Acknowledgement of Country

- The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.
- We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.
- We recognise their valuable contributions to Australian and global society.

Image: Digital reproduction of *A guidance through time* by Casey Coolwell and Kyra Mancktelow

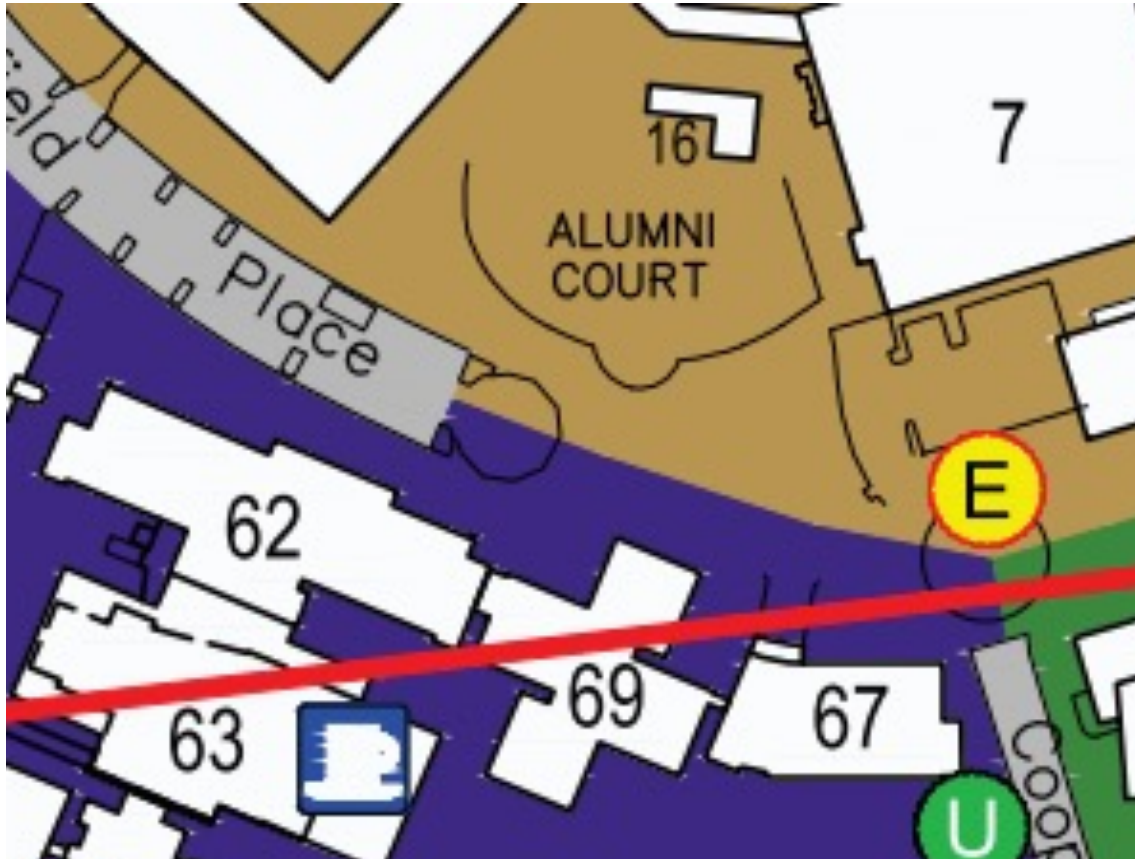
General Information:

- We are currently located in Building 69

Emergency evacuation point

- Food court and bathrooms are in Building 63

- 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



Data Agreement

To maximize your learning experience, we will be working with genuine human genetic data, during this module.

Access to this data requires agreement to the following in to comply with human genetic data ethics regulations

Please email pctgadmin@imb.uq.edu.au with your name and the below statement to confirm that you agree with the following:

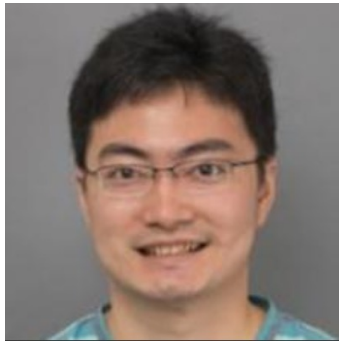
“I agree that access to data is provided for educational purposes only and that I will not make any copy of the data outside the provided computing accounts.”

Plan for the Module: Genetic Mapping

DAY 1 – GWAS, theory & practice	
8am	1. Introduction
9am	2. Phenotypes
10am	3. Genotypes
11am	4. Statistical tests
12pm	**Lunch break**
1pm	4. Statistical tests cont.
1:30pm	5. Lecture - data Cleaning
2pm	5. Data cleaning Prac
3pm	6. GWAS using unrelated indiv.
4pm	7. GWAS with relatives
5pm	Walk to dinner for 5:30pm

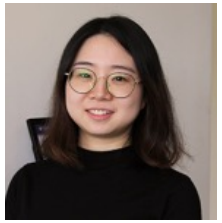
DAY 2 – Post-GWAS analysis	
8am	8. Meta-analysis & 9. Finding independent loci
9:45am	10. Gene based tests & 11. Fine mapping
11:30am	Q&A session with Dr Loic Yengo
12pm	**End of module, lunch**


Kath


JZ


Alesha


Tian


Ang

Li

Acknowledgements

I wish to acknowledge the following people for allowing me to use/adapt their material (slides and practical) in this module:

Allan McRae
Ben Hayes
Joanna Revez
Jian Zeng
Alesha Hatton
Ang Li
Naomi Wray

Various internet resources & GWAS protocol papers

Prac testers: Giulio, Tian, JZ, Em, Li

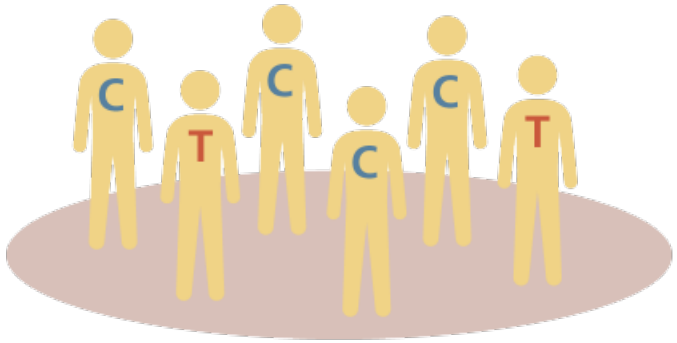
Introduction to Genome Wide Association Studies (GWAS)

What is a GWAS?

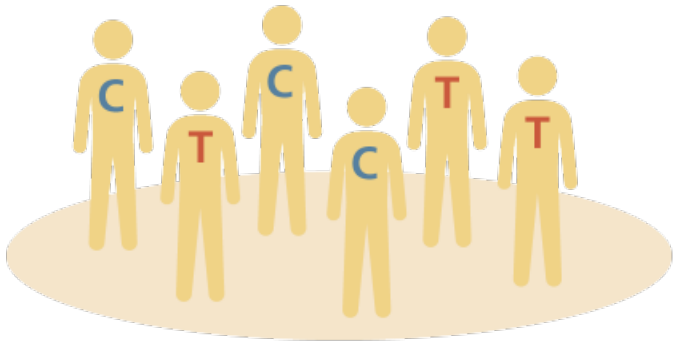
- A **G**enome **W**ide **A**ssociation **S**tudy is a hypothesis-free method for identifying associations between locations in the genome and a trait of interest

- Three key parts to a GWAS:
  - A trait of interest or phenotype
  - Genetic markers measured across the genome
  - Statistical test of association between markers & phenotype

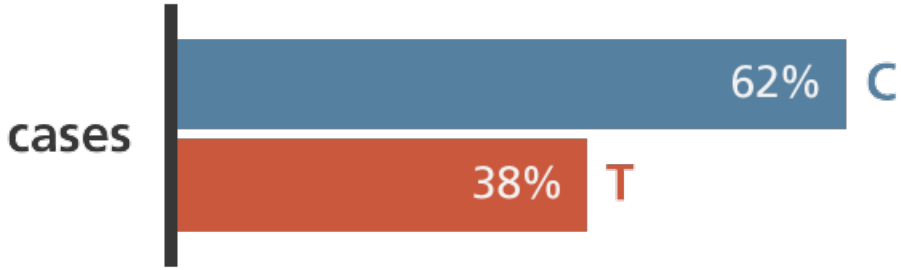
Example: Binary trait



cases (n=1,000)
people with heart disease



controls (n=1,000)
people without heart disease

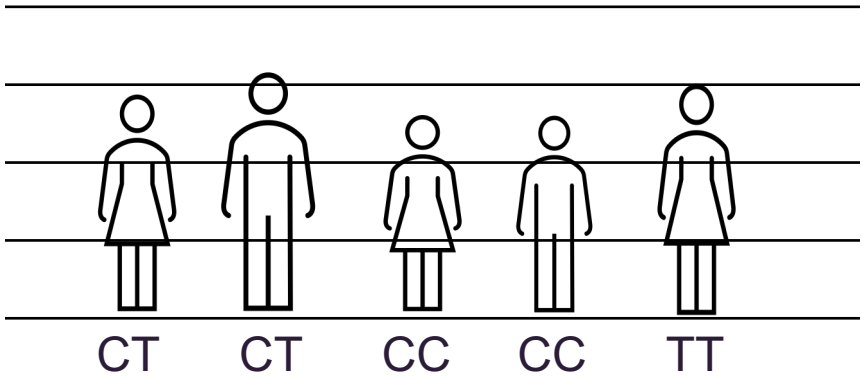


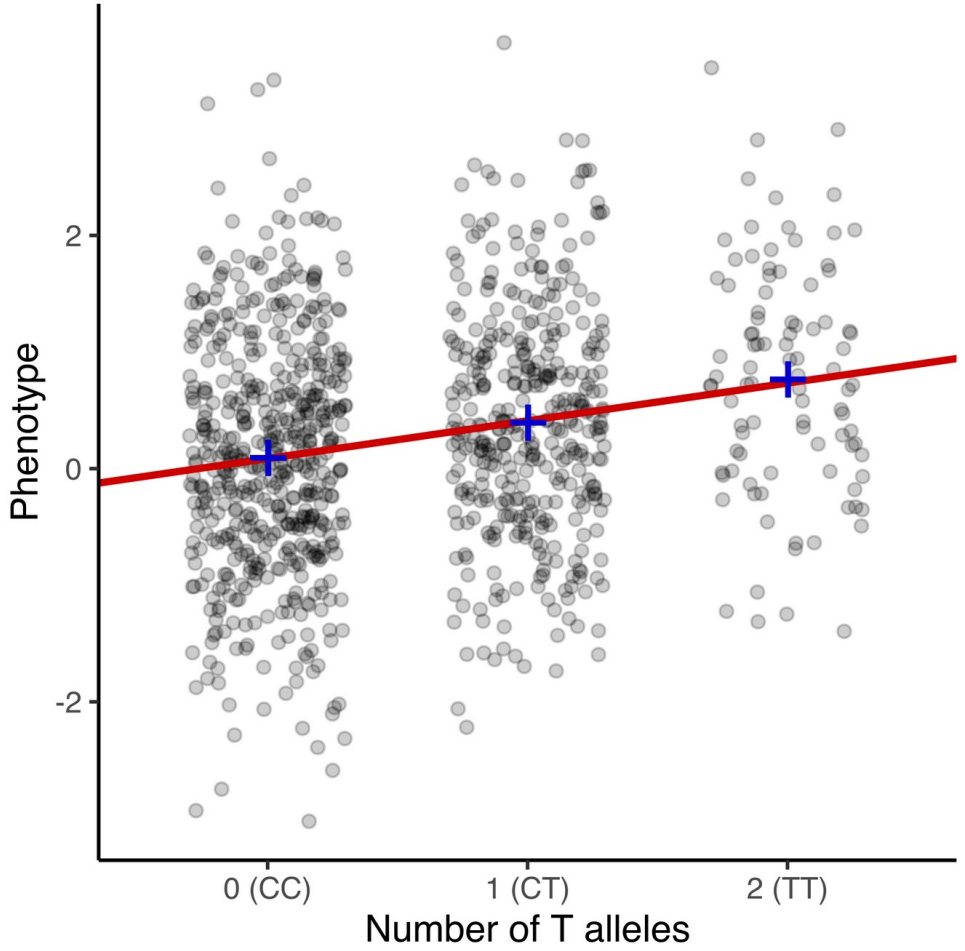
Test of association:

"Is the frequency of the 'C' allele different in cases vs. controls"

P = 0.0012

Example: Quantitative trait





- Linear model:

$$y = \alpha + x\beta + e$$

phenotypes

intercept

genotypes

SNP effect

error

e.g.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways

Elena López-Isac ✉, Marialbert Acosta-Herrera, Martin Kerick, Shervin Assassi, Ansuman T. Satpathy,

Nature Communications **10**, Article number: 4955 (2019) | Cite this article

Abstract

Systemic sclerosis (SSc) is an autoimmune disease that shows one of the highest mortality rates among rheumatic diseases. We perform a large genome-wide association study (GWAS), and meta-analysis with previous GWASs, in 26,679 individuals and identify 27 independent genome-wide associated signals, including 13 new risk loci. The novel associations nearly double the number of genome-wide hits reported for SSc thus far. We define 95% credible sets of less than 5 likely causal variants in 12 loci. Additionally, we identify specific SSc subtype-associated signals. Functional analysis of high-priority variants shows the potential function of SSc signals, with the identification of 43 robust target genes through HiChIP. Our results point towards molecular pathways potentially involved in vasculopathy and fibrosis, two main hallmarks in SSc, and highlight the spectrum of critical cell types for the disease. This work supports a better understanding of the genetic basis of SSc and provides directions for future functional experiments.

e.g.

GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways

Elena López-Isac, Marialbert Acosta-Herrera, Martin Kerick, Shervin Assassi, Ansuman T. Satpathy,

Nature Communications **10**, Article number: 4955 (2019) | Cite this article

Abstract

Systemic sclerosis (SSc) is an autoimmune disease that shows one of the highest mortality rates among rheumatic diseases. We perform a large genome-wide association study (GWAS), and meta-analysis with previous GWASs, in 26,679 individuals and identify 27 independent genome-wide associated signals, including 13 new risk loci. The novel associations nearly double the number of genome-wide hits reported for SSc thus far. We define 95% credible sets of less than 5 likely causal variants in 12 loci. Additionally, we identify specific SSc subtype-associated signals. Functional analysis of high-priority variants shows the potential function of SSc signals, with the identification of 43 robust target genes through HiChIP. Our results point towards molecular pathways potentially involved in vasculopathy and fibrosis, two main hallmarks in SSc, and highlight the spectrum of critical cell types for the disease. This work supports a better understanding of the genetic basis of SSc and provides directions for future functional experiments.

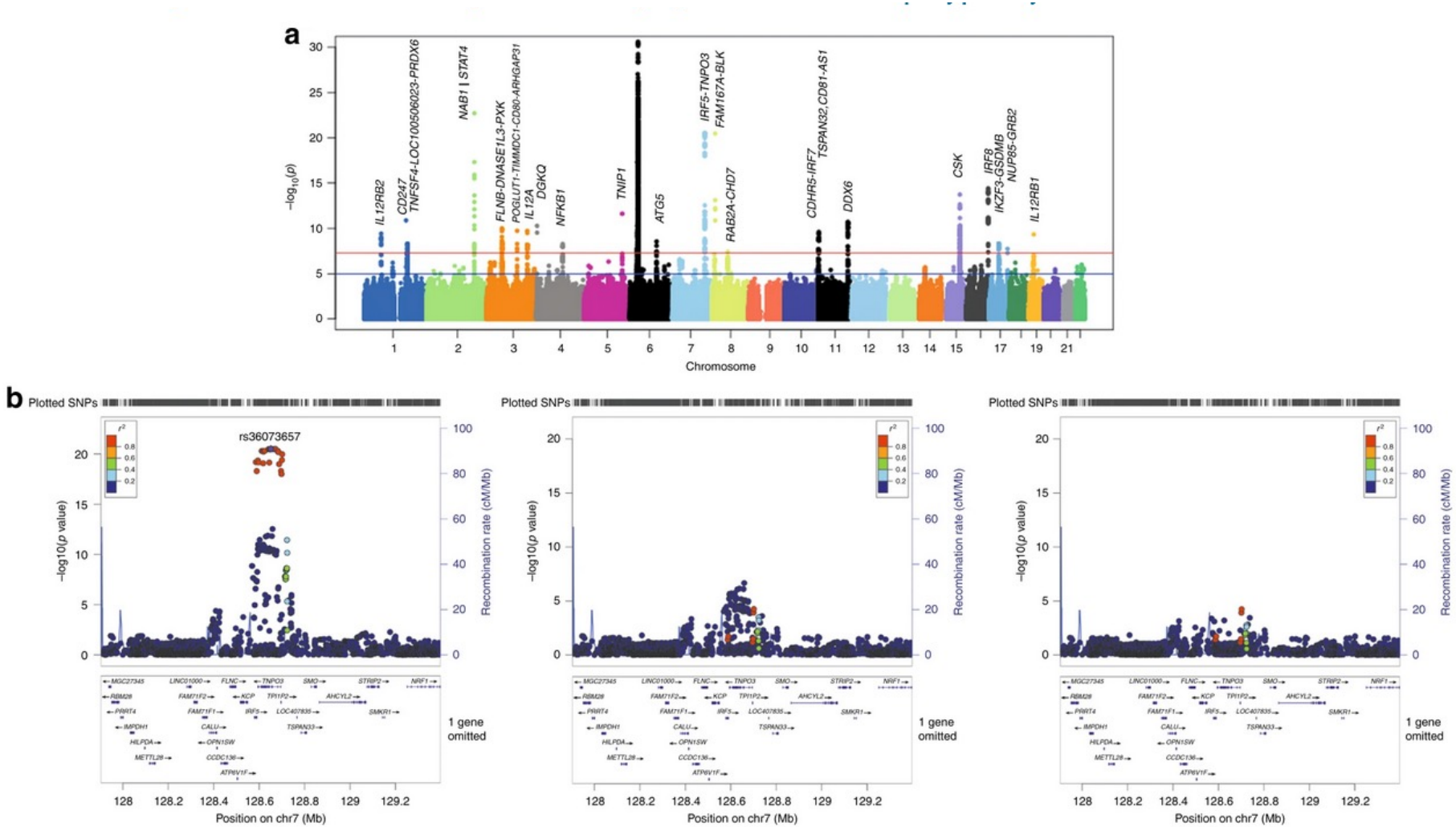
study design

key result

fine mapping & biological interpretation

GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways

Elena López-Isac ✉, Marialbert Acosta-Herrera, Martin Kerick, Shervin Assassi, Ansuman T. Satpathy,



What is a GWAS?

- Conceptually, a GWAS is straight forward. However, LOTS can go wrong!

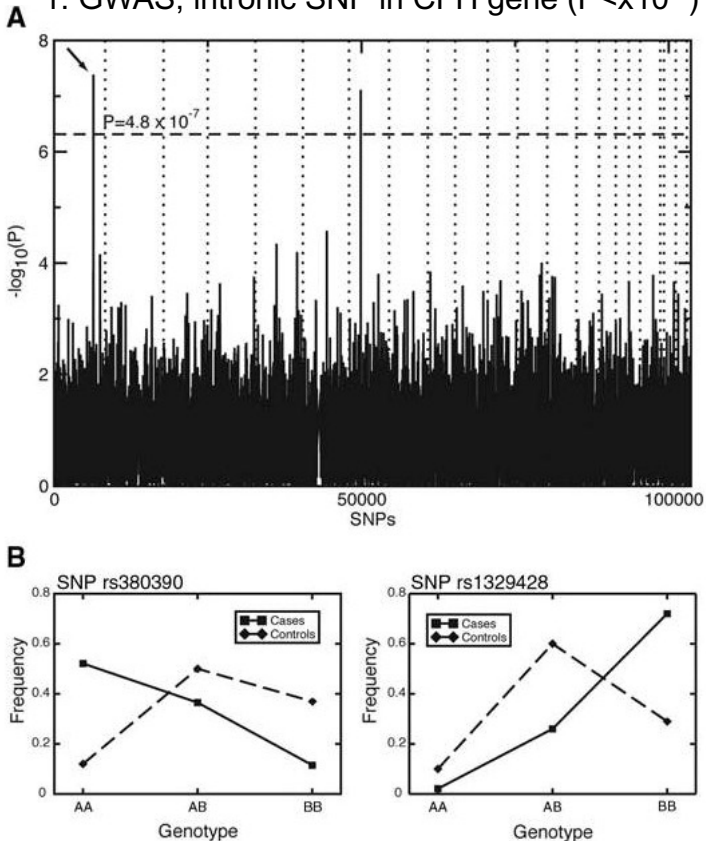
- **Most of the time in GWAS is spent in preparing the data to avoid various pitfalls**

Brief history of GWAS

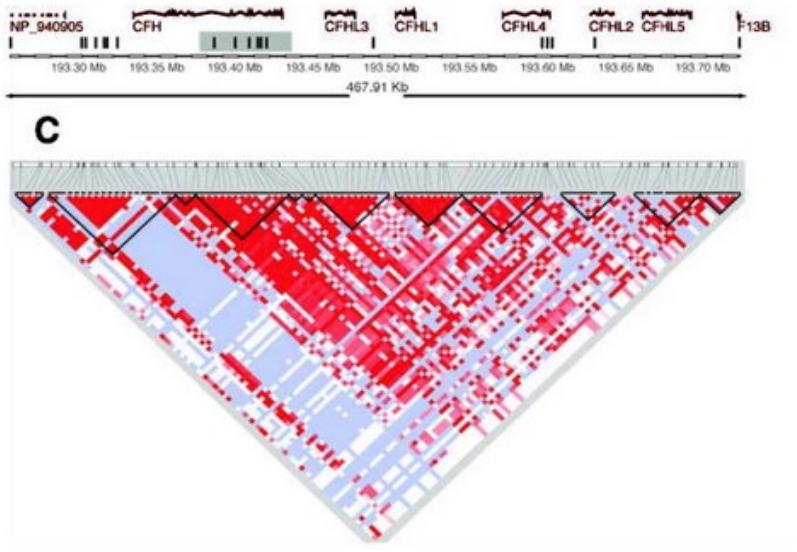
2005: one of the first GWAS, age-related macular degeneration

- Klein et al. 2005 *Science;* 96 cases and 50 controls; 116,204 SNPs

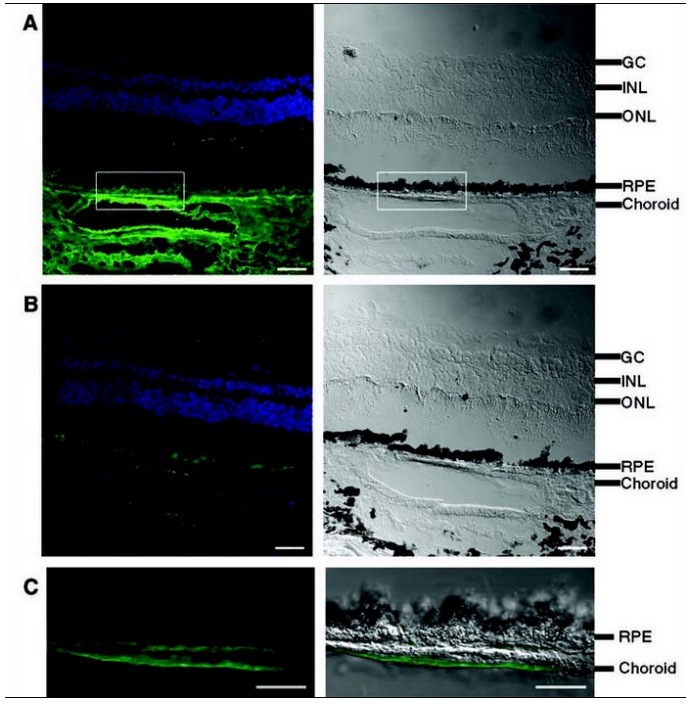
1. GWAS, intronic SNP in CFH gene ($P<x10^{-7}$)



2. re-sequencing region to identify missense variant



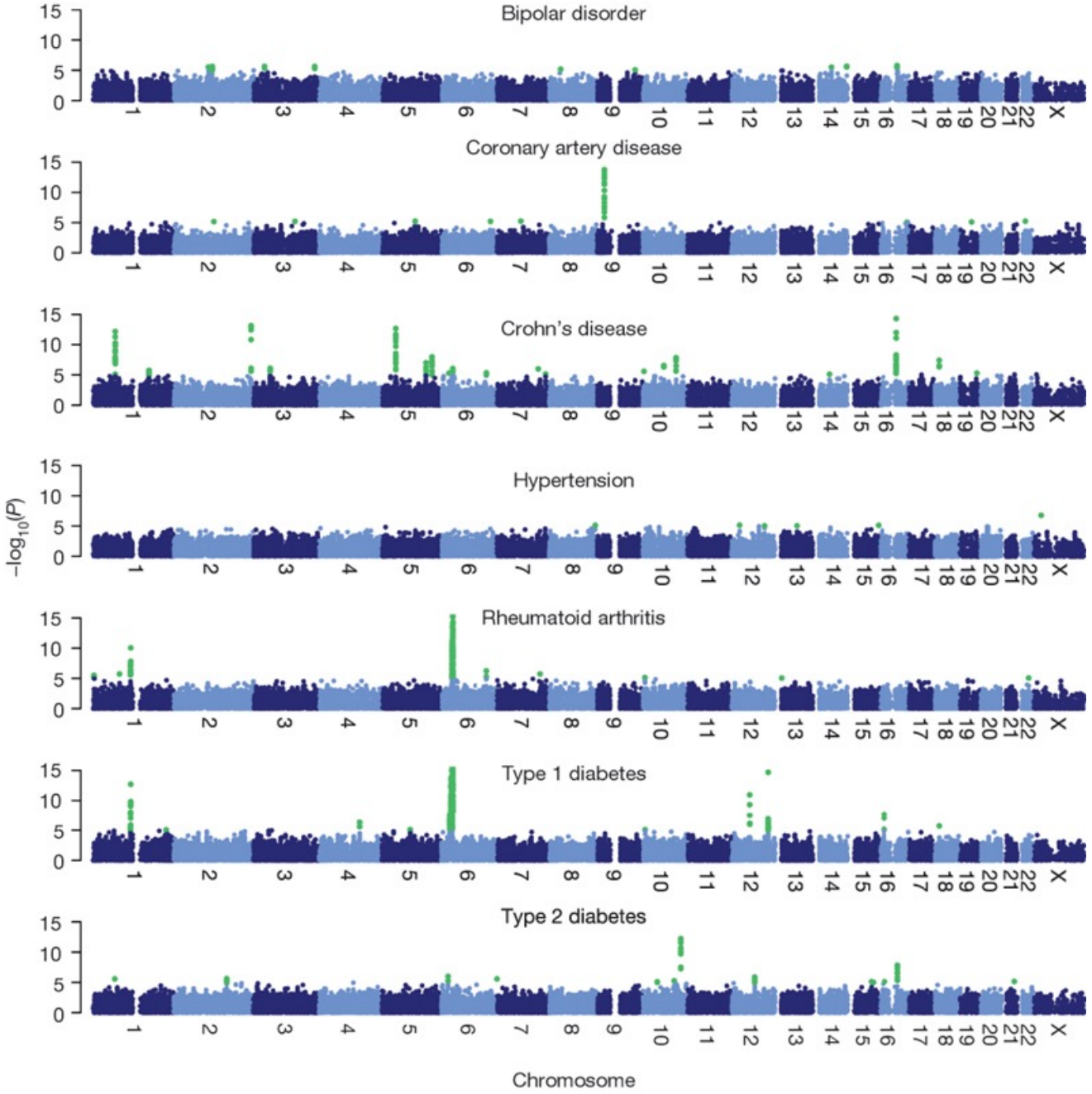
3. functional follow-up of CFH gene in retina



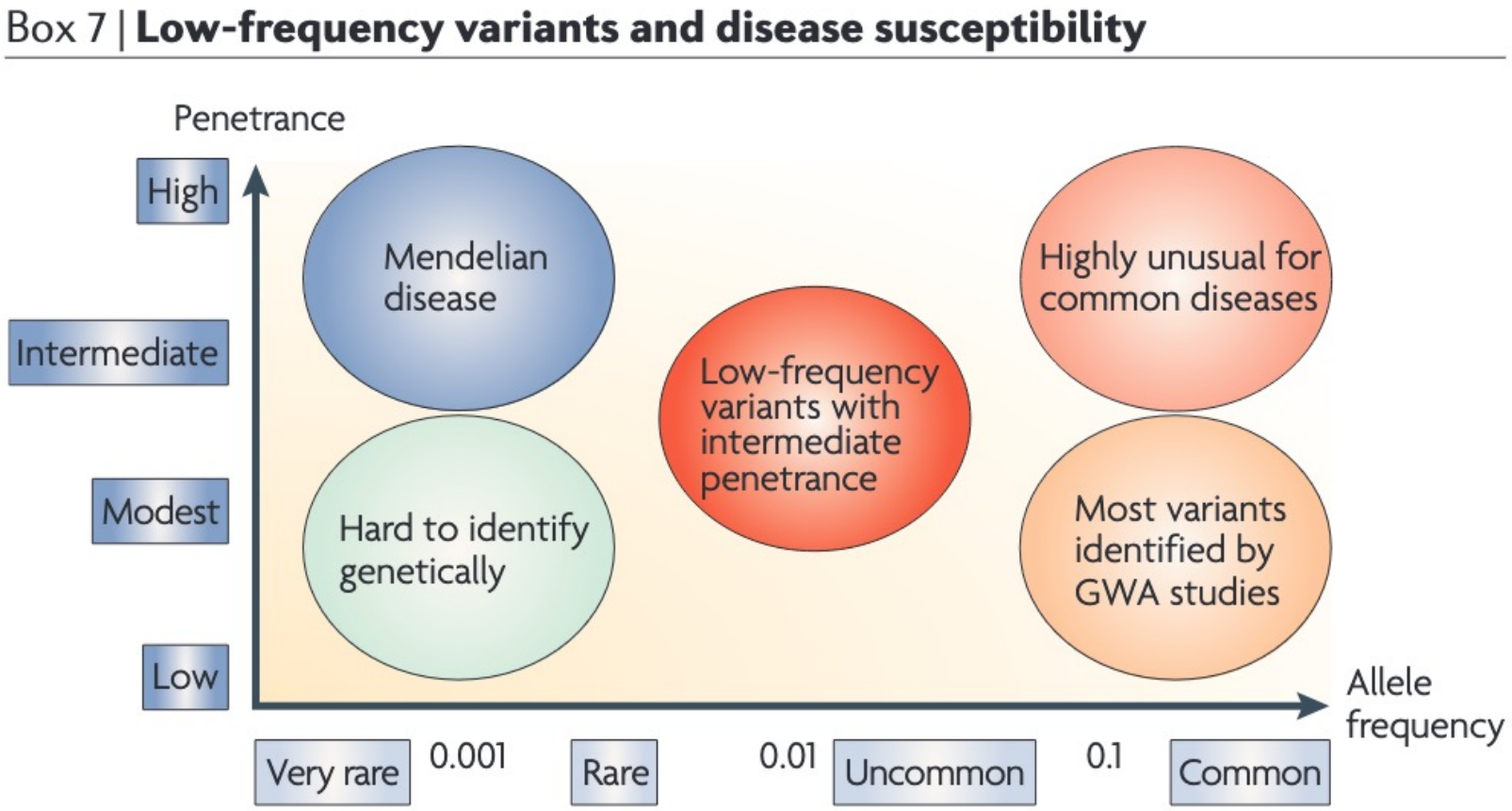
WTCCC

Wellcome Trust Case-Control Consortium

- First large scale GWAS (2007)
- 14,000 cases over 7 diseases
- 3,000 shared controls
- 500K Affymetrix GeneChip



GWAS
– ‘hypothesis free’ but dependent on frequency



McCarthy et al. (2008) *Nat Reviews Genetics*

GWAS
– ‘hypothesis free’ but dependent on frequency

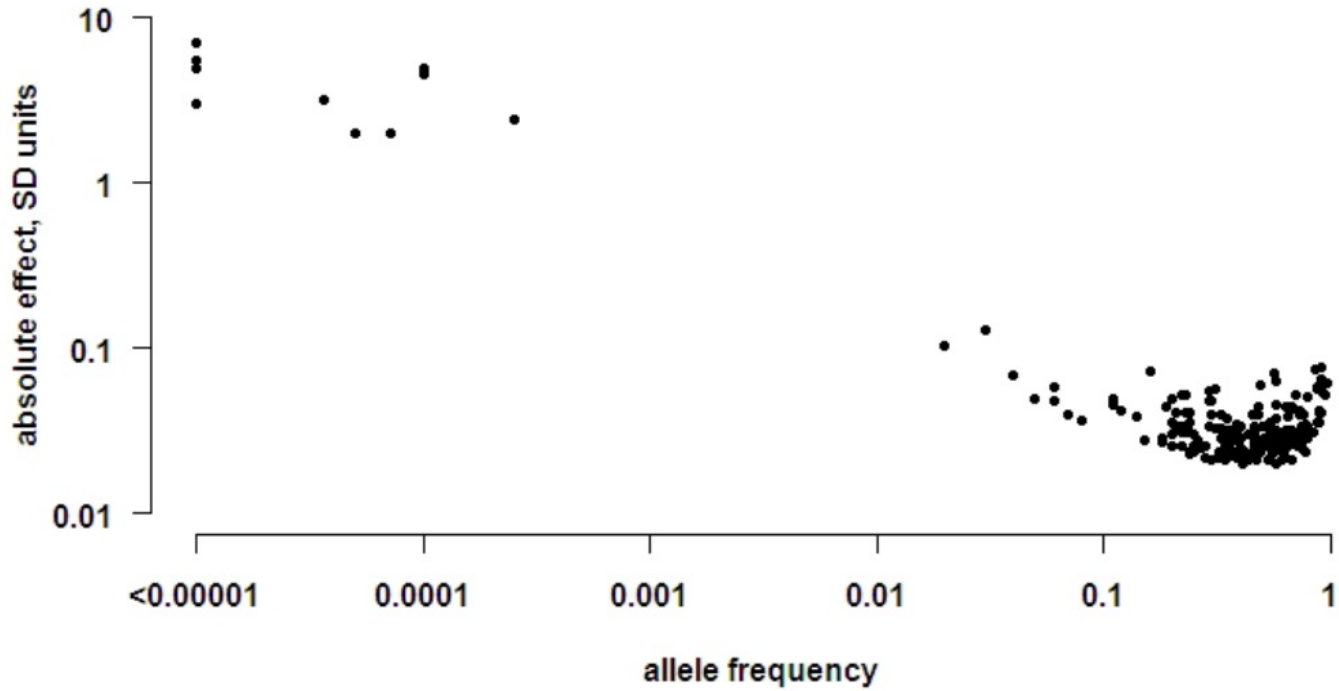
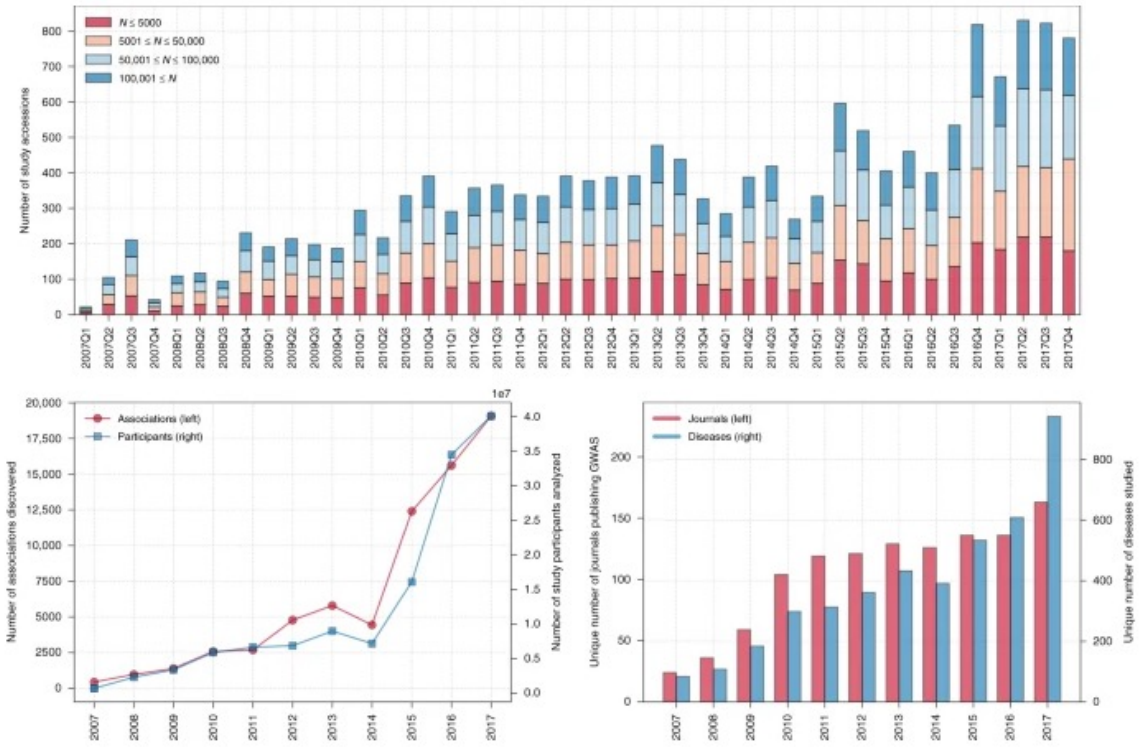


Figure 2. Mutations with intermediate effect (0.1 to 1 SD units) and low frequency (0.01 to 0.001) are not detected efficiently by either linkage or genome-wide association studies. Shown are the allele effect size (x-axis) and frequency (y-axis) for GWAS results from Lango Allen *et al.* [7] and for the sample of mutations described in Table 1. Assumptions for frequencies and effect of mutations from Table 1 are noted in the table footer.

Kemper et al. (2005) *Genome Biology*

Trends in GWAS

Fig. 1



The growth of GWAS, 2007–2017. The upper panel shows the number of study accessions published per quarter over time colored according to sample size to show the growth of larger ($100{,}001 \leq N$) GWAS. The lower left panel shows the strong positive correlation between the number of associations found and the number of participants used in GWAS over time. The lower right panel shows the growth in the number of unique traits examined as well as the number of unique journals publishing GWAS over time. 2007–2017 is selected since only 10 entries occurred before 2007. Each panel contains full calendar years only. Source: NHGRI-EBI GWAS Catalog

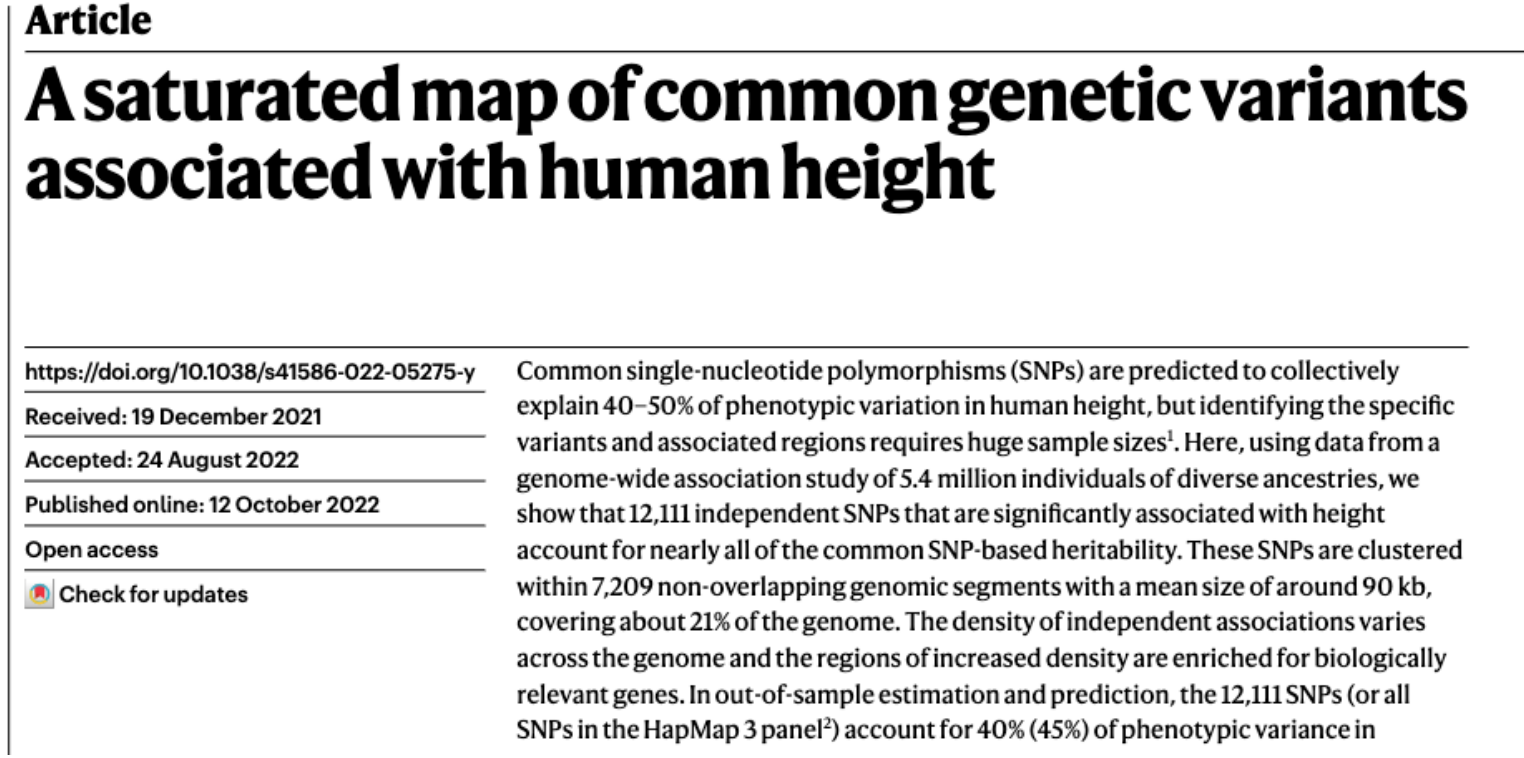
- More genetic markers
- Bigger sample sizes + meta-analysis
- New traits & diseases
- More discoveries + insights

- Predominately EUR ancestry recruited from US, UK & Iceland

Mills & Rahal (2019) *Communications Biology*

Latest (published) GWAS has 5.4M people!

- Q&A with Loic Yengo (1st author) tomorrow



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[1]. 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[2]) account for 40% (45%) of phenotypic variance in

Yengo et al. (2023) *Nature*

In conclusion

3 key considerations in GWAS design

phenotypes, genotypes, statistical tests

GWASs have led to important discoveries since early 2000's

Primary goal -> biological insight into disease (requires follow-up)
Can learn about genetic architecture of trait
Limitations

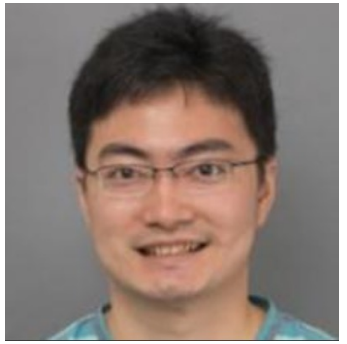
Conceptually a simple idea but care is required

Plan for the Module: Genetic Mapping

DAY 1 – GWAS, theory & practice	
8am	1. Introduction
9am	2. Phenotypes
10am	3. Genotypes
11am	4. Statistical tests
12pm	**Lunch break**
1pm	4. Statistical tests cont.
1:30pm	5. Lecture - data Cleaning
2pm	5. Data cleaning Prac
3pm	6. GWAS using unrelated indiv.
4pm	7. GWAS with relatives
5pm	Walk to dinner for 5:30pm

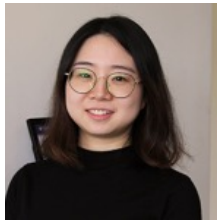
DAY 2 – Post-GWAS analysis	
8am	8. Meta-analysis & 9. Finding independent loci
9:45am	10. Gene based tests & 11. Fine mapping
11:30am	Q&A session with Dr Loic Yengo
12pm	**End of module, lunch**


Kath


JZ


Alesha


Tian


Ang

Li

Objective for the Module: Genetic Mapping

- during the module you will *not* find recipes and/or pipelines

- I hope you will *understand* what your doing & be able to critique others

Objective for the Module: Genetic Mapping

Take home messages for today...

- Beware of biases
- Multiple testing