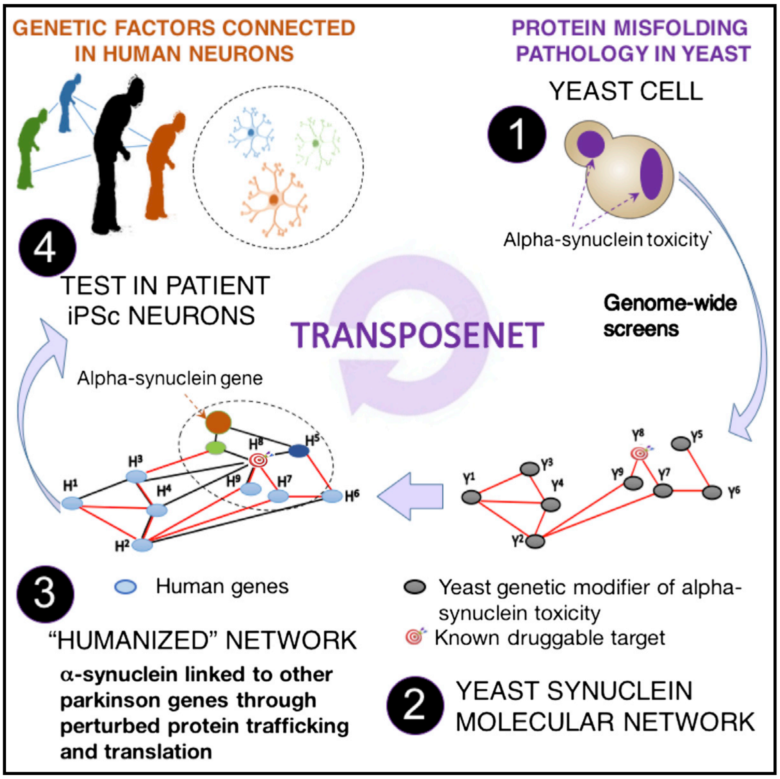
Genome-Scale Networks Link Neurodegenerative Disease Genes to #-Synuclein through Specific Molecular Pathways

Cell Systems

Article

Genome-Scale Networks Link Neurodegenerative Disease Genes to α-Synuclein through Specific Molecular Pathways

Graphical Abstract



Highlights

- 332 modifiers of α-syn toxicity in yeast are assembled in a gene network
- The computational method TransposeNet transposes molecular interactions across species
- A "humanized" toxicity map links parkinsonism genes and druggable targets to α-syn
- Common pathologies identified in patient neurons harboring distinct disease genes

Authors

Vikram Khurana, Jian Peng, Chee Yeun Chung, ..., Ernest Fraenkel, Bonnie Berger, Susan Lindquist

Correspondence

vkhurana@bwh.harvard.edu (V.K.), fraenkel-admin@mit.edu (E.F.), bab@csail.mit.edu (B.B.)

In Brief

Khurana et al. develop a computational approach to transpose molecular networks across species. Modifiers from genome-wide screens against alpha-synuclein toxicity in yeast yield a "humanized" network that interconnects genetic forms of parkinsonism and predicts pathologies in patient-derived neurons.

Khurana et al., 2017, Cell Systems *4*, 157–170
February 22, 2017 © 2017 The Authors. Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.cels.2016.12.011

CellPress

Cell Systems
Article

Genome-Scale Networks Link Neurodegenerative Disease Genes to α-Synuclein through Specific Molecular Pathways

Vikram Khurana,[1,2,3,14,17,*] Jian Peng,[1,4,14,15] Chee Yeun Chung,[1,14,16] Pavan K. Auluck,[1] Saranna Fanning,[1] Daniel F. Tardiff,[1,16] Theresa Bartels,[1] Martina Koeva,[1,5] Stephen W. Eichhorn,[1] Hadar Benyamini,[1] Yali Lou,[1] Andy Nutter-Upham,[1] Valeriya Baru,[1] Yelena Freyzon,[1] Nurcan Tuncbag,[5] Michael Costanzo,[6] Bryan-Joseph San Luis,[6] David C. Schöndorf,[7] M. Inmaculada Barrasa,[1] Sepehr Ehsani,[1] Neville Sanjana,[8,9] Quan Zhong,[10] Thomas Gasser,[7] David P. Bartel,[1] Marc Vidal,[11,12] Michela Deleidi,[7] Charles Boone,[6] Ernest Fraenkel,[5,*] Bonnie Berger,[3,*] and Susan Lindquist[1,13]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Ann Romney Center for Neurologic Disease, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[3]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[4]Computer Science and Artificial Intelligence Laboratory and Department of Mathematics, MIT, Cambridge, MA 02139, USA
[5]Department of Biological Engineering, MIT, Cambridge, MA 02139, USA
[6]Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5G 1L6, Canada
[7]Department of Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), and Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, 72076, Germany
[8]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[9]New York Genome Center and Department of Biology, New York University, New York, NY 10013, USA
[10]Department of Biological Sciences, Wright State University, Dayton, OH 45435, USA
[11]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[12]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[13]HHMI, Department of Biology, MIT, Cambridge, MA 02139, USA
[14]Co-first author
[15]Present address: Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL 61801, USA
[16]Present address: Yumanity Therapeutics, Cambridge, MA 02139, USA
[17]Lead Contact
*Correspondence: vkhurana@bwh.harvard.edu (V.K.), fraenkel-admin@mit.edu (E.F.), bab@csail.mit.edu (B.B.)
http://dx.doi.org/10.1016/j.cels.2016.12.011

SUMMARY

Numerous genes and molecular pathways are implicated in neurodegenerative proteinopathies, but their inter-relationships are poorly understood. We systematically mapped molecular pathways underlying the toxicity of alpha-synuclein (α-syn), a protein central to Parkinson's disease. Genome-wide screens in yeast identified 332 genes that impact α-syn toxicity. To "humanize" this molecular network, we developed a computational method, TransposeNet. This integrates a Steiner prize-collecting approach with homology assignment through sequence, structure, and interaction topology. TransposeNet linked α-syn to multiple parkinsonism genes and druggable targets through perturbed protein trafficking and ER quality control as well as mRNA metabolism and translation. A calcium signaling hub linked these processes to perturbed mitochondrial quality control and function, metal ion transport, transcriptional regulation, and signal transduction. Parkinsonism gene interaction profiles spatially opposed in the network (*ATP13A2/PARK9* and *VPS35/PARK17*) were highly distinct, and network relationships for specific genes (*LRRK2/PARK8*, *ATXN2*, and *EIF4G1/PARK18*) were confirmed in patient induced pluripotent stem cell (iPSC)-derived neurons. This cross-species platform connected diverse neurodegenerative genes to proteinopathy through specific mechanisms and may facilitate patient stratification for targeted therapy.

INTRODUCTION

Common neurodegenerative diseases result in the loss of distinct neuronal populations and abnormal accumulation of misfolded proteins. Synucleinopathies, including Parkinson's disease (PD), dementia with Lewy bodies, and multiple system atrophy, are associated with abnormal intracellular aggregation of alpha-synuclein (α-syn). Alzheimer's disease (AD) is associated with beta-amyloid (Aβ) and tau accumulation, while amyotrophic lateral sclerosis (ALS) is associated with altered localization and accumulation of TAR DNA-binding protein 43 (TDP-43). The richest source of hypotheses regarding the pathogenesis of these diseases has derived from neuropathology of postmortem brain. While providing pivotal insights, these observations are made decades after disease inception.

A revolution in human genetic analysis over the last 20 years has uncovered disease-causing mutations that connect protein misfolding to the neurodegenerative process. For instance, point mutations and gene multiplication at the α-syn (*SNCA*) locus lead to rare but early-onset, highly penetrant forms of PD and dementia. Common polymorphisms in regulatory regions of the *SNCA* locus that increase gene expression confer increased risk for late-onset PD (Fuchs et al., 2008; Nalls et al., 2014). These studies enabled the creation of animal and cellular disease models and enriched our understanding of disease mechanisms. However, with this knowledge, a new set of challenges has emerged.

First, seemingly disparate genes have been tied to particular clinical phenotypes. For example, parkinsonism is characterized by slowness (bradykinesia), rigidity, tremor, and postural instability. The most common form is PD, defined by α-syn pathology and loss of dopaminergic neurons. However, numerous other disease entities—tied to distinct genetic signatures and neuropathology—can lead to parkinsonism, demonstrating that there is not a simple correspondence among genotype, neuropathology, and clinical presentation (Martin et al., 2011; Shulman et al., 2011; Verstraeten et al., 2015). Those few genetic loci with parkinsonism as the primary clinical phenotype have been given a numeric "*PARK*" designation (for example, the *SNCA/PARK1* locus itself and *LRRK2/PARK8*), but even mutations in the same gene can produce distinct neuropathology and diverse clinical presentations (Martin et al., 2011; Rajput et al., 2006; Shulman et al., 2011; Verstraeten et al., 2015). Understanding the inter-relationship between genetic risk factors for parkinsonism, and their relationship to α-syn itself, is vital for patient stratification and targeted therapeutic strategies.

Second, human genetic studies have sometimes produced ambiguous and controversial data. For rare variants, substantial recent genetic divergence of human populations may render traditional methods of cross-validation between different populations unfeasible (Nelson et al., 2012; Tennessen et al., 2012). Inconsistencies in the literature abound; for example, studies implicating UCHL1 as "*PARK5*" and the translation initiation factor EIF4G1 as "*PARK18*" have failed to reproduce. For common polymorphisms, the challenge is distinguishing between multiple candidate gene loci in linkage to a SNP. It is becoming clear that biological validation will be required to fully establish which genetic factors are causally related to disease processes and how (Casals and Bertranpetit, 2012).

One approach to validating candidate gene variants, and understanding their relationship to proteinopathy, is to systematically screen the entire genome to identify every gene that modifies proteotoxicity when overexpressed or deleted. This is achievable in baker's yeast (*Saccharomyces cerevisiae*), a unicellular eukaryote of unparalleled genetic tractability. Yeast has proved highly informative for understanding the cytotoxicity induced by misfolded proteins (Khurana and Lindquist, 2010). This is not surprising, because human genetic data for neurodegenerative diseases heavily implicate cellular pathways that are among the most highly conserved in eukaryotic evolution, including protein homeostasis and quality control, protein trafficking, RNA metabolism, and mitochondrial function (Brás et al., 2015; Guerreiro et al., 2015).

Expressing toxic proteins relevant to neurodegeneration in yeast creates a robust and easily scorable growth and viability defect amenable to genome-wide phenotypic screening. Toxicities of α-syn, Aβ, and TDP-43 have been screened by individually overexpressing each one of ~5,500 open reading frames (ORFs) that comprise the majority of the yeast genome (Khurana and Lindquist, 2010; Kim et al., 2014; Treusch et al., 2011; Yeger-Lotem et al., 2009). These screens have guided the discovery of cellular pathologies in neurons and animal models (Cooper et al., 2006; Dhungel et al., 2015; Khurana and Lindquist, 2010; Kim et al., 2014), provided important insights on the relationship of genetic modifier data to gene-expression analysis (Yeger-Lotem et al., 2009), and led to the identification of novel human disease genes (Elden et al., 2010). Recently, processes pinpointed by phenotypic screening in a yeast synucleinopathy model led to the discovery of cellular pathologies in induced pluripotent stem cell (iPSC)-derived neurons from patients with PD due to mutations at the α-syn locus (Chung et al., 2013). In that study, integrating high-throughput genetic and small-molecule screens identified genes and small molecules that could correct pathologies from yeast to neurons (Chung et al., 2013; Tardiff et al., 2013, 2014).

Here, we build genome-scale networks of α-syn and other proteotoxicities by combining a new computational approach with substantially broader yeast genetic screens. To discover meaningful molecular connections in yeast and patient-derived neurons, we develop a TransposeNet algorithm that (1) maps yeast hits to their human homologs by considering sequence, structure, and molecular interactions; (2) builds networks by linking yeast hits and hidden human genes through an optimization framework based on the prize-collecting Steiner forest algorithm (SteinerForest Ensemble); and (3) transposes molecular interactions across species from yeast to human, utilizing the high density of known molecular interactions in yeast to compensate for the relative sparseness of the human interactome. The networks linked many parkinsonism and neurodegenerative disease risk factors to α-syn toxicity through specific molecular pathways, most notably vesicle trafficking and mRNA metabolism.

RESULTS

SteinerForest Ensemble Networks Uncover Biological Connections between α-Syn Screen Hits

One conventional approach to creating a network from a gene list is to connect them via known genetic or physical protein-protein interactions. To illustrate this, we considered 77 genes that modify α-syn toxicity in our previous overexpression screen (Tables S1 and S2). Even with the rich yeast interactome, 30 hits were not incorporated into the network (Figure 1A, top). Moreover, some genes ("hubs") occupied a central position in the network, not because of their importance to proteotoxicity but because they were connected to more genes. For example, PMR1 is a hub that has 955 annotated interactions in BioGRID compared to the median of 70 interactions across the 77 modifiers (Figure 1A, top right; Table S2).

To build more inclusive networks, we adapted the "Prize-collecting Steiner Forest (PCSF) algorithm," which connects gene or protein "nodes" through molecular interactions, or "edges" (Huang and Fraenkel, 2009; Tuncbag et al., 2013, 2016) (Figure 1A, bottom). Edges can include genetic or physical

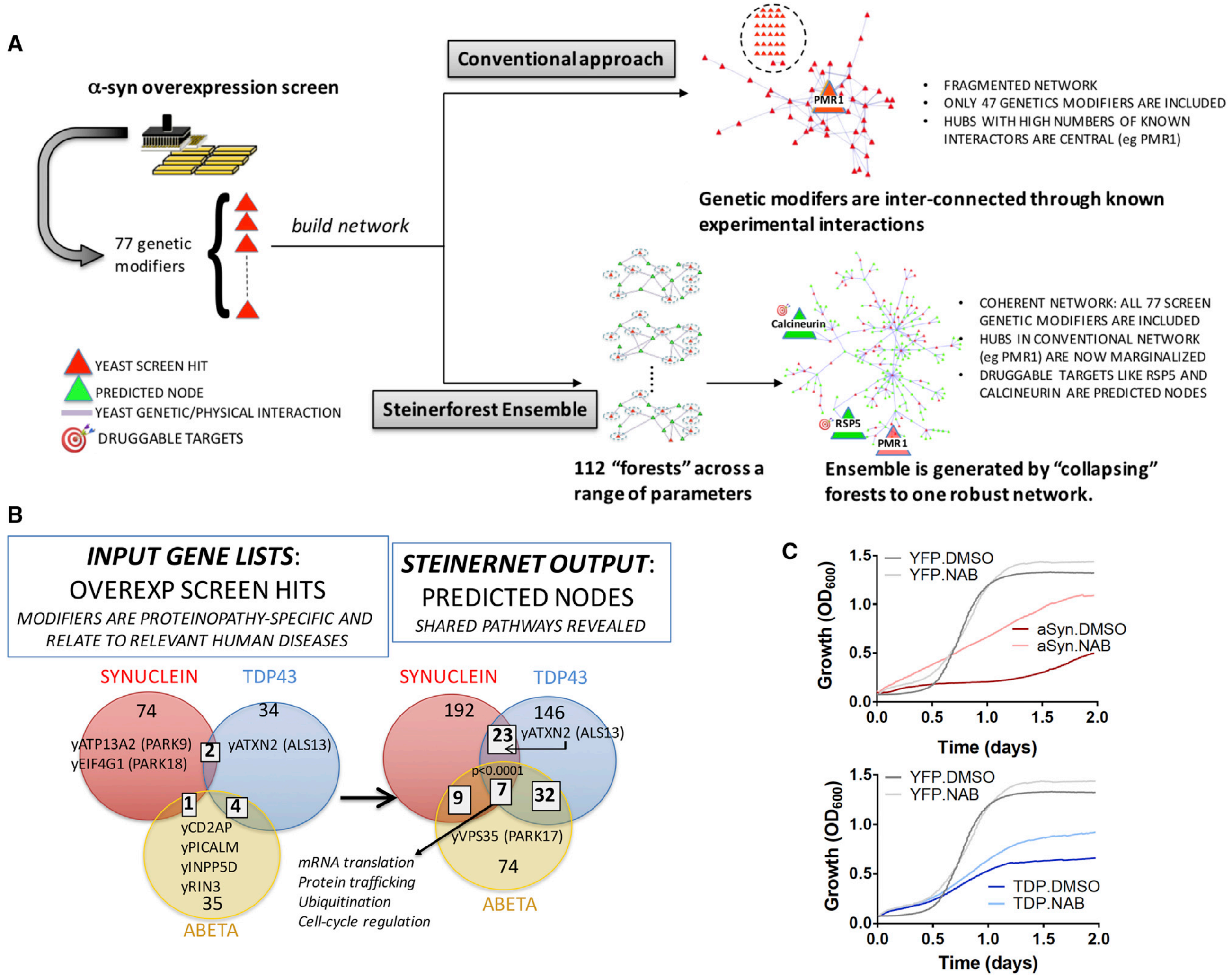


Figure 1. SteinerForest Ensemble Builds Proteotoxicity Networks from Yeast Genetic Screens and Uncovers Druggable Targets

(A) SteinerForest Ensemble methodology versus conventional approach. 77 genetic modifiers ("hits") from a previous overexpression screen against α-syn toxicity are mapped to the yeast interactome. The conventional approach misses 30 genetic hits and overemphasizes "hub" genes like PMR1. SteinerForest Ensemble includes all 77 hits and predicts additional nodes of biological relevance, including the druggable targets Rsp5 and Calcineurin (Cnb1).

(B) Left: hits from three published overexpression screens for α-syn, Aβ, and TDP-43 proteotoxicities in yeast. Venn diagrams indicate the numbers of genetic modifiers recovered. Right: a comparison of the output SteinerForest Ensemble networks generated from inputting these three sets of screen hits (empirical p value for 1,000 similarly connected random networks is shown for triple-wise comparison).

(C) Growth curves demonstrating effects of a compound that activates Rsp5, NAB, on TDP-43 (bottom) and α-syn (top) toxicity. Yeast expressing yellow fluorescent protein (YFP), TDP-43(TDP), or α-syn were treated with 20 μM (for TDP-43) or 10 μM (for α-syn) NAB. Growth was monitored over time by optical density (OD) at 600 nm. Results are representative of three experiments.

interactions or annotated pathways from curated databases (Szklarczyk et al., 2015) and are refined by minimizing "cost." Costs increase (1) when a "prized" node (an original hit from a genetic screen) is excluded, (2) when an "edge" connecting two nodes derives from a low-confidence interaction, or (3) when edges connect to hubs. To ensure that our PCSFs were not dependent on specific parameterization, we generated an ensemble of 112 individual forests with different algorithm parameters and created an averaged, or "collapsed," representative network through a maximum spanning-tree algorithm ("SteinerForest Ensemble"; Figure 1A, bottom right).

To encompass the largest number of prized nodes while avoiding unlikely interactions, the PCSF method introduces "predicted nodes," which are proteins or genes not part of the original prized hit list, (Figure 1A, green triangles). Predicted nodes will occur between two nodes within the network. However, as the final network is a superposition of many different networks, these may be at the periphery in the final Ensemble output. Predicted nodes can add biological value, because any high-throughput screen will miss many true biological connections.

When we applied SteinerForest Ensemble to our previous α-syn overexpression screen data, the fragmented networks became more coherently connected. All 77 modifier genes were now incorporated in the network, (Figure 1A, bottom right; Tables S1 and Table S3; Data S1). By penalizing the exclusion of genetic modifiers and the inclusion of hubs, the PCSF algorithm favored the biological context at the expense of hubs. To establish specificity of the network output, we generated ensembles of forests from 1,000 sets of 77 genes randomly chosen from the yeast genome with identical connectivity (degree distribution) to the α-syn modifier list. An empiric p value for each node (based on probability of occurring in a network by chance) was significant ($p = 0.025$, $SD = 0.021$).

Importantly, predicted nodes (Figure 1A, green triangles) included genetic modifiers of α-syn toxicity not hit in the original screen but uncovered through other studies, including Sec14 (phospholipase D) (Outeiro and Lindquist, 2003), and Pbp1 (yeast homolog of ataxin 2; see below and Figure 3). This network also identified two druggable targets: Cnb1 (Calcineurin subunit B) and Rsp5 (Figure 1A, bottom right). Cnb1 is targeted by FK506, a drug that ameliorates α-syn toxicity (Caraveo et al., 2014). Rsp5 is the target of a specific N-arylbenzimidazole (NAB) that protects against α-syn toxicity (Tardiff et al., 2013). The SteinerForest Ensemble methodology thus connects genetic screen hits through biologically relevant pathways, including druggable targets.

Cross-Comparison of α-Syn, TDP-43, and Aβ Proteinopathies Reveals Distinct and Shared Mechanisms

To cross-compare different proteinopathies, we examined previous Aβ and TDP-43 overexpression screens (Figure 1B; "yeast over-expression networks" in Table S1; Data S1) and found virtually no overlap (Figure 1B, left; Table S2). There was, however, reassuring overlap between the yeast genetic modifiers and disease genes associated with the human disorders, including putative parkinsonism genes recovered in the α-syn screen (*ATP13A2* [*PARK9*] and *EIF4G1* [*PARK18*]), AD risk factors in the Aβ screen (*PICALM*, *CD2AP*, *INPP5D*, and *RIN3*), and an ALS genetic risk factor (*ATXN2*) in the TDP-43 screen (Elden et al., 2010).

SteinerForest Ensembles from these screen hits revealed more biological overlap between these proteinopathies including protein trafficking, mRNA translation, ubiquitination, and cell-cycle genes (Tables S3 and S4; Figure 1B, right). There was also a crossover between genetic risk factors for distinct human diseases: the *ATXN2* homolog was a predicted node in the α-syn network (confirmed as a modifier of α-syn toxicity; Figures 3 and 4), and the homolog of the parkinsonism gene *VPS35* (*PARK17*) was a predicted node in the yeast Aβ network. VPS35 encodes a key component of the retromer complex, and defective retromer function has been identified in AD brain and animal models (Small et al., 2005). These overlaps were unrelated to increasing the number of genes. Empirical p values for 1,000 similarly connected random networks were statistically significant, whether considered pairwise ($p \leq 0.002$) or triplewise ($p \leq 0.001$).

One trafficking gene predicted to be a common node between all three proteinopathies was Rsp5, a ubiquitin ligase activated by NAB. Indeed, NAB was originally recovered in a small-molecule screen against TDP-43 proteinopathy in yeast. We utilized a sensitive bioscreen assay to test NAB on growth defects induced by these proteinopathies. NAB rescued all three proteinopathies as predicted by the network. It was most effective for α-syn (Figure 1C) and only rescued against Aβ toxicity synergistically in combination with other compounds known to protect from Aβ toxicity (unpublished data). NAB failed to provide significant rescue for 20 unrelated toxic yeast strains (Figure S1).

TransposeNet Generates a "Humanized" Network

It would be desirable to identify connections between our yeast molecular networks to human genes, including human disease genes that have no straightforward homologs in yeast. We therefore developed TransposeNet, a suite of computational methods to "humanize" yeast molecular networks (Figure 2A).

The first step in TransposeNet is assignment of yeast-to-human homology by considering sequence similarity (BLAST and DIOPT scores; Hu et al., 2011), yeast-to-human structure alignments (using the *HHpred* tool; Söding et al., 2005), and incorporating network topology (Figure 2A, top left). Network topology assesses neighborhoods of genetic and physical molecular interactions around a given protein, positing "guilt-by-association" logic that the topological place within a network relates to biological function (Cho et al., 2016). Thus, sharing similar neighbors should be a factor in determining whether two proteins are homologs. The relative weight of each homology method was carefully tuned (see STAR Methods and Figure S2 for full details), providing a more comprehensive and unified protein homology score (Berger et al., 2013; Singh et al., 2008; Söding et al., 2005). The underlying framework that relates genes according to these different features is known as diffusion-component analysis (DCA). DCA has also been used as the core algorithm in Mashup, a tool for integrating multiple hetergeneous interactomes. More information can be found in the Method Details section of STAR Methods and in Cho et al., (2016).

Our method assigned 4,923 yeast proteins to human homologs and conversely predicted yeast homologs for 15,200 human proteins, a substantial improvement over BLAST (4,023 yeast to human and 7,248 human to yeast) or BLAST with HHpred (4,312 yeast to human and 9,577 human to yeast). Additionally, our method improved predictions as determined by gene ontology (GO) accuracy and Jaccard similarity scores (STAR Methods; Figure S3) and did not introduce false positives for high-confidence yeast-human protein pairs (EnsemblCompara; STAR Methods).

There is high conservation of core eukaryotic biology from yeast to man, and pivotal complementation studies in yeast have determined the functions of many genes in other species, including human (Osborn and Miller, 2007; Kachroo et al., 2015). On this basis, we used our homology tool to augment the human interactome with interactions inferred from the much richer yeast interactome. This was the key advance that enabled TransposeNet. Importantly, this cross-species "edge" transposition did not increase the rate of false-positive hits; rather, it substantially improved network performance. In fact, for our screen hits the PCSF-based SteinerForest Ensemble outperformed two alternative network-building methodologies, DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009) (STAR Methods; Figure S4).

In our humanized networks (indexed in Table S1; Figure 2A, right), each yeast gene (red triangle) is connected to one or more human homologs (blue circles) based on our homology-tool-generated score. SteinerForest Ensemble then interconnects

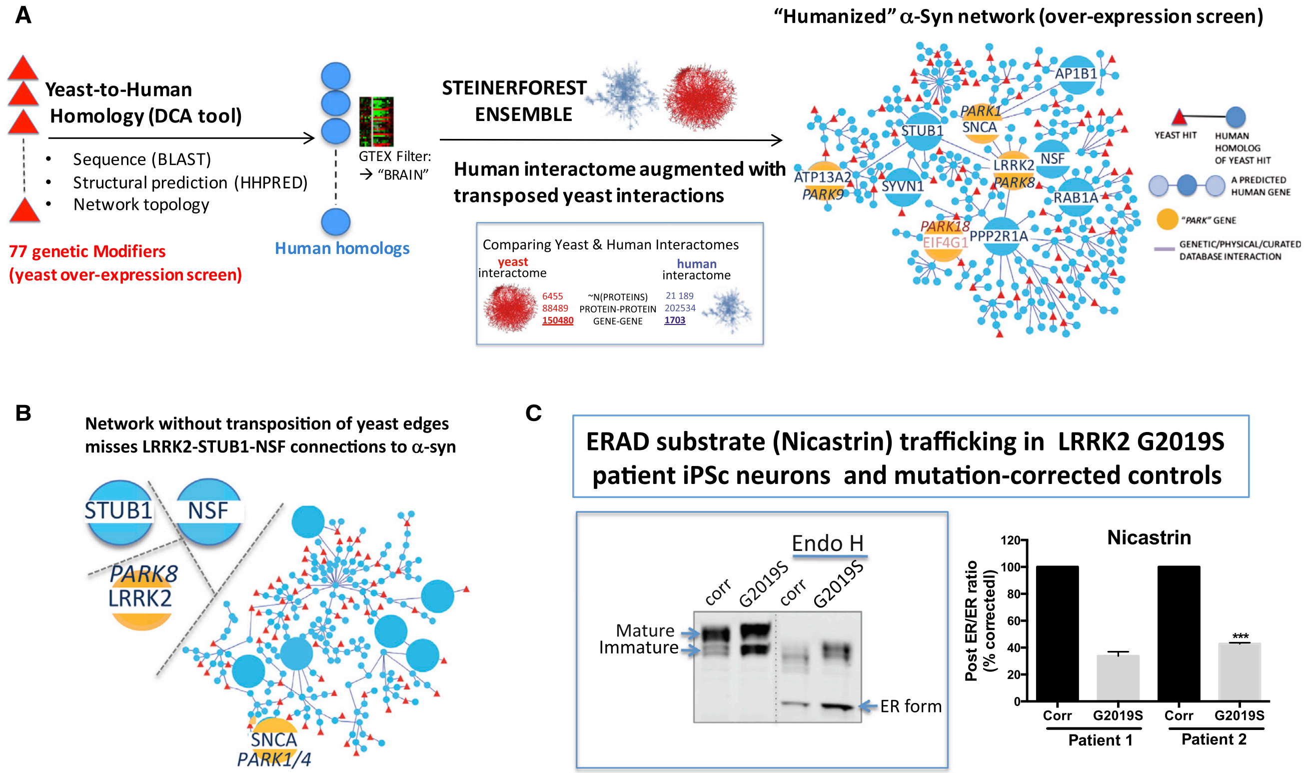


Figure 2. A "Humanized" TransposeNet Network Incorporates LRRK2 into the α-Syn Proteotoxicity Network

(A) A humanized network is generated from the 77 α-syn overexpression screen hits by TransposeNet. Each yeast gene (red triangle) is linked to its human homolog(s) (blue circle) by a weight proportional to the homology strength. Edges are weighted based on their experimental level of confidence. Certain nodes are enlarged for emphasis. LRRK2 is linked within network via NSF1 and STUB1. EIF4G1 is marked in red because it is a controversial PD gene. Inset: density of known molecular interactions in yeast and human (Biogrid: https://wiki.thebiogrid.org/doku.php/statistics). Abbreviations: DCA, diffusion component analysis; PARK, known "parkinsonism" gene. See the Supplemental Information for complete network.

(B) The effect on the humanized network of withholding yeast edge augmentation.

(C) Accumulation of Nicastrin in the endoplasmic reticulum (ER) in LRRK2^{G2019S} mutant iPSC-derived dopaminergic neurons compared to mutation-corrected control neurons. Endoglycosidase H (Endo H) removes post-ER glycosylation and reveals the ER form of Nicastrin, an ER-associated degradation substrate. The post ER-to-ER ratio was calculated using the ratio of the mature form over the deglycosylated ER form. Data represent mean ± SEM (n = 2 for patient 1 and n = 3 for patient 2; ***p < 0.0001, two-tailed t test).

each resulting human gene or protein, through edges generated from the human interactome augmented with the humanized yeast molecular interactome. If a particular human homolog of a yeast genetic modifier had been implicated as a parkinsonism gene, a small inclusion weight is given. However, no special preference was given to any human disease genes other than close homologs of our yeast hits.

Humanized Network from Overexpression Screen Connects α-Syn to Other Human Disease Genes

We tested the humanized network approach on the 77 modifiers from the α-syn overexpression screen ("α-syn over-expression humanized network"; Tables S1, S9, and S11; Data S1; Figure 2A, right). Several predicted human nodes in the resultant humanized network had no obvious homologs in the yeast proteome, the most striking example being α-syn itself. α-syn was connected to endoplasmic reticulum (ER) quality control and protein trafficking modifiers through a predicted node Ap1b1 (Figure 2A, right), a component of the clathrin adaptor complex that localizes in the immediate vicinity of α-syn in neurons (Chung et al., 2017, in this issue of *Cell Systems*). The emergence of α-syn in the humanized network strongly indicates that a functional, highly interconnected relationship between our original yeast genetic hits and α-syn is conserved from yeast to man.

LRRK2 and α-Syn Are Connected through ER Stress and Vesicle Trafficking

The kinase/GTPase LRRK2, another PD-gene-encoded protein without an obvious yeast homolog, was centrally incorporated into the humanized network (Figure 2A, right). We tested the robustness and specificity of this finding by computationally generating ensembles of humanized Steiner forests from 1,000 lists of genes that were randomly selected (matching the size of our original α-syn genetic modifier list). LRRK2 and α-syn (SNCA) occurred together in 72% of humanized networks generated through SteinerNet Ensemble from our input list (individually, SNCA appeared in 86% and LRRK2 in 76% of networks). Neither was incorporated in any of the humanized networks generated from Aβ or TDP-43 screen hits ("TDP-43"- and "Aβ"-"overexpression humanized networks" in Table S1; Data S1). LRRK2 and α-syn appeared together in 0/1,000 of the randomly generated network ensembles. Without transposition of yeast interaction information into our networks, α-syn was peripherally placed and its connection to Ap1b1 (see above) was lost and LRRK2 was entirely absent (Figure 2B). Thus, the inclusion of LRRK2 and α-syn is robust, specific, and dependent upon augmentation of human networks with yeast interaction data.

LRRK2 was related to the humanized α-syn network through proteins involved in ER-to-Golgi trafficking (Nsf1 and Rab1a) and ER quality control (Stub1/Chip/Scar6, Sgk1, and Syvn1), pathways previously implicated among many others in LRRK2 (Cho et al., 2014; Liu et al., 2012) and α-syn (Chung et al., 2013; Cooper et al., 2006)-induced toxicity. Our data pinpointed these pathways as key points of convergence. We previously showed that the A53T mutation and triplication of wild-type α-syn leads to pathologic accumulation of specific trafficked proteins in the ER of patient-derived neurons, including Nicastrin (Chung et al., 2013). Using previously generated LRRK2 mutant iPSC (Figure S5), we recapitulated this phenotype. As early as 4 weeks after initiating differentiation, Nicastrin accumulated in the ER of LRRK2^{G2019S} dopaminergic neurons compared to isogenic mutation-corrected controls (Figure 2C), phenocopying the previously described pathology in neurons derived from patients with α-syn mutations. Thus, the humanized α-syn network correctly predicted convergence of cellular pathologies in distinct forms of parkinsonism. A Nicastrin trafficking defect has also been demonstrated in LRRK2 knockout mouse fibroblasts (Cho et al., 2014), raising the possibility that the G2019S mutation may lead to deficiency of a LRRK2-related function in protein trafficking.

Genome-wide Pooled Overexpression and Deletion Screens against α-Syn Toxicity

For a more comprehensive view, we executed two additional genome-wide screens against α-syn toxicity: (1) a genome-wide deletion screen to identify nonessential genes that, when deleted, lead to an extreme sensitivity to low levels of α-syn that would otherwise not be toxic (Figure S6A number 2; Table S5); (2) a pooled screen in which the galactose-inducible overexpression library was transformed en masse into our α-syn HiTox strain (Figure S6A, number 3; Table S6). For pooled screens, we sequenced plasmid DNA to identify genes specifically over- or under-represented under selective conditions. These are putative suppressors and enhancers of toxicity, respectively. We compared a similarly transformed YFP strain as a control. Pooled screens are far more rapid, and theoretically more sensitive, than individually transforming each library plasmid into the α-syn strain and measuring growth.

These screens encompassed tests of approximately 10,000 potential genetic interactions (~5,500 overexpression, ~4,500 deletion). After extensive validation of the hits (Figures S6C and S7B), we recovered 318 genetic modifiers. Very little overlap existed between the specific genes recovered in the deletion versus the overexpression screens (Figure 3A). However, we found considerable overlap in the biological pathways represented (see network analysis below). 16 modifiers have emerged in independent work from our laboratory (Caraveo et al., 2014; Chung et al., 2013) or were identified herein (Table S7). 14 of these were distinct from our screen hits, leading to a total of 332 genetic modifiers of α-syn toxicity (Figure 3A).

Homologs of *PARK* and Other Neurodegeneration Genes Modify α-Syn Toxicity in Yeast

Modifiers of α-syn toxicity included homologs of many known genetic risk factors for parkinsonism and other neurodegenerative disease phenotypes (Figure 3B; Table S14). These included genes involved in vesicle trafficking (*yRAB7L1*, *yRAB39B*, *ySORL1*, *ySYNJ1/PARK20*, and *yVPS35/PARK17*), mRNA translation (*yATXN2* and *yEIF4G1/PARK18*), mitochondrial quality control and function (*yCHCHD2/10*), metal ion transport (*yATP13A2*), transcriptional regulation (*yATXN7*), metabolism (*yDHDDS*), and signaling (*yPDE8B* and *yPPP2R2B/ATXN12)*, among others. Many of these genes, including those at so-called *PARK* loci, have been implicated in neuronal pathologies quite distinct from the α-syn pathology that defines PD. Their recovery in our screens suggested that mechanisms of neurotoxicity related to diverse neurodegenerative disease genes might be shared.

Of the 19 *PARK* loci, 9 have clear yeast homologs (Table S8). Four of these emerged in our screens: *yATP13A2* (*PARK9*) (*YPK9* in yeast), *yVPS35* (*PARK17*) (*VPS35*), *yEIF4G1* (*PARK18*) (*TIF4631*, *TIF4632*), and *ySYNJ1* (*PARK20*) (*INP53*). A fifth putative *PARK* gene, *yRAB7L1* (*PARK16*) (*YPT7*), emerged as a genetic modifier when tested as a candidate (see below). The probability of recovering homologs of these genes by chance is low ($p = 0.00013$; hypergeometric test). None of these *yPARK* genes were modifiers in the Aβ or TDP-43 overexpression screens (Table S2). These findings underscore the biological specificity of the α-syn screen hits in yeast.

TransposeNet Generates a Genome-Scale “Map” of α-Syn Toxicity

We applied TransposeNet to homologs of the 332 α-syn toxicity modifiers to generate a humanized network, or “map” (“Complete α-syn humanized network” in Tables S1, S10, and S11; Figures 3B and S8; Data S1). Multiple genes implicated in neurodegeneration emerged in this α-syn network by direct homology to yeast hits or as predicted network nodes (Figure 3B; Figure S8; Table S14).

We superimposed gene ontologies onto “branches” in our map (Figure 3B), and various vesicle-mediated transport processes dominated. Genetic risk factors associated with typical PD (*SNCA* itself, *LRRK2*, *RAB7L1*, and *VPS35*) were concentrated in the subnetwork enriched in vesicle trafficking genes (Figure 3B). In contrast, the majority of neurodegeneration genes associated with non-Lewy neuropathology, atypical parkinsonism, or non-parkinsonian neurodegenerative phenotypes (Table S14) were distant from the vesicle trafficking network. A full analysis of the biological processes enriched in the network branches is provided in Table S12. Notably, this humanized network elucidated the molecular context in which the previously identified druggable targets NEDD4 (Tardiff et al., 2013) and Calcineurin (Caraveo et al., 2014) impact α-syn toxicity, and identified the transcription factor NFAT, through which Calcineurin exerts its toxicity (Caraveo et al., 2014), as a hub (Figure 3B).

Furthermore, both α-syn itself and LRRK2 were predicted as nodes, just as in the overexpression network (Figure 2A). In the

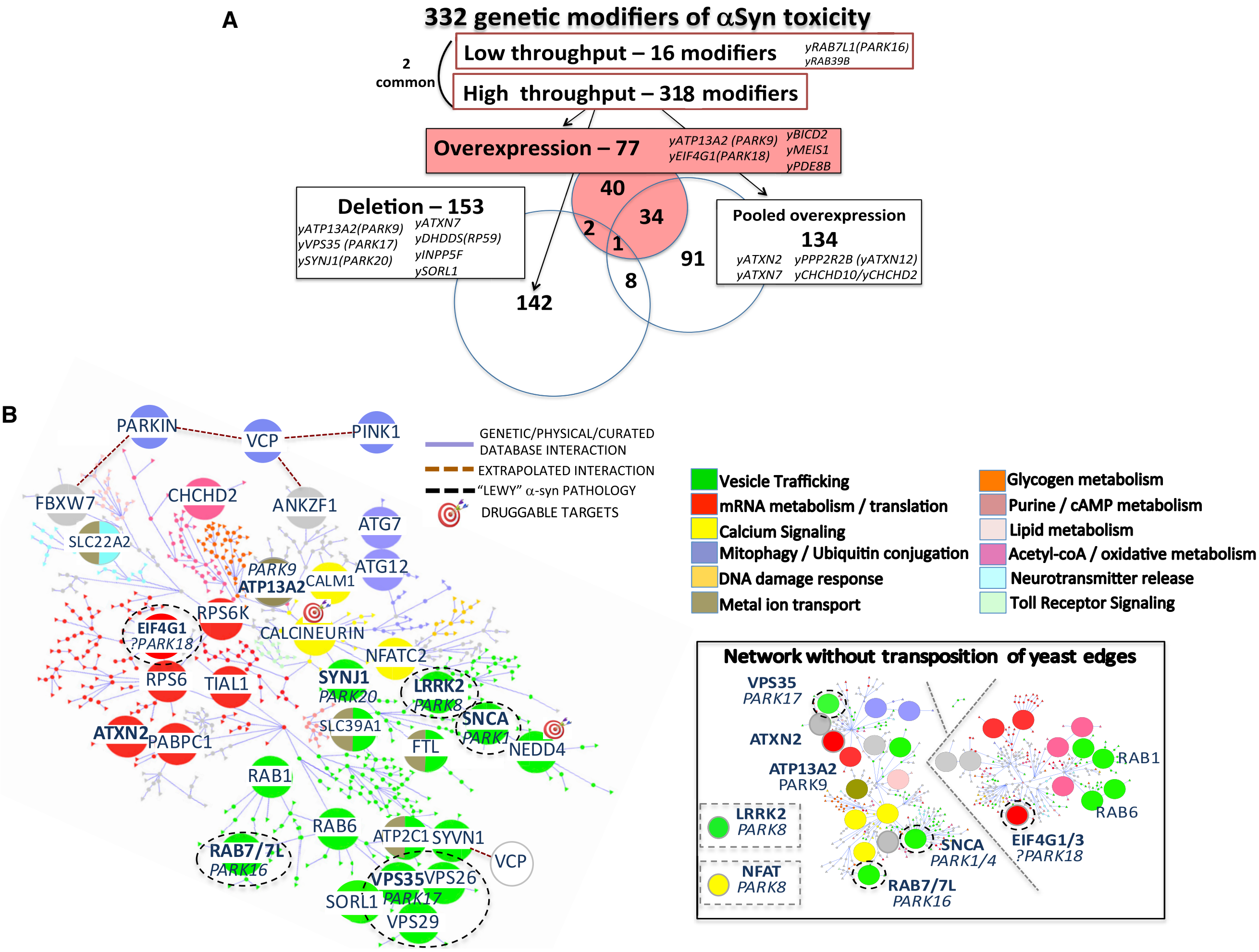


Figure 3. TransposeNet Builds Genome-Scale Molecular Network for α-Syn Toxicity from Genome-wide Deletion and Overexpression Yeast Screens

(A) Summary of genetic modifiers recovered in screens. 16 genetic modifiers (14 unique) from low-throughput investigations were also incorporated. Yeast homologs of genes linked to PD and other neurodegenerative disorders are listed. A "y" preceding the human gene name indicates "the yeast homolog of".

(B) A humanized network is generated from the 332 α-syn screen hits by TransposeNet. Genes of interest are enlarged, including multiple neurodegeneration-related disease genes (see also Figure S8 and Table S14). Gene ontology process enrichment within stems of the network are color-coded (full details in Table S12; gray portions were not enriched). Brown lines indicate extrapolated connections to VCP/Cdc48 through Vms1 (the yeast homolog of Ankzf1) and Hrd1 (the yeast homolog of Syvn1) and from VCP to Parkin/PARK2 and Pink1/PARK6. A target symbol marks two druggable nodes, Calcineurin (Caraveo et al., 2014) and Nedd4 (a target of NAB; Tardiff et al., 2013). Inset: network without transposition of yeast edges. *LRRK2* and *NFAT* are not included. Ontologically connected proteins (for example, Rab proteins) are dispersed.

ensemble of Steiner forests generated from our list of 332 modifiers, α-syn appeared in 100% and LRRK2 in 70%. In 1,000 random sets of 332 genes, even when we forced the incorporation of five *yPARK* genes recovered in our genetic experiments (*yPARK9*, *yPARK16*, *yPARK17*, *yPARK18*, and *yPARK20*), α-syn and LRRK2 appeared together in only 0.6% of humanized networks. Thus, yeast modifiers of α-syn toxicity generated a specific humanized network in which the PD-associated proteins α-syn and LRRK2 emerged as critical network nodes.

TransposeNet generated a coherent network: 295 out of 332 yeast modifier genes were included in a single tree network (Table S10) with biologically intuitive "stems" comprising genes of similar ontology (Figure 3B). Networks generated from these 332 modifiers without transposition of yeast interactome data produced three fragmented networks comprising 136, 2, and 122 yeast genes, respectively (Figure 3B, inset; Data S1). Genes that should be related biologically through involvement in common cellular processes were separated (Figure 3B). Moreover, LRRK2 and NFAT were not incorporated. Testable hypotheses, such as the relationship of EIF4G1 to ATXN2 (Figure 5), did not emerge because these proteins were not included in the same network. DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009) also produced highly fragmented or dense "hairball" networks of limited utility for hypothesis generation (Figure S9) and these algorithms did not incorporate critical nodes like LRRK2 (Figure S9). Thus, transposition of yeast networks to augment the human

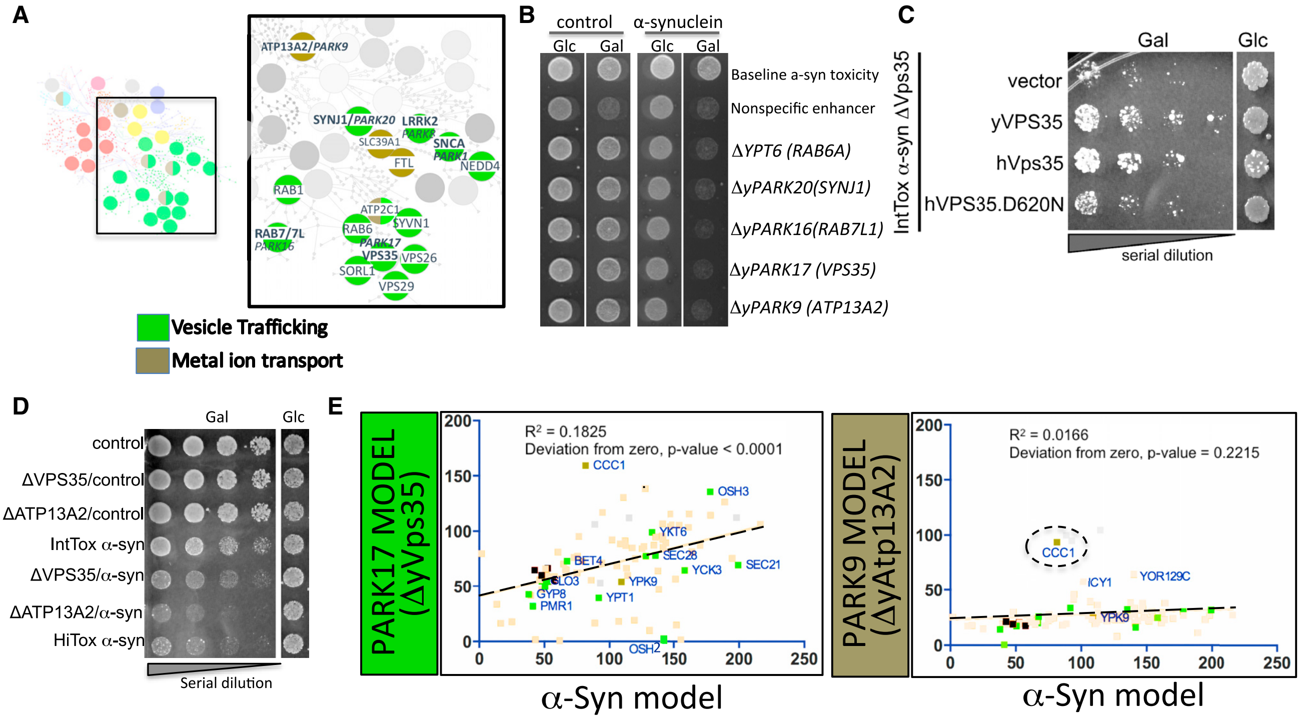


Figure 4. Genetic Dissection of Parkinsonism Susceptibility Genes Reveals Distinct Biology

(A) Vesicle trafficking subnetwork within the α-syn map (from Figure 3B) and location of *PARK9* (*ATP13A2*) outside of this subnetwork. Green, trafficking proteins; brown, metal ion transporters.

(B) Synthetic toxic interactions between trafficking genes and α-syn (spotting assays on agar plates). αsyn transgene is expressed from a galactose-inducible promoter ("on" in galactose [Gal]; "off" in glucose [Glc]). A "y" ahead of the human gene name indicates "the yeast homolog of". Δ*GAL2* strain ("nonspecific enhancer") serves as a (+) control because it grows less well on galactose (regardless of α-syn expression). The (−) control, a deletion (ΔYMR191W), has no deleterious effect in presence of α-syn ("baseline toxicity").

(C) Expression of yeast VPS35 (yVps35), human VPS35 (hVps35), and human mutant (D620N) VPS35 in α-syn-expressing Vps35-deleted "IntTox" cells (yeast spotting assay, showing serial 5× dilutions from left to right; transgenes are expressed from a galactose-inducible promoter).

(D and E) Cross-comparison of genetic interactors with similarly toxic HiTox α-syn, α-syn-ΔVPS35/PARK17, and α-syn-ΔATP13A2/PARK9 strains. Spotting assay demonstrates relative levels of toxicity among these three strains (D; 5× serial dilutions from left to right). In (E), data are shown on dot plots comparing the efficacy of 77 known α-syn modifiers (see Figure 1) in HiTox α-syn (x axis) versus ΔPARK17/α-syn (y axis; left) ΔPARK9/α-syn (y axis; right). Green, vesicle trafficking genetic modifiers; brown, metal ion transport modifiers. Axis scales represent growth relative to Mig1/Mig3-positive controls (100, black). Mig1/Mig3 overexpression represses the galactose promoter driving α-syn expression.

Each spot assay in this figure was repeated two to four times. The dot plot is representative of two experiments performed on separate days with biological replicates. Transformants were plated in quadruplicate for each experiment.

interactome created a coherent, biologically meaningful α-syn network.

An Endocytic and Retrograde Trafficking Subnetwork in the α-Syn Toxicity Map Incorporates Yeast Homologs of *RAB7L1* (*PARK16*) and *VPS35* (*PARK17*)

In the α-syn map, homologs of some parkinsonism genes coalesced in a sub-network around YPT6, the yeast homolog of RAB6A (Soper et al., 2011) (Figure 3B). Included were *YPT7*, *VTH1*, and *VPS35*, which encode proteins involved in endosomal trafficking. *YPT7* is a close homolog of *RAB7L1*, a leading candidate for the *PARK16* locus (MacLeod et al., 2013; Nalls et al., 2014), and also of the Mendelian parkinsonism risk factor *RAB39B* (Wilson et al., 2014). *VTH1* is a close yeast homolog of *SORL1*, an established AD risk modifier (Rogaeva et al., 2007) that encodes a protein involved in intracellular sorting (Nykjaer and Willnow, 2012). *VPS35* is homologous to the Mendelian risk factor for classic PD, *VPS35* (*PARK17*) (Zimprich et al., 2011). *VPS35*, *VPS26*, and *VPS29* comprise the retromer complex that transports cargo from endosomal to Golgi compartments. In our accompanying study (Chung et al., 2017), we show that deletion of the VSP26 and VPS29 core retromer components strongly enhances α-syn toxicity. A fourth gene involved in Golgi-to-endosome and endocytic trafficking, *INP53*, is homologous to the Mendelian parkinsonism gene *SYNJ1* (*PARK20*) (Olgiati et al., 2014). Deletion of any one of these genes was not toxic in a wild-type strain. However, deletion of any one of these genes in a strain expressing low (nontoxic) levels of α-syn produced a strong and synergistic growth defect (Table S5; Figures 4B and S10A). Importantly, ectopic expression of yeast or human *VPS35* rescued the toxicity induced by deleting *VPS35*, but expression of a disease-causing mutation (*VPS35-D620N*) did not (Figure 4C). Finally, yRAB7L1 enhanced α-syn toxicity when deleted but rescued from this toxicity when overexpressed (Figure S10B).

The α-Syn Map Predicts Diverging Genetic Interaction Profiles in Δ*PARK9* (*ATP13A2*)- and Δ*PARK17* (*VPS35*)-Sensitized Yeast Models

To test functional consequences of being located in distinct subnetworks of our α-syn map, we compared *VPS35* (*PARK17*) and *ATP13A2* (*PARK9*). ATP13A2 is a type 5 P-ATPase implicated in cation transport and metal ion homeostasis (Kong et al., 2014; Park et al., 2014; Ramonet et al., 2012; Tsunemi and Krainc, 2014). Mutations in *ATP13A2* lead to juvenile-onset parkinsonism or Kufor-Rakeb syndrome, which is distinct from PD (Schneider et al., 2010). *yATP13A2* suppressed α-syn toxicity in our overexpression screen (Figure 1B), and deletion of *yATP13A2* strongly enhanced α-syn toxicity (Figure 4B). In our humanized network, *ATP13A2* was spatially distant from *VPS35*, lying well outside the vesicle trafficking subnetwork (Figures 3B and 4A). We asked whether this spatial separation reflected differences in underlying biology.

We generated three strains with similar toxicities (Figure 4D). In one strain, toxicity resulted from overexpression of α-syn (HiTox). In two other strains, mild toxicity induced by intermediate levels of α-syn expression was enhanced by deletion of yeast *ATP13A2* (hereafter, ΔATP13A2/α-syn) or *VPS35* (hereafter, ΔVPS35/α-syn). These three yeast strains thus modeled cellular pathologies related to three forms of familial parkinsonism: two with typical α-syn pathology (PD related to α-syn multiplication and *VPS35* (*PARK17*)-associated parkinsonism) and one with strikingly different pathology (*PARK9* (*ATP13A2*)).

While ΔATP13A2 is known to sensitize cells to metal ion stress (Kong et al., 2014), ΔVPS35 strains exhibit retrograde trafficking defects (Seaman et al., 1997), suggesting that ΔATP13A2 and ΔVPS35 strains are differentially sensitized to α-syn toxicity. We asked whether our 77 α-syn overexpression screen hits differentially modified the toxicity of our ΔVPS35/α-syn and ΔATP13A2/α-syn models.

We expressed these α-syn toxicity modifiers in each of the yeast models and monitored growth. For the α-syn HiTox and ΔVPS35/α-syn models, 69 out of 77 genes overlapped (Figure 4E, left), correlating well with the similar pathology associated with these genetic forms of parkinsonism. Notably, the overlapping modifiers were enriched in vesicle trafficking genes (Table S13). In contrast, there were only 3 out of 77 modifiers in common between α-syn HiTox and ΔATP13A2/α-syn (Figure 4E, right). These were involved in iron and manganese homeostasis (CCC1) and actin cytoskeleton rearrangements (ICY1 and AFI1 (YOR129C)), respectively. Notably, metal ion homeostasis is strongly implicated in Kufor-Rakeb syndrome (Schneider et al., 2010) and its mammalian models (Park et al., 2014). Thus, neurodegenerative diseases that are genetically, clinically, and neuropathologically distinct may nonetheless share some key molecular pathologies that can be uncovered through genetic network analysis.

mRNA Translation Subnetwork Links α-Syn to PABPC1, EIF4G1, and ATXN2

In our overexpression screen against α-syn toxicity, *TIF4632* (hereafter *yEIF4G1-2*) emerged as a suppressor. *yEIF4G1-2* is a yeast homolog of the translational initiation factor *EIF4G1.* The genome-wide deletion and pooled overexpression screens identified additional genetic modifiers related to mRNA translation, including initiation factors and multiple ribosomal subunits (Figures 3B and 5A; Tables S5 and S6). These included the homolog of human PABPC1, PAB1, the *ATXN2* homolog *PBP1*, and the second *EIF4G* family homolog in yeast, *TIF4631* (hereafter *yEIF4G1-1*). Enrichment of these hits in the pooled screen was confirmed by qPCR (Figure 5B, left), and overexpression of these genes suppressed α-syn toxicity in bioscreen (Figure 5B, right) and/or spot (Figure S11) growth assays. Genetic experiments in different proteinopathy models revealed that the effects of these modifiers on α-syn toxicity were specific (Figure S11). Thus, perturbation of mRNA translation in synucleinopathy was not simply a generic proteotoxic response.

Protein Translation Is Perturbed in PD-Patient-Derived Neurons

Because we recovered numerous genetic modifiers in the mRNA translation and mRNA processing pathways (Figures 3 and 5), we asked whether protein translation was perturbed in cellular synucleinopathy models, including PD-patient-derived neurons. Bulk changes in protein production were assessed by determining the rate at which S^{35}-radiolabeled methionine and cysteine are incorporated into protein over time after a brief "pulse labeling." Overexpression of α-syn in HEK cells and primary rat cortical neurons reduced the accumulation of S^{35}-Met/Cys without changing amino acid uptake (Figure S12). Similarly, in 6- to 8-week-old iPSC neurons harboring the α-syn^{A53T} mutation, S^{35}-Met/Cys incorporation into protein was reduced compared to isogenic mutation-corrected controls (Figure 5C). Thus, our α-syn screens and network analysis identified a strong effect of α-syn toxicity on bulk mRNA translation in cellular models of synucleinopathy. This effect was not attributable to a canonical ER stress response, because phosphorylation of EIF2A (Figures S12D and S13A) and XBP1 splicing (Figure S13B) was not altered in these neurons.

Conserved Genetic Interactions of *ATXN2 and EIF4G1* from Yeast to Patient Neurons

We next tested whether human homologs of two translation factors that suppressed α-syn toxicity when overexpressed in yeast, *ATXN2* and *EIF4G1*, could similarly reverse the protein translation defect in neurons. We generated transcription activator-like effector transcription factor (TALE-TF) constructs to transcriptionally upregulate neuronal isoforms of *EIF4G1* and *ATXN2* (Sanjana et al., 2012) (Figure 5D, left). These constructs were then delivered with an adeno-associated viral vector to differentiated α-syn^{A53T} iPSC-derived neuronal cultures.

10 days after transduction, endogenous *EIF4G1* and *ATXN2* mRNA levels increased by ~4-fold, as measured by qPCR (Figure 5D, right). This increased expression substantially reversed the defect in bulk protein translation we had observed in α-syn^{A53T} neurons (Figure 5E). Overexpression of *EIF4G1* increased translation in A53T neurons, but not in mutation-corrected controls. *ATXN2* overexpression equally increased bulk translation in mutant and control cells (Figure 5E).

Thus, our cross-species molecular network predicted a biological interaction between α-syn and mRNA translation factors in PD-patient-derived neurons. These data strengthen the argument that perturbed mRNA translation is an important aspect of α-syn toxicity. Interestingly, we identified a strong signature of

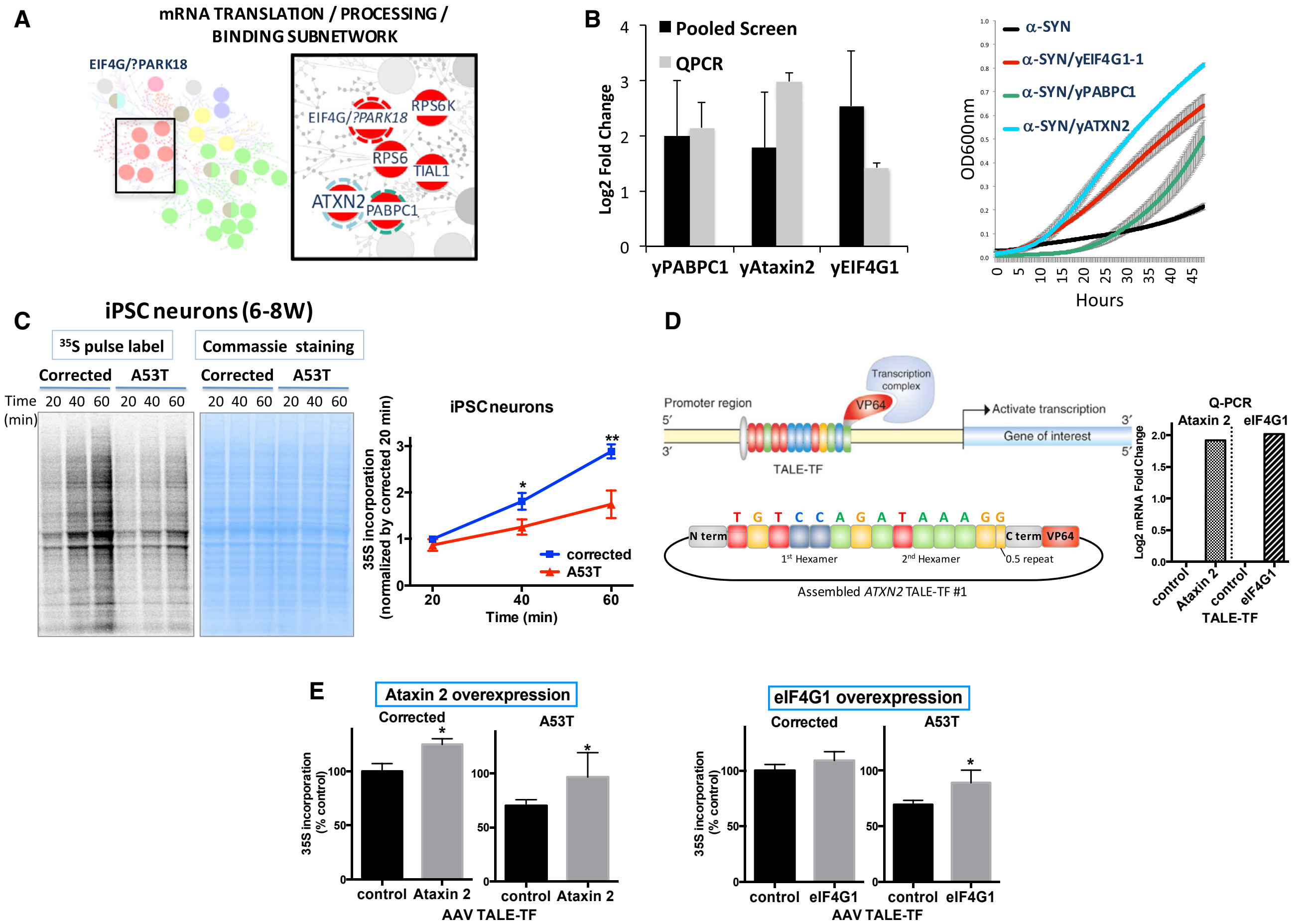


Figure 5. mRNA Translation Factors Impact α-Syn Toxicity from Yeast to Patient-Derived Neurons

(A) mRNA translation subnetwork in α-syn toxicity (from Figure 3B), including *ATXN2*, *EIF4G1* (*PARK18*), and *PABPC1*.

(B) Effects of yPABPC1, yAtaxin2, and yEIF4G1-1 on α-syn toxicity (left: qPCR to verify enrichment of these genes in pooled over-expression screen; right: bioscreen growth assay). A "y" ahead of the human gene name indicates "yeast homolog of".

(C) Bulk mRNA translation in mutant α-syn^{A53T} iPSC neurons compared to isogenic mutation-corrected control neurons as measured by ^{35}S-cysteine and ^{35}S-methionine incorporation over time (phosphorimager scan). Coomassie staining shows loading of protein. At right, the graph shows two subclones of the mutation-corrected line compared to α-syn^{A53T} cells (n = 4; *p < 0.05; **p < 0.01 (two-tailed t test).

(D) TALE-TFs were designed to elevate the endogenous levels of *ATXN2* or *EIF4G1* genes. These bind to the 5′ UTR of the target genes and recruit a transcriptional activator (Sanjana et al., 2012). qPCR indicates transcript levels after AAV-mediated TALE-TF delivery into A53T iPSC neurons.

(E) Effect of increasing endogenous *EIF4G1* or *ATXN2* levels on bulk translation in A53T neurons (n = 3).

Data represent mean ± SEM. *p < 0.05 (two-tailed t test).

decreased translation of mRNA translation-related transcripts in ribosomal footprinting experiments of α-syn^{A53T} iPSC-derived cortical neurons at 4 and 12 weeks of differentiation compared to isogenic mutation-corrected control neurons (Figures S14 and S15; Tables S16, S17, S18, and S19). Indeed, mRNA-related translation factors, ribonucleoproteins, and ribosomal proteins were enriched not only in our genetic and translational maps of α-syn toxicity but also among proteins in the immediate vicinity of α-syn in neurons (Figure 4 in accompanying manuscript, Chung et al., 2017; Figure S14A, "spatial α-syn map"). Moreover, a number of mRNA translation proteins directly complexed with α-syn. This convergence of genetic, translational, and spatial maps suggests the connection between α-syn toxicity and mRNA metabolism is deeply rooted in α-syn biology.

DISCUSSION

We describe a coherent, systems-level analysis of how α-syn misfolding and mistrafficking perturbs cell biology. Genome-wide screens identified modifiers of the toxic consequences of α-syn expression in baker's yeast (*S. cerevisiae*). Our key computational contribution, TransposeNet, coupled richly annotated molecular interactions in yeast with a Steiner prize-collecting algorithm and a sophisticated cross-species homology tool to visualize the screen hits as a humanized molecular network. TransposeNet revealed that α-syn pathology is deeply connected to human genetic risk factors for parkinsonism and parsed out the molecular pathways through which these connections occur. We envisage TransposeNet as a valuable resource for the community, easily generalizable to the modeling of any physiologic or pathologic process in genetically tractable organisms.

A pressing challenge in neurodegeneration is to determine whether genes associated with highly distinct pathologies, but that nevertheless converge on similar clinical phenotypes, are related at a molecular level or not. Our network tied α-syn not only to genes that cause classical PD (Ogaki et al., 2015) but also to genes that cause parkinsonism with different pathologies and genes associated with other neurodegenerative phenotypes altogether (Table S14). The relationships were highly specific to α-syn. Moreover, genes tied to classical PD or α-syn pathology (like RAB7L1, VPS35, and LRRK2) were concentrated in a vesicle-trafficking-associated subnetwork, while genes tied to "atypical" parkinsonism (like ATP13A2 and ATXN2) were in separate subnetworks. For a few examples, we experimentally validated the convergent and divergent molecular pathologies indicated by the spatial location on the map. Thus, LRRK2 and α-syn pathologies were connected at the level of perturbed protein trafficking, which was confirmed in patient-derived neurons. In another example, *VPS35* and *ATP13A2* exhibited highly distinct genetic modifier profiles in yeast. Other network and model-organism studies provide important support for our results, including connections between α-syn and LRRK2 (Cho et al., 2014; Liu et al., 2012), RAB7L1 and LRRK2 (MacLeod et al., 2013; Beilina et al., 2014), and VPS35 and α-syn (Dhungel et al., 2015).

For some genes connected to α-syn toxicity by our network, including *EIF4G1* (*PARK18*) and *CHCHD2*, human genetic data are limited or controversial (Funayama et al., 2015; Liu et al., 2015; Ogaki et al., 2015; Chartier-Harlin et al., 2011; Nuytemans et al., 2013). Another gene, *RAB7L1*, is one of two candidates in linkage with a parkinsonism-associated common gene variant (*PARK16*). Our analysis affirms a connection between such genes and α-syn proteinopathy and provides the biological context in which to place these interactions. For example, we make no claim that the translation factor *EIF4G1* should be designated a "PD gene." However, EIF4G1 and α-syn toxicity are connected in the context of an important and previously unrecognized direct effect of α-syn on mRNA biology and protein translation. This was confirmed by multiple hits in our genetic analysis (Figure 3) and in our mRNA translational profiling of neurons (Figure S14). Moreover, a connection to mRNA translation and metabolism was also confirmed in our spatial mapping of α-syn in neurons (Chung et al., 2017). This map revealed that α-syn is in the immediate vicinity and complexed to proteins involved in mRNA translation and protein trafficking, suggesting that these perturbations may be upstream or proximal events in α-syn toxicity. Interestingly, a connection is emerging between mRNA translation and other parkinsonism-related genes (Gehrke et al., 2015; Martin et al., 2014).

Finally, by identifying connections between druggable targets and gene networks, our approach provides a glimpse of how future treatments might be targeted to specific genetic lesions. We envisage that the inflexibility of a single clinical or pathologic diagnosis will yield to a more nuanced molecular diagnosis. In this scenario, genetic lesions will be matched to compound targets and confirmed in "personalized" cellular models in which combinatorial genetic lesions are introduced to reflect specific genetic risk and biology. Emerging genome-editing technologies will enable such models to be developed in patient-derived cells, and for genome-wide screening to be carried out in these cells also (Hasson et al., 2013; Khurana et al., 2015; Shalem et al., 2014; Wang et al., 2014). These will unquestionably be welcome advances, but impressive developments will continue in simple model organisms. Variomic libraries in yeast, for example, now enable mutagenesis at single-amino-acid resolution across the entire yeast proteome (Huang et al., 2013), unlocking enormous potential for target identification in phenotypic screens. We envision multi-faceted, cross-species approaches will continue to reveal critical insights into many complex diseases and perhaps fulfill therapeutic promises in the post-genomics era.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Yeast Strains
  - Human iPSC Lines
  - Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines
  - Human Pluripotent Stem Cell Culture for α-Syn Mutant Lines
  - Primary Rat Cortical Cultures
- METHOD DETAILS
  - Yeast-to-Human Homology
  - Preprocessing of Interactomes
  - Augmentation of Human Interactome with Yeast-to-Yeast Edges (for Humanized Networks Only)
  - Prize-Collecting Steiner Forest Algorithm
  - Node and Edge Setup for Yeast and Humanized Steiner Networks
  - Parameter Tuning for Computational Pipelines
  - Spotting Assays
  - Screening against Known α-Syn Modifiers in ΔPARK17/α-Syn and ΔPARK9/α-Syn Strains
  - Small Molecule (NAB2) Treatment
  - Pooled α-Syn Overexpression Screen
  - Pooled Screen-QPCR Verification
  - Pooled Screen-Growth Curve Analysis
  - Genome-Wide Deletion Screen (Synthetic Gene Array Methodology)
  - Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines
  - Human Pluripotent Stem Cell Culture for α-Syn Mutant Lines
  - Human Neural Induction by Embryoid Body (EB) Formation
  - Human Cortical Neural Differentiation
  - Primary Rat Cortical Cultures
  - AAV-1 Transduction of iPS Neurons
  - Antibodies
  - Protein Labeling with ^{35}S-Methionine/-Cysteine

- Free ^{35}S-Methionine/-Cysteine in the Cytosol
- Cell Lysis and Endoglycosidase H Digestion
- Western Blotting
- TALE-TF Design
- TALE-TF Assembly
- Ribosomal Footprint Profiling
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Comparison with Existing Homology Prediction Approaches
  - Evaluation of PCSF and Humanized Steiner Networks
  - Statistical Methods and Data Analysis for Cell-Based Assays
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes 15 figures, 19 tables, and 1 data file and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.12.011.

AUTHOR CONTRIBUTIONS

The study was designed and supervised by V.K. and S.L. TransposeNet was conceived by V.K. and executed by J.P., with key contributions by M.K. and N.T., B.B., and E.F, and supervision by B.B. and E.F. C.Y.C. led the neuronal assay development efforts. P.K.A., S.F., D.F.T., T.B., S.W.E, H.B., Y.L., A.N.-U., V.B., Y.F., M.C., B.-J.S.L., D.C.S., M.I.B., S.E., N.S., Q.Z., T.G., D.P.B., M.V., M.D., and C.B. performed bench or computational experiments or supervised these efforts. The paper was written by V.K., C.Y.C., J.P., B.B., E.F., and S.L.

ACKNOWLEDGMENTS

Research was supported by an HHMI Collaborative Innovation Award (V.K., C.Y.C., and S.L.), the JPB Foundation (V.K., C.Y.C., and S.L.), an American Brain Foundation and Parkinson's Disease Foundation Clinician-Scientist Development Award (V.K.), the Multiple System Atrophy Coalition (V.K.), the NIH/NIA (grant K01 AG038546, C.Y.C.), the Eleanor Schwartz Charitable Foundation (S.L.), the NIH (grants R01 NS 089076 and U54 NS 091046, E.F.; R01GM081871, B.B.; HG006061, M.V. and S.L.; and P50HG004233 and U01HG001715, M.V.). We thank Drs. Prathapan Thiru and George Bell, Bioinformatics and Research Computing (BaRC; Whitehead Institute) for expert assistance; Kristina Fischer for important technical assistance; Drs. Luke Whitesell, Philip de Jager, Mel Feany, Aftabul Haque, David Hill, Priyanka Narayan, and Linda Clayton for stimulating discussion and critical comments on the manuscript; and Dr. Julie Valastyan for yeast strain construction. We dedicate this manuscript to Dr Susan Lindquist, our incomparable mentor and colleague, who passed away while this manuscript was in its final stages of revision.

V.K., C.Y.C., D.F.T., and S.L. are scientific co-founders of Yumanity Therapeutics, a company developing neurodegenerative disease therapeutics.

Received: January 12, 2016
Revised: August 5, 2016
Accepted: December 14, 2016
Published: January 25, 2017

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Bailly-Bechet, M., Borgs, C., Braunstein, A., Chayes, J., Dagkessamanskaia, A., François, J.-M., and Zecchina, R. (2011). Finding undetected protein associations in cell signaling by belief propagation. Proc. Natl. Acad. Sci. USA *108*, 882–887.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat. Methods *7*, 1017–1024.

Beilina, A., Rudenko, I.N., Kaganovich, A., Civiero, L., Chau, H., Kalia, S.K., Kalia, L.V., Lobbestael, E., Chia, R., Ndukwe, K., et al.; International Parkinson's Disease Genomics Consortium; North American Brain Expression Consortium (2014). Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. Proc. Natl. Acad. Sci. USA *111*, 2626–2631.

Berger, B., Peng, J., and Singh, M. (2013). Computational solutions for omics data. Nat. Rev. Genet. *14*, 333–346.

Brás, J., Guerreiro, R., and Hardy, J. (2015). SnapShot: Genetics of Parkinson's disease. Cell *160*, 570–570.e1.

Caraveo, G., Auluck, P.K., Whitesell, L., Chung, C.Y., Baru, V., Mosharov, E.V., Yan, X., Ben-Johny, M., Soste, M., Picotti, P., et al. (2014). Calcineurin determines toxic versus beneficial responses to α-synuclein. Proc. Natl. Acad. Sci. USA *111*, E3544–E3552.

Casals, F., and Bertranpetit, J. (2012). Genetics. Human genetic variation, shared and private. Science *337*, 39–40.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Chartier-Harlin, M.-C., Dachsel, J.C., Vilariño-Güell, C., Lincoln, S.J., Leprêtre, F., Hulihan, M.M., Kachergus, J., Milnerwood, A.J., Tapia, L., Song, M.-S., et al. (2011). Translation initiator EIF4G1 mutations in familial Parkinson disease. Am. J. Hum. Genet. *89*, 398–406.

Cho, H.J., Yu, J., Xie, C., Rudrabhatla, P., Chen, X., Wu, J., Parisiadou, L., Liu, G., Sun, L., Ma, B., et al. (2014). Leucine-rich repeat kinase 2 regulates Sec16A at ER exit sites to allow ER-Golgi export. EMBO J. *33*, 2314–2331.

Cho, H., Peng, J., and Berger, B. (2015). Diffusion Component Analysis: Unraveling Functional Topology in Biological Networks (RECOMB).

Cho, H., Berger, B., and Peng, J. (2016). Compact integration of multi-network topology for functional analysis of genes. Cell Syst. Published online November 22, 2016. http://dx.doi.org/10.1016/j.cels.2016.10.017.

Chung, C.Y., Khurana, V., Auluck, P.K., Tardiff, D.F., Mazzulli, J.R., Soldner, F., Baru, V., Lou, Y., Freyzon, Y., Cho, S., et al. (2013). Identification and rescue of α-synuclein toxicity in Parkinson patient-derived neurons. Science *342*, 983–987.

Chung, C.Y., Khurana, V., Yi, S., Sahni, N., Loh, K.H., Auluck, P.K., Baru, V., Udeshi, N.D., Freyzon, Y., Carr, S.A., et al. (2017). In situ peroxidase labeling and mass-spectrometry connects alpha-synuclein directly to endocytic trafficking and mRNA metabolism in neurons. Cell Syst. *4*, 242–250.

Cooper, A.A., Gitler, A.D., Cashikar, A., Haynes, C.M., Hill, K.J., Bhullar, B., Liu, K., Xu, K., Strathearn, K.E., Liu, F., et al. (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. Science *313*, 324–328.

Dhungel, N., Eleuteri, S., Li, L.-B., Kramer, N.J., Chartron, J.W., Spencer, B., Kosberg, K., Fields, J.A., Stafa, K., Adame, A., et al. (2015). Parkinson's disease genes VPS35 and EIF4G1 interact genetically and converge on α-synuclein. Neuron *85*, 76–87.

Elden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature *466*, 1069–1075.

Fuchs, J., Tichopad, A., Golub, Y., Munz, M., Schweitzer, K.J., Wolf, B., Berg, D., Mueller, J.C., and Gasser, T. (2008). Genetic variability in the SNCA gene influences alpha-synuclein levels in the blood and brain. FASEB J. *22*, 1327–1334.

Funayama, M., Ohe, K., Amo, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., Yoshino, H., et al. (2015). CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. Lancet Neurol. *14*, 274–282.

Geetha, V., Di Francesco, V., Garnier, J., and Munson, P.J. (1999). Comparing protein sequence-based and predicted secondary structure-based methods for identification of remote homologs. Protein Eng. *12*, 527–534.

Gehrke, S., Wu, Z., Klinkenberg, M., Sun, Y., Auburger, G., Guo, S., and Lu, B. (2015). PINK1 and Parkin control localized translation of respiratory chain component mRNAs on mitochondria outer membrane. Cell Metab. *21*, 95–108.

Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. *20*, 1425.

Gietz, R.D., Schiestl, R.H., Willems, A.R., and Woods, R.A. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast *11*, 355–360.

GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

Guerreiro, R., Brás, J., and Hardy, J. (2015). SnapShot: Genetics of ALS and FTD. Cell *160*, 798.e1.

Hasson, S.A., Kane, L.A., Yamano, K., Huang, C.-H., Sliter, D.A., Buehler, E., Wang, C., Heman-Ackah, S.M., Hessa, T., Guha, R., et al. (2013). High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. Nature *504*, 291–295.

Hu, Y., Rolfs, A., Bhullar, B., Murthy, T.V.S., Zhu, C., Berger, M.F., Camargo, A.A., Kelley, F., McCarron, S., Jepson, D., et al. (2007). Approaching a complete repository of sequence-verified protein-encoding clones for Saccharomyces cerevisiae. Genome Res. *17*, 536–543.

Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics *12*, 357.

Huang, S.-S.C., and Fraenkel, E. (2009). Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. Sci. Signal. *2*, ra40.

Huang, Z., Chen, K., Zhang, J., Li, Y., Wang, H., Cui, D., Tang, J., Liu, Y., Shi, X., Li, W., et al. (2013). A functional variomics tool for discovering drug-resistance genes and drug targets. Cell Rep. *3*, 577–585.

Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., and Marcotte, E.M. (2015). Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science *348*, 921–925.

Khurana, V., and Lindquist, S. (2010). Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? Nat. Rev. Neurosci. *11*, 436–449.

Khurana, V., Tardiff, D.F., Chung, C.Y., and Lindquist, S. (2015). Toward stem cell-based phenotypic screens for neurodegenerative diseases. Nat. Rev. Neurol. *11*, 339–350.

Kim, J.-E., O'Sullivan, M.L., Sanchez, C.A., Hwang, M., Israel, M.A., Brennand, K., Deerinck, T.J., Goldstein, L.S.B., Gage, F.H., Ellisman, M.H., and Ghosh, A. (2011). Investigating synapse formation and function using human pluripotent stem cell-derived neurons. Proc. Natl. Acad. Sci. USA *108*, 3005–3010.

Kim, H.-J., Raphael, A.R., LaDow, E.S., McGurk, L., Weber, R.A., Trojanowski, J.Q., Lee, V.M.-Y., Finkbeiner, S., Gitler, A.D., and Bonini, N.M. (2014). Therapeutic modulation of eIF2α phosphorylation rescues TDP-43 toxicity in amyotrophic lateral sclerosis disease models. Nat. Genet. *46*, 152–160.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Kong, S.M.Y., Chan, B.K.K., Park, J.-S., Hill, K.J., Aitken, J.B., Cottle, L., Farghaian, H., Cole, A.R., Lay, P.A., Sue, C.M., and Cooper, A.A. (2014). Parkinson's disease-linked human PARK9/ATP13A2 maintains zinc homeostasis and promotes α-Synuclein externalization via exosomes. Hum. Mol. Genet. *23*, 2816–2833.

Kriks, S., Shim, J.-W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., et al. (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature *480*, 547–551.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Liu, G., Aliaga, L., and Cai, H. (2012). α-synuclein, LRRK2 and their interplay in Parkinson's disease. Future Neurol. *7*, 145–153.

Liu, Z., Guo, J., Li, K., Qin, L., Kang, J., Shu, L., Zhang, Y., Wei, Y., Yang, N., Luo, Y., et al. (2015). Mutation analysis of CHCHD2 gene in Chinese familial Parkinson's disease. Neurobiol. Aging *36*, 3117.e7–3117.e8.

MacLeod, D.A., Rhinn, H., Kuwahara, T., Zolin, A., Di Paolo, G., McCabe, B.D., Marder, K.S., Honig, L.S., Clark, L.N., Small, S.A., and Abeliovich, A. (2013). RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. Neuron *77*, 425–439.

Martin, I., Dawson, V.L., and Dawson, T.M. (2011). Recent advances in the genetics of Parkinson's disease. Annu. Rev. Genomics Hum. Genet. *12*, 301–325.

Martin, I., Kim, J.W., Lee, B.D., Kang, H.C., Xu, J.-C., Jia, H., Stankowski, J., Kim, M.-S., Zhong, J., Kumar, M., et al. (2014). Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. Cell *157*, 472–485.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al.; International Parkinson's Disease Genomics Consortium (IPDGC); Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI); 23andMe; GenePD; NeuroGenetics Research Consortium (NGRC); Hussman Institute of Human Genomics (HIHG); Ashkenazi Jewish Dataset Investigator; Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE); North American Brain Expression Consortium (NABEC); United Kingdom Brain Expression Consortium (UKBEC); Greek Parkinson's Disease Consortium; Alzheimer Genetic Analysis Group (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. *46*, 989–993.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.-A., Fraser, D., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science *337*, 100–104.

Nuytemans, K., Bademci, G., Inchausti, V., Dressen, A., Kinnamon, D.D., Mehta, A., Wang, L., Züchner, S., Beecham, G.W., Martin, E.R., et al. (2013). Whole exome sequencing of rare variants in EIF4G1 and VPS35 in Parkinson disease. Neurology *80*, 982–989.

Nykjaer, A., and Willnow, T.E. (2012). Sortilin: a receptor to regulate neuronal viability and function. Trends Neurosci. *35*, 261–270.

Ogaki, K., Koga, S., Heckman, M.G., Fiesel, F.C., Ando, M., Labbé, C., Lorenzo-Betancor, O., Moussaud-Lamodière, E.L., Soto-Ortolaza, A.I., Walton, R.L., et al. (2015). Mitochondrial targeting sequence variants of the CHCHD2 gene are a risk for Lewy body disorders. Neurology *85*, 2016–2025.

Olgiati, S., De Rosa, A., Quadri, M., Criscuolo, C., Breedveld, G.J., Picillo, M., Pappatà, S., Quarantelli, M., Barone, P., De Michele, G., and Bonifati, V. (2014). PARK20 caused by SYNJ1 homozygous Arg258Gln mutation in a new Italian family. Neurogenetics *15*, 183–188.

Osborn, M.J., and Miller, J.R. (2007). Rescuing yeast mutants with human genes. Brief. Funct. Genomics Proteomics *6*, 104–111.

Outeiro, T.F., and Lindquist, S. (2003). Yeast cells provide insight into alpha-synuclein biology and pathobiology. Science *302*, 1772–1775.

Park, J.-S., Koentjoro, B., Veivers, D., Mackay-Sim, A., and Sue, C.M. (2014). Parkinson's disease-associated human ATP13A2 (PARK9) deficiency causes zinc dyshomeostasis and mitochondrial dysfunction. Hum. Mol. Genet. *23*, 2802–2815.

Rajput, A., Dickson, D.W., Robinson, C.A., Ross, O.A., Dächsel, J.C., Lincoln, S.J., Cobb, S.A., Rajput, M.L., and Farrer, M.J. (2006). Parkinsonism, Lrrk2 G2019S, and tau neuropathology. Neurology *67*, 1506–1508.

Ramonet, D., Podhajska, A., Stafa, K., Sonnay, S., Trancikova, A., Tsika, E., Pletnikova, O., Troncoso, J.C., Glauser, L., and Moore, D.J. (2012). PARK9-associated ATP13A2 localizes to intracellular acidic vesicles and regulates cation homeostasis and neuronal integrity. Hum. Mol. Genet. *21*, 1725–1743.

Reinhardt, P., Schmid, B., Burbulla, L.F., Schöndorf, D.C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S.A., Dhingra, A., et al. (2013). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. Cell Stem Cell *12*, 354–367.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat. Genet. *39*, 168–177.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212–1226.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., and Daly, M.J.; International Inflammatory Bowel Disease Genetics Constortium (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. *7*, e1001273.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. Nat. Protoc. *7*, 171–192.

Schneider, S.A., Paisan-Ruiz, C., Quinn, N.P., Lees, A.J., Houlden, H., Hardy, J., and Bhatia, K.P. (2010). ATP13A2 mutations (PARK9) cause neurodegeneration with brain iron accumulation. Mov. Disord. *25*, 979–984.

Schöndorf, D.C., Aureli, M., McAllister, F.E., Hindley, C.J., Mayer, F., Schmid, B., Sardi, S.P., Valsecchi, M., Hoffmann, S., Schwarz, L.K., et al. (2014). iPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. Nat. Commun. *5*, 4028.

Seaman, M.N., Marcusson, E.G., Cereghino, J.L., and Emr, S.D. (1997). Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. J. Cell Biol. *137*, 79–92.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shulman, J.M., De Jager, P.L., and Feany, M.B. (2011). Parkinson's disease: genetics and pathogenesis. Annu. Rev. Pathol. *6*, 193–222.

Singh, R., Xu, J., and Berger, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. Proc. Natl. Acad. Sci. USA *105*, 12763–12768.

Small, S.A., Kent, K., Pierce, A., Leung, C., Kang, M.S., Okada, H., Honig, L., Vonsattel, J.-P., and Kim, T.-W. (2005). Model-guided microarray implicates the retromer complex in Alzheimer's disease. Ann. Neurol. *58*, 909–919.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244-8.

Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318–331.

Soper, J.H., Kehm, V., Burd, C.G., Bankaitis, V.A., and Lee, V.M.-Y. (2011). Aggregation of α-synuclein in S. cerevisiae is associated with defects in endosomal trafficking and phospholipid biosynthesis. J. Mol. Neurosci. *43*, 391–405.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome Res. *21*, 2213–2223.

Tardiff, D.F., Jui, N.T., Khurana, V., Tambe, M.A., Thompson, M.L., Chung, C.Y., Kamadurai, H.B., Kim, H.T., Lancaster, A.K., Caldwell, K.A., et al. (2013). Yeast reveal a "druggable" Rsp5/Nedd4 network that ameliorates α-synuclein toxicity in neurons. Science *342*, 979–983.

Tardiff, D.F., Khurana, V., Chung, C.Y., and Lindquist, S. (2014). From yeast to patient neurons and back again: powerful new discovery platform. Mov. Disord. *29*, 1231–1240.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.

Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol. Biol. *313*, 171–192.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808–813.

Treusch, S., Hamamichi, S., Goodman, J.L., Matlack, K.E.S., Chung, C.Y., Baru, V., Shulman, J.M., Parrado, A., Bevis, B.J., Valastyan, J.S., et al. (2011). Functional links between Aβ toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. Science *334*, 1241–1245.

Tsunemi, T., and Krainc, D. (2014). Zn^{2+} dyshomeostasis caused by loss of ATP13A2/PARK9 leads to lysosomal dysfunction and alpha-synuclein accumulation. Hum. Mol. Genet. *23*, 2791–2801.

Tu, Z., Argmann, C., Wong, K.K., Mitnaul, L.J., Edwards, S., Sach, I.C., Zhu, J., and Schadt, E.E. (2009). Integrating siRNA and protein-protein interaction data to identify an expanded insulin signaling network. Genome Res. *19*, 1057–1067.

Tuncbag, N., Braunstein, A., Pagnani, A., Huang, S.-S.C., Chayes, J., Borgs, C., Zecchina, R., and Fraenkel, E. (2013). Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. J. Comput. Biol. *20*, 124–136.

Tuncbag, N., Gosline, S.J.C., Kedaigle, A., Soltis, A.R., Gitter, A., and Fraenkel, E. (2016). Network-based interpretation of diverse high-throughput datasets through the Omics Integrator software package. PLoS Comput. Biol. *12*, e1004879.

Verstraeten, A., Theuns, J., and Van Broeckhoven, C. (2015). Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. Trends Genet. *31*, 140–149.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327–335.

Voevodski, K., Teng, S.-H., and Xia, Y. (2009). Finding local communities in protein networks. BMC Bioinformatics *10*, 297.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Wilson, G.R., Sim, J.C.H., McLean, C., Giannandrea, M., Galea, C.A., Riseley, J.R., Stephenson, S.E.M., Fitzpatrick, E., Haas, S.A., Pope, K., et al. (2014). Mutations in RAB39B cause X-linked intellectual disability and early-onset Parkinson disease with α-synuclein pathology. Am. J. Hum. Genet. *95*, 729–735.

Yeger-Lotem, E., Riva, L., Su, L.J., Gitler, A.D., Cashikar, A.G., King, O.D., Auluck, P.K., Geddie, M.L., Valastyan, J.S., Karger, D.R., et al. (2009). Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nat. Genet. *41*, 316–323.

Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., et al. (2011). A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am. J. Hum. Genet. *89*, 168–175.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-Carboxypeptidase Y	Life Technologies	A66428
Rabbit anti-Nicastrin	Cell Signaling	3632; RRID: AB_2149581
phospho eIF2A	Cell Signaling	9721; RRID: AB_330951
total eIF2A	Cell Signaling	2103; RRID: AB_836874
LRRK2	Abcam	Ab133474
Experimental Models: Cell Lines		
Induced pluripotent stem cells (SNCA-A53T-1)	V.K. Lab, Chung et al. (2013)	N/A
Induced pluripotent stem cells (SNCA-A53T-CORR-1, SNCA-A53T-CORR-2)	V.K. Lab, Chung et al. (2013)	N/A
Induced pluripotent stem cells (LRRK2-G2019S-1, LRRK2-G2019S-2)	M.D. Lab, Reinhardt et al. (2013)	N/A
Induced pluripotent stem cells (LRRK2-G2019S-CORR-1, LRRK2-G2019S-CORR-2)	M.D. Lab, Reinhardt et al. (2013)	N/A
Experimental Models: Organisms/Strains		
Yeast strains: BY4741, deletion collection	C. Boone Lab, Tong and Boone (2006), Baryshnikova et al. (2010)	N/A
Yeast strain: W303 background, α-syn strains (NoTox), *MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1*	V.K. Lab	N/A
Yeast strain: W303 background, α-syn strain (IntTox), *MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1*	V.K. Lab	N/A
Yeast strain: W303 background, α-syn strain (HiTox), *MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1*	V.K. Lab	N/A
Yeast strain: W303 background, α-syn/delta-VPS35, W303 *MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1*	V.K. Lab	N/A
Yeast strain: W303 background, α-syn/delta-ATP13A2, *MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1*	V.K. Lab	N/A
Recombinant DNA		
TALE cloning backbone	V.K. Lab	N/A
TALE-TF (Ataxin-2)	V.K. Lab	N/A
TALE (EIF4G1)	V.K. Lab	N/A
SNCA expression construct collection (GFP-tagged, untagged)	V.K. Lab	N/A
Software and Algorithms		
TransposeNet Pipeline	web portal	http://transposenet.csail.mit.edu
DCA/Mashup	web portal	http://mashup.csail.mit.edu.
PCSF	web portal	http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/
Other		
Cytoscape Files	These are provided in the supplement (data files S20 through S30).	

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the Lead Contact, Vikram Khurana, by email at vkhurana@bwh.harvard.edu.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Yeast Strains

For the deletion screen, strains were in the BY4741 background and have been described in detail elsewhere (Baryshnikova et al., 2010; Tong and Boone, 2006).

For all experiments except the deletion screen and validation, the yeast strains used were in the w303 background (MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1). The vector control strain contained empty vector at the trp and ura loci (pAG304Gal, pAG306GAL). The NoTox α-syn strain contained α-syn fused to green fluorescent protein (α-syn-GFP) inserted at the his locus (pAG303Gal-α-syn-GFP). IntTox and HiTox α-syn strains contained multiple tandem copies of α-syn-GFP inserted at this and trp loci (pRS303GAL-α-syn-GFP, pRS304GAL-α-syn-GFP). IntTox strains have 4-5 copies of α-syn while HiTox cells have > 6 copies of α-syn. The ΔPARK17/α-syn and ΔPARK9/α-syn were generated by replacing the PARK17/VPS35 or PARK9/SPF1 gene loci in IntTox α-syn strains with a kanamycin resistance cassette (VPS35::kanMX or SPF1::kanMX).

Human iPSC Lines

iPSC from control individuals and PD patients carrying G2019S *LRRK2* along with isogenic gene-corrected controls were generated as previously described (Reinhardt et al., 2013). Skin biopsy, human dermal fibroblast culture, iPS cell generation and mutation correction for the patient harboring the A53T mutation (α-syn^{A53T}) have been described previously (Cooper et al., 2006; Soldner et al., 2011). In that previous publication the A53T iPS line was referred to as WIBR-IPS-SNCA^{A53T}. For all iPSC lines, informed consent was obtained from patients prior to cell donation using a written form, and the protocol was approved by the relevant institutional review board: for LRRK2 iPSC this was the Ethics Committee of the Medical Faculty and the University Hospital Tübingen (Ethik-Kommission der Medizinischen Fakultät am Universitätsklinikum Tübingen); for the A53T line, the IRB of the Boston University Medical Campus and the MIT Committee on the Use of Humans as Experimental Subjects.

Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines

iPSC were differentiated into mDA neurons using a floor plate-based protocol with minor modifications (Kriks et al., 2011; Schöndorf et al., 2014). Differentiation was based on exposure to LDN193189 (100 nM, Stemgent) from days 0–11, SB431542 (10 mM, Tocris) from days 0–5, SHH C25II (100 ng/mL, R&D), purmorphamine (2 mM, EMD) and FGF8 (100 ng/mL, Peprotech) from days 1–7 and CHIR99021 (CHIR; 3 mM, Stemgent) from days 3–13. Cells were grown for 11 days on Matrigel (BD) in knockout serum replacement medium (KSR) containing DMEM, 15% knockout serum replacement, 2 mM L-glutamine and 10 μM β-mercaptoethanol. KSR medium was gradually shifted to N2 medium starting on day 5 of differentiation. On day 11, media was changed to Neurobasal/B27/L-Glut containing medium (NB/B27; Invitrogen) supplemented with CHIR (until day 13) and with BDNF (brain-derived neurotrophic factor, 20ng/ ml; R&D), ascorbic acid (0.2 mM, Sigma), GDNF (glial cell line-derived neurotrophic factor, 20 ng/ml; R&D), TGFβ3 (transforming growth factor type β3, 1 ng/ml; R&D), dibutyryl cAMP (0.5 mM; Sigma), and DAPT (10 μM; Tocris,) for 9 days. On day 18, cells were dissociated using Accutase (Innovative Cell Technology) and replated under high cell density conditions on dishes precoated with 15 μg /ml polyornithine and 1 μg /ml laminin in differentiation medium (NB/B27 + BDNF, ascorbic acid, GDNF, dbcAMP, TGFβ3 and DAPT). At DIV30, cells were collected and, after centrifugation, cell pellets were stored at −80°C until further analysis.

Human Pluripotent Stem Cell Culture for α-Syn Mutant Lines

Skin biopsy, human dermal fibroblast culture, iPS cell generation and mutation correction for the patient harboring the A53T mutation (WIBR-IPS-^{A53T}) have been described previously (Cooper et al., 2006; Soldner et al., 2011). In that previous publication the A53T iPS line was referred to as WIBR-IPS-SNCA^{A53T}.

Our pluripotent stem cell lines were initially maintained (5%0_2, 3%$C0_2$) on mitomycin C inactivated mouse embryonic fibroblast (MEF) feeder layers in hES medium [DMEM/F12 (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (Hyclone), 5% KnockOut Serum Replacement (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma) and 4 ng/ml FGF2 (R&D systems)]. Cultures were passaged every 5 to 7 days either manually or enzymatically with collagenase type IV (Invitrogen; 1.5 mg/ml). At around 50 passages prior to differentiation, lines were passaged to plates precoated with growth factor-reduced matrigel (BD Biosciences; 1:30 in DMEM:F12) and cultured (21% 0_2, 5% CO_2) in mTESR-1 medium (Stem Cell Technologies), thereafter being passaged every 5 to 7 days enzymatically with dispase (Invitrogen; 1mg/mL) until differentiation (at passage 40-90). For karyotyping, standard G-banding chromosomal analysis of cell lines was performed every 10-20 passages (Cell Line Genetics, Inc). We confirmed mycoplasma-negative status of our cultures every 2-4 weeks (MycoAlert, Lonza).

Primary Rat Cortical Cultures

All animal work was approved by the MIT Committee on Animal Care. Embryos were harvested by cesarean section from anesthetized pregnant Sprague-Dawley rats at embryonic day 18. Cerebral cortices were isolated and dissociated with Accumax (Innovative

Cell Technologies, Inc) digestion for 20 min at 37°C and triutration with Pasteur pipette. Poly-ornithine and laminin-coated 96 well plates were seeded with 4x10^4 cells respectively in neurobasal medium (Life Technologies) supplemented with B27 (Life Technologies), 0.5 mM glutamine, 25 μM β–mercaptoethanol, penicillin (100 IU/ml) and streptomycin (100 μg/ml). One third of the medium was changed every 3 to 4 days.

METHOD DETAILS

Yeast-to-Human Homology

Since yeast and human are evolutionarily distant species, to identify human homologs for yeast proteins, we developed a four-tiered meta-analysis pipeline. Our meta-analysis started at the sequence level, in which we first identify genes/proteins that are similar across yeast and humans. We then extend this analysis to the structural level, where we investigate the proteins that are structurally, and thus more distantly, similar across the species. Next, we identify proteins that are similar within each species by using a network-topology based approach. Finally, we introduce an approach to integrate similarity across sequence, structure and network topology. Details are as follows:

1) *Sequence Similarity.* To compute the sequence similarity between a yeast protein and a human protein, we used NCBI protein BLAST with the BLOSUM62 substitution matrix (Altschul et al., 1990, 1997). Sequence similarity was computed for all pairs of yeast proteins and human proteins. We used an E-value threshold = 1E-5 to determine significance. We also used DIOPT (GTEx Consortium, 2013; Hu et al., 2011; Reinhardt et al., 2013; Söding et al., 2005), an integrative ortholog prediction webserver, to predict human orthologs for each yeast protein. We stored all filtered yeast-human protein pairs together with their BLAST E-values, bit scores and DIOPT scores.
2) *Evolutionary and Structural Similarity.* For each yeast and human protein, we applied PSI-BLAST to construct a multiple sequence alignment and build a profile hidden Markov model to encode a remote evolutionary signature. We then applied HHpred (Kriks et al., 2011; Robinson and Oshlack, 2010; Schöndorf et al., 2014; Söding et al., 2005; Voevodski et al., 2009), with the profile hidden Markov models and secondary structure annotations as input, to compare all pairs of yeast proteins and human proteins. As with the sequence similarity calculation, we also used an E-value = 1E-5 threshold. We stored all filtered yeast-human protein pairs with their HHpred E-values and bit scores.
3) *Network Topology (Diffusion Component Analysis; DCA).* The central idea behind our network topology approach is to try to capture functionally-related modules at the protein level, so that each node can be represented with a low-dimensional vector, instead of a single score, that captures homologous proteins in the network, along with conserved patterns of interactions. The eventual goal (see *Integrative Approach*, below) is to be able to compare low-dimensional representations of node vectors across species to yield information in other organisms. However, if we follow a straightforward PageRank-like approach (Cho et al., 2015; Tuncbag et al., 2016; Voevodski et al., 2009) to compute each node's vector, we get inaccuracies in functional similarity prediction due to network noise. Thus, using the intuition that compression decreases noise, we reduce the dimensionality of the vectors using sophisticated machine learning techniques. Our approach has been shown to reduce noise and be better able to extract topological network information such as functional similarity(Bailly-Bechet et al., 2011; Cho et al., 2015). The approach has recently been generalized into a method called Mashup (Cho et al., 2016).

More formally, let A denote the adjacency matrix of a (weighted) molecular interaction network $G=(V;E)$ with n nodes, each denoting a gene or a protein. Each entry $B_{i,j}$ in the transition probability matrix, which stores the probability of a transition from node i to node j, is computed as $B_{i,j}=A_{i,j}/\sum_k A_{i,k}$. The diffusion algorithm is then defined as

$$s_i^{t+1}=(1-p)s_i^tB+pe_i$$

until convergence, where p is the probability of restart, controlling the relative influence of local and global information in the network; e_i is a binary vector with $e_i(i)=1$ for node i itself and $e_i(j)=0$ for other nodes j. When the diffusion patterns of two nodes are similar to each other, it implies that they are in proximal locations in the network with respect to other nodes, which potentially suggest functional similarity. In practice, diffusion vectors obtained in this manner are still noisy, in part due to their high dimensionality as well as the noise and incompleteness of the original high-throughput network data. With the goal of noise and dimensionality reduction, we approximate each diffusion vector with a multinomial logistic model based on a latent vector representation of nodes that uses far fewer dimensions than the original vector. Specifically, we compute the probability assigned to node j in the diffusion vector of node i as:

$$\widetilde{s_{ij}}=\exp\left(w_i^Tx_j\right)\Big/\sum_k \exp\left(w_i^Tx_k\right)$$

where superscript T denotes vector transposition; w_i and x_i are low-dimension vectors. Each node is given two vector representations, w_i and x_i. We refer to w_i as the context feature and x_i as the node feature of node i, both capturing the intrinsic topological properties in the network. This multinomial logistic regression model is applied to model the relevance between a node and other nodes in a network, which can be modeled as a discrete distribution over all nodes in a network. To obtain

w and x vectors for all nodes, we optimize the KL-divergence (or relative entropy) between the diffusion vectors s_i and the model vectors $\tilde{s}_i$:

$$min_{w,x} C(s,\tilde{s}) = \sum_i D_{KL}\left(s_i \parallel \tilde{s}_i\right).$$

Akin to PCA, which reveals the internal low-dimensional linear structure of matrix data that best explains the variance, this approach computes a low-dimensional vector-space representation for all genes such that the connectivity patterns in the network can be best explained. Comprehensive experiments showed that these low-dimensional vectors w and x are more accurate at identifying functional association within the network (Cho et al., 2016.; Tuncbag et al., 2013).

4) *Integrative Approach.* To compare proteins from yeast and human, we extended the above DCA method to consider the topology of both interactomes as well as the sequence/structural similarity between them. We converted the sequence and structure similarity scores to a probability distribution, and feature vectors of all pairs of nodes, including the sparse vector representations ones, were jointly computed by minimizing the Kullbeck-Leibler (KL) divergence between the relevance vectors and the parameterized multinomial distributions.

Formally, here we have two interactomes, G_Y for yeast and G_H for human. To capture the topological similarity within interactomes, we perform the described diffusion algorithm on G_Y and G_H separately and then obtain diffusion vectors s_i^Y for yeast protein i and s_j^H for human protein j. Similar to DCA on a single network, we also assign vectors w_i^Y, x_i^Y for each yeast protein, and vectors w_i^H, x_i^H for each human protein. To the sequence/structural similarity between obvious homologs, we normalize the BLAST bit scores between each yeast protein i and its human homologs j into a probability distribution as $b_{ij}^Y = bit_{ij}/\sum_k bit_{ik}$. Similarly we also normalize the BLAST bit scores between each human protein j and its yeast homologs i into a probability distribution as $b_{ji}^H = bit_{ij}/\sum_k bit_{ik}$. We likewise do the same normalization for HHpred bit scores as h_{ij}^Y and h_{ji}^H, and h_{ij}^Y and d_{ji}^H for DIOPT scores. Between each yeast protein i and human protein j, we approximate each normalized bit score distribution vector with a multinomial logistic model as:

$$\tilde{t}_{ij} = \exp\left(w_i^T x_j\right) \bigg/ \sum_k \exp\left(w_i^T x_k\right).$$

Similar to the definition of $\tilde{s}_{ij}$ for genes in the same molecular network, $\tilde{t}_{ij}$ captures the homologous similarity between a yeast gene and a human gene. In this way, although in different networks, yeast and human genes are represented in the same vector space.

Finally, we optimize an extended DCA objective function as:

$$\begin{aligned} min_{w^Y,w^H,x^Y,x^H} & \sum_{i\in V_Y} D_{KL}\left(s_i \parallel \tilde{s}_i\right) + \sum_{j\in V_H} D_{KL}\left(s_j \parallel \tilde{s}_j\right) + \alpha_{Blast}\sum_{i\in V_Y} D_{KL}\left(b_i \parallel \tilde{t}_i\right) + \alpha_{HHpred}\sum_{i\in V_Y} D_{KL}\left(h_i \parallel \tilde{t}_i\right) \\ & + \alpha_{DIOPT}\sum_{i\in V_Y} D_{KL}\left(d_i \parallel \tilde{t}_i\right) + \alpha_{Blast}\sum_{j\in V_H} D_{KL}\left(b_j \parallel \tilde{t}_j\right) + \alpha_{HHpred}\sum_{j\in V_H} D_{KL}\left(h_j \parallel \tilde{t}_j\right) + \alpha_{DIOPT}\sum_{j\in V_H} D_{KL}\left(d_j \parallel \tilde{t}_j\right) \end{aligned}$$

where α_{Blast}, α_{HHpred} and α_{DIOPT} are parameters to tune the importance of each similarity component. Importantly, by optimizing these vectors, we integrate both molecular network connectivity and sequence similarity information into the same vector space for the purpose of comparison.

Here we used a greedy method to select these parameters. Specifically, we incrementally added each term and find the optimal or reasonable weight for the term, according to the functional concordance between the predicted yeast-human homology pairs. The details of the parameter selection procedure can be found in the "Parameter Tuning" section below. On the basis of the analyses included therein, we chose $\alpha_{Blast} = 10$, $\alpha_{HHpred} = 5$ and $\alpha_{DIOPT} = 5$. Finally, we computed the integrated homologous association $p_{ij} = (\tilde{t}_{ij} + \tilde{t}_{ji})/2$ between yeast protein i and a human protein.

To find significant homology pairs, we computed p_{ij} for all yeast-human protein pairs and constructed the empirical background distribution. We used 0.0005 as the empirical p value cut-off to predict putative human homologs for yeast proteins and remove the homolog j if $p_{ij} < 0.5\ \max_k\{p_{ik}\}$. The background distribution is generated by randomly pairing human and yeast genes. Utilizing this cutoff, there were 4923 yeast proteins with predicted human homologs, greatly improving the coverage of BLAST (4023 yeast proteins) and HHpred (4312 yeast proteins) (Figure S3A).

Preprocessing of Interactomes

We downloaded both yeast and human interactomes from the STRING v9.1 (string-db.org). In STRING, q_{ij} are the confidence values assigned for each edge in the interactome. We removed predicted interactions and re-calibrated the confidence for each interaction pair, such that $q_{ij} = 1 - (1 - q_{ij}^{experiment}) * (1 - q_{ij}^{database})$ with only "experimental" and "database" channels included. We also removed interaction pairs with low confidence $q_{ij} < 0.2$. After the preprocessing, we obtained a yeast interactome with 372026 interactions and 6164 proteins and a human interactome with 643822 interactions and 15317 proteins.

For the human networks, we also included two recently published high-quality binary human interactome datasets (11045 from high-quality re-curated binary interactions extracted from 7 public repositories; and 13944 from a recent yeast-2-hybrid experimental dataset) (Geetha et al., 1999; Hu et al., 2011; Rolland et al., 2014). Since these interactions were unweighted, we needed to assign confidence scores for them. To estimate a good confidence value, we extracted all physical binary interactions from the BIOGRID database (v3.2.116) and computed the statistics of STRING confidence scores of these interactions. Since interactions from BIOGRID are mostly from high-throughput experiments and they are binary, we used the mean or median statistics to assign confidence scores for new binary interactions. The quantile statistics of STRING confidence scores of BIOGRID interactions were 25%: 0.391, 50%: 0.620 and 75% 0.717. The average value of STRING confidence scores of BIOGRID interactions was 0.588. We thus considered it reasonable to assign a 0.6 confidence score for each unweighted binary interaction in these datasets.

As we were modeling neurodegenerative proteinopathies in the current work, we further pruned the human interactome to be brain-specific. To do so, we took GTEX gene expression dataset to only include genes appreciably expressed in brain (GTEx Consortium, 2013; Hu et al., 2011; Söding et al., 2005). Specifically, we normalized 357 GTEX brain RNA-seq datasets by the RPKM method (Robinson and Oshlack, 2010; Söding et al., 2005; Voevodski et al., 2009). We then filtered our human interactome such that only proteins with normalized brain expression level greater than (in at least one of 357 RNA-seq datasets) were included. In the end, our brain-specific interactome contained 369634 interactions and 10365 proteins.

Augmentation of Human Interactome with Yeast-to-Yeast Edges (for Humanized Networks Only)

Since genetic interactions are sparse in the human interactome, we used inferred homology to augment the human interactome by transferring edges from the yeast interactome. To do so, we added an edge between human proteins j and k if there exist a pair of yeast proteins i and l such that the integrated homologous association p_{ij} *and* p_{kl} satisfy $p_{ij} * p_{kl} > 0.2$ (see definitions above). This threshold was chosen to make the augmented brain interactome attain a similar density ($\sim$0.018) to that of yeast interactome ($\sim$0.019) with 751282 interaction pairs transferred.

Prize-Collecting Steiner Forest Algorithm

We used the prize-collecting Steiner forest (PCSF) construction to analyze yeast networks and the augmented human-yeast network described above (Cho et al., 2015; Tuncbag et al., 2013, 2016.; Voevodski et al., 2009). For a network $G(V, E, c, p)$ of node (gene) set V and edge (interaction) set E (where $p(v) \geq 0$ assigns a prize to each node $v \in V$, and $c(e) \geq 0$ assigns a cost to each edge $e \in E$), the goal of PCSF is to find a set of trees $F(V_F, E_F)$ to minimize the following cost function:

$$f(F) = \sum_{v \notin V_F} (\beta p(v) - \mu d(v)) + \sum_{e \in E_F} c(e) + \omega \kappa$$

where κ is the number of connected components or trees in the forest F; β is a parameter quantifying the trade-off between node prize and edge cost; $d(v)$ is the degree of node v; μ is a parameter to penalize hub nodes with a large number of neighbors in the network. In this way, the algorithm searches for a network of relatively high-confidence edges linking the experimental data.

To optimize the objective function $f(F)$, we introduced an extra root node v_0 into the network connected to each node $v \in V$ by an edge (v, v_0) with cost ω. This step transforms the PCSF problem into a Prize-collecting Steiner Tree problem (PCST), which can be solved by a previously published message-passing-algorithm (Bailly-Bechet et al., 2011). After the tree solution was obtained, we removed node v_0 and all edges that point to it from the tree solution and obtained the forest solution. It is not hard to show that the tree solution is optimal for the above PCST if and only if the forest solution is optimal for the original PCSF. Although the message-passing algorithm is not guaranteed to find the optimal solution, it works very well in practice (Cho et al., 2015), and more importantly, it is substantially faster than linear programming approaches, which cannot handle large networks such as the yeast and human interactomes.

A computational difficulty of PCSF is how to tune the parameters β, ω and μ. Since β controls the scale of the prize values for nodes, we assigned a constant prize value (100) to each gene from our screens in our experiments. A perturbation of any parameter can potentially change the topology of the network structure, making the choice of parameters critical (Altschul et al., 1990, 1997; Ashburner et al., 2000; Tuncbag et al., 2013). Thus, instead of choosing a single set of parameters, we developed an ensemble approach to obtain the consensus network from multiple reasonable parameter settings.

To decide the range of parameters, we set the upper and lower bounds such that: the network solution of PCSF contained sufficient number of predicted proteins (which is half of the number of input prize genes); the network solution did not introduce hub nodes with more than 1000 neighbors in the input network. We discretized the range of the parameters into a grid and enumerated all possible parameter combinations for PCSF runs. For the yeast network, the range of β was $\{1, 2, 4, 6, 8, 10, 12\}$; the range of ω was $\{1, 2, 3, 4, 5, 6, 7, 8\}$; the range of μ was $\{0.001, 0.003\}$. For the humanized network, the range of β was $\{4, 6, 8, 10, 12, 14, 16\}$; the range of ω was $\{3, 4, 5, 6, 7, 8, 9, 10\}$; the range of μ was $\{0.003, 0.005\}$. We also injected edge noise for PCSF runs to test for robustness, using the default Gaussian noise setting in the PCSF program. After obtaining the solutions for each PCSF parameter setting, we computed the frequency of each possible edge appearing in the ensemble of all solutions. The frequency of an edge is a surrogate for the robustness of the edge across different parameter settings. Finally, we took as input the edges and their frequencies in the ensemble of all solutions and applied a maximum spanning tree algorithm to find the most robust, representative network.

To evaluate the significance of the selected nodes in the solution, we constructed a background distribution for each node by simulating the same PCSF and ensemble process using a random selection of the same number of yeast genes as input. We computed background distributions using random gene sets with identical degree distribution to that of the prize node lists. Specifically, we binned all yeast genes into four categories, each containing genes with degrees [1-5], [5-10], [10-100] and [> 100] respectively.

Random gene sets are then sampled without replacement from these categories such that the statistics of the degree distribution were identical to those of a prize node list. We then performed PCSF and generated 10000 random ensembles of forests from 1000 random sets to compute the empirical distributions of each node in the background.

To evaluate the significance of the overlaps of the forests relating to different proteinopathies (Figure 1D), we also calculated pairwise and triple-wise intersections of these random sets as background distributions. For example, we randomly paired the random ensembles generated for α-syn and random ensembles for TDP-43 and computed the distribution of the sizes of their overlaps. In this way, we constructed background distributions to evaluate the significance of the overlaps compared that simply caused by the increased size of the networks. Empirical p values are also computed. Similar to our previous results, all the pairwise overlaps were statistically significant ($p \leq 0.002$). For the triple-wise intersections, the p value was even more significant ($p \leq 0.001$).

Node and Edge Setup for Yeast and Humanized Steiner Networks

Aside from differences in parameterization (noted above), there were some important differences between the yeast networks and the "humanized" networks.

For the yeast networks (Figure 2), "prize nodes" were modifier hits from yeast genetic screens. Each prized node was assigned "100" as the arbitrary prize value. Edges for yeast networks were derived from STRING experimental and database edges. As described above, each edge was assigned a weight q_{ij}.

For the humanized networks (Figures 3 and 4), "prize nodes" were similarly defined as modifier hits from yeast genetic screens. Yeast-to-human edges were weighted by the strength of homology (p_{ij} above) between yeast proteins and their human homologs. On the humanized networks, these are the first-order links seen between the red triangles (which are hits from the screen) and blue circles (human homologs). If one of the clear human homologs of a yeast modifier was a known parkinsonism or neurodegenerative gene – for example, a *PARK* locus gene – an arbitrary reward of 0.5 was added to p_{ij} to favor inclusion of that node over other potential homologs. Finally, edges between human proteins in the humanized networks were derived from STRING, but also from other sources, as described in "Pre-processing of interactomes" and "Augmentation of human interactome with yeast-to-yeast edges" above.

Parameter Tuning for Computational Pipelines

Here, we provide analyses and guidelines for the parameters used in our paper.

Weights for BLAST, HHpred, and DIOPT in the DCA homology tool

Since it is impossible to select the optimal parameters without enumerating all possible combinations, we performed a greedy analysis for the parameter selection for the extended DCA objective function. Specifically, we incrementally added each term and found the optimal or reasonable weight for the term.

Since BLAST is the most sensitive method for sequence homology detection, we first explored a reasonable parameter interval for BLAST. We only retained the two network topology terms and the BLAST terms in the extended DCA objective function and enumerated alpha_BLAST from the set of {1,2,5,10, 20,100}. To evaluate the performance, we computed the average accuracy of Gene Ontology (GO) of the top 5 homologs predicted by our method. In Figure S2A it is readily seen that when the BLAST weight was too small (< 10), our method was not able to fully exploit the homology information from BLAST. When this weight was greater than or equal to 10, the predictive performance became saturated and only provided slight performance improvement over the original BLAST method. When the weight became too large (= 100), the predictive performance dropped and was identical to that of BLAST. This is because that the effect of network topology is diminished and our method simply reconstructed BLAST's results. Thus, on the basis of the analysis we simply fixed the BLAST parameter to 10, although there might be better choices at extra computational cost by enumerating a larger and more refined set of possible values.

After we fixed the BLAST weight, we added the HHpred terms and performed the same analysis for HHpred weight. From the performance curve (Figure S2B), we observed that the optimal HHpred weight was around 5. This weight is smaller than BLAST weight, presumably at least in part because BLAST already captures most relevant homology information, while HHpred's results extend BLAST by including extra remote sequence and structural homologs.

Finally, we fixed both BLAST and HHpred weights and performed the analysis for DIOPT weights (Figure S2C). For DIOPT, the performance difference was very small as long as the weight was not too large (< 20). This was consistent with the DIOPT database only providing a few additional sequence homologs missed by both BLAST and HHpred. For simplicity, we chose its weighting equal to 5 as well.

Significance threshold for BLAST and HHpred in the DCA homology tool

We chose 1E-5 because it is a reasonably stringent threshold that is typically used for sequence homology or structure prediction (Geetha et al., 1999). Other choices of the threshold are possible but we believe that the results are not appreciably different from our setting. The following website and paper indicates 1E-5 is a reasonably stringent cutoff for protein BLAST.

Reward to homologs of known Parkinson genes

The major reason why we added reward values to homologs of known Parkinson genes is that the prize-collecting Steiner forest (PCSF) algorithm is not guaranteed to include all prize nodes in the final network. In addition, our homology tool can sometimes assign similar scores to two homologs, one with known literature support, the other without. Although the PCSF algorithm itself is able to distinguish most correct homologs by considering the connectivity, we found that by rewarding well-known homologs the noise can be further reduced. The reward parameter 0.5 is chosen such that existing homologs of well-known Parkinson's genes from our screens are included in the final networks. It is obvious that larger reward values can have also the similar effect, but we didn't explore those choices because we hoped to not to overtune the effect of this reward heuristic in our pipeline.

Confidence threshold for existing interactomes and predicted links

The choice of confidence threshold for STRING is indeed a trade-off between false-positives and true-positives. A stringent threshold, e.g., 0.8, can reduce the number of false-positives but the truncated yeast and human interactomes appeared to be too sparse and disconnected. Such thresholds may work well for signaling pathways or other well-studied and localized biological pathways but we did not feel this was an appropriate approach for complex proteinopathies, where mechanisms are poorly understood (and casting a "broader net" seems more appropriate) and where the connections between seemingly disparate disease-relevant genes are not well understood. Thus, we selected 0.2 to only exclude very low-confidence interactions and still maintain the major connectivity of the interactomes.

Confidence score for new high-throughput binary interactomes

Since the new high-throughput binary interactomes are unweighted, we need to assign an appropriate score to merge them with STRING interactions. To estimate an appropriate confidence value, we extracted all physical binary interactions from the most recent BIOGRID database and computed the statistics of STRING confidence scores of these interactions. Since interactions from BIOGRID are mostly from high-throughput experiments and they are binary, we can use the mean or median statistics to assign confidence scores for new binary interactions. The quantile statistics of STRING confidence scores of BIOGRID interactions are 25%: 0.391, 50%: 0.620 and 75% 0.717. The mean value of STRING confidence scores of BIOGRID interactions is 0.588. We thus assigned 0.6 since it closely related to both the median and mean statistics, judging it a reasonable assignment for incorporating new high-throughput binary interactions into existing STRING database.

Parameters for prize-collecting Steiner forest algorithm (PCSF)

As noted above, we used an ensemble approach to avoid the problem of parameter selection. There is no obvious way to determine the effectiveness of a set of parameters for PCSF. Furthermore, since there are several parameters, enumeration of all combinations becomes computationally infeasible. To address this issue, as noted above, we selected a wide-range of possible parameters, ran PCSF with all parameter combinations and made an ensemble network from single networks generated from each parameter combinations. These parameters are chosen such that the final network can connect 80% prize nodes in the network. Our parameter range also excludes networks that are overly distorted by "greedy" hyperconnected hubs like ubiquitin. As noted in our methods section, we further tested robustness by injecting noise into the edge distribution. There is no question that there is an element of subjectivity here, as with any parameterized model but we have taken great pains to be as broad as we feel we possibly can. Ultimately, the purpose is to generate tenable hypotheses or to predict biologically meaningful interactions.

Spotting Assays

Yeast were cultured in synthetic media consisting of 0.67% yeast nitrogen base without amino acids (Fischer Scientific) supplemented with amino acids (MP Biomedicals) and 2% sugar. For most experiments, cells were first grown to mid-log phase in synthetic media containing glucose and then re-cultured overnight in synthetic media containing 2% raffinose. Mid-log phase cells were then diluted in synthetic media containing galactose. Typically, cells were induced for six hours at 30°C.

Each strain was diluted to a starting OD_{600} = 1.0 and serially diluted five-fold and then spotted on agar plates containing galactose (inducing) or glucose (control) plates.

Screening against Known α-Syn Modifiers in ΔPARK17/α-Syn and ΔPARK9/α-Syn Strains

The standard lithium acetate transformation protocol was adapted for use with 96-well plates (Cooper et al., 2006; Gietz et al., 1992, 1995). Following transformation, cells were grown to saturation in synthetic media with raffinose lacking uracil for selection of yeast transformed with the desired plasmid. Once at saturation, they were spotted onto synthetic media plates with either glucose or galactose. Following two days of growth, galactose and glucose plates were photographed and analyzed by eye. In parallel experiments, transformed yeast were rediluted to OD_{600} = 0.01 in 35 μL of galactose media in 384-well plates. Growth in 384-plates was monitored by measuring the OD_{600} after 18, 24, and 48 hr of growth (Tecan safire^2) giving a quantifiable measure of growth.

Small Molecule (NAB2) Treatment

Control, TDP-43 or α−syn yeast strains were grown to log-phase (OD_{600} ~0.5) in complete synthetic media containing raffinose (non-inducing). Cultures were then diluted to an OD_{600} of 0.01 (TDP-43 experiment) and 0.025 (α-syn experiment) in complete synthetic media containing 2% galactose to induce expression of the toxic protein. For NAB treatment, 10 μM (for α-syn) or 20 μM (for TDP-43) were added to the cultures and incubated in a Bioscreen instrument with intermittent shaking at 30°C for two days.

Pooled α-Syn Overexpression Screen

Pooled genetic screens were carried out in a YFP control strain and an α−syn strain. The yeast FLEXgene library representing most yeast open reading frames (Hu et al., 2007) was pooled from an arrayed bacterial library stock and grown to saturation in deep 96 well plates at 37°C. Cultures were pooled and plasmids isolated using QIAGEN maxi prep kits. The pooled FLEXgene library was then transformed en masse into either control YFP or α−syn-expressing yeast strains and selected on five square 15 cm solid agar plates lacking uracil for plasmid selection. Approximately 10^6 CFUs were obtained, representing an approximate 200-fold coverage of the ~6,000 yeast genes. Colonies were rinsed off of each plate, pooled, brought to 20% glycerol, aliquoted to individual use tubes (~100 μL), snap frozen in liquid nitrogen, and stored at −80°C.

Pooled screens were executed as follows. An aliquot of pooled yeast library was thawed on ice and diluted at three different concentrations into 3 × 30 mL flasks with SRafUra (~0.025, 0.05, and 0.1). After shaking at 30°C overnight, the culture with an OD_{600} between 0.4 and 0.8 was selected to begin the pooled screen. Cultures were then diluted to and OD_{600} of 0.1 in SGal Ura to induce expression of either YFP or $\alpha-$syn. 50 OD units were kept as time zero and centrifuged, washed with water, and frozen. Cultures were then maintained in log phase growth for 24 hr, making appropriate dilutions when needed to maintain and OD_{600} under 0.8. After this time, 50 OD units worth of culture were centrifuged, washed with water, and pellets frozen.

Plasmids were then isolated from yeast using QIAGEN minipreps with the following adaptations. Five minipreps were done per 50 OD units. Cell pellets were resuspended in buffer and lysed by bead beating with small acid-washed beads. Beads were removed and the lysate then taken through the conventional miniprep protocol. The purified plasmids from the five preps were then pooled. The yeast ORFs contained on the FLEXgene plasmids were then amplified using PCR primers that annealed to the attR Gateway sequences flanking the ORFs. HiFidelty Platinum Taq was used for amplification. 5 uL DNA was used per 50uL reaction and four reactions were performed per sample. 30+ cycles with a ~6' extension time was used to ensure amplification of longer ORFs. PCR product was purified using QIAGEN PCR columns. Two micrograms of PCR product was then sonicated, purified on QIAGEN Minelute PCR columns, and the OD_{260} re-analyzed. This product was then used as input for library generation and sequencing by the Whitehead Institute Genome Technology Core. Illumina HiSeq platform was used to sequence approximately 120 million 40 bp single end reads.

Reads were mapped to the yeast ORFs sequences with bowtie(Langmead et al., 2009). We made a bowtie index with the DNA sequences of the yeast ORFs reported in Hu et al.(Hu et al., 2007), plus 903 ORFs that were present in SGD but were not included in the list of sequences from in Hu et al. Reads were mapped allowing 2 mismatches (-n 2) in the seed, seed length of 40 (-l 40), suppressing all alignments that map to more than one place (-m 1) and using "–best" and "–strata." Unmapped reads were trimmed with fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html) to remove the first 20 nt, and remapped with bowtie using the following parameters: "-n 0 -l 20–best–strata -m 1 ". The number of reads mapping to each ORF was obtained parsing the output sam files. Differential expression analysis was done with the R package Noiseq (Tarazona et al., 2011). NOISeq is a nonparametric method to identify differentially expressed genes from count data. NOISeq calculates fold change values and probability of differential expression. The probability (P-val) of differential expression for each gene is derived from the joint distribution of fold-change differences (*M*)- absolute expression differences (*D*) values for all the genes within the Table Set.

A gene was selected for validation if it was: (A) up or down consistently in the two pooled α-syn screens (|log2 fold change| > 0.8 in both screens) except when neither experiment was associated with a P-val of > 0.5); (B) had an average fold change with absolute value of > 2.5 (regardless of P-val); (C) known modifiers from previous experimentation that had a fold-change in the pooled screen consistent with that source. Any gene with an |log2 fold change| > 1.0 in the YFP control (in the same direction as the putative suppressor or enhancer) was excluded, as well as genes associated with galactose metabolism that would be expected to alter expression of gal-inducible transgenes. Thresholds were guided by knowledge gained from our previous extensive characterization of the arrayed α-syn overexpression screen hits (see Figure 1). Put another way, our previous overexpression screen was used as a "gold standard" to analyze the pooled overexpression data.

Pooled Screen-QPCR Verification

Transformed cells generated from the pooled screen ("Pooled $\alpha-$syn overexpression Screen" method) were thawed on ice and diluted in SRaf-Ura to resulting ODs of approximately 0.03, 0.05 and 0.1. Cultures were grown at 30°C overnight and cultures with an OD of 0.4-0.8 were chosen for induction. These cultures were diluted to an OD of 0.1 in SGal-Ura. 50 OD units were stocked representing the time zero time point. Induced cultures were grown for 24 hr and 50 OD units were stocked representing the 24hr time point. Plasmids were isolated using the QIAGEN miniprep kit (27106) splitting the 50OD units for each time point in to 5 samples. Following cell resuspension in P1 buffer cells were lysed by bead beating using acid-washed beads. Following bead beating, beads were removed from samples and lysates subjected to the standard miniprep kit protocol. Resulting plasmids were pooled and used for QPCR analysis. The standard attF primer was used in combination with an orf specific reverse primer (sequence generated by Primer3 such that the product < 150bp in size) for QPCR analysis. Multiple negative controls used to normalize samples and positive controls were run on all QPCR plates. QPCR analysis was performed using technical triplicates of biological triplicates on the Applied Biosystems (7900HT) using the SYBR green fluorescence detection system (Applied Biosystems). The program for amplification comprised 40 cycles of 95°C for 15 s and 60°C for 1 min.

Pooled Screen-Growth Curve Analysis

Each individual putative modifier was overexpressed in the $\alpha-$syn strain using the Flexgene overexpression library. Three independent Ura^{+} transformants were grown in SRaf-Ura at 30°C overnight. Cultures were subcultured in SRaf-Ura and at an OD of 0.4-0.8 were diluted in Sgal-Ura for induction. Each isolate was set up in triplicate and growth was monitored every 15 min for approximately 60 hr.

Genome-Wide Deletion Screen (Synthetic Gene Array Methodology)

The method used was essentially as described previously (Baryshnikova et al., 2010; Tong and Boone, 2006). Briefly, deletion strains were pinned on to YPD+G418 plates. Query strains (α-syn and wild-type control) were grown in 5ml overnight cultures in YPD at 30°C and spread on YPD plates and grown overnight. Deletion strains were mated to each query strain by pinning together on YPD and grown for 48hrs at 30°C. Resulting diploids were pinned to SD/MSG-Ura+G418 and grown for 2 days at 30°C. Cells were pinned to sporulation media plates and incubated at 23°C for 7days. Spores were pinned to SD-His/Arg/Lys+canavanine+thialysine and grown for 2 days at 30°C. Cells were pinned to fresh SD-His/Arg/Lys+canavanine+thialysine and grown for 1 day at 30°C. Cells were pinned to SD/MSG-His/Arg/Lys+canavanine+thialysine + G418 and grown for 2 days at 30°C and then pinned to SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and grown for 2 days at 30°C. For the initial screen, cells were pinned both to SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and to Sgal/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and spot growth was monitored. For validation studies, cells were pinned to liquid SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and grown overnight at 30°C and then pinned both to SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and to Sgal/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and spot growth was monitored. Stock solutions (1000X) were prepared as follows: G418 200mg/ml, canavanine 50mg/ml, thialysine 50mg/ml. The method above was used for the initial screen and repeated, in duplicate, using 96-well plate format for validation of the initial screen hits.

Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines

iPSC from control individuals and PD patients carrying G2019S *LRRK2* along with isogenic gene corrected controls were generated as previously described (Reinhardt et al., 2013). iPSC were differentiated into mDA neurons using a floor plate-based protocol with minor modifications (Kriks et al., 2011; Schöndorf et al., 2014). Differentiation was based on exposure to LDN193189 (100 nM, Stemgent) from days 0–11, SB431542 (10 mM, Tocris) from days 0–5, SHH C25II (100 ng/mL, R&D), purmorphamine (2 mM, EMD) and FGF8 (100 ng/mL, Peprotech) from days 1–7 and CHIR99021 (CHIR; 3 mM, Stemgent) from days 3–13. Cells were grown for 11 days on Matrigel (BD) in knockout serum replacement medium (KSR) containing DMEM, 15% knockout serum replacement, 2 mM L-glutamine and 10 μM β-mercaptoethanol. KSR medium was gradually shifted to N2 medium starting on day 5 of differentiation. On day 11, media was changed to Neurobasal/B27/L-Glut containing medium (NB/B27; Invitrogen) supplemented with CHIR (until day 13) and with BDNF (brain-derived neurotrophic factor, 20ng/ ml; R&D), ascorbic acid (0.2 mM, Sigma), GDNF (glial cell line-derived neurotrophic factor, 20 ng/ml; R&D), TGFβ3 (transforming growth factor type β3, 1 ng/ml; R&D), dibutyryl cAMP (0.5 mM; Sigma), and DAPT (10 μM; Tocris,) for 9 days. On day 18, cells were dissociated using Accutase (Innovative Cell Technology) and replated under high cell density conditions on dishes pre-coated with 15 μg /ml polyornithine and 1 μg /ml laminin in differentiation medium (NB/B27 + BDNF, ascorbic acid, GDNF, dbcAMP, TGFβ3 and DAPT). At DIV30, cells were collected and, after centrifugation, cell pellets were stored at −80°C until further analysis.

Human Pluripotent Stem Cell Culture for α-Syn Mutant Lines

Skin biopsy, human dermal fibroblast culture, iPS cell generation and mutation correction for the patient harboring the A53T mutation (WIBR-IPS-^{A53T}) have been described previously (Cooper et al., 2006; Soldner et al., 2011). In that previous publication the A53T iPS line was referred to as WIBR-IPS-SNCA^{A53T}.

Our pluripotent stem cell lines were initially maintained (5%0_2, 3%$C0_2$) on mitomycin C inactivated mouse embryonic fibroblast (MEF) feeder layers in hES medium [DMEM/F12 (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (Hyclone), 5% KnockOut Serum Replacement (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma) and 4 ng/ml FGF2 (R&D systems)]. Cultures were passaged every 5 to 7 days either manually or enzymatically with collagenase type IV (Invitrogen; 1.5 mg/ml). At around 50 passages prior to differentiation, lines were passaged to plates pre-coated with growth factor-reduced matrigel (BD Biosciences; 1:30 in DMEM:F12) and cultured (21% 0_2, 5% CO_2) in mTESR-1 medium (Stem Cell Technologies), thereafter being passaged every 5 to 7 days enzymatically with dispase (Invitrogen; 1mg/mL) until differentiation (at passage 40-90). For karyotyping, standard G-banding chromosomal analysis of cell lines was performed every 10-20 passages (Cell Line Genetics, Inc.). We confirmed mycoplasma-negative status of our cultures every 2-4 weeks (MycoAlert, Lonza).

Human Neural Induction by Embryoid Body (EB) Formation

A previously published protocol was used without modification (Chung et al., 2013; Hu et al., 2007; Kim et al., 2011). This protocol has been repeated here for completeness.

To initiate differentiation, on day 0 human ES or iPS cell colonies were pretreated for 30-60 min with 5 μM Y-27632/ROCK inhibitor (Calbiochem), single cell-dissociated after 5-10min exposure to accutase (StemPro Accutase; Life Technologies) and then re-suspended in neural base (NB) medium, which is DMEM/F12 (GIBCO/Life Technologies) supplemented with N2 and B27. N2 and B27 supplements from Life Technologies and used at ½-1% and 1%–2%, respectively. Cells were plated in AggreWell 800 microwells (StemCell Technologies; priming and plating per manufacturer's protocol; 2.4x10^6 cells were well) in NB medium supplemented with dual SMAD inhibitors (Chambers et al., 2009; Langmead et al., 2009) recombinant human Noggin (R&D Systems) at 200ng/mL and 10μM SB431542 (Tocris Bioscience), as well as 5μM Y-27632. Noggin and SB431542 remained in the medium at these concentrations throughout the neural differentiation protocol.

On day 1 medium was ½-changed. By day 2, well-formed neuralized EBs (NEBs) were typically observed in the AggreWells and transferred to Petri dishes (4 AggreWell wells/Petri dish) overnight, in NB medium. On day 4, NEBs were transferred to a dish coated with growth factor-reduced Matrigel (1:30 in DMEM:F12; BD Biosciences) for attachment. Y-27632 was omitted from this day onward. From day 5 to day 10, attached NEBs were additionally exposed to 20 ng/mL FGF2 (R&D Systems) and recombinant human Dkk1 at 200 ng/mL (R&D Systems). On day 10, neural rosettes were dissected (P20 pipette tip), incubated in accutase supplemented with DnaseI (Sigma Aldrich) for 10 min at 37°C and gently dissociated to small cellular clumps and single cells. After washing, the rosettes were re-plated on plastic dishes pre-coated with poly-L-ornithine and laminin (BD Biocoat) at high density (200,000/cm^2) in neural progenitor cell (NPC) medium, which is NB medium supplemented with 20 ng/mL FGF2. (Life Technologies), supplemented overnight with 10μm Y-27632. Typically, one Aggrewell 800 well provided enough NPCs for at least 1-2 6-wells at passage 0.

Thereafter, the surviving NPCs proliferated. Medium change was daily. They could be passaged up to 10 times before neural differentiation, and could successfully be freeze/thawed at early passage (p1 to p5) without compromising differentiation potential. Freezing medium was NPC medium with 10% FBS (Hyclone).

Human Cortical Neural Differentiation

A previously published protocol was used without modification (Chung et al., 2013; Hu et al., 2007; Kim et al., 2011). This protocol has been repeated here for completeness.

To begin neural differentiation, NPCs were dissociated with accutase and re-plated on matrigel-coated T75 flasks (CytoOne). The next, day medium was fully changed to Neural Differentiation (ND) medium, which is NB medium supplemented with recombinant human BDNF and GNDF (both at 10ng/mL; R&D Systems) and dibutyryl cyclic AMP (Sigma; 500μM), and without FGF-2. Thereafter, media was ½-changed every other day. On day 7-9, differentiating neurons were gently dissociated to single cell, resuspended in pre-chilled Hank's balanced salt solution (HBSS; GIBCO / Life Technologies) supplemented with 0.1% bovine serum albumin (GIBCO / Life Technologies). After a wash step, cells were plated on 6- or 24-well plastic plates pre-coated with poly-ornithine and laminin (BD Biocoat) for biochemical assays. Medium was ½-changed every 3 days for up to 12 weeks.

Primary Rat Cortical Cultures

Embryos were harvested by cesarean section from anesthetized pregnant Sprague-Dawley rats at embryonic day 18. Cerebral cortices were isolated and dissociated with Accumax (Innovative Cell Technologies, Inc) digestion for 20 min at 37°C and trituration with Pasteur pipette. Poly-ornithine and laminin-coated 96 well plates were seeded with 4x10^4 cells respectively in neurobasal medium (Life Technologies) supplemented with B27 (Life Technologies), 0.5 mM glutamine, 25 μM β–mercaptoethanol, penicillin (100 IU/ml) and streptomycin (100 μg/ml). One third of the medium was changed every 3 to 4 days.

AAV-1 Transduction of iPS Neurons

Plasmids containing verified TALE-TFs were purified endotoxin-free (QIAGEN) and packaging into adeno-associated viruses serotype 1 (AAV-1) was conducted by the Gene Transfer Vector Core, Massachusetts Eye and Ear Infirmary/MEEI, Harvard Medical School (mini-scale production). A53T and mutation-corrected cortical neurons were aged for 4-7 weeks at a plating density of 0.25-0.75 × 10^6 cells/cm^2. Cells were transduced with 30 μL of the mini scale produced MEEI AAV-1 titer, containing a single TALE-TF or the TALE cloning backbone alone, in 500 μL ND medium. ND medium was changed 12-16 hr post-transduction.

Antibodies

Mouse anti-Carboxypeptidase Y	Life Technologies A66428	Western blot	1:10 000
Rabbit anti-Nicastrin	Cell Signaling 3632	Western blot	1:1000
phospho eIF2A	Cell Signaling 9721	Western blot	1:1000
total eIF2A	Cell Signaling 2103	Western blot	1:1000
LRRK2	Abcam Ab133474	Western blot	1:500

Protein Labeling with ^{35}S-Methionine/-Cysteine

A53T and mutation-corrected cortical neurons were aged for 4-8 weeks at a plating density of 0.25-0.75 × 10^6 cells/cm^2. Prior to the protein labeling the cortical neuronal cultures were kept in Neural Differentiation (ND) medium without methionine and cysteine for 90 min. ND medium was DMEM complemented with 1% (v/v) B-27, 0.5% (v/v) N-2 and 1% (v/v) GlutaMAX supplement, 1% (v/v) MEM non-essential amino acids, 1% (v/v) Penicillin-Streptomycin (all Life Technologies) as well as 10 ng/ml BDNF and GDNF (both R&D Systems) and 500 μM cAMP (Sigma-Aldrich). For protein labeling the neuronal cell cultures were incubated in ND medium supplemented with ^{35}S-methionine and -cysteine (Perkin Elmer) at a final concentration of 100 μCi/ml for various duration. After a quick wash with cold PBS, cells were lysed in a buffer containing 50 mM Tris-HCl and 2% (w/v) SDS, supplemented with protease inhibitor cocktail (Sigma-Aldrich). The samples were boiled at 100°C for 5 min and spun down at 10,000 g for 15 min. The supernatant was collected and the protein concentration was determined using BCA assay (Pierce, Thermo Fisher Scientific). ^{35}S labeled samples were run in 4%–12% Nupage Bis-Tris gel (Life Technologies). As a loading control, gels were stained with SimplyBlue SafeStain (Life Technologies), and destained by incubation in water. Thereafter, the gels were incubated in 11.2% (v/v) salicylic acid and 10% glycerol (v/v) for 15 min. The gels were dried and exposed to a phosphor screen (Fujifilm) for a minimum of 48 hr. The screen was scanned using the phosphorimager BAS-2500 (Fujifilm) and ^{35}S incorporation was determined by measuring the intensity of each lane (MultiGauge Analysis Software v2.2, Fujifilm).

Free ^{35}S-Methionine/-Cysteine in the Cytosol

Rat primary neurons overexpressing either GFP or αSyn-GFP were incubated with ^{35}S-methionine and –cysteine at 100 μCi/ml for various durations. After a quick wash with cold PBS, cells were lysed in RIPA buffer for 20 min on ice and the debris was removed by centrifugation. Proteins in the lysates were precipitated by adding 1 volume 100% TCA to 4 volume of lysate and incubate 10 min at 4°C. After centrifugation at 14K rpm for 10 min, supernatant was collected to measure a cytosolic pool of free ^{35}S-methionine/-cysteine. ^{35}S incorporation was determined by quantifying using an LS 6500 liquid scintillation counter (Beckman Coulter) with 5 μL of the sample being immersed in 7 mL scintillation cocktail (National Diagnostics).

Cell Lysis and Endoglycosidase H Digestion

Cells were lysed in a buffer containing 20 mM HEPES, 150 mM NaCl, 10% (v/v) glycerol, 1 mM EGTA, 1.5 mM $MgCl_2$, 1% (v/v) Triton X-100, pH to 7.4, protease inhibitor cocktail (Sigma-Aldrich), and protein phosphatase inhibitor cocktail 1 and 2 (Sigma-Aldrich), and incubated in an ice/water slurry for 20 min, followed by 2 freeze-thaw cycles (−80°C/37°C, ~1 min each). Supernatant was collected after ultracentrifugation at 100,000 g, 4°C, for 30 min. Protein concentration was determined using BCA assay (Pierce, Thermo Fisher Scientific). Endoglycosidase (Endo) H (New England Biolabs) digestion was performed based on the manufacturer's instructions. Briefly, 20-40 μg bulk protein was assembled in 15.3 μL reaction volume; 1.7 μL denaturing buffer was added and samples were boiled for 10 min at 100°C. Then 2 μL of G5 buffer and 1 μL of Endo H or 1 μL H_2O were added to the denatured reaction and incubated for 2 hr at 37°C.

Western Blotting

For protein trafficking after Endo H digestion, protein samples were denatured in sample buffer (20 mM Tris-Cl pH 6.8, 4% (v/v) glycerol, 180 mM 2-mercaptoethanol, 0.0003% (v/v) bromophenol blue and 2% (v/v) SDS), run in 10% Tris-glycine gel, and wet transferred with 20% methanol onto PVDF membranes (Bio-Rad). Blots were blocked in a 1:1 dilution of Odyssey blocking buffer (Li-Cor Biosciences) and PBS for 1 hr at room temperature, followed by incubation with primary antibodies in a 1:1 dilution of Odyssey blocking buffer (Li-Cor Biosciences) and PBS containing 0.1% Tween 20 (PBST) at 4°C overnight with gentle rocking. After three 5 min washes with PBST, blots were incubated with secondary antibodies such as anti-mouse or -rabbit IgG conjugated to IRDye 680 or 800 (1:10,000, Rockland) in a 1:1 dilution of Odyssey blocking buffer and PBST for 2 hr at room temperature. After three 5 min washes with PBST and two with water, blots were scanned using the Odyssey quantitative fluorescent imaging system (Li-Cor Biosciences) and bands were quantitated using Odyssey Software v2.1 (Li-Cor Biosciences).

For other western blots, samples were lysed in RIPA buffer and run in either 8 or 10% Nupage Bis-Tris gel (Life Technologies) and transferred using iBlot (Life Technologies). Blocking was in 5% nonfat dry milk in PBST. As for the secondary antibodies and chemiluminescent detection, anti-mouse, -rabbit IgG or avidin conjugated to HRP was used with SuperSignal West Pico chemiluminescent substrate (Thermo Fisher Scientific).

TALE-TF Design

TALE-TFs were designed to target between 200bp upstream (5′) and 50bp downstream (3′) of the transcription start site (TSS) of ATXN2 or EIF2G transcripts. Within these regions near the TSS, we identified DNaseI hypersensitive regions from human ventromedial prefrontal cortex samples (PMID: 22955617). Within these DNaseI HS regions, we designed 5 TALE-TFs for each transcript.

Each TALE-TF was designed to target a 14bp genomic sequence consisting of an initial thymidine (T) plus 12 full repeats and 1 half repeat. For each TALE-TF, the TALE repeats were cloned into an rAAV transfer plasmid using a PCR-based, Golden Gate cloning strategy as previously described (Konermann et al., 2015; Sanjana et al., 2012; Tarazona et al., 2011). The rAAV transfer plasmid contained the TALE backbone fused to the synthetic VP64 activator domain along with a 2A-linked EGFP that is cleaved during translation.

TALE-TF Assembly

14-mer transcription activator-like effector transcription factors (TALE-TFs) were constructed using Golden Gate cloning as described previously (Sanjana et al., 2012). For each gene, *ATXN2* and *eIF4G1* (transcript variant 7), five different TALE-TFs were designed with the 14 bp long target loci being located in the proximal promoter region (*ATXN2* TALE-TF #1: 5′-TGTCCAGA TAAAGG-3′, #2: 5′-TGAACCTATGTTCC-3′, #3: 5′-TGCCAGATTCAGGG-3′, #4: 5′-TGGAGCGAGCGCCA-3′, #5: 5′-TAGCTGGT CATGGT-3′; *eIF4G1* TALE-TF #1: 5′-TGTCACGTGACGGG-3′, #2: 5′-TGTGGCTGTCACGT-3′, #3: 5′-TCAAAGTTCGGGAG-3′, #4: 5′-TCGCGGAACAGAGA-3′, #5: 5′-TCTCCTGCCTCAGC-3′). For each TALE-TF the correct sequence of the DNA-binding domain was verified by Sanger sequencing and all TALE-TF clones with non-silent mutations were excluded.

Ribosomal Footprint Profiling

For ribosome footprint profiling, 12-week old cells were treated with cycloheximide (100 ug/mL) for 5 min at 37°C to stop translation elongation. Cells were washed twice with ice-cold 9.5 mM PBS, pH 7.3, containing 100 μg ml^{-1} cycloheximide, and lysed by adding lysis buffer (10 mM Tris-HCl, pH 7.4, 5 mM MgCl2, 100 mM KCl, 2 mM dithiothreitol, 100 μg ml^{-1} cycloheximide, 1% Triton X-100, 500 U ml^{-1} RNasin Plus, and protease inhibitor (1x complete, EDTA-free, Roche)), scrapping cells from the plate, and then triturating four times with a 26-gauge needle. After centrifuging the crude lysate at 1,300*g* for 10 min at 4 °C, the supernatant was removed and flash-frozen in liquid nitrogen. The lysate was thawed on ice, after which ribosome profiling and mRNA-seq were performed as described previously (Subtelny et al., 2014) using a detailed protocol available at http://bartellab.wi.mit.edu/protocols.html. The 4-week old cells were washed twice with 37°C growth media, then after removing the media by aspiration the plates were sealed and then plunged into liquid nitrogen. Cells were then lysed with lysis buffer as described above, but cycloheximide was excluded from all solutions including the sucrose gradients. After thawing on ice, a small amount of cycloheximide-free zebrafish RPF lysate was spiked into the 4-week old cell lysates (10-fold less based on A_{260}) prior to digestion with RNase I.

RPF and RNA-seq tags were mapped to the ORFs, as described previously (Subtelny et al., 2014). To account for the zebrafish reads present in the 4-week old samples, indexes comprising both the zebrafish and human genomes or transcriptomes were created and these data were mapped to the combined indexes. Only reads mapping uniquely were considered, and those mapping to zebrafish were excluded from the analysis.

Enriched pathways in the translational profiling for the 4-week and 12-week datasets were computed with the Gene Set Enrichment Analysis tool, available at the Broad Institute website (http://software.broadinstitute.org/gsea/index.jsp).

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison with Existing Homology Prediction Approaches

To evaluate the functional association between yeast proteins and the predicted human homologs, we computed the average accuracy of Gene Ontology (GO) of the top 5 homologs predicted by our method, HHpred and BLAST (Altschul et al., 1990, 1997; Ashburner et al., 2000; Tuncbag et al., 2013) (Figure S3B). We chose the top 5 homologs since yeast proteins often have more than one good human homolog. The accuracy of a homolog was calculated as the percentage of overlapped GO labels between the yeast protein and the putative homolog. We noted that the number of assigned GO labels per gene varied considerably between yeast and human proteomes, so that the GO accuracy metric favored predicted homologs with a large number of labels and query proteins with a small number of GO labels, potentially biasing the analysis. Furthermore, false positives were not considered by this metric. To address these issues, we computed the widely used Jaccard similarity score, which is the number of overlapping GO labels divided by the total number of unique GO labels of the yeast (or human) gene and its human (or yeast) homolog. BLAST's accuracy for 4023 yeast proteins was 31.1%. HHpred in conjunction with BLAST achieved of 32.6% for accuracy for 4312 yeast proteins. Our method obtained 31.6% accuracy for a significantly greater number, 4923, of yeast proteins. It also outperformed BLAST on 4023 yeast proteins with BLAST output (32.0% versus 31.1% accuracy and 25.2% versus 24.3% Jaccard similarity) and HHpred on 4312 proteins with HHpred output (34.1% versus 32.6% accuracy and 26.9% versus 24.9% Jaccard Similarity). The improvements over BLAST and HHpred were significant (paired t test p values < 0.01).

We then tested our method on finding yeast homologs for human proteins (Figures 3C and 3D). The improvement of the coverage over BLAST and HHpred was even more substantial than for generating human homologs from yeast proteins. Our method predicted homologs for 15200 proteins but BLAST and HHpred only covered a relatively small portion of human proteome (7248 and 9577 respectively). Accuracy metrics also favored the DCA method. Our method improved the predictive power over BLAST (57.6% versus 57% accuracy and 26% versus 26.6% Jaccard similarity) and HHpred (56% versus 54.9% accuracy and 25% versus 24.2% Jaccard similarity) on proteins which BLAST or HHpred can find yeast homologs on both GO accuracy and Jaccard similarity score. These comparisons were all statistically significant (all p values < 0.01 by paired t test).

We also compared our homology tool to the state-of-the-art Ensembl Compara method. Ensembl Compara identifies high confidence homolog pairs through phylogenetic tree-based clustering and analysis across multiple species. This sequence-based method sacrifices coverage for accuracy, and these pairs are considered a gold standard for traditional analyses (Vilella et al., 2009). We downloaded the Ensembl Compara v85, and mapped gene ids to the gene names used in our homology tool, identifying 5093 high-confidence yeast/human pairs for 2409 yeast genes. Among these pairs, there are three major categories: "one-to-one," "one-to-many" and "many-to-many." To evaluate our DCA homology tool, we checked whether it performed at least as well for high-confidence yeast/human pairs, whether predicted as one-to-one, one-to-many or many-to-many by Ensembl Compara. Since orthology relationships between human and yeast genes can be ambiguous due to their remote evolutionary distance, DCA and Ensemble Compara may predict different putative homologs, especially for the many-to-many case. For such cases, we also computed the GO accuracy as the percentage of overlapping GO labels between a yeast protein and the predicted homolog. For clear one-to-one pairs by Ensembl Compara, DCA differed in only 25 of 1040 genes. Of those 25 genes that differed, our method achieved comparable accuracy in ontology prediction (0.394) as compared to Ensembl Compara (0.388) based on ontology matching. There were 1518 entries in the "many2many" prediction category. For these, our method achieved a correct pairing (0.414) equivalent to Ensembl Compara (0.412). Finally, for the yeast genes in which a one-to-many correspondence was predicted, there were 2535 entries. Again, our method identified homologs by gene ontology (0.391) similar to Ensembl Compara (0.390). Among the differences, we observed most of them to be similar genes within the same family; moreover, these differences are not statistically significant. Thus, our approach does not disrupt homology prediction for high-confidence orthology pairs, a surrogate for false-positivity in the absence of any other gold standard yeast-to-human homolog pairing. From these results, we demonstrated that DCA provides comparable yeast-to-human accuracy as Ensembl Compara for the same input yeast genes.

Recently, Kachroo et al.(Kachroo et al., 2015)carefully tested 414 essential yeast genes for complementation by homologs that were clear by sequence. Thus, for each of these 414 yeast/human gene pairs, the complementation assay provided a binary and experimentally strong readout of homology. Kachroo et al. developed a method to predict which of these high confidence pairs were likely to be actual positive complementation pairs. They utilized more than 100 features, including careful manual curation of sequence properties, network features, transcriptional and translational features, and expression abundances, to establish a predictive tool. They showed that this predictive tool could be trained on a subset of the experimentally tested yeast/human pairs to correctly identify functional replaceability in a separate test set. To demonstrate the effectiveness of integration of DCA, we trained a DCA-based classifier, using only sequence and network information, to predicted true yeast-human complementation pairs. In particular, to check the predictive power of our DCA pipeline, we built a classifier based on the low-dimensional gene vectors obtained from our joint DCA learning pipeline. For each pair of yeast/human genes, we built features based on the gene vectors to consider their sequence similarity and topological roles in their molecular networks. These features, including element-wise product and difference and sum of two gene vectors, were used as input to a gradient boosted forest classifier. We tested whether this classifier, for our more elemental, automated DCA tool based on only sequence/network features, could be tuned to also predict the functional complementation between yeast and human. When we trained our DCA classifier via 5-fold cross-validation on the yeast/human pairs from Kachroo et al. we achieved a high rate of prediction accuracy (AUC = 0.82, SD = 0.08). This was comparable to the intricate, manual integrated method of Kachroo et al., demonstrating that our automated homology tool, based on only sequence and network topology, is sufficient for training a classifier for this specific homology task. It is worth noting that methods utilizing sequence-similarity alone, including BLAST and HHpred, performed considerably worse than DCA (0.70 and 0.69, respectively). It is clear that our DCA-based classifier, which effectively integrates network topology and sequence similarity, is just as effective as the method in Kachroo et al. that utilizes more than 100 features, thus overcoming the barrier of major time-consuming manual feature curation.

Evaluation of PCSF and Humanized Steiner Networks

We tested PCSF on two separate datasets and demonstrate vastly superior performance when compared to existing methods. For comparison, we identified two popular algorithms, DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009), and implemented them. Both methods take seed genes and identify subnetworks that span the seed genes to reveal possible functional interconnectedness of these genes. The first algorithm, DAPPLE, identifies significant direct and one-hop indirect edges in the human interactome to connect as many seed genes as possible. The second algorithm, PEXA, utilizes existing pathway annotations, such as KEGG or Reactome, to cover seed genes. Merging and pruning are then applied to link connected components and remove hanging genes. For these comparisons, we provided each algorithm with yeast-to-human homology links and injected yeast interaction edges into the human network, just as we provide for our PCSF method. For DAPPLE, we used the predicted dense network with significant one-hop indirect edges, since the sparse direct network is not able to identify hidden genes. We curated hits from 15 complete screens in yeast (Tong et al., 2004). In these screens, a gene is deleted as well as its genetic interactors or modifiers. We used these genetic modifiers as input for the network algorithms. The inactivated gene was hidden from the algorithm, and was used to evaluate the predicted network. Taking cues from previously-published methods (Yeger-Lotem et al., 2009), here we considered an algorithm successful in discovering the cellular response if the predicted hidden human genes were significantly enriched for specific gene ontology biological process terms attributed to the hidden inactivated yeast gene (hypergeometric test; p value < 0.01). We generated humanized networks with PCSF, and two alternative methods: DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009). For these screens, the success rate of PCSF was 47%, as compared to DAPPLE and PEXA which were 6.6% and 13%, respectively. These results suggest superior performance of PCSF over DAPPLE and PEXA.

To better understand the relevance of genes and predicted pathways recovered by PCSF, DAPPLE and PEXA, we designed a well-controlled simulation. To mimic genetic screens of perturbed pathways, we selected individual pathways from the well-known human pathway database KEGG and identified all genes in each pathway (Table S15). We then identified yeast homologs via stringent Ensembl one-to-one mapping. We treated those human genes with clear yeast homologs as "perturbed" and picked their homologs' genetic interaction neighboring genes as hits from a "virtual yeast genetic screen." Virtual screens like these minimize experimental noise as a confounding factor and enable cleaner evaluation of algorithm performance. Since we know the "true" pathway information, this method can be used to test the sensitivity and specificity of algorithms by quantifying how often "relevant" genes in the original KEGG pathway are recovered as predicted (non-seed) genes. We chose 50 KEGG pathways that had at least 5 human genes with clear yeast homologs and created 50 associated "virtual" screens for testing (Table S15). We used two performance metrics: precision, i.e., the percentage of predicted hidden genes shown in the original KEGG pathway, and recall, i.e., the percentage of the original KEGG genes shown as hidden nodes in the predicted pathway. Ideally, these values would be 100% for perfect predictions. For PCSF, the average precision and recall values are 63% and 74% resp. In contrast, for DAPPLE, the average precision and recall values are 6% and 47% resp., whereas for PEXA, they are 8% and 83% resp. The differences between three precision values are substantial: PCSF has much higher precision within very compact subnetworks, while both DAPPLE and PEXA predict huge "hair ball" networks with low precision. It is worth noting that PEXA has a very high recall value likely because it uses the KEGG pathways to build networks, and thus predictably has high recall (because the simulated screens here are generated from KEGG pathways); however, its precision metric is very low.

Further, we tested the effectiveness of injected yeast genetic interactions into networks through the simulated yeast genetic screens we generated, and cross-compare our PCSF method with the other algorithms, DAPPLE and PEXA. First, we tested performance by removing all injected yeast interactions. For PCSF, the average precision and recall values are 37% and 54% resp. For DAPPLE, the average precision and recall values are 8% and 27% resp. Compared to the precision and recall results (i.e., 63% and 74% for PCSF versus 6% and 47% for DAPPLE), it is clear that both PCSF and DAPPLE have much lower recall if yeast interactions are excluded. This analysis thus confirms with data that injection of yeast interactions into "humanized" networks provide key connections between genetic modifiers to the perturbed genes. For PEXA, the average precision value is 9%, similar to that with yeast injection, whereas the recall rate is again predictably very high. Second, we tested the effects of randomly removing a portion of injected genetic interactions over 10 trials. The average precision and recall values are shown in Figure S4, demonstrating the relationship between the accuracy of these methods and the percentage of injected yeast interactions. A notable observation is that the performance becomes reasonable when >40% of interactions are injected. The performance of PEXA remains relatively unchanged because it utilizes the human KEGG pathway information in its algorithm, as noted above. In terms of false-positives and -negatives, there is clearly a trade-off between the different methods. PCSF works best for our current work, as PCSF identifies a small set of relevant genes for cost-effective experimental explorations.

Statistical Methods and Data Analysis for Cell-Based Assays

Sample sizes for all experimentation were chosen based on our previous extensive experience with the methods and assays in these studies. For most experiments in mammalian cells, robustness and consistency of the results are typically established after three biological replicates are analyzed. Unless otherwise stated in the figure legends, this was the standard number of replicates required for all experiments. For all human and rat cellular experiments, significance was then determined by appropriate statistical tests that are standard in the field. The two-tail t test was applied when there were only two conditions to compare within the experiments. One-Way ANOVA with a multiple comparisons post hoc test was performed when experiments include multiple conditions. Data points were excluded based on the following pre-established criteria: 1) errors were introduced to the particular sample while performing the experiments, 2) the values are greater or less than two standard deviation from the mean. For yeast spot assays, results were considered significant when three biological replicates (unless otherwise stated) demonstrated the same trend by eye. Methods used for Figure 5E are outlined in the figure legend. For the pooled screen yeast assay (Figures 4A and S6) detailed statistical methods for reads and cutoff thresholds are supplied above in the methods. The statistical methods for the computational analysis are described in detail in the methods sections above.

DATA AND SOFTWARE AVAILABILITY

All cytoscape files depicting networks are provided in the supplement and referred to in the text. The TransposeNet pipeline is described at http://transposenet.csail.mit.edu. The DCA/Mashup web portal is http://mashup.csail.mit.edu. The PCSF web portal is http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/.