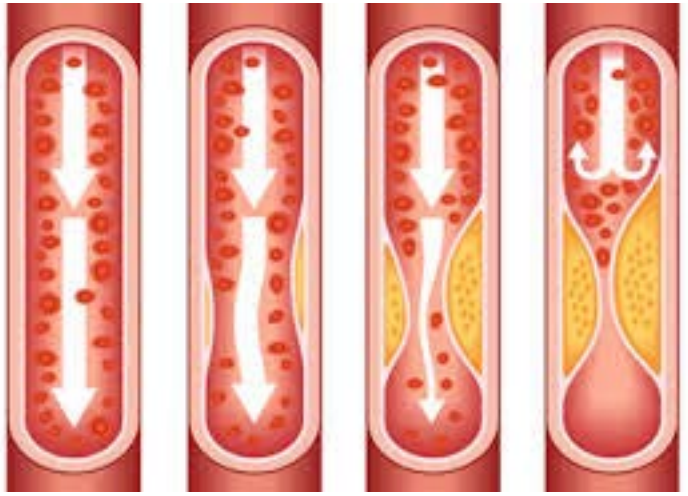
Genetic variation and common human diseases

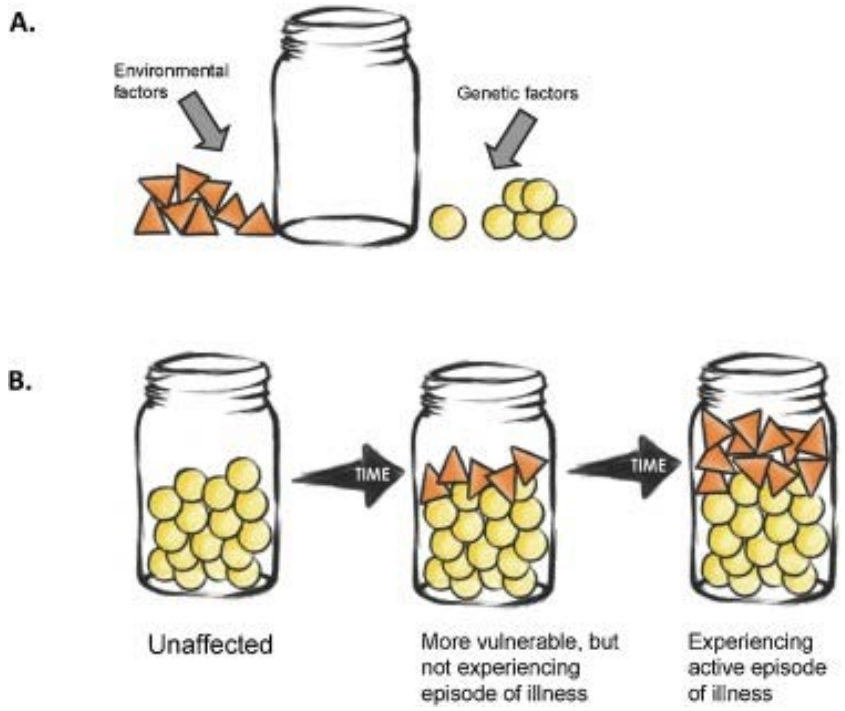
Pak Sham
Director, Centre for Genomic Sciences
The University of Hong Kong
27th November 2018

Common diseases

- Single-gene diseases are distinct entities – patients are clear “outliers” from the general population
- In contrast, common diseases often reflect underlying pathology that is continuous in nature.
- The underlying continuous pathology may be determined by multiple genetic and environmental factors



Multifactorial model for disease



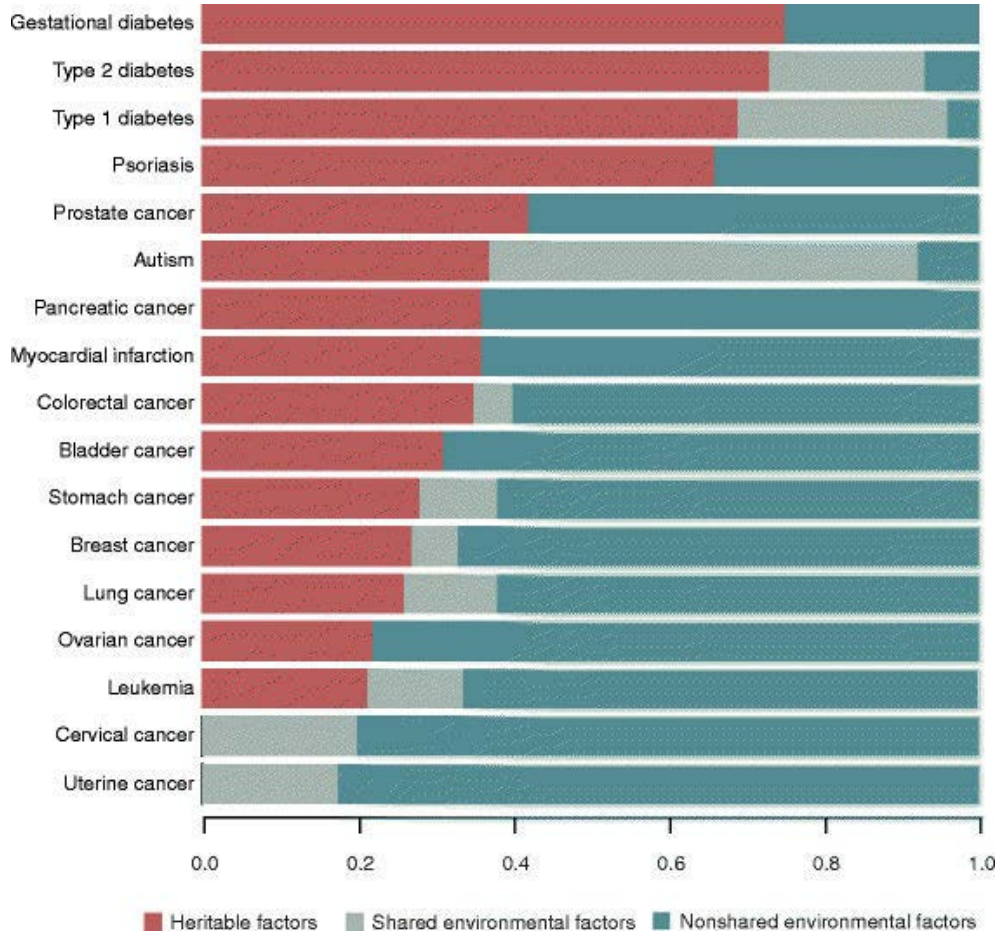
From Hoang, Cytrynbaum & Scherer, 2017

- Numerous genetic and environmental factors contribute to disease
- When total burden of risk factors (liability) reaches a particular level (the threshold), disease occurs

Heritability of common diseases

Total genetic contribution to a disease (heritability) can be estimated from twin or adoption studies

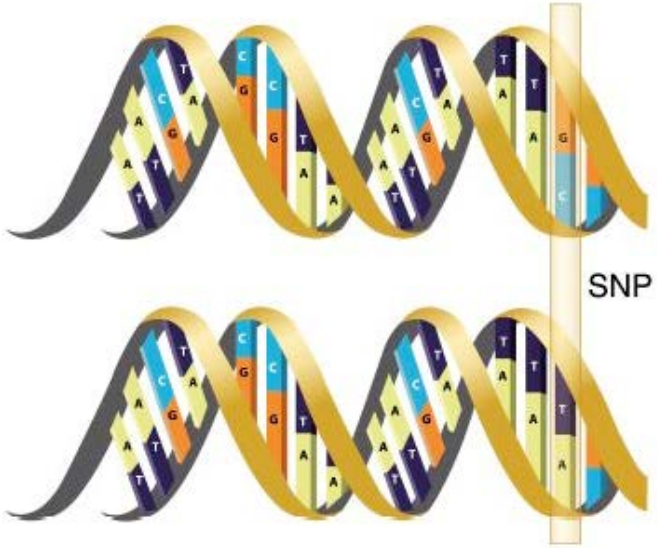
Most common human diseases have moderate to high heritability



Castillo-Fernandez, Spector & Bell, 2014

Genetic variations and disease

- Genetic factors: individual differences in genome sequence that influence disease risk
- Genetic differences (variants / polymorphisms) arose from ancestral mutations.
- Some mutations have large impact and cause rare “monogenic” diseases
- Variants that contribute to common diseases have milder effects and can be common or rare.



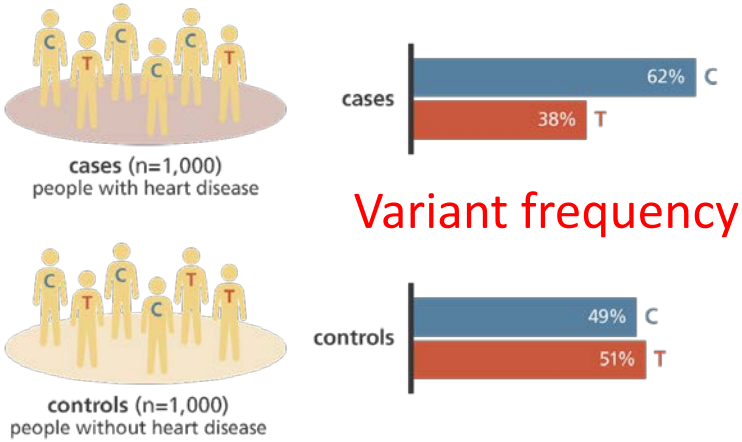
G/T (or C/A) single nucleotide polymorphism (SNP)

http://www.mdsupport.org/images/geneticsexplained2.jpg

Genotyping / Sequencing Technologies

- Almost all common variants in human genome now documented by HapMap and 1K Genomes Projects.
- Efficient SNP arrays (e.g. Illumina iScan) cover nearly all common variants.
- Many rare variants undocumented despite large databases such as ExAC and gnomAD
- Comprehensive rare variants analysis requires high-throughput sequencing technologies (e.g. Illumina NovaSeq)
- High-throughput sequencing remains an order of magnitude more expensive than SNP arrays.

Genome-wide association studies

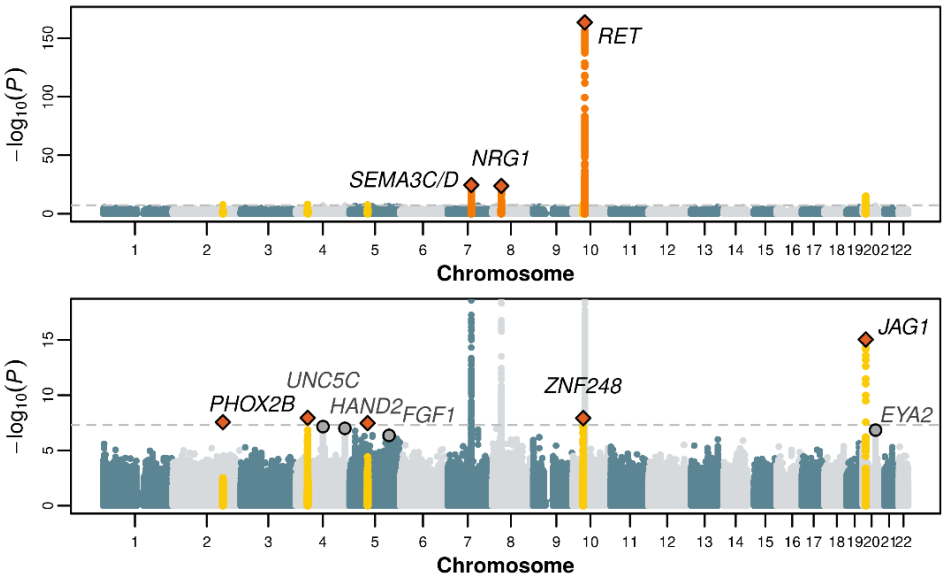


Variant frequency difference

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

"Manhattan" Plots

- Stringent significance threshold to control the number of false positive associations
- Large sample size required to achieve adequate statistical power



Example: schizophrenia

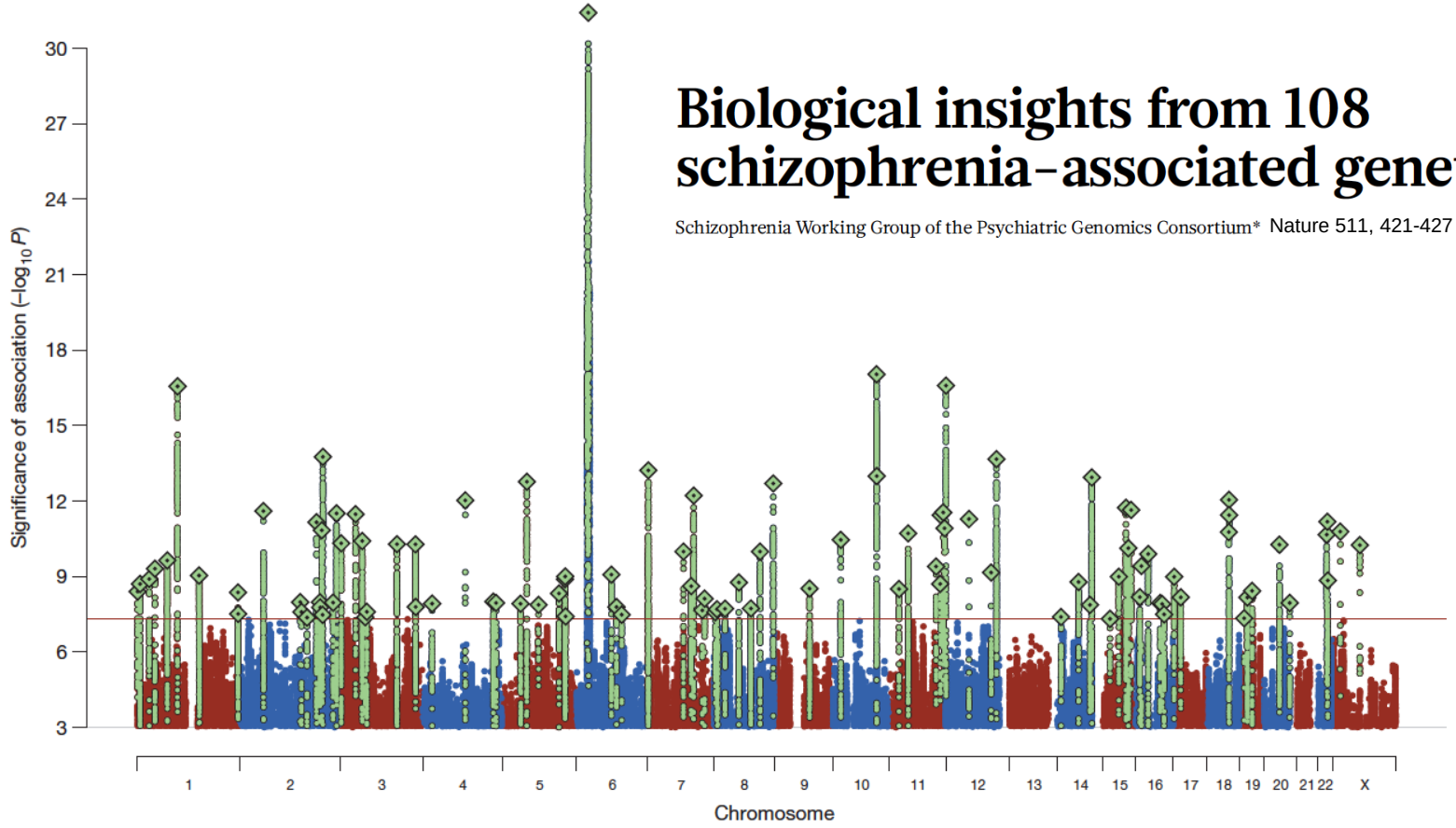
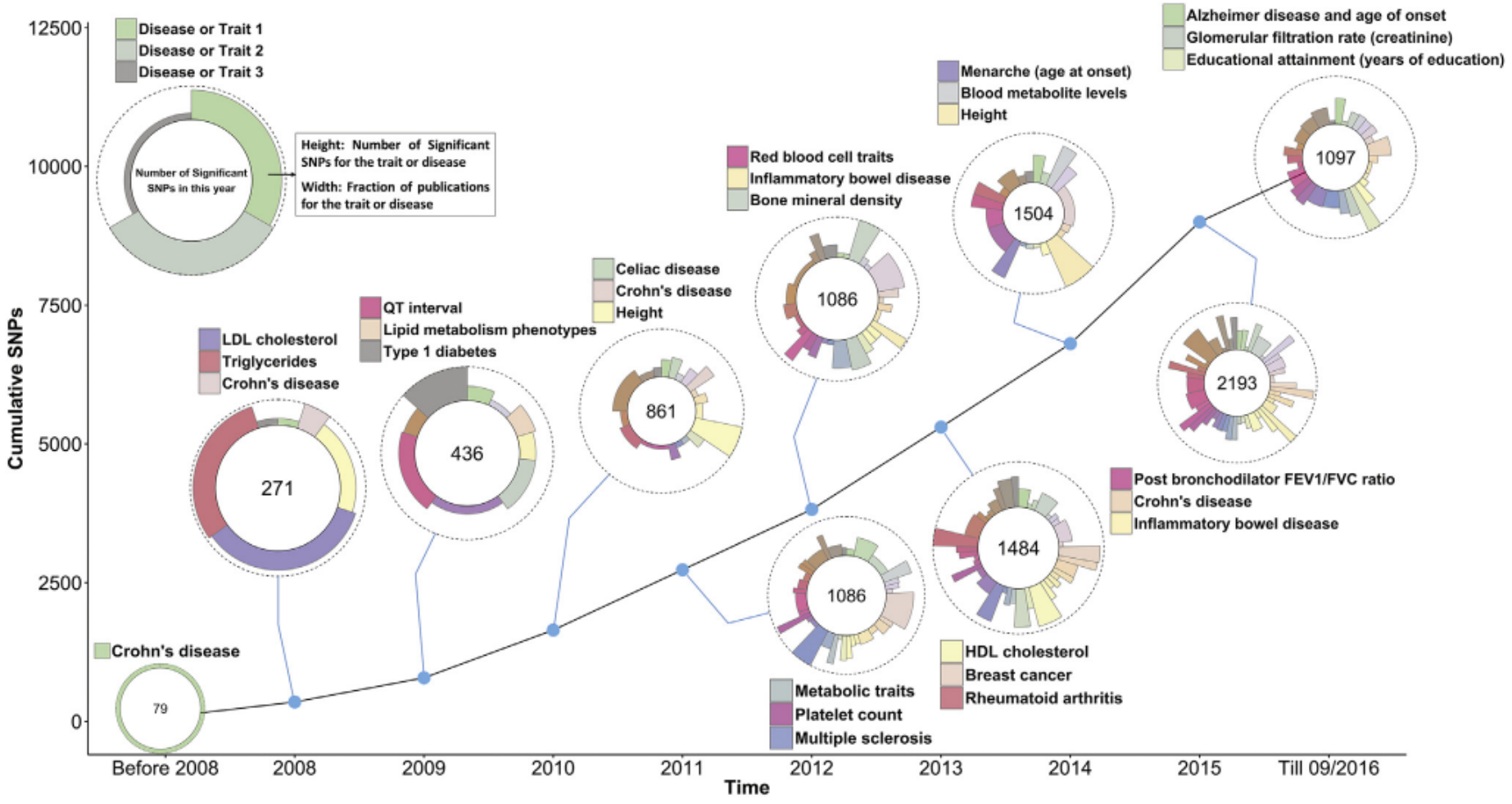


Figure 1 | Manhattan plot showing schizophrenia associations. Manhattan plot of the discovery genome-wide association meta-analysis of 49 case control samples (34,241 cases and 45,604 controls) and 3 family based association studies (1,235 parent affected-offspring trios). The *x* axis is chromosomal position and the *y* axis is the significance ($-\log_{10} P$; 2-tailed) of association derived by logistic regression. The red line shows the genome-wide significance level (5×10^{-8}). SNPs in green are in linkage disequilibrium with the index SNPs (diamonds) which represent independent genome-wide significant associations.

Accumulation of GWAS findings



>10,000 genome-wide significant associations with common diseases / traits

Visscher et al, 2017

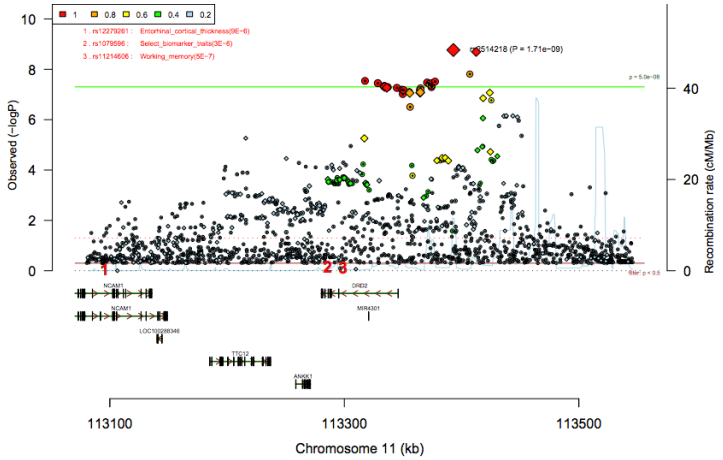
Current status of GWAS

Figure source: https://www.ebi.ac.uk/gwas/

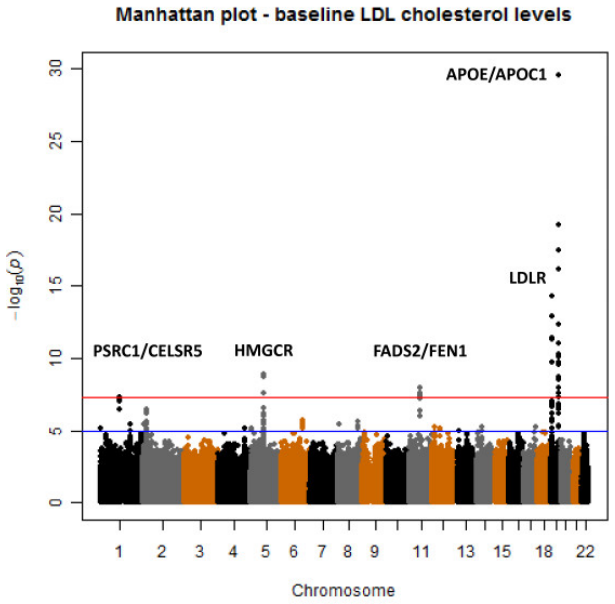
>78,000 robust associations with >300 complex diseases and traits

Predicting drug targets

- Dopamine D2 receptor gene (*DRD2*) associated with schizophrenia
- Dopamine D2 blockade is the mechanism of current antipsychotic drugs

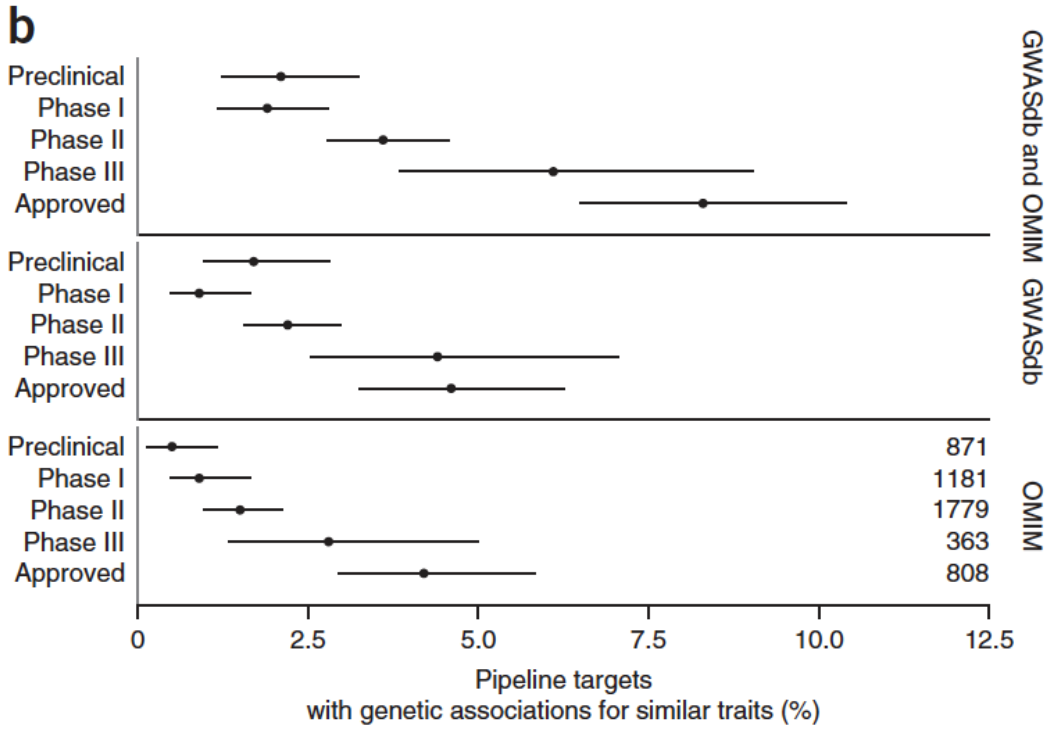


- LDL-associated genetic variant near *HMGCR* decreases LDL-C levels by 2.5 mg/dL
- Statins inhibit enzyme encoded by *HMGCR* and typically decreases LDL-C levels by 14-70 mg/dL



Trompet et al, 2011

Predicting successful drug targets



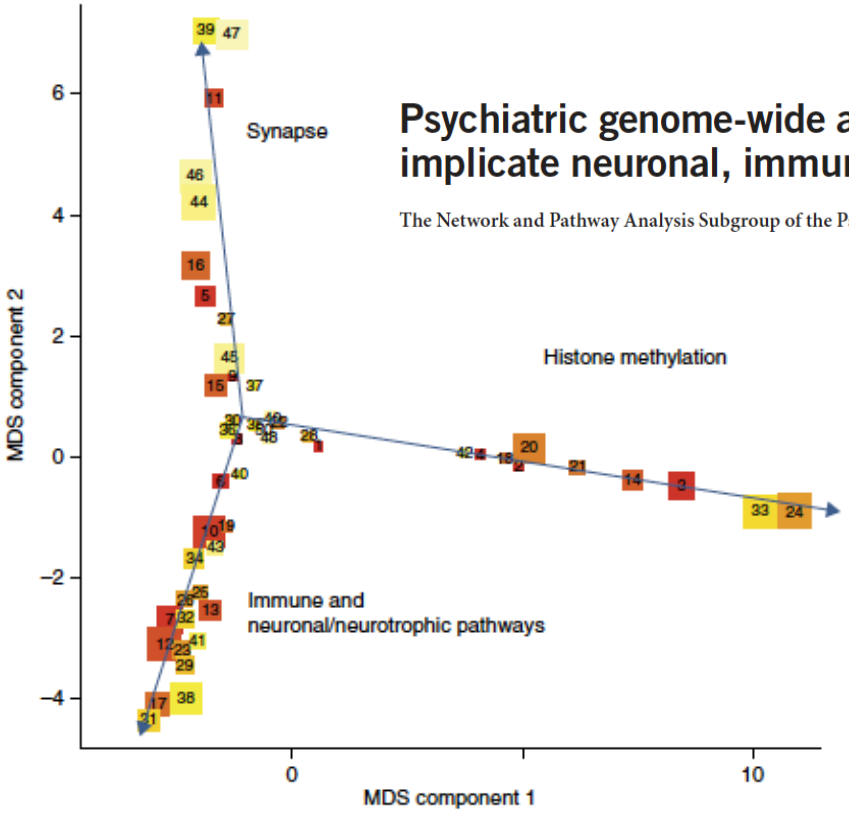
- The proportion of drug targets with direct genetic support increases from 2.0% at the preclinical stage to 8.2% among targets for approved drugs

The support of human genetic evidence for approved drug indications

Matthew R Nelson[1], Hannah Tipney[2], Jeffery L Painter[1], Judong Shen[1], Paola Nicoletti[3], Yufeng Shen[3,4], Aris Floratos[3,4], Pak Chung Sham[5,6], Mulin Jun Li[6,7], Junwen Wang[6,7], Lon R Cardon[8], John C Whittaker[2] & Philippe Sanseau[2]

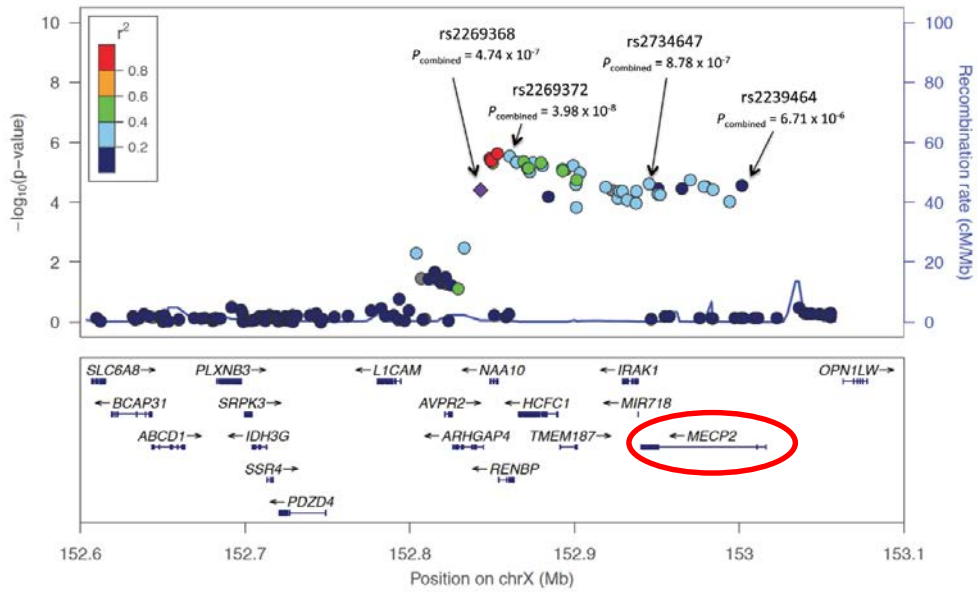
nature genetics

Implicating biological pathways



Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways

The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium*



Common Variants on Xq28 Conferring Risk of Schizophrenia in Han Chinese

Emily H. M. Wong[1,11], Hon-Cheong So[1,11], Miaoxin Li[1,2], Quang Wang[3], Amy W. Butler[1,4], Basil Paul[1], Hei-Man Wu[1], Tomy C. K. Hui[1], Siu-Chung Choi[1], Man-Ting So[5], Maria-Mercè Garcia-Barcelo[2,5], Grainne M. McAlonan[6], Eric Y. H. Chen[1], Eric F. C. Cheung[7], Raymond C. K. Chan[8], Shaun M. Purcell[9], Stacey S. Cherny[1,2,10], Ronald R. L. Chen[1], Tao Li[3], and Pak-Chung Sham*[,1,2,10]

Schizophrenia Bulletin vol. 40 no. 4 pp. 777–786, 2014

Converging rare & common variants

Table 1 Significant CNV loci from gene-based association test

Chr.	Start	End	Locus (gene)	Status	Putative mechanism	CNV test	Direction	FWER	BH-FDR	CAS	CON	Regional *P*	OR (95% CI)
22	17400000	19750000	22q11.21	Previously implicated	NAHR	Loss	Risk	Yes	3.54×10^{-15}	64	1	5.70×10^{-18}	67.7 (9.3–492.8)
16	29560000	30110000	16p11.2, proximal	Previously implicated	NAHR	Gain	Risk	Yes	5.82×10^{-10}	70	7	2.52×10^{-12}	9.4 (4.2–20.9)
2	50000992	51113178	2p16.3 (*NRXN1*)	Previously implicated	NHEJ	Loss	Risk	Yes	3.52×10^{-7}	35	3	4.92×10^{-9}	14.4 (4.2–46.9)
15	28920000	30270000	15q13.3	Previously implicated	NAHR	Loss	Risk	Yes	2.22×10^{-5}	28	2	2.13×10^{-7}	15.6 (3.7–66.5)
1	144646000	146176000	1q21.1	Previously implicated	NAHR	Loss + gain	Risk	Yes	0.00011	60	14	1.50×10^{-6}	3.8 (2.1–6.9)
3	197230000	198840000	3q29	Previously implicated	NAHR	Loss	Risk	Yes	0.00024	16	0	1.86×10^{-6}	INF
16	28730000	28960000	16p11.2, distal	Previously reported	NAHR	Loss	Risk	Yes	0.0029	11	1	5.52×10^{-5}	20.6 (2.6–162.2)
7	72380000	73780000	7q11.23	Previously reported	NAHR	Gain	Risk	Yes	0.0048	16	1	1.68×10^{-4}	16.1 (3.1–125.7)
X	153800000	154225000	Xq28, distal	Novel	NAHR	Gain	Risk	No	0.049	18	2	3.61×10^{-4}	8.9 (2.0–39.9)
22	17400000	19750000	22q11.21	Previously reported	NAHR	Gain	Protective	No	0.024	3	16	4.54×10^{-4}	0.15 (0.04–0.52)
7	64476203	64503433	7q11.21 (*ZNF92*)	Novel	NAHR	Loss + gain	Protective	No	0.033	131	180	6.71×10^{-4}	0.66 (0.52–0.84)
13	19309593	19335773	13q12.11 (*ZMYM5*)	Novel	NHAR	Gain	Protective	No	0.024	15	38	7.91×10^{-4}	0.36 (0.19–0.67)
X	148575477	148580720	Xq28 (*MAGEA11*)	Novel	NAHR	Gain	Protective	No	0.044	12	36	1.06×10^{-3}	0.35 (0.18–0.68)
15	20350000	20640000	15q11.2	Previously implicated	NAHR	Loss	Risk	No	0.044	98	50	1.34×10^{-3}	1.8 (1.2–2.6)
9	831690	959090	9p24.3 (*DMRT1*)	Novel	NHEJ	Loss + gain	Risk	No	0.049	13	1	1.35×10^{-3}	12.4 (1.6–98.1)
8	100094670	100958984	8q22.2 (*VPS13B*)	Novel	NHEJ	Loss	Risk	No	0.048	7	1	1.74×10^{-3}	14.5 (1.7–122.2)
7	158145959	158664998	7*p36.3* (*VIPR2*; *WDR60*)	Previously reported	NAHR	Loss + gain	Risk	No	0.046	20	6	5.79×10^{-3}	3.5 (1.3–9.0)

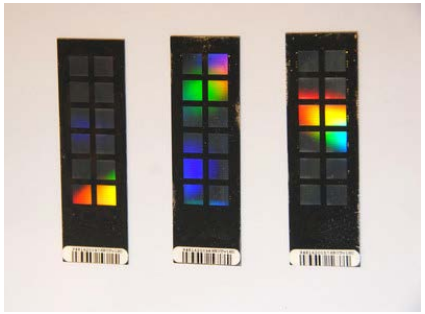
- Presynaptic adhesion molecules *NRXN1*, *NRXN2*
- Postsynaptic scaffolding proteins *DLG1*, *DLG2*, *DLGAP1*, *SHANK1*, *SHANK2*
- Glutamatergic ionotropic receptors *GRID1*, *GRID2*, *GRIN1m, GRIA4*
- Dystrophin and its synaptic interacting partners *DMD*, *DTNB*, *SNTB1*, *UTRN*

Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects

CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium*

Associations specific to population

- Genetic variants vary in frequency and effect in different populations
- Analysis of 47,532 East Asians using Exome Chip revealed 14 significant associations not present in other populations.
- Of these, 12 had much lower variant frequency (<1%) in Europeans, and the other 2 had smaller effect size (1 being in the opposite direction)



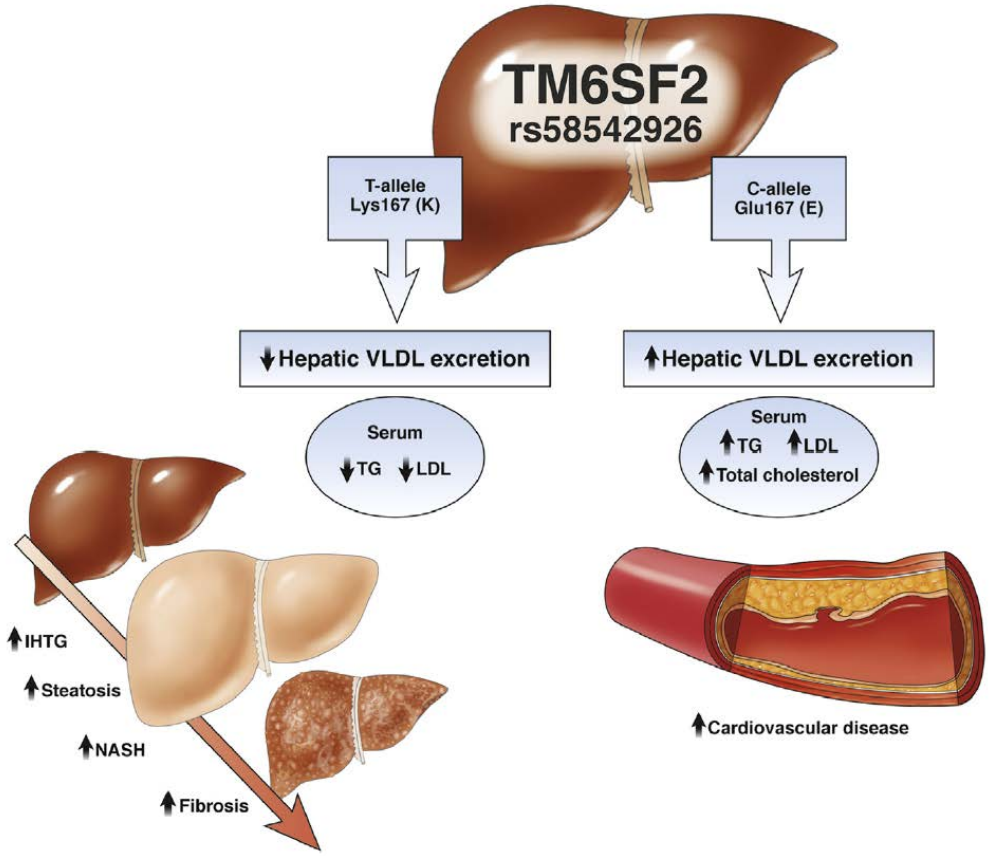
ARTICLES

nature genetics

Exome chip meta-analysis identifies novel loci and East Asian–specific coding variants that contribute to lipid levels and coronary artery disease

Most genome-wide association studies have been of European individuals, even though most genetic variation in humans is seen only in non-European samples. To search for novel loci associated with blood lipid levels and clarify the mechanism of action at previously identified lipid loci, we used an exome array to examine protein-coding genetic variants in 47,532 East Asian individuals. We identified 255 variants at 41 loci that reached chip-wide significance, including 3 novel loci and 14 East Asian–specific coding variant associations. After a meta-analysis including >300,000 European samples, we identified an additional nine novel loci. Sixteen genes were identified by protein-altering variants in both East Asians and Europeans, and thus are likely to be functional genes. Our data demonstrate that most of the low-frequency or rare coding variants associated with lipids are population specific, and that examining genomic data across diverse ancestries may facilitate the identification of functional genes at associated loci.

Variants affected multiple traits



Kahalli et al, Gastroenterology 2015

- *TM6SF2* involved in VLDL efflux from liver to blood
- Increased activity raises blood VLDL level and increases coronary heart disease risk
- Decreased activity leads to increased lipid accumulation in liver and non-alcholic fatty liver disease
- A “Catch-22” situation

Diseases affected by same genes

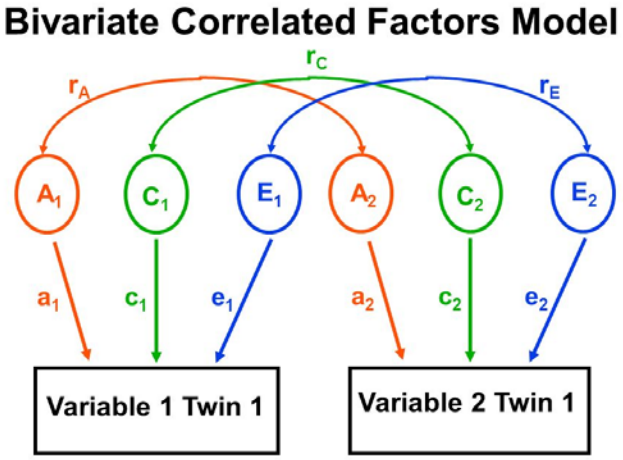
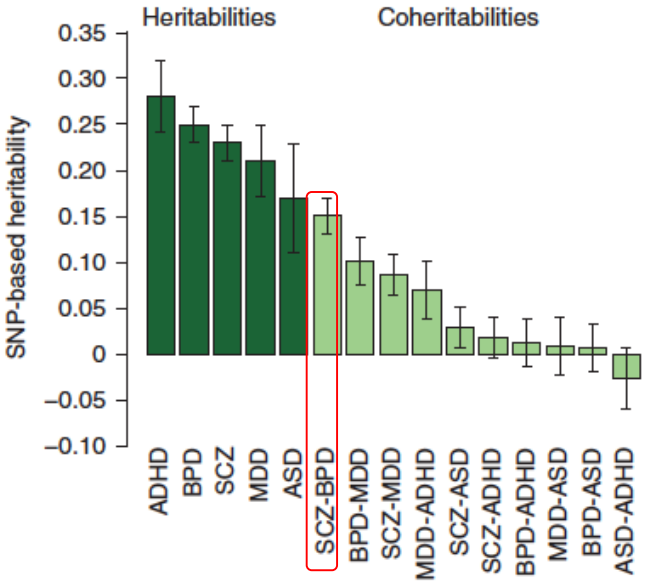


TABLE 3. Parameter Estimates for the Best-Fitting Independent Pathway Model of Liability to Lifetime-Ever RDC Schizophrenic, Schizoaffective, and Manic Syndromes in Monozygotic and Dizygotic Twin Pairs[a]

	Proportion of Variance in Liability					
	Additive Genetic Effects			Individual-Specific Environmental Effects		
Syndrome	Total	Common	Specific	Total	Common	Specific
Schizophrenic	0.82	0.49	0.33	0.18	0.09	0.09
Schizoaffective	0.85	0.85	—	0.15	—	0.15
Manic	0.87	0.68	0.19	0.13	0.06	0.07

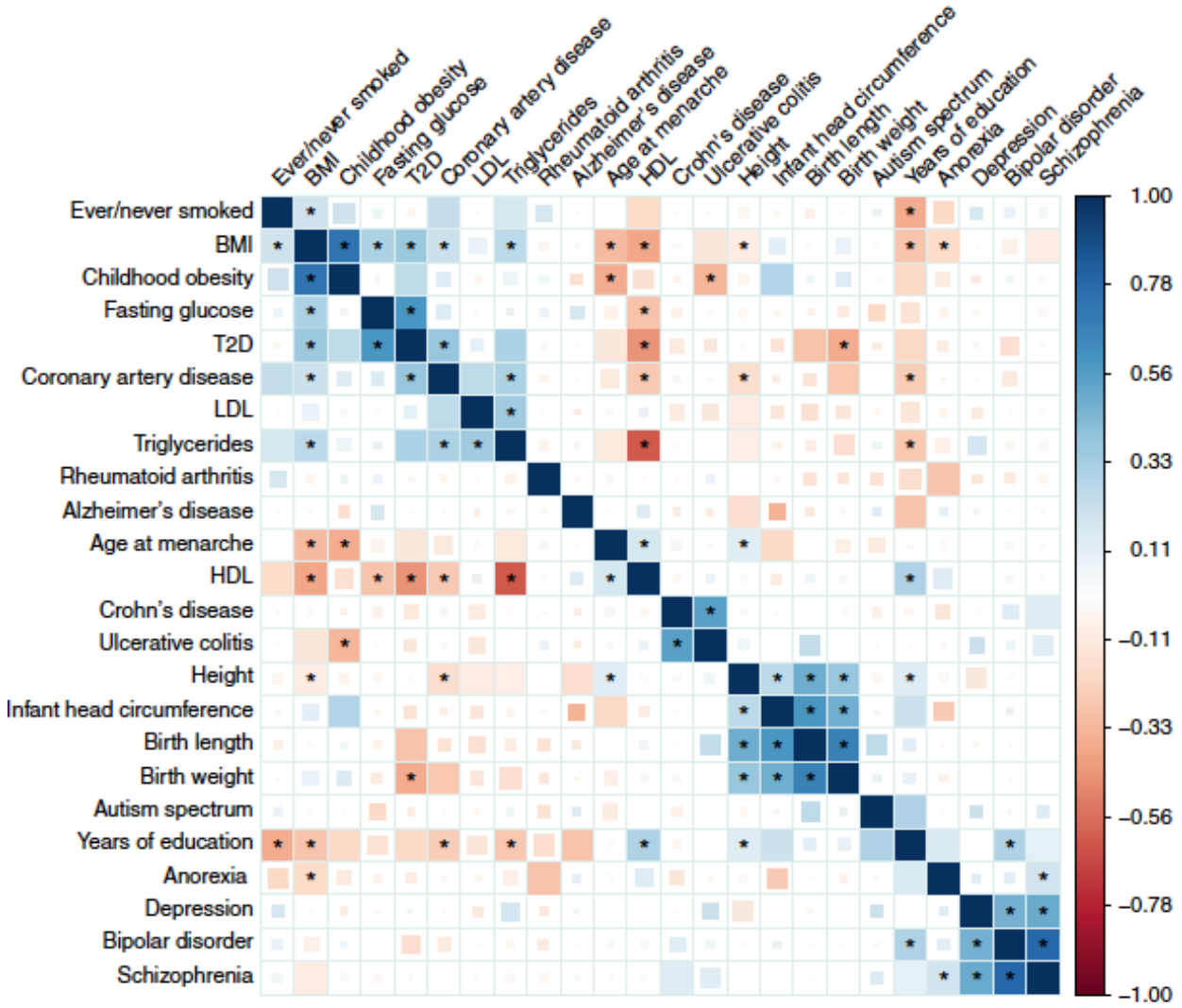
Cardno, Rijsdijk, et al, Am J Psychiat, 2002



- Twin studies indicated substantial genetic sharing between schizophrenia and bipolar affective disorder
- This has been confirmed by GWAS

Cross Disorder Group of the PGC, Nature Genetics 2013

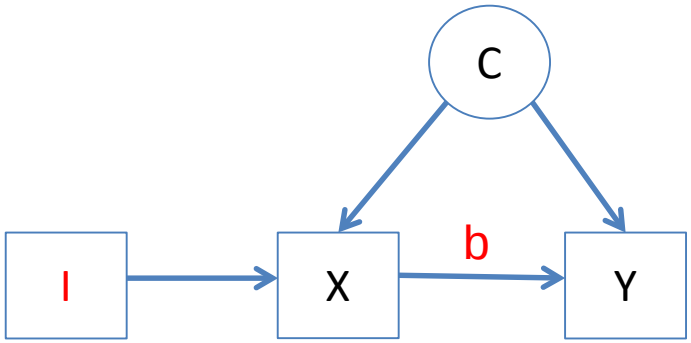
Map of genetically related diseases



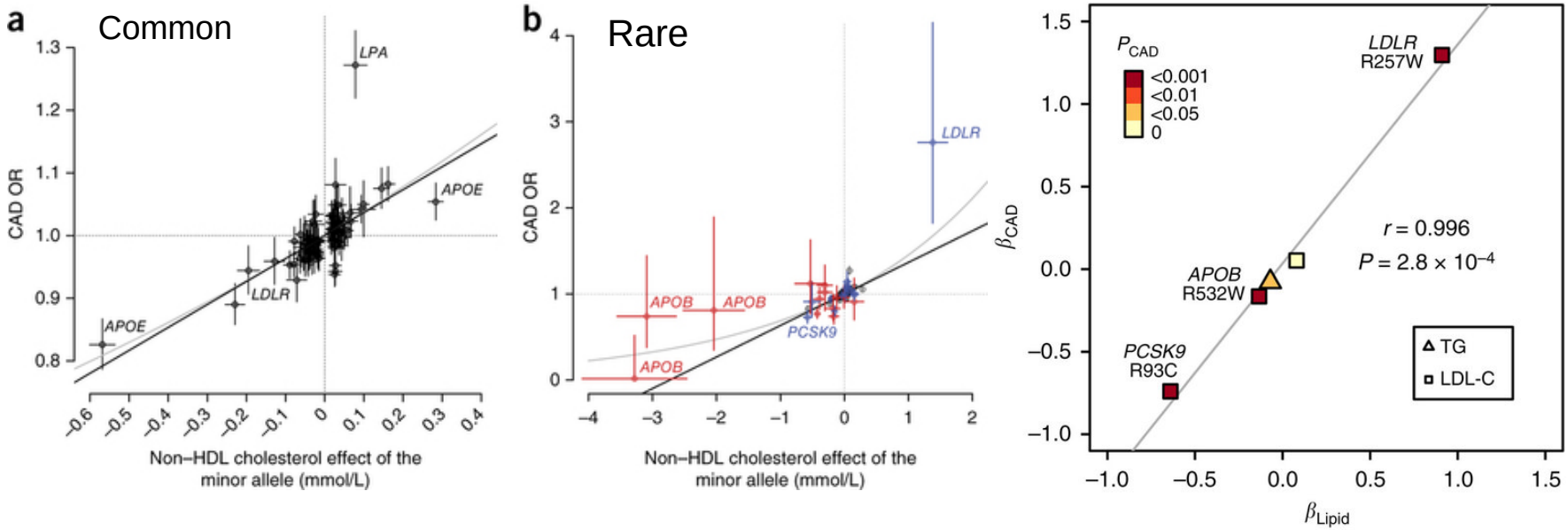
Genetic sharing is greatest for diseases affecting the same cell type / organ / system, e.g. metabolic disorders, psychiatric disorders

Bulik-Sullivan et al, Nature Genetics, 2015

Resolving causal relationships



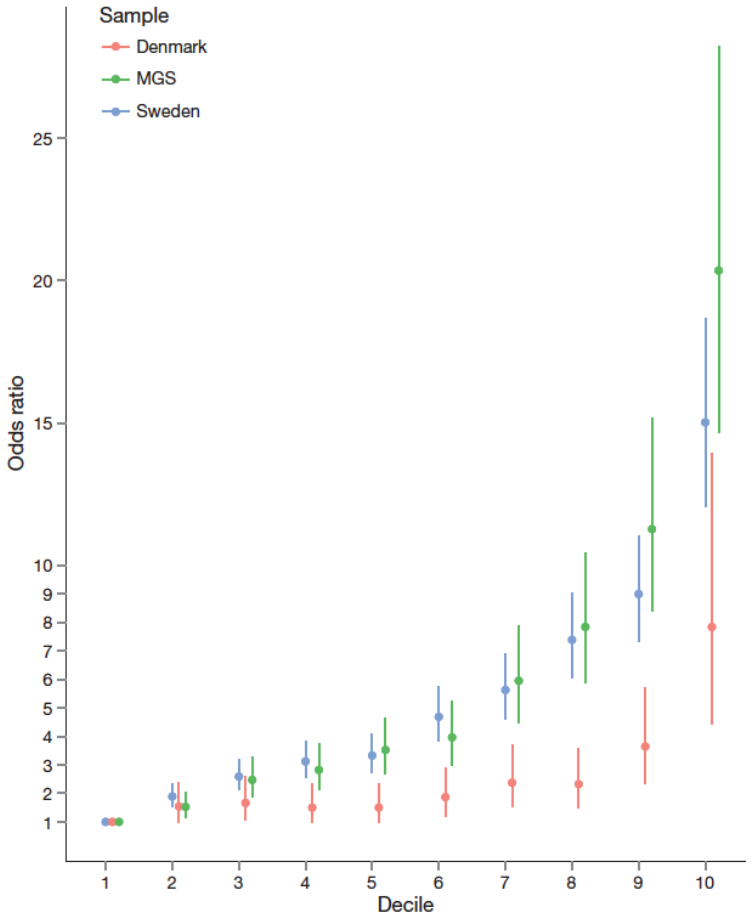
- Causal inference complicated by confounding
- Mendelian randomization: instrumental variable related to outcome only through exposure
- Genetic variants provide multiple valid instruments
- Consistent estimates implicating causal relationship of non-HDL cholesterol on coronary heart disease



Helgadottir *et al*. 2016

Tang *et al*. (2015)

Predicting disease risk

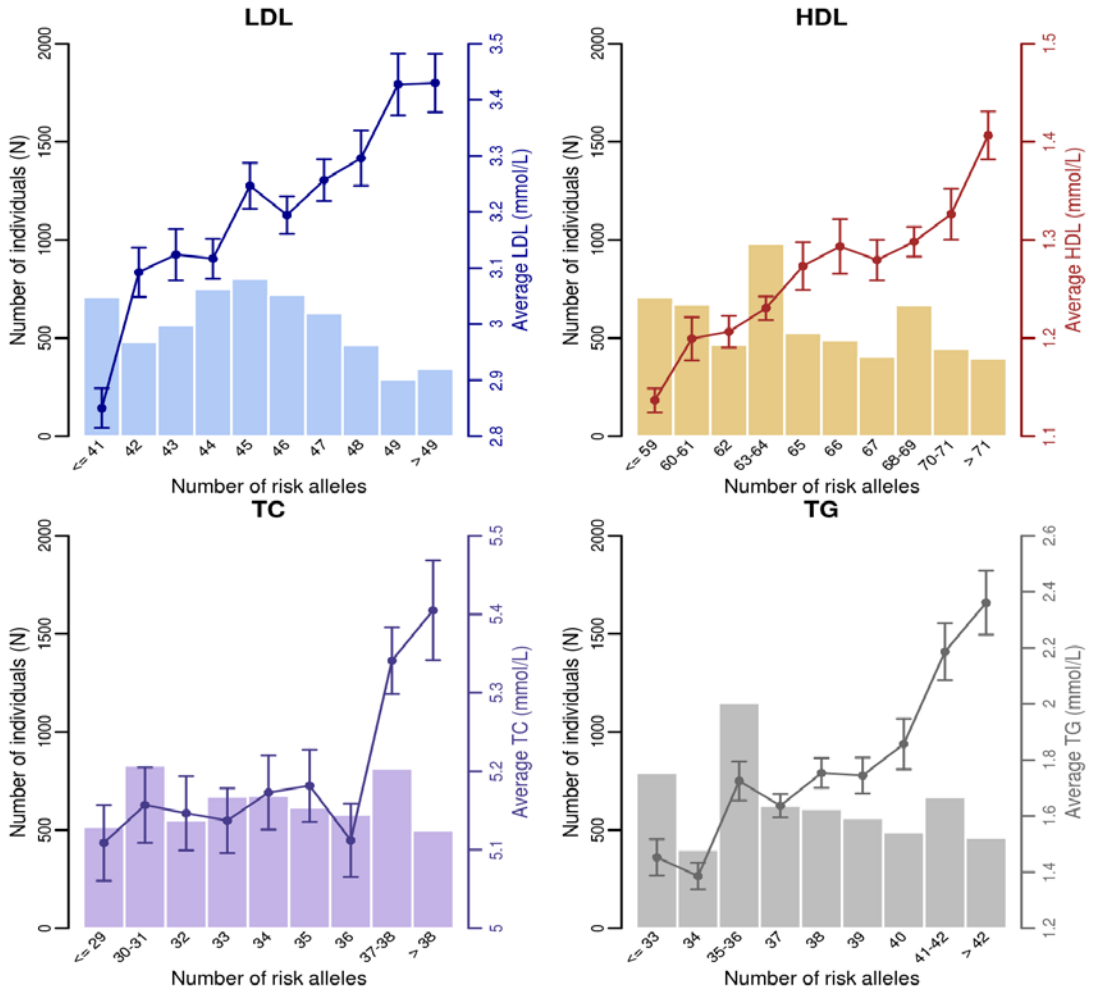


- For a polygenic disease, every person carries a certain number of high-risk genetic variants
- Cumulative effect of all high-risk variants present in a person: "polygenic risk score"
- Population can be stratified by polygenic risk score, e.g. into deciles, in ascending disease risk
- For schizophrenia, the highest and lowest deciles have 8-20 fold risk difference (similar effect as a positive family history)

Biological insights from 108 schizophrenia-associated genetic loci

Schizophrenia Working Group of the Psychiatric Genomics Consortium*

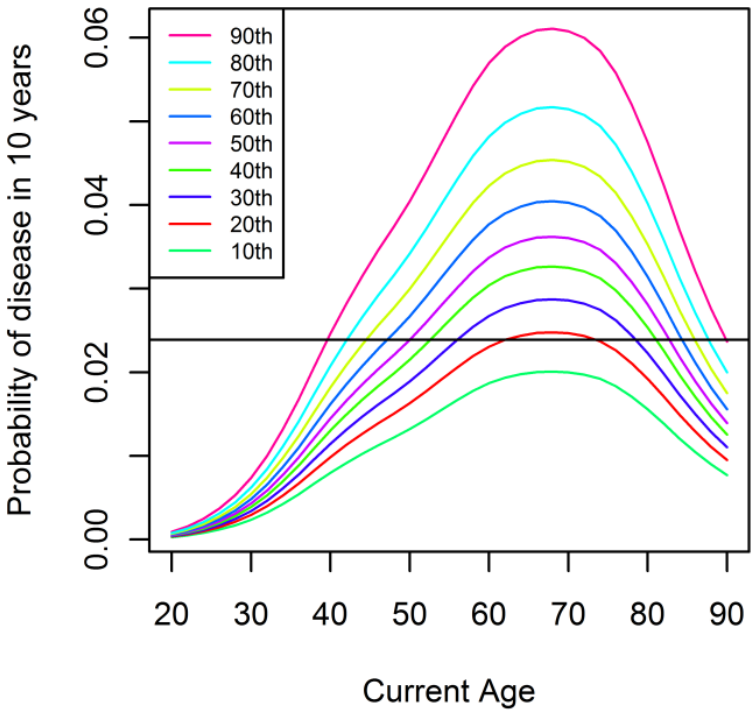
Polygenic scores for risk factors



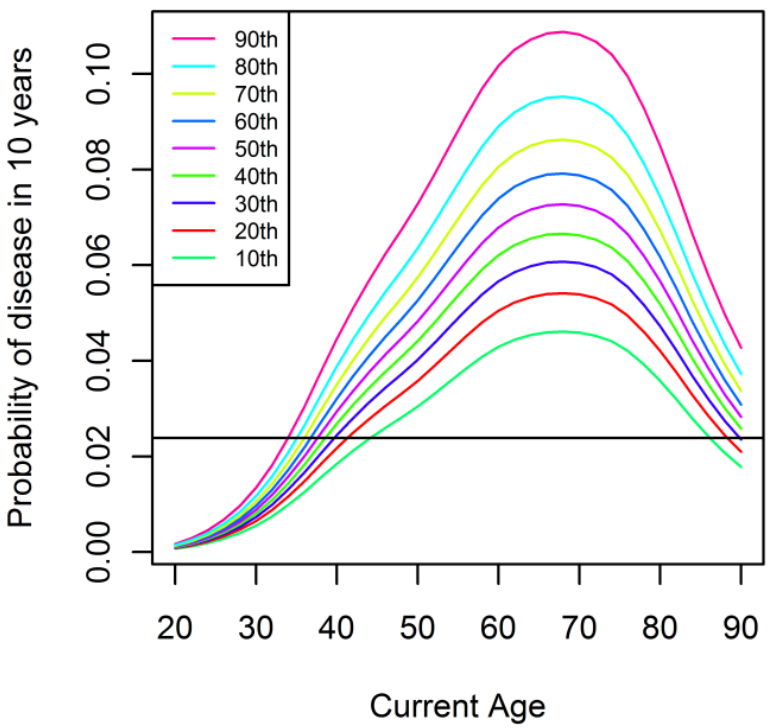
- The effects of some genetic variants on disease risk may be mediated through known risk factors
- Blood lipid levels are risk factors for coronary heart disease (CAD)
- "Polygenic scores" for blood lipid levels predict CAD risk

Application to cancer screening

General population



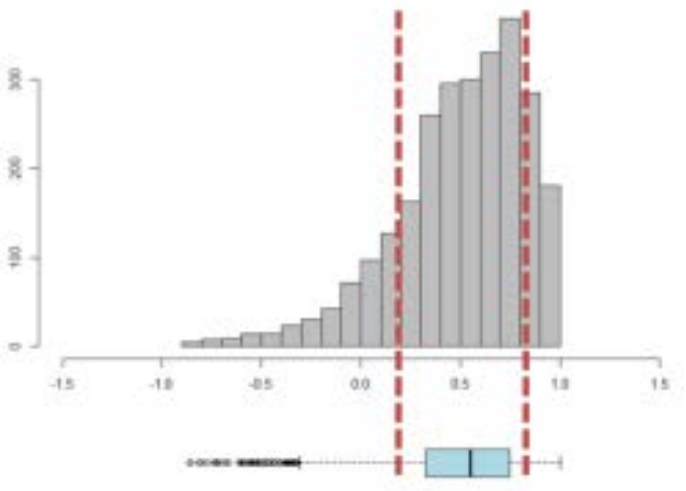
Having affected first degree relative

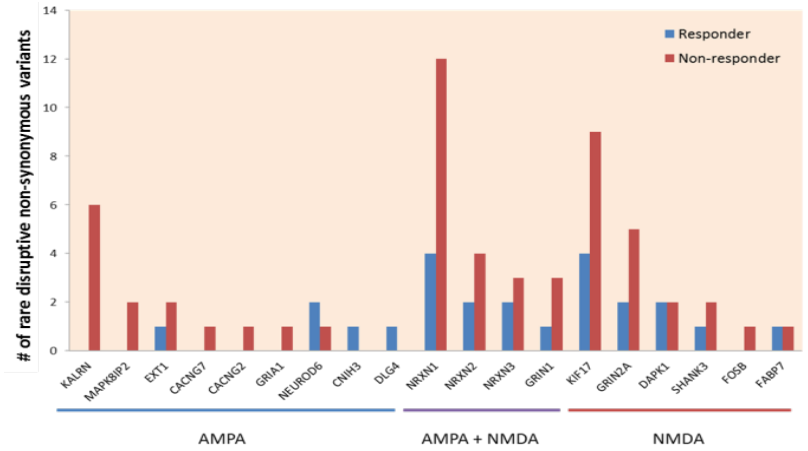


Breast cancer screening may commence earlier in women with high polygenic scores and / or positive family history

So et al, American Journal of Human Genetics 2012

Predicting treatment response





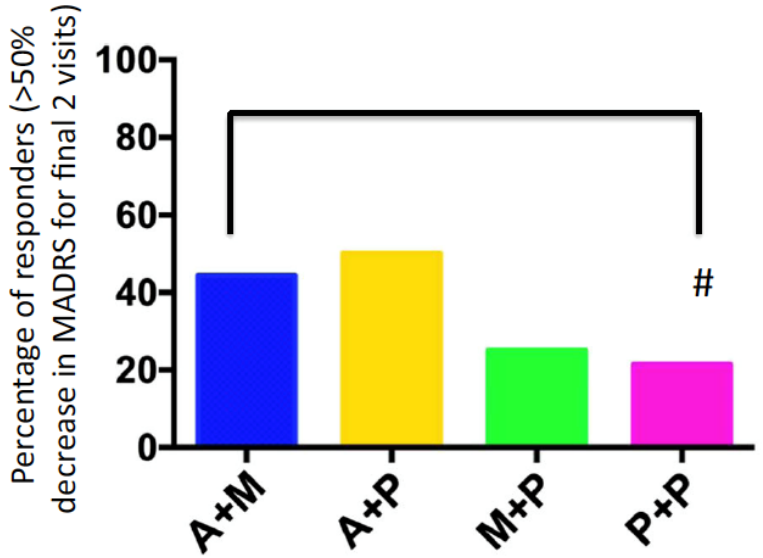
- Two gene-sets with excess rare damaging variants in schizophrenia patients not responsive to antipsychotic medications
  - Reduced NMDA-mediated synaptic currents
  - Reduced AMPA-mediated synaptic currents

JAMA Psychiatry | Original Investigation

Effect of Damaging Rare Mutations in Synapse-Related Gene Sets on Response to Short-term Antipsychotic Medication in Chinese Patients With Schizophrenia

Repositioning current drugs

- Using tissue-specific e-QTL data (e.g. GTEx), GWAS results can predict disease-associated gene expression changes
- Many existing drugs have documented effects on gene expression.
- A drug that reverses disease-related gene expression changes may be effective in treating the disease
- Example: aspirin was identified as a possible treatment of bipolar disorder.



Savitz et al. Translational Psychiatry (2018)8:27
DOI 10.1038/s41398-017-0073-7
Translational Psychiatry

Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry

Hon-Cheong So[1,2], Carlos Kwan-Long Chau[1], Wan-To Chiu[3], Kin-Sang Ho[3], Cho-Pong Lo[3], Stephanie Ho-Yue Yim[4] & Pak-Chung Sham[5–8]

Nature Neuroscience, 2017

ARTICLE Open Access

Treatment of bipolar depression with minocycline and/or aspirin: an adaptive, 2×2 double-blind, randomized, placebo-controlled, phase IIA clinical trial

Future Trends

- Large population cohorts with comprehensive assessments of multiple clinical outcomes, intermediate phenotypes, biomarkers, lifestyle and environmental factors (e.g. UK Biobank)
- Large-scale whole-genome sequencing (e.g. UK 100,000K Project), enabling comprehensive assessment of rare variants
- Linking up genetic data with clinical databases and other data domains
- Greater integration of genetic data with functional annotation (e.g. eQTL, Roadmap Epigenome, ENCODE)
- Greater use of functional assays for evaluating the consequences of genetic mutations

Acknowledgments

Innovation and Technology Commission

Research Grants Council
of Hong Kong, China
中國香港
研究資助局

Suen Chi-Sun Foundation

Centre for Genomic Sciences
The University of Hong Kong