nature portfolio

Peer Review File

eQTL in diseased colon tissue identifies potential target genes associated with IBD

Corresponding Author: Dr Terrence Furey

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)
This manuscript covers a comprehensive set of computational approaches to investigate the cis-eQTLs in context of IBD tissues, compared to non-IBD samples. A higher enrichment of GWAS loci has been identified in eQTLs in IBD, with novel, interesting examples such as ABO and TNFRSF14 showing different regulatory patterns across datasets. This study reveals the importance of using disease relevant tissue to uncover the IBD-associated genetic effect on transcription levels. It is appreciated that the authors performed nice analysis to approve this conclusion but I do have extra concerns. Moreover, the manuscript is well-written in general, but with some difficulties to follow for certain result sections.

1. The authors compared the eQTL signals and eGenes across different datasets, followed by GWAS loci enrichment analysis, cross-colocalization and mash analysis, providing statistical evidence of different eQTLs between IBD and non-IBD tissues. My major concern is how the data heterogeneity affects the results. For example, the different tissue sampling locations, the different bioinformatic methods used in UNC, GTEx and BarcUVa (this can be reflected by the lower replication rate between GTEx and BarcUVa), as well as the power issue. I understand all the analysis of non-IBD eQTLs is based on summary statistics, but I would suggest to perform power calculation and comparison across datasets to investigate to what extent, the unique UNC eQTLs have been influenced by detection power.
2. My second concern is still about the conclusion “IBD associated eQTL unique to diseased tissue had distinct regulatory and functional characteristics with increased effect sizes”. I notice that the IBD tissues are collected from non-inflamed ascending colon. Ideally, the authors should incorporate paired inflamed and non-inflamed tissues. It would be fantastic to see a pure “increased effect size” (announced by authors) from IBD non-inflamed to IBD inflamed tissues. Then I would be more convinced that the disease plays a role here, not confounded by others, e.g. data resource or allele frequency.
3. The whole manuscript reads well, but some results sections are difficult to follow. For example, the comparison of eQTL effect size. What does it mean in relation to biology, e.g. inflammation or disease severity? It would be helpful for readers if the authors link these observations to clinical phenotype and add some biological explanation.
4. The different colocalization patterns of SLC45A1 is only presented in figure 1 but not mentioned in the main text.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This paper presents an eQTL analysis in 252 colon samples from IBD patients. The results are compared to published colon (non-diseased) eQTL and IBD GWAS hits. Overall I think the paper is clearly written and provides interesting analyses.

Major comments:

The only analysis that I think the paper would really benefit from is a meta analysis of the 3 colon datasets. Given all the evidence that there is substantial overlap, it seems natural to meta-analyse and use that set of summary stats to coloc with IBD GWAS? This also helps address challenges in interpreting whether individual loci are seen in each dataset, since they have different sample sizes, and therefore statistical power. For example, the number of unique colocs in Fig 2b goes

directly proportionally to sample size of the 3 eQTL datasets.

I think eQTL coloc with a GWAS hit is not proof that the GWAS hit acts via that gene/mechanism. There are now many papers detailing the abundant coloc/overlap of eQTLs and GWAS hits, which I think makes it clear not all are causal. This line of argument starts, e.g. on line 64 "with over 40% of GWAS loci explained by eQTL colocalizations". It continues in the results & discussion, and although I think these are very valuable findings, they need to be interpreted in light of other evidence, and so I think shouldn't be taken to be definitively identifying effector genes.

I think the results specific to the UNC dataset are too readily interpreted as being driven by the fact that the tissue samples came from IBD patients. (e.g. line 158, "suggests unique regulatory activity in the presence of disease uncovers alternate eQTL which may reveal new insights into IBD-associated loci beyond our present knowledge using only non-IBD eQTL."). And some of the differences may well be due to that reason, but there are many other differences: age distribution of the donors, ethnicities, where the biopsies were taken, how the RNA seq was done, etc. There's likely at least some part driven by disease state, given the additional enrichment of IBD GWAS colocs, but there are also pieces of evidence in the paper that argue against this, most notably that the non-coloc rate between the two non-disease colon datasets are _lower_ than the disease/non-disease concordance (line 144).

Minor comments:

- On line 127, "We found that 15% of lead eVariant-eGene pairs in UNC eQTL were also lead pairs in GTEx and/or BarcUVa eQTL". This number is not really meaningful, because lead variants can differ due to LD, even in truly coloc'd signals. It's addressed in the next paragraph, but then why even show this number, since it could potentially be misleading (a casual glance suggests there is not a lot of sharing).

- are the Y axes of Fig 1c -log10(p)? If so this locus seems pretty weakly associated in all three datasets, raising the possibility that the lack of coloc could be coincidence.

- I like the use of "Carolina blue" in Figure 2.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thanks for all the additional analysis and thoughtful discussion.

(Remarks on code availability)
Code has been well-organized.

Reviewer #2

(Remarks to the Author)
Both reviews started by raising the same major concern: that interpretation of many of the results is affected by differing statistical power of the 3 eQTL datasets considered. I asked for a meta-analysis, and the other reviewer asked for power calculations. The authors declined to perform either analysis. I don't entirely follow the authors' explanation for why a meta-analysis would be such a "considerable effort". As they do colocs (both between eQTL studies and with IBD studies), surely they have already addressed issues of matching "genome builds and gene annotations"? I can appreciate there will be some complexities of non-overlapping genes etc, but there are 15,000 commonly tested genes, and that's a quite powerful analysis.

As for why this is relevant when "we were less motivated by maximizing the number of IBD-associated eQTL than by showing that analyses in IBD patient tissue revealed unique IBD eQTL.", I would say a key question is how many are truly "unique IBD eQTL" versus "eQTL with bigger/more detectable effect sizes". I think the latter is biologically more plausible (IBD is not a homgenous disease, certainly not across UC & CD, so each group of "healthy" and "diseased" is a mix of people with regulatory programs activated to different extents). If you combine Barcuva and Gtex, how many of the UNC eQTL do you find? What if you had 10,000 healthy specimens? I think the quantitative nature of the added value of disease tissue is central to how broadly applicable the inferences of this study are. The TNFRSF14 signal is striking, but is it the only one of that magnitude?

The additional analyses that were performed do not, in my view, directly address either set of questions, and therefore a substantial gap in the paper remains. I do agree those new analyses provide greater support for the authors' view that eQTL from tissue from individuals with the disease in question do, in some loci, have stronger effect sizes, and are therefore a

valuable addition to interpreting GWAS hits.

My other comments were addressed, and in particular with respect to the second part of my major comment 2 (about coloc rates between gtex and barcuva, vs unc) I agree the authors are right, that's almost certainly about differential power. But I think that actually supports the points above that differential power in the 3 datasets is relevant to many results in the paper!

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have indicated that an individual-level re-analysis of the three constituent datasets is too much effort:

"To perform a rigorous meta-analysis requires having access to and re-processing individual-level data from each of the cohorts. This would require significant effort to: harmonize genotype arrays, impute missing genotypes, align and quantify RNA-seq reads for each individual based on a single gene annotation, perform additional covariate analyses such as genotype PCA to control for population structure and hidden covariate analysis via PEER or RUVseq, as well as perform the appropriate model testing. Each of these steps and analyses are non-trivial and time-consuming."

I agree! This is not what I was asking them to do; I was asking for a summary-statistic level meta-analysis. Which I think the authors have now done for 25 genes (though they have only described it as "rudimentary", and chosen not to describe it in the paper, so it is not entirely clear how the analysis was done).

They have not included those results because "the methodology [is not] the most robust". I find this puzzling, as the assumptions for an inverse-variance weighted summary statistic meta-analysis are exactly the same as those needed for COLOC analysis. Namely that the entity being measured as phenotype (i.e. a gene) is the same in both studies, that population structure and other hidden covariates have been controlled for, and that differential imputation accuracy or genotyping error does not materially affect the association pattern in the region. If these assumptions are not met in the summary statistics, and such a meta-analysis is "not the most robust", then this undermines the whole paper -- the coloc results must overcome all the same objections.

I think the results of the paper strongly suggest this isn't the case, and that therefore a summary statistic meta-analysis of the two non-IBD cohorts is both methodologically sound, and a reasonable request for revision.

I also think the partial analysis the authors have done demonstrates that the point I've been making (and which the other reviewer made) is germane to the overall conclusions of the paper, and so I can't see any a compelling reason not to have included it in the paper.

The authors have made it clear they disagree, so we are, respectfully, at an impasse.

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
To facilitate the interpretation and reproducibility of the findings, I strongly encourage the public release of the summary statistics for the UNC eQTLs.

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This manuscript covers a comprehensive set of computational approaches to investigate the cis-eQTLs in context of IBD tissues, compared to non-IBD samples. A higher enrichment of GWAS loci has been identified in eQTLs in IBD, with novel, interesting examples such as ABO and TNFRSF14 showing different regulatory patterns across datasets. This study reveals the importance of using disease relevant tissue to uncover the IBD-associated genetic effect on transcription levels.

We appreciate that the reviewer understands the importance of this work.

It is appreciated that the authors performed nice analysis to approve this conclusion but I do have extra concerns. Moreover, the manuscript is well-written in general, but with some difficulties to follow for certain result sections.

1. The authors compared the eQTL signals and eGenes across different datasets, followed by GWAS loci enrichment analysis, cross-colocalization and mash analysis, providing statistical evidence of different eQTLs between IBD and non-IBD tissues. My major concern is how the data heterogeneity affects the results. For example, the different tissue sampling locations, the different bioinformatic methods used in UNC, GTEx and BarcUVa (this can be reflected by the lower replication rate between GTEx and BarcUVa), as well as the power issue. I understand all the analysis of non-IBD eQTLs is based on summary statistics, but I would suggest to perform power calculation and comparison across datasets to investigate to what extent, the unique UNC eQTLs have been influenced by detection power.

We agree with the reviewer that inevitably there will be differences between studies conducted by different groups, including those listed, that will influence what eQTL are detected. That being said, we believe that the GTEx and BarcUVa studies used here are the best possible comparisons given currently available data. We have added additional text to the Discussion better acknowledging that these factors may play a role in the differences in detection.

We believe both this concern and the first concern from Reviewer #2 are related and important as they essentially question whether the unique UNC eQTLs are due to disease and not some other factor. We agree that the specific cohorts used by each study will influence what eQTLs are found, both due to numbers of individuals and minor allele frequencies in the cohort. The strong correlations in effect sizes across the three studies (Fig 1b) and the colocalization of eQTL signals between studies (Figure 1c) attest to the high degree of similarity in overall genetic effects on expression. But it is the differences that require more attention, and we have attempted to address this by providing additional evidence that the unique UNC eQTLs are arising due to presence of disease.

The reviewer suggests performing power calculations. We believe this is motivated by the fact that for detection of eQTL with smaller effects, even slight changes in observed effect sizes due to expression measurements and to minor allele frequencies in a population could render it significant in one study but not in another. In general, we believe that all three of the studies are underpowered to detect eQTLs with small effect sizes, and that a significant amount of the variation in eQTL detected are due to these small effect eQTL. Power analyses in general show that for small effect eQTL, the probability of detecting these eQTL decreases with smaller sample sizes (Huang et al., GTEx Consortium) – we argue it is not necessary to explicitly do a power analysis to show this. For two studies even of equal sample size, if they are underpowered, it will not be surprising, then, that they do not detect the same small effect eQTL. But, there should be more consistency in detection for larger effect eQTL.

Thus, this comment inspired us to perform additional analyses to verify this intuition and add evidence that unique UNC eQTL detection is not simply due to small effects and lack of power. To test more directly that small effect sizes drive differences in eQTL detection, we looked at the effect sizes of eQTL found in at least two studies ("shared") vs eQTL that are found in only one ("unique"), before and after applying mash (Figure 5). Recall that the mash analysis is meant to adjust effect sizes in individual studies to allow better comparison across studies. As expected, when looking at all eQTL, the calculated effect sizes for the shared eQTL group are significantly higher than the unique eQTL group in all three studies (see Figure 5b, 5c). This supports that overall, differences in eQTL detection can be driven by smaller effect sizes.

To determine whether this was also driving the unique IBD GWAS colocalizing eQTL found in the UNC IBD patients, we repeated this analysis using just colocalized eQTL. Again, we see that detected shared colocalizing eQTL had significantly higher effect sizes within the GTEx and BarcUVa studies compared to detected unique colocalizing eQTL (Figure 5d, 5e). Importantly, the exception to this was for eQTL found only in UNC IBD patients, where the effect sizes for these unique colocalizing eQTL are higher than the shared colocalizing eQTL.

We believe these analyses are an important addition to the paper and provide strong evidence that small effect sizes and reduced detection due to power do not completely explain why we find some IBD-associated eQTL only in IBD patients. In appreciation of the importance of these results, an expanded description of this analysis is now included at the end of the Results. We believe this supports our assertion that while the differences in GWAS-associated eQTL detected between the two non-IBD studies are likely driven by power dynamics, at least some of those only found in IBD patients are driven by increased effect sizes in the presence of disease.

2. My second concern is still about the conclusion "IBD associated eQTL unique to diseased tissue had distinct regulatory and functional characteristics with increased effect sizes". I notice that the IBD tissues are collected from non-inflamed ascending colon. Ideally, the authors should incorporate paired inflamed and non-inflamed tissues.

It would be fantastic to see a pure “increased effect size” (announced by authors) from IBD non-inflamed to IBD inflamed tissues. Then I would be more convinced that the disease plays a role here, not confounded by others, e.g. data resource or allele frequency.

We appreciate that we did not explain the context of this statement sufficiently. First, we want to point out that as IBD is currently considered an incurable disease, the absence of inflammation at the time of sample collection does not indicate absence of disease. We have included additional text in the Introduction to provide the reader with some additional background on IBD.

More generally, the model we suggest by this statement is that a variant that affects the expression of a gene (i.e. an eQTL) is doing so by altering the transcriptional regulation of that gene (Figure 1d). IBD GWAS loci are primarily in non-coding regions, suggesting many are similarly involved in altering transcriptional regulation (Huang et al.). The size of the effect of an eQTL variant on gene regulation is going to be due, in part, by how much the affected regulatory mechanism, such as the binding of a transcription factor in an accessible chromatin region, is actively regulating expression of a target gene at the time of sample collection.

We assert that the presence of disease changes the overall regulatory program. This can include increasing or decreasing the activity of regulatory elements also used in non-IBD tissue, but additionally using regulatory elements not normally used in the absence of disease. Previous work from our group has shown that the alteration in this gene regulatory program is present even in the absence of active pathological markers of disease (Simon et al., Weiser et al., Keith et al.). This has also been shown in other diseases. For example, Natri et al. pointed out that 21% of disease status eGenes were not differentially expressed between samples from interstitial lung disease patients and healthy controls, but that their analysis does suggest that those eGenes were differentially affected by regulatory loci. Overall, this leads to the “distinct regulatory and functional characteristics with increased effect sizes” that can be seen in the presence of disease.

The exact topic of how active disease affected eQTL in IBD was addressed in a previous publication where “inflammation” eQTL were determined in IBD patients (Hu et al.). They determined that inflammation status can alter the genetic effects for a very small subset of genes (2% of eGenes). Interestingly, they concluded that eQTL in IBD patients that colocalized with IBD GWAS loci were not significantly different from those identified previously in non-IBD patients. Moreover, they determined that inflammation-dependent eQTL were mostly associated with shifts in cell types rather than an increase in an inflammatory gene signature.

We specifically chose here to focus on uninflamed tissue for two reasons: (i) to remove effects of inflammation as a confounding variable when comparing to non-IBD samples; and (ii) to show that presence of active disease is not necessary to show changes due to disease but rather these represented more generally fundamental changes in

expression programs in individuals with disease. We agree we did not provide this context and have added clarifying text throughout the paper, especially in the first section of the Results (see additional comments on this below).

3. The whole manuscript reads well, but some results sections are difficult to follow. For example, the comparison of eQTL effect size. What does it mean in relation to biology, e.g. inflammation or disease severity? It would be helpful for readers if the authors link these observations to clinical phenotype and add some biological explanation.

We appreciate this comment as we agree, it can be difficult to draw the connection between a specific eQTL and disease consequence. This is directly related to the discussion above where we argue presence of disease can alter the regulatory program, changing which regulatory elements are being used or the extent to which they are used for regulating a particular gene. These changes will uncover regulatory processes that are not used, or not used as strongly, in the absence of disease, some of which are additionally affected by genetic variation.

More concretely, consider *TNFRSF14* that has been shown to mediate the immune response in the colon and the absence of which has been linked to accelerated intestinal inflammation (Shui et al., Steinberg et al.). If decreased expression of *TNFRSF14* contributes to an increase in the inflammatory response leading to increased risk of IBD, and if genetic variation contributes to the expression of *TNFRSF14* (i.e. there exists an eQTL), then individuals with one or more variant alleles associated with decreased *TNFRSF14* expression will have higher risk of developing IBD compared to those with only variant alleles associated with higher expression of *TNFRSF14* (Figure 3d, 3e). If the casual variant affects a regulatory element not strongly used (i.e does not have a strong effect on expression of the target gene) under healthy conditions, then it would be difficult to detect in unaffected samples (Figure 5h). If onset of disease (inflammation) activates this regulatory element, and this regulator remains active (has a stronger effect) even in the absence of active disease, then you are better powered to detect this in IBD patients.

We have added text and provided a model for this in Figure 1d to clarify this concept.

4. The different colocalization patterns of SLC45A1 is only presented in figure 1 but not mentioned in the main text.

We apologize for this oversight. We had meant to use SLC45A1 as a representative example of this phenomena, thus we did not explicitly name the gene in the text. We have added this for clarity.

Reviewer #2 (Remarks to the Author):

This paper presents an eQTL analysis in 252 colon samples from IBD patients. The

results are compared to published colon (non-diseased) eQTL and IBD GWAS hits. Overall I think the paper is clearly written and provides interesting analyses.

We appreciate the positive comments about our study.

Major comments:

1) The only analysis that I think the paper would really benefit from is a meta analysis of the 3 colon datasets. Given all the evidence that there is substantial overlap, it seems natural to meta-analyse and use that set of summary stats to coloc with IBD GWAS? This also helps address challenges in interpreting whether individual loci are seen in each dataset, since they have different sample sizes, and therefore statistical power. For example, the number of unique colocs in Fig 2b goes directly proportionally to sample size of the 3 eQTL datasets.

We believe that this comment is motivated by similarly wanting more evidence for our claims of eQTL found in IBD patients only, especially associated with GWAS loci, as was in the first concern for Reviewer #1 above. The primary goal of a meta-analysis is to address the lack of power in individual studies by trying to combine information across multiple cohorts. Meta-analyses do not increase our ability to detect eQTL in any particular study, nor to better understand potentially missed signals in single studies. We agree that this meta-analysis may reveal additional colocalized loci, but these analyses are difficult due to differences across studies, as Reviewer 1 detailed above. To combine information across studies, considerable effort is involved to harmonize gene annotations, genome builds, adjustments for technical variation, etc. This is why most analyses rely on summary statistics calculated for each study.

In this study, we were less motivated by maximizing the number of IBD-associated eQTL than by showing that analyses in IBD patient tissue revealed unique IBD eQTL. Therefore, our analyses, especially the mash analyses, the colocalization across eQTL data, and now the newly included analyses on small effect eQTL, aimed to reveal that these unique eQTL in IBD patients existed. We believe the new analyses on effect sizes described above addresses the issue of how statistical power affects identification of small effect eQTL, but that these new eQTL in IBD patients are not explained simply by small effect sizes.

2) I think eQTL coloc with a GWAS hit is not proof that the GWAS hit acts via that gene/mechanism. There are now many papers detailing the abundant coloc/overlap of eQTLs and GWAS hits, which I think makes it clear not all are causal. This line of argument starts, e.g. on line 64 "with over 40% of GWAS loci explained by eQTL colocalizations". It continues in the results & discussion, and although I think these are very valuable findings, they need to be interpreted in light of other evidence, and so I think shouldn't be taken to be definitively identifying effector genes.

We agree that we have stated this too strongly, and in fact we show examples in this paper where analysis of IBD patients results in different colocalizing eQTL compared to

nonIBD samples. Thus we know well that simply colocalizing does not imply it has been “explained”. We have modified this text to indicate that these colocalizations provide “potential” target genes for GWAS loci.

I think the results specific to the UNC dataset are too readily interpreted as being driven by the fact that the tissue samples came from IBD patients. (e.g. line 158, "suggests unique regulatory activity in the presence of disease uncovers alternate eQTL which may reveal new insights into IBD-associated loci beyond our present knowledge using only non-IBD eQTL."). And some of the differences may well be due to that reason, but there are many other differences: age distribution of the donors, ethnicities, where the biopsies were taken, how the RNA seq was done, etc. There's likely at least some part driven by disease state, given the additional enrichment of IBD GWAS colocs, but there are also pieces of evidence in the paper that argue against this, most notably that the non-coloc rate between the two non-disease colon datasets are _lower_ than the disease/non-disease concordance (line 144).

We believe that these concerns are again similar to comments from Reviewer #1 addressed above. We acknowledge that study-specific characteristics may influence eQTL detection, but we do not feel these fully explain all of the IBD-associated eQTL that we detected, such the highlighted examples of *ABO* and *TNFRS14.* We believe that our additions to the first Results section describing the model by which this could happen more clearly provides the context under which we believe these eQTL may only be found in IBD patients. We agree that this may not be the only reason why eQTL are only found in IBD patients, and we have added text to acknowledge this, as indicated in our response to comment 1 from Reviewer #1. We believe that the reduced rate of colocalization between GTEx and BarCUVa is more due to the fact that given their increased sample sizes, each study were able to detect more unique small effect eQTL that other studies did not. In contrast, since the UNC cohort was the smallest, it was not as powered to detect unique small effect eQTL.

Line 144 points out that through the eQTL-eQTL colocalization, we were able to provide evidence that many eQTL found in GTEx and/or BarcUVa but not in UNC were likely due to reduced power, as this colocalization indicates that a similar eQTL signal is present in the UNC data. It simply was not strong enough to be considered significant.

Minor comments:

- On line 127, "We found that 15% of lead eVariant-eGene pairs in UNC eQTL were also lead pairs in GTEx and/or BarcUVa eQTL". This number is not really meaningful, because lead variants can differ due to LD, even in truly coloc'd signals. It's addressed in the next paragraph, but then why even show this number, since it could potentially be misleading (a casual glance suggests there is not a lot of sharing).

We initially focused on these as they were cases in which unequivocally the same variant was identified as the lead variant by the other studies, and not just considered

the same based on LD. The 15% number was not meant to be of import, but rather we felt concentrating on these first was necessary to establish the concordance of the UNC data analysis with the other studies. We have added text to better explain the purpose of this analysis.

- are the Y axes of Fig 1c -log10(p)? If so this locus seems pretty weakly associated in all three datasets, raising the possibility that the lack of coloc could be coincidence.

The y-axes of (now) Figure 1e are -log10(nominal p-value). While these signals are not extremely strong in all three studies, they were all independently considered significant based on standard multiple hypothesis testing correction and false discovery rate thresholding. Furthermore, there is no LD between the index variants, and the UNC signal is located 300 kb upstream of the GTEx and BarcUVa signals, so it is unlikely these signals would colocalize even with smaller p-values. We have included additional text as well as eQTL summary and coloc statistics pulled from the Supplementary Tables into the main text to provide additional support.

- I like the use of "Carolina blue" in Figure 2.

Of course! Go Heels!

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Thanks for all the additional analysis and thoughtful discussion.

We appreciate that the reviewer acknowledges the importance of the revisions in addressing their previous concerns.

Reviewer #2 (Remarks to the Author):

I don't entirely follow the authors' explanation for why a meta-analysis would be such a "considerable effort". As they do colocs (both between eQTL studies and with IBD studies), surely they have already addressed issues of matching "genome builds and gene annotations"?

We agree we did not adequately elaborate on this concern. To perform a rigorous meta-analysis requires having access to and re-processing individual-level data from each of the cohorts. This would require significant effort to: harmonize genotype arrays, impute missing genotypes, align and quantify RNA-seq reads for each individual based on a single gene annotation, perform additional covariate analyses such as genotype PCA to control for population structure and hidden covariate analysis via PEER or RUVseq, as well as perform the appropriate model testing. Each of these steps and analyses are non-trivial and time-consuming.

In contrast, colocalization analyses can be performed using already available summary statistics from the published eQTL and GWAS studies, where variant coordinates can be lifted over to the appropriate genome build for analysis. This did not require the extensive re-processing and harmonization of the data required of a robust meta-analysis.

I would say a key question is how many are truly "unique IBD eQTL" versus "eQTL with bigger/more detectable effect sizes". I think the latter is biologically more plausible (IBD is not a homgenous disease, certainly not across UC & CD, so each group of "healthy" and "diseased" is a mix of people with regulatory programs activated to different extents).

We do not disagree with the reviewer on this point, and given the samples sizes, we acknowledge it is impossible to state with high confidence that an eQTL is “unique to IBD”. It would likely take thousands if not tens of thousands of samples to answer this with great certainty, as the reviewer implies. But we were careful not to claim that a eQTL was unique to IBD as a whole, but rather we state that the eQTL was unique to

the analysis of the IBD-patient cohort, even though with less patients it was less powered than the other two cohorts. Text in the manuscript has been modified to make this distinction clearer to the readers. We believe that this is still a critical advance as it supports that scaling up studies such as ours in IBD patients will have much better power to comprehensively detect relevant disease associations than attempting to do this in non-affected individuals.

If you combine Barcuva and Gtex, how many of the UNC eQTL do you find? What if you had 10,000 healthy specimens?

In the initial review, the reviewer requested a “meta analysis of the 3 colon datasets”, not just the BarcUVa and GTEx data. We argued that this was not an appropriate analysis, as including all 3 would not answer the question of whether the eQTL found only in the IBD dataset was only due to power since the IBD data would be included and would confound this analysis. This new question, which suggests using just the BarcUVA and GTEx data, is more appropriate.

As requested, we performed a rudimentary meta-analysis of the non-IBD data sets, focused on the set of 25 eGenes from our UNC cohort that were found to colocalize with GWAS, as these eQTL are likely the most relevant to Reviewer 2’s concerns. For this analysis, we required data for each gene to be available in both non-IBD studies. Based on the available summary statistic information, we were able to individually meta-analyze 19/25 genes. Focusing on the signals we detected using our IBD cohort, 7/19 meta-analyzed signals did not change when combining information across GTEx and BarcUVa (i.e. no change in effect size). An additional 5/19 meta-analyzed signals resulted in larger p-values compared to their individual studies, suggesting either significant heterogeneity between studies or that there is less evidence of the eQTL effect than initially indicated. Of the remaining 7 signals we tested, 3 continued to show stronger effect sizes and smaller p-values in the UNC cohort, despite less than 1/3 sample size (n_{UNC} = 252, n_{meta} = 813). Only 2/4 remaining meta-analyzed signals appeared to colocalize with our UNC cohort eQTL, suggesting that some signals observed in our cohort may be detected in an unaffected population given sufficiently large sample sizes. Despite this, none of the 19 meta-analyzed signals colocalized with GWAS loci (3.09e-56 ≤ coloc H4 shared signal posterior probabilities ≤ 0.343).

Importantly, the eQTL we highlighted in the manuscript (*ABO, TNFRSF14*) still did not colocalize with the meta-analyzed non-IBD signals for these genes (coloc H3 distinct signal posterior probabilities: *ABO* = 1.0; *TNFRSF14* = 0.913). In summary, these results suggest that performing a meta-analysis would not significantly alter the importance of our work in aiding our interpretation of GWAS loci. We argue that extending this meta-analysis approach to all commonly tested genes would require a significant amount of effort, as each gene must be meta-analyzed separately.

Furthermore, we warrant this unnecessary effort as the primary goal of this study was to focus on IBD-relevant eQTL and not colon eQTL as a whole. We have chosen not to include the above analyses in the paper due to the methodology not being the most robust, despite the results providing support for our findings.

The additional analyses that were performed do not, in my view, directly address either set of questions, and therefore a substantial gap in the paper remains. I do agree those new analyses provide greater support for the authors' view that eQTL from tissue from individuals with the disease in question do, in some loci, have stronger effect sizes, and are therefore a valuable addition to interpreting GWAS hits.

We disagree, as does Reviewer #1, that the previous new analyses did not address at least their previous questions. We hope that the analyses provided in this revision and in this response will be found to address all remaining concerns.

RESPONSE TO REVIEWER REQUESTS

We completely agree with the reviewers that a robust meta-analysis of the two non-IBD cohorts would provide additional support for our claims that some eQTL identified only in the IBD patients are due to altered regulatory usage in this IBD cohort. As detailed below, we have now included in the manuscript our meta-analysis of the 25 genes only found in the IBD cohort that were found to colocalize with GWAS loci. We provide our reasoning why a larger scale study is not feasible, and we hope that the reviewers agree that our limited meta-analysis that targets those genes most relevant to our study is sufficient to address their concerns.

Reviewer #2: They have not included those results because "the methodology [is not] the most robust". I find this puzzling, as the assumptions for an inverse-variance weighted summary statistic meta-analysis are exactly the same as those needed for COLOC analysis. Namely that the entity being measured as phenotype (i.e. a gene) is the same in both studies, that population structure and other hidden covariates have been controlled for, and that differential imputation accuracy or genotyping error does not materially affect the association pattern in the region. If these assumptions are not met in the summary statistics, and such a meta-analysis is "not the most robust", then this undermines the whole paper -- the coloc results must overcome all the same objections.

I think the results of the paper strongly suggest this isn't the case, and that therefore a summary statistic meta-analysis of the two non-IBD cohorts is both methodologically sound, and a reasonable request for revision.

I also think the partial analysis the authors have done demonstrates that the point I've been making (and which the other reviewer made) is germane to the overall conclusions of the paper, and so I can't see any a compelling reason not to have included it in the paper.

Reviewer #1: In the primary manuscript, the major concern was the statistical issue of comparing an in-house dataset with two public datasets. Such heterogeneity inevitably affects the identification of unique IBD eQTLs. However, in the revised version, the authors have provided additional evidence showing that at least the IBD-associated eQTLs are unlikely to be solely due to dataset differences. This strengthens the most intriguing aspect of the study—that IBD samples may provide added value in identifying potential causal SNPs from GWAS mediated by eQTL effects.

That said, one limitation remains. The extent to which general eQTLs (beyond IBD-associated loci) are influenced by differences between datasets is still unresolved, and cannot be fully addressed due to the study design (in-house vs. public datasets). Nevertheless, Reviewer 2's suggestion of conducting a meta-analysis of two public non-IBD eQTL datasets represents a meaningful step to reduce false positives or missed signals (e.g., 2 of 19 meta signals in non-IBD are colocalized with IBD, from the revised response letter).

Therefore, I support Reviewer 2's suggestion. Incorporating this meta-analysis would strengthen the findings, particularly in the section from lines 132–218, which discusses general colon eQTLs.

We appreciate that Reviewer #2 feels that this analysis is highly relevant to the overall conclusions we are making. We also appreciate that Reviewer #1 felt that the additional evidence included in our previous revision strengthened our assertation that using IBD samples is important in identifying casual SNPs.

We have now included our eQTL meta-analysis in the manuscript and described the methods we used to perform this meta-analysis (see new text below for Results [lines 279-290], Discussion [lines 566-582], and Methods [lines 697-702] sections). Briefly, we used the software METAL, which was designed to perform meta-analyses for GWAS studies using summary statistics. METAL is very efficient for GWAS because there is a single trait: the GWAS trait of interest. In eQTL studies, though, the expression of each gene is a trait of interest, and thus a meta-analysis needs to be run individually for each gene when using METAL. To our knowledge, a robust method or software specifically for efficient meta-analyses of eQTL data using summary statistics has not been previously developed. Given the similarities of meta-analyses between GWAS and eQTL, we felt this was an appropriate solution.

For meta-analyses, the goal is to determine whether combining information from multiple studies provides more power to detect significant associations and the resulting effect size of these associations. The latter is important for performing colocalization, which is necessary to determine if the meta-analyzed locus now reflects the same pattern of association as in GWAS. Therefore, we used the inverse variance approach implemented in METAL, which weights the effect sizes for each study by their standard error. This approach does allow us to generate new summary statistics that can be interpreted by coloc.

However, there are some limitations to this approach, which we feel makes it less appropriate for performing a genome-wide analysis. Running METAL separately for expression of each gene means that p-values cannot be properly adjusted in a genome-wide manner. METAL does have an option to perform a genomic control correction by calculating the inflation of test statistics for each individual study as well as meta-analyzed results, however, this requires full genomic information and cannot be applied to individual loci and is strongly discouraged by the developers. Also, since we are only meta-analyzing 2 studies, METAL is not as accurate for cases where a variant is present in only one study. METAL will report summary statistics for these variants, where the effect size and standard error is unchanged, but because of the inverse weighting, the p-value tends to become smaller than the original study. Thus, we excluded these cases. Due to these limitations, we feel it is best to focus on our targeted results that are informative and provide additional evidence for these eQTL being uniquely colocalized within the IBD patient cohort, given the data that is available.

Results [lines 279-290]: “We employed a cross-colocalization strategy to determine whether these signals were also detected in colon tissue in the absence of disease, even if the eQTL signals were not prioritized for colocalization based on our initial LD lookup. When eGenes associated with colocalizing UNC eQTL signals were tested in the GTEx and BarcUVa data, they did not colocalize with GWAS loci, despite significant eQTL reported by GTEx and/or BarcUVa for 18 out of these 25 eGenes. To interrogate this further, we meta-analyzed 19 out of the 25 non-IBD eQTL signals for which we had summary statistic information available in both studies (**Supplementary Table 11**). Importantly, even with an effective sample size over 200% greater than our UNC study, none of the meta-analyzed signals colocalized with IBD GWAS loci ($3.09e\text{-}56 \leq PP4 \leq 0.343$; **Supplementary Table 12**). This suggests it is plausible that regulatory programs for some genes may be altered due to the presence of disease and that we are able to more effectively detect these effects by leveraging diseased tissue.”

Discussion [lines 566-582]: “Comparing and combining independent eQTL studies is also challenging due to limited access to individual-level data and reliance on summary data. Meta-analyses can strengthen power to detect small or weak effect eQTL that may be difficult to detect in a single study alone, but they fail to provide additional information of study-unique eQTL signals. We performed a small, summary statistics-based meta-analysis of non-IBD signals on a subset of genes to determine whether increasing the effective sample size in a non-diseased population would bolster the power to detect some of the unique signals that were uncovered in our UNC IBD cohort. As expected, our meta-analysis showed that while the signals detected by GTEx and BarcUVa are robust, they seem to differ from the driving signals we detected in our UNC cohort. Further work is needed to more comprehensively meta-analyze individual-level data to more conclusively determine to what extent signals may be shared across diseases states in the colon. In addition to this, we also performed multiple parallel statistical comparisons to support our conclusions, and we have shown that by using multiple independent eQTL studies, we can provide robust evidence for new genetic hypotheses. We acknowledge that to more definitively determine whether disease state alters regulatory mechanisms for some genes resulting in new eQTL or increased eQTL effect sizes, we need better powered studies in both diseased and non-diseased tissue, ideally within the same study.”

Methods [lines 697-702]: “Non-IBD eQTL meta-analysis
We used METAL (v2020-05-05) to meta-analyzed a subset of eQTL signals, requiring summary statistics to be available for each gene in both GTEx and BarcUVa78. Briefly, we separated out full summary statistics for each gene of interest for each study. We performed a fixed effects inverse variance weighted meta-analysis on each gene of interest individually. We then performed colocalizations using the meta-analyzed summary statistics.”

Response to Reviewers

Reviewer #1:

To facilitate the interpretation and reproducibility of the findings, I strongly encourage the public release of the summary statistics for the UNC eQTLs.

Our UNC eQTL summary statistics are available through dbGaP.