Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Incorporating cell states for mapping eQTLs in single-cell studies

Yao Lu
Tsinghua University
Nayang Shan
Capital University of Economics and Business
Zuoheng Wang
Yale University
Lin Hou
houl@tsinghua.edu.cn

Tsinghua University

Method Article

Keywords: eQTL, single-cell sequencing, retrospective association test, functional mixed model

Posted Date: September 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10993235/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Incorporating cell states for mapping eQTLs in single-cell studies

Yao Lu[1], Nayang Shan[2], Zuoheng Wang[3,4,*], and Lin Hou[1,5,*]

[1]Department of Statistics and Data Science, Tsinghua University, Beijing, China

[2]School of Statistics and Data Science, Capital University of Economics and Business, Beijing, China

[3]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

[4]Department of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA

[5]MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China

[*]Correspondence to: houl@tsinghua.edu.cn and zuoheng.wang@yale.edu

September 9, 2026

Abstract

Single-cell expression quantitative trait loci (eQTL) studies have revealed genetic regulation of gene expression with cell-type specificity. However, methods for genome-wide eQTL mapping that account for cell-state-dependent effects remain limited. To fill this gap, we propose scarf-QTL, which maps eQTLs across cell states and implements retrospective association tests for robust statistical inference. Simulations showed that scarf-QTL controls type I error rates and improves statistical power compared with pseudobulk methods. Applied to the OneK1K dataset, scarf-QTL identified 30% more cell-type-specific eGenes (genes with eQTLs) than the pseudobulk analysis and revealed distinct cell-state-dependent eQTL patterns in dendritic cells.

Keywords: eQTL, single-cell sequencing, retrospective association test, functional mixed model

1 Introduction

Expression quantitative trait loci (eQTLs) are genetic variants that regulate gene expression [1]. Bulk eQTL studies have advanced our understanding of gene regulation across tissues [2, 3] and cell types [4, 5], providing insights into the molecular mechanisms of disease-associated variants [6, 7]. Recently, single-cell eQTL (sc-eQTL) studies based on single-cell RNA sequencing (scRNA-seq) in genotyped populations have enabled direct characterization of cell-type-specific eQTLs [8–12]. By measuring

bulk or single-cell gene expression at multiple time points, some eQTLs have been demonstrated to have dynamic effects along continuous biological processes like cell differentiation [13–18]. Moreover, sc-eQTL studies have shown that eQTL effects can vary across fine-grained cell states within cell types [19, 20]. In practice, such cell states are often represented by discrete cell subpopulations, continuous pseudotime trajectories [12, 21], low-dimensional embeddings [19, 20, 22], and gene signature scores [23].

Typically, sc-eQTL analyses follow a pseudobulk strategy, aggregating gene expression across cells of predefined cell populations for each individual and testing for association between genetic variants and pseudobulk expression levels [8, 17]. Recent methods model gene expression at the single-cell level using mixed-effects models. For example, Nathan *et al.* and Kang *et al.* assumed that the read counts of a gene in a cell follow Poisson or negative binomial distributions with individual-level random effects, thus accommodating the sparsity of single-cell measurements and the within-individual correlation [19, 22]. Alternative frameworks assumed approximate normality for transformed expression values and then applied linear mixed models [24–26]. Moreover, several sc-eQTL methods consider cell-state-dependent genetic regulation by modeling interactions between cell state and genotype [24, 25].

As more and more sc-eQTL datasets become available [27, 28], critical challenges for sc-eQTL mapping remain. First, pseudobulk-based analyses may have limited power when genetic effects vary across cell states, because aggregating cells into overly coarse groups can attenuate such effects, whereas defining overly fine groups may lead to too few cells in many individuals [29]. Moreover, when the numbers

of cells differ substantially across individuals, pseudobulk gene expression can become heteroskedastic, further reducing statistical power [29, 30]. Second, violations of model assumptions remain a concern, especially for the normality assumption, because scRNA-seq data are sparse and the distributions of read counts vary across genes and datasets. Such model violations can compromise error control of association tests [22, 24, 30, 31]. Third, most existing single-cell models, particularly those accounting for cell-state-dependent effects, are computationally intensive. Consequently, cell-state-dependent eQTL analysis is often performed as a *post hoc* step following initial genome-wide eQTL screening [19, 24, 26, 32].

Here, we present scarf-QTL (**S**calable single-**C**ell **A**ssociation with **R**etrospective testing and **F**unctional mixed model for e**QTL**), a statistical framework for genome-wide eQTL mapping that accommodates both constant and cell-state-dependent genetic effects. Scarf-QTL represents cellular variation using a one-dimensional cell-state coordinate, uses a functional mixed model of gene expression along this coordinate to construct score functions, and performs retrospective association tests for robust eQTL mapping. Simulation studies show that scarf-QTL maintains well-calibrated type I error rates even when gene expression counts are generated from Poisson or negative binomial models, and achieves high power for eQTLs with constant effects and diverse patterns of cell-state-dependent effects, while remaining computationally efficient and scalable. Applied to the OneK1K dataset [12], scarf-QTL identifies 30% more cell-type-specific eGenes than the pseudobulk approach and uncovers cell-state-dependent genetic regulation, exemplified by eQTLs exhibiting distinct regulatory patterns across dendritic cell (DC) subsets.

2 Results

2.1 Overview of scarf-QTL

Scarf-QTL is a statistical framework for identifying associations between genetic markers and gene expression levels in sc-eQTL studies, while allowing genetic effects to vary smoothly along the cell-state coordinate (Fig. 1; methodological details are provided in the "Methods" section). The framework is especially designed to investigate the genetic regulation of genes whose expression levels may vary across cell states. Scarf-QTL represents gene expression along the cell-state coordinate using a functional mixed model [33,34], which is used to construct score functions for detecting cell-state-dependent genetic effects (Fig. 1a). Specifically, the functional mixed model decomposes gene expression into three components. The first component describes baseline expression and additive genetic effects, both represented as linear combinations of spline functions of the cell-state coordinate. The second component accounts for covariate effects, including age, sex, population stratification factors, and measures of sequencing quality. The third component consists of individual-level random effects and cell-level noise, capturing within-individual correlation with stronger correlations between cells in similar cell states, together with additional cell-specific variation.

To identify significant eQTL effects, scarf-QTL uses score functions of the functional mixed model to construct retrospective association tests [35] (Fig. 1b). The significance of retrospective association tests is evaluated by treating the genotype as a random variable conditional on the observed gene expression and covariates.

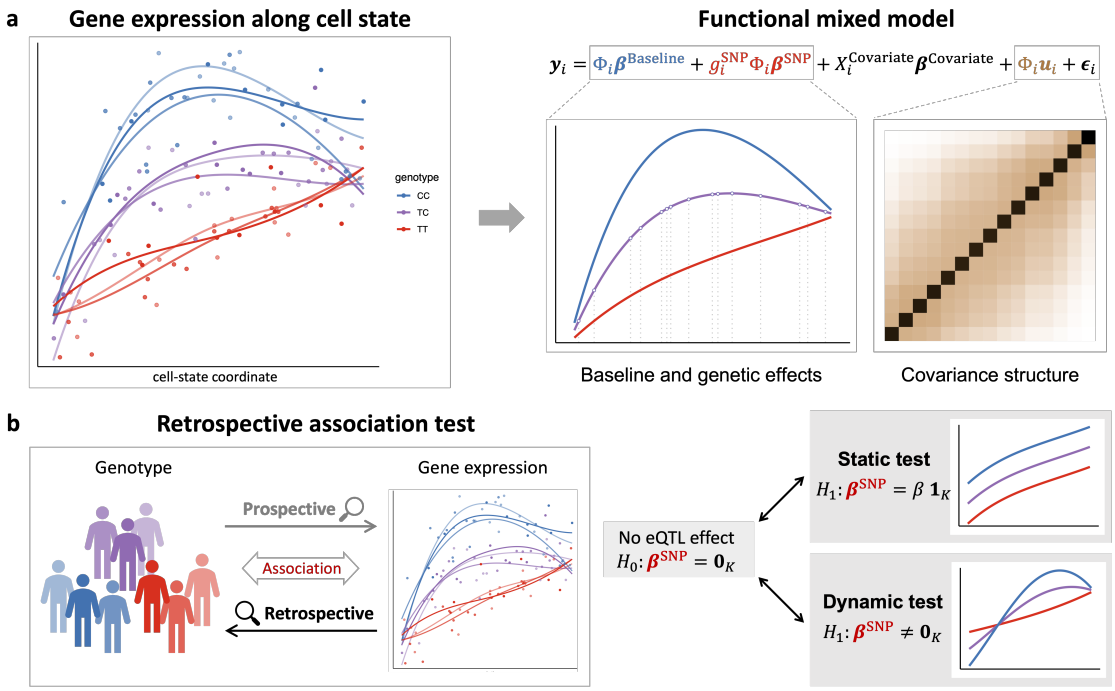


Figure 1: **Overview of scarf-QTL. (a)** Modeling gene expression along cell states. The left panel illustrates single-cell gene expression measurements (dots) from multiple individuals (colors) with different genotypes, plotted against the cell-state coordinate. The curves illustrate individual-level gene expression levels along the cell-state coordinate. Scarf-QTL models gene expression using a functional mixed model (right panel). Using B-spline basis expansions of the cell-state coordinate (Φ_i is the matrix of basis evaluations for the cells of individual i), cell-state-dependent baseline expression and genetic effects are modeled as smooth functions of the cell-state coordinate. Fixed covariate effects are included, while the individual-level random effects ($\Phi_i \mathbf{u}_i$) and the random noise ($\boldsymbol{\varepsilon}_i$) induce a cell-state-dependent covariance structure and capture additional variation. **(b)** Association testing in scarf-QTL. Scarf-QTL tests the association between genotype and gene expression using retrospective association tests. The retrospective strategy treats genotype as a random variable conditional on the observed gene expression and covariates. This retrospective formulation differs from prospective score tests built on gene expression models conditional on genotype and covariates, and provides improved robustness to misspecification of the gene expression model. Two alternative hypotheses are evaluated against a common null hypothesis of no eQTL effect. The alternative of the static test assumes a constant genetic effect across cell states, whereas the dynamic test allows genetic effects to vary across cell states.

Under the null hypothesis of no genetic effect, the conditional mean and variance of genotype do not depend on the observed gene expression. This retrospective approach is robust to misspecification of the gene expression model [35], a property that is particularly important for single-cell data, where normality assumptions are often violated. Scarf-QTL implements the association test in two modes: static and dynamic. The static mode tests a constant effect as in conventional eQTL studies. The dynamic mode tests the overall effect of a genetic marker while allowing the effect to vary along the cell-state coordinate. As both tests assess the presence of an eQTL effect but target different regulatory scenarios, we combine their p-values using the Cauchy combination method [36] to provide a unified test statistic, which we refer to as the combined test. For genetic markers with significant association, the effect sizes along the cell-state coordinate are further estimated using penalized splines [37]. The resulting smooth effect-size curves can be clustered to summarize and visualize representative patterns of cell-state-dependent genetic regulation.

2.2 Scarf-QTL controls type I error rate under Gaussian and count-based gene expression models

We conducted simulation studies to evaluate the type I error rate of scarf-QTL. As benchmarks, we also considered the pseudobulk approach (see the "Methods" section (4.6.3) for details) and the prospective score tests (see the "Methods" section (4.3.2)). Normalized expression values without genetic regulation were simulated under the normal distribution for three representative immune cell types with large ($CD4_{NC}$ cells), moderate (B_{IN} cells), and small (plasma cells) numbers of cells (see

the "Methods" section (4.6.1)). As shown in Tables 1, S1, and S2, the static and dynamic tests of scarf-QTL maintained well-calibrated type I error rates across all cell types. For the combined test, type I error was well controlled at smaller α levels, although slight inflation was observed at $\alpha = 0.05$ and 0.01. This pattern is consistent with the fact that the Cauchy combination is more accurate for sufficiently small p-values [36, 38]. Similarly, both the pseudobulk approach and the prospective score tests controlled type I error well under the normal expression model.

Table 1: Evaluation of type I error for scarf-QTL and benchmark methods

Model	Level	Scarf-QTL			Prospective Score Test		Pseudobulk
		Static	Dynamic	Combined	Static	Dynamic	
Normal	0.05	5.01×10^{-2}	4.96×10^{-2}	$\mathbf{5.20 \times 10^{-2}}$	5.04×10^{-2}	5.04×10^{-2}	4.99×10^{-2}
	0.01	1.00×10^{-2}	9.78×10^{-3}	1.02×10^{-2}	1.01×10^{-2}	1.01×10^{-2}	9.64×10^{-3}
	0.001	9.66×10^{-4}	9.33×10^{-4}	9.87×10^{-4}	9.98×10^{-4}	9.67×10^{-4}	8.86×10^{-4}
	0.0001	8.70×10^{-5}	8.00×10^{-5}	8.50×10^{-5}	9.80×10^{-5}	8.50×10^{-5}	7.30×10^{-5}
Poisson	0.05	5.01×10^{-2}	4.95×10^{-2}	$\mathbf{5.22 \times 10^{-2}}$	$\mathbf{1.88 \times 10^{-1}}$	$\mathbf{5.19 \times 10^{-2}}$	4.99×10^{-2}
	0.01	1.02×10^{-2}	9.80×10^{-3}	$\mathbf{1.03 \times 10^{-2}}$	$\mathbf{8.63 \times 10^{-2}}$	$\mathbf{1.51 \times 10^{-2}}$	1.00×10^{-2}
	0.001	1.01×10^{-3}	9.55×10^{-4}	1.00×10^{-3}	$\mathbf{3.04 \times 10^{-2}}$	$\mathbf{3.36 \times 10^{-3}}$	9.61×10^{-4}
	0.0001	1.13×10^{-4}	7.80×10^{-5}	8.90×10^{-5}	$\mathbf{1.14 \times 10^{-2}}$	$\mathbf{8.68 \times 10^{-4}}$	8.80×10^{-5}
Negative Binomial	0.05	4.98×10^{-2}	4.95×10^{-2}	$\mathbf{5.17 \times 10^{-2}}$	$\mathbf{1.20 \times 10^{-1}}$	4.80×10^{-2}	5.02×10^{-2}
	0.01	9.99×10^{-3}	9.84×10^{-3}	$\mathbf{1.03 \times 10^{-2}}$	$\mathbf{4.29 \times 10^{-2}}$	$\mathbf{1.08 \times 10^{-2}}$	9.92×10^{-3}
	0.001	9.82×10^{-4}	9.06×10^{-4}	9.31×10^{-4}	$\mathbf{1.05 \times 10^{-2}}$	$\mathbf{1.37 \times 10^{-3}}$	9.61×10^{-4}
	0.0001	1.00×10^{-4}	8.00×10^{-5}	9.40×10^{-5}	$\mathbf{2.78 \times 10^{-3}}$	$\mathbf{2.08 \times 10^{-4}}$	8.50×10^{-5}

Empirical type I error rates that fall outside the corresponding acceptance region with significance level 1% are shown in bold. Specifically, the 1% acceptance regions are $(0, 5.05 \times 10^{-2})$ for nominal level $\alpha = 0.05$, $(0, 1.02 \times 10^{-2})$ for $\alpha = 0.01$, $(0, 1.07 \times 10^{-3})$ for $\alpha = 0.001$, and $(0, 1.23 \times 10^{-4})$ for $\alpha = 0.0001$. Results are presented for the B_{IN} cell type; results for $CD4_{NC}$ and plasma cells are provided in Tables S1 and S2.

Additionally, to assess the robustness of scarf-QTL to model misspecification, we simulated raw gene expression counts from Poisson and negative binomial distributions. Under these non-Gaussian settings, scarf-QTL and the pseudobulk analysis

maintained well-calibrated type I error rates, whereas the prospective score tests showed severe inflation (Tables 1, S1, and S2). These results demonstrate that scarf-QTL maintains valid type I error control even when the working gene expression model is misspecified.

2.3 Scarf-QTL achieves high power across diverse patterns of genetic regulation

Beyond constant eQTL effects, we curated various patterns of dynamic and cell-state-dependent eQTLs from previous studies (Fig. 2A), including linear [12, 16], two-stage [16, 18, 20], peak [20, 21], and cyclic [13, 20], and evaluated the statistical power of scarf-QTL across these patterns (see the "Methods" section (4.6.1)). For benchmarking, we compared the performance of scarf-QTL with that of the pseudobulk approach.

For the static pattern, in which the regulatory effect does not depend on cell states (Fig. 2(a)), the static scarf-QTL test had the highest power, whereas the dynamic test showed reduced power, and the performance of the combined test was comparable to the static test. The pseudobulk approach exhibited lower power than the static and combined tests. Similarly, for the linear pattern (Fig. 2(b)) and the two-stage pattern (Fig. 2(c)), the static test remained the best, despite the genetic effects varying across cell states. This was expected, as the dynamic test has more degrees of freedom. However, the gap between the static and dynamic tests diminished.

The dynamic and combined tests were advantageous for the peak pattern (Fig. 2(d)) and the two-stage pattern with reverse signs (Fig. 2(e)). In the peak pattern, while

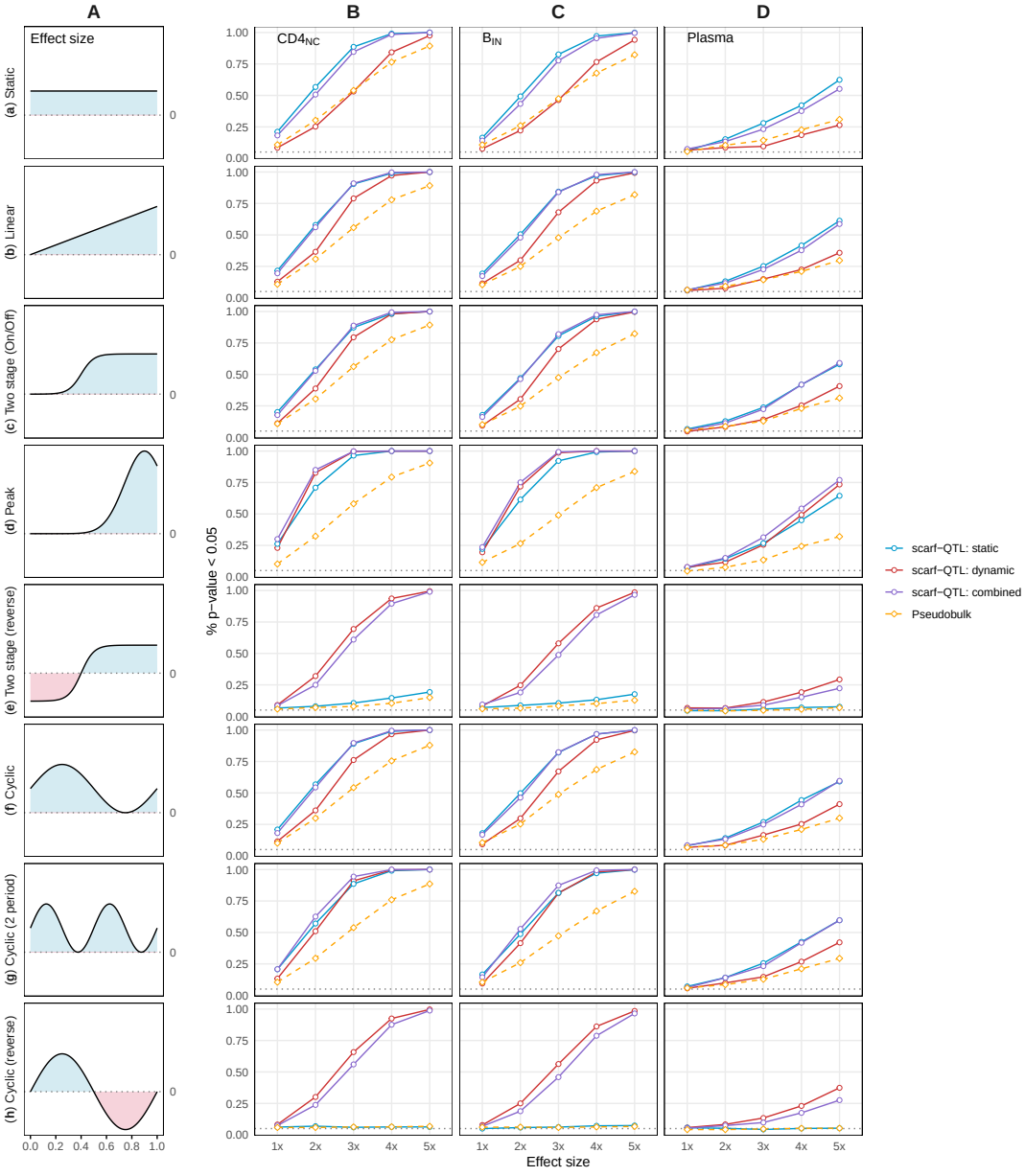


Figure 2: Power evaluation of scarf-QTL across diverse patterns of genetic regulation. **(A)** Patterns of eQTL effects used in simulations, including **(a)** static effects, **(b)** linearly increasing effects along the cell-state coordinate, **(c)** two-stage (On/Off) effects confined to a subset of cells, **(d)** peak (pulse-like) effects, **(e)** two-stage effects with reverse directions, **(f)** cyclic effects with one period, **(g)** cyclic effects with two periods, **(h)** cyclic effects with sign reversals. Curves are normalized such that the area under each curve equals one. **(B–D)** Empirical power of scarf-QTL (static, dynamic, and combined tests) and the pseudobulk approach across three representative immune cell types ($CD4_{NC}$, B_{IN} and plasma cells) at nominal level $\alpha = 0.05$.

the empirical power of the static and pseudobulk tests was similar to that in the other patterns, the dynamic and combined tests achieved markedly higher power, as they captured genetic effects concentrated in a narrow range of cell states. In the two-stage pattern with reverse signs, the loss of power in the static and pseudobulk tests was due to cancellation when the genetic effects had opposite signs across cell states. In contrast, the dynamic test retained superior power because the functional mixed model captured the varying genetic effects without cancellation across cell states.

We further examined cyclic patterns of genetic regulation (Fig. 2(f-h)), motivated by periodic regulation along the cell cycle or other oscillatory biological processes. Again, the scarf-QTL tests were more powerful than the pseudobulk approach. The cyclic patterns with consistent effect directions (Fig. 2(f,g)) behaved similarly to the two-stage pattern (Fig. 2(c)), whereas the cyclic pattern with sign reversal (Fig. 2(h)) behaved similarly to the two-stage pattern with reverse signs (Fig. 2(e)).

We also simulated gene expression levels using Poisson and negative binomial models, and the superior performance of scarf-QTL over the pseudobulk approach remained consistent (Fig. S1 and S2). In practice, we recommend using the combined test, especially when the underlying genetic model is unknown, as it leverages the strengths of both the static and dynamic tests and consistently achieves the best or second-best power across different settings. In addition to hypothesis testing, scarf-QTL recovers specific patterns of cell-state-dependent effects, as the estimated effect-size curves align well with the true patterns (Fig. S3).

2.4 Scarf-QTL identifies more cell-type-specific eGenes and reveals cell-state-dependent genetic regulation in the OneK1K dataset

We applied scarf-QTL to the OneK1K dataset [12], which comprises 1,267,758 peripheral blood mononuclear cells profiled by scRNA-seq from 982 genotyped healthy individuals across 14 annotated immune cell types. Here, an eGene is defined as a gene with at least one significant eQTL at a false discovery rate (FDR) threshold of 5%, implemented using local false discovery rates [39]. Across these cell types, scarf-QTL identified more eGenes than the pseudobulk approach [12], especially with the combined test (Fig. 3a and S4), while all these tests showed extensive overlap with each other (Fig. 3a). We detected more eGenes in cell types with larger numbers of cells, likely reflecting increased statistical power (Fig. S4). Overall, the combined test identified 21,558 eGenes across cell types, representing a substantial increase of approximately 30% over the 16,597 eGenes previously reported [12].

Next, we investigated whether scarf-QTL enables the detection of cell-state-dependent genetic regulatory effects. We took dendritic cells as an example, as they comprise multiple specialized subtypes including conventional DCs (cDCs) and plasmacytoid DCs (pDCs). Villani *et al.* identified six distinct DC subsets and reported discriminative marker genes for these subsets [40]. Although our inferred cell-state coordinate did not perfectly recapitulate these discrete subsets, the expression of the marker genes varied systematically along this coordinate, broadly separating pDCs and cDCs (Fig. S5). Thus, the cell-state coordinate provided a useful surrogate for

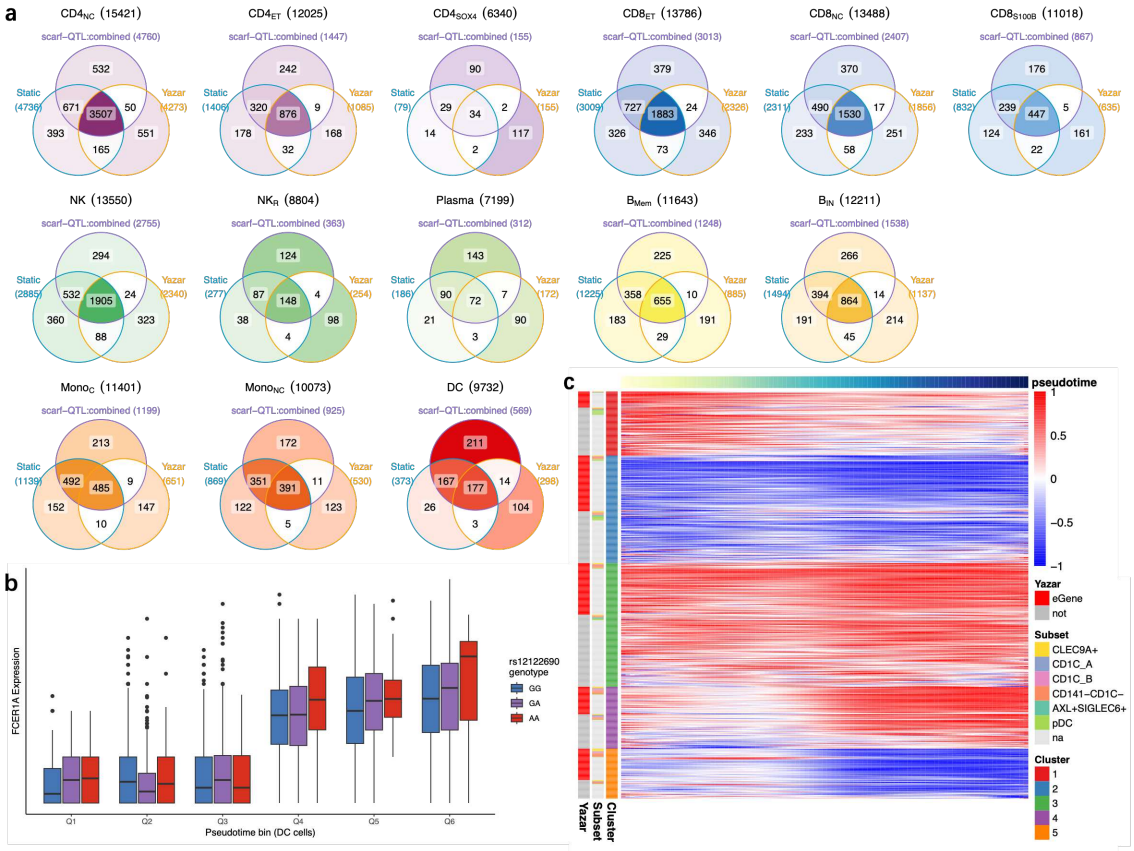


Figure 3: eQTL discovery by scarf-QTL in the OneK1K dataset. **(a)** Comparison of eGenes identified by three approaches across 14 immune cell types. Each Venn diagram shows the overlap among eGenes identified by the pseudobulk approach [12] (labeled Yazar), the static test of scarf-QTL, and the combined test of scarf-QTL. **(b)** Cell-state-dependent regulatory effect of variant rs12122690 on *FCER1A* expression in dendritic cells. Cells were grouped along the cell-state coordinate, and expression values were summarized per individual within each group. Boxplots display expression across genotypes. Here and in subsequent figures, pseudotime is used as the one-dimensional cell-state coordinate, without requiring a temporal interpretation. **(c)** Cell-state-dependent regulatory patterns of eQTLs (rows) identified by the combined test in dendritic cells. The heatmap shows fitted eQTL effect sizes as smooth functions of the cell-state coordinate. eQTLs were clustered by k-means based on their effect-size curves. For visualization, one eQTL per eGene is shown within each cluster, and the effect curves are scaled so that the maximum absolute effect size equals 1. Yazar: eGenes identified by the pseudobulk approach in DCs [12]. Subset: DC subset marker genes [40].

DC subset structure. Because these subset markers capture subset-specific transcriptional programs, they represented biologically motivated candidates for investigating cell-state-dependent genetic regulation across DC subsets [41].

Notably, the eGenes identified by the combined test of scarf-QTL were significantly enriched for DC subset markers (odds ratio = 1.89; one-sided Fisher's exact test p-value $= 1.7 \times 10^{-6}$). As an illustrative example, *FCER1A*, a marker of CD1C$^+$ conventional DCs [40], showed increasing expression along the cell-state coordinate in the OneK1K dataset. Its association with rs12122690 became more evident at larger values of the cell-state coordinate (Fig. 3b; combined test p-value $= 5.9 \times 10^{-7}$). This gene was not identified as an eGene by the pseudobulk approach (minimum p-value $= 3.3 \times 10^{-6}$ across variants; local FDR = 0.080), illustrating how scarf-QTL captures genetic effects that vary across cell states and are attenuated in pseudobulk analysis.

To further characterize cell-state-dependent genetic regulation in DCs, we estimated and clustered eQTL effect-size curves. Five major clusters were identified (Fig. 3c), representing distinct regulatory patterns. Clusters 4 and 5 exhibited increasing effect sizes along the cell-state coordinate, in the positive and negative directions, respectively, and genes appearing in these clusters were strongly enriched for conventional DC subset markers (odds ratio = 6.12; one-sided Fisher's exact test p-value $= 1.4 \times 10^{-5}$). These genes were also enriched for antigen processing and presentation and broader immune-related GO biological processes; for example, the term "antigen processing and presentation" showed a fold enrichment of 2.87 (FDR $= 3.5 \times 10^{-5}$), and the overlapping genes were dominated by MHC class II genes.

This pattern is consistent with the antigen-presenting function of conventional dendritic cells. In contrast, cluster 1 showed decreasing effect sizes, and genes appearing in this cluster were enriched for plasmacytoid DC subset markers (odds ratio = 2.83; one-sided Fisher's exact test p-value = 0.019). Clusters 2 and 3 showed approximately constant effect sizes with opposite directions, corresponding to static eQTLs. Similar patterns were observed in other cell types (Fig. S6, S7).

2.5 Scalability of scarf-QTL for large-scale sc-eQTL studies

Scarf-QTL employs several algorithmic optimizations to reduce the computational cost of null model fitting and retrospective association testing (see the "Methods" section (4.4)). To assess empirical runtime, we performed resampling from the OneK1K dataset by varying the numbers of variants, individuals, and cells. Under a typical cis-eQTL setting (around 3,000 variants per gene), the total runtime of scarf-QTL was comparable to performing Spearman correlation tests in the pseudobulk approach (Fig. 4a). As the number of variants increased, scarf-QTL became faster than the pseudobulk approach, because the computationally intensive null model fitting is performed only once per gene, while retrospective association testing adds relatively little computational cost.

We then varied the number of individuals and the number of cells per individual to assess how runtime scales with sample size (Fig. 4b). When the number of cells per individual was 200, the total runtime of scarf-QTL was comparable to that of the pseudobulk approach across the numbers of individuals considered. Increasing the number of cells per individual increased the runtime of scarf-QTL but not that

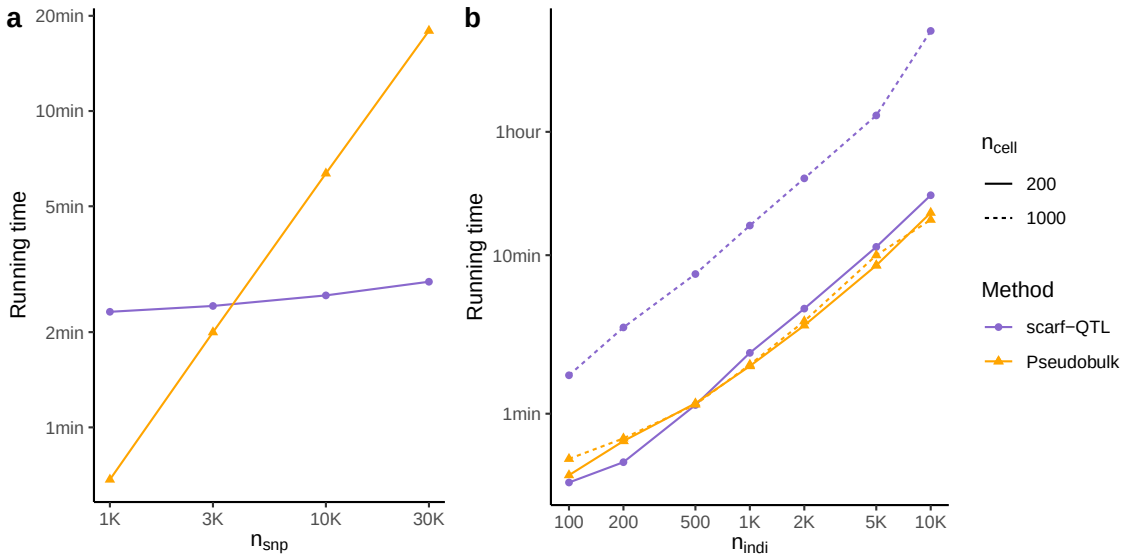


Figure 4: Runtime scalability of scarf-QTL compared with the pseudobulk approach. Shown are empirical runtimes for eQTL mapping in resampled datasets with 100 genes. **(a)** Runtime is shown as a function of the number of variants per gene (n_{snp}), while fixing the number of individuals (n_{indi}) at 1,000 and the number of cells per individual (n_{cell}) at 200. **(b)** Runtime is shown as a function of the number of individuals (n_{indi}) under different numbers of cells per individual (n_{cell}), while fixing n_{snp} at 3,000. Both axes are shown on logarithmic scales.

of the pseudobulk approach, as expected because scarf-QTL directly models single-cell expression, whereas the pseudobulk approach operates on individual-level gene expression. Importantly, the runtime of scarf-QTL showed only an approximately linear increase with the total number of cells ($n_{\text{indi}} \times n_{\text{cell}}$; Fig. S8).

3 Discussion

Here, we present scarf-QTL, a statistically rigorous and computationally efficient framework for eQTL mapping in large-scale single-cell studies. Compared with pseudobulk approaches, scarf-QTL improves power by directly leveraging single-cell expression variation, revealing cell-state-dependent genetic regulation while maintaining well-calibrated type I error rates. The retrospective association tests in scarf-QTL are theoretically robust to misspecification of the gene expression model, which is important because no single distributional assumption is expected to fully capture scRNA-seq expression across genes, cell types, and cellular contexts [19, 22]. In practice, the retrospective formulation allows scarf-QTL to use a computationally efficient linear modeling framework while preserving valid type I error control, even when the gene expression distribution deviates from the normality assumption.

The power gains of scarf-QTL over pseudobulk methods observed in both simulations and real data likely reflect several complementary advantages. First, modeling gene expression at the single-cell level improves sensitivity by avoiding variance distortions that can arise from unbalanced cell numbers in pseudobulk approaches [29]. Similar advantages of single-cell modeling have been reported previously [31]. Sec-

ond, the dynamic test leverages cell-to-cell variation to capture variation in baseline expression and genetic effects across cell states. Our results show that scarf-QTL detects nonlinear and cell-state-dependent eQTL effects along the inferred cell-state coordinate, which may capture continuous biological variation or discrete cell subsets. Third, the combined test integrates evidence from the static and dynamic tests, allowing genome-wide eQTL screening sensitive to both constant and cell-state-dependent effects, whereas existing approaches often relegate cell-state-dependent eQTL analysis to a *post hoc* step. Together, these features explain the improved power of scarf-QTL while preserving a principled testing framework.

The cell-state-dependent eQTLs identified in dendritic cells illustrate how predefined cell grouping in pseudobulk analyses can obscure biologically meaningful heterogeneity, particularly for rare cell populations, and how scarf-QTL mitigates this limitation. DCs are relatively rare in peripheral blood, which likely limits the feasibility of further stratifying them into subsets such as cDCs and pDCs in pseudobulk-based eQTL analyses, as such approaches rely on sufficient numbers of cells to achieve adequate statistical power. When DCs are treated as a single cell group, genetic regulation specific to particular DC subsets may already be lost during the initial pseudobulk screening, and therefore cannot be recovered by subsequent analyses of cell-state-dependent effects restricted to eQTLs with robust main effects. This limitation is illustrated by *FCER1A*, whose genetic effect is substantially stronger at cDC-like cell states, resulting in an association that fails to pass multiple-testing correction under the pseudobulk framework. In contrast, scarf-QTL recovers such associations by modeling heterogeneity in gene expression and genetic regulation

along the inferred cell-state coordinate. Accordingly, the number of detected eGenes in DCs increased from 298 under the pseudobulk analysis to 569 using scarf-QTL, representing more than a 90% increase. Nevertheless, downstream analyses, including estimation of effect-size curves, remain essential for mechanistic interpretation of cell-state-dependent genetic regulation.

There are several limitations of the current work. First, while scarf-QTL allows eQTL effects to vary across cell states, it does not directly test whether an eQTL effect varies across cell states, as specifying the retrospective genotype distribution under a null model with a static eQTL effect is challenging. Second, the retrospective association test is grounded in large-sample theory and relies on assumptions about genotype distributions under the null hypothesis. In settings with limited sample sizes or rare variants, more tailored approximations to the null distribution of the score test statistics may be needed, rather than relying directly on a chi-square approximation. Third, the retrospective framework does not directly provide point estimates of eQTL effects, so it is not straightforward to perform downstream analyses such as fine mapping, colocalization, or Mendelian randomization. Finally, modeling cell-state-dependent effects in scarf-QTL requires somewhat arbitrary choices regarding the specification of the cell-state coordinate and its basis representation. Although these choices do not compromise type I error control under the null hypothesis, they may influence statistical power and the interpretation of cell-state-dependent eQTL effects. Future work could consider data-adaptive basis selection and sensitivity analyses with respect to cell-state inference to assess the robustness of eQTL detection in specific datasets.

In conclusion, scarf-QTL provides a scalable framework for detecting genetic regulation while allowing eQTL effects to vary across cell states. Scarf-QTL can be applied to increasingly large and diverse sc-eQTL datasets, and its retrospective framework may also be extended to other molecular QTL studies.

4 Methods

4.1 Notations and model

For a given gene and SNP, suppose that normalized gene expression levels $\boldsymbol{y}$ are available in C cells from I individuals,

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_I \end{pmatrix}, \ \boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{ic_i} \end{pmatrix}, \ i = 1, \ldots, I, \tag{1}$$

where c_i denotes the number of cells from individual i, and $C = \sum_{i=1}^{I} c_i$. Each cell is assigned a one-dimensional cell-state coordinate t_{ic} inferred from its transcriptomic profile. In this work, pseudotime inference is used to obtain this coordinate.

Scarf-QTL models the normalized gene expression across the continuous cell-state coordinate using a functional mixed model [33, 34]:

$$y_{ic} = \boldsymbol{x}_{ic}^{\top} \boldsymbol{\gamma}(t_{ic}) + \alpha_i(t_{ic}) + \epsilon_{ic}, \ i = 1, \ldots, I; \ c = 1, \ldots, c_i, \tag{2}$$

where $\boldsymbol{x}_{ic}$ is a vector of predictors, $\boldsymbol{\gamma}(t)$ is a vector of smooth functions over the

cell-state coordinate for cell-state-dependent fixed coefficients, $\alpha_i(t)$ is an individual-specific random smooth function, and $\epsilon_{ic} \sim \mathcal{N}(0, \tau^{-1})$ is independent noise. Here, the predictor vector $\boldsymbol{x}_{ic}$ includes an intercept term (a constant 1) for baseline expression, the genotype g_i^{SNP} of the candidate SNP, and a covariate vector $\boldsymbol{w}_{ic} \in \mathbb{R}^J$, which comprises individual-level covariates (e.g., age, sex, genotype principal components, and PEER factors) and cell-level covariates (e.g., mitochondrial RNA percentage and UMI counts).

We represent both the smooth fixed effects for baseline expression and genotype effects and the random individual effects using cubic B-splines with K basis functions over equally spaced knots, denoted as $\boldsymbol{\phi}(t) = [\phi_1(t), \ldots, \phi_K(t)]^\top$. Throughout the paper, we fix $K = 10$, although K can also be chosen according to the number of cells or via cross-validation [34]. Coefficients of covariates are assumed to be constant over the cell-state coordinate, although extending to cell-state-dependent covariate effects is straightforward and described in the supplementary materials. This formulation yields the following functional mixed model:

$$y_{ic} = \sum_{k=1}^{K} \phi_k(t_{ic}) \left(\beta_k^{\text{Baseline}} + g_i^{\text{SNP}} \beta_k^{\text{SNP}} \right) + \sum_{j=1}^{J} w_{ic,j} \beta_j^{\text{Covariate}} + \sum_{k=1}^{K} \phi_k(t_{ic}) u_{ik} + \epsilon_{ic}, \quad (3)$$

where $u_{ik} \sim \mathcal{N}(0, \eta\tau^{-1})$ are independent individual-level random effects. Under this model, the expectation of gene expression is decomposed into a cell-state-dependent baseline expression pattern and a cell-state-dependent genetic effect, together with constant covariate effects. Furthermore, the model induces stronger within-individual correlation between cells with similar cell-state coordinates.

The observed gene expression of the candidate gene in all cells from individual i can be expressed in matrix form as

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \Phi_i\boldsymbol{u}_i + \boldsymbol{\epsilon}_i, \tag{4}$$

where

$$X_i = \begin{bmatrix} X_i^{\text{Covariate}} & X_i^{\text{Baseline}} & X_i^{\text{SNP}} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \boldsymbol{w}_{i1}^\top \\ \vdots \\ \boldsymbol{w}_{ic_i}^\top \end{bmatrix} & \Phi_i & g_i^{\text{SNP}}\Phi_i \end{bmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^{\text{Covariate}} \\ \boldsymbol{\beta}^{\text{Baseline}} \\ \boldsymbol{\beta}^{\text{SNP}} \end{pmatrix}. \tag{5}$$

Here, $\boldsymbol{\beta}^{\text{Covariate}} \in \mathbb{R}^J$, $\boldsymbol{\beta}^{\text{Baseline}} \in \mathbb{R}^K$, and $\boldsymbol{\beta}^{\text{SNP}} \in \mathbb{R}^K$ are vectors of fixed coefficients. The matrix $\Phi_i \in \mathbb{R}^{c_i \times K}$ collects the spline basis evaluations, with rows $\boldsymbol{\phi}^\top(t_{ic})$ for $c = 1, \ldots, c_i$. $\boldsymbol{u}_i \in \mathbb{R}^K$ and $\boldsymbol{\epsilon}_i \in \mathbb{R}^{c_i}$ are random vectors.

Stacking all individuals yields the full model:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \Phi\boldsymbol{u} + \boldsymbol{\epsilon}, \tag{6}$$

and this model takes the standard form of a linear mixed model. The design matrices are given by

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_I \end{bmatrix} \in \mathbb{R}^{C \times (J+2K)}, \ \Phi = \begin{bmatrix} \Phi_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Phi_I \end{bmatrix} \in \mathbb{R}^{C \times IK}. \tag{7}$$

Throughout the paper, $\mathbf{0}$ denotes a vector or matrix of zeros of compatible dimension. We assume independence across individuals, and that the random effects $\boldsymbol{u} \in \mathbb{R}^{IK}$ and the noise $\boldsymbol{\epsilon} \in \mathbb{R}^{C}$ are independent of each other.

4.2 Maximum likelihood estimation using the IDUL algorithm

We estimate the parameters of the functional mixed model (6) using maximum likelihood estimation (MLE). Let $\theta = \{\boldsymbol{\beta}, \tau, \eta\}$ denote the model parameters. After integrating out the random effects $\boldsymbol{u}$, the gene expression vector $\boldsymbol{y}$ follows a multivariate normal distribution:

$$\boldsymbol{y} \sim \mathrm{MVN}\left(X\boldsymbol{\beta}, \tau^{-1}(\eta\Phi\Phi^{\top} + I_C)\right), \tag{8}$$

where I_C is the $C \times C$ identity matrix.

To efficiently fit this linear mixed model, we adopt the IDUL framework [42]. Following the strategy of FaST-LMM [43], IDUL first diagonalizes the covariance matrix via eigen decomposition

$$\Phi\Phi^{\top} = QDQ^{\top}. \tag{9}$$

For notational convenience, we use the subscript Q to denote quantities left-multiplied by $Q^{\top}$, that is, for any matrix or vector A, we define $A_Q = Q^{\top}A$. Rotating the model

with $Q^\top$ yields

$$\boldsymbol{y}_Q \sim \mathrm{MVN}\left(X_Q\boldsymbol{\beta}, \tau^{-1}(\eta D + I_C)\right). \tag{10}$$

The covariance matrix $\tau^{-1}(\eta D + I_C)$ is now diagonal. Although the likelihood of $\boldsymbol{y}_Q$ is algebraically identical to that of $\boldsymbol{y}$, the diagonal covariance structure enables substantially more efficient computation.

The MLE is then obtained by maximizing the marginal log-likelihood with respect to the variance ratio η. The IDUL algorithm and its gradient-ascent variant (IDUL†) iteratively update η through dispersion updates. Both methods are more computationally efficient than the commonly used Newton-Raphson method [42]. Further implementation details of IDUL are provided in the supplementary materials.

4.3 Hypothesis testing

4.3.1 Fitting the null gene expression model

Scarf-QTL implements several hypothesis tests for detecting eQTL effects that may either vary across the cell-state coordinate or remain constant across cell states. All tests share a common null hypothesis, namely, that the candidate SNP has no effect on gene expression. The gene expression model (6) is fitted under the null hypothesis ($\boldsymbol{\beta}^{\mathrm{SNP}} = \mathbf{0}$) once for each gene, and the fitted null model is reused for all candidate cis-SNPs in the subsequent score tests.

For convenience, we partition the design matrix and the corresponding coefficient

vector as follows:

$$X = \begin{bmatrix} X^{\text{Covariate}} & X^{\text{Baseline}} & X^{\text{SNP}} \end{bmatrix} = \begin{bmatrix} X^{\text{COV}} & X^{\text{SNP}} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^{\text{COV}} \\ \boldsymbol{\beta}^{\text{SNP}} \end{pmatrix}. \tag{11}$$

We fit the model (6) with the constraint $\boldsymbol{\beta}^{\text{SNP}} = \mathbf{0}$ using the IDUL algorithm. Let $\hat{\theta}_0 = (\hat{\boldsymbol{\beta}}_0, \hat{\tau}_0, \hat{\eta}_0)$ denote the MLE under the null hypothesis, where $\hat{\boldsymbol{\beta}}_0 = \left(\left(\hat{\boldsymbol{\beta}}_0^{\text{COV}} \right)^\top, \mathbf{0}^\top \right)^\top$.

For individual i, the fitted mean and variance of gene expression under H_0 are

$$\hat{\boldsymbol{y}}_i = X_i^{\text{COV}} \hat{\boldsymbol{\beta}}_0^{\text{COV}}, \qquad \widehat{\text{Var}}(\boldsymbol{y}_i) = \hat{\tau}_0^{-1} (\hat{\eta}_0 \Phi_i \Phi_i^\top + I_{c_i}), \tag{12}$$

and we define the corresponding weighted residual vector as

$$\boldsymbol{e}_i = \left(\widehat{\text{Var}}(\boldsymbol{y}_i) \right)^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i). \tag{13}$$

These residuals are used to construct both the prospective score tests and retrospective association tests below.

4.3.2 Prospective score tests for cell-state-dependent or constant eQTL effects

Under the functional mixed model, it is straightforward to test for an eQTL whose effect may vary across the cell-state coordinate. Since the eQTL effect size is modeled as a smooth function of the cell-state coordinate,

$$\gamma^{\text{SNP}}(t) = \sum_{k=1}^{K} \phi_k(t) \beta_k^{\text{SNP}}, \tag{14}$$

we test the hypotheses $H_0 : \boldsymbol{\beta}^{\text{SNP}} = \mathbf{0}$ v.s. $H_1 : \boldsymbol{\beta}^{\text{SNP}} \neq \mathbf{0}$.

In addition, we consider another alternative hypothesis $\boldsymbol{\beta}^{\text{SNP}} = \beta^{\text{SNP}}_{\text{static}} \mathbf{1}_K$ corresponding to a constant eQTL effect, where $\beta^{\text{SNP}}_{\text{static}}$ is a scalar and $\mathbf{1}_K$ is a K-dimensional vector of ones. Under this alternative, the eQTL effect size becomes

$$\gamma^{\text{SNP}}(t) = \sum_{k=1}^{K} \phi_k(t) \beta^{\text{SNP}}_{\text{static}} = \beta^{\text{SNP}}_{\text{static}}, \ \forall t, \tag{15}$$

due to the partition-of-unity property of B-spline basis.

Here, we introduce the prospective score tests, which model gene expression as a random outcome conditional on observed genotype and covariates. Note that score tests avoid fitting the alternative model. For the dynamic test, following standard score test procedures (see the supplementary materials), we have the score function and the Fisher information matrix of $\boldsymbol{\beta}$ evaluated at $\hat{\theta}_0$

$$U_{\boldsymbol{\beta}}(\hat{\theta}_0) = \begin{pmatrix} \mathbf{0} \\ U_{\text{dynamic}} \end{pmatrix}, \ I_{\boldsymbol{\beta}}(\hat{\theta}_0) = \hat{\tau}_0 X^\top \left(\hat{\eta}_0 \Phi \Phi^\top + I_C \right)^{-1} X, \tag{16}$$

where $U_{\text{dynamic}} \in \mathbb{R}^K$ is referred to as the dynamic score function

$$U_{\text{dynamic}} = \hat{\tau}_0 \left(X^{\text{SNP}} \right)^\top \left(\hat{\eta}_0 \Phi \Phi^\top + I_C \right)^{-1} \left(\boldsymbol{y} - X \hat{\boldsymbol{\beta}}_0 \right) = \sum_i g_i^{\text{SNP}} \Phi_i^\top \boldsymbol{e}_i. \tag{17}$$

The score statistic is then

$$T^{\text{prosp}}_{\text{dynamic}} = U^\top_{\text{dynamic}} \left[\widehat{\text{Var}}_0 \left(U_{\text{dynamic}} \right) \right]^{-1} U_{\text{dynamic}}, \tag{18}$$

where $\left[\widehat{\mathrm{Var}}_0\left(U_{\mathrm{dynamic}}\right)\right]^{-1}$ is the lower-right submatrix of $\left[I_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right)\right]^{-1}$ corresponding to $\boldsymbol{\beta}^{\mathrm{SNP}}$. The score statistic asymptotically follows a χ^2_K distribution under H_0, and the resulting p-value is denoted by $p^{\mathrm{prosp}}_{\mathrm{dynamic}}$.

For the static test, a similar procedure leads to the static score function

$$U_{\mathrm{static}} = \sum_i g_i^{\mathrm{SNP}} \mathbf{1}_{c_i}^\top \boldsymbol{e}_i = \mathbf{1}_K^\top U_{\mathrm{dynamic}} \in \mathbb{R}, \tag{19}$$

and the corresponding prospective variance is

$$\widehat{\mathrm{Var}}_0\left(U_{\mathrm{static}}\right) = \mathbf{1}_K^\top \widehat{\mathrm{Var}}_0\left(U_{\mathrm{dynamic}}\right) \mathbf{1}_K. \tag{20}$$

The resulting score test statistics $T^{\mathrm{prosp}}_{\mathrm{static}} = U^2_{\mathrm{static}} / \widehat{\mathrm{Var}}_0\left(U_{\mathrm{static}}\right)$ asymptotically follows a χ^2_1 distribution under H_0, and the p-value is denoted by $p^{\mathrm{prosp}}_{\mathrm{static}}$.

4.3.3 Retrospective association tests

Instead of prospective score tests, scarf-QTL adopts retrospective association tests to improve robustness against model misspecification, following the framework of RVMMAT [35] originally developed for longitudinal GWAS. The retrospective approach considers the conditional distribution of genotype given the observed gene expression $\boldsymbol{y}$ and covariates X^{COV}.

Let $G^{\mathrm{SNP}} = [g_1^{\mathrm{SNP}}, \ldots, g_I^{\mathrm{SNP}}]^\top \in \mathbb{R}^I$ denote the vector of genotypes across individuals. Under the null hypothesis of no eQTL effect, following previous work [35], we assume that the conditional expectation and variance of genotype do not depend

on the observed gene expression and are given by

$$\mathrm{E}_0\left(G^{\mathrm{SNP}} \mid \boldsymbol{y}, X^{\mathrm{COV}}\right) = 2p\mathbf{1}_I,\ \mathrm{Var}_0\left(G^{\mathrm{SNP}} \mid \boldsymbol{y}, X^{\mathrm{COV}}\right) = \sigma_g^2\Sigma, \tag{21}$$

where p denotes the minor allele frequency (MAF) of the SNP, σ_g^2 is an unknown variance parameter, and Σ is the genetic relationship matrix (GRM).

Our retrospective association tests do not explicitly model the conditional distribution of G^{SNP} under the alternative hypothesis. Instead, we use the dynamic score function U_{dynamic} defined in equation (17) to test for an eQTL while allowing its effect to vary along the cell-state coordinate. Let $E^{\mathrm{BD}} \in \mathbb{R}^{C\times I}$ be the block-diagonal matrix whose i-th diagonal block is the weighted residual vector $\boldsymbol{e}_i$. The score function can then be rewritten as $U_{\mathrm{dynamic}} = \left(X^{\mathrm{Baseline}}\right)^\top E^{\mathrm{BD}} G^{\mathrm{SNP}}$. The conditional expectation of the score function is

$$\begin{aligned}\mathrm{E}_0\left(U_{\mathrm{dynamic}} \mid \boldsymbol{y}, X^{\mathrm{COV}}\right) &= \left(X^{\mathrm{Baseline}}\right)^\top E^{\mathrm{BD}}\, \mathrm{E}_0\left(G^{\mathrm{SNP}} \mid \boldsymbol{y}, X^{\mathrm{COV}}\right) &(22)\\ &= 2p\hat{\tau}_0 \left(X^{\mathrm{Baseline}}\right)^\top \left(\hat{\eta}_0 \Phi\Phi^\top + I_C\right)^{-1}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_0\right) = \mathbf{0}, &(23)\end{aligned}$$

where the last equality holds according to the likelihood equation under the null hypothesis. Furthermore, the conditional variance of the score function is

$$\begin{aligned}\mathrm{Var}_0\left(U_{\mathrm{dynamic}} \mid \boldsymbol{y}, X^{\mathrm{COV}}\right) &= \left(X^{\mathrm{Baseline}}\right)^\top E^{\mathrm{BD}}\, \mathrm{Var}_0\left(G^{\mathrm{SNP}} \mid \boldsymbol{y}, X^{\mathrm{COV}}\right)\left(E^{\mathrm{BD}}\right)^\top X^{\mathrm{Baseline}} &(24)\\ &= \sigma_g^2 \left(X^{\mathrm{Baseline}}\right)^\top E^{\mathrm{BD}} \Sigma \left(E^{\mathrm{BD}}\right)^\top X^{\mathrm{Baseline}}. &(25)\end{aligned}$$

We estimate p by the sample MAF ($\hat{p}$) and σ_g^2 by $\hat{\sigma}_g^2 = 2\hat{p}(1-\hat{p})$, assuming Hardy–Weinberg equilibrium. The GRM Σ is estimated using genome-wide genotype data through the make-rel command in the "PLINK" software [44], and the resulting estimate is denoted by $\hat{\Sigma}$. Plugging $\hat{\sigma}_g^2$ and $\hat{\Sigma}$ into equation (25) gives the estimated retrospective variance $\hat{\mathrm{Var}}_0(U_{\text{dynamic}} \mid \boldsymbol{y}, X^{\text{COV}})$. The retrospective score test statistic is then constructed as

$$T_{\text{dynamic}} = U_{\text{dynamic}}^{\top} \left[\widehat{\mathrm{Var}}_0 \left(U_{\text{dynamic}} \mid \boldsymbol{y}, X^{\text{COV}}\right)\right]^{-1} U_{\text{dynamic}}. \tag{26}$$

Under the null hypothesis, conditional on the observed gene expression and covariates, $T_{\text{dynamic}} \sim \chi_K^2$ asymptotically. The corresponding p-value is denoted by p_{dynamic}. We refer to this test as the dynamic test hereafter.

Similarly, the static score function U_{static} is used for the static test. The conditional expectation of U_{static} is zero, and its conditional variance is

$$\widehat{\mathrm{Var}}_0 \left(U_{\text{static}} \mid \boldsymbol{y}, X^{\text{COV}}\right) = \mathbf{1}_K^{\top} \widehat{\mathrm{Var}}_0 \left(U_{\text{dynamic}} \mid \boldsymbol{y}, X^{\text{COV}}\right) \mathbf{1}_K. \tag{27}$$

The resulting score statistic T_{static} asymptotically follows a χ_1^2 distribution under the null hypothesis, yielding the p-value p_{static} for the static test.

Finally, scarf-QTL also implements a combined test to leverage information from both tests. The Cauchy combination method [36] is applied as follows:

$$T_{\text{combined}} = \frac{1}{2} \left\{\tan\left[\left(0.5 - p_{\text{static}}\right)\pi\right] + \tan\left[\left(0.5 - p_{\text{dynamic}}\right)\pi\right]\right\}. \tag{28}$$

Under the null hypothesis, T_{combined} follows an approximately standard Cauchy distribution in the right tail, and the corresponding combined p-value is denoted p_{combined}.

It is worth noting that the retrospective association tests are robust to misspecification of the gene expression model used to construct the score function. Specifically, the score function can be written in the form $U = f(\boldsymbol{y}, X^{\text{COV}})G^{\text{SNP}}$, where $f(\boldsymbol{y}, X^{\text{COV}})$ is determined by the gene expression model. Therefore, conditioning on the observed gene expression and covariates, $f(\boldsymbol{y}, X^{\text{COV}})$ becomes a fixed matrix, and the validity of the test depends only on the assumptions about the genotype distribution under the null hypothesis. However, an appropriate gene expression model may improve statistical power.

4.4 Algorithmic accelerations

Efficient computation of scarf-QTL relies on the eigen decomposition of $\Phi\Phi^\top$, which not only diagonalizes the variance structure of gene expression but also reduces the dimensionality of score statistic computation, owing to the block-diagonal and low-rank structure of $\Phi\Phi^\top$.

We first perform the singular value decomposition on Φ_i for each individual i:

$$\Phi_i = \begin{bmatrix} U_i & V_i \end{bmatrix} \begin{bmatrix} \Lambda_i^{1/2} \\ \mathbf{0}_{n_i^V \times n_i^U} \end{bmatrix} R_i^\top = U_i \Lambda_i^{1/2} R_i^\top, \tag{29}$$

where $\Lambda_i^{1/2} \in \mathbb{R}^{n_i^U \times n_i^U}$ is a diagonal matrix of singular values, with $n_i^U = \min\{c_i, K\}$. $U_i \in \mathbb{R}^{c_i \times n_i^U}$ contains the left-singular vectors corresponding to $\Lambda_i^{1/2}$, and $V_i \in \mathbb{R}^{c_i \times n_i^V}$ contains the remaining left-singular vectors, which correspond to zero singular values,

with $n_i^V = c_i - n_i^U$. $R_i \in \mathbb{R}^{K \times n_i^U}$ is the matrix of right-singular vectors. It follows that

$$\Phi_i \Phi_i^\top = U_i \Lambda_i U_i^\top = \begin{bmatrix} U_i & V_i \end{bmatrix} \begin{bmatrix} \Lambda_i & \mathbf{0}_{n_i^U \times n_i^V} \\ \mathbf{0}_{n_i^V \times n_i^U} & \mathbf{0}_{n_i^V \times n_i^V} \end{bmatrix} \begin{bmatrix} U_i^\top \\ V_i^\top \end{bmatrix}. \tag{30}$$

Let $U \in \mathbb{R}^{C \times n^U}$ be the block-diagonal matrix composed of $\{U_i, i = 1, \ldots, I\}$, $V \in \mathbb{R}^{C \times n^V}$ the block-diagonal matrix of $\{V_i, i = 1, \ldots, I\}$, and $\Lambda \in \mathbb{R}^{n^U \times n^U}$ be the block diagonal matrix of $\{\Lambda_i, i = 1, \ldots, I\}$, with $n^U = \sum_{i=1}^I n_i^U \leq IK$, and $n^V = \sum_{i=1}^I n_i^V = C - n^U$. Then

$$Q = \begin{bmatrix} U & V \end{bmatrix} \in \mathbb{R}^{C \times C}, \; D = \begin{bmatrix} \Lambda & \mathbf{0}_{n^U \times n^V} \\ \mathbf{0}_{n^V \times n^U} & \mathbf{0}_{n^V \times n^V} \end{bmatrix}, \tag{31}$$

gives the eigen decomposition $\Phi\Phi^\top = QDQ^\top$.

In the IDUL algorithm, the rotation of $\boldsymbol{y}$ and X in equation (10) is implemented as

$$\boldsymbol{y}_Q = \begin{pmatrix} \boldsymbol{y}_U \\ \boldsymbol{y}_V \end{pmatrix}, \; \boldsymbol{y}_U = \begin{pmatrix} U_1^\top \boldsymbol{y}_1 \\ \vdots \\ U_I^\top \boldsymbol{y}_I \end{pmatrix}, \; \boldsymbol{y}_V = \begin{pmatrix} V_1^\top \boldsymbol{y}_1 \\ \vdots \\ V_I^\top \boldsymbol{y}_I \end{pmatrix}, \tag{32}$$

$$X_Q = \begin{pmatrix} X_U \\ X_V \end{pmatrix}, \; X_U = \begin{pmatrix} U_1^\top X_1 \\ \vdots \\ U_I^\top X_I \end{pmatrix}, \; X_V = \begin{pmatrix} V_1^\top X_1 \\ \vdots \\ V_I^\top X_I \end{pmatrix}. \tag{33}$$

Notably, the rotation of $\boldsymbol{y}$ for multiple genes can be computed simultaneously by rotating the gene-expression matrix (cells by genes) for each individual. More importantly, the columns of X_V corresponding to individual-level predictors are exactly

zero. For instance,

$$V_i^\top X_i^{\text{Baseline}} = V_i^\top \Phi_i = V_i^\top U_i \Lambda_i^{1/2} R_i^\top = \mathbf{0}, \tag{34}$$

since U_i and V_i are orthogonal. Similarly, if $w_{ic,j} = w_{i,j}$ for all $c \in \{1, \ldots, c_i\}$ (an individual-level covariate), the j-th column of $V_i^\top X_i$ is equal to $V_i^\top w_{i,j} \mathbf{1}_{c_i} = w_{i,j} V_i^\top \Phi_i \mathbf{1}_K = \mathbf{0}$, due to the partition-of-unity property of B-splines ($\sum_k \phi_k(t) = 1$ for all t).

Eigen transformation also facilitates the computation of the score function and retrospective variance. For the score function, using equations (12) and (13), we have

$$U_{\text{dynamic}} = \sum_i g_i^{\text{SNP}} \left\{ R_i \Lambda_i^{1/2} U_i^\top \right\} \left\{ \hat{\tau}_0 \begin{bmatrix} U_i & V_i \end{bmatrix} \begin{bmatrix} \left(\hat{\eta}_0 \Lambda_i + I_{n_i^U}\right)^{-1} & \\ & I_{n_i^V} \end{bmatrix} \begin{bmatrix} U_i^\top \\ V_i^\top \end{bmatrix} \right\} (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i) \tag{35}$$

$$= \sum_i g_i^{\text{SNP}} R_i \Lambda_i^{1/2} \hat{\tau}_0 (\hat{\eta}_0 \Lambda_i + I_{n_i^U})^{-1} U_i^\top (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i) = \sum_i g_i^{\text{SNP}} R_i \Lambda_i^{1/2} \boldsymbol{e}_i^U, \tag{36}$$

with

$$\boldsymbol{e}_i^U = \hat{\tau}_0 \left(\hat{\eta}_0 \Lambda_i + I_{n_i^U}\right)^{-1} \left(U_i^\top \boldsymbol{y}_i - U_i^\top \hat{\boldsymbol{y}}_i\right) \in \mathbb{R}^{n_i^U}. \tag{37}$$

Here, $\left(\hat{\eta}_0 \Lambda_i + I_{n_i^U}\right)^{-1}$ and $U_i^\top \boldsymbol{y}_i - U_i^\top \hat{\boldsymbol{y}}_i \in \mathbb{R}^{n_i^U}$ are slices of the weights and residuals in model (10), respectively, and $\boldsymbol{e}_i^U$ is referred to as the rotated weighted residual vector. We stack $R_i \Lambda_i^{1/2} \boldsymbol{e}_i^U \in \mathbb{R}^K$ across individuals into the matrix $M_U \in \mathbb{R}^{I \times K}$, yielding $U_{\text{dynamic}} = M_U^\top G^{\text{SNP}}$. This formulation allows parallel computation of the

score function for thousands of SNPs per gene via matrix multiplication between M_U and the genotype matrix.

For the retrospective variance, since the i-th column of $\left(X^{\text{Baseline}}\right)^\top E^{\text{BD}}$ equals $\Phi_i^\top \boldsymbol{e}_i = R_i \Lambda_i^{1/2} \boldsymbol{e}_i^U$, we have $\left(X^{\text{Baseline}}\right)^\top E^{\text{BD}} = M_U^\top$, and thus

$$\widehat{\text{Var}}_0\left(U_{\text{dynamic}} \mid \boldsymbol{y}, X^{\text{COV}}\right) = \hat{\sigma}_g^2 M_U^\top \hat{\Sigma} M_U. \tag{38}$$

For thousands of cis-SNPs of a given gene, inversion of the retrospective variance requires only a single matrix inversion of $M_U^\top \hat{\Sigma} M_U$. Computation of the prospective variance, although somewhat slower, can also be simplified (see the supplementary materials).

In summary, the scarf-QTL algorithm (see the supplementary materials) consists of three main steps: eigen transformation, null model fitting, and retrospective association test. First, scarf-QTL performs eigen decomposition of the gene-expression covariance matrix under the functional mixed model and uses the resulting eigenvectors to rotate both the gene-expression matrix and the corresponding design matrices. Next, scarf-QTL fits the model under the null hypothesis of no eQTL effect using the IDUL algorithm [42] for each gene. Finally, for each gene, scarf-QTL computes score functions and their variances for all cis-SNPs and derives the corresponding p-values.

4.5 Penalized estimation of cell-state-dependent eQTL effects

While it is feasible to estimate the eQTL effect under the full functional mixed model using MLE, the computational cost can be substantial, as the model would need to be refitted for each SNP. To reduce computational burden, we fix the variance components and covariate effects at their MLEs obtained under the null model, namely $\hat{\tau}_0$, $\hat{\eta}_0$, and $\hat{\boldsymbol{\beta}}_0^{\mathrm{COV}}$. The likelihood is then maximized with respect to $\boldsymbol{\beta}^{\mathrm{SNP}}$ only, yielding a pseudo-MLE of $\boldsymbol{\beta}^{\mathrm{SNP}}$, in the form of a weighted least squares estimator:

$$\tilde{\boldsymbol{\beta}}^{\mathrm{SNP}} = \left[\left(X^{\mathrm{SNP}}\right)^\top W_0^{-1} X^{\mathrm{SNP}}\right]^{-1} \left(X^{\mathrm{SNP}}\right)^\top W_0^{-1} \left(\boldsymbol{y} - X^{\mathrm{COV}} \hat{\boldsymbol{\beta}}_0^{\mathrm{COV}}\right) \tag{39}$$

$$= \arg\min_{\boldsymbol{b}} \left(\boldsymbol{y} - X^{\mathrm{COV}} \hat{\boldsymbol{\beta}}_0^{\mathrm{COV}} - X^{\mathrm{SNP}} \boldsymbol{b}\right)^\top W_0^{-1} \left(\boldsymbol{y} - X^{\mathrm{COV}} \hat{\boldsymbol{\beta}}_0^{\mathrm{COV}} - X^{\mathrm{SNP}} \boldsymbol{b}\right), \tag{40}$$

where $W_0 = \hat{\eta}_0 \Phi \Phi^\top + I_C$.

Furthermore, in practice, collinearity in the design matrix and potential model misspecification in real data can cause the effect-size curves obtained from the standard MLE or the pseudo-MLE in equation (39) to fluctuate sharply along the cell-state coordinate, particularly when the number of cells is limited. To stabilize estimation, we adopt a smoothness-penalized approach [37]. Specifically, we introduce a second-order finite-difference operator $D_{(2)} \in \mathbb{R}^{(K-2)\times K}$, leading to the penalty term

$$\boldsymbol{b}^\top D_{(2)}^\top D_{(2)} \boldsymbol{b} = \sum_{k=1}^{K-2} [(b_{k+2} - b_{k+1}) - (b_{k+1} - b_k)]^2. \tag{41}$$

The penalized estimator with penalty parameter λ is then given by

$$\tilde{\boldsymbol{\beta}}_{\lambda}^{\mathrm{SNP}} = \arg\min_{\boldsymbol{b}} \left(\boldsymbol{y} - X^{\mathrm{COV}}\hat{\boldsymbol{\beta}}_{0}^{\mathrm{COV}} - X^{\mathrm{SNP}}\boldsymbol{b}\right)^{\top} W_0^{-1} \left(\boldsymbol{y} - X^{\mathrm{COV}}\hat{\boldsymbol{\beta}}_{0}^{\mathrm{COV}} - X^{\mathrm{SNP}}\boldsymbol{b}\right) + \lambda \boldsymbol{b}^{\top} D_{(2)}^{\top} D_{(2)} \boldsymbol{b} \tag{42}$$

$$= \left[\left(X^{\mathrm{SNP}}\right)^{\top} W_0^{-1} X^{\mathrm{SNP}} + \lambda D_{(2)}^{\top} D_{(2)}\right]^{-1} \left(X^{\mathrm{SNP}}\right)^{\top} W_0^{-1} \left(\boldsymbol{y} - X^{\mathrm{COV}}\hat{\boldsymbol{\beta}}_{0}^{\mathrm{COV}}\right). \tag{43}$$

We evaluated multiple values of the penalty parameter λ in simulations to characterize its impact on effect-size estimation (Fig. S3). Under the normal model, λ exhibited a clear bias–variance tradeoff: smaller λ yielded more fluctuating estimates with higher variance, whereas larger λ led to smoother estimates with increased bias. Under count-based models (Poisson and negative binomial), adopting $\lambda = 1$ recovered the overall effect patterns well when the genetic signal was sufficiently strong. Since effect-size estimation is used only for downstream clustering and visualization rather than for inference, we use $\lambda = 1$ throughout our analyses, namely $\hat{\boldsymbol{\beta}}^{\mathrm{SNP}} = \tilde{\boldsymbol{\beta}}_{1}^{\mathrm{SNP}}$. The estimated eQTL effect as a function of the cell-state coordinate is then given by

$$\hat{\gamma}^{\mathrm{SNP}}(t) = \sum_{k=1}^{K} \phi_k(t)\, \hat{\beta}_k^{\mathrm{SNP}}. \tag{44}$$

4.6 Simulation studies

4.6.1 Simulation settings

Simulation analyses were based on the OneK1K dataset [12] and conducted for three representative cell types ($\mathrm{CD4_{NC}}$, $\mathrm{B_{IN}}$, and Plasma cells), representing cell types

with large, intermediate, and small numbers of cells, respectively. We first simulated gene expression without cell-level random noise, with the eQTL effect specified as a smooth function of the cell-state coordinate, according to the model (3):

$$\nu_{ic} = \sum_{j=1}^{J} w_{ic,j}\beta_j^{\text{Covariate}} + \sum_{k=1}^{K} \phi_k(t_{ic}) \left(\beta_k^{\text{Baseline}} + u_{ik}\right) + g_i^{\text{SNP}}\gamma^{\text{SNP}}(t_{ic}). \tag{45}$$

Here, covariates ($w_{ic,j}$) and the cell-state coordinates (t_{ic}, obtained using pseudotime inference) were taken from the OneK1K dataset described in the "Application to OneK1K dataset" section. For each individual, the genotype of the causal SNP g_i^{SNP} was independently sampled from a binomial distribution $\mathcal{B}(2, p)$ with minor allele frequency $p = 0.2$. Individual-level random effects were generated as $u_{ik} \sim \mathcal{N}(0, 1)$. Covariate coefficients were sampled from $\beta_j^{\text{Covariate}} \sim \mathcal{N}(0, 1)$, and baseline expression coefficients were fixed as $\beta_k^{\text{Baseline}} = k/K$ for $k = 1, \ldots, K$.

For type I error evaluation, we set $\gamma^{\text{SNP}}(t) = 0$. For power analysis, we simulated a range of eQTL effect patterns by specifying $\gamma^{\text{SNP}}(t)$ as follows:

- Static: $\gamma^{\text{SNP}}(t) = 1$.
- Linear: $\gamma^{\text{SNP}}(t) = t$.
- Two-stage (on/off): $\gamma^{\text{SNP}}(t) = \{\exp[-20(t - 0.4)] + 1\}^{-1}$.
- Two-stage (switch): $\gamma^{\text{SNP}}(t) = \{\exp[-20(t - 0.4)] + 1\}^{-1} - 0.5$.
- Peak: $\gamma^{\text{SNP}}(t) = \exp\{-20(t - 0.9)^2\}$.
- Cyclic: $\gamma^{\text{SNP}}(t) = \sin(2\pi t) + 1$.

- Cyclic (switch): $\gamma^{\text{SNP}}(t) = \sin(2\pi t)$.
- Cyclic (two periods): $\gamma^{\text{SNP}}(t) = \sin(4\pi t) + 1$.

We further scaled the eQTL effect size, quantified by $\int_0^1 |\gamma^{\text{SNP}}(t)|dt$, such that it equaled $0.02 \times m$ for $m = 1, \ldots, 5$, denoted by "$m\times$" in Fig. 2.

Observed gene expression y_{ic} was then generated from ν_{ic} using the following distributions:

- Normal: $y_{ic} \sim \mathcal{N}(\nu_{ic}, 1)$.
- Poisson: $y_{ic} \sim \mathcal{P}(\exp\{\nu_{ic}\})$.
- Negative binomial: $y_{ic} \sim \mathcal{NB}(\exp\{\nu_{ic}\}, 1)$, where the second parameter denotes the dispersion parameter.

4.6.2 Type I error evaluation

To assess type I error, we generated 1,000 replicates of gene expression and tested for eQTL associations between gene expression and genotypes of 1,000 randomly sampled common SNPs (MAF > 0.05) from real data. Both prospective score tests and retrospective association tests were evaluated under different generative models. When gene expression was generated from Poisson or negative binomial distributions, count data were normalized using the rank-based inverse normal transformation (the RankNorm function from the "RNOmni" R package). As a benchmark, we additionally evaluated a pseudobulk-based method, which tested for Spearman correlation between genotype and the residuals from regressing pseudobulk gene expression on

individual-level covariates. Pseudobulk expression was calculated as the mean of single-cell normalized expression when gene expression was generated from the normal distribution, or as the mean of $\log(x+1)$-transformed cell-level counts when gene expression was generated from Poisson or negative binomial distributions. Empirical type I error rates were calculated as the proportion of gene–SNP pairs with p-values below nominal significance levels $\alpha = 0.05, 0.01, 0.001$, and 0.0001.

Using the central limit theorem, we constructed a one-sided hypothesis test for whether the type I error rate p_0 exceeds the nominal level α, with null hypothesis $H_0 : p_0 \leq \alpha$ versus alternative $H_1 : p_0 > \alpha$. At a 1% significance level, the corresponding rejection region is

$$\left[\alpha + Z_{0.99}\sqrt{\frac{\alpha(1-\alpha)}{N}},\ \infty\right), \tag{46}$$

where $Z_{0.99}$ denotes the 99th percentile of the standard normal distribution and $N = 10^6$ is the total number of tested gene–SNP pairs.

4.6.3 Power evaluation and eQTL pattern recovery

To evaluate the power of scarf-QTL, we simulated gene expression data for 1,000 genes under each combination of eQTL effect pattern and effect size. We then tested each gene for association with the corresponding causal SNP using scarf-QTL and the pseudobulk-based benchmark described above. Empirical power was defined as the proportion of tests with p-values below $\alpha = 0.05$.

In addition, eQTL effects were estimated using multiple values of the smoothing parameter λ. To assess pattern recovery, estimated effects were plotted against the true underlying patterns, with both rescaled so that $\int_0^1 |\hat{\gamma}^{\mathrm{SNP}}(t)|\, dt = 1$ and

$\int_0^1 |\gamma^{\mathrm{SNP}}(t)|\, dt = 1$, respectively.

4.6.4 Runtime evaluation

To evaluate the computational efficiency of scarf-QTL on large-scale datasets, we simulated datasets for 100 genes with varying numbers of SNPs, individuals, and cells by resampling the OneK1K dataset. The number of cis-SNPs per gene, individuals, and cells per individual are denoted by n_{snp}, n_{indi}, and n_{cell}, respectively.

The default values are $n_{\mathrm{snp}} = 3{,}000$, $n_{\mathrm{indi}} = 1{,}000$, and $n_{\mathrm{cell}} = 200$. We varied one or two parameters to assess their effects on runtime.

- The number of cis-SNPs per gene, $n_{\mathrm{snp}} \in \{10^3, 3 \times 10^3, 10^4, 3 \times 10^4\}$.
- The numbers of individuals and cells per individual jointly,

$$(n_{\mathrm{indi}}, n_{\mathrm{cell}}) \in \{100, 200, 500, 10^3, 2 \times 10^3, 5 \times 10^3, 10^4\}^2. \tag{47}$$

For each resampled dataset, we applied scarf-QTL and recorded the total runtime. We also applied the pseudobulk-based method and recorded the runtime of the Spearman correlation test only. All runtime experiments were conducted on a single CPU core of an Intel Xeon Platinum 8375C processor (2.90 GHz) with multi-threaded BLAS disabled.

4.7 Application to OneK1K dataset

4.7.1 Data preprocessing

We applied scarf-QTL to the OneK1K dataset [12], in which cells were annotated into 14 immune cell types. We used the normalized gene expression matrix, imputed genotypes, and covariate files provided directly by the study authors. Following the authors' recommendations, we excluded two additional individuals and corrected a previously mislabelled donor annotation. Individual-level covariates, including sex, age, six genotype principal components, and two PEER factors per cell type, as well as cell-level covariates, including total UMI count and mitochondrial UMI percentage, were incorporated into downstream analyses. Gene expression values were further normalized using the rank-based inverse normal transformation.

For each cell type, we computed Pearson residuals of 500 highly variable genes using the SCTransform function from the "Seurat" R package [45], regressing out pool (batch effect) and total UMI count. Dimensionality reduction was performed using the phate function from the "phateR" R package [46] on the top 10 principal components, followed by trajectory inference using the slingshot function from the "slingshot" R package [47]. Cells were ordered according to their position along the inferred trajectory, and pseudotime values were rescaled to the interval $(0, 1)$ based on their ranks. We note that the resulting pseudotime is used as a one-dimensional cell-state coordinate. This coordinate provides a continuous representation of cellular variation and may capture either continuous biological variation or discrete cell subsets. When an underlying biological trajectory exists, it may correspond to a monotonic but potentially nonlinear transformation of such a trajectory.

4.7.2 eQTL mapping and false discovery rate control

Scarf-QTL was applied to all gene–cis-SNP pairs tested in the original OneK1K analysis. While we performed the static, dynamic, and combined tests, our downstream analyses mainly focused on the combined test.

False discovery rate was controlled separately for each cell type and chromosome, and for each test. We applied the qvalue procedure implemented in the R package "qvalue" [39] with tuning parameter $\lambda = 0.5$. Gene–SNP pairs with local FDR below 5% were considered significant eQTLs. Genes with at least one significant eQTL were referred to as eGenes.

4.7.3 Enrichment of eGenes with dendritic cell subset marker genes

Villani *et al.* [40] defined transcriptionally distinct dendritic cell (DC) subsets based on single-cell RNA sequencing and reported discriminative genes for each subset (referred to as subset markers). Enrichment analysis was performed to evaluate the overlap between these subset markers and eGenes identified by the static and combined tests. A one-sided Fisher's exact test was used, with the background gene set consisting of all genes tested for eQTL mapping in dendritic cells.

4.7.4 Cell-state-dependent eQTL effect estimation and clustering

For significant eQTLs identified by the combined test of scarf-QTL, we estimated smooth functions of eQTL effects along the cell-state coordinate. Estimated effect

curves were rescaled such that

$$\int_0^1 |\hat{\gamma}^{\mathrm{SNP}}(t)|\, dt = 1$$

to focus on effect patterns rather than effect magnitude. Rescaled eQTL effect curves were then clustered using k-means. The number of clusters was selected automatically according to the strategy of Hou et al. [34].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The OneK1K single-cell RNA-seq and genotype data are available through the Gene Expression Omnibus under accession GSE196830. The analyses in this study used normalized gene expression matrices, imputed genotype data, and covariate files provided directly by the study authors. The dendritic cell subset marker genes used in the enrichment analysis were obtained from the supplementary tables of Villani

et al. [40]. The scarf-QTL source code and analysis scripts are available at GitHub: https://github.com/luy21-thu/scarf-QTL/.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (T2322017) to L.H.

Authors' contributions

L.H. and Y.L. conceived the study. Y.L. developed the method and conducted the study. N.S. contributed to method development and interpretation of the results. Z.W. and L.H. supervised all aspects of the project. All authors edited and approved the final manuscript.

Acknowledgements

We thank the investigators of the original OneK1K study for making the data available.

References

[1] 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, 2008.

[2] GTEx Consortium. The gtex consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, 2020.

[3] Nurlan Kerimov, James D Hayhurst, Kateryna Peikova, Jonathan R Manning, Peter Walter, Liis Kolberg, Marija Samoviča, Manoj Pandian Sakthivel, Ivan Kuzmin, Stephen J Trevanion, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nature genetics*, 53(9):1290–1299, 2021.

[4] Harm-Jan Westra, Danny Arends, Tõnu Esko, Marjolein J Peters, Claudia Schurmann, Katharina Schramm, Johannes Kettunen, Hanieh Yaghootkar, Benjamin P Fairfax, Anand Kumar Andiappan, et al. Cell specific eqtl analysis without sorting cells. *PLoS genetics*, 11(5):e1005223, 2015.

[5] Zepeng Mu, Wei Wei, Benjamin Fair, Jinlin Miao, Ping Zhu, and Yang I Li. The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome biology*, 22(1):122, 2021.

[6] Zhihong Zhu, Futao Zhang, Han Hu, Andrew Bakshi, Matthew R Robinson, Joseph E Powell, Grant W Montgomery, Michael E Goddard, Naomi R Wray,

Peter M Visscher, et al. Integration of summary data from gwas and eqtl studies predicts complex trait gene targets. *Nature genetics*, 48(5):481–487, 2016.

[7] Farhad Hormozdiari, Martijn Van De Bunt, Ayellet V Segre, Xiao Li, Jong Wha J Joo, Michael Bilow, Jae Hoon Sul, Sriram Sankararaman, Bogdan Pasaniuc, and Eleazar Eskin. Colocalization of gwas and eqtl signals detects target genes. *The American Journal of Human Genetics*, 99(6):1245–1260, 2016.

[8] Hyun Min Kang, Meena Subramaniam, Sasha Targ, Michelle Nguyen, Lenka Maliskova, Elizabeth McCarthy, Eunice Wan, Simon Wong, Lauren Byrnes, Cristina M Lanata, et al. Multiplexed droplet single-cell rna-sequencing using natural genetic variation. *Nature biotechnology*, 36(1):89–94, 2018.

[9] Monique GP Van Der Wijst, Harm Brugge, Dylan H De Vries, Patrick Deelen, Morris A Swertz, LifeLines Cohort Study, BIOS Consortium, and Lude Franke. Single-cell rna sequencing identifies celltype-specific cis-eqtls and co-expression qtls. *Nature genetics*, 50(4):493–497, 2018.

[10] Drew Neavin, Quan Nguyen, Maciej S Daniszewski, Helena H Liang, Han Sheng Chiu, Yong Kiat Wee, Anne Senabouth, Samuel W Lukowski, Duncan E Crombie, Grace E Lidgerwood, et al. Single cell eqtl analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome biology*, 22:1–19, 2021.

[11] Julien Bryois, Daniela Calini, Will Macnair, Lynette Foo, Eduard Urich, Ward Ortmann, Victor Alejandro Iglesias, Suresh Selvaraj, Erik Nutma, Manuel

721 Marzin, et al. Cell-type-specific cis-eqtls in eight human brain cell types identify
722 novel risk genes for psychiatric and neurological disorders. *Nature neuroscience*,
723 25(8):1104–1112, 2022.

724 [12] Seyhan Yazar, Jose Alquicira-Hernandez, Kristof Wing, Anne Senabouth,
725 M Grace Gordon, Stacey Andersen, Qinyi Lu, Antonia Rowson, Thomas RP
726 Taylor, Linda Clarke, et al. Single-cell eqtl mapping identifies cell type–specific
727 genetic control of autoimmune disease. *Science*, 376(6589):eabf3041, 2022.

728 [13] Mirko Francesconi and Ben Lehner. The effects of genetic variation on gene
729 expression dynamics during development. *Nature*, 505(7482):208–211, 2014.

730 [14] Avital Brodt, Maya Botzman, Eyal David, and Irit Gat-Viks. Dissecting dy-
731 namic genetic variation that controls temporal gene response in yeast. *PLoS*
732 *Computational Biology*, 10(12):e1003984, 2014.

733 [15] Mina Samad-Zamini, Wolfgang Schweiger, Thomas Nussbaumer, Klaus FX
734 Mayer, and Hermann Buerstmayr. Time-course expression qtl-atlas of the global
735 transcriptional response of wheat to fusarium graminearum. *Plant Biotechnology*
736 *Journal*, 15(11):1453–1464, 2017.

737 [16] BJ Strober, Reem Elorbany, Katherine Rhodes, Nirmal Krishnan, Karl Tayeb,
738 Alexis Battle, and Yoav Gilad. Dynamic genetic regulation of gene expression
739 during cellular differentiation. *Science*, 364(6447):1287–1290, 2019.

740 [17] Anna SE Cuomo, Daniel D Seaton, Davis J McCarthy, Iker Martinez, Marc Jan
741 Bonder, Jose Garcia-Bernardo, Shradha Amatya, Pedro Madrigal, Abigail Isaac-

son, Florian Buettner, et al. Single-cell rna-sequencing of differentiating ips cells reveals dynamic genetic effects on gene expression. *Nature communications*, 11(1):810, 2020.

[18] Reem Elorbany, Joshua M Popp, Katherine Rhodes, Benjamin J Strober, Kenneth Barr, Guanghao Qi, Yoav Gilad, and Alexis Battle. Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLoS genetics*, 18(1):e1009666, 2022.

[19] Aparna Nathan, Samira Asgari, Kazuyoshi Ishigaki, Cristian Valencia, Tiffany Amariuta, Yang Luo, Jessica I Beynor, Yuriy Baglaenko, Sara Suliman, Alkes L Price, et al. Single-cell eqtl models reveal dynamic t cell state dependence of disease loci. *Nature*, 606(7912):120–128, 2022.

[20] Joshua M Popp, Katherine Rhodes, Radhika Jangi, Mingyuan Li, Kenneth Barr, Karl Tayeb, Alexis Battle, and Yoav Gilad. Cell type and dynamic state govern genetic regulation of gene expression in heterogeneous differentiating cultures. *Cell Genomics*, 4(12), 2024.

[21] Wenhe Lin, Mingyuan Li, Olivia Allen, Jonathan Burnett, Joshua M Popp, Matthew Stephens, Alexis Battle, and Yoav Gilad. Disease-associated loci share properties with response eqtls under common environmental exposures. *bioRxiv*, pages 2025–04, 2025.

[22] Joyce B Kang, Amber Z Shen, Saisriram Gurajala, Aparna Nathan, Laurie Rumker, Vitor RC Aguiar, Cristian Valencia, Kaitlyn A Lagattuta, Fan Zhang,

Anna Helena Jonsson, et al. Mapping the dynamic genetic regulatory architecture of hla genes at single-cell resolution. *Nature genetics*, 55(12):2255–2268, 2023.

[23] Richard K Perez, M Grace Gordon, Meena Subramaniam, Min Cheol Kim, George C Hartoularos, Sasha Targ, Yang Sun, Anton Ogorodnikov, Raymund Bueno, Andrew Lu, et al. Single-cell rna-seq reveals cell type–specific molecular and genetic associations to lupus. *Science*, 376(6589):eabf1970, 2022.

[24] Anna SE Cuomo, Tobias Heinen, Danai Vagiaki, Danilo Horta, John C Marioni, and Oliver Stegle. Cellregmap: a statistical framework for mapping context-specific regulatory variants using scrna-seq. *Molecular Systems Biology*, 18(8):e10663, 2022.

[25] Natsuhiko Kumasaka, Raghd Rostom, Ni Huang, Krzysztof Polanski, Kerstin B Meyer, Sharad Patel, Rachel Boyd, Celine Gomez, Sam N Barnett, Nikolaos I Panousis, et al. Mapping interindividual dynamics of innate immune response at single-cell resolution. *Nature Genetics*, 55(6):1066–1075, 2023.

[26] Benjamin J Strober, Karl Tayeb, Joshua Popp, Guanghao Qi, M Grace Gordon, Richard Perez, Chun Jimmie Ye, and Alexis Battle. Surge: uncovering context-specific genetic-regulation of gene expression from single-cell rna sequencing using latent-factor models. *Genome Biology*, 25(1):28, 2024.

[27] Anna SE Cuomo, Eleanor Spenceley, Hope A Tanudisastro, Blake Bowen, Albert Henry, Hao Lawrence Huang, Angli Xue, Wei Zhou, Matthew J Welland,

Arthur S Lee, et al. Impact of rare and common genetic variation on cell type-specific gene expression. *medRxiv*, pages 2025–03, 2025.

[28] Kian Hong Kock, Kyung Yeon Han, Yoshinari Ando, Damita Jevapatarakul, Ankita Chatterjee, Quy Xiao Xuan Lin, Eliora Violain Buyamin, Radhika Sonthalia, Deepa Rajagopalan, Yoshihiko Tomofuji, et al. Asian diversity in human immune cells. *Cell*, 188(8):2288–2306, 2025.

[29] Joyce B Kang, Alessandro Raveane, Aparna Nathan, Nicole Soranzo, and Soumya Raychaudhuri. Methods and insights from single-cell expression quantitative trait loci. *Annual review of genomics and human genetics*, 24(1):277–303, 2023.

[30] Zixuan Eleanor Zhang, Artem Kim, Noah Suboc, Nicholas Mancuso, and Steven Gazal. Efficient count-based models improve power and robustness for large-scale single-cell eqtl mapping. *medRxiv*, pages 2025–01, 2025.

[31] Wei Zhou, Anna SE Cuomo, Angli Xue, Masahiro Kanai, Grant Chau, Chirag Krishna, Ramnik J Xavier, Daniel G MacArthur, Joseph E Powell, Mark J Daly, et al. Efficient and accurate mixed model association tool for single-cell eqtl analysis. *medRxiv*, pages 2024–05, 2024.

[32] Matthew W Funk, Yuhe Wang, and Lingfei Wang. Airqtl dissects cell state-specific causal gene regulatory networks with efficient single-cell eqtl mapping. *Nature Communications*, 2025.

[33] Wensheng Guo. Functional mixed effects models. *Biometrics*, 58(1):121–128, 2002.

[34] Wenpin Hou, Zhicheng Ji, Zeyu Chen, E John Wherry, Stephanie C Hicks, and Hongkai Ji. A statistical framework for differential pseudotime analysis with multiple single-cell rna-seq samples. *Nature communications*, 14(1):7286, 2023.

[35] Gang Xu, Amei Amei, Weimiao Wu, Yunqing Liu, Linchuan Shen, Edwin C Oh, and Zuoheng Wang. Retrospective varying coefficient association analysis of longitudinal binary traits: application to the identification of genetic loci associated with hypertension. *The annals of applied statistics*, 18(1):487, 2024.

[36] Yaowu Liu and Jun Xie. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529):393–402, 2020.

[37] Paul HC Eilers and Brian D Marx. Flexible smoothing with b-splines and penalties. *Statistical science*, 11(2):89–121, 1996.

[38] Yaowu Liu, Sixing Chen, Zilin Li, Alanna C Morrison, Eric Boerwinkle, and Xihong Lin. Acat: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics*, 104(3):410–421, 2019.

[39] John D Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

[40] Alexandra-Chloé Villani, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, Andrew Butler, Shiwei Zheng, Suzan Lazo, Laura Jardine, David Dixon, Emily Stephenson, Emil Nilsson, Ida Grundberg, David McDonald, Andrew Filby, Weibo Li, Philip L. De Jager, Orit Rozenblatt-Rosen, Andrew A. Lane, Muzlifah Haniffa, Aviv Regev, and Nir Hacohen. Single-cell rna-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, 356(6335):eaah4573, 2017.

[41] Christophe Macri, Ee Shan Pang, Timothy Patton, and Meredith O'Keeffe. Dendritic cell subsets. In *Seminars in cell & developmental biology*, volume 84, pages 11–21. Elsevier, 2018.

[42] Yongtao Guan and Daniel Levy. Asymptotically exact fit for linear mixed model in genetic association studies. *Genetics*, 228(2):iyae143, 2024.

[43] Christoph Lippert, Jennifer Listgarten, Ying Liu, Carl M Kadie, Robert I Davidson, and David Heckerman. Fast linear mixed models for genome-wide association studies. *Nature methods*, 8(10):833–835, 2011.

[44] Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1):s13742–015, 2015.

[45] Christoph Hafemeister and Rahul Satija. Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression. *Genome biology*, 20(1):296, 2019.

[46] Kevin R Moon, David Van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen, Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492, 2019.

[47] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):477, 2018.

5 Supplementary Materials

5.1 Key steps of the IDUL algorithm

In order to diagonalize the covariance matrix of $\boldsymbol{y}$, we perform the eigen decomposition $\Phi\Phi^\top = QDQ^\top$, where Q is orthogonal and D is diagonal. It follows that $LL(\theta)$ equals the log-likelihood of $Q^\top \boldsymbol{y}$, whose covariance matrix $\tau^{-1}(\eta D + I_C)$ is diagonal,

$$LL(\theta) = -\frac{1}{2}\left\{C\log\frac{2\pi}{\tau} + \log\left|Q(\eta D + I_C)Q^\top\right| + \tau(\boldsymbol{y} - X\boldsymbol{\beta})^\top[Q(\eta D + I_C)Q^\top]^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})\right\} \tag{48}$$

$$= -\frac{1}{2}\left\{C\log\frac{2\pi}{\tau} + \log|\eta D + I_C| + \tau(\boldsymbol{y} - X\boldsymbol{\beta})^\top Q(\eta D + I_C)^{-1}Q^\top(\boldsymbol{y} - X\boldsymbol{\beta})\right\}. \tag{49}$$

Based on the transformed log-likelihood, the marginal log-likelihood function for

η is

$$f_{ml}(\eta) = LL(\hat{\tau}(\eta), \eta, \hat{\boldsymbol{\beta}}(\eta)) = -\frac{1}{2}\left\{C \log \frac{2\pi}{C} + \log|H| + C + C \log(\boldsymbol{y}_Q^\top P_X \boldsymbol{y}_Q)\right\},$$

where $\hat{\tau}(\eta)$ and $\hat{\boldsymbol{\beta}}(\eta)$ are obtained analytically by maximum likelihood for a given η, and $H = \eta D + I_C$, $P_X = H^{-1} - H^{-1}X_Q(X_Q^\top H^{-1} X_Q)^{-1} X_Q^\top H^{-1}$.

Iterative dispersion updates are applied to maximize $f_{ml}(\eta)$. In short, η is updated by repeating the following steps:

- For a given η, weighted least squares of $\boldsymbol{y}_Q$ on X_Q with weight H^{-1} is performed, and the residual vector $\boldsymbol{r}$ is used to compute $\hat{t} = \frac{1}{C}\sum_i r_i^2 / H_{ii}$.

- Weighted least squares of $\boldsymbol{r}^2$ on $\mathrm{diag}\{D\}$ with weight H^{-2} is performed, and the resulting slope and intercept estimates, $\hat{\gamma}$ and $\hat{\mu}$, are used to compute $\eta^\dagger = \hat{\gamma}/\hat{t} + (1 - \hat{\mu}/\hat{t})\eta$.

In practice, we used the gradient-ascent version of IDUL (denoted IDUL†), together with the lazy update described in [42]. Taking advantage of the sparse structure of X_Q, we performed weighted least squares via Cholesky decomposition using the "SparseM" R package, rather than the QR decomposition typically used for linear regression in R.

5.2 Derivation of the prospective score function and Fisher information matrix

To implement the prospective dynamic score test, the score function and the Fisher information matrix are derived from the first and second derivatives of the log-likelihood function for model (8), taking the form

$$U(\theta) = \begin{pmatrix} U_\tau(\theta) \\ U_\eta(\theta) \\ U_{\boldsymbol{\beta}}(\theta) \end{pmatrix}, \; I(\theta) = \begin{bmatrix} I_\tau(\theta) & I_{\tau,\eta}(\theta) & I_{\tau,\boldsymbol{\beta}}(\theta) \\ I_{\eta,\tau}(\theta) & I_\eta(\theta) & I_{\eta,\boldsymbol{\beta}}(\theta) \\ I_{\boldsymbol{\beta},\tau}(\theta) & I_{\boldsymbol{\beta},\eta}(\theta) & I_{\boldsymbol{\beta}}(\theta) \end{bmatrix}, \tag{50}$$

and the score statistic is given by

$$T^{\text{prosp}}_{\text{dynamic}} = U^\top\left(\hat{\theta}_0\right)\left[I\left(\hat{\theta}_0\right)\right]^{-1} U\left(\hat{\theta}_0\right), \tag{51}$$

where $\hat{\theta}_0$ is the MLE under the null hypothesis.

Direct calculation shows that

$$I_{\tau,\boldsymbol{\beta}}(\theta) = \mathbf{0}, \; I_{\eta,\boldsymbol{\beta}}(\theta) = \mathbf{0}, \; I_{\boldsymbol{\beta}}(\theta) = \tau X^\top\left(\eta\Phi\Phi^\top + I_C\right)^{-1} X, \tag{52}$$

and, when evaluated at the MLE under the null hypothesis, the score functions satisfy

$$U_\tau\left(\hat{\theta}_0\right) = \mathbf{0}, \; U_\eta\left(\hat{\theta}_0\right) = \mathbf{0}, \; U_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right) = \begin{pmatrix} \mathbf{0} \\ U_{\text{dynamic}} \end{pmatrix}, \tag{53}$$

$$U_{\text{dynamic}} = \hat{\tau}_0\left(X^{\text{SNP}}\right)^\top\left(\hat{\eta}_0\Phi\Phi^\top + I_C\right)^{-1}\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}_0\right) = \sum_i g_i^{\text{SNP}}\Phi_i^\top \boldsymbol{e}_i. \tag{54}$$

The score statistic therefore simplifies to

$$T_{\text{dynamic}}^{\text{prosp}} = U_{\boldsymbol{\beta}}^{\top}\left(\hat{\theta}_0\right)\left[I_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right)\right]^{-1} U_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right) = U_{\text{dynamic}}^{\top}\left[\widehat{\text{Var}}_0\left(U_{\text{dynamic}}\right)\right]^{-1} U_{\text{dynamic}}, \quad (55)$$

where $\left[\widehat{\text{Var}}_0\left(U_{\text{dynamic}}\right)\right]^{-1}$ is the lower-right submatrix of $\left[I_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right)\right]^{-1}$ corresponding to $\boldsymbol{\beta}^{\text{SNP}}$.

5.3 Simplification of the computation of the prospective variance

For convenience, we use the subscripts U and V to denote quantities left-multiplied by these eigenvector matrices in this section. The core step of calculating the prospective variance is to compute $I_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right)$:

$$I_{\boldsymbol{\beta}}\left(\hat{\theta}_0\right) = \hat{\tau}_0 X^{\top}\left(\hat{\eta}_0 \Phi\Phi^{\top} + I_C\right)^{-1} X = \hat{\tau}_0 X_Q^{\top}\left(\hat{\eta}_0 D + I_C\right)^{-1} X_Q \quad (56)$$

$$= \hat{\tau}_0 \begin{bmatrix} X_U^{\top} & X_V^{\top} \end{bmatrix} \begin{bmatrix} \left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & I_{n^V} \end{bmatrix} \begin{bmatrix} X_U \\ X_V \end{bmatrix} = \hat{\tau}_0 X_V^{\top} X_V + \hat{\tau}_0 X_U^{\top}\left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U. \quad (57)$$

The first term does not involve the genotype because $X_V^{\text{SNP}} = \mathbf{0}$, as genotypes are individual-level predictors (see equation (34)). For the second term, we have

$$X_U^{\top}\left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U = \begin{bmatrix} \left(X_U^{\text{COV}}\right)^{\top}\left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\text{COV}} & \left(X_U^{\text{COV}}\right)^{\top}\left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\text{SNP}} \\ \left(X_U^{\text{SNP}}\right)^{\top}\left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\text{COV}} & \left(X_U^{\text{SNP}}\right)^{\top}\left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\text{SNP}} \end{bmatrix}, \quad (58)$$

and to enable parallel computation for multiple SNPs, we rewrite

$$\left(X_U^{\mathrm{COV}}\right)^\top \left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\mathrm{SNP}} = \sum_i g_i^{\mathrm{SNP}} \left(X_{U,i}^{\mathrm{COV}}\right)^\top \left(\hat{\eta}_0\Lambda_i + I_{n_i^U}\right)^{-1} U_i^\top \Phi_i. \quad (59)$$

Here, $X_{U,i}^{\mathrm{COV}} = U_i^\top X_i^{\mathrm{COV}}$, and $M_{V^{12},i} = \left(X_{U,i}^{\mathrm{COV}}\right)^\top \left(\hat{\eta}_0\Lambda_i + I_{n_i^U}\right)^{-1} U_i^\top \Phi_i$ is a $p \times K$ matrix for each i, where p is the number of columns in X^{COV}. Then, we reshape each matrix $M_{V^{12},i}$ into a vector of length pK, and stack the resulting vectors row-wise to obtain $M_{V_{12}}$. Thus, $M_{V_{12}}^\top G^{\mathrm{SNP}}$ can be reshaped to obtain $\left(X_U^{\mathrm{COV}}\right)^\top \left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\mathrm{SNP}}$, enabling simultaneous computation for multiple SNPs through matrix multiplication. Similarly, since

$$\left(X_U^{\mathrm{SNP}}\right)^\top \left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\mathrm{SNP}} = \sum_i \left(g_i^{\mathrm{SNP}}\right)^2 \left(U_i^\top \Phi_i\right)^\top \left(\hat{\eta}_0\Lambda_i + I_{n_i^U}\right)^{-1} U_i^\top \Phi_i, \quad (60)$$

we construct $M_{V_{22}}$ using the reshaped matrices $\left(U_i^\top \Phi_i\right)^\top \left(\hat{\eta}_0\Lambda_i + I_{n_i^U}\right)^{-1} U_i^\top \Phi_i$, and multiply the squared genotype matrix (consisting of $G^{\mathrm{SNP}} \odot G^{\mathrm{SNP}}$ for multiple SNPs, where $\odot$ represents element-wise multiplication) by $M_{V_{22}}$ to obtain the reshaped $\left(X_U^{\mathrm{SNP}}\right)^\top \left(\hat{\eta}_0\Lambda + I_{n^U}\right)^{-1} X_U^{\mathrm{SNP}}$.

5.4 Algorithms

Inputs of scarf-QTL:

- Gene expression matrix $Y = \left[Y_1^\top, \ldots, Y_I^\top\right]^\top \in \mathbb{R}^{C \times G}$, where each column corresponds to $\boldsymbol{y}$ in the "Methods" section and $Y_i \in \mathbb{R}^{c_i \times G}$ is the gene expression matrix for individual i. The rank-based inverse normal transformation is per-

formed on raw counts of each gene by default

- Cell-state coordinates $T \in \mathbb{R}^C$, and number of B-spline basis functions K
- Covariates including
  - Individual-level covariates with static effects $W^{\text{indi}} \in \mathbb{R}^{I \times J_1}$
  - Individual-level covariates with cell-state-dependent effects $Z^{\text{indi}} \in \mathbb{R}^{I \times J_2}$
  - Cell-level covariates with static effects $W^{\text{cell}} \in \mathbb{R}^{C \times J_3}$
  - Cell-level covariates with cell-state-dependent effects $Z^{\text{cell}} \in \mathbb{R}^{C \times J_4}$
- Genotype matrix for S SNPs $G^{\text{SNPs}} \in \mathbb{R}^{I \times S}$, and the list of cis-SNPs for each gene
- Genetic relationship matrix estimate $\hat{\Sigma}$

For implementation, we distinguish between individual-level and cell-level covariates. To align with the notations in the "Methods" section, the design matrix for individual i, X_i, is given by

$$X_i = \begin{bmatrix} X_i^{\text{indi}} & X_i^{\text{cell}} \end{bmatrix} = \left[\begin{bmatrix} W_i^{\text{indi}} \otimes \mathbf{1}_{c_i} & Z_i^{\text{indi}} \otimes \Phi_i \end{bmatrix} \quad \begin{bmatrix} W_i^{\text{cell}} & Z_i^{\text{cell}} \otimes_{\text{row}} \Phi_i \end{bmatrix} \right] \in \mathbb{R}^{c_i \times (J_1 + KJ_2 + J_3 + KJ_4)}, \tag{61}$$

where $\otimes$ represents Kronecker product, and $\otimes_{\text{row}}$ denotes the row-wise Kronecker product.

Algorithm of scarf-QTL:

1. (Eigen transformation) Perform eigen decomposition and rotate the gene expression and covariate matrices

(a) Evaluate K B-spline functions for T, resulting in Φ_i for all i

(b) Perform singular value decomposition of Φ_i to obtain U_i, V_i, Λ_i, R_i for all i

(c) Rotate Y to Y_Q as equation (32), where the vector $\boldsymbol{y}_i$ is replaced by the matrix $Y_i \in \mathbb{R}^{c_i \times G}$

(d) Construct X_i^{cell} for all i as in equation (61), and rotate them to obtain X_U^{cell} and X_V^{cell} as in equation (33)

(e) Compute $\Phi_{U,i} = U_i^\top \Phi_i = \Lambda_i^{1/2} R_i^\top$, and $\mathbf{1}_{U,i} = U_i^\top \mathbf{1}_{c_i}$, and construct X_U^{indi} from $U_i^\top X_i^{\text{indi}} = \begin{bmatrix} W_i^{\text{indi}} \otimes \mathbf{1}_{U,i} & Z_i^{\text{indi}} \otimes \Phi_{U,i} \end{bmatrix}$

(f) Combine $X_Q^{\text{COV}} = \begin{bmatrix} X_U^{\text{cell}} & X_U^{\text{indi}} \\ X_V^{\text{cell}} & \mathbf{0} \end{bmatrix}$

(g) Concatenate the eigenvalues in Λ_i to form D

2. (Null model fitting) For each gene g, select the column $\boldsymbol{y}_Q^{(g)}$ from Y_Q, and fit the null model by performing IDUL or IDUL† on

$$\boldsymbol{y}_Q^{(g)} \sim \text{MVN}\left(X_Q^{\text{COV}} \boldsymbol{\beta}, \tau^{-1}(\eta D + I_C)\right). \tag{62}$$

Obtain the rotated weighted residual vector $\boldsymbol{e}_i^{U(g)}$ in equation (37) from the weights and residuals of the IDUL regression, and return $\hat{\tau}^{(g)}$

3. (Retrospective association test) For each gene g, perform the retrospective association test:

(a) Calculate $\Phi_{U,i} \boldsymbol{e}_i^{U(g)}$ for each i, and stack them row-wise to obtain $M_{U(g)}$

(b) Select the cis-SNPs for this gene from the genotype matrix, namely $G^{\mathrm{SNP}(g)}$. Estimate MAF and variance ($\hat{p}$ and $\hat{\sigma}_g^2$) for each SNP

(c) Compute $U^{(g)}_{\mathrm{dynamic}} = M_{U(g)}^{\top} G^{\mathrm{SNP}(g)}$ by matrix multiplication for the score function, and $V^{(g)} = M_{U(g)}^{\top} \hat{\Sigma} M_{U(g)}$ for the common component of the retrospective variance

(d) For static test, calculate $U^{(g)}_{\mathrm{static}} = \mathbf{1}_K^{\top} U^{(g)}_{\mathrm{dynamic}}$ and calculate $\mathbf{1}_K^{\top} V^{(g)} \mathbf{1}_K$

(e) Calculate score statistics as in equation (26) and obtain p-values

(f) Apply the Cauchy combination method

4. (Optional) Perform the prospective score test:

(a) The score function has already been calculated in the retrospective association test

(b) Calculate $X_V^{\top} X_V$ using X_V^{cell}

(c) For each gene g, calculate $\left(X_U^{\mathrm{COV}}\right)^{\top} \left(\hat{\eta}_0 \Lambda + I\right)^{-1} X_U^{\mathrm{COV}}$, $M_{V_{12}}$, and $M_{V_{22}}$

(d) Perform the matrix multiplications $M_{V_{12}}^{\top} G^{\mathrm{SNP}(g)}$ and $M_{V_{22}}^{\top} \left(G^{\mathrm{SNP}(g)} \odot G^{\mathrm{SNP}(g)}\right)$ for each gene g, which yield $I_{\beta}(\hat{\theta}_0)$ for each cis-SNP (see the "Simplification of the computation of the prospective variance" section for details)

(e) Compute the prospective variance, score statistics, and p-values (see the "Methods" section "Prospective score tests for cell-state-dependent or constant eQTL effects" for details)

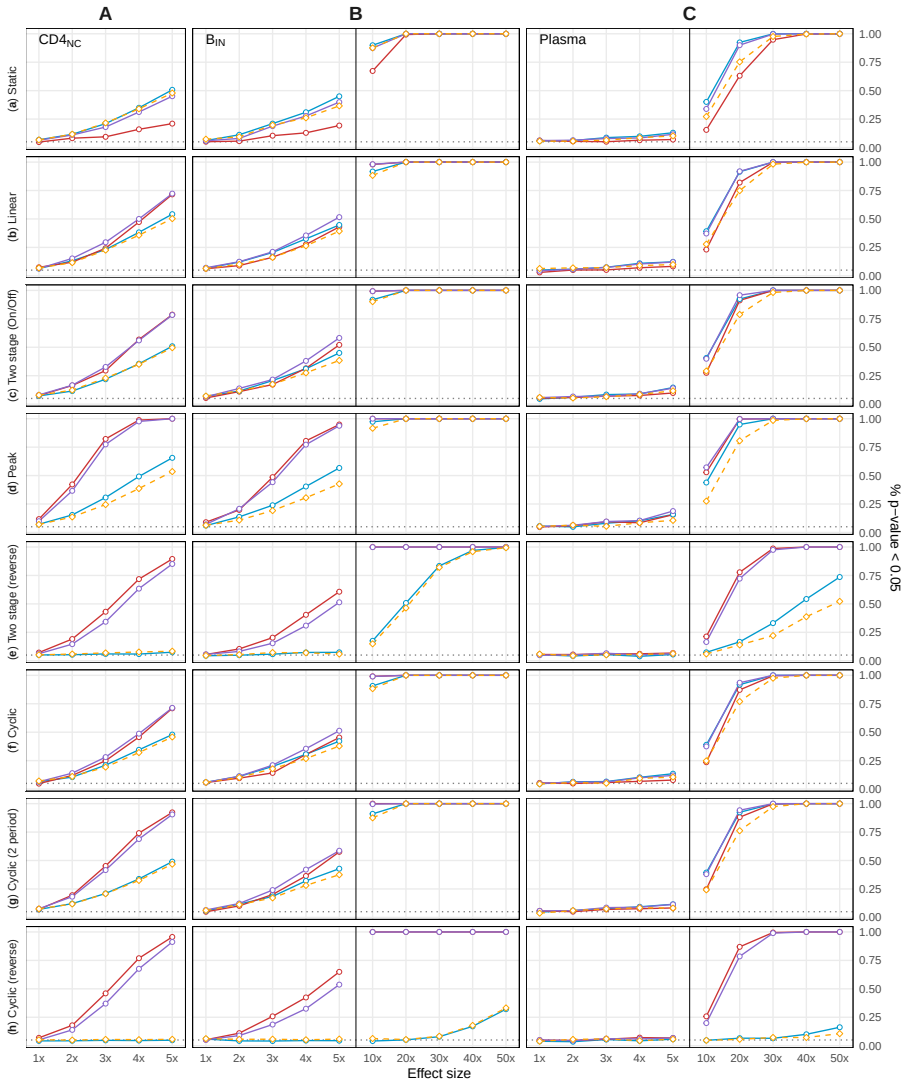


Figure S1: Statistical power of scarf-QTL under Poisson-generated count data. Power is evaluated across the same eQTL effect patterns considered in Fig. 2. For the two cell types with smaller numbers of cells (B_{IN} and plasma), additional effect size magnitudes (10×–50×) were considered to account for lower signal-to-noise ratios. All other simulation settings followed those in Fig. 2.

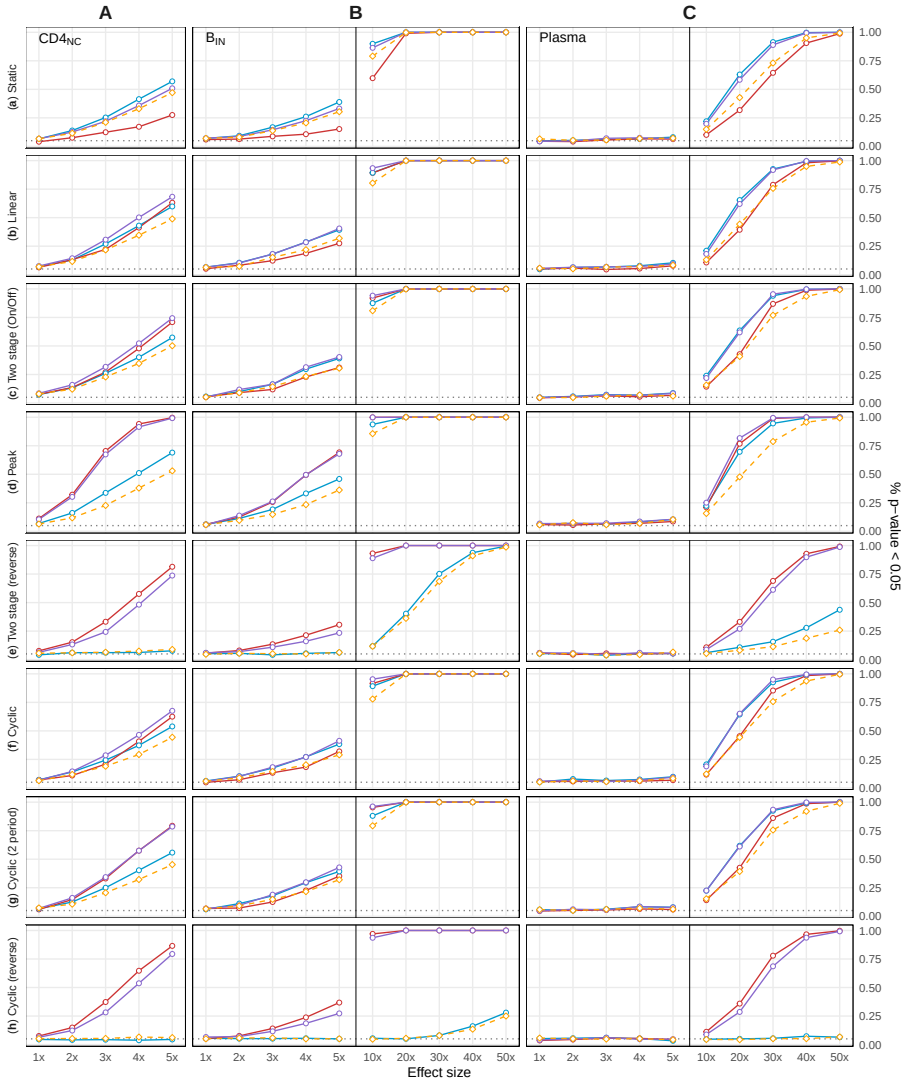


Figure S2: Statistical power of scarf-QTL under negative binomial-generated count data. Simulation settings are identical to those in Fig. S1, except that gene expression counts were generated from a negative binomial distribution.

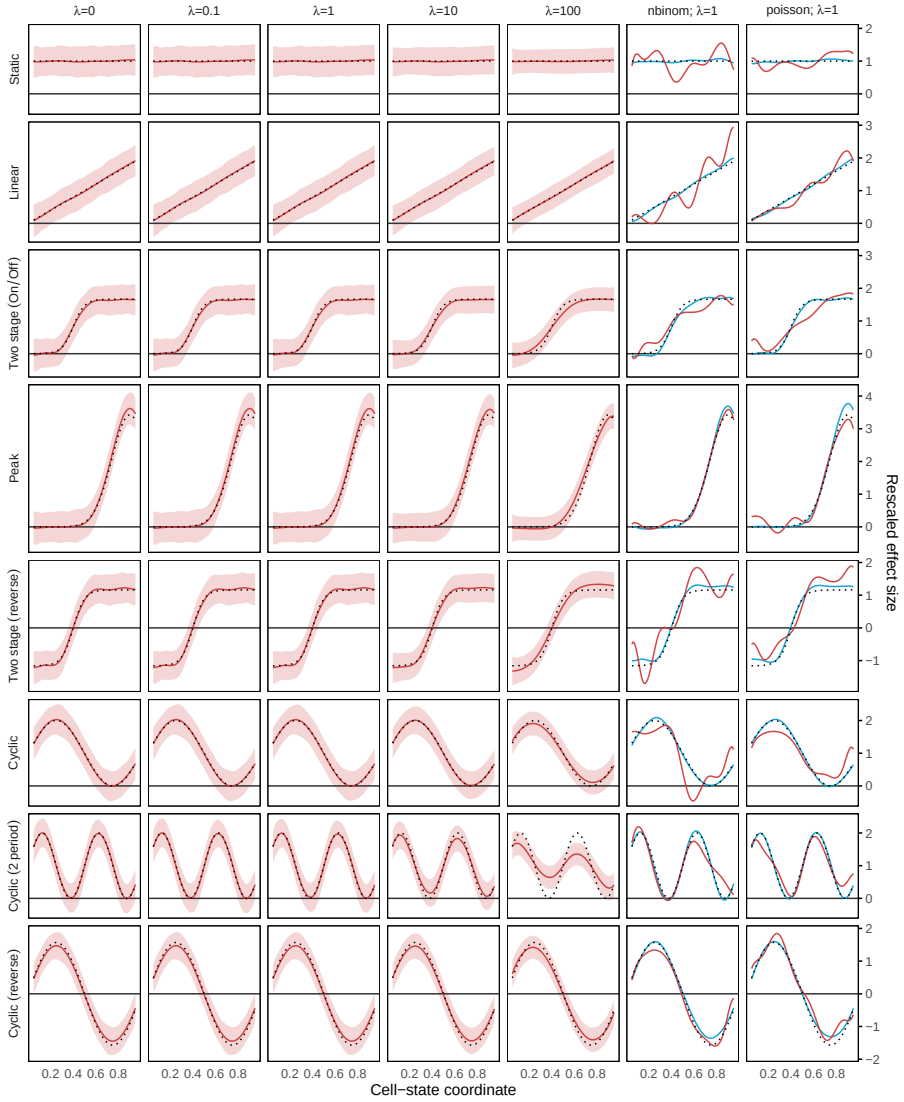


Figure S3: Evaluation of cell-state-dependent eQTL effect estimation across representative simulation settings. Rows correspond to the eQTL effect patterns used in the power analysis. Results are shown for the B_{IN} cell type, with effect size fixed at 5× unless otherwise specified. Columns represent combinations of the gene expression model and the penalty parameter λ used in penalized spline estimation. Columns 1–5 show the normal model with $\lambda \in \{0, 0.1, 1, 10, 100\}$, whereas columns 6 and 7 show the Poisson and negative binomial models with $\lambda = 1$. For the latter two columns, effect sizes of 5× and 50× are shown in red and blue, respectively. In each panel, the curve is the pointwise mean of the estimated eQTL effects across simulation replicates after rescaling the estimated effects such that $\int_0^1 |\hat{\gamma}^{\mathrm{SNP}}(t)|\, dt = 1$. For columns 1–5, the shaded band indicates one standard deviation across replicates. The true underlying effect pattern is shown as a black dotted line.

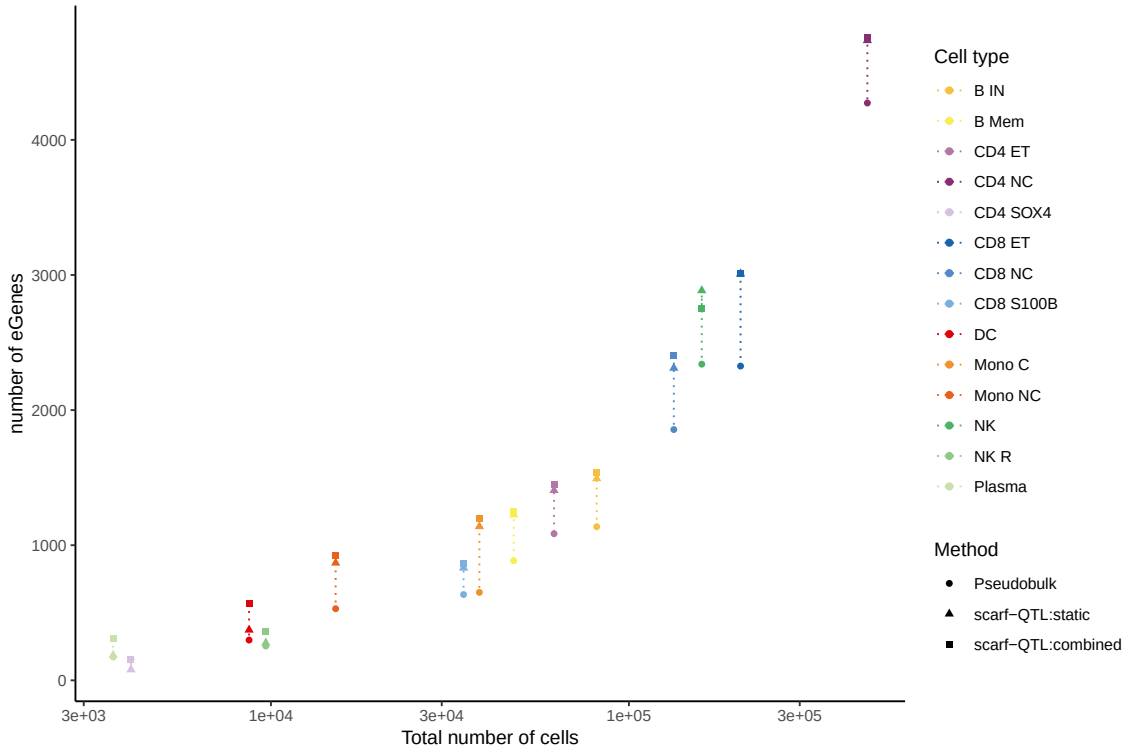


Figure S4: Number of eGenes identified by different approaches across cell types with varying numbers of cells.

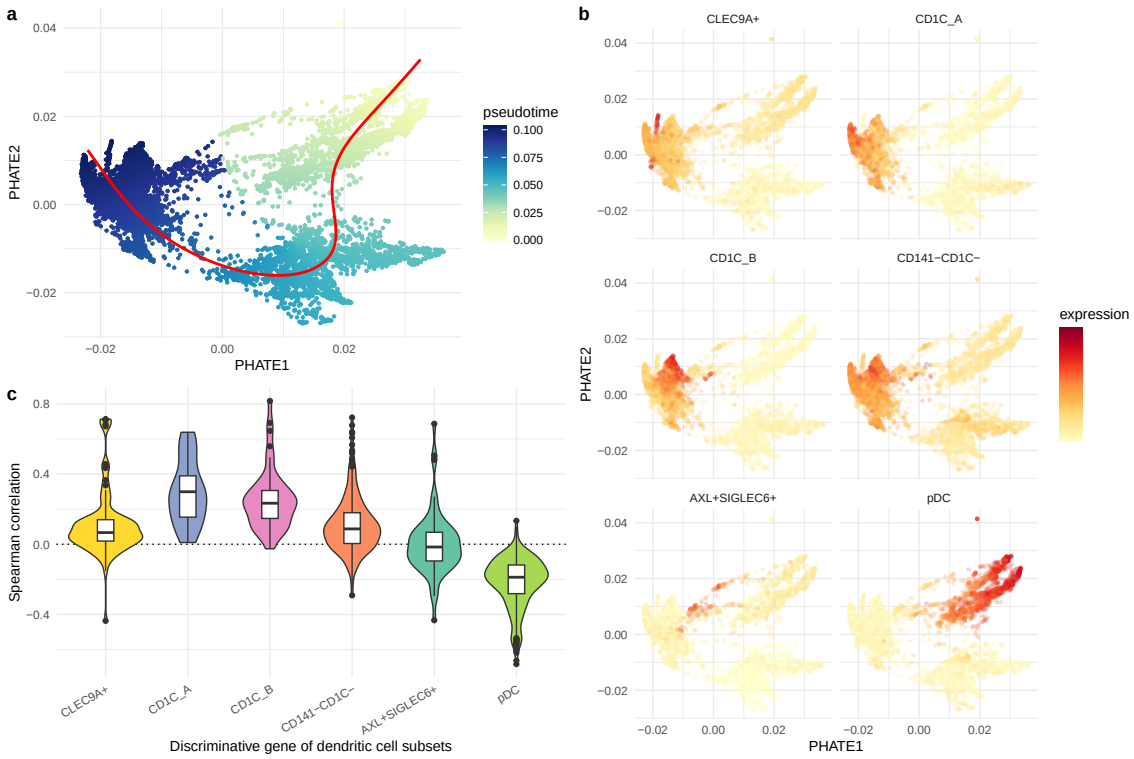


Figure S5: The pseudotime-derived cell-state coordinate captures dendritic-cell (DC) subset structure. **(a)** PHATE embedding of DCs colored by the cell-state coordinate; the red curve indicates the fitted trajectory. **(b)** Mean normalized expression across the top five subset markers for each DC subset, projected onto the same PHATE embedding. **(c)** Violin plots and boxplots of Spearman correlations between the cell-state coordinate and expression of subset markers for each DC subset.

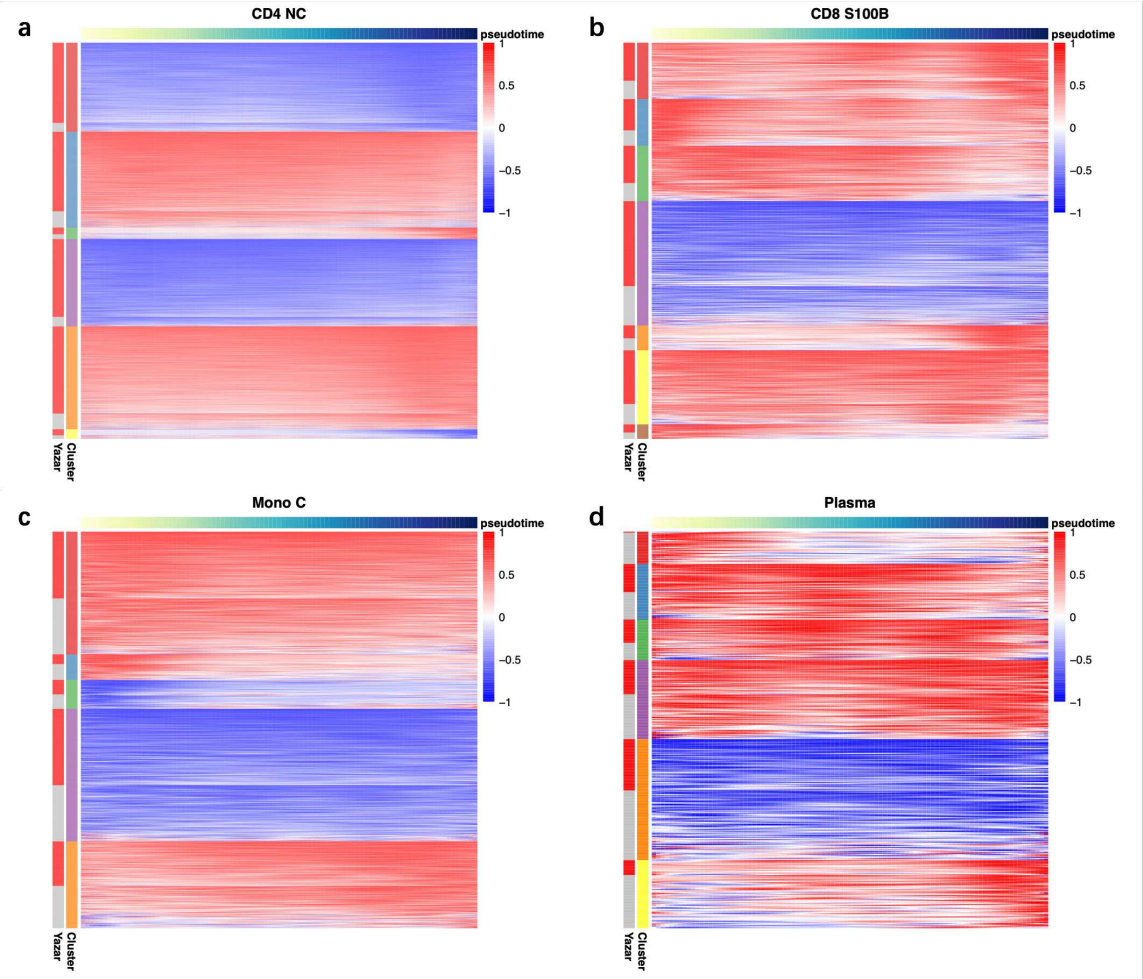


Figure S6: Cell-state-dependent regulatory patterns of eQTLs in four example cell types. The heatmap is constructed and interpreted in the same way as Fig. 3c, showing fitted eQTL effects as smooth functions of the cell-state coordinate.

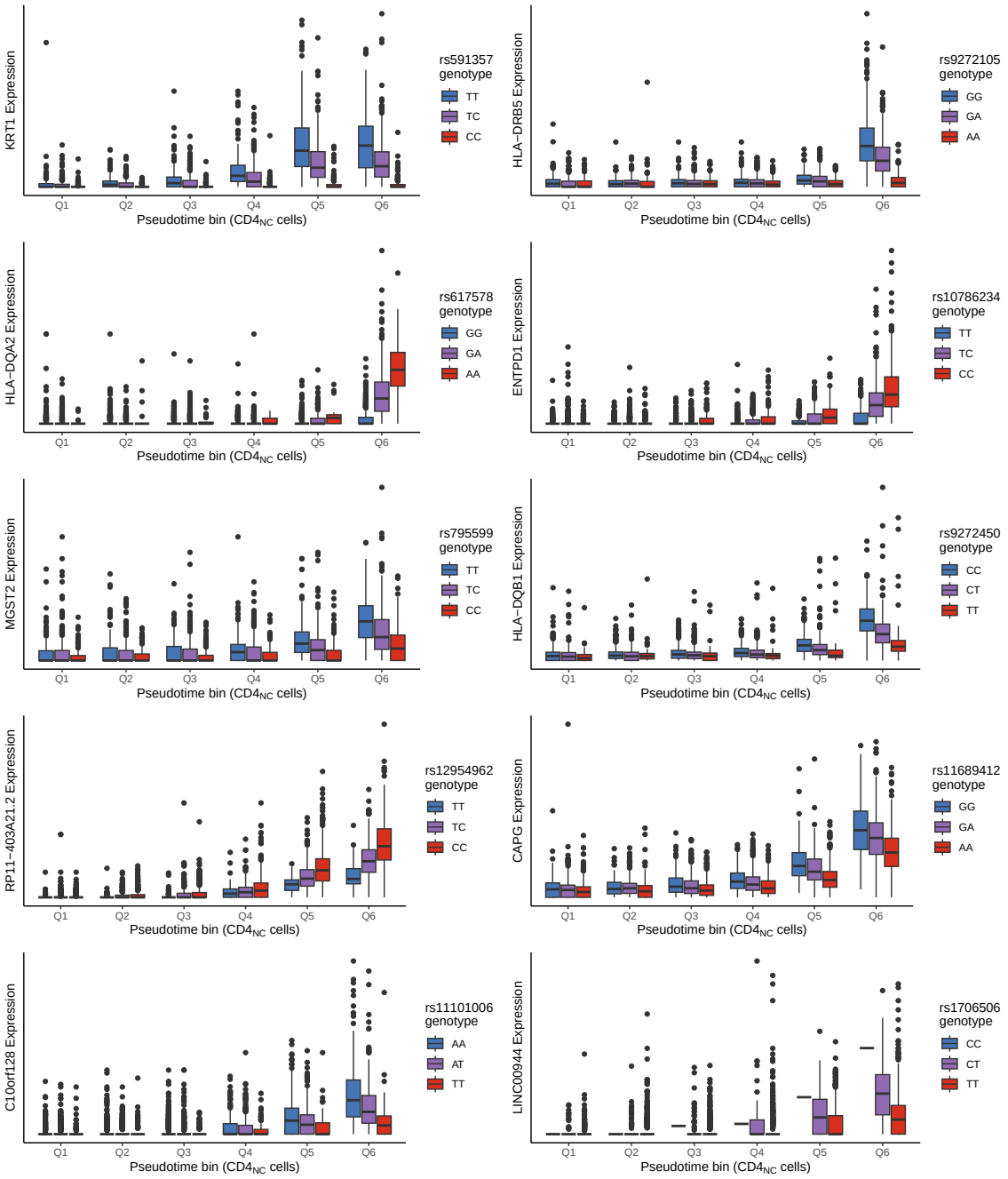


Figure S7: Representative cell-state-dependent eQTL examples in $CD4_{NC}$ T cells. Boxplots are constructed in the same manner as in Fig. 3b. Shown are representative eQTLs from clusters 3 and 6 in Fig. S6.

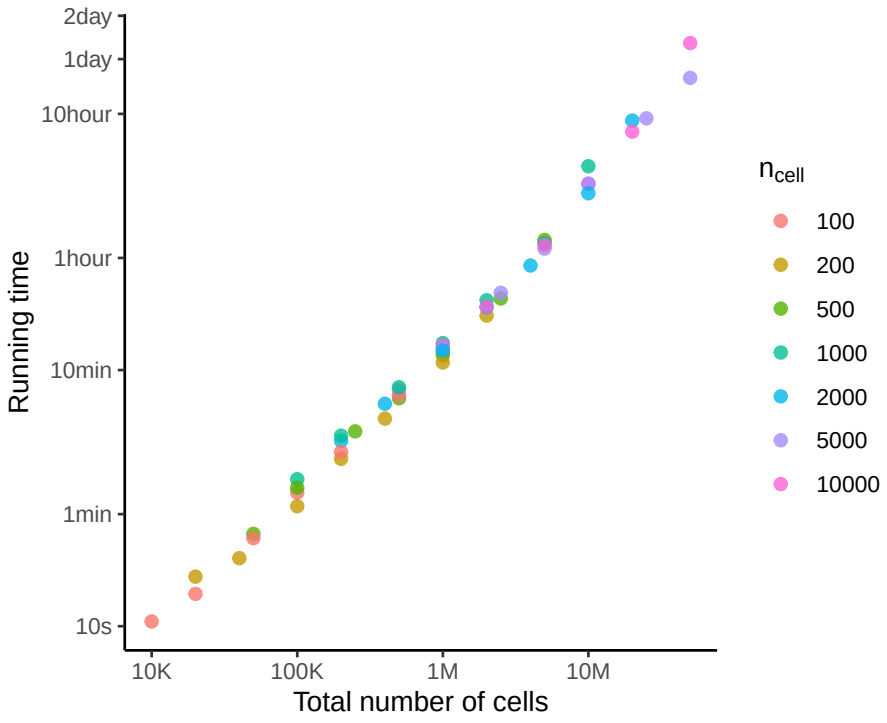


Figure S8: Runtime scalability of scarf-QTL across different sample sizes. Shown are empirical runtimes for eQTL mapping in resampled datasets with 100 genes. Runtime is shown as a function of the total number of cells ($n_{\text{indi}} \times n_{\text{cell}}$), across different combinations of n_{indi} and n_{cell}, while fixing n_{snp} at 3,000. Points are colored according to n_{cell}. Both axes are shown on logarithmic scales. A log-log regression yielded an empirical scaling relationship of runtime $\propto (n_{\text{indi}} \times n_{\text{cell}})^{1.057}$ ($R^2 = 0.991$).

5.5 Supplementary figures

5.6 Supplementary Tables

Table S1: Evaluation of type I error for scarf-QTL and benchmark methods ($CD4_{NC}$)

Model	Level	Scarf-QTL			Prospective Score Test		Pseudobulk
		Static	Dynamic	Combined	Static	Dynamic	
Normal	0.05	5.02×10^{-2}	4.96×10^{-2}	$\mathbf{5.19 \times 10^{-2}}$	5.04×10^{-2}	5.04×10^{-2}	4.73×10^{-2}
	0.01	1.01×10^{-2}	9.78×10^{-3}	$\mathbf{1.03 \times 10^{-2}}$	1.02×10^{-2}	1.01×10^{-2}	9.21×10^{-3}
	0.001	1.02×10^{-3}	9.65×10^{-4}	1.01×10^{-3}	1.03×10^{-3}	9.91×10^{-4}	9.00×10^{-4}
	0.0001	1.03×10^{-4}	7.60×10^{-5}	8.20×10^{-5}	1.04×10^{-4}	9.30×10^{-5}	9.40×10^{-5}
Poisson	0.05	5.05×10^{-2}	4.94×10^{-2}	$\mathbf{5.20 \times 10^{-2}}$	$\mathbf{2.68 \times 10^{-1}}$	$\mathbf{6.90 \times 10^{-2}}$	4.96×10^{-2}
	0.01	1.01×10^{-2}	9.83×10^{-3}	$\mathbf{1.04 \times 10^{-2}}$	$\mathbf{1.48 \times 10^{-1}}$	$\mathbf{2.80 \times 10^{-2}}$	9.79×10^{-3}
	0.001	1.07×10^{-3}	9.62×10^{-4}	1.07×10^{-3}	$\mathbf{6.76 \times 10^{-2}}$	$\mathbf{9.79 \times 10^{-3}}$	9.41×10^{-4}
	0.0001	1.10×10^{-4}	9.60×10^{-5}	1.07×10^{-4}	$\mathbf{3.24 \times 10^{-2}}$	$\mathbf{3.92 \times 10^{-3}}$	9.00×10^{-5}
Negative Binomial	0.05	5.01×10^{-2}	4.98×10^{-2}	$\mathbf{5.22 \times 10^{-2}}$	$\mathbf{1.96 \times 10^{-1}}$	$\mathbf{5.45 \times 10^{-2}}$	4.94×10^{-2}
	0.01	1.01×10^{-2}	9.93×10^{-3}	$\mathbf{1.03 \times 10^{-2}}$	$\mathbf{9.24 \times 10^{-2}}$	$\mathbf{1.65 \times 10^{-2}}$	9.70×10^{-3}
	0.001	1.03×10^{-3}	1.01×10^{-3}	9.96×10^{-4}	$\mathbf{3.40 \times 10^{-2}}$	$\mathbf{4.04 \times 10^{-3}}$	9.40×10^{-4}
	0.0001	1.17×10^{-4}	9.00×10^{-5}	9.30×10^{-5}	$\mathbf{1.33 \times 10^{-2}}$	$\mathbf{1.20 \times 10^{-3}}$	8.80×10^{-5}

Empirical type I error rates that fall outside the corresponding acceptance region with significance level 1% are shown in bold.

Table S2: Evaluation of type I error for scarf-QTL and benchmark methods (Plasma)

Model	Level	Scarf-QTL			Prospective Score Test		Pseudobulk
		Static	Dynamic	Combined	Static	Dynamic	
Normal	0.05	5.02×10^{-2}	4.99×10^{-2}	$\mathbf{5.23 \times 10^{-2}}$	$\mathbf{5.08 \times 10^{-2}}$	$\mathbf{5.09 \times 10^{-2}}$	4.86×10^{-2}
	0.01	9.94×10^{-3}	9.93×10^{-3}	$\mathbf{1.03 \times 10^{-2}}$	1.02×10^{-2}	$\mathbf{1.03 \times 10^{-2}}$	9.80×10^{-3}
	0.001	9.58×10^{-4}	1.02×10^{-3}	9.89×10^{-4}	9.72×10^{-4}	9.78×10^{-4}	9.33×10^{-4}
	0.0001	9.40×10^{-5}	1.10×10^{-4}	1.09×10^{-4}	1.12×10^{-4}	1.00×10^{-4}	9.80×10^{-5}
Poisson	0.05	4.96×10^{-2}	4.99×10^{-2}	$\mathbf{5.22 \times 10^{-2}}$	$\mathbf{7.54 \times 10^{-2}}$	4.70×10^{-2}	4.93×10^{-2}
	0.01	1.01×10^{-2}	1.01×10^{-2}	$\mathbf{1.04 \times 10^{-2}}$	$\mathbf{1.98 \times 10^{-2}}$	9.27×10^{-3}	9.90×10^{-3}
	0.001	1.01×10^{-3}	1.01×10^{-3}	1.01×10^{-3}	$\mathbf{3.06 \times 10^{-3}}$	8.85×10^{-4}	9.35×10^{-4}
	0.0001	1.02×10^{-4}	9.40×10^{-5}	9.20×10^{-5}	$\mathbf{4.83 \times 10^{-4}}$	9.30×10^{-5}	1.14×10^{-4}
Negative Binomial	0.05	4.96×10^{-2}	5.00×10^{-2}	$\mathbf{5.20 \times 10^{-2}}$	$\mathbf{5.96 \times 10^{-2}}$	$\mathbf{5.16 \times 10^{-2}}$	4.94×10^{-2}
	0.01	9.88×10^{-3}	1.01×10^{-2}	$\mathbf{1.05 \times 10^{-2}}$	$\mathbf{1.33 \times 10^{-2}}$	$\mathbf{1.05 \times 10^{-2}}$	9.81×10^{-3}
	0.001	9.79×10^{-4}	1.01×10^{-3}	1.04×10^{-3}	$\mathbf{1.57 \times 10^{-3}}$	$\mathbf{1.15 \times 10^{-3}}$	9.77×10^{-4}
	0.0001	1.06×10^{-4}	1.05×10^{-4}	1.13×10^{-4}	$\mathbf{1.93 \times 10^{-4}}$	$\mathbf{1.38 \times 10^{-4}}$	9.50×10^{-5}

Empirical type I error rates that fall outside the corresponding acceptance region with significance level 1% are shown in bold.