

SCIENTIFIC DATA

OPEN Data Descriptor: Systematic chemical-genetic and chemical-chemical interaction datasets for prediction of compound synergism

Jan Wildenhain^{1,2}, Michaela Spitzer^{1,2}, Sonam Dolma^{3,†}, Nick Jarvik^{3,†}, Rachel White^{1,†}, Marcia Roy^{1,†}, Emma Griffiths^{2,†}, David S. Bellows^{3,†}, Gerard D. Wright² & Mike Tyers^{1,4}

Received: 08 April 2016
Accepted: 23 September 2016
Published: 22 November 2016

The network structure of biological systems suggests that effective therapeutic intervention may require combinations of agents that act synergistically. However, a dearth of systematic chemical combination datasets have limited the development of predictive algorithms for chemical synergism. Here, we report two large datasets of linked chemical-genetic and chemical-chemical interactions in the budding yeast *Saccharomyces cerevisiae*. We screened 5,518 unique compounds against 242 diverse yeast gene deletion strains to generate an extended chemical-genetic matrix (CGM) of 492,126 chemical-gene interaction measurements. This CGM dataset contained 1,434 genotype-specific inhibitors, termed cryptagens. We selected 128 structurally diverse cryptagens and tested all pairwise combinations to generate a benchmark dataset of 8,128 pairwise chemical-chemical interaction tests for synergy prediction, termed the cryptagen matrix (CM). An accompanying database resource called ChemGRID was developed to enable analysis, visualisation and downloads of all data. The CGM and CM datasets will facilitate the benchmarking of computational approaches for synergy prediction, as well as chemical structure-activity relationship models for anti-fungal drug discovery.

Design Type(s)	molecular interaction analysis objective • strain comparison design
Measurement Type(s)	gene chemical interaction • chemical chemical interaction
Technology Type(s)	compound library screening
Factor Type(s)	selectively maintained organism
Sample Characteristic(s)	<i>Saccharomyces cerevisiae</i>

¹Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK. ²Michael G. DeGroot Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada L8N 3Z5. ³Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada M5G 1X5. ⁴Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Quebec, Canada H3C 3J7. †Present addresses: Division of Neurosurgery, Arthur and Sonia Labatt Brain Tumor Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada M5G 1X8 (S.D.); Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada M5S 3E1 (N.J.); Edinburgh Cancer Research Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XR, UK (R.W.); Centre for Clinical Brain Sciences and Centre for Neuroregeneration, University of Edinburgh, Edinburgh EH16 4SB, UK (M.R.); Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6 (E.G.); School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand (D.S.B.). Correspondence and requests for materials should be addressed to M.T. (email: md.tyers@umontreal.ca).

Background & Summary

The network-based organization of biological systems suggests that combinations of small molecules will be needed to achieve therapeutic efficacy and selectivity for infectious disease, cancer and many other disorders^{1,2}. Cellular networks are strongly buffered against loss of gene function, manifest as synthetic lethal genetic interactions³. A genetic interaction occurs when a phenotype caused by a mutation in one gene is exacerbated (or suppressed) by a mutation in another gene⁴. Genome-wide screens in the budding yeast *S. cerevisiae* have uncovered over 200,000 genetic interactions, in contrast to the ~1,000 essential genes^{3,5}. Analogous to genetic interactions, combinations of chemicals that individually cause minimal phenotypes may exhibit greater than additive effects, termed synergism^{2,6–9}. Notably, compounds that phenocopy the effect of mutations in non-essential genes may have no discernable effect on wild-type cells, but might inhibit growth in a given genetic context². By definition, the biological activity of such compounds would not be detected in many high-throughput screens used in modern drug discovery. Compounds with such latent activities have been termed cryptagens or dark chemical matter^{10,11}.

We recently generated a systematic chemical-genetic dataset in *S. cerevisiae* to allow the discovery and prediction of synergistic interactions between cryptagens that do not have obvious effects on cell proliferation on their own¹¹. Various algorithmic approaches have been developed to predict synergistic compound combinations^{1,12,13}. However, in most cases such predictions have been made on focused datasets and/or known chemical activities, which inherently constrains the development of general methods¹⁴. The dearth of fully factorial drug combination data matrices has hampered the systematic testing and comparisons of different predictive approaches¹. To address this shortfall, we generated two large-scale data sets: a chemical-genetic matrix (CGM) of 356,500 pairwise chemical-gene interaction tests and a derived cryptagen matrix (CM) of 8,128 chemical-chemical interaction tests¹¹. Based on this data, we developed a machine learning approach that integrates structural features of compounds with chemical-genetic interactions to predict compound synergism^{11,15}. This systematic approach identified many novel synergistic anti-fungal combinations, many of which also exhibited species-selective effects against clinical isolates of pathogenic fungi¹¹. The CM represents a benchmark dataset for the development and refinement of synergy prediction algorithms.

Here, we describe the CGM and CM datasets in detail to facilitate use of this data for synergy prediction by computational approaches. The original CGM was generated by screening 4,915 compounds drawn from four different chemical libraries (LOPAC, Maybridge Hitskit 1000, Spectrum Collection and an in-house collection called Bioactive 1). These libraries were screened against 195 diverse *S. cerevisiae* deletion strains, which we termed sentinel strains for their ability to detect otherwise hidden chemical activities¹¹. The updated CGM described here is an extended version of the dataset reported previously: the number of sentinels has been increased from 195 to 242 yeast deletion strains and the cohort of chemical libraries has been expanded to include a second in-house collection of 892 compounds with bioactivity in yeast, termed Bioactive 2. This extended CGM dataset contains data for 5,518 unique compounds, 242 sentinel strains and duplicate measurements for 492,126 pairwise chemical-gene interaction tests, which represent an additional 135,626 duplicate interaction tests compared to the original CGM dataset (Figs 1,2; Table 1 (available online only)). As previously, we defined cryptagens as compounds that were active against more than 4 and less than 2/3 of tested sentinel strains. Out of the 5,518 compounds in the expanded CGM, 1,434 compounds were categorized as cryptagens (Table 2). From the original CGM dataset¹¹, we selected a subset of 128 cryptagens that were used to generate a complete single concentration combination matrix, termed the cryptagen matrix (CM) (Fig. 3a). All 8,128 possible combinations between the 128 cryptagens were tested for synergy at 10 μ M concentration for each compound in a drug pump-deficient *S. cerevisiae* strain (Fig. 3b). Bliss independence values were calculated for each compound pair in the CM dataset (see Methods for details). Independent dose-response surface (checkerboard) assays demonstrated a 65% confirmation rate of synergistic compound interactions from the CM dataset. The full CGM and CM datasets can be accessed at ChemGRID, a web portal that also houses a suite of tools that enable the interrogation and visualization of the chemical interaction datasets (Fig. 4). The CGM dataset and a detailed accompanying description of the yeast cell growth assay have been deposited at NCBI PubChem BioAssay (Data citations 1,2).

Methods

The methods detailed below are expanded versions of descriptions in our related work¹¹.

Compound libraries used to generate the CGM

Compound libraries used were the LOPAC (Sigma), Maybridge Hitskit 1000 (Ryan Scientific) and the Spectrum Collection (MicroSource Discovery Systems Inc). We also screened two custom Yeast Bioactive Collections, termed Bioactive 1 and Bioactive 2, both of which were derived from screens of a 53,000 compound synthetic library (Ryan Scientific) in an *S. cerevisiae* cell proliferation assay at 10 μ M (refs 16,17). The Bioactive 1 library contained 678 compounds that inhibited growth of a *pdrl Δ pdrl3 Δ* strain between 20% and 80%. Bioactive 2 contained 892 compounds that inhibited growth of a *pdrl Δ pdrl3 Δ* strain by at least 80%. Both of the Bioactive 1 and Bioactive 2 collections were selected to maximize chemical structural diversity compared to approved drugs listed by the World Health

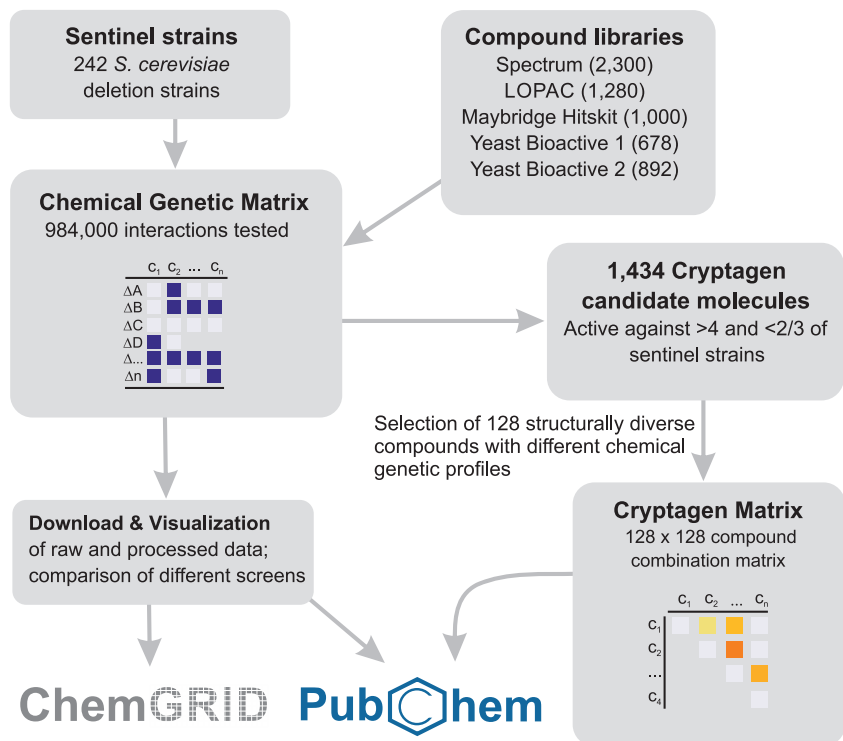


Figure 1. Schematic overview of experimental workflow for CGM and CM dataset generation and data deposition.

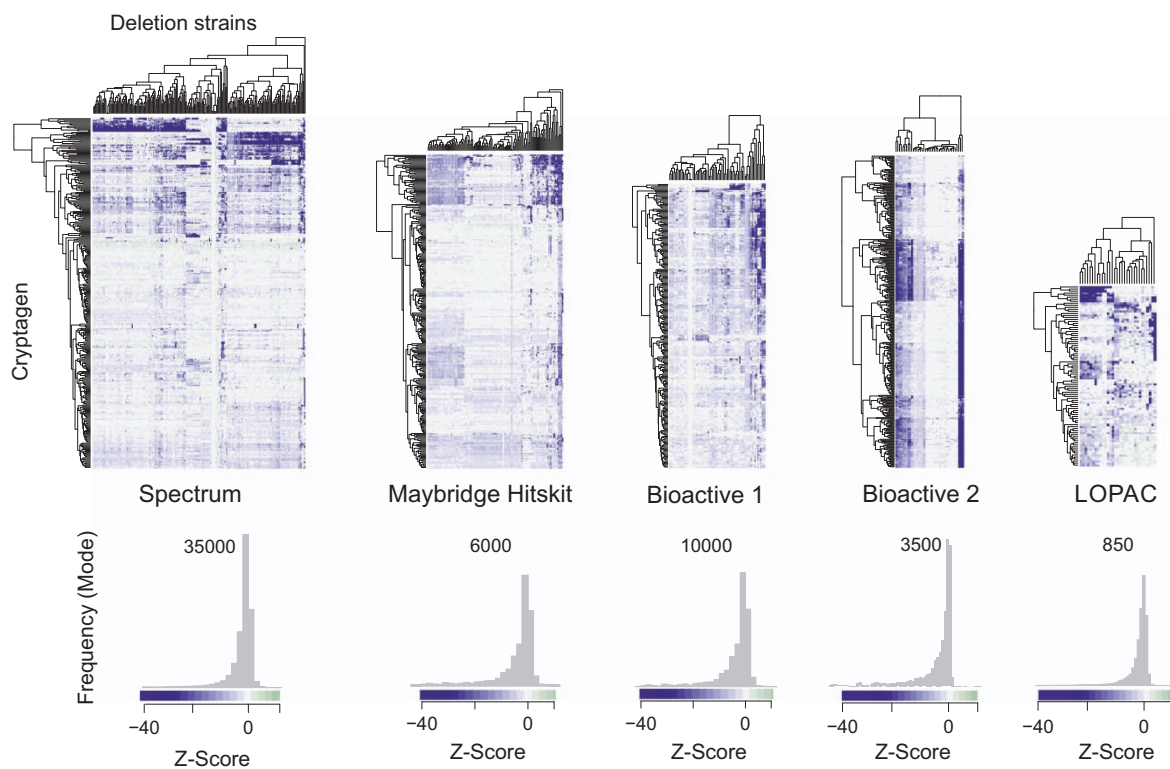


Figure 2. CGM heatmaps for activity (Z-score) of cryptagens contained in each of the five different compound libraries screened against the sentinel deletion strains. Corresponding histograms with Z-score distributions are shown below.

Library	Compounds	Unique cryptagen hits	Sentinel strains	Description
Spectrum03	2000	394	89	Microsource Spectrum Collections consisting of 60% drug-like compounds, 25% natural products and 15% other bioactive molecules
Spectrum05	2000	54	87	
Spectrum08	2000	31	7	
LOPAC	1267	67	20	Library Of Pharmacologically Active Compounds
Maybridge	1000	254	32	Maybridge Hitskit 1000 selection of synthetic compounds representative of Maybridge Hitfinder library
Bioactive 1	678	421	69	Selected from Maybridge Hitfinder 50,000 compound library for growth inhibitory activity in yeast
Bioactive 2	892	213	14	Selected from Maybridge Hitfinder 50,000 compound library for growth inhibitory activity in yeast
Total unique	5518	1434	242	

Table 2. Library screen strains and hits that comprise the chemical genetic matrix.

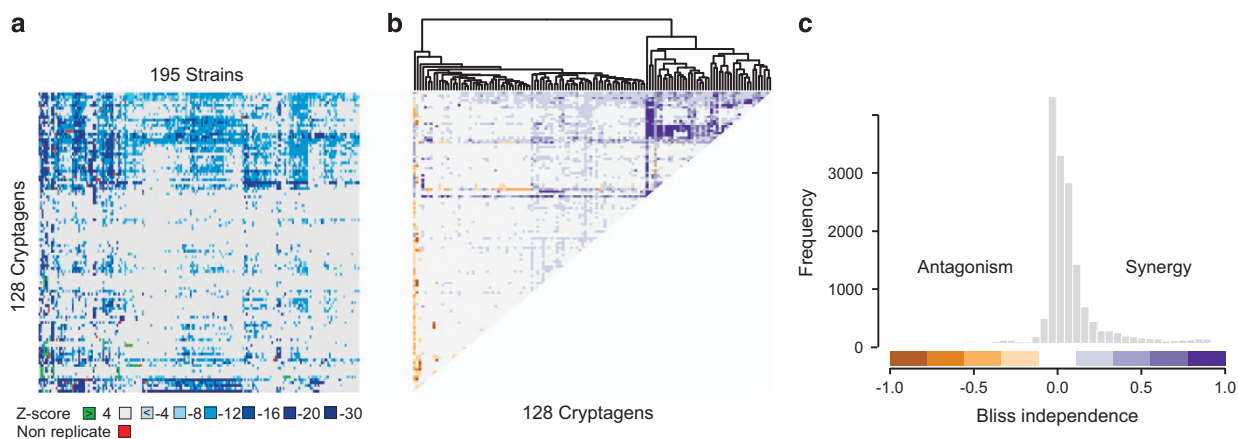


Figure 3. Representations of CM dataset. (a) Heatmap of growth inhibition for the 128 cryptagen compounds from the Microsource Spectrum library against 195 sentinel strains. (b) Heatmap of Bliss scores for all pairwise combinations of 128 cryptagen compounds. (c) Histogram of the Bliss score distribution.

Organization (WHO). The approved drug list contains about 1,500 compounds that represent 50 Bemis-Murcko Fragments¹⁸. In contrast, the Bioactive 1 and Bioactive 2 collections contain 78 novel Bemis-Murcko scaffolds with uncharacterized modes of action. For all screens in the CGM, compound library stocks of 10 mM were diluted to 1 mM working stocks in DMSO in 96 well plates. Over the course of the study, the Spectrum Collection was re-purchased twice: the original library was called Spectrum03, and the two repurchased versions were called Spectrum05 and Spectrum08. As the composition of each release of this library differed somewhat, this non-redundancy resulted in 2,300 unique compounds in the combined Spectrum Collection used in this study. Library compositions and chemical structures are available at www.chemgrid.org.

Yeast strains used to generate the CGM and the CM

The 242 different *S. cerevisiae* deletion strains used as sentinels to generate the CGM (Table 1 (available online only)) were obtained from the Euroscarf deletion collection and are isogenic to BY4741, which has the genotype *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*. For the CM, an isogenic *pdr1Δpdr3Δ* strain (MT2481) was generated from BY4741 using *PDR1::nat* and *PDR3::URA3* deletion cassettes.

Generation of the chemical genetic matrix (CGM)

All screens were conducted in synthetic complete (SC) medium with 2% glucose. Yeast deletion strains (Table 1 (available online only)) were seeded at 50,000 cells per well from fresh overnight cultures in a screening volume of 100 μ l in 96 well plates. 2 μ l of 1 mM compound stock was added to each well for a final compound concentration of 20 μ M. Screens were conducted in technical duplicates using a Biomek FX liquid handling workstation with an integrated stacker carousel. 10 μ M cycloheximide positive controls and DMSO solvent-only controls were set up in columns 1 and 12 of each 96 well plate. All plates were incubated at 30 $^{\circ}$ C without shaking for approximately 18 h or until solvent-treated control cultures were saturated. Cultures were then resuspended by shaking on the robotic platform before reading OD₆₀₀ values on either Tecan Sunrise or Tecan M1000 plate readers¹⁹. Biological repeats were

generated every 8 months over the course of the study for both the *ptr1Δptr3Δ* strain and a wild-type *S. cerevisiae* strain (BY4741) to ensure consistency of compound activity and data reproducibility.

Statistical analysis of the CGM data

All screening data was subjected to the following analysis workflow:

1. LOWESS regression was used to correct spatial effects on growth across all plates for all screens performed with the LOPAC, Maybridge Hitskit 1000 and Spectrum Collection libraries, and for all but seven screens with the Bioactive 1 library. An empirically estimated sliding window of 1/3 was used and data normalization was based on the plate median. The LOWESS normalization method effectively removed variable plate edge effects within the non-active fraction of compounds, which sometimes occurred when plates were read at late time points.
2. For the seven Bioactive 1 screens with higher hit rates, and for all Bioactive 2 screens, the data was not LOWESS corrected but was instead normalized to DMSO controls. The Bioactive 1 library was selected for enrichment of compounds with moderate bioactivity against yeast, whereas the Bioactive 2 library was selected for compounds with strong bioactivity.
3. Median-normalization was applied to all plates and experiments.
4. Z-scores for growth inhibition were calculated based on the median and the interquartile range (IQR) by fitting a normal distribution with $N(1, IQR)$ to the experimental data. We note that the IQR was intentionally chosen as a conservative estimate of variance to reduce the risk of false positives among the weakly active compounds in the screen. This approach slightly underestimates the significance of the Z-scores. In addition, Z-scores, per cent inhibition and normalized OD values were calculated for manual data validation.
5. Data points with high variation between replicates with growth inhibition up to 30% (>3 MAD) were flagged as inconsistent outliers and removed from further analysis.
6. Classification of compounds into 'active' and 'enhanced growth' was based on Z-score cut-offs. Compounds with Z-scores < -4 were classified as 'active', i.e., with reduced OD values compared to the negative solvent-only control. Positive Z-scores >4 were classified as 'enhanced growth', i.e., with OD values greater than the negative control.

All raw and processed CGM data are available online at <http://chemgrid.org/cgm> and from NCBI PubChem BioAssay (Data citation 1).

Selection of cryptagen compounds for the cryptagen matrix (CM)

From the CGM dataset, we identified 1,434 cryptagen compounds. These compounds inhibited growth of at least four and less than two-thirds of the yeast deletion strains. Each of the four chemical libraries yielded cryptagen compounds: the LOPAC at a hit rate of 5%, the Maybridge Hitskit at 27%, the Spectrum Collection at 18%, Bioactive 1 at 10% and Bioactive 2 at 23%. For the generation of the CM, cryptagens were selected from the Microsource Spectrum Collection based on activities against 58 different sentinel strains (Table 1 (available online only)). All cryptagens from the Spectrum Collection were clustered (average linkage hierarchical clustering) based on chemical structure and a structurally diverse set of 128 compounds with diverse chemical-genetic profiles was selected for the generation of the CM.

Generation of the CM

The 128 Spectrum compounds used for the CM were resupplied from MicroSource Discovery Systems Inc. (Groton, CT). Compounds were diluted to a stock concentration of 0.5 mM and arrayed in two 96 well plates. The 128×128 matrix was generated at 10 μ M per compound in duplicate experiments (i.e., biological replicates) using a *ptr1Δptr3Δ S. cerevisiae* strain in 96 well plates. Yeast cultures were seeded at 50,000 cells per well in synthetic complete (SC) medium with 2% glucose at a screening volume of 100 μ l. For the combination screens, 2 μ l of 0.5 mM compound stock was added for both compounds for a final concentration of 10 μ M per compound. The 128 compounds were also screened in combination with 2 μ l DMSO to obtain growth inhibition data for each compound alone for compound interaction calculations. DMSO solvent-only controls were set up in columns 1 and 12 as well as four wells with a 10 μ M cycloheximide positive control for complete growth inhibition in column 12. Plates were incubated for approximately 18 h or until solvent control cultures were saturated at 30 °C without shaking. Cultures were then resuspended by shaking on the robotic platform before reading OD₆₀₀ values on either Tecan Sunrise or Tecan M1000 plate readers¹⁹.

Analysis of CM data

OD₆₀₀ measurements for the CM were normalized to DMSO controls and data was averaged between the biological replicates. Bliss independence²⁰ for compounds X and Y was calculated using the equation $I_{xy} = I_x + I_y - (I_x \times I_y)$ where I_x and I_y correspond to growth inhibition in the presence of 10 μ M compound X and Y, respectively. The expected growth inhibition for combination treatment with X and Y was compared to the actual growth inhibition observed in the CM to obtain the Bliss independence values. Bliss independence values within 90% density kernel fit represented additive effects. Based on the density

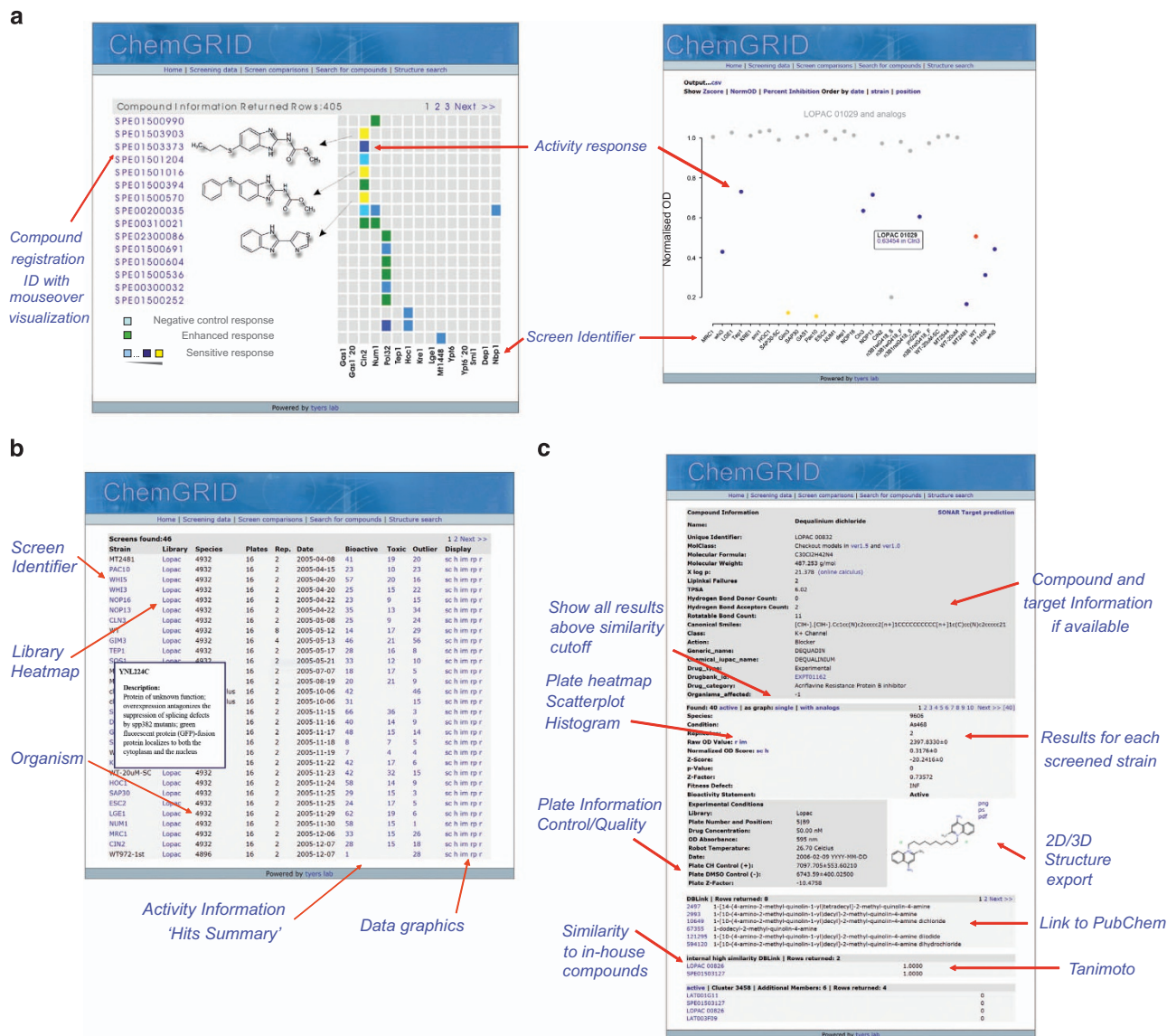


Figure 4. Screenshots of ChemGRID web portal functions. (a) Heatmap view of representative screens in ChemGRID (left) and accompanying data plot for a single screen (right). Mouseover functions allow interactive exploration of small molecule activities across different screens. (b) Overview page for the LOPAC screen component of the CGM. (c) The compound view page summarizes chemical properties, structure, and activity across all screens in the CGM, with automated internal links to similar compounds.

kernel density estimation, Bliss independence values larger than 0.25 represented synergism and values below -0.18 , antagonism. All raw and processed CM data are available online at <http://chemgrid.org/cgm> and from NCBI PubChem BioAssay (Data citation 2).

Note about the Bliss independence model

The Bliss independence model is based on testing two compounds and the pairwise combination at single concentrations²⁰. The Bliss model does not account for possible non-linear concentration effects of either drug, which would instead require assessment over a two dimensional dose-response surface to determine Loewe additivity²¹, also sometimes calculated as the fractional inhibitory concentration index or FICI⁶. The Loewe additivity model requires extensive single and combination drug inhibition measurements that are not practical for large-scale surveys of drug combinations. Therefore, our initial estimates for synergism in the CM relied on the Bliss independence model, as calculated from single concentrations for each drug and the pairwise combination, using the equation $I_{xy} = I_x + I_y - (I_x I_y)$ ⁶. The Bliss independence model for growth inhibition is equivalent to the multiplicative fitness model used to quantify genetic interactions. If two genes A and B do not interact, the expected fitness F of the double deletion strain is $F_{\Delta\Delta B} = F_{\Delta A} \times F_{\Delta B}$ where $F_{\Delta A}$ and $F_{\Delta B}$ represent the fitness defects of the two single deletion strains.

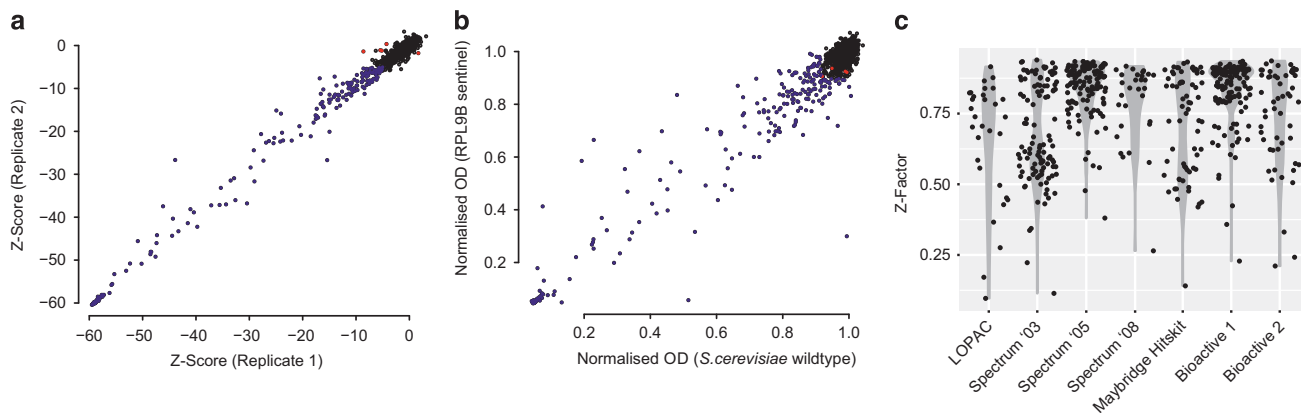


Figure 5. Technical validation of CGM data. (a) Technical replicates of *S. cerevisiae* wild type screen data. (b) Scatter plot of *S. cerevisiae* wild type versus *rbl9Δ* screen data. (c) Z-factor distributions for each individual compound library screened in the CGM.

Replacing fitness F with growth inhibition ($F = 1 - I$) in the genetic model yields: $1 - I_{\Delta A \Delta B} = (1 - I_{\Delta A}) \times (1 - I_{\Delta B})$ which can be simplified to $1 - I_{\Delta A \Delta B} = 1 - I_{\Delta B} - I_{\Delta A} + I_{\Delta A} \times I_{\Delta B}$. Subtracting '1' and multiplication with (-1) yields the Bliss independence formula for growth inhibition: $I_{\Delta A \Delta B} = I_{\Delta A} + I_{\Delta B} - I_{\Delta A} \times I_{\Delta B}$. For example, if two single deletion strains, ΔA and ΔB , have fitness values of 0.6 and 0.8, compared to a wild type fitness of 1.0, the expected fitness of the double deletion strain $\Delta A \Delta B$ is $F_{\Delta A \Delta B} = 0.6 \times 0.8 = 0.48$, corresponding to a growth inhibition of 52%. Similarly, if two small molecules A and B inhibit growth by 40 and 20%, their expected combined growth inhibition based on the Bliss independence model is $I_{AB} = 0.4 + 0.2 - (0.5 \times 0.2) = 0.6 - 0.08 = 0.52$. Thus, if compounds A and B do not interact (i.e., have an additive effect), we expect a growth inhibition of 52% for this compound combination. Two genes are said to display a negative genetic interaction when the observed fitness of the double deletion strain is lower than expected. Similarly, synergy between two small molecules is observed when growth inhibition in response to the combination treatment is higher than expected.

Code availability

The data analysis procedures for the CGM were implemented in R using the additional packages RMySQL, outliers, matlab, amap, RSvgDevice and RSvgTipsDevice. We developed ChemGRID (<http://chemgrid.org/cgm>) as a webportal for the upload, processing and visualisation of chemical-genetic screen data (Fig. 4a,b) using PHP, PEAR, Perl and MySQL. The cheminformatics functionalities to register structural information were implemented with Python and FROWNS, PerlMol and MolDB4 (ref. 22). All code for data analysis is available at <https://github.com/jwildenhain/chemical-genetic-matrix>.

Data Records

Data record 1—chemical genetic matrix (CGM)

The chemical genetic data set of five different compound libraries screened against different subsets of 242 different *S. cerevisiae* deletion strains. In total, 492,126 chemical-genetic interaction tests (i.e., 984,252 independent measurements since each test was performed in duplicate) between 5,518 compounds and 242 *S. cerevisiae* deletion strains are represented in the CGM dataset. All raw and processed CGM data can be downloaded and visualized using ChemGRID (<http://chemgrid.org/cgm>) and are available from NCBI PubChem BioAssay (Data citation 1).

Data record 2—cryptagen matrix (CM)

The raw data as well as the Bliss independence values for all 8,128 pairwise combinations of 128 cryptagen compounds are available from NCBI PubChem BioAssay (Data citation 2). Bliss independence value calculations are described in detail in the Method section above.

Technical Validation

In-plate controls

For the CGM and the CM, each screening plate contained DMSO solvent-only controls as well as cycloheximide positive controls to ensure that inoculum preparation, compound pipetting and screening conditions were not compromised. DMSO solvent-only controls were used for data normalization to account for plate-to-plate variation within each screen and for variation between different screens.

Technical and biological replicates

During generation of the CGM, each *S. cerevisiae* deletion strain was screened in a technical or biological replicate. In addition, we repeatedly screened the compound collections against the *pdr1Δpdr3Δ* strain and a wild-type *S. cerevisiae* strain (BY4741) to ensure data reproducibility (Fig. 5a,b).

Z-factor

For each library screen, Z-factors were calculated to ensure quality of the screening data (Fig. 5c).

Quality filter

Small molecules that did not reproduce between replicate screens were filtered out and not used for further analysis of the CGM data.

Biological replicates

Every compound combination was tested in biological replicate on different days.

Surface dose-response (checkerboard) assays

A set of 75 combinations from the CM was tested in 4 × 4 checkerboard assays. 40 of these combinations initially showed synergistic interactions and 35 did not. This checkerboard analysis confirmed 23 interactions as true positives and 26 as true negatives, with 9 false negatives and 17 false positives. Overall, these confirmatory assays resulted in a 65% confirmation rate of synergistic compound interactions from the CM dataset.

Usage Notes

Dataset applications

The linked CGM chemical-genetic interaction and CM chemical-chemical interaction datasets were designed to serve as resources for computational predictions of chemical synergism, and were used in our original study to predict synergistic antifungal combinations¹¹. To our knowledge, the CM is the first unbiased large-scale benchmark dataset for chemical-chemical interactions. The CGM and CM datasets can also be used to derive chemical structure-activity relationships for anti-fungal drug discovery. Furthermore, the CGM data can be used to infer compound mode of action^{11,13,23,24}.

ChemGRID functionality

The ChemGRID webportal is designed to allow the upload, processing and visualisation of chemical-genetic screen data (Fig. 4a,b). All of the raw and processed CGM data is available on ChemGRID and the dataset for each individual chemical library can be browsed, visualized and downloaded. Screens against different *S. cerevisiae* deletion strains with the same compound collection can be compared and a mouse-over functionality allows for interactive data exploration, including chemical structures and properties. Scatterplots can be generated to compare data from different screens, for example between a wild type and a specific deletion strain. Compounds can also be searched by name and chemical structure. Data for single compounds across different screens can be retrieved and viewed on a summary page together with chemical properties, structures, and links to similar compounds (Fig. 4c). ChemGRID also provides linkouts to other webportals for drug target and compound activity information, including Drugbank²⁵, PubChem²⁶, MolClass¹⁵ and BioGRID⁵.

References

1. Feala, J. D. *et al.* Systems approaches and algorithms for discovery of combinatorial therapies. *Wiley Interdiscip Rev Syst Biol Med* **2**, 181–193 (2010).
2. Sharom, J. R., Bellows, D. S. & Tyers, M. From large networks to small molecules. *Curr Opin Chem Biol* **8**, 81–90 (2004).
3. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
4. Mani, R., St Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc Natl Acad Sci USA* **105**, 3461–3466 (2008).
5. Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res* **43**, D470–D478 (2015).
6. Greco, W. R., Bravo, G. & Parsons, J. C. The search for synergy: a critical review from a response surface perspective. *Pharmacol Rev* **47**, 331–385 (1995).
7. Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. & Sorger, P. K. Systems biology and combination therapy in the quest for clinical efficacy. *Nat Chem Biol* **2**, 458–466 (2006).
8. Kitano, H. A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* **6**, 202–210 (2007).
9. Roemer, T. & Boone, C. Systems-level antimicrobial drug and drug synergy discovery. *Nat Chem Biol* **9**, 222–231 (2013).
10. Wassermann, A. M. *et al.* Dark chemical matter as a promising starting point for drug lead discovery. *Nat Chem Biol* **11**, 958–966 (2015).
11. Wildenhain, J. *et al.* Prediction of synergism from chemical-genetic interactions by machine learning. *Cell Systems* **1**, 383–395 (2015).
12. Cokol, M. *et al.* Systematic exploration of synergistic drug pairs. *Mol Syst Biol* **7**, 544 (2011).
13. Spitzer, M. *et al.* Cross-species discovery of synergistic drug combinations that potentiate the antifungal fluconazole. *Mol Syst Biol* **7**, 499 (2011).
14. Bansal, M. *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* **32**, 1213–1222 (2014).
15. Wildenhain, J., Fitzgerald, N. & Tyers, M. MolClass: a web portal to interrogate diverse small molecule screen datasets with different computational models. *Bioinformatics* **28**, 2200–2201 (2012).

16. Ishizaki, H. *et al.* Combined zebrafish-yeast chemical-genetic screens reveal gene-copper-nutrition interactions that modulate melanocyte pigmentation. *Dis Model Mech* **3**, 639–651 (2010).
17. Zhou, L. *et al.* ALDH2 mediates 5-nitrofurantoin activity in multiple species. *Chem Biol* **19**, 883–892 (2012).
18. Bemis, G. W. & Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J Med Chem* **39**, 2887–2893 (1996).
19. Wong, L. H. *et al.* A yeast chemical genetic screen identifies inhibitors of human telomerase. *Chem Biol* **20**, 333–340 (2013).
20. Bliss, C. I. The toxicity of poisons applied jointly. *Ann Appl Biol* **26**, 585–615 (1939).
21. Loewe, S. Die quantitativen Probleme der Pharmakologie. *Ergebnisse der Physiologie* **27**, 47–187 (1928).
22. Haider, N. Functionality pattern matching as an efficient complementary structure/reaction search tool: an open-source approach. *Molecules* **15**, 5079–5092 (2010).
23. Hillenmeyer, M. E. *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365 (2008).
24. Parsons, A. B. *et al.* Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol* **22**, 62–69 (2004).
25. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. *Nucleic Acids Res* **39**, D1035–D1041 (2011).
26. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res* **44**, D1202–D1213 (2016).

Data Citations

1. NCBI PubChem BioAssay 1159580 (2015).
2. NCBI PubChem BioAssay 1159603 (2016).

Acknowledgements

We thank S. Giguere, G. Giaever and L. Fischer for discussions. This work was supported by Canada Research Chairs in Molecular Studies of Antibiotics (G.D.W.) and in Systems and Synthetic Biology (M.T.), a Howard Hughes Medical Institute International Research Scholar Award (M.T.), a Royal Society Wolfson Research Merit Award (M.T.), a Scottish Universities Life Sciences Alliance Research Chair (M.T.), and by grants from the Canadian Institutes for Health Research (MOP 119572 to G.D.W. and M.T.), the European Research Council (SCG-233457 to M.T.), the Wellcome Trust (085178 to M.T.), the National Institutes of Health (R01RR024031 to M.T.), an award from the Ministère de l'enseignement supérieur, de la recherche, de la science et de la technologie du Québec through Génome Québec (M.T.), and the US Office of the Assistant Secretary of Defence for Health Affairs through the Breast Cancer Research Program (DAMD17-03-1-0471 to D.B.).

Author Contributions

J.W. designed the study and experimental approach, developed all software, analyzed the data and wrote the manuscript. M.S. performed experiments, analyzed data and wrote the manuscript. S.D. performed experiments. N.J. performed experiments. R.W. performed experiments. M.R. performed experiments. E. G. performed experiments. D.S.B. designed the study and experimental approach and performed experiments. G.D.W. wrote the manuscript. M.T. designed the study and experimental approach and wrote the manuscript.

Additional Information

Table 1 is only available in the online version of this paper.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wildenhain, J. *et al.* Systematic chemical-genetic and chemical-chemical interaction datasets for prediction of compound synergism. *Sci. Data* **3**:160095 doi: 10.1038/sdata.2016.95 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>

Metadata associated with this Data Descriptor is available at <http://www.nature.com/sdata/> and is released under the CC0 waiver to maximize reuse.

© The Author(s) 2016