Genetic Analysis of a Metazoan Pathway using Transcriptomic Phenotypes

immediate

This manuscript was compiled on January 23, 2017

RNA-Seq is a technology that is commonly used to identify genetic modules that are responsive to a perturbation. In theory, global gene expression could also be used as a phenotype in complex metazoans, with all the implications that has for genetic analysis. To that end, we sequenced the transcriptome of four single mutants and two double mutants of the hypoxia pathway in *C. elegans*. We successfully analyzed the single mutants in a blinded fashion to predict the genetic relationships between the genes, and used the double mutants as a test of our predictions and to infer the directionality of the relationship. We show that genes along a pathway tend to decorrelate as a result of alternative regulatory modes and crosstalk with other pathways; and that this decorrelation accurately reflects functional distance between genes. As a by-product of our analysis, we predict 133 genes under the regulation of *hif-1*, and 36 genes under the regulation of *vhl-1*. Transcriptomic perturbations suggest an important role of *hif-1*-dependent response in chromatin remodelling in *C. elegans*. Interactive graphics for this paper can be found at www.wormlabcaltech.github.io/mprsq.

genetics | RNA-Seq | epistasis | hypoxia | transcriptomics | systems biology

Genetic analysis of molecular pathways has traditionally been performed through epistatic analysis. Epistasis occurs when two genes interact, either directly (biochemical interaction) or through a molecular pathway or physical interaction (genetic interaction). If two genes interact, and the mutants of these genes have a quantifiable phenotype, the double mutant will have a phenotype that is not the sum of the phenotypes of the single mutants that make up its genotype. Epistatic analysis remains a cornerstone of genetics today [1].

Previous work in *S. cerevisiae* and *D. discoideum* using microarrays has shown that transcriptomes can be used to infer genetic relationships in simple eukaryotes [2, 3]. Developments in the area of transcriptomics have brought forward new protocols, such as RNA-Seq [4], and have also made important progress towards cheaper sequencing [5], better and faster abundance quantification [6? , 7] and improved differential analysis of gene expression [8, 9]. As a result, RNA-Seq has been successfully used to identify genetic modules involved in a variety of processes, including T-cell regulation [10, 11], the *C. elegans* linker cell migration [12], or planarian stem cell maintenance [13, 14]. However, even in these novel applications, transcriptional profiling largely serves a descriptive role in target gene identification.

To investigate the ability of transcriptomes to serve as quantitative phenotypes, we selected mutants in the *C. elegans* hypoxia pathway for transcriptome sequencing. Metazoans depend on the presence of oxygen in sufficient concentrations to support aerobic metabolism. Genetic pathways evolved to rapidly respond to any acute or chronic changes in oxygen levels at the cellular or organismal level. These oxygen sensitive pathways are involved in a broad range of human pathologies and they have been subject to investigation biochemical and genetic approaches [15]. These approaches identified the Hypoxia Inducible Factors (HIFs) as an important group of oxygen responsive genes.

Hypoxia Inducible Factors are highly conserved in metazoans []. A common mechanism for hypoxia-response induction is heterodimerization between a HIFα and a HIFβ subunit. The heterodimer then initiates transcription of target genes []. The number and complexity of HIFs varies throughout metazoans, with humans having three HIFα subunits and two HIFβ subunits, whereas in the roundworm *Caenorhabditis elegans* (*C. elegans*) there is a single HIFα gene, *hif-1* and a single HIFβ gene, *ahr-1*. HIF target genes have been implicated in a wide variety of cellular and extracellular processes such as glycolysis, extracellular matrix modification, autophagy and immunity [].

Levels of HIFα proteins tend to be tightly regulated. Under conditions of normoxia, HIF-1α exists in the cytoplasm and partakes in a futile cycle of continuous protein production and rapid degradation with a half-life of minutes []. HIF-1α is hydroxylated by three proline hydroxylases in humans (PHD1, PHD2 and PHD3) but is only hydroxylated by one proline hydroxylase (*egl-9*) in *C. elegans* []. HIF-1 hydroxylation increases its binding affinity to Von Hippel Lindau Tumor

Significance Statement

Measurements of global gene expression are often used as descriptive tools capable of identifying genes that are downstream a perturbation. In theory, there is no reason why measurements of global transcriptomes could not be used as a quantitative phenotype for genetic analysis in multicellular organisms. In fact, qPCR measurements of single or a few reporter genes are already used to perform genetic network analysis. Here, we show that transcriptomes can be used for epistasis analysis in a metazoan, and that transcriptomes afford far more information per experiment than classic genetic analysis. By using transcriptomes as quantitative phenotypes, we can accurately predict interactions between genes, while at the same time identifying genes common to a pathway. When pathways branch, it is also possible to identify gene batteries that are associated with each end of the branch point. Finally, genes that would result in invisible visible phenotypes in an animal are not likely to be invisible at the transcriptome phenotype due to the exquisite granularity present in these structures, which represents an important advance towards studying small effect genes that make up the majority of animals' genetic repertoire.

DA, CPR and PWS designed the experiments. CPR selected the genes and extracted mRNA from all mutants. BW made the libraries. IA performed all sequencing. DA developed the mathematical theory. DA wrote all computer code and performed all analyses. DA made all the reporter strains and performed all microscopy. DA, CPR and PWS wrote the manuscript.

The authors declare no conflict of interest.

[2] To whom correspondence should be addressed. E-mail: pws@caltech.edu

www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXXXX

Suppressor 1 (VHL-1), which allows ubiquitination of HIF-1 leading to its subsequent degradation. In *C. elegans*, EGL-9 activity is inhibited by binding of CYSL-1, and CYSL-1 activity is in turn inhibited at the protein level by RHY-1, possibly by post-translational modifications to CYSL-1 [16].

Here, we show that transcriptomes contain strong, robust signals that can be used to infer relationships between genes in complex metazoans by reconstructing a the hypoxia pathway in *C. elegans* using RNA-Seq. To reconstruct this pathway, we developed new analytic mathematical tools to query genetic interactions at the transcriptome level. We used these analytic tools to generate a method for analyzing transcriptome genetics data, which we termed Metazoan RNA-based Gene Analysis (MoRGAn). We show that MoRGAn provides a sound theoretical and experimental framework with which to dissect genetic pathways. Using MoRGAn, we were able to reconstruct interactions between genes in the hypoxia pathway. Furthermore, we show that the phenomenon of phenotypic epistasis, a hallmark of genetic interaction, holds at the molecular systems level. We also demonstrate that transcriptomes contain sufficient information, under certain circumstances, to order genes in a pathway using only single mutants. Finally, we were able to identify genes that appear to be downstream of *egl-9* and *vhl-1*, but are almost certainly not targets of *hif-1*. A complete, interactive version of the analysis is also available at www.wormlabcaltech.github.io/mprsq.

Results

Development of a New Genetic Logic for Vectorial Phenotypes. Transcriptomes can be understood as information-rich phenotypes that are fine-grained representations of an organism's internal state (citation?) []. As transcriptomes have become more prominent, the single-cell community in particualr has begun to use them to define and understand cellular identities []. However, most approaches rely on data-driven, computational methods to define the important or relevant aspects of a transcriptomes and the community is actively developing new algorithms for increasingly complex data. On the other hand, whole-organism transcriptomes have not been analyzed quite so exhaustively, partially because these transcriptomes contain the identities of multiple cell-types in one convoluted measurement.

Although computational advances for RNA-seq analysis are important, an equally important facet of transcriptome analysis is the development of analytical mathematical tools with which to understand these objects. In an effort to understand transcriptome genetics at a profound level, we developed extensions to standard genetic theory borrowing from concepts in group and operator theory that enabled us to think logically about genetics using a vectorial phenotype.

Briefly, we envisioned that a single gene is always associated with the same specific transcriptome under a set of specified conditions (age of the organism, ambient temperature, food status, etc. . .). We refer to the set of genes that are differentially expressed in response to a perturbation of a gene as a specific transcriptome. A specific transcriptome can be formally thought of as a group of genes with an associated metric for each gene (such as abundance, fold change or log-fold change), or equivalently as a function that takes as input a gene ID and outputs a scalar. However, although a global transcriptome is fundamentally an observable quantity, the specific transcriptome associated with a gene is not, since we do not know the identities of the genes that incorporate it. In order to observe a specific transcriptome, it is necessary to perform at least two experiments: first, we must measure a basal transcriptome; second, we must measure a transcriptome in a set of organisms that have a perturbed activity of the gene in question.

Having established the concept of a specific transcriptome as a group of ordered tuples of the form (gene, measurement), we reasoned that transcriptomes should obey the laws of genetic networks. Namely, we reasoned that if two hypothetical genes, A and B, act in a linear unbranched pathway, then A and B should have equivalent specific transcriptomes. We also reasoned that if A acts upon two different pathways, one which involves B and one which doesn't, the specific transcriptome of A should itself be separable into two independent parts: the specific transcriptome associated with the pathway $A \to B$, and the specific transcriptome associated with A but not B.

In order to solve genetic pathways, it is not sufficient to measure two transcriptomes and compare the overlap between the genes they contain. To solve complex genetic pathways, we must understand the valence of a genetic interaction between genes. In order to do this, we created linear operators to model perturbations that increase or decrease the activity of a gene relative to the basal measured state. These linear operators allow us to propagate a perturbation in one gene through a network and predict the expected changes in specific transcriptomes associated with the other genes in a network subject to the genetic model that has been specified by the investigator.

Using this formalism, we are, in theory, able to infer genetic relationships between genes. However, this formalism also provides a rational framework with which to identify specific transcriptomes associated with a pathway, and it also provides a theoretical manner with which to order genes acting along a pathway if the pathway branches at multiple points. Finally, our methods can allow us to identify regulatory relationships that are acting on different levels (either at the transcriptional level or at the protein level). For example, assuming that the mode of interaction between *rhy-1* and *egl-9* is the same as the mode of interaction between *hif-1* and *rhy-1* (transcriptional control) leads to aberrant behavior that is not observed empirically. Only by assuming that the regulatory mode between *rhy-1* and *egl-9* is different than between *hif-1* and *rhy-1* does the model generate predictions that parallel reality. Our new genetic notation, and the concepts contained therein, represent an important advance towards harnessing the power of transcriptomes as quantitative phenotypes. For a full description of this notation, see the S.I..

Clustering visualizes epistatic relationships between genes. As a first step in our analysis, we analyzed our data using a general linear model with a genotype term (see 1) on logarithm-transformed counts. Genes that are significantly altered between wild-type and a given mutant have a genotype coefficient that is statistically significantly different from 0. We refer to these coefficients through the greek letter β. These coefficients are not identical to the average log-fold change per gene, although they are loosely related to this quantity. In general, larger β magnitudes correspond to larger perturbations. These coefficients can be used to study the RNA-Seq data in question.

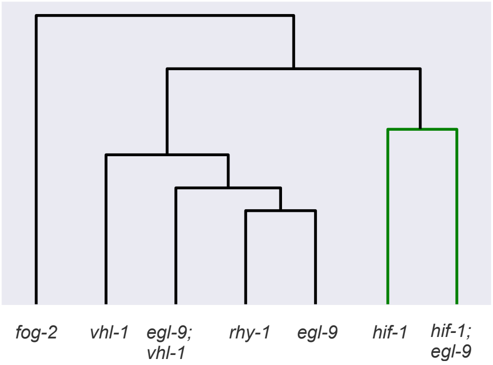

Fig. 1. Unsupervised aggregative clustering of various *C. elegans* mutants. Genes cluster in a manner that is biologically intuitive. Genes that inhibit *hif-1* (i.e, *egl-9*, *vhl-1*, and *rhy-1*) cluster far from *hif-1*. *hif-1* clusters with the suppressed *egl-9*; *hif-1* double mutant. A a mutant *fog-2* transcriptome, used as an outgroup, clusters farthest away.

Clustering is a well-known technique in bioinformatics that is used to identify relationships between high dimensional data points [17]. We wanted to make sure that clustering by differential expression yielded genetically relevant information. *hif-1* exhibits no obvious phenotypes under normoxic conditions, in contrast to *egl-9*, which exhibits an egg-laying (*egl*) phenotype in the same environment. *egl-9*; *hif-1* mutants suppress the *egl* phenotype. If transcriptomic phenotypes behave similarly to their macroscopic counterparts, *hif-1* should cluster with the *egl-9*; *hif-1* double mutant, whereas *egl-9* should cluster away from the *hif-1* mutant. Indeed, when blind, unsupervised clustering was performed on the data, three clusters emerged. *hif-1* and *egl-9*;*hif-1* clustered together, indicating suppression of the *egl-9* phenotype; whereas *egl-9*, *egl-9*;*vhl-1*, *vhl-1* and *rhy-1* all clustered separately. Finally, our negative control *fog-2* was in its own cluster (see Fig. 1). We conclude that expression data contains enough signal to cluster genes in a meaningful manner in complex metazoans.

Transcriptomic correlations can predict genetic regulation. Theoretically, two genes that share linear positive regulation should be positively correlated in their overlapping transcriptomes, whereas two genes that share linear negative regulation should be negatively correlated in their transcriptomes. Conversely, it follows that if two mutants have overlapping transcriptomes that are strongly positively correlated, it is likely that these two genes share a positive regulatory association. In other words, transcriptomic correlation is a good predictor of genetic regulation. For a formal introduction to the genetic logic, see S.I..

Although transcriptomic correlations could theoretically be used for the purposes of identifying genetic regulation, noise can cause serious interference with any inferences. Additionally, genes sometimes experience multiple modes of regulation, including positive and negative regulation, from the same gene or pathway. Because we are measuring the system at steady state, both modes of regulation will be measured simultaneously. If a positive and a negative signal of equal strength are both present in a transcriptome, running a naive regression may result in a value close to zero. Therefore, we took steps to mitigate noise emanating from frequent outliers and to identify multiple regulatory signals.

To mitigate noise, we rank-tranformed the β coefficients for each mutant. This has the effect of mitigating outliers by resetting the difference between adjacent coefficients to unity. Next, we performed robust Bayesian regressions using a Student-T distribution as a prior. A Student-T distribution decays less quickly than a normal distribution, which causes the model to consider outliers to be less informative than traditional regressions.

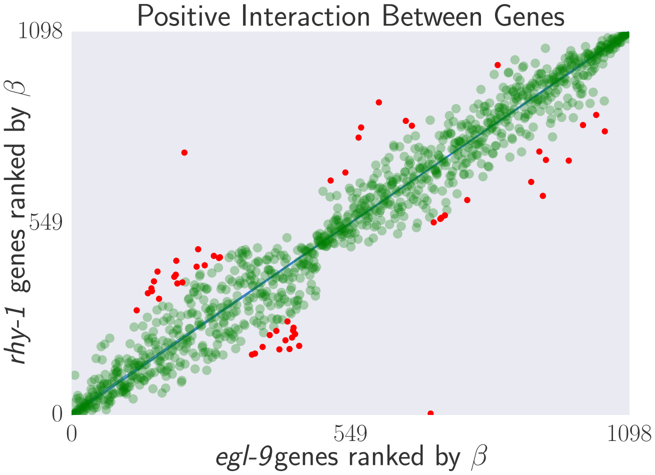

Fig. 2. Strong transcriptional correlations can be identified between genes that share a positive regulatory connection. We took the *egl-9* and the *rhy-1* transcriptomes, identified differentially expressed genes common to both transcriptomes and ranked each gene according to its differential expression coefficient β. We then plotted the rank of each gene in *rhy-1* versus the rank of the same gene in the *egl-9* transcriptome. The result is an almost perfect correlation. Green, transparent large points mark inliers to the regression (blue line); red, opaque, small points mark outliers to the regression. The two furthest outliers are annotated as pseudogenes in WormBase.

We saw that certain gene pairs correlated very well when genes were ranked by their expression changes (see Fig. 2). We generated all pairwise correlations between transcriptomes and we weighted the correlations by the number of genes that participated in the correlation (that were not outliers) divided by the total number of genes detected in all samples. We were able to identify a strong positive interaction between *egl-9* and *rhy-1*. The transcriptomes for these genes consisted of 1,487 and 1,816 significantly altered genes respectively and the overlap between both transcriptomes was extensive. On the other hand, none of the primary correlations between *hif-1* and its controlling genes were negative. We were unable to definitively determine the reason for behind this. The overlap between *hif-1* and all other genes was relatively small, and each overlap involved different sets of genes, which suggests that we did not sequence deeply enough to identify the nature of these positive interactions. See SI A for an exhaustive analysis of the expected and observed correlation between each gene pair in this circuit respectively. The regression slopes recapitulated a network with three 'modules': A control module, a responder module and an uncorrelated module (see Fig. 3). We were able to identify a strong positive interaction between *egl-9* and *rhy-1*. The magnitude of this weighted correlation is derived from the fact that the transcriptomes for these genes consisted of 1,487 and 1,816 significantly altered genes respectively and the overlap between both genes was quite extensive, which makes the weighting factor considerably larger than other pairs.

Previous work in the hypoxia pathway has found extensive

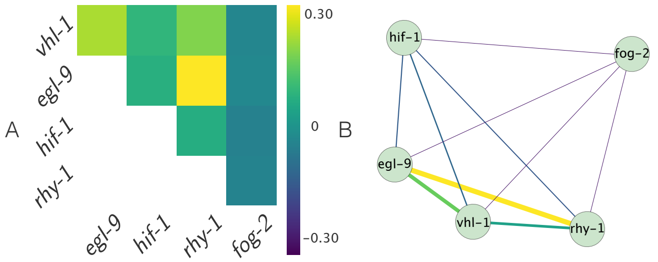

Fig. 3. A: Heatmap showing pairwise regression values between all single mutants. **B**: Correlation network drawn from the diagram. Edge width is proportional to the logarithm of the magnitude of the weighted correlation between two nodes divided by absolute value of the weighted correlation value of smallest magnitude. Edges are also colored according to the heatmap in **A**.

feedback loops in this pathway. Using the genetic formalism we developed, we realized that due to the fine-grained nature of interactomes we can use them to measure two regulatory interactions of opposite sign simultaneously in a single gene pair. This should lead to a characteristic X pattern in the ranked data. Such cross patterns can be indicative of feedback loops or of incoherent regulation at two distinct levels of gene expression (see Fig 4). For any gene that has a X-patterned diagram, we refer to the correlation that contains the largest number of participating isoforms as the primary correlation, and the correlation that contains the lesser number of points is referred to as a secondary correlation. In our dataset, all primary correlations were positive.

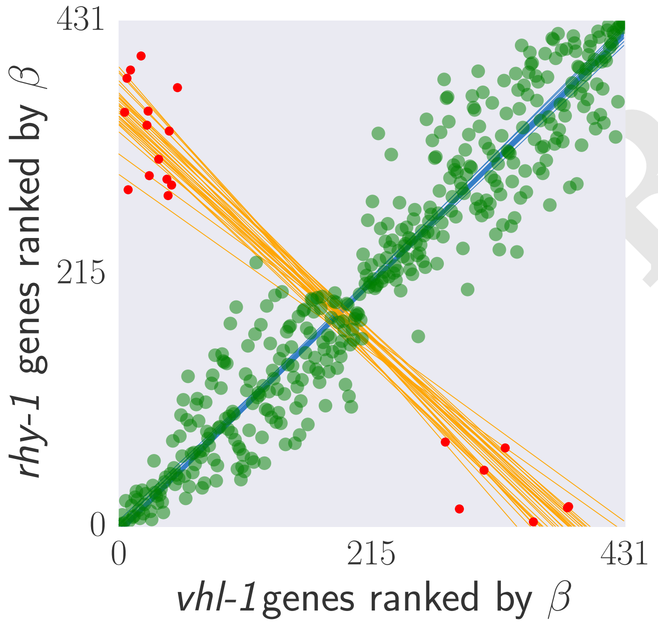

Fig. 4. Top: A feedback loop can generate transcriptomes that are both correlated and anti-correlated. **Bottom**: *hif-1* transcriptome correlated to the *rhy-1* transcriptome. Green large points are inliers to the first regression. Red small points are outliers to the first regression. Only the red small points were used for the secondary regression. Blue lines are representative samples of the primary bootstrapped regression lines. Orange lines are representative samples of the secondary bootstrapped regression lines.

We investigated whether any pairwise comparisons between our single mutants generated this cross pattern. Indeed, we found that comparing *hif-1* with *rhy-1*, and *hif-1* with *egl-9* yielded negative correlations, as did *rhy-1* and *vhl-1*. While the number of genes that lead these negative correlations is small (10–30 genes in any comparison) and is not significantly different from random expectation as assessed by a hypergeometric test, these outliers are expected for this circuit (see SI). On the other hand, unweighted secondary correlations had coefficients near unity for most comparisons, and they are predicted by theory. Statistical information should be integrated holistically with genetic models to assess whether outliers are meaningful or not. In our case, we believe that these outliers reflect meaningful interactions between genes in the hypoxia circuit, not random noise.

***in silico* qPCR reveals extensive feedback in the hypoxia pathway.** Our dataset enables us to perform a sort of *in silico* qPCR by selectively looking at expression of a few genes at a time. To verify the quality of our data, we queried the changes in expression of *nhr-57*. This reporter has been shown to have *hif-1* dependent expression [18–21]. In our dataset, this gene be upregulated in *egl-9*, *rhy-1* and *vhl-1*, but remains unchanged in *hif-1*. The *egl-9*;*vhl-1* had an expression level similar to *egl-9*; whereas the *egl-9*;*hif-1* mutant showed suppression of the reporter expression. All of these interactions reflect the literature.

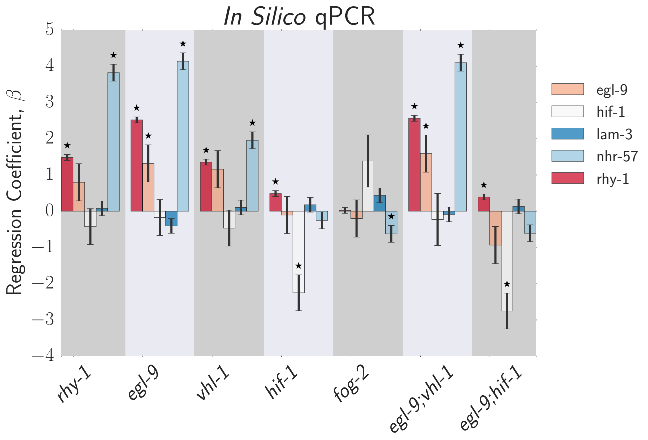

Fig. 5. Top: *In silico* qPCR results. *nhr-57* is an expression reporter that has been used previously to identify *hif-1* regulators [18, 22]. The *nhr-57* mRNA levels replicate what is observed in the literature. *lam-3* is shown here as a negative control that should not be altered by mutations in this pathway. The increases in the levels of *egl-9* and *rhy-1* when repressors of *hif-1* are knocked out are in agreement with previous literature [23]. We measured modest increases in the levels of *rhy-1* mRNA when *hif-1* is knocked out. The mechanism behind this is unclear. Negative and positive feedback loops from *hif-1* into its inhibiting genes could be a homeostatic mechanism.

We also performed *in silico* qPCR of every gene under scrutiny to get a clearer idea of the relationships between them (see Fig. 5). We observed changes in *rhy-1* expression consistent with previous literature [] when *hif-1* is activated. We also observed changes in *egl-9* expression when *egl-9* was mutated, and previous literature has identified *egl-9* as a hypoxia responsive gene []. Although *egl-9* was not increased in *rhy-1* and *vhl-1* mutants, the mRNA levels of *egl-9* trended towards increased expression. As with *nhr-57*, the *egl-9* and *rhy-1* expression phenotypes were abrogated in the *egl-9*;*hif-1* mutant; whereas the *egl-9*;*vhl-1* mutant showed expression phenotypes identical to the *egl-9* mutant, in support of *egl-9* and *vhl-1* acting in an AND-gated manner. Our dataset also

Table 1. Response Modeling of Double Mutants to Single Mutants

Double Mutant	Single Mutant	Δ	SE	p-value
1. *egl-9*;*vhl-1*	*egl-9*	0.00	0.01	0.81
2. *egl-9*;*vhl-1*	*vhl-1*	0.28	0.033	10^{-15}
3. *egl-9*;*hif-1*	*egl-9*	-0.85	0.074	10^{-13}
4. *egl-9*;*hif-1*	*hif-1*	-0.18	0.10	0.10

Table showing changes between single and double mutants. Δ is the result of a weighted-linear regression (WLS) between $\beta_{\text{Single Mutant}}$ and $\Delta = \beta_{\text{Double Mutant}} - \beta_{\text{Single Mutant}}$. $\Delta > 0$ represents a more severe phenotype than the single mutant. $\Delta < 0$ represents a suppressed phenotype relative to the single mutant. $\Delta = 0$ is expected for linear pathways or genes that are acting in linear or AND-gated fashion. $\Delta > 0$ is expected for genes that are acting additively on a pathway. WLS were performed only on genes that were significantly altered in both single mutants and the double mutant. $1 + \Delta$ is a very close approximation to the line of best fit between single mutant and double mutant.

shows that knockout of *hif-1* resulted in a modest increase in the levels of *rhy-1*. This suggests that *hif-1* is also a negative regulator of *rhy-1*.

Epistasis effects can be detected and quantified.. To quantify any epistasis between *egl-9* and *vhl-1* in our dataset, we identified the genes that were shared between each single mutant and the double mutant *egl-9*;*vhl-1*. If two genes act, for example, in a linear manner, then the double mutant should exhibit an identical phenotype to each single mutant. To test such a relationship, we can plot the difference in a gene i between the log change in expression between the two mutants, $\Delta_i = \beta_{\text{Double Mutant},i} - \beta_{\text{Single Mutant},i}$, against the log change in the single mutant, $\beta_{\text{Single Mutant},i}$. We can then fit a weighted linear regression to measure the slope of best fit. Genes that act in a linear pathway should yield lines with a slope of 0. Genes that have some additive flavor should have slopes greater than 0. Suppression, a hallmark of inhibition, should yield a slope less than 0.

We observe that the *egl-9*;*vhl-1* mutant has an identical phenotype to the *egl-9* single mutant (slope = 0; see Table. ??). On the other hand, *vhl-1* has a positive slope, indicating that *egl-9* is additive to *vhl-1*. Such partial additivity can be explained if *egl-9* is inhibiting *hif-1* in a *vhl-1*-dependent as well as a *vhl-1*-independent manner [22].

On the other hand, comparison of the *egl-9*;*hif-1* double mutant showed suppression of the *egl-9* transcriptomic phenotype. This suppression is expressed in various ways. First, the double mutant shows less statistically significantly differentially expressed genes than either single mutant. Secondly, the genes that are common to the *egl-9* and *egl-9*;*hif-1* transcriptomes show decreased expression in the *egl-9*;*hif-1* mutant than they do in *egl-9* on average (see Fig. 6). Likewise, the genes that are common to *hif-1* and *egl-9*;*hif-1* show no change in expression on average between these two mutants.

Because of the feedback between *hif-1* and *egl-9*, we expected a small subset of genes to be up-regulated or down-regulated consistently in every hypoxia mutant. Therefore, we searched for genes that were differentially expressed in all our hypoxia mutants (except the *hif-1*;*egl-9* mutant because it has a suppressed phenotype), reasoning that these genes should constitute an extremely high-quality picture of the hypoxia response, and should filter out other pathways (see SI). We identified 53 genes that satisfied these conditions, of which 10 genes were up-regulated in every mutant, and 13 genes were down-regulated (see SI for gene identities). These genes constitute a core response around the circuit in question, and their behaviour should reflect the genetic relationships in our system the best. Although we performed the regressions using only the overlapped genes between the single and double mutants, when we plotted only these high-quality genes, we can see that they show beautiful agreement with the global regressions (see www.wormlabcaltech.github.io/mprsq for all interactive graphics).

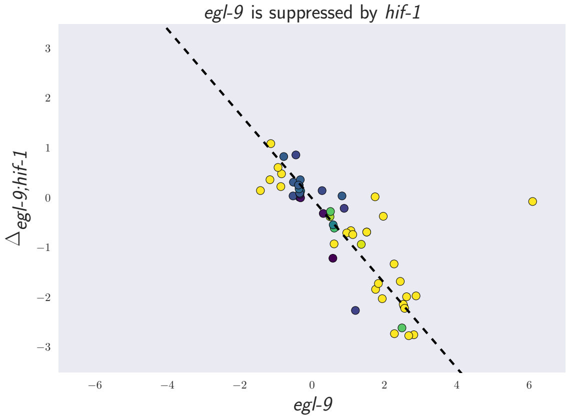

Fig. 6. The mutant *egl-9* transcriptomic phenotype is suppressed by mutations in *hif-1*. The graph shows the β coefficients for *egl-9* in the x-axis, and the change in β coefficient between the *egl-9*;*hif-1* and *egl-9* mutant. The dotted line is the regression line between the complete *egl-9* and *egl-9*;*hif-1* shared transcriptome. For clarity, only genes that were differentially expressed in the *egl-9*, *rhy-1*, *vhl-1*, *hif-1* and *egl-9*;*vhl-1* datasets are shown. These points constitute a very high-quality subset of the measured hypoxia response, as each isoform was identified as differentially expressed in 5 independent genotypes. The single outlier near (6, 0) is *nog-1*, and it is probably downstream of *egl-9*, and is not likely a *hif-1* target.

Transcriptomic decorrelation can be used to infer functional distance. We were interested in figuring out whether RNA-Seq could be used to identify functional interactions within a genetic pathway. Although there is no *a priori* reason why global gene expression should reflect functional interactions, the strength of the unweighted correlations between genes in the hypoxia pathway made us wonder how much information can be extracted from this dataset. Single genes are often regulated by multiple independent sources. The connection between two nodes can in theory be characterized by the strength of the edges connecting them (the thickness of the edge); the fraction of sources that regulate both nodes (the fraction of common inputs); and the fraction of genes that are regulated by both nodes (the fraction of common outputs). In other words we expected that expression profiles associated with a pathway would respond quantitatively to quantitative changes in activity of the pathway. Targeting a pathway at multiple points would lead to expression profile divergence as we compare nodes that are separated by more degrees of freedom, reflecting the flux in information between them.

We investigated the possibility that transcriptomic signals do in fact contain relevant information about the degrees of separation by weighting the robust bayesian regression of each pair of genes by $N_{\text{Intersection}}/N_{\text{Union}}$. We plotted the

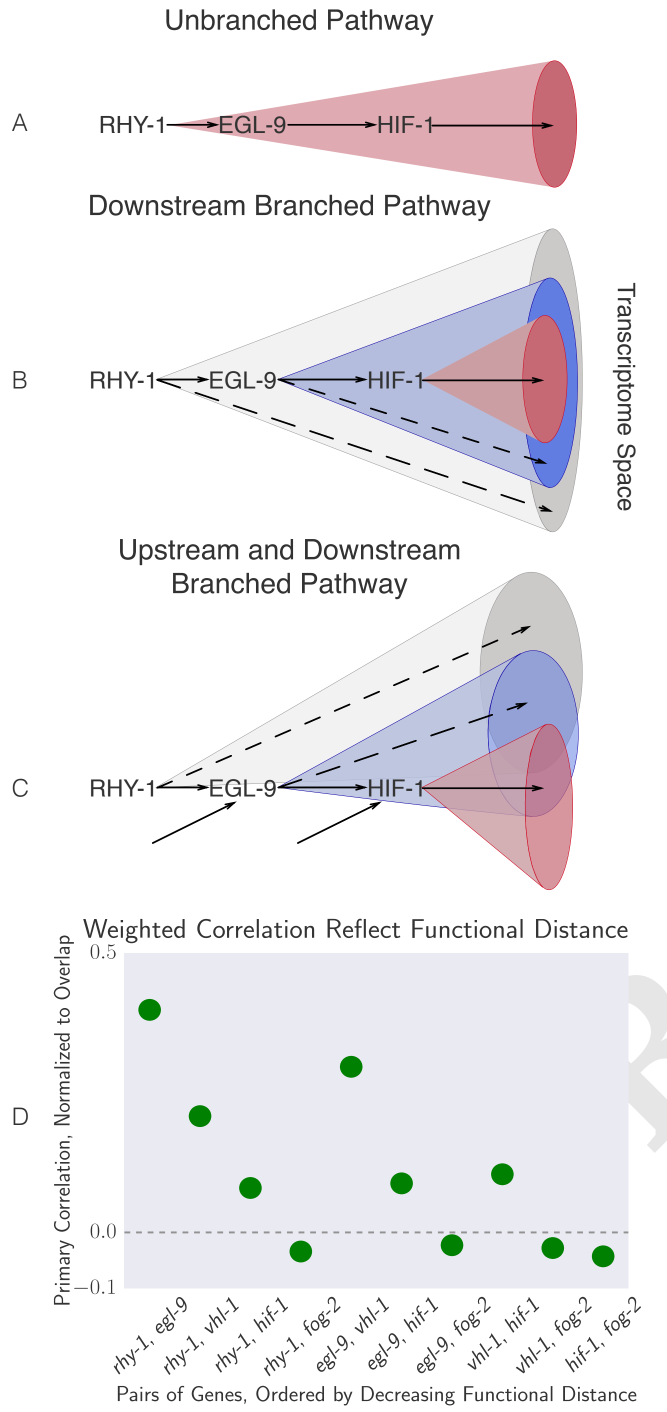

Fig. 7. Theoretically, transcriptomes can be used to order genes in a pathway under certain assumptions. Arrows in the diagrams above are intended to show the direction of flow, and do not indicate valence. **A** A linear pathway in which *rhy-1* is the only gene controlling *egl-9*, which in turn controls *hif-1* does not contain transcriptomes with enough information to infer the order between genes. **B** On the other hand, if *rhy-1* and *egl-9* have transcriptomic effects that are separable from *hif-1*, then the *rhy-1* transcriptome should contain contributions from *egl-9*, *hif-1* and *egl-9*- and *hif-1*-independent pathways. This pathway contains enough information to infer order. **C** If a pathway is branched in both upstream and downstream directions, observed transcriptomes will show even faster decorrelation. Nodes that are separated by many edges may begin to behave almost independently of each other with marginal transcriptomic overlap or correlation, reflecting the weak control distant nodes exert on each other. **D** The hypoxia pathway can be ordered according to functional distance. The rapid decay in correlation is probably due to a mixture of upstream and downstream branching that happens along this pathway.

weighted correlation of each gene pair, ordered by increasing functional distance (see Fig. 7). In every case, we see that the weighted correlation decreases monotonically due mainly, but not exclusively, to decreasing N_{Overlap}. We believe that this result is not due to random noise or insufficiently deep sequencing. Instead, we propose a framework in which every gene is regulated by multiple different molecular species, which induces progressive decorrelation. This decorrelation in turn has two consequences. First, decorrelation within a pathway implies that two nodes may be almost independent of each other if the functional distance between them is large. Second, it may be possible to use decorrelation dynamics to infer gene order in a pathway, as we have done with the hypoxia pathway[1].

Identification of novel targets and biological processes in the hypoxia response. So far, our analysis has focused mainly on extracting genetic relationships between the set of mutants we sequenced. Next, we attempted to leverage the information gained from modeling the genetic system using our new genetic notation to extract specific interactomes for each gene. First, we tried to identify genes that are candidates for *hif-1* binding by identifying genes which expression goes down in response to loss of *hif-1* (or viceversa) and are not downstream of *egl-9* or *rhy-1*. We identified 133 genes that are activated by HIF-1. We verified that the genes we identified are actually downstream of *hif-1* by searching for a set of 20 gold-standard genes from the literature [18, 19] in our gene set. We found that *hif-1* targets were significantly enriched in these genes ($p < 10^{-7}$). GO term enrichment using DAVID indicated that this list was significantly enriched in metabolic pathways (Enrichment Score 5.17, all p-values in cluster < 0.01). Although no other clusters were statistically significantly enriched, the next possibly enriched clusters included terms involving oxidoreduction, lectins, mitochondria and proteolysis. Although not statistically significantly enriched, we consider these terms informative, since they point at known actors within the hypoxia pathway.

We were also able to identify a 36 genes associated specifically with *vhl-1*. These 36 genes include *pole-1*, an ortholog of human polymerase ϵ catalytic subunit; *F33H2.6*, an ortholog of the human regulator of microtubule dynamics 1 (RMDN1)[]; and many solute carriers. *vhl-1* has been previously implicated as a controller of mitotic fidelity in renal cell carcinoma [24]. We analyzed this gene list using Tissue Enrichment Analysis (TEA), which showed enrichment of multiple neuronal terms ('ADE', 'postdeirid sensillum', 'PDE', 'posterior lateral ganglion') and in 'head muscle' ($q < 0.05$ for all). Neuronal enrichment could be consistent with reports that *vhl-1* promotes neuronal apoptosis in a HIF-independent manner []. Our findings support a role of *vhl-1* in chromosomal integrity and mitotic fidelity.

Finally, we tried to identify specific transcriptomes associated with *egl-9* and *rhy-1*. In order to identify specific transcriptomes for these two genes, we parted from the assumption that both of them might have non-empty specific transcriptomes. Using our newly developed logic to test this

[1] An important question is whether a looped circuit like the hypoxia pathway can be ordered in the way we have ordered it in Fig. 7 since a loop does not technically have a beginning. One explanation is that we studied the hypoxia pathway under normoxic conditions, and therefore the control of *hif-1* over *rhy-1* and *egl-9* is weak, effectively turning the looped pathway into a linear one. Probably, under hypoxic conditions the pathway would effectively be reversed.

suggests that comparing the *egl-9* and *rhy-1* should show a cross pattern as a result of the feedback loops into *rhy-1* and *egl-9*, and this cross-pattern should reflect the contribution of the *rhy-1* transcriptome (*egl-9* causes *rhy-1* mRNA levels to go up, which we assume leads to increased activity of the gene, whereas mutating *rhy-1* should decrease activity of the gene). However, no such cross pattern is observed (with the exception of two genes that are clear outliers, both of which are annotated as pseudogenes, see Fig. 2). Therefore, one of the following must be true: *rhy-1* does not have a specific transcriptome, increased expression of *rhy-1* does not lead to increased activity of *rhy-1* or we failed to measure the *rhy-1* transcriptome, possibly because the signal for it is much weaker than for other genes.

Once we established that we could not measure a *rhy-1* specific transcriptome, we identified the *egl-9* transcriptome as consisting of 432 genes. We subjected this list of genes to GO analysis using Panther, which showed enrichment of terms involving 'chromatin assembly' ($p < 0.01$), 'cellular amino acid catabolic process' ($p < 0.01$), 'ectoderm development' ($p < 10^{-5}$), 'translation' ($p < 10^{-12}$) and 'cell-cell adhesion' ($p < 10^{-4}$). Tissue Enrichment Analysis (TEA) showed that these genes reflect enrichment of 'anal depressor muscle' ($q < 10^{-2}$) and 'intestinal muscle' ($q < 0.1$). In spite of known involvement of *egl-9* in neuronal tissues, no neuronal enrichment was observed. On the other hand, when we performed hypergeometric enrichment test on WormBase's phenotype ontology, we found that the *egl-9* specific transcriptome is enriched for terms related to behavior, such as 'male mating defective', 'organism region behavior variant' and 'positive chemotaxis defective' ($q < 10^{-19}$ for all). Not surprisingly, *egl-9* has been implicated in behavioral modification previously [25, 26], although a male-mating phenotype has not previously been reported.

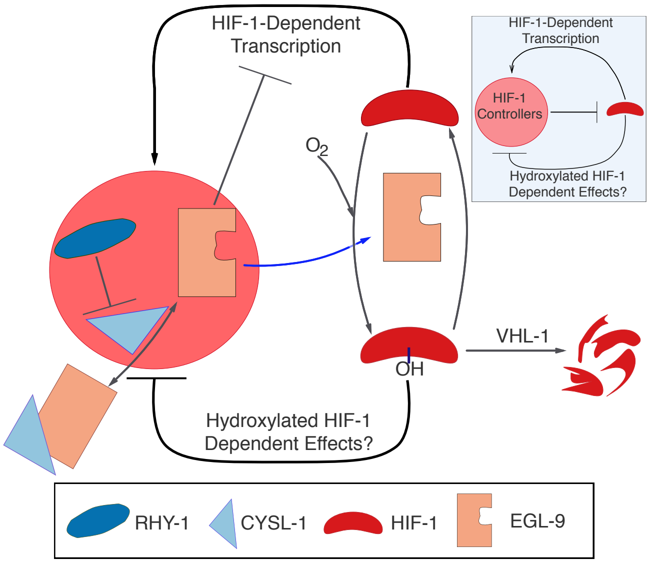
HIF-1-Dependent Transcription
O$_2$
OH
VHL-1
Hydroxylated HIF-1 Dependent Effects?
HIF-1-Dependent Transcription
HIF-1 Controllers
Hydroxylated HIF-1 Dependent Effects?
RHY-1
CYSL-1
HIF-1
EGL-9

Fig. 8. A schematic of the hypoxia pathway in *C. elegans*. RHY-1 likely inhibits a protein-protein interaction between EGL-9 and CYSL-1 [16]. This interaction inhibits the activity of EGL-9. When active, EGL-9 can inhibit HIF-1 by catalyzing the hydroxylation of HIF-1, which leads to its recognition and ubiquitination by VHL-1 and ultimately leads to rapid degradation of HIF-1 protein. Independently of its enzymatic function, EGL-9 can also inhibit HIF-1 transcriptional activity. Our results identified *rhy-1*, *egl-9* and *cysl-1* as downstream targets of the *hif-1*-dependent transcriptional response. However, we also identified increased in the mRNA levels of *rhy-1* and *cysl-1* when *hif-1* was knocked out. Moreover, the *hif-1* knock-out transcriptome correlated positively with *rhy-1*, *egl-9* and *vhl-1*. One plausible explanation for these observations is that hydroxylated HIF-1 has transcriptional consequences. Inset shows a simplified diagram of the hypoxia pathway.

Discussion

Reconstructing Hypoxia Circuit in *C. elegans* using Transcriptomic Profiling. Previous work has established a circuit in which *rhy-1* leads to the activation of *egl-9*, and *egl-9* inhibits *hif-1* in an oxygen-dependent manner. Hydroxylated HIF-1 can then be degraded in a *vhl-1*-dependent manner. There is also evidence that *egl-9* and *rhy-1* are in turn activated by *hif-1* [23, 27]. Finally, there is evidence that although the interaction between *egl-9* and *vhl-1* is important for *hif-1* repression, *egl-9* can also act in a non-*vhl-1* and non-oxygen dependent manner (see Fig. 8 top).

Using only information gathered from transcriptomic profiling, we were able to reconstruct the known regulatory relationships between *egl-9*, *rhy-1* and *vhl-1* using a combination of inter-transcriptome correlations and epistasis measurements. Using clustering as a proxy for phenotype, we were able to infer the relationship between *egl-9* and *hif-1*. Alternatively, we could have used our epistasis measurements to conclude that *egl-9* inhibits *hif-1*. By looking at single gene measurements using *in silico* qPCR, we were able to determine that *rhy-1* transcription is stimulated by HIF-1. Single gene measurements also suggested that *egl-9* is transcriptionally downstream of *hif-1*, although only the measurements from *egl-9* mutants were statistically significantly different from 0 and all other *hif-1*-constitutive mutants only trended towards increased *egl-9* expression. In addition, our dataset also identified *cysl-1* as a *hif-1*-responsive gene. We were unable to detect *vhl-1* transcripts in our RNA-seq measurements, and therefore we cannot rule out effects of *hif-1* on *vhl-1* using this data. All of these measurements are consistent with the previous literature.

Although we reconstructed the reported pathway, some of our data cannot be easily explained under a model in which HIF-1 is transcriptionally active but hydroxylated HIF-1 is not. For example, the *hif-1* transcriptome correlated positively with the *egl-9*, *rhy-1* and *vhl-1* transcriptomes. Moreover, our *in silico* qPCR suggests that *rhy-1* expression levels are increased upon mutation of the *hif-1* gene, which is hard to explain under the standard model, since *hif-1* is an activator of *rhy-1*, most HIF-1 is expected to be hydroxylated under normoxic conditions, and total HIF-1 are not anywhere near their hypoxic levels. The simplest explanation that could explain these results without involving another gene would be to postulate that hydroxylated HIF-1 is an active player with downstream transcriptional consequences, and that one of these consequences is repression of *rhy-1*.

Such a model is appealing because the network that emerges has certain characteristics of homeostatic pathways. First, under oxygen deplete conditions, HIF-1 will no longer be hydroxylated and it will initiate transcription of the hypoxic response. Upregulation of its controller genes generates a negative feedback loop which has positive consequences for the response time of a HIF-1 growth curve [28]. On the other hand, we can envision an environment in which HIF-1 is completely hydroxylated. Presumably, an organism always needs non-hydroxylated HIF-1, or alternatively, needs hydroxylated HIF-1 levels to be below a certain threshold. If hydroxylated HIF-1 can mediate a transcriptional response, then the organism benefits from complete information about the iron and oxygen balance within it, as both species of HIF-1 signal with strengths proportional to their total levels in the cell.

A model in which the hydroxylation of HIF-1 alters, but does not abrogate, the transcriptional effects of this gene would explain the positive correlations between *hif-1*, *egl-9*, and *rhy-1*. However, we must emphasize that the poor agreement between overlaps (only 5 genes appear in both intersections, and each intersection consists of less than 200 genes) prevents us from making a more complete statement. Moreover, this newly proposed model does not explain why *vhl-1* has a positive correlation with *hif-1*. It may be the case then that another gene is involved in this pathway that positively links *vhl-1* and *hif-1*. An interaction of this sort would explain why the correlation between *vhl-1* and *hif-1* is mediated by genes that are completely different from *egl-9* and *rhy-1*.

Towards A Genetic Theory of Transcriptomics. We have shown that transcriptomes contain sufficient information to be used as semi-quantitative phenotypes in complex metazoans. These phenotypes can be interpreted globally via correlation tests, clustering or other probabilistic methods; alternatively, they can be used to query single reporter genes in a manner similar to qPCR today. As a result of this dynamic range, transcriptomic phenotypes have distinct advantages over physical traits. Firstly, due to their increased complexity, the genotype-phenotype mapping degeneracy ought to be greatly reduced, which facilitates predictions of genetic interaction. Secondly, genes that result in subtle or no visible traits when mutated may have detectable, reproducible phenotypes at the transcriptomic level, which would facilitate the study of small-effect genes and other quantitative traits.

We have formalized the concept of genetic transcriptome by developing a controlled language and formal notation to study these objects. In our language, perturbations are unary operators that act on a single gene, which has consequences for the function it controls (its transcriptome). This notation makes it possible to think about transcriptome genetics in a manner analogous to classical genetics using scalar phenotypes, and also makes identification of specific transcriptomes easier and more rigorous. For example, we concluded that *rhy-1* does not have a specific transcriptome. All of its transcriptomic consequences appear to emanate from the downstream gene *egl-9*. We arrived at this conclusion through theoretical considerations that suggested that a *rhy-1* specific transcriptome should manifest in a negative correlation when the *egl-9* and *rhy-1* transcriptomes are plotted on a rank-plot. Using simple binary exclusion principles one might arrive at the conclusion that the *rhy-1* specific transcriptome is defined as the genes that are not present in the intersection between *egl-9* and *rhy-1*, which would result in almost 400 genes being assigned to *rhy-1*. Although such substractive logic is attractive at first sight, it is fundamentally untestable due to the large number of genes involved. On the other hand, using our notation and logic, we can directly query the data for signs of a specific transcriptome. If those signs are not missing, we have some degree of confidence that the specific transcriptome was not measured, could not be identified, was not perturbed or does not exist. Further tests can be envisioned to test each possibility.

Of particular interest to us was the idea that the logic necessary to understand transcriptome genetics is fundamentally simple. It builds in a logical manner from genetics principles derived from macroscopic observations, and makes predictions that are easily testable through linear regressions. This contrasts starkly with preconceived notions of complicated bioinformatics that are believed to be necessary to extract information from next-generation sequencing data. Even more striking is the fact that transcriptomes seem to exhibit quantitative epistatic behaviour. Finally, we note that the linear structure of the logic we have developed hints at the possibility that an algebra exists that encompasses these rules. Finding such an algebra would constitute an important development which could have important consequences for the kind of algorithms that are developed to deconvolute, dissect and understand transcriptomes in the future.

Materials and Methods

Nematode strains and culture. Strains used were N2 wild-type Bristol, CB5602 *vhl-1*(ok161), CB6088 *egl-9*(sa307) *hif-1*(ia4), CB6116 *egl-9*(sa307) *vhl-1*(ok161), JT307 *egl-9*(sa307), ZG31 *hif-1*(ia4), RB1297 *rhy-1*(ok1402). ZG31*hif-1*(ia4) is a null mutant of *hif-1* which deletes 1231 bp of the second, third and fourth exons. JT307 contains the null mutant *egl-9*(sa307) which is a 243 bp deletion. RB1297 contains null mutation *rhy-1*(ok1402) with an estimated 700 bp deletion constructed by the OMRF Knockout Group. CB5602 contains the deletion mutation of *vhl-1*(ok161). CB6088 contains *egl-9*(sa307);*hif-1*(ia4). CB6116 contains *egl-9*(sa307) *vhl-1*(ok161). All strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). All lines were grown on standard nematode growth media (NGM) plates with seeded with OP50 *E. coli* at 20°C (Brenner 1974).

RNA Isolation. Unsynchronized lines were grown on NGM plates at 20C and eggs harvested by sodium hypochlorite treatment. Eggs were plated on 6 to 9 small 5cm NGM plates with ample OP50 *E. coli* at a density chosen to avoid starvation and grown at 20°C. Worms were staged and harvested based on the time after plating, vulva morphology and the absence of eggs. Approximately 30–50 non-gravid young adults (YA) were picked and placed in 100μL of TE pH 8.0 at 4°C in 0.2mL PCR tubes. After settling and a brief spin in microfuge approximately 80μL of TE was removed from the top of the sample and individual replicates were snap frozen in liquid N2. These replicate samples were then digested with Proteinase K for 15min at 60° in the presence of 1% SDS and 1.25μL RNA Secure (Ambion AM 7005). RNA samples were then taken up in 5 Volumes of Trizol (Tri Reagent Zymo Research) and processed and treated with DNAase I using Zymo MicroPrep RNA Kit (Zymo Research Quick-RNA MicroPrep R1050). RNA was eluted in dH2O and divided into aliquots and stored at -80°C. One aliquot of each replicate was analyzed by both NanoDrop for impurities, Qubit for concentration and then analyzed on an Agilent 2100 BioAnalyzer. Replicates were selected that had RNA integrity numbers (RIN) equal or greater than 9.0 and showed no evidence of bacterial ribosomal bands, except for the ZG31 mutant where one of three replicates had a RIN of 8.3.

Library Preparation and Sequencing. Forthcoming.

Read Alignment and Differential Expression Analysis. We used Kallisto to perform read pseudo-alignment and performed differential analysis using Sleuth. We fit a generalized linear model for a transcript t in sample i:

$$y_{t,i} = \beta_{t,0} + \beta_{t,genotype} \cdot X_{t,i} + \beta_{t,batch} \cdot Y_{t,i} + \epsilon_{t,i} \qquad [1]$$

where $y_{t,i}$ are the logarithm transformed counts; $\beta_{t,genotype}$ and $\beta_{t,batch}$ are parameters of the model, and which can be interpreted as biased estimators of the log-fold change; $X_{t,i}, Y_{t,i}$ are indicator variables describing the conditions of the sample; and $\epsilon_{t,i}$ is the noise associated with a particular measurement.

ACKNOWLEDGMENTS. This article was written with support of the Howard Hughes Medical Institute. We would like to thank Jonathan Liu, Han Wang, and Porfirio Quintero for helpful discussion.

1. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9(11):855–867.
2. Hughes TR et al. (2000) Functional Discovery via a Compendium of Expression Profiles. *Cell* 102(1):109–126.
3. Van Driessche N et al. (2005) Epistasis analysis with global transcriptional phenotypes. *Nature genetics* 37 VN - r(5):471–477.
4. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621–628.
5. Metzker ML (2010) Sequencing technologies - the next generation. *Nature reviews. Genetics* 11(1):31–46.
6. Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature biotechnology* 32(5):462–464.
7. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2016) Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *bioRxiv* p. 021592.
8. Pimentel HJ, Bray NL, Puente S, Melsted P, Pachter L (2016) Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* p. 058164.
9. Trapnell C et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* 31(1):46–53.
10. Singer M et al. (2016) A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. *Cell* 166(6):1500–1511.e9.
11. Shalek AK et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498(7453):236–40.
12. Schwarz EM, Kato M, Sternberg PW (2012) Functional transcriptomics of a migrating cell in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* 109(40):16246–51.
13. Van Wolfswinkel JC, Wagner DE, Reddien PW (2014) Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell* 15(3):326–339.
14. Scimone ML, Kravarik KM, Lapan SW, Reddien PW (2014) Neoblast specialization in regeneration of the planarian schmidtea mediterranea. *Stem Cell Reports* 3(2):339–352.
15. Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. *Cell* 148(3):399–408.
16. Ma DK, Vozdek R, Bhatla N, Horvitz HR (2012) CYSL-1 Interacts with the O 2-Sensing Hydroxylase EGL-9 to Promote H 2S-Modulated Hypoxia-Induced Behavioral Plasticity in C. elegans. *Neuron* 73(5):925–940.
17. Yeung KY, Medvedovic M, Bumgarner RE (2003) Clustering gene-expression data with repeated measurements. *Genome biology* 4(5):R34.
18. Shen C, Shao Z, Powell-Coffman JA (2006) The Caenorhabditis elegans rhy-1 Gene Inhibits HIF-1 Hypoxia-Inducible Factor Activity in a Negative Feedback Loop That Does Not Include vhl-1. *Genetics* 174(3):1205–1214.
19. Shen C, Nettleton D, Jiang M, Kim SK, Powell-Coffman JA (2005) Roles of the HIF-1 hypoxia-inducible factor during hypoxia response in Caenorhabditis elegans. *Journal of Biological Chemistry* 280(21):20580–20588.
20. Ackerman D, Gems D (2012) Insulin/IGF-1 and hypoxia signaling act in concert to regulate iron homeostasis in Caenorhabditis elegans. *PLoS Genetics* 8(3).
21. Park EC et al. (2012) Hypoxia regulates glutamate receptor trafficking through an HIF-independent mechanism. *The EMBO Journal* 31(6):1618–1619.
22. Shao Z, Zhang Y, Powell-Coffman JA (2009) Two Distinct Roles for EGL-9 in the Regulation of HIF-1-mediated gene expression in Caenorhabditis elegans. *Genetics* 183(3):821–829.
23. Powell-Coffman JA (2010) Hypoxia signaling and resistance in C. elegans. *Trends in Endocrinology and Metabolism* 21(7):435–440.
24. Hell MP, Duda M, Weber TC, Moch H, Krek W (2014) Tumor Suppressor VHL Functions in the Control of Mitotic Fidelity. *Cancer Research* 74(9):2422–2431.
25. Chang AJ, Bargmann CI (2008) Hypoxia and the HIF-1 transcriptional pathway reorganize a neuronal circuit for oxygen-dependent behavior in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* 105(20):7321–7326.
26. Gray JM et al. (2004) Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. *Nature* 430(6997):317–322.
27. Bishop T et al. (2004) Genetic Analysis of Pathways Regulated by the von Hippel-Lindau Tumor Suppressor in Caenorhabditis elegans. *PLoS Biology* 2(10).

Genetic Analysis. Genetic analysis of the processed data was performed in Python 3.5. Our scripts made extensive use of the Pandas, Matplotlib, Scipy, Seaborn, Sklearn, Networkx, Bokeh, PyMC3, and TEA libraries [29–37]. Our analysis is available in a Jupyter Notebook [38]. All code and required data (except the raw reads) are available at https://github.com/WormLabCaltech/mprsq along with version-control information. Our Jupyter Notebook and interactive graphs for this project can be found at https://wormlabcaltech.github.io/mprsq/. Raw reads were deposited at XXXXXXXXXXX

28. Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8(6):450–61.
29. Bokeh Development Team (2014) Bokeh: Python library for interactive visualization.
30. McKinney W (2011) pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* pp. 1–9.
31. Oliphant TE (2007) SciPy: Open source scientific tools for Python. *Computing in Science and Engineering* 9:10–20.
32. Pedregosa F et al. (2012) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
33. Salvatier J, Wiecki T, Fonnesbeck C (2015) Probabilistic Programming in Python using PyMC. *PeerJ Computer Science* 2(e55):1–24.
34. Van Der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering* 13(2):22–30.
35. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* 9(3):99–104.
36. Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW (2016) Tissue enrichment analysis for C. elegans genomics. *BMC Bioinformatics* 17(1):366.
37. Waskom M et al. (2016) seaborn: v0.7.0 (January 2016).
38. Pérez F, Granger B (2007) IPython: A System for Interactive Scientific Computing Python: An Open and General- Purpose Environment. *Computing in Science and Engineering* 9(3):21–29.