Hierarchical hidden community detection for protein complex prediction (*Supplementary Material*)

Chao Li[1], Kun He[1,*], Guangshuai Liu[1], Bart Selman[2], and John E. Hopcroft[2]

[1] *School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China*
[2] *Department of Computer Science, Cornell University, Ithaca NY, 14853, U.S.A.*
[*] *brooklet60@hust.edu.cn*

SUPPLEMENTAL MATERIALS

The source code of HirHide, the most updated version and documents are at https://github.com/JHL-HUST/HirHide/

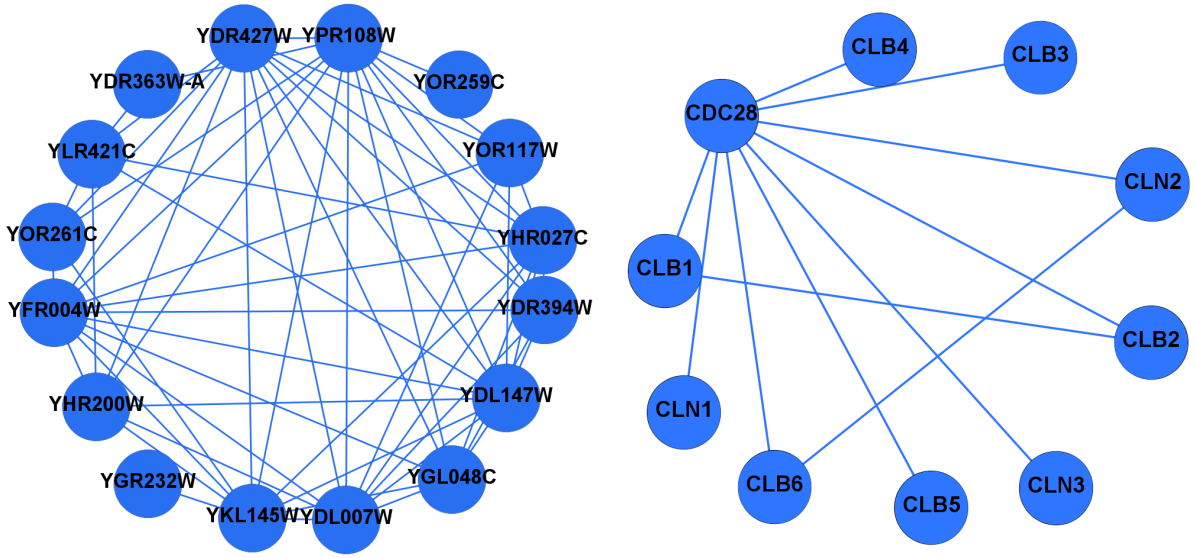
SUPPLEMENTAL TABLE AND FIGURE

We compare the information of hidden complexes with dominant complexes about gene ontology in AmiGO 2 [1]. The hidden complexes contain all complexes with at least 25 percent of proteins belong to stronger complexes while the dominant complexes contain no protein that belongs to stronger complexes. For fairness, the two sets have the same number of complexes. Table S1 illustrates the great difference between the two sets.

Table S1. Comparison of complexes with high hiddenness values and complexes with low hiddenness values. 'Ancestor evidence' indicates the number of the biological aspect of ancestor evidence used in manual assertion.

Complexes with high hiddenness values (hiddenness values)	Ancestor evidence	Complexes with low hiddenness values (hiddenness values)	Ancestor evidence
U5 snRNP complex (0.50)	1535	Ada2p/Gcn5p/Ada3 transcription activator (0.00)	307
commitment complex (0.81)	781	Arp2/3 protein complex (0.00)	847
DNA replication factor C complex (0.57)	759	cAMP-dependent protein kinase (0.00)	391
Nuclear exosome complex (0.83)	1388	GINS complex (0.00)	340
protein phosphatase type 2A complex (0.67)	1077	HIR complex (0.00)	143
Swr1p complex (0.46)	713	ribonuclease H2 complex (0.00)	295
SAGA complex (0.80)	1175	Ssh1p translocon complex (0.00)	179
transcription factor TFIIH complex (0.47)	1800	TRAMP complex (0.00)	246
U2 snRNP complex (0.83)	2226	transcription factor TFIIIC complex (0.00)	682
U4/U6 x U5 tri-snRNP complex (0.68)	2319	TRAPP complex (0.00)	1552
transcription factor TFIID complex (0.47)	1800	RNA polymerase I transcription factor complex (0.00)	143
Swr1p complex (0.46)	713	prefoldin complex (0.00)	478
eIF3 (0.43)	172	COMPASS complex (0.00)	666
DNA-directed RNA polymerase III complex (0.41)	2321	transcription export complex (0.00)	729
ubiquitin ligase ERAD-L complex (0.40)	534	mitochondrial inner membrane peptidase complex (0.00)	270
multi-eIF complex (0.38)	320	nucleotide-excision repair factor 3 complex (0.00)	144
DNA-directed RNA polymerase I complex (0.36)	1520	nuclear condensin complex (0.00)	503
PAM complex (0.33)	487	RNA polymerase I upstream activating factor complex (0.00)	20
mitochondrial sorting and assembly machinery (0.25)	365	Sec61p translocon complex (0.00)	269
Ino80p complex (0.25)	872	ESCRT III complex (0.00)	1032
nuclear ubiquitin ligase complex (0.25)	2536	HOPS complex (0.00)	540
glycine cleavage complex (0.25)	169	RecQ helicase-Topo III complex (0.00)	106
Average	**1128**	**Average**	**449**
Median	**975**	**Median**	**324**

In Fig. S1, we illustrate some examples of the predicted protein complexes identified by HirHide. They are sparse and covered by stronger complexes but can potentially be protein complexes if the datasets become complete.



(a) (b)

Fig. S1. Examples of predicted protein complexes identified by HirHide. The name of the proteins making up complexes are shown in the blue nodes.

REFERENCES

1. S. Carbon, A. Ireland, C. J. Mungall, S. Shu, B. Marshall, S. Lewis, A. Hub, and W. P. W. Group, "Amigo: online access to ontology and annotation data," Bioinformatics **25**, 288–289 (2009).