Supplementary Information For:

Conserved Abundance and Topological Features in Chromatin Remodeling Protein Interaction Networks

Mihaela E. Sardiu[1#], Joshua M. Gilmore[1#], Brad D. Groppe[1#], Damir Herman[2], Sreenivasa R. Ramisetty[3], Yong Cai[4], Jingji Jin[4], Ronald C. Conaway[1,5], Joan W. Conaway[1,5], Laurence Florens[1], and Michael P. Washburn[1,6‡]

[1]Stowers Institute for Medical Research, Kansas City, MO 64110 U.S.A.
[2]Ayasdi, Inc.,636 Ramona St, Palo Alto, CA 94301
[3]Idexx Laboratories, 1 Idexx Drive, Westbrook, ME-04092
[4]College of Life Sciences, Jilin University, 2699 Qianjin St., Changchun Jilin Province 130012, China
[5]Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA
[6]Department of Pathology and Laboratory Medicine, The University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA

# These authors contributed equally to this work

To whom correspondence should be addressed:

Michael P. Washburn, Ph.D.

Stowers Institute for Medical Research

1000 E. 50th St

Kansas City, MO 64110

mpw@stowers.org

816-926-4457

Running Title: Conserved Abundance and Topology in Chromatin Networks

Subject Category: Methods and Resources; Chromatin, Epigenetics, Genomics, & Functional Genomics; Computational Biology; Proteomics

Supplementary Methods: MudPIT mass spectrometry and data analysis

In order to analyze the purified protein complexes, TCA-precipitation, LysC/Trypsin digestion, and multidimensional protein identification technology (MudPIT) analyses were performed as previously described [1, 2]. RAW files were converted to the ms2 format using RAWDistiller v. 1.0, an in-house developed software. The ms2 files were subjected to database searching using SEQUEST (version 27 (rev.9) [3]. Tandem mass spectra of proteins purified from *S. cerevisiae* were compared to 11677 amino acid sequences consisting of 5880 non-redundant *S. cerevisiae* protein sequences obtained from the National Center for Biotechnology (2009-10-27 release). Data sets obtained from *H. sapiens* were searched against 29375 amino acid sequences (2010-11-22 release). The databases also included 172 common contaminant proteins including human keratins, IgGs, and proteolytic enzymes. Randomized versions of each non-redundant protein entry were included in the databases to estimate the false discovery rates (FDR) [4]. All SEQUEST searches were performed with a static modification of +57 Daltons added to cysteine residues to account for carboxamidomethylation, and dynamic searches of +16 Daltons for oxidized methionine.

Spectra/peptide matches were filtered using DTASelect/CONTRAST [5]. In this dataset, spectrum/peptide matches only passed filtering if they were at least 7 amino acids in length and fully tryptic. The DeltCn was required to be at least 0.08, with minimum XCorr value of 1.8 for singly-, 2.0 for doubly-, and 3.0 for triply-charged spectra, and a maximum Sp rank of 10. Proteins that were subsets of others were removed using the parsimony option in DTASelect [5] on the proteins detected after merging all runs. Proteins that were identified by the same set of peptides (including at least one peptide unique to such protein group to distinguish between

isoforms) were grouped together, and one accession number was arbitrarily considered as representative of each protein group.

Quantitation was performed using label-free spectral counting. The number of spectra identified for each protein was used for calculating the distributed normalized spectral abundance factors (dNSAF) [6] and nonspecific proteins removed as previously described [4]. NSAF v7 (an in-house developed software) was used to create the final report on all non-redundant proteins detected across the different runs, estimate false discovery rates (FDR), and calculate their respective distributed Normalized Spectral Abundance Factor (dNSAF) values.

Supplementary Legends

Figure S1. Data analysis workflow. The main dataset is a list of proteins detected in subunits purified from yeast and human of the following chromatin complexes: Sc_INO80/H_INO80, Sc_NuA4/H_TIP60, and Sc_SWR/H_SRCAP. After the removal of non-specific proteins, 26 orthologs proteins specific to three complexes are identified. Spearman correlations were performed between yeast and human baits for the 26 proteins. Using yeast data, we further estimate missing values in human purifications. To identify new conserved interactions associated with chromatin remodeling complexes we extended the analysis to all the orthologs proteins identified in the entire datasets. Three proteins were selected and used for validation by performing reciprocal purifications.

Figure S2. Topological Data Analysis of a Merged Yeast and Human Dataset. Topological data analysis was performed on the protein abundance of a merged yeast and human dataset. *L-infinity centrality* filter with normalized correlation was used (resolution 45 , gain 4.0x). Proteins are colored based on the *L-infinity centrality*. Color bar: red: high values, blue: low values. Biological functions of proteins in the respective flares which exhibit the lowest p-values or

highest count as determined by DAVID annotation tool are listed. For simplicity we show the GO terms that statistically best explain the protein list.

Figure S3. Hierarchical clustering of 26 core proteins in the complexes. Hierarchical clustering was performed on the spectral counts of the 26 conserved proteins between yeast and human. For this clustering we used all the baits from yeast and human that were analyzed here. The Ward method was used and Pearson correlation as the distance. The three chromatin complexes are colored as follows: red represents the yeast INO80/human INO80 complex, blue the yeast Swr1/human SRCAP complex, and green the yeast NuA4/human TIP60 complex. Proteins that are shared between all three complexes are colored in purple.

Figure S4. Clustering analysis on yeast and human baits before and after missing value estimation. Hierarchical clustering was performed on the baits using (A) original spectral counts for the 26 orthologous proteins and (B) using fitted values from the regression analysis (human vs. yeast) for the proteins in the INO80 and INO80C human baits. Baits were separated in three chromatin complexes: red represents the yeast INO80/ human INO80 complex, blue the yeast SWR1/ human SRCAP complex, and green the yeast NuA4/ human TIP60 complexes. INO80 bait represented in the black box with an asterisk was correctly placed after correction in (B) in close proximity with the protein baits belonging to the yeast/human INO80 complex.

Figure S5. Clustering analysis of all orthologous proteins identified in the yeast and human datasets. Unsupervised hierarchical biclustering analysis of the baits and prey proteins using normalized abundance levels. The bait proteins pertaining to the three chromatin complexes are colored as follows: red represents the yeast INO80/human INO80 complex, blue the yeast Swr1/human SRCAP complex, and green the yeast NuA4/human TIP60 complex. Proteins that are shared between all three complexes are colored in purple.

Figure S6. Statistical analysis on TMA19, YAP1802 and DDH1. (A) Qspec [7] analysis was performed on the wild-type and control purifications to determine the differential abundance change. Significant fold changes (in blue) and z-statistics (in orange) were obtained for these proteins. (B) Changes in gene expressions. Significant P-values of these proteins were observed in mutants of the INO80, NuA4 and Swr complexes.

Figure S7. Gene Ontology enrichment analysis. Gene ontology analysis was performed on associated proteins of yeast TMA19/ human TPT1, yeast YAP1802/ human PICALM, and yeast DDH1/ human DDX6 using DAVID tools for functional data annotation.

Table S1. List of proteins detected in the human wild-type using baits from: human INO80, human SRCAP, and human TIP60 before contaminant extraction, after contaminant extraction, and list of a subset proteins detected in the human wild-type using baits from human Ino80, human SRCAP, and human TIP60.

Table S2. List of proteins detected in the yeast wild-type using baits from: yeast Ino80, yeast NuA4, and yeast Swr1, before contaminant extraction, after contaminant extraction, and a list of a subset proteins detected in the yeast wild-type using baits from yeast Ino80, yeast NuA4, and yeast Swr1.

Table S3. List of peptides of a subset proteins detected in the yeast and human wild-type using baits from human INO80, human SRCAP, human TIP60, yeast Ino80, yeast NuA4, and yeast Swr1. List of a spectral counts for subset proteins detected in the yeast and human wild-type using baits from human INO80, human SRCAP, human TIP60, yeast INO80, yeast NuA4, and yeast Swr1.

Table S4. Spearman Correlation coefficients based peptides calculated from 26 orthologs between yeast and human bait purifications. Spearman Correlation coefficients based spectral count calculated from 26 orthologs between yeast and human bait purifications.

Table S5. List of orthologous proteins detected in yeast and human topological network for each of the three main network flares.

Table S6. List of orthologs between proteins identified in the baits of human INO80, human SRCAP, human TIP60, yeast INO80, yeast NuA4, and yeast Swr1. The dNSAF values are included in the table after extracting contaminants. List of expression profiles from a subset of mutants of yeast INO80, yeast NuA4, and yeast Swr1 complexes obtained from Lenstra *et al.* [8]. List of orthologs between proteins identified in at least two purifications of human INO80, human SRCAP, human TIP60, yeast INO80, yeast NuA4, and yeast Swr1. List of orthologs between proteins identified in at least two purifications of human INO80, human SRCAP, human TIP60, yeast INO80, yeast NuA4, and yeast Swr1. List of proteins identified in at least two purifications of human INO80, human SRCAP, human TIP60, yeast INO80, yeast NuA4, and yeast Swr1 and have a P_ value (in microarray data) less than 0.08.

Table S7. List of protein abundance in the DHH1, TMA19, and YAP1802 yeast purifications, list of protein abundance in the DDX6, TPT1, and PICALM human purifications.

Supplementary References

1. Florens L, Washburn MP (2006) Proteomic analysis by multidimensional protein identification technology. *Methods Mol Biol* **328:** 159-175
2. Gilmore JM, Sardiu ME, Venkatesh S, Stutzman B, Peak A, Seidel CW, Workman JL, Florens L, Washburn MP (2012) Characterization of a highly conserved histone related protein, Ydl156w, and its functional associations using quantitative proteomic analyses. *Mol Cell Proteomics* **11:** M111 011544
3. Eng JK, McCormack AL, Yates JR, 3rd (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5:** 976-989

4. Mosley AL, Sardiu ME, Pattenden SG, Workman JL, Florens L, Washburn MP (2011) Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes. *Mol Cell Proteomics* **10:** M110 000687
5. Tabb DL, McDonald WH, Yates JR, 3rd (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1:** 21-26
6. Zhang Y, Wen Z, Washburn MP, Florens L (2010) Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem* **82:** 2272-2281
7. Choi H, Fermin D, Nesvizhskii AI (2008) Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics* **7:** 2373-2385
8. Lenstra TL *et al* (2011) The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* **42:** 536-549

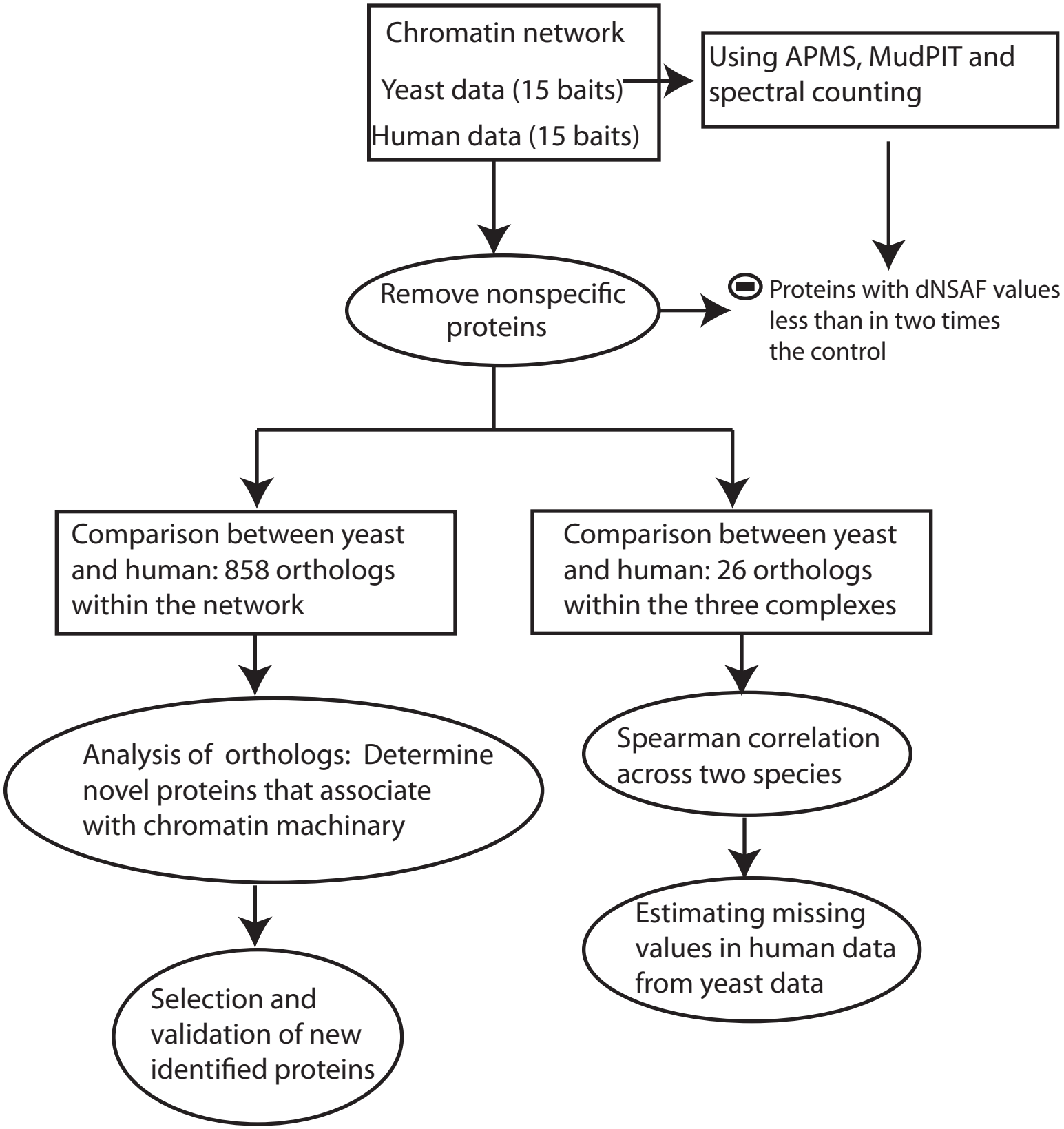


Figure S1. Sardiu et al.

Human-Yeast merged data

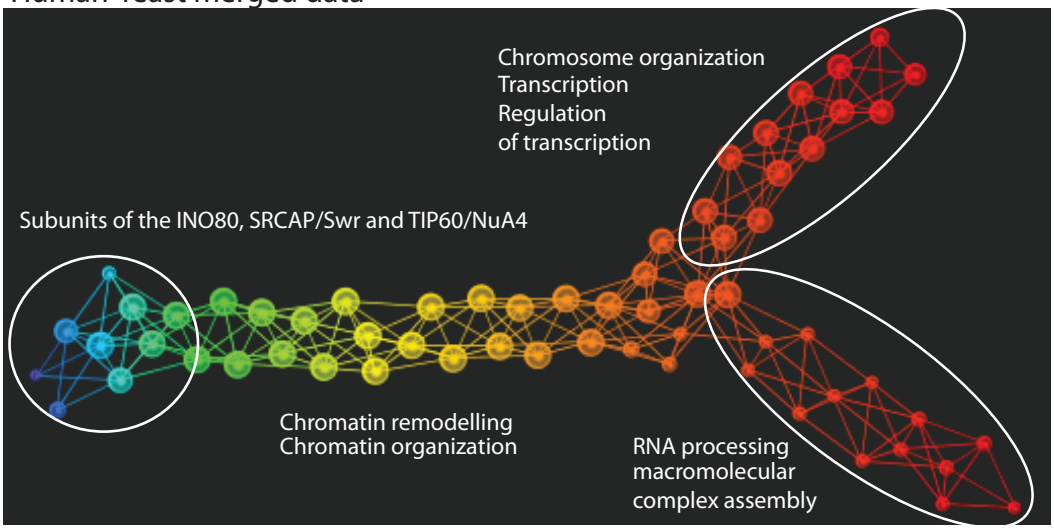


Figure S2. Sardiu et al.

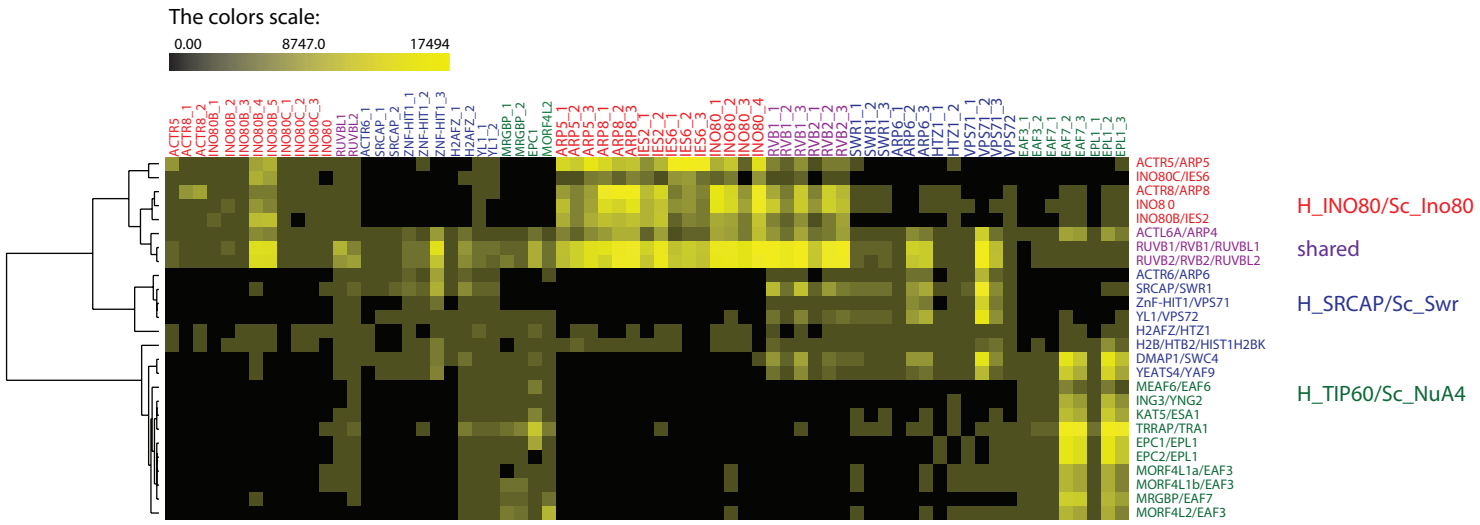


Figure S3. Sardiu et al.

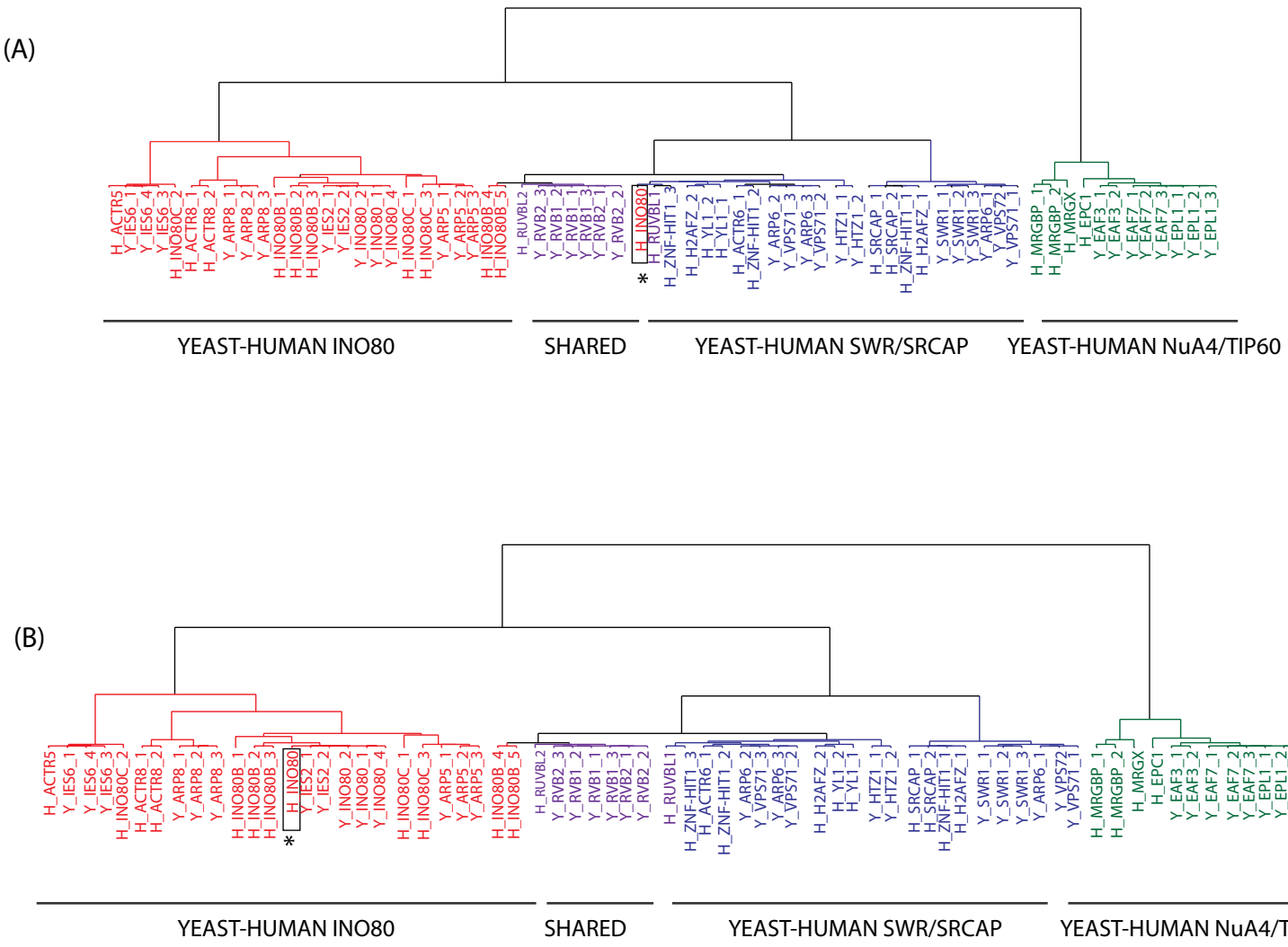


Figure S4. Sardiu et al.

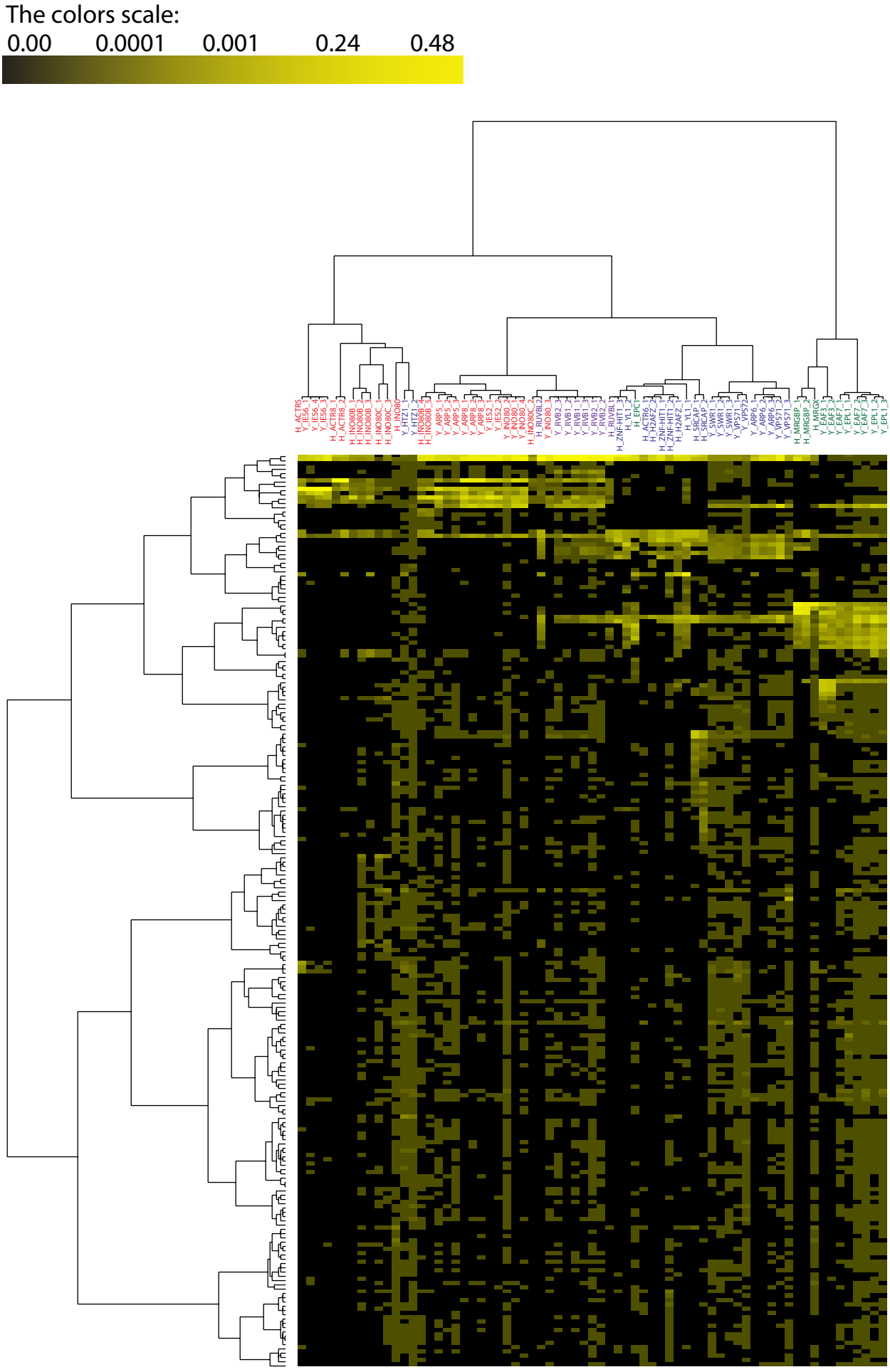
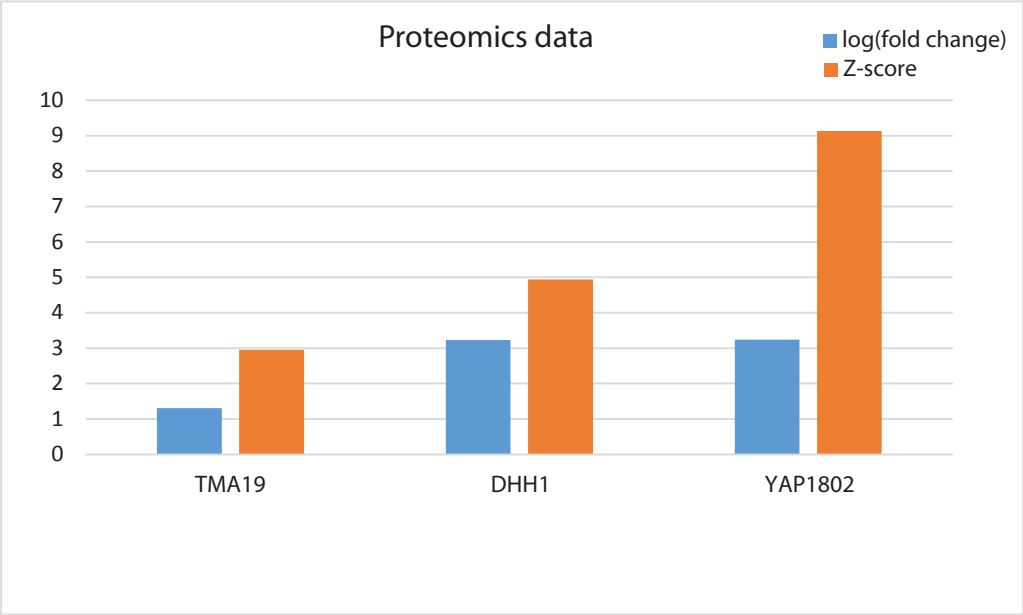


Figure S5. Sardiu et al.

(A)



(B)

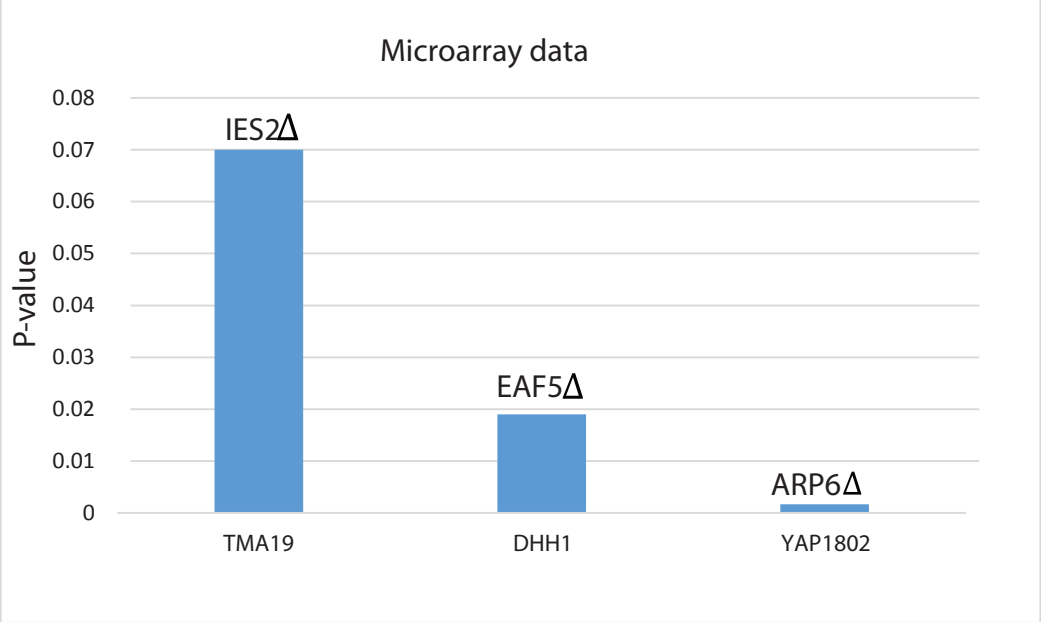


Figure S6. Sardiu et al.

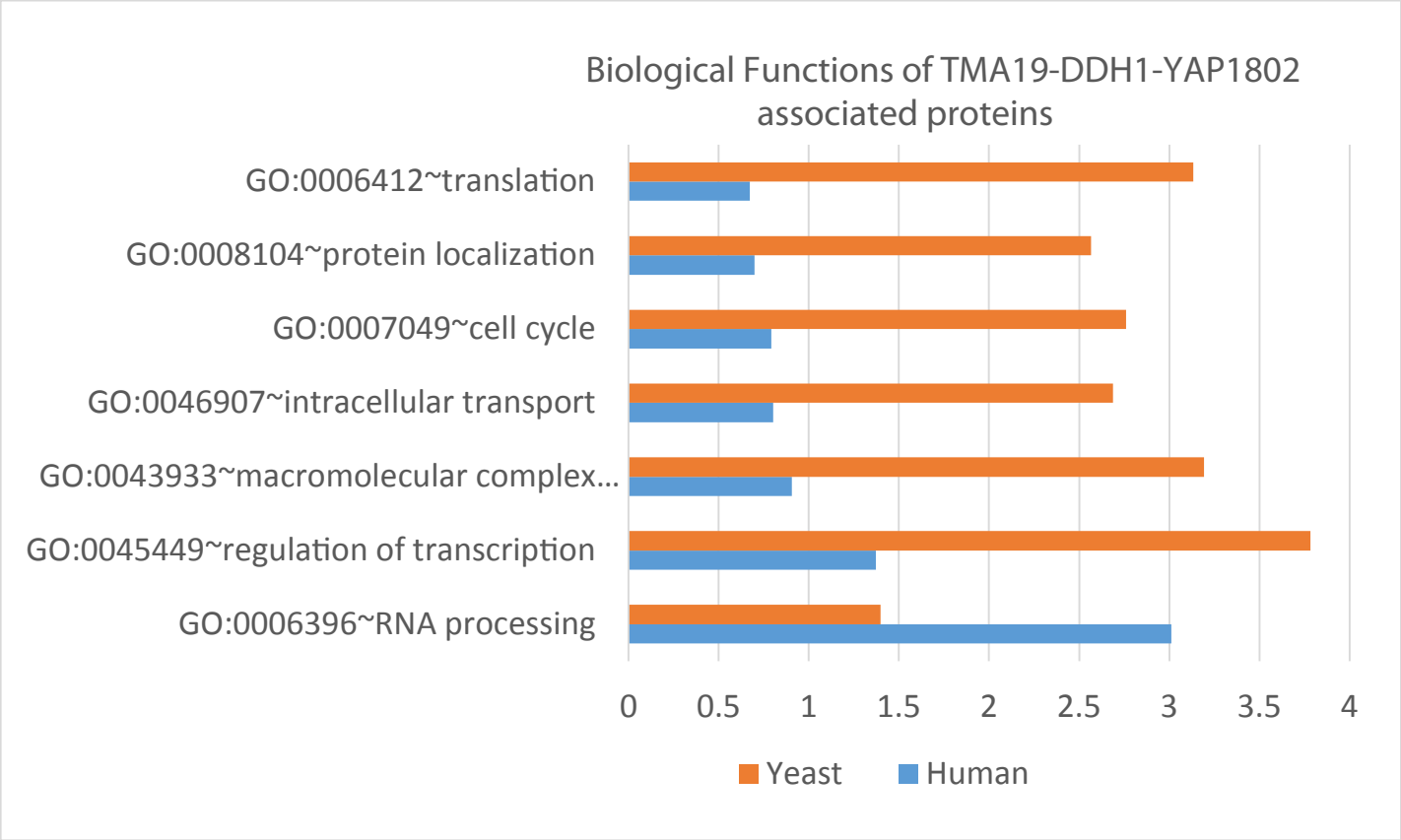


Figure S7. Sardiu et al.