

Understanding the genetics of complex traits II

Gavin Band gavin.band@well.ox.ac.uk

BA Human Sciences

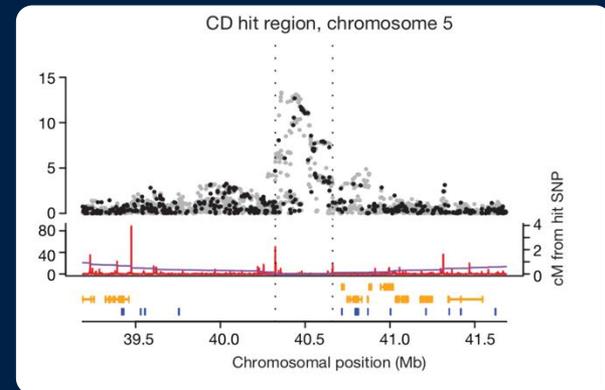
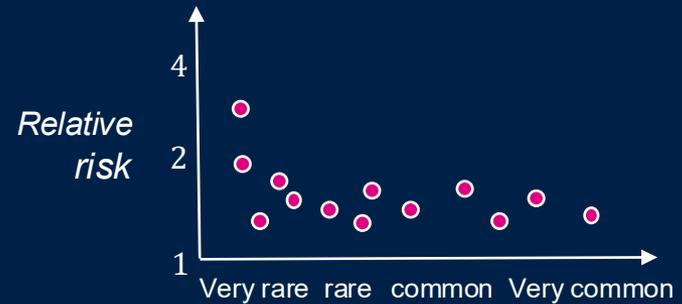
Friday 6th March 2026



Main points in this lecture

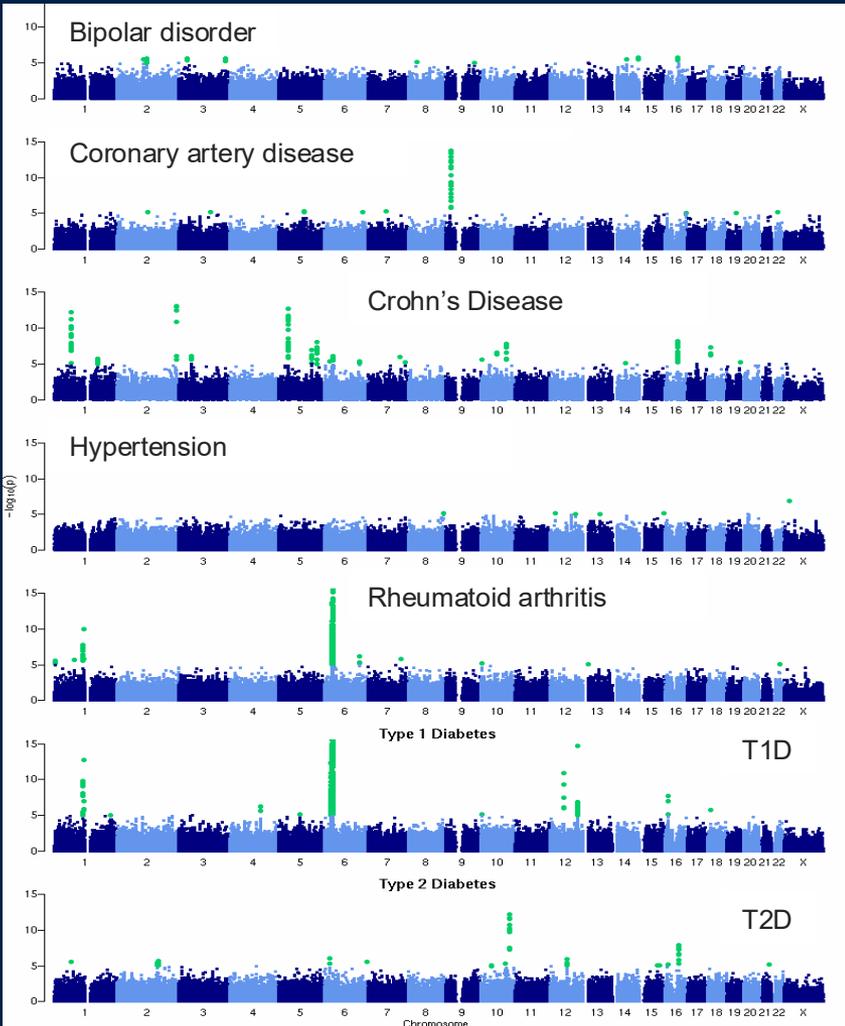
How polygenic do traits get,
anyway?

Extracting biological information
from GWAS

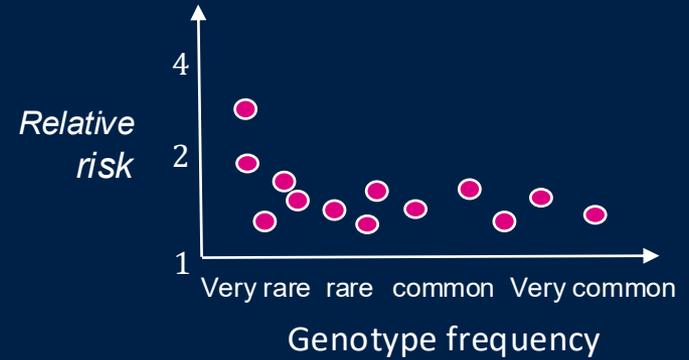


Actual results
from the Wellcome Trust Case-Control
Consortium study:

Number of
signals



My cartoon:



“Common variant, common trait” hypothesis

Are you convinced?

Additive model, N=2,000 cases + 3000 controls per phenotype

Actual results
from the Wellcome Trust Case-Control
Consortium study:

Number of
signals

1

1

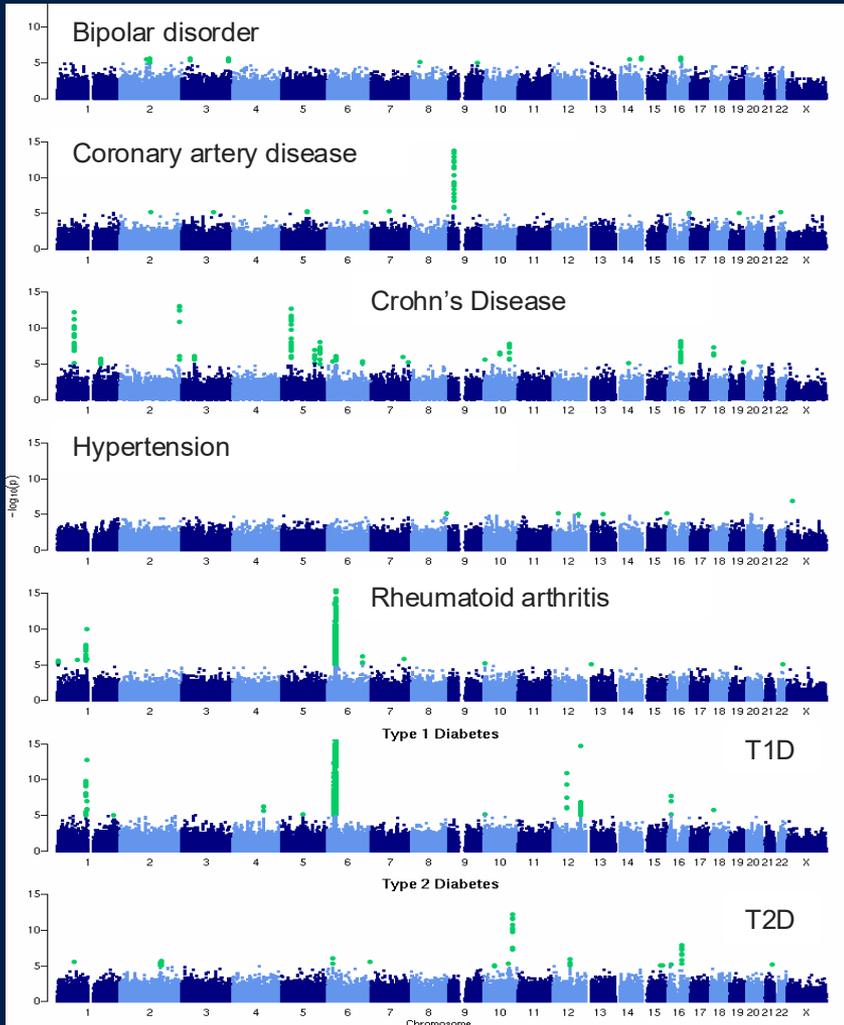
9

0

3

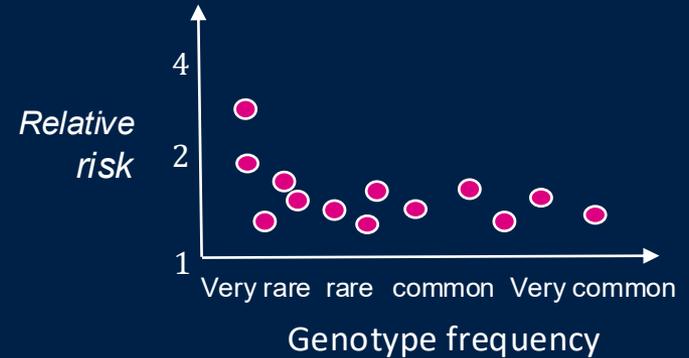
7

3



Additive model, N=2,000 cases + 3000 controls per phenotype

My cartoon:



“Common variant, common trait” hypothesis

Are you convinced?

Maybe we haven't found them all -
how could we find more?

Remember the formula

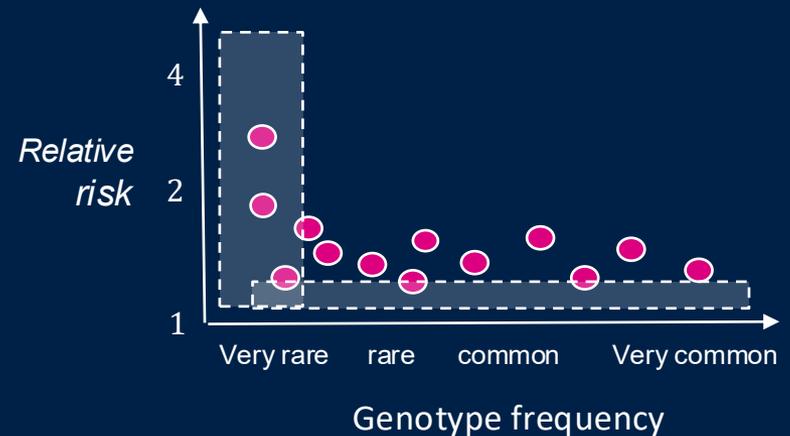
$$\text{Standard error}(\log OR) \approx \frac{1}{\sqrt{2N \times f(1-f) \times \phi(1-\phi)}}$$

N = sample size

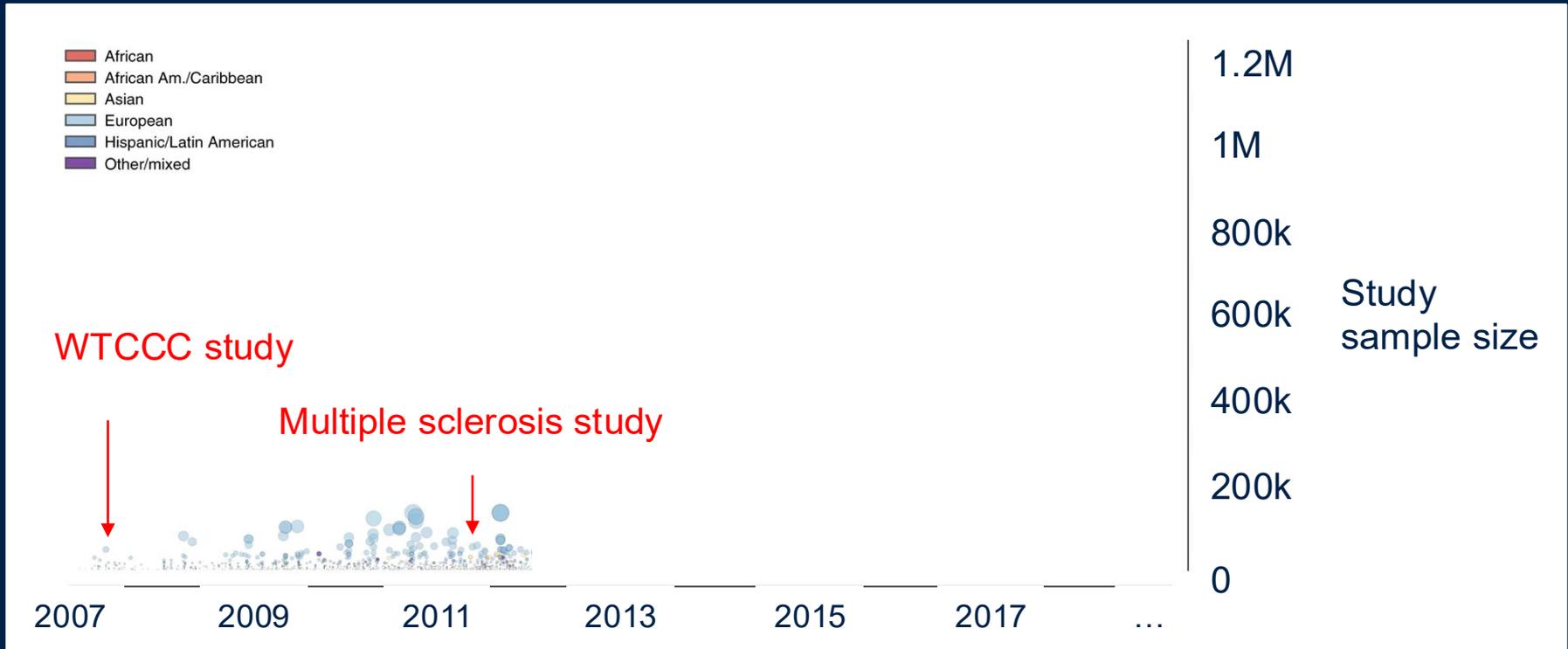
f = frequency of allele

ϕ = proportion of cases

To find more associations we should:
increase the sample size



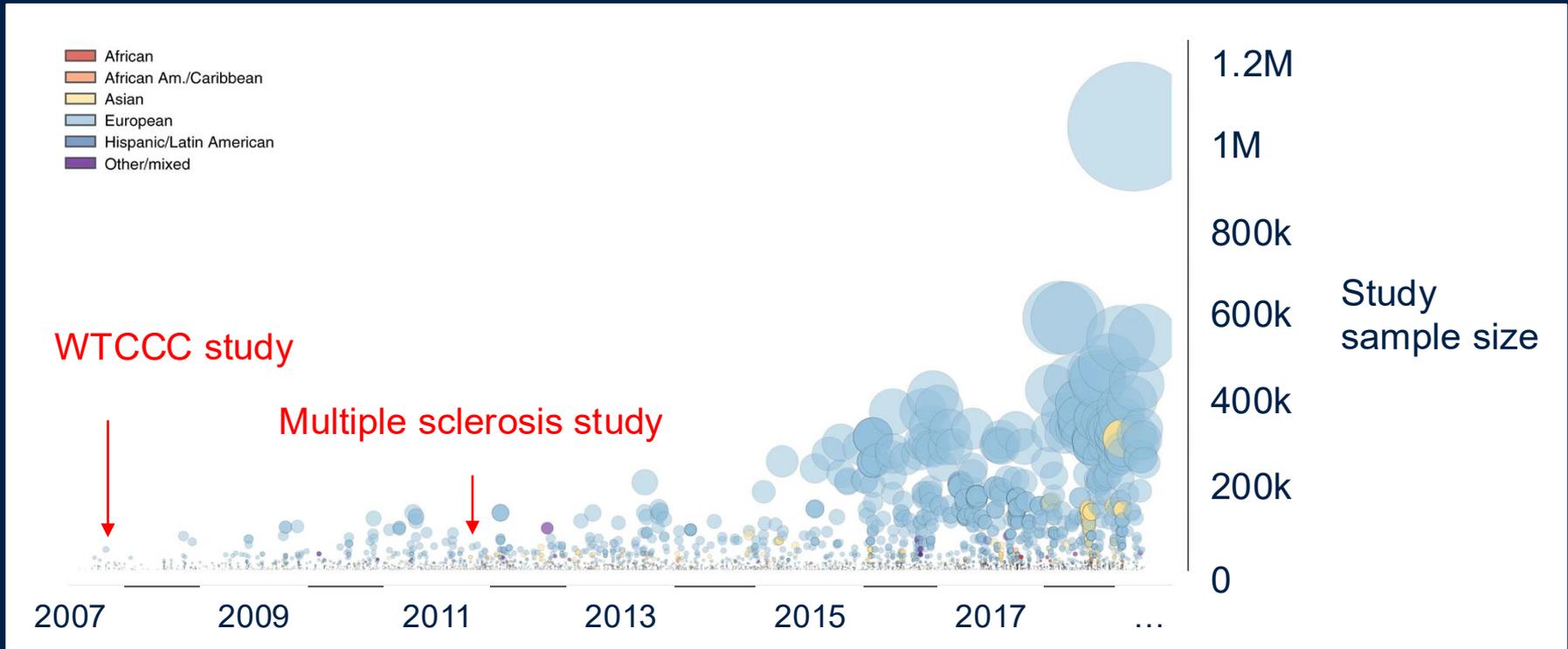
GWAS revolution



Mills & Rahal, "A scientometric review of genome-wide association studies", Communications Biology 2019

NHGRI GWAS Catalog: <https://www.ebi.ac.uk/gwas/>

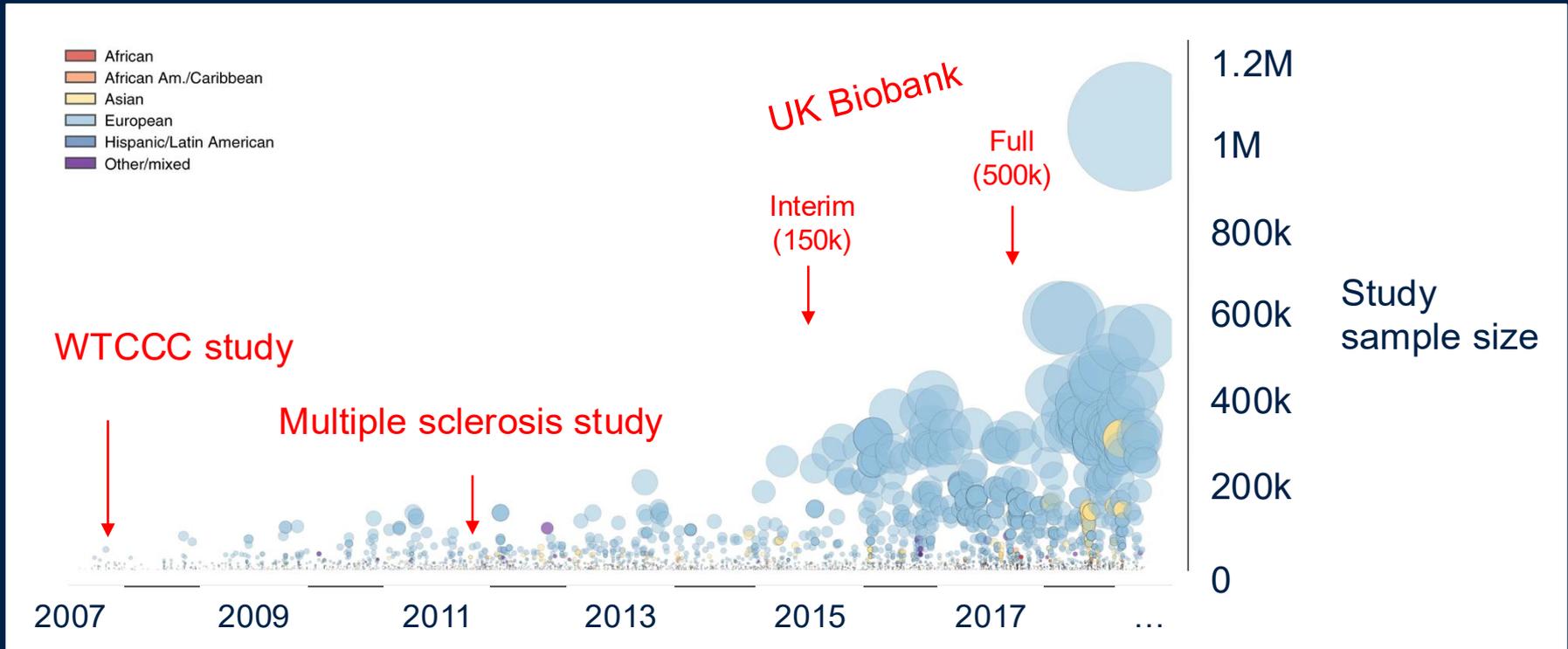
GWAS revolution



Mills & Rahal, "A scientometric review of genome-wide association studies", Communications Biology 2019

NHGRI GWAS Catalog: <https://www.ebi.ac.uk/gwas/>

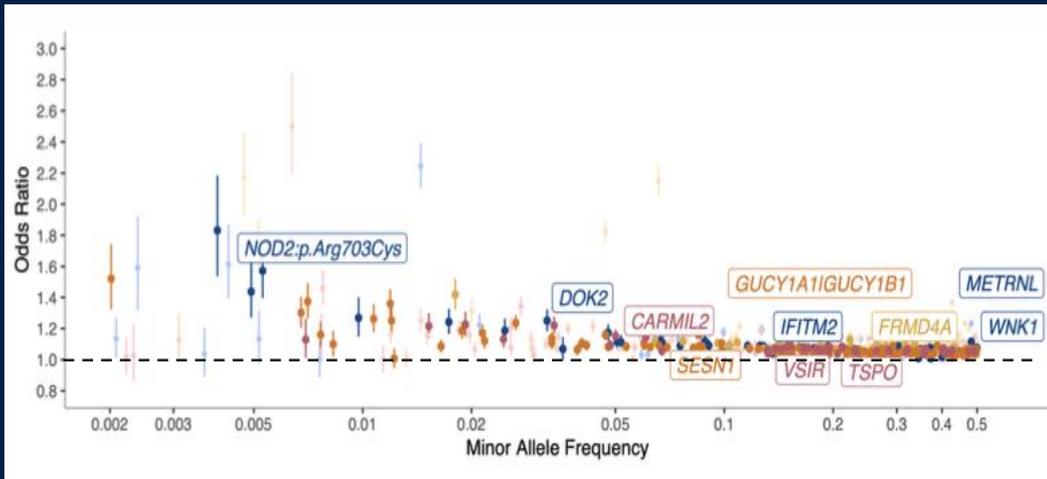
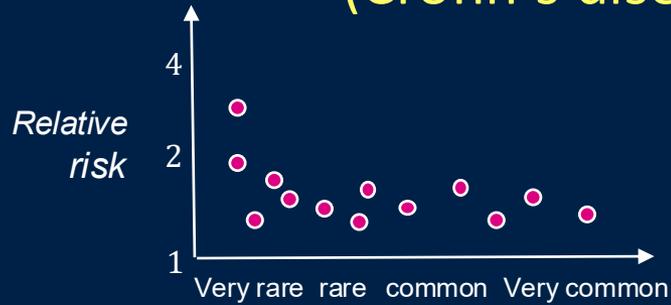
GWAS revolution



Mills & Rahal, "A scientometric review of genome-wide association studies", Communications Biology 2019

NHGRI GWAS Catalog: <https://www.ebi.ac.uk/gwas/>

Inflammatory bowel disease (Crohn's disease and ulcerative colitis)



N = 125,992 IBD cases
1.2 million controls

> 600 association signals.

Abstract citation ID: jjad212.0008

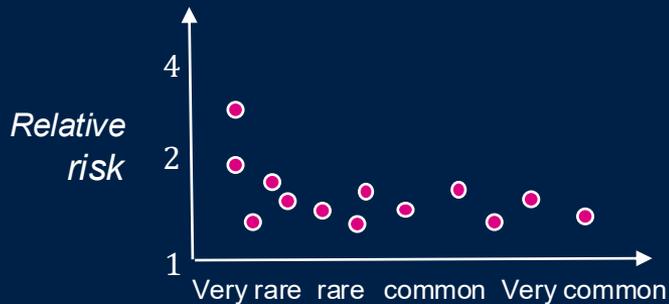
OP08

Multi-ancestry genome-wide association study of inflammatory bowel disease identifies 125 novel loci and directly implicates new genes in disease susceptibility

L. Fachal¹, on behalf of the International IBD Genetics Consortium

¹Wellcome Sanger Institute, Human Genetics, Hinxton- Saffron Walden, United Kingdom

Type 2 diabetes



N = 74,000 T2D cases
And 824,000 controls

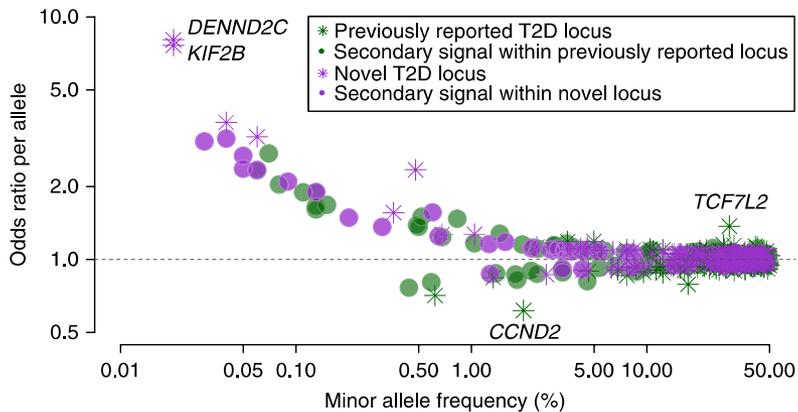


Fig. 5 | The relationship between effect size and MAF. Conditional- and joint-analysis effect size (y axis) and MAF (x axis) for 403 conditionally independent SNPs. Previously reported T2D-associated variants are shown in green, and novel variants are shown in purple. Stars and circles represent the 'strongest regional lead at a locus' and 'lead variants for secondary signals', respectively.

403 signals

"conditionally independent" meaning
some of them overlap the same regions

nature
genetics

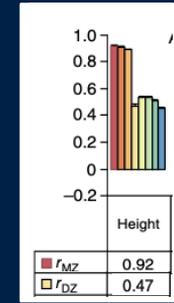
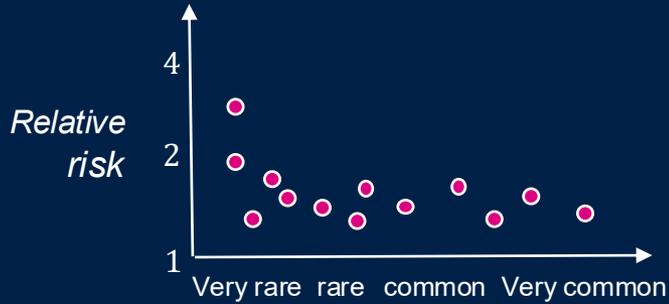
ARTICLES

<https://doi.org/10.1038/s41588-018-0241-6>

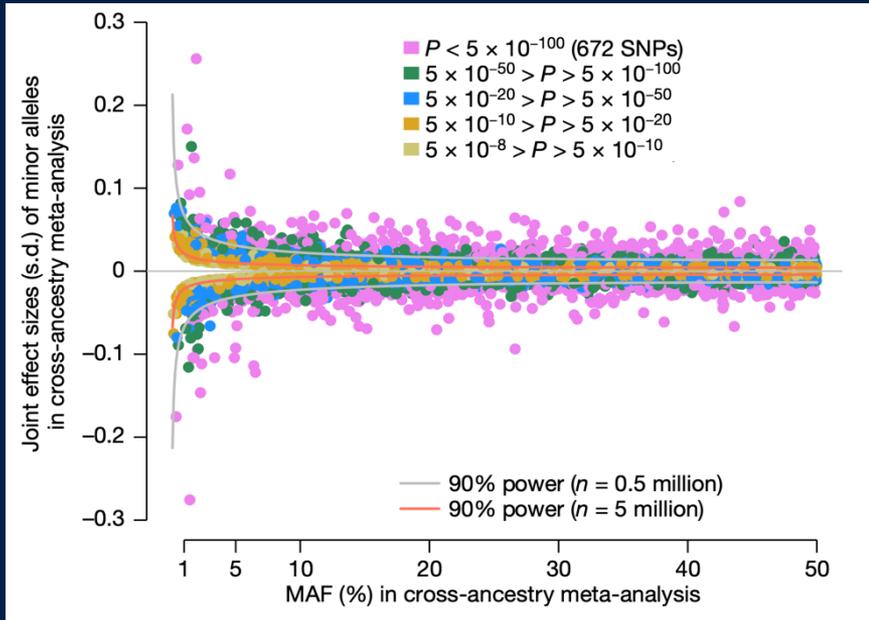
Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps

GWAS of human height

In 5.4 million individuals



~90% heritability



? signals

Article

A saturated map of common genetic variants associated with human height

<https://doi.org/10.1038/s41586-022-05275-y>

Received: 19 December 2021

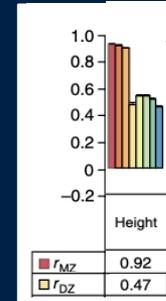
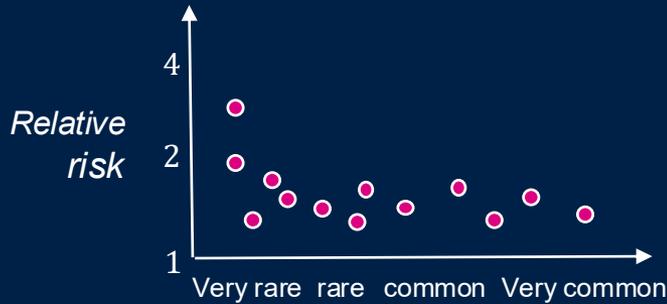
Accepted: 24 August 2022

Common single-nucleotide polymorphisms (SNPs) are predicted to collectively explain 40–50% of phenotypic variation in human height, but identifying the specific variants and associated regions requires huge sample sizes'. Here, using data from a genome-wide association study of 5.4 million individuals of diverse ancestries, we

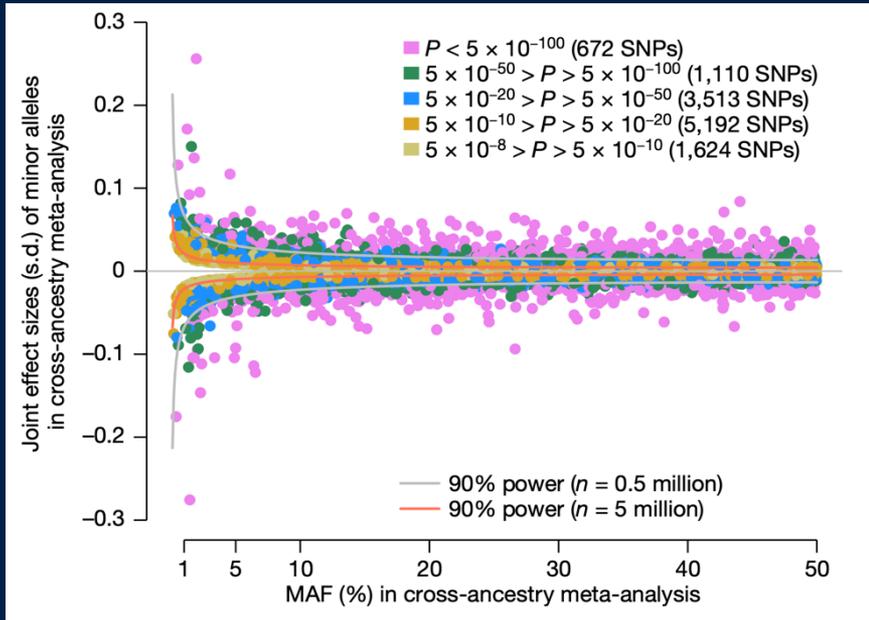
N = 5.4 million

GWAS of human height

In 5.4 million individuals



~90% heritability



12,111 independent signals

Collectively explaining 50% of heritability

Article

A saturated map of common genetic variants associated with human height

<https://doi.org/10.1038/s41586-022-05275-y>

Received: 19 December 2021

Accepted: 24 August 2022

Published online: 12 October 2022

Open access

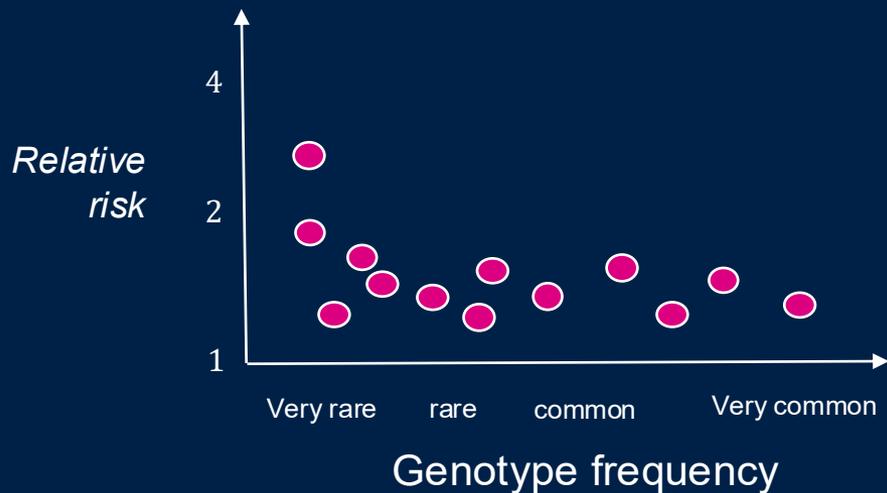
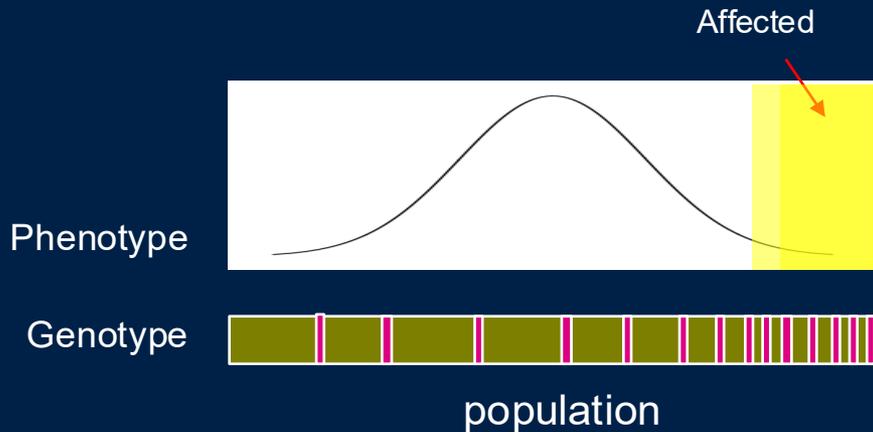
Check for updates

Common single-nucleotide polymorphisms (SNPs) are predicted to collectively explain 40–50% of phenotypic variation in human height, but identifying the specific variants and associated regions requires huge sample sizes. Here, using data from a genome-wide association study of 5.4 million individuals of diverse ancestries, we show that 12,111 independent SNPs that are significantly associated with height account for nearly all of the common SNP-based heritability. These SNPs are clustered within 7,209 non-overlapping genomic segments with a mean size of around 90 kb, covering about 21% of the genome. The density of independent associations varies across the genome and the regions of increased density are enriched for biologically relevant genes. In out-of-sample estimation and prediction, the 12,111 SNPs (or all SNPs in the HapMap 3 panel) account for 40% (45%) of phenotypic variance in populations of European ancestry but only around 10–20% (14–24%) in populations of other ancestries. Effect sizes, associated regions and gene prioritization are similar

N = 5.4 million

They collectively explain ~ 50% of heritability
In European ancestry people

Common variant, common disease hypothesis



A complex trait.

Caused by many factors, each having a small overall effect. Including

- Many genetic variants, including common ones
- Environmental factors
- Gene-environment or gene-gene interactions
- ...

How are these studies possible?
Consortia and meta-analysis

Consolidation question from last lecture

WTCCC2 GWAS of multiple sclerosis (9,772 cases and 7,376 controls).

For further information about terms used below, hover over the red question marks.

Region

dbSNP id: [rs11581062](#)
 status: novel association
 physical position: 01:101,180,107
 association region: [01:100,983,315-101,455,310](#)
 functional tag: N/A
 nearest gene: [SLC30A7](#)
 candidate gene: [VCAM1](#)*

Signal

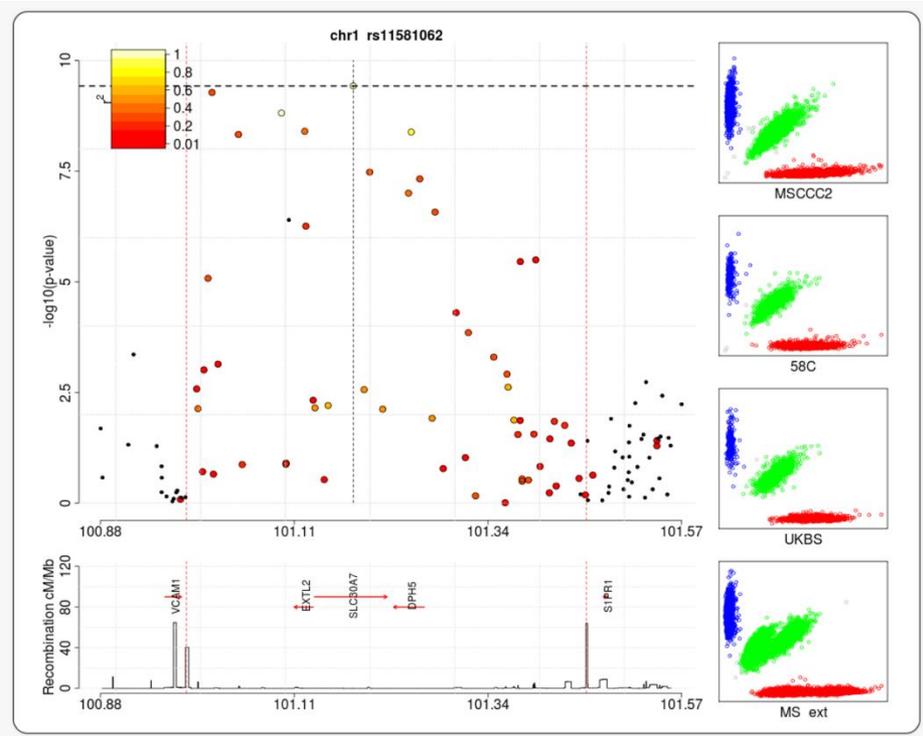
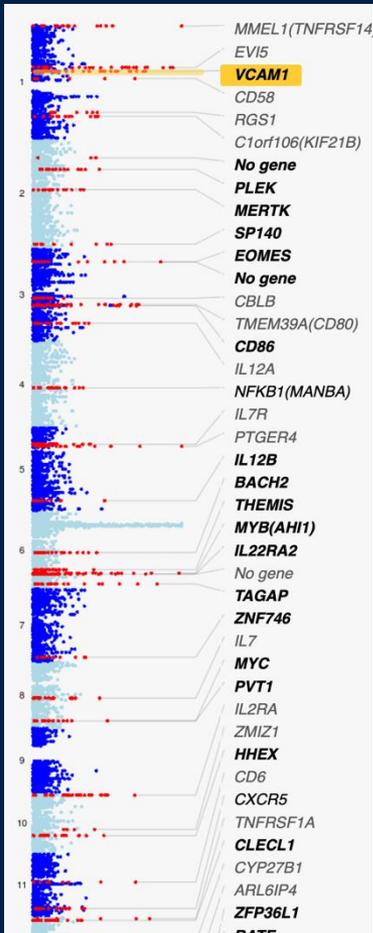
p-value discovery: 3.7e-10
 OR discovery (95% CI): 1.13 (1.09-1.18)
 p-value replication: 4.20e-02 (one-sided)
 OR replication (95% CI): 1.07 (0.99-1.15)
 p-value combined: 2.50e-10
 OR combined (95% CI): 1.12 (1.1-1.13)
 Risk (non-risk) allele: G(A)

Allele frequencies

Country	controls / cases	control / case frequency
Australia	- / 647	- / 0.32
Belgium	- / 544	- / 0.33
Denmark	- / 332	- / 0.32
Finland	2165 / 581	0.23 / 0.24
France	347 / 479	0.31 / 0.34
Germany	1699 / 1100	0.29 / 0.31
Ireland	- / 61	- / 0.34
Italy	571 / 745	0.30 / 0.33
Norway	121 / 953	0.26 / 0.28
Poland	- / 58	- / 0.27
Spain	- / 205	- / 0.36
Sweden	1928 / 685	0.27 / 0.28
UK	5175 / 1854	0.29 / 0.32
USA	5370 / 1382	0.29 / 0.32

Proximal genes

[DPH5](#), [EXTL2](#), [S1PR1](#), [SLC30A7](#), [VCAM1](#)*

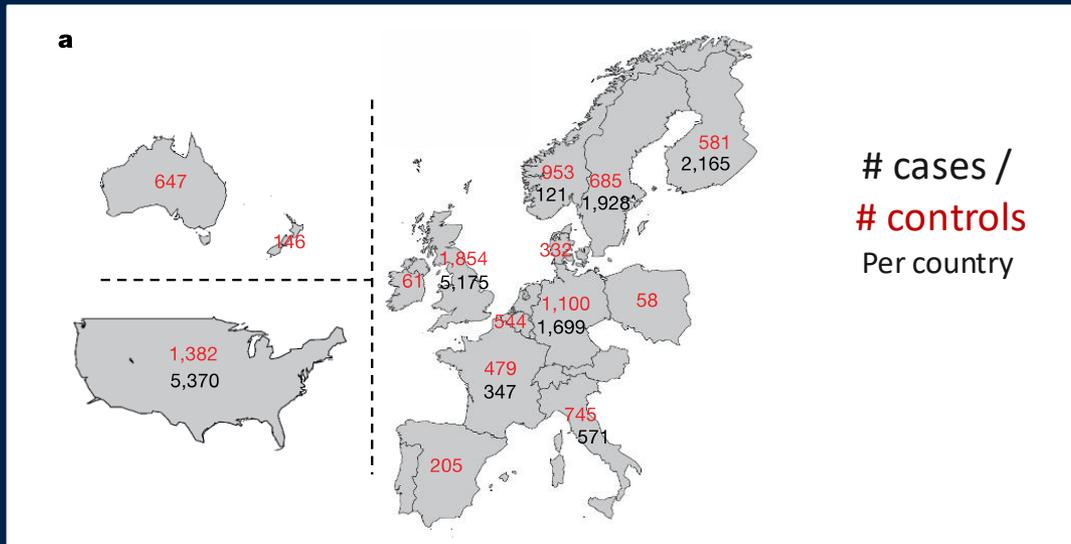


Can you explain?

Consortia and meta-analysis

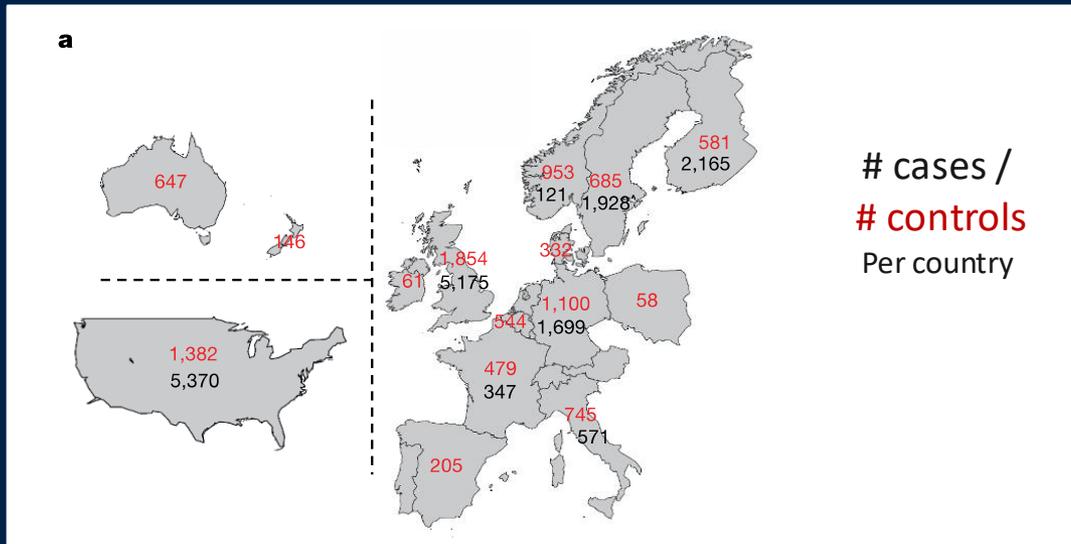
To generate such large sample sizes for “common” (but still relatively rare) diseases, requires setting up large multi-centre collaborations. This is fun to be involved in but comes with its own analysis challenges....

Dealing with population structure

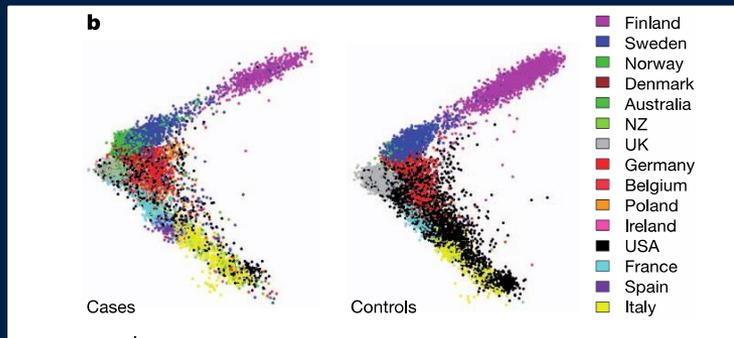


This study suffered from a key problem. Can you see what it is?

Dealing with population structure



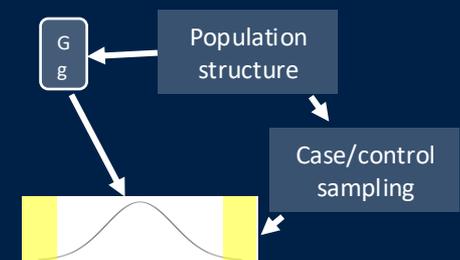
This study suffered from a key problem. Can you see what it is?



First two “principal components” obtained purely from the genotypes

Case-control sampling is correlated with genome-wide genetic variation.

“Confounding by population structure”

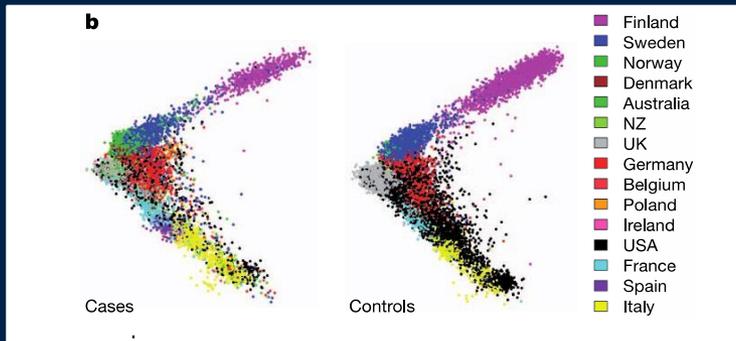


Population structure: solutions

Instead of simple 2x2 table

1. Regression including principal components

$$\text{outcome} \sim \text{genotype} + PCs$$

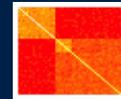


Plot of first two principal components obtained from the genome-wide genotypes

Uses just the strongest directions of variation in relatedness (population structure)

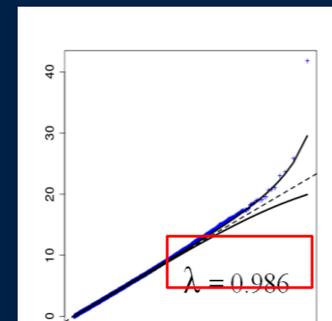
2. Linear mixed model

$$\text{outcome} \sim \text{genotype} +$$



Include a **genetic relatedness matrix** computed from **genome-wide genotypes** in the association test

Uses the **entire matrix** of relationships



MS study

Most p-values are now not inflated

Anatomy of an association analysis

All GWAS should report data in a way that can be re-used by future studies.

This study used several previous GWAS to conduct replication. All the details are given in a supplementary table:

Gene	Risk Allele	WAS + replication			GWAS			UK only GWAS			non-UK only GWAS		combined replication		GeneMSEA NL replication			GeneMSEA US replication			GeneMSEA CH replication			ANZ replication			BWH replication		
		pval	OR (95% CI)	log10 (BayesFactor)	pval	OR (95% CI)	log10 (BayesFactor)	pval	OR (95% CI)	pval	OR (95% CI)	pval*	OR (95% CI)	pval*	OR (95% CI)	inf	pval*	OR (95% CI)	inf	pval*	OR (95% CI)	inf	pval*	OR (95% CI)	inf	pval*	OR (95% CI)	inf	
MMEL1	C	1.00E-14	1.14 (1.11-1.17)	11.39	3.10E-14	1.16 (1.13-1.19)	7.43	0.0073	1.12 (1.09-1.15)	7.10E-13	1.17 (1.14-1.20)	0.0085	1.08 (1.01-1.15)	0.26	1.1 (1.07-1.13)	0.94	0.18	1.1 (1.07-1.13)	1.01	0.24	1.11 (1.08-1.14)	1.03	0.006	1.15 (1.12-1.18)	1.1	0.41	1.02 (1.00-1.04)	1.0	
EVI5	A	5.80E-15	1.15 (1.12-1.18)	9.15	6.50E-12	1.15 (1.12-1.18)	9.15	2.90E-05	1.2 (1.17-1.23)	2.70E-08	1.14 (1.11-1.17)	1.00E-04	1.14 (1.06-1.22)	0.088	1.23 (1.19-1.27)	1.05	0.59	0.97 (0.94-1.00)	0.91	0.71	0.92 (0.89-0.95)	0.94	0.023	1.12 (1.09-1.15)	1.1	0.97	0.0059	1.18 (1.15-1.21)	1.1
SLC30A7	G	2.50E-10	1.12 (1.09-1.15)	7.43	3.70E-10	1.13 (1.10-1.16)	7.43	0.00047	1.16 (1.13-1.19)	1.70E-07	1.13 (1.10-1.16)	0.042	1.07 (0.99-1.15)	0.57	0.99 (0.96-1.02)	1.01	0.095	1.09 (1.06-1.12)	0.99	0.013	1.18 (1.15-1.21)	0.91	0.57	0.99 (0.96-1.02)	1.01	0.095	1.09 (1.06-1.12)	1.0	
EXTL2	A	4.00E-08	1.09 (1.06-1.12)	4.52	3.70E-07	1.1 (1.07-1.13)	4.52	0.00096	1.14 (1.11-1.17)	6.00E-05	1.08 (1.05-1.11)	0.017	1.08 (1.01-1.15)	0.025	1.11 (1.08-1.14)	1.1	0.088	1.09 (1.06-1.12)	1.01	0.45	1.01 (0.98-1.04)	0.88	0.025	1.11 (1.08-1.14)	1.1	0.88	1.09 (1.06-1.12)	1.0	

Discovery and overall data as on web page

Evidence for the same effect direction was seen separately in both arms of the discovery...

...and in the combined replication...

...and in most of the individual replication studies.

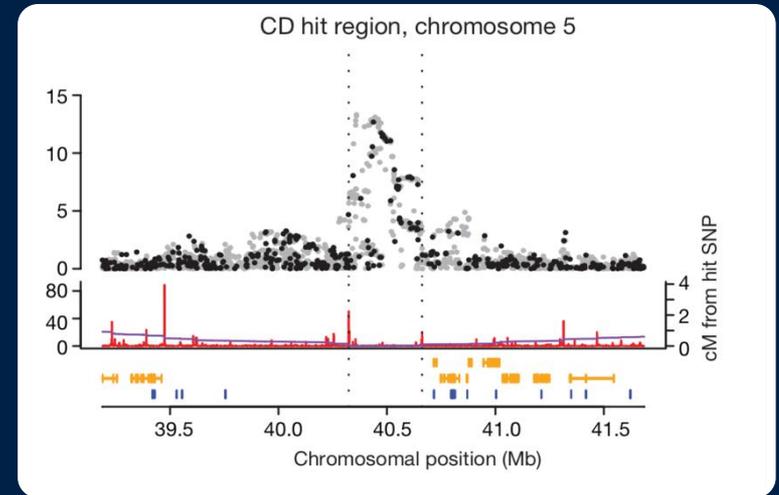
This is a common analysis approach: to gain sample size, use meta-analysis to combine results across several component studies. Then look for consistency between the studies.

$$v_{meta} = 1 / \left(\sum_i \frac{1}{v_i} \right) \quad \beta_{meta} = \left(\sum_i \frac{\beta_i}{v_i} \right) \times v_{meta} \quad (\text{Where } v \text{ denotes squared standard error})$$

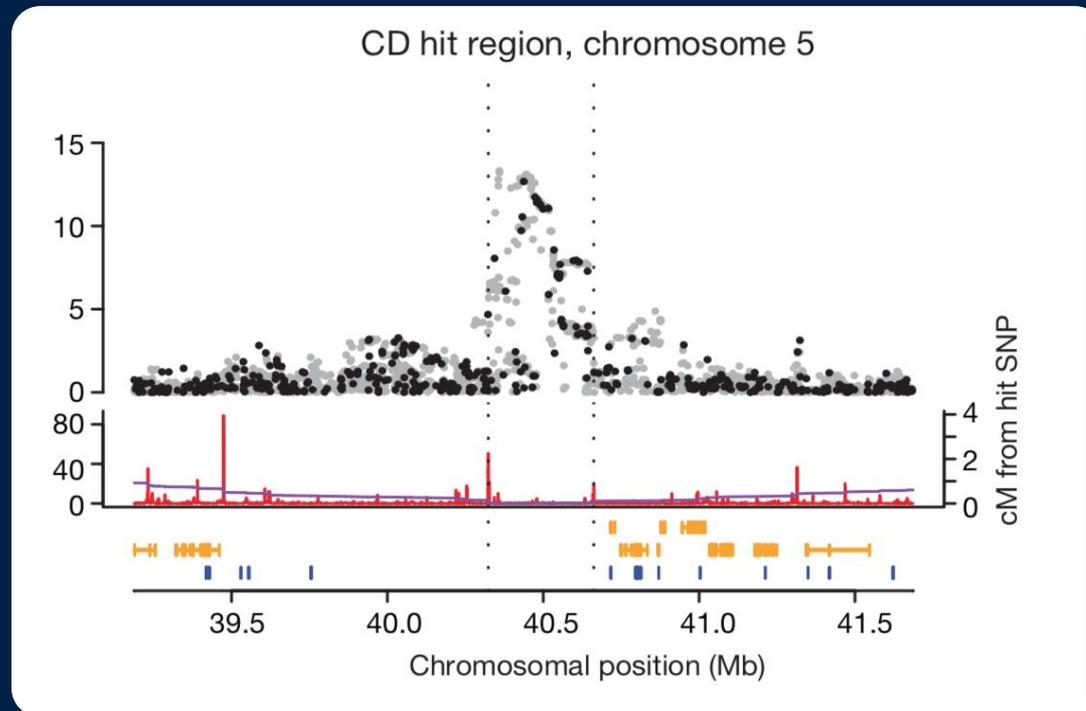
"Inverse variance weighted fixed-effect meta-analysis", gives results approximately equal to joint analysis of genotype data.

Main points in this lecture

Extracting biological information
from GWAS



We now have thousands of GWAS signals across thousands of traits. What do they teach us about the underlying biology?

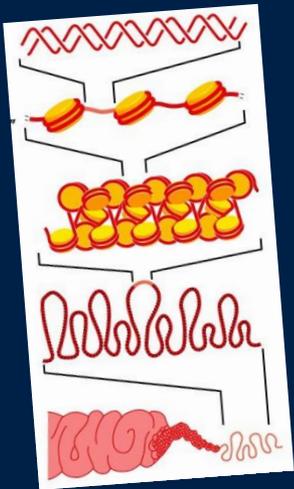


?

The circle of genetic causation



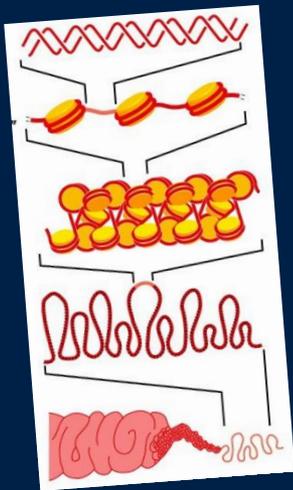
DNA gets physically
packaged up into
chromosomes...



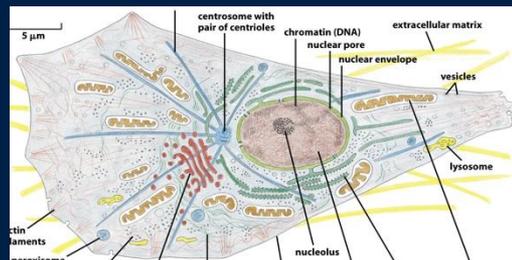
The circle of genetic causation



DNA gets physically packaged up into chromosomes...



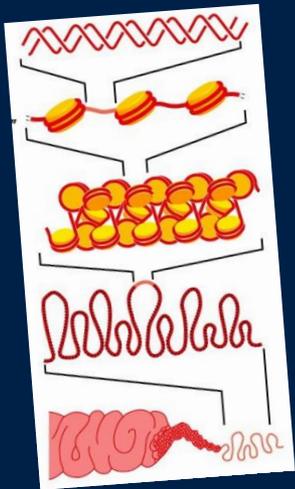
...inside cells, where it is **transcribed** to form proteins and other molecules...



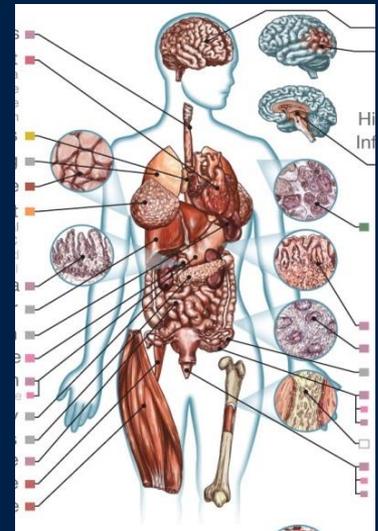
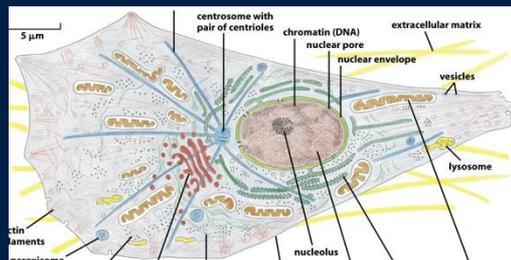
The circle of genetic causation



DNA gets physically packaged up into chromosomes...



...inside cells, where it is **transcribed** to form proteins and other molecules...



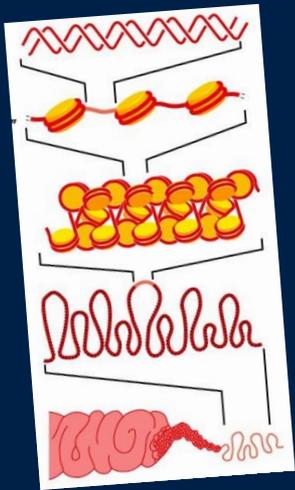
...that combine to make individuals...

...that affect how the cells behave, forming different organs...

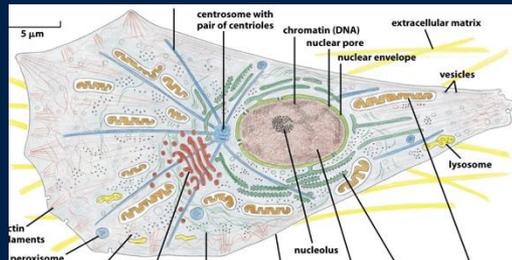
The circle of genetic causation



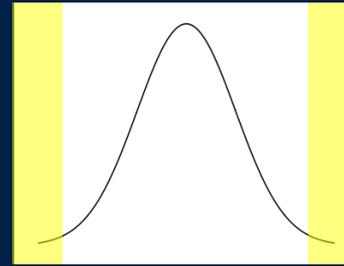
DNA gets physically packaged up into chromosomes...



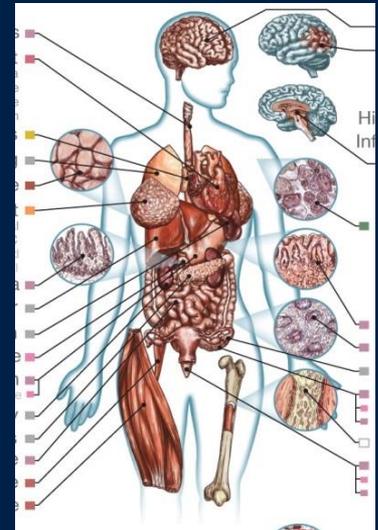
...inside cells, where it is transcribed to form proteins and other molecules...



...that affect how the cells behave, forming different organs...



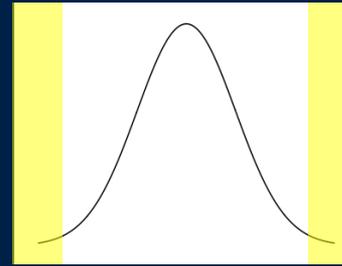
...whose success is affected by the traits they have...



...that combine to make individuals...

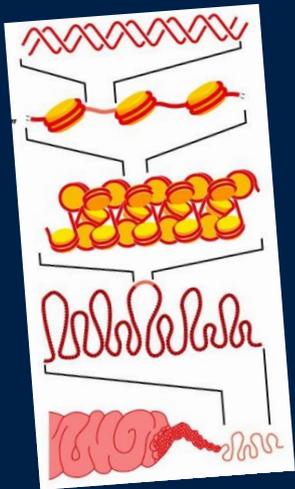
The circle of genetic causation

...passing on DNA, with **mutations** and **recombination**, to new generations...

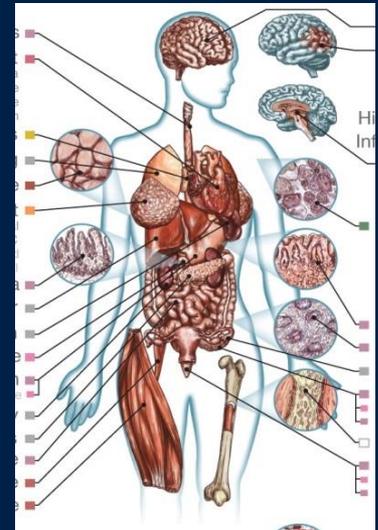


...whose success is affected by the traits they have...

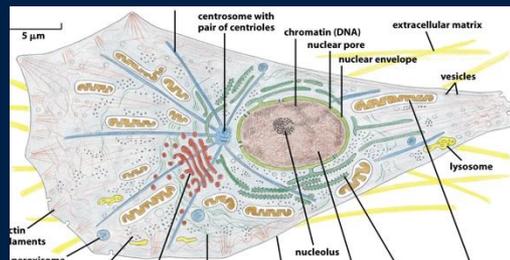
...that gets physically packaged up into chromosomes...



There is complex biology at all stages



...inside cells, where it is **transcribed** to form proteins and other molecules...

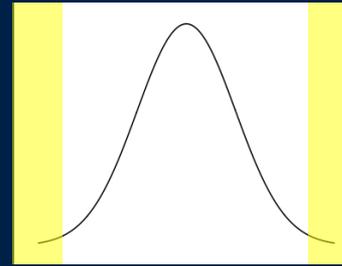


...that affect how the cells behave, forming different organs...

...that combine to make individuals...

The circle of genetic causation

...passing on DNA, with **mutations** and **recombination**, to new generations...



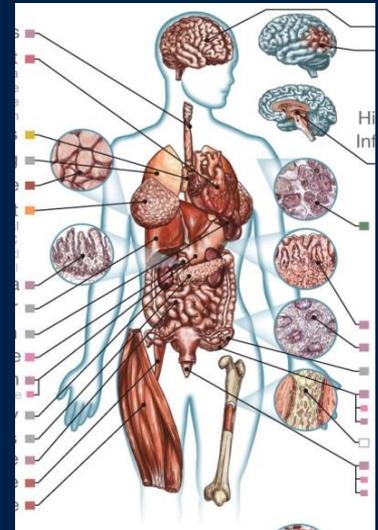
...whose success is affected by the traits they have...

...that gets physically packaged up into chromosomes...

*microarrays,
genome sequencing*

*Clinical phenotype
measurements*

*Biomarker
measurements*



There is **complex biology** at all stages and we can measure it

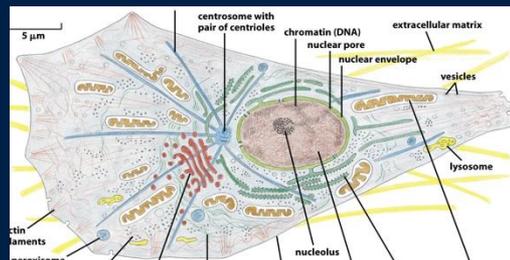
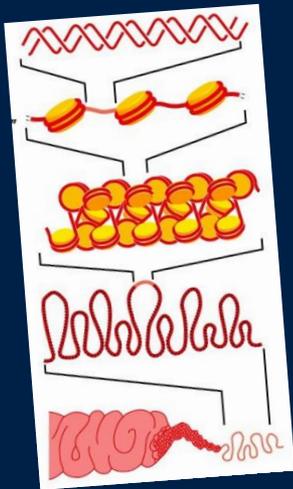
*Chromatin state
marker assays,
ChIP-seq, ...*

*RNA-seq,
spectroscopy, antibody
binding*

...that combine to make individuals...

...inside cells, where it is **transcribed** to form proteins and other molecules...

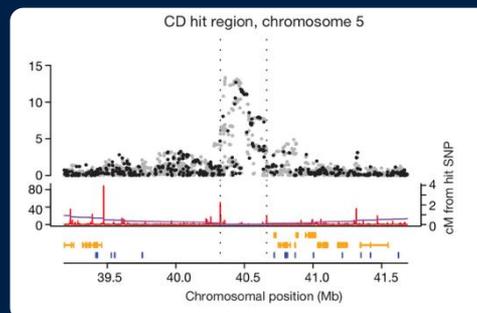
...that affect how the cells behave, forming different organs...



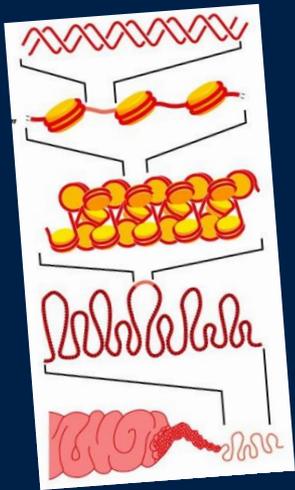
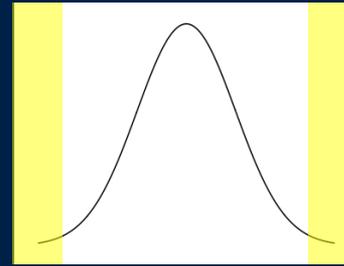
Gaining biological knowledge from GWAS

There are several ways we can try to translate knowledge of associations into new biological insights. I will try to describe a few of these.

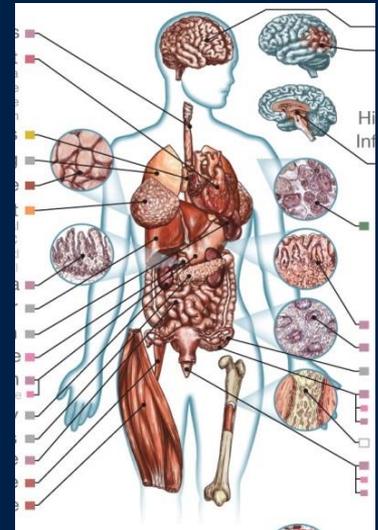
- **Pathway analysis** – even if we can't fine-map, we can still try to assess whether associations group into particular biological pathways that might shed light on biology
- **Pleiotropy** – how are associations shared between traits?
- **Fine-mapping** – can we identify the actual causal variants underlying these associations, and hence discover specific proteins and disease pathways?



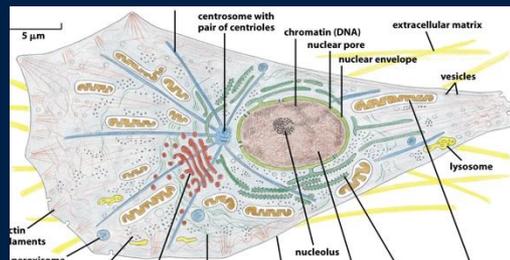
The circle of genetic causation



Pathway analysis
and pleiotropy

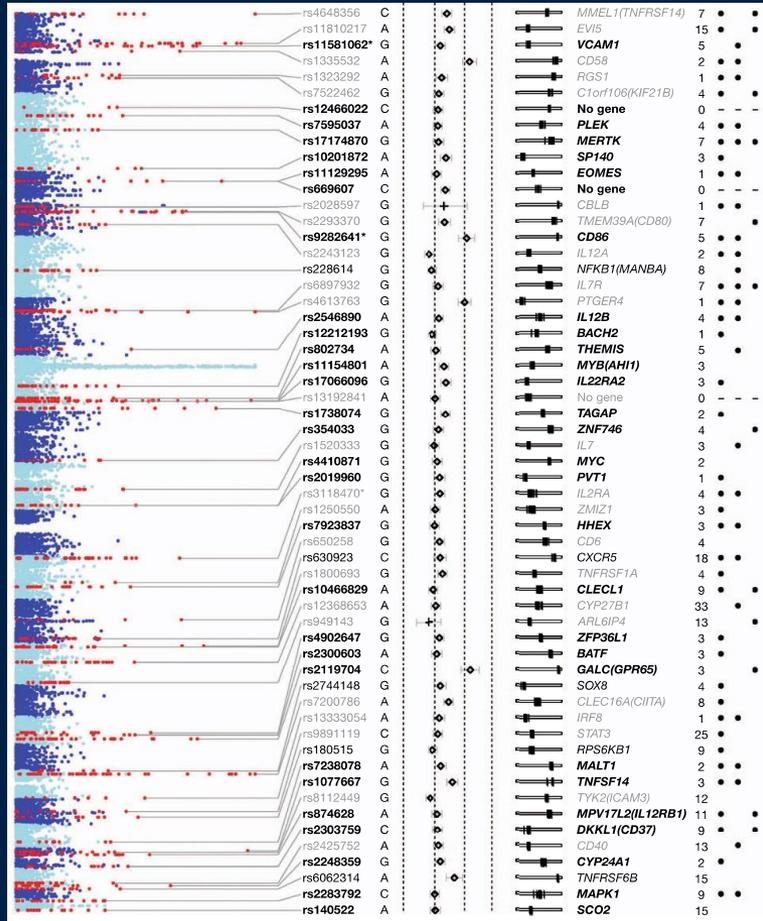


...that combine to make
individuals...



Pathway analysis example

The primary cause of MS has typically been thought to be inflammation causing downstream neurodegeneration – with some debate about this. Can the GWAS of MS we discussed shed light on this?



Clinical and Experimental Neuroimmunology 1 (2010) 2–11

REVIEW ARTICLE

What drives disease in multiple sclerosis: Inflammation or neurodegeneration?

Hans Lassmann

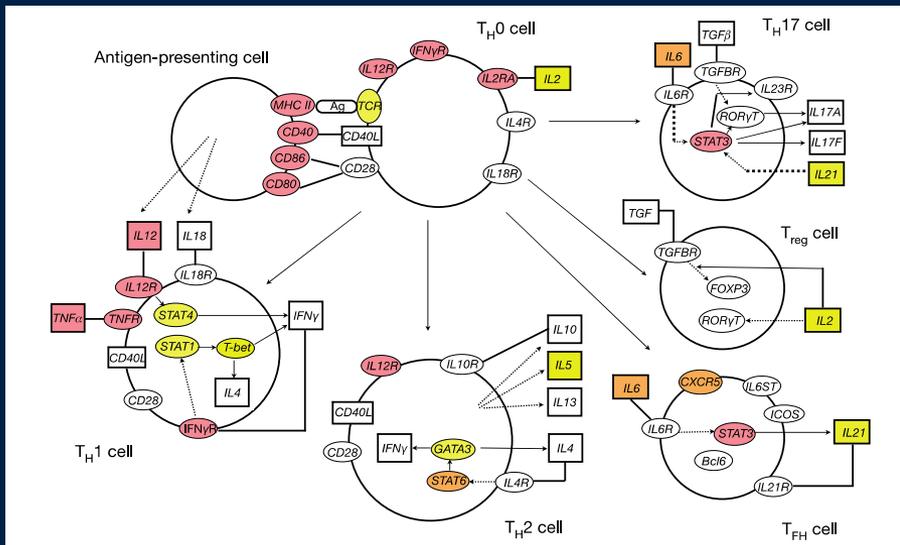
Center for Brain Research, Medical University of Vienna, Vienna, Austria

As the main figure shows, many of the association signals looked like they were near immune-system related genes.

Pathway analysis example

We:

- Assigned SNPs to their nearest gene using the available annotation
- Used the Gene Ontology Project to classify genes into functionally related groups
- Conducted a statistical test (Fisher's exact test) to identify whether the nearest genes were enriched in each group.



T-helper-cell differentiation pathway
(from Ingenuity Pathway Analysis software)

Particularly strong enrichment was observed for immune system pathways – notably in “T cell activation and proliferation” ($P=1.9 \times 10^{-9}$)

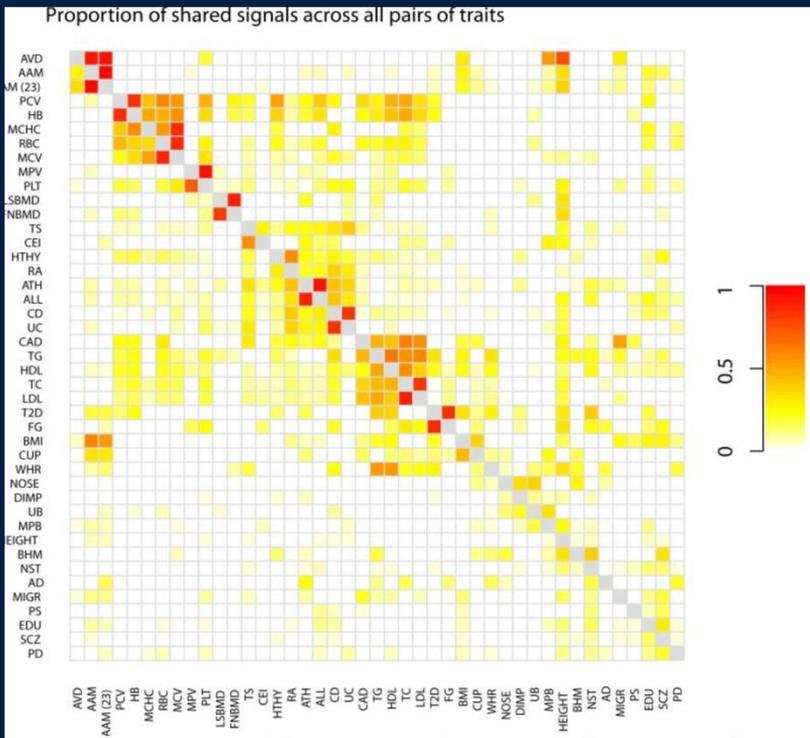
“Although GO immune system genes only account for 7% of human genes, in 30% of our association regions the nearest gene to the lead SNP is an immune system gene”

Published: 10 August 2011

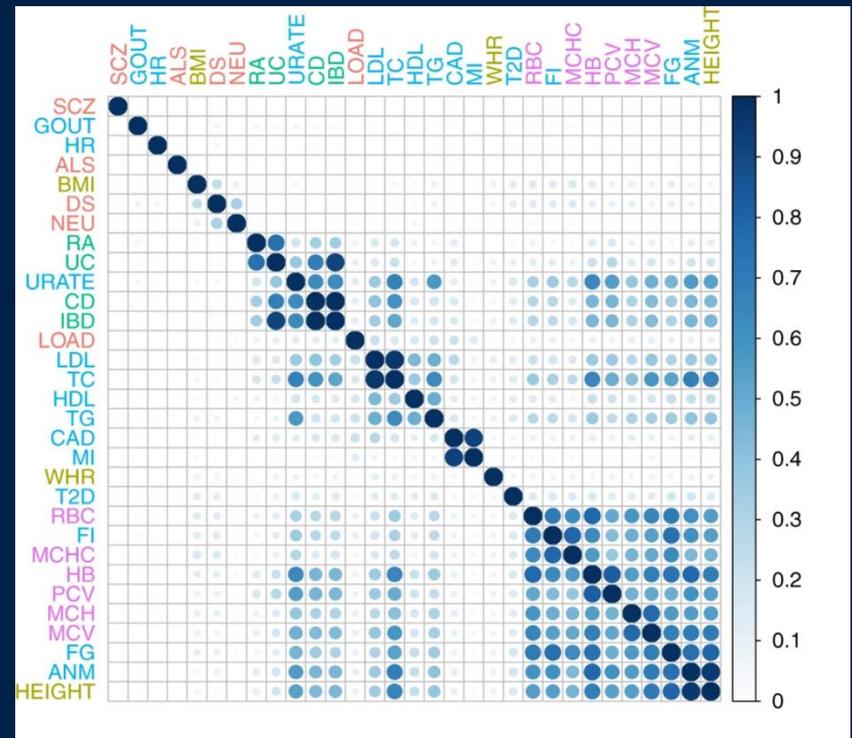
Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis

The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium

2



Sharing of associations between traits
 ("proportion of associations for trait i (row) that are also associations for trait j (column)")



Sharing of *pathways* between traits
 ("proportion of pathways that are enriched in both traits, among pathways enriched in at least one of the traits")

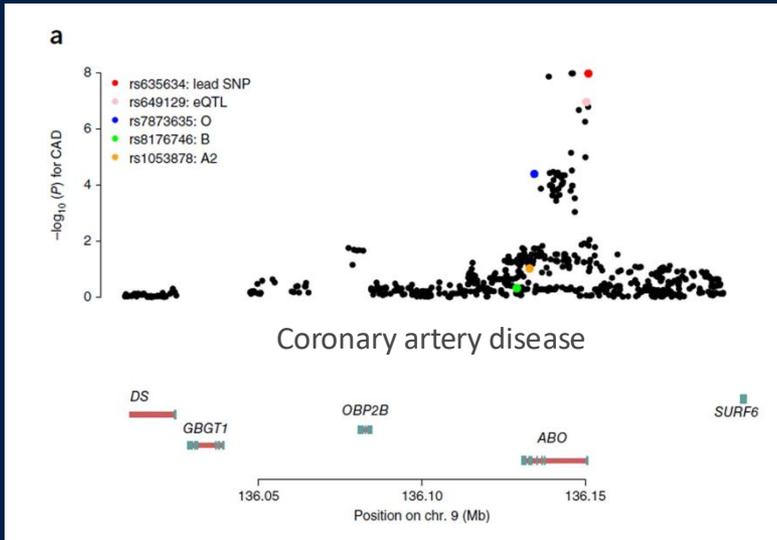
Detection and interpretation of shared genetic influences on 42 human traits

Joseph K Pickrell^{1,2}, Tomaz Berisa¹, Jimmy Z Liu¹, Laure Séguérel¹, Joyce Y Tung⁴ & David A Hinds⁴

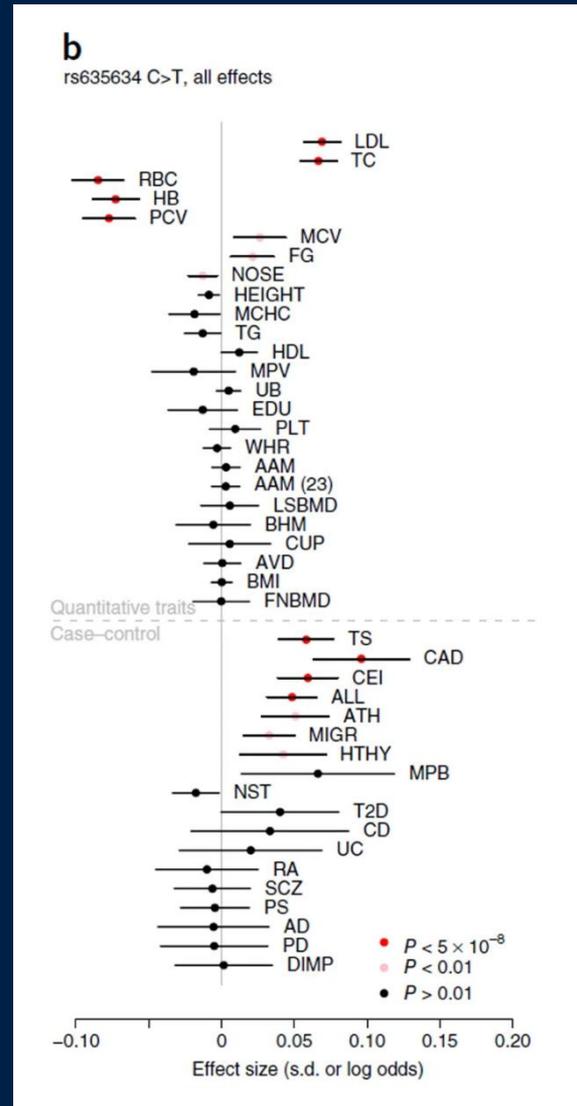
Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes

Xiang Zhu^{1,2} & Matthew Stephens^{2,3}

From GWAS to PheWAS



GWAS
= one trait, across SNPs



PheWAS
= one SNP,
across traits

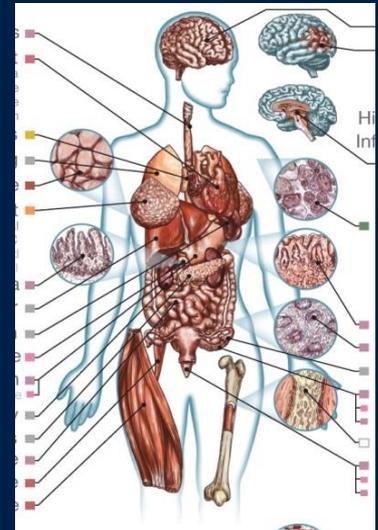
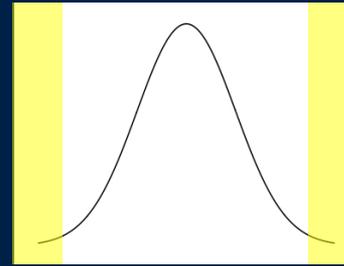
Fine-mapping

“Fine-mapping” is the general term used for attempts to narrow down association signals to the underlying causal variants. A typical process involves:

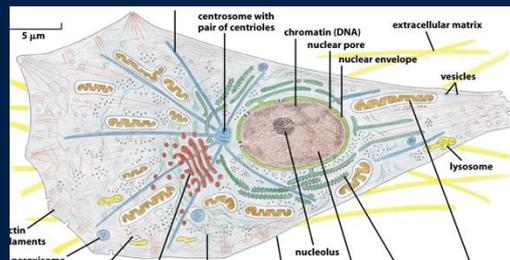
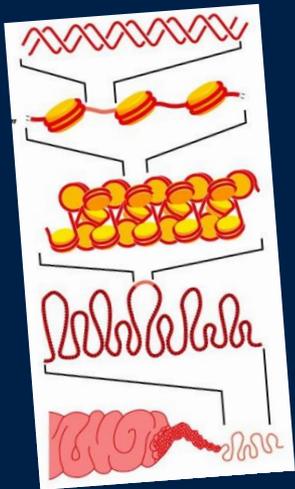
- Gathering complete information on genetic variation in the region of interest – for example by deep-sequencing a large number of individuals. (Large databases such as gnomAD / TopMed now make this easier.)
- Gathering information on genome function – including gene structure and regulatory regions.
- Potentially leveraging data from different ancestral backgrounds, hoping that differences in LD patterns will help narrow down signals.
- Fitting models that attempt to parse apart multiple associations in the same region

Possible underlying mechanisms are pretty diverse and a healthy dose of genomic detective work is often needed.

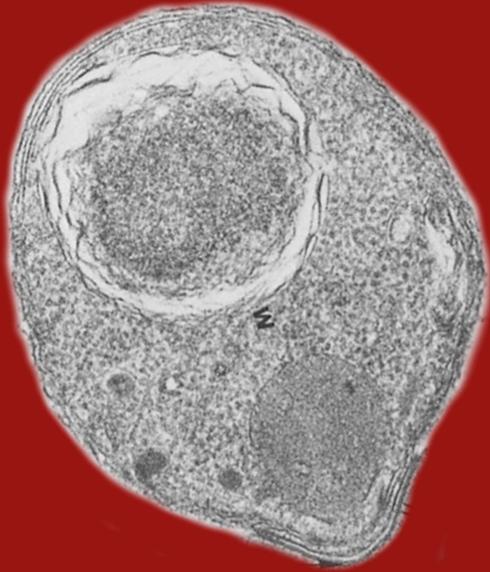
The circle of genetic causation



Fine-mapping example
Complex genetic variation

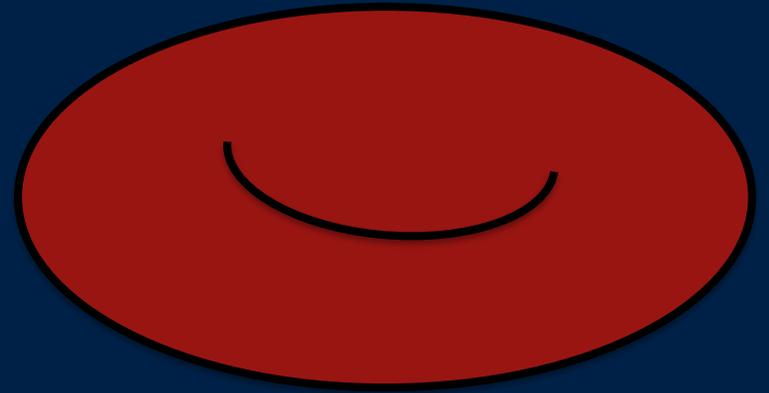


...that combine to make
individuals...



Plasmodium falciparum

VS

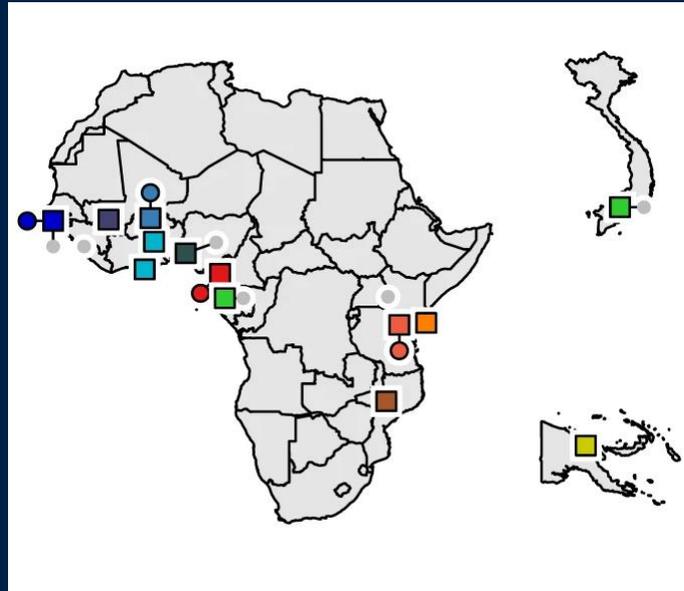


humans

GWAS of susceptibility to severe malaria

Study samples

Group	Cases	Controls	TOTAL
Africa			
■ Gambia	2567	2605	5172
■ Mali	274	183	457
■ Burkina Faso	733	596	1329
■ Ghana	399	320	719
■ Nigeria	113	22	135
■ Cameroon	592	685	1277
■ Malawi	1182	1317	2499
■ Tanzania	416	403	819
■ Kenya	1681	1615	3296
Asia			
■ Vietnam	718	546	1264
Oceania			
■ PNG	402	374	776



a

Whole-genome sequences

Group	Trios	Duos	Other	TOTAL
● Gambia				
FULA	31	1	5	100
JOLA	32	1	2	100
MANDINKA	33	0	1	100
WOLLOF	32	1	3	98
● Burkina Faso				
MOSSI	0	0	57	57
● Cameroon				
BANTU	5	3	11	31
SEMIBANTU	8	0	7	32
● Tanzania				
CHAGGA	21	2	13	80
PARE	22	2	7	77
WASAAMBA	23	6	9	90

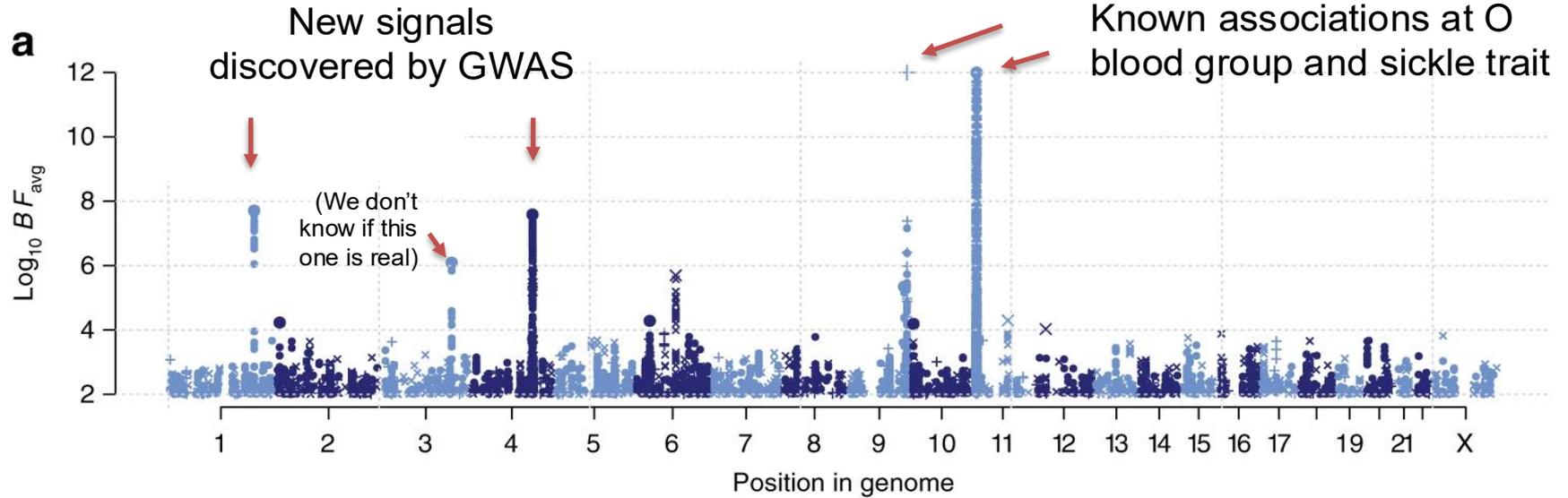
GWAS in 17,000 severe malaria cases and population controls
 From 12 sites in Africa, Oceania, and SE Asia.
 Genotyped on the Illumina Omni 2.5M array

+ whole-genome sequences
 for imputation

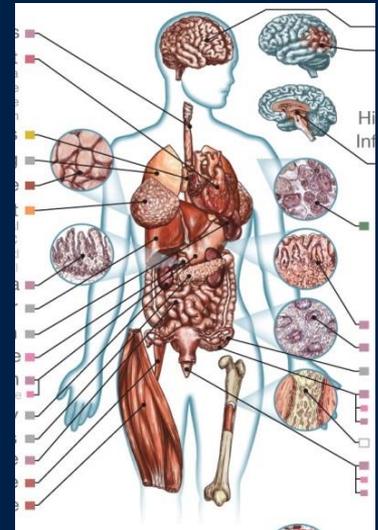
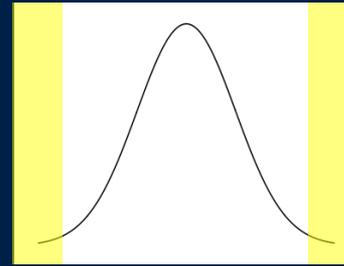
Malaria Genomic Epidemiology Network. *“Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania”*.

Nature Communications (2019). <https://doi.org/10.1038/s41467-019-13480-z>

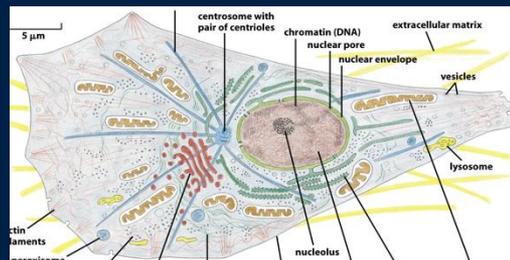
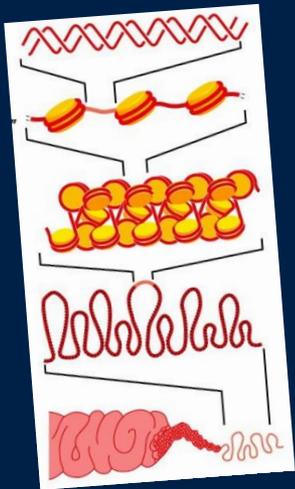
Natural resistance is driven by red blood cell variation



The circle of genetic causation

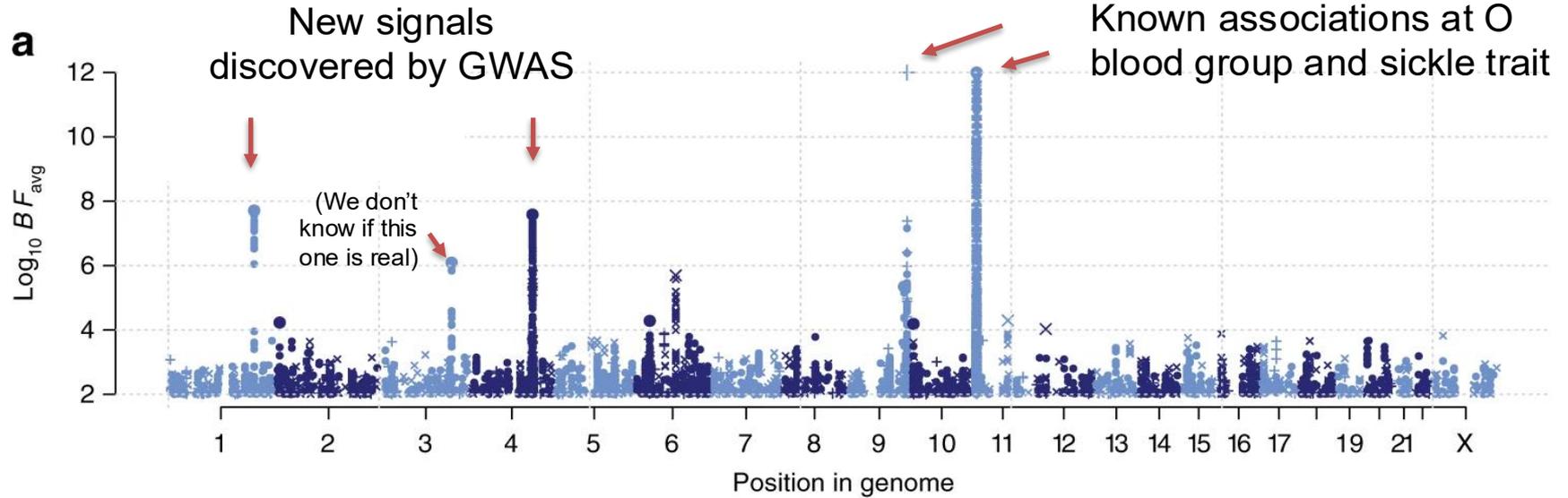


Fine-mapping example 2
Cell-specific gene regulation

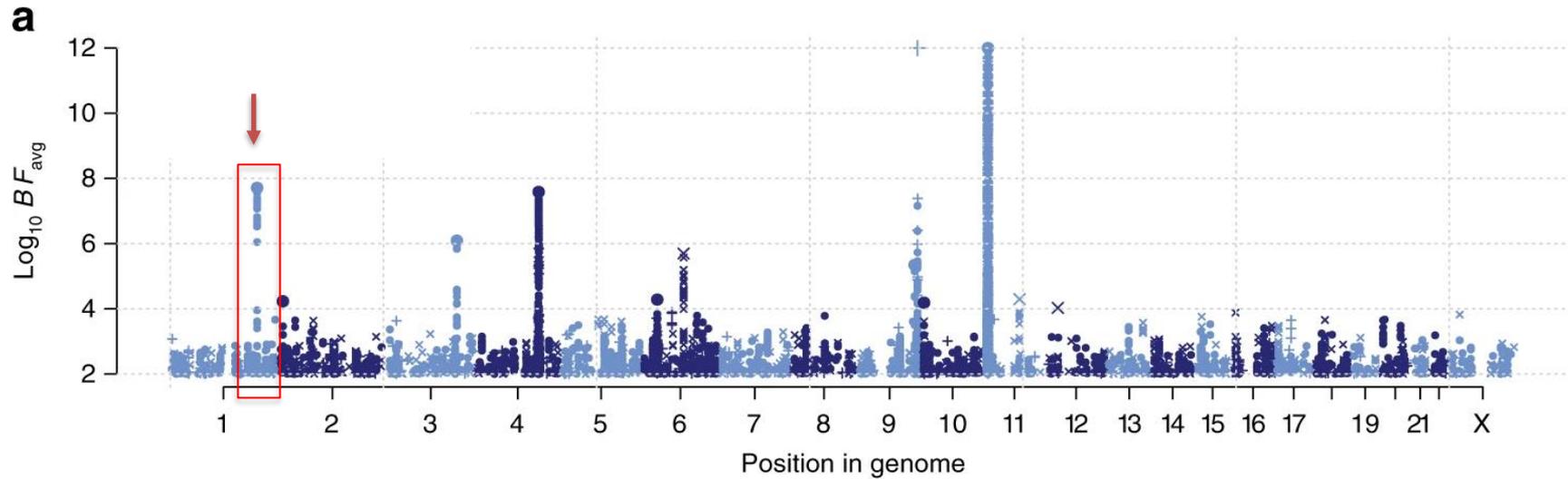


...that combine to make
individuals...

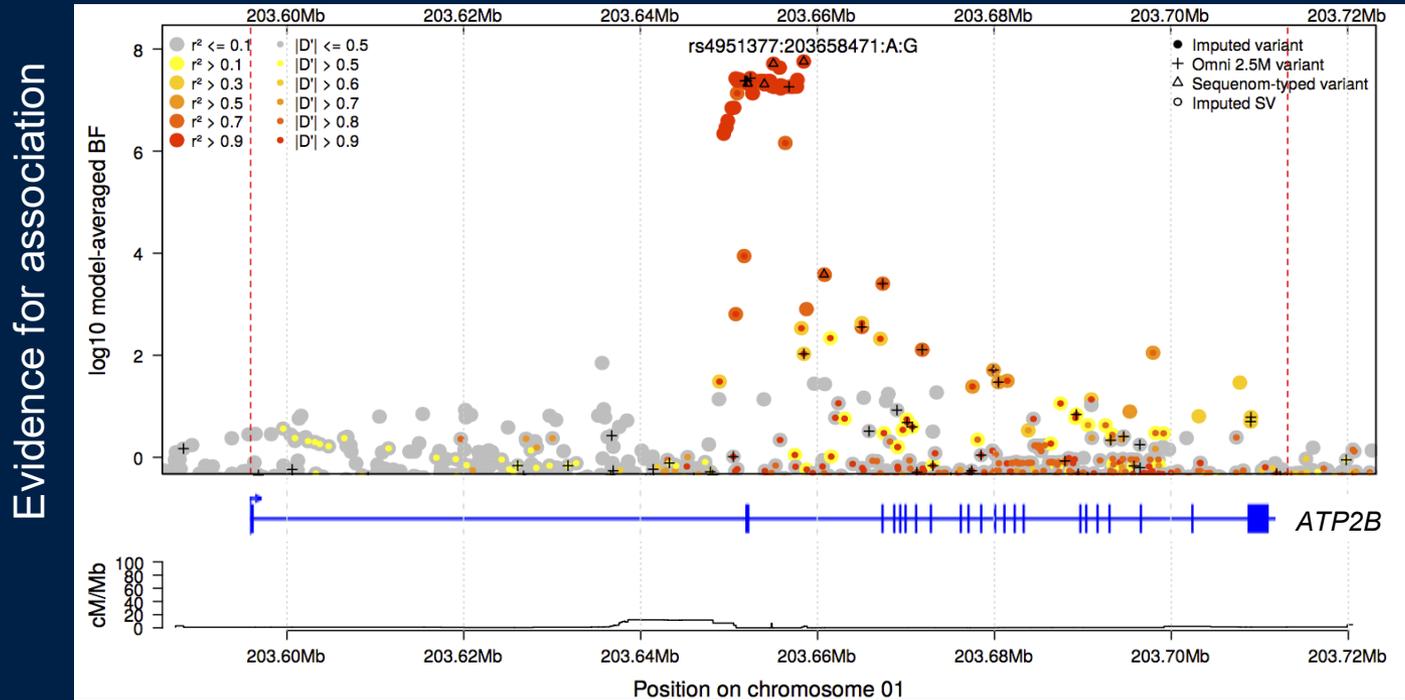
Natural resistance is driven by red blood cell variation



Natural resistance is driven by red blood cell variation

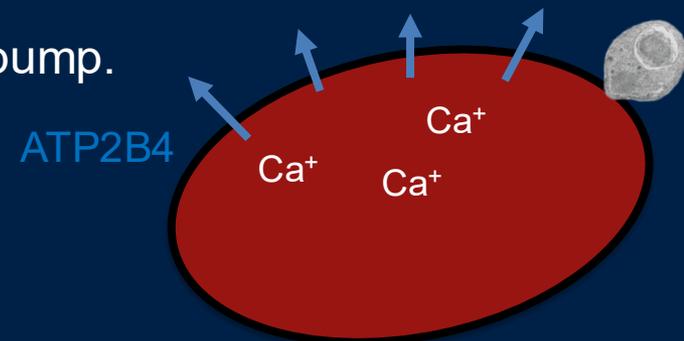


Association near 2nd exon of *ATP2B4*

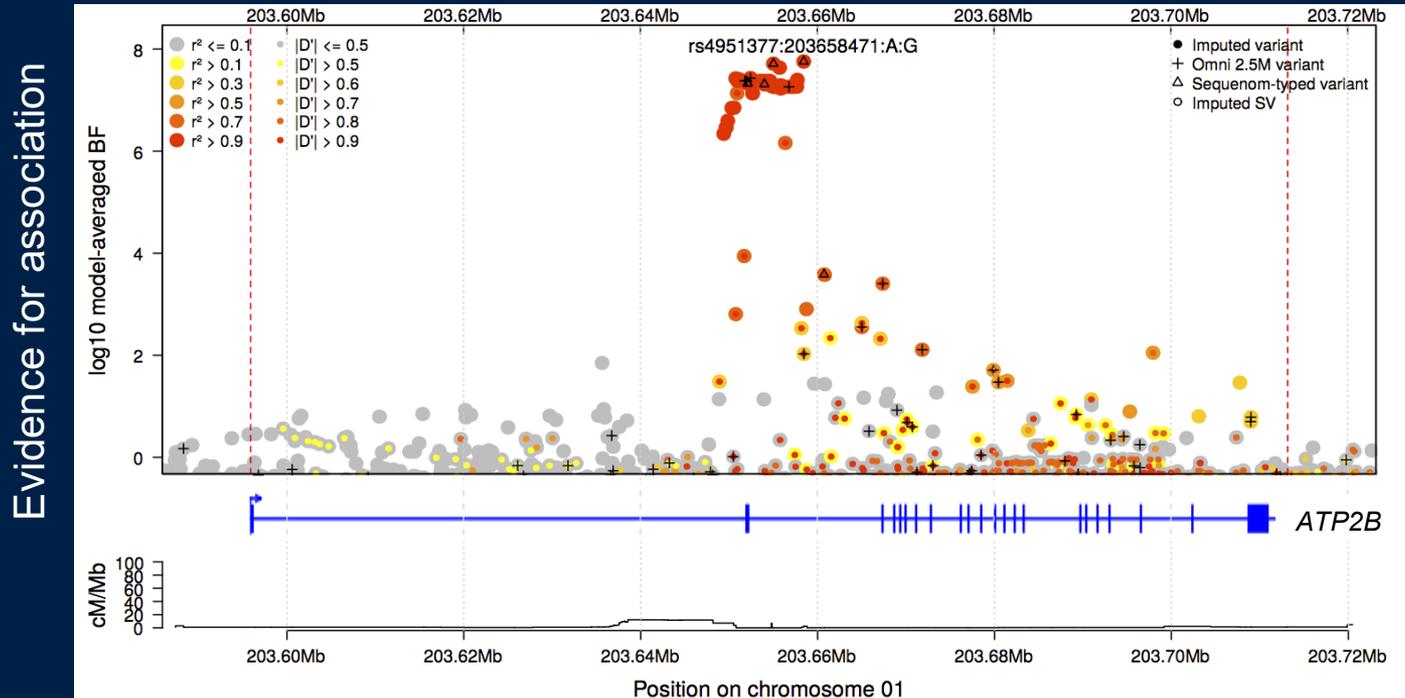


The associated SNPs cover a region around the second exon.

In a gene called *ATP2B4*, which encodes a calcium pump.



Association near 2nd exon of *ATP2B4*

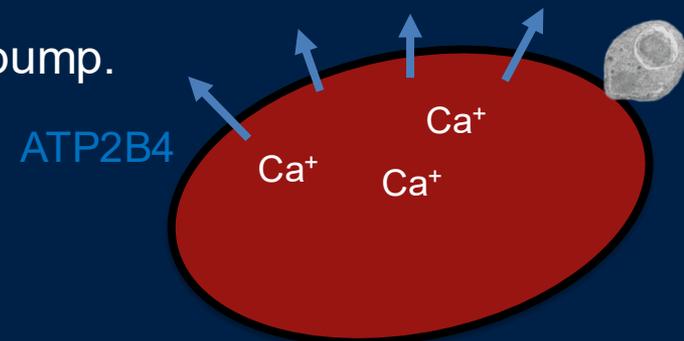


The associated SNPs cover a region around the second exon.

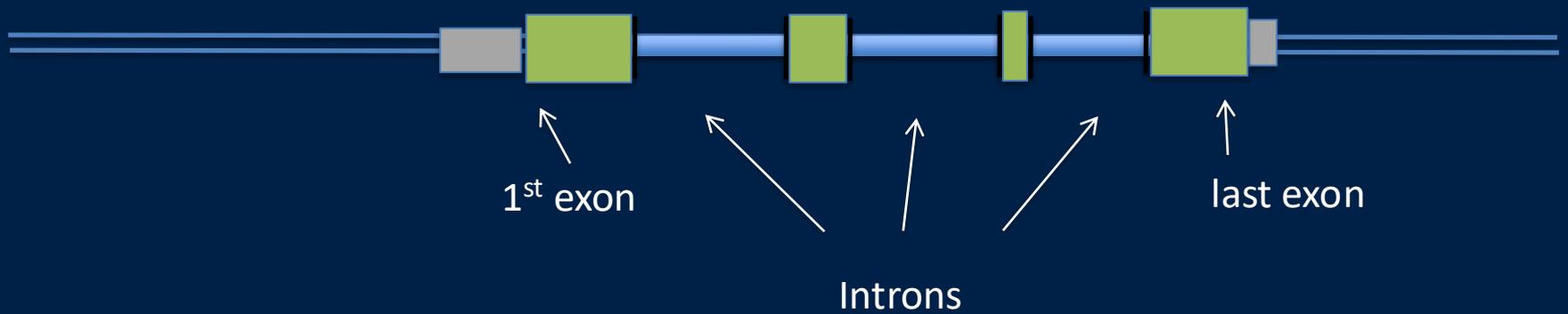
In a gene called *ATP2B4*, which encodes a calcium pump.

None of these SNPs make changes to the protein.

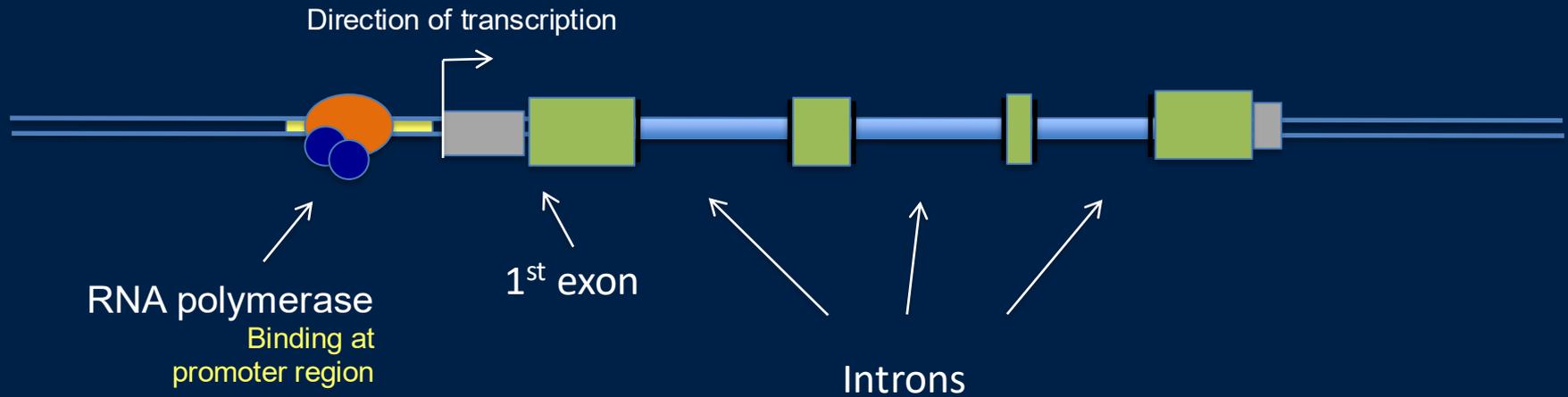
What could be going on?



Cartoon of a gene

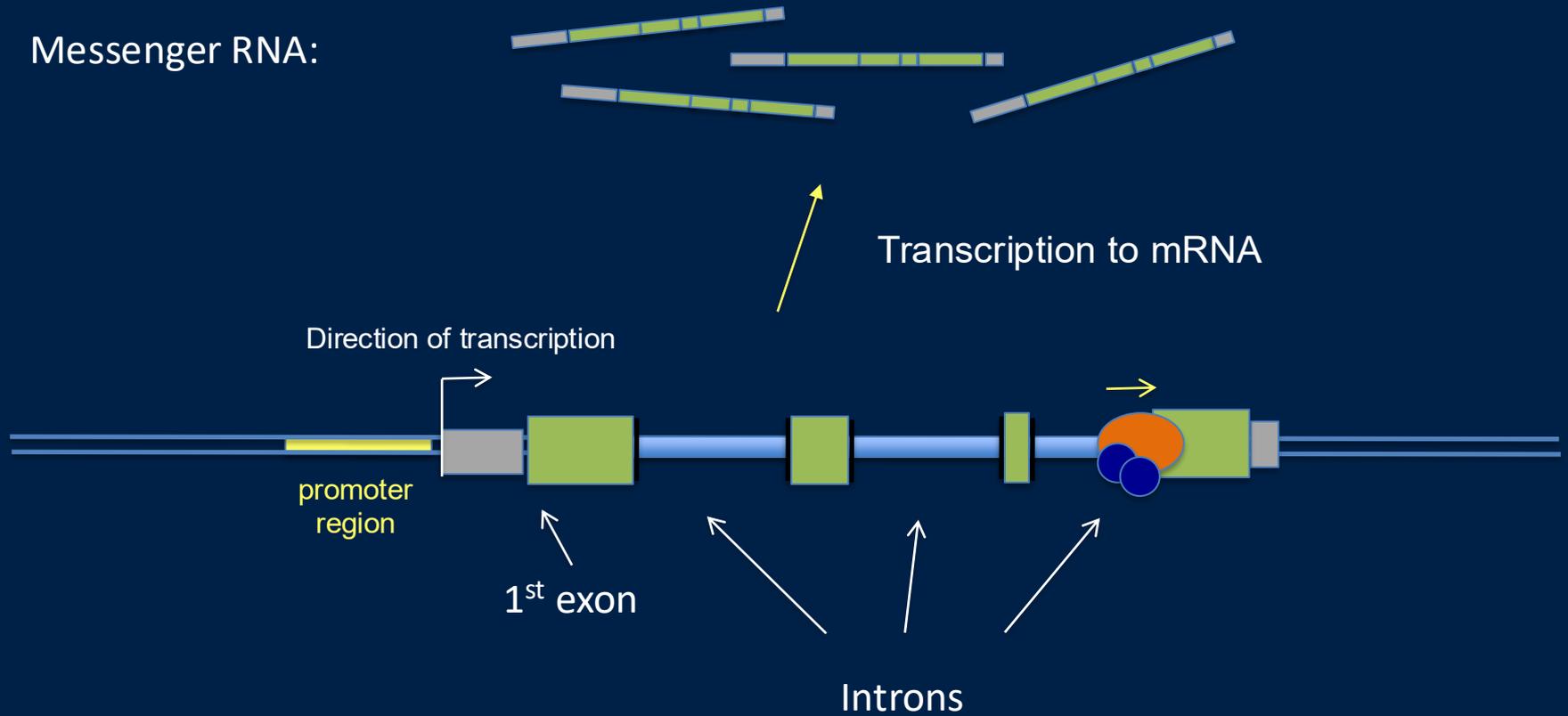


Cartoon of a gene



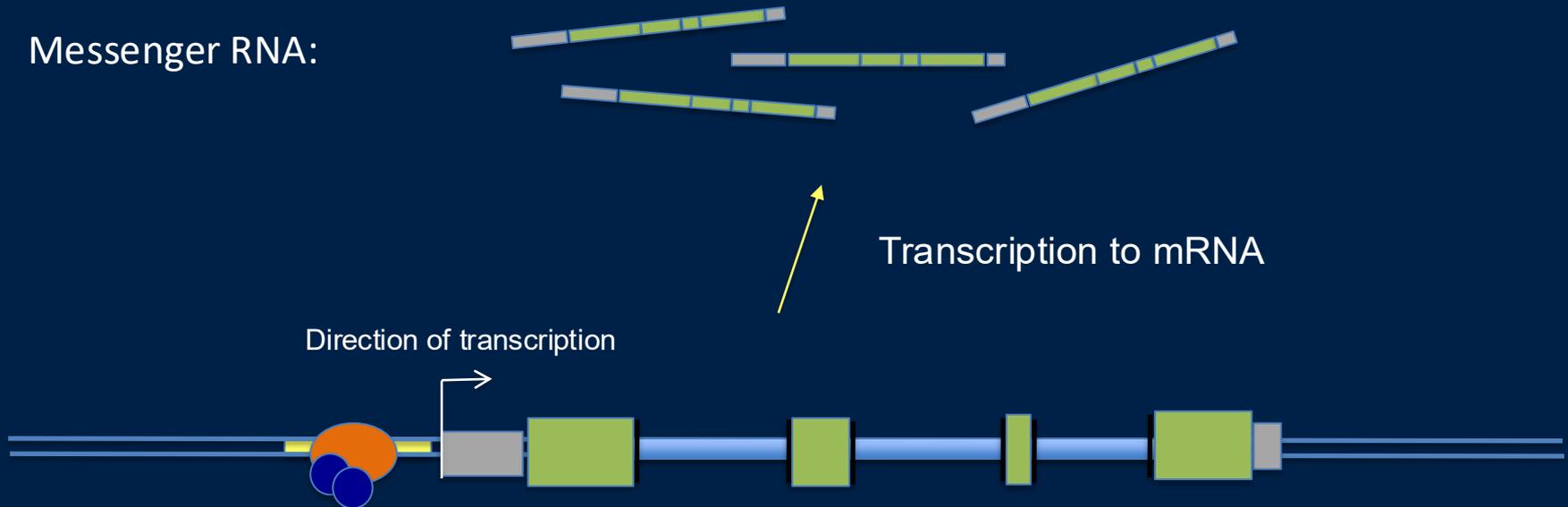
Cartoon of a gene

Messenger RNA:



Cartoon of a gene

Messenger RNA:



What happens at the promoter can really affect transcription.

Cartoon of a gene

Messenger RNA:



Transcription to mRNA



Direction of transcription



“Transcription factors”
must be able to bind.

Cartoon of a gene

Messenger RNA:



Transcription to mRNA



Direction of transcription



DNA must be
“accessible”
(not wound round
nucleosomes)

“Transcription factors”
must be able to bind.

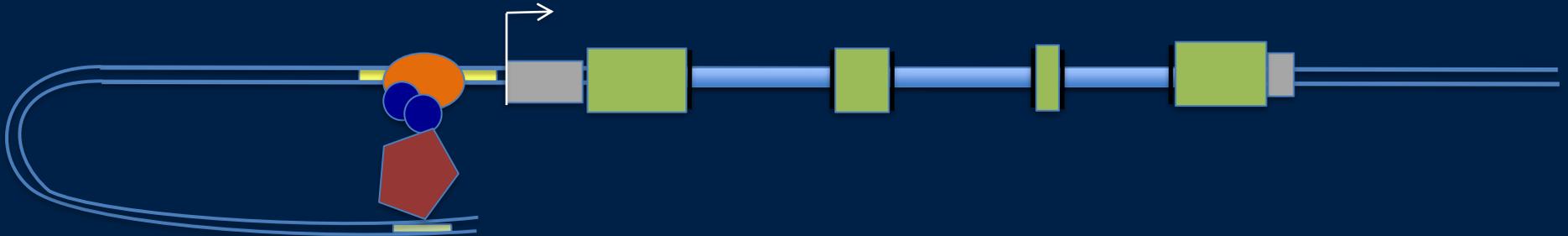
Cartoon of a gene

Messenger RNA:



Transcription to mRNA

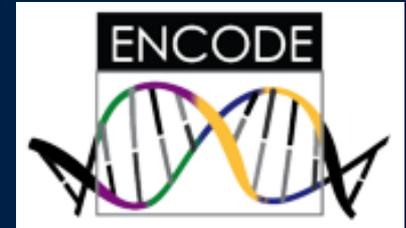
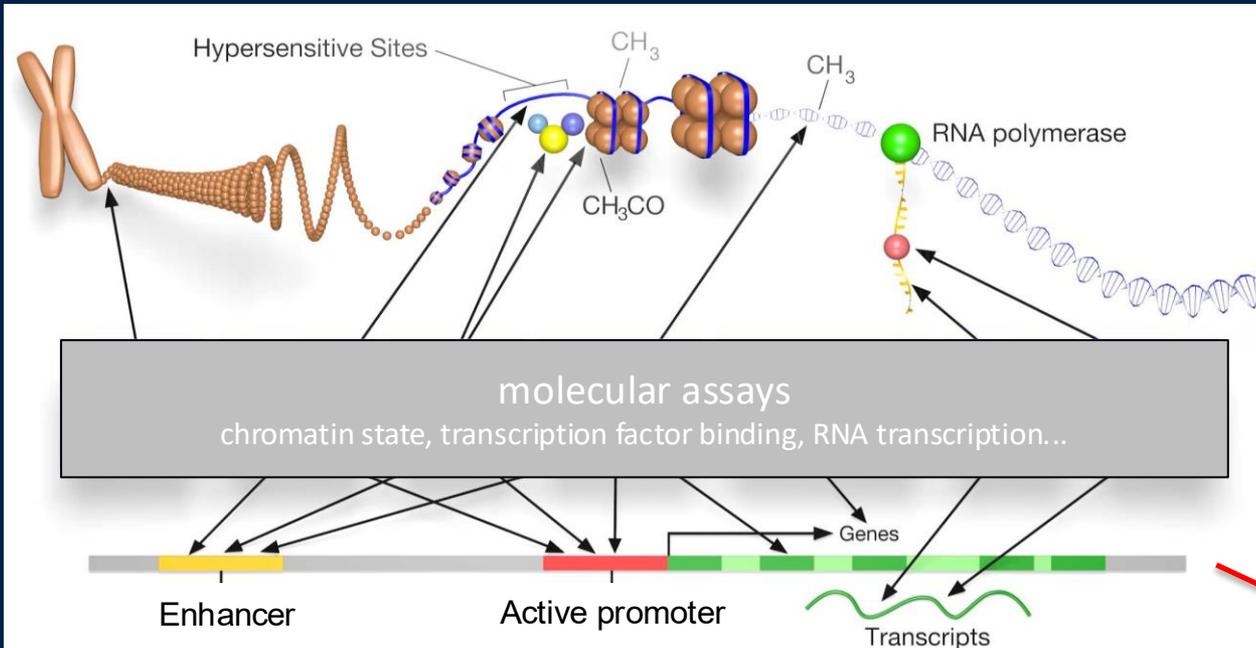
Direction of transcription



“Enhancer” – promoter interactions

All this can be measured

Two ways to look at transcription



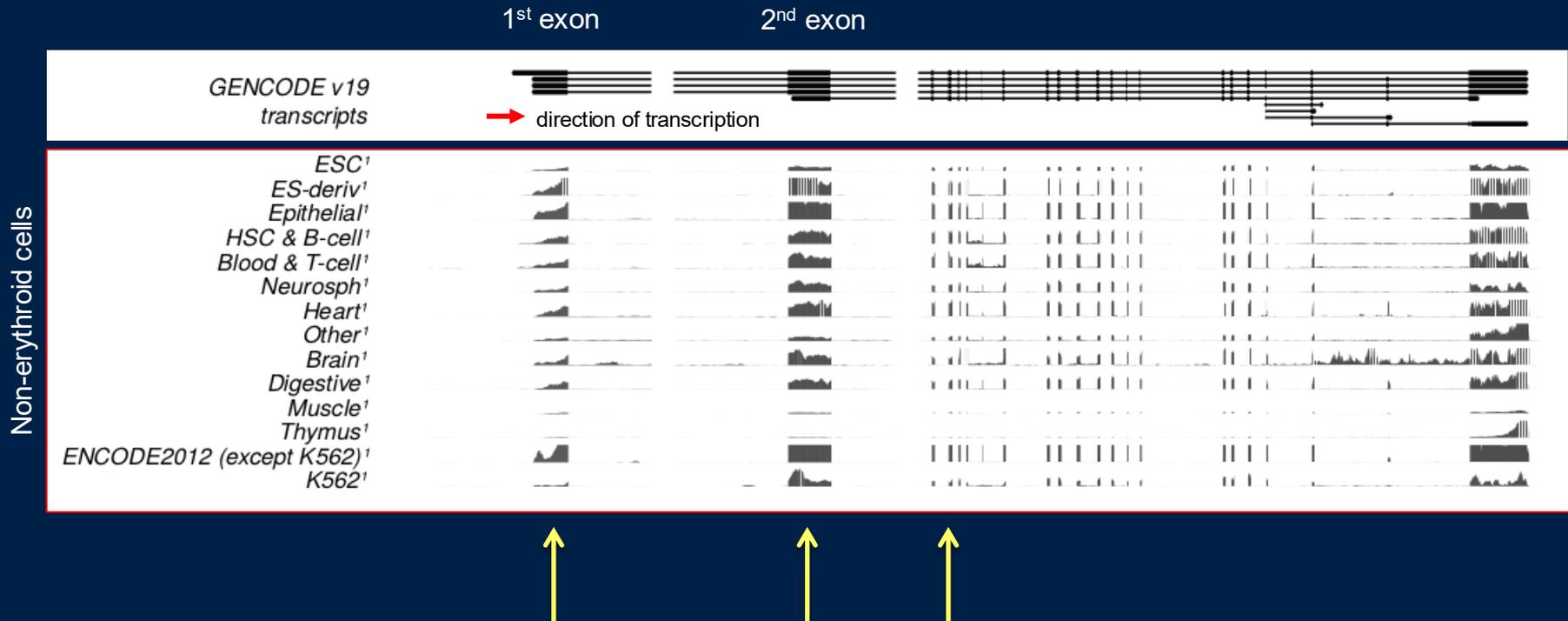
<https://www.encodeproject.org>

Can look at chromatin state

RNA expression



Expression data (RNA-seq) across cell types

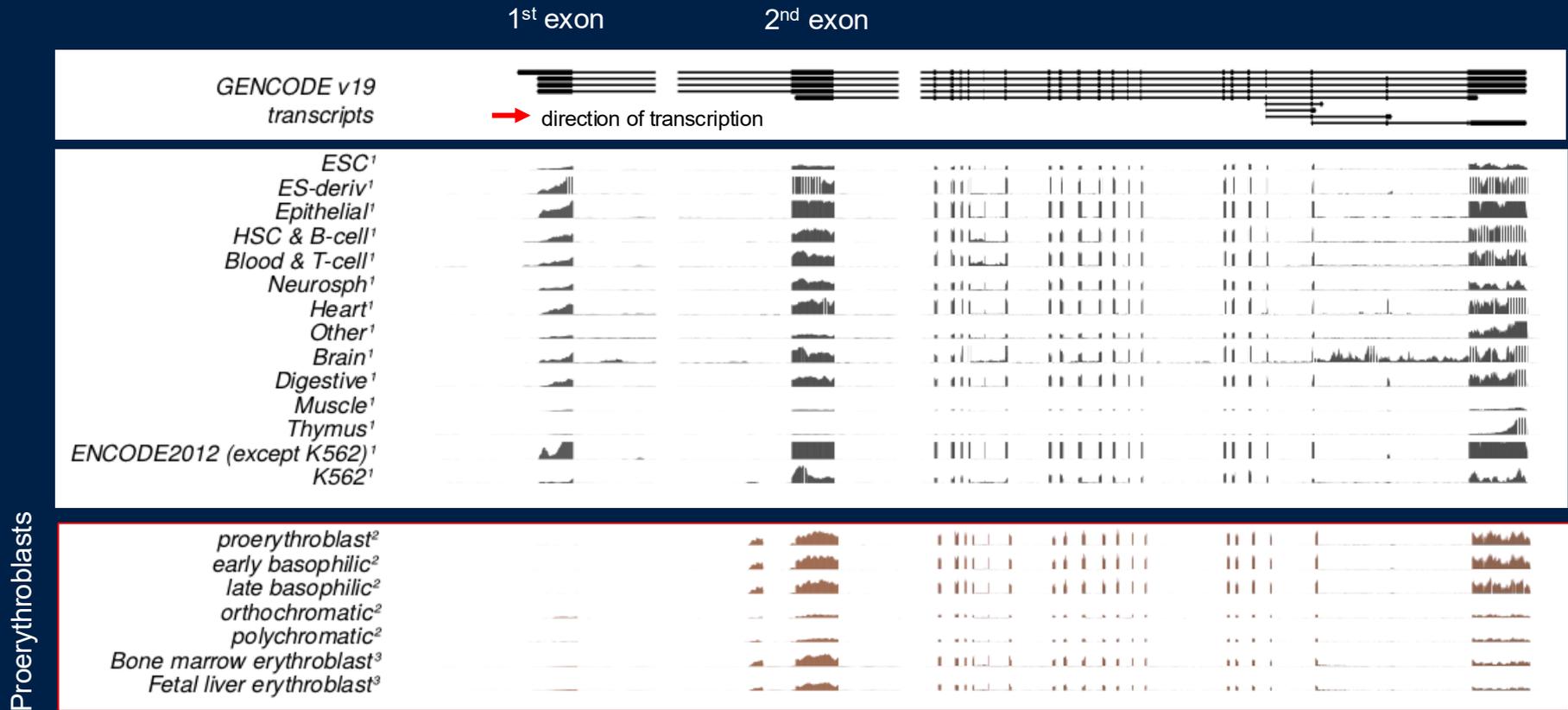


ATP2B4 is expressed essentially in all cells in the body

It looks like the gene model says it should

What if we look in red cells?

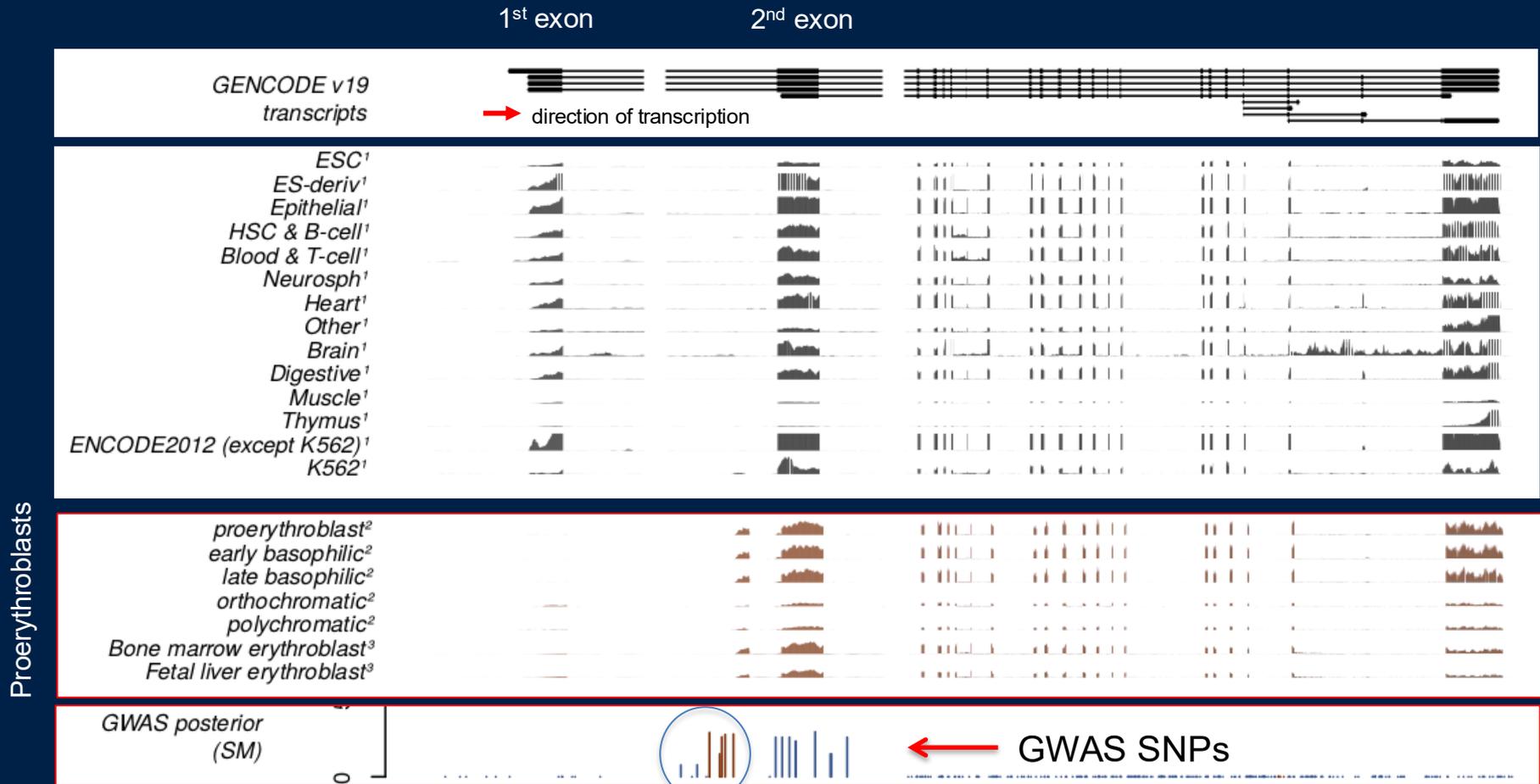
Expression data (RNA-seq) across cell types



Red cell (precursors) show a different pattern

There's a new 1st exon!

Expression data (RNA-seq) across cell types



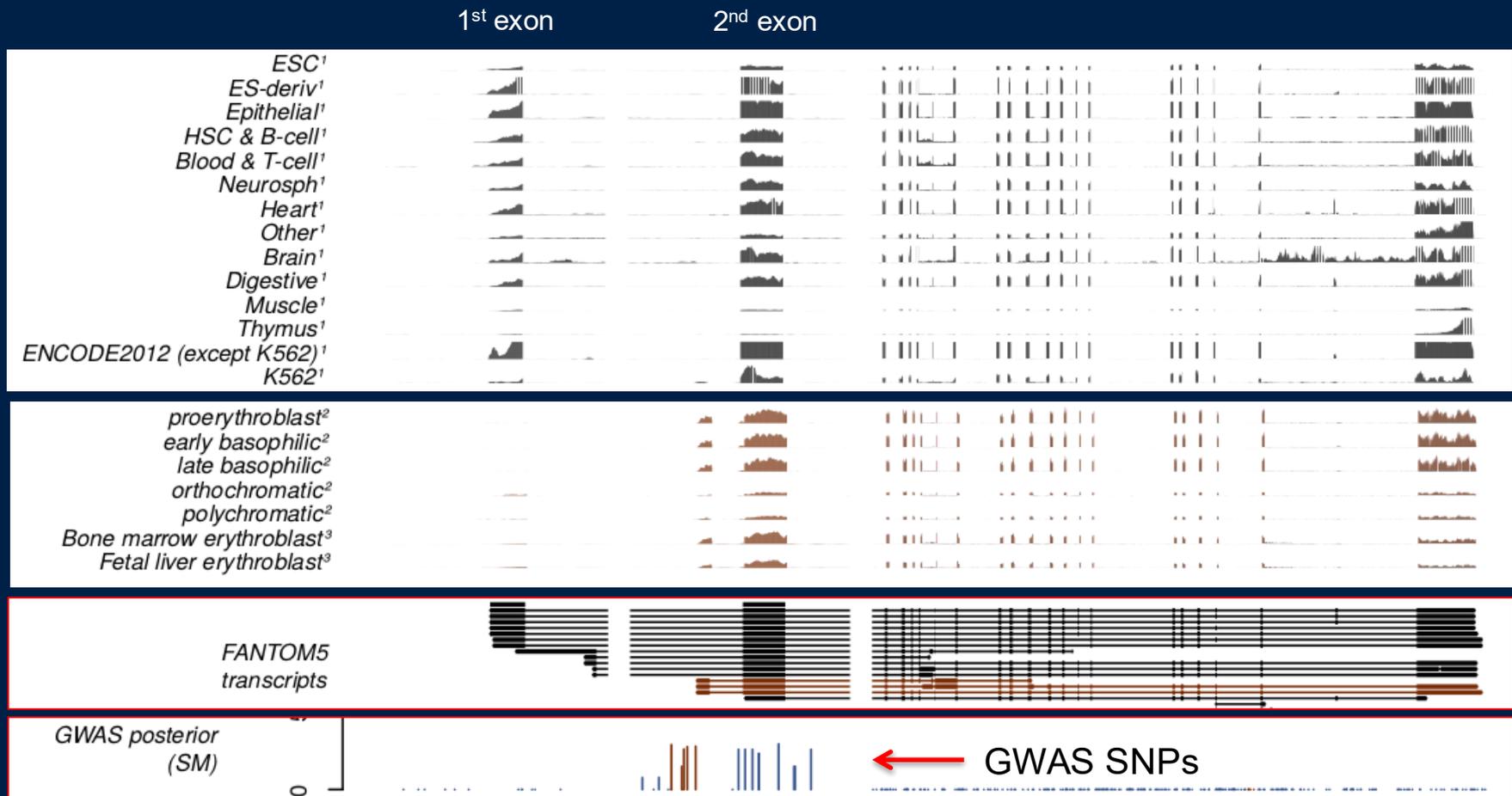
Red cell (precursors) show a different pattern

There's a new 1st exon!

Could these upstream GWAS SNPs be affecting transcription in red cells?

ATP2B4 has an erythroid-specific transcript

Measured RNA transcription (RNA-seq)



Putting together data from a variety of sources suggests the existence of an *alternative transcription start site* near the GWAS signal, but only active in erythrocytes. How can this be?

What is different about RBCs?



GATA1

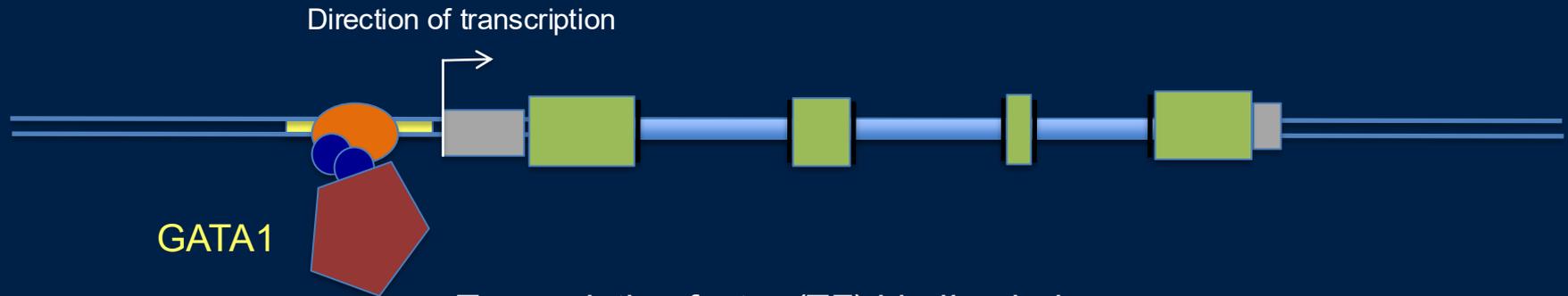
Transcription factor (TF) binding helps promote transcription

A major TF in red cells is **GATA1**.

“GATA1 regulates the expression (i.e. formation of the genes' products) of an ensemble of genes that mediate the development of red blood cells and platelets.”

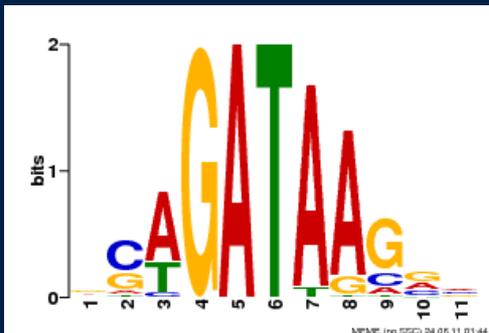
- Wikipedia

What is different about RBCs?



Transcription factor (TF) binding helps promote transcription

A major TF in red cells is **GATA1**.

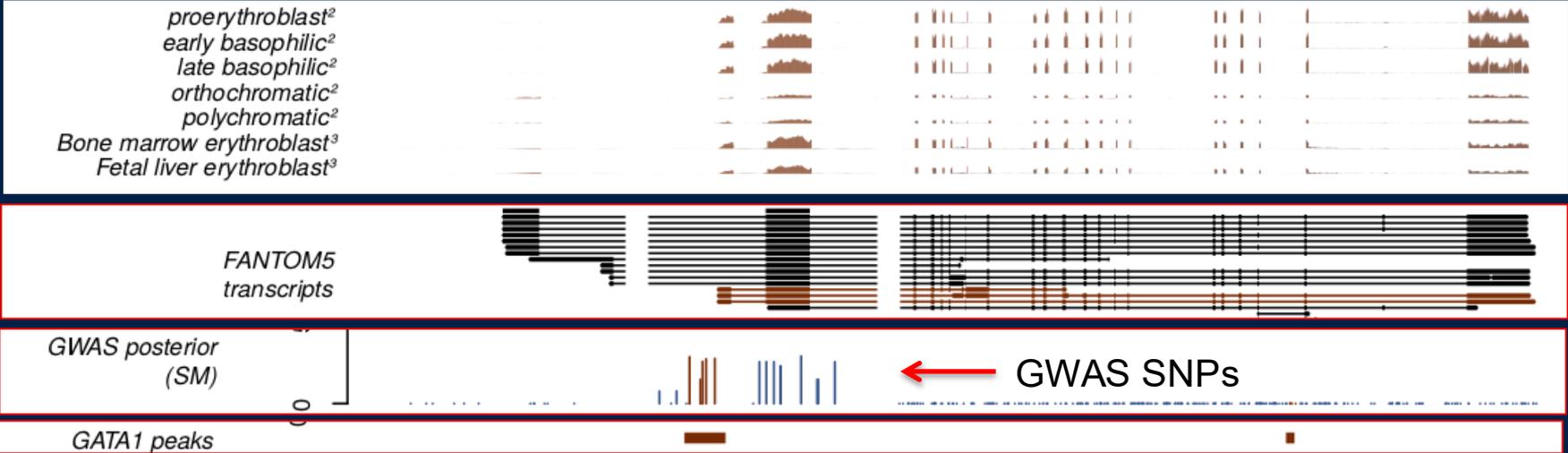


...so-called because of the **DNA motif** it binds to.

"GATA1 regulates the expression (i.e. formation of the genes' products) of an ensemble of genes that mediate the development of red blood cells and platelets."

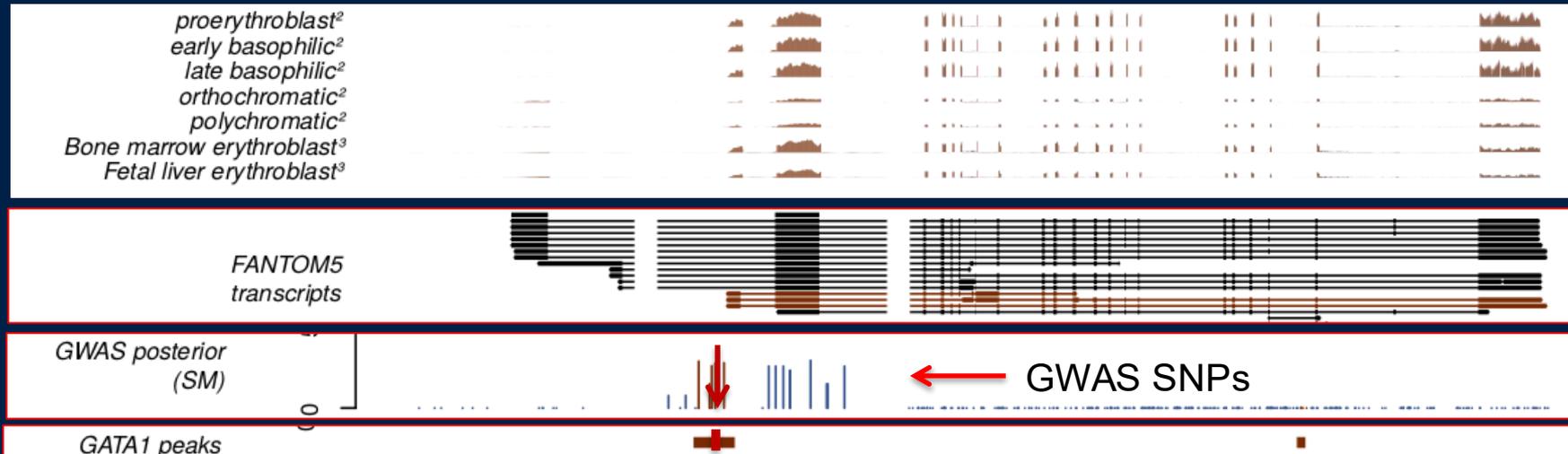
- Wikipedia

Putting the story together



GATA1 does bind this region in erythroid cells
measured using ChIP-seq

Putting the story together



rs10715451

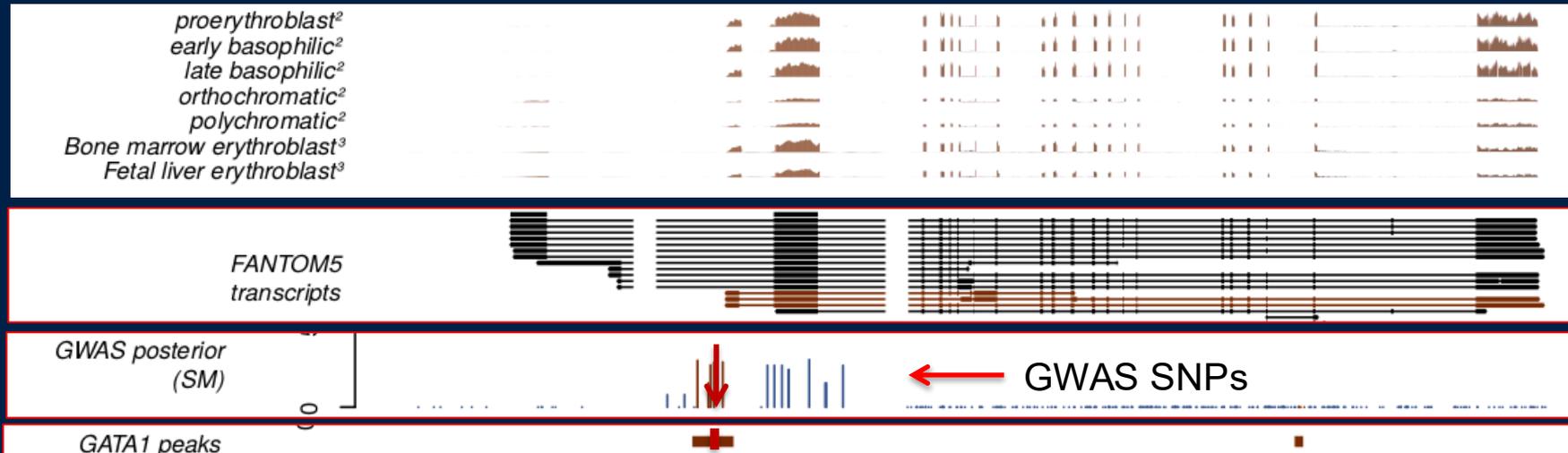
(malaria-protective allele)

...GGAGCGGTAAGATA...

(malaria risk allele)

...GGAGCGATAAGATA...

Putting the story together



Prediction from all this detective work:

The malaria risk allele causes higher expression of ATP2B4 – and more calcium pumping – in red cells

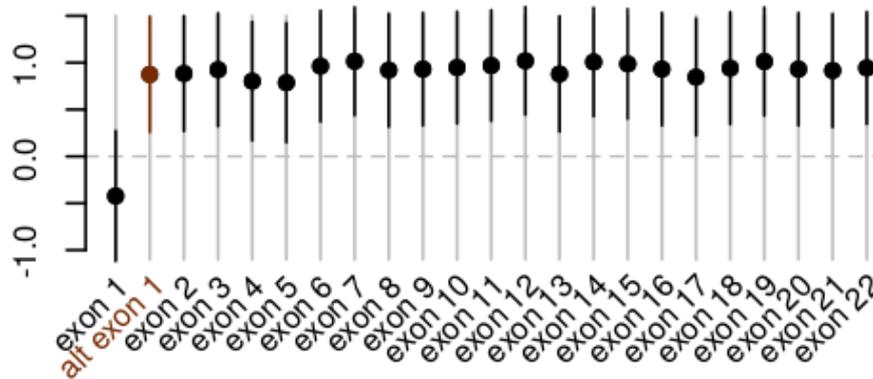
Probably without affecting other cells in the body

Genetic prediction

This prediction is borne out by RNA-seq data in cells from people with different genotypes.



Relative level of
expression
(A vs. G allele)



$N = 24$

Only limited data is
available

This example illustrates the complexities of fine-mapping

We've gone from a GWAS signal and (putatively) narrowed it down to a possible causal effect of a single SNP, in a single relevant cell type.

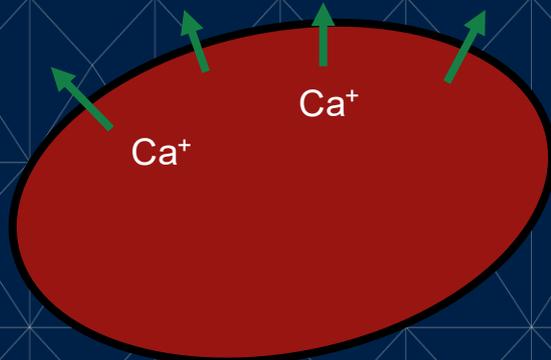
To get there we used a raft of functional genomic data – RNA-seq, ChIP-seq, chromatin state models – across different cell types.



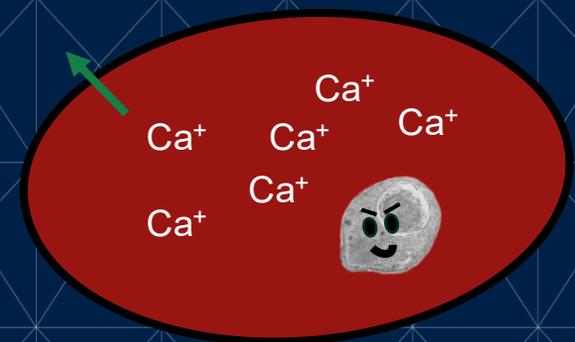
There is one remaining conundrum...

ATP2B4 encodes a calcium pump. It removes Ca^+ from the cell.

Parasites need calcium to grow. They *should* like the cell on the right...



rs10715451 G allele
associated with **malaria risk**

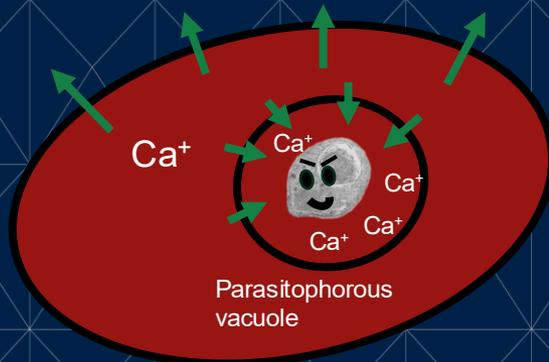


rs10715451 A allele
associated with **malaria protection**

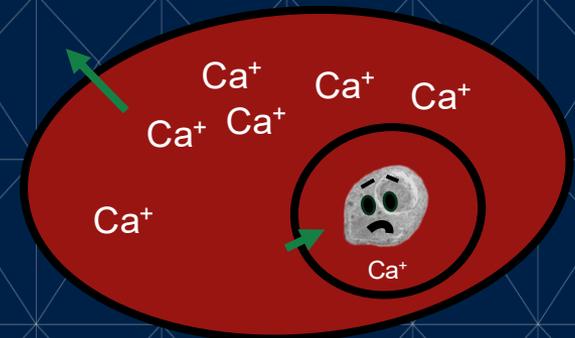
There is one remaining conundrum...

ATP2B4 encodes a calcium pump. It removes Ca^+ from the cell.

Hypothesis: pump gets inverted in the membrane of the parasitophorous vacuole.



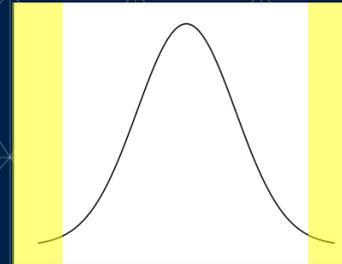
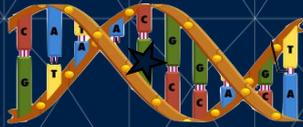
rs10715451 G allele
associated with **malaria risk**



rs10715451 A allele
associated with **malaria protection**

The circle of genetic causation

...passing on DNA, with **mutations** and **recombination**, to new generations...



...whose success is affected by the traits they have...

...that gets physically packaged up into chromosomes...

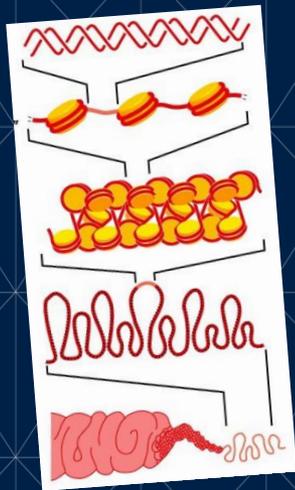
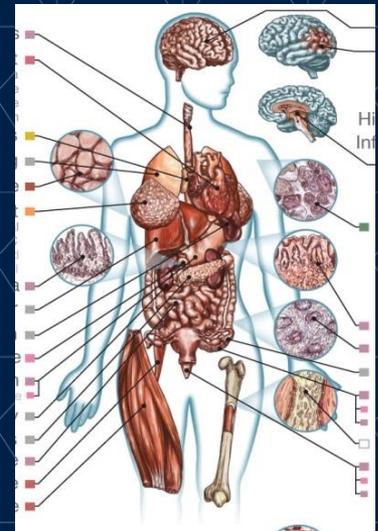
*microarrays,
genome sequencing*

*Clinical phenotype
measurements*

There is complex biology at all stages

And we can measure it.

*Biomarker
measurements*

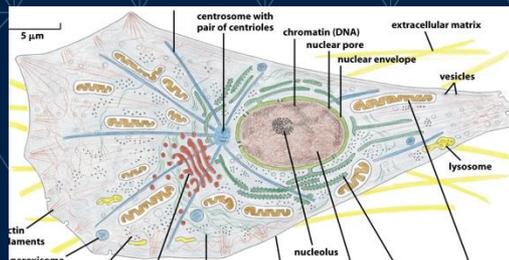


*Chromatin state
marker assays,
CHIP-seq, ...*

*RNA-seq,
spectroscopy, antibody
binding*

...that combine to make individuals...

...inside cells, where it is **transcribed** to form proteins and other molecules...



...that affect how the cells behave, forming different organs...

Learning biology from GWAS - summary

Have highlighted two of the complexities that could occur when trying to fine-map genetic association signals.

They are pretty fascinating and luckily there is lots more of this type of thing to find!

**Anything that can happen, does happen.
...and there is lots of data!**

Learning biology from GWAS - summary

Non-coding variants

Long-distance interactions in the genome

Changes to gene expression

Polygenic effects (lots of variants involved)

Cell-type / tissue heterogeneity

Pleiotropy (a variant affects lots of phenotypes at once)

Genetic interactions

Host-pathogen interactions

Repetitive DNA / repeat expansions

Genome structural variation

Genome evolution

Anything that can happen, it does happen.

Clinical translation?

It is safe to say that there has been only **limited translation** of GWAS outputs into clinically useful results so far.

But there are some success stories, e.g.:

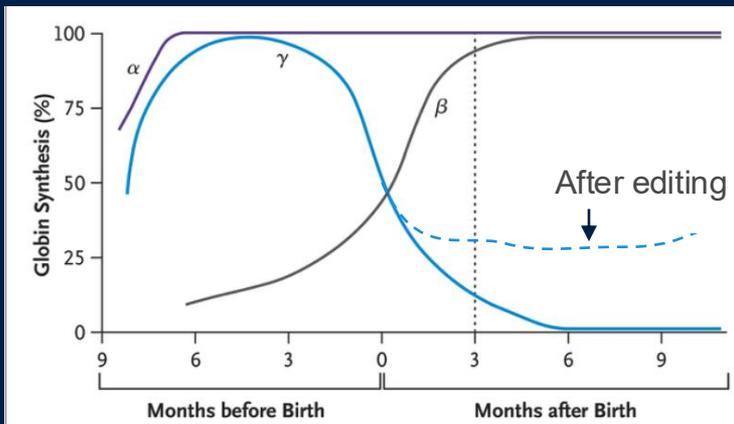
ORIGINAL ARTICLE | BRIEF REPORT

CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia

Authors: Haydar Frangoul, M.D., David Altshuler, M.D., Ph.D., M. Domenica Cappellini, M.D., Yi-Shan Chen, Ph.D., Jennifer Domm, M.D., Brenda K. Eustace, Ph.D., Juergen Foell, M.D., [+18](#), and Selim Corbacioglu, M.D. [Author Info & Affiliations](#)

Published December 5, 2020 | N Engl J Med 2021;384:252-260 | DOI: 10.1056/NEJMoa2031054 | VOL. 384 NO. 3

Copyright © 2020



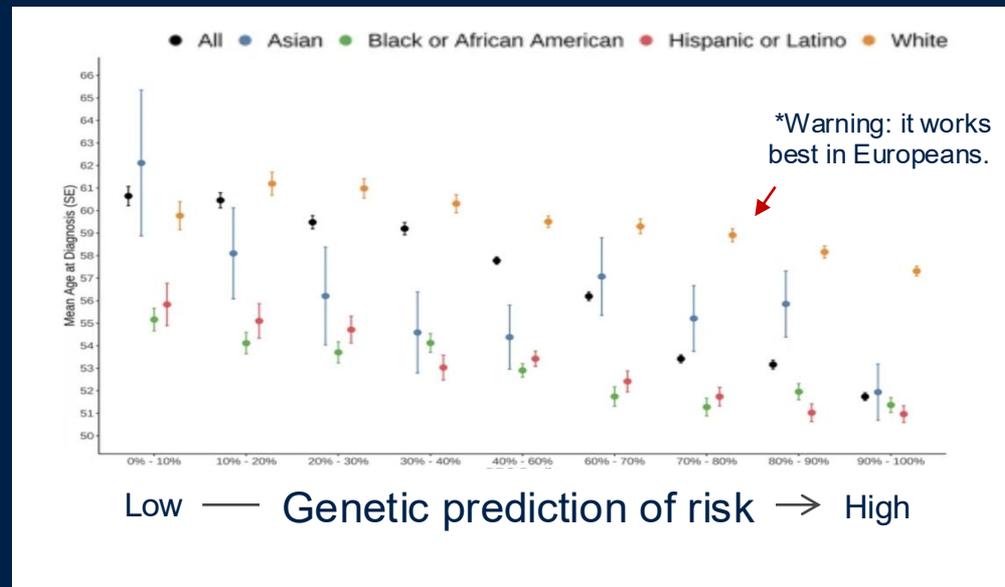
Gene editing of GWAS SNPs on chromosome 2 in adults with severe **sickle cell disease** or **beta thalassaemia**.

The SNP turns off suppression of **fetal haemoglobin expression**

Clinical translation?

It is safe to say that there has been only **limited translation** of GWAS outputs into clinically useful results so far.

But there are some success stories, e.g.:



Polygenic risk score modelling of Type II Diabetes

Armed with the GWAS results, you can **predict*** lifetime risk of T2D just from genome-wide genetics

Prospective cohort studies

A new crop of studies aims to create a database of deep genotype, phenotype, and exposure data across large cohorts of individuals sampled from the population or from health services. Examples:



Precision Medicine Initiative,
All of Us (US)



CartaGene (Canada)



China Kadoorie Biobank



FinnGen (Finland)



UK Biobank



The 100,000 genomes project (UK)

+
Our
Future
Health

In partnership with



0079618

0075506
398008/00TL1C9Q800196

1,885,350

people are already taking part
in the UK's largest health
research programme.

(As of 26 September 2024)

To all residents,

An opportunity to take part in research and learn new information about your blood pressure and future risk of disease.

You are invited to take part in Our Future Health, the UK's largest ever health research programme. If you take part, you will have the chance to find out more about your health now, and your risk of developing some diseases in the future.

Today, too many people spend many years of their life in poor health. Our Future Health aims to help prevent, detect and treat diseases earlier. Diseases like dementia, cancer, diabetes, heart disease and stroke.

Our Future Health needs up to five million people. Everyone aged 18 and over living in the UK is eligible to take part.

Taking part includes answering some online questions about yourself, providing a blood sample, and having your blood pressure measured at a local clinic.

In the future you will have the option to receive information on your risk of some diseases including diabetes, heart disease and some cancers. This will be calculated using the information you provide and analysis of the DNA in your blood sample.



Scan this QR code for more info
and to sign up

Or visit ourfuturehealth.org.uk/join/0518

£10 voucher

Sign up using the QR code or website link above and you will be eligible for a £10 voucher to recognise the time and effort of volunteering. You can find more information on the back of this letter.

You can also share this invitation with other members of your household.

If you have any questions, please call 0808 501 5634 or email support@ourfuturehealth.org.uk

Yours sincerely,

Raghib Ali

Raghib Ali OBE MD FRCP(UK)
Chief Medical Officer, Our Future Health
NHS Consultant in Acute Medicine

John Bell

Professor Sir John Bell GBE, FRS
Chairman, Our Future Health

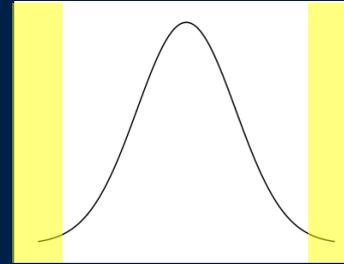
+
Our
Future
Health

<https://ourfuturehealth.org.uk>

Recruiting now

Conclusions and summary

- Most human traits are highly heritable
- For 'complex' traits, the effects are made up of many genetic variants often with modest effects - **polygenicity**
- GWAS study designs can find these variants. They rely on **large samples and dense genotyping**, and **patterns of linkage disequilibrium** to detect signals.
- A major frontier is to **understand the biology** and translate these findings into clinically useful insights and predictions.



Thanks for listening!

