Supporting Information

Wang *et al.* 10.1073/pnas.0806427106

SI Text

SI Methods. Method of frequent-pattern mining (1, 2).

Table S1 is an example for the transaction database, DB, in which the minimum support threshold is 3.

First, a scan of DB derives a list of frequent items, (*F*:4), (*C*:4), (*A*:3), (*B*:3), (*M*:3) (the number after the colon indicates the support), in which items are ordered in frequency-descending order.

Second, the root of a tree is created and labeled with "null." The FP-tree is constructed as follows by scanning the transaction database DB the second time.

1. The scan of the first transaction leads to the construction of the first branch of the tree: ⟨(*F*:1), (*C*:1), (*A*:1), (*M*:1)⟩. Notice that the frequent items in the transaction are listed according to the order in the list of frequent items.
2. For the second transaction, since its (ordered) frequent item list ⟨*F*, *C*, *A*, *B*, M⟩ shares a common prefix ⟨*F*, *C*, *A*⟩ with the existing path ⟨*F*, *C*, *A*, *M*⟩, the count of each node along the prefix is incremented by 1, and 1 new node (*B*:1) is created and linked as a child of (*A*:2) and another new node (*M*:1) is created and linked as the child of (*B*:1).
3. For the third transaction, since its frequent item list ⟨*F*, *B*⟩ shares only the node ⟨*F*⟩ with the F -prefix subtree, *F*′s count is incremented by 1, and a new node (*B*:1) is created and linked as a child of (*F*:3).
4. The scan of the 4th transaction leads to the construction of the second branch of the tree, ⟨(*C*:1), (*B*:1)⟩.
5. For the last transaction, since its frequent item list ⟨*F*, *C*, *A*, *M*⟩ is identical to the first one, the path is shared with the count of each node along the path incremented by 1.

To facilitate tree traversal, an item header table is built in which each item points to its first occurrence in the tree via a node-link. Nodes with the same item-name are linked in sequence via such node-links. After scanning all of the transactions, the tree, together with the associated node-links, are shown in Fig. S1.

Bayesian network method. Bayesian network (BN) provides a tool for representing joint probability distributions of many random variables. It is particularly effective in domains where the interactions between variables are fairly local: each variable directly depends on a small set of other variables. Bayesian networks have been applied extensively for modeling complex domains in different fields. This success is due both to the flexibility of the models and to the naturalness of incorporating expert knowledge into the domains (3).

A Bayesian network, also called causal network, consists of the following:

1. A set of variables and a set of directed edges between variables.
2. Each variable has a finite set of directed edges between variables.
3. The variables together with the directed edges form a directed acyclic graph(DAG).
4. To each variable A with parents $B_1, \ldots, B_n$, there is attached the potential table $P(A|B_1, \ldots, B_n)$.

A bayesian network must have 2 parts. One of them is directed acyclic graph G composed of k nodes, and the other is one conditional probability table (CPT) (4). Conditional probability could be expressed in $P(V_i\,\Pi(V_i))$, to represent relationship of nodes and their parents. Probabilities of the nodes with no parents are defined to transcendental probability. A united probability could be expressed according to a conditional probability chain. The common format is:

$$P(V_1, V_2, \ldots, V_k) = \sum_{i=1}^{n} P(V_i|(V_{i-1} \ldots V_1)) \qquad [1]$$

From the transcendental probability, according to Bayesian rules, one can calculate the distribution of conditional probability of nodes we are interested in. In theory, given the complete united probability function of a random variable, any lower-order united probability could be computed according to the following format (5–7):

$$P(V_1, V_2, \ldots, V_6) = \sum_{i=1}^{6} P(V_i|\Pi(V_i)) = P(V_6|V_5)P(V_5|V_2, V_3)$$
$$P(V_4|V_2)P(V_3|V_1)P(V_2|V_1)P(V_1) \qquad [2]$$

Logistic regression. Denote the gold standards by random variable *Y* and the other genomic features by *X*1, *X*2, . . . , *X*n. Let $Y = 1$ when 2 proteins interact, i.e., they are in the same complex, and $Y = 0$ when the 2 proteins do not interact with each other. The logistic model is of the form eqation (3) where the random vector *X* consists of *X*1, *X*2, . . . , *X*n and their interaction terms (8).

$$\log \frac{Pr(Y = 1)}{1 - Pr(Y = 1)} = \alpha + \beta X \qquad [3]$$

ROC curve analysis. Receiving operator characteristic (ROC) curve is a graphical representation used to assess the discriminatory ability of a dichotomous classifier by showing the tradeoffs between sensitivity and specificity. Sensitivity is calculated by dividing the number of true positives (TP) through the number of all positives, which equals the sum of the true positives and the false negatives (FN); specificity is calculated by dividing the number of true negatives (TN) through the number of all negatives, which equals the sum of the true negatives and the false positives (FP).

$$Sensitivity = TP/(TP + FN),\ Specificity = TN/(TN + FP) \qquad [4]$$

The ROC plot shows specificity on the *x* axis and sensitivity on the *y* axis. A good classifier has its ROC curve climbing rapidly toward upper left hand corner of the graph. This can also be quantified by measuring the area under the curve. The closer the area is to 1.0, the better the classifier is; and the closer the area is to 0.5, the worse the classifier is (8).

KS analysis. Komogorov and Smirnov Statistics is a statistical measure of the discriminations or classifications of one group with another. If we divide the samples into 2 populations, one good and one bad, then KS is defined as the percentage of accumulations of good populations subtracts the percentage of accumulated bad populations (at specific bins or intervals).

$$KS = |\%accumulated_{negativepopulations} - \%accumulated_{positivepopulations}| \qquad [5]$$

If the KS values is closer to 1, the discrimination is strong and there is a clear way to distinguish good populations with bad populations. If the KS value is closer to 0, there is no discrimination. There is no statistical way of differentiate the good

populations with bad ones. KS value provides a quantitative measure of the discrimination or the performance of the classification.

SI Results. ***Comparison of different method.***

$$p(o/m,g,s,p,r) = \frac{p(p/s,o)p(s/m,g,o)p(o/m,g)p(r/m,s,o)}{\sum_o p(p/s,o)p(s/m,g,o)p(o/m,g)p(r/m,s,o)} \quad [6]$$

We got the prediction graph as shown in Fig. S2*B* (4). In the mean time, the conditional probability table is obtained. Therefore, according to the formula (1), we can obtain the predicted results and compare its performance with FPT. In the equation (1), o,m,g,s,p,r separately represent variable output,mips,go,e-s,exp,expr in the Fig. S2*B*.

Add new features. Fig. S3 shows the roc curve comparisons of 5 features and 13 features, and Fig. S4 *a* and *b* shows the KS value comparisons of 5 features and 13 features. From the results we can see that integrating 13 features performs a little better than 5 features. Fig. S5 shows the roc curve comparisons of 4 different methods for 13 features. The results of ROC curve shows that Bayesian network method performs badly when using 13 features. It is probably because Bayesian network method encounters difficulties when used to create too many nodes network. In the meantime, FPT performs a little better than logistic regression and much better than SNB in both training and testing samples as illustrated in Fig. S5. Therefore, we further compare KS values of FPT and logistic regression method in Fig. S4 *c* and *d*. The KS values show that FPT does better when hit rate is between 0.5 and 0.8 whereas logistic regression performs better when hit rate is >0.8. Our goal here is to predict whether there are interactions between 2 proteins, and 0.5 should be an appropriate cut, so FPT is a more useful approach. Finally, we also compare their correct prediction rates when hit rate is 0.5. Fig. S6 shows that FPT predicts more accurately. This supports our conclusion that FPT is a better predictor among all of the other methods included.

Protein–protein interactions in 26S proteasome complex and cytoplasmic ribosome complex. Figs. S7 and S8 are illustrations of protein–protein interactions predicted by mFPT. As we can see mFPT predict more proteins than SNB for 26S proteasome complex and cytoplasmic ribosome complex. See also Table S3 and S4. (blue nodes represent proteins that mFPT and SNB both predict, cyan nodes represent proteins mFPT predict while SNB do not).

For 26S proteasome complex, mFPT predicts 29 proteins more than SNB′13 proteins, and database search tells us that these newly predicted proteins all belong to 26S proteasome or 20S proteansome, which are associated to 26S proteansome. See Fig. S7 and Table S3.

Among these newly predicted proteins, YDL147W, YDR427W, YFR004W YFR010W, YGL011C, YGR232W, YHR200W, YOR117W, YOR261C, YPR108W are regulatory particles of 26S proteasome, YER012W, YER094C, YFR050C, YGR135W, YJL001W, YML092C, YMR314W, YOL038W, YOR362C, YPR103W are regulatory particles of 20S proteasome. The proteasome is an essential component of the ATP-dependent proteolytic pathway in eukaryotic cells and is responsible for the degradation of most cellular proteins. The 20S proteasome contains multiple peptidase activities that function through a new type of proteolytic mechanism involving a threonine active site. The 26S complex, which degrades ubiquitinated proteins, contains in addition to the 20S proteasome a 19S regulatory complex composed of multiple ATPases and components necessary for binding protein substrates (9). Therefore, 20S proteasome is naturally associated with 26S complex.

For Cytoplasmic ribosome complex we can see in Table S4 and Fig. S8, mFPT predicts 122 proteins more than SNB′98 proteins, and in the newly predicted proteins, YDL082W, YMR142C, YHL001W, YMR121C, YNL069C, YNL301C, YGL135W, YBR191W, YHR010W, YDR471W, YFR032CA, YFR031CA, YOR063W, YPL143W, YPL249CA, YBR031W are Protein component of the large (60S) ribosomal subunit, YJL191W, YDL083C, YDR450W, YKR057W, YLR367W, YPR132W, YLR287CA, YPL081W are protein component of the small (40S) ribosomal subunit. YDL130W is Ribosomal protein P1 beta. YDL208W is Nuclear protein related to mammalian high mobility group (HMG) proteins (9).

Ribosomes are highly conserved large ribonucleoprotein (RNP) particles, consisting in yeast of a large 60S subunit and a small 40S subunit, that perform protein synthesis. The 60S subunit contains 42 proteins and 3 RNA molecules. The 40S subunit has a single 18S RNA of 1798 nt(Nucleic Length) and 32 proteins (11). These two kinds of proteins that we found all belong to ribosome proteins.

Tables of FPT trees example and proteins in each protein–protein interaction complex. In Table S1.1, we listed the FPT tree examples. In Table S1.2, we listed proteins predicted in Mitocondrial ribosome complex. In Table S1.3, we listed proteins predicted in 26S complex. In Table S1.4, we listed proteins predicted in Cytoplasmic ribosome complex. In Table S1.5, we listed proteins predicted in new complex with biological function of premRNA splicing. The tables are predictions with both FPT and SNB methods. 1 means there are interactions and 0 means there are no interactions. The name of the proteins and the associated biological functions are listed too.

1. Han J, Pei J, Yin Y, Mao R (2004) Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. *Data Mining and Knowledge Discovery* 8:53–87.
2. Han J, Pei J (2000) Mining Frequent Patterns by Pattern-Growth: Methodology and Implications. *ACM SIGKDD Explorations Newsletter* 2:14–20.
3. Nachman I (2004) Probabilistic Modeling of Gene Regulatory Networks From Data. Ph.D. dissertation, Heberw University.
4. Chen XQ (2004) A data mining technology based on bayesian network. *Journal of Sanming College* 21:47–52.
5. Stephenson TA (2000) An introduction to Bayesian network theory and usage. http://www.idiap.ch/publications/todd00a.bib.abs.html
6. Friedman N, Koller D (2003) Being Bayesian about Network Structure: A Bayesian Approach toStructure Discovery in Bayesian Networks. *Machine Learning* 50:95–126.
7. Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian Network to Analyze Expression Data. *Computational Biology* 7:601–620.
8. Lin N, Wu BL, Jansen R, Gerstein M, Zhao HY (2004) Information assessment on predicting protein–protein interactions. *BMC Bioinformatics* 5:154.
9. Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem* 65:801–847.
10. Venema J, Tollervey D (1999) Ribosome synthesis in Saccharomyces cerevisiae. *Annu Rev Genet* 33:261–311.
11. Vershoor1998 Verschoor A, *et al.* (1998) Three-dimensional structure of the yeast ribosome. *Nucleic Acids Res* 26:655–661.

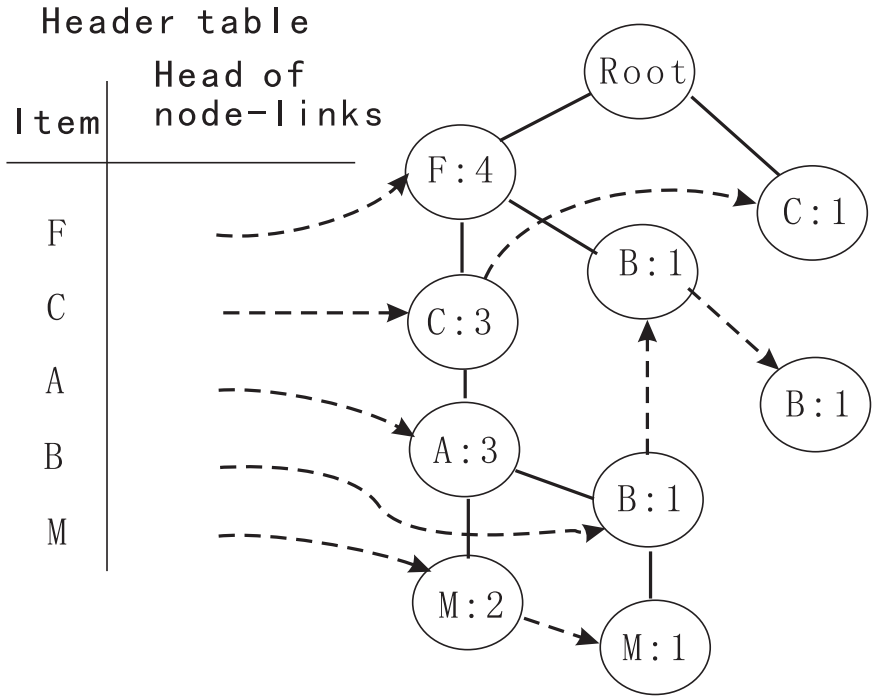

Fig. S1. The FP-tree in Example 1

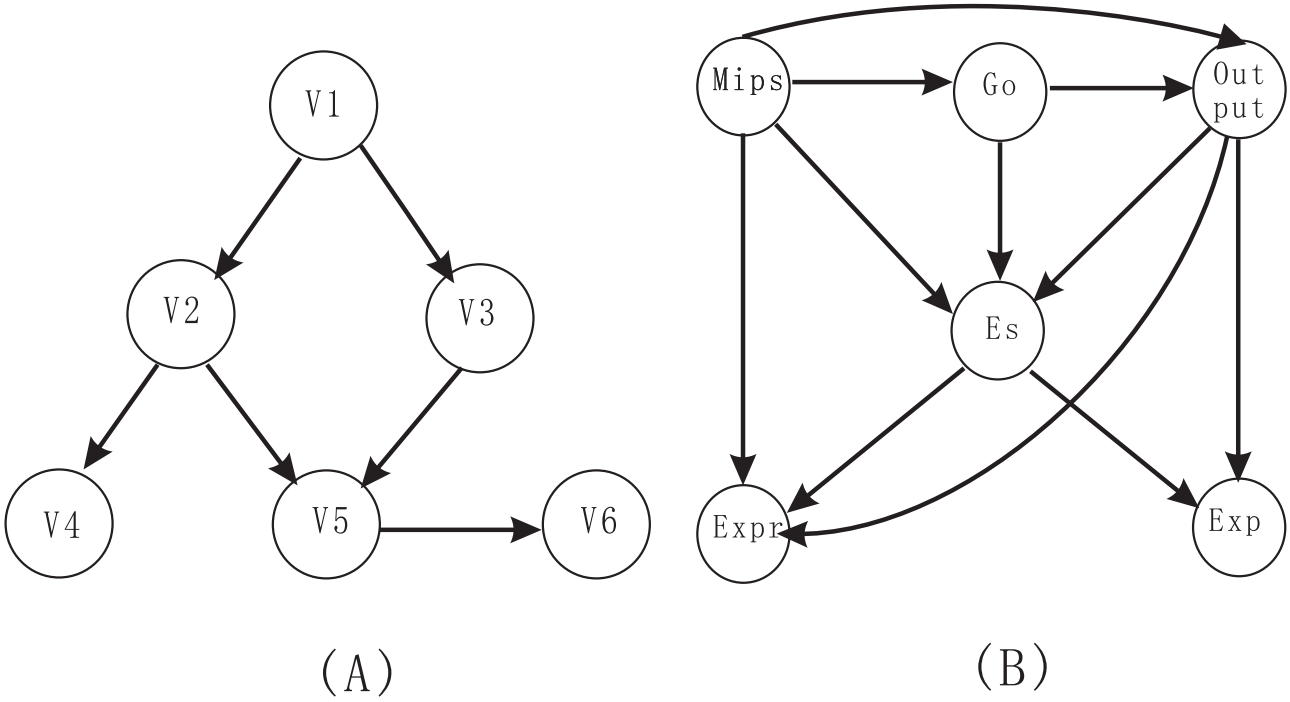

Fig. S2. Bayesian network graph (*A*), and Bayesian network graph of predicting results (*B*).

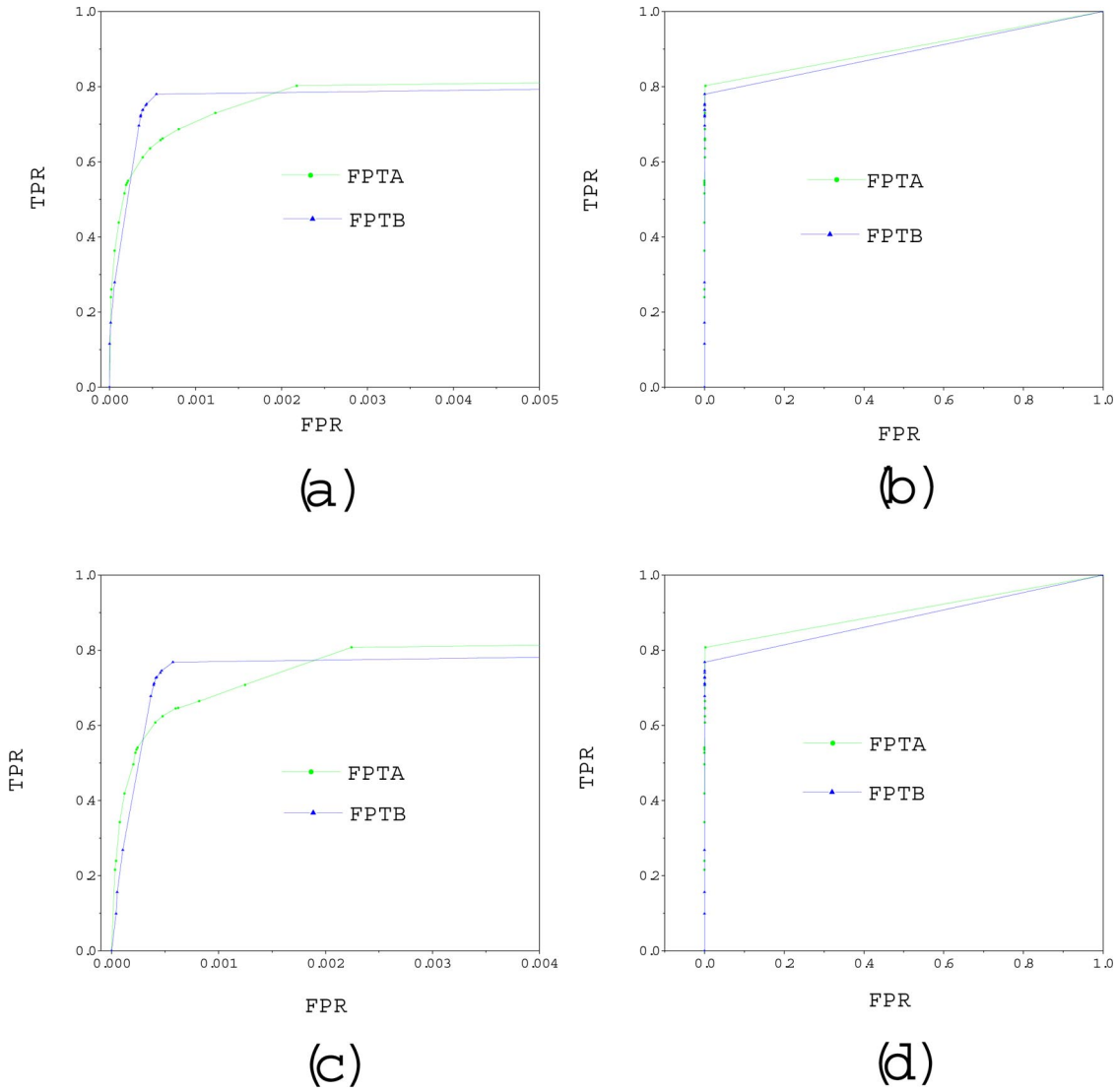

Fig. S3. (*a* and *b*) Roc curve comparisons of 5 features and 13 features for training sample. (*c* and *d*) Roc curve comparisons of 5 features and 13 features for testing sample.

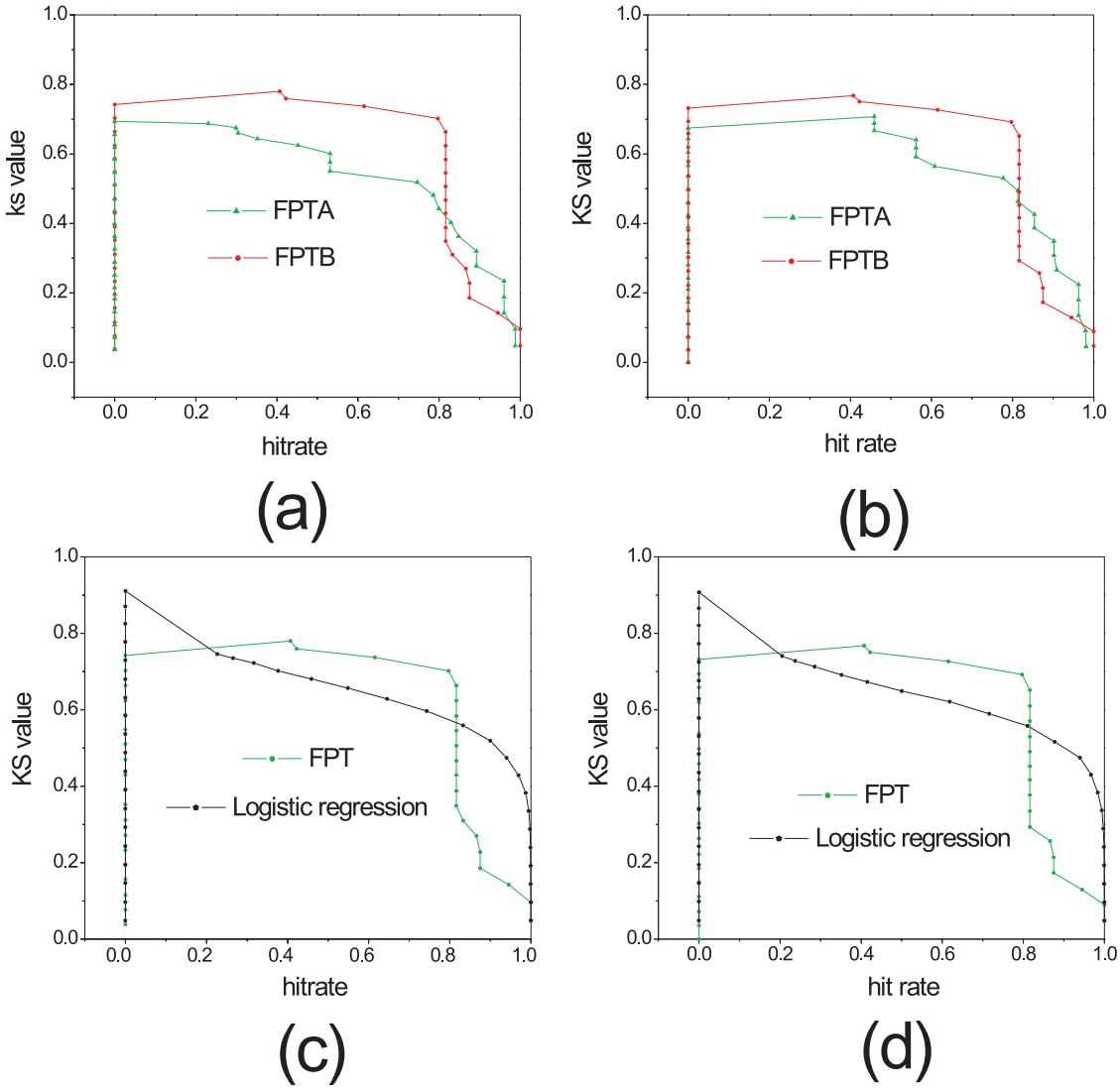

Fig. S4. (*a* and *b*)KS value comparisons of 5 (*a*) and 13 (*b*) features for training (*a*) and testing (*b*) samples. (*c* and *d*) KS value comparisons of different methods for 13 features, respectively, for training (*c*) and testing (*d*) samples.

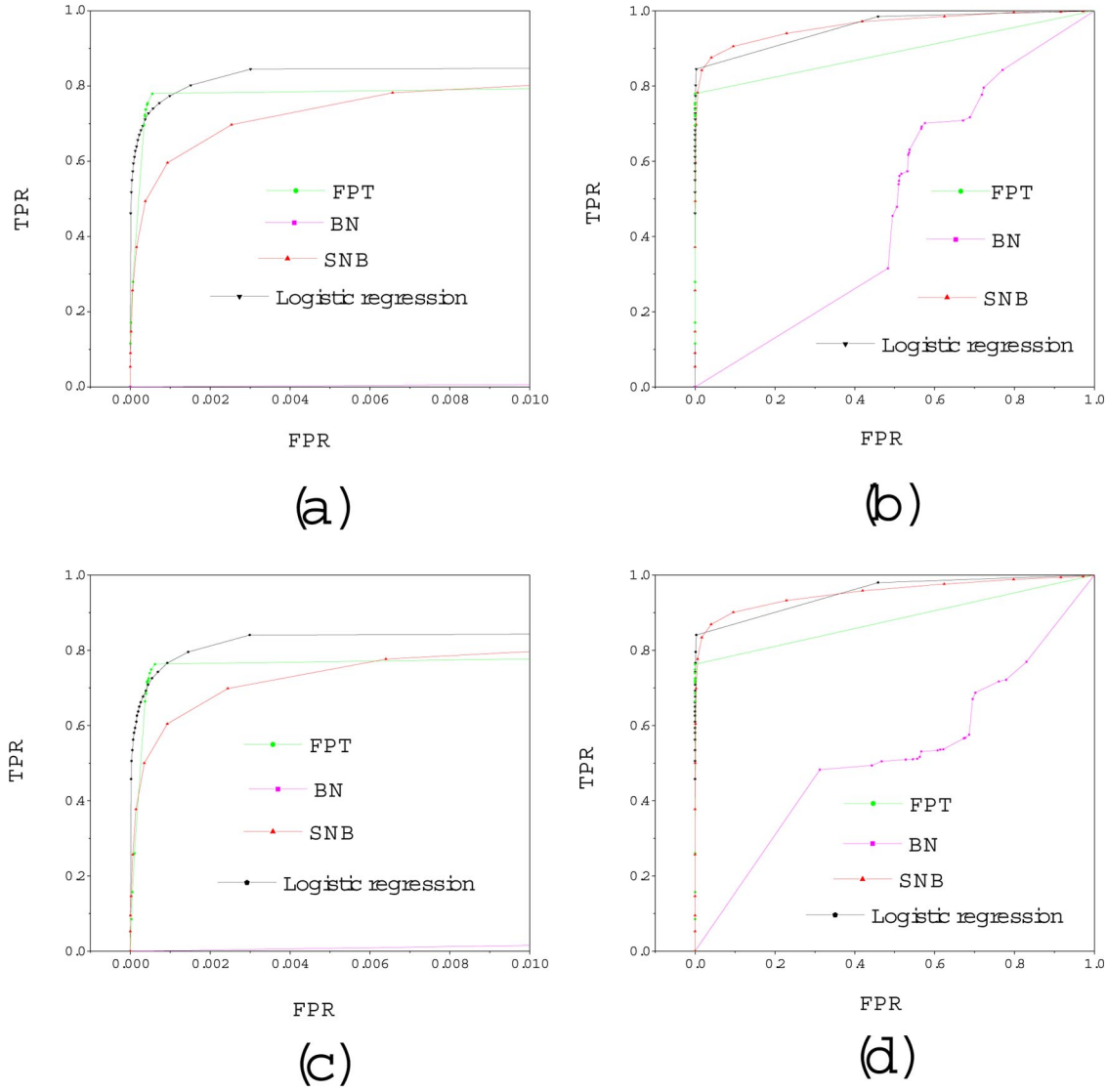

Fig. S5. (*a* and *b*) Roc curve comparisons of the training samples for 4 methods for 13 features. (*c* and *d*) Roc curve comparisons of the testing samples for 4 methods for 13 features.

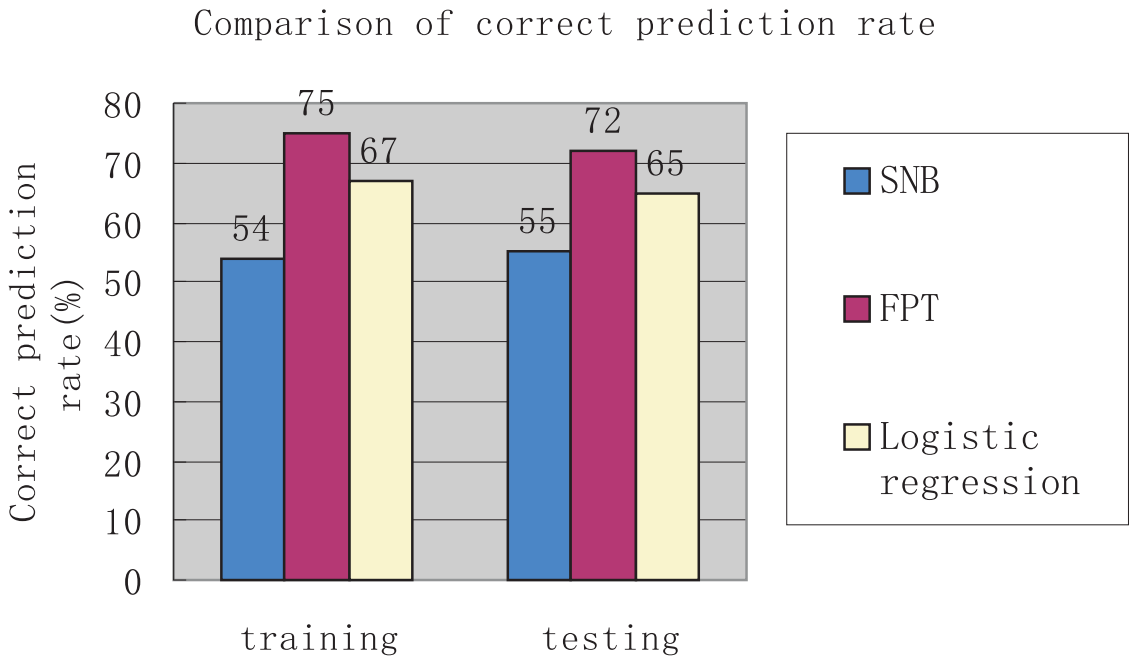

Fig. S6. Comparisons of correct prediction rate for 3 methods for 13 features for both training and testing data samples.

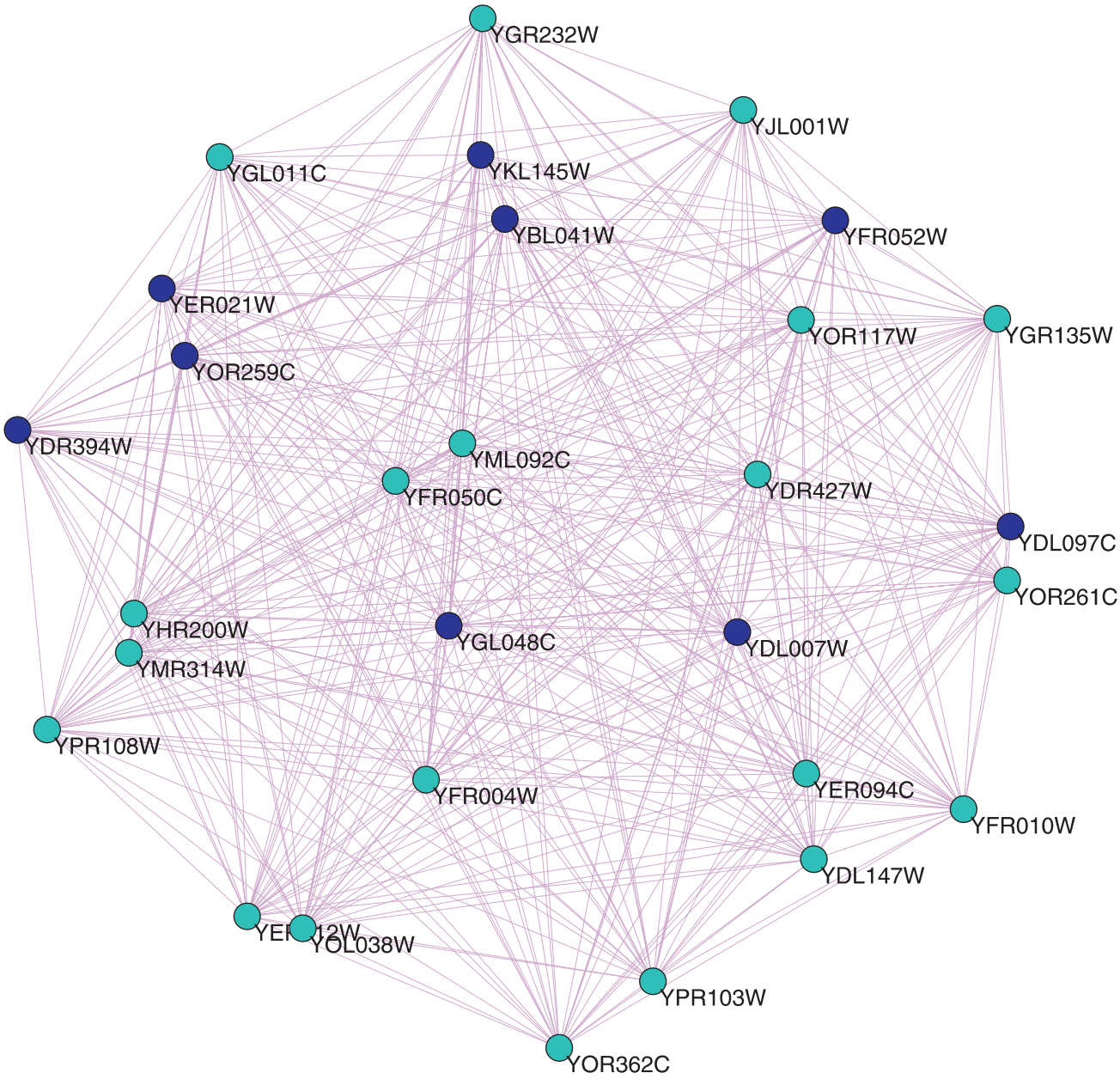

Fig. S7. The 26S proteasome complex.

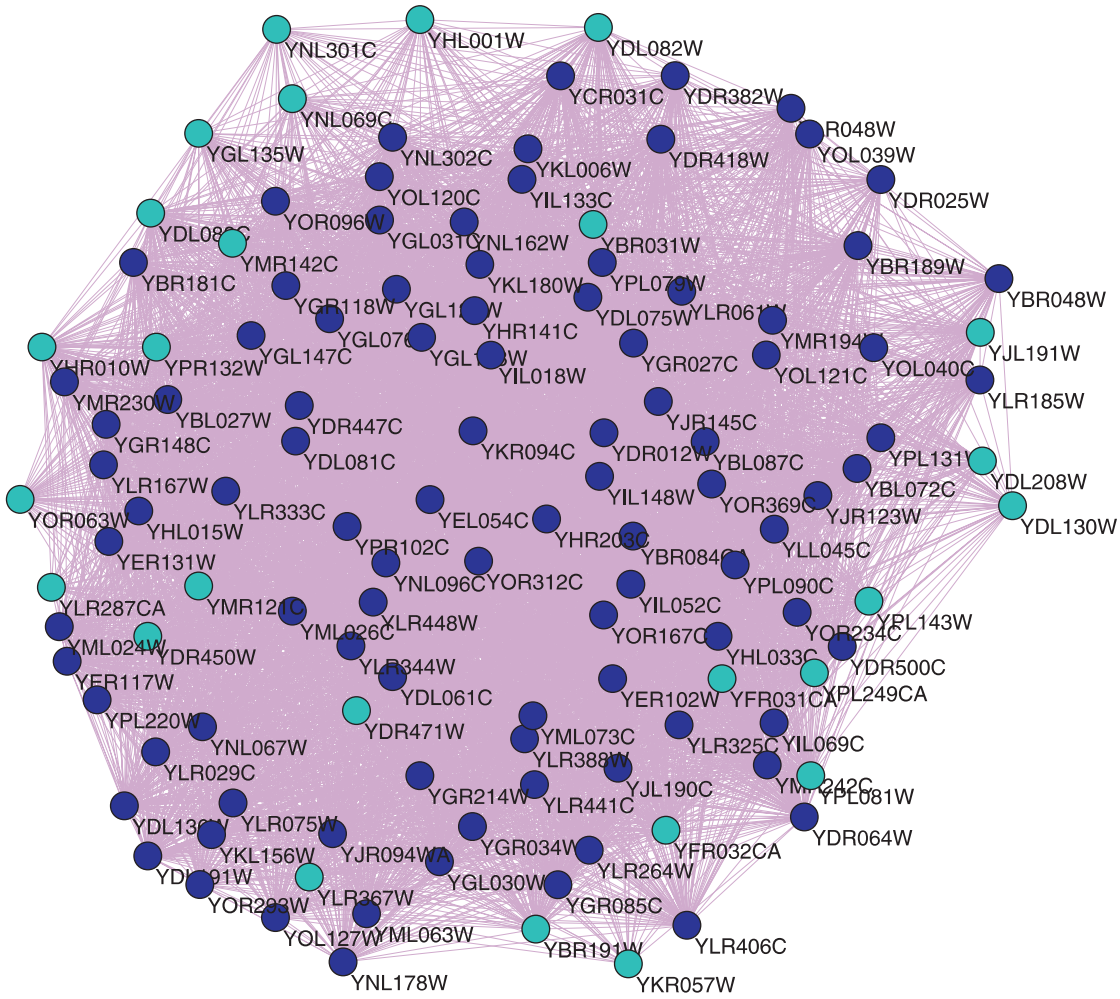

Fig. S8. Cytoplasmic ribosome complex.

Table S1. The FP-treein Example 1

Tid	Items bought	(Ordered) frequent items
1	*F, A, C, D, G, I, M*	*F, C,A,M*
2	*A, B, C, F, L, M, O*	*F, C,A,B,M*
3	*B, F, H, J, O*	*F, B*
4	*B, C, K, S*	*C, B*
5	*A, F, C, E, L, M, N*	*F, C,A,M*

Table S2. Mitocondrial ribosome complex. MRP represents Mitochondrial ribosomal protein

Systematic name	Standard name	FPT	SNB	Discription
YDR115W	YDR115W	1	1	Putative MRPL
YLR069C	MEF1	1	1	Mitochondrial elongation factor
YBL090W	MRP21	1	1	MRPL
YBR146W	YBR146W	1	1	MRPS
YBR282W	MRPL27	1	1	MRPL
YCR003W	MRPL32	1	1	MRPL
YDR116C	MRPL1	1	1	MRPL
YDR237W	MRPL7	1	1	MRPL
YDR337W	MRPS28	1	1	MRPS
YHR147C	MRPL6	1	1	MRPL
YJL096W	MRPL49	1	1	MRPL
YKL138C	MRPL31	1	1	MRPL
YKR006C	MRPL13	1	1	MRPL
YKR085C	MRPL20	1	1	MRPL
YMR193W	MRPL24	1	1	MRPL
YNL252C	MRPL17	1	1	MRPL
YNL306W	MRPL18	1	1	MRPL
YOR150W	MRPL23	1	1	MRPL
YOR158W	PET123	1	1	MRPS
YBR024W	SCO2	1	0	Mitochondrial inner membrane protein
YBR120C	CBP6	1	0	Mitochondrial translational activator
YGL143C	MRF1	1	0	Mitochondrial translation release factor
YLR203C	MSS51	1	0	Mitochondrial Splicing Suppressor
YOL023W	IFM1	1	0	Mitochondrial translation initiation factor
YPL104W	MSD1	1	0	Mitochondrial aspartyl-tRNA synthetase
YPL183WA	YPL183W-A	1	0	Likely to be a MRP
YPR047W	MSF1	1	0	Mitochondrial aminoacyl-tRNA Synthetase
YBL038W	MRPL16	1	0	MRPL
YBR122C	MRPL36	1	0	MRPL
YBR251W	MRPS5	1	0	MRPS
YBR268W	MRPL37	1	0	MRPL
YCR046C	IMG1	1	0	MRPL
YDL045WA	MRP10	1	0	MRPS
YDL202W	MRPL11	1	0	MRPL
YDR322W	MRPL35	1	0	MRPL
YDR347W	MRP1	1	0	MRPS
YDR405W	MRP20	1	0	MRPL
YDR462W	MRPL28	1	0	MRPL
YGR076C	MRPL25	1	0	MRPL
YGR084C	MRP13	1	0	MRPS
YGR220C	MRPL9	1	0	MRPL
YHL004W	MRP4	1	0	MRPS
YJL063C	MRPL8	1	0	MRPL
YKL167C	MRP49	1	0	MRPL
YKL170W	MRPL38	1	0	MRPL
YLR312WA	MRPL15	1	0	MRPL
YLR439W	MRPL4	1	0	MRPL
YML009C	MRPL39	1	0	MRPL
YML025C	YML6	1	0	MRPL
YMR024W	MRPL3	1	0	MRPL
YMR225C	MRPL44	1	0	MRPL
YMR257C	PET111	1	0	Mitochondrial translational activator
YNL005C	MRP7	1	0	MRPL
YNL137C	NAM9	1	0	MRPS
YNL185C	MRPL19	1	0	MRPL
YNL284C	MRPL10	1	0	MRPL
YNR037C	RSM19	1	0	MRPS
YPR166C	MRP2	1	0	MRPS
YGL068W	MNP1	0	1	Mitochondrial-Nucleoid Protein
YMR188C	MRPS17	0	1	MRPS
YNL081C	SWS2	0	1	Putative MRPS

MRPL, mitochondrial ribosomal protein of the large subunit; MRPS, mitochondrial ribosomal protein of the small subunit.

Table S3. 26S complex

Systematic name	Standard name	FPT	SNB	Discription
YBL041W	PRE7	1	1	20S proteasome
YDL007W	RPT2	1	1	26S proteasome
YDL097C	RPN6	1	1	26S proteasome
YDR394W	RPT3	1	1	26S proteasome
YER021W	RPN3	1	1	26S proteasome
YFR052W	RPN12	1	1	26S proteasome
YGL048C	RPT6	1	1	26S proteasome
YKL145W	RPT1	1	1	26S proteasome
YOR259C	RPT4	1	1	26S proteasome
YDL147W	RPN5	1	0	26S proteasome
YDR427W	RPN9	1	0	26S proteasome
YER012W	PRE1	1	0	20S proteasome
YER094C	PUP3	1	0	20S proteasome
YFR004W	RPN11	1	0	26S proteasome
YFR010W	UBP6	1	0	26S proteasome
YFR050C	PRE4	1	0	20S proteasome
YGL011C	SCL1	1	0	26S proteasome
YGR135W	PRE9	1	0	20S proteasome
YGR232W	NAS6	1	0	26S proteasome
YHR200W	RPN10	1	0	26S proteasome
YJL001W	PRE3	1	0	20S proteasome
YML092C	PRE8	1	0	20S proteasome
YMR314W	PRE5	1	0	20S proteasome
YOL038W	PRE6	1	0	20S proteasome
YOR117W	RPT5	1	0	26S proteasome
YOR261C	RPN8	1	0	26S proteasome
YOR362C	PRE10	1	0	20S proteasome
YPR103W	PRE2	1	0	20S proteasome
YPR108W	RPN7	1	0	26S proteasome
YDL126C	CDC48	0	1	ATPase in ER
YGR253C	PUP2	0	1	20S proteasome
YHR027C	RPN1	0	1	26S proteasome
YIL075C	RPN2	0	1	26S proteasome

Table S4. Citoplasmic ribosome complex. RPL represents Protein component of the large (60S) ribosomal subunit. RPS represents Protein component of the small (40S) ribosomal subunit

Systematic name	Standard name	FPT	SNB	Discription
YLR075W	RPL10	1	1	RPL
YPR102C	RPL11A	1	1	RPL
YGR085C	RPL11B	1	1	RPL
YEL054C	RPL12A	1	1	RPL
YDR418W	RPL12B	1	1	RPL
YKL006W	RPL14A	1	1	RPL
YLR029C	RPL15A	1	1	RPL
YIL133C	RPL16A	1	1	RPL
YKL180W	RPL17A	1	1	RPL
YOL120C	RPL18A	1	1	RPL
YBR084CA	RPL19A	1	1	RPL
YBL027W	RPL19B	1	1	RPL
YPL220W	RPL1A	1	1	RPL
YMR242C	RPL20A	1	1	RPL
YOR312C	RPL20B	1	1	RPL
YPL079W	RPL21B	1	1	RPL
YLR061W	RPL22A	1	1	RPL
YBL087C	RPL23A	1	1	RPL
YER117W	RPL23B	1	1	RPL
YGL031C	RPL24A	1	1	RPL
YGR148C	RPL24B	1	1	RPL
YOL127W	RPL25	1	1	RPL
YLR344W	RPL26A	1	1	RPL
YGR034W	RPL26B	1	1	RPL
YGL103W	RPL28	1	1	RPL
YIL018W	RPL2B	1	1	RPL
YGL030W	RPL30	1	1	RPL
YDL075W	RPL31A	1	1	RPL
YLR406C	RPL31B	1	1	RPL
YOR234C	RPL33B	1	1	RPL
YIL052C	RPL34B	1	1	RPL
YDL191W	RPL35A	1	1	RPL
YDL136W	RPL35B	1	1	RPL
YMR194W	RPL36A	1	1	RPL
YLR185W	RPL37A	1	1	RPL
YDR500C	RPL37B	1	1	RPL
YLR325C	RPL38	1	1	RPL
YIL148W	RPL40A	1	1	RPL
YKR094C	RPL40B	1	1	RPL
YNL162W	RPL42A	1	1	RPL
YHR141C	RPL42B	1	1	RPL
YJR094WA	RPL43B	1	1	RPL
YDR012W	RPL4B	1	1	RPL
YPL131W	RPL5	1	1	RPL
YML073C	RPL6A	1	1	RPL
YLR448W	RPL6B	1	1	RPL
YGL076C	RPL7A	1	1	RPL
YHL033C	RPL8A	1	1	RPL
YLL045C	RPL8B	1	1	RPL
YGL147C	RPL9A	1	1	RPL
YNL067W	RPL9B	1	1	RPL
YDL081C	RPP1A	1	1	Ribosomal Protein P1 Alpha
YOL039W	RPP2A	1	1	Ribosomal Protein P2 Alpha
YDR382W	RPP2B	1	1	Ribosomal Protein P2 Beta
YGR214W	RPS0A	1	1	RPS
YLR048W	RPS0B	1	1	RPS
YOR293W	RPS10A	1	1	RPS
YMR230W	RPS10B	1	1	RPS
YDR025W	RPS11A	1	1	RPS
YBR048W	RPS11B	1	1	RPS
YOR369C	RPS12	1	1	RPS
YDR064W	RPS13	1	1	RPS
YCR031C	RPS14A	1	1	RPS

Systematic name	Standard name	FPT	SNB	Discription
YOL040C	RPS15	1	1	RPS
YML024W	RPS17A	1	1	RPS
YDR447C	RPS17B	1	1	RPS
YML026C	RPS18B	1	1	RPS
YOL121C	RPS19A	1	1	RPS
YNL302C	RPS19B	1	1	RPS
YLR441C	RPS1A	1	1	RPS
YML063W	RPS1B	1	1	RPS
YGL123W	RPS2	1	1	RPS
YHL015W	RPS20	1	1	RPS
YJL190C	RPS22A	1	1	RPS
YGR118W	RPS23A	1	1	RPS
YIL069C	RPS24B	1	1	RPS
YGR027C	RPS25A	1	1	RPS
YLR333C	RPS25B	1	1	RPS
YER131W	RPS26B	1	1	RPS
YKL156W	RPS27A	1	1	RPS
YOR167C	RPS28A	1	1	RPS
YLR264W	RPS28B	1	1	RPS
YLR388W	RPS29A	1	1	RPS
YDL061C	RPS29B	1	1	RPS
YNL178W	RPS3	1	1	RPS
YLR167W	RPS31	1	1	RPS
YJR145C	RPS4A	1	1	RPS
YHR203C	RPS4B	1	1	RPS
YJR123W	RPS5	1	1	RPS
YPL090C	RPS6A	1	1	RPS
YBR181C	RPS6B	1	1	RPS
YOR096W	RPS7A	1	1	RPS
YNL096C	RPS7B	1	1	RPS
YBL072C	RPS8A	1	1	RPS
YER102W	RPS8B	1	1	RPS
YBR189W	RPS9B	1	1	RPS
YDL208W	NHP2	1	0	rRNA processing
YDL082W	RPL13A	1	0	RPL
YMR142C	RPL13B	1	0	RPL
YHL001W	RPL14B	1	0	RPL
YMR121C	RPL15B	1	0	RPL
YNL069C	RPL16B	1	0	RPL
YNL301C	RPL18B	1	0	RPL
YGL135W	RPL1B	1	0	RPL
YBR191W	RPL21A	1	0	RPL
YHR010W	RPL27A	1	0	RPL
YDR471W	RPL27B	1	0	RPL
YFR032CA	RPL29	1	0	RPL
YFR031CA	RPL2A	1	0	RPL
YOR063W	RPL3	1	0	RPL
YPL143W	RPL33A	1	0	RPL
YPL249CA	RPL36B	1	0	RPL
YBR031W	RPL4A	1	0	RPL
YDL130W	RPP1B	1	0	Ribosomal Protein P1 Beta
YJL191W	RPS14B	1	0	RPS
YDL083C	RPS16B	1	0	RPS
YDR450W	RPS18A	1	0	RPS
YKR057W	RPS21A	1	0	RPS
YLR367W	RPS22B	1	0	RPS
YPR132W	RPS23B	1	0	RPS
YLR287CA	RPS30A	1	0	RPS
YPL081W	RPS9A	1	0	RPS
YHR010W	RPL27A	0	1	RPL
YLR340W	RPP0	0	1	ribosomal protein P0

Table S5. New complex

Systematic name	Standard name	FPT	SNB	Discription
YLR298C	YHC1	1	0	U1 snRNP
YGR013W	SNU71	1	0	U1 snRNP
YIL061C	SNP1	1	0	U1 snRNP
YDR240C	SNU56	1	0	U1 snRNP
YBL074C	AAR2	1	0	U5 snRNP
YER029C	SMB1	1	0	Sm B
YGR074W	SMD1	1	0	Sm D1
YLR275W	SMD2	1	0	Sm D2
YLR147C	SMD3	1	0	Sm D3
YOR159C	SME1	1	0	Sm E
YPR182W	SMX3	1	0	Sm F
YMR213W	CEF1	1	0	splicing factor
YLR117C	CLF1	1	0	splicesome assembly factor
YKL173W	SNU114	1	0	U5 snRNP
YMR125W	STO1	1	0	nuclear mRNA degradation
YBL026W	LSM2	1	0	Sm-like protein
YER112W	LSM4	1	0	Sm-like
YBR055C	PRP6	1	0	pre-mRNA processing
YDL030W	PRP9	1	0	pre-mRNA processing
YDL043C	PRP11	1	0	pre-mRNA processing
YMR268C	PRP24	1	0	pre-mRNA processing
YDR473C	PRP3	1	0	pre-mRNA processing
YPR178W	PRP4	1	0	pre-mRNA processing
YML046W	PRP39	1	0	pre-mRNA processing
YDR243C	PRP28	1	0	Pre-mRNA Processing
YGL120C	PRP43	1	0	Pre-mRNA Processing
YGR006W	PRP18	1	0	Pre-mRNA Processing
YGR075C	PRP38	1	0	Pre-mRNA Processing
YGR091W	PRP31	1	0	Pre-mRNA Processing
YHR165C	PRP8	1	0	Pre-mRNA Processing
YJL203W	PRP21	1	0	Pre-mRNA Processing
YKL012W	PRP40	1	0	Pre-mRNA Processing
YLL036C	PRP19	1	0	Pre-mRNA Processing
YDR235W	PRP42	1	0	U1 snRNP
YML049C	RSE1	1	0	pre-mRNA splicing
YMR240C	CUS1	1	0	U2 snRNP
YHR086W	NAM8	1	0	U1 snRNP
YDR088C	SLU7	1	0	RNA splicing factor
YPR057W	BRR1	1	0	pre-mRNA splicing
YBR119W	MUD1	1	0	U1 snRNP
YOR319W	HSH49	1	0	U2-snRNP
YIR009W	MSL1	1	0	U2 snRNP