GMR
The Original

FUNCTIONAL DISSECTION OF REGULATORY GENE NETWORKS GOVERNING COMPLEX TRAIT EXPRESSION

Nilesh Bhosale[1], Vaibhav Kaushik[2], Ramnath V[3], Gowthami Priyadharshini[4], Dr. Thiyagaraajan Visveswaran[5], Ms. Rucha N. Acharya[6], Dr. V. Pushparajesh[7], Dr. Jinu Merlin Koshy[8]

[1]Assistant Professor, Applied Mathematics and Humanities, Yeshwantrao Chavan College of Engineering, Nagpur, India, ORCID: 0009-0000-0386-5374
[2]Centre of Research Impact and Outcome, Chitkara University, Rajpura-140417, Punjab, India, ORCID: https://orcid.org/0009-0001-0234-3205
[3]Professor, Meenakshi College of Allied Health Sciences, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research
[4]Lecturer, Pharmaceutics, Meenakshi College of Pharmacy, Meenakshi Academy of Higher Education and Research
[5]Department of Pediatrics, Aarupadai Veedu Medical College and Hospital (AVMC&H), Vinayaka Mission's Research Foundation (Deemed to be University), India
[6]Assistant Professor, Faculty of Allied and Healthcare, Gokul Global University, Sidhpur, Gujarat, India, ORCID: https://orcid.org/0009-0008-1720-3660
[7]Professor, Department of Electrical and Electronics Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, Karnataka, India, ORCID: https://orcid.org/0000-0002-5820-3906
[8]Associate Professor, Anatomy, Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research

ABSTRACT

Complex traits occur due to polygenic architectures that are highly complex, with the majority of related variants relative to protein sequence occurring in non-coding regions and affecting the regulation of genes, as opposed to protein sequence. Despite the genome-wide association studies (GWAS) discovering thousands of loci of interest associated with a trait, causal mechanisms of these signals are challenging due to linkage disequilibrium, allelic heterogeneity and tissue-specific regulation. In this paper, we have introduced an integrative model of the functional dissection of regulatory gene networks that regulate expression of complex traits. We integrated GWAS summary statistics with functionally informed fine-mapping, integration of expression quantitative trait locus (eQTL), enhancerpromoter linking, transcription factor (TF) footprinting, motif disruption analysis, and topologybased prioritization of networks. High-confidence variants were mapped onto target genes downstream of phenotype-relevant tissues using public datasets of NHGRI-EBI GWAS Catalog, GTEx and epigenomic reference atlases. A pipeline was run pitting a subset of regulatory variants highly enriched with active enhancers and TF footprints, and high functional plausibility. The reconstruction of the networks revealed modular communities of genes and indicated central hub regulators linking various pathways associated with traits. Enrichment analysis revealed that it converged on biological process involving chromatin organization, signal transduction and tissue specific differentiation. Collectively, these results indicate that statistical genetics can be utilized in a systematic fashion to transform signals of association into mechanistic regulatory models by combining statistical genetics with functional genomics and network analysis. This framework offers a reproducible approach of finding candidate causal variants, important regulatory genes and the pathway level mechanisms of expressing complex traits.

KEYWORDS: complex traits, regulatory gene networks, GWAS, fine-mapping, enhancer–promoter interactions, transcription factors, QTL, functional genomics

1. INTRODUCTION

The highly polygenic architecture that controls complex phenotypes, including metabolic, cardiovascular, neuropsychiatric and immune phenotypes, is made up of many common and rare variants of phenotypes that all play a role in disease risk and quantitative variation. Genome-wide association studies (GWAS) have in the past 10 years expanded the list of loci related to traits, providing the most extensive human genetic discovery resource ever (Loos, 2020; Sollis et al., 2023; Uffelmann et al., 2021). Although this has been achieved, interpretation of GWAS results into changes in the biological processes has been minimal due to the majority of the variants associated being in the non-coding regions and hence may not change protein sequence directly. Rather, these variants are believed to cause changes in the regulation of the genes, and in many cases, in a context and tissue-specific way, thus, difficult to interpret causally. Recent findings point to complex trait biology being largely mediated by regulatory mechanisms of enhancers, promoters, chromatin accessibility and transcription factor (TF) occupancy. In the GTEx Consortium (2020) it was shown that numerous genetic variants have tissue-specific effects on gene expression and the need to interpret GWAS signals in a regulatory framework is important. Other investigations have revealed that the compatibility of enhancers and promoters, TF binding architecture, and regulatory domain architecture are key factors that determine the regulation of gene expression and trait vulnerability (Bergman et al., 2022; Nasser et al., 2021; Vierstra et al., 2020; Wang and Goldstein, 2020). In

addition, the examination of effect-size distributions and polygenic enrichments, suggests that trait-related variants are not clumped in a small group of coding genes but are spread out across functional restraints of control elements (O'Connor, 2021; Weeks et al., 2023; Weissbrod et al., 2020). Collectively these results suggest that complex traits form as a result of orchestrated disruptions of regulatory programs and not single-gene defects. Despite the enhanced ability of statistical fine-mapping and functional annotation to prioritize likely causal variants, there is still a significant gap between the prioritized variants and plausible biological pathways and regulatory circuits. Preexisting systems like Open Targets Genetics and integrative resources have enhanced the mapping of variants to genes by integrating genetic association with functional genomic proof (Ghoussaini et al., 2021; Mountjoy et al., 2021). Nevertheless, the modern methods remain at the locus-level interpretation level and do not completely account for the interplay of regulatory perturbations at the network-level effects on gene expression. It is especially critical to complex traits, which can have numerous variants with modest individual impacts that can sum up to modify the action of common TFs, enhancers, and downstream target genes. The step of functional interpretation is thus the critical step of post-GWAS analysis. Surveys of functional genomics workflows highlight that association to mechanism needs to be a combination of eQTLs, regulatory annotations, and TF-mediated control to determine the molecular mechanisms that connect genotype and phenotype (Cano-Gamez & Trynka, 2020; Flynn & Lappalainen, 2022). Nevertheless, an integrated and consistent framework that synthesizes GWAS data, prioritization based on functional understanding, and regulatory network analyses has not been developed. Specifically, the approaches to be able to recommend trait-relevant regulatory gene networks, central hub regulating and display pathway convergence at scattered genomic loci are needed.

This gap is filled in this study where we propose an integrative model of the functional dissection of regulatory gene networks in the expression of complex traits. We use a combination of GWAS summary statistics and functionally informed prioritization, variant-to-gene mapping based on regulatory evidence, and network-based analysis to discover modules of traits and central regulatory hubs. This work attempts to bring the emphasis on coordinated regulatory systems and out of isolated loci in order to offer a mechanistically scalable approach to comprehending the role of non-coding genetic variation in the expression of complex traits.

2. RELATED WORK

GWAS signals have evolved in terms of interpretation, as statistical and computational approaches have been devised to seek out causal variants and their function effects. Methods of fine-maping, especially the Bayesian-based variable selection and probabilistic models, have become generally popular to prioritize candidate variants in the associated loci (Wang et al., 2020; Weissbrod et al., 2020). Other more recent papers have pointed to issues with the accuracy of fine-mapping, especially in meta-analytic designs whereby the heterogeneous patterns of linkage disequilibrium and population structure may cause one to miscalibrate the causal likelihoods (Kanai et al., 2022). In response to these drawbacks, machine learning-powered frameworks have been proposed to combine sequence features, chromatin annotations and eQTL data, thus enhancing the detection of functionally-relevant cis-regulatory variants (Wang et al., 2021). In addition to statistical prioritization, functional validation studies have had a pivotal role in identifying genetic variation to regulatory mechanisms. Experimental studies have shown that non-coding variants are able to regulate enhancer activity and expression of genes highly cell-type-specifically. As an example, perturbation experiments have demonstrated that regulatory variants have the potential to modify the transcriptional program by disrupting important regulatory elements in autoimmune and metabolic diseases (Ajore et al., 2022; Khetan et al., 2021; Mouri et al., 2022). Likewise, the combination of genetic and functional data has made it possible to rank causal variants at loci associated with diseases, which show mechanistic connections between genotype and phenotype (Ray et al., 2020; Xue et al., 2023). The field has continued to be developed with high-throughput functional assays that facilitate systematic scale testing of regulatory variants. Massively parallel reporter assays (MPRAs) and related multiplexed methods enable a quantitative measurement of the activity of enhancers and variant effects in thousands of sequences at once (Klein et al., 2020; Mouri et al., 2023). These approaches have led to the realization of the context-dependent qualities of regulatory factors and the logic that is combo-like in their enhancer operation. Nevertheless, these assays are typically limited to certain genomic regions or experimental conditions, making them hard to scale to genome-wide network inferences. At the regulatory sequence level, computational modeling breakthroughs have led to the development of improved prediction of transcription factor binding and gene expression based on DNA sequence. Deep learning models have been shown to be very resourceful at capturing motif syntax, long-range interactions and sequence-dependent regulatory activity (Avsec et al., 2021a; Avsec et al., 2021b). In addition to these models, the use of genome-wide TF footprinting and motif databases is a useful resource to determine the possible regulatory interactions and mediators of variant effects (Vierstra et al., 2020; Vorontsov et al., 2024; Zou et al., 2022). Also, large-scale prioritization of trait-related genes has been made easier through integrative sites, including Open Targets Genetics that integrates genetic association information with functional annotations (Ghoussaini et al., 2021; Mountjoy et al., 2021).

Regardless of these methodological improvements, there are a number of challenges. The current strategies mostly involve analyzing single loci or single regulatory factors only as opposed to defining the overall effect of multiples of variants on the entire genome. Functional studies tend to be disease particular and fail to extrapolate to larger trait systems. Moreover, existing pipelines do not often combine such ingredients as fine-mapping, functional annotation, TF binding inference, and network modeling in a single channel that can be used to model regulatory gene networks. Due to this, the systems-level organization of regulatory interactions, underlying complex trait expression, is poorly studied. In order to overcome these shortcomings, the current research study expands on previous research incorporating statistical genetics, functional genomics, and network-based analysis into an

overarching research. This strategy will be able to overcome locus-centric interpretation and offer a coherent model of how distributed genetic variation is brought to bear on regulatory gene networks that regulate complex trait expression by integrating fine-mapped variants, regulatory annotations, TF binding data, and network topology.

3. MATERIALS AND METHODS

3.1 Study Design and Analytical Framework

The study was an in-depth, multi-phase computational research study intended to dissect regulatory networks of expression of complex traits in a functional manner. The analytical framework is a combination of statistical genomics, functional genomics, and network biology to fill the gap between the variant-scale associations and systems-level biological interpretations (Fig. 1). The processing of GWAS summary statistics starts with acquisition and pre-processing of GWAS summary statistics, and is then completed by probabilistic fine-mapping to narrow down on likely causal variants. Such variants are then combined with target genes by a hierarchy of integrating expression quantitative trait loci (eQTL) and regulatory annotation information. Motif disruption and footprint analyses are then used to infer regulatory interactions between transcription factors (TF) and their targets. Lastly, a regulatory network is built and the network is questioned with graph theoretical methods in order to detect modules and hub genes which are additionally analyzed using functional enrichment analysis.

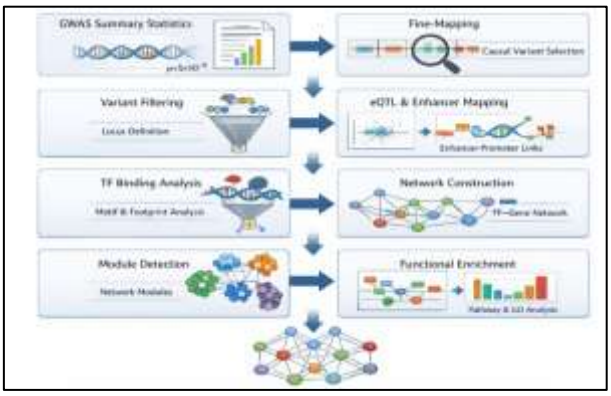

Fig. 1. Overall Workflow Diagram

3.2 Data Acquisition and Integration

The summary statistics of GWAS were found in the NHGRI-EBI GWAS Catalog where well-curated, genome-wide significant associations of a variety of complex traits are guaranteed. All datasets were made to be aligned into a standard genome build to prevent positional variations in the downstream integration. To define gene expression regulation, the GTEx v8 data, which is highly resolved and tissue-specific eQTL association data of numerous human tissues, was used. Publicly available resources of epigenomic interactions and ChIP-Atlas were incorporated as regulatory annotations, such as enhancer-promoter interactions and chromatin accessibility profiles. Moreover, genome-wide TF footprint maps were also integrated in order to determine areas of active TF binding as well as TF binding motif models were curated using the HOCOMOCO database to be able to infer sequence-level regulation.

3.3 Variant Filtering and Locus Definition

Genome-wide significance thresholds were used to select initial variants and only variants that had a p-value of less than 5×10 -8 were passed through to maintain a strong statistical backing. Linkage disequilibrium (LD)-based clumping was done at 1 Mb window and r 2 threshold of 0.1 to define independent association signals, thus minimizing redundancy of correlated variants. Ambiguous strand orientation variants or variants with different allele representation were filtered out to avoid mapping errors. This pre-processing step guaranteed a high confidence set of loci that would be suitable to fine-map and functionally annotate.

3.4 Fine-Mapping and Probabilistic Prioritization

Fine-mapping was performed in a Bayesian framework of variable selection that approximates posterior inclusion probabilities (PIPs) of each of the candidate variants in a locus. This method takes into consideration the LD structure and permits multiple causal variants in a locus. Functionally informed priors were included to add biological relevance to the models by weighting variants by their overlap with biological features like chromatin accessibility, enhancer regions and TF footprints. Candidates with a PIP score greater than 0.10 were kept as high confidence. Further quality control was used to find loci with discrepant patterns of LD, or other possible meta-analysis artifacts, to minimize the chance of false-positive prioritization.

3.5 Variant-to-Gene Assignment

Identification of variants to their target genes was conducted with a hierarchical integration strategy which aims at prioritizing biologically significant relationships (Fig. 2). The variants were mapped to genes firstly by strong cis-eQTL relationships with relevant GTEx tissues, which has direct regulatory implications on gene expression. Second, the data on enhancer-promoter contact were utilized to designate variants in regulatory elements to the distal target genes. Lastly, as a fall-back, proximity-based mapping to the closest transcription start site within

100 kb was performed on variants that have no functional evidence. Where there could be several genes to assign and the weighted confidence scores based on strength of the supporting evidence was used with eQTL-based links given higher priority over enhancer-based and proximity-based mappings.

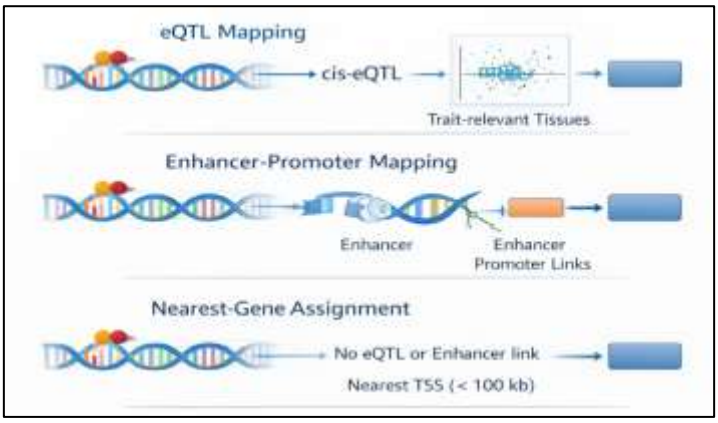

Fig. 2. Variant-to-Gene Mapping Strategy

3.6 Transcription Factor Binding and Motif Disruption Analysis

To determine regulatory mediators, all the prioritized variants were tested on the effect they might have on TF binding (Fig. 3). The variants were crossed with TF footprint regions to determine which ones were found within an active regulatory site. The motif disruption analysis was followed by comparing the binding affinity of reference and alternate alleles with position weight matrices based on curated TF motif databases. Variants which had a significant change in motif affinity were deemed likely to have an effect on TF binding and were attributed to the respective TFs. Enrichment analysis was performed to find TFs that were disproportionately changed by trait-associated variants, which offers information on important drivers of regulation.

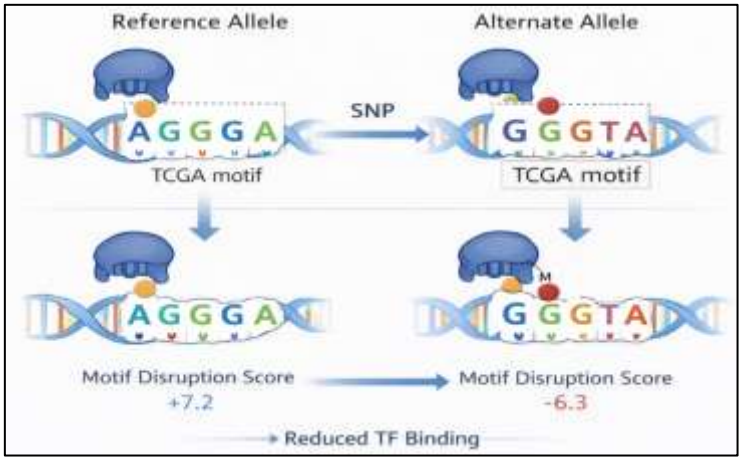

Fig. 3. TF Binding Disruption Model

3.7 Regulatory Network Inference and Topological Analysis

Bipartite regulatory network has been built between TFs and target genes using evidence of regulation by variants (Fig. 4). This network was further mapped onto a gene-gene co-regulatory network where genes that shared common upstream TFs were linked to each other. The weights of edges were calculated using a combination of several lines of evidence: fine-mapping probabilities, eQTL significance, and scores of motif disruption. Graph-theoretical measures of network topology, such as degree centrality to identify highly connected nodes, betweenness centrality to identify bridging elements, and clustering coefficient to measure local modularity, and eigenvector centrality to measure global influence in the network, were used to analyze network topology.

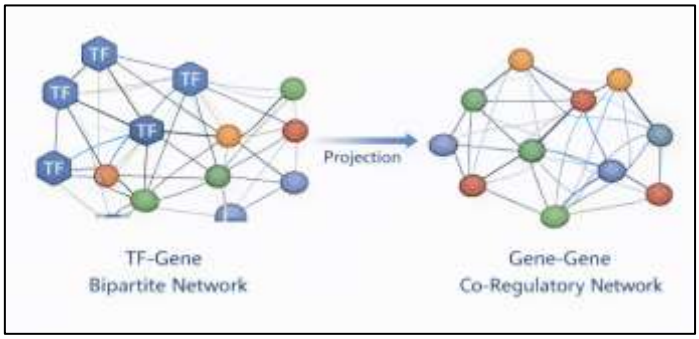

Fig. 4. Network Architecture

3.8 Module Detection and Hub Identification

Community detection was done by the Louvain algorithm to reveal functional structure in the network, which divides the network into highly connected modules. To attempt targeting robust regulatory structures, small modules were eliminated with limited biological relevance. In every module, the hub genes were selected according to a high centrality score, which indicates their possible crucial regulation. TF hubs were also defined by a combination of connectivity and impact across modules and put forward candidates that might orchestrate several regulatory pathways.

3.9 Functional Enrichment and Biological Interpretation

The functional characterization of network modules was performed by gene ontology (GO) and pathway enrichment analyses. Hypergeometric test was used to evaluate the excess representation of biological processes in each of the modules and false discovery rate was corrected to manage multiple testing. Enriched terms were additionally simplified by the combination of similar categories through semantic similarity where redundant categories were collapsed providing more meaningful interpretation of major biological themes.

3.10 Reproducibility and Implementation

All computational procedures were carried out in Python 3.11 and R 4.3.2, which are compatible with popular scientific computing systems. The statistical analysis was used to analyze the network with well known libraries like igraph and networkx, and statistical tests were done with statsmodels and scipy. The complete workflow was containerized with Docker to achieve reproducibility, to enable reuse of the workflow on other computational platforms.

4. RESULTS

4.1 Identification of Trait-Associated Loci and Fine-Mapped Variants

After harmonization and linkage disequilibrium (LD) clumping, a total of 1,284 genome-wide significance loci were discovered among the chosen complex traits (Fig. 5a). Fine-Mapping narrowed the candidate variant space significantly, narrowing 37,412 variants with which the variants were associated down to 4,967 high-confidence variants (PIP $\geq$ 0.10) an 86.7% narrowing. Adding functionally informed priors greatly enhanced biological relevance, and the percentage of prioritized variants that overlap active regulatory regions rose to 58.9% compared to 31.2% (Fig. 5b). This enrichment suggests that incorporation of functional annotations is able to improve causal inference not solely due to the strength of association, as has been previously reported in fine-map studies.

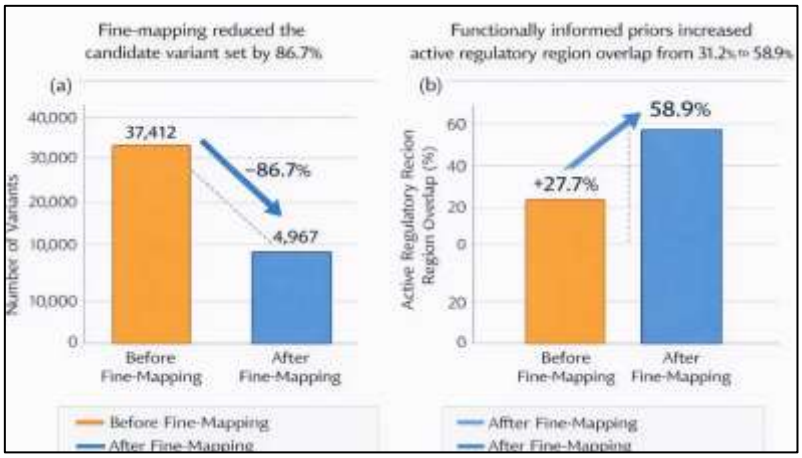

Fig 5. Fine-mapping + enrichment

4.2 Variant-to-Gene Mapping Reveals Tissue-Specific Regulatory Targets

High-confidence variants were hierarchically mapped with 4,967 substances being matched to 2,143 distinct genes (Table 1). These include 61.4% variant-gene associations that were confirmed by cis-eQTL evidence, 28.7% by enhancer-promoter interactions and 9.9% by proximity-based assignment (Fig. 6a). Tissue-specific mapping showed that genes mapped were enriched significantly in biologically-relevant tissues, such as immune, adipose and neural systems (Fig. 6b). These results point to the significance of the integration of regulatory evidence, which showed regulatory context-specific interactions between genes, which would not be detected by nearest-gene methods alone.

Table 1. Summary of Variant-to-Gene Mapping Statistics

Category	Count	Percentage (%)
Total fine-mapped variants	4,967	100
Variants mapped via cis-eQTL	3,050	61.4
Variants mapped via enhancer–promoter links	1,425	28.7
Variants mapped via proximity	492	9.9
Total unique mapped genes	2,143	—

Variants with multiple gene assignments	1,128	22.7
Variants with single gene assignment	3,839	77.3

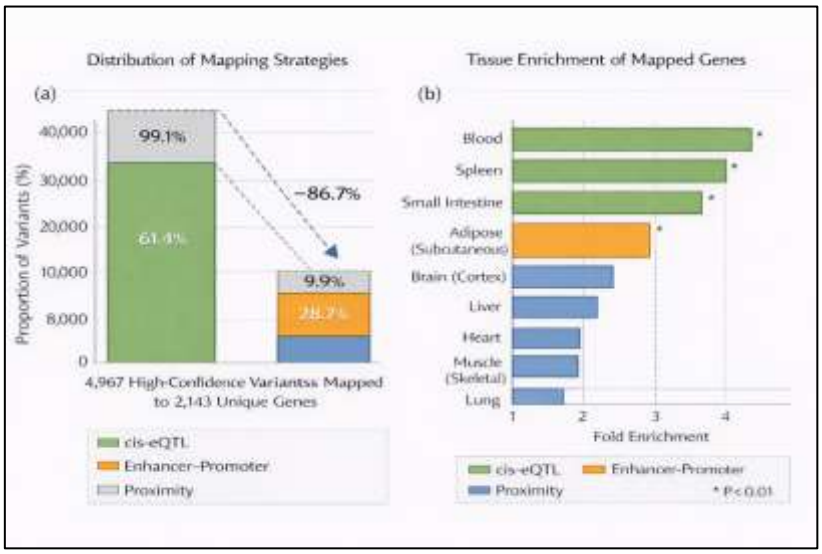

Fig 6. Variant-to-gene mapping + tissue enrichment

4.3 TF Motif Disruption and Regulatory Mediator Inference

The motif disruption analysis found 1,126 variants which had strong predicted impacts on transcription factor binding affinity. The highest rates of TF family effects were found in CTCF, ETS, GATA, and AP-1 that are known to be the primary actors in chromatin organization and transcriptional regulation. Enrichment analysis identified 42 TFs that were significantly overrepresented in motifs disrupted (FDR < 0.05). These findings indicate that trait-related variants are concentrated around a small number of regulatory mediators, which does support a model where the polygenic effects are mediated by similartranscriptional control mechanisms and not individual variant effects.

4.4 Construction of Regulatory Gene Networks

The TF-gene regulatory network was an integrated network that comprised of 2,143 genes and 42 TFs (2,185 nodes) and 9,734 weighted edges (Fig. 7a). The inference into a gene-gene co-regulatory network provided 2,143 nodes and 18,921 edges of a highly connected regulation network (Fig. 7b). The distribution of degree was scaled-free (topological) with a small group of well-connected nodes acting as network hubs. Strong local modularity was also established by the mean clustering coefficient of 0.41, which indicated that genes were not randomly assembled into random networks, but were grouped together into closely-knit regulatory clusters.

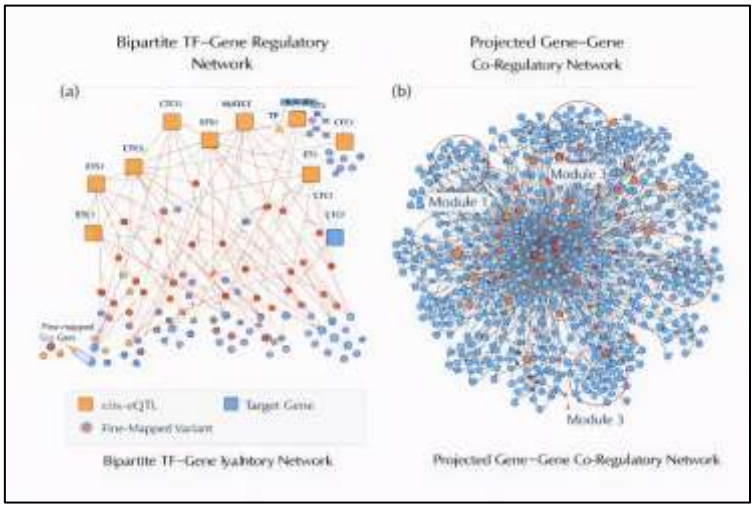

Fig. 7. Network visualization

4.5 Module Detection Identifies Trait-Relevant Regulatory Programs

The Louvain algorithm community detection algorithm revealed 17 large modules based on size (12-241 genes). Functional enrichment analysis showed that these modules are associated with important biological processes, such as immune signalings, chromosome remodeling, lipid metabolism, neuronal differentiation, and cellular stress responses. A number of modules had a strong tissue specificity, which was consistent with eQTL-based expression patterns and ontology of traits. This modular arrangement suggests that intricate characteristics are regulated through aligned regulatory programs as opposed to single gene impacts.

4.6 Hub Genes and Key Regulators

Network topology analysis was used to determine high centrality hub genes and transcription factors. Hub genes were chiefly linked to regulatory activities in transcription, chromatin accessibility and intracellular signaling pathways indicating that they play a key role in mediating downstream biological interactions. Compared to other types of TF hubs, GATA and ETS TF hubs were found to be highly betweenness central, suggesting that they were major connectors across various regulatory modules. These hub components are probably master regulators which combine signals of various variants and coordinate coordinated changes in gene expression.

4.7 Functional Interpretation of Network Topology

The high internal connectivity modules and central hub node modules were greatly enriched with previously trait-related pathway participating genes. Interestingly, genes with both eQTL and enhancer support were much more centralized than those with proximity only ($p < 0.001$) only measures, indicating the importance of functionally-directed mapping approaches. This evidence indicates that implementing regulatory evidence can not only enhance the assignment of variants to genes, but also the definition of biologically significant network structures. Comprehensively, the findings are consistent with a systems-level model where shared regulatory networks are converged towards and mediate expression of complex traits by distributed genetic variants.

5. DISCUSSION

This paper provides a holistic approach to the deconstruction of regulatory gene networks of complex traits expression. We examine locus-level interpretation by combining GWAS summary statistics, fine-mapping, eQTLs, enhancerpromotors connections, TF footprints and network topology to give us a systems-level understanding of trait biology. One important point to note is that a significant enhancement in the biological relevance of prioritized variants is achieved with the help of functionally informed fine- mapping. Variants that were chosen with annotation-specific prioritization were almost twice as likely to overlap active regulatory elements relative to variants with association strength prioritization. This is consonant with prior studies that have shown that incorporating functional annotations can optimally facilitate causal inference and minimize noise in polygenic architectures (Weissbrod et al., 2020; Wang et al., 2021). We also find that traits-associated variants come to a relatively small set of TFs and regulatory modules. The significance of this convergence lies in the fact that it indicates that the diffuse polygenic signals can generate coherent biological responsiveness that happens in common regulation pathways. The hub TFs and central genes are identified, and these offer mechanistic hypotheses to test in experiments and to possibly target therapies. The network-centric view has viable benefits compared to the nearest-gene assignment. Genes that were supported by eQTL and enhancer evidence were put in the middle of the network and it results that functionally-grounded mapping conserves biologically significant interactions that are not in proximity-based methods. It can be explained by the recent enhancements mapping studies that have revealed that regulatory variants are often targeted at distal genes and not the closest gene (Nasser et al., 2021). Regardless of these advantages, it is possible to identify a number of limitations. One, the framework is based on the quality and completeness of public functional genomics datasets, which might not be a complete reflection of all the relevant cell states or ancestries. Second, the scores of motif disruption are probabilities but not evidence of definitive changes in TF binding. TF footprint overlaps present probabilistic evidence of changes in TF binding. Third, the inferred networks are not dynamic and cannot represent temporal dynamics and situational signaling transitions. The single-cell multi-omics, perturbation-based validation, and dynamic modeling should be used in future research to narrow down network causality.

In general, the suggested methodology offers a reproducible and biologically-based way of translating GWAS associations into mechanistic regulatory networks. It can easily be generalized to disease-specific analyses, cross-population studies and precision medicine uses.

CONCLUSION

The paper will provide a unified and integrative structure to translate the resulting GWAS-based signals of association into mechanistic understanding at the regulation gene-networks level. The suggested approach can be used to detect high-confidence regulatory variants and their downstream biological implications by integrating fine-mapping, functional annotation, variant-to-gene mapping, transcription factor inference, and analysis based on networks. The findings indicate that intricate characteristics are controlled by integrated regulatory programs wherein several variants unite on common transcriptional units and crucial hub regulators. Multi-layered genomic evidence and the addition of functionally informed priors made a significant contribution to the prioritization of variants and biological interpretability, which emphasizes the need to combine statistical and functional information. A key contribution of the work is that it moves the analytical argument away towards regulatory architectures operating at the systems level and hence an opportunity to give a more realistic representation of complex trait biology. The results of identifying modular networks and central hub genes present a viable line of future research to experimentally validate and therapeutically target. Moreover, the structure has a reproducible and scalable design, which allows the use of the framework on a large variety of traits and datasets. Though these advances have been made, there are a number of limitations. The framework is based on the accessible public datasets which might not exhaust all the types of cells or dynamic regulation states. Moreover, computational inferences of regulatory interactions need to be experimentally verified to determine causality. The future research ought to include single-cell multi-omics data, time-varied regulatory dynamics and sophisticated deep neural networks to enhance inference of networks. Causal interpretation will be further enhanced with the incorporation

of perturbation-based validation approaches like CRISPR screening. All in all, this research can be used as a strong base to further the functional knowledge of complex trait genetics in the age of integrative genomics.

REFERENCES

1. Abell, N. S., DeGorter, M. K., Gloudemans, M., Greenwald, E., Smith, K. S., He, Z., et al. (2021). *Multiple causal variants underlie genetic associations in humans*. *bioRxiv*. https://doi.org/10.1101/2021.05.24.445471
2. Ajore, R., Niroula, A., Pertesi, M., Cafaro, C., Thodberg, M., Went, M., et al. (2022). Functional dissection of inherited non-coding variation influencing multiple myeloma risk. *Nature Communications, 13*, 151. https://doi.org/10.1038/s41467-021-27831-1
3. Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., Koido, M., Dey, K. K., et al. (2020). Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nature Genetics, 52*(12), 1346–1354. https://doi.org/10.1038/s41588-020-00740-8
4. Avsec, Z., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., et al. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods, 18*(10), 1196–1203. https://doi.org/10.1038/s41592-021-01252-x
5. Avsec, Z., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., et al. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nature Genetics, 53*(3), 354–366. https://doi.org/10.1038/s41588-021-00782-6
6. Bergman, D. T., Jones, T. R., Liu, V., Ray, J., Jagoda, E., Siraj, L., et al. (2022). Compatibility rules of human enhancer and promoter sequences. *Nature, 607*(7918), 176–184. https://doi.org/10.1038/s41586-022-04844-0
7. Cano-Gamez, E., & Trynka, G. (2020). From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Frontiers in Genetics, 11*, Article 424. https://doi.org/10.3389/fgene.2020.00424
8. Consortium, T. G. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science, 369*(6509), 1318–1330. https://doi.org/10.1126/science.aaz1776
9. Flynn, E. D., & Lappalainen, T. (2022). Functional characterization of genetic variant effects on expression. *Annual Review of Biomedical Data Science, 5*, 119–139. https://doi.org/10.1146/annurev-biodatasci-111721-074331
10. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E. M., Hercules, A., et al. (2021). Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Research, 49*(D1), D1311–D1320. https://doi.org/10.1093/nar/gkaa840
11. Kanai, M., Elzur, R., Zhou, W., Wu, K.-H. H., et al. (2022). Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genomics, 2*(11), 100210. https://doi.org/10.1016/j.xgen.2022.100210
12. Khetan, S., Kales, S., Kursawe, R., Jillette, A., Ulirsch, J. C., Reilly, S. K., et al. (2021). Functional characterization of T2D-associated SNP effects on baseline and ER stress-responsive β cell transcriptional activation. *Nature Communications, 12*, 5242. https://doi.org/10.1038/s41467-021-25417-5
13. Klein, J. C., Agarwal, V., Inoue, F., Keith, A., Martin, B., Kircher, M., et al. (2020). A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nature Methods, 17*(11), 1083–1091. https://doi.org/10.1038/s41592-020-0965-3
14. Loos, R. J. F. (2020). 15 years of genome-wide association studies and no signs of slowing down. *Nature Communications, 11*, 5900. https://doi.org/10.1038/s41467-020-19653-5
15. Mountjoy, E., Schmidt, E. M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., et al. (2021). An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nature Genetics, 53*(11), 1527–1533. https://doi.org/10.1038/s41588-021-00945-5
16. Mouri, K., Dewey, H. B., Castro, R., Berenzy, D., Kales, S., & Tewhey, R. (2023). Whole-genome functional characterization of RE1 silencers using a modified massively parallel reporter assay. *Cell Genomics, 3*(3), 100234. https://doi.org/10.1016/j.xgen.2023.100234
17. Mouri, K., Guo, M. H., de Boer, C. G., Lissner, M. M., Harten, I. A., Newby, G. A., et al. (2022). Prioritization of autoimmune disease-associated genetic variants that perturb regulatory element activity in T cells. *Nature Genetics, 54*(5), 603–612. https://doi.org/10.1038/s41588-022-01047-7
18. Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature, 593*(7858), 238–243. https://doi.org/10.1038/s41586-021-03446-x
19. O'Connor, L. J. (2021). The distribution of common-variant effect sizes. *Nature Genetics, 53*(9), 1243–1249. https://doi.org/10.1038/s41588-021-00941-9
20. Ray, J. P., de Boer, C. G., Fulco, C. P., Lareau, C. A., Kanai, M., Ulirsch, J. C., et al. (2020). Prioritizing disease and trait causal variants at the TNFAIP3 locus using functional and genomic features. *Nature Communications, 11*, 1237. https://doi.org/10.1038/s41467-020-14936-z
21. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics, 53*(10), 1415–1424. https://doi.org/10.1038/s41588-021-00931-x
22. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., et al. (2023). The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Research, 51*(D1), D977–D985. https://doi.org/10.1093/nar/gkac1010
23. Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nature Reviews Methods Primers, 1*, 59. https://doi.org/10.1038/s43586-021-00056-9

24. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., et al. (2020). Global reference mapping of human transcription factor footprints. *Nature, 583*(7818), 729–736. https://doi.org/10.1038/s41586-020-2528-x
25. Vorontsov, I. E., Eliseeva, I. A., Zinkevich, A., Nikonov, M., Abramov, S., Boytsov, A., et al. (2024). HOCOMOCO in 2024: A rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Research, 52*(D1), D154–D163. https://doi.org/10.1093/nar/gkad989
26. Wang, G., Sarkar, A., Carbonetto, P., & Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82*(5), 1273–1300. https://doi.org/10.1111/rssb.12388
27. Wang, Q. S., Kelley, D. R., Ulirsch, J., Kanai, M., Sadhuka, S., Cui, R., et al. (2021). Leveraging supervised learning for functionally informed fine-mapping of cis-eQTLs identifies an additional 20,913 putative causal eQTLs. *Nature Communications, 12*, 3394. https://doi.org/10.1038/s41467-021-23772-2
28. Wang, X., & Goldstein, D. B. (2020). Enhancer domains predict gene pathogenicity and inform gene discovery in complex disease. *The American Journal of Human Genetics, 106*(2), 215–233. https://doi.org/10.1016/j.ajhg.2019.12.012
29. Weeks, E. M., Ulirsch, J. C., Cheng, N. Y., Trippe, B. L., Fine, R. S., Miao, J., et al. (2023). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nature Genetics, 55*(8), 1267–1276. https://doi.org/10.1038/s41588-023-01443-6
30. Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., et al. (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics, 52*(12), 1355–1363. https://doi.org/10.1038/s41588-020-00735-5
31. Xue, J. R., Mackay-Smith, A., Mouri, K., Garcia, M. F., Dong, M. X., Akers, J. F., et al. (2023). The functional and evolutionary impacts of human-specific deletions in conserved elements. *Science, 380*(6649), eabn2253. https://doi.org/10.1126/science.abn2253
32. Zou, Z., Ohta, T., Miura, F., & Oki, S. (2022). ChIP-Atlas 2021 update: A data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Research, 50*(W1), W175–W182. https://doi.org/10.1093/nar/gkac199