1 **SUPPLEMENTARY INFORMATION**

2 **Functional Annotation of Chemical Libraries across Diverse Biological Processes**

3 Jeff S. Piotrowski*[1,2], Sheena C. Li*[1], Raamesh Deshpande*[3], Scott W. Simpkins*[4], Justin
4 Nelson[4], Yoko Yashiroda[1], Jacqueline M. Barber[1], Hamid Safizadeh[3,5], Erin Wilson[3], Hiroki
5 Okada[6], Abraham A. Gebre[6], Karen Kubo[6], Nikko P. Torres[7], Marissa A. LeBlanc[1], Kerry
6 Andrusiak[7], Reika Okamoto[1], Mami Yoshimura[1], Eva DeRango-Adem[7], Jolanda van Leeuwen[7],
7 Katsuhiko Shirahige[8], Anastasia Baryshnikova[9, 10], Grant W. Brown[7,11], Hiroyuki Hirano[1],
8 Michael Costanzo[7], Brenda Andrews[7], Yoshikazu Ohya[6], Hiroyuki Osada[1§], Minoru Yoshida[1§],
9 Chad L. Myers[3,4§], Charles Boone[1,7§]

10

11 1. RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan
12 2. Yumanity Therapeutics, Cambridge, MA, USA
13 3. University of Minnesota-Twin Cities, Department of Computer Science and Engineering,
14 Minneapolis, Minnesota, USA
15 4. University of Minnesota-Twin Cities, Bioinformatics and Computational Biology Program,
16 Minneapolis, Minnesota, USA
17 5. University of Minnesota-Twin Cities, Department of Electrical and Computer Engineering,
18 Minneapolis, Minnesota, USA
19 6. University of Tokyo, Department of Integrated Biosciences, Graduate School of Frontier
20 Sciences, Kashiwa, Chiba, Japan
21 7. University of Toronto, Donnelly Centre, Toronto, Ontario, Canada
22 8. The University of Tokyo, Institute of Molecular and Cellular Biosciences, Center for
23 Epigenetic Disease, Laboratory of Genome Structure and Function, Yayoi 1-1-1, Bunkyo-ku,
24 Tokyo 113-0032, Japan
25 9. Princeton University, Lewis-Sigler Institute for Integrative Genomics, Princeton, NJ 08544,
26 USA
27 10. Calico Life Sciences, South San Francisco, CA 94080, USA
28 11. University of Toronto, Department of Biochemistry, Toronto, Ontario, Canada
29
30

31 *Authors contributed equally to this work
32 § Correspondence to charlie.boone@utoronto.ca, chadm@umn.edu,
33 osadahiro@riken.jp, yoshidam@riken.jp
34

SUPPLEMENTARY RESULTS

Supplementary Table 1. Comparison of chemical-genetic interaction of common compounds between this dataset, Lee et al. 2014, and Hoepfner et al. 2014.

Piotrowski Drug names	Hoepfer Drug names	Giaever Drug names	Piotrowski number drugs	Hoepfner number drugs	Giaever number drugs	Piotrowski-Hoepfner Pearson correlation	Piotrowski-Giaever Pearson correlation	Hoepfner Giaever Pearson correlation	Threeway Pearson correlation
Fluconazole, CPD000471882	1129	SGTC_1787, SGTC_1788, SGTC_227	1	2	3	0.293	0.243	0.064	0.200
Nocodazole	2390	SGTC_1875	1	3	1	0.681	0.721	0.615	0.672
Benomyl	991	SGTC_229	1	97	1	0.704	0.824	0.671	0.733
Chlorpromazine	956	SGTC_2728	1	3	1	0.243	−0.054	−0.009	0.060
MMS	2878	SGTC_915	1	1	1	0.579	0.338	0.347	0.422
NSC19893, CPD000038082	1119	SGTC_242, SGTC_423	2	1	2	0.137	0.046	0.042	0.075
5-Fluorocytosine, CPD000059047	3331	SGTC_1077, SGTC_1698	1	1	2	0.279	0.092	0.026	0.132
Caffeine	1080	SGTC_228	1	4	1	0.446	0.236	0.154	0.279
Hydroxyurea, NSC32065	1089	SGTC_273	3	4	1	0.070	0.155	0.103	0.109

35

36

37 **Supplementary Table 2.** Top 20 compounds with distinct dual-target predictions

| Compound	Prediction coherence	Target process #1	Target process #2
NPD7992	−0.13	cell wall organization or biogenesis	negative regulation of transcription from RNA polymerase I promoter
NPE593	−0.11	Golgi to plasma membrane transport	cellular macromolecule localization
NPD6955	−0.1	receptor recycling	endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA)
NPD887	−0.05	proteasomal ubiquitin-independent protein catabolic process	hydrogen transport
NPE738	−0.02	meiotic DNA double-strand break processing	fungal-type cell wall biogenesis
NPD7879	−0.01	DNA catabolic process	fungal-type cell wall biogenesis
NPE164	0	telomere maintenance	fungal-type cell wall biogenesis
NPD5925	0	DNA catabolic process	fungal-type cell wall biogenesis
NPE614	0	fungal-type cell wall biogenesis	DNA catabolic process
Desipramine	0.01	RNA polymerase II transcriptional preinitiation complex assembly	retrograde vesicle-mediated transport, Golgi to ER
NPD7371	0.02	positive regulation of cytoskeleton organization	phospholipid transport
NPE81	0.03	energy coupled proton transmembrane transport, against electrochemical gradient	DNA-dependent transcriptional preinitiation complex assembly
NPD1256	0.03	cytokinesis	vesicle targeting
NPD6024	0.04	meiotic DNA double-strand break processing	response to metal ion
NPD5954	0.04	nuclear migration	RNA export from nucleus
NPE1081	0.06	fungal-type cell wall biogenesis	response to metal ion
NPD838	0.06	establishment of protein localization to membrane	organelle localization
NPD3577	0.06	nuclear-transcribed mRNA poly(A) tail shortening	DNA catabolic process
NP214	0.07	energy coupled proton transmembrane transport, against electrochemical gradient	DNA replication
NPD401	0.07	glycosylation	cytoplasmic translation

38

39

40

41 **Supplementary Table 3.** Compounds with described modes-of-action and targets in the high-confidence
42 predictions set

| Drug	P-value	GO Process	GO rank	Top 5 targets
MMS	$< 2 \times 10^{-5}$	DNA metabolic process	1	TSA1, RRM3, RNR4, RNH203, MMS4
Benomyl	$< 2 \times 10^{-5}$	tubulin complex assembly	1	TUB3, GIM3, GIM4, YKE2, CIN1
Nocodazole	$< 2 \times 10^{-5}$	tubulin complex assembly	1	TUB3, GIM3, YKE2, GIM4, CIN1
Latrunculin B	$< 2 \times 10^{-5}$	cellular component movement	1	PFY1, MYO5, TAF1, BIM1, CIN8
FK228	$< 2 \times 10^{-5}$	negative regulation of chromatin silencing	1	HCS1, YPL150W, SSU72, SIN3, RPD3
Hedamycin	$< 2 \times 10^{-5}$	DNA metabolic process	1	POL3, RNR4, TSA1, RAD27, RFA3
Mycophenolic acid	$< 2 \times 10^{-5}$	DNA metabolic process	1	RAD55, RAD57, RNR4, AFT1, YPL077C
Trichostatin A	$< 2 \times 10^{-5}$	negative regulation of chromatin silencing	3	TRS33, SSU72, GRH1, YPL150W, HCS1
Tunicamycin	$< 2 \times 10^{-5}$	glycosylation	1	ALG14, ALG5, YIL102C, ERI1, ALG13
Micafungin	$< 2 \times 10^{-5}$	cellular component movement	1	FKS1, PFY1, ARC18, YNL181W, ILM1
Furazolidone	$< 2 \times 10^{-5}$	Maintenance of fidelity in DNA replication	1	MMS4, RNR4, SLX1, MUS81, RNH203
Brefeldin A	$< 2 \times 10^{-5}$	RNA transport	1	HIR2, TAF7, MED1, THP2, SUB2
Acriflavine	$< 2 \times 10^{-5}$	base-excision repair	1	RFA2, RAD27, RFC4, HOM6, POL32
Haloperidol	$< 2 \times 10^{-5}$	homoserine biosynthetic process	1	HOM2, ERG25, UBP3, HOM3, BRE5
Aclacinomycin A	$< 2 \times 10^{-5}$	DNA conformation change	2	NUP60, NUP84, SUB2, TLD3, TOP1
Cinerubin B	$< 2 \times 10^{-5}$	DNA packaging	1	FAL1, SLD3, PDS1, SCC4, LSR4
Rapamycin	$< 2 \times 10^{-5}$	energy coupled proton transmembrane transport	1	CNB1, NBP2, SWA2, FLC2, VMA11
Daunorubicin	$< 2 \times 10^{-5}$	Ubiquitin-dependent macromolecule catabolic process	2	RPT2, SPT3, UBP6, RPN11, UBX4
Hydroxyurea	$< 2 \times 10^{-5}$	DNA conformation change	2	TOP1, TAF7, TAF1, WSS1, MMS4
5-Fluorocytosine	$< 2 \times 10^{-5}$	tRNA modification	1	SWC4, ADA2, CDC7, YTA7, DBF4
Camptothecin	4×10^{-5}	DNA geometric change	3	SCC2, TCP1, POL2, DCC1, KTI11
Caffeine	4×10^{-5}	TOR signaling cascade	1	TOR1, YMR018W, STE24, KOG1, KRS1
Tyrocidin B	4×10^{-5}	regulation of pH	1	GAS1, SWA2, KRE1, KRE5, BIG1
Blasticidin S	6×10^{-5}	transcription from RNA polymerase I promoter	1	VPS28, VPS25, BTS1, YGR012W, SNM1
Fluconazole	8×10^{-5}	mitotic sister chromatid cohesion	1	POC4, PRP16, RPN10, HDA3, YPL150W
Itraconazole	1.8×10^{-4}	steroid biosynthetic process	16	ERG3, RPN10, EMC6, POC4, ERG25
Cisplatin	2.4×10^{-4}	DNA replication	3	SLX1, WSS1, MMS4, RNH202, RNH203
Podophyllotoxin	2.4×10^{-4}	RNA splicing	1	MED1, MRP7, BIM1, SLU7, YHC1
Bortezomib		No Prediction		IRC25, POC4, EMC6, SEM1, UBX4
Nigericin		No Prediction		NUP84, SEC27, SED5, COG6, ARL1
OligomycinA		No Prediction		PDB1, POP7, MET30, MOB2, RIM8
Griseofluvin		No Prediction		SPF1, MNN2, YND1, SSS1, ERG4
Polyoxin D		No Prediction		YTA7, IPK1, BIM1, RSE1, MPS1

43

44

45
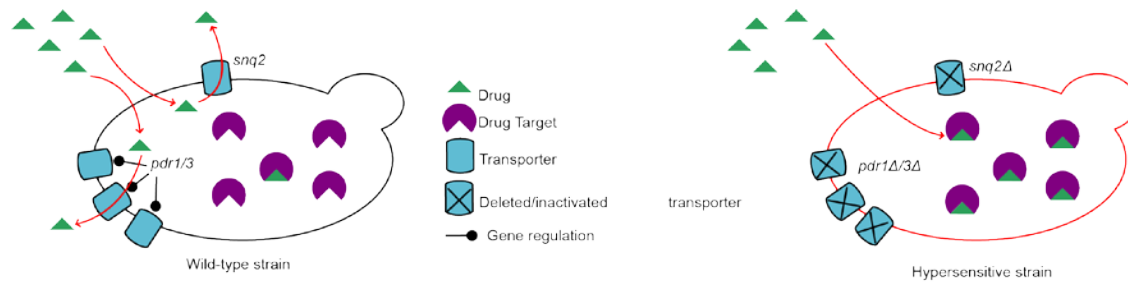
Supplementary Figure 1. Constructing a drug sensitized yeast strain. To construct a drug sensitized yeast strain for chemical genomics assays, we deleted the transcription factors Pdr1p and Pdr3p, which control much of the yeast pleiotropic drug response as well as the multidrug transporter, Snq2p, in strain Y7092, which encodes markers and reporters necessary for SGA analysis.

51
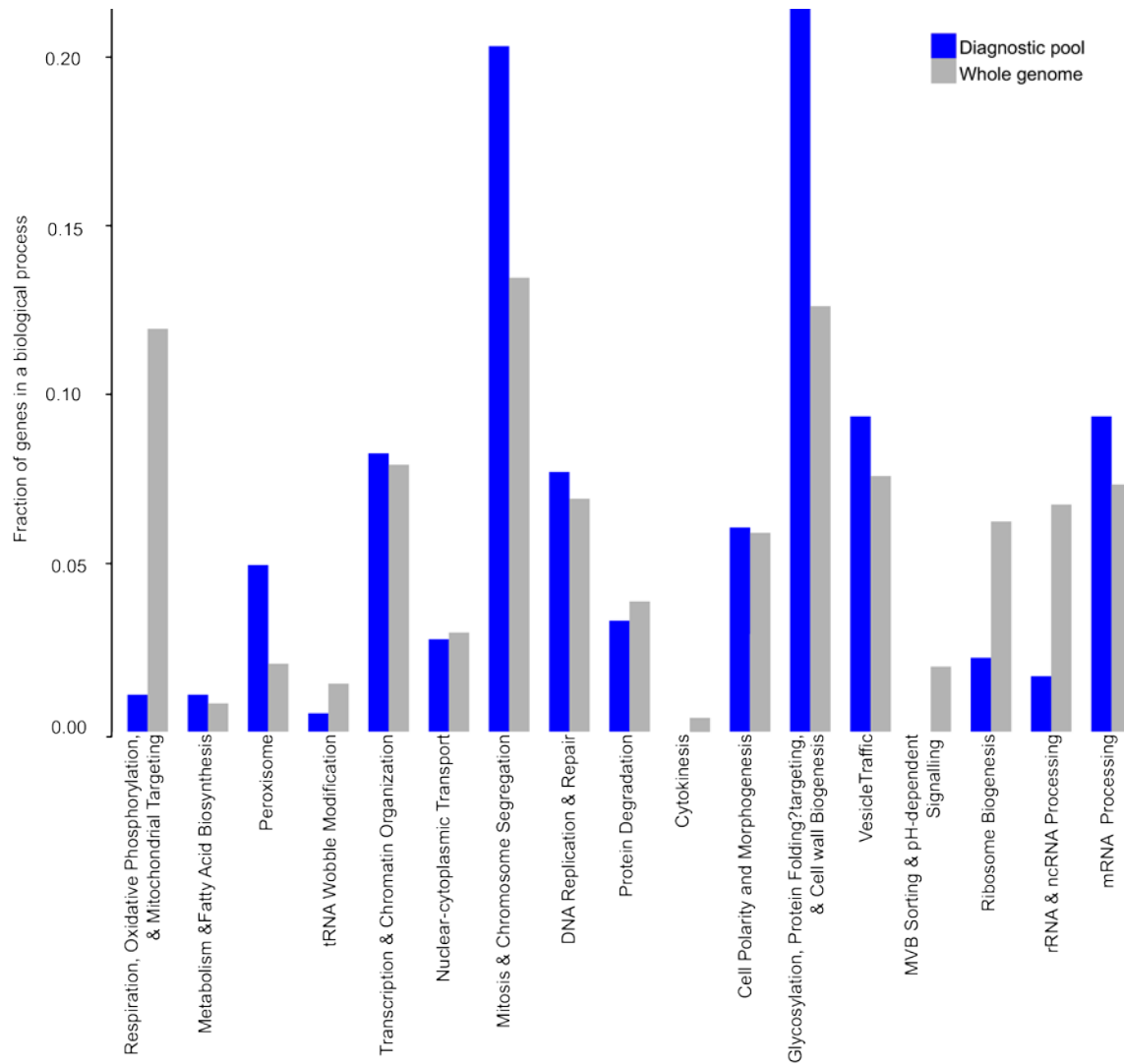
Supplementary Figure 2. Functional distribution of genes in the diagnostic screening collection. Distribution of genes comprising the diagnostic set (blue) compared to that of genes in the complete genome-wide deletion collection (grey) across the 17 major bioprocesses of the cell.
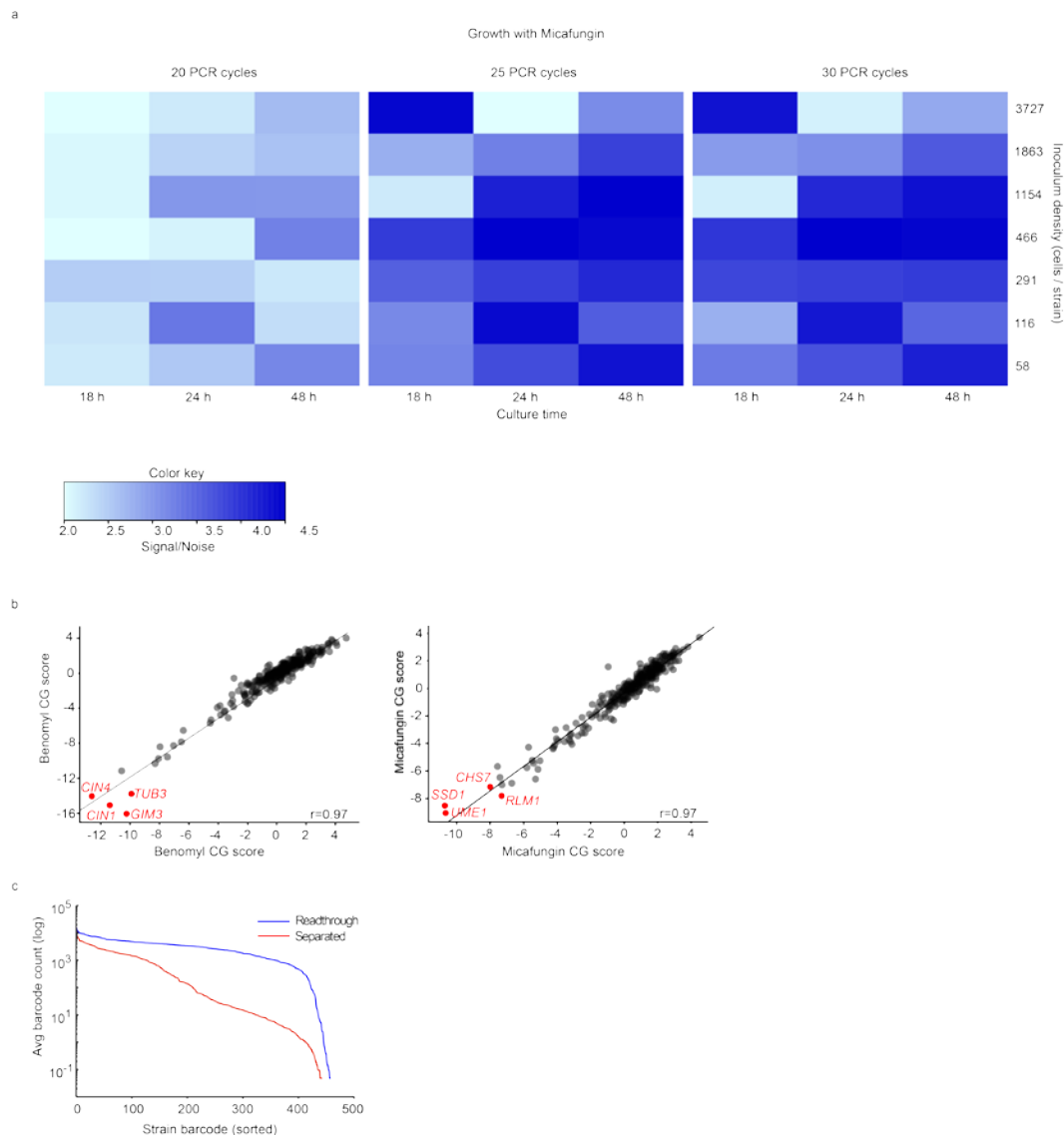
Supplementary Figure 3. Optimizing detection of chemical-genetic interactions. (a) The effect of culture time, inoculum density, and PCR cycle number on the signal-to-noise ratio within the chemical genomic profile of micafungin (25 nM). Darker blue indicates improved signal detection of the top sensitive strains relative to the entire pool. (b) Correlation of independent, replicate chemical genomic profiles for benomyl and micafungin. (c) Total barcode read count yield when sequencing of the amplicon containing the multiplex tag barcode, and mutant strain barcode is done in a single priming step (Read-through), versus yield when sequencing is done with separate priming steps for the multiplex tag barcode and mutant strain barcode (Separated).
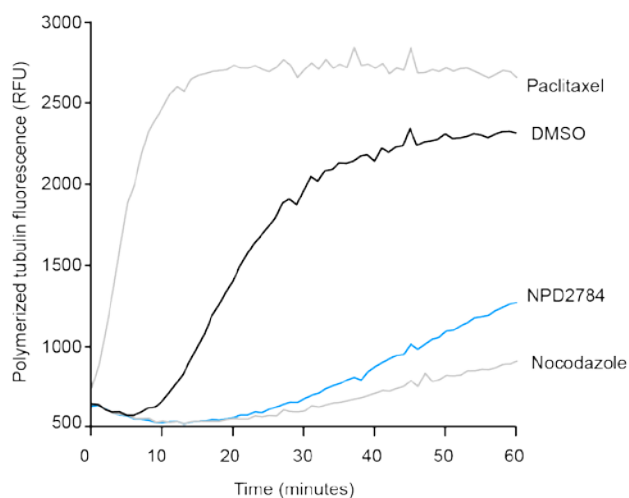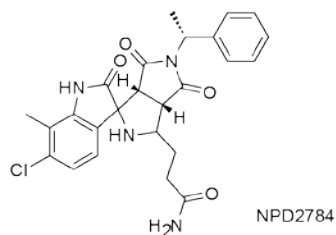
Supplementary Figure 4. *In vitro* **tubulin polymerization in the presence of paclitaxel, nocodazole, NPD2784 and DMSO**. Tubulin polymerization was observed in a fluorescence-based *in vitro* assay (Cytoskeleton, Cat. #BK011P) using 10 µM paclitaxel, 10 µM nocodazole, 21 µM NPD2784, and 1% DMSO. In this assay, tubulin polymerization incorporates a fluorescent analog that accurately reports microtubule polymer mass in terms of relative fluorescence units (RFU). The rate of tubulin polymerization increases in the presence of paclitaxel, an anti-mitotic drug, and substantially decreases in the presence of compounds that inhibit tubulin polymerization such as nocodazole and NPD2784.
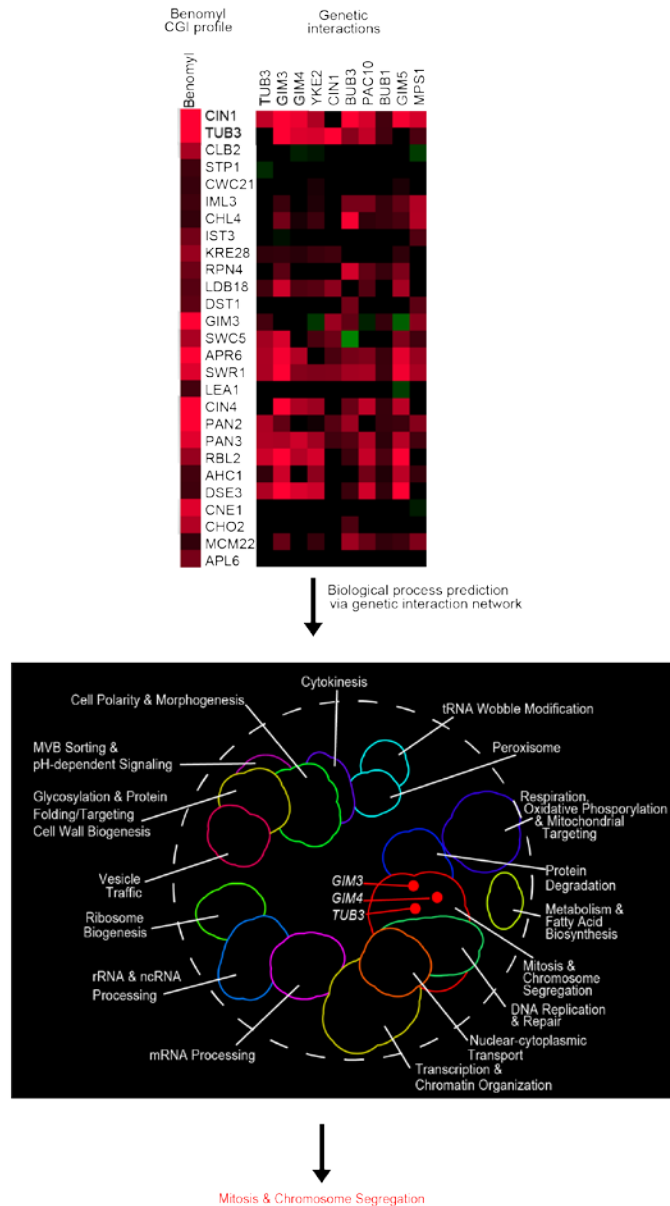
Supplementary Figure 5. Schematic of target bioprocess prediction method. *Top*: Heat maps illustrate the set of gene mutants that are sensitive to benomyl exposure (red signal in Benomyl CGI profile) and a subset of negative (red) and positive (green) genetic interactions associated with the benomyl sensitive mutants (Genetic Interactions). *Bottom*: The chemical genetic profile of a compound is correlated with the known genetic interaction network of yeast. The genes (red nodes) whose genetic interaction profiles have the greatest correlation with the benomyl chemical-genetic interaction profile and represent "gene-level" target predictions are indicated on the global genetic profile similarity network. Functional enrichment among the "gene-level" target predictions is calculated using Gene Ontology as a functional standard to provide "process-level" target predictions.

93
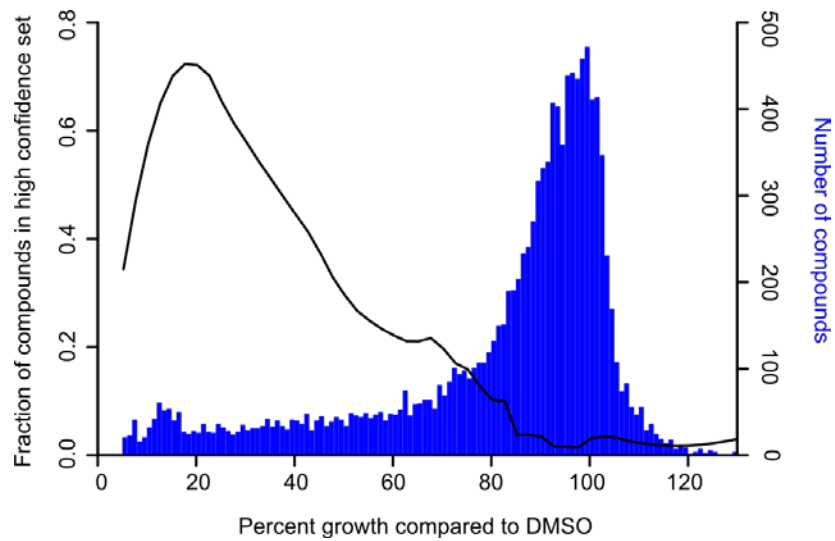
94
95 **Supplementary Figure 6. Bioactivities of all compounds in the final dataset and their**
96 **relationship to high-confidence target prediction**. Bioactivity distribution of screened
97 compounds (blue) and the effect of bioactivity on inclusion into the high confidence set (black)
98 based on our false discovery rate (FDR). Greater bioactivity is correlated with confidence of
99 target predictions, but drops off at very high bioactivity (> 80% growth inhibition).
100
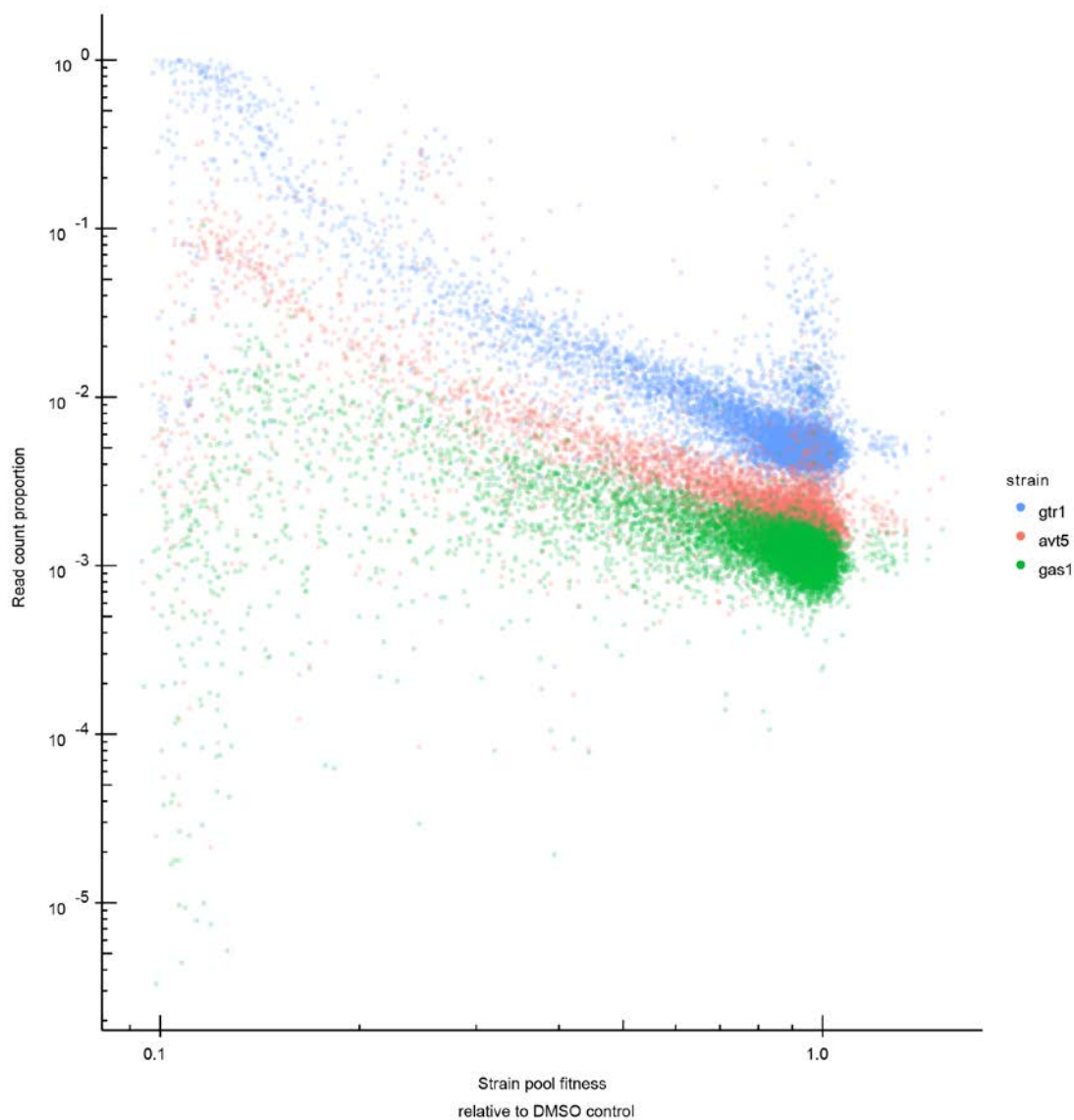
Supplementary Figure 7. Relationship between bioactivity and pool overrepresentation of *gtr1Δ*, *avt5Δ* and *gas1Δ* mutants. The proportion of read counts mapped to *gtr1Δ*, *avt5Δ* and *gas1Δ* strains is inversely related to the fitness of the pooled collection in the presence of compounds (x-axis, growth relative to DMSO).
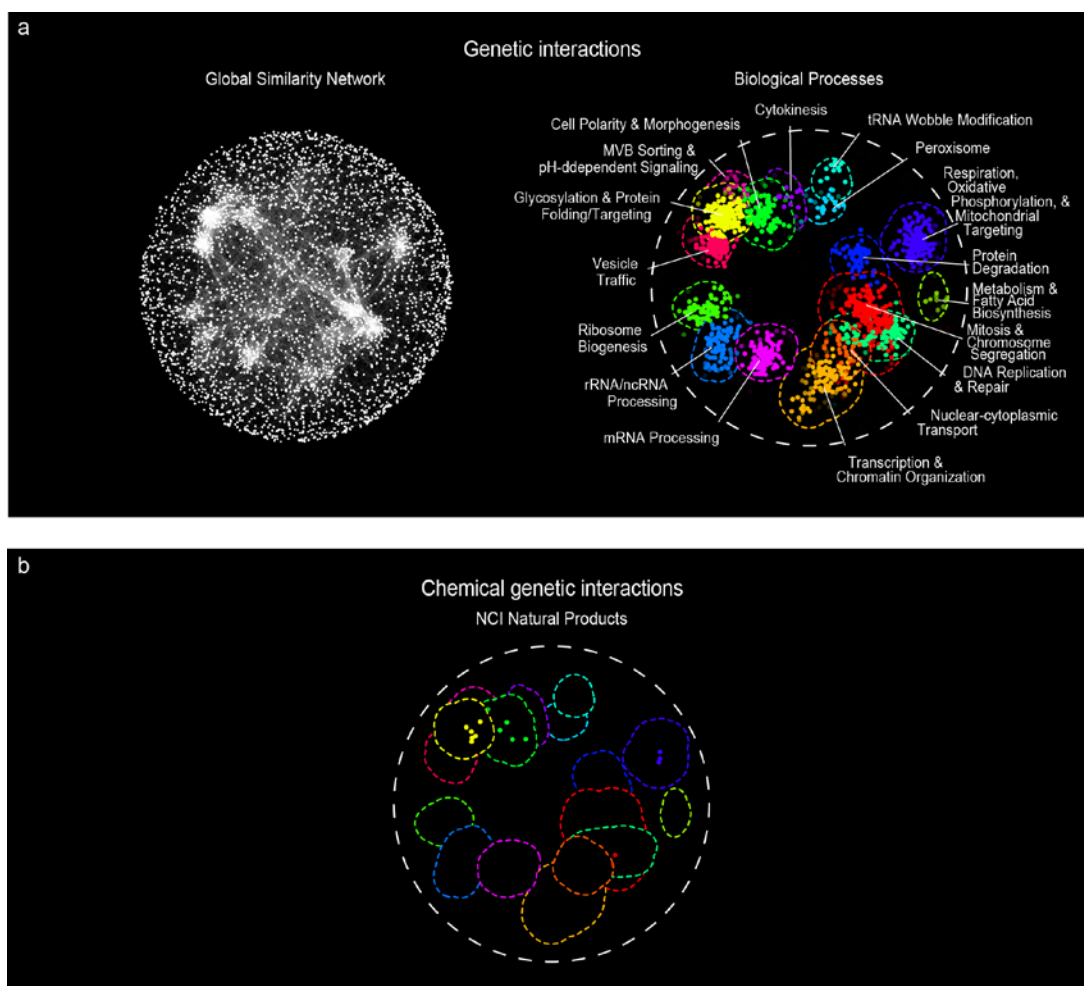
**Supplementary Figure 8**. **Functional space covered by the NCI natural products collection.**
(a) The global genetic interaction similarity network. Genes (nodes) that share similar genetic
interaction profiles are connected by an edge and are proximal to each other; less-similar genes
are positioned further apart. Densely connected network clusters enriched for genes with similar
functional annotations and corresponding to distinct biological processes are indicated and color
coded. **(b)** Compounds of the NCI Natural Product collection predicted to target specific genes
are presented as a node on the network map, where each node corresponds to the top gene-level
target in one compound's top process-level target. Compounds are colored according to the color
of the biological process-enriched cluster shown in **(a)**.
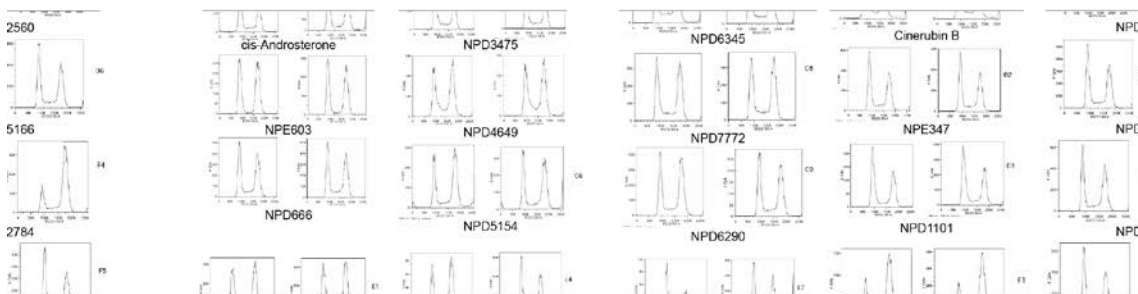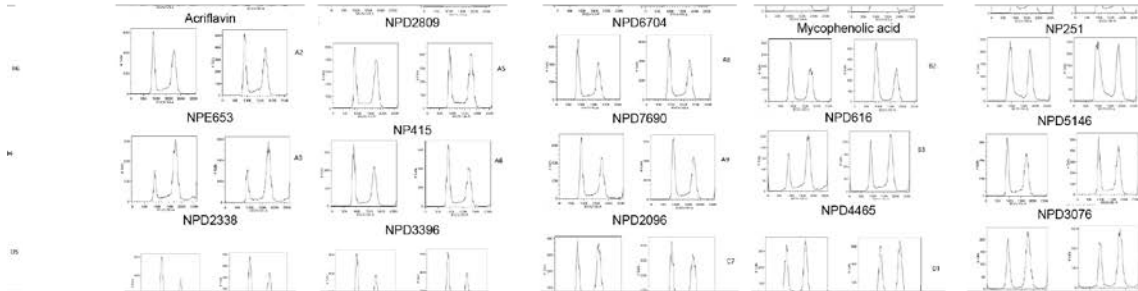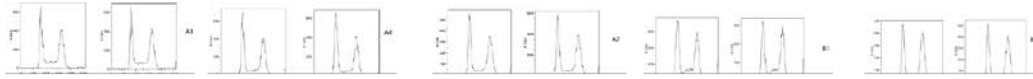
122
Supplementary Figure 9. Cell cycle progression phenotypes associated with 71 compounds.
Asynchronous log phase cells were treated with each compound for 4 h then prepared for flow
cytometry analysis to assess DNA content and identify compounds that result in in a G1-phase
delay or arrest, S-phase delay or arrest, or G2-phase delay arrest phenotypes. Two biological
replicates presented.

13

129

130
131
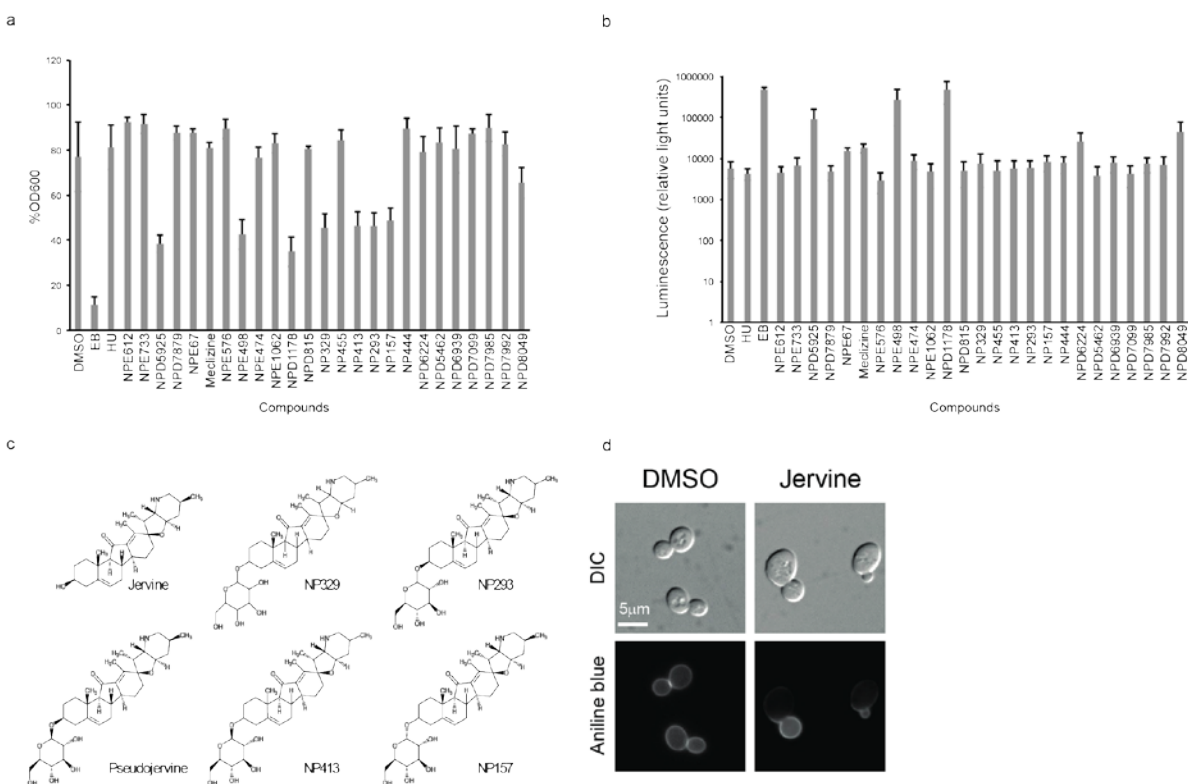
Supplementary Figure 10. Phenotypic analysis of cells treated with predicted cell wall targeting agents. Effect of zymolyase treatment on cells treated with predicted cell wall targeting compounds; reduced OD600 indicated increased cell lysis in the presence of zymolyase (n=3, mean ± S.E.). **(b)** Leakage of adenylate kinase from cells treated with predicted cell wall targeting compounds; increase in luminescence indicates leakage of cytosolic adenylate kinase resulting from compromised cell surface integrity (n=3, mean ± S.E.). **(c)** Structural comparison of jervine, pseudojervine, and related RIKEN NPDepo compounds **(d)**. Aniline blue staining of cells treated with either jervine or DMSO.

140
141
142

143
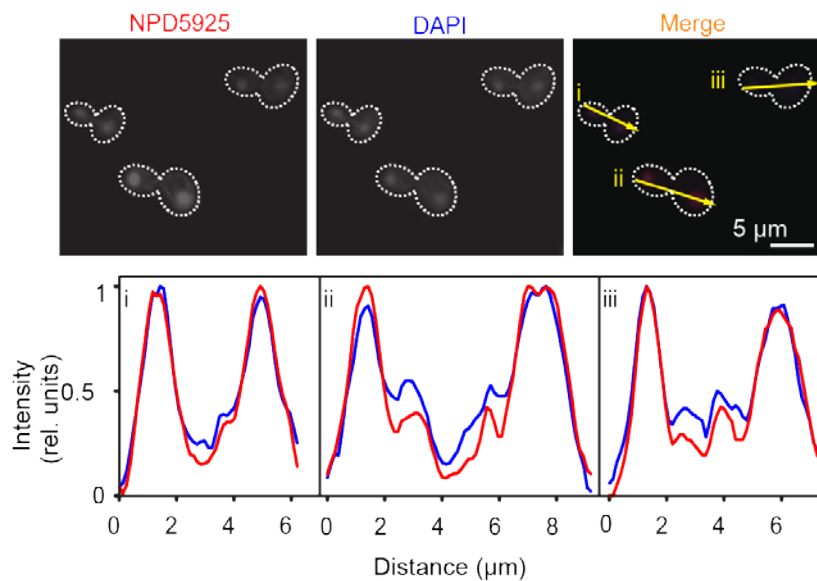
144
Supplementary Figure 11. Overlay of fluorescence signals of NPD5925 and DAPI. Yeast cells were exposed to NPD5925 and DAPI, and visualized by fluorescence microscopy. The fluorescence signals across the cells from bud to mother cell (arrows) of NPD5925 (red) and DAPI (blue) are displayed as a histogram, and are merged in right most panel.

149
150
151
152
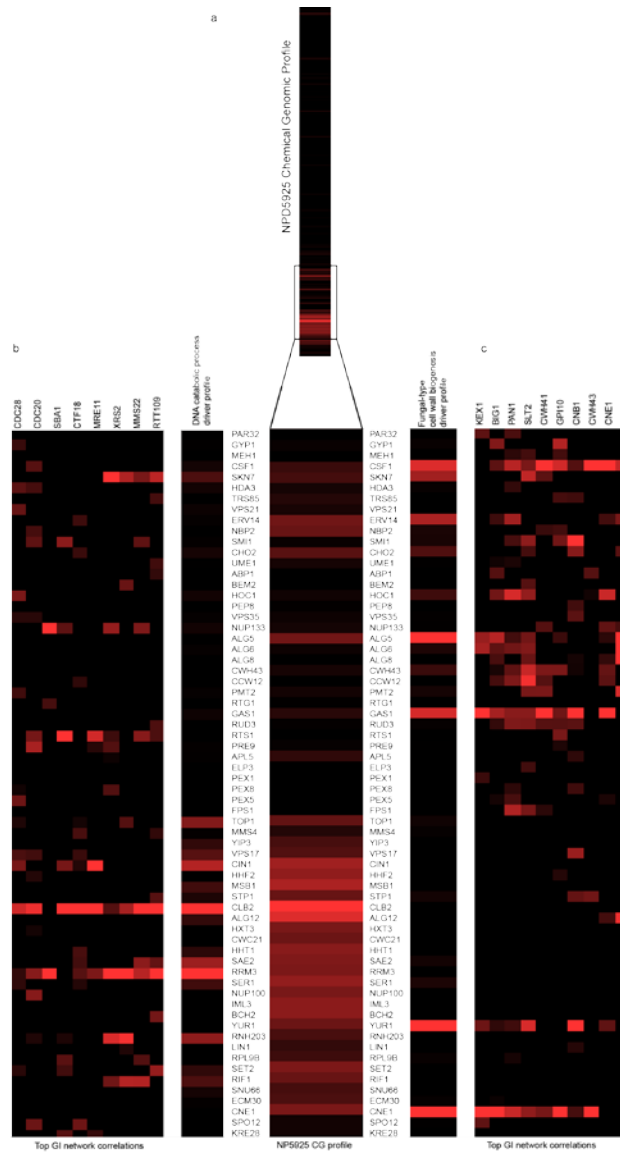
Supplementary Figure 12. Detailed investigation of the genetic interaction profiles that drive dual bioprocess target predictions of NPD5925. A heatmap illustrating the dual target nature of NPD5925. (**a**) The mean chemical genetic interaction profile of NPD5925 (n=3, technical replicates). Chemical-genetic interactions are shown in red. The heatmap visualizing the pleiotropy or dual target nature of NPD5925 was created by performing target prediction on the whole chemical genetic profiles, to compare for pleiotropy, or driver profiles of either GO process (**b**) A portion of the NPD5925 chemical-genetic profile overlaps genetic interactions associated with genes involved in DNA catabolic processes. (**c**) A different set of chemical genetic interactions in the NPD5925 chemical genetic profile overlaps genetic interactions associated with genes involved in fungal-type cell wall biogenesis.
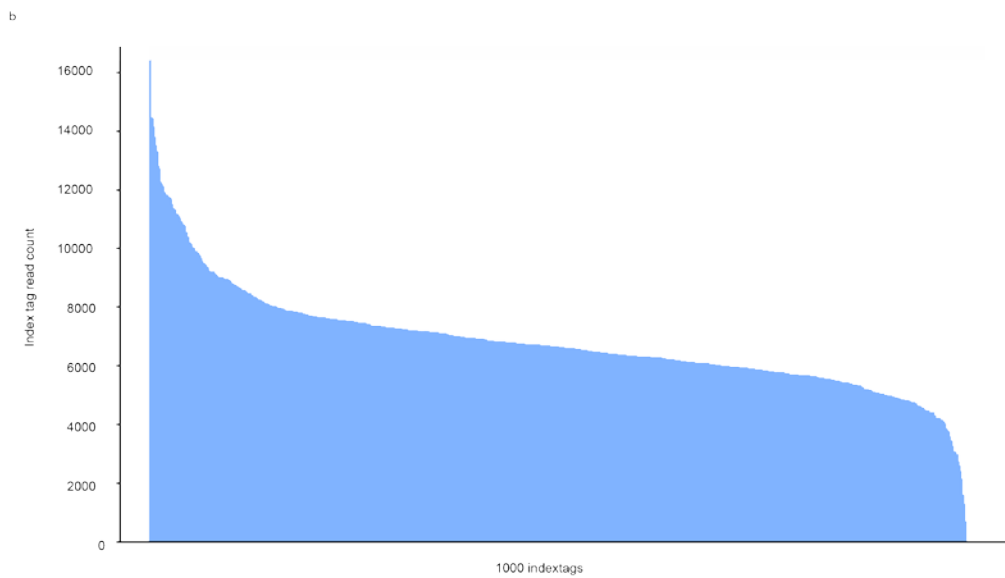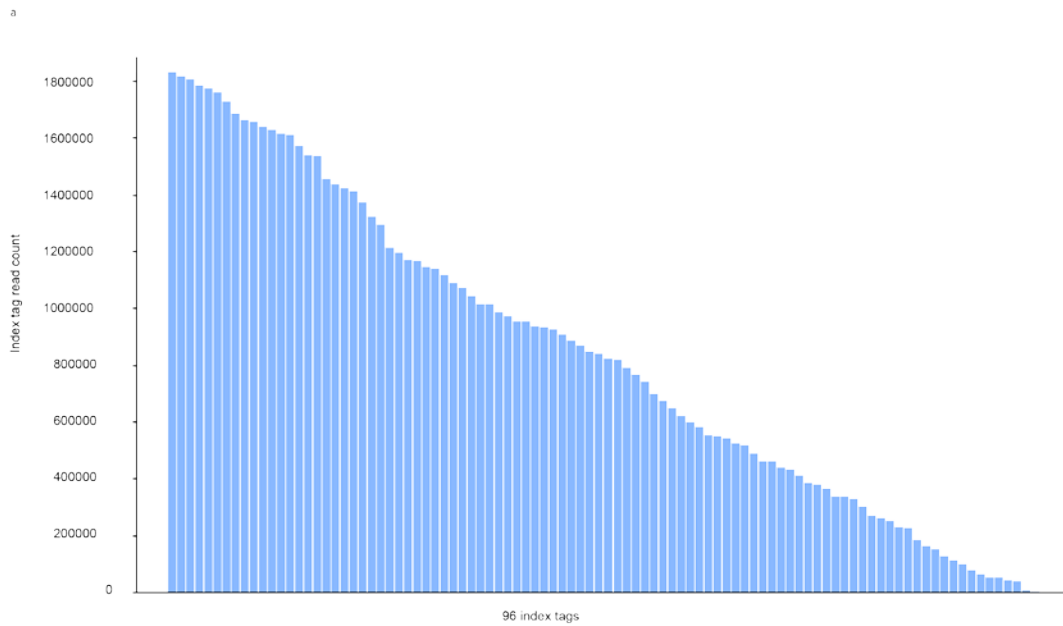
Supplementary Figure 13. **Barcode amplification performance of 96 multiplex tag primers, and rational selection of 768 well-performing multiplex tag primers.** (**a**) Read count distribution of 96 multiplex index tags used in initial pilot experiments. (**b**) Read count distribution of 1000 index tags from identical DMSO treated pools. A set of 768 index tags with near-uniform performance was selected from the 1000 index tags to give the most consistent read counts, and tags yielding very high or low read counts were excluded.
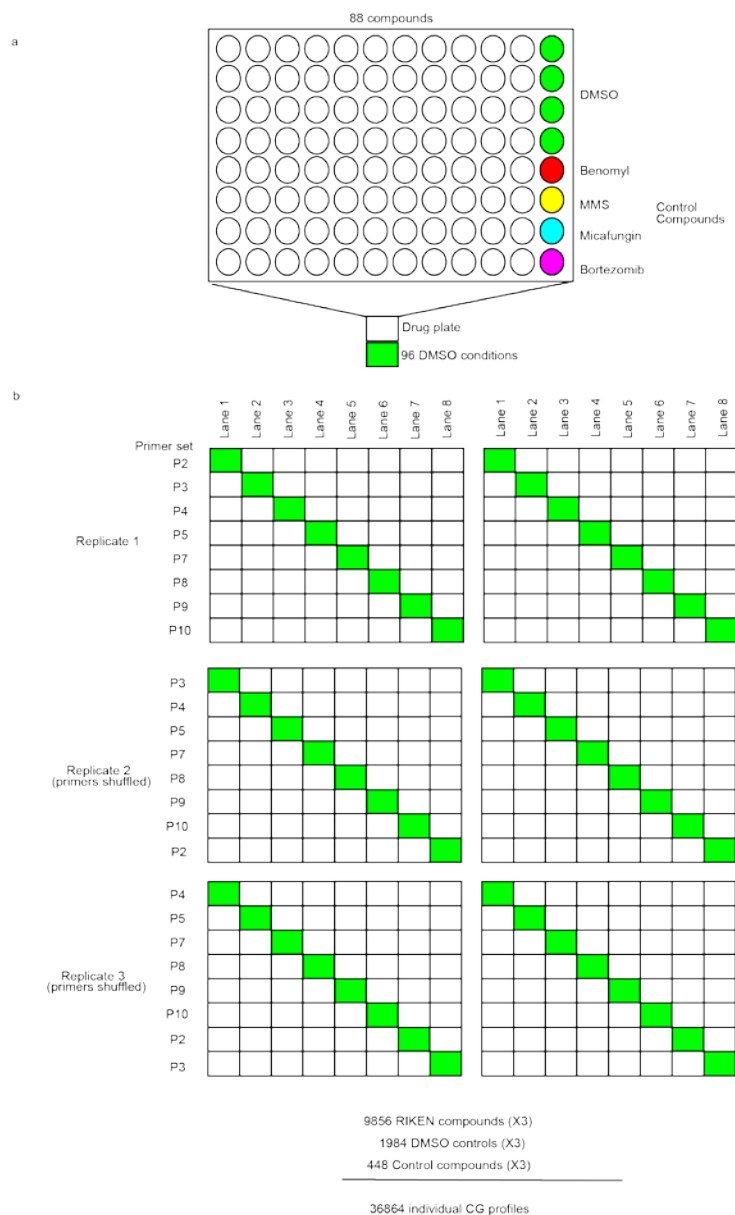
**Supplementary Figure 14. Design of our screening plates and sequencing strategy.** Each of our screening plates had 88 unique compounds, 4 control compounds, and 4 DMSO control conditions. The control compounds were benomyl, MMS, micafungin, and bortezomib for the RIKEN screen. For the NCI/NIH/GSK screens, tunicamycin was added as a fifth control in place of a DMSO. The control compounds and DMSO controls were included to ensure proper plate orientation and assess any plate-specific batch effects. Each sequencing lane had 7 compound plates and one DMSO only plate to serve as the solvent control. Our 768 indexed primers were shuffled for each of the 3 replicates to ensure that no compound plate had the same primer set in any replicates, to ensure we could detect and correct any potential primer biases.
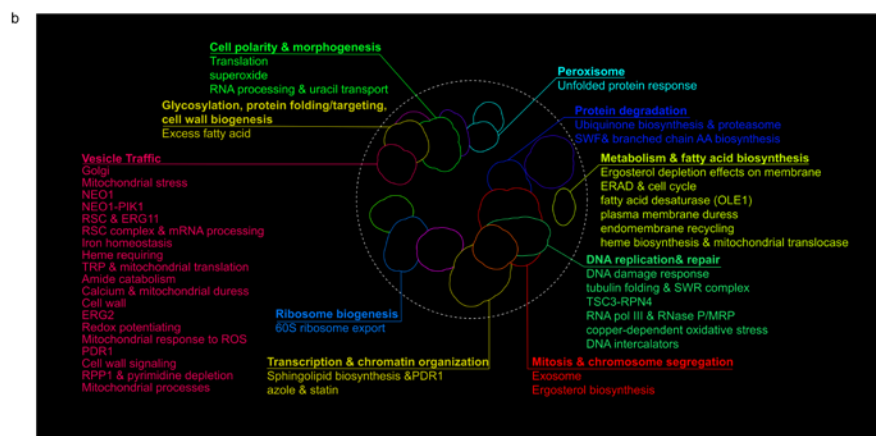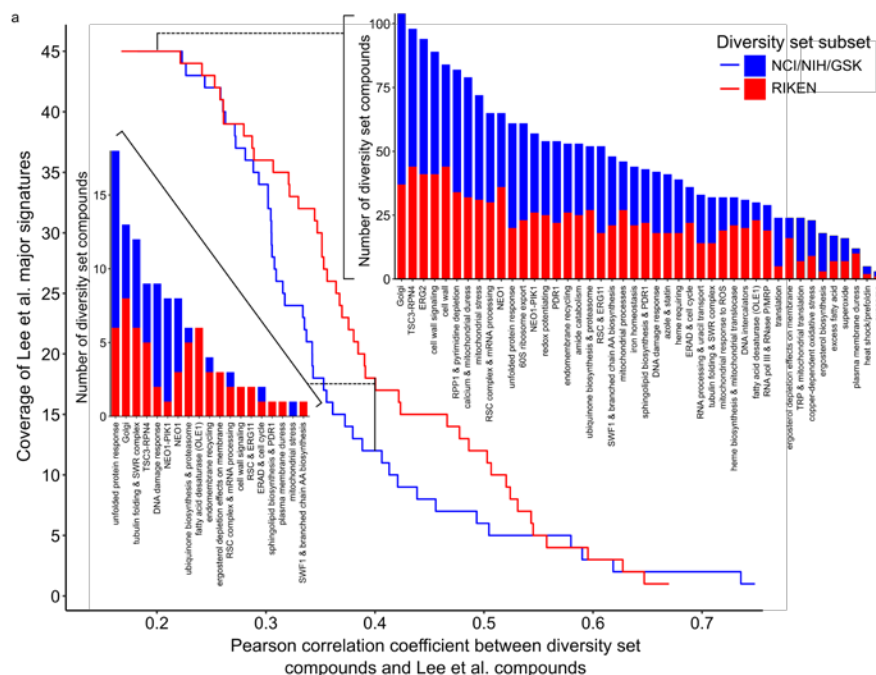
Supplementary Figure 15. Overlap of diversity set compounds with previously defined chemical-genetic signatures. Compounds in each diversity set (RIKEN and NCI/NIH/GSK) were annotated to the major chemical-genetic signatures defined in Lee et al. 2014[8] if they possessed significant Pearson correlation coefficients (PCCs) above a given cutoff value to at least one compound in a Lee et al. major signature (see **Methods**). (**a**) Coverage of Lee et al. major chemical-genetic signatures by at least one RIKEN or NCI/NIH/GSK diversity set compound as a function of PCC cutoff value. The distributions of diversity set compound annotations across Lee et al. major chemical-genetic signatures are shown in the upper-right and lower-left insets for PCC cutoffs of 0.2 and 0.4, respectively. (**b**) We mapped 43/45 chemical-genetic signatures defined in Lee et al. 2014 to the 17 bioprocesses of Costanzo et al. 2016[11].