



Published in final edited form as:

Science. 2015 May 22; 348(6237): 921–925. doi:10.1126/science.aaa0769.

## Systematic humanization of yeast genes reveals conserved functions and genetic modularity

Aashiq H. Kachroo<sup>1</sup>, Jon M. Laurent<sup>1</sup>, Christopher M. Yellman<sup>1</sup>, Austin G. Meyer<sup>1,2</sup>, Claus O. Wilke<sup>1,2,3</sup>, and Edward M. Marcotte<sup>1,2,4,\*</sup>

<sup>1</sup>Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712

<sup>2</sup>Center for Computational Biology and Bioinformatics, University of Texas at Austin, Austin, TX 78712

<sup>3</sup>Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712

<sup>4</sup>Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

### Abstract

To determine whether genes retain ancestral functions over a billion years of evolution and to identify principles of deep evolutionary divergence, we replaced 414 essential yeast genes with their human orthologs, assaying for complementation of lethal growth defects upon loss of the yeast genes. Nearly half (47%) of the yeast genes could be successfully humanized. Sequence similarity and expression only partly predicted replaceability. Instead, replaceability depended strongly on gene modules: genes in the same process tended to be similarly replaceable (e.g., sterol biosynthesis) or not (e.g., DNA replication initiation). Simulations confirmed selection for specific function can maintain replaceability despite extensive sequence divergence. Critical ancestral functions of many essential genes are thus retained in a pathway-specific manner, robust to drift in sequences, splicing, and protein interfaces.

---

The ortholog-function conjecture posits that orthologous genes in diverged species perform similar or identical functions (1). The conjecture is supported by comparative analyses of gene-expression patterns, genetic interaction maps, and chemogenomic profiling (2-6), and it is widely used to predict gene function across species. However, even if two genes perform similar functions in different organisms, it may not be possible to replace one for the other, in particular if the organisms are widely diverged. To what extent deeply divergent orthologs can stand in for each other, and which principles govern such functional equivalence across species, is largely unknown.

Here, we systematically addressed these questions by replacing a large number of yeast genes with their human orthologs. Humans and the baker's yeast *Saccharomyces cerevisiae* diverged from a common ancestor approximately one billion years ago (7). They share several thousand orthologous genes, accounting for more than 1/3 of the yeast genome (8).

---

\*Correspondence to: marcotte@icmb.utexas.edu.

Yeast and human orthologs tend to be recognizable but often highly diverged; amino-acid identity ranges from 9% to 92%, with a genome-wide average of 32%. While we know of individual examples of human genes capable of replacing their fungal orthologs (9-12), the extent and specific conditions under which human genes can substitute for their yeast orthologs are generally not known.

We focused on the set of genes essential for yeast cell growth under standard laboratory conditions (13, 14) and for which the yeast-human orthology is 1:1, *i.e.* genes without lineage-specific duplicate genes that might mask the effects. Based on availability of full-length human cDNA recombinant clones (15, 16) and matched yeast strains with conditionally null alleles of the test genes (17-19), we selected 469 human genes to study (Fig. 1A).

We first sub-cloned and sequence-verified each human protein coding sequence into a single-copy, centromeric yeast plasmid under the transcriptional control of either an inducible (GAL) or constitutively active (GPD) promoter. We assembled a matched set of yeast strains in which each orthologous yeast gene could be conditionally down-regulated (*via* a tetracycline-repressible promoter (17)), inactivated (*via* a temperature sensitive allele (18)), or segregated away genetically (following sporulation of a heterozygous diploid deletion strain (13, 19)) (Fig. 1A; Fig. S1). After verifying that loss of the relevant yeast gene conferred a strong growth defect, we tested whether expression of the human ortholog could complement the growth defect, as illustrated for several examples in Fig. 1B (also Figs. S2-4). 73 of the human genes exhibited toxicity when expressed in the permissive condition; reducing the genes' expression levels allowed us to assay replacement in 66 cases (Table S1).

Overall, we performed 652 informative growth assays surveying 414 human/yeast orthologs (Figs. 1A, C). In total, 176 yeast genes (43%) could be replaced by their human orthologs in at least one of the three strain backgrounds, while 238 (57%) could not (Table S1). We collated previously published reports of yeast gene complementation by human genes; our assays recapitulated these cases with 90% precision, 72% recall (Table S1), and incorporating the literature data for subsequent analyses brought the observed complementation rate to 47% (Fig. 1C). For randomly selected subsets of strains, we additionally validated the assays by sub-cloning the yeast test genes into the assay vectors and confirming positive complementation assays (Table S2), by confirming human protein expression using Western blot analysis (Fig. S5), and confirming complementation by tetrad dissection (Table S1).

Given that roughly half of the tested human genes successfully replaced and half did not, we next investigated factors determining replaceability. We assembled 104 quantitative features of the genes or ortholog pairs, including calculated properties of the genes' sequences (e.g., gene and protein lengths, sequence similarities, codon usage, and predicted protein aggregation potential) and properties such as protein interactions, mRNA and protein abundances, transcription and translation rates, and mRNA splicing features (Table S3). We then quantified how well each feature predicted replaceability (Fig. 2A, Table S3).

Notably, sequence similarity only partly predicted replaceability. This tendency was strongest for highly similar (>50% amino acid identity) or dissimilar (<20%) ortholog pairs. However, most pairs fell into an intermediate range of 20-50% sequence identity, which only poorly predicted replaceability (Fig. 2B). Instead, replaceability was best predicted by properties of specific gene modules. In particular, proteins in the same pathway or complex tended to be similarly replaceable (Fig. 2A). Replaceable genes also tended to be shorter and more highly expressed. Using these features in a supervised Bayesian network classification algorithm (Fig. S6), we achieved a high overall cross-validated prediction rate (area under the receiver operating characteristic curve of 0.825, Fig. 2A) and correct prediction of 8 of 10 literature cases withheld from all computational analyses (Table S4). Properties such as human-gene splice forms counts, yeast 5' and 3' UTR lengths, codon adaptation indices, and yeast mRNA half-lives showed little relationship with replaceability (Fig. 2A, Table S3).

The strong association between replaceability and gene modules led us to investigate this phenomenon in more depth, examining replaceability as a function of specific protein complexes and pathways. Broad KEGG (20) pathway classes showed highly differential replaceability: metabolic enzymes (e.g., enzymes participating in lipid, amino-acid, and carbohydrate metabolism) tended to be replaceable, while proteins involved in DNA replication and repair or in cell growth tended not to be replaceable (Fig. 2C).

Among large protein complexes and pathways, we observed both extremes of replaceability. Some were entirely non-replaceable: for example, we did not observe a single successful replacement among 13 tested members of the TriC chaperone complex, the DNA replication initiation origin recognition complex, or its interacting MCM complex (Figs. 3A, B). In contrast, some pathways were almost entirely replaceable: among 19 components of the sterol biosynthesis pathway (which catalyzes the conversion of acetyl-CoA to cholesterol in humans and ergosterol in yeast) only the human farnesyl-diphosphate farnesyltransferase 1 enzyme (*FDFT1*) and farnesyl diphosphate synthase (*FDPS*) failed to replace their yeast orthologs. All other tested components were replaceable, suggesting that yeast and humans both retain the same essential complement of ancestral sterol biosynthesis functionality (Figs. 3C, S7).

The modular nature of replaceability was particularly evident in the case of the 26S proteasome complex. Of 28 tested subunits, 21 human genes replaced their yeast counterparts (Fig. 4A). However, the non-replaceable subunits were not randomly distributed; rather, they clustered in two physically-interacting groups—one consisting of the 19S lid components Rpn3 and Rpn12 and one consisting of the 20S inner core heptameric beta ring subunits  $\beta 1$ ,  $\beta 2$ ,  $\beta 5$ ,  $\beta 6$ , and  $\beta 7$ . Thus, of the two central heteroheptameric rings, all testable components of the alpha ring replaced, while most of the beta ring did not.

An examination of the alpha and beta subunit structures showed that subunit-subunit interfacial amino acids were conserved to similar degrees between yeast and human subunits (Fig. S8A), although beta subunits exhibited elevated rates of non-synonymous substitutions compared to alpha subunits (Fig. S8B). Even when interfacial amino acids were only partly conserved, modeling human alpha subunits into the known structure of the yeast proteasome

(21) revealed that human proteins could be sterically accommodated into the yeast intersubunit-interface, as shown for human  $\alpha 6$  (Fig. 4B) packing against yeast  $\beta 6$ , in spite of only sharing 50% identical amino acids at the interface (Fig. S8A). Only orthologous alpha subunits replaced; non-orthologs failed (Fig. S9).

We further confirmed this trend across alpha and beta proteasome subunits by cloning and assaying subunits from additional organisms, including another yeast (*Saccharomyces kluyverii*), the nematode *C. elegans*, and several beta subunits from the frog *X. laevis*. In all cases, alpha subunits complemented loss of the yeast orthologs, while beta subunits generally failed to complement (Fig. 4C). The pattern of replaceability across species suggests that alpha and beta subunits experienced different evolutionary pressures, in each case operating at the level of the system of genes (the alpha or beta heteroheptamer).

To determine further why proteasome alpha subunits were replaceable while beta subunits were not, we isolated human  $\beta 2$  subunit mutants that complemented the yeast defect (Figs. S10-12). A single serine to glycine substitution (S214G) was sufficient to rescue growth (Fig. S11).  $\beta 2$  subunits act as proteases, but yeast  $\beta 2$  catalytic activity is dispensable if the proteasome assembles with other functioning protease subunits (22). Notably, a catalytically dead (T44A) human  $\beta 2$  failed to complement, while an S214G, T44A double mutant complemented successfully (Fig. S11). We conclude the S214G mutant is competent to assemble an intact proteasome, although the subunit may not be catalytically active. Thus, native human  $\beta 2$  needs only one amino acid change to pack within the yeast proteasome.

Theory predicts that evolutionary divergence creates Dobzhansky-Muller incompatibilities, since novel mutations in one species are untested in the other species' genetic background and may be deleterious there (23, 24). To better understand how proteins retain the ability to interact with their ortholog's interaction partners, even when they have diverged substantially, we developed a biochemically realistic divergence model in which we simulated the evolution of two physically interacting proteins, which both diverge over time. We considered three distinct scenarios: (i) both thermodynamic stability and binding to the extant partner were selected at ancestral levels; (ii) binding was selected at ancestral levels but stability was not; (iii) stability was selected at ancestral levels but binding was not. Thermodynamic stability ( $G^{\text{folding}}$ ) and binding ( $G^{\text{interaction}}$ ) were calculated using the empirical FoldX energy function (25). Under all scenarios, we evaluated whether an evolved member of the pair could still bind to its ancestral partner, for which binding was not enforced. We found that ancestral binding decayed rapidly under scenario (iii) but much more slowly under the other two scenarios (Figs. 4D, S13-15). Natural selection for a protein interaction thus preserves the interaction interface in a manner consistent with binding to the ancestral partner (Figs. S16-17), even though many lineages will eventually accumulate mutations that cause incompatibilities with the ancestral interactor.

Our data demonstrate that a substantial portion of conserved yeast and human genes perform much the same roles in both organisms—to an extent that the protein-coding DNA of a human gene can actually substitute for that of the yeast. The strong pathway-specific pattern of individual replacements suggests that group-wise replacement of the genes should be feasible, raising the possibility of humanizing entire cellular processes in yeast. Such strains

would simplify drug discovery against human proteins, enable studies of the consequences of human genetic polymorphisms (as in (26) and Fig. S7), and empower functional studies of entire human cellular processes in a simplified organism.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

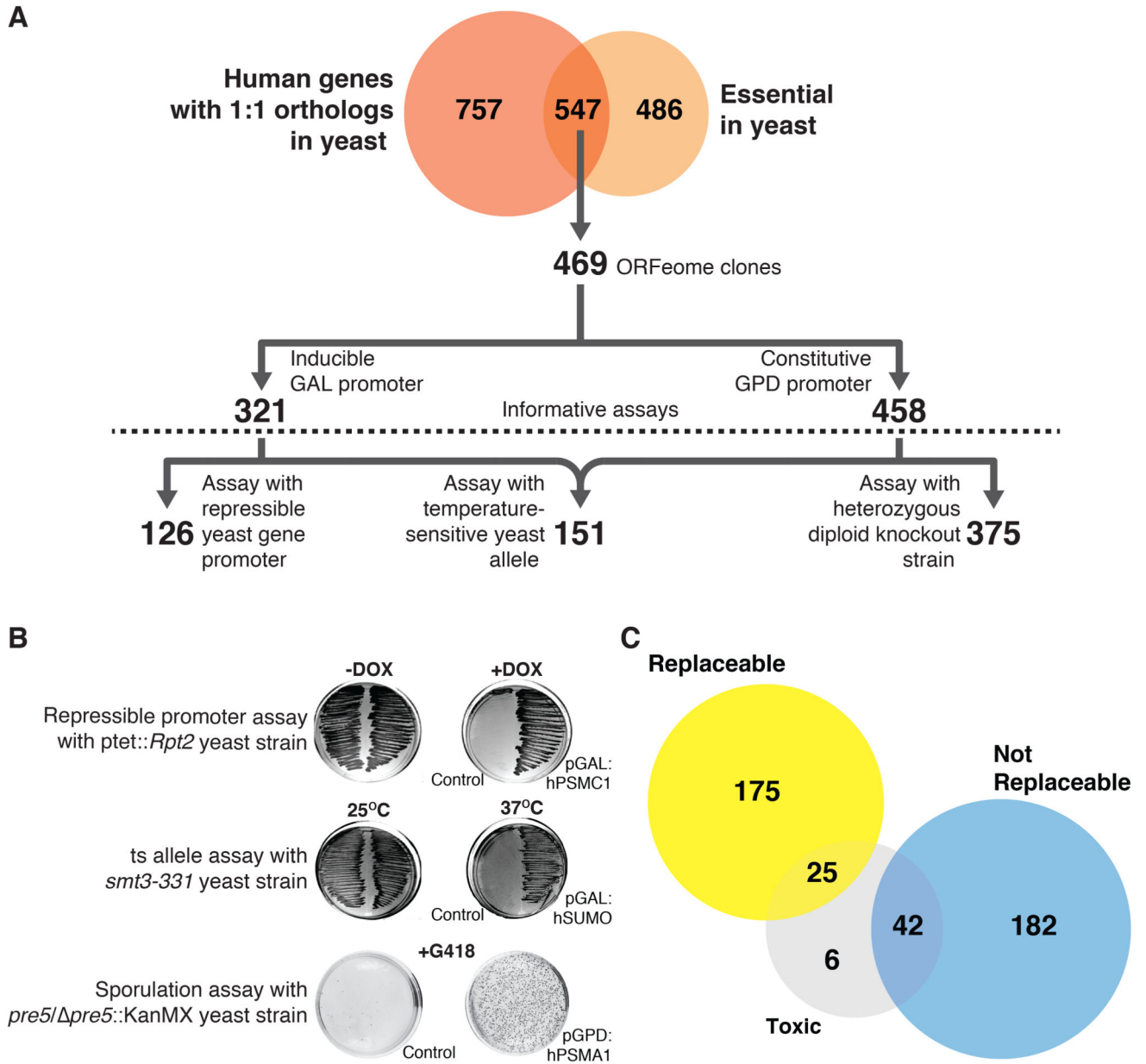
## Acknowledgments

We thank Megan Minnix and Ariel Royall for assistance with cloning and assays, Kevin Drew for structural modeling assistance, Mark Tsechansky for TANGO assistance, and Charlie Boone for providing the temperature sensitive yeast strain collection. This work was supported by CPRIT research fellowships to A.H.K. and J.M.L., NIH grant R01 GM088344, DTRA grant HDTRA1-12-C-0007, and NSF STC BEACON funds (DBI-0939454) to C.O.W., and grants from the NIH, NSF, CPRIT, and Welch foundation (F-1515) to E.M.M.

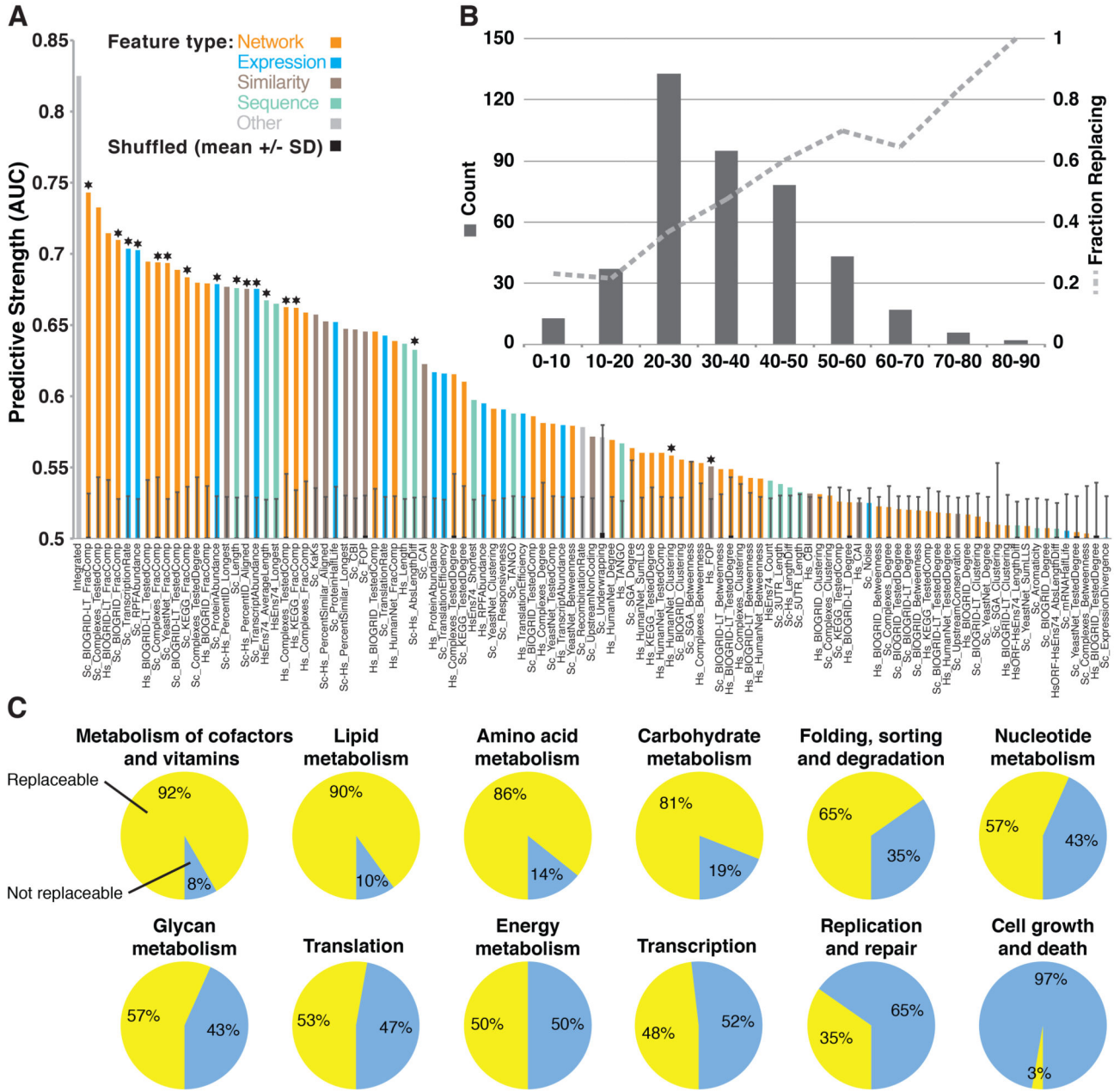
## References and Notes

- Gabalton T, Koonin EV. *Nat. Rev. Genet.* 2013; 14:360–366. [PubMed: 23552219]
- Nehrt NL, Clark WT, Radivojac P, Hahn MW. *PLoS Comput. Biol.* 2011; 7:e1002073. [PubMed: 21695233]
- Chen X, Zhang J. *PLoS Comput. Biol.* 2012; 8:e1002784. [PubMed: 23209392]
- Kapitzky L, et al. *Mol. Syst. Biol.* 2010; 6:451. [PubMed: 21179023]
- Ryan CJ, et al. *Mol. Cell.* 2012; 46:691–704. [PubMed: 22681890]
- Frost A, et al. *Cell.* 2012; 149:1339–1352. [PubMed: 22682253]
- Douzery EJP, Snell EA, Baptiste E, Delsuc F, Philippe H. *Proc Natl Acad Sci U A.* 2004; 101:15386–15391.
- O'Brien KP, Remm M, Sonnhammer ELL. *Nucleic Acids Res.* 2005; 33:D476–80. [PubMed: 15608241]
- Dolinski K, Botstein D. *Annu. Rev. Genet.* 2007; 41:465–507. [PubMed: 17678444]
- Basson ME, Thorsness M, Finer-Moore J, Stroud RM, Rine J. *Mol Cell Biol.* 1988; 8:37973808.
- Cormack BP, Strubin M, Stargell LA, Struhl K. *Genes Dev.* 1994; 8:1335–1343. [PubMed: 7926734]
- Osborn MJ, Miller JR. *Brief. Funct. Genomic. Proteomic.* 2007; 6:104–111. [PubMed: 17698857]
- Tong AH, et al. *Science.* 2001; 294:2364–2368. [PubMed: 11743205]
- Winzeler EA, et al. *Science.* 1999; 285:901–906. [PubMed: 10436161]
- MGC Project Team. et al. *Genome Res.* 2009; 19:2324–2333. [PubMed: 19767417]
- Rual J-F, et al. *Nature.* 2005; 437:1173–1178. [PubMed: 16189514]
- Mnaimneh S, et al. *Cell.* 2004; 118:31–44. [PubMed: 15242642]
- Li Z, et al. *Nat. Biotechnol.* 2011; 29:361–367. [PubMed: 21441928]
- Pan X, et al. *Mol Cell.* 2004; 16:487–496. [PubMed: 15525520]
- Kanehisa M, Goto S. *Nucleic Acids Res.* 2000; 28:27–30. [PubMed: 10592173]
- Groll M, et al. *Nature.* 1997; 386:463–471. [PubMed: 9087403]
- Groll M, et al. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:10976–10983. [PubMed: 10500111]
- Orr HA. *Genetics.* 1995; 139:1805–1813. [PubMed: 7789779]
- Welch JJ. *Evol. Int. J. Org. Evol.* 2004; 58:1145–1156.
- Guerois R, Nielsen JE, Serrano L. *J. Mol. Biol.* 2002; 320:369–387. [PubMed: 12079393]
- Marini NJ, Thomas PD, Rine J. *PLoS Genet.* 2010; 6:e1000968. [PubMed: 20523748]
- Houten SM, Waterham HR. *Mol. Genet. Metab.* 2001; 72:273–276. [PubMed: 11243736]
- Alberti S, Gitler AD, Lindquist S. *Yeast.* 2007; 24:913–919. [PubMed: 17583893]
- Liang X, Peng L, Baek C-H, Katzen F. *BioTechniques.* 2013; 55:265–268. [PubMed: 24215642]

30. Burke, D.; Cold Spring Harbor Laboratory. *Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual*. 2000. Cold Spring Harbor Laboratory Press; Plainview, N.Y.: 2000.
31. Kaiser, C.; Cold Spring Harbor Laboratory. *Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual*. 1994. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 1994.
32. Remm M, Storm CE, Sonnhammer EL. *J. Mol. Biol.* 2001; 314:1041–1052. [PubMed: 11743721]
33. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. *Cell.* 2009; 138:198–208. [PubMed: 19596244]
34. Fernandez-Escamilla A-M, Rousseau F, Schymkowitz J, Serrano L. *Nat. Biotechnol.* 2004; 22:1302–1306. [PubMed: 15361882]
35. Stark C, et al. *Nucleic Acids Res.* 2006; 34:D535–539. [PubMed: 16381927]
36. Ruepp A, et al. *Nucleic Acids Res.* 2008; 36:D646–650. [PubMed: 17965090]
37. Hart GT, Lee I, Marcotte ER. *BMCBioinformatics.* 2007; 8:236.
38. Ogata H, et al. *Nucleic Acids Res.* 1999; 27:29–34. [PubMed: 9847135]
39. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. *Genome Res.* 2011; 21:1109–1121. [PubMed: 21536720]
40. Lee I, Date SV, Adai AT, Marcotte EM. *Science.* 2004; 306:1555–1558. [PubMed: 15567862]
41. Costanzo M, et al. *Science.* 2010; 327:425–431. [PubMed: 20093466]
42. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. *Nat. Methods.* 2014; 11:319–324. [PubMed: 24487582]
43. Guo H, Ingolia NT, Weissman JS, Bartel DP. *Nature.* 2010; 466:835–840. [PubMed: 20703300]
44. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. *Science.* 2009; 324:218–223. [PubMed: 19213877]
45. Frank E, Hall M, Trigg L, Holmes G, Witten IH. *Bioinforma. Oxf. Engl.* 2004; 20:2479–2481.
46. Escalante-Chong R, et al. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:1636–1641. [PubMed: 25605920]
47. Sella G, Hirsh AE. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:9541–9546. [PubMed: 15980155]
48. Katoh K, Standley DM. *Mol. Biol. Evol.* 2013; 30:772–780. [PubMed: 23329690]
49. Houten SM, van Woerden CS, Wijburg FA, Wanders RJA, Waterham HR. *Eur. J. Hum. Genet. EJHG.* 2003; 11:196–200. [PubMed: 12634869]
50. D’Osueldo A, et al. *Eur. J. Hum. Genet. EJHG.* 2005; 13:314–320. [PubMed: 15536479]
51. Leyva-Vega M, et al. *Am. J. Med. Genet. A.* 2011; 155A:1461–1464. [PubMed: 21548022]
52. Kone-Paut I, Sanchez E, Le Quellec A, Manna R, Touitou I. *Ann. Rheum. Dis.* 2007; 66:832834.
53. Nevyjel M, et al. *Pediatrics.* 2007; 119:e523–527. [PubMed: 17261617]
54. Bayes A, et al. *Nat. Neurosci.* 2011; 14:19–21. [PubMed: 21170055]
55. Unno M, et al. *Struct. Lond. Engl.* 1993. 2002; 10:609–618.



**Fig. 1. Systematic functional replacement of essential yeast genes by their human counterparts** (A) Of 547 human genes with 1:1 orthology to essential yeast genes, 469 human ORFs were subcloned into single copy yeast expression vectors under control of either the GAL or GPD promoters. Using three distinct assay classes (repressible yeast-gene promoter, temperature-sensitive yeast allele, heterozygous diploid knockout strain), we obtained 126, 151, and 375 informative replaceability assays, respectively. (B) Representative examples of the three assay classes. (C) Combining assays and literature, 200 human genes could functionally replace their yeast orthologs and 224 genes could not. Some human genes were toxic using GAL induction but replaced their yeast orthologs upon reducing expression.



**Fig. 2. Properties of gene modules can predict replaceability**  
 (A) 104 quantitative features of proteins or ortholog pairs were evaluated for their ability to explain replaceability, assessing each feature’s predictive strength as the area under a ROC curve (AUC) and determining significance by shuffling replacement status 1,000 times, measuring mean AUCs +/- 1 standard deviation (s.d.). AUCs above 0.58 were generally individually significant with 95% confidence. Starred features were included in the integrated classifier (left-most bar). (B) Distribution of amino-acid identities among the tested ortholog pairs (left y axis) and fraction of replaceable genes in each sequence-identity

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



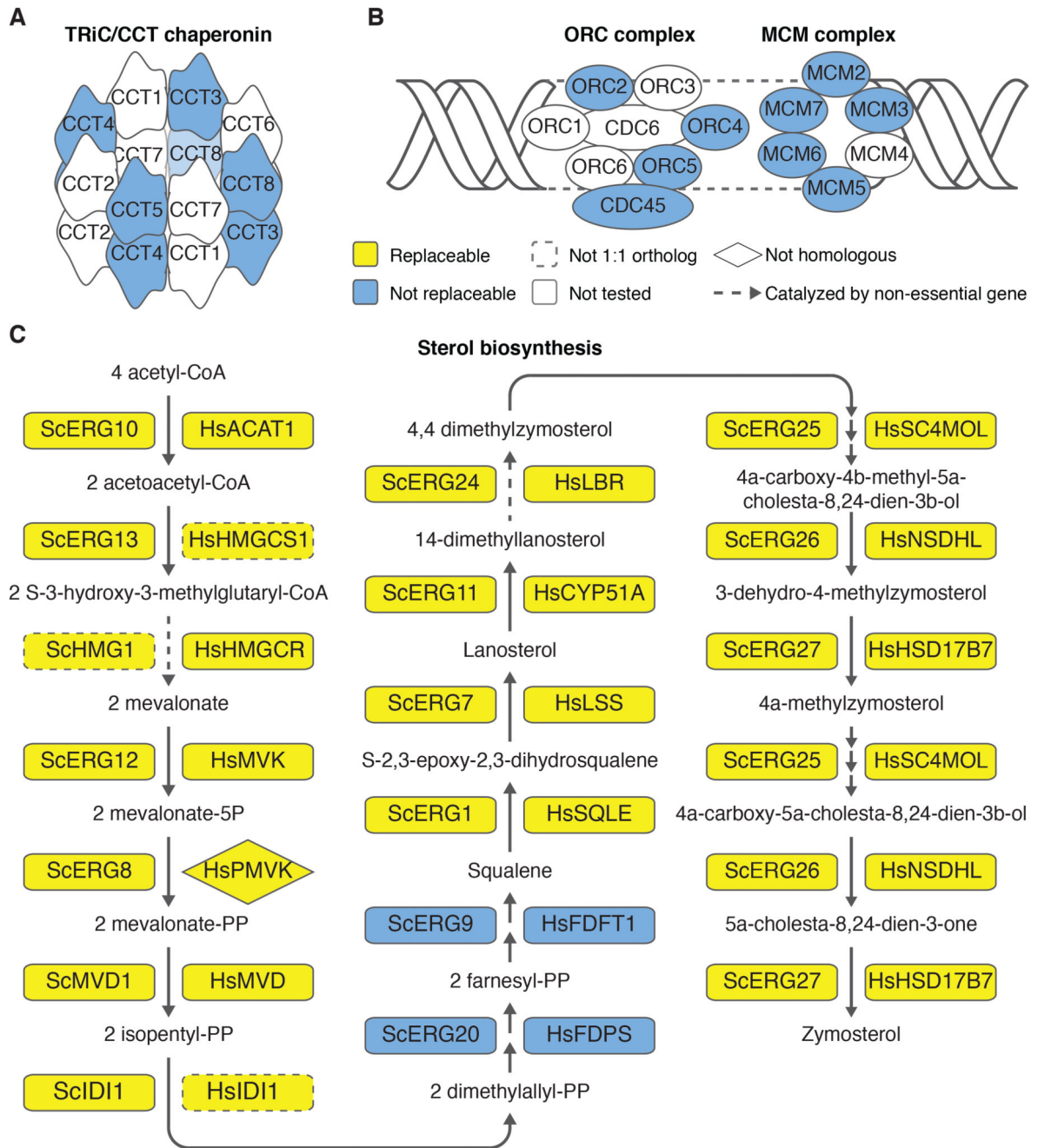
bin (right y axis). (C) Relative proportion of replaceable and non-replaceable genes among 12 broad KEGG (20) pathway classes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 3. The modular nature of functional replacement**

(A) None of the four tested human TRiC/CCT chaperonin genes replaced their yeast counterparts. (B) Similarly, no genes tested in the Origin Recognition Complex (ORC) or the Mini-Chromosome Maintenance (MCM) complex were replaceable. (C) In contrast, 17 of 19 sterol biosynthesis genes were replaceable. In two cases, the yeast gene had two human orthologs, but only one could complement. Human *HMGCS1* but not *HMGCS2* replaced yeast *ERG13*; human *ID11* but not *ID12* replaced yeast *ID11*. Human *PMVK*, a non-

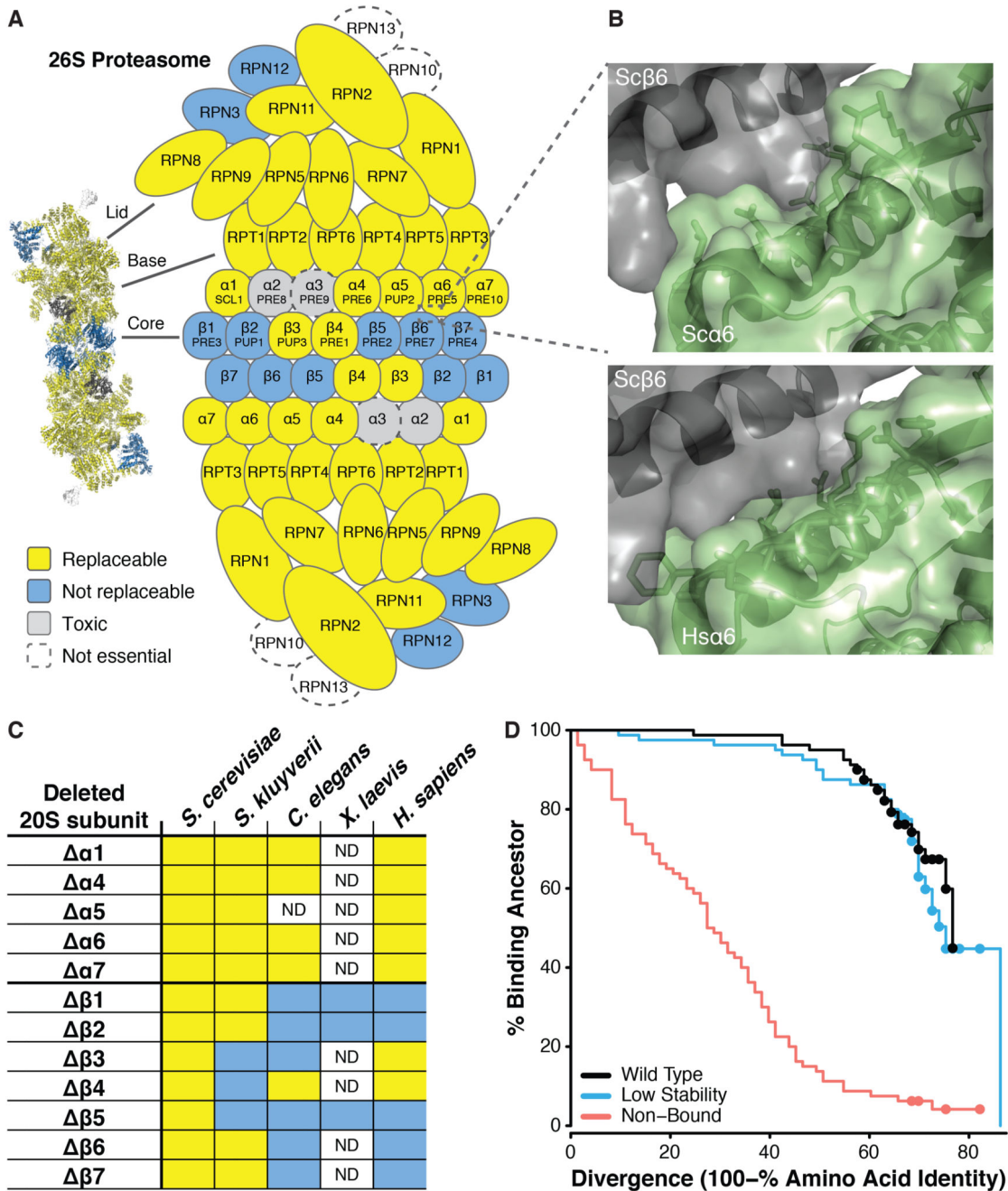
homologous protein that carries out the same reaction as yeast Erg8 (27), complemented temperature sensitive allele *erg8-1*.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 4. Proteasome subunits are differentially replaceable**

(A) Yeast 26S proteasome genes were generally replaceable, except for two interacting clusters, in the 19S regulatory “lid” particle and in the 20S core  $\beta$ -subunit ring. (B) The yeast  $\alpha 6$ - $\beta 6$  subunit interface (top panel) sterically accommodates the human subunit (bottom panel), showing superposition of human  $\alpha 6$  onto the yeast  $\alpha 6$  despite 50% sequence identity at the interface. (C) Alpha subunits from diverse eukaryotes generally complemented the yeast mutant, but not beta subunits (unlike plasmid-expressed *S. cerevisiae* genes, included as positive controls). (D) In simulated evolution of interacting proteins Ubc9 and Smt3, if

binding to the extant partner is not enforced (“Non-Bound”) a protein’s ability to bind its ancestral partner decays rapidly as sequences diverge. However, if extant binding is enforced (“Wild Type” and “Low Stability”), even highly diverged proteins often still bind to their ancestral partners. (Dots indicate right-censored data; see Fig. S14.)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript