Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses

Gabriel Cuellar-Partida[1*], Mischa Lundberg[1], Pik Fang Kho[2], Shannon D'Urso[1], Luis F. Gutierrez-Mondragon[3], Liang-Dar Hwang[1]

1. The University of Queensland Diamantina Institute, Brisbane, Queensland, Australia

2. QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

3. Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

*Corresponding Author

Abstract

Background: Genome-wide association studies (GWAS) are an important method for mapping genetic variation underlying complex traits and diseases. Tools to visualize, annotate and analyse results from these studies can be used to generate hypotheses about the molecular mechanisms underlying the associations.

Findings: The Complex-Traits Genetics Virtual Lab (CTG-VL) integrates over a thousand publicly-available GWAS summary statistics, a suite of analysis tools, visualization functions and diverse data sets for genomic annotations. CTG-VL also makes available results from gene, pathway and tissue-based analyses from over 1,500 complex-traits allowing to assess pleiotropy not only at the genetic variant level but also at the gene, pathway and tissue levels. In this manuscript, we showcase the platform by analysing GWAS summary statistics of mood swings derived from UK Biobank. Using analysis tools in CTG-VL we highlight hippocampus as a potential tissue involved in mood swings, and that pathways including neuron apoptotic process may underlie the genetic associations. Further, we report a negative genetic correlation with educational attainment rG = -0.41 ± 0.018 and a potential causal effect of BMI on mood swings OR = 1.01 (95% CI = 1.00–1.02). Using CTG-VL's database, we show that pathways and tissues associated with mood swings are also associated with neurological traits including reaction time and neuroticism, as well as traits such age at menopause and age at first live birth.

Conclusions: CTG-VL is a platform with the most complete set of tools to carry out post-GWAS analyses. The CTG-VL is freely available at https://genoma.io as an online web application.

Keywords

GWAS, Web application, Complex-Traits, Genetics, Bioinformatics

Background

Genome-wide association studies (GWAS) have revolutionized the genetics research of complex traits and diseases over the last decade [1, 2]. As these studies become more common thanks to initiatives such as UK Biobank [3] and international genetics research consortia, there is growing need for improved data sharing and more accessible large-scale collaborative genomic analyses tools for users across different technical backgrounds.

To this end, we present the first release of the Complex-Traits Genetics Virtual Lab (CTG-VL), a web platform to annotate, analyse and share GWAS and post-GWAS results. The CTG-VL complements already widely used web applications including LD-Hub [4], FUMA-GWAS [5] and MR-Base [6] by incorporating functions available in these platforms in addition to many others. In this manuscript, we introduce the first public release of the CTG-VL, compare it with other platforms and apply some of its functions to GWAS summary statistics of mood swings derived from UK Biobank.

Findings

Visualization and annotations

The current release of the CTG-VL facilitates the visual inspection of GWAS results through LocusTrack (Figure 1A) [7] and Manhattan plots (Figure 1B). These visualizations are fully interactive and allow users to annotate and query results with a single click. Specifically, we incorporated epigenetic data from ENCODE [8] and Roadmap epigenomics projects [9], eQTL data from the GTEx project [10], GWAS catalogue [11] and polyphen [12] for annotation. An example of an annotated LocusTrack plot is shown in Figure 1A. Utilities such as obtaining the closest genes for a given genetic variant or CpG, obtaining gene information, SNPs in linkage disequilibrium (LD) and population frequencies are also actionable through a single click. The CTG-VL also provides other graphing/plotting functions, including heatmaps, karyograms and network visualizations (Figure 1C). For the network-visualization, we have incorporated functions to analyse the topology of networks including density, degree distribution, mean degree, number of edges of the giant component as well as functions to estimate shortest paths between nodes.

Analysis tools

The CTG-VL aims to facilitate downstream analyses of GWAS summary statistics. As such, we are actively integrating tools and data that will allow the research community to speed up their analyses. The current release of the platform (0.31-alpha) includes some of the newest and most commonly used analysis tools for downstream analyses of GWAS results. Briefly, we have integrated LD-score regression [13, 14] to estimate the heritability of traits and genetic correlation between traits using GWAS summary statistics. Additionally, the user can check genetic correlation against any of the

(over 1500) traits with publicly available GWAS summary statistics integrated in the platform. The platform also implements Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT) [15] that prioritizes the most likely causal genes at associated loci, and identifies enriched pathways and tissues/cell types that may underlie the associations. Another popular tool incorporated in CTG-VL is MetaXcan [16] which obtains gene-trait association by testing if the predicted expression levels (based on predicting models derived from eQTL data of selected tissues) of particular genes underlie the associations. Similarly, we integrated fastBAT [17], a fast set-based association analysis that uses GWAS summary statistics and a linkage-disequilibrium reference to summarize genetic associations with a trait of interest at the gene level. Finally, we have included the Summary-data-based Mendelian Randomization (SMR) [18] to test the causal effect of the expression level of a gene on a trait of interest, and the Generalized Summary-data-based Mendelian Randomization (GSMR) [19] to test the causal relationship between two traits.

Data

The CTG-VL aims to function as an aggregator of GWAS summary statistics and post-GWAS analyses results. As such, it complements the recent release of GWAS-ATLAS [20], a titanic effort that integrates results from post-GWAS analyses (gene-based and gene-enrichment analyses results from MAGMA [21] as well as genetic correlations and heritability derived from LD-score and SumHer [22]) on over 3500 complex traits in a browsable platform. The CTG-VL has made available post-GWAS results (DEPICT and MetaXcan) of over 1,500 traits, and we are actively updating this. Furthermore, the CTG-VL allows users to upload their own GWAS summary statistics or to link the over 1,500 publicly available GWAS summary statistics in the platform to their user profile where they can perform further analyses.

Web applications ecosystem

Comparing CTG-VL with currently available web platforms can be challenging as each web platform facilitates specific tasks and is optimized for particular pipelines/analysis methods. While all these web platforms are aggregators of GWAS summary statistics, their functions are largely segregated. CTG-VL is a great complement to web platforms already available, by integrating their key analysis methods into a single web platform. We briefly describe the differences and similarities between CTG-VL and some of these widely used web tools (i.e. FUMA-GWAS, LD-Hub and MR-Base) in Table 1.

Table 1. Some of the differences and similarities between CTG-VL, FUMA-GWAS, LD-HUB and MR-Base.

	CTG-VL	FUMA-GWAS	LD-Hub	MR-Base
Aggregator of GWAS summary statistics	Yes	Yes	Yes	Yes
Visualization	All visualizations are interactive: Manhattan plot Regional plot Heatmaps Network plots Karyograms	Manhattan plot Regional plot (interactive) Heatmap (only for gene expression)	No	No
Data for annotation	ENCODE Roadmap Epigenomics GTEx Polyphen	ENCODE Roadmap Epigenomics GTEx Polyphen Multiple others (http://fuma.ctglab.nl/links)	No	No
Heritability	LD-score regression	No	LD-score regression	No
Genetic correlations	LD-score regression The users choose traits of interest from available GWASs in the platform (over 1500 complex traits) or their own GWAS dataset.	No	LD-score regression Automated to run against over 1000 complex traits and diseases simultaneously.	No
Gene-based analysis	MetaXcan, fastBAT, SMR, DEPIC gene-prioritization	MAGMA	No	No
Tissue specificity analysis	DEPICT	MAGMA	No	No
Gene-enrichment analysis	DEPICT	MAGMA	No	No
Mendelian Randomization	GSMR, SMR	No	No	Over 10 different statistical methods to perform Two-sample mendelian randomization (excluding SMR and GSMR).

Case study

In this section, we present a series of analyses that can be performed in the CTG-VL. As an example, we extract the GWAS summary statistics for mood swings derived from UK Biobank [23]. This dataset (in addition to over 1500 others) is available in the CTG-VL. In this example, we visualize the GWAS results and estimate its genetic correlation with Neuroticism score and Educational Attainment [24]. We then identify candidate genes, pathways and tissues/cell types that are most likely to underlie the genetic associations and use the CTG-VL's large database to identify what other complex-traits are also associated with those genes, pathways and tissues/cell types. Finally, we assess the causal role of body mass index (BMI) in mood swings. A diagram showing the steps to

complete these tasks inside the CTG-VL is displayed in Figure 2. Supplementary Figure 1-8 display screen shots of each step.

The first step is to obtain the GWAS summary statistics for mood swings, neuroticism score, educational attainment and BMI. To achieve this, we go into the platform, and after signing in we go to "Data -> Public data". Here, we look for those traits and click "Add to my data". We then go to "Visualization -> Manhattan plot". The CTG-VL implements an interactive Manhattan plot that facilitates the annotation of the SNPs displayed by clicking on them. In addition, users can highlight all the SNPs that are in LD with the lead-SNP of each locus. Figure 3 shows the Manhattan plot of mood swings with 3 SNPs labelled.

In the "Analysis" tab we then select LD-score, where we estimate the genetic correlation between mood swings and neuroticism score to be rG = 0.87 (s.e. = 0.009). Such high genetic correlation is not surprising since neuroticism score is calculated based on 12 questions that include whether an individual experiences mood swings. More interestingly, using the same steps, we estimate a negative genetic correlation between educational attainment and mood swings rG = -0.41 (s.e. = 0.018).

Thereafter, we run DEPICT under the tab "Analyses" to obtain the most likely tissues/cell types and biological pathways that may underlie the genetic associations. Tables 2 and 3 summarize the top 5 results of each of these analyses. Supplementary Tables 1 and 2 show all biological pathways, other gene sets and tissues with an FDR < 5%.

Table 2. Top 5 tissues and cell types likely to underlie the associations.

Tissue	Group	P-value	FDR<5%
Hippocampus	Nervous System	0.00000211	Yes
Retina	Sense Organs	0.00000378	Yes
Limbic System	Nervous System	0.00000884	Yes
Cerebral Cortex	Nervous System	0.0000136	Yes
Cerebrum	Nervous System	0.0000173	Yes

Table 3. Top 5 pathways from Gene-Ontology likely to underlie the associations.

Pathway ID	Pathway Description	P-value	FDR<5%
GO:0043523	regulation of neuron apoptotic process	0.0000125	Yes
GO:0051402	neuron apoptotic process	0.0000173	Yes
GO:0070997	neuron death	0.0000209	Yes
GO:0017016	Ras GTPase binding	0.0000370	Yes
GO:0008017	microtubule binding	0.0000565	Yes

By using the "check overlap" function in the platform and the big database of post-GWAS analyses (tissue, pathway and gene-based test) on over 1,500 complex traits we then assess in what other traits and diseases the associated tissues and pathways are also involved. We observe that tissues/cell types with an FDR < 5% for mood swings are also associated to traits such as neuroticism score, age at first birth, anxious feelings and reaction time (Supplementary Table 3). Similarly, biological pathways with an FDR < 5% for mood swings are also associated to neuroticism, reaction time, age at menopause and hand grip strength among many others (Supplementary Table 4).

We then run MetaXcan and SMR analyses under the tab "Analysis" using gene-expression prediction models and eQTL data derived from the hippocampus (the strongest tissue association derived from DEPICT as shown in Table 2) to prioritize genes that are likely to be associated with mood swings. We examine 8440 genes through MetaXcan and 1161 through SMR. Top 5 results for MetaXcan and SMR are displayed in Tables 4 and 5 respectively. Supplementary Tables 5 and 6 show all genes that reach statistical significance based on Bonferroni correction (P < 6e-6 for MetaXcan and 4e-5 for SMR).

Table 4. Top 5 results from MetaXcan using gene-expression prediction models for the hippocampus.

Gene ID	Gene Name	Z-score	P-value	#SNPs used in model
ENSG00000238083	*LRRC37A2*	-7.789	6.73e-15	38
ENSG00000196628	*TCF4*	5.47	4.51e-8	8
ENSG00000213619	*NDUFS3*	5.192	2.08e-7	17
ENSG00000171044	*XKR6*	-5.152	2.57e-7	11
ENSG00000205882	*DEFB134*	5.122	3.03e-7	4

Table 5. Top 5 results from SMR using eQTL data from hippocampus.

Gene ID	Gene Name	Stronger eQTL	Beta	P-value	HEIDI P-value*	SNPs in HEIDI test
ENSG00000214425	*LRRC37A4P*	rs62063676	-0.01	1.10e-10		
ENSG00000214401	*KANSL1-AS1*	rs55974014	0.01	1.35e-10		
ENSG00000238083	*LRRC37A2*	rs2040845	0.01	7.47e-10		
ENSG00000263503	*RP11-707O23.5*	rs55974014	0.01	1.54e-9		
ENSG00000264070	*DND1P1*	rs55974014	0.01	1.94e-9	0.033	4

*HEIDI (HEterogeneity In Depedent Instruments) test P-value. HEIDI test is only performed if the number of eQTLs for a specific gene is more than 3.

Finally, using GSMR, we assess the causal role of BMI on the risk of mood swings. The results suggest a potential causal effect of BMI on mood swings, with each unit (kg/m^2) higher in BMI leading to a higher risk of mood swings (OR = 1.01, s.e. = 0.006). However, the HEIDI-outlier test result (P-value =

0.053) suggests potential pleiotropic effects of the genetic variants used in the test on both exposure (BMI) and the outcome (mood swings).

Future developments

We update the CTG-VL on a weekly basis with new data, utilities and analysis tools. At this stage (CTG-VL version alpha-0.31), we are collecting feedback from the research community to improve the performance of the platform. In the suture, we aim to encapsulate a more stable version of the CTG-VL as Machine Image / Docker container to deploy in any cloud provider.

Remarks and Conclusion

We present the first release of the CTG-VL. This platform integrates the most comprehensive set of post-GWAS analysis tools available in other web-based applications such as LDHub [4] and FUMA-GWAS [5] in addition to many others into a single platform, allowing users to test and derive novel hypothesis out of GWAS summary statistics. A catalogue of GWAS and post-GWAS analysis results on over 1,500 complex traits in CTG-VL provides users the ability to explore the pleiotropy at gene-, pathway- and tissue- level as shown in our model of mood swings, which is currently not available in any other web platforms. This is of great utility for the interpretation of genetic correlations. For example, LD-score regression (also implemented in the platform) can be used to assess the extent to which two traits are affected by the same SNPs in the same (positive genetic correlation) or opposite (negative genetic correlation) direction. However, this analysis is unable to implicate the genes and biological pathways that may underlie the correlation. To aid this endeavor researchers can use the "check overlap" function in the platform to identify shared genes, tissues and pathways, leading to novel hypothesis and interpretation of the genetic correlations.

This work was inspired by big efforts from multiple research groups making phenome-wide GWAS summary statistics available (i.e. Neale's lab at the Broad Institute, GeneAtlas [20], and those GWAS summary statistics released along with the SAIGE software [25]) that await to be analyzed further, and more importantly, interpreted.

The CTG-VL intends to be a user-friendly platform that will allow research teams with different expertise to perform common post-GWAS analyses and to aid in the interpretation of this huge amount of data. The CTG-VL aims to be a community-driven platform where novel analysis tools and data are regularly incorporated as these become available. As such, this report aims to gauge interest from the complex-traits genetics research community.

Methods

Implementation

The web interface and server are written in Angular 6 and Node.js respectively. Scripts to analyze networks were made with python (2.7) using NetworkX (1.8.1) and Igraph (0.7.1) libraries. Tabix [26] is used to index and query user GWAS data. Graphics were made with in-house d3.js v5 scripts. Genomic annotations and user data are stored at a server in the Queensland Research and Innovation Services Cloud. We use Google's firebase for authentication and meta-data storage. All communications between the CTG-VL platform back- and front-end are encrypted through SSL (Secure Sockets Layer).

Analysis software

The CTG-VL uses PLINK [27] to clump and calculate LD between SNPs. These two functions are mainly used to obtain LD data for the regional plots (LocusTrack), identify SNPs in LD with lead SNPs (those with an association P-value <5e-8) and identify independent loci to be used by DEPICT.

MetaXcan [16] is implemented in the platform with the default options along with gene expression prediction models from PredictDB [15] to carry out gene-based analyses. LD-score regression [13, 14] is implemented along with LD-scores derived from 1000 Genomes [14] to estimate the heritability and genetic correlation based on GWAS summary statistics. Prior to LD-score analyses, GWAS summary statistics are run through the munge.py utility and merged to HapMap SNPs to ensure that alleles match between different datasets.

DEPICT [15] is implemented to perform gene prioritization, gene-enrichment and identify the tissue/cell type most likely to underlie the associations. DEPICT takes as input independently associated SNPs below a user-specified P-value threshold.

SMR [18] is implemented along with eQTL data for 48 cell types derived from GTEx and downloaded from the SMR website [28]. GSMR [19] is a function in GCTA [29] and is implemented with the default parameters.

Each of the individual commands to run the analyses are included in the platform.

Data

Currently, genotypes from 1000 Genomes phase 3 [30] are used throughout the platform as a reference for LD. eQTL data from the GTEx Project [10] is used for annotation in regional plots, SMR and MetaXcan. Chromatin states [31] derived from ENCODE [8] and Roadmap Epigenomics Mapping Consortium [9] are used to annotate regional plots. Genes and SNPs position information is based on

the GRCh37 human genome assembly. Descriptions of genes are extracted from RefSeq [32] and UniProt [33].

GWAS summary statistics integrated within the CTG-VL platform are updated regularly, and both the references and links can be found in the platform [34]. GWAS summary statistics are integrated as published by the authors with exception to the summary statistics derived from Neale lab [23] where we removed SNPs with a MAF < 0.001 or Minor Allele Count < 25 in Cases (for case-control GWAS) as recommended in their documentation.

Post-GWAS analyses in over 1500 traits

We downloaded 3670 GWAS summary statistics from Neale Lab [23] and removed variants as described in the previous sections. We ran LD-score regression against each of these GWAS summary statistics. Those traits with a statistically significant heritability of P-value <0.05 (1,747) were selected to be analyzed with DEPICT and MetaXcan. As a heuristic, we ran MetaXcan using whole-blood gene-expression prediction models as these models were derived from a larger sample compared to other tissues. To run DEPICT, we clumped the summary statistics with the recommended parameters (LD-window 500Kb, and R^2 < 0.05) and used SNPs with the recommended P-value threshold of <1e-5. For traits with less than 10 independent loci, we relaxed the threshold to P-value <1e-4.

Updates

The CTG-VL is updated on a weekly basis, and this short report may not list/reference all the tools and data in the current version. However, this information can be found on the main page of the platform.

Availability of supporting source code and requirements

Project name: Complex-Traits Genetics Virtual Lab

Project home page: https://genoma.io

Operating system(s): Platform independent

Programming language: JavaScript, Python, R, C

Other requirements: Google Chrome version 65 or above; Firefox version 60 or above.

License: The CTG-VL home page contain links to each analysis tool repository with license information.

Acknowledgements

GCP is supported by an Australian Research Council Discovery Early Career Research Award [DE180100976]. ML is supported by a Research Training Program Scholarship from The University of

Queensland. PFK is supported by a Research Training Program Scholarship from Queensland University of Technology and QIMR Berghofer Medical Research Institute.

Conflict of interest

All authors declare that there are no conflicts of interest.

Contributions

GCP designed and developed the platform, implemented workflows, built databases and ran analyses. ML aggregated GWAS summary statistics, implemented workflows and ran analyses. PFK implemented workflows. SDU ran analyses. LFGM implemented workflows. LDH implemented workflows and build databases. Everyone wrote and review this manuscript.

References

1. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation.* Am J Hum Genet, 2017. **101**(1): p. 5-22.
2. Cloney, R., *Complex traits: Integrating gene variation and expression to understand complex traits.* Nat Rev Genet, 2016. **17**(4): p. 194.
3. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.
4. Zheng, J., et al., *LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis.* Bioinformatics, 2017. **33**(2): p. 272-279.
5. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nat Commun, 2017. **8**(1): p. 1826.
6. Hemani, G., et al., *The MR-Base platform supports systematic causal inference across the human phenome.* Elife, 2018. **7**.
7. Cuellar-Partida, G., M.E. Renteria, and S. MacGregor, *LocusTrack: Integrated visualization of GWAS results and genomic annotation.* Source Code Biol Med, 2015. **10**: p. 1.
8. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
9. Bernstein, B.E., et al., *The NIH Roadmap Epigenomics Mapping Consortium.* Nat Biotechnol, 2010. **28**(10): p. 1045-8.
10. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.
11. MacArthur, J., et al., *The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).* Nucleic Acids Res, 2017. **45**(D1): p. D896-D901.
12. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Curr Protoc Hum Genet, 2013. **Chapter 7**: p. Unit7 20.
13. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nat Genet, 2015. **47**(11): p. 1236-41.
14. Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.* Nat Genet, 2015. **47**(3): p. 291-5.
15. Pers, T.H., et al., *Biological interpretation of genome-wide association studies using predicted gene functions.* Nat Commun, 2015. **6**: p. 5890.
16. Barbeira, A.N., et al., *Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics.* Nat Commun, 2018. **9**(1): p. 1825.

17. Bakshi, A., et al., *Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits.* Sci Rep, 2016. **6**: p. 32894.
18. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.* Nat Genet, 2016. **48**(5): p. 481-7.
19. Zhu, Z., et al., *Causal associations between risk factors and common diseases inferred from GWAS summary data.* Nat Commun, 2018. **9**(1): p. 224.
20. Kyoko Watanabe, S.S., Oleksandr Frei, , *A global view of pleiotropy and genetic architecture in complex traits.* bioRxiv, 2018.
21. de Leeuw, C.A., et al., *MAGMA: generalized gene-set analysis of GWAS data.* PLoS Comput Biol, 2015. **11**(4): p. e1004219.
22. Speed, D. and D.J. Balding, *SumHer better estimates the SNP heritability of complex traits from summary statistics.* Nat Genet, 2018.
23. Lab, N., *UK Biobank GWAS summary statistics.* http://www.nealelab.is/uk-biobank/, 2018.
24. Lee, J.J., et al., *Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals.* Nat Genet, 2018. **50**(8): p. 1112-1121.
25. Zhou, W., et al., *Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies.* Nat Genet, 2018. **50**(9): p. 1335-1341.
26. Li, H., *Tabix: fast retrieval of sequence features from generic TAB-delimited files.* Bioinformatics, 2011. **27**(5): p. 718-9.
27. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.
28. *SMR eQTL datasets.* http://cnsgenomics.com/software/smr/#DataResource.
29. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* Am J Hum Genet, 2011. **88**(1): p. 76-82.
30. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
31. Ernst, J. and M. Kellis, *Chromatin-state discovery and genome annotation with ChromHMM.* Nat Protoc, 2017. **12**(12): p. 2478-2492.
32. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.
33. UniProt, C., *The universal protein resource (UniProt).* Nucleic Acids Res, 2008. **36**(Database issue): p. D190-5.
34. *Complex-Traits Genetics Virtual Lab.* https://genoma.io.

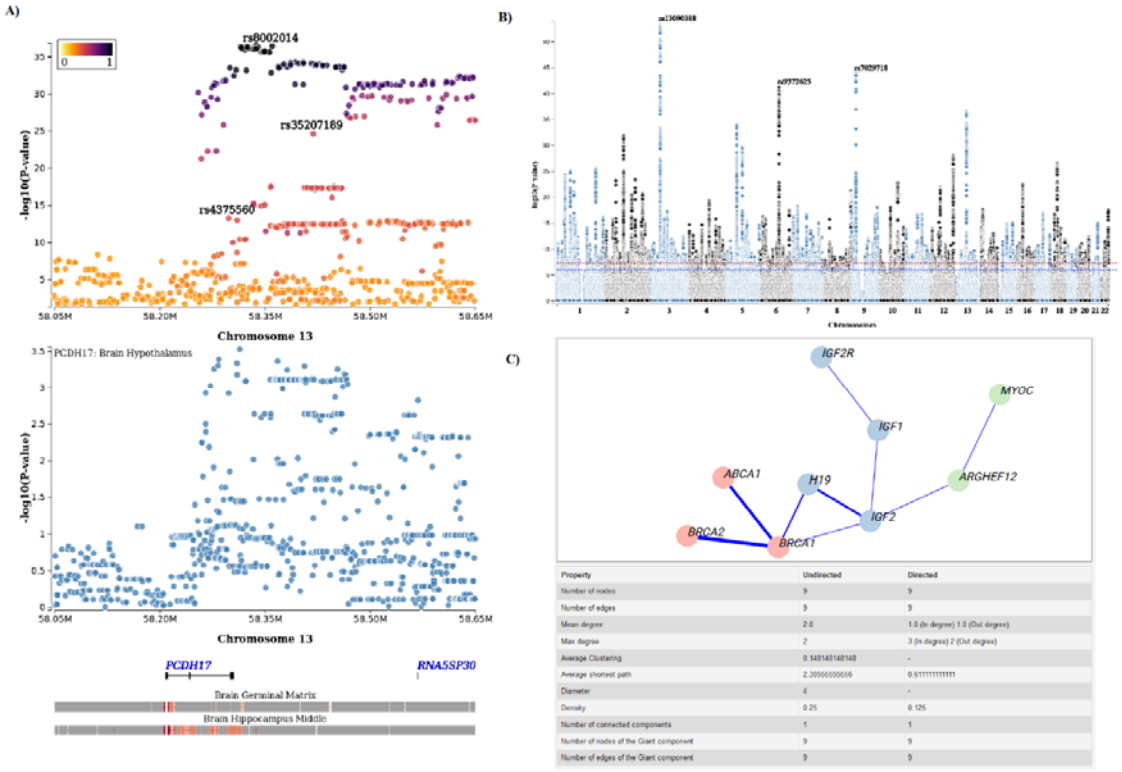

Figure 1. Examples of the visualizations available in the CTG-VL. **(A)** An example of a LocusTrack plot that displays association results within a specified region (upper panel), along with eQTLs for a selected gene (PCDH17 in this example) (middle panel) and annotation tracks such as chromatin information for selected tissues (lower panel). Both, **(A)** LocusTrack and **(B)** Manhattan plots are fully interactive, allowing the user to annotate and query SNPs by simply clicking on them. **(C)** An example of a network visualization and its properties.

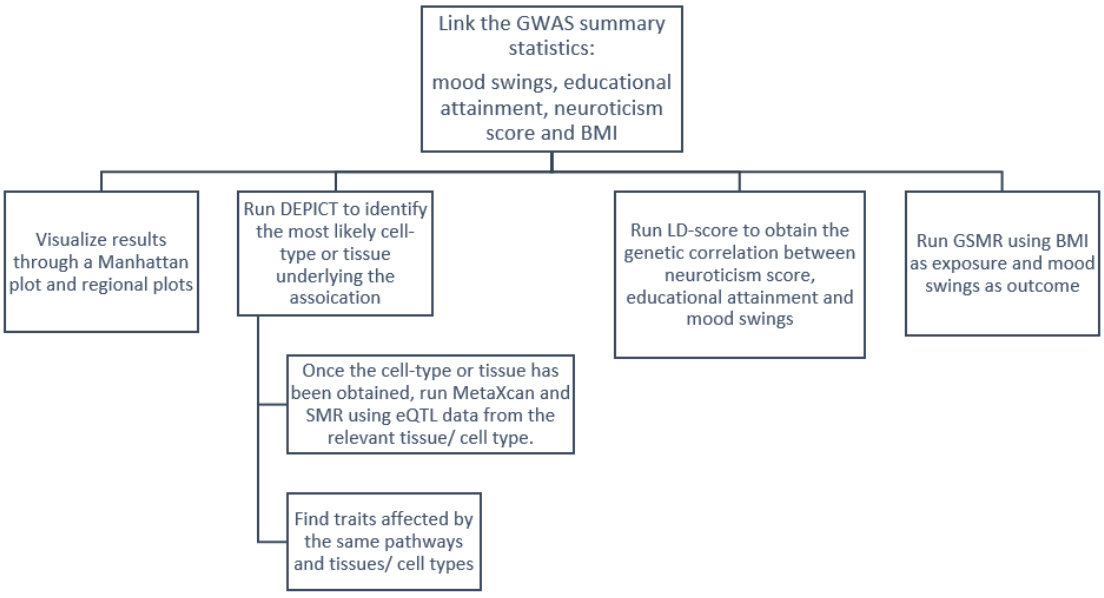

Figure 2. Diagram of the steps to analyse the mood swings GWAS summary statistics.

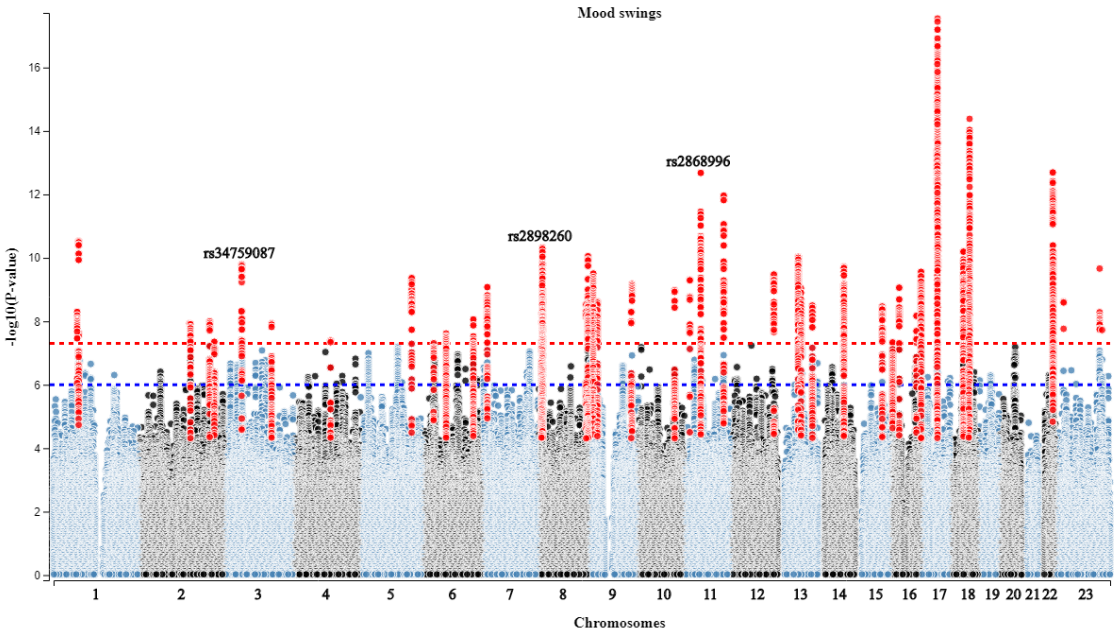

Figure 3. Manhattan plot of mood swings. The CTG-VL enables the user to click on each of the SNPs to obtain the summary statistics, annotate and query the SNP. In addition, the user can highlight SNPs in LD with the lead SNP, as can be seen in red.