

Supplementary Table 1: List of aggregating functions combining protein-level signals to complex-level features

For each protein-level pairwise signal, we use the aggregating functions in the first table to create the complex-level features. The cutoffs used by some of the aggregating functions are different for different signals. Those cutoffs are listed in the second table.

Summation of all values
Maximum of all values
Average of all values
Minimum of all values
Number of values above a certain cutoff where the cutoff for different data sources are listed below
“number above cutoff” divided by total number of values
“number above cutoff” divided by square root of total number of values
Average of top three values
Decayed max: weighted average with the largest value has weight 1, second largest weight 1/2, and each subsequent one reducing the weight by half
Decayed min: weighted average with the smallest value has weight 1, second smallest weight 1/2, and each subsequent one reducing the weight by half

Data source	Cutoff
direct PE score	3
indirect PE score	3
scaled PE score	0.5
PE score correlation	0.2
GO cell component distance	7.5
trans-membrane proteins	0.5
co-expression (SMD)	0.5
InSite	0.5
correlation of growth fitness	0
TF mutual information	0.01
time-series correlation	0.5

Supplementary Table 2: List of aggregating functions chosen

We plot the ROC curve of each aggregating function, applied to a data source, in predicting interactions and pick the aggregating function that has the maximum area

under the curve. The data source is listed in the first column and the corresponding best aggregating function is listed in the second column. Among all global features, we also pick the one that has the maximum area under the ROC curve. The first table is for complex-complex interactions and the second table is for the unified interaction network involving both proteins and complexes.

Data source	Aggregating function chosen
direct PE score	max
indirect PE score	max
scaled PE score	sum
PE score correlation	sum
GO cell component distance	min
trans-membrane proteins	sum
co-expression (SMD)	top 3 average
InSite	decayed min
correlation of growth fitness	top 3 average
TF mutual information	sum
time-series correlation	max
GFP localization	# pairs with different localization
Global	# protein pairs

Data source	Aggregating function chosen
direct PE score	max
indirect PE score	max
scaled PE score	sum
PE score correlation	sum
GO cell component distance	decayed min
trans-membrane proteins	sum
co-expression (SMD)	top 3 average
InSite	decayed min
correlation of growth fitness	top 3 average
TF mutual information	sum
time-series correlation	max
GFP localization	fraction of pairs with different localization
Global	# protein pairs

Supplementary Table S3: Parametric family in the model for the complex-complex interaction network

Shown here are the parametric families of distributions for the feature values given the two complexes interact and given the two complexes do not interact. The parametric families are picked based on examining the reference set of complex-complex interactions and non-interactions. The parameters for the distributions are learned from the data. Exponential distribution starts at 0 and goes to the right. “Reverse exponential” starts at 1 and goes to the left. “point 0” and “point 1” refers to discrete distributions with all mass at the point 0 and point 1 respectively.

Data source	Distribution given non-interaction	Distribution given interaction
direct PE score	Mixture of uniform if negative and exponential if positive	Mixture of uniform if negative and Gaussian if positive
indirect PE score	Exponential	Exponential
scaled PE score	Exponential	Mixture of uniform and exponential
PE score correlation	Mixture of two exponentials with different means	Mixture of two exponentials with different means
GO cell component distance	Mixture of three uniforms with boundary 6.7 and 7.5	Mixture of three uniforms with boundary 6.7 and 7.5
trans-membrane proteins	Exponential	Exponential
co-expression (SMD)	Mixture of point 0 and Gaussian	Gaussian
InSite	Reverse exponential	Reverse exponential
correlation of growth fitness	Mixture of point 0 and Gaussian	Mixture of point 0 and Gaussian
TF mutual information	Mixture of point 0 and exponential	Mixture of point 0 and exponential

time-series correlation	Mixture of point 0 and Gaussian	Mixture of point 0 and Gaussian
GFP localization	Exponential	Exponential
# protein pairs	Exponential	Exponential

Supplementary Table S4: Parametric family in the model for the unified interaction network

Same as Supplementary Table S3 except here the parametric families are picked based on examining the reference set for the unified interaction network involving both proteins and complexes.

Data source	Distribution given non-interaction	Distribution given interaction
direct PE score	Mixture of two exponentials with different means	Mixture of exponential and Gaussian
indirect PE score	Mixture of point 0 and Gaussian	Mixture of point 0 and Gaussian
scaled PE score	Mixture of point 0 and exponential	Mixture of point 0 and exponential
PE score correlation	Mixture of two exponentials with different means	Mixture of two exponentials with different means
GO cell component distance	Mixture of three uniforms with boundary 6.7 and 8.4	Mixture of three uniforms with boundary 6.7 and 8.4
trans-membrane proteins	Exponential	Exponential
co-expression (SMD)	Mixture of point 0 and Gaussian	Gaussian
InSite	Reverse exponential	Reverse exponential
correlation of growth fitness	Mixture of point 0 and	Mixture of point 0 and

	Gaussian	Gaussian
TF mutual information	Mixture of point 0 and exponential	Mixture of point 0 and exponential
time-series correlation	Mixture of point 0 and Gaussian	Gaussian
GFP localization	Mixture of point 0, point 1, and Gaussian	Mixture of point 0, point 1, and Gaussian
# protein pairs	Exponential	Exponential