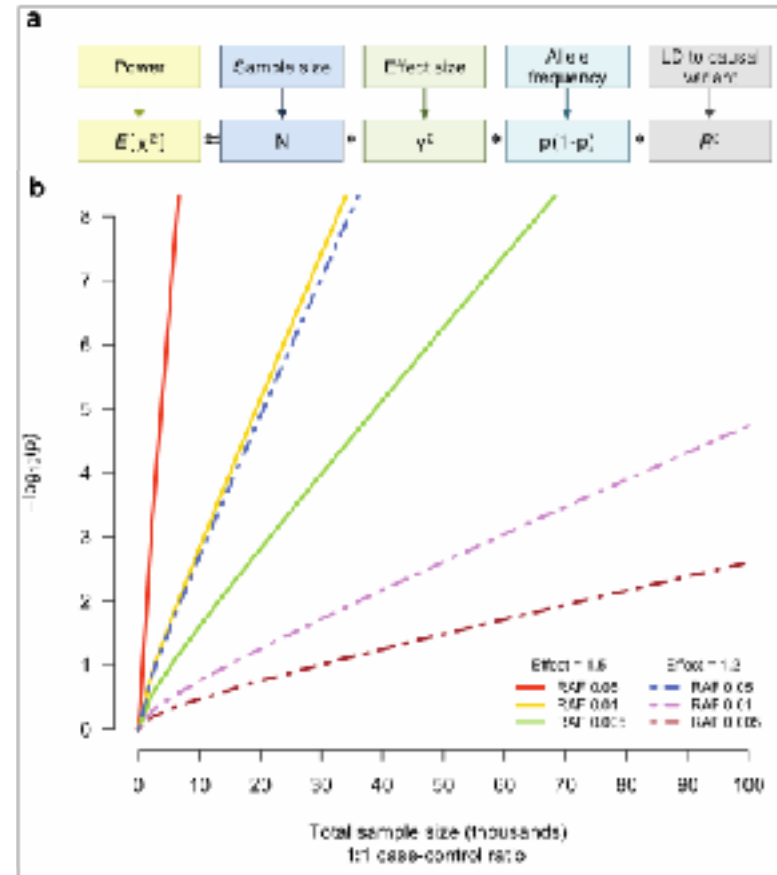
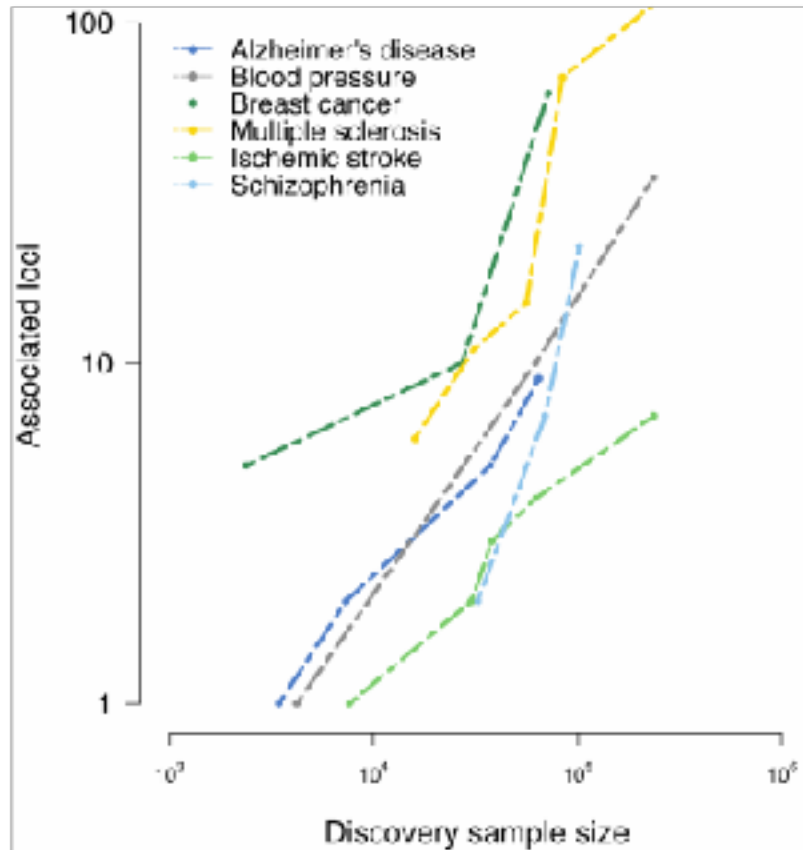


Figure 1 Genome-wide Manhattan plots for the DIAGRAM+ stage 1 meta-analysis. Top panel summarizes the results of the unconditional meta-analysis. Previously established loci are denoted in red and loci identified by the current study are denoted in green. The ten signals in blue are those taken forward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

Power, Effect size, Sample size...



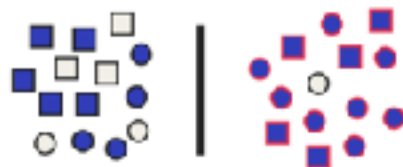
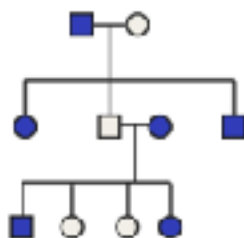


Linkage
analysis

Candidate
gene
studies

GWAS

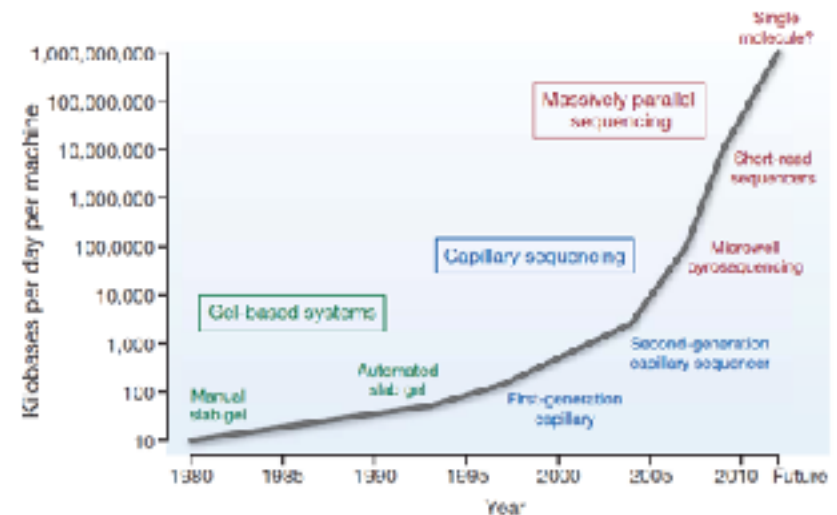
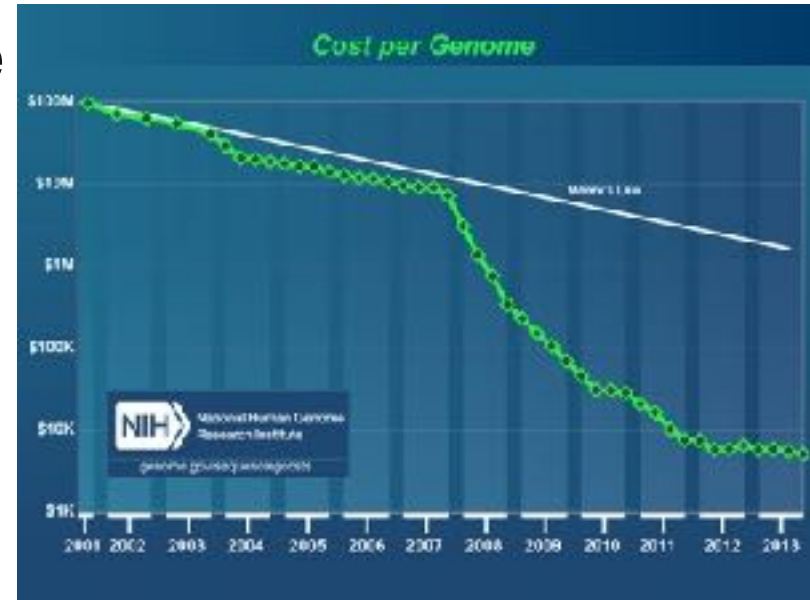
Sequencing



Next-generation sequencing

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

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



Summary: what's been (being) done?

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

