

GWAS as a window to genetic architecture

NHGRI GWAS Catalog Webinar

July 18th 2013

Benjamin Neale



Massachusetts
General Hospital



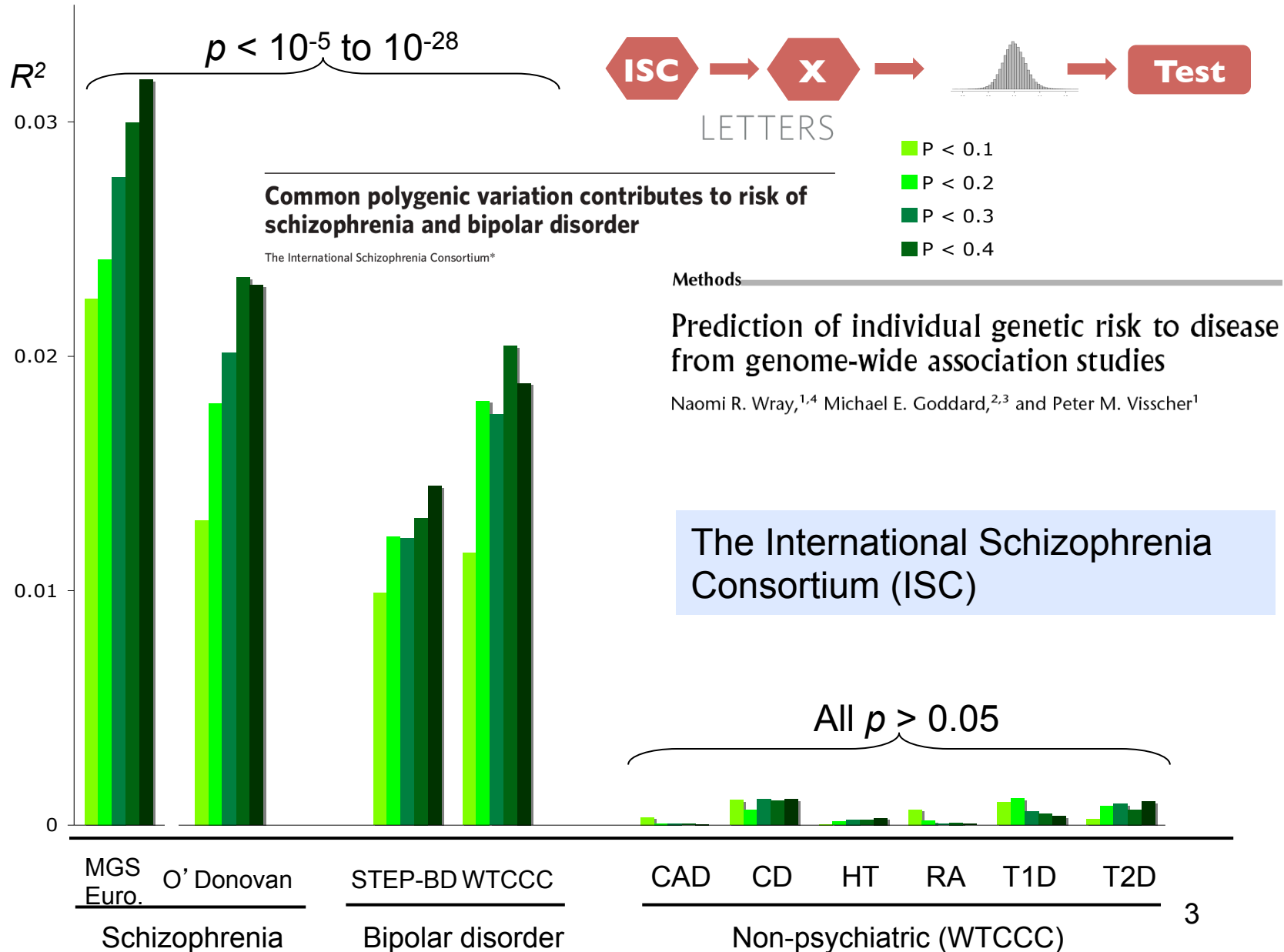
Harvard
Medical
School



Slide of contents

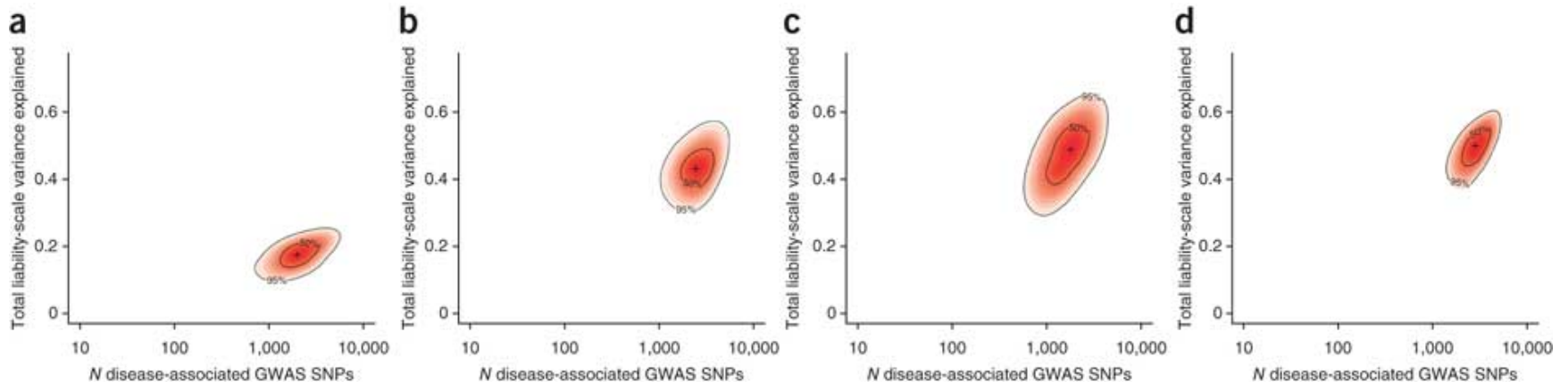
- Complex trait genetic architecture
 - Polygenic prediction
 - Heritability
- Mendelian randomization
 - Example from lipids

Polygenic prediction



Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis

Eli A Stahl^{1-3*}, Daniel Wegmann⁴, Gosia Trynka⁵, Javier Gutierrez-Achury⁵, Ron Do^{2,6}, Benjamin F Voight⁷, Peter Kraft⁸, Robert Chen¹⁻³, Henrik J Kallberg⁹, Fina A S Kurreeman¹⁻³, Diabetes Genetics Replication and Meta-analysis Consortium¹⁰, Myocardial Infarction Genetics Consortium¹⁰, Sekar Kathiresan^{2,6}, Cisca Wijmenga⁵, Peter K Gregersen¹¹, Lars Alfredsson⁹, Katherine A Siminovitch¹², Jane Worthington¹³, Paul I W de Bakker^{2,3,14,15}, Soumya Raychaudhuri^{1-3,16} & Robert M Plenge^{1-3,16}



Rheumatoid Arthritis

Celiac Disease

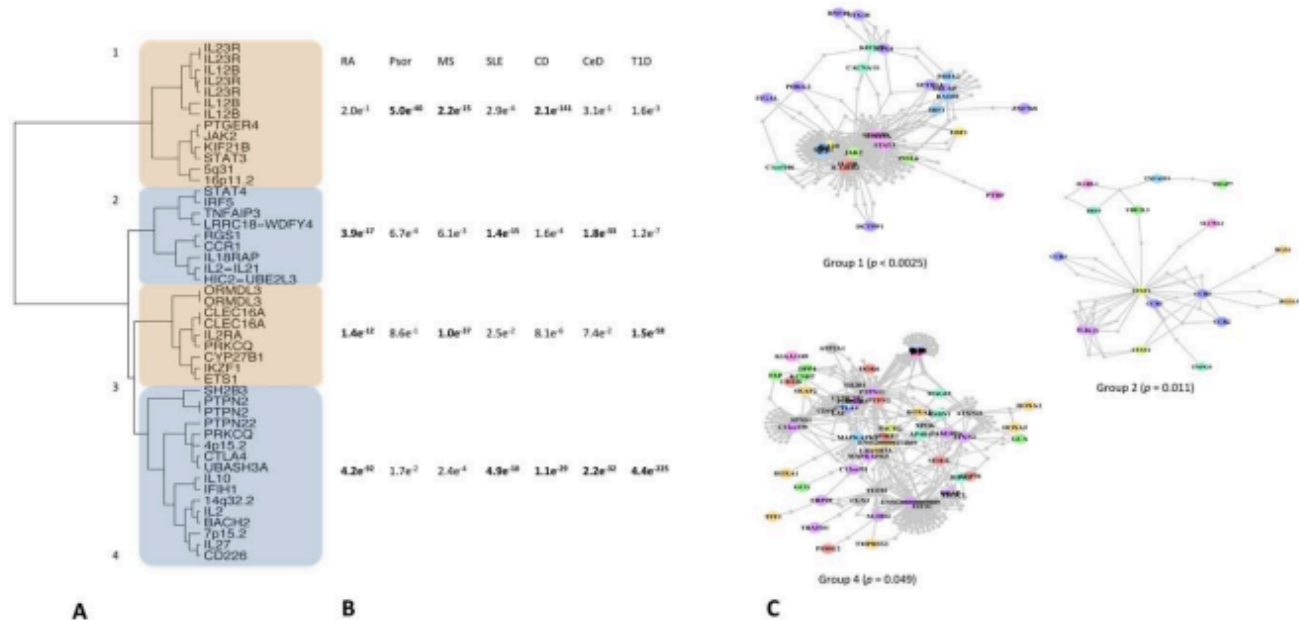
MI/CAD

Type II Diabetes

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

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 Wallace¹¹, Gonalo R. Abecasis¹², Jeffrey C. Barrett¹³, Timothy Behrens¹⁴, Judy Cho^{5,15}, Philip L. De Jager^{3,16}, James T. Elder¹⁷, Robert R. Graham¹⁴, Peter Gregersen¹⁸, Lars Klareskog¹⁹, Katherine A. Siminovitch²⁰, David A. van Heel²¹, Cisca Wijmenga²², Jane Worthington²³, John A. Todd¹¹, David A. Hafler⁴, Stephen S. Rich²⁴, 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. Visscher¹

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

Jian Yang¹, Beben Benyamin¹, Brian P McEvoy¹, Scott Gordon¹, Anjali K Henders¹, Dale R Nyholt¹, Pamela A Madden², Andrew C Heath², Nicholas G Martin¹, Grant W Montgomery¹, Michael E Goddard³ & Peter M Visscher¹

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. Visscher¹

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

Jian Yang¹, Beben Benyamin¹, Brian P McEvoy¹, Scott Gordon¹, Anjali K Henders¹, Dale R Nyholt¹,
Pamela A Madden², Andrew C Heath², Nicholas G Martin¹, Grant W Montgomery¹, Michael E Goddard³ &
Peter M Visscher¹

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. Visscher¹

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

Jian Yang¹, Beben Benyamin¹, Brian P McEvoy¹, Scott Gordon¹, Anjali K Henders¹, Dale R Nyholt¹,
Pamela A Madden², Andrew C Heath², Nicholas G Martin¹, Grant W Montgomery¹, Michael E Goddard³ &
Peter M Visscher¹

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

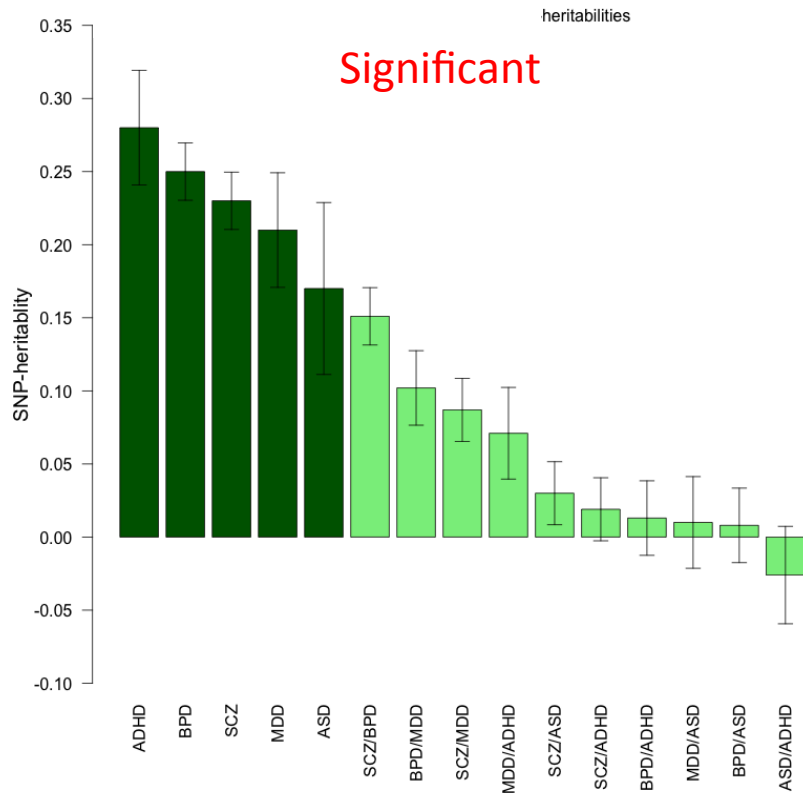
S.H. Lee^{1,*}, J. Yang², M.E. Goddard³, P.M. Visscher^{1,2} and N.R. Wray¹

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

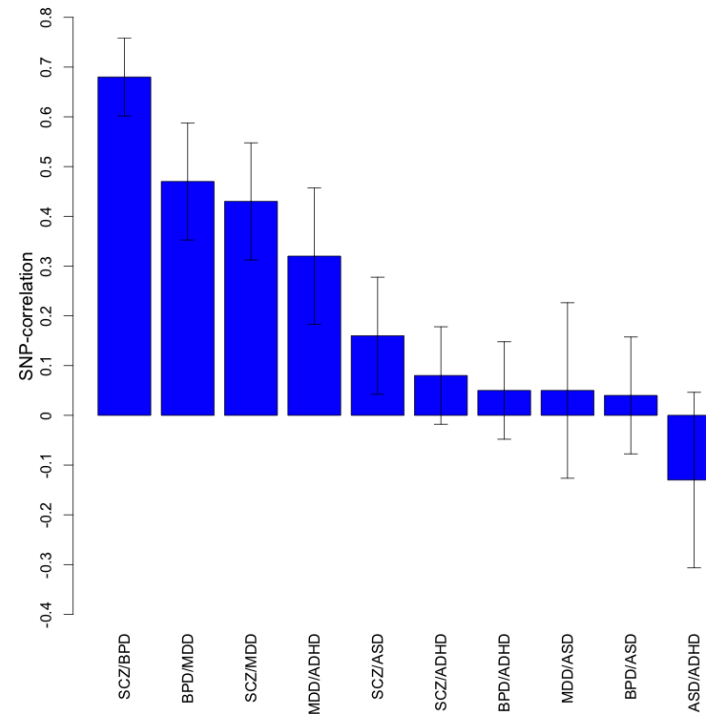
²University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia

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

SNP-chip heritabilities



SNP-chip genetic correlation
A ratio of estimates



Public availability of full results

BROAD INSTITUTE Partnerships Contribute Careers Contact Us

What Is Broad News and Publications For the Scientific Community

Welcome To Ricopili

Ricopili is a tool for visualizing regions of interest...

The following data sets are currently available for
[\(click on the consortium name for more information\)](#)

- Schizophrenia, [PGC - Psychiatric Genetics Consortium](#)
- Bipolar disorder, [PGC - Psychiatric Genetics Consortium](#)
- Major depressive disorder, [PGC - Psychiatric Genetics Consortium](#)
- ADHD, [PGC - Psychiatric Genetics Consortium](#)
- Psychiatric Cross Disorder Analysis [PGC - Psychiatric Genetics Consortium](#)
- Inflammatory Bowel Disease, [International IBD Genetics Consortium](#)
- Host control of HIV-1, [International HIV Control Consortium](#)
- GWAS, Tobacco and Genetics (TAG) Consortium
- GWAS, rheumatoid arthritis risk: [public domain](#)
- Antidepressant Efficacy in Major Depressive Disorder, [American Journal of Psychiatry, 2013](#)

ENIGMA

EnigmaVis

Welcome to EnigmaVis. EnigmaVis is a tool for generating interactive plots that you can use to visualize and navigate datasets from ENIGMA. To get begin, make a query using the fields below. Hit "submit" to generate your plot. Additional documentation of features will be displayed below the plot, after it is generated.

EnigmaVis is free

Novak, N. et al.
through Meta-Analysis

Social Science Genetic Association Consortium

Home About Us Principal Investigators Advisory Board Participating Cohorts Publications **Data** Phenotypes Events Contact

Home **Data**

On this page, we release summary data from past studies of the SSGAC, in order to enable other researchers to replicate our results and to conduct follow-up research. To protect subject confidentiality, we are not releasing sample allele frequencies, but HapMap2-CEU allele frequencies instead. The "Read me" file contains details about the data. When you report results of research that utilizes the data posted below in any way, it is our policy that you mention the SSGAC in your paper and cite the relevant publication of the original results as listed below. If you would like additional results (e.g. a meta-analysis of a subset of cohorts included in the original paper), please submit a short, informal research proposal to the principal investigators of the SSGAC (contact AT ssgac DOT org).

- "GWAS of 126,559 individuals identifies genetic variants associated with educational attainment", Rietveld et al., *Science*, 314, 1467-1471, 2013. doi:10.1126/science.1235488

- Summary data file
- Read me file
- Answers to frequently asked questions about the article (FAQs)
- Supplementary information







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

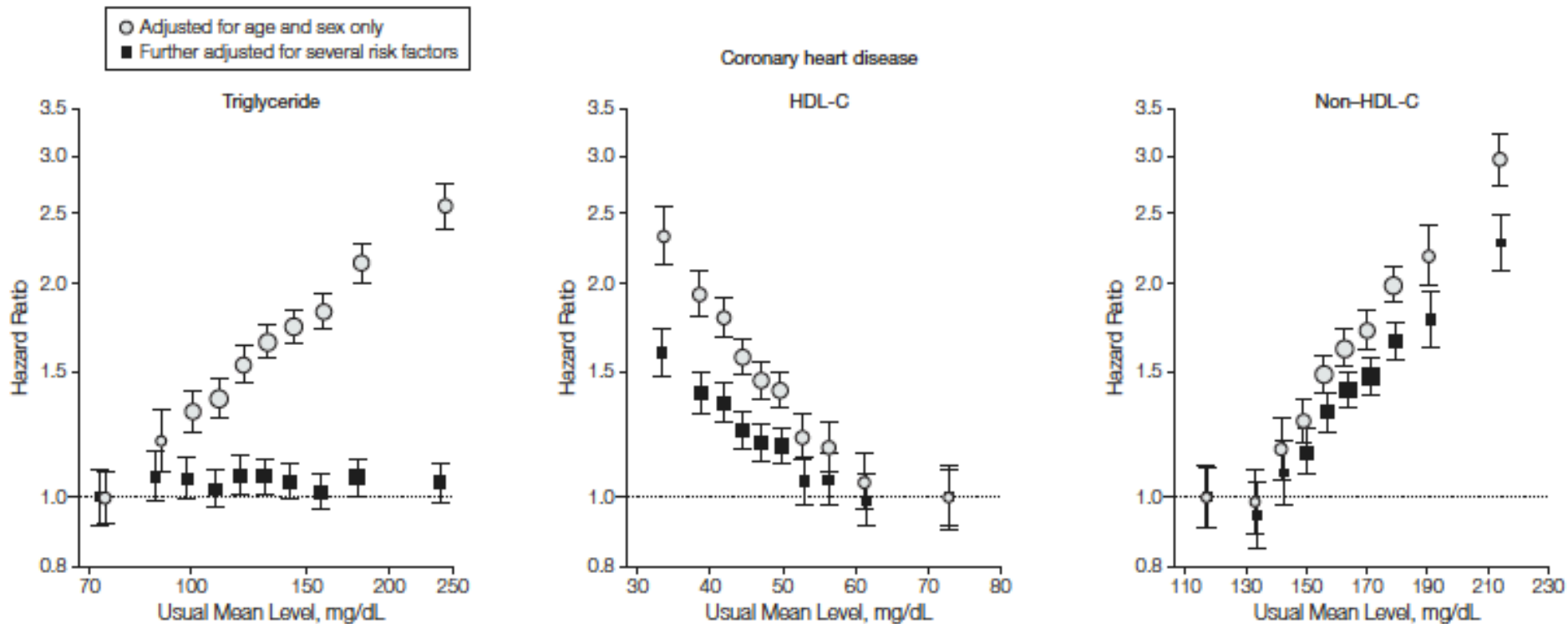
Leveraging genetics to understand the relationship between lipid levels and myocardial infarction

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

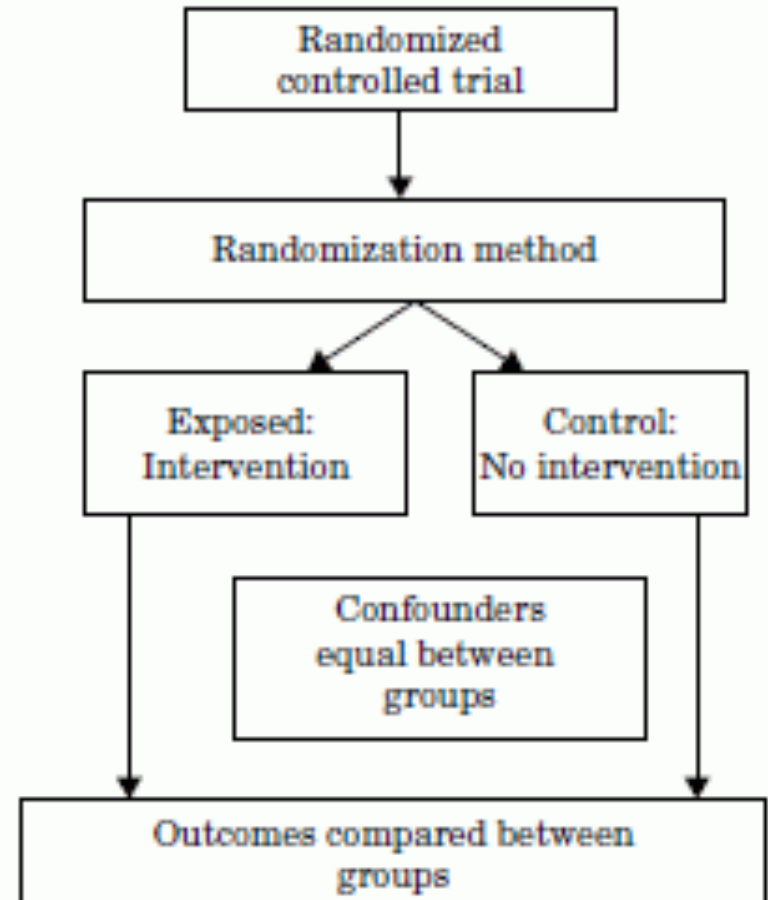
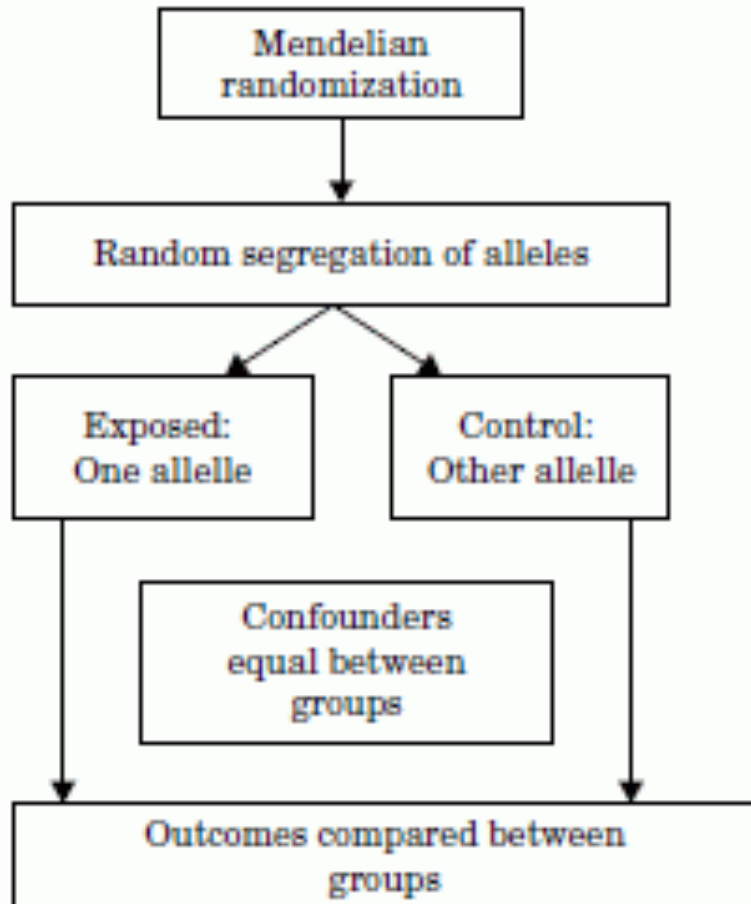
Hazard ratios of coronary heart disease across triglyceride, HDL-C and non-HDL-C levels



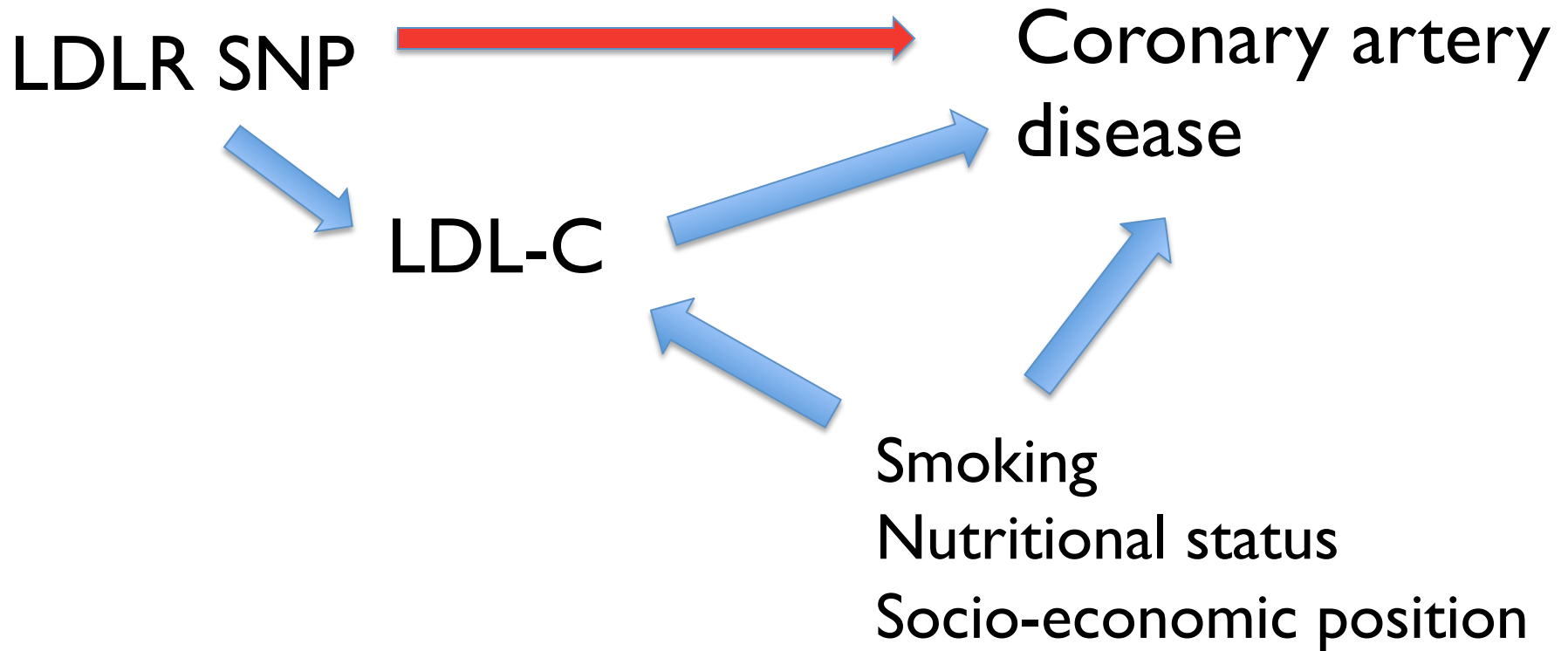
Observational epidemiology

- 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



Plasma HDL cholesterol and risk of MI

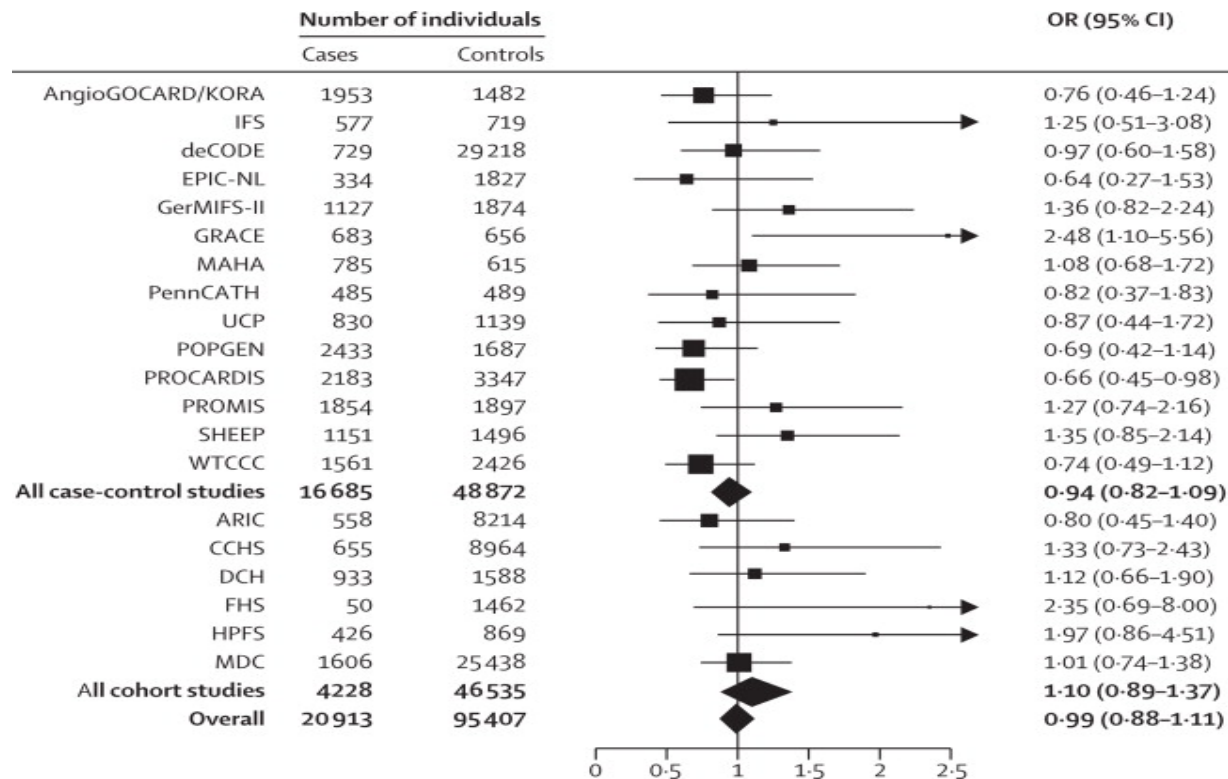


Figure 2 Association of LIPG Asn396Ser with myocardial infarction in 116320 participants from 20 studies. In each study, the HDL-cholesterol-raising serine allele was modelled.

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

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=2×10 ⁻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

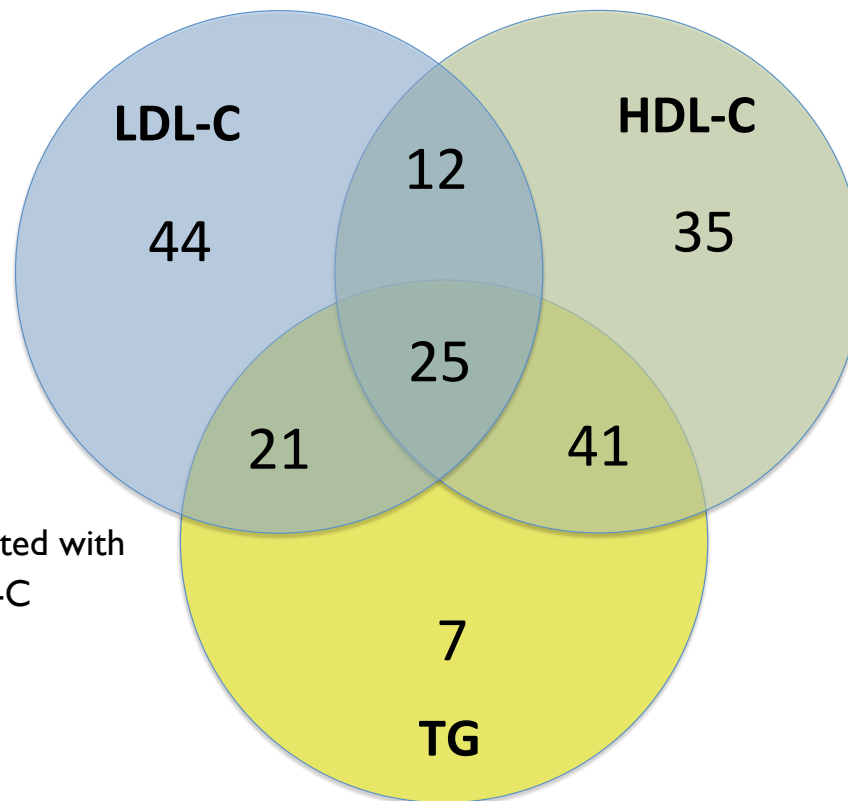
Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study

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 $P < 5 \times 10^{-8}$ for TG, LDL-C, or HDL-C
- GWAS for lipids involves >180,000 individuals
- GWAS for MI involves >95000 individuals
- We examined β_{TG} , $\beta_{\text{LDL-C}}$, $\beta_{\text{HDL-C}}$, with β_{CAD}

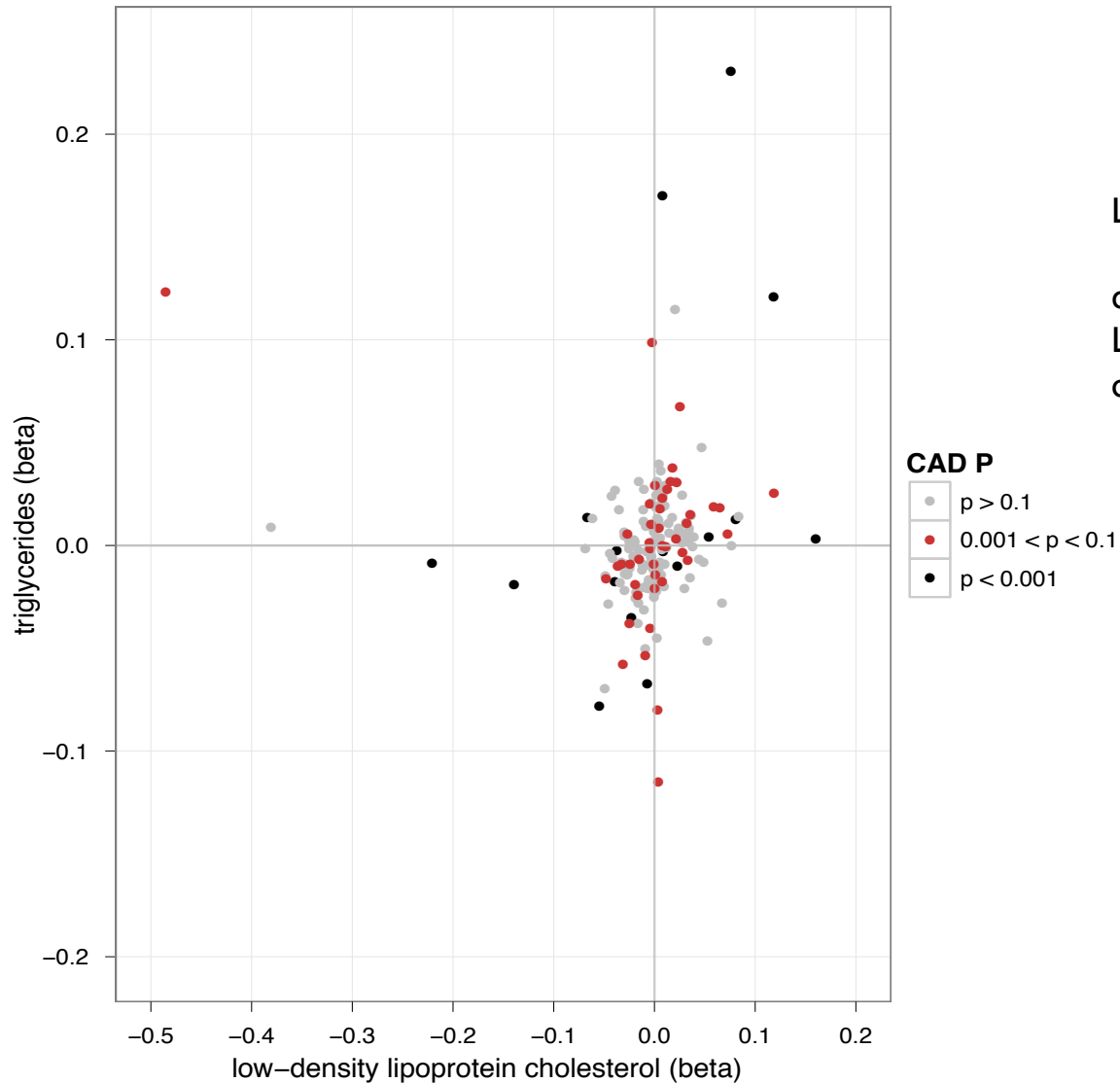
SNPs with consistent direction of effect for TG and LDL-C on CAD risk

			LDL-C		TG		CAD	
Gene	rs ID	AI	$\beta_{\text{LDL-C}}$	P	β_{TG}	P	β_{CAD}	P
<i>TRIB1</i>	rs2954022	A	-0.055	4×10^{-51}	-0.078	2×10^{-124}	-0.056	6×10^{-5}
<i>MEF2B</i>	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

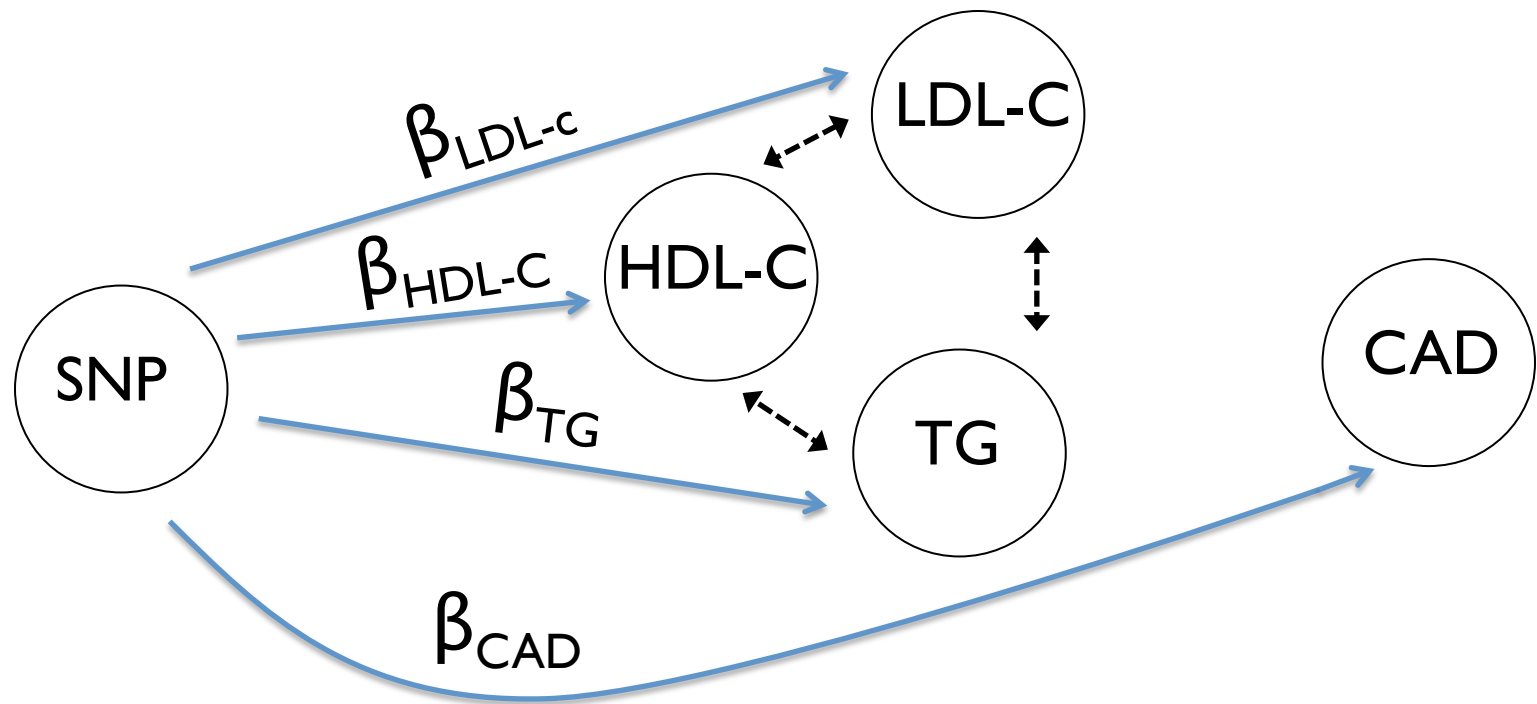
Gene	rs ID	AI	LDL-C		TG		CAD	
			$\beta_{\text{LDL-C}}$	P	β_{TG}	P	β_{CAD}	P
<i>NFE2L3</i>	rs4722551	T	-0.039	7×10^{-16}	0.027	2×10^{-9}	-0.033	0.23
<i>GPAM</i>	rs2255141	A	0.030	7×10^{-14}	-0.021	1×10^{-8}	-0.0076	0.63
<i>SYT7</i>	rs1535	A	0.053	3×10^{-43}	-0.046	1×10^{-40}	0.0019	0.90

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



$$1) y = \beta x + b_1 \quad \begin{array}{l} x = \beta_{LDL-C} \\ y = \beta_{CAD} \end{array}$$

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

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

Model	Outcome	Predictor	Covariate
1	β CAD	β LDL-C	-
2	β CAD	β LDL-C	β HDL-C
3	β CAD	β LDL-C	β TG
4	β CAD	β LDL-C	β HDL-C, β TG
5	β CAD	β HDL-C	-
6	β CAD	β HDL-C	β LDL-C
7	β CAD	β HDL-C	β TG
8	β CAD	β HDL-C	β LDL-C, β TG
9	β CAD	β TG	-
10	β CAD	β TG	β LDL-C
11	β CAD	β TG	β HDL-C
12	β CAD	β TG	β LDL-C, β HDL-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}	β_{LDL-C}	-	0.41	0.039	4×10^{-20}
2	β_{CAD}	β_{LDL-C}	β_{HDL-C}	0.38	0.039	9×10^{-19}
3	β_{CAD}	β_{LDL-C}	β_{TG}	0.40	0.034	1×10^{-23}
4	β_{CAD}	β_{LDL-C}	$\beta_{HDL-C}, \beta_{TG}$	0.38	0.034	2×10^{-22}
5	β_{CAD}	β_{HDL-C}	-	-0.18	0.052	0.0006
6	β_{CAD}	β_{HDL-C}	β_{LDL-C}	-0.12	0.041	0.005
7	β_{CAD}	β_{HDL-C}	β_{TG}	-0.09	0.048	0.057
8	β_{CAD}	β_{HDL-C}	$\beta_{LDL-C}, \beta_{TG}$	-0.04	0.037	0.35
9	β_{CAD}	$\beta_{TRIGLYCERIDES}$	-	0.44	0.074	2×10^{-8}
10	β_{CAD}	$\beta_{TRIGLYCERIDES}$	β_{LDL-C}	0.42	0.057	5×10^{-12}
11	β_{CAD}	$\beta_{TRIGLYCERIDES}$	β_{HDL-C}	0.36	0.074	3×10^{-6}
12	β_{CAD}	$\beta_{TRIGLYCERIDES}$	$\beta_{LDL-C}, \beta_{HDL-C}$	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, ($0 < \beta_{\text{LDL-C}} < 0.1$ and $-0.1 < \beta_{\text{TG}} < 0$) or ($-0.1 < \beta_{\text{LDL-C}} < 0$ and $0 < \beta_{\text{TG}} < 0.1$)
-
- The association of $\beta_{\text{LDL-C}}$ on β_{CAD} is not significant after restricting to SNPs with opposite directions of $\beta_{\text{LDL-C}}$ and β_{TG} , suggesting that the CAD association is attenuated due to competing $\beta_{\text{LDL-C}}$ 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, ($-0.01 < \beta_{LDL-C} < 0.01$) and ($\beta_{TG} < -0.01$ or $\beta_{TG} > 0.01$)
-
- The association of β_{LDL-C} on β_{CAD} is not significant after restricting to SNPs with opposite directions of β_{LDL-C} and β_{TG} , suggesting that the CAD association is attenuated due to competing β_{LDL-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