UNIVERSITY OF CAMBRIDGE

Integrating genomics and multi-omics at population scale for disease insights

Professor Adam Butterworth
BHF Cardiovascular Epidemiology Unit
Department of Public Health & Primary Care

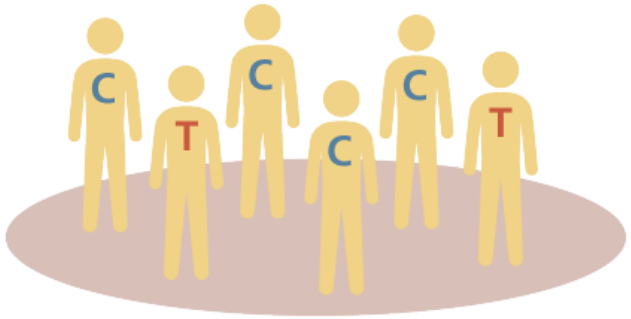
CMIH Academic Engagement Event

26th July 2022

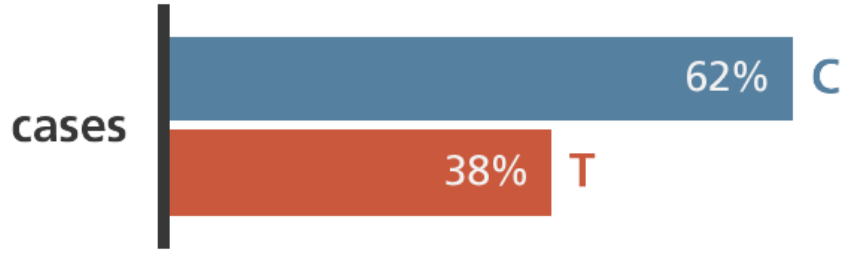
asb38@medschl.cam.ac.uk

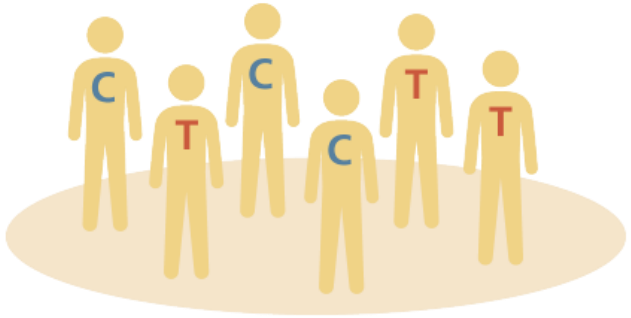
@aidanbutty

Associating genetic variants with risk of disease



cases
people with heart disease

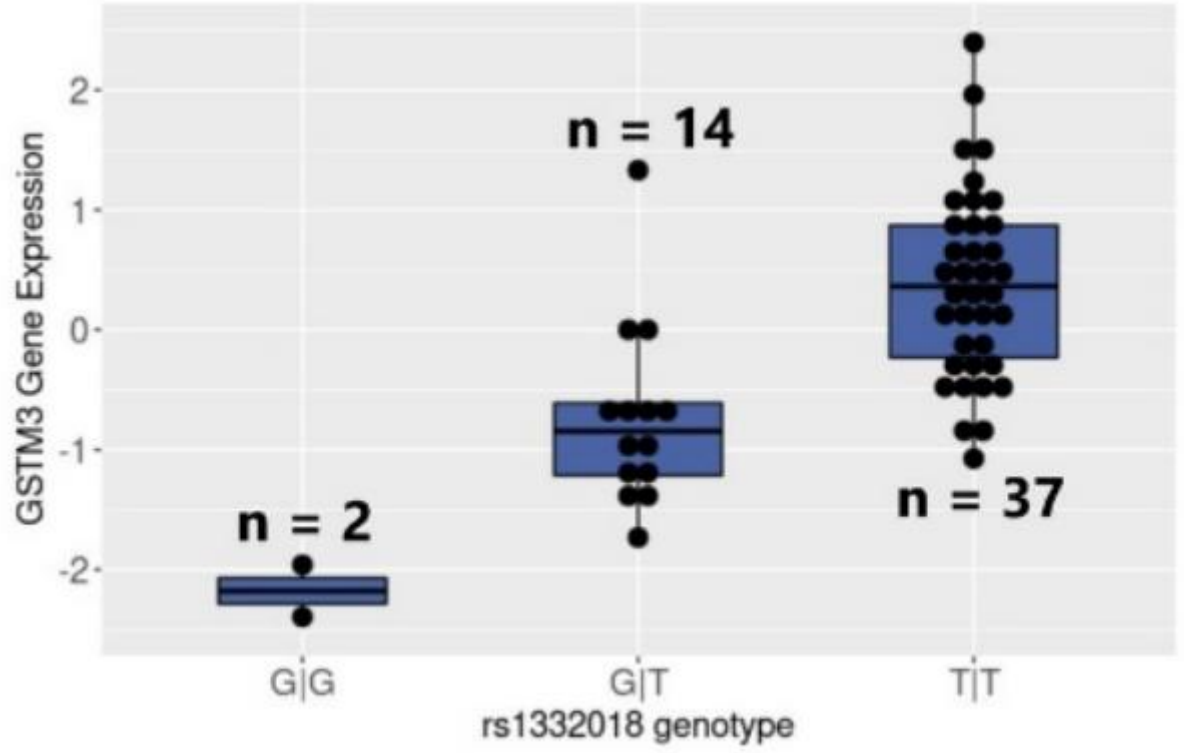




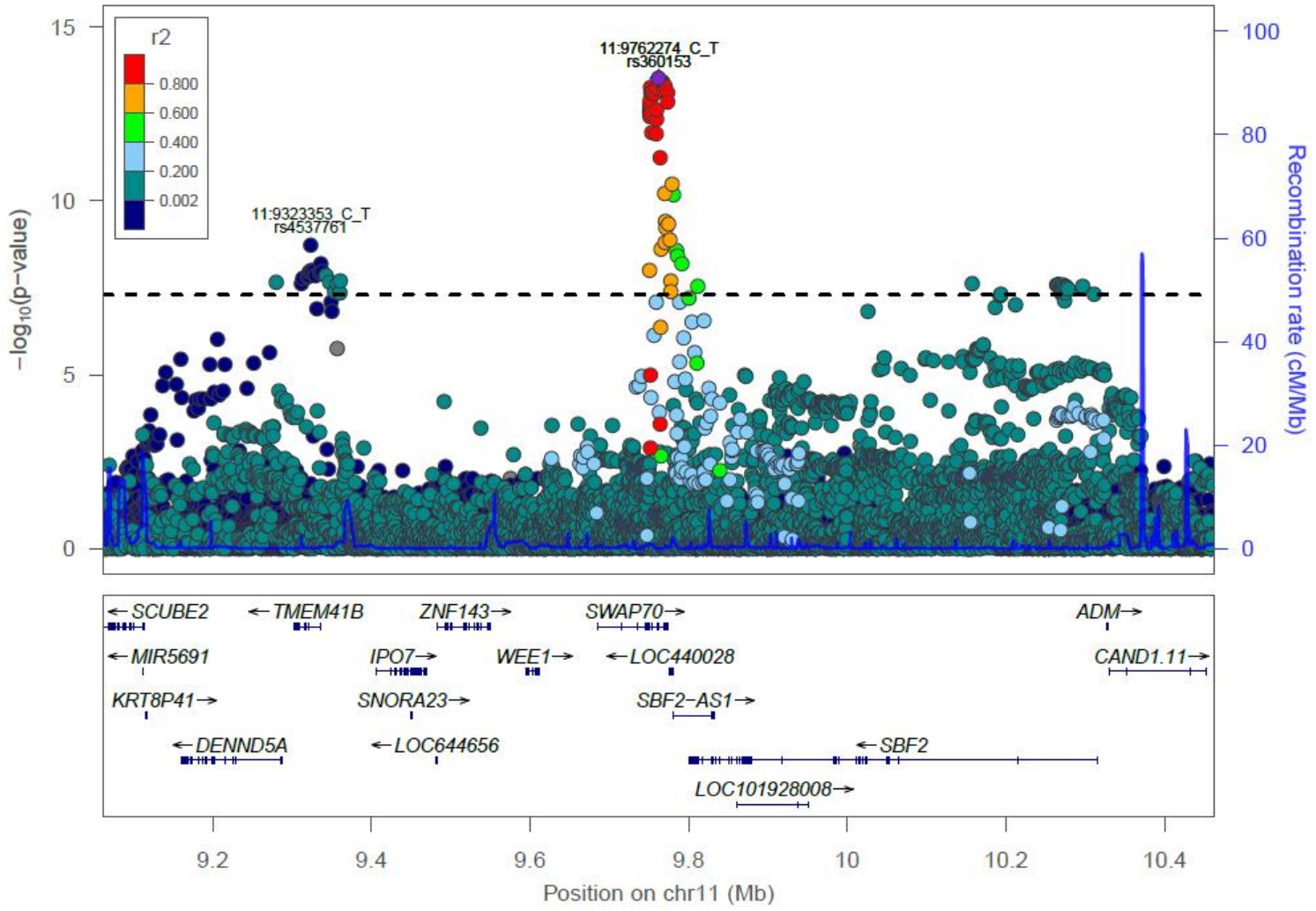
controls
people without heart disease

https://www.yourgenome.org/stories/genome-wide-association-studies/

Associating genetic variants with quantitative molecular traits

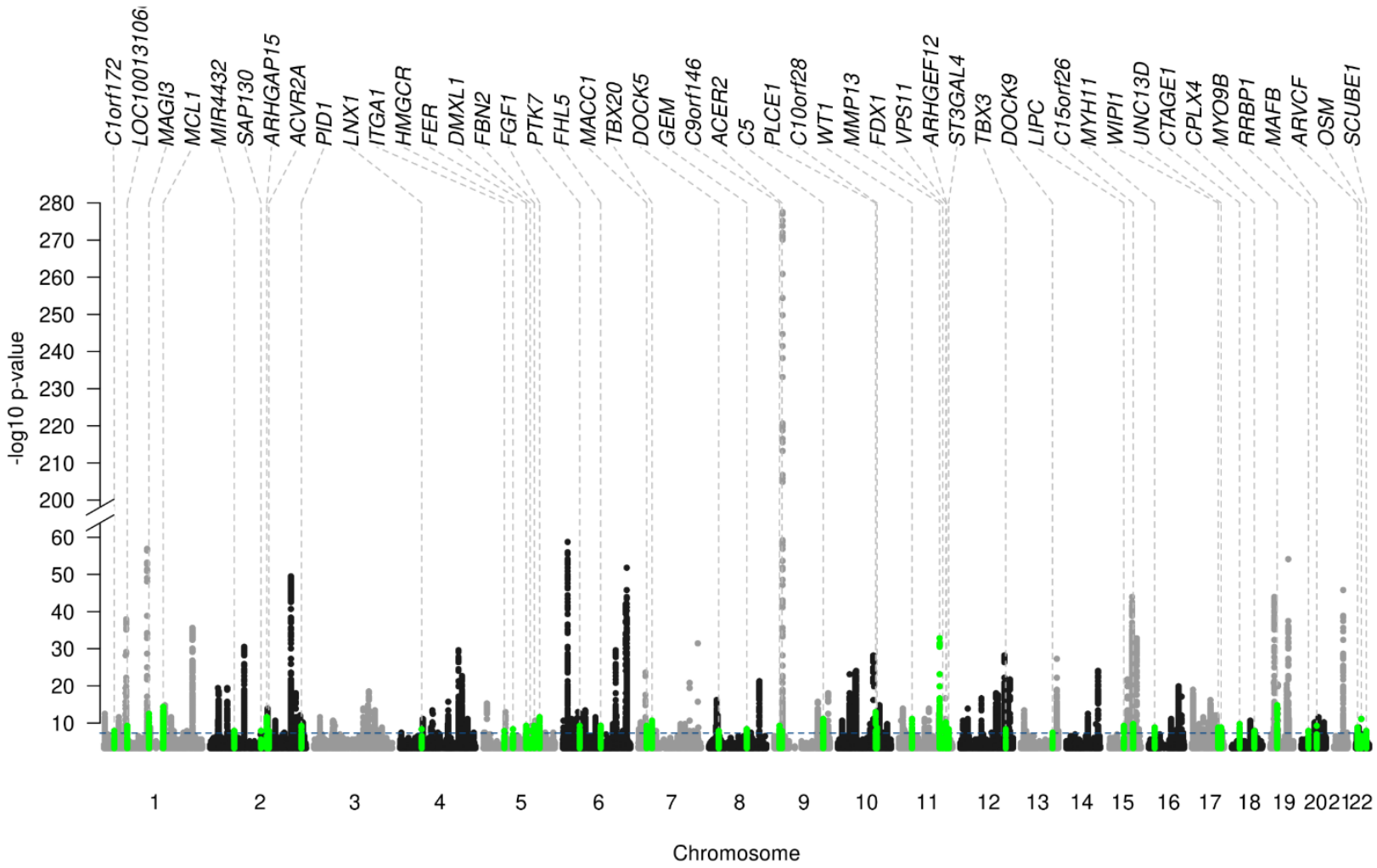


Linking genomic regions with coronary artery disease



Aragam *et al, Nature Genetics*, *in press*

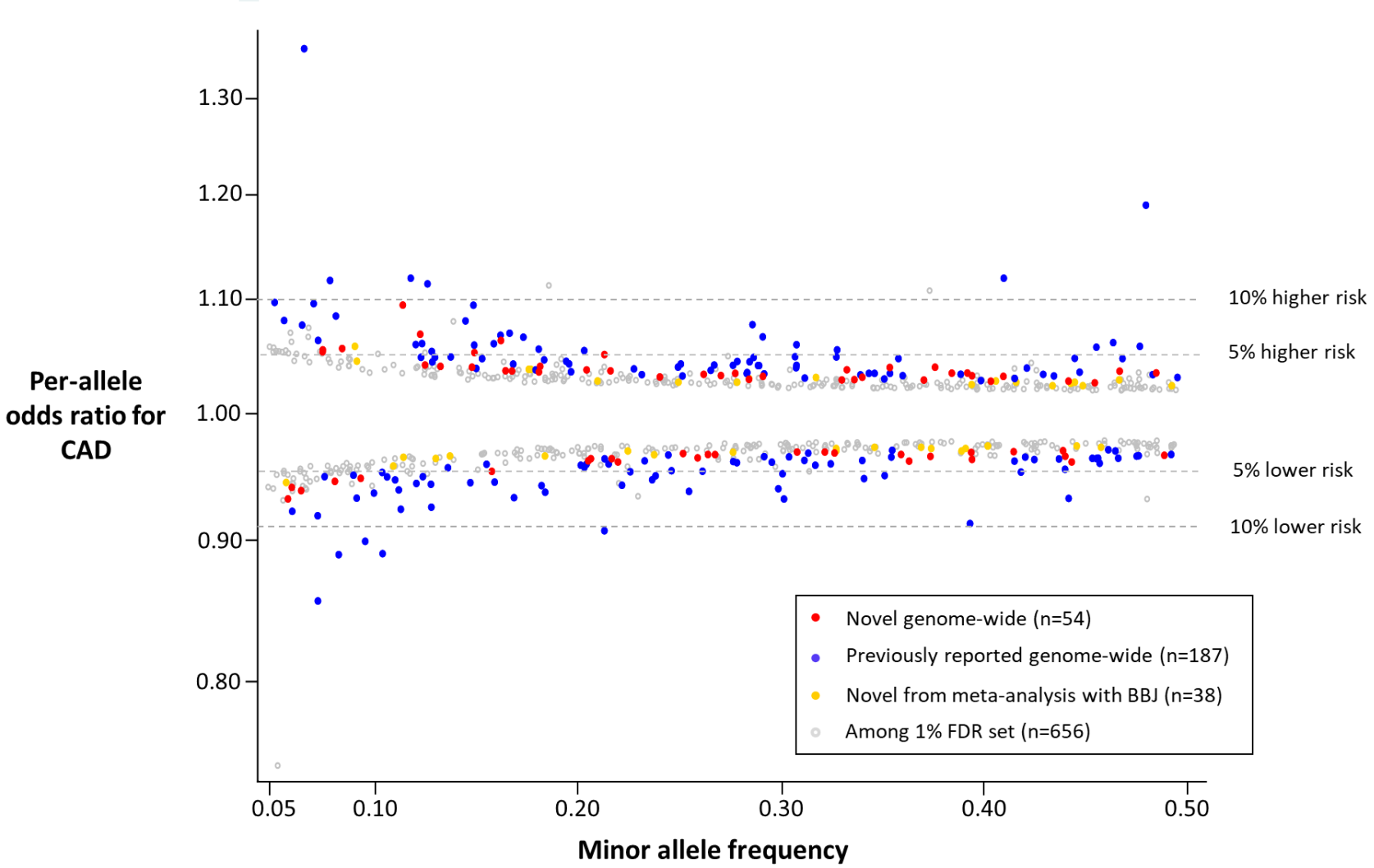
Genome-wide association study (GWAS) of coronary artery disease involving >1M participants



- >11 million variants tested
- 279 signals with p<5x10e-08
- 68 novel (in green)
- ~900 signals at 1% FDR

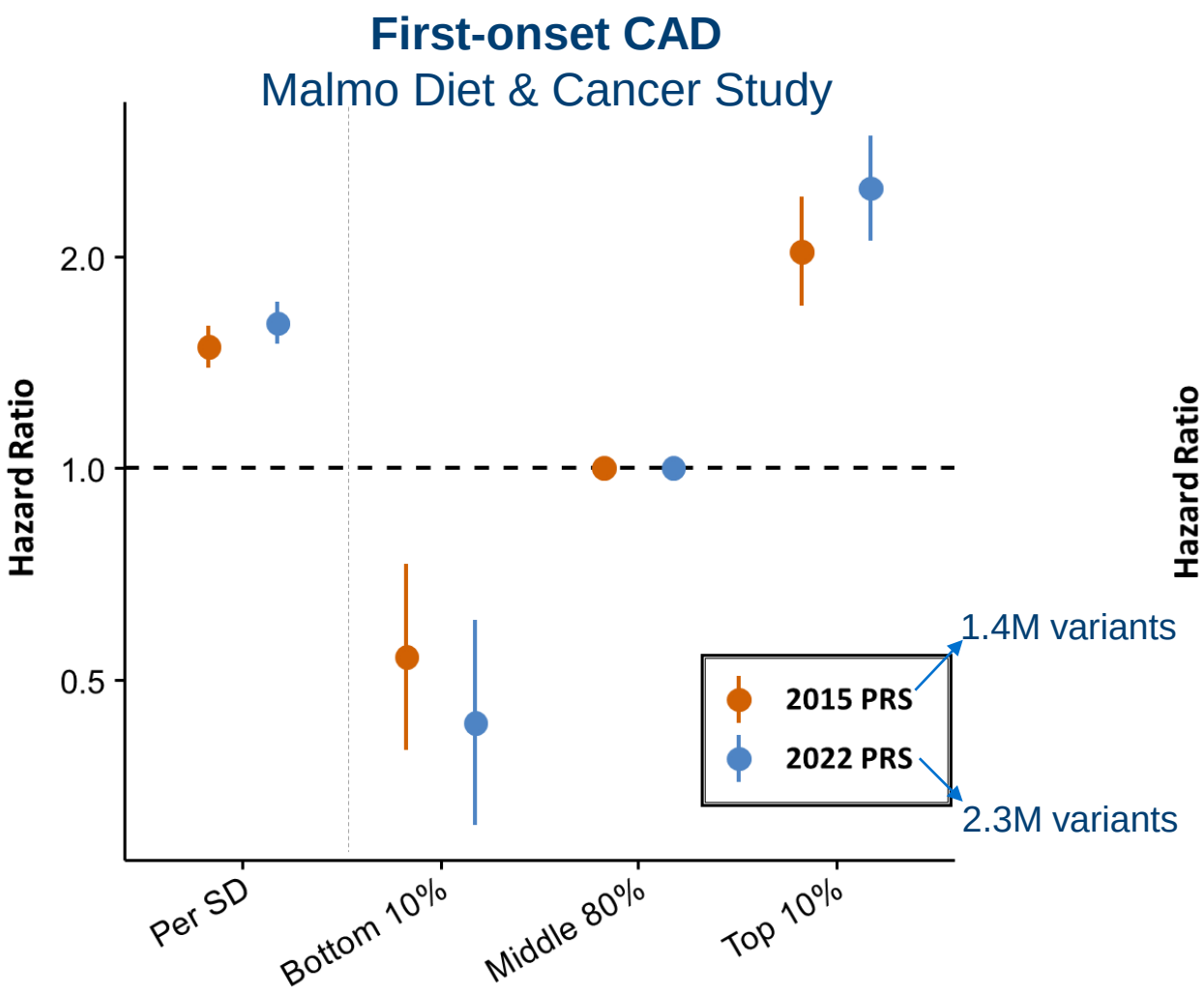
Aragam *et al, Nature Genetics*, *in press*

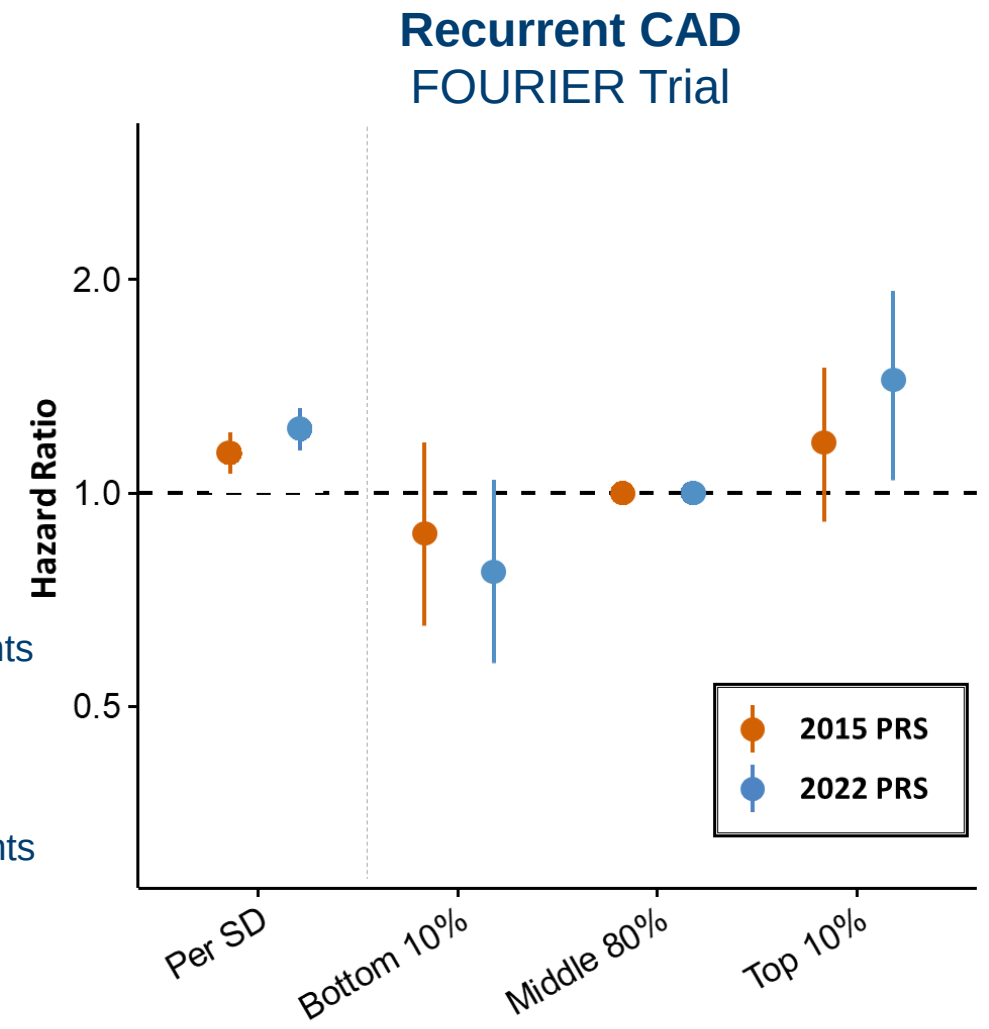
Effect sizes for most variants are very small!



Aragam *et al, Nature Genetics, in press*

New polygenic risk score (PRS) improves prediction of first-onset and recurrent CAD risk



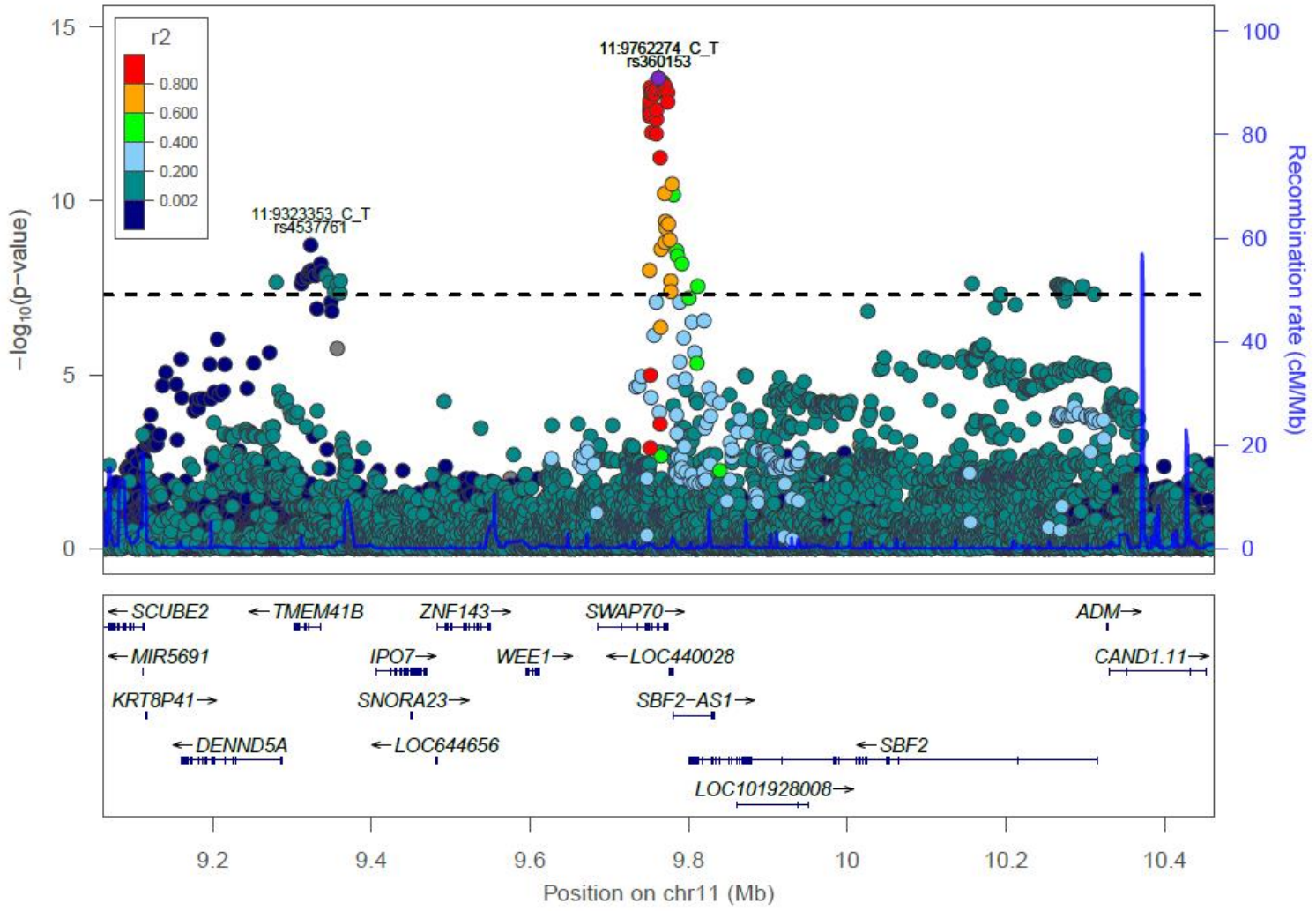


Aragam *et al, Nature Genetics, in press*

Challenge: enhanced methods to improve PRS?

- Combine multiple existing PRS methods?
- Allow for interaction effects?
- Combine PRS for risk factors with PRS for disease?
- **Translate PRS across ethnicities**

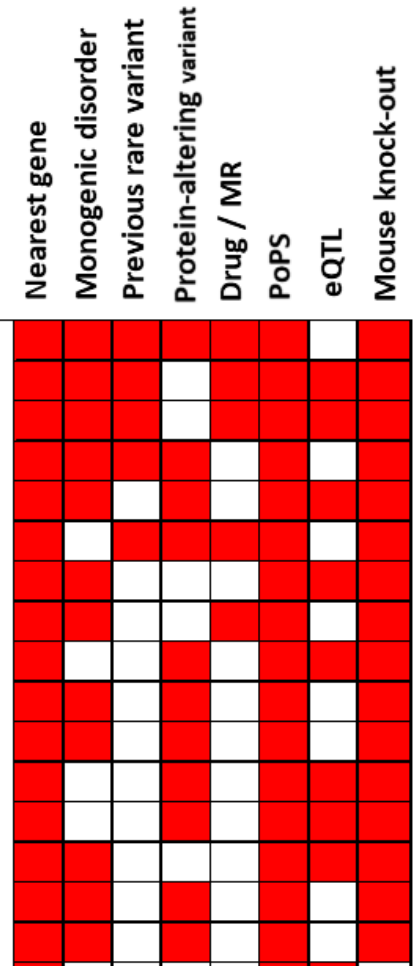
Linking genomic regions with coronary artery disease

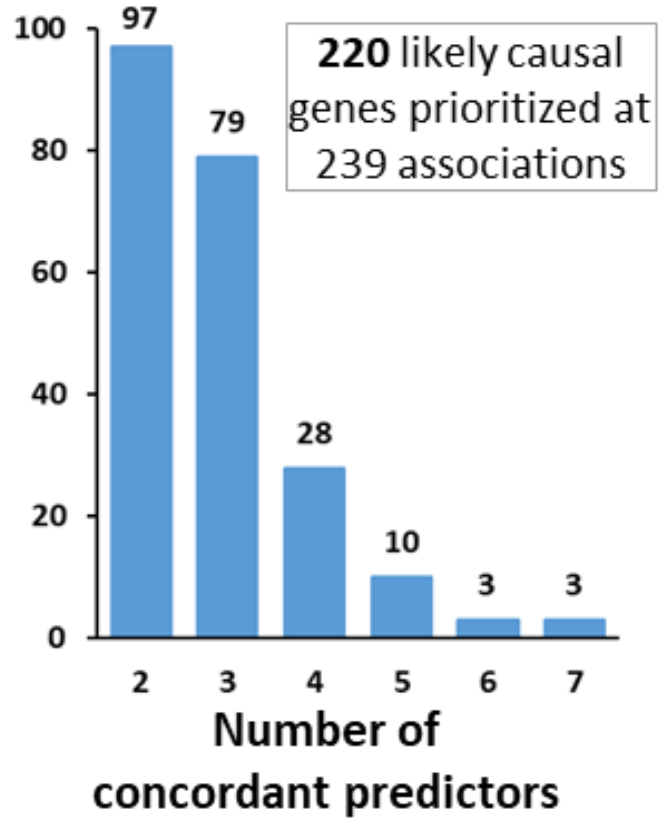


Aragam *et al, Nature Genetics*, *in press*

Integrating predictive features to identify ~200 potential causal effector genes

Sentinel variant	Prioritized gene(s)	Number of supporting predictors
rs11591147	*PCSK9*	7
rs3918226	*NOS3*	7
rs72658867	*LDLR*	7
rs268	*LPL*	6
rs1051338	*LIPA*	6
rs116843064	*ANGPTL4*	6
rs12864131	*CDK8*	5
rs515135	*APOB*	5
rs1250247	*FN1*	5
rs2107732	*CCM2*	5
rs2244608	*HNF1A*	5
rs738408	*PNPLA3*	5
rs781663	*REST*	5
rs7485656	*SCARB1*	5
rs7412	*APOE*	5
rs76866386	*ABCG8*	5
.....	.....	..

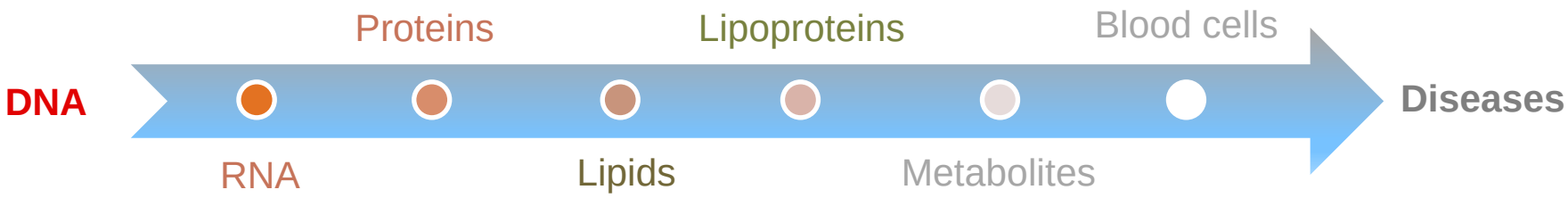




Challenge: better strategies for causal gene prediction?

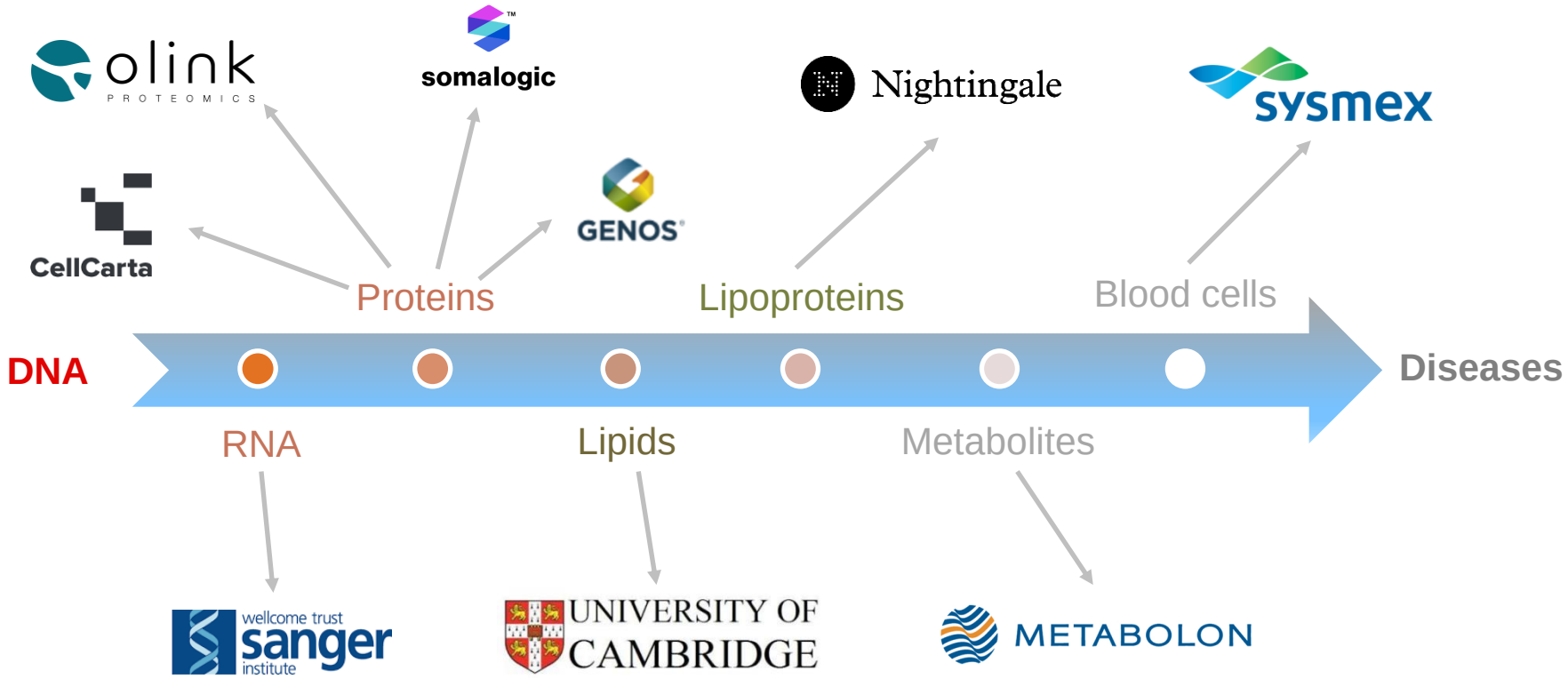
- Weighting different predictors differently?
- Incorporate predictions from CVD risk factors (e.g. blood pressure, lipids)
- Machine-learning approaches like OpenTargets?

Integrating genomics and multi-omics to understand disease aetiology



1. **Colocalisation** with disease GWAS signals
2. **Mendelian randomisation** (both specific & wide-angle)

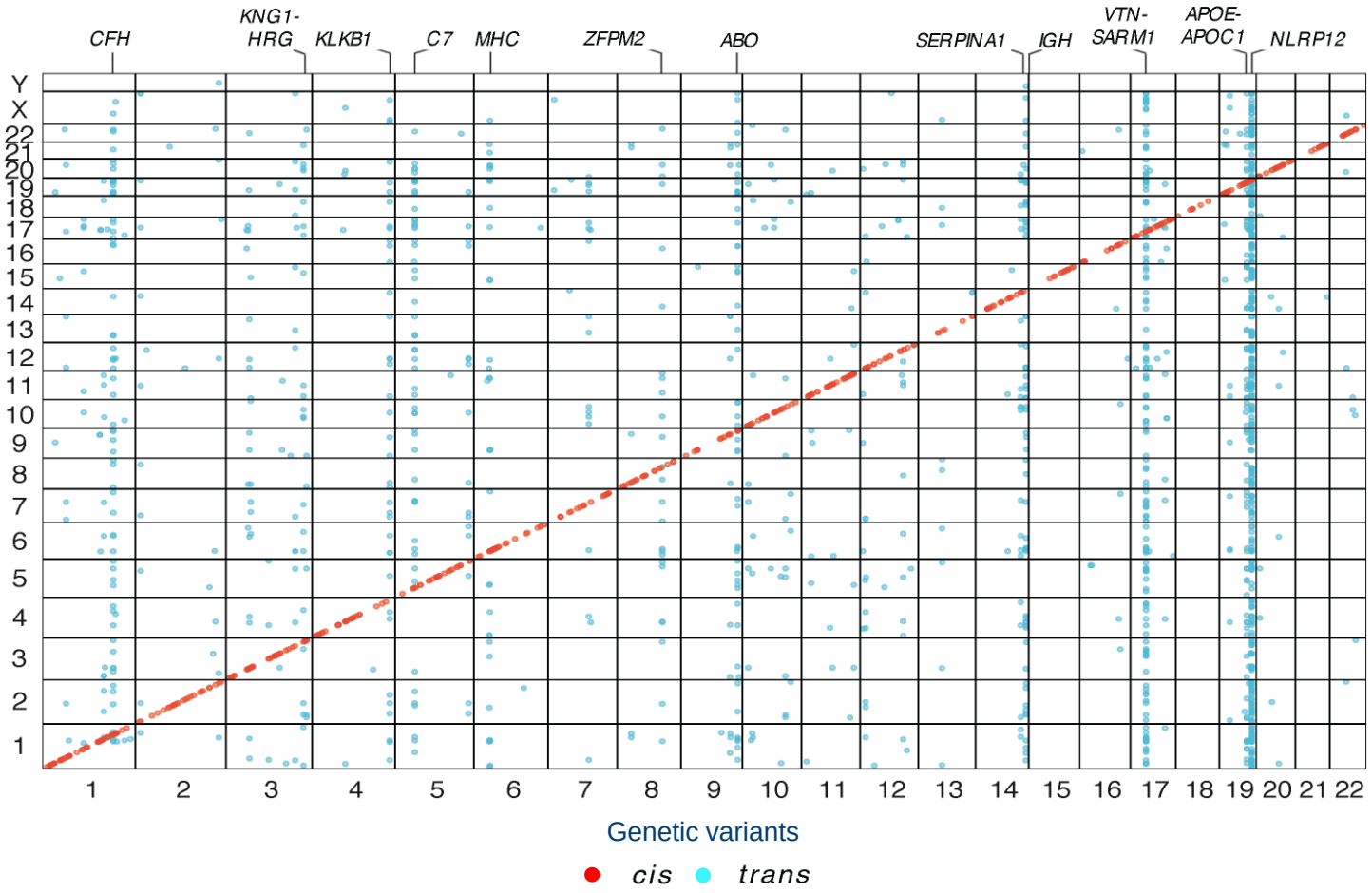
INTERVAL – a 50,000 person multi-omics cohort



GWAS of ~3500 plasma proteins

somalogic

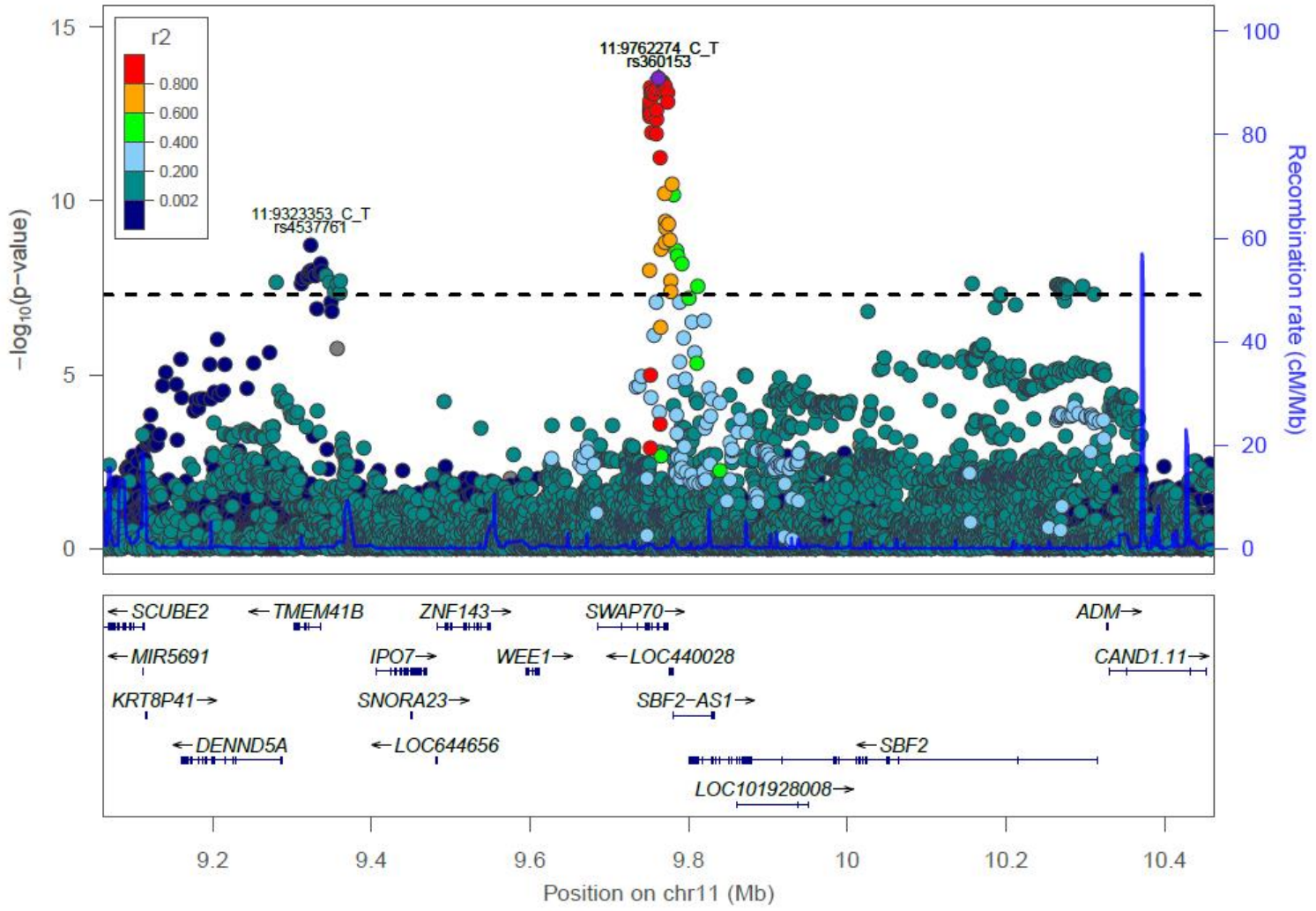
Genes encoding target proteins



3500 proteins
+
3200 participants
=
1900 pQTLs

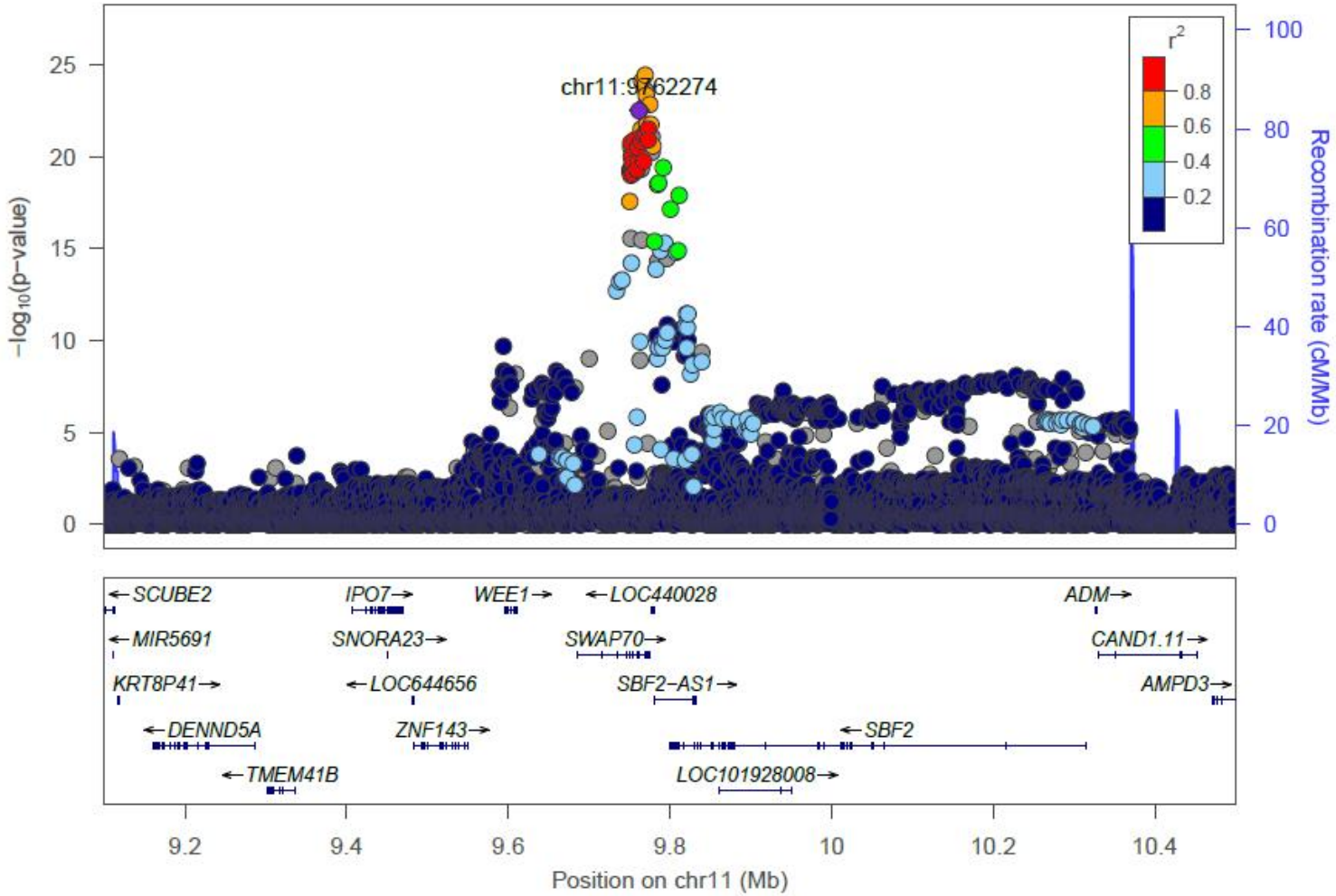
Sun *et al*, *Nature* 2018

Linking genomic regions with coronary artery disease



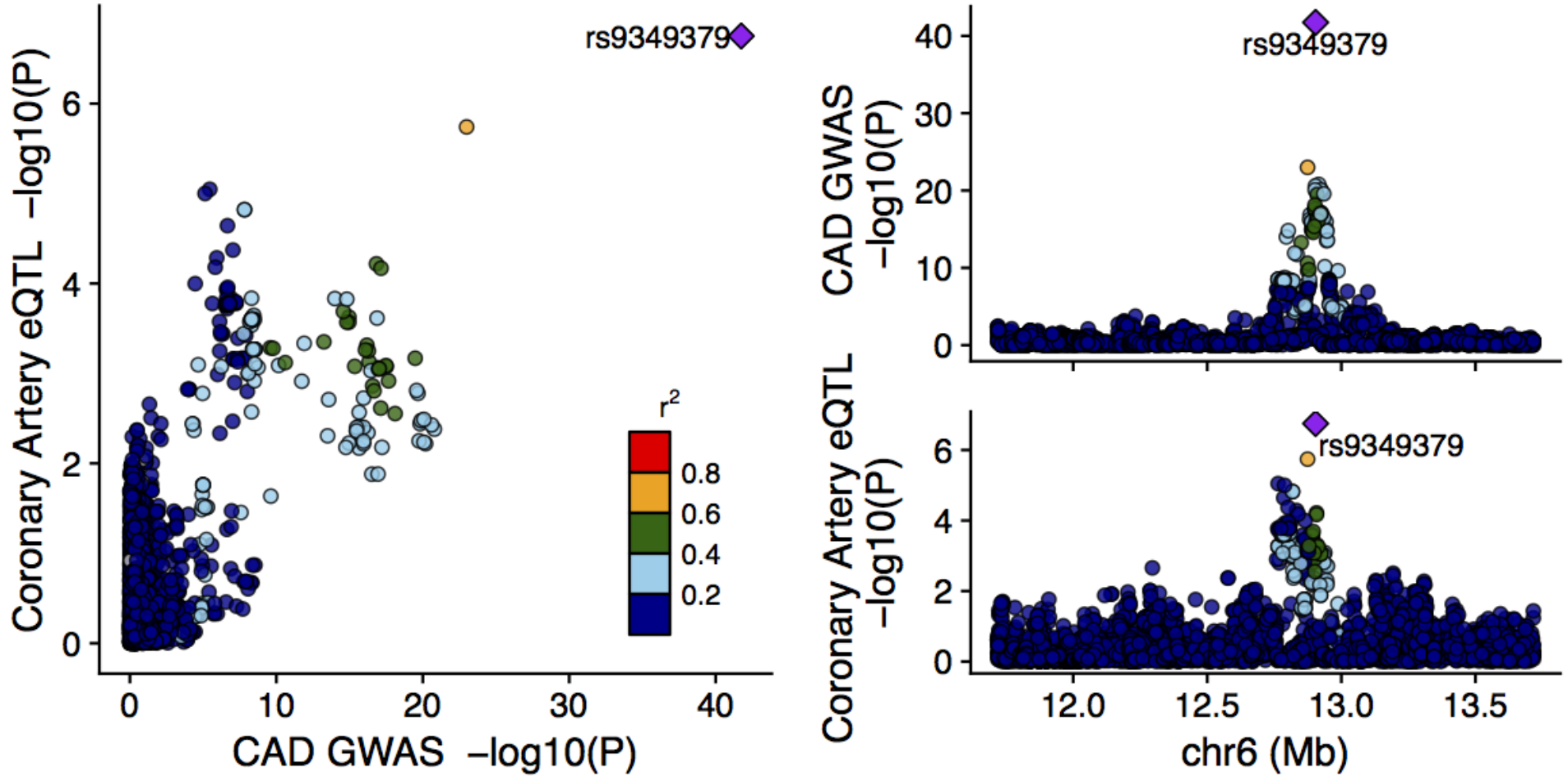
Aragam *et al, Nature Genetics*, *in press*

Coronary disease variants also associate with plasma SWAP70 protein levels

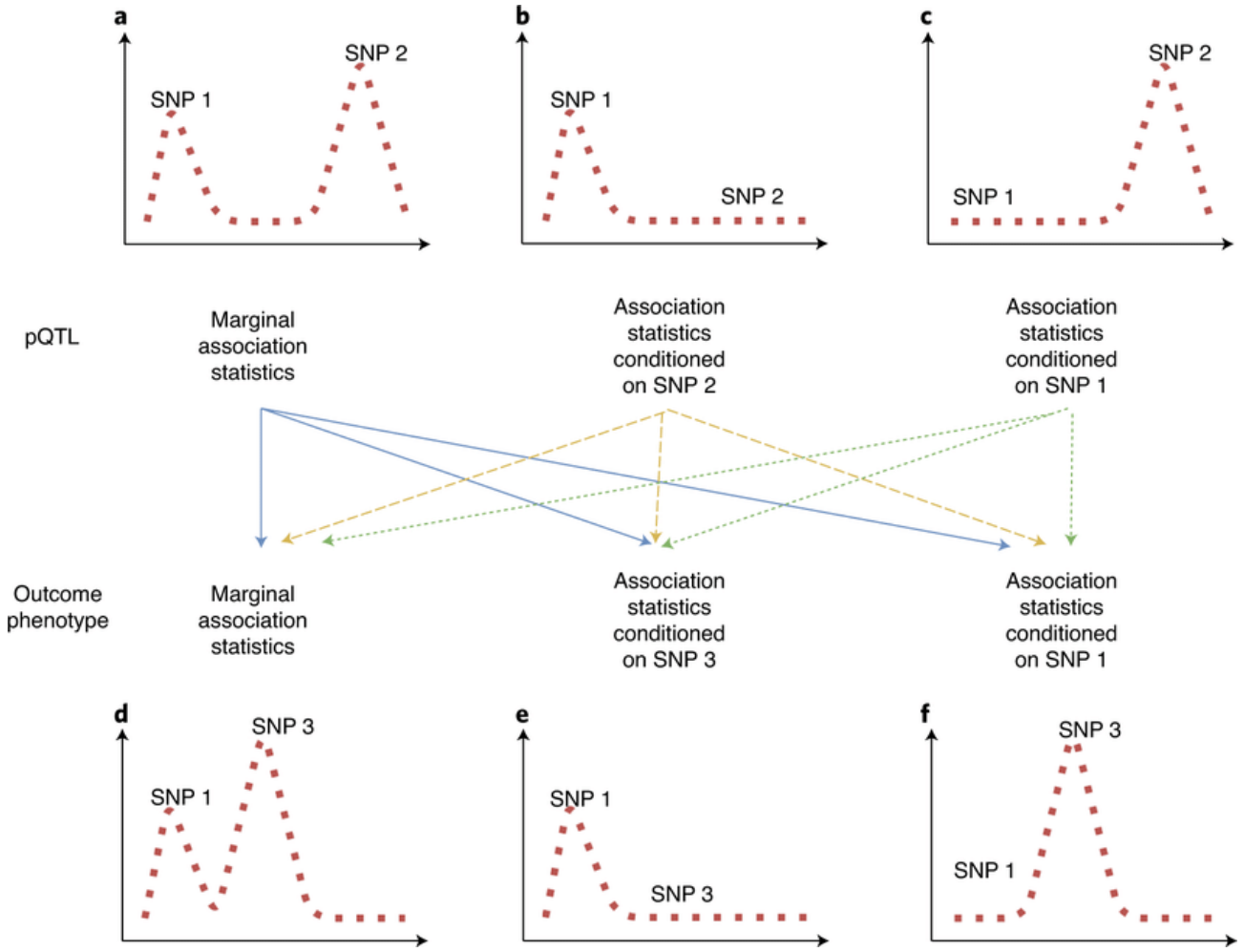


Sun *et al, Nature*, 2018

Statistical colocalisation testing

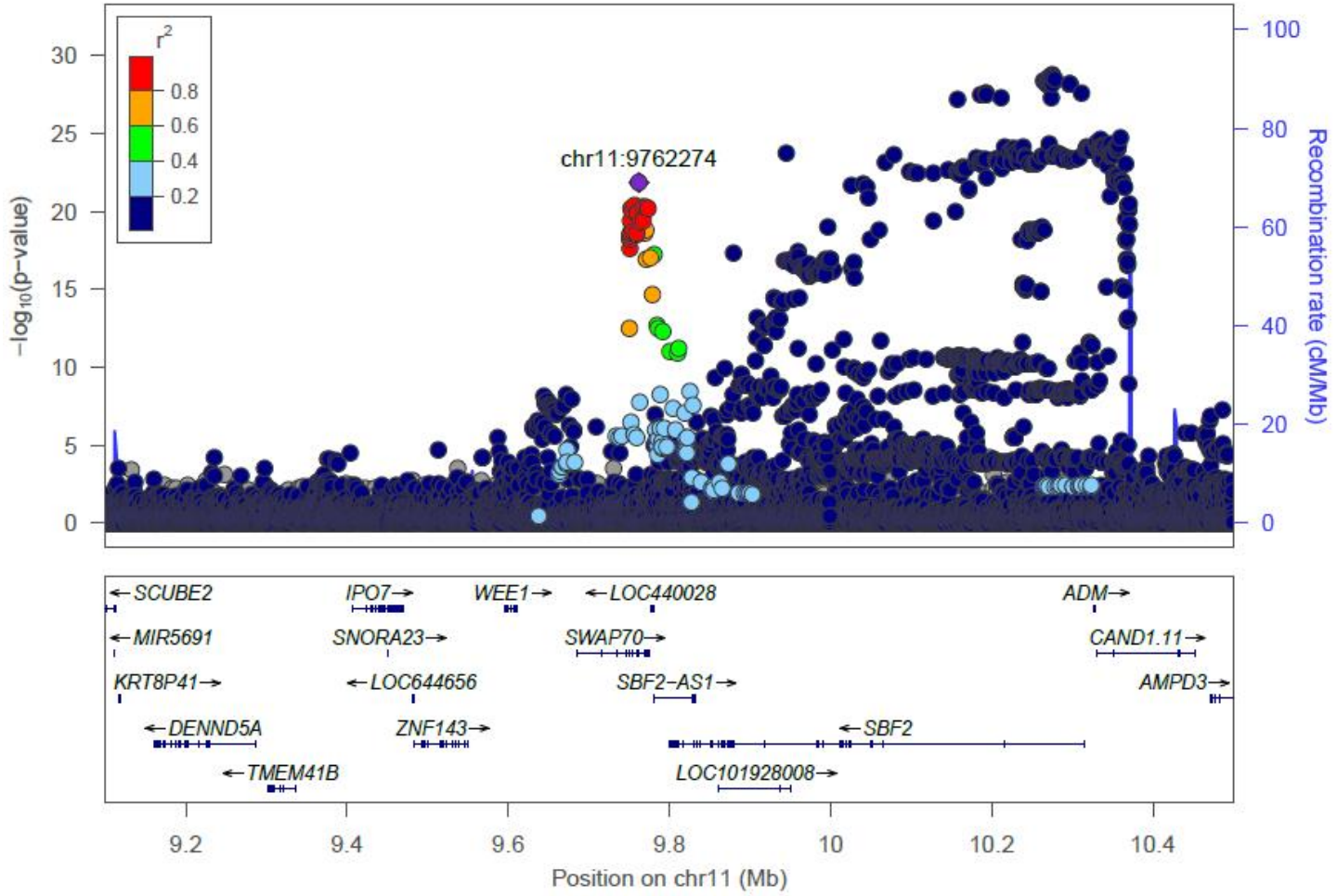


Pairwise conditional colocalisation (PwCoCo) testing



Zheng *et al, Nature Genetics*, 2021

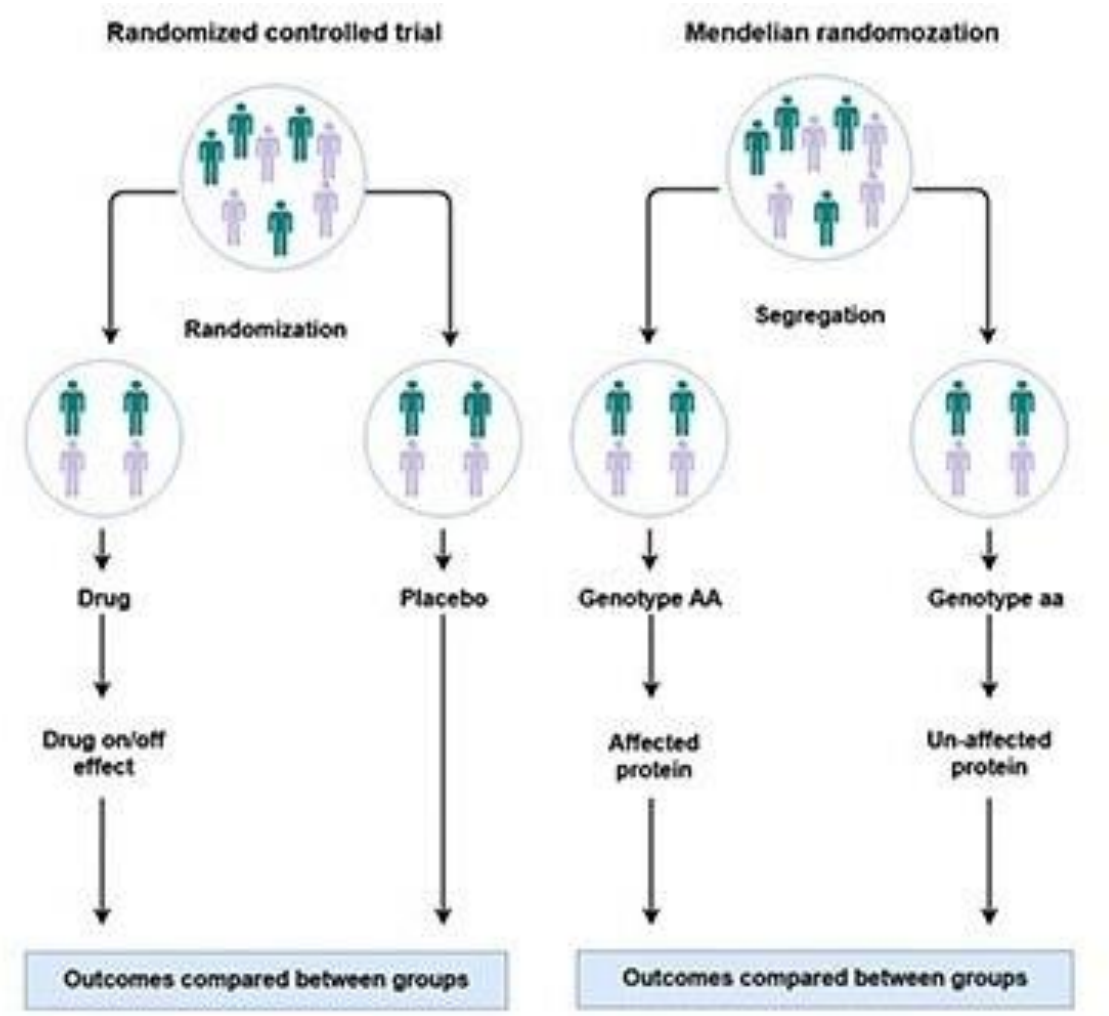
Systolic blood pressure signal at SWAP70 colocalises too…… once the independent neighbouring signal is conditioned out



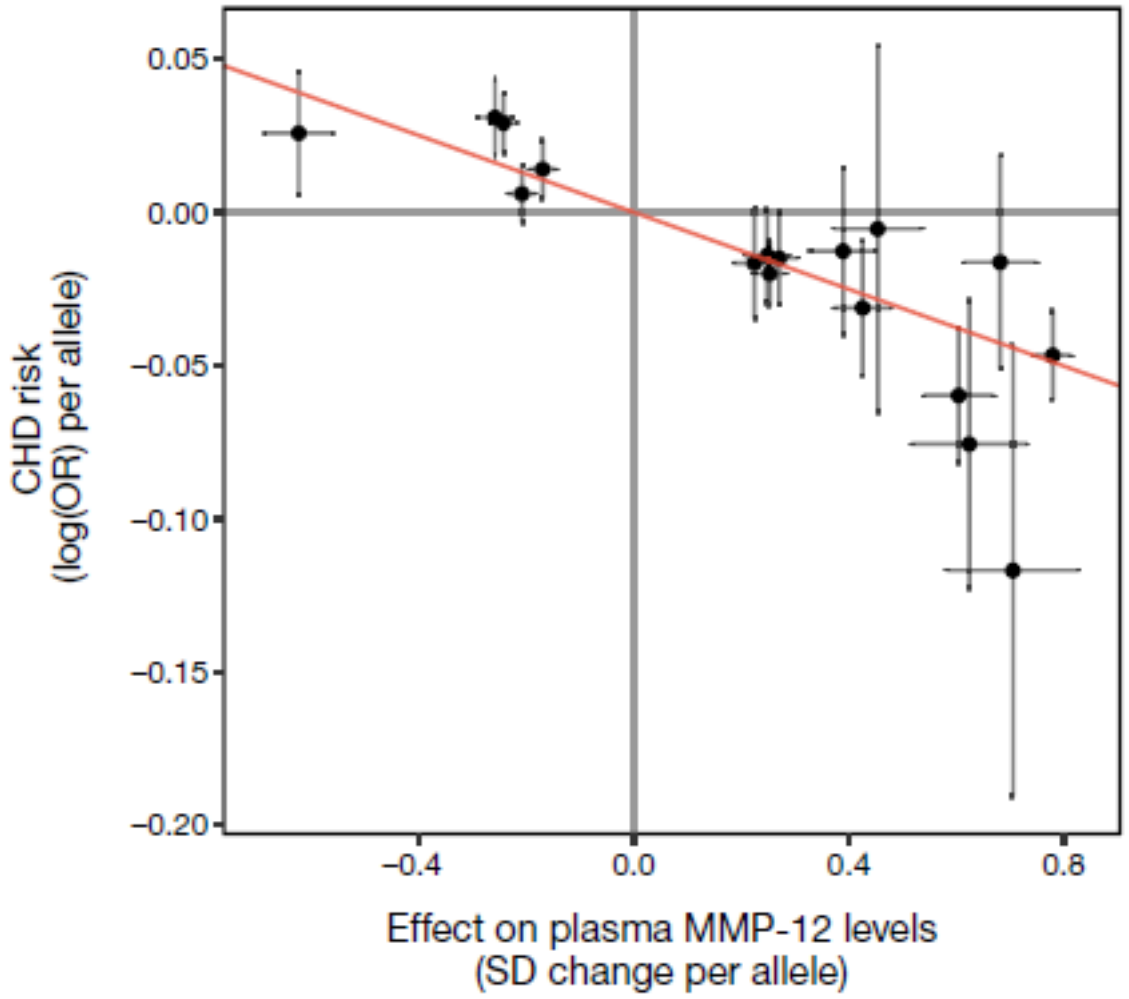
Challenge: enhancing colocalisation methods?

- Methods that account for multiple independent associations?
- Methods that appropriately handle >3 traits?
- Methods that handle different patterns of linkage disequilibrium (between-variant correlation)?

Mendelian randomisation to infer causality of molecular risk factors



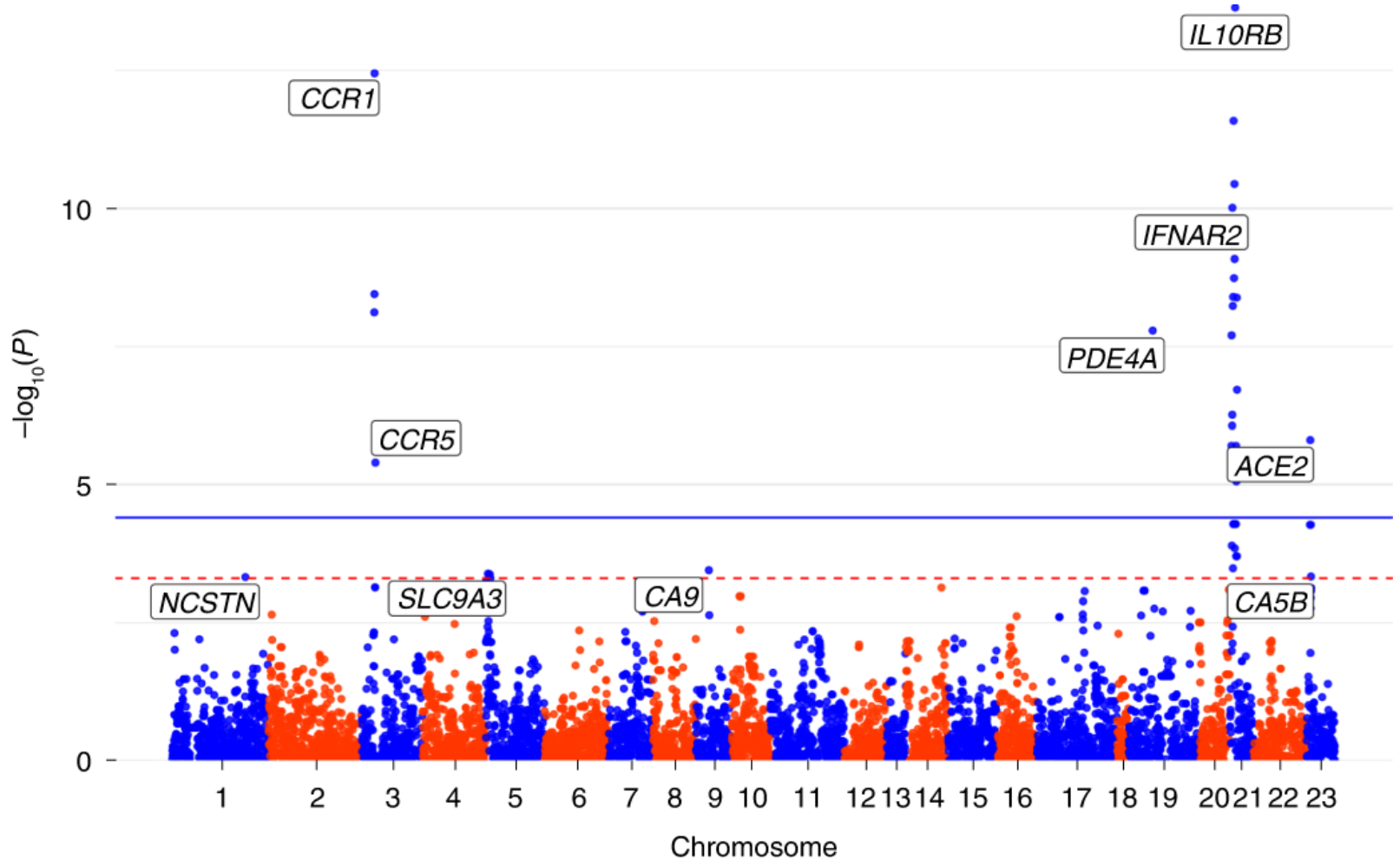
MR suggests inverse causal association of plasma matrix metalloproteinase-12 levels with CAD



- Plasma MMP-12 levels are positively associated with risk of first-onset or recurrent CAD events
- Suggests that MMP-12 is likely to be a protective biomarker, produced in response to vascular damage

Sun *et al., Nature* (2018)

Integrating genomics, transcriptomics, proteomics and COVID-19 outcomes



- *Cis*-MR for eQTLs and pQTLs
- Integrate with COVID-HGI GWAS summary statistics
- Test colocalization to avoid confounding by LD

Gaziano *et al, Nature Medicine*, 2021

Summary

- Large-scale genetic studies are revealing links between molecular risk factors and disease aetiology
- Despite small effect sizes of individual variants, useful predictive tools and causal pathways can be derived
- Genetic epidemiologists are still (largely) using very basic statistical models - huge scope for improvements through enhanced approaches…….

Acknowledgements

- Current and former members of **my group**, especially those who led the projects described (Praveen Surendran, Tao Jiang, Liam Gaziano, Ben Sun, Savita Karthikeyan, Emily Kawabata)
- Members of the **BHF Cardiovascular Epidemiology Unit**, particularly Yu Xu, Mike Inouye, Jimmy Peters, Emanuele di Angelantonio, John Danesh, Clare Oliver-Williams, Dirk Paul
- Collaborators from across the **HDR-UK Multiomics Cohorts Consortium**, particularly Maik Pietzner, Isobel Stewart, Claudia Langenberg, Aroon Hingorani, Ioanna Tzoulaki, Chris Yau, Emily Jefferson, Chris Orton
- **International collaborators**, including Johannes Kettunen, Minna Karjalainen, Krishna Aragam, Sam Lessard, Yannis Hara, Juan-Pablo Casas, CARDIoGRAMplusC4D Consortium, SCALLOP Consortium
- **Funders**, especially the British Heart Foundation, Health Data Research UK, UKRI and industry funders/collaborators