

Supplementary Text: Performance Evaluation of the TLM Network

We used the following performance measures to compare ANAT to other network construction tools (Table 2).

Cross validation

In each iteration, we hid a subset of proteins that is unique to one of the three screens and used the remaining of the TLM set to build a network model. We based the cross validation score on two sets: (i) the union of all proteins that were recovered when hidden, and (ii) the union of all proteins that were included in any one of the networks and were not part of the input target set. We report the ratio between the sizes of these two sets. This index indicates how "enriched" the inferred subnetwork is in true hits that were hidden.

Functional enrichment

Given a network model, we compute for each submodel (a path that connects a certain target protein to the TLM anchor) a functional enrichment score based on the GO "Biological Process" annotation. The enrichment is computed by a hypergeometric p -value as described in the Notes and Remarks (under "Analysis of Submodels"). The reported functional enrichment scores for the TLM network are the fraction of subnetworks with at least four nodes that were significantly coherent (empirical p -value lower than 0.05).

Monochromaticity

The monochromaticity of a subnetwork P is the larger of the fraction of proteins representing genes that when deleted results in shortened telomers in P and the fraction of proteins representing genes that when deleted results in elongation of telomers in P . We report the mean monochromaticity across all subnetworks that had at least four annotated nodes.

Phenotype prediction

We defined the effect of a subnetwork P according to the phenotype of the participating TLM proteins. The effect was either "long" if most proteins were encoded by genes that had a lengthening effect on telomeres, short in the opposite case, or undecided if it contained equal numbers of proteins with shortening and elongating effects. We then classified a given protein to the long class if all the subnetworks to which it belonged were either long or undecided and at least one of them is long. We applied the opposite rule to predict a short phenotype. If none of these conditions were true, then the protein remained classified as undecided. We separately analyzed each of the proteins in the TLM target set by hiding its phenotype and using the decision rule to predict its phenotype. The reported precision value is the number of correct predictions divided by the overall number of predictions. The reported recall is the number of correct predictions divided by the overall number of TLM proteins that have a recorded phenotype.

Supplementary Table Legends

Table S1. Network sizes, data sources, and gene identifier types. For every organism in ANAT's database, the table lists the number of nodes and edges in the respective network. (If the organism has protein-DNA interaction (PDI) data, then the numbers of nodes and edges in the protein-protein interaction (PPI) and the PDI networks are listed as well). The table includes the data sources from which interaction data was taken. The sources for the identifiers used for the genes in the networks are also provided. The list of identifiers and synonyms for each gene in ANAT's database is provided as a series of text files in the compressed archive Gene_identifier.zip in the Supplementary Materials.

Table S2. Constraints used for the autophagy-apoptosis crosstalk network. This table lists all the edge constraints (removal, or forced directionality) used in the PCD background network (used to build the autophagy-apoptosis cross talk network). The first two columns list the names of the interacting proteins. The third column lists Pubmed references that explain the respective constraint.