Bridging the SNP-to-Gene Gap using eQTL Mapping and Network Analysis

Dinesha Jayawardana[1], Vivi Arief[1], Scott Chapman[1], Yidi Deng[2], Emi Tanaka[2], Huan Zhao[3], Claire Huang[1] and David Kainer[1]

[1]School of Agriculture and Food Sustainability, The University of Queensland, St Lucia QLD 4072, Australia
[2]Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, Australian Capital Territory, Australia
[3]School of Agriculture, Food and Wine, Adelaide University, Waite Campus, Urrbrae, Australia
Contact: dinesha.jayawardana@uq.edu.au

BACKGROUND AND MOTIVATION

- Genome-wide association studies (GWAS) can identify genomic regions associated with complex traits, but often fail to pinpoint the causal genes underlying these associations. To address this challenge, AAGI-UQ is developing an integrated eQTL mapping and network analysis pipeline that links genetic variants to gene expression, helping to bridge the SNP-to-gene gap in the sorghum diversity panel. By integrating eQTL networks with GWAS signals, this framework enables the biological interpretation of sorghum genotype data by connecting SNPs to regulatory genes and gene networks that influence trait variation.
- This research work is directly motivated by the Grain Data Fusion project 'Building the Bridge from GWAS to Breeder through Multi-Omic Network Data Fusion (UOQ2503-015RTX)', and the identified eQTLs will be an input for the GWAS boosting analysis of the project as an AAGI service and support work.

STUDY AIM

Developing a transferable eQTL mapping and network analysis pipeline to improve SNP-to-gene assignment in post-GWAS studies using public sorghum datasets.

PIPELINE

Four linked stages take raw data to eQTL network; 1) primary data acquisition; 2) quality control, population structure analysis and factor analysis; 3) cis and trans-eQTL detection and two-stage FDR control and 4) network analysis.

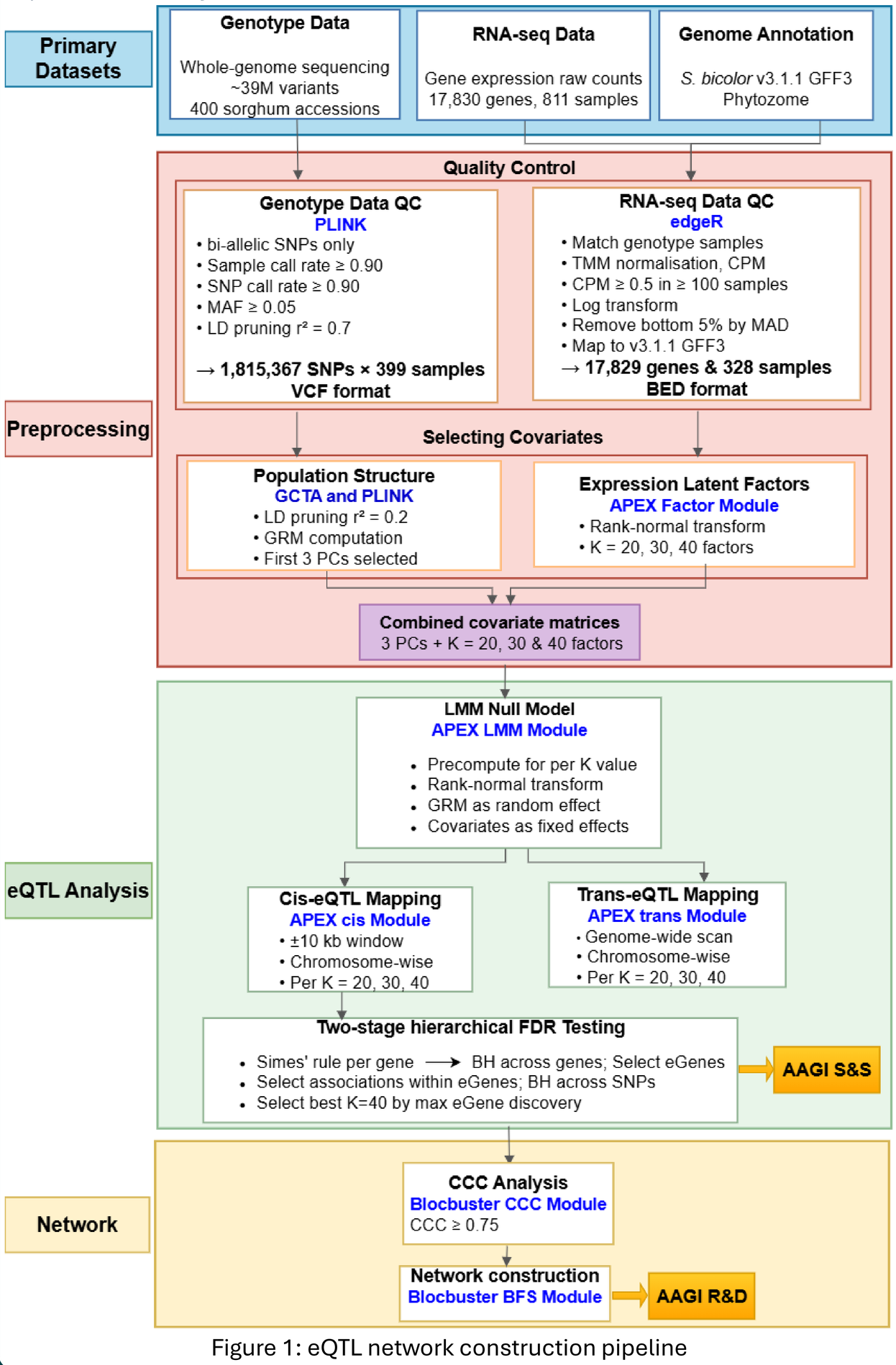


Figure 1: eQTL network construction pipeline

RESULTS

- A total of 13,743 genes (77.1%) were identified as significant eGenes, with K = 40 yielding the highest number of eGene associations.
- Following filtering, 138,357 significant cis-eQTL SNPs retained for downstream CCC analysis.

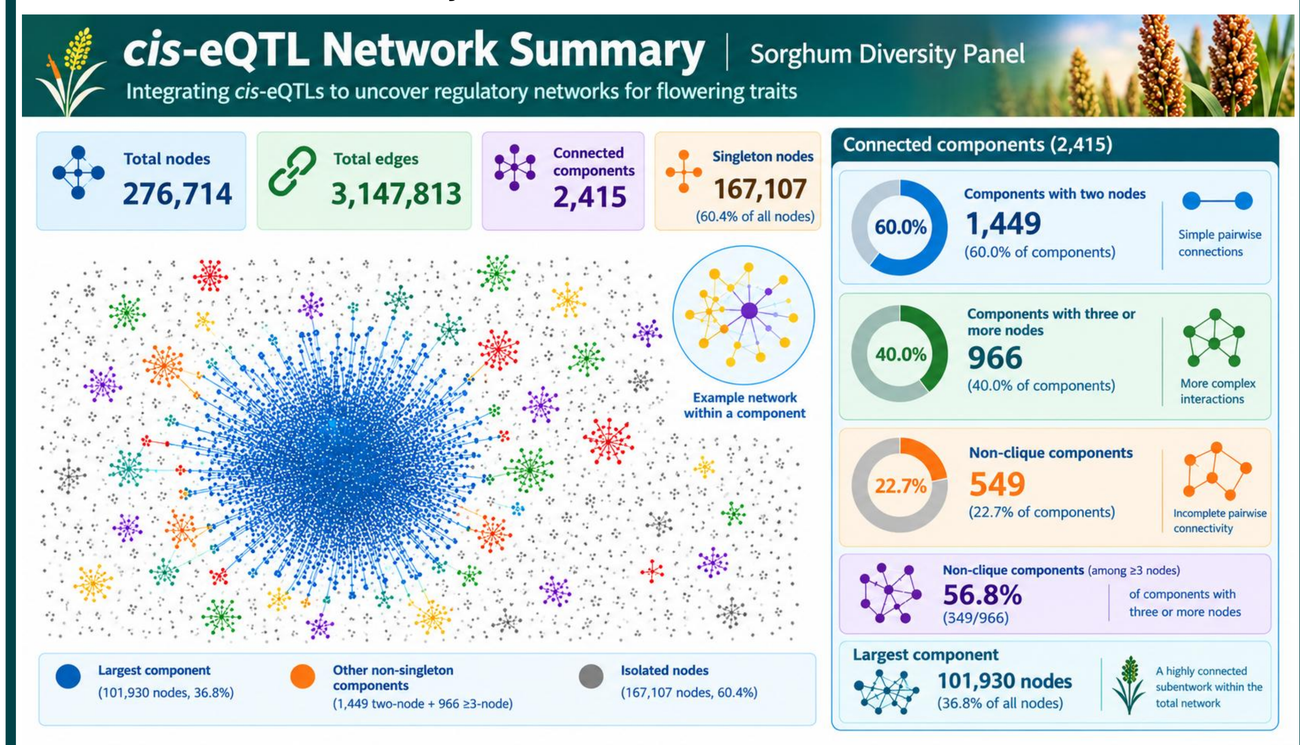


Figure 2: Summary of cis-eQTL network

DISCUSSION

- **Large connected network:** The largest component contained 101,930 nodes (36.8%), suggesting a substantial interconnected structure that may capture relationships among SNP regions and haplotypes, warranting further investigation of its biological significance.
- **Diverse connectivity:** 60.0% of connected components contained two nodes, while 56.8% of components with ≥3 nodes were non-cliques, indicating that many larger subnetworks exhibit more complex connectivity patterns than simple, fully connected groups.
- **High fragmentation:** 60.4% of nodes were isolated, highlighting substantial sparsity in the inferred network.
- **Biological relevance:** The network structure provides a foundation for prioritising biologically meaningful subnetworks and evaluating their connectivity and potential relevance to flowering-time variation in sorghum.

FUTURE WORKS

Integrate network modules with eQTL, GWAS, and gene annotations to identify key regulatory networks, hub SNPs, candidate genes, and pathways associated with flowering-time variation, supporting marker validation and sorghum breeding improvement.

The support of the Grains Research and Development Corporation (GRDC) provided through the Analytics for the Australian Grains Industry (AAGI) project is acknowledged.