GWAS as a window to genetic architecture

NHGRI GWAS Catalog Webinar

July 18th 2013

Benjamin Neale

Massachusetts General Hospital

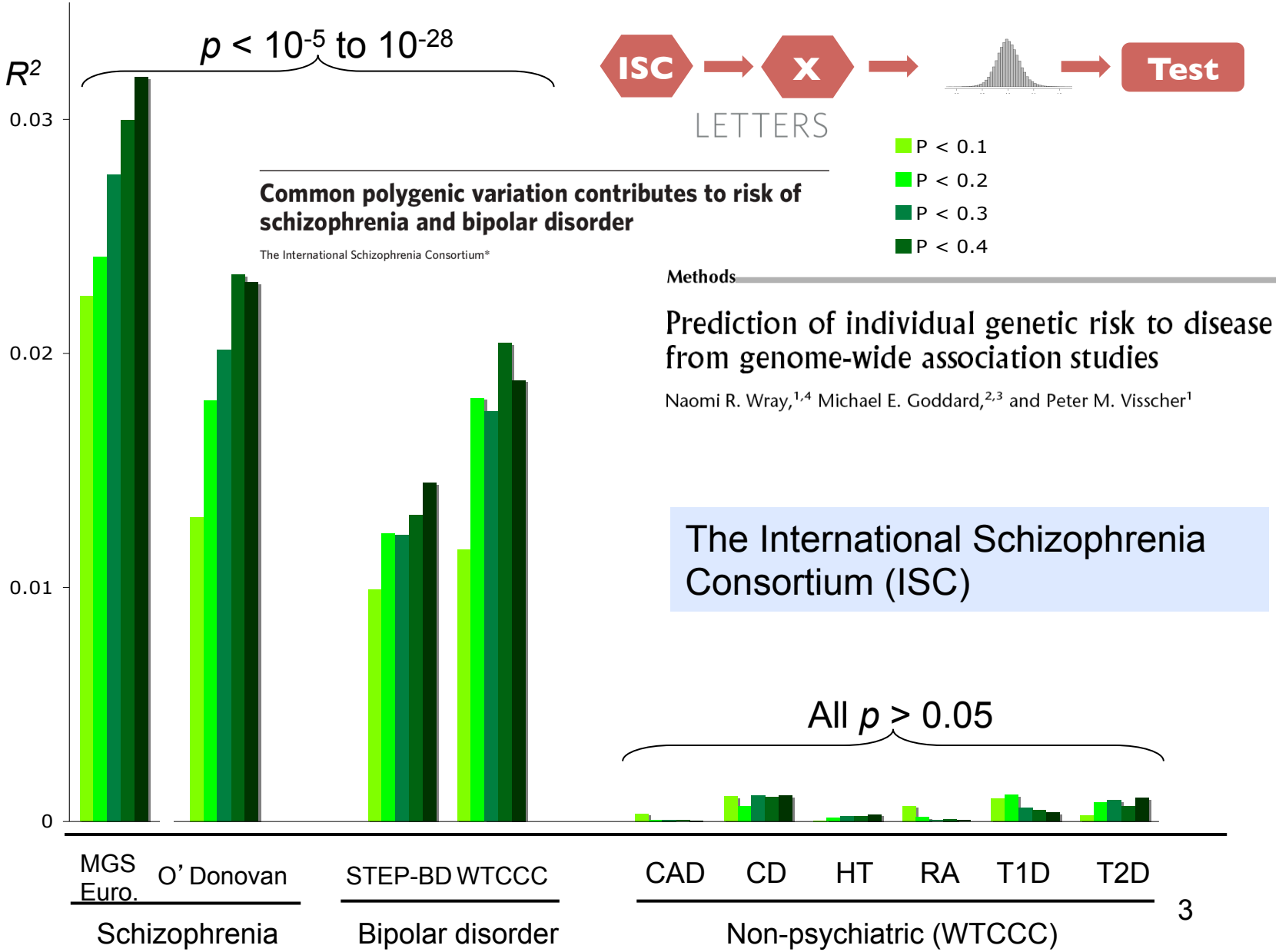
Harvard Medical School

BROAD INSTITUTE

Slide of contents

- Complex trait genetic architecture
  - Polygenic prediction
  - Heritability
- Mendelian randomization
  - Example from lipids

Polygenic prediction



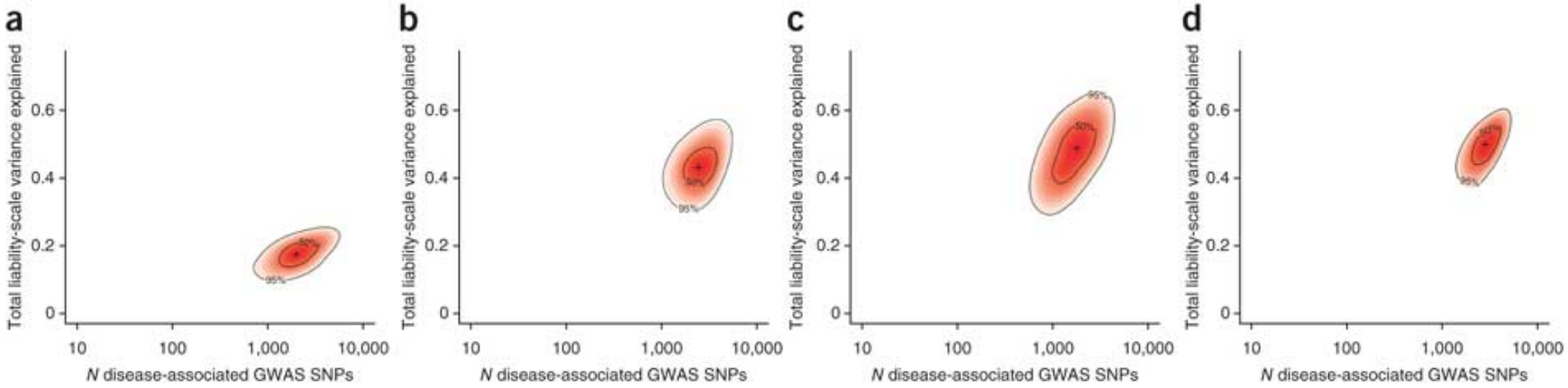
[ISC 2009]

ANALYSIS

nature genetics

Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis

Eli A Stahl[1–3]*, Daniel Wegmann[4], Gosia Trynka[5], Javier Gutierrez-Achury[5], Ron Do[2,6], Benjamin F Voight[7], Peter Kraft[8], Robert Chen[1–3], Henrik J Kallberg[9], Fina A S Kurreeman[1–3], Diabetes Genetics Replication and Meta-analysis Consortium[10], Myocardial Infarction Genetics Consortium[10], Sekar Kathiresan[2,6], Cisca Wijmenga[5], Peter K Gregersen[11], Lars Alfredsson[9], Katherine A Siminovitch[12], Jane Worthington[13], Paul I W de Bakker[2,3,14,15], Soumya Raychaudhuri[1–3,16] & Robert M Plenge[1–3,16]



Rheumatoid Arthritis Celiac Disease MI/CAD Type II Diabetes

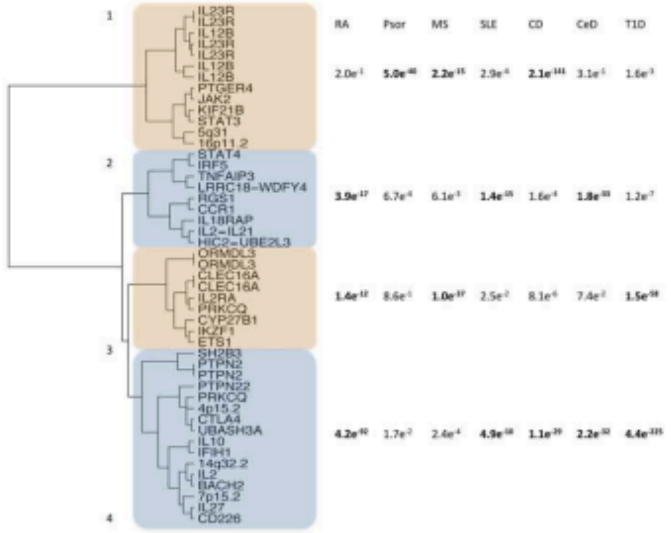
Overlap at locus and disease level can provide insight into biological overlap

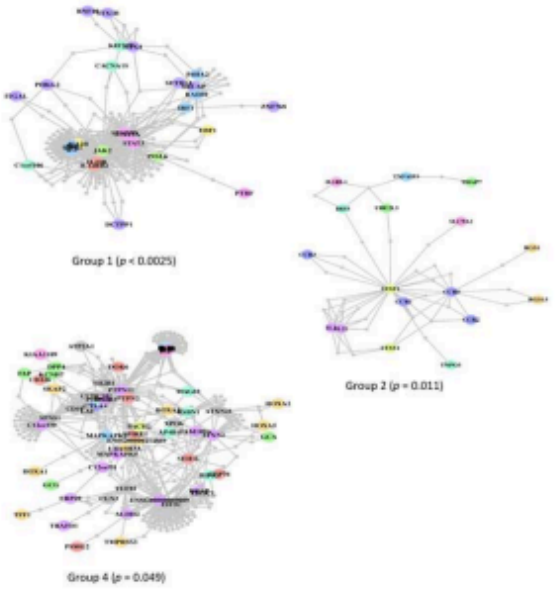
OPEN ACCESS Freely available online

PLoS GENETICS

Pervasive Sharing of Genetic Effects in Autoimmune Disease

Chris Cotsapas[1,2,3,4,5*], Benjamin F. Voight[1,2,3*], Elizabeth Rossin[1,2,3,6,7], Kasper Lage[2,8,9], Benjamin M. Neale[1,2,3,10], Chris Wallace[11], Gonçalo R. Abecasis[12], Jeffrey C. Barrett[13], Timothy Behrens[14], Judy Cho[5,15], Philip L. De Jager[3,16], James T. Elder[17], Robert R. Graham[14], Peter Gregersen[18], Lars Klareskog[19], Katherine A. Siminovitch[20], David A. van Heel[21], Cisca Wijmenga[22], Jane Worthington[23], John A. Todd[11], David A. Hafler[4], Stephen S. Rich[24], Mark J. Daly[1,2,3,10*], on behalf of the FOCiS Network of Consortia





Genome-wide Complex Trait Analysis (GCTA)

nature genetics

REPORT

GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

ARTICLE

Estimating Missing Heritability for Disease from Genome-wide Association Studies

Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

Genome-wide Complex Trait Analysis (GCTA)

nature genetics

REPORT

GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

ARTICLE

Estimating Missing Heritability for Disease from Genome-wide Association Studies

Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

Genetic similarity should correlate with phenotypic similarity

Genome-wide Complex Trait Analysis (GCTA)

nature genetics

REPORT

GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

ARTICLE

Estimating Missing Heritability for Disease from Genome-wide Association Studies

Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

Genetic similarity should correlate with phenotypic similarity

Caution: any artifact in the data that makes cases appear more similar to other cases or controls more similar to other controls will inflate estimates of heritability (e.g. batch effects)

Bivariate models to estimate genome-wide pleiotropy between disorders

Two traits

- Trait 1 = Cases and controls of disorder 1
- Trait 2 = Cases and controls of disorder 2

Traits measured on different sets of people

– linked through genetic relationships.

Can explore genetic relationships between disorders that are simply not possible with family data

- Low prevalence
- Ascertainment
- Confounding with common environment

Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood

S.H. Lee[1,*], J. Yang[2], M.E. Goddard[3], P.M. Visscher[1,2] and N.R. Wray[1]

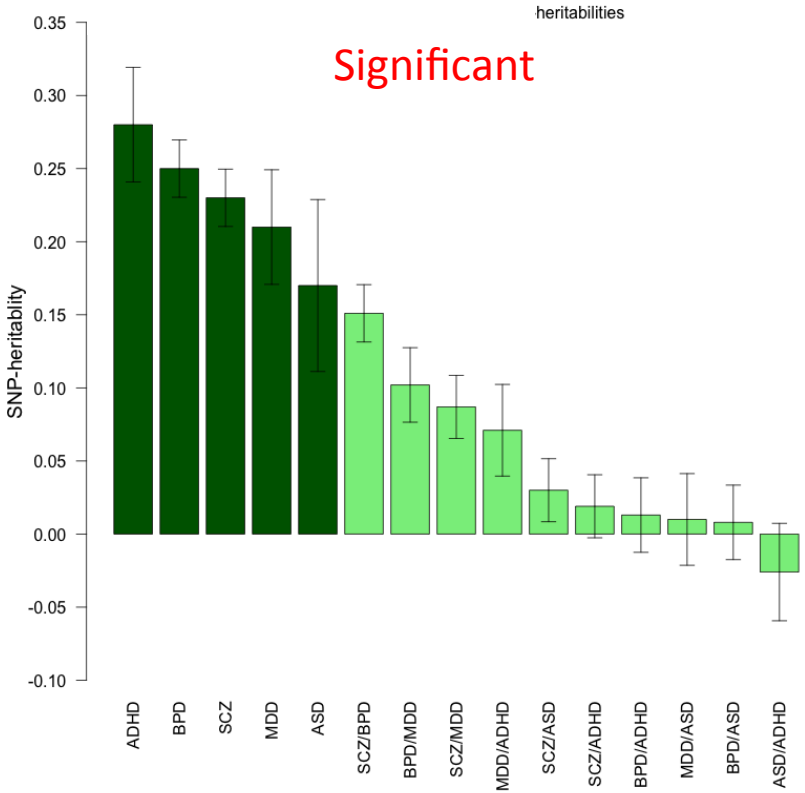
[1]Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia

[2]University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia

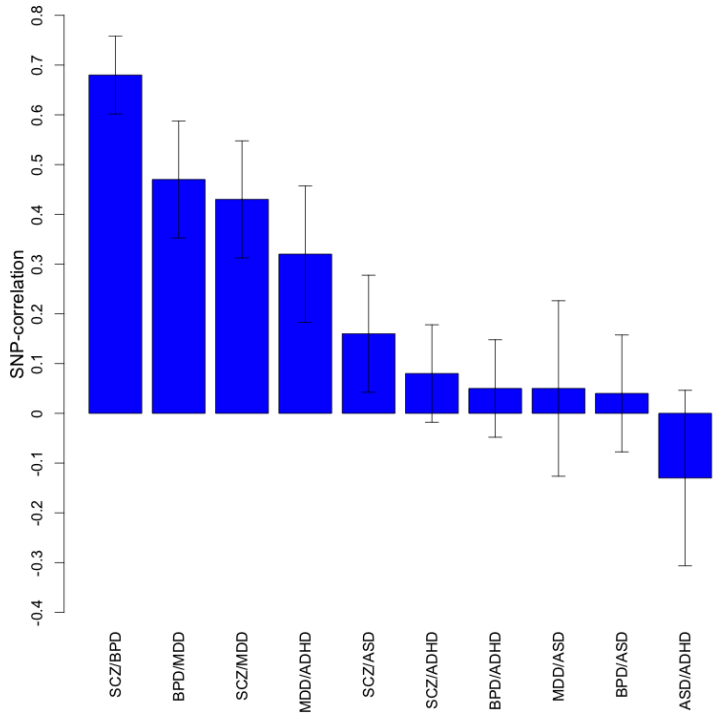
Bioinformatics Advance Access published July 26, 2012

Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs

SNP-chip heritabilities



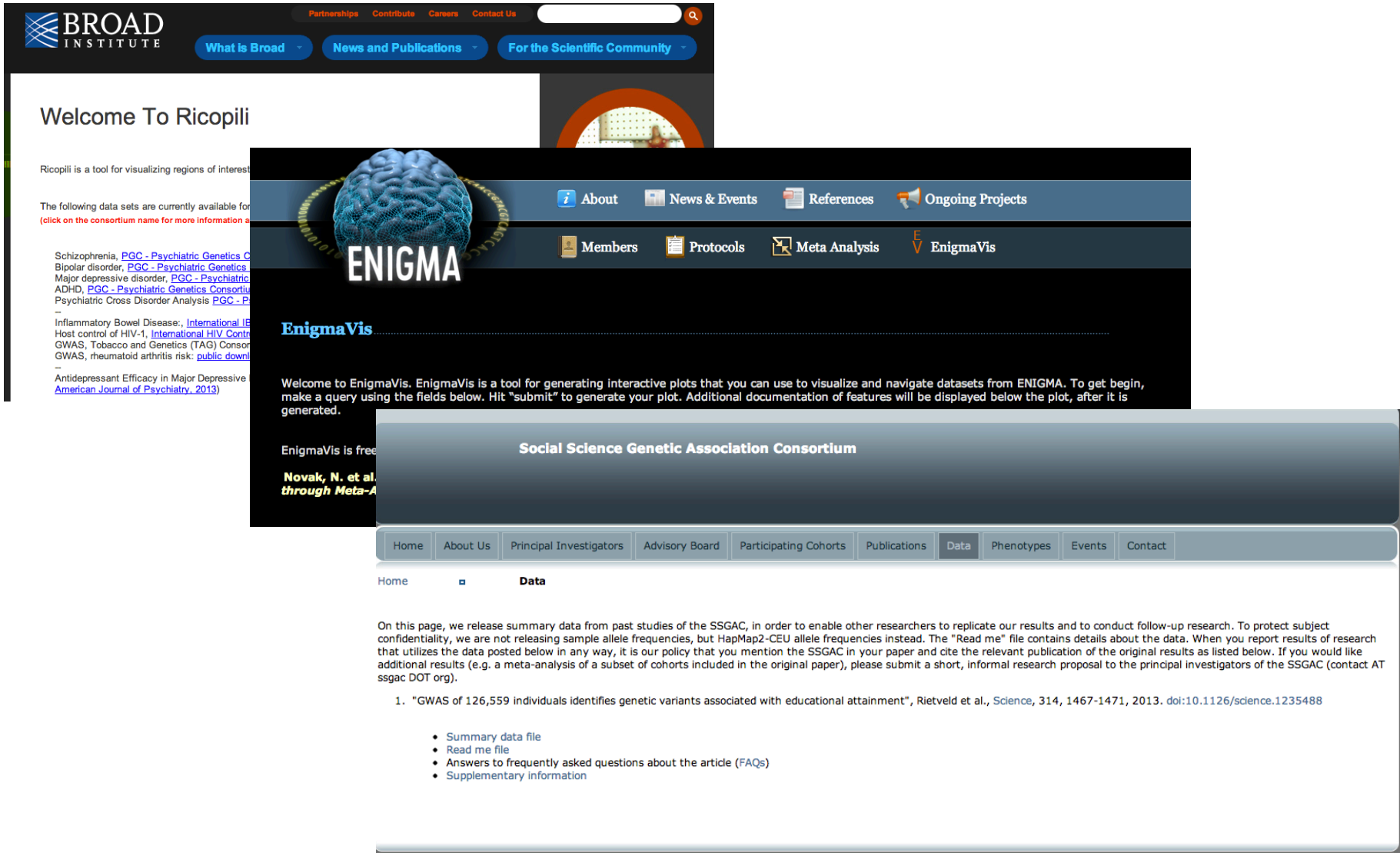
SNP-chip genetic correlation
A ratio of estimates



Accepted for publication in Nature Genetics

Thanks to the PGC cross-disorder group
Naomi Wray, Hong Lee and Ken Kendler

Public availability of full results



Implications for future of GWAS

- Increased meta-analysis will identify additional significant loci
- Predictive ability will continue to improve as sample size increases
- Challenges for GWAS catalog include the management of iterations of meta-analyses and presenting robustness of results
- Increasing uptake of full result hosting

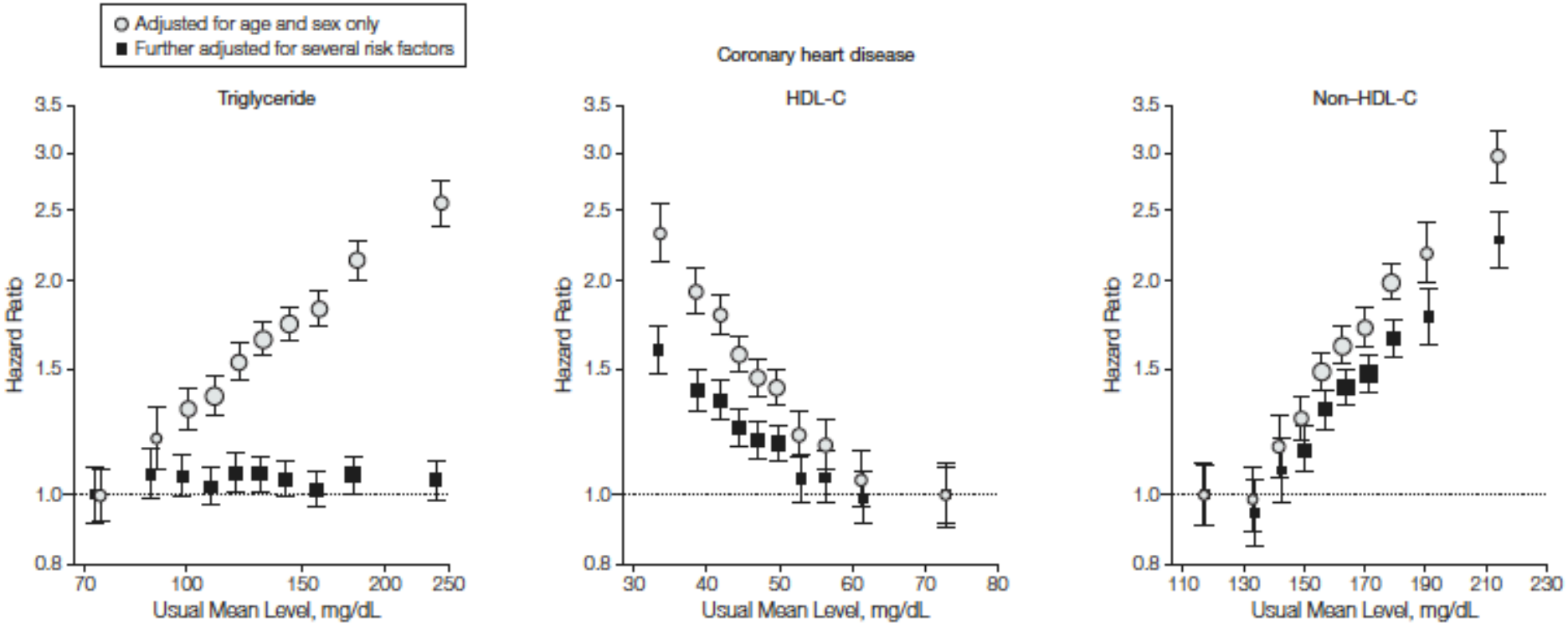
Leveraging genetics to understand the relationship between lipid levels and myocardial infarction

With thanks to Ron Do, Mark Daly, Sek Kathiresan and the Global Lipids Consortium

Introduction

- In observational epidemiological studies:
- ↑ LDL-C is associated with ↑ risk to CAD
- ↓ HDL-C is associated with ↓ risk to CAD
- ↑ TG is associated with ↑ risk to CAD

Hazard ratios of coronary heart disease across triglyceride, HDL-C and non-HDL-C levels



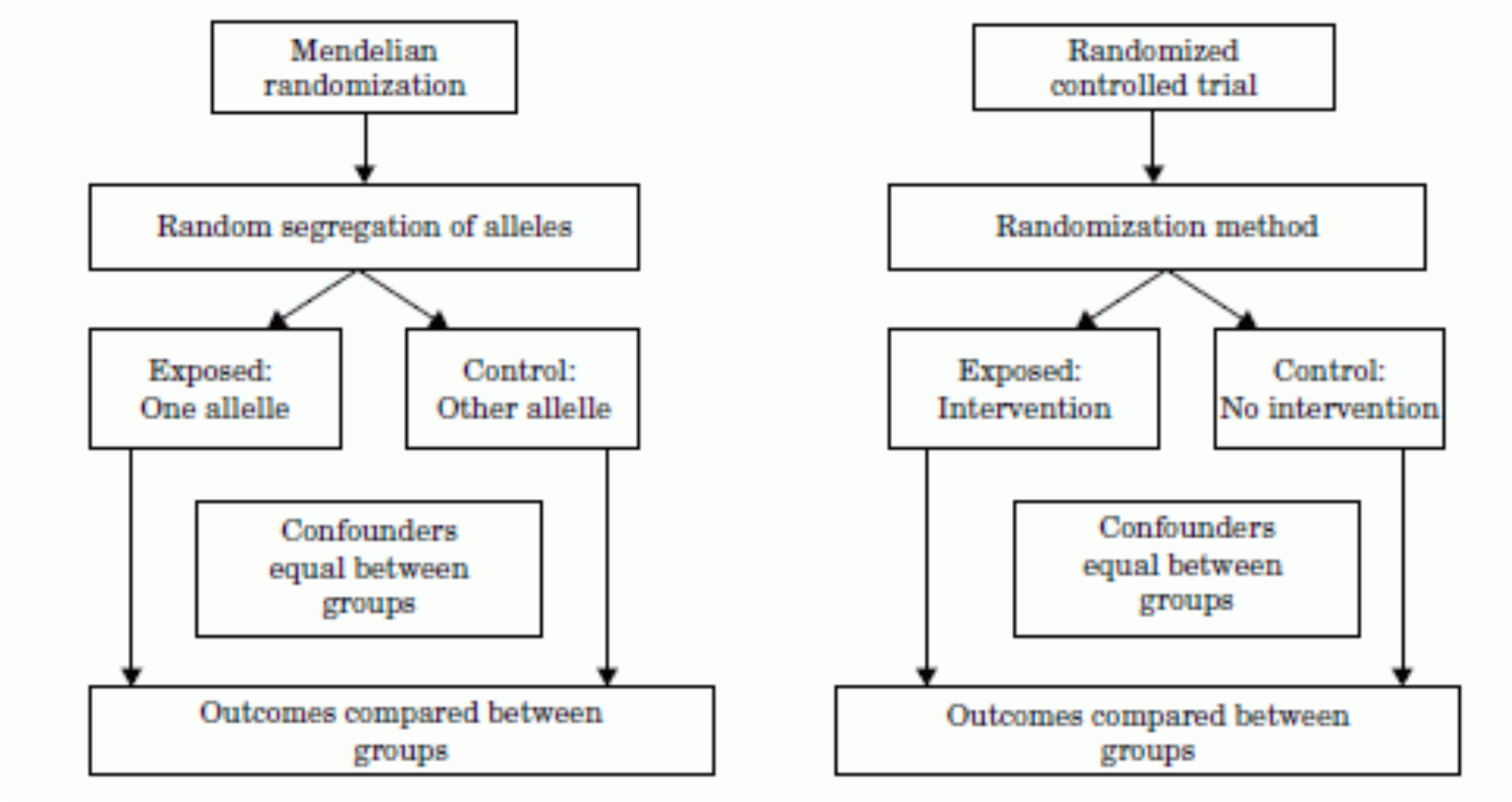
Di Angelantonio et al. JAMA. 2009.

Observational epidemiology

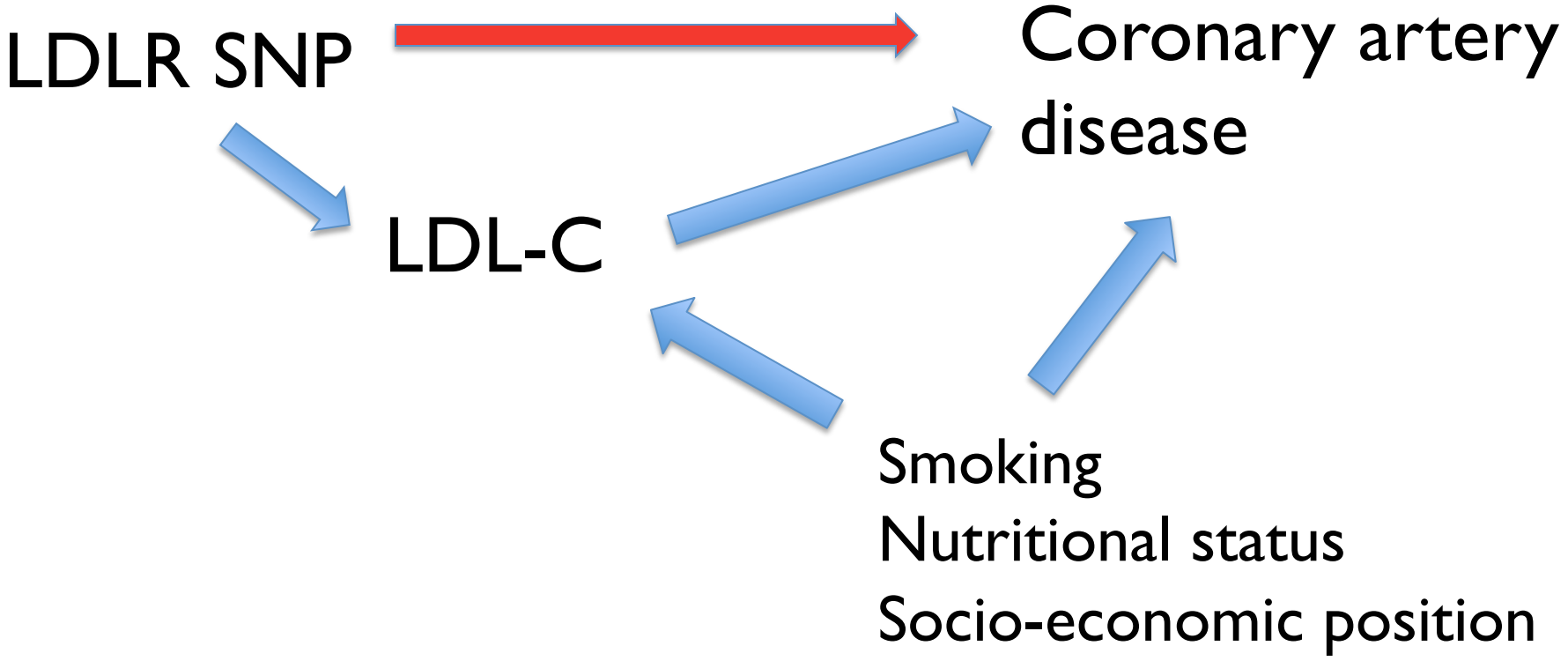
- A limitation of observational epidemiological studies is that it is difficult to establish causal inference.

- Problem is exaggerated by correlation among TG, LDL-C and HDL-C.

Mendelian randomization



LDL-C and CAD


LDLR SNP
Coronary artery disease
LDL-C
Smoking
Nutritional status
Socio-economic position

Plasma HDL cholesterol and risk of MI

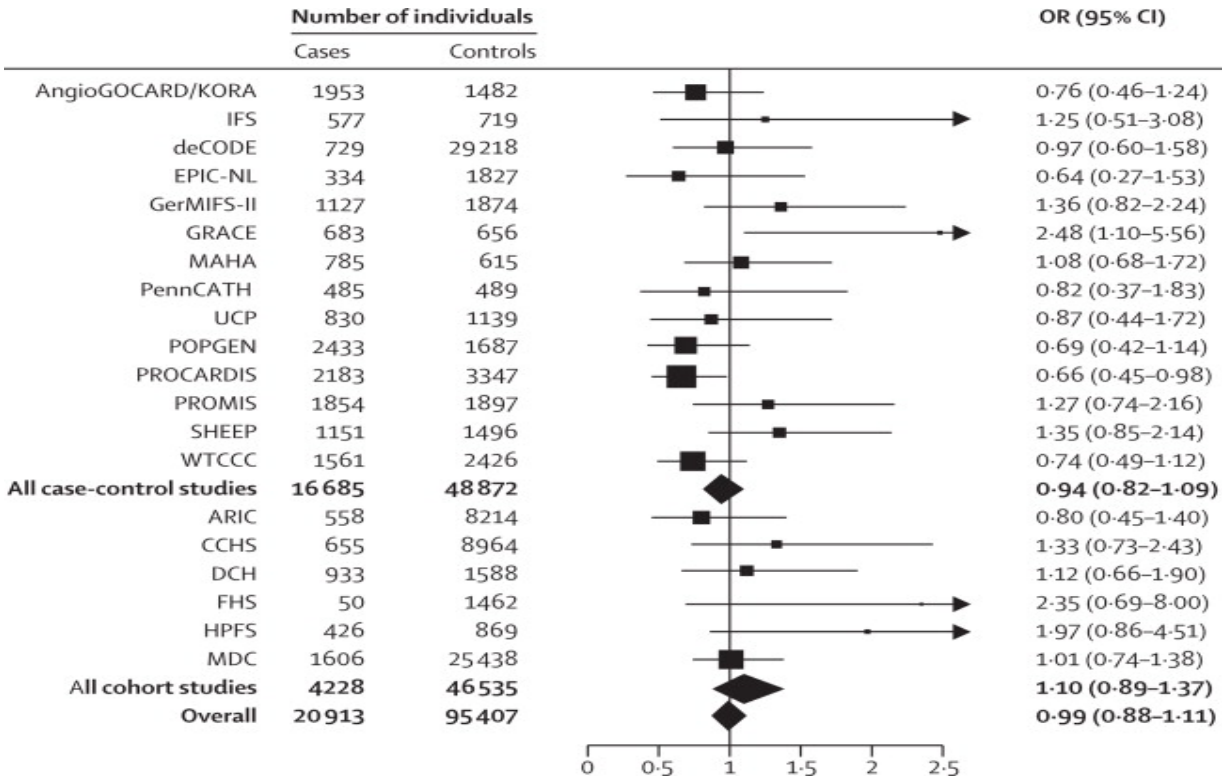


Figure 2 Association of LIPG Asn396Ser with myocardial infarction in 116320 participants from 20 studies In each study, the HDL-cholesterol-raising serine allele was modelled.

Benjamin F Voight* , Gina M Peloso*, Marju Orho-Melander , Ruth Frikke-Schmidt , Maja Barbalic , Majken K Jensen , .., Sekar Kathiresan.

Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study

The Lancet Volume 380, Issue 9841 2012 572 - 580

Plasma HDL cholesterol and risk of MI

	Odds ratio (95% CI) per SD increase in plasma lipid based on observational epidemiology*	Odds ratio (95% CI) per SD increase in plasma lipid conferred by genetic score†
LDL cholesterol	1·54 (1·45–1·63)	2·13 (1·69–2·69), $p=2\times10^{-10}$
HDL cholesterol	0·62 (0·58–0·66)	0·93 (0·68–1·26), p=0·63

Table 4. Estimate of the association of genetically raised LDL cholesterol or HDL cholesterol and risk of myocardial infarction using multiple genetic variants as instruments.

Benjamin F Voight* , Gina M Peloso*, Marju Orho-Melander , Ruth Frikke-Schmidt , Maja Barbalic , Majken K Jensen , ... Sekar Kathiresan

Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study

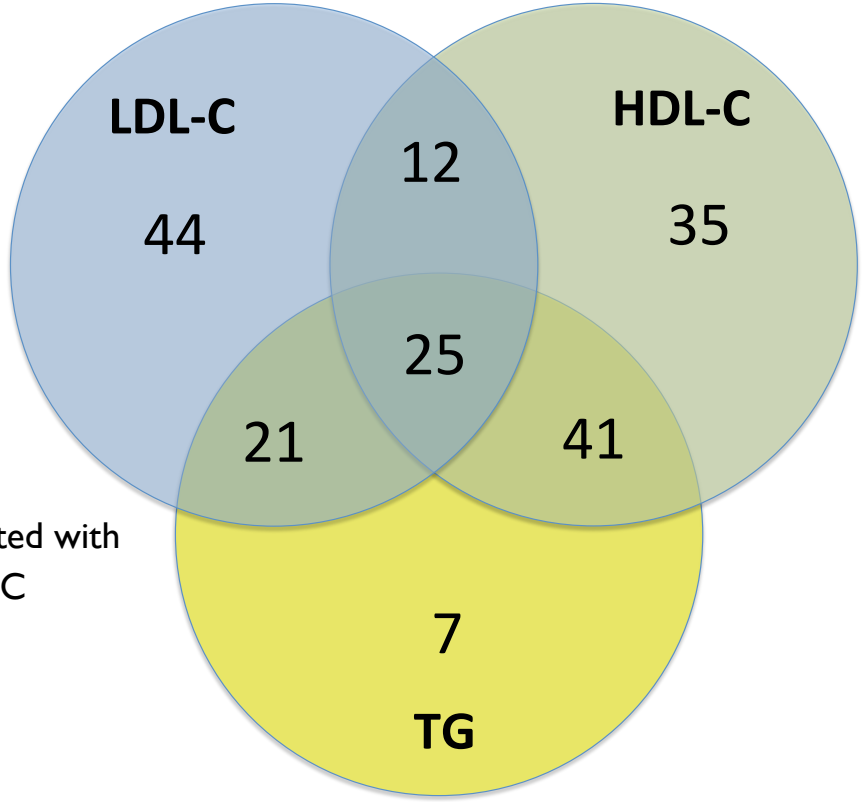
The Lancet Volume 380, Issue 9841 2012 572 - 580

Genetics of lipids and CAD

- LDL-C ⬆ CAD - YES
- HDL-C ⬇ CAD – Not really
- TG ⬆ CAD - ??

Problems with mendelian randomization for TG and CAD

- Nearly all SNPs associated with TG are also associated with LDL-C and HDL-C



Number of SNPs associated with TG, LDL-C and/or HDL-C ($P < 0.001$) are shown.

Methods

- We studied 185 SNPs at 157 loci with association P < 5×10^{-8} for TG, LDL-C, or HDL-C
- GWAS for lipids involves >180,000 individuals
- GWAS for MI involves >95000 individuals
- We examined β_{TG}, $\beta_{LDL\text{-}C}$, $\beta_{HDL\text{-}C}$, with β_{CAD}

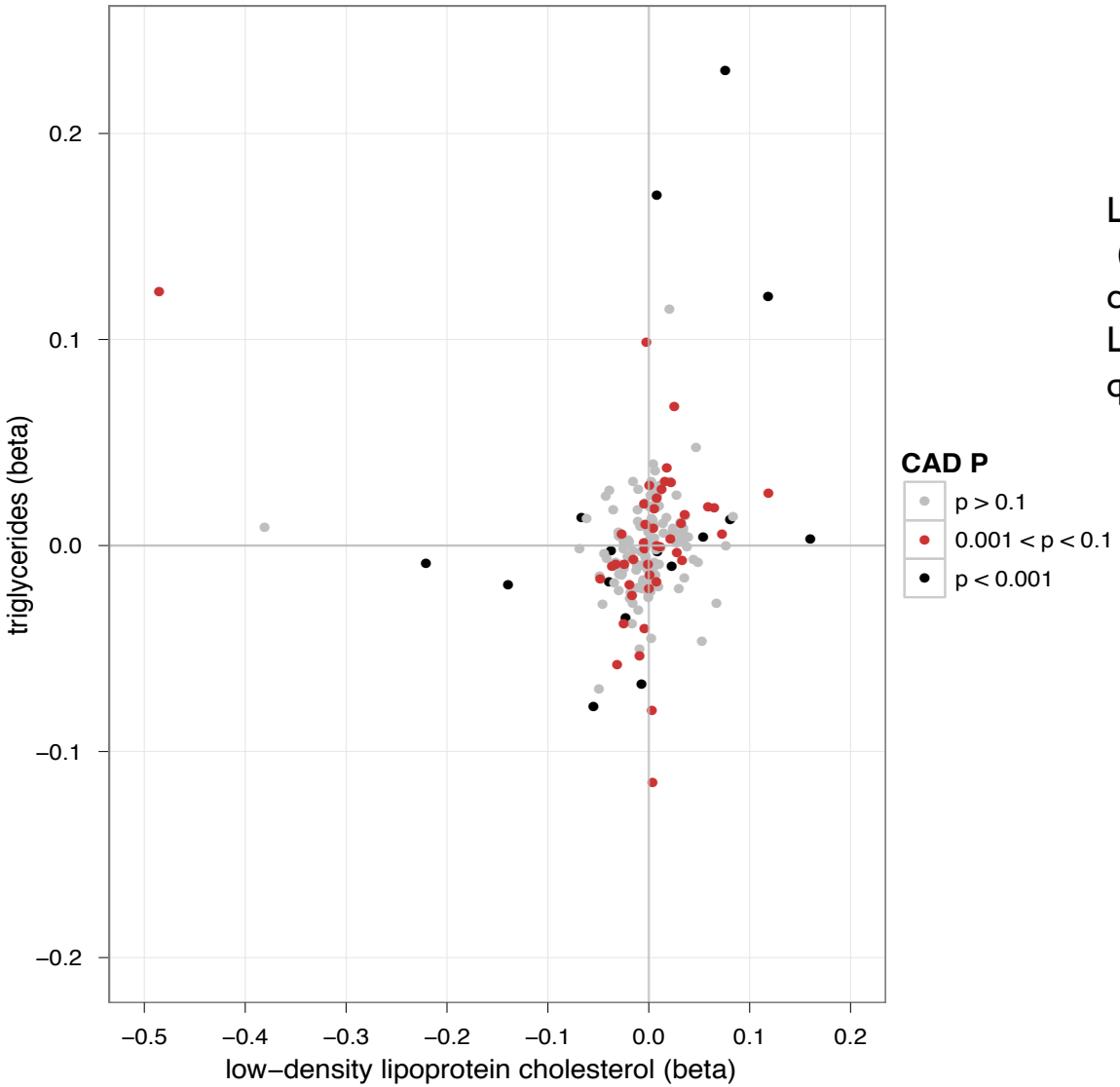
SNPs with consistent direction of effect for TG and LDL-C on CAD risk

			LDL-C		TG		CAD	
Gene	rs ID	A1	$\beta_{LDL\text{-}C}$	*P*	β_{TG}	*P*	β_{CAD}	*P*
TRIB1	rs2954022	A	-0.055	4×10^{-51}	-0.078	2×10^{-124}	-0.056	6×10^{-5}
MEF2B	rs10401969	T	0.12	2×10^{-60}	0.12	3×10^{-76}	0.11	2×10^{-4}

SNPs with opposite direction of effect for TG and LDL-C on CAD risk

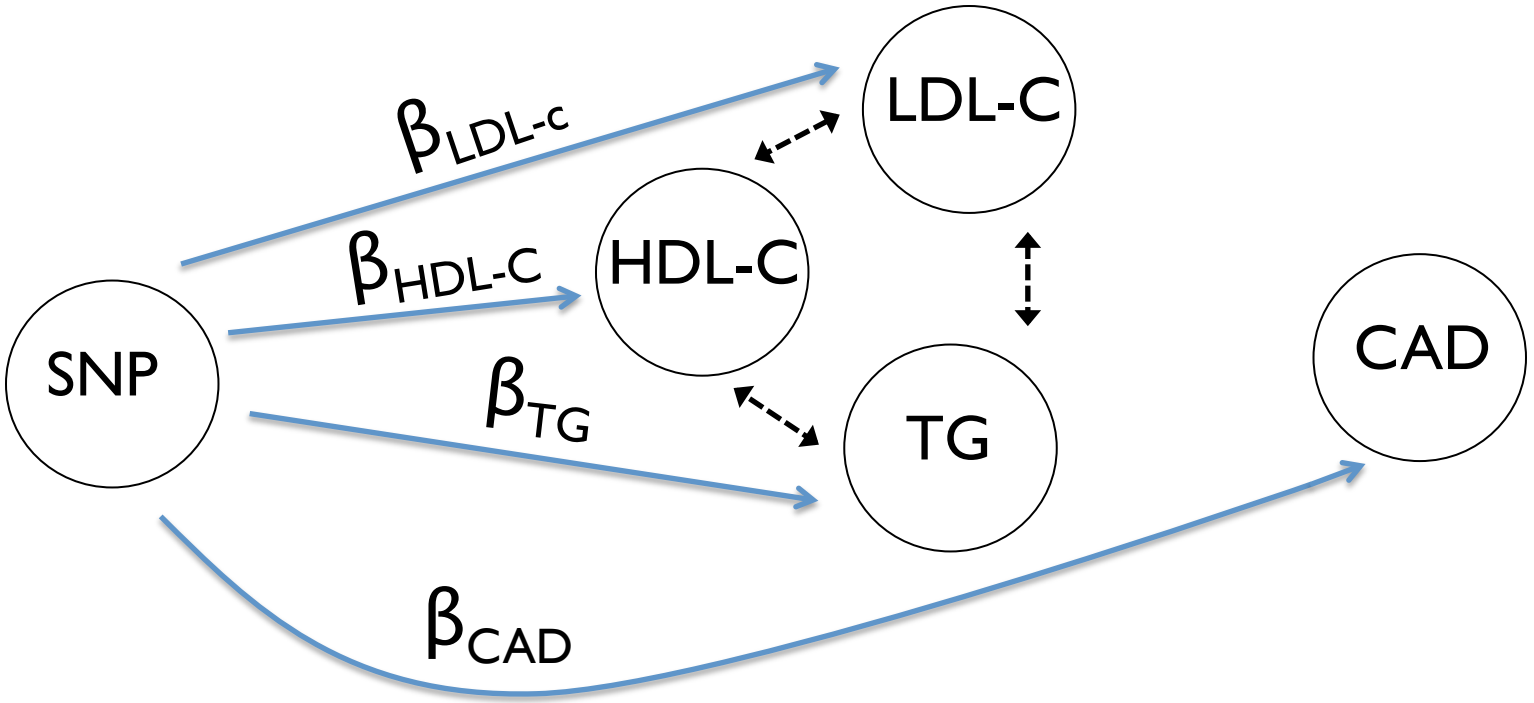
			LDL-C		TG		CAD	
Gene	rs ID	A1	$\beta_{LDL\text{-}C}$	*P*	β_{TG}	*P*	β_{CAD}	*P*
NFE2L3	rs4722551	T	-0.039	7×10^{-16}	0.027	2×10^{-9}	-0.033	0.23
GPAM	rs2255141	A	0.030	7×10^{-14}	-0.021	1×10^{-8}	-0.0076	0.63
SYT7	rs1535	A	0.053	3×10^{-43}	-0.046	1×10^{-40}	0.0019	0.90

Effect of a SNP on TG, LDL-C and risk for CAD



Loci strongly associated with CAD (dots in red or black) tend to have consistent directions for both TG and LDL-C (bottom left and top right quadrants).

Approach to evaluate the effect of a SNP on three lipid fractions and CAD



1) $y = \beta x + b_1$ $\quad x = \beta_{LDL\text{-}C}$, $y = \beta_{CAD}$

2) $e_y = \beta z + b_2$ $\quad z = \beta_{TG}$

Approach to evaluate the effect of a SNP on three lipid fractions and CAD

Model	Outcome	Predictor	Covariate
1	β_{CAD}	$\beta_{LDL\text{-}C}$	-
2	β_{CAD}	$\beta_{LDL\text{-}C}$	$\beta_{HDL\text{-}C}$
3	β_{CAD}	$\beta_{LDL\text{-}C}$	β_{TG}
4	β_{CAD}	$\beta_{LDL\text{-}C}$	$\beta_{HDL\text{-}C}$, β_{TG}
5	β_{CAD}	$\beta_{HDL\text{-}C}$	-
6	β_{CAD}	$\beta_{HDL\text{-}C}$	$\beta_{LDL\text{-}C}$
7	β_{CAD}	$\beta_{HDL\text{-}C}$	β_{TG}
8	β_{CAD}	$\beta_{HDL\text{-}C}$	$\beta_{LDL\text{-}C}$, β_{TG}
9	β_{CAD}	β_{TG}	-
10	β_{CAD}	β_{TG}	$\beta_{LDL\text{-}C}$
11	β_{CAD}	β_{TG}	$\beta_{HDL\text{-}C}$
12	β_{CAD}	β_{TG}	$\beta_{LDL\text{-}C}$, $\beta_{HDL\text{-}C}$

Association of the strength of a SNP's effect on plasma lipids with its strength of effect on CAD risk.

Model	Outcome	Predictor	Covariate	Beta	SE	P
1	β_{CAD}	$\beta_{LDL\text{-}C}$	-	0.41	0.039	4×10^{-20}
2	β_{CAD}	$\beta_{LDL\text{-}C}$	$\beta_{HDL\text{-}C}$	0.38	0.039	9×10^{-19}
3	β_{CAD}	$\beta_{LDL\text{-}C}$	β_{TG}	0.40	0.034	1×10^{-23}
4	β_{CAD}	$\beta_{LDL\text{-}C}$	$\beta_{HDL\text{-}C}$, β_{TG}	0.38	0.034	2×10^{-22}
5	β_{CAD}	$\beta_{HDL\text{-}C}$	-	-0.18	0.052	0.0006
6	β_{CAD}	$\beta_{HDL\text{-}C}$	$\beta_{LDL\text{-}C}$	-0.12	0.041	0.005
7	β_{CAD}	$\beta_{HDL\text{-}C}$	β_{TG}	-0.09	0.048	0.057
8	β_{CAD}	$\beta_{HDL\text{-}C}$	$\beta_{LDL\text{-}C}$, β_{TG}	-0.04	0.037	0.35
9	β_{CAD}	$\beta_{TRIGLYCERIDES}$	-	0.44	0.074	2×10^{-8}
10	β_{CAD}	$\beta_{TRIGLYCERIDES}$	$\beta_{LDL\text{-}C}$	0.42	0.057	5×10^{-12}
11	β_{CAD}	$\beta_{TRIGLYCERIDES}$	$\beta_{HDL\text{-}C}$	0.36	0.074	3×10^{-6}
12	β_{CAD}	$\beta_{TRIGLYCERIDES}$	$\beta_{LDL\text{-}C}$, $\beta_{HDL\text{-}C}$	0.36	0.057	1×10^{-9}

Association of SNPs with opposite effects of LDL-C and TG on CAD effect.

Trait	Beta	SE	*P*
$\beta_{\text{LDL-C}}$	0.23	0.17	0.20

- N=58 SNPs, $(0 < \beta_{LDL\text{-}C} < 0.1$ and $-0.1 < \beta_{TG} < 0)$ or $(-0.1 < \beta_{LDL\text{-}C} < 0$ and $0 < \beta_{TG} < 0.1)$
-
- The association of $\beta_{LDL\text{-}C}$ on β_{CAD} is not significant after restricting to SNPs with opposite directions of $\beta_{LDL\text{-}C}$ and β_{TG}, suggesting that the CAD association is attenuated due to competing $\beta_{LDL\text{-}C}$ and β_{TG}.

Association of SNPs with moderate effect on TG but minimal effect on LDL-C on CAD effect.

Trait	Beta	SE	*P*
β_{TG}	0.51	0.11	0.00003

- N=44 SNPs, ($-0.01 < \beta_{LDL\text{-}C} < 0.01$) and ($\beta_{TG} < -0.01$ or $\beta_{TG} > 0.01$)
-
- The association of $\beta_{LDL\text{-}C}$ on β_{CAD} is not significant after restricting to SNPs with opposite directions of $\beta_{LDL\text{-}C}$ and β_{TG}, suggesting that the CAD association is attenuated due to competing $\beta_{LDL\text{-}C}$ and β_{TG}.

Conclusions

- GWAS is improving our understanding of the genetic architecture of complex traits
- Biological relationships between different phenotypes are clearly being exposed
- As sample sizes continue to increase so too will the number of significant loci for common complex disease