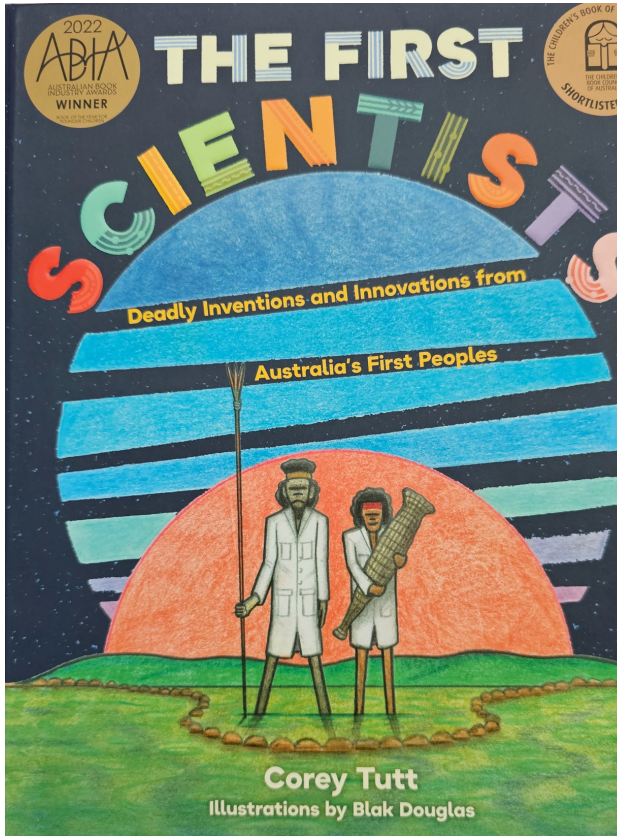
# UQ Genetics and Genomics Winter School 2026

## Systems Genomics and Pharmacogenomics
## Module 6

## Day 1 – QTL mapping

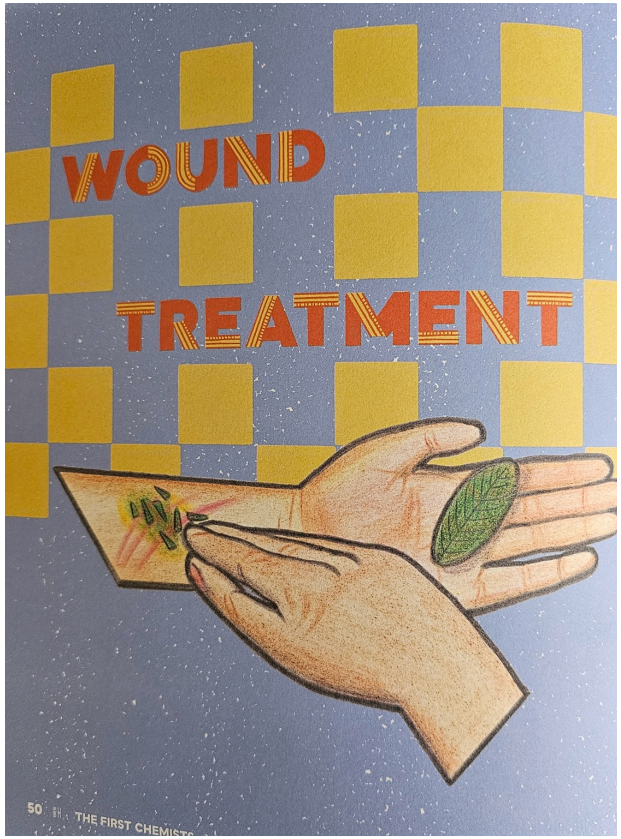
# Acknowledgement of Country




I acknowledge the Traditional Custodians of the lands on which we meet today.

I pay my respects to Elders past and present, who continue cultural and spiritual connections to Country.

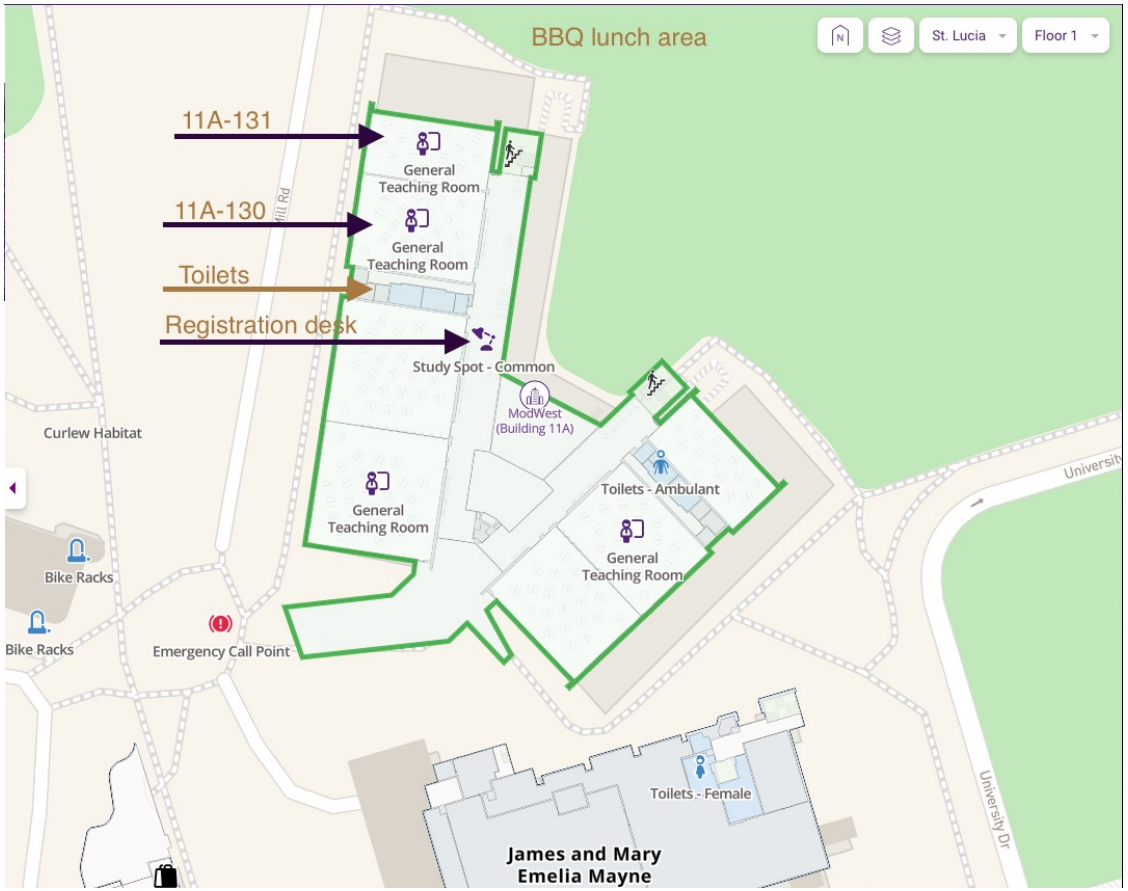
I recognise their valuable contributions to Australian and global society.




Celebrating the histories, cultures, and achievements of Aboriginal and Torres Strait Islander peoples.
Visit the UQ Art Museum to see some incredible Aboriginal and Torres Strait Islander artworks

# General Information

- We are currently located in Building 11A MODWEST
  - Bathrooms
  - Vending machines
- Food court and other bathrooms are located in Building 63 or Building 21B
- If you are experiencing cold/flu symptoms or have had COVID in the last 7 days please ensure you are wearing a mask for the duration of the module



# Learning materials

Instructions to access WiFi/desktop/server:

https://cnsgenomics.com/data/teaching/GNGWS26/module0/

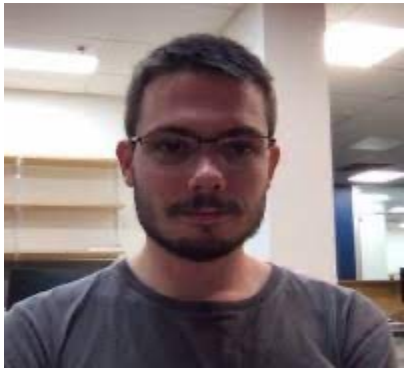
The winter school server is available until **24th July 2026** (2 weeks after the course)

Slides and practical notes for this module:

https://cnsgenomics.com/data/teaching/GNGWS26/module5/

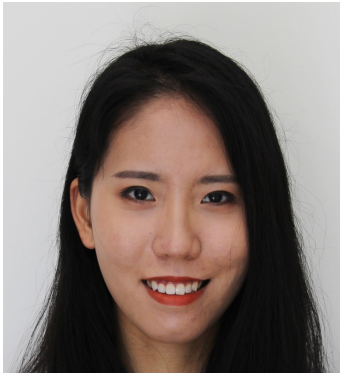
# Tutors

Lauren Barker



Solal Chauquet

Sonia Shah



Clara Jiang

Zoe Hunter

# Why do we do GWAS?

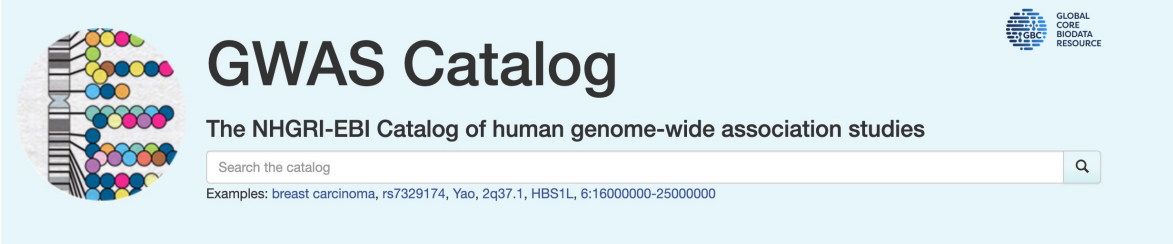
GWAS identifies variants in the genome that are associated with disease

Next steps:

- Understand how variants in this region impact disease
- Identifying genes and biological pathways involved in disease
- Identifying which cell types and tissues are most relevant to disease
- Identifying new avenues for drug development
- Risk stratification (polygenic risk scores)
- Understanding modifiable causal risk factors (genetic epidemiology), which can inform intervention

**GOAL: Progress from genetic maps to mechanism to medicine**

# We’re great at generating genetic maps




a freely accessible curated collection of all human genome-wide association studies

**As of 2026-06-01**

7714 publications

1,142,122 top SNP associations

176,855 full summary statistics available

# BUT…moving from genetic maps to mechanism and medicine is challenging

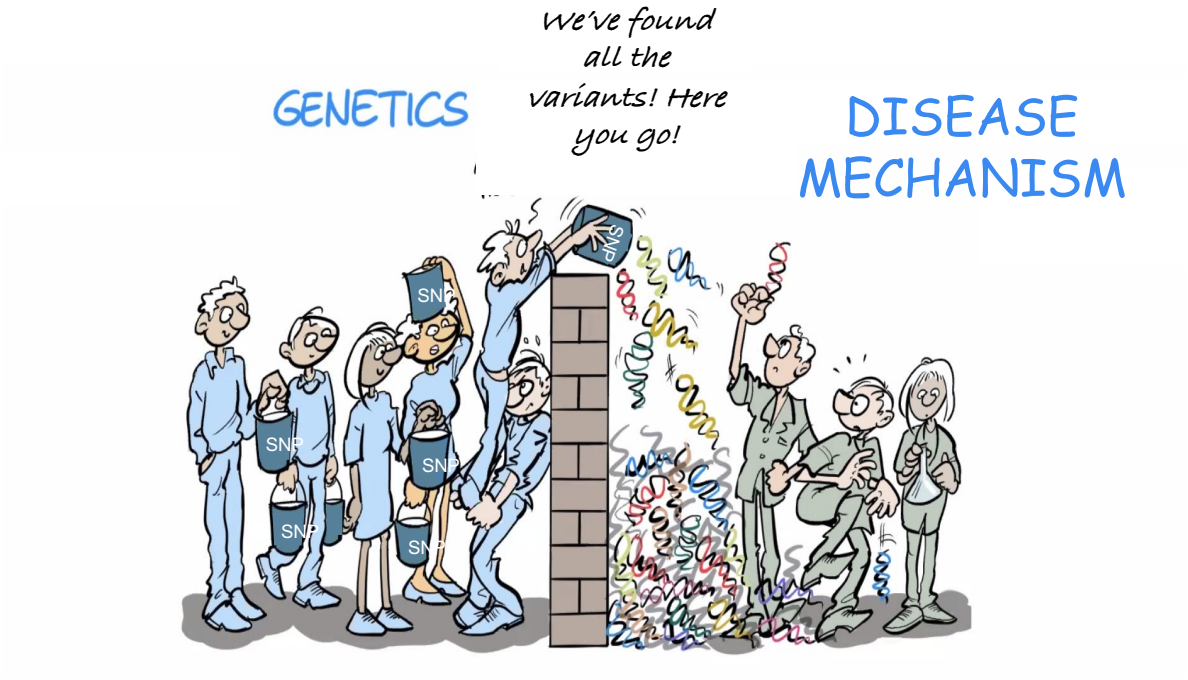
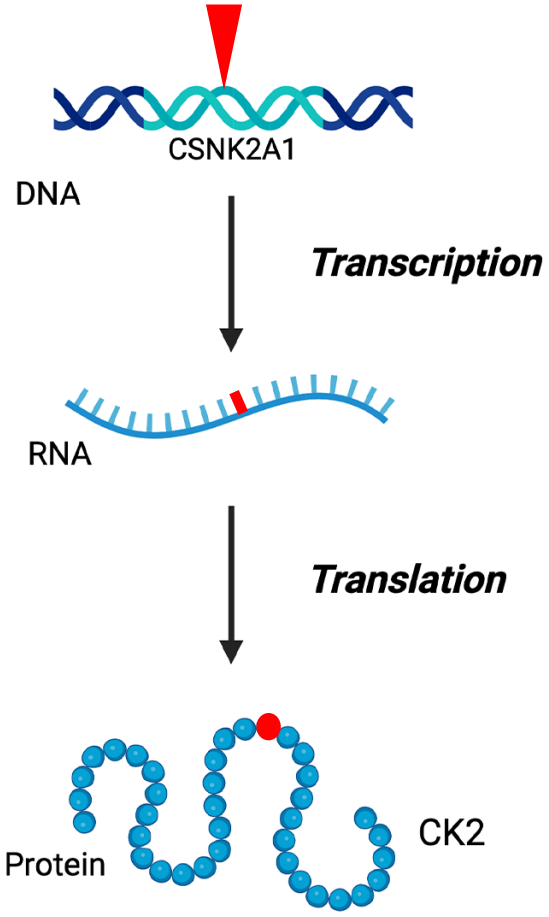
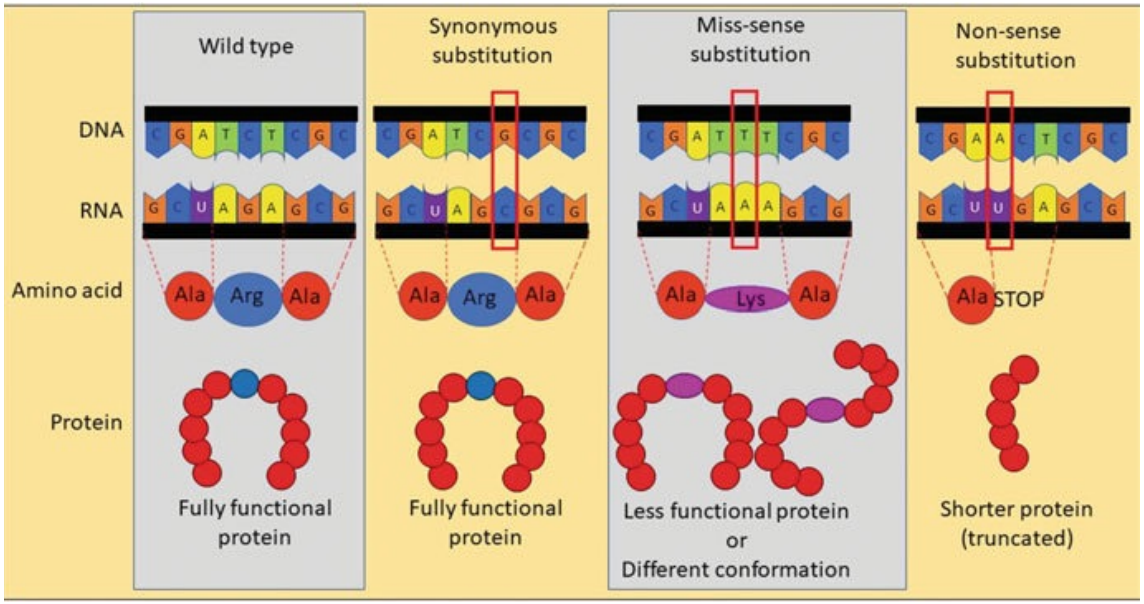



IMAGE CREDIT: BRAINSCAPES

- What are the causal variants?
- What are the causal genes?
- What are the relevant cell types and tissues?
- How does the variant and gene affect the disease?

# SNP to mechanism – protein-coding variants (low hanging fruit)



# SNP to mechanism – protein-coding variants

- *PCSK9* c.426C>G (p.Tyr142X) associated with lower risk of coronary artery disease
- Premature termination codon and truncation of the encoded protein or absence of the protein due to nonsense mediated decay i.e. Loss-of-function (LOF) variant
- LOF carriers have substantially lower plasma LDL-C

- Predicting functional consequence of coding variants – SIFT, PolyPhen

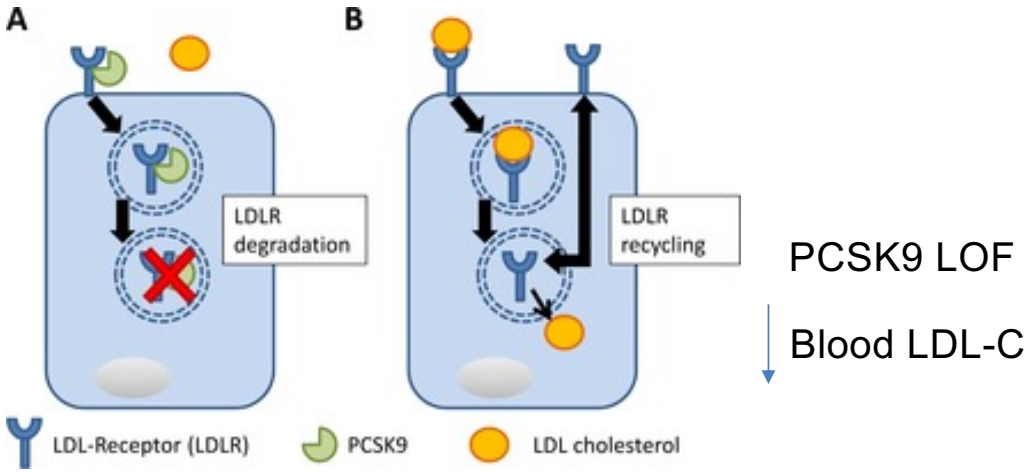



Image source: https://www.biovendor.com/

# REVEL

Am J Hum Genet. 2016 Sep 22;99(4):877–885. doi: 10.1016/j.ajhg.2016.08.016

## REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants

Nilah M Ioannidis [1,2,34], Joseph H Rothstein [2,3,4,34], Vikas Pejaver [5], Sumit Middha [6], Shannon K McDonnell [7], Saurabh Baheti [7], Anthony Musolf [8], Qing Li [8], Emily Holzinger [8], Danielle Karyadi [9], Lisa A Cannon-Albright [10], Craig C Teerlink [10], Janet L Stanford [11], William B Isaacs [12], Jianfeng Xu [13], Kathleen A Cooney [10,14], Ethan M Lange [15], Johanna Schleutker [16,17], John D Carpten [18], Isaac J Powell [19], Olivier Cussenot [20], Geraldine Cancel-Tassin [20], Graham G Giles [21,22], Robert J MacInnis [21,22], Christiane Maier [23,24], Chih-Lin Hsieh [25], Fredrik Wiklund [26], William J Catalona [27], William D Foulkes [28], Diptasri Mandal [29], Rosalind A Eeles [30], Zsofia Kote-Jarai [30], Carlos D Bustamante [1,31], Daniel J Schaid [7], Trevor Hastie [31,32], Elaine A Ostrander [9], Joan E Bailey-Wilson [8], Predrag Radivojac [5], Stephen N Thibodeau [33], Alice S Whittemore [2,31], Weiva Sieh [2,3,4,*]

Author information · Article notes · Copyright and License information

PMCID: PMC5065685 PMID: 27666373

Ensembl learning method that combines scores from 13 coding variant functional prediction tools, including PolyPhen and SIFT

# SNP to mechanism – protein-coding variants


- unsupervised protein language modelling to learn amino acid distributions conditioned on sequence context
- incorporates evolutionary conservation as well as structural context
- incorporates population frequency data, thereby avoiding bias from human-curated annotations.
- Provide a database of predictions for all possible single amino acid substitutions in the human proteome.
- Classifies 32% of all missense variants as likely pathogenic and 57% as likely benign

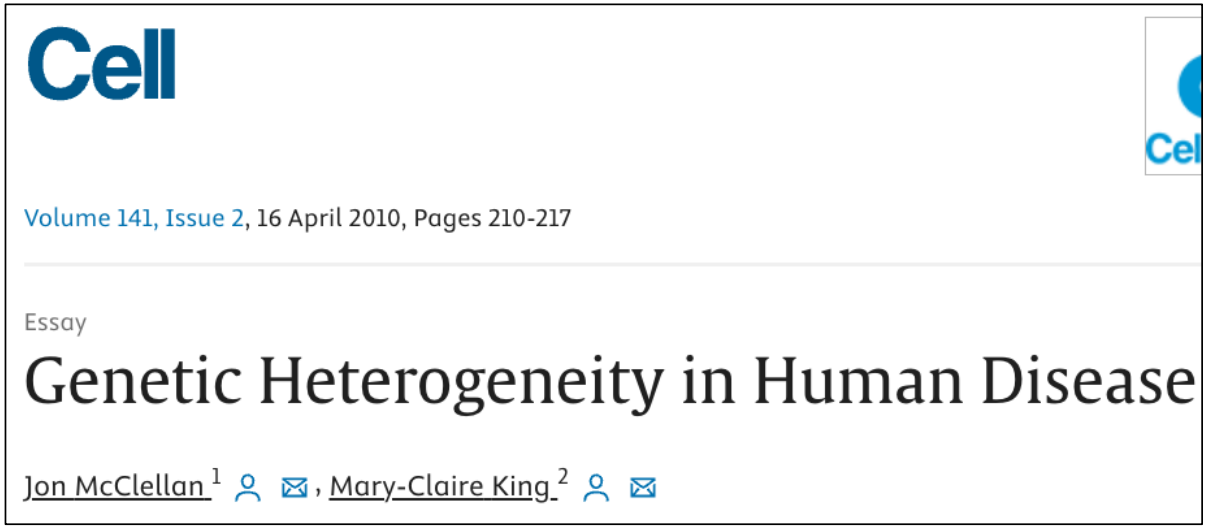
**Results** Download

**PCSK9** (PCSK9_HUMAN, Q8NBP7, ENST000003021

| Protein variant | SNV | Pathogenicity | Class | REVEL |
|---|---|---|---|---|
| p.Leu324Trp | | 0.129 | likely_benign | |
| p.Leu324Tyr | | 0.202 | likely_benign | |
| p.Tyr325Ala | | 0.792 | likely_pathogenic | |
| p.Tyr325Cys | y | 0.443 | ambiguous | l_pathogenic (0.716) |
| p.Tyr325Asp | y | 0.767 | likely_pathogenic | |
| p.Tyr325Glu | | 0.93 | likely_pathogenic | |
| p.Tyr325Phe | y | 0.159 | likely_benign | |
| p.Tyr325Gly | | 0.705 | likely_pathogenic | |
| p.Tyr325His | y | 0.4 | ambiguous | l_pathogenic (0.755) |
| p.Tyr325Ile | | 0.81 | likely_pathogenic | |
| p.Tyr325Lys | | 0.903 | likely_pathogenic | |
| p.Tyr325Leu | | 0.661 | likely_pathogenic | |
| p.Tyr325Met | | 0.886 | likely_pathogenic | |
| p.Tyr325Asn | y | 0.494 | ambiguous | l_pathogenic (0.752) |

Are any of the GWAS-significant SNPs or SNPs in LD with them predicted to impact protein function?

<2% of the human genome is protein-coding
>90% GWAS significant SNPs are in non-coding regions
GWAS SNPs have small effect sizes



Cell

Volume 141, Issue 2, 16 April 2010, Pages 210-217

Essay

## Genetic Heterogeneity in Human Disease

Jon McClellan [1], Mary-Claire King [2]

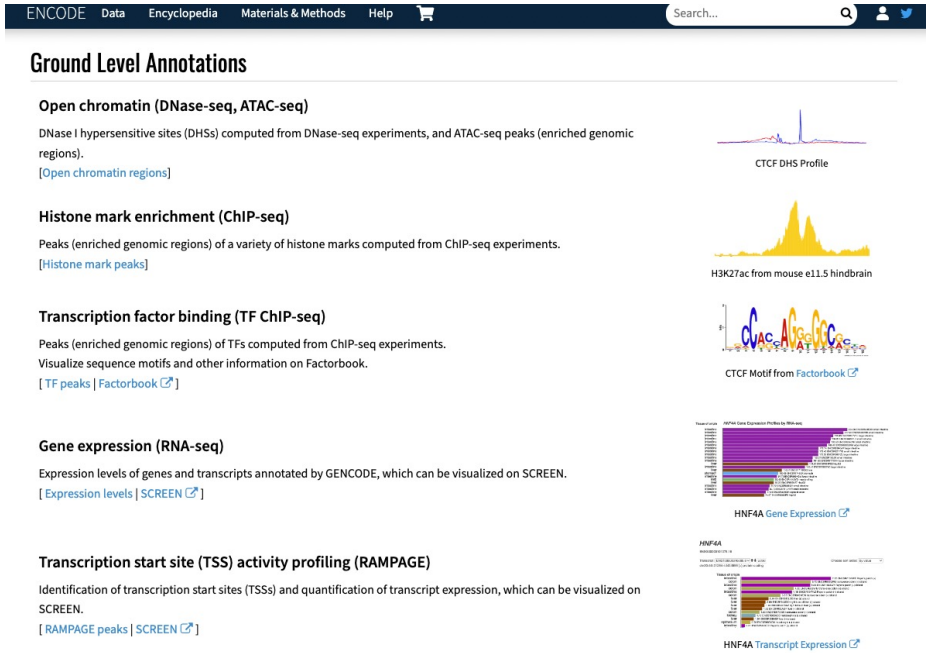
*"To date, GWAS have published hundreds of common variants…*
*However, the vast majority of such variants*
***have no biological relevance to disease or clinical utility…"***

# 98% of the genome is not junk!


Image source www.biocomicals.com

## The Encyclopedia of DNA Elements (ENCODE)

Goal: Build a comprehensive list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.



# DNase I hypersensitive sites

- Regions that are sensitive to cleavage by the DNase I enzyme
- When genes are transcribed, chromatin structure loses its condensed format so that TFs and regulatory proteins necessary for transcription can bind to DNA.
- This also raises the availability of DNA to degradation by enzymes, such as DNase I.
- Open chromatin therefore associated with regulatory elements (promoters, enhancers, insulators, silencers)
- Maps to regulatory elements
- Sites are highly cell type- and state-selective

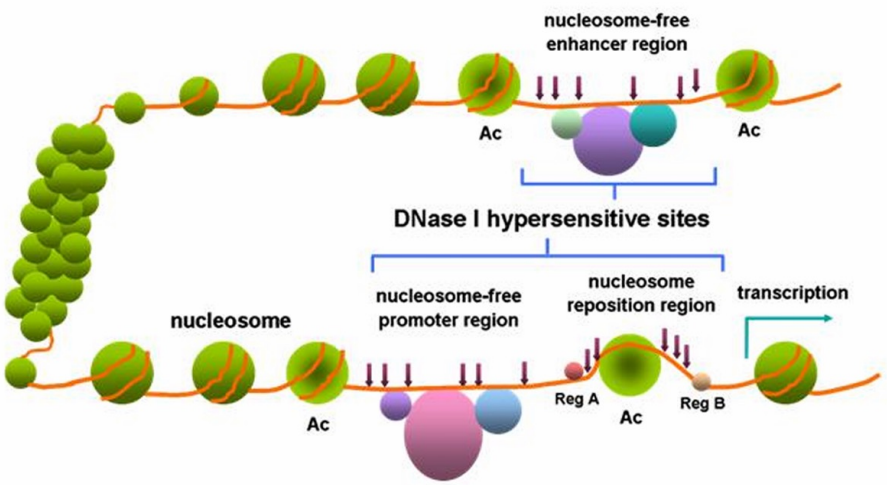



image source: Wikipedia

# ENCODE Data Snapshot

**Showing 20671 results**

List | Report | Download | Visualize | {;}

ASSAY ⟶ / ⟵ BIOSAMPLE

| | TF ChIP-seq | Histone ChIP-seq | DNase-seq | total RNA-seq | Mint-ChIP-seq | polyA plus RNA-seq | ATAC-seq | long read RNA-seq | microRNA-seq | snRNA-seq | snATAC-seq | intact Hi-C | eCLIP | DNAme array | WGBS | small RNA-seq | PRO-cap | ChIA-PET | RAMPAGE | RNA microarray | genotyping array | CAGE | microRNA counts |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **tissue** | 439 | 2175 | 1605 | 453 | | 393 | 225 | 344 | 305 | 390 | 292 | 110 | 2 | 121 | 162 | 67 | 141 | 1 | 104 | 2 | 7 | 17 | 101 |
| dorsolateral prefrontal cortex | 59 | 188 | 110 | 120 | | | | 13 | 43 | | | 2 | | | | | | | | | | | |
| heart | 6 | 81 | 41 | 21 | | 16 | 7 | 25 | 25 | 49 | 17 | 7 | | | 10 | 1 | | | 2 | | | | 8 |
| adrenal gland | 8 | 39 | 35 | 27 | | 10 | 8 | 26 | 23 | 59 | 15 | 3 | 2 | 5 | 5 | 2 | 3 | | 4 | | | | 2 |
| heart left ventricle | 16 | 84 | 14 | 9 | | 2 | 15 | 4 | 5 | 54 | 65 | 10 | | 3 | 4 | | 6 | | 2 | | | | 2 |
| liver | 42 | 91 | 29 | 3 | | 20 | 8 | 8 | 7 | 8 | 8 | | | 1 | 9 | 1 | 5 | | 2 | | | | 7 |
| **cell line** | 2831 | 773 | 583 | 735 | 54 | 210 | 123 | 61 | 39 | 2 | 32 | 161 | 250 | 87 | 18 | 110 | 27 | 108 | 29 | 67 | 73 | 50 | 8 |
| K562 | 749 | 19 | 59 | 17 | | 15 | 4 | 4 | 2 | | 3 | 8 | 145 | 3 | 1 | 7 | 7 | 9 | 1 | 9 | 2 | 9 | 1 |
| HepG2 | 814 | 15 | 3 | 6 | | 11 | 2 | 3 | 2 | | 3 | 3 | 105 | 3 | 2 | 3 | | 4 | | 6 | 2 | 6 | 1 |
| GM12878 | 188 | 15 | 4 | 5 | | 14 | 2 | 4 | 2 | | 3 | 68 | | 3 | 1 | 6 | | 6 | 1 | 7 | 2 | 6 | 1 |
| MCF-7 | 150 | 18 | 66 | 1 | | 4 | 1 | 1 | 2 | | 3 | 1 | | 2 | | 7 | | 4 | | 2 | 2 | 3 | 1 |
| HEK293 | 204 | 6 | | | | | | | | | | | | 2 | | | 1 | 2 | | 1 | 2 | | |
| **primary cell** | 74 | 506 | 1194 | 654 | 715 | 132 | 76 | 6 | 52 | 1 | 29 | 26 | | 38 | 8 | 24 | 25 | 60 | 16 | 57 | 37 | 30 | 1 |
| CD4-positive, alpha-beta T cell | 1 | 12 | 160 | 191 | 6 | 2 | 1 | | 1 | | 1 | 2 | | | | | | 2 | | | | | |
| T-cell | | 11 | 75 | 51 | 18 | | 7 | | | | | | | | 1 | | 7 | 4 | | | | | |
| stimulated activated CD4-positive, alpha-beta T cell | | | 133 | 6 | | | | | | | | | | | | | | | | | | | |
| naive thymus-derived CD4-positive, alpha-beta T cell | | 16 | 20 | 6 | 74 | 8 | 2 | | 2 | | 2 | 2 | | | | | | 2 | | | | | |
| CD14-positive monocyte | 2 | 21 | 39 | 12 | 41 | 7 | | | | | 2 | 2 | | | 1 | 1 | | | | | | 1 | |

# >90% GWAS significant SNPs are in non-coding regions

- SNP heritability $h^2$ partitioned by functional category across 11 common diseases (psychiatric, metabolic, immune and cardiovascular)
- **Coding variants:** explain 8% of SNP $h^2$
- **DHSs** from 217 cell types explain ~80% of SNP $h^2$

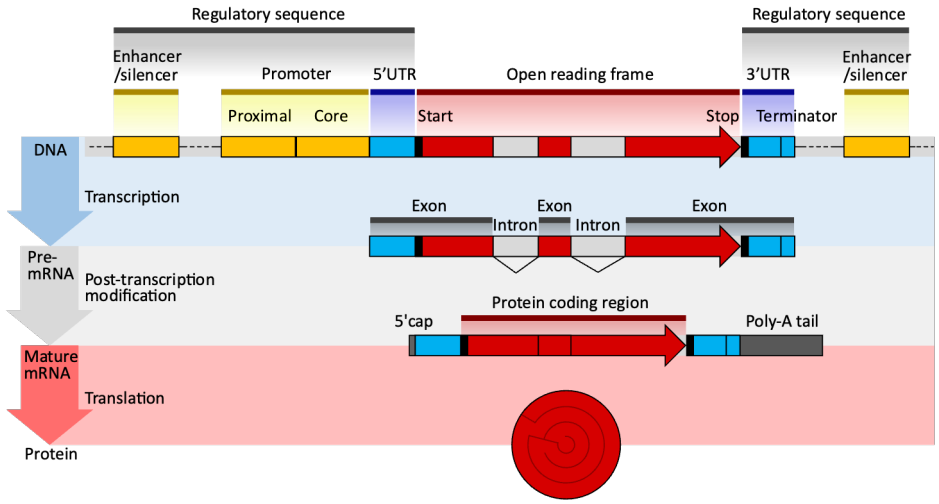
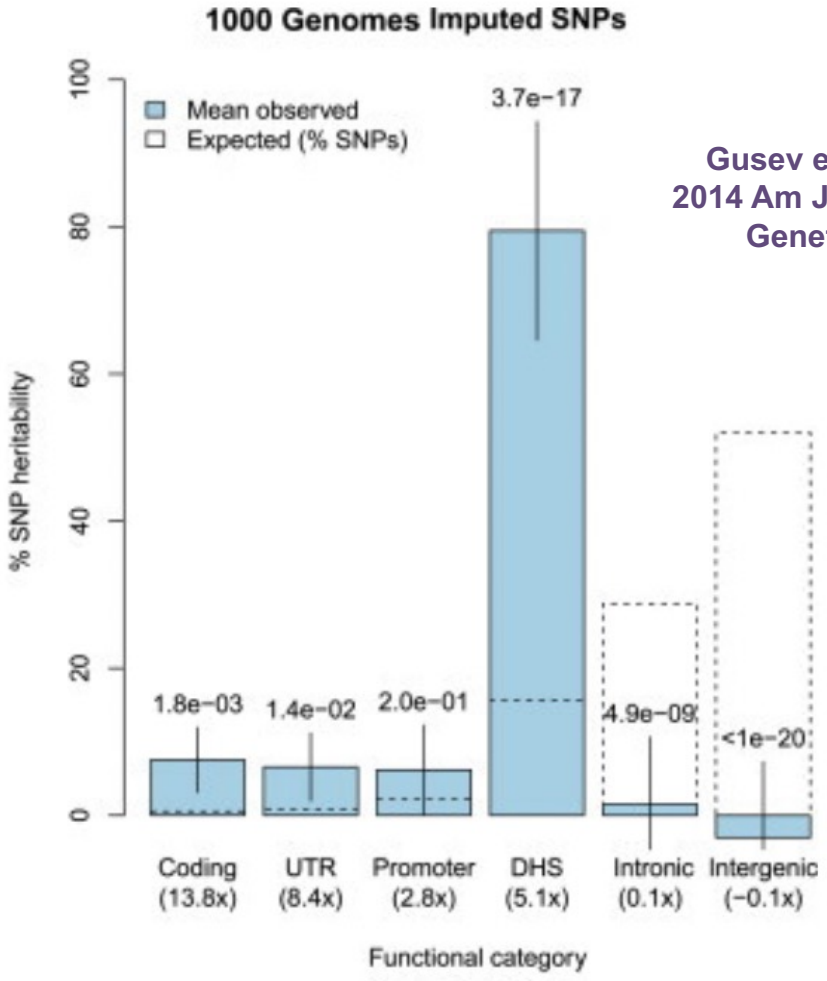
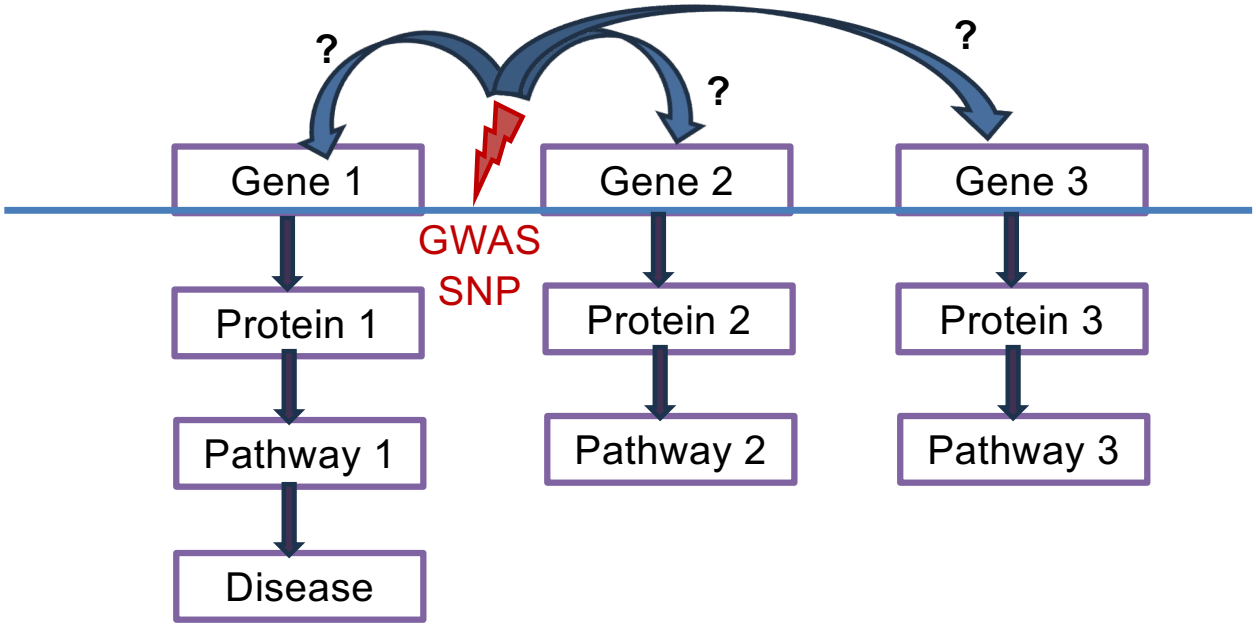


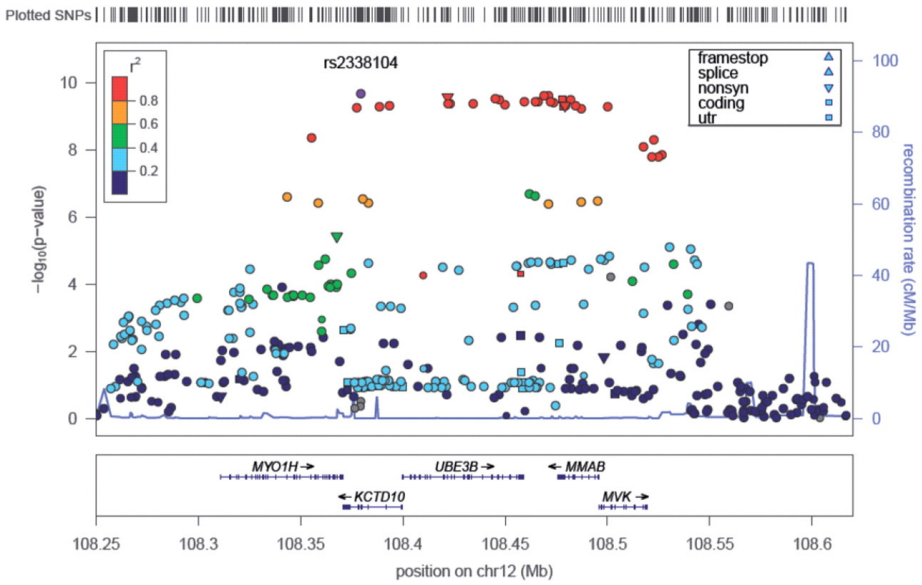

image source: Wikipedia




Gusev et al 2014 Am J Hum Genet

# Do these SNPs affect gene expression, and if so, which gene?



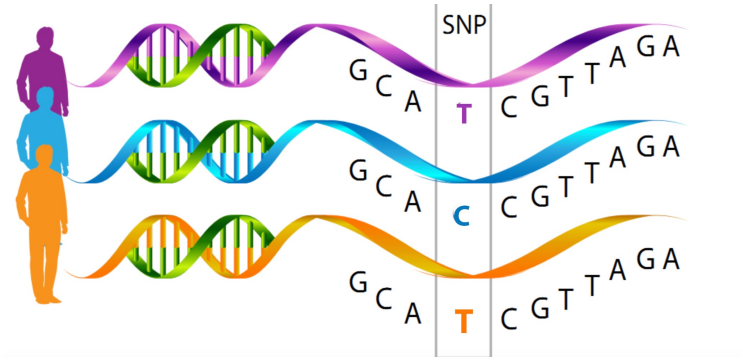




HDL cholesterol-associated region
(Kathiresan et al Nature Genetics 2009)

# Quantitative Trait Loci (QTL) Mapping

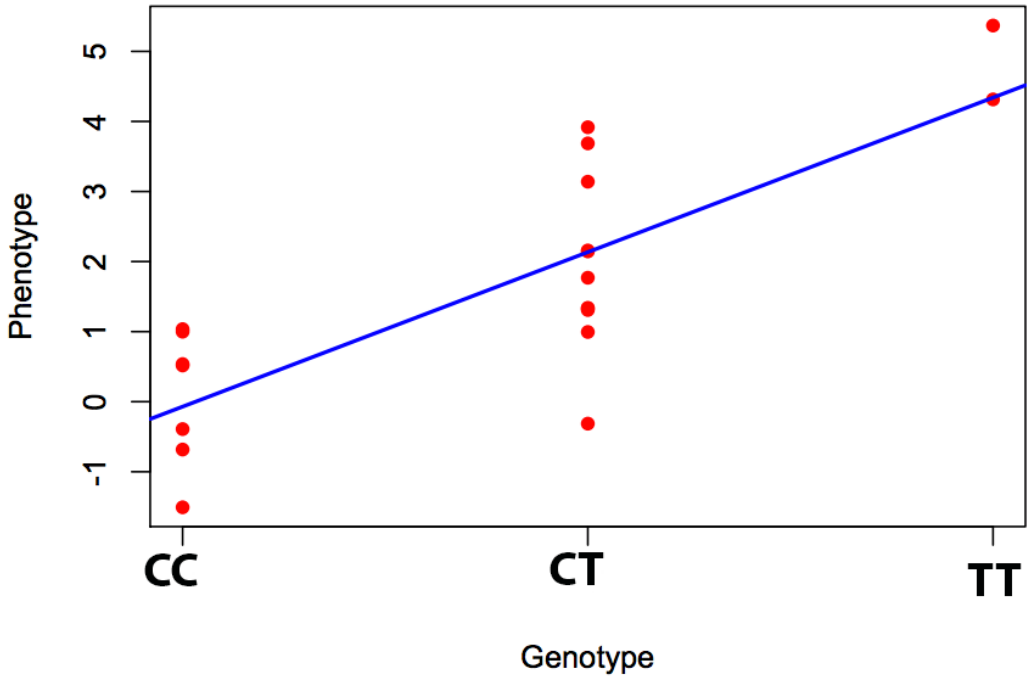
QTLs are genetic variants that are associated with gene expression (eQTLs) or protein expression (pQTLs) or splicing (spliceQTLs)

# Genetic association – quantitative trait



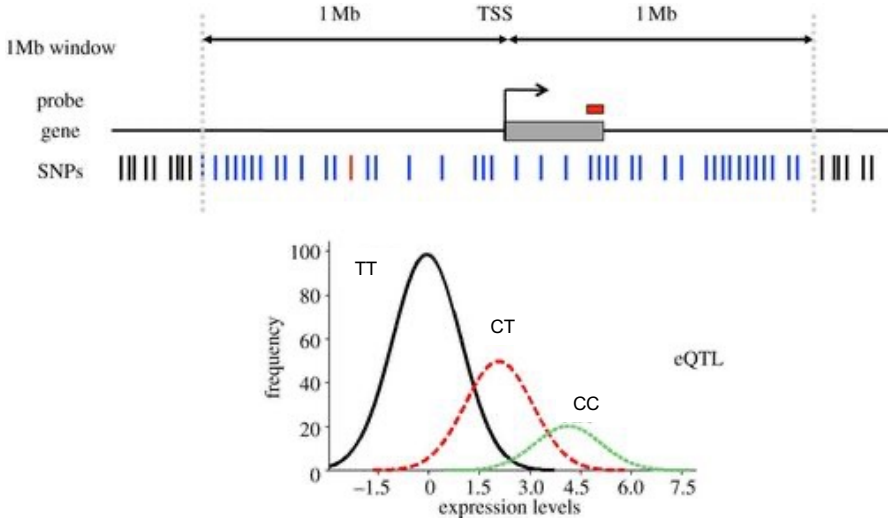

$$Y = b0 + b1X + e$$

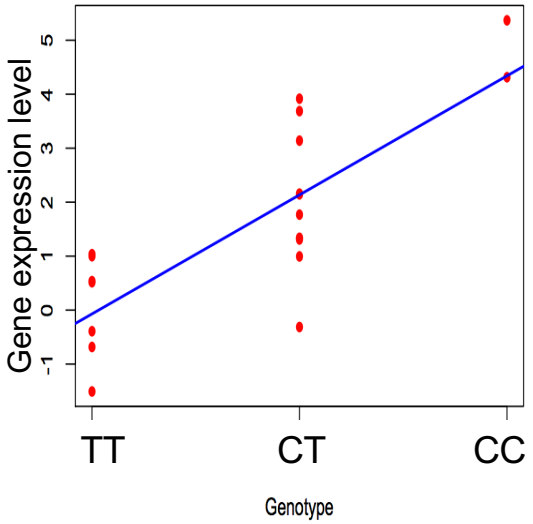
***Y*** phenotype e.g. LDL-cholesterol levels
***b0*** intercept
***b1*** effect of each copy of the risk allele on the mean phenotype
***e*** noise or the part of y that is not explained by the SNP (e.g., environmental effect)



# eQTL analysis

- Gene expression is a complex phenotype with both genetic and environmental determinants
- eQTL - variant that contributes to inter-individual variation in gene expression




Test for an association between genotype group and **mean** gene expression

$Y = b0 + b1X + e$

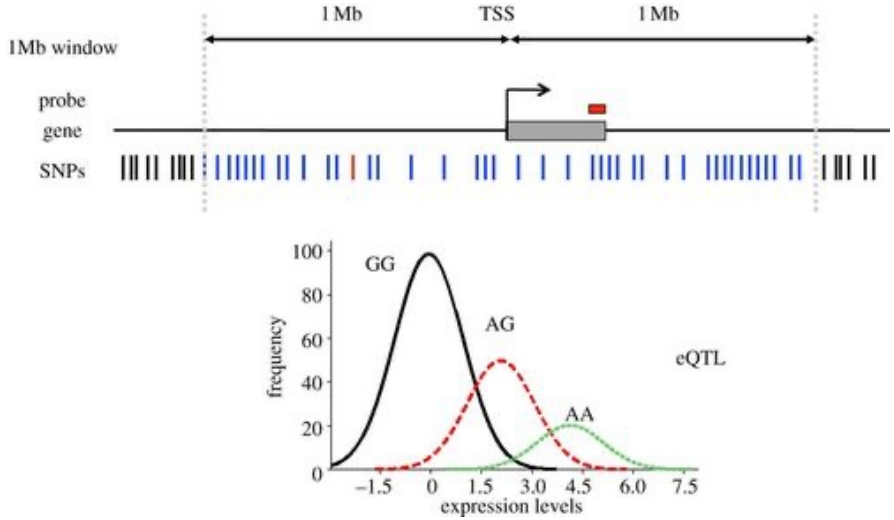
***Y*** gene expression
***b0*** intercept
***b1*** effect of risk allele on mean expression
***e*** noise or the part of y that is not explained by the SNP x (e.g. environmental, batch effect)

https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0362

# Cis vs trans QTLs

Cis-eQTL: SNP affects a gene located < 1Mb away

Trans-eQTL: SNP affects a gene located > 1Mb away (could be on a different chromosome)




https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0362

# Why a 1Mb cut-off for identifying cis vs trans QTLs?

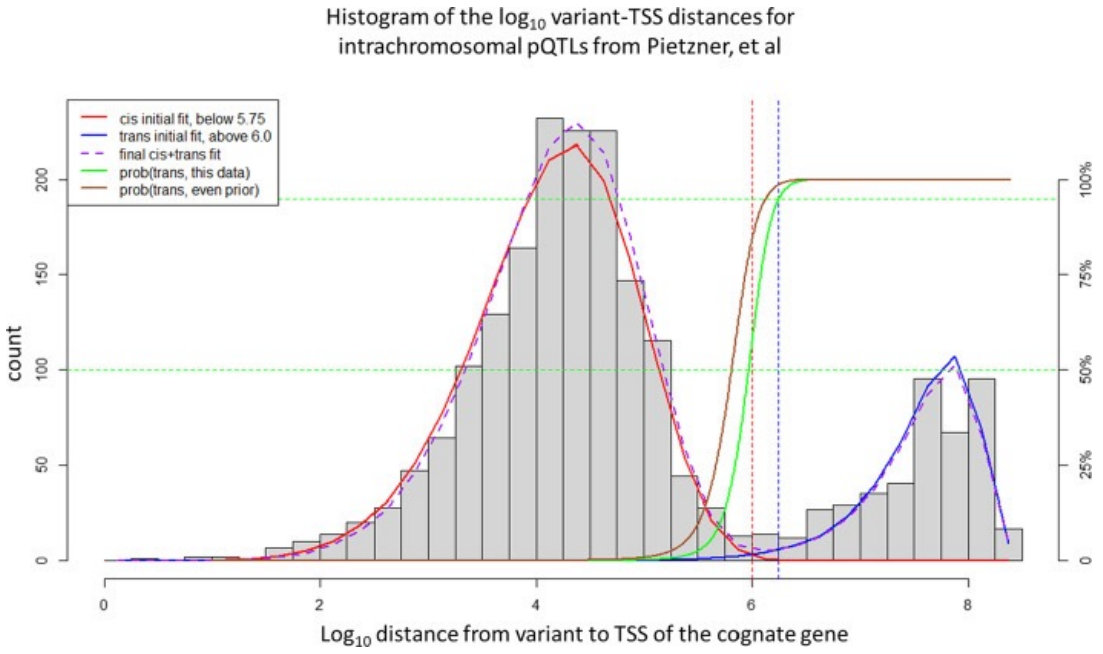
BMC Bioinformatics. 2022 May 8;23:169. doi: 10.1186/s12859-022-04706-x

**An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs**

Eric B Fauman [1,✉], Craig Hyde [2]

Author information · Article notes · Copyright and License information

PMCID: PMC9082853 PMID: 35527238




Looked at the distribution of variants to the target gene from a large protein QTL study

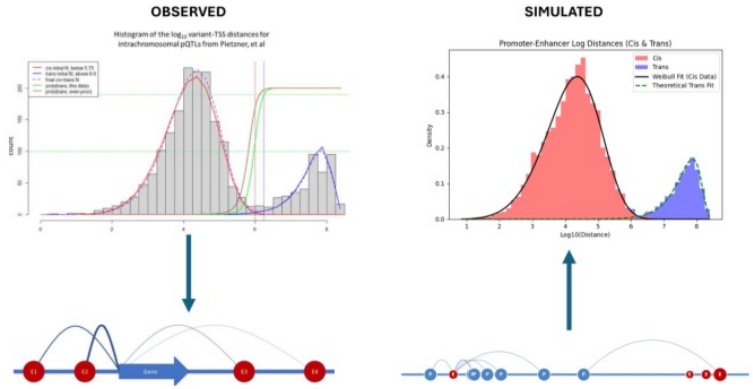
Distribution of intrachromosomal pQTL variant-TSS distances fell into a bimodal distribution

- First population is contained within the interval from 0 to 1 Mb
- Second distribution is similar to the observed distances for two randomly selected points on a chromosome - most chromosomes are over 100 Mb and two randomly selected intrachromosomal points are almost always (99%) more than 1 Mb apart.

From this finding I (Eric Fauman) typically offer the following advice:
*If you're trying to find the causal gene for a GWAS SNP and you've ventured more than 1 Mb from the lead SNP and you haven't found it, go back and look again.*

Search

Home My Network Jobs




## Science at the Speed of Thought: Reflections of a Computational Biologist in the AI Era

**Eric Fauman**
Executive Director, AI Strategy for Pharma R&D

May 10, 2026

While on the train from London to Cambridge last week I kept wondering:

> "What about the Hox gene cluster?"

This was a challenge put to me the day before when presented my work on "When GWAS meets Biochemistry" at Queen Mary University, at the invitation of my collaborator, Claudia Langenberg

The University Of Queensland
AUSTRALIA

The 2022 paper presented an explanation for the observed distribution.

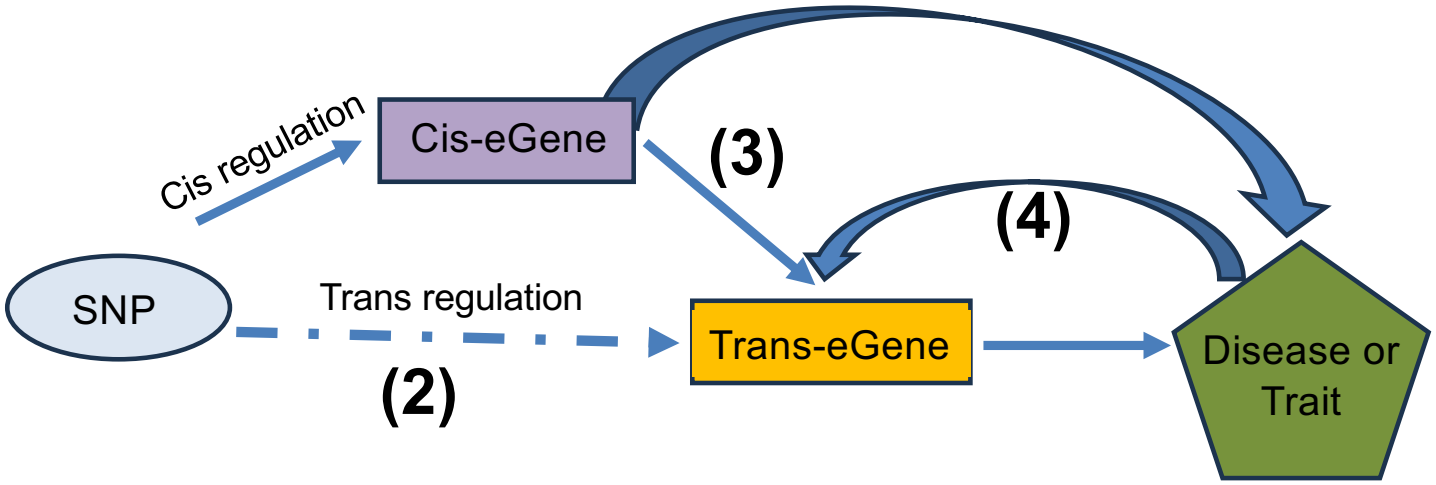
"Somewhere north of Finsbury Park I decided to put this to the test. I opened up my laptop and fed this explanation and my figure into Antigravity, Google's new agentic development platform."

"I asked Antigravity to cook up a simulation based on reasonable parameters regarding promoters, enhancers and their distribution across the genome."

"And all wrapped up by the time I reached Royston." (around 45min train ride!!)

Eric Fauman
Executive Director, AI Strategy for Pharma R&D at Pfizer

# Mechanisms by which a non-coding SNP affects a trait




**Yao et al 2017 AJHG**
Most trans-eQTLs are on the same chromosome - likely mediated by mechanism 3

**Mediation Mechanisms of eQTLs (Yao et al 2017 AJHG)**

(1) non-coding SNP affects expression of nearby gene (*cis-eQTL*) (e.g. SNP in promoter region)
(2) non-coding SNP affects remote gene expression directly (*trans-eQTL*) (e.g. SNP in enhancer region)
(3) *cis*-eGene mediation of the *trans*-eGene (e.g. if cis-eGene is a TF for trans-eGene);
(4) reverse causality (trait has feedback effect on gene expression).

# Performing eQTL mapping

# Performing genome-wide QTL mapping

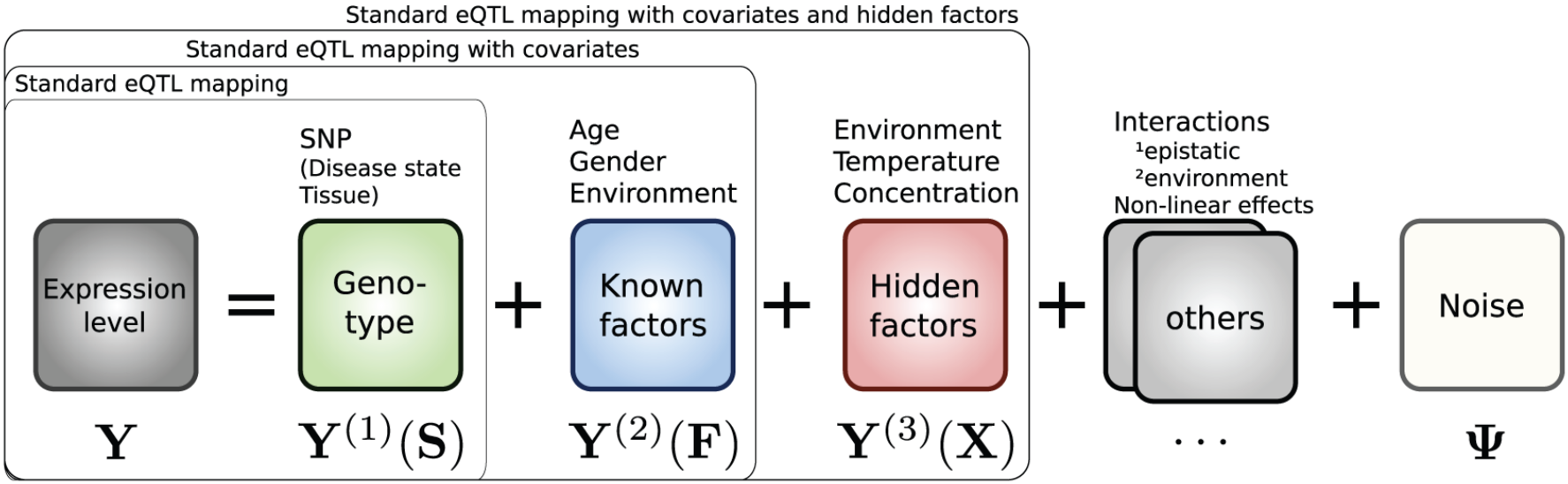
1. Need to measure transcriptome-wide gene expression (bulk tissue or single cell using arrays or RNAseq).

2. Need to generate genotype data in the same individuals.

3. Separate processing of genotype and expression data. (Expression data requires quality control, normalisation and correcting for batch factors and unmeasured confounders) – exact workflow is dependent on technological platform used.

# eQTL Mapping – Covariate adjustment

- Adjusting for measured and unmeasured (hidden) confounders can increase power to detect QTLs
- PEER is a method used to estimate unmeasured confounders




Stegle et al PLOS Comput Biol 2010

# Performing genome-wide QTL mapping

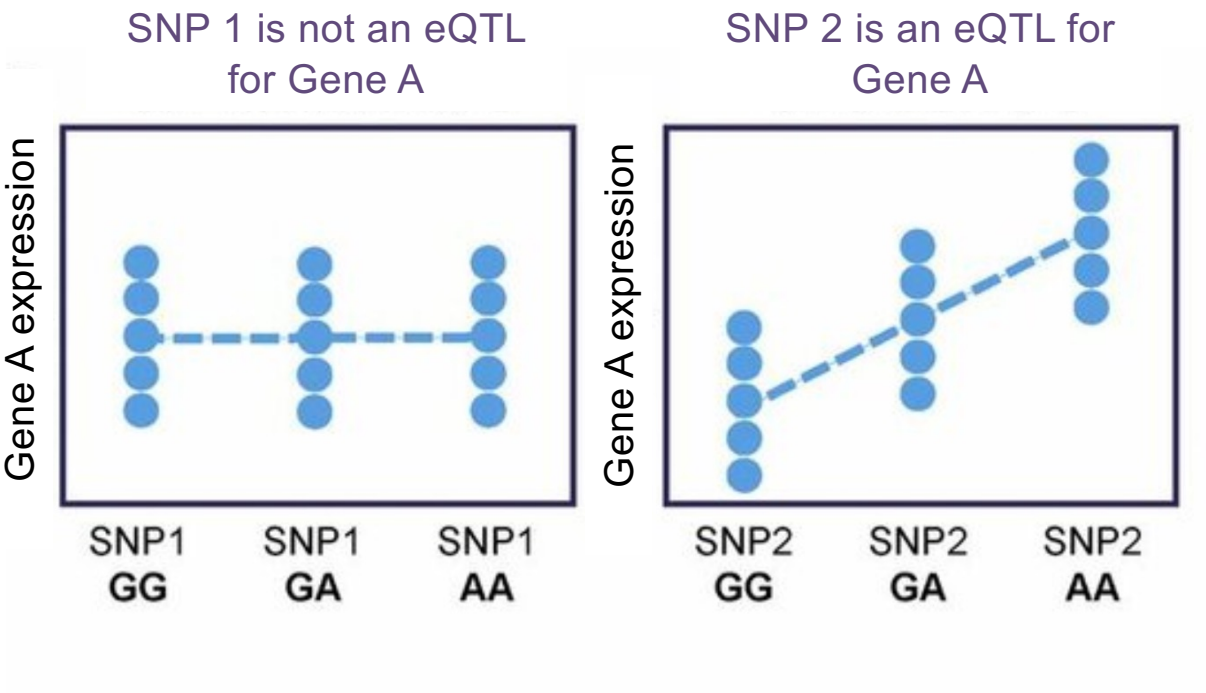
1. Need to measure transcriptome-wide gene expression (bulk tissue or single cell using arrays or RNAseq).

2. Need to generate genotype data in the same individuals.

3. Separate processing of genotype and expression data. (Expression data requires quality control, normalisation and correcting for batch factors and unmeasured confounders) – exact workflow is dependent on technological platform used.

4. Conduct a GWAS for every single gene, where gene expression levels are your phenotype i.e. run ~22,000 GWAS

# QTL mapping




- Test every single SNP with every single gene – A LOT of tests (high multiple testing and computational burden)

- Test only cis-SNPs (SNPs within 1Mb of a gene)

# Genome-wide QTL mapping software

**Plink** – most-commonly used software for GWAS - legacy approach

- default software to manipulate genetic files and run genetic association analysis

**matrix eQTL** (2012)
http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/

- Computationally efficient
- fast performance is achieved by special data pre-processing and using matrix operations
- Calculates FDR only for gene-SNP pairs that pass a user-defined significance

**fastQTL** (2016) https://hpc.nih.gov/apps/FastQTL.html

- faster processing time (16× faster than matrix QTL).
- Different permutation schemes available for multiple testing correction

**Table 1.** FastQTL and Matrix eQTL running times

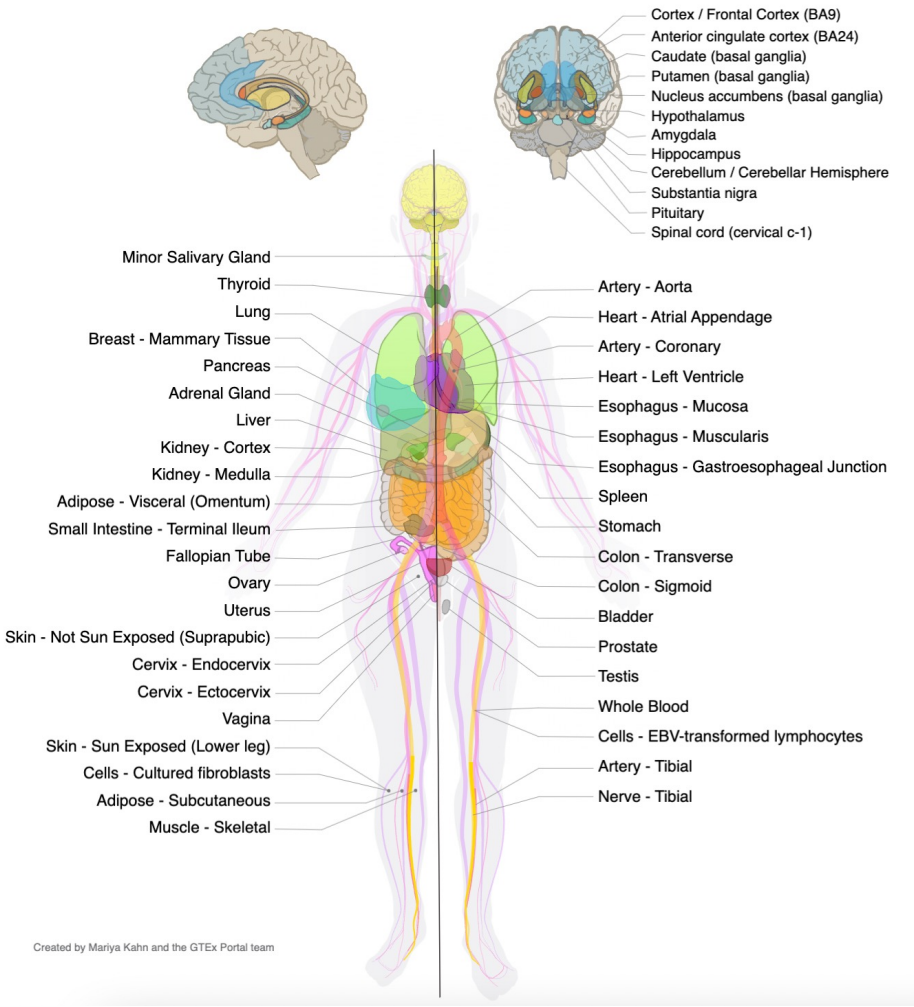
| Number of permutations | Matrix eQTL | FastQTL | | |
|---|---|---|---|---|
| | 1000 | 1000 | 500 | 100 |
| GTEx_AS | 337.4 | 19 | 10.8 | 3.5 |
| GTEx_AT | 330.3 | 21 | 10.8 | 3.6 |
| GTEx_HLV | 312.4 | 15 | 8 | 3 |
| GTEx_L | 364.9 | 25.6 | 13.1 | 4 |
| GTEx_MS | 335.9 | 23.6 | 12.6 | 3.9 |
| GTEx_NT | 343.8 | 18.4 | 9.5 | 3.4 |
| GTEx_SSEL | 349.7 | 20.7 | 10.8 | 3.6 |
| GTEx_T | 358.1 | 22.3 | 11.8 | 3.9 |
| GTEx_WB | 340.5 | 25.5 | 13.7 | 4.1 |
| ALL | 3073 | 191.1 | 101.1 | 33 |

Table 1 shows the running times in CPU hours to produce the results shown in Figure 2e; nine GTEx datasets (column 1) processed with 1000 Matrix eQTL permutations (column 2) and FastQTL with 1000 (column 3), 500 (column 4) and 100 permutations (column 5). Total running times for all nine datasets together are shown in the last row.

# eQTL data resources

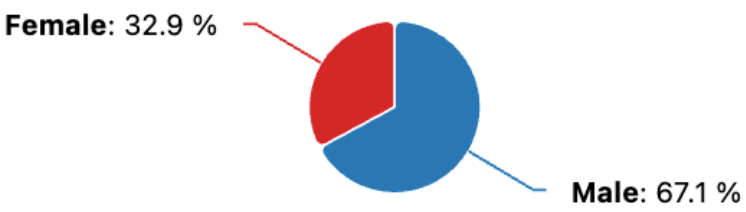
# The GTEx Project

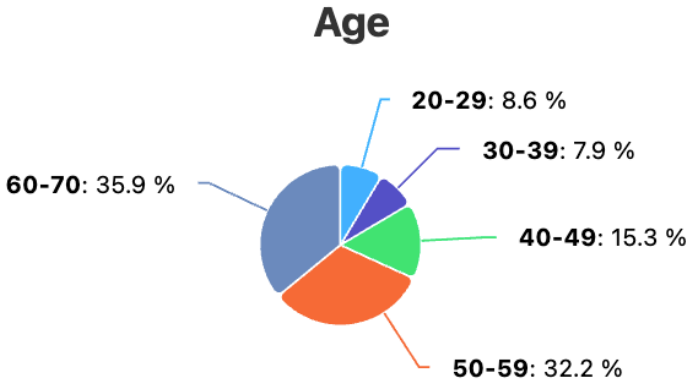
- Launched 2010
- Now at GTEx v10
- Catalogue of genetic effects on gene expression across a large number of human tissues
- eQTL data from 9436donors and 19,788 samples in 54 tissues
- Gene expression (RNA-seq) and genotype data (WGS data)

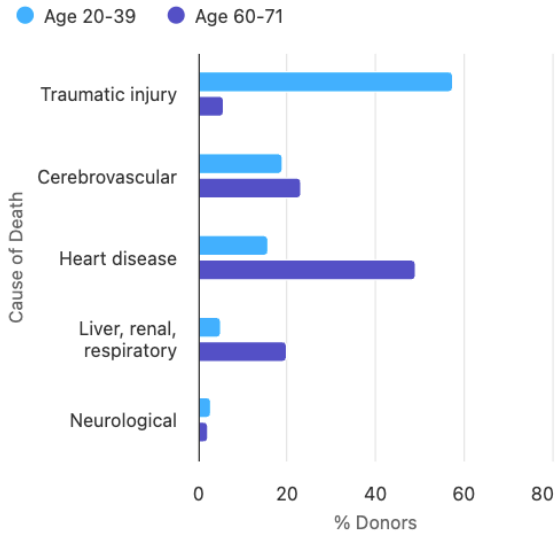



**The GTEx Consortium atlas of genetic regulatory effects across human tissues**
**Science 2020** DOI: 10.1126/science.aaz1776

# The GTEx Project

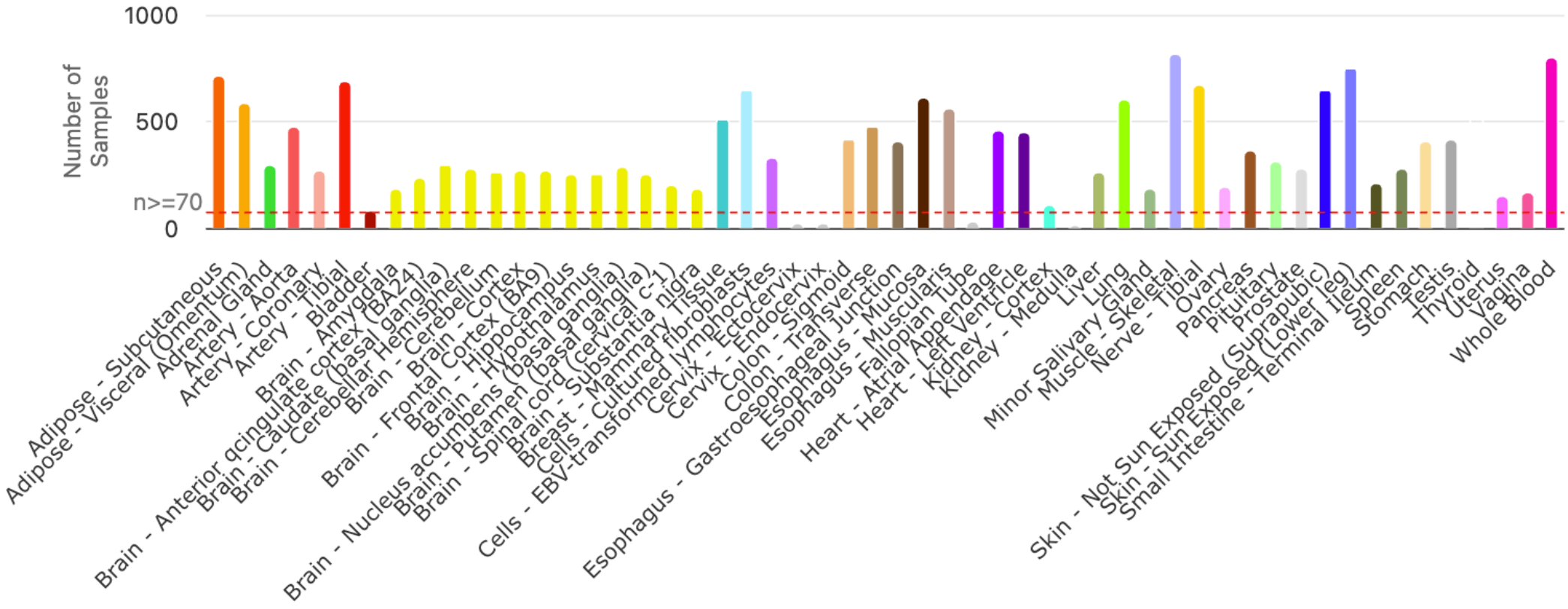



| Cause of Death | Age 20 - 39 | Age 60 - 71 |
|---|---|---|
| Traumatic injury | 57.4% | 5.6% |
| Cerebrovascular | 19.1% | 23.2% |
| Heart disease | 15.6% | 49.2% |
| Liver, renal, respiratory | 5.0% | 19.8% |
| Neurological | 2.8% | 2.2% |

# Tissue sample size



# The GTEx Project – some key findings

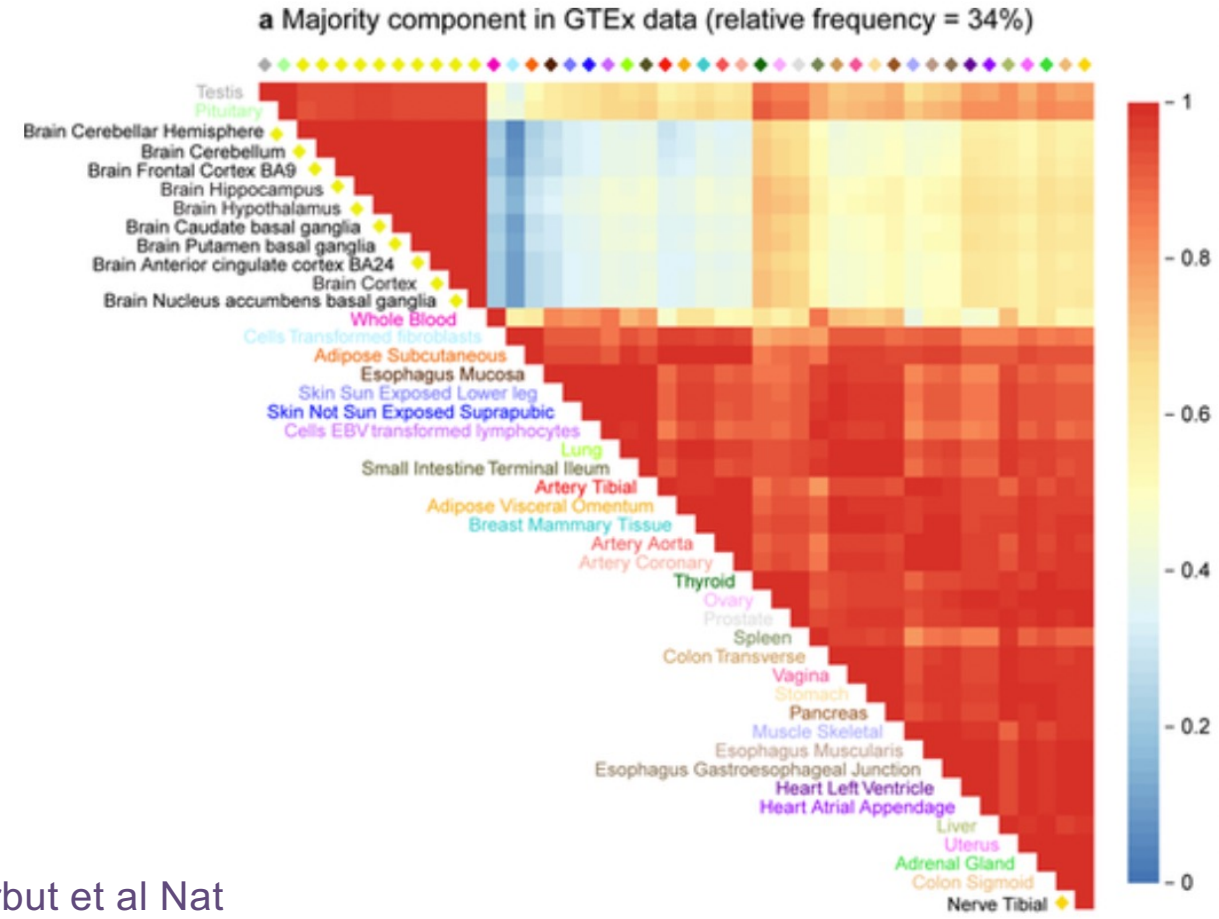
- cis-eQTLS identified (at 5% FDR per tissue) for 94.7% of all protein-coding and 67.3% of all lincRNA genes detected in at least one tissue

- most cis-eQTLs had small effect sizes: ~22% of cis-eQTLs had allelic Fold Change > 2-fold

- Genes lacking a cis-eQTL enriched for those not expressed in the tissues analysed, including genes involved in early development
  - Bulk RNAseq – may lose cell-specific effects

- Interchromosomal trans-eQTLs for 143 trans eGenes

# Many eQTLs are shared across tissues




Urbut et al Nat Genetics 2018

Correlation of eQTL effect estimates for 16,069 (genes expressed and have effect estimates in all 44 tissues)

(1) effects are **positively** correlated among all tissues;

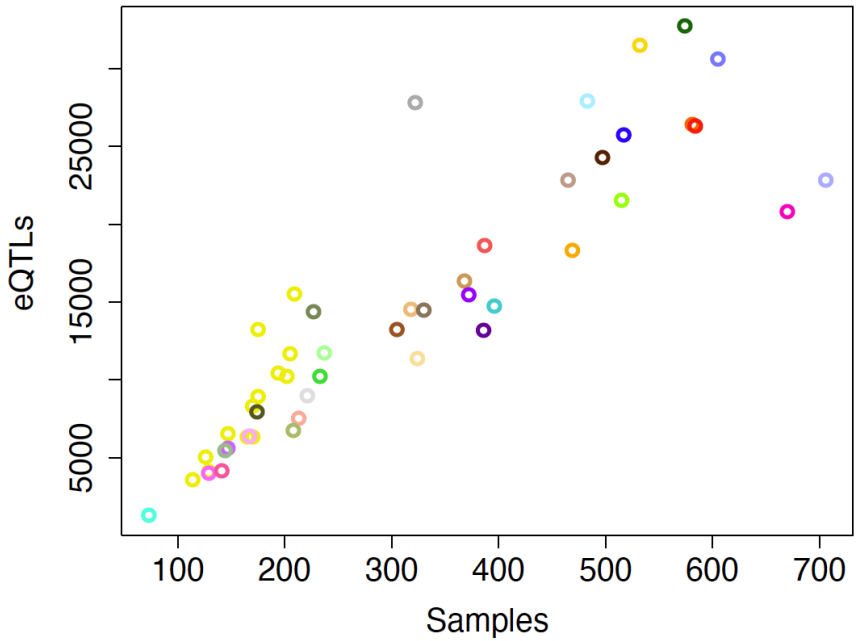
(2) the brain tissues—and, to a lesser extent, testis and pituitary—are particularly strongly correlated with one another, and less correlated with other tissues;

(3) effects in whole blood less are not as strongly correlated with other tissues

# True tissue-specific eQTL vs power to detect an eQTL

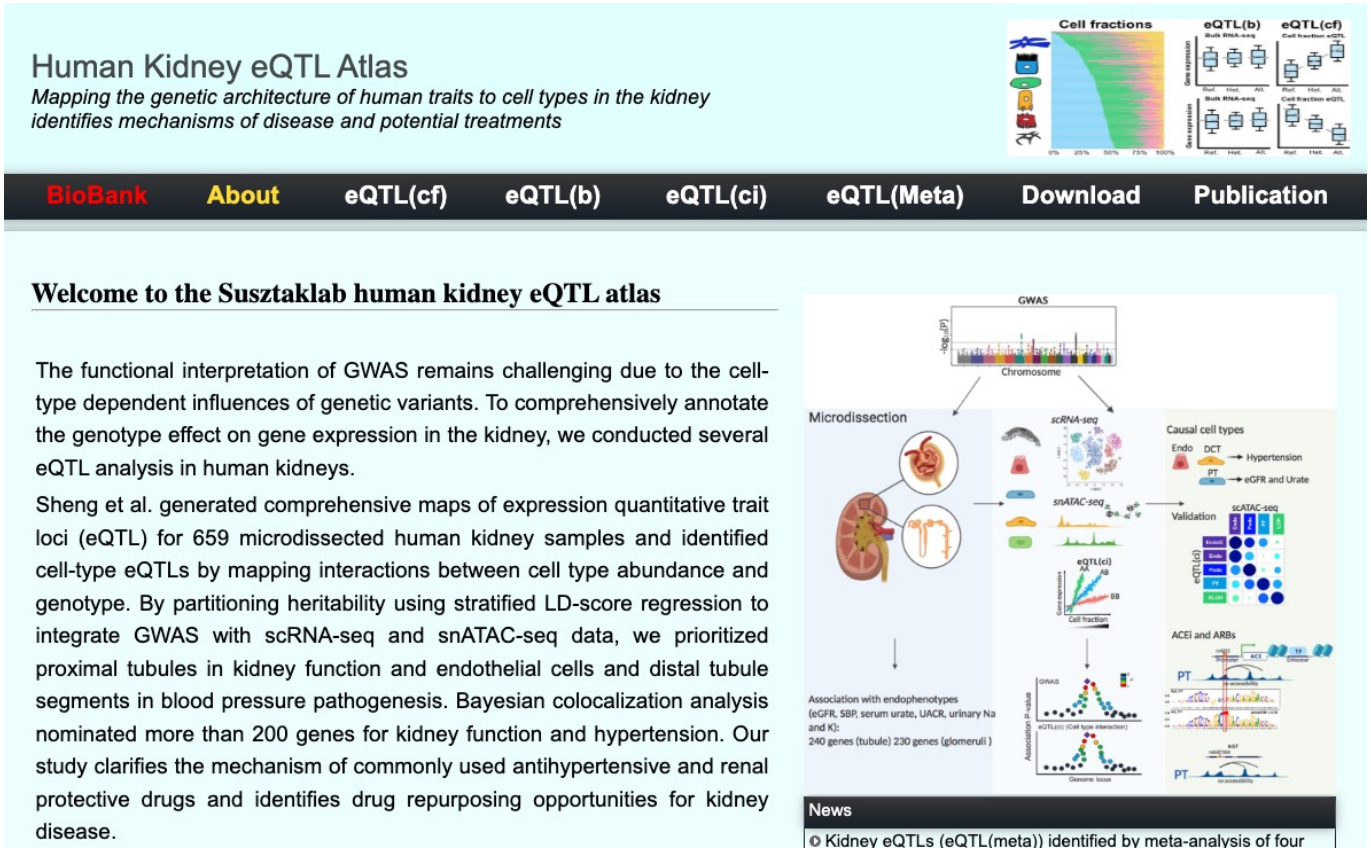
Urbut et al 2019 Nature Genetics

- eQTL discovery related to sample size
- Increased discovery in larger samples driven by increased power to detect small effects
- Papers that identify tissue-specific genes often use p-value threshold for detecting tissue-specific eqtls
- Power is an important consideration when trying to identify tissue-specific eQTLs
  - **Down-sampling analysis to check if tissue-specificity is real or simply a reflection of sample size**



# GTEx web browser

# Kidney eQTL dataset - 659 samples



# eQTL catalogue

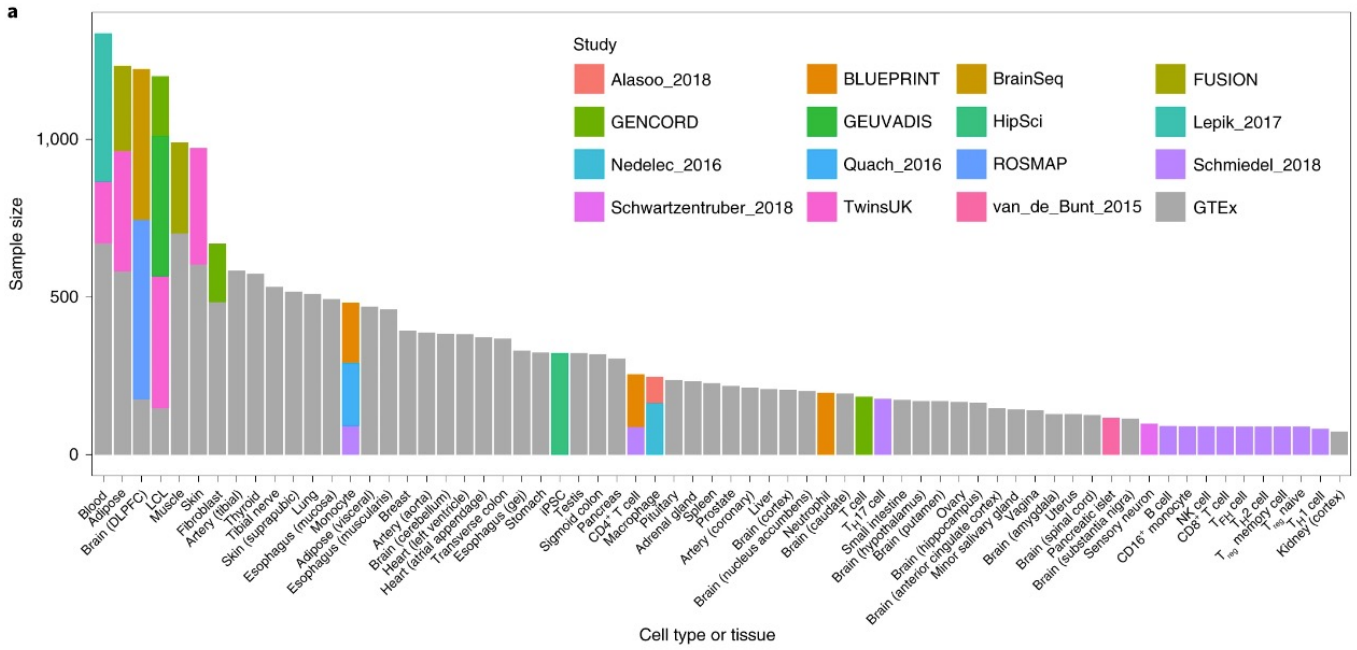
https://www.ebi.ac.uk/eqtl/
Provides uniformly processed cis-eQTLs and sQTLs from all available public studies on human.

Article | Open Access | Published: 06 September 2021

## A compendium of uniformly processed human gene expression and splicing quantitative trait loci

Nurlan Kerimov, James D. Hayhurst, Kateryna Peikova, Jonathan R. Manning, Peter Walter, Liis Kolberg, Marija Samovičа, Manoj Pandian Sakthivel, Ivan Kuzmin, Stephen J. Trevanion, Tony Burdett, Simon Jupp, Helen Parkinson, Irene Papatheodorou, Andrew D. Yates, Daniel R. Zerbino ✉ & Kaur Alasoo ✉

*Nature Genetics* **53**, 1290–1299 (2021) | Cite this article



# eQTL mapping could give clues to causal genes

Nat Genet. Author manuscript; available in PMC: 2018 Dec 28.

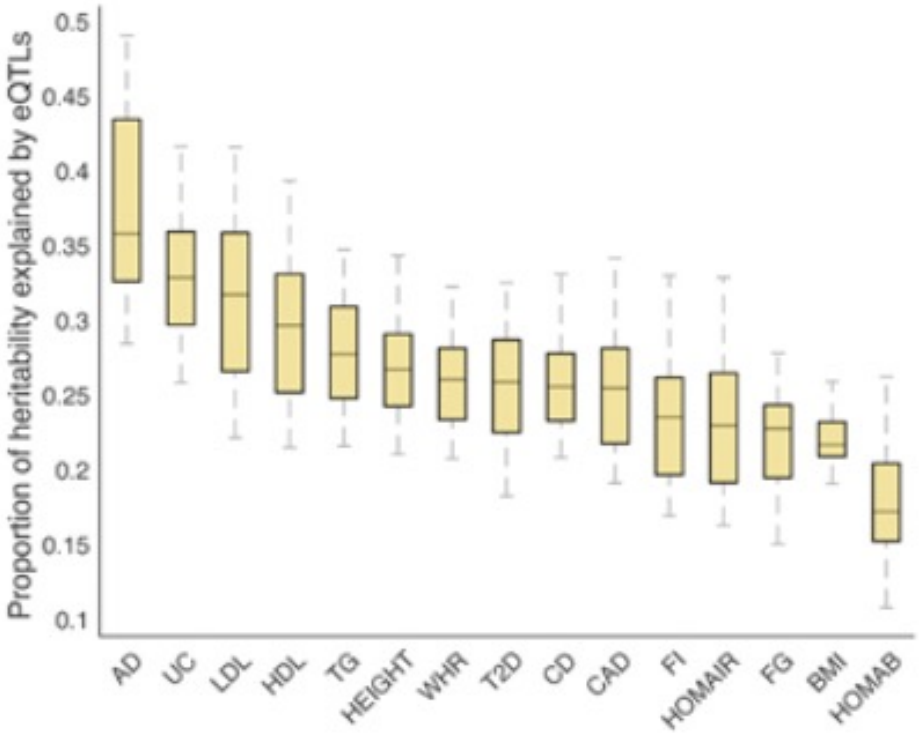
*Published in final edited form as:* Nat Genet. 2018 Jun 28;50(7):956–967. doi: 10.1038/s41588-018-0154-4

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

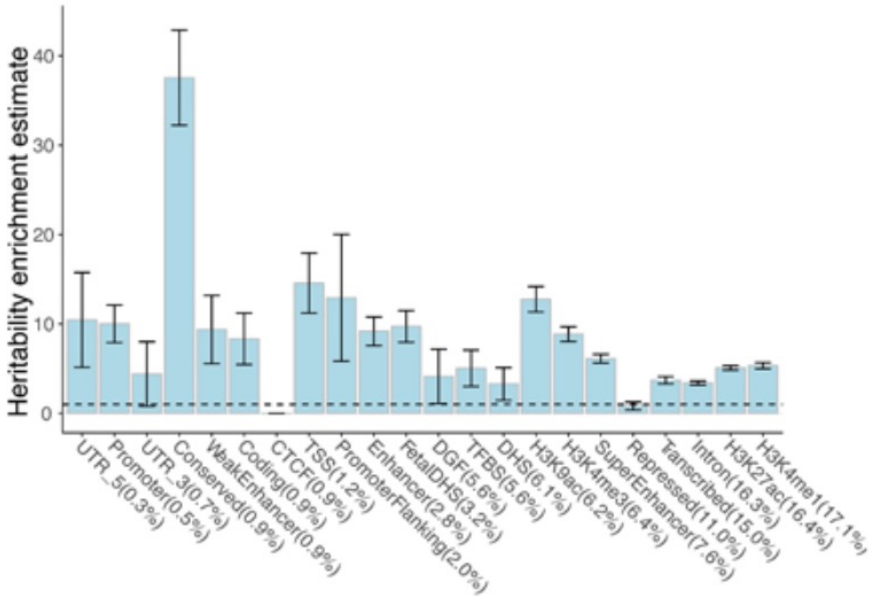
# Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation

Eric R Gamazon [1,2,§,*], Ayellet V Segrè [3,4,§,*], Martijn van de Bunt [5,6,§], Xiaoquan Wen [7], Hualin S Xi [8], Farhad

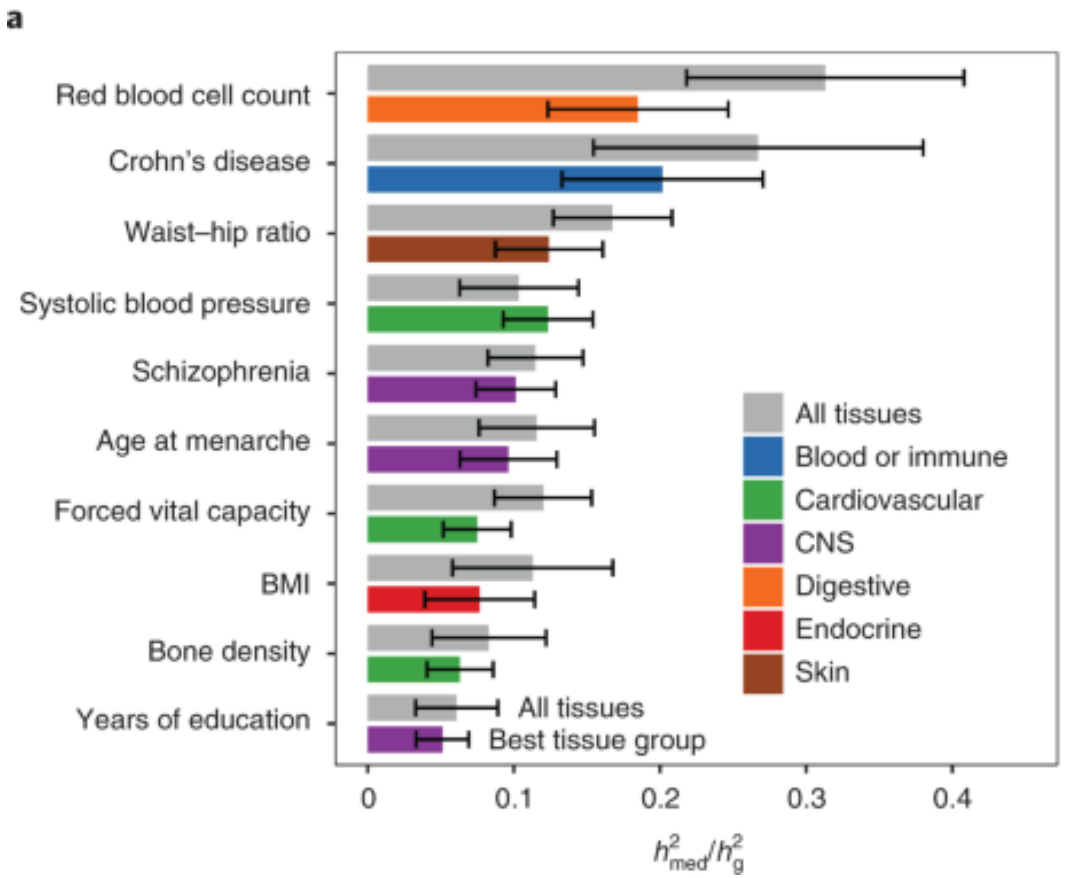
## Distribution of proportion of heritability of 15 traits explained by eQTLs in 44 tissues (GTEx data)

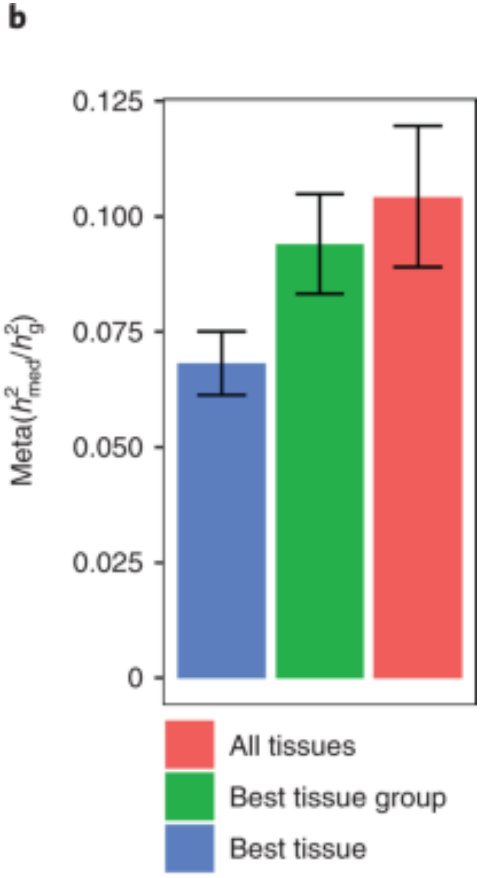



## Heritability enrichment estimate computed for subsets of eQTLs that fall in different genomic features



# *What proportion of complex trait heritability is **mediated** by the cis-eQTLs*




Potential under-estimation if assayed gene expression levels do not adequately capture expression levels in causal cell types/contexts.

Yao et al Nature Genetics 2020

# eQTLs can be context-dependent and may be missed in resources such as GTEx

# Context-specific eQTLs

The University Of Queensland
AUSTRALIA

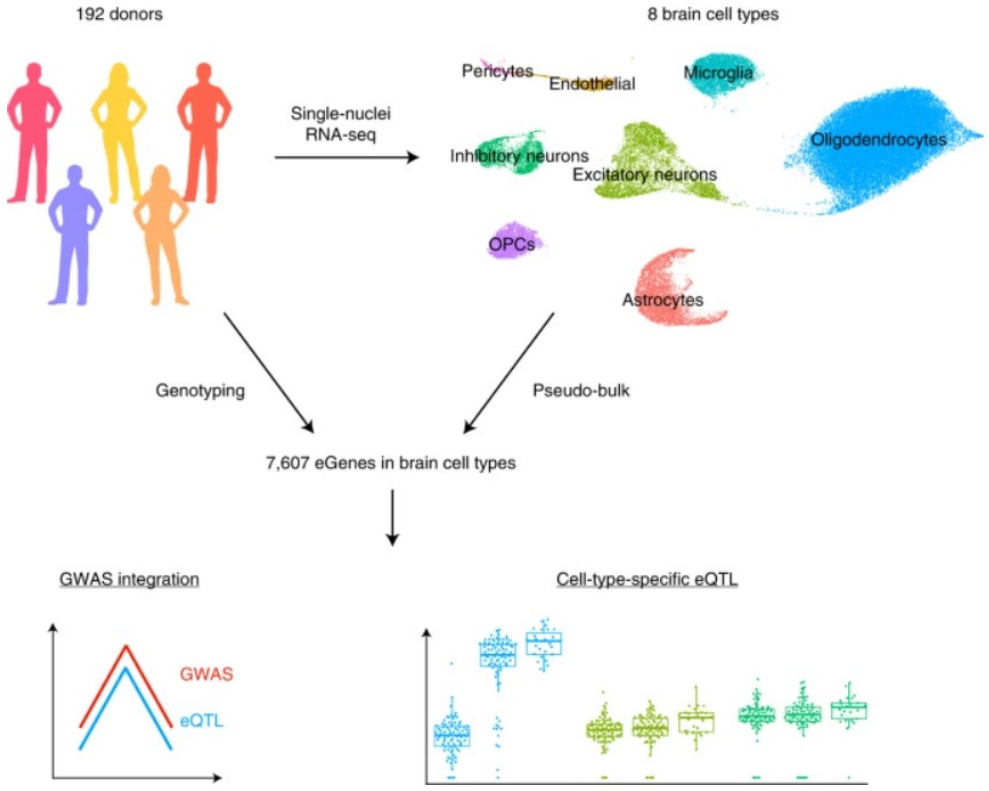
- eQTLs can be time-dependent or environment-specific
  - Cell-type-specific expression
  - Response to treatment
  - Developmental stage

- Most eQTL studies measure gene expression
  - At a **single timepoint**
  - In adult tissues
  - Using bulk RNA seq methods (no information on cell type differences)

# Cell-type-specific QTLs

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

**Fig. 1: Study summary.**




We performed single-nuclei RNA-seq on brain samples from 192 genotyped donors. We mapped *cis*-eQTLs for eight major brain cell types and identified a total of 7,607 *cis*-eQTL genes. We identified cell-type-specific genetic effects and leveraged our results to identify risk genes for brain disorders.

nature › nature neuroscience › resources › article

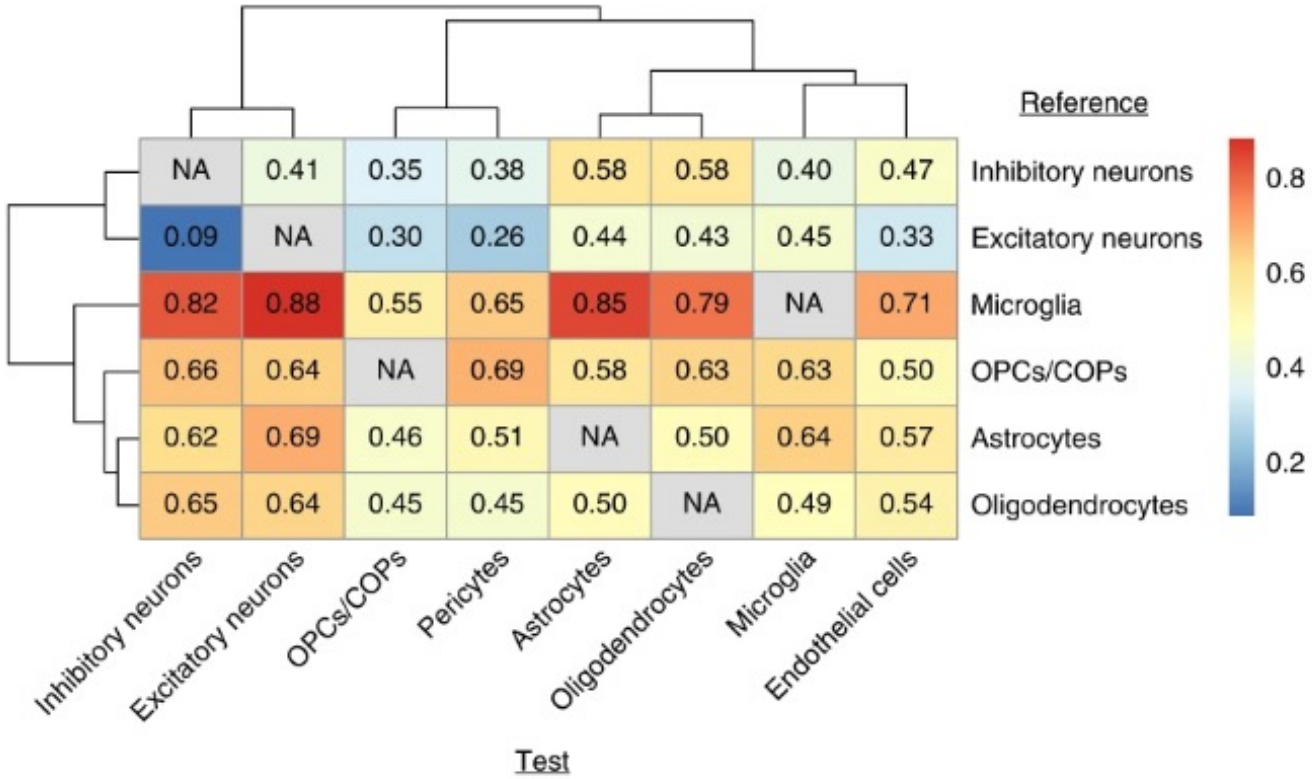
Resource | Published: 01 August 2022

## Cell-type-specific *cis*-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders

Julien Bryois, Daniela Calini, Will Macnair, Lynette Foo, Eduard Urich, Ward Ortmann, Victor Alejandro Iglesias, Suresh Selvaraj, Erik Nutma, Manuel Marzin, Sandra Amor, Anna Williams, Gonçalo Castelo-Branco, Vilas Menon, Philip De Jager & Dheeraj Malhotra

*Nature Neuroscience* **25**, 1104–1112 (2022) | Cite this article

# Cell-type-specific QTLs




- Almost all eQTLs detected in excitatory neurons have relatively similar effect sizes in inhibitory neurons (pi1 = 0.09)
- 88% of the microglia eQTLs have a different effect size in excitatory neurons (pi1 = 0.88)

Estimates of the proportions of cis-eQTLs on one brain cell type that have a different genetic effect in another brain cell type.


nature › nature neuroscience › resources › article

Resource | Published: 01 August 2022

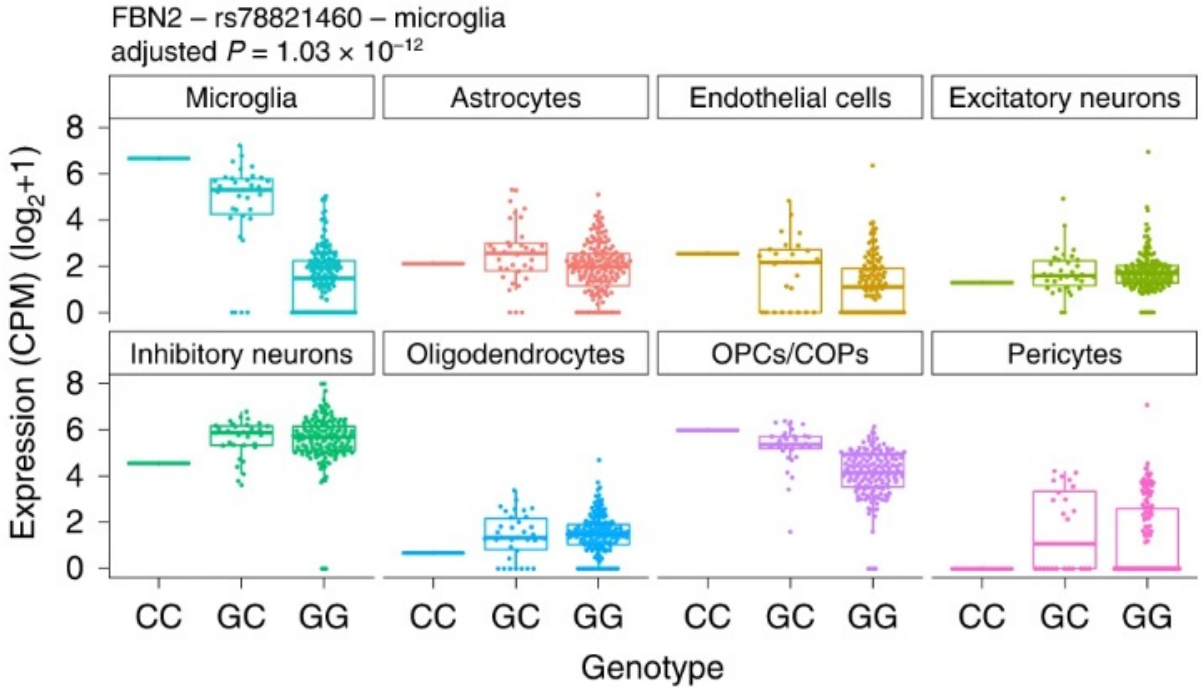

**Cell-type-specific *cis*-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders**


Julien Bryois, Daniela Calini, Will Macnair, Lynette Foo, Eduard Urich, Ward Ortmann, Victor Alejandro Iglesias, Suresh Selvaraj, Erik Nutma, Manuel Marzin, Sandra Amor, Anna Williams, Gonçalo Castelo-Branco, Vilas Menon, Philip De Jager & Dheeraj Malhotra



*Nature Neuroscience* **25**, 1104–1112 (2022) | Cite this article

# Cell-type-specific QTLs




nature › nature neuroscience › resources › article

Resource | Published: 01 August 2022

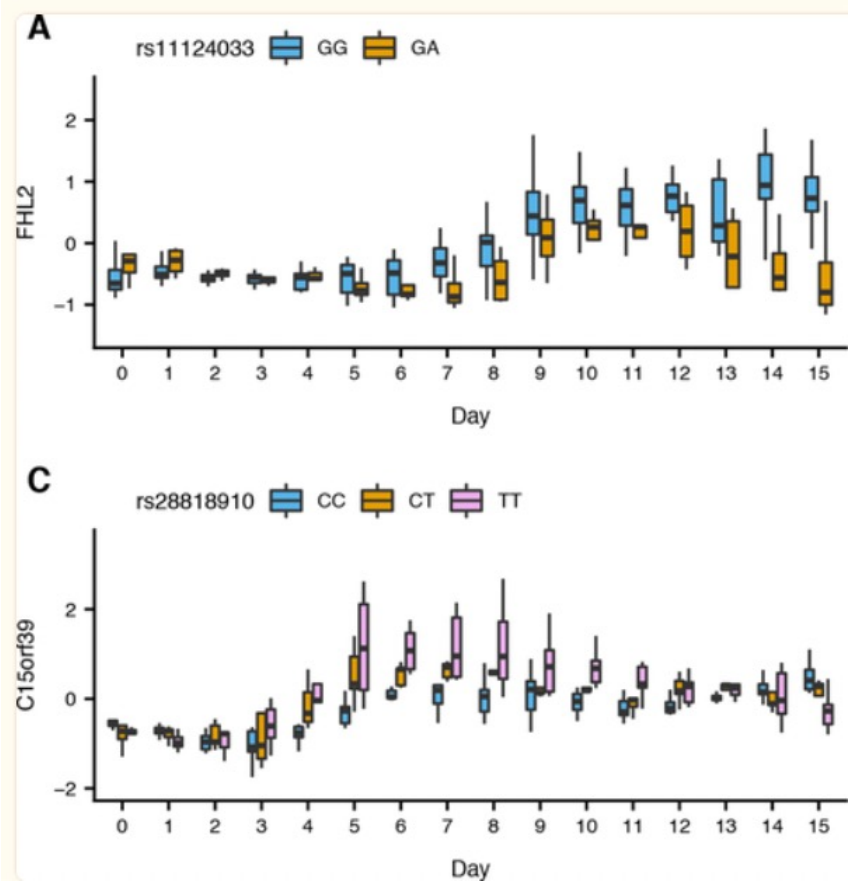
**Cell-type-specific *cis*-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders**

Julien Bryois, Daniela Calini, Will Macnair, Lynette Foo, Eduard Urich, Ward Ortmann, Victor Alejandro Iglesias, Suresh Selvaraj, Erik Nutma, Manuel Marzin, Sandra Amor, Anna Williams, Gonçalo Castelo-Branco, Vilas Menon, Philip De Jager & Dheeraj Malhotra

*Nature Neuroscience* **25**, 1104–1112 (2022) | Cite this article

# Developmental stage-specific QTLs


- iPSC differentiation into cardiomyocytes.
- eQTL analysis at 16 time points in 19 human cell lines



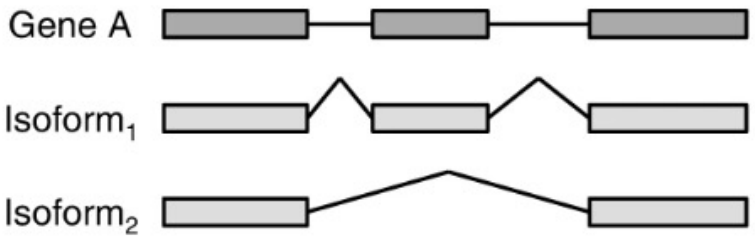
# Splice QTLs

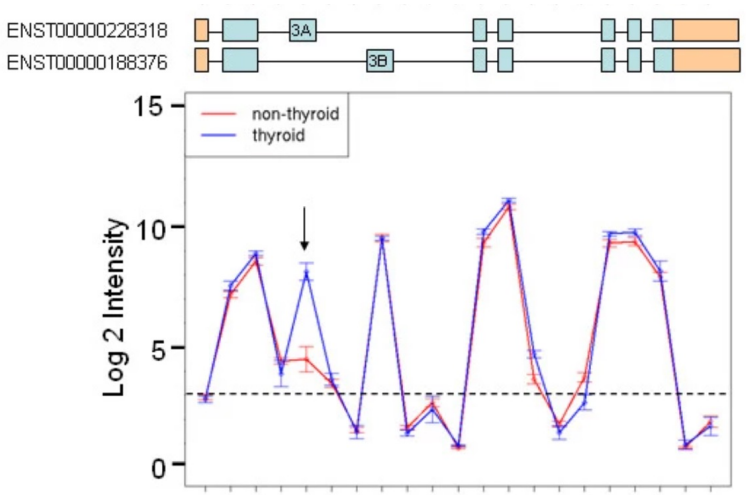
# splice QTLs (sQTLs)

- Alternative splicing (AS) produces multiple transcript isoforms from a single gene
  tissue-, cell type-, or condition-specific

- sQTLs - genetic variants that regulate AS

- sQTLs may change:
  - Coding sequence by skipping or inclusion of coding exons, affecting protein structure and function
  - UTRs, affecting RNA stability or translational efficacy




Shah & Pallas BMC Bioinformatics 2009

# Splice variants

## A Deep Intronic Variant in *LDLR* in Familial Hypercholesterolemia

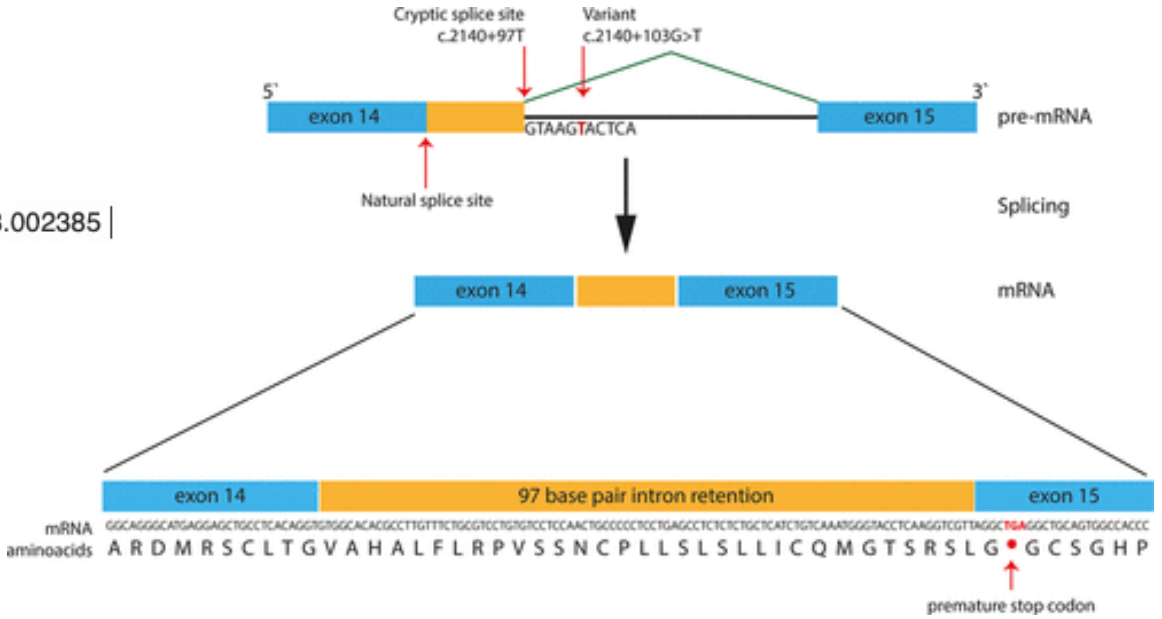
### Time to Widen the Scope?

Laurens F. Reeskamp, Merel L. Hartgers, Jorge Peter, Geesje M. Dallinga-Thie, Linda Zuurbier, Joep C. Defesche, Aldo Grefhorst and G. Kees Hovingh

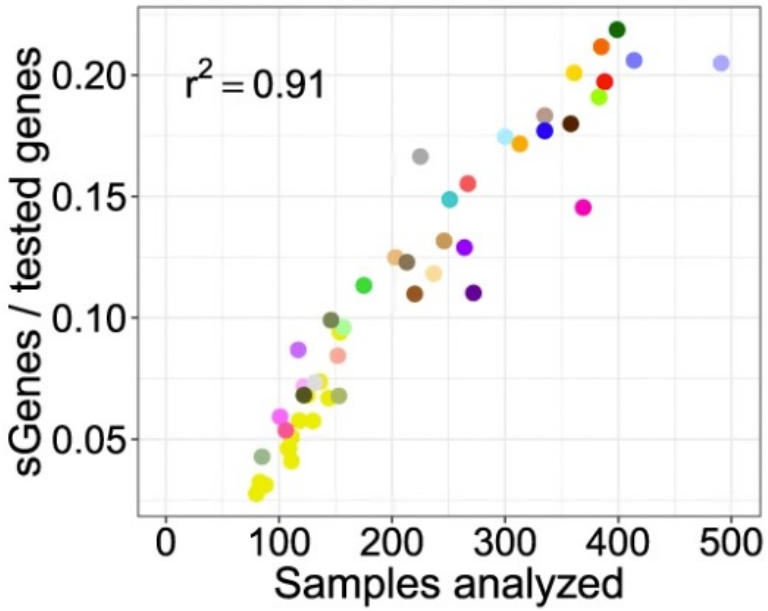
**Originally published** 11 Dec 2018 | https://doi.org/10.1161/CIRCGEN.118.002385 |
Circulation: Genomic and Precision Medicine. 2018;11:e002385

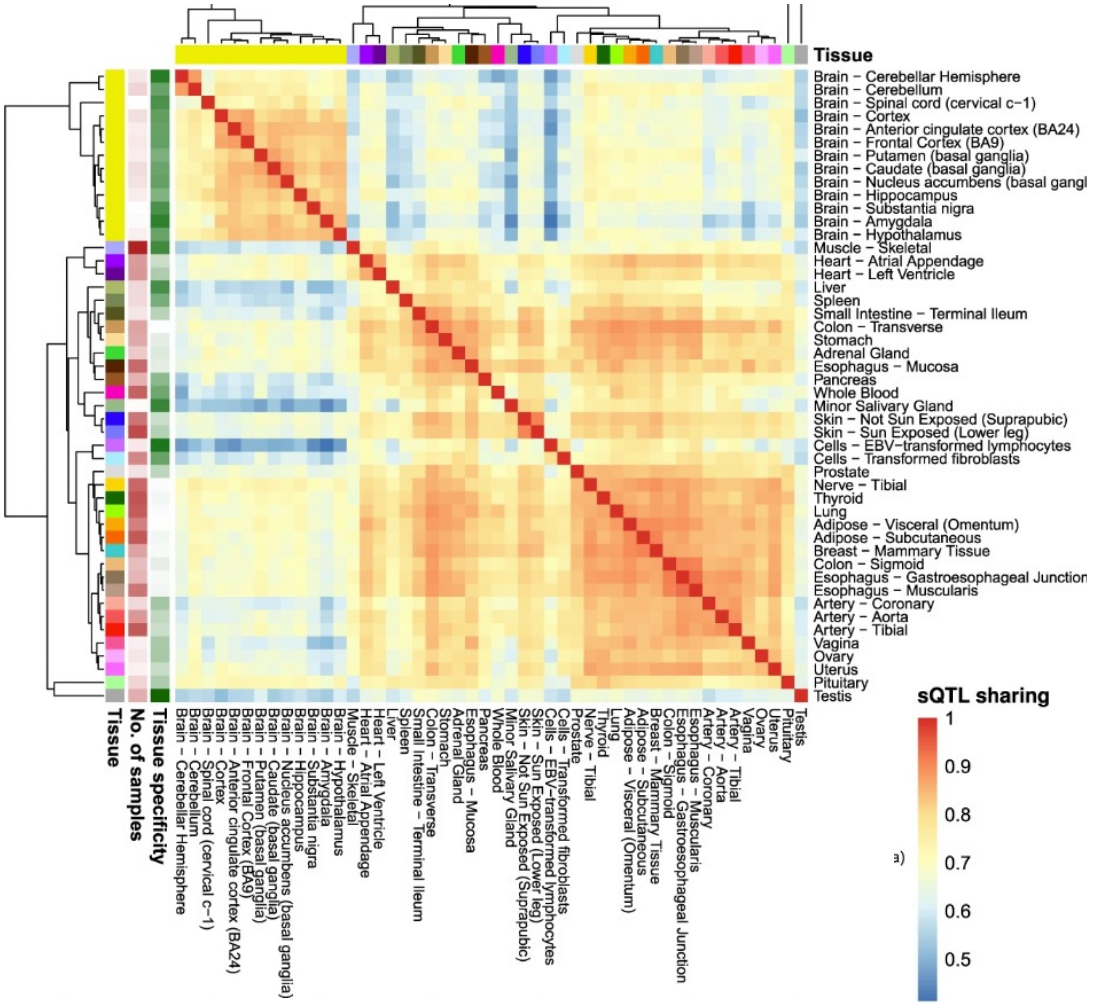
A novel deep intronic variant in LDLR (c.2140+103G>T) was found to cosegregate with high LDL-C in 5 patients but was not present in 4 unaffected family members.

Intronic variant (c.2140+103G>T) in *LDLR* results in a 97-nucleotide insertion leading to a frameshift and premature stop codon in exon 15 of *LDLR*.



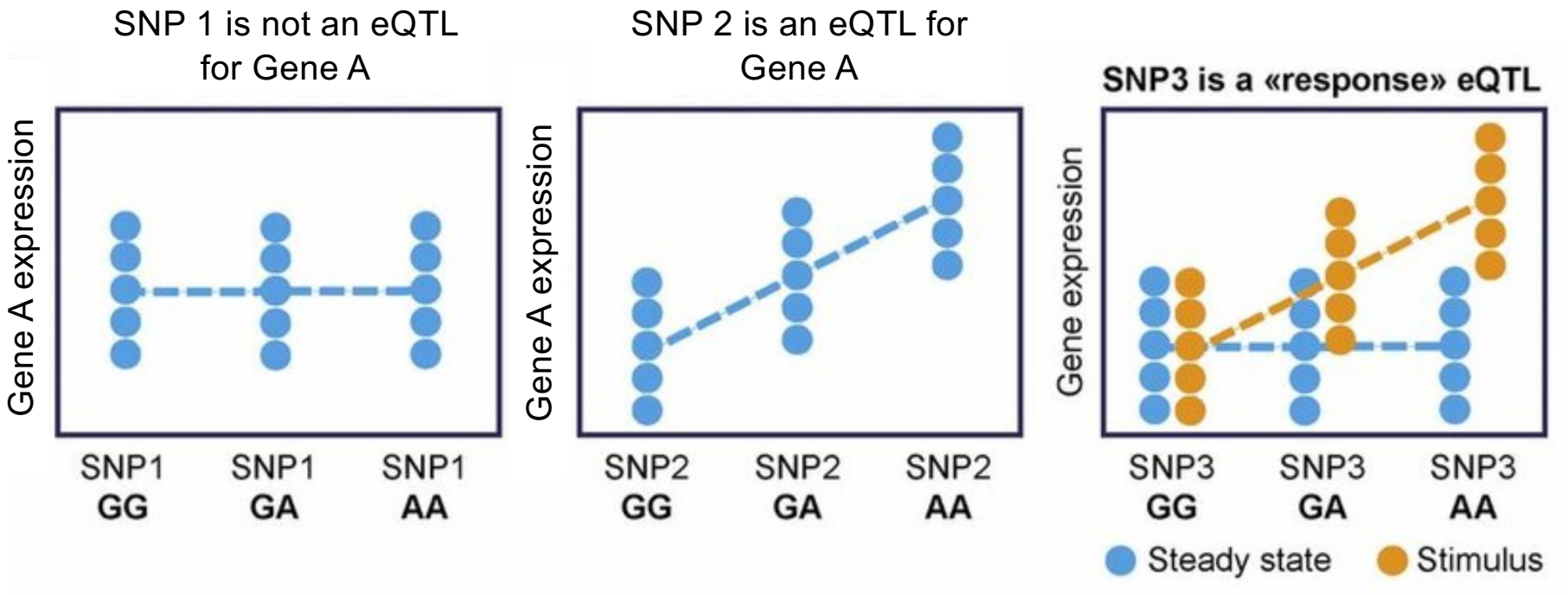
# sQTL sharing across tissues




Garrido-Martin et al Nat comm 2021

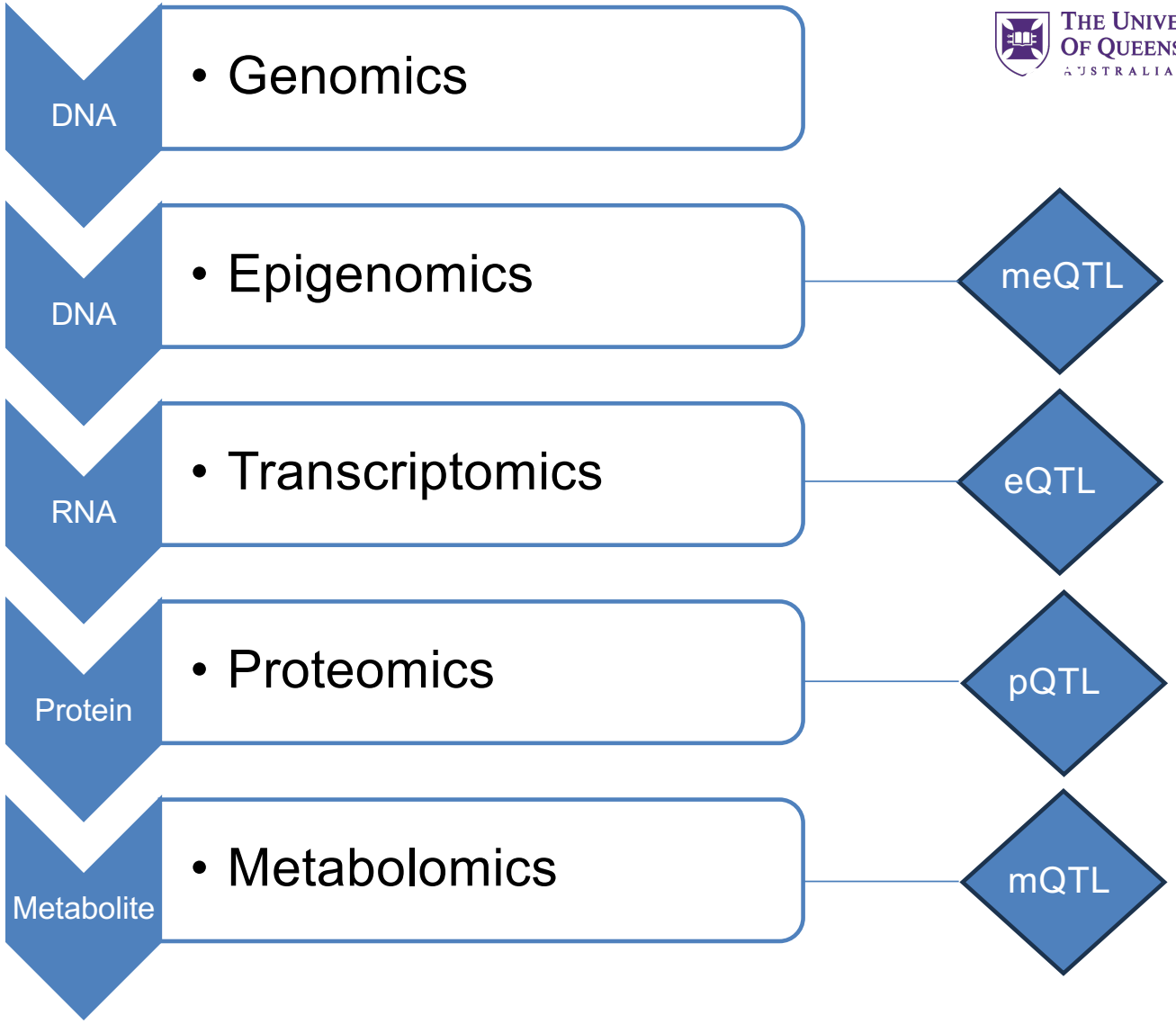
# Stimulus-dependent eQTL



Changes in gene expression doesn't necessarily translate to changes in protein levels

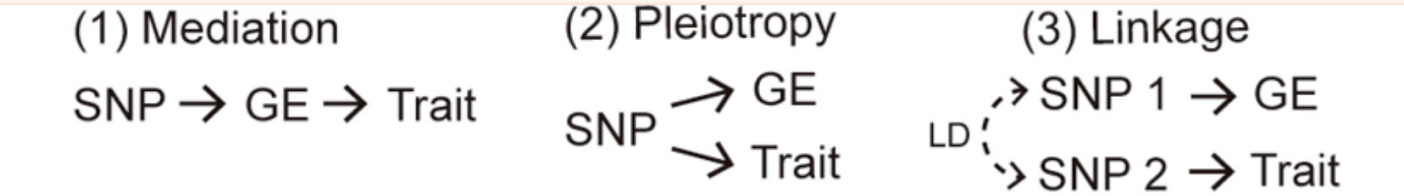
Integration of genetic and omic data for greater understanding of variant consequence.

Limitation – most other omic QTLs are measured in blood



# Association does not mean causality

Several different scenarios can result in an overlap between GWAS loci and eQTLs




- Co-regulation of nearby genes – multiple eGenes for the same SNPs
- Unclear which is the causal gene just based on overlapping association
- Several statistical methods to determine if a variant impacts phenotype through gene expression change (SMR, coloc)

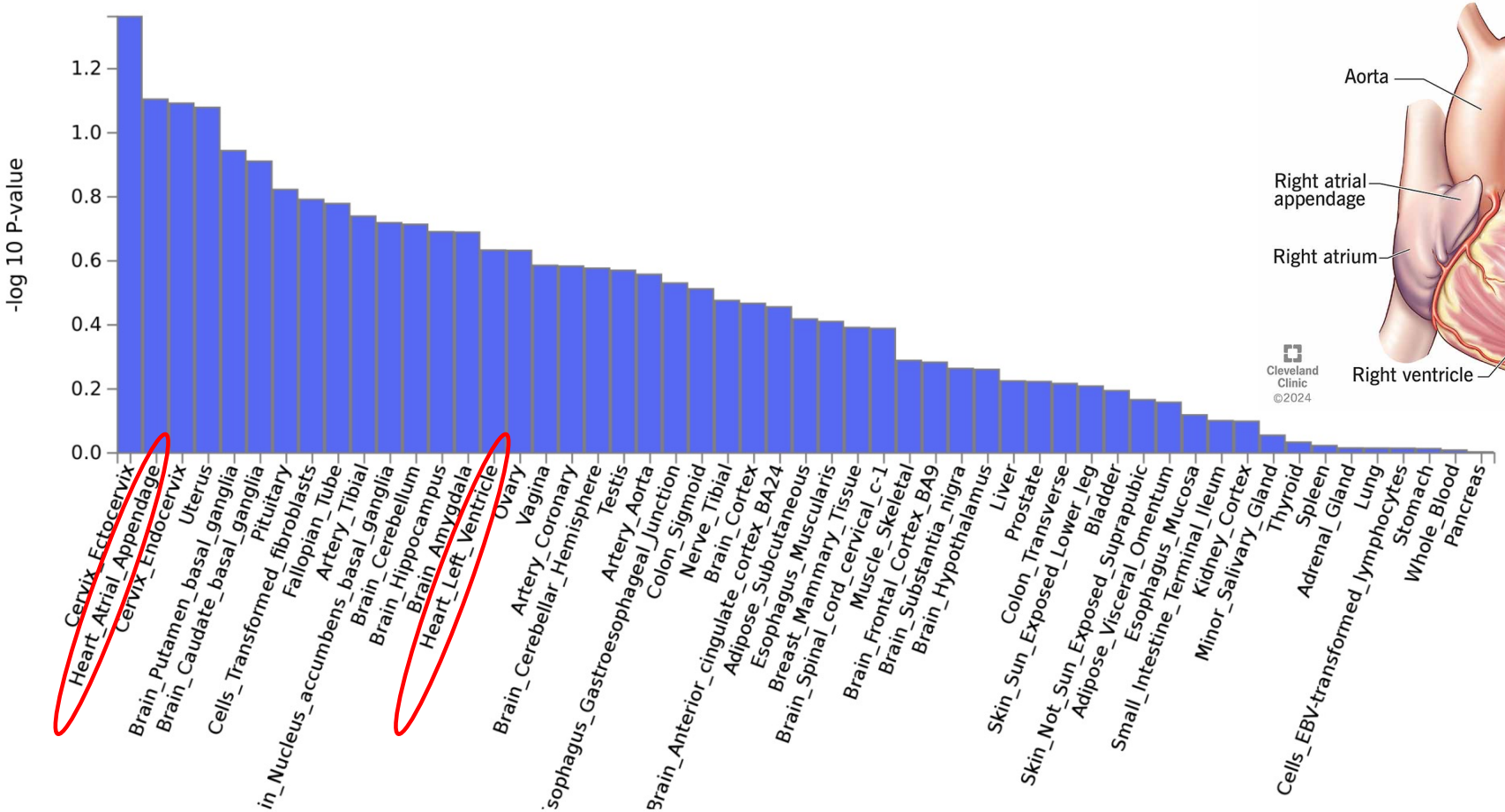
Article | Open access | Published: 09 January 2020

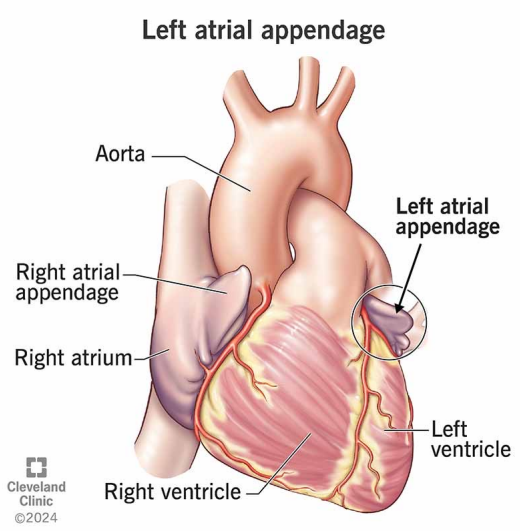
# Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure

Sonia Shah, Albert Henry, Carolina Roselli, Honghuang Lin, Garðar Sveinbjörnsson, Ghazaleh Fatemifar, Åsa K. Hedman, Jemma B. Wilk, Michael P. Morley, Mark D. Chaffin, Anna Helgadottir, Niek Verweij, Abbas Dehghan, Peter Almgren, Charlotte Andersson, Krishna G. Aragam, Johan Ärnlöv, Joshua D. Backman, Mary L. Biggs, Heather L. Bloom, Jeffrey Brandimarto, Michael R. Brown, Leonard Buckbinder, David J. Carey, Regeneron Genetics Center, ... R. Thomas Lumbers ✉ + Show authors

*Nature Communications* **11**, Article number: 163 (2020) | Cite this article

# Tissue enrichment analysis



# Shah et al. Nat Comms 2020 - Heart Failure GWAS

| Sentinel variant | eQTL Gene | GTEXv7 | | | | | | MAGnET | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Heart Left Ventricle [N=272] | | | Heart Right Atrial Appendage [N=264] | | | Left Atrium [N=101] | | |
| | | Effect | SE | P-value | Effect | SE | P-value | Effect | SE | P-value |
| rs600038 | ABO | 0.324 | 0.068 | **1.87E-06** | 0.356 | 0.077 | **4.24E-06** | 0.867 | 0.153 | **1.35E-08** |
| rs600038 | SURF1 | -0.269 | 0.063 | **1.93E-05** | -0.321 | 0.065 | **8.18E-07** | -0.306 | 0.164 | **6.21E-02** |

*ABO* determines blood group

SURF1 encodes a protein localized to the inner mitochondrial membrane and thought to be involved in the biogenesis of the cytochrome c oxidase complex

# Shah et al. Nat Comms 2020 - Heart Failure GWAS

| Sentinel variant | eQTL Gene | GTEXv7 | | | | | | | | MAGnET | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Heart Left Ventricle [N=272] | | | | Heart Right Atrial Appendage [N=264] | | | | Left Atrium [N=101] | | | |
| | | Effect | SE | P-value | Coloc PP | Effect | SE | P-value | Coloc PP | Effect | SE | P-value | Coloc PP |
| rs600038 | ABO | 0.324 | 0.068 | **1.87E-06** | 0.02 | 0.356 | 0.077 | **4.24E-06** | **0.76** | 0.867 | 0.153 | **1.35E-08** | **0.83** |
| rs600038 | SURF1 | -0.269 | 0.063 | **1.93E-05** | 7.5E-06 | -0.321 | 0.065 | **8.18E-07** | 3.3E-05 | -0.306 | 0.164 | **6.21E-02** | 0.04 |

*ABO* determines blood group

SURF1 encodes a protein localized to the inner mitochondrial membrane and thought to be involved in the biogenesis of the cytochrome c oxidase complex

**Coloc PP4:** Association with both traits, driven by exactly *one shared* causal variant.

nature



Article | Open access | Published: 28 January 2026

# Advancing regulatory variant effect prediction with AlphaGenome

Žiga Avsec, Natasha Latysheva, Jun Cheng, Guido Novati, Kyle R. Taylor, Tom Ward, Clare Bycroft, Lauren Nicolaisen, Eirini Arvaniti, Joshua Pan, Raina Thomas, Vincent Dutordoir, Matteo Perino, Soham De, Alexander Karollus, Adam Gayoso, Toby Sargeant, Anne Mottram, Lai Hong Wong, Pavol Drotár, Adam Kosiorek, Andrew Senior, Richard Tanburn, Taylor Applebaum, ... Pushmeet Kohli + Show authors

*Nature* **649**, 1206–1218 (2026) | Cite this article