Flexible and efficient count-distribution and mixed-model methods for eQTL mapping with quasar

Jeffrey M. Pullin[1, *] and Chris Wallace[1, 2]

[1] *MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom*
[2] *Cambridge Institute of Therapeutic Immunology and Infectious Disease, University of Cambridge, Cambridge, United Kingdom*

[*] *Corresponding author* `jp2045@cam.ac.uk`

Abstract

Identifying genetic variants that affect gene expression, expression quantitative trait loci (eQTLs), is a major focus of modern genomics. Today, various methods exist for eQTL mapping, each using different statistical and methodological approaches. However, it is unclear which approaches lead to better performance, and challenges, particularly scalability as datasets continue to increase in size, remain. Here, we introduce quasar, a flexible and efficient C++ software program for eQTL mapping. Compared to existing eQTL mapping methods, quasar implements a wider variety of statistical models, including the linear model, Poisson and negative binomial generalised linear models, linear mixed model and Poisson and negative binomial generalised linear mixed models. Methodologically, we introduce and implement a simple, analytic approximation to the score test variance in mixed models. Furthermore, we highlight that difficulties with accurately estimating the negative binomial dispersion parameter, previously identified in the context of RNA-seq differential expression analysis, also apply to eQTL mapping. Therefore, quasar implements the Cox-Reid adjusted profile likelihood which enables unbiased estimation of the negative binomial dispersion parameter. We assess quasar's performance and compare it to three existing eQTL mapping methods: apex, jaxQTL and tensorQTL, on the OneK1K dataset. We demonstrate that quasar's output agrees with established methods where their models aligns but that quasar is between 5 to over 45 times faster. We exploit the range of models implemented in quasar to compare statistical models for eQTL mapping without confounding by implementation. We find that: count-based models have higher power, that mixed models do not show better performance in a dataset without substantial relatedness, and that the adjusted profile likelihood improves Type 1 error control when using the negative binomial distribution. Additionally, we investigate the relative performance of Poisson and negative binomial mixed models and the use of different approaches for gene-level FDR control. Overall, quasar provides a performant and versatile program for eQTL mapping and we nominate the negative binomial GLM model, incorporating adjusted profile likelihood dispersion estimation, as the statistical model with the best performance.

Introduction

A major focus of modern genomics is understanding how genetic variation affects gene expression. In particular, statistical methods have been used to identify genetic variants associated with the expression of genes: expression quantitative trait loci (eQTLs) [1]. Traditionally, gene expression has been measured using microarray or RNA-seq in bulk samples, but the cell-type specificity of

expression has motivated a recent shift to using single-cell RNA-seq [2, 3, 4]. Datasets of single-cell eQTLs are rapidly growing in size with the recent releases of the OneK1K dataset, including just under 1,000 individuals, and Phase 1 of the TenK10K project, which aims to map single-cell eQTLs in 10,000 individuals [3, 5]. A major use of eQTLs is linking variants identified by GWAS to affect disease risk to their causal gene and context, for example using colocalisation analysis [6]. Currently, however, there is still a substantial 'colocalisation gap', where many GWAS variants cannot be associated with eQTLs, indicating a need for more extensive eQTL discovery across additional contexts and at increased sample sizes [7, 8]. In this context, there is an urgent need for scalable eQTL mapping methods that make optimal methodological choices to maximise the power while controlling false discoveries.

Today, while many methods exist for eQTL mapping, there are substantial differences in their methodology and little empirical comparison of these methodological choices has been performed. A significant difference between methods are the different statistical models of gene expression they use. Earlier methods, such as FastQTL and tensorQTL, modelled transformed expression values using the normal distribution [9, 10]. More recently, methods have modelled the counts produced by RNA-seq with count distribution models, particularly the Poisson and negative binomial distributions, often demonstrating more power compared to methods that use a normal distribution [11, 12]. Use of the count distribution has allowed some methods, particularly SAIGE-QTL, to model count expression at single-cell resolution. Here, however we focus on modelling bulk or pseudobulk data. Furthermore, despite the success of count distribution methods, eQTL methods do not incorporate methodological improvements developed for fitting count models to RNA-seq data for differential expression testing and it is unclear whether these improvements could enhance eQTL mapping performance [13, 14].

Another difference between methods is whether they use mixed models, which include random effects modelling relatedness between individuals. Methods implement both linear mixed models (LMM) (apex, limix) and Poisson generalised linear mixed model (SAIGE-QTL) which incorporate a random effect with covariance matrix set to an estimated genetic relatedness matrix [15, 16]. This approach has been used extensively in the GWAS literature, for example by SAIGE and BOLT-LMM, and is designed to account for cryptic relatedness between samples [17, 18]. However, given the smaller size of eQTL datasets it is unclear whether using mixed models to model relatedness is needed. In addition, it is unclear how different count distribution perform when used as part of a mixed model. Recently, the Poisson generalised linear mixed model (GLMM) has been found to give the equivalent results to a negative binomial GLMM, while while being much faster, when used for single-cell eQTL mapping [2]. However, this finding contradicts extensive work on the importance of accounting for overdispersion using the negative binomial when modelling RNA-seq data [13, 14].

Finally, an ongoing challenge for QTL mapping methods is scalability, especially as datasets increase in size. A bottleneck in cis-eQTL mapping has been the approach used to calculate the significance of genes, where computationally intensive permutations are used to evaluate the significance of the minimum per-gene p-value, followed by gene-level false discovery rate correction using q-values [9, 19]. More recently, the apex and SAIGE-QTL methods have instead used the aggregated Cauchy associated test (ACAT) to summarise significance by aggregating the variant-level p-values for each gene [20]. This approach does not require permutations, massively speeding up computation. However, how ACAT interacts with the use of q-values, which have previously been used to control false discovery rates in cis-eQTL mapping, has not been fully described [9, 19].

Here, we introduce quasar, a software package for bulk and pseudobulk eQTL mapping. Compared to existing methods quasar has several advantages. First, it implements a wide variety of methods:

linear model, Poisson GLM, negative binomial GLM, linear mixed model, Poisson GLMM and negative binomial GLMM. This wide range of models allows analysts to choose the most appropriate model for their data and allows us, in this paper, to systematically compare different models without the other, potentially confounding inter-package differences. Second, quasar is highly performant and scalable to large datasets, using the ACAT approach to compute gene-level significance, and implements a novel, faster variance approximation to score test variance in mixed models [20]. Finally, quasar implements the Cox-Reid adjusted profile likelihood for dispersion estimation in the negative binomial model [14, 21]. We compare the performance of quasar to three existing methods for eQTL mapping - tensorQTL, jaxQTL and apex - on the OneK1K dataset. Exploiting quasar's extensive range of implemented models we also compare different approaches for eQTL mapping and study the interactions between methodological choices.

Results

Overview of quasar methods

The quasar software package is a command line program, implemented in C++ and uses the Eigen library for performant linear algebra operations [22] and supports the common plink bed/bim/fam format for genotype data. The package is freely available from from GitHub (See Code Availability). The quasar software supports both cis- and trans-QTL mapping. Like other QTL mapping methods, quasar uses a score test framework to test for associations between variants and features. Briefly, in this framework, a statistical regression model is first fit for each feature to model and correct for both nuisance covariates, such as age, sex, genotype PCs and gene expression PCs, and, if a mixed model is used, correlation between individual's measurements. Then, a score test is performed, using the residualised feature values, for each variant and each feature. In cis-eQTL mapping, the overall significance of genes is commonly assessed. In quasar, the ACAT method is used to combine variant-level p-values to determine significance at the gene level. (See Methods and Supplementary Methods for details)

In addition to the range of models it implements, quasar improves on eQTL mapping methodology in two key ways. First, a key challenge in fitting mixed models using score tests in both QTL and GWAS applications is approximating the variance of the score test. Naively computing the score test variance for each variant requires a $O(n^2)$ matrix multiplication, where n is the number of individuals, which is too computationally intensive to perform for each variant when n is large. Therefore, methods, including SAIGE and SAIGE-QTL, use a variance-ratio-based approximation, which was first developed in GRAMMAR-Gamma (See Methods for details) [11, 18, 23]. Here, we derive and implement in quasar, a simple analytic trace-based approximation, which we also show is closely related to the residual-variance-based variance approximation used in the regenie software package for GWAS analysis [24]. Second, while recent methods have used the negative binomial model for eQTL mapping, it is not well understood whether the improved estimation methods developed for negative binomial models in the context of RNA-seq differential expression analysis may be needed [12]. We implement in quasar the Cox-Reid adjusted profile likelihood (APL), which allows unbiased estimation of the dispersion parameter in the negative binominal model. [14, 21]. The use of the APL is a core feature of the statistical methodology of the DESeq2 and edgeR packages for differential expression analysis in RNA-seq data [14, 25]. We explore whether this methodology can improve Type 1 error control in the context of eQTL mapping.

Method validation

To assess the performance of quasar, we applied it to the recent large OneK1K dataset with single-cell RNA-seq of just under 1000 individuals [3]. Briefly, we performed standard filtering of variants and pseudobulk aggregation following recently described best practices [26]. (See Methods for details). We applied quasar's range of models to the OneK1K dataset and, in addition, applied three existing bulk/pseudobulk eQTL mapping methods: the LMM model in apex, the linear model from tensorQTL, and the negative binomial model from jaxQTL [9, 12, 16]. Both the LMM in apex and the negative binomial model in jaxQTL were shown to be the most effective model implemented in their respective packages [12, 16]. To reduce the computational overhead and environmental impact of our comparison we focus our analysis on three cell-types (Plasma, B IN - immature and naive B, and CD4 NC - CD4 naive and central memory T cells) that have previously been used as representative cell types for analysis of methods on the OneK1K dataset [11]. The Plasma and CD4 NC are the smallest and largest cell types by total number of cells, while the B IN cell type is of average size.

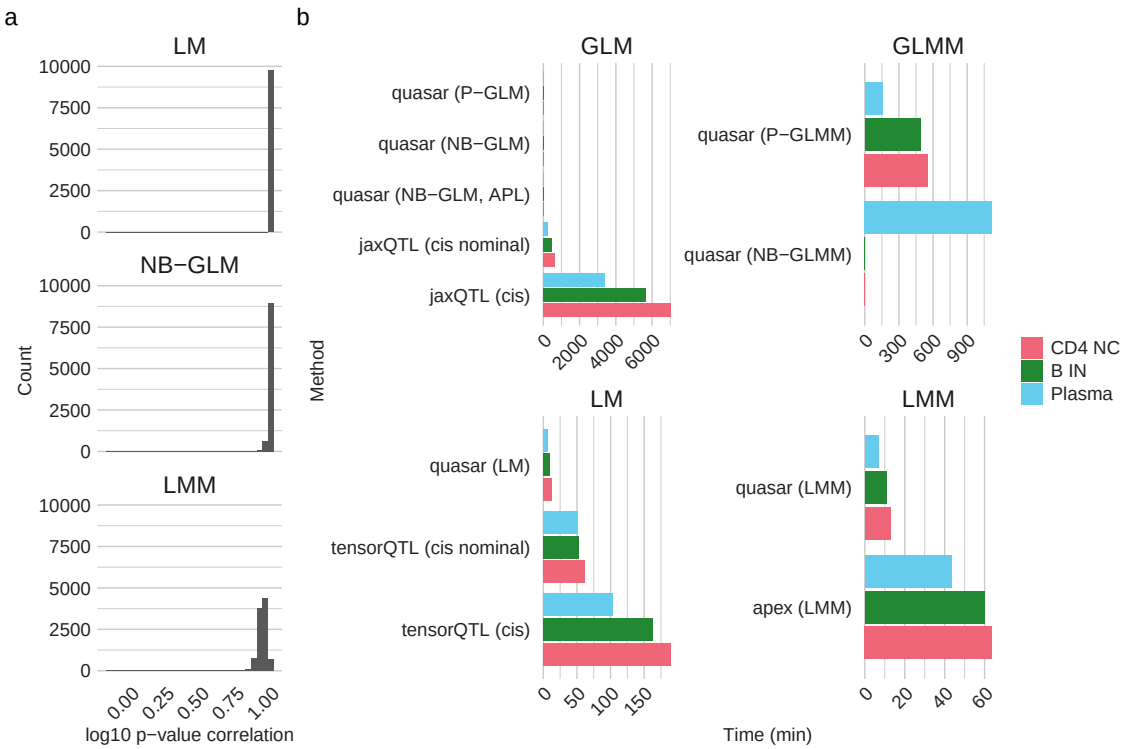


Fig 1: **Method concordance and computational performance** a) Histograms of Pearson correlation of $-\log_{10}$ transformed variant-level p-values for each gene, correlating the output of output of quasar against that uses the same statistical model (LM: tensorQTL, NB-GLM : jaxQTL, LMM: apex. All results are computed for the B IN cluster. b) Speed of methods across the three representative cell types. All methods were run on CPUs. Methods are labelled by the options used to run them: for tensorQTL and jaxQTL 'cis' computes significance at the level of genes while 'cis nominal' computes significance at the level of variants.

First, we validated quasar by comparing its output to that of existing methods where the statistical model aligns. In particular we compared quasar's LM method to tensorQTL, the NB-GLM method to jaxQTL and the LMM method to apex. To assess this concordance we computed the Pearson correlation between the $-\log_{10}$ variant-level p-values computed by quasar and by the other method for each gene (Figure 1a). In this analysis all or almost all genes displayed correlations of 1 for the LM and NB-GLM models, indicating very high concordance between the results for quasar and other methods. For the LMM model, we observed generally very high but not perfect correlations. This is likely due to the fact that apex uses a different method of approximating the mixed model score test variance. Overall, this analysis demonstrates that quasar's output is concordant with that of other widely used eQTL mapping methods.

In addition, while running the negative binomial GLMM we observed that the estimated values of the dispersion parameter (denoted by ϕ) are very close to zero (Figure S1). This observation is consistent with the previous use of heuristic algorithms to better estimate ϕ in negative binomial GLMM models and reports that the Poisson GLMM model is sufficient for QTL mapping, as when $\phi \approx 0$ the negative binomial GLMM model reduces to the Poisson GLMM model. Theoretically, if the variance-covariance matrix of the random effect is the identity matrix, which would occur if all individuals are completely unrelated, then it can be shown that the Poisson GLMM induces the same form quadratic mean-variance relationship as a negative binomial model without random effects (see Supplementary Methods for details). This quadratic mean-variance relationship is one of the main factors motivating the use of negative binomial distribution in RNA-seq data. We also show that in the negative binomial GLMM model, if the random effects are uncorrelated the dispersion parameter and variance of the random effect are not identifiable, and therefore the standard estimation algorithm will therefore lead to the variance of the random effect accounting for all overdispersion (See Supplementary Methods for details). As individuals in the OneK1K dataset do not show strong relatedness this effect explains the small ϕ values. Therefore, due to its equivalence with the Poisson GLMM when ϕ is small we assess the speed of the negative binomial GLMM model but do not otherwise assess it in this paper.

Computational performance

We next assessed the speed of quasar's models and compared it to other methods for eQTL mapping. The speed of a method has a large impact on it's usability, especially as datasets continue to grow in size. We measured the speed of the methods on the three representative cell types and, for consistency, ran all methods using CPUs, which are generally less expensive to use and more available than GPUs in research and commercial high performance computing platforms. The quasar package is designed to compute both variant and gene level p-values in one run, as is apex, while the tensorQTL and jaxQTL methods compute gene and variant-level p-values in separate commands. Here, tensorQTL and jaxQTL methods use permutations to compute gene-level p-values while apex and quasar use the ACAT method. Due to its poor computational performance and equivalence of the Poisson GLMM method, we only ran the Negative Binomial method on the small Plasma cell type.

First, we compared the performance of quasar to the comparison eQTL mapping methods, where the statistical model aligns (Figure 1b) When considering the time taken to compute both variant and gene level p-values the quasar models were substantially faster across cell-types when compared to the same model implemented in the comparison eQTL mapping packages (linear model (tensorQTL): 21-23x, negative binomial GLM (jaxQTL): 484-542x, linear mixed model (apex): 5-6x). In addition, quasar was also faster, though to a smaller extent, when comparing it's speed to the time taken to

compute variant p-values only (linear model: 5-8x, negative binomial GLM: 37-45x) This finding demonstrates that while a major determinant of quasar's efficiency is its use of the ACAT test, its variant-level p-value calculation is also more efficient. Second, we compared the performance of the different models implements in quasar. The fastest models were the Poisson GLM model and the linear model, followed by the linear mixed model and the negative binomial GLM. Using the adjusted profile likelihood in the negative binomial GLM was slower than MLE estimation, at 1.4-1.6 slower across cell types. The slowest quasar methods were the GLMM model methods, with the Poisson GLMM being 25-50x slower than the Poisson GLM. This reduction is performance is caused by the complexity of the PQL algorithm implemented in quasar for fitting GLMMs, which involves inversion of the relatedness matrix an $O(n^3)$ operation, where n is the number of individuals, at each iteration. In the Plasma cluster the negative binomial GLMM was approximately 7 times slower than the Poisson GLMM, demonstrating the increased computational burden of fitting the negative binominal distribution. Finally, as expected all models and methods were slower as the number of sample available for the cell type increased, with the mixed model methods showing a substantially larger increase compared to other methods. Overall, these results demonstrate that quasar is more efficient than other eQTL mapped methods, and illustrated the large impact of both statistical model and gene-level significance aggregation strategy on computational performance.

Type 1 error rate control

We next assessed the ability of methods to control the Type 1 error rate, that is to not produce false discoveries. To assess this aspect of method performance, we used a permutation-based approach. Using permutations allow us to generate null datasets that maintain properties of the data such as the distribution of counts in the single-cell data and correlation between SNPs, which are challenging to accurately simulate [27]. Specifically, we filtered the data to variants on chromosome 22 (chosen as the chromosome with the fewest variants) and permuted the genotypes 10 times with respect to the gene expression data (See Method for details). All quasar models, except the negative binomial GLMM, and the three comparison methods were then run on the three representative cell types.

First, we assessed the ability of methods to produce well calibrated p-values when calculating p-values for variants (Figure 2). Across quasar models, the LM and LMM controlled displayed a uniform distribution of p-values. The NB-GLM model showed p-values that were slightly smaller than expected and a larger number of very small p-values in the B IN cell type. These issue were reduced when the adjusted profile likelihood was used to estimate the negative binomial dispersion. This improved calibration is due to the less biased estimation of the dispersion when using the adjusted profile likelihood, leading to generally larger values of the dispersion, increasing the estimated variance (Figure S2). The quasar negative binomial methods were slightly conservative in the small Plasma cell type. The Poisson GLMM model displayed uniform p-values for the B IN and CD4 NC cell types but had a large excess of small p-values for the small Plasma cluster. This effect is in line with previous reports of the Poisson mixed model failing to control the Type 1 error when the event rate is low ([28]). Across all cell types the Poisson GLM produced a very poorly calibrated p-values, demonstrating that it could not robustly control Type 1 error. This observation is consistent with previous work and is likely caused by the Poisson distribution's inability to model overdispersion, unlike the negative binomial model [12]. Due to its failure to control the Type 1 error rate we do not use the model in other comparisons and running quasar specifying the Poisson GLM generates warnings. Considering the three comparison methods, tensorQTL produced uniform p-values across all cell types while jaxQTL showed slightly smaller p-values than excepted, similar to the quasar NB-GLM model. In our hands, apex produced overly small p-values across all cell-types, possibly

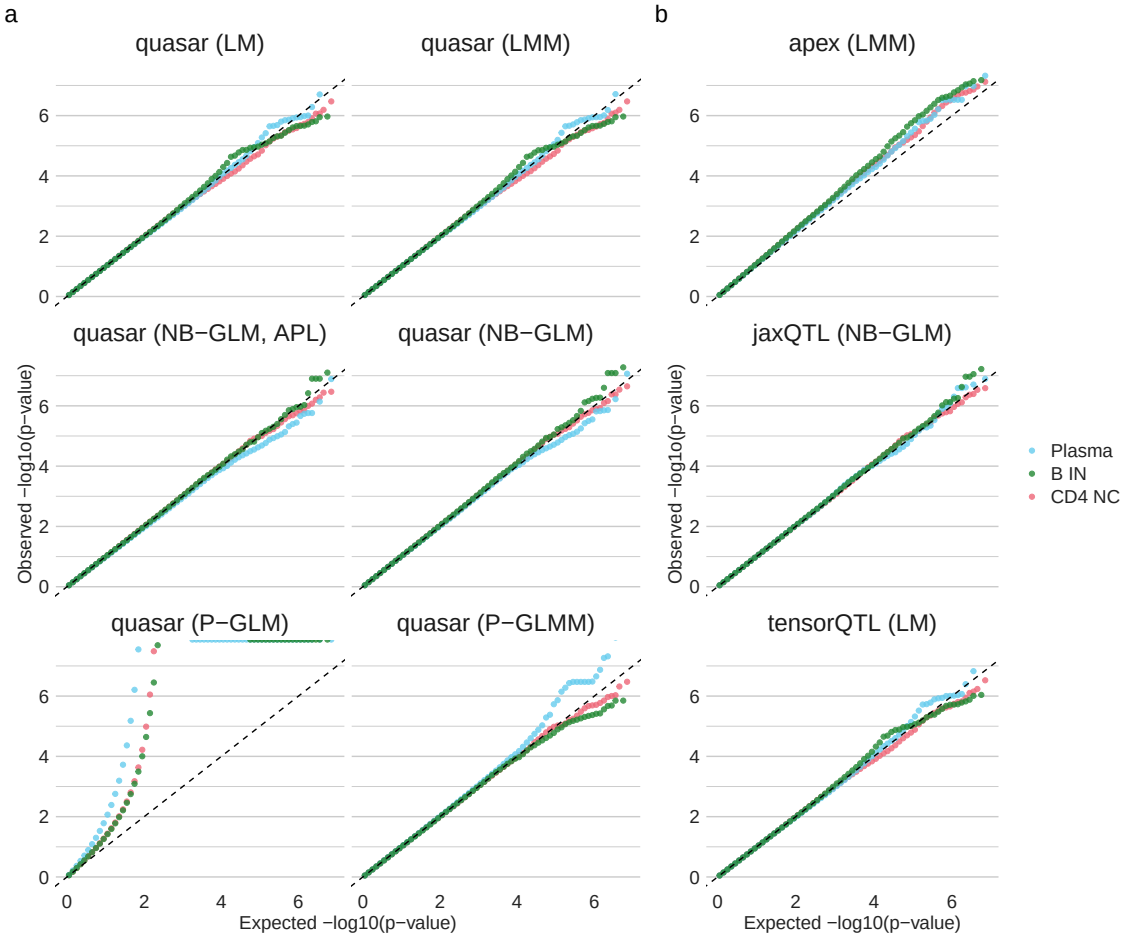


Fig 2: **Distribution of variant-level p-values under a permutation null** a) Results across the statistical models implements in quasar, showing results for the three representative cell types, b) Results for the comparison eQTL mapping methods, also across the three representative cell-types

due to issues with processing of the genetic relatedness matrix. (See Method for details) Next, to more formally assess the Type 1 error control of the methods, we computed the proportion of variants declared significant at different significance thresholds (Figure S3a). Overall, these results were concordant with the p-value calibration results, with the Poisson GLMM failing to control Type 1 errors in the Plasma cell type and the adjusted profile likelihood providing better control of the Type 1 error compared to jaxQTL and quasar's negative binomial GLM method.

Second, we assessed the ability of methods to produce well calibrated p-values at the level of genes (Figure 3). Across quasar methods, which all use the ACAT method to compute per-gene p-values, the results at the gene-level were similar to those at the variant level. In particular, the use of the APL also decreased the number of small p-values relative to the expectation. These results for quasar demonstrate the applicability of the ACAT method across statistical models. For the comparison methods, jaxQTL and tensorQTL showed well calibrated distributions while apex had an excess of

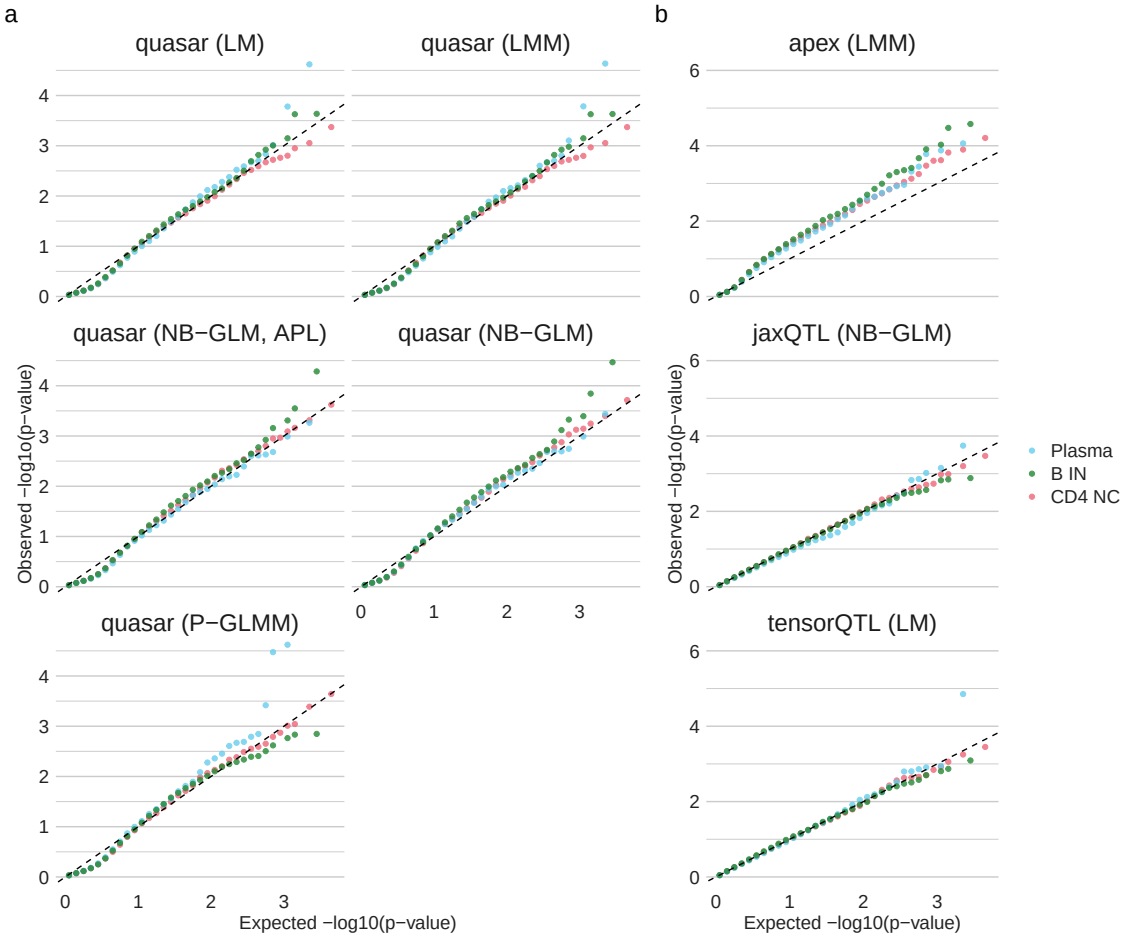


Fig 3: **Distribution of gene-level p-values under a permutation null** a) Results across the statistical models implements in quasar, showing results for the three representative cell types, b) Results for the comparison eQTL mapping methods

small p-values. The poor performance of apex at the gene-level shows that the ACAT method's performance is dependent on the calibration of variant-level p-values. In addition, we computed the proportion of variants declared significant at different significance thresholds (Figure S3b). This analysis highlighted similar results to the p-value calibration analysis but while the adjusted profile likelihood improved performance, the negative binomial GLM still displayed a slightly inflated Type 1 error rate.

While performing these analyses, we noticed that the distribution of ACAT p-values under the permutation null was substantially non-uniform, instead displaying a bi-model distribution (Figure S4). This distribution is different to that of permutation-computed p-values which do show the expected uniform distribution (Figure S4b) We confirmed that this pattern is not due to issues in quasar, as it occurs when ACAT is computed from the output of tensorQTL (as in the figure) and it can also be observed in log-transformed Q-Q plots in other methods that use the ACAT, as a small bump

below the line when the p-values are close to one. The bimodal distribution under the null does not satisfy the assumptions of the q-value estimation procedure, which has previously been used to assess the significance of gene-level p-values computed using permutations [29]. We further investigated the effect of this misspecification when comparing the power of methods. Overall, however, we found that the ACAT method displayed uniform p-value under a permutation null in the left tail, demonstrating that the ACAT distribution is well-calibrated under the null for the small p-values of interest in eQTL mapping analysis.

Power analysis

Finally, we assessed the statistical power of all methods. We assessed power at the level of both variants - how many eQTLs a method identifies, and genes - how many eGenes a method identifies. To quantify the the power of methods to detect eQTLs, we computed the number of variants that are significant at a $p < 5 \times 10^{-6}$ threshold (Figure 4a). In the three representative cell types, we observed that count distribution methods (quasar NB-GLM with and without the APL, Poisson GLMM, and jaxQTL) displayed the greatest power to detect eQTLs, as previously described [12]. In the the Plasma cell type the Poisson GLMM model mapped more eQTLs than other methods, consistent with it's poor Type 1 error in that cell type. Also in the Plasma cell type only, the jaxQTL method had more power than the NB-GLM quasar methods. The next most powerful type of methods were linear model methods, with LMM-based methods surprisingly showing the lowest power. For the the LM and LMM models implemented in quasar, the number of significant variants was highly concordant with the tensorQTL and apex methods, respectively.

Next, we quantified the power of methods to detect eGenes, computing the number of genes which were significant at a 5% FDR threshold (Figure 4b). The power of methods at the gene level is affected by both the model and gene-level aggregation procedure and the ACAT test has been previously been shown to have greater power than the permutation approach [16]. To assess the impact of ACAT in our data and study the use of q-values with ACAT, where assumptions of its estimation procedure are not met, we applied all combinations of procedures to the output of tensorQTL (Figure S5a). This analysis had three main findings. First, as previously reported, the ACAT method generally detected more eGenes than the permutation-computed p-values [16]. Second, applying the Benjamini-Hochberg (BH) method and the q-value procedure to the ACAT p-values produced the same set of eGenes. Third, when the BH and q-value procedure were applied to the permutation p-values, the q-value procedure detected more eGenes and the increase was larger for the larger cell types. Reassuringly, the overall overlap between all sets was very high (Figure S5b). Based on these findings, we recommend that the BH procedure be used with ACAT p-values, as the technical assumptions of the q-value estimation procedure are not met and the methods produce equivalent results. However, for the permutation approaches our results validate the use of the q-value method as increasing power. Therefore, in our overall comparison of methods we use the BH procedure for methods that compute ACAT p-values and the q-values procedure for methods that compute permutation p-values.

Overall, the gene-level results were concordant with the variant-level results with count-distribution methods again displaying higher power. This increase in power decreased as the sample size of the cell type increased. The apex method displayed a much higher power than in the variant-level analysis but this effect is due to it's poor calibration under the null, highlighting the ACAT method's tendency to produced false discovery if the variant-level p-values are not well calibrated. As the sample size of the cell type increased the methods using q-value corrected permutation p-values - jaxQTL and tensorQTL - had higher power. Indeed, while they had lower power in the

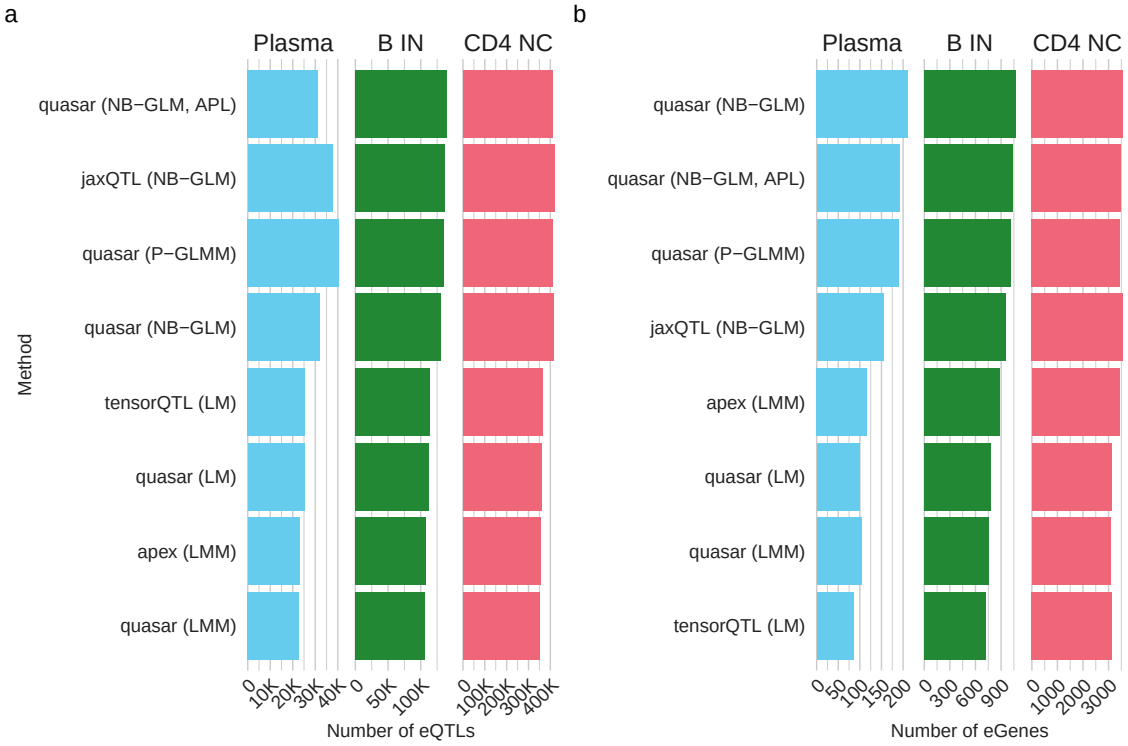


Fig 4: **Statistical power of all methods in the three representative cell types** a) The number of eQTLs, at a $p < 5 \times 10^{-6}$ threshold, found across the three representative cell-types by quasar, with different models, and the comparison eQTL mapping methods. b) The number of eGenes, at a 5% FDR threshold, found across the three representative cell-types by quasar, with different models, and the comparison eQTL mapping methods.

Plasma and B IN cell types, in the CD4 NC cell type (the largest in the OneK1K dataset) both methods had comparable or slightly higher power than the equivalent quasar method.

To confirm that the quasar models were finding a concordant set of eGenes, we visualised the overlap between the eGenes selected by different models on the B IN cell type (Figure S6a). This analysis showed that, as expected, most eGenes were found by all models with a subset being found only by count-distribution methods. The eGenes only identified by the count distribution methods had lower total counts compared to eGenes identified by all methods ($p < 9.5 \times 10^{-4}$, two sided Wilcoxon rank sum test) (Figure S6b). This observation, which is in line with previous literature, suggests that the count-distributions method's higher power is caused by better performance in genes with lower expression.

Finally, we sought to expand our analysis of the finding that count-distribution methods have more power for eQTL mapping. To enhance our analysis we additionally ran the quasar linear model and quasar negative binomial GLM with the adjusted profile likelihood in the 11 other cell types in the OneK1K dataset (Figure 5). As in the previous analysis we computed the number of eQTLs and eGenes found by each method. The negative binominal GLM had more power than the linear

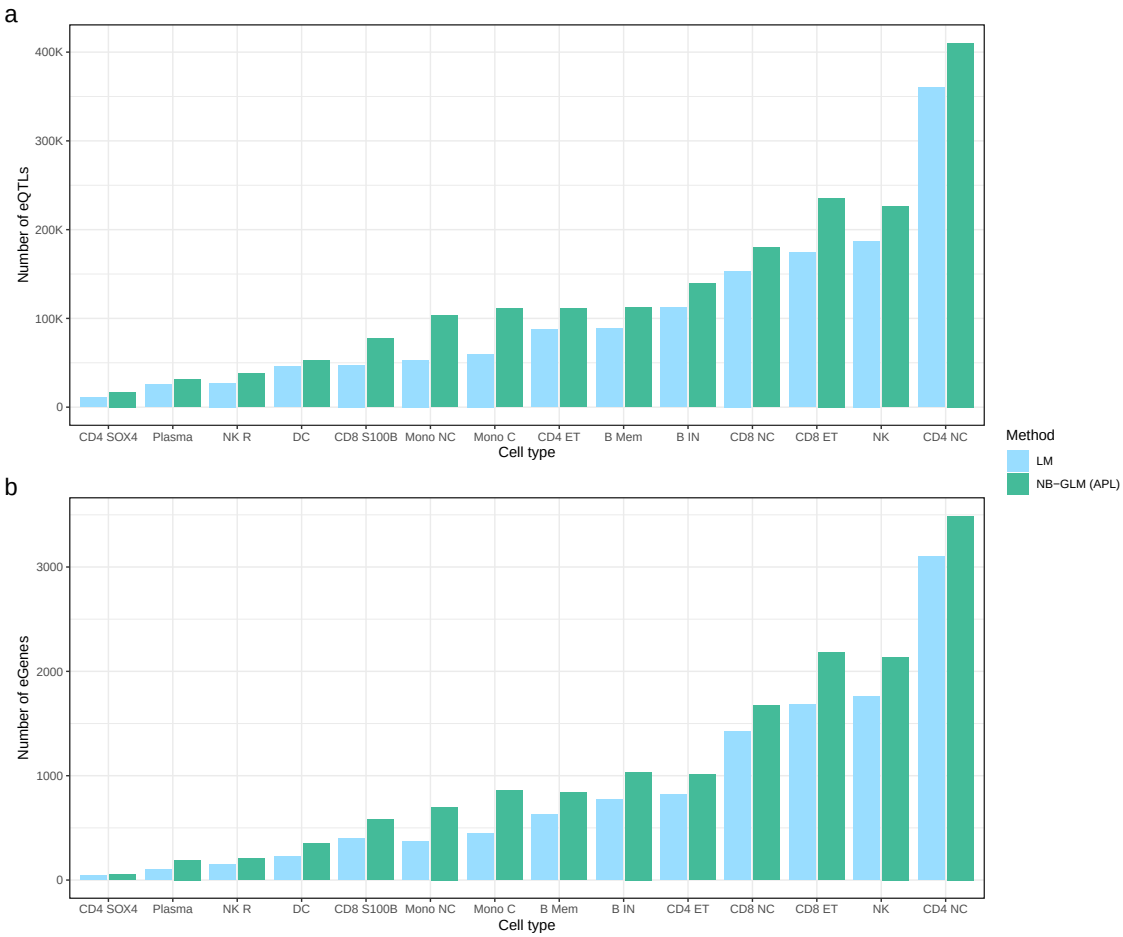


Fig 5: **Statistical power of negative binomial and linear model methods across the OneK1K dataset** a) Number of eQTLs detected by the quasar linear model and negative binomial GLM with adjusted profile likelihood dispersion estimation methods across all cell types in the OneK1K dataset. b) Number of eGenes detected by the quasar linear model and negative binomial GLM with adjusted profile likelihood dispersion estimation methods across all cell types in the OneK1K dataset.

model across all cell types. Overall, across cell types the negative binomial GLM found 29% more eQTLs and 29% more eGenes. The two monocyte cell types showed the largest gain in power for the negative binomial GLM method.

Discussion

In this paper we introduced quasar, a flexible and efficient C++ package for performing eQTL mapping. Compared to existing eQTL mapping methods, quasar implements a wider variety of statistical models, including both count distribution and mixed models. Here, we also introduce and implement

in quasar an analytic approximation to the score-test variance in mixed models. Furthermore, we highlighted that difficulties in estimating the negative binomial dispersion, identified in the context of RNA-seq differential expression testing, also apply to eQTL mapping. Therefore, quasar implements the Cox-Reid adjusted profile likelihood, which enables unbiased estimation of the dispersion. We demonstrated quasar in a pseudobulk analysis of the OneK1K dataset and compared it to three existing eQTL mapping tools: tensorQTL, jaxQTL and apex.

The quasar software package has several advantages relative to existing eQTL mapping software. Compared to tensorQTL, quasar's linear model implementation gives highly similar results and is around 6x faster for computing variant-level p-values on CPUs and replaces permutations with the ACAT method. Compared to jaxQTL, quasar's negative binomial GLM gives highly similar results, is around 40x faster on CPUs for computing variant-level p-values and additionally implements the adjusted profile likelihood for estimating the negative binomial dispersion. Finally, compared to apex's linear mixed model, quasar's linear mixed model implementation is faster and provides robust control of Type 1 error.

We exploited quasar's flexibility to compare different statistical models for eQTL mapping without confounding due to other method implementation details. In this analysis, and consistent with previous reports, we observed that count-distribution approaches have more power, with around 30% more eGenes found across cell-types. We observed that negative binomial GLM could fail to produce well-calibrated p-values and control the Type 1 error rate at the variant level but that this effect could be fixed by using the adjusted profile likelihood to estimation the dispersion parameter. Furthermore, in the datasets we considered mixed models did not display better performance but were substantially slower. Overall, we recommend the use of the negative binomial model with the adjusted profile likelihood. The linear model can also be used as it is slightly faster then NB-GLM at the cost of reduced power. Mixed model approaches should be considered if there is known to be substantial relatedness between samples.

Our conclusions about the utility of mixed models are specific to current datasets. It is possible that in larger datasets, where cryptic relatedness is more likely, or in datasets that are known to contain related individuals mixed models could have better performance and we encourage users of quasar to exploit its flexibility and try mixed models in these situations. However, our conclusions should hold in all datasets with low relatedness and less than 1000 individuals, which accounts for a large proportion of all current eQTL datasets. We also note that analysis of Phase 1 of the TenK10K project, including approximately 2000 individuals, did not use the ability of SAIGE-QTL to model sample relatedness [5]. Mixed models became standard in GWAS analysis as sample sizes increased to biobank scale so the performance of mixed models will likely need to be revisited as larger eQTL datasets, and datasets collected from established biobanks, become available. Furthermore, our findings about the use of mixed models for relatedness are not relevant to the use of mixed models to model the repeated measurement structure of single cells nested within individuals.

Our findings have implications for the broader landscape of QTL mapping methods. First, we observed empirically and derived theoretically that in situations with low relatedness Poisson GLMMs can model a quadratic mean-variance relationship. This finding justifies the lack of negative binomial model in SAIGE-QTL and suggests negative binomial mixed models may be broadly superfluous except in situations with known strong sample relatedness. Second, our use of ACAT across all methods in quasar highlights its wide applicability and we showed that using ACAT increased power. This increase may reflect ACAT's use of multiple SNPs, while the permutation approach considers only the minimum p-value per gene. Overall, we recommend that ACAT can be widely applied in place of permutations in eQTL mapping. However, we note that ACAT does not have

a uniform distribution under the null, breaking the assumptions of the q-value fitting procedure, so we instead recommend Benjamini-Hochberg FDR correction be used to perform multiple testing correction when the ACAT is used. Third, while in this paper we focused on eQTL mapping, quasar is likely to be effective for mapping other molecular QTLs. For example, chromatin accessibility QTLs could be modelled using the negative binomial model as ATAC-seq data is known to follow a count distribution. Furthermore, quasar could be used to map other types of QTLs such as cell-state or cell-state abundance QTLs [30, 31].

Finally, quasar has some limitations. First, quasar is not designed for larger, biobank scale, datasets (approximately $> 20,000$ individuals), where direct inversion of the genetic relatedness matrix required when fitting the mixed models is not computationally feasible. To scale to such datasets quasar could implement the preconditioned gradient descent algorithm used for matrix inversion by SAIGE and SAIGE-QTL. Second, unlike apex and SAIGE-QTL, quasar does not support sparse or low-rank GRMs. Third, unlike jaxQTL and tensorQTL, quasar does not currently have the ability to accelerate performance with GPUs if they are available. Integration of OpenCL would allow quasar to also use GPUs to accelerate its performance. Finally, quasar is designed to work on bulk or pseudobulk single-cell data, unlike SAIGE-QTL which models single-cell level counts directly. Extending quasar to efficiently map eQTLs at single-cell resolution is a promising direction for future research. An additional limitation of our comparison of methods was that we were unable to run jaxQTL and tensorQTL with GPU due to computational constraints.

Methods

Overview of quasar

The quasar method uses the following overall algorithm. (More details about the overall algorithm and all aspect of the quasar's methodology are available in the Supplementary Methods)

1. For each feature fit a 'null' model where the feature observations are modelled as a function of nuisance covariates, such as age, sex or genotype PCs, supplied by the user.
2. The estimated model parameters are used to residualise the feature observations, removing the effect of the nuisance covariates.
3. All genetic variants tested for association have the nuisance covariates projected out. That is, we regress the genotype vector on the nuisance covariates and subtract the predicted values.
4. For each genetic variant-feature combination, a score test is performed using the processed feature and genetics variants described above. When a mixed model is used in step 1 a novel trace-based approximation for the score test variance is used.

Statistical models

The quasar software implements the following statistical models for residualising the phenotype:

- Linear model,
- Poisson generalised linear model,
- Negative binomial generalised linear model,
- Linear mixed model,

- Poisson generalised liner mixed model and
- Negative Binomial generalised linear mixed model.

To fit the linear model we use the standard analytic MLE. To fit the Poisson GLM we use the standard iteratively least squares (IRLS) algorithm, the same algorithm used by the R function `stats::glm()`. To fit the NB GLM we use the standard algorithm, as implemented in the R function `MASS::glm.nb()`, that alternates between estimating the dispersion parameter ϕ and fitting a NB GLM where ϕ is known. To fit the LMM we use the FaST-LMM algorithm [32], with the C++ implementation based on that of apex. To fit the Poisson GLMM we use the penalized quasi-likelihood algorithm (PQL), which has been previously applied to fit logistic GLMM for GWAS SAIGE and for single-cell level QTL mapping in SAIGE-QTL. The algorithm has also been applied to estimate heritability for count distributed responses with PQLSeq/PQLSeq2 and we use methodological elements of the PQLSeq algorithm in quasar [33] (See Supplementary methods for details). To fit the NB GLMM we use an analogous algorithm to that used for the NB GLM, replacing the Poisson GLM and Negative Binominal GLM with ϕ fixed with their mixed model equivalents. This is the same algorithm used in the R function `glmer.nb()`, from the widely used `lme4` package for mixed models, which has previously been used to map single-cell eQTL. [2, 34]

Score test variance approximation

When mixed models are used to residualise the feature outcome the variance of the score statistic has the form,

$$\boldsymbol{g}^T \boldsymbol{P} \boldsymbol{g},$$

where $\boldsymbol{g}$ is the n-vector of genotypes after covariate residualisation and $\boldsymbol{P}$ is a dense $n \times n$ residualisation matrix that arises in the mixed model fitting procedure (See the Supplementary Methods for details). This quantity must be computed for each variant, but this is not feasible as computing it has computational complexity $O(n^2)$ and n, and the number of variants, may be large. This computational problem has been addressed by approximating the score test variance with approximations of the form

$$\boldsymbol{g}^T \boldsymbol{P} \boldsymbol{g} \approx \alpha \boldsymbol{g}^T \boldsymbol{g}$$

where the constant α is chosen to make the approximation as accurate as possible. (Here we consider the approximation in the LMM case, for the GLMM case see the Supplementary Methods) The most common approach for choosing α is the variance ratio approach, also known as the GRAMMAR-GAMMA approximation where we set

$$\alpha_{\text{vr}} = \frac{1}{m} \sum_{i=1}^{m} \frac{\boldsymbol{g}_i^T \boldsymbol{P} \boldsymbol{g}_i}{\boldsymbol{g}_i^T \boldsymbol{g}_i},$$

where m is generally taken to be small ≈ 30 and the variants, $\boldsymbol{g}_i$, are drawn at random genome-wide. [23]. This approach is now used across GWAS and eQTL methods. [11, 18] A different approach, used by the software package for GWAS analysis regenie, is to set α as

$$\alpha_{\text{reg}} = \hat{\sigma}^2, \quad \hat{\sigma}^2 = \frac{\tilde{\boldsymbol{y}}^T \tilde{\boldsymbol{y}}}{n-p},$$

where $\tilde{\boldsymbol{y}}$ is the value of the response after nuisance covariates have been regressed out, n is the number of individual and p is the number of nuisance covariates [24]. The regenie estimator is also

closely related to original GRAMMAR estimator [35]. Here, we propose to set α using the trace of $\boldsymbol{P}$, specifically, we propose

$$\alpha_q = \frac{\mathrm{tr}(\boldsymbol{P})}{n-c}.$$

This expression is an approximate expectation of the variance-ratio approximation, treating $\boldsymbol{g}$ as random. (See the Supplementary Methods for a derivation.) This new estimator can be calculated from $\boldsymbol{P}$ in $O(n)$, removing the need to perform any $O(n^2)$ complexity $\boldsymbol{Pg}$ multiplications. In the case of the GLMM, the need for a second-order correction makes the trace-based approximation also has computational complexity $O(n^2)$, see the Supplementary Methods for details.

This novel estimator also helps to clarify why $\alpha_{\mathrm{reg}} \approx \alpha_{\mathrm{vr}}$, despite appearing to estimate different quantities. For a LMM, we show that under null of the genetic variant having no effect on the feature

$$\tilde{\boldsymbol{y}} = \boldsymbol{Py} \sim N(\boldsymbol{0}, \boldsymbol{P}) \text{ so that } \mathbf{E}(\hat{\sigma}^2) = \frac{\mathbf{E}(\tilde{\boldsymbol{y}}^T\tilde{\boldsymbol{y}})}{n-p} = \frac{\mathrm{tr}(\boldsymbol{P})}{n-c}$$

so the regenie approximation is effective because it estimates the trace of $\boldsymbol{P}$. In quasar, we use the regenie approximation for the LMM case as the full $\boldsymbol{P}$ matrix is not explicitly formed by the algorithm used to fit the LMM. For the GLMM methods we use the novel approximation (described in details in the supplementary methods.)

Adjusted profile likelihood

In a frequentist setting the negative binominal dispersion parameter, ϕ, is estimated using maximum likelihood. However, the maximum likelihood estimator of ϕ is biased, systematically underestimating the dispersion [13]. This issue has been previously identified in the context of differential expression testing in RNA-seq data, where instead maximising the Cox-Reid adjusted profile likelihood has been shown to alleviate it [14, 21]. The quasar package implements the option to estimate the dispersion parameter ϕ in the NB GLM and NB GLMM models by maximising the adjusted profile likelihood, which is the penalized log-likelihood,

$$\mathrm{APL}(\phi) = l(\phi) - \frac{1}{2}\log(\det(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}))$$

where $l(\phi)$ is the usual log likelihood , det denotes the determinant, $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$ is the information matrix, where $\boldsymbol{X}$ is the matrix of covariates and $\boldsymbol{W}$ is the diagonal matrix of working weights. We maximise $\mathrm{APL}(\phi)$ using Brent's algorithm, which is chosen for convenience as it is also used to fit the LMM.

ACAT

To compute gene-level p-values quasar uses the aggregated Cauchy association test (ACAT) method [20]. The ACAT method is a metanalysis approach specifically designed to handle situations where the p-values are highly correlated, as in eQTL mapping.The ACAT was first used for this purpose in APEX and is also used by SAIGE-QTL [11, 16] Prior to the use of the ACAT, a permutation approach was used [9], however this is computationally expensive. For a vector of p-values $p_1, \ldots, p_k$ the ACAT statistic is defined as

$$T_{\mathrm{ACAT}} = \sum_{i=1}^{k} \tan\left(\left(\frac{1}{2} - p_i\right)\pi\right),$$

where $\tan(\cdot)$ is the standard tan function. If $p_i \sim U(0,1)$, i.e. under H_0, then $\tan((\frac{1}{2} - p_i)\pi)$ follows a standard Cauchy distribution. Under H_0, the distribution of T_{ACAT} is well approximated by a Cauchy distribution with location parameter 0 and scale parameter k. Therefore, the associated p-value of the T_{ACAT}, p_{ACAT} is approximately

$$p_{\text{ACAT}} \approx \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{T_{\text{ACAT}}}{k}\right).$$

OneK1K dataset processing

The variant data was processed in accordance with previous descriptions [3]. We filtered the variant data to only variants with MAF > 0.05 and passing a Hardy-Weinberg equilibrium test at $p > 10^{-6}$. Following recommendations for pseduobulk eQTL analysis [26], for methods that require count data as input we compute the sum for each individual of the raw counts, while for methods that require normalised input we normalise the counts with Scanpy and then compute the mean for each individual. We performed this process for each cell type, in each cell type filtering to the set of genes defined for that cell type in the original OneK1K publication [3]. We used genotype data from chromosomes: 1, 3, 4, 5, 6, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 22.

Computation of genetic relatedness matrices

To compute the genetic relatedness matrix for use in quasar mixed model methods we filtered the data to remove variants in LD using the command `plink2 --indep-pairwise 250 50 0.2` then computed the GRM the recommended `plink2 --make-king` command. To ensure the computed GRM was positive definite we used the R function `Matrix::nearPD()` to compute the nearest positive definite matrix to the matrix returned by plink and used that matrix as the GRM. We computed one GRM across for all chromosomes. Initially, we used the same GRM in apex, however with this estimated GRM apex produced a very large number of false positives variants with small p-values. While trying to fix this issue we observed that apex performed better when we instead used the GRM estimated with `plink --make-rel`, so we used this GRM with apex. We note that the `--make-king` is recommended over `--make-rel` in the plink documentation as it is more robust to population stratification. As the code used to run apex on real data was not released with the paper describing apex we were unable to determine which GRM apex's authors intended it to be used with.

Covariates for QTL mapping

We used sex, age, 5 expression principle components and 5 genotype PC components as covariates for all QTL mapping. The genotype PCs were computed using plink after filtering to variants using the command `plink2 --indep-pairwise 50000 200 0.05`. The expression PCs were computed using scanpy applied to the pseudobulked Scanpy-normalised gene expression data. Our use of principle components to correct the expression data follows recent recommendations for pseduobulk eQTL mapping [26].

Comparison QTL mapping methods

tensorQTL

We ran tensorQTL in both `cis` and `cis_nominal` modes with all parameters at their default values. The pseudobulked gene expression data was quantile-normalised to a normal distribution before

being input to tensorQTL.

jaxQTL

We ran jaxQTL using the `nb` model, which was found by the authors of jaxQTL to have the highest power. Following the package documentation, we ran parralelised jaxQTL by running it on chunks of 100 genes. Following the example code, we used 1000 permutations to define gene-level p-values. We changed the default window size to 1000000 (1Mb), as in apex, tensorQTL and quasar.

apex

We ran apex using the LMM mode, passing a GRM with the `--grm` flag and also running with the `--rankNormal` flag to quantile normalised the transformed expression values to a standard normal distribution. The apex program only implements the ACAT method to compute gene-level p-values. As described above, we used a GRM computed with `plink --make-rel`.

Permutation analysis

For the permutation analysis we subsetted the data to chromosome 22 and permuted the data 10 times. To permute the plink data we permuted the ids in the `.fam` file using R, while for the vcf we permuted the sample IDs using `bcftools` and the `shuf` command. While most methods take plink .bim/.bed/.fam as input the apex methods required .vcf file as input. This analysis was performed for each of the three representative cell types. We visualise the results of the permutation analysis using a quantile-quantile plot of the -log10 transformed p-values.

Power analysis

To quantify the power of the methods we computed:

- The number of eQTLs, defined as variants with p-values $< 5 \times 10^{-6}$
- The number of eGene, defined as genes with p-values with Benjamini-Hochberg FDR < 0.05 for methods with ACAT p-values and genes with q-values < 0.05 for methods with permutation p-values

We used the `p.adjust()` function with `method = "BH"` to compute the Benjamini-Hochberg FDR values and the `qvalue()` function to compute q-values.

Data availability

Single-cell RNA-seq and genotype data of OneK1K are available via Gene Expression Omnibus (GSE196830). Single-cell gene expression data we used the RNA-seq data available on Human Cell Atlas: https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1. Processed genotype data were provided by the authors of the OneK1K paper.

Code availability

The quasar package is available at: https://github.com/jeffreypullin/quasar. Documentation and instructions for installing and running quasar are available at: https://jeffreypullin.github.io/quasar/. Code used to reproduce the analysis described in this paper is available at: https://github.com/jeffreypullin/quasar-paper.

We downloaded jaxQTL from GitHub https://github.com/mancusolab/jaxqtl/, commit f4f52a. We downloaded tensorQTL from Github at https://github.com/broadinstitute/tensorqtl, commit 812040. We downloaded apex from Github at https://github.com/corbinq/apex, commit 63b605.

Funding

C.W. acknowledges funding from the Wellcome Trust (WT220788) and Medical Research Council (MC_UU_00040/01). J.P. was supported by a Gates Cambridge scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

C.W. is a part time employee of GSK and holds shares. GSK had no influence or involvement in this work.

Acknowledgements

We thank members of the Wallace group for helpful discussions, particularly Andy Bass for discussion about the assumptions of q-value estimation procedures. We thank Joseph Powell for his assistance with the OneK1K dataset.

References

[1] Alexandra C. Nica and Emmanouil T. Dermitzakis. Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620): 20120362, June 2013. doi: 10.1098/rstb.2012.0362. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0362. Publisher: Royal Society.

[2] Aparna Nathan, Samira Asgari, Kazuyoshi Ishigaki, Cristian Valencia, Tiffany Amariuta, Yang Luo, Jessica I. Beynor, Yuriy Baglaenko, Sara Suliman, Alkes L. Price, Leonid Lecca, Megan B. Murray, D. Branch Moody, and Soumya Raychaudhuri. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature*, 606(7912):120–128, June 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04713-1. URL https://www.nature.com/articles/s41586-022-04713-1. Publisher: Nature Publishing Group.

[3] Seyhan Yazar, Jose Alquicira-Hernandez, Kristof Wing, Anne Senabouth, M. Grace Gordon, Stacey Andersen, Qinyi Lu, Antonia Rowson, Thomas R. P. Taylor, Linda Clarke, Katia Maccora, Christine Chen, Anthony L. Cook, Chun Jimmie Ye, Kirsten A. Fairfax, Alex W. Hewitt, and Joseph E. Powell. Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. *Science*, 376(6589):eabf3041, April 2022. doi: 10.1126/science.abf3041. URL https://www.science.org/doi/10.1126/science.abf3041. Publisher: American Association for the Advancement of Science.

[4] Anna S. E. Cuomo, Aparna Nathan, Soumya Raychaudhuri, Daniel G. MacArthur, and Joseph E. Powell. Single-cell genomics meets human genetics. *Nature Reviews Genetics*,

24(8):535–549, August 2023. ISSN 1471-0064. doi: 10.1038/s41576-023-00599-5. URL https://www.nature.com/articles/s41576-023-00599-5. Number: 8 Publisher: Nature Publishing Group.

[5] Anna S. E. Cuomo, Eleanor Spenceley, Hope A. Tanudisastro, Blake Bowen, Albert Henry, Hao Lawrence Huang, Angli Xue, Wei Zhou, Matthew J. Welland, Arthur S. Lee, Kristof Wing, Owen Tang, Michael P. Gray, Michael Franklin, Michael Harper, Michael Silk, Katalina Bobowik, Alexander Stuckey, John Marshall, Vivian Bakiris, Caitlin Uren, Bindu Swapna Madala, Amy Miniter, Caitlin Bartie, Drew R. Neavin, Zhen Qiao, Eyal Ben-David, Ling Chen, Kyle Kai-How Farh, Stuart M. Grieve, Tung Nguyen, Jennifer Piscionere, Owen M. Siggs, Hannah Nicholas, Katrina M. de Lange, Alex H. Hewitt, Gemma A. Figtree, Daniel G. MacArthur, and Joseph E. Powell. Impact of Rare and Common Genetic Variation on Cell Type-Specific Gene Expression, March 2025. URL https://www.medrxiv.org/content/10.1101/2025.03.20.25324352v1. Pages: 2025.03.20.25324352.

[6] Claudia Giambartolomei, Damjan Vukcevic, Eric E. Schadt, Lude Franke, Aroon D. Hingorani, Chris Wallace, and Vincent Plagnol. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics*, 10(5):e1004383, May 2014. ISSN 1553-7404. doi: 10.1371/journal.pgen.1004383. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004383. Publisher: Public Library of Science.

[7] Benjamin D. Umans, Alexis Battle, and Yoav Gilad. Where Are the Disease-Associated eQTLs? *Trends in Genetics*, 37(2):109–124, February 2021. ISSN 0168-9525. doi: 10.1016/j.tig.2020.08.009. URL https://www.sciencedirect.com/science/article/pii/S0168952520302092.

[8] Hakhamanesh Mostafavi, Jeffrey P. Spence, Sahin Naqvi, and Jonathan K. Pritchard. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nature Genetics*, pages 1–10, October 2023. ISSN 1546-1718. doi: 10.1038/s41588-023-01529-1. URL https://www.nature.com/articles/s41588-023-01529-1. Publisher: Nature Publishing Group.

[9] Halit Ongen, Alfonso Buil, Andrew Anand Brown, Emmanouil T. Dermitzakis, and Olivier Delaneau. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*, 32(10):1479–1485, May 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv722. URL https://doi.org/10.1093/bioinformatics/btv722.

[10] Amaro Taylor-Weiner, François Aguet, Nicholas J. Haradhvala, Sager Gosai, Shankara Anand, Jaegil Kim, Kristin Ardlie, Eliezer M. Van Allen, and Gad Getz. Scaling computational genomics to millions of individuals with GPUs. *Genome Biology*, 20(1):228, November 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1836-7. URL https://doi.org/10.1186/s13059-019-1836-7.

[11] Wei Zhou, Anna S. E. Cuomo, Angli Xue, Masahiro Kanai, Grant Chau, Chirag Krishna, Ramnik J. Xavier, Daniel G. MacArthur, Joseph E. Powell, Mark J. Daly, and Benjamin M. Neale. Efficient and accurate mixed model association tool for single-cell eQTL analysis, May 2024. URL https://www.medrxiv.org/content/10.1101/2024.05.15.24307317v1. Pages: 2024.05.15.24307317.

[12] 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,

January 2025. URL https://www.medrxiv.org/content/10.1101/2025.01.18.25320755v1. Pages: 2025.01.18.25320755.

[13] Mark D. Robinson and Gordon K. Smyth. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, 9(2):321–332, April 2008. ISSN 1465-4644. doi: 10.1093/biostatistics/kxm030. URL https://doi.org/10.1093/biostatistics/kxm030.

[14] Davis J. McCarthy, Yunshun Chen, and Gordon K. Smyth. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10):4288–4297, May 2012. ISSN 0305-1048. doi: 10.1093/nar/gks042. URL https://doi.org/10.1093/nar/gks042.

[15] Christoph Lippert, Franceso Paolo Casale, Barbara Rakitsch, and Oliver Stegle. LIMIX: genetic analysis of multiple traits, May 2014. URL https://www.biorxiv.org/content/10.1101/003905v2. Pages: 003905 Section: New Results.

[16] Corbin Quick, Li Guan, Zilin Li, Xihao Li, Rounak Dey, Yaowu Liu, Laura Scott, and Xihong Lin. A versatile toolkit for molecular QTL mapping and meta-analysis at scale, December 2020. URL https://www.biorxiv.org/content/10.1101/2020.12.18.423490v1. Pages: 2020.12.18.423490 Section: New Results.

[17] Po-Ru Loh, George Tucker, Brendan K. Bulik-Sullivan, Bjarni J. Vilhjálmsson, Hilary K. Finucane, Rany M. Salem, Daniel I. Chasman, Paul M. Ridker, Benjamin M. Neale, Bonnie Berger, Nick Patterson, and Alkes L. Price. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*, 47(3):284–290, March 2015. ISSN 1546-1718. doi: 10.1038/ng.3190. URL https://www.nature.com/articles/ng.3190. Publisher: Nature Publishing Group.

[18] Wei Zhou, Jonas B. Nielsen, Lars G. Fritsche, Rounak Dey, Maiken E. Gabrielsen, Brooke N. Wolford, Jonathon LeFaive, Peter VandeHaar, Sarah A. Gagliano, Aliya Gifford, Lisa A. Bastarache, Wei-Qi Wei, Joshua C. Denny, Maoxuan Lin, Kristian Hveem, Hyun Min Kang, Goncalo R. Abecasis, Cristen J. Willer, and Seunggeun Lee. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, 50(9):1335–1341, September 2018. ISSN 1546-1718. doi: 10.1038/s41588-018-0184-y. URL https://www.nature.com/articles/s41588-018-0184-y. Publisher: Nature Publishing Group.

[19] Olivier Delaneau, Halit Ongen, Andrew A. Brown, Alexandre Fort, Nikolaos I. Panousis, and Emmanouil T. Dermitzakis. A complete tool set for molecular QTL discovery and analysis. *Nature Communications*, 8(1):15452, May 2017. ISSN 2041-1723. doi: 10.1038/ncomms15452. URL https://www.nature.com/articles/ncomms15452. Publisher: Nature Publishing Group.

[20] 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, January 2020. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2018.1554485. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2018.1554485.

[21] D. R. Cox and N. Reid. Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1):1–39, 1987. ISSN 0035-

9246. URL https://www.jstor.org/stable/2345476. Publisher: [Royal Statistical Society, Oxford University Press].

[22] Gaël Guennebaud, Benoît Jacob, et al. Eigen v3. http://eigen.tuxfamily.org, 2010.

[23] Gulnara R. Svishcheva, Tatiana I. Axenovich, Nadezhda M. Belonogova, Cornelia M. van Duijn, and Yurii S. Aulchenko. Rapid variance components–based method for whole-genome association analysis. *Nature Genetics*, 44(10):1166–1170, October 2012. ISSN 1546-1718. doi: 10.1038/ng.2410. URL https://www.nature.com/articles/ng.2410. Publisher: Nature Publishing Group.

[24] Joelle Mbatchou, Leland Barnard, Joshua Backman, Anthony Marcketta, Jack A. Kosmicki, Andrey Ziyatdinov, Christian Benner, Colm O'Dushlaine, Mathew Barber, Boris Boutkov, Lukas Habegger, Manuel Ferreira, Aris Baras, Jeffrey Reid, Goncalo Abecasis, Evan Maxwell, and Jonathan Marchini. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, 53(7):1097–1103, July 2021. ISSN 1546-1718. doi: 10.1038/s41588-021-00870-7. URL https://www.nature.com/articles/s41588-021-00870-7. Publisher: Nature Publishing Group.

[25] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, December 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8. URL https://doi.org/10.1186/s13059-014-0550-8.

[26] Anna S. E. Cuomo, Giordano Alvari, Christina B. Azodi, Davis J. McCarthy, Marc Jan Bonder, and single-cell eQTLGen consortium. Optimizing expression quantitative trait locus mapping workflows for single-cell studies. *Genome Biology*, 22(1):188, June 2021. ISSN 1474-760X. doi: 10.1186/s13059-021-02407-x. URL https://doi.org/10.1186/s13059-021-02407-x.

[27] Helena L. Crowell, Sarah X. Morillo Leonardo, Charlotte Soneson, and Mark D. Robinson. The shaky foundations of simulating single-cell RNA sequencing data. *Genome Biology*, 24(1):62, March 2023. ISSN 1474-760X. doi: 10.1186/s13059-023-02904-1. URL https://doi.org/10.1186/s13059-023-02904-1.

[28] Rounak Dey, Wei Zhou, Tuomo Kiiskinen, Aki Havulinna, Amanda Elliott, Juha Karjalainen, Mitja Kurki, Ashley Qin, Seunggeun Lee, Aarno Palotie, Benjamin Neale, Mark Daly, and Xihong Lin. Efficient and accurate frailty model approach for genome-wide survival association analysis in large-scale biobanks. *Nature Communications*, 13(1):5437, September 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32885-x. URL https://www.nature.com/articles/s41467-022-32885-x. Publisher: Nature Publishing Group.

[29] John D. Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9440–9445, August 2003. ISSN 0027-8424. doi: 10.1073/pnas.1530509100.

[30] Kluzer Alice, John C. Marioni, and Michael D. Morgan. Milo2.0 unlocks population genetic analyses of cell state abundance using a count-based mixed model, November 2023. URL https://www.biorxiv.org/content/10.1101/2023.11.08.566176v1. Pages: 2023.11.08.566176 Section: New Results.

[31] Laurie Rumker, Saori Sakaue, Yakir Reshef, Joyce B. Kang, Seyhan Yazar, Jose Alquicira-Hernandez, Cristian Valencia, Kaitlyn A. Lagattuta, Annelise Mah-Som, Aparna Nathan, Joseph E. Powell, Po-Ru Loh, and Soumya Raychaudhuri. Identifying genetic variants that

influence the abundance of cell states in single-cell data. *Nature Genetics*, 56(10):2068–2077, October 2024. ISSN 1546-1718. doi: 10.1038/s41588-024-01909-1. URL https://www.nature.com/articles/s41588-024-01909-1. Publisher: Nature Publishing Group.

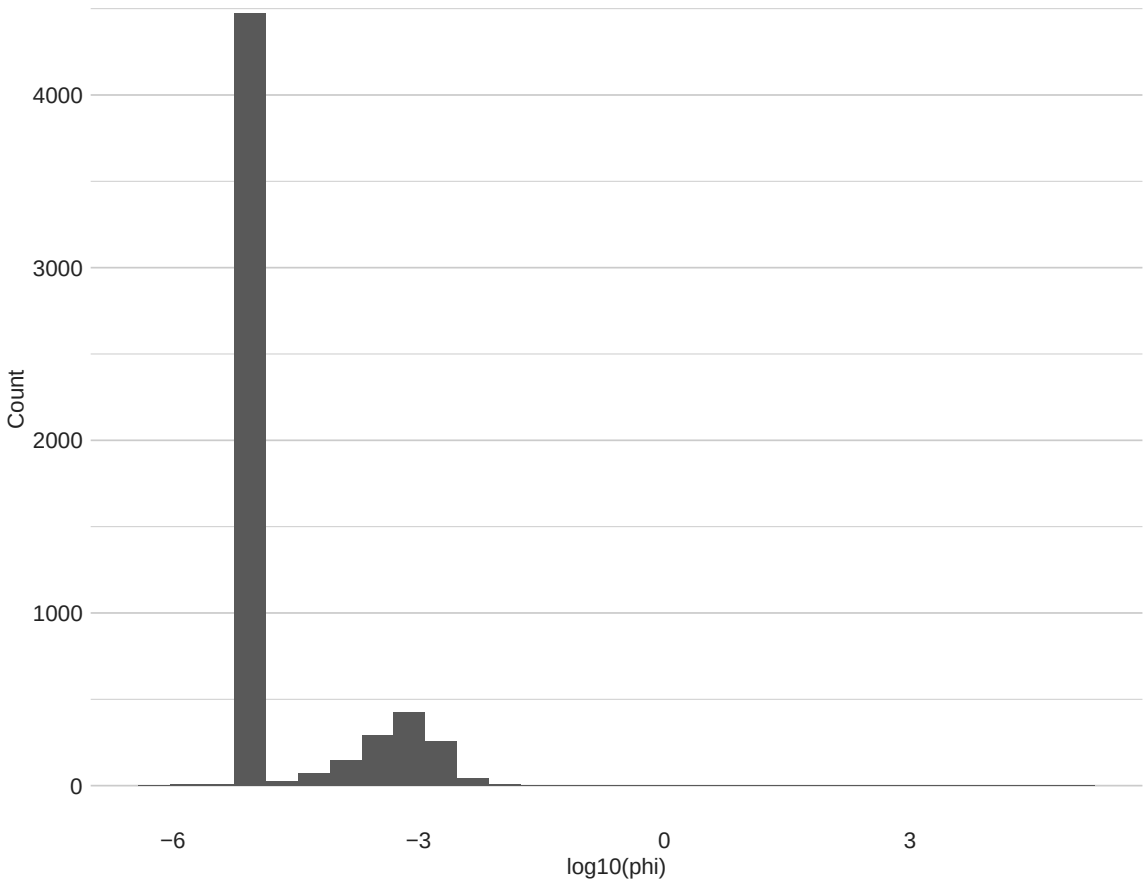
[32] Christoph Lippert, Jennifer Listgarten, Ying Liu, Carl M. Kadie, Robert I. Davidson, and David Heckerman. FaST linear mixed models for genome-wide association studies. *Nature Methods*, 8(10):833–835, October 2011. ISSN 1548-7105. doi: 10.1038/nmeth.1681. URL https://www.nature.com/articles/nmeth.1681. Publisher: Nature Publishing Group.

[33] Shiquan Sun, Jiaqiang Zhu, Sahar Mozaffari, Carole Ober, Mengjie Chen, and Xiang Zhou. Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies. *Bioinformatics*, 35(3):487–496, February 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty644. URL https://doi.org/10.1093/bioinformatics/bty644.

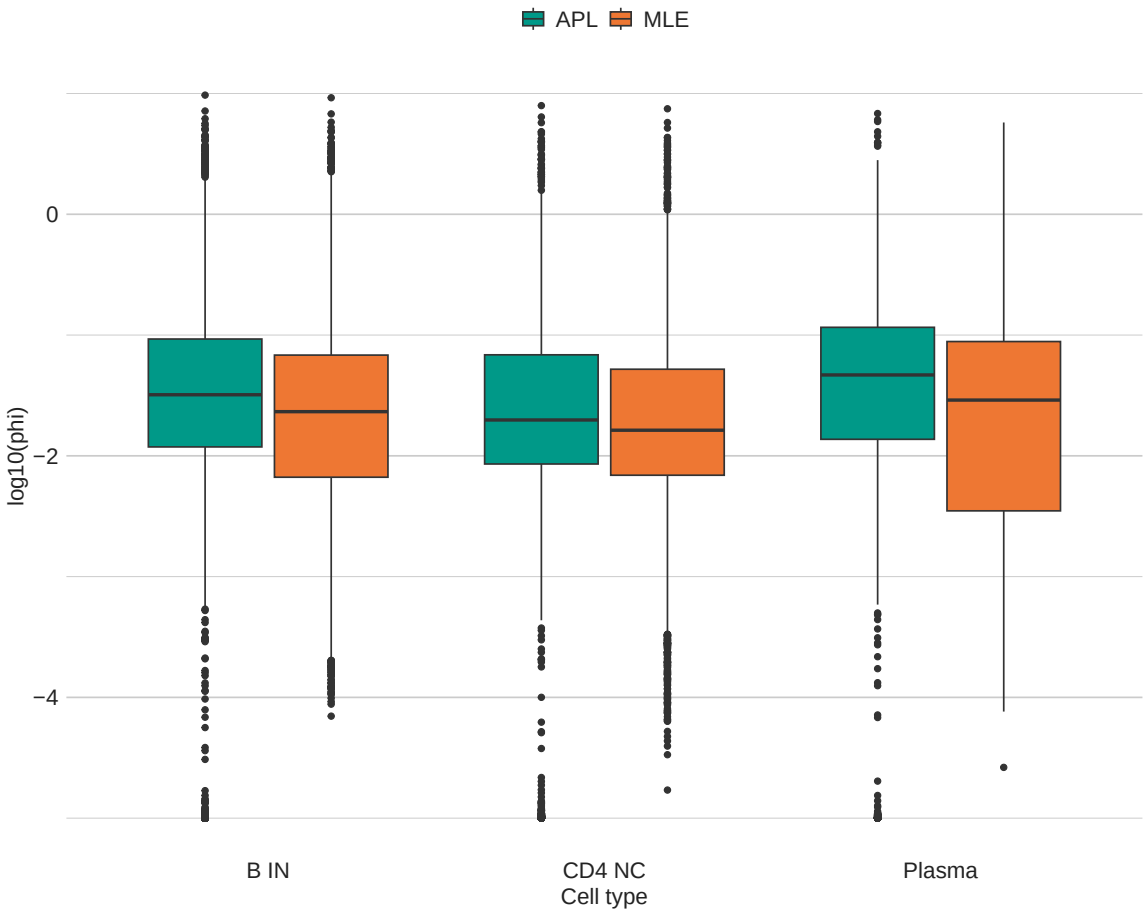
[34] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67:1–48, October 2015. ISSN 1548-7660. doi: 10.18637/jss.v067.i01. URL https://doi.org/10.18637/jss.v067.i01.

[35] Yurii S Aulchenko, Dirk-Jan de Koning, and Chris Haley. Genomewide Rapid Association Using Mixed Model and Regression: A Fast and Simple Method For Genomewide Pedigree-Based Quantitative Trait Loci Association Analysis. *Genetics*, 177(1):577–585, September 2007. ISSN 1943-2631. doi: 10.1534/genetics.107.075614. URL https://doi.org/10.1534/genetics.107.075614.

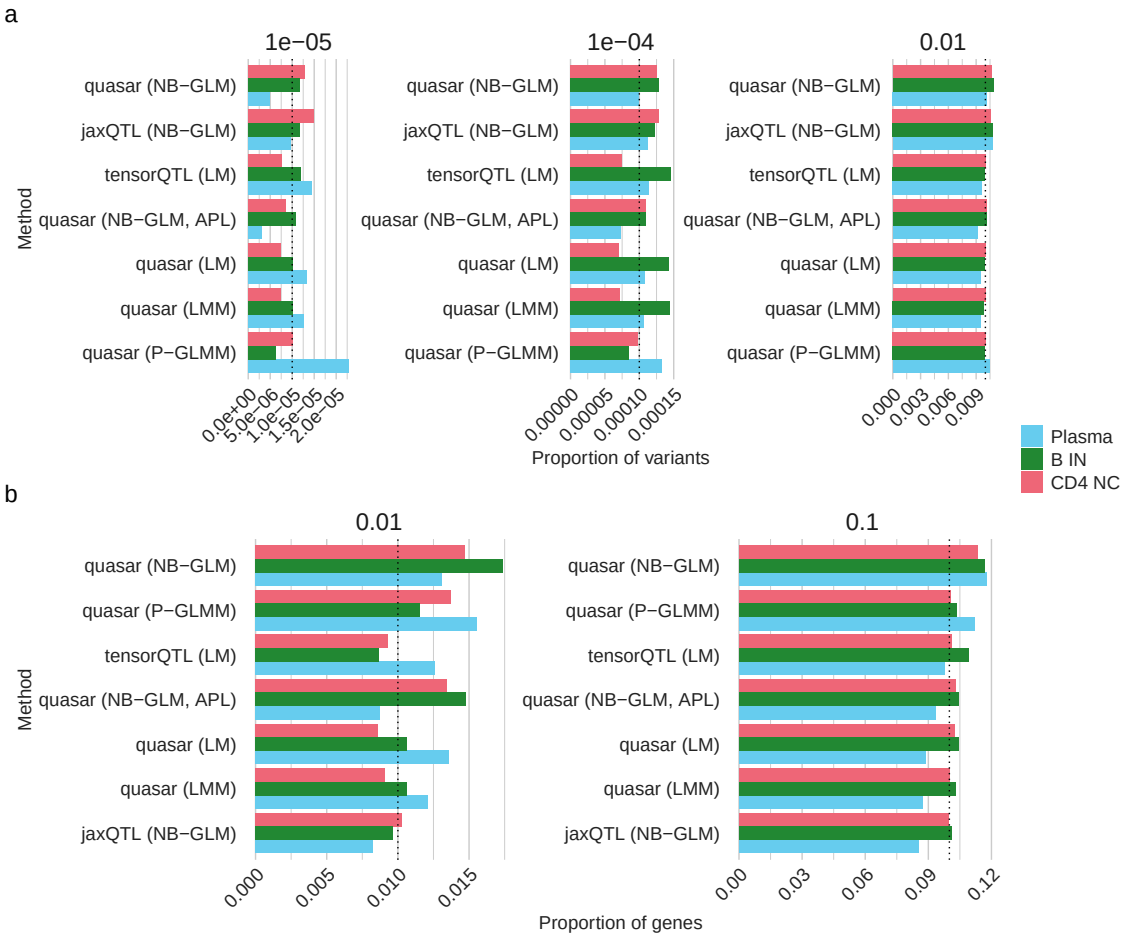
Supplementary Figures



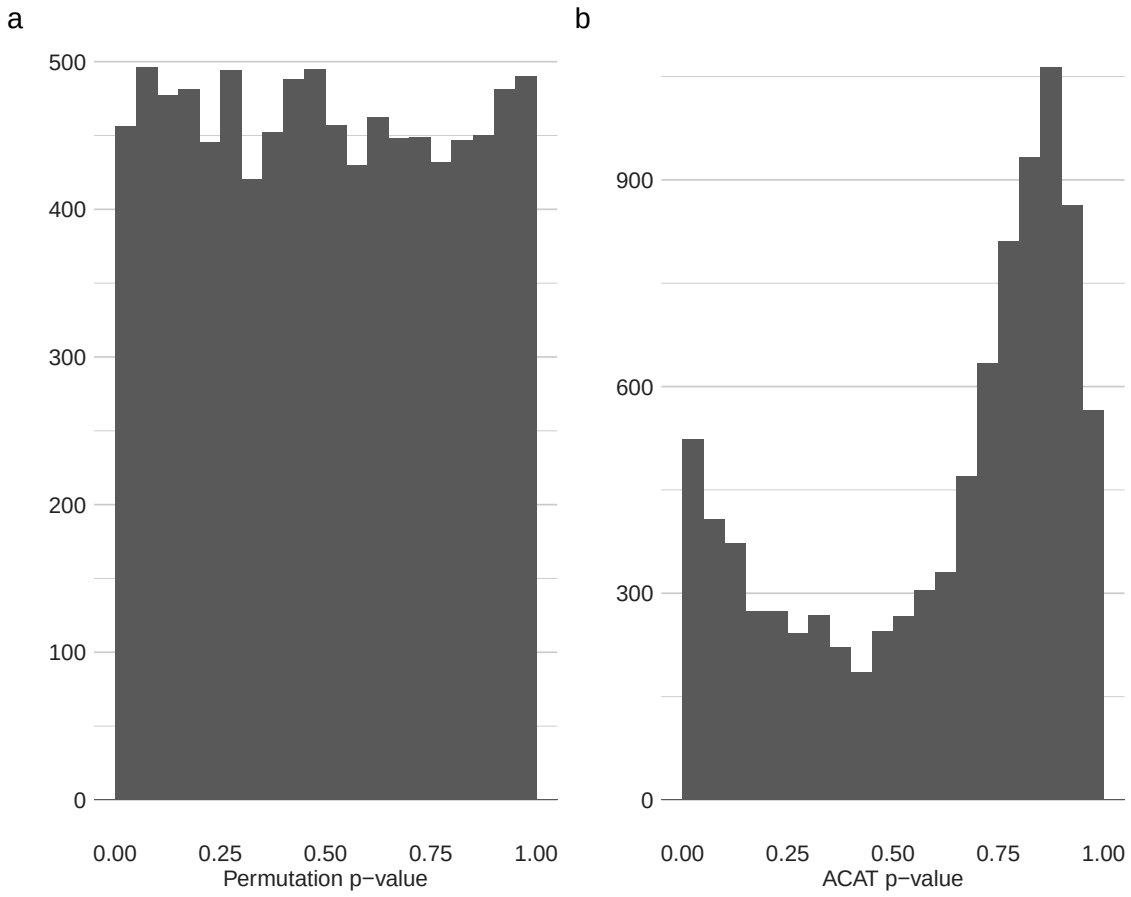
Supplementary Figure S1: **Distribution of the negative binomial dispersion estimates in the NB-GLMM model** Distribution of the estimated negative binomial dispersion, ϕ, for the negative binomial GLMM in the Plasma cell type. Almost all values are very small.



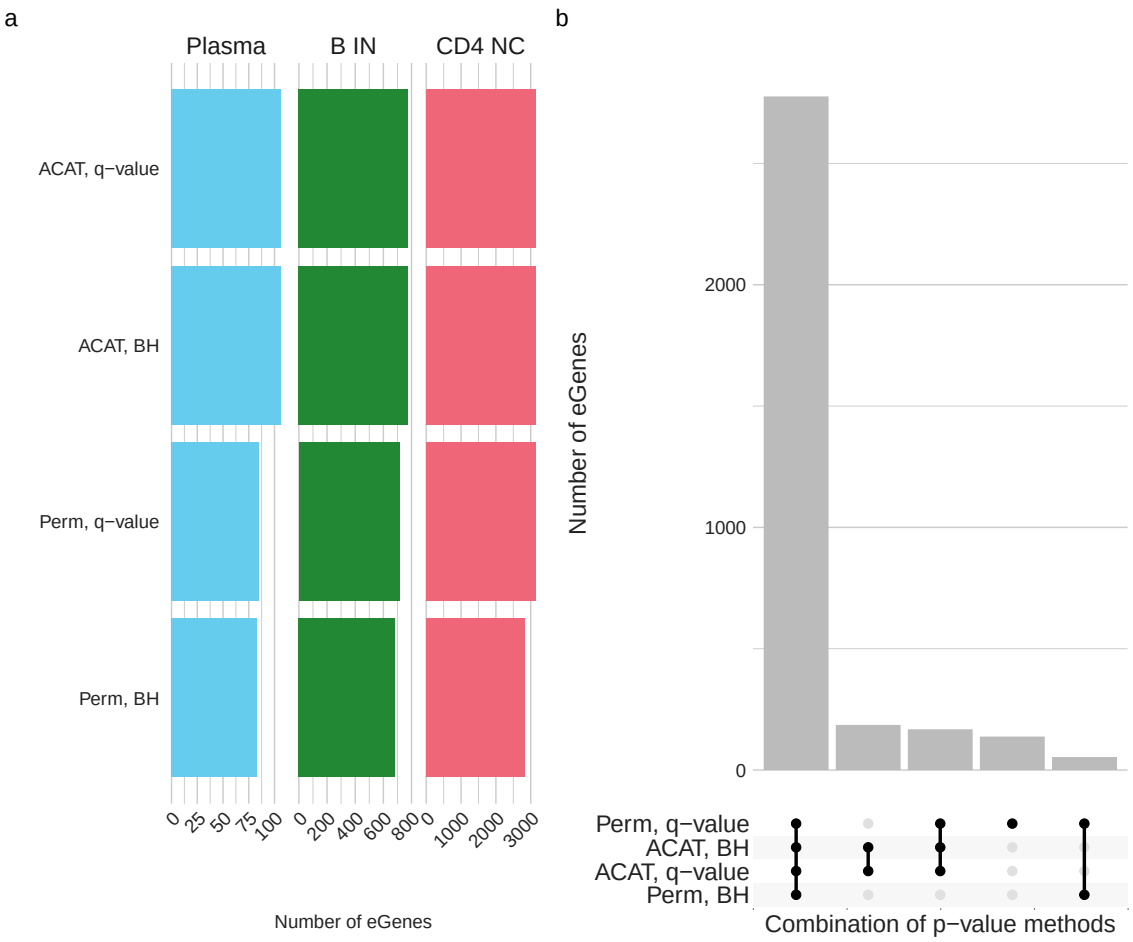
Supplementary Figure S2: **Distribution of negative binomial dispersion estimates with estimation methods** $\log_{10}$ transformed values of the negative binomial dispersion parameter, ϕ, for genes across the three representative cell types, estimated using both the Cox-Reid adjusted profile likelihood and maximum likelihood.



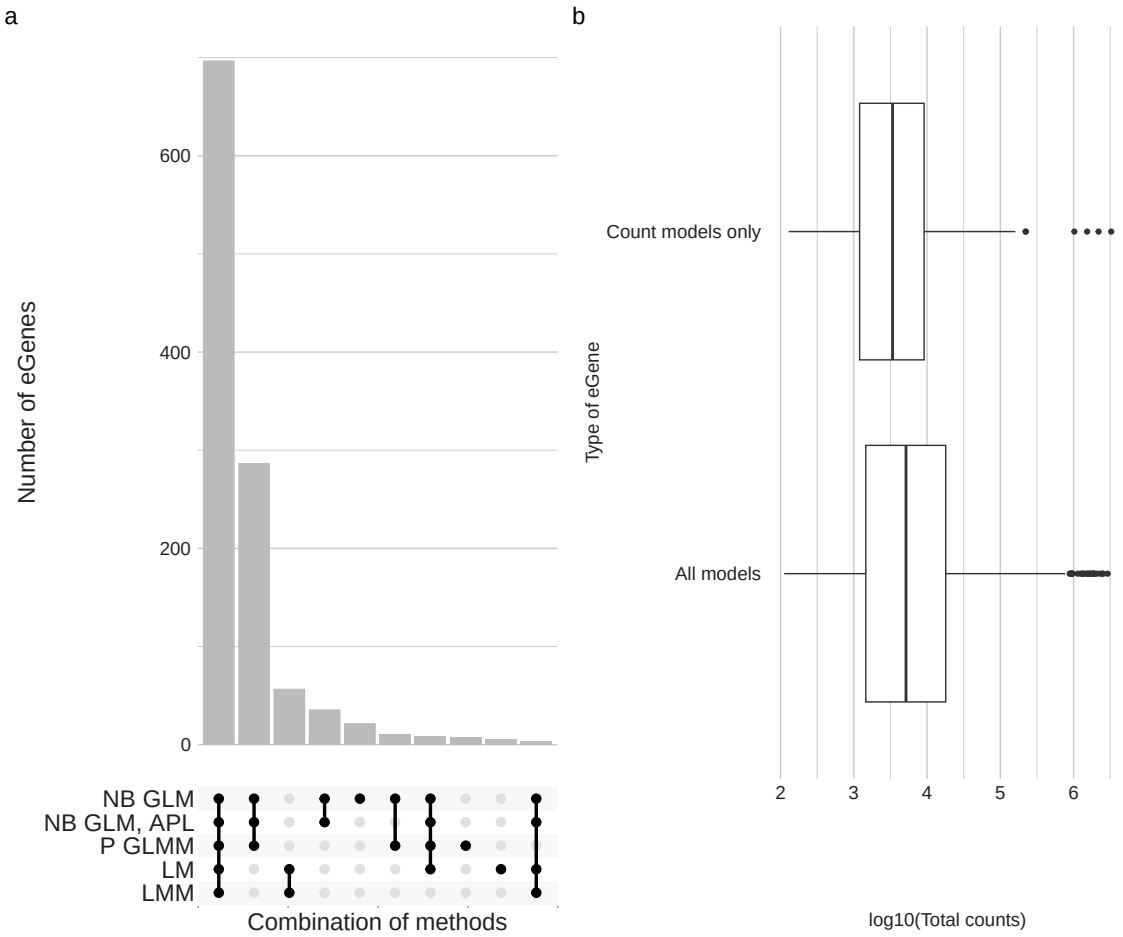
Supplementary Figure S3: **Empirical significance proportions at different thresholds** a) Proportion of significant variants at $\alpha = 10^{-5}, 10^{-4}, 10^{-2}$ under a permutation null for the three representative cell types. b) Proportion of signifiant genes at $\alpha = 10^{-1}, 10^{-2}$ under a permutation null across the three representative cell types. The dotted lines show the α level.



Supplementary Figure S4: **Distribution of gene-level p-values under a permutation null** a) p-values calculated by deriving a permutation p-values for the minimum per-gene p-value b) p-values calculated using ACAT applied to on variant-level p-values. Both panels show the results for tensorQTL run on all three representative cell types combined.



Supplementary Figure S5: **Comparison of different strategies for computing gene-level p-values** a) The number of eGenes called by different methods of computing gene-level p-values at a 5% FDR threshold. b) An upset plot showing the overlap between sets of eGene computed using different methods in the CD4 NC cell type. There are two methodological choices, leading to four combinations of methods. The p-values can be computed using either ACAT or permutations and FDR correction can be performed for using either the Benjamini-Hochberg procedure or with q-values. All results are for tensorQTL computed p-values. The permutation p-values are computed by tensorQTL and the ACAT p-values are computed from the variant-level p-values.



Supplementary Figure S6: **Characteristic of eGenes mapped by quasar** a) Upset plot of eGenes found across quasar methods in the B IN cell type. Most eGenes found are consistent across methods but a minority are found only by the count distribution methods. b) Boxplots of the log_{10} total counts for eGenes found by all methods or by count methods only