May 30, 2019

Evolutionary decomposition of genetic effects

Li Liu[#], Mengdi Liu[#], Di Zhang, Shanjun Deng, Piaopiao Chen, Jing Yang, Yunhan Xie & Xionglei He

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

[#] These authors contributed equally to this work.

[*] Correspondence should be addressed to X.H. (hexiongl@mail.sysu.edu.cn).

Abstract - Genetics has long been confused by the pleiotropic mutational effects associated with even clearly-annotated genes. To understand such genetic effects, we here examine HAP4, a well-studied transcription factor in *Saccharomyces cerevisiae* that forms a tetramer with HAP2, HAP3, and HAP5. Deletion of HAP4 results in highly pleiotropic gene expression responses, some of which are clustered in related cellular processes (clustered effects) while most are distributed randomly across diverse cellular processes (distributed effects). The clustered effects tend to be conserved in related yeasts while the distributed effects, which account for the pleiotropy of HAP4, are genetically non-heritable and evolutionarily volatile. There are substantial overlaps among clustered effects, but not distributed effects, of the four genes encoding the HAP2/3/4/5 tetramer, and this pattern holds for other biochemically characterized protein complexes and KEGG pathways. We further examine a large set of cell morphological traits of the gene deletion lines and obtain consistent results. Hence, an expanded framework for reverse genetics is suggested, in which genetic effects are further divided into evolutionarily selected and unselected ones, with the former supporting related biochemistry understandings and the latter accounting for cryptic pleiotropy.

It has been increasingly recognized that a gene can affect various seemingly unrelated traits[1-6], a phenomenon termed pleiotropy[7,8]. For instance, a mapping of quantitative trait loci (QTL) for 70 skeletal traits in mice showed that on average a QTL

55 affects about 10% of the traits[9]. Attempts to understand pleiotropic genetic effects
56 were primarily from mechanistic perspectives[10-14], by considering the focal gene's
57 multiple molecular functions or multiple cellular consequences associated with a single
58 molecular function[15]. Since biological phenomena would make little sense except in
59 the light of evolution[16], a dissection of genetic effects from an evolutionary perspective
60 may provide novel insights[17].

61 The yeast *Saccharomyces cerevisiae* gene HAP4 is a non-essential transcription
62 factor that has been subjected to extensive studies since its discovery 30 years ago[18].
63 We deleted the open reading frame of HAP4 in *S. cerevisiae* strain BY4741, and
64 checked the expression trait of the other yeast genes (Table S1). We found 195
65 responsive genes each with a significant expression change under a stringent statistical
66 cutoff (Table S2). Gene ontology (GO) analysis of the 195 genes revealed one third
67 of them (65/195) clustered more than expected by chance in dozens of GO terms.
68 These GO terms are related with each other and reflect well the functional annotations
69 of HAP4 as a regulator of mitochondria activities[19,20] (Fig. 1A). The remaining two
70 thirds (130/195) distribute rather randomly across diverse biological processes,
71 underscoring the strong pleiotropy of HAP4. The two sets of genes are all functionally
72 well characterized, with comparable fitness importance ($P = 0.1$, Mann-Whitney U-test;
73 Fig. S1 and Table S2). Notably, the 65 clustered deletion effects and 130 distributed
74 deletion effects are supported by similar P-values and fold changes ($P = 0.20$ and 0.46,
75 respectively, Mann-Whitney U-test; Fig. 1B).

76 Genetic effects observed in an isogenic background should be heritable in

77 population to be evolutionarily effective[21]. We crossed two *S. cerevisiae* strains to
78 obtain a population of yeast segregants. Specifically, a wild-type strain
79 BY4742(MATapha), which is identical to BY4741 except at the mating locus, was
80 crossed with the HAP4 deletion line of DBVPG1373(MATa) (Fig. 1C). This way, the
81 comparison between wild-type and null alleles at the HAP4 locus would match the
82 comparison conducted in the isogenic BY4741 background. We dissected six tetrads
83 of the hybrid and obtained 12 HAP4 wild-type and 12 HAP4 null segregants. For each
84 of the 195 deletion effects we defined a heritability score HS = Var(wt, null) / Var(total),
85 where Var(wt, null) is the variance of the focal gene's expression level between the two
86 segregant groups, and Var(total) is the total variance across the segregants. We found
87 that the 65 clustered effects in general have much higher HS than the 130 distributed
88 effects ($P = 1.1\times10^{-3}$, Mann-Whitney U-test; Fig. 1D). Approximately 26.2% (17/65)
89 of the clustered effects are significantly heritable under a stringent statistical cutoff,
90 while the number is 0.8% (1/130) for the distributed effects ($P = 2.1\times10^{-8}$, Fisher's
91 exact test; Fig. 1E).

92 To test how the genetic effects are maintained by natural selection, we examined
93 their evolutionary conservation. We obtained, under the same statistical cutoff as in
94 BY4741, the deletion effects of HAP4 in *Saccharomyces paradoxus* strain N17, a
95 closely related yeast species diverged from *S. cerevisiae* ~10 million years ago[22] (Fig.
96 1F). Only 5.4% (7/130) of the distributed effects found in *S. cerevisiae* were also
97 observed in the HAP4 deletion line of *S. paradoxus* N17, while the number was 27.7%
98 (18/65) for the clustered effects ($P = 3.1\times10^{-5}$, Chi-square test). The difference is

99 robust as evidenced by plotting the expression responses of the 195 genes in N17(Δhap4)
100 (Fig. 1G). We further examined intra-species conservation of the HAP4 deletion
101 effects by looking at two other *S. cerevisiae* strains DBVPG1373 and GIL104, the
102 former of which is 0.35% diverged from the strain BY4741 and the latter 0.07%
103 diverged at the genomic level[23]. Similar1y, the 130 distributed effects were largely
104 unreproducible in strains DBVPG1373(Δhap4) and GIL104(Δhap4), while the 65
105 clustered effects were much more conserved ($P = 0.5\times10^{-4}$ for DBVPG1373(Δhap4)
106 and $P = 2.6\times10^{-5}$ for GIL104(Δhap4), Chi-square test; Fig. S2). As expected, both
107 types of the deletion effects are more reproducible in more related yeasts (Fig. 1H).
108 We further confirmed that the results cannot be explained by different detectability of
109 expression changes between the two gene sets by excluding those genes lowly
110 expressed in wild-type BY4741 (Fig. S3). These data, together with the heritability
111 analysis, suggest that the clustered effects of HAP4 tend to be selected in evolution; on
112 the contrary, the distributed genetic effects appear non-evolutionary, likely representing
113 context-dependent *ad hoc* responses of cell to the gene deletion[24].

114 HAP4 functions by forming a tetramer with HAP2, HAP3 and HAP5 as a result of
115 evolution[25]. We hypothesized that clustered genetic effects should support this
116 biochemistry understanding better than distributed genetic effects, because the latter is
117 non-evolutionary. To test the hypothesis, we deleted the other three genes that encode
118 the tetramer, respectively, in *S. cerevisiae* BY4741, and measured the expression
119 profiles of the deletion lines. We defined clustered effects and distributed effects for
120 each of the lines using the same method as in the HAP4 deletion line. We obtained

121 43, 150 and 50 clustered effects, and 61, 306 and 111 distributed effects for the deletions
122 of HAP2, HAP3 and HAP5, respectively (Fig. 2A-C; Table S3). Consistent with the
123 hypothesis, we found 20 overlapped clustered effects across the four gene deletion lines,
124 14.5 times higher than that of the distributed effects ($P < 0.001$, simulation test, Fig.
125 2D). Notably, the 20 overlapped clustered effects are not the strongest in
126 BY4741(Δhap4) (Fig. S4). To avoid the potential bias that expression responses to
127 the tetramer may have been considered in the GO annotations of the responsive genes,
128 we excluded all expression-related evidences for GO annotations to define new
129 clustered and distributed effects. We obtained essentially the same result (Fig. S5).
130 Because there are publicly available microarray data for HAP2, HAP3, HAP4 and
131 HPA5 deletion lines[26], we also repeated the analysis using the public expression data
132 and observed a similar pattern (Fig. S6).

133 In addition to considering the protein complex formed by HAP4, we could also
134 consider protein-DNA interactions since HAP4 is a transcription factor. Data from a
135 chromatin immuno-precipitation (CHIP) assay of the promoters bound by HAP4 show
136 that, among the 195 responsive genes observed in BY4741(Δhap4), 13 are direct targets
137 of HAP4 (Fig. 2E)[27,28]. Interestingly, there are 24-fold enrichment of direct targets in
138 the clustered effects relative to the distributed effects ($P = 8.6\times10^{-6}$, Fisher's exact test);
139 among the 20 overlapped clustered effects 50% (10/20) are direct targets of HAP4,
140 while the genomic background is 0.64% (33/5146) ($P = 4.6\times10^{-18}$, Hypergeometric test)
141 (Fig. 2F). Hence, the CHIP data well support the distinction of the two effect types.

142 Collectively, these results are consistent with a recently proposed model that the

143 null phenotype of a gene can be ascribed to either the loss of native functions of the
144 gene, or the gain of spurious functions arising from the genetic manipulation due to
145 passive adjustments of the cell system[24]. The key difference of the two function types
146 is their evolutionary effectiveness: native functions are historical and evolutionarily
147 selected while spurious functions are ahistorical and non-evolutionary [29-32].
148 Accordingly, the distributed effects examined here likely represent spurious functions
149 created by the HAP4 deletion, and the clustered effects could be in a large part ascribed
150 to the native functions of HAP4. It is intriguing how the two effect types defined by
151 GO could fit the two function types described in the model. We reason that
152 evolutionarily optimized native functions are likely to regulate specific pathways or
153 processes; losing them would thus cause clustered changes of the related genes, which
154 are detected by GO analysis. In contrast, spurious functions may affect the
155 transcriptome in a rather random way, resulting in distributed changes across diverse
156 cellular processes, most of which cannot be covered by overrepresented GO terms.
157 This may explain why GO clustering here could echo evolutionary effectiveness.

158 Regardless of the underlying logic, clustered genetic effects seem to be well
159 matched with related biochemistry understandings. This would help resolve a long-
160 standing challenge to molecular biology - the gap between genetic analysis and
161 biochemistry analysis[33,34]; specifically, genes with intimate biochemical interactions do
162 not have common genetic effects and genes with common genetic effects do not show
163 intimate biochemical interactions[35,36]. To test the generality of the finding that was
164 based on the HAP2/3/4/5 tetramer, we examined other biochemically characterized

165 protein complexes by using publicly available expression data. To avoid bias we
166 considered a single dataset comprising microarray-based expression profiles of over
167 one thousand yeast gene deletion lines[26]. There are 54 protein complexes annotated
168 by a previous study suitable for our analysis[37] . In 24 cases the overlaps of clustered
169 effects are significantly more than what would be expected from distributed effects at
170 a 99% confidence level, and the enrichments range from 2.7-fold to over 100-fold with
171 a median 5.3-fold (Fig. 3A and Table S4). The overlapped clustered effects of each
172 protein complex represent specific functions (Fig. 3B-C and Fig. S7). For example,
173 the ten overlapped clustered effects of the elongator holoenzyme complex are tens to
174 hundreds times overrepresented in a few transcription-related GO terms as well as
175 proteasome-related GO terms (Fig. 3B), the former of which echo well the annotated
176 functions of the complex while the latter appear to suggest new understandings[38]. As
177 another example, the genes encoding the protein kinase CK2 complex have 21
178 overlapped clustered effects that appear to affect specifically the metabolism of various
179 amino acids (Fig. 3C), a functional insight not been well recognized[39].

180 We also checked genes on the same KEGG pathways. There are 41 pathways that
181 are related to metabolism, genetic information processing, cellular processes, and so on,
182 suitable for our analysis (Table S5). The rate of overlaps of clustered effects is
183 significant higher than that of distributed effects in nine cases, and the enrichments
184 range from 5.6-fold to over 100-fold with a median 46.9-fold (Fig. 3D and Table S5).
185 Consistently, the overlapped clustered effects of each pathway represent distinct
186 functions (Fig. 3E-F and Fig. S7). For the many cases in which clustered effects show

187 no more overlaps than distributed effects, the involved genes may execute additional
188 functions irrelevant to the focal complex or pathway. Notably, in none of the cases
189 distributed effects represent related biochemistry understandings better than clustered
190 effects, highlighting the cryptic nature of them. Taken together, focusing on clustered
191 effects is a readily operational approach to narrowing the gap between genetic analysis
192 and biochemistry data.

193 The above analyses considered gene expression traits. We next examined the
194 yeast cell morphological traits that are based on the microscopic images of cells stained
195 by fluorescent dyes[4]. With the help of a computer software as many as 405
196 quantitative traits can be obtained from cell wall and nuclear stained cell images[40].
197 These traits are typically about area, distance, and angle calculated based on dozens of
198 coordinate points, lines and angles that describe the shape of mother cell and bud, and
199 the shape and localization of the nuclei in mother cell and bud (Fig. 4A). This large
200 set of yeast traits had served as a great resource for studying genotype-phenotype
201 relationships[17,41,42]. Deletion of HAP4 in *S. cerevisiae* significantly altered 78
202 morphological traits, among which 24 are also significantly affected in *S. paradoxus* by
203 HAP4 deletion (Table S6). To test if the evolutionarily conserved effects of HAP4 are
204 shared with HAP2, HAP3 and HAP5 more than the non-conserved effects, we also
205 measured the morphological traits affected by each of the other three genes, respectively,
206 in *S. cerevisiae*. We found that 58.3% (14/24) of the conserved effects are shared with
207 all the other three genes, which is significantly higher than the number (18/54 = 33.3%)
208 for the non-conserved effects of HAP4 (P = 0.035, Fisher's exact test; Fig. 4B). The

difference remains largely unchanged when only traits with small measuring noise are considered (Fig. S8). Hence, the cell morphology data also support the role of evolution in separating genetic effects.

In summary, by re-assessing genetic causality in the light of evolution this study suggests an expanded framework for reverse genetics (Fig. 4C). Specifically, the conventional framework aims to find causality only: it relies on statistical tests to separate the mutant versus wild-type differences into significant and insignificant effects. In the expanded framework evolutionary effectiveness is included to further separate significant effects into evolutionarily selected and unselected one, the former of which would support related biochemistry understandings while the latter account for cryptic pleiotropy. When expression traits are examined, GO clustering serves as a proxy for evolutionary effectiveness, which offers a readily operational means to delineate expression network using genetic data.

Methods

Yeast gene deletions

Three *S. cerevisiae* (SC) strains BY4741 (*MATa, his3, leu2, met15, ura3*), GIL104 (*MATa, URA3, leu2, trp1, CAN1, ade2, his3, bar1Δ::ADE2*; derived from the W303) and DBVPG1373 (*MATa, ura3*), and one S. paradoxus (SP) strain N17 (*MATa, ura3*) were included in the study. Unless otherwise stated, the S. cerevisiae strains were cultured in the rich medium YPD (1%Yeast extract, 1%Peptone, 2% Dextrose) at 30°C, and *S. paradoxus* N17 was cultured in YPD at 25°C. The wild type URA3 in GIL104 was first replaced by a LEU2 cassette. HAP4 was replaced by a URA3 cassette in each of the four strains. HAP2, HAP3, and HAP5 were also replaced, respectively, by a URA3 cassette in BY4741. Notably, for all gene replacements the whole open reading frame from the starting codon to the stop codon of a focal gene was replaced. As described in our previous study[43], the standard LiAc transformation method[44] was used to transform DNA into the yeast cells and gene replacements were achieved by homologous recombination. The transformation protocol was slightly modified for *S. paradoxus*[45]; specifically, heat shock was performed for seven minutes at 37°C.

Synthetic medium deprived of uracil or leucine was used to select the clones with successful replacement for the target gene. All gene replacements were verified using polymerase chain reaction (PCR). For each gene deletion line, 3-5 independent clones were obtained for further examination, which effectively controlled the potential effects of secondary mutations introduced during the gene replacement.

Because haploid yeast cells tend to flocculate, which is not suitable for cell morphology characterization, diploid yeasts are required in the analysis of morphological traits. We first obtained haploid gene deletion strains (SC-BY4741 or SP-N17 background; MATa), which were then crossed with the corresponding MATalpha wild-type strain, respectively. The diploid heterozygous gene deletion strains were sporulated by following the method of a previous study[46]. Specifically, the cells were incubated in YEP (1% yeast extract,1% Bacto peptone,0.05% NaCl) containing 2% potassium acetate) for five hours at 30 °C to start the sporulation process. The culture was centrifuged (2,000g, for 2 min), the cell pellet was washed three times by sterile water, and re-suspended in sporulation media (10g/L potassium acetate and 50mg/L zinc acetate) for five days at 25°C with shaking. The products were incubated with 200U/ml lyticase (Sigma #L4025) for 30 min at 30°C followed by 15 min at 50°C. The products were washed by sterile water and then plated on synthetic medium deprived of uracil for two days at 30°C for SC strains, or 25°C for SP strains. The genotypes of the colonies were determined by PCR. For each gene the haploid deletion strains of both mating types were obtained. A pair of MATa and MATalpha strains with the same gene deletion were crossed to obtain a diploid homozygous gene deletion strain. For each gene deletion line three independent clones were obtained and examined.

Obtain a population of segregants

A wild-type strain BY4742 (MATapha), which is identical to BY4741 except at the mating locus, was crossed with a HAP4 deletion stain of DBVPG1373 (MATa). Two biological replications were carried out. The diploid heterozygous deletion strains were sporulated for 3-5 days in sporulation medium on a shaking table at 25°C. Tetrads were obtained and incubated with 200U/ml lyticase for 3-5 min at 30°C, and then streaked onto a YPD plate for tetrad dissection using the MSM400 dissection microscope (Singer Instrument Company Ltd). Spores were grown on YPD plates at 30°C for two days, and the genotypes of the colonies were determined by PCR. We selected only those tetrads that produce four segregants with genotypes MATa+HAP4, MATa+ Δhap4, MATalpha+HAP4, and MATalpha+Δhap4, respectively. A total of 24 segregants from six such tetrads were obtained for the heritability analysis.

RNA sequencing and data analysis

For each strain a single colony on agar plate was picked and cultured in YPD liquid overnight at 30 °C with shaking. Approximately 200uL saturated culture was added into 10mL fresh YPD, which resulted in an optical density OD600 ~ 0.1 (UNICO UV/VIS Spectrophotometer), and cells of 3mL culture at OD600 = 0.5-0.65 were harvested. Total RNA was extracted by QIAGEN RNeasy Plus mini kit (Cat

No.74136). The mRNA sequencing was performed using the paired-end module on a HiSeq platform at Genwiz by following the standard procedure. To ensure the high quality of expression analysis, we sequenced the mRNA of 3-6 independent clones for each wild-type or gene deletion line.

RNA-seq reads were mapped to reference yeast genomes using STAR (Version 2.6.0c)[47]. For BY4741 and GIL104, we used the genome of *S. cerevisiae* strain S288C as the reference (version R64-2-1_20150113; http://www.yeastgenome.org). The reference genomes of SC-DBVPG1373 and SP-N17 were downloaded from SGRP (https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html). For a typical clone there were about 6.5 million paired-end reads mapped to the coding sequences. Gene expression levels were determined by Featurecounts (version 1.6.2)[48] with default settings and RPKM (reads per kilobase per million) of each gene were calculated by R package edgeR. The wild-type versus mutant differential expression analysis was performed by DESeq2[49] with default parameters, and genes with an adjusted P-value smaller than 0.05 and a fold change (FC) greater than 1.5 were defined as significantly changed genes. In the conservation, heritability or overlapping analysis, an effect is called conserved, heritable or overlapped only when it shows the same direction in the various conditions examined. Genes of the uracil biosynthesis pathway (YBL039C, YEL021W, YJL130C, YJR103W, YKL024C, YKL216W, YLR420W, YML106W, YMR271C) were excluded from further analysis. The expression level measured by RPKM of HAP2, HAP3, HAP4 and HAP5 in wild-type BY4741 is 84.4, 64, 194 and 90.5, respectively, and all becomes zero in the corresponding deletion lines. The fitness of yeast gene deletion lines was produced by a previous study[50]. Table S1 contains details of the RNA-seq expression information of all yeast lines examined in this study.

GO analysis

The GO analyses of the responsive genes derived from our RNA-seq data were conducted in the SGD website using GO Term Finder (Version 0.86; https://www.yeastgenome.org/goTermFinder), by excluding computational analysis evidences and other less reliable evidences: IBA, IC, IEA, IKR, IRD, ISA, ISM, ISO, ISS, NAS, ND, TAS. In a strict analysis which required the exclusion of all expression-related evidences, only three GO evidence codes IDA, HDA and IPI that represent direct experimental assays were considered. For the GO analyses of public microarray data the R package clusterProfiler[51] was used with default settings. The cutoffs used to define an enriched GO term include an adjusted P-value smaller than 0.01 and a fold enrichment greater than 2. To improve specificity only GO terms containing less than 200 genes were considered. The fold enrichment was calculated as (number of changed genes in the GO term / number of all changed genes) / (number of genes in the GO term / number of genes in all GO terms of the class). The GO semantic similarity scores were calculated by R package GOSemSim[52].

Heritability analysis

To estimate the heritability of the 195 HAP4 deletion effects, a population of 24

327 segregants were divided into two groups: 12 with the wild-type allele of HAP4 and 12
328 with the null allele of HAP4. The heritability score H was the variation between the
329 two groups divided by the total variation of the population. To test the statistical
330 significant of an H, we compared the expression levels of the focal gene between the
331 two segregant groups using DESeq2. The obtained 195 raw P-values were adjusted
332 for multiple testing using the Benjamini-Hochberg controlling procedure. An
333 adjusted P-value smaller than 0.05 was considered significant.
334
335 **Analyze protein complexes and metabolic pathways**
336 The public microarray data of ~1,400 yeast gene deletion lines were obtained from a
337 previous study[26], and P-values and fold changes (FC) provided in the data were directly
338 used. Specifically, $P < 0.05$ and absolute $FC > 1.2$ were used to define genes with
339 significant expression changes; if the number of significantly changed genes was over
340 1,500, a more stringent cutoff $P < 0.01$ was used. To avoid the effects of genes with
341 ubiquitous expression responses we excluded from further analyses the top 10% genes
342 that each show significant changes in at least 12% of the gene deletion lines. GO
343 analyses were performed by R package clusterProfiler to define clustered and
344 distributed effects for each deletion line, with the results summarized in Table S7. To
345 examine the overlapped clustered effects between genes of the same protein complex
346 or pathway, we only considered the deletion lines with at least 20 clustered effects,
347 resulting in a set of 422 deletion lines suitable for further analyses.
348 Information of 518 protein complexes was obtained from a previous study [37]. The
349 KEGG pathways of the yeast *S. cerevisiae* were downloaded from KEGG website
350 (https://www.genome.jp/kegg-bin/get_htext?sce00001). There are 54 complexes and
351 41 pathways each with at least two member genes found in the above defined mutant
352 set.
353 For each protein complex or pathway, the overlaps of clustered effects and the
354 overlaps of distributed effects were compared in number. The numbers of clustered
355 effects and distributed effects of the involved genes were normalized to make the
356 overlaps between the two effect types comparable. To estimate the confidence
357 interval of a comparison we used random samplings. If clustered effects are less than
358 distributed effects in all genes, which is true in most of the cases examined, we sampled
359 (without replacements) a random subset of distributed effects to ensure the two effect
360 types of a gene equal in number. If clustered effects are more than distributed effects
361 in all genes, we sampled (without replacements) a random subset of distributed effects
362 to ensure the two effect types of a gene equal in number. If the above consistent
363 patterns do not exist, we sampled consistently from one side (either clustered effects or
364 distributed effects) but with replacements for the gene with an insufficient number of
365 effects on this side. For each complex or pathway 1,000 such random samplings were
366 carried out to derive the 99% confidence interval, and an observed difference is called
367 significant if it is not within the interval. Table S4 and Table S5 have details about the
368 protein complexes and KEGG pathways examined, respectively.
369
370 **Analyze cell morphological traits**

371 Diploid yeast cells were examined by following the protocol of previous studies with
372 slight modifications[4,40]. In brief, a single yeast colony was picked and cultured in
373 YPD liquid overnight with shaking to the saturation phase. Then, 1.5uL culture was
374 transferred to 100uL fresh YPD in a 96-well plate and grew for 3-4 hours at 30°C for
375 SC strains or 25°C for SP strains. Cells were fixed with 3.7% formaldehyde solution.
376 Cell wall was stained by FITC-ConA (fluorescein isothiocyanate-conjugated,
377 concanavalin A, Sigma-Aldrich C7642). Cell nucleus was stained by hochest-mix
378 (Thermo Fisher, Hoechst 33342 Solution) instead of DAPI to enhance the specificity.
379 We did not stain actin because the dye Rhodamine phalloidin was not stable enough to
380 support the following high-throughput automated image capturing which takes about
381 10 hours for scanning 96 wells of a plate. The stained cells were plated into a
382 microplate (Greiner 781091) with ~5.0×10^4 cells per well and images were captured by
383 IN Cell Analyzer 2200 (GE Healthcare) using the 60× objective lens.
384 Five SC lines (all diploid with BY4741 background: wild-type, Δhap2, Δhap3,
385 Δhap4 and Δhap5) and two SP lines (all diploid with N17 background: wild-type and
386 Δhap4) were examined. For each line three independent clones were available. For
387 each clone there were 6-8 technical replicates in the staining and image capturing. The
388 images were analyzed by CalMorph[4,40] with default settings, and only 405 rather than
389 501 traits were extracted in this study because actin is not stained. At least 1,000 cells
390 were captured and analyzed (with at least 100 informative cells for each cell-cycle stage)
391 for a high-quality replicate. In the end, there were 13~23 high-quality replicates for
392 each of the lines included in further analysis. Trait values were compared between
393 replicates of a gene deletion line and replicates of the corresponding wild-type line
394 using T-test, and the resulting 405 P-values were adjusted for multiple testing using the
395 Benjamini-Hochberg controlling procedure. Because of the many replicates included
396 in the comparison, many traits showed a statistically significant but biologically
397 negligible difference between wild-type and mutant lines. Hence, a trait is called
398 affected by a gene only when the adjusted $P < 0.05$ and the difference between wild-
399 type and mutant is large than 5%. Table S8 has complete information regarding the
400 morphological trait analysis.
401

402 **Acknowledgments**

403 We are grateful to financial support from NSFC (grant #31630042 and #91731302 to
404 X. H.), technical support from Z. Zhou and X. Chen, and helpful discussions with J.
405 Yang, J. Zhang, W. Qian, and Y. Zhang.
406

407 **References**

408 1 Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy
409 and phenotypically derived gene function in yeast. *Mol Syst Biol* **1**, 2005 0001,
410 doi:10.1038/msb4100004 (2005).
411 2 Sonnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in Caenorhabditis
412 elegans. *Nature* **434**, 462-469, doi:10.1038/nature03353 (2005).
413 3 Mackay, T. F. Mutations and quantitative genetic variation: lessons from Drosophila. *Philos
414 Trans R Soc Lond B Biol Sci* **365**, 1229-1239, doi:10.1098/rstb.2009.0315 (2010).

415 4 Ohya, Y. *et al.* High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad*
416 *Sci U S A* **102**, 19015-19020, doi:10.1073/pnas.0509436102 (2005).
417 5 Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *Am J Hum*
418 *Genet* **89**, 607-618, doi:10.1016/j.ajhg.2011.10.004 (2011).
419 6 Wang, Z., Liao, B. Y. & Zhang, J. Genomic patterns of pleiotropy and the evolution of complexity.
420 *Proc Natl Acad Sci U S A* **107**, 18034-18039, doi:10.1073/pnas.1004666107 (2010).
421 7 Caspari, E. J. E. PLEIOTROPIC GENE ACTION. **6**, 1-18 (1952).
422 8 Stearns, F. W. One hundred years of pleiotropy: a retrospective. *Genetics* **186**, 767-773,
423 doi:10.1534/genetics.110.122549 (2010).
424 9 Wagner, G. P. *et al.* Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature* **452**,
425 470-472, doi:10.1038/nature06756 (2008).
426 10 Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype-phenotype map: the
427 evolvability of complex organisms. *Nat Rev Genet* **12**, 204-213, doi:10.1038/nrg2949 (2011).
428 11 Paaby, A. B. & Rockman, M. V. The many faces of pleiotropy. *Trends Genet* **29**, 66-73,
429 doi:10.1016/j.tig.2012.10.010 (2013).
430 12 Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits:
431 challenges and strategies. *Nat Rev Genet* **14**, 483-495, doi:10.1038/nrg3461 (2013).
432 13 Hodgkin, J. Seven types of pleiotropy. *Int J Dev Biol* **42**, 501-505 (1998).
433 14 Foster, K. R., Shaulsky, G., Strassmann, J. E., Queller, D. C. & Thompson, C. R. Pleiotropy as a
434 mechanism to stabilize cooperation. *Nature* **431**, 693-696, doi:10.1038/nature02894 (2004).
435 15 He, X. L. & Zhang, J. Z. Toward a molecular understanding of pleiotropy. *Genetics* **173**, 1885-
436 1891, doi:10.1534/genetics.106.060269 (2006).
437 16 Ayala, F. J. "Nothing in biology makes sense except in the light of evolution": Theodosius
438 Dobzhansky: 1900-1975. *J Hered* **68**, 3-10 (1977).
439 17 Chen, H., Wu, C. I. & He, X. The genotype-phenotype relationships in the light of natural
440 selection. *Molecular biology and evolution*, doi:10.1093/molbev/msx288 (2017).
441 18 Forsburg, S. L. & Guarente, L. Identification and characterization of HAP4: a third component
442 of the CCAAT-bound HAP2/HAP3 heteromer. *Genes Dev* **3**, 1166-1178 (1989).
443 19 Lascaris, R. *et al.* Hap4p overexpression in glucose-grown Saccharomyces cerevisiae induces
444 cells to enter a novel metabolic state. *Genome Biol* **4**, R3 (2003).
445 20 Buschlen, S. *et al.* The S. Cerevisiae HAP complex, a key regulator of mitochondrial function,
446 coordinates nuclear and mitochondrial gene expression. *Comparative and functional genomics*
447 **4**, 37-46, doi:10.1002/cfg.254 (2003).
448 21 Pinzon, N. *et al.* microRNA target prediction programs predict many false positives. *Genome*
449 *Res* **27**, 234-245, doi:10.1101/gr.205146.116 (2017).
450 22 Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of
451 yeast species to identify genes and regulatory elements. *Nature* **423**, 241-254,
452 doi:10.1038/nature01644 (2003).
453 23 Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-341,
454 doi:10.1038/nature07743 (2009).
455 24 He, X. The Biology Complicated by Genetic Analysis. *Molecular biology and evolution* **33**, 2177-
456 2181, doi:10.1093/molbev/msw111 (2016).
457 25 Bolotin-Fukuhara, M. Thirty years of the HAP2/3/4/5 complex. *Biochim Biophys Acta Gene*
458 *Regul Mech* **1860**, 543-559, doi:10.1016/j.bbagrm.2016.10.011 (2017).

459 26 Kemmeren, P. *et al.* Large-scale genetic perturbations reveal regulatory networks and an
460 abundance of gene-specific repressors. *Cell* **157**, 740-752, doi:10.1016/j.cell.2014.02.054
461 (2014).
462 27 Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-
463 104, doi:10.1038/nature02800 (2004).
464 28 MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for Saccharomyces
465 cerevisiae. *BMC Bioinformatics* **7**, 113, doi:10.1186/1471-2105-7-113 (2006).
466 29 Neander, K. Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of*
467 *Science* **58**, 168-184 (1991).
468 30 Millikan, R. G. In Defense of Proper Functions. *Philosophy of Science* **56**, 288-302 (1989).
469 31 Amundson, R. & Lauder, G. V. Function without purpose. *Biology and Philosophy* **9**, 443-469,
470 doi:10.1007/BF00850375 (1994).
471 32 Graur, D. *et al.* On the Immortality of Television Sets: "Function" in the Human Genome
472 According to the Evolution-Free Gospel of ENCODE. *Genome Biol Evol* **5**, 578-590,
473 doi:10.1093/gbe/evt028 (2013).
474 33 Griffiths, A. J. F. *An Introduction to genetic analysis*. 5th edn, (W.H. Freeman, 1993).
475 34 Wong, S. L., Zhang, L. V. & Roth, F. P. Discovering functional relationships: biochemistry versus
476 genetics. *Trends Genet* **21**, 424-427, doi:10.1016/j.tig.2005.06.006 (2005).
477 35 He, X., Qian, W., Wang, Z., Li, Y. & Zhang, J. Prevalent positive epistasis in Escherichia coli and
478 Saccharomyces cerevisiae metabolic networks. *Nat Genet* **42**, 272-276, doi:10.1038/ng.524
479 (2010).
480 36 Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular
481 function. *Science* **353**, doi:10.1126/science.aaf1420 (2016).
482 37 Benschop, J. J. *et al.* A consensus of core protein complex compositions for Saccharomyces
483 cerevisiae. *Mol Cell* **38**, 916-928, doi:10.1016/j.molcel.2010.06.002 (2010).
484 38 Svejstrup, J. Q. Elongator complex: how many roles does it play? *Current Opinion in Cell Biology*
485 **19**, 331-336, doi:10.1016/j.ceb.2007.04.005 (2007).
486 39 Litchfield, D. W. Protein kinase CK2: structure, regulation and role in cellular decisions of life
487 and death. *The Biochemical journal* **369**, 1-15, doi:10.1042/BJ20021469 (2003).
488 40 Okada, H., Ohnuki, S. & Ohya, Y. Quantification of cell, actin, and nuclear DNA morphology with
489 high-throughput microscopy and CalMorph. *Cold Spring Harb Protoc* **2015**, 408-412,
490 doi:10.1101/pdb.prot078667 (2015).
491 41 Ho, W. C. & Zhang, J. The genotype-phenotype map of yeast complex traits: basic parameters
492 and the role of natural selection. *Molecular biology and evolution* **31**, 1568-1580,
493 doi:10.1093/molbev/msu131 (2014).
494 42 Nogami, S., Ohya, Y. & Yvert, G. Genetic complexity and quantitative trait loci mapping of yeast
495 morphological traits. *Plos Genet* **3**, e31, doi:10.1371/journal.pgen.0030031 (2007).
496 43 Chen, P., Wang, D., Chen, H., Zhou, Z. & He, X. J. G. r. The non-essentiality of essential genes in
497 yeast provides therapeutic insights into a human disease. *Genome Res.* **26**, 1355-1362 (2016).
498 44 Gietz, R. D. & Schiestl, R. H. J. N. p. High-efficiency yeast transformation using the LiAc/SS
499 carrier DNA/PEG method. *Nat. Protoc.* **2**, 31-34 (2007).
500 45 Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome
501 Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3-Genes Genom*
502 *Genet* **1**, 11-25, doi:10.1534/g3.111.000273 (2011).

503 46 Enyenihi, A. H. & Saunders, W. S. Large-scale functional genomic analysis of sporulation and
504 meiosis in Saccharomyces cerevisiae. *Genetics* **163**, 47-54 (2003).
505 47 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
506 doi:10.1093/bioinformatics/bts635 (2013).
507 48 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
508 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
509 doi:10.1093/bioinformatics/btt656 (2014).
510 49 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
511 RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
512 50 Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The Genomic Landscape and Evolutionary
513 Resolution of Antagonistic Pleiotropy in Yeast. *Cell reports* **2**, 1399-1410,
514 doi:https://doi.org/10.1016/j.celrep.2012.09.017 (2012).
515 51 Yu, G. C., Wang, L. G., Han, Y. Y. & He, Q. Y. clusterProfiler: an R Package for Comparing Biological
516 Themes Among Gene Clusters. *Omics-a Journal of Integrative Biology* **16**, 284-287,
517 doi:10.1089/omi.2011.0118 (2012).
518 52 Yu, G. C. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms
519 and gene products. *Bioinformatics* **26**, 976-978, doi:10.1093/bioinformatics/btq064 (2010).
520
521
522

523 **Figure legends**

524 **Fig. 1. Tests of evolutionary effectiveness of HAP4 deletion effects.**

525 **(A):** The 195 responsive genes are enriched in dozens of GO terms that are related and
526 also reflect well the functional annotations of HAP4. The heatmap shows the pairwise
527 similarity of the enriched GO terms, with three subclasses each corresponding to certain
528 biological processes that are summarized at the right.

529 **(B):** The P-value (adjusted for multiple testing) and fold change (FC) of the 65 clustered
530 effects and 130 distributed effects. Each dot represents a responsive gene (i.e., an
531 effect).

532 **(C):** Obtain a population of segregants with different genetic backgrounds to test the
533 heritability of HAP4 deletion effects.

534 **(D):** The 65 clustered effects have greater heritability scores than the 130 distributed
535 effects ($P = 0.0011$, Mann-Whitney U-test). Each dot represents an effect, and P-value
536 measures the statistical significance of a heritability score, with the vertical dashed line
537 showing adjusted P = 0.05.

538 **(E):** The proportions of deletion effects that are significantly heritable with adjusted P
539 < 0.05. Error bars represent SE.

540 **(F):** A dendrogram showing the phylogeny the four yeast strains examined in this study.

541 **(G):** Conservation analysis of the HAP4 deletion effects. The 195 responsive genes
542 defined in BY4741(Δhap4) are examined with respect to their expression responses in
543 *S. paradoxus* N17 (Δhap4). The horizontal dashed line shows adjusted P = 0.05 and
544 vertical dashed lines show $\log_2 FC = \pm 0.58$. (cyan: clustered effects; red: distributed
545 effects; cycle: down-regulated in BY4741(Δhap4); triangle: up-regulated in
546 BY4741(Δhap4))

547 **(H):** The rate of conservation in the three related yeasts for the 65 clustered effects and
548 130 distributed effects defined in BY4741(Δhap4), respectively. Error bars represent
549 SE.

550

551 **Fig. 2. Clustered effects of the four genes encoding the HAP2/3/4/5 tetramer**
552 **overlap a lot more than their distributed effects do.**

553 **(A, B, C):** The P-value (adjusted for multiple testing) and fold change (FC) of the
554 clustered effects and distributed effects defined in the three BY4741 strains Δhap2,
555 Δhap3, and Δhap5, respectively. Each dot represents a responsive gene (i.e., an effect),
556 and the total number of responsive genes is shown at the bottom next to the effect type.

557 **(D):** There are 20 overlapped clustered effects for the four genes encoding the
558 HAP2/3/4/5 tetramer, which is significantly higher than expectation. The expectation
559 is estimated by random sampling of the distributed effects of the four genes to calculate
560 overlaps, and 1,000 such simulations were conducted.

561 **(E):** Among the 195 responsive genes found in BY4741(Δhap4) 13 are direct target of
562 HAP4 according to a chromatin immune-precipitation assay.

563 **(F):** The proportion of direct target of HAP4 in different gene sets. Error bars
564 represent SE.

565

566 **Fig. 3. Clustered effects support related biochemistry understandings much**

better than distributed effects in a variety of protein complexes and KEGG pathways.

(A): The clustered effects of genes encoding a protein complex in general overlap more than their distributed effects. Each circle represents a complex, and the filled ones are significant at a 99% confidence level estimated by random sampling. A total of 54 protein complexes are included here, with 24 cases showing at least twice more overlapped clustered effects than overlapped distributed effects (below the line y = 0.5x). The numbers of effects have been normalized such that in each case the overlaps of clustered effects and the overlaps of distributed effects can be directly compared.

(B): The representative GO terms of the overlapped clustered effects of the elongator holoenzyme complex. Only four genes encoding the complex, which are highlighted in orange, have suitable expression data for the analysis. There are 10 overlapped clustered effects of the four genes, which are over 300 times more than expected. The expectation is estimated by random sampling of the distributed effects of the focal genes to calculate overlaps. The orange circle each represents an overlapped clustered effect, and the blue circles represent the enriched GO terms of the overlapped clustered effects with the number inside showing the fold enrichment in the given term.

(C): The representative GO terms of the overlapped clustered effects of the protein kinase CK2 complex. There are 21 overlapped clustered effects, 83.3 times more than expected.

(D): The clustered effects of genes in the same KEGG pathway also tend to overlap more than their distributed effects. Each circle represents a pathway, and the filled ones are significant at a 99% confidence level. A total of 41 pathways are included, with nine showing at least five times more overlapped clustered effects than overlapped distributed effects (below the line y = 0.2x). The number of effects have been normalized such that in each case the overlaps of clustered effects and the overlaps of distributed effects can be directly compared.

(E): The representative GO terms of the overlapped clustered effects of the four genes in the metabolic pathway sce00260. There are six overlapped clustered effects, over 100 times more than expected.

(F): The representative GO terms of the overlapped clustered effects of the four genes in the genetic information processing pathway sce03010. There are 10 overlapped clustered effects, 21.4 times more than expected.

Fig. 4. Examination of cell morphological traits also supports the role of evolution in separating genetic effects.

(A): The yeast cell morphology outlined by coordinate points, lines and angles (only some are shown) based on which a total of 405 quantitative traits can be derived by a computer software.

(B): The traits affected by HAP4 in both *S. cerevisiae* and *S. paradoxus* (i.e., conserved effects) are more likely to overlap with those affected by HAP2, HAP3 and HAP5 than the traits affected by HAP4 only in *S. cerevisiae* (non-conserved effects) (P = 0.035, one-tailed Fisher's exact test). A total of 78 morphological traits significantly affected

by HAP4 deletion in *S. cerevisiae* are examined, among which 24 are conserved effects and 54 non-conserved effects. Overlaps refer to traits significantly affected by all four gene deletions in *S. cerevisiae.* Error bars represent SE.

(C): Proposition of an expanded framework for reverse genetics. Statistically significant genetic effects defined in conventional framework are further separated into evolutionarily selected and unselected, with the former supporting related biochemistry understandings and the latter accounting for pleiotropy.

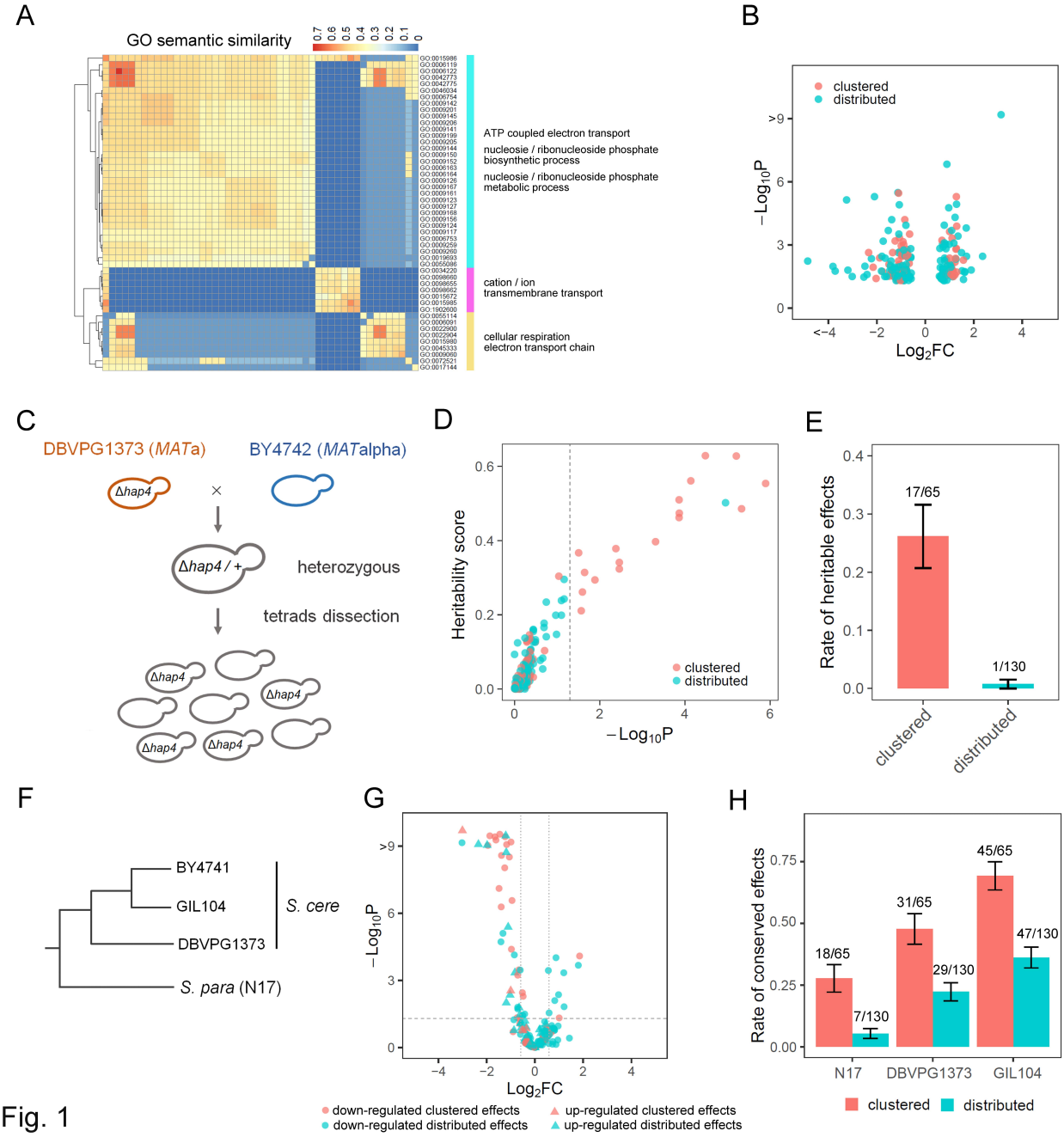

A
GO semantic similarity
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
ATP coupled electron transport
nucleosie / ribonucleoside phosphate biosynthetic process
nucleosie / ribonucleoside phosphate metabolic process
cation / ion transmembrane transport
cellular respiration electron transport chain
B
clustered
distributed
>9
6
3
0
−Log10P
<−4
−2
0
2
4
Log2FC
C
DBVPG1373 (MATa)
BY4742 (MATalpha)
Δhap4
×
Δhap4 / +
heterozygous
tetrads dissection
Δhap4
D
Heritability score
0.6
0.4
0.2
0.0
0
2
4
6
−Log10P
clustered
distributed
E
Rate of heritable effects
0.4
0.3
0.2
0.1
0.0
17/65
1/130
clustered
distributed
F
BY4741
GIL104
DBVPG1373
S. cere
S. para (N17)
G
>9
6
3
0
−Log10P
−4
−2
0
2
4
Log2FC
down-regulated clustered effects
down-regulated distributed effects
up-regulated clustered effects
up-regulated distributed effects
H
Rate of conserved effects
0.75
0.50
0.25
0.00
18/65
7/130
31/65
29/130
45/65
47/130
N17
DBVPG1373
GIL104
clustered
distributed

Fig. 1

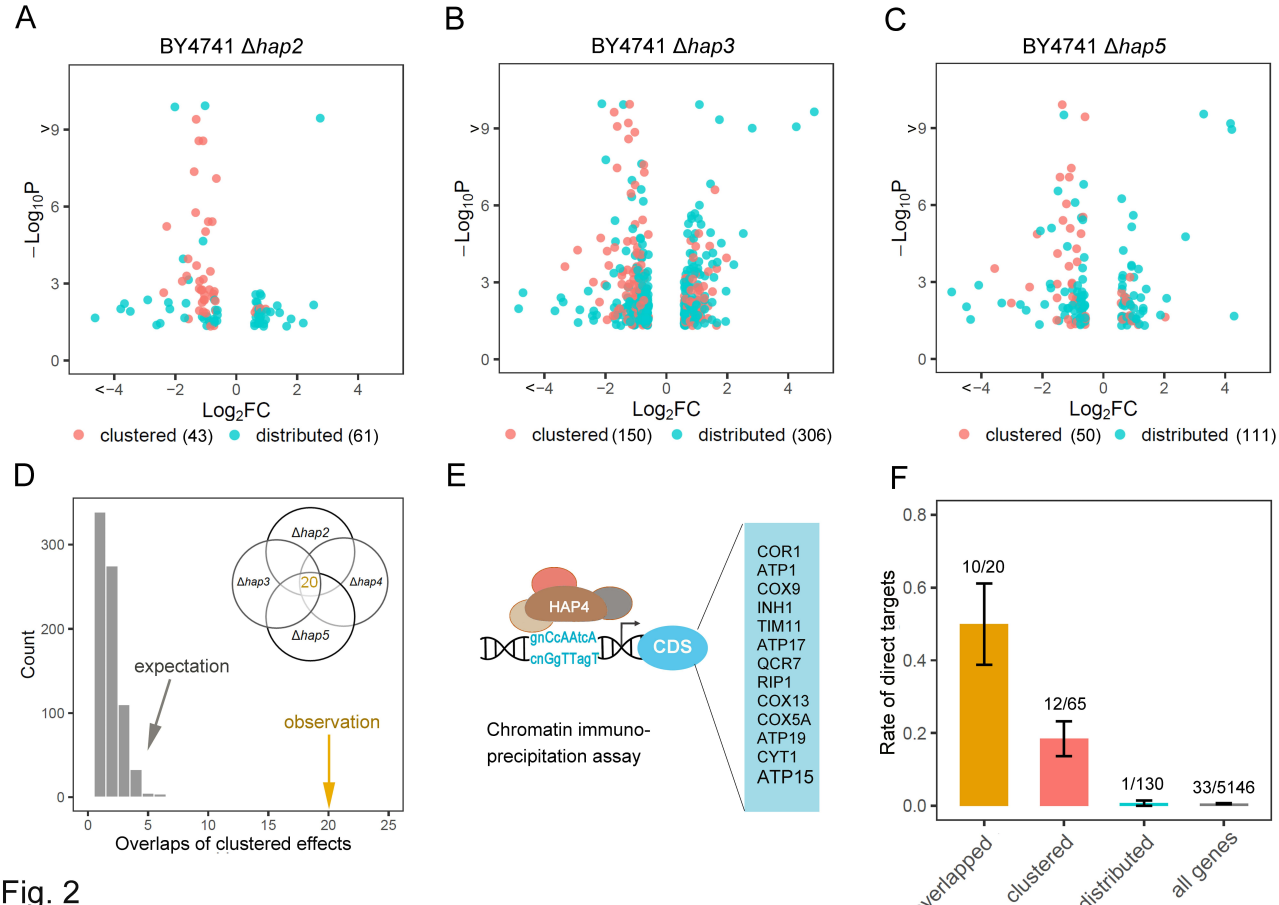

A
BY4741 Δhap2
−Log10P
>9
6
3
0
<−4
−2
0
2
4
Log2FC
clustered (43)
distributed (61)
B
BY4741 Δhap3
clustered (150)
distributed (306)
C
BY4741 Δhap5
clustered (50)
distributed (111)
D
Count
300
200
100
0
Δhap2
Δhap3
Δhap4
Δhap5
20
expectation
observation
0
5
10
15
20
25
Overlaps of clustered effects
E
HAP4
gnCcAAtcA
cnGgTTagT
CDS
Chromatin immuno-
precipitation assay
COR1
ATP1
COX9
INH1
TIM11
ATP17
QCR7
RIP1
COX13
COX5A
ATP19
CYT1
ATP15
F
Rate of direct targets
0.8
0.6
0.4
0.2
0.0
10/20
12/65
1/130
33/5146
overlapped
clustered
distributed
all genes

Fig. 2

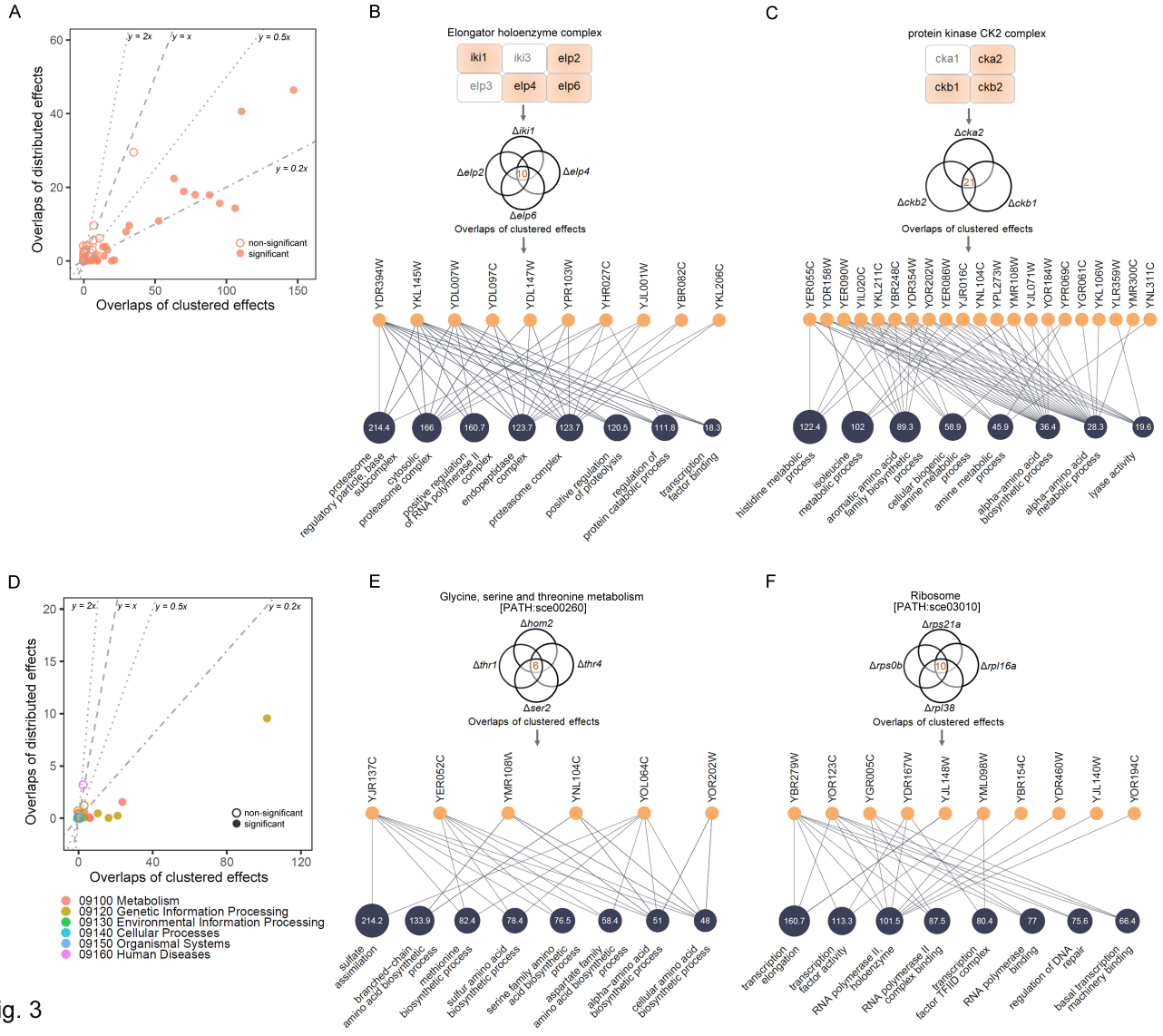
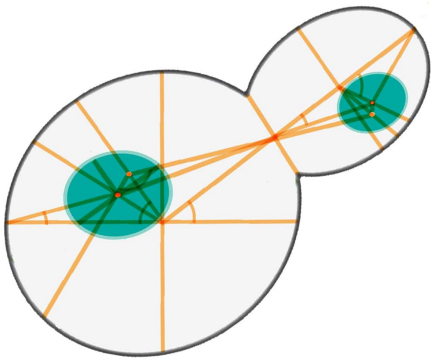
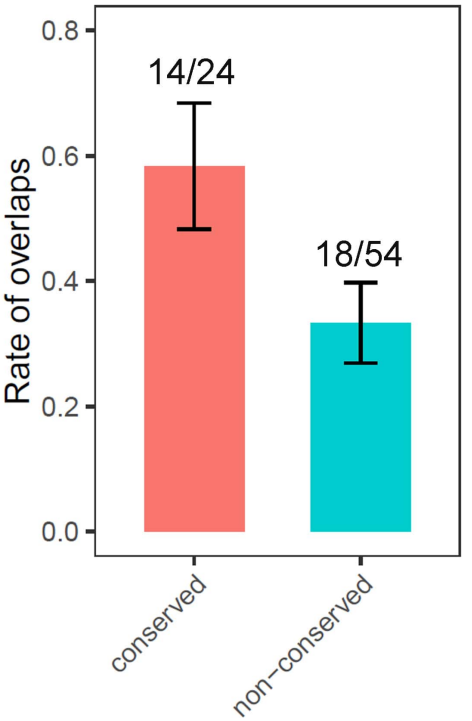

A
y = 2x
y = x
y = 0.5x
y = 0.2x
Overlaps of distributed effects
Overlaps of clustered effects
non-significant
significant
B
Elongator holoenzyme complex
iki1
iki3
elp2
elp3
elp4
elp6
Δiki1
Δelp2
Δelp4
Δelp6
10
Overlaps of clustered effects
YDR394W
YKL145W
YDL007W
YDL097C
YDL147W
YPR103W
YHR027C
YJL001W
YBR082C
YKL206C
214.4
166
160.7
123.7
123.7
120.5
111.8
18.3
proteasome regulatory particle base subcomplex
cytosolic proteasome complex
positive regulation of RNA polymerase II complex
endopeptidase complex
proteasome complex
positive regulation of proteolysis
regulation of protein catabolic process
transcription factor binding
C
protein kinase CK2 complex
cka1
cka2
ckb1
ckb2
Δcka2
Δckb2
Δckb1
21
Overlaps of clustered effects
YER055C
YDR158W
YER090W
YIL020C
YKL211C
YBR248C
YDR354W
YOR202W
YER086W
YJR016C
YNL104C
YPL273W
YMR108W
YJL071W
YOR184W
YPR069C
YGR061C
YKL106W
YLR359W
YMR300C
YNL311C
122.4
102
89.3
58.9
45.9
36.4
28.3
19.6
histidine metabolic process
isoleucine metabolic process
aromatic amino acid family biosynthetic process
cellular biogenic amine metabolic process
amine metabolic process
alpha-amino acid biosynthetic process
alpha-amino acid metabolic process
lyase activity
D
y = 2x
y = x
y = 0.5x
y = 0.2x
Overlaps of distributed effects
Overlaps of clustered effects
non-significant
significant
09100 Metabolism
09120 Genetic Information Processing
09130 Environmental Information Processing
09140 Cellular Processes
09150 Organismal Systems
09160 Human Diseases
E
Glycine, serine and threonine metabolism
[PATH:sce00260]
Δhom2
Δthr1
Δthr4
Δser2
6
Overlaps of clustered effects
YJR137C
YER052C
YMR108W
YNL104C
YOL064C
YOR202W
214.2
133.9
82.4
78.4
76.5
58.4
51
48
sulfate assimilation
branched-chain amino acid biosynthetic process
methionine biosynthetic process
sulfur amino acid biosynthetic process
serine family amino acid biosynthetic process
aspartate family amino acid biosynthetic process
alpha-amino acid biosynthetic process
cellular amino acid biosynthetic process
F
Ribosome
[PATH:sce03010]
Δrps21a
Δrps0b
Δrpl16a
Δrpl38
10
Overlaps of clustered effects
YBR279W
YOR123C
YGR005C
YDR167W
YJL148W
YML098W
YBR154C
YDR460W
YJL140W
YOR194C
160.7
113.3
101.5
87.5
80.4
77
75.6
66.4
transcription elongation
transcription factor activity
RNA polymerase II, holoenzyme
RNA polymerase II complex binding
transcription factor TFIID complex
RNA polymerase binding
regulation of DNA repair
basal transcription machinery binding

Fig. 3

A



B


0.8
0.6
0.4
0.2
0.0
Rate of overlaps
14/24
18/54
conserved
non−conserved

C

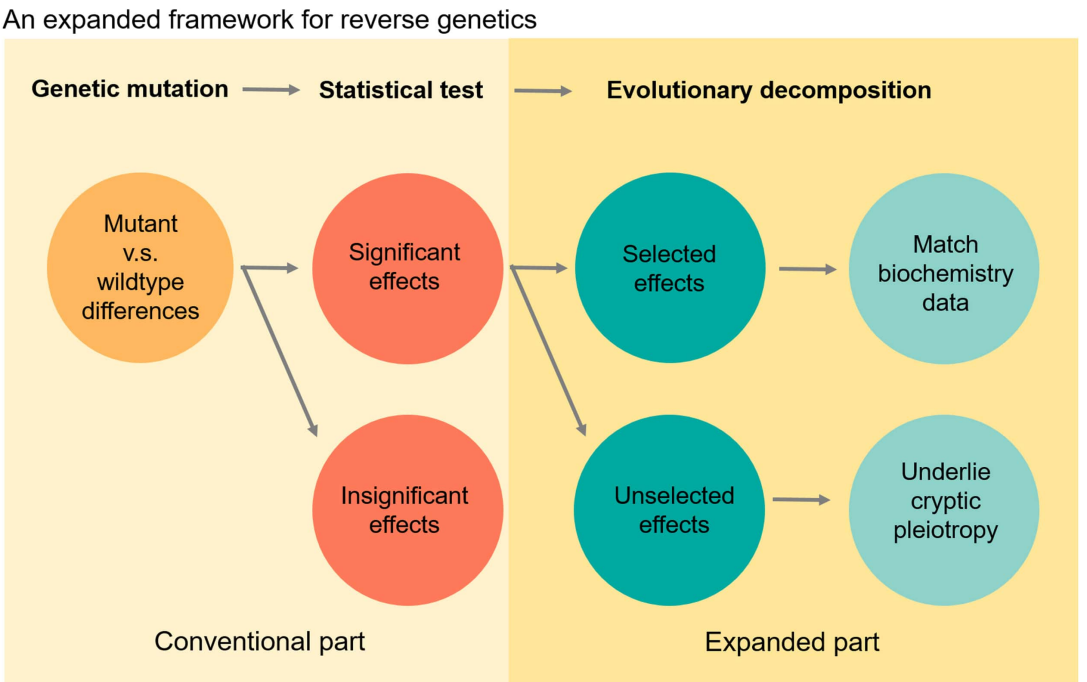

An expanded framework for reverse genetics
Genetic mutation
Statistical test
Evolutionary decomposition
Mutant v.s. wildtype differences
Significant effects
Insignificant effects
Selected effects
Unselected effects
Match biochemistry data
Underlie cryptic pleiotropy
Conventional part
Expanded part

Fig. 4

Supplementary Information of

"Evolutionary decomposition of genetic effects"

Li Liu[#], Mengdi Liu[#], Di Zhang, Shanjun Deng, Piaopiao Chen, Jing Yang, Yunhan Xie & Xionglei He*

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

This file contains:

Legends of Tables S1 to S8

Figs. S1 to S8

Legends of supplementary tables

Table S1: RNA-seq-based gene expression levels for each HAP4 deletion and wild-type strain.

Table S2: RNA-seq-based expression changes (P-value and FC) of the 195 responsive genes in BY4741(Δhap4) and other deletion lines, with the heritability scores also included.

Table S3: RNA-seq-based gene expression changes after deleting HAP2, HAP3, HAP4 and HAP5, respectively, in BY4741.

Table S4: Summary of the analyses of protein complexes in this study.

Table S5: Summary of the analyses of KEGG pathways in this study.

Table S6: The affected morphological traits in a variety of gene deletion lines.

Table S7: Summary of the clustered effects and distributed effects defined in each mutant that has public microarray data.

Table S8: Summary of the trait information of each diploid gene deletion or wild-type yeast strain, with the 405 trait values, the number of examined cells, and the number of replications included.

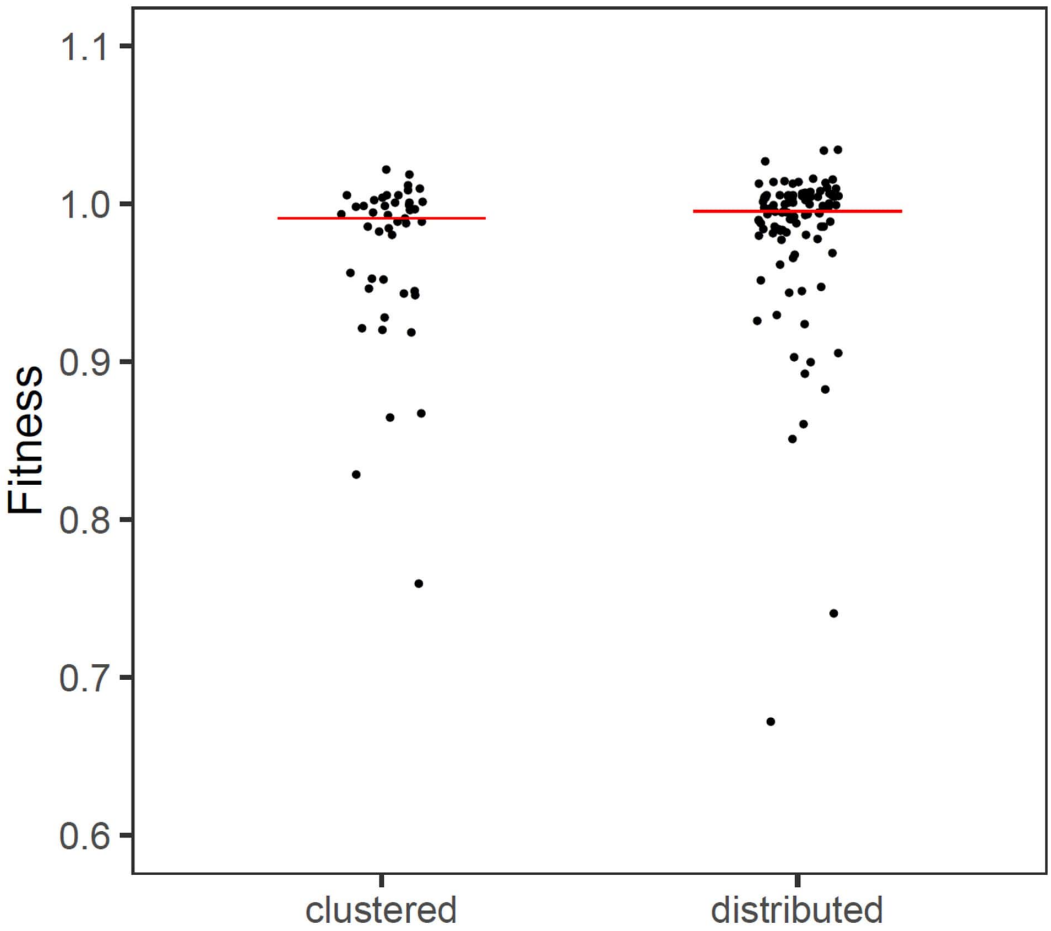


Fig. S1. Fitness importance is comparable between the clustered responsive genes and distributed responsive genes defined in BY4741(Δhap4) (*P* = 0.10, Mann-Whitney U-test). Fitness importance of a gene is measured by the relative growth rate of the gene deletion line to wild-type. The horizontal line shows the median.

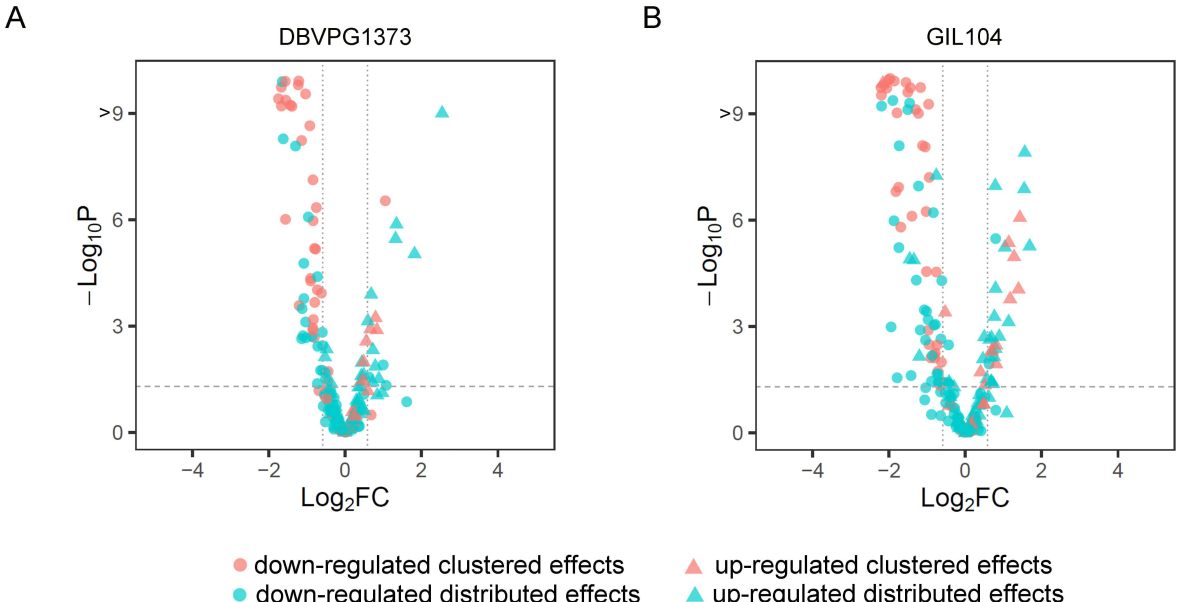


Fig. S2. The intra-species conservation analysis of the HAP4 deletion effects. The 195 responsive genes defined in BY4741(Δhap4) are examined with respect to their expression responses in *S. cerevisiae* DBVPG1373(Δhap4) and GIL104(Δhap4), respectively. The horizontal dashed line shows adjusted P = 0.05 and vertical dashed lines show log2FC = ±0.58. (cyan: clustered effects; red: distributed effects; cycle: down-regulated in BY4741(Δhap4); triangle: up-regulated in BY4741(Δhap4)).

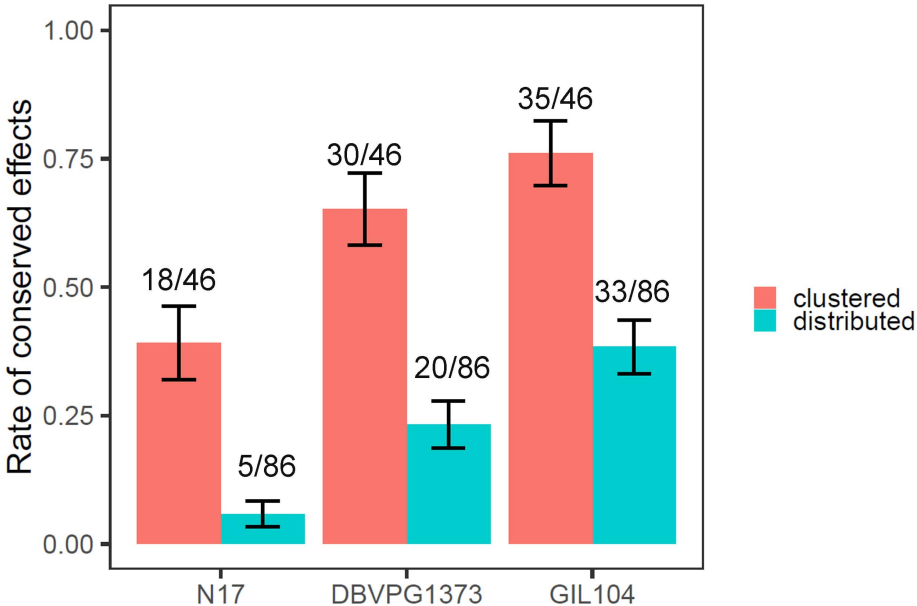


Fig. S3. Conservation analysis of the HAP4 deletion effects by considering only genes with a strong expression level in wild-type BY4741. This analysis is to address the concern that lowly expressed genes in wild-type tend not to have detectable down-regulation due to technical bias. Hence, for the 195 responsive genes defined in BY4741(Δhap4) only those with log2RPKM > 5 in wild-type BY4741 are considered here, leaving 46 clustered effects and 86 distributed effects. Error bars represent SE.

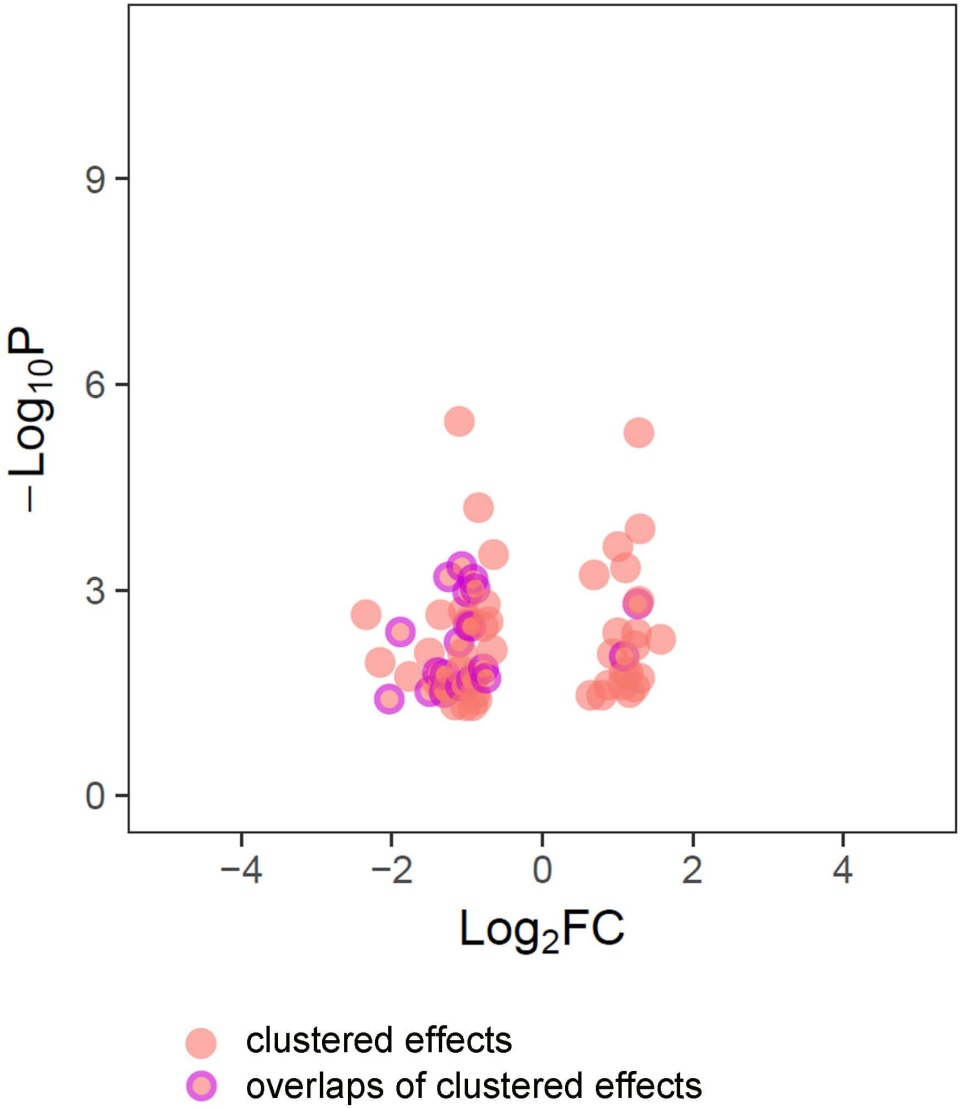
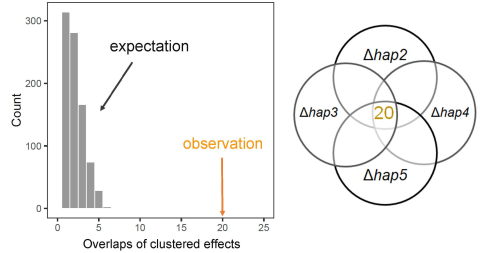


Fig. S4. The 20 overlapped clustered effects are comparable to the rest 45 (65-20) clustered effects defined in BY4741(Δhap4) with regard to their effect size. The differences are not statistically significant for both the P-values and fold changes observed in BY4741(Δhap4) (P = 0.75 and 0.51, respectively, Mann-Whitney U-test)

A

Deleted gene	Number of clustered effects	Number of distributed effects
HAP2	46	58
HAP3	140	316
HAP4	70	125
HAP5	53	108

B



C

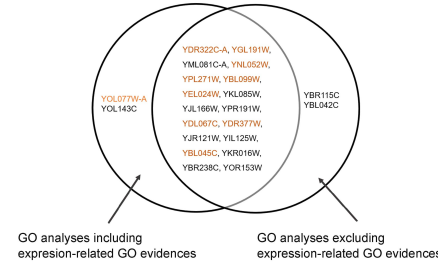
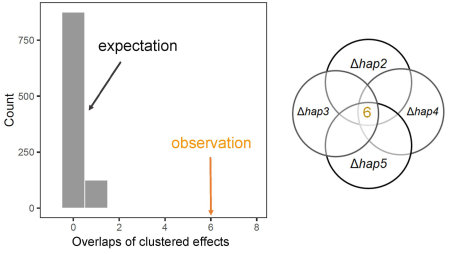


Fig. S5. The strong overlaps of clustered effects of the HAP2/3/4/5 complex genes are not biased by expression-related GO evidences for annotating the deletion effects. (A) Only the evidences IDA, HDA and IPI are used to re-define the clustered and distributed deletion effects of the four genes, respectively. (B) There are 20 overlapped clustered effects for the four genes encoding the HAP2/3/4/5 tetramer, which is significantly higher than expectation. The expectation is estimated by random sampling of the distributed effects of the four genes to calculate overlaps, and 1,000 such simulations were conducted. (C) The overlapped clustered effects are largely the same before and after excluding expression-related GO evidences. Genes that are the direct target of HAP4 are highlighted in yellow.

A

Deleted gene	Number of clustered effects	Number of distributed effects
HAP2	130	173
HAP3	33	41
HAP4	81	119
HAP5	12	39

B



C

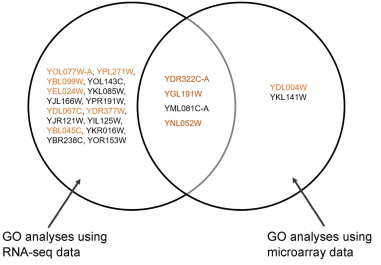


Fig. S6. The enrichment of overlapped clustered effects in HAP2/3/4/5 complex is reproduced by using public microarray data of the four gene deletion lines. (A) Microarray-based expression data are used to re-define the clustered and distributed deletion effects of the four genes, respectively. (B) There are only six overlapped clustered effects for the four genes encoding the HAP2/3/4/5 tetramer, which is significantly higher than expectation. The reduced number is primarily due to the small number (12) of clustered effects observed in HAP5 deletion. The expectation is estimated by random sampling of the distributed effects of the four genes to calculate overlaps, and 1,000 such simulations were conducted. (C) Comparison of the six overlapped clustered effects defined using microarray data with the 20 overlapped clustered effects defined using RNA-seq data. Genes that are the direct target of HAP4 are highlighted in yellow.

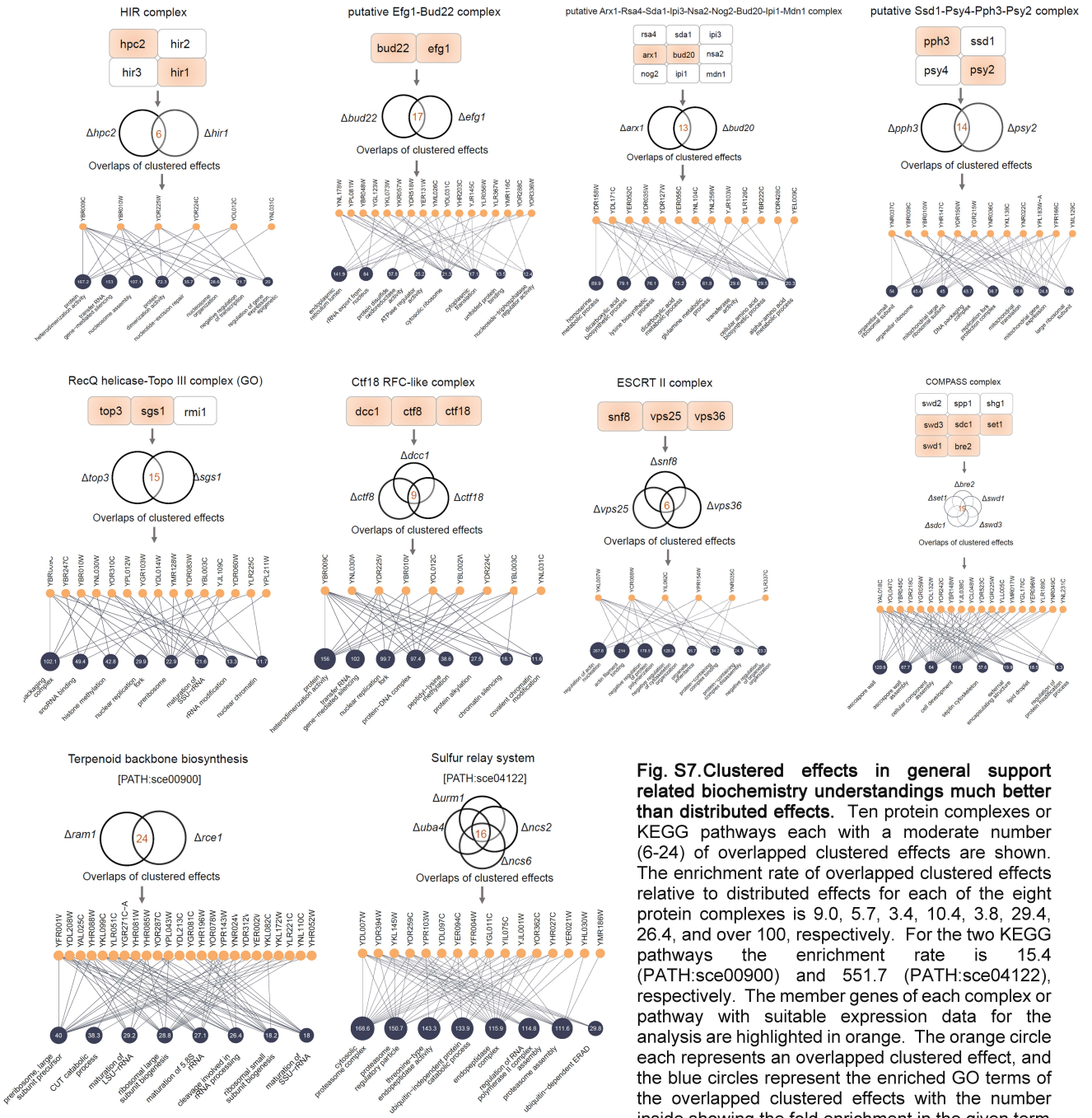


Fig. S7. Clustered effects in general support related biochemistry understandings much better than distributed effects. Ten protein complexes or KEGG pathways each with a moderate number (6-24) of overlapped clustered effects are shown. The enrichment rate of overlapped clustered effects relative to distributed effects for each of the eight protein complexes is 9.0, 5.7, 3.4, 10.4, 3.8, 29.4, 26.4, and over 100, respectively. For the two KEGG pathways the enrichment rate is 15.4 (PATH:sce00900) and 551.7 (PATH:sce04122), respectively. The member genes of each complex or pathway with suitable expression data for the analysis are highlighted in orange. The orange circle each represents an overlapped clustered effect, and the blue circles represent the enriched GO terms of the overlapped clustered effects with the number inside showing the fold enrichment in the given term.

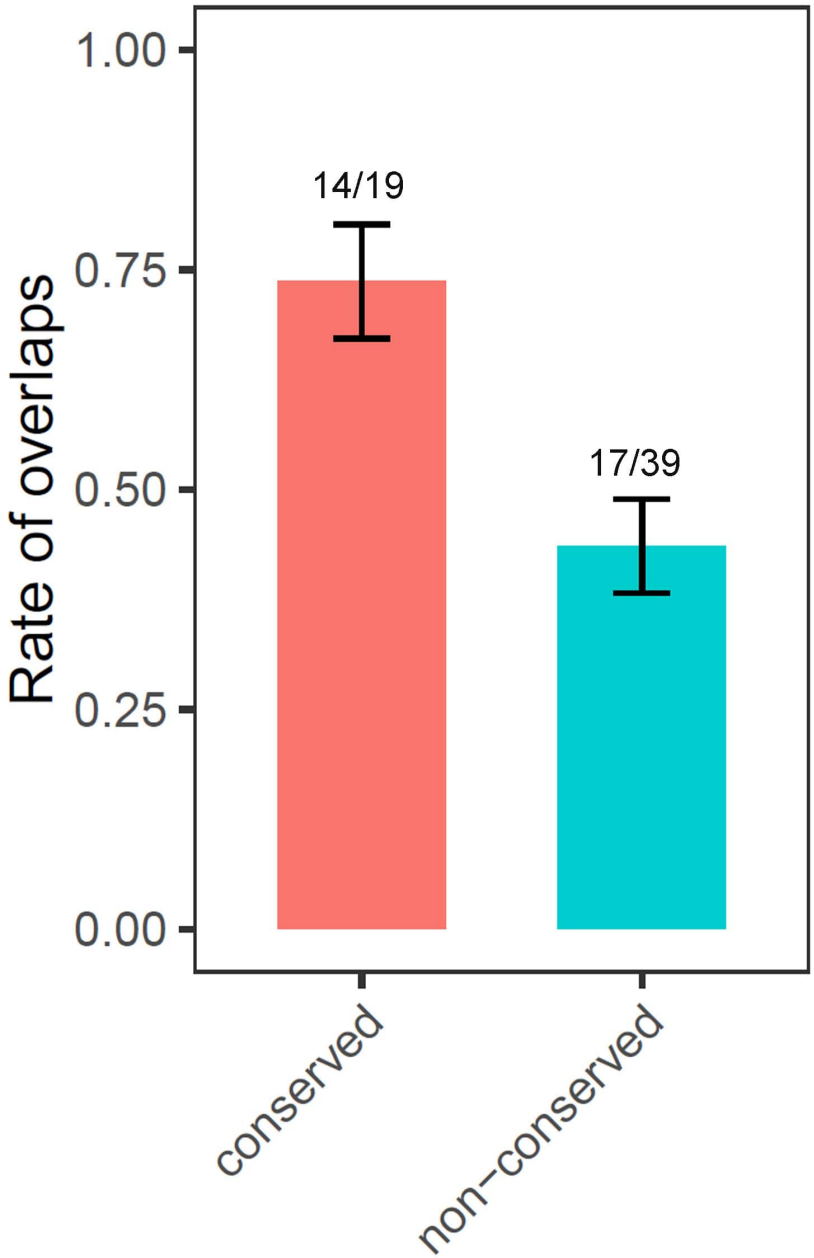


Fig. S8. The comparison in Fig. 4B is robust against trait measuring noise. To address the potential technical bias that traits with large measuring noise tend to be both non-conserved and non-overlapping, only traits with measuring CV < 0.1 across the replicates in wild-type BY4741 are considered. This results in 58 traits that are significantly affected by HAP4 deletion in S. cerevisiae, among which 19 are conserved effects and 39 non-conserved effects. The rate of overlaps in the conserved set remains significantly higher than the non-conserved set ($P = 0.029$, one-tailed Fisher's exact test). Overlaps refer to traits significantly affected by all four gene deletions in *S. cerevisiae*.