Introduction to Add Health GWAS Data Part I

Christy Avery

Department of Epidemiology

University of North Carolina at Chapel Hill

Outline

- Introduction to genome-wide association studies (GWAS)
- Research enabled by GWAS
- Obtaining Add Health data
- Further considerations

Genetics: Difficult to Escape

Genetics: Difficult to Escape

Genetic Research Definition: Research into the cause, transmission, amelioration, elimination, or enhancement of inherited disorders and traits.

“Inherited Disorders” Encompasses a Broad Spectrum of Diseases and Traits

Molecular Psychiatry (2014) 19, 41

www.nature.com/mp

IMMEDIATE COMMUNICATION

Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci

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

LETTER

doi:10.1038/nature17671

Genome-wide association study identifies 74 loci associated with educational attainment

Defining the role of common variation in the genomic and biological architecture of adult human height

Using genome-wide data from 253,288 individuals, we identified 697 variants at genome-wide significance that together explained one-fifth of the heritability for adult height. By testing different numbers of variants in independent studies, we show that the most strongly associated ~2,000, ~3,700 and ~9,500 SNPs explained ~21%, ~24% and ~29% of phenotypic variance. Furthermore, all common variants together captured 60% of heritability. The 697 variants clustered in 423 loci were enriched for genes, pathways and tissue types known to be involved in growth and together implicated genes and pathways not highlighted in earlier efforts, such as signaling by fibroblast growth factors, WNT/β-catenin and chondroitin sulfate–related genes. We identified several genes and pathways not previously connected with human skeletal growth, including mTOR, osteoglycin and binding of hyaluronic acid. Our results indicate a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants.

OPEN ACCESS Freely available online

PLOS | ONE

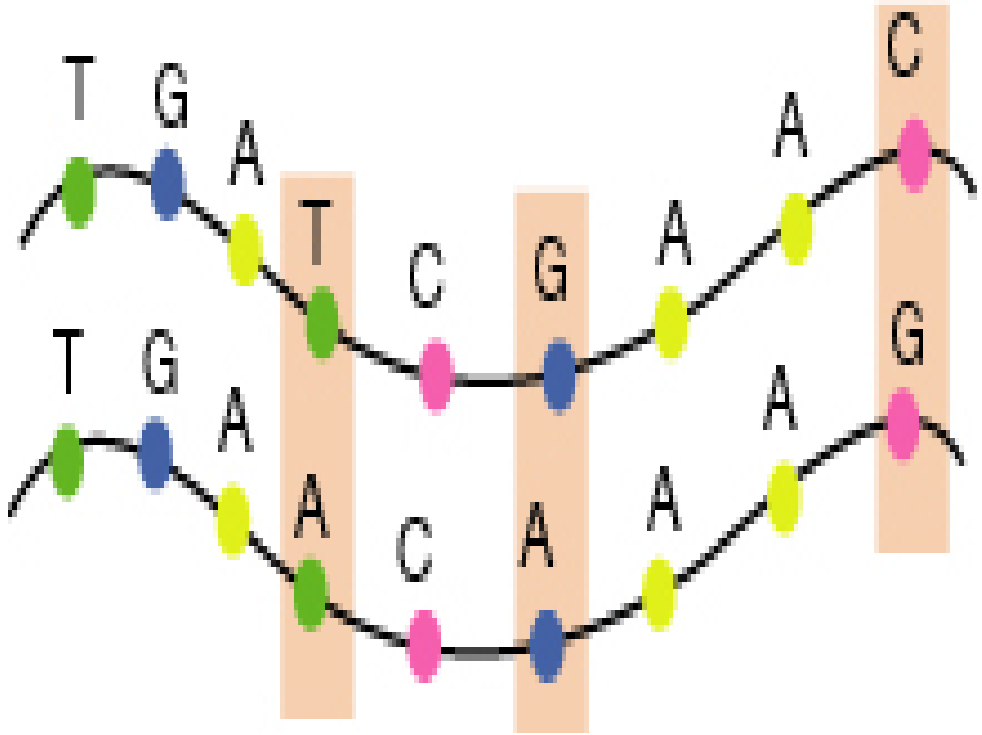
Genome-Wide Association Study of Proneness to Anger

Definition: Genome-Wide Association Study (GWAS)

- ***One*** of many contemporary tools to evaluate the genetic basis of disease/phenotypes
- Study that surveys ***most*** of the genome for genetic causal variants.
- Capitalizes on the strengths of association studies without having to guess the identity of candidate genes.
- Enables testing of multiple, genome-wide (~40 million) variants without any prior hypothesis (other than the trait is heritable)
- GWAS genetic metric: the SNP

Hirschhorn JN et al. (2005) Nature Reviews Genetics 6: 95-100.

Single Nucleotide Polymorphisms (SNPs)



- Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide (A,T,C,or G) in the genome sequence is altered
- Millions of SNPs in the genome!

LETTER

doi:10.1038/nature17671

Genome-wide association study identifies 74 loci associated with educational attainment

A list of authors and their affiliations appears in the online version of the paper.

Educational attainment is strongly influenced by social and other environmental factors, but genetic factors are estimated to account for at least 20% of the variation across individuals[1]. Here we report the results of a genome-wide association study (GWAS) for educational attainment that extends our earlier discovery sample[1,2] of 101,069 individuals to 293,723 individuals, and a replication study in an independent sample of 111,349 individuals from the UK Biobank. We identify 74 genome-wide significant loci associated with the number of years of schooling completed. Single-nucleotide polymorphisms associated with educational attainment are disproportionately found in genomic regions regulating gene expression in the fetal brain. Candidate genes are preferentially expressed in neural tissue, especially during the prenatal period, and enriched for biological pathways involved in neural development. Our findings demonstrate that, even for a behavioural phenotype that is mostly environmentally determined, a well-powered GWAS identifies replicable associated genetic variants that suggest biologically relevant pathways. Because educational attainment is measured in large numbers of individuals, it will continue to be useful as a proxy phenotype in efforts to characterize the genetic influences of related phenotypes, including cognition and neuropsychiatric diseases.

Educational attainment is measured in all main analyses as the number of years of schooling completed (EduYears, $n = 293{,}723$, mean = 14.3, s.d. = 3.6; Supplementary Information sections 1.1–1.2). All GWAS were performed at the cohort level in samples restricted to individuals of European descent whose educational attainment was assessed at or above age 30. A uniform set of quality-control procedures was applied to the cohort-level summary statistics. In our GWAS meta-analysis of ~9.3 million SNPs from the 1000 Genomes Project, we used sample-size weighting and applied a single round of genomic control at the cohort level.

Our meta-analysis identified 74 approximately independent genome-wide significant loci. For each locus, we define the 'lead SNP' as the SNP in the genomic region that has the smallest P value (Supplementary Information section 1.6.1). Figure 1 shows a Manhattan plot with the lead SNPs highlighted. This includes the three SNPs that reached genome-wide significance in the discovery stage of our previous GWAS meta-analysis of educational attainment[1]. The quantile–quantile (Q–Q) plot of the meta-analysis (Extended Data Fig. 1) exhibits inflation ($\lambda_{GC} = 1.28$), as expected under polygenicity[3].

Extended Data Fig. 2 shows the estimated effect sizes of the lead SNPs. The estimates range from 0.014 to 0.048 standard deviations per allele (2.7 to 9.0 weeks of schooling), with incremental R^2 in the range 0.01% to 0.035%.

To quantify the amount of population stratification in the GWAS estimates that remains even after the stringent controls used by the cohorts (Supplementary Information section 1.4), we used linkage-disequilibrium (LD) score regression[4]. The regression results indicate that ~8% of the observed inflation in the mean χ^2 is due to bias rather than polygenic signal (Extended Data Fig. 3a), suggesting that stratification effects are small in magnitude. We also found evidence for polygenic association signal in several within-family analyses, although these are not powered for individual SNP association testing (Supplementary Information section 2 and Extended Data Fig. 3b).

To further test the robustness of our findings, we examined the within-sample and out-of-sample replicability of SNPs reaching genome-wide significance (Supplementary Information sections 1.7–1.8). We found that SNPs identified in the previous educational attainment meta-analysis replicated in the new cohorts included here, and conversely, that SNPs reaching genome-wide significance in the new cohorts replicated in the old cohorts. For the out-of-sample replication analyses of our 74 lead SNPs, we used the interim release of the UK Biobank[5] (UKB) ($n = 111{,}349$). As shown in Extended Data Fig. 4,

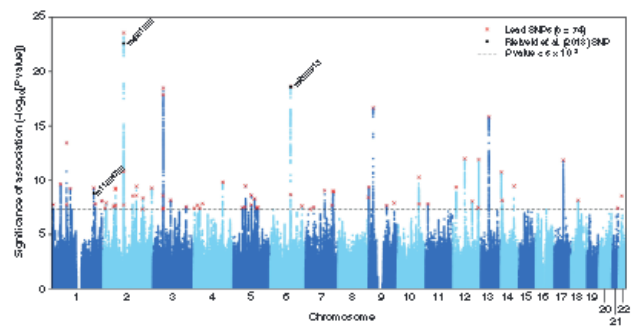


Figure 1 | Manhattan plot for EduYears associations ($n = 293{,}723$). The x axis is chromosomal position, and the y axis is the significance on a $-\log_{10}$ scale (two-tailed test). The black dashed line shows the genome-wide significance level (5×10^{-8}). The red crosses are the 74 approximately independent genome-wide significant associations (lead SNPs). The black dots labelled with rs numbers are the three SNPs identified in ref. 1.

doi:10.1038/nature17671

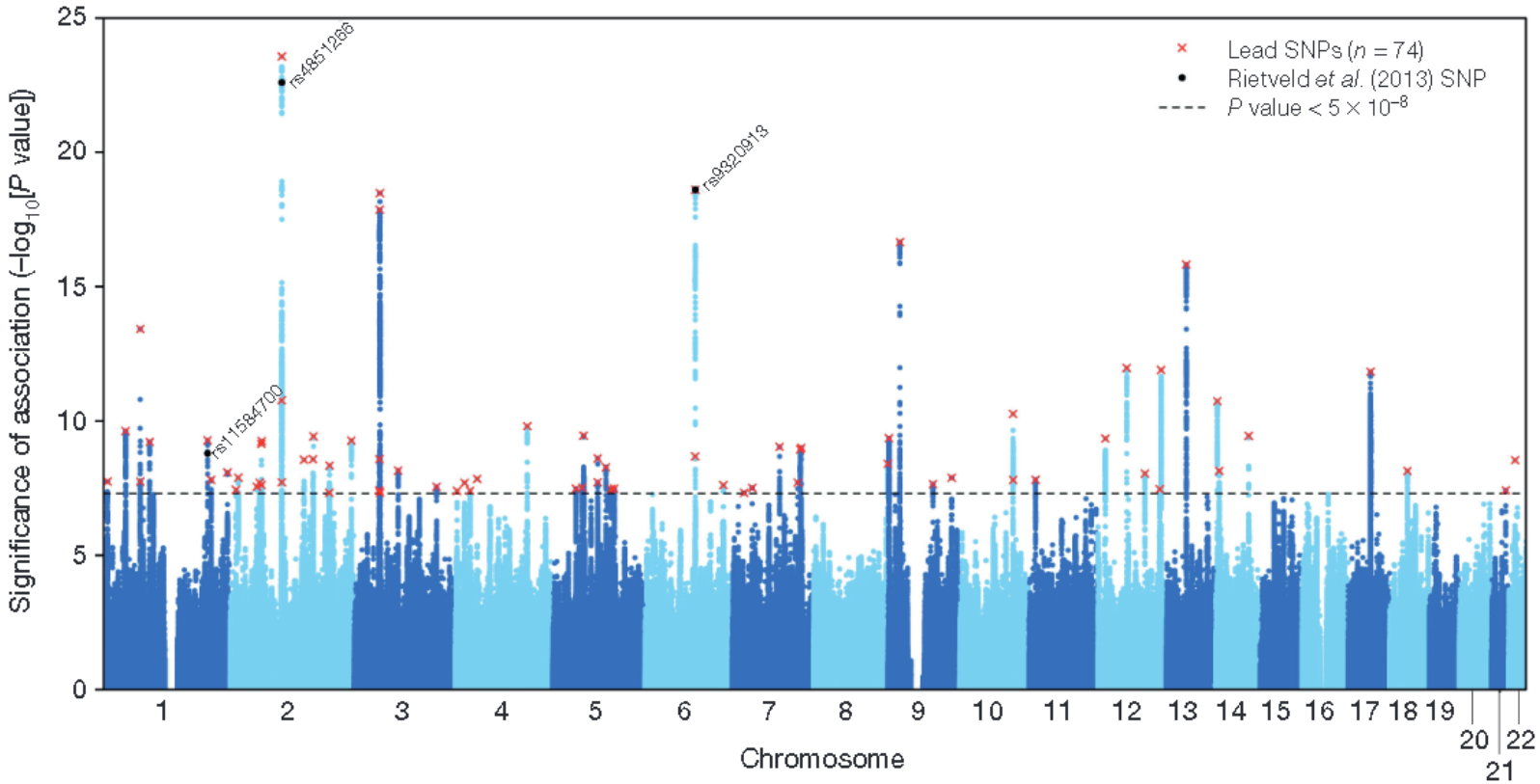


Figure 1 | Manhattan plot for EduYears associations ($n = 293{,}723$). The x axis is chromosomal position, and the y axis is the significance on a $-\log_{10}$ scale (two-tailed test). The black dashed line shows the genome-wide significance level (5×10^{-8}). The red crosses are the 74 approximately independent genome-wide significant associations (lead SNPs). The black dots labelled with rs numbers are the three SNPs identified in ref. 1.

Figure 1 | Manhattan plot for EduYears associations ($n = 293{,}723$). The x axis is chromosomal position, and the y axis is the significance on a $-\log_{10}$ scale (two-tailed test). The black dashed line shows the genome-wide significance level (5×10^{-8}). The red crosses are the 74 approximately independent genome-wide significant associations (lead SNPs). The black dots labelled with rs numbers are the three SNPs identified in ref. 1.

Published GWAS through 01/2016

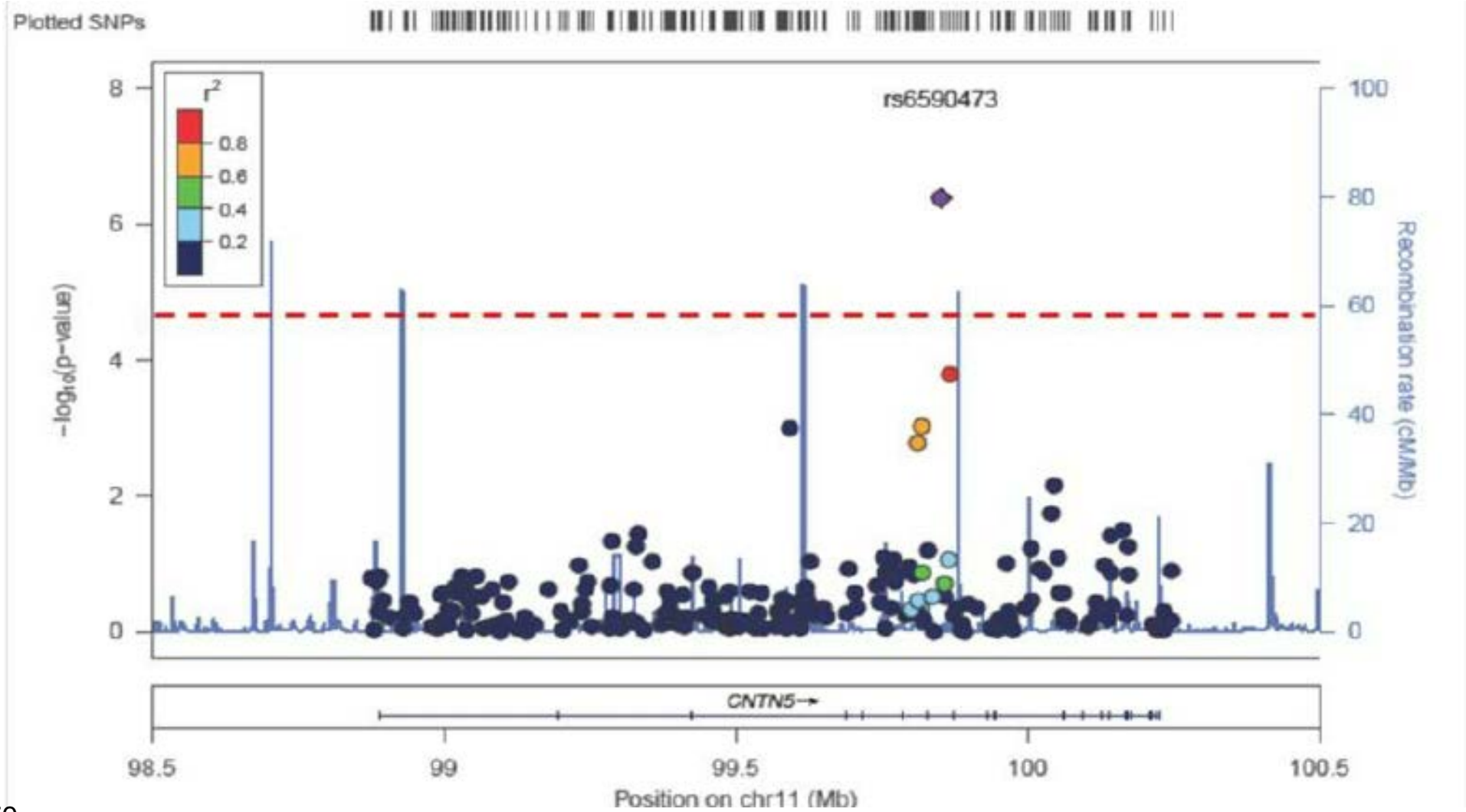
NHGRI-EBI Catalog, http://www.ebi.ac.uk/gwas/

- Abdominal aortic aneurysm
- Acute lymphoblastic leukemia
- Adhesion molecules
- Adverse response to carbamapezine
- Adiponectin levels
- Age-related macular degeneration
- AIDS progression
- Alcohol dependence
- Alopecia areata
- Alzheimer disease
- Amyloid A levels
- Amyotrophic lateral sclerosis
- Angiotensin-converting enzyme activity
- Ankylosing spondylitis
- Arterial stiffness
- Asparagus anosmia
- Asthma
- Atherosclerosis in HIV
- Atrial fibrillation
- Attention deficit hyperactivity disorder
- Autism
- Basal cell cancer
- Behcet's disease
- Bipolar disorder
- Biliary atresia
- Bilirubin
- Bitter taste response
- Birth weight
- Bladder cancer
- Bleomycin sensitivity
- Blond or brown hair
- Blood pressure
- Blue or green eyes
- BMI, waist circumference
- Bone density
- Breast cancer
- C-reactive protein
- Calcium levels
- Cardiac structure/function
- Carnitine levels
- Carotenoid/tocopherol levels
- Celiac disease
- Cerebral atrophy measures
- Chronic lymphocytic leukemia
- Cleft lip/palate
- Cognitive function
- Conduct disorder
- Colorectal cancer
- Corneal thickness
- Coronary disease
- Creutzfeldt-Jakob disease
- Crohn's disease
- Cutaneous nevi
- Dermatitis
- Drug-induced liver injury
- Endometriosis
- Eosinophil count
- Eosinophilic esophagitis
- Erectile dysfunction and prostate cancer treatment
- Erythrocyte parameters
- Esophageal cancer
- Essential tremor
- Exfoliation glaucoma
- Eye color traits
- F cell distribution
- Fibrinogen levels
- Folate pathway vitamins
- Follicular lymphoma
- Fuch's corneal dystrophy
- Freckles and burning
- Gallstones
- Gastric cancer
- Glioma
- Glycemic traits
- Hair color
- Hair morphology
- Handedness in dyslexia
- HDL cholesterol
- Heart failure
- Heart rate
- Height
- Hemostasis parameters
- Hepatic steatosis
- Hepatitis
- Hepatocellular carcinoma
- Hirschsprung's disease
- HIV-1 control
- Hodgkin's lymphoma
- Homocysteine levels
- Hypospadias
- Idiopathic pulmonary fibrosis
- IgA levels
- IgE levels
- Inflammatory bowel disease
- Intracranial aneurysm
- Iris color
- Iron status markers
- Ischemic stroke
- Juvenile idiopathic arthritis
- Keloid
- Kidney stones
- LDL cholesterol
- Leprosy
- Leptin receptor levels
- Liver enzymes
- Longevity
- LP (a) levels
- LpPLA(2) activity and mass
- Lung cancer
- Magnesium levels
- Major mood disorders
- Malaria
- Male pattern baldness
- Matrix metalloproteinase levels
- MCP-1
- Melanoma
- Menarche & menopause
- Meningococcal disease
- Metabolic syndrome
- Migraine
- Moyamoya disease
- Multiple sclerosis
- Myeloproliferative neoplasms
- N-glycan levels
- Narcolepsy
- Nasopharyngeal cancer
- Neuroblastoma
- Nicotine dependence
- Obesity
- Open angle glaucoma
- Open personality
- Optic disc parameters
- Osteoarthritis
- Osteoporosis
- Otosclerosis
- Other metabolic traits
- Ovarian cancer
- Pancreatic cancer
- Pain
- Paget's disease
- Panic disorder
- Parkinson's disease
- Periodontitis
- Peripheral arterial disease
- Phosphatidylcholine levels
- Phosphorus levels
- Photic sneeze
- Phytosterol levels
- Platelet count
- Polycystic ovary syndrome
- Primary biliary cirrhosis
- Primary sclerosing cholangitis
- PR interval
- Progranulin levels
- Prostate cancer
- Protein levels
- PSA levels
- Psoriasis
- Psoriatic arthritis
- Pulmonary funct. COPD
- QRS interval
- QT interval
- Quantitative traits
- Recombination rate
- Red vs.non-red hair
- Refractive error
- Renal cell carcinoma
- Renal function
- Response to antidepressants
- Response to antipsychotic therapy
- Response to hepatitis C treat
- Response to metaformin
- Response to statin therapy
- Restless legs syndrome
- Retinal vascular caliber
- Rheumatoid arthritis
- Ribavirin-induced anemia
- Schizophrenia
- Serum metabolites
- Skin pigmentation
- Smoking behavior
- Speech perception
- Sphingolipid levels
- Statin-induced myopathy
- Stroke
- Systemic lupus erythematosus
- Systemic sclerosis
- T-tau levels
- Tau AB1-42 levels
- Telomere length
- Testicular germ cell tumor
- Thyroid cancer
- Tooth development
- Total cholesterol
- Triglycerides
- Tuberculosis
- Type 1 diabetes
- Type 2 diabetes
- Ulcerative colitis
- Urate
- Venous thromboembolism
- Ventricular conduction
- Vertical cup-disc ratio
- Vitamin B12 levels
- Vitamin D insuffiency
- Vitiligo
- Warfarin dose
- Weight
- White cell count
- YKL-40 levels

SNP Data Enable a Wide Range of Investigations in Addition to Genome-Wide Scans

- Abdominal aortic aneurysm
- Acute lymphoblastic leukemia
- Adhesion molecules
- Adverse response to carbamapezine
- Adiponectin levels
- Age-related macular degeneration
- AIDS progression
- Alcohol dependence
- Alopecia areata
- Alzheimer disease
- Amyloid A levels
- Amyotrophic lateral sclerosis
- Angiotensin-converting enzyme activity
- Ankylosing spondylitis
- Arterial stiffness
- Asparagus anosmia
- Asthma
- Atherosclerosis in HIV
- Atrial fibrillation
- Bladder cancer
- Bleomycin sensitivity
- Blond or brown hair
- Blood pressure
- Blue or green eyes
- BMI, waist circumference
- Bone density
- Breast cancer
- C-reactive protein
- Calcium levels
- Cardiac structure/function
- Carnitine levels
- Carotenoid/tocopherol levels
- Celiac disease
- Cerebral atrophy measures
- Chronic lymphocytic leukemia
- Cleft lip/palate
- Cognitive function
- Conduct disorder
- Colorectal cancer
- Corneal thickness
- Coronary disease
- Creutzfeldt-Jakob disease
- Crohn's disease
- Cutaneous nevi
- Dermatitis
- Drug-induced liver injury
- Endometriosis
- Eosinophil count
- Eosinophilic esophagitis
- Erectile dysfunction and prostate cancer treatment
- Erythrocyte parameters
- Esophageal cancer
- Essential tremor
- Exfoliation glaucoma
- Glioma
- Glycemic traits
- Hair color
- Hair morphology
- Handedness in dyslexia
- HDL cholesterol
- Heart failure
- Heart rate
- Height
- Hemostasis parameters
- Hepatic steatosis
- Hepatitis
- Hepatocellular carcinoma
- Hirschsprung's disease
- HIV-1 control
- Hodgkin's lymphoma
- Homocysteine levels
- Hypospadias
- Idiopathic pulmonary fibrosis
- IgA levels
- IgE levels
- Inflammatory bowel disease
- Intracranial aneurysm
- Iris color
- Iron status markers
- Ischemic stroke
- Juvenile idiopathic arthritis
- Keloid
- Kidney stones
- LDL cholesterol
- Leprosy
- Leptin receptor levels
- Liver enzymes
- Longevity
- LP (a) levels
- Menarche & menopause
- Meningococcal disease
- Metabolic syndrome
- Migraine
- Moyamoya disease
- Multiple sclerosis
- Myeloproliferative neoplasms
- N-glycan levels
- Narcolepsy
- Nasopharyngeal cancer
- Neuroblastoma
- Nicotine dependence
- Obesity
- Open angle glaucoma
- Open personality
- Optic disc parameters
- Osteoarthritis
- Osteoporosis
- Otosclerosis
- Other metabolic traits
- Ovarian cancer
- Pancreatic cancer
- Pain
- Paget's disease
- Panic disorder
- Parkinson's disease
- Periodontitis
- Peripheral arterial disease
- Phosphatidylcholine levels
- Phosphorus levels
- Photic sneeze
- Phytosterol levels
- Platelet count
- Polycystic ovary syndrome
- Primary biliary cirrhosis
- QRS interval
- QT interval
- Quantitative traits
- Recombination rate
- Red vs.non-red hair
- Refractive error
- Renal cell carcinoma
- Renal function
- Response to antidepressants
- Response to antipsychotic therapy
- Response to hepatitis C treat
- Response to metaformin
- Response to statin therapy
- Restless legs syndrome
- Retinal vascular caliber
- Rheumatoid arthritis
- Ribavirin-induced anemia
- Schizophrenia
- Serum metabolites
- Skin pigmentation
- Smoking behavior
- Speech perception
- Sphingolipid levels
- Statin-induced myopathy
- Stroke
- Systemic lupus erythematosus
- Systemic sclerosis
- Type 1 diabetes
- Type 2 diabetes
- Ulcerative colitis
- Urate
- Venous thromboembolism
- Ventricular conduction
- Vertical cup-disc ratio
- Vitamin B12 levels
- Vitamin D insuffiency
- Vitiligo
- Warfarin dose
- Weight
- White cell count
- YKL-40 levels

E.g, Limit Inference to Specific Genes



PMID:25337070

E.g., Polygenic Scores (PGS)

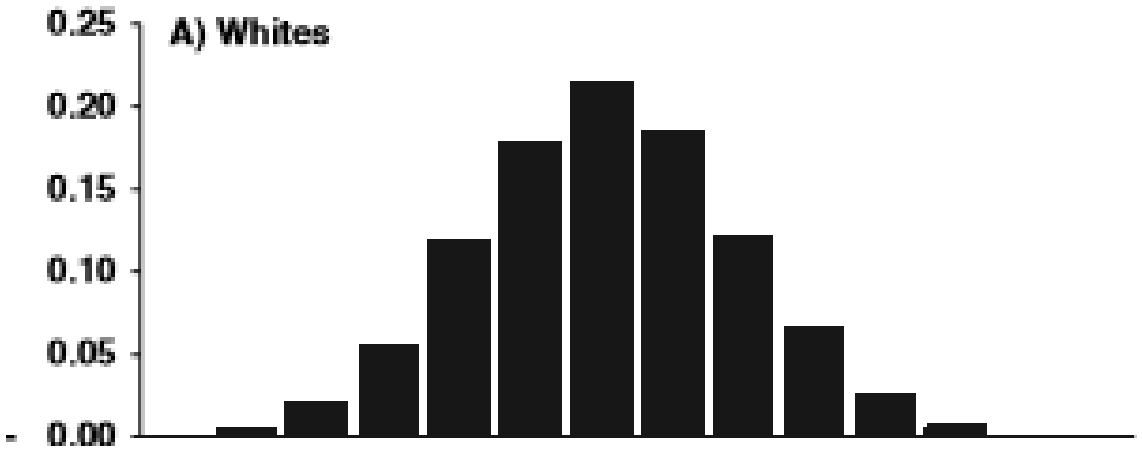
- One-variable summary score constructed from SNPs previously associated with phenotype/disease of interest (i.e. via large GWAS)
- AKA genetic risk score
  - Estimation strategy:
    - Locate dataset with large-scale genotyping and measure of phenotype of interest (e.g. blood pressure in Add Health)
    - Identify published GWAS, including associated lead SNPs and effect estimates
    - Predict participant-specific phenotype (e.g. blood pressure) using participant-specific genotypes and published lead SNP effect estimates

Estimation of a Cardiovascular PGS

Equation 1

$$\text{Cardiovascular genetic risk score} = \sum_i \frac{1}{Odds\ Ratio_{SNPi}} (SNP_i\ \ dosage)$$

where i is the index of SNPs included in Appendix A, Table 2.

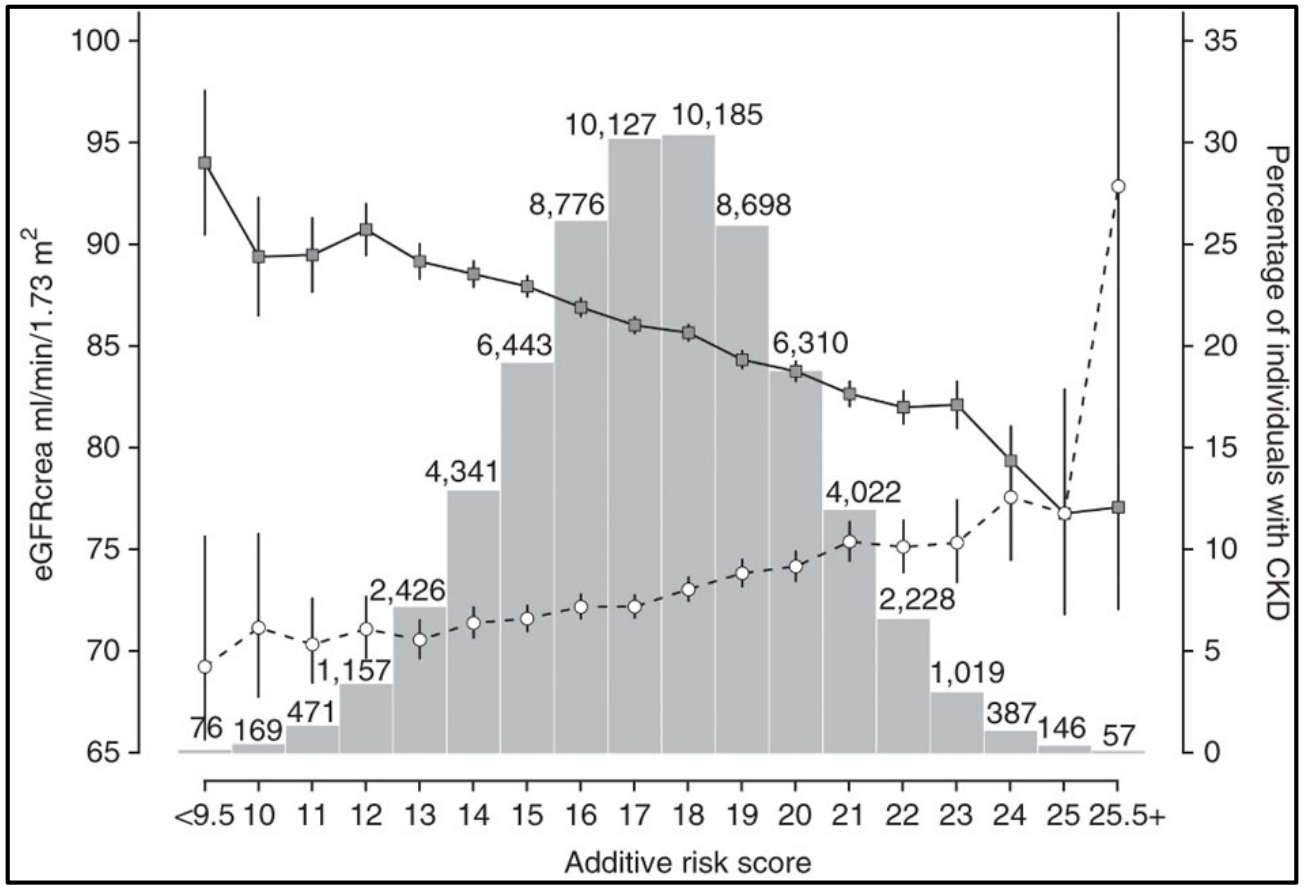


Appendix A, Table 2: Lead SNPs and ORs for CHD used to calculate the cardiovascular genetic risk score (cGRS)[16]

Gene	Lead SNP	Odds Ratio for coronary heart disease	Risk allele
1p13.3 (*SORT1*)	rs646776	1.19	T
1p32.3 (*PPAP2B*)	rs17114036	1.17	A
1p32.3 (*PCSK9*)	rs11206510	1.15	T
1q41 (*MIA3*)	rs17465637	1.14	C
2q33.1 (*WDR12*)	rs6725887	1.17	C
6p21.31 (*ANKS1A*)	rs17609940	1.07	G
6p24.1 (*PHACTR1*)	rs9349379	1.12	G
6q23.2 (*TCF21*)	rs12190287	1.08	C
6q25.3 (*LPA*)	rs3798220	1.47	C
6q25.3 (*LPA*)	rs10455872	1.70	G
7q32.3 (*ZC3HC1*)	rs11556924	1.09	C
9p21.3 (*CDKN2A*)	rs4977574	1.29	G
9q34.2 (*ABO*)	rs9411489	1.10	T
10q11.21 (*CXCL12*)	rs1746048	1.17	C
10q24.32 (*CYP17A1*)	rs12413409	1.12	G
11q23.3 (*APOA5*)	rs964184	1.13	G
12q2.4 (*HNF1A*)	rs2259816	1.08	T
12q24.12 (*SH2B3*)	rs3184504	1.13	T
13q3.4 (*COL4A1*)	rs4773144	1.07	G
14q32.2 (*HHPL1*)	rs2895811	1.07	C
15q25.1 (*ADAMTS7*)	rs3825807	1.08	T
17p11.2 (*RASD1*)	rs12936587	1.07	G
17p13.3 (*SMG6*)	rs216172	1.07	C
17q21.32 (*UBE2Z*)	rs46522	1.06	T
19p13.2 (*LDLR*)	rs1122608	1.15	G
21q22.11 (*KCNE2*)	rs9982601	1.20	T

PGS: Application to Kidney Disease

- One-variable summary score constructed from SNPs associated with phenotype/disease of interest



PMID:20383146

E.g., Polyge

Polygenic risk for coronary artery disease is associated with cognitive ability in older adults

Saskia P. Hagenaars,[1,2,3] Sarah E. Harris,[1,4] Toni-Kim Clarke,[3] Lynsey Hall,[3] Michelle Luciano,[1,2] Ana Maria Fernandez-Pujals,[3] Gail Davies,[1,2] Caroline Hayward,[4] Generation Scotland,[4] John M. Starr,[1,5] David J. Porteous,[1,4] Andrew M. McIntosh[1,3] and Ian J. Deary[1,2]*

[1]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK [2]Department of Psychology, University of Edinburgh, Edinburgh, UK, [3]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK, [4]Institute for Genetics and Molecular Medicine, Western General Hospital, University of Edinburgh, Edinburgh, UK and [5]Geriatric Medicine Unit, University of Edinburgh, Royal Infirmary of Edinburgh, Edinburgh, UK

*Corresponding author. Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. E-mail: i.deary@ed.ac.uk

Accepted 8 December 2015

Abstract

Background: Coronary artery disease (CAD) is associated with cognitive decrements and risk of later dementia, but it is not known if shared genetic factors underlie this association. We tested whether polygenic risk for CAD was associated with cognitive ability in community-dwelling cohorts of middle-aged and older adults.

Methods: Individuals from Generation Scotland: Scottish Family Health Study (GS:SFHS, $N = 9865$) and from the Lothian Birth Cohorts of 1921 (LBC1921, $N=517$) and 1936 (LBC1936, $N=1005$) provided cognitive data and genome-wide genotype data. Polygenic risk profile scores for CAD were calculated for all of the cohorts using the largest available genome-wide association studies (GWAS) data set, the CARDIoGRAM consortium (22 233 cases and 64 762 controls). Polygenic risk profile scores for CAD were then tested for their association with cognitive abilities in the presence and absence of manifest cardiovascular disease.

Results: A meta-analysis of all three cohorts showed a negative association between CAD polygenic risk and fluid cognitive ability ($\beta = 0.022$, $P=0.016$), verbal intelligence ($\beta = 0.024$, $P=0.011$) and memory ($\beta = 0.021$, $P=0.028$).

Conclusions: Increased polygenic risk for CAD is associated with lower cognitive ability in older adults. Common genetic variants may underlie some of the association between age-related cognitive decrements and the risk for CAD.

E.g., Gene-Environment Interaction

Genetic variants in *ABCB1* and *CYP2C19* and cardiovascular outcomes after treatment with clopidogrel and prasugrel in the TRITON–TIMI 38 trial: a pharmacogenetic analysis

Jessica L Mega, Sandra L Close*, Stephen D Wiviott, Lei Shen, Joseph R Walker, Tabassome Simon, Elliott M Antman, Eugene Braunwald, Marc S Sabatine*

Lancet 2010; 376: 1312–19
Published Online August 29, 2010
DOI:10.1016/S0140-6736(10)61273-1

See Comment page 1278
See Articles page 1320

*Authors contributed equally

TIMI Study Group, Cardiovascular Division, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA (J L Mega MD, S D Wiviott MD, Prof E M Antman MD, Prof E Braunwald MD, M S Sabatine MD); **Indiana University, Indianapolis, IN, USA** (S L Close PhD); **Eli Lilly and Company, Indianapolis, IN, USA** (L Shen PhD, S L Close PhD); **Daiichi Sankyo Inc, Edison, NJ, USA** (J R Walker PharmD); **and Assistance Publique-Hôpitaux de Paris, UPMC-Paris06, France** (Prof T Simon MD)

Correspondence to: Dr Jessica L Mega or Dr Marc S Sabatine, Brigham and Women's Hospital, TIMI Study Group, Cardiovascular Division, 350 Longwood Ave, Boston, MA 02115, USA
jmega@partners.org
msabatine@partners.org

Summary

Background Clopidogrel and prasugrel are subject to efflux via P-glycoprotein (encoded by *ABCB1*, also known as *MDR1*). *ABCB1* polymorphisms, particularly 3435C→T, may affect drug transport and efficacy. We aimed to assess the effect of this polymorphism by itself and alongside variants in *CYP2C19* on cardiovascular outcomes in patients treated with clopidogrel or prasugrel in TRITON–TIMI 38. We also assessed the effect of genotype on the pharmacodynamic and pharmacokinetic properties of these drugs in healthy individuals.

Methods We genotyped *ABCB1* in 2932 patients with acute coronary syndromes undergoing percutaneous intervention who were treated with clopidogrel (n=1471) or prasugrel (n=1461) in the TRITON–TIMI 38 trial. We evaluated the association between *ABCB1* 3435C→T and rates of the primary efficacy endpoint (cardiovascular death, myocardial infarction, or stroke) until 15 months. We then assessed the combined effect of *ABCB1* 3435C→T genotype and reduced-function alleles of *CYP2C19*. 321 healthy individuals were also genotyped, and we tested the association of genetic variants with reduction in maximum platelet aggregation and plasma concentrations of active drug metabolites.

Findings In patients treated with clopidogrel, *ABCB1* 3435C→T genotype was significantly associated with the risk of cardiovascular death, myocardial infarction, or stroke (p=0·0064). TT homozygotes had a 72% increased risk of the primary endpoint compared with CT/CC individuals (Kaplan-Meier event rates 12·9% [52 of 414] *vs* 7·8% [80 of 1057 participants]; HR 1·72, 95% CI 1·22–2·44, p=0·002). *ABCB1* 3435C→T and *CYP2C19* genotypes were significant, independent predictors of the primary endpoint, and 681 (47%) of the 1454 genotyped patients taking clopidogrel who were either *CYP2C19* reduced-function allele carriers, *ABCB1* 3435 TT homozygotes, or both were at increased risk of the primary endpoint (HR 1·97, 95% CI 1·38–2·82, p=0·0002). In healthy participants, 3435 TT homozygotes had an absolute reduction in maximum platelet aggregation with clopidogrel that was 7·3 percentage points less than for CT/CC individuals (p=0·0127). *ABCB1* genotypes were not significantly associated with clinical or pharmacological outcomes in patients with an acute coronary syndrome or healthy individuals treated with prasugrel, respectively.

Interpretation Individuals with the *ABCB1* 3435 TT genotype have reduced platelet inhibition and are at increased risk of recurrent ischaemic events during clopidogrel treatment. In patients with acute coronary syndromes who have undergone percutaneous intervention, when both *ABCB1* and *CYP2C19* are taken into account, nearly half of the population carries a genotype associated with increased risk of major adverse cardiovascular events while on standard doses of clopidogrel.

Funding Daiichi Sankyo Company Ltd and Eli Lilly and Company.

E.g., G

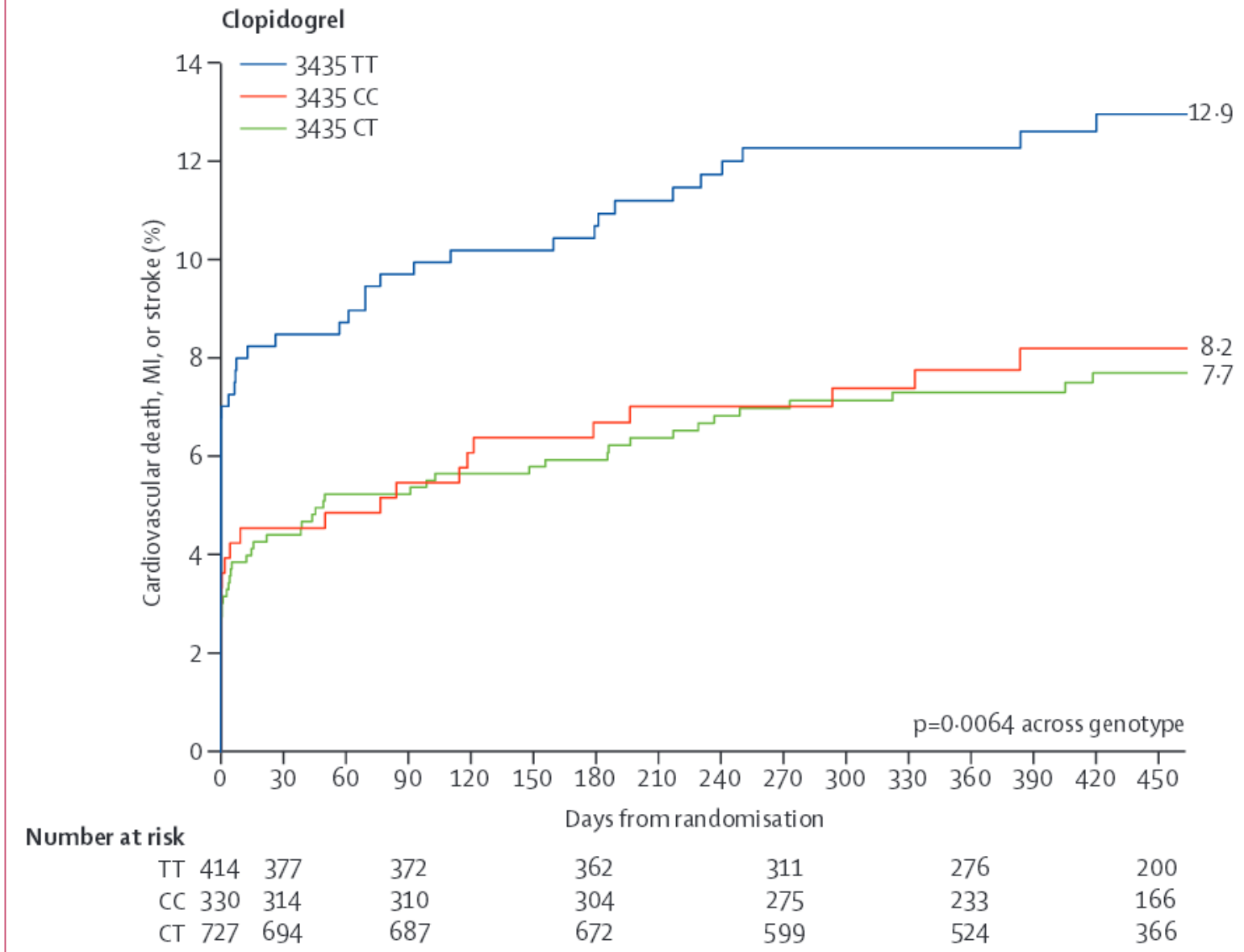
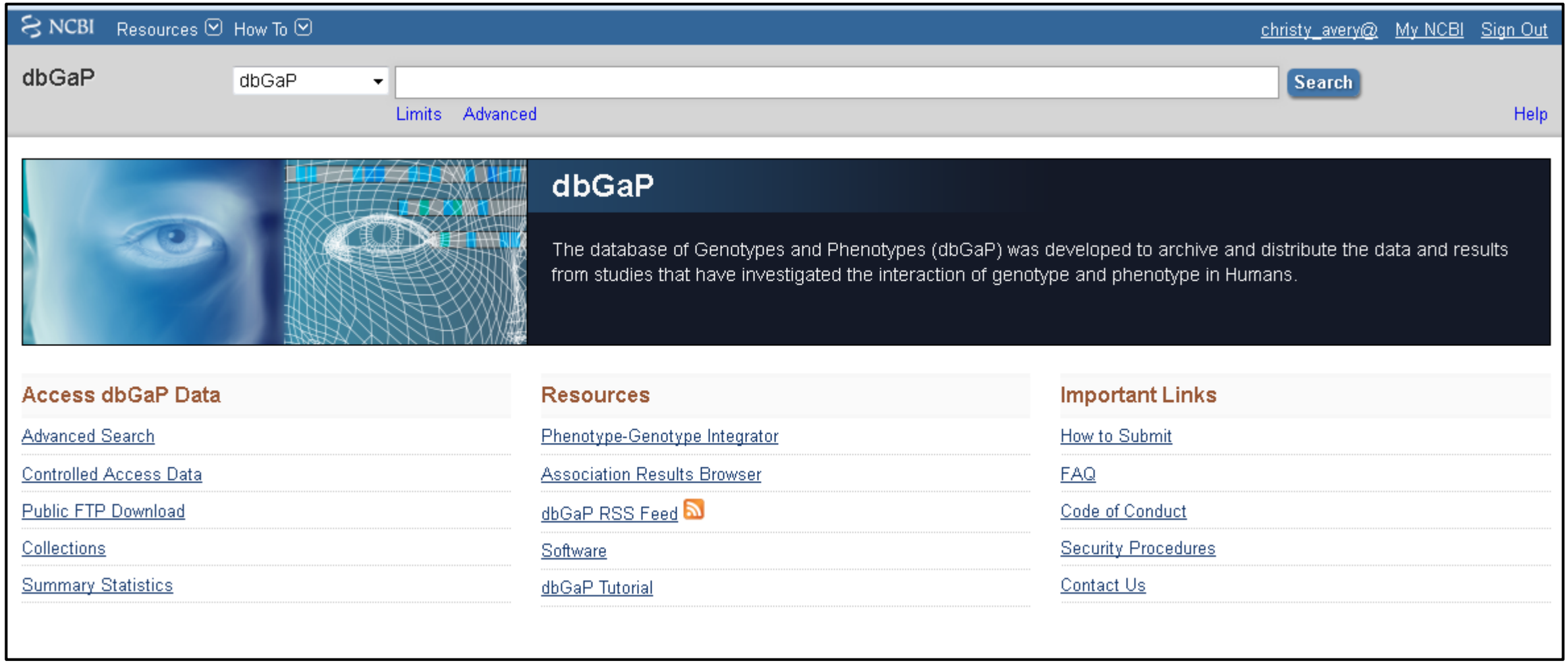


***Figure 1:* *ABCB1* 3435C→T and cardiovascular outcomes in patients treated with clopidogrel**
Cumulative risk of cardiovascular death, myocardial infarction (MI), or stroke for each genotype, with a p value across genotype.

How Can You Obtain Add Health GWAS Data?

dbGaP: Add Health Genotype Warehouse



Genome-wide Association Study of Adiposity in Samoans

dbGaP Study Accession: phs000914.v1.p1

Show BioProject list

Study | Variables | Documents | Analyses | Datasets | Molecular Data

Jump to: Authorized Access | Attribution | Authorized Requests

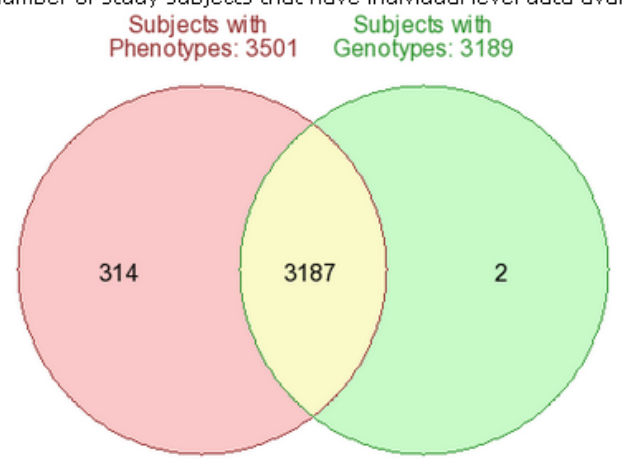
Study Description

The research goal of this study is to identify genetic variation that increases susceptibility to obesity and cardiometabolic phenotypes among adult Samoans using genome-wide association (GWAS) methods. DNA from peripheral blood and phenotypic information were collected from 3,119 adult Samoans, 23 to 70 years of age. The participants reside throughout the independent nation of Samoa, which is experiencing economic development and the nutrition transition. Genotyping was performed with the Affymetrix Genome-Wide Human SNP 6.0 Array using a panel of approximately 900,000 SNPs. Anthropometric, fasting blood biomarkers and detailed dietary, physical activity, health and socio-demographic variables were collected. We are replicating the GWAS findings in an independent sample of 2,500 Samoans from earlier studies. After replication of genomic regions and informative SNPs in those regions, we will determine sequences of the important genes, and determine the specific genetic variants in the sequenced genes that are associated with adiposity and related cardiometabolic conditions. We will also identify gene by environment interactions, focusing on dietary intake patterns and nutrients.

Important Links and Information

- Request access via Authorized Access
  - Instructions for requestors
  - Data Use Certification (DUC) Agreement
- Talking Glossary of Genetic Terms

- Study Types: Cross-Sectional, Population
- Number of study subjects that have individual level data available through Authorized Access: 3501



NCBI Resources How To

Genome-wide Association Study of Adiposity in Samoans

dbGaP Study Accession: phs000914.v1.p1

Show BioProject list

Study | Variables | Documents | Analyses | Datasets | Molecular Data

Jump to: Authorized Access | Attribution | Authorized Requests

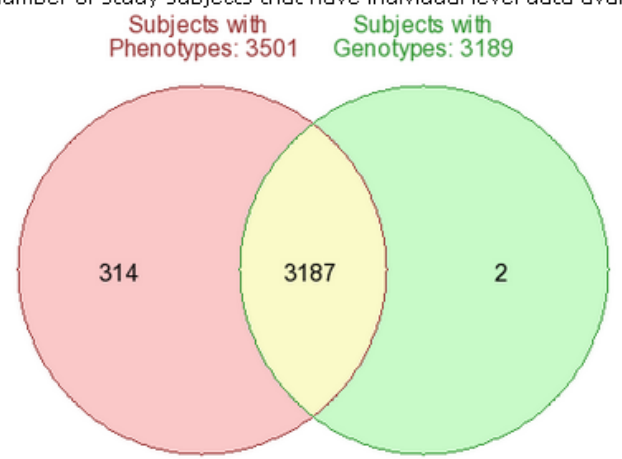
Study Description

The research goal of this study is to identify genetic variation that increases susceptibility to obesity and cardiometabolic phenotypes among adult Samoans using genome-wide association (GWAS) methods. DNA from peripheral blood and phenotypic information were collected from 3,119 adult Samoans, 23 to 70 years of age. The participants reside throughout the independent nation of Samoa, which is experiencing economic development and the nutrition transition. Genotyping was performed with the Affymetrix Genome-Wide Human SNP 6.0 Array using a panel of approximately 900,000 SNPs. Anthropometric, fasting blood biomarkers and detailed dietary, physical activity, health and socio-demographic variables were collected. We are replicating the GWAS findings in an independent sample of 2,500 Samoans from earlier studies. After replication of genomic regions and informative SNPs in those regions, we will determine sequences of the important genes, and determine the specific genetic variants in the sequenced genes that are associated with adiposity and related cardiometabolic conditions. We will also identify gene by environment interactions, focusing on dietary intake patterns and nutrients.

Important Links and Information

- Request access via Authorized Access
  - Instructions for requestors
  - Data Use Certification (DUC) Agreement
- Talking Glossary of Genetic Terms

- Study Types: Cross-Sectional, Population
- Number of study subjects that have individual level data available through Authorized Access: 3501



NCBI Resources How To

dbGaP GENOTYPES and PHENOTYPES

Genome-wide Association Study of Adiposity in Samoans

dbGaP Study Accession: phs000914.v1.p1

Show BioProject list

Study | Variables | Documents | Analyses | Datasets | Molecular Data

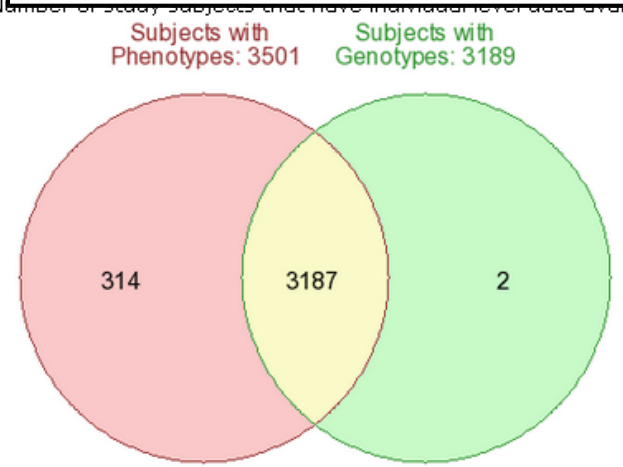
Jump to: Authorized Access | Attribution | Authorized Requests

Study Description

The research goal of this study is to identify genetic variation that increases susceptibility to obesity and cardiometabolic phenotypes among adult Samoans using genome-wide association (GWAS) methods. DNA from peripheral blood and phenotypic information were collected from 3,119 adult Samoans, 23 to 70 years of age.

Important Links and Information

Add Health Phenotype Data Are Available Through the Add Health Study (UNC)



Analysis Best Practices/Hints

- Do not discount the large number of existing resources!
- Team science/consortia
- Statistical power
- Race/ethnicity heterogeneity and admixture
- Intergenic regions
- Family structure/clustering
- Analytic pipeline

Summary Results from *Many Large Consortia* Are Available Online

Meta-Analyses of Glucose and Insulin-related traits Consortium

MAGIC

- Home page
- Steering committee
- Cohorts
- Publications
- Data downloads
- Ongoing projects

Data available for Download

We are releasing the summary data from our genome-wide meta analyses of glycaemic traits, in order to enable other researchers to examine particular variants or loci of their interest for association with these traits. The files include p-values and direction of effect at over 2 million directly genotyped or imputed single nucleotide polymorphisms (SNPs), as well as frequency information from the HapMap project (release 27).

Acknowleding the data

When using data from the downloadable meta-analyses results please acknowledge the source of the data as follows: **"Data on glycaemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org".**

In addition to the above acknowledgement, please cite the relevant paper.

Downloading the data

The data can be downloaded from the magic directory on the Sanger FTP site:

Summary Results from *Many Large Consortia* Are Available Online

GIANT CONSORTIUM

Page | Discussion | Read | View source | View history | Go | Searc

Navigation

- Main page
- Data Release
- Community portal
- Recent changes
- Help

Toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

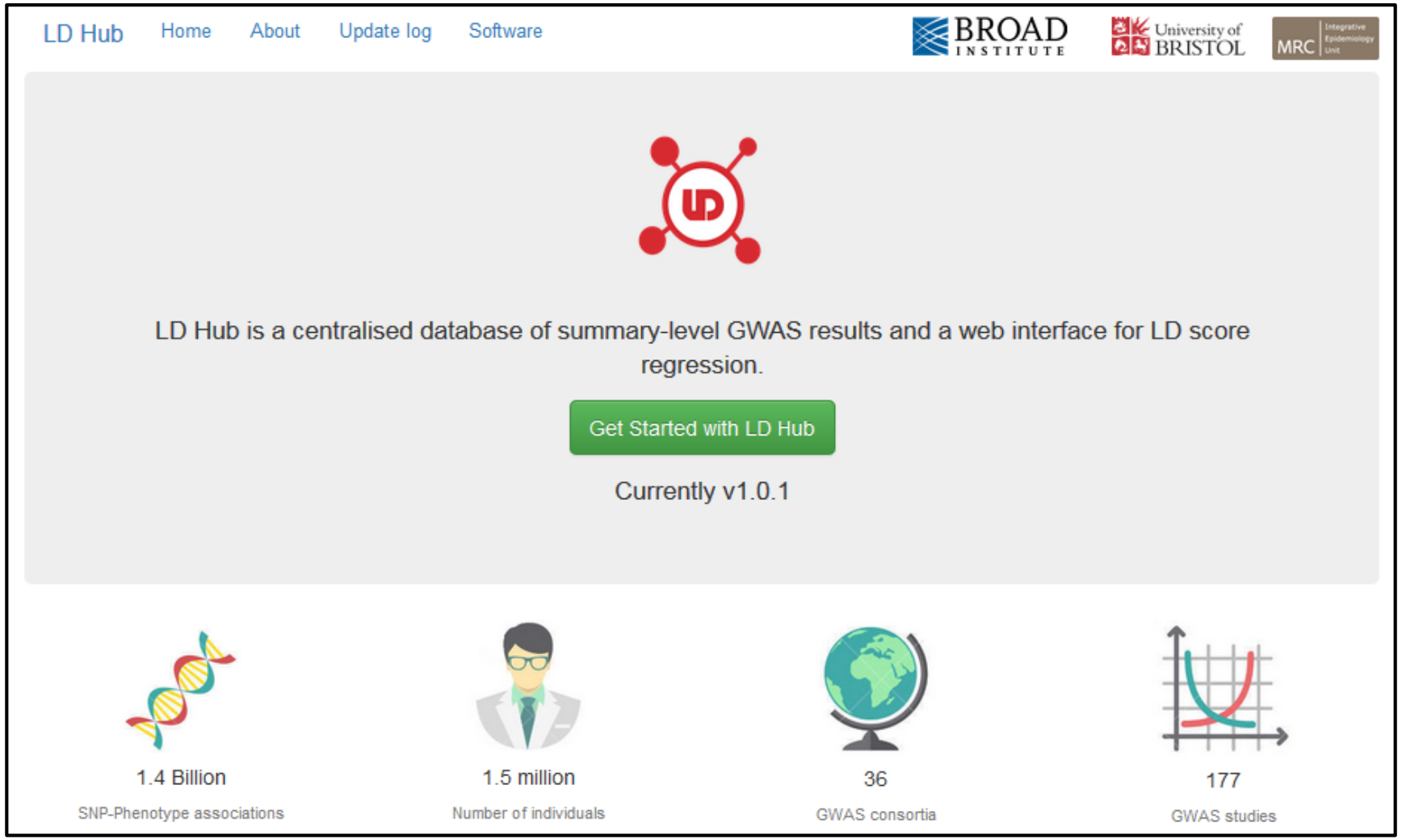
GIANT consortium data files

We are releasing the summary data from our 2010-2013 meta-analyses of Genome-wide Association (GWA) data, in order to enable other researchers to examine particular variants or loci for their evidence of association with anthropometric traits. The files include p-values and direction of effect at over 2 million directly genotyped or imputed single nucleotide polymorphisms (SNPs). To prevent the possibility of identification of individuals from these summary results, we are not releasing allele frequency data from our samples.

Contents [hide]

GIANT Consortium 2010 GWAS Metadata is Available Here for Download

Summary Results from *Many Large Consortia* Are Available Online



Summary Results from *Many Large Consortia* Are Available Online

OXFORD JOURNALS

Human Molecular Genetics

For large-scale or genome-wide genetic studies, we feel that widespread availability of the complete set of results is highly desirable. Authors of manuscripts describing new genome-wide association or similar data must indicate in their cover letter whether at least a minimal set of summary results (p value and direction of effect) will be made freely available for all variants, either as supplementary material, by being publicly posted, or by being deposited in a database that is accessible to researchers with minimal restrictions on access. Making these results available will not necessarily be required in all cases for acceptance of a manuscript for publication, but the availability of results after publication will be considered in decisions regarding publication. Accordingly, authors are strongly encouraged to make available complete lists of summary statistics for large scale genetic studies.

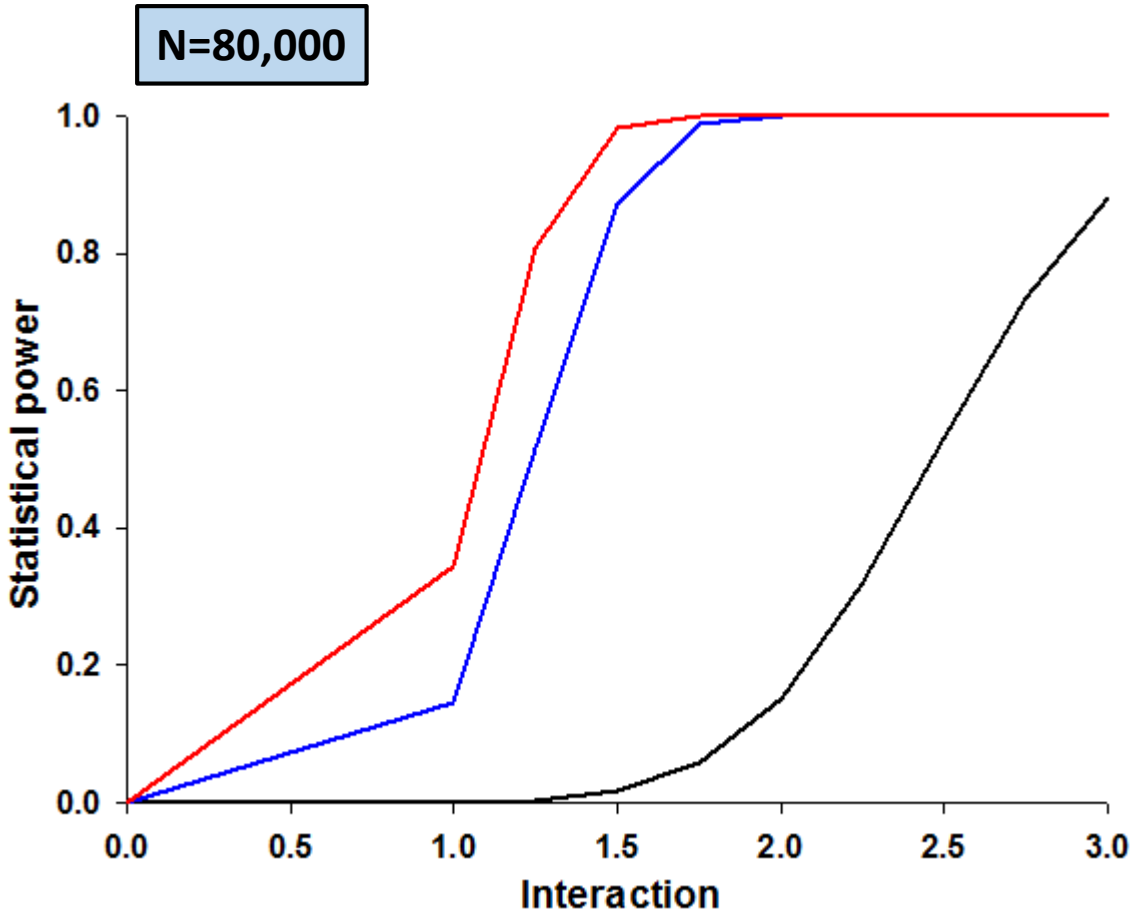
Team Science/Consortia

- Joining a consortium is often the first step in ***<u>GWAS</u>***

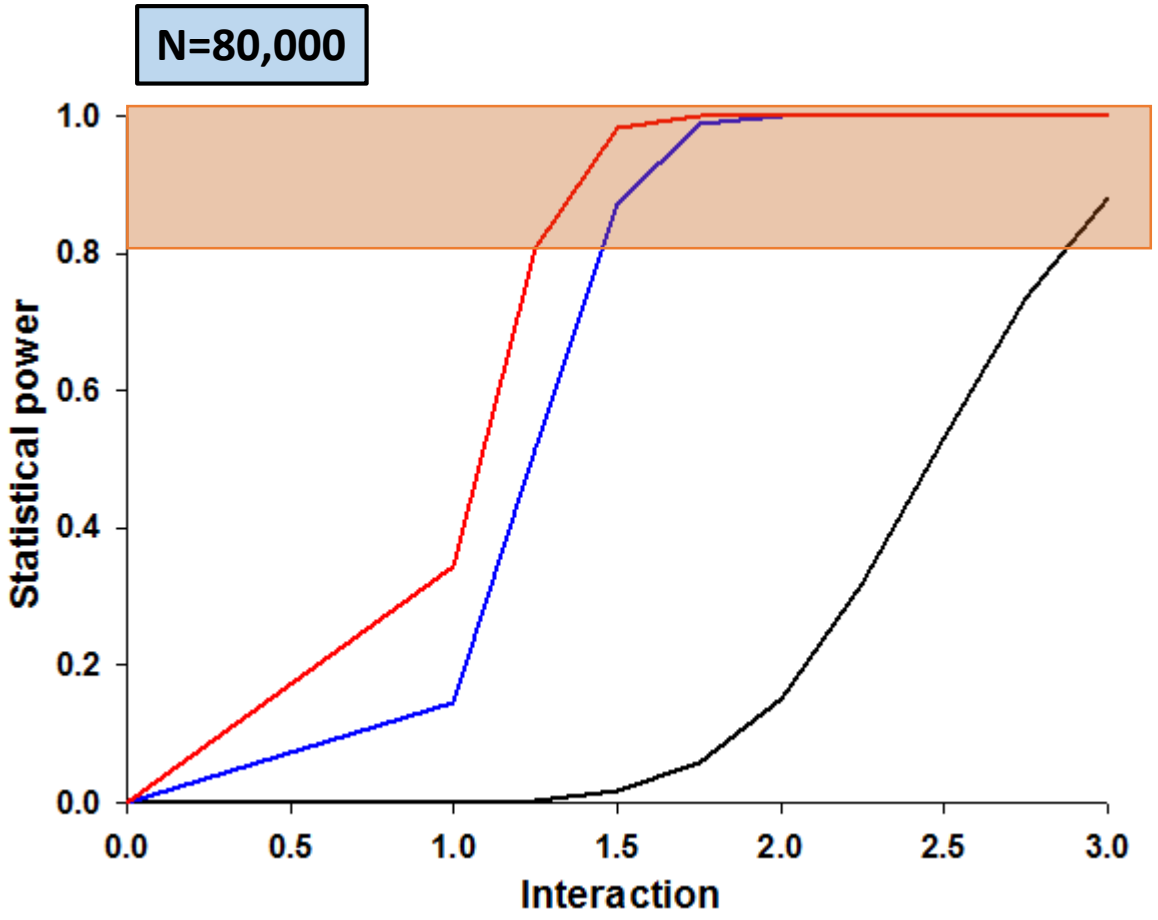
Team Science/Consortia

- Joining a consortium is often the first step in ***<u>GWAS</u>***

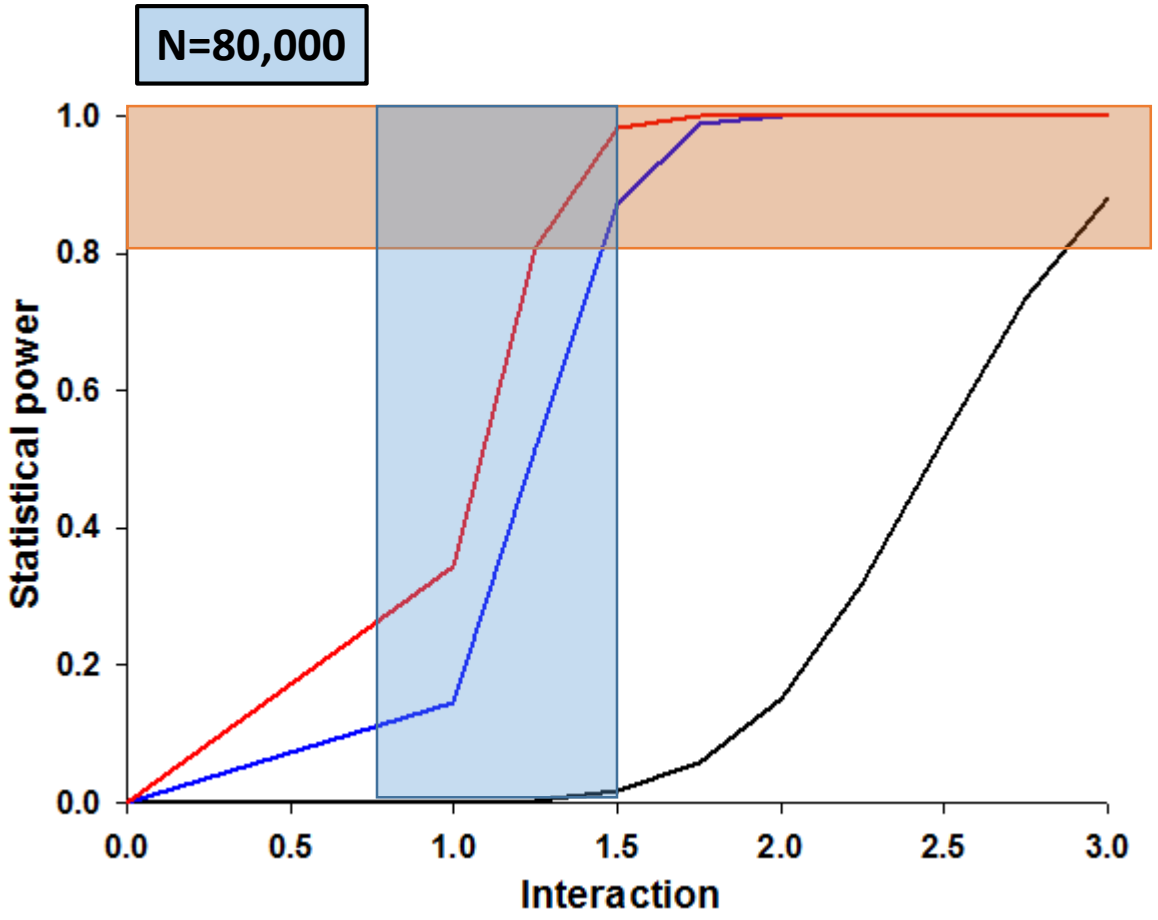
Statistical Power: Gene-Environment GWAS



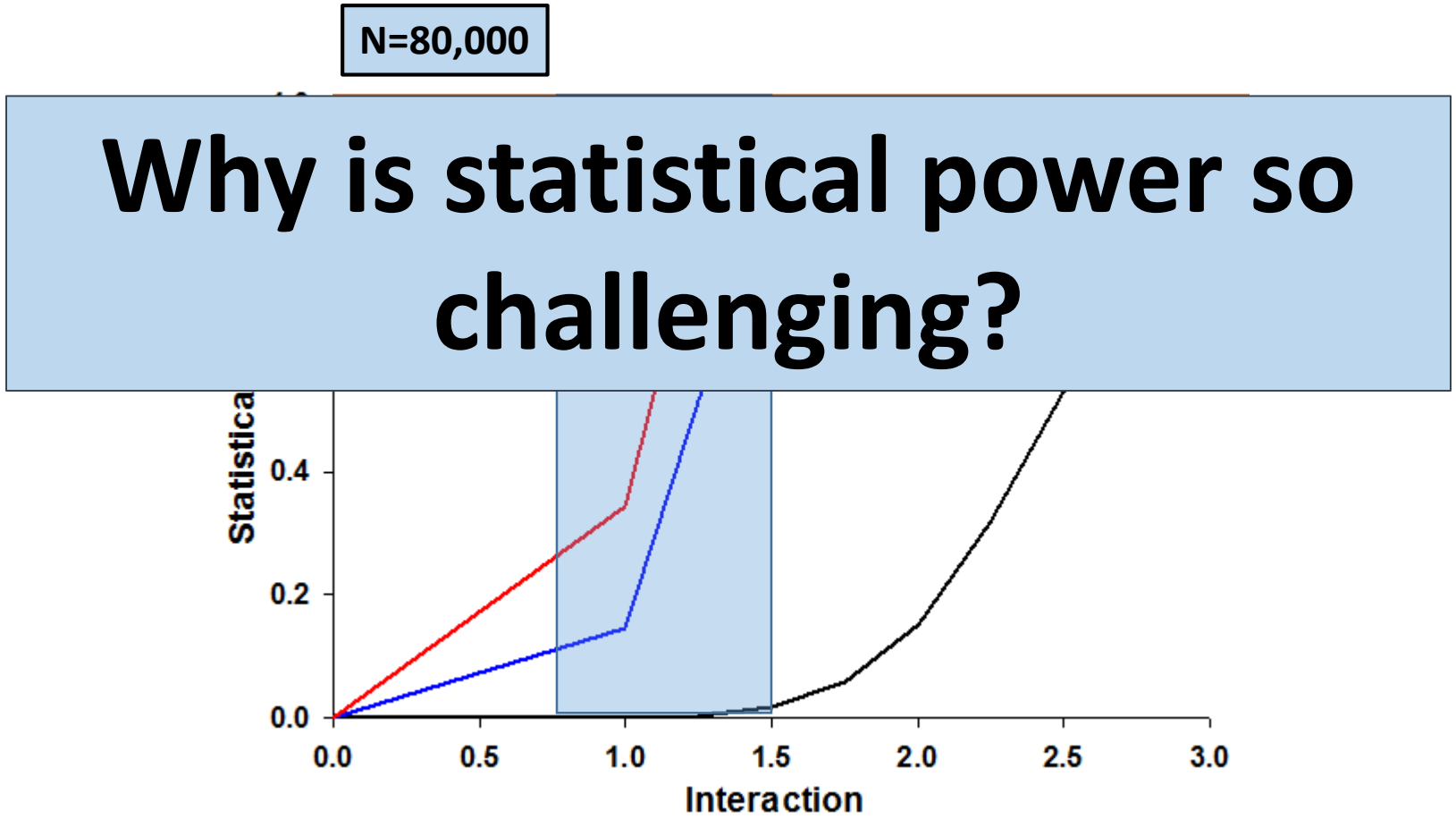
Statistical Power: Gene-Environment GWAS



Statistical Power: Gene-Environment GWAS



Statistical Power: Gene-Environment GWAS



Why is statistical power so challenging?

Statistical Power: Gene-Environment GWAS

N=80,000

We need to correct for 1,000,000 statistical tests when interrogating genome!

$\alpha = 0.05/1M$ or $5x10^{-8}$

Correction for only 1M tests given correlation in human genome

Race/Ethnicity Heterogeneity

Genomics for the world

Medical genomics has focused almost entirely on those of European descent. Other ethnic groups must be studied to ensure that more people benefit, say **Carlos D. Bustamante, Esteban González Burchard** and **Francisco M. De La Vega.**

In the past decade, researchers have dramatically improved our understanding of the genetic basis of complex chronic diseases, such as Alzheimer's disease and type 2 diabetes, through more than 1,000 genome-wide association studies (GWAS). These scan the genomes of thousands of people for known genetic variants, to find out which are associated with a particular condition.

Yet the findings from such studies are likely to have less relevance than was previously thought for the world's population as a whole. Ninety-six per cent of subjects included in the GWAS conducted so far are people of European descent[1] (see 'Sampling bias'). And a recent *Nature* survey suggests that this bias is likely to persist in the upcoming efforts to sequence people's entire genomes[2].

Geneticists worldwide must investigate a much broader ensemble of populations, including racial and ethnic minorities. If we do not, a biased picture will emerge of which variants are important, and genomic medicine will largely benefit a privileged few.

SUMMARY

- Those most in need must not be the last to benefit from genetic research
- Reviewers and granting bodies must demand racial and ethnic diversity in genome studies
- Global genomics needs the financial support of governments and non-profits

PMID:21753830

Race/Ethnicity Heterogeneity

Why are genomic studies in non-European populations necessary?

Medical genomics has focused almost entirely on those of European descent. Other ethnic groups must be studied to ensure that more people benefit, say **Carlos D. Bustamante, Esteban González Burchard** and **Francisco M. De La Vega.**

In the past decade, researchers have dramatically improved our understanding of the genetic basis of complex chronic diseases, such as Alzheimer's disease and type 2 diabetes, through more than 1,000 genome-wide association studies (GWAS). These scan the genomes of thousands of people for known genetic variants, to find out which are associated with a particular condition.

Yet the findings from such studies are likely to have less relevance than was previously thought for the world's population as a whole. Ninety-six per cent of subjects included in the GWAS conducted so far are people of European descent[1] (see 'Sampling bias'). And a recent *Nature* survey suggests that this bias is likely to persist in the upcoming efforts to sequence people's entire genomes[2].

Geneticists worldwide must investigate a much broader ensemble of populations, including racial and ethnic minorities. If we do not, a biased picture will emerge of which variants are important, and genomic medicine will largely benefit a privileged few.

SUMMARY

- Those most in need must not be the last to benefit from genetic research
- Reviewers and granting bodies must demand racial and ethnic diversity in genome studies
- Global genomics needs the financial support of governments and non-profits

PMID:21753830

Limited Studies Suggest that Genes Generalize Across Global Populations…

A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry

Genome-wide association studies (GWAS) have identified 36 loci associated with body mass index (BMI), predominantly in populations of European ancestry. We conducted a meta-analysis to examine the association of >3.2 million SNPs with BMI in 39,144 men and women of African ancestry and followed up the most significant associations in an additional 32,268 individuals of African ancestry. We identified one new locus at 5q33 (*GALNT10*, rs7708584, $P = 3.4 \times 10^{-11}$) and another at 7p15 when we included data from the GIANT consortium (*MIR148A-NFE2L3*, rs10261878, $P = 1.2 \times 10^{-10}$). We also found suggestive evidence of an association at a third locus at 6q16 in the African-ancestry sample (*KLHL32*, rs974417, $P = 6.9 \times 10^{-8}$). Thirty-two of the 36 previously established BMI variants showed directionally consistent effect estimates in our GWAS (binomial $P = 9.7 \times 10^{-7}$), five of which reached genome-wide significance. These findings provide strong support for shared BMI loci across populations, as well as for the utility of studying ancestrally diverse populations.

of *GALNT10*, $P = 8.02 \times 10^{-9}$), has not been previously associated with BMI at genome-wide significant levels in any population.

We subsequently selected the 1,500 most significantly associated SNPs from stage 1 ($P < 1.19 \times 10^{-3}$) and examined associations with BMI in an independent sample of 6,817 men and women of African ancestry from seven additional studies (stage 2) (Online Methods, **Supplementary Tables 1–3** and **Supplementary Note**). Of these 1,500 SNPs, 179 replicated at nominal significance ($P < 0.05$) and had effects that were directionally consistent with those in stage 1 (**Supplementary Table 4**). A meta-analysis of stages 1 and 2 revealed a second new locus, 6q16 (rs974417, located in an intronic region of *KLHL32*; stage 2 $P = 3.5 \times 10^{-3}$, combined stages 1 and 2 $P = 2.2 \times 10^{-8}$), and confirmed our finding at rs7708584 on 5q33 near *GALNT10* (stage 2 $P = 9.4 \times 10^{-3}$, combined stages 1 and 2 $P = 2.2 \times 10^{-10}$). We further examined the associations of these two variants in a third stage composed of 25,451 individuals of African ancestry from an additional 12 studies. We found support for an association with both variants, although the strength of the association was

Limited Studies Suggest that Genes Generalize Across

A meta-analysi
mass index in i

Genome-wide association studies
36 loci associated with body mass
in populations of European ancest
analysis to examine the associatio
with BMI in 39,144 men and wom
followed up the most significant a
32,268 individuals of African anc
new locus at 5q33 (*GALNT10*, rs7
and another at 7p15 when we inc
consortium (*MIR148A-NFE2L3*, rs
We also found suggestive evidenc
third locus at 6q16 in the African-
rs974417, $P = 6.9 \times 10^{-8}$). Thirty-
established BMI variants showed
estimates in our GWAS (binomial
reached genome-wide significanc
strong support for shared BMI loc
as for the utility of studying ances

Generalization of Associations of Kidney-Related Genetic Loci to American Indians

Nora Franceschini, Karin Haack,† Laura Almasy,† Sandra Laston,† Elisa T. Lee,‡ Lyle G. Best,§ Richard R. Fabsitz,‖ Jean W. MacCluer,† Barbara V. Howard,¶ Jason G. Umans,¶** and Shelley A. Cole†*

Summary

Background and objectives CKD disproportionally affects American Indians, who similar to other populations, show genetic susceptibility to kidney outcomes. Recent studies have identified several loci associated with kidney traits, but their relevance in American Indians is unknown.

Design, setting, participants, & measurements This study used data from a large, family-based genetic study of American Indians (the Strong Heart Family Study), which includes 94 multigenerational families enrolled from communities located in Oklahoma, the Dakotas, and Arizona. Individuals were recruited from the Strong Heart Study, a population-based study of cardiovascular disease in American Indians. This study selected 25 single nucleotide polymorphisms in 23 loci identified from recently published kidney-related genome-wide association studies in individuals of European ancestry to evaluate their associations with kidney function (estimated GFR; individuals 18 years or older, up to 3282 individuals) and albuminuria (urinary albumin to creatinine ratio; n=3552) in the Strong Heart Family Study. This study also examined the association of single nucleotide polymorphisms in the *APOL1* region with estimated GFR in 1121 Strong Heart Family Study participants. GFR was estimated using the abbreviated Modification of Diet in Renal Disease Equation. Additive genetic models adjusted for age and sex were used.

Results This study identified significant associations of single nucleotide polymorphisms with estimated GFR in or nearby *PRKAG2*, *SLC6A13*, *UBE2Q2*, *PIP5K1B*, and *WDR72* ($P<2.1 \times 10^{-3}$ to account for multiple testing). Single nucleotide polymorphisms in these loci explained 2.2% of the estimated GFR total variance and 2.9% of its heritability. An intronic variant of *BCAS3* was significantly associated with urinary albumin to creatinine ratio. *APOL1* single nucleotide polymorphisms were not associated with estimated GFR in a single variant test or haplotype analyses, and the at-risk variants identified in individuals with African ancestry were not detected in DNA sequencing of American Indians.

Conclusion This study extends the genetic associations of loci affecting kidney function to American Indians, a population at high risk of kidney disease, and provides additional support for a potential biologic relevance of these loci across ancestries.

Clin J Am Soc Nephrol 9: 150–158, 2014. doi: 10.2215/CJN.02300213

Limited Studies Suggest
Across

A meta-analysi
mass index in i

Genome-wide association studies
36 loci associated with body mass
in populations of European ancest
analysis to examine the associatio
with BMI in 39,144 men and wom
followed up the most significant a
32,268 individuals of African anc
new locus at 5q33 (*GALNT10*, rs7
and another at 7p15 when we inc
consortium (*MIR148A-NFE2L3*, rs
We also found suggestive evidenc
third locus at 6q16 in the African-
rs974417, $P = 6.9 \times 10^{-8}$). Thirty-
established BMI variants showed
estimates in our GWAS (binomial
reached genome-wide significanc
strong support for shared BMI loc
as for the utility of studying ances

Generalization of Associations o
Genetic Loci to American Indian

Nora Franceschini, Karin Haack,† Laura Almasy,† Sandra Laston,† Elis
Jean W. MacCluer,† Barbara V. Howard,¶ Jason G. Umans,¶** and Sl*

Summary

Background and objectives CKD disproportionally affects Americ
show genetic susceptibility to kidney outcomes. Recent studies ha
kidney traits, but their relevance in American Indians is unknown

Design, setting, participants, & measurements This study used da
American Indians (the Strong Heart Family Study), which include
from communities located in Oklahoma, the Dakotas, and Arizon
Heart Study, a population-based study of cardiovascular disease
25 single nucleotide polymorphisms in 23 loci identified from rec
association studies in individuals of European ancestry to evaluat
(estimated GFR; individuals 18 years or older, up to 3282 individ
creatinine ratio; n=3552) in the Strong Heart Family Study. This s
nucleotide polymorphisms in the *APOL1* region with estimated G
ticipants. GFR was estimated using the abbreviated Modification
genetic models adjusted for age and sex were used.

Results This study identified significant associations of single nucl
or nearby *PRKAG2*, *SLC6A13*, *UBE2Q2*, *PIP5K1B*, and *WDR72* (P-
Single nucleotide polymorphisms in these loci explained 2.2% of th
heritability. An intronic variant of *BCAS3* was significantly associ
APOL1 single nucleotide polymorphisms were not associated wit
haplotype analyses, and the at-risk variants identified in individu
DNA sequencing of American Indians.

Conclusion This study extends the genetic associations of loci affe
population at high risk of kidney disease, and provides additiona
these loci across ancestries.

Clin J Am Soc Nephrol

OPEN ACCESS Freely available online PLoS GENETICS

Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study

Logan Dumitrescu[1], Cara L. Carty[2], Kira Taylor[3], Fredrick R. Schumacher[4], Lucia A. Hindorff[5], José L. Ambite[6], Garnet Anderson[2], Lyle G. Best[7], Kristin Brown-Gentry[1], Petra Bůžková[8], Christopher S. Carlson[2], Barbara Cochran[9], Shelley A. Cole[10], Richard B. Devereux[11], Dave Duggan[12], Charles B. Eaton[13], Myriam Fornage[14,15], Nora Franceschini[3], Jeff Haessler[2], Barbara V. Howard[16], Karen C. Johnson[17], Sandra Laston[10], Laurence N. Kolonel[18], Elisa T. Lee[19], Jean W. MacCluer[10], Teri A. Manolio[5], Sarah A. Pendergrass[1], Miguel Quibrera[20], Ralph V. Shohet[21], Lynne R. Wilkens[18], Christopher A. Haiman[4], Loïc Le Marchand[18], Steven Buyske[22], Charles Kooperberg[2], Kari E. North[3,23], Dana C. Crawford[1,24]*

1 Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **3** Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, United States of America, **4** Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **5** Office of Population Genomics, National Human Genome Research Institute, Bethesda, Maryland, United States of America, **6** Information Sciences Institute, University of Southern California, Los Angeles, California, United States of America, **7** Missouri Breaks Industries Research, Timber Lake, South Dakota, United States of America, **8** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, **9** Sponsored Programs, Baylor College of Medicine, Houston, Texas, United States of America, **10** Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America, **11** Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, **12** The Translational Genomics Research Institute, Phoenix, Arizona, United States of America, **13** Department of Family Medicine and Community Health, Alpert Medical School of Brown University School of Medicine, Providence, Rhode Island, United States of America, **14** Institute of Molecular Medicine, University of Texas Health Sciences Center at Houston, Texas, United States of America, **15** Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center, Houston, Texas, United States of America, **16** Medstar Research Institute, Washington, D.C., United States of America, **17** Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **18** Epidemiology Program, University of Hawaii Cancer Center, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, **19** University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, **20** School of Public Health, University of North Carolina, Chapel Hill, North Carolina, United States of America, **21** Center of Cardiovascular Research, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, **22** Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, United States of America, **23** Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina, United States of America, **24** Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America

Abstract

For the past five years, genome-wide association studies (GWAS) have identified hundreds of common variants associated with human diseases and traits, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglyceride (TG) levels. Approximately 95 loci associated with lipid levels have been identified primarily among populations of European ancestry. The Population Architecture using Genomics and Epidemiology (PAGE) study was established in 2008 to characterize GWAS–identified variants in diverse population-based studies. We genotyped 49 GWAS–identified SNPs associated with one or more lipid traits in at least two PAGE studies and across six racial/ethnic groups. We performed a meta-analysis testing for SNP associations with fasting HDL-C, LDL-C, and ln(TG) levels in self-identified European American (~20,000), African American (~9,000), American Indian (~6,000), Mexican American/Hispanic (~2,500), Japanese/East Asian (~690), and Pacific Islander/Native Hawaiian (~175) adults, regardless of lipid-lowering medication use. We replicated 55 of 60 (92%) SNP associations tested in European Americans at $p<0.05$. Despite sufficient power, we were unable to replicate *ABCA1* rs4149268 and rs1883025, *CETP* rs1864163, and *TTC39B* rs471364 previously associated with HDL-C and *MAFB* rs6102059 previously associated with LDL-C. Based on significance ($p<0.05$) and consistent direction of effect, a majority of replicated genotype-phentoype associations for HDL-C, LDL-C, and ln(TG) in European Americans generalized to African Americans (48%, 61%, and 57%), American Indians (45%, 64%, and 77%), and Mexican Americans/Hispanics (57%, 56%, and 86%). Overall, 16 associations generalized across all three populations. For the associations that did not generalize, differences in effect sizes, allele frequencies, and linkage disequilibrium offer clues to the next generation of association studies for these traits.

Limited Studies Suggest
Across

A meta-analysi
mass index in i

Genome-wide association studies
36 loci associated with body mass
in populations of European ancest
analysis to examine the associatio
with BMI in 39,144 men and wom
followed up the most significant a
32,268 individuals of African anc
new locus at 5q33 (*GALNT10*, rs7
and another at 7p15 when we inc
consortium (*MIR148A-NFE2L3*, rs
We also found suggestive evidenc
third locus at 6q16 in the African-
rs974417, $P = 6.9 \times 10^{-8}$). Thirty-
established BMI variants showed
estimates in our GWAS (binomial
reached genome-wide significanc
strong support for shared BMI loc
as for the utility of studying ances

Generalization of Associations o
Genetic Loci to American Indian

Nora Franceschini, Karin Haack,† Laura Almasy,† Sandra Laston,† Elis
Jean W. MacCluer,† Barbara V. Howard,¶ Jason G. Umans,¶** and Sl*

Summary

Background and objectives CKD disproportionally affects Americ
show genetic susceptibility to kidney outcomes. Recent studies ha
kidney traits, but their relevance in American Indians is unknown

Design, setting, participants, & measurements This study used da
American Indians (the Strong Heart Family Study), which include
from communities located in Oklahoma, the Dakotas, and Arizon
Heart Study, a population-based study of cardiovascular disease
25 single nucleotide polymorphisms in 23 loci identified from rec
association studies in individuals of European ancestry to evaluat
(estimated GFR; individuals 18 years or older, up to 3282 individ
creatinine ratio; n=3552) in the Strong Heart Family Study. This s
nucleotide polymorphisms in the *APOL1* region with estimated G
ticipants. GFR was estimated using the abbreviated Modification
genetic models adjusted for age and sex were used.

Results This study identified significant associations of single nucl
or nearby *PRKAG2*, *SLC6A13*, *UBE2Q2*, *PIP5K1B*, and *WDR72* (P-
Single nucleotide polymorphisms in these loci explained 2.2% of th
heritability. An intronic variant of *BCAS3* was significantly associ
APOL1 single nucleotide polymorphisms were not associated wit
haplotype analyses, and the at-risk variants identified in individu
DNA sequencing of American Indians.

Conclusion This study extends the genetic associations of loci affe
population at high risk of kidney disease, and provides additiona
these loci across ancestries.

Clin J Am Soc Nephrol

OPEN ACCESS Freely available online PLoS GENETICS

Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study

Logan Dumitrescu[1], Cara L. Carty[2], Kira Taylor[3], Fredrick R. Schumacher[4], Lucia A. Hindorff[5], José L. Ambite[6], Garnet Anderson[2], Lyle G. Best[7], Kristin Brown-Gentry[1], Petra Bůžková[8], Christopher S. Carlson[2], Barbara Cochran[9], Shelley A. Cole[10], Richard B. Devereux[11], Dave Duggan[12], Charles B. Eaton[13], Myriam Fornage[14,15], Nora Franceschini[3], Jeff Haessler[2], Barbara V. Howard[16], Karen C. Johnson[17], Sandra Laston[10], Laurence N. Kolonel[18], Elisa T. Lee[19], Jean W. MacCluer[10], Teri A. Manolio[5], Sarah A. Pendergrass[1], Miguel Quibrera[20], Ralph V. Shohet[21], Lynne R. Wilkens[18], Christopher A. Haiman[4], Loïc Le Marchand[18], Steven Buyske[22], Charles Kooperberg[2], Kari E. North[3,23], Dana C. Crawford[1,24]*

1 Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **3** Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, United States of America, **4** Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **5** Office of Population Genomics, National Human Genome Research Institute, Bethesda, Maryland, United States of America, **6** Information Sciences Institute, University of Southern California, Los Angeles, California, United States of America, **7** Missouri Breaks Industries Research, Timber Lake, South Dakota, United States of America, **8** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, **9** Sponsored Programs, Baylor College of Medicine, Houston, Texas, United States of America, **10** Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America, **11** Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, **12** The Translational Genomics Research Institute, Phoenix, Arizona, United States of America, **13** Department of Family Medicine and Community Health, Alpert Medical School of Brown University School of Medicine, Providence, Rhode Island, United States of America, **14** Institute of Molecular Medicine, University of Texas Health Sciences Center at Houston, Texas, United States of America, **15** Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center, Houston, Texas, United States of America, **16** Medstar Research Institute, Washington, D.C., United States of America, **17** Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **18** Epidemiology Program, University of Hawaii Cancer Center, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, **19** University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, **20** School of Public Health, University of North Carolina, Chapel Hill, North Carolina, United States of America, **21** Center of Cardiovascular Research, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, **22** Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, United States of America, **23** Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina, United States of America, **24** Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America

Abstract

For the past five years, genome-wide association studies (GWAS) have identified hundreds of common variants associated with human diseases and traits, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglyceride (TG) levels. Approximately 95 loci associated with lipid levels have been identified primarily among populations of European ancestry. The Population Architecture using Genomics and Epidemiology (PAGE) study was established in 2008 to characterize GWAS–identified variants in diverse population-based studies. We genotyped 49 GWAS–identified SNPs associated with one or more lipid traits in at least two PAGE studies and across six racial/ethnic groups. We performed a meta-analysis testing for SNP associations with fasting HDL-C, LDL-C, and ln(TG) levels in self-identified European American (~20,000), African American (~9,000), American Indian (~6,000), Mexican American/Hispanic (~2,500), Japanese/East Asian (~690), and Pacific Islander/Native Hawaiian (~175) adults, regardless of lipid-lowering medication use. We replicated 55 of 60 (92%) SNP associations tested in European Americans at $p<0.05$. Despite sufficient power, we were unable to replicate *ABCA1* rs4149268 and rs1883025, *CETP* rs1864163, and *TTC39B* rs471364 previously associated with HDL-C and *MAFB* rs6102059 previously associated with LDL-C. Based on significance ($p<0.05$) and consistent direction of effect, a majority of replicated genotype-phenotype associations for HDL-C, LDL-C, and ln(TG) in European Americans generalized to African Americans (48%, 61%, and 57%), American Indians (45%, 64%, and 77%), and Mexican Americans/Hispanics (57%, 56%, and 86%). Overall, 16 associations generalized across all three populations. For the associations that did not generalize, differences in effect sizes, allele frequencies, and linkage disequilibrium offer clues to the next generation of association studies for these traits.

GENETIC ANALYSIS OF AFRICAN POPULATIONS: HUMAN EVOLUTION AND COMPLEX DISEASE

Sarah A. Tishkoff and Scott M. Williams*[‡§]

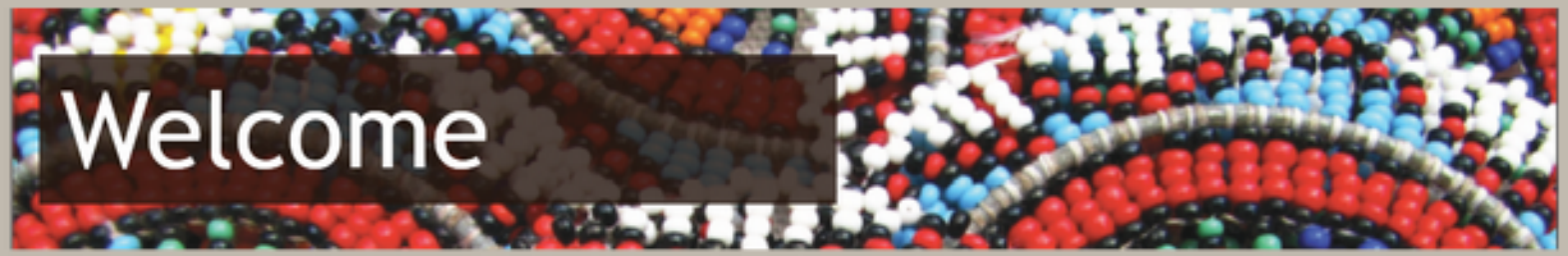
Africa is one of the most ethnically and genetically diverse regions of the world. It is thought to be the ancestral homeland of all modern humans, and is the homeland of millions of people of the recent African diaspora. Because of the central role of African populations in human history, characterizing their patterns of genetic diversity and linkage disequilibrium is crucial for reconstructing human evolution and for understanding the genetic basis of complex diseases.

PMID:12154384

H3Africa

Human Heredity & Health in Africa

Home | About | Consortium | Resources | Events | Contacts | Login



The Human Heredity and Health in Africa (H3Africa) Initiative aims to facilitate a contemporary research approach to the study of genomics and environmental determinants of common diseases with the goal of improving the health of African populations. To accomplish this, the H3Africa Initiative aims to contribute to the development of the necessary expertise among African scientists, and to establish networks of African investigators.

Search... Search

Current Funding Opportunities

For bioinformatics queries please contact the **H3ABioNet helpdesk**

Join our mailing list here!

News

Professor Alash'le Abimiku

OPEN ACCESS Freely available online PLoS GENETICS

Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study

Logan Dumitrescu[1], Cara L. Carty[2], Kira Taylor[3], Fredrick R. Schumacher[4], Lucia A. Hindorff[5], José L. Ambite[6], Garnet Anderson[2], Lyle G. Best[7], Kristin Brown-Gentry[1], Petra Bůžková[8], Christopher S. Carlson[2], Barbara Cochran[9], Shelley A. Cole[10], Richard B. Devereux[11], Dave Duggan[12], Charles B. Eaton[13], Myriam Fornage[14,15], Nora Franceschini[3], Jeff Haessler[2], Barbara V. Howard[16], Karen C. Johnson[17], Sandra Laston[10], Laurence N. Kolonel[18], Elisa T. Lee[19], Jean W. MacCluer[10], Teri A. Manolio[5], Sarah A. Pendergrass[1], Miguel Quibrera[20], Ralph V. Shohet[21], Lynne R. Wilkens[18], Christopher A. Haiman[4], Loïc Le Marchand[18], Steven Buyske[22], Charles Kooperberg[2], Kari E. North[3,23], Dana C. Crawford[1,24]*

1 Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **3** Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, United States of America, **4** Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **5** Office of Population Genomics, National Human Genome Research Institute, Bethesda, Maryland, United States of America, **6** Information Sciences Institute, University of Southern California, Los Angeles, California, United States of America, **7** Missouri Breaks Industries Research, Timber Lake, South Dakota, United States of America, **8** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, **9** Sponsored Programs, Baylor College of Medicine, Houston, Texas, United States of America, **10** Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America, **11** Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, **12** The Translational Genomics Research Institute, Phoenix, Arizona, United States of America, **13** Department of Family Medicine and Community Health, Alpert Medical School of Brown University School of Medicine, Providence, Rhode Island, United States of America, **14** Institute of Molecular Medicine, University of Texas Health Sciences Center at Houston, Texas, United States of America, **15** Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center, Houston, Texas, United States of America, **16** Medstar Research Institute, Washington, D.C., United States of America, **17** Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **18** Epidemiology Program, University of Hawaii Cancer Center, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, **19** University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, **20** School of Public Health, University of North Carolina, Chapel Hill, North Carolina, United States of America, **21** Center of Cardiovascular Research, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, **22** Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, United States of America, **23** Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina, United States of America, **24** Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America

Abstract

For the past five years, genome-wide association studies (GWAS) have identified hundreds of common variants associated with human diseases and traits, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglyceride (TG) levels. Approximately 95 loci associated with lipid levels have been identified primarily among populations of European ancestry. The Population Architecture using Genomics and Epidemiology (PAGE) study was established in 2008 to characterize GWAS–identified variants in diverse population-based studies. We genotyped 49 GWAS–identified SNPs associated with one or more lipid traits in at least two PAGE studies and across six racial/ethnic groups. We performed a meta-analysis testing for SNP associations with fasting HDL-C, LDL-C, and ln(TG) levels in self-identified European American (~20,000), African American (~9,000), American Indian (~6,000), Mexican American/Hispanic (~2,500), Japanese/East Asian (~690), and Pacific Islander/Native Hawaiian (~175) adults, regardless of lipid-lowering medication use. We replicated 55 of 60 (92%) SNP associations tested in European Americans at $p<0.05$. Despite sufficient power, we were unable to replicate *ABCA1* rs4149268 and rs1883025, *CETP* rs1864163, and *TTC39B* rs471364 previously associated with HDL-C and *MAFB* rs6102059 previously associated with LDL-C. Based on significance ($p<0.05$) and consistent direction of effect, a majority of replicated genotype-phentoype associations for HDL-C, LDL-C, and ln(TG) in European Americans generalized to African Americans (48%, 51%, and 57%), American Indians (45%, 64%, and 77%), and Mexican Americans/Hispanics (57%, 56%, and 86%). Overall, 16 associations generalized across all three populations. For the associations that did not generalize, differences in effect sizes, allele frequencies, and linkage disequilibrium offer clues to the next generation of association studies for these traits.

E.g., G

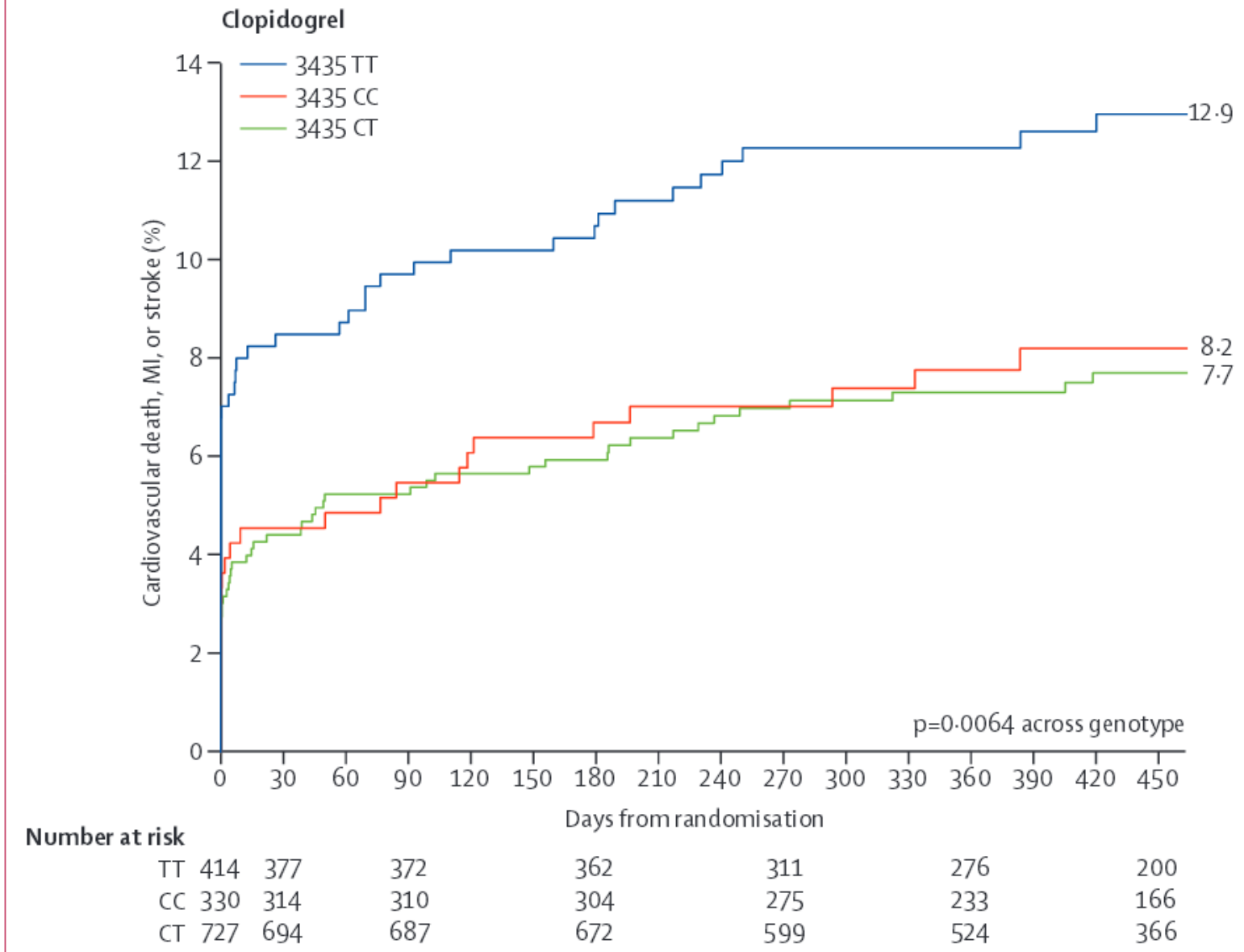
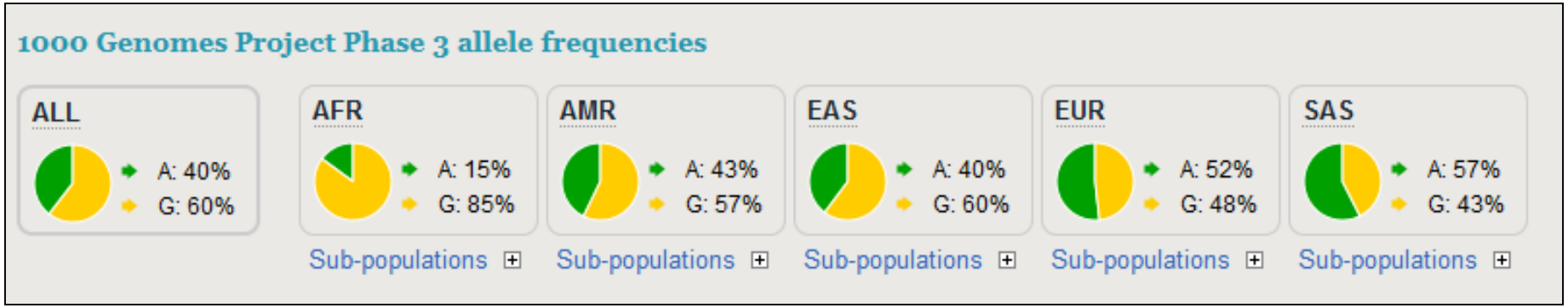


***Figure 1:* *ABCB1* 3435C→T and cardiovascular outcomes in patients treated with clopidogrel**
Cumulative risk of cardiovascular death, myocardial infarction (MI), or stroke for each genotype, with a p value across genotype.

The rs1045642 A Allele: Substantial Variation Across Global Populations

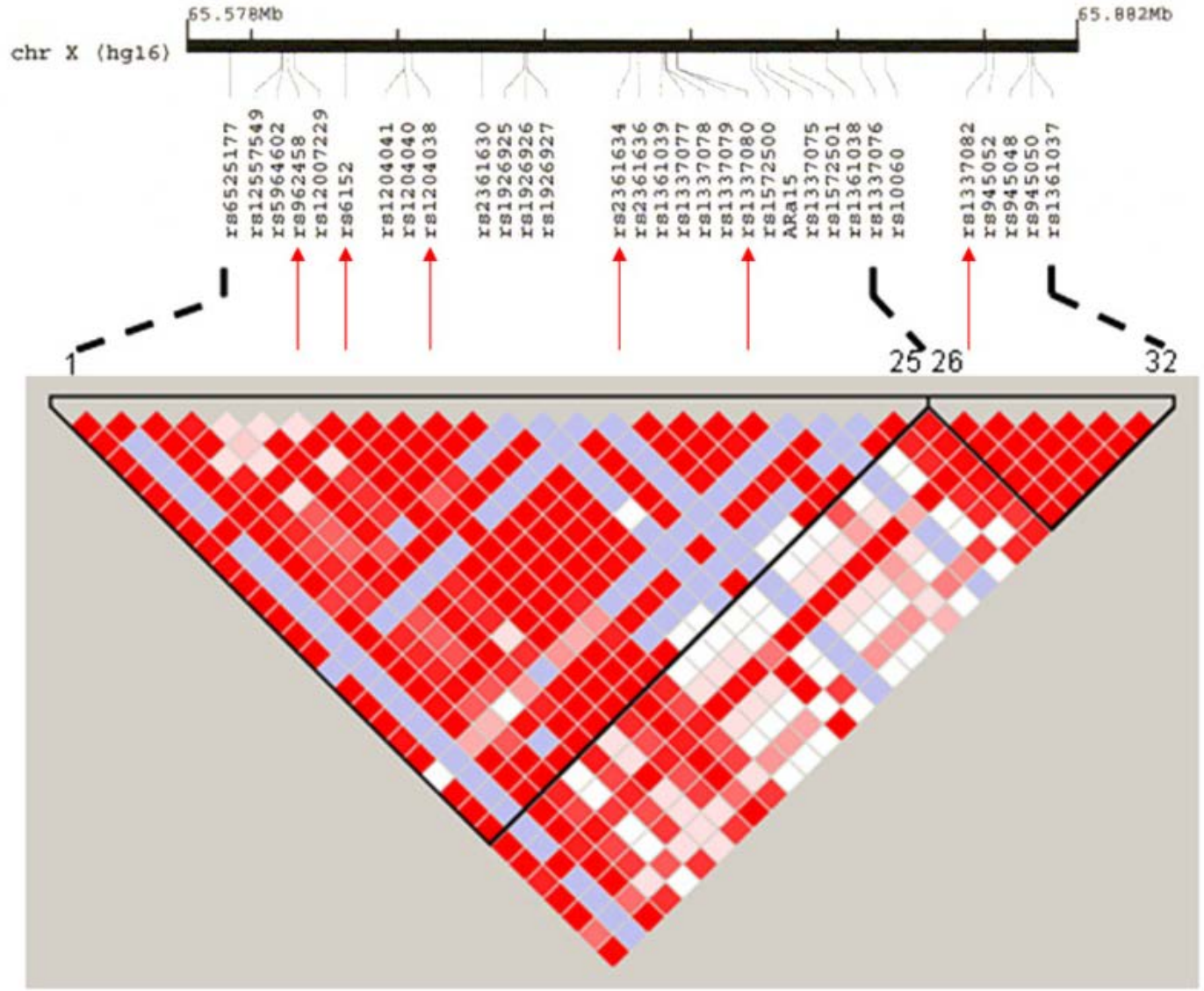


Approximately **33%** (i.e. 0.57^2) of the SAS population is homozygous for the causal allele compared to **2.3%** (i.e. 0.15^2) of the AFR population.

Linkage Disequilibrium (LD)

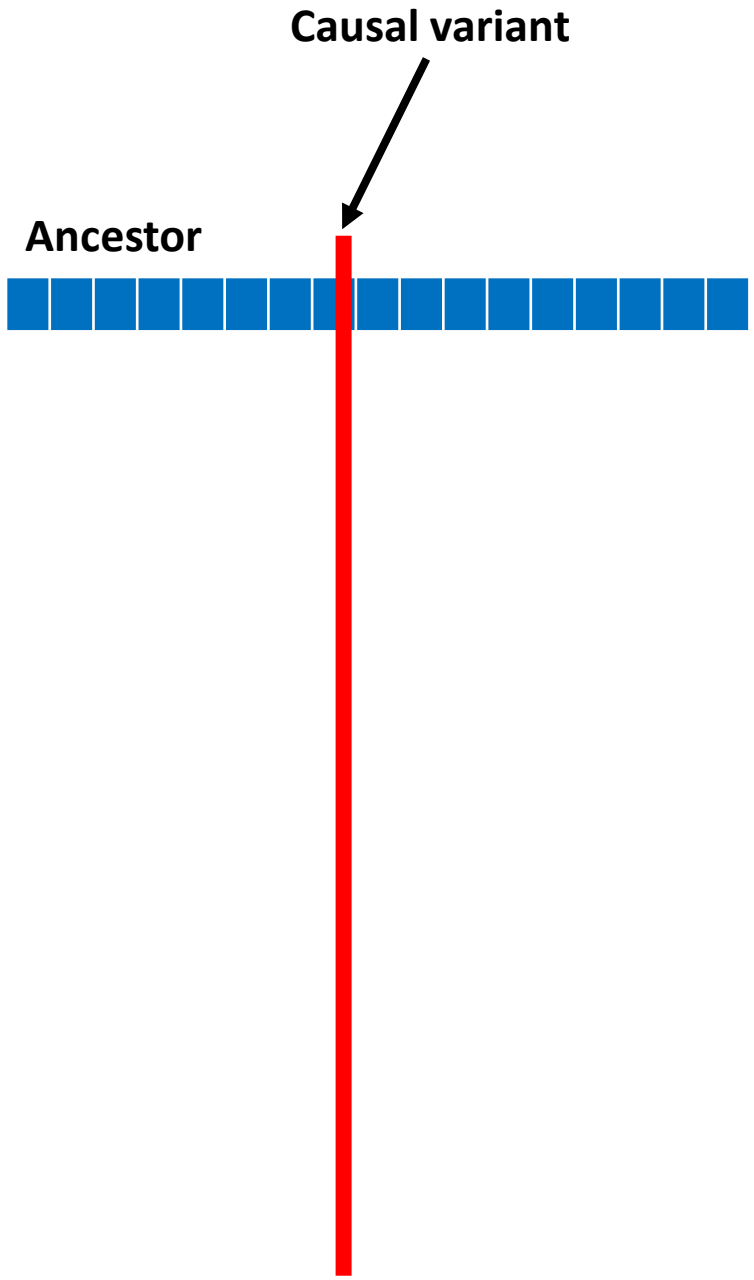
- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.

Linkage Disequilibrium (LD): SNPs are Inherited in Blocks



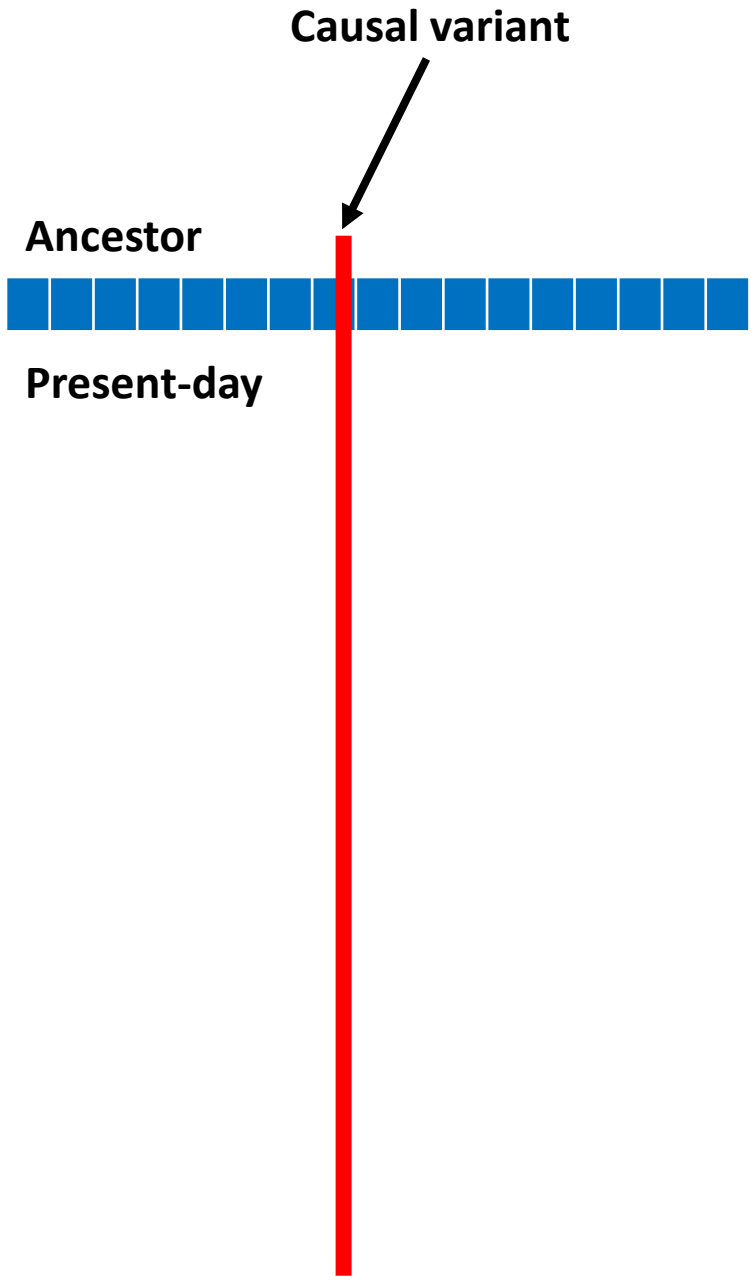
Linkage Disequilibrium (LD)

- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.



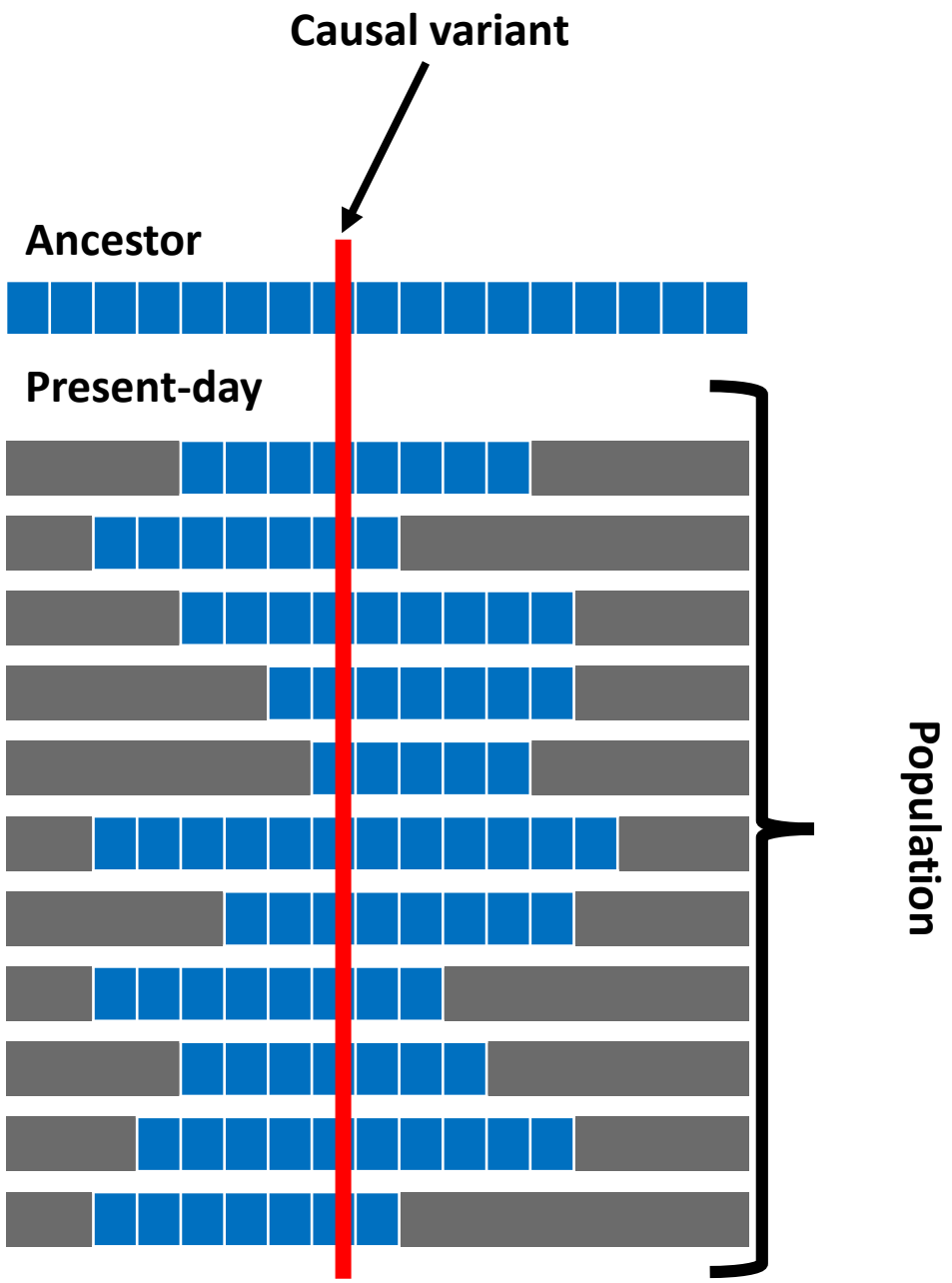
Linkage Disequilibrium (LD)

- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.



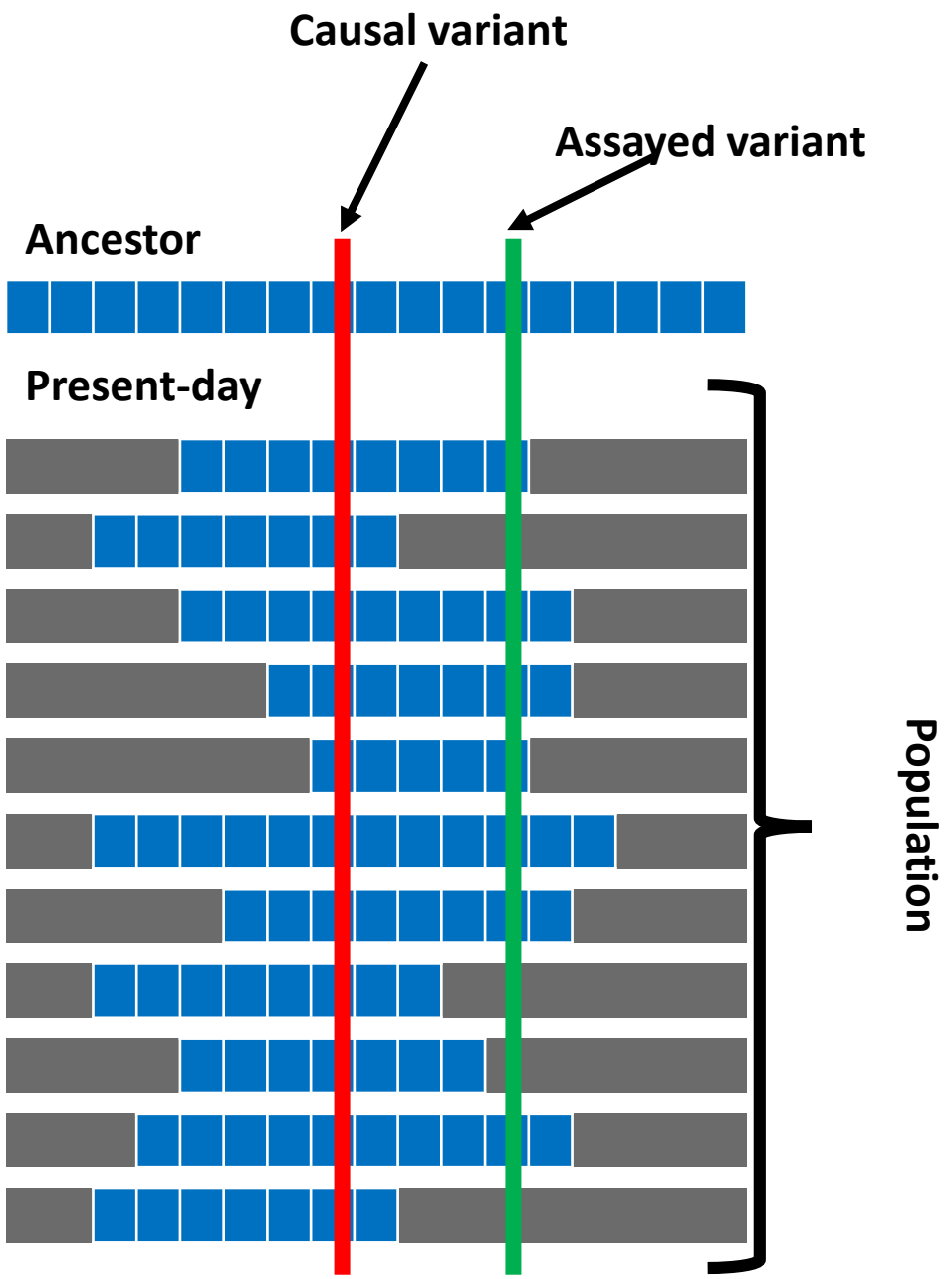
Linkage Disequilibrium (LD)

- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.



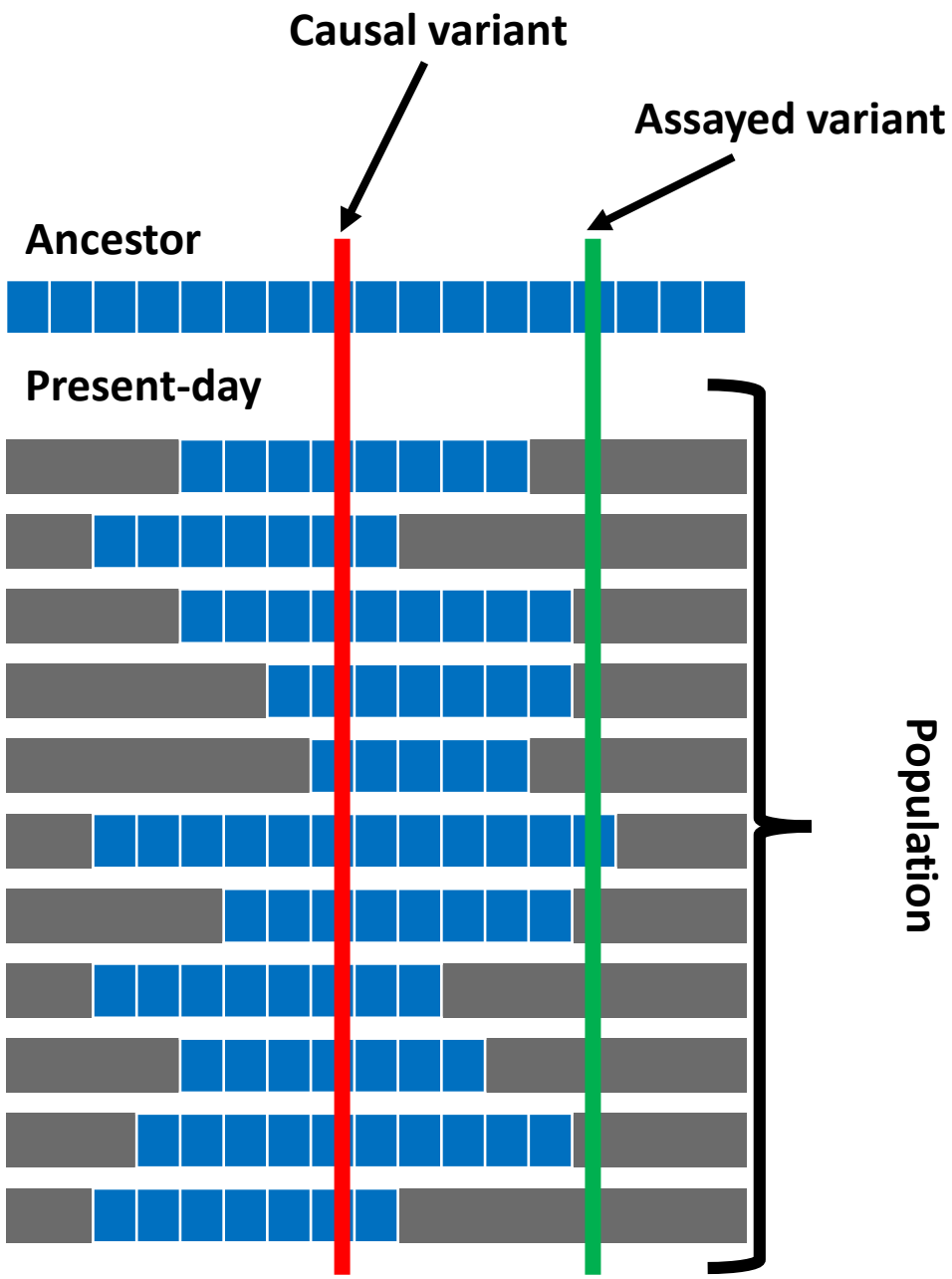
Linkage Disequilibrium (LD)

- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.



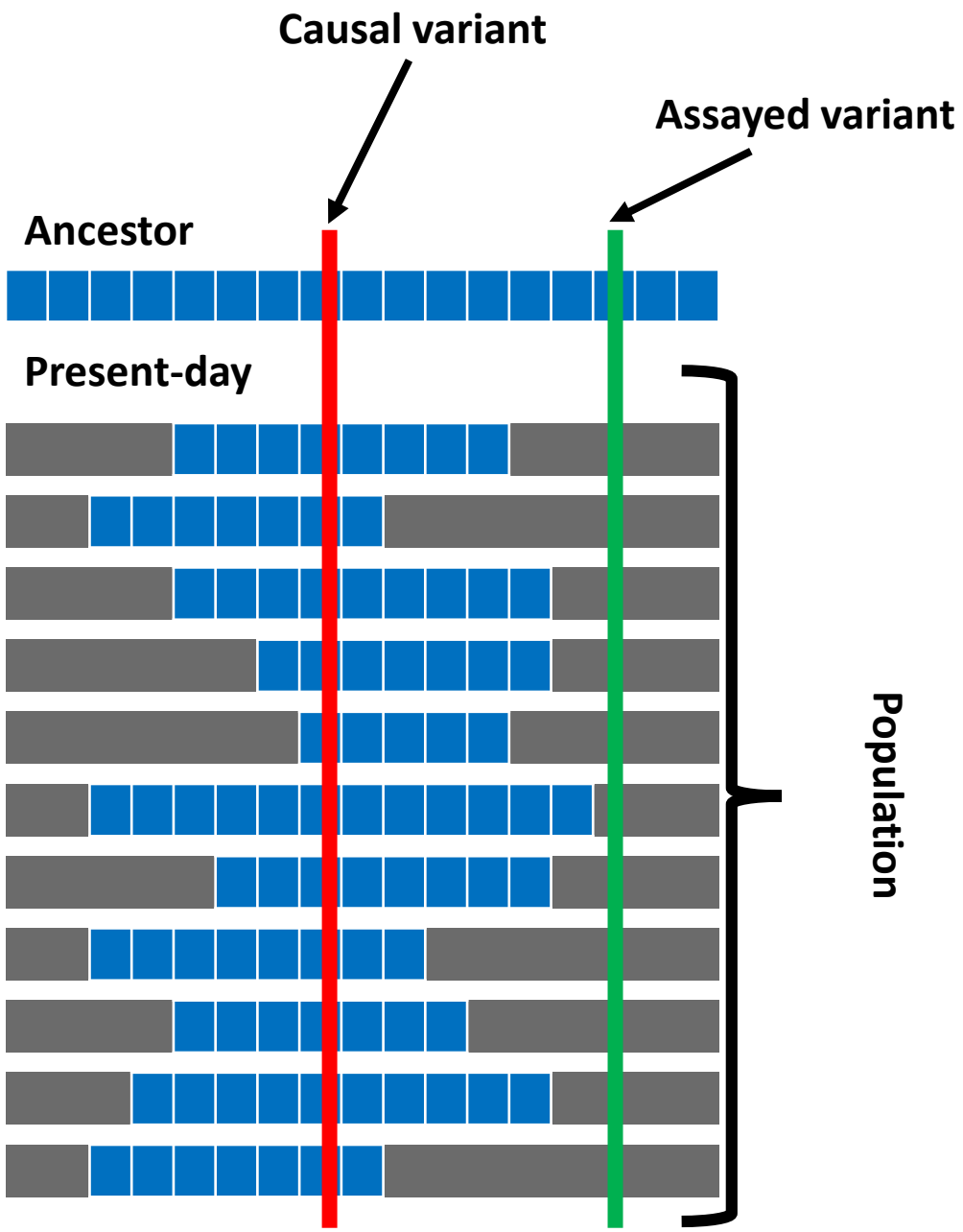
Linkage Disequilibrium (LD)

- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.



Linkage Disequilibrium (LD)

- Non-random assortment of alleles at 2+ SNPs
- Population-specific!
- The closer the SNPs, the stronger the LD since recombination will have occurred at a lower rate
- Two markers are in LD if knowing the allele at one marker allows you to predict the allele at the other marker
  - E.g. in a population where there are AB, Ab, and aB haplotypes at adjacent markers, but no ab haplotypes, if we know an individual has a b allele, we know that s/he also has at least one A allele.



Race/Ethi

OPEN ACCESS Freely available online PLOS BIOLOGY

Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study

Christopher S. Carlson[1*], Tara C. Matise[2], Kari E. North[3], Christopher A. Haiman[4], Megan D. Fesinmeyer[5], Steven Buyske[6], Fredrick R. Schumacher[4], Ulrike Peters[1], Nora Franceschini[3], Marylyn D. Ritchie[7], David J. Duggan[8], Kylee L. Spencer[9], Logan Dumitrescu[10], Charles B. Eaton[11], Fridtjof Thomas[12], Alicia Young[1], Cara Carty[1], Gerardo Heiss[3], Loic Le Marchand[13], Dana C. Crawford[10], Lucia A. Hindorff[14],

Take-home messages:

1 – Genes generalize, but variation in SNPs exist.

2 – Studies in non-European populations are needed.

A. Implications for gene-environment?

3 – Genetic analyses should be population-specific.

A. Analyses also need to address within-population variation (e.g. with ancestral principal components.)

Citation: Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, et al. (2013) Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. PLoS Biol 11(9): e1001661. doi:10.1371/journal.pbio.1001661

nature Vol 456|6 November 2008|doi:10.1038/nature07331

LETTERS

Genes mirror geography within Europe

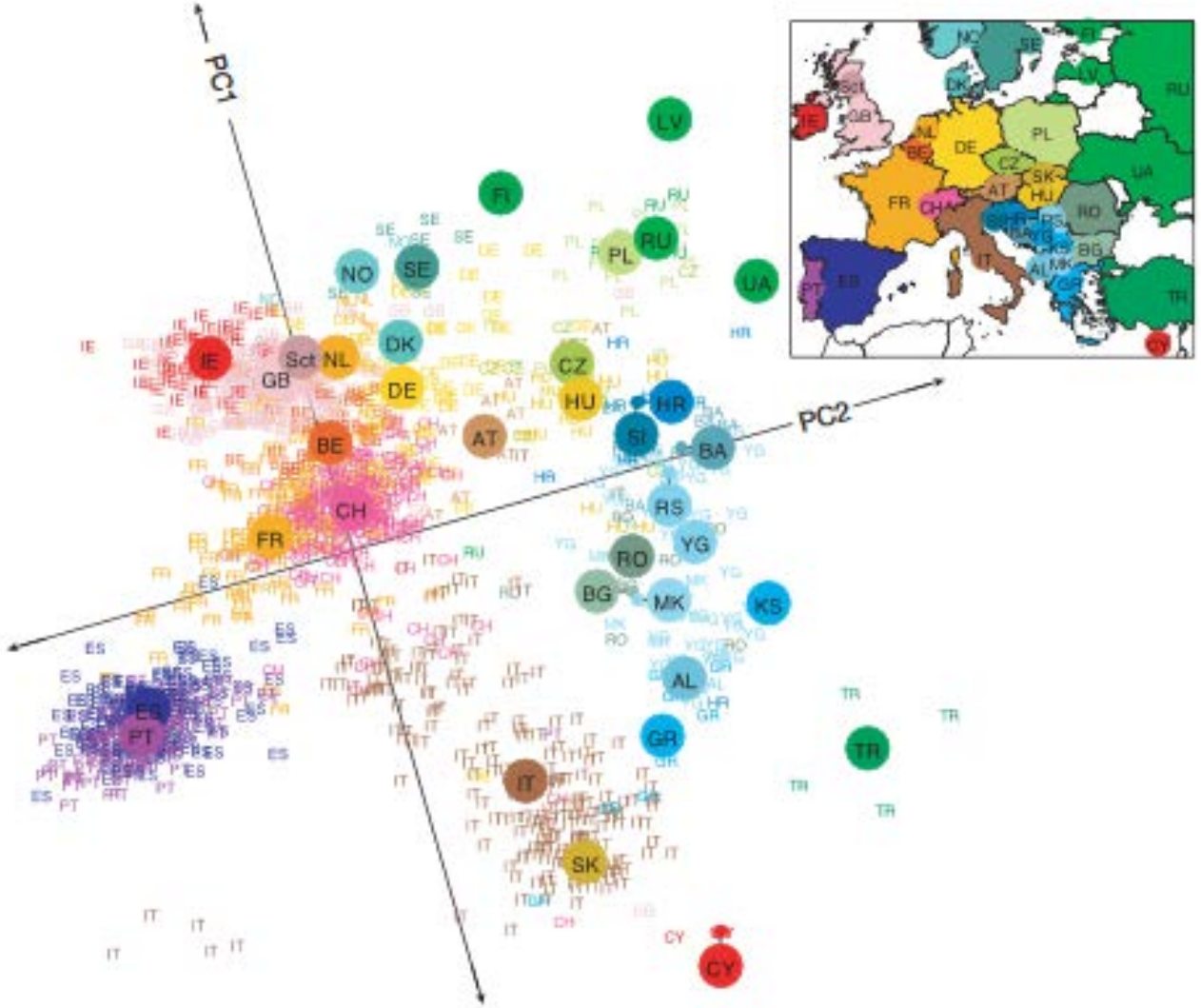
John Novembre[1,2], Toby Johnson[4,5,6], Katarzyna Bryc[7], Zoltán Kutalik[4,6], Adam R. Boyko[7], Adam Auton[7], Amit Indap[7], Karen S. King[8], Sven Bergmann[4,6], Matthew R. Nelson[8], Matthew Stephens[2,3] & Carlos D. Bustamante[7]

Understanding the genetic structure of human populations is of fundamental interest to medical, forensic and anthropological sciences. Advances in high-throughput genotyping technology have markedly improved our understanding of global patterns of human genetic variation and suggest the potential to use large samples to uncover variation among closely spaced populations[1–5]. Here we characterize genetic variation in a sample of 3,000 European individuals genotyped at over half a million variable DNA sites in the human genome. Despite low average levels of genetic differentiation among Europeans, we find a close correspondence between genetic and geographic distances; indeed, a geographical map of Europe arises naturally as an efficient two-dimensional summary of genetic variation in Europeans. The results emphasize that when mapping the genetic basis of a disease phenotype, spurious associations can arise if genetic structure is not properly accounted for. In addition, the results are relevant to the prospects of genetic ancestry testing[6]; an individual's DNA can be used to infer their geographic origin with surprising accuracy—often to within a few hundred kilometres.

The resulting figure bears a notable resemblance to a geographic map of Europe (Fig. 1a). Individuals from the same geographic region cluster together and major populations are distinguishable. Geographically adjacent populations typically abut each other, and recognizable geographical features of Europe such as the Iberian peninsula, the Italian peninsula, southeastern Europe, Cyprus and Turkey are apparent. The data reveal structure even among French-, German- and Italian-speaking groups within Switzerland (Fig. 1b), and between Ireland and the United Kingdom (Fig. 1a, IE and GB). Within some countries individuals are strongly differentiated along the principal component (PC) axes, suggesting that in some cases the resolution of the genetic data may exceed that of the available geographic information.

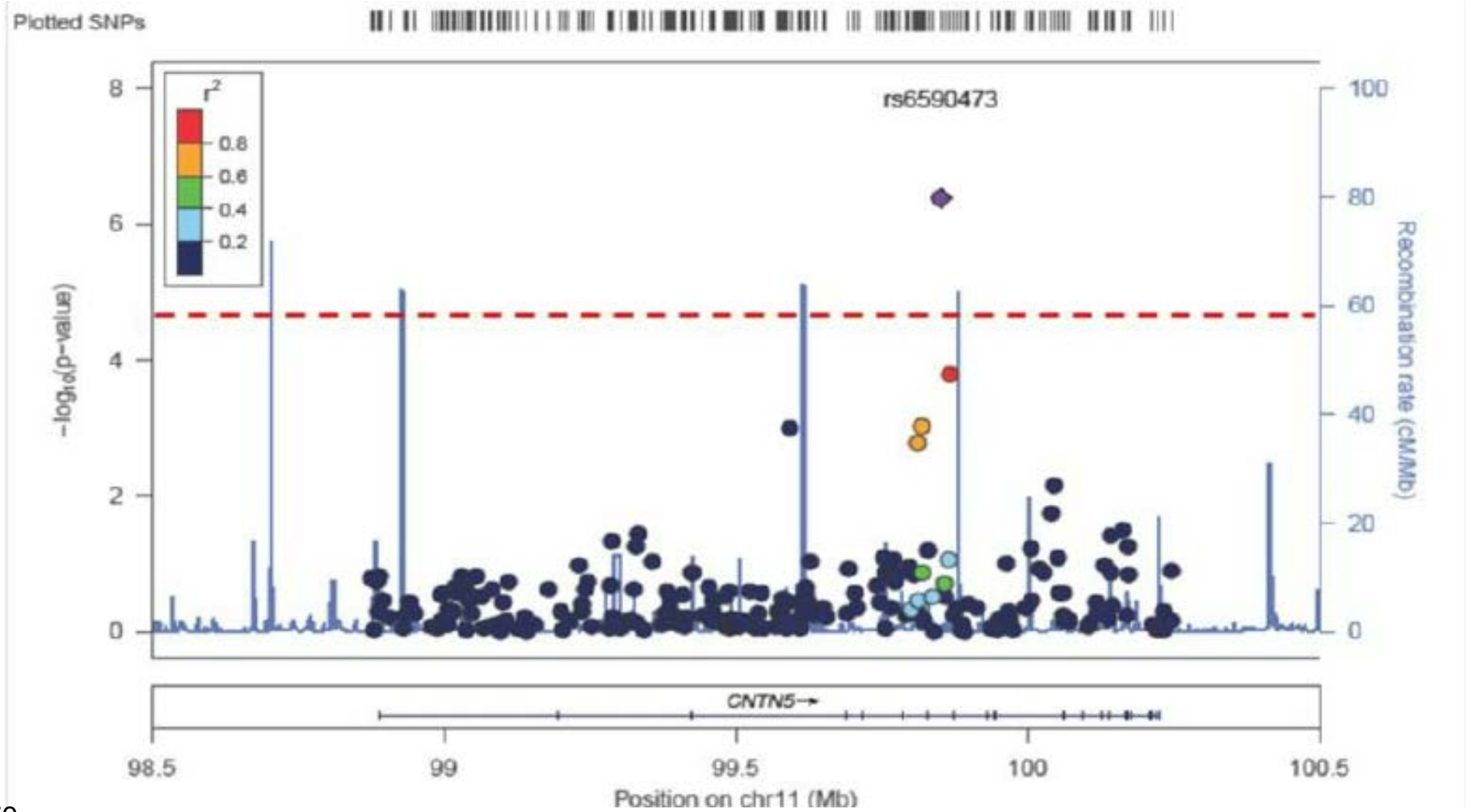
When we quantitatively compare the geographic position of countries with their PC-based genetic positions, we observe few prominent differences between the two (Supplementary Fig. 1), and those that exist can be explained either by small sample sizes (for example, Slovakia (SK)) or by the coarseness of our geographic data (a problem for large countries, for example, Russia (RU)); see

Population Structure within Europe



PMID:18758442

Do Not Ignore Intergenic Regions



PMID:25337070

Coronary Heart Disease GWAS and 9p21

The NEW ENGLAND JOURNAL of MEDICI

ESTABLISHED IN 1812 AUGUST 2, 2007

Genomewide Association Analysis of Coronary A

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

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.

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

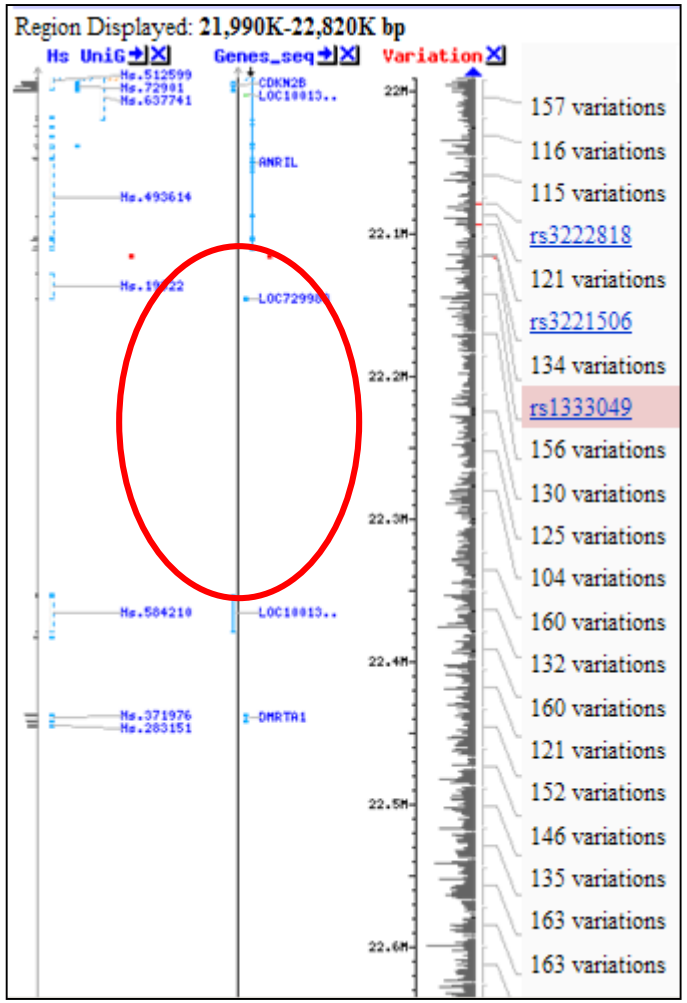
A Common Allele on Chromosome 9 Associated with Coronary Heart Disease

Ruth McPherson,[1*†] Alexander Pertsemlidis,[2*] Nihan Kavaslar,[1] Alexandre Stewart,[1]
Robert Roberts,[1] David R. Cox,[3] David A. Hinds,[3] Len A. Pennacchio,[4,5] Anne Tybjaerg-Hansen,[6]

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

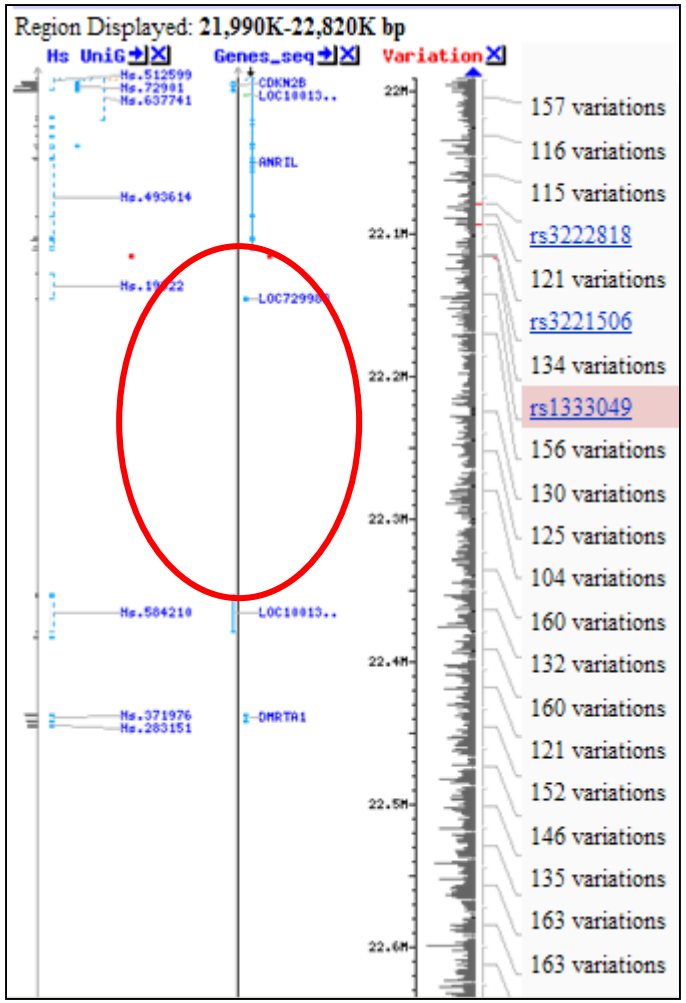
Anna Helgadottir,[1*] Gudmar Thorleifsson,[1*] Andrei Manolescu,[1*] Solveig Gretarsdottir,[1]
Thorarinn Blondal,[1] Aslaug Jonasdottir,[1] Adalbjorg Jonasdottir,[1] Asgeir Sigurdsson,[1]
Adam Baker,[1] Arnar Palsson,[1] Gisli Masson,[1] Daniel F. Gudbjartsson,[1] Kristinn P. Magnusson,[1]
Karl Andersen,[2] 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 Thorgeirsson,[2] Unnur Thorsteinsdottir,[1]
Augustine Kong,[1†] Kari Stefansson[1†]

9p21.3: Replicated Locus with Zero Prior Biologic Plausibility



The risk interval narrowed to a block approximately 58 kb wide that did not contain any annotated genes.

9p21.3: Replicated Locus with Zero Prior Biologic Plausibility



Coronary Heart Disease

9p21.3 Coronary Artery Disease Risk Variants Disrupt TEAD Transcription Factor–Dependent Transforming Growth Factor β Regulation of p16 Expression in Human Aortic Smooth Muscle Cells

Naif A. M. Almontashiri, PhD; Darlène Antoine, MSc; Xun Zhou, MSc; Ragnar O. Vilmundarson, MSc; Sean X. Zhang; Kennedy N. Hao; Hsiao-Huei Chen, PhD; Alexandre F. R. Stewart, PhD

Background—The mechanism whereby the 9p21.3 locus confers risk for coronary artery disease remains incompletely understood. Risk alleles are associated with reduced expression of the cell cycle suppressor genes CDKN2A (p16 and p14) and CDKN2B (p15) and increased vascular smooth muscle cell proliferation. We asked whether risk alleles disrupt transcription factor binding to account for this effect.

Methods and Results—A bioinformatic screen was used to predict which of 59 single nucleotide polymorphisms at the 9p21.3 locus disrupt (or create) transcription factor binding sites. Electrophoretic mobility shift and luciferase reporter assays examined the binding and functionality of the predicted regulatory sequences. Primary human aortic smooth muscle cells (HAoSMCs) were genotyped for 9p21.3, and HAoSMCs homozygous for the risk allele showed reduced p15 and p16 levels and increased proliferation. rs10811656 and rs4977757 disrupted functional TEF-1 TEC1 AbaA domain (TEAD) transcription factor binding sites. TEAD3 and TEAD4 overexpression induced p16 in HAoSMCs homozygous for the nonrisk allele, but not for the risk allele. Transforming growth factor β, known to activate p16 and also to interact with TEAD factors, failed to induce p16 or to inhibit proliferation of HAoSMCs homozygous for the risk allele. Knockdown of TEAD3 blocked transforming growth factor β–induced p16 mRNA and protein expression, and dual knockdown of TEAD3 and TEAD4 markedly reduced p16 expression in heterozygous HAoSMCs.

Conclusions—Here, we identify a novel mechanism whereby sequences at the 9p21.3 risk locus disrupt TEAD factor binding and TEAD3-dependent transforming growth factor β induction of p16 in HAoSMCs. This mechanism accounts, in part, for the 9p21.3 coronary artery disease risk. (***Circulation***.2015;132:1969-1978.DOI:10.1161/CIRCULATIONAHA.114.015023.)

Key Words: atherosclerosis ■ coronary disease ■ genetics ■ molecular biology ■ smooth muscle cells

Family Structure/Clustering

- Add Health GWAS data has a non-negligible number of related participants
  - Failure to address lack of independence between family members leads to anti-conservative *P*-values
  - Most “canned” software (e.g. PLINK, ProbAbel) does not address relatedness
- Option 1 (easiest): exclude all but one member of each first-degree relative set (kinship matrix provided on dbGap) and proceed as unrelated.
- Option 2 (more work, more power): model the family structure
- School clustering also requires extension of models to include additional variance components

Analytic Pipeline: Addresses Add Health Data Challenges

- GWAS tools have been published that can accommodate Add Health analysis challenges
- Implementation may be challenging if modest Unix/R/python expertise
- Scalability remains a challenge in GWAS setting.
  - Linear mixed models run locally can be used when examining a limited number of SNPs

Analytic Pipeline: Addresses Add Health Data Challenges

ARTICLE

Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models

Han Chen,[1,8] Chaolong Wang,[1,2,8] Matthew P. Conomos,[3] Adrienne M. Stilp,[3] Zilin Li,[1,4] Tamar Sofer,[3] Adam A. Szpiro,[3] Wei Chen,[5] John M. Brehm,[5] Juan C. Celedón,[5] Susan Redline,[6] George J. Papanicolaou,[7] Timothy A. Thornton,[3] Cathy C. Laurie,[3] Kenneth Rice,[3] and Xihong Lin[1,*]

Linear mixed models (LMMs) are widely used in genome-wide association studies (GWASs) to account for population structure and relatedness, for both continuous and binary traits. Motivated by the failure of LMMs to control type I errors in a GWAS of asthma, a binary trait, we show that LMMs are generally inappropriate for analyzing binary traits when population stratification leads to violation of the LMM's constant-residual variance assumption. To overcome this problem, we develop a computationally efficient logistic mixed model approach for genome-wide analysis of binary traits, the generalized linear mixed model association test (GMMAT). This approach fits a logistic mixed model once per GWAS and performs score tests under the null hypothesis of no association between a binary trait and individual genetic variants. We show in simulation studies and real data analysis that GMMAT effectively controls for population structure and relatedness when analyzing binary traits in a wide variety of study designs.

odate Add Health

/R/python

ining a limited

Ana a
Cha alth

Control f
for Binary
via Logist

Han Chen,[1,8]
Adam A. Szpi
George J. Pap

Linear mixed mod
edness, for both c
trait, we show tha
LMM's constant-r
approach for genc
logistic mixed mo
individual genetic
ture and relatedne

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Search:

Home | Install | Help | Developers | About

Home » Bioconductor 3.3 » Software Packages » GWASTools

GWASTools

platforms all | downloads top 5% | posts 4 / 1 / 0.2 / 1 | in Bioc 4.5 years
build ok | commits 1.50 | test coverage 69%

Tools for Genome Wide Association Studies

Bioconductor version: Release (3.3)

Classes for storing very large GWAS data sets and annotation, and functions for GWAS data cleaning and analysis.

Author: Stephanie M. Gogarten, Cathy Laurie, Tushar Bhangale, Matthew P. Conomos, Cecelia Laurie, Caitlin McHugh, Ian Painter, Xiuwen Zheng, Jess Shen, Rohit Swarnkar, Adrienne Stilp, Sarah Nelson

Maintainer: Stephanie M. Gogarten <sdmorris at u.washington.edu>, Adrienne Stilp <amstilp at u.washington.edu>

Citation (from within R, enter `citation("GWASTools")`):

Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, Zheng X, Crosslin DR, Levine D, Lumley T, Nelson SC, Rice K, Shen J, Swarnkar R, Weir BS and Laurie CC (2012). "GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies." *Bioinformatics*, **28**(24), pp. 3329-3331.

Documentation »

Bioconductor

- Package vignettes and manuals.
- Workflows for learning and use.
- Course and conference material.
- Videos.
- Community resources and tutorials.

R / CRAN packages and documentation

Support »

Please read the posting guide. Post questions about Bioconductor to one of the following locations:

- Support site - for questions about Bioconductor packages
- Bioc-devel mailing list - for package developers

Conclusions

- Add Health GWAS data offer a wealth of opportunities for advancing the understanding of human phenotypes and traits
  - Very unique resource: few studies of nationally representative populations beginning in adolescence are available
- Genomics data are challenging at first to use, but numerous resources exist
  - Consider establishing relationships with existing consortia/engaging a genetic epidemiologist etc.
- Genetics of “social science” traits, gene-environment interactions etc. remain largely unexplored