Computational and Systems Biology

Nonlinear transcriptional responses to gradual modulation of transcription factor dosage

Júlia Domingo, **Mariia Minaeva**, **John A Morris**, **Samuel Ghatan**, **Marcello Ziosi**, **Neville E Sanjana**, **Tuuli Lappalainen**

New York Genome Center, New York, United States • Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden • Department of Biology, New York University, New York, United States

Reviewed Preprint
v2 • November 25, 2025
Revised by authors

Reviewed Preprint
v1 • December 17, 2024

eLife Assessment

This **important** work develops a new protocol to experimentally perturb target genes across a quantitative range of expression levels in cell lines. The evidence supporting their new perturbation approach is **convincing**, and we propose that focusing on single modality (activation or inhibition) would be sufficient to draw their conclusions. The study will be of broad interest to scientists in the fields of functional genomics and biotechnology.

https://doi.org/10.7554/eLife.100555.2.sa3

Abstract

Genomic loci associated with common traits and diseases are typically non-coding and likely impact gene expression, sometimes coinciding with rare loss-of-function variants in the target gene. However, our understanding of how gradual changes in gene dosage affect molecular, cellular, and organismal traits is currently limited. To address this gap, we induced gradual changes in gene expression of four genes using CRISPR activation and inactivation. Downstream transcriptional consequences of dosage modulation of three master trans-regulators associated with blood cell traits (*GFI1B*, *NFE2*, and *MYB*) were examined using targeted single-cell multimodal sequencing. We showed that guide tiling around the TSS is the most effective way to modulate *cis* gene expression across a wide range of fold-changes, with further effects from chromatin accessibility and histone marks that differ between the inhibition and activation systems. Our single-cell data allowed us to precisely detect subtle to large gene expression changes in dozens of *trans* genes, revealing that many responses to dosage changes of these three TFs are nonlinear, including non-monotonic behaviours, even when constraining the fold-changes of the master regulators to a copy number gain or loss. We found that the dosage properties are linked to gene constraint and that some of these nonlinear responses are enriched for disease and GWAS genes. Overall, our study provides a straightforward and scalable method to precisely modulate gene expression and gain insights into its downstream consequences at high resolution.

Introduction

Precision control of gene expression levels plays a pivotal role in defining cell type specificity and coordinating responses to external stimuli. Imbalances in this intricate regulation can underlie the genetic basis of both common and rare human diseases. The vast majority of genetic variants associated with complex disease, as revealed by genome-wide association studies (GWAS), are located in noncoding regions, with likely gene regulatory effects [1]. Previous studies have attempted to elucidate these effects by mapping genetic associations to gene expression [2,3], and more recently, CRISPR-based perturbations of GWAS loci have provided insights into their functional consequences [4]. A major driver of rare genetic diseases is loss-of-function variants affecting one or both copies of the gene, leading to disease via dramatic reduction of functional gene dosage [5]. The substantial overlap [6,7] and potential joint effects [8,9] of rare and common variants indicate a general link between different degrees of perturbation of gene dosage and disease phenotypes.

However, our understanding of the quantitative relationship between gradual changes in gene dosage and downstream phenotypes remains elusive for most human genes. Practical applications of the compelling allelic series concept to identify genes where increasingly deleterious mutations have increasing phenotypic effects have been limited by the sparsity of segregating variants with an impact on a given gene in the human population [10]. Experimental characterization of gene function in model systems has predominantly relied on gene knock-out or knock-down approaches [11]. While these studies have proven useful to identify dosage-sensitive genes involved in cellular functions and disease [12–16], these approaches only provide a limited discrete relationship between the number of functional gene copies and a certain phenotype (eg. loss-of-function consequence vs. wild-type). However, such relationships are in fact determined by continuous dosage-to-phenotypes functions that, as suggested by a small number of previous experimental studies [17–19], can be complex and thus are challenging to infer from loss-/gain-of-function data.

Recently, new methods have enabled the gradual modulation of gene dosage in model systems [18,20–22], while large-scale insights into the downstream effects of dosage modulation have largely come from yeast [17] and bacteria [19,23], demonstrating that nonlinear relationships between gene dosage and phenotype are common. In humans, the relationship between dosage and downstream phenotypes is largely unexplored. Only a few limited studies [17–19] have dissected these consequences. For instance, the disease-associated transcription factor *SOX9* [24] showed a nonlinear relationship between dosage and multiple tiers of phenotypes, including DNA accessibility, RNA expression of downstream targets, raising the question of whether this phenomenon occurs with other transcription factors. More recently, similar evidence has been shown in the case of the *NKX2-1* lineage factor with an oncogenic role in lung adenocarcinoma [25]. Generally, transcription factors represent a particularly compelling target for the characterization of gene dosage effects. They are key regulators of cellular functions, enriched for disease associations [26] and often classified as haploinsufficient [27]. Additionally, their effects can be measured by transcriptome analysis. However, our knowledge of their dosage-dependent effects on regulatory networks still remains limited, particularly regarding subtle dosage variation within their natural range [22].

In this study, we developed and characterised a scalable novel sgRNA design approach for gradually decreasing and increasing gene dosage with the CRISPR interference (CRISPRi) and activation (CRISPRa) systems. We applied this to four genes, with single-cell RNA-sequencing (scRNA-seq) as a cellular readout of downstream effects. While classic Perturb-Seq analyses have focused on gene knockdown effects, we assess the effects of gradual up- and down-regulation of

target genes. We uncovered quantitative patterns of how gradual changes in transcription dosage lead to linear and nonlinear responses in downstream genes. Many downstream genes are associated with rare and complex diseases, with potential effects on cellular phenotypes.

Results

Precise modulation and quantification of gene dosage using CRISPR and targeted multimodal single-cell sequencing

We selected four genes for gradual modulation of their dosage in the human erythroid progenitor cell line K562 [28]: *GFI1B*, *NFE2*, *MYB* and *TET2*. Two of the genes, *GFI1B* and *NFE2*, have been implicated in blood diseases and traits [29–31], and in our earlier work, we identified a broad transcriptional response to inhibition of GWAS-overlapping enhancers to these genes [4]. *MYB* is a key transcription factor [32] and a downstream target of *GFI1B* [4].*TET2* has a role in DNA demethylation and is unrelated to these transcriptional networks and is included in this study as a control with minimal expected *trans* effects. We refer to these four genes, targeted in *cis* for modulation of their regulation, as *cis* genes (**Figure 1A**).

To modulate the gene expression of the *cis* genes, we use K562 cells expressing CRISPRi (KRAB-dCas9-MECP2) and CRISPRa (dCas9-VPR) systems (see Methods), both cell lines were hashed with DNA conjugated antibodies against different surface proteins that allow pooled experiments. To obtain a wide range of dosage effects we used four different single guide RNA (sgRNA) design strategies (**Figure 1B**): 1) targeting the transcription start site (TSS) as in the standard CRISPRi/CRISPRa approach, 2) tiling sgRNAs +/- 1000 bp from the TSS in approximately 100bp intervals, 3) targeting known *cis*-regulatory elements (CREs), and 4) using attenuated guides that target the TSS but contain mismatches to modulate their activity [18]. We further included 5 non-targeting control (NTC) sgRNAs as negative controls.

The library of altogether 96 guides was transduced to a pool of K562-CRISPRi and K562-CRISPRa cells at low multiplicity of infection (MOI). After eight days, we performed ECCITE-seq (see Methods) to capture three modalities: cDNA, sgRNAs and surface protein hashes (oligo-tagged antibodies with unique barcodes against ubiquitously expressed surface proteins). Instead of sequencing the full transcriptome, we used target hybridization to capture a smaller fraction of the cDNA and obtain more accurate expression readouts at a feasible cost. The subset of selected transcripts were picked from the transcriptional downstream regulatory networks of *GFI1B* and *NFE2* identified previously [4], maintaining similar patterns of co-expression correlation across co-expression clusters (see Methods, Figure S1A). We targeted a total of 94 transcripts (**Figure 1A**), including the four *cis* genes, 86 genes that represent trans targets of *GFI1B* and/or *NFE2* [4] (Figure S1A), *LXH3* that is not expressed in blood progenitors, *GAPDH* that is highly expressed and often considered an invariable housekeeping gene and the dCas9-VPR or KRAB-dCas9-MeCP2 transcripts.

We used the protein hashes and the dCas9 cDNA (indicating the presence or absence of the KRAB domain) to demultiplex and determine the cell line—CRISPRi or CRISPRa. Cells containing a single sgRNA were identified using a Gaussian mixture model (see Methods). Standard quality control procedures were applied to the scRNA-seq data (see Methods). To confirm that the targeted transcript capture approach worked as intended, we assessed concordance across capture lanes (Figure S1C). The final data set had 20,001 cells (10,647 CRISPRi and 9,354 CRISPRa), with an average of 81 and 86 cells with a unique sgRNA for the CRISPRa and CRISPRi, respectively (Figure S1D).

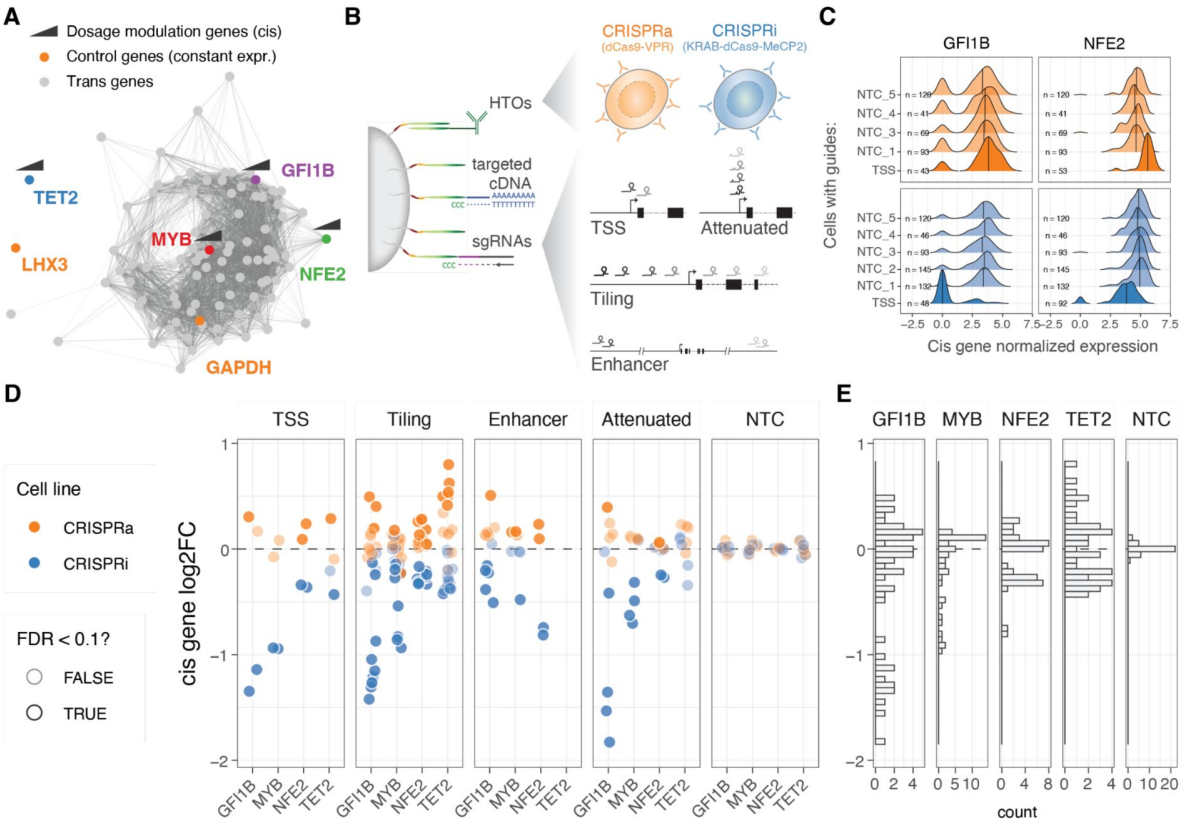

Figure 1

Modulation and quantification of gene dosage using CRISPR and targeted multimodal single-cell sequencing.

A. Co-expression network representation of the 92 selected genes under study. Genes (nodes) are connected by edges when their co-expression across single cells was above 0.5 (data used from Morris *et al*. 2023). Highlighted in colour are the two control highly (GAPDH) and lowly (LHX3) constantly expressed genes, as well as cis genes for which dosage was modulated with CRISPRi/a. B. Design of the multimodal single cell experiment (HTO = hash-tag oligos). C. Distribution of the GFI1B (left) or NFE2 (right) normalised expression across single cells for different classes of sgRNAs (NTC = Non-targeting controls, TSS = transcription start site). D. Resulting relative expression change (log2 fold change) of the 4 cis genes upon each unique CRISPR perturbation when grouped across different classes of sgRNAs. E. Distribution of cis gene log2FC across all sgRNA perturbations.

Gradual modulation of gene expression across a broad range with CRISPRi/a

Next, we calculated the expression fold change for each of the four *cis* genes targeted by each sgRNA in the two cell lines (CRISPRi/a), comparing each group of cells with its respective NTC sgRNA group (see Methods). We first confirmed that the sgRNAs targeting the transcription start site (TSS) up- and down-regulated their targets (**Figure 1C**, **Figure S1F**). When looking at all sgRNAs at once, across the four genes, we observed a 2.3 fold range (**Figure 1E**), with a minimum 72% reduction and maximum 174% increased expression (log2(FC) values from -1.83 to 0.80). However, the range varied between the genes, with *GFI1B* covering the widest range of gene expression changes (gene expression ranging between 0.28 to 1.42 fold), while *MYB* expression could not be pushed higher than 1.13 fold (**Figure 1E**). The direction of the effects were consistent with the cell lines of origin, where 98.88% of the significant perturbations (Wilcoxon rank test at 10% FDR, n = 89) were correctly predicted based on the direction of the target gene fold change. The predicted on- and off-target properties of the guides [33–35] did not correlate with the fold changes in the *cis* genes (Figure S2A), suggesting that the observed effects represent true *cis*-regulatory changes. The fold changes were also robust to the number of cells containing a particular sgRNA (Figure S2B, top).

We verified that the fold change estimation was not biased depending on the expression level of the target gene at the single-cell level, which can vary due to drop-out effects or binary on/off effects of the KRAB-based CRISPRi system [20]. By splitting cells with the same sgRNA based on the normalised expression of the *cis* gene (0 vs. >0 normalised UMIs, Figure S3A), we observed highly concordant transcriptome gene expression effects between the two groups (Figure S3B). This indicates that the dosage changes per guide were not primarily driven by the changing frequency of binary on/off effects, and the use of pseudo-bulk fold changes provides a robust estimation of *cis* gene fold changes. These patterns are further supported by the cells forming a gradient rather than distinct clusters on a UMAP (Figure S3C).

The fold change patterns differed between sgRNA designs (**Figure 1D**, left). As expected, sgRNAs targeting the TSS showed strong perturbations in gene expression. However, sgRNAs tiled +/- 1kb from the TSS provided a broader and more gradual range of up- and downregulation across the target genes, sometimes surpassing the effects of TSS-targeting sgRNAs. Attenuated sgRNAs with mismatch mutations resulted in a range of gene silencing effects in the CRISPRi line, as expected based on their original design [18]. However, these attenuated sgRNAs did not exhibit such a dynamic range in the CRISPRa modality, although a significant correlation existed between the silencing or activating effect size and the distance of the mismatch from the protospacer adjacent motif (PAM) when considering all data points together (Figure S2C). The sgRNAs targeting distal *cis*-regulatory elements (CREs) showed both inhibiting and activating effects, even though both the CRISPRi and CRISPRa constructs were initially designed to inhibit or activate transcription from the promoter and initial gene body region. Nonetheless, the number of known CREs per gene is typically limited. Given its simplicity and the ability to achieve both up- and downregulation of the target gene, we consider the tiling sgRNA approach, with a simple design that only requires annotation of the TSS, as a useful method for gradually modulating gene dosage with CRISPRi/a systems.

Cis determinants of dosage

Having designed guides targeting both distal and local neighbouring regulatory regions of the four transcription factors (TFs) and ensuring minimal bias in fold-changes due to sgRNA's biochemical properties, we investigated the *cis* features that determine the strength of dosage perturbation. We observed substantial differences in the effects of the same guide on the CRISPRi and CRISPRa backgrounds, with no significant correlation between *cis* gene fold-changes (**Figure 2A**). However, in both modalities, the strongest effects on gene expression were observed when the

guides were close to the transcription start site (TSS) (**Figure 2B**, excluding NTC and attenuated sgRNAs), although the peaks of strongest activation or repression differed between the modalities. In the CRISPRi modality, the maximum effect was located within the gene body at +238 bp from the TSS (**Figure 2B**, bottom), consistent with previous studies that used essentiality as a proxy for expression [36]. However, in the CRISPRa modality, the maximum average fold changes occurred closer to the TSS at around -99 bp (**Figure 2B**, bottom), as also shown for CD45 [37].

Enhancer, tiling, and TSS sgRNAs targeted regions of the genome with different compositions of histone marks, annotated by ENCODE, in K562 cells[38] (**Figure 2C**), which allowed us to investigate the impact of chromatin state on the strength of *cis* gene dosage modulation. The magnitude of *cis* gene fold changes varied significantly depending on the presence of specific marks or peaks, which again differed between the two modalities (**Figure 2D**). In the CRISPRa cell line, the strongest effects were observed when guides were located in regions with open chromatin marks, such as DNase or ATAC peaks. In contrast, the strongest repression by CRISPRi occurred in genomic regions with the presence of H3K27ac, H3K4me3, and H3K9ac marks. These differences may be explained by the distinct mechanisms of action of the activator and repressor domains. MeCP2 and KRAB repressor domains recruit additional repressors that silence gene expression through chromatin remodelling activities such as histone deacetylation [39]. On the other hand, the VPR activation fusion domain may only require Cas9 to scan the open chromatin and recruit RNA polymerase and additional transcription factors to activate transcription. Overall, while a few sgRNAs have a strong effect in both CRISPRi and CRISPRa cell lines, a single guide library containing guides optimised for both modalities enables a range of gradual dosage regulation. However, larger data sets are needed for more careful modelling of the ideal dosage modulation designs and to understand how both *cis*-regulatory features, feedback loops, and other mechanisms contribute to the outcomes.

Trans responses of transcription factor dosage modulation

We then turned our attention to the remaining 91 genes captured by our custom panel and determined the relative expression fold change of each *trans* gene, compared to NTC in each unique guide perturbation (see Methods). Principal component analysis (PCA) performed on all pseudo-bulk fold changes demonstrated the removal of batch effects from the cell lines and revealed a clear direction of the *cis* gene dosage effect in the first three principal components (Figure S4B). This finding suggests that dosage modulation is the primary determinant of *trans* effects. The PCA indicated that the dosage modulation of *GFI1B* and *MYB* is reflected in opposite directions in PC1 and PC2, while the *trans* responses of *NFE2* are captured by PC3.

Using a false discovery rate (FDR) cutoff of 0.05, all 91 *trans* genes except for the neural-specific TF *LHX3* (negative control) exhibited a significant change in expression upon perturbation of any of the TFs. The observed trans-effects were well correlated with perturbations of these genes in other data sets (Figure S4C,D). Among all measured fold changes, the most extreme negative effect sizes were observed in *cis* genes, with the top 10 being predominantly reductions in *GFI1B* expression. This indicates that *cis* downregulation tended to surpass the endogenous expression limits. In contrast, the largest increases in gene expression were observed through *trans* mechanisms, where *KLK1* and *TUBB1* reached the largest expression values when *GFI1B* was strongly upregulated, or *SPI1* and *DAPK1* when *GFI1B* was strongly downregulated. These findings suggest that the CRISPRa approach did not reach a biological ceiling of overexpression.

Inspecting *trans* responses as a function of *cis* gene modulation, we observed that the number of expressed genes and the mean absolute expression changes of *trans* genes exhibited gene-specific correlations with *cis*-gene dosage (**Figure 3A**, **Figure S4E**). Perturbations in *GFI1B* led to the most pronounced *trans* responses, with positive dosage changes resulting in larger effect sizes compared to decreasing TF gene expression, where the effect plateaued. *NFE2* exhibited similar patterns but with a smaller magnitude. In the case of *MYB*, *trans* responses were observed when decreasing the expression of this TF, but the effects of upregulation are largely unknown as we

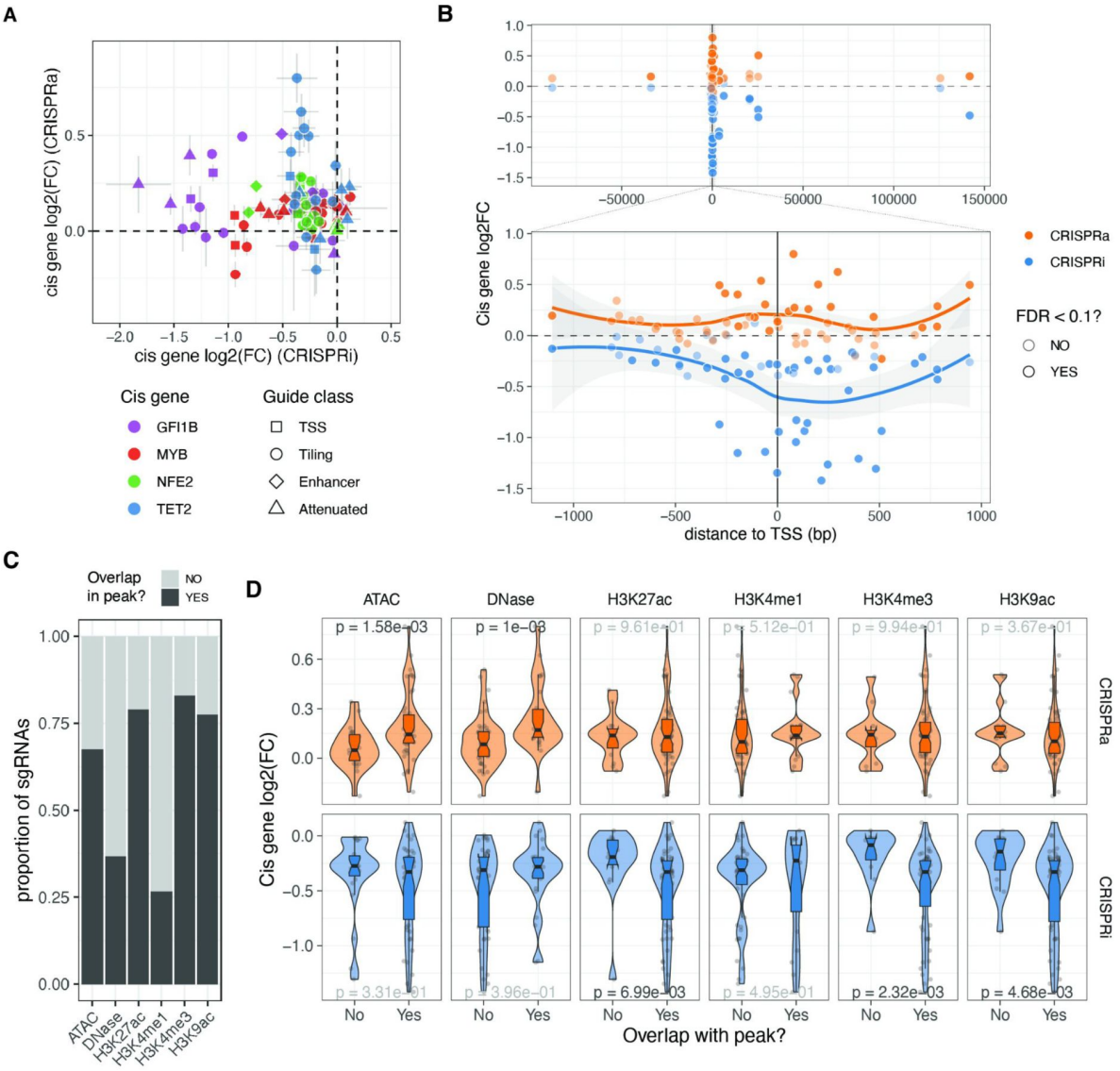

Figure 2

Cis determinants of dosage.

A. Comparison of the relative expression change (log2FC) from the same sgRNA between the two different CRISPR modalities. Vertical and horizontal bars represent CRISPRa and CRISPRi standard errors, respectively. B. Relative expression change of the targeted cis gene based on distance from transcription start site (TSS). Top plot excluded attenuated and NTC sgRNAs, while bottom plot also excludes enhancer sgRNAs. C. Number of sgRNAs that overlap with the different epigenetic or open chromatin peaks. D. Relative expression change to NTC sgRNAs (log2(FC)) of all cis genes when their sgRNAs fall or not in the different epigenetic or open chromatin peaks. P-value result from Wilcoxon rank-sum tests, with nominally significant p-values shown in black.

were unable to increase *MYB* expression beyond 0.35. As expected, given the unrelatedness of *TET2* to the *trans* network, dosage modulation of this gene had minimal *trans* effects with the least pronounced trend when compared to *TET2* dosage, so we excluded it from subsequent analyses.

Widespread nonlinear dosage responses in *trans* regulatory networks

Upon clustering the changes in expression of *trans* genes based on the *cis* gene dosage change linked to each sgRNA, we identified distinct clusters exhibiting different dosage-response patterns (**Figure 3B** for *GFI1B*, **Figure S5-8A** for all *cis* genes). Further examination of the gene expression fold changes for each individual trans gene in relation to the TF fold changes revealed a diverse range of response patterns (**Figure 3C**, **Figure S5-8B** for all *cis* genes). These responses exhibited both linear and nonlinear forms, including some instances of non-monotonic gene expression responses for certain *trans* genes within the *GFI1B* trans network (e.g., *GATA2* in **Figure 3C**, **Figure S9E**).

To accurately characterise the dosage response, we employed both linear and nonlinear modelling approaches (**Figure 3D**), which allowed us to quantitatively assess the extent of nonlinear responses by comparing the goodness of fit of these models using the Akaike Information Criterion (AIC). For the nonlinear model, we utilised a sigmoid function with four free parameters (**Figure 3D**, right). These parameters represented the slope at the inflection point (*b*, indicating the rate of increase or decrease in expression), the minimum and maximum asymptotes (*c* and *d*, representing the lower and upper limits of fold change), and the value of *cis* gene expression at which the inflection point occurs (*a*). To prevent overfitting, we implemented a 10-fold cross-validation scheme, which yielded reliable predictions on the left-out data (Pearson r = 0.71 to 0.88 for all *trans* genes in the *GFI1B*, *MYB*, and *NFE2* networks, **Figure S9C**). Additionally, the predicted parameter *a* was centred around zero, as expected since the input data represents relative fold changes (**Figure S10**). Since a sigmoid function cannot capture non-monotonic responses, we employed a loess regression as an alternative approach for the few genes that exhibited non-monotonic responses (see Methods, **Figure S9D, E**). For the vast majority of genes, the sigmoid (or loess) fit was remarkably good, partially due to the low level of noise in the targeted scRNA-seq data.

We compared the performance of the linear vs. nonlinear models with the ΔAIC (AIC_{linear} - AIC_r), where a positive ΔAIC means that the sigmoid model captures the variance better in the dosage response than in the linear model. This showed that most *GFI1B*-dependent dosage expression responses are better fit by the sigmoid model (median ΔAIC = 18.7, with 70.4% of all *trans* genes with a significant response having ΔAIC >2, **Figure 3D**). The responses to dosage modulation of *MYB* and *NFE2* were also better captured by the nonlinearities, but to a lesser extent (0.14 and 3.4 median ΔAIC, with 20.8% and 40.7% of all

trans genes dosage responses having ΔAIC > 2 for *MYB* and *NFE2*, respectively, **Figure S9A**). The broader range of *GFI1B* expression modulation, providing more data to detect nonlinear trends, likely contributes to this difference. When ignoring those genes classified as unresponsive (genes that their expression did not change upon the TF modulator, see Methods), even more responses of the remaining *trans* genes were better explained by a sigmoidal model with 83.6%, 26.3% and 63.2% of these having a ΔAIC > 2, for *GFI1B*, *MYB* and *NFE2* respectively. A similar trend holds even when limiting the models to be fitted to those data points that correspond to a hypothetical one copy loss or gain of the *cis* gene (**Figure S9B**), where the median ΔAIC of responsive genes are 7.05, 0.05, and 3.6 for *GFI1B*, *MYB,* and *NFE2 trans* responses. Overall, this shows that *trans* responses to TF dosage are dominated by nonlinear behaviours even when the TF dosage changes are not extreme but within biologically plausible ranges.

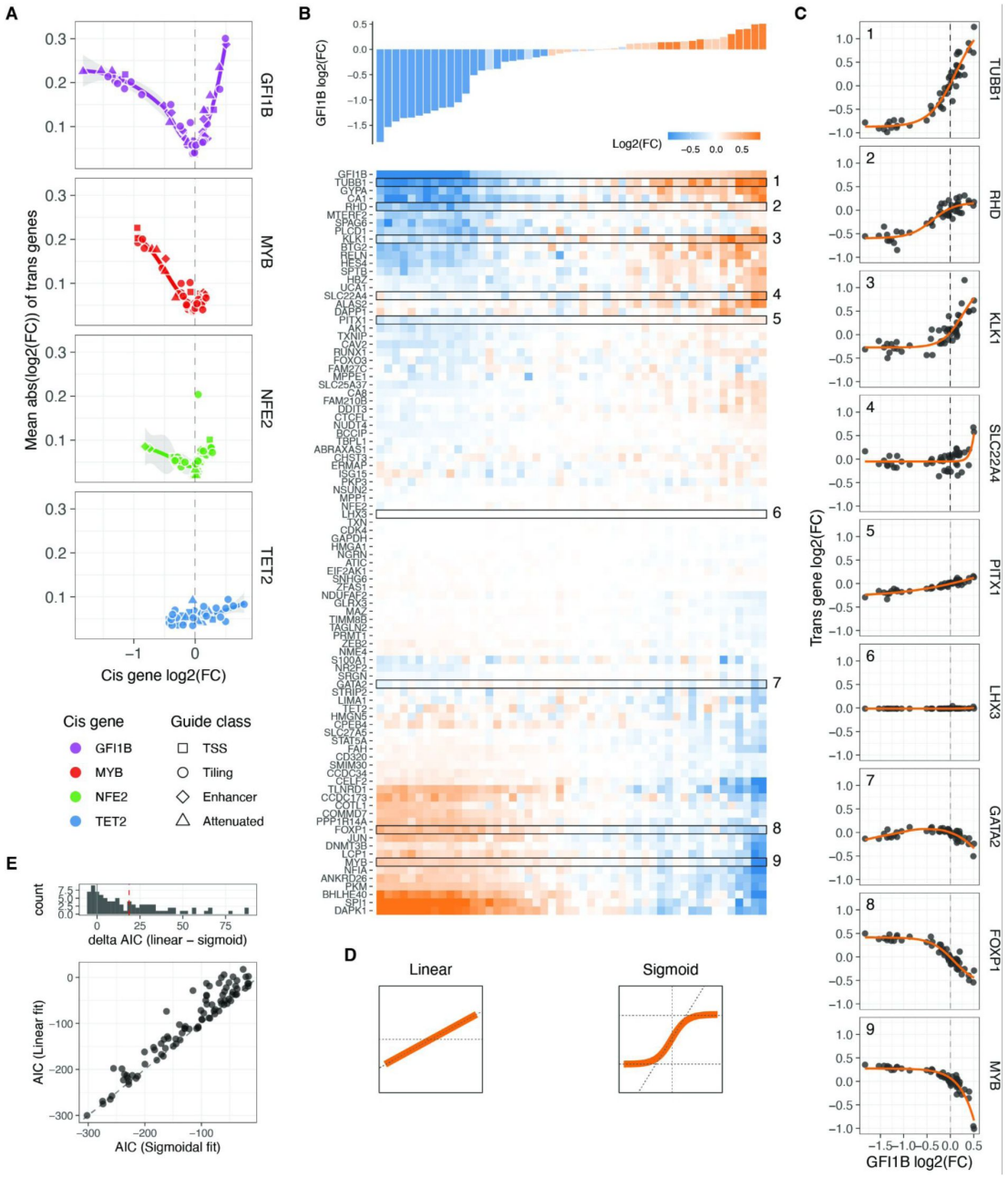

Figure 3

Trans responses of transcription factor dosage modulation

A. Average absolute expression change of all trans genes relative to the changes in expression of the cis genes. B. Changes in relative expression of all trans genes (bottom heatmap) in response to GFI1B expression changes (top barplot) upon each distinct targeted sgRNA perturbation, in comparison to NTC cells. The rows of the heatmap (trans genes) are hierarchically clustered based on their expression fold change linked to alterations in GFI1B dosage. Highlighted rows are selected dosage response examples shown in panel C. C. Dosage response curves of the highlighted trans gene in B as a function of changes in GFI1B expression. The orange line represents the sigmoid model fit, except for GATA2, which display a non-monotonic response and are fitted with a loess curve. D. Illustration of the linear and sigmoid models and equations used to fit the dosage response curves. E. Distribution of the difference in Akaike Information Criterion ($\Delta AIC_{linear\text{-}sigmoid}$) after fitting the sigmoidal or linear model for each trans gene upon GFI1B dosage modulation (top panel), and the direct comparison of the AIC of each fit (bottom panel).

Gene and transcriptional network properties of dosage response

Utilising a model that effectively captures the variance in our data provided the ability to predict unmeasured TF dosage points and facilitated a direct comparison of *trans* effects across different cis genes. Employing the sigmoid model (and loess for those with non-monotonic responses), we estimated the continuous expression of *trans* genes on a uniform fold-change scale across the spectrum of *GFI1B*, *MYB*, and *NFE2* expression changes (**Figure 4A**). This estimation was carried out within the empirically observed range of all three *cis* genes, spanning from log2(FC) -1.83 to 0.51. Subsequent hierarchical clustering of *trans* gene responses revealed six major clusters of distinct response patterns. For the majority of *trans* genes, the response to *GFI1B* and *MYB* was opposite, with only two small clusters displaying exceptions. Notably, *GFI1B* generally induced the most substantial response, while *NFE2* triggered the smallest range of *trans* gene response.

Next, we collected diverse annotations for the *trans* genes to explore the connections between their regulatory properties, disease associations, and selective constraints concerning their response to TF dosage (**Figure 4B, C**). To quantify these relationships, we assessed significant differences in belonging to these qualitative annotations using the Wilcoxon rank test (**Figure 4D**) and correlated parameters from the sigmoid model with quantitative gene metrics (**Figure 4E**). We hypothesised that genes with annotated selective constraint, numerous regulatory elements, and central positions in regulatory networks would exhibit greater robustness to TF changes. Indeed, housekeeping genes demonstrated a considerably smaller dosage response range (**Figure 4F**). Moreover, genes classified as unresponsive were enriched in the housekeeping category (odds ratio = 2.14, Fisher test p-value = 0.024). The link between selective constraint and response properties is most apparent in the *MYB* trans network. Specifically, the probability of haploinsufficiency (pHaplo) shows a significant negative correlation with the dynamic range of transcriptional responses (**Figure 4G**): genes under stronger constraint (higher pHaplo) display smaller dynamic ranges, indicating that dosage-sensitive genes are more tightly buffered against changes in *MYB* levels. This pattern was not reproduced in the other trans networks (**Figure 4E**).

The relationship between the response of *trans* genes and intrinsic gene properties differed between *GFI1B*, *MYB*, and *NFE2 trans* network responses. We also performed a similar analysis comparing the sigmoid parameters to network properties using the approach outlined by Minaeva et al.[40] and obtained inconsistent results between TF regulons (Figure S11A, B). This suggests that the link between commonly annotated gene properties and the gene responses are complex and highly context specific, as in our data from a single cell line, they differed between the upstream regulators that were manipulated. Thus, much more data is needed before transcriptional responses can be predicted from gene properties, and conversely, to understand the cellular mechanisms that lead to the annotated gene properties.

Nonlinear dosage responses in complex traits and disease

Moving beyond the characterization of mechanisms of transcriptome regulation, a key question is how gradual dosage variation links to downstream cellular phenotypes, and whether these responses exhibit analogous nonlinear patterns. To address this question, we correlated our findings with the expression profiles of various cell types in order to study the myeloid differentiation process, a phenotype well-characterised for our K562 model that has been used as a reliable system for investigating erythroid differentiation within myeloid lineages [41] and blood tumours [42]. Specifically, leveraging single-cell expression data for bone marrow cell types from the Human Cell Atlas and Human Biomolecular Atlas Project [43], we filtered the expression data to the targeted genes in our study. After aggregating data across donors and normalising expression across cell types (Figure S12A), we compared the expression patterns resulting from each unique transcription factor dosage modulation in relation to each unique cell type expression

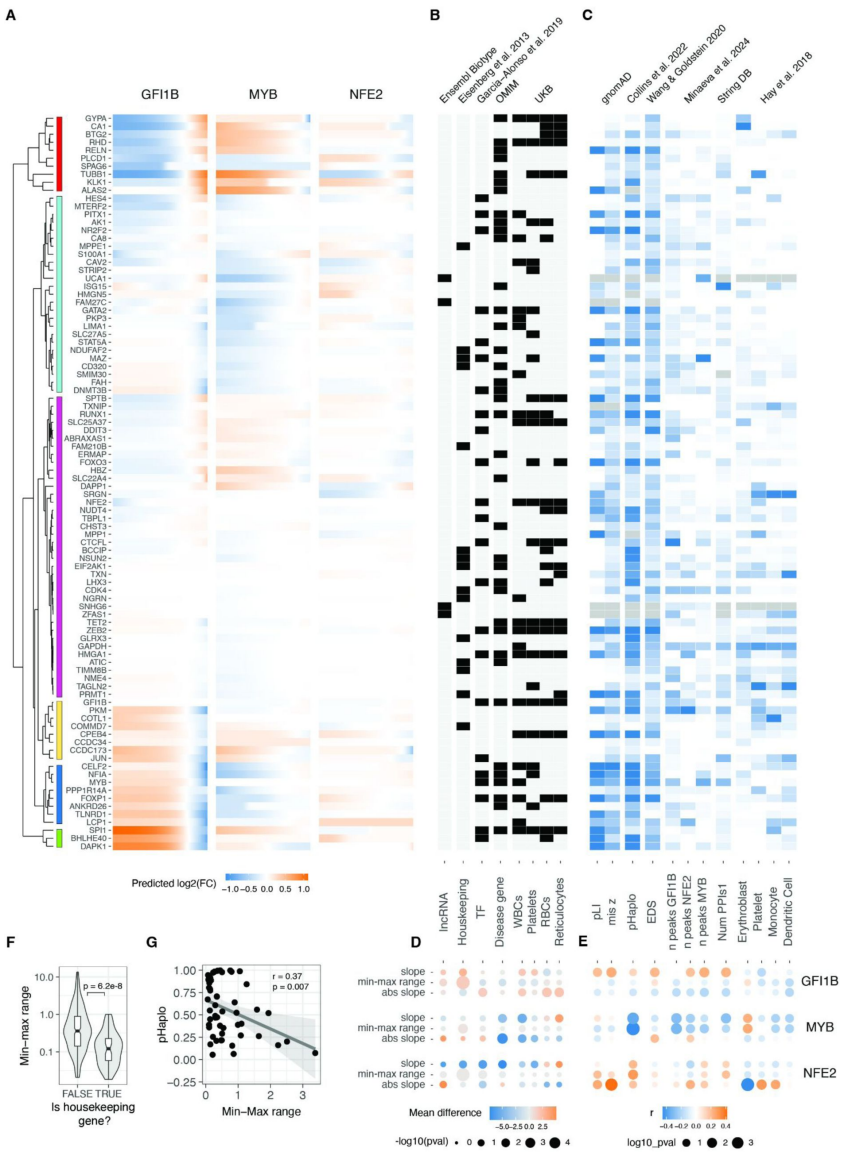

Figure 4

Relationship between gene and dosage response properties

A. Predicted changes (using sigmoid or loess fits for monotonic and non-monotonic responses, respectively) in relative expression of all trans genes in response to changes of the GFI1B, MYB and NFE2 expression. Trans genes (rows) were hierarchically clustered based on their expression fold change linked to alterations of all TF's dosage. Dendrogram of the resulting clustering shown in the left. B. Heatmap showing the qualitative properties of each trans gene. The x-axis indicates specific gene features. The top labels specify the source of the data, while the bottom labels describe the corresponding gene properties. WBCs, platelets, RBCs, and reticulocytes refer to GWAS of white blood cells, platelets, red blood cells, and reticulocytes, respectively. C. Heatmap indicating the z-scaled quantitative gene features of each transgene. The x-axis indicates specific gene features. The top labels specify the source of the data, while the bottom labels describe the corresponding gene properties. Erythroblast, platelets, monocytes, and dendritic cells refer to cell types from Hay et al. 2018. Grey cells indicate missing data. D. Difference in the average value of the sigmoid parameter indicated in right between the genes qualified into the no/yes category of the gene properties indicated in B. E. Pearson correlation coefficient of the quantitative trans gene features (shown in C) with the sigmoid parameter value for each trans gene in the response of the modulation of dosage of the TF indicated on the left. Size of the points are inversely related to significance of correlation, and colour indicates the direction of correlation. F. Differences in the range of expression response for Housekeeping vs. non-Housekeeping trans genes with changes of dosage of MYB, GFI1B and NFE2. G. Negative correlation between haploinsufficiency score (pHaplo) and the range of the response of trans genes to the modulation of MYB.

state. The ensuing correlation can then be construed as a "phenotype," signifying the similarity between the transcriptional state induced by the TF increase or decrease and the transcriptional state of a specific blood lineage cell type.

Such analyses recapitulate known biology, with *GFI1B* upregulation [29] and *MYB* downregulation [44] being crucial factors promoting erythrocyte maturation (**Figure 5A**). The downregulation of *NFE2* instead was negatively related to platelet differentiation. Analysing the correlations as inferred phenotypes suggests potential nonlinear relationships (**Figure S12B**), but these trends should be considered hypotheses that require experimental validation. In summary, this points to cellular phenotypes resulting from gradual TF dosage modulation.

Many of the analysed *trans* genes are associated with physiological traits and diseases (**Figure 4**). Understanding the nonlinear trends in the expression of these genes is of particular interest, as it helps comprehend how genes with physiological impacts may be buffered against upstream regulatory changes, and how their dosage changes as a response to upstream regulators contrasts with genetic variants that contribute to diseases and traits. Additionally, knowing the underlying dosage-to-phenotype curve of a gene can be crucial if this is considered a biomarker for identifying or treating disease. To investigate this, we analysed whether OMIM genes for rare diseases and Mendelian traits or GWAS genes for different blood cell traits (**Figure 5B**) that are part of the *trans* networks of genes affected by *GFI1B*, *MYB,* or *NFE2* perturbation are enriched for nonlinear dosage responses. As seen in **Figure 4**, the *trans* response properties of each gene are highly specific to the regulators and thus analysed in parallel for each *cis* gene network. An enrichment for nonlinear responses was observed for *MYB trans* network genes associated with disease and for blood traits related to white blood cells and reticulocytes. These enrichments are particularly interesting given that most *trans* genes that were sensitive to *MYB* dosage modulation did not respond with a nonlinear trend (**Figure S9A**).

Despite nonlinear responses not being significantly enriched among disease genes across all *trans* networks, the responses of the same *trans* gene can show very different dosage responses depending on the upstream regulator being tuned. In **Figure 5C**, we highlight several disease-associated genes (linked to one or more disease phenotypes [45]. *FOXP1*, a haploinsufficient and potentially triplosensitive transcription factor implicated in intellectual disability, exhibited a strong and dose-dependent response, particularly to varying levels of *GFI1B*. A similar pattern was observed for *NFIA*, another haploinsufficient gene involved in developmental disorders.. However, it is difficult to interpret their expression response in K562 cells when their most apparent phenotypic effects likely derive from other cell types. *RHB* is the Rhesus blood type gene, where a common deletion of the gene causes the Rh-blood type in homozygous individuals, with a strong nonlinear response to *GFI1B* levels. A particularly interesting gene is *TUBB1*, part of β-tubulin, that causes autosomal dominant macrothrombocytopenia or abnormally large platelets. Here, K562 cells are a reasonable model system, being closely related precursors to megakaryocytes that produce platelets. Interestingly, *GFI1B* loss also causes a macrothrombocytopenia phenotype in mice [46], and in our data, *TUBB1* expression decreases quickly as a function of decreased *GFI1B* expression but then plateaus at a level that corresponds to loss of one copy of *TUBB1*. This raises the hypothesis that low *GFI1B* levels may cause macrothrombocytopenia at least partially via reducing *TUBB1* expression.

Discussion

In this paper, we have investigated how gradual dosage modulation of transcription factors contributes to dosage-sensitive transcriptional regulation and investigated its potential phenotypic consequences. First, we set up an easily scalable and generalizable CRISPRi/CRISPRa approach with tiling sgRNAs for gradual titration of gene expression, with reagents that can be designed with data only of the TSS and easily ordered at scale. Alternative approaches that rely e.g., on

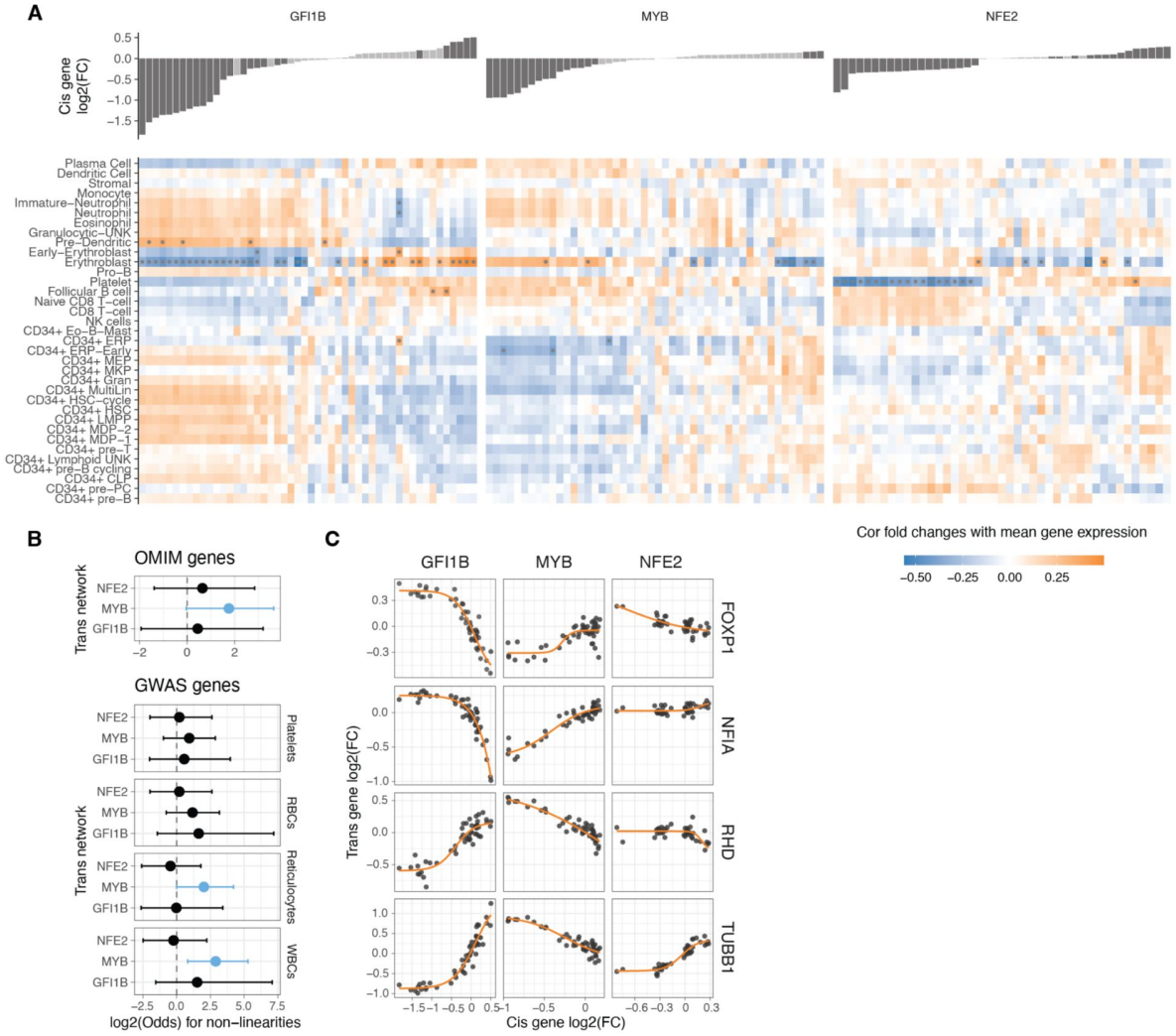

Figure 5

Non-linearities in TF dosage responses of complex traits and disease genes

A. Heatmap illustrating the correlation between the mean expression of cell types and the changes in expression linked to individual TF dosage perturbations. The barplot on the top panel represents cis gene dosage perturbation. Asterisks (*) denote correlations with 10% FDR. B. Enrichment log(odds) ratio of non-linear TF dosage responses ($\Delta AIC_{linear\text{-}sigmoid} > 0$) in disease related genes (OMIM genes linked to 1 or more diseases, top panel) or in GWAS blood traits associated genes (closest expressed gene to lead GWAS variant, bottom panel). Log(odds) with Fisher's exact test at FDR < 0.05 are highlighted in blue. C. Examples of TF dosage response curves of genes both associated with disease (OMIM) and complex traits (Blood GWAS).

targeting CREs that are often unknown, dramatic overexpression, or laborious setting up of constructs for each gene are less practical for large-scale analyses. Our approach appears best suited for expression modulation in the biologically reasonable range, and other methods would be needed for dramatic overexpression or complete silencing of the target genes. Our inability to substantially increase *MYB* expression indicates the need for further work and larger *cis* gene sets to fully understand how widespread this is and to what extent this depends on *cis*-regulatory properties versus feedback and buffering mechanisms. Nevertheless, we believe that the approach proposed here is a useful complement to the diversifying set of tools for dosage modulation for different purposes [18–22].

In this work, we made use of targeted transcriptome sequencing to avoid complications from the sparsity of single-cell data. While highly accurate targeted readout of the *cis* gene expression linked to each sgRNA is a core component of our approach, analysing *trans* responses could also be achieved by standard single-cell sequencing of the full transcriptome, possibly in combination with a targeted readout of transcripts of particular interest. In this study, the targeted genes were selected based on prior data of responding to *GFI1B*, *NFE2,* or *MYB* regulation and thus do not represent an unbiased or random sample of genes. An interesting future extension would be the addition of single-cell protein quantification to confirm that the detected mRNA levels correspond to protein levels, but this remains technically challenging.

Our results show that nonlinear responses to gradual up- and down-regulation of TF dosage are widespread and can be detected even without extreme overexpression or full knockout of the TF. The patterns of transcriptional responses are highly context-specific and vary between upstream regulators. Further work with larger sets of *cis* and *trans* genes, as well as direct quantification of cellular readouts, will be needed to fully characterise the patterns and mechanisms of downstream impacts on gene dosage. However, our findings indicate important directions for future research. First, the widespread nonlinearity suggests that inferring gene function from classical molecular biology approaches—such as drastic knockouts or knockdowns—may be limited, as these perturbations can produce effects that are both quantitatively and qualitatively different from the more modest changes that occur naturally. This may be particularly relevant for essential and highly dosage-sensitive genes, where applying our gradual dosage modulation framework can provide opportunities for functional characterization at perturbation levels that do not kill the cells. Secondly, we show that the effects of up- and downregulation are qualitatively and quantitatively different, which calls for increased attention to analysing both directions of effect, which also occur in natural responses to genetic variants and environmental stimuli.

Gene dosage sensitivity has typically been studied by human genetics and genomics methods [47–49]. The experimental approach pursued in this study and the computational approaches are fundamentally different and complement each other: while human genetics is powerful for capturing the functional importance of physiological phenotypes via patterns of population variation and selective constraint, experimental approaches provide more granularity and insights into cellular mechanisms. Furthermore, while the convergence of disease effects of common and rare variants affecting the same gene is a well-known phenomenon [6,7], the sparsity of variants makes it difficult to properly model allelic series as a continuous dosage-to-phenotype function for individual genes. Experimental approaches can provide a powerful complement to this. Altogether, we envision that combining these perspectives into true systems genetics approaches will be a powerful way to understand how gene dosage variation contributes to human phenotypes from molecular to cellular and eventually physiological levels.

Data availability and materials

All code used in this study is available at *https://github.com/LappalainenLab/d2n_ms*. Raw sequencing data has been submitted to GEO (accession number GSE257547).

Acknowledgements

This work was funded by NIH grants R01MH106842, R01AG057422, DP2HG010099, R01HG012790, and R01GM122924; a grant from the Knut and Alice Wallenberg Foundation to SciLifeLab for research in Data-driven Life Science, DDLS (KAW 2020.0239); funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101043238); a European Molecular Biology Organization Postdoctoral Fellowship (ALTF 345-2021) to J.D.; a Canadian Institutes of Health Research Banting Postdoctoral Fellowship and NIH/NHGRI (K99HG012792) to J.A.M.

Additional information

Author Contributions

J.D. and T.L. conceived the study. J.D. performed the experiments, with contributions from J.A.M. and M.Z.. N.S. contributed experimental resources to the study. J.D. performed the computational analyses, with contributions from M.M., S.G. and J.A.M.. J.D. and T.L. wrote the manuscript with contributions and review from all the authors.

Funding

National Institutes of Health (R01MH106842)

National Institutes of Health (R01AG057422)

National Institutes of Health (DP2HG010099)

National Institutes of Health (R01HG012790)

National Institutes of Health (R01GM122924)

National Institutes of Health (K99HG012792)

Knut and Alice Wallenberg Foundation (KAW 2020.0239)

European Research Council

https://doi.org/10.3030/101043238

European Molecular Biology Organization (ALTF 345-2021)

Canadian Institutes of Health Research (Banting Postdoctoral Fellowship)

- John A Morris

Additional files

Supplemental material

References

1. Maurano M. T., et al. (2012) **Systematic localization of common disease-associated variation in regulatory DNA** *Science* **337**:1190–1195 Google Scholar

2. GTEx Consortium (2020) **The GTEx Consortium atlas of genetic regulatory effects across human tissues** *Science* **369**:1318–1330 Google Scholar

3. Claussnitzer M., et al. (2020) **A brief history of human disease genetics** *Nature* **577**:179–189 Google Scholar

4. Morris J. A., et al. (2023) **Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens** *Science* **380**:eadh7699 Google Scholar

5. Zschocke J., Byers P. H., Wilkie A. O. M (2023) **Mendelian inheritance revisited: dominance and recessiveness in medical genetics** *Nat. Rev. Genet* **24**:442–463 Google Scholar

6. Backman J. D., et al. (2021) **Exome sequencing and analysis of 454,787 UK Biobank participants** *Nature* **599**:628–634 Google Scholar

7. Freund M. K., et al. (2018) **Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits** *Am. J. Hum. Genet* **103**:535–552 Google Scholar

8. Castel S. E., et al. (2018) **Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk** *Nat. Genet* **50**:1327–1334 Google Scholar

9. Fahed A. C., et al. (2020) **Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions** *Nat. Commun* **11**:3635 Google Scholar

10. McCaw Z. R., et al. (2023) **An allelic-series rare-variant association test for candidate-gene discovery** *Am. J. Hum. Genet* **110**:1330–1342 Google Scholar

11. Sanjana N. E (2017) **Genome-scale CRISPR pooled screens** *Anal. Biochem* **532**:95–99 Google Scholar

12. Collins R. L., et al. (2021) **A cross-disorder dosage sensitivity map of the human genome** *medRxiv* Google Scholar

13. Fehrmann R. S. N., et al. (2015) **Gene expression analysis identifies global gene dosage sensitivity in cancer** *Nat. Genet* **47**:115–125 Google Scholar

14. Wang T., et al. (2015) **Identification and characterization of essential genes in the human genome** *Science* **350**:1096–1101 Google Scholar

15. Hart T., et al. (2015) **High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities** *Cell* **163**:1515–1526 Google Scholar

16. Cowley G. S., et al. (2014) **Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies** *Sci Data* **1**:140035 Google Scholar

17. Keren L., et al. (2016) **Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness** *Cell* **166**:1282–1294 https://doi.org/10.1016/j.cell.2016.07.024 | Google Scholar

18. Jost M., et al. (2020) **Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs** *Nat. Biotechnol* **38**:355–364 Google Scholar

19. Hawkins J. S., et al. (2020) **Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in Escherichia coli and Bacillus subtilis** *Cell Syst* **11**:523–535 Google Scholar

20. Noviello G., Gjaltema R. A. F., Schulz E. G (2023) **CasTuner is a degron and CRISPR/Cas-based toolkit for analog tuning of endogenous gene expression** *Nat. Commun* **14**:3225 Google Scholar

21. Chiarella A. M., et al. (2020) **Dose-dependent activation of gene expression is achieved using CRISPR and small molecules that recruit endogenous chromatin machinery** *Nat. Biotechnol* **38**:50–55 Google Scholar

22. Liu W., et al. (2024) **Dissecting reprogramming heterogeneity at single-cell resolution using scTF-seq** *bioRxiv* https://doi.org/10.1101/2024.01.30.577921 | Google Scholar

23. Lalanne J.-B., Parker D. J., Li G.-W (2021) **Spurious regulatory connections dictate the expression-fitness landscape of translation factors** *Mol. Syst. Biol* **17**:e10302 Google Scholar

24. Naqvi S., et al. (2023) **Precise modulation of transcription factor levels identifies features underlying dosage sensitivity** *Nat. Genet* https://doi.org/10.1038/s41588-023-01366-2 | Google Scholar

25. Pulice J. L., Meyerson M. (2023) **Dosage amplification dictates oncogenic regulation by the NKX2-1 lineage factor in lung adenocarcinoma** *bioRxiv* https://doi.org/10.1101/2023.10.26.563996 | Google Scholar

26. Mostafavi H., Spence J. P., Naqvi S., Pritchard J. K (2023) **Systematic differences in discovery of genetic effects on gene expression and complex traits** *Nat. Genet* **55**:1866–1875 Google Scholar

27. van der Lee R., Correard S., Wasserman W. W (2020) **Deregulated Regulators: Disease-Causing cis Variants in Transcription Factor Genes** *Trends Genet* **36**:523–539 Google Scholar

28. Ulirsch J. C., et al. (2016) **Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits** *Cell* **165**:1530–1545 Google Scholar

29. Möröy T., Vassen L., Wilkes B., Khandanpour C (2015) **From cytopenia to leukemia: the role of Gfi1 and Gfi1b in blood formation** *Blood* **126**:2561–2569 Google Scholar

30. Polfus L. M., et al. (2016) **Whole-Exome Sequencing Identifies Loci Associated with Blood Cell Traits and Reveals a Role for Alternative GFI1B Splice Variants in Human Hematopoiesis** *Am. J. Hum. Genet* **99**:481–488 Google Scholar

31. Jutzi J. S., et al. (2019) **Altered NFE2 activity predisposes to leukemic transformation and myelosarcoma with AML-specific aberrations** *Blood* **133**:1766–1777 Google Scholar

32. Baker S. J., et al. (2014) **B-myb is an essential regulator of hematopoietic stem cell and myeloid progenitor cell development** *Proc. Natl. Acad. Sci. U. S. A* **111**:3122–3127 Google Scholar

33. Doench J. G., et al. (2014) **Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation** *Nat. Biotechnol* **32**:1262–1267 Google Scholar

34. Doench J. G., et al. (2016) **Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9** *Nat. Biotechnol* **34**:184–191 Google Scholar

35. McKenna A., Shendure J (2018) **FlashFry: a fast and flexible tool for large-scale CRISPR target design** *BMC Biol* **16**:74 Google Scholar

36. Sanson K. R., et al. (2018) **Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities** *Nat. Commun* **9**:5416 Google Scholar

37. Legut M., et al. (2020) **High-Throughput Screens of PAM-Flexible Cas9 Variants for Gene Knockout and Transcriptional Modulation** *Cell Rep* **30**:2859–2868 Google Scholar

38. Luo Y., et al. (2020) **New developments on the Encyclopedia of DNA Elements (ENCODE) data portal** *Nucleic Acids Res* **48**:D882–D889 Google Scholar

39. Lupo A., et al. (2013) **KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions** *Curr. Genomics* **14**:268–278 Google Scholar

40. Minaeva M., Domingo J., Rentzsch P., Lappalainen T (2025) **Specifying cellular context of transcription factor regulons for exploring context-specific gene regulation programs** *NAR Genom Bioinform* **7**:lqae178 https://doi.org/10.1093/nargab/lqae178 | Google Scholar

41. Wang H., et al. (2013) **Dynamic transcriptomes of human myeloid leukemia cells** *Genomics* **102**:250–256 Google Scholar

42. Salvadores M., Fuster-Tormo F., Supek F (2020) **Matching cell lines with cancer type and subtype of origin via mutational, epigenomic, and transcriptomic patterns** *Sci Adv* **6** Google Scholar

43. Granja J. M., et al. (2019) **Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia** *Nat. Biotechnol* **37**:1458–1465 Google Scholar

44. Jayapal S. R., et al. (2010) **Down-regulation of Myc is essential for terminal erythroid maturation** *J. Biol. Chem* **285**:40252–40265 Google Scholar

45. Amberger J. S., Bocchini C. A., Schiettecatte F., Scott A. F., Hamosh A. OMIM (2015) **org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders** *Nucleic Acids Res* **43**:D789–98 Google Scholar

46. Beauchemin H., et al. (2017) **Gfi1b controls integrin signaling-dependent cytoskeleton dynamics and organization in megakaryocytes** *Haematologica* **102**:484–497 Google Scholar

47. Collins R. L., et al. (2020) **A structural variation reference for medical and population genetics** *Nature* **581**:444–451 Google Scholar

48. Mohammadi P., et al. (2019) **Genetic regulatory variation in populations informs transcriptome analysis in rare disease** *Science* **366**:351–356 Google Scholar

49. Dong D., et al. (2023) **An RNA-informed dosage sensitivity map reflects the intrinsic functional nature of genes** *Am. J. Hum. Genet* **110**:1509–1521 Google Scholar

Author information

Júlia Domingo[†]
New York Genome Center, New York, United States
ORCID iD: 0000-0001-6359-2785

For correspondence: jdomingo@nygenome.org
[†]Present address: Allostery Exploration Technologies, S.L., Barcelona, Spain

Mariia Minaeva[#]
Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden
ORCID iD: 0000-0001-7028-1264

[#]Present address: Institute of Computational Biology, Helmholtz Center, Munich, Germany

John A Morris
New York Genome Center, New York, United States, Department of Biology, New York University, New York, United States
ORCID iD: 0000-0003-2769-8202

Samuel Ghatan
New York Genome Center, New York, United States

Marcello Ziosi
New York Genome Center, New York, United States
ORCID iD: 0000-0001-5128-6158

Neville E Sanjana
New York Genome Center, New York, United States, Department of Biology, New York University, New York, United States
ORCID iD: 0000-0002-1504-0027

Tuuli Lappalainen
New York Genome Center, New York, United States, Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden
ORCID iD: 0000-0002-7746-8109

For correspondence: tlappalainen@nygenome.org

Editors

Reviewing Editor
Murim Choi
Seoul National University, Seoul, Republic of Korea

Senior Editor
Murim Choi
Seoul National University, Seoul, Republic of Korea

Reviewer #1 (Public review):

In this manuscript, Domingo et al. present a novel perturbation-based approach to experimentally modulate the dosage of genes in cell lines. Their approach is capable of gradually increasing and decreasing gene expression. The authors then use their approach to perturb three key transcription factors and measure the downstream effects on gene expression. Their analysis of the dosage response curve of downstream genes reveals marked non-linearity.

One of the strengths of this study is that many of the perturbations fall within the physiological range for each cis gene. This range is presumably between a single-copy state of heterozygous loss-of-function (log fold change of -1) and a three-copy state (log fold change of ~0.6). This is in contrast with CRISPRi or CRISPRa studies that attempt to maximize the effect of the perturbation, which may result in downstream effects that are not representative of physiological responses.

Another strength of the study is that various points along the dosage-response curve were assayed for each perturbed gene. This allowed the authors to effectively characterize the degree of linearity and monotonicity of each dosage-response relationship. Ultimately, the study revealed that many of these relationships are non-linear, and that the response to activation can be dramatically different than the response to inhibition.

To test their ability to gradually modulate dosage, the authors chose to measure three transcription factors and around 80 known downstream targets. As the authors themselves point out in their discussion about MYB, this biased sample of genes makes it unclear how this approach would generalize genome-wide. In addition, the data generated from this small sample of genes may not represent genome-wide patterns of dosage response. Nevertheless, this unique data set and approach represents a first step in understanding dosage-response relationships between genes.

Another point of general concern in such screens is the use of the immortalized K562 cell line. It is unclear how the biology of these cell lines translates to the in vivo biology of primary cells. However, the authors do follow up with cell-type-specific analyses (Figures 4B, 4C, and 5A) to draw correspondence between their perturbation results and the relevant biology in primary cells and complex diseases.

The conclusions of the study are generally well supported with statistical analysis throughout the manuscript. As an example, the authors utilize well-known model selection methods to identify when there was evidence for non-linear dosage response relationships.

Gradual modulation of gene dosage is a useful approach to model physiological variation in dosage. Experimental perturbation screens that use CRISPR inhibition or activation often use guide RNAs targeting the transcription start site to maximize their effect on gene expression. Generating a physiological range of variation will allow others to better model physiological conditions.

There is broad interest in the field to identify gene regulatory networks using experimental perturbation approaches. The data from this study provides a good resource for such analytical approaches, especially since both inhibition and activation were tested. In addition, these data provide a nuanced, continuous representation of the relationship between effectors and downstream targets, which may play a role in the development of more rigorous regulatory networks.

Human geneticists often focus on loss-of-function variants, which represent natural knock-down experiments, to determine the role of a gene in the biology of a trait. This study demonstrates that dosage response relationships are often non-linear, meaning that the effect of a loss-of-function variant may not necessarily carry information about increases in gene dosage. For the field, this implies that others should continue to focus on both inhibition and activation to fully characterize the relationship between gene and trait.

Comments on revisions:

Thank you for responding to our comments. We have no further comments for the authors.

https://doi.org/10.7554/eLife.100555.2.sa2

Reviewer #2 (Public review):

Summary:

This work investigates transcriptional responses to varying levels of transcription factors (TFs). The authors aim for gradual up- and down-regulation of three transcription factors GFI1B, NFE2 and MYB in K562 cells, by using a CRISPRa- and a CRISPRi line, together with sgRNAs of varying potency. Targeted single-cell RNA sequencing is then used to measure gene expression of a set of 90 genes, which were previously shown to be downstream of GFI1B and NFE2 regulation. This is followed by an extensive computational analysis of the scRNA-seq dataset. By grouping cells with the same perturbations, the authors can obtain groups of cells with varying average TF expression levels. The achieved perturbations are generally subtle, not reaching half or double doses for most samples, and up-regulation is generally weak below 1.5-fold in most cases. Even in this small range, many target genes exhibit a non-linear response. Since this is rather unexpected, it is crucial to rule out technical reasons for these observations.

Strengths:

The work showcases how a single dataset of CRISPRi/a perturbations with scRNA-seq readout and an extended computational analysis can be used to estimate transcriptome dose-responses, a general approach that likely can be built upon in the future.
Moreover, the authors highlight tiling of sgRNAs +/-1000bp around TSS as a useful approach. Compared with conventional direct TSS-targeting (+/- 200 bp), the larger sequence window allows placing more sgRNAs. Also it requires little prior knowledge of CREs, and avoids using "attenuated" sgRNAs which would require specialized sgRNA design.

Weaknesses:

The experiment was performed in a single replicate and it would have been reassuring to see an independent validation of the main findings, for example through measuring individual dose-response curves .

Much of the analysis depends on the estimation of log-fold changes between groups of single cells with non-targeting controls and those carrying a guide RNA driving a specific knockdown. Generally, biological replicates are recommended for differential gene expression testing (Squair et al. 2021, https://doi.org/10.1038/s41467-021-25960-2). When using the FindMarkers function from the Seurat package, the authors divert from the recommendations for pseudo-bulk analysis to aggregate the raw counts (https://satijalab.org/seurat/articles/de_vignette.html). Furthermore, differential gene expression analysis of scRNA-seq data can suffer from mis-estimations (Nguyen et al. 2023, https://doi.org/10.1038/s41467-023-37126-3), and different computational tools or versions can affect these estimates

strongly (Pullin et al. 2024, https://doi.org/10.1186/s13059-024-03183-0 and Rich et al. 2024, https://doi.org/10.1101/2024.04.04.588111). Therefore it would be important to describe more precisely in the Methods how this analysis was performed, any deviations from default parameters, package versions, and at which point which values were aggregated to form "pseudobulk" samples.

Two different cell lines are used to construct dose-response curves, where a CRISPRi line allows gene down-regulation and the CRISPRa line allows gene upregulation. Although both lines are derived from the same parental line (K562) the expression analysis of Tet2, which is absent in the CRISPRi line, but expressed in the CRISPRa line (Fig. S1F, S3A) suggests clonal differences between the two lines. Similarly, the UMAP in S3C and the PCA in S4A suggest batch effects between the two lines. These might confound this analysis, even though all fold changes are calculated relative to the baseline expression in the respective cell line (NTC cells). Combining log2-fold changes from the two cell lines with different baseline expression into a single curve (e.g. Fig. 3) remains misleading, because different data points could be normalized to different base line expression levels.

The study estimates the relationship between TF dose and target gene expression. This requires a system that allows quantitative changes in TF expression. The data provided does not convincingly show that this condition is met, which however is an essential prerequisite for the presented conclusions. Specifically, the data shown in Fig. S3A shows that upon stronger knock-down, a subpopulation of cells appear, where the targeted TF is not detected any more (drop-outs). Also in Fig. 3B (top) suggests that the knock-down is either subtle (similar to NTCs) or strong, but intermediate knock-down (log2-FC of 0.5-1) does not occur. Although the authors argue that this is a technical effect of the scRNA-seq protocol, it is also possible that this represents a binary behavior of the CRISPRi system. Previous work has shown that CRISPRi systems with the KRAB domain largely result in binary repression and not in gradual down-regulation as suggested in this study (Bintu et al. 2016 (https://doi.org/10.1126/science.aab2956), Noviello et al. 2023 (https://doi.org/10.1038/s41467-023-38909-4)).

One of the major conclusions of the study is that non-linear behavior is common. It would be helpful to show that this observation does not arise from the technical concerns described in the previous points. This could be done for instance with independent experimental validations.

Did the authors achieve their aims? Do the results support the conclusions?:

Some of the most important conclusions, such as the claim that non-linear responses are common, are not well supported because they rely on accurately determining the quantitative responses of trans genes, which suffers from the previously mentioned concerns.

Discussion of the likely impact of the work on the field, and the utility of the methods and data to the community:

Together with other recent publications, this work emphasizes the need to study transcription factor function with quantitative perturbations. The computational code repository contains all the valuable code with inline comments, but would have benefited from a readme file explaining the repository structure, package versions, and instructions to reproduce the analyses, including which input files or directory structure would be needed.

https://doi.org/10.7554/eLife.100555.2.sa1

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public review):

In this manuscript, Domingo et al. present a novel perturbation-based approach to experimentally modulate the dosage of genes in cell lines. Their approach is capable of gradually increasing and decreasing gene expression. The authors then use their approach to perturb three key transcription factors and measure the downstream effects on gene expression. Their analysis of the dosage response curve of downstream genes reveals marked non-linearity.

One of the strengths of this study is that many of the perturbations fall within the physiological range for each cis gene. This range is presumably between a single-copy state of heterozygous loss-of-function (log fold change of -1) and a three-copy state (log fold change of ~0.6). This is in contrast with CRISPRi or CRISPRa studies that attempt to maximize the effect of the perturbation, which may result in downstream effects that are not representative of physiological responses.

Another strength of the study is that various points along the dosage-response curve were assayed for each perturbed gene. This allowed the authors to effectively characterize the degree of linearity and monotonicity of each dosage-response relationship. Ultimately, the study revealed that many of these relationships are non-linear, and that the response to activation can be dramatically different than the response to inhibition.

To test their ability to gradually modulate dosage, the authors chose to measure three transcription factors and around 80 known downstream targets. As the authors themselves point out in their discussion about MYB, this biased sample of genes makes it unclear how this approach would generalize genome-wide. In addition, the data generated from this small sample of genes may not represent genome-wide patterns of dosage response. Nevertheless, this unique data set and approach represents a first step in understanding dosage-response relationships between genes.

Another point of general concern in such screens is the use of the immortalized K562 cell line. It is unclear how the biology of these cell lines translates to the in vivo biology of primary cells. However, the authors do follow up with cell-type-specific analyses (Figures 4B, 4C, and 5A) to draw a correspondence between their perturbation results and the relevant biology in primary cells and complex diseases.

The conclusions of the study are generally well supported with statistical analysis throughout the manuscript. As an example, the authors utilize well-known model selection methods to identify when there was evidence for non-linear dosage response relationships.

Gradual modulation of gene dosage is a useful approach to model physiological variation in dosage. Experimental perturbation screens that use CRISPR inhibition or activation often use guide RNAs targeting the transcription start site to maximize their effect on gene expression. Generating a physiological range of variation will allow others to better model physiological conditions.

There is broad interest in the field to identify gene regulatory networks using experimental perturbation approaches. The data from this study provides a good resource for such analytical approaches, especially since both inhibition and activation were tested. In addition, these data provide a nuanced, continuous representation of the relationship between effectors and downstream targets, which may play a role in the development of more rigorous regulatory networks.

> *Human geneticists often focus on loss-of-function variants, which represent natural knock-down experiments, to determine the role of a gene in the biology of a trait. This study demonstrates that dosage response relationships are often non-linear, meaning that the effect of a loss-of-function variant may not necessarily carry information about increases in gene dosage. For the field, this implies that others should continue to focus on both inhibition and activation to fully characterize the relationship between gene and trait.*

We thank the reviewer for their thoughtful and thorough evaluation of our study. We appreciate their recognition of the strengths of our approach, particularly the ability to modulate gene dosage within a physiological range and to capture non-linear dosage-response relationships. We also agree with the reviewer's points regarding the limitations of gene selection and the use of K562 cells, and we are encouraged that the reviewer found our follow-up analyses and statistical framework to be well-supported. We believe this work provides a valuable foundation for future genome-wide applications and more physiologically relevant perturbation studies.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *This work investigates transcriptional responses to varying levels of transcription factors (TFs). The authors aim for gradual up- and down-regulation of three transcription factors GFI1B, NFE2, and MYB in K562 cells, by using a CRISPRa- and a CRISPRi line, together with sgRNAs of varying potency. Targeted single-cell RNA sequencing is then used to measure gene expression of a set of 90 genes, which were previously shown to be downstream of GFI1B and NFE2 regulation. This is followed by an extensive computational analysis of the scRNA-seq dataset. By grouping cells with the same perturbations, the authors can obtain groups of cells with varying average TF expression levels. The achieved perturbations are generally subtle, not reaching half or double doses for most samples, and up-regulation is generally weak below 1.5-fold in most cases. Even in this small range, many target genes exhibit a non-linear response. Since this is rather unexpected, it is crucial to rule out technical reasons for these observations.*

We thank the reviewer for their detailed and thoughtful assessment of our work. We are encouraged by their recognition of the strengths of our study, including the value of quantitative CRISPR-based perturbation coupled with single-cell transcriptomics, and its potential to inform gene regulatory network inference. Below, we address each of the concerns raised:

> *Strengths:*
>
> *The work showcases how a single dataset of CRISPRi/a perturbations with scRNA-seq readout and an extended computational analysis can be used to estimate transcriptome dose responses, a general approach that likely can be built upon in the future.*
>
> *Weaknesses:*
>
> *(1) The experiment was only performed in a single replicate. In the absence of an independent validation of the main findings, the robustness of the observations remains unclear.*

We acknowledge that our study was performed in a single pooled experiment. While additional replicates would certainly strengthen the findings, in high-throughput single-cell CRISPR screens, individual cells with the same perturbation serve as effective internal

replicates. This is a common practice in the field. Nevertheless, we agree that biological replicates would help control for broader technical or environmental effects.

> *(2) The analysis is based on the calculation of log-fold changes between groups of single cells with non-targeting controls and those carrying a guide RNA driving a specific knockdown. How the fold changes were calculated exactly remains unclear, since it is only stated that the FindMarkers function from the Seurat package was used, which is likely not optimal for quantitative estimates. Furthermore, differential gene expression analysis of scRNA-seq data can suffer from data distortion and mis-estimations (Heumos et al. 2023 (https://doi.org/10.1038/s41576-023-00586-w), Nguyen et al. 2023 (https://doi.org/10.1038/s41467-023-37126-3)). In general, the pseudo-bulk approach used is suitable, but the correct treatment of drop-outs in the scRNA-seq analysis is essential.*

We thank the reviewer for highlighting recent concerns in the field. A study benchmarking association testing methods for perturb-seq data found that among existing methods, Seurat's FindMarkers function performed the best (T. Barry et al. 2024).

In the revised Methods, we now specify the formula used to calculate fold change and clarify that the estimates are derived from the Wilcoxon test implemented in Seurat's FindMarkers function. We also employed pseudo-bulk grouping to mitigate single-cell noise and dropout effects.

> *(3) Two different cell lines are used to construct dose-response curves, where a CRISPRi line allows gene down-regulation and the CRISPRa line allows gene upregulation. Although both lines are derived from the same parental line (K562) the expression analysis of Tet2, which is absent in the CRISPRi line, but expressed in the CRISPRa line (Figure S3A) suggests substantial clonal differences between the two lines. Similarly, the PCA in S4A suggests strong batch effects between the two lines. These might confound this analysis.*

We agree that baseline differences between CRISPRi and CRISPRa lines could introduce confounding effects if not appropriately controlled for. We emphasize that all comparisons are made as fold changes relative to non-targeting control (NTC) cells within each line, thereby controlling for batch- and clone-specific baseline expression. See figures S4A and S4B.

> *(4) The study uses pseudo-bulk analysis to estimate the relationship between TF dose and target gene expression. This requires a system that allows quantitative changes in TF expression. The data provided does not convincingly show that this condition is met, which however is an essential prerequisite for the presented conclusions. Specifically, the data shown in Figure S3A shows that upon stronger knock-down, a subpopulation of cells appears, where the targeted TF is not detected anymore (drop-outs). Also Figure 3B (top) suggests that the knock-down is either subtle (similar to NTCs) or strong, but intermediate knock-down (log2-FC of 0.5-1) does not occur. Although the authors argue that this is a technical effect of the scRNA-seq protocol, it is also possible that this represents a binary behavior of the CRISPRi system. Previous work has shown that CRISPRi systems with the KRAB domain largely result in binary repression and not in gradual down-regulation as suggested in this study (Bintu et al. 2016 (https://doi.org/10.1126/science.aab2956), Noviello et al. 2023 (https://doi.org/10.1038/s41467-023-38909-4)).*

Figure S3A shows normalized expression values, not fold changes. A pseudobulk approach reduces single-cell noise and dropout effects. To test whether dropout events reflect true binary repression or technical effects, we compared trans-effects across cells with zero

versus low-but-detectable target gene expression (Figure S3B). These effects were highly concordant, supporting the interpretation that dropout is largely technical in origin. We agree that KRAB-based repression can exhibit binary behavior in some contexts, but our data suggest that cells with intermediate repression exist and are biologically meaningful. In ongoing unpublished work, we pursue further analysis of these data at the single cell level, and show that for nearly all guides the dosage effects are indeed gradual rather than driven by binary effects across cells.

> *(5) One of the major conclusions of the study is that non-linear behavior is common. This is not surprising for gene up-regulation, since gene expression will reach a plateau at some point, but it is surprising to be observed for many genes upon TF down-regulation. Specifically, here the target gene responds to a small reduction of TF dose but shows the same response to a stronger knock-down. It would be essential to show that his observation does not arise from the technical concerns described in the previous point and it would require independent experimental validations.*

This phenomenon—where relatively small changes in cis gene dosage can exceed the magnitude of cis gene perturbations—is not unique to our study. This also makes biological sense, since transcription factors are known to be highly dosage sensitive and generally show a smaller range of variation than many other genes (that are regulated by TFs). Empirically, these effects have been observed in previous CRISPR perturbation screens conducted in K562 cells, including those by Morris et al. (2023), Gasperini et al. (2019), and Replogle et al. (2022), to name but a few studies that our lab has personally examined the data of.

> *(6) One of the conclusions of the study is that guide tiling is superior to other methods such as sgRNA mismatches. However, the comparison is unfair, since different numbers of guides are used in the different approaches. Relatedly, the authors point out that tiling sometimes surpassed the effects of TSS-targeting sgRNAs, however, this was the least fair comparison (2 TSS vs 10 tiling guides) and additionally depends on the accurate annotation of TSS in the relevant cell line.*

We do not draw this conclusion simply from observing the range achieved but from a more holistic observation. We would like to clarify that the number of sgRNAs used in each approach is proportional to the number of base pairs that can be targeted in each region: while the TSS-targeting strategy is typically constrained to a small window of a few dozen base pairs, tiling covers multiple kilobases upstream and downstream, resulting in more guides by design rather than by experimental bias. The guides with mismatches do not have a great performance for gradual upregulation.

We would also like to point out that the observation that the strongest effects can arise from regions outside the annotated TSS is not unique to our study and has been demonstrated in prior work (referenced in the text).

To address this concern, we have revised the text to clarify that we do not consider guide tiling to be inherently superior to other approaches such as sgRNA mismatches. Rather, we now describe tiling as a practical and straightforward strategy to obtain a wide range of gene dosage effects without requiring prior knowledge beyond the approximate location of the TSS. We believe this rephrasing more accurately reflects the intent and scope of our comparison.

> *(7) Did the authors achieve their aims? Do the results support the conclusions?: Some of the most important conclusions are not well supported because they rely on accurately determining the quantitative responses of trans genes, which suffers from the previously mentioned concerns.*

We appreciate the reviewer's concern, but we would have wished for a more detailed characterization of which conclusions are not supported, given that we believe our approach actually accounts for the major concerns raised above. We believe that the observation of non-linear effects is a robust conclusion that is also consistent with known biology, with this paper introducing new ways to analyze this phenomenon.

> *(8) Discussion of the likely impact of the work on the field, and the utility of the methods and data to the community:*

Together with other recent publications, this work emphasizes the need to study transcription factor function with quantitative perturbations. Missing documentation of the computational code repository reduces the utility of the methods and data significantly.

Documentation is included as inline comments within the R code files to guide users through the analysis workflow.

> ***Reviewer #1 (Recommendations for the authors):***
>
> *In Figure 3C (and similar plots of dosage response curves throughout the manuscript), we initially misinterpreted the plots because we assumed that the zero log fold change on the horizontal axis was in the middle of the plot. This gives the incorrect interpretation that the trans genes are insensitive to loss of GFI1B in Figure 3C, for instance. We think it may be helpful to add a line to mark the zero log fold change point, as was done in Figure 3A.*

We thank the reviewer for this helpful suggestion. To improve clarity, we have added a vertical line marking the zero log fold change point in Figure 3C and all similar dosage-response plots. We agree this makes the plots easier to interpret at a glance.

> *Similarly, for heatmaps in the style of Figure 3B, it may be nice to have a column for the non-targeting controls, which should be a white column between the perturbations that increase versus decrease GFI1B.*

We appreciate the suggestion. However, because all perturbation effects are computed relative to the non-targeting control (NTC) cells, explicitly including a separate column for NTC in the heatmap would add limited interpretive value and could unnecessarily clutter the figure. For clarity, we have emphasized in the figure legend that the fold changes are relative to the NTC baseline.

> *We found it challenging to assess the degree of uncertainty in the estimation of log fold changes throughout the paper. For example, the authors state the following on line 190: "We observed substantial differences in the effects of the same guide on the CRISPRi and CRISPRa backgrounds, with no significant correlation between cis gene fold-changes." This claim was challenging to assess because there are no horizontal or vertical error bars on any of the points in Figure 2A. If the log fold change estimates are very noisy, the data could be consistent with noisy observations of a correlated underlying process. Similarly, to our understanding, the dosage response curves are fit assuming that the cis log fold changes are fixed. If there is excessive noise in the estimation of these log fold changes, it may bias the estimated curves. It may be helpful to give an idea of the amount of estimation error in the cis log fold changes.*

We agree that assessing the uncertainty in log fold change estimates is important for interpreting both the lack of correlation between CRISPRi and CRISPRa effects (Figure 2A) and the robustness of the dosage-response modeling.

In response, we have now updated Figure 2A to include both vertical and horizontal error bars, representing the standard errors of the log2 fold-change estimates for each guide in the CRISPRi and CRISPRa conditions. These error estimates were computed based on the differential expression analysis performed using the FindMarkers function in Seurat, which models gene expression differences between perturbed and control cells. We also now clarify this in the figure legend and methods.

> *The authors mention hierarchical clustering on line 313, which identified six clusters. Although a dendrogram is provided, these clusters are not displayed in Figure 4A. We recommend displaying these clusters alongside the dendrogram.*

We have added colored bars indicating the clusters to improve the clarity. Thank you for the suggestion.

> *In Figures 4B and 4C, it was not immediately clear what some of the gene annotations meant. For example, neither the text nor the figure legend discusses what "WBCs", "Platelets", "RBCs", or "Reticulocytes" mean. It would be helpful to include this somewhere other than only the methods to make the figure more clear.*

To improve clarity, we have updated the figure legends for Figures 4B and 4C to explicitly define these abbreviations.

> *We struggled to interpret Figure 4E. Although the authors focus on the association of MYB with pHaplo, we would have appreciated some general discussion about the pattern of associations seen in the figure and what the authors expected to observe.*

We have changed the paragraph to add more exposition and clarification:

“The link between selective constraint and response properties is most apparent in the MYB trans network. Specifically, the probability of haploinsufficiency (pHaplo) shows a significant negative correlation with the dynamic range of transcriptional responses (Figure 4G): genes under stronger constraint (higher pHaplo) display smaller dynamic ranges, indicating that dosage-sensitive genes are more tightly buffered against changes in MYB levels. This pattern was not reproduced in the other trans networks (Figure 4E)”.

> *Line 71: potentially incorrect use of "rending" and incorrect sentence grammar.*

Fixed

> *Line 123: "co-expression correlation across co-expression clusters" - authors may not have intended to use "co-expression" twice.*

Original sentence was correct.

> *Line 246: "correlations" is used twice in "correlations gene-specific correlations."*

Fixed.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) To show that the approach indeed allows gradual down-regulation it would be important to quantify the know-down strength with a single-cell readout for a subset of sgRNAs individually (e.g. flowfish/protein staining flow cytometry).*

We agree that single-cell validation of knockdown strength using orthogonal approaches such as flowFISH or protein staining would provide additional support. However, such experiments fall outside the scope of the current study and are not feasible at this stage. We note that the observed transcriptomic changes and dosage responses across multiple perturbations are consistent with effective and graded modulation of gene expression.

> *(2) Similarly, an independent validation of the observed dose-response relationships, e.g. with individual sgRNAs, can be helpful to support the conclusions about non-linear responses.*

Fig. S4C includes replication of trans-effects for a handful of guides used both in this study and in Morris et al. While further orthogonal validation of dose-response relationships would be valuable, such extensive additional work is not currently feasible within the scope of this study. Nonetheless, the high degree of replication in Fig. S4C as well as consistency of patterns observed across multiple sgRNAs and target genes provides strong support for the conclusions drawn from our high-throughput screen.

> *(3) The calculation of the log2 fold changes should be documented more precisely. To perform a pseudo-bulk analysis, the raw UMI counts should be summed up in each group (NTC, individual targeting sgRNAs), including zero counts, then the data should be normalized and the fold change should be calculated. The DESeq package for example would be useful here.*

We have updated the methods in the manuscript to provide more exposition of how the logFC was calculated:

"In our differential expression (DE) analysis, we used Seurat's FindMarkers() function, which computes the log fold change as the difference between the average normalized gene expression in each group on the natural log scale:

Logfc = log_e(mean(expression in group 1)) - log_e(mean(expression in group 2))

This is calculated in pseudobulk where cells with the same sgRNA are grouped together and the mean expression is compared to the mean expression of cells harbouring NTC guides. To calculate per-gene differential expression p-value between the two cell groups (cells with sgRNA vs cells with NTC), Wilcoxon Rank-Sum test was used".

> *(4) A more careful characterization of the cell lines used would be helpful. First, it would be useful to include the quality controls performed when the clonal lines were selected, in the manuscript. Moreover, a transcriptome analysis in comparison to the parental cell line could be performed to show that the cell lines are comparable. In addition, it could be helpful to perform the analysis of the samples separately to see how many of the response behaviors would still be observed.*

Details of the quality control steps used during the selection of the CRISPRa clonal line are already included in the Methods section, and Fig. S4A shows the transcriptome comparison of CRISPRi and CRISPRa lines also for non-targeting guides. Regarding the transcriptomic comparison with the parental cell line, we agree that such an analysis would be informative; however, this would require additional experiments that are not feasible within the scope of the current study. Finally, while analyzing the samples separately could provide further insight into response heterogeneity, we focused on identifying robust patterns across perturbations that are reproducible in our pooled screening framework. We believe these aggregate analyses capture the major response behaviors and support the conclusions drawn.

> *(5) In general we were surprised to see such strong responses in some of the trans genes, in some cases exceeding the fold changes of the cis gene perturbation more than 2x, even at the relatively modest cis gene perturbations (Figures S5-S8). How can this be explained?*

This phenomenon—where trans gene responses can exceed the magnitude of cis gene perturbations—is not unique to our study. Similar effects have been observed in previous CRISPR perturbation screens conducted in K562 cells, including those by Morris et al. (2023), Gasperini et al. (2019), and Replogle et al. (2022).

Several factors may contribute to this pattern. One possibility is that certain trans genes are highly sensitive to transcription factor dosage, and therefore exhibit amplified expression changes in response to relatively modest upstream perturbations. Transcription factors are known to be highly dosage sensitive and generally show a smaller range of variation than many other genes (that are regulated by TFs). Mechanistically, this may involve non-linear signal propagation through regulatory networks, in which intermediate regulators or feedback loops amplify the downstream transcriptional response. While our dataset cannot fully disentangle these indirect effects, the consistency of this observation across multiple studies suggests it is a common feature of transcriptional regulation in K562 cells.

> *(6) In the analysis shown in Figure S3B, the correlation between cells with zero count and >0 counts for the cis gene is calculated. For comparison, this analysis should also show the correlation between the cells with similar cis-gene expression and between truly different populations (e.g. NTC vs strong sgRNA).*

The intent of Figure S3B was not to compare biologically distinct populations or perform differential expression analyses—which we have already conducted and reported elsewhere in the manuscript—but rather to assess whether fold change estimates could be biased by differences in the baseline expression of the target gene across individual cells. Specifically, we sought to determine whether cells with zero versus non-zero expression (as can result from dropouts or binary on/off repression from the KRAB-based CRISPRi system) exhibit systematic differences that could distort fold change estimation. As such, the comparisons suggested by the reviewer do not directly relate to the goal of the analysis which Figure S3B was intended to show.

> *(7) It is unclear why the correlation between different lanes is assessed as quality control metrics in Figure S1C. This does not substitute for replicates.*

The intent of Figure S1C was not to serve as a general quality control metric, but rather to illustrate that the targeted transcript capture approach yielded consistent and specific signal across lanes. We acknowledge that this may have been unclear and have revised the relevant sentence in the text to avoid misinterpretation.

“We used the protein hashes and the dCas9 cDNA (indicating the presence or absence of the KRAB domain) to demultiplex and determine the cell line—CRISPRi or CRISPRa. Cells containing a single sgRNA were identified using a Gaussian mixture model (see Methods). Standard quality control procedures were applied to the scRNA-seq data (see Methods). To confirm that the targeted transcript capture approach worked as intended, we assessed concordance across capture lanes (Figure S1C)”.

> *(8) Figures and legends often miss important information. Figure 3B and S5-S8: what do the transparent bars represent? Figure S1A: color bar label missing. Figure S4D: what are the lines?, Figure S9A: what is the red line? In Figure S8 some of the fitted curves do not*

> *overlap with the data points, e.g. PKM. Fig. 2C: why are there more than 96 guide RNAs (see y-axis)?*

We have addressed each point as follows:

Figure 3B: The figure legend has been updated to clarify the meaning of the transparent bars.

Figures S5–S8: There are no transparent bars in these figures; we confirmed this in the source plots.

Figure S1A: The color bar label is already described in the figure legend, but we have reformulated the caption text to make this clearer.

Figure S4D: The dashed line represents a linear regression between the x and y variables. The figure caption has been updated accordingly.

Figure S9A: We clarified that the red line shows the median ΔAIC across all genes and conditions.

Figure S8: We agree that some fitted curves (e.g., PKM) do not closely follow the data points. This reflects high noise in these specific measurements; as noted in the text, TET2 is not expected to exert strong trans effects in this context.

Figure 2C: Thank you for catching this. The y-axis numbers were incorrect because the figure displays the proportion of guides (summing to 100%), not raw counts. We have corrected the y-axis label and updated the numbers in the figure to resolve this inconsistency.

> *(9) The code is deposited on Github, but documentation is missing.*

Documentation is included as inline comments within the R code files to guide users through the analysis workflow.

> *(10) The methods miss a list of sgRNA target sequences.*

We thank the reviewer for this observation. A complete table containing all processed data, including the sequences of the sgRNAs used in this study, is available at the following GEO link:

https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE257547&format=file&file=GSE257547%5Fd2n%5Fprocessed%5Fdata%2Etxt%2Egz

> *(11) In some parts, the language could be more specific and/or the readability improved, for example:*
>
> *Line 88: "quantitative landscape".*

Changed to “quantitative patterns”.

> *Lines 88-91: long sentence hard to read.*

This complex sentence was broken up into two simpler ones:

“We uncovered quantitative patterns of how gradual changes in transcription dosage lead to linear and non-linear responses in downstream genes. Many downstream genes are associated with rare and complex diseases, with potential effects on cellular phenotypes”.

> *Line 110: "tiling sgRNAs +/- 1000 bp from the TSS", could maybe be specified by adding that the average distance was around 100 or 110 bps?*
>
> *Lines 244-246: hard to understand.*

We struggle to see the issue here and are not sure how it can be reworded.

> *Lines 339-342: hard to understand.*

These sentences have been reworded to provide more clarity.

> *(12) A number of typos, and errors are found in the manuscript:*
>
> *Line 71: "SOX2" -> "SOX9".*

FIXED

> *Line 73: "rending" -> maybe "raising" or "posing"?*

FIXED

> *Line 157: "biassed".*

FIXED

> *Line 245: "exhibited correlations gene-specific correlations with".*

FIXED

> *Multiple instances, e.g. 261: "transgene" -> "trans gene".*

FIXED

> *Line 332: "not reproduced with among the other".*

FIXED

> *Figure S11: betweenness.*

This is the correct spelling

> *There are more typos that we didn't list here.*

We went through the manuscript and corrected all the spelling errors and typos.

https://doi.org/10.7554/eLife.100555.2.sa0