Update on Cardiovascular Genomic Epidemiology: Using Genomics in Risk Assessment & Prevention of Coronary Heart Disease

Christopher J. O'Donnell, MD MPH FACC FAHA

Public Health Genomics Interest Group (PHGIG)

2010 NIH Seminar Series: Advances in Public Health Genomics

February 24, 2010

DISCLOSURES: NONE

Department of Health and Human Services • National Institutes of Health

National **Heart Lung and Blood** Institute

People Science Health

Genomic Studies and Public Health

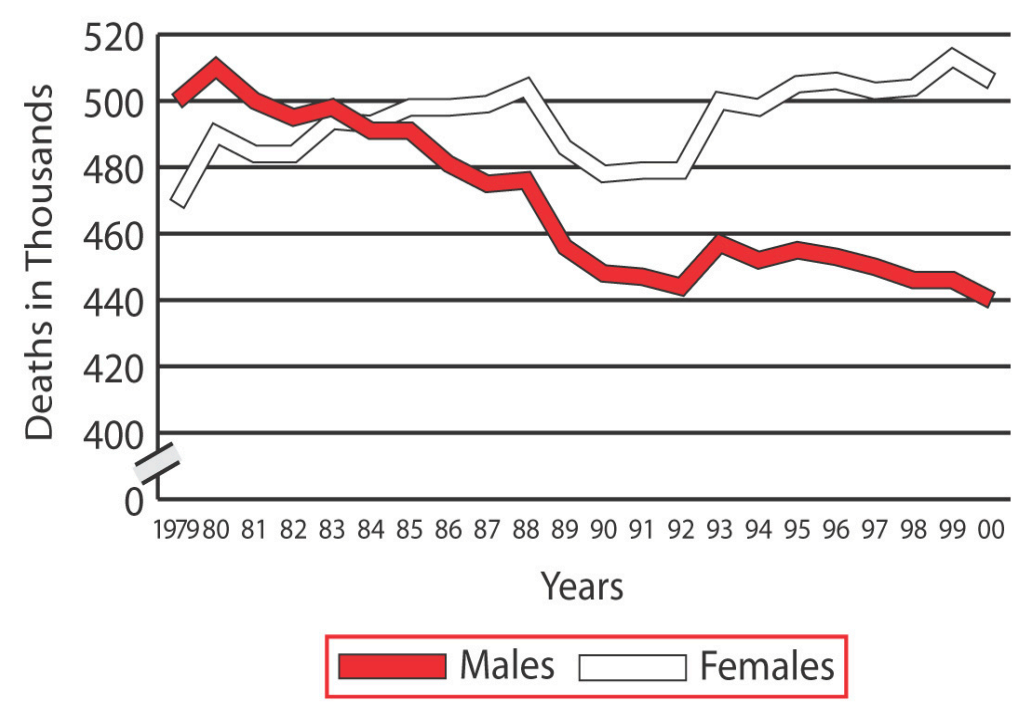
- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

Genomic Studies and Public Health

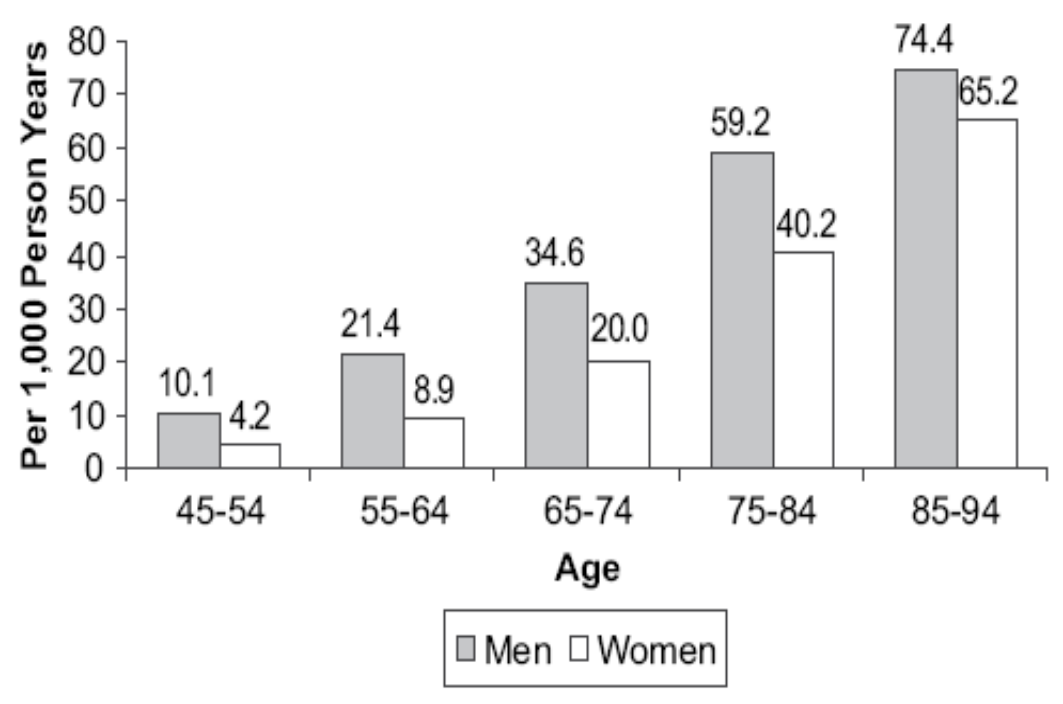
- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

Deaths from Cardiovascular Disease: Decreasing Overall, Increases with Age

CVD Death Trends*



Incident CVD and Age**



AHA Heart and Stroke Facts: *US, 1979-2000, **US 1988-1994.
Rosamond et al. Circulation 2008.

Framingham Heart Study

Downtown Framingham, MA (circa 1960)

1948 → 1958 → 1968 → 1978 → 1988 → 1998 → 2008

1948 → → → → → → → 2008

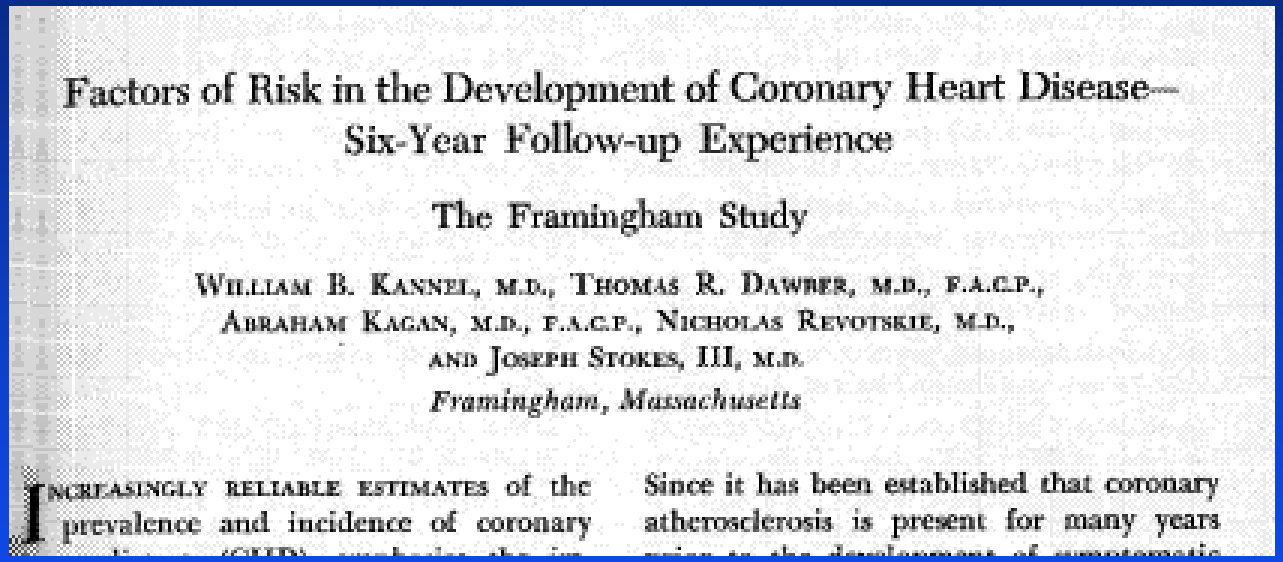
Original cohort: N = 5209 men and women (ages 28-62)
1644 spouse pairs, 596 extended families

1972 → → → → 2008

Offspring study: N = 5124 men and women (ages 5-70)
1576 spouse pairs, 3514 biological offspring

Third Generation study: 2002 →
N=~4000 men and women

“Factors of Risk” for CVD



Factors of Risk in the Development of Coronary Heart Disease—Six-Year Follow-up Experience

The Framingham Study

WILLIAM B. KANNEL, M.D., THOMAS R. DAWBER, M.D., F.A.C.P., ABRAHAM KAGAN, M.D., F.A.C.P., NICHOLAS REVOTSKIE, M.D., AND JOSEPH STOKES, III, M.D.

Framingham, Massachusetts

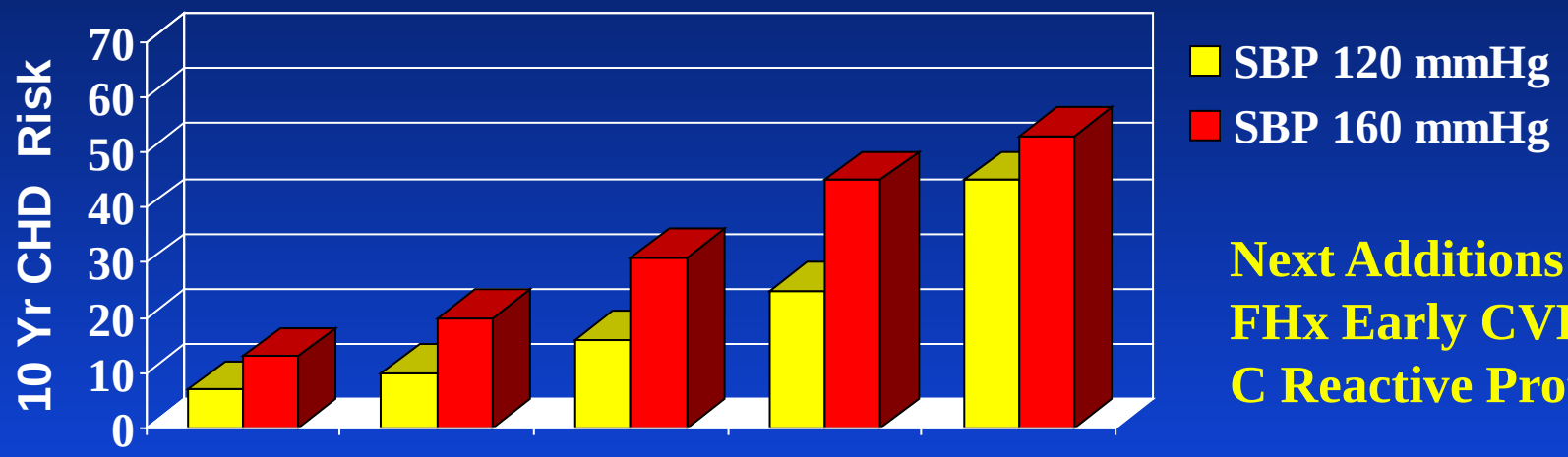
INCREASINGLY RELIABLE ESTIMATES of the prevalence and incidence of coronary

Since it has been established that coronary atherosclerosis is present for many years

Annals of Internal Medicine
November **1961**
Vol. 55, No.1

- High Blood Pressure
- Increased Cholesterol
- Smoking
- Diabetes
- Family History
- Male Gender

Framingham 10 Year CHD Risk Algorithm



Tot Chol	180	240	240	240	240
HDL Chol	50	50	35	35	35
Smoking	No	No	No	Yes	Yes
Diabetes	No	No	No	No	Yes

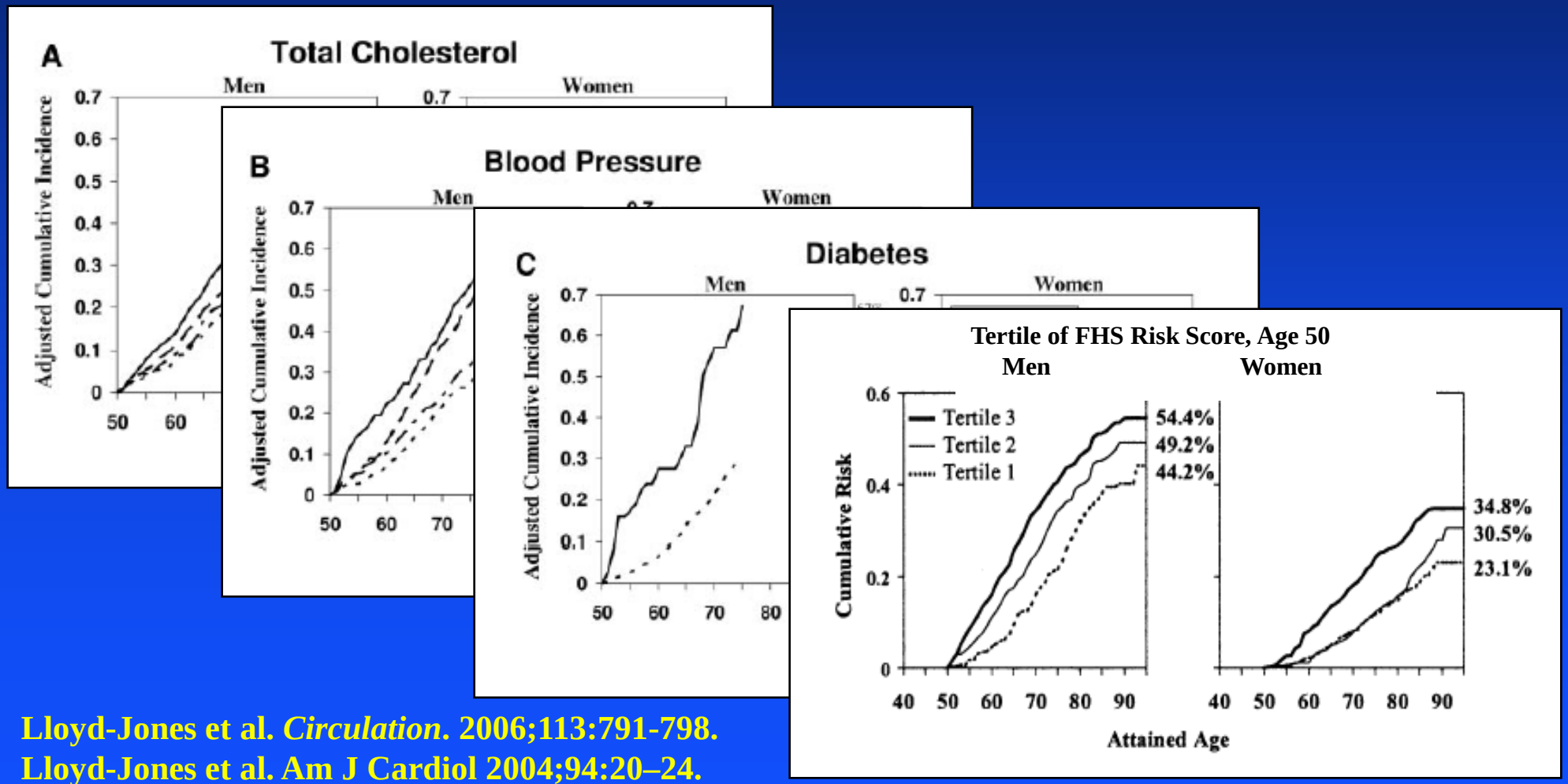
Source: Framingham Heart Study, Wilson et al. Circulation 1998;97:1837.
Example shown for Typical 55 Year-Old Man.

Next Additions:
FHx Early CVD?
C Reactive Protein?

NHLBI ATP LIPID GUIDELINES:

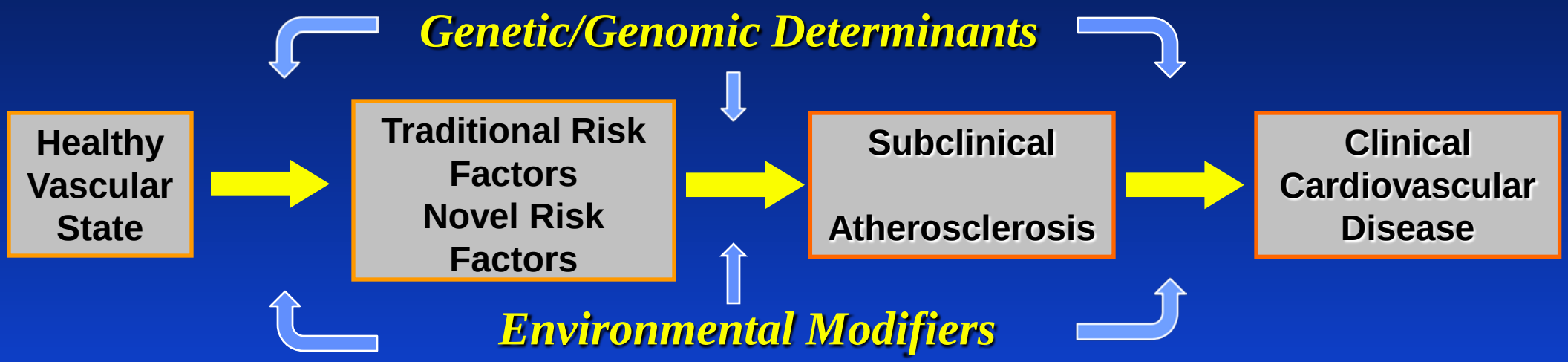
STEP 4: If 2+ risk factors (other than LDL) present without CHD or CHD risk equivalent, assess 10-year (short-term) CHD risk (use Framingham risk score).

Lifetime CHD Risk for Risk Factors and Overall FHS CHD Risk Score



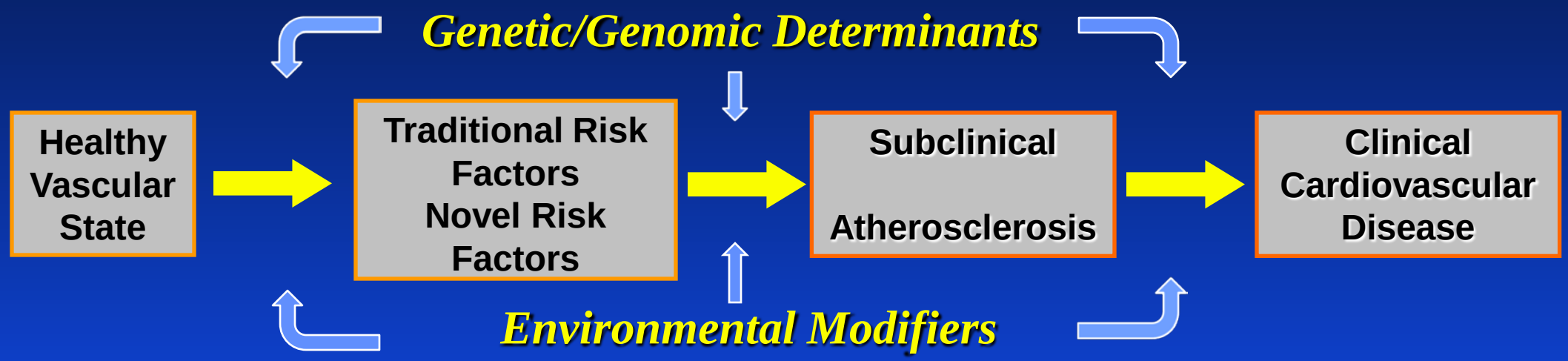
Lloyd-Jones et al. *Circulation*. 2006;113:791-798.
Lloyd-Jones et al. Am J Cardiol 2004;94:20–24.

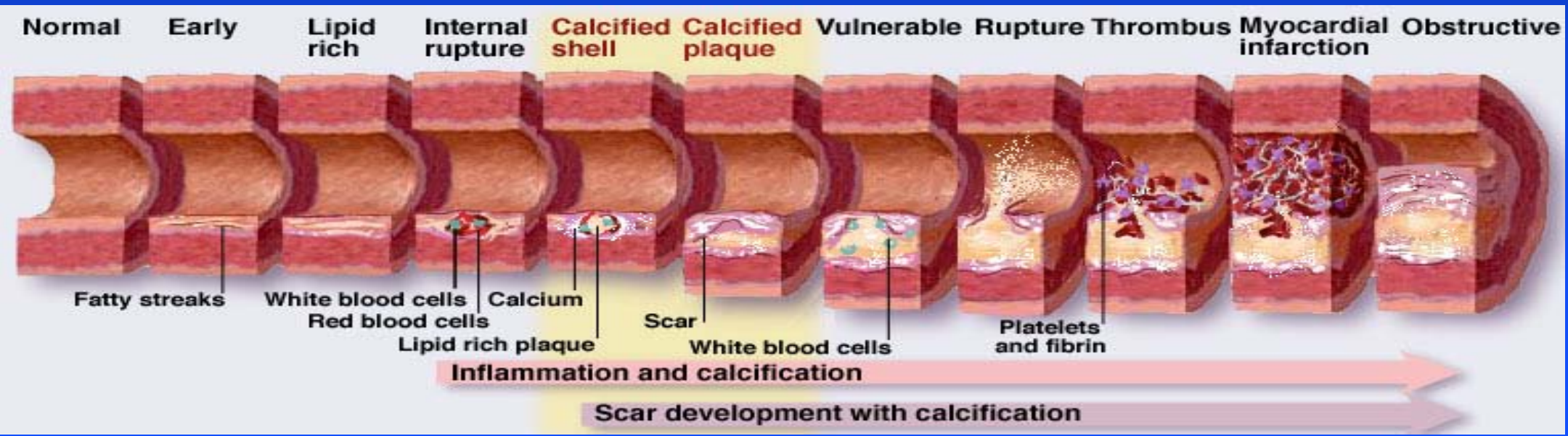
Atherosclerotic Plaque Development: From Healthy Vessel to Clinical CVD



Time Magazine, Feb 22 '10: www.time.com/time/covers

Atherosclerotic Plaque Development: From Healthy Vessel to Clinical CVD





Genomic Studies and Public Health

- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

Premature CVD in Parents* or Siblings** Leads to Increased Relative Risk of CVD

	Positive Parental History of Premature CVD		
Model Adjustment	**None**	**Both**	**≥1 Parent**
Men			
Age-adjusted	1.0	**3.1**	**2.6**
Multivariable-adjusted*	1.0	**2.4**	**2.0**
Women			
Age-adjusted	1.0	**4.1**	**2.3**
Multivariable-adjusted*	1.0	**2.8**	**1.7**

*Lloyd-Jones et al. JAMA 2004;291:2204. Adj. for age, total/HDL, SBP, anti-HTN Rx, AODM, BMI, current smoking.

Heritability of CVD Risk Factors & Biomarkers of Subclinical Disease: Framingham Study

Variable	Heritability
Risk Factors*	
Fasting Glucose	32%
Systolic Blood Pressure	42%
Maximum BMI	40%
Protein Biomarkers	
CRP	**26%**
PAI-1 Antigen	**26%**
Subclinical LV & Conduction Disease**	
LV Hypertrophy	**26%**
QT Duration	**35%**
Imaging Biomarkers of Atherosclerosis	
Wall Thickness,Carotid Artery†	**38%**
Calcium, Abdominal Aorta‡	**38%**
Calcium, Coronary Arteries‡	**38%**

*L. A. Cupples, ASHG 1996. *E. Benjamin, AHA 2000. †C Fox, Stroke 2002,
‡P Peyser, Circulation 2002, ††C O'Donnell, Circulation 2002, Newton-Cheh, Heart Rhythm 2005.

The Human Genome and Hapmap Projects

Nature October 21 2004;431:931.

articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

* *A list of authors and their affiliations appears in the* Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

"2.85 billion nucleotides …encode only 20,000-25,000 protein-coding genes"

The genome consists of ~ 1 SNP per 1000 bp and "a block-like structure of linkage disequilibrium…leading to substantial correlations of SNPs with…their neighbours"

Science October 27 2005;431:931

nature

ARTICLES

A haplotype map of the human genome

The International HapMap Consortium*

Inherited genetic variation has a critical but as yet largely uncharacterized role in human disease. Here we report a public database of common variation in the human genome: more than one million single nucleotide polymorphisms (SNPs) for which accurate and complete genotypes have been obtained in 269 DNA samples from four populations, including ten 500-kilobase regions in which essentially all information about common DNA variation has been extracted. These data document the generality of recombination hotspots, a block-like structure of linkage disequilibrium and low haplotype diversity, leading to substantial correlations of SNPs with many of their neighbours. We show how the HapMap resource can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution.

Translation From Genomics Studies to Personalized Genomic Medicine

Transcriptomic, Proteomic, Metabalomic

Genomewide: GWAS, Genome Sequencing

↓ ↓

Genome-Phenotype Association

↓

Sequencing & Functional Studies

↓

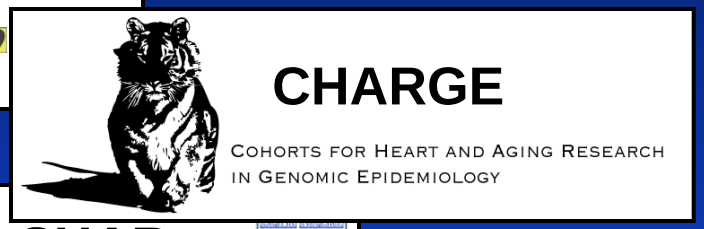
Genomics-Based Clinical Trials

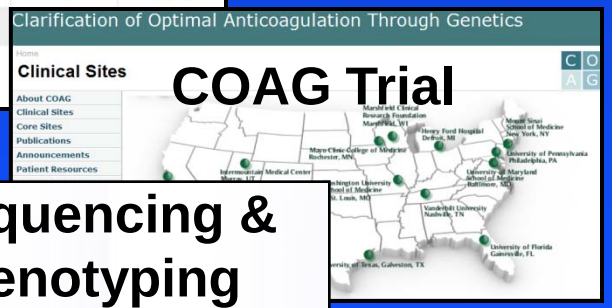
↓

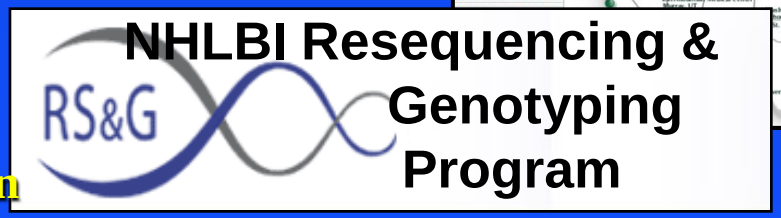
Applications for Genomic Medicine

↓ ↓ ↓

Predict	Prevent	Personalize
CVD Incidence & Recurrence	CVD Incidence & Recurrence	CVD PharmGen & Treatments







NHLBI GWAS Programs 2010

Program	Sample Size	Population(s)	Phenotype(s)
Framingham SHARe	~9,400	M+F; EA	☑ Many CVD & RFs
Asthma SHARe	~5,000	M+F; EA	☑ Asthma
MESA SHARe	~8,500	M+F; EA, AA, HA, CA	☑ Many CVD & RFs
Women's Health Initiative SHARe	~12,000	F; AA, HA	☑ Many CVD & RFs
STAMPEED	>50,000	M+F; EA, EU, AA, HA	☑ Many CVD & RFs, Asthma, Blood
CARe	~11,000	M+F; AA	☑ Many CVD&RFs, Pulm & Sleep Dz's
Women's Genome Health GWAS	~28,000	F; Largely EA	☑ Many CVD & RFs
COPD Gene	~10,000	M+F; EA, AA	☑ COPD
Total Subjects:	**>140,000**	M=male, F=female	☑ Yes Biomarker Data

Framingham SHARe (SNP Health Association Resource) GWAS Design

- All consenting subjects with DNA from 3 generations (n=9,500)
- Genotype 550,000 SNPs (Affymetrix) in each subject
- Include all N≥21,250 available 'phenotype' datasets
  - Original Cohort, Offspring, Third Generation
  - All available exams, Quantitative and qualitative
  - Primary traits (e.g., CVD and pulmonary)
  - Traits from ancillary studies
- Comprehensive, web-based database (dbGaP) accessible to qualified investigators since October 2007
- Over 100 applications approved by NHLBI's DAC

NCBI's dbGaP Main Page

dbGaP Home - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap

Most Visited | Getting Started | Latest Headlines | Top Websites | Citi Performing Arts C...

dbGaP Home | JAMA -- Abstract: Ass... | CHD_-_C'ODonnell.pd... | Filters and False Positi... | Capture and Subasse... | DDIR

NCBI

dbGaP GENOTYPES and PHENOTYPES

My NCBI [Sign In] [Register]

Search dbGaP for Go Clear

Limits | Preview/Index | History | Clipboard | Details

NCBI
dbGaP

Search dbGaP FAQ Go

About dbGaP
Browse dbGaP
Authorized Access
Email Alerts
dbGaP Tutorial
Security Procedures
FTP Download
Publications
Contact dbGaP
dbGaP Alert NEW

Other Services

MeSH Browser
Clinical Queries

Browse dbGaP TUTORIAL ABOUT dbGAP

By Studies | By Diseases | Advanced Search

Project ▲	Study	Embargo Release	Details	Participants	Type of Study
University of Oxford	SEARCH GWA Study of Statin-Induced Myopathy		V D A	175	Case-Control
University of Michigan	The Finland-United States Investigation of NIDDM Genetics (FUSION) study		V D A	2335	Case-Control
University of Chicago	Genome-Wide Association Study of Serum YKL-40 Levels		V D A	632	Population
SHARe	Framingham SHARe	Version 1: Oct 19, 2008 Version 2: Feb 01, 2009 Version 3: Jul 08, 2009 Version 4: Dec 04, 2009 Version 5: Mar 30, 2010 Version 6: Jun 04, 2010 Version 7: Oct 05, 2010 Version 8: Dec 03, 2010	V D A	14276	Longitudinal
SHARe	GAW16 Framingham and Simulated Data	Version 1: Oct 19, 2008 Version 2: Oct 19, 2008	V D A	7126	Longitudinal
SHARe	Framingham SHARe Social Network	Jun 26, 2010	V D A	14276	Longitudinal
SHARe	SNP Health Association Resource (SHARe) Asthma Resource Project (SHARP)	Version 1: Dec 11, 2009 Version 2: Dec 11, 2009	V D A	4046	Longitudinal
SHARe	Women's Health Initiative SHARe	Jan 13, 2011	V D A	12008	Partial Factorial Randomized
SHARe	MESA SHARe	Jan 14, 2011	V D A	8306	Longitudinal
NINDS	NINDS Parkinsonism Study	Oct 12, 2007	V D A	1283	Case Set

Find: morga Next Previous Highlight all Match case

Done

Genomic Studies and Public Health

- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

GWA Studies with Strongly Replicated Gene Associations 2005-2009

Cancer

- Prostate Cancer
- Breast Cancer
- Colon Cancer
- Thyroid Cancer

Lung

- Asthma
- COPD

Eye

- Macular Degeneration
- Glaucoma (XFG)

Gastrointestinal

- Crohn's and IBD
- Celiac Disease

Infectious Disease

- HIV Host Control

Cardiovascular Disease and Its Risk Factors

- MI, CAD, Stroke
- Obesity
- Cigarette Use
- Hypertension
- Diabetes, Type 2
- CRP, Biomarkers
- LV Mass, Diameter
- ECG QT interval
- Atrial Fibrillation

Rheumatic Diseases

- SLE (Lupus)
- Celiac Disease
- Rheumatoid Arthritis
- Crohn's Disease
- Sarcoidosis

Bone

- Osteoarthritis

Blood

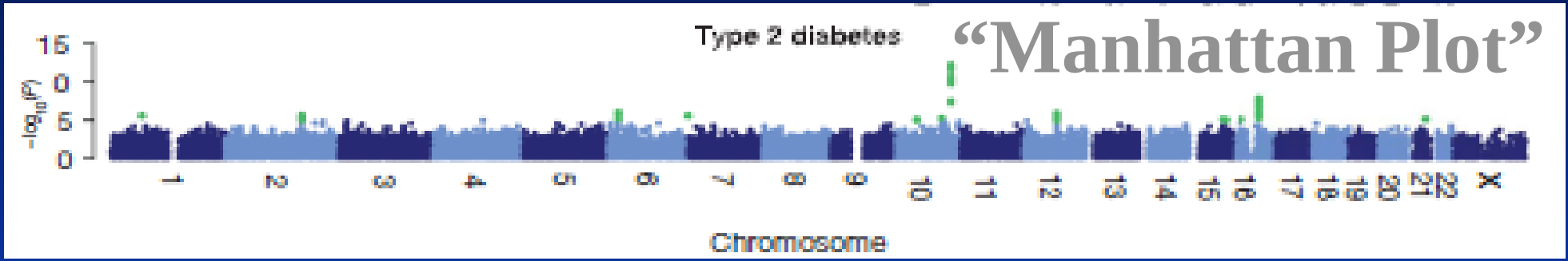
- Red Blood Cell Indices
- WBC, Platelet Count

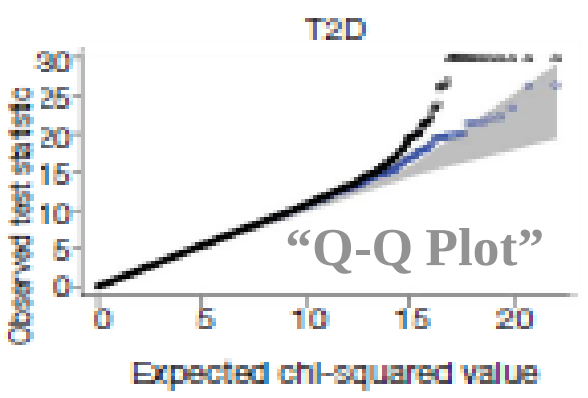
Psychiatry and Neurology

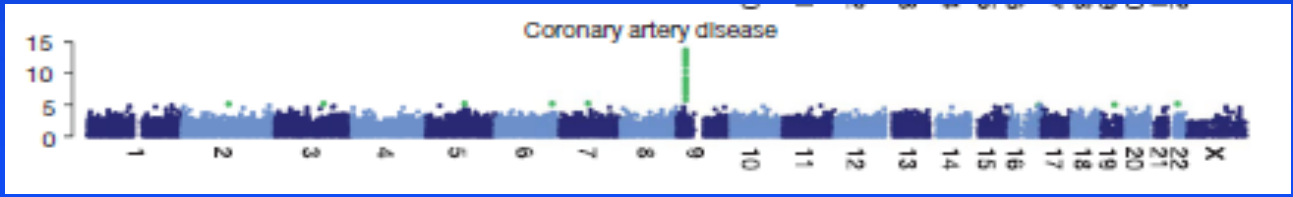
- Bipolar Disorder
- Schizophrenia
- Restless Leg Syndrome
- Multiple Sclerosis
- ALS

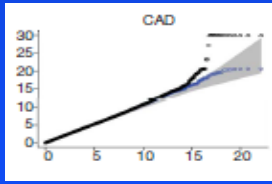
As of 1/15/10, 472 GWAS reports and 2204 SNPs, http://www.genome.gov/26525384.

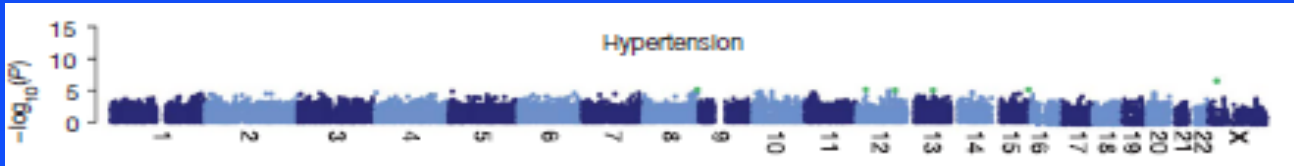
WTCCC: Manhattan Plot, Q-Q Plot

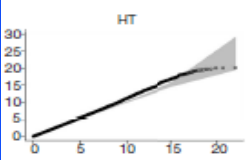












Common Allele on Chr 9 Associated with Coronary Heart Disease: GWAS 2007

Sciencexpress Report

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

Anna Helgadottir,[1]* Gudmar Thorleifsson,[1]* Andrei Manolescu,[1]* Solveig Gretarsdottir,[1] Thorarinn Blondal,[1] Aslaug Jonasdottir,[1] Adalbjorg Jonasdottir,[1] Asgeir Sigurdsson,[1] Adam Baker,[1] Arnar Palsson,[1] Gisli Masson,[1] Daniel Gudbjartsson,[1] Kristinn P. Magnusson,[1] Karl Andersen,[2] Allan I. Levey,[3] Valgerdur M. Backman,[1] Sigurborg Matthiasdottir,[1] Thorbjorg Jonsdottir,[1] Stefan Palsson,[1] Helga Einarsdottir,[1] Steinunn Gunnarsdottir,[1] Arnaldur Gylfason,[1] Viola Vaccarino,[3] W. Craig Hooper,[3] Muredach P. Reilly,[4] Christopher B. Granger,[5] Harland Austin,[3] Daniel J. Rader,[4] Svati H. Shah,[5] Arshed A. Quyyumi,[3] Jeffrey R. Gulcher,[1] Gudmundur Thorg

[1]deCODE genetics Inc, Reykjavik Medicine, Atlanta, GA 30322, US University School of Medicine, D

*These authors contributed equal

†To whom correspondence shoul

The global endemic of cardiova improved risk assessment and t describe an association between (MI) and a common sequence v 9p21. This study included a tota controls. The identified variant, suppressor genes *CDKN2A* and to the disease with high significa Approximately 21% of individu homozygous for this variant an 1.64-fold greater risk of sufferin than non-carriers. The correspo early onset cases. The populatio is 21% for MI in general and 31

A Common Allele on Chromosome 9 Associated with Coronary Heart

Ruth McPherson,[1]*† Alexander Pertsemlidis,[2]* Nihan Kavaslar,[1] Alexandre Stewart,[1] Robert Roberts,[1] Hinds,[3] Len A. Pennacchio,[4] Anne Tybjaerg-Hansen,[5] Aaron R. Folsom,[6] Eric Boerwinkle,[7] Helen H. H Cohen[2,8]†

[1]Division of Cardiology, University of Ottawa Heart Institute, Ottawa K1Y4W7, Canada. [2]Donald W. R Clinical Research Center and the Eugene McDermott Center for Human Growth and Development, Uni Southwestern Medical Center, Dallas, TX 75390, USA. [3]Perlegen Sciences, Mountain View, CA 94043 Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA & U.S. Department of E Institute, Walnut Creek, CA 94598, USA. [5]Department of Clinical Biochemistry, Rigshospitalet, Copen Hospital, Copenhagen DK-2100, Denmark. [6]Division of Epidemiology and Community Health, Univers Minneapolis, MN 55454, USA. [7]Human Genetics Center and Institute for Molecular Medicine, Univers Science Center, Houston, TX 77030, USA. [8]Center for Human Nutrition and the [9]Howard Hughes Medi University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: jonathan.cohen@utsouthwestern.edu or rmcph

Coronary heart disease (CHD) is a major cause of death in Western countries. Here we used genome-wide association scanning to identify a 58 kilobase interval on chromosome 9p21 that was consistently associated with CHD in six independent samples (n > 23,000 participants) from four Caucasian populations. This interval, which is located near the *CDKN2A* and *CDKN2B* genes, contains no annotated genes and is not associated with established CHD risk factors such as plasma lipoproteins, hypertension or diabetes. Homozygotes for the risk allele comprise 20-25% of Caucasians and have a ~30-40% increased risk of CHD.

Weinberg equilibrium (P<0.001) or d control criteria (3) and 17,500 were s allele frequency <1%) in the sample. SNPs were entered into the analysis a associated with CHD at a nominal sig 0.025 (table S2). These 2,586 SNPs w independent sample of 311 cases and Ottawa (OHS-2) using the same crite Of these, 50 were associated with CH significance threshold of 0.025, with effect (Table S2).

The NEW ENGLAND JOURNAL of MEDICINE

Genomewide Association Analysis of Coronary Artery Disease

Nilesh J. Samani, F.Med.Sci., Jeanette Erdmann, Ph.D., Alistair S. Hall, F.R.C.P., Christian Hengstenberg, M.D., Massimo Mangino, Ph.D., Bjoern Mayer, M.D., Richard J. Dixon, Ph.D., Thomas Meitinger, M.D., Peter Braund, M.Sc., H.-Erich Wichmann, M.D., Jennifer H. Barrett, Ph.D., Inke R. König, Ph.D., Suzanne E. Stevens, M.Sc., Silke Szymczak, M.Sc., David-Alexandre Tregouet, Ph.D., Mark M. Iles, Ph.D., Friedrich Pahlke, M.Sc., Helen Pollard, M.Sc., Wolfgang Lieb, M.D., Francois Cambien, M.D., Marcus Fischer, M.D., Willem Ouwehand, F.R.C.Path., Stefan Blankenberg, M.D., Anthony J. Balmforth, Ph.D., Andrea Baessler, M.D., Stephen G. Ball, F.R.C.P., Tim M. Strom, M.D., Ingrid Braenne, M.Sc., Christian Gieger, Ph.D., Panos Deloukas, Ph.D., Martin D. Tobin, M.F.P.H.M., Andreas Ziegler, Ph.D., John R. Thompson, Ph.D., and Heribert Schunkert, M.D., for the WTCCC and the Cardiogenics Consortium*

ABSTRACT

BACKGROUND

Modern genotyping platforms permit a systematic search for inherited components of complex diseases. We performed a joint analysis of two genomewide association studies of coronary artery disease.

METHODS

We first identified chromosomal loci that were strongly associated with coronary artery disease in the Wellcome Trust Case Control Consortium (WTCCC) study (which involved 1926 case subjects with coronary artery disease and 2938 controls) and looked for replication in the German MI (Myocardial Infarction) Family Study (which involved 875 case subjects with myocardial infarction and 1644 controls). Data on other single-nucleotide polymorphisms (SNPs) that were significantly associated with coronary artery disease in either study (P<0.001) were then combined to identify additional loci with a high probability of true association. Genotyping in both studies was performed with the use of the GeneChip Human Mapping 500K Array Set (Affymetrix).

RESULTS

Of thousands of chromosomal loci studied, the same locus had the strongest association with coronary artery disease in both the WTCCC and the German studies: chromosome 9p21.3 (SNP, rs1333049) ($P=1.80\times10^{-14}$ and $P=3.40\times10^{-6}$, respectively). Overall, the WTCCC study revealed nine loci that were strongly associated with coronary artery disease ($P<1.2\times10^{-5}$ and less than a 50% chance of being falsely positive). In addition to chromosome 9p21.3, two of these loci were successfully replicated (adjusted P<0.05) in the German study: chromosome 6q25.1 (rs6922269) and chromosome 2q36.3 (rs2943634). The combined analysis of the two studies identified four additional loci significantly associated with coronary artery disease ($P<1.3\times10^{-6}$) and a high probability (>80%) of a true association: chromosomes 1p13.3 (rs599839), 1q41 (rs17465637), 10q11.21 (rs501120), and 15q22.33 (rs17228212).

CONCLUSIONS

We identified several genetic loci that, individually and in aggregate, substantially affect the risk of development of coronary artery disease.

From the University of Leicester, Leicester (N.J.S., M.M., R.J.D., P.B., S.E.S., H.P., M.D.T., J.R.T.); University of Leeds, Leeds (A.S.H., J.H.B., M.M.I., A.J.B., S.G.B.); University of Cambridge and National Health Service Blood and Transplant, Cambridge (W.O.); and the Wellcome Trust Sanger Institute, Hinxton (P.D.) — all in the United Kingdom; Universität zu Lübeck, Lübeck (J.E., B.M., I.R.K., S.S., F.P., W.L., I.B., A.Z., H.S.); Universität Regensburg, Regensburg (C.H., M.F., A.B.); GSF–Nationales Forschungszentrum für Umwelt und Gesundheit, Neuherberg (T.M., H.-E.W., T.M.S., C.G.); Technische Universität München, Munich (T.M.); Ludwig Maximilians University, Munich (H.-E.W., C.G.); and Johannes Gutenberg University Mainz, Mainz (S.B.) — all in Germany; and INSERM, UMRS525, Université Pierre et Marie Curie, Paris (D.-A.T., F.C.). Address reprint requests to Dr. Samani at the Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, United Kingdom, or at nja@le.ac.uk or to Dr. Schunkert at Medizinische Klinik II, Universität zu Lübeck, 23538 Lübeck, Germany, or at heribert.schunkert@innere2.uni-luebeck.de.

*Members of the Wellcome Trust Case Control Consortium (WTCCC) and the Cardiogenics Consortium are listed in the Supplementary Appendix, available with the full text of this article at www.nejm.org.

This article [10.1056/NEJMoa072366] was published at www.nejm.org on July 18, 2007.

N Engl J Med 2007;357.

Copyright © 2007 Massachusetts Medical Society.

nature genetics

Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants

Myocardial Infarction Genetics Consortium*

Myocardial Infarction Genetics Consortium (MIGen)

GWAS + Replication, Early MI
12,713 Cases : 12,812 Controls

Stage I: 1.0 million SNPs (Affy 6.0) in 6 cohorts
2967 Cases : 3075 Controls
All SNPs with p<0.001

↓

Stage II:
1433 SNPs in 4 studies

↓

Stage III:
25 SNPs in 6 studies:

↓

Stage IV, 13 SNPs

MIGen Consortium GWAS Study*:

- **Unknown:** ***9p21, PHACTR1, KCNE2***
- **LDL:** ***SORT1, MIA3, LDLR, PCSK9, WDR12***
- **Inflammation:** ***CXCL12***
- **CNV:** **No associations noted**

Sister GWAS Studies:

- **LDL:** ***MRAS, HNF1A****
- **Cell Adhesion:** ***SH2B3*****

Total of Twelve Strongly Replicated MI Loci

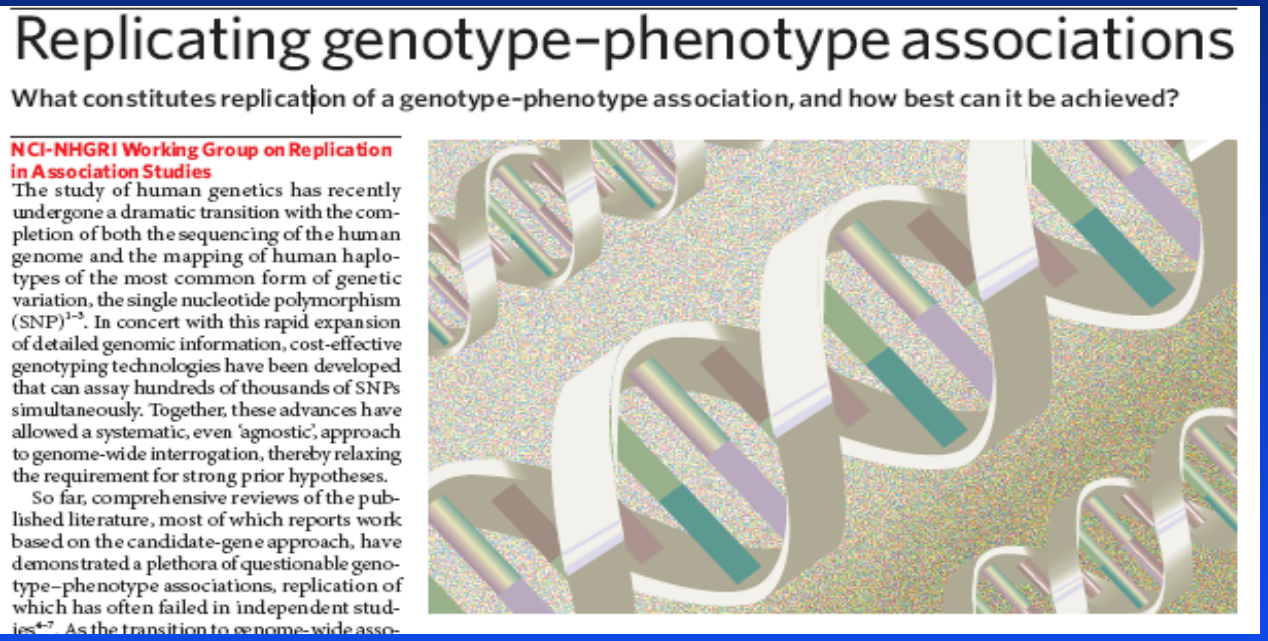
*MIGen Consortium. Nat Gen 2009;41:334. **Erdmann et al. Nat Gen 2009;41:280. ***Soranzo et al. Nat Gen 2009;41:1182.

Pre-Genome Genetics: Weakly Replicated Candidate Gene Variants for CHD

Gene	Variant	Relative Risk	N of Studies	N of Cases / Controls
Apo E	ε4/ε4	1.42	48	15,492 / 32,965
MTHFR	C677T	1.21	72	12,193 / 11,945
ACE	D/I	1.21	19	2,848 / 10,256
Apolipo B	Ins/Del	1.19	22	6,007 / 5,609
PAI-1	4G/5G	1.20	10	1,515 / 1,866
Fibrinogen B	G-455A	0.68	4	745 / 816
ENOS	Intron-4	1.34	16	6,212 / 6,737

Adapted from Ginsburg G et al. JACC 200546:1615.

Criteria for Replicating Genotype-Phenotype Associations in GWAS



Replicating genotype–phenotype associations

What constitutes replication of a genotype–phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies

The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)[1–3]. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even 'agnostic', approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype–phenotype associations, replication of which has often failed in independent studies[4–7]. As the transition to genome-wide asso-

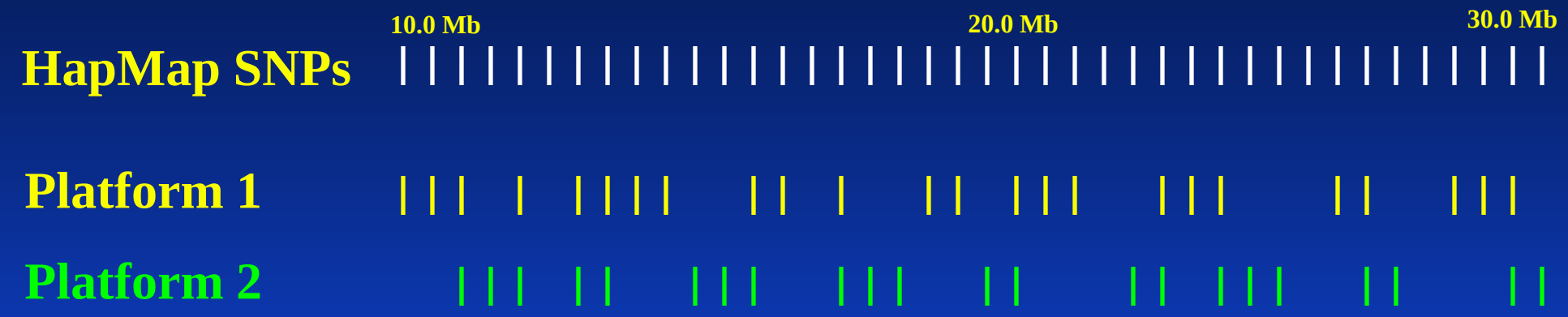
- Sufficiently large N
- Independent data sets not split one study into two
- Same phenotype
- Similar population
- Similar magnitude of effect & significance, same direction of effect
- Same genetic model
- Joint or combined analysis→ smaller p-value than 1st report
- Detailed description of study design and analysis in replications reports

Chanock SJ, Manolio T et al. NCI-NHGRI Working Group on Replication in Association Studies. Nature 2007;447:655-60.

Post-Genome Genetics: Strongly Replicated Genetic Variants for CHD

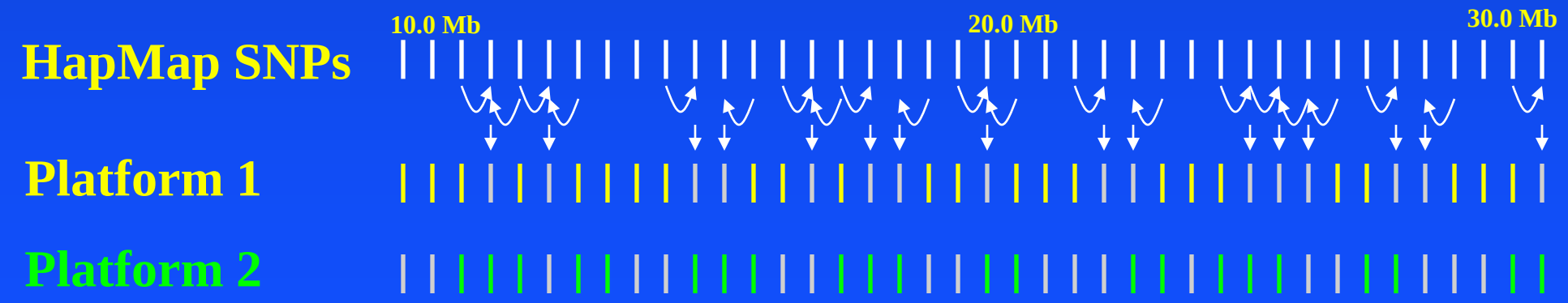
Gene/Locus	SNP	Relative Risk	N of Studies	N of Cases / Controls
CDKN2B, 9p21	rs4977574	1.29	7+	>20,000 / 20,000
PHACTR1, 6p24	Rs12526453	1.12	1	>12,000 / 12,000
KCNE2, 21q22	rs9982601	1.20	1	>12,000 / 12,000
SORT1, 1p13	rs646776	1.19	2	>12,000 / 12,000
MIA3, 1q41	rs17465637	1.14	2	>12,000 / 12,000
LDLR, 19p13	rs1122608	1.15	2+	>12,000 / 12,000
PCSK9, 1p32	rs11206510	1.15	2+	>12,000 / 12,000
WDR12, 2q33	rs6725887	1.17	1	>12,000 / 12,000
MRAS, 3q22	rs9818870	1.15	1	>19,000 / 21,000
CXCL12, 10q11	rs1746048	1.17	2+	>12,000 / 12,000

In Silico GWAS Meta-Analysis Using HapMap LD: Cross-Array Imputation

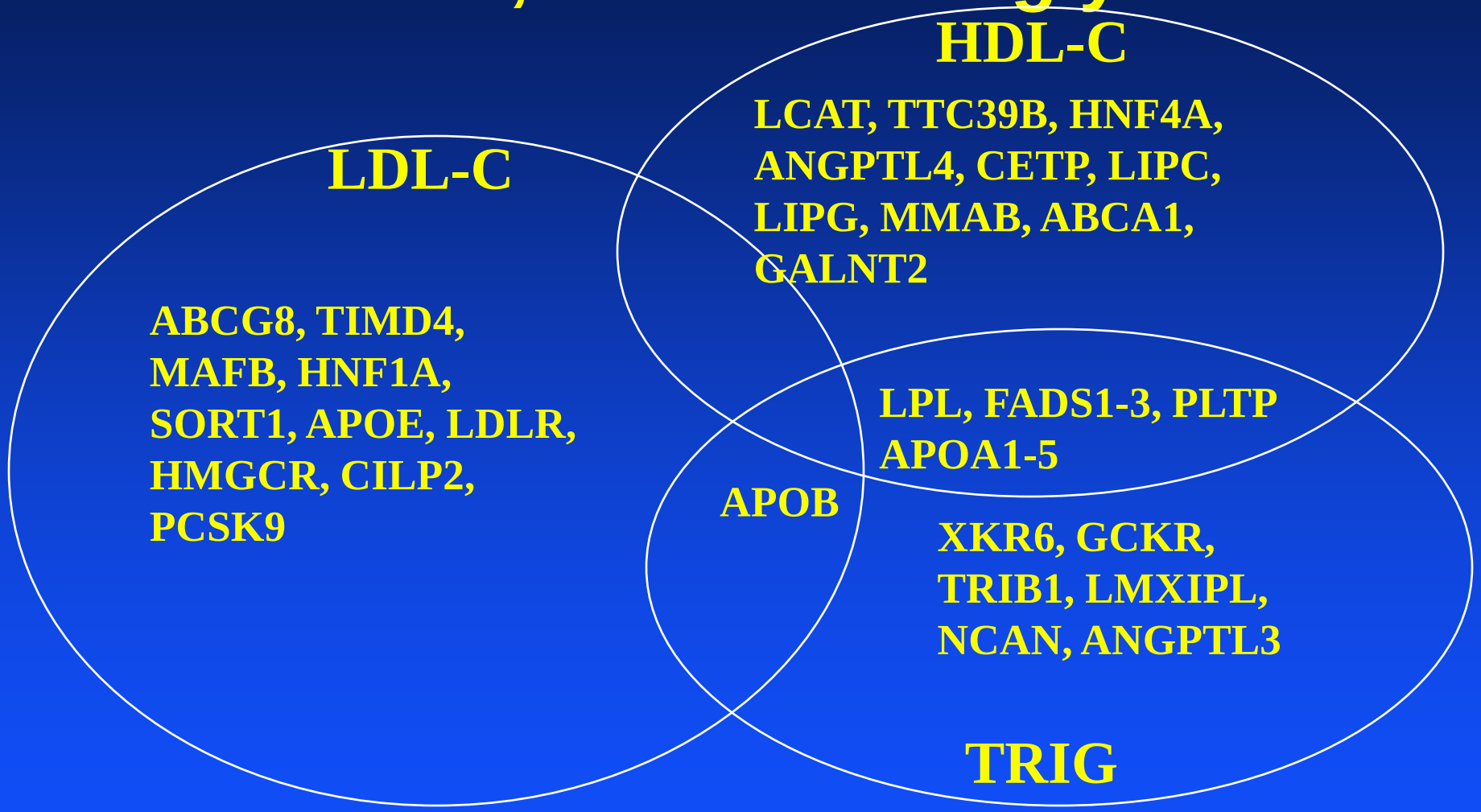


Option 1: Intersecting SNPs:

Option 2: Imputed SNPs—HapMap to Impute Frequency of SNPs:



30 Genomewide Significant Loci Modify LDL-C, HDL-C or Triglycerides



Kathiresan et al. Nature Genetics 2008;40:189-197. Kathiresan et al. Nature Genetics 2009;41:56-65.

Diabetes Mellitus GWAS 2009: Multiple Loci to Date

Sciencexpress Report

Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes

Eleftheria Zeggini,[1,2*] Michael N. Weedon,[3,4*] Cecilia M. Lindgren,[1,2*] Timothy M. Frayling,[3,4*] Katherine S. Elliott,[2] Hana Lango,[3,4] Nicholas J. Timpson,[2,5] John R. B. Perry,[3,4] Nigel W. Rayner,[1,2]

A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Laura J. Scott,[1] Karen L. Mohlke,[2] Lori L. Bonnycastle,[3] Cristen J. Willer,[1] Yun Li,[1] William L. Duren,[1] Michael R. Erdos,[3] Heather M. Stringham,[1] Peter S. Chines,[3] Anne U. Jackson,[1] Ludmila Prokunina-Olsson,[3] Chia-Jen Ding,[1] Amy J. Swift,[3] Narisu Narisu,[3] Tianle Hu,[1] Randall Pruim,[4] Rui Xiao,[1] Xiao-Yi Karen N. Conneely,[1] Nancy L. Riebow,[3] Andrew G. Sprau,[3] Maurine Tong,[3] Peggy P. White,[1] Kurt N. Hetrick,[5] Michael W. Barnhart,[5] Craig W. Bark,[5] Janet L. Goldstein,[5] Lee Watkins,[5] Fang Xiang,[1] Jouko Saramies,[6] Thomas A. Buchanan,[7] Richard M. Watanabe,[8,9] Timo T. Valle,[10] Leena Kinnunen,[10,11] Gong

Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels

Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes for BioMedical Research*†

Locus	SNP	Chr	Risk Allele	OR per allele	P value
NOTCH2	rs10923931	1	T	1.13	6.7x10-6
THADA	rs7578597	2	T	1.15	1.1x10-4
BCL11A	rs10490072	2	T	1.05	4.9x10-5
IGF2BP2	rs1470579	3	C	1.14	8.9x10-16
PPARg	rs1801282	3	C	1.14	1.7x10-6
ADAMTS9	rs4607103	3	C	1.09	3.1x10-4
CDKAL1	rs7754840	6	C	1.12	4.1x10-11
VEGFA	rs9472138	6	T	1.06	4.1x10-5
JAFZ1	rs864745	7	T	1.10	4.6x105
SLC30A8	rs13266634	8	C	1.12	5.3x10-8
CDKN2B	rs10811661	9	T	1.20	7.8x10-15
TCF7L2	rs7903146	10	T	1.37	1.0x10-48
CDC123,CAMK1D	rs12779790	10	G	1.11	4.7x10-5
HHEX	rs1111875	10	C	1.13	5.7x10-10
KCNJ11	rs5219	11	T	1.14	6.7x10-11
TSPAN8,LGR5	rs7961581	12	C	1.09	3.7x10-5
DCD	rs1153188	12	A	1.08	3.2x10-5

Subsequent Studies: Zeggini E et al.Meta-analysis of GWAS and large-scale replication identifies additional susceptibility loci for AODM. Nat Genet. 2008 May;40(5):638-45.

GWAS for Blood Pressure and HTN

					CHARGE meta-analysis results				Global BPgen meta-analysis results			CHARGE + Global BPgen meta-analysis		
SNP identifier	Chr	Position	Nearest gene	Alleles (coded/other)	Freq. coded allele	Beta	s.e.	*P* value	Beta	s.e.	*P* value	Beta	s.e.	*P* value
Systolic blood pressure														
rs12046278	1	10,722,164	*CASZ1*	T/C	0.64	−0.84	0.18	1.84E−06	−0.29	0.15	5.71E−02	−0.53	0.12	4.77E−06
rs7571613	2	190,513,907	*PMS1*	A/G	0.82	−0.96	0.19	7.24E−07	−0.23	0.16	1.59E−01	−0.54	0.13	1.90E−05
rs448378	3	170,583,593	*MDS1*	A/G	0.52	−0.71	0.15	1.28E−06	−0.36	0.13	4.76E−03	−0.51	0.10	1.18E−07
rs2736376	8	11,155,175	*MTMR9*	C/G	0.13	−1.08	0.23	1.90E−06	−0.06	0.19	7.36E−01	−0.48	0.15	9.15E−04
rs1910252	8	49,569,915	*EFCAB1*	T/C	0.18	−0.93	0.19	1.70E−06	−0.07	0.17	6.80E−01	−0.43	0.13	6.13E−04
rs11014166	10	18,748,804	*CACNB2*	A/T	0.66	0.74	0.16	2.11E−06	0.33	0.13	1.31E−02	0.50	0.10	7.03E−07
rs1004467	**10**	**104,584,497**	***CYP17A1***	**A/G**	**0.90**	**1.20**	**0.25**	**1.99E−06**	**0.94**	**0.21**	**1.08E−05**	**1.05**	**0.16**	**1.28E−10**
rs381815	**11**	**16,858,844**	***PLEKHA7***	**T/C**	**0.26**	**0.84**	**0.17**	**5.76E−07**	**0.52**	**0.14**	**2.72E−04**	**0.65**	**0.11**	**1.89E−09**
rs2681492	**12**	**88,537,220**	***ATP2B1***	**T/C**	**0.80**	**1.26**	**0.19**	**3.01E−11**	**0.50**	**0.17**	**4.07E−03**	**0.85**	**0.13**	**3.76E−11**
rs3184504	**12**	**110,368,991**	***SH2B3***	**T/C**	**0.48**	**0.75**	**0.15**	**5.73E−07**	**0.45**	**0.13**	**6.36E−04**	**0.58**	**0.10**	**4.52E−09**
Diastolic blood pressure														
rs13423988	2	68,764,770	*GPR73-ARHGAP25*	T/C	0.17	0.59	0.12	1.09E−06	0.11	0.11	3.22E−01	0.33	0.08	5.00E−05
rs13401889	2	190,618,804	*MSTN*	T/C	0.79	−0.54	0.11	9.70E−07	−0.10	0.11	3.55E−01	−0.31	0.08	4.82E−05
rs9815354	**3**	**41,887,655**	***ULK4***	**A/G**	**0.17**	**0.60**	**0.12**	**7.80E−07**	**0.40**	**0.11**	**3.79E−04**	**0.49**	**0.08**	**2.54E−09**
rs7016759	8	49,574,969	*EFCAB1*	T/C	0.83	0.57	0.12	1.87E−06	0.06	0.11	5.79E−01	0.30	0.08	2.29E−04
rs11014166	**10**	**18,748,804**	***CACNB2***	**A/T**	**0.66**	**0.46**	**0.09**	**8.69E−07**	**0.28**	**0.09**	**1.46E−03**	**0.37**	**0.06**	**1.24E−08**
rs11024074	11	16,873,795	*PLEKHA7*	T/C	0.72	−0.50	0.10	2.83E−07	−0.17	0.09	6.82E−02	−0.33	0.07	1.20E−06
rs2681472	**12**	**88,533,090**	***ATP2B1***	**A/G**	**0.83**	**0.64**	**0.12**	**3.74E−08**	**0.36**	**0.12**	**2.43E−03**	**0.50**	**0.08**	**1.47E−09**
rs3184504	**12**	**110,368,991**	***SH2B3***	**T/C**	**0.48**	**0.50**	**0.09**	**1.68E−08**	**0.45**	**0.09**	**2.83E−07**	**0.48**	**0.06**	**2.58E−14**
rs2384550	**12**	**113,837,114**	***TBX3-TBX5***	**A/G**	**0.35**	**−0.48**	**0.09**	**1.32E−07**	**−0.23**	**0.09**	**1.06E−02**	**−0.35**	**0.06**	**3.75E−08**
rs6495122	**15**	**72,912,698**	***CSK-ULK3***	**A/C**	**0.42**	**0.45**	**0.09**	**8.05E−07**	**0.35**	**0.09**	**3.98E−05**	**0.40**	**0.06**	**1.84E−10**
Hypertension														
rs17806132	2	190,416,532	*PMS1*	A/G	0.16	0.14	0.03	1.14E−05	0.04	0.04	2.87E−01	0.10	0.02	4.70E−05
rs305489	3	11,986,163	*SYN2*	A/G	0.55	0.10	0.02	4.20E−06	0.01	0.03	7.75E−01	0.06	0.02	1.70E−04
rs7640747	3	37,571,809	*ITGA9*	C/G	0.62	−0.12	0.02	4.81E−07	−0.02	0.03	4.12E−01	−0.08	0.02	1.24E−05
rs11775334	8	10,109,030	*MSRA*	A/G	0.32	0.10	0.02	1.03E−05	0.05	0.03	5.86E−02	0.08	0.02	4.05E−06
rs899364	8	11,366,954	*FAM167A-BLK*	T/G	0.32	−0.10	0.02	6.95E−06	−0.04	0.03	1.32E−01	−0.08	0.02	1.01E−05
rs11014166	10	18,748,804	*CACNB2*	A/T	0.66	0.11	0.02	7.79E−07	0.07	0.03	1.06E−02	0.09	0.02	5.72E−08
rs2681472	**12**	**88,533,090**	***ATP2B1***	**A/G**	**0.83**	**0.16**	**0.03**	**1.65E−08**	**0.13**	**0.04**	**2.15E−04**	**0.15**	**0.02**	**1.75E−11**
rs278126	12	118,620,100	*CIT*	T/G	0.28	0.11	0.02	8.34E−06	−0.01	0.03	6.72E−01	0.06	0.02	1.74E−03
rs11612893	12	129,290,572	*FZD10-PIWIL1*	T/G	0.10	−0.19	0.04	7.62E−06	−0.04	0.06	4.19E−01	−0.14	0.03	5.50E−05
rs16982520	20	57,192,115	*ZNF831-EDN3*	A/G	0.88	−0.14	0.03	4.95E−06	−0.11	0.04	7.44E−03	−0.13	0.02	1.59E−07

SNPs in boldface attained $P < 5 \times 10^{-8}$ in meta-analysis of CHARGE and Global BPgen.

Levy D et al. Nature May 10 2009.

GWAS for Other CHD Risk Factors

- Nicotine Abuse
  - Nature 2008;452:638.
- Obesity and BMI
  - Willer et al. Nat Genetics 2009;41:25.
  - Thorleifsson G et al. Nat Genetics 2009;41:18.
  - Loos RJ et al. Nat Genetics 2008;40:768.
  - Frayling T et al. Science 2007;316:889.
- C-Reactive Protein
  - Ridker et al. AJHG 2008;82:1185 and 2008;82:1193.

Observations on GWAS: June 2009

- IT WORKS: N of strongly replicated novel loci:
  - 2006: Handful;
  - 2007: Dozens;
  - 2008-09: Hundreds*
- For CHD/MI and its risk factors—T2DM, lipids, cigarette use, CRP (and obesity/BMI)—exceeding100 novel loci
- Case-control → Prospective population cohorts
- Small effect sizes (RR's 1.1-1.3)→need large N's
- Evolving replication paradigm: "in silico" using other GWAS data, research by consortia
- Mostly Caucasian data—we must study other ethnicities
- Only captures common variants (MAF>5%)
- Defining the functional variant of replicated GWAS variants is essential but highly incomplete
- Applications for clinical practice are largely untested

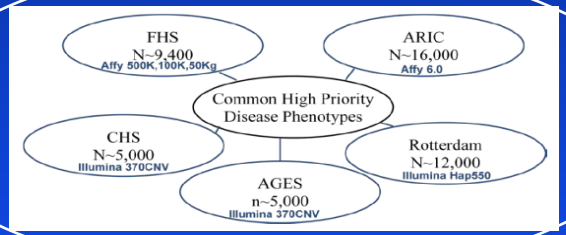
***As of February 2, 2010, 492 GWAS papers and 2317 SNPs reported; see http://www.genome.gov/26525384**

Consortia of Consortia*

GIANT Consortium: **BMI, Obesity-Related Traits**

Cardiogenics Consortium: **Myocardial Infarction**

Global BP Consortium: **BP, HTN-Related Traits**



MAGIC Consortium: **T2DM and Glycemia**

Global Lipids: LDL, HDL, Triglycerides

TAG Consortium: **Cigarette Smoking**

***"Metabochip"**

Genomic Studies and Public Health

- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

Extensive, Prospective Phenotypes in Multiple NHLBI Population Cohorts

Framingham SHARe GWA Study
N~10,000 Subjects
MESA +Women's Health Initiative
+Pulmonary SHARe GWAS
N>25,000 Subjects

Hyperlipidemia
- HDL
- LDL
- Triglycerides
- Subfractions

Ancillary Traits
- Bone density
- Cancer
- Dementia

CVD Outcomes
- CHD, stroke
- Heart failure
- Atrial fibrillation

LV Hypertrophy And Dilation
- LVH and LV size
- Valve disease
- LA size

Hypertension
- Systolic BP
- Diastolic BP
- Pulse pressure

Thrombosis, Inflammation
- CRP, IL-6
- Fibrinogen
- tPA and PAI-1

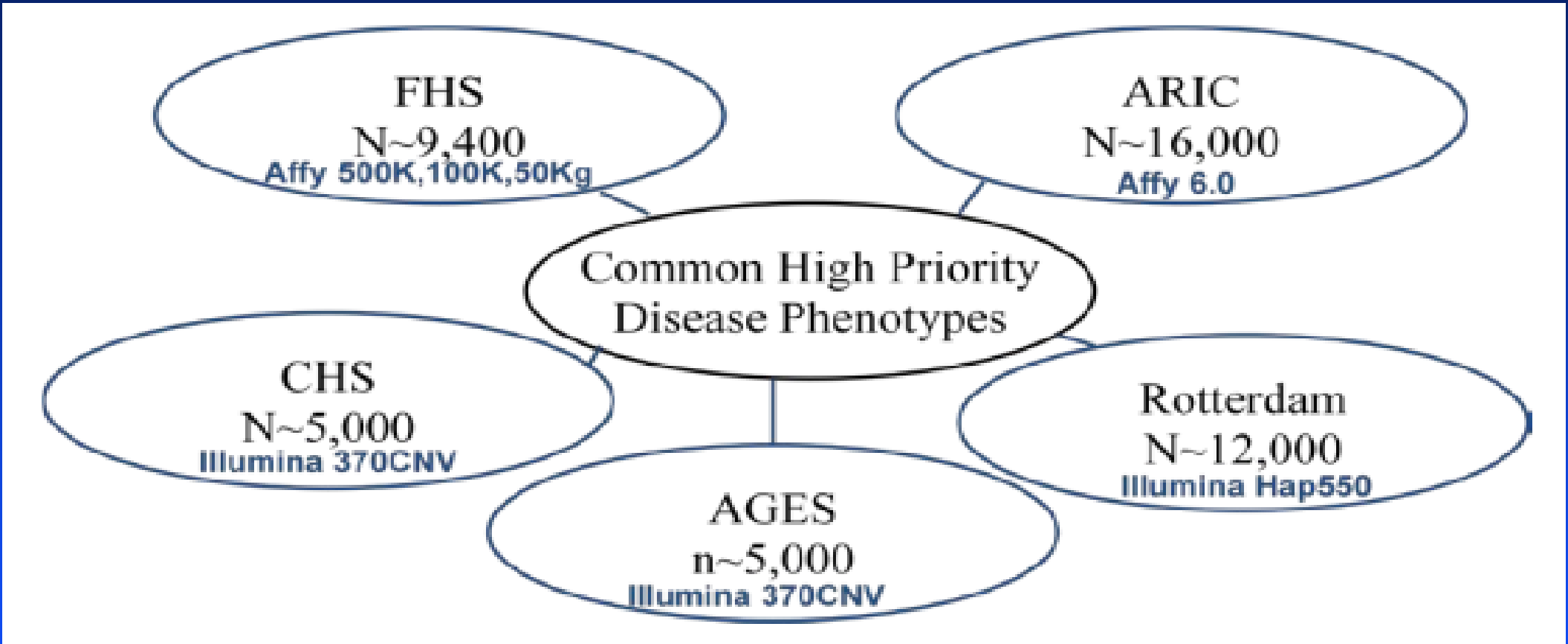
Diabetes, Metabolic Syndrome
- Fasting BS
- Insulin
- Glucose Tolerance

Obesity
- BMI
- Waist-hip
- Visceral fat
- Adiponectin

Pulmonary
- FEV1
- FVC
- Apnea

Subclinical Atherosclerosis
- Carotid IMT
- Vascular calcium
- MRI Plaque

CHARGE (Cohorts for Heart & Aging Research in Genome Epidemiology) Consortium, N~40,000



Affiliated cohorts:

GENOA, KORA, Family Heart, Health ABC, 1958 BC, Guttenberg Heart

***Psaty BM, O'Donnell CJ, et al. Circulation CV Genetics 2009.**

CHARGE: Phenotypes for GWAS in Selected Working Groups

- Coronary heart disease/MI
- Stroke*
- ECG QT Interval*
- Heart Failure
- Chronic kidney disease*
- Biomarkers:CRP, Fibrinogen*
- Adiposity*
- Hypertension/SBP/DBP*
- Menopause*
- Diabetes/glycemic status*
- Subclinical athero: CAC, carotid IMT, ABI
- ECHO*
- Aging
- Neurology traits
- Atrial fibrillation
- Bone density

…and many more

***Manuscripts published or in press**

Translation From Genomics Studies to Personalized Genomic Medicine

2009 Publications

nature genetics
Common variants at ten loci influence QT interval duration in the QTGEN Study

nature genetics
Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants

nature genetics
Variants in *ZFHX3* are associated with atrial fibrillation in individuals of European ancestry

nature genetics
Common variants at ten loci modulate the QT interval duration in the QTSCD Study

ORIGINAL CONTRIBUTION
JAMA®
Genetic Variants Associated With Cardiac Structure and Function

nature genetics
New susceptibility locus for coronary artery disease on chromosome 3q22.3

We present a three-stage analysis of genome-wide SNP data in 1,222 German individuals with myocardial infarction and 1,298 controls, *in silico* replication in three additional genome-wide datasets of coronary artery disease (CAD) and subsequent replication in ~25,000 subjects. We identified one new CAD

ORIGINAL CONTRIBUTION
JAMA®
Association of Cytochrome P450 2C19 Genotype With the Antiplatelet Effect and Clinical Efficacy

The NEW ENGLAND JOURNAL of MEDICINE
ORIGINAL ARTICLE
Genomewide Association Studies of Stroke

nature genetics
Variants in *MTNR1B* influence fasting glucose levels

nature genetics
Common variants at 30 loci contribute to polygenic dyslipidemia

nature genetics
Genome-wide association study identifies eight loci associated with blood pressure

nature genetics
Genome-wide association study of blood pressure and hypertension

nature genetics
Multiple loci associated with indices of renal function and chronic kidney disease

nature genetics
Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium

nature genetics
Genome-wide association studies identify loci associated with age at menarche and age at natural menopause

ARTICLE
Genome-wide Association Analysis Identifies *PDE4D* as an Asthma-Susceptibility Gene

NHLBI-Supported Minority Cohorts with GWAS

- CHARGE African American
  - ARIC
  - CHS
- CARe African American:
  - ARIC
  - CARDIA
  - Jackson Heart Study
  - MESA
- SHARe MultiEthnic
  - MESA
  - Women's Health Initiative
- STAMPEED African American
  - Hopkins Twins
  - GENOA
  - Family Heart Study

Genomic Studies and Public Health

- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

Finding the missing heritability of complex diseases

Teri A. Manolio[1], Francis S. Collins[2], Nancy J. Cox[3], David B. Goldstein[4], Lucia A. Hindorff[5], David J. Hunter[6], Mark I. McCarthy[7], Erin M. Ramos[5], Lon R. Cardon[8], Aravinda Chakravarti[9], Judy H. Cho[10], Alan E. Guttmacher[1], Augustine Kong[11], Leonid Kruglyak[12], Elaine Mardis[13], Charles N. Rotimi[14], Montgomery Slatkin[15], David Valle[9], Alice S. Whittemore[16], Michael Boehnke[17], Andrew G. Clark[18], Evan E. Eichler[19], Greg Gibson[20], Jonathan L. Haines[21], Trudy F. C. Mackay[22], Steven A. McCarroll[23] & Peter M. Visscher[24]

Table 1 | Estimates of heritability and number of loci for several complex traits

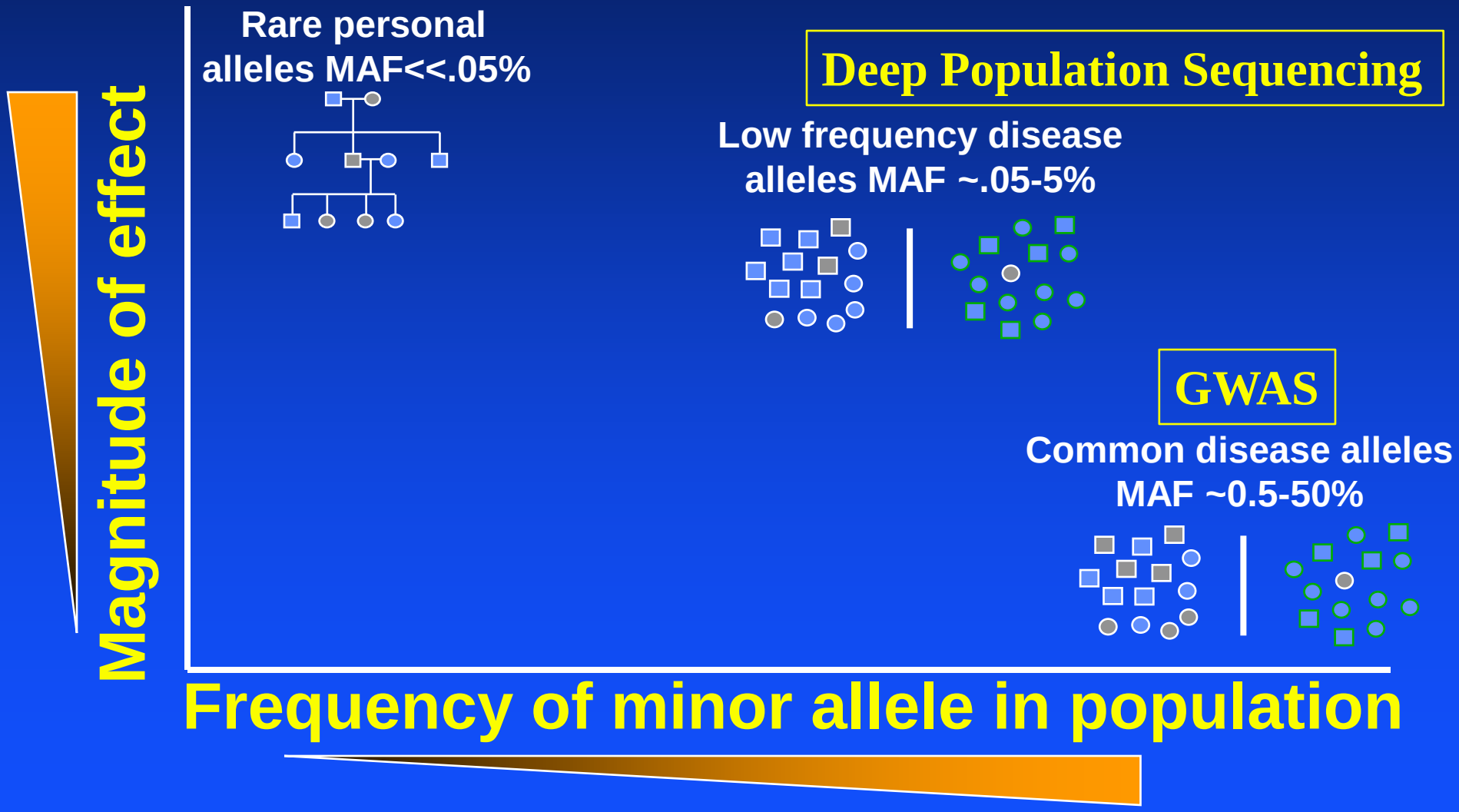
Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration[72]	5	50%	Sibling recurrence risk
Crohn's disease[21]	32	20%	Genetic risk (liability)
Systemic lupus erythematosus[73]	6	15%	Sibling recurrence risk
Type 2 diabetes[74]	18	6%	Sibling recurrence risk
HDL cholesterol[75]	7	5.2%	Residual* phenotypic variance
Height[15]	40	5%	Phenotypic variance
Early onset myocardial infarction[76]	9	2.8%	Phenotypic variance
Fasting glucose[77]	4	1.5%	Phenotypic variance

* Residual is after adjustment for age, gender, diabetes.

Ten Next Steps for GWAS

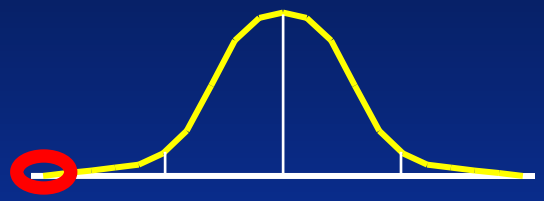
- Missing Variants
- Structural Variants
- Rare Variants: Targeted & Genomewide Sequencing
- Genetic Modifiers
- Epigenetic Effects
- Gene Function, Epistatic Effects
- Prediction and Risk Scores
- Transcription of RNA & siRNA: Profiling and Sequencing
- Pharmacogenomics
- Genome Ethics

Allelic Architecture for the Genetic Determinants of HLB Diseases

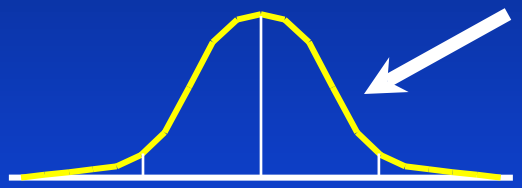


Sequence Variations in PCSK9, Low LDL, and Protection against Coronary Heart Disease

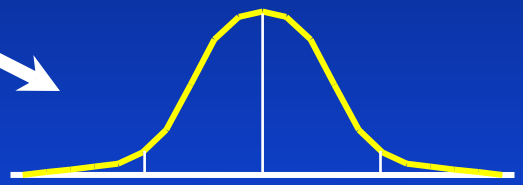
Sequence PCSK9 exons in patients with low LDL in Dallas Heart Study



N=128, Nonsense Mutations Y142X & C679X



N=3363 AA NHLBI ARIC Mutations



N= 9524 White NHLBI ARIC Mutations

	No	Yes (2.6%)
LDL (mg/dL):**	138	~100
CIMT (mm):	0.73	~0.70
CHD (%):*	10%	~1%

	No	Yes (2.6%)
LDL (mg/dL):**	137	116
CIMT (mm):*	0.73	0.71
CHD (%):*	12	6

*p<0.001; **p<0.01

Reference: Cohen JC et al. NEJM 2006;354:1264.

Sequencing Strategies to Define Common and Low Frequency Variants for CVD

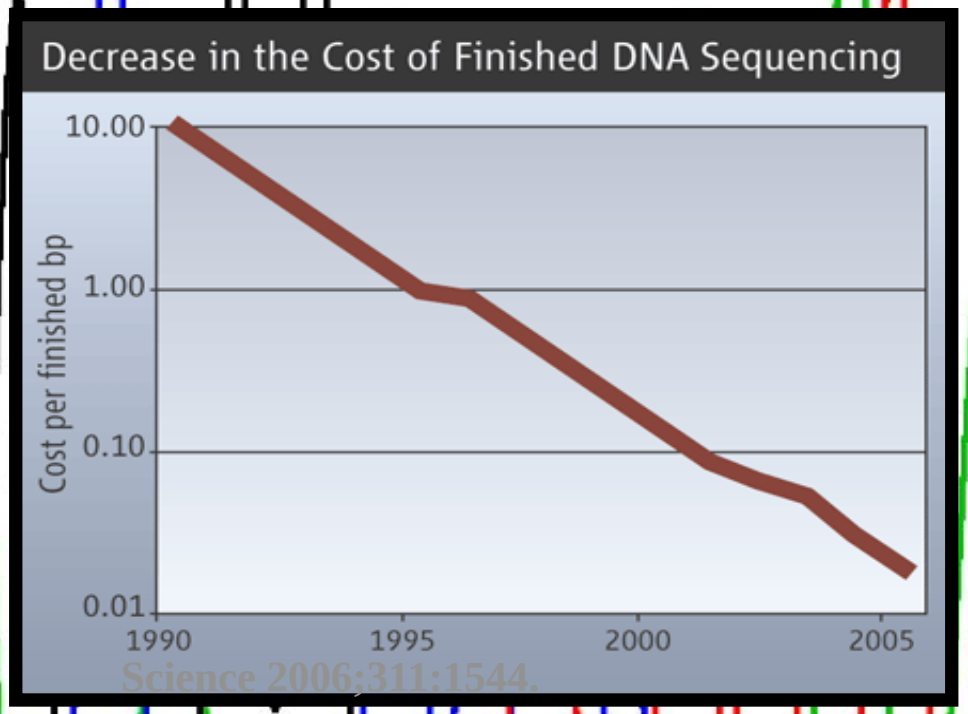
Strategies

- Targeted Resequencing
- Exome Resequencing
- Whole Genome Resequencing

Programs

- NHLBI Resequencing and Genotyping Program
- ENCODE
- 1000 Genomes
- NHLBI Exome Resequencing (RFA HL 08-004)
- CHARGE RC2 Targeted Sequencing of GWAS
- NHLBI Large-Scale (Exome) Sequencing

Decline in DNA Sequencing Costs Towards the “$1000 Genome”



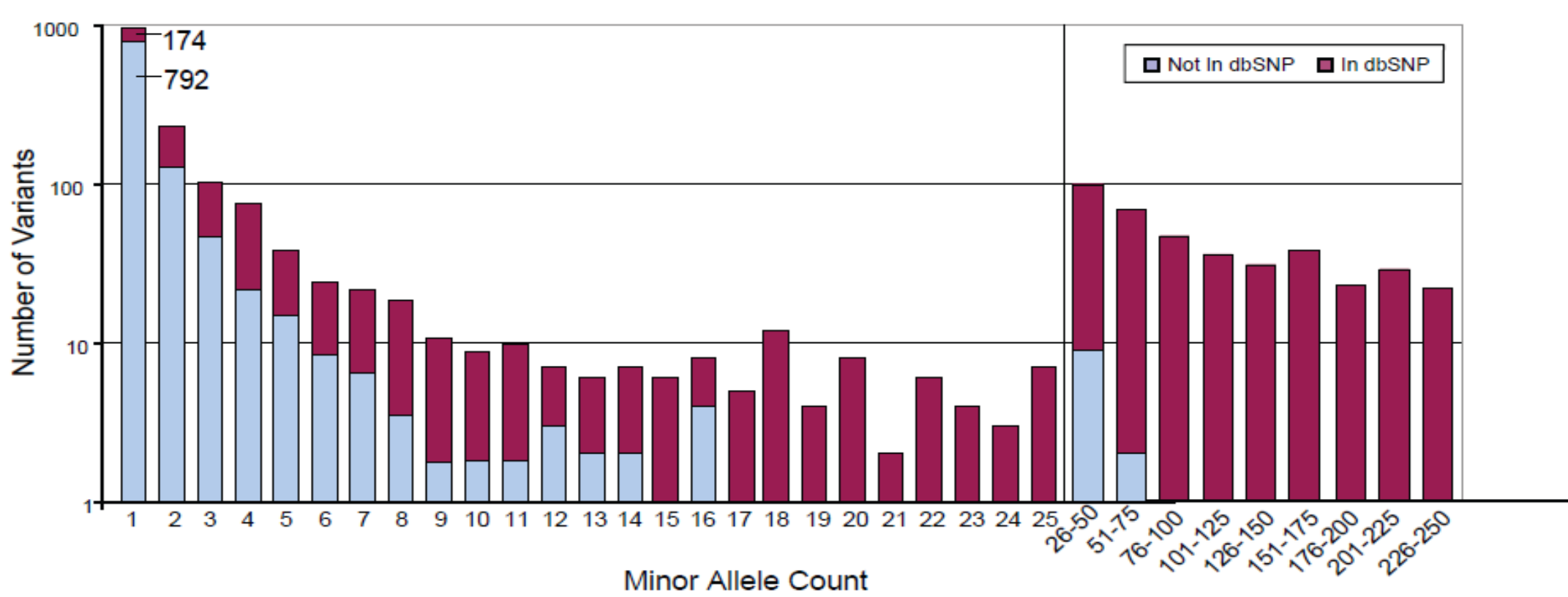
Science 2006;311:1544.

Searching for Cheaper Genome Sequencers

Company	Format	Read Length (bases)	Expected Throughput Mb (million bases)/day
454 Life Sciences	Parallel bead array	100	96
Agencourt Bioscience	Sequencing by ligation	50	200
Applied Biosystems	Capillary electrophoresis	1000	3-4
LI-COR Biosciences	Electronic microchip	20,000	14,000
Microchip Biotechnologies	Parallel bead array	850-1000	7
Network Biosystems	Biochip	800+	5
NimbleGen Systems	Map and survey microarray	30	100
Solexa	Parallel microchip	35	500
VisiGen Biotechnologies	Single-molecule array	NA	1000

ClinSeq CVD Sequencing Program

N of times minor allele detected in sequencing of 219 cardiovascular candidate genes in 250 subjects.



Biesecker L et al. Genome Research July 2009 (online).

Translation From Genomics Studies to Personalized Genomic Medicine

Transcriptomic, Proteomic, Metabalomic

Genomewide: GWAS, Genome Sequencing

↓ ↓

Genome-Phenotype Association

↓

Sequencing & Functional Studies

↓

Genomics-Based Clinical Trials

↓

Applications for Genomic Medicine

↓ ↓ ↓

Predict **Prevent** **Personalize**

The NHLBI Exome Pilot Program

nature genetics

Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng[1,10], Kati J Buckingham[2,10], Choli Lee[1], Abigail W Bigham[3], Holly K Tabor[2,3], Karin M Dent[4], Chad D Huff[5], Paul T Shannon[6], Ethylin Wang Jabs[7,8], Deborah A Nickerson[1], Jay Shendure[1] & Michael J Bamshad[1,2,9]

nature Vol 461|10 September 2009| doi:10.1038/nature08250

LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[3], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

ARRA-Funded Grand Opportunity Genomics Projects

LargeScale DNA (Exome) Sequencing:

- Lung Cohorts
- Myelodysplastic Syndrome
- CVD Cohorts: HeartGO and WHISP

Exome sequence in N~7,500;
Replication in N~many 10,000's

Phase 1 Phenotypes:

Early MI and its Risk Factors
(High/low LDL, T2DM, Obesity)
Rare Pulmonary Diseases

Whole Exome Sequencing Identifies Variants for Miller Syndrome in DHODH*

	Any 1 Exome	Any 2 Exome	Any 3 Exome
Dominant Model			
NS/SS/I**	6943	5167	3920
Predicted damaging	787	126	11
Recessive Model			
NS/SS/I	4293	3172	2329
Predicted damaging	44	4	1***

****NS/SS/I=nonsynonymous variant, splice acceptor or donor site variant or coding indel (NS/SS/I)**
***** *DHODH*; encodes a key enzyme in the pyrimidine de novo biosynthesis pathway**

Mutations in *DHODH* in Discovery and Follow-up Kindreds:

Discovery Kindreds: Kindred 1a G454A, exon 4, and G605C, exon 5; Kindred 2a C403T, exon 3, and C1036T, exon 8; Kindred 3a C595T, exon 5, and 611ΔT, exon 5
Follow-up Kindreds: Kindred 4 G605A, exon 5, and C730T, exon 6; Kindred 5 G56A, exon 2, and C1036T, exon 8; Kindred 6 C851T, exon 7, and A1175G, exon 9

***Ng S et al. Nature Genetics November 13, 2009 advance online.**

Translation From Genomics Studies to Personalized Genomic Medicine

Transcriptomic, Proteomic, Metabalomic

Genomewide: GWAS, Genome Sequencing

↓ ↓

Genome-Phenotype Association

↓

Sequencing & Functional Studies

↓

Genomics-Based Clinical Trials

↓

Applications for Genomic Medicine

↓ ↓ ↓

Predict **Prevent** **Personalize**

The NHLBI Exome Pilot Program

nature genetics

Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng[1,10], Kati J Buckingham[1,10], Choli Lee[1], Abigail W Bigham[2], Holly K Tabor[2,3], Karin M Dent[4], Chad D Huff[5], Paul T Shannon[6], Ethylin Wang Jabs[7,8], Deborah A Nickerson[1], Jay Shendure[1] & Michael J Bamshad[1,2,9]

nature Vol 461|10 September 2009| doi:10.1038/nature08250

LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[3], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

ARRA-Funded Grand Opportunity Genomics Projects

Building on GWAS for NHLBI Disease: Sequencing in the CHARGE Consortium

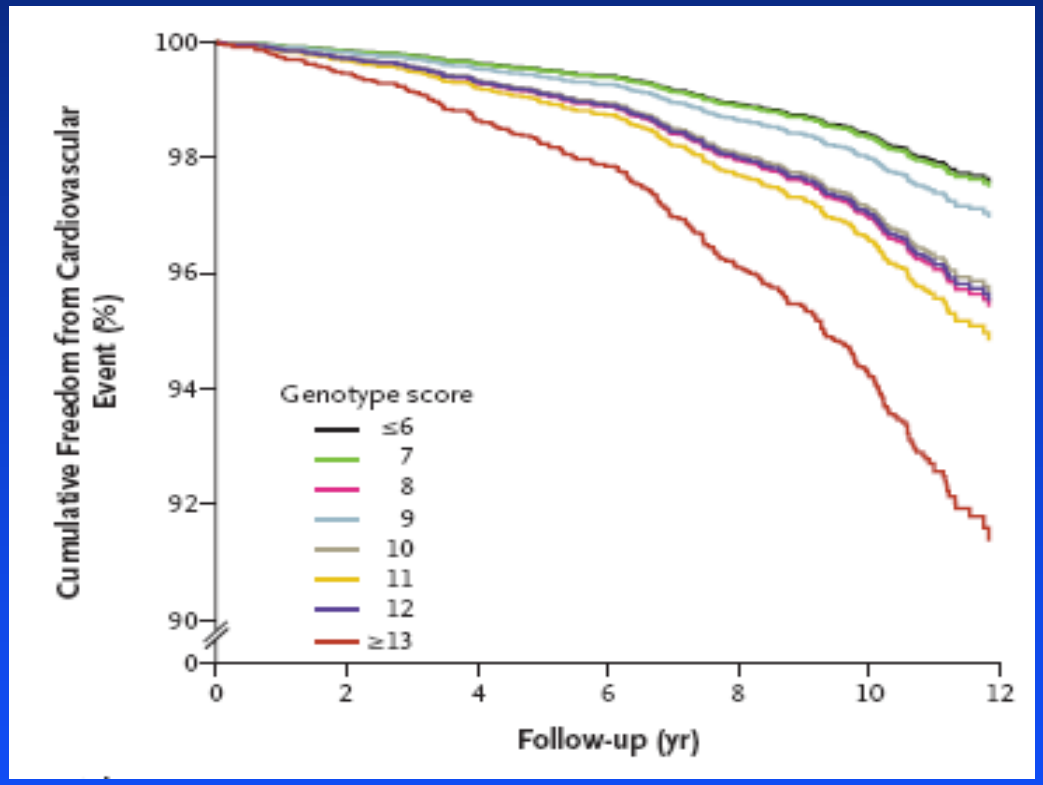
Targeted sequencing of loci implicated in GWAS for each of >=10 CVD Traits and Diseases (eg, hypertension, atrial fibrillation, biomarker levels)

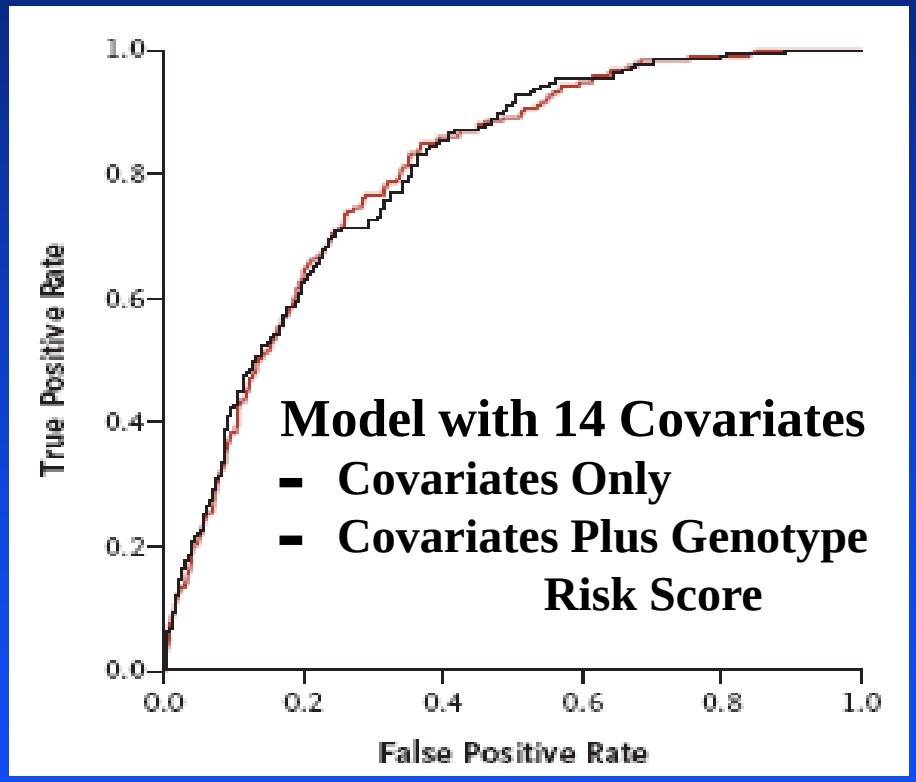
Exome sequencing for traits with multiple (>10) hits by GWAS

Coordination with ESP/HeartGO

Genotype Risk Score for CVD Associations with 9 Lipid SNPs in 5414 Men and Women

Incident MI, Ischemic Stroke or CHD Death





Freedom from Events **Area under ROC Curve**

Kathiresan et al. NEJM 2008;358:1240. SNPs in *ApoE, ApoB, LDLR, PCSK9, HMGCR, CETP, LPL, LIPC,* and *ABCA1.*

AUC Using GWAS SNPs with Liberal P-Values vs Genome-Wide SNPs

	P<0.000001 to P<0.80	Known Variants	Combined
Coronary Heart Disease	**0.56 – 0.60**	**0.57**	**0.58 – 0.62**
Hypertension	**0.52 – 0.61**	**Not Available from WTCCC**	**Not Available from WTCCC**
Type 2 Diabetes	**0.54 – 0.60**	**0.67**	**0.67 – 0.71**
Type 1 Diabetes	0.75 – 0.72	0.78	0.78 – 0.79
Bipolar Disorder	0.53 – 0.67	0.55	0.56 – 0.68
Crohn's Disease	0.61 – 0.61	0.77	0.77 – 0.78
Rheumatoid Arthritis	0.66 – 0.65	0.70	0.70 – 0.72

Evans D et al. Hum Mol Genetics 2009;18:3525.

Caution on SNPs and Risk Prediction

OPEN ACCESS Freely available online PLoS GENETICS

Review

Interpretation of Genetic Association Studies: Markers with Replicated Highly Significant Odds Ratios May Be Poor Classifiers

Johanna Jakobsdottir[1]*, Michael B. Gorin[2], Yvette P. Conley[3,4], Robert E. Ferrell[4], Daniel E. Weeks[1,4]

Genetic Risk Prediction — Are We There Yet?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

Number of Risk Alleles Needed to Produce a Sibling Relative Risk of 1.5, 2.0, or 3.0.*

Relative Risk Per Allele	Sibling Relative Risk		
	1.5	2.0	3.0
		no. of risk alleles	
1.10	203–507	347–867	550–1374
1.20	51–135	87–231	138–367

Association Between a Literature-Based Genetic Risk Score and Cardiovascular Events in Women

Nina P. Paynter, PhD
Daniel I. Chasman, PhD
Guillaume Paré, MD, MS
Julie E. Buring, ScD
Nancy R. Cook, ScD
Joseph P. Miletich, MD, PhD
Paul M Ridker, MD, MPH

Context While multiple genetic markers associated with cardiovascular disease have been identified by genome-wide association studies, their aggregate effect on risk beyond traditional factors is uncertain, particularly among women.

Objective To test the predictive ability of a literature-based genetic risk score for cardiovascular disease.

Design, Setting, and Participants Prospective cohort of 19 313 initially healthy white women in the Women's Genome Health Study followed up over a median of 12.3 years (interquartile range, 11.6-12.8 years). Genetic risk scores were constructed from the National Human Genome Research Institute's catalog of genome-

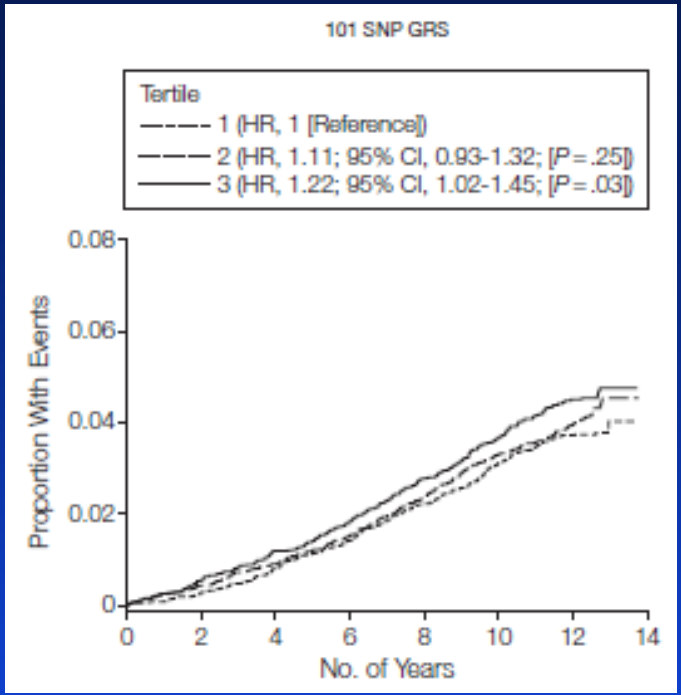
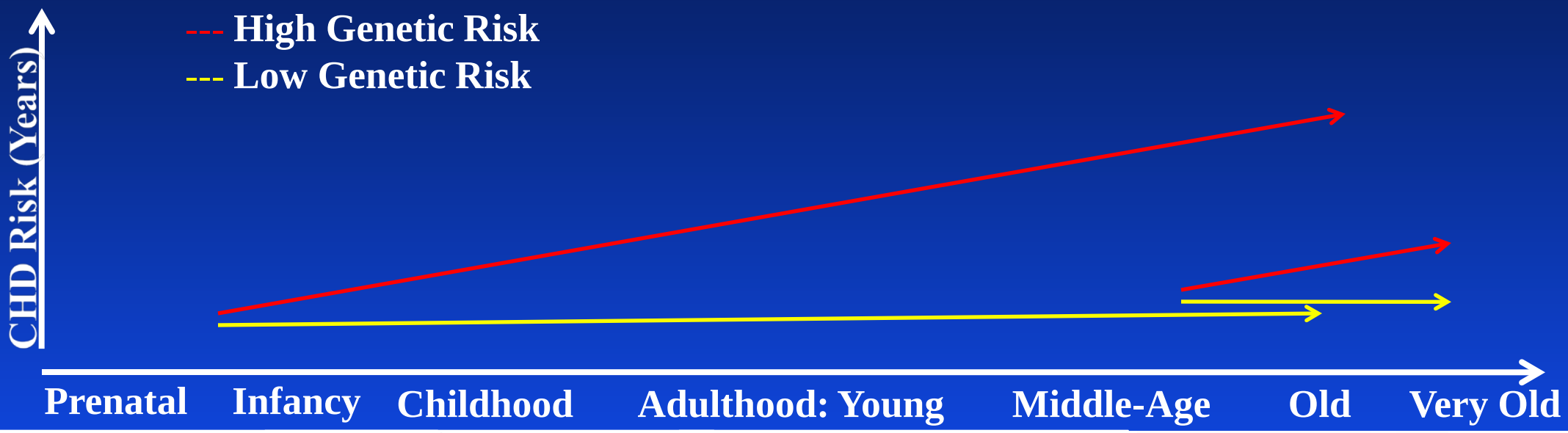


Table 1. Association of Genetic Risk Score (GRS) and Family History of Cardiovascular Disease (CVD)

	101 SNP GRS[a]		12 SNP GRS[b]		Family History of Premature MI	
	HR/Allele (95% CI)	*P* Value	HR/Allele (95% CI)	*P* Value	HR (95% CI)	*P* Value
Age	1.02 (1.00-1.03)	.006	1.05 (1.01-1.09)	.01	1.67 (1.39-1.03)	<.001
Covariates						
ATP III [c]	1.00 (0.99-1.01)	.63	1.04 (1.00-1.08)	.05	1.57 (1.31-1.89)	<.001
Reynolds[d]	1.00 (0.99-1.01)	.76	1.04 (1.00-1.07)	.06	NA	NA

Paynter NP et al. JAMA 2010;303:631-37. SNPs derived from literature; of 101 SNPs, 46 were genotyped directly and 55 were imputed (minimum R2 0.6). Baseline age 53 years with ~12 years followup for CVD.

Atherosclerotic Plaque Development: Lifetime Risk vs Short-Term Late Adult Risk



Time Magazine, Feb 22 '10: www.time.com/time/covers

Providing Genetic Results to Participants: NHLBI Working Groups I and II

Working Group I Criteria for Notification*:

1. Established analytic validity
2. Disease risk is replicable and significant
   - E.g. RR > 2, high penetrance, younger onset
3. Important health implications
   - E.g. premature death or substantial morbidity

Working Group II**: Considered the impact of GWAS, genomewide resequencing, growing N of risk variants and evidence base on risk/benefit of notification

***Bookman. *Am J Med Genet* 2006;140A:1033. **January 2009; report in preparation.**

Personalized Genomics

It's All About Me

Along with the flood of discoveries in human genetics, 2007 saw the birth of a new industry: personal genomics. Depending on your budget, you can either buy a rough scan of your genome or have the whole thing sequenced. The companies say the information will help customers learn about themselves and improve their health. But researchers worry that these services open up a Pandora's box of ethical issues.

At $300,000 to $1 million per genome, sequencing all 3 billion base pairs is still too costly for all but a few. Although dozens more personal genomes will probably be sequenced in the coming year, most will be done by public and private research organizations—including the institute run by genome maverick J. Craig Venter, whose personal genome was one of three completed in 2007 in the United States and China. In a lower-budget effort, Harvard's George Church this month will deliver initial DNA sequences for the protein-coding sections (1% of the genome) to the first 10 volunteers for his Personal Genome Project. Meanwhile, a new company called Knome is offering full-genome sequencing to 20 customers willing to pay $350,000.

A glimpse of one's genome is already within the reach of ordinary people, thanks to several companies. They include 23andMe, which has financing from Google and may let users link to others with shared traits; Navigenics, which will screen for about 20 medical conditions; and deCODE Genetics in Iceland, a pioneer in disease gene hunting. For $1000 to $2500, these companies will have consumers send in a saliva sample or cheek swab, then use "SNP chips" to scan their DNA for as many as 1 million markers. The companies will then match the results with the latest publications on traits, common diseases, and ancestry.

Although many customers may view this exercise as a way to learn fun facts about themselves—recreational genomics, some call it—bioethicists are wary. Most common disease markers identified so far raise risks only slightly, but they could cause needless worry. At the same time, some people may be terrified to learn they have a relatively high risk for an incurable disease such as Alzheimer's.

Pandora's box? This cheek-swab kit could reveal your intimate secrets.

The rush toward personal genome sequences also sharpens long-held worries about discrimination. A bill to prevent insurers and employers from misusing genetic data is stalled in Congress. Complicating matters, your genetic information exposes your relatives' DNA, too.

The most profound implications of having one's genome analyzed may not be what it reveals now—which isn't much—but what it may show later on. Perhaps to sidestep such questions, some companies will limit which markers to disclose. Others, however, will hand customers their entire genetic identity, along with all the secrets it may hold.

—JOCELYN KAISER

- Several commercial companies are offering to scan a person's genome using genomewide SNP testing (cost $985-2500)
- These companies include 23andMe, DeCode Genetics, Navigenics (NY Times Nov 17, 2007) and Lumigenix
- Information about specific diseases as well as the entire genome is made available
- Clinical utility (risks, benefits, costs to patient, costs to healthcare system) of such testing is unknown

Debate Regarding Utility of GWAS SNPs for Personal Medicine

Genetic risk information for common diseases may indeed be already useful for prevention and early detection

J. Gulcher and K. Stefansson

deCODE Genetics, Reykjavik, Iceland

Eur J Clin Invest 2010; 40 (1): 56–63.

Personal genomics: information can be harmful

D. F. Ransohoff* and M. J. Khoury†

*Departments of Medicine and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, †Office of

Eur J Clin Invest 2010; 40 (1): 64–68.

Genomic Studies and Public Health

- The Risk Factor Paradigm
- Rationale for Genome Studies for CVD & Other Diseases
- What we Have Learned from CVD Genomewide Association Studies (GWAS)
- Population-Based GWAS Programs
- Genomewide Sequencing and Next Steps Post-GWAS
- Translating GWAS to Personalized Medicine

Translation From Genomics Studies to Personalized Genomic Medicine

Transcriptomic, Proteomic, Metabalomic | **Genomewide: GWAS, Genome Sequencing**

↓ ↓

Genome-Phenotype Association

↓

Sequencing & Functional Studies

↓

Genomics-Based Clinical Trials

↓

Applications for Genomic Medicine

↓ ↓ ↓

Predict	**Prevent**	**Personalize**
CVD Incidence & Recurrence	**CVD Incidence & Recurrence**	**CVD PharmGen & Treatments**

- Biology: Initial legacy of GWAS and GWSeq
- Patience: 50 Years for Risk Factor Paradigm
- Time Horizon: Decades
- Prospective cohorts across lifespan
- iPS & nanotechnology
- Incremental Knowledge
  - Bioinformatics
  - New clinical paradigm

Current Evidence-Based Approaches to Predict, Prevent, and Pre-empt CVD

Clinical Risk Factor or CVD	Evidence-Based Testing and Treatment Strategy: Predict CVD	Prevent CVD*	Pre-Empt Recurrent CVD*
Hypertension	Office or Ambulatory Blood Pressure	Diuretic, Beta-blockers, ACE-inhibitors, ARBs	Beta-blockers, ACE-inhibitors, ARBs
Hyperlipidemia	Lipid Panel	Statins, Resins, Fibrates	Statins
Diabetes	Fasting BS, HgA1C	Treat Diabetes??	Treat Diabetes??
Coagulation and Inflammation	Measure CRP? Other Biomarkers?	Aspirin or Plavix	Aspirin or Plavix
Subclinical vascular disease	ABI? CAC? IMT?	Intensify Therapy?	NA
Heart Attack, Stroke, & other CVDs	Multiple Risk Factor Algorithm	Rx Hypertension, Hypercholesterolemia, quit smoking	Rx Hypertension, Use Statin, Use ACE-inhibitor, quit smoking, Use ASA or Plavix

***Lifestyle modification is essential: diet, exercise, weight loss**

Current Evidence-Based Approaches to Predict, Prevent, and Pre-empt CVD

Clinical Risk Factor or CVD	Evidence-Based Testing and Treatment Strategy: Predict CVD	Prevent CVD*	Pre-Empt Recurrent CVD*
Hypertension	Office or Ambulatory Blood Pressure **Biomarker Testing**	Diuretic, Beta-blockers, ACE-inhibitors, ARBs **Biomarker Stratification**	Beta-blockers, ACE-inhibitors, ARBs **Biomarker Stratification**
Hyperlipidemia	Lipid Panel **Biomarker Testing**	Statins, Resins, Fibrates **Biomarker Stratification**	Statins **Biomarker Stratification**
Diabetes	Fasting BS, HgA1C **Biomarker Testing**	Treat Diabetes?? **Biomarker Stratification**	Treat Diabetes?? **Biomarker Stratification**
Coagulation and Inflammation	Measure CRP? Other Biomarkers? **Biomarker Testing**	Aspirin or Plavix **Biomarker Stratification**	Aspirin or Plavix **Biomarker Stratification**
Subclinical vascular disease	ABI? CAC? IMT? **Biomarker Testing**	Intensify Therapy? **Biomarker stratification**	NA
Heart Attack, Stroke, & other CVDs	Multiple Risk Factor Algorithm **Global Predictive** **Biomarker Testing**	Rx Hypertension, Hypercholesterolemia, quit smoking **Biomarker Stratification**	Rx Hypertension, Use Statin, Use ACE-inhibitor, quit smoking, Use ASA or Plavix **Biomarker Stratification**

***Lifestyle modification is essential: diet, exercise, weight loss**

Acknowledgements

- NHLBI's Framingham Heart Study
  - BU-NHLBI Fellows: Andrew Johnson, Jie Huang, George Thanassoulis, Wolfgang Lieb, Sarah Rosner Preis
  - BU-NHLBI FHS Investigators: Larry Atwood, L.A. Cupples, J. Murabito C. Fox, R. D'Agostino, PA. Wolf, MG. Larson, D. Levy, E. Benjamin, V. Ramachandran, LA. Cupples, SJ. Hwang, Q. Yang, K. Lunetta, J. Dupuis, A. DeStefano,
  - FHS SHARe Collaborators
- MIGen Consortium
- CHARGe Consortium
  - NHLBI CHS
  - NHLBI ARIC
  - Iceland and NIA AGES
  - Netherlands Rotterdam Study
  - Other European and US Cohorts
- NHLBI
  - C. Jaquish, R. Fabsitz, D. Paltoo, P. Sholinsky, J. Rossouw, G. Papanicalou, A. Michelson, P. Sorlie, D. Applebaum-Bowden, W. Gan
- NCBI
  - J. Ostell, S. Sherry, M. Feolo
- Framingham Heart Study Participants

National Heart Lung and Blood Institute
People Science Health

NATIONAL INSTITUTES OF HEALTH