Genome-wide fine-mapping improves identification of causal variants

Yang Wu[1,2,*], Zhili Zheng[2,3,4], Loic Thibaut[2], Tian Lin[2], Qian Feng[2], Hao Cheng[5], Loic Yengo[2], Michael E. Goddard[6,7], Naomi R. Wray[2,8], Peter M. Visscher[2,9], Jian Zeng[2,*]

[1] Institute of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China

[2] Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

[3] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[4] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[5] Department of Animal Science, University of California, Davis, CA, USA

[6] Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria, Australia

[7] Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia

[8] Department of Psychiatry, University of Oxford, Oxford, UK

[9] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK

[*] Correspondence: Yang Wu (yang.wu@wchscu.edu.cn), Jian Zeng (j.zeng@uq.edu.au)

Abstract

Fine-mapping refines genotype-phenotype association signals to identify causal variants underlying complex traits. However, current methods typically focus on individual genomic loci and do not account for the global genetic architecture. Here, we demonstrate the advantages of performing genome-wide fine-mapping (GWFM) with functional annotations and develop methods to facilitate GWFM. In simulations and real data analyses, GWFM outperforms current methods across multiple metrics, including error control, mapping power, resolution, precision, replication rate, and trans-ancestry phenotype prediction. Across 48 complex traits, we identify credible sets that collectively explain 18% of the SNP-based heritability (h^2_{SNP}) on average, with 30% credible sets located outside genome-wide significant loci. Leveraging the genetic architecture estimated from GWFM, we predict that fine-mapping over 50% of h^2_{SNP} would require an average of 2 million samples. Finally, as proof-of-principle, we highlight a known causal variant at FTO for body mass index and identify novel missense causal variants for schizophrenia and Crohn's disease.

Introduction

Despite the success of genome-wide association studies (GWAS) in identifying trait-associated variants[1], the causal variants underlying complex traits remain unresolved due to extensive linkage disequilibrium (LD) between SNPs[2] and the polygenic nature of these traits[3,4]. Statistical fine-mapping, often employing a Bayesian mixture model (BMM) that jointly fit multiple SNPs, offers a direct approach to identifying candidate causal variants[5]. However, current fine-mapping methods focus on genome-wide significant loci only (e.g., 1-2 Mb windows centred on lead SNPs after LD clumping[6-12]) or consider one genomic region at a time (e.g., a LD block[13]), in isolation from the rest of the genome.

While widely used, region-specific analysis has several limitations. First, restricting fine-mapping to GWAS loci alone, which often explain only a small fraction of the total genetic variance[14,15], omits meaningful signals that have not yet reached the stringent GWAS significance threshold. Second, the prior probability of association can be influential but is often conservatively predetermined (e.g., as the inverse of the number of SNPs in the region[6,7,16]) due to challenges in estimating the genetic architecture within a region. Third, fine-mapping can benefit from incorporating functional genomic annotations[10-12,17], but region-specific methods often estimate functional priors separately before performing functionally-informed fine-mapping[10-12,17], rather than jointly modelling GWAS data and functional annotations. Finally, none of the current methods estimate the power of identifying the causal variants for a trait, which is critical to inform the experimental design of prospective studies (such a power analysis is available in GWAS[18] but absent in fine-mapping).

These limitations can be addressed with genome-wide fine-mapping (GWFM) analysis. Genome-wide Bayesian mixture models (GBMMs), extensively used for predicting breeding values in agriculture[19-21] and complex trait phenotypes in humans[22-24], have recently emerged as a method for GWFM[25,26]. Unlike conventional GWAS and region-specific fine-mapping approaches, GBMMs jointly fit genome-wide SNPs in the model, simultaneously estimating the genetic architecture and functional priors through "information borrowing"[24,25]. For example, SNPs sharing the same functional annotation across the genome collectively prioritize that annotation based on their aggregated evidence of trait association, which, in turn, informs the estimation of individual SNP effects. This learning process is often performed iteratively using Markov chain Monte Carlo (MCMC) sampling, leading to posterior inference with superior asymptotic accuracy compared to variational inference[27,28], though MCMC sampling can be computationally intensive with high-density SNPs. Moreover, GBMMs estimate polygenicity and the distribution

of variant effect sizes[3,20,22-24,29], enabling the prediction of power for prospective studies with larger sample sizes. However, relevant theory and methods have not yet been developed.

In this study, we comprehensively evaluated the performance of GWFM analysis using SBayesRC[24], a state-of-the-art GBMM that enables efficient MCMC-based fitting of all common SNPs with functional annotations. Through extensive simulations under various genetic architectures, we compared methods using multiple metrics, including posterior inclusion probability (PIP) calibration, fine-mapping power, mapping precision, credible set size, replication rate in independent samples, and out-of-sample prediction accuracy using fine-mapped variants. We developed an LD-based method to construct local credible sets (α-LCSs) for GBMM, each capturing a causal variant with posterior probability α, and estimated the proportion of SNP-based heritability (h^2_{SNP}) explained by these LCSs. To quantify overall fine-mapping uncertainty, we provided a global credible set (α-GCS) that captures α% of all causal variants for the trait. Leveraging the genetic architecture estimated from SBayesRC, we further developed a method to predict fine-mapping power and the h^2_{SNP} explained by identified variants in prospective studies, enabling estimation of the minimal sample size required to capture a desired proportion of causal variants or h^2_{SNP}. Finally, we applied SBayesRC to 599 complex traits and diseases with 13 million SNPs and compared the fine-mapping results for 48 well-powered traits, primarily using data from the UK Biobank[30] (UKB).

Results

Method overview

We selected SBayesRC as the method for GWFM (**Fig. 1**) because it outperforms other GBMMs in polygenic prediction[24]. SBayesRC is a multi-component mixture model that simultaneously analyzes all SNPs across approximately independent LD blocks[13,31], using a hierarchical prior to borrow information from functional annotations across the genome for fine-mapping within blocks, where both annotation and SNP effects are jointly estimated from the data (**Methods**). Unlike existing fine-mapping methods that focus on specific GWAS loci, SBayesRC accounts for long-range LD (variable LD block sizes with a minimum of 1 Mb, including the major histocompatibility complex (MHC) region) and maps causal signals over the entire genome (**Supplementary Table 1**). To optimize fine-mapping performance, we heuristically estimated the number of mixture components, adapted a tempered Gibbs sampling algorithm[32] to improve mixing properties, and implemented a method[33] to assess MCMC convergence using multiple independent chains (**Methods**). Further methodological differences between our approach and existing methods are discussed in **Supplementary Note 1**.

To capture the uncertainty in identifying causal variants driven by LD, we developed an LD-based method for constructing α-LCSs (**Methods; Supplementary Fig. 1**). Our approach forms an α-LCS for each candidate causal variant with a leading PIP by grouping it with SNPs in LD (r^2 > 0.5) until their combined PIPs sum to α. To refine the LCSs, we further filter them based on the posterior h^2_{SNP} enrichment probability (PEP > 0.7) to ensure that the selected LCS explains more h^2_{SNP} than a random set of SNPs with the same size, using MCMC samples of SNP effects (**Methods**). Additional justification for the r^2 and PEP thresholds is provided in **Supplementary Note 2** and **Supplementary Fig. 2-3**.

Since SBayesRC estimates the genetic architecture of the trait, this framework lends itself to quantifying overall uncertainty in fine-mapping and predicting power for future GWFM studies. The overall uncertainty in the current study is captured by our α-GCS, which is expected to cover α% of all causal variants for the trait, with the size of α-GCS reflecting power and genetic architecture (**Methods**). Fine-mapping power prediction for prospective studies is achieved by deriving the sampling distribution of PIP given a sample size (**Methods**). This method allows us to estimate the minimal sample size required to achieve the desired power to identify all causal variants or those explaining a specific proportion of h^2_{SNP}. Although we focused on European ancestry in this study, our power prediction method is, in principle, applicable to other ancestries; however, the genetic architecture would need to be estimated using ancestry-specific GWAS data, as it may vary across populations. Our fine-mapping power prediction method is analytically tractable and available as a publicly accessible online tool (https://sbayes.pctgplots.cloud.edu.au/shiny/power/).

Calibration of fine-mapping methods under various genetic architectures

We performed extensive genome-wide simulations to evaluate SBayesRC for GWFM, comparing to FINEMAP[7], SuSiE[6], FINEMAP-inf[8], SuSiE-inf[34], PolyFun+SuSiE[10], and another two GWFM methods, SBayesR[22] and SBayesC (two-component SBayesR). For a fair comparison, we defined fine-mapping regions consistently across all methods using independent LD blocks and applied the default parameter settings recommended for each method (**Methods**). Using 100,000 individuals with ~1 million HapMap3 SNPs from the UKB[30], we simulated three genetic architectures: 1) a sparse architecture with 1% randomly selected causal SNPs explaining 50% of phenotypic variance, 2) a large-effects architecture where 10 random causal variants contributed 10% of phenotypic variance and the rest 40%, and 3) an LD-and-MAF-stratified (LDMS) architecture with causal variants sampled from high LD score and high minor allele frequency (MAF) SNPs (**Methods**).

144 Results showed that GWFM methods generally performed better in terms of PIP calibration
145 compared to region-specific methods, with PIPs from SBayesRC closely aligning with true
146 probabilities of causality (measured by true discovery rate, TDR) across all genetic
147 architectures (**Fig. 2a,d,g**). In contrast, SuSiE and FINEMAP exhibited notable inflation in high-
148 PIP SNPs (**Fig. 2b,e,h**), indicating poor control of false discovery rate (FDR = 1-TDR). While
149 SuSiE-inf and FINEMAP-inf performed reasonably well under the sparse and large-effects
150 architectures (despite some deflation in low-PIP SNPs), their PIPs failed to track TDR accurately
151 under the LDMS architecture, where causal variants were not randomly distributed (**Fig. c,f,i**).
152 Polyfun+SuSiE, when incorporating the same LD/MAF annotations as SBayesRC, showed
153 improved performance over SuSiE and FINEMAP but still struggled with FDR control (**Fig. 2i**).
154 The alternative GWFM methods, SBayesC and SBayesR, were inferior to SBayesRC under the
155 large-effects and LDMS architectures, respectively, highlighting the importance of using
156 multiple mixture components and integrating informative genomic annotations for robust
157 calibration.
158
159 **Assessing mapping power, resolution, and precision via simulations**
160 We first compared the power of different fine-mapping strategies, GWFM and GWAS loci-based
161 fine-mapping, by evaluating the proportion of causal variants captured by α-LCSs. For a fair
162 comparison, GWAS loci-based fine-mapping used the same PIP estimates as GWFM with
163 SBayesRC but was restricted to 2Mb regions around GWAS lead SNPs (P-value < 5×10^{-8}). As
164 expected, GWFM was significantly more powerful, with 46%-61% improvement across genetic
165 architectures (**Fig. 3a-c**). This highlights the advantage of considering the entire genome,
166 including regions not genome-wide significant in GWAS, in fine-mapping.
167
168 Next, we compared different fine-mapping methods, focusing on SBayesRC and SuSiE-inf, as
169 SuSiE-inf had the best PIP calibration and least FDR inflation among the region-specific methods.
170 Our results showed that SBayesRC was significantly more powerful (**Fig. 3d-f**), with similar TDR
171 (**Supplementary Fig. 4a-c**) but smaller LCS sizes at the same α threshold, indicating superior
172 mapping resolution (**Fig. 3g-i**). For α=0.9, SBayesRC outperformed SuSiE-inf by up to 194%
173 increase in power and 21% reduction in average LCS size across the three genetic architectures.
174 Polyfun+SuSiE, which incorporated LD/MAF annotations through a stepwise analysis,
175 performed slightly better than SuSiE-inf under the LDMS architecture but remained inferior to
176 SBayesRC. Mapping precision was assessed as the distance from an identified variant (PIP > 0.9)
177 to the nearest causal variant. Under the sparse architecture, 98% of SNPs identified by
178 SBayesRC were causal, and 99% of SNPs with PIP > 0.9 were within 18kb of a causal variant,
179 leading to a 2% increase in TDR and a 68% reduction in distance compared to other methods

180 (**Fig. 3j**). The advantage of SBayesRC was even more pronounced under the large-effects and
181 LDMS architecture (**Fig. 3k-l**), likely due to its ability to differentially weigh SNPs based on LD
182 and MAF annotations, thereby improving causal variants identification.
183
184 Furthermore, we used these simulations to evaluate GCS (**Supplementary Fig. 5**) and
185 conducted sensitivity analyses to confirm the robustness of SBayesRC to missing annotations
186 (**Supplementary Fig. 6),** unobserved causal variants (**Supplementary Fig. 7),** and the chosen
187 value of LD matrix factorization parameter (**Supplementary Fig. 8).** To understand why GWFM
188 had higher power, we investigated power gains in causal variants with different LD and MAF
189 properties (**Supplementary Fig. 9)** and assessed SBayesRC within each LD block separately
190 (**Supplementary Fig. 10**). These results are discussed in **Supplementary Note 3**. Overall, these
191 simulation results suggested that SBayesRC is a reliable method for GWFM and can substantially
192 improve fine-mapping performance.
193
194 **Assessing replication rate, effect size estimation, and prediction accuracy in real data**

195 In real data analysis, direct evaluation of mapping power and precision is challenging since true
196 causal variants are often unknown. Thus, we assessed methods using metrics that do not
197 require knowledge of causal variants. First, we evaluated replication rate by calculating the
198 proportion of variants with PIP > 0.9 in a GWAS sample that were also identified in an
199 independent replication sample at the same or a lower PIP threshold. Assuming no systematic
200 confounding, a higher replication rate suggests that the method identifies more genuine causal
201 variants, as false positives are less likely to replicate. Using UKB height data (n=100,000) as a
202 GWAS sample, SBayesRC achieved the highest replication rate in each PIP threshold, improving
203 by 1% (compared to SuSiE-inf) to 11% (compared to FINEMAP) at PIP > 0.9 when replication
204 *n*=100,000 (**Fig. 4a**). When replication sample size was doubled, the improvement increased to
205 4% compared to SuSiE-inf and 19% compared to FINEMAP (**Supplementary Fig. 11**).
206
207 We assessed bias in effect size estimates of putative causal variants through regressing their
208 marginal effect sizes from the replication sample on the joint effect sizes estimated from the
209 GWAS sample, expecting a regression slope of one for unbiased estimation. In the UKB height
210 analysis, SBayesRC produced the minimal bias, with a regression slope of 0.98, outperforming
211 all the other methods (**Fig. 4b**). This likely reflects the capacity of SBayesRC to estimate genetic
212 architecture genome-wide, whereas other methods estimate genetic architecture locally or use
213 preset parameters.
214
215 Given that common causal variants and their effect sizes are mostly shared across ancestries[35,36],
216 knowing the causal variants should improve trans-ancestry prediction in complex traits.
217 Therefore, we used fine-mapped variants and their posterior effect sizes from UKB European
218 (EUR) ancestry individuals to predict phenotypes in African (AFR), East Asian (EAS), and South
219 Asian (SAS) ancestry groups. We selected 6 complex traits that had at least 50 SNPs with PIP >
220 0.9. Compared to SuSiE-inf, SBayesRC improved the trans-ancestry prediction accuracy, with
221 nearly a 10-fold increase in mean relative prediction R^2 across traits and ancestries (**Fig. 4c**).
222 Similar improvements were observed when using 0.9-LCSs instead of individual SNPs
223 (**Supplementary Fig. 12)** or when comparing to Polyfun+SuSiE with the same functional
224 annotations (**Supplementary Fig. 13**). We further quantified the transferability of fine-mapped
225 SNPs by computing the ratio of per-SNP prediction accuracy in a hold-out EUR sample versus a
226 different ancestry. The result showed that relative prediction accuracy increased with PIP in the
227 EUR GWAS sample (**Fig. 4d**), consistent with a model of shared genetic effects across ancestries.
228
229 We validated these advantages of SBayesRC through simulations, confirming its superior power
230 (**Supplementary Fig. 14**), replication rate (**Supplementary Fig. 15a**), effect size estimation

(**Supplementary Fig. 15b**), and out-of-sample prediction using fine-mapped variants (**Supplementary Fig. 16**), with further discussions in **Supplementary Note 4**. We also quantified the replication rate in the reverse case where the GWAS sample size was 200,000 but the replication sample size was only 100,000, to mimic the reality that the sample size of replication data is often much smaller than that of discovery, and again observed superior replication rate of our SBayesRC (**Supplementary Fig. 17a,b**). These results suggested that SNPs identified by SBayesRC are more likely to be causal, as indicated by higher replication rates, while the improved prediction accuracy likely arises from both better causal variant identification and more accurate effect size estimation.

Prediction of fine-mapping power and variance explained for future studies

As a unique feature of the GWFM approach, the genetic architecture estimated from SBayesRC provides key information to predict the proportion of causal variants identified (fine-mapping power) and the proportion of h^2_{SNP} explained by these variants (PHE) in future studies (**Methods**). To evaluate our approach, we computed the predicted values of power and PHE across varying GWAS sample sizes and projected the outcome of GWFM with SBayesRC onto the prediction for a simulated trait (under the sparse architecture), height[30,37], high density lipoprotein (HDL)[30,38], schizophrenia (SCZ)[39,40], and Crohn's disease (CD)[30,41], which represented diverse genetic architectures, using two datasets with different sample sizes for each trait (**Fig. 5a-c**). Despite some variation across traits, GWFM outcomes were broadly consistent with the theoretical predictions (**Fig. 5d,e**).

Using the latest GWAS summary statistics from the Psychiatric Genomics Consortium (PGC)[39], we identified 13 SNPs and 174 LCSs for SCZ, collectively explaining 4.2% of h^2_{SNP} on the liability scale (converted from observed $\hat{h}^2_{SNP}$ based on ref[42]). These estimates aligned closely with our theoretical prediction, based on the 53,386 cases and 77,258 controls in their study[39], which is equivalent to a sample size of 228,810 on the liability scale (ref[43]; **Methods**). For a prospective SCZ study using GWFM, we predict that ~180k cases (with equal controls and a population prevalence of 0.01) would be required to fine-map 1,000 common causal variants (estimated to be 1.2% of all common causal variants), collectively explaining ~20% of h^2_{SNP} (**Fig. 5**). Increasing the sample size to ~550k cases would allow for the identification of 10% of causal variants explaining 50% of h^2_{SNP}, while fine-mapping variants accounting for 80% of h^2_{SNP} would require 1.4 million cases according to our prediction.

For height, based on the UKB data (n=350k), we predicted that with 5 million samples, ~10,000 variants would be identified with individual PIP > 0.9 or ~30,000 when considering LCSs,

267 explaining up to 95% of the genetic variance (**Fig. 5**). This prediction is consistent with a recent
268 GWAS of 5 million individuals, which reported 12,111 independently significant SNPs identified
269 from COJO accounting for nearly all of the common h^2_{SNP} in height[36].
270
271 **Applying genome-wide fine-mapping to a range of complex traits**
272 We applied GWFM with SBayesRC to 599 complex traits (597 UKB traits plus SCZ[39] and CD[41])
273 and developed an online resource to query these fine-mapping results (see **Data Availability**;
274 **Supplementary Table 3**). The 597 UKB traits were selected based on z-score > 4 and high
275 confidence for heritability estimates using LD score regression[44]
276 (https://zenodo.org/records/7186871). To better capture causal variants, we used 13 million
277 imputed SNPs with functional genomic annotations from Finucane et al.[45]. Here, we focus on 48
278 well-powered traits, including SCZ, CD[39,41] and 46 UKB traits measured in the European
279 ancestry individuals (**Methods**).
280
281 Across the 48 traits, we identified 1,820 SNPs with PIP > 0.9, with the number of fine-mapped
282 SNPs correlated strongly with the number of GWAS loci identified by LD clumping
283 (**Supplementary Fig. 18-19**). However, 1,158 of these SNPs were not GWAS lead SNPs, and
284 14.9% located outside GWAS loci (1Mb regions around independently significant SNPs) (**Fig.**
285 **6a**). Consistent with our simulation results (**Fig. 3a-c**), this highlights the importance of
286 conducting genome-wide fine-mapping to capture all relevant signals. Notably, 469 fine-
287 mapped SNPs (25.8%) were associated with multiple traits, suggesting prevalent pleiotropy in
288 the human genome. Moreover, the MAF of pleiotropic SNPs decreased as the number of affected
289 traits increased (**Supplementary Fig. 20**), consistent with a model of negative selection, where
290 highly pleiotropic variants are either purged from the population or kept at low frequencies[46].

We identified 19,863 0.9-LCSs with a median size of 5 SNPs, of which 29.8% were located outside the genome-wide significant loci (**Fig. 6a**). While these LCSs captured only 0.9% of causal variants, they explained 17.7% of h^2_{SNP}, with 2.7% PHE (relative proportion of PHE = 2.7%/17.7=15.3%) attributed to LCSs in non-significant GWAS regions (**Fig. 6a**). Across trait categories, cognitive traits had the highest proportion of LCSs and PHE observed outside GWAS loci (**Fig. 6b**), consistent with the high polygenicity of cognitive traits reported previously[3]. On average, the 0.1-GCS (expected to contain 10% of all causal variants) comprised ~1.5% of the SNPs, explaining 44.7% of the h^2_{SNP} (**Fig. 6c,d**). Reproductive traits had the largest 0.1-GCSs (2.3% of SNPs for 10% causal variant coverage), while blood cell traits had the smallest (0.55%) but explained a large proportion of h^2_{SNP} (54.6%), highlighting the impact of genetic architecture on fine-mapping uncertainty.

Our results recapitulated previous findings and led to new discoveries. For example, we identified 13 SNPs at PIP > 0.9 for SCZ from the latest meta-analysis, 5 of which overlapped with SNPs identified by FINEMAP in their study[39], and all 8 FINEMAP-identified SNPs were included in our 0.9-LCSs. Among the 5 novel fine-mapped SNPs not identified by FINEMAP based on individual PIP, 3 were missense variants (**Supplementary Fig. 21a-d**). We also identified novel putative causal variants for CD (**Supplementary Table 4**), fine-mapping 15 SNPs, including 5 missense variants, and successfully recapitulated all 3 variants identified in a previous study using the same dataset[41]. Notably, at the *NOD2* locus, we identified 4 putative causal variants (**Supplementary Fig. 22**), consistent with the well-established role of this locus in CD etiology. As further support for our method, we identified non-synonymous variants in 11 well-established causal genes for Crohn's disease, schizophrenia, height, and HDL (**Supplementary Table 5**).

316

317 Leveraging the estimated genetic architecture for these 48 traits, we predicted the power of
318 prospective fine-mapping studies. With a sample size of 2 million individuals, we predict an
319 average fine-mapping power of 10.4% (**Fig. 6e**) and an average PHE of 57.9% (**Fig. 6f**). The
320 predicted power and PHE varied substantially across trait categories: blood cell traits had both
321 the highest power (12.8%) and PHE (69.4%), while cognitive traits had the lowest (5.2% power,
322 35.9% PHE). To achieve a PHE of 50%, blood cell traits required 1 million individuals, while
323 cognitive traits required 4 million. For PHE = 80%, the required sample sizes increased to 3
324 million for blood cell traits and 8 million for cognitive traits (**Fig. 6f**).

325

326 **Incorporating functional annotations improves fine-mapping**

327 SBayesRC incorporates functional annotations by learning their weights from the data, rather
328 than relying on preassigned values. We observed multiple lines of evidence supporting the
329 contribution of functional annotations to fine-mapping. First, although the impact of
330 annotations varied across traits, certain functional annotations consistently had stronger
331 impact on SNP effect weighting (**Supplementary Fig. 23**). Second, the 1,820 fine-mapped SNPs
332 showed greater enrichment than GWAS significant SNPs in functional genomic regions, such as
333 coding sequences, transcription start sites, non-synonymous variants, and conserved regions,
334 and were more depleted in repressed regions (**Fig. 7a**). Third, fine-mapping power enrichment
335 was strongly correlated with per-SNP heritability enrichment (r = 0.93), which is not subject to
336 ascertainment bias (**Supplementary Fig. 24**). Fourth, among 13 fine-mapped SNPs for SCZ,
337 running SBayesR (without annotations) identified 4 at PIP > 0.9 and the other 7 at lower but
338 still meaningful PIP values (>0.5), suggesting that functional annotations enhanced fine-
339 mapping power. Finally, some putative causal variants could not be distinguished from SNPs in
340 high LD without considering functional annotations, as demonstrated in the examples below.

341

342 One notable example is the variant rs1421085 at the *FTO* locus, a known causal variant for body
343 mass index (BMI)[47], which was fine-mapped in a credible set from our analysis for body fat
344 percentage (BFP), hip circumference (HC) and waist circumference (WC). Experimental
345 evidence supports its causal role in regulating adipocyte thermogenesis[47]. Unlike the standard
346 GWAS results, where many SNPs in *FTO* reached genome-wide significance, both SBayesRC and
347 SuSiE-inf identified a 5-SNP 0.9-LCS that included rs1421085 for BMI, along with four additional
348 SNPs in almost complete LD (minimum LD r^2=0.997) and with strong annotation overlaps (**Fig.
349 7b**). Notably, rs1421085 was prioritized in SBayesRC (PIP=0.47), likely due to its cross-species
350 conservation annotations (**Fig. 7c**), whereas SuSiE-inf prioritized the GWAS lead SNP instead,
351 underscoring the importance of incorporating functional annotations into fine-mapping.

352 Additionally, SBayesRC identified a secondary signal with another 0.9-LCS, which aligned with
353 the result of SuSiE-inf but had not been previously reported. The lead SNP rs76488452
354 (PIP=0.41) resided in a conserved region across primates and was significant in COJO analysis[48]
355 (P=1.8×10^{-17}) when conditioned on rs1421085. Notably, rs76488452 was only nominally
356 significantly in GWAS marginal analysis (P=3.6×10^{-4}), and its trait-increasing allele was in
357 negative LD (r=-0.16) with that of rs1421085, indicating a masked effect at the SNP[49] (estimated
358 masked effect size $b_2 - r \times b_1$ = 0.02, consistent with the reported GWAS marginal effect size).
359
360 We highlight additional examples where functional annotations improved fine-mapping. We
361 recapitulated a missense variant (rs13107325) in *SLC39A8*, a gene implicated in the latest SCZ
362 analysis for its function in regulating dendritic spine density[50,51]. This variant was identified
363 through an aggregated effect from multiple key functional annotations at the SNP (**Fig. 7d-e**).
364 Another notable SCZ variant is rs11854184 in *SECISBP2L* (**Supplementary Fig. 21a**), a gene
365 highly expressed in brain-related tissues (**Supplementary Fig. 25**). *SECISBP2L* is specifically
366 expressed in differentiating oligodendrocytes, where the SECISBP2L-DIO2-T3 pathway
367 regulates oligodendrocyte differentiation during myelin development[52]. SNP rs11854184 is a
368 missense variant conserved across primates and mammals, with a high PIP (0.91) despite not
369 reaching genome-wide significance (P-value = 1.6e-6). Although the over-expression of
370 *SECISBP2L* in the brain can be coincidental and further experimental validation is needed to
371 confirm causality, this finding suggests that variants below genome-wide significance can still
372 be biologically relevant and merit fine-mapping analysis. Moreover, compared to a recent
373 exome-wide association study for CD[53], we identified a novel gene (*SLAMF8*) with missense
374 putative causal variants (**Supplementary Fig. 26**). These results highlight the power of
375 SBayesRC in identifying putative causal variants and provide a valuable resource for
376 downstream analysis and functional validation.
377
378 **Discussion**
379 In this study, we comprehensively evaluated GWFM using SBayesRC through extensive
380 simulations and real data analyses, compared with existing fine-mapping methods that analyze
381 one genomic region at a time. SBayesRC demonstrated superior PIP calibration, leading to
382 better FDR control, particularly when the genetic architecture deviated from model
383 assumptions (**Fig. 2**). Across various metrics, SBayesRC outperformed the best alternative
384 method (SuSiE-inf) by 194% in mapping power, 21% in resolution, and 68% in precision (**Fig.
385 3**). In real trait analyses, SBayesRC achieved higher replication rates, lower estimation bias, and
386 greater prediction accuracy using fine-mapped SNPs in independent validation samples (**Fig. 4**),
387 identifying putative causal variants missed by other methods (**Fig. 7**). To date, most fine-

388 mapping studies have focused exclusively on genome-wide significant GWAS loci[6-12]. However,
389 our analysis revealed that 30% of the local credible sets (median size=5) identified by SBayesRC
390 were beyond genome-wide significant loci, contributing to 15% of the h^2_{SNP} captured by all local
391 credible sets (**Fig. 6a**). These findings suggest GWFM as a more effective approach than current
392 fine-mapping strategy that focuses solely on GWAS loci or individual regions, and highlight
393 SBayesRC as a reliable and superior method for GWFM analysis.
394
395 The advantage of GWFM over GWAS loci-based fine-mapping arise from several key aspects.
396 First, incorporating functional annotations improves fine-mapping power. Unlike GWAS, which
397 does not consider functional annotations, SBayesRC incorporates functional genomic
398 information, improving SNP effect weighting and the ability to identify causal variants that
399 would otherwise be overlooked in GWAS. Second, fine-mapping can identify causal variants
400 masked by LD that do not reach genome-wide significance in GWAS. Previous works[49,55] have
401 shown that the LD masking between causal variants is common in complex traits, which can
402 weaken their marginal association signals. However, fine-mapping methods that assess joint
403 association signals across SNPs can recover these masked effects, identifying causal variants
404 with strong PIPs despite their weaker marginal GWAS signals. Third, P-value and PIP serve
405 different purposes in managing false positives. The genome-wide significance threshold in
406 GWAS is designed to control of false positive rate under the null hypothesis, suffering from
407 multiple testing burdens[2]. In contrast, PIP directly controls the FDR within a Bayesian
408 framework, automatically accounting for multiple comparisons[54]. For example, with a PIP
409 threshold of 0.9, 90% of identified variants are expected to be true positives, while 10% may be
410 false positives. In comparison, GWAS significance threshold is often more conservative,
411 potentially leading to missed true causal variants. Together, we recommend directly performing
412 GWFM analysis, rather than restricting fine-mapping to genome-wide significant loci identified
413 from GWAS.
414
415 The advantages of SBayesRC over the region-specific fine-mapping methods is likely due to the
416 following factors. First, SBayesRC involves a genome-wide analysis fitting all SNPs jointly across
417 LD blocks. Compared to the existing practice focusing on 1-2 Mb regions[6-12], using LD blocks
418 better accounts for long-range LD, including the MHC region which is often excluded from the
419 analysis[10,39] (**Supplementary Fig. 27**). Although in this study all method comparisons were
420 based on using same LD blocks as fine-mapping regions, SBayesRC benefits from genome-wide
421 fitting which allows for better estimation of genetic architecture parameters from the full set of
422 SNPs. Second, SBayesRC assumes a more realistic distribution of SNP effects through using
423 MAF/LD groups along with other functional annotations. Unlike other fine-mapping methods[6-12],

424 which either ignore annotation data or lack a unified framework to integrate it with GWAS data,
425 SBayesRC estimates the impact of functional annotations and SNP effects within the same model,
426 enabling a proper Bayesian learning process. This framework estimates, instead of preassign,
427 weights to trait-relevant annotations, and is robust to the inclusion of irrelevant annotations
428 (**Fig. 2a**). Third, SBayesRC utilizes Gibbs sampling algorithm to estimate model parameters and
429 PIPs, which is asymptotically exact. In contrast, FINEMAP and FINEMAP-inf use shotgun
430 stochastic search, while both SuSiE and SuSiE-inf rely on variational Bayes. Previous studies
431 have demonstrated that MCMC sampling generally leads to a higher accuracy in capturing the
432 posterior distribution than variational Bayes[27].
433
434 In addition to the improved fine-mapping performance, we introduced new features for GWFM
435 using MCMC sampling. First, based on the posterior samples of SNP effects, we refined LCS by
436 PEP, ensuring that all included SNPs explain more h^2_{SNP} than a random set. This approach is
437 similar but more interpretable than the "purity" parameter in SuSiE (**Supplementary Note 1;**
438 **Supplementary Fig. 28-29**). Second, while a LCS quantifies local uncertainty in identifying a
439 causal variant, a GCS quantifies uncertainty in identifying all causal variants, reflecting both the
440 power of current study and the genetic architecture of the trait. However, computing GCS
441 requires an estimate of the total number of causal variants, which is better achieved when
442 analysing all SNPs jointly in the model. Third, similar to estimating genome-wide polygenicity
443 and h^2_{SNP}, we estimated the proportion of polygenicity and h^2_{SNP} explained by LCSs and GCSs.
444 The analysis of 48 complex traits showed that although 19,863 0.9-LCS were identified, they
445 captured only 1% of all common causal variants, explaining 18% genetic variance, indicating
446 that many causal variants with small effects remain undiscovered. Fourth, the genetic
447 architecture estimation enables power calculations for fine-mapping, informing the sample size
448 required to identify a certain proportion of causal variants or h^2_{SNP} explained in future studies.
449 The robustness of this approach is supported by real data projections aligning with theoretical
450 predictions (**Fig. 5**). Finally, to address poor mixing in standard Gibbs sampling when analyzing
451 SNPs with extremely high LD, we implemented a scalable tempered Gibbs sampling algorithm[32]
452 and assessed the convergence of each SNP's PIP using multiple independent chains, facilitating
453 robust posterior inference.
454
455 We note several limitations of this work. First, there are certainly more complicated scenarios
456 about effect size distribution for causal variants that have not been investigated in our
457 simulations. However, SBayesRC remains one of the most flexible models available, as its
458 assumed multi-component Gaussian mixture is well-suited to accommodate various scenarios.
459 Second, unlike individual-level models where each PIP is estimated conditional on all other

460 SNPs jointly, SBayesRC is a summary-level model that ignores LD between LD blocks. As a result,
461 SNPs outside the block influence PIP estimates only through the mixture distribution of SNP
462 effects. Third, we applied SBayesRC only to GWAS summary data from relatively homogenous
463 populations (inferred European ancestry) and the robustness of the methods on GWAS data
464 based on trans-ancestry meta-analyses is not investigated. Future work should evaluate its
465 performance in diverse populations. Fourth, SBayesRC requires LD information from a
466 reference cohort that matches with GWAS ancestry without substantial sampling variation. Our
467 simulation showed that when using a LD reference of thousands of samples (n=3,642),
468 SBayesRC remained robust (**Supplementary Fig. 30**). However, with a smaller LD reference
469 (n=500), we observed some inflation in PIPs, indicating the importance of sufficient LD
470 reference sample size for maintaining accuracy. Fifth, to construct LCS, a LD threshold was
471 arbitrarily chosen to define a set of SNPs in LD with the candidate causal variant. Although we
472 have explored the robustness to this threshold (**Supplementary Fig. 3**), future methodological
473 advancements, such as Bayesian hypothesis testing with hierarchical clustering[56], could be used
474 to relax this condition. Sixth, our power prediction is based on the genetic architecture
475 estimates given a SNP set. However, as sample sizes increases, more common SNPs may be
476 detected, potentially affecting polygenicity and SNP-based heritability estimates. Seventh,
477 although we incorporated a comprehensive set of functional annotations from BaselineLD
478 (v2.2)[57], these annotations remain incomplete in genome coverage and lack tissue and cell-type
479 diversity. Incorporation of context-dependent annotations, such as those derived from single-
480 cell data, could further enhance fine-mapping accuracy by capturing cell-type-specific
481 regulatory effects. Despite these limitations, our study provides a powerful and robust GWFM
482 framework for identifying causal variants, highlighting the advantages of this approach over
483 existing region-specific fine-mapping methods. With its capacity to enhance mapping power in
484 the current study and to predict mapping power for future studies, we anticipate that GWFM
485 using a state-of-the-art GBMM will become the preferred method for fine-mapping complex
486 traits.

Methods

Ethics approval

The University of Queensland Human Research Ethics Committee B (2011001173) provides approval for analysis of human genetic data used in this study on the high-performance cluster of the University of Queensland.

Low-rank GBMM

We used state-of-the-art GBMM that employed a low-rank model to improve computational efficiency and robustness[24]. As described below, the low-rank GBMM can be derived from the individual-level model. Consider a multiple linear regression of phenotypes on genotypes:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of complex trait phenotypes and $\mathbf{X}$ is an $n \times m$ matrix of mean-centred genotypes, $\boldsymbol{\beta}$ is $m \times 1$ vector of SNP effects on the trait, and $\mathbf{e}$ is $n \times 1$ vector of residuals with $\mathrm{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$. Let

$$\mathbf{K} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\mathbf{X}'n^{-1} \tag{2}$$

where $\boldsymbol{\Lambda}$ and $\mathbf{U}$ are diagonal matrix of eigenvalues and matrix of eigenvectors for the LD correlation matrix $\mathbf{R} = \mathbf{X}'\mathbf{X}/n$, respectively. It follows that $\mathbf{K}'\mathbf{K} = \mathbf{P}n^{-1}$, where $\mathbf{P}$ is the projection matrix of $\boldsymbol{y}$ on the column space of $\mathbf{X}$, and $\mathbf{K}\mathbf{K}' = \mathbf{I}n^{-1}$. Multiplying both sides of Eq (2) by $\mathbf{K}$ gives

$$\mathbf{K}\mathbf{y} = \mathbf{K}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{e} \tag{3}$$

or

$$\mathbf{w} = \mathbf{Q}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4}$$

When only the top q principal components of $\mathbf{R}$ are used, the dimension of $\mathbf{w}$ is $q \times 1$ and $\mathbf{Q}$ is $q \times m$. By selecting $q \ll n$, this model would have a substantially lower rank than Eq (1), improving the computational efficiency for the estimation of $\boldsymbol{\beta}$. By default, our method automatically determines q through pseudo cross-validation based on GWAS summary statistics (Supplementary Note 10 of Zheng et al. [24]). With a recognition that $\mathbf{b} = \mathbf{X}'\mathbf{y}/n$ is the GWAS marginal effect estimates, $\mathbf{w}$ can be directly computed from the GWAS summary statistics

$$\mathbf{w} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\mathbf{b} \tag{5}$$

with $\boldsymbol{\Lambda}$ and $\mathbf{U}$ obtained from a LD reference sample, and $\mathbf{Q}$ can be computed as

$$\mathbf{Q} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\mathbf{R} = \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}' \tag{6}$$

Essentially, the low-rank model transforms the mutually correlated GWAS marginal effects ($\mathbf{b}$) into a set of independent data points ($\mathbf{w}$), while maintaining the ability to estimate individual SNP joint effects ($\boldsymbol{\beta}$). In practice, we compute $\mathbf{w}$ and $\mathbf{Q}$ within quasi-independent LD blocks in the human genome. With this low-rank model, we can estimate $\boldsymbol{\beta}$ for all common variants

522 jointly through considering $\boldsymbol{\beta}$ as random effects. In addition, this model allows a direction
523 estimation of the residual variance, as $\mathrm{var}(\boldsymbol{\epsilon}) = \mathbf{I}\sigma_e^2 n^{-1}$, which can be used as a nuisance
524 parameter to accommodate heterogeneity in the summary statistics and LD reference[24].
525
526 **SBayesC, SBayesR, and SBayesRC**
527 GBMMs are flexible in specifying the prior distribution of SNP effects. In SBayesC, the prior for
528 the effect of SNP j is given by

529 $$\beta_j \sim \pi N\left(0,\ \sigma_\beta^2\right) + (1-\pi)\phi \tag{7}$$

530 where σ_β^2 is the common variance across all causal variants, π is the proportion of SNPs with
531 nonzero effects, and ϕ is a point mass at zero. Both σ_β^2 and π are considered as unknown, with
532 σ_β^2 following a scaled inverse chi-squared prior distribution and π following a uniform prior
533 distribution[24]. This is also the underlying model for most fine-mapping methods, assuming that
534 only a fraction of SNPs are causal with nonzero effects.
535
536 SBayesR[22] extends SBayesC by assuming a more flexible prior distribution for SNP effects, using
537 a multi-component Gaussian mixture:

538 $$\beta_j \sim \sum_{k=1}^{5} \pi_k N\left(0,\ \gamma_k \sigma_g^2\right) \tag{8}$$

539 where σ_g^2 is the total genetic variance, $\boldsymbol{\gamma} = [0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}]'$ are pre-specified scale
540 factors that constrain the variance of each effect size distribution, representing zero, very small,
541 small, medium, and large effect size categories, and π_k is the prior probability of SNP j belonging
542 to the kth distribution. With different π_k values, this multi-component mixture model can
543 approximate almost any distribution of SNP effects. This allows SBayesR to better model genetic
544 architectures with a wide range of effect sizes compared to SBayesC.
545
546 To further account for the stratification of causal variants and their effects based on functional
547 annotations, SBayesRC[24] assumes a SNP-specific prior probability of distribution membership,
548 π_{jk}, which depends on the annotations at each SNP through a generalised linear model:

549 $$f\left(\pi_{jk}\right) = \mu_k + \sum_{l=1}^{c} A_{jl}\, \alpha_{kl} \tag{9}$$

550 where $f(\cdot)$ is the probit link function, μ_k is the intercept, A_{jl} is the value of annotation l on SNP j
551 (either binary or continuous annotations), and α_{kl} is the effect of annotation l on the prior
552 probability of SNP j belonging to the kth distribution. To enable a data-driven estimation of
553 annotation effects, we assume an independent and identical normal prior for each annotation
554 effect, $\alpha_{kl} \sim N(0, \sigma_{\alpha_k}^2)$. This prior assumes, *a priori,* an equal contribution of all functional
555 annotations to the SNP prior probability of membership. Details of the method and the MCMC
556 sampling scheme can be found in ref[24].

In this study, we ran SBayesRC with 4 MCMC chains, each consisting of 3,000 iterations with the first 1,000 discarded as burn-in. Analysing 13 million SNPs with 96 annotations, SBayesRC required only 150 GB of RAM and 13 hours of computation using 24 CPU cores, which are commonly available in a standard computing cluster.

LD blocks

We partitioned the genome into 1,588 approximately independent LD blocks, following the methods of Li et al[13] and Berisa and Pickrell[58]. Specifically, we merged the LD blocks from Li et al. to ensure a minimum length of 1 Mb and treated the MHC region as a single LD block due to its complex LD patterns, which had the largest number of SNPs. As a result, our LD blocks ranged in size from 1 to 30 Mb, with a median length of 1.5 Mb.

Calculation of PIP

We assessed the strength of joint association of each SNP using PIP, i.e., the probability of a SNP being included with a nonzero effect in the model, given the data ($\mathbf{w}$). Let δ_j be the indicator variable for the distribution membership for SNP j, with $\delta_j = 1$ indicating a null effect and $\delta_j = 2, \ldots, 5$ indicating a nonzero component. In SBayesRC, we computed PIP for SNP j as

$$\mathrm{PIP}_j = \sum_{k=2}^{5} \Pr\left(\delta_j = k \middle| \mathbf{w}\right) = 1 - \Pr\left(\delta_j = 1 \middle| \mathbf{w}\right) \tag{10}$$

In the literature, $\Pr\left(\delta_j = 1 \middle| Data\right)$, the probability of SNP j having a zero effect, is often computed by counting the frequency of $\delta_j = 1$ across MCMC samples[23,25]. To improve precision, we use Rao-Blackwellized estimates[59,60] and compute the posterior mean of $[1 - \Pr\left(\delta_j = 1 \middle| \mathbf{w}, \boldsymbol{\theta}\right)]$ conditional on data and all the other parameters except β_j ($\boldsymbol{\theta}$) across T iterations. That is, suppose $p_j^{(t)} = 1 - \Pr\left(\delta_j = 1 \middle| \mathbf{w}, \boldsymbol{\theta}^{(t)}\right)$, we have

$$\mathrm{PIP}_j = E\left[p_j^{(t)}\right] = \frac{1}{T}\sum_{t=1}^{T}\left[1 - \Pr\left(\delta_j = 1 \middle| \mathbf{w}, \boldsymbol{\theta}^{(t)}\right)\right] \tag{11}$$

where

$$\Pr\left(\delta_j = 1 \middle| \mathbf{w}, \boldsymbol{\theta}^{(t)}\right) = \frac{f(\mathbf{w}|\delta_j = 1, \boldsymbol{\theta}^{(t)})\pi_{j1}}{\sum_{k=1}^{5} f(\mathbf{w}|\delta_j = k, \boldsymbol{\theta}^{(t)})\pi_{jk}} \tag{12}$$

with $f(\mathbf{w}|\delta_j = k, \boldsymbol{\theta}^{(t)})$ being the likelihood given δ_j and the sampled values of other parameters. A detailed derivation can be found in **Supplementary Note 5**.

To further improve robustness, we run multiple MCMC chains simultaneously, randomly shuffling the sampling order of SNP effects in every iteration. For SNP j in chain s at iteration t, we have

$$p_j^{(s,t)} = 1 - \Pr\left(\delta_j = 1 \middle| \mathbf{w}, \boldsymbol{\theta}^{(s,t)}\right) \tag{13}$$

Following Gelman and Rubin (1992)[61], we calculated the potential scale reduction factor (PSRF) for each SNP PIP:

$$\hat{R}_j = \sqrt{\frac{\frac{L-1}{L}W_j + \frac{1}{L}B_j}{W_j}} \tag{14}$$

where $W_j = E_s[Var_t(p_j^{(s,t)})]$ is the averaged within-chain variance across chains, and $B_j = Var_s[E_t(p_j^{(s,t)})]$ is the variance of the means of the chains. By comparing the between-chain variance to the within-chain variance, PSRF assesses whether the MCMC chains have converged to a stationary distribution. Empirically, a PSRF value below 1.2 indicates adequate convergence[61]. By pooling samples across chains, our final PIP estimate is

$$\mathrm{PIP}_j = E_{s,t}[p_j^{(s,t)}] \tag{15}$$

We report the PSRF value to imply if the SNP PIP is converged or not. SNPs with high PIP but large PSRF should be interpreted with caution. Running longer chains is recommended if a remarkable proportion of fine-mapped variants have high PSRG values. Across 48 well-powered complex traits analyzed, we found that none of fine-mapped individual SNPs (singleton LCSs) had PSRF < 1.2, while 99.4% of SNPs included in the local credible sets had PSRF < 1.2 and 99.9% had PSRF < 2, suggesting adequate convergence in our analyses.

Heuristic estimation of mixture components

The standard SBayesRC requires pre-specifying the number of mixture components before analysis. While polygenic prediction is generally robust to this choice[24], fine-mapping may be affected if an unnecessary small-effect component is included, leading to null SNPs being misclassified and inflating PIP. This could bias the estimated number of causal variants. To address this, we implemented a heuristic procedure for determining the number of mixture components in SBayesRC. Multiple models, varying from 2 to 5 mixture components, were run in parallel for 500 MCMC iterations. The posterior distribution of h^2_{SNP} was then computed to assess model fit. The model with the highest posterior mean, where the lower bound (posterior mean minus standard error) was greater than that of the second-best model, was selected for formal analysis. If no significant difference was found, the model with the fewest components was chosen.

617 **Tempered Gibbs sampling**

618 Joint analysis of all common variants presents a challenge for MCMC mixing when using the
619 standard Gibbs sampling (GS) algorithm. For example, in a finite number of iterations, a
620 common causal variant may fail to enter the model if the SNP in complete LD with it has already
621 been selected. To address this issue, we incorporated a tempered Gibbs sampling (TGS)
622 algorithm[32] into our SBayesRC model. Essentially, TGS improves mixing by (1) strategically
623 selecting SNPs for updating ("informed choice") – focusing on those whose indicator variable δ_j
624 is currently in the tail of the full conditional distribution, and (2) allowing for longer jumps
625 across local maxima by sampling δ_j from a tempered distribution. For computational efficiency,
626 we applied TGS only to SNPs with nonzero effects sampled from the standard GS to assess
627 whether alternative SNPs in high LD (r^2>0.95) could provide a better fit. Further details on our
628 TGS implementation and the corresponding pseudocode are provided in **Supplementary Note**
629 **6**.

630

631 **Local credible sets**

632 Following prior work[9], we defined LCS at confidence level α (α-LCS) as the minimum set of SNPs
633 that contains at least one causal variant with probability α. To construct an α-LCS, SNPs were
634 ranked by their PIPs. For a focal SNP that has the highest PIP and not yet assigned to an LCS, a
635 candidate set was formed by including all remaining SNPs in high LD (r^2 > 0.5) with the focal
636 SNP. The α-LCS was then determined by summing PIPs in decreasing order until the total
637 reached at least α. This process was iterated over SNPs until no possible LCS can be formed. For
638 each α-LCS, we computed a posterior SNP-heritability enrichment probability (PEP) that the
639 LCS explains more h^2_{SNP} than a random set of SNPs with the same size. We reported all 0.9-LCSs
640 with PEP > 0.7 for each LD block. A schematic of the procedure is shown in **Supplementary Fig.**
641 **1**, differences to existing approaches are discussed in **Supplementary Note 1**, and further
642 justification of these thresholds based on simulations is provided in **Supplementary Note 2**.

643

644 **Global credible sets**

645 The α-GCS identifies the smallest set of SNPs that capture α% of the causal variants, quantifying
646 the overall uncertainty in fine-mapping and reflecting the genetic architecture of the trait. The
647 size of α-GCS decreases with increasing power and a smaller fraction of small-effect variants.
648 We computed the α-GCS as the cumulative sum of decreasingly ranked PIPs until the total
649 exceeded $\alpha \times m_c$, where m_c was the estimated number of causal variants from GBMM for the
650 trait. Since each SNP is causal with probability PIP_j, the expected number of causal variants in a

set Ω is E[Number of causal variants in Ω] $= \sum_{j\in\Omega} \mathrm{PIP}_j$. Since the α-GCS is constructed such that $\sum_{j\in GCS} \mathrm{PIP}_j \geq \alpha \times m_c$, this directly implies that E[Number of causal variants in α-GCS] $\geq \alpha \times m_c$.

Estimation of power and variance explained given the data

For the identified SNPs using individual PIP or LCS, we estimated the true positive rate (TPR or power) of identifying the causal variants at a given threshold α,

$$\mathrm{TPR}_\alpha = \frac{\sum_j [\mathrm{PIP}_j | \ \mathrm{PIP}_j \geq \alpha]}{M\pi} \tag{16}$$

where M is the total number of SNPs, and π is the proportion of causal SNPs. A formal derivation is given in **Supplementary Note 7**.

The proportion of SNP-based heritability explained (PHE) by LCSs or GCS is estimated based on the MCMC samples of SNP effects. For a focal set (Ω) of SNPs in iteration t, we computed PHE as

$$\mathrm{PHE}_\Omega^{(t)} = \frac{\sum_{j\in\Omega} \beta_{jt}^2}{\sum_{m=1}^{M} \beta_{mt}^2} \tag{17}$$

where $\beta_{\cdot t}$ is the sampled effect for a SNP in iteration t in scaled genotype units. Finally, we computed the posterior mean across MCMC iterations as the estimate for PHE_Ω

$$\mathrm{PHE}_\Omega = \frac{1}{L} \sum_{t=1}^{L} \mathrm{PHE}_\Omega^{(t)} \tag{18}$$

where L is the total number of MCMC iterations.

Prediction of power and variance explained for prospective studies

We aim to predict the power of identifying a certain proportion of the causal variants in a prospective fine-mapping analysis, given a sample size (n) and the genetic architecture of the trait, when PIP from a GBMM is used as the test statistic. As shown in **Supplementary Note 8**, assuming that variance explained by the causal variant is v, the sampling distribution of PIP from the multi-component mixture model, e.g., SBayesRC, is

$$PIP = 1 - \frac{1}{1+\sum_{k=2}^{5} A_k \exp\{B_k Z\}} \tag{19}$$

where $A_k = \frac{\pi_k}{\pi_1} \lambda_k^{\frac{1}{2}} C_k^{-\frac{1}{2}}$ and $B_k = \frac{n\sigma_e^2}{2C_k}$ are two constants given the genetic architecture parameters ($\boldsymbol{\pi}$ and h_{SNP}^2), with $\lambda_k = \frac{\sigma_e^2}{\gamma_k \sigma_g^2}$ and $C_k = n + \lambda_k$, and Z is a data-dependent random variable following a non-central Chi-square distribution with the non-centrality parameter $NCP = \frac{nv}{\sigma_e^2}$:

$$Z \sim \chi_1^2\left(\frac{nv}{\sigma_e^2}\right) \tag{20}$$

Given the threshold of PIP being α for the hypothesis test, the power to detect this causal variant can be calculated as

$$Power_v = \Pr(PIP > \alpha | v) = \int_\alpha^1 f(PIP|v) \mathrm{d}PIP \tag{21}$$

where $f(PIP|v)$ is Eq (19) above. To compute the power for identifying any causal variant, we integrated out v by

$$Power = \int_{\alpha}^{1}\int_{0}^{\infty} f(PIP|v)f(v)\mathrm{d}v\mathrm{d}PIP \tag{22}$$

where

$$f(v) = f_{\beta}(v^{\frac{1}{2}})v^{-\frac{1}{2}} \tag{23}$$

and $f_{\beta}(\cdot)$ is the distribution of β estimated from the SBayesRC model.

Therefore, given a sample size, the expected number of causal variants identified from fine-mapping is

$$E[NCV] = m(1-\pi_1) \times Power \tag{24}$$

The expected proportion of genetic variance explained by the fine-mapped variants is

$$E[PHE] = m(1-\pi_1)\sigma_g^{-2}\int_{0}^{\infty} Power_v \times vf(v)\mathrm{d}v \tag{25}$$

Since it is computationally challenging to obtain an analytical solution, we opted to estimate these quantities through Monte Carlo simulation (**Supplementary Note 8**). While our theoretical prediction does not model LD between SNPs, the extent to which the observed values were consistent with the predicted suggests that LD had been effectively, albeit not perfectly, accounted for by our LCSs.

Disease sample size at the liability scale

For diseases or binary traits, we converted the GWAS summary statistics from the linear mixed model to the liability scale prior to running GBMM. Based on the method in Yang et al.[43], we estimated the sample size at the liability scale that gives an equivalent power to detect a locus affecting a quantitative trait with the same properties,

$$N_{eq} = \frac{i^2 v(1-v)N_{01}}{(1-K)^2} \tag{26}$$

where K is the disease prevalence, v is the sample prevalence, $i = h/K$ with h being the height of standard normal curve at the truncation point $t = 1-K$, and N_{01} is the total number of cases and controls. Given the z-score (z_j) from the original GWAS summary data for SNP j, the marginal effect and its standard error at the liability scale can be estimated as following ref[62]

$$SE_j = \sqrt{\frac{1}{2p_j(1-p_j)(N_{eq}+z_j^2)}} \tag{27}$$

$$b_j = z_j \times SE_j \tag{28}$$

where p_j is the minor allele frequency of the SNP.

The results from GBMM using the converted summary statistics will be directly comparable across traits regardless of the sample prevalence and the type of traits. In our prediction

717 analysis of power, we compared results between diseases and quantitative traits based on the
718 equivalent sample size estimated from Eq (26). Similarly, to estimate the number of cases
719 required, in a case-control study with equivalent number of controls, to achieve a certain power,
720 we rearranged the same equation so that

721 $$N_{cases} = \frac{1}{2}\frac{(1-K)^2 N_{eq}}{i^2 \nu(1-\nu)} \quad (29)$$

722

723 **Simulations based on imputed genotype data from the UK Biobank**

724 To evaluate the performance of GBMM, we ran simulations using imputed genotype data from
725 the UK Biobank after quality control (QC). We selected 300,000 unrelated individuals and
726 included ~1.2 million HapMap3 SNPs with MAF > 0.01, Hardy-Weinberg equilibrium test P >
727 1×10^{-6}, genotyping rate > 0.95, and imputation information score > 0.8 for simulations.

728

729 We randomly sampled $m_c = 10{,}000$ casual variants from the genome for 100,000 individuals
730 and simulated complex trait phenotypes based on the following model:

731 $$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e} \quad (30)$$

732 where $\mathbf{X}$ is the genotype matrix for the causal variants, $\boldsymbol{\beta}$ is the vector of causal variant effects,
733 and $\boldsymbol{e} \sim N(0, \mathbf{I}var(\mathbf{X}\boldsymbol{\beta})/(1/h^2 - 1))$ with $h^2 = 0.5$ being the proportion of phenotypic variance
734 explained by all the causal variants. Based on this underlying model, we simulated three genetic
735 architectures: sparse architecture, large-effects architecture, and LD-MAF stratified (LDMS)
736 architecture. In simulations under the sparse genetic architecture, causal variant effects were
737 drawn from $\beta_i \sim N(0, h^2/m_c)$, while the remaining SNPs had zero effect. For the large-effects
738 genetic architecture, effect sizes for causal variants were drawn from two distributions, 10
739 random causal variant with effects from $N(0, 0.2\, h^2/10)$ and the remaining causal variants with
740 effects from $N(0, 0.8\, h^2/9990)$. For the LDMS model, we partitioned all genome-wide SNPs into
741 four LD and MAF groups (by their median values) and only sampled the causal variants from the
742 high LD and high MAF group with the same normal distribution as in the sparse genetic
743 architecture.

744

745 We ran a standard GWAS using the genotypes and the simulated phenotypes under different
746 genetic architectures. We then used the GWAS summary data to perform GBMM (SBayesRC[24],
747 SBayesR[22], and SBayesC) implemented in GCTB, along with SuSiE[6], FINEMAP[7], SuSiE-inf[8],
748 FINEMAP-inf[8] and PolyFun+SuSiE[10], to compute PIPs for detecting causal variants and estimate
749 their effect sizes. For all methods, we followed the recommended default parameter settings.
750 Specifically, for SuSiE and SuSiE-inf, we set the purity parameter to 0.5, the maximum number of
751 SNPs per credible set (n_purity) to 100, and max number of causal variants to 10. We used

752 imputed genotypes from 10,000 random samples of European ancestry from the UKB as the LD
753 reference. To ensure a fair comparison, we used the same LD reference sample and the same
754 independent LD blocks to define fine-mapping regions for all fine-mapping methods,
755 irrespective of p-value significance. We repeated the whole process 100 times and quantified
756 the true discovery rate (TDR), power, mapping precision, and replication rate for each method.
757 TDR was quantified as the proportion of identified α-LCS containing at least one causal variant,
758 while power was defined as the proportion of simulated causal variants included in the
759 identified α-LCS. Mapping precision was computed as the physical distance between the
760 identified SNPs and nearest causal variants. The replication rate was the proportion of variants
761 with PIP > 0.9 in a GWAS sample that were identified in an independent replication sample at
762 the same or a lower PIP threshold.
763
764 **Real data analysis**
765 We analysed 597 complex traits from UK Biobank using GWAS summary data from Neale's lab
766 (**Data Availability**) and schizophrenia[39] and Crohn's disease[41] using previously published
767 GWAS summary data. The 597 UKB traits were selected based on z-score > 4 and high-
768 confidence heritability estimates using LD score regression[57]. We used annotations from the
769 baseline model BaseLineLD v2.2[57] and extract imputed SNPs with MAF > 0.001 and that
770 overlapped with the annotations, resulting in 13,065,104 imputed SNPs passed quality control.
771 We used 10,000 random samples from the UKB as the LD reference to run the SBayesRC and
772 other region-specific fine-mapping analysis. Additionally, we selected 48 well-powered traits
773 with relatively large sample size (n > 100, 000), high heritability estimate (h^2 > 0.1), and at least
774 one fine-mapped SNP with PIP > 0.9.
775
776 For the polygenic score prediction analysis using fine-mapped variants only, we performed
777 quality control on the imputed genotype data provided by the UKB analysis team[30]. Following
778 ref[24], we kept SNPs with MAF > 0.01, Hardy-Weinberg Equilibrium test P > 10^{-10}, imputation info
779 score > 0.6 within each ancestry samples. We removed samples with mismatched sex
780 information, samples withdrawn from participation, and cryptic related samples. The final UKB
781 dataset consisted of 4 ancestries: European (EUR, N= 347,800), East Asian (EAS, N=2,252),
782 South Asian (SAS, N=9,436) and African (AFR, N=7,006). For continuous traits, the phenotypes
783 were filtered within the range of mean +/- 7SD, followed by rank-based inverse-normal
784 transformed within each ancestry and sex group. GWAS were performed using PLINK2[63] with
785 sex, age, and the first 10 principal component as covariates. Linear regression was used for
786 continuous traits, while logistic regression was applied for binary traits.
787

Supplementary Information

Supplementary data include 29 supplementary figures, 4 supplementary tables and 8 supplementary notes.

Data Availability

Our SBayesRC-enabled genome-wide fine-mapping results for 599 complex traits are available at link (https://sbayes.pctgplots.cloud.edu.au/data/SBayesRC/share/Finemap/v1.1/). The UK Biobank data are available through formal application to the UK Biobank (http://www.ukbiobank.ac.uk). The GWAS summary data for 597 complex traits in UK Biobank are from http://www.nealelab.is/uk-biobank/. The LD data used in this study are available at https://cnsgenomics.com/software/gctb/#Download. All the other datasets used in this study are available in the public domain.

Code Availability

Summary-data-based genome-wide Bayesian mixture models are implemented in a public available software GCTB at https://cnsgenomics.com/software/gctb/#Download. Methods to compute LCS and GCS have also been implemented in GCTB (https://cnsgenomics.com/software/gctb/#Genome-wideFine-mappinganalysis). Online tool for predicting fine-mapping power: https://sbayes.pctgplots.cloud.edu.au/shiny/power/.

Acknowledgements

This research was supported by the Fundamental Research Funds for the Central Universities, the 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYYC24006, ZYJC20002), Australian National Health and Medical Research Council (1177268, 1113400,1173790), the Australian Research Council (FL180100072, DP220101947, DP230101352) and the National Institute of Mental Health (5R01MH121545-05). This study makes use of data from the UK Biobank (project ID: 12505).

Author Contributions

J.Z. conceived and supervised the study. J.Z. and Y.W. developed the methods. J.Z., P.M.V. and Y.W. designed the experiment. Y.W. conducted all analyses with the assistance or guidance from J.Z., Z.Z., L.T., T.L., Q.F., H.C., L.Y., P.M.V., N.R.W. and M.E.G.. Y.W., and J.Z. wrote the manuscript with the participation of all authors. All the authors approved the final version of the manuscript.

Competing Interests

The authors declare no competing interests.

824

825 **References**

826 1 Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery:
827 Realizing the promise. *Am J Hum Genet* **110**, 179-194, doi:10.1016/j.ajhg.2022.12.011
828 (2023).

829 2 Uffelmann, E. *et al.* Genome-wide association studies. *Nature Reviews Methods Primers* **1**,
830 59, doi:10.1038/s43586-021-00056-9 (2021).

831 3 Zeng, J. *et al.* Widespread signatures of natural selection across human complex traits
832 and functional genomic categories. *Nature Communications* **12**, 1164,
833 doi:10.1038/s41467-021-21446-3 (2021).

834 4 O'Connor, L. J. *et al.* Extreme Polygenicity of Complex Traits Is Explained by Negative
835 Selection. *The American Journal of Human Genetics* **105**, 456-476,
836 doi:https://doi.org/10.1016/j.ajhg.2019.07.003 (2019).

837 5 Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate
838 causal variants by statistical fine-mapping. *Nature Reviews Genetics* **19**, 491-504,
839 doi:10.1038/s41576-018-0016-z (2018).

840 6 Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A Simple New Approach to Variable
841 Selection in Regression, with Application to Genetic Fine Mapping. *Journal of the Royal*
842 *Statistical Society Series B: Statistical Methodology* **82**, 1273-1300,
843 doi:10.1111/rssb.12388 (2020).

844 7 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-
845 wide association studies. *Bioinformatics* **32**, 1493-1501,
846 doi:10.1093/bioinformatics/btw018 (2016).

847 8 Cui, R. *et al.* Improving fine-mapping by modeling infinitesimal effects. *Nature Genetics*,
848 doi:10.1038/s41588-023-01597-3 (2023).

849 9 Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with
850 the "Sum of Single Effects" model. *PLOS Genetics* **18**, e1010299,
851 doi:10.1371/journal.pgen.1010299 (2022).

852 10 Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of
853 complex trait heritability. *Nature Genetics* **52**, 1355-1363, doi:10.1038/s41588-020-
854 00735-5 (2020).

855 11 Kichaev, G. *et al.* Integrating Functional Data to Prioritize Causal Variants in Statistical
856 Fine-Mapping Studies. *PLOS Genetics* **10**, e1004722, doi:10.1371/journal.pgen.1004722
857 (2014).

858 12 Yang, Z. *et al.* CARMA is a new Bayesian model for fine-mapping in genome-wide
859 association meta-analyses. *Nature Genetics* **55**, 1057-1065, doi:10.1038/s41588-023-
860 01392-0 (2023).
861 13 Li, X., Sham, P. C. & Zhang, Y. D. A Bayesian fine-mapping model using a continuous
862 global-local shrinkage prior with applications in prostate cancer analysis. *The American*
863 *Journal of Human Genetics* **111**, 213-226,
864 doi:https://doi.org/10.1016/j.ajhg.2023.12.007 (2024).
865 14 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human
866 height. *Nature Genetics* **42**, 565-569, doi:10.1038/ng.608 (2010).
867 15 So, H.-C., Gui, A. H. S., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by
868 known susceptibility variants: a survey of ten complex diseases. *Genetic Epidemiology*
869 **35**, 310-317, doi:https://doi.org/10.1002/gepi.20579 (2011).
870 16 Chen, W. *et al.* Fine Mapping Causal Variants with an Approximate Bayesian Method
871 Using Marginal Test Statistics. *Genetics* **200**, 719-736, doi:10.1534/genetics.115.176107
872 (2015).
873 17 Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-
874 Mapping Studies. *The American Journal of Human Genetics* **97**, 260-271,
875 doi:https://doi.org/10.1016/j.ajhg.2015.06.007 (2015).
876 18 Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size
877 distributions using summary-level statistics from genome-wide association studies
878 across 32 complex traits. *Nature Genetics* **50**, 1318-1326, doi:10.1038/s41588-018-
879 0193-x (2018).
880 19 Erbe, M. *et al.* Improving accuracy of genomic predictions within and between dairy
881 cattle breeds with imputed high-density single nucleotide polymorphism panels. *J Dairy*
882 *Sci* **95**, 4114-4129, doi:10.3168/jds.2011-5019 (2012).
883 20 MacLeod, I. M. *et al.* Exploiting biological priors and sequence variants enhances QTL
884 discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144,
885 doi:10.1186/s12864-016-2443-6 (2016).
886 21 Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian
887 alphabet for genomic selection. *BMC Bioinformatics* **12**, 186, doi:10.1186/1471-2105-
888 12-186 (2011).
889 22 Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression
890 on summary statistics. *Nature Communications* **10**, 5086, doi:10.1038/s41467-019-
891 12653-0 (2019).

892 23 Moser, G. *et al.* Simultaneous Discovery, Estimation and Prediction Analysis of Complex
893 Traits Using a Bayesian Mixture Model. *PLOS Genetics* **11**, e1004969,
894 doi:10.1371/journal.pgen.1004969 (2015).
895 24 Zheng, Z. *et al.* Leveraging functional genomic annotations and genome coverage to
896 improve polygenic prediction of complex traits within and between ancestries. *Nature*
897 *Genetics*, doi:10.1038/s41588-024-01704-y (2024).
898 25 Xiang, R. *et al.* Genome-wide fine-mapping identifies pleiotropic and functional variants
899 that predict many traits across global cattle populations. *Nature Communications* **12**,
900 860, doi:10.1038/s41467-021-21001-0 (2021).
901 26 Shrestha, M. *et al.* Evaluation of Bayesian Linear Regression Models as a Fine Mapping
902 tool. *bioRxiv*, 2023.2009.2001.555889, doi:10.1101/2023.09.01.555889 (2024).
903 27 Carbonetto, P. & Stephens, M. Scalable Variational Inference for Bayesian Variable
904 Selection in Regression, and Its Accuracy in Genetic Association Studies. *Bayesian*
905 *Analysis* **7**, 73-108, 136 (2012).
906 28 Blei, D. M., Alp, K. & and McAuliffe, J. D. Variational Inference: A Review for Statisticians.
907 *Journal of the American Statistical Association* **112**, 859-877,
908 doi:10.1080/01621459.2017.1285773 (2017).
909 29 Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human
910 complex traits. *Nature Genetics* **50**, 746-753, doi:10.1038/s41588-018-0101-4 (2018).
911 30 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
912 *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
913 31 Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in
914 human populations. *Bioinformatics* **32**, 283-285, doi:10.1093/bioinformatics/btv546
915 (2016).
916 32 Zanella, G. & Roberts, G. Scalable Importance Tempering and Bayesian Variable Selection.
917 *Journal of the Royal Statistical Society Series B: Statistical Methodology* **81**, 489-517,
918 doi:10.1111/rssb.12316 (2019).
919 33 Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences.
920 *Statistical Science* **7**, 457-472, 416 (1992).
921 34 Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-
922 wide association studies using whole-genome sequencing data. *Genome Biology* **18**, 86,
923 doi:10.1186/s13059-017-1216-0 (2017).
924 35 Hou, K. *et al.* Causal effects on complex traits are similar for common variants across
925 segments of different continental ancestries within admixed individuals. *Nature Genetics*
926 **55**, 549-558, doi:10.1038/s41588-023-01338-6 (2023).

927 36 Yengo, L. *et al.* A saturated map of common genetic variants associated with human
928 height. *Nature* **610**, 704-712, doi:10.1038/s41586-022-05275-y (2022).
929 37 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body
930 mass index in ⍰700000 individuals of European ancestry. *Human Molecular Genetics* **27**,
931 3641-3649, doi:10.1093/hmg/ddy271 (2018).
932 38 Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of
933 lipids. *Nature* **600**, 675-679, doi:10.1038/s41586-021-04064-3 (2021).
934 39 Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
935 schizophrenia. *Nature* **604**, 502-508, doi:10.1038/s41586-022-04434-5 (2022).
936 40 Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant
937 genes and in regions under strong background selection. *Nature Genetics* **50**, 381-389,
938 doi:10.1038/s41588-018-0059-2 (2018).
939 41 de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of
940 multiple integrin genes in inflammatory bowel disease. *Nature Genetics* **49**, 256-261,
941 doi:10.1038/ng.3760 (2017).
942 42 Lee, Sang H., Wray, Naomi R., Goddard, Michael E. & Visscher, Peter M. Estimating
943 Missing Heritability for Disease from Genome-wide Association Studies. *The American*
944 *Journal of Human Genetics* **88**, 294-305, doi:10.1016/j.ajhg.2011.02.002 (2011).
945 43 Yang, J., Wray, N. R. & Visscher, P. M. Comparing apples and oranges: equating the power
946 of case-control and quantitative trait association studies. *Genet Epidemiol* **34**, 254-257,
947 doi:10.1002/gepi.20456 (2010).
948 44 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
949 polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295,
950 doi:10.1038/ng.3211 (2015).
951 45 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
952 wide association summary statistics. *Nature Genetics* **47**, 1228-1235,
953 doi:10.1038/ng.3404 (2015).
954 46 Novo, I., López-Cortegano, E. & Caballero, A. Highly pleiotropic variants of human traits
955 are enriched in genomic regions with strong background selection. *Human Genetics* **140**,
956 1343-1351, doi:10.1007/s00439-021-02308-w (2021).
957 47 Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans.
958 *New England Journal of Medicine* **373**, 895-907, doi:10.1056/NEJMoa1502214 (2015).
959 48 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics
960 identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369-375,
961 doi:10.1038/ng.2213 (2012).

962 49 Li, A. *et al.* mBAT-combo: A more powerful test to detect gene-trait associations from
963 GWAS data. *The American Journal of Human Genetics* **110**, 30-43,
964 doi:https://doi.org/10.1016/j.ajhg.2022.12.006 (2023).
965 50 Li, S. *et al.* The schizophrenia-associated missense variant rs13107325 regulates
966 dendritic spine density. *Transl Psychiatry* **12**, 361, doi:10.1038/s41398-022-02137-z
967 (2022).
968 51 Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia.
969 *Nature* **604**, 509-516, doi:10.1038/s41586-022-04556-w (2022).
970 52 Dai, Z.-M. *et al.* SECISBP2L-Mediated Selenoprotein Synthesis Is Essential for
971 Autonomous Regulation of Oligodendrocyte Differentiation. *The Journal of Neuroscience*
972 **42**, 5860-5869, doi:10.1523/jneurosci.2141-21.2022 (2022).
973 53 Sazonovs, A. *et al.* Large-scale sequencing identifies multiple genes and rare variants
974 associated with Crohn's disease susceptibility. *Nature Genetics* **54**, 1275-1283,
975 doi:10.1038/s41588-022-01156-2 (2022).
976 54 Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275-294,
977 doi:10.1093/biostatistics/kxw041 (2016).
978 55 Zhang, M. J. *et al.* Pervasive correlations between causal disease effects of proximal SNPs
979 vary with functional annotations and implicate stabilizing selection. *medRxiv*,
980 2023.2012.2004.23299391, doi:10.1101/2023.12.04.23299391 (2023).
981 56 Samaddar, A., Maiti, T. & de los Campos, G. Bayesian Hierarchical Hypothesis Testing in
982 Large-Scale Genome-Wide Association Analysis. *bioRxiv*, 2024.2002.2026.582204,
983 doi:10.1101/2024.02.26.582204 (2024).
984 57 Gazal, S. *et al.* Functional architecture of low-frequency variants highlights strength of
985 negative selection across coding and non-coding annotations. *Nature Genetics* **50**, 1600-
986 1607, doi:10.1038/s41588-018-0231-8 (2018).
987 58 Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in
988 human populations. *Bioinformatics* **32**, 283-285, doi:10.1093/bioinformatics/btv546
989 (2015).
990 59 Casella, G. & Robert, C. P. Rao-Blackwellisation of Sampling Schemes. *Biometrika* **83**, 81-
991 94 (1996).
992 60 Zhu, X. & Stephens, M. Bayesian large-scale multiple regression with summary statistics
993 from genome-wide association studies. *Ann Appl Stat* **11**, 1561-1592, doi:10.1214/17-
994 aoas1046 (2017).
995 61 Andrew, G. & Donald, B. R. Inference from Iterative Simulation Using Multiple Sequences.
996 *Statistical Science* **7**, 457-472, doi:10.1214/ss/1177011136 (1992).

997 62 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts
998 complex trait gene targets. *Nature Genetics* **48**, 481-487, doi:10.1038/ng.3538 (2016).
999 63 Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
1000 Linkage Analyses. *The American Journal of Human Genetics* **81**, 559-575,
1001 doi:https://doi.org/10.1086/519795 (2007).

1002

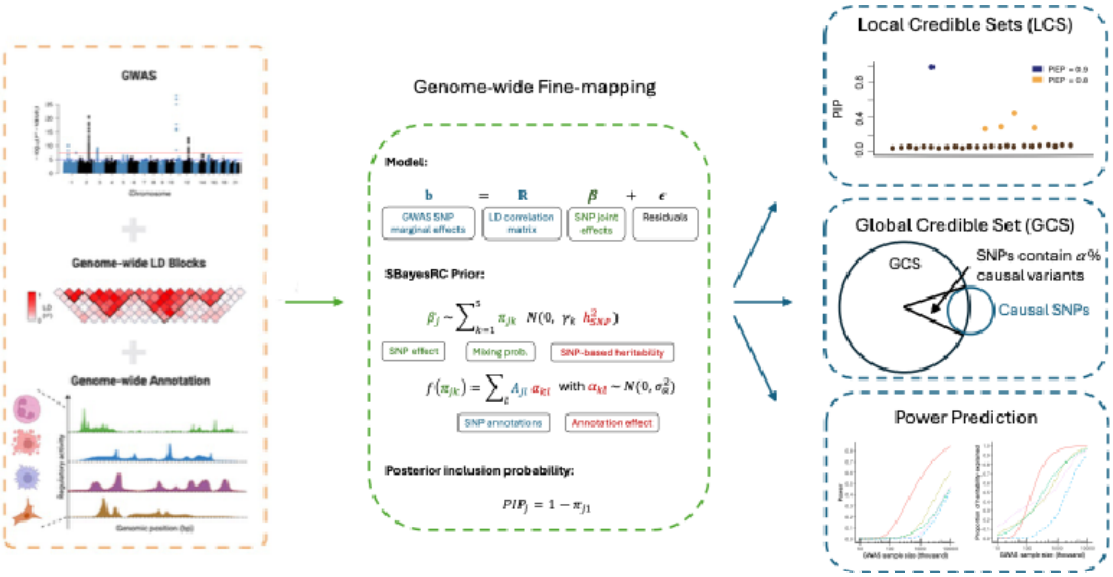

Figure 1 Schematic overview of genome-wide fine-mapping (GWFM) analysis using GBMM. GBMM utilizes GWAS summary statistics and genome-wide LD reference to fine-map candidate causal variants for complex traits, incorporating functional annotations when available. Unlike region-specific fine-mapping approaches, GBMM maps causal signals across the entire genome and provides greater flexibility in modelling the distribution of causal effects by estimating priors using genome-wide SNPs (**Table S1**). Different GBMMs share the same fitted model but differ in prior specification for SNP effects. As shown in the middle box, SBayesRC assumes a mixture prior consisting of a point mass at zero and multiple normal distributions with variance $\gamma_k h^2_{SNP}$, where $\gamma_k = [0, 1^{-5}, 1^{-4}, 1^{-3}, 1^{-2}]'$. The mixing probability for SNP j in distribution k (π_{jk}) is modelled as a linear combination of SNP annotations through a probit link. The color scheme is as follows: blue denotes input data, green denotes SNP-specific parameters, and red denotes global parameters shared across SNPs. The GBMM output includes SNP PIPs, local and global credible sets, and prediction of fine-mapping power for future studies. The illustration in the left box was created with BioRender.com.

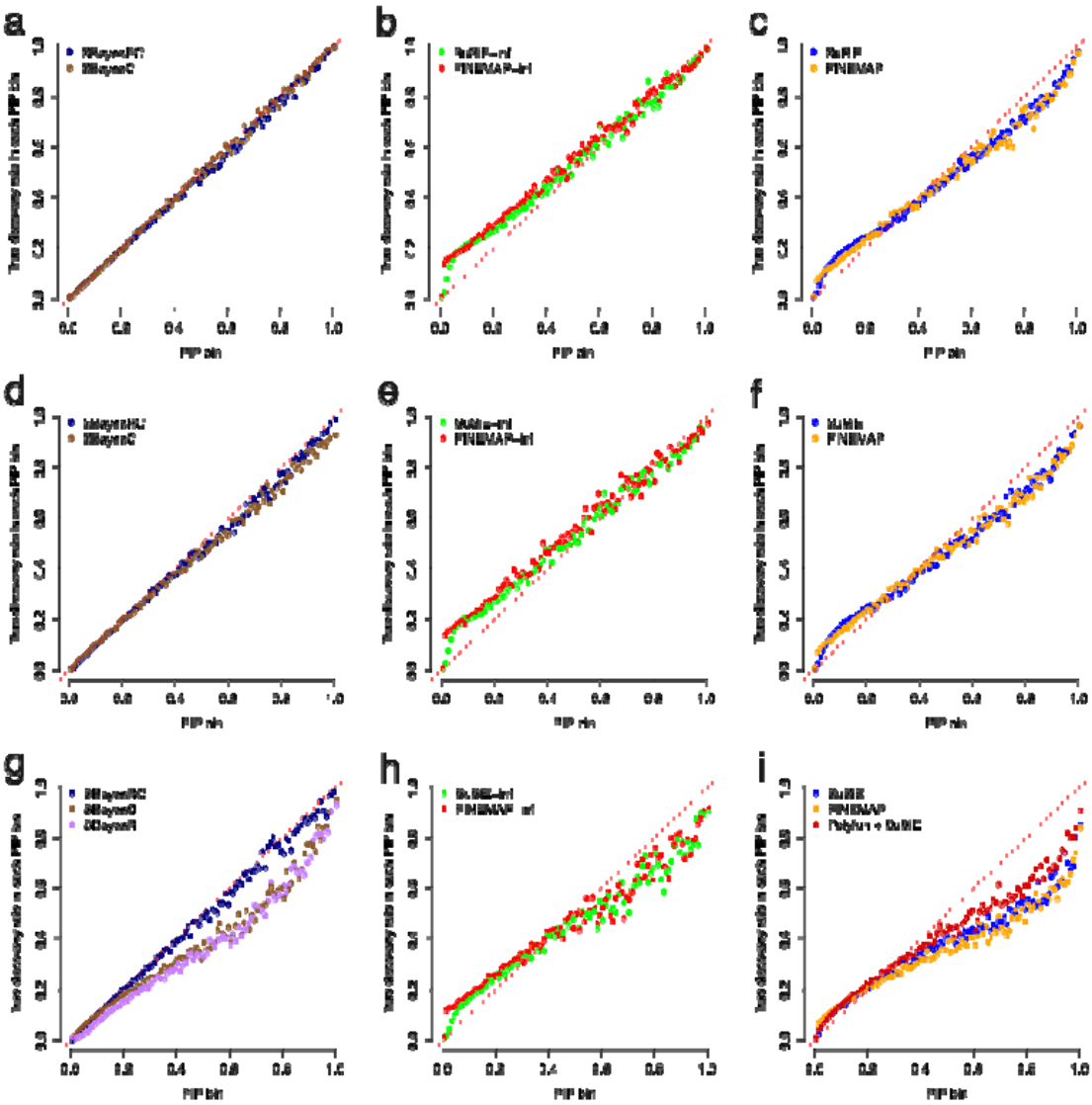

Figure 2 Comparison in the calibration of PIP between GWFM and existing fine-mapping methods under simulations with various genetic architectures. Shown are relationships between PIP and the true discovery rate (TDR) across 100 PIP bins. Left column: SBayesC, SBayesR, and SBayesRC; middle column: SuSiE-inf and FINEMAP-inf; right column: SuSiE, FINEMAMP, and Polyfun + SuSiE. Top row: sparse genetic architecture; middle row: large-effects genetic architecture; Bottom row: LDMS architecture.

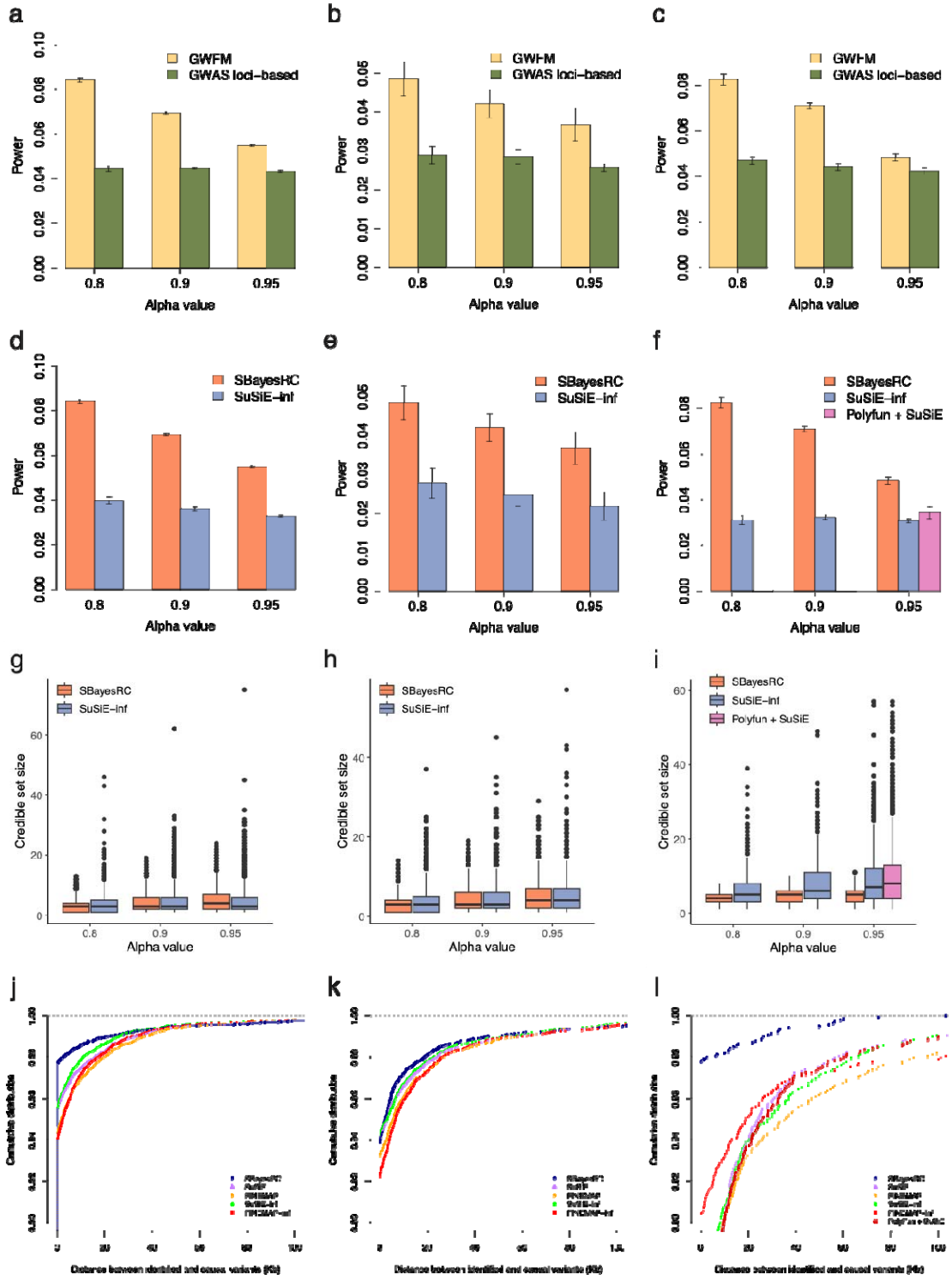

Figure 3 Comparison in mapping power, resolution, and precision between methods based on simulations. Panels (a-c) compare mapping power between GWFM and GWAS loci-based strategies at the same confidence level (α) for LCSs. For a fair comparison, both strategies used the same PIP estimates from SBayesRC, whereas GWAS loci-based fine-mapping constructed α-LCSs restricted to 2Mb regions around GWAS lead SNPs (P-value < 5×10^{-8}). Panels (d-f) compare

1031 mapping power across different methods, including SBayesRC, SuSiE-inf, Polyfun + SuSiE.
1032 Power was quantified as the proportion of simulated causal variants included in the credible
1033 sets reported by each method. Panels (g-i) compare mapping resolution across SBayesRC,
1034 SuSiE-inf, and Polyfun + SuSiE, measured by the credible set size. Panels (j-l) compare mapping
1035 precision across different methods, measured by the physical distance between the causal
1036 variants and the SNPs identified at PIP ≥ 0.9 in each method. Columns correspond to
1037 simulations under the sparse genetic architecture (left column), large-effects genetic
1038 architecture (middle column), and LDMS genetic architecture (right column).

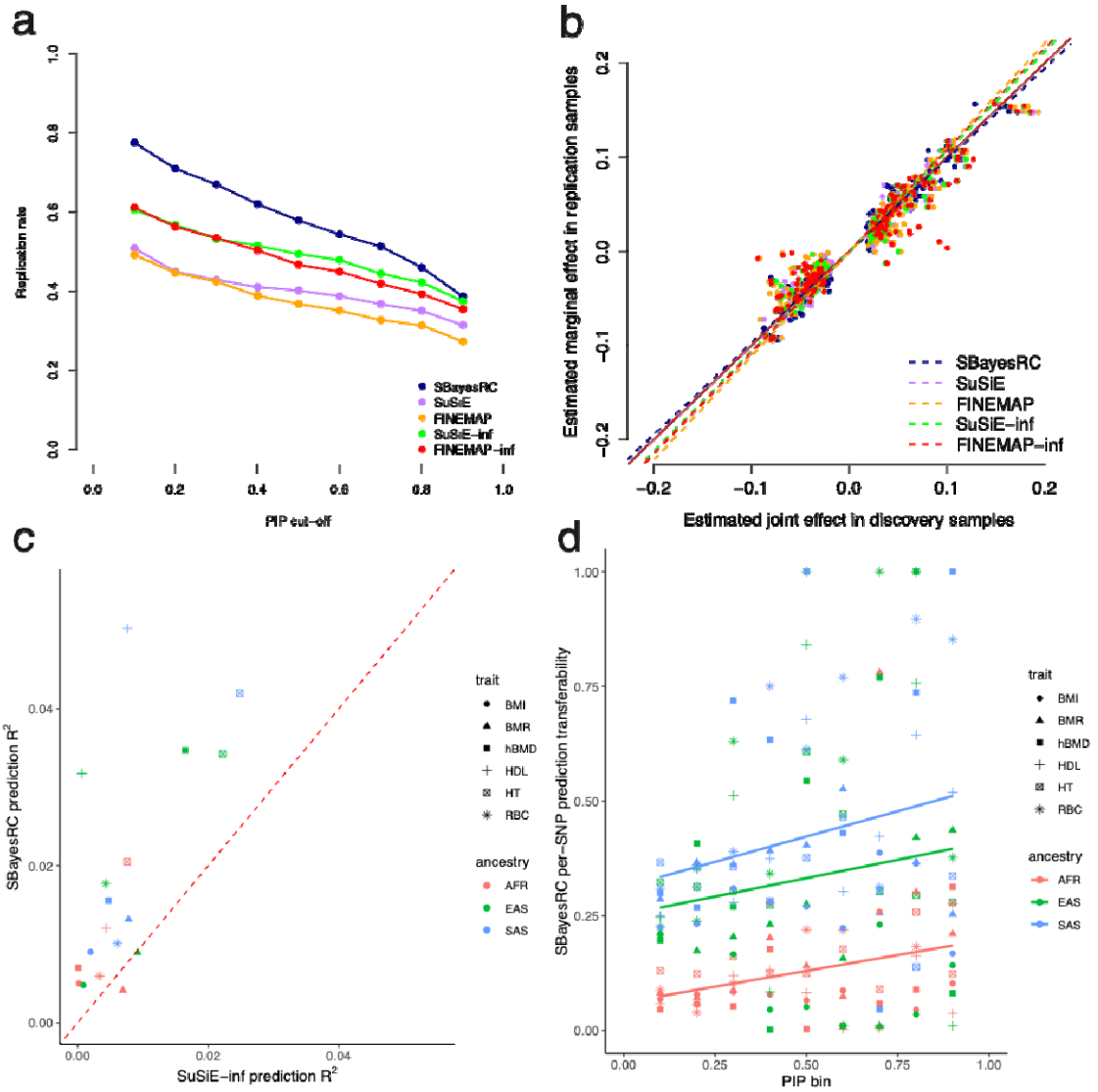

Figure 4 Comparison of independent sample replication, effect size estimation bias, and prediction accuracy using fine-mapped variants across fine-mapping methods in UK biobank traits. (a) Replication rate of discovery using different methods at a given PIP threshold in the replication sample (x-axis) for height in the UKB. (b) Regression of the estimated marginal effect sizes in replication samples on the estimated joint effect sizes in discovery samples using different fine-mapping methods. Dash lines show the regression slopes, where values closer to one indicate less bias, as the marginal effect estimated in the independent replication sample is an unbiased estimate to the true effect. The brown solid line is y=x. (c) Comparison of trans-ancestry prediction accuracy using fine-mapped variants (PIP > 0.9) from SBayesRC and SuSiE-inf, based on the analysis using samples of European ancestry for GWAS and the other ancestries for validation across 6 UKB traits. BMI: body mass index; BMR: base metabolic rate; hBMD: heel bone mineral density; HDL: high-density lipoprotein level; HT: height; RBC: red

1052 blood cell count. AFR: African ancestry; EAS: East Asian ancestry; SAS: South Asian ancestry. (d)
1053 Relationship between trans-ancestry prediction transferability and PIP estimates in the
1054 European ancestry. The transferability was computed as non-EUR-R^2/ EUR-R^2. The solid lines
1055 are regression lines across traits in each ancestry.

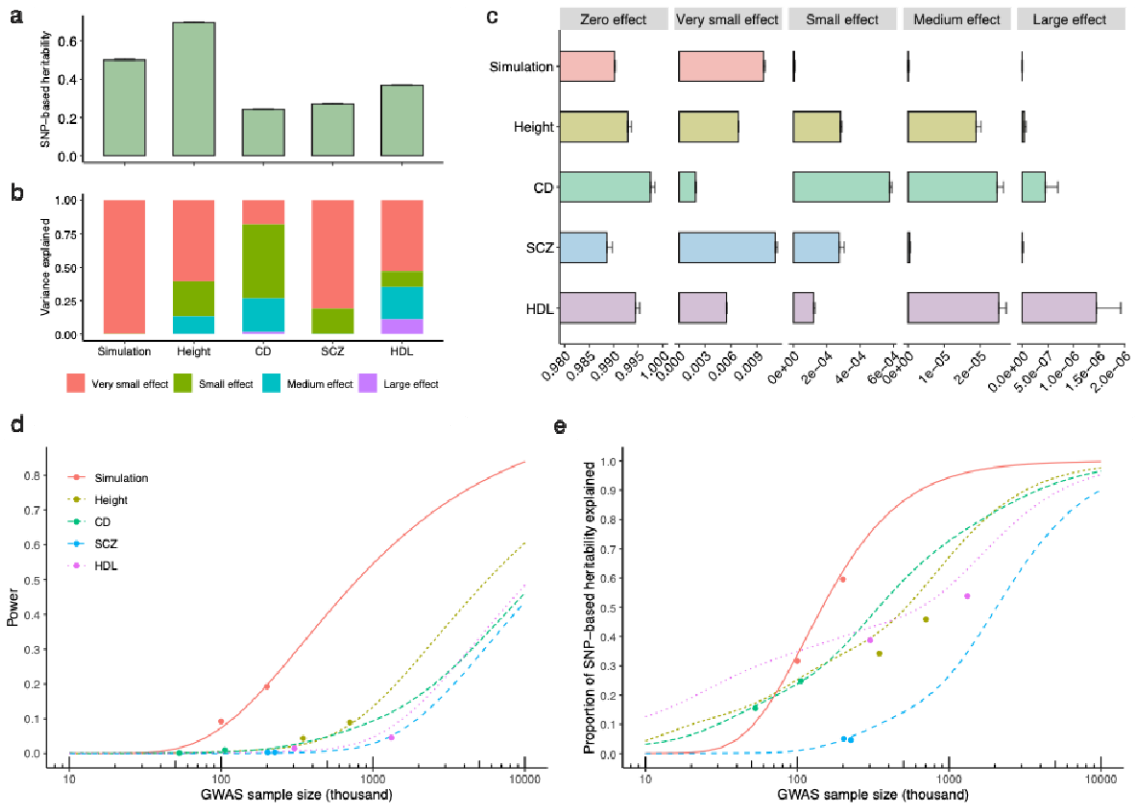

Figure 5 Projection of genome-wide fine-mapping outcomes to the theoretical power prediction in complex traits with diverse genetic architectures. (a) Estimates of SNP-based heritability (h^2_{SNP}) obtained using SBayesRC for height, Crohn's disease (CD), schizophrenia (SCZ), high density lipoprotein (HDL), and a simulated trait with the sparse genetic architecture. (b) Proportions of h^2_{SNP} explained by different mixture components in the SBayeRC model. (c) Proportions of SNPs assigned to different mixture components ($\boldsymbol{\pi}$) based on their estimated effect sizes. The "zero", "very small", "small", "medium" and "large" effect categories correspond to the five variance components in SBayesRC, which explain 0, 0.001, 0.01, 0.1, and 1% h^2_{SNP}, respectively. For the simulated traits, all 10,000 causal variants had small effects sampled from a single normal distribution, leading to all estimated effect sizes being assigned to the very small effect category. (d-e) Theoretical prediction of the power of identifying causal variants (d) and the proportion of h^2_{SNP} explained by the identified causal variants (e) across different GWAS sample sizes for these traits. Dots show the empirical observations based on 0.9-LCSs identified from SBayesRC. The genetic architecture estimates from (a, c) were used as input data for the fine-mapping power prediction (d-e). To check consistency, two datasets with different GWAS sample sizes were analyzed for each trait, with genetic architecture estimates derived from the dataset with the smaller sample size.

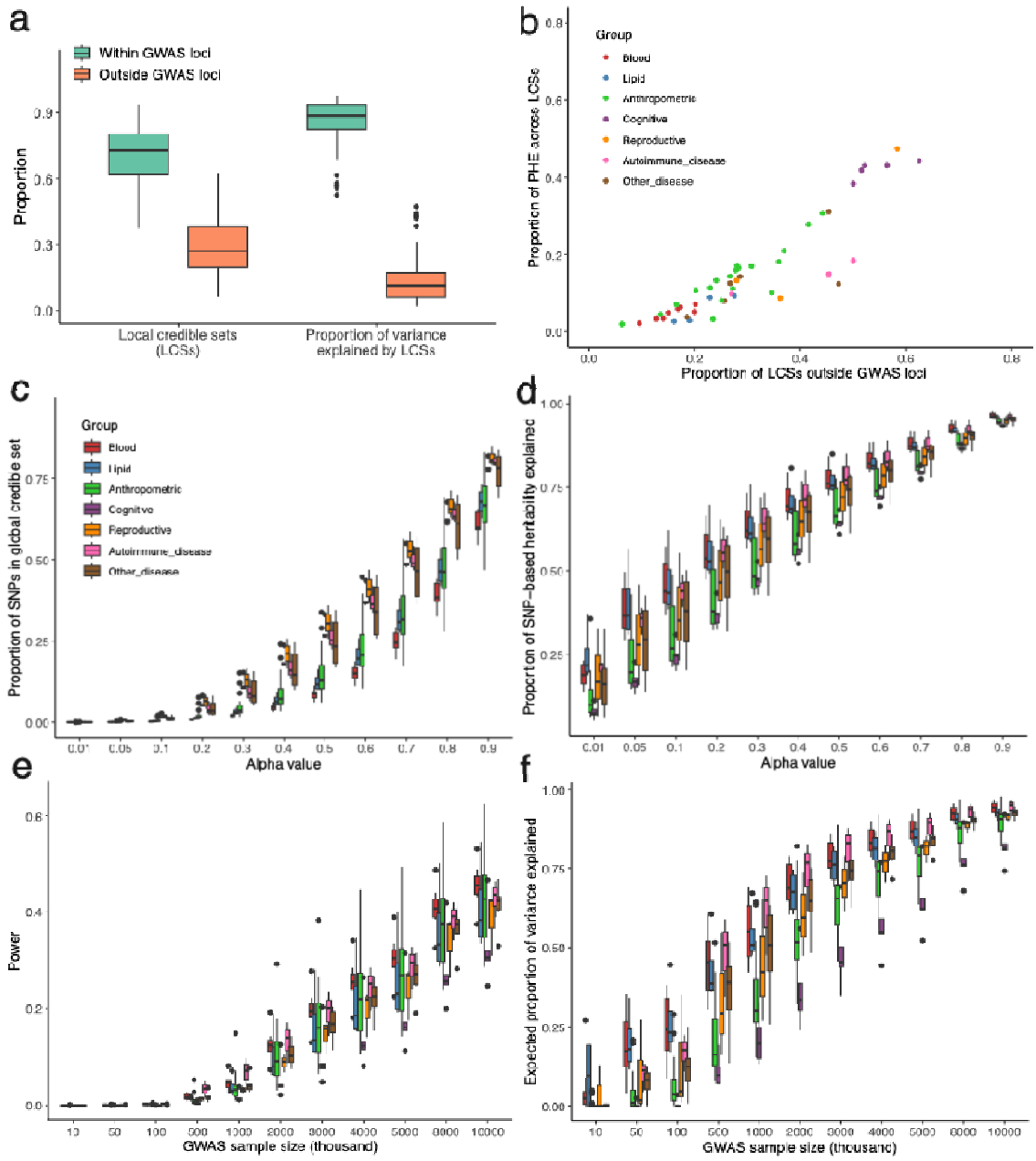

Figure 6 Identification of local credible set and genome-wide credible set SNPs, and theoretical prediction of fine-mapping power for 48 independent complex traits. Panel (a) shows the proportion of the identified local credible sets (LCSs) within or outside the GWAS loci (left) and the proportion of variance explained by the identified LCSs within or outside the GWAS loci (right). Panel (b) shows the relationship between the proportion of LCSs outside GWAS loci and proportion of PHE across the 48 complex traits. Panel (c) shows the proportion of identified GCS SNPs at different alpha threshold (proportion of causal variants captured) for the 48 complex traits (average sample size = 291K). Panel (d) shows the proportion of h^2_{SNP} explained by the GCS SNPs at different alpha threshold. Panel (e-f) show theoretical prediction of power and

1084 proportion of SNP-based heritability explained by LCS SNPs at different GWAS sample sizes for
1085 the 48 complex traits, respectively. Colours indicate different trait categories.

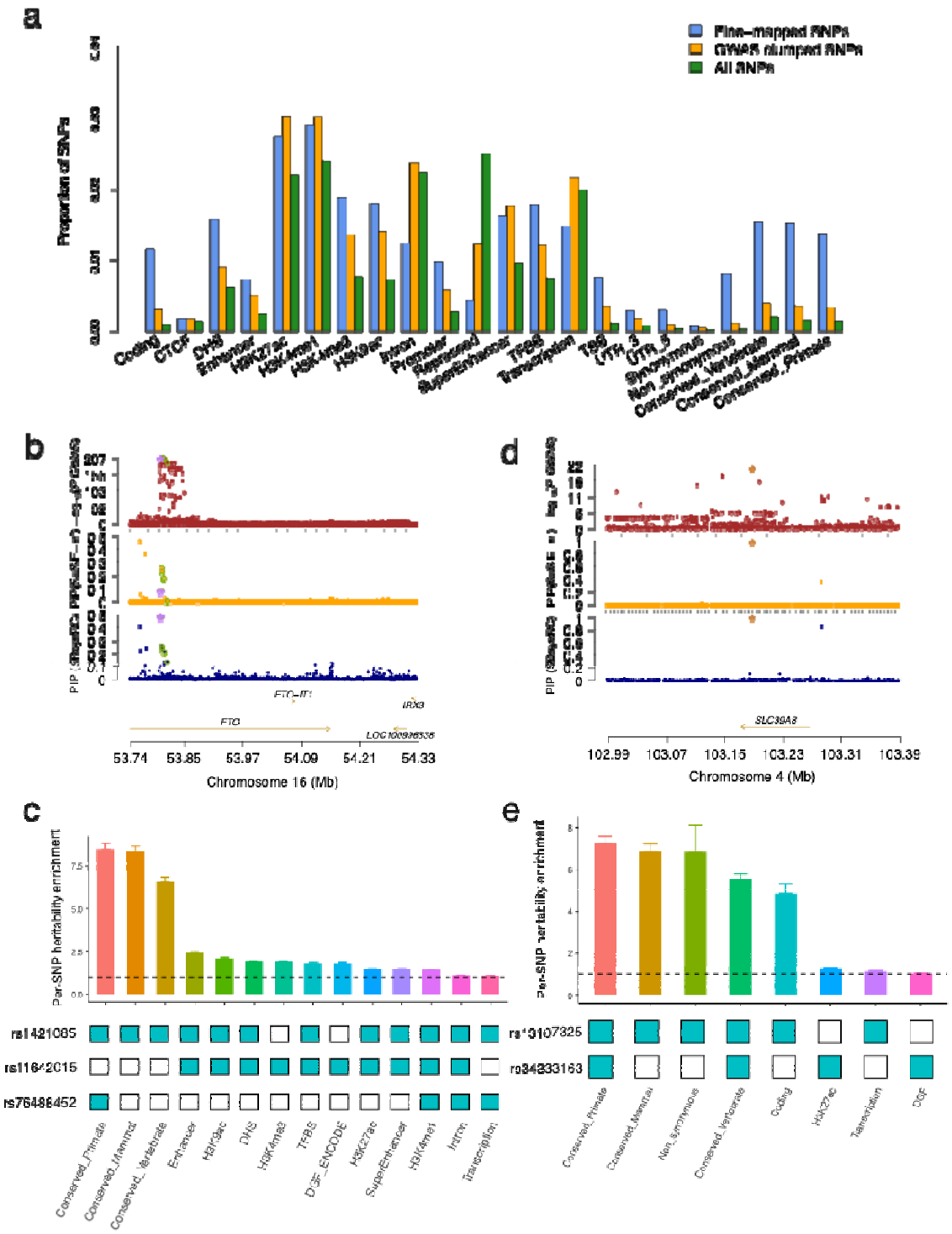

Figure 7 Genome-wide fine-mapping with functional annotations helped pinpoint the putative causal variants. Panel (a) shows enrichment of the genome-wide fine-mapped SNPs from SBayesRC and GWAS clumped SNPs in the 22 main functional categories defined in the LDSC baseline model. Panel (b) shows the prioritized causal variant at the *FTO* locus for BMI. The top track shows the *FTO* locus plot of the standard GWAS for BMI, the second track shows the

similar plot but with the PIP from SBayesRC for BMI, and the third track shows the similar plot with PIP from SuSiE-inf. The dots with green circles are credible set SNPs identified by SBayesRC. The starred SNP is the known causal variant (rs1421085) for obesity at the *FTO* locus supported by previous functional studies. Panel (c) shows the per-SNP heritability enrichment for the causal variant (rs1421085), the GWAS lead variant (rs11642015) and the secondary signal (rs76488452) for BMI at the *FTO* locus. The annotations on the x-axis were those present at least once in these three variants, excluding quantitative annotations and duplicated annotations with flanking windows. Panel (d) shows the prioritized causal variant at the *SLC39A8* locus for SCZ. The top track shows the *SLC39A8* locus plot of the standard GWAS for SCZ, and the second track shows the similar plot but with the PIP from SBayesRC for SCZ. The starred SNP is the missense variant (rs13107325) fine-mapped for SCZ at the *SLC39A8* locus. Panel (e) shows the per-SNP heritability enrichment for the causal variant (rs13107325) and the secondary signal (rs34333163) for SCZ at the *SLC39A8* locus.