flashfm-ivis: interactive visualisation for fine-mapping of multiple quantitative traits

Feng Zhou[1,*], Adam S Butterworth[2,3,4,5] and Jennifer L Asimit[1,*]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, CB2 0SR, UK, [2]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK, [3] National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge CB1 8RN, UK, [4]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge CB1 8RN, UK, [5]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge CB10 1SA, UK.
*To whom correspondence should be addressed.

Abstract

Summary: *flashfm-ivis* provides a suite of interactive visualisation plots to view potential causal genetic variants that underlie associations that are shared or distinct between multiple quantitative traits and compares results between single- and multi-trait fine-mapping. Unique features include network diagrams that show joint effects between variants for each trait and regional association plots that integrate fine-mapping results, all with user-controlled zoom features for an interactive exploration of potential causal variants across traits.

Availability and Implementation: *flashfm-ivis* is an open-source software under the MIT license. It is available as an interactive web-based tool (http://shiny.mrc-bsu.cam.ac.uk/apps/flashfm-ivis/) and as an R package format. Code and documentation are available at https://github.com/fz-cambridge/flashfm-ivis and https://zenodo.org/record/6376244#.YjnarC-l2X0. Additional features can be downloaded as standalone R libraries to encourage reuse.

Contact: feng.zhou@mrc-bsu.cam.ac.uk or jennifer.asimit@mrc-bsu.cam.ac.uk

Introduction

Genome-wide association studies (GWAS) have successfully identified many genetic variants that are associated with diseases and traits (Claussnitzer et al., 2020). Identifying the causal variants that underlie genetic associations is key to translating these findings into new therapeutic targets or revealing new biological insights for diseases. Statistical fine-mapping aids this by identifying potential causal variants with the aim of reducing the number of genetic variants for follow-up in downstream functional validation experiments (Spain and Barrett, 2015; Hutchinson et al., 2020). As biologically related traits often have shared causal variants, multi-trait fine-mapping that shares information between traits can improve precision over single-trait fine-mapping of each trait independently.

There are few multi-trait fine-mapping methods that allow multiple causal variants at a single locus due to the computational complexity of many possible model combinations between traits. One approach is to restrict traits to have the same causal variants and allow for different effect sizes, as in fastPAINTOR (Kichaev et al., 2017). In contrast, flashfm (Hernandez et al., 2021) makes no such

restrictions and uses a Bayesian framework that upweights joint models with shared causal variants. In extensive simulation comparisons, flashfm was shown to have higher precision than fastPAINTOR.

Bioinformatics tools are moving in the direction of dynamic interaction between GWAS data and plots (**Table_S1**), but most require some programming knowledge; none of these help to explore fine-mapping results. Non-interactive fine-mapping visualisation tools include PAINTOR-CANVIS for visualising a single set of fine-mapping results (and linkage disequilibrium (LD) structure) from PAINTOR (Kichaev et al., 2017), and echolocatoR (Schilder et al., 2021) for single-trait fine-mapping results from several methods; both require some programming knowledge.

Flashfm-ivis provides interactive exploration and publication-ready plots to summarise fine-mapping results from multiple traits. Linked from *flashfm-ivis,* users may use *finemap-ivis* to interact with and plot single-trait results. **Table_1** compares the key features of the tools most similar to *flashfm-ivis*.

Flashfm multi-trait fine-mapping uses single-trait fine-mapping results from FINEMAP (Benner et al., 2016) or JAM (Newcombe et al., 2016), and we refer to either of these methods as "fm". For ease in interpretation, variant groups are constructed in flashfm for both fm and flashfm such that variants in the same group can be viewed as exchangeable, i.e., they are in high LD and rarely appear in the same model together. Key features of *flashfm-ivis* are shown in **Fig.1** (also listed in **Table_S2),** with an overview here:

- Interactive visualisations encourage users to compare results for multiple traits
- Downloadable summary output from different plots, based on users' selections (e.g., credible sets, variant groups, etc.), simplifying complex results and giving insight for further analyses (e.g., discovering shared variants in credible sets for different trait combinations).
- Flexible ways of viewing the joint effects of variantS on one or more traits.

Implementation

Flashfm-ivis is currently built in R and minimises the complex (inter-)dependency with other R packages (**Table_S2**), making it a standalone tool (easier to maintain in the long-term). The web-based version (http://shiny.mrc-bsu.cam.ac.uk/apps/flashfm-ivis/) does not require users to have any programming skills, encouraging a wide spectrum of researchers to interact with and visualise their own results. It includes six Dashboard tabs that give access to various comparisons and summaries, links to download all codes (https://github.com/fz-cambridge/flashfm-ivis) and a direction to *finemap-ivis* (http://shiny.mrc-bsu.cam.ac.uk/apps/finemap-ivis/).

2.1 Data inputs

A pre-loaded example from a cardiometabolic GWAS of a Ugandan cohort for four lipid traits and the *APOE* region (Hernandez et al. 2022) is included in *flashfm-ivis* to illustrate its main features (**Figure_S1**).

To view both single and multi-trait fine-mapping results, users need to upload their output from flashfm and its input files to *flashfm-ivis* (details in **Supplementary Material**). To view only single-trait fine-mapping results from FINEMAP (**Figure_S13)**, users upload the standard FINEMAP output files to *finemap-ivis*.

Users can verify their inputs on a sub-dashboard (**Figure_S1** and **Figure_S12**). All plots are interactive, allowing the user to control zoom-in focus areas in regional association plots and to drag network nodes to change the perspective of plots. An overview of key plots follows (also see **Table_S2**, all detailed plots in **Supplementary Material**).

2.2 Control widgets on the sidebar panel

Control widgets help users to refine the input data selection and then focus on key results; both credible sets and *MPP* (marginal posterior probability of variant causality – the posterior probability that the variant is included in a model) values can be controlled (**Figure_S2**).

2.3 Fine-map integrated regional association plots

Individual regional association plots are presented for each trait in **Figure_S2** and MPPs from fine-mapping are denoted by the dot size. In **Figure_S8**, all regional association plots are linked together for easy comparisons between traits and between methods – users may select a subset of variants to focus on in all plots.

2.4 Network plots of joint variant effects

Figures S4-S7 present both variant group- and individual-level dynamic networks. The key features are: (1) users interact with the networks by specifically defining the PP thresholds for variant models, so the network can be expanded or refined; (2) the size and colour of both nodes and edges indicate evidence strength and associated traits, as detailed in **Table_S2**; (3) sub-networks can be formed and moved. When viewing models based on variant groups, we consider the group model PP (PPg), which is the sum of the PPs for all models that have exactly one variant from each group listed in the model; e.g., model A+B consists of all variant models with one variant from each of groups A and B (i.e. fm in "Ch_1_Single_trait'' and flashfm in "Ch_2_Multi_trait").

2.5 Venn diagrams of shared potential causal variants

In **Fig.1** and **Figure_S10**, through interactive Venn diagrams, users see and download lists of variants in each combination of overlapping credible sets.

2.6 Spider/radar diagram of credible set sizes

In **Fig.1** and **Figure_S9**, the spider chart compares the number of variants in credible sets of fm/flashfm methods based on different traits.

2.7 Sankey diagrams of variant groups

In **Fig.1** and **Figure_S11**, Sankey flowcharts show the connections between variants and variant groups in fm/flashfm methods and based on different traits.

Conclusion

We provide a user-friendly fully interactive web tool, *flashfm-ivis*, that is accessible without any programming knowledge; it is available as an R package for those who prefer to run it from their own machine. It promotes exploration of fine-mapping results of several traits and helps with interpretation, as well as identification of variants for further follow-up. Users can interact with plots and decide on the final version of publication-ready plots for download. *Flashfm-ivis* output, such as lists of variants in credible sets for selected traits, is easily downloaded for further follow-up. If users are interested in only a single trait, there is an option to produce plots based on FINEMAP output only, *finemap-ivis*. We hope that *flashfm-ivis* will become standard practice for exploring fine-mapping results and will contribute to revealing the underlying mechanisms of diseases.

Acknowledgements

Funding

Data Availability

The example datasets can be found in the web-based version of flashfm.ivis (http://shiny.mrc-bsu.cam.ac.uk/apps/flashfm-ivis/) and its GitHub repository (https://github.com/fz-cambridge/flashfm-ivis), as well as https://zenodo.org/record/6376244#.YjnarC-I2X0.

Supplementary Material

See the online appendix and the README tab of *flashfm-ivis* (http://shiny.mrc-bsu.cam.ac.uk/apps/flashfm-ivis/), which also includes YouTube video demonstrations of the tool.

Conflict of Interest: none declared.

References

Benner, C., Spencer, C. C., Havulinna, A. S., Salomaa, V., Ripatti, S., & Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics, 32(10), 1493-1501.* https://doi.org/10.1093/bioinformatics/btw018.

Boughton, A. P., Welch, R. P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G. R., & Boehnke, M. (2021) LocusZoom. js: interactive and embeddable visualization of genetic association study results. *Bioinformatics, 37(18), 3017-3018.* https://doi.org/10.1093/bioinformatics/btab186.

Claussnitzer, M., Cho, J.H., Collins, R. *et al.* A brief history of human disease genetics. *Nature* 577, 179–189 (2020). https://doi.org/10.1038/s41586-019-1879-7

George, G., Gan, S., Huang, Y., Appleby, P., Nar, A. S., Venkatesan, R., ... & Doney, A. S. (2020). PheGWAS: a new dimension to visualize GWAS across multiple phenotypes. *Bioinformatics, 36(8), 2500-2505.* https://doi.org/10.1093/bioinformatics/btz944.

He, F., Ding, S., Wang, H., & Qin, F. (2020). IntAssoPlot: An R Package for Integrated Visualization of Genome-Wide Association Study Results With Gene Structure and Linkage Disequilibrium Matrix. *Frontiers in genetics, 11, 260.* https://doi.org/10.3389/fgene.2020.00260

Hernandez, N., Soenksen, J., Newcombe, P., Sandhu, M., Barroso, I., Wallace, C., Asimit, J.L. (2021). The flashfm approach for fine-mapping multiple quantitative traits. *Nat Commun* 12, 6147. https://doi.org/10.1038/s41467-021-26364-y

Hutchinson, A., Asimit, J., Wallace, C. (2020). Fine-mapping genetic associations, *Human Molecular Genetics*, Volume 29, Issue R1, Pages R81–R88, https://doi.org/10.1093/hmg/ddaa148

Khramtsova, E. A., & Stranger, B. E. (2017). Assocplots: a Python package for static and interactive visualization of multiple-group GWAS results. *Bioinformatics, 33(3), 432-434.* https://doi.org/10.1093/bioinformatics/btw641.

Kichaev, G., Roytman, M., Johnson, R., Eskin, E., Lindstroem, S., Kraft, P., & Pasaniuc, B. (2017). Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics, 33(2), 248-255.* https://doi.org/10.1093/bioinformatics/btw615.

Kierczak, M., Jablonska, J., Forsberg, S. K., Bianchi, M., Tengvall, K., Pettersson, M., et al. (2015). Cgmisc: enhanced genome-wide association analyses and visualization. *Bioinformatics 31, 3830–3831.* https://doi.org/10.1093/bioinformatics/btv426.

Machiela, M. J., and Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics 31, 3555–3557.* https://doi.org/10.1093/bioinformatics/btv402.

Newcombe, P. J., Conti, D. V. & Richardson, S. (2016). JAM: a scalable Bayesian framework for joint analysis of marginal SNP effects. *Genet. Epidemiol.* 40, 188–201 https://doi.org/10.1002/gepi.21953

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., ... & Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics, 26(18), 2336-2337.* https://doi.org/10.1093/bioinformatics/btq419.

Schilder, B. M., Humphrey, J., & Raj, T. (2021). echolocatoR: an automated end-to-end statistical and functional genomic fine-mapping pipeline. *Bioinformatics. 1-4,* https://doi.org/10.1093/bioinformatics/btab658.

Spain, S. L., Barrett, J. C. (2015). Strategies for fine-mapping complex traits. *Human molecular genetics, 24(R1), R111-R119.* https://doi.org/10.1093/hmg/ddv260

Table_1: Comparison of visualisation features and tools for GWAS and fine-mapping results. A "+" indicates that the tool has the feature, possibly with a few modifications.

Features and tools	Flashfm-ivis	echolocatoR	PAINTOR-CANVIS	PheGWAS	LocusZoom.js	Cgmisc	LDlink	Assocplots	IntAssoPlot
Input data in/from different formats/methods	+								
Display fine-mapping results	+	+	+						
Multi-panel comparison with linked data	+	+						+	
Joint SNP effects (fine-mapping model PPs)	+								
Interactive features/tools	+			+	+		+	+	
Download outputs/plots	+	+	+	+			+		
A standalone web or R package	+			+	+		+		
Regional association plot of GWAS results	+	+	+	+	+	+	+	+	+
Scatter plot of SNP PPs (MPP) only	+[a]	+	+						
Regional association plot with SNP PPs	+								
LD with lead SNP	+	+			+	+	+[b]		+
LD heatmap and plot	+		+				+[b]		+
Link between GWAS and LD matrix	+								+
Integrated display of multiple GWAS	+			+				+	+

No programming knowledge required	+				+		+		
Allow iPad or touch screen	+								
Can be used with other R/Python packages	+	+		+	+	+			
Reference		Schilder et al., 2021	Kichaev et al., 2017	George et al., 2020	Boughton et al., 2021; Pruim et al., 2010	Kierczak et al., 2015	Machiel and Chanock, 2015	Khramtsova and Stranger, 2017	He et al. 2020

[a]*Flashfm-ivis* integrates the SNP PPs with the regional association plots. [b]LDlink only uses built-in reference panels.

Fig.1: Selected *flashfm-ivis* interactive plots for 4 lipid traits from Ugandan GWAS data in the *APOE* region[a]. (A) Interactive and Linked regional association plots with integrated fine-mapping results. Four lipid traits (LDL-cholesterol, total cholesterol (TC), triglycerides (TG), and HDL-cholesterol) were single- (left panel) and multi-trait (right panel) fine-mapped; the rows correspond to each trait: LDL, TC, TG, HDL. The regional association plots (-log10(p) against variant position) integrate the variant MPPs (see Section 2.2) from fine-mapping, by displaying the diameter of each variant point as proportional to its MPP. Colours indicate variant groups, which consist of variants in high LD and could be treated as interchangeable in the causal models.

Users may click on particular groups to focus on (or to remove) in all plots, or they may draw their selection of points in one plot as a focus for all plots. **(B) Network diagram of variant group causal models.** variant group models are displayed for the four lipid traits. Each node represents a variant group, and edges between nodes indicate variant groups that appear together in a model; only models with PPg (as defined in Section 2.4) within the range of the widget selection are included. The colours of the nodes indicate which traits include the variant group in their models, and their diameter is proportional to the frequency that they appear in models over all traits. Edge colours indicate which traits include the pair of variant groups in a model, while edge thickness is proportional to the PPg values. **(C) Radar Chart and Venn diagram of credible sets for traits.** These plots show and compare the number of variants that appear in multiple credible sets for different methods and traits. For example, in the Venn diagram the four traits share 2 variants in their credible sets constructed from flashfm, while the Radar chart compares the total number of variants based on FINEMAP (cs_1) and flashfm (cs_M). Hovering on that segment shows the variant ids, and details of each segment are in a downloadable table. **(D) Sankey diagram connects individual variant ids and their groups in flashfm or FINEMAP (fm).** This links the individual variants and their groups in different methods. The size of the edge reflects the MPP value of the variant in each trait (e.g., trait_2 cholesterol here, users can interact with the plots to show the values).
[a]Hernandez et al. 2021.
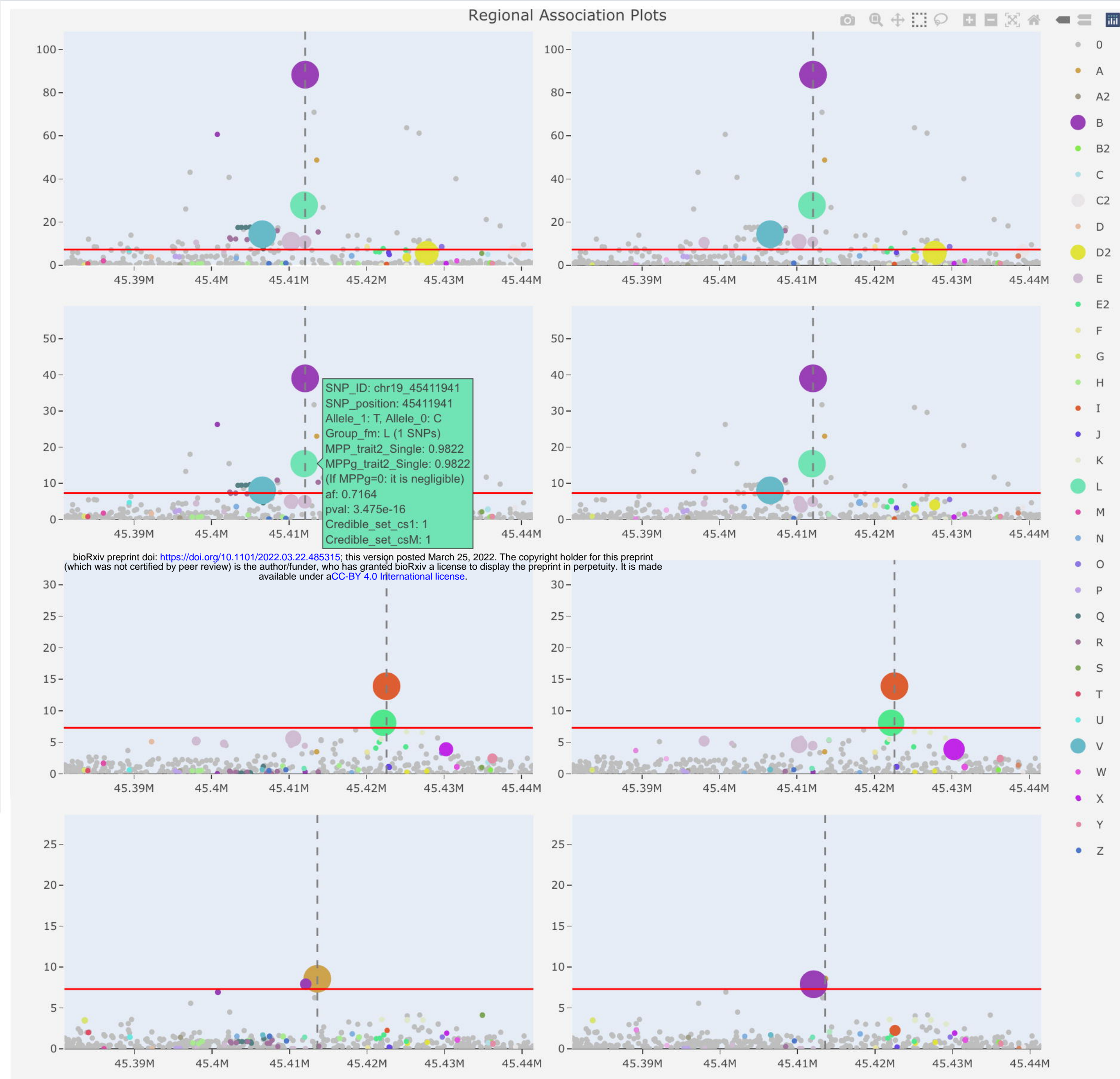
(A)

snpGroup$group.sizes

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Color	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
FINEMAP	1	2	1	1	5	1	1	9	1	2	3	1	1	2	1	2	6	9	1	1	1	1	1	1	3	2
flashfm	1	1	0	0	5	1	1	0	1	1	4	1	0	2	1	2	0	1	0	0	1	1	1	1	1

How many traits: (NOTE - left panel is fm, right panel is flashfm, rows show individual traits)

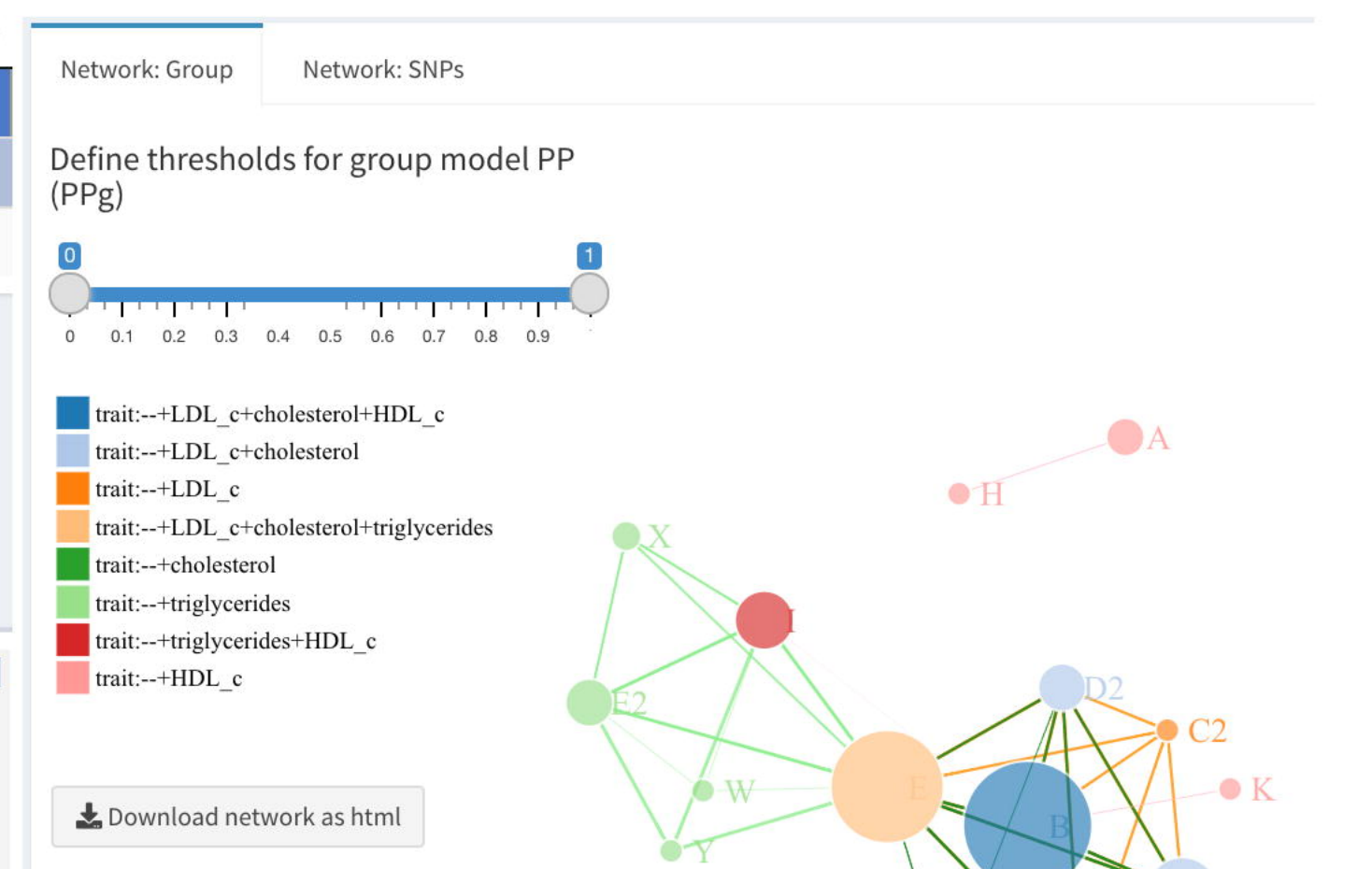○ 2 traits ○ 3 traits ● 4 traits ○ 5 traits ○ 6 traits

Show a dashed line at the maximum value:

● Yes ○ No

Regional Association Plots

Legend: 0, A, A2, B, B2, C, C2, D, D2, E, E2, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z

SNP_ID: chr19_45411941
SNP_position: 45411941
Allele_1: T, Allele_0: C
Group_fm: L (1 SNPs)
MPP_trait2_Single: 0.9822
MPPg_trait2_Single: 0.9822
(If MPPg=0: it is negligible)
af: 0.7164
pval: 3.475e-16
Credible_set_cs1: 1
Credible_set_csM: 1

(B)

Network: Group   Network: SNPs

Define thresholds for group model PP (PPg)

0 — 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

■ trait:--+LDL_c+cholesterol+HDL_c
■ trait:--+LDL_c+cholesterol
■ trait:--+LDL_c
■ trait:--+LDL_c+cholesterol+triglycerides
■ trait:--+cholesterol
■ trait:--+triglycerides
■ trait:--+triglycerides+HDL_c
■ trait:--+HDL_c

⬇ Download network as html

(C)

Radar chart: Credible Sets   Venn diagram 1: cs1   Venn diagram 2: csM

● cs_1
● cs_M

cholesterol, LDL_c, triglycerides, HDL_c

count
40
30
20
10
0

cholesterol 20
triglycerides 21
LDL_c 41 / 3 / 6
26 / 6 / 0 / 43
HDL_c
2 / 6 / 2 / 0 / 12
0
chr19_45381443 chr19_45401579
chr19_45389381

(D)

Sankey Diagram of SNP groups (trait 2): cholesterol

chr19_45406673
chr19_45412079 — flashfm_B — fm_B
flashfm_Q — fm_Q
chr19_45435661 — fm_C
chr19_45392254 — fm_D
chr19_45422160 — flashfm_E2 — fm_E2
chr19_45425066 — fm_U
chr19_45389381
chr19_45413760
chr19_45408475 — flashfm_R — fm_R
chr19_45407409
chr19_45409595 — flashfm_Z — fm_Z
chr19_45413576 — flashfm_A — fm_A
chr19_45395909 — flashfm_A2 — fm_A2
chr19_45436010 — flashfm_B2 — fm_B2
chr19_45439163 — flashfm_C2 — fm_C2
chr19_45427780 — flashfm_D2 — fm_D2
chr19_45425212
chr19_45410548
chr19_45410273 — flashfm_E — fm_E
source: chr19_45412040
target: flashfm_E 0.2
chr19_45398014
chr19_45401587
chr19_45420079 — flashfm_F — fm_F
chr19_45383562 — flashfm_G — fm_G
chr19_45411941 — flashfm_L — fm_L
chr19_45418086 — flashfm_N — fm_N
chr19_45403192
chr19_45429708 — flashfm_O — fm_O
chr19_45395816 — flashfm_P — fm_P
chr19_45395330
chr19_45406538 — flashfm_V — fm_V
chr19_45393144 — flashfm_F2
chr19_45438575 — flashfm_I2