nature portfolio

# Peer Review File

# Genetic variants associated with cell-type-specific intra-individual gene expression variability reveal new mechanisms of genome regulation

Corresponding Author: Professor Joseph Powell

**This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.**

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Angli Xue and colleagues present a new approach to detect genetic variants associated with intra individual variability in single-cell RNA data, and an application of this approach to the OneK1K cohort. While this study represents an impressive amount of work, the authors do not sufficiently motivate or justify their approach in the context of previous methods, and several key details of their approach are not sufficiently explained.
1. Overall, the relationship between the authors' MEOTIVE approach and previous approaches is not very clear. The manuscript would benefit from a clearer enumeration of which features of the approach are novel and how they differ from previous approaches. In particular, I would like to see a more complete description of how other cited studies have parameterized the relationship between mean and variance or dispersion, what advantages and disadvantages each parameterization has, and what advantages and disadvantages the authors' approach has relative to previous methods. Without a straightforward comparison, is difficult to evaluate the novelty and importance of the authors' work.
2. Throughout the manuscript, the authors use the terms "dispersion" and "variance" somewhat interchangeably, using both to refer to the variability of expression counts data. While "variance" is a precise and well-defined term, "dispersion" does not have a universal definition, and the manuscript would benefit from a more precise and explicit definition. The authors define a "dispersion" parameter $\theta$ by the equation $\sigma^2 = \mu + \theta\mu^2$, but I am not aware of any use of this theta parameter except as a parameterization of the negative binomial distribution. It is possible that there is literature I am not familiar with that uses this definition; in my experience it is more typical to measure "dispersion" using the variance-mean ratio (VMR). Even in the context of the negative binomial distribution, while the theta parameter can be thought of as the excess variance compared to the mean, it is completely independent of the VMR and has an inverse relationship with both the mean and variance (perhaps explaining the observation that dispersion eQTLs tend to be in the opposite direction from variance eQTLs). This is dramatically at odds with the way the text describes dispersion as a measure of variability controlling for the mean-variance relationship in a conceptually similar way to the VMR and the coefficient of variation (CV). The authors should be clearer about the definition of "dispersion" they are using, describe its derivation and its properties, and argue explicitly that their definition is more appropriate than more commonly-used measures such as variance, VMR, or CV.
3. Throughout the manuscript, the authors discuss different probability distributions that can be fitted to expression counts data, including the zero-inflated negative binomial (ZINB) distribution, the standard (non-zero-inflated) negative binomial (NB) distribution, the Poisson distribution, and the Gaussian distribution. However, there is no mention of how the appropriate distribution for each gene or genotype is selected, and no p-values or any other goodness-of-fit statistics for distribution choice are given. The Methods also do not describe how any distribution other than the NB distribution is fitted or parameterized or how the dispersion parameter is defined for these distributions, referring to the Supplement for these details. Since the authors invoke the choice of distribution in several places to explain key observations, the details of how these distributions are selected, parameterized, and fitted are critical and it is very difficult to evaluate the authors' claims without them. Even without these details, the fact that the authors observe dramatic differences in expression patterns between different distributions suggests that these differences are a statistical artifact of how distributions are fitted rather than a meaningful difference in biology. In particular, the Poisson distribution is the limit of the NB distribution for low mean and dispersion, the Gaussian distribution is the limit of the NB distribution for large mean, and the NB distribution is the limit of the ZINB distribution for low zero-inflation, so it seems very likely that the choice between these distributions is determined by the values of these parameters rather than the parameter values being determined by the underlying distribution. In the absence of a hypothesis ascribing different biological meaning to different distributions, statistical

distributions should not be treated as biologically meaningful categories.
4. One of the key results the authors present, accounting for 2 of the paper's 5 figures, is the identification of rs1131017 as a deQTL for RPS26, which they hypothesize is due to that SNP's location in a transcription factor binding site and confirm this mechanism by co-expression analysis with transcription factors targeting that site. While this is a nice result, it is not obvious that it really required the MEOTIVE methodology. Both Figure 4 and Table 2 indicate that rs1131017 has large effects on both mean and variance of RPS26 expression, so it seems likely that the same hypothesis would have arisen from standard eQTL or veQTL analysis without the need to calculate deQTLs. The co-expression analysis performed for these genes appears to use mean expression levels without reference to dispersion; confusingly, the term "co-deQTL" seems to be used interchangeably with "co-eQTL" to describe eQTLs controlling co-expression in this analysis; while this implies that the authors may have performed some calculation of co-dispersion or dispersion-aware co-expression, I could not find a description of any such methodology in either the main text or the Methods. Altogether, it appears that this analysis does not rely on the new methodology reported in this paper. The authors should either argue for why this analysis would not have been possible without their new methodology or remove it from this paper and report it elsewhere.
5. Figure 1A is very confusing. The sc-eQTL panel has genotype on the y-axis while all other panels have genotype on the x-axis, and the sc-veQTL panel has expression variance as the dependent variable while all other panels have expression counts; this makes it difficult to visually compare these panels to the other panels. The panels are also not grouped in an intuitive order, with no grouping of single-cell vs bulk and no grouping of eQTLs vs vQTLs. Altogether, it makes it extremely difficult to see what is similar and what is different between these different kinds of analyses. There are also no panels illustrating the deQTL method that is the subject of this manuscript, which contributes to the lack of clarity about what the dispersion parameter measures and how it is expected to behave.
In addition to these major problems, I have three more minor comments:
6. (Results, p.10). "Most deQTLs are in the intergenic or intronic regions, implying they affect intra-individual variation via regulatory mechanisms." This is also true of ordinary eQTLs and GWAS loci generally, and what this implies about their mechanisms is still an open question in the field. In general, it is not safe to assume that the variants discovered in the analysis are the actual causal variants, rather than simply being markers linked to the causal variants. Here and in the paragraphs that follow, be a little more cautious about language that implies that these variants are actually causal.
7. (Results, p.11) "Our results suggest that the dispersion effects on the gene expression across cells within individuals are enriched in the biological process of immune response and viral infection." / (Discussion, p.15) "These results suggest that the transcriptional variability at the single-cell level could arise due to immune and/or external stimulus…." This is potentially a very interesting result, and I would like to see more discussion of it. Is there a clear mechanistic hypothesis of why this might be true? As an alternative explanation, most of the cell types profiled in this study are immune cells, and we should expect an enrichment for immune response and viral infection among genes with sufficient expression to calculate deQTLs in these cell types; can this possibility be ruled out? Similarly, could there be a confounding effect of contamination by exogenous RNA, including viral RNA?
8. (Results, p.15) In the replication analysis, only 6 out of 34 dGenes replicate. On its face, this is not very impressive and suggests that the dGenes may not be doing a good job of capturing any genuine biological signal; however, the authors argue that the replication dataset is very small and therefore it may not be expected for many genes to replicate. A permutation analysis would establish how many dGenes are expected to replicate by chance and give a baseline for evaluating the performance of this method. It would also be useful to compare the rate of replication of eGenes and vGenes across the same datasets.

Reviewer #2

(Remarks to the Author)
This paper examines the important question of how much cell-to-cell variation there is in gene expression and whether we can determine the underlying genetic basis for differences in expression variability itself. To this end, the authors develop a new statistical approach to analyze variance in expression from single-cell data, using a previously published dataset on expression in human blood cells across nearly 1000 individuals as the exemplar. The question addressed in this paper is important and will become increasingly relevant as more data of this type becomes available. The overall idea and framework of mapping expression variation using QTL or GWAS approaches is not itself novel, but the application of these ideas to a single-cell setting is definitely very worthwhile, as well as likely to gain increasing import over time. Thus the question is both timely and important, so the question then sits on whether the proposed method works as advertised and can be seen as an important advance. It seems as though the answers to these questions are yet, although the paper itself is fairly confusing in parts. This partly due to the potentially confusing nature of analyzing variance itself and partly because some important elements seem to be missing from the presentation.

The major contributions of the paper are:
1. A software pipeline that implements the variance analysis and pipeline approach.
2. An examination of the relationship between mean and variance expression levels and how to potentially correct for that in these analyses.
3. Application of the approach to an existing human expression data set.

From what I can tell from the github site, the authors do not invent or implement a new statistical approach, but instead apply existing statistical packages to explore this specific question. There is nothing wrong with this, but the naming of the approach (MEOTIVE) suggested something a bit more in terms of inventing a new method. The authors are clear to call this a "framework" and I think that that is appropriate. I leave it to the editors to decide how that falls in terms of expectations for Nature Communications.

Here are some suggestions for improving/clarifying the work:
1. The paper might benefit from being even more explicit about a hierarchical approach to what expression variation is. The authors are interested in "individual variation" but look to distinguish variation that is environmental or stochastic in nature from that which is due to genetic differences among individuals. Looking at variation within individuals but *among* cells could be a good approach, especially for humans. But we expect different cells to have different expression patterns for functional reasons (e.g., intestine vs brain), so we are really talking about cells in the same state of terminal differentiation. That is a point that probably could be made a bit more explicit. This is important because cells do not come with differentiation labels, but are assigned identities based on their expression patterns. This generates some potential circularity. The authors must rely on a subset of genes to not be variable enough to obscure the cellular assignment of a given cell. This is a potential source of error in the data and we might expect that this error would vary in a tissue-specific manner. This may not be an issue for the blood samples analyzed here, but could be more important in other samples in which more heterogeneous cell types might be collected (e.g., a biopsy).
2. So the authors are interested in among individual variation in gene expression among cells with identical functional differentiation states, which we then expected to have identical patterns of expression under the null hypothesis. Being a bit clearer about all of this might help when the authors start spinning out the numerous (and seemingly occasionally overlapping) labels that they use for their statistical quantities.
3. For example, the spends a great deal of time introducing various terms including eQTL, veQTL, sc-eQTL, sc-veQTL, etc. Then it proceeds to use TensorQTL and identify vGenes, without sufficient discussion of what this package is doing and what a vGene is. E.g., the text in line 153-157 does not explain what a vGene is after illustrating with with figures and using the term sc-veQTL in the same sentence. Presumably this is a gene with a significant QTL, but later it is then unclear what a dGene is. Also, the same gene could have multiple SNPs/QTL, and it is unclear how often this occurs.
4. Figure 1A is illustrative, but it would be helpful to better group the bulk eQTL/veQTL and the sc-eQTL/sc-veQTL, since the sc-eQTL and sc-veQTL are plotted against each other in the final panel. In the final panel, it is unclear what that dotted line represents. The units for expression (CPM, FPKM…) are also not labeled. It would be helpful to also present deQTL here and schematics of how these relate to dGenes, eGenes, vGenes because these are the major points of discussion later in the paper. Additionally, "sc-" is used as a prefix in this figure to refer to single cells here, but later in the paper this prefix is dropped while analysis is still presumably being done on single cells.
5. Figure 2 seems as though it could be a supplementary figure. It is already known that the mean and variance are highly correlated. It is good that this is the case here as well, but I think there needs to be more rationale for why this is a main figure.
6. Figure 2B indicates that the genes that have the strongest correlation between mean and variance are expressed at the lowest levels. Are genes with low expression reliable to be included in mean/variance estimates? In Figure 1, each point for the mean was from a single individual, but Figure 2 does not clarify this point. Is it the same here? Also, some of the points are extremely faint. How many points are being plotted for Figures 2B-C?
7. What is a dGene in Figure 2D? It is not labeled in the figure, only in the legend. The graph makes it appear that all genes are classified as either eGene or vGene, but it seems like this must be the breakdown once a gene is significant. My assumption had been that dGenes related to deQTL, which are distinct from eQTL and veQTL, but if dGenes are part of this figure then this must not be the correct assumption.
8. In Line 182 and following: "This lack of overlap validates that testing or such residual-eQTL is not valid for identifying true veQTL." How does this show that the method doesn't identify veQTL? This seems to simply show that two reasonable sounding approaches get very different answers, which highlights the problem for identifying veQTL reliably. How can the authors assert what are "true" veQTL? That would seem to require some sort of simulation/power analysis.
9. The results section has substantial discussion of potential mechanisms that are not tested and may be more suitable for the discussion.
10. The analysis in Figure 5 seems like a nice way to test a mechanistic hypothesis; however, it is still somewhat unclear what is being plotted in Figure 5. It would seem that Figure 5A is that transcription factor expression is affected by the eQTL in a way that is correlated with RPS26, and the beta estimate reflects the similar effect of the QTL on both genes (the TF and RPS26). Then in Figure 5B, the correlations are broken down by the genotype of the individuals. If that is the case, then it is confusing why several of the transcription factor effects in B seem to cluster near 0. Four outlier points were also removed, but from which graph, and to illustrate what?

Minor comments:
1. Line 48: parentheses with nothing inside
2. Line 61 should be 'a' SNP's genotype
3. Line 89 doesn't need 'the' before gene expression
4. Line 164 "the latter group is merely above the significant threshold, suggesting this group of signals are more likely to be random noises due to arbitrary threshold". The nature of the argument here is not entirely clear; it makes sense to indicate the q-values are close to 0.05 for vGenes without significant eGenes, but it is less clear what the 'random noises' part of the sentence is trying to suggest. Please reword for clarity.
5. Line 175 "verifying that mean veQTLs can explain most effects". Is there is some switching between nomenclature happening here? Earlier sentences referred to sc-eQTL and sc-veQTL, but is this now switching back to bulk experiments? Since veQTL refer to effects on variance, I believe the sentence should be indicate that eQTLs (effects on the mean) can explain most of the effects of veQTLs.
6. Line 211: No explanation of what a dGene is (this appears to be the first time the word 'dGene' this is mentioned).
7. Line 213: Do all deQTLs have a corresponding veQTL? deQTL also needs to be defined (only eQTL and veQTL have been defined)
8. Line 241: "Most deQTLs are in intergenic or intronic regions" – are these deQTL assigned to a 'dGene'? How does this assignment work? Earlier it was mentioned that cis-SNPs were analyzed to find deQTLs; what is the rationale for excluding potential trans-SNPs?

9. Line 268: ‘which’ shouldn’t be italicized
10. Figure 3: How many points are plotted in Figure 3A (i.e. how many SNP-gene pair tests are done for each cell type)? The color-coding by cell type and faceting by cell type are redundant.
11. What is the relationship between Figure 3A and B? It might be helpful to identify the dGenes plotted in B on the plots in Figure 3A by color-coding significant genes.
12. Line 317-318: The parenthetical comment “but it is only 0 vs 1 so it is not a large difference” is not really needed.
13. Line 395 which refers to Figure 5 references co-deQTL but this is not addressed in the figure.
14. Lines 459-461: What simulations are being referred to here, and how were they assessed for accuracy?

Reviewer #3

(Remarks to the Author)
Summary:

Building a more complete picture of how DNA sequence variants influence gene expression (eQTLs) is a critical challenge for 21st century biology. Single-cell RNA-sequencing technology has recently been applied to address this challenge and it is now accepted that many eQTLs are cell-type specific. Most of these studies have focused on identifying genetic variants that influence mean expression levels, however, there has been a long-standing interest in identifying variants that influence the variance in gene expression in addition to or independently of the mean. Studies addressing this question have focused on the genetic effects on inter-individual variance in gene expression, which in many cases reflect unmodelled GxG and GxE. Xu et al. took a different approach and instead focused on how genetic variants influence the intra-individual variance in gene expression. The authors took advantage of the fact scRNA-sequencing generates many more cells than individuals and identified thousands of genetic variants associated with intra-individual variance of gene expression (sc-veQTL and sc-deQTLs) in the OneK1K sc-RNA-seq dataset. The authors detected 4,642 sc-veQTLs across all 14 cell types, but found that the vast majority of these sc-veQTLs influence mean expression in a consistent direction and are thus not independent of the mean.

To identify genetic variants that influence the intra-individual variance independently of the mean the authors developed a framework (MEOTIVE) that estimates the effect of variants on the intra-individual dispersion of gene expression (sc-deQTLS). This method uses a negative binomial framework to estimate the dispersion and mean per person, and then, using regression, they map genetic variants associated with the intra-individuals dispersion of each gene. They applied their methodology to the OneK1K data and identified a modest number (55) of sc-deQTLs. The genes influenced by these sc-deQTLs were enriched in biological processes related to viral infection and immune responses with the specific examples of RPS26 and LYZ being explored in detail. Finally, the authors attempted to replicate their results in another study, an effort that was modestly successful even though the replication dataset they used was small relative to the OneK1K data.

The paper is well written and the methodology was described carefully where importantly the authors seem well aware of many of the pitfalls of analyzing sparse count-based data. Overall, the conclusions are well supported by the data, however, I do have some remaining questions about the statistical methodology, and the presentation of the data. Below are my comments in approximate order of importance.

Major comments

1. Work by Resztak et al. (PMID 35313584) used scRNA-seq data to map variance eQTLs with single-cell RNA-seq and found that even when modelling the inter-individual dispersion in a negative binomial framework there remains a negative linear relationship between the dispersion parameter and the mean (see Fig 6F). This observation was also reported in the seminal DeSeq2 manuscript which was one of the first papers to introduce the negative binomial model for RNA-seq analysis (PMID 25516281, Fig 1). Given this expected relationship between the dispersion parameter and the mean is it possible to conclude that the sc-deQTLs are independent of the mean as the authors claim between lines 203 and 207 when introducing their MEOTIVE approach? Along these lines of reasoning, do any of the 55 deQTLs violate this global relationship between the mean and dispersion parameters.
2. Based on my reading of the methods, it looks like the authors fed the processed counts from the package ‘sctransform’ into their per person estimators of dispersion and mean that were calculated with their method MEOTIVE. I understand the motivation behind the choice because the authors wanted an easy way to adjust the counts for batch effects prior to running their per sample estimator of the mean and dispersion of each gene. As far as I understand it, the ‘sctransform’ method involves fitting a regularized negative binomial model to the original counts and outputting new counts that are adjusted for the batch effects and the UMI count of each cell. I wonder whether this procedure to adjust the counts could result in artificial signals being observed. One idea I had to overcome this possible issue would be to learn the ‘batch’ and other parameters across their entire dataset with the ‘sctransform’ framework’ or another regression approach and then fix them as offsets along with a size factor in MEOTIVE.
3. In the section on the functional analysis of the deQTLs lines 290-303, the authors mention that if they remove all the five genes from the MHC from their deQTL list that the immune enrichment goes away. The MHC locus contains many complex haplotypes. Could mapping issues be influencing the estimates of dispersion in this region?
On lines 233-236 the authors mention that most dGenes are vGenes, but that there was minimal overlap with residual genes. Do the authors have an explanation for what kind of sources of variance the residual method captures and how that differs from MEOTIVE beyond the small statement they made in the text.

Minor comments

4. In the discussion at lines 443 to 451 the authors propose an explanation for the 'zero-inflated' model in single-cell RNA-seq data. There has been some debate about the use of zero-inflated models, and whether single-cell data is distributed in a zero-inflated negative binomial fashion at all (PMID 31937974). Given the relationship between the dispersion parameter and the mean as discussed above in comment #1, is another explanation that a single negative binomial doesn't model the number of zeros correctly when the mean differences between genotypes groups are large as in the case of rs1131017-RPS26. Rather, one needs to model the differences in dispersion as a function of genotype in addition to the mean, as is possible in the R package 'glmmTMB' with the dispformula option. I would suggest analyzing the data using this package, but it may be that fitting a negative binomial regression model with the amount of data used in this study is likely impractical.
5. I found some of the figures difficult to follow and the captions were a little sparse on details that were necessary to understand the plots. More detailed comments follow.
a. In the subplot of Figure 1A titled "sc-EQTL" the genotype is on the Y-axis, but for the rest of the subplots the genotype is the X-axis. The caption of Figure 1A is lacking some detail.
b. In Figure 1B, one of the 3 individuals has Poisson distributed gene expression. It makes sense to me this could be modelled in the negative binomial framework but might not be obvious to a reader. Another of the individuals looks like they have normally distributed gene expression, a process which isn't going to generate counts. I found that a little confusing. I wonder if all three of those examples could be drawn from negative binomial distributions to make it easier to understand.
c. In general the axis titles could be clearer. For example, in Figure 2A 'rho_s' should be converted to s, and in Figure 1, the underscores should be removed from the axis labels.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have performed an impressive amount of new analyses and done an admirable job of addressing reviewer comments. I have a few comments remaining.

1. While the authors did add compelling text and analysis describing the shortcomings of variance, VMR, and CV as measures of expression variability, the manuscript still does not convincingly argue that the proposed dispersion measurement does any better. In particular, it is stated that these metrics have high overlap with eGenes and (especially for variance) strong correlation with eQTL effect sizes. However, no direct comparison is made to dGenes and deQTLs, and I would like to see evidence that dGenes perform meaningfully better at these metrics. Based on the data presented it certainly looks like they do not: a majority of dGenes are also eGenes and deQTL effect sizes are very strongly negatively correlated with eQTL effect sizes. In the current state of the manuscript, I don't see a convincing argument that the MEOTIVE framework is really capturing an effect of dispersion independent of mean expression level in a way that veQTL analysis does not. This is not helped by the fact that every specific dGene discussed in the Results section (RPS26, LYZ, IL32, GNLY, GZMH, IFITM2, HLA-B) is also seen in Table 2 to be a highly significant eGene with an inverted direction of effect. I encourage the authors to focus on the minority of dGenes that are not also eGenes, and see if it is possible to identify an effect that is clearly independent of mean expression level.

2: On a similar note, I am still not convinced that the rs1131017-RPS26 analysis required the sc-deQTL approach, and I did not find any additions to the text or figures that address this. The authors argue in their response to my comment that the hypothesis does not arise without the deQTL analysis, but it is not clear to me why. Isn't the dispersion-reducing C allele also strongly associated with an increase in mean expression (as shown in Table 2 and Figure 3), and wouldn't this be discovered by a straightforward sc-eQTL analysis with no reference to variance or dispersion? Which of the analyses described are specific to the hypothesis that the C allele increases disease risk by reducing dispersion, and inconsistent with the hypothesis that the C allele increases disease risk by increasing mean expression? The co-eQTL analysis in particular actually seems to show that these TFs control mean expression rather than dispersion, as it measures a relationship between TF expression and RPS26 expression, NOT a relationship between TF expression and dispersion of RPS26 expression. I would like to see a clear statement in the text stating which part of the hypothesis requires the sc-deQTL analysis to formulate, and which part of the analysis suggests that the effect is due to dispersion rather than mean expression.

All my other comments were adequately addressed and I am very pleased in general with the responses. I have a few more minor suggestions:

3. The definition of the theta parameter is much clearer and easier to follow. I identified one statement that still needs to be rephrased to make this clearer: ll.211-217: "For example, when the expression level is low for a specific genotype group, the count data distribution is more left-skewed, driving the dispersion to be higher. For heterozygous individuals, the distribution approaches Gaussian; thus, the dispersion becomes much smaller, and so on for individuals with alternative homozygous alleles (see an example of RPS26 below, Figure 3). This feature results in a negative relationship between intra-individual variance and intra-individual dispersion for many genes with significant sc-veQTLs and sc-deQTLs." This is all correct and makes sense, but it is presenting dispersion as a function of the mean, when your argument is that dispersion should be considered independent of the mean while the variance is not. I suggest rephrasing this to clarify your argument that the low-mean, left-skewed distribution has deflated variance (not inflated dispersion), and that the dispersion is a more accurate

measure of variability because it accounts for the way the shape of the distribution changes with the mean.

4. I am now well convinced that the immune connection is real, but I think it would improve the argument to show some p-values in the Results section where it is discussed ("Functional annotations of sc-deQTLs and dGenes"). Most straightforwardly, I would like to see some p-values for the gene enrichment analysis in the main text, rather than just referring to the supplemental table. I also am very convinced by the analysis described in response to my comment, where a random subset of genes were selected and found not to have the same level of enrichment for immune, inflammatory, and viral genes, and I suggest using this to report a permutation-based p-value: take 100 or 1000 random samples of 34 genes and report the fraction of them containing at least 31 genes annotated as immune, inflammatory, or viral as the p-value.

5. I think the new framing of how distributions are discussed is much better and makes the statistical framework much easier to follow. I identified one place where the previous framing is preserved and is now even more confusing: ll.381-385, "We observe that the UMI count distribution for individuals in the CC genotype group mostly follows a negative binomial distribution. In contrast, the UMI count distributions for CG and GG individuals follow a Poisson distribution. This suggests that even for the same gene, genetic effects could impact the distribution type across the cells within an individual." Later, in the Discussion, it is described more clearly that the "Poisson-distributed" cells are a population with low dispersion, leading to the appearance of a zero-inflated / mixture distribution, when each individual population of cells can actually be modeled with negative binomial distributions with their own dispersion parameters. I suggest rephrasing this statement in the Results section to make a similar point.

Reviewer #3

(Remarks to the Author)
Xue et al have responded sufficiently to our comments. We have no further concerns with their manuscript.

Reviewer #4

(Remarks to the Author)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have once gain put a lot of work into responding to comments from all reviewers, including mine. I am mostly satisfied with the responses to my more minor comments, with one minor correction: the way empirical p-values are usually expressed would give you $p < 0.01$ for the gene set permutation test, not $p = 1$.

However, I do not think the answers to either of my major comments are adequate, and this leaves me with some concerns about the novelty and general usefulness of this approach. At the risk of repeating myself, I essentially have two major concerns left unaddressed:

1. The authors emphasize the difference between Poisson and negative binomial distributions and the ability of this method to distinguish between these distributions. I do not understand why this distinction is interesting or important, since the Poisson distribution is just a negative binomial distribution with dispersion set to 0. Choosing between a Poisson distribution and a negative binomial distribution can be rephrased as choosing between a negative binomial distribution with dispersion = 0 and a negative binomial distribution with dispersion > 0. I cannot think of any reason why choosing between negative binomial distributions with dispersion = 0 and dispersion > 0 would be meaningfully different from a standard negative binomial model that allows dispersion to be 0, and as far as I can tell none is given in the paper. If this is the main application of this method, I do not think it is of general interest.
2. The primary motivation given for using the dispersion statistic over other more commonly used statistics is that it should in principle be independent of the mean. However, in real data dispersion appears to be highly correlated with the mean. The authors point to the difference between Poisson and negative binomial distributions to explain this, but I reiterate that a Poisson distribution is just a negative binomial distribution with dispersion 0. The observation that negative binomial distributions with dispersion = 0 have systematically lower means than negative binomial distributions with dispersion > 0 is an expected consequence of the observation that dispersion is strongly correlated with mean, not an explanation of it. As currently presented, the theta statistic does not appear to do a good job of separating mean effects from dispersion effects, and the authors do not present a convincing argument that it does.

Reviewer #4

(Remarks to the Author)
Overall I feel this article is a nice contribution to the field.

I think the authors should consider that their response to one aspect of my previous critique may be unclear. I'll explain the details below.

I previously said:
Reviewer #2 correctly identified that the sentence “This lack of overlap validates that testing or such residual-eQTL is not valid for identifying true veQTL” is problematic.

Reviewer #2 had originally said:
How does this show that the method doesn’t identify veQTL? This seems to simply show that two reasonable sounding approaches get very different answers, which highlights the problem for identifying veQTL reliably. How can the authors assert what are “true” veQTL?

In my first review (of the initial resubmission), I noted:
The authors’ response does not seem to fully address the concern. They explain that they found residuals with a spurious quadratic relationship with the mean. They provide some example plots of these relationships that don’t look very convincing to me — the plots look like they are either dominated by a few outlier points, or perhaps linear and noisy. More generally, why is this proof that these genes are not in fact veQTL?

Now, in the new submission and in their rebuttal:
The authors explain that the outliers in the mean vs variance plot are those with higher variance. And that, looking at the mean vs residuals plots, there is generally a negative relationship between mean and residuals. I don’t understand how this shows a quadratic relationship. Shouldn’t the authors be asserting that their mean vs residuals plots show that the data don’t fit model assumptions for the residuals mapping model? And that this model mis-specification in fact would be consistent with the residuals-QTL method not being well suited to identifying true veQTL?

Version 3:

Reviewer comments:

Reviewer #4

(Remarks to the Author)
I am satisfied by the reviewers' revisions to their manuscript.

---
Here are a few comments I had in response to the authors' rebuttal comments to Reviewer #1.

In response to Reviewer #1 point 1: In the authors' revised MS, lines 629-632 state:
"In sparse count regimes typical of single-cell data, dispersion estimates are known to be biased upward for lowly expressed genes, which can lead to apparent overdispersion even when the underlying count process is effectively Poisson"
Please clarify if "known to be biased upward" is a result derived from your analyses (e.g. Supplementary Fig 17) or also known elsewhere in the literature (if so, include references).

In the authors' rebuttal to Reviewer #1 point 2:
"We agree that, under the negative binomial (NB) model, the mean and dispersion are not theoretically independent in empirical datasets"
Unclear -- contradiction between theoretical and empirical here...
And "Our objective in using the dispersion statistic was ... to identify genetic effects on expression variability beyond those attributable to mean expression
differences."
But the authors just said that the mean and dispersion are not independent in empirical datasets. So why are they explicitly using dispersion stat to identify effects on variability beyond the mean? They need to clarify or perhaps further hedge their claims.
The additional analysis where dispersion estimates are conditioned on mean prior to mapping is good. It is encouraging that many of their deQTLs remain significant. However, >40% of their deQTLs are not significant in this analysis. I think their unconditioned analysis includes both variants that affect expression variability in a mean-independent fashion as well as some that increase variability at least partially through an effect on mean (the latter accounting for incomplete overlap with mean-conditioned analysis, although power is certainly a component as well).

Overall the work is good but I think the reviewers need to be a little more nuanced to clarify that their method is not completely able to disentangle effects on mean vs dispersion in all cases.

(Remarks on code availability)
Code appears to be structured clearly and somewhat straightforward to run. I noticed that there are some places where this group's specific absolute paths to the data are hardcoded. It would be ideal if the authors could e.g. have these stored in variables that clearly indicate how they would be modified in the case of someone working with data downloaded from Zenodo into $DIR.

Reviewer #1 (Remarks to the Author):

Angli Xue and colleagues present a new approach to detect genetic variants associated with intra individual variability in single-cell RNA data, and an application of this approach to the OneK1K cohort. While this study represents an impressive amount of work, the authors do not sufficiently motivate or justify their approach in the context of previous methods, and several key details of their approach are not sufficiently explained.

1. Overall, the relationship between the authors' MEOTIVE approach and previous approaches is not very clear. The manuscript would benefit from a clearer enumeration of which features of the approach are novel and how they differ from previous approaches. In particular, I would like to see a more complete description of how other cited studies have parameterized the relationship between mean and variance or dispersion, what advantages and disadvantages each parameterization has, and what advantages and disadvantages the authors' approach has relative to previous methods. Without a straightforward comparison, is difficult to evaluate the novelty and importance of the authors' work.

Re: Thank you for raising this point; we agree that comparing *MEOTIVE* to alternative methods/approaches is important, and we did not adequately address this in our original submission. We have now done so and included the new results in the revised manuscript (lines 182-185) and more explicit text comparing the advantages and disadvantages (lines 542-549, and Supplementary Note 2). We have compared *MEOTIVE* against two previously used methods to estimate the dispersion in single-cell data, including coefficient of variation (CV) and variance-to-mean ratio (VMR, aka Fano factor). It is worth noting here that CV and VMR are not independent of the mean, which was a factor in initially deciding not to apply these methods. However, they are easy to estimate and computationally efficient.

As with the mean, dispersion and variance models, we calculated the CV and VMR for each gene/cell type that passed the QC thresholds and tested against the same SNPs (methods, lines 635-637). For individuals with a mean of zero gene expression, we fixed the CV and VMR to 1, mimicking the scenario where the mean is equal to the variance under the Poisson distribution. We observed substantially lower power to detect effects with VMR (Supplementary Table 2, and see below) compared to CV, and the majority (70.0 ~ 94.1% across cell types) of the genes detected overlapped with eGenes. These results show lower power than using the variance (veQTL) method and a high concordance with the mean effect eQTL. Overall, we conclude that there are better metrics to capture the variability in gene expression that is not induced by mean differences.

| Cell_type | eGene | cvGene | vmrGene | overlap_e_cv | overlap_e_vmr |
|---|---|---|---|---|---|
| B_IN | 843 | 319 | 15 | 284 | 13 |
| B_MEM | 672 | 223 | 19 | 206 | 15 |
| CD4_NC | 4252 | 2559 | 58 | 2371 | 50 |
| CD4_ET | 986 | 358 | 19 | 337 | 11 |
| CD4_SOX4 | 37 | 9 | 2 | 8 | 0 |
| CD8_NC | 1760 | 756 | 28 | 691 | 22 |
| CD8_ET | 1884 | 913 | 40 | 804 | 31 |
| CD8_S100B | 461 | 155 | 12 | 131 | 8 |
| DC | 228 | 51 | 11 | 44 | 8 |
| Mono_C | 517 | 173 | 16 | 146 | 12 |
| Mono_NC | 497 | 161 | 13 | 140 | 10 |
| NK_R | 179 | 45 | 4 | 39 | 3 |
| NK | 2099 | 927 | 48 | 853 | 43 |
| Plasma | 70 | 10 | 3 | 7 | 0 |

Significant effects are shown per cell type and the overlap between different models. eGenes are the number of eGenes detected using the mean effect model. Meanwhile, cvGenes and vmrGenes are the significant genes detected using the CV and VMR metrics. The overlaps are the number of cvGenes and vmrGenes observed in the eGene list.

2. Throughout the manuscript, the authors use the terms “dispersion” and “variance” somewhat interchangeably, using both to refer to the variability of expression counts data. While “variance” is a precise and well-defined term, “dispersion” does not have a universal definition, and the manuscript would benefit from a more precise and explicit definition. The authors define a “dispersion” parameter $\theta$ by the equation $\sigma^2 = \mu + \theta\mu^2$, but I am not aware of any use of this theta parameter except as a parameterization of the negative binomial distribution. It is possible that there is literature I am not familiar with that uses this definition; in my experience it is more typical to measure “dispersion” using the variance-mean ratio (VMR). Even in the context of the negative binomial distribution, while the theta parameter can be thought of as the excess variance compared to the mean, it is completely independent of the VMR and has an inverse relationship with both the mean and variance (perhaps explaining the observation that dispersion eQTLs tend to be in the opposite direction from variance eQTLs). This is dramatically at odds with the way the text describes dispersion as a measure of variability controlling for the mean-variance relationship in a conceptually similar way to the VMR and the coefficient of variation (CV). The authors should be clearer about the definition of “dispersion” they are using, describe its derivation and its properties, and argue explicitly that their definition is more appropriate than more commonly-used measures such as variance, VMR, or CV.

Re: We acknowledge that using the terms “dispersion” and “variance” interchangeably throughout the manuscript is confusing. We have clarified this throughout the manuscript, including an initial clarification of the terms (lines 107-110).

We apologise for the lack of clarity on how we have aligned dispersion definitions with the overall aim of linking genetic variants to within-individual variability in gene expression. For background, in the original manuscript, when we claimed that the *dispersion was a measure of variability controlling for the mean-variance relationship*, we intended to convey that the estimation itself is not a function of the mean compared to VMR or CV. In one of our recent papers (Xue et al. *Genome Biology*. 2023), we have shown that the standard index of dispersion, such as the coefficient of variation and variance to mean ratio, is highly dependent on the mean expression of the genes (Panel C of Supp. Figure 1 in that paper). Thus, indices like VMR or CV cannot capture the variability of intra-individual gene expression, which is not at least partly explained by the mean expression levels. We have added new results for QTL associations with VMR and CV (lines 182-185) and justified our choice of using the dispersion parameter in a negative binomial model in the main text (lines 542-549) and Supplementary Note 2. Here, we have followed the definition of dispersion for single-cell data as per Love et al. *Genome Biology* and Hafemeister et al. *Genome Biology,* where dispersion assumes a negative binomial distribution of the data.

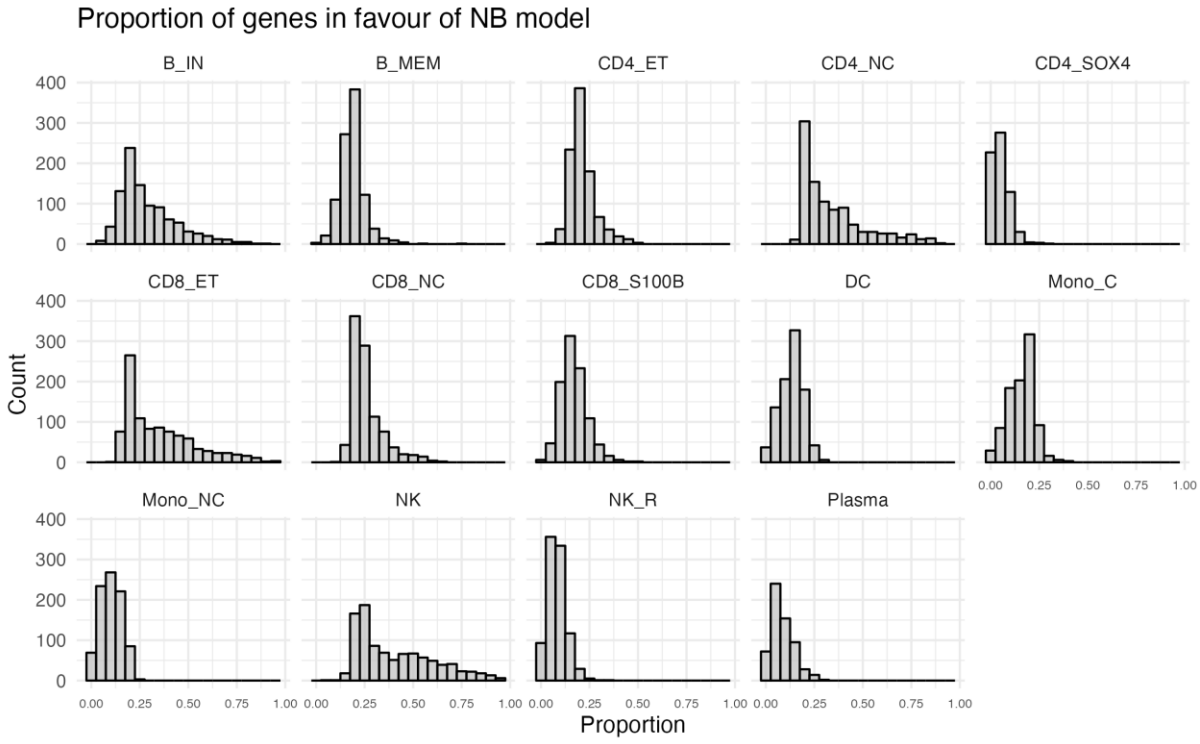
3. Throughout the manuscript, the authors discuss different probability distributions that can be fitted to expression counts data, including the zero-inflated negative binomial (ZINB) distribution, the standard (non-zero-inflated) negative binomial (NB) distribution, the Poisson distribution, and the Gaussian distribution. However, there is no mention of how the appropriate distribution for each gene or genotype is selected, and no p-values or any other goodness-of-fit statistics for distribution choice are given. The Methods also do not describe how any distribution other than the NB distribution is fitted or parameterized or how the dispersion parameter is defined for these distributions, referring to the Supplement for these details. Since the authors invoke the choice of distribution in several places to explain key observations, the details of how these distributions are selected, parameterized, and fitted are critical and it is very difficult to evaluate the authors’ claims without them. Even without these details, the fact that the authors observe dramatic differences in expression patterns between different distributions suggests that these differences are a statistical artifact of how distributions are fitted rather than a meaningful difference in biology. In particular, the Poisson distribution is the limit of the NB distribution for low mean and dispersion, the Gaussian distribution is the limit of the NB distribution for large mean, and the NB distribution is the limit of the ZINB distribution for low zero-inflation, so it seems very likely that the choice between these distributions is determined by the values of these parameters rather than the parameter values being determined by the underlying

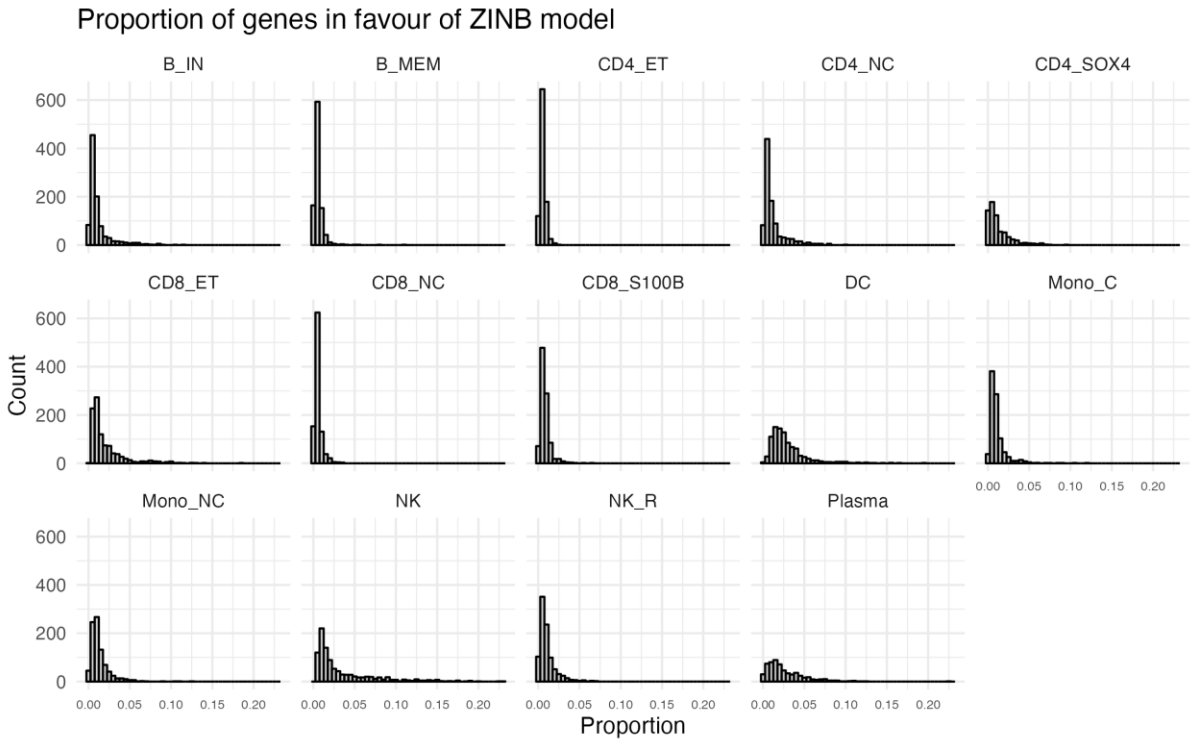
distribution. In the absence of a hypothesis ascribing different biological meaning to different distributions, statistical distributions should not be treated as biologically meaningful categories.

Re: We thank the reviewer for raising the importance of evaluating alternative distributions and their effect on the parametrisation of the model and apologise for not being clearer in the choice of parameterisation. We have addressed this and updated the manuscript in lines 579-581 and Supplementary Note 3. We chose to use NB model to fit the expression counts data based on the following reasoning:

First, we estimated every individual's intra-individual mean and variance for all genes. We found that, except for genes with zero expression, most intra-individual variance estimates are larger or equal to the intra-individual mean estimates.

Second, we ran a likelihood ratio test (LRT) to evaluate the goodness-of-fit between the Poisson and NB models for all the genes in every individual in each cell type. On average, 21.1% of genes showed nominal evidence ($P < 0.05$) favouring the NB model rather than the Poisson model (Supplementary Figure 14). We also ran an LRT between NB and ZINB, which showed that only a very small proportion of genes (~1.46%) showed evidence in favour of ZINB distribution (Supplementary Figure 15). Based on these tests, we chose to use NB model as the default model and estimate the dispersion parameter ($\theta$) as the indicator for intra-individual variability.






4. One of the key results the authors present, accounting for 2 of the paper's 5 figures, is the identification of rs1131017 as a deQTL for RPS26, which they hypothesize is due to that SNP's location in a transcription factor binding site and confirm this mechanism by co-expression analysis with transcription factors targeting that site. While this is a nice result, it is not obvious that it really required the MEOTIVE methodology. Both Figure 4 and Table 2 indicate that rs1131017 has large effects on both mean and variance of RPS26 expression, so it seems likely that the same hypothesis would have arisen from standard eQTL or veQTL analysis without the need to calculate deQTLs. The co-expression analysis performed for these genes appears to use mean expression levels without reference to dispersion; confusingly, the term "co-deQTL" seems to be used interchangeably with "co-eQTL" to describe eQTLs controlling co-expression in this analysis; while this implies that the authors may have performed some calculation of co-dispersion or dispersion-aware co-expression, I could not find a description of any such methodology in either the main text or the Methods. Altogether, it appears that this analysis does not rely on the new methodology reported in this paper. The authors should either argue for why this analysis would not have been possible without their new methodology or remove it from this paper and report it elsewhere.

Re: Thank you for raising this point, and we apologise that the text was unclear concerning how this observation was discovered. The same hypothesis for rs1131017-*RPS26* would not simply arise without conducting deQTL analysis. This is because even if both eQTL and veQTL are detected, given the high correlation between the

mean and variance estimates, we would not know the genetic effect on the dispersion level and whether its effect is in the same or opposite direction to the mean effect. Also, if we did not estimate the dispersion of each individual, we would not be able to identify the mixture distribution issue for this gene (Figure 4).

We also apologise for the confusion about “co-deQTL” and “co-eQTL”. We confirm the “co-deQTL” is a typo and should be “co-eQTL”. We have corrected it throughout the main text.

5. Figure 1A is very confusing. The sc-eQTL panel has genotype on the y-axis while all other panels have genotype on the x-axis, and the sc-veQTL panel has expression variance as the dependent variable while all other panels have expression counts; this makes it difficult to visually compare these panels to the other panels. The panels are also not grouped in an intuitive order, with no grouping of single-cell vs bulk and no grouping of eQTLs vs vQTLs. Altogether, it makes it extremely difficult to see what is similar and what is different between these different kinds of analyses. There are also no panels illustrating the deQTL method that is the subject of this manuscript, which contributes to the lack of clarity about what the dispersion parameter measures and how it is expected to behave.

Re: Thanks for the suggestion. We have regrouped the panels in Figure 1 and explicitly separate the pseudobulk and single-cell QTLs. We also add a panel to describe the deQTL identification.

In addition to these major problems, I have three more minor comments:

6. (Results, p.10). “Most deQTLs are in the intergenic or intronic regions, implying they affect intra-individual variation via regulatory mechanisms.” This is also true of ordinary eQTLs and GWAS loci generally, and what this implies about their mechanisms is still an open question in the field. In general, it is not safe to assume that the variants discovered in the analysis are the actual causal variants, rather than simply being markers linked to the causal variants. Here and in the paragraphs that follow, be a little more cautious about language that implies that these variants are actually causal.

Re: Agreed. We have revised the wording to avoid causal inference (line 243).

7. (Results, p.11) “Our results suggest that the dispersion effects on the gene expression across cells within individuals are enriched in the biological process of immune response and viral infection.” / (Discussion, p.15) “These results suggest that

the transcriptional variability at the single-cell level could arise due to immune and/or external stimulus….” This is potentially a very interesting result, and I would like to see more discussion of it. Is there a clear mechanistic hypothesis of why this might be true? As an alternative explanation, most of the cell types profiled in this study are immune cells, and we should expect an enrichment for immune response and viral infection among genes with sufficient expression to calculate deQTLs in these cell types; can this possibility be ruled out? Similarly, could there be a confounding effect of contamination by exogenous RNA, including viral RNA?

Re: We appreciate the reviewer's insightful comment regarding the potential enrichment of immune response and viral infection-related pathways. We do not believe that we expect an enrichment in viral infection just because most of the cell types profiled in this study are immune cells. We base this on the following: None of the genes in the enriched ‘viral gene expression’ pathway is expressed explicitly in blood immune cells and ubiquitously across tissues (according to GTEx database portal). Also, we randomly subset 34 genes (matching with the number of dGenes) from the eGene list, and ran the enrichment analysis with exactly the same parameters. No significant enrichment was observed for immune response and viral infection pathways, again supporting the conclusion that the enrichment of dGenes is not simply because we used immune cells as the main samples.

Contamination by exogenous RNA, including viral RNA, also seems unlikely because the enrichment signal came from a list of 7 genes, and they are either C-C motif chemokine ligand (*CCL3*, *CCL4*) or ribosome genes (*RPLP2*, *RPS26*, *RPS9*, *RPS18*, *RPS10*). Contamination by exogenous RNA will not change the gene expression levels from the single-cell RNA-seq protocols because exogenous RNA would not be aligned to the human reference genome in Cell Ranger software. We have updated the discussion to include these points (lines 437-439).

8. (Results, p.15) In the replication analysis, only 6 out of 34 dGenes replicate. On its face, this is not very impressive and suggests that the dGenes may not be doing a good job of capturing any genuine biological signal; however, the authors argue that the replication dataset is very small and therefore it may not be expected for many genes to replicate. A permutation analysis would establish how many dGenes are expected to replicate by chance and give a baseline for evaluating the performance of this method. It would also be useful to compare the rate of replication of eGenes and vGenes across the same datasets.

Re: Thank you very much for this suggestion. Thanks for pointing out the potential issues with the replication analysis. We want to clarify that we also believe that the small number of the replicated dGenes is due to differences in the genetic ancestral

diversity between the cohorts (and thus different allele frequencies), disease status (Lupus in the replication and healthy in the discovery), in addition to the small sample size (N = 97). We also found that the FDR was calculated based on all the SNP-gene pairs across each chromosome for all genes (not just 34 dGenes).

To improve our replication analysis, we obtained the other two sub-cohorts from Perez et al. of the European ancestry (i.e., EUR CLUES cohort and ImmVar Consortium), one of which also has a similar sample size (N = 89 and 46). We reran the FDR adjustment just for the pairs between SNP and dGenes. In the updated results, we replicated 12 (EAS CLUES), 17 (EUR CLUES), and 9 (ImmVar Consortium) dGenes at FDR < 5%. By meta-analysing three cohorts in Perez et al. dataset (total N = 232), we replicated 26/34 (76.5%) dGenes in the corresponding cell types. The results and methods have been updated with this new analysis (lines 412-421 and 697-706).

In the OneK1K paper (Yazar et al. *Science*. 2022), the replication analysis for eSNP-eGene pairs has been done using the Perez et al. data and 87% in the European cohort and 78% in the Asian cohort replicated at the FDR threshold of 5%, when correcting the FDR distributions under the assumptions of equal sample size. This showed that we have a similar and reasonably great replication rate of dGenes compared to eGenes.

Reviewer #2 (Remarks to the Author):

This paper examines the important question of how much cell-to-cell variation there is in gene expression and whether we can determine the underlying genetic basis for differences in expression variability itself. To this end, the authors develop a new statistical approach to analyze variance in expression from single-cell data, using a previously published dataset on expression in human blood cells across nearly 1000 individuals as the exemplar. The question addressed in this paper is important and will become increasingly relevant as more data of this type becomes available. The overall idea and framework of mapping expression variation using QTL or GWAS approaches is not itself novel, but the application of these ideas to a single-cell setting is definitely very worthwhile, as well as likely to gain increasing import over time. Thus the question is both timely and important, so the question then sits on whether the proposed method works as advertised and can be seen as an important advance. It seems as though the answers to these questions are yet, although the paper itself is fairly confusing in parts. This partly due to the potentially confusing nature of analyzing variance itself and partly because some important elements seem to be missing from the presentation.

The major contributions of the paper are:

1. A software pipeline that implements the variance analysis and pipeline approach.
2. An examination of the relationship between mean and variance expression levels and how to potentially correct for that in these analyses.
3. Application of the approach to an existing human expression data set.

From what I can tell from the github site, the authors do not invent or implement a new statistical approach, but instead apply existing statistical packages to explore this specific question. There is nothing wrong with this, but the naming of the approach (MEOTIVE) suggested something a bit more in terms of inventing a new method. The authors are clear to call this a “framework” and I think that that is appropriate. I leave it to the editors to decide how that falls in terms of expectations for Nature Communications.

Re: Thank you very much for your time and efforts in reviewing our manuscript. We appreciate the comments and suggestions, which are very constructive and helped us improve the manuscript.

Regarding MEOTIVE, as observed, we refer to this as a framework rather than a new statistical approach. The rationale for doing so is that during the development of this project, we realised that adequately testing for allelic effects on within-individual gene expression variance involves many steps and numerous steps that would not easily be reproducible for others.

Here are some suggestions for improving/clarifying the work:

1. The paper might benefit from being even more explicit about a hierarchical approach to what expression variation is. The authors are interested in “individual variation” but look to distinguish variation that is environmental or stochastic in nature from that which is due to genetic differences among individuals. Looking at variation within individuals but *among* cells could be a good approach, especially for humans. But we expect different cells to have different expression patterns for functional reasons (e.g., intestine vs brain), so we are really talking about cells in the same state of terminal differentiation. That is a point that probably could be made a bit more explicit. This is important because cells do not come with differentiation labels, but are assigned identities based on their expression patterns. This generates some potential circularity. The authors must rely on a subset of genes to not be variable enough to obscure the cellular assignment of a given cell. This is a potential source of error in the data and we might expect that this error would vary in a tissue-specific manner. This may not be an issue for the blood samples analyzed here, but could be more important in other samples in which more heterogeneous cell types might be collected (e.g., a biopsy).

Re: We thank the reviewer for pointing out the clarity issue of expression variation. We agree that we are evaluating genetic control of variation amongst cells that are of the same type/state, and that this could be made more explicit. We have revised the introduction to make this clearer (lines 107-110).

The point about potential circularity is well taken and one that we grappled with early in this project. To overcome this, we used an individual-level cell classification approach, where each cell was labelled based on total transcriptional fit to a reference dataset, labelled using orthogonal-protein-based markers (Yazar et al. Science 2022). Therefore, cell types (where we test for genetic effects within individual variance) are not identified and classified based on their unique, highly variable genes. However, it is also important to consider that intra-individual expression variation is a different concept from the variation used to define cell types. For example, a gene marker for CD4 Naïve cells is highly expressed in CD4 Naïve cells but not in other cell types. Within CD4 Naïve cells, different individuals could still have different levels of variations of this gene, and what we are trying to detect is the genetic variants that affect such variations.

2. So the authors are interested in among individual variation in gene expression among cells with identical functional differentiation states, which we then expected to have identical patterns of expression under the null hypothesis. Being a bit clearer about all of this might help when the authors start spinning out the numerous (and seemingly occasionally overlapping) labels that they use for their statistical quantities.

Re: Thank you for flagging this. As identified by other reviewers, the terminology of different QTLs could be clearer. We have made significant efforts to improve this in the manuscript, especially when introducing terms (see lines 107-110). We would like to note that since we are testing the genetic association with intra-individual variability of gene expression, even under hypothesis, different individuals do not have to have identical patterns of expression. Individuals can have different levels of intra-individual variability as long as the average level of this intra-individual variability in each genotype group is similar. The alternative hypothesis is depicted in the 'Bulk veQTL' in Figure 1A.

3. For example, the spends a great deal of time introducing various terms including eQTL, veQTL, sc-eQTL, sc-veQTL, etc. Then it proceeds to use TensorQTL and identify vGenes, without sufficient discussion of what this package is doing and what a vGene is. E.g., the text in line 153-157 does not explain what a vGene is after illustrating with with figures and using the term sc-veQTL in the same sentence.

Presumably this is a gene with a significant QTL, but later it is then unclear what a dGene is. Also, the same gene could have multiple SNPs/QTL, and it is unclear how often this occurs.

Re: Thank you. We have added the definition of vGene and dGene where they were first mentioned (lines 149-151 and lines 211-212). We also added to the description how TensorQTL is used and defined the significance levels in the Methods section (line 551-571).

The same gene could have multiple SNPs/QTL associated with its expression levels. This is because we tested all SNPs located within the 1Mb window of the genes. This is a common scenario, but we often report the top SNP/QTL of the genes in the main tables.

4. Figure 1A is illustrative, but it would be helpful to better group the bulk eQTL/veQTL and the sc-eQTL/sc-veQTL, since the sc-eQTL and sc-veQTL are plotted against each other in the final panel. In the final panel, it is unclear what that dotted line represents. The units for expression (CPM, FPKM…) are also not labeled. It would be helpful to also present deQTL here and schematics of how these relate to dGenes, eGenes, vGenes because these are the major points of discussion later in the paper. Additionally, “sc-” is used as a prefix in this figure to refer to single cells here, but later in the paper this prefix is dropped while analysis is still presumably being done on single cells.

Re: Thank you for this suggestion. We have revised the illustration of Figure 1A and regrouped the panels based on bulk approach and single-cell approach. This figure is illustrative so the units for expression are not like CPM or FPKM but metrics derived from raw counts without scaling. We also added “sc-” prefix wherever necessary to avoid confusion.

5. Figure 2 seems as though it could be a supplementary figure. It is already known that the mean and variance are highly correlated. It is good that this is the case here as well, but I think there needs to be more rationale for why this is a main figure.

Re: Thanks for your suggestion. We have now moved Figure 2 to the supplementary figure and added the relationship between different variance indicators (variance, CV, VMR, theta).

6. Figure 2B indicates that the genes that have the strongest correlation between mean and variance are expressed at the lowest levels. Are genes with low expression reliable to be included in mean/variance estimates? In Figure 1, each point for the mean was from a single individual, but Figure 2 does not clarify this point. Is it the same here? Also, some of the points are extremely faint. How many points are being plotted for Figures 2B-C?

Re: We filtered out the genes with expression in less than 10% of individuals or extremely low inter-individual abundance ($\mu < 0.001$) (lines 502-504). Among the remaining genes, some were still vulnerable to estimation bias in the intra-individual dispersion parameter. Therefore, we further excluded these genes from being classified as dGenes, with the cut-off threshold for mean expression determined using a simulation strategy (lines 610-622).

In Figure 2 (now Supplementary Figure 1), each dot represents a gene rather than an individual. It is the mean of intra-individual mean or variance across all individuals. There are 24,364 points/genes plotted for Figures 2B-C. We have clarified this in the figure legend.

7. What is a dGene in Figure 2D? It is not labeled in the figure, only in the legend. The graph makes it appear that all genes are classified as either eGene or vGene, but it seems like this must be the breakdown once a gene is significant. My assumption had been that dGenes related to deQTL, which are distinct from eQTL and veQTL, but if dGenes are part of this figure then this must not be the correct assumption.

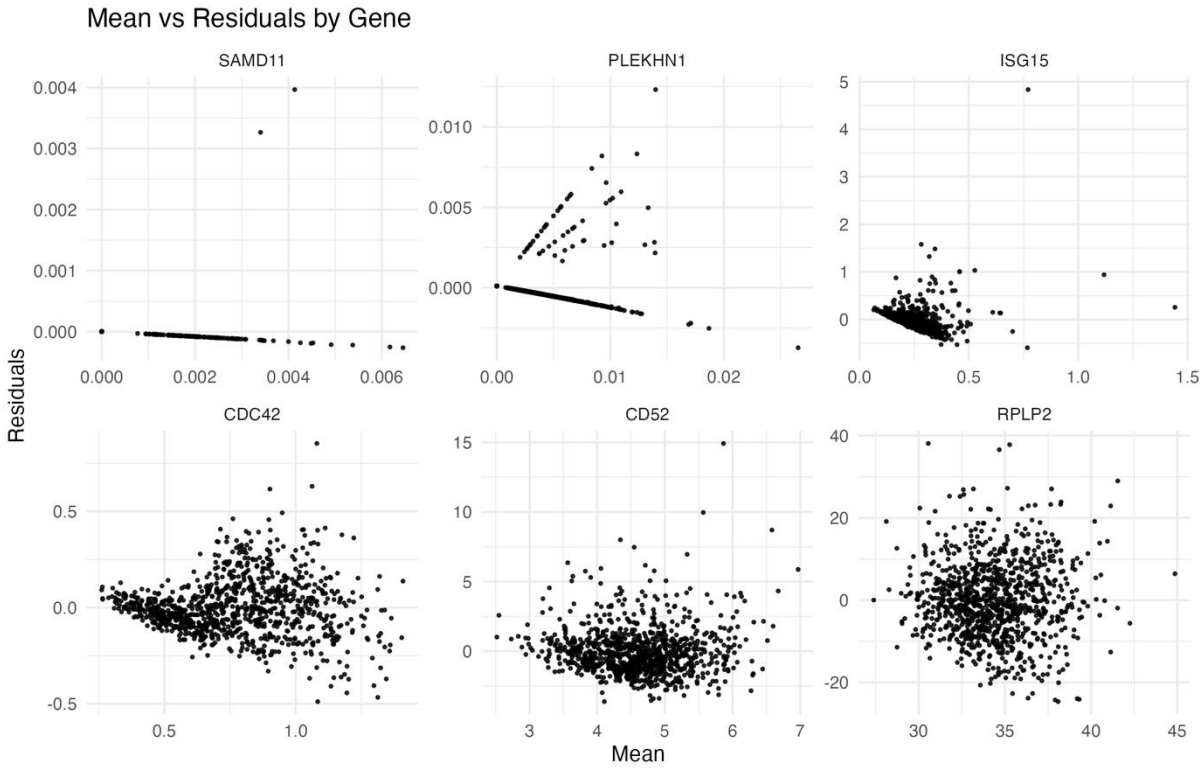
Re: Sorry for overlooking this typo. It should be “vGene that is not identified as eGene”.

The y-axis of Figure 2D (now Supplementary Figure 1D) is the proportion, and each cell type has a different absolute number of significant eGene/vGene. We chose to show the proportion rather than the absolute number due to the large difference between cell types (as seen in Figure 2E). We realised this could be confusing, so we revised the y-axis as the absolute number of genes and annotated the actual number in each bar. We also moved this figure to **Supplementary Figure 1** per the reviewer’s suggestion.

8. In Line 182 and following: “This lack of overlap validates that testing or such residual-eQTL is not valid for identifying true veQTL.” How does this show that the method doesn’t identify veQTL? This seems to simply show that two reasonable sounding approaches get very different answers, which highlights the problem for

identifying veQTL reliably. How can the authors assert what are “true” veQTL? That would seem to require some sort of simulation/power analysis.

Re: Thank you for identifying this issue, and we apologise for the confusion from our statement in the original submission. When we derived the residuals by regressing the mean from the variance for each gene, we found that the residuals showed a spurious quadratic relationship with the mean. After some investigation, we determined that this is because the count data of many genes follow a negative binomial distribution. The relationship between mean and variance can be formulated as $\sigma^2 = \mu + \theta \cdot \mu^2$. Thus, when we regress the mean out of variance, the residuals would contain a term that is linear to $\mu^2$, which explains the observed quadratic relationship. We agree that the original statement could be confusing without showing the relevant plots. We have added several examples (genes with different expression levels) of such spurious quadratic relationships as **Supplementary Figure 7** and revised this sentence (lines 178-179).




9. The results section has substantial discussion of potential mechanisms that are not tested and may be more suitable for the discussion.

Re: We have moved some of the discussion in the **Results** section to the **Discussion** section.

10. The analysis in Figure 5 seems like a nice way to test a mechanistic hypothesis; however, it is still somewhat unclear what is being plotted in Figure 5. It would seem that Figure 5A is that transcription factor expression is affected by the eQTL in a way that is correlated with RPS26, and the beta estimate reflects the similar effect of the QTL on both genes (the TF and RPS26). Then in Figure 5B, the correlations are broken down by the genotype of the individuals. If that is the case, then it is confusing why several of the transcription factor effects in B seem to cluster near 0. Four outlier points were also removed, but from which graph, and to illustrate what?

Re: We would like to clarify that the Figure 5A (now Figure 4A) is not the transcription factor expression affected by the eQTL, but how much an eQTL affect the co-expression strength between TF and *RPS26*. For example, in CD4_NC cells, the purple round dot indicates the co-eQTL effect of rs1131017 on the co-expression strength between *RBM39* and *RPS26*. This value is a positive number, which is why in the zoom-in plot in panel B, we see G allele shows an increasing effect on the co-expression strength. In Figure 5B (now Figure 4B), the y-axis represents the standardised co-expression (measured by Spearman's correlation across cells per individual) between *RPS26* and the TF gene. This is why the values are clustered near 0. We have emphasised that the values have been standardised in the legend (page #27).

Minor comments:
1. Line 48: parentheses with nothing inside

Re: Sorry, this was a formatting issue when converting docx to PDF in the submission system. We have double-checked it and make sure the parentheses are removed.

2. Line 61 should be 'a' SNP's genotype

Re: Updated.

3. Line 89 doesn't need 'the' before gene expression

Re: Updated.

4. Line 164 "the latter group is merely above the significant threshold, suggesting this group of signals are more likely to be random noises due to arbitrary threshold". The nature of the argument here is not entirely clear; it makes sense to indicate the q-values are close to 0.05 for vGenes without significant eGenes, but it is less clear what the 'random noises' part of the sentence is trying to suggest. Please reword for clarity.

Re: Thank you for the suggestion. We were trying to convey the message that the vGenes without significant eGenes are showing weak evidence of the association. If the mean effect does not induce the genetic effect on the intra-individual variability, we expect to see some vGenes to have significant p-values even if they are not eGenes. We have revised the wording of this sentence.

5. Line 175 “verifying that mean veQTLs can explain most effects”. Is there is some switching between nomenclature happening here? Earlier sentences referred to sc-eQTL and sc-veQTL, but is this now switching back to bulk experiments? Since veQTL refer to effects on variance, I believe the sentence should be indicate that eQTLs (effects on the mean) can explain most of the effects of veQTLs.

Re: To clarify, our claim is “verifying that most veQTLs can be explained by mean effects”, which means the veQTL is significant because of the mean effect. This is similar to the reviewer’s interpretation that eQTLs (effects on the mean) can explain most of the effects of veQTLs. We have updated this sentence.

6. Line 211: No explanation of what a dGene is (this appears to be the first time the word ‘dGene’ this is mentioned).
Re: We have added the definition of what a dGene is after this line.

7. Line 213: Do all deQTLs have a corresponding veQTL? deQTL also needs to be defined (only eQTL and veQTL have been defined)
Re: Of the 55 deQTLs we identified, 31 have $p$-value < 5 x $10^{-8}$ in the veQTL association test. The deQTL has been defined in line 124-126 in the last paragraph of the Introduction.

8. Line 241: “Most deQTLs are in intergenic or intronic regions” – are these deQTL assigned to a ‘dGene’? How does this assignment work? Earlier it was mentioned that cis-SNPs were analyzed to find deQTLs; what is the rationale for excluding potential trans-SNPs?
Re: Each deQTL has a corresponding dGene (as shown in **Table 2**) since the definition of deQTL is an SNP significantly associated with a gene's dispersion level. The functional annotation of the deQTLs is assigned by ANNOVAR (Wang et al. *Nucleic Acids Res*. 2010) built in the FUMA platform (Watanabe et al. *Nat Communs*. 2017). We did not exclude potential trans-SNPs but put trans-analysis in another section (lines 307-331) and used trans-regulation as a potential explanation of the detected deQTLs.

9. Line 268: ‘which’ shouldn’t be italicized

Re: Thank you. This has been updated.

10. Figure 3: How many points are plotted in Figure 3A (i.e. how many SNP-gene pair tests are done for each cell type)? The color-coding by cell type and faceting by cell type are redundant.

Re: There are 162,062 points in total in Figure 3A (now Figure 2A), and for each cell type, it ranges from 8,234 (CD4_SOX4) to 15,433 (CD4_NC). We acknowledge that the colour coding by cell type and facets is redundant, but we keep this for consistent colour schemes across the paper.

11. What is the relationship between Figure 3A and B? It might be helpful to identify the dGenes plotted in B on the plots in Figure 3A by color-coding significant genes.

Re: Thank you for the suggestion. We have highlighted the dGenes in Figure 3A (now Figure 2A). The combination of the two panels will help understand how those sc-deQTL behaves when estimating their effects as sc-eQTL or sc-veQTL in the corresponding cell types. We also provide more detailed information of this comparison in the Supplementary Table 3.

12. Line 317-318: The parenthetical comment “but it is only 0 vs 1 so it is not a large difference” is not really needed.

Re: We have removed the parenthetical comment and rephrased the sentence as “The comparison also showed that no dGene has more independent *cis*-veQTLs than *cis*-eQTLs”

13. Line 395 which refers to Figure 5 references co-deQTL but this is not addressed in the figure.

Re: We apologise for the typos in this section. It should be “co-eQTL” rather than “co-deQTL”. We have updated the term and highlighted the two occurrences in yellow.

14. Lines 459-461: What simulations are being referred to here, and how were they assessed for accuracy?

Re: The simulation is described in **Supplementary Note 4** and referred to in lines 610-622 in the **Methods** section. The simulation is to assess the performance of Cox-Reid adjusted MLE for dispersion parameters, and we conducted a simulation with different numbers of cells ($M$), mean ($\mu$), and dispersion ($\theta$).

Reviewer #3 (Remarks to the Author):

Summary:

Building a more complete picture of how DNA sequence variants influence gene expression (eQTLs) is a critical challenge for 21st century biology. Single-cell RNA-sequencing technology has recently been applied to address this challenge and it is now accepted that many eQTLs are cell-type specific. Most of these studies have focused on identifying genetic variants that influence mean expression levels, however, there has been a long-standing interest in identifying variants that influence the variance in gene expression in addition to or independently of the mean. Studies addressing this question have focused on the genetic effects on inter-individual variance in gene expression, which in many cases reflect unmodelled GxG and GxE. Xue et al. took a different approach and instead focused on how genetic variants influence the intra-individual variance in gene expression. The authors took advantage of the fact scRNA-sequencing generates many more cells than individuals and identified thousands of genetic variants associated with intra-individual variance of gene expression (sc-veQTL and sc-deQTLs) in the OneK1K sc-RNA-seq dataset. The authors detected 4,642 sc-veQTLs across all 14 cell types, but found that the vast majority of these sc-veQTLs influence mean expression in a consistent direction and are thus not independent of the mean.

To identify genetic variants that influence the intra-individual variance independently of the mean the authors developed a framework (MEOTIVE) that estimates the effect of variants on the intra-individual dispersion of gene expression (sc-deQTLS). This method uses a negative binomial framework to estimate the dispersion and mean per person, and then, using regression, they map genetic variants associated with the intra-individuals dispersion of each gene. They applied their methodology to the OneK1K data and identified a modest number (55) of sc-deQTLs. The genes influenced by these sc-deQTLs were enriched in biological processes related to viral infection and immune responses with the specific examples of RPS26 and LYZ being explored in detail. Finally, the authors attempted to replicate their results in another study, an effort that was modestly successful even though the replication dataset they used was small relative to the OneK1K data.

The paper is well written and the methodology was described carefully where importantly the authors seem well aware of many of the pitfalls of analyzing sparse count-based data. Overall, the conclusions are well supported by the data, however, I do have some remaining questions about the statistical methodology, and the presentation of the data. Below are my comments in approximate order of importance.

Re: Thank you very much for your time and efforts in reviewing our manuscript. We appreciate the comments and suggestions, which are very constructive and helped us improve the manuscript. We have revised the manuscript and replied to the specific points below.

Major comments

1. Work by Resztak et al. (PMID 35313584) used scRNA-seq data to map variance eQTLs with single-cell RNA-seq and found that even when modelling the inter-individual dispersion in a negative binomial framework there remains a negative linear relationship between the dispersion parameter and the mean (see Fig 6F). This observation was also reported in the seminal DeSeq2 manuscript which was one of the first papers to introduce the negative binomial model for RNA-seq analysis (PMID 25516281, Fig 1). Given this expected relationship between the dispersion parameter and the mean is it possible to conclude that the sc-deQTLs are independent of the mean as the authors claim between lines 203 and 207 when introducing their MEOTIVE approach? Along these lines of reasoning, do any of the 55 deQTLs violate this global relationship between the mean and dispersion parameters.

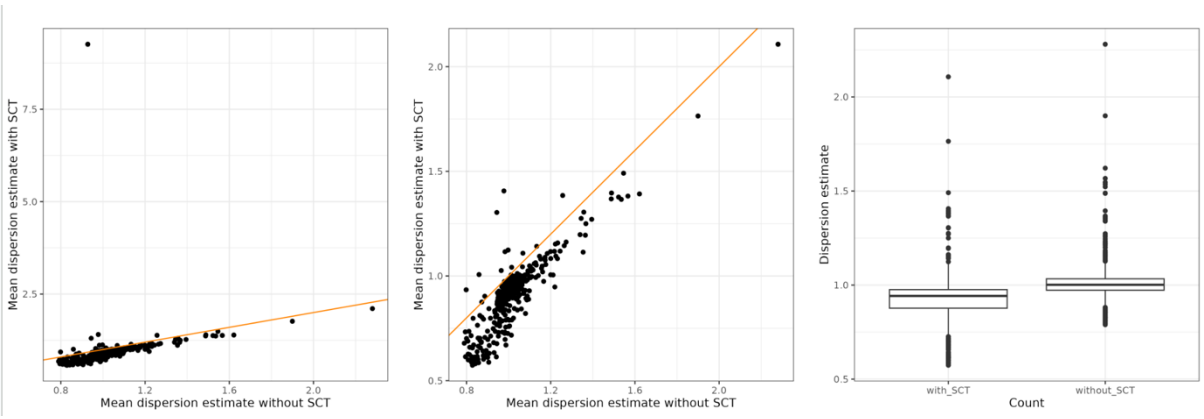
Re: Thank you for raising this point. We acknowledge that there could still be a negative linear relationship between the dispersion estimates and the mean. In the current study, we adopted the simulation-based strategy to rule out the possibility that biased estimates didn't cause this negative relationship. This will ensure the relationship between intra-individual dispersion and mean levels is not a technical artifact signal.

For all 55 sc-deQTLs we detected, 5 of them (with 4 genes: *HNRNPH1*, *HLA-C*, *RPS10*, *RPLP2*) violate this global negative relationship (**Table 2**). We have updated the manuscript with this observation.

2. Based on my reading of the methods, it looks like the authors fed the processed counts from the package 'sctransform' into their per person estimators of dispersion and mean that were calculated with their method MEOTIVE. I understand the motivation behind the choice because the authors wanted an easy way to adjust the counts for batch effects prior to running their per sample estimator of the mean and dispersion of each gene. As far as I understand it, the 'sctransform' method involves fitting a regularized negative binomial model to the original counts and outputting new counts that are adjusted for the batch effects and the UMI count of each cell. I wonder whether this procedure to adjust the counts could result in artificial signals

being observed. One idea I had to overcome this possible issue would be to learn the 'batch' and other parameters across their entire dataset with the 'sctransform' framework' or another regression approach and then fix them as offsets along with a size factor in MEOTIVE.

Re: Thank you for raising this point also. It is something we looked into when developing our analysis approach, but we had not documented this in the original submission. To make sure the estimation for the dispersion parameter is not affected by the 'sctransform', we performed an analysis to compare the dispersion estimates before and after 'sctransform' in CD4 naïve cells (**Supplementary Figure 20**). There is one outlier gene, *MALAT1*, a highly expressed and highly variable gene, but no SNPs were associated with its dispersion levels. The results show that the dispersion estimates highly correlate before and after the SCTransformation (Pearson's cor = 0.82, excluding the outlier). From left panel, we can observe an outlier, which is *MALAT1*. This gene has been discussed in the sctransform paper (Hafemeister & Satija. *Genome Biology*. 2019), and the method can "appropriately modelled technically driven heterogeneity in this gene". This gene is the most expressed but was not significant in the sc-deQTL association test (p-value = 0.919), so it is very unlikely that sctransform would result in artificial signals.



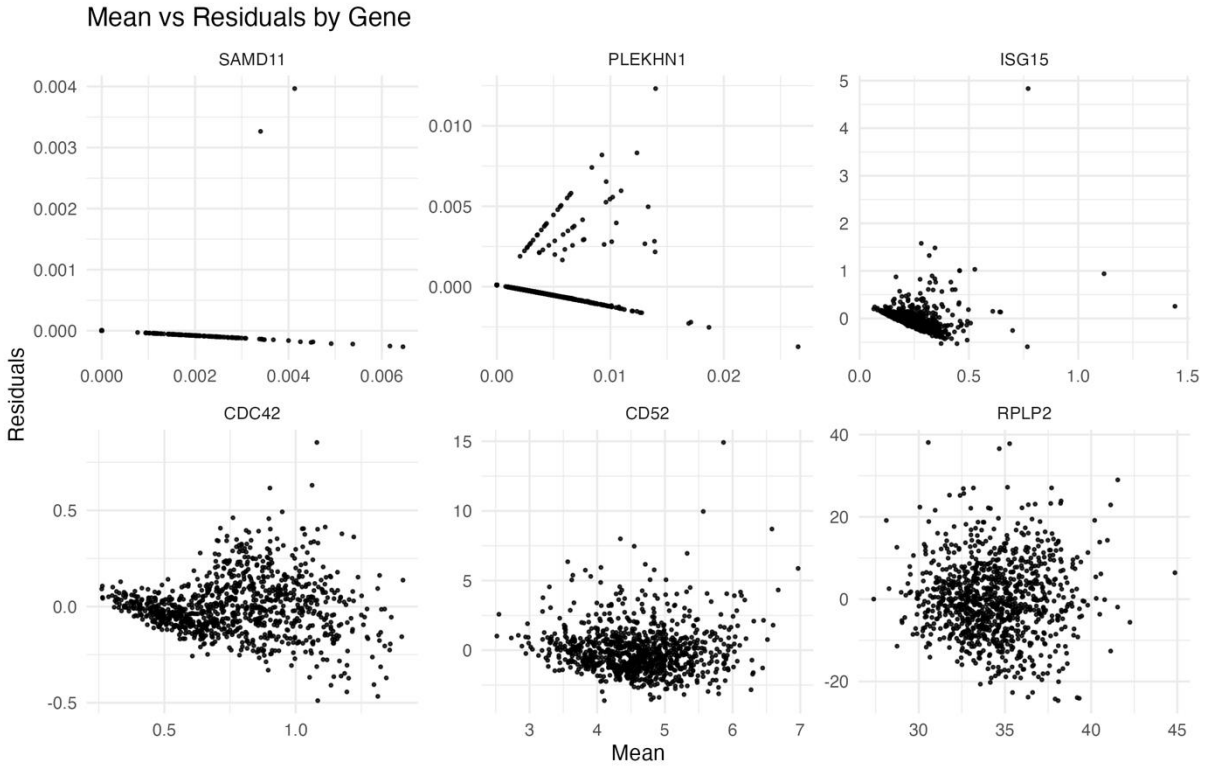

When we ran the sctransform, we already fitted the batch as covariate. Also, the sctransform algorithm assumes the count data follows a negative binomial distribution, and it will learn the dispersion parameter for each gene across all cells. Overall, our detected dGenes will not be affected by the sctransform. We have updated the manuscript (lines 638-642) with these analyses.

3. In the section on the functional analysis of the deQTLs lines 290-303, the authors mention that if they remove all the five genes from the MHC from their deQTL list that the immune enrichment goes away. The MHC locus contains many complex haplotypes. Could mapping issues be influencing the estimates of dispersion in this region?

Re: The current way to quantify single-cell MHC gene expression is based on the cell ranger pipeline provided by 10x genomics. A recent study (Kang et al. Nature Genetics. 2023) showed that performing alignment with the support of individual's unique classical *HLA* alleles can increase the estimated expression by rescuing previously unmapped reads. However, there is no evidence that the remapping would change the dispersion levels, and such change is associated with genetic variants. Thus, we believe that mapping issues are very unlikely to affect our sc-deQTL detection.

On lines 233-236 the authors mention that most dGenes are vGenes, but that there was minimal overlap with residual genes. Do the authors have an explanation for what kind of sources of variance the residual method captures and how that differs from MEOTIVE beyond the small statement they made in the text.

Re: This is a good point! When we estimated the residual of (variance ~ mean) for each gene, we found some distributions of the residuals are skewed, multimodal, and show a spurious relationship with the mean expression levels. This is because 1) mean and variance are highly correlated; 2) the data is sparse. Below are examples of mean-residual relationships for six genes (*SAMD11*, *PLEKHN1*, *ISG15*, *CDC42*, *CD52*, *RPLP2*) with different expression levels in CD4 naïve cells. We can observe spurious relationships for those genes with relatively low expression levels. As we discussed in a comment from Reviewer #1, we found that the residuals contain a term that is linear, the mean-squared, which means the residuals still capture the variance from a function of the mean.



Minor comments

4. In the discussion at lines 443 to 451 the authors propose an explanation for the 'zero-inflated' model in single-cell RNA-seq data. There has been some debate about the use of zero-inflated models, and whether single-cell data is distributed in a zero-inflated negative binomial fashion at all (PMID 31937974). Given the relationship between the dispersion parameter and the mean as discussed above in comment #1, is another explanation that a single negative binomial doesn't model the number of zeros correctly when the mean differences between genotypes groups are large as in the case of rs1131017-RPS26. Rather, one needs to model the differences in dispersion as a function of genotype in addition to the mean, as is possible in the R package 'glmmTMB' with the dispformula option. I would suggest analyzing the data using this package, but it may be that fitting a negative binomial regression model with the amount of data used in this study is likely impractical.

Re: We agree with the reviewer's interpretation. For those genes with large genetic effects on the expression levels, it will be challenging to understand the distribution type without modelling the genotype information. In our analysis, we modelled the count distribution of each individual so that the large mean difference between genotype groups would not impact our intra-individual dispersion estimate.

We also wanted to clarify that the model in 'glmmTMB' is unsuitable for our sc-deQTL detection because we are detecting the genetic variants associated with the average changes in intra-individual dispersion levels. This differs from the 'glmmTMB model, which directly models the count data. In the discussion (lines 443-451), we tried to convey that a seemingly zero-inflated NB distribution could be explained by a mixture of NB distributions with different mean values. This will serve as an additional caution for researchers when analysing such datasets.

5. I found some of the figures difficult to follow and the captions were a little sparse on details that were necessary to understand the plots. More detailed comments follow.
a. In the subplot of Figure 1A titled "sc-EQTL" the genotype is on the Y-axis, but for the rest of the subplots the genotype is the X-axis. The caption of Figure 1A is lacking some detail.

Re: Thanks for raising up this. We have restructured Figure 1 to make it more reader-friendly. We regrouped the panel based on bulk and single-cell approach and also added a panel describing sc-deQTL.

b. In Figure 1B, one of the 3 individuals has Poisson distributed gene expression. It makes sense to me this could be modelled in the negative binomial framework but

might not be obvious to a reader. Another of the individuals looks like they have normally distributed gene expression, a process which isn't going to generate counts. I found that a little confusing. I wonder if all three of those examples could be drawn from negative binomial distributions to make it easier to understand.

Re: Thanks for pointing out the confusion here. We intended to show a distribution that looks similar to a normal distribution but is generated from a negative binomial distribution with a larger mean and smaller dispersion. We have updated the figure with three negative binomial distributions (1) small mean and zero dispersion; (2) small mean and large dispersion (3) large mean and small dispersion. The axis titles are also updated accordingly. We have updated the Figure 1B accordingly.

c. In general the axis titles could be clearer. For example, in Figure 2A 'rho_s' should be converted to s, and in Figure 1, the underscores should be removed from the axis labels.

Re: We have updated the axis titles in Figures 1 & 2 to make it them clearer. As mentioned in the above comments, we have moved Figure 2 to the Supplementary Figure 1 and replaced with a new one explaining the relationship between different dispersion indices.

Reviewer's Comments:

Reviewer #1 (Remarks to the Author)

The authors have performed an impressive amount of new analyses and done an admirable job of addressing reviewer comments. I have a few comments remaining.

1. While the authors did add compelling text and analysis describing the shortcomings of variance, VMR, and CV as measures of expression variability, the manuscript still does not convincingly argue that the proposed dispersion measurement does any better. In particular, it is stated that these metrics have high overlap with eGenes and (especially for variance) strong correlation with eQTL effect sizes. However, no direct comparison is made to dGenes and deQTLs, and I would like to see evidence that dGenes perform meaningfully better at these metrics. Based on the data presented it certainly looks like they do not: a majority of dGenes are also eGenes and deQTL effect sizes are very strongly negatively correlated with eQTL effect sizes. In the current state of the manuscript, I don't see a convincing argument that the MEOTIVE framework is really capturing an effect of dispersion independent of mean expression level in a way that veQTL analysis does not. This is not helped by the fact that every specific dGene discussed in the Results section (RPS26, LYZ, IL32, GNLY, GZMH, IFITM2, HLA-B) is also seen in Table 2 to be a highly significant eGene with an inverted direction of effect. I encourage the authors to focus on the minority of dGenes that are not also eGenes, and see if it is possible to identify an effect that is clearly independent of mean expression level.

Re:

We thank the reviewer for the helpful comment and suggestion. First, we added additional analyses and focus on dGenes that are not eGenes: *KLRB1* (CD4_ET), *CCL3* (NK), and *LGALS1* (NK). *KLRB1* is a cell-type marker for CD4 effector memory or central memory cells. This gene encodes a killer-like receptor and is upregulated in CD4+ T cells from primary Sjögren's syndrome patients (Zhang et al., Int Immuno pharmacol. 2024). CCL3 is C-C motif chemokine ligand 3 and is rapidly upregulated during influenza infection (Ruben Ferrero et al. Front. Immunol. 2022). *LGALS1* encodes galectin-1, which plays a key role in establishing and maintaining immune tolerance at the maternal–fetal interface by directly acting on NK cells (Novák et al., Semin Immunopathol. 2025). We have added these functional annotations in #line 292-298

Second, the fact that a majority of dGenes are also eGenes and deQTL effect sizes are strongly negatively correlated with eQTL effect sizes is more likely to be driven by the nature (distribution) of the single-cell count data. We have now shown that, even though the dispersion estimate is unbiased (**Sup Fig. 16**), there is still a negative trend between the mean and dispersion estimates. We believe this is due to the distribution of single-cell count data, in which many intra-individual distributions follow either the Poisson or negative binomial distributions – results shown in **Sup Fig. 17**. We have included text covering this topic in the discussion. As we are unable to resolve this confounding directly, we have addressed this more clearly in the text and highlighted examples of the dGenes that are not eGenes.

In addition, we would like to clarify that the MEOTIVE framework is not only helpful for identifying genetic effects on intra-individual expression variability but also for assessing the underlying distributional type of the features. The parameters (mean and dispersion) it estimates will improve our understanding of intra-individual gene expression patterns across cells and help build more sophisticated models.

2: On a similar note, I am still not convinced that the rs1131017-RPS26 analysis required the sc-deQTL approach, and I did not find any additions to the text or figures that address this. The authors argue in their response to my comment that the hypothesis does not arise without the deQTL analysis, but it is not clear to me why. Isn't the dispersion-reducing C allele also strongly associated with an increase in mean expression (as shown in Table 2 and Figure 3), and wouldn't this be discovered by a straightforward sc-eQTL analysis with no reference to variance or dispersion? Which of the analyses described are specific to the hypothesis that the C allele increases disease risk by reducing dispersion, and inconsistent with the hypothesis that the C allele increases disease risk by increasing mean expression? The co-eQTL analysis in particular actually seems to show that these TFs control mean expression rather than dispersion, as it measures a relationship between TF expression and RPS26 expression, NOT a relationship between TF expression and dispersion of RPS26 expression. I would like to see a clear statement in the text stating which part of the hypothesis requires the sc-deQTL analysis to formulate, and which part of the analysis suggests that the effect is due to dispersion rather than mean expression.

All my other comments were adequately addressed and I am very pleased in general with the responses. I have a few more minor suggestions:

Re:

We thank the reviewer for further discussion on this SNP-gene link. We agree that additional text clarifying this observation is needed. We want to clarify that the sc-deQTL approach is not required to identify the association between rs1131017 and RPS26. As the reviewer pointed out, even based on the eQTL results, we can detect

this association. We highlighted this example because the sc-deQTL signal helps us dissect the underlying count distribution model within genotype groups, and we observed that the allelic change drives the negative binomial distribution towards a Poisson-like distribution. This further observation (rather than a typical boxplot for visualising eQTL signals) would not have been made if we had only looked at eQTL or veQTL patterns.

We have added the following additional text (#lines 393-400) to clearly state that only the dissection of the underlying count model requires the sc-deQTL analysis.

“We observe that the UMI count distribution for individuals in the CC genotype group mostly follows a negative binomial distribution, while the UMI count distributions for CG and GG individuals follow a Poisson-like distribution, which is statistically equivalent to a negative binomial distribution with an extremely tiny dispersion parameter and can be modelled by a negative binomial model. This suggests that even for the same gene, genetic effects can affect the distribution type across cells within an individual. This observation would only be seen if we dissect the count distribution per genotype group when performing sc-deQTL analysis.”

3. The definition of the theta parameter is much clearer and easier to follow. I identified one statement that still needs to be rephrased to make this clearer: ll.211-217: "For example, when the expression level is low for a specific genotype group, the count data distribution is more left-skewed, driving the dispersion to be higher. For heterozygous individuals, the distribution approaches Gaussian; thus, the dispersion becomes much smaller, and so on for individuals with alternative homozygous alleles (see an example of RPS26 below, Figure 3). This feature results in a negative relationship between intra-individual variance and intra-individual dispersion for many genes with significant sc-veQTLs and sc-deQTLs." This is all correct and makes sense, but it is presenting dispersion as a function of the mean, when your argument is that dispersion should be considered independent of the mean while the variance is not. I suggest rephrasing this to clarify your argument that the low-mean, left-skewed distribution has deflated variance (not inflated dispersion), and that the dispersion is a more accurate measure of variability because it accounts for the way the shape of the distribution changes with the mean.

Re:

We thank the reviewer for this insightful comment and suggestion. We have added the following sentences (lines 214-223) to clarify our definition of the theta parameter.

“This is because the distribution type of gene expression count per genotype group varies. For example, when the expression level is low for a specific genotype group,

the distribution of the count data is more left-skewed, leading to deflated variance estimates. For heterozygous individuals, the distribution approaches Gaussian; thus, the dispersion is much smaller, and the same holds for individuals with alternative homozygous alleles (see the example of *RPS26* below, Figure 3). This feature results in a negative relationship between intra-individual variance and intra-individual dispersion for many genes with significant sc-veQTLs and sc-deQTLs. In such cases, dispersion is a more reliable metric of variability in this context because it explicitly accommodates the mean-dependent changes in the shape of the underlying count distribution."

4. I am now well convinced that the immune connection is real, but I think it would improve the argument to show some p-values in the Results section where it is discussed ("Functional annotations of sc-deQTLs and dGenes"). Most straightforwardly, I would like to see some p-values for the gene enrichment analysis in the main text, rather than just referring to the supplemental table. I also am very convinced by the analysis described in response to my comment, where a random subset of genes were selected and found not to have the same level of enrichment for immune, inflammatory, and viral genes, and I suggest using this to report a permutation-based p-value: take 100 or 1000 random samples of 34 genes and report the fraction of them containing at least 31 genes annotated as immune, inflammatory, or viral as the p-value.

Re:

We have now added corresponding *p*-values to the gene enrichment analyses in the main text (# lines 301-310).

Regarding the permutation-based results, we randomly selected 34 eGenes 100 times, and none of the samples returned significant enrichment in either immune response or viral infection pathways. So, the permutation-based *p*-value = 1.

5. I think the new framing of how distributions are discussed is much better and makes the statistical framework much easier to follow. I identified one place where the previous framing is preserved and is now even more confusing: ll.381-385, "We observe that the UMI count distribution for individuals in the CC genotype group mostly follows a negative binomial distribution. In contrast, the UMI count distributions for CG and GG individuals follow a Poisson distribution. This suggests that even for the same gene, genetic effects could impact the distribution type across the cells within an individual." Later, in the Discussion, it is described more clearly that the "Poisson-distributed" cells are a population with low dispersion, leading to the appearance of a zero-inflated / mixture distribution, when each individual population of cells can

actually be modeled with negative binomial distributions with their own dispersion parameters. I suggest rephrasing this statement in the Results section to make a similar point.

Re: Thanks for the suggestion. We have rephrased this sentence to keep a consistent description for the distribution types.

Now the line #393-397 reads
"We observe that the UMI count distribution for individuals in the CC genotype group mostly follows a negative binomial distribution, while the UMI count distributions for CG and GG individuals follow a Poisson-like distribution, which is statistically equivalent to a negative binomial distribution with an extremely tiny dispersion parameter and can be modelled by a negative binomial model."

Reviewer #3 (Remarks to the Author)

Xue et al have responded sufficiently to our comments. We have no further concerns with their manuscript.

Re: We thank the reviewer for their constructive comments which helped us significantly improve the manuscript.

Reviewer #4 (Remarks to the Author):

This review is focused only on addressing the author comments in response to Reviewer #2's initial review. Overall, I think this is a solid manuscript. Broadly speaking, the concerns of reviewer #2 are addressed well.

I have one major and one minor point related to Reviewer #2's comments that I feel are not fully addressed.

Re: We thank the reviewer for taking time to read the original comments from Reviewer #2 and providing additional feedback and comments.

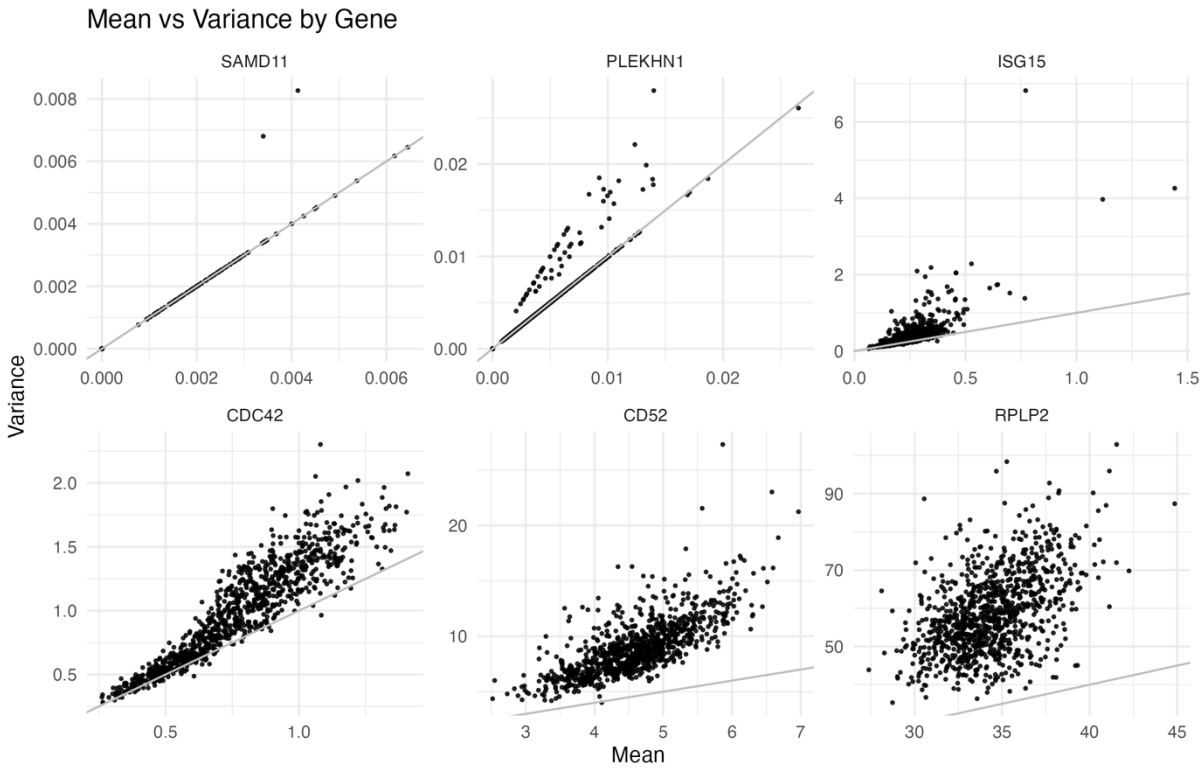
Major:
Reviewer #2 correctly identified that the sentence "This lack of overlap validates that testing or such residual-eQTL is not valid for identifying true veQTL" is problematic. The authors' response does not seem to fully address the concern. They explain that they found residuals with a spurious quadratic relationship with the mean. They

provide some example plots of these relationships that don't look very convincing to me — the plots look like they are either dominated by a few outlier points, or perhaps linear and noisy. More generally, why is this proof that these genes are not in fact veQTL? I'm willing to accept that's the case, it's just that the authors don't seem to have provided any explicit reasoning or rationale for such.

Re: We thank you for the continuing discussion on the residual-eQTL, which we agree needed additional explanation. We wanted to clarify that the few outlier points presented do not just dominate the example plots due to linearity and noise, but rather to a systematic bias.

Those dots look like "outliers" in *SAMD11* and *PLEKHN11,* but they are actually due to the modelling of the relationship between intra-individual mean and variance. As mentioned in line #187-189, the residuals from the regression of variance out of the mean may have a spurious quadratic relationship with the mean. When the expression level is extremely low like *SAMD11* and *PLEKHN11*, there will be a lot of zeros in the regression model and most intra-individual mean will equal to the variance, and only a few data intra-individual variance will be slightly higher than the intra-individual mean.

Below are the scatter plots of the mean-variance relationship for those genes.




When we look at the pattern for *SAMD11* and *PLEKHN11,* it is very clear that the outliers in the mean vs residual plot are those with slightly higher variance. However, those individual with equal mean and variance (dots on the diagonal line) is expected to have no dispersion. When we regressed the mean out of variance, most residuals

do not capture the “excessive variability” but show an artificial negative relationship. If we take a closer look at this **Sup Fig.7**, this artificial negative relationship is more severe in lowly expression gene but still observable in moderately expressed genes like *ISG15* and *CDC42*. This is why we argued that residual-eQTL is not a valid approach for identifying true veQTL.

We have included this new plot in the Sup Fig. 7 as well.

Minor:
The authors should change the y-axis label in Fig 4A from “beta estimates” to something more informative. Reviewer #2 misinterpreted it and I did too.

Re: Thanks for the suggestion. We have now updated the y-axis label to “beta estimates of co-expression between *RPS26* and its TF”.

**REVIEWER COMMENTS**

Reviewer #1 (Remarks to the Author):

The authors have once gain put a lot of work into responding to comments from all reviewers, including mine. I am mostly satisfied with the responses to my more minor comments, with one minor correction: the way empirical p-values are usually expressed would give you $p < 0.01$ for the gene set permutation test, not $p = 1$.

Thank you for the suggestion and for flagging this point. We apologise for the oversight. We have excluded insignificant results with p-values > 0.05 and corrected the text (lines 448-451) to clarify the thresholds for *p*-values, which now reads:

*These mechanisms are unlikely to reflect bias from analysing immune cells, as permutation-based enrichment analyses using 34 randomly selected eGenes (100 permutations) showed no enrichment for immune response or viral infection pathways (permuted P > 0.05).*

However, I do not think the answers to either of my major comments are adequate, and this leaves me with some concerns about the novelty and general usefulness of this approach. At the risk of repeating myself, I essentially have two major concerns left unaddressed:

1. The authors emphasize the difference between Poisson and negative binomial distributions and the ability of this method to distinguish between these distributions. I do not understand why this distinction is interesting or important, since the Poisson distribution is just a negative binomial distribution with dispersion set to 0. Choosing between a Poisson distribution and a negative binomial distribution can be rephrased as choosing between a negative binomial distribution with dispersion = 0 and a negative binomial distribution with dispersion > 0. I cannot think of any reason why choosing between negative binomial distributions with dispersion = 0 and dispersion > 0 would be meaningfully different from a standard negative binomial model that allows dispersion to be 0, and as far as I can tell none is given in the paper. If this is the main application of this method, I do not think it is of general interest.

**Re:** We thank the reviewer for raising this important point. We agree that, formally, the Poisson distribution is a special case of the negative binomial (NB) distribution corresponding to dispersion equal to zero. In principle, therefore, the question could be framed as testing whether the NB dispersion parameter lies at the boundary of the parameter space ($\theta = 0$) versus $\theta > 0$.

However, the motivation for explicitly distinguishing these regimes arises from the statistical behaviour of dispersion estimation in sparse single-cell count data. In practice, dispersion estimation under the NB model is known to be unstable when the true dispersion lies near the boundary of the parameter space, particularly for low mean counts and limited sample sizes (in our case, the number of cells per donor). Under these conditions, maximum likelihood estimators of dispersion are biased upward and can exhibit substantial variability. Consequently, genes that are effectively Poisson-distributed (i.e. variance determined entirely by the mean) are frequently assigned non-zero dispersion estimates, even when the true generating process has $\theta = 0$.

This behaviour is especially pronounced in single-cell RNA-seq data, where many genes are expressed at very low levels and the number of observations per individual can vary substantially. In this regime,

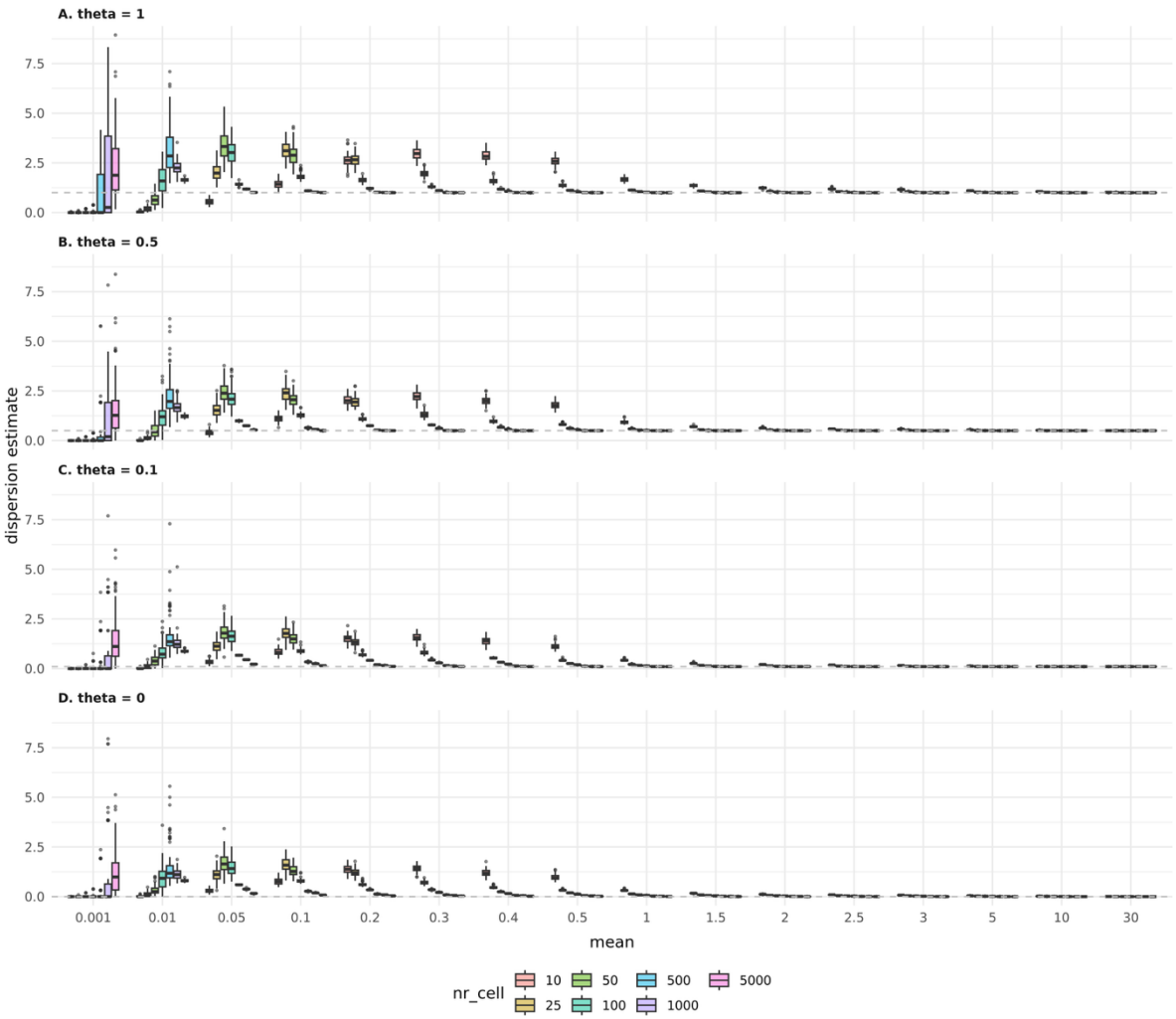
a standard NB model that simply allows $\theta \geq 0$ does not reliably distinguish between Poisson-like and genuinely overdispersed count processes, because the dispersion estimator rarely collapses exactly to zero and tends to absorb sampling noise as apparent overdispersion.

This distinction is important for the objective of our method, which is to identify genetic variants associated with intra-individual variability of gene expression rather than mean effects. If Poisson-like genes are incorrectly assigned non-zero dispersion, the variance component becomes confounded with estimation noise, obscuring true biological variability signals.

To illustrate this point empirically, we have added a simulation analysis in the revised manuscript (**Supplementary Fig. 17**). In this simulation, counts were generated under a negative binomial model with dispersion fixed at $\theta = 0$ (i.e. a Poisson-generating process). We varied the number of cells (M) and the mean expression level ($\mu$), and estimated dispersion using the Cox–Reid adjusted maximum likelihood estimator implemented in glmGamPoi::glm_gp(). For each parameter setting, we simulated 1,000 genes across M cells and repeated the procedure 100 times. Details of the methodological approach are covered in **Supplementary Note 4**.

As shown in the new figure (**Supplementary Fig. 17, copied below)**, even when the true dispersion is zero, the estimated dispersion is systematically inflated for lowly expressed genes, with the magnitude of inflation depending strongly on both the mean and the number of cells. These results demonstrate that, in realistic single-cell settings, the practical distinction between Poisson-like and overdispersed count behaviour is non-trivial and cannot be reliably inferred from standard NB dispersion estimates alone.

We have clarified this motivation in the revised manuscript (lines 614-641) and added the simulation results to better illustrate the statistical behaviour of dispersion estimation in this regime.




**Supplementary Figure 17. Behaviour of CR-MLE dispersion estimates across simulated expression regimes.**

Dispersion estimates obtained using the Cox–Reid adjusted maximum likelihood estimator (CR-MLE) under simulated count data. For each simulation, gene expression counts were generated under a negative binomial model with specified mean expression (x-axis) and dispersion parameter (panel label). The y-axis shows the estimated dispersion values obtained from CR-MLE. Colours indicate different numbers of cells per individual (sample size). The horizontal grey dashed line denotes the true dispersion parameter used to generate the data. Results illustrate that dispersion estimates become increasingly inflated and variable at low mean expression levels and small sample sizes, particularly when the true dispersion lies near zero.

2. The primary motivation given for using the dispersion statistic over other more commonly used statistics is that it should in principle be independent of the mean. However, in real data dispersion appears to be highly correlated with the mean. The authors point to the difference between Poisson and negative binomial distributions to explain this, but I reiterate that a Poisson distribution is just a negative binomial distribution with dispersion 0. The observation that negative binomial distributions with dispersion = 0 have systematically lower means than negative binomial distributions with dispersion > 0 is an expected consequence of the observation that dispersion is strongly correlated with mean, not

an explanation of it. As currently presented, the theta statistic does not appear to do a good job of separating mean effects from dispersion effects, and the authors do not present a convincing argument that it does.

Re: We thank the reviewer for raising this important point regarding the relationship between mean expression and dispersion. We agree that, under the negative binomial (NB) model, the mean and dispersion are not theoretically independent in empirical datasets, and that estimator behaviour in sparse count regimes can induce additional coupling between these quantities. Our objective in using the dispersion statistic was therefore not to claim strict theoretical independence from the mean, but rather to identify genetic effects on expression variability beyond those attributable to mean expression differences.

As discussed in the revised manuscript, dispersion estimation is particularly unstable for lowly expressed genes, where the true dispersion may lie near the boundary of the parameter space ($\theta \approx 0$). In this regime, CR-MLE dispersion estimates tend to be biased upward and highly variable, which can induce apparent relationships between mean and dispersion that reflect estimator behaviour rather than biological signal. To mitigate this effect, we used simulation-based analyses to determine expression thresholds under which dispersion estimates reliably recover the true parameter (**Supplementary Note 4**; **Supplementary Figures 17–18**). Only genes exceeding these cell-type-specific mean thresholds were retained for downstream analysis, which substantially reduces mean-dependent inflation of dispersion estimates.

After applying these filters, we still observe a negative relationship between mean expression and estimated dispersion. We believe that, at least in part, this reflects genuine biological heterogeneity rather than a purely technical artefact, as the analysis is restricted to genes within regimes where dispersion estimation is stable according to the simulations.

Nevertheless, to directly address the reviewer's concern, we performed an additional analysis in which the mean–dispersion relationship was explicitly removed by regressing mean expression out of the dispersion estimates prior to QTL mapping. While we were initially cautious about this approach because it imposes a specific functional form on the mean–dispersion relationship and may introduce additional noise (similar to issues observed with mean–variance residual approaches), we included it here as a robustness check.

Encouragingly, the majority of deQTL signals remained significant after this adjustment. Of the 55 deQTL pairs (34 unique genes) identified in the primary analysis, 32 pairs (24 unique genes) remain significant (q-value < 0.05) when using mean-adjusted dispersion estimates. These results indicate that the detected deQTL signals are not simply driven by genetic effects on mean expression. The results of this analysis have been added to the revised manuscript (lines #640-660) and are reported in a new **Supplementary Table 9**.

Reviewer #4 (Remarks to the Author):

Overall, I feel this article is a nice contribution to the field.

I think the authors should consider that their response to one aspect of my previous critique may be unclear. I'll explain the details below. I previously said:

Reviewer #2 correctly identified that the sentence “This lack of overlap validates that testing or such residual-eQTL is not valid for identifying true veQTL” is problematic.

Reviewer #2 had originally said:

How does this show that the method doesn’t identify veQTL? This seems to simply show that two reasonable sounding approaches get very different answers, which highlights the problem for identifying veQTL reliably. How can the authors assert what are “true” veQTL?

In my first review (of the initial resubmission), I noted:

The authors’ response does not seem to fully address the concern. They explain that they found residuals with a spurious quadratic relationship with the mean. They provide some example plots of these relationships that don’t look very convincing to me — the plots look like they are either dominated by a few outlier points, or perhaps linear and noisy. More generally, why is this proof that these genes are not in fact veQTL?

Now, in the new submission and in their rebuttal:

The authors explain that the outliers in the mean vs variance plot are those with higher variance. And that, looking at the mean vs residuals plots, there is generally a negative relationship between mean and residuals. I don’t understand how this shows a quadratic relationship. Shouldn’t the authors be asserting that their mean vs residuals plots show that the data don’t fit model assumptions for the residuals mapping model? And that this model mis-specification in fact would be consistent with the residuals-QTL method not being well suited to identifying true veQTL?

Re: We thank the reviewer for this thoughtful clarification and agree that the key issue is whether the residual-based approach is appropriately specified for modelling variability in count data. We apologise if our previous explanation was unclear.

Our intention was not to claim that the plots alone demonstrate that these loci are definitively not veQTL, but rather that the residual-based approach is likely to be model-misspecified for count data whose variance depends on the mean, which can generate misleading signals.

In **Supplementary Fig. 7**, the mean–variance relationship for genes such as *PLEKHN1* shows that most individuals lie close to the Poisson expectation (variance ≈ mean), while a subset of individuals exhibits higher variance consistent with negative binomial overdispersion. This mixture of regimes is expected for count data and reflects the well-known mean–variance relationship for the negative binomial distribution:

$$\mathrm{Var}_i = \mu_i + \theta_i \mu_i^2$$

where $\theta_i = 0$ corresponds to Poisson-like behaviour and $\theta_i > 0$ corresponds to overdispersed counts. Importantly, this relationship is quadratic in $\mu$.

The residual-based veQTL approach instead assumes a linear relationship between variance and mean, typically modelling

$$\mathrm{Var} = \hat{\alpha} + \hat{\beta}\mu.$$

When variance is regressed on the mean under this linear model, the residual for individual *i* becomes

$$e_i = \theta_i \mu_i^2 + (1 - \hat{\beta})\mu_i - \hat{\alpha}.$$

Thus, even when the underlying variability follows the standard negative binomial mean–variance relationship, the residuals will retain a systematic quadratic dependence on μ. In practice, this manifests as structured patterns in the mean–residual plots (e.g. the *CDC42* panel in **Supplementary Fig. 7**), where residuals increase with $\mu^2$ for individuals with overdispersed counts. The visibility of this curvature depends on the mixture of Poisson-like and overdispersed observations, which explains why the pattern is more pronounced for some genes than others.

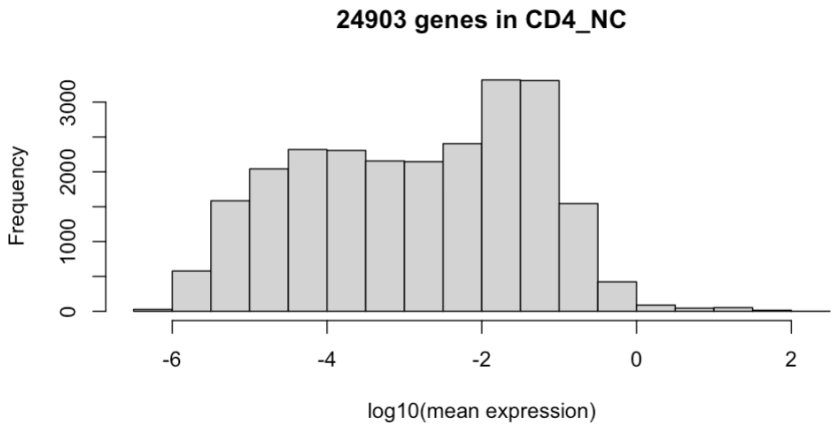
This behaviour reflects model mis-specification of the residual approach rather than a small number of influential outliers. In particular, when the true mean–variance relationship is quadratic (as expected for NB-like count data), forcing a linear correction of variance on mean can generate structured residual patterns that do not correspond to genuine biological variability.

Because a large fraction of genes in our dataset are expressed at very low levels (19,237 of 24,903 genes have mean expression < 0.03), many genes follow Poisson-like behaviour with variance ≈ mean. In this regime, the residual approach tends to produce residual patterns dominated by technical noise rather than meaningful variability signals.

For these reasons, we conclude that the residual-based veQTL framework is not well suited for identifying variability QTLs in sparse single-cell count data, where the mean–variance relationship is inherently nonlinear. We have revised the manuscript to clarify that our interpretation is based on model mis-specification of the residual approach, rather than asserting that the genes identified by that method cannot represent true veQTL (lines 179-183).

We thank the reviewer for the constructive discussion on the interpretation of residual-QTL. We would like to clarify that in **Supplementary Fig. 7**, the mean vs residual relationship is not simply dominated by a few outlier points, or perhaps linear and noisy.

If we take a close look at *PLEKHN1* panel, we will see most individuals have equal mean and variance. However, because the nature of the count distribution, there are a number of individuals have higher variance than mean (i.e., dots above the diagonal line). When we regress the mean out of variance, we observed that most data points were on a negative linear line (**lower panel of Supp Fig. 7**), but they actually are those distribution fit more for Poisson distribution. This means the residual did not capture the extra variations of the dataset, but generated large technical noise for most individuals. Given the expression level range, the majority of genes (19,237/24,903 with mean < 0.03) will fit the pattern of *SAMD11* or *PLEKHN1*, rather than other four genes (see histogram of mean expression per gene below).




As for the quadratic relationship between mean and residuals, our reasoning is as follows. For a mixture of Poisson and NB donors, the true mean-variance relationship is:

$$var_i = \mu_i + \theta_i \mu_i^2$$

Where $\theta_i = 0$ for Poisson donors and $\theta_i > 0$ for NB donors. This is inherently quadratic in $\mu$.

When we regress the mean out of the variance, we are estimating

$$var = \hat{\alpha} + \hat{\beta}\mu$$

So, if we write out the residual for donor i, it will be

$$e_i = \theta_i \mu_i^2 + (1 - \hat{\beta})\mu - \hat{\alpha}$$

This is where the quadratic relationship emerges. When we examine the gene *CDC42* in Supplementary Figure 7, we observe such a quadratic relationship. Depending on how many donors have a Poisson or NB distribution, the quadratic relationship could be more or less obvious. This is consistent with what the reviewer mentioned about the model mis-specification.

Thus, we concluded that the residuals-QTL method is not a suitable approach for identifying true veQTL.

**REVIEWER COMMENTS**

Reviewer #4 (Remarks to the Author):

I am satisfied by the reviewers' revisions to their manuscript.

We thank the reviewer for their time and expertise in reviewing our paper.

---

Here are a few comments I had in response to the authors' rebuttal comments to Reviewer #1.

**Comment 1**. In response to Reviewer #1 point 1: In the authors' revised MS, lines 629-632 state:
"In sparse count regimes typical of single-cell data, dispersion estimates are known to be biased upward for lowly expressed genes, which can lead to apparent overdispersion even when the underlying count process is effectively Poisson"
Please clarify if "known to be biased upward" is a result derived from your analyses (e.g. Supplementary Fig 17) or also known elsewhere in the literature (if so, include references).

**Response 1:** Thank you for flagging this, and we apologise for the use of the term 'known', which is not technically accurate. Instead, we have edited this sentence to read: "In sparse count regimes typical of single-cell data, we observe an upward bias in dispersion estimates for lowly expressed genes"

**Comment 2**. In the authors' rebuttal to Reviewer #1 point 2:
"We agree that, under the negative binomial (NB) model, the mean and dispersion are not theoretically independent in empirical datasets"
Unclear -- contradiction between theoretical and empirical here...
And "Our objective in using the dispersion statistic was ... to identify genetic effects on expression variability beyond those attributable to mean expression differences."
But the authors just said that the mean and dispersion are not independent in empirical datasets. So why are they explicitly using dispersion stat to identify effects on variability beyond the mean? They need to clarify or perhaps further hedge their claims.
The additional analysis where dispersion estimates are conditioned on mean prior to mapping is good. It is encouraging that many of their deQTLs remain significant. However, >40% of their deQTLs are not significant in this analysis. I think their unconditioned analysis includes both variants that affect expression variability in a mean-independent fashion as well as some that increase variability at least partially through an effect on mean (the latter accounting for incomplete overlap with mean-conditioned analysis, although power is certainly a component as well).

**Response 2:** Thank you, we agree, and have added the following test to the discussion to ensure that we recognise the nuance. "MEOTIVE addresses the mean-variance relationship by modelling dispersion directly and applying additional filtering and sensitivity analyses.

However, because mean expression and dispersion are intrinsically linked under count-based models such as the negative binomial, complete separation of mean and variability effects is not always possible in empirical data. Consistent with this, while most deQTLs remained significant after conditioning on mean expression, a subset did not, suggesting that some signals may be partially mediated through effects on mean expression."

**Comment 3**: Overall, the work is good, but I think the authors need to be a little more nuanced to clarify that their method is not completely able to disentangle effects on mean vs dispersion in all cases.

**Response 3**: Thank you again, and we agree with the need to be nuanced on the ability to disentangle effects on the mean and dispersion. As per above, we have addressed this with edits to the text.

**Comment 4**: (Remarks on code availability)
Code appears to be structured clearly and somewhat straightforward to run. I noticed that there are some places where this group's specific absolute paths to the data are hardcoded. It would be ideal if the authors could e.g. have these stored in variables that clearly indicate how they would be modified in the case of someone working with data downloaded from Zenodo into $DIR.

**Response 4:** We have modified the code so that users can include any paths as variable inputs.