Supplementary Materials: Assessing the role of multi-protein complexes in determining phenotype

Nolwenn Le Meur and Robert Gentleman
Fred Hutchinson Cancer Center Research, Program in Computational Biology,
Division of Public Health Sciences, 1100 Fairview Avenue North -
M2-B876, P.O. Box 19024, Seattle, WA 98109, USA

This document is the supplemental file for the manuscript entitled "Assessing the role of multi-protein complexes in determining phenotype" by Le Meur and Gentleman (submitted).

Data sources

Experimental phenotypic datasets

We made use of 5 different *Saccharomyces cerevisiae* phenotypic datasets. First, we used 2 lethal phenotype datasets published by Giaever *et al.* [1] and Dudley *et al.* [2], and presented in the main manuscript. Then, we investigated the environmental stress conditions tested by Giaever *et al.* [1] and Dudley *et al.* [2], as well as an other, but less well-known, gene deletion set by Kastenmayer *et al.* [3].

Cellular organizational units

S. cerevisiae multi-protein complex co-membership was determined from GO [4, 5], MIPS [6], protein-protein interactions data obtained from the IntAct database (http://www.ebi.ac.uk/intact/site) and estimates from tandem affinity purification-mass spectrometry experiments (AP-MS) [7–11]. This resulted in an estimated interactome of 398 curated multi-protein complexes from the online databases (GO, MIPS, IntAct), and 549 multi-protein complexes estimated from the AP-MS experiments [12]. The curated complexes are named by their database ID. The estimated multi-protein complexes are named using the prefix *apComplex* followed by the author and year of the experiment, and an arbitrary identification number [12].

Pathways were extracted from the Kyoto Encyclopedia of Genes and Genomes, KEGG [13]. KEGG contains 99 pathways specific to *S. cerevisiae* out of 918 referenced, corresponding to 1,205 unique genes in which 332 are essential and 94 are haploinsufficient.

Mappings between yeast genes, GO categories, and KEGG pathways were obtained from the R metadata package, *YEAST*, available from the Bioconductor Project (http://www.bioconductor.org).

Gene coverage

The genome of *S. cerevisiae* is not entirely annotated and only a small proportion of the genes are covered by our interactome or the KEGG pathways (Table S1). As mentioned above, we can divide our interactome in 2 subsets: the curated multi-protein complexes,*i.e.,* annotated in the GO, MIPS or IntAct databases, and the predicted multi-protein complexes, estimated from high throughput proteomic experiments using the Bioconductor package *apComplex* [12]. The curated interactome currently counts

1,629 genes and 398 multi-protein complexes. The whole interactome (curated and predicted multi-protein complexes) is now composed of 947 multi-protein complexes and 1,803 genes (including 695 essential genes and 100 haploinsufficient genes).

	Genes	KEGG pathways	Curated interactome	Curated and Predicted interactome
Total ORFs	6609	1205	1629	1803
Essential	1101	332	587	695
Haploinsufficient	184	94	84	100

Table S1: Genes coverage. The first row of the table presents the total number of ORFs in S. cerevisiae genome (charaterized, uncharaterized, and dubious) and their coverage in KEGG pathways and our protein interactome (curated and predicted multi-protein complexes databases). The 2 following rows show the representation of the essential, and haploinsufficient genes.

Protein complexes and phenotype

Lethal phenotypes

The following tables present the results of the Hypergeometric test applied on the essential genes (Table S2) and the haploinsufficient genes (Table S3) using the global interactome (*i.e.,* curated and predicted multi-protein complexes). We note that only the GO and MIPS complexes are annotated and have an description as the other are predicted complexes.

Complex	Observed	Expected	Size	Odds	P-value	Description
GO:0005732	42	21.59	56	5.03	1.98e−08	small nucleolar ribo...
GO:0005666	17	6.55	17	Inf	8.11e−08	DNA-directed RNA pol...
MIPS-410.30	16	6.17	16	Inf	2.13e−07	Pre-replication comp...
apCompGavin2002: 228	18	7.32	19	29.43	3.75e−07	-
GO:0005656	15	5.78	15	Inf	5.61e−07	pre-replicative comp...
MIPS-360	28	13.88	36	5.77	1.50e−06	Proteasome
MIPS-410.35	18	7.71	20	14.70	2.40e−06	Replication complex
apCompGavin2002: 231	18	7.71	20	14.70	2.40e−06	-
MIPS-510.120	13	5.01	13	Inf	3.87e−06	RNA polymerase III
apCompGavin2002: 224	14	5.78	15	22.76	1.43e−05	-
GO:0046540	22	10.79	28	6.00	1.63e−05	U4/U6 x U5 tri-snRNP...
apCompGavin2002: 203	11	4.24	11	Inf	2.66e−05	-
apCompGavin2002: 50	19	9.25	24	6.20	5.36e−05	-
apCompGavin2002: 12	16	7.32	19	8.68	5.50e−05	-
GO:0000172	10	3.85	10	Inf	6.96e−05	ribonuclease MRP com...
GO:0005847	13	5.78	15	10.54	1.70e−04	mRNA cleavage and po...
GO:0005669	13	5.78	15	10.54	1.70e−04	transcription factor...
MIPS-360.10.10	13	5.78	15	10.54	1.70e−04	20S proteasome
apCompGavin2002: 43	13	5.78	15	10.54	1.70e−04	-
GO:0005849	9	3.47	9	Inf	1.82e−04	mRNA cleavage factor...
GO:0005655	9	3.47	9	Inf	1.82e−04	nucleolar ribonuclea...
apCompGavin2002: 205	9	3.47	9	Inf	1.82e−04	-

Continued on next page

Complex	Observed	Expected	Size	Odds	P-value	Description
GO:0005681	22	11.95	31	3.99	2.29e−04	spliceosome
apCompGavin2002: 61	12	5.40	14	9.72	3.88e−04	-
apCompGavin2002: 229	12	5.40	14	9.72	3.88e−04	-
apCompKrogan2004: 39	12	5.40	14	9.72	3.88e−04	-
GO:0043614	8	3.08	8	Inf	4.75e−04	multi-eIF complex
GO:0000145	8	3.08	8	Inf	4.75e−04	exocyst
MIPS-130	8	3.08	8	Inf	4.75e−04	Chaperonine containi...
MIPS-410.20	8	3.08	8	Inf	4.75e−04	Replication initiati...
MIPS-440.12.20	8	3.08	8	Inf	4.75e−04	RNase MRP
apCompGavin2002: 71	8	3.08	8	Inf	4.75e−04	-
apCompGavin2002: 73	8	3.08	8	Inf	4.75e−04	-
apCompKrogan2004: 9	8	3.08	8	Inf	4.75e−04	-
apCompKrogan2004: 13	8	3.08	8	Inf	4.75e−04	-
apCompGavin2002: 250	10	4.24	11	16.16	5.00e−04	-
GO:0000177	9	3.85	10	14.52	1.19e−03	cytoplasmic exosome ...
GO:0005675	9	3.85	10	14.52	1.19e−03	holo TFIIH complex
apCompGavin2002: 53	9	3.85	10	14.52	1.19e−03	-
apCompGavin2002: 126	9	3.85	10	14.52	1.19e−03	-
apCompHo2002: 86	9	3.85	10	14.52	1.19e−03	-
GO:0019774	7	2.70	7	Inf	1.24e−03	proteasome core comp...
MIPS-160	7	2.70	7	Inf	1.24e−03	Exocyst complex
MIPS-440.12.10	7	2.70	7	Inf	1.24e−03	Exosome complex
apCompGavin2002: 52	7	2.70	7	Inf	1.24e−03	-
apCompGavin2002: 202	7	2.70	7	Inf	1.24e−03	-
apCompKrogan2004: 14	7	2.70	7	Inf	1.24e−03	-
apCompGavin2002: 14	12	5.78	15	6.47	1.26e−03	-
GO:0005665	10	4.63	12	8.07	1.96e−03	DNA-directed RNA pol...
apCompGavin2002: 200	10	4.63	12	8.07	1.96e−03	-
apCompGavin2002: 223	10	4.63	12	8.07	1.96e−03	-
apCompKrogan2004: 2	10	4.63	12	8.07	1.96e−03	-
apCompKrogan2004: 4	11	5.40	14	5.92	2.67e−03	-
apCompGavin2002: 181	18	10.41	27	3.25	2.69e−03	-
MIPS-510.180.10.30	8	3.47	9	12.89	2.82e−03	NEF3 complex
apCompKrogan2004: 61	8	3.47	9	12.89	2.82e−03	-
MCM2-7 heterohexamer	6	2.31	6	Inf	3.24e−03	-
Signal recognition particle	6	2.31	6	Inf	3.24e−03	-
GO:0000127	6	2.31	6	Inf	3.24e−03	transcription factor...
GO:0005664	6	2.31	6	Inf	3.24e−03	nuclear origin of re...
apCompGavin2002: 62	6	2.31	6	Inf	3.24e−03	-
apCompHo2002: 237	6	2.31	6	Inf	3.24e−03	-
apCompKrogan2004: 56	6	2.31	6	Inf	3.24e−03	-
apCompHo2002: 76	12	6.17	16	4.85	3.29e−03	-
apCompGavin2002: 158	13	6.94	18	4.20	3.79e−03	-
apCompGavin2002: 56	9	4.24	11	7.26	4.32e−03	-
apCompGavin2002: 217	9	4.24	11	7.26	4.32e−03	-

Continued on next page

Complex	Observed	Expected	Size	Odds	P-value	Description
GO:0000176	10	5.01	13	5.38	5.57e−03	nuclear exosome (RNa...
MIPS-510.40.10	10	5.01	13	5.38	5.57e−03	RNA polymerase II
apCompGavin2002: 91	10	5.01	13	5.38	5.57e−03	-
apCompGavin2002: 209	10	5.01	13	5.38	5.57e−03	-
apCompHo2002: 21	11	5.78	15	4.44	6.55e−03	-
GO:0030915	7	3.08	8	11.26	6.60e−03	Smc5-Smc6 complex
apCompKrogan2004: 55	7	3.08	8	11.26	6.60e−03	-
apCompGavin2002: 42	12	6.55	17	3.88	7.29e−03	-
apCompGavin2002: 182	14	8.09	21	3.23	8.14e−03	-
DNA replication factor C complex	5	1.93	5	Inf	8.44e−03	-
GO:0000120	5	1.93	5	Inf	8.44e−03	RNA polymerase I tra...
GO:0000799	5	1.93	5	Inf	8.44e−03	nuclear condensin co...
GO:0042765	5	1.93	5	Inf	8.44e−03	GPI-anchor transamid...
GO:0032040	5	1.93	5	Inf	8.44e−03	small subunit proces...
MIPS-445.10	5	1.93	5	Inf	8.44e−03	SCF-CDC4 complex
apCompGavin2002: 51	5	1.93	5	Inf	8.44e−03	-
apCompGavin2002: 70	5	1.93	5	Inf	8.44e−03	-
apCompGavin2002: 99	5	1.93	5	Inf	8.44e−03	-
apCompGavin2002: 175	5	1.93	5	Inf	8.44e−03	-
apCompGavin2002: 194	5	1.93	5	Inf	8.44e−03	-
apCompGavin2002: 207	5	1.93	5	Inf	8.44e−03	-
apCompHo2002: 34	5	1.93	5	Inf	8.44e−03	-
apCompHo2002: 166	5	1.93	5	Inf	8.44e−03	-
apCompKrogan2004: 43	5	1.93	5	Inf	8.44e−03	-
apCompKrogan2004: 52	5	1.93	5	Inf	8.44e−03	-
apCompKrogan2004: 81	5	1.93	5	Inf	8.44e−03	-
GO:0008541	8	3.85	10	6.44	9.34e−03	proteasome regulator...
apCompGavin2002: 155	8	3.85	10	6.44	9.34e−03	-
apCompGavin2002: 196	8	3.85	10	6.44	9.34e−03	-
apCompGavin2002: 199	8	3.85	10	6.44	9.34e−03	-
apCompGavin2002: 204	8	3.85	10	6.44	9.34e−03	-
apCompGavin2002: 221	8	3.85	10	6.44	9.34e−03	-
apCompHo2002: 81	8	3.85	10	6.44	9.34e−03	-
apCompKrogan2004: 30	8	3.85	10	6.44	9.34e−03	-
apCompKrogan2004: 82	8	3.85	10	6.44	9.34e−03	-

Table S2: Essentiality can be attributed to some multi-protein complexes. These complexes (curated and predicted) present an over-representation of essential genes(p-value<0.01). Observed: number of essential genes in the complex; Expected: expected number of essential genes in the complex; Size: total number of genes in the complex; Odds: odds ratios; P-value: p-value of the Hypergeometric test; Description: fullname. Note that when the multi-protein complex is entirely composed of essential genes (Observed = Size) the odds ratio are infinite (Inf).

	Observed	Expected	Size	Odds	P-value	Description
MIPS-130	7	0.44	8	128.11	1.01e−08	Chaperonine containi...
GO:0005732	16	3.11	56	7.92	1.62e−08	small nucleolar ribo...
GO:0005832	7	0.61	11	31.97	3.61e−07	chaperonin-containin...
GO:0005665	7	0.67	12	25.56	8.28e−07	DNA-directed RNA pol...
MIPS-510.40.10	7	0.72	13	21.29	1.71e−06	RNA polymerase II
apCompGavin2002: 223	6	0.67	12	18.05	1.77e−05	-
GO:0000176	6	0.72	13	15.47	3.14e−05	nuclear exosome (RNa...
GO:0000177	5	0.55	10	17.87	9.60e−05	cytoplasmic exosome ...
apCompHo2002: 141	3	0.17	3	Inf	1.66e−04	-
MIPS-440.12.10	4	0.39	7	23.61	2.74e−04	Exosome complex
apCompGavin2002: 12	6	1.05	19	8.30	3.77e−04	-
apCompKrogan2004: 5	5	0.72	13	11.15	4.29e−04	-
MIPS-510.40	8	1.94	35	5.40	4.57e−04	RNA polymerase II ho...
apCompGavin2002: 181	7	1.50	27	6.33	4.64e−04	-
apCompKrogan2004: 10	4	0.44	8	17.70	5.25e−04	-
apCompKrogan2004: 22	4	0.44	8	17.70	5.25e−04	-
apCompGavin2002: 143	3	0.22	4	52.64	6.36e−04	-
apCompHo2002: 32	3	0.22	4	52.64	6.36e−04	-
apCompGavin2002: 49	5	0.78	14	9.91	6.37e−04	-
apCompGavin2002: 14	5	0.83	15	8.91	9.14e−04	-
apCompGavin2002: 79	5	0.83	15	8.91	9.14e−04	-
apCompGavin2002: 175	3	0.28	5	26.30	1.53e−03	-
apCompKrogan2004: 52	3	0.28	5	26.30	1.53e−03	-
apCompGavin2002: 42	5	0.94	17	7.42	1.72e−03	-
apCompKrogan2004: 2	4	0.67	12	8.83	3.13e−03	-
apCompGavin2002: 41	5	1.11	20	5.92	3.78e−03	-
MIPS-510.120	4	0.72	13	7.84	4.32e−03	RNA polymerase III
apCompKrogan2004: 74	3	0.39	7	13.14	4.92e−03	-
GO:0005736	4	0.78	14	7.05	5.80e−03	DNA-directed RNA pol...
GO:0043614	3	0.44	8	10.50	7.56e−03	multi-eIF complex
apCompKrogan2004: 13	3	0.44	8	10.50	7.56e−03	-
apCompGavin2002: 50	5	1.33	24	4.66	8.68e−03	-
GO:0005850	2	0.17	3	34.73	8.81e−03	eukaryotic translati...
GO:0000928	2	0.17	3	34.73	8.81e−03	gamma-tubulin small ...
apCompGavin2002: 251	2	0.17	3	34.73	8.81e−03	-
apCompHo2002: 44	2	0.17	3	34.73	8.81e−03	-
apCompHo2002: 118	2	0.17	3	34.73	8.81e−03	-
apCompHo2002: 205	2	0.17	3	34.73	8.81e−03	-

Table S3: Haploinsufficiency can be attributed to some multi-protein complexes. These complexes (curated and predicted) present an over-representation of haploinsufficient genes (p-value<0.01). Observed: number of haploinsufficient genes in the complex; Expected: expected number of haploinsufficient genes in the complex; Size: total number of genes in the complex; Odds: odds ratios; P-value: p-value of the Hypergeometric test; Description: fullname. Note that when the multi-protein complex is entirely composed of haploinsufficient genes (Observed = Size) the odds ratio are infinite (Inf).

Kastenmayer *et al.* (2006) undertook the first functional studies of small open reading frames (sORFs), using *S. cerevisiae* as a model. Phenotypic analyses of the new gene-deletion strains identified 22 sORFs required for haploid growth, growth at high temperature, growth in the presence of a non-fermentable carbon source, or growth in the presence of DNA damage and replication-arrest agents. We looked if those 22 sORFs are randomly distributed among protein complexes or if they cluster within a set of them. We identified 11 critical complexes with an over-representation of sORFs (listed below). As observed by Kastenmayer *et al.* (2006), the sORFs are well conserved across species compare to the other genes that compose the protein complexes (Figure S1). We also note that 6 of the identified complexes were also found in our analysis of the essential genes.

```
---------Condition: sORF -----------

GO:0046540  U4/U6 x U5 tri-snRNP complex
GO:0042729  DASH complex
GO:0005732  small nucleolar ribonucleoprotein complex
MIPS-270.20.30  Dam1 protein complex
GO:0000776  kinetochore
MIPS-270.20  Outer Kinetochor Protein Complex
GO:0005665  DNA-directed RNA polymerase II, core complex
MIPS-510.40.10  RNA polymerase II
MIPS-510.120  RNA polymerase III
GO:0005736  DNA-directed RNA polymerase I complex
GO:0005666  DNA-directed RNA polymerase III complex
```

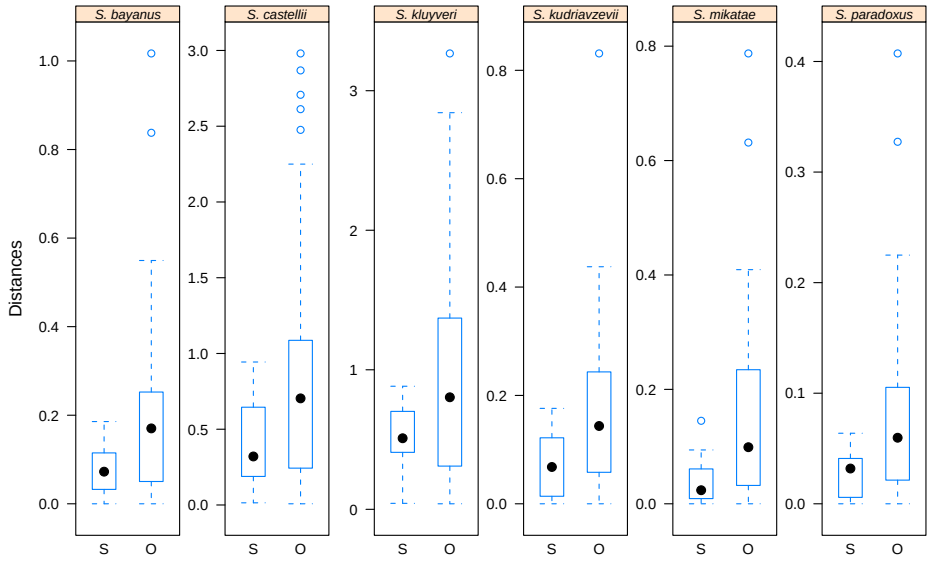

Figure S1: sORFs from critical complexes are highly conserved across species. Each panel presents a comparison between *S. cerevisiae* and one other species (named in the panel strip). In each panel, each boxplot shows the distribution of the gene evolution distances between the 2 species, calculated using the RSD approach [14]. The 'S' boxplot represents the distribution of distances for the genes inducing a lethal phenotype (Sensitive) and the 'O' represents the other set of genes. In all cases the median distance and the spread (IQR) are smaller for the sORFs.

Environmental stress conditions and fitness growth defect phenotypes

While our approach focuses on understanding the functional roles that underly lethal phenotypes in rich media, these methods can also be used to investigate other environmental conditions or other phenotypic changes. As an example, using the curated subset of our interactome, we looked at phenotypic datasets by Giaever *et al.* [1] and Dudley *et al.* [2] where they measured fitness growth defects in various environmental stress conditions.

Giaever *et al.* (2002)

As a first intention, we used the same filtering parameters as described in Giaever *et al.* [1] to select genes that present a significant growth defect according to the condition and generation time under study.

Table S4 resumes our results and shows that some but not all experimental conditions give rise to phenotype that can be attributed to some multi-protein complexes. In reality, we can not rule out the fact the phenotype originated from the other experimental conditions are not associated to any multi-protein complexes as our view and coverage of the *S. cerevisae* interactome is incomplete.

	Giaever et al. (2002)	Interactome	p.value	Complexes
pH8g15	225	41	0	10
pH8g5	275	54	0.002	9
nystatin15	46	8	0.006	3
ypg15	30	3	0.009	1
nystatin5	171	37	0.016	-
minimalPlus15	93	14	0.034	-
ypg5	23	6	0.036	-
minimalC5	183	27	0.162	-
trpM5	343	49	0.166	-
sorbitol15	59	6	0.21	-
NaCl15	334	47	0.269	-
sorbitol5	356	48	0.283	-
lysM5	304	36	0.286	-
NaCl5	175	22	0.345	-
minimalPlus5	262	33	0.86	-

Table S4: Some phenotypic changes induced in environmental stress conditions (Giaever et al. 2002) are tightly associated with multi-protein complexes. Each row corresponds to an environmental stress condition and different generation time (5, 15). The first column indicates the number of mutants with growth defect in Giaever's experiment. The second columns indicates the number of those deleted genes in the interactome. The third columns presents the p-value obtained by the graph theory test. A p-value < 0.01 indicates that those deleted genes are not randomly distributed in the multi-protein complexes of the interactome. The fourth column indicates the number of multi-protein complexes involved. The differents conditions are: ypg: yeast/peptone/galactose 5 gen. rep. a and b; sorbitol: 1.5M Sorbitol (sugar, osmotic stress); NaCl: 1M NaCl (salt, osmotic stress); lysM: lysine minus (lack of required AA); thM: threonine minus (lack of required AA); trpM: tritophanee minus (lack of required AA); minimalPlus: minimal + histidine/leuvine/uracile; minimalC: minimal complete; nystatin: Nystatin (antifungal drug); pH8: pH 8 (alkali stress).

The following list shows the different multi-protein complexes involved in the different experimental conditions.

```
---------Condition: nystatin15 -----------
GO:0000813  ESCRT I complex
MIPS-260.70  Vps4p ATPase complex (Vps protein complex)
GO:0000815  ESCRT III complex
---------Condition: pH8g15 -----------
MIPS-260.20  Clathrin-associated protein (AP) complex
GO:0030122  AP-2 adaptor complex
GO:0030121  AP-1 adaptor complex
GO:0005955  calcineurin complex
GO:0030123  AP-3 adaptor complex
MIPS-260.20.10  AP-1 complex
EBI-1249909  Calcineurin variant 1
GO:0048188  COMPASS complex
MIPS-140.30.30.30  Dynactin complex
GO:0005869  dynactin complex
---------Condition: pH8g5 -----------
MIPS-260.20  Clathrin-associated protein (AP) complex
GO:0000812  SWR1 complex
GO:0005955  calcineurin complex
GO:0030123  AP-3 adaptor complex
GO:0030121  AP-1 adaptor complex
EBI-1249909  Calcineurin variant 1
GO:0016593  Cdc73/Paf1 complex
GO:0030122  AP-2 adaptor complex
MIPS-260.20.10  AP-1 complex
---------Condition: ypg15 -----------
MIPS-510.190.80  GAL80 complex
```

Under nystatin condition, we identify or example the three ESCRT complexes or Endosomal Sorting Complexes Required for Transport (*GO:0000813, GO:0000814, GO:0000815*) as significantly related to the growth defect. At pH8, one of the interesting complex seems to be the Gene Ontology complex GO:0000812 or SWR1 complex, composed of 13 genes. It is a multi-subunit protein complex that is involved in chromatin remodeling. It is required for the incorporation of the histone variant H2AZ into chromatin. In *S. cerevisiae*, the complex contains Swr1p, a Swi2/Snf2-related ATPase, and 12 additional subunits.

As described in the main manuscript, we completed our analysis by testing the conservation level of the genes inducing those phenotypes and over-represented in some, thus critical, multi-protein complexes. We tested whether in a protein complex, genes inducing the observed phenotype are more conserved than the other genes of the complex. Figure S2 shows that it might be the case in some of the environmental conditions.

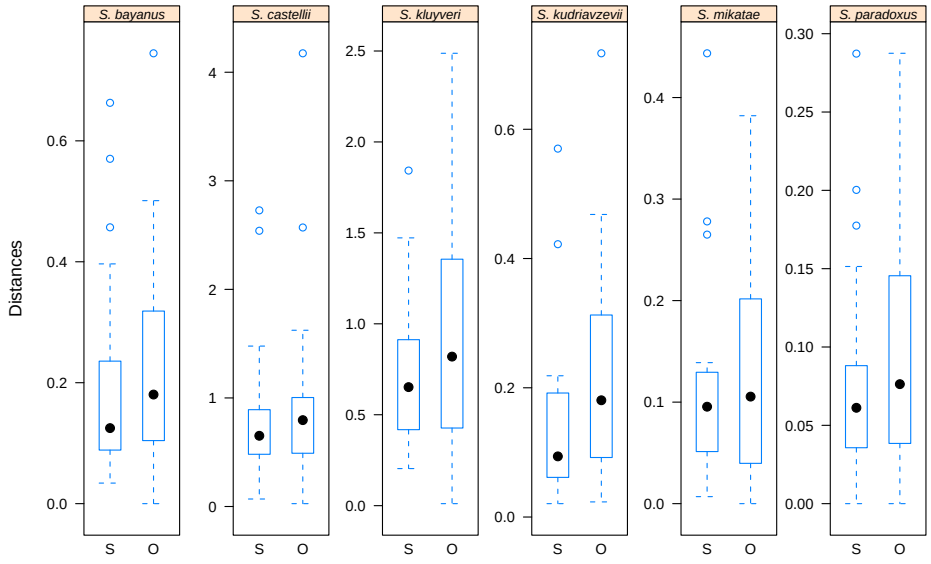

Figure S2: Non-essential genes inducing fitness growth defect phenotype under the stress conditions studied by Giaever *et al.* (2002) and over-represented in some, thus critical, complexes seems well conserved across species. Each panel presents a comparison between *S. cerevisiae* and one other species (named in the panel strip). In each panel, each boxplot shows the distribution of the gene evolution distances between the 2 species, calculated using the RSD approach [14]. The 'S' boxplot represents the distribution of distances for the genes inducing a lethal or growth defect phenotype (Sensitive) and the 'O' represents the other set of genes. In many cases the median distance and the spread (IQR) are smaller for the sensitive genes.

Dudley *et. al* (2005)

Dudley *et al.* [2] created a collection of gene-deletion mutants to determine genes that contribute to a particular phenotype in 21 different environmental conditions. Table S6 shows the results of the graph theory approach and the Hypergeometric test applied to each of the condition. The first two columns indicates the number of genes that were identified as sensitive by Dudley *et al.* [2] and how many of those genes are actually present in our interactome. The p-value column correspond to the result of the graph theory approach. A p-value ≤ 0.01 in a row shows that in that condition the sensitive genes are not randomly distributed among multi-protein complexes. Finally, the number of complexes involved in the phenotypic changes are listed in the last column.

	Dudley et al (2005)	Interactome	p.value	Complexes
CaCl2	180	73	0	14
CAD	83	36	0	15
cyclohex	164	62	0	20
FeLim	35	15	0	3
HU	87	46	0	11
MPA	11	5	0	5
Paraq	36	21	0	7
YPGal	41	15	0	7
YPGly	206	53	0	16
YPRaff	31	13	0	4
YPLac	159	33	0.001	8
UV	32	19	0.006	5
HygroB	264	85	0.007	16
lowPO4	34	6	0.015	-
pH3	16	4	0.016	-
rap	119	39	0.016	-
EtOH	75	43	0.018	-
NaCl	57	23	0.133	-
Caff	208	90	0.22	-
benomyl	34	15	0.465	-
Sorb	8	-	-	-

Table S5: Dudley et al. (2005) environmental stress conditions. Each row corresponds an environmental stress condition. The first column indicates the number of mutants with growth defect in Dudley's experiment. The second column indicates the number of those deleted genes in the interactome. The third column presents the p-value obtained by the graph theory test. A p-value <= 0.01 indicates that those deleted genes are not randomly distributed in the multi-protein complexes of the interactome. The fourth column indicates the number of multi-protein complexes involved. The 22 environmental conditions listed are: benomyl: 15ug/ml benomyl,microtubule function; CaCl2: 0.7M calcium chloride, divalent cation; CAD: 55uM Cadmium, heavy metal; Caff: 2mg/ml Caffeine; cyclohex: 0.18ug/ml cycloheximide, protein synthesis; DTT: unknown; EtOH YPD + 6% Ethanol; FeLim: irion limited,nutrient limited condition; HU: 11.4mg/ml Hudroxyurea, DNA replication and repair; HygroB: 50ug/ml hygromycin B, aminoglycosides; lowPO4: low phosphate, nutrient limited condition; MPA: 20ug/ml mycophenolic acid, transcriptional elongation; NaCl: 1.2M sodium chloride, general stress condition; Paraq: 1mM paraquat, oxidative stress; pH3: low pH, general stress condition; rap: 0.1ug/ml rapamycin, protein synthesis; Sorb: 1.2M sorbitol, general stress condition; UV: 100J/m2 ultra-violet, DNA replication and repair; YPGal 2% galactose, carbon source; YPGly 3% glycerol, carbon source; YPLac 2% lactate, carbon source; YPRaff 2% raffinose, carbon source.

```
---------Condition: CaCl2 -----------
MIPS-220  H+-transporting ATPase, vacuolar
GO:0000815  ESCRT III complex
GO:0000814  ESCRT II complex
MIPS-260.70  Vps4p ATPase complex (Vps protein complex)
GO:0016593  Cdc73/Paf1 complex
GO:0000221  vacuolar proton-transporting V-type ATPase, V1 domain
GO:0000938  GARP complex
GO:0008023  transcription elongation factor complex
GO:0030897  HOPS complex
MIPS-90.20  Vacuolar assembly complex
MIPS-90.30  ER assembly complex
GO:0000220  vacuolar proton-transporting V-type ATPase, V0 domain
MIPS-510.40  RNA polymerase II holoenzyme
MIPS-510.190.50  SWI/SNF transcription activator complex
---------Condition: CAD -----------
GO:0000815  ESCRT III complex
MIPS-260.70  Vps4p ATPase complex (Vps protein complex)
GO:0000938  GARP complex
GO:0030904  retromer complex
GO:0030897  HOPS complex
MIPS-230.20.10  ADA complex
MIPS-90.20  Vacuolar assembly complex
MIPS-260.30.30.10  Vps35/Vps29/Vps26 complex
EBI-1250344
GO:0005838  proteasome regulatory particle (sensu Eukaryota)
GO:0005671  Ada2/Gcn5/Ada3 transcription activator complex
GO:0046695  SLIK (SAGA-like) complex
MIPS-230.20.20  SAGA complex
GO:0033263  CORVET complex
EBI-1251060
---------Condition: cyclohex -----------
MIPS-230.20.10  ADA complex
GO:0000508  Rpd3L complex
GO:0000119  mediator complex
GO:0016593  Cdc73/Paf1 complex
EBI-1250344
GO:0046695  SLIK (SAGA-like) complex
MIPS-230.20.20  SAGA complex
GO:0005671  Ada2/Gcn5/Ada3 transcription activator complex
EBI-1251060
MIPS-510.40.20  Kornberg's mediator (SRB) complex
MIPS-90.30  ER assembly complex
MIPS-370  Protein N-acetyltransferase
MIPS-510.40  RNA polymerase II holoenzyme
GO:0000220  vacuolar proton-transporting V-type ATPase, V0 domain
MIPS-220  H+-transporting ATPase, vacuolar
GO:0031415  NatA complex
MIPS-510.190.50  SWI/SNF transcription activator complex
GO:0000445  THO complex part of transcription export complex
GO:0005838  proteasome regulatory particle (sensu Eukaryota)
```

GO:0016514 SWI/SNF complex
---------Condition: FeLim -----------
MIPS-220 H+-transporting ATPase, vacuolar
GO:0000220 vacuolar proton-transporting V-type ATPase, V0 domain
GO:0000221 vacuolar proton-transporting V-type ATPase, V1 domain
---------Condition: HU -----------
MIPS-510.40 RNA polymerase II holoenzyme
GO:0000119 mediator complex
MIPS-510.40.20 Kornberg's mediator (SRB) complex
GO:0016593 Cdc73/Paf1 complex
GO:0033062 Rhp55-Rhp57 complex
MIPS-510.190.50 SWI/SNF transcription activator complex
MIPS-260.70 Vps4p ATPase complex (Vps protein complex)
GO:0016514 SWI/SNF complex
GO:0000445 THO complex part of transcription export complex
GO:0000815 ESCRT III complex
GO:0031207 Sec62/Sec63 complex
---------Condition: MPA -----------
MIPS-230.20.20 SAGA complex
EBI-1251060
GO:0046695 SLIK (SAGA-like) complex
MIPS-370 Protein N-acetyltransferase
GO:0031415 NatA complex
---------Condition: Paraq -----------
MIPS-220 H+-transporting ATPase, vacuolar
GO:0000220 vacuolar proton-transporting V-type ATPase, V0 domain
GO:0000814 ESCRT II complex
GO:0000221 vacuolar proton-transporting V-type ATPase, V1 domain
MIPS-90.30 ER assembly complex
GO:0000813 ESCRT I complex
GO:0000815 ESCRT III complex
---------Condition: YPGal -----------
MIPS-220 H+-transporting ATPase, vacuolar
MIPS-510.190.80 GAL80 complex
GO:0000119 mediator complex
MIPS-510.40.20 Kornberg's mediator (SRB) complex
GO:0000220 vacuolar proton-transporting V-type ATPase, V0 domain
GO:0000221 vacuolar proton-transporting V-type ATPase, V1 domain
MIPS-510.40 RNA polymerase II holoenzyme
---------Condition: YPGly -----------
MIPS-220 H+-transporting ATPase, vacuolar
GO:0009353 mitochondrial oxoglutarate dehydrogenase complex
GO:0016602 CCAAT-binding factor complex
EBI-1225194
MIPS-420.50 F0/F1 ATP synthase (complex V)
MIPS-230.20.10 ADA complex
GO:0000220 vacuolar proton-transporting V-type ATPase, V0 domain
GO:0005754 mitochondrial proton-transporting ATP synthase, catalytic core
MIPS-90.30 ER assembly complex
MIPS-440.40.10 mitochondrial 3'-to-5' exoribonuclease (mtEXO)
GO:0000221 vacuolar proton-transporting V-type ATPase, V1 domain

```
MIPS-420.40  Cytochrome c oxidase (complex IV)
GO:0005751  mitochondrial respiratory chain complex IV
EBI-1250344
GO:0005671  Ada2/Gcn5/Ada3 transcription activator complex
GO:0005967  mitochondrial pyruvate dehydrogenase complex
---------Condition: YPRaff -----------
MIPS-220  H+-transporting ATPase, vacuolar
GO:0000220  vacuolar proton-transporting V-type ATPase, V0 domain
MIPS-90.30  ER assembly complex
GO:0000221  vacuolar proton-transporting V-type ATPase, V1 domain
---------Condition: YPLac -----------
GO:0009353  mitochondrial oxoglutarate dehydrogenase complex
EBI-1225194
GO:0016602  CCAAT-binding factor complex
MIPS-220  H+-transporting ATPase, vacuolar
GO:0000220  vacuolar proton-transporting V-type ATPase, V0 domain
MIPS-420.50  F0/F1 ATP synthase (complex V)
MIPS-440.40.10  mitochondrial 3'-to-5' exoribonuclease (mtEXO)
EBI-1225074
---------Condition: UV -----------
MIPS-510.180.10  Nucleotide excision repairosome
GO:0000110  nucleotide-excision repair factor 1 complex
GO:0000445  THO complex part of transcription export complex
GO:0000108  repairosome
GO:0000502  proteasome complex (sensu Eukaryota)
---------Condition: HygroB -----------
MIPS-260.20  Clathrin-associated protein (AP) complex
GO:0000815  ESCRT III complex
GO:0000814  ESCRT II complex
MIPS-260.70  Vps4p ATPase complex (Vps protein complex)
GO:0017119  Golgi transport complex
GO:0000938  GARP complex
GO:0031902  late endosome membrane
GO:0000136  alpha-1,6-mannosyltransferase complex
GO:0030897  HOPS complex
GO:0031416  NatB complex
MIPS-90.20  Vacuolar assembly complex
MIPS-90.30  ER assembly complex
GO:0000119  mediator complex
GO:0016593  Cdc73/Paf1 complex
GO:0031417  NatC complex
GO:0043529  GET complex
```

We note that some of the multi-protein complexes involved in phenotypic changes induced by the anti-fungal drug Paraquat are similar to the one found for the anti-fungal drug Nystatin tested by Giaever *et al.* (2002).

Similar to the previous analysis, Figure S3 presents the results of the evolution analysis. According to those results, the genes inducing the phenotypes and over-represented in the critical complexes are not evolutionary different to the other genes of the complexes.

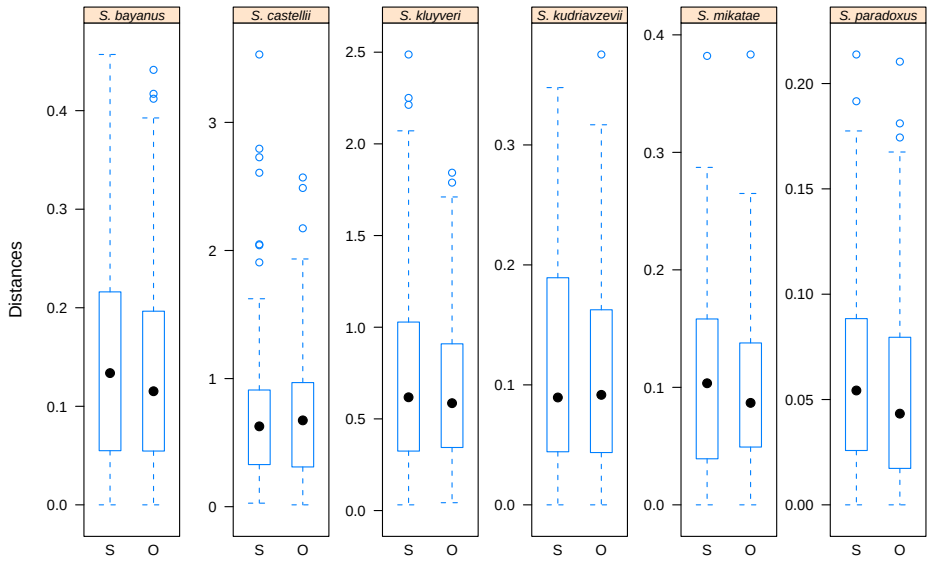

Figure S3: Non-essential genes inducing fitness growth defect phenotype under the stress conditions studied by Dudley *et al.* (2005) and over-represented in some, thus critical, complexes do not seem especially more conserved across species. Each panel presents a comparison between *S. cerevisiae* and one other species (named in the panel strip). In each panel, each boxplot shows the distribution of the gene evolution distances between the 2 species, calculated using the RSD approach [14]. The 'S' boxplot represents the distribution of distances for the genes inducing a fitness growth defect phenotype (Sensitive) and the 'O' represents the other set of genes.

Predicted multi-protein complexes

One way to assess the quality of the predicted multi-protein complexes is to search whether those complexes are involved in multiple phenotypic changes as we have shown that in the case of essentiality and haploinsufficiency several known complexes contribute to both phenotypes. Table S6 shows that indeed 17 predicted multi-protein complexes can be attributed to at least 2 phenotypes.

	cyclohex	FeLim	Paraq	YPRaff	HU	YPGal	MPA	YPGly	CaCl2	pH8g15	pH8g5	Ess	HI
PC1	X						X						
PC2	X				X								
PC3	X			X	X	X		X	X				
PC4	X								X		X		
PC5	X						X						
PC6		X	X										
PC7										X	X		
PC8												X	X
PC9												X	X
PC10												X	X
PC11												X	X
PC12												X	X
PC13												X	X
PC14												X	X
PC15												X	X
PC16												X	X
PC17												X	X

Table S6: Predicted multi-protein complexes contributing to more than one phenotype. Rows are predicted multi-protein complexes and columns indicates the enviromnental conditions. The 'X' marks when a particular multi-protein complexes is involved. The first 9 conditions were studied by Dudley et al. (2005) (cyclohex: 0.18ug/ml cycloheximide, protein synthesis; FeLim: irion limited,nutrient limited condition; Paraq: 1mM paraquat, oxidative stress; YPRaff: 2HU: 11.4mg/ml Hudroxyurea, DNA replication and repair; YPGal: 2MPA: 20ug/ml mycophenolic acid, transcriptional elongation; YPGly 3CaCl2: 0.7M calcium chloride, divalent cation). The next two conditions were tested by Giaever et al. (2002) (pH8: pH 8 (alkali stress) after 5 and 15 generations; ESS: essentiality). The last column is an experiment by Deutschbauer et al. (2005), HI: haploinsufficiency.

Complex ID	Gene	Commun Name	Current Gene Annotation
PC1 - apCompGavin2002: 5	YDL040C	NAT1	Subunit of the N-terminal acetyltransfer...
	YHR013C	ARD1	Subunit of the N-terminal acetyltransfer...
	YMR116C	ASC1	G-beta protein for Gpa2p; involved in tr...
PC2 - apCompKrogan2004: 18	YDR378C	LSM6	Lsm (Like Sm) protein; part of heterohep...
	YJL124C	LSM1	Lsm (Like Sm) protein; forms heterohepta...
	YGR054W		Eukaryotic initiation factor (eIF) 2A; a...
	YCR077C	PAT1	Topoisomerase II-associated deadenylatio...
	YGL173C	KEM1	Evolutionarily-conserved 5'-3' exonuclea...
PC3 - apCompHo2002: 31	YCR009C	RVS161	Amphiphysin-like lipid raft protein; sub...
	YDR388W	RVS167	Actin-associated protein, subunit of a c...
PC4 - apCompKrogan2004: 1	YBR279W	PAF1	RNAP II-associated protein; defines larg...
	YGL244W	RTF1	Subunit of the RNA polymerase II-associa...
	YLR418C	CDC73	Constituent of Paf1 complex with RNA pol...
	YOL145C	CTR9	Component of the Paf1p complex, which is...
	YOR123C	LEO1	Component of the Paf1 complex, which ass...

Continued on next page

Complex ID	Gene	Commun Name	Current Gene Annotation
	YGL207W	SPT16	Subunit of the heterodimeric FACT comple...
	YML069W	POB3	Subunit of the heterodimeric FACT comple...
PC5 - apCompGavin2002: 6	YGR090W	UTP22	Possible U3 snoRNP protein involved in m...
	YHR013C	ARD1	Subunit of the N-terminal acetyltransfer...
	YMR116C	ASC1	G-beta protein for Gpa2p; involved in tr...
PC6 - apCompGavin2002: 256	YEL051W	VMA8	Subunit D of the eight-subunit V1 periph...
	YGR020C	VMA7	Subunit F of the eight-subunit V1 periph...
	YKL080W	VMA5	Subunit C of the eight-subunit V1 periph...
	YOR332W	VMA4	Subunit E of the eight-subunit V1 periph...
	YPR036W	VMA13	Subunit H of the eight-subunit V1 periph...
	YKR001C	VPS1	Dynamin-like GTPase required for vacuola...
PC7 - apCompGavin2002: 3	YBR288C	APM3	Mu3-like subunit of the clathrin associa...
	YGR261C	APL6	Beta3-like subunit of the yeast AP-3 com...
	YPL195W	APL5	Delta adaptin-like subunit of the clathr...
PC8 - apCompGavin2002: 50	YPR023C	EAF3	Esa1p-associated factor, nonessential co...
	YDL014W	NOP1	Nucleolar protein, component of the smal...
	YDR324C	UTP4	Nucleolar protein, component of the smal...
	YHR052W	CIC1	Essential protein that interacts with pr...
	YBR095C	RXT2	Subunit of the histone deacetylase Rpd3L...
	YGR134W	CAF130	Part of the evolutionarily-conserved CCR...
	YLR002C	NOC3	Protein that forms a nuclear complex wit...
	YNL030W	HHF2	One of two identical histone H4 proteins...
	YLR249W	YEF3	Translational elongation factor 3, stimu...
	YPL012W	RRP12	Protein required for export of the ribos...
	YPR016C	TIF6	Constituent of 66S pre-ribosomal particl...
	YKR081C	RPF2	Essential protein involved in the proces...
	YNL110C	NOP15	Constituent of 66S pre-ribosomal particl...
	YLR276C	DBP9	ATP-dependent RNA helicase of the DEAD-b...
	YGL171W	ROK1	ATP-dependent RNA helicase of the DEAD b...
	YFL002C	SPB4	Putative ATP-dependent RNA helicase, nuc...
	YER006W	NUG1	GTPase that associates with nuclear 60S ...
	YGL111W	NSA1	Constituent of 66S pre-ribosomal particl...
	YGR103W	NOP7	Nucleolar protein involved in rRNA proce...
	YHR197W	RIX1	Essential component of the Rix1 complex ...
	YLR106C	MDN1	Huge dynein-related AAA-type ATPase (mid...
	YNL182C	IPI3	Essential component of the Rix1 complex ...
	YNR053C	NOG2	Putative GTPase that associates with pre...
	YOR272W	YTM1	Constituent of 66S pre-ribosomal particl...
PC9 - apCompGavin2002: 12	YGR158C	MTR3	3'5' exoribonuclease, exosome subunit; n...
	YDL014W	NOP1	Nucleolar protein, component of the smal...
	YDR324C	UTP4	Nucleolar protein, component of the smal...
	YHR052W	CIC1	Essential protein that interacts with pr...
	YOR206W	NOC2	Protein that forms a nucleolar complex w...
	YNL030W	HHF2	One of two identical histone H4 proteins...
	YLR074C	BUD20	Protein involved in bud-site selection; ...

Continued on next page

Complex ID	Gene	Commun Name	Current Gene Annotation
PC9 - apCompGavin2002: 12	YPR016C	TIF6	Constituent of 66S pre-ribosomal particl...
	YER126C	NSA2	Protein constituent of 66S pre-ribosomal...
	YGR245C	SDA1	Highly conserved nuclear protein require...
	YKR081C	RPF2	Essential protein involved in the proces...
	YNL110C	NOP15	Constituent of 66S pre-ribosomal particl...
	YGL171W	ROK1	ATP-dependent RNA helicase of the DEAD b...
	YER006W	NUG1	GTPase that associates with nuclear 60S ...
	YGR103W	NOP7	Nucleolar protein involved in rRNA proce...
	YHR197W	RIX1	Essential component of the Rix1 complex ...
	YLR106C	MDN1	Huge dynein-related AAA-type ATPase (mid...
	YNL182C	IPI3	Essential component of the Rix1 complex ...
	YNR053C	NOG2	Putative GTPase that associates with pre...
PC10 - apCompKrogan2004: 13	YDR324C	UTP4	Nucleolar protein, component of the smal...
	YDR398W	UTP5	Nucleolar protein, component of the smal...
	YGR128C	UTP8	Nucleolar protein required for export of...
	YHR196W	UTP9	Nucleolar protein, component of the smal...
	YJL109C	UTP10	Nucleolar protein, component of the smal...
	YMR093W	UTP15	Nucleolar protein, component of the smal...
	YPL126W	NAN1	U3 snoRNP protein, component of the smal...
	YEL055C	POL5	DNA Polymerase phi; has sequence similar...
PC11 - apCompGavin2002: 14	YGR158C	MTR3	3'5' exoribonuclease, exosome subunit; n...
	YDL014W	NOP1	Nucleolar protein, component of the smal...
	YDR324C	UTP4	Nucleolar protein, component of the smal...
	YHR052W	CIC1	Essential protein that interacts with pr...
	YLR074C	BUD20	Protein involved in bud-site selection; ...
	YPR016C	TIF6	Constituent of 66S pre-ribosomal particl...
	YFR001W	LOC1	Nuclear protein involved in asymmetric l...
	YGL171W	ROK1	ATP-dependent RNA helicase of the DEAD b...
	YER006W	NUG1	GTPase that associates with nuclear 60S ...
	YGR103W	NOP7	Nucleolar protein involved in rRNA proce...
	YHR085W	IPI1	Essential component of the Rix1 complex ...
	YHR197W	RIX1	Essential component of the Rix1 complex ...
	YLR106C	MDN1	Huge dynein-related AAA-type ATPase (mid...
	YNL182C	IPI3	Essential component of the Rix1 complex ...
	YNR053C	NOG2	Putative GTPase that associates with pre...
PC12 - apCompGavin2002: 223	YBR154C	RPB5	RNA polymerase subunit ABC27, common to ...
	YOR224C	RPB8	RNA polymerase subunit ABC14.5, common t...
	YPR187W	RPO26	RNA polymerase subunit ABC23, common to ...
	YMR146C	TIF34	Subunit of the core complex of translati...
	YDR404C	RPB7	RNA polymerase II subunit B16; forms two...
	YGL070C	RPB9	RNA polymerase II subunit B12.6; contact...
	YIL021W	RPB3	RNA polymerase II third largest subunit ...
	YJL140W	RPB4	RNA polymerase II subunit B32; forms two...
	YOR151C	RPB2	RNA polymerase II second largest subunit...
	YDR045C	RPC11	RNA polymerase III subunit C11; mediates...

Continued on next page

Complex ID	Gene	Commun Name	Current Gene Annotation
	YGR005C	TFG2	TFIIF (Transcription Factor II) middle s...
	YDL166C	FAP7	Essential NTPase required for small ribo...
PC13 - apCompKrogan2004: 2	YCR052W	RSC6	Component of the RSC chromatin remodelin...
	YDR303C	RSC3	Component of the RSC chromatin remodelin...
	YFR037C	RSC8	Component of the RSC chromatin remodelin...
	YIL126W	STH1	ATPase component of the RSC chromatin re...
	YKR008W	RSC4	Component of the RSC chromatin remodelin...
	YLR033W	RSC58	Component of the RSC chromatin remodelin...
	YLR321C	SFH1	Component of the RSC chromatin remodelin...
	YLR357W	RSC2	Component of the RSC chromatin remodelin...
	YML127W	RSC9	Component of the RSC chromatin remodelin...
	YMR033W	ARP9	Component of both the SWI/SNF and RSC ch...
	YMR091C	NPL6	Component of the RSC chromatin remodelin...
	YPR034W	ARP7	Component of both the SWI/SNF and RSC ch...
PC14 - apCompGavin2002: 181	YOL148C	SPT20	Subunit of the SAGA transcriptional regu...
	YHR069C	RRP4	3'-5' exoribonuclease involved in rRNA p...
	YPR023C	EAF3	Esa1p-associated factor, nonessential co...
	YOL051W	GAL11	Subunit of the RNA polymerase II mediato...
	YDL014W	NOP1	Nucleolar protein, component of the smal...
	YDR324C	UTP4	Nucleolar protein, component of the smal...
	YHL030W	ECM29	Major component of the proteasome; tethe...
	YHR052W	CIC1	Essential protein that interacts with pr...
	YGR134W	CAF130	Part of the evolutionarily-conserved CCR...
	YOR206W	NOC2	Protein that forms a nucleolar complex w...
	YLR002C	NOC3	Protein that forms a nuclear complex wit...
	YNL030W	HHF2	One of two identical histone H4 proteins...
	YLR249W	YEF3	Translational elongation factor 3, stimu...
	YDL160C	DHH1	Cytoplasmic DExD/H-box helicase, stimula...
	YNL061W	NOP2	Probable RNA m(5)C methyltransferase, es...
	YKR081C	RPF2	Essential protein involved in the proces...
	YLR276C	DBP9	ATP-dependent RNA helicase of the DEAD-b...
	YPL093W	NOG1	Putative GTPase that associates with fre...
	YGL171W	ROK1	ATP-dependent RNA helicase of the DEAD b...
	YFL002C	SPB4	Putative ATP-dependent RNA helicase, nuc...
	YER006W	NUG1	GTPase that associates with nuclear 60S ...
	YER139C		Putative protein of unknown function; YE...
	YGL111W	NSA1	Constituent of 66S pre-ribosomal particl...
	YGR103W	NOP7	Nucleolar protein involved in rRNA proce...
	YHR197W	RIX1	Essential component of the Rix1 complex ...
	YNR053C	NOG2	Putative GTPase that associates with pre...
	YOR272W	YTM1	Constituent of 66S pre-ribosomal particl...
PC15 - apCompGavin2002: 42	YHR052W	CIC1	Essential protein that interacts with pr...
	YGR134W	CAF130	Part of the evolutionarily-conserved CCR...
	YDR060W	MAK21	Constituent of 66S pre-ribosomal particl...
	YOR206W	NOC2	Protein that forms a nucleolar complex w...

Continued on next page

Complex ID	Gene	Commun Name	Current Gene Annotation
PC15 - apCompGavin2002: 42	YNL030W	HHF2	One of two identical histone H4 proteins...
	YDL160C	DHH1	Cytoplasmic DExD/H-box helicase, stimula...
	YPR016C	TIF6	Constituent of 66S pre-ribosomal particl...
	YMR049C	ERB1	Protein required for maturation of the 2...
	YNL061W	NOP2	Probable RNA m(5)C methyltransferase, es...
	YHR066W	SSF1	Constituent of 66S pre-ribosomal particl...
	YKR081C	RPF2	Essential protein involved in the proces...
	YNL110C	NOP15	Constituent of 66S pre-ribosomal particl...
	YPL093W	NOG1	Putative GTPase that associates with fre...
	YGL171W	ROK1	ATP-dependent RNA helicase of the DEAD b...
	YGL111W	NSA1	Constituent of 66S pre-ribosomal particl...
	YGR103W	NOP7	Nucleolar protein involved in rRNA proce...
	YOR272W	YTM1	Constituent of 66S pre-ribosomal particl...
PC16 - apCompGavin2002: 175	YCR035C	RRP43	Protein involved in rRNA processing; com...
	YGR095C	RRP46	Protein involved in rRNA processing; com...
	YGR195W	SKI6	3'-to-5' phosphorolytic exoribonuclease ...
	YPR137W	RRP9	Protein involved in pre-rRNA processing,...
	YOR326W	MYO2	One of two type V myosin motors (along w...
PC17 - apCompKrogan2004: 52	YCR057C	PWP2	Conserved 90S pre-ribosomal component es...
	YDR449C	UTP6	Nucleolar protein, component of the smal...
	YJL069C	UTP18	Possible U3 snoRNP protein involved in m...
	YLR222C	UTP13	Nucleolar protein, component of the smal...
	YLR409C	UTP21	Possible U3 snoRNP protein involved in m...

Table S7: Predicted critical multi-protein complexes and associated genes related to more than one phenotype. Gene annotation was truncated for formatting purposes. Full annotation can be retrieve using the SGD database.

Critical multi-protein complex stability and robustness

Are sensitive genes more conserved?

We have seen in the various analysis about gene conservation that only the dataset by Giaever *et al.* (2002) and by Kastenmayer *et al.* (2006) seem different. Gene inducing a phenotype and over-represented in some multi-protein complexes seem extremely well conserved across species. To assess the results of our graphical approach, we perform a two samples t-test for every comparisons. Our null hypothesis was that there is no difference between the mean evolutionary distances of the sensitive and non-sensitive genes. Table S8 presents the results. At p-value ≤ 0.05, this analysis shows that only the essential genes [1] and the sORFs [3] are different and well conserved across species. The null hypothesis is true for the other experiments.

	S.bayanus	S.castellii	S.kluyveri	S.kudriavzevii	S.mikatae	S.paradoxus
essential	0.00	0.01	0.08	0.03	0.05	0.01
sORF	0.00	0.00	0.01	0.00	0.00	0.00
nystatin15	0.52	0.30	0.26	0.29	0.78	0.33
pH8g15	0.43	0.08	0.79	0.58	0.25	0.67
pH8g5	0.81	0.70	0.93	0.73	0.82	0.60
ypg15	0.33	0.53	0.33	0.30	0.87	0.22
CaCl2	0.60	0.45	0.40	0.41	0.20	0.46
cyclohex	0.34	0.80	0.82	0.53	0.40	0.23
FeLim	0.28	0.52	0.35	0.27	0.21	0.14
MPA	0.33	0.61	0.56	0.34	0.53	0.19
Paraq	0.29	0.49	0.21	0.51	0.21	0.11
YPGal	0.53	0.88	0.49	0.43	0.19	0.32
YPRaff	0.38	0.55	0.44	0.39	0.25	0.16
HU	0.48	0.76	0.87	0.45	0.61	0.56
YPGly	0.24	0.65	0.42	0.11	0.12	0.11
CAD	0.41	0.58	0.44	0.55	0.51	0.22
YPLac	0.25	0.70	0.42	0.20	0.20	0.11
UV	0.96	0.79	0.92	0.83	0.73	0.76
pH3	0.61	0.72	0.58	0.52	0.55	0.31

Table S8: Evolutionary distances not necessarily correlate with genes inducing a phenotype and their co-membership in a protein complex. P-value of two sample t-tests between evolutionary distances of genes inducing a phenotype and co-member of a complex and the other member of genes of the complex. Each column represent the comparison between S. cerevisiae and the specified species. Each row is a different environmental condition. The first 2 rows are the lethal phenotypes studied by Kastenmayer et al. (2006) (sORF: small Open Reading Frames) and Giaever et al. (2002)(ESS: essentiality). The next 3 conditions were also tested by Giaever et al. (2002) (nystatin: Nystatin (antifungal drug) after 5 and 15 generations; pH8: pH 8 (alkali stress) after 5 and 15 generations; ypg15: yeast/peptone/galactose 15 gen.). The remaining rows report the experimental conditions studied by Dudley et al. (2005) (CaCl2: 0.7M calcium chloride, divalent cation; cyclohex: 0.18ug/ml cycloheximide, protein synthesis; FeLim: iron limited,nutrient limited condition; MPA: 20ug/ml mycophenolic acid, transcriptional elongation; Paraq: 1mM paraquat, oxidative stress; YPGal: 2YPRaff: 2HU: 11.4mg/ml Hudroxyurea, DNA replication and repair; YPGly: 3CAD: 55uM Cadmium, heavy metal; YPLac: 2UV: 100J/m2 ultra-violet, DNA replication and repair; pH3: low pH, general stress condition.)

Critical multi-protein complexes robustness

We proposed to make use of a synthetic lethal approach [15, 16] to further investigate the concept of essentiality. To that aims, we selected 14 critical complexes that have not been much studied and for which not all genes are essential (at least 2) (Table S9). We tested whether two or more deletions can effectively disrupt the functioning of the complex and hence refine the role of the complex. We found

that in those critical complexes

In vivo experiment

	gene1	gene2	score
apCompGavin2002: 14	ARX1	BUD20	1
	ARX1	LHP1	1
	ARX1	MRT4	0
	BUD20	LHP1	1
	BUD20	MRT4	1
	LHP1	MRT4	1
apCompGavin2002: 49	LOC1	PUF6	E
	MRT4	PUF6	E
	LOC1	MRT4	1
apCompGavin2002: 181	FPR4	PUF6	0
	PUF6	YER139C	1
	ARX1	FPR4	0
	ARX1	YER139C	1
	FPR4	YER139C	0
apCompGavin2002: 22	PUF6	TIF4632	1
	GAR1	PUF6	1
	GAR1	TIF4632	0
apCompKrogan2004: 82	RPL8B	RPS8A	0
apCompGavin2002: 182	NOP12	PUF6	1
apCompGavin2002: 41	PUF6	SSF1	1
apCompGavin2002: 158	NOP12	SSF1	0
apCompGavin2002: 75	MRPL10	SEC28	1
apCompHo2002: 73	PRB1	SEC28	1
apCompGavin2002: 50	ARX1	PUF6	1
apCompHo2002: 87	MSS116	NOP6	0
apCompGavin2002: 157	BRR1	STO1	1
apCompHo2002: Ho174	CKA1	CKB2	0

Table S9: The 27 pairs of non-essential genes tested for synthetic lethality interaction in not well studied but probably essential protein complexes. The score column records the outcome of the experiment: 0 could not create diploid, 1 no growth defect, E experimental error.

KEGG pathways and phenotype

As described in the article, we computed the two omnibus tests to evaluate whether there is an over-abundance of KEGG pathways with low or high proportions of genes associated with essentiality [1] and haploinsuffiency [17]. The smoothed density estimates are shown in Figure S4 for essential genes, Panel (a), and for haploinsufficient genes, Panel (b). These figures suggest that there are many more pathways with zero genes in them than should be observed under the null hypothesis, suggesting that the null hypothesis is not tenable for either the haploinsufficient genes or the essential genes.

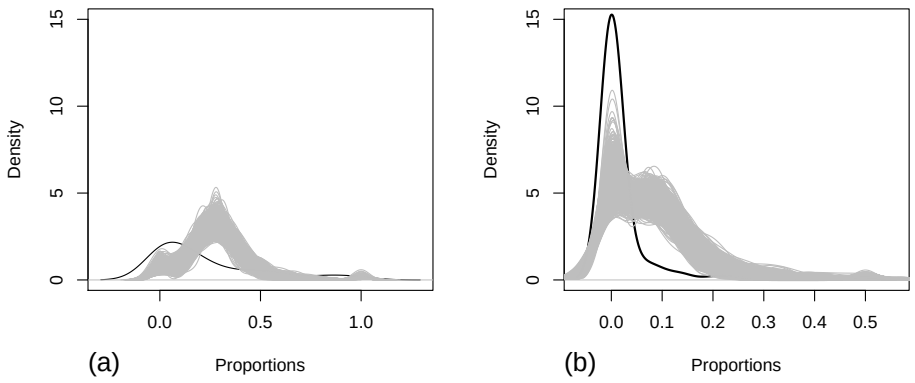

Figure S4: Essential and haploinsufficient genes not well represented in KEGG pathways. Smoothed histograms of the proportion of genes per KEGG pathway that are associated to a phenotype. The dark line represents the observed data and the light curves represent the permuted data. Only the first 50 simulated density estimates out of 1000 permutations are displayed for visualization efficiency. Panel (a) corresponds to the essential genes and Panel (b) to the haploinsufficient genes.

Using the graph theory approach [18], we observed the permutation results shown in Figure S5. Panel (a) presents the results for essentiality and Panel (b) presents the results for haploinsufficiency. In both cases the observed number of edges in the intersection is far larger than any value from the permutations and hence the permutational p-value is less that 1 in 1,000 providing some evidence against the null hypothesis and indicating that there is an association between the genes associated with a particular phenotype and the KEGG pathways used in our analysis.

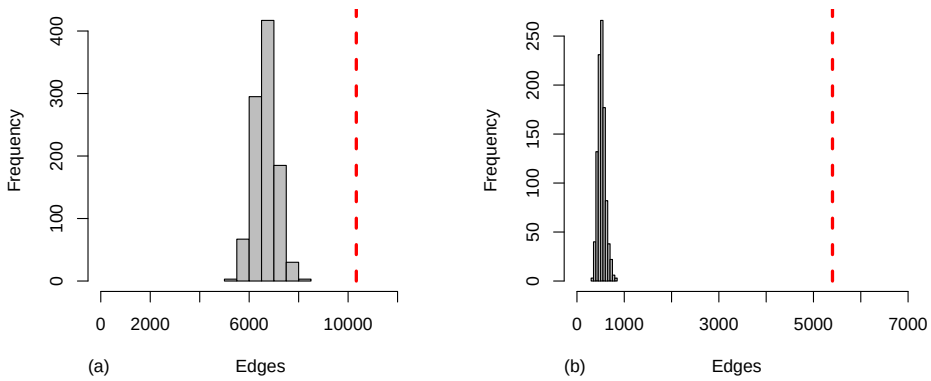

Figure S5: Associations exist between KEGG pathways and the genes associated to essentiality and haploinsufficiency. The distribution of the number of edges, under the null distribution of genes randomly distributed (1,000 permutations), grey histogram, in KEGG pathways compared to the number of observed edges, dashed red line. Panel (a) shows the results for the essential genes; Panel (b) for the haploinsufficient genes.

Since the overall tests provided some evidence against the null hypothesis the second part of the analysis was performed and Hypergeometric tests were used to identify specific pathways that had an over abundance of genes for each of the different phenotypes. Tables S10 and S11 present the KEGG pathways having an overabundance of essential genes and haploinsufficient genes (p-value ≤ 0.01), respectively.

	Observed	Expected	Size	Odds	P.value	Description
sce03050	28	8	32	20.01	1.02e−12	Proteasome
sce00563	20	5	21	55.90	6.53e−11	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis
sce03022	21	6	23	29.41	1.55e−10	Basal transcription factors
sce03020	23	7	29	10.76	6.33e−09	RNA polymerase
sce04111	56	30	109	3.14	2.61e−08	Cell cycle - yeast
sce01031	18	5	21	16.62	3.33e−08	Glycan structures - biosynthesis 2
sce00240	37	19	70	3.19	3.45e−06	Pyrimidine metabolism
sce00970	21	10	37	3.62	1.39e−04	Aminoacyl-tRNA biosynthesis
sce04120	17	8	30	3.57	6.40e−04	Ubiquitin mediated proteolysis
sce03060	7	2	9	9.38	2.41e−03	Protein export
sce00100	12	5	21	3.60	3.82e−03	Biosynthesis of steroids
sce00230	36	24	89	1.88	4.28e−03	Purine metabolism

Table S10: KEGG pathways associated with Essentiality. KEGG and essential genes. KEGG identification number; Observed: number of essential genes in the pathway; Expected: number of essential expected to be in the pathway; Odds: odds ratio; Size: total number of genes in the pathway; P-value: p-value of Hypergeometric test that was used to identify whether or not some pathways have a significant number of essential genes; Description: annotation given to the pathway.

	Observed	Expected	Size	Odds	P-value	Description
sce03010	73	11	148	48.02	1.33e−55	Ribosome
sce03020	7	2	29	3.98	5.28e−03	RNA polymerase

Table S11: KEGG pathways associated with Haploinsufficiency. KEGG in the interactome (all the genes). KEGG identification number; Observed: number of haploinsufficient genes in the pathway; Expected: number of haploinsufficient genes expected to be in the pathway; Odds: odds ratio; Size: total number of genes in the pathway; P-value: p-value of Hypergeometric test that was used to identify whether or not some pathways have a significant number of haploinsufficient genes; Description: annotation given to the pathway.

References

[1] Giaever G, et al.: **Functional profiling of the Saccharomyces cerevisiae genome.** *Nature* 2002, **418**(6896):387–391.

[2] Dudley A, Janse D, Tanay A, Shamir R, Church GM: **A global view of pleiotropy and phenotypically derived gene function in yeast**. *Molecular Systems Biology* 2005, **1**:E1–E11.

[3] Kastenmayer J, Ni L, Chu A, Kitchen L, Au W, Yang H, Carter C, Wheeler D, Davis R, Boeke J, Snyder M, Basrai M: **Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae.** *Genome Research* 2006, **16**(3):365–373.

[4] The Gene Ontology Consortium: **Gene Ontology: tool for the unification of biology**. *Nature Genetics* 2000, **25**:25–29.

[5] Camon E, Magrane M, Barrell D, et al.: **The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology.** *Nucleic Acids Res* 2004, **32**(Database issue):D262–D266.

[6] Guldener U, Munsterkotter M, Kastenmuller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak S, Garcia-Martinez J, Perez-Ortin J, et al.: **CYGD: the Comprehensive Yeast Genome Database**. *Nucleic Acids Research* 2005, **33**(Database Issue):D364.

[7] Gavin AC, Aloy P, Grandi P, et al.: **Proteome survey reveals modularity of the yeast cell machinery.** *Nature* 2006.

[8] Krogan N, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis A, et al.: **Global landscape of protein complexes in the yeast Saccharomyces cerevisiae**. *Nature* 2006, **440**:637–643.

[9] Gavin A, Boesche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick J, Michon A, Cruciat C, et al.: **Functional organization of the yeast proteome by systematic analysis of protein complexes**. *Nature* 2002, **415**(6868):141–147.

[10] Krogan N, Peng W, Cagney G, Robinson M, Haw R, Zhong G, Guo X, Zhang X, Canadien V, Richards D, et al.: **High-definition macromolecular composition of yeast RNA-processing complexes**. *Mol. Cell* 2004, **13**:225–239.

[11] Ho Y, Gruhler A, Heilbut A, Bader G, Moore L, Adams S, Millar A, Taylor P, Bennett K, Boutilier K, et al.: **Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry**. *Nature* 2002, **415**(6868):180–183.

[12] Scholtens D, Gentleman R: **Making Sense of High throughput Protein-Protein Interaction Data**. *Statistical Applications in Genetics and Molecular Biology* 2004, **3**:39.

[13] Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic Acids Research* 2000, **28**:27–30.

[14] Wall D, Fraser H, Hirsh A: **Detecting putative orthologs**. *Bioinformatics* 2003, **19**(13):1710–1711.

[15] Tong AH, Evangelista M, Parsons AB, et al.: **Systematic genetic analysis with ordered arrays of yeast deletion mutants.** *Science* 2001, **294**(5550):2364–2368.

[16] Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD: **A robust toolkit for functional profiling of the yeast genome.** *Mol Cell* 2004, **16**(3):487–496.

[17] Deutschbauer A, Jaramillo D, Proctor M, et al.: **Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in Yeast**. *Genetics* 2005, **169**:1915–1925.

[18] Balasubramanian R, LaFramboise T, Scholtens D, Gentleman R: **A graph-theoretic approach to testing associations between disparate sources of functional genomics data.** *Bioinformatics* 2004, **20**(18):3353–3362.