Evaluation of Fine Mapping Approaches using Bayesian Linear Regression Models

Shrestha, Merina; Rohde, Palle Duun; Sørensen, Peter

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Evaluation of Fine Mapping Approaches using Bayesian Linear Regression Models

M. Shrestha[1], P. D. Rohde[2], P. Sørensen[1]

1 Center for Quantitative Genetics and Genomics (QGG), Faculty of Technical Sciences, Aarhus University, Aarhus, Denmark

2 Department of Health Science and Technology (HST), Aalborg University, Aalborg, Denmark

Corresponding author: mesh@qgg.au.dk

AU AARHUS UNIVERSITY

Background and Aims

Fine mapping identifies the underlying causal genetic variants and is essential for the future of precision medicine. We propose a novel statistical method, Bayesian linear regression (BLR) models, and aim to investigate their fine mapping efficiency (power and precision) based on definition of:

1. Prior distribution for marker effects
2. Marker sets
3. Marker set association statistics

Results

- Larger number of true positive (TP) SNPs along with low number of false positive (FP) SNPs were identified by Bayes C and Bayes R prior distribution models.
- Highest average F_1 score for the marker set, linkage disequilibrium (LD) ($r^2 > 0.7$) across all prior models and marker set association statistics suggests inclusion of LD information while designing a marker set.
- Higher F_1 score of the marker set association statistic, sum of posterior inclusion probability (T_{PIP}) in the stringent significance threshold (*p<1e-05*) suggests use of the PIP information compared to the genetic variance information.

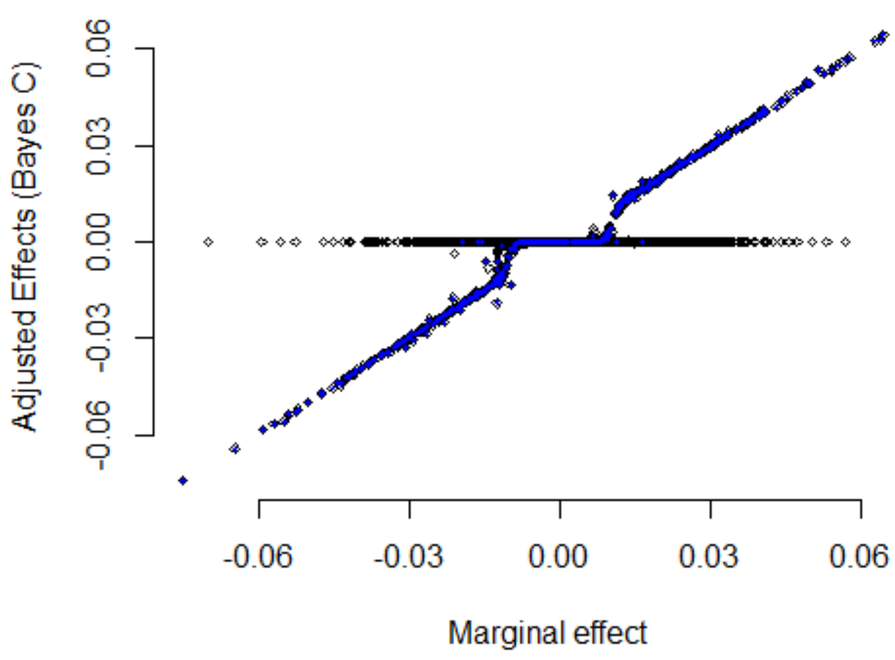

Figure 1. Scatter plot for marginal effect and adjusted effect obtained from Bayes C model for causal SNPs (π) 0.1% and genetic architecture scenario (GA1).

Conclusions

BLR fine mapping is influenced by the choice of the marker sets definition, the prior distributions, and type of marker set association statistic. Further plan includes comparison of BLR models to other fine mapping methods.

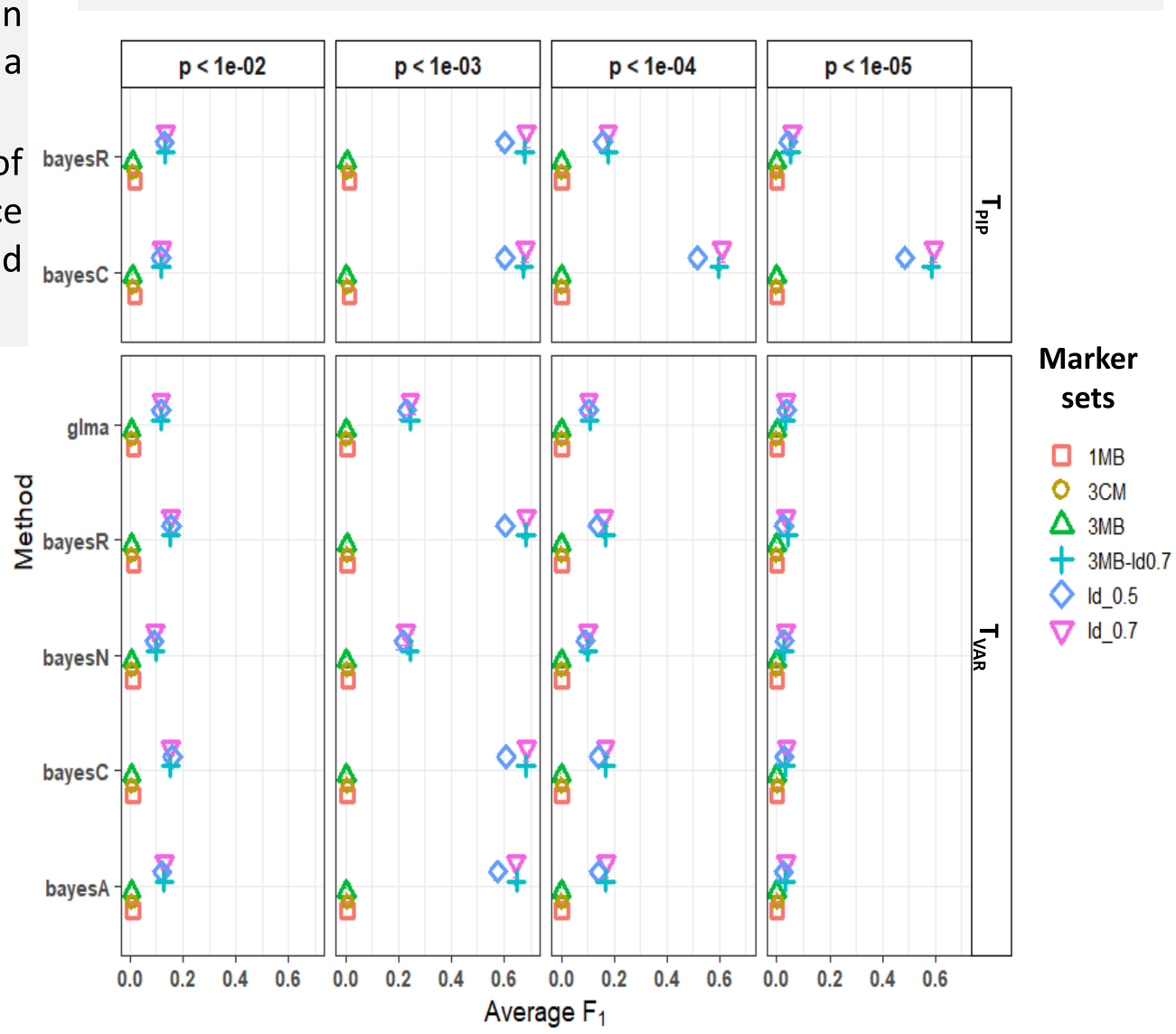

Figure 2. Average F_1 score of marker set association statistics, Sum of genetic variance (T_{VAR})and Sum of posterior inclusion probability (T_{PIP}) for π 0.1% and GA1.

Materials and Methods

Data

- UK biobank (UKB) genotypes (ID: ***96479***)
- QC: MAF, HWE, MHC region, multiallelic and indel markers
- 335,744 White British Unrelated (WBU) and 533,679 SNPs

Simulation

- Quantitative phenotype for 268,595 WBU (10 replicates)
- h^2 30% and Causal SNPs (π): 0.1%, 1% and 5%
- Two genetic architecture (GA) scenarios: GA1 (Bayes C) and GA2 (Bayes R)

Fine mapping

- Single SNP linear regression → marginal β estimates
- marginal β estimates + linkage disequilibrium (LD) matrix → BLR models → adj. β estimates
- Estimate association test statistics for marker sets
- ***qgg*** package in R

Marker sets

- LD ($r^2 > 0.7$ and $r^2 > 0.5$)
- Physical map – 1MB, 3MB
- Genetic map – 1CM, 3CM

Prior distribution of marker effects

- Low – Bayes C
- Moderate – Bayes R and Bayes A
- High – Bayes N

Marker set association statistics

- Sum of genetic variance (T_{VAR})
- Sum of posterior inclusion probability (T_{PIP})

F_1 classification score $= \frac{2pr}{p + r}$,

Precision: $p = TP/(TP + FP)$

Power: $r = TP/(TP + FN)$

*Authors would like to acknowledge **Open Discovery Innovation Network (ODIN)** for funding this study (grant no.: NNF20SA0061466).*