

# Refining Context-Specific eQTL Detection: A Comparative Evaluation of Correction Methods in FastGxC

Jiachen Ai<sup>1</sup>, Lena Krockenberger<sup>2</sup>, and Brunilda Balliu<sup>\*1,3,4</sup>

<sup>1</sup>Department of Biostatistics, University of California Los Angeles, Los Angeles, CA, USA  
<sup>2</sup>Bioinformatics Interdepartmental Graduate Program, University of California Los Angeles, Los Angeles, CA, USA

<sup>3</sup>Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA, USA

<sup>4</sup>Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA, USA

## Abstract

Expression quantitative trait loci (eQTL) mapping helps uncover how genetic variation regulates gene expression across biological contexts. Many eQTLs exhibit context-specific effects that standard approaches often fail to detect efficiently. FastGxC is a recently developed pipeline that improves detection of both shared and context-specific eQTLs by decomposing gene expression, eQTL mapping, and multiple testing correction. In this study, we apply FastGxC to simulated datasets and focus on evaluating and refining its multiple testing correction step. We compare three methods: TreeQTL, TreeBH, and mashR, which represent hierarchical correction and empirical Bayes shrinkage approaches. Simulation results show that TreeQTL provides strong false discovery rate (FDR) control, TreeBH offers a balance between power and precision, and mashR improves sensitivity in small sample settings with heterogeneous effects. These findings provide practical guidance for choosing multiple testing strategies in context-aware eQTL analyses and demonstrate how FastGxC, combined with appropriate correction methods, enables scalable and accurate detection of regulatory variants across diverse conditions.

**Keywords:** context-specific eQTL; multiple testing correction; FastGxC

---

\*Corresponding author: bballiu@ucla.edu

# 1 Introduction

Genome-wide association studies (GWAS) is a method used to identify genetic variants associated with complex traits and diseases by scanning the genome for single nucleotide polymorphisms (SNPs) that occur more frequently in individuals with a particular trait or disease compared to those without<sup>1–3</sup>. These studies have been instrumental in uncovering the genetic basis of various conditions over the past 15 years, including cardiovascular diseases<sup>4</sup>, diabetes<sup>5</sup>, and psychiatric disorders<sup>6</sup>. Notably, the majority of these variants reside in non-coding regions of the genome, frequently overlapping with regulatory elements such as enhancers and promoters<sup>7</sup>. This observation suggests that many trait-associated variants exert their effects through transcriptional regulation rather than through alterations in protein-coding sequences<sup>7–10</sup>. Therefore, understanding the transcriptional regulation of gene expression is crucial for elucidating the mechanisms underlying complex traits and diseases<sup>11</sup>.

To bridge the gap between GWAS associations and gene regulatory mechanisms<sup>12–14</sup>, considerable attention has been directed toward identifying eQTLs<sup>13,14</sup>. eQTLs are genetic variants that influence gene expression levels, providing insights into how genetic variation can affect phenotypes<sup>15,16</sup>. By mapping eQTLs, researchers can uncover the regulatory networks that govern gene expression and their contributions to complex traits<sup>15</sup>. Advances in next-generation sequencing technologies have enabled the measurement of gene expression across various biological contexts, including different tissues<sup>17</sup>, cell types<sup>18–20</sup>, or environmental conditions<sup>21–25</sup>, using both bulk and single-cell RNA sequencing. These developments have reinforced the hypothesis that many eQTLs exhibit context-specific effects, wherein a variant may regulate gene expression in one tissue or condition but not in another<sup>17</sup>. Context-specificity plays a critical role in understanding the functional impact of genetic variants and their contributions to trait variation<sup>26</sup>. Many conventional eQTL mapping methods can detect context-specific effects, but they often analyze each tissue or condition separately or assume that effects are shared in simple ways<sup>27</sup>. This can lead to suboptimal detection of genetic variants that behave differently across biological contexts. These challenges are especially pronounced in repeated measures studies, where gene expression is profiled across multiple conditions or time points within the same individuals, requiring methods that account for within-subject correlations and context-dependent regulatory effects<sup>28</sup>.

To address the limitations of these conventional approaches, several methods have been developed to identify context-specific eQTLs in repeated measures studies, falling into two primary categories. One class includes joint modeling approaches that incorporate genotype-by-context (GxC) interaction terms within (generalized) linear mixed models<sup>29–34</sup> or non-linear<sup>35</sup> to detect context-specific regulatory effects. While these methods are powerful, they are often computationally intensive and may produce latent contexts that are difficult to interpret in biological terms. The second class involves context-by-context (CxC) approaches, which map eQTLs separately within each context and assess context-specificity post hoc<sup>17,18,20,36,37</sup>.

Although computationally efficient, CxC methods often suffer from reduced statistical power and rely on ad hoc criteria for defining context-specificity, leading to potential false positives or negatives. These limitations hinder the ability to interpret the functional relevance of genetic variants in a context-dependent manner. To address these challenges, Krockenberger et al.<sup>38</sup> developed FastGxC, a novel method for efficient mapping of context-specific eQTLs that accounts for repeated measures, offering both computational scalability and biological interpretability. The method innovatively decomposes gene expression into context-shared and context-specific components, estimates genetic effects via separate linear regressions, and detects eQTLs through appropriate multiple testing correction. FastGxC offers three key advantages over previous methods: (1) it directly identifies specific eQTLs without relying on post hoc thresholds<sup>38</sup>, (2) it accounts for intra-individual correlation (e.g. sex, age, population stratification, or sequencing batch<sup>39–41</sup>) to reduce confounding and enhance power, (3) it leverages ultra-fast linear regression implementations<sup>42</sup>, reducing computational time from years to minutes. FastGxC is broadly applicable to continuous molecular traits and integrates seamlessly with modern eQTL mapping frameworks.

Multiple testing correction is a statistical procedure used when conducting multiple hypothesis tests to reduce the inflated Type I errors caused by multiple times of testing<sup>43–45</sup>. In the context of eQTL mapping, where thousands of genes and variants are tested simultaneously, multiple testing correction is crucial to ensure that the identified associations are statistically robust and not due to chance<sup>44</sup>. Two categories of multiple testing correction methods are commonly employed in recent functional genomics studies, including differential expression and eQTL mapping analyses: hierarchical correction-based methods and empirical Bayes shrinkage-based approaches. Hierarchical correction methods control the false discovery rate (FDR)<sup>44</sup> by adjusting p-values according to their rank within a predefined hierarchical structure. These methods are especially useful in large-scale studies and can effectively balance discovery power and error control<sup>44,45</sup>. On the other hand, empirical Bayes shrinkage methods<sup>43,46</sup> leverage prior information to improve estimation accuracy and reduce variance in p-value estimates. These methods can be particularly advantageous in scenarios with limited sample sizes or when dealing with sparse data. In the established FastGxC pipeline, the multiple testing correction step is crucial for ensuring the reliability of eQTL associations. The original FastGxC implementation uses a hierarchical correction-based method, which has shown effectiveness in controlling false discovery rates in large-scale eQTL mapping studies<sup>38</sup>. However, the choice of multiple testing correction method can significantly impact the results and interpretations of eQTL analyses. Therefore, it is essential to evaluate and compare different correction methods to identify the most suitable approach for specific research contexts.

This report begins by outlining the three-step architecture of the FastGxC pipeline: expression decomposition, eQTL mapping, and multiple testing correction. We then focus on evaluating the multiple testing correction step by systematically comparing two hierarchical FDR control methods (TreeQTL and TreeBH) and an empirical Bayes shrinkage approach

(mashR). Through simulation studies under varying intra-individual correlation structures and sample sizes, we assess each method’s performance in detecting context-specific eQTLs while effectively controlling false discoveries. Based on these evaluations, we provide practical recommendations for selecting correction strategies tailored to different study designs. In summary, the refined FastGxC framework offers a scalable, efficient, and statistically robust approach for context-aware eQTL mapping, advancing our understanding of the regulatory mechanisms that underlie complex human diseases.

## 2 Methods

### 2.1 FastGxC

Gene expression data with repeated measures are frequently collected in a context-specific framework, where each context corresponds to a distinct biological condition or environment. In this setting, gene expression can be conceptualized as a three-dimensional tensor, with dimensions corresponding to genes, individuals, and contexts. Let  $i = 1, \dots, I$  index individuals,  $j = 1, \dots, J$  index genes, and  $c = 1, \dots, C$  index contexts. The expression tensor is denoted as  $E_{ijc}$  (Figure 1), where each entry represents the expression level of gene  $j$  in individual  $i$  under context  $c$ . Let  $s = 1, \dots, S$  index genotype of SNP, the genotype is represented as a matrix  $G_{is}$ , where each entry corresponds to the genotype of SNP  $s$  for individual  $i$ . The goal of FastGxC is to model the relationship between the genotype and context-shared and context-specific gene expression across the sample, allowing for the identification of context-specific eQTLs.

#### 2.1.1 Decomposition

The FastGxC pipeline comprises three principal steps: expression decomposition, eQTL mapping, and multiple testing correction (Figure 1). In the first step, gene expression data are decomposed into context-shared and context-specific components. Drawing on the framework proposed by Cronbach and Webb<sup>47</sup>, for each individual, we assume that the observed expression of a gene reflects both shared expression across contexts and variation that is specific to individual contexts. Let  $E_{ic}$  denote the expression of any gene in context  $c$  for individual  $i$ , where  $i = 1, \dots, I$  and  $c = 1, \dots, C$ . The expression can be decomposed as equation (1).  $E_{..} = \frac{\sum_{j=1}^J \sum_{c=1}^C E_{ic}}{I \times C}$  is the overall mean expression across all individuals and contexts (i.e., the population mean),  $E_{i.} = \frac{\sum_{c=1}^C E_{ic}}{C}$  represents the mean expression of the gene for individual  $j$  across all contexts (i.e., the individual mean). Then new term  $E_i^{sh} = E_{i.} - E_{..}$  captures the context-shared component of expression for individual  $i$ . And  $E_{ic}^{sp} = E_{ic} - E_{i.}$  captures the context-specific expression in context  $c$  for individual  $i$ .

$$E_{ic} = E_{..} + \underbrace{(E_{i.} - E_{..})}_{E_i^{sh}} + \underbrace{(E_{ic} - E_{i.})}_{E_{ic}^{sp}} \quad (1)$$

#### 2.1.2 eQTL mapping

In the second step of the FastGxC pipeline, eQTL mapping is performed by regressing cis-acting single nucleotide polymorphisms (cis-SNPs), which are defined as variants located within 1,000 base pairs of a given gene, on the context-shared and context-specific components of gene expression. This step leverages ultra-fast fixed-effect linear regression models implemented via the **MatrixEQTL**<sup>42</sup> framework, which is designed for high-throughput genomic association analyses.

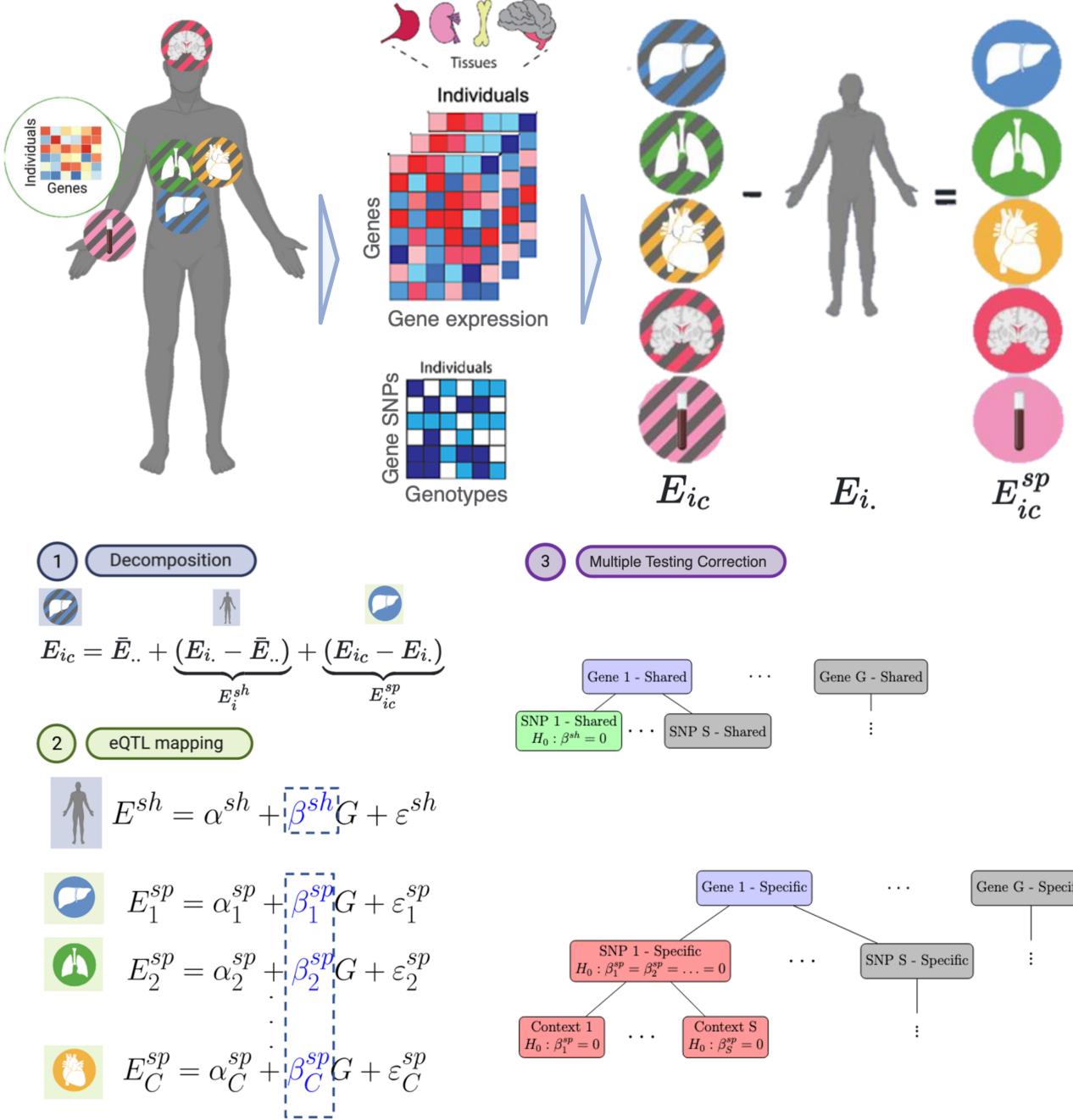


Figure 1: **Overview of the FastGxC pipeline.** The top row illustrates the data structure, where genotype and gene expression matrices are collected across multiple individuals and biological contexts (e.g., tissues), forming a three-dimensional tensor. This layout facilitates the decomposition of gene expression into context-shared and context-specific components for downstream eQTL mapping. FastGxC consists of three steps: (1) Decomposition of gene expression into context-shared ( $E_i^{sh}$ ) and context-specific ( $E_{ic}^{sp}$ ) components; (2) eQTL mapping via linear regression of genotypes on each component to estimate shared ( $\beta^{sh}$ ) and context-specific ( $\beta_c^{sp}$ ) genetic effects; and (3) Multiple testing correction using hierarchical procedures to control the false discovery rate across genes, SNPs, and contexts.

Let  $G_j \in \{0, 1, 2\}^J$  denote the vector of genotypes for a cis-SNP associated with gene  $i$ , where  $J$  is the number of individuals and each genotype value represents the number of copies of the reference allele: 0 for homozygous alternate (e.g.,  $bb$ ), 1 for heterozygous (e.g.,  $Bb$  or  $bB$ ), and 2 for homozygous reference (e.g.,  $BB$ ). The context-shared expression components are modeled as equation (2) and context-specific expression components are modeled as equation (3). Here,  $E_{ij}^{sh}$  denotes the context-shared expression of gene  $i$  for individual  $j$ , and  $E_{ijc}^{sp}$  denotes the context-specific expression for gene  $i$  in individual  $j$  under context  $c$ . The coefficients  $\beta_i^{sh}$  and  $\beta_{ic}^{sp}$  represent the genetic effects on the shared and context-specific expression, respectively. The residual terms  $\varepsilon_{ij}^{sh}$  and  $\varepsilon_{ijc}^{sp}$  capture the unexplained variation and are assumed to follow a normal distribution. All models are estimated under the Gauss–Markov assumptions, ensuring that the estimators are unbiased, consistent, and efficient.

$$E_{ij}^{sh} = \alpha_i^{sh} + \beta_i^{sh} G_j + \varepsilon_{ij}^{sh}, \quad \varepsilon_{ij}^{sh} \sim \mathcal{N}(0, \sigma_{sh}^2) \quad (2)$$

$$E_{ijc}^{sp} = \alpha_{ic}^{sp} + \beta_{ic}^{sp} G_j + \varepsilon_{ijc}^{sp}, \quad \varepsilon_{ijc}^{sp} \sim \mathcal{N}(0, \sigma_{sp,c}^2) \quad (3)$$

### 2.1.3 Multiple testing correction

In the third step of the FastGxC pipeline, significant eQTLs are identified based on p-values derived from  $t$ -tests on the estimated regression coefficients  $\beta_j^{sh}$  and  $\beta_{jc}^{sp}$ . To control for multiple testing while preserving statistical power, we employ TreeQTL<sup>44</sup>, a hierarchical FDR correction method that organizes hypothesis testing into multiple levels, as shown in step (3) in Figure 1.

For the identification of *shared eQTLs*, TreeQTL first conducts a Level 1 gene-level test to determine whether any cis-SNPs are associated with the context-shared expression of a given gene. Specifically, for gene  $j$ , the null and alternative hypotheses are:  $H_0^{(j)} : \beta_{js}^{sh} = 0$  for all cis-SNPs  $s$  vs.  $H_A^{(j)} : \exists s \text{ such that } \beta_{js}^{sh} \neq 0$ . If  $H_0^{(j)}$  is rejected, the gene proceeds to Level 2 testing, where each cis-SNP  $s$  is individually tested for association with the gene's context-shared expression:  $H_0^{(js)} : \beta_{js}^{sh} = 0$ . Rejection of  $H_0^{(js)}$  indicates that SNP  $s$  is a significant shared eQTL for gene  $j$ . FDR correction is applied independently at each level to ensure control of false discoveries.

For the identification of *context-specific eQTLs*, TreeQTL extends this structure to accommodate multiple biological contexts. The Level 1 test evaluates whether any cis-SNP exhibits a context-specific effect for gene  $j$ :  $H_0^{(j)} : \beta_{jsc}^{sp} = 0$  for all  $s$  and  $c$  vs.  $H_A^{(j)} : \exists s, c \text{ such that } \beta_{jsc}^{sp} \neq 0$ . If  $H_0^{(j)}$  is rejected, the gene proceeds to Level 2 testing, which evaluates whether each SNP  $s$  has any context-specific effect across all contexts:  $H_0^{(js)} : \beta_{jsc}^{sp} = 0$  for all  $c$  vs.  $H_A^{(js)} : \exists c \text{ such that } \beta_{jsc}^{sp} \neq 0$ . SNPs passing Level 2 proceed to Level 3, where the effect of SNP  $s$  on gene  $j$  is tested within each specific context  $c$ :

$$H_0^{(jsc)} : \beta_{jsc}^{sp} = 0 \quad \text{vs.} \quad H_A^{(jsc)} : \beta_{jsc}^{sp} \neq 0.$$

This structured multi-level testing approach enables TreeQTL to rigorously control the FDR while enhancing power to detect both shared and context-specific eQTLs. Its hierarchical design makes it well-suited for studies involving repeated measures and multiple biological conditions, as in FastGxC.

## 2.2 TreeBH

The `TreeBH`<sup>45</sup> (Tree-Structured Benjamini–Hochberg) procedure is a hierarchical multiple testing method designed to control the selective FDR (sFDR) across multiple levels of a hypothesis tree. It generalizes the classical BH procedure to accommodate structured hypotheses, where each node corresponds to a hypothesis and each parent represents the logical intersection of its child hypotheses. The algorithm proceeds in a top-down, recursive fashion. At each level  $\ell$ , the BH procedure is applied to families of hypotheses  $\mathcal{F}_i^{(\ell)}$ , where each family consists of the children of a hypothesis  $H_i$  that was rejected at the preceding level  $\ell - 1$ . To account for the hierarchical selection, the significance level for testing each family is scaled down according to the rejections made along its ancestral path via equation (4).  $q^{(\ell)}$  is the target FDR level at layer  $\ell$ ,  $\mathcal{A}(i)$  is the set of ancestor nodes on the path from the root to hypothesis  $H_i$ ,  $S_j$  is the number of rejections in family  $\mathcal{F}_j$ , and  $|\mathcal{F}_j|$  is the size of the family  $\mathcal{F}_j$ . The adjusted threshold  $q_i^{(\ell)}$  determines how likely a family is to be tested.

$$q_i^{(\ell)} = q^{(\ell)} \cdot \prod_{j \in \mathcal{A}(i)} \frac{S_j}{|\mathcal{F}_j|} \quad (4)$$

To evaluate the overall false discovery rate while respecting this selection process, TreeBH defines selective FDR at level  $\ell$  as equation (5).  $S_{\ell-1}$  is the set of rejected hypotheses at level  $\ell - 1$ ,  $\text{FDP}(\mathcal{F}_i^{(\ell)})$  is the false discovery proportion within family  $\mathcal{F}_i^{(\ell)}$ ,  $w_i^{(\ell)}$  down-weights the contribution of a family based on how selectively it was tested. The weight  $w_i^{(\ell)}$  in the sFDR calculation reflects how much influence that family's false discovery proportion should have, based on its selection path. Together, this process ensure that testing is done cautiously in deep branches, and that error rates are fairly and rigorously controlled across the entire hierarchy.

$$\text{sFDR}_{\ell} = \mathbb{E} \left[ \sum_{i \in S_{\ell-1}} w_i^{(\ell)} \cdot \text{FDP}(\mathcal{F}_i^{(\ell)}) \right], \quad \text{where} \quad w_i^{(\ell)} = \left( \prod_{j \in \mathcal{A}(i)} S_j \right)^{-1} \quad (5)$$

## 2.3 mashR

To compare the performance of hierarchical correction methods, we also evaluate an empirical Bayes shrinkage approach using the `mashR`<sup>43</sup> (multivariate adaptive shrinkage in R) framework, which enables joint modeling of multivariate effects and multiple testing correction in a unified probabilistic model. Given the estimates of the observed effects  $\hat{\boldsymbol{b}}_j$  for the gene-SNP

pair  $j$  in biological contexts  $R$ , the observation model is defined in equation (6).  $\mathbf{b}_j$  is the true latent effect vector and  $V_j$  is the known covariance matrix of estimation error.

$$\hat{\mathbf{b}}_j \mid \mathbf{b}_j \sim \mathcal{N}_R(\mathbf{b}_j, V_j) \quad (6)$$

To flexibly model both shared and condition-specific patterns of genetic regulation, mashR specifies a mixture-of-normals prior over  $\mathbf{b}_j$  as equation (7).  $U_k$  are covariance matrices capturing canonical or data-driven sharing structures. Figure 2A shows an example for  $U_k$ .  $\omega_\ell$  are scale parameters representing effect magnitudes, and  $\pi_{k\ell}$  are the mixture weights to be estimated.

$$\mathbf{b}_j \sim \sum_{k,\ell} \pi_{k\ell} \mathcal{N}_R(\mathbf{0}, \omega_\ell U_k) \quad (7)$$

Posterior inference yields the conditional distribution of  $\mathbf{b}_j$  given observation  $\hat{\mathbf{b}}_j$  by equation (8).  $\tilde{\pi}_{j,k\ell}$  are posterior weights,  $\tilde{\mu}_{j,k\ell}$  are posterior means, and  $\tilde{U}_{j,k\ell}$  are posterior covariances.

$$\mathbf{b}_j \mid \hat{\mathbf{b}}_j \sim \sum_{k,\ell} \tilde{\pi}_{j,k\ell} \mathcal{N}_R(\tilde{\mu}_{j,k\ell}, \tilde{U}_{j,k\ell}) \quad (8)$$

$$\mathbb{E}[\mathbf{b}_j \mid \hat{\mathbf{b}}_j] = \sum_{k,\ell} \tilde{\pi}_{j,k\ell} \tilde{\mu}_{j,k\ell} \quad (9)$$

The final estimate of the true effect vector is given by equation (9), which provides adaptive shrinkage by borrowing strength across conditions while retaining the flexibility to model both global and context-specific effects. The mixture weights  $\pi_{k\ell}$  are estimated via maximum likelihood using the expectation–maximization (EM) algorithm<sup>43,48</sup>. Significant eQTLs are identified using the local false sign rate (lfsr), which quantifies the probability of incorrectly inferring the direction of an effect<sup>43,46</sup>. Compared to the local false discovery rate (lfdr), the lfsr is considered a more robust and broadly applicable measure of significance, particularly in settings where accurately estimating effect direction is crucial<sup>46</sup>.

## 2.4 Simulation

To evaluate the performance of the proposed method, we conducted simulations that reflect key features of eQTL studies across multiple biological contexts. For each individual  $j$  and SNP  $s$ , genotypes were independently sampled from a binomial distribution:  $G_{js} \sim \text{Binomial}(2, \text{MAF})$ . We assume the pre-defined minor allele frequency (MAF) as 0.2. Gene expression levels were generated under a linear model of genetic regulation given by equation (10).  $\mu_c$  is the context-specific baseline expression,  $\beta_{ic}$  is the genetic effect size in context  $c$ ,  $G_{ji}$  is the genotype at the causal SNP, and  $\varepsilon_{jc}$  is the residual noise. To capture intra-individual correlation across contexts, the residual vector  $\boldsymbol{\varepsilon}_j = (\varepsilon_{j1}, \dots, \varepsilon_{jC})$  was drawn from a multivariate normal distribution:  $\boldsymbol{\varepsilon}_j \sim \mathcal{N}_C(\mathbf{0}, \Sigma)$ ,  $\Sigma = (1 - w_{\text{corr}})I_C + w_{\text{corr}}\mathbf{1}_C\mathbf{1}_C^\top$ , where  $w_{\text{corr}}$  represents the intra-individual correlation across contexts. The effect sizes  $\beta_{ic}$  were de-

terminated from the heritability  $h_{ic}^2$  using equation (11) with  $v_e$  denoting the residual variance. Under the *null* scenario, we set  $h_{ic}^2 = 0$  for all  $i$  and  $c$ , implying no genetic contribution to expression. Under the *single context heterogeneity* scenario, each gene  $i$  was assigned a dominant context  $c_i^*$  with heritability  $h_{ic}^2 = 0.2$  in that context and  $h_{ic}^2 = 0.1$  in others, simulating context-specific regulation. Each gene was assumed to have exactly one causal SNP (set as the first SNP) that influenced expression in all contexts.

$$Y_{jic} = \mu_c + \beta_{ic} G_{ji} + \varepsilon_{jc}, \quad \boldsymbol{\varepsilon}_j = (\varepsilon_{j1}, \dots, \varepsilon_{jC}) \sim \mathcal{N}_C(\mathbf{0}, \Sigma) \quad (10)$$

$$\beta_{ic} = \sqrt{\frac{h_{ic}^2 v_e}{(1 - h_{ic}^2) \cdot \text{Var}(G_{ji})}} \quad (11)$$

This simulation generated a genotype matrix  $G \in \mathbb{N}^{I \times (S \times J)}$ , a gene expression tensor  $Y \in \mathbb{R}^{I \times J \times C}$ , and associated SNP and gene location files. In this study, we varied the intra-individual correlation parameter  $w_{\text{corr}} \in \{0, 0.2, 0.4, 0.6, 0.8\}$ , number of individuals  $I \in \{100, 500\}$ , simulation scenarios (*null* and *single context heterogeneity*), number of genes  $J = 1000$ , number of SNPs per gene  $S = 100$ , and number of biological contexts  $C = 10$ .

## 2.5 Performance assessment

To assess the performance of multiple testing correction methods, we applied `mashR`, `TreeBH`, and `TreeQTL` to simulated datasets under both the *null* and *single context heterogeneity* scenarios across varying levels of intra-individual correlation. Performance was evaluated based on two criteria: Type I error, defined as the proportion of false positives among all discoveries under the *null*, and power, defined as the proportion of true positives correctly identified under the *single context heterogeneity*. We also assessed precision, which is the proportion of true positives among all discoveries, and recall, which is the proportion of true positives among all true signals, to quantify each method's ability to accurately detect eQTLs.

`TreeQTL`, `TreeBH`, and `mashR` are three candidates for the refinement of multiple testing correction step in `FastGxC`. `TreeQTL` and `TreeBH` are both hierarchical multiple testing procedures developed to control FDR in structured hypothesis settings. `TreeQTL` was specifically designed for eQTL analysis and implements a multi-level hierarchy to control FDR at both the gene and variant–gene pair levels. It uses Simes' method<sup>49</sup> to compute p-values for higher-level hypotheses and applies Benjamini–Hochberg (BH) procedures within selected groups to control FDR, allowing for selective inference. `TreeBH` generalizes this framework to accommodate arbitrarily deep tree structures and introduces selective FDR, a new error metric that accounts for the selection process and adapts to signal heterogeneity across the tree. `TreeBH` provides theoretical guarantees of FDR control at all levels under mild dependency assumptions and can be viewed as an extension of `TreeQTL`, formalizing and broadening

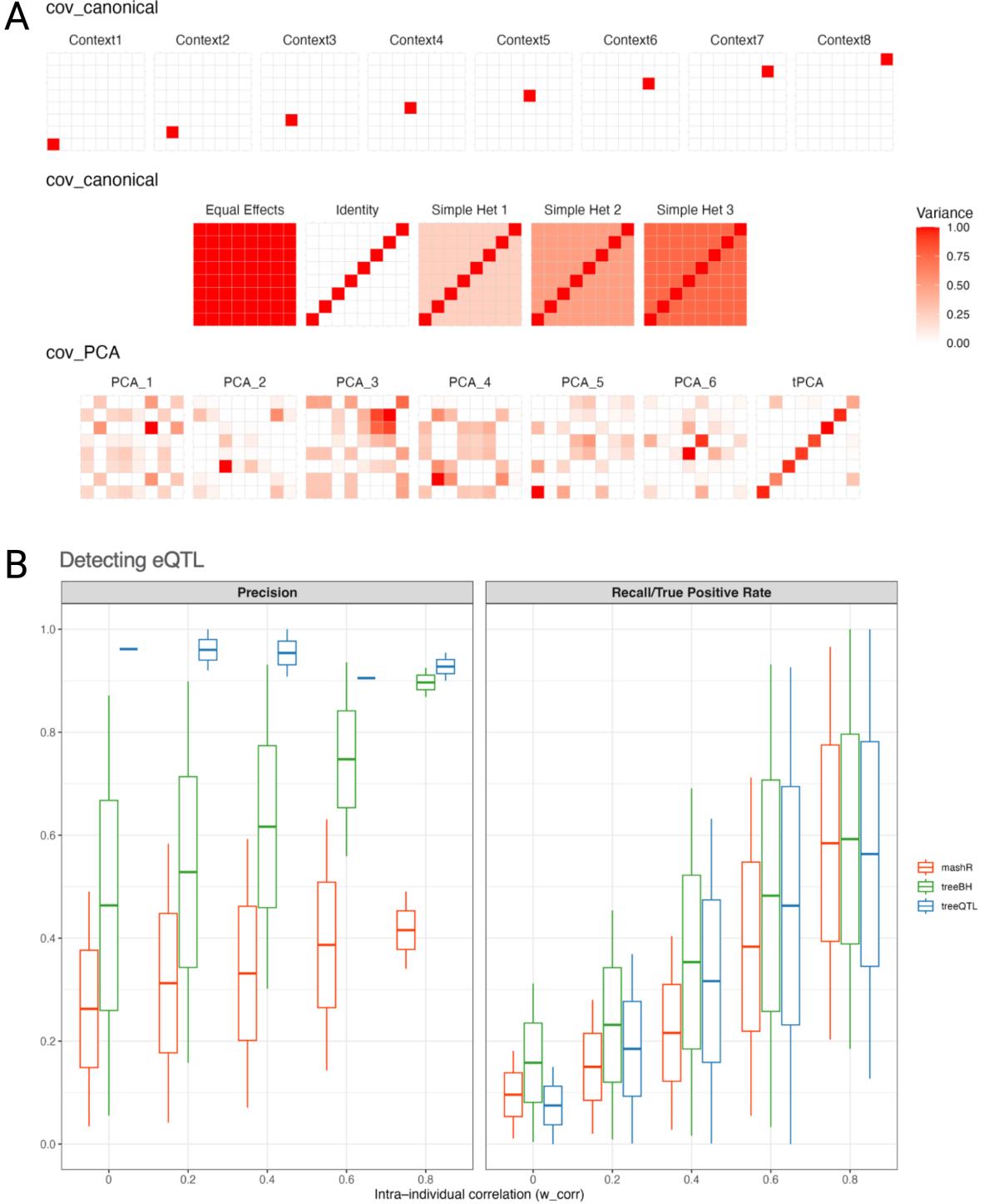
its error control capabilities. The multiple testing correction in `mashR` is grounded in an empirical Bayes framework that estimates both the effect sizes and their uncertainty across multiple conditions. Instead of testing each hypothesis individually, `mashR` models the joint distribution of observed statistics using a mixture of multivariate normal distributions with pre-specified and data-driven covariance structures  $U_k$  like Figure 2A. This approach allows `mashR` to borrow strength across conditions and shrink noisy estimates toward shared patterns, increasing power and improving the false sign rate (FSR) and local false sign rate (lfsr) control. Unlike TreeQTL and TreeBH, which explicitly control the FDR through hierarchical hypothesis testing, `mashR` provides posterior summaries such as lfsr for each test, allowing users to set significance thresholds based on posterior probabilities rather than p-values. Conceptually, `mashR` complements TreeQTL and TreeBH by shifting the focus from hierarchical p-value correction to a model-based estimation of effect patterns, offering a more flexible and adaptive alternative for identifying shared or specific signals across multiple contexts.

### 3 Results

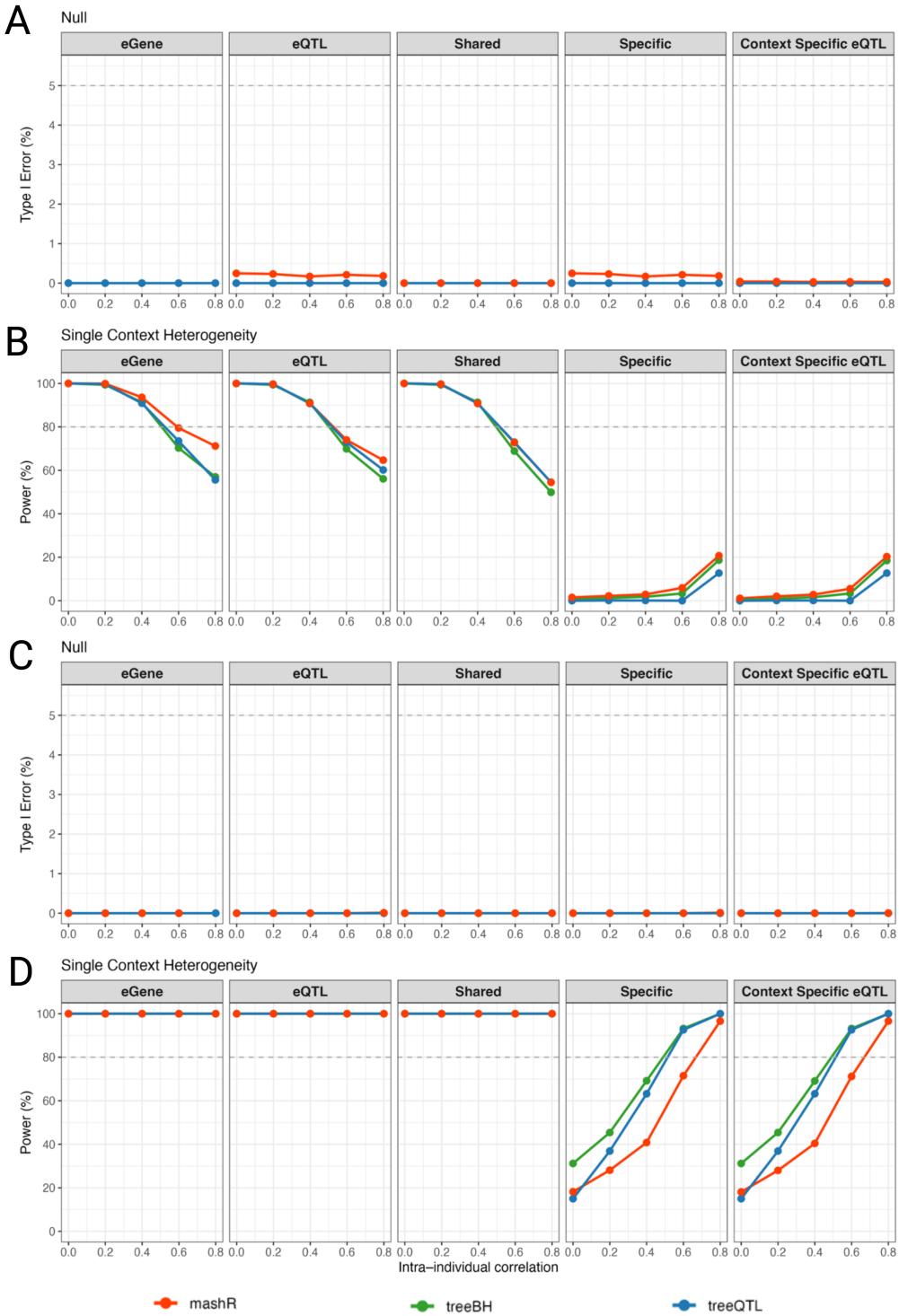
We first compared the performance of `mashR`, TreeBH, and TreeQTL in detecting eQTLs under the *single context heterogeneity* scenario, focusing on their precision and recall across varying intra-individual correlation levels (Figure 2B). TreeQTL consistently achieved high precision across all levels of intra-individual correlation, reflecting its conservative approach and strong control of false discoveries. TreeBH demonstrated moderate to high precision that improved with increasing correlation, while `mashR` showed lower precision, particularly under low correlation. In terms of recall, hierarchical correction based methods outperform empirical Bayes shrinkage based method, especially at moderate to high correlation levels. Specifically, TreeBH maintained high recall rates across all correlation levels.

To comprehensively evaluate statistical validity and power, we conducted simulation studies applying TreeQTL, TreeBH, and `mashR` under both *null* and *single context heterogeneity* scenarios, examining Type I error control and detection power across intra-individual correlation gradients (Figure 3). In the *null* scenario with 100 individuals (Figure 3A), all three methods effectively controlled Type I error in all discovery categories, including eGenes, eQTLs, shared eQTLs, specific eQTLs and context-specific eQTLs. Type I error rates remained well below the conventional 5% threshold, demonstrating the validity of each method's multiple testing correction procedure under the global null. When assessing power under *single context heterogeneity* with 100 individuals (Figure 3B), all methods showed decreasing power with increasing intra-individual correlation for detecting eGenes, eQTLs, and shared eQTLs. `mashR` achieved the highest power for shared eQTLs and eGenes across most correlation levels. For specific and context-specific eQTLs, `mashR` outperformed the other two methods, reflecting its strength in detecting context-specific effects under limited sample size and heterogeneity.

With a larger sample size of 500 individuals under the *null* (Figure 3C), Type I error



**Figure 2: Example variance-covariance matrix in `mashR` and precision and recall comparison.** **(A)** Example set up of  $U_k$  in `mashR`. Each `cov_canonical` matrix is an  $8 \times 8$  structure over contexts 1–8. Single-context matrices have a single nonzero diagonal, modeling effects specific to one context. The Equal Effects matrix models fully shared effects; the Identity matrix represents independent effects; and Simple Het 1–3 introduce increasing cross-context correlation with diagonal dominance. `cov_pca` matrices are derived from the top 6 principal components (PCs) of strong signals, resulting in 7 matrices: one using all PCs and six using each PC individually. **(B)** Comparison of precision and recall for three methods across intra-individual correlation



**Figure 3: Comparison of multiple testing correction methods: TreeQTL, TreeBH, and mashR.** (A) Simulation under the null scenario with 100 individuals, evaluating Type I error. (B) Simulation under single-context heterogeneity with 100 individuals, evaluating power. (C) Simulation under the null scenario with 500 individuals, evaluating Type I error. (D) Simulation under single-context heterogeneity with 500 individuals, evaluating power. In each subpanel, an eGene is defined as a gene with at least one associated eQTL. An eQTL refers to any significant variant identified from either the shared or specific models. Shared eQTLs are significant SNPs detected from the context-shared model. Specific eQTLs are significant SNPs identified from the context-specific models, and context-specific eQTLs are those that are significant within one specific context.

control remained stable and conservative for all three methods, further confirming their robustness. In the power evaluation with 500 individuals under *single context heterogeneity* (Figure 3D), all methods demonstrated near-maximal power for detecting eGenes, eQTLs, and shared eQTLs, regardless of correlation level. Differences emerged more clearly in the detection of specific and context-specific eQTLs. TreeBH consistently achieved higher power than TreeQTL and mashR as correlation increased.

## 4 Conclusions

In this study, we evaluated multiple testing correction strategies for context-specific eQTL detection within the FastGxC framework. We compared two hierarchical FDR control methods, TreeQTL and TreeBH, with the empirical Bayes shrinkage approach implemented in mashR, using simulation scenarios that varied in intra-individual correlation and sample size.

Our findings indicate that all three methods effectively control Type I error under the global null, validating their reliability in high-dimensional eQTL analyses. TreeQTL exhibited strong FDR control and high precision but tended to be conservative, limiting its sensitivity in detecting weaker or context-specific effects. TreeBH, by generalizing the hierarchical structure and incorporating selective FDR, achieved a better balance between precision and recall, particularly in moderate-to-high correlation settings. mashR, leveraging adaptive shrinkage and multivariate modeling, demonstrated superior power in small-sample scenarios and was particularly effective at identifying context-specific eQTLs under heterogeneous effect architectures with a small sample size.

With larger sample sizes, hierarchical methods, especially TreeBH, showed improved power and competitive performance relative to mashR, suggesting that sample size and correlation structure should inform the choice of correction method. Overall, the integration of flexible multiple testing correction into FastGxC enhances its utility for large-scale, context-aware eQTL studies, offering researchers a principled and scalable framework for detecting regulatory variation across diverse biological conditions.

## 5 Discussion

The FastGxC pipeline integrates multiple testing correction methods to enhance the accuracy and reliability of eQTL mapping<sup>38</sup>. By comparing TreeQTL<sup>44</sup>, TreeBH<sup>45</sup>, and mashR<sup>43</sup>, we demonstrate the strengths and weaknesses of hierarchical correction versus empirical Bayes shrinkage approaches. The results indicate that while hierarchical methods like TreeQTL and TreeBH excel in controlling false discovery rates<sup>45,50</sup>, mashR offers a flexible alternative that can adapt to complex data structures<sup>43</sup>.

TreeQTL implements a hierarchical testing framework with stringent control of the FDR at both the gene and SNP levels<sup>44</sup>. TreeQTL prioritizes control over false positives, which inherently reduces power in detecting weaker or context-specific effects. Its conservative nature, while yielding high precision, leads to reduced sensitivity, particularly under scenarios of high correlation or subtle effect heterogeneity.

TreeBH, as a generalized extension of TreeQTL, incorporates deeper tree structures and introduces selective FDR to adjust for the selection process at each level<sup>45</sup>. This structure enables TreeBH to adapt more flexibly to signal complexity and varying levels of effect sharing. As observed in the simulations, TreeBH consistently achieved a favorable trade-off between precision and recall, especially in detecting shared and context-specific effects at moderate to high correlation. Its advantage lies in maintaining rigorous error control while capturing more signals compared to TreeQTL, although it may still be conservative in detecting subtle, localized effects under limited sample sizes.

In contrast, mashR adopts an empirical Bayes approach that models the joint distribution of effect sizes across conditions<sup>43,46</sup>. By learning data-driven covariance structures and shrinkage patterns from the strongest signals, mashR is particularly effective in borrowing strength across conditions and stabilizing estimates. This feature explains its superior performance in detecting shared and context-specific eQTLs under limited sample sizes (e.g., 100 individuals), where hierarchical methods struggle due to reduced signal strength per context. However, mashR's probabilistic framework does not directly control FDR in the classical sense; instead, it estimates local false sign rates (lfsr), which, while more informative in some settings, may lead to reduced interpretability or precision under sparse effect architectures<sup>46</sup>.

The results from larger sample simulations (500 individuals) further highlight these trade-offs. All methods achieve strong Type I error control and high power for shared effects, but mashR's relative advantage diminishes as TreeQTL and TreeBH gain statistical power with increased data. Notably, TreeBH's strong performance in identifying context-specific signals at high correlation levels suggests that, given sufficient sample, hierarchical correction can effectively detect subtle patterns and outperform mashR.

The choice between hierarchical correction methods like TreeQTL and TreeBH versus empirical Bayes approaches like mashR should be guided by the study's design, sample size, and the expected structure of genetic effects. When stringent control of false discoveries is critical and effect architectures are relatively simple, the original TreeQTL provides robust performance<sup>44</sup>. TreeBH is better suited for settings involving complex hierarchical hypotheses or moderate heterogeneity, offering a practical balance between sensitivity and specificity<sup>45</sup>. In contrast, mashR is advantageous in high-dimensional or small-sample settings, where borrowing strength across conditions and modeling effect sharing can substantially improve power,

especially for context-specific signals<sup>43,46</sup>.

By integrating these correction methods within the FastGxC framework, we provide a scalable and modular solution that accommodates diverse biological scenarios. This flexibility empowers researchers to tailor the analysis pipeline to their data, ensuring both statistical rigor and biological relevance in the detection of eQTLs across varied experimental contexts. Despite its advantages, FastGxC has certain limitations that warrant consideration. First, the current evaluation relies primarily on simulation studies conducted under a limited number of settings due to time constraints. While these simulations capture key features of context-specific eQTL architectures, they may not fully represent the complexity and heterogeneity encountered in real-world datasets. Second, the updated FastGxC pipeline has not yet been applied to empirical data, and its performance in large-scale, biologically diverse datasets remains to be assessed. As a result, the generalizability and practical utility of the framework across tissues, platforms, and population structures are yet to be validated. Moreover, the computational scalability of the current implementation may still pose challenges for extremely high-dimensional contexts, such as multi-modal single-cell eQTL mapping, without further optimization.

Future work will focus on enhancing the FastGxC framework across several dimensions. We will improve the computational efficiency of TreeBH and mashR through parallelization and memory-optimized algorithms, enabling the pipeline to scale to biobank-scale and single-cell datasets with high dimensionality<sup>51,52</sup>. We will also explore adaptive prior learning strategies in mashR to better accommodate signal heterogeneity observed in empirical data<sup>43</sup>. To validate the pipeline's practical utility, we will apply it to large-scale eQTL datasets from consortia such as GTEx<sup>51</sup> and eQTLGen<sup>53</sup>, benchmarking the performance of TreeQTL, TreeBH, and mashR across tissues, cell types, and environmental conditions. These analyses will inform method selection guidelines and motivate development of hybrid strategies that combine the interpretability of hierarchical testing with the flexibility of empirical Bayes shrinkage<sup>45,46</sup>. Lastly, we aim to deliver a user-friendly software package with modular components for data processing, model fitting, and visualization, fostering broad adoption of FastGxC in the genomics research community<sup>54,55</sup>.

In conclusion, the updated FastGxC pipeline represents a significant advancement in eQTL mapping methodologies, providing a flexible and robust framework for detecting genetic effects across multiple contexts. By integrating hierarchical correction methods and empirical Bayes shrinkage approaches, FastGxC empowers researchers to uncover complex genetic architectures with greater accuracy and reliability, paving the way for deeper insights into gene regulation and its implications for health and disease.

## Acknowledgments

The author gratefully acknowledges the guidance and support of Dr. Brunilda Balliu and Lena Krockenberger from the Balliu Lab, whose insights were instrumental to the development of this project. The author also thanks Anthony Zhao for his collaboration and contributions related to the TreeBH methodology.

## Data and Code Availability

The FastGxC software, along with all data and code used to reproduce the figures in this manuscript, is publicly available at: <https://github.com/BalliuLab/FastGxC>.

## References

- [1] Emil Uffelmann, Qin Qin Q. Huang, Natasha S. Munung, et al. Genome-wide association studies. *Nature Reviews Methods Primers*, 1:59, 2021.
- [2] Teri A. Manolio. Genomewide association studies and assessment of the risk of disease. *New England Journal of Medicine*, 363(2):166–176, 2010.
- [3] Thomas A. Pearson and Teri A. Manolio. How to interpret a genome-wide association study. *JAMA*, 299(11):1335–1344, March 2008.
- [4] Camille Tcheandjieu, Xiaotong Zhu, Ashley T. Hilliard, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nature Medicine*, 28:1679–1692, 2022.
- [5] Angli Xue, Yang Wu, Zhihong Zhu, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications*, 9(1):2941, 2018.
- [6] Elliot Sollis, Abayomi Mosaku, et al. The nhgri-ebi gwas catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, 51(D1):D977–D985, Nov 2022.
- [7] Matthew T. Maurano, Richard Humbert, et al. Systematic localization of common disease-associated variation in regulatory dna. *Science*, 337(6099):1190–1195, Sep 2012.
- [8] Eric R. Gamazon, Ayellet V. Segrè, and et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics*, 50(7):956–967, Jul 2018.
- [9] Alexander Gusev, S. Hong Lee, and et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, 95(5):535–552, Nov 2014.
- [10] Dan L. Nicolae, Eric R. Gamazon, and et al. Trait-associated snps are more likely to be eqtls: Annotation to enhance discovery from gwas. *PLoS Genetics*, 6(4):e1000888, Apr 2010.
- [11] Ting Qi, Liyang Song, Yazhou Guo, Chang Chen, and Jian Yang. From genetic associations to genes: methods, applications, and challenges. *Trends in Genetics*, 40(8):642–667, 2024.
- [12] Noah J. Connally, Sumaiya Nazeen, et al. The missing link between genetic association and regulatory function. *eLife*, 11, 2022.
- [13] Benjamin D. Umans, Alexis Battle, and Yoav Gilad. Where are the disease-associated eqtls? *Trends in Genetics*, 37(2):109–124, 2021.

- [14] Hakhamanesh Mostafavi, Jeffrey P. Spence, et al. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nature Genetics*, 55(11):1866–1875, 2023.
- [15] Frank W. Albert and Leonid Kruglyak. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, 16:197–212, 2015.
- [16] George N. Goulielmos, Maria I. Zervou, Effie Myrthianou, Agata Burska, Timothy B. Niewold, and Frederique Ponchel. Genetic data: The new challenge of personalized medicine, insights for rheumatoid arthritis patients. *Gene*, 583(2):90–101, 2016.
- [17] The GTEx Consortium. The gtex consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, September 2020. Accessed: 2020-11-04.
- [18] Masashi Fujita, Zongmei Gao, et al. Cell subtype-specific effects of genetic variation in the alzheimer’s disease brain. *Nature Genetics*, 56(4):605–614, March 2024.
- [19] Richard K. Perez, M. Grace Gordon, et al. Single-cell rna-seq reveals cell type–specific molecular and genetic associations to lupus. *Science*, 376(6589):eabf1970, April 2022.
- [20] Seyhan Yazar, Jose Alquicira-Hernandez, et al. Single-cell eqtl mapping identifies cell type–specific genetic control of autoimmune disease. *Science*, 376(6589):eabf3041, April 2022.
- [21] Benjamin P. Fairfax, Peter Humburg, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, 343(6175), March 2014.
- [22] Chun Jimmie Ye, Jenny Chen, et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of erap2 transcripts under balancing selection. *Genome Research*, 28(12):1812–1825, December 2018.
- [23] Diego Calderon, Michelle L. T. Nguyen, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Genetics*, 51(10):1494–1505, September 2019.
- [24] Maria Gutierrez-Arcelus, Yuriy Baglaenko, et al. Allele-specific expression changes dynamically during t cell activation in hla and other autoimmune loci. *Nature Genetics*, 52(3):247–253, February 2020.
- [25] Nana Matoba, Brandon D. Le, et al. Stimulating wnt signaling reveals context-dependent genetic effects on gene regulation in primary human neural progenitors. *Nature Neuroscience*, 27(12):2430–2442, September 2024.
- [26] Ino H. Jonkers and Cisca Wijmenga. Context-specific effects of genetic variants associated with autoimmune disease. *Human Molecular Genetics*, 26(R2):R185–R192, October 2017.

- [27] Yoav Gilad, Scott A. Rifkin, and Jonathan K. Pritchard. Revealing the architecture of gene regulation: the promise of eqtl studies. *Trends in Genetics*, 24(8):408–415, August 2008.
- [28] Meng Chen and Christina Kendziorski. A statistical framework for expression quantitative trait loci mapping. *Genetics*, 177(2):761–771, October 2007. Epub 2007 Jul 29.
- [29] Benjamin J. Strober, Karl Tayeb, et al. Surge: uncovering context-specific genetic regulation of gene expression from single-cell rna sequencing using latent-factor models. *Genome Biology*, 25(1), 2024.
- [30] Anna S. E. Cuomo, Tobias Heinen, et al. Cellregmap: a statistical framework for mapping context-specific regulatory variants using scrna-seq. *Molecular Systems Biology*, 18(8), 2022.
- [31] Wancen Mu, Hirak Sarkar, et al. Airport: interpretable statistical models for analyzing allelic imbalance in single-cell datasets. *Bioinformatics*, 38(10):2773–2780, 2022.
- [32] Tobias Heinen, Stefano Secchia, et al. scdali: modeling allelic heterogeneity in single cells reveals context-specific genetic regulation. *Genome Biology*, 23(1):8, 2022.
- [33] Guanghao Qi, Benjamin J. Strober, et al. Single-cell allele-specific expression analysis reveals dynamic and cell-type-specific regulatory effects. *Nature Communications*, 14(1), 2023.
- [34] Lingfei Wang. Airqlt dissects cell state-specific causal gene regulatory networks with efficient single-cell eqtl mapping. *bioRxiv*, January 2025.
- [35] Natsuhiko Kumasaka, Raghd Rostom, et al. Mapping interindividual dynamics of innate immune response at single-cell resolution. *Nature Genetics*, 55(6):1066–1075, June 2023.
- [36] Wei Zhou, Anna S. E. Cuomo, et al. Efficient and accurate mixed model association tool for single-cell eqtl analysis. *medRxiv*, May 2024.
- [37] Zixuan Eleanor Zhang, Artem Kim, et al. Efficient count-based models improve power and robustness for large-scale single-cell eqtl mapping. *bioRxiv*, January 2025.
- [38] Lena Krockenberger, Andrew Lu, Mike Thompson, Brunilda Balliu, et al. Fastgxc: Fast and powerful context-specific eqtl mapping in bulk and single-cell data. *bioRxiv*, March 2025.
- [39] Kimberly R. Kukurba, Princy Parsana, and et al. Impact of the x chromosome and sex on regulatory variation. *Genome Research*, 26(6):768–777, 2016.
- [40] Brunilda Balliu, Matthew Durrant, and et al. Genetic regulation of gene expression and splicing during a 10-year period of human aging. *Genome Biology*, 20(1):230, Nov 2019.

- [41] Meritxell Oliva, Manuel Muñoz-Aguirre, and et al. The impact of sex on gene expression across human tissues. *Science*, 369(6509), Sept 2020.
- [42] Andrey A. Shabalin. Matrix eqtl: ultra fast eqtl analysis via large matrix operations. *Bioinformatics*, 28(10):1353–1358, May 2012.
- [43] Stephen M Urbut, Gao Wang, Peter Carbonetto, and Matthew Stephens. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics*, 51:187–195, 2019.
- [44] Christine B. Peterson, Marina Bogomolov, Yoav Benjamini, and Chiara Sabatti. Treeqtl: hierarchical error control for eqtl findings. *Bioinformatics*, 32(16):2556–2558, August 2016.
- [45] Marina Bogomolov, Chris B Peterson, Yoav Benjamini, and Chiara Sabatti. Hypotheses on a tree: new error rates and testing strategies. *Biometrika*, 108(3):575–590, 2021.
- [46] Matthew Stephens. False discovery rates: a new deal. *Biostatistics*, 18(2):275–294, April 2017.
- [47] Lee J. Cronbach and N. M. Webb. Between-class and within-class effects in a reported aptitude \* treatment interaction: Reanalysis of a study by g. l. anderson. *Journal of Educational Psychology*, 67(6):717–724, 1975.
- [48] Xiao-Li Meng and David Van Dyk. The em algorithm—an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Methodological)*, 59(3):511–567, 1997.
- [49] Sanat K. Sarkar and Chung-Kuei Chang. The simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, 92(440):1601–1608, 1997.
- [50] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300, 1995.
- [51] GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, Sep 2020.
- [52] Tabula Muris Consortium and et al. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372, Oct 2018.
- [53] Urmo Võsa, Annique Claringbould, Harm-Jan Westra, and et al. Large-scale cis- and trans-eqtl analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics*, 53(9):1300–1310, Sep 2021.

- [54] Robert C. Gentleman, Vincent J. Carey, Douglas M. Bates, and et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80, 2004.
- [55] Steffen Durinck, Paul T. Spellman, Ewan Birney, and et al. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature Protocols*, 4(8):1184–1191, 2009.