nature portfolio

Peer Review File

Large-scale causal discovery using interventional data sheds light on gene network structure in k562 cells

Corresponding Author: Dr Brielin Brown

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

Version 0:

Reviewer comments:

Reviewer #4

(Remarks to the Author)
The manuscript by Brown et al. presents a novel approach learning causal networks based on interventional data which builds on inverse sparse regression (inspre). The procedure first estimates the pairwise average causal effects between all features. As the true effects matrix cannot be determined exactly, the authors calculate the noisy, sparse inverse effect matrix instead, from which a causal graph can be estimated. The authors benchmarked their method against existing methods for causal discovery from observational and interventional data, showing superior performance in terms of speed, precision, Hamming distance, and other measures.
The authors then applied their method on interventional data of a perturb-seq experiment to improve the understanding of complex-trait genetics and the underlying architecture of complex regulatory networks and pathways. For this they reconstructed an acyclic network based 788 genes from genome-wide perturb-seq experiments on K562 cells. From the network, the authors related shortest paths to effects, identifying central nodes and their importance in the network with albeit borderline significance. Correlating genes associated with blood traits and shortest paths and upstream effects but failed to do so.

The authors mention two reasons for the limited insight above. On the one hand, the K562 cancer cell line might behave differently compared to normal tissue. On the other hand, they econclude that gene networks are highly connected such that perturbation effects ramify through the whole network. Therefore, it is hard to impossible to assign core pathways and genes to individual perturbations, which limits the insights from their inspre algorithm.

Overall, the paper is well-written and concise. It deals with a timely and important subject on the inference of large-scale gene regulatory networks from interventional data. The inspre algorithm is – undoubtedly - an interesting step forward in this direction. However, its biological usefulness still needs to be shown. In the manuscript, the authors have confirmed prior biological knowledge, but new insights are missing. This should be added through another application of their method to biological data.

Major points

- The reason why the authors picked exactly 788 genes from the perturb-seq data should be explained in more detail. How would the results change, if the authors used fewer or more genes? Are the findings a result of the limited number of genes used for network inference or are they a general feature of the omnigene model, small world and scale-free property of networks?
- The authors benchmarked the inspre algorithm using a network of 50 nodes. From the the perturb-seq data they inferred a network with 788 nodes, which is one order of magnitude larger. Given the fact that the authors could not infer new biological insights from the large network, it would be advisable to infer a biological network having a similar number of nodes, say from a smaller chemical compound or siRNA screen? Would it be possible to use genes in well-defined pathways and their perturbation as a proof of principle for a smaller inspre-inferred network?
- The GRN inferred from the perturb-seq data using the inspre algorithm confirms existing knowledge on the small world property of GRNs and regulatory properties of hub genes, such as the anti-proportionality between dosage sensitivity and centrality. However, it remains open whether the network also shows scale-free properties concerning connectivity. This should be shown/discussed.

- The authors find that shortest-path measures explain only a fraction of the effects. In light of this, what is the distribution of eigencentrality (Fig. 2d)? Will it drop off quickly beyond the genes shown or it is distributed broadly across many nodes? In the latter case, this would confirm the results from Fig. 2c.
- In a small world network, it is hard to gain information from shortest path analysis. Instead, network propagation/diffusion methods have been proposed. The authors might want to try this approach on their network to find meaningful pathways or gene groups.

Minor points

- In lines 30 and 32, the authors should refer to Fig. 2d, correct?
- While the paper is written concisely, it should be checked whether the outline conforms with that of Nature Communications.
- The headline of the paper "Large-scale causal discovery using interventional data sheds light on the regulatory network architecture of blood traits" does not match the results from the analysis, in which the authors did not find any specific gene networks distinguishing the blood traits. This should be corrected.

Reviewer #5

(Remarks to the Author)
Interventional datasets offer a unique opportunity to infer causal relationships between variables, exemplified by the use of perturb-seq data for constructing gene interaction networks. Causal graphs can be inferred through the following two-step procedure: (1) estimate the average causal effect (ACE) of each feature on all others, and (2) derive the causal graph from the inverse of the ACE matrix. However, ACE estimation is inherently noisy and often results in ill-conditioned or non-invertible matrices. The authors propose a novel approach to estimate a sparse approximate inverse of the ACE matrix, addressing these challenges.

Following methodological benchmarking through simulations, the approach is applied to a perturb-seq dataset. The paper delineates the general characteristics of the inferred gene network. Next, the validation of the network involves examining the properties of genes with high eigencentrality and comparing the network distances among genes associated with identical versus different blood traits.

The manuscript poses important questions and employs sound methods. However, there are significant concerns regarding the validation of the method with real data:

1) The authors' validation of the inferred network primarily hinges on two aspects: (i) the association between gene eigencentrality and various gene annotations, and (ii) the observed statistical differences in "On-Network Distances" for genes linked to identical versus differing blood traits. This approach, while informative, could benefit from additional validation strategies. For instance, leveraging genetics to assess whether genes with high eigencentrality exhibit greater conservation or reduced likelihood of loss-of-functon(LoF) mutations. Furthermore, the authors might explore gene-set identification through graph clustering, investigate specific pathways, or evaluate the impact of pathway-specific genetic variations on phenotypes through pathway-based burden tests;

2) The manuscript's validation strategy could be significantly strengthened by incorporating analyses from additional, varied datasets. For example, the dataset in [1] provides an ideal context for further validating the proposed method.

Minor comment:
- The manuscript would benefit from addressing the relationship and distinctions between its proposed method and recent advancements in the field, such as the work in [2]. This comparison, or a reasoned discussion on the non-applicability of such methods to the current study, is essential for situating the work within the broader research landscape.
- The manuscript's reliance on a t-test to assess mean path length differences between gene sets prompts questions about test calibration under the null hypothesis. The reported marginal differences underscore the need for a clear baseline expectation of differences by chance within the network, especially given that this is one of the few quantitative analyses for validation in real data.

[1] T. M. Norman, M. A. Horlbeck, J. M. Replogle, A. Y. Ge, A. Xu, M. Jost, L. A. Gilbert, and J. S. Weissman. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science, 365(6455):786–793, 2019
[2] Bereket M, Karaletsos T. Modelling Cellular Perturbations with the Sparse Additive Mechanism Shift Variational Autoencoder. arXiv preprint arXiv:2311.02794. 2023 Nov 5.

Version 1:

Reviewer comments:

Reviewer #4

(Remarks to the Author)

The authors have substantially revised the manuscript and made significant efforts to include additional data for validation while exploring new methods for pathway identification, heritability, and functional enrichment. As a result, the approach and findings are now presented more clearly and concisely.

That said, the authors have moderated many of their claims compared to the previous version of the manuscript. According to the discussion, their primary conclusion is that their algorithm identifies a scale-free and small-world network, which they suggest "likely reflects general properties of large causal networks." While this finding aligns with prior knowledge about the topology of gene regulatory networks (GRNs), the authors were unable to infer or uncover causal relationships or trait-relevant genes, despite considerable effort.

Although their inspre algorithm represents a notable scientific achievement, its broader impact on biology remains uncertain, particularly in the light of the uncertainty of the graph estimates. The central question is whether the limited outcomes are due to insufficient data and methodology, as the authors suggest, or whether they stem from an intrinsic feature of GRNs. The concept of "complexity for simplicity" (Lauffenburger, PNAS, 2000) posits that cellular complexity allows for reliable decision-making under perturbations, leading to simple behavioral outputs. In this context, causal genes and pathways might be obscured in experimental outputs.

Indeed, GRNs with scale-free topology are designed to be robust in the face of uncertainty (Carson and Doyle, Phys. Rev. E, 1999). This raises a daunting question: can advanced experimental techniques and algorithms ever fully uncover new relationships, or is this inherently unachievable because regulatory causality remains hidden in interventional experiments such as perturb-seq? This should be discussed.

(Remarks on code availability)
I could install and run the code in my R environment and provide an output according to the example data and tests as explained in the Readme file. The code contains sufficient information to be used by the community.

Reviewer #5

(Remarks to the Author)
I would like to be more positive about this revision because, as I previously noted, the manuscript poses important questions, and the methodology is interesting. However, the changes made by the authors do not strengthen the validation—in fact, they introduce the following major concerns:

1) The authors shuffle node labels on the estimated graph, which does not strictly test whether their method identifies meaningful causal edges. For example, a method that correctly identifies hub genes as highly connected but assigns incorrect edges may still show high heritability under this permutation strategy. This is because hub genes retain a larger number of edges in the real network compared to the permuted one, leading to inflated heritability estimates.

A more appropriate control would involve permuting the causal gene assignments for each analyzed focal gene while keeping its number of edges intact. This would provide a more direct test of whether the inferred causal edges explain more variance than random ones—still a limited approach, but at least one that can be clearly interpreted.

2) Edge replication across datasets is weak, despite using the same biological system. The authors claim that global network properties are similar across datasets, but the actual edge overlap is poor. Given that both datasets come from the same perturb-seq system in K562 cells, this raises concerns about the robustness of their method. If their inference is stable, they should at least demonstrate some level of replication within the same dataset, for example, by splitting the data and assessing whether inferred networks are consistent in this simpler setting.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #4

(Remarks to the Author)
I would like to thank the authors for the honest feedback and discussion to my points raised.
The authors have thoroughly addressed all points raised with new analyses, and manuscript revisions that significantly improve the clarity, rigor, and impact of the work. The manuscript now presents a robust methodological contribution to the field of causal network inference, with a clear relevance to ongoing debates about GRN analysis. The added cross-validation, updated variance analysis, and discussion of GRN robustness enhance the study's scientific value, while the authors' acknowledgment of limitations will certainly spark further discussions and research.

(Remarks on code availability)
The provided code could be successfully downloaded, installed and the minimal example could be executed.

Reviewer #5

(Remarks to the Author)
Thank you for the revised manuscript.

On point 1, I acknowledge the addition of alternative validation strategies. While the chosen control (node label shuffling) has known limitations, it reflects current practice and is acceptable under field norms.

On point 2, the narrative around replication remains ambiguous. The second dataset is framed as validation, yet when edge overlap is low, the interpretation shifts to biological variability. This pivot muddies the claim of robustness. I suggest revising the language to clarify whether the second dataset is meant as validation, contextual insight, or both, and adjusting the framing accordingly.

(Remarks on code availability)

Dear reviewers,

We would like to thank you for the helpful and thorough feedback. We believe that our revisions based on these comments have substantially improved the clarity and comprehensiveness of our study. In summary, to address reviewer feedback we have

1. Added additional analyses of graph degree and eigencentrality distribution.
2. Added an analysis connecting the variance explained by the network a gene with that gene's genome-wide (*cis* + *trans*) heritability in whole blood.
3. Performed a comprehensive replication analysis comparing networks estimated on the essential-scale screen and the independent genome-wide screen.
4. Removed analyses of differential regulation of blood-trait associated genes (see discussion below).

These new analyses and discussion amount to eight additional main text figure panels, two additional supplemental figures and 1.5 pages of text. We also corrected a minor bug in the network edges that were being included in the eigencentrality calculation, resulting in more robust estimates and slight changes to the associations with gene annotations.

Finally, we would like to apologize for the delay in submission of this revision. Receipt of these reviews corresponded with the latter stages of the first author's faculty job search, followed subsequently by second visits, negotiations, relocation and starting a new position with new responsibilities. In addition, we took note of Reviewer #2's minor point about the null distribution of our test for differential regulation of blood trait associated genes, ultimately concluding that our estimates were not properly calibrated. We spent substantial time attempting to derive, sample from or approximate a proper null distribution but were ultimately unsuccessful and made the decision to pull these analyses and leave them for future work instead of spending additional time on what was amounting to a new methods paper in itself. Please see the discussion below. We thank the reviewers for their patience.

Below, we have copied the text of the reviewers comments and have provided a response to each comment in blue. Similarly, in the attached manuscript you will find additions in blue and deletions in red.

Reviewer #4 (Remarks to the Author):

The manuscript by Brown et al. presents a novel approach learning causal networks based on interventional data which builds on inverse sparse regression (inspre). The procedure first estimates the pairwise average causal effects between all features. As the true effects matrix cannot be determined exactly, the authors calculate the noisy, sparse inverse effect matrix instead, from which a causal graph can be estimated. The authors benchmarked their method against existing methods for causal discovery from observational and interventional data, showing superior performance in terms of speed, precision, Hamming distance, and other measures.
The authors then applied their method on interventional data of a perturb-seq experiment to improve the understanding of complex-trait genetics and the underlying architecture of complex

regulatory networks and pathways. For this they reconstructed an acyclic network based 788 genes from genome-wide perturb-seq experiments on K562 cells. From the network, the authors related shortest paths to effects, identifying central nodes and their importance in the network with albeit borderline significance. Correlating genes associated with blood traits and shortest paths and upstream effects but failed to do so.

The authors mention two reasons for the limited insight above. On the one hand, the K562 cancer cell line might behave differently compared to normal tissue. On the other hand, they econclude that gene networks are highly connected such that perturbation effects ramify through the whole network. Therefore, it is hard to impossible to assign core pathways and genes to individual perturbations, which limits the insights from their inspre algorithm.

Overall, the paper is well-written and concise. It deals with a timely and important subject on the inference of large-scale gene regulatory networks from interventional data. The inspre algorithm is – undoubtedly - an interesting step forward in this direction. However, its biological usefulness still needs to be shown. In the manuscript, the authors have confirmed prior biological knowledge, but new insights are missing. This should be added through another application of their method to biological data.

We thank the reviewer for their positive feedback and discuss additional new insights below.

Major points

- The reason why the authors picked exactly 788 genes from the perturb-seq data should be explained in more detail. How would the results change, if the authors used fewer or more genes? Are the findings a result of the limited number of genes used for network inference or are they a general feature of the omnigene model, small world and scale-free property of networks?

We have edited the text to more clearly reflect our inclusion criteria (p. 3, l. 21-27). In response to this and to Reviewer #2's point (2) below, we constructed a larger network of 1429 genes from the independent genome-wide-scale screen. We observed nearly identical properties in this network as the essential screen data (p. 5, l. 14-20; Figure S3). We conclude that these findings are likely to be a general feature of large gene networks and have added this point to the discussion (p. 6, l. 1-2).

- The authors benchmarked the inspre algorithm using a network of 50 nodes. From the the perturb-seq data they inferred a network with 788 nodes, which is one order of magnitude larger. Given the fact that the authors could not infer new biological insights from the large network, it would be advisable to infer a biological network having a similar number of nodes, say from a smaller chemical compound or siRNA screen? Would it be possible to use genes in well-defined pathways and their perturbation as a proof of principle for a smaller inspre-inferred network?

This is an important point and interesting suggestion. In benchmarking, we are limited to network sizes that are tractable for comparable methods, which often take orders of magnitude longer to run than *inspre*. This makes it infeasible to benchmark methods on large datasets. Moreover, we chose to focus on CRISPR-inhibition data due to the advantages it presents over previous screening approaches including knowledge of the target (as compared to chemical screens), reduced off-target effects (as compared to siRNA screens) and presence of both perturbation and response data for all modelled genes, all of which are required by our method.

- The GRN inferred from the perturb-seq data using the inspre algorithm confirms existing knowledge on the small world property of GRNs and regulatory properties of hub genes, such as the anti-proportionality between dosage sensitivity and centrality. However, it remains open whether the network also shows scale-free properties concerning connectivity. This should be shown/discussed.

We thank the reviewer for pointing out this omission and have added degree and eigencentrality distribution plots to Figure 2. We observe that our network indeed shows scale-free properties (p. 3, l. 29-31).

- The authors find that shortest-path measures explain only a fraction of the effects. In light of this, what is the distribution of eigencentrality (Fig. 2d)? Will it drop off quickly beyond the genes shown or it is distributed broadly across many nodes? In the latter case, this would confirm the results from Fig. 2c.

We have added the eigencentrality distribution to Figure 2. We note that while there are only a few genes with very high eigencentrality, there are many other genes with non-trivial centrality. Specifically, there are 125 genes with eigencentrality above 0.2. We have added this observation to the text (p. 3, l39-41).

- In a small world network, it is hard to gain information from shortest path analysis. Instead, network propagation/diffusion methods have been proposed. The authors might want to try this approach on their network to find meaningful pathways or gene groups.

We thank the reviewer for pointing us to this interesting line of work. In our review of this literature, we noticed 1) these approaches have been developed primarily for undirected graphs, rather than directed/causal graphs and 2) these approaches typically rely on fixed graph structures, rather than estimated ones. Still, we reasoned that a directed extension of network propagation would result in analysis of the total causal effect network (that we call U). We spent substantial time attempting analyses of this network, ultimately realizing that rigorous analysis required an understanding of the sampling distribution of U. Despite devising several approaches to understand this sampling distribution, we were ultimately unsuccessful and decided to move on and leave this point to future research. Please see this discussion (p. 6, l. 15-21) and response to Reviewer #2, minor point #2 below.

Minor points

- In lines 30 and 32, the authors should refer to Fig. 2d, correct?

These labels have been corrected.

- While the paper is written concisely, it should be checked whether the outline conforms with that of Nature Communications.

Initially, we planned a brief correspondence style submission, but given the expanded analyses in review we have adjusted the formatting to a full article.

- The headline of the paper "Large-scale causal discovery using interventional data sheds light on the regulatory network architecture of blood traits" does not match the results from the analysis, in which the authors did not find any specific gene networks distinguishing the blood traits. This should be corrected.

We have changed the title.

Reviewer #5 (Remarks to the Author):

Interventional datasets offer a unique opportunity to infer causal relationships between variables, exemplified by the use of perturb-seq data for constructing gene interaction networks. Causal graphs can be inferred through the following two-step procedure: (1) estimate the average causal effect (ACE) of each feature on all others, and (2) derive the causal graph from the inverse of the ACE matrix. However, ACE estimation is inherently noisy and often results in ill-conditioned or non-invertible matrices. The authors propose a novel approach to estimate a sparse approximate inverse of the ACE matrix, addressing these challenges.

Following methodological benchmarking through simulations, the approach is applied to a perturb-seq dataset. The paper delineates the general characteristics of the inferred gene network. Next, the validation of the network involves examining the properties of genes with high eigencentrality and comparing the network distances among genes associated with identical versus different blood traits.

The manuscript poses important questions and employs sound methods. However, there are significant concerns regarding the validation of the method with real data:

We thank the reviewer for their feedback and believe we have addressed their concerns.

1) The authors' validation of the inferred network primarily hinges on two aspects: (i) the association between gene eigencentrality and various gene annotations, and (ii) the observed

statistical differences in "On-Network Distances" for genes linked to identical versus differing blood traits. This approach, while informative, could benefit from additional validation strategies. For instance, leveraging genetics to assess whether genes with high eigencentrality exhibit greater conservation or reduced likelihood of loss-of-functon(LoF) mutations. Furthermore, the authors might explore gene-set identification through graph clustering, investigate specific pathways, or evaluate the impact of pathway-specific genetic variations on phenotypes through pathway-based burden tests;

We have added new analyses that support our claims. Indeed, several of the annotations that are significantly associated with eigencentrality are constructed from conservation and loss of function intolerance. This includes gnomAD pLI (probability of loss of function intolerance), and sHet (Estimated selection coefficient on heterozygous loss-of-function mutations in ExAC), (p. 4, l. 17-24; Figure 2c, Figure S1). We also added a new analysis of whole blood genome-wide (*cis*+*trans*) heritability using the INTERVAL dataset, showing that genes with high genome-wide heritability also have high variance explained by the network we estimate in K562 cells (p. 5, l.27-37; Figure 3d). We extensively explored gene set and pathway identification, but ultimately concluded that robust and calibrated inference was difficult without understanding the sampling distribution of the estimator for the graph. See the response to minor comment #2 below.

2) The manuscript's validation strategy could be significantly strengthened by incorporating analyses from additional, varied datasets. For example, the dataset in [1] provides an ideal context for further validating the proposed method.

We thank the reviewer for pointing us to this interesting study. While that study uses CRISPRa and mainly attempts to identify guide-guide interactions, it helped realize that a validation dataset already exists in the Replogle et al dataset we analyze in this manuscript. In particular, we had initially analyzed the essential-scale screen, but not the independent genome-wide screen. We added an extensive analysis of this dataset (p. 5, l. 5-25; Figure 4, Figure S2, Figure S3). In brief, we found: 1) the graph estimated on the essential screen data could explain a large amount of variance in the genome-wide screen data, 2) network properties between graphs estimated on the two datasets were extremely similar, and 3) gene centralities between the two graphs were similar. However, we also found that the overlap in actual included edges between the two graphs was limited. We have added our thoughts on this to the Discussion (p. 6, l. 39-43).

Minor comment:
- The manuscript would benefit from addressing the relationship and distinctions between its proposed method and recent advancements in the field, such as the work in [2]. This comparison, or a reasoned discussion on the non-applicability of such methods to the current study, is essential for situating the work within the broader research landscape.

We have added text to the discussion regarding the relationship between our method and this (and similar) methods (p. 6, l. 22-29). In brief, our approach and those like it aim to directly

identify the causal graph, while the referenced approach aims to identify gene modules without identifying the graph. These are complementary but distinct analysis goals.

- The manuscript's reliance on a t-test to assess mean path length differences between gene sets prompts questions about test calibration under the null hypothesis. The reported marginal differences underscore the need for a clear baseline expectation of differences by chance within the network, especially given that this is one of the few quantitative analyses for validation in real data.

This turned out to be more complicated than we initially imagined. After some experimentation we realized that this null distribution was indeed improperly calibrated. We attempted multiple permutation strategies involving sampling sets of random genes or shuffling node labels, but none were able to produce well-calibrated null distributions. Ultimately, we realized that the test statistic distribution depended on the sampling distribution of the graph. Specifically, edge inclusions are not independent events, and thus to perform such analyses we needed to be able to estimate the joint distribution of all edge inclusion events resulting from our algorithm. As graph inference is computationally intensive, we were not able to feasibly conduct large-scale permutation tests to assess this distribution. We attempted several other ways of approximately sampling graphs and modeling the edge joint distribution, but were unable to obtain null distributions that were clearly properly calibrated. We also realized solving this problem would require additional simulations to validate any approach that we did come up with. After spending over a month on this, we ultimately decided to leave this important question to future research in the interest of time. This observation has interesting implications. Any method, not just inspre, that attempts to estimate a graph and then perform path identification or gene set enrichment identification must be able to quantify uncertainty in the graph structure. Please see the discussion in the text (p. 6, l. 15-21).

[1] T. M. Norman, M. A. Horlbeck, J. M. Replogle, A. Y. Ge, A. Xu, M. Jost, L. A. Gilbert, and J. S. Weissman. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. Science, 365(6455):786–793, 2019
[2] Bereket M, Karaletsos T. Modelling Cellular Perturbations with the Sparse Additive Mechanism Shift Variational Autoencoder. arXiv preprint arXiv:2311.02794. 2023 Nov 5.

Dear Reviewers,

We would like to thank you for the helpful comments on our manuscript. To address this feedback, we have added a discussion about the potential robustness of biological systems to perturbation, changed our method for calculating the variance explained in each gene to control for in-degree, and added a cross-validation analysis.

Reviewer #4 (Remarks to the Author):

The authors have substantially revised the manuscript and made significant efforts to include additional data for validation while exploring new methods for pathway identification, heritability, and functional enrichment. As a result, the approach and findings are now presented more clearly and concisely.

We appreciate these positive comments.

That said, the authors have moderated many of their claims compared to the previous version of the manuscript. According to the discussion, their primary conclusion is that their algorithm identifies a scale-free and small-world network, which they suggest "likely reflects general properties of large causal networks." While this finding aligns with prior knowledge about the topology of gene regulatory networks (GRNs), the authors were unable to infer or uncover causal relationships or trait-relevant genes, despite considerable effort.

Respectfully, we disagree with the characterization that this is the primary contribution of our work, and the reviewer here refers only to a small section of the Discussion that in its entirety describes our findings and their implications more broadly. First, we do in fact estimate causal relationships between genes. This is in contrast to prior studies on the large-scale topology of GRNs which are based on co-expression analysis. Second, while uncovering trait-relevant genes was never our goal per se, we successfully integrate with gene essentiality metrics and identify important relationships between gene essentiality and network centrality in the causal graph, which are highly relevant features of rare disease genes in particular. These results were replicated in an independent dataset. Finally, we also integrate with blood eQTL data and find an association between *trans*-eQTL heritability and variance in gene expression explained by the network. This is an important result given that K562 is commonly used to study blood traits, and there is an ongoing debate about the informativeness of regulatory networks in K562 (and other cancer cell lines) in the interpretation of gene regulation in primary cells. Altogether, these results demonstrate multiple lines of biological insights obtained from our approach, demonstrating its potential for further applications with increasingly powerful data.

Although their inspre algorithm represents a notable scientific achievement, its broader impact on biology remains uncertain, particularly in the light of the uncertainty of the graph estimates. The central question is whether the limited outcomes are due to insufficient data and methodology, as the authors suggest, or whether they stem from an intrinsic feature of GRNs. The concept of "complexity for simplicity" (Lauffenburger, PNAS, 2000) posits that cellular complexity allows for reliable decision-making under perturbations, leading to simple behavioral outputs. In this context, causal genes and pathways might be obscured in experimental outputs.

Indeed, GRNs with scale-free topology are designed to be robust in the face of uncertainty (Carson and Doyle, Phys. Rev. E, 1999). This raises a daunting question: can advanced experimental techniques and algorithms ever fully uncover new relationships, or is this inherently unachievable because regulatory causality remains hidden in interventional experiments such as perturb-seq? This should be discussed.

We agree with the reviewer that this is an incredibly important point. We began to address this in our discussion, saying "It is unclear whether the network connections discovered via said experiments are the same as those that the molecular effects of GWAS variants propagate along. Instead, large-scale gene promoter inhibition may result in changes to cellular state that brings a concomitant change in network structure". However, we did not discuss the possibility that robustness to perturbation may preclude causal inference from these experiments. We have added a discussion of this to the manuscript (p6, l26-34).
This may also explain why we see high concordance of network estimates in cross-validation but lower concordance with external validation (see response to Reviewer #5 point 2 below).

Most importantly, we would like to point out that the majority of this critique is not about our contribution but rather about the value of the broader scientific effort to use Perturb-seq to understand regulatory interactions. This is an incredibly popular and fast-growing field with substantial investment from both the public and private sectors. Due to widespread interest in this topic, studies that assess the feasibility of using Perturb-seq data to infer regulatory relationships are critical, perhaps especially if they demonstrate that getting consistent estimates from current data is challenging, as this can drive improved study designs that addresses the current caveats, as we state in the Discussion. Thus, we believe that our work is going to be of broad interest to the scientific community, as it provides a rigorous method to analyze a novel data type at scale and demonstrates what can and cannot be inferred from the current data, enabling and inspiring further inquiry.

Reviewer #5 (Remarks to the Author):

I would like to be more positive about this revision because, as I previously noted, the manuscript poses important questions, and the methodology is interesting. However, the changes made by the authors do not strengthen the validation—in fact, they introduce the following major concerns:

1) The authors shuffle node labels on the estimated graph, which does not strictly test whether their method identifies meaningful causal edges. For example, a method that correctly identifies hub genes as highly connected but assigns incorrect edges may still show high heritability under this permutation strategy. This is because hub genes retain a larger number of edges in the real network compared to the permuted one, leading to inflated heritability estimates.

A more appropriate control would involve permuting the causal gene assignments for each analyzed focal gene while keeping its number of edges intact. This would provide a more direct test of whether the inferred causal edges explain more variance than random ones—still a limited approach, but at least one that can be clearly interpreted.

We agree with the reviewer that assessing identification of meaningful causal edges is challenging. We explored several possibilities, including this exact suggestion, before settling on the one we reported. Ultimately, we chose the approach that we did because it had been used in a recently-published and peer reviewed study on a similar topic (Ružičková *et al* PNAS 2024), however we understand the concern about preserving the number of edges. We have adjusted our analysis to use both approaches. Specifically, when estimating the variance explained in an *individual* gene, we keep the in-degree intact while sampling random genes that were determined not to be directly causal by our algorithm. This results in 605 genes with significant variance explained at FDR 5%, a substantially larger number than our previous analysis randomizing the entire graph (100 genes). We have changed Figure 4 to reflect these results and added these results on page 4, lines 29-36. Note that because this per-gene edge resampling does not yield valid graph structures, we still take the label permutation approach when estimating the variance in the *entire dataset* explained by our graph versus random graphs.

2) Edge replication across datasets is weak, despite using the same biological system. The authors claim that global network properties are similar across datasets, but the actual edge overlap is poor. Given that both datasets come from the same perturb-seq system in K562 cells, this raises concerns about the robustness of their method. If their inference is stable, they should at least demonstrate some level of replication within the same dataset, for example, by splitting the data and assessing whether inferred networks are consistent in this simpler setting.

We have added cross-validation analyses to assess stability within the essential-screen dataset. In the initial revision, we followed the reviewers prior suggestion to validate with external data. However we had already used cross-validation to select our sparsity hyperparameter, and have now added additional analyses of the graphs produced in each fold. Specifically, we used 5-fold cross-validation and calculated the consistency of edge inclusion in the cross-validated graphs (p6 l2-9). We found generally high concordance between our reported graph and the consensus graph (F1=0.68), with slightly lower but still strong concordance between our reported graph and the individual cross-validation graphs (mean F1=0.58). The concordance within a dataset but discordance between datasets indicates the method is robust, but the experimental system may not be. This is an important insight, especially given reviewer #4s concern about the utility of Perturb-seq as a general system for inferring regulatory relationships, and we have added this to discussion (p6, l26-34). Given the widespread interest in deriving causal networks from large-scale Perturb-seq datasets, we believe this differential consistency will be of high interest to readers.