

# Genetic Analysis of a Metazoan Pathway using Transcriptomic Phenotypes

immediate

This manuscript was compiled on February 7, 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*. Interactive graphics for this paper can be found at [www.wormlabcaltech.github.io/mprsq](http://www.wormlabcaltech.github.io/mprsq).**

**Abstract must be re-written.**

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

**G**enetic 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 [4].

**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 four loss of function mutants in the *C. elegans* hypoxia pathway that have been extensively characterized previously [19–22] 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 [23]. These approaches identified the Hypoxia Inducible Factors (HIFs) as an important group of oxygen responsive genes.

Hypoxia Inducible Factors are highly conserved in metazoans [24]. A common mechanism for hypoxia-response induction is heterodimerization between a HIF $\alpha$  and a HIF $\beta$  subunit. The heterodimer then initiates transcription of target genes [25]. The number and complexity of HIFs varies throughout metazoans, with humans having three HIF $\alpha$  subunits and two HIF $\beta$  subunits, whereas in the roundworm *Caenorhabditis elegans* (*C. elegans*) there is a single HIF $\alpha$  gene, *hif-1* [22] and a single HIF $\beta$  gene, *ahr-1* [26]. 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.

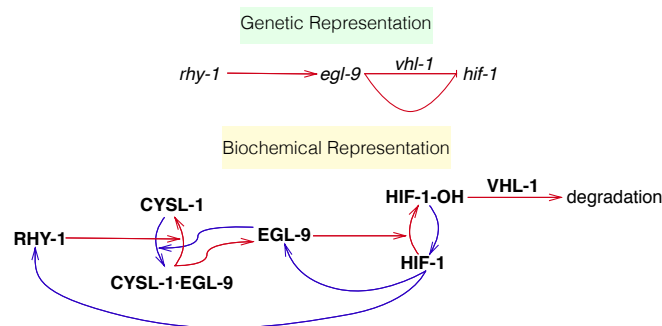
DA, CPR and PWS designed the experiments. CPR selected the genes and extracted mRNA from all mutants. BW made the libraries. IA performed all sequencing. DA developed the mathematical theory. DA wrote all computer code and performed all analyses. DA made all the reporter strains and performed all microscopy.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

The authors declare no conflict of interest.

<sup>2</sup>To whom correspondence should be addressed. E-mail: [pws@caltech.edu](mailto:pws@caltech.edu)

processes such as glycolysis, extracellular matrix modification, autophagy and immunity [23, 27–30].

Levels of HIF $\alpha$  proteins tend to be tightly regulated. Under conditions of normoxia, HIF-1 $\alpha$  exists in the cytoplasm and partakes in a futile cycle of continuous protein production and rapid degradation [31]. HIF-1 $\alpha$  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* [32]. 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 [33] (see Fig. 1).

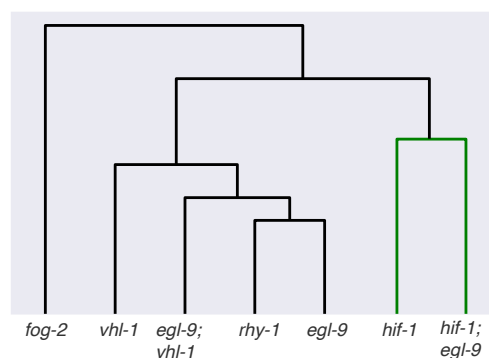


**Fig. 1.** Genetic and biochemical representation of the hypoxia pathway in *C. elegans*. Red arrows are arrows that lead to inhibition of HIF-1, and blue arrows are arrows that increase HIF-1 activity or are the result of HIF-1 activity. EGL-9 is known to exert *vhl-1*-dependent and independent repression on HIF-1 as shown in the genetic diagram. The biochemical diagram does not reflect the *vhl-1*-independent repression of HIF-1 by EGL-9 because that pathway is considerably less well understood.

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 do not appear to be 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/mpsq](http://www.wormlabcaltech.github.io/mpsq).

## 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 ( $\sim 7$  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* (*lf*) transcriptome showed differential expression of 1,487



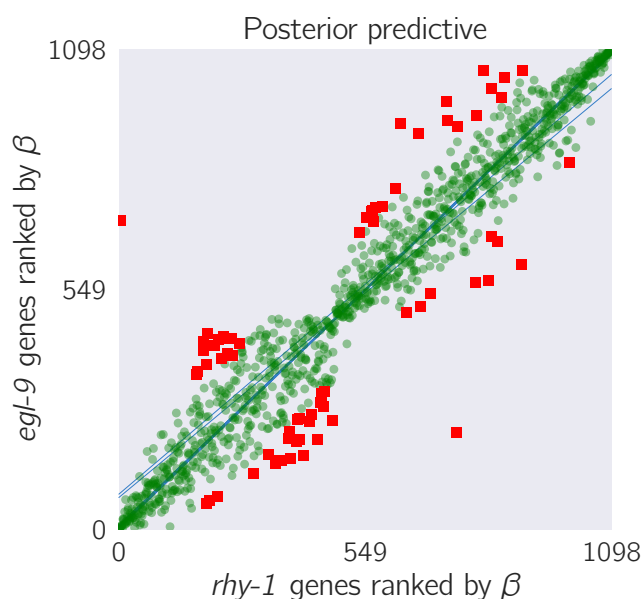
**Fig. 2.** Unsupervised aggregative clustering of various *C. elegans* mutants. Genes cluster in a manner that is biologically intuitive. Genotypes that have an activated hypoxia response (i.e., *egl-9* (*lf*), *vhl-1* (*lf*), and *rhy-1* (*lf*)) cluster far from *hif-1* (*lf*). *hif-1* (*lf*) clusters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

genes, similarly to the 1,816 genes differentially expressed in *rhy-1* (*lf*) mutants. The *vhl-1* (*lf*) transcriptome showed considerably fewer differentially expressed genes (605), possibly reflecting the known fact that it is a weaker controller of *hif-1* (*lf*) than *egl-9* (*lf*) [21]. The *egl-9* (*lf*)/*vhl-1* (*lf*) double mutant transcriptome showed 1,989 differentially expressed genes. The *hif-1* (*lf*) mutant also showed a transcriptomic phenotype involving 481 genes. The *egl-9*/*hif-1* (*lf*) 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 ( $\beta$ ) 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  $\beta$  correspond to larger perturbations. These coefficients can be used to study the RNA-Seq data in question.

Clustering is a well-known technique in bioinformatics that is used to identify relationships between high dimensional data points [34]. We wanted to make sure that clustering by differential expression yielded genetically relevant information. *hif-1* (*lf*) exhibits no obvious phenotypes under normoxic conditions, in contrast to *egl-9* (*lf*), which exhibits an egg-laying (Egl) phenotype in the same environment. *egl-9*/*hif-1* (*lf*) mutants suppress the Egl phenotype. If transcriptomic phenotypes correlate with their identified phenotypes, *hif-1* (*lf*) should cluster with the *egl-9*/*hif-1* (*lf*) double mutant, whereas *egl-9* (*lf*) should cluster away from the *hif-1* (*lf*) mutant. Indeed, when blind, unsupervised clustering was performed on the data, three clusters emerged. *hif-1* (*lf*) and *egl-9*/*hif-1* (*lf*) clustered together, indicating suppression of the *egl-9* (*lf*) phenotype; whereas *egl-9* (*lf*), *egl-9*/*vhl-1* (*lf*), *vhl-1* (*lf*) and *rhy-1* (*lf*) all clustered separately. Finally, our negative control *fog-2* (*lf*) was in its own cluster (see Fig. 2). We conclude that expression profiling measures enough signal to cluster genes in a meaningful manner in complex metazoans.



**Fig. 3.** Strong transcriptional correlations can be identified between genes that share a positive regulatory connection. We took the *egl-9 (lf)* and the *rhy-1 (lf)* transcriptomes, identified differentially expressed genes common to both transcriptomes and ranked each gene according to its differential expression coefficient  $\beta$ . We plotted the rank of each gene in *rhy-1 (lf)* versus the rank of the same gene in the *egl-9 (lf)* transcriptome. The result is an almost perfect correlation. Green, transparent large points mark inliers to the primary regressions (blue lines); red squares mark outliers to the primary regressions.

**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 (lf)* and *egl-9;hif-1 (lf)* to a maximum of 1,249 shared genes between *egl-9 (lf)* and *egl-9;vhl-1 (lf)*). For comparison, we also analyzed a previously published *fog-2 (lf)* transcriptome [18]. The *fog-2* gene is involved in masculinization of the *C. elegans* germline, which enables sperm formation, and is not known to be involved in the hypoxia pathway. The hypoxia pathway transcriptomes and the *fog-2 (lf)* transcriptome showed significant overlap (123–618 genes). Given the similar overlaps between known interactors and an unknown transcriptome, we conclude that the *fog-2 (lf)* mutant we studied acts on the same phenotype as mutants from the hypoxia pathway.

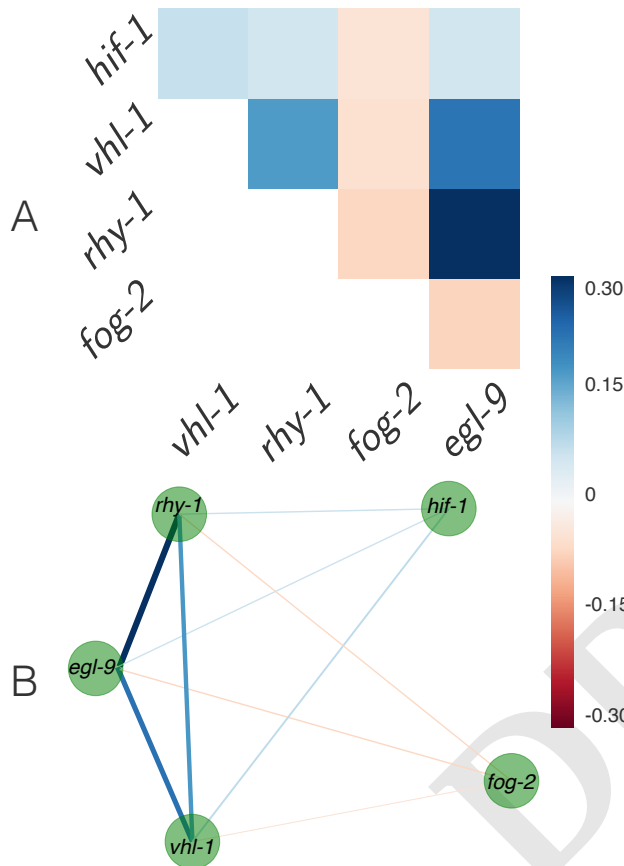
Although overlapping transcriptomes may be enough to conclude that a set of mutants share a phenotype, we wanted

to know whether it was informative to look at quantitative agreement between perturbations. We rank-transformed the regression coefficients  $\beta$  for each transcriptome, and calculated lines of best fit using Bayesian regression with a Student-T distribution to mitigate noise from outliers (see Fig 3). For transcriptomes associated with the hypoxia pathway, we found that these correlations tended to have values higher than 0.9 with a tight distribution around the line of best fit. The correlations for mutants from the hypoxia pathway with the *fog-2 (lf)* mutant were considerably weaker, with magnitudes between 0.6–0.85 and a greater variance around the line of best fit. Although *hif-1* is known to be genetically repressed by *egl-9*, *rhy-1* and *vhl-1* [19, 20], all the correlations between mutants of these genes and *hif-1 (lf)* were positive. The overlap between *hif-1 (lf)* and all other mutants 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,  $N_{\text{overlap}}/N_{g1 \cup g2}$ . The weighted regressions recapitulated a network with three ‘modules’: A control module, a responder module and an uncorrelated module (see Fig. 4). We identified a strong positive interaction between *egl-9 (lf)* and *rhy-1 (lf)*. The magnitude of this weighted correlation derives from the magnitude of the transcriptomes for these mutants (1,487 and 1,816 differentially expressed genes respectively) and the overlap between both genes was extensive, which makes the weighting factor considerably larger than other pairs. The weak correlation between *hif-1 (lf)* and *egl-9 (lf)* is derived from the small size of the *hif-1 (lf)* transcriptome and the small overlap between the transcriptomes. The fine-grained nature of transcriptional phenotypes means that these weighted correlations between transcriptomes of single mutants are predictive of genetic interaction.

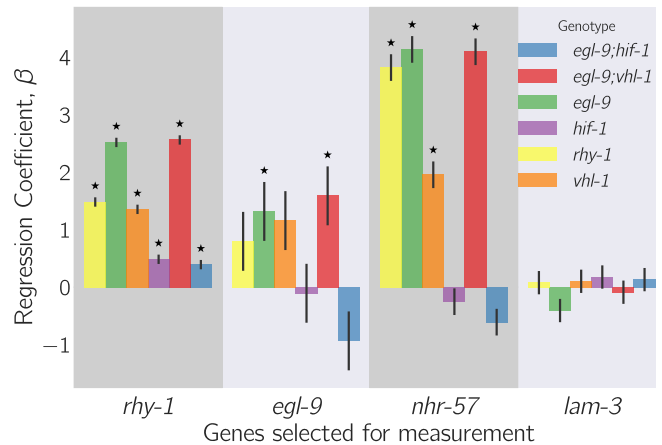
**A quality check of the transcriptomic data reveals excellent agreement with the literature.** One way to establish whether genes are acting additively or epistatically to each other is to perform qPCR of a reporter gene in the single and double mutants. This approach was used to successfully map the relationships within the hypoxia pathway (see, for example [20, 21]). A commonly used reporter is *nhr-57*, which is known to exhibit a several fold increase in mRNA expression when HIF-1 accumulates [20, 29, 35, 36]. Likewise, HIF-1 is known to increase transcription of *rhy-1* and *egl-9* [37].

Our dataset enables us to perform an equivalent computational 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 representative negative control. In our dataset, *nhr-57* is upregulated in *egl-9 (lf)*, *rhy-1 (lf)* and *vhl-1 (lf)*, but remains unchanged in *hif-1 (lf)*. *egl-9;vhl-1 (lf)* had an expression level similar to *egl-9 (lf)*; whereas the *egl-9;hif-1 (lf)* mutant showed wild-type levels of the reporter expression, as reported previously [20].

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 (lf)* expression consistent with previous literature [20] when HIF-1 accumulates. We also observed changes in *egl-9* expression in



**Fig. 4. 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**.



**Fig. 5. 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,  $\beta$ , 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  $\beta$ . *nhr-57* is an expression reporter that has been used previously to identify *hif-1* regulators [20, 21]. 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* (*lf*) 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.

*egl-9* (*lf*). *egl-9* is known as a hypoxia responsive gene [37]. Although changes in *egl-9* expression were not statistically significantly different from the wild-type in *rhy-1* (*lf*) and *vhl-1* (*lf*) mutants, the mRNA levels of *egl-9* still trended towards increased expression in these genotypes. As with *nhr-57*, *egl-9* and *rhy-1* expression was wild-type in *egl-9;hif-1* (*lf*); whereas *egl-9;vhl-1* (*lf*) mutant showed expression phenotypes identical to *egl-9* (*lf*). This dataset also showed that knockout of *hif-1* resulted in a modest increase in the levels of *rhy-1*. This suggests that *hif-1*, in addition to being a positive regulator of *rhy-1*, also inhibits it, which constitutes a novel observation. Taken together, these results indicate that RNA-seq data is 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.** Although it may be sufficient to extract the regression coefficients of a known reporter gene and use it to rebuild a genetic pathway, we felt that by relying on just a single gene, or even a handful of genes to rebuild the pathway discards most of the valuable information present in RNA-seq datasets. 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 plot transcriptome data onto an epistasis plot. In an epistasis plot, the X-axis represents the expected expression of a double mutant  $X^-Y^-$  if  $X$  and  $Y$  interact additively. In other words, it is the sum of the regression coefficients for an isoform calculated from the single mutants  $X^-$  and  $Y^-$ . The Y-axis represents the deviations from the additive (null) model, and can be calculated as the difference between the observed regression coefficient and the predicted regression coefficient. Only genes that are differentially expressed in all three genotypes are plotted.

Epistasis plots can be understood intuitively for simple cases of genetic interactions. If two genes are acting additively on their shared transcriptomes, then all the plotted points will fall along the line  $y = 0$ . If two genes interact in a single, unbranched pathway, then  $X^-$  and  $Y^-$  should have identical phenotypes for  $X^-$ ,  $Y^-$  and  $X^-Y^-$ . It follows that the data points should fall along a line with slope equal to  $-\frac{1}{2}$ . On the other hand, cases where  $X$  is a very strong inhibitor of  $Y$  should result in points that fall along the line with slope equal to  $-1$ . Finally, genes that have a synthetic interaction between them will fall along lines with slopes  $> 0$ . Finally, we can also create models, using only the data from the single mutants, to predict the slope that results for each case stated above, and for each epistatic combination ( $X^-Y^- = X^-$  or  $X^-Y^- = Y^-$ ). Given the biological relevance of the slope of the lines of best fit to the biological relationship between the genes under study, we refer to it as the genome-wide epistasis coefficient, because it integrates information from many different genes into a single number.

In our experiment, we studied two double mutants, *egl-9;hif-1 (lf)* and *egl-9;vhl-1 (lf)*. We wanted to understand how well the global epistasis agreed with the literature based on qPCR of single reporters. Therefore, we performed orthogonal distance regression on to the two gene combinations we studied (*egl-9* and *vhl-1*; and *egl-9* and *hif-1*) to determine the epistasis coefficient for each gene pair. We generated models for a number of pathways using the single mutant data. For every simulation, as well as for the observed data, we used parametric bootstraps to generate probability distributions of the epistasis coefficients. This enables us to test whether two coefficients are statistically distinguishable from each other.

We generated predictions for the epistasis coefficient that the double mutant should have if *egl-9* was epistatic over *vhl-1* (if the double mutant looked like *egl-9 (lf)*,  $s_{Pred} = -0.42$ ) using only the single mutant data. Next, we plotted the double mutant onto an epistasis plot and calculated the line of best fit. We observed that the observed slope agreed exactly with the predicted slope ( $s_{Obs} = -0.42$ ) (see Fig. 6). The epistasis coefficient is close to  $-0.5$ , which suggests that *egl-9* and *vhl-1* interact primarily in a linear pathway. However, when we simulated a linear pathway using the single mutant data, we observed that the linear pathway has negligible probability density at  $-0.42$ , ( $p = 0$ ). Therefore, *egl-9* acts on the downstream phenotype effector (HIF-1) in a more complex manner than a linear pathway with *vhl-1 (lf)*. A genetic pattern that can lead to epistasis coefficients between  $-\frac{1}{2}$  and 0 is a branched pattern, where the gene that is epistatic interacts with the output phenotype via two arrows of the same valency. Overall, this suggests that *egl-9* and *vhl-1* form a branched circuit with *hif-1*. Since *egl-9* was epistatic over *vhl-1*, this

suggests that *egl-9* acts to inhibit HIF-1 in *vhl-1*-dependent and independent ways [21].

We repeated our calculations for *egl-9* and *hif-1*. We predicted that these genes would exhibit an epistatic coefficient equal to  $-0.72$ , and we observed a coefficient of  $-0.84$  ( $p = 0.06$ ). We compared the observed distribution with the expected distribution for an pathway where *hif-1 (lf)* is strongly repressed by *egl-9 (lf)*. We found that this distribution overlapped extensively with the empirical distribution ( $p = 0.53$ ). Therefore, we conclude that *hif-1* is very strongly suppressed by *egl-9*.

### Transcriptomic decorrelation can be used to infer functional distance.

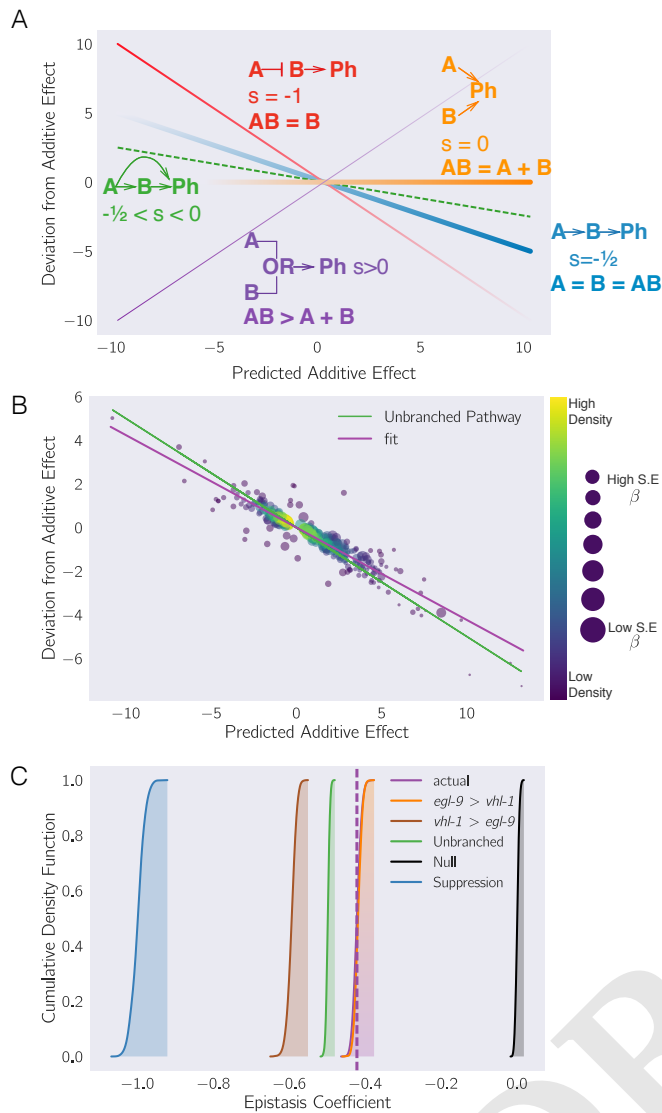
We were interested in figuring out whether RNA-Seq could be used to identify functional interactions within a genetic pathway. Although there is no *a priori* reason why global gene expression should reflect functional interactions, the strength of the unweighted correlations between genes in the hypoxia 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  $N_{Intersection}/N_{Union}$ . We plotted the 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<sup>1</sup>.

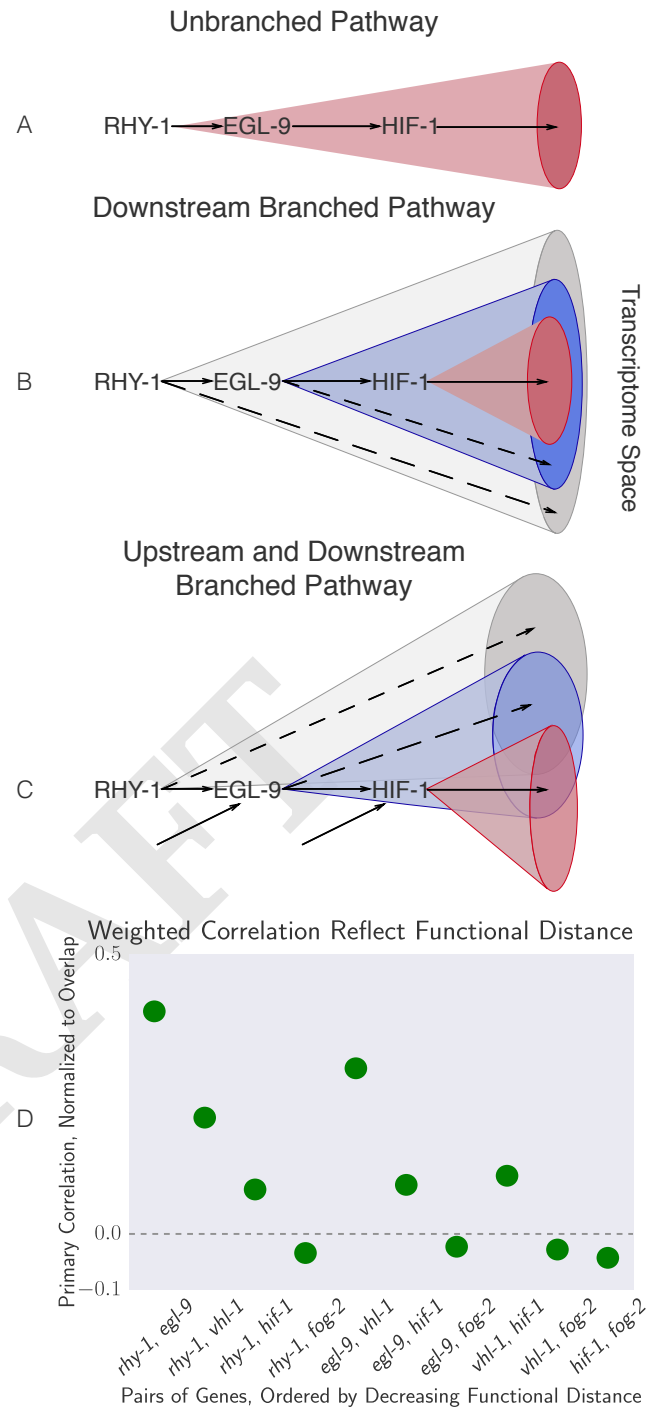
### The circuit topology of the hypoxia pathway explains patterns in the data.

We noticed that while some of the rank-plots contained a clear positive correlation (see Fig. 3), some of the other rank-plots showed a discernible cross-pattern (see Fig. 8). In particular, this cross-pattern emerged between *vhl-1 (lf)* and *rhy-1 (lf)* or between *vhl-1 (lf)* and *egl-9 (lf)*, even though

<sup>1</sup>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.

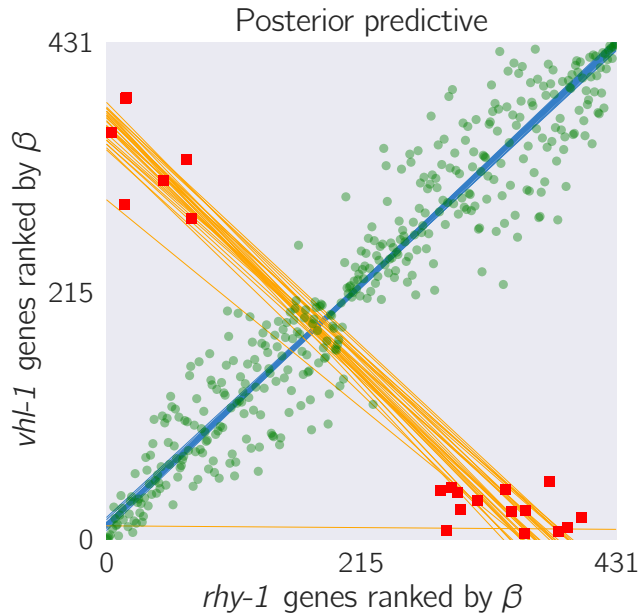


**Fig. 6. (A)** Schematic diagram of an epistasis plot. The X-axis on an epistasis plot is the expected coefficient for a double mutant under an additive model (null model). The Y-axis plots deviations from this model. Double mutants that deviate in a systematic manner from the null model exhibit genome-wide epistasis ( $s$ ). To measure  $s$ , we perform a linear regression on the data. The slope of the line of best fit is  $s$ . This coefficient is related to genetic architectures. Genes that act additively on a phenotype (Ph) will have  $s = 0$ ; whereas genes that act along a single unbranched pathway will have  $s = -1/2$ . Strong repression is reflected by  $s = -1$ . Cases where  $s > 0$  correspond to synthetic interactions, and in the limit as  $s \rightarrow \infty$ , the synthetic interaction is most likely to represent an OR-gate. Cases where  $0 < s < -1/2$  correspond to circuits that have positive branches; whereas cases where  $-1/2 < s < -1$  correspond to cases where the branches have different valency. Cases where  $s < -1$  represent inhibitory branches. **(B)** Epistasis plot showing that the *egl-9;vhl-1* (lf) transcriptome deviates significantly from a null additive. Points are colored qualitatively according to density (purple—low, yellow—high) and size is inversely proportional to the standard error of the y-axis (larger points, higher accuracy). The purple line is the line of best fit from an orthogonal distance regression. **(C)** Bootstrapped cumulative density function for the observed genome-wide epistasis coefficient for *egl-9* and *vhl-1*. Dashed purple line shows the mean value of the data. Using the single mutants, we simulated coefficient distributions for a linear model; an additive model; a model where either *egl-9* or *vhl-1* suppresses the other gene strongly and epistasis models where the double mutant has a phenotype equal to one of the single mutants. We find that the double mutant matches the predicted epistasis curve for *egl-9;vhl-1* (lf) = *egl-9* (lf) (orange and purple). The lack of overlap between the purple/blue curve (observed epistasis) and the distribution for the linear pathway strongly suggests that EGL-9 acts on HIF-1 in *vhl-1*-dependent and independent ways.



**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. We hypothesize the rapid decay in correlation is probably due to a mixture of upstream and downstream branching that happens along this pathway.

genetically *vhl-1*, *rhy-1* and *egl-9* are all inhibitors of *hif-1* (*lf*). 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. 8. Top:** A feedback loop can generate transcriptomes that are both correlated and anti-correlated. **Bottom:** *hif-1* (*lf*) transcriptome correlated to the *rhy-1* (*lf*) transcriptome. Green large points are inliers to the first regression. Red squares 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 be in the overlap because it was not identified due to chance in one of the two transcriptomes. In aggregate, 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 [33]. 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

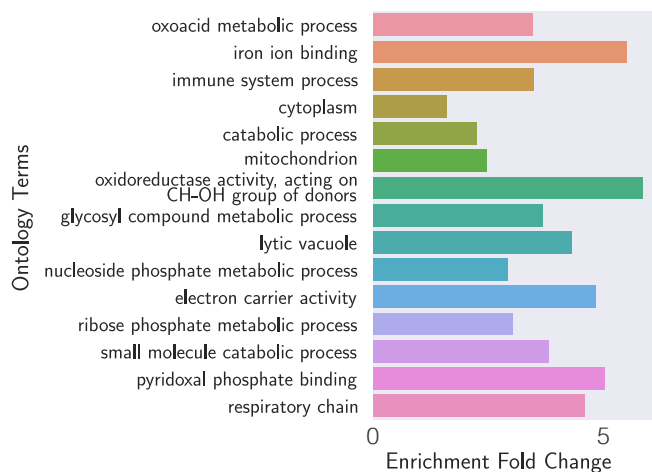
for *egl-9* (*lf*), except that RHY-1 is not inactive, and therefore the observed transcriptome is the result of RHY-1 ‘up’; EGL-9 ‘null’; and HIF-1 ‘on’. From this pattern, we conclude that the *egl-9* (*lf*) and *rhy-1* (*lf*) transcriptomes should exhibit a cross-pattern when plotted against each other: 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 or possibly all the transcriptomic effects of RHY-1 in this dataset are downstream of *egl-9* (*lf*).

Next, we wanted to know whether our dataset was able to capture *hif-1*-independent transcriptomic effects of *egl-9*. We have observed that *hif-1* (*lf*) 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* (*lf*) mutant than in the wild-type. Therefore, we searched for genes that were regulated in opposite manner between *hif-1* (*lf*) and *egl-9*;*hif-1* (*lf*), and that were regulated in the same direction between the *egl-9*;*hif-1* (*lf*) and *egl-9* (*lf*) (or *rhy-1* (*lf*)) mutants. We were only able to find a single gene, *clec-88*, which was down-regulated in *hif-1* (*lf*), 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* (*lf*) animal 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). Although the hypoxic response is likely to involve between five and ten times more genes, this is a conservative estimate that minimizes false negative results, since these changes were identified in four independent genotypes with three replicates each. 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* (*lf*)) 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. 9). Indeed, *hif-1* (*lf*) 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’



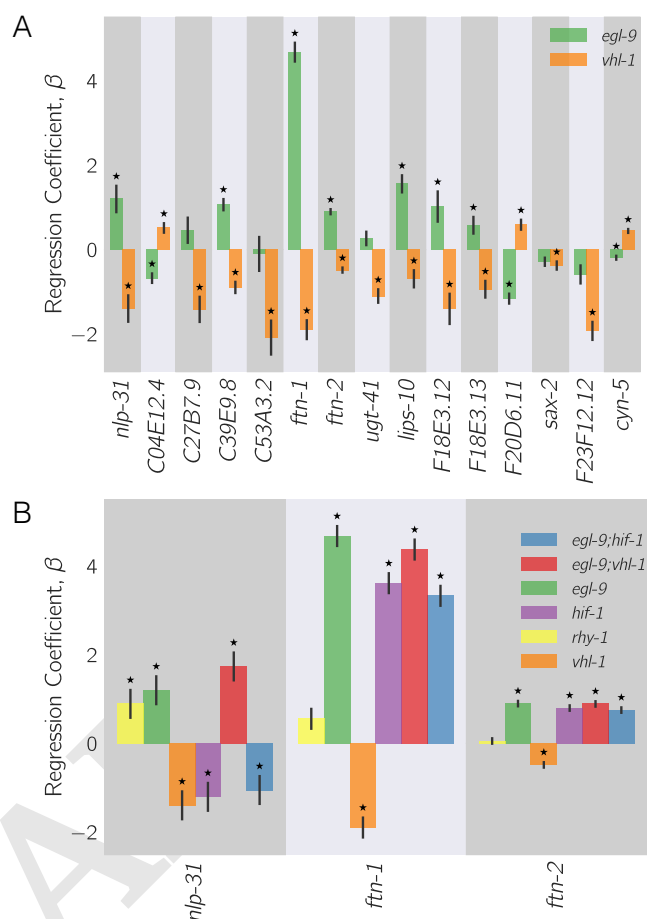
**Fig. 9.** GEA of genes associated with the main hypoxia response. A number of terms reflecting catabolism and bioenergetics are enriched.

( $q < 10^{-2}$ , 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.

**Identification of non-classical epistatic interactions.** *hif-1* (*lf*) 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* (*lf*) expression levels and *egl-9* (*lf*), *vhl-1* (*lf*) and *rhy-1* (*lf*) 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 *egl-9* (*lf*) and *vhl-1* (*lf*), or between *rhy-1* (*lf*) and *vhl-1* (*lf*) (we assume that all results from the *rhy-1* (*lf*) transcriptome reflect a complete loss of *egl-9* activity). We found 27 that satisfied this condition (see Fig. 10). 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. 10.** (A) 27 genes in *C. elegans* exhibit non-classical epistasis in the hypoxia pathway, characterized by an opposite phenotypes of the *vhl-1* (*lf*) and *egl-9* (*lf*) (or *rhy-1* (*lf*)) mutants. Shown are a random selection of the 27 genes for illustrative purposes. (B) Representative genes showing that non-canonical epistasis shows a consistent pattern. *vhl-1* (*lf*) mutants have an opposite effect to *egl-9* (*lf*), but *egl-9* remains epistatic to *vhl-1* and loss of function mutations in *hif-1* suppress the *egl-9* (*lf*) phenotype.

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* (*lf*) and *vhl-1* (*lf*). Probably as a reflection of the oddity of these results, these studies showed that loss of *vhl-1* (*lf*) 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* 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*, and that *egl-9* and *hif-1* both 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 in *hif-1 (lf)*, and increased in *egl-9 (lf)*. However, *egl-9* is epistatic to *hif-1*. Like *ftn-1* and *ftn-2*, *vhl-1* has the opposite effect to *egl-9*, yet is epistatic to *egl-9*.

**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 (see Fig 11). 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 [23, 27]).

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?

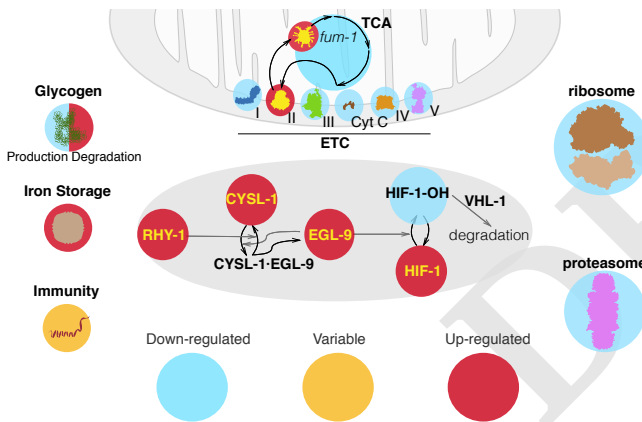


Fig. 11. A graphic summary of the genome-wide effects of HIF-1 from our RNA-seq data.

#### Bio-energetic pathways.

Bio-energetic or bioenergetic?

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 [23].

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. Complex II has been identified as a source of reserve respiratory capacity in neonatal rat cardiomyocytes previously [43].

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 [44] 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-linked 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 (lf)* 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 [45]. 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 (lf)*, *vhl-1 (lf)*, *rhy-1 (lf)* and *egl-9 (lf);vhl-1 (lf)* 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, *hif-1 (lf)* 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.** We have presented the first genetic pathway reconstruction in a multicellular organism using whole-organism RNA-seq to measure transcriptomic phenotypes. We were able to reconstruct first-order and second-order interactions. We were able to infer order of action (*rhy-1* activates *egl-9*, *egl-9* and *vhl-1* inhibit *hif-1*), and we were able to infer from genome-wide epistatic measurements that *egl-9* exerts *vhl-1*-dependent and independent inhibition on *hif-1*.

**HIF-1 and the cellular environment.** In addition to reconstructing the pathway, our dataset afforded us the opportunity to observe a wide variety of physiologic changes that occur when the HIF-1-dependent hypoxia response is activated. In particular, we observed down-regulation of most components of the TCA cycle and the mitochondrial electron transport chain. As an exception, *fum-1* and the mitochondrial complex II, which are involved in fumarate metabolism within these pathways

were up-regulated. The mitochondrial complex II catalyses the reaction of succinate into fumarate. Complex II is known to be important for hypoxic survival in rat cardiomyocyte cells [43]. Complex II may play a similar role in *C. elegans*. The product of complex II activity is fumarate. In mouse embryonic fibroblasts, fumarate has been shown to antagonize HIF-1 prolyl hydroxylase domain (PHD) enzymes, which are orthologs of EGL-9. Upregulation of complex II by HIF-1 during hypoxia may therefore result in increased intracellular levels of fumarate, which in turn could lead to artificially high levels of HIF-1 (if the inhibitory role of fumarate is conserved in *C. elegans*) even after hypoxic conditions have vanished in the absence of concurrent metabolic changes.

Under this framework, the up-regulation of *fum-1* agrees with intuition. By up-regulating *fum-1*, we speculate that *C. elegans* may be capable of harnessing reserve respiratory capacity via complex II, while rapidly degrading the excess fumarate that is generated. Degrading fumarate rapidly may allow *C. elegans* to maintain plasticity in the hypoxia pathway, keeping the pathway sensitive to oxygen levels.

**Non-classical epistasis in the hypoxia pathway.** The observation of almost 30 genes that exhibit a specific pattern of non-classical epistasis reveals new aspects of the pathway. Some of these non-classical epistases had been observed previously, but no satisfactory mechanism has been proposed to explain this biology. [40] and [35] suggest that HIF-1 integrates information on iron concentration in the cell to bind to the *ftn-1* promoter, but could not definitively establish a mechanism. In particular, it is unclear why deletion of *hif-1* induces *ftn-1* expression, deletion of *egl-9* also causes induction of *ftn-1* expression, but deletion of *vhl-1* removes this inhibition. Moreover, [39] have previously reported that certain genes important for the *C. elegans* immune response against pathogens reflect similar expression patterns. Their interpretation was that *swan-1*, a binding partner to EGL-9 [46], is important for modulating HIF-1 activity in some manner. The lack of a conclusive double mutant analysis in this work means the role of SWAN-1 in modulation of HIF-1 activity remains to be demonstrated. At any rate, mechanisms that call for additional transcriptional modulators become more unlikely given our data the large number of genes with different biological functions that exhibit the same pattern.

Genotype	Interpretation
wt	HIF-1 → <i>ftn-1</i> ⇐ HIF-1-OH
<i>egl-9</i>	HIF-1 ⇨ <i>ftn-1</i> HIF-1 activates/ HIF-1-OH represses
<i>vhl-1</i>	HIF-1 ⇨ <i>ftn-1</i> ⇨ HIF-1-OH      HIF-1-OH represses
<i>hif-1</i>	<i>ftn-1</i> Depends on specific activities/ concentrations at S.S.
<i>egl-9; hif-1</i>	<i>ftn-1</i> Depends on specific activities/ concentrations at S.S.
<i>egl-9; vhl-1</i>	HIF-1 ⇨ <i>ftn-1</i> HIF-1 activates/ HIF-1-OH represses

**Fig. 12.** A toy model showing that an interpretation where HIF-1-hydroxyl is biochemically active can potentially explain how genes that exhibit non-canonical epistasis are regulated.

One way to resolve this problem without invoking additional genes is to model HIF-1 as a protein with both activating and

inhibiting states. In fact, HIF-1 already exists in two states in *C. elegans*: unmodified HIF-1 and HIF-1-hydroxyl. Under this model, HIF-1-hydroxyl would inhibit gene expression, whereas HIF-1 drives it. Loss of *vhl-1* stabilizes HIF-1-hydroxyl, which will cause inhibition of genes to which both forms of the protein bind; whereas *egl-9* (*lf*) selectively removes all HIF-1-hydroxyl, indirectly stimulating accumulation of HIF-1 and promoting gene activity. Whether deletion of *hif-1* (*lf*) is activating or inhibiting will depend on the relative contributions of each protein activity under normoxia (see Fig. 12).

The possibility that HIF-1-hydroxyl has a function has not been previously considered in the existing literature, although experts have wondered about the possibility that HIF-1-hydroxyl may have transcriptional effects independent of HIF-1 (William Kaelin, pers. comm.). Here, we draw multiple, varied lines of circumstantial evidence to suggest that HIF-1 hydroxylation plays a role in the functionality of the hypoxia pathway. First, HIF-1-hydroxyl is challenging to study genetically because no mimetic mutations are available with which to study the pure hydroxylated HIF-1 species. Moreover, mutations in the Von-Hippel Landau gene stabilize the hydroxyl species, but also increase the quantity of HIF-1 by mass action. Since HIF-1 is detected low levels in cells under normoxic conditions, total HIF-1 protein (unmodified HIF-1 plus HIF-1-hydroxyl) is often tacitly assumed to be vanishingly rare.

Our data shows that there are hundreds of genes that change expression in response to loss of *hif-1* under normoxic conditions. This establishes that there is sufficient total HIF-1 protein to be biologically active. Previous literature showing substantial changes in mRNA expression of *ftn-1* support our claim that under normoxia *hif-1* (*lf*) is biologically relevant. Moreover, our analysis of the hypoxia pathway using transcriptomic phenotypes reveals that *hif-1* (*lf*) shares main positive correlations with *egl-9* (*lf*), *rhy-1* (*lf*) and *vhl-1* (*lf*), and that each of these genotypes also shows a secondary negative rank-ordered expression correlation with each other. These cross-patterns between all loss of function of inhibitors of HIF-1 and *hif-1* (*lf*) can be most easily explained if HIF-1-hydroxyl is biologically active.

An additional argument in favor of the activity of HIF-1-hydroxyl is a homeostatic argument. At any point in time, the cell must measure the levels of multiple small molecules at once. Strictly speaking, the *hif-1*-dependent hypoxia response integrates information from O<sub>2</sub>, α-ketoglutarate (2-oxoglutarate) and iron concentrations in the cell. One way to encode this information is by encoding it only in the effective hydroxylation rate of HIF-1 by EGL-9. Then the dynamics in this system will evolve exclusively as a result of the total amount of HIF-1 in the cell. Such a system can be sensitive to fluctuations in the absolute concentration of HIF-1 [47]. In the case of severe hypoxia, when the levels of HIF-1 are expected to rise enormously within the cell, simple information integration via EGL-9 would probably be sufficient encoding for a subset of protective genes.

For yet other set of genes that must change expression in response to the hypoxia pathway, it may not make as much sense to integrate metabolite information exclusively via EGL-9-dependent hydroxylation of HIF-1. In particular, genes that may increase survival in mild hypoxia may benefit from homeostatic regulation that is not susceptible to transient changes

in protein copy number. Likewise, genes that are involved in iron or  $\alpha$ -ketoglutarate metabolism (such as *ftn-1*) may benefit from being able to sense, accurately, small and consistent deviations from basal concentrations of these metabolites. For these genes, the information may be better encoded by using HIF-1 and HIF-1-hydroxyl as an activator/repressor pair. Such paradoxical circuits are known to possess distinct advantages for controlling output in a manner that is robust to transient fluctuations in the levels of their components [48, 49].

Our RNA-seq data suggests that one of the targets that HIF-1 may target paradoxically is RHY-1. Although *rhy-1* does not exhibit non-classical epistasis, *hif-1 (lf)* and *egl-9;hif-1 (lf)* both had increased expression levels of *rhy-1*. We speculate that if *rhy-1* is controlled by both HIF-1 and HIF-1-hydroxyl, then this might mean that HIF-1 regulates the expression of its pathway (and therefore itself) in a manner that is robust to total HIF-1 levels.

**Looking forward.** We have demonstrated the first complete reconstruction of a genetic pathway using whole-organism RNA-seq in a complex multicellular organism. Future work must rigorously demonstrate that macroscopically derived genetic rules of interaction hold genome-wide, and if not, how genetic rules deviate from expectations.

## Materials and Methods

**Nematode strains and culture.** Strains used were N2 wild-type Bristol, CB5602 *uhl-1 (lf)* (ok161), CB6088 *egl-9 (lf)* (sa307) *hif-1 (lf)* (ia4), CB6116 *egl-9 (lf)* (sa307) *uhl-1 (lf)* (ok161), JT307 *egl-9 (lf)* (sa307), ZG31 *hif-1 (lf)* (ia4), RB1297 *rhy-1 (lf)* (ok1402). ZG31 *hif-1 (lf)* (ia4) is a null mutant of *hif-1 (lf)* which deletes 1231 bp of the second, third and fourth exons. JT307 contains the null mutant *egl-9 (lf)* (sa307) which is a 243 bp deletion. RB1297 contains null mutation *rhy-1 (lf)* (ok1402) with an estimated 700 bp deletion constructed by the OMRF Knockout Group. CB5602 contains the deletion mutation of *uhl-1 (lf)* (ok161). CB6088 contains *egl-9 (lf)* (sa307); *hif-1 (lf)* (ia4). CB6116 contains *egl-9 (lf)* (sa307) *uhl-1 (lf)* (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 20°C 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 100 $\mu$ L of TE pH 8.0 at 4°C in 0.2mL PCR tubes. After settling and a brief spin in microfuge approximately 80 $\mu$ L of TE was removed from the top of the sample and individual replicates were snap frozen in liquid N<sub>2</sub>. These replicate samples were then digested with Proteinase K for 15min at 60° in the presence of 1% SDS and 1.25 $\mu$ L 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 dH<sub>2</sub>O 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.** 10 ngs of quality checked (RIN > 9.0), total RNA from each sample were reverse-transcribed into cDNA using the Clontech SMARTer Ultra Low Input RNA for Sequencing v3 kit reagents (catalog #634848) in the SMARTSeq2 protocol [50]. RNA was denatured at 70°C for 3 minutes in the presence of dNTPs, oligo dT primer and spiked-in quantitation standards (NIST/ERCC from Ambion, catalog #4456740). After chilling to 4°C, the first strand reaction was assembled using the LNA TSO primer described in [50], and run at 42°C for 90 minutes, followed by denaturation at 70°C for 10 minutes. The entire first strand reaction was then used as template for 13 cycles of PCR using the Clontech v3 kit. Reactions were cleaned up with 1.8X volume of Ampure XP SPRI beads (catalog #A63880) according to the manufacturer's protocol. After quantification using the Qubit High Sensitivity DNA assay, a 3ng aliquot of the amplified cDNA was run on the Agilent HS DNA chip to confirm the length distribution of the amplified fragments. The median value for the average cDNA lengths from all length distributions was 1076bp. Tagmentation of the full length cDNA for sequencing was performed using the Illumina/Nextera DNA library prep kit (catalog #FC-121-1030). Following Qubit quantitation and Agilent BioAnalyzer profiling, the tagmented libraries were sequenced. 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. Spearman correlation of the transcripts per million (TPM) for each genotype showed that every pairwise correlation within genotype was > 0.9.

**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  $i$ :

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

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, Overview.** 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, 51–58]. Our analysis is available in a Jupyter Notebook [59]. 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 XXXXXXXXXXXX

**Weighted Correlations.** Pairwise correlations between transcriptomes where calculated by first identifying the set of differentially expressed genes (DEGs) common to both transcriptomes under analysis. DEGs were then rank-ordered according to their regression coefficient,  $\beta$ . Bayesian robust regressions were performed using a Student-T distribution. Bayesian analysis was performed using the PyMC3 library [55] (`pm.glm.families.StudentT()` in Python). If the correlation has an average value >1, the correlation coefficient was set to 1.

Weights were calculated as the proportion of genes that were < 1.5 standard deviations away from the primary regression out of the entire set of shared DEGs for each transcriptome.

**Epistatic Analysis.** For a double mutant  $X^-Y^-$ , we used the single mutants  $X^-$  and  $Y^-$  to find expected value of the coefficient for a double mutant under an additive model for each isoform  $i$ . Specifically,

$$\beta_{Exp,i} = \beta_{X,i} + \beta_{Y,i}. \quad [2]$$

Next, we found the deviation of the double mutant from the additive model ( $\delta$ ) for each isoform  $i$ :

$$\Delta_i = \beta_{XY,i} - \beta_{Exp,i} \quad [3]$$



To calculate the genome-wide epistasis coefficient, we plotted ( $\beta_{\text{Exp},i}, \delta_i$ ) and found the line of best fit using orthogonal distance regression using the `scipy.odr` package. We performed a parametric bootstrap sampling the ordered tuples with replacement using 5,000 iterations to generate a probability distribution of slopes of best fit.

To understand the epistatic relationships completely, we needed to simulate the possible relationships. There are as many models as epistatic relationships. The epistatic relationships are such that:

$$X^-Y^- = Z, \quad [4]$$

where  $Z$  is some combination of  $X^-$  and  $Y^-$ . Specifically,  $Z$  can take on the values  $X^-$  ( $X$  is epistatic to  $Y$ );  $Y^-$  ( $Y$  is epistatic to  $X$ );  $X^- + Y^-$  (additive);  $k^-, k^- = X^- = Y^-$  (unbranched pathway) and  $X^-, Y^- = WT$  (strong repression, where  $Y^-$  is wild-type).

For each case, we modified equation 3:

$$\Delta_{Z,i} = \beta_{Z,i} - \beta_{\text{Exp},i}. \quad [5]$$

In each case,  $\Delta_{Z,i}$  reduces to  $-\beta_{X,i}$  or  $-\beta_{Y,i}$ . For each model, we performed parametric bootstraps as previously described to generate probability distributions for each model.

We tested whether a given model could be rejected by identifying the fraction of the slopes that were equal to a critical value  $x$  or more extreme under the model. Models were rejected if  $p < 0.05$ .

**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.

- 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.
- Hughes TR et al. (2000) Functional Discovery via a Compendium of Expression Profiles. *Cell* 102(1):109–126.
- Van Driessche N et al. (2005) Epistasis analysis with global transcriptional phenotypes. *Nature genetics* 37(5):471–477.
- 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.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nature reviews. Genetics* 11(1):31–46.
- 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.
- Bray NL, Pimentel HJ, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* 34(5):525–7.
- Patro R, Duggal G, Love MI, Izziyari RA, Kingsford C (2016) Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *bioRxiv* p. 021592.
- Pimentel HJ, Bray NL, Puente S, Melsted P, Pachter L (2016) Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv* p. 058164.
- Trapnell C et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* 31(1):46–53.
- 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.
- Shalek AK et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498(7453):236–40.
- 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.
- 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.
- 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.
- 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.
- 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.
- Angeles-Albores D et al. (2016) Transcriptomic Description of an Endogenous Female State in *C. elegans*. *bioRxiv*.
- 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.
- 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.
- 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.
- 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.

- Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. *Cell* 148(3):399–408.
- 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.
- 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.
- 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.
- 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.
- Bishop T et al. (2004) Genetic Analysis of Pathways Regulated by the von Hippel-Lindau Tumor Suppressor in *Caenorhabditis elegans*. *PLoS Biology* 2(10).
- 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.
- 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).
- 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.
- Kaelin WG, Ratcliffe PJ (2008) Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway. *Molecular Cell* 30(4):393–402.
- Ma DK, Vozdek R, Bhatia N, Horvitz HR (2012) CYSL-1 Interacts with the O<sub>2</sub>-Sensing Hydroxylase EGL-9 to Promote H<sub>2</sub>S-Modulated Hypoxia-Induced Behavioral Plasticity in *C. elegans*. *Neuron* 73(5):925–940.
- Yeung KY, Medvedovic M, Bumgarner RE (2003) Clustering gene-expression data with repeated measurements. *Genome biology* 4(5):R34.
- 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).
- Park EC et al. (2012) Hypoxia regulates glutamate receptor trafficking through an HIF-independent mechanism. *The EMBO Journal* 31(6):1618–1619.
- Powell-Coffman JA (2010) Hypoxia signaling and resistance in *C. elegans*. *Trends in Endocrinology and Metabolism* 21(7):435–440.
- Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW (2016) Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics* 17(1):366.
- 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.
- 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).
- 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.
- 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.
- Pfleger J, He M, Abdellatif M (2015) Mitochondrial complex II is a source of the reserve respiratory capacity that is regulated by metabolic sensors and promotes cell survival. *Cell Death and Disease* 6(7):1–14.
- 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.
- 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.
- Shao Z, Zhang Y, Ye Q, Saldanha JN, Powell-Coffman JA (2010) *C. elegans* swan-1 binds to egl-9 and regulates hif-1-mediated resistance to the bacterial pathogen *Pseudomonas aeruginosa* pao1. *PLoS Pathogens* 6(8):91–92.
- Goentoro L, Shoval O, Kirschner MW, Alon U (2009) The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. *Molecular Cell* 36(5):894–899.
- Hart Y, Antebi YE, Mayo AE, Friedman N, Alon U (2012) Design principles of cell circuits with paradoxical components. *Proceedings of the National Academy of Sciences of the United States of America* 109(21):8346–8351.
- Hart Y, Alon U (2013) The Utility of Paradoxical Components in Biological Circuits. *Molecular Cell* 49(2):213–221.
- Picelli S et al. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* 9(1):171–81.
- Bokeh Development Team (2014) Bokeh: Python library for interactive visualization.
- McKinney W (2011) pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* pp. 1–9.
- Oliphant TE (2007) SciPy: Open source scientific tools for Python. *Computing in Science and Engineering* 9:10–20.
- Pedregosa F et al. (2012) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
- Salvatièr J, Wiecki T, Fonnesbeck C (2015) Probabilistic Programming in Python using PyMC. *PeerJ Computer Science* 2(e55):1–24.
- 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.
- Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* 9(3):99–104.
- Waskom M et al. (2016) seaborn: v0.7.0 (January 2016).
- Pérez F, Granger B (2007) IPython: A System for Interactive Scientific Computing Python: An



DRAFT