Under consideration for publication in Network Science

Supplementary Material for: Networks' Characteristics Are Important *for Systems Biology*

Andrew K. Rider[1,3,4], Tijana Milenković[1,3,4], Geoffrey H. Siwo[2,3,4], Richard S. Pinapati[2,3,4], Scott J. Emrich[1,3,4], Michael T. Ferdig[2,3,4], Nitesh V. Chawla[1,3,4,*]

[1]Department of Computer Science and Engineering

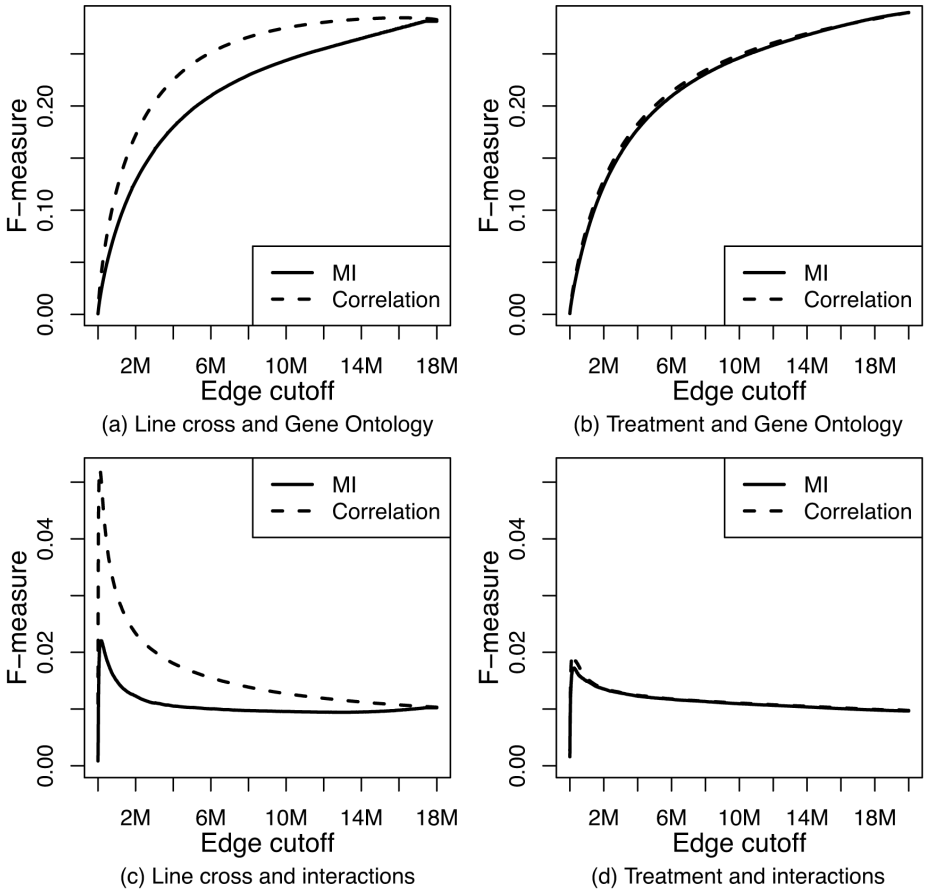
[2]Department of Biological Sciences

[3]ECK Institute for Global Health

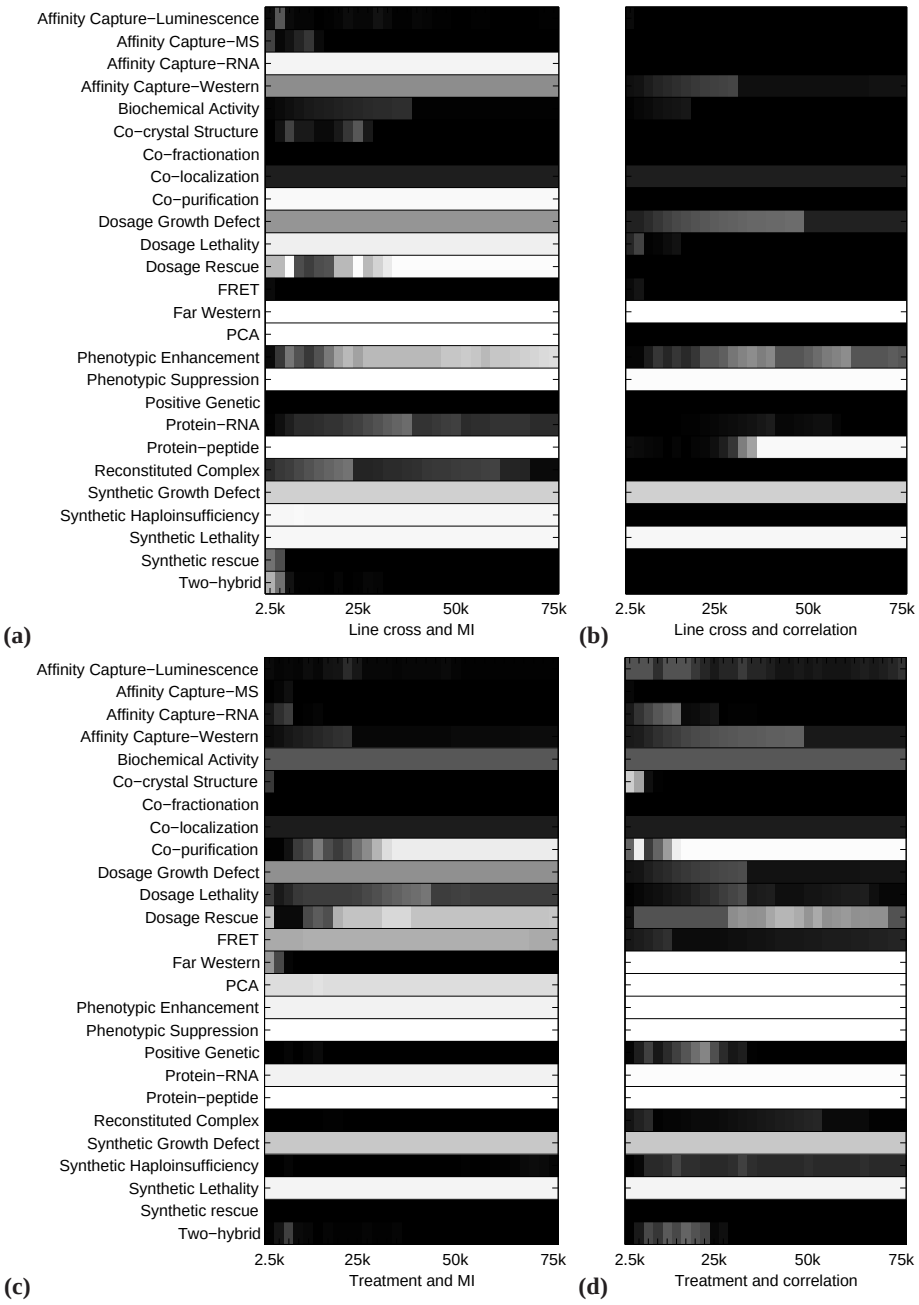
[4]Interdisciplinary Center for Network Science and Applications (iCeNSA)

University of Notre Dame, Notre Dame IN 46556, USA

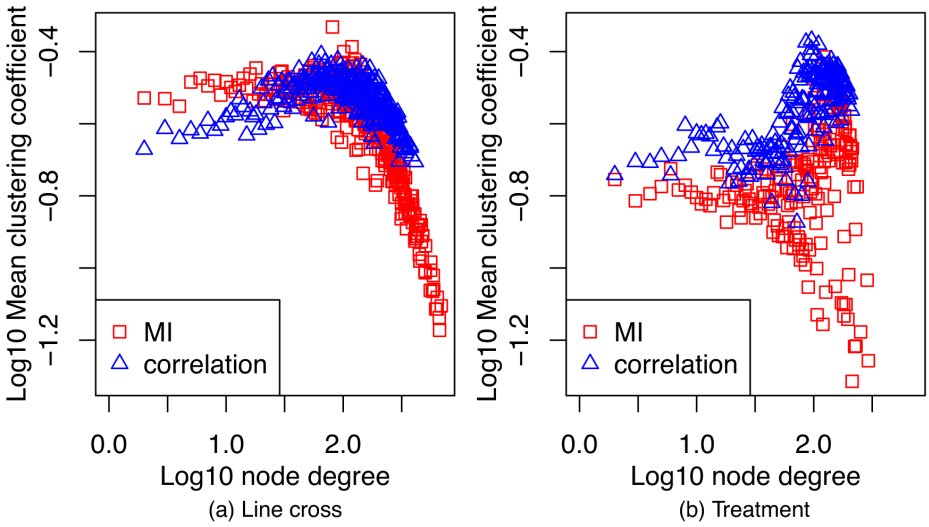
[*]Corresponding Author (E-mail: nchawla@nd.edu)



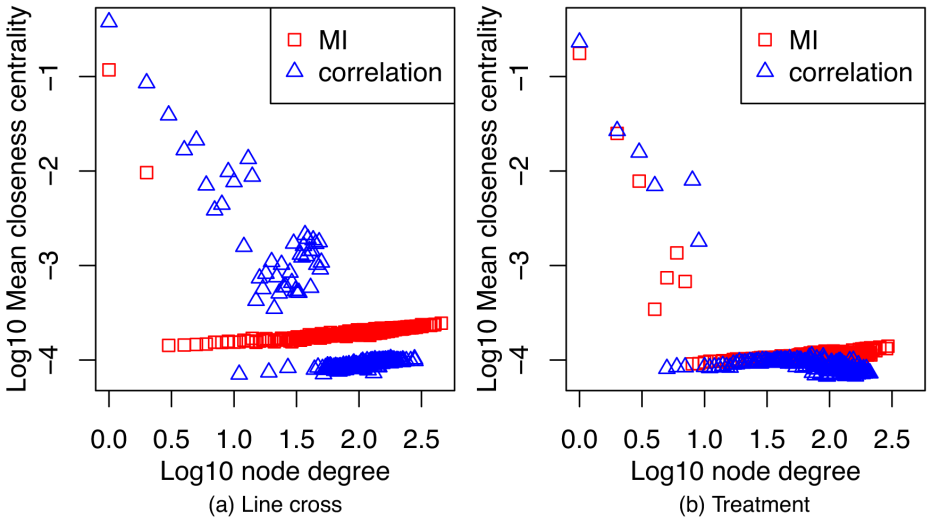
Supporting Figure 1**.** F-scores at different edge cut-offs measuring how accurately networks constructed in different ways capture known biological knowledge. Panel **(a)** shows F-scores for networks constructed from the line cross data with respect to shared GO terms. Panel **(b)** shows F-scores for networks constructed from the treatment data with respect to shared GO terms. Panel **(c)** shows F-scores for networks constructed from the line cross data with respect to known interactions. Panel **(d)** shows F-scores for networks constructed from the treatment data with respect to known interactions.



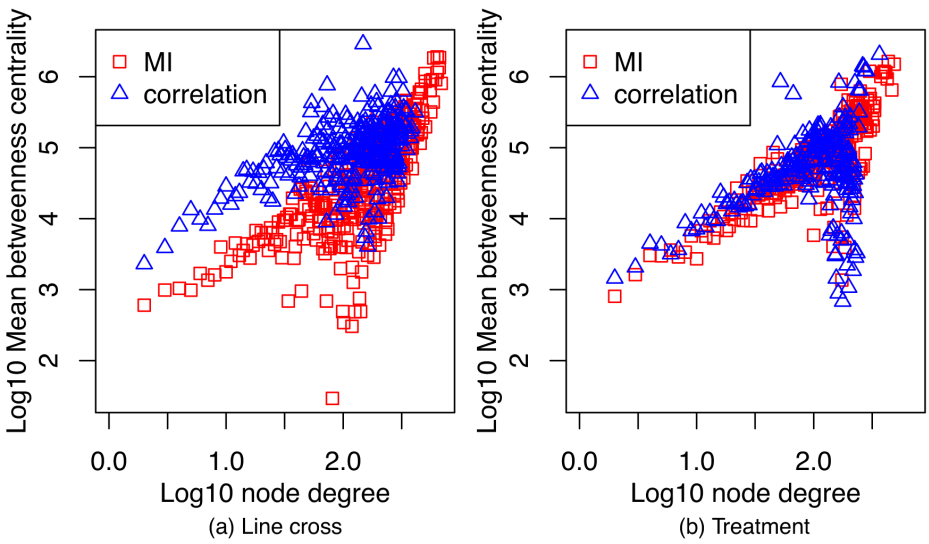
Supporting Figure 2**.** Heat maps showing the significance of the enrichment of a network of a given size (*x*-axis) in known interactions of a given type (*y*-axis) according to the hypergeometric test (see Methods). Significance is denoted by the darkness of the color, black being the most significant; significance diminishes as the color approaches white. Panel **(a)** shows enrichment results for networks constructed from the line cross data using mutual information. Panel **(b)** shows enrichment results for networks constructed from the line cross data using correlation. Panel **(c)** shows enrichment results for networks constructed from the treatment data using mutual information. Panel **(d)** shows enrichment results for networks constructed from the treatment data using correlation.



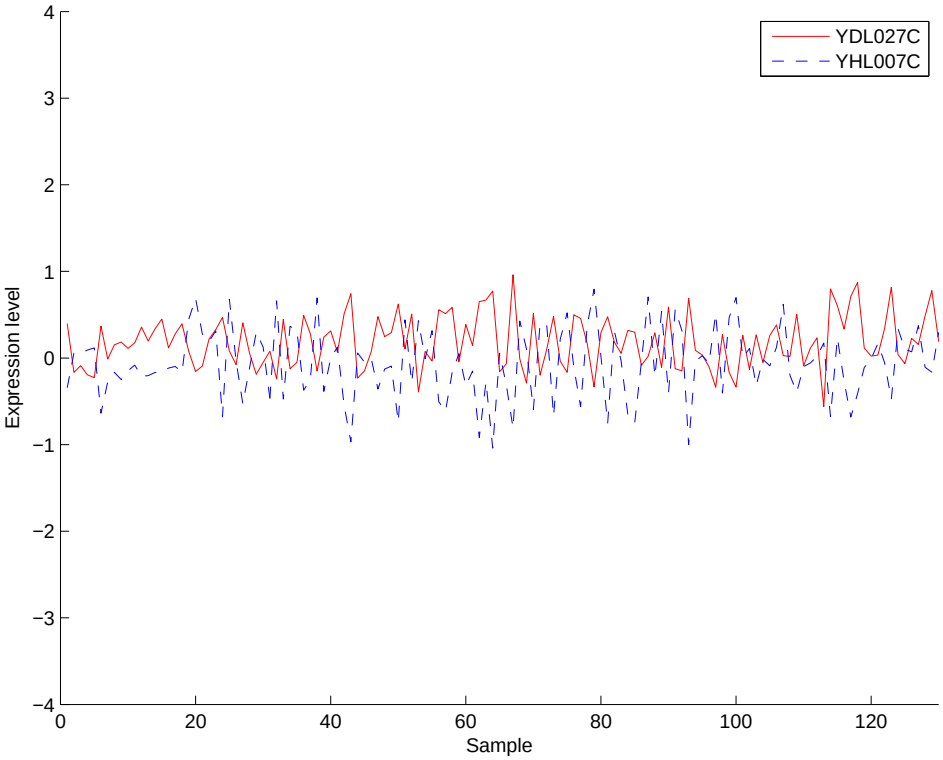
Supporting Figure 3**.** Clustering spectra for networks constructed from the line cross data (a) and the treatment data (b) at the edge cut-off of 25,000. We compared with a *t*-test pairs of blue and red spectra within panels, which share the same data type but differ in the edge weighting method. We also compared pairs of blue and blue spectra or red and red spectra across panels, which share the same edge weighting method but differ in the data type. The two spectra in panel **(a)** as well as the two spectra in panel **(b)** were statistically significantly different with *p*-values $< 2.2 \times 10^{-16}$. The two correlation-based spectra (blue) across the two panels were statistically significantly different with a *p*-value of 0.008. The two mutual information-based spectra (red) across the two panels were statistically significantly different with a *p*-value of 1.07×10^{-15}.



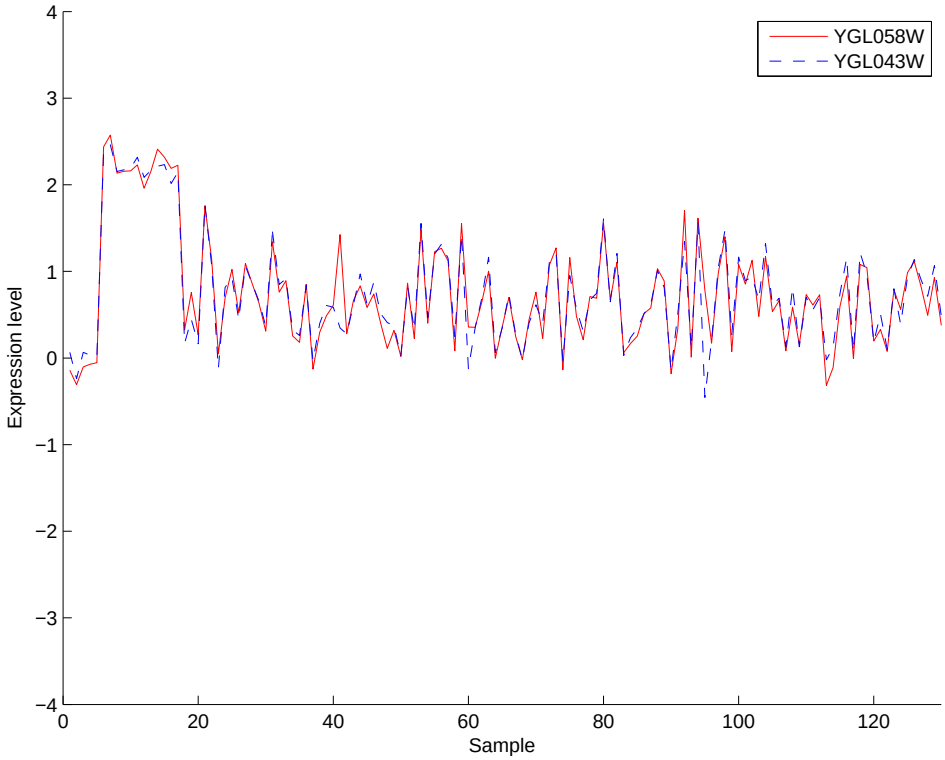
Supporting Figure 4**.** Closeness spectra for networks constructed from the the line cross data **(a)** and the treatment data **(b)** at the edge cut-off of 25,000. We compared with a *t*-test pairs of blue and red spectra within panels, which share the same data type but differ in the edge weighting method. We also compared pairs of blue and blue spectra or red and red spectra across panels, which share the same edge weighting method but differ in the data type. The two spectra in panel **(a)** were not statistically significantly different (*p*-value of 0.3511). The two spectra in panel **(b)** were statistically significantly different with a *p*-value of 0.0063. The two correlation-based spectra (blue) across the two panels were statistically significantly different with a *p*-value of 2.81×10^{-6}. The two mutual information-based spectra (red) across the two panels were statistically significantly different with a *p*-value of 1.6×10^{-12}.



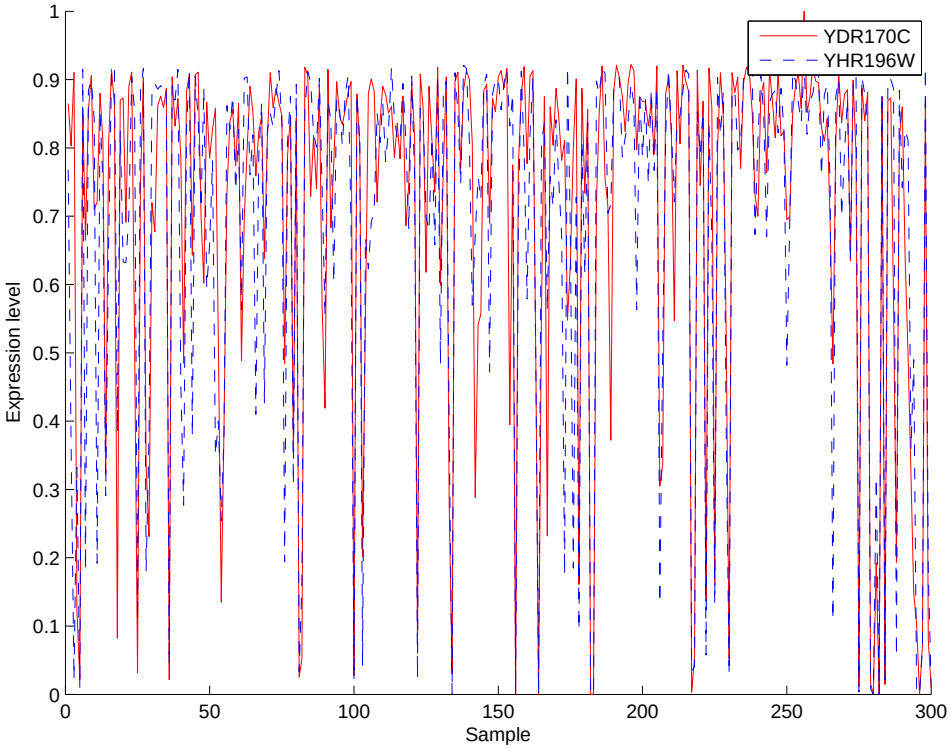
Supporting Figure 5**.** Betweenness spectra for networks constructed from the the line cross data **(a)** and the treatment data **(b)** at the edge cut-off of 25,000. We compared with a *t*-test pairs of blue and red spectra within panels, which share the same data type but differ in the edge weighting method. We also compared pairs of blue and blue spectra or red and red spectra across panels, which share the same edge weighting method but differ in the data type. The two spectra in panel **(a)** as well as the two spectra in panel **(b)** were statistically significantly different with p-values $< 2.2 \times 10^{-16}$. The two correlation-based spectra (blue) across the two panels were statistically significantly different with a p-value of 3.3×10^{-7}. The two mutual information-based spectra (red) across the two panels were statistically significantly different with a p-value of 9.6×10^{-11}.



Supporting Figure 6**.** Expression levels in the line cross data of two genes that share a known Biochemical Activity interaction and for which the correlation is low while the mutual information is high. Namely, the correlation has a value of -0.742330 and the mutual information has a value of 0.157230. The correlation between these two genes is greater than the correlation between 0% of all pairs of genes in the data. The mutual information between these two genes is greater than the mutual information between 99.2% of all pairs of genes in the data.



Supporting Figure 7**.** Expression levels in the line cross data of two genes that share a known Synthetic Lethality interaction and for which both the correlation and the mutual information are high. Namely, the correlation has a value of 0.958520 and the mutual information has a value of 0.863000. The correlation between these two genes is greater than the correlation of 99.8% of all pairs of genes in the data. The mutual information between these two genes is greater than the mutual information between 99.9% of all pairs of genes in the data.



Supporting Figure 8**.** Expression levels in the treatment data of two genes that share a known Synthetic Lethality interaction and for which both the correlation and the mutual information are high. Namely, the correlation has a value of 0.858660 and the mutual information has a value of 0.347460. The correlation between these two genes is greater than the correlation of 99.7% of all pairs of genes in the data. The mutual information between these two genes is greater than the mutual information between 99.9% of all pairs of genes in the data.

1 Tables

Supporting Table 1. *P*-values from signed-rank tests comparing different edge weighting methods and data types with respect to the proportion of known interactions of a given type (out of the total number of edges in the network) across 30 networks corresponding to the 30 cut-offs. For each of the 26 known interaction types, for each combination of the edge weighting method and data type, for each of the 30 cut-offs, we compute the proportion of known interactions of the given type out of all edges in the network constructed using the given edge weighting method, data type, and cut-off. Then, we compare the 30 resulting values corresponding to the 30 cut-offs between networks constructed from linecross data using correlation and networks constructed from linecross data using mutual information, between networks constructed from treatment data using correlation and networks constructed from treatment data using mutual information, between networks constructed from linecross data using correlation and networks constructed from treatment data using correlation, and between networks constructed from linecross data using mutual information and networks constructed from treatment data using mutual information. The *p*-value on the left of a given cell in the table tests whether the median rank of the first set of the 30 values is greater than or equal to the median rank of the second set of the 30 values. The *p*-value on the right of the cell tests whether the median rank of the second set of the 30 values is greater than or equal to the median rank of the first set of the 30 values. If the *p*-value is below a given cut-off (see below), the difference in the median ranks between two given sets is considered to be statistically significant (and is bolded in the table). We used the Šidák correction for multiple testing to identify a stringent *p*-value cut-off, corresponding to the 0.05 cut-off. The Šidák correction is similar to the Bonferroni correction but assumes independence of individual tests (Šidák, 1967). The extent to which interaction types are independent is unclear but it is common to assume independence in the case of uncertainty. Correcting for 26 tests corresponding to the 26 known interaction types, the *p*-value cut-off is 1.7×10^{-03}. The "NAs" correspond to no observations being made for the given interaction type. The last row counts the number of known interaction types out of 26 of them for which at least one of the two *p*-values is below the cut-off.

	Line cross & MI v. Line cross & Correlation		Treatment & MI v. Treatment & Correlation		Treatment & MI v. Line cross & MI		Treatment & Correlation v. Line cross & Correlation	
Affinity Capture-Luminescence	$1.0\times10^{+00}$	$\mathbf{1.3}\times10^{-06}$	$\mathbf{1.5}\times10^{-05}$	$1.0\times10^{+00}$	$\mathbf{3.5}\times10^{-04}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Affinity Capture-MS	$1.0\times10^{+00}$	$\mathbf{9.3}\times10^{-10}$	9.8×10^{-01}	2.4×10^{-02}	9.8×10^{-01}	2.4×10^{-02}	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Affinity Capture-RNA	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{4.4}\times10^{-06}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{2.9}\times10^{-06}$	$\mathbf{2.9}\times10^{-05}$	$1.0\times10^{+00}$
Affinity Capture-Western	$1.0\times10^{+00}$	$\mathbf{1.1}\times10^{-04}$	3.0×10^{-03}	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{3.2}\times10^{-05}$	1.1×10^{-02}	9.9×10^{-01}
Biochemical Activity	$1.0\times10^{+00}$	$\mathbf{1.6}\times10^{-04}$	NA	NA	$\mathbf{3.6}\times10^{-04}$	$1.0\times10^{+00}$	$\mathbf{1.4}\times10^{-05}$	$1.0\times10^{+00}$
Co-crystal Structure	$1.0\times10^{+00}$	$\mathbf{9.3}\times10^{-10}$	$\mathbf{3.3}\times10^{-04}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{1.2}\times10^{-05}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Co-fractionation	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{3.2}\times10^{-04}$	$1.0\times10^{+00}$	$\mathbf{1.4}\times10^{-06}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Co-localization	NA	NA	NA	NA	NA	NA	NA	NA
Co-purification	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{3.6}\times10^{-04}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{3.2}\times10^{-05}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Dosage Growth Defect	$1.0\times10^{+00}$	1.9×10^{-03}	$1.0\times10^{+00}$	$\mathbf{1.6}\times10^{-04}$	NA	NA	9.9×10^{-01}	1.8×10^{-02}
Dosage Lethality	$1.0\times10^{+00}$	$\mathbf{2.0}\times10^{-06}$	$1.0\times10^{+00}$	$\mathbf{8.3}\times10^{-04}$	$1.0\times10^{+00}$	$\mathbf{1.3}\times10^{-06}$	$\mathbf{7.8}\times10^{-04}$	$1.0\times10^{+00}$
Dosage Rescue	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	9.6×10^{-01}	4.1×10^{-02}	2.8×10^{-01}	7.3×10^{-01}	$\mathbf{1.3}\times10^{-06}$	$1.0\times10^{+00}$
Far Western	$1.0\times10^{+00}$	1.9×10^{-03}	$\mathbf{1.3}\times10^{-06}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{1.3}\times10^{-06}$	$1.0\times10^{+00}$	$\mathbf{4.8}\times10^{-05}$
FRET	$\mathbf{1.1}\times10^{-04}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{2.2}\times10^{-05}$	$\mathbf{1.3}\times10^{-06}$	$1.0\times10^{+00}$	$\mathbf{2.0}\times10^{-06}$	$1.0\times10^{+00}$
PCA	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{1.3}\times10^{-06}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{1.2}\times10^{-05}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Phenotypic Enhancement	$1.0\times10^{+00}$	$\mathbf{2.1}\times10^{-04}$	$\mathbf{1.3}\times10^{-06}$	$1.0\times10^{+00}$	$\mathbf{1.6}\times10^{-05}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Phenotypic Suppression	$1.0\times10^{+00}$	$\mathbf{3.0}\times10^{-06}$	$\mathbf{6.5}\times10^{-06}$	$1.0\times10^{+00}$	$\mathbf{9.7}\times10^{-05}$	$1.0\times10^{+00}$	$\mathbf{3.0}\times10^{-06}$	$1.0\times10^{+00}$
Positive Genetic	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{3.3}\times10^{-06}$	$1.0\times10^{+00}$	8.9×10^{-01}	1.1×10^{-01}	$\mathbf{9.3}\times10^{-10}$	$1.0\times10^{+00}$

Continued on next page

	Line cross & MI v. Line cross & Correlation		Treatment & MI v. Treatment & Correlation		Treatment & MI v. Line cross & MI		Treatment & Correlation v. Line cross & Correlation	
Protein-peptide	$1.0\times10^{+00}$	$\mathbf{9.3}\times10^{-10}$	$\mathbf{1.6}\times10^{-05}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{1.4}\times10^{-06}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Protein-RNA	$1.0\times10^{+00}$	$\mathbf{4.4}\times10^{-06}$	$\mathbf{1.3}\times10^{-03}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Reconstituted Complex	$1.0\times10^{+00}$	$\mathbf{3.0}\times10^{-06}$	$\mathbf{6.5}\times10^{-06}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{4.4}\times10^{-06}$	$1.0\times10^{+00}$
Synthetic Growth Defect	NA	NA	NA	NA	NA	NA	NA	NA
Synthetic Haploinsufficiency	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{4.3}\times10^{-05}$	$1.0\times10^{+00}$	$1.0\times10^{+00}$	$\mathbf{9.3}\times10^{-10}$	$\mathbf{2.0}\times10^{-06}$	$1.0\times10^{+00}$
Synthetic Lethality	NA	NA	NA	NA	NA	NA	NA	NA
Synthetic Rescue	$1.0\times10^{+00}$	$\mathbf{9.3}\times10^{-10}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	9.5×10^{-01}	5.5×10^{-02}	$\mathbf{9.3}\times10^{-10}$	$1.0\times10^{+00}$
Two-hybrid	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$\mathbf{4.0}\times10^{-04}$	$1.0\times10^{+00}$	$\mathbf{1.2}\times10^{-03}$	$1.0\times10^{+00}$	$\mathbf{9.1}\times10^{-07}$	$1.0\times10^{+00}$
Significant differences	21		19		18		21	

Supporting Table 2**.** *The number of known interactions of a given type and the number of genes from each of the two data sets that are involved in the corresponding interactions. There are 5,829 shared genes between the two data sets, with a total of 5,913 genes in the line cross data and a total of 6,207 genes in the treatment data.*

Known interaction type	Number of interactions	Number of genes from line cross data	Number of genes from treatment data
Affinity Capture-Luminescence	32	11	11
Affinity Capture-MS	58,861	3,972	4,123
Affinity Capture-RNA	6,961	3,044	3,170
Affinity Capture-Western	12,165	2,383	2,470
Biochemical Activity	9,447	1,820	1,888
Co-purification	2,238	871	898
Co-crystal Structure	324	220	224
Co-fractionation	1,120	557	572
Co-localization	719	353	366
Dosage Growth Defect	476	258	269
Dosage Lethality	1,128	505	521
Dosage Rescue	6,554	1,930	1,999
Far Western	100	74	80
FRET	194	90	92
PCA	8,569	1,474	1,521
Phenotypic Enhancement	7,959	1,885	1,952
Phenotypic Suppression	5,672	1,363	1,406
Positive Genetic	24,810	2,806	2,915
Protein-RNA	772	380	391
Protein-peptide	208	112	116
Reconstituted Complex	3,996	1,374	1,421
Synthetic Growth Defect	26,944	2,869	2,974
Synthetic Haploinsufficiency	396	199	202
Synthetic Lethality	20,899	2,691	2,792
Synthetic Rescue	5,606	1,616	1,673
Two-hybrid	13,863	3,152	3,268

Bibliography

Šidák, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, **62**(318), 626–633.