Integrating protein-protein interaction networks with phenotypes reveals signs of interactions

Arunachalam Vinayagam[1], Jonathan Zirin[1], Charles Roesel[2,3,5], Yanhui Hu[1,2], Bahar Yilmazel[2,3], Anastasia A Samsonova[1], Ralph A Neumüller[1], Stephanie E Mohr[1,2] & Norbert Perrimon[1,4]

A major objective of systems biology is to organize molecular interactions as networks and to characterize information flow within networks. We describe a computational framework to integrate protein-protein interaction (PPI) networks and genetic screens to predict the 'signs' of interactions (i.e., activation-inhibition relationships). We constructed a *Drosophila melanogaster* signed PPI network consisting of 6,125 signed PPIs connecting 3,352 proteins that can be used to identify positive and negative regulators of signaling pathways and protein complexes. We identified an unexpected role for the metabolic enzymes enolase and aldo-keto reductase as positive and negative regulators of proteolysis, respectively. Characterization of the activation-inhibition relationships between physically interacting proteins within signaling pathways will affect our understanding of many biological functions, including signal transduction and mechanisms of disease.

© 2013 Nature America, Inc. All rights reserved.

Objectives of systems biology research include organizing molecular interactions as networks and characterizing their structure, dynamics and controllability. Tremendous progress has been made using 'omics' data sets to identify the parts and connections of these networks. For example: PPIs, identified from yeast-two hybrid or affinity purification–mass spectrometry (AP-MS) approaches, have provided information on the biophysical interactions occurring between two or more proteins[1–5]. Similarly, systematic loss-of-function analyses such as RNAi screens have identified sets of genes implicated in specific biological processes[6]. Integration of omics data sets and inferring information flow are critical aspects of the reconstruction of signaling networks[7]. Such reconstructions reveal how proteins communicate and coordinate cellular functions, and they allow researchers to explore the emergent properties of networks.

There is a need for systematic approaches to infer causal relationships between interacting proteins, by which we refer to the 'direction' (edge direction), 'sign' (activation or inhibition) and 'mode' (such as phosphorylation or ubiquitination) of signal flow in PPI networks. Genome-scale reconstruction of signaling networks remains a challenge[8] largely because of the difficulty of predicting such causal relationships, although small-scale networks have been successfully reconstructed. Furthermore, databases of signaling pathways are incomplete, and annotations are inconsistent across databases[9]. Recent studies have attempted to infer direction of information flow[10–14] and to reconstruct kinase-substrate networks[15], but few attempts have been made to predict activation-inhibition relationships among interacting proteins.

Here we have developed a computational framework to predict the signs (positive or negative) of physical interactions using RNAi screens. In a positive PPI, proteins A and B interact to form a functional complex in which A activates B (or vice versa). In a negative PPI, proteins A and B interact to form a protein complex in which A inhibits B (or vice versa), such that one of the proteins is a negative regulator of the complex. We applied this framework to construct a *D. melanogaster* signed PPI network and thereby identified unexpected roles for the metabolic enzymes enolase and aldo-keto reductase (AKR) as positive and negative regulators, respectively, of proteolysis in *Drosophila*. Finally, we built a database, the signed protein-protein interaction network (SignedPPI), to access, build and navigate signed interaction networks (http://www.flyrnai.org/SignedPPI/; **Supplementary Software**).

RESULTS

Development of a signed prediction framework

We compiled RNAi screens recording 42 phenotypes from various resources including the *Drosophila* RNAi Screening Center[16], GenomeRNAi[17], Neuroblasts Screen online databases[18] and Bristle Screen online database[19] (Online Methods and **Supplementary Table 1**). We also included results from an image-based RNAi screen measuring nucleolus size[20] and six other phenotypes (R.A.N. and N.P., unpublished data). With respect to the hits, the screens showed an average 14% similarity with each other (**Supplementary Fig. 1**). Each screen identifies positive and negative regulators of a particular phenotype, allowing us to construct a phenotypic matrix in which the rows correspond to genes and columns correspond to 49 different phenotypes (**Fig. 1a**);

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [2]Drosophila RNAi Screening Center, Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [3]Bioinformatics Program, Northeastern University, Boston, Massachusetts, USA. [4]Howard Hughes Medical Institute, Boston, Massachusetts, USA. [5]Present address: Marine Science Center, Northeastern University, Nahant, Massachusetts, USA. Correspondence should be addressed to A.V. (vinu@genetics.med.harvard.edu) or N.P. (perrimon@receptor.med.harvard.edu).

RECEIVED 6 JUNE; ACCEPTED 25 OCTOBER; PUBLISHED ONLINE 17 NOVEMBER 2013; DOI:10.1038/NMETH.2733

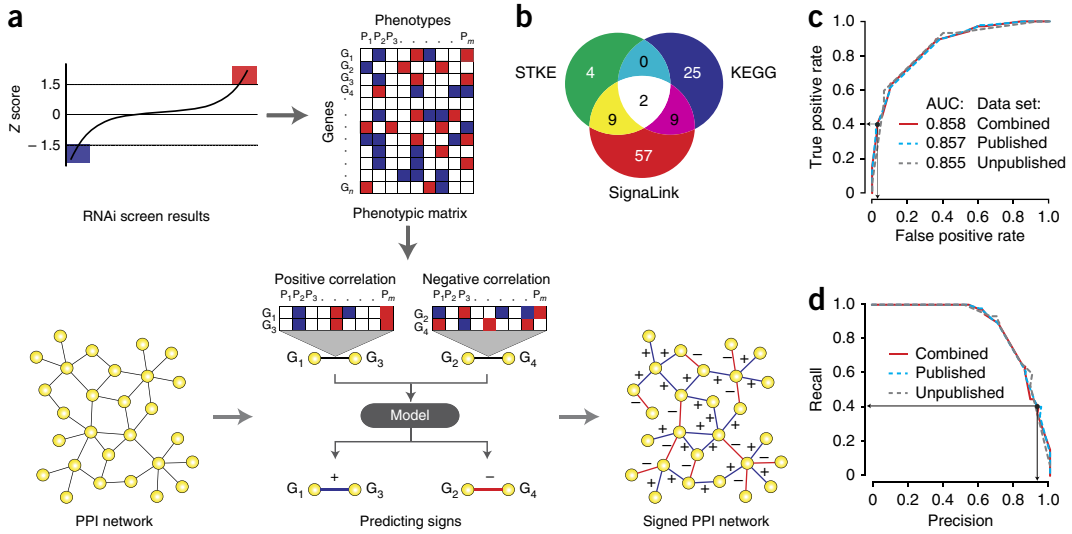

Figure 1 | Framework to predict the signs of protein interactions. (**a**) Schematic representation of the framework. (**b**) Sources of signaling PPIs with known edge signs (the number of PPIs from each is listed). STKE, Database of Cell Signaling (via the Signal Transduction Knowledge Environment); KEGG, Kyoto Encyclopedia of Genes and Genomes. (**c**,**d**) Receiver operating characteristic (ROC) plot (**c**) and precision-recall curve (**d**) showing the performance of the sign-prediction model. Black dots and the arrows show the chosen S_{sign} cutoff ($S_{sign} \geq 1$ or $S_{sign} \leq -1$). AUC, area under the ROC curve.

positive and negative regulators are color coded differently. Next we use a simple correlation of phenotypes to predict activation-inhibition relationships, with positive correlations occurring when both genes have the same color and negative correlation when they have different colors. We compute a sign score (S_{sign}) when both of the interacting proteins in a pair score a nonzero value in two or more screens (**Fig. 1a** and Online Methods). The sign score determines whether the phenotypes have positive or negative correlations. We predict a positive edge sign (activation) if the S_{sign} is positive and a negative edge sign (inhibition) if the S_{sign} is negative.

We used interactions with known activation-inhibitory relations from the literature to test our model and find an appropriate cutoff value for the sign score. We compiled such interactions from signaling pathway databases such as SignaLink[21], the Database of Cell Signaling (http://stke.sciencemag.org/cm/) and the Kyoto Encyclopedia of Genes and Genomes (KEGG)[22] (**Supplementary Table 2** and Online Methods). We selected 106 literature-based interactions in which both proteins scored in two or more RNAi screens, and we defined these as the positive reference set (PRS) (**Fig. 1b** and **Supplementary Table 3**). By reversing the original signs, we created a negative reference set (NRS; 106 interactions; see Online Methods). Next we used our model to predict signs and compared the results to the original annotations. We used three different RNAi data sets to assess the robustness of our model: published (42 phenotypes), unpublished (7 phenotypes) and combined (49 phenotypes). The results show that our model has good predictive power (area under the receiver operating characteristic curve = 0.858) and is robust to various subsets of RNAi screen data (**Fig. 1c**,**d** and **Supplementary Table 4**). However, the predictive performance of sign score was impaired with a randomized phenotype matrix (**Supplementary Fig. 2**). We found that the performance was comparable between a subset of reference interactions with respect to the source database or considering only positive or negative interaction signs (**Supplementary Fig. 3**). A minimum of nine RNAi screens was needed to make a reliable prediction, but coverage increased with an increasing number of screens (**Supplementary Fig. 4a**,**b**). Further, the relationship between the RNAi screens and number of hits in RNAi screens influenced coverage but not the predictive performance (**Supplementary Fig. 4c**–**f**).

We identified an appropriate sign score cutoff value of $S_{sign} \geq 1$ for positive signs and $S_{sign} \leq -1$ for negative signs (**Fig. 1c**,**d**). At this cutoff value, we achieved 90% precision and 41% recall (2.8% false positive rate and 59% false negative rate). Note that we compared the performance of this simple model to various classifiers trained to predict signs using the phenotype matrix as features and found that the simple model performed better (**Supplementary Table 5**).

Constructing a signed *Drosophila* PPI network

We collected PPIs from major databases such as BioGrid[23], IntAct[24], DIP[25], MINT[26], DroID[27] and DPiM[1] (**Supplementary Table 6**), selecting PPIs identified as binary interactions (for example, from yeast two-hybrid screens), high-confidence AP-MS interactions and AP-MS interactions predicted to be direct interactions (Online Methods). The resulting integrated *Drosophila* network consists of 47,293 PPIs among 9,107 proteins. We next predicted signs for these *Drosophila* PPIs on the basis of the 49 phenotypic data sets. The signed network consists of 6,125 PPIs connecting 3,352 proteins, among which 4,135 PPIs are positive interactions and 1,990 PPIs are negative (**Fig. 2a** and **Supplementary Table 7**), with the sign score of each interaction indicating the confidence of the predicted sign. Our predicted sign network consists of 13-fold more interactions than literature-based signed interactions (434 PPIs).

We systematically analyzed various properties of the entire signed network and of subnetworks consisting of only positive or negative interactions. We observed a positive correlation between the number of phenotypes regulated by a gene and the number of interactions ('degree') of the gene (**Supplementary Fig. 5**). Further, subnetworks with positive and negative interactions showed a

© 2013 Nature America, Inc. All rights reserved.

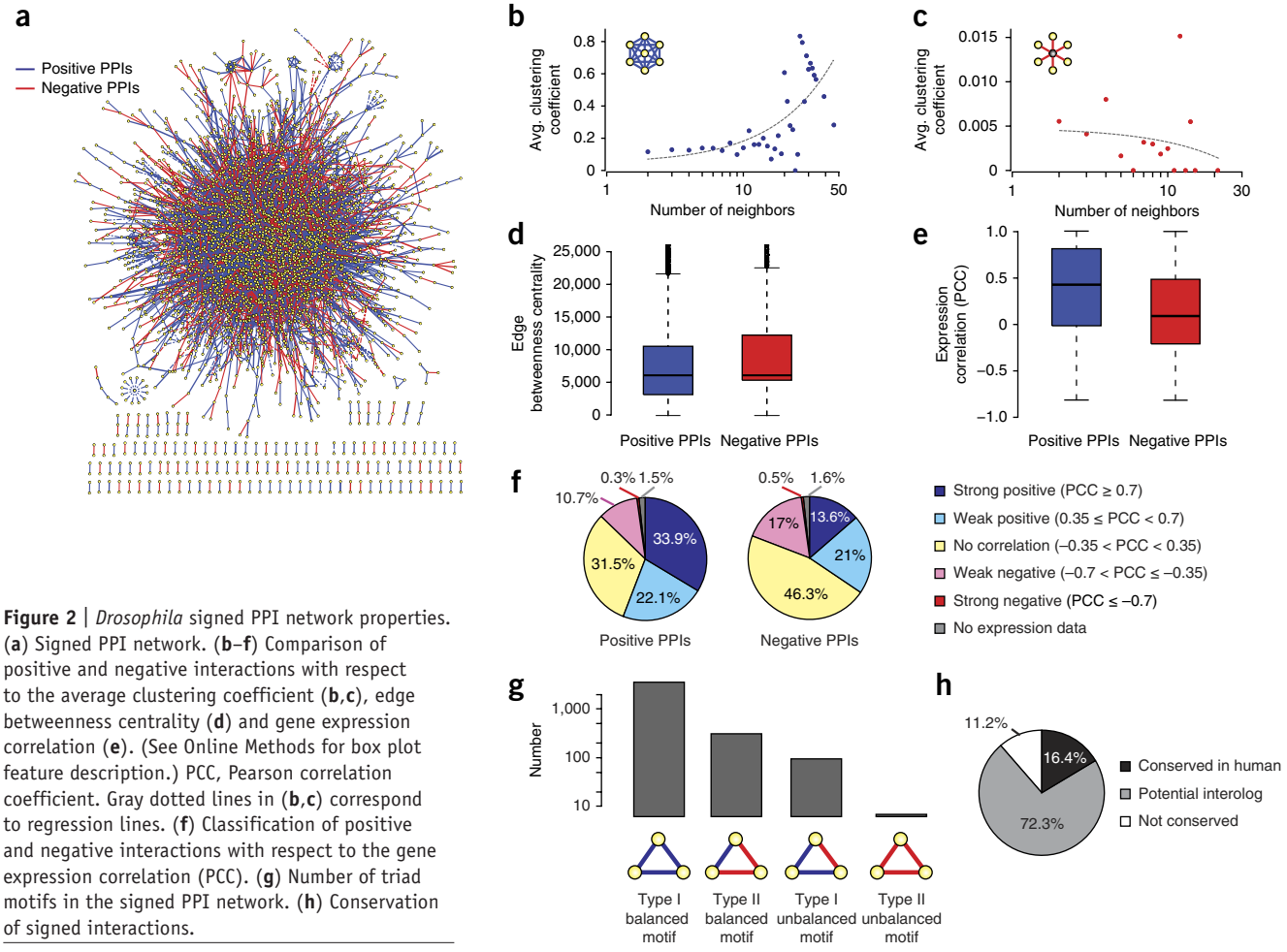

Figure 2 | *Drosophila* signed PPI network properties. (**a**) Signed PPI network. (**b**–**f**) Comparison of positive and negative interactions with respect to the average clustering coefficient (**b**,**c**), edge betweenness centrality (**d**) and gene expression correlation (**e**). (See Online Methods for box plot feature description.) PCC, Pearson correlation coefficient. Gray dotted lines in (**b**,**c**) correspond to regression lines. (**f**) Classification of positive and negative interactions with respect to the gene expression correlation (PCC). (**g**) Number of triad motifs in the signed PPI network. (**h**) Conservation of signed interactions.

similar degree distribution (**Supplementary Fig. 6**). Functional-enrichment analysis revealed that kinases tend to be hubs with similar proportions of positive and negative interactions, whereas transcription co-regulators tend to be hubs with primarily positive interactions and transcriptional co-repressors are hubs with primarily negative interactions (**Supplementary Table 8**).

Correlating the number of neighbors and average clustering coefficient showed that hubs with positive interactions tend to cluster (**Fig. 2b**,**c**). Next we compared positive and negative interactions with respect to the edge 'betweenness centrality', a measure based on the number of shortest paths that passes through an edge in the network. The intermodular interactions, bridging different biological processes, tend to have high edge betweenness-centrality scores, whereas intramodular interactions, such as interactions within a protein complex, tend to have low edge betweenness centrality[28]. Our analysis revealed that negative interactions tend to have high edge betweenness centrality, meaning that that they are likely to be intermodular interactions, in contrast to positive interactions, which are likely intramodular interactions ($P = 2.2 \times 10^{-16}$, Wilcoxon test; **Fig. 2d**).

Comparison of edge signs with gene expression from *Drosophila* developmental time-course data[29] revealed that the positive interactions are more likely to show positive expression correlation (Pearson correlation coefficient, PCC) than are negative interactions (P value = 2.2×10^{-16}, Wilcoxon test; **Fig. 2e** and **Supplementary Fig. 7**). Although half of the negative interactions showed no expression correlation (–0.35 < PCC < 0.35), ~13.6% of these interactions had strong positive expression correlations (**Fig. 2f**), which suggests potential tight negative regulation. Furthermore, over half of the positive interactions showed positive expression correlation.

Including signs on a PPI network allows the application of structural balance theory, which is based on the ratios of balanced and unbalanced triad motifs, enabling us to measure the stability of the network in a given condition[30]. In a triad motif, if the product of the signs is positive, the motif is defined as a balanced motif. Our analysis reveals that, similarly to social networks, signed PPI networks have more balanced than unbalanced motifs (**Fig. 2g**, **Supplementary Table 9** and Online Methods). Unbalanced motifs are particularly interesting because they are highly dynamic and unstable. For instance, type I unbalanced motifs, consisting of two positive and one negative interaction, could potentially function as negative feedback loops or incoherent feed-forward loops, which are both associated with adaptation responses and are crucial for system controllability[31]. We identified 95 type I unbalanced motifs in the signed network (**Supplementary Table 10**). The network can also be used to systematically explore larger unbalanced motifs (four nodes or more) that could contribute to the network dynamics. Finally, 16% of the signed *Drosophila* interactions are conserved in human and another 72% are potential human interologs (**Fig. 2h**). 35% of the conserved interactions are linked directly to human disease

© 2013 Nature America, Inc. All rights reserved.

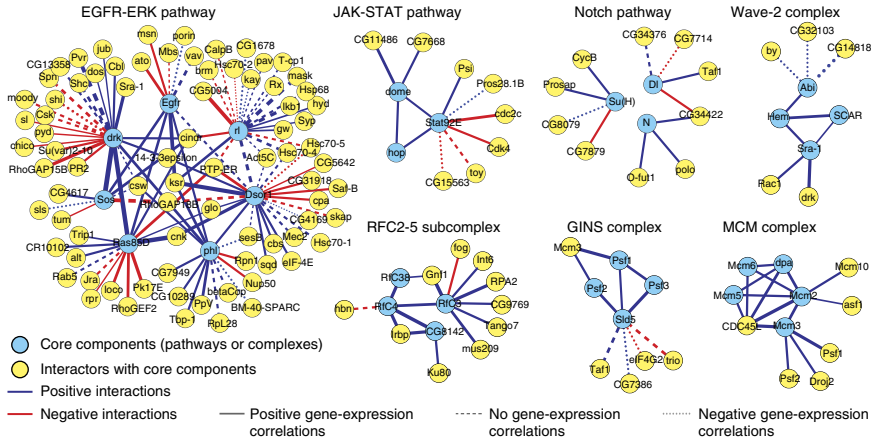

Figure 3 | Network representation of signed PPIs connecting known signaling pathways and protein complexes.

proteins, thereby implying the signed network is relevant to human diseases (**Supplementary Table 11**).

We constructed subnetworks focusing on major signaling pathways[9] and known protein complexes[32] to explore novel (previously unidentified) positive and negative interactions linked to core components (**Fig. 3** and **Supplementary Fig. 8**). Compared to other signaling pathways, the EGFR-ERK pathway is densely connected because of the availability of both PPI networks and functional data sets for this pathway[6,33]. Consistent with our prior results (**Fig. 2b**), members of the same protein complexes are primarily connected to one another via positive interactions.

Validating novel regulators of the proteasome

We selected the proteasome complex to further investigate positive and negative interactions inferred using signed PPI networks. The proteasome is essential for regulating proteostasis via degradation of proteins modified by polyubiquitin. Moreover, deregulation of proteasome function is relevant to many human diseases, including neurodegeneration, cancer and cachexia[34].

First we constructed a proteasome subnetwork with 51 nodes, including 29 proteins that are part of the proteasome complex and 22 proteins that interact with it (**Fig. 4a**). We selected these 22 interacting proteins on the basis of high-confidence positive or negative interactions with the proteasome ($S_{sign} \geq 1.73$ and $S_{sign} \leq -1.73$, respectively) and for which we have three or more independent RNAi reagents for tissue culture experiments. Next we knocked down the genes encoding these 51 selected proteins (the 29 proteasome components served as controls) in primary embryonic *D. melanogaster* muscle cells and assessed the effect on proteasome activity by measuring the accumulation of ubiquitinated proteins (**Fig. 4b** and **Supplementary Data**). Knocking down positive regulators should increase the accumulation of ubiquitinated proteins, whereas knocking down the negative regulators should decrease their accumulation. We identified genes as proteasome regulators if two or more independent RNAi reagents met the fold-change cutoff (Online Methods and **Supplementary Table 12**). Out of ten putative hits, tests in primary cell culture showed that eight of them could regulate the proteasome in a manner that was consistent with the predicted edge signs (**Fig. 4b**,**c** and **Supplementary Table 12**).

Next we used a luminescence assay that measures the protease activities associated with the proteasome complex in S2R+ cultured cells to further validate our putative regulators. We observed that knockdown of candidate positive regulators (*Enolase* (*Eno*), *polo*, *Hsc70Cb* and *Pomp*) with independent RNAi reagents decreased proteasome activity (**Fig. 4d** and **Supplementary Table 13**), whereas knockdown of the candidate negative regulators (*MRP*, *CG32039*, *CG15717* and *CG10638)* increased this activity (**Fig. 4d**). Knockdown efficiency of the RNAi reagents determined by qPCR is shown in **Supplementary Figure 9**. Our experiments, together with the evidence of direct physical interactions (**Supplementary Table 14**), show that the hits we identified using the signed PPI network are bona fide candidate regulators of the proteasome.

To further validate these regulators *in vivo*, we selected two metabolic enzymes *Eno* and *CG10638*, for which two or more transgenic RNAi fly lines were available. Using the muscle-specific *Mef2* (*Dmef2*)-*GAL4* line to drive the expression of the upstream activating sequence–RNAi hairpins, we assayed the formation of ubiquitinated protein aggregates in the RNAi-treated muscles compared to in control *Rpn1* (encoding regulatory particle non-ATPase 1, a proteasome component) and *white* RNAi knockdown samples. All RNAi constructs targeting *Eno* (three of three) and *CG10638* (two of two) gave consistent phenotypes. In agreement with our predicted signs, knockdown of *Eno* resulted in an increase in ubiquitin aggregates, whereas knockdown of *CG10638* led to a decrease in aggregates (**Fig. 4e**). Eno is a multifunctional protein with a key role in glycolysis[35], and its role in proteasome regulation had not previously been established. *CG10638* is an AKR family member. AKRs catalyze the NADPH-dependent reduction of aldehydes and ketones to alcohols[36]. A subset of mammalian AKRs have previously been shown to be regulated by drug-induced proteasome inhibition[37]. However, direct regulation of the proteasome or proteolysis by an AKR has not been previously reported. We note that knockdown of *Eno* or *CG10638* had no effect on ubiquitin gene expression in S2R+ cell lines (**Supplementary Fig. 10**). Altogether, using the proteasome complex as an example, our results demonstrate the usefulness of predicted edge signs to discover protein function.

Database to navigate signed PPI networks

We created the SignedPPI database (http://www.flyrnai.org/SignedPPI/) to build and navigate signed interaction networks (**Supplementary Fig. 11**). In addition to focusing on PPI networks, we also predicted signs for *Drosophila* functional interaction derived from the STRING (search tool for the retrieval of interacting genes/proteins) database[38]. We successfully predicted signs for 40,216 functional interactions, including 31,178 positive and 9,038 negative interactions; these data are accessible via the SignedPPI database (**Supplementary Table 15**). We created a prediction tool called SignPredictor that accepts a phenotype matrix and PPIs as input and predicts signed PPIs (**Supplementary Software**).

© 2013 Nature America, Inc. All rights reserved.

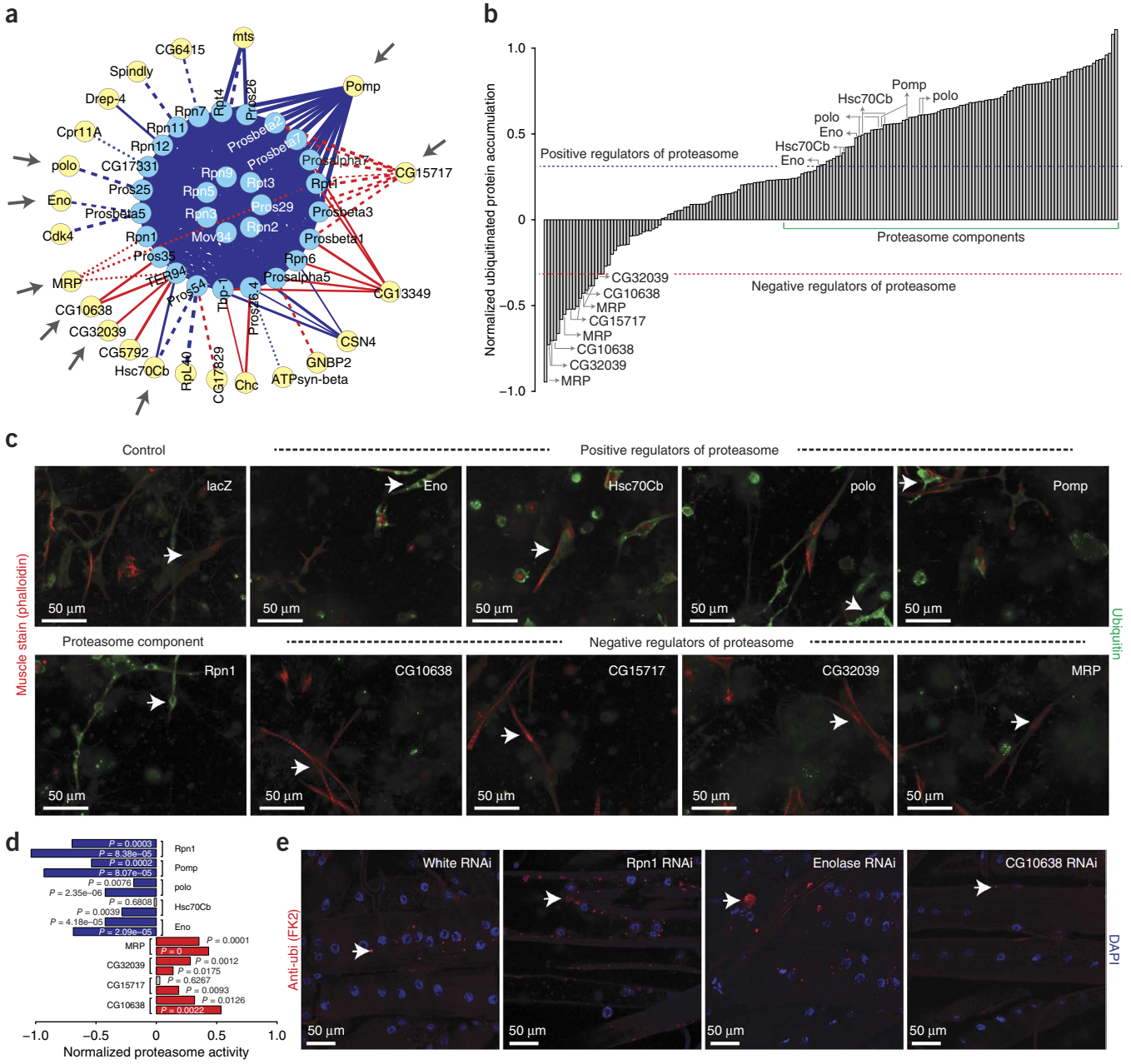

Figure 4 | Validation of predicted proteasome regulators. (**a**) Subnetwork of a proteasome complex (as in **Fig. 3**). Gray arrows highlight subsequently validated proteins. (**b**) Results from the image-based RNAi screen measuring the accumulation of ubiquitinated proteins in primary muscle cells in which regulators shown in **a** have been knocked down with RNAi. Blue and red dotted lines indicate the cutoff values used for positive and negative regulators, respectively. The green line highlights the region corresponding to most proteasome core components. (**c**) Muscle cells stained with phalloidin (red) and anti-ubiquitin (green); the indicated candidate regulators have been knocked down with RNAi. Arrows point to ubiquitinated proteins in cells. (**d**) Enzymatic activity of the proteasome upon the indicated RNAi treatment in S2R+ cultured cells. Blue and red bars corresponds to significant reductions and increases in proteasome activity, respectively. Independent RNAi reagents are shown for each gene. (**e**) Third instar larval longitudinal muscles expressing the indicated RNAi hairpins under control of the muscle-specific driver line *Mef2-GAL4* (red, anti-ubiquitin (FK2); blue, DAPI). Arrows point to ubiquitin-labeled aggregates.

The tool is implemented as a Perl module that can be downloaded from the SignedPPI database and installed locally.

DISCUSSION

Unlike previous studies that have used genetic interaction correlation[39] or phenotype similarity[40] to predict functional interactions, we used phenotype correlation to predict the function of physical interactions (signs). Our method is robust with respect to inherent noise in RNAi screens and has high predictive power. It is limited, however, to predicting context-dependent signs such as asymmetric bidirectional signs (for example, the negative feedback loop between Cdc2 and the anaphase-promoting complex (APC), wherein Cdc2 activates the APC, which in turn inactivates Cdc2).

The signed network we constructed for *Drosophila* covers only ~10% of known PPIs owing to the limited number of RNAi data

© 2013 Nature America, Inc. All rights reserved.

sets available. As more RNAi screens become available, it will become possible to further expand the scope and utility of the constructed network. In addition, the resource will benefit from additional PPI data sets such as ongoing interactome mapping projects[41]. Finally, the sign-prediction approach could be easily applied to other species.

Our analysis of the *Drosophila* signed network revealed insights into the design principles of network organization and identified unexpected roles for two metabolic enzymes, Eno and AKR, in regulating proteasome function. The signed network opens new possibilities for network analysis, such as the application of structural balance theory. Further integration of other information-flow properties such as edge direction would enable sophisticated flow-based network analysis.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank B. Housden, Y. Kwon, I.T. Flockhart and S. Rajagopal for helpful suggestions for tool development and manuscript preparation. This work was financially supported by P01-CA120964, R01-GM067761 and R01DK088718. S.E.M. is also supported in part by the Dana-Farber/Harvard Cancer Center (P30-CA06516). N.P. is supported by the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

A.V. and N.P. conceived of and designed the project. A.V. developed the computational method and built and analyzed the signed network. J.Z. designed and performed experimental validation. C.R. and B.Y. developed the SignedPPI database. A.V. and S.E.M. coordinated the development of the SignedPPI database. A.V. and Y.H. compiled RNAi screens. A.A.S. analyzed modENCODE expression data. R.A.N. contributed the unpublished RNAi screen. All authors contributed to manuscript preparation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Guruharsha, K.G. *et al.* A protein complex network of *Drosophila melanogaster*. *Cell* **147**, 690–703 (2011).
2. Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
3. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
4. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
5. Rual, J.F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
6. Friedman, A. & Perrimon, N. Genetic screening for signal transduction in the era of network biology. *Cell* **128**, 225–231 (2007).
7. Kholodenko, B., Yaffe, M.B. & Kolch, W. Computational approaches for analyzing information flow in biological networks. *Sci. Signal.* **5**, re1 (2012).
8. Hyduke, D.R. & Palsson, B.O. Towards genome-scale signalling network reconstructions. *Nat. Rev. Genet.* **11**, 297–307 (2010).
9. Kirouac, D.C. *et al.* Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. *BMC Syst. Biol.* **6**, 29 (2012).
10. Vinayagam, A. *et al.* A directed protein interaction network for investigating intracellular signal transduction. *Sci. Signal.* **4**, rs8 (2011).
11. Liu, W. *et al.* Proteome-wide prediction of signal flow direction in protein interaction networks based on interacting domains. *Mol. Cell Proteomics* **8**, 2063–2070 (2009).
12. Yeger-Lotem, E. *et al.* Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat. Genet.* **41**, 316–323 (2009).
13. Gitter, A., Carmi, M., Barkai, N. & Bar-Joseph, Z. Linking the signaling cascades and dynamic regulatory networks controlling stress responses. *Genome Res.* **23**, 365–376 (2013).
14. Singh, R. Algorithms for the analysis of protein interaction networks. Ch. 6, 91–102 PhD thesis, Massachusetts Institute of Technology (2012).
15. Linding, R. *et al.* Systematic discovery of *in vivo* phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
16. Flockhart, I.T. *et al.* FlyRNAi.org—the database of the *Drosophila* RNAi screening center: 2012 update. *Nucleic Acids Res.* **40**, D715–D719 (2012).
17. Gilsdorf, M. *et al.* GenomeRNAi: a database for cell-based RNAi phenotypes. 2009 update. *Nucleic Acids Res.* **38**, D448–D452 (2010).
18. Neumüller, R.A. *et al.* Genome-wide analysis of self-renewal in *Drosophila* neural stem cells by transgenic RNAi. *Cell Stem Cell* **8**, 580–593 (2011).
19. Mummery-Widmer, J.L. *et al.* Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. *Nature* **458**, 987–992 (2009).
20. Neumüller, R.A. *et al.* Conserved regulators of nucleolar size revealed by global phenotypic analyses. *Sci. Signal.* **6**, ra70 (2013).
21. Korcsmáros, T. *et al.* Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* **26**, 2042–2050 (2010).
22. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
23. Stark, C. *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* **39**, D698–D704 (2011).
24. Kerrien, S. *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40**, D841–D846 (2012).
25. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32**, D449–D451 (2004).
26. Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **40**, D857–D861 (2012).
27. Murali, T. *et al.* DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res.* **39**, D736–D743 (2011).
28. Girvan, M. & Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002).
29. Graveley, B.R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
30. Facchetti, G., Iacono, G. & Altafini, C. Computing global structural balance in large-scale signed social networks. *Proc. Natl. Acad. Sci. USA* **108**, 20953–20958 (2011).
31. Ma, W., Trusina, A., El-Samad, H., Lim, W.A. & Tang, C. Defining network topologies that can achieve biochemical adaptation. *Cell* **138**, 760–773 (2009).
32. Vinayagam, A. *et al.* Protein complex-based analysis framework for high-throughput data sets. *Sci. Signal.* **6**, rs5 (2013).
33. Friedman, A.A. *et al.* Proteomic and functional genomic landscape of receptor tyrosine kinase and ras to extracellular signal-regulated kinase signaling. *Sci. Signal.* **4**, rs10 (2011).
34. Dahlmann, B. Role of proteasomes in disease. *BMC Biochem.* **8** (suppl. 1), S3 (2007).
35. Díaz -Ramos, A., Roig-Borrellas, A., García-Melero, A. & López-Alemany, R. α-Enolase, a multifunctional protein: its role on pathophysiological situations. *J. Biomed. Biotechnol.* **2012**, 156795 (2012).
36. Mindnich, R.D. & Penning, T.M. Aldo-keto reductase (AKR) superfamily: genomics and annotation. *Hum. Genomics* **3**, 362–370 (2009).
37. Ebert, B., Kisiela, M., Wsól, V. & Maser, E. Proteasome inhibitors MG-132 and bortezomib induce AKR1C1, AKR1C3, AKR1B1, and AKR1B10 in human colon cancer cell lines SW-480 and HT-29. *Chem. Biol. Interact.* **191**, 239–249 (2011).
38. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2013).
39. Collins, S.R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
40. Fuchs, F. *et al.* Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol. Syst. Biol.* **6**, 370 (2010).
41. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).

© 2013 Nature America, Inc. All rights reserved.

ONLINE METHODS

Compiling RNAi screens. We compiled *Drosophila* RNAi screens that cover 49 phenotypes from the following resources: (i) *Drosophila* RNAi Screening Center (DRSC, http://www.flyrnai.org/), (ii) GenomeRNAi (http://genomernai.de/GenomeRNAi/), (iii) Neuroblasts Screen online database (http://neuroblasts.imba.oeaw.ac.at/), (iv) Bristle Screen online database (http://bristlescreen.imba.oeaw.ac.at/), and (v) unpublished data (seven phenotypes). Refer to **Supplementary Table 1** for more details. The data sets are preprocessed according to general guidelines provided by the DRSC to handle potential off-targets (only for data sets where amplicon level information is available). This includes (i) filtering out the amplicons that have predicted off-target effects, (ii) averaging values of multiple RNAi reagents targeting the same gene to consider the most consistent phenotype, and (iii) filtering out the hits that are not shown to be expressed in a given cell line. The screens reported the phenotypic contributions as either a *Z* score or as categorical data. For screens with *Z* scores, we used a standard cutoff of 1.5/−1.5 and define those genes with *Z* score ≥1.5 as negative regulators and genes with *Z* score ≤−1.5 as positive regulators of the phenotype. For the categorical data, we directly used the annotation: for example, "Down regulation of Wg Pathway" or "Upregulation of Wg pathway," to define positive and negative regulators of the phenotypes. Note that we selected only screens that include both positive and negative regulators of a specific phenotype.

Model for predicting signs. For each RNAi screen, the positive and negative regulators were distinguished with the values +1 and −1, respectively. Genes that did not score in a particular screen were assigned the value 0. We constructed a phenotype matrix by combining multiple RNA screens, where the rows correspond to genes and columns correspond to the RNAi screens. In a given RNAi screen, if both interacting proteins have nonzero values, then the relationship is classified as either a positive correlation (both +1 or both −1) or a negative correlation (one is +1 and another is −1). For each interacting pair, we computed the total number of positive and negative correlations. Then we used a simple model to calculate a sign score (S_{sign}) for each interaction as follows

$$S_{sign} = \frac{P_c - N_c}{T_n}\sqrt{T_p}$$

P_c and N_c correspond to the number of positive and negative correlations, respectively. T_p is the total number of matching phenotypes ($P_c + N_c$). Note that T_p should be ≥2 in order to be considered for sign predictions (to compute the correlation, a minimum of two data points are needed). $\sqrt{T_p}$ is the weight factor to assign more confidence for signs predicted on the basis of a larger number of phenotypes. If a score has a positive value, the interaction gets a positive sign (activation); similarly, interactions with negative values are assigned a negative sign (inhibition).

Validation of the model. To validate the model, we compiled known signaling PPIs from SignaLink (http://signalink.org/), KEGG (http://www.genome.jp/kegg/pathway.html) and the Database of Cell Signaling (STKE) (http://stke.sciencemag.org/cm/). All the data correspond to the versions available online at those sites as of September 2012. The three data sets were integrated, and we selected signaling PPIs with two or more matching phenotypes in the phenotype matrix as the reference sets (**Supplementary Tables 2** and **3**). Next we manually curated the entire reference set by verifying the literature for the validity of the signs assigned by the databases. The sign annotation from the signaling databases was used as a PRS. To construct the NRS, the signs from the signaling databases were reversed. As our objective is to predict signs of physical interactions, both the PRS and the NRS include physically interacting pairs. However, the PRS has true signs and the NRS has incorrect signs. Consider the example reference interactions Dsor1-rolled from the MAPK signaling pathway and Akt-Foxo from the insulin signaling pathway. In the first example, Dsor1 (the *Drosophila* ortholog of MAP2K) activates rolled (the *Drosophila* ortholog of ERK/MAPK) by phosphorylation. In other words, whenever an interaction between these two proteins occur, Dsor1 only activates rolled and never inhibits it, and the interaction sign is always positive. In the next example, Akt inhibits Foxo by phosphorylation, hence the interaction sign is negative. In the PRS the Dsor1-rolled interaction is assigned a positive sign and the Akt-Foxo interaction is assigned a negative sign, but in the NRS, they are assigned inverse signs (negative sign for Dsor1-rolled and positive sign for Akt-Foxo). On the basis of the current literature, Dsor1 never inhibits rolled and Akt never activates Foxo, thus making valid negative reference sets. With the PRS and NRS, the true positive rate, false positive rate, precision and recall were calculated at various S_{sign} cutoff values. We then plotted receiver operating characteristic (ROC) and precision-recall (PR) curves.

Compiling PPI networks. We compiled experimentally identified PPIs from BioGrid, IntAct, DIP, MINT, DroID and DPiM (**Supplementary Table 6**). The data correspond to versions available in September 2012. Next we grouped the PPIs as direct (for example, those identified from yeast-two hybrid screens) or indirect (for example, those identified from AP-MS) according to the experimental approach used to detect PPIs. We constructed a binary interaction network as follows. (i) All interactions identified as direct interactions were selected. (ii) High-confidence AP-MS interactions reported in the literature were selected. (iii) We analyzed the rest of the AP-MS interactions network to look for additional evidence such as domain-domain interactions, kinase-substrate interactions, interologs and genetic interactions as described in ref. 33. We selected the AP-MS interactions that overlap with any of these networks and considered them to be direct interactions.

Network analysis and visualization. To analyze properties of the signed network, we used various publicly available tools. Networks were visualized using Cytoscape, an open-source platform for network analysis and visualization[42]. NetworkAnalyzer, a Cytoscape plug-in for analysis of network properties, was used to analyze the degree distribution, clustering coefficient, and edge betweenness centrality[43]. Both Cytoscape and NetworkAnalyzer were downloaded from http://www.cytoscape.org/ and installed locally. The triad-motif enrichment analysis was performed using FANMOD[44] (**Supplementary Table 9**). Triad motifs were extracted using Perl scripts developed in-house; these are freely available upon request.

© 2013 Nature America, Inc. All rights reserved.

For the box plots in **Figure 2d**,**e**, upper whiskers were calculated as min(max(x), (Q3 + 1.5 × IQR)) and lower whiskers were calculated as max(min(x), (Q1 − 1.5 × IQR)), where Q1 is the lower quartile, Q3 is the upper quartile and IQR is the interquartile range.

Computing gene expression correlation. To compute correlation coefficients from gene expression profiles, we used *Drosophila* developmental time-course data from the modENCODE Consortium[29]. A gene expression matrix was generated from RNA-seq data using the latest fly genome annotation as provided by modENCODE (http://www.modencode.org/). The profiles characterize expression dynamics for 15,998 coding and noncoding genes during 27 distinct stages of development, including 12 embryonic, 6 larval, 6 pupal and 3 sexed adult stages (30 data points in total). The Pearson correlation coefficient was computed for 6,125 signed PPIs across 30 data points.

Constructing subnetworks around signaling pathways and protein complexes. Annotation of signaling pathways was obtained from the SignaLink database[21]. The protein-complex annotations are used as defined in the protein complex enrichment analysis tool (COMPLEAT) resource[32]. For each pathway/protein complex, direct signed interactions were extracted and integrated with gene expression correlation. Note that for this analysis we excluded signed interactions from the literature but not those predicted by our approach.

Preparation of dsRNAs. Gene-specific amplicons (~200–500 bp) were amplified by PCR using Choice Taq Mastermix (Denville Scientific, CB4070-8) from genomic DNA, using synthesized oligos with an attached T7 sequence. dsRNA was then synthesized from PCR templates using the T7 Megascript kit (Ambion), and product size was confirmed by gel electrophoresis. Following purification with Millipore Multiscreen PCR plates (#MANU03050), dsRNAs were quantified by measurement of the OD260 (Nano-drop 8000, Fisher Scientific) and then stored at −20 °C until use.

Primary cell culture RNAi experiments. Embryonic primary cell cultures were isolated from gastrulating Oregon R embryos as described previously[45–47] and seeded in 384-well plates at 4×10^4 cells (10-µl volume) per well. Each well contained 5 µl dsRNA (0.25 µg dsRNA in water) targeting a gene from the proteasome network or control dsRNAs targeting *lacZ* or *thread*. Following a 20-h incubation in serum-free M3 medium at 18 °C, 30 µl of serum-containing medium was added to each well for a final FCS concentration of 10%. Primary cells were then cultured for an additional 5 d at 18 °C before fixation for 2 h in 2% formaldehyde. Cells were stained overnight at 4 °C with anti-ubiquitin mAb FK2 (Enzo Life Sciences, 1:400), washed and stained for 2 h with Alexa Fluor 594 goat anti-mouse (Molecular probes, 1:1,000), phalloidin Alexa Fluor 635 (Molecular Probes, 1:2,000), and 4,6-diamidino-2-phenylindole (DAPI; Sigma, 1:5,000), and then washed again. All antibody incubations and washes were performed in PBT, except for a final rinse in PBS before image analysis. This experiment was repeated to give four biological replicates.

Image analysis. Acquisition of high-quality images of the primary cell culture was performed with the Evotec Opera microscope at the DRSC (http://www.flyrnai.org/). With the use of a 20× water-immersion lens, 24 microscope fields were obtained per well for both the anti-ubiquitin and phalloidin stains. Images were analyzed with MetaXpress high-content image acquisition and analysis software (Molecular Devices). For each microscope field, muscles were identified from the mixed population of cells by positive phalloidin staining. Then ubiquitin aggregates were identified within these muscles using the MetaXpress granularity application module. Total ubiquitin aggregate area was divided by total muscle area for each field. These values were then combined to give a measure of the total area of aggregates per muscle cell area per well (ubiquitin accumulation). Normalized ubiquitinated protein accumulation was computed by taking the $\log_2$(fold change) of ubiquitin accumulation over the lacZ control. For each amplicon (unique dsRNA design), the replicates (four) were combined and median values were obtained. A gene was considered a regulator of proteolysis if two or more amplicons targeting that gene met the threshold (±0.32 $\log_2$(fold change)). We did not consider a gene to be a regulator if different dsRNAs gave inconsistent results, i.e., one scored as positive and another as negative. The dsRNAs used for the screening and the results are shown in **Supplementary Table 12**.

Proteasome activity assay. For each experiment, 5×10^3 S2R+ cells in serum-free Schneider's medium were added to each well (in a 384-well tissue culture plate) and incubated with 0.25 µg dsRNA for 30 min before serum was added to a final volume of 30 µl. Following 1 d of incubation, 30 µl of Proteasome-Glo trypsin-like cell-based reagent (Promega G8760) was added to each well, and the plate was incubated for 30 min at room temperature. Luminescence was measured with a SpectraMax Paradigm plate reader. Readings from cell culture medium control wells (no cells) were subtracted from all experimental values. Four replicates were performed for each dsRNA, and readings were normalized to those of the lacZ control RNAi wells. Normalized proteasome activity was computed by taking $\log_2$ median values from four replications.

qPCR. For each experiment, 2×10^6 S2R+ cells in serum-free Schneider's medium were added to each well (in a six-well tissue culture plate) and incubated with 20 µg dsRNA for 30 min before serum was added. Two days later, cells were harvested with Trizol (Invitrogen), and this step was followed by phenol-chloroform extraction and purification with the RNeasy kit (Qiagen). cDNA was synthesized with the iScript cDNA Synthesis kit (Bio-Rad), and quantitative RT-PCR was performed with the iQ SYBR Green Supermix (Bio-Rad). Rp49 was used as normalization reference. Relative quantitation of mRNA expression was calculated using the comparative CT method. The primers used were Rp49, 5′-ATCGGTTACGGATCGAACAA-3′ (forward) and 5′-GACAATCTCCTTGCGCTTCT-3′ (reverse); Eno, 5′-CCGAGAACAAGAGCAAGTTCG-3′ (forward) and 5′-CATGGCCTCTGTGAAGCTGG-3′ (reverse); CG10638, 5′-AGCTCGCTCCGACTGTTAAG-3′ (forward) and 5′-AGGCCCAGAATTGGCATCTC-3′ (reverse); Rpn1, 5′-CCGACGCTGGAGAGTATGG-3′ (forward) and 5′-GCATGAACTTCAAAGGCTTGG-3′ (reverse); Hsc70Cb, 5′-CGTGGCCGCTAAGAACCAG-3′ (forward) and 5′-ATGCTCGTGAGTTCGTGTTGT-3′ (reverse); CG32039, 5′-GAGCCTGTCTGTCCTGCTG-3′ (forward) and 5′-AGGCGA

© 2013 Nature America, Inc. All rights reserved.

doi:10.1038/nmeth.2733

TGGCATCAGGTT-3′ (reverse); pomp, 5′-TATCAGCCATCAC TGAAAGTCCA-3′ (forward) and 5′-GTTGCGGTTGTACTGG TGC-3′ (reverse); MRP, 5′-CGCCTTCTACTGGGCGTTC-3′ (forward) and 5′-ACCAGAGCTTTGCTCACGTTC-3′ (reverse); mts, 5′-ACGGTCAGTTTCACGACCTC-3′ (forward) and 5′-CT CCACGGAGTAGTATCCACG-3′ (reverse); polo, 5′-TCACCGC AGCCTTAACCATC-3′ (forward) and 5′-ACAGCTCCAGCA CAATGTAGAT-3′ (reverse); GNBP2, 5′-CCGCCCAAACGATA GTGAG-3′ (forward) and 5′-GATGTCATGCTTCCAGGTGGT-3′ (reverse); CSN4, 5′-AAGTTGCCTGACGATCTGTCC-3′ (forward) and 5′-TATGCCAGCCACTTGCTCTTC-3′ (reverse); CG15717, 5′-AGTCCCTGCAGAATCCCTTT-3′ (forward) and 5′-GGCTTT CGCCTTGTACTGTC-3′ (reverse); Ubi-p5E, 5′-TCTTCACTTG GTCCTGCGTC-3′ (forward) and 5′-ATGGCTCGACCTCCAAA GTG-3′ (reverse); Ubi-p63E, 5′-ACGCACCCTGTCCGATTAC-3′ (forward) and 5′-TGGTCTTTCCGGTCAAAGTCTT-3′ (reverse).

Larval muscle histology. Wandering third instar larvae were dissected in ice-cold PBS and fixed for 20 min in 4% formaldehyde in PBS. After being washed, body-wall muscles were incubated overnight at 4 °C with anti-ubiquitin mAb FK2 (Enzo Life Sciences, 1:250), washed and stained for 2 h with Alexa Fluor 594 goat anti-mouse (Molecular probes, 1:1,000), and DAPI (Sigma, 1:5,000), and then washed again and mounted in 50% glycerol/PBS. All antibody incubations and washes were performed in PBT, except for a final rinse in PBS before mounting. The ventral longitudinal muscles from segment 3 or 4 were imaged with a Leica TCS SP2 confocal laser-scanning microscope.

Fly stocks. *Mef2-GAL4* (ref. 48) was used to drive transgene expression specifically in the larval muscles via the Gal4/UAS system[49]. The following RNAi hairpin lines targeting proteasome network components were obtained from the NIG Japan: *Enolase* (17655R-1 and 17654R-2) and *CG10638* (32101R-2 and 32101R-3). An additional line for *Enolase* (JF02070) was obtained from the DRSC/TRiP at Harvard Medical School. All RNAi experiments were performed at 25 °C, and all lines gave phenotype.

Implementation of SignedPPI. The SignedPPI user interface was implemented as a collection of Java servlets, JavaScript, and Adobe Flash components. SignedPPI integrates existing tools, including Cytoscape Web for complex visualization (http://cytoscapeweb.cytoscape.org/), Lucene (http://lucene.apache.org/) for searches, JSON (http://www.json.org/) for communication between Java servlets and JavaScript, and WebFX for parameter adjustment sliders (http://webfx.eae.net/). The application is hosted on the Orchestra cluster supported by the Research IT Group (RITG) at Harvard Medical School. PPIs are maintained as flat files and indexed by Lucene. The Java servlets and Java Server Pages run within an instance of Tomcat 6.0.18 on the Orchestra cluster. The Entrez Gene IDs, gene symbols, and alias names for *Drosophila* genes were retrieved from NCBI (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/). FlyBase IDs, CG numbers, gene symbols, and synonyms were retrieved from FlyBase (ftp://ftp.flybase.net/releases/current/). Protein IDs were retrieved from UniProt (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/). A program developed in-house automatically downloads via FTP and processes these files on a monthly basis.

42. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
43. Doncheva, N.T., Assenov, Y., Domingues, F.S. & Albrecht, M. Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.* **7**, 670–685 (2012).
44. Wernicke, S. & Rasche, F. FANMOD: a tool for fast network motif detection. *Bioinformatics* **22**, 1152–1153 (2006).
45. Bai, J., Sepp, K.J. & Perrimon, N. Culture of *Drosophila* primary cells dissociated from gastrula embryos and their use in RNAi screening. *Nat. Protoc.* **4**, 1502–1512 (2009).
46. Bai, J. *et al.* RNA interference screening in *Drosophila* primary cells for genes involved in muscle assembly and maintenance. *Development* **135**, 1439–1449 (2008).
47. Sepp, K.J. *et al.* Identification of neural outgrowth genes using genome-wide RNAi. *PLoS Genet.* **4**, e1000111 (2008).
48. Ranganayakulu, G., Schulz, R.A. & Olson, E.N. Wingless signaling induces *nautilus* expression in the ventral mesoderm of the *Drosophila* embryo. *Dev. Biol.* **176**, 143–148 (1996).
49. Brand, A.H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).

© 2013 Nature America, Inc. All rights reserved.