Supporting Information

Kampmann et al. 10.1073/pnas.1307002110

Definition of Buffering and Synergistic Genetic Interactions

High-density genetic interaction (GI) maps not only reveal functional groups of genes based on the correlation of their GI patterns but also comprehensively quantify GIs, which can be interpreted directly to gain insight into the nature of the relationship between genes, and, ideally, to reconstruct entire pathways (1, 2). One classical example of an interpretable GI is the case of two genes that act in parallel pathways and partially compensate for each other's loss. Depletion of either gene product will have a moderate effect; depletion of both will have a much stronger effect, which typically is referred to as "synergistic" or "synthetic sick/synthetic lethal" GI. The opposite type of GI is characteristic of genes acting in a linear pathway: Depletion of either gene product interferes with the pathway and causes a given phenotype. In combination, depletion of both gene products together has no additive effect on the phenotype; such an interaction is referred to as a "buffering" GI. Genes encoding subunits of a physical complex often are connected by one type of GI, either buffering or synergistic, a phenomenon referred to as "monochromaticity" (3).

In the case of GIs between genes whose knockdowns have deleterious effects ("negative" phenotypes), positive GIs are buffering and negative GIs are synergistic. Conversely, in the case of GIs between genes whose knockdowns have beneficial effects ("positive" phenotypes), negative GIs are buffering and positive GIs are synergistic (4). GIs between genes of mixed phenotypes or with paradoxical double-mutant phenotypes [sometimes referred to as "sign epistasis" (5)] are more difficult to interpret. A qualitative classification of different cases of GIs has been proposed (6), but to our knowledge, a method for mapping quantitative GIs between mixed-phenotype genes onto a continuum of synergistic to buffering GIs has not previously been developed.

"Raw" GIs generally are defined as follows:

$$\begin{aligned}\text{GI} = {} & \text{Observed double-shRNA phenotype} \\ & - \text{Expected double-shRNA phenotype}.\end{aligned}$$

We explored two possible definitions for synergistic and buffering GIs that differ in their interpretation of sign epistasis (Fig. S4):

Buffering/synergistic GI definition 1:

$$\begin{aligned}\text{Buffering GI} = {} & \text{sign}(\text{Expected double-shRNA phenotype}) \\ & \times (\text{Expected double-shRNA phenotype} \\ & - \text{Observed double-shRNA phenotype}).\end{aligned}$$

Buffering/synergistic GI definition 2:

$$\begin{aligned}\text{Buffering GI} = {} & |\text{Expected double-shRNA phenotype}| \\ & - |\text{Observed double-shRNA phenotype}|.\end{aligned}$$

To evaluate whether these definitions are biologically meaningful, we determined the distribution of buffering and synergistic GIs between shRNAs targeting the same gene, genes encoding subunits of the same complex, and other shRNAs (Fig. S4). For both definitions, shRNAs targeting the same gene or subunits of complexes generally were connected by buffering GIs, whereas the distribution of GIs for other shRNAs was centered around 0 (Fig. S4). We investigated whether clustering genes according to the correlation of buffering/synergistic GIs, as opposed to raw GIs, would improve the clustering of biologically meaningful groups of genes, but this was not the case for our dataset. Therefore, we created GI maps by clustering genes based on raw GIs but colored them using a heatmap based on buffering/synergistic GIs (according to definition 2) to make individual GIs interpretable.

1. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9(11):855–867.
2. Battle A, Jonikas MC, Walter P, Weissman JS, Koller D (2010) Automated identification of pathways from quantitative genetic interaction data. *Mol Syst Biol* 6:379.
3. Segrè D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. *Nat Genet* 37(1):77–83.
4. Phillips PC, Otto SP, Whitlock MC (2000) Beyond the average - the evolutionary importance of gene interactions and variability of epistatic effects. Epistasis and the Evolutionary Process, eds Wolf JB, Brodie ED, Wade MJ (Oxford Univ Press, Oxford, UK).
5. Weinreich DM, Watson RA, Chao L (2005) Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59(6):1165–1174.
6. Drees BL, et al. (2005) Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol* 6(4):R38.

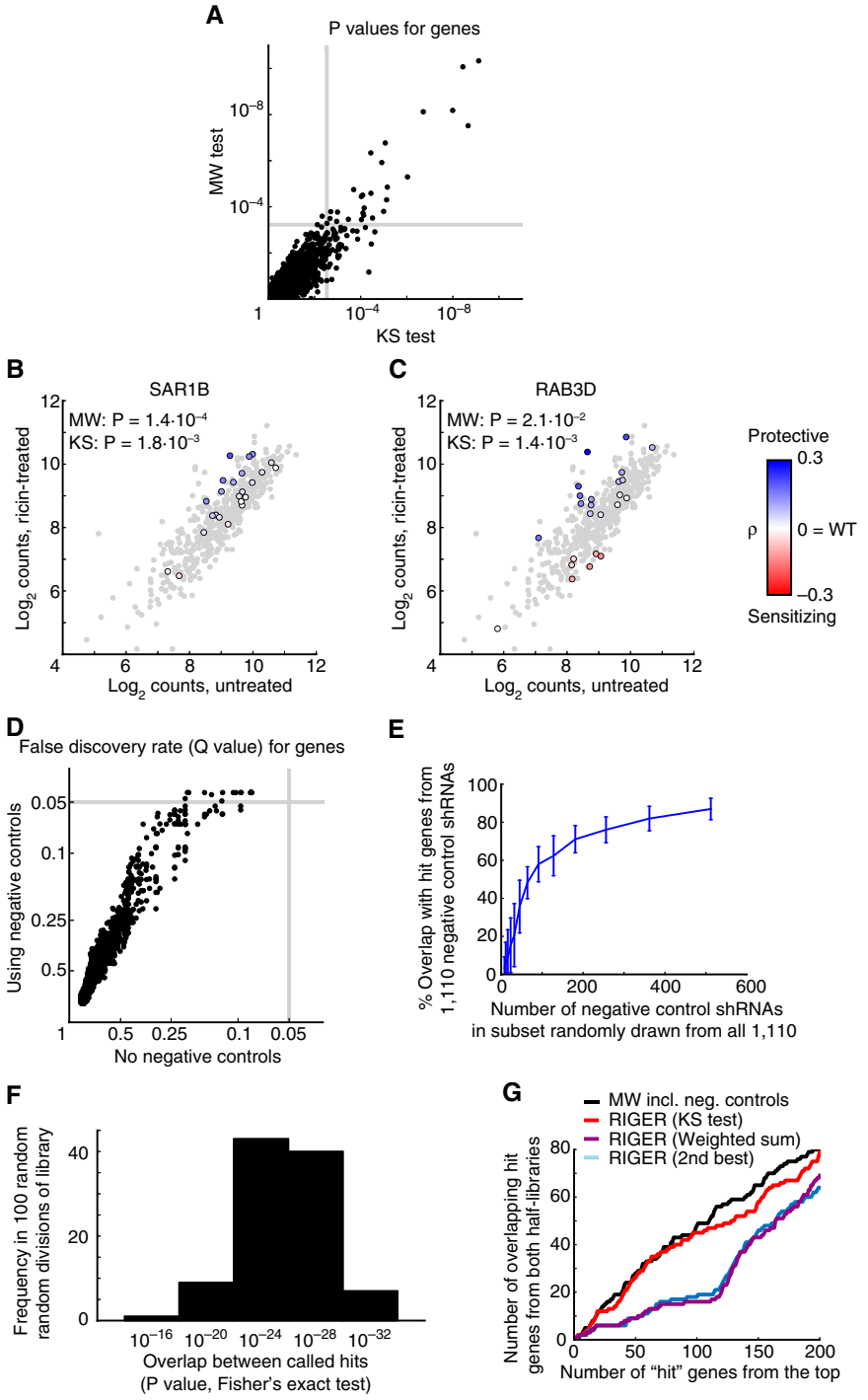

Fig. S1. Statistical testing for primary screen hits. (*A*) *P* values for genes targeted by a subgenomic shRNA library were calculated by comparing ricin resistance phenotypes of shRNAs targeting each gene with the phenotypes of negative control shRNAs. *P* values calculated using the Mann–Whitney *U* test (MW test) show good overall agreement with *P* values calculated using the two-sample Kolmogorov–Smirnow test (KS test). Gray lines indicate a 5% false-discovery rate (FDR). (*B* and *C*) Atypical examples for shRNA phenotype distributions resulting in divergent *P* values from the MW and KS tests are presented. (*B*) The MW test ranked genes highly for which phenotypes of all shRNAs were shifted consistently. (*C*) The KS test ranked genes highly for which many shRNAs had extreme phenotypes, even if other shRNAs had strong opposite phenotypes. (*D*) Illustration of the advantage of using separate negative-control shRNAs. For a subgenomic pilot screen for ricin resistance, *P* values were calculated by comparing shRNAs targeting each gene either with negative control shRNAs or with the distribution of all shRNAs in the library. Q values representing the FDR were calculated to correct for multiple hypothesis testing. When negative controls are used, higher Q values are obtained, and a number of hits are significant at an FDR below 5% (indicated by gray lines). (*E*) Hit genes (defined as genes with FDR <5%) in a subgenomic test library were determined by the MW test either using all 1,110 negative control shRNAs present in the test library or using random subsets of negative control shRNAs with varying sizes. The graph shows the percent overlap in hit genes, defined as 100 × (number of shared hit genes/number of hit genes called by at least one of the tests). Random subset generation was repeated 100 times for each subset size; average values ± SD are shown. (*F*) Random division of shRNAs in a test library, as in Fig. 3*B*, was repeated 100 times, and this histogram shows the distribution of *P* values for overlap between called hits from the two half-libraries (Fisher's exact test). (*G*) Different hit-calling methods are compared on two random half-libraries (as in Fig. 3*B*). Genes

Legend continued on following page

were ranked using our approach (the MW test and negative controls) or different versions of the RIGER algorithm (www.broadinstitute.org/cancer/software/GENE-E/). The number of overlapping hit genes from two half-libraries is shown for hit gene sets of different sizes.

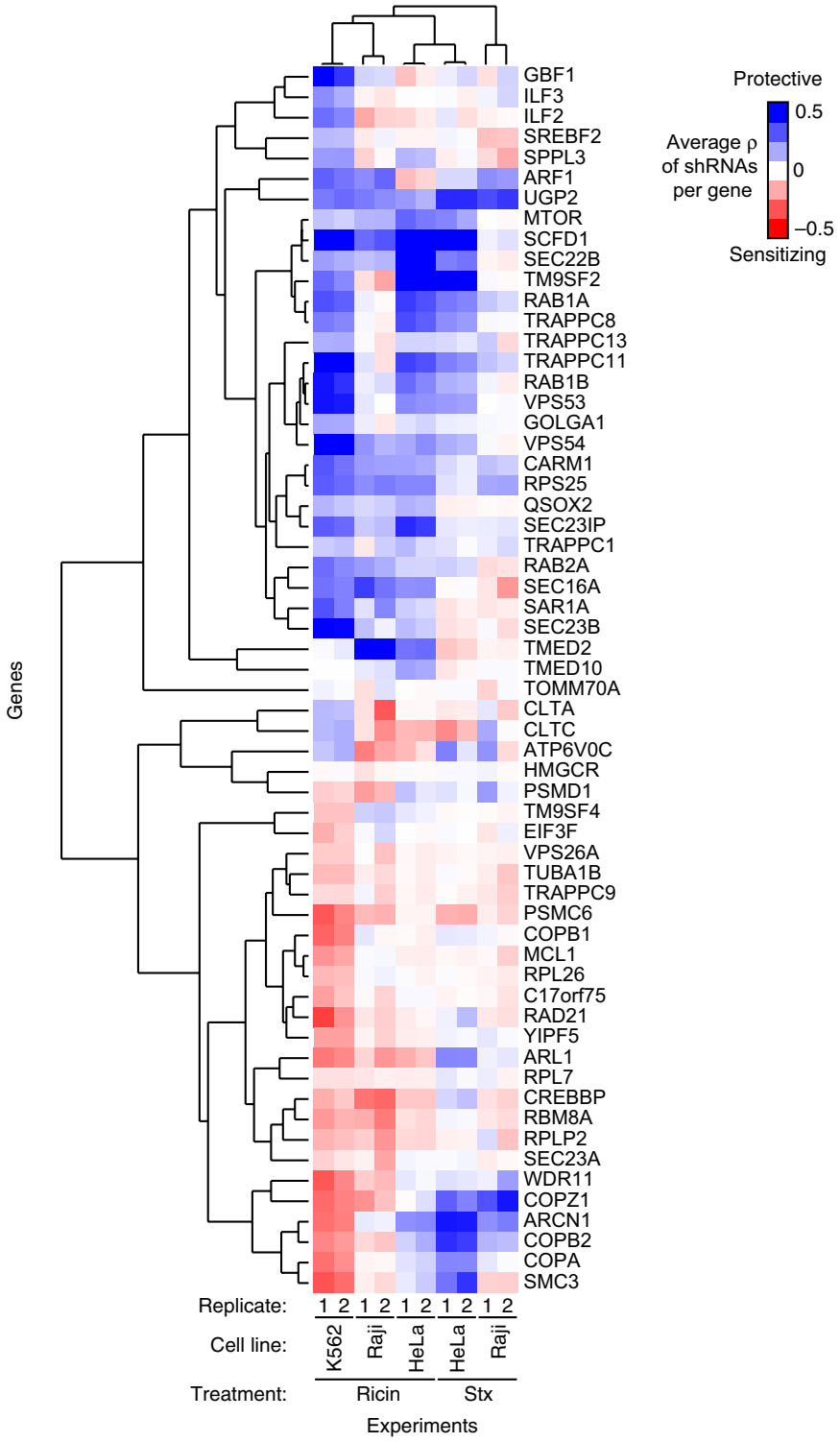

Fig. S2. Alternative presentation of the data from Fig. 4*B*. Averaged ricin resistance phenotypes (ρ) per gene and experimental replicates are displayed separately.

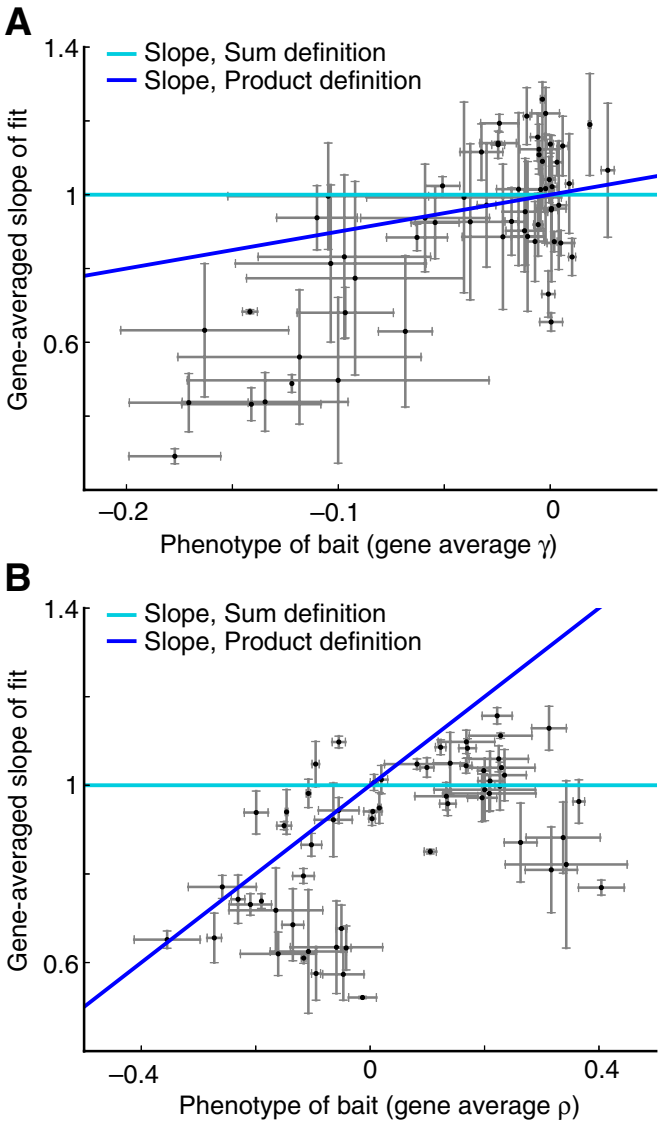

Fig. S3. Definitions of the expected double-shRNA phenotypes. The relationship between single-shRNA phenotypes and double-shRNA phenotypes for growth and ricin resistance in K562 cells was fitted linearly as in Fig. 5*A* for all shRNAs ("baits") in the double-shRNA library. Slopes of these linear fits are plotted as a function of the single-shRNA phenotype of the bait; dots and error bars denote averages and SD, respectively, of shRNAs targeting the same gene. The slopes obtained by the sum of product definitions for expected double-shRNA phenotypes are shown as light and dark blue lines, respectively. (*A*) Double-shRNA phenotypes for growth-based screen. (*B*) Double-shRNA phenotypes for ricin resistance-based screen.

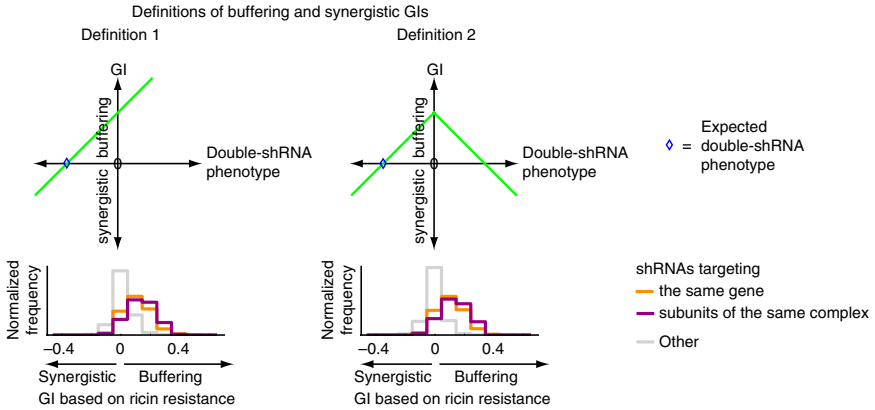

Fig. S4. Defining buffering and synergistic GIs. (*Upper*) Two possible definitions (green lines) as a function of double-shRNA phenotypes and expected double-shRNA phenotype (blue diamond). (*Lower*) Distribution of GIs according to these two definitions between shRNAs targeting the same gene (orange), subunits of the same complex (purple), or other pairs of shRNAs (gray).

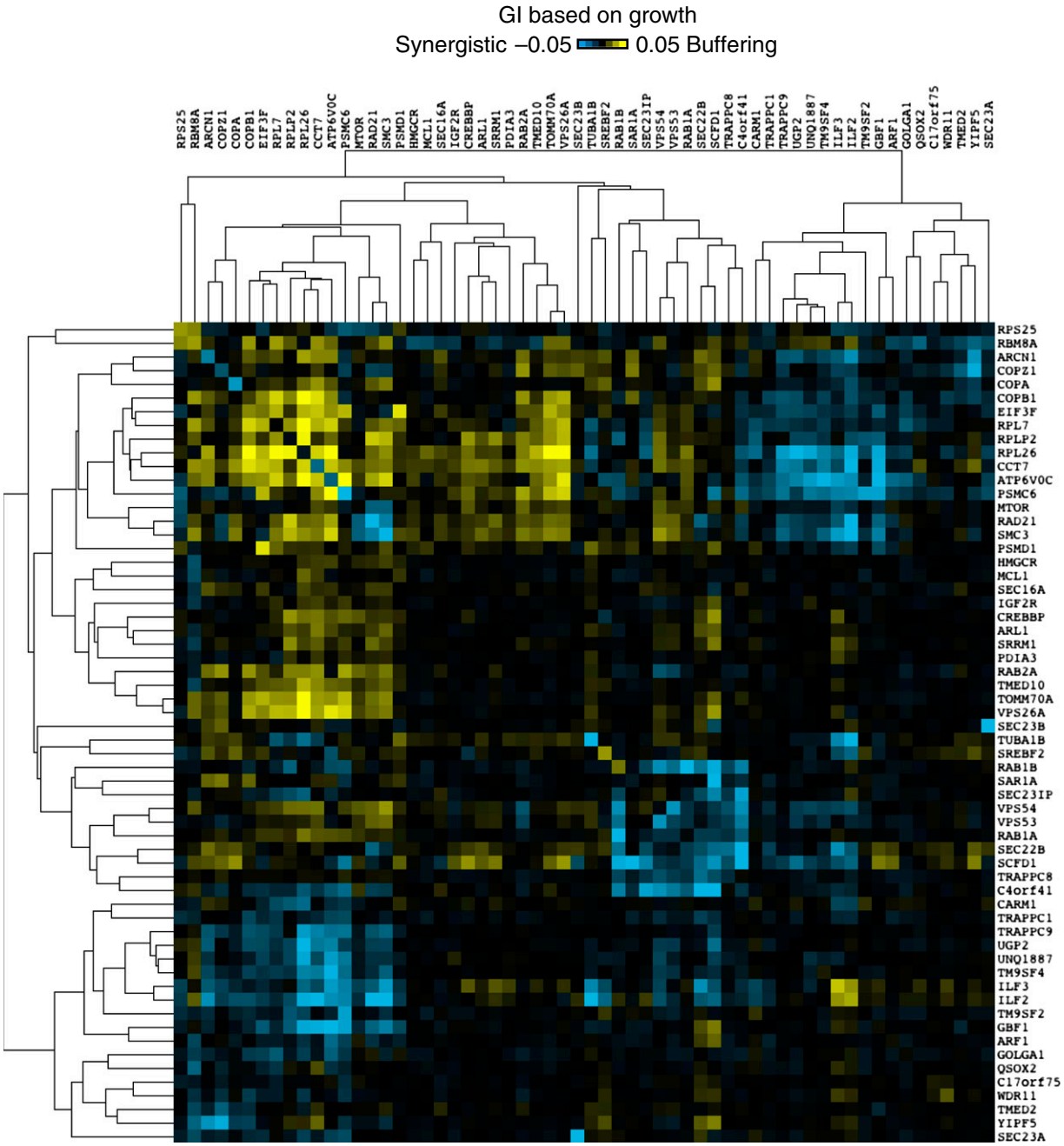

Fig. S5. GI map based on growth phenotypes. This figure is a version of Fig. 6*A* labeled with gene names.

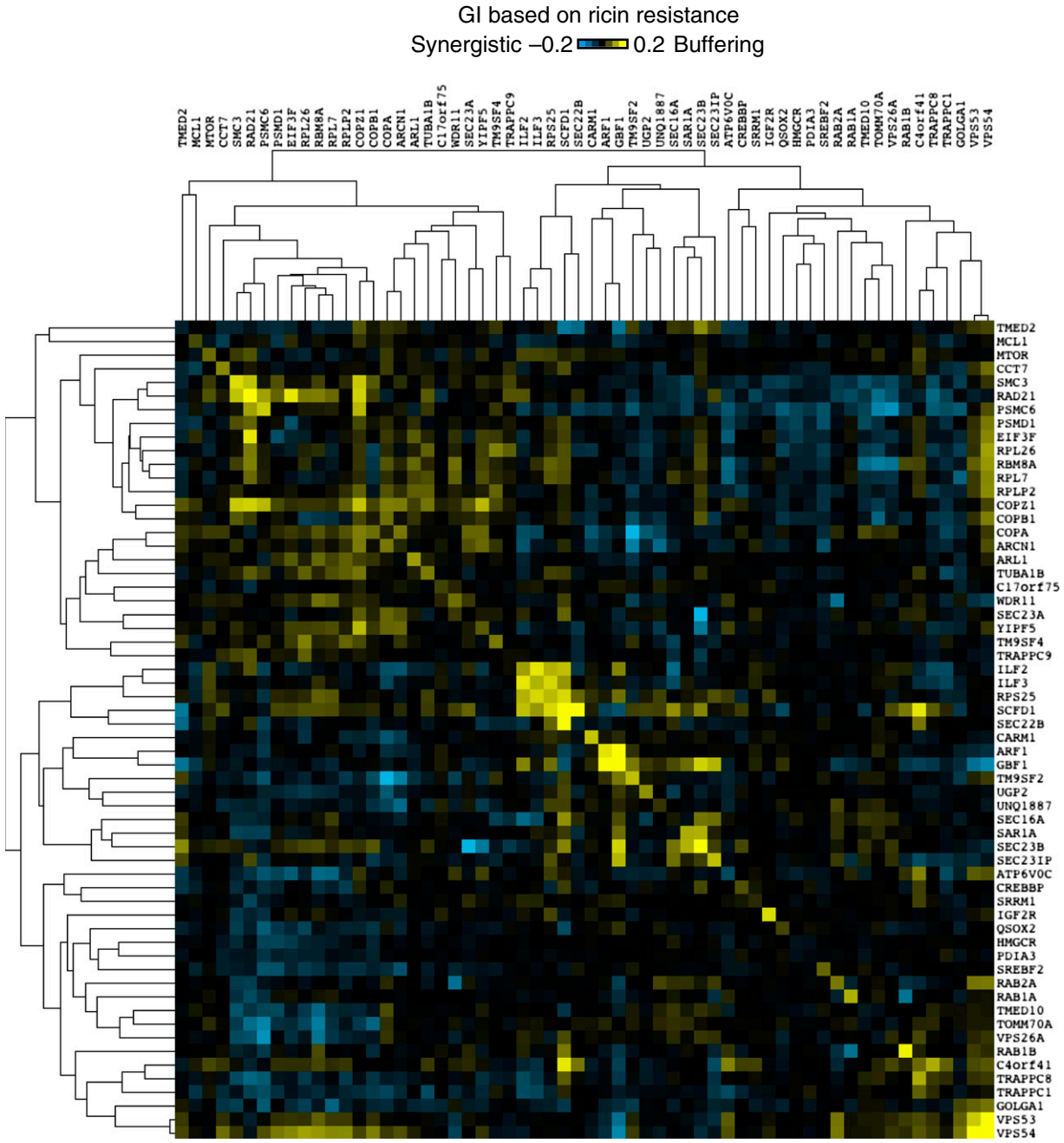

Fig. S6. Genetic interaction map based on ricin resistance phenotypes. This figure is a version of Fig. 6*B* labeled with gene names.

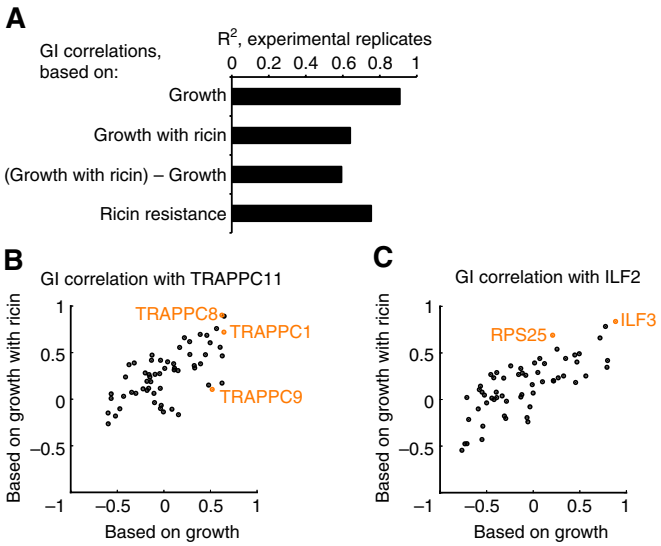

Fig. S7. Differential GI maps. (*A*) GIs were calculated for the same set of shRNAs based on growth (under standard conditions), growth in the presence of ricin, ricin resistance, or by subtracting growth GIs from growth with ricin GIs. R^2 quantifies reproducibility between independent experimental replicates. (*B* and *C*) As in Fig. 6 *C* and *D*, respectively, but comparing GI correlation based on growth with GI correlation based on growth in the presence of ricin.

Table S1. Overlap of hits detected based on independent shRNA half-libraries with different algorithms

Method	No. of hits shared between half-libraries	No. of hits only in half-library 1	No. of hits only in half-library 2	No. of non-hits shared between half-libraries	*P* value (Fisher's exact test)	Overlap of hit genes, %
Mann–Whitney *U* test using negative controls	28	23	22	1,006	$5.6{\cdot}10^{-28}$	38.4
RIGER using KS test	41	53	53	932	$1.2{\cdot}10^{-22}$	27.9
RIGER using second best rank	22	46	53	958	$5.4{\cdot}10^{-11}$	18.2
RIGER using weighted sum	20	46	51	962	$6.5{\cdot}10^{-10}$	17.1

An FDR cutoff of 5% is used to define hits. For RIGER output, which lists one-tailed FDR as *P* values, enriched and depleted hits up to an FDR cutoff of 2.5% were added.