eQTpLot: an R package for the visualization and colocalization of eQTL and GWAS signals

Theodore G. Drivas[1,2], Anastasia Lucas[2], Marylyn D. Ritchie[2]

[1]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[2]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA , USA

Summary: Genomic studies increasingly integrate expression quantitative trait loci (eQTL) information into their analysis pipelines, but few tools exist for the visualization of colocalization between eQTL and GWAS results. To address this issue, we developed the intuitive R package eQTpLot, which takes as input GWAS and eQTL summary statistics to generate a series of plots visualizing colocalization, correlation, and enrichment between eQTL and GWAS signals for a given gene-trait pair. We believe eQTpLot will prove a useful tool for investigators seeking a convenient and customizable visualization of genomic data colocalization.
Availability and Implementation: the eQTpLot R package and tutorial are available at https://github.com/RitchieLab/eQTpLot
Contact: theodore.drivas@pennmedicine.upenn.edu

INTRODUCTION

Non-protein-coding genetic variants make up the majority of significant associations identified by genome wide association studies (GWAS). As these variants typically do not have obvious consequences for gene function, it can be difficult to map their effects to specific genes. To address this issue, genomic studies have increasingly begun to integrate expression quantitative trait loci (eQTL) information into their analysis pipelines, with the thought that non-coding variants might be exerting their effects on patient phenotypes through the modulation of expression levels of nearby genes. Through this approach, indirect evidence for causality can be obtained if a genetic locus significantly associated with candidate gene expression levels is found to colocalize with a genetic locus statistically significantly associated with the phenotype of interest.

A number of excellent tools have been developed to discover and analyze colocalization between eQTL and phenotype association signals,(Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016; He *et al.*, 2013; Liu *et al.*, 2019; Yao *et al.*, 2020; Nica *et al.*, 2010; Zhu *et al.*, 2016) but few packages provide the necessary tools to visualize these colocalizations in an intuitive and informative way. For this reason, we present eQTpLot, an R package developed for the visualization of colocalization between eQTLs and GWAS signals. The major contributions of eQTpLot lie in its ability to easily generate a series of three customizable plots clearly illustrating,

for a given gene-trait pair: 1) colocalization between GWAS and eQTL signals, 2) correlation between GWAS and eQTL p-values, and 3) enrichment of eQTLs among trait-significant variants. These clear and comprehensive plots provide a unique view of eQTL-GWAS colocalization, allowing for a more complete understanding of the interaction between gene expression and trait associations.

eQTpLot DESRIPTION

eQTplot is a freely available R package (https://github.com/RitchieLab/eQTpLot) that takes as input two data frames, one of GWAS summary statistics and the other of eQTL summary statistic, such as the data available through the GTEx Portal(GTEx Consortium, 2013). The user specifies the HGNC symbol of the gene to be analyzed, the GWAS trait to be analyzed which is useful if the GWAS summary statistics contains information on multiple traits, as one might obtain from a Phenome-Wide Association Study (PheWAS), and the tissue type to use for the eQTL analysis which is useful if eQTL data is available on multiple tissue types. A PanTissue analysis can be specified as well, combining data across tissue types for each variant. Additional parameters may be specified, including the p-value thresholds for GWAS or eQTL significance, the genomic range to be displayed, axis/layout modifications for the resultant graphs, etc. These data are then used to generate and output a series of plots visualizing colocalization, correlation, and enrichment between eQTL and GWAS signals for a given gene-trait pair (Figure 1), discussed in detail below.

eQTL-GWAS Colocalization Plot

To generate the main eQTL-GWAS Colocalization Plot (Figure 1A-B), a locus of interest (LOI) is defined to include the specified gene's coordinates, along with a default 200kb of flanking genomic material. GWAS summary statistics for SNPs within the LOI are plotted in chromosomal space along the horizontal axis, with the inverse log of the p-value of association with the GWAS trait (p_{trait}) along the vertical axis (as one would plot a standard GWAS Manhattan plot). GWAS variants for which eQTL data are not available for the given gene, or for which the eQTL p-value (p_{eQTL}) does not meet the specified significance threshold, are plotted as grey boxes. On the other hand, to colocalize eQTL and GWAS signals, GWAS variants that are also eQTLs for the specified gene (p_{eQTL} < specified eQTL significance threshold) are plotted as colored triangles, with a color gradient corresponding to the inverse magnitude of p_{eQTL}. The size of each triangle is set to correspond to the eQTL normalized effect size (NES) for the variant, while the directionality of each triangle is set to correspond to the direction of effect of the variant for the GWAS trait. A depiction of the genomic positions of all genes within the LOI is generated below the plot using the package Gviz for R(Hahne and Ivanek, 2016) (Figure 1B).

It should be noted, for all analyses, we divide variants into two groups – congruous (those with the same direction of effect on gene expression and the GWAS trait (e.g., both negative)) in blue, and incongruous (those with opposite directions of effect on gene expression and the GWAS trait (e.g., one negative, one positive)), in red. The division between congruous and incongruous variants provides a more nuanced view of the relationship between gene expression level and GWAS associations – a variant associated with increased expression of a candidate gene and increased risk for a given GWAS trait would seem to be operating through different mechanisms

that a variant that is similarly associated with increased expression of the same candidate gene, but a decreased risk for the same GWAS trait.

eQTL Enrichment Plot

For variants within the LOI with p_{trait} meeting the specified GWAS significance threshold, the proportion that are also eQTLs (p_{eQTL} < specified eQTL significance threshold) for the gene of interest are calculated and plotted, and the same is done for variants with p_{trait} larger than the specified GWAS significance threshold (Figure 1C). Enrichment of eQTLs among GWAS-significant variants is determined by Fisher's exact test. Variants with congruent and incongruent directions of effect, as described above, are considered separately.

P-P Correlation Plot

Each variant within the LOI is plotted with p_{eQTL} along the horizontal axis, and p_{trait} along the vertical axis. Correlation between the two probabilities is visualized by plotting a best-fit linear regression over the points, with the line equation displayed on the plot. The Pearson correlation coefficient and p-value of correlation are computed and displayed on the plot as well. Separate plots are made and superimposed over each other for variants with congruent and incongruent directions of effect, as described above.

CONCLUSIONS

eQTpLot provides a unique, user-friendly, and intuitive means of visualizing eQTL and GWAS signal colocalization in a single figure. As plotted by eQTpLot, colocalization between GWAS and eQTL data for a given gene-trait pair is immediately visually obvious, and can be compared across candidate genes to quickly generate hypotheses about the underlying causal mechanisms driving GWAS association peaks. We have made eQTpLot freely available along with a brief tutorial describing its implementation. We believe eQTpLot will prove a useful tool for investigators seeking a convenient and customizable visualization of genomic data colocalization.

FUNDING

TGD is supported in part by the NIH T32 training grant 5T32GM008638-23. MDR is supported in part by NIH R01 AI077505.

REFERENCES

Giambartolomei,C. *et al.* (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics*, **10**, e1004383.

GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.

Hahne,F. and Ivanek,R. (2016) Visualizing Genomic Data Using Gviz and Bioconductor. In, Mathé,E. and Davis,S. (eds), *Statistical Genomics: Methods and Protocols*, Methods in Molecular Biology. Springer, New York, NY, pp. 335–351.

He,X. *et al.* (2013) Sherlock: Detecting Gene-Disease Associations by Matching Patterns of Expression QTL and GWAS. *The American Journal of Human Genetics*, **92**, 667–680.

Hormozdiari,F. *et al.* (2016) Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.*, **99**, 1245–1260.

Liu,B. *et al.* (2019) Abundant associations with gene expression complicate GWAS follow-up. *Nature Genetics*, **51**, 768–769.

Nica,A.C. *et al.* (2010) Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. *PLoS Genet*, **6**.

Yao,D.W. *et al.* (2020) Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.*, **52**, 626–633.

Zhu,Z. *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics*, **48**, 481–487.

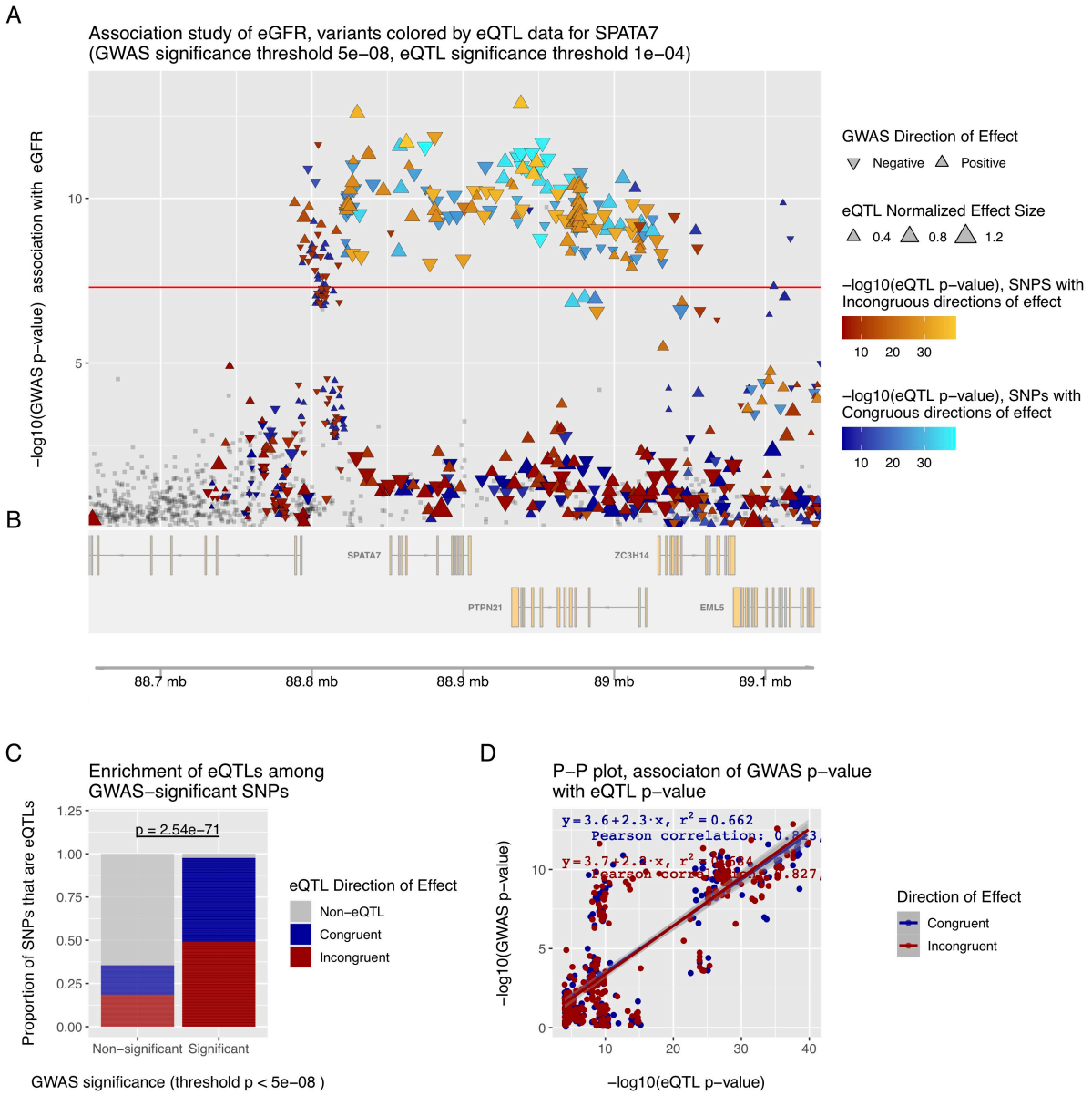

Figure 1. Example eQTpLot

An example eQTpLot illustrating A. colocalization between the eQTL signal for *SPATA7* (PanTissue analysis) and a GWAS significance peak for estimated glomerular filtration rate (eGFR). B. Map of genes present in the designated genomic range. C. Plot demonstrating significant enrichment of *SPATA7* eQTLs among GWAS-significant SNPs. D. P-P plot illustrating significant correlation between p_{eQTL} and p_{trait} for *SPATA7* and eGFR.