Genetic Analysis of a Metazoan Pathway using Transcriptomic Phenotypes

David Angeles-Albores[a,b,*], **Carmie Puckett Robinson**[a,b,*], **Brian Williams**[a], **Igor Antoshechkin**[a], **and Paul W Sternberg**[a,b]

[a]Department of Biology and Biological Engineering, Caltech, Pasadena, USA, 91125; [b]Howard Hughes Medical Institute; [*]These authors contributed equally to this manuscript

This manuscript was compiled on September 29, 2016

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, with all the implications that has for genetic analysis. To that end, we sequenced the transcriptome of four single mutants and two double mutants of the hypoxia pathway in *C. elegans*. We successfully analyzed the single mutants in a blinded fashion to predict the genetic relationships between the genes, and used the double mutants as a test of our predictions and to infer the directionality of the relationship. We show that genes along a pathway tend to decorrelate as a result of alternative regulatory modes and crosstalk with other pathways; and that this decorrelation accurately reflects functional distance between genes. As a by-product of our analysis, we predict 133 genes under the regulation of *hif-1*, and 25 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 | *C. elegans* | hypoxia | transcriptomics

By definition, phenotypes are measurable traits that are related to genotypes via a formal mapping function. A requirement for a genetic relationship to exist is that two genes must act on the same phenotype. However, the converse, that two genes share a phenotype, does not imply that these genes interact. One way to prove genetic interaction is to measure epistasis in a double mutant. Epistasis refers to changes in a phenotype that are not additive and epistatic analysis remains an important cornerstone of genetics today [1].

Previous work in *S. cerevisiae* and *D. discoideum* has shown that transcriptomes contain sufficient information to infer genetic relationships in a simple eukaryote [2, 3]. This early work was performed using microarray technology, which suffered from drawbacks related to sensitivity. New technologies such as RNA-seq [4] do not suffer from these drawbacks. Developments in the area of transcriptomics have also made important progress towards cheaper sequencing [5], better read alignment [6–8] and differential analysis [9, 10]. As a result, RNA-seq has been used to identify key regulatory modules involved in a variety of processes, including T-cell regulation [11, 12], the *C. elegans* linker cell [13], or planarian stem cell identification [14, 15]. However, even in these novel applications, transcriptomes largely serve a descriptive role, and are important for hypothesis generation and target acquisition as opposed to hypothesis testing and model creation.

To investigate the ability of transcriptomes to serve as quantitative phenotypes, we selected mutants in the *C. elegans* hypoxia pathway for transcriptome sequencing. The hypoxia pathway is a conserved pathway that is found in all metazoans [16]. It plays an important role in oxygen and iron homeostasis and in the immune response among others [17, 18], and it is believed to play an important role in cancer appearance and progression, making it an attractive therapeutic target for disease [19]. In *C. elegans* and other systems, HIF-1 is constitutively degraded by a futile cycle that involves hydroxylation by the EGLN1 ortholog EGL-9, followed by ubiquitination by the von Hippel-Lindau Suppressor 1, VHL-1 [20–23]. Inhibition of hydroxylation leads to accumulation of activated HIF-1 [20]. Among the known targets of HIF-1 activation are *rhy-1* and *egl-9*. Increased RHY-1 levels lead to activation of EGL-9 protein by inhibition of CYSL-1 which is an inhibitor of EGL-9 [24, 25].

Here, we show that transcriptomes contain extremely strong, robust signals that can be used to infer relationships between genes in complex metazoans by reconstructing a the hypoxia pathway in *C. elegans* using RNA-seq in a blinded manner. Our goal is not to generate a high-quality database of hypoxia-related genes, but rather to perform a quantitative genetic analysis analogous to classical genetics. Using this experimental setup, we show that various techniques, including pairwise comparisons, clustering or *in silico* qPCR can be used to generate a testable model of genetic interactions. A complete, interactive version of the analysis is also available at www.wormlabcaltech.github.io/mprsq.

Significance Statement

Measurements of global gene expression are often used as descriptive tools that identify 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. 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 wrote the manuscript. CPR performed all experiments. BW performed library preparation. IA performed sequencing. DA performed all analysis and blinded genetic reconstruction. DA, CPR and PWS designed the experiments.

The authors declare no conflict of interest.

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

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

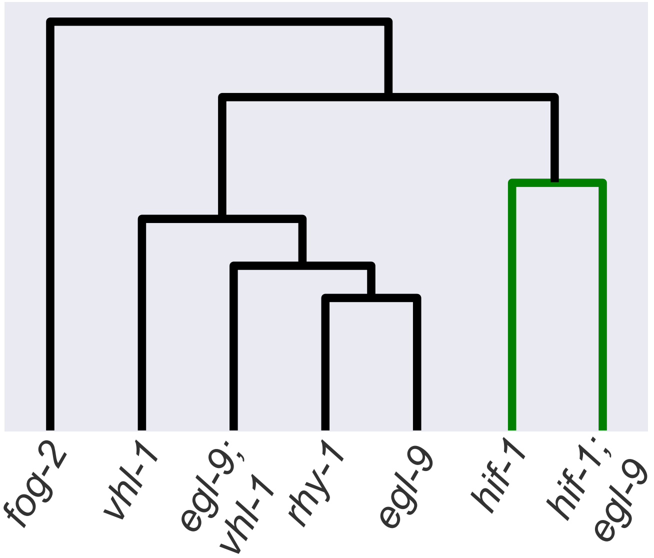

Fig. 1. Blind unsupervised clustering of various *C. elegans* mutants. Genes cluster in a manner that is biologically intuitive. Genes that inhibit *hif-1* (i.e, *egl-9*, *vhl-1*, and *rhy-1*) cluster far from *hif-1*. *hif-1* clusters with the suppressed *egl-9*; *hif-1* double mutant. A control gene,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

Results

Clustering visualizes epistatic relationships between genes.

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

Clustering is a well-known technique in bioinformatics that is used to identify relationships between data [26]. We wanted to make sure that clustering by differential expression yielded genetically relevant information. *hif-1* is known to be inhibited by a pathway involving *egl-9*, *vhl-1* and *rhy-1*. *egl-9*; *hif-1* mutants exhibit suppression of the *egl* phenotype. Given this information, *hif-1* should cluster near the *egl-9*; *hif-1* double 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). Thus, we conclude that expression data contains enough signal to cluster genes in a meaningful manner.

Transcriptomic correlations can predict genetic regulation.

Theoretically, two genes that share linear positive regulation should be positively correlated in their overlapping transcriptomes, whereas two genes that share linear negative regulation should be negatively correlated in their transcriptomes. Formally, if we consider that a gene A has a transcriptome $\{A\}$ associated with it, and if we consider a second gene B with an associated transcriptome $\{B\}$ that is activated by A (that is, $B \in \{A\}$, such that $\{B\} \subset \{A\}$), then it follows that genetic knockout of A or B should both lead to the same perturbation of the transcriptome $\{B\}$. Conversely, it follows that if two mutants have overlapping transcriptomes, and if these transcriptomes have a strong positive association, it is likely that these two genes share a positive regulatory association. In other words, transcriptomic correlation is a good predictor of genetic regulation. For a formal introduction to the genetic logic, see S.I..

Although transcriptomic correlations could theoretically be used for the purposes of identifying genetic regulation, noise from measuring 20,000 genes in multiple different genotypes can cause serious interference with any inferences. Additionally, genes sometimes experience multiple modes of regulation, including positive and negative regulation, from the same gene or pathway. Because we are measuring the system at steady state, both modes of regulation will be measured simultaneously. If a positive and a negative signal are both present in a transcriptome, running a naive regression may result in a value close to zero. Therefore, we took steps to mitigate noise emanating from frequent outliers. As a first mitigation attempt, we rank-tranformed the β coefficients for each mutant. This has the effect of mitigating outliers by resetting the difference between adjacent coefficients to unity. Secondly, we performed robust Bayesian regressions using a Student T distribution as a prior. A Student T distribution decays less quickly than a normal distribution, which causes the model to consider outliers to be less informative than traditional frequentist regressions which effectively use a normal prior.

Having mitigated the effect of outliers, we saw that for certain gene pairs, their transcriptomes correlated very well when genes were ranked by their expression changes (see Fig. 2). Having confirmed that we can extract strong signals from these transcriptomes, we proceeded to generate all pairwise correlations between transcriptomes and we weighted the correlations by the number of genes that participated in the correlation (that were not outliers) divided by the total number of genes detected in all samples. The regression slopes recapitulated a network with three ‘modules’: A control module, a responder module and an uncorrelated module (see Fig. 3). We were able to identify a strong positive interaction between *egl-9* and *rhy-1*. Part of the reason for this lies in the fact that the transcriptomes for these genes consisted of 1487 and 1816 significantly altered genes respectively and the overlap between both genes was quite extensive. On the other hand, none of the primary correlations between *hif-1* and its controlling genes are negative. We suspect that this is a result of the profound control that *hif-1* exerts on *egl-9* and *rhy-1*. See SI A and B for an exhaustive analysis of the expected and observed correlation between each gene pair in this circuit respectively.

Previous work in the hypoxia pathway suggests that this pathway may have feedback loops. Using the same genetic formalism as above, we realized that interactomes due to the fine-grained nature of the data can identify two regulatory interactions if they are of opposite sign. Consider a system in which an arbitrary gene A activates a gene B, which in turn blocks a gene C. Each gene X has a specific transcriptome $\{X\}$.Under this system, B and C should have transcriptomes

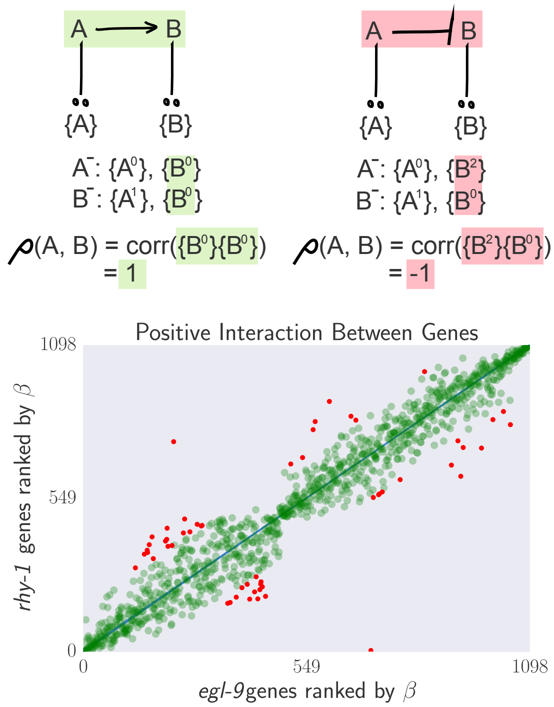

Fig. 2. Top Schematic Diagram showing that genes that interact positively should have a positive transcriptomic correlation, whereas genes that interact negatively should have a negative correlation. Single genes are referred to by their names (A, B), and the transcriptome associated only with gene X is referred to as {X}. We use superscripts to denote expression level. In this case, 0 = no expression (knockout); 1 = WT level; 2 = Greater than WT level. **Bottom** Empirical demonstration that transcriptomes between two interacting genes can be extremely well correlated when genes are ranked by expression changes relative to a wild-type.

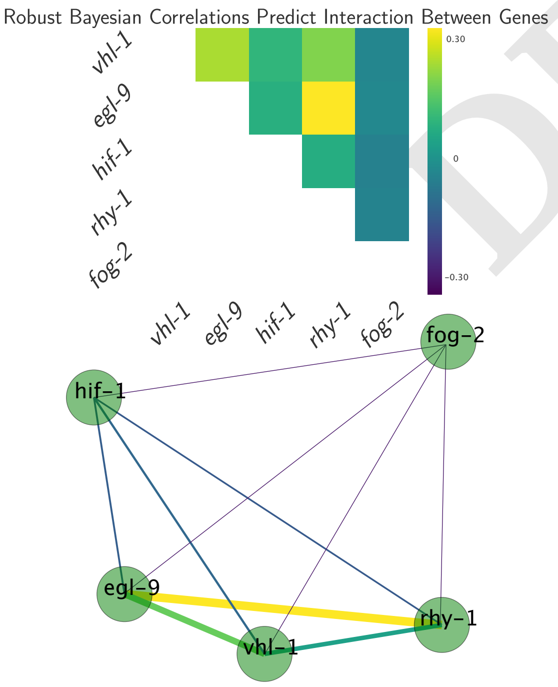

Fig. 3. Top: Heatmap showing pairwise regression values between all single mutants. **Bottom**: Correlation network drawn from the diagram. Edge width is directly proportional to the regression value.

that are negatively correlated. If C activates A, however, then knocking out B should augment expression of C, which should in turn increase expression of A. However, knocking out C should lead to less A, which in turn will lead to less B. Under this thought experiment, suppose that we know the specific transcriptomes associated with A, B and C:$\{A\}$, $\{B\}$, $\{C\}$. Then it must be the case that the genetic knockout of B must have a perturbed transcriptomes $\{A^2\}, \{B^0\}, \{C^2\}$-in other words, knocking out B increases the levels of A, which leads to an overexpression perturbation of the specific transcriptome associated with A, and so forth. On the other hand, knocking out C must lead to the perturbed transcriptomes $\{A^0\}, \{B^0\}, \{C^0\}$. Now, if we were able to correlate each specific transcriptome between correlations, we would find that the specific transcriptomes associated with A and C are anti-correlated; whereas the specific transcriptome associated with B is correlated between both genotypes. This should lead to a characteristic X pattern in the ranked data. Although in this particular example the cross is due to feedback loops, it is important to point out that there are other patterns that could generate crosses.

We investigated whether any pairwise comparisons between our single mutants generated this cross pattern. Indeed, we found that comparing *hif-1* with *rhy-1*, and *hif-1* with *egl-9* yielded negative correlations, as did *rhy-1* and *vhl-1*. While the number of genes that lead these negative correlations is not significant as assessed by a hypergeometric test, these outliers are expected for this circuit (see SI). Statistical information should be integrated holistically with genetic models to assess whether outliers are meaningful or not. Since *hif-1* was off in the conditions under which we performed our experiment, measurement of the *hif-1* transcriptome is difficult and we suspect the small number of outliers is the result of this low resolution.

***in silico* qPCR reveals extensive feedback in the hypoxia pathway.** We realized that our dataset enabled us to perform a sort of *in silico* qPCR. To verify the quality of our data and the veracity of *in silico* qPCR, we first queried the changes in expression of *nhr-57*. This particular reporter has been shown to be under direct control of *hif-1*. Thus, we expected that this gene should go up in *egl-9*, *rhy-1* and *vhl-1*, and it should be unchanged in *hif-1*. The epistasis test, using the double *egl-9*;*hif-1* double mutant should result in no change; whereas the *egl-9*;*vhl-1* double mutant should have a similar change to the *vhl-1* and the *egl-9* mutants. In fact, our datasets reflected these known interactions, showing that the RNA-seq measurements can be used in a semi-quantitative fashion to perform inferences on genetic regulation.

Next, we decided to perform *in silico* qPCR of every gene under scrutiny to get a clearer idea of the relationships between them (see Fig. 5). We found that *rhy-1* transcription levels, and to a lesser extent *egl-9* levels were increased by mutations in *egl-9*, *rhy-1* and *vhl-1*. This suggests that *hif-1* is a positive regulator of *rhy-1*. Given that *rhy-1* post-translationally controls *egl-9* [25], it is unlikely that the increase in *egl-9* is driven by the increase in *rhy-1* levels. Therefore, our experiment also suggests that *hif-1* is a positive regulator of *egl-9*. On the other hand, we also discovered that mutation of *hif-1* increased levels of *rhy-1*. This suggests that *hif-1* is also a negative regulator of *rhy-1*. One potential mechanism through which *hif-1* could be both a positive and a negative regulator would

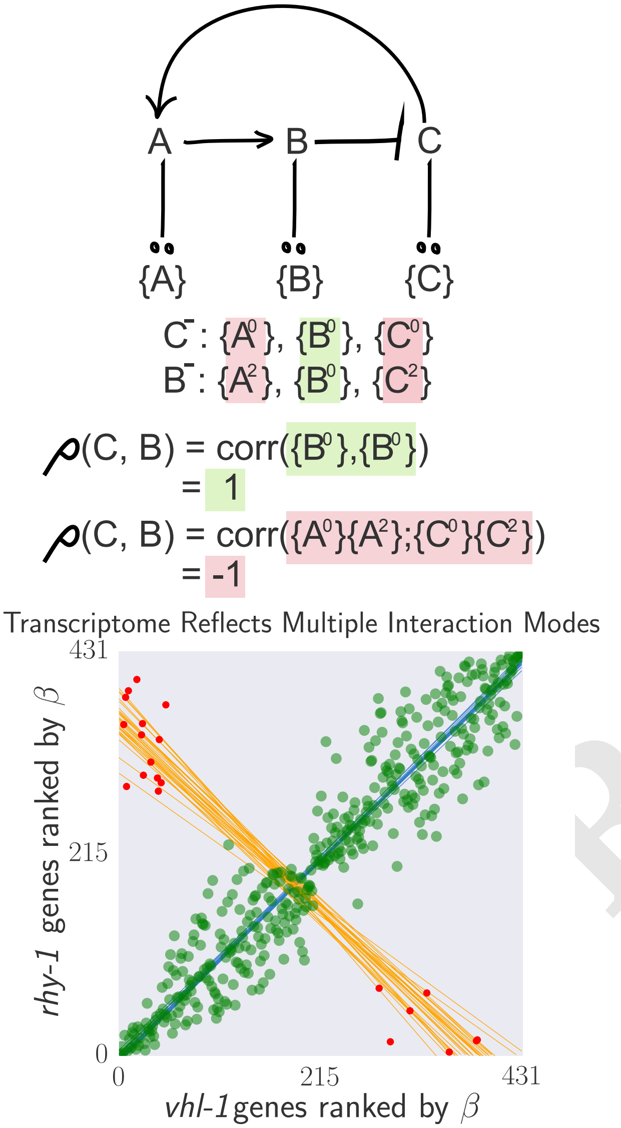

Fig. 4. Top: A feedback loop can generate transcriptomes that are both correlated and anti-correlated. **Bottom**: *hif-1* transcriptome correlated to the *rhy-1* transcriptome. Green large points are inliers to the first regression. Red small points are outliers to the first regression. Only the red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

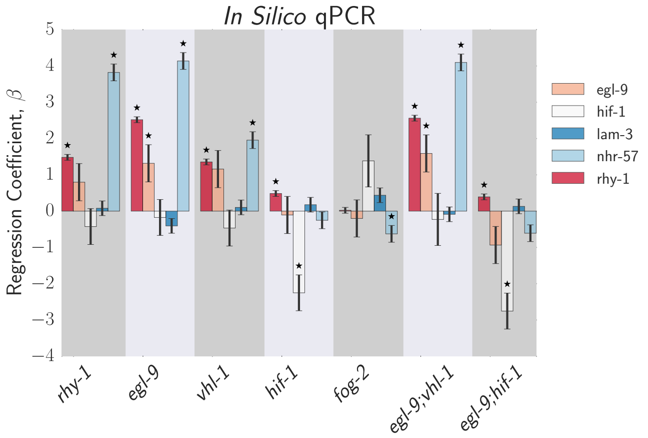

Fig. 5. Top: *In silico* qPCR results. *nhr-57* is an expression reporter that has been used previously to identify *hif-1* regulators [21, 24]. The *nhr-57* mRNA levels replicate what is observed in the literature and serves as a quality control for our dataset. *lam-3* is a negative control that should not be involved in this pathway. Changes in the hypoxia pathway suggest that *hif-1* activates *rhy-1*, and possibly *egl-9*, when it is not hydroxylated. *hif-1* also appears to autoactivate in a hydroxylation-dependent manner.

be for hydroxylation of *hif-1* to change its activity. Under this mechanism, loss of *hif-1* hydroxylation leads to activation of *rhy-1* and *egl-9* as a homeostatic mechanism, whereas excessive hydroxylation causes inhibition of these genes.

Whereas loss of hydroxylation seems to lead to overexpression of *rhy-1* and *egl-9*, there is no change in *hif-1* levels. The only change in expression level of this gene occurs in the *hif-1* mutant[1]. Therefore, we postulate that *hif-1* positively autoregulates itself only in its non-hydroxylated state. We propose that increasing HIF-1 levels through inactivation of EGL-9 does not increase *hif-1* mRNA levels because *hif-1* is already autoactivating at its maximum rate under normoxia, which should happen if HIF-1 has a very strong tendency to autoactivate.

Performing the *in silico* experiment with the *egl-9*;*vhl-1* double mutant shows a similar increase in activity of the two genes in question. This provides confirmatory evidence that *hif-1* up-regulates *egl-9*, but also suggests that *egl-9* and *vhl-1* are epistatic to one another. Such epistasis can only occur in one of two ways: Either the genes are acting linearly, or they are acting in AND gated fashion, with both genes required to mediate an effect. Similarly, the *egl-9*;*hif-1* double mutant exhibits the same expression profile as *hif-1*, which means *egl-9* is an inhibitor of *hif-1*.

In summary, the *in silico* qPCR results suggest that *egl-9* and *vhl-1* act in concert to inhibit *hif-1*. Likewise, these results taken together with the transcriptome-wide cross-patterns that emerge from pairwise comparisons between genes in the hypoxia pathway suggest that there are positive and negative feedback loops feeding into *rhy-1* and possibly *egl-9*. These feedack loops explain why *hif-1* is positively transcriptomically correlated with *egl-9*.

Epistasis effects can be detected and quantified.. To discover whether there is evidence of *egl-9* and *vhl-1* acting independently of each other in our dataset, we identified the genes that were shared between each single mutant and the double

[1] We cannot discard the possibility that this decrease in mRNA levels is not due to NMD or some other decay mechanism, but we consider such a large change unlikely

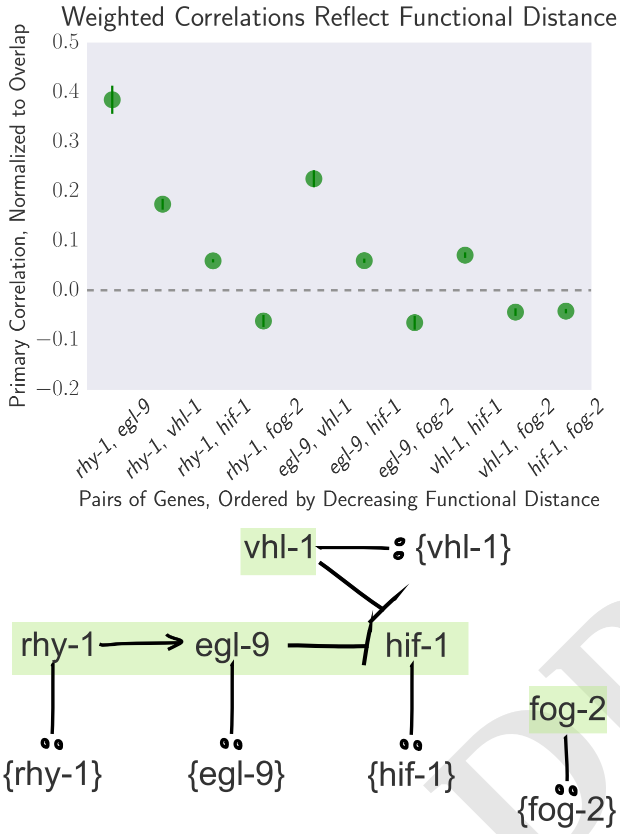

Fig. 6. Top: Pairwise weighted correlations between transcriptomes can be used to infer functional distance between interacting genetic partners. Pairwise correlations are ordered by increasing network distance between genes. Correlations were weighted by the fraction of genes that overlapped between the two genes being compared. Notice that correlations involving the *fog-2* negative control are very near zero. Error bars represent standard deviation of the weighted correlation. Bottom: Simplified schematic of the hypoxia pathway shown to illustrate functional distance between genes in the pathway.

Table 1. Response Modeling of Double Mutants to Single Mutants

Double Mutant	Single Mutant	Δ	SEM	p-value
1. *egl-9*;*vhl-1*	*egl-9*	0.0	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	0.097

Table showing changes between single and double mutants. Δ is the result of a weighted-linear regression (WLS) between β_{single} and $\beta_{\text{double}} - \beta_{\text{single}}$. $\Delta > 0$ 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.

mutant *egl-9*;*vhl-1*. If two genes act, for example, in a linear manner, then the double mutant should exhibit an identical phenotype to each single mutant. To test such a relationship, we can plot the change in β coefficients between a single and double mutant versus the β coefficient in the respective single mutant and fitting a weighted linear regression to measure the slope of best fit. Genes that act in a linear pathway should yield lines with a slope of 0. Genes that have some additive flavor should have slopes greater than 0. Suppression, a hallmark of inhibition, should yield a slope less than 0.

We observe that the *egl-9*;*vhl-1* mutant has an identical phenotype to the *egl-9* single mutant (slope = 0; see Fig. 1). On the other hand, *vhl-1* has a positive slope, indicating that *egl-9* is additive to *vhl-1*. Such partial additivity can be explained if *egl-9* is inhibiting *hif-1* in a *vhl-1*-dependent as well as a *vhl-1*-independent manner [21].

On the other hand, comparison of the *egl-9*;*hif-1* double mutant showed suppression of the *egl-9* transcriptomic phenotype. This suppression is expressed in various ways. First, the double mutant shows less statistically significantly differentially expressed genes than either single mutant. Secondly, these genes change less on average than they do in *egl-9*, but the average change is the same as *hif-1*.

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.

We investigated the possibility that transcriptomic signals might contain relevant information about the degrees of separation by weighting the robust bayesian regression of each pairwise analysis by $N_{\text{Overlap}}/N_{\text{detected}}$. We then 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.

Even in unbranched pathways, this would induce progressive decorrelation between genes.

Identification of novel targets and biological processes in the hypoxia response. So far, our analysis has focused mainly on extracting genetic relationships between the set of mutants we sequenced. It has not escaped our attention that our dataset also provides us with a unique view of the *hif-1*-dependent response in *C. elegans*. In total, we identified 3211 differentially expressed genes that are altered in any of the hypoxia pathway mutants. Of these 3211 genes, 53 genes were differentially expressed in all the hypoxia mutants. Because of the extensive feedback between *hif-1* and *egl-9*, we expected to identify a small subset of genes that were up-regulated or down-regulated consistently in every hypoxia mutant except the *egl-9*;*hif-1* double mutant. We identified 10 genes that were up-regulated in this manner, and 13 genes that were down-regulated (see SI for gene identities). These genes likely constitute a core response around the circuit in question, and their behaviour should reflect the genetic relationships the best. Indeed, graphing these genes shows beautiful agreement with predictions (see www.wormlabcaltech.github.io/mprsq for interactive graphics).

In order to identify affected biological processes, we performed an in-house gene ontology enrichment analysis using annotations provided by WormBase, followng the procedure shown in TEA [27]. Top enriched terms included 'hydrolase activity' (869 observed hits; 7.8 fold change; p-value $< 10^{-10}$); 'organic anion transport' (803 hits; 7.5; p-value $< 10^{-10}$); 'spliceosomal complex' (647 hits; 8.2 p-value $< 10^{-10}$); 'SAM-depdendent methyltransferase activity' (1215 hits; 6.6; p-value $< 10^{-10}$); and 'cell division' (1251 hits; 7.9; p-value $< 10^{-10}$). In mammals, the mammalian target of rapamycin pathway, which is intimately associated with the hypoxia pathway, has been previously linked to osmotic stress responses [28]. Our findings also suggest that the *hif-1*-dependent response causes important changes in chromatin structure via activation or recruitment of chromatin remodelling factors.

Next, we attempted to identify direct targets of the genes we studied. *vhl-1* targets were particularly easy to isolate: because *vhl-1* has a direct role in protein degradation, direction of change is known; and because *vhl-1* does not seem to participate in the *rhy-1*, *egl-9*, *hif-1* feedback circuit, it is easy to isolate targets for that are *hif-1*-independent. We found 25 genes that are putative candidates for *vhl-1*-targeted degradation. These 25 genes include *pole-1*, an ortholog of human polymerase ϵ catalytic subunit; *F33H2.6*, an ortholog of the human regulator of microtubule dynamics 1 (RMDN1); and many solute carriers. Reflecting this, enriched GO terms were 'ion binding', 'growth', 'cell division', 'cell projection assembly' as well as 'ion binding' and 'divalent metal ion transport'. *vhl-1* has been previously implicated as a controller of mitotic fidelity in renal cell carcinoma [29]. Our findings support a role of *vhl-1* in chromosomal integrity and mitotic fidelity. Furthermore, recent reports suggest that solute carriers may be associated with poor prognosis in clear-cell renal carcinoma [30], which highlights the biological relevance of our predictions.

We identified 133 genes that are activated by HIF-1. We verified that the genes we identified are actually *hif-1* targets by searching for a set of 20 gold-standard genes from the literature [] in our gene-set (see SI), and found that *hif-1* targets were significantly enriched in these genes ($p < 10^{-7}$). GO term enrichment indicated that this list was associated with 'cell division', 'SAM methyltransferase activity' and 'cellular modified amino acid metabolic processes'. A full list of *hif-1* targets can be found in S.I..

Discussion

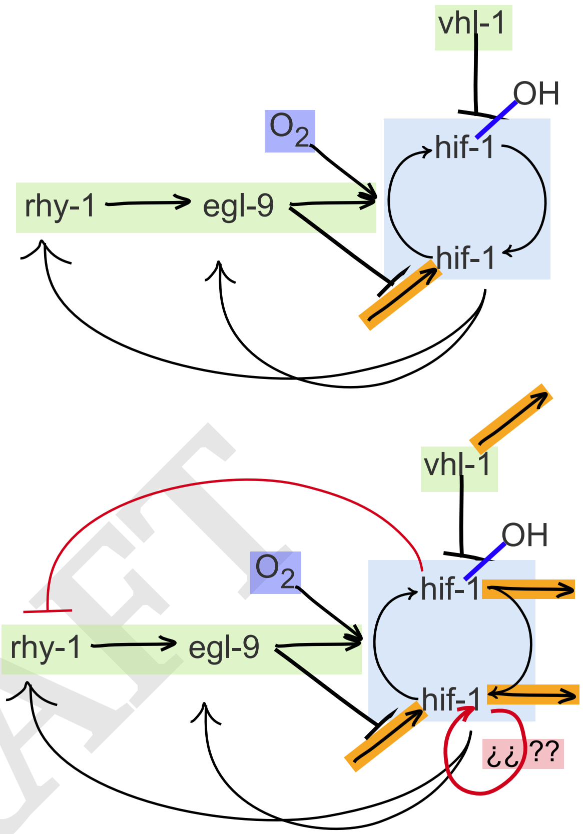

Fig. 7. Top: Previous model. Bottom: Current model derived from RNA-seq data.

Previous work has established a circuit in which *rhy-1* activates *egl-9* in a linear pathway, and *egl-9* inhibits *hif-1* in an oxygen-dependent manner. Hydroxylated HIF-1 can then be degraded in a *vhl-1*-dependent manner. There is also evidence that *egl-9* and *rhy-1* are in turn activated by *hif-1* [20, 31]. Finally, there is evidence that although the interaction between *egl-9* and *vhl-1* is important for *hif-1* repression, *egl-9* can also act in a non-*vhl-1* dependent manner (see Fig. 7 top).

From our data, we were able to impute the positive regulatory relationship between *egl-9* and *rhy-1*. We would not have been able to infer the order of the regulation without additional information. Using clustering as a proxy for phenotype, we were able to infer the relationship between *egl-9* and *hif-1*. We were also able to infer a positive (linear or AND) relationship between *egl-9* and *vhl-1* using clustering. Alternatively, we gained the same information by performing *in silico* qPCR on the genes under study. *In silico* qPCR also revealed that *hif-1* has two states with different activities: Non-hydroxylated HIF-1 increases levels of *rhy-1*, and hydroxylated HIF-1 inhibits *rhy-1* and possibly *egl-9* as well, although the double mutant did not recapitulate that interaction. We also revealed that *hif-1* is an autoregulator.

These discoveries are consistent with a homeostatic circuit. By autoregulating itself, *hif-1* can mantain appropriate protein levels both in normoxic and hypoxic conditions. Inhibition of *rhy-1*, and possibly of *egl-9*, ensures that an appropriate equilibrium is maintained between hydroxylated and non-hydroxylated protein, which may have functional consequences for the cell if both forms are active.

In addition to these biological findings, our dataset allows us to generate predictions of genes that may be under direct *hif-1* regulation. Assuming that non-hydroxylated HIF-1 has different activities from hydroxylated HIF-1, we identified 5 genes that are candidates for activation by hydroxylated HIF-1. These genes have been implicated in the *C. elegans* immune response, or have behavioural phenotypes, underscoring the importance of *hif-1* in neurobiology and immunology [32–35]. We have shown that transcriptomes contain sufficient information to be used as semi-quantitative phenotypes in metazoans. These phenotypes can be interpreted globally via correlation tests, clustering or other probabilistic methods; alternatively, they can be used to query single reporter genes in a manner similar to qPCR today. Transcriptomic phenotypes have distinct advantages over physical traits. Firstly, due to their increased complexity, the genotype-phenotype mapping degeneracy ought to be greatly reduced, which facilitates predictions of genetic interaction. Secondly, genes that result in subtle or no visible traits when mutated may have strong (detectable), reproducible phenotypes at the transcriptomic level, which would make the study of small-effect genes significantly easier.

RNA-seq and microarray datasets have been used previously by bioinformaticians to generate high-throughput predictions of genetic interactions and consortiums such as the The Cancer Genome Atlas have sequenced RNA from many different cancers in the hope of identifying clinically or biologically relevant interactions []. By correlating many different datasets in many different conditions, it is possible in theory to predict genetic interaction. Our approach differs from these high-throughput methods in that we are not attempting to generate large scale networks. Rather, the strength in our analysis derives from our experimental design, which allows us to ask and answer a large number of questions about the functional interactions between genes. As a by-product, we are also able to identify genes related to the core circuit studied in question, but our main goal is not to generate databases or predict large numbers of interactions between a large number of genes. We have shown that transcriptomic phenotypes can capture distinct interaction modes in a single experiment, making it possible to infer complex regulatory relationships between genes. By measuring these transcriptomes under a rigorous experimental design, it is possible to identify many relationships simultaneously. With the advent of fast pseudo-alignment tools and ever cheaper sequencing techniques, biologists should consider using global transcriptomes as a tool beyond hypothesis generation or target acquisition.

Materials and Methods

RNA-seq. Tagmentation etc

We used Kallisto to perform pseudo-read 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. 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 [27, 36–43]. Our analysis is available in a Jupyter Notebook[44]. 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 article was written with support of the Howard Hughes Medical Institute.

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. TL - 37. *Nature genetics* 37 VN - r(5):471–477.
4. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621–628.
5. Metzker ML (2010) Sequencing technologies - the next generation. *Nature reviews. Genetics* 11(1):31–46.
6. Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature biotechnology* 32(5):462–464.
7. Bray NL, Pimentel H, Melsted P, Pachter L (2015) Near-optimal RNA-Seq quantification. *aRxiv*.
8. Patro R, Duggal G, Kingsford C (2015) Salmon: Accurate, Versatile and Ultrafast Quantification from RNA-seq Data using Lightweight-Alignment. *bioRxiv* p. 021592.
9. Pimentel HJ, Bray N, 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. Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. *Cell* 148(3):399–408.
17. Nizet V, Johnson RS (2009) Interdependence of hypoxic and innate immune responses. *Nature reviews. Immunology* 9(9):609–17.
18. 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).
19. Semenza GL (2003) Targeting HIF-1 for cancer therapy. *Nature reviews. Cancer* 3(10):721–32.
20. Bishop T et al. (2004) Genetic analysis of pathways regulated by the von Hippel-Lindau tumor suppressor in Caenorhabditis elegans. *PLoS Biology* 2(10).
21. 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.
22. Tanimoto K, Makino Y, Pereira T, Poellinger L (2000) Mechanism of regulation of the hypoxia-inducible factor-1 alpha by the von Hippel-Lindau tumor suppressor protein. *Embo J* 19(16):4298–309.
23. Jaakkola P et al. (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* 292(5516):468–472.
24. 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.
25. 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.
26. Yeung KY, Medvedovic M, Bumgarner RE (2003) Clustering gene-expression data with repeated measurements. *Genome biology* 4(5):R34.
27. Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW (2016) Tissue enrichment analysis for C. elegans genomics. *BMC Bioinformatics* 17(1):366.

28. Zhou B et al. (2007) Hypertonic induction of aquaporin-5: novel role of hypoxia-inducible factor-1alpha. *Am J Physiol Cell Physiol* 292(4):C1280–90.
29. Hell MP, Duda M, Weber TC, Moch H, Krek W (2014) Tumor suppressor vhl functions in the control of mitotic fidelity. *Cancer Research* 74(9):2422–2431.
30. Liu Y et al. (2015) High expression of Solute Carrier Family 1, member 5 (SLC1A5) is associated with poor prognosis in clear-cell renal cell carcinoma. *Scientific reports* 5(October):16954.
31. Powell-Coffman JA (2010) Hypoxia signaling and resistance in C. elegans. *Trends in Endocrinology and Metabolism* 21(7):435–440.
32. Gray JM et al. (2004) Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. *Nature* 430(6997):317–322.
33. Cheung BHH, Cohen M, Rogers C, Albayram O, De Bono M (2005) Experience-dependent modulation of C. elegans behavior by ambient oxygen. *Current Biology* 15(10):905–917.
34. Chang AJ, Bargmann CI (2008) Hypoxia and the HIF-1 transcriptional pathway reorganize a neuronal circuit for oxygen-dependent behavior in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* 105(20):7321–7326.
35. Ma DK et al. (2013) Cytochrome P450 drives a HIF-regulated behavioral response to reoxygenation by C. elegans. *Science (New York, N.Y.)* 341(6145):554–8.
36. Team BD (2014) Bokeh: Python library for interactive visualization.
37. McKinney W (2011) pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* pp. 1–9.
38. Oliphant TE (2007) SciPy: Open source scientific tools for Python. *Computing in Science and Engineering* 9:10–20.
39. Pedregosa F et al. (2012) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.
40. Salvatier J, Wiecki T, Fonnesbeck C (2015) Probabilistic Programming in Python using PyMC. *Arxiv* pp. 1–24.
41. 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.
42. Developers M (year?) matplotlib: v1.5.3.
43. Waskom M et al. (year?) seaborn: v0.7.0 (January 2016).
44. 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.