


Needles, haystacks, and genetic fine mapping in immune-mediated diseases

Chris Wallace  chr1swallace  chr1swallace.github.io



Jenn Asimit



Dan Rainbow



Linda Wicker

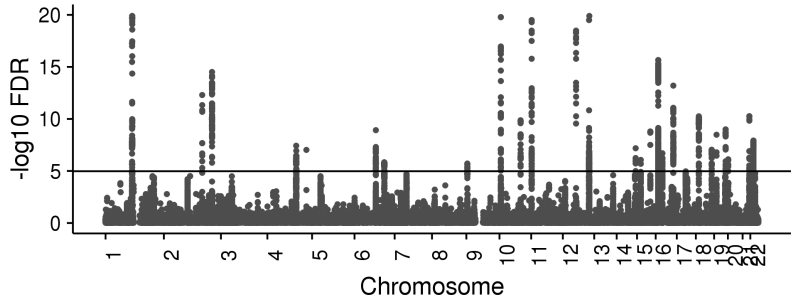


UNIVERSITY OF
CAMBRIDGE

MRC

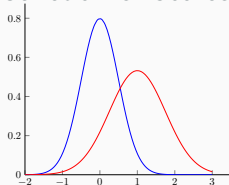
Biostatistics Unit

Manhattan plots (haystack plots?)

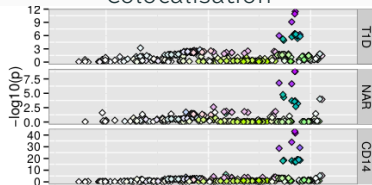


Post-GWAS

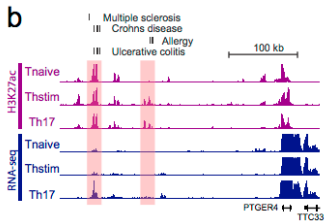
Genetic Risk Scores



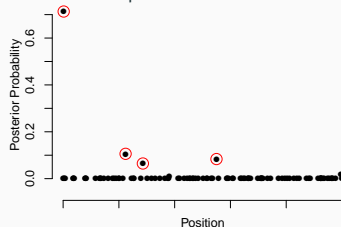
Colocalisation



Locational integration



Fine map causal variants

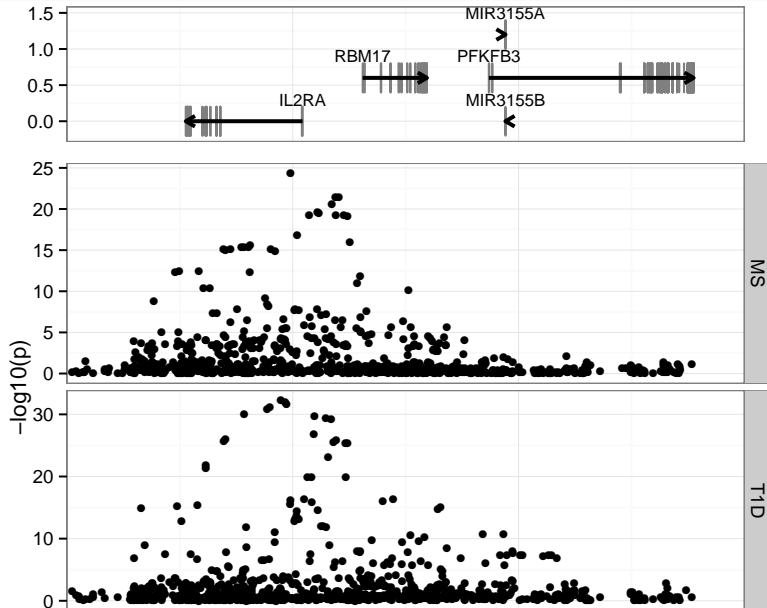


Overview

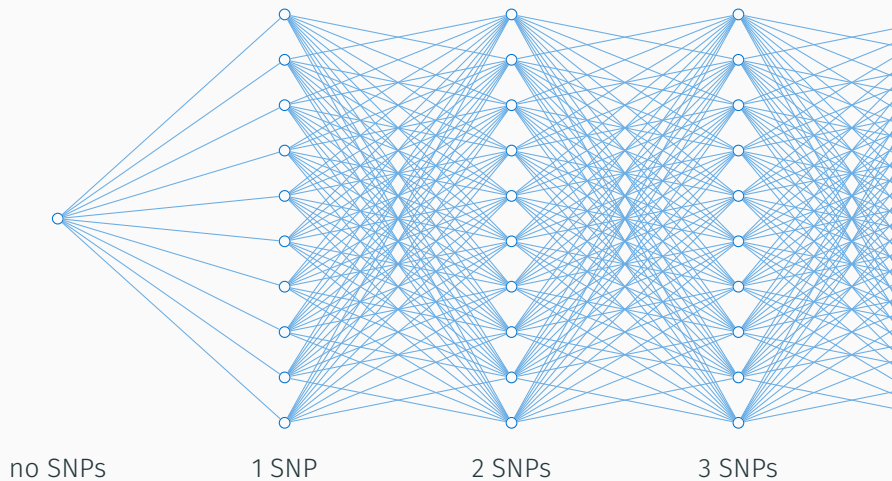
- Fine mapping causal variants
- Multiple-disease fine mapping
- Functional validation of causal effects on *IL2RA*

Fine mapping causal variants

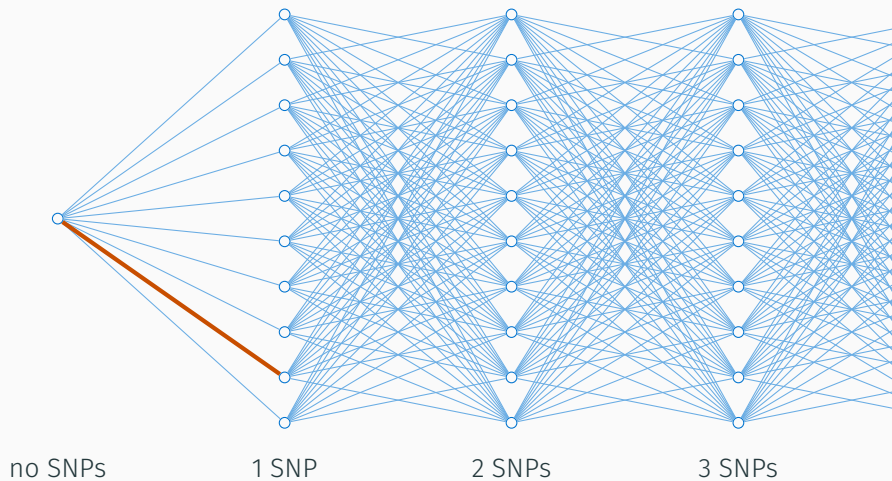
Association of MS and T1D in *IL2RA* region



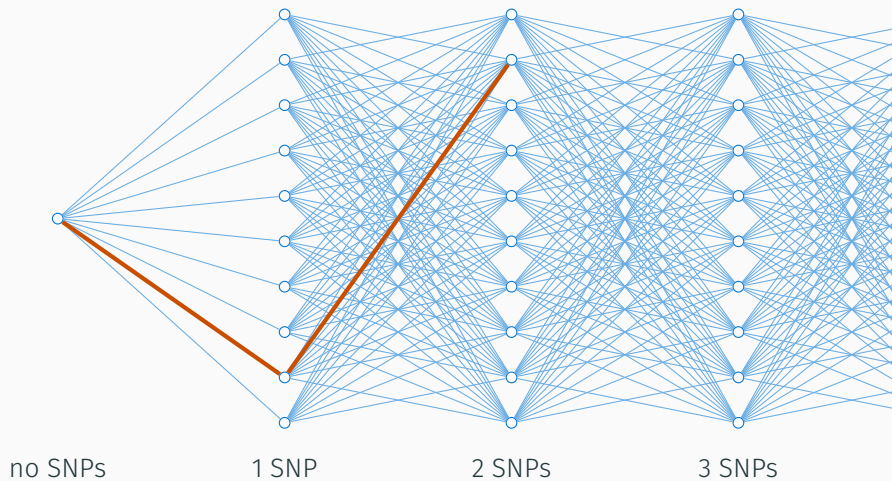
Stepwise search



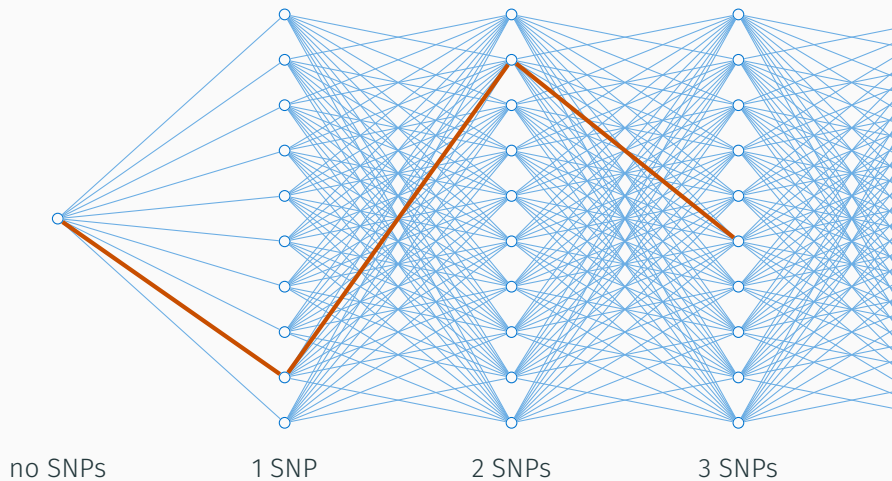
Stepwise search



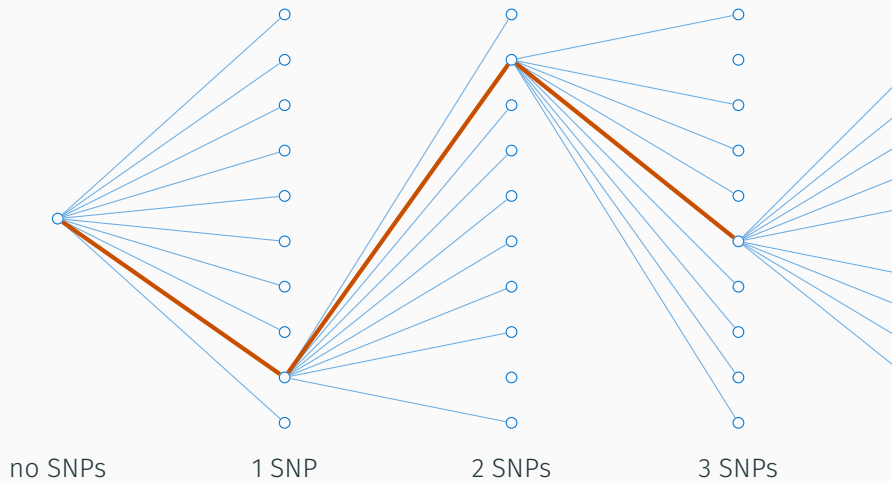
Stepwise search



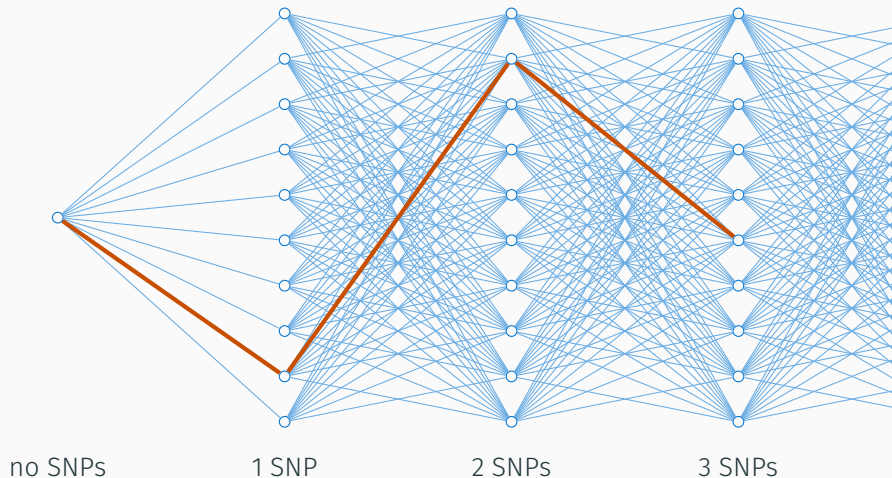
Stepwise search



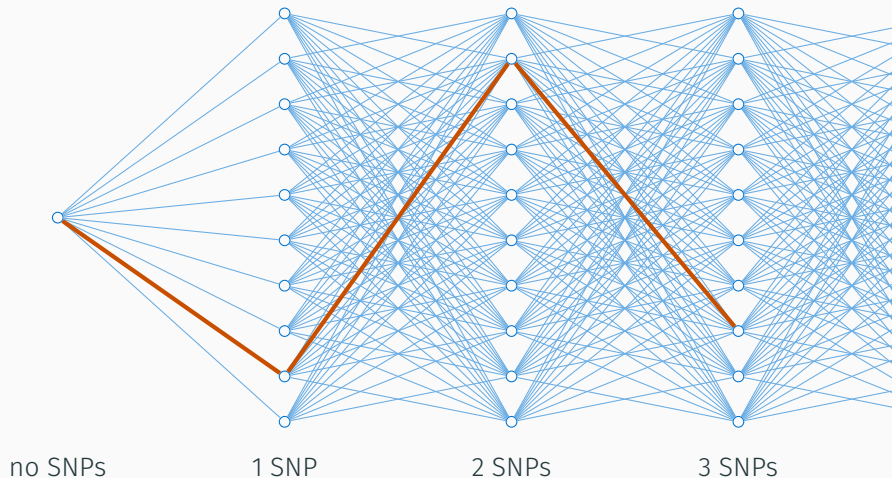
Stepwise search



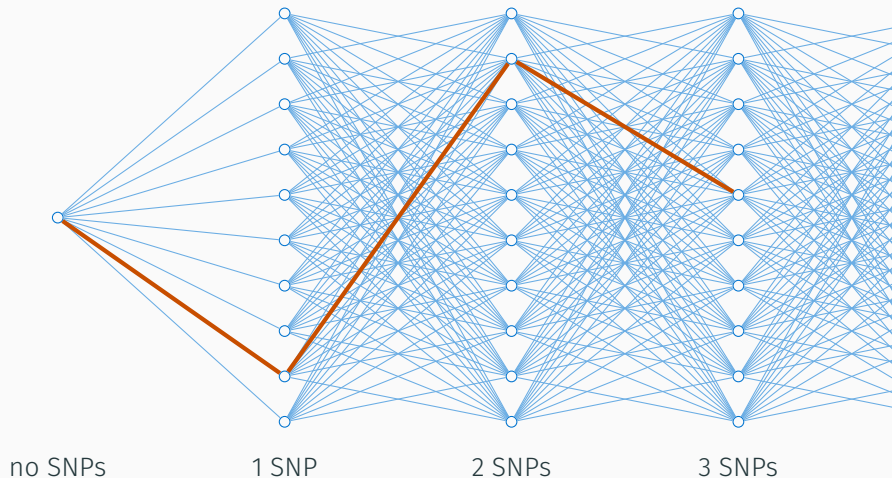
Alternative: evolutionary stochastic search



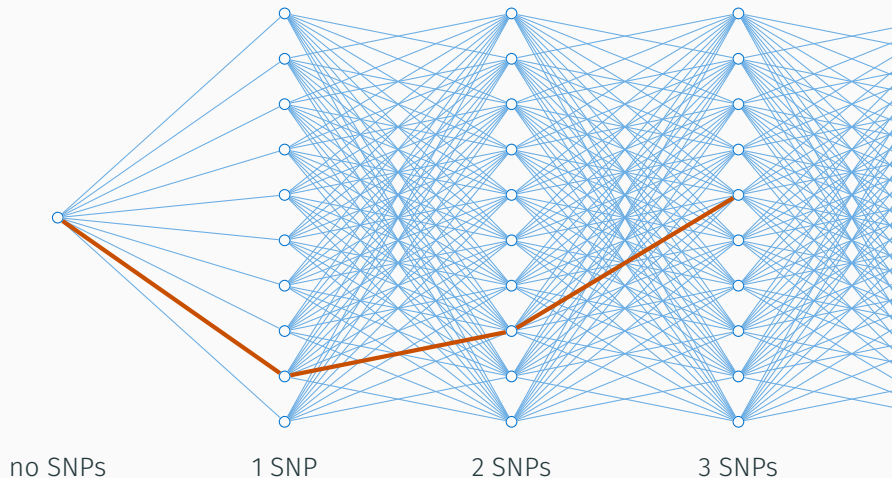
Alternative: evolutionary stochastic search



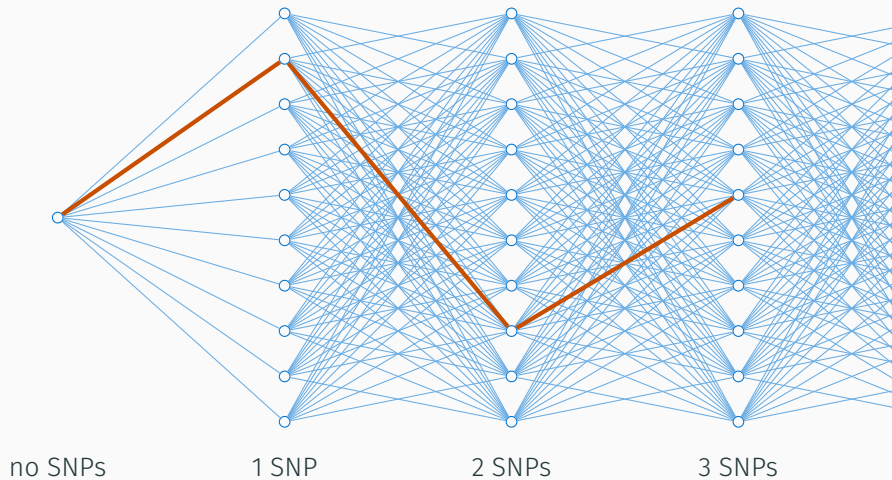
Alternative: evolutionary stochastic search



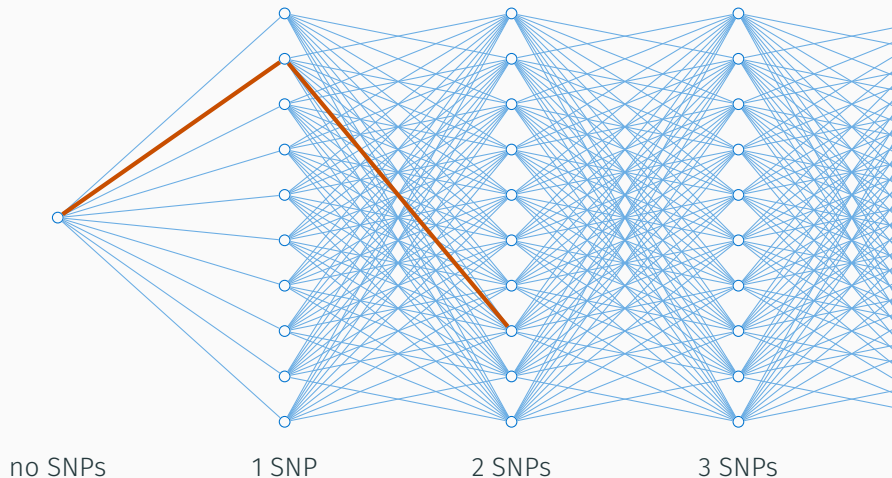
Alternative: evolutionary stochastic search



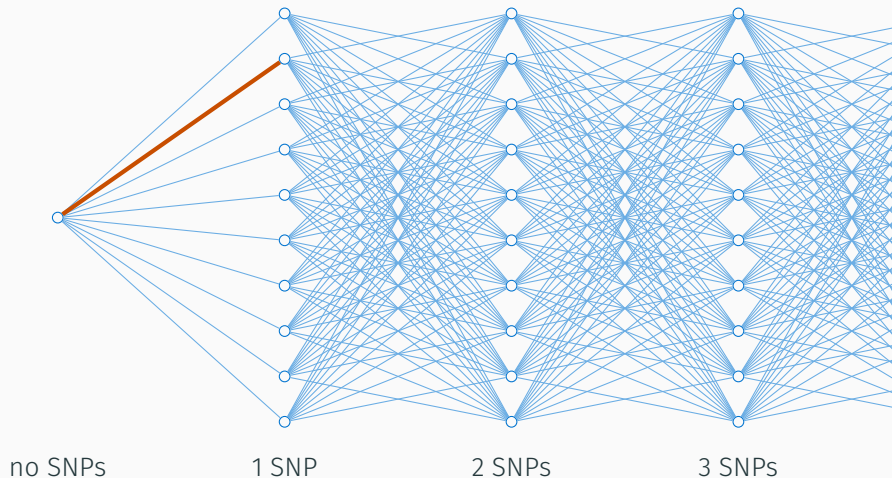
Alternative: evolutionary stochastic search



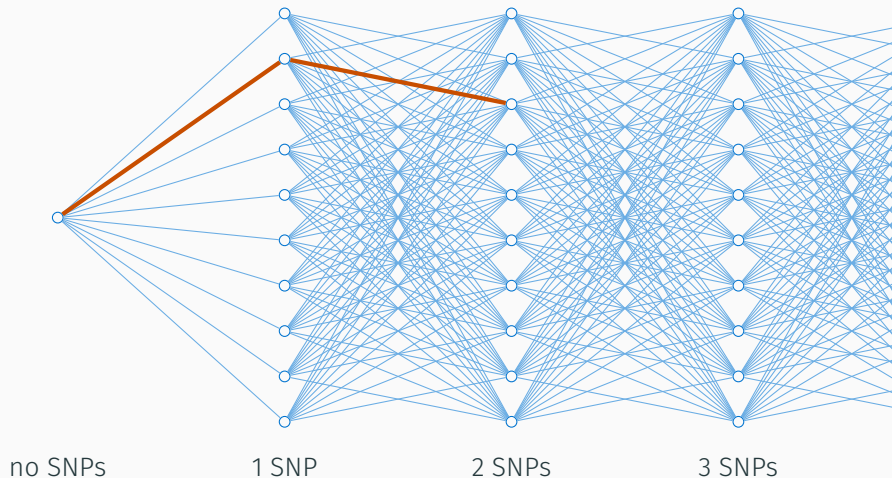
Alternative: evolutionary stochastic search



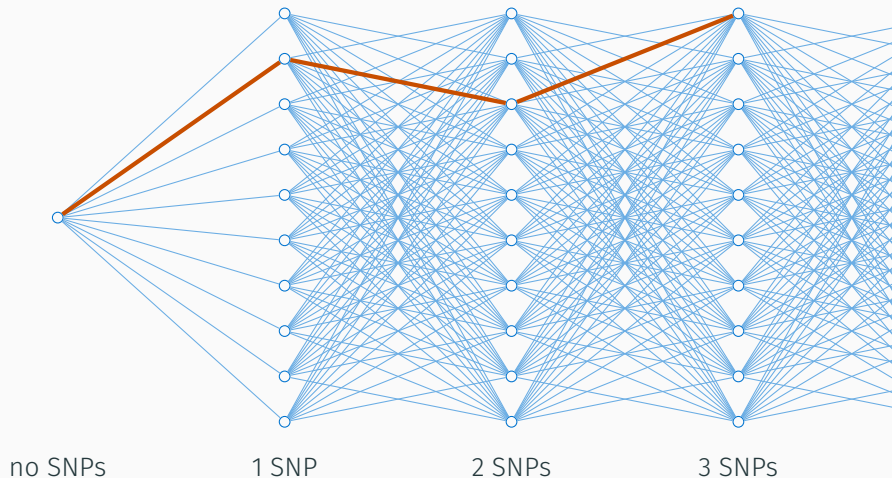
Alternative: evolutionary stochastic search



Alternative: evolutionary stochastic search



Alternative: evolutionary stochastic search



Fine mapping multiple causal variants



Effect

—

Frequency



Fine mapping multiple causal variants



Effect

–

Frequency



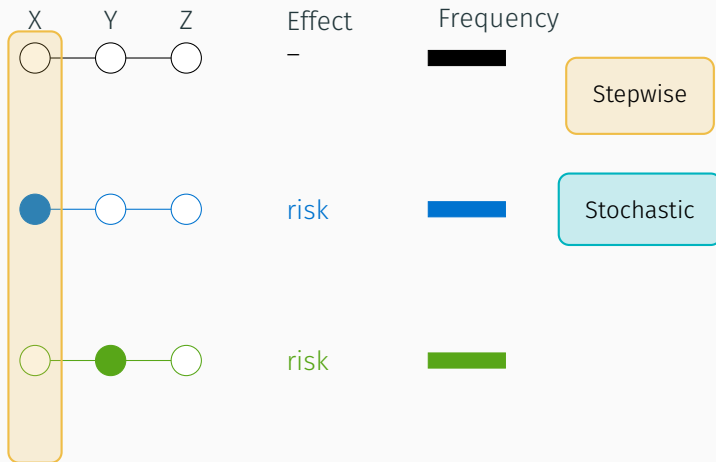
risk



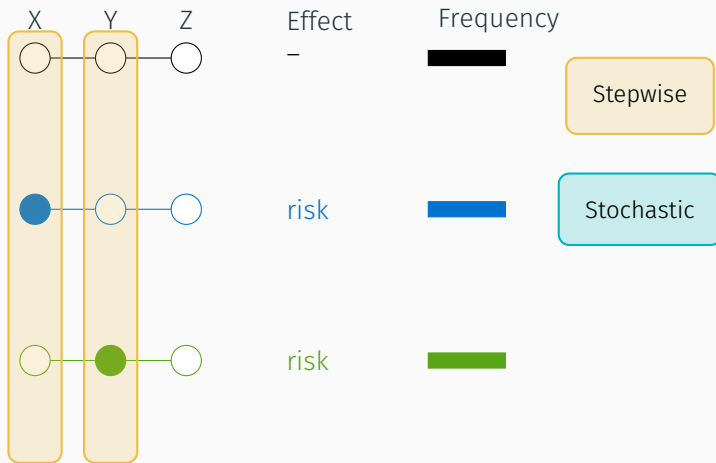
risk



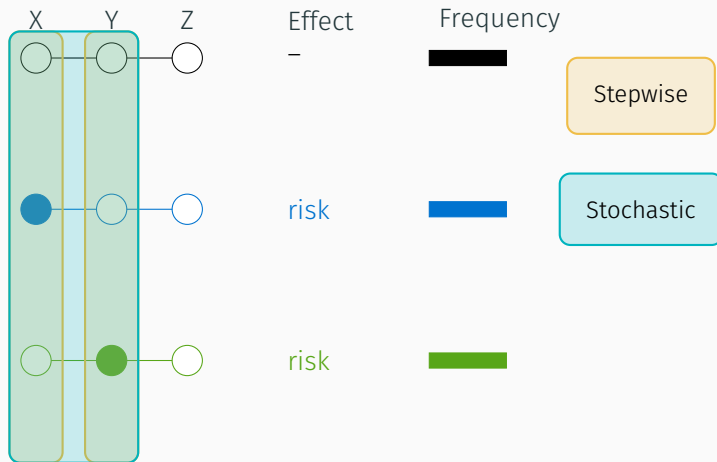
Fine mapping multiple causal variants



























Fine mapping multiple causal variants



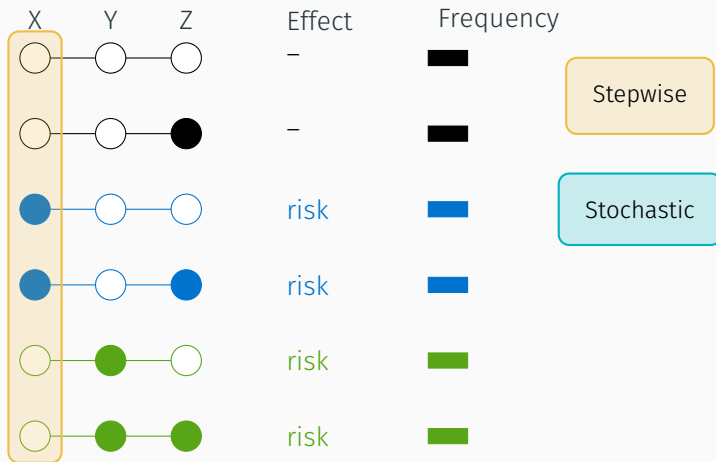
Fine mapping multiple causal variants



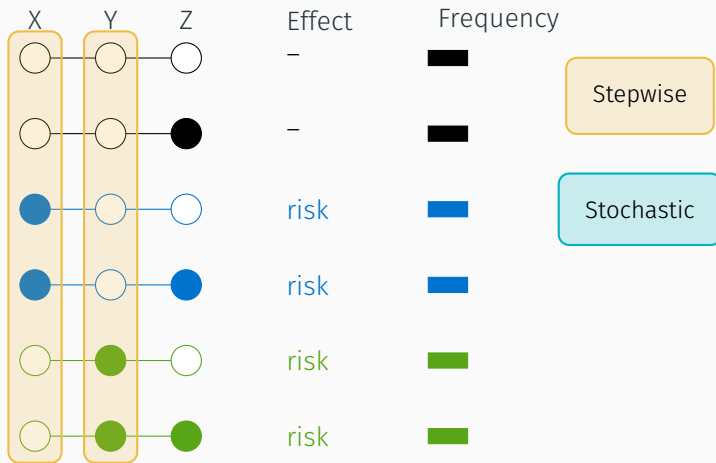
Fine mapping multiple causal variants

X	Y	Z	Effect	Frequency
			-	
			-	
			risk	
			risk	
			risk	
			risk	

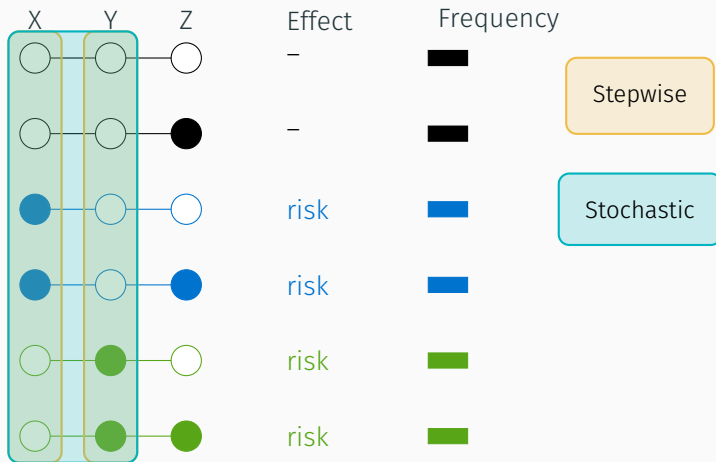
Fine mapping multiple causal variants



Fine mapping multiple causal variants



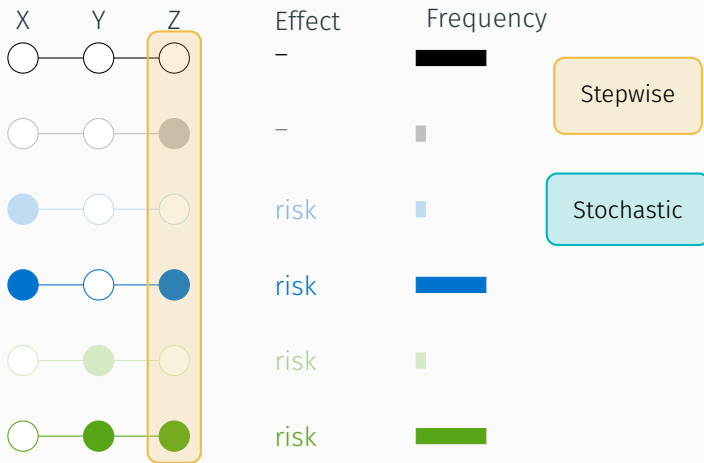
Fine mapping multiple causal variants



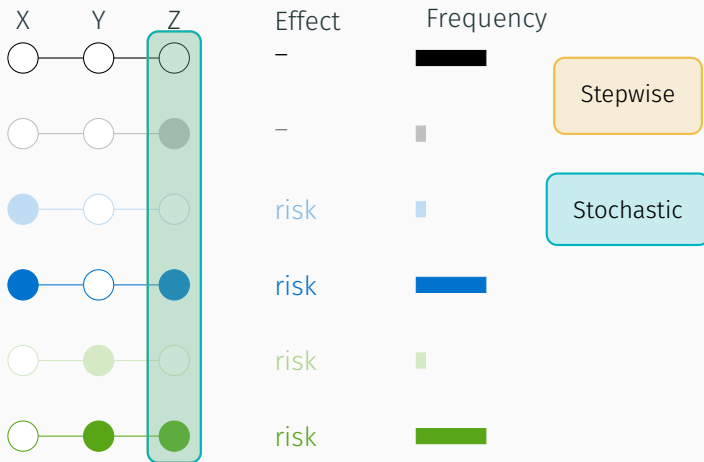
Fine mapping multiple causal variants



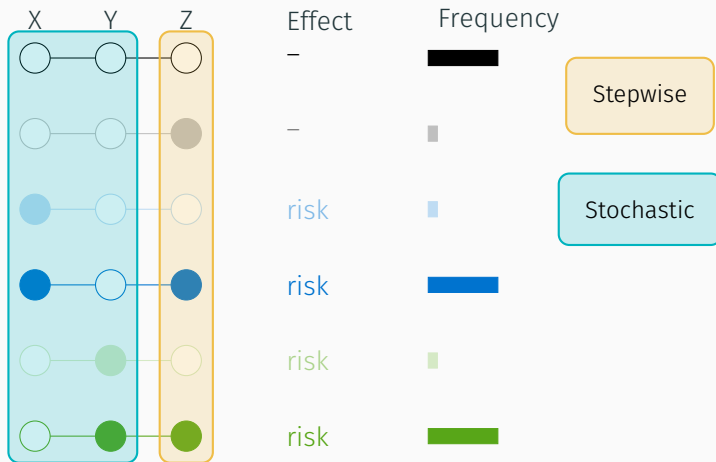
Fine mapping multiple causal variants



Fine mapping multiple causal variants



Fine mapping multiple causal variants



Systematic comparison: 89 genetic regions, 6 diseases

Group	Number	Group	Number
Autoimm. Thyroid Disease, ATD	2772	Celiac Disease, CEL	12041
Juvenile Idiopathic Arthritis, JIA	1214	Multiple Sclerosis, MS	4461
Rhemuatoid Arthritis, RA	11475	Type 1 Diabetes, T1D	6681
CONTROL	22997		

201 region/disease pairs showing association (min. $p < 10^{-6}$)

Systematic comparison: 89 genetic regions, 6 diseases

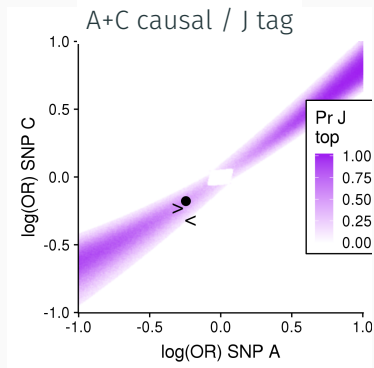
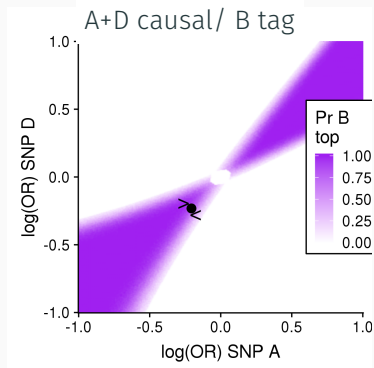
Group	Number	Group	Number
Autoimm. Thyroid Disease, ATD	2772	Celiac Disease, CEL	12041
Juvenile Idiopathic Arthritis, JIA	1214	Multiple Sclerosis, MS	4461
Rhemuatoid Arthritis, RA	11475	Type 1 Diabetes, T1D	6681
CONTROL	22997		

201 region/disease pairs showing association (min. $p < 10^{-6}$)

Regions	Region-disease pairs		
62	171	matched	
2	2	stochastic null ($p \simeq 1 \times 10^{-6}$)	
15	17	stepwise nested in stochastic	
5	5	different top SNP (two weak signals)	
5	6	non-nested mismatch	

What do mismatches look like?

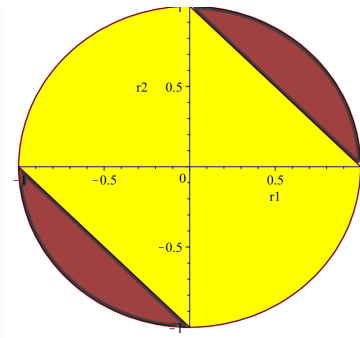
1. Similar effect sizes



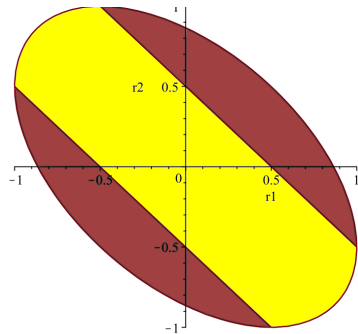
What do mismatches look like?

2. Tag correlated with both effect alleles

causal variants uncorrelated $r = 0$

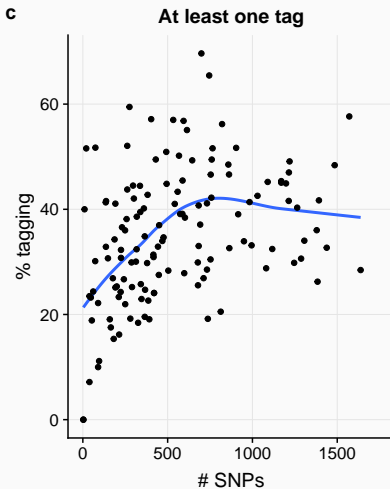
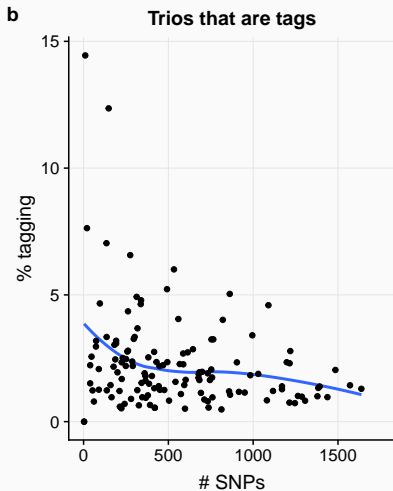


causal variants $r = 0.5$



$$|r_1 + r_2| > |1 + r_{12}|. \quad (1)$$

Frequency of joint tag pattern



Mismatch example: around *IL2RA*

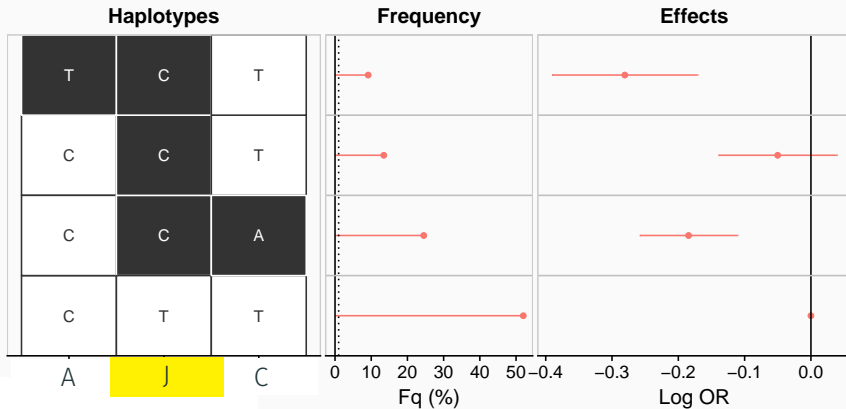
Disease	Stepwise	Stochastic
ATD	J	A+C
MS-UK	B	B
MS-international	B	A+D
RA-international	I	I
T1D	A+C+E	A+C+E+F

Mismatch example: around *IL2RA*

Disease	Stepwise	Stochastic
ATD	J	A+C
MS-UK	B	B
MS-international	B	A+D
RA-international	I	I
T1D	A+C+E	A+C+E+F

A = rs61839660 also associated with IBD, asthma
(opposite risk allele)

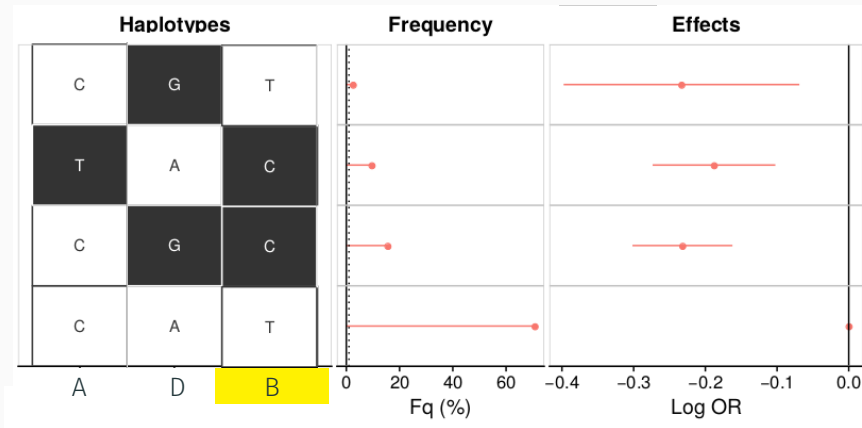
Haplotype analysis of ATD in *IL2RA* region



Disease	Stepwise	Stochastic
---------	----------	------------

ATD	J	A+C
-----	---	-----

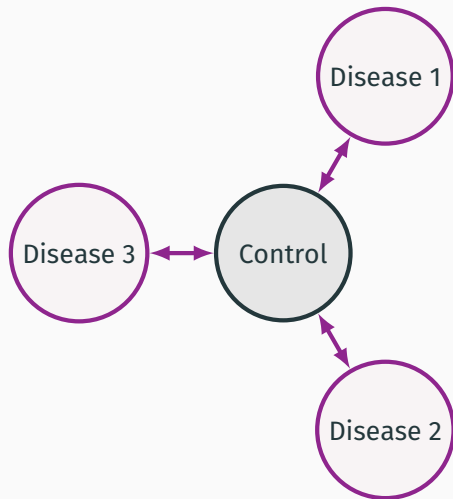
Haplotype analysis of MS in *IL2RA* region



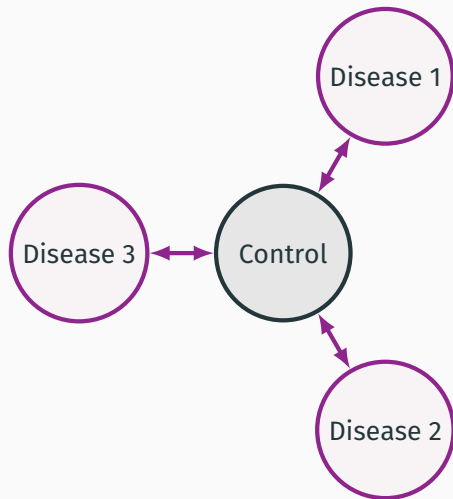
Disease	Stepwise	Stochastic
MS-UK	B	B
MS-international	B	A+D

Multiple-disease fine mapping

New multi-disease fine mapping method



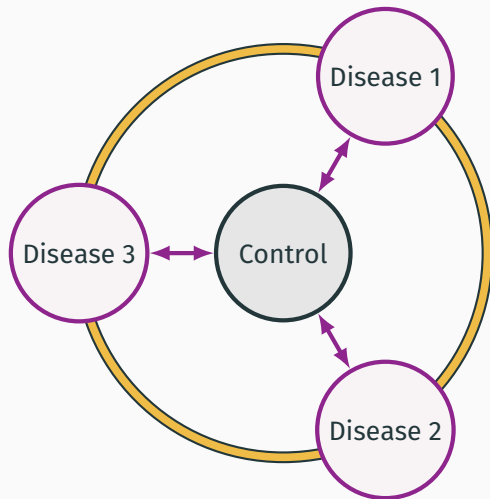
New multi-disease fine mapping method



Disease	Stochastic
ATD	A+C
MS-UK	B
MS-int	A+D
RA-int	I
T1D	A+C+E+F

20/30 regions with
> 1 associated disease
had a shared signal

New multi-disease fine mapping method



Disease	Stochastic
ATD	A+C
MS-UK	B
MS-int	A+D
RA-int	I
T1D	A+C+E+F

20/30 regions with
> 1 associated disease
had a shared signal

Bayesian fine mapping

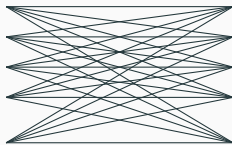
Single disease

Model	Prior	Data	Posterior
A	π_A	BF_A	$\propto \pi_A BF_A$
B	π_B	BF_B	$\propto \pi_B BF_B$
D	π_D	BF_D	$\propto \pi_D BF_D$
B+D	π_{B+D}	BF_{B+D}	$\propto \pi_{B+D} BF_{B+D}$
\vdots	\vdots	\vdots	\vdots

Bayesian fine mapping

Two diseases

Disease 1	
Model	Data
A	BF_A
B	BF_B
D	BF_D
B+D	BF_{B+D}
\vdots	\vdots

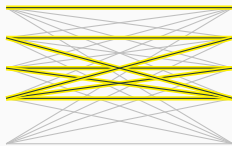


Disease 2	
Model	Data
A	BF_A
B	BF_B
D	BF_D
B+D	BF_{B+D}
\vdots	\vdots

Bayesian fine mapping

Two diseases

Disease 1	
Model	Data
A	BF_A
B	BF_B
D	BF_D
B+D	BF_{B+D}
\vdots	\vdots

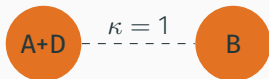


Disease 2	
Model	Data
A	BF_A
B	BF_B
D	BF_D
B+D	BF_{B+D}
\vdots	\vdots

Use prior to borrow information between diseases

Define **configurations**: sets of models for each disease

$\{M_i \text{ for disease 1}, M_j \text{ for disease 2}\}$



Prior: $Pr(M_i)Pr(M_j) \times \tau_{ij}$



$Pr(M_i)Pr(M_j) \times \tau_{ij} \times \kappa$

κ : upweighting factor

τ_{ij} : normalisation factor, keeps prior on total number of causal variants fixed

Computational challenges of Bayesian fine mapping

Single disease

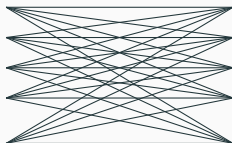
Model	Prior	Data	Posterior
A	π_A	BF_A	$\propto \pi_A BF_A$
B	π_B	BF_B	$\propto \pi_B BF_B$
D	π_D	BF_D	$\propto \pi_D BF_D$
B+D	π_{B+D}	BF_{B+D}	$\propto \pi_{B+D} BF_{B+D}$
\vdots	\vdots	\vdots	\vdots

Model space: exponential in number of causal variants

Computational challenges of Bayesian fine mapping

Two diseases

Disease 1	
Model	Data
A	BF_A
B	BF_B
D	BF_D
B+D	BF_{B+D}
\vdots	\vdots



Disease 2	
Model	Data
A	BF_A
B	BF_B
D	BF_D
B+D	BF_{B+D}
\vdots	\vdots

Model space: $(\text{exp. causal variants})^{\text{number of diseases}}$

Challenges: memory, computational time

Fast, memory efficient calculation of marginal posteriors

Speed: Joint Bayes factor approximated by function of single disease Bayes factors

$$BF(\{M_i, M_j\}) \propto BF(M_i) \times BF(M_j) \times \eta$$

η = function of numbers of cases, shared controls and causal variants

Memory: linear (not exponential) in number of diseases, by storing only marginal single disease posteriors

Fast, memory efficient calculation of marginal posteriors

Speed: Joint Bayes factor approximated by function of single disease Bayes factors

$$BF(\{M_i, M_j\}) \propto BF(M_i) \times BF(M_j) \times \eta$$

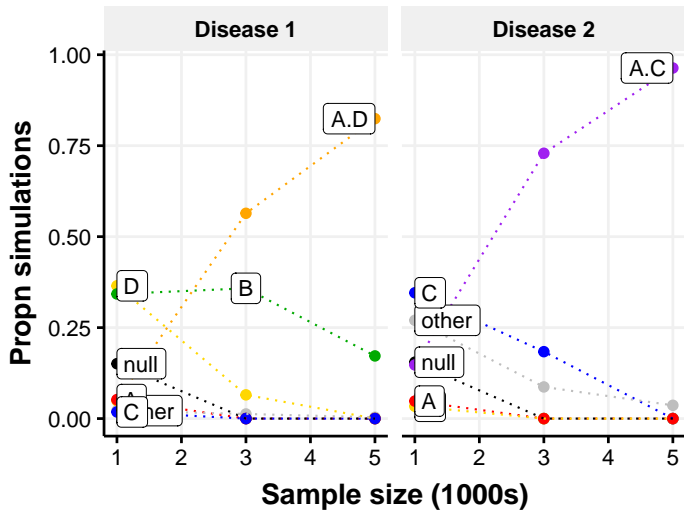
η = function of numbers of cases, shared controls and causal variants

Memory: linear (not exponential) in number of diseases, by storing only marginal single disease posteriors

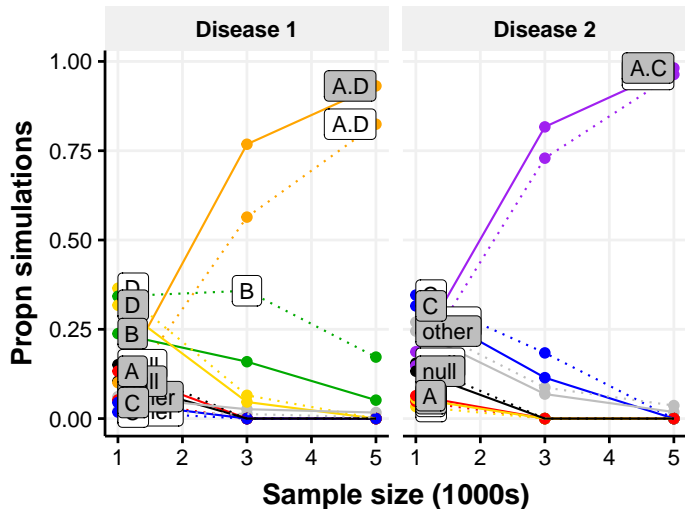
🕒 Running time: 15 seconds (2 diseases) — 83 seconds (6 diseases)

 <https://github.com/jennasimit/MFM>

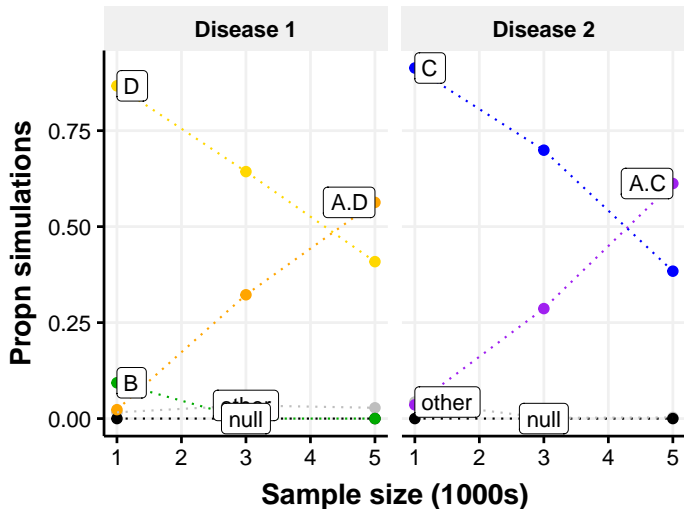
Joint fine mapping improves accuracy at smaller sample sizes



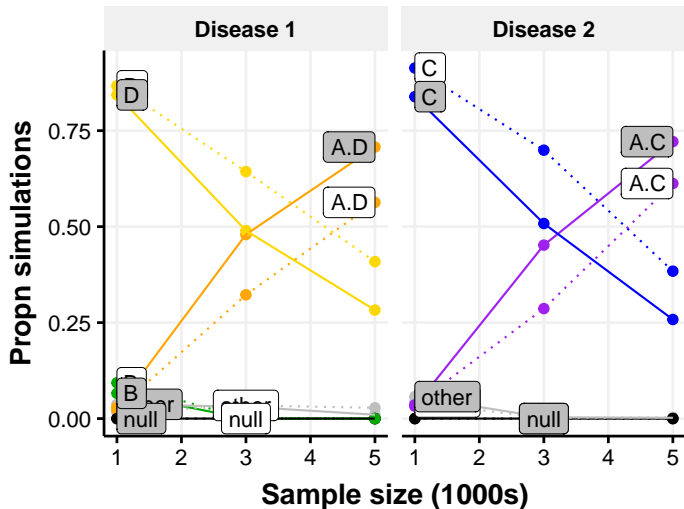
Joint fine mapping improves accuracy at smaller sample sizes



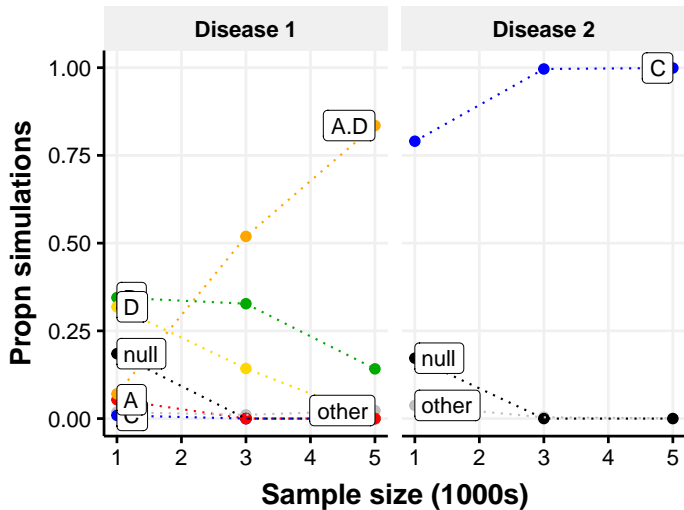
Joint fine mapping improves accuracy at smaller sample sizes



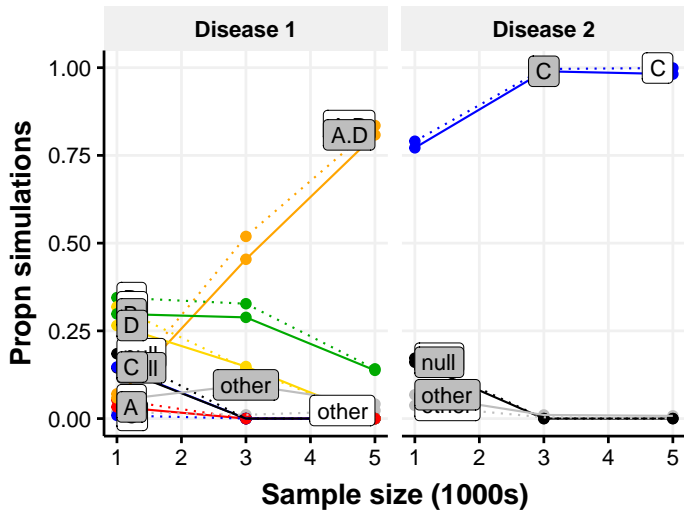
Joint fine mapping improves accuracy at smaller sample sizes



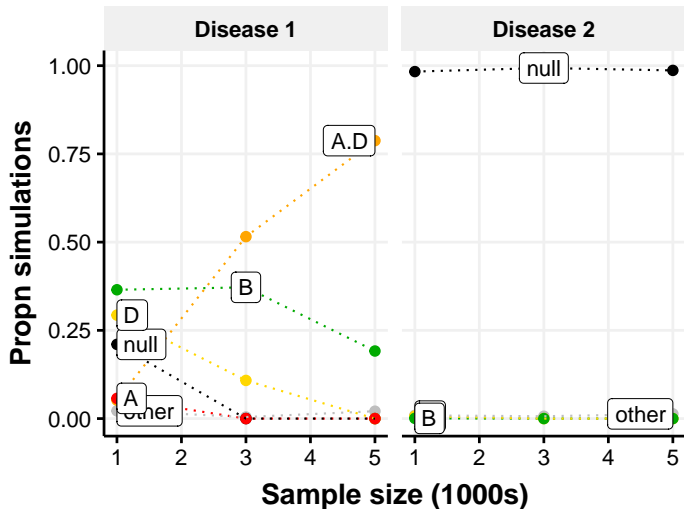
No free lunch, but loss small if no sharing



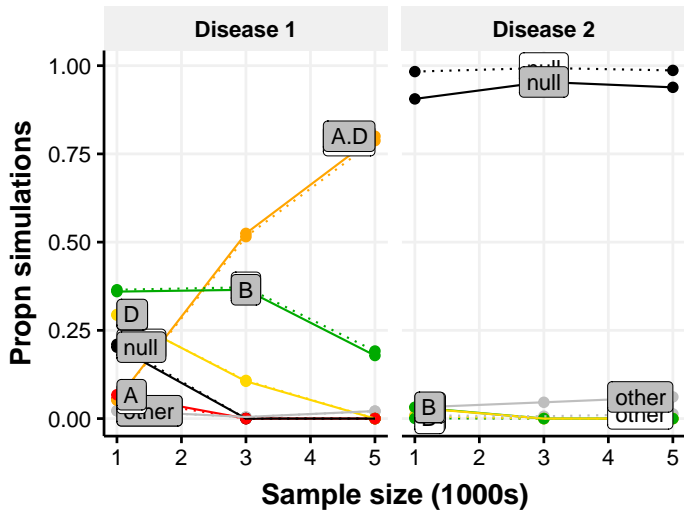
No free lunch, but loss small if no sharing



No free lunch, but loss small if no sharing



No free lunch, but loss small if no sharing



7/30 regions showed differences between single and multi-disease results

4/8 corresponded to UK subsets of larger UK+international samples

Region	Disease	Others	Single analysis		Multiple analysis	
			UK	UK+Int'al	UK	UK+Int'al
1p TNFRSF14	RA	CEL, MS	D	C		
6q BACH2	RA	ATD, T1D	G	C		
18p PTPN2	CEL	RA, T1D	F	C		
10p IL2RA	MS	JIA, RA, T1D	B	A+D		

7/30 regions showed differences between single and multi-disease results

4/8 corresponded to UK subsets of larger UK+international samples

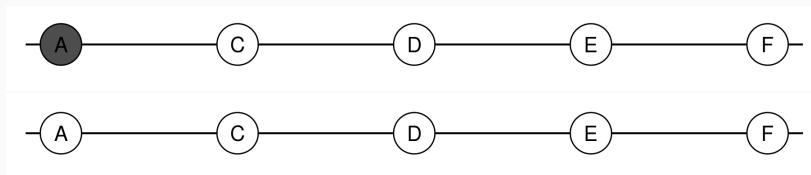
Region	Disease	Others	Single analysis		Multiple analysis	
			UK	UK+Int'al	UK	UK+Int'al
1p TNFRSF14	RA	CEL, MS	D	C	C	C
6q BACH2	RA	ATD, T1D	G	C	C	C
18p PTPN2	CEL	RA, T1D	F	C	C	C
10p IL2RA	MS	JIA, RA, T1D	B	A+D	A+D	A+D

Functional validation of causal effects on *IL2RA*

Allele specific expression

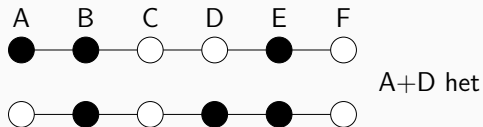
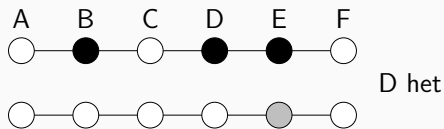
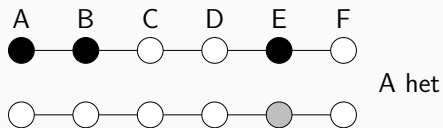
Allele specific expression: quantify relative expression of two chromosomes using targeted PCR and sequencing

Within-individual: controls for between individual variation in environment, other genetics etc

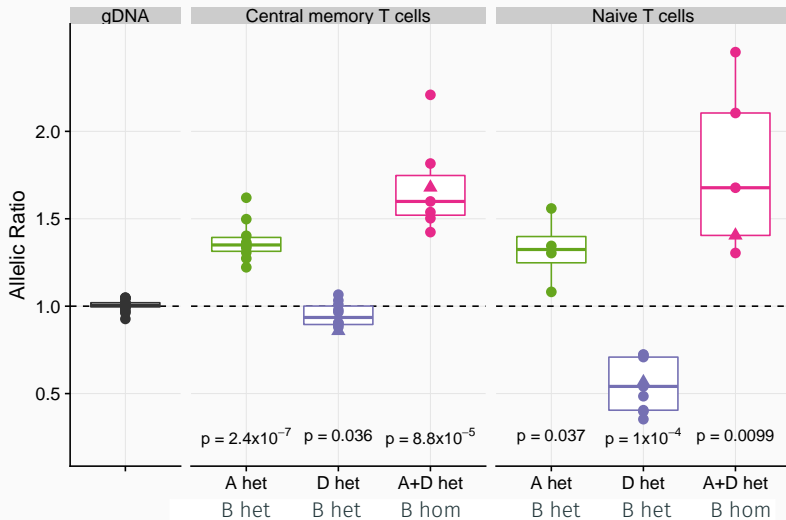


Cambridge NIHR BioResource

Effects of A, D and B



Effects of A, D and B



Reasons for caution

- Key assumption in stepwise search: no other SNPs act as a lower dimensional summary. Not about causal variants themselves!

Reasons for caution

- Key assumption in stepwise search: no other SNPs act as a lower dimensional summary. Not about causal variants themselves!
- *IL2RA* “famous”: multiple, complex associations
- Other regions of greatest a-priori interest show strongest associations, learning they are also complex (e.g. *IL2*, *CTLA4*)
- Most regions likely to contain > 1 causal variants as sample numbers increase



Reasons for optimism

- Borrowing information between related diseases can help overcome sample size limitations
- Correct fine mapping enables design of functional experiments
- Allows testing each effect, while controlling for others



results: chr1swallace.github.io/MFM-output @

software: [chr1swallace/GUESSFM](https://github.com/chr1swallace/GUESSFM) [jennasimit/MFM](https://github.com/jennasimit/MFM)

Thanks to:



Jenn Asimit



Dan Rainbow



Mary Fortune



Stasia Grinberg



Linda Wicker

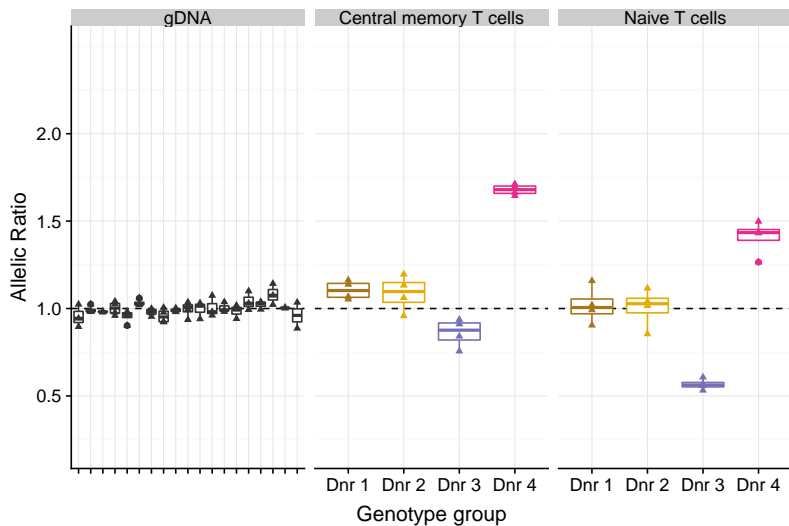
Disease investigators Steve Eyre (RA), Steve Rich, John Todd (T1D), Stephen Sawcer, IMSGC (MS), Wendy Thomson (JIA), David van Heel (Coeliac), Stephen Gough (ATD)



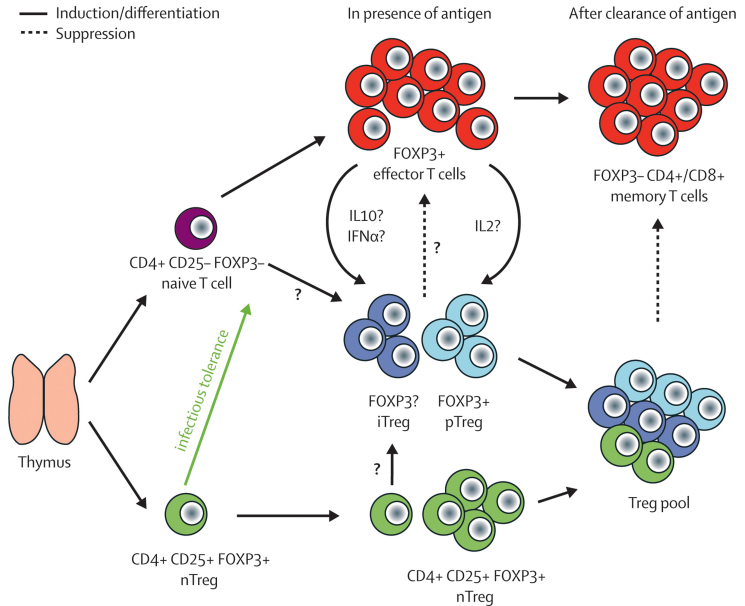
Cambridge NIHR BioResource



Individuals with rare recombination events



T cell subsets in immune-mediated diseases



T cell subsets in immune-mediated diseases

Disease	Model
T1D	A + C + E + F
MS	A + D
ATD	A + other
Celiac	A+C? / A / C / other

