Seminars in Immunopathology (2022) 44:101–113
https://doi.org/10.1007/s00281-021-00902-8

REVIEW

Methods for statistical fine-mapping and their applications to auto-immune diseases

Qingbo S. Wang[1,2,3,4] · **Hailiang Huang**[3,5,6]

Received: 6 July 2021 / Accepted: 22 October 2021 / Published online: 18 January 2022

Abstract

Although genome-wide association studies (GWAS) have identified thousands of loci in the human genome that are associated with different traits, understanding the biological mechanisms underlying the association signals identified in GWAS remains challenging. Statistical fine-mapping is a method aiming to refine GWAS signals by evaluating which variant(s) are truly causal to the phenotype. Here, we review the types of statistical fine-mapping methods that have been widely used to date, with a focus on recently developed functionally informed fine-mapping (FIFM) methods that utilize functional annotations. We then systematically review the applications of statistical fine-mapping in autoimmune disease studies to highlight the value of statistical fine-mapping in biological contexts.

Keywords Statistical fine-mapping · Functionally informed fine-mapping · Bayesian · Autoimmune disorders · Inflammatory bowel diseases · IBD genetics

Introduction

Genome-wide association studies (GWAS) have identified thousands of loci in the human genome that are associated with different traits such as height, body mass index (BMI), or susceptibility to different diseases [1–3]. In typical GWAS, for phenotype and genotype of interest, their relationship is modeled in a generalized linear model such that the phenotype (either quantitative or logit of binary outcome) is the sum of the genotype times its effect size (slope), the effects of covariates such as sex, age, and principal components accounting for the population structure, intercept, and error term [4, 5] (Box. 1). The null hypothesis that the slope is zero (i.e., the genotype of interest is not associated with the phenotype of interest) is tested for each variant where the genotype is available. In other words, each variant will have a *p*-value that characterizes the evidence that the variant is associated with the phenotype in a frequentist approach. With proper quality control and rigorous correction for multiple test, the variants passing significance threshold [6] (typically 5×10^{-8}, so called "genome-wide significance") is considered to be associated with the phenotype of interest. However, there is a clear difference between association and causation. This is true in GWAS as well: studies [7, 8] suggest that the majority of variants with significant *p*-value (i.e., "associated" with phenotypes) will have no detectable effect on the phenotype when perturbed (i.e., "causal" to phenotype). Such observations motivate us to differentiate between "association" and "causality," to pinpoint the causal variant(s) in a locus. Fine-mapping [9, 10] is such an effort to pinpoint causal variants (either experimentally or computationally), and statistical

This article is a contribution to the special issue on: Genetics and functional genetics of Autoimmune diseases - Guest Editors: Yukinori Okada & Kazuhiko Yamamoto

✉ Qingbo S. Wang
qingbow@sg.med.osaka-u.ac.jp

✉ Hailiang Huang
hhuang@atgu.mgh.harvard.edu

1 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan

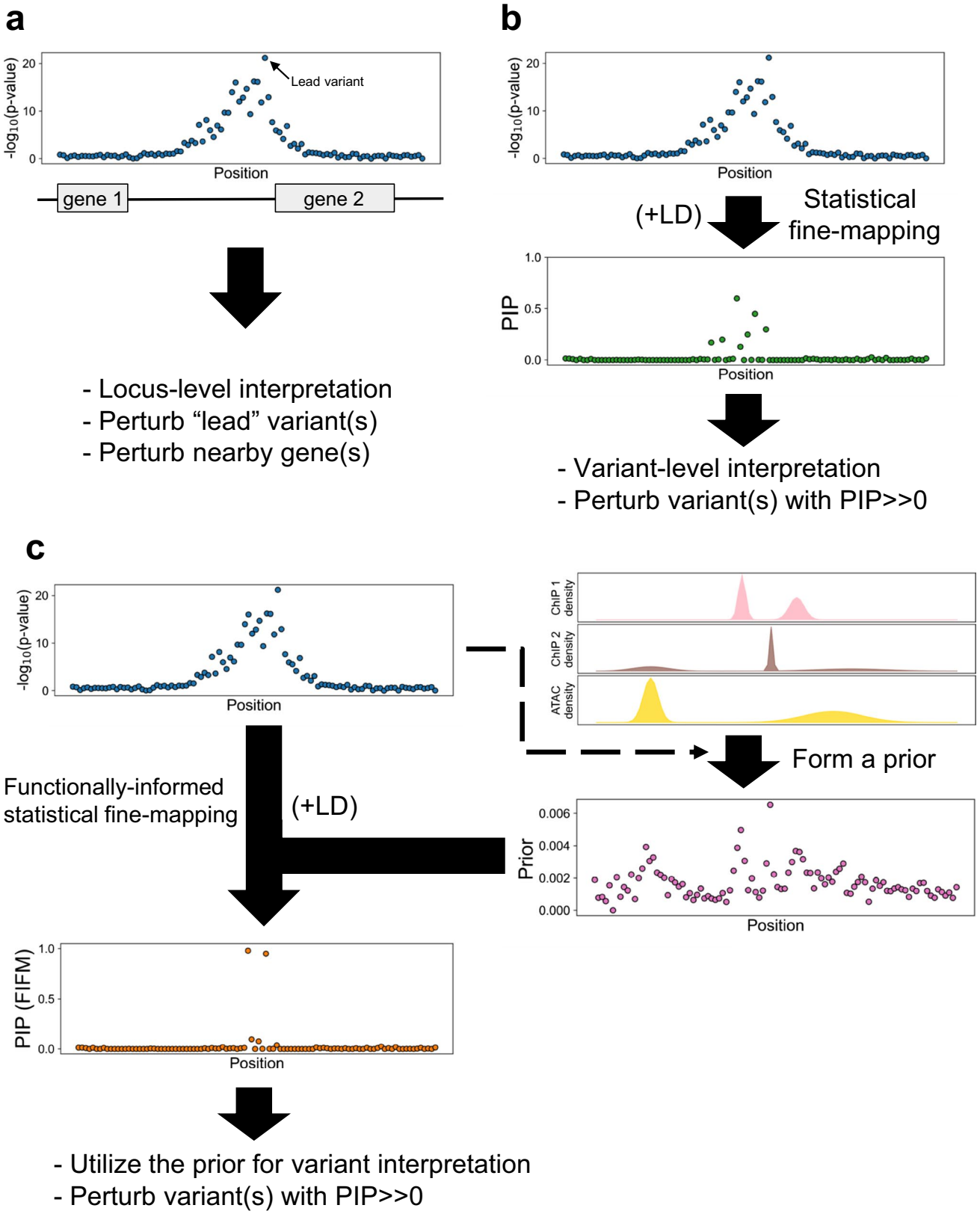
2 Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

3 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

4 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

5 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

6 Department of Medicine, Harvard Medical School, Boston, MA, USA



fine-mapping [9, 11] is a subset of fine-mapping studies that utilizes statistical framework. In this review, we will discuss the nature of statistical fine-mapping with four focuses. First, we will briefly review the challenges of GWAS as well as

◂**Fig. 1** Schematic overview of the statistical fine-mapping methods with uniform or functionally informed prior, in comparison with direct experimental approaches. **a** Downstream experiments following GWAS without statistical fine-mapping often assume the variant with the most significant *p*-value ("lead" variant) as the causal variant and proceed to perturbation of the lead variant and/or nearby gene(s). **b** Statistical fine-mapping is utilized to prioritize a small number of variants for downstream perturbation, which can be different from what *p*-value in GWAS suggests. This facilitates variant-level interpretation of GWAS results. **c** In a functionally informed fine-mapping (FIFM) framework, functional annotations are used (often together with the GWAS data) to form a prior. FIFM often results in an increase of power in prioritizing putative causal variants, which is typically characterized by higher maximum posterior inclusion probability (PIP) and/or lower credible set size[14]. The functional annotations used to form the prior are often directly used to interpret the biological mechanisms of causal variant(s)

experimental perturbation approaches to further clarify the motivation of statistical fine-mapping. Second, we will review the types of statistical fine-mapping methods that have been widely used to date (Fig. 1). Since high-quality reviews that achieve the same purpose already exist [9, 11, 12], we will make this second section brief without deep-diving into individual methods. On the other hand, a number of large-scale statistical fine-mapping studies [13, 14] have emerged recently. Since a large majority of such studies utilize functional annotations to perform functionally informed fine-mapping (FIFM), our third focus will be on the application of FIFM on large-scale studies. Finally, to highlight the value of statistical fine-mapping in biological contexts, we will systematically review the applications of statistical fine-mapping in autoimmune disease studies.

GWAS is not designed for causal variant identification

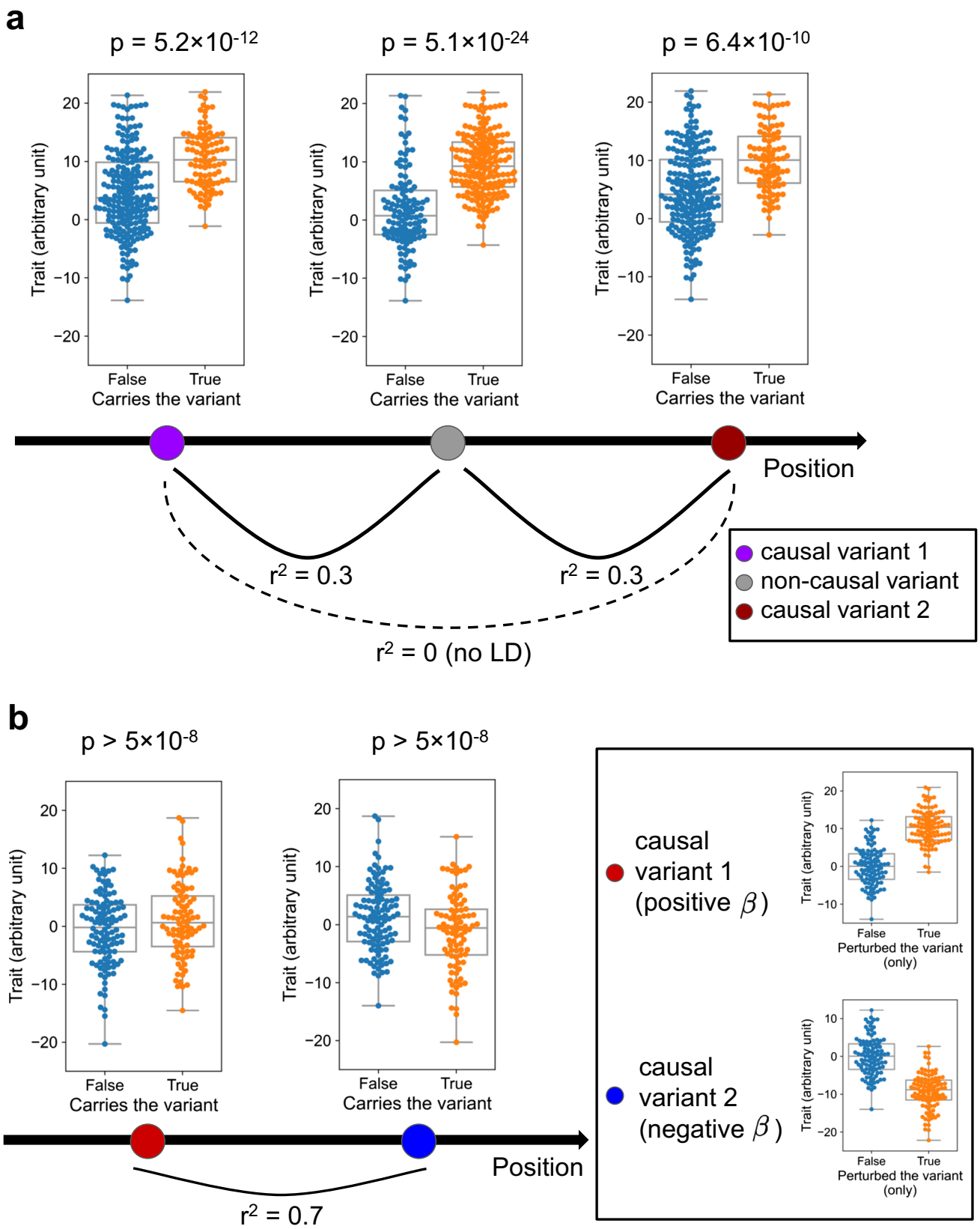
GWAS is not designed for identifying causal variants at single-variant resolution — instead, GWAS is designed to identify regions in the genome that are associated with the trait of interest [15, 16]. The main factor that makes association and causality different in GWAS is the existence of linkage disequilibrium (LD). LD is a term that describes the non-random association between nearby genomic variants [17, 18] — variants that are nearby tend to appear together more (or less) often than by chance, because the probability that a recombination event occurs at a position between two variants are typically smaller when they are nearby, compared to when they are far away. The strength of LD between two variants is typically denoted by the Pearson correlation coefficient (r) [18]. Because of LD, even if there is only one causal variant in a locus, hundreds or thousands of non-causal variants can be associated with the phenotype in GWAS, just because they are associated with the causal variant [19, 20] (i.e., "tagged"). In fact, the set of variants tested for association in GWAS, either directly measured in a genotyping array [21, 22] or imputed from a population reference panel [15, 23], are typically common variants that are designed to hopefully "tag" the causal variant. They are, most of the time, no more than "markers" that are in LD with causal variant(s), and the causal variant(s) themselves can even be missing from the set of variants tested. In addition, due to the limited sample size and stochastic noises, the variant with the strongest association (i.e., "lead variant," the variant with lowest *p*-value) is not always the causal variant. Understanding which variant(s) are truly causal (in other words, identifying the true causal configuration) can be even more complicated since there are often more than one causal variant in a locus (a locus is typically defined as a set of variants with (r^2 against the lead variant) > threshold, or simply within a certain distance window [24]). In such cases, the effect size and direction we observe for a variant can largely vary from the true causal effect (Fig. 2). Statistical fine-mapping, in a way, can be thought of as an exercise to disentangle the effect of LD from the GWAS data.

These factors remind us of the fact that even with orders of magnitude larger sample sizes, GWAS alone is, by nature, still not suited for causal variant identification, and highlight the value of statistical fine-mapping methods.

Experimental approaches are valuable but limited

Since one of the goals of GWAS is to nominate a set of regions for downstream biological experiments, one natural suggestion would be to directly move to experimental validations after GWAS, without performing statistical fine-mapping. One caveat of such approaches is that it often ambiguates the biological mechanisms underlying the GWAS signal. As a toy example, if a locus of gene X is associated with phenotype Y, we can validate that X is causal for Y by knocking out the entire gene X. However, if we can statistically fine-map the causal variant V on gene X and validate it by introducing variant V at single base-pair resolution followed by different biological assays, we can highlight different scenarios such as V introducing a stop codon, V being a missense variant that changes the protein 3D conformation, and V introducing aberrant splicing (Fig. 1a,b).

Recent developments in such genome perturbation at single-variant resolution have been remarkable. (1) Massive parallel reporter assay [7, 8, 25, 26] enables us to test the effect of mutations on gene expression in vitro with high throughput. (2) Genome engineering tools such as base editors enable introduction of single base-pair mutations in vivo [27–29]. However, they are still limited in that (1) is not a perfect proxy of human physiology and (2) is limited in its throughput, such that saturation mutagenesis at genome-wide


a
p = 5.2×10-12
p = 5.1×10-24
p = 6.4×10-10
Trait (arbitrary unit)
20
10
0
−10
−20
False
True
Carries the variant
Position
r2 = 0.3
r2 = 0.3
r2 = 0 (no LD)
causal variant 1
non-causal variant
causal variant 2
b
p > 5×10-8
p > 5×10-8
Trait (arbitrary unit)
Carries the variant
Position
r2 = 0.7
causal variant 1 (positive β)
causal variant 2 (negative β)
Perturbed the variant (only)

◂**Fig. 2** Two simplified examples where marginal *p*-value fails to prioritize the true causal variants. **a** The non-causal variant (center), frequently tagging one of the two true causal variants, has the most significant association *p*-value (in *F*-test) as well as the highest marginal effect size β (7.8 vs 5.8 and 5.3). **b** Two nearby causal variants in LD harboring high true effect sizes to the opposite direction, both have limited marginal association *p*-values that do not reach the statistical significance under multiple test correction ($p = 0.06$ and 0.007). Synthetic samples of $n = 300$ for **a** and $n = 200$ for **b** were generated, with true $|\beta| = 10$ and ϵ drawn from a normal distribution with SD = 5 for simplicity (therefore, *y* axis has no unit). $r = 0.317$, 0.317, and 0.0353 for **a** and 0.734 for **b** The code is available at http://github.com/QingboWang/fm-toy

scale is still far away. Performing statistical fine-mapping before such experimental validation is thus a natural way to maximize the value of downstream experimental approaches. Developments in the methods so called co-localization [30–34] further enhanced the value of statistical fine-mapping, by analyzing the results of statistical fine-mapping on complex traits and gene expression regulation (expression quantitative loci, or eQTL) studies simultaneously to elucidate the mechanisms from variant to gene to complex trait in a streamlined manner, making the downstream experimental validation easily designable and interpretable.

From *p*-value to Bayes factor and Posterior inclusion probability

Although *p*-value characterizes the evidence of a variant being associated with the phenotype, it does not allow us to compare one model likelihood (e.g., a model that variant V1 is causal) to another (e.g., a model that variant V2 is causal) in a direct and quantitative way. Bayes factor (BF) is a notion that quantifies the relative likelihood of one model over another [35] (Box. 2). Early studies [23, 36] included such a Bayesian approach and reported BF in addition to canonical *p*-value, with a hint to use it as a means to directly quantify the "probability of being a causal variant." Wakefield (2007, 2009) [37, 38] later showed that BF can be approximated from summary statistics (such as *p*-value, point estimation, and standard error of the effect size of each variant from GWAS) without individual-level genotype data ("Approximate Bayes Factor," or "ABF"). With these developments in Bayesian approaches to GWAS, Maller et al. [39] showed that, under a simplified scenario that there is exactly one causal variant in a locus of interest, we can directly compute the probability of a variant V being causal (nowadays called posterior inclusion probability, or PIP), as the BF of the model that V is causal (compared to the model that no variant is causal; they showed that this BF can be calculated by the genotype of variant V alone), divided by the sum of BF that each of the other variants is causal. They also introduced the term credible set, defined as the smallest set of variants that the PIPs sum up to a certain threshold value.

Statistical fine-mapping assuming a single causal variant in a locus has been valuable in its simplicity and interpretability, but it relies on a very strong assumption that does not necessarily hold true. Maller et al. [39], fully aware of the fact, also suggested that jointly modeling multiple causal variants is theoretically possible, but the implementation is challenging (when there are *n* variants, there would be 2^n causal configurations). One of the main focuses in the later development of statistical fine-mapping methods lays on modeling multiple causal variants. We also note that, although Bayesian approaches force us to specify the prior by nature (which could introduce biases) and non-Bayesian statistical fine-mapping approaches exist [40], our methods review will be focused on Bayesian approaches that are most commonly used.

Multiple causal variants

One possible approach for dealing with locus that may harbor more than one causal variant is to divide up the set of variants into independent signals, such that each set of variants would contain exactly one causal variant. Although this intuitively could be achieved by a series of conditional analysis (i.e., condition on the lead variant by including it as a covariate, do GWAS again to find remaining signal, add the lead variant in that conditioned GWAS, do the GWAS again, … until we see no more signal), it introduces practical challenges; setting the *p*-value threshold to determine that there is no more signal is non-trivial, and running GWAS iteratively is computationally expensive [11]. An early study [41] avoided such challenges and practiced a simple approach to define a locus simply based on the r^2 against the lead variant (i.e., clumping and merging) and to assume that each locus contains exactly one causal variant. Yang et al. [42] showed that without such extreme model simplification, conditioning can be achieved with summary statistics and LD matrix without requiring individual genotype data in a scalable manner. Once a set of variants likely containing exactly one causal variant is defined, ABF can be applied also from summary statistics. Such a COJO + ABF approach allows the whole process of identifying multiple causal variants tractable, by dividing up the problem into two steps: (1) identify a set of variants harboring exactly one causal variant, and (2) perform ABF for each variant set. The COJO + ABF approach has been widely used since then [43, 44].

However, over the time, there has been increasing amounts of evidence that conditional analysis often results in sub-optimal solutions, in simulations [11, 45–47] and real data [48, 49]. The simplest intuition [45] is that a non-causal variant that is in strong LD with two causal variants can have the most significant *p*-value and thus be mistakenly prioritized as the causal variant (Fig. 2a). One of the major methods that overcomes this limitation was presented in Hormozdiari et al. [45] (CAVIAR). In CAVIAR, they took

the approach of jointly modeling multiple causal variants rather than sequentially, and allowed directly calculating the BF for the case of > 1 causal variant. They dealt with the computational complexity by limiting the maximum number of causal variants as well as the number of variants in the locus. Other approaches that are distinct from naive conditional approach includes [BIMBAM [50]] (requires individual genotype data), [pi-MASS [51]] (utilizes MCMC), [JAM [47]] (utilizes matrix decomposition), the extensions of CAVIAR that is more scalable and widely applicable [CAVIARBF [52], eCAVIAR [31]], and those used in autoimmune disease studies that are discussed later [53].

Scalable methods

Although methods such as CAVIAR allowed a joint model of multiple causal variants, scaling such methods to a genome-wide level remained challenging. To implement a scalable fine-mapping method, Benner et al. [54] applied a shotgun stochastic search of the possible causal configurations instead of exhaustively enumerating the BFs for 2^n causal configurations. Their method, FINEMAP, either adds, exchanges, or deletes one putative causal variant in the locus in each iteration to generate a new causal configuration to evaluate. The method further utilizes a hash table to avoid re-computation of the same causal configuration and terminates the iteration once nearly all the causal configurations with non-negligible probability are searched. Wen et al. [55] (DAP-G) used a similar idea of avoiding the enumeration of all 2^n causal configurations by focusing only on non-negligible ones but used a deterministic method instead; their Deterministic Approximation of Posterior (DAP) algorithm restricts the search space based on two assumptions: (1) The true causal variants should have medium to highly significant association p-value. (2) The fraction of causal variants in a locus should be small (sparsity assumption) and allows tractable computation. Another widely used method developed by Wang et al. [56] (Sum of Single Effects = SuSiE) takes an iterative approach; analogous to conditional analysis, SuSiE takes single effect regression (a regression model where there is exactly one causal variant in a locus) as a building block to perform iterative Bayesian stepwise selection (IBSS). The algorithm (1) explicitly specifies L single effect vectors (initialized with uniform probability of being causal for each variant in each single effect vector, when the prior is uniform) to begin with, updates the 1st single effect vector based on the data, and (2) repeats the process of updating the 2nd, 3rd, …, L-th, 1st, 2nd, … single effect vector based on the data plus all the other single effect vectors until convergence.

Each of these methods is highly scalable and has been applicable to different large-scale studies (e.g., DAP-G in GTEx v8 study [57] and FINEMAP for UKBB biomarkers study in Sinnott-Armstrong et al. [58]) to elucidate the detailed biological mechanisms of GWAS signals, highlighting the value of scalable methods.

Functionally informed fine-mapping

A variant falling in a histone mark peak is more (or less) likely to be causal to a phenotype compared to another variant. A missense variant is more likely to be causal than an intron variant. Such additional biological information about the variants (e.g., epigenetic information, conservation, or other scores, which are called "functional annotations") are informative to identify causal variants, even before investigating specific GWAS data. In other words, we have a "prior" knowledge about the variants. As one strength of Bayesian methods is that it can flexibly incorporate different priors, a number of methods including those highlighted in the previous section [55, 59–63] allow incorporating such biological functional annotations as priors to increase the power of statistical fine-mapping. For example, distance to transcription starting site (dTSS) was incorporated as a prior in DAP-G to perform *cis*-eQTL fine-mapping in GTEx v8 [57]. We call such a series of methods that use functional annotations to form a prior (rather than using functional annotations post-statistical fine-mapping only to interpret the results; Fig. 1c) as functionally informed (statistical) fine-mapping (FIFM). Among various FIFM methods, this review focuses on two recent large-scale FIFM methods: (1) Polyfun [13] that was applied to UKBB phenotypes and (2) EMS [14] that was applied to GTEx v8 eQTLs. These two methods, rather than performing expectation–maximization (EM) iteration in the fine-mapping process as in PAINTOR [59], take a two-step approach of first calibrating the functional prior and then using the functional prior to perform FIFM using scalable methods [FINEMAP [54], SuSiE [56]].

The first method, Polyfun, allows the incorporation of functional features by stratified ld-score regression (S-LDSC [20]). First, it uses S-LDSC to estimate the heritability enrichment of each of the functional annotations for a phenotype of interest (with proper regularization and training-test split to avoid overfitting). Second, it estimates the per-SNP heritability (heritability explained by a single nucleotide polymorphism = SNP) by adding up the heritability enrichment of the functional annotations that the variant (SNP) of interest belongs to. Following the calibration step (binning the SNPs and re-calculating the per-SNP heritability for each bin), the functional prior is defined to be proportional to the per-SNP heritability. Then they use the functional prior for downstream statistical fine-mapping using SuSiE or FINEMAP. By applying the method to 49 UKBB traits, the authors validated the power gain of FIFM compared to

canonical methods and also discussed the polygenic localization of common trait heritability (i.e., how many variants are needed to explain a certain percentage of trait heritability).

The second method, Expression Modifier Score (EMS), first trains a random forest (RF)-based predictor that uses > 6,000 functional annotations, to prioritize putative causal eQTLs that are nominated with high confidence in uniform prior fine-mapping. The method also includes deep-neural network-based variant activity prediction scores [64, 65] as a set of features and shows that those features collectively present high feature importance in addition to dTSS. In the subsequent step, the output scores (EMS) from the RF model are scaled and used to re-weight the single effect vectors in SuSiE. Functionally informed PIP and credible sets are then quantified from the weighted vectors for 49 GTEx tissues individually. The method was also applied for a large-scale co-localization analysis to elucidate > 300 additional candidate genes for UKBB phenotypes.

These results both showed an improvement compared to the canonical methods in terms of the number of putative causal variants discovered, without loss of accuracy, and thus together highlighted the value of performing FIFM on a large scale.

Further extension of statistical fine-mapping methods

Although not deeply covered in this review, a more diverse set of applications exist in recent development of statistical fine-mapping methods [66–75]. First is the cross-population fine-mapping (xpop-FM) approach that utilizes different LD structures between populations. Such approaches [71, 72] rely on an assumption (supported by biological observations) that the true causal variant is, most of the time, shared between different populations [76, 77]. By simple intuition, when variant V0 and V1 each has PIP = 0.5 in population 1, variant V0 and V2 has PIP = 0.5 in population 2 and variant V0 and V3 has PIP = 0.5 in population 3, one would gain confidence that variant V0 is the true causal variant. One challenge in such xpop methods is the model misspecification, i.e., a causal variant may not be shared or has very different effects across populations for some loci (e.g., the *TNFSF15* locus, with Crohn's disease OR of 1.15 and 1.75 for Europeans and East Asians respectively [78]). While such heterogeneity across populations can be properly modeled for GWAS using methods such as MANTRA [79], MR-MEGA [80], MAMA [81], or random effect models [82], the ability of xpop fine-mapping methods to model such heterogeneity has not been fully evaluated in real data. With further methodology developments as well as the increase in population diversity of the available genome, we envision the value of such xpop methods will increase. Similarly, harmonizing heterogeneous datasets with different underlying technologies (such as different arrays, whole exome, or genome sequencing) and including low-frequency variants is thought to be fruitful for further discovery of putative causal variants underlying human complex disorders by increasing the statistical power and the coverage of the genome. Another direction is the optimization of the prior distribution of the causal effect sizes [83, 84] (not the causal configuration); for example, Walters et al. [83] suggested Laplace prior could increase the statistical power compared to the commonly used normal distribution. As optimizing the prior is a non-trivial problem in Bayesian analysis in general, it could be also valuable to discuss the possibility of moving outside of the Bayesian world to practice statistical fine-mapping in a frequentist approach. As a general note, no single method for statistical fine-mapping today serves as a "gold standard," and different methods rely on different assumptions. Interpreting the results from multiple different aspects, as will be discussed in the next sections, is of high importance.

Application of statistical fine-mapping in autoimmune diseases

Many autoimmune disorders are highly heritable [85]. GWAS and statistical fine-mapping have thus been very effective in finding genetic variants underlying these disorders. Here, we review methods and findings for ten major autoimmune disorders including rheumatoid arthritis (RA), type 1 diabetes (T1D), the inflammatory bowel diseases (IBD) including Crohn's disease (CD) and ulcerative colitis (UC), systemic lupus erythematosus (SLE), ankylosing spondylitis (AS), psoriasis (PSOR), autoimmune thyroid disease (THY), celiac disease (CeD), and multiple sclerosis (MS). We chose these disorders because they are sufficiently powered with at least 10,000 cases. The number of genetic loci associated with these disorders ranges from 40 (CeD) to 240 (IBD) and is influenced by the sample size, the heritability, and the genetic architecture of the disorder (Table 1).

Farh et al. [53] performed the first genome-wide statistical fine-mapping on several autoimmune disorders using Probabilistic Identification of Causal SNPs (PICS), an algorithm estimating the probability that an individual variant is causal considering the haplotype structure and observed pattern of association at the genetic locus. This fine-mapping analysis was performed on data available prior to July 2013. For some disorders (AS, PSOR, THY, CeD, and MS), this study remains the best available fine-mapping study. For other disorders (RA, T1D, CD, UC, and SLE), subsequent fine-mapping studies have been performed on data with larger sample size and higher quality (e.g., higher imputation quality and higher genomic coverage). These studies also used more sophisticated fine-mapping methods. RA and T1D used conditional analysis for multiple independent

Table 1 GWAS and fine-mapping analyses across ten autoimmune disorders. All studies were performed on European subjects except for SLE, which combined European and East Asian subjects in the fine-mapping analysis

Disorder	Abbreviation	Heritability (CIs) (c)	GWAS			Fine-mapping					
			# case	# loci	PMID	# case	# loci	1-SNP set	5-SNP set	Method	PMID
Ankylosing spondylitis	AS	0.97 (0.92–0.99)	10,417	48	26974007	10,619	28	0	-	PICS	25363779
Autoimmune thyroid disease	THY	0.79	30,234	93	32581359	2,747	10	0	-	PICS	25363779
Celiac disease	CeD	0.57 (0.32–0.93)	12,041	40	22057235	12,041	40	1	-	PICS	25363779
Inflammatory bowel diseases—Crohn's disease (a)	IBD—CD	1.00 (0.34–1.00)	25,042	240	28067908	20,155	94	18	42	-	28658209
Inflammatory bowel diseases—Ulcerative colitis (a)	IBD—UC	0.67 ± 0.13				15,191					
Multiple sclerosis	MS	0.25 (0.00–0.88)	47,429	233	31604244	14,498	87	2	-	PICS	25363779
Psoriasis	PSOR	0.66 (0.52–0.77)	19,032	63	28537254	10,588	36	7	-	PICS	25363779
Rheumatoid arthritis	RA	0.68 (0.55–0.79)	22,628	121	33310728	11,475	46	0	5	ABF	30224649
Systemic lupus erythematosus	SLE	0.66	11,283	132	33536424	11,283	132	5	17	PAINTOR	33536424
Type 1 diabetes (b)	T1D	0.88 (0.78–0.94)	11,644	51	25751624	9,334	49	1	10	ABF	30224649

[a]CD and UC are two subtypes of IBD and are often analyzed together for their extensively shared genetic architecture

[b]The GWAS study included both case–control and family samples

[c]Heritability estimates compiled from multiple sources with detailed provided in Maria Gutierrez-Arcelus et al. Nature Reviews Genetics 2016 (PMID: 26907721)

Table 2 Putative causal variants with PIP > 95% for autoimmune disorders. See Table 1 for trait abbreviations

Trait	Variant	Gene	Function	PIP
CD	rs2066844	*NOD2*	R702W	99.9%
CD	rs2066845	*NOD2*	G908R	99.9%
CD	rs5743293	*NOD2*	Fs1007insC	99.9%
CD	rs61839660	*IL2RA*	Intronic	99.9%
CD	rs7307562	*LRRK2*	Intronic	99.9%
CD	rs5743271	*NOD2*	N289S	99.3%
CD	rs72796367	*NOD2*	Intronic	98.3%
CD	rs41313262	*IL23R*	V362I	97.3%
CD	rs28701841	*PRDM1*	Intergenic	97.1%
UC	rs6017342	*HNF4A*	Intergenic	99.9%
UC	rs35667974	*IFIH1*	I923V	99.4%
UC	rs4676408	*GPR35*	Intergenic	99.4%
IBD	rs6062496	*RTEL1-TNFRSF6B*	Intronic	99.6%
IBD	rs141992399	*CARD9*	1434 + 1G > C	99.5%
IBD	rs74465132	*IKZF1*	Intergenic	99.4%
IBD	rs10748781	*NKX2-3*	Intergenic	99.0%
IBD	rs35874463	*SMAD3*	I170V	98.9%
IBD	rs1887428	*JAK2*	Intergenic	97.4%
SLE	rs2736100	TERT	Intronic	100.0%
SLE	rs2431697	*PTTG1-MIR146A*	Intergenic	99.9%
SLE	rs2297550	*IKBKE*	TF binding site	99.7%
SLE	rs7097397	*WDFY4*	Arg1816Gln	99.3%
SLE	rs2205960	*TNFSF4*	Intergenic	95.7%
T1D	rs34536443	*TYK2*	P1104A	100.0%
MS	rs533259	*RNASEL*	Intronic	100.0%
MS	rs733724	*HACE1*	Intronic	98.0%
PSOR	rs17716942	*KCNH7*	Intronic	100.0%
PSOR	rs12188300	*IL12B*	Intergenic	100.0%
PSOR	rs33980500	*TRAF3IP2-AS1*	D10N	100.0%
PSOR	rs11795343	*DDX58*	Intronic	99.7%
PSOR	rs8016947	*NFKBIA*	Intergenic	100.0%
PSOR	rs28998802	*NOS2*	Intronic	100.0%
PSOR	rs34536443	*TYK2*	P1104A	99.6%%
CeD	rs1893592	*UBASH3A*	Intronic	98.0%

associations, and ABF to compute the credible sets. IBD used three fine-mapping methods specifically designed in order to capture the disease subtypes (CD and UC). Both the stepwise conditional analysis and MCMC were used to infer the independent associations for IBD. Results from the three methods were then harmonized which served as a quality control filter. SLE used conditional analysis for loci hosting multiple independent associations and PAINTOR [59] to compute the credible sets combining subjects of both European and East Asian ancestries. All fine-mapping studies for these disorders were performed without the functional priors.

Outcome of fine-mapping studies is dependent on the disease heritability, the sample size, and the disease genetic architecture (Tables 1 and 2). THY, MS, T1D, RA, and PSOR only mapped a subset of the genome-wide significant loci using a subset of subjects because the largest GWASs were published after the fine-mapping studies. None of the THY loci was mapped to a small credible set, likely because only less than 3,000 cases were used in fine-mapping. MS had two loci mapped to single-variant resolution, located in the introns of *RNASEL* and *HACE1*. T1D had one locus mapped to a single-variant credible set (*TYK2* P1104A) and nine more to credible sets with five or fewer variants. Fine-mapping for RA and PSOR was more productive: RA had five loci mapped to credible sets with five or fewer variants, and PSOR had seven loci mapped to a single causal variant, including the *TYK2* P1104A (also the T1D putative causal variant), a missense variant for *TRAF3IP2-AS1* (D10N), and variants in the introns of *KCNH7*, *DDX58*, and *NOS2*. AS, CeD, and SLE used all available GWAS samples. None of the AS loci was mapped to a small credible set likely because the effect sizes for AS loci are small thus are less powered for fine-mapping. One locus for CeD was mapped to a single variant (in the intron of *UBASH3A*), and 17 SLE loci were mapped to credible sets with five or fewer variants, among which five loci were mapped to a single causal variant, including a variant upstream of *TNFSF4* and a *WDFY4* missense variant (R1816Q). Driven by the sample size and the heritability, IBD fine-mapping is the most productive among the ten autoimmune disorders: 42 associations were mapped to credible sets with five or fewer variants, and 18 to a single causal variant, including multiple missense variants (fs1007insC, R702W, G908R, N289S) in *NOD2*, a *CARD9* essential splicing variant (1434 + 1G > C) and so on.

Coding variants play a critical role in autoimmune disorders. We have observed a clear enrichment of coding causal variants for IBD compared with synonymous variants. This observation is consistent with the allelic series observed in earlier IBD genetics studies, for example in *NOD2* and *CARD9*. Coding variants in general have larger effect sizes on diseases (e.g., fs1007insC has OR close to 3 for CD) and are particularly valuable in connecting genetic findings to their biological mechanisms [86]. Coding variants have also been fine-mapped for other autoimmune disorders revealing key mechanistic insights. For example, the *IFIH1* I923V variant was mapped as the putative causal variant for T1D and UC (though only to the single-variant resolution in UC), suggesting the antiviral response pathway could be relevant to onset of these disorders. Genes with fine-mapped coding variants, such as *NOD2* and *TYK2*, are also historically known to be responsible for Blau syndrome [87] (dominant) and immunodeficiency [88] (recessive) respectively, suggesting converging biological mechanisms between polygenic and Mendelian immune disorders.

The majority of autoimmune GWAS loci implicate the noncoding genome. Farh et al. [53] first connected these noncoding genetic variations to immune-cell enhancers and found many of them gain histone acetylation or transcribe enhancer-associated RNA upon immune stimulation. Huang et al. [89] further investigated the noncoding IBD putative causal variants and found them disrupt transcription factor binding sites, implicating epigenetic marks in specific immune cells in CD patients and in gut mucosa in UC patients. The IBD noncoding variants were also found to regulate gene expressions but only in cell types or tissues relevant to the disease, not in whole blood. Despite these initial insights, the biological and molecular mechanism for most fine-mapped causal variants is still unclear, reflecting our limited knowledge in the noncoding genome.

We note that several IBD genes have multiple independent variants associated with the disease [89]. The most notable one is *NOD2*, the first reported IBD genetic association, which hosts more than ten variants contributing to the IBD risk (mostly CD). The other notable gene is *IL23R*, hosting five independent causal variants (three coding and two noncoding) that confer protection to IBD. Such a spectrum of disease-associated alleles, or allele series, can be used to establish the function-phenotype dose–response relationship, which has been shown to be important in revealing the disease genetic mechanism and facilitates the discovery and validation of therapeutic targets [86, 90, 91].

We also note that many autoimmune disease causal variants are highly pleiotropic [89]. For example, the *TYK2* P1104A variant confers protection to CD, MS, PSOR, RA, and T1D (though only mapped to single-variant resolution for T1D and PSOR). Interestingly, one causal variant can sometimes confer different directions of effects for different autoimmune or infectious disorders. For example, the *IFIH1* I923V variant increases an individual's risk for UC but decreases the risk for T1D; an *IL2RA* intronic variant, rs61839660, increases the disease risk for CD and SLE but confers protection to T1D; the *TYK2* P1104A variant, despite being protective for several autoimmune disorders, increases homozygous carriers' risk to tuberculosis across diverse ancestral populations [92, 93]. These observations reflect the shared biological pathways underlying autoimmune disorders and the delicate balance between tolerance and autoimmunity in the human immune system.

Future perspectives

We have reviewed the basis of statistical fine-mapping methods, key fine-mapping studies in autoimmune disorders, and their important findings. These studies have revealed important causal variants underlying the human autoimmune disorders, and the mechanisms through which they modify individual's risk to the diseases. Despite these successes, we

note that not every autoimmune disease genetic loci have been fine-mapped and not all resources available have been leveraged in fine-mapping. This is partially because a high-quality fine-mapping typically requires a sample size larger than that of GWAS, and genetic data of higher quality to allow every variant to be assessed for their causality (while GWAS is typically tolerant to missing a few variants). Future investigations into how to properly perform fine-mapping across studies with different design factors (e.g., xpop) or genomic technology (various arrays, whole exome, or genome sequencing), as discussed in the "Further extension of statistical fine-mapping methods," is key for fine-mapping studies to be more inclusive and powerful.

We noted that although the causal variant can be identified without ambiguity from statistical fine-mapping, they often have no clearly known functional implications if located in the noncoding genome, especially when functional priors are not incorporated. Expanding regulatory genome resources across diverse human cell types [94, 95] to advance our knowledge in the noncoding genome and incorporating those into fine-mapping frameworks are necessary to translate the putative causal variants from fine-mapping into mechanistic insights.

Lastly, MHC is a locus of paramount importance to autoimmune disorders [96, 97] but often excluded in recent statistical fine-mapping studies. This is because the MHC locus is very complex: with linkage-disequilibrium over megabses of genomes, and with complicated structural and copy number variations not often observed in other parts of the genome [98]. Thus, fine-mapping using arrays or shotgun sequencing technologies tends to be less productive. A strategy imputing the HLA alleles using data from the high density genotyping array and a set of reference individuals with HLA alleles has been shown to be productive for RA [99] and IBD [100].

Overall, fine-mapping studies for autoimmune disorders have been very productive. They have pinpointed disease causal variants and revealed key insights into the disorders. Building on this success, developments in fine-mapping methods to incorporate studies with various design factors, and resources to interpret the functional impact of causal variants on the molecular and physiological levels, will likely further advance fine-mapping studies and facilitate the therapeutics translation of their findings.

Appendix

Box. 1 Overview of GWAS and statistical fine-mapping models.

(We are not including intercept and covariates term in the below equations, for simplicity).

In GWAS, we test one variant at a time for its association with the phenotype of interest:

$$\boldsymbol{y} = \beta \cdot \boldsymbol{x} + \epsilon$$

where $\boldsymbol{y} =$ is the $n \times 1$ phenotype vector corresponding to n individuals, x is the $n \times 1$ vector denoting the genotype dosages of n individuals at a specific variant position (for each individual, 1 if the individual carries the alternate allele of the variant in heterozygote, 2 if homozygote, and 0 otherwise), β is the effect size of the variant (scalar), and ϵ is a noise term vector (typically normally distributed) of size $n \times 1$.

In contrast, when we perform statistical fine-mapping, we consider a set of m variants in a locus at a time:

$y = \boldsymbol{X} \cdot \boldsymbol{\beta} + \epsilon$

where X is a matrix of size $n \times m$, and β is now a vector of size $m \times 1$.

Not all the variants in a locus are likely causal. We typically assume sparse causal configuration, which means most of the elements of β are zero:

$\beta = \gamma \cdot b$

where γ is the causal indicator vector (1 if a variant is causal, 0 otherwise) with most of the elements being 0, and b is the (true) effect sizes vector when the variant is causal for the phenotype.

In a typical Bayesian statistical fine-mapping, we set a prior distribution for the parameters γ and b such that all the elements of γ have an equal probability of being non-zero, and each element of b follows a normal distribution with pre-specified mean (=0) and variance, to evaluate different sparse causal configurations (γ s). In contrast, functionally informed fine-mapping corresponds to letting the prior distribution of γ to be non-uniform depending on the variant annotations.

Box. 2 Bayesian method overview.

(In this box, we are assuming uniform prior).

Let X be the genotypes in a locus of interest (and the phenotypes), and X_i be the genotype of the variant i in a locus, Maller et al. [39] showed that, the Bayes factor corresponding to a model that variant i is the only causal variant in the locus of interest (M_i) over the model that no variant in the locus is causal (M_0) depends only on the genotype data of the variant i:

$$BF_i = \frac{P(X|M_i)}{P(X|M_0)} = \frac{P(X_i|M_i)}{P(X_i|M_0)}$$

and if we assume there is exactly one causal variant in a locus of interest (i.e., the model $M = M_1 \cup M_2 \cup ... \cup M_i \cup ...$), the posterior probability of that variant being causal is simply proportional to the Bayes factor:

$$P(M_i|X, M) \propto BF_i$$

Wakefield (2007,2009)[37,38] showed that the Bayes factor can be approximated using summary statistics of the variant i alone as:

$$ABF_i = \frac{P(\widehat{\theta}_i|M_i)}{P(\widehat{\theta}_i|M_0)} = \sqrt{1-r_i} \cdot exp(\frac{Z_i{}^2 r_i}{2})$$

where $\widehat{\theta}_i$ is the marginal effect size, Z_i is the z-score, and r_i is the ratio of the prior variance (W) to the total variance of the effect size of variant i ($W + V_i$) in GWAS (we have flipped the denominator and the numerator compared to the original notation of ABF in Wakefield (2007,2009), for convenience). Then the posterior inclusion probability (PIP) can be simply given as:

$$PIP_i = \frac{ABF_i}{\sum_{j=1}^{k} ABF_j}$$

for a locus harboring k variants.

For > 1 causal variants, we cannot use such simple approximations. Let Γ be all possible causal configurations, and Γ_i be a subset that includes variant i in the causal variant set,

$PIP_i = \frac{\sum_{C \in \Gamma_i} BF_C \cdot P(C)}{\sum_{C \in \Gamma} BF_C \cdot P(C)}$. Calculating the Bayes factor for a causal configuration C

$BF_C = \frac{P(X|M_C)}{P(X|M_0)}$ over all the 2^k possible causal configurations as required in the calculation of denominator could be computationally expensive, and different fine-mapping methods have been developed to overcome the computational challenge (e.g., many methods restrict the number of causal variants in the model. FINEMAP[54] performs stochastic search to avoid considering all the possible causal configurations).

Declarations

Competing interests The authors declare no competing interests.

References

1. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108
2. Visscher PM et al (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101:5–22
3. Buniello A et al (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47:D1005–D1012
4. Bůžková P (2013) Linear regression in genetic association studies. PLOS ONE 8:e56976
5. Price AL et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909
6. Jannot A-S, Ehret G, Perneger T (2015) P < 5 × 10–8 has emerged as a standard of statistical significance for genome-wide association studies. J Clin Epidemiol 68:460–465
7. Ulirsch JC et al (2016) Systematic functional dissection of common genetic variation affecting red blood cell traits. Cell 165:1530–1545
8. Tewhey R et al (2016) Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. Cell 165:1519–1529
9. Spain SL, Barrett JC (2015) Strategies for fine-mapping complex traits. Hum Mol Genet 24:R111–R119
10. Broekema RV, Bakker OB, Jonkers IH. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. Open Biol. 10, 190221.
11. Schaid DJ, Chen W, Larson NB (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19:491–504
12. Hutchinson A, Asimit J, Wallace C (2020) Fine-mapping genetic associations. Hum Mol Genet 29:R81–R88
13. Weissbrod O, et al (2020) Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat. Genet. 1–9
14. Wang QS et al (2021) Leveraging supervised learning for functionally informed fine-mapping of cis-eQTLs identifies an additional 20,913 putative causal eQTLs. Nat Commun 12:3394
15. Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11:499–511
16. McCarthy MI et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356–369
17. Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231
18. Wray NR (2005) Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. Twin Res Hum Genet 8:87–94
19. Bulik-Sullivan BK et al (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47:291–295
20. Finucane HK et al (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47:1228–1235
21. Kim S, Misra A (2007) SNP genotyping: technologies and biomedical applications. Annu Rev Biomed Eng 9:289–320
22. Perkel J (2008) SNP genotyping: six technologies that keyed a revolution. Nat Methods 5:447–453
23. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39:906–913

24. Pruim RJ et al (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26:2336–2337
25. Kircher M et al (2019) Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. Nat Commun 10:3583
26. van Arensbergen J et al (2019) High-throughput identification of human SNPs affecting regulatory element activity. Nat Genet 51:1160–1169
27. Findlay GM et al (2018) Accurate classification of BRCA1 variants with saturation genome editing. Nature 562:217–222
28. Rees HA, Liu DR (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet 19:770–788
29. Anzalone AV, Koblan LW, Liu DR (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol 38:824–844
30 Giambartolomei C et al (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLOS Genet. 10:e1004383
31. Hormozdiari F et al (2016) Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet 99:1245–1260
32. Wen X, Pique-Regi R, Luca F (2017) Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. PLOS Genet. 13:e1006646
33. Giambartolomei C et al (2018) A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics 34:2538–2545
34. Foley CN et al (2021) A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. Nat Commun 12:764
35 Goodman SN (1999) Toward evidence-based medical statistics 2: The Bayes Factor. Ann Intern Med. 130:1005–1013
36. Burton PR et al (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678
37. Wakefield J (2007) A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 81:208–227
38. Wakefield J (2009) Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol 33:79–86
39. Maller JB et al (2012) Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet 44:1294–1301
40. Brown AA et al (2017) Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. Nat Genet 49:1747–1751
41. Beecham AH et al (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet 45:1353–1360
42. Yang J et al (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44:369–375
43 Horikoshi M et al (2015) Discovery and fine-mapping of glycaemic and obesity-related trait loci using high-density imputation. PLOS Genet. 11:e1005230
44. Teumer A et al (2019) Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. Nat Commun 10:4130
45. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E (2014) Identifying causal variants at loci with multiple signals of association. Genetics 198:497–508
46. Faye LL, Machiela MJ, Kraft P, Bull SB, Sun L (2013) Re-ranking sequencing variants in the post-GWAS era for accurate causal variant identification. PLOS Genet. 9:e1003609
47. Newcombe PJ, Conti DV, Richardson S (2016) JAM: a scalable Bayesian framework for joint analysis of marginal SNP effects. Genet Epidemiol 40:188–201
48. Udler MS et al (2009) FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. Hum Mol Genet 18:1692–1703
49. Dadaev T et al (2018) Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. Nat Commun 9:2256
50. Servin B, Stephens M (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. PLOS Genet. 3:e114
51. Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association studies and other large-scale problems. Ann Appl Stat 5:1780–1815
52. Chen W et al (2015) Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. Genetics 200:719–736
53. Farh KK-H et al (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518:337–343
54. Benner C et al (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinforma Oxf Engl 32:1493–1501
55. Wen X, Lee Y, Luca F, Pique-Regi R (2016) Efficient integrative multi-SNP association analysis via deterministic approximation of posteriors. Am J Hum Genet 98:1114–1129
56. Wang G, Sarkar A, Carbonetto P, Stephens M (2020) A simple new approach to variable selection in regression, with application to genetic fine mapping. J R Stat Soc Ser B Stat Methodol 82:1273–1300
57. Consortium, GTEx (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330
58. Sinnott-Armstrong N et al (2021) Genetics of 35 blood and urine biomarkers in the UK Biobank. Nat Genet 53:185–194
59 Kichaev G et al (2014) Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 10:e1004722
60. Chen W, McDonnell SK, Thibodeau SN, Tillmans LS, Schaid DJ (2016) Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. Genetics 204:933–958
61. Jiang J et al (2019) Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. Commun Biol 2:1–12
62. Li Y, Kellis M (2016) Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. Nucleic Acids Res. 44:e144
63. Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet 94:559–573
64. Kelley DR et al (2018) Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res 28:739–750
65 Kelley DR (2020) Cross-species regulatory sequence activity prediction. PLOS Comput. Biol. 16:e1008050
66. Hutchinson A, Watson H, Wallace C (2020) Improving the coverage of credible sets in Bayesian genetic fine-mapping. PLOS Comput. Biol. 16:e1007829
67. Schilder BM, Humphrey J, Raj T (2020) echolocatoR: an automated end-to-end statistical and functional genomic fine-mapping pipeline. bioRxiv 2020.10.22.351221. https://doi.org/10.1101/2020.10.22.351221.
68. Liu L, et al. (2020) TreeMap: a structured approach to fine map-880 ping of eQTL variants. Bioinformatics 37:1125–1134

69. Zheng J et al (2017) HAPRAP: a haplotype-based iterative method for statistical fine mapping using GWAS summary statistics. Bioinformatics 33:79–86
70. Kichaev G et al (2017) Improved methods for multi-trait fine mapping of pleiotropic risk loci. Bioinforma Oxf Engl 33:248–255
71. Wen X, Luca F, Pique-Regi R (2015) Cross-population joint analysis of eQTLs: fine mapping and functional annotation. PLOS Genet. 11:e1005176
72. Kichaev G, Pasaniuc B (2015) Leveraging functional-annotation data in trans-ethnic fine-mapping studies. Am J Hum Genet 97:260–271
73. Zou J et al (2019) Leveraging allelic imbalance to refine fine-mapping for eQTL studies. PLOS Genet. 15:e1008481
74 Wallace C et al (2015) Dissection of a complex disease susceptibility region using a Bayesian stochastic search approach to fine mapping. PLoS Genet. 11:e1005272
75. Asimit JL et al (2019) Stochastic search and joint fine-mapping increases accuracy and identifies previously unreported associations in immune-mediated diseases. Nat Commun 10:3216
76. Lam M et al (2019) Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet 51:1670–1678
77. Shi H et al (2021) Population-specific causal disease effect sizes in functionally important regions impacted by selection. Nat Commun 12:1098
78. Liu JZ et al (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 47:979–986
79. Morris AP (2011) Transethnic meta-analysis of genomewide association studies. Genet Epidemiol 35:809–822
80. Mägi R et al (2017) Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. Hum Mol Genet 26:3639–3650
81. Turley P, et al. (2021) Multi-Ancestry Meta-Analysis yields novel genetic discoveries and ancestry-specific associations. bioRxiv 2021.04.23.441003. https://doi.org/10.1101/2021.04.23.441003
82. Lee CH, Eskin E, Han B (2017) Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. Bioinformatics 33:i379–i388
83. Walters K, Cox A, Yaacob H (2019) Using GWAS top hits to inform priors in Bayesian fine-mapping association studies. Genet Epidemiol 43:675–689
84. Walters K, Cox A, Yaacob H (2021) The utility of the Laplace effect size prior distribution in Bayesian fine-mapping studies. Genet Epidemiol.
85. Seldin MF (2015) The genetics of human autoimmune disease: a perspective on progress in the field and future directions. J Autoimmun 64:1–12
86. Plenge RM, Scolnick EM, Altshuler D (2013) Validating therapeutic targets through human genetics. Nat Rev Drug Discov 12:581–594
87. PaÇ Kisaarslan A et al (2020) Blau syndrome and early-onset sarcoidosis: a six case series and review of the literature. Arch. Rheumatol. 35:117–127
88. Kreins AY et al (2015) Human TYK2 deficiency: mycobacterial and viral infections without hyper-IgE syndrome. J Exp Med 212:1641–1662
89. Huang H et al (2017) Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 547:173–178
90. Okada Y et al (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506:376–381
91. Sazonovs A, et al (2021) Sequencing of over 100,000 individuals identifies multiple genes and rare variants associated with Crohns disease susceptibility. medRxiv 2021.06.15.21258641. doi:https://doi.org/10.1101/2021.06.15.21258641.
92. Boisson-Dupuis S, et al (2018) Tuberculosis and impaired IL-23–dependent IFN-γ immunity in humans homozygous for a common TYK2 missense variant. Science Immunology 3.
93. Kerner G et al (2019) Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of patients in a cohort of European ancestry. PNAS 116:10430–10434
94. Dunham I et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
95. Kundaje A et al (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330
96. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A (2017) The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol. 18.
97. Deitiker P, Atassi MZ (2015) MHC genes linked to autoimmune disease. Crit. Rev. Immunol. 35.
98. Miretti MM et al (2005) A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. Am J Hum Genet 76:634–646
99. Raychaudhuri S et al (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 44:291–296
100. Goyette P et al (2015) High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet 47:172–179
101. Zeng B, et al (2017) Constraints on eQTL fine mapping in the presence of multisite local regulation of gene expression. G3 Bethesda Md. 7, 2533–2544.