# Cross-trait genetics in immune-mediated diseases

Chris Wallace  chr1swallace  chr1swallace.github.io

18 May 2020


James Liley

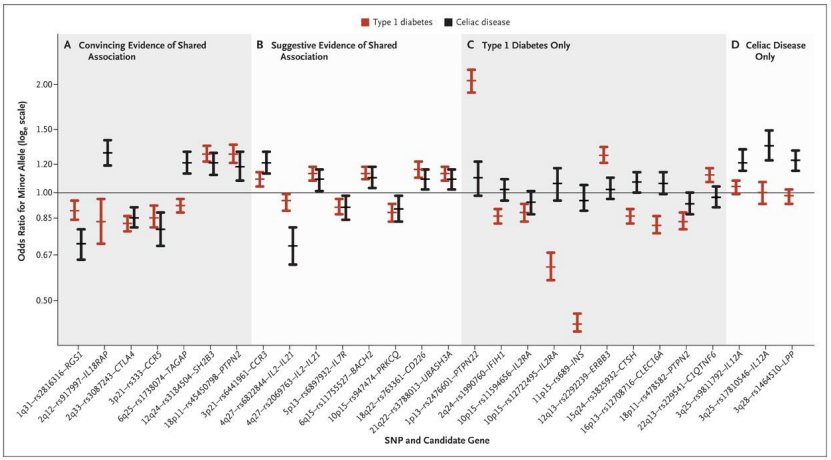

Stasia Grinberg


Olly Burren


Guillermo Reales

# Shared genetics underpin the established co-occurence of IMD

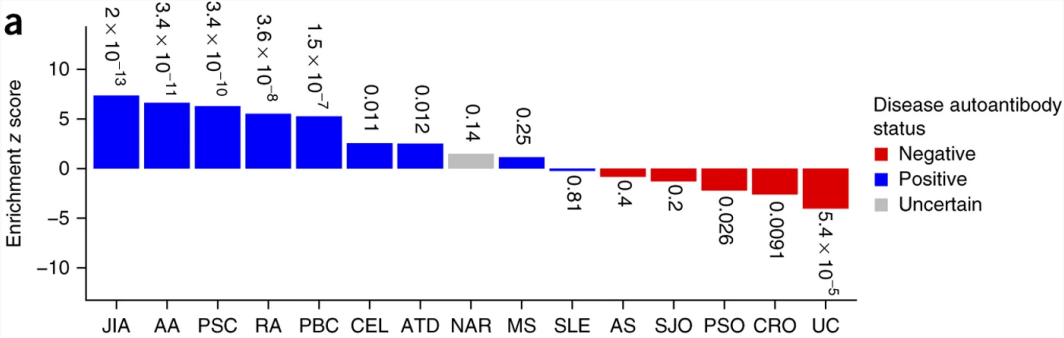
# Comparison of pairs of diseases: type 1 diabetes and celiac disease




For each disease, more “hits” were shared than unique

*Smyth et al, NEJM 2008*

# Compare one disease to many: Type 1 diabetes *versus*




JIA=juvenile idiopathic arthritis
AA = alopecia
PSC = primary sclerosing cholangitis
RA=rheumatoid arthritis

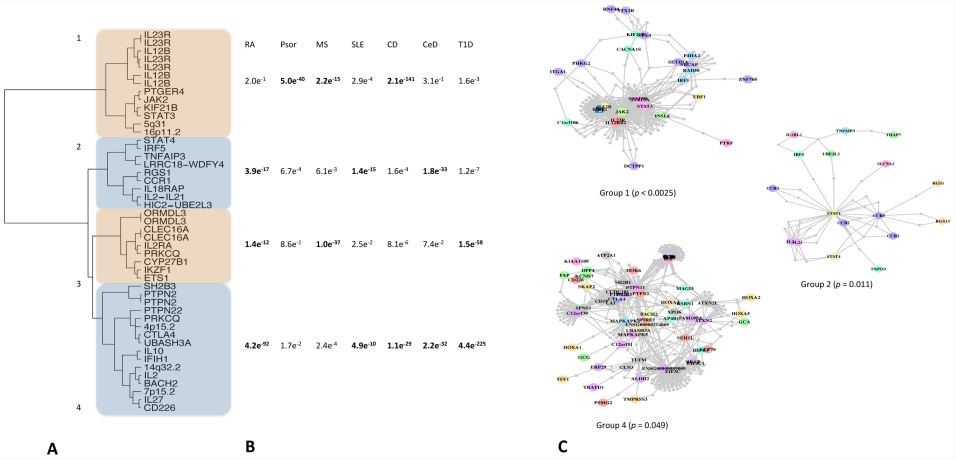
PBC = primary biliary cholangitis
CEL = celiac disease
ATD = autoimmune thyroid disease
NAR = narcolepsy

MS = multiple sclerosis
SLE = lupus
AS = ankylosing spondylitis
SJO = Sjogren's

PSO = psoriasis
CRO = Crohn's disease
UC = ulcerative colitis

*Onengut et al, Nat Genet 2015*

# Compare many diseases, clustering lead SNPs



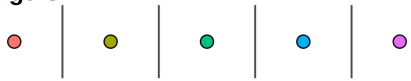

SNPs clustered by disease susceptibility patterns

Genes near clustered SNPs tended to be biologically related (help pick causal genes?)

*Cotsapas et al, PLoS Genet 2011*

# Goals of cross-trait genetics

**Distinguish**



What makes disease A, A?
(and not B?)

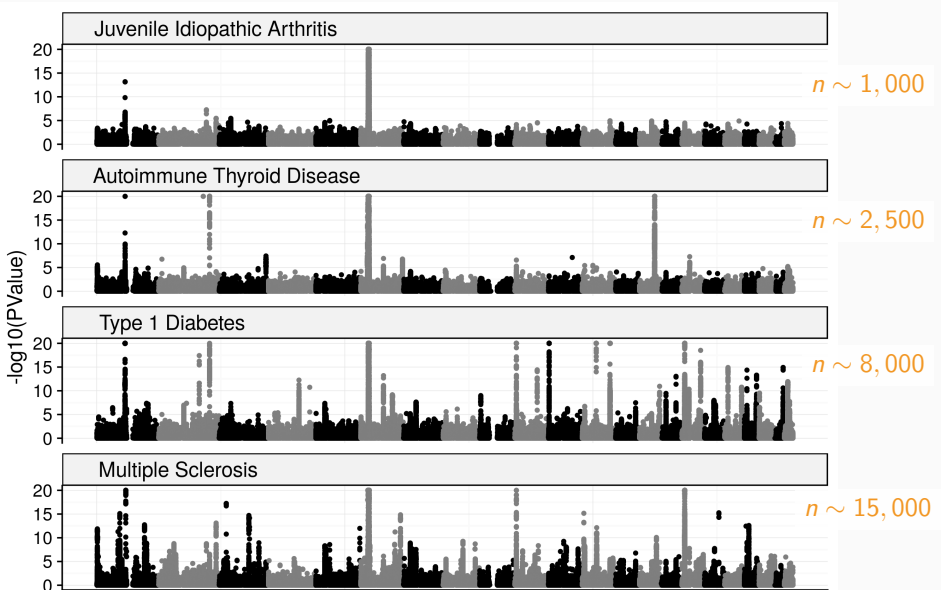
**Pool**

Transfer knowledge from better studied disease
(often to rare disease)

**Transfer**

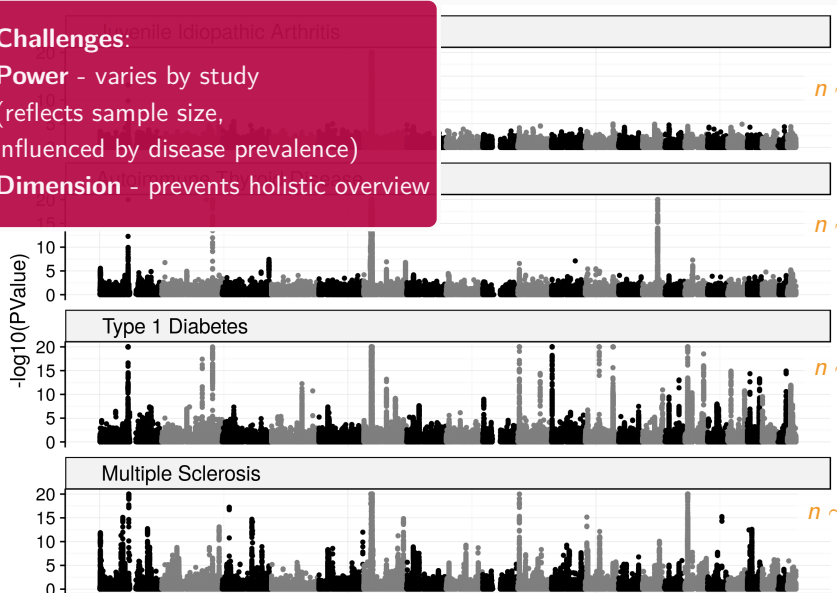
Share resources, increase knowledge

# Challenges for cross-trait genetics



# Challenges for cross-trait genetics

**Challenges**:
**Power** - varies by study
(reflects sample size,
influenced by disease prevalence)
**Dimension** - prevents holistic overview




$n \sim 1,000$

$n \sim 2,500$

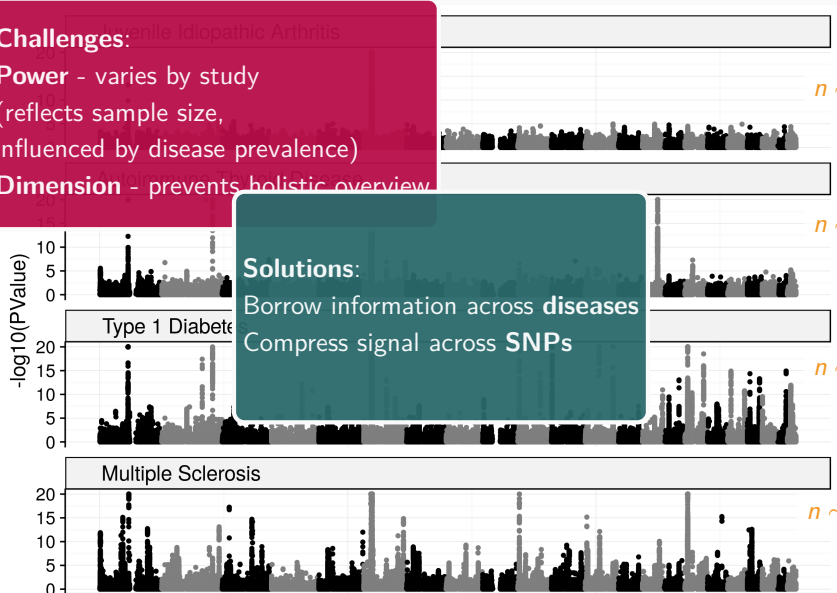
$n \sim 8,000$

$n \sim 15,000$

# Challenges for cross-trait genetics

**Challenges**:
**Power** - varies by study
(reflects sample size,
influenced by disease prevalence)
**Dimension** - prevents holistic overview

**Solutions**:
Borrow information across **diseases**
Compress signal across **SNPs**




$n \sim 1,000$

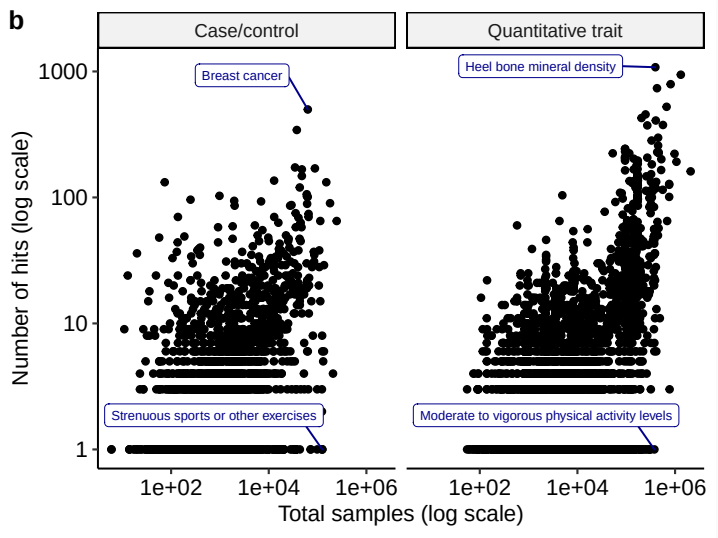
$n \sim 2,500$

$n \sim 8,000$

$n \sim 15,000$

# Information borrowing

# Association count is a function of sample size




*GWAS catalog, Feb 2020*

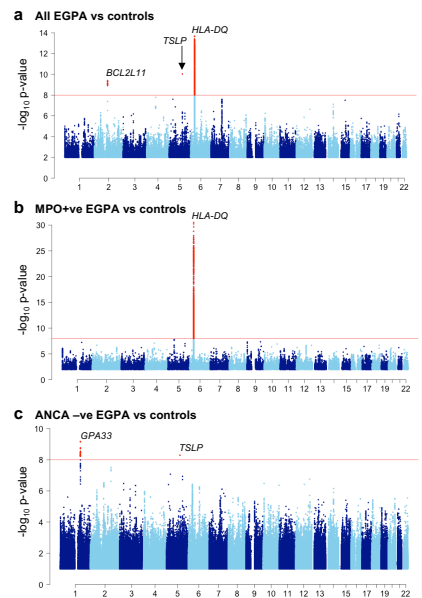
# Eosinophilic Granulomatosis with Polyangiitis (EGPA)

- ANCA-associated vasculitis + eosinophilia, asthma
- not all patients are ANCA positive
- considered a single disease
- incidence of 0.5-3.7 / million / year
- pan-European study collected 519 cases

| | |
|---|---|
| ANCA positive (MPO) | 161 |
| ANCA negative | 358 |
| Controls | 6717 |

*Lyons et al, Nat Commun 2019*

# Eosinophilic Granulomatosis with Polyangiitis (EGPA)




all EGPA

MPO+ EGPA

ANCA- EGPA

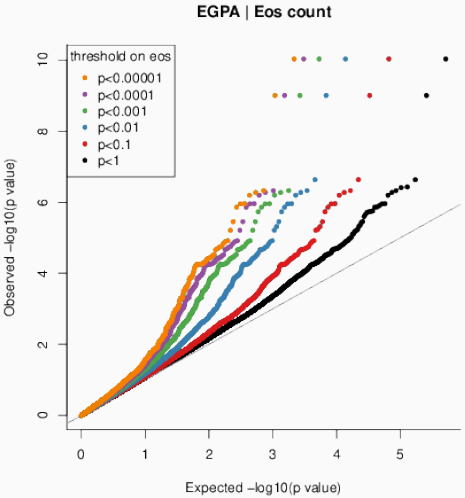
**EGPA genetics**

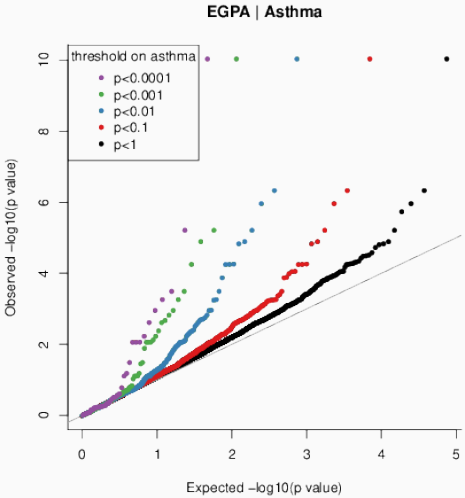
polygenic

subtype-specific associations

# EGPA significance enriched at both asthma and eosinophilia associated variants

Reminder: EGPA prodrome: **eosinophilia**, **asthma**



# cFDR: Conditional false discovery rate

**False discovery rate (FDR)**

$$FDR(\alpha) = \Pr(H_0|p < \alpha) = \frac{\Pr(p < \alpha|H_0)\Pr(H_0)}{\Pr(p < \alpha)}$$

# cFDR: Conditional false discovery rate

**False discovery rate (FDR)**

$$FDR(\alpha) = \Pr(H_0|p < \alpha) = \frac{\Pr(p < \alpha|H_0)\Pr(H_0)}{\Pr(p < \alpha)}$$

**conditional false discovery rate (cFDR)**

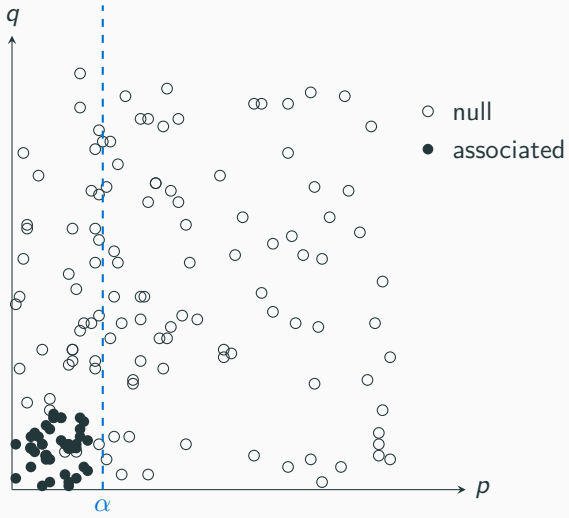
Two sets of GWAS P values:

- $p$ from *test disease* (EGPA)
- $q$ from *conditional trait* (asthma/eosinophil count)

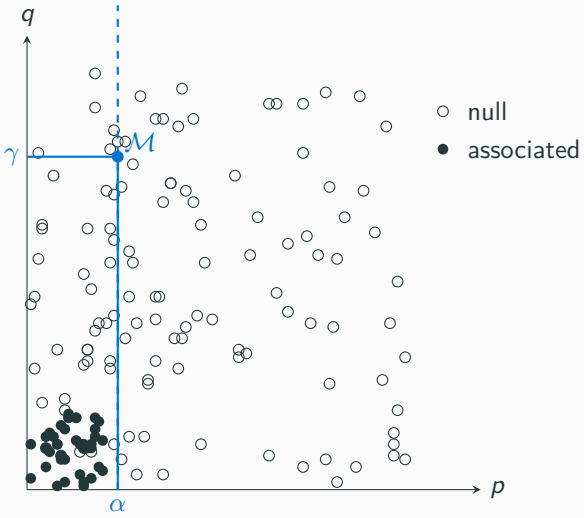
$$cFDR(\alpha, \gamma) = \Pr(H_0|p < \alpha, q < \gamma) = \frac{\Pr(p < \alpha|q < \gamma, H_0)\Pr(H_0|q < \gamma)}{\Pr(p < \alpha|q < \gamma)}$$

*Andreassen et al, PLOS Genet 2013*

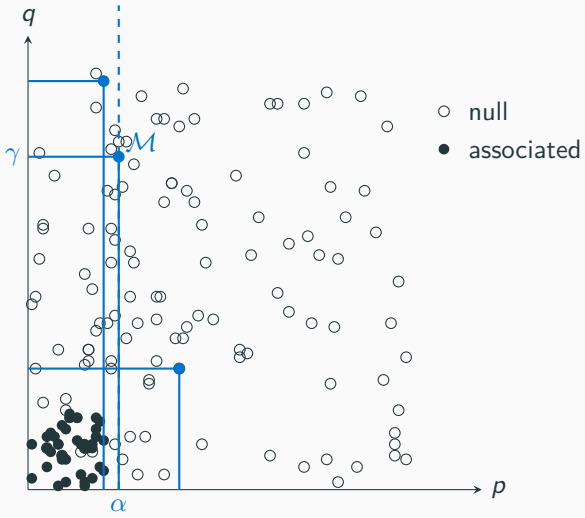
# Problem: the cFDR does not control the FDR



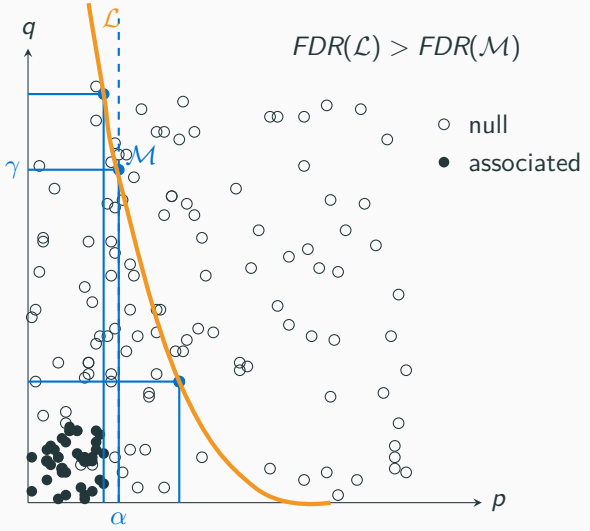
# Problem: the cFDR does not control the FDR



# Problem: the cFDR does not control the FDR



# Problem: the cFDR does not control the FDR




**Two solutions**

1. Asymptotically, ratio of $FDR(\mathcal{L})/FDR(\mathcal{M})$ is bounded
2. Generate new $v$ values: $\Pr((p,q) \in \mathcal{L} | H_0)$

# cFDR identifies 7 additional associations for EGPA

| | MPO+ EGPA | ANCA- EGPA |
|---|---|---|
| *Distinct associations* | HLA-DQ | GPA33 |
| *Shared associations* | BCL2L11, TSLP | |
| *Conditional associations* | | |
| Asthma | IL5, BACH2, 10p14 (GATA3?) | |
| Eosinophilia | CDK6, SOCS1, LPP, TBX3 | |
| *Differential symptoms (%)* | | |
| Glomerulonephritis | 29 | 9 |
| Neuropathy | 79 | 57 |
| *Treatment response (%)* | | |
| **Rituximab response** | **80** | **38** |

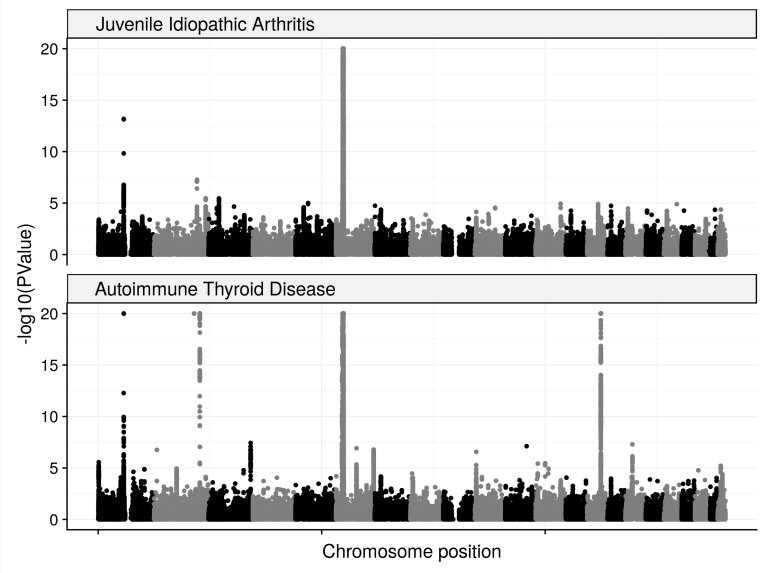
## Conclusions

1. EGPA is polygenic
2. Symptomatic differences between subtypes supported by genetics
3. asthma, eosinophilia in EGPA driven (partly) by same mechanisms as in general population

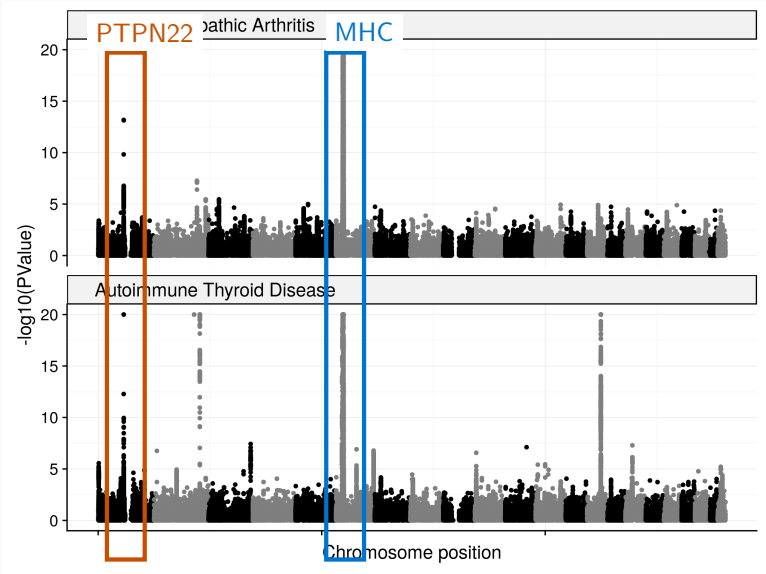
*Lyons et al, Nat Commun 2019*

# Compress information across SNPs in same region

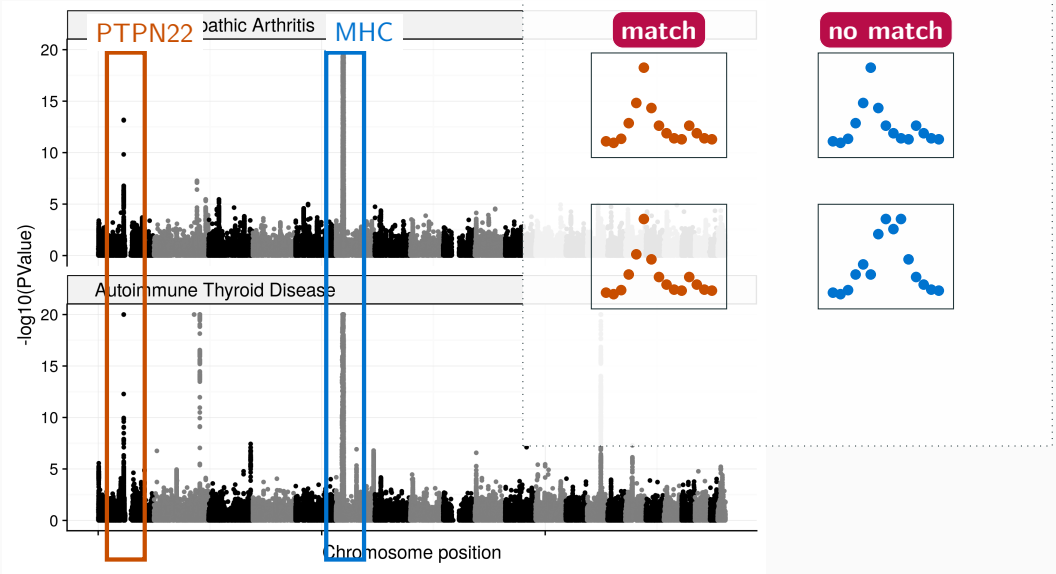
# Colocalisation: assessing evidence for shared/distinct causal variants



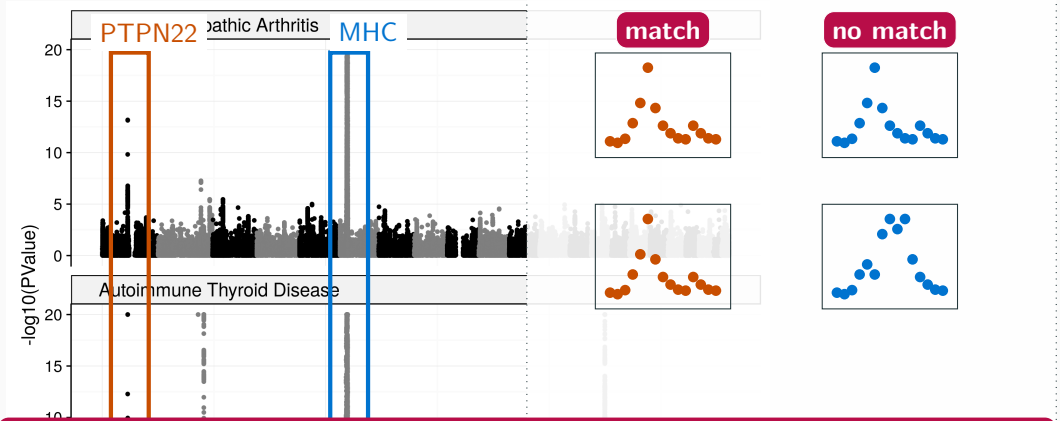
# Colocalisation: assessing evidence for shared/distinct causal variants



# Colocalisation: assessing evidence for shared/distinct causal variants



# Colocalisation: assessing evidence for shared/distinct causal variants




## Colocalisation as compression

Pair of diseases, hundreds of variants → single summary

(Potentially more if single causal variant assumption relaxed)

## ImmunoChip: 11 IMD / 188 genetic regions

| Disease | Study | Cases | Controls |
|---|---|---|---|
| Narcolepsy | Faraco | 1,886 | 10,421 |
| Psoriatic arthritis | Bowes | 1,962 | 8,923 |
| Autoimmune thyroid disease | Cooper | 2,733 | 9,364 |
| Juvenile arthritis | Hinks | 2,816 | 13,056 |
| Primary billiary cholangitis | Liu | 2,861 | 8,514 |
| Type 1 diabetes | Onengut | 6,670 | 12,262 |
| Coeliac disease | Trynka | 10,413 | 8,274 |
| Rheumatoid arthritis | Eyre | 11,475 | 15,870 |
| Systemic lupus erythmatosus | Langefeld | 11,590 | 15,984 |
| Ulcerative colitis | Liu | 13,449 | 31,766 |
| Multiple sclerosis | IMSGC | 14,498 | 24,091 |
| Crohn's disease | Liu | 16,619 | 31,766 |

# Colocalising more than two traits

## moloc

Formally extends coloc pairwise likelihood to $n$ traits, exponential increase in combinations limits utility beyond 3-4 traits

*Giambartolomei et al, Bioinformatics 2018*

## HyPrColoc

Can analyse 100s of traits, by deterministic collapse and clustering

*Foley et al, bioRxiv 592238*

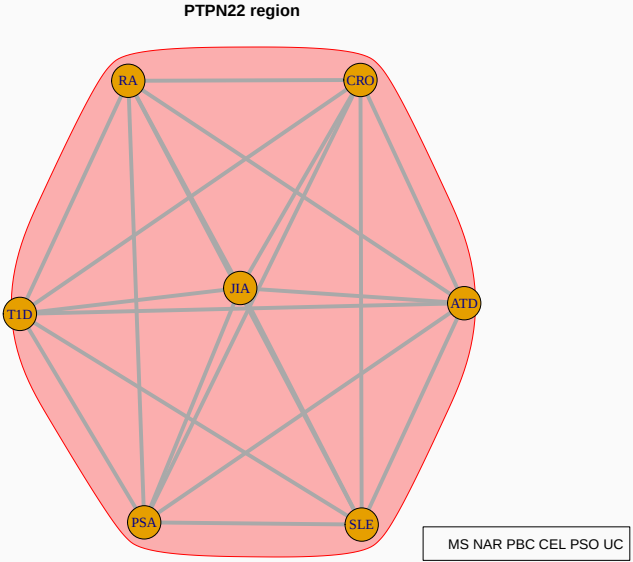
## Use coloc posteriors as a distance measure

Use standard hierarchical methods to cluster traits according to a symmetric distance measure

Takes advantage of recent extension to relax single causal variant assumption

*Wallace, PLoS Genetics 2020*

# Shared signals

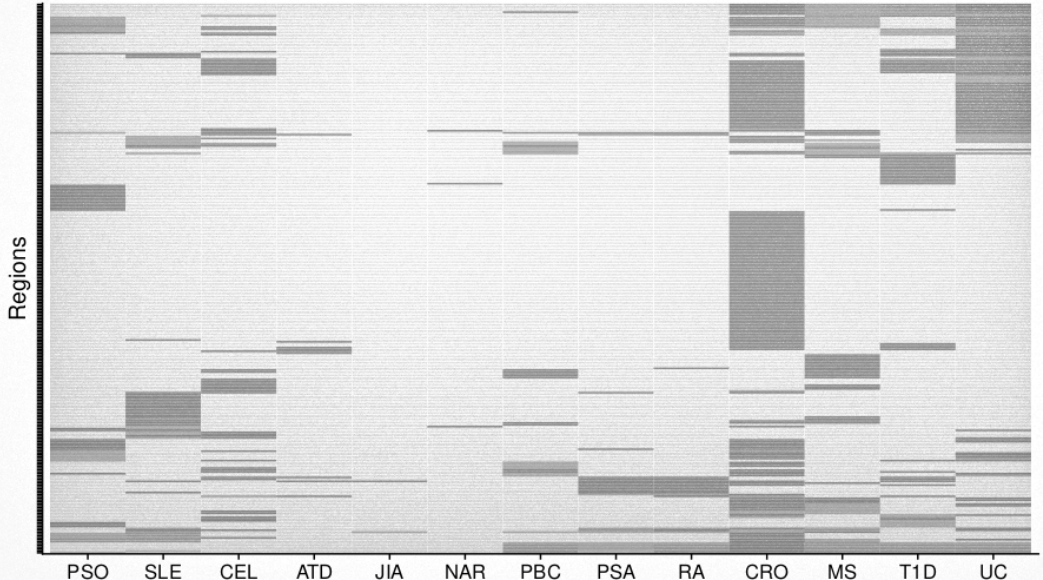


# Single ATD signal in *TSHR* region

**THSR region**


RA T1D JIA MS NAR PBC CEL PSO SLE CRO UC PSA

# Overview



# Overview


**Sample size still an issue**

Cannot distinguish low power/no association

# Informed dimension reduction and information borrowing

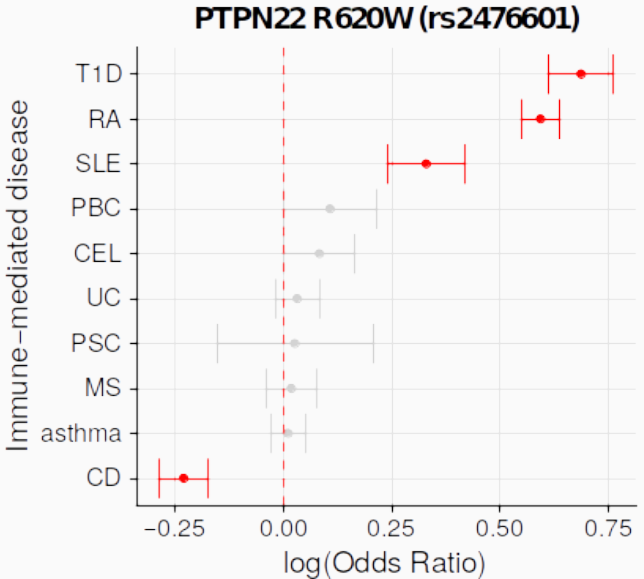
**Aim**: combine dimension reduction with information borrowing

- study genetics of diseases and subtypes too rare for standard GWAS
- pull apart IMD mechanisms

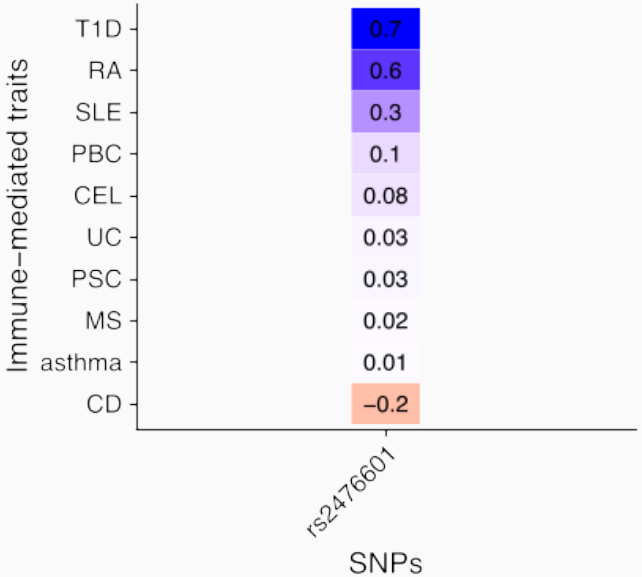
# Learn IMD-relevant dimensions from larger GWAS

| Disease | Study | Cases | Controls |
|---|---|---|---|
| Asthma | Demenais | 19954 | 107715 |
| Rheumatoid arthritis | Okada | 14361 | 43923 |
| Ulcerative colitis | de Lange | 12366 | 33609 |
| Crohn's disease | de Laange | 12194 | 28072 |
| Multiple sclerosis | IMSGC | 9772 | 17376 |
| Type 1 diabetes | Cooper | 5913 | 8829 |
| Primary sclerosing cholangitis | Ji | 4796 | 19955 |
| Coeliac disease | Dubois | 4533 | 10750 |
| Systemic lupus erythematosus | Bentham | 4036 | 6959 |
| IgA Nephropathy | Kiryluk | 3952 | 2747 |
| Vitiligo | Jin | 2853 | 37405 |
| Primary biliary cholangitis | Cordell | 2764 | 10475 |
| Latent autoimmune diabetes in adults | Cousminer | 2634 | 8581 |

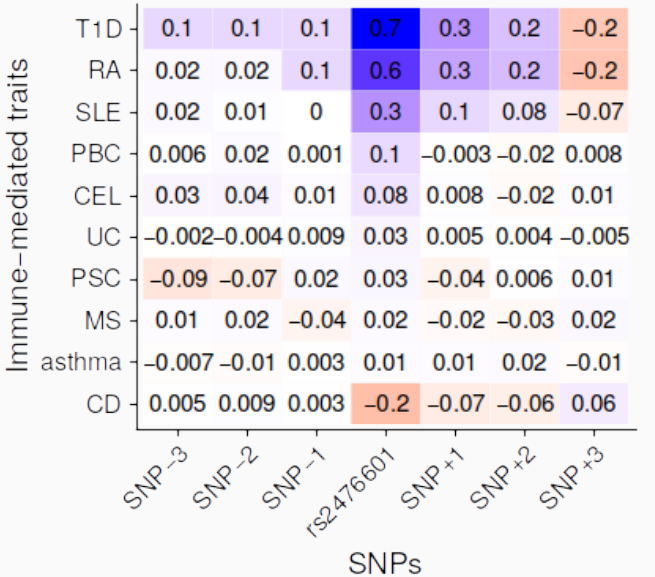
# Patterns of heterogeneity across diseases



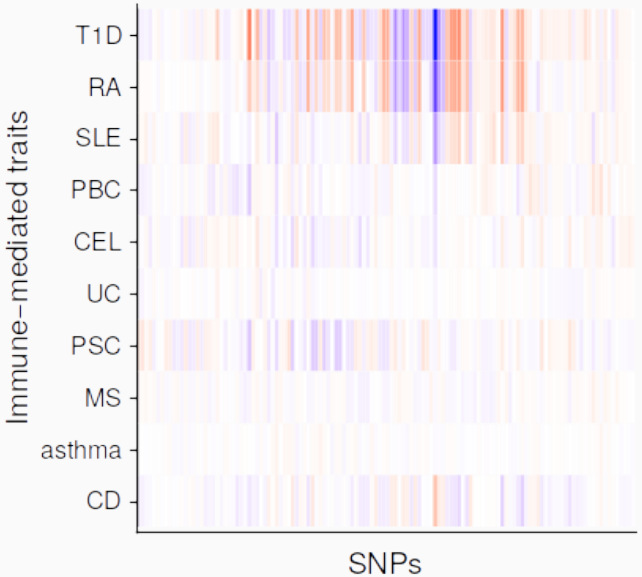
# Patterns of heterogeneity across diseases



# Patterns of heterogeneity across diseases

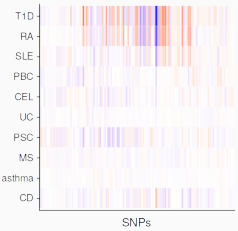


# Patterns of heterogeneity across diseases



# Principal components analysis to generate a new “basis”

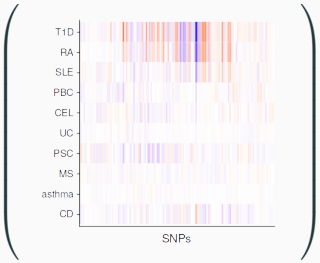
300,000 SNPs



# Principal components analysis to generate a new “basis”

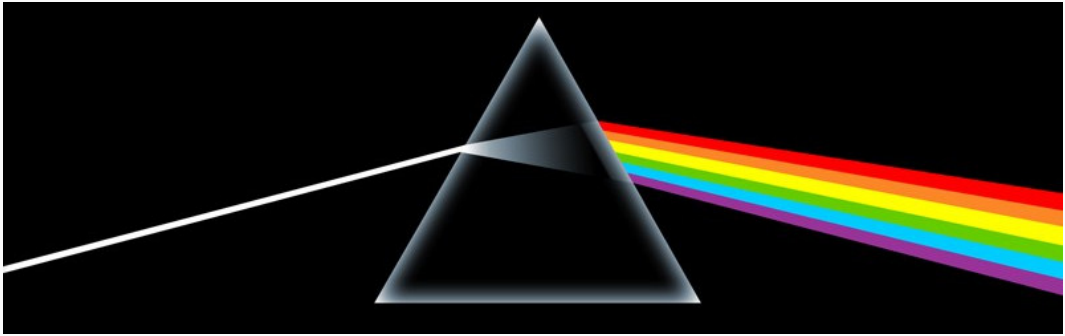
300,000 SNPs

PCA




=

# Principal components analysis to generate a new “basis”

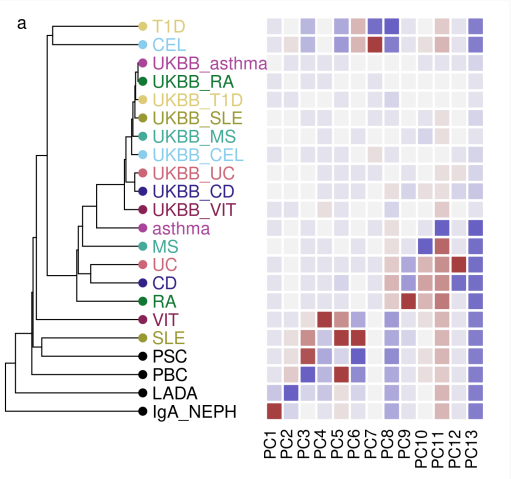


# Naive basis captures dataset, not disease

| UK Biobank | Cases |
|---|---|
| Vitiligo | 154 |
| Type 1 diabetes | 286 |
| Systemic lupus erythematosus | 366 |
| Multiple sclerosis | 1,228 |
| Coeliac disease | 1,452 |
| Ulcerative colitis | 1,795 |
| Crohn's disease | 1,032 |
| Rheumatoid arthritis | 3,730 |
| Asthma | 39,049 |

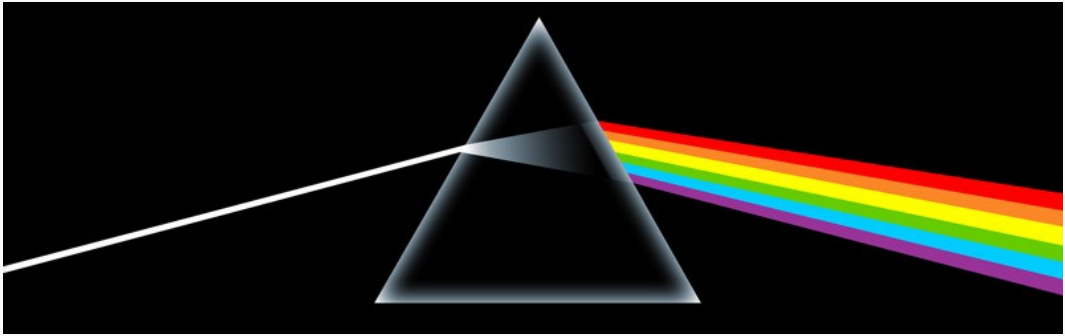
# Naive basis captures dataset, not disease

| UK Biobank | Cases |
|---|---|
| Vitiligo | 154 |
| Type 1 diabetes | 286 |
| Systemic lupus erythematosus | 366 |
| Multiple sclerosis | 1,228 |
| Coeliac disease | 1,452 |
| Ulcerative colitis | 1,795 |
| Crohn's disease | 1,032 |
| Rheumatoid arthritis | 3,730 |
| Asthma | 39,049 |

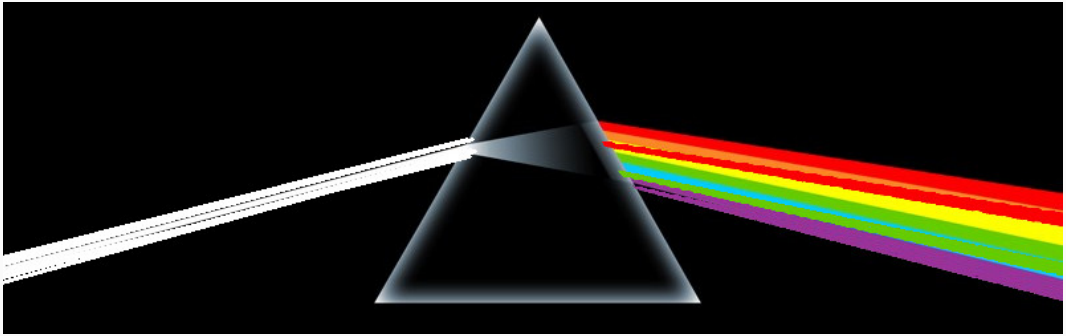



*UKBB summary data: http://www.nealelab.is/uk-biobank*

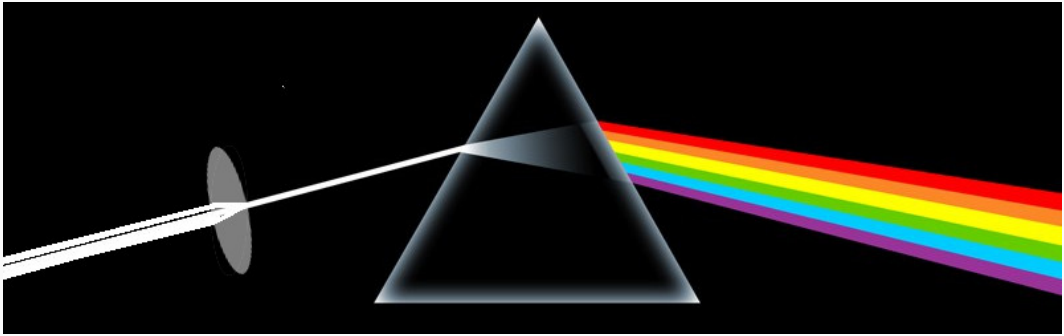
# Solution: focus data before principal components analysis



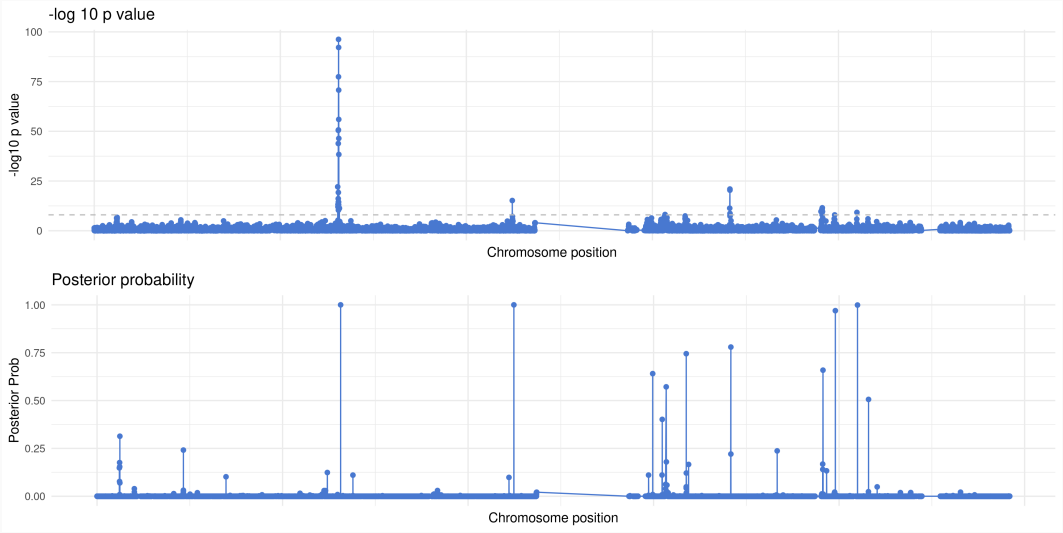
# Solution: focus data before principal components analysis



# Solution: focus data before principal components analysis

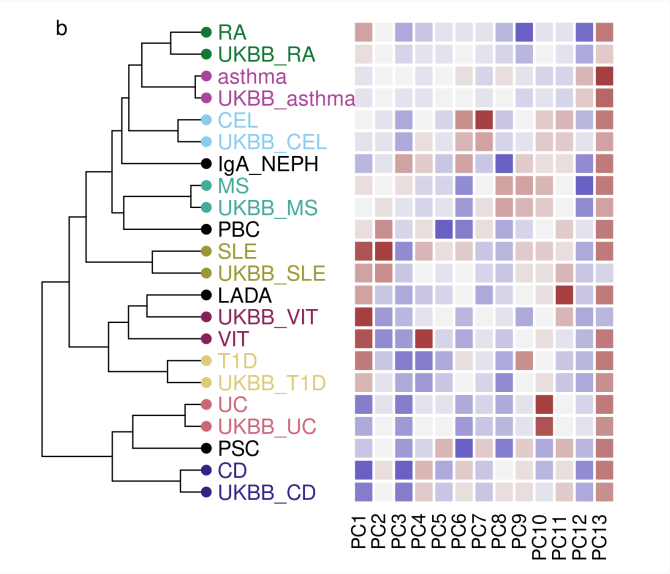


# Use fine mapping techniques to create a lens

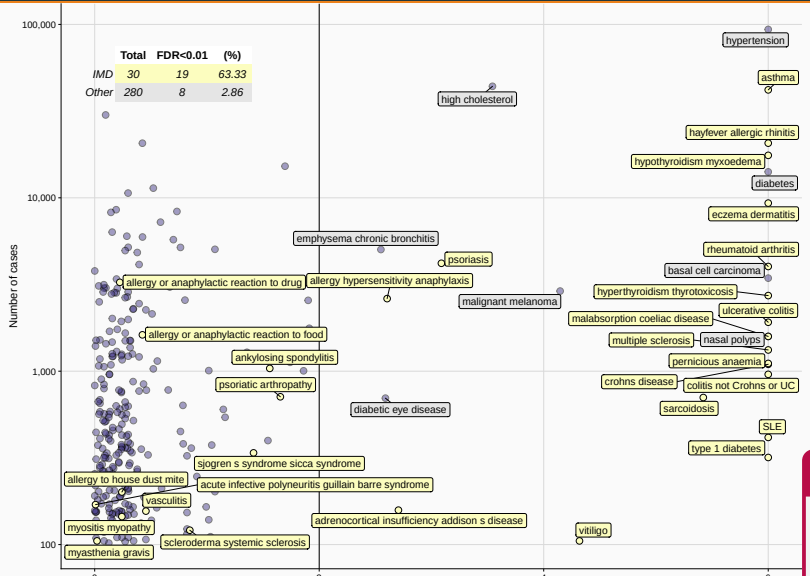


# Focused basis captures disease information

| UK Biobank | Cases |
|---|---|
| Vitiligo | 154 |
| Type 1 diabetes | 286 |
| Systemic lupus erythematosus | 366 |
| Multiple sclerosis | 1,228 |
| Coeliac disease | 1,452 |
| Ulcerative colitis | 1,795 |
| Crohn's disease | 1,032 |
| Rheumatoid arthritis | 3,730 |
| Asthma | 39,049 |



# Projection of 310 UKBB traits, global test for significance




## Significant traits

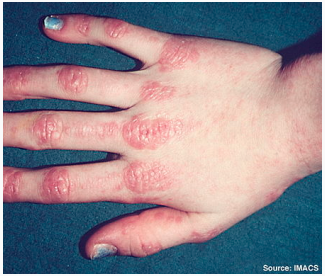
Mostly IMD

As few as 150 cases

## Idiopathic inflammatory myopathies

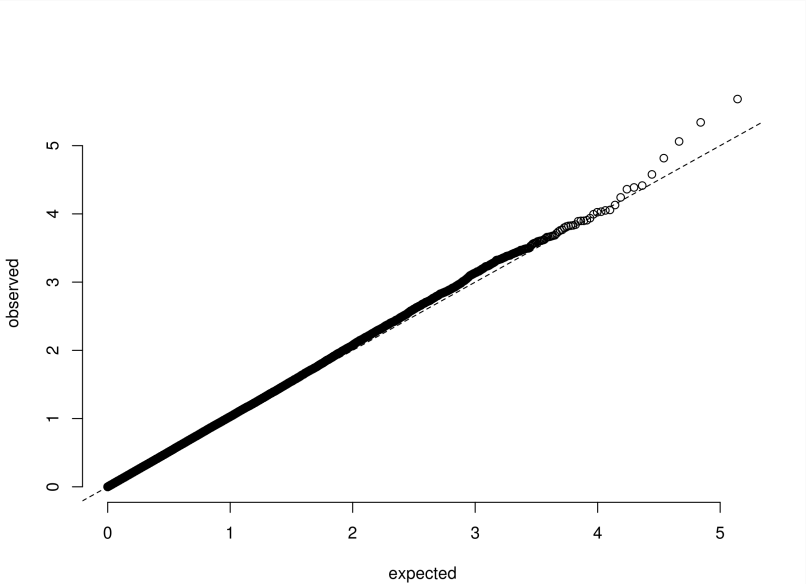
Weakness and inflammation of muscles, cause unknown
HLA association established
Incidence 2-7/million/year

| subtype | n. cases |
|---|---|
| adult dermatomyositis | 705 |
| juvenile dermatomyositis | 473 |
| polymyositis | 532 |

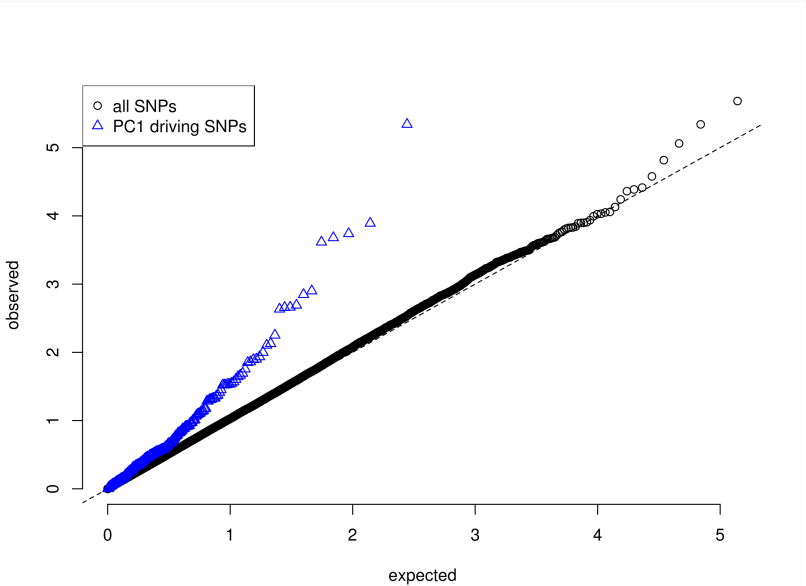



*Miller et al. 2015*

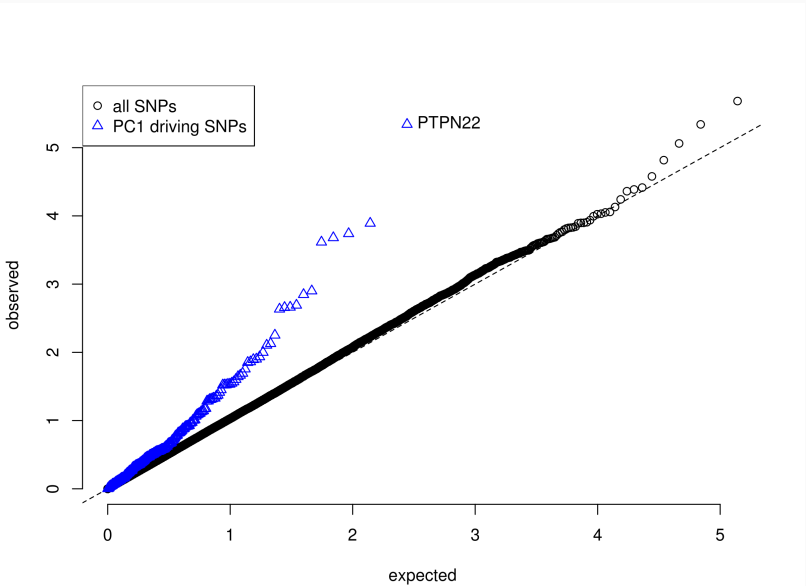
# qqplot of IIM GWAS results



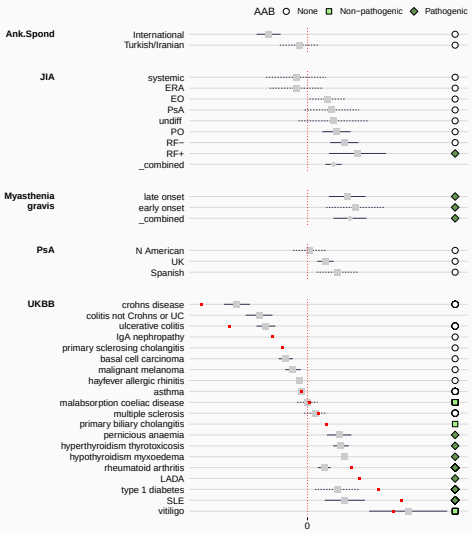
# qqplot of IIM GWAS results



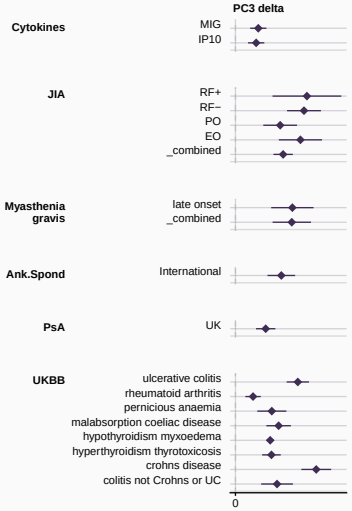
# qqplot of IIM GWAS results



# Component 1: sero+/- / autoimmune/inflammatory disease

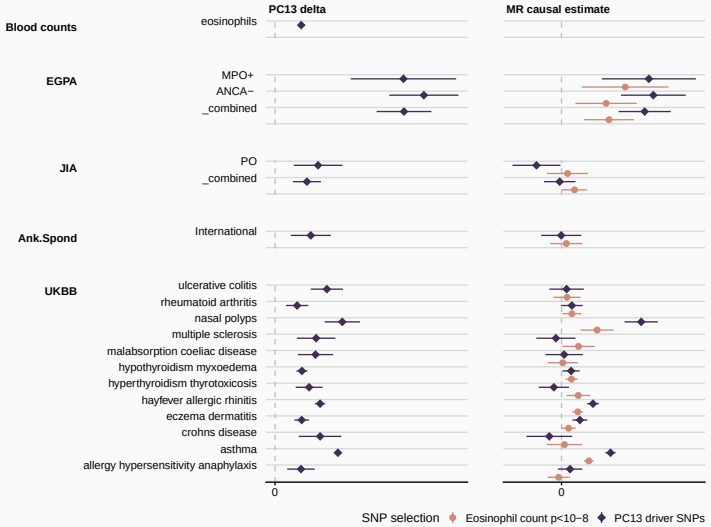


# Component 3: serum IP-10 (CXCL10), MIG (CXCL9)

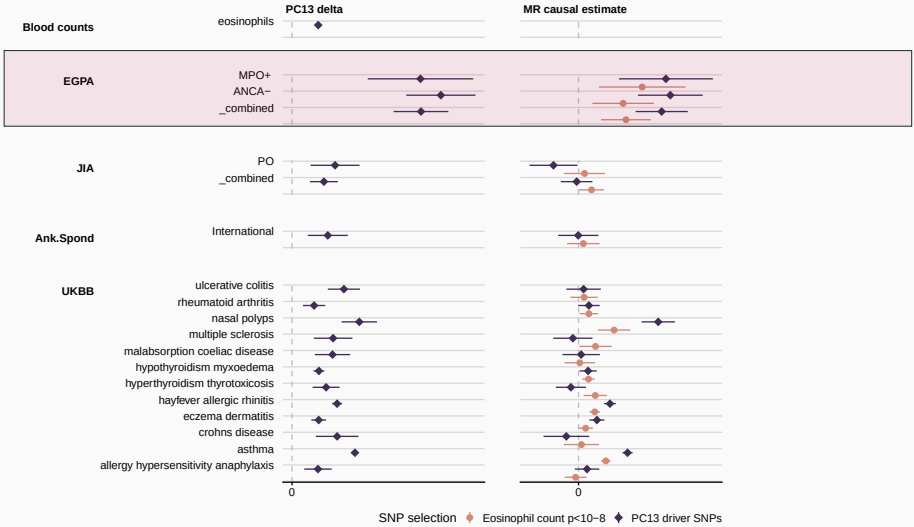



- secreted by epithelial and dendritic cells
- chemoattractants for CXCR3-expressing immune cells, including Th1
- Expression at site of autoimmune target previously implicated in the development of autoimmunity
- serum IP-10 raised in patients with recent-onset/more active autoimmune disease
- monoclonal antibody to IP-10, MDX1100, clinically effective in RA
- suggest IP-10 blockade might also be considered in patients with myasthenia gravis, JIA, AS, and sarcoidosis.

# Component 13: raised eosinophil count

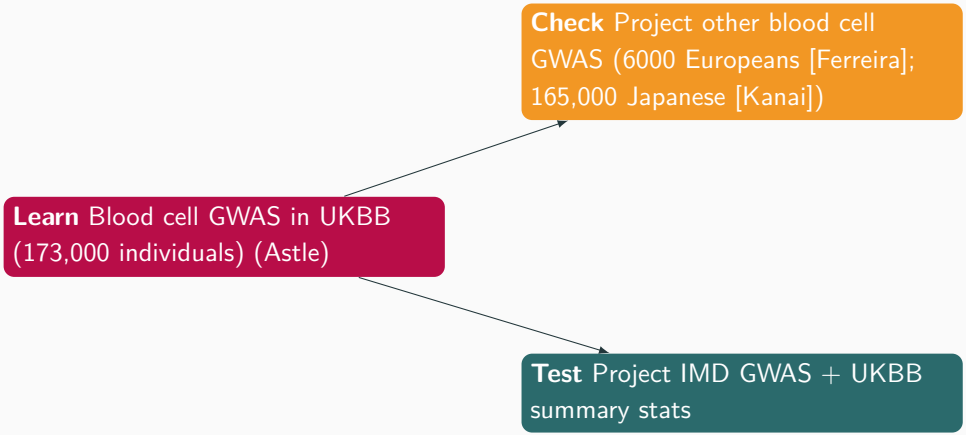


# Component 13: raised eosinophil count



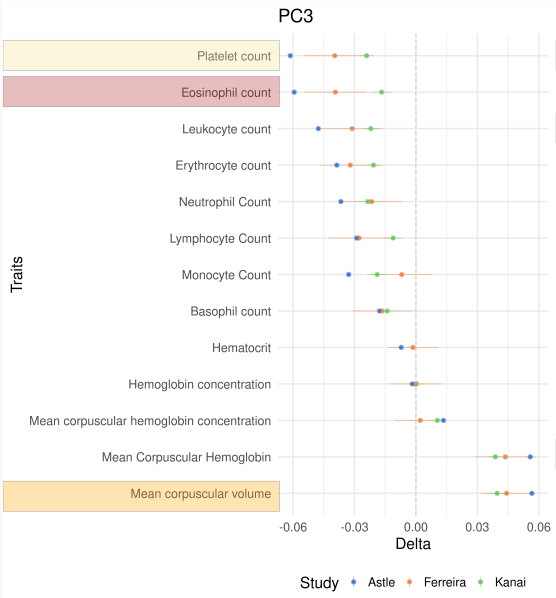
# What if we started with blood cells?

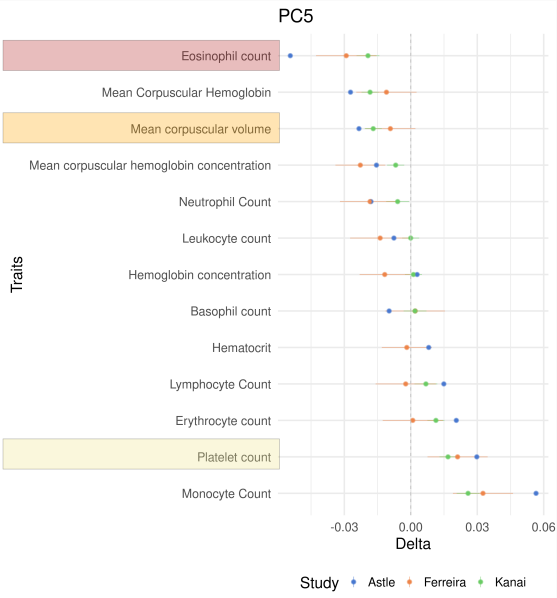
# A reduced dimension basis for blood cell count genetics



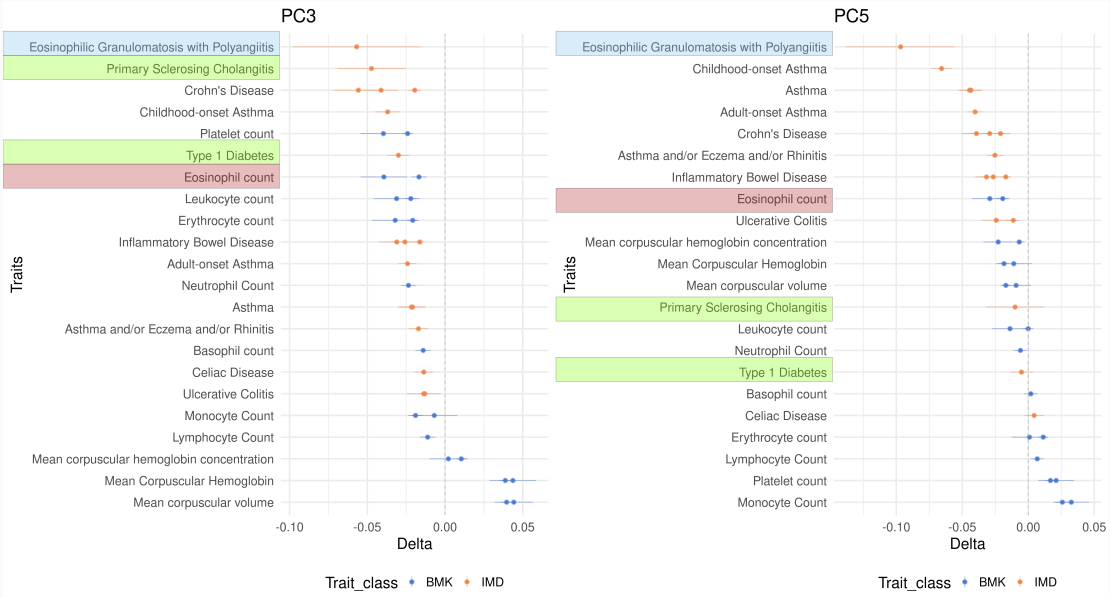

*Astle et al, Cell 2016; Kanai et al, Nat Genet 2018; Ferreira et al, AmJHG 2009*

# PC3, PC5 relate most strongly to eosinophils



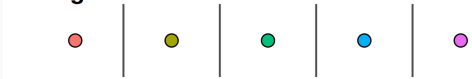



# Differential association with IMD



# Summary

**Distinguish**



What makes disease A, A?
(and not B?)

**Pool**

Transfer knowledge from better studied disease
(often to rare disease)

**Transfer**

Share resources, increase knowledge

# Summary

- Relationships between IMD are complex
- *Exploiting* relationships particularly useful to analyse genetics of rare diseases
- Must be able to control for between-study noise
- *Understanding* relationships harder
- We would like to deal directly with the mediating phenotypes, but typically only proxy or composite measures available
- Expert domain knowledge required, joint analysis of disease and composite phenotypes may help

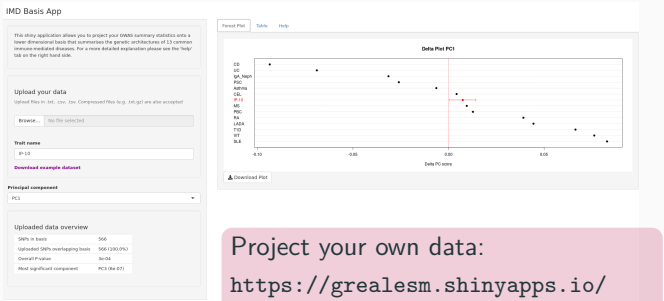
# Tools and references

IMD basis (Burren et al):
bioRxiv 2020.01.14.905869v3

Cook your own basis:
ollyburren/cupcake

Colocalisation update:
Wallace, PLoS Gen 2020

coloc v4:
chr1swallace/coloc




Project your own data:
https://grealesm.shinyapps.io/IMDbasisApp/

# Thanks to...

James Liley

Stasia Grinberg

Olly Burren

Guillermo Reales

## Patients, families, study PIs who shared data

**European Vasculitis Genetics Consortium**
Ken Smith  Paul Lyons

**Frankfurt**
Urs Christen

**MYOGEN consortum**
Fred Miller  Chris Amos
Claire Deakin  Lucy Wedderburn

wellcome

UNIVERSITY OF CAMBRIDGE

MRC | Biostatistics Unit