nature genetics

Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity

Esti Yeger-Lotem[1,2,8], Laura Riva[1,8], Linhui Julie Su[2], Aaron D Gitler[2,7], Anil G Cashikar[2,7], Oliver D King[2,7], Pavan K Auluck[2,3], Melissa L Geddie[2], Julie S Valastyan[2,4], David R Karger[5], Susan Lindquist[2,6] & Ernest Fraenkel[1,5]

Cells respond to stimuli by changes in various processes, including signaling pathways and gene expression. Efforts to identify components of these responses increasingly depend on mRNA profiling and genetic library screens. By comparing the results of these two assays across various stimuli, we found that genetic screens tend to identify response regulators, whereas mRNA profiling frequently detects metabolic responses. We developed an integrative approach that bridges the gap between these data using known molecular interactions, thus highlighting major response pathways. We used this approach to reveal cellular pathways responding to the toxicity of alpha-synuclein, a protein implicated in several neurodegenerative disorders including Parkinson's disease. For this we screened an established yeast model to identify genes that when overexpressed alter alpha-synuclein toxicity. Bridging these data and data from mRNA profiling provided functional explanations for many of these genes and identified previously unknown relations between alpha-synuclein toxicity and basic cellular pathways.

The cellular response to perturbations including environmental changes, toxins and mutations is typically complex and comprises signaling and metabolic changes, as well as changes in gene expression. Revealing the molecular mechanisms underlying cellular response to a specific perturbation may determine the nature of the perturbation, thus illuminating disease mechanisms[1] or a drug's mode of action[2,3], and identify points of intervention with potential therapeutic value[4].

High-throughput experimental techniques are commonly used for finding components of these response pathways because they provide a genome- and proteome-wide view of molecular changes. mRNA profiling experiments rapidly identify genes that are differentially expressed following stimuli. Genetic screening, including deletion, overexpression and RNAi library screens, identify genetic 'hits', genes whose individual manipulation alters the phenotype of stimulated cells. However, each technique has obvious limitations for identifying the full nature of cellular responses. mRNA profiling experiments do not target the series of events that led to the differential expression. Genetic screens provide strong evidence that a gene is functionally related to the response process, but this relationship is often indirect and hard to decipher, especially in high-throughput experiments that typically result in scores of relevant genes with various functions.

It has been noted previously in a few specific instances[2,5–9] that genetic screens do not identify the same genes as mRNA assays conducted in the same conditions. Here we show that this discrepancy is, in fact, a general rule. Furthermore, we find a marked bias in each technique. We bridge this gap between the two forms of high-throughput data by using an algorithm that exploits molecular interactions data to reveal the functional context of genetic hits and additional proteins that participate in the response but that were not detected by either the genetic or the mRNA profiling assays themselves.

We applied the algorithm to identify cellular responses to increased expression of alpha-synuclein, a small human protein implicated in Parkinson's disease whose native function and role in the etiology of the disease remain unclear[10]. We screened an established yeast model for alpha-synuclein toxicity[11,12] using an additional set of 3,500 overexpression yeast strains, exposing the multifaceted toxicity of alpha-synuclein. Application of our approach to the genetic hits from the screen and to transcriptional data of the yeast model provides the first cellular map of the proteins and genes responding to alpha-synuclein expression.

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. [3]Departments of Pathology and Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, and Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [6]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [7]Present addresses: Department of Cell and Developmental Biology, The University of Pennsylvania, Philadelphia, Pennsylvania, USA (A.D.G.), Medical College of Georgia, Augusta, Georgia, USA (A.G.C.) and Boston Biomedical Research Institute, Watertown, Massachusetts, USA (O.D.K.). [8]These authors contributed equally to this work. Correspondence should be addressed to S.L. (lindquist_admin@wi.mit.edu) or E.F. (fraenkel-admin@mit.edu).

Received 7 August 2008; accepted 27 January 2009; published online 22 February 2009; doi:10.1038/ng.337

Table 1 Measured responses to cellular perturbations

Perturbation[a]	Number of differentially expressed genes[b]	Number of genetic hits[c]	Overlap	*P* value
Growth arrest (HU)	59	86	0	1
DNA damage (MMS)	198	1,448	43	0.81
ER stress (tunicamycin)	200	127	5	0.42
Fatty acid metabolism (oleate)	269	103	9	0.041
ATP synthesis block (arsenic)	828	50	9	0.25
Protein biosynthesis (cycloheximide)	20	164	0	1
Gene inactivation, screen complete (24 data sets)[d]	27	130	0	1
Gene inactivation, screen incomplete (149 data sets)[d]	24	12	0	1

[a]See **Supplementary Table 1a** for data sources. [b]Differentially expressed genes were defined as those showing at least a twofold change in expression following the perturbation or as defined in the original papers. [c]Number of genes whose genetic manipulation affects the phenotype of perturbed cells as defined in the original papers. [d]Median results are shown.

RESULTS

Comparing genetic hits and differentially expressed genes

We analyzed published mRNA profiles and genetic hits for 179 distinct perturbations in yeast (Methods). The perturbations included chemical and genetic insults affecting a multitude of cellular processes. Thirty of the genetic screens are complete, typically identifying >100 genetic hits. In almost all cases the overlap was small and statistically insignificant (**Table 1** and **Supplementary Table 1a** online).

We used Gene Ontology (GO) enrichment analysis to check whether each assay may be biased toward distinct aspects of cellular responses (**Supplementary Table 1b** and **Supplementary Fig. 1a** online). The combined genetic hits from all 179 genetic screens were highly enriched for several annotations, among the most frequent of which were biological regulation (23.3%, $P < 10^{-82}$), including transcription (14%, $P < 10^{-44}$) and signal transduction (6.3%, $P < 10^{-31}$). In contrast, the differentially expressed genes from all perturbations were enriched mostly for various metabolic processes (for example, organic acid metabolic process 7.1%, $P < 10^{-18}$) and oxidoreductase activities (7.2%, $P < 10^{-34}$). We observed the same enrichment trends upon focusing only on the 30 perturbations for which complete data were available when analyzed individually or when combined (**Supplementary Tables 1 c,d** and **Supplementary Note** online). Thus, we find that genetic assays tend to probe the regulation of cellular responses, whereas mRNA profiling assays tend to probe the metabolic aspects of cellular responses.

The differences in annotation between genetic hits and differentially expressed genes imply that each gene set alone often provides a limited and biased view of cellular responses. This hypothesis was confirmed in pathways that were well-studied by more classical methods. In the yeast DNA-damage response pathway, for example, a genetic screen[4] detected proteins that sense DNA damage (Mec3, Ddc1, Rad17 and Rad24), whereas mRNA profiling detected repair enzymes such as Rnr4 (ref. 13). Yet core components that had been uncovered by intense investigations over many years, such as the signal transducers Mec1 and Rad53 and the transcription factor Rfx1, remained undetected by either high-throughput assay.

To fully reap the benefits of applying high-throughput methods to new problems and underexplored biological processes, it is essential to find new routes to connect these data and obtain a true picture of the regulation of cellular responses. Judging from characterized pathways such as the DNA-damage response discussed above, we expect that some of the genetic hits, which are enriched for response regulators, will be connected via regulatory pathways to the differentially expressed genes, which are the output of such pathways, via components of the response that are missing from the experimental data (**Fig. 1**).

ResponseNet algorithm for identifying response networks

We devised the ResponseNet algorithm to identify molecular interaction paths connecting genetic hits and differentially expressed genes, including components of the response that are otherwise hidden (**Fig. 1**). The yeast *Saccharomyces cerevisiae* provides a powerful model system for such analysis owing to the extensive molecular interactions data now available (Methods and **Supplementary Table 2a** online). We assembled an integrated network model of the yeast interactome that contains protein–protein interactions, metabolic relations and protein–DNA interactions detected by various methods with different levels of reliability[14]. The resulting interactome relates 5,622 interacting proteins and 5,510 regulated genes, which are represented by network nodes, via 57,955 molecular interactions, which are represented by network edges.

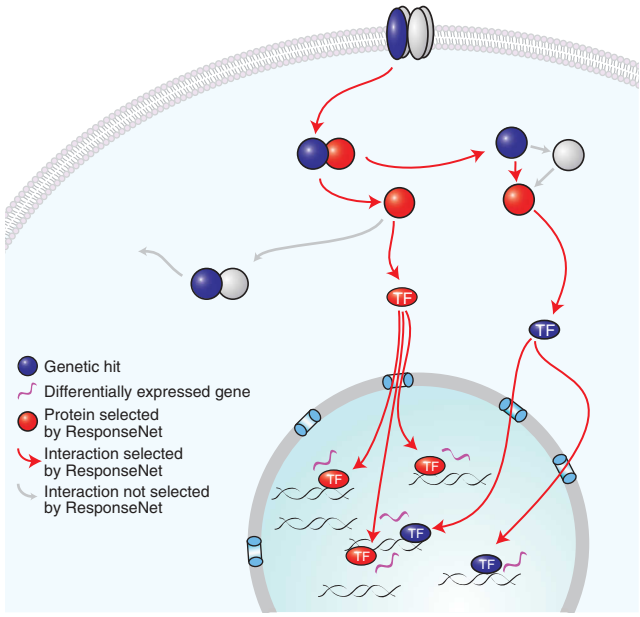

Figure 1 Regulatory relationships between genetic and transcriptional data. Cellular response is depicted through a general signaling pathway, including receptor binding, transcription factor (TF) translocation into the nucleus and gene expression. Genetic screens and mRNA profiling identify only some of these molecular components and often do not identify the same genes, as shown. We find that the proteins products of genes identified in genetic screens (colored blue) tend to be molecules with regulatory roles. We therefore hypothesize that they may directly or indirectly contribute to the regulation of the observed change in gene expression (colored magenta). ResponseNet identifies the likely regulatory pathways and predicts proteins that are part of these pathways even if they are not identified in either screen (colored red).

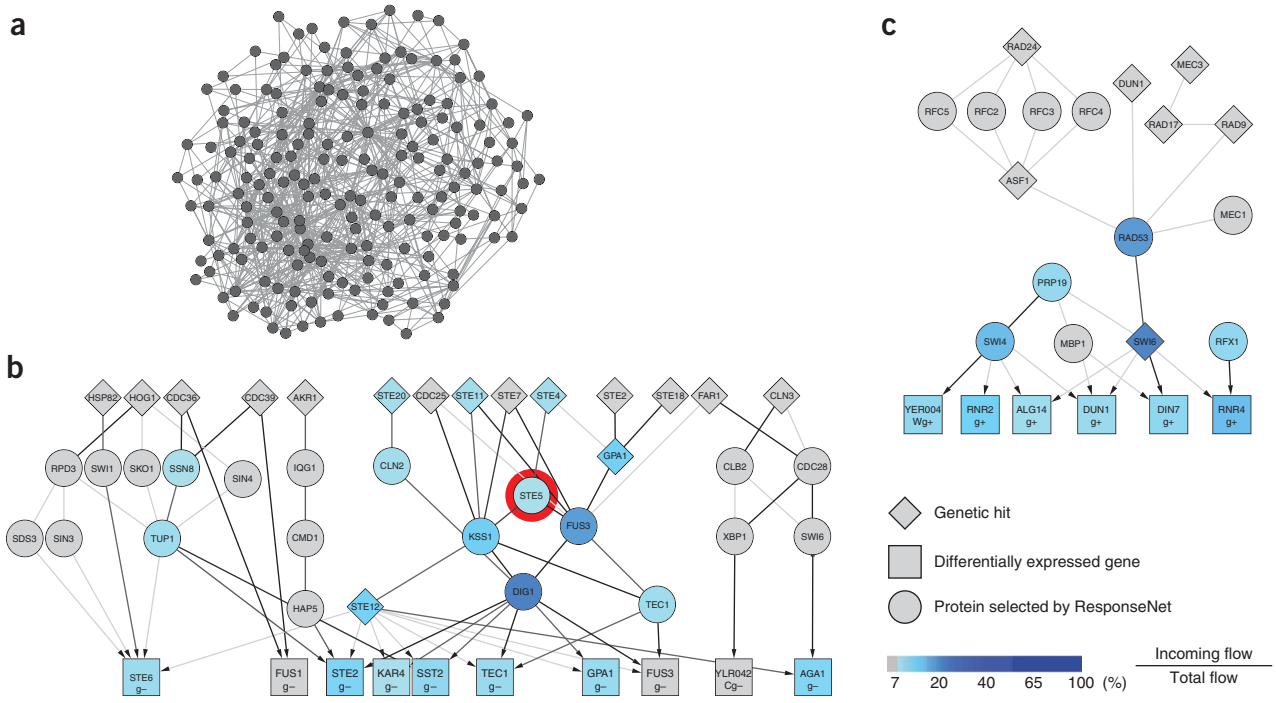

Figure 2 Interactome subnetworks connecting genetic and transcriptional data. (**a**) A network connecting genetic and transcriptional[19] data of *STE5* deletion strain via paths with length of three edges or fewer finds 193 nodes and 778 edges. (**b**) The network created by ResponseNet connects the genetic and transcriptional[19] data of *STE5* deletion strain via 23 intermediary nodes and 96 edges. Higher ranked nodes, as determined by ResponseNet, appear in darker shades of blue and include core components of the pheromone response pathway. Ste5 itself, marked by a red circle, is ranked ninth among the top predicted proteins. (**c**) The highly ranked part of the network created by ResponseNet upon connecting genetic hits[4,20] to DNA-damage signature genes[21] identified in yeast treated with the DNA-damaging agent methyl methanesulfonate (MMS). The highest ranking intermediate nodes predicted by ResponseNet include core components of the DNA-damage–response pathway. The complete network appears in **Supplementary Figure 4** online. Each node represents either a protein or a gene, and edges represent protein–protein, metabolic and protein–DNA interactions. The darkness of an edge increases with the amount of flow it carries. Differentially expressed genes are labeled with a suffix of g+ for upregulation and g– for downregulation. Networks were visualized using Cytoscape.

Our interactome representation has two important features that facilitate identification of pathways relating genetic hits to transcriptional changes. First, we highlighted the transcriptional regulatory role of proteins by representing differentially expressed genes and their protein products as separate gene and protein nodes, respectively. The only connection between protein and gene nodes is through edges representing observed protein–DNA interactions between transcriptional regulators and their target genes. Edges between two protein nodes represent other interaction types. Consequently, pathways connecting genetic hits to differentially expressed genes must pass through transcriptional regulators (**Supplementary Fig. 1b**). Second, because interactions vary in their reliability, each edge was given a weight that represents the probability that the connected nodes interact in a response pathway. Probabilities were computed using a Bayesian method that considers the experimental evidence supporting an interaction, and that favors interactions among proteins acting in a common cellular response pathway (Methods and **Supplementary Table 2b**).

Because of the vast number of edges, a search for all interaction paths connecting the genetic hits to the differentially expressed genes typically results in 'hairball' networks that are very hard to interpret (**Fig. 2a**). Pioneering approaches that searched an interactome for high-probability paths had to limit the output path lengths to three edges for computational complexity issues[15,16]. We aimed for a solution that would (i) pick the subset of genetic hits most likely to modulate the differentially expressed genes without limiting it a priori to known regulatory genes, (ii) identify and rank intermediary proteins that are likely to be part of response pathways but escaped detection by high-throughput methods and (iii) give preference to proteins that lie on high-probability paths connecting the genetic hits to the differentially expressed genes without imposing constraints on the network topology.

These requirements were met with a 'flow algorithm', a computational method used previously to analyze known signaling or metabolic pathways (for example, see ref. 17). Basically, flow goes from a source node to a sink node through the graph edges; edges are associated with a capacity that limits the flow and with a cost. (As a loose analogy, this resembles water finding the path of least resistance through a complex landscape.) To identify response pathways we required that flow pass from genetic hits through interactome edges to differentially expressed genes (**Supplementary Fig. 1b**). We then formulated our goal as a minimum-cost flow optimization problem[18]: Cost was defined as the negative log of the probability of an edge. Hence, minimizing the cost gives preference to high-probability paths (Methods).

The solution to the optimization problem is a relatively sparse network connecting many of the genetic hits to many of the differentially expressed genes through known interactions and intermediary proteins (**Fig. 2b**). Although these intermediary proteins escaped detection by either high-throughput genetic analysis or mRNA profiling, they are predicted by the algorithm to participate in the response. All proteins in the solution are ranked by the amount of flow they

Table 2 Yeast genes that modify α-syn toxicity when overexpressed

Gene class	α-syn toxicity suppressors	α-syn toxicity enhancers
Amino acid transport	*Avt4, Dip5, Lst8*	
Autophagy	*Nvj1*	
Cytoskeleton	*Icy1, Icy2*	
Manganese transport	*Ccc1*	*Pmr1*
Protein phosphorylation	*Cdc5, Gip2, Ime2, Ptp2, Ptc4, Rck1, Yck3*	*Cax4, Ppz1, Ppz2, Sit4*
Transcription or translation	*Cup9, Fzf1, Hap4, Jsn1, Mga2, Stb3, Tif4632, Vhr1*	*MATALPHA1, Mks1, Sut2*
Trehalose biosynthesis	*Nth1, Tps3, Ugp1*	
Ubiquitin-related	*Cdc4, Hrd1, Uip5*	*Ubp7, Ubp11*
Vesicular transport, ER-Golgi	*Bre5, Erv29, Sec21, Sec28, Sft1, Ubp3, Ykt6, Ypt1*	*Bet4, Glo3, Gos1, Gyp8, Sec31, Sly41, Trs120, Yip3*
Other cellular processes	*Isn1, Mum2, Osh2, Osh3, Pde2, Pho80, Pfs1, Qdr3*	*Eps1, Ids2, Izh3, Tpo4*
Unknown function	*YBR030W, YDL121C, YDR374C, YKL063C, YKL088W, YML081W, YML083C, YMR111C, YNR014W, YOR129C, YOR291W (Ypk9)*	

carry. The more flow that passes through a protein, the more important it is in connecting the input sets.

Validation of the ResponseNet algorithm

To determine whether ResponseNet provides valid biological insights, we used it to analyze data from perturbations of well-studied pathways. For example, we used ResponseNet to connect genetic hits associated with Ste5 (from the *Saccharomyces* Genome Database) and differentially expressed genes[19] collected from a strain lacking Ste5, a scaffold protein that coordinates the MAP kinase cascade activated by pheromone (**Fig. 2b**). Nodes selected by ResponseNet were highly enriched for proteins functioning in the pheromone response pathway (46%, $P < 10^{-18}$), thus revealing the perturbed biological process. The highly ranked intermediary proteins included key regulators of the pheromone response including Ste5, the source of perturbation.

ResponseNet also performed well in analyzing the complex cellular response to DNA damage[4,20,21]. Nodes discovered by ResponseNet were highly enriched for the GO categories response to DNA damage stimulus (21%, $P < 10^{-14}$) and DNA repair (19%, $P < 10^{-14}$). The highly ranked part of the network contained core pathway proteins that were uncovered by years of intense investigation but escaped detection by high-throughput screens, including signal transducers (Mec1, Rad53), members of the RFC complex (Rfc2, Rfc3, Rfc4, Rfc5) and the transcriptional regulator Rfx1 (**Fig. 2c**). Statistical evaluation of the performance of ResponseNet on data for less well-characterized pathways is described in the **Supplementary Note**.

Mapping the cellular responses to alpha-synuclein toxicity

Having established the validity of our method to uncover connections between otherwise disparate high-throughput datasets, we applied ResponseNet to investigate the cellular toxicity associated with alpha-synuclein (α-syn). α-Syn is a small lipid-binding protein that is natively unfolded when not bound to lipids and prone to forming toxic oligomers[22]. It has been implicated in several neurodegenerative disorders, particularly Parkinson's disease (PD): it is the main component of Lewy bodies, locus duplication or triplication of α-syn lead to familial forms of PD, and increased expression of α-syn leads to neurodegeneration in several animal models[23]. Despite immense efforts, the cellular pathways by which α-syn leads to cell death are just beginning to emerge.

The yeast *Saccharomyces cerevisiae* provides a powerful system for studying the toxicities of α-syn that result from its intrinsic physical properties. Expression of human α-syn in yeast yields dosage-dependent defects also found in mammalian systems, including cytosolic-lipid-droplet accumulation, reactive-oxygen-species production and ubiquitin-proteasome system impairment[11]. An initial screen for yeast genes that modify α-syn toxicity when overexpressed identified genes involved in ER-to-Golgi vesicle trafficking and led to the observation that α-syn blocks ER-to-Golgi vesicle trafficking[12].

We now report the results of screening 5,500 overexpression yeast strains, thereby covering 85% of the yeast proteome. We identified 55 suppressors and 22 enhancers of α-syn toxicity, many with clear human orthologs, including the homolog of human PD gene *ATP13A2* (also known as *PARK9*; **Table 2** and **Supplementary Table 3a** online). As demonstrated in the accompanying article (Gitler *et al.*[24]), PARK9 and the human homologs of eight other genetic modifiers with diverse functions (Ypt1, Hrd1, Ubp3, Pde2, Cdc5, Yck3, Sit4 and Pmr1) are efficacious in neuronal models, validating the yeast model as meaningful to α-syn toxicity in neurons[12,24]. Major classes of genes that emerged include vesicle-trafficking genes, kinases and phosphatases, ubiquitin-related proteins, transcriptional regulators, manganese transporters and trehalose-biosynthesis genes (**Supplementary Table 3a,b**). Notably, trehalose was recently shown to promote the clearance of misfolded mutant α-syn[25], and manganese exposure has been linked with Parkinson's-like symptoms, albeit with a distinct underlying pathology[26]. The genes identified by the screen point to causal relations between α-syn expression and toxicities previously associated with PD but not specifically linked to α-syn (**Supplementary Note**).

mRNA profiling of the yeast model was determined in a separate study (unpublished data and **Supplementary Table 3b,c**). Upregulated genes prominently included genes with oxidoreductase activities (13%, $P < 10^{-9}$). Downregulated genes included ribosomal genes (28%, $P < 10^{-30}$), as commonly observed under stress[27]. More specific to α-syn toxicity, the downregulated genes were markedly enriched for genes encoding proteins localized to the mitochondria (60%, $P < 10^{-44}$).

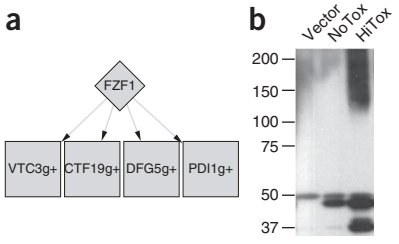

Figure 3 Nitrosative stress response to α-syn expression in yeast. (**a**) The predicted subnetwork containing Fzf1 and its differentially expressed target genes. Graphical representation is similar to **Figure 2**. (**b**) Immunoblotting against S-nitrosocysteine performed on a control strain (vector), on a strain expressing one copy of α-syn (NoTox) and on a high-toxicity strain (HiTox) expressing several copies of α-syn reveals that increasing levels of α-syn increase the amount of S-nitrosylated proteins.

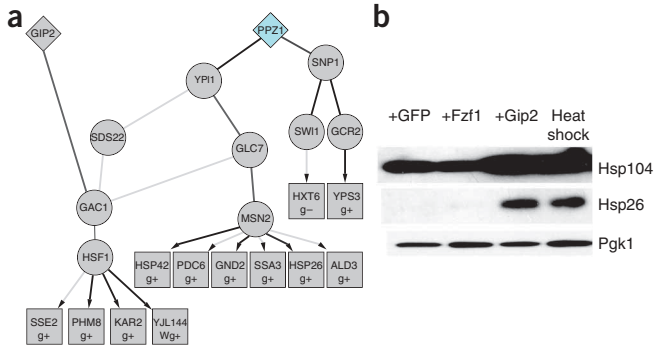

Figure 4 Overexpression of Gip2 causes induced expression of Hsf1 targets. (**a**) The predicted subnetwork links the toxicity suppressor Gip2 and the toxicity enhancer Ppz1 to Hsf1 and Msn2 via components of type 1 protein phosphatase complex (Gac1, Glc7, Ypi1, Sds22). Graphical representation is similar to **Figure 2**. (**b**) Immunoblotting of vector cells overexpressing GFP, Fzf1 or Gip2 with antibodies against Hsp104 and Hsp26. Overexpression of Gip2 is sufficient to activate Hsf1 and induce higher protein levels of both its targets Hsp104 and Hsp26, similar to that of vector cells subjected to heat shock. In contrast, overexpression of another genetic suppressor, Fzf1, does not activate Hsf1. Immunoblotting against Pgk1 was used as a loading control.

The genetic and mRNA profiling data exemplify both the power and the limitations of the current approaches. Although they reveal the wide range of cellular functions altered by α-syn, the precise roles of the identified genes in the cellular response are unclear. For example, we checked whether the ubiquitin-related genetic hits affect α-syn degradation. However, in strains overexpressing these ubiquitin-related genes, we did not detect changes in steady-state α-syn protein concentrations (**Supplementary Fig. 2** online). As with our analyses above, the overlap between the genetic hits and the differentially expressed genes was minor (four genes, $P = 0.96$).

Application of ResponseNet to these disparate datasets gave a more coherent view of the cellular response (**Supplementary Fig. 3a** online). The resulting network provided context to a large portion of the data: 34 (44%) genetic hits and 166 (27%) differentially expressed genes were linked to each other through 106 intermediary proteins. These include two-thirds of the protein kinase, phosphatase and ubiquitin-related genetic hits, illuminating their intricate role in the response to α-syn.

The major cellular pathways identified by ResponseNet included ubiquitin-dependent protein degradation, cell cycle regulation and vesicle-trafficking pathways, all of which have previously been associated with PD (**Supplementary Note** and **Supplementary Fig. 3a**). Four examples illustrate the ability of ResponseNet to clarify aspects of α-syn responses relevant to PD and uncover others whose relationship to α-syn was completely unknown.

Nitrosative stress

Fzf1 was the only genetic hit related to nitrosative stress[28]. However, ResponseNet connected it to four upregulated transcripts, including that encoding Pdi1, a protein disulfide isomerase (PDI) (**Fig. 3a**). Notably, the upregulation of human PDI protects neuronal cells from neurotoxicity associated with ER stress and protein misfolding (both of which are linked to α-syn expression in yeast and neurons), and PDI is one of a small number of specific proteins S-nitrosylated in PD that activate protective pathways, in addition to the generalized nitrosative damage that is a hallmark of the disease[29]. We found that increased expression of α-syn causes both specific and general increases in S-nitrosylation of proteins (**Fig. 3b**). This was highly surprising because the yeast genome does not encode a canonical nitric oxide synthase and, until very recently, yeast were not thought to produce nitric oxide[30]. Our results indicate that the nitrosylation of specific proteins and generalized nitrosylation is a highly conserved and deeply rooted response to cellular perturbations created by α-syn.

Heat shock

The induction of the heat-shock response directly or via chemical inhibition of Hsp90 (ref. 31) suppresses α-syn toxicity in many model systems including yeast, flies, mice and human cells (for example, see refs. 32,33). However, heat-shock–related genes were conspicuously absent among the list of genetic suppressors. Nonetheless, ResponseNet predicted the involvement of two highly conserved heat-shock regulators, the chaperone Hsp90 (isoform Hsp82, **Supplementary Fig. 3a**, panel a) and the heat-shock transcription factor Hsf1 (**Fig. 4a**). Hsf1 appeared downstream of the toxicity suppressor Gip2, a putative regulatory subunit of the Glc7 phosphatase, which interacts with Gac1. Gac1 is a regulatory subunit of the Glc7 complex that is known to activate Hsf1 (ref. 34). These connections suggested that Gip2 overexpression might induce a heat-shock response. Indeed, we found that strains overexpressing Gip2 show elevated concentrations of heat-shock proteins (**Fig. 4b**). ResponseNet therefore provided a mechanistic explanation for the suppression of α-syn toxicity achieved by Gip2 overexpression and identified a new regulator of the highly conserved heat-shock response.

The mevalonate-ergosterol biosynthesis pathway

This pathway, which is targeted by the cholesterol-lowering statin drugs, synthesizes sterols as well as other products with connections to α-syn toxicity, such as farnesyl groups required for vesicle trafficking proteins and ubiquinone required for mitochondrial respiration. ResponseNet ranked highly Hrd1, which regulates the protein target of statins, and the predicted intermediary Hap1, a proposed transcriptional regulator of the pathway[35] (**Supplementary Fig. 3a**, panel a). In addition, the α-syn mRNA profile modestly correlated with the profile of yeast treated with lovastatin ($r = 0.32$, $P < 10^{-93}$, L.J.S. and S.L., unpublished data), and several genetic hits also could be associated with products of the pathway (enzymes Bet4 and Cax4, farnesylated proteins Ypt1 and Ykt6 and putative sterol carriers Sut2, Osh2 and Osh3). We therefore tested the effect of lovastatin, which selectively inhibits the highly conserved HMG-CoA reductase protein in yeast and in mammalian cells, on α-syn toxicity. Addition of 5 μM lovastatin to the media caused a further reduction in growth to strains overexpressing α-syn (**Fig. 5a**), but did not reduce growth of either wild-type controls or of cells expressing another toxic protein, a glutamine-expansion variant of huntingtin exon I[36] (**Supplementary Fig. 3b**). We further tested ubiquinone, a downstream output of this pathway, reasoning that its downregulation through the action of α-syn might increase cellular vulnerability. Indeed, the addition of ubiquinone-2 to the media provided a modest suppression against α-syn toxicity. Ubiquinone is an antioxidant, but this was not a nonspecific antioxidant response, as the antioxidant N-acetylcysteine had no effect (data not shown).

The target of rapamycin (TOR) pathway

ResponseNet identified the TOR pathway proteins Tor1, Tor2 and their target transcription factors as intermediary between the genetic suppressor Lst8, a positive regulator of the TOR pathway, and several upregulated genes involved in spore wall formation (a vectorially directed secretory process in yeast) and vacuolar protein degradation

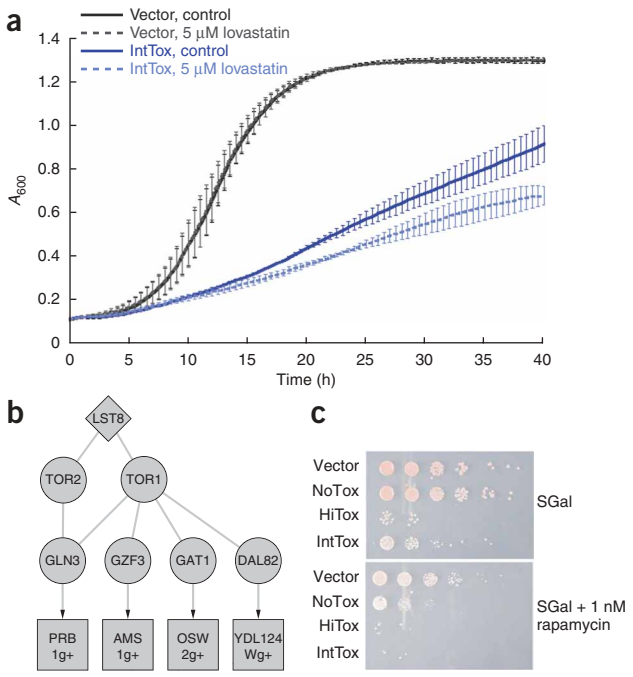

Figure 5 Effects of the small molecules lovastatin and rapamycin on α-syn toxicity. (**a**) Lovastatin inhibits growth of the yeast strain expressing an intermediate level of α-syn. Growth of a control strain (vector) and an intermediate toxicity strain (IntTox) expressing several copies of α-syn was measured in a galactose containing media with and without 5 μM lovastatin. Each growth curve reflects the average of three individual runs, each of which is indicated by a bar. (**b**) The predicted subnetwork containing TOR pathway components includes the predicted proteins Tor1 and Tor2. Graphical representation is similar to **Figure 2c**. (**c**) The effect of rapamycin on growth of different yeast strains. The upper panel shows the growth of a control strain (vector), a strain expressing one copy of α-syn (NoTox), a high-toxicity strain (HiTox) and an intermediate toxicity strain (IntTox) both expressing several copies of α-syn, in a galactose containing media (SGal) that is used to induce expression of α-syn. The lower panel shows the same strains grown in media that also contains 1 nM rapamycin, showing that rapamycin inhibits growth of all α-syn–expressing strains but not the control strain, as observed by the difference in the number of colonies per drop. The different columns correspond to serial dilutions.

(**Fig. 5b**). We found that addition of the TOR-inhibitor rapamycin to the media markedly enhanced the toxicity of α-syn. Indeed, a low dose of α-syn, which is otherwise innocuous, became toxic (**Fig. 5c**). Establishing the specificity of this effect to α-syn, rapamycin did not reduce growth of cells expressing glutamine expansion variants of huntingtin exon I (**Supplementary Fig. 3c**). As other studies have suggested benefits of rapamycin treatment in PD models, these results call for further investigation and suggest a complexity to the response to rapamycin that is potentially due to the vast range of processes affected by TOR activation.

DISCUSSION

We provide a novel framework in which genetic, physical and transcriptional data naturally complement each other in the context of cellular response to biological perturbations. Although the complementary nature of these data has been noted[2,5–9,37], a systematic analysis of the relationship between stimulus-specific genetic modifiers and transcriptional responses has been lacking. By examining over 150 distinct stimuli we find that differentially expressed genes and genetic hits are consistently disparate (**Table 1**); genetic hits are biased toward regulatory proteins, whereas the differentially expressed genes are biased toward metabolic processes. Indeed, each assay has inherent 'blind spots'. Many yeast regulatory proteins are not detected by transcriptional assays because either they are predominantly regulated post-transcriptionally, they have a low transcript concentration[38] or their differential expression is transient, making changes hard to measure. Conversely, the genes that are differentially transcribed are often involved in metabolic processes or redundant functions, which tend to be robust against single mutations[39].

The discordance between genetic hits and differentially expressed genes has implications for the search for therapeutic strategies. In yeast, inactivating a differentially expressed gene is no more likely to affect cell viability than targeting a randomly chosen gene. Bridging the gap between these data using techniques like ResponseNet can potentially reveal intervention points not discovered in the high-throughput assays themselves (**Fig. 2**) that may be targeted by drugs.

Our computational approach is based on a flow algorithm to connect the genetic hits and differentially expressed genes. Unlike studies that link a target gene with its causal transcriptional change[13,15,16,40–43], a flow-based approach allows for a global, efficient and simultaneous solution for multiple target genes that puts no a priori bounds on the structure of the output. Indeed, the predicted output networks have rich structures with half of all paths having a length of three edges or more. The ability of ResponseNet to analyze interactome data containing tens of thousands of nodes and edges make it well suited to analyzing the accumulating data from other species or other techniques.

We applied our approach to a yeast model for α-syn pathobiology implicated in PD. Our unbiased screen identified 77 genes whose overexpression altered α-syn toxicity (**Table 2**). These included genes involved in vesicle trafficking (as previously reported), protein degradation, cell cycle regulation, nitrosative stress, osmolyte biosynthesis and manganese transport. This screen established an interface between α-syn and a large number of cellular and environmental factors previously linked to neuropathology and, in some cases, specifically to parkinsonism, but not specifically linked to α-syn. Many of the genes we identified are highly conserved in humans, where they may exert similar effects. Indeed, eight out of nine toxicity modifiers tested had similar effects on α-syn toxicity in yeast and in neuronal systems[24].

Application of ResponseNet to the α-syn model successfully provided functional context to many of the genetic hits identified in our yeast screen (**Supplementary Fig. 3a**) and pointed to the involvement of several cellular pathways (**Figs. 3–5**). Of these, the mevalonate-ergosterol pathway is of special interest as its perturbation could potentially alter a variety of downstream pathways, including protein farnesylation and ubiquinone biosynthesis, which are closely related to the vesicle trafficking defects and mitochondrial dysfunction observed in the yeast model. Indeed, a link between sterol biosynthesis and the etiology of PD has recently emerged. Individuals with PD have significantly lower concentrations of low-density lipoprotein (LDL) cholesterol than their spouses[44], and low concentrations of LDL preceded the appearance of PD in a group of men of Japanese ancestry[45]. Our work provides a molecular framework for elucidating this connection.

The global picture obtained by integrating high-throughput genetic, transcriptional and physical data demonstrates the power of integrative approaches to illuminate underexplored cellular processes. As high-throughput assays are becoming routine in the study of complex

disease and developmental processes, approaches for deciphering these data based on their underlying characteristics are vital.

METHODS

Genetic and transcriptional datasets. Chemical perturbation data were downloaded from original papers. Genetic hits for gene inactivation included proteins that genetically interact with the inactivated gene according to *Saccharomyces cerevisiae* Genome Databases (SGD). Differentially expressed genes included genes that showed at least a twofold change in expression with a P value ≤ 0.05 (ref. 19), or else as defined according to the original papers. Genetic and mRNA profiling assays for chemical perturbations were paired if the chemical concentrations were comparable.

Interactome data description. The interactome was represented as a graph $G = (V, E)$ where nodes V represent genes and proteins and edges E represent their interactions. Different nodes represent a gene and its corresponding protein.

Bidirectional edges between protein nodes represent physical protein–protein interactions or metabolic interactions between enzymes if the substrate of one is the product of the other.

Directed edges represent regulatory interactions. Outgoing edges connected protein nodes to gene nodes if there was evidence from literature or ChIP-chip assays that the proteins may regulate the genes. Proteins nodes were connected if both proteins were transcriptional regulators and one regulated the other.

The data sources appear in the **Supplementary Note**. **Supplementary Table 2a** lists the number of interacting pairs per interaction type in the interactome.

Weighting scheme for interactome edges. *Interactions between protein nodes.* Each interacting protein pair $p_i p_j$ was associated with an interaction vector $Ip_i p_j$; vector entry $I_k p_i p_j$ is an indicator function for interaction evidence of type k. Interactions are weighted (w_{ij}) to reflect the probability that $p_i p_j$ function in a randomly selected response pathway (denoted $RP_{Pi,Pj} = 1$) as follows:

$$w_{ij} = \mathrm{P}(RP_{p_i p_j} = 1 | I_{p_i p_j}) = \mathrm{P}(I_{p_i p_j} | RP_{p_i p_j} = 1)\mathrm{P}(RP_{p_i p_j} = 1)/\mathrm{P}(I_{p_i p_j}),$$

where

$$\begin{aligned}\mathrm{P}(I_{p_i p_j}) = {} & \mathrm{P}(I_{p_i p_j} | RP_{p_i p_j} = 1)\mathrm{P}(RP_{p_i p_j} = 1) \\ & + \mathrm{P}(I_{p_i p_j} | RP_{p_i p_j} = 0)P(RP_{p_i p_j} = 0)\end{aligned}$$

We assumed conditionally independence between different types of evidence:

$$\mathrm{P}(I_{p_i p_j} | RP_{p_i p_j}) = \prod\nolimits_k \mathrm{P}(I_{kp_i p_j} | RP_{p_i p_j})$$

Interactions between protein and gene nodes. Weights were designed to reflect the reliability of the interaction on the basis of experimental evidence and binding-site conservation.

The scheme for calculating P(RP) and P(I_k| RP) and the weights per interaction type appear in the **Supplementary Note**. Because high edge weights could indicate unusually well-studied proteins[46] or imperfectness of the assumption of conditional independence, all weights were capped to a maximum value of 0.7.

Linear programming formulation. For each perturbation, the input to ResponseNet consisted of the weighted interactome $G = (V, E)$, the genetic hits $Gen \subset V$ and the differentially expressed genes $Tra \subset V$ identified following the perturbation. Each edge $(i, j) \in E$ was characterized by a weight w_{ij} and a capacity $c_{ij} = 1$.

The graph G was updated as follows:

1. $V' = V \cup \{S, T\}$, where S and T are auxiliary nodes representing the source and sink, respectively.

2. $E' = E \cup (S,i)_{\forall i \in Gen} \cup (i,T)_{\forall i \in Tra}$, connecting S to the genetic hits and T to the differentially expressed genes by directed edges.

3.

$$c_{Si} = \frac{|strength_i|}{\sum_{j \in Gen} |strength_j|},$$

$\forall i \in Gen$, where the strength of each genetic hit was measured by the variation it conferred on the number of colonies per drop if available; otherwise, strengths were uniform.

4.

$$c_{iT} = \frac{|\log_2(strength_i)|}{\sum_{j \in Tra} |\log_2(strength_j)|},$$

$\forall i \in Tra$, where the strength was measured by either the relative change in its transcript level or the P value associated with it, depending on their availability.

5. $w_{Si} = c_{Si}\ \forall i \in Gen$ and $w_{iT} = c_{iT}\ \forall i \in Tra$

Letting f_{ij} denote the flow from node i to node j and for any given $\gamma \geq 0$, the following optimization problem was solved using LOQO[47]:

$$\min_f((\sum_{i \in V', j \in V'} -\log(w_{ij}) * f_{ij}) - (\gamma * \sum_{i \in Gen} f_{Si}))$$

Subject to:

$$\sum_{j \in V'} f_{ij} - \sum_{j \in V'} f_{ji} = 0 \quad \forall i \in V' - \{S, T\}$$

$$\sum_{i \in Gen} f_{Si} - \sum_{i \in Tra} f_{iT} = 0$$

$$0 \leq f_{ij} \leq c_{ij} \quad \forall (i,j) \in E'$$

The solution $F = \{f_{ij} > 0\}$ defined the predicted response network. For enrichment analysis only protein nodes were considered, and genetic hits were included only if they received flow from nodes other than the source. Protein nodes were ranked in decreasing order according to the total amount of their incoming flow. Although the solution to the optimization problem is a directed network, this directionality only reflects the way in which the algorithm directed flow from the genetic hits to the differentially expressed genes and does not represent the causal order of events (**Supplementary Fig. 1b**).

Additional information regarding the formulation, space of solutions, setting γ value and ResponseNet performance appear in the **Supplementary Note**. For ResponseNet validation $\gamma = 10$.

Statistical analysis. Probabilities of overlap between genetic hits and differentially expressed genes were calculated using Fisher's exact test, given a total of 6,000 yeast genes. Enrichment analysis was done using the Gene Ontology Term Finder from SGD.

α-Syn toxicity modifier screen The high-throughput yeast transformation protocol appears elsewhere[12].

Immunoblotting. Phosphoglycerate kinase 1(Pgk1) mouse monoclonal antibody was used at 1:5000. Hsp26 rabbit polyclonal antibody (gift from J. Buchner, Center for Integrative Protein Science and Department of Chemistry, Technische Universität München) was used at 1:5000. Hsp104 mouse monoclonal antibody (4B; ref. 48) was used at 1:5000. S-nitosocysteine rabbit polyclonal antibody (Sigma) was used at 1:10,000.

α-Syn ResponseNet analysis. Differentially expressed genes had at least a twofold change in expression with P value ≤ 0.05 (**Supplementary Table 3c**). Capacities of edges connecting the source to genetic hits were relative to the absolute strength of the genetic hits (**Supplementary Table 3a**). Capacities of edges connecting differentially expressed genes to the sink were relative to the absolute log of the change in expression. We repeated the analysis excluding nonspecific stress responses (**Supplementary Note**). ResponseNet was run with $\gamma = 12$.

α-Syn growth in presence of small molecules. For spotting assays, yeast strains were initially grown to saturation in media containing raffinose, normalized for their A_{600} and serially diluted by fivefold before spotting onto appropriate yeast media. Growth curves were monitored using the Bioscreen instrument. Yeast strains were pre-grown in 2% raffinose medium and induced in 2% galactose medium in presence of either the compound or vehicle control (1% DMSO final) with starting A_{600} of 0.1. Cells were grown at 30 °C, with plates shaken

every 30 s to ensure proper aeration and A_{600} measurements taken every half hour over a 2-d period. The resulting data (A_{600} versus time) were plotted using Kaleidagraph. At least three independent runs were conducted for each growth condition.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

E.Y.-L. has been supported by an EMBO long-term postdoctoral fellowship and by a research grant from the National Parkinson Foundation. L.R. has been supported by Roberto Rocca doctoral fellowship and the CSBi Merck-MIT postdoctoral fellowship. L.J.S. was supported by an American Cancer Society postdoctoral fellowship. A.D.G. was a Lilly Fellow of the Life Sciences Research Foundation. M.L.G is supported by a research grant from the National Parkinson Foundation. S.L. is a founder of and has received consulting fees from FoldRx Pharmaceuticals, a company that investigates drugs to treat protein folding diseases. A.D.G., A.G.C. and S.L. are inventors on patents and patent applications that have been licensed to FoldRx. E.F. is the recipient of the Eugene Bell Career Development Chair. This work was supported in part by HHMI and by MGH/MIT Morris Udall Center of Excellence in PD Research NS38372. We thank M. Taipale, S. Treusch and G. Caraveo Piso for helpful discussions and comments and T. DiCesare for help with figures. L.R. thanks G. Casari and S. Cerutti for support and helpful discussions.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Calvano, S.E. *et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032–1037 (2005).
2. Haugen, A.C. *et al.* Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol.* **5**, R95 (2004).
3. Parsons, A.B. *et al.* Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* **22**, 62–69 (2004).
4. Begley, T.J., Rosenbach, A.S., Ideker, T. & Samson, L.D. Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. *Mol. Cell* **16**, 117–125 (2004).
5. Deutschbauer, A.M., Williams, R.M., Chu, A.M. & Davis, R.W. Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **99**, 15530–15535 (2002).
6. Fry, R.C., Begley, T.J. & Samson, L.D. Genome-wide responses to DNA-damaging agents. *Annu. Rev. Microbiol.* **59**, 357–377 (2005).
7. Birrell, G.W. *et al.* Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proc. Natl. Acad. Sci. USA* **99**, 8778–8783 (2002).
8. Winzeler, E.A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
9. Smith, J.J. *et al.* Expression and functional profiling reveal distinct gene classes involved in fatty acid metabolism. *Mol. Syst. Biol.* **2**, 2006.0009 (2006.).
10. Schiesling, C., Kieper, N., Seidel, K. & Kruger, R. Review: familial Parkinson's disease–genetics, clinical phenotype and neuropathology in relation to the common sporadic form of the disease. *Neuropathol. Appl. Neurobiol.* **34**, 255–271 (2008).
11. Outeiro, T.F. & Lindquist, S. Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* **302**, 1772–1775 (2003).
12. Cooper, A.A. *et al.* Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science* **313**, 324–328 (2006).
13. Workman, C.T. *et al.* A systems approach to mapping DNA damage response pathways. *Science* **312**, 1054–1059 (2006).
14. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.* **8**, 699–710 (2007).
15. Yeang, C.H., Ideker, T. & Jaakkola, T. Physical network models. *J. Comput. Biol.* **11**, 243–262 (2004).
16. Ourfali, O., Shlomi, T., Ideker, T., Ruppin, E. & Sharan, R. SPINE: a framework for signaling-regulatory pathway inference from cause-effect experiments. *Bioinformatics* **23**, i359–i366 (2007).
17. Dasika, M.S., Burgard, A. & Maranas, C.D. A computational framework for the topological analysis and targeted disruption of signal transduction networks. *Biophys. J.* **91**, 382–398 (2006).
18. Cormen, T.H., Leiserson, C.E., Rivest, R.L. & Stein, C. *Introduction to Algorithms* (The MIT Press, Cambridge, Massachusetts, 2001).
19. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
20. Chang, M., Bellaoui, M., Boone, C. & Brown, G.W. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *Proc. Natl. Acad. Sci. USA* **99**, 16934–16939 (2002).
21. Gasch, A.P. *et al.* Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell* **12**, 2987–3003 (2001).
22. Tofaris, G.K. & Spillantini, M.G. Physiological and pathological properties of alpha-synuclein. *Cell. Mol. Life Sci.* **64**, 2194–2201 (2007).
23. Lee, V.M. & Trojanowski, J.Q. Mechanisms of Parkinson's disease linked to pathological alpha-synuclein: new targets for drug discovery. *Neuron* **52**, 33–38 (2006).
24. Gitler, A.D. *et al.* α-Synuclein is part of a diverse and highly conserved interaction network that includes PARK9 and manganese toxicity. *Nat. Genet.* advance online publication, doi:10.1038/ng.300 (1 February 2009).
25. Sarkar, S., Davies, J.E., Huang, Z., Tunnacliffe, A. & Rubinsztein, D.C. Trehalose, a novel mTOR-independent autophagy enhancer, accelerates the clearance of mutant huntingtin and alpha-synuclein. *J. Biol. Chem.* **282**, 5641–5652 (2007).
26. Olanow, C.W. Manganese-induced parkinsonism and Parkinson's disease. *Ann. NY Acad. Sci.* **1012**, 209–223 (2004).
27. Gasch, A.P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
28. Sarver, A. & DeRisi, J. Fzf1p regulates an inducible response to nitrosative stress in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **16**, 4781–4791 (2005).
29. Uehara, T. *et al.* S-nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration. *Nature* **441**, 513–517 (2006).
30. Almeida, B. *et al.* NO-mediated apoptosis in yeast. *J. Cell Sci.* **120**, 3279–3288 (2007).
31. Zou, J., Guo, Y., Guettouche, T., Smith, D.F. & Voellmy, R. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell* **94**, 471–480 (1998).
32. Flower, T.R., Chesnokova, L.S., Froelich, C.A., Dixon, C. & Witt, S.N. Heat shock prevents alpha-synuclein-induced apoptosis in a yeast model of Parkinson's disease. *J. Mol. Biol.* **351**, 1081–1100 (2005).
33. Auluck, P.K., Meulener, M.C. & Bonini, N.M. Mechanisms of Suppression of {alpha}-Synuclein Neurotoxicity by Geldanamycin in *Drosophila*. *J. Biol. Chem.* **280**, 2873–2878 (2005).
34. Lin, J.T. & Lis, J.T. Glycogen synthase phosphatase interacts with heat shock factor to activate CUP1 gene transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 3237–3245 (1999).
35. Hickman, M.J. & Winston, F. Heme levels switch the function of Hap1 of *Saccharomyces cerevisiae* between transcriptional activator and transcriptional repressor. *Mol. Cell. Biol.* **27**, 7414–7424 (2007).
36. Duennwald, M.L., Jagadish, S., Muchowski, P.J. & Lindquist, S. Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc. Natl. Acad. Sci. USA* **103**, 11045–11050 (2006).
37. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* **23**, 561–566 (2005).
38. Holstege, F.C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
39. Deutscher, D., Meilijson, I., Kupiec, M. & Ruppin, E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat. Genet.* **38**, 993–998 (2006).
40. Shachar, R., Ungar, L., Kupiec, M., Ruppin, E. & Sharan, R. A systems-level approach to mapping the telomere length maintenance gene circuitry. *Mol. Syst. Biol.* **4**, 172 (2008).
41. Bromberg, K.D., Ma'ayan, A., Neves, S.R. & Iyengar, R. Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. *Science* **320**, 903–909 (2008).
42. Tu, Z., Wang, L., Arbeitman, M.N., Chen, T. & Sun, F. An integrative approach for causal gene identification and gene regulatory pathway inference. *Bioinformatics* **22**, e489–e496 (2006).
43. Suthram, S., Beyer, A., Karp, R.M., Eldar, Y. & Ideker, T. eQED: an efficient method for interpreting eQTL associations using protein networks. *Mol. Syst. Biol.* **4**, 162 (2008).
44. Huang, X. *et al.* Lower low-density lipoprotein cholesterol levels are associated with Parkinson's disease. *Mov. Disord.* **22**, 377–381 (2007).
45. Huang, X., Abbott, R.D., Petrovitch, H., Mailman, R.B. & Ross, G.W. Low LDL cholesterol and increased risk of Parkinson's disease: prospective results from Honolulu-Asia Aging Study. *Mov. Disord.* **23**, 1013–1018 (2008).
46. Hoffmann, R. & Valencia, A. Life cycles of successful genes. *Trends Genet.* **19**, 79–81 (2003).
47. Vanderbei, R.J. LOQO User's Manual–Version 3.10. *Optimization Methods and Software* **12**, 485–514 (1999).
48. Cashikar, A.G. *et al.* Defining a pathway of communication from the C-terminal peptide binding domain to the N-terminal ATPase domain in a AAA protein. *Mol. Cell* **9**, 751–760 (2002).