

# Defining genetic interaction

Mani *et al.* 10.1073/pnas.0712255105.

## Supporting Information

### Files in this Data Supplement:

- SI Figure 3
- SI Table 2 (/content/suppl/2008/02/21/0712255105.DC1/12255Table2.pdf)
- SI Table 3 (/content/suppl/2008/02/21/0712255105.DC1/12255Table3.pdf)
- SI Figure 4
- SI Table 4 (/content/suppl/2008/02/21/0712255105.DC1/12255Table4.pdf)
- SI Table 5 (/content/suppl/2008/02/21/0712255105.DC1/12255Table5.pdf)
- SI Table 6 (/content/suppl/2008/02/21/0712255105.DC1/12255Table6.pdf)
- SI Table 7 (/content/suppl/2008/02/21/0712255105.DC1/12255Table7.pdf)
- SI Table 8 (/content/suppl/2008/02/21/0712255105.DC1/12255Table8.pdf)
- SI Figure 5
- SI Figure 6
- SI Figure 7
- SI Figure 8
- SI Figure 9
- SI Figure 10
- SI Figure 11
- SI Figure 12
- SI Text

- SI Figure 3A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig3A.pdf)
- SI Figure 3B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig3B.pdf)

**Fig. 3.** Coefficients of variation for Study J data. Plot *a* shows a plot of coefficient of variation versus the mean for each single and double mutant that was studied in Study J. Plot *b* shows a region of plot *a* with a high density of points superimposed with a running median calculated with window size of 30.

- SI Figure 4A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig4A.pdf)
- SI Figure 4B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig4B.pdf)
- SI Figure 4C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig4C.pdf)
- SI Figure 4D (/content/suppl/2008/02/21/0712255105.DC1/12255Fig4D.pdf)

**Fig. 4.** Comparing distributions of  $\epsilon_{\text{Min}}$ ,  $\epsilon_{\text{Product}}$ ,  $\epsilon_{\text{Log}}$ , and  $\epsilon_{\text{Additive}}$  using data from Study J. (*a*) Pairs from Study J for which the single and double mutants exhibited slower than wild-type growth. (*b*) The subset of pairs from *a* for which the phenotypes of both single mutants was also  $>0.9$ . (*c*) The subset of pairs from *a* for which the phenotypes of both single mutants was  $<0.9$  but  $>0.75$ . (*d*) The subset of pairs for which phenotype of at least one of the single mutants was  $>0.75$ , and the phenotype of the other was  $>0.9$ .

- SI Figure 5A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig5A.pdf)
- SI Figure 5B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig5B.pdf)
- SI Figure 5C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig5C.pdf)

**Fig. 5.** Comparing distributions of  $\epsilon_{\text{Min}}$ ,  $\epsilon_{\text{Product}}$ ,  $\epsilon_{\text{Log}}$ , and  $\epsilon_{\text{Additive}}$ , using data from Study S in the absence of MMS. (*a*) All pairs from Study S (MMS-). (*b*) The subset of pairs from *a* for which the phenotypes of both single mutants was also  $>0.9$ . (*c*) The subset of pairs from *a* for which the phenotypes of both single mutants was  $<0.9$ .



(/content/current)

- Current Issue (/content/current)
- E-mail Alerts (/site/misc/PNASalerts.xhtml)
- Subscribe (/site/subscriptions/index.xh)
- RSS (/site/aboutpnas/rss.xhtml)

**Published online before print**  
**February 27, 2008, doi:**  
**10.1073/pnas.0712255105**  
PNAS (Proceedings of the National Academy of Sciences) **March 4, 2008** vol. 105 no. 9 **3461-3466**

## Access

- Abstract (/content/105/9/3461.abstract)
- Full Text (HTML) (/content/105/9/3461.full)
- Full Text (PDF) (/content/105/9/3461.full.pdf+html)
- Full Text + SI (Combined PDF) (/content/105/9/3461.full.pdf+html?with-ds=yes)
- Figures Only (/content/105/9/3461.figures-only)
- Supporting Information

## MOST READ    MOST CITED

1. Involvement of Toso in activation of monocytes, macrophages, and granulocytes (/cgi/content/short/110/7/2593?rss=1&ssource=mfr)
2. Higher social class predicts increased unethical behavior (/cgi/content/short/109/11/4086?rss=1&ssource=mfr)
3. Revised standards for statistical evidence (/cgi/content/short/110/48/19313?rss=1&ssource=mfr)
4. HSP72 protects against obesity-induced insulin resistance

SI Figure 6A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig6A.pdf)  
SI Figure 6B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig6B.pdf)  
SI Figure 6C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig6C.pdf)

**Fig. 6.** Comparing distributions of  $\epsilon_{\text{Min}}$ ,  $\epsilon_{\text{Product}}$ ,  $\epsilon_{\text{Log}}$ , and  $\epsilon_{\text{Additive}}$  using data from Study S in the presence of MMS. (a) All pairs from Study S. (b) The subset of pairs from a for which the phenotypes of both single mutants was  $<0.9$  but  $>0.75$ . (c) The subset of pairs from a for which phenotype of at least one of the single mutants was  $<0.75$ , and the phenotype of the other was  $<0.9$ .

SI Figure 7A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig7A.pdf)  
SI Figure 7B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig7B.pdf)  
SI Figure 7C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig7C.pdf)

**Fig. 7.** Comparing distributions of  $|\epsilon_{\text{Min}}|$ ,  $|\epsilon_{\text{Product}}|$ ,  $|\epsilon_{\text{Log}}|$ , and  $|\epsilon_{\text{Additive}}|$  using data from Study S in the absence of MMS. (a) All pairs from Study S (MMS-). (b) The subset of pairs from a for which the phenotypes of both single mutants was also  $>0.9$ . (c) The subset of pairs from a for which the phenotypes of both single mutants was  $<0.9$ .

SI Figure 8A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig8A.pdf)  
SI Figure 8B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig8B.pdf)  
SI Figure 8C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig8C.pdf)

**Fig. 8.** Comparing distributions of  $|\epsilon_{\text{Min}}|$ ,  $|\epsilon_{\text{Product}}|$ ,  $|\epsilon_{\text{Log}}|$ , and  $|\epsilon_{\text{Additive}}|$  using data from Study S in the presence of MMS. (a) All pairs from Study S. (b) The subset of pairs from a for which the phenotypes of both single mutants was  $<0.9$  but  $>0.75$ . (c) The subset of pairs from a for which phenotype of at least one of the single mutants was  $<0.75$ , and the phenotype of the other was  $<0.9$ .

SI Figure 9A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig9A.pdf)  
SI Figure 9B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig9B.pdf)  
SI Figure 9C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig9C.pdf)  
SI Figure 9D (/content/suppl/2008/02/21/0712255105.DC1/12255Fig9D.pdf)  
SI Figure 9E (/content/suppl/2008/02/21/0712255105.DC1/12255Fig9E.pdf)  
SI Figure 9F (/content/suppl/2008/02/21/0712255105.DC1/12255Fig9F.pdf)

**Fig. 9.** Differences in  $\epsilon$  between Product, Min, and Additive definitions. The six plots show differences in  $\epsilon$  between each pair of definitions for various single-mutant fitness values  $W_x$  and  $W_y$ : (a) Min and Product definitions, (b) Min and Log definitions, (c) Min and Additive definitions, (d) Product and Log definitions, (e) Product and Additive definitions, and (f) Log and Additive definitions. Dark plot regions indicate that the pair of definitions agree closely on the double-mutant fitness predicted for a non-interacting gene pair. These plots show that all four definitions produce identical results when at least one of the single mutants has wild-type fitness. They also show that Product and Log definitions yield practically equivalent results under all circumstances.

SI Figure 10 (/content/suppl/2008/02/21/0712255105.DC1/12255Fig10.pdf)

**Fig. 10.** Comparison of Product and Log definitions. This plot shows Fig. 7d with more resolution on the contours.

SI Figure 11A (/content/suppl/2008/02/21/0712255105.DC1/12255Fig11A.pdf)  
 SI Figure 11B (/content/suppl/2008/02/21/0712255105.DC1/12255Fig11B.pdf)  
 SI Figure 11C (/content/suppl/2008/02/21/0712255105.DC1/12255Fig11C.pdf)

**Fig. 11.** Distributions of  $\epsilon$  for interactions reported by Study T and Study P. Plots show how interactions identified in Study T (plot *a*) and Study P (plots *b* and *c*) map on to the corresponding  $\epsilon$  calculated for the Product definition in the Study S. Plot *b* shows only Study P<sub>severe</sub> interactions. With this restriction, both Study T and Study P<sub>severe</sub> (which use the Min definition) generally find more severe synthetic interactions that are also identified by the Product definition. Plot *c* shows only Study P<sub>slight</sub> interactions. Those interactions reported by Study P<sub>slight</sub> (which used the Min definition) and not by Study S<sub>Product</sub> are near  $\epsilon = 0$ , suggesting that the differences between these studies arise from the definition of genetic interaction.

SI Figure 12 (/content/suppl/2008/02/21/0712255105.DC1/12255Fig12.pdf)

**Fig. 12.** Comparison of number of synthetic interactions reported in Study P and Study T. Results either include (*Upper*) or exclude (*Lower*) the SF and SF-slight interaction classes in Study P that were most affected by interaction definition choice, and consider only gene pairs tested in both studies (see Table 1).

## SI Text

### Computing distributions of $\epsilon$

For subsets of gene pairs, medians and median absolute deviations (defined as the median of the absolute deviation of the samples from sample median) were computed using the "median" and "mad" programs in MATLAB (MathWorks). For subsets of pairs, the standard error on the median was computed using 10,000 bootstrapped samples obtained using the "bootstrap" program in MATLAB. Distributions of  $\epsilon$  were plotted by computing histograms with a bin width of 0.04.

### Measuring confidence intervals

Confidence intervals reported for confirmation and rejection rates were computed using the Clopper and Pearson exact method as implemented by the "binofit" program in MATLAB.

### Defining shared function or functional links

A set of additional studies used the methodology described in the main manuscript for defining functional links between genes (1-3).

### Deriving the Product definition

The Product definition is obtained by applying a multiplicative neutrality function to the fitness measure  $W$  defined for a strain as

$$m / m_{wt}$$

, where  $m$  is the exponential growth rate of that strain and  $m_{wt}$  is the exponential growth rate of the wild-type strain (4). For a pair of genes ( $x$ ,  $y$ ), fitness values are

$$\bar{W}_x \equiv m_x / m_{wt}$$

,

$$\bar{W}_y \equiv m_y / m_{wt}$$

, and

$$\bar{W}_{xy} \equiv m_{xy} / m_{wt}$$

for the two single mutants and double mutant, respectively. Here,

$$m_x$$

,

$$m_y$$

, and

$$m_{xy}$$

are the corresponding exponential growth rates. If  $x$  and  $y$  are noninteracting genes, then the multiplicative neutrality function used in these studies predicts that

$$W_w \cdot W_{xy} = W_x \cdot W_y$$

, where

$$W_w \equiv m_w / m_{wt} = 1$$

. Thus the Product definition is

$$W_{xy} = W_x \cdot W_y$$

.

### Deriving the Additive definition

The Additive definition is obtained by applying a multiplicative neutrality function to the fitness measure  $F_{exp}$  defined for a strain as

$$e^m$$

where  $m$  is the exponential growth rate of that strain (5, 6). For a pair of genes ( $x$ ,  $y$ ), fitness is

$$F_x = e^{m_x}$$

,

$$F_y = e^{m_y}$$

, and

$$F_{xy} = e^{m_{xy}}$$

respectively, for the two single mutants and double mutant. Here

$$m_x$$

,

$$m_y$$

, and

$$m_{xy}$$

are the corresponding exponential growth rates. If  $x$  and  $y$  are noninteracting genes, then the multiplicative neutrality function used in these studies predicts that

$$F_w \cdot F_{xy} = F_x \cdot F_y$$

where

$$F_w = e^{m_{wt}}$$

is the fitness of the wild-type strain. In terms of exponential growth rate  $m$ , the multiplicative expectation above becomes:

$$m_{wt} + m_{xy} = m_x + m_y$$

(6).

Now, applying the relative-growth-rate fitness definition,  $W$ , fitness values for strains  $x$ ,  $y$ , and  $xy$  are

$$W_x = \frac{m_x}{m_{wt}}$$

,

$$W_y = \frac{m_y}{m_{wt}}$$

, and

$$W_{xy} = \frac{m_{xy}}{m_{wt}}$$

. The expectation

$$m_{wt} + m_{xy} = m_x + m_y$$

may be expressed as:

$$W_{xy} = W_x + W_y - 1$$

. This equation forms the basis for the Additive definition. The corresponding quantitative measure of genetic interaction for this definition is  $\epsilon_{Additive} =$

$$W_{xy} - [W_x + W_y - 1]$$

### Deriving the Log definition

The Log definition is obtained by applying a multiplicative neutrality function to the fitness measure

$$F = (2^W - 1)$$

(7). For a pair of genes ( $x, y$ ), fitness measures are

$$F_x = (2^{W_x} - 1)$$

,

$$F_y = (2^{W_y} - 1)$$

, and

$$F_{xy} = (2^{W_{xy}} - 1)$$

respectively, for the two single mutants and double mutant. Here,  $W_x$ ,  $W_y$  and  $W_{xy}$  are the relative-growth-rate measures of fitness corresponding to the mutant strains. If  $x$  and  $y$  are non-interacting genes, then the multiplicative neutrality function predicts

$$F_{wt} \cdot F_{xy} = F_x \cdot F_y$$

. Using the fact that

$$W_{wt} = 1$$

, this reduces to

$$(2^{W_{xy}} - 1) = (2^{W_x} - 1) \cdot (2^{W_y} - 1)$$

, which in turn can be represented as

$$W_{xy} = \log_2 \left[ (2^{W_x} - 1)(2^{W_y} - 1) + 1 \right]$$

. The corresponding measure of genetic interaction for the Log definition is then  $\epsilon_{\text{Log}} =$

$$W_{xy} - \log_2 \left[ (2^{W_x} - 1)(2^{W_y} - 1) + 1 \right]$$

.

### Genetic interaction definitions used previously but not considered here

Some other fitness measures used in conjunction with the multiplicative neutrality function (8-11) have not been considered here, either due to our inability to replicate the fitness scoring steps carried out in the corresponding studies across the quantitative datasets we wanted to compare in our work).

### Bias shift for pairs involving extreme single-mutant fitness defects.

As described in the main text, we observed a positive shift in the distribution of  $\epsilon$  under each definition for gene pairs involving an extremely deleterious mutation, relative to the distribution (for the same definition) for pairs involving less severely deleterious mutations. We cannot rule out the possibility that definitions of genetic interaction truly have a shifted bias in the context of more deleterious mutations, as has been recently suggested (12). Another explanation is the preferential existence of compensatory mutations in extremely slow-growing mutant strains. This is supported by the fact that the positive shift in  $\epsilon$  bias we observed in Study J was more severe than that observed in Study S. Specifically, the more extreme positive shift in  $\epsilon$  for Study J (relative to Study S) under every genetic interaction definition may be explained by compensatory mutations that have arisen more frequently in Study J strains. Strains carrying compensatory mutations (e.g., aneuploidy) may overtake a population over time and affect measured growth rates, and this effect will be more pronounced for populations of doubly mutant strains with more severe fitness defects. Study S was less prone to compensatory mutations occurring either before, during, or after the meiosis combining the deletions. All of the parental strains with the nourseothricin resistance marker gene ( $\text{Nat}^r$ ) were freshly generated from wild type and would have had less opportunity for aneuploidy or other compensatory mutation than strains from the preexisting library containing strains with the kanamycin resistance marker gene ( $\text{Kan}^r$ ). Furthermore, growth rates for  $\text{Kan}^r$  and  $\text{Nat}^r$  strains corresponding to the same deleted gene were compared and where differences were observed, the  $\text{Kan}^r$  and/or  $\text{Nat}^r$  strain was rederived from wild type. If this did not correct the difference in growth rates, the corresponding gene was eliminated from further analyses. Thus, Study S was less prone to compensatory mutations occurring before the meiotic combination of two deletion alleles. Furthermore, Study S strains were generated in the absence of MMS. Because no strains in this Study S had severe growth defects in the absence of MMS, there was less selection for compensatory mutations both before and after the meiosis combining the two deletion alleles. In Study J, it was argued that compensatory mutations present before the meiosis would not lead to any bias since the compensatory mutations would randomly segregate to wild-type, single- and double-mutant strains. Accepting this argument, the positive shift in  $\epsilon$  observed for extreme relative to moderate mutations could still be due to compensatory mutations occurring after the meiosis. Alternatively, mutations that suppress the growth defect of a mutation with which it was coselected may not impact growth similarly in the absence of that mutation. Thus, the fact that the study with the greater positive bias in  $\epsilon$  was also the study more prone to compensatory mutation is consistent with (but by no means proves) the idea that compensatory mutations are causing the bias.

### Data

## Difference between Product and Log definitions

For a given pair of deleterious single mutations, the predictions for double mutant phenotype from the Product and Log definitions presented herein are practically indistinguishable. In SI Fig. 3, we demonstrate this fact using a heatmap of the absolute difference between the double mutant fitness predictions for the two definitions over all deleterious fitness values of  $W_x$  and  $W_y$ . In SI Fig. 10, we show the same plot but with the scale modified. It is clear from this plot the numerical difference between the two definitions has a peak at  $W_x = W_y = 0.5$ . The maximum difference in  $\epsilon$  between Product and Log definitions is 0.02155. Note that the difference between Log and Product definitions becomes significantly larger for advantageous mutations, so that Product and Log models are only practically equivalent for deleterious mutations.

## Overlap of Study J with Study T and Study P

Among the 45 pairs tested by both Study T and Study J is the pair *NUP60-CTF18* that is labeled as synthetic by Study T. This pair has negative  $\epsilon$  for all four definitions. Among the 56 pairs tested by both Study J and Study P are two interactions labeled as synthetic by Study P: *VAC14-CCR4* (interaction type SL/SF) and *RAD52-RML2* (interaction type SF). The value of  $\epsilon$  in Study J is negative for all four definitions for *VAC14-CCR4*, but for *RAD52-RML2* only the Min definition has a negative  $\epsilon$ .

1. Lee I, Date SV, Adai AT, Marcotte EM (2004) A probabilistic functional network of yeast genes. *Science* 306:1555-1558.
2. Wong SL, Zhang LV, Roth FP (2005) Discovering functional relationships: biochemistry versus genetics. *Trends Genet* 21:424-427.
3. Rual JF, *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437:1173-1178.
4. St. Onge RP, *et al.* (2007) Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat Genet* 39:199-206.
5. Szafraniec K, Wloch DM, Sliwa P, Borts RH, Korona R (2003) Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast *Saccharomyces cerevisiae*. *Genet Res* 82:19-31.
6. Jasnos L, Korona R (2007) Epistatic buffering of fitness loss in yeast double deletion strains. *Nat Genet* 39:550-554.
7. Sanjuan R, Elena SF (2006) Epistasis correlates to genomic complexity. *Proc Natl Acad Sci USA* 103:14402-14405.
8. Hartman JL, Tippery NP (2004) Systematic quantification of gene interactions by phenotypic array analysis. *Genome Biol* 5:R49.
9. Schuldiner M, *et al.* (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123:507-519.
10. Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol* 7:R63.
11. Collins SR, *et al.* (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446:806-810.
12. Beerenwinkel N, *et al.* (2007) Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol Biol* 7:60.