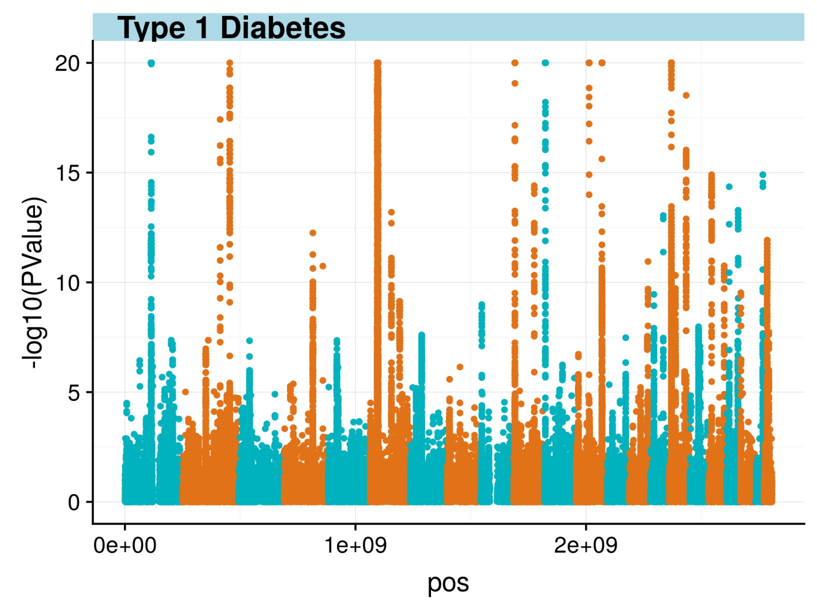
Using genetics and genomics to distinguish immune-mediated disease mechanisms and subtypes

Chris Wallace chr1swallace chr1swallace.github.io

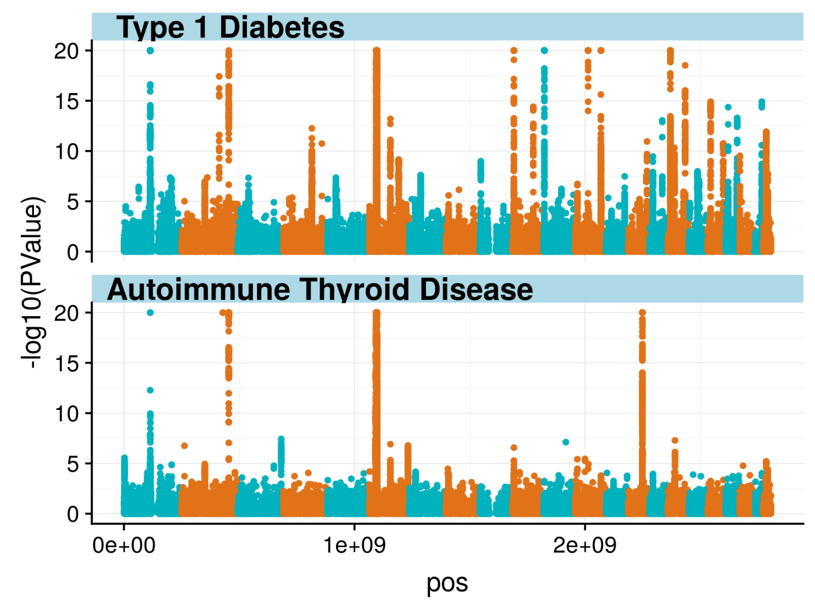
14 February 2019

University of Cambridge

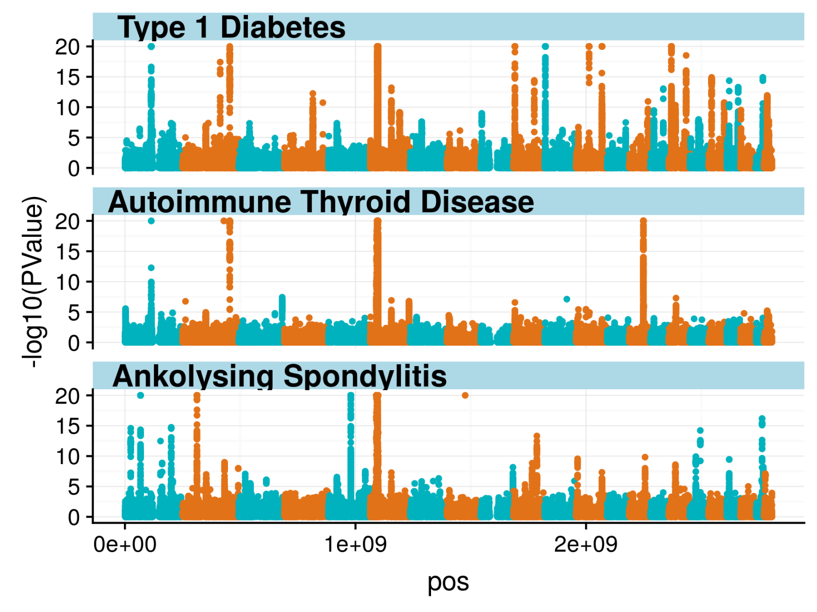
Immune mediated disease genetics



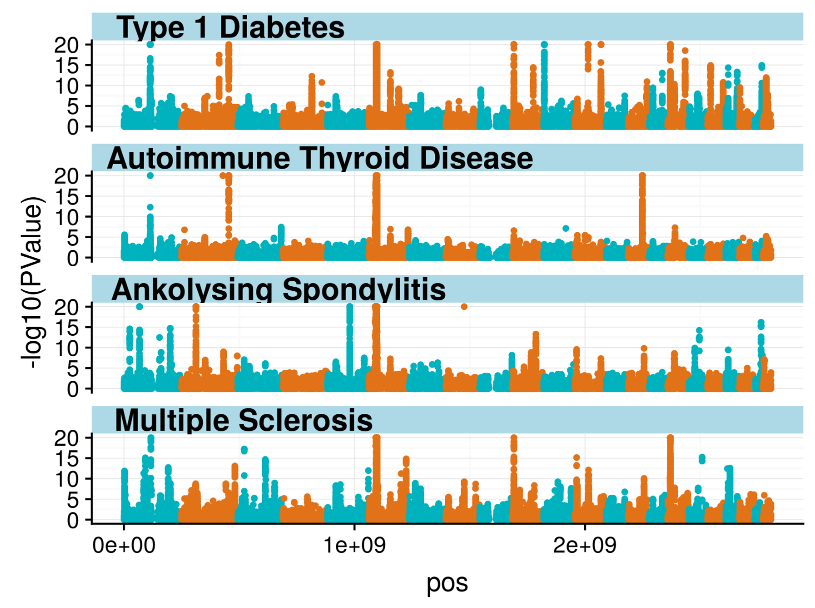
Immune mediated disease genetics



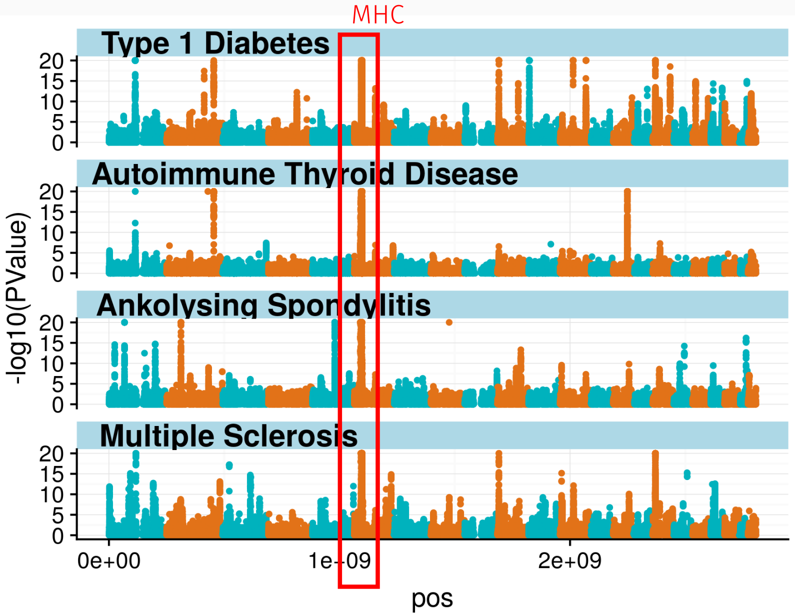
Immune mediated disease genetics



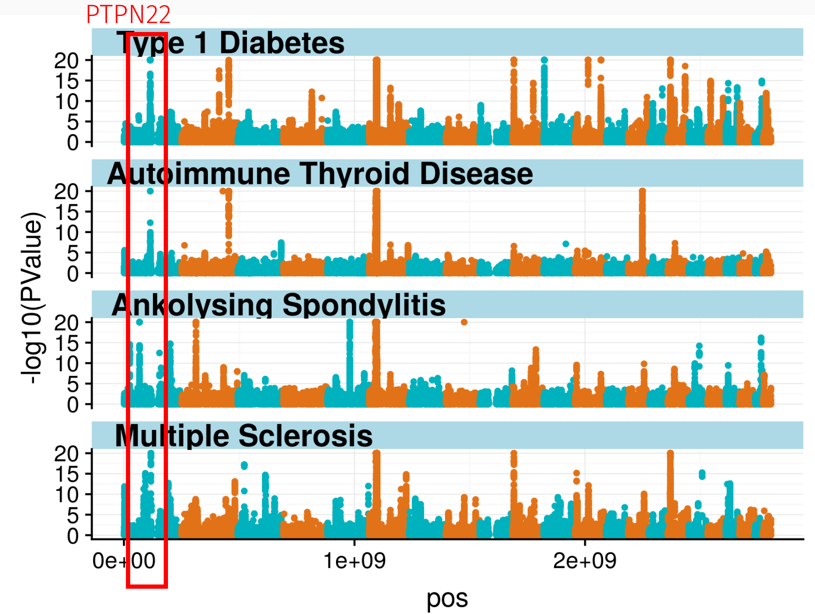
Immune mediated disease genetics



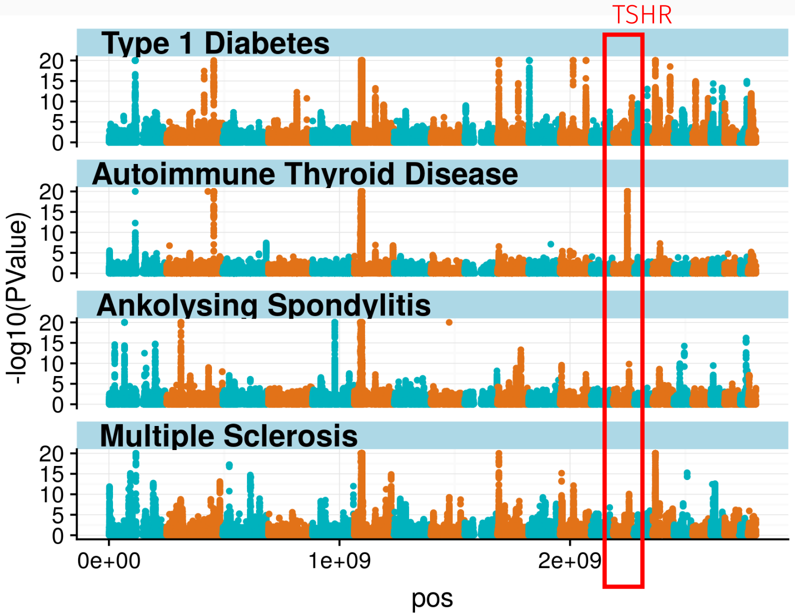
Immune mediated disease genetics



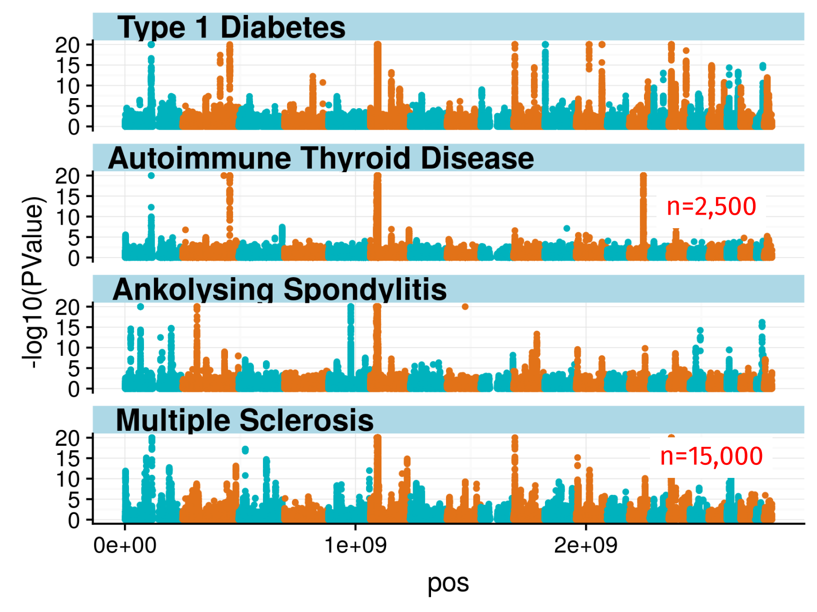
Immune mediated disease genetics



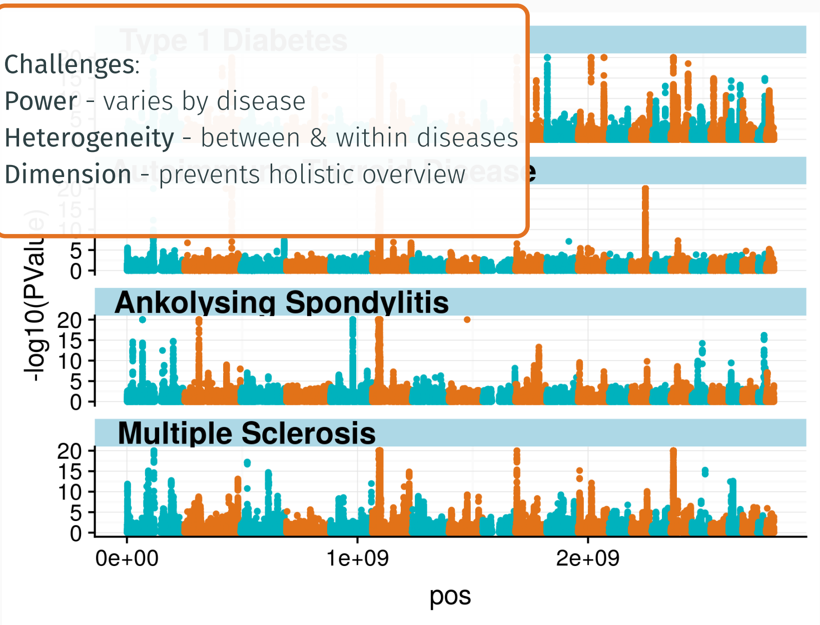
Immune mediated disease genetics



Immune mediated disease genetics



Immune mediated disease genetics



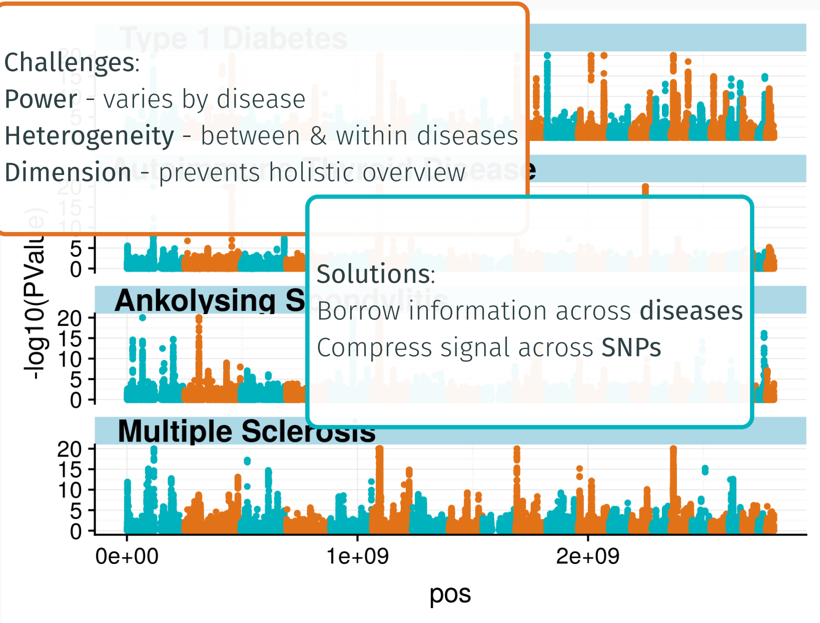
Challenges:

Power - varies by disease

Heterogeneity - between & within diseases

Dimension - prevents holistic overview

Immune mediated disease genetics



Challenges:

Power - varies by disease

Heterogeneity - between & within diseases

Dimension - prevents holistic overview

Solutions:

Borrow information across **diseases**

Compress signal across **SNPs**

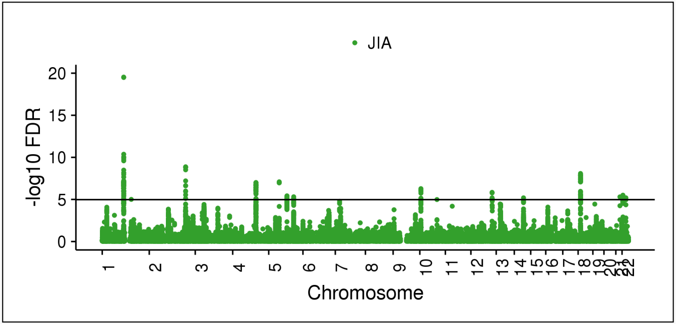
Table of contents

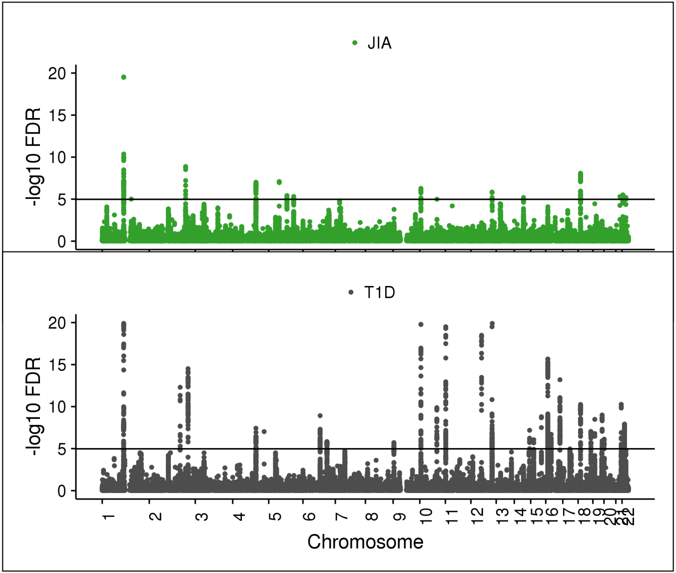
1. Sharing information between diseases to increase power

2. Compress information across SNPs in same region

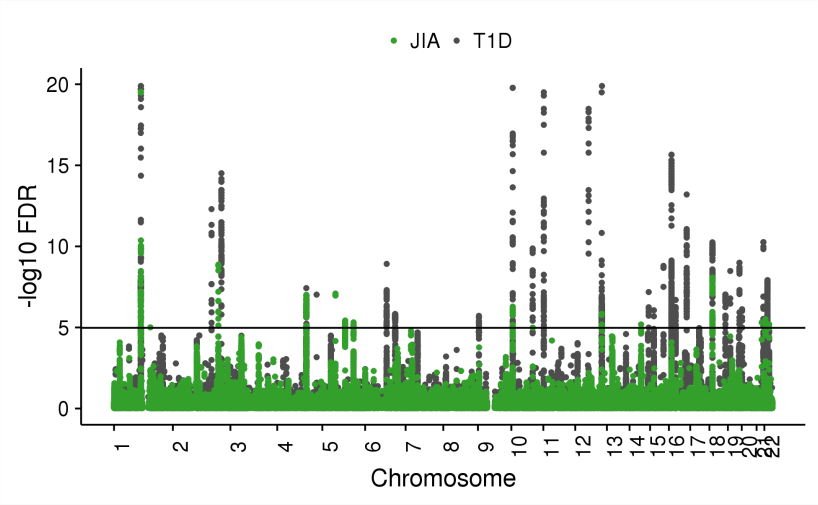
3. Informed dimension reduction

Sharing information between diseases to increase power


JIA
-log10 FDR
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

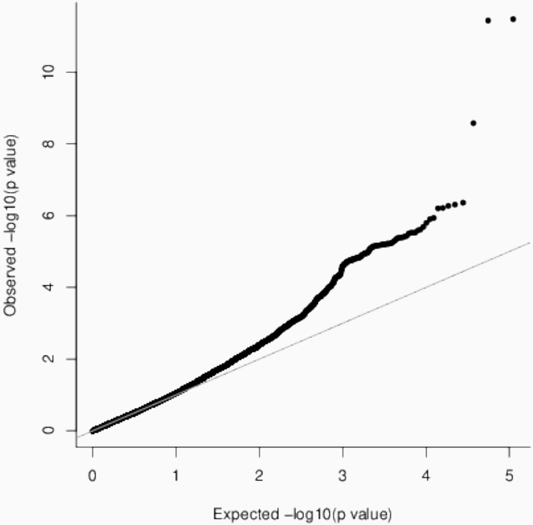


Hinks et al, 2013; Onengut et al 2015

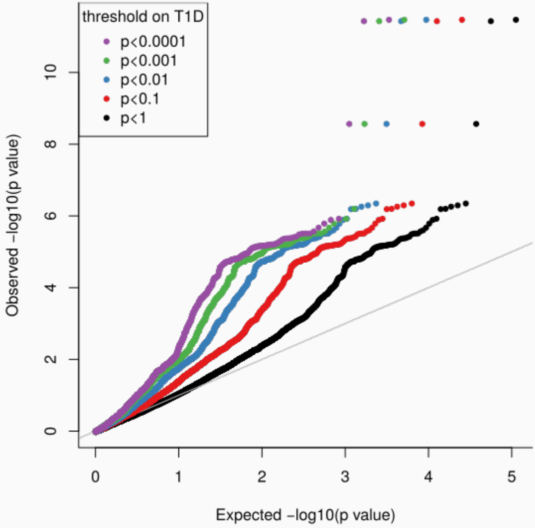


Hinks et al, 2013; Onengut et al 2015

T1D association is informative for JIA



T1D association is informative for JIA



False discovery rates $\rightarrow$ conditional false discovery rates

False discovery rate, FDR

$$Pr(H_0|P < \alpha) \propto \alpha \times Pr(H_0)$$

False discovery rates $\rightarrow$ conditional false discovery rates

False discovery rate, FDR

$$Pr(H_0|P < \alpha) \propto \alpha \times Pr(H_0)$$

Conditional false discovery rate, cFDR

$$Pr(H_0|P < \alpha, Q < \gamma) \propto \alpha \times Pr(H_0|Q < \gamma)$$

False discovery rates $\rightarrow$ conditional false discovery rates

False discovery rate, FDR

$$Pr(H_0|P < \alpha) \propto \alpha \times Pr(H_0)$$

Conditional false discovery rate, cFDR

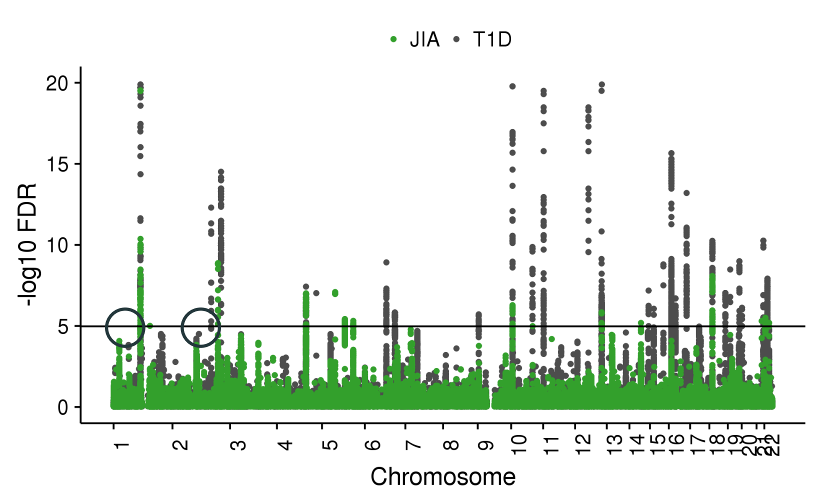
$$Pr(H_0|P < \alpha, Q < \gamma) \propto \alpha \times Pr(H_0|Q < \gamma)$$

If Q is independent of P then

$$cFDR = FDR$$

False discovery rates → conditional false discovery rates

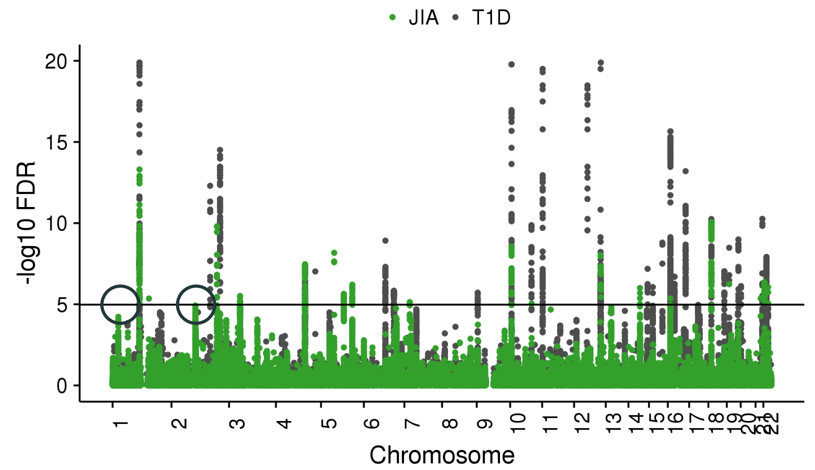
FDR



Hinks et al, 2013; Onengut et al 2015

False discovery rates → conditional false discovery rates

conditional FDR



Hinks et al, 2013; Onengut et al 2015

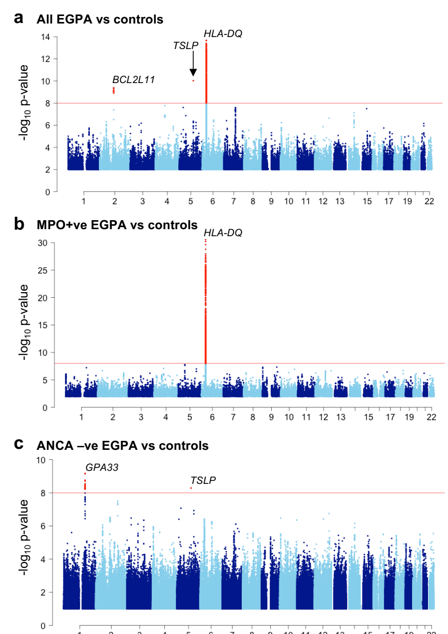
Example: Eosinophilic Granulomatosis with Polyangiitis (EGPA)

- ANCA-associated vasculitis + eosiniphilia, asthma
- incidence of 0.5-3.7 per million
- considered a single disease
- not all patients are ANCA positive

ANCA positive (MPO)	161
ANCA negative	358
Controls	6717

Lyons et al, biorxiv 491837

Eosinophilic Granulomatosis with Polyangiitis (EGPA)

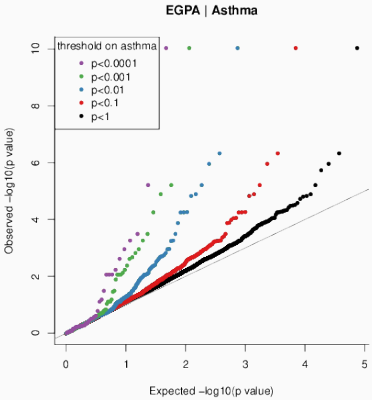


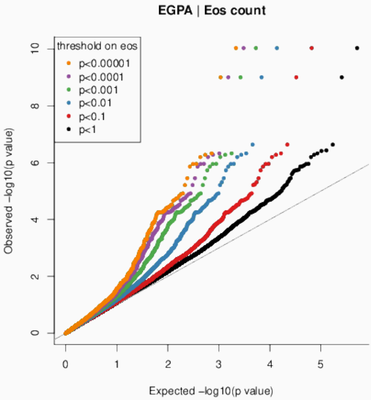
all EGPA

MPO+ EGPA

ANCA- EGPA

Both asthma and eosinophilia are informative for EGPA



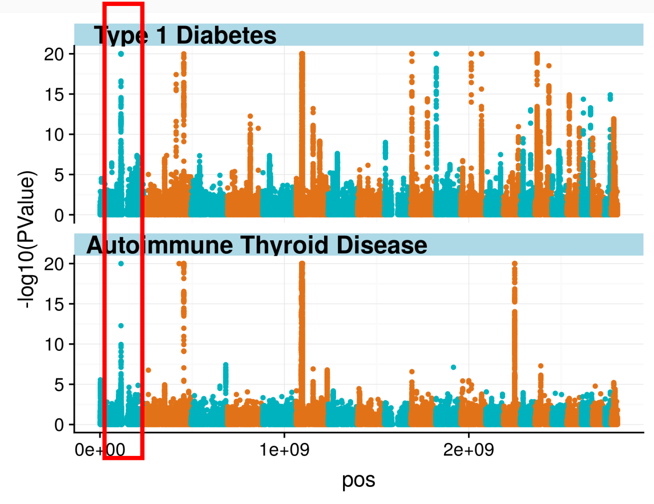


EGPA: Genetically distinct clinical subsets, shared associations with asthma and eosinophil count

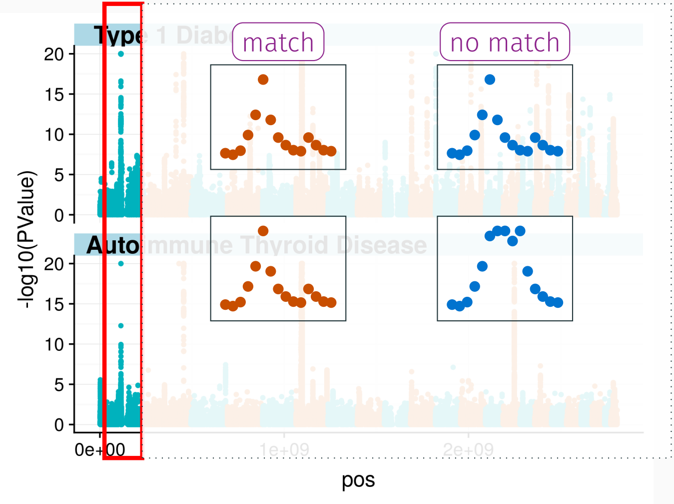
	MPO+ EGPA	ANCA- EGPA
Distinct associations	HLA-DQ	GPA33, (IL5)
Shared associations	BIM, TSLP	
Conditional associations		
Asthma	IL5, BACH2, 10p14 (GATA3?)	
Eosiniphilia	CDK6, SOCS1, LPP, TBX3	
Differential symptoms (%)		
Glomerulonephritis	29	9
Neuropathy	79	57
Treatment response (%)		
Rituximab response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

Compress information across SNPs in same region

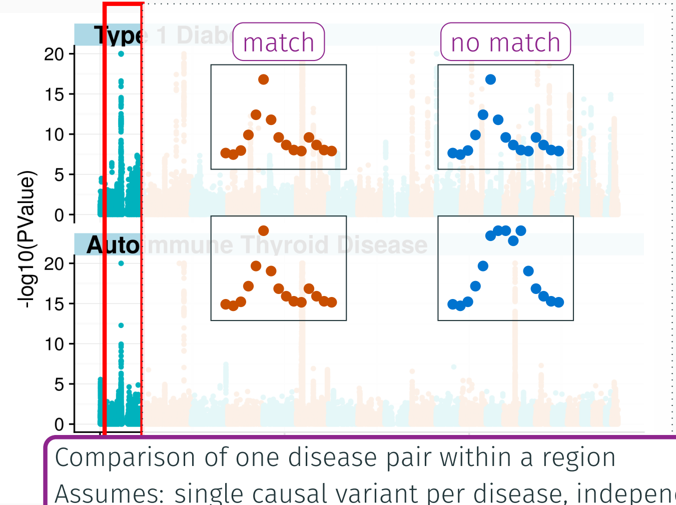
Colocalisation



Colocalisation



Colocalisation



Comparison of one disease pair within a region

Assumes: single causal variant per disease, independent datasets

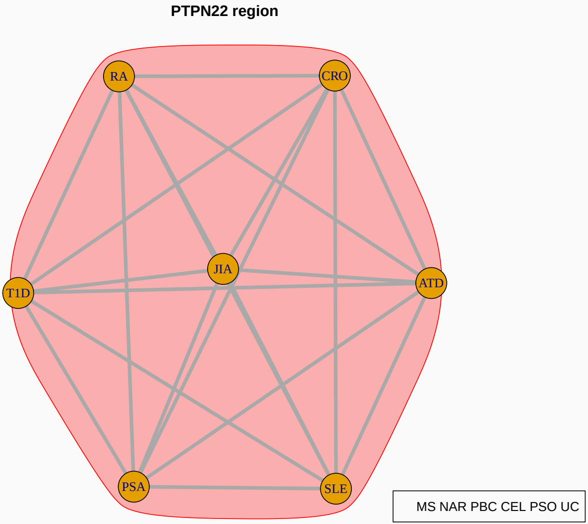
New: multiple uncorrelated causal variants, dependent controls

Giambartolomei et al, 2014. chr1swallace/coloc

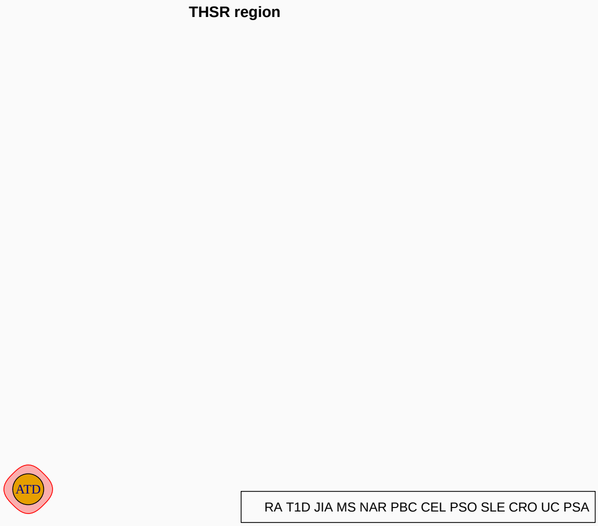
ImmunoChip: 13 diseases, 188 genetic regions with dense coverage

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
Psorias	Tsoi	10,588	22,806
Ankolysing spondylitis	IGAS	10,619	15,145
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

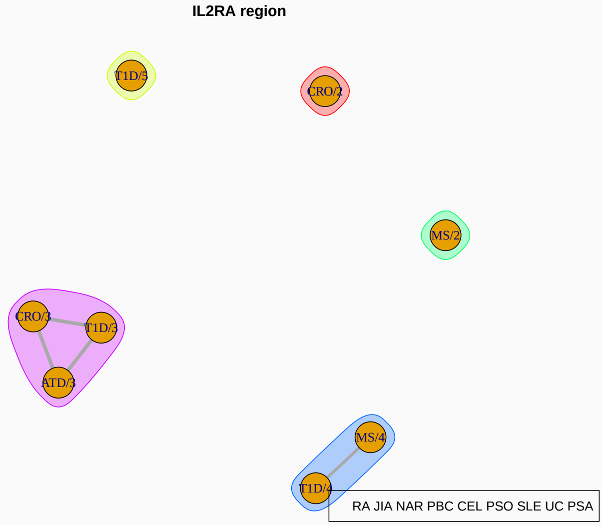
Shared signal in *PTPN22* region



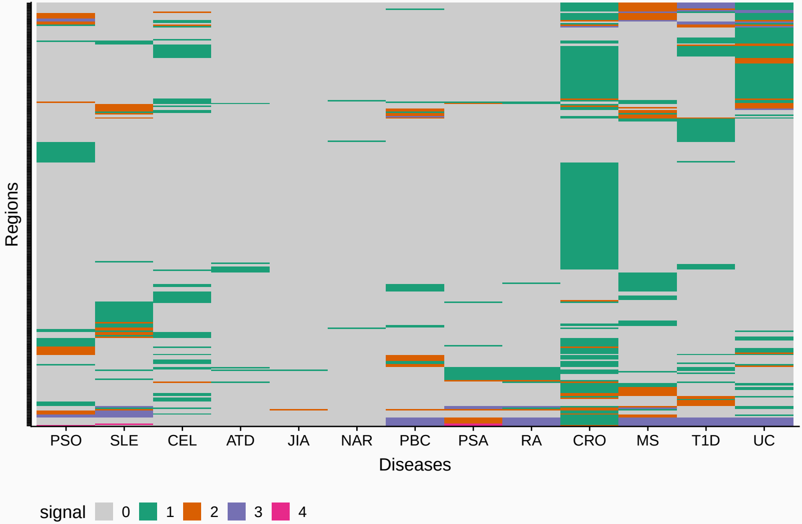
Single ATD signal in *TSHR* region



Multiple signals in *IL2RA* region

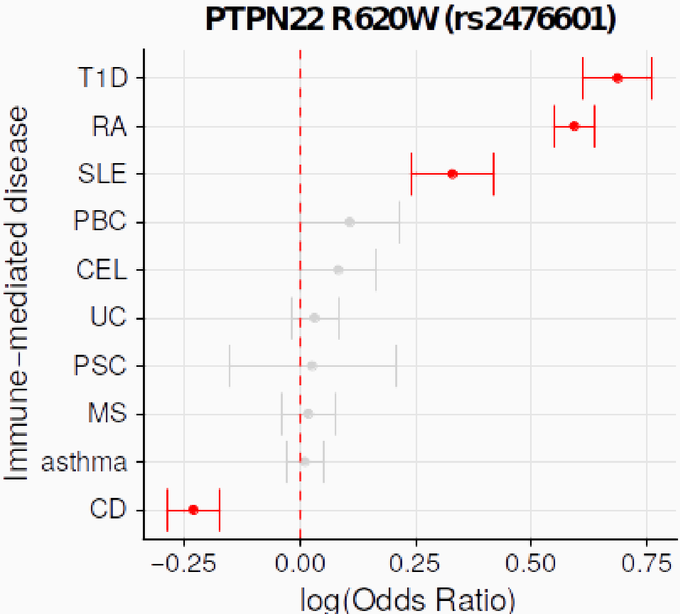


Overview

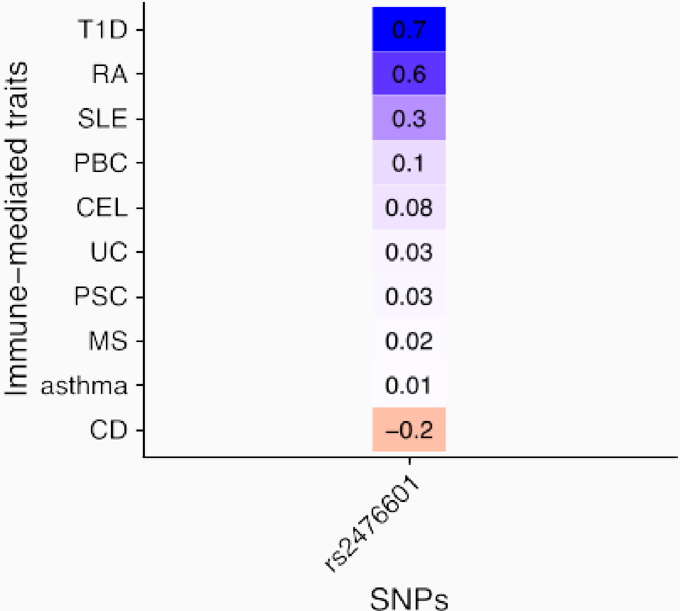

Regions
PSO
SLE
CEL
ATD
JIA
NAR
PBC
PSA
RA
CRO
MS
T1D
UC
Diseases
signal
0
1
2
3
4

Informed dimension reduction

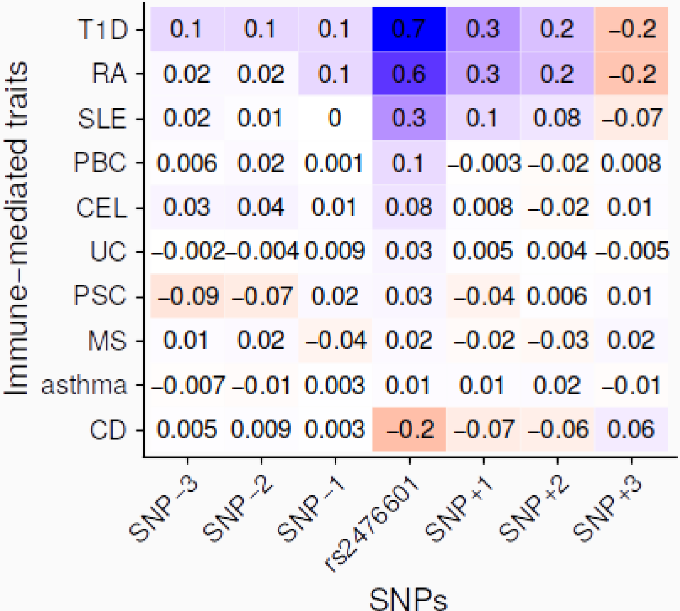
Heterogeneity across diseases



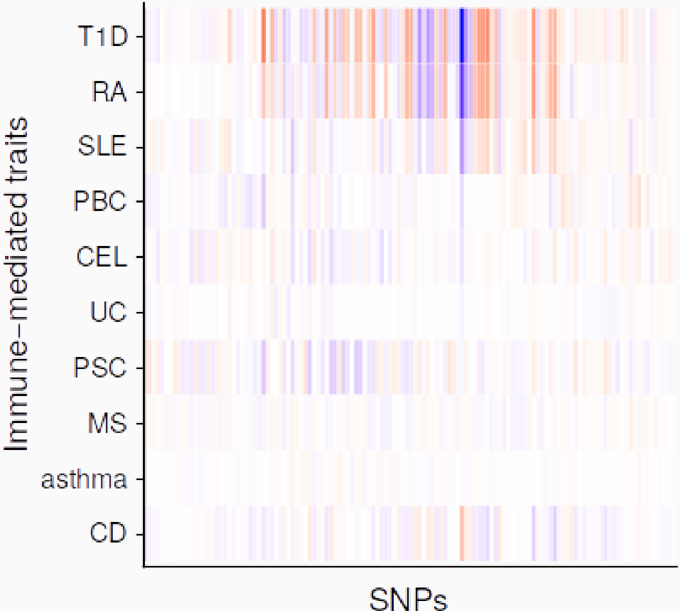
Heterogeneity across diseases



Heterogeneity across diseases

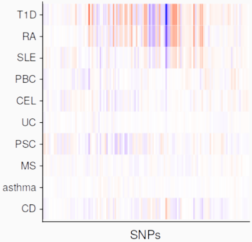


Heterogeneity across diseases

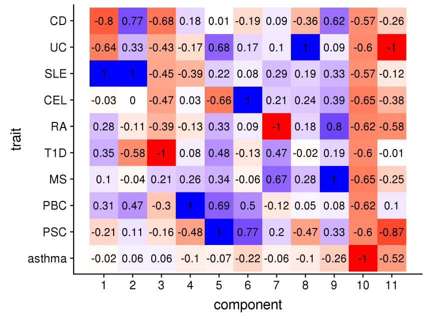


Principal components analysis to generate a new “basis”

PCA

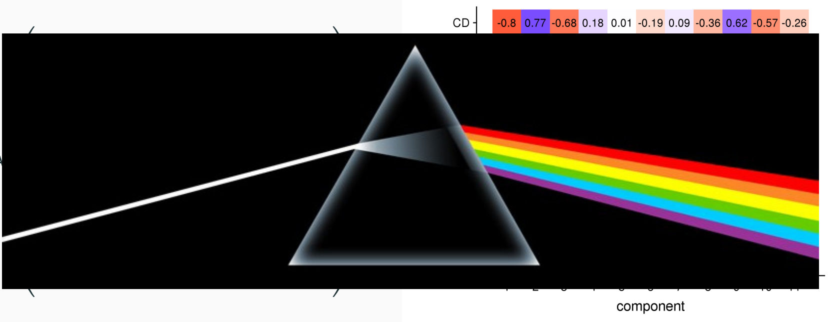


=



Principal components analysis to generate a new “basis”

PCA

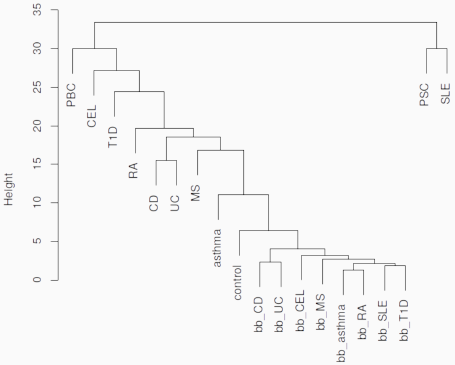


Basis diseases

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
Primary biliary cholangitis	Cordell	2764	10475

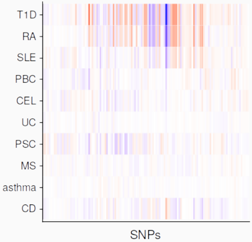
Naive basis captures dataset, not disease

UK Biobank	Cases
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

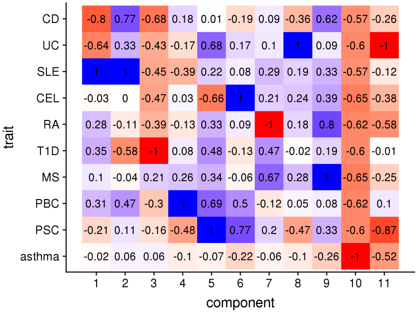


Solution: focus data before principal components analysis

PCA

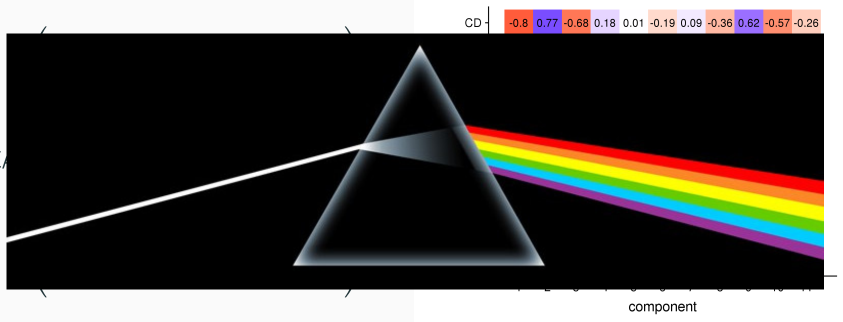


=



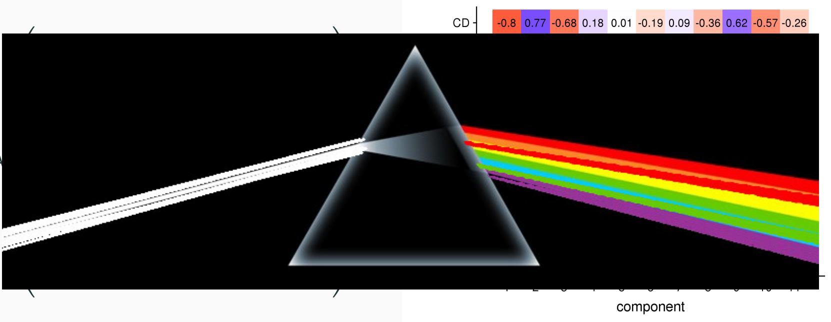
Solution: focus data before principal components analysis

PC



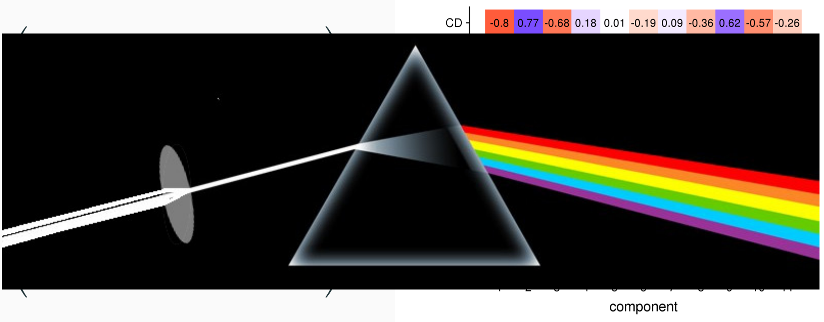
Solution: focus data before principal components analysis

PCA



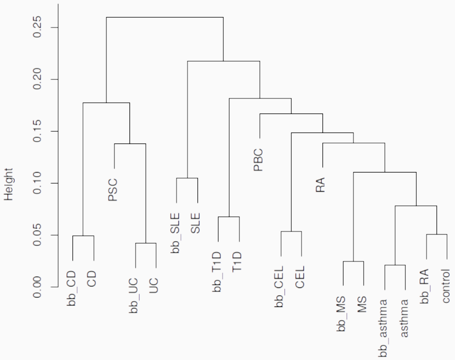
Solution: focus data before principal components analysis

PCA

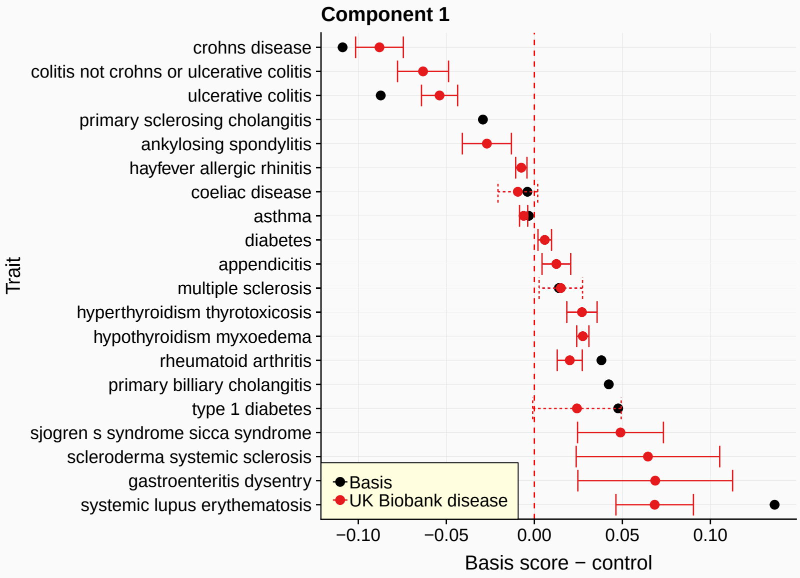


Focused basis captures disease information

UK Biobank	Cases
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

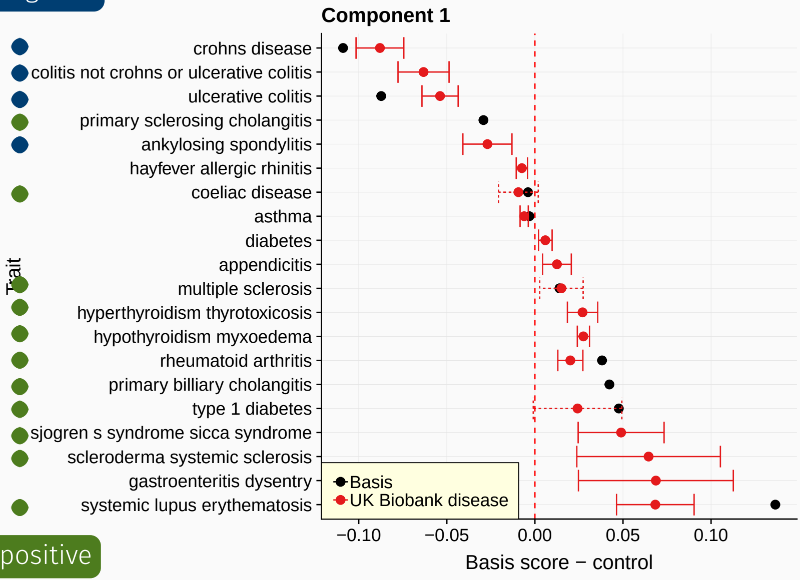


Component 1: sero+/- disease, innate/adaptive



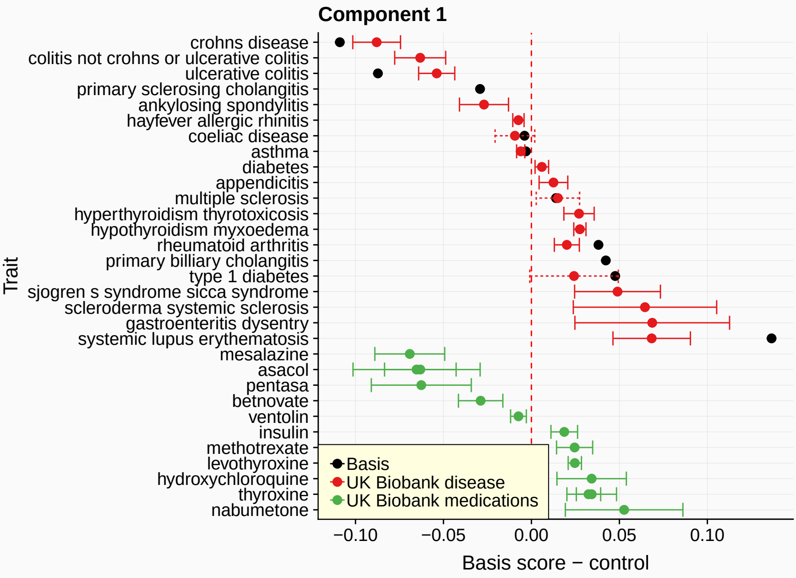
Component 1: sero+/- disease, innate/adaptive

seronegative

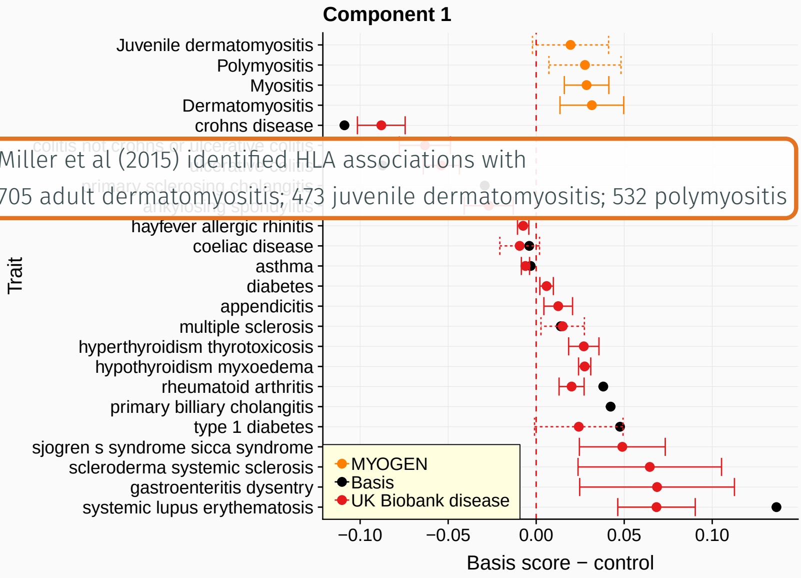


seropositive

Component 1: sero+/- disease, innate/adaptive



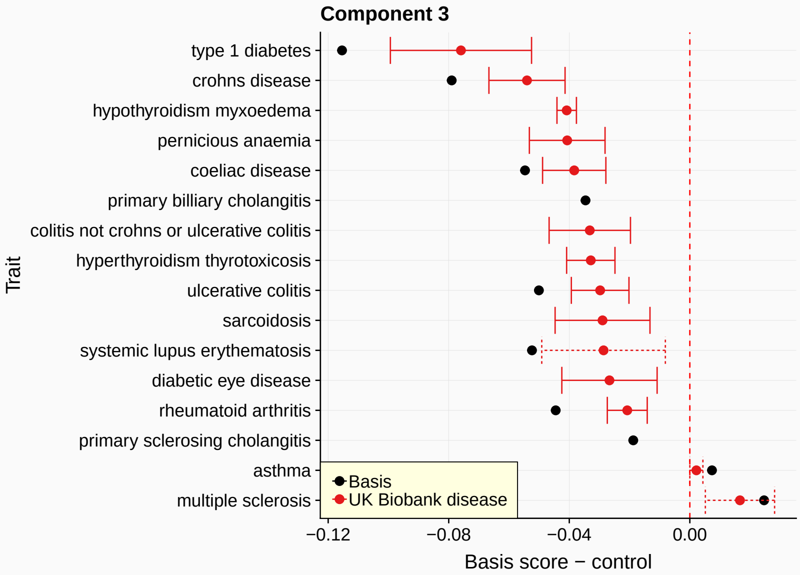
Component 1: sero+/- disease, innate/adaptive



Miller et al (2015) identified HLA associations with

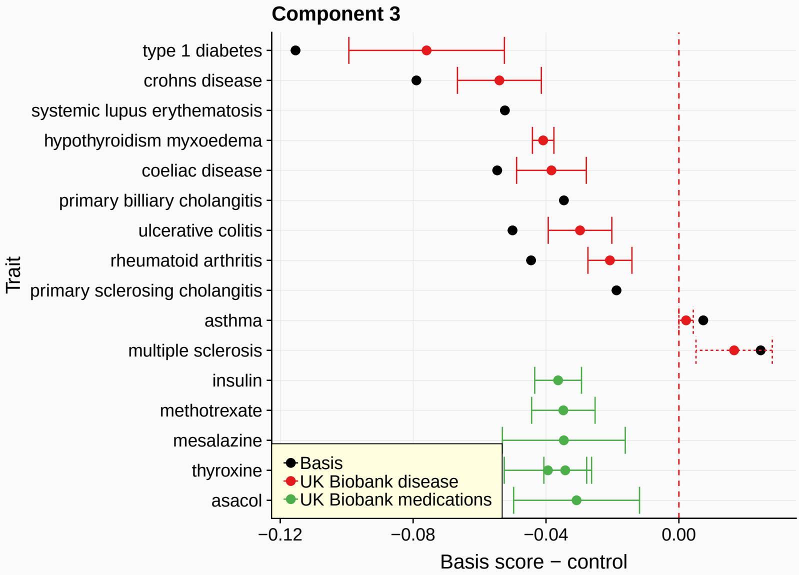
705 adult dermatomyositis; 473 juvenile dermatomyositis; 532 polymyositis

Component 3



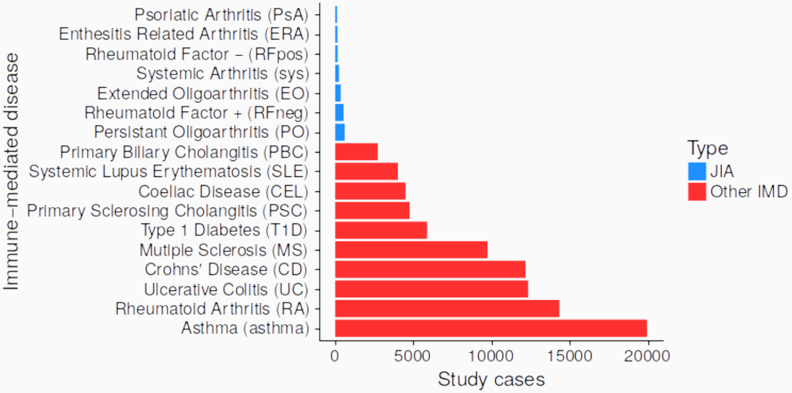
Component 3
type 1 diabetes
crohns disease
hypothyroidism myxoedema
pernicious anaemia
coeliac disease
primary billiary cholangitis
colitis not crohns or ulcerative colitis
hyperthyroidism thyrotoxicosis
ulcerative colitis
sarcoidosis
systemic lupus erythematosis
diabetic eye disease
rheumatoid arthritis
primary sclerosing cholangitis
asthma
multiple sclerosis
Trait
Basis
UK Biobank disease
−0.12
−0.08
−0.04
0.00
Basis score − control

Component 3



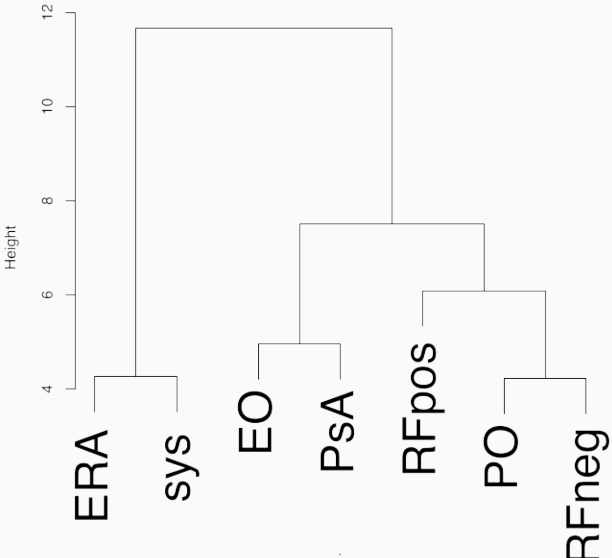
Component 3
Trait
type 1 diabetes
crohns disease
systemic lupus erythematosis
hypothyroidism myxoedema
coeliac disease
primary billiary cholangitis
ulcerative colitis
rheumatoid arthritis
primary sclerosing cholangitis
asthma
multiple sclerosis
insulin
methotrexate
mesalazine
thyroxine
asacol
Basis
UK Biobank disease
UK Biobank medications
−0.12
−0.08
−0.04
0.00
Basis score − control

Component 3

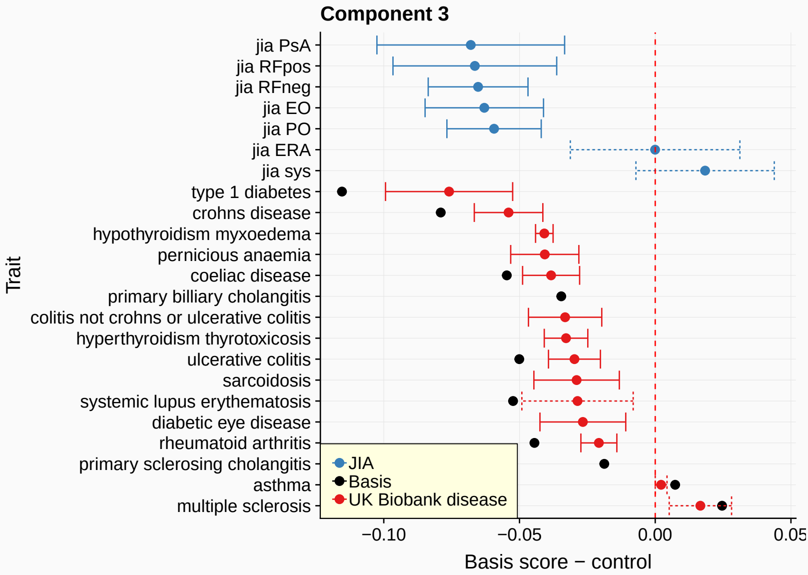


UK JIA genetics consortium

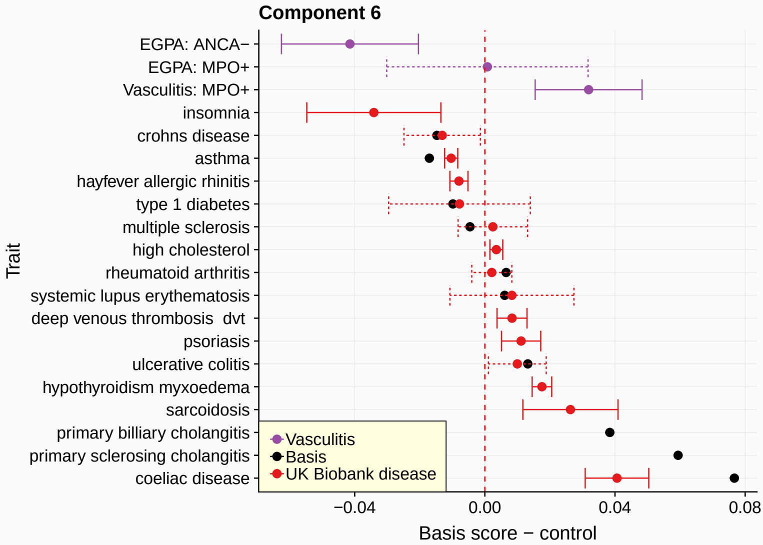
Component 3



Component 3

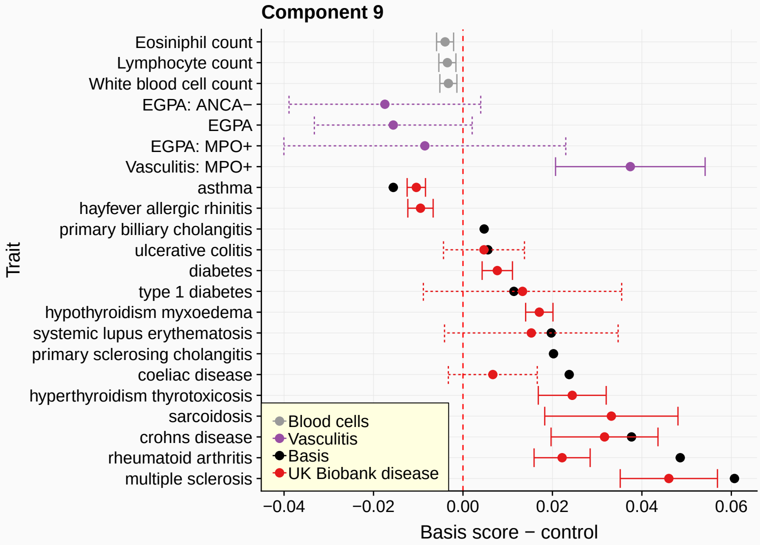

Component 3
jia PsA
jia RFpos
jia RFneg
jia EO
jia PO
jia ERA
jia sys
type 1 diabetes
crohns disease
hypothyroidism myxoedema
pernicious anaemia
coeliac disease
primary billiary cholangitis
colitis not crohns or ulcerative colitis
hyperthyroidism thyrotoxicosis
ulcerative colitis
sarcoidosis
systemic lupus erythematosis
diabetic eye disease
rheumatoid arthritis
primary sclerosing cholangitis
asthma
multiple sclerosis
Trait
JIA
Basis
UK Biobank disease
−0.10
−0.05
0.00
0.05
Basis score − control

Vasculitis



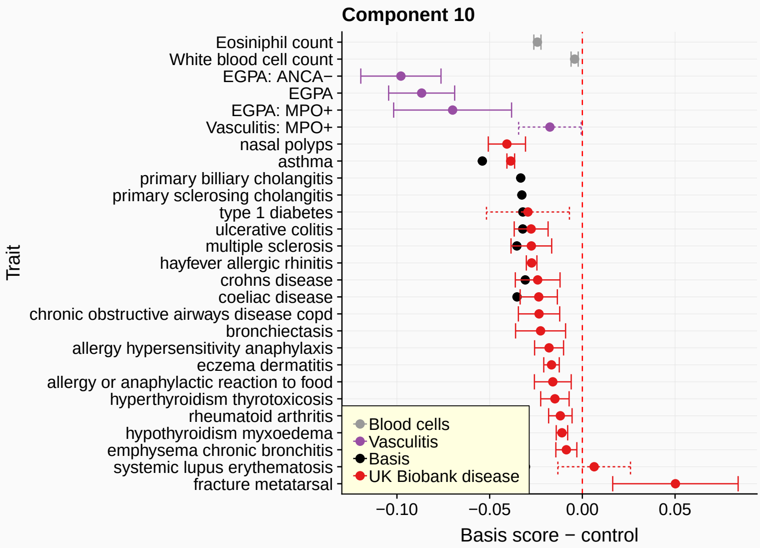
Component 6
EGPA: ANCA−
EGPA: MPO+
Vasculitis: MPO+
insomnia
crohns disease
asthma
hayfever allergic rhinitis
type 1 diabetes
multiple sclerosis
high cholesterol
rheumatoid arthritis
systemic lupus erythematosis
deep venous thrombosis dvt
psoriasis
ulcerative colitis
hypothyroidism myxoedema
sarcoidosis
primary billiary cholangitis
primary sclerosing cholangitis
coeliac disease
Trait
Vasculitis
Basis
UK Biobank disease
−0.04
0.00
0.04
0.08
Basis score − control

Vasculitis



Astle et al, 2016

Vasculitis



Astle et al, 2016

Summary

- Genetic sharing across immune mediated diseases is pervasive, but complex

Summary

- Genetic sharing across immune mediated diseases is pervasive, but complex
- Sharing **should** be exploited to learn about less common diseases

Summary

- Genetic sharing across immune mediated diseases is pervasive, but complex
- Sharing **should** be exploited to learn about less common diseases
- Region-wise overlap analysis useful to “zoom in”

Summary

- Genetic sharing across immune mediated diseases is pervasive, but complex
- Sharing **should** be exploited to learn about less common diseases
- Region-wise overlap analysis useful to “zoom in”
- Dimension reduction obscures individual effects, offers manageable holistic view of multiple diseases

Summary

- Genetic sharing across immune mediated diseases is pervasive, but complex
- Sharing **should** be exploited to learn about less common diseases
- Region-wise overlap analysis useful to “zoom in”
- Dimension reduction obscures individual effects, offers manageable holistic view of multiple diseases
  - Focusing on OR rather than significance puts everything on same scale
  - Improves power: **many** fewer tests, stronger prior for association
  - Components are directions, from x to y, but interpretation difficult
  - Given difference on component, can identify driving SNPs - “zoom out”

Thanks to...

James Liley

Stasia Grinberg

Olly Burren

Patients, families, study PIs who shared data

Sample Cohorts

- UK JIA Genetics Consortium
- Childhood Arthritis Prospective Study (CAPS)
- Childhood Arthritis Response to Medication Study (CHARMS)
- MYOGEN consortium
- European Vasculitis Genetics Consortium

Cambridge

Ken Smith, Paul Lyons

MYOGEN consortum

Fred Miller, Chris Amos

Manchester

Wendy Thomson, Anne Hinks, Joanna Cobb, John Bowes, Annie Yarwood, Sam Smith, Kimme Hyrich

University College London

Lucy Wedderburn, Claire Deakin

wellcome

UNIVERSITY OF CAMBRIDGE

MRC Biostatistics Unit