Genetic Analysis of a Metazoan Pathway using Transcriptomic Phenotypes

immediate

This manuscript was compiled on January 26, 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 120 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 of their gene products) or indirectly. 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]. Additionally, eQTL studies in *C. elegans* and *Drosophila melanogaster* have established the usefulness of transcriptomic phenotypes for population genetics studies [].

Need citations.

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 [6–8] and improved differential analysis of gene expression [9, 10]. As a result, RNA-Seq has been successfully used to identify genetic modules involved in a variety of processes, including T-cell regulation [11, 12], the *C. elegans* linker cell migration [13], or planarian stem cell maintenance [14, 15]. For the most part, the role of transcriptional profiling has been restricted to target gene identification. In cell culture, single-cell RNA-seq has seen significant progress towards using transcriptomes as phenotypes with which to test genetic interactions [16, 17]. More recently, we have shown the first identification of a developmental state of *C. elegans* using whole-organism transcriptome profiling [18].

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 [19]. These approaches identified the Hypoxia Inducible Factors (HIFs) as an important group of oxygen responsive genes.

Hypoxia Inducible Factors are highly conserved in metazoans [20]. A common mechanism for hypoxia-response induction is heterodimerization between a HIFα and a HIFβ subunit. The heterodimer then initiates transcription of target genes [21]. 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* [22] and a single HIFβ gene, *ahr-1* [23]. HIF target genes have been implicated in a wide variety of cellular and extracellular

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

processes such as glycolysis, extracellular matrix modification, autophagy and immunity [19, 24–27].

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 [28]. 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* [29]. HIF-1 hydroxylation increases its binding affinity to Von Hippel Lindau Tumor 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 [30].

Here, we show that transcriptomes contain strong, robust signals that can be used to infer relationships between genes in complex metazoans by reconstructing the hypoxia pathway in *C. elegans* using RNA-Seq. 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*. Using a single set of genome-wide measurements, we were able to observe and quantitatively assess significant fraction of the known transcriptional effects of *hif-1* in *C. elegans*. A complete, interactive version of the analysis is also available at www.wormlabcaltech.github.io/mprsq.

Results

The hypoxia pathway controls thousands of genes in *C. elegans*. We performed whole-organism RNA-seq of the hypoxia pathway at a moderate sequencing depth (7 million mapped reads for each individual replicate) under normoxic conditions, which allowed us to measure 13,598 isoforms across all replicates and genotypes, which constitutes over half of all isoforms in *C. elegans*. In spite of the low sequencing depth, transcriptome profiling of the hypoxia pathway revealed that this pathway controls thousands of genes in *C. elegans*. The *egl-9* transcriptome showed differential expression of 1,487 genes, similarly to the 1,816 genes differentially expressed in *rhy-1* mutants. The *vhl-1* transcriptome showed considerably fewer differentially expressed genes (605), possibly reflecting the known fact that it is a weaker controller of *hif-1* than *egl-9* [31]. The *egl-9*;*vhl-1* double mutant transcriptome showed 1,989 differentially expressed genes. The *hif-1* mutant also showed a transcriptomic phenotype involving 481 genes. The *egl-9*;*hif-1* double mutant showed a similar number of genes with altered expression (364).

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

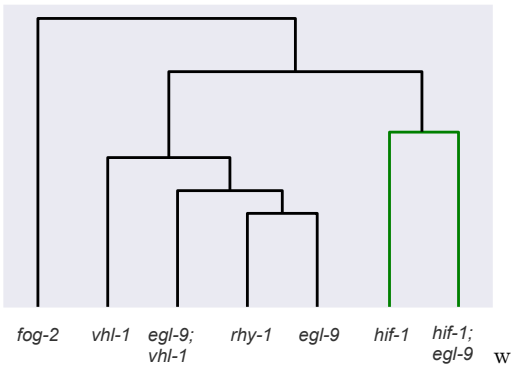

Fig. 1. Unsupervised aggregative clustering of various *C. elegans* mutants. Genes cluster in a manner that is biologically intuitive. Genes that inhibit *hif-1* (i.e, *egl-9*, *vhl-1*, and *rhy-1*) cluster far from *hif-1*. *hif-1* clusters with the suppressed *egl-9*; *hif-1* double mutant. A a mutant *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 [32]. 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.

Reconstruction of the hypoxia pathway from first genetic principles. Having shown that the signal in the mutants we selected was strong enough to cluster mutants using the regression coefficients, we set out to reconstruct the hypoxia pathway from first genetic principles. In general, to reconstruct a pathway, we must assess whether two genes act on the same phenotype (independence); then we must measure whether these genes act additively or epistatically on the measured phenotype; and if there is epistasis we must measure whether it is positive or negative, in order to assess whether the epistatic regulation is a genetic suppression or a synthetic interaction.

Genes in the hypoxia mutant act on the same transcriptional phenotype. We observed that all the hypoxia mutants had significant overlap between their differentially expressed transcriptomes relative to a wild-type control (fraction of shared transcriptomes ranged from a minimum of 65 genes shared between *hif-1* and *egl-9*;*hif-1* to a maximum of 1,249 shared genes between *egl-9* and *egl-9*;*vhl-1*). For comparison, we also analyzed a previously published *fog-2* transcriptome [18]. *fog-2* is involved in masculinization of the *C. elegans* germline, which

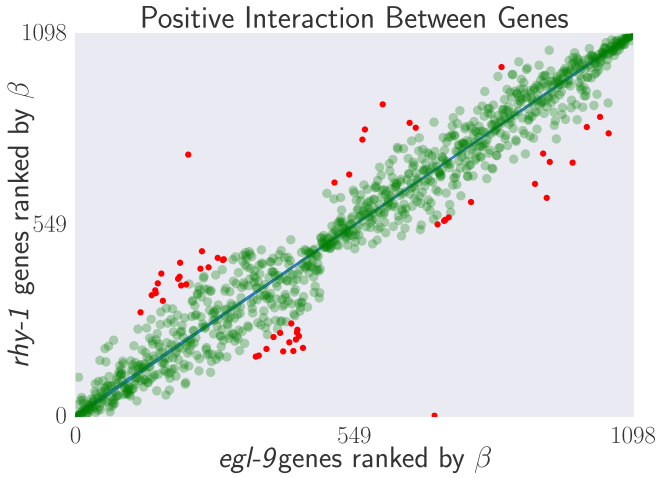

Fig. 2. Strong transcriptional correlations can be identified between genes that share a positive regulatory connection. We took the *egl-9* and the *rhy-1* transcriptomes, identified differentially expressed genes common to both transcriptomes and ranked each gene according to its differential expression 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.

enables sperm formation, and has not been described to be involved in the hypoxia pathway. The hypoxia pathway transcriptomes and the *fog-2* transcriptome showed similar overlap as the hypoxia pathway to itself (123–618 genes). Given the similar overlaps between known interactors and an unknown transcriptome, we conclude that the *fog-2* mutant we studied acts on the same phenotype as mutants from the hypoxia pathway.

Although overlapping transcriptomes may be enough to conclude that a set of mutants share a phenotype, we wanted to know whether we could draw out more information from looking at quantitative agreement between perturbations. To this end, we rank-transformed the regression coefficients β for each transcriptome, and calculated lines of best fit using Bayesian regression with a Student-T distribution to mitigate noise from outliers (see Fig 2). For transcriptomes associated with the hypoxia pathway, we found that these correlations tended to have values as high as 0.98 with a tight distribution around the line of best fit, whereas the correlations for mutants from the hypoxia pathway with the *fog-2* mutant were considerably weaker, with magnitudes between 0.6–0.85 and a considerably larger spread around the line of best fit. Although *hif-1* is known to be genetically repressed by *egl-9*, *rhy-1* and *vhl-1* [33], all the correlations between these genes and *hif-1* were negative. The overlap between *hif-1* and all other genes was 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. After we calculated the pairwise correlation between each transcriptome, we weighted the result of each regression by the number of differentially expressed isoforms shared by two transcriptomes and divided by the total number of differentially expressed isoforms present in the two transcriptomes, $N_{\text{overlap}}/N_{\text{g1}\cup\text{g2}}$. The weighted regressions 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 extensive, which makes the weighting factor considerably larger than other pairs. Likewise, the weak correlation between *hif-1* and *egl-9*, *vhl-1* and *rhy-1* is at least partially the result of its relatively weak transcriptomic phenotype relative to the other genes, particularly *egl-9* and *rhy-1*. The fine-grained nature of transcriptional phenotypes means that these weighted correlations between transcriptomes of single mutants are predictive of genetic interaction.

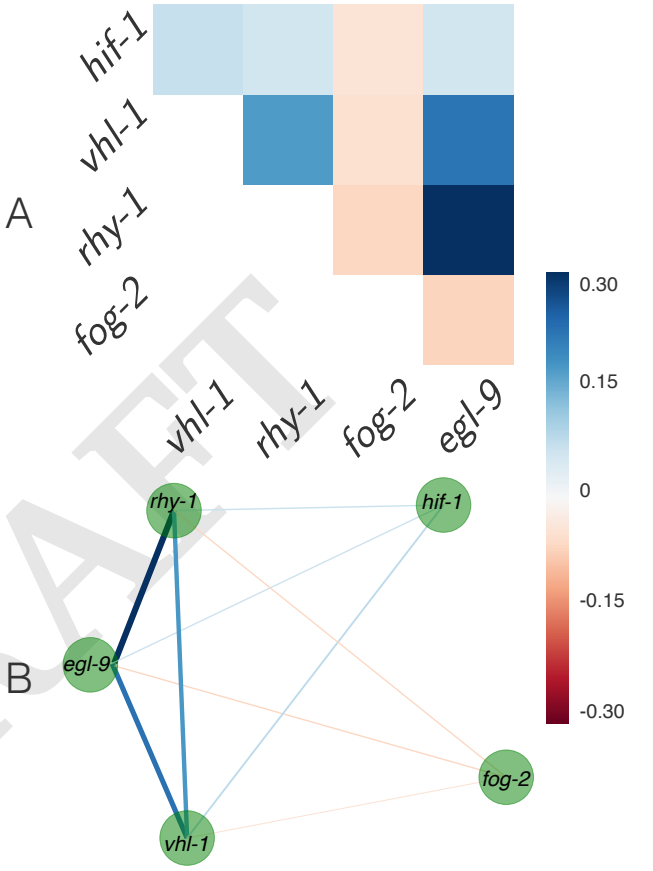

Fig. 3. A: Heatmap showing pairwise regression values between all single mutants. **B**: Correlation network drawn from the diagram. Edge width is proportional to the logarithm of the magnitude of the weighted correlation between two nodes divided by absolute value of the weighted correlation value of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

A quality check of the transcriptomic data reveals excellent agreement with the literature. One way to establish whether genes are acting additively or epistatically to each other is to perform qPCR of a reporter gene in the single and double mutants. This approach was used to successfully map the relationships within the hypoxia pathway (see, for example [31, 34]). A commonly used reporter is *nhr-57*, which is known to exhibit large changes in expression upon induction of HIF-1[26, 34–36]. Likewise, *rhy-1* and *egl-9* are both known to be up-regulated when HIF-1 becomes common in the cell [37].

Our dataset enables us to perform an equivalent computa-

tional experiment to qPCR by selectively looking at expression of a few genes at a time. Therefore, we queried the changes in expression of *rhy-1*, *egl-9*, *nhr-57* and *lam-3* as a negative control. 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.

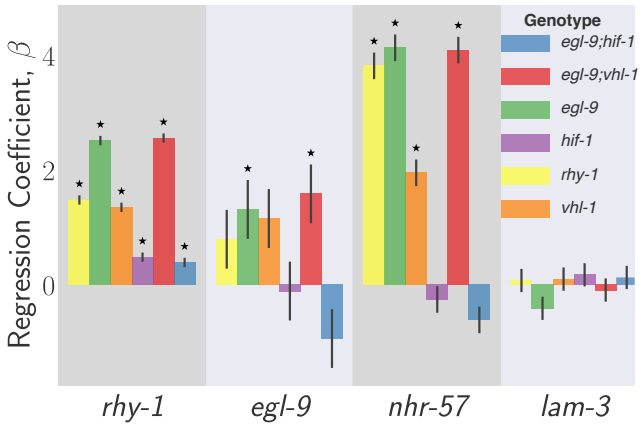

Fig. 4. Top: *In silico* qPCR. We extracted four genes (*rhy-1*, *egl-9*, *nhr-57* and *lam-3*, shown on the x-axis) and plotted their regression coefficients, β, as measured for every genotype (represented by one of six colors) to study the epistatic relationships between each gene. Stars above a bar represent a regression coefficient statistically significantly different from 0, meaning that expression is altered relative to a wild-type control. Error bars show standard error of the mean value of β. *nhr-57* is an expression reporter that has been used previously to identify *hif-1* regulators [31, 34]. 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 [37]. 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. 4). We observed changes in *rhy-1* expression consistent with previous literature [34] when *hif-1* is activated. We also observed changes in *egl-9* expression when *egl-9* was mutated. *egl-9* is known as a hypoxia responsive gene [37] Although changes in *egl-9* expression were not statistically significant in *rhy-1* and *vhl-1* mutants, the mRNA levels of *egl-9* trended towards increased expression in these genotypes. 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. Our dataset also 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*, which constitutes a novel observation. Taken together, these results indicate that RNA-seq data is at least equivalent to qPCR for purposes of comparing gene expression of a reporter between genotypes. Using a single reporter we would have been able to reconstruct an important fraction of the genetic relationships between the genes in the hypoxia pathway.

Genes in the hypoxia pathway exhibit genome-wide epistasis. As we have shown, it may be sufficient to extract the regression coefficients of a previously known reporter gene and study just that pattern in order rebuild a genetic pathway from RNA-seq data. However, we felt that by relying on a single gene, or even a handful of genes to rebuild the pathway was throwing out all of the valuable information present in our dataset. Therefore, we decided to explore a new epistatic metric—genome-wide epistasis.

Ideally, any measurement of genome-wide epistasis should conform to certain expectations. First, it should make use of the regression coefficients of as many genes as possible. Second, it should be summarizable in a single, well-defined number. Third, it should have an intuitive behavior, such that the special values of the statistic (maximum, minimum, zero) should have an unambiguous interpretation.

One way of defining genome-wide epistasis is to use linear regressions to describe the relationship between the change in expression for a set of genes caused by a single mutant and the change in expression in the same set of genes caused by a double mutant containing the single mutant. The set of genes to be studied can be defined as the set of differentially expressed genes common to both genotypes. Once the set is defined, the regression coefficient of each gene in the single mutant can be plotted against the difference between the regression coefficients of the double mutant and the single mutant. We reasoned that under ideal conditions, such a plot would have an intuitive explanation. If two genes are acting entirely independently of each other, the plot will show a line with slope equal to 0. This is because is acting on entirely different sets of genes, so the perturbation caused a gene X^- is unchanged in the double mutant, X^-Y^-. If two genes are acting only additively, then the plot will show a line with slope > 0 (and in fact, the slope should be equal to the slope between a plot of the single mutants X^- and Y^-). If two genes share a negative regulatory interaction, then epistasis will be reflected in the plot as a line with a negative slope that should approach -1, because the double mutant, X^-Y^- should have regression coefficients near 0, such that the y-axis becomes equal to $-X^-$. If the two genes have a synthetic interaction, we would expect that the slope must be positive and it must be greater than the slope predicted by an additive model.

In our experiment, we studied two double mutants, *egl-9*;*hif-1* and *egl-9*;*vhl-1*. We wanted to understand how well the global epistasis agreed with the literature based on qPCR of single reporters. Therefore, we fit weighted linear regressions to each of the four possible combinations (*egl-9* vs. *egl-9*;*hif-1*; *hif-1* vs. *egl-9*;*hif-1*; *egl-9* vs. *egl-9*;*vhl-1*; and *vhl-1* vs. *egl-9*;*vhl-1*) to measure the slopes of the lines of best fit.

We observe that the *egl-9*;*vhl-1* mutant has an identical phenotype to the *egl-9* single mutant (slope = 0; see Table. 1). On the other hand, *vhl-1* has a positive slope, indicating that *egl-9* is additive to *vhl-1*. However, this positive slope has to be less than the slope that would be predicted by an additive model because the slope between *egl-9*;*vhl-1* and *egl-9* is not statistically different from zero. 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, as is well-documented in the literature [31].

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 differen-

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.

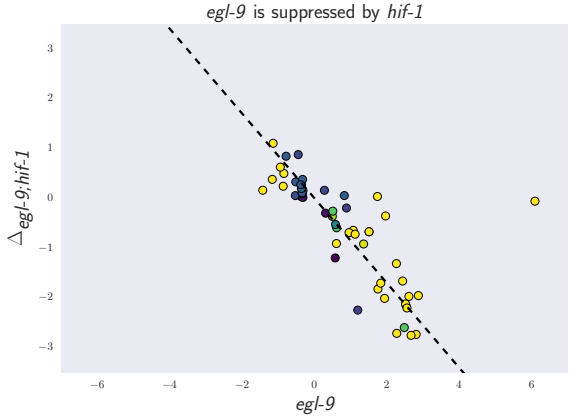

Fig. 5. The mutant *egl-9* transcriptomic phenotype is suppressed by mutations in *hif-1*. The graph shows the β coefficients for *egl-9* in the x-axis, and the change in β coefficient between the *egl-9*;*hif-1* and *egl-9* mutant. The dotted line is the regression line between the complete *egl-9* 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*. It is probably downstream of *egl-9*, and is not likely a *hif-1* target.

tially 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. 5). 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 differentially expressed in every hypoxia pathway 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 the least number of differentially expressed genes), reasoning that these genes should constitute an extremely high-quality picture of the hypoxia response, and should filter out other pathways. We identified 53 genes that satisfied these conditions, of which 10 genes were up-regulated in every mutant, and 13 genes were down-regulated. These genes constitute a core response around the circuit in question, and their behavior should reflect the genetic relationships in our system the best. Although we performed the regressions using all 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).

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 weighted correlation of each gene pair, ordered by increasing functional distance (see Fig. 6). 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

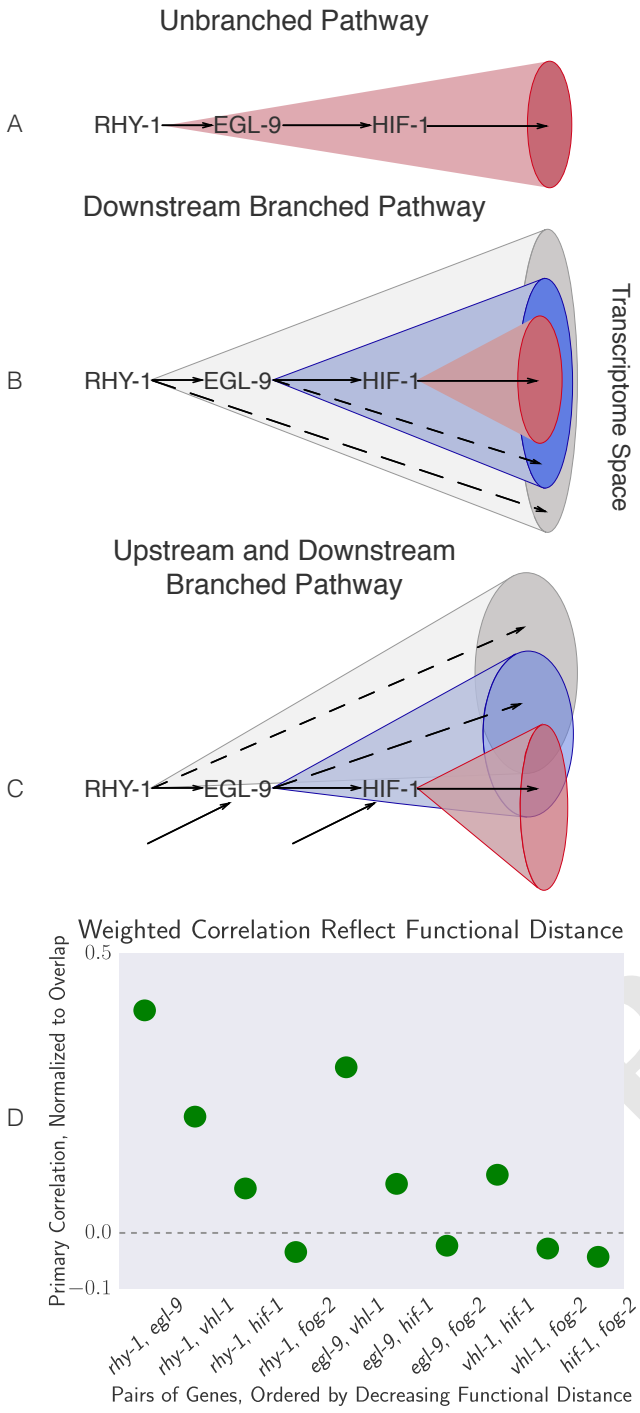

Fig. 6. Theoretically, transcriptomes can be used to order genes in a pathway under certain assumptions. Arrows in the diagrams above are intended to show the direction of flow, and do not indicate valence. **A** A linear pathway in which *rhy-1* is the only gene controlling *egl-9*, which in turn controls *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.

gene order in a pathway, as we have done with the hypoxia pathway[1].

The circuit topology of the hypoxia pathway explains patterns in the data. We noticed that while some of the rank-plots contained a clear positive correlation (see Fig. 2), some of the other rank-plots showed a discernible cross-pattern (see Fig. 7). In particular, this cross-pattern emerged between *vhl-1* and *rhy-1* or between *vhl-1* and *egl-9*, even though genetically *vhl-1*, *rhy-1* and *egl-9* are all inhibitors of *hif-1*. We reasoned that it could be possible that these cross-patterns reflected multiple interaction modes between genes Therefore, we hypothesized that patterns in the rank-plots contained valuable information for decoding more interactions in our circuit.

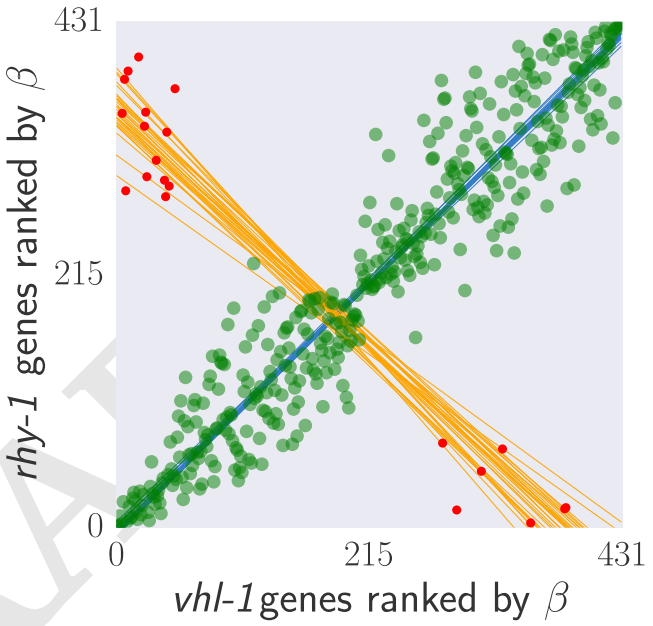

Fig. 7. Top: A feedback loop can generate transcriptomes that are both correlated and anti-correlated. **Bottom**: *hif-1* transcriptome correlated to the *rhy-1* transcriptome. Green large points are inliers to the first regression. Red small points are outliers to the first regression. Only the red 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.

If the logic above is correct, then it should be possible to decouple transcriptomes in a logically consistent way. Currently, transcriptomes are decoupled via subtractive logic. In other words, to identify the *rhy-1*-specific transcriptome (the effects of *rhy-1* not dependent on *egl-9*), subtractive logic might suggest to find the overlap between the two transcriptomes. The genes that are differentially expressed but are not in the overlap would then be considered *rhy-1*-specific transcriptomes. Such a gene set would consider of almost 700 genes. However, this approach suffers from a number of drawbacks, principally that it does not take into account the relationship between the two genes in question. Moreover, these genes have no testable properties: i.e., a gene might not

[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. 6 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.

be in the overlap because it was not identified due to chance in one of the two transcriptomes. In aggreggate, there is no pattern that is present in these genes that can be used to identify them beyond overlapping the two transcriptomes.

rhy-1 and *egl-9* share a well-defined relationship. *rhy-1* inhibits *cysl-1*, which in turn inhibits *egl-9* [30]. Therefore, loss of *rhy-1* leads to inactivation of *egl-9*, which leads to increase in the cellular levels of HIF-1. HIF-1 in turn causes the mRNA levels of *rhy-1* and *egl-9* to increase, as they are involved in the *hif-1*-dependent hypoxia response. However, since *rhy-1* has been mutated, the observed transcriptome is *rhy-1* null; *egl-9* null; HIF-1 on. The situation is similar for a knockout of *egl-9*, except that *rhy-1* is not inactive, and therefore the observed transcriptome is the result of RHY-1 up; *egl-9* null; and HIF-1 on. From this pattern, we conclude that the *egl-9* and *rhy-1* transcriptomes should exhibit a cross-pattern: The positive arm of the cross is the result of the *egl-9* null; HIF-1 on dynamics; and the negative arm reflects the different direction of RHY-1 activity between transcriptomes. However, no negative arm is visible (with the exception of two outliers, which are annotated as pseudogenes in WormBase). Therefore, it is likely that a large portion of all the transcriptomic effects of RHY-1 in this dataset are downstream of *egl-9*.

Next, we wanted to know whether our dataset was able to capture *egl-9 hif-1*-independent transcriptomic effects. We have observed that deletion of *hif-1* leads to a modest increase in the transcription of *rhy-1*, from which we concluded that EGL-9 would be more active in the *hif-1* mutant than in the wild-type. Therefore, we searched for genes that were regulated in opposite manner between the *hif-1* and *hif-1*;*egl-9* mutants, and that were regulated in the same direction between the *hif-1*;*egl-9* and *egl-9* (or *rhy-1*) mutants. We were only able to find a single gene, *clec-88*, which was down-regulated in *hif-1* mutants, but upregulated in every other mutant we studied. Although it may be the case that *egl-9* does not have a *hif-1*-independent transcriptomic phenotype, it is also possible that the change in HIF-1 dosage between a wild-type normoxic animal and a *hif-1* mutant is not sufficient to alter the activity of EGL-9 to a consistently detectable level given our read-depth.

We leveraged this genetic logic to identify a main hypoxia response induced by removing inhibition on *hif-1* (260 genes). We also identified a *vhl-1*-specific response, resulting in 36 genes. We searched for candidates directly regulated by *hif-1*. Initially, we generated this list using the most stringent pattern matching, but this revealed only 2 genes (*R08E5.3* and *nit-1*). A relaxed set of conditions (target genes should go up in all mutants that induce HIF-1, and should not be up in *hif-1* mutants) identified 120 candidate genes.

Enrichment analysis of the hypoxia response. In order to validate that our transcriptomes were correct, and to understand how functionalities may vary between them, we subjected each decoupled response to enrichment analysis using the WormBase Enrichment Suite [38].

Gene enrichment analysis (GEA) showed that the terms 'oxoacid metabolic process' ($q < 10^{-3}$, 3.4 fold-change, 19 genes), 'iron ion binding' ($q < 10^{-3}$, 5.5 fold-change, 10 genes), and 'immune system process' ($q < 10^{-3}$, 3.4 fold-change, 17 genes) were enriched with the lowest q-values. GEA also showed enrichment of terms including 'electron carrier activity' ($q < 10^{-1}$, 4.8 fold-change, 5 genes), 'mitochondrion' ($q < 10^{-2}$, 2.5 fold-change, 20 genes) and 'respiratory chain' ($q < 10^{-1}$, 4.6 fold-change, 4 genes) (see Fig. 8). Indeed, *hif-1* has been implicated in all of these biological and molecular functions [35, 39–41]. Phenotype Enrichment Analysis (PEA) revealed that this gene list was enriched in two phenotypes: 'oxygen response variant' ($q < 10^{-2}$, 5.8 fold-change, 7 genes) and 'pleiotropic defects severe early embryo' ($q < 10^{-2}$, 4.4 fold-change, 9 genes). The overrepresented terms from PEA and GEA are biologically directly connected to the process we are studying, which suggests that we have correctly identified the main hypoxic response. As a final test to guarantee the quality of our data, we selected a set of 21 known reporters from the literature and asked whether these reporters were present in our list. We found 5/21 known reporters, which constitutes a statistically significant result ($p < 10^{5}$). The small number of reporters found in this list probably reflects the conservative nature of our estimates. We also analyzed the list of predicted *hif-1* direct targets. Phenotype Enrichment Analysis revealed that this list was significantly enriched in 'oxygen response variant' ($q < 10^{-2}$, 12.3 fold-change, 4 genes) and Tissue Enrichment Analysis (TEA) showed enrichment of the 'coelomic system' ($q < 10^{-1}$, 2.7 fold-change, 16 genes). The *vhl-1*, *hif-1*-independent specific transcriptome was also submitted for enrichment analysis but no terms were significantly enriched.

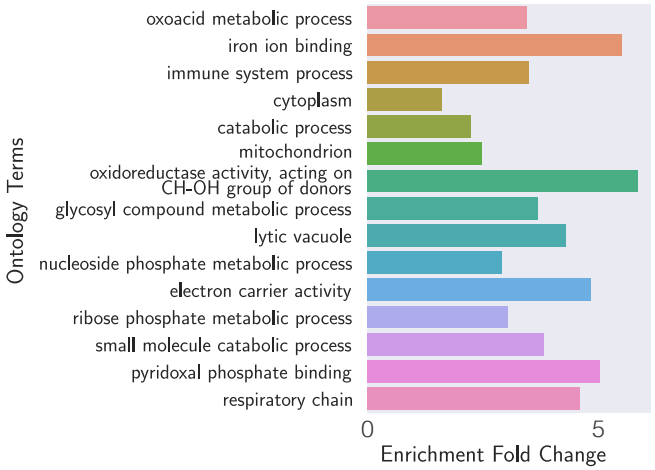

Fig. 8. GEA of genes associated with the main hypoxia response. A number of terms reflecting catabolism and bioenergetics are enriched.

closer is missing.

Identification of non-classical epistatic interactions. *hif-1* has traditionally been viewed as existing in a genetic OFF state under normoxic conditions. However, our dataset indicates that 481 genes show altered expression when it is removed in normoxic conditions. Moreover, we observed positive genome-wide expression correlations between *hif-1* expression levels and *egl-9*, *vhl-1* and *rhy-1* expression levels in spite of the negative regulatory relationships between these genes and *hif-1*. Such positive relationships could indicate a different relationship between these genes than has previously been reported. We wanted to explore whether these genome-wide positive correlations were substantiated by epistatic analyses.

To perform epistatic analyses, we first identified genes that exhibited violations of the canonical genetic model of the hypoxia pathway. To this end, we searched for genes that

exhibited different behaviors between the *egl-9* and the *vhl-1* mutants, or between the *rhy-1* and *vhl-1* (we assume that all results from the *rhy-1* transcriptome reflect a complete loss of *egl-9* activity). We found 27 that satisfied this condition (see Fig. 9). Additionally, many of these genes exhibited new kinds of epistasis. Namely, *egl-9* was epistatic to *vhl-1*. Identification of a set of genes that have a consistent set of relationships with between themselves suggests that we have identified a new aspect of the hypoxia pathway.

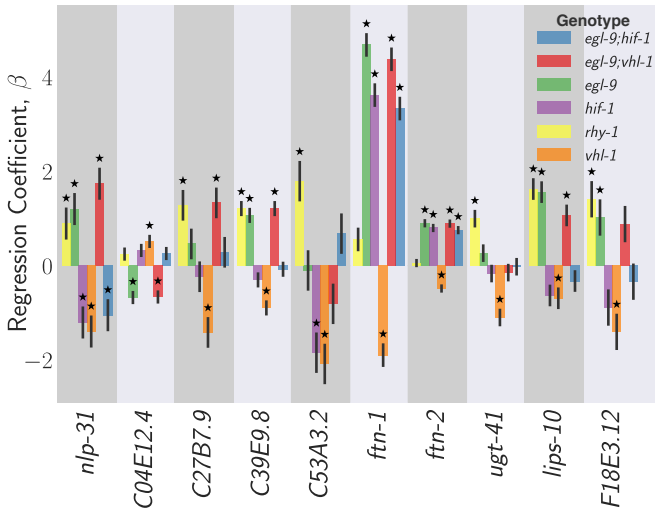

Fig. 9. Genes that have altered differential expression between *egl-9* and *vhl-1* also often exhibit the same epistatic patterns between *egl-9* and *hif-1*, and *egl-9* and *vhl-1*, which suggests they are the result of the same biological effect.

In particular, we focused on three genes, *nlp-31*, *ftn-1* and *ftn-2*, which epistasis patterns that we felt reflected the population well. As a sanity check, we reviewed the literature, and found that *ftn-1* and *ftn-2* are both described in the literature as genes that are responsive to mutations in the hypoxia pathway. Moreover, these genes have been previously described to have aberrant behaviors previously [35, 40], specifically documenting the opposite effects of *egl-9* and *vhl-1*. Probably as a reflection of the oddity of these results, these studies showed that loss of *vhl-1* suppresses *ftn-1* and *ftn-2* using both RNAi and alleles, which allays concerns of strain-specific interference. Moreover, one of these studies showed that *vhl-1*;*hif-1* is epistatic to *hif-1* [35], and that loss of *hif-1* is associated with increased expression of *ftn-1* and *ftn-2*. We observe that *hif-1* is epistatic to *egl-9*;*hif-1*, and that *egl-9* and *hif-1* promote *ftn-1* and *ftn-2* expression.

see Romney2011 figure 2b; Ackerman2012 Fig. 3; Fig.6

This further validates the quality of our RNA-seq data and the analysis, and highlights the power of RNA-seq to identify novel interactions.

A qualitative epistatic analysis of *ftn-1* and *ftn-2* reveals that *egl-9* is epistatic to *hif-1*; that *vhl-1* has opposite effects to *egl-9*; and *vhl-1* is epistatic to *egl-9*. Epistatic analysis of *nlp-31* reveals similar relationships. *nlp-31* expression is decreased by loss of *hif-1*, and promoted by *egl-9*. However, *egl-9* is epistatic to *hif-1*. Like *ftn-1* and *ftn-2*, *vhl-1* has the opposite effect to *egl-9*, but is also epistatic to *egl-9*.

This section is missing figures.

Genome-wide effects of *hif-1*. The high quality of this dataset also provides us directly with a high-level overview of the transcriptional responses that lead to physiologic and metabolic changes in hypoxia. We wanted to better understand the transcriptional changes associated with bioenergetic pathways in *C. elegans*. To this end, we extracted from WormBase all genes associated with the tricarboxylic acid (TCA) cycle, the electron transport chain (ETC) and with the *C. elegans* energy reserve (glycogen metabolism, fatty acid metabolism, etc...). Previous research has described the effects of mitochondrial dysfunction in eliciting the hypoxia response [42], but transcriptional feedback from *hif-1* into bioenergetic pathways has not been well described in *C. elegans*, although it has been extensively described in other organisms (see, for example [19, 24]).

I have searched the literature fairly thoroughly, and haven't seen a complete transcriptional description of hif-1 induced changes in the TCA, ETC. Can anyone let me know if I missed something?

Bio-energetic pathways. Our data shows that most of the enzymes involved in the TCA cycle and in the ETC are down-regulated when HIF-1 is induced in agreement with the previous literature [19].

Bio-energetic or bioenergetic?

Missing ETC citation

However, *fum-1* and the mitochondrial complex II stood out as notable exceptions to the trend, as they were up-regulated in every single genotype that causes deployment of the hypoxia response. *fum-1* catalyses the reaction of fumarate into malate, and complex II catalyses the reaction of succinate into fumarate.

Missing discussion of why this is biologically relevant. Fumarate is a poison for egl-9.

We found two energy reserve genes that were down-regulated by HIF-1. *aagr-1* and *aagr-2*, which are predicted to function in glycogen catabolism [43] were both down-regulated in all the relevant mutants. Three distinct genes involved in energy reserve were up-regulated. These genes were *ogt-1*, an O-linled GlcNac Transferase; *T04A8.7*, an ortholog of human glucosidase acid beta (GBA); and *T22F3.3*, an ortholog of human glycogen phosphorylase isozymes.

Protein synthesis and degradation. *hif-1* is also known to inhibit protein synthesis and translation in varied ways. For example, *hif-1* is known to control the translational machinery indirectly via inhibition of mTOR [44]. However, most reported effects of *hif-1* on the translation machinery are posttranslational, and no reports to date show decreases in transcription of the ribosomal machinery in *C. elegans*. We used the WormBase Enrichment Suite Gene Ontology dictionary [] to extract 143 genes annotated as 'structural constituents of the ribosome' and we queried whether they were differentially expressed in our mutants. *egl-9*, *vhl-1*, *rhy-1* and *egl-9*;*vhl-1* mutants showed differential expression of 91 distinct ribosomal constituents (not all constituents were detected in all genotypes). For every one of these genotypes, these genes were always down-regulated. In contrast, the *hif-1* mutant showed up-regulation of a single ribosomal constituent.

Next, we wanted to know whether *hif-1* has any transcriptional effects on the proteasomal constituents, because no such effects of *hif-1* on the proteasome have been reported

in *C. elegans*. Out of 40 WormBase annotated proteasomal constituents, we found 31 constituents that were differentially expressed in at least one of the four genotypes that induce a hypoxic response. Every gene we found was down-regulated in at least two out of the four genotypes we studied, although in each case the down-regulation was minor. It is impossible to distinguish whether these animals exhibit a decrease in proteasome expression is due to a lower requirement for degradation in these animals due to constitutively depressed translation rates or whether the decrease in expression is a direct result of HIF-1 stabilization.

Discussion

The *C. elegans* hypoxia pathway can be reconstructed entirely from RNA-seq data.

Non-classical epistasis in the hypoxia pathway.

Looking forward.

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. Libraries were sequenced on Illumina HiSeq2500 in single read mode with the read length of 50 nt to an average depth of 15 million reads per sample following manufacturer's instructions. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4.

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.

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 [38, 45–52]. Our analysis is available in a Jupyter Notebook [53]. 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

ACKNOWLEDGMENTS. This work was supported by HHMI with whom PWS is an investigator and by the Millard and Muriel Jacobs Genetics and Genomics Laboratory at California Institute of Technology. This article was written with support of the Howard Hughes Medical Institute. We thank Hillel Schwartz for his help in the fun analysis of ftn-1 epistasis. 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(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. Bray NL, Pimentel HJ, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34(5):525–7.
8. 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.
9. Pimentel HJ, Bray NL, Puente S, Melsted P, Pachter L (2016) Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* p. 058164.
10. Trapnell C et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* 31(1):46–53.
11. 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.
12. Shalek AK et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498(7453):236–40.
13. 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.
14. 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.
15. 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.
16. Adamson B et al. (2016) A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167(7):1867–1882.e21.
17. Dixit A et al. (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167(7):1853–1866.e17.
18. Angeles-Albores D et al. (2016) Transcriptomic Description of an Endogenous Female State in C. elegans. *bioRxiv*.
19. Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. *Cell* 148(3):399–408.
20. Loenarz C et al. (2011) The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO reports* 12(1):63–70.
21. Jiang BH, Rue E, Wang GL, Roe R, Semenza GL (1996) Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *The Journal of biological chemistry* 271(30):17771–17778.
22. Jiang H, Guo R, Powell-Coffman JA (2001) The Caenorhabditis elegans hif-1 gene encodes a bHLH-PAS protein that is required for adaptation to hypoxia. *Proceedings of the National Academy of Sciences of the United States of America* 98(14):7916–7921.

23. Powell-Coffman JA, Bradfield CA, Wood WB (1998) Caenorhabditis elegans Orthologs of the Aryl Hydrocarbon Receptor and Its Heterodimerization Partner the Aryl Hydrocarbon Receptor Nuclear Translocator. *Proceedings of the National Academy of Sciences* 95(6):2844–2849.
24. Semenza GL, Roth PH, Fang HM, Wang GL (1994) Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *The Journal of biological chemistry* 269(38):23757–63.
25. Bishop T et al. (2004) Genetic Analysis of Pathways Regulated by the von Hippel-Lindau Tumor Suppressor in Caenorhabditis elegans. *PLoS Biology* 2(10).
26. 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.
27. Bellier A, Chen CS, Kao CY, Cinar HN, Aroian RV (2009) Hypoxia and the hypoxic response pathway protect against pore-forming toxins in C. elegans. *PLoS Pathogens* 5(12).
28. Huang LE, Arany Z, Livingston DM, Franklin Bunn H (1996) Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its alpha subunit. *Journal of Biological Chemistry* 271(50):32253–32259.
29. Kaelin WG, Ratcliffe PJ (2008) Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway. *Molecular Cell* 30(4):393–402.
30. 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.
31. 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.
32. Yeung KY, Medvedovic M, Bumgarner RE (2003) Clustering gene-expression data with repeated measurements. *Genome biology* 4(5):R34.
33. Epstein ACR et al. (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107(1):43–54.
34. 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.
35. 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).
36. Park EC et al. (2012) Hypoxia regulates glutamate receptor trafficking through an HIF-independent mechanism. *The EMBO Journal* 31(6):1618–1619.
37. Powell-Coffman JA (2010) Hypoxia signaling and resistance in C. elegans. *Trends in Endocrinology and Metabolism* 21(7):435–440.
38. Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW (2016) Tissue enrichment analysis for C. elegans genomics. *BMC Bioinformatics* 17(1):366.
39. Luhachack LG et al. (2012) EGL-9 Controls C. elegans Host Defense Specificity through Prolyl Hydroxylation-Dependent and -Independent HIF-1 Pathways. *PLoS Pathogens* 8(7):48.
40. Romney SJ, Newman BS, Thacker C, Leibold EA (2011) HIF-1 regulates iron homeostasis in caenorhabditis elegans by activation and inhibition of genes involved in iron uptake and storage. *PLoS Genetics* 7(12).
41. Semenza GL (2011) Hypoxia-inducible factor 1: Regulator of mitochondrial metabolism and mediator of ischemic preconditioning. *Biochimica et Biophysica Acta - Molecular Cell Research* 1813(7):1263–1268.
42. Lee SJ, Hwang AB, Kenyon C (2010) Inhibition of respiration extends C. elegans life span via reactive oxygen species that increase HIF-1 activity. *Current Biology* 20(23):2131–2136.
43. Sikora J et al. (2010) Bioinformatic and biochemical studies point to AAGR-1 as the ortholog of human acid ??-glucosidase in Caenorhabditis elegans. *Molecular and Cellular Biochemistry* 341(1-2):51–63.
44. Brugarolas J et al. (2004) Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1 / TSC2 tumor suppressor complex. *Genes and Development* 18(23):1–12.
45. Bokeh Development Team (2014) Bokeh: Python library for interactive visualization.
46. McKinney W (2011) pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* pp. 1–9.
47. Oliphant TE (2007) SciPy: Open source scientific tools for Python. *Computing in Science and Engineering* 9:10–20.
48. Pedregosa F et al. (2012) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
49. Salvatier J, Wiecki T, Fonnesbeck C (2015) Probabilistic Programming in Python using PyMC. *PeerJ Computer Science* 2(e55):1–24.
50. 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.
51. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* 9(3):99–104.
52. Waskom M et al. (2016) seaborn: v0.7.0 (January 2016).
53. 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.