Large-scale interpretation of chemical-genetic interaction profiles using a genetic interaction network

Scott W. Simpkins[1], Justin Nelson[1], Raamesh Deshpande[2], Sheena C. Li[3], Jeff S. Piotrowski[3,4], Erin H. Wilson[2], Abraham A. Gebre[5], Reika Okamoto[3], Yoshikazu Ohya[5], Hiroyuki Osada[3], Minoru Yoshida[3], Charles Boone[3,6†], Chad L. Myers[1,2†]

1. University of Minnesota-Twin Cities, Bioinformatics and Computational Biology Graduate Program, Minneapolis, Minnesota, USA
2. University of Minnesota-Twin Cities, Department of Computer Science and Engineering, Minneapolis, Minnesota, USA
3. RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan
4. Yumanity Therapeutics, Cambridge, MA, USA
5. University of Tokyo, Department of Integrated Biosciences, Graduate School of Frontier Sciences, Kashiwa, Chiba, Japan
6. University of Toronto, Donnelly Centre, Toronto, Ontario, Canada

†Correspondence to: chadm@umn.edu, charlie.boone@utoronto.ca

ABSTRACT

Genetic interactions provide a key for interpreting the functional information contained in chemical-genetic interaction profiles. However, they have remained underutilized in this capacity across recent chemical-genetic interaction screening efforts and their ability to interpret chemical-genetic interaction profiles on a large scale has not been tested. We developed a method, which we refer to as CG-TARGET (Chemical Genetic Translation via A Reference Genetic nETwork), that integrates the data from large-scale chemical-genetic interaction screens with genetic interaction data to predict the biological processes perturbed by compounds. CG-TARGET compared favorably to a standard enrichment approach across a variety of benchmarks, achieving similar performance on measures of accuracy and substantial improvement in the ability to control the false discovery rate of its predictions. We found that one-third to one-half of gene mutants in the data contribute to the highest-confidence biological process predictions and that these contributions overwhelmingly come from negative chemical-genetic interactions. This method was used to prioritize over 1500 out of over 13,000 compounds for further study in a recently-completed chemical-genetic interaction screen in *Saccharomyces cerevisiae*, enabling the rapid functional annotation of unknown compounds to biological processes through targeted biological validations. We present here a detailed characterization of

the method and further biological validations to demonstrate the utility of genetic interactions in the interpretation of chemical-genetic interaction profiles and the effectiveness of our implementation of this concept.

INTRODUCTION

The ability to discover chemical compounds with desirable and/or interesting biological activity is essential to understanding the way compounds and biological systems interact. One way to characterize the biological activity of a compound in an unbiased manner is to profile its activity across a genome-wide array of genetic mutants, also known as chemical-genetic interaction screening [1]. In the resulting chemical-genetic interaction profiles, the identities of the gene mutations that confer sensitivity or resistance to a compound provide functional information regarding the actions performed by that compound inside the cell.

Genetic interaction profiles provide analogous information regarding gene function, and as such can be used to interpret the functional information contained in chemical-genetic interaction profiles [2]. Specifically, shared interactions between chemical-genetic and genetic interaction profiles may implicate a particular gene or group of genes (e.g. a biological process or protein complex) as the target of a compound's actions in the cell (Figure 1). This scheme for interpreting chemical-genetic interaction profiles does not depend on the existence of chemical-genetic interaction profiles for well-characterized compounds, and thus enables the discovery of compounds with novel modes of action.

Recent advances in whole-genome chemical-genetic interaction screening technology have opened the possibility of using chemical genomics as a high-throughput screening approach [3–5]. This would, for example, enable functional profiling of compound bioactivity at earlier points in the drug discovery process, providing an additional means of prioritizing promising compounds (and discarding compounds with less obvious, yet undesirable activities) before investing into them large amounts of resources. However, genetic interactions remained essentially unused in these screens for the systematic interpretation of chemical-genetic interaction profiles. As such, a systematic interpretation of chemical-genetic interaction profiles using genetic interaction profiles has not been demonstrated on a large scale (1,000s to 10,000s of compounds). A study of this type would provide insights into the compatibility between chemical-genetic and genetic interaction profiles and the ability of a genetic interaction-based

method to prioritize compounds with high-confidence predictions while controlling for issues typically associated with high-throughput chemical screening.

In this manuscript, we present the use of genetic interaction profiles to systematically interpret chemical-genetic interaction profiles on a large-scale. Specifically, we developed a method, called CG-TARGET (Chemical Genetic Translation via A Reference Genetic nETwork), that incorporates genetic interaction data and different sources of experimental variation to predict the biological processes perturbed by compounds. We applied this method to a high-throughput chemical-genetic interaction screen of more than 13,000 compounds in *S. cerevisiae* [6], using profiles from the corresponding yeast genetic interaction network [7,8] to interpret the chemical-genetic interaction profiles. CG-TARGET recapitulated known information for well-characterized compounds and showed a marked improvement in the ability to control the false discovery rate – and as a result, prioritize interesting compounds – compared to a baseline approach. We also confirmed, through a global analysis, the compatibility between chemical-genetic and genetic interaction profiles for the purpose of predicting perturbed biological processes. CG-TARGET is available, free for academic use and licensed for commercial use, at github.com/csbio/CG-TARGET.

RESULTS

Predicting perturbed biological processes from chemical-genetic interaction profiles

When developing a method that uses genetic interaction profiles to interpret chemical-genetic interaction profiles obtained at scale, it was important to consider scenarios in which experimental artifacts or common signatures in the chemical-genetic interaction profiles could strongly influence the similarities between chemical-genetic and genetic interaction profiles, leading to biased and inaccurate process predictions. For example, common similarity measures are blind to the variance of individual gene mutants across all chemical-genetic interaction profiles. As a result, gene mutants with highly variable interaction scores in chemical-genetic interaction experiments possess the potential to drive the prediction of processes in a nonspecific manner. While this can be addressed with negative experimental controls, it is not inconceivable that certain gene mutants would respond nonspecifically only in the presence of compound, requiring a correction derived from the dataset itself. Additionally, spurious correlations introduced by normalized similarity measures on weak chemical-genetic interaction profiles

(Pearson correlation coefficient, cosine correlation) can be further amplified by the redundancy in the genetic interaction network, leading to false discoveries.

We developed CG-TARGET (Chemical Genetic Translation via A Reference Genetic nETwork) to address these concerns surrounding the prediction of perturbed biological processes at scale (Figure 1). Predicting a compound's perturbed biological processes using CG-TARGET requires three input datasets (chemical-genetic interaction profiles, genetic interaction profiles, and a mapping from the query genes in the genetic interaction profiles to biological processes) and involves four distinct steps. First, a set of resampled chemical-genetic interaction profiles is generated, each of which consists of one randomly sampled interaction score for each gene mutant across all compound treatment profiles in the chemical-genetic interaction dataset. Second, scores reflecting both the strength of each compound's chemical-genetic interaction profile and its similarity to the profile of each gene mutant are obtained by computing a dot product between all chemical-genetic interaction profiles (comprising compound treatment, experimental control, and resampled profiles) and all L_2-normalized query genetic interaction profiles. These "gene-level" prediction scores, which possess per-compound ranks equivalent to those obtained using cosine similarity but prioritize compounds with stronger profiles, are then aggregated into process predictions; the z-score and empirical p-value for each compound-process prediction are obtained by mapping the gene-level prediction scores to the genes in the process of interest and comparing these scores to those from shuffled gene-level prediction scores and to distributions of the scores derived from experimental control and resampled profiles. Finally, the false discovery rates for these predictions are estimated by calculating the frequency at which experimental control and resampled profiles predict processes across a range of significance thresholds, compared to the compound treatment profiles.

Application to and evaluation on large-scale chemical-genetic interaction screening data

We applied CG-TARGET to the problem of predicting biological target processes from two recent large-scale chemical-genetic interaction screens in *S. cerevisiae*. The first screen was performed on 9850 compounds from the RIKEN Natural Product Depository [9] (the "RIKEN" screen) and the second was performed on 4116 compounds from the NCI Open Chemical Repository's plated compound libraries, the NIH Clinical Collection, and GlaxoSmithKline's Published Kinase Inhibitor Set (the "NCI/NIH/GSK" screen) [10]. The number of chemical-

genetic interaction profiles obtained from each screen was 8418 and 3565, respectively. In both screens, the chemical-genetic interaction profiles were determined across a diagnostic set of approximately 300 haploid gene deletion mutants. Genetic interaction profiles were obtained from a compendium of genetic interaction profiles in *S. cerevisiae* [7], with the query genes mapped to propagated Gene Ontology biological process terms [11,12] to define the process targets.

To provide a baseline approach for benchmarking the performance of CG-TARGET on these large screens, we implemented a standard enrichment approach that tests for the enrichment of processes in each compound's top-k gene-level prediction scores. Using experimental control and resampled profiles, we assessed the ability of the enrichment-based prediction method to control the false discovery rate at various values of k. The best-performing value of k was then used to further benchmark the accuracy of the process predictions made by CG-TARGET.

CG-TARGET was successful in controlling the false discovery rate across both chemical-genetic interaction screens, identifying 848 out of 8418 compounds from the RIKEN screen (10%) and 705 compounds from the NCI/NIH/GSK screen (20%) with at least one prediction that achieved false discovery rates of 25 and 27%, respectively (Table 1, Figure 2A-D). In contrast, the best-performing top-k enrichment approach (k=100) identified only 57 compounds with an equivalent false discovery rate when applied to the profiles from the RIKEN screen (Figure 2E-F). In all cases, the false discovery rates derived from resampled profiles were more conservative than those derived from experimental controls, suggesting that some sources of variance in each gene mutant's interaction scores arise only upon treatment with compound and therefore cannot be corrected using only negative experimental controls. The compounds from the RIKEN and NCI/NIH/GSK screens with at least one prediction at or below the respective false discovery rate cutoff set for each screen will be referred to as the RIKEN and NCI/NIH/GSK "high-confidence sets," respectively.

We benchmarked the accuracy of CG-TARGET's process predictions against a set of 35 gold-standard compound-process annotations curated from the literature and observed favorable performance to that on predictions generated using the top-100 enrichment approach. More specifically, we computed the rank of each compound's gold-standard process within its set of process predictions, and compared this rank to those obtained from randomly shuffled predictions. CG-TARGET performed slightly worse overall when comparing the ranks of the

gold-standard processes (12 versus 14 in the top ten, Figure 3A) and the number of compounds with significant ranks (22 versus 23, Table 2). However, only 2 out of 23 significantly-ranked gold-standard predictions made using the top-100 enrichment approach achieved a false discovery rate of 25% or less, while 16 out of 22 predictions from CG-TARGET achieved this same false discovery rate. As such, CG-TARGET discovered 8-fold more compounds with significantly-ranked gold-standard process annotations within the RIKEN high-confidence set. This result provides further evidence supporting the utility of genetic interaction profiles in the interpretation of chemical-genetic interaction profiles, while simultaneously demonstrating that the predictive power of genetic interaction profiles improved when combined with additional experimental data.

In a more global benchmarking effort, we also observed that CG-TARGET improved the prioritization of process predictions when applied to simulated chemical-genetic interaction profiles. The set of simulated profiles was designed contain three compounds that target each query gene in the genetic interaction dataset; each simulated profile thus inherited the process annotations of its parent genetic interaction profile, providing a gold standard with which to evaluate their predictions. Evaluation performed on the top process prediction for each simulated compound revealed that CG-TARGET captured 15% more gold-standard annotations than did top-100 enrichment. While both methods only captured a gold-standard annotation in the top process prediction for approximately 30% of the simulated compounds, this still represents a 56-fold enrichment over the background expectation of 0.00533. In addition, CG-TARGET more successfully prioritized its true positive annotations, as shown by the consistent improvement in precision over top-100 enrichment, especially at low recall values (Figure 3B). The diversity of the true positive process predictions, when mapped to a set of 17 broad functional neighborhoods, was also improved using CG-TARGET (Shannon index = 2.58 for CG-TARGET versus 2.33 for top-100 enrichment predictions), likely due to a substantial reduction in the number of compounds mapped to the “vesicle traffic” neighborhood (150 for CG-TARGET vs. 367 for top-100 enrichment).

In addition to benchmarking, we investigated the potential to expand the use cases of CG-TARGET to the prediction of perturbed protein complexes. For protein complex prediction on the RIKEN screen data, 714 compounds were identified with at least one prediction achieving a false discovery rate of 25% or less. 603 of these 714 compounds were also identified in the high

confidence RIKEN process predictions, suggesting the potential to map prioritized process predictions to more specific, defined predictions of compound targets in the cell. For example, the top protein complex prediction for NPD1409 was "Kornberg's mediator (SRB) complex," which plays an important role in the initiation of transcription as well as chromatin looping [13,14]. This prediction agrees with the compound's top process predictions from the RIKEN screen to perturb "chromosome organization," "DNA metabolic process," and/or "RNA polymerase II transcriptional preinitiation complex assembly," and points to a more direct target for testing in experimental validation efforts.

Insights into the compatibility of chemical-genetic and genetic interaction profiles

Given previous demonstrations [2,7] and the evaluations presented here, it should be clear that the use of genetic interaction profiles to interpret chemical-genetic interaction profiles is both appropriate and useful. However, a further investigation of the inner workings of this approach was warranted to more comprehensively understand the extent to which these two types of profiles can be combined and how this affects the prediction of processes. Here we present visualizations that reveal insights into 1) the interpretation of a chemical-genetic interaction profile to predict a biological process and 2) the different ways in which a process can be predicted using chemical-genetic and genetic interaction profiles. Finally, we quantify, across the RIKEN high confidence set of compounds, the relationship between chemical-genetic interactions and their importance to the prediction of perturbed biological processes.

To better understand the interpretation of chemical-genetic interaction profiles, we quantified, for every compound, the contribution of each gene mutant to the prediction of individual biological processes. For a single compound and predicted process, these "importance scores" were obtained by 1) computing the Hadamard product (elementwise multiplication) between the compound's chemical-genetic interaction profile and each L_2-normalized query genetic interaction profile mapped to the predicted process and 2) for each gene mutant, computing the mean of this product across the genetic interaction profiles. These scores can be positive, indicating agreement in the sign of chemical-genetic and genetic interactions for a particular gene mutant, or they can be negative, indicating that the interactions do not agree for that gene mutant. As such, the importance scores summarize the concordance between chemical-genetic and genetic interaction profiles, conditioned on an individual compound and a perturbed process of interest.

The prediction of NPD4142, a compound from the RIKEN Natural Product Depository, to the "mRNA transport" process can be used to illustrate how the overlap between chemical-genetic and genetic interactions leads to process predictions (Figure 4A). A qualitative examination revealed that, indeed, NPD4142 possesses a pattern of chemical-genetic interactions similar to the genetic interactions for the query genes annotated to mRNA transport. However, a quantitative assessment achieved more nuance in this comparison. While the *POM152* deletion mutant possessed the strongest negative interaction with NPD4142, the importance scores revealed that it was not the most important gene mutant for making this prediction; instead, the deletion mutant for *NUP133*, which possessed a weaker chemical-genetic interaction score but more genetic interactions with the mRNA transport-annotated query genes, emerged as the most important for predicting mRNA transport.

We also compared the concordance of chemical-genetic and genetic interaction profiles across multiple compounds predicted to the same process, revealing that individual processes were predicted by both homogenous and heterogeneous sets of chemical-genetic interaction profiles. For example, all predictions made to "proteasome assembly" depended almost entirely on a strong negative chemical-genetic interaction with *RPN4*, which was captured most clearly by the relevant importance scores (Figure 4B). This uniformity in the prediction of a process is contrasted by the diversity of profiles captured within "fungal-type cell wall organization" predictions (Figure 4C). Here, filtering on the importance scores showed that chemical-genetic interactions with four genes – *GAS1*, *SMI1*, *ABP1*, and *DFG5* – were primarily responsible for predictions to this term, but with low agreement regarding their relative importance for each compound's prediction. In the lattermost case, the concordance of chemical-genetic and genetic interactions was not particularly obvious, yet was sufficient to enable the prediction of a perturbed process.

More globally, we found broad contribution across a large fraction of observed chemical-genetic interactions – primarily negative interactions – to the prediction of perturbed processes (Figure 4D). By comparing the chemical-genetic interactions for each compound to their corresponding importance scores for that compound's top process prediction, we observed that nearly one-third (5398 / 16464) of chemical-genetic interactions contributed to top process predictions, the fraction of which increased to nearly one-half (5087 / 10281) when considering only negative interactions. While positive chemical-genetic interactions were much less

frequently observed overall (and only 7% positively contributed to a top process prediction), they were 3.5 times more likely to contribute negatively to a process prediction than were negative interactions (1.47% vs. 0.42%). Overall, 199 gene mutants (72%) contributed to at least one top process prediction, while half (143) contributed to at least five predictions (importance score > 0.1). Based on a more stringent threshold on importance scores (> 1.0), 65 gene mutants (23% of mutants) were observed to be very strong contributors to certain process predictions, with 17 of these contributing strongly to the top process prediction of at least five compounds. While some gene mutants certainly contributed disproportionately to a subset of predictions, the broad majority of predictions required contributions from a much larger fraction of gene mutants.

Experimental validation of compound-process predictions

Phenotypic analysis of cell cycle progression

Several compounds from the RIKEN dataset were predicted to perturb the process related ot the cell cycle. We chose to test 19 of these compounds, 14 of which had high-confidence to the "spindle assembly checkpoint" process active in the M phase of the cell cycle, to determine if our predictions captured the biological activity of these compounds. Indeed, we observed that 7 of the 19 compounds induced a cell cycle phenotype, with 6 of the 14 compounds annotated to spindle assembly checkpoint inducing abnormally large buds on cells, changes in the budding index of cells, and increases in cellular DNA content, indicative of arrest in G2/M phase (Figure 5A-C). These phenotypes were not observed when performing the same experiments on a set of 10 active compounds from the high-confidence set whose predictions were to processes unrelated to the cell cycle. This difference in the rate of validation between predicted active compounds and negative controls (6 / 14 vs. 0 / 10) was statistically significant ($p < 0.03$, proportion test). Two of the selected compounds were predicted to perturb "cell cycle phase," one of which induced phenotypes consistent with G1 arrest (Fig 5A-C). This provided one demonstration of CG-TARGET's ability to prioritize compounds that perturb a particular function in the cell.

Inhibition of tubulin polymerization

Compounds that disrupt microtubules are useful for studying cell organization and division, and remain promising candidates as antitumor agents [15–17]. We therefore chose to experimentally validate our predictions in a way that might identify compounds that possess such

activities. All compounds with the strongest predictions (FDR = 0%) to "tubulin complex assembly" were selected for biochemical validation in an *in vitro* tubulin polymerization assay (Figure 5D). Similar to the previous validation, a negative control set of compounds was selected to contain active compounds (process predictions with FDR $\leq$ 25%) whose predictions were not related to microtubules or related processes. We observed that the novel compound NPD2784 strongly inhibited tubulin polymerization nearly as well as the drug nocodazole and more strongly than the microtubule probe benomyl. In addition, the entire set of compounds predicted to perturb tubulin complex assembly showed significantly increased inhibition of tubulin polymerization when compared to the negative control compounds ($p < 0.005$, Wilcoxon ranksum test). These confirmatory results showed the translation of our predictions to a specific biochemical validation, even in the context of a different species.

DISCUSSION

The scaling of chemical-genetic interaction screens from tens or hundreds of compounds to tens of thousands of compounds has provided the opportunity, and the necessity, to more comprehensively characterize appropriate methods for interpreting the interaction profiles and prioritizing high-confidence compounds. We developed a method, CG-TARGET, to address this need and used it to predict perturbed biological processes for more than 13,000 interaction profiles from a recent high-throughput chemical-genetic interaction screen [6]. CG-TARGET demonstrated the ability to recapitulate known compound function while controlling the false discovery rate, prioritizing 1522 compounds for further study. Further investigation of the profiles from these high-confidence compounds revealed broad compatibility between chemical-genetic and genetic interaction profiles. In addition to these findings, the predictions made using CG-TARGET were experimentally validated on a large scale for 67 compounds in an orthogonal cell cycle assay and revealed insights into the distribution of functions perturbed by compounds in large compound libraries [6].

In high-throughput chemical screens, it is important to prioritize the compounds most likely to demonstrate desired biological activity in further studies. While CG-TARGET and a baseline approach performed similarly on the task of ranking gold-standard compound-process annotations, CG-TARGET was 8 times better at prioritizing these compounds as high-confidence predictions. Surprisingly, CG-TARGET outperformed the same baseline method when predicting and prioritizing perturbed processes for simulated chemical-genetic interaction

profiles derived from genetic interaction profiles across the genome, providing evidence for its value in discovering compounds with modes of action not previously characterized in the literature. This is particularly important, as our gold standard set of compound-process annotations, consisting of 35 compounds across 17 biological processes (4 of which are DNA-related), does not enable prediction across the range of biological processes present in the cell. Genetic interactions thus provide the most comprehensive reference for interpreting chemical-genetic interaction profiles in an unbiased, genome-wide manner.

While we demonstrated the ability to predict perturbed processes for compounds and prioritize the highest-confidence predictions, many further steps are required to identify lead compounds and ultimately develop molecular probes or even pharmaceutical agents. Perturbing a biological process does not necessarily require perturbing a specific protein target, and as such, further refinements to our methods are needed to prioritize the compounds most likely to perturb a small number of defined targets in the cell. Different modes of chemical-genetic interaction screening can provide support in this endeavor, as the profiling of heterozygous diploid strains provides evidence for the direct, essential cellular target(s) of a compound [1,4]. Regardless of these limitations in predicting defined targets, information about the processes perturbed by an entire library of would be useful in selecting the compounds most amenable to activity optimization and off-target effect minimization in the development of a pharmaceutical agent or molecular probe.

Moving forward, this work supports the idea of performing both genetic and chemical-genetic interaction screens in other species for which obtaining functional information on compounds would be useful. For example, genome-wide deletion collections have been developed for *Escherichia coli* [18] and *Schizosaccharymyces pombe* [19] and used to perform chemical-genetic interaction screens [20,21] as well as genetic interaction mapping [22–25]. Such efforts are even underway in human cell lines, enabled by genome-wide CRISPR knockout screens [26–29]. Furthermore, future efforts to interpret chemical-genetic interaction profiles in a new species need not wait for the completion of an all-by-all genetic interaction network, as this work highlights the ability of a diagnostic set of gene mutants to capture functional information and predict perturbed biological processes. From the discovery of urgently-needed antibacterial or antifungal agents, to the treatment of orphan diseases or a better understanding of drug and chemical toxicity, the combination of chemical-genetic and genetic interactions in a high-

throughput format, with appropriate analysis tools, offers a means to achieve these goals via the discovery of new compounds with previously uncharacterized mechanisms of action.

MATERIALS AND METHODS

Datasets

Chemical-genetic interaction data

Chemical-genetic interaction profiles were obtained from a recent study [6], in which nearly 14,000 compounds were screened for chemical-genetic interactions across approximately 300 haploid yeast gene deletion strains. Each profile contains z-scores that reflect the deviation of each strain's observed fitness from its expected fitness in the presence of compound. The chemical-genetic interaction profiles were obtained in two batches and divided as such into: 1) the "RIKEN" dataset, which contains chemical-genetic interaction profiles across 289 deletion strains for 8418 compounds from the RIKEN Natural Product Depository [9] and 5724 negative experimental controls (solvent control, DMSO) ; and 2) the "NCI/NIH/GSK" dataset, which contains chemical-genetic interactions across 282 deletion strains for 3565 compounds from the NCI Open Chemical Repository, the NIH Clinical Collection, and the GSK kinase inhibitor collection [10], as well as 2128 negative experimental control profiles. The solvent control profiles consisted of biological and technical replicate profiles.

Genetic interaction data

The genetic interaction dataset consisted of quantitative fitness observations for the double mutants obtained upon crossing between 1505 high-signal query gene mutants into an array of 3827 array gene mutants, obtained from a recently assembled *S. cerevisiae* genetic interaction map [7,8]. These observations are represented as epsilon scores, which quantify the difference between each double mutant's observed and its expected fitness values and are analogous to the chemical-genetic interaction z-scores. The procedure for selecting the 1505 high-signal query genes out of the larger pool of 4382 is described in [6]. Briefly, each query profile was required to possess at least 40 significant genetic interactions, a sum of cosine similarity scores with all other query profiles greater than 2, and a sum of dot products with all other query profiles greater than 2. When comparing with chemical-genetic interactions, the genetic interaction dataset was filtered to contain only array strains present in the chemical-genetic interaction datasets.

372 *GO Biological Processes and protein complexes*

373 A subset of terms from the "biological process" ontology within the Gene Ontology
374 annotations [12] were used as the processes. Query genes from the *S. cerevisiae* genetic
375 interaction dataset were mapped to genes to biological process terms using annotations from the
376 Saccharomyces cerevisiae Genome Database [11]. Both gene ontology and *S. cerevisiae*
377 annotations were downloaded from their respective databases via Bioconductor in R [30]. Terms
378 were propagated using "is_a" relationships, such that all child terms were annotated to their
379 parent terms as well. The final set of processes consisted of the terms with 4 – 200 gene
380 annotations from the set of 1505 high-signal query genes in the genetic interaction dataset.
381 Protein complex annotations were obtained from [8]. Complexes with 3 or more genes
382 annotated to them were used as the input biological processes for CG-TARGET-based protein
383 complex predictions.

384 *Gold standard compound-process annotations*

385 Biological processes were assigned to 35 primarily antifungal compounds with chemical-
386 genetic interaction profiles in the RIKEN dataset, based on known information about their
387 mechanisms of action. Process terms were selected to be specific to the compound's mechanisms
388 of action where applicable.

389 *Mapping biological processes to functional neighborhoods*

390 We expanded an initial standard of 488 Gene Ontology biological process terms annotated to
391 17 functional neighborhoods [31] using a k-nearest-neighbors approach. For each previously
392 unannotated process in our set of processes, we assigned similarity scores for the 3 most similar
393 (Jaccard overlap on gene annotations) processes in the process-neighborhood standard to their
394 respective functional neighborhoods and annotated the new process to the functional
395 neighborhood with the highest sum of similarity scores. In the case of a tie, the process was
396 annotated to both functional neighborhoods.

397 **Predicting the biological processes perturbed by compounds**

398 Our method to predict biological processes perturbed by compounds is described in the
399 recent study from which the chemical-genetic interaction profiles were obtained [6]. We describe
400 here the modifications to this approach to implement the top-k enrichment method for
401 benchmarking.

Given a set of gene-level similarity scores for each compound and a set of gene-process annotations, the enrichment of each process in the set of the top-k most similar genes for each compound was computed. This is reflected in an enrichment factor (the fraction of the *k* selected genes annotated to a particular process divided by the fraction of total genes annotated to that process) and a p-value obtained using a hypergeometric test. These enrichment factors and p-values were substituted in place of the z-scores and p-values obtained using CG-TARGET for subsequent analyses.

Computational evaluation of process predictions

Performance on gold-standard compounds

The predicted perturbed processes for each of the gold standard compounds were sorted, first by their p-value (ascending) and then by their z-score (for CG-TARGET, descending) or enrichment factor (top-100 enrichment, descending), and the rank of each of their gold-standard process annotations was recorded. To assess the significance of each rank, each pair of p-value and z-score was assigned to a new process, the lists re-ordered, and the ranks of each compound's target process re-computed. The empirical p-value for each gold-standard compound-process pair was computed as the number of times the rank from the shuffled processes achieved the same or better rank as the observed rank.

Performance on genetic interaction profiles

We generated a set of simulated chemical-genetic interaction profiles derived from the genetic interaction profiles [6]. Each simulated chemical-genetic interaction profile was a query genetic interaction profile augmented with noise sampled from a Gaussian distribution with a mean of 0 and a variance for each array gene twice that of the same array gene in the genetic interaction dataset. Three simulated profiles were generated based on each query gene, resulting in 4515 total profiles. Because each simulated chemical-genetic interaction profile was derived from a query genetic interaction profile, it inherited the gold standard process annotations from its parent genetic interaction profile in subsequent benchmarking efforts.

We then used CG-TARGET and the top-100 enrichment method to predict perturbed processes for this set of 4515 simulated chemicals x 289 deletion mutants. For each simulated chemical, its top process prediction was compared to the set of inherited gold-standard process annotations, counting as a true positive if the top prediction matched an existing annotation and a

false positive if it did not. Precision-recall curves were then generated by sorting the list of each simulated chemical's top process predictions (p-value ascending, z-score or enrichment factor descending) and computing the precision (true positives / (true positives + false positives)) and recall (true positives) at each point in this list.

The set of true positive process predictions from both methods was mapped to functional neighborhoods via the expanded process-neighborhood mapping ("Mapping biological processes to functional neighborhoods"). The proportion of processes mapped to each neighborhood was used to compute diversity via Shannon Index.

Analysis of process prediction drivers in chemical-genetic interaction data

Given a compound and a predicted process, a profile of "importance scores" describes the contribution of each gene mutant that compound's process prediction. To obtain this score, a Hadamard product (elementwise multiplication) is first computed between the compound's genetic interaction profile and each L_2-normalized genetic interaction profile for which the dot product between them is 2 or greater and the genetic interaction profile is annotated to the process of interest. The final importance profile consists of the mean of each gene's elementwise products across all selected genetic interaction profiles.

Experimental validation of process predictions

Phenotypic assessment of cell cycle

To examine the effect of compounds on arresting cells in G2/M phase, we looked for differences in budding index and cell DNA content between compounds predicted to perturb the cell cycle versus negative control compounds. Nineteen compounds with high-confidence predictions to cell cycle-related biological processes, 14 of them to "spindle assembly checkpoint" were selected for validation, while ten compounds with predictions of false discovery rate $\leq 25\%$ to processes not mapped to the functional neighborhoods of "Cell Cycle Signaling and Progression" and "Mitosis and Chromosome Segregation" (see "Mapping biological processes to functional neighborhoods") were selected as bioactive negative controls. Two compounds predicted to perturb "cell cycle phase" were also tested in these experiments. All compounds were tested at a concentration of 10 μg/mL, which was also the concentration used to obtain their chemical-genetic interaction profiles [6].

To quantify budding index, logarithmically-growing *pdr1*Δ*pdr3*Δ*snq2*Δ cells were transferred to fresh galactose-containing medium (YPGal) containing compounds and incubated at 25^0C for 4 h. The budding status of at least 200 cells was visually determined under the microscope. The percentage of the budded cells in no compound or compound-treated cells was counted.

For flow cytometry analysis, the *pdr1*Δ*pdr3*Δ*snq2*Δ cells were grown in YPGal media in the presence or absence of a compound until log phase and then fixed in 70% ethanol for 1 h at 25^0C. Cells were collected by centrifugation, treated with RNase A solution (0.25 mg/mL in 50 mM Tris pH7.5) for 1.5 h, and 20 µl of 20 mg/ml proteinase K further incubated at 50^0C for 1h. Samples were then stained with propidium iodide, briefly sonicated, and measured using FACSCalibur ver 2.0 (Becton Dickinson, CA, USA).

The proportions of predicted active compounds and negative controls with positive phenotypic results were compared using the prop.test function in R to assess significance.

Tubulin polymerization assay and analysis

We performed in vitro tubulin polymerization assays using the Cytoskeleton fluorescent-based porcine tubulin polymerization assay (BK011P) following manufacturer specifications. We tested the compounds at a concentration of 10 µg/ml, which was identical to the concentration at which they were screened to generate their chemical-genetic interaction profiles. Nine out of the 10 compounds predicted to perturb "tubulin complex assembly" with an estimated false discovery rate of 0% were selected for testing in the tubulin polymerization assay. Twelve compounds with predictions of false discovery rate ≤ 25% to processes not mapped to the "Mitosis and Chromosome Segregation" functional neighborhood were selected as bioactive negative controls.

We used the V_{max} of tubulin polymerization between the tubulin-predicted compounds and the negative controls to determine if the tubulin-predicted compounds inhibited polymerization to a greater degree than the controls. V_{max} for each compound's fluorescence time-course was calculated as the maximum mean change in fluorescence across 8 consecutive time points. V_{max} values were then compared using a Wilcoxon rank-sum test to determine the significance of polymerization inhibition increases caused by the compounds predicted to perturb tubulin complex assembly.

ACKNOWLEDGEMENTS

This work was partially supported by the National Institutes of Health (R01HG005084, R01GM104975) and the National Science Foundation (DBI 0953881). SWS is supported by an NSF Graduate Research Fellowship (00039202), an NIH Biotechnology training grant (T32GM008347), and a BICB one-year fellowship. CLM and CB are fellows in the Canadian Institute for Advanced Research (CIFAR) Genetic Networks Program. Computing resources and data storage services were partially provided by the Minnesota Supercomputing Institute and the UMN Office of Information Technology, respectively.

REFERENCES

1. Giaever G, Shoemaker DD, Jones TW, Liang H, Winzeler EA, Astromoff A, et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat Genet. 1999 Mar;21(3):278–83.

2. Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, et al. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol. 2004 Jan;22(1):62–9.

3. Hoepfner D, Helliwell SB, Sadlish H, Schuierer S, Filipuzzi I, Brachat S, et al. High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions. Microbiol Res. 2014 Mar;169(2-3):107–20.

4. Lee AY, St Onge RP, Proctor MJ, Wallace IM, Nile AH, Spagnuolo PA, et al. Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science. 2014 Apr 11;344(6180):208–11.

5. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Quantitative phenotyping via deep barcode sequencing. Genome Res. 2009 Oct;19(10):1836–42.

6. Piotrowski JS, Li SC, Deshpande R, Simpkins SW, Nelson J, Yashiroda Y, et al. Functional Annotation of Chemical Libraries across Diverse Biological Processes. bioRxiv [Internet]. 2017 Feb 28; Available from: http://biorxiv.org/content/early/2017/02/28/112557.abstract

7. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010 Jan 22;327(5964):425–31.

8. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016 Sep 23;353(6306).

9. Kato N, Takahashi S, Nogawa T, Saito T, Osada H. Construction of a microbial natural product library for chemical biology studies. Curr Opin Chem Biol. 2012 Apr;16(1-2):101–8.

10. Drewry DH, Willson TM, Zuercher WJ. Seeding collaborations to advance kinase science with the GSK Published Kinase Inhibitor Set (PKIS). Curr Top Med Chem. 2014;14(3):340–2.

11. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012 Jan;40(Database issue):D700–5.

12. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015 Jan;43(Database issue):D1049–56.

13. Flanagan PM, Kelleher RJ, Sayre MH, Tschochner H, Kornberg RD. A mediator required for activation of RNA polymerase II transcription in vitro. Nature. 1991 Apr 4;350(6317):436–8.

14. Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol. 2015 Mar;16(3):155–66.

15. Denning DP, Hirose T. Anti-tubulins DEPendably induce apoptosis. Nat Cell Biol. 2014 Aug;16(8):741–3.

16. Jackson JR, Patrick DR, Dar MM, Huang PS. Targeted anti-mitotic therapies: can we improve on tubulin agents? Nat Rev Cancer. 2007 Feb;7(2):107–17.

17. La Regina G, Bai R, Coluccia A, Famiglini V, Pelliccia S, Passacantilli S, et al. New pyrrole derivatives with potent tubulin polymerization inhibiting activity as anticancer agents including hedgehog-dependent cancer. J Med Chem. 2014 Aug 14;57(15):6531–52.

18. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008.

19. Kim D-U, Hayles J, Kim D, Wood V, Park H-O, Won M, et al. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol. 2010 Jun;28(6):617–23.

20. Kapitzky L, Beltrao P, Berens TJ, Gassner N, Zhou C, Wüster A, et al. Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action. Mol Syst Biol [Internet]. 2010 Dec 21 [cited 2017 Feb 28];6. Available from: http://msb.embopress.org/cgi/doi/10.1038/msb.2010.107

21. French S, Mangat C, Bharat A, Côté J-P, Mori H, Brown ED. A robust platform for chemical genomics in bacterial systems. Mol Biol Cell. 2016 Mar 15;27(6):1015–25.

22. Babu M, Arnold R, Bundalovic-Torma C, Gagarinova A, Wong KS, Kumar A, et al. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in Escherichia coli. PLoS Genet. 2014 Feb;10(2):e1004120.

23. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science. 2008 Oct 17;322(5900):405–10.

24. Frost A, Elgort MG, Brandman O, Ives C, Collins SR, Miller-Vedam L, et al. Functional repurposing revealed by comparing S. pombe and S. cerevisiae genetic interactions. Cell. 2012 Jun 8;149(6):1339–52.

25. Ryan CJ, Roguev A, Patrick K, Xu J, Jahari H, Tong Z, et al. Hierarchical modularity and the evolution of genetic interactomes across species. Mol Cell. 2012 Jun 8;46(5):691–704.

567 26. Estoppey D, Hewett JW, Guy CT, Harrington E, Thomas JR, Schirle M, et al. Identification
568 of a novel NAMPT inhibitor by CRISPR/Cas9 chemogenomic profiling in mammalian cells.
569 Sci Rep. 2017 Feb 16;7:42728.

570 27. Blomen VA, Májek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, et al. Gene
571 essentiality and synthetic lethality in haploid human cells. Science. 2015 Nov
572 27;350(6264):1092–6.

573 28. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, et al. Identification and
574 characterization of essential genes in the human genome. Science. 2015 Nov
575 27;350(6264):1096–101.

576 29. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-
577 Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities.
578 Cell. 2015 Dec 3;163(6):1515–26.

579 30. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating
580 high-throughput genomic analysis with Bioconductor. Nat Methods. 2015 Jan 29;12(2):115–
581 21.

582 31. Baryshnikova A. Systematic Functional Annotation and Visualization of Biological
583 Networks. Cell Syst. 2016 Jun 22;2(6):412–21.

584

585 **FIGURE LEGENDS**

586 **Figure 1**. **Overview of genetic interaction-based interpretation of chemical-genetic**
587 **interactions on a large scale.** Chemical-genetic interaction profiles, obtained by measuring the
588 sensitivity or resistance of a library of gene mutants to a particular compound, are compared
589 against genetic interaction profiles consisting of double mutant interaction scores. The resulting
590 similarities are aggregated at the level of biological processes to predict the process(es) perturbed
591 by the compound. Better agreement between chemical-genetic and genetic interaction profiles
592 leads to stronger process predictions.
593
594 **Figure 2. Rate of compound discovery and control of the false discovery rate for the**
595 **prediction of biological processes from chemical-genetic interaction profiles.** Biological
596 processes were predicted for compounds, negative controls (DMSO), and resampled compound
597 profiles from the RIKEN and NCI/NIH/GSK datasets. (A,C,E) The number of compounds,

experimental controls, and resampled compound profiles discovered across different significance thresholds. (B,D,F) The respective translation of panels A, C, E into estimates of the false discovery rate of biological process predictions. (A,B) Estimation of false discovery rate for predictions using CG-TARGET on the RIKEN dataset. (C,D) Same as (A,B), but for the NCI/NIH/GSK dataset. (E,F) Same as (A,B), but for predictions made using the top-100 enrichment-based process prediction method on the RIKEN data.

Figure 3. Performance of biological process prediction for well-characterized and simulated compounds. (A) Biological processes were predicted for compounds in the RIKEN dataset using both CG-TARGET and the top-100 enrichment method. For each of 35 well-characterized compounds in the RIKEN dataset with gold-standard biological process annotations, we determined the rank of its gold-standard process within its list of predictions. The number of compounds for which a rank (or better) was achieved is plotted for each process prediction method. (B) Biological processes were also predicted using both prediction methods for a set of 4500 simulated compounds, each of which was designed to target one query gene in the genetic interaction network. As such, each simulated compound inherited gold-standard process annotations from its target gene, and the ability to recapitulate gold-standard annotations within the set of rank 1 predictions was assessed using precision and recall measures.

Figure 4. Detailed analysis of the contribution of individual gene mutants to biological process predictions. (A) Schematic showing prediction of the "mRNA transport" process for NPD4142. An abbreviated chemical-genetic interaction profile for NPD4142 is shown, aligned with the corresponding array genes from the genetic interaction data and scores that reflect the importance of each of these gene mutants in making the mRNA transport prediction. Query genes were selected from the set of genes mapped to mRNA transport and were sorted from left to right in order of increasing similarity to NPD4142 (gene-level prediction score ≥ 2). Array genes were included if they possessed a chemical-genetic interaction score < -2.5 or variance > 0.02 across the selected query genes. Compounds are sorted from left to right in order of decreasing process prediction strength. (B) Schematic showing the prediction of "proteasome assembly" for several compounds. Query genes (columns) were selected from the set of genes mapped to proteasome assembly and sorted from left to right in order of increasing average similarity with the compounds predicted to "proteasome assembly" (gene-level prediction score ≥ 2). Array genes were included if they possessed a mean chemical-genetic interaction score < -2.5 or variance > 0.02 across the selected query genes in the genetic interaction data. Compounds are sorted from left to right in order of decreasing process prediction strength. Importance scores, as in (A), reflect the importance of each strain for each compound's prediction to proteasome assembly. (C) same as (B), but for the "fungal-type cell wall organization" process. (D) Global characterization of the relationship between chemical-genetic interactions and their importance to process predictions for the RIKEN dataset. Each point compares the interaction score between one gene mutant and one compound to the importance of that gene mutant in predicting the top biological process for that compound. Black points represent gene-compound pairs that possess chemical-genetic interactions (z-score $\geq \pm 2.5$) and contribute nontrivially (importance score $\geq \pm 0.1$) to the compound's top process prediction. Data for predictions made with high confidence (FDR $\leq 25\%$) were included.

Figure 5. In vivo and in vitro experimental validations of biological process predictions. (A,B,C) Phenotypic validation of cell cycle-related predictions, performed on drug-hypersensitive yeast treated with solvent control (DMSO) or compounds predicted to perturb the cell cycle. (A) Differential interference contrast microscopy (DIC) and fluorescence upon DAPI staining showing bud size and DNA localization, respectively, after compound treatment. (B) FACS analysis of cell populations in different cell cycle phases at 0h, 2h, and 4h after compound treatment. The green curve overlay represents the estimated cell population in G1, S and G2/M phases. (C) Budding index percentages induced by treatment with compound or solvent control. (D) In vitro inhibition of tubulin polymerization by compounds predicted to perturb "tubulin complex assembly" (red) compared to negative control compounds (blue). Vmax values reflecting the maximum rate of tubulin polymerization for each compound and negative control are plotted. Assay positive and negative control compounds are colored grey. A Wilcoxon rank-sum test was used to assess the significance of the difference in Vmax between the predicted active compounds and the negative controls.

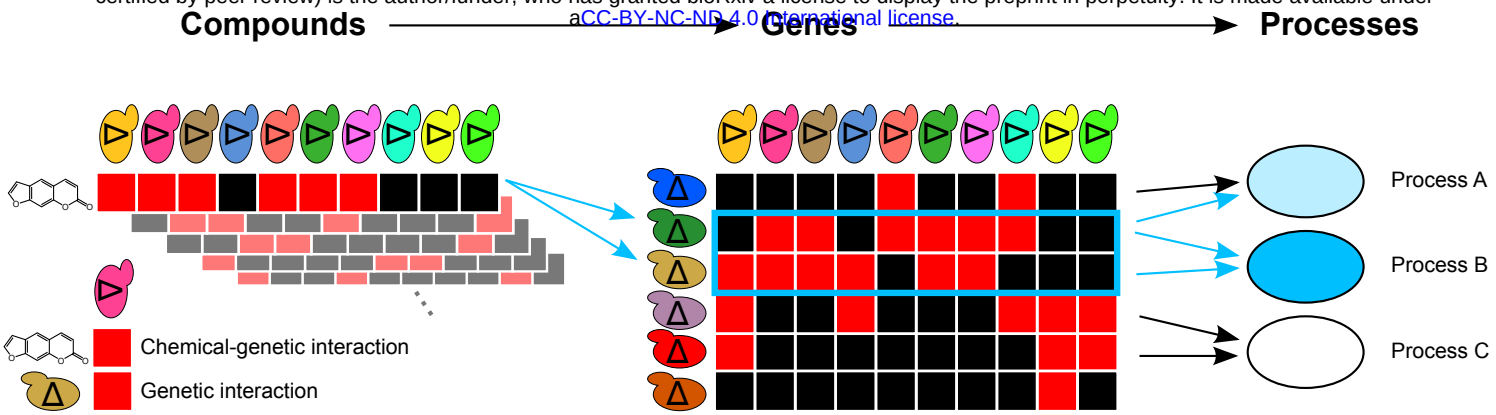


Figure 1

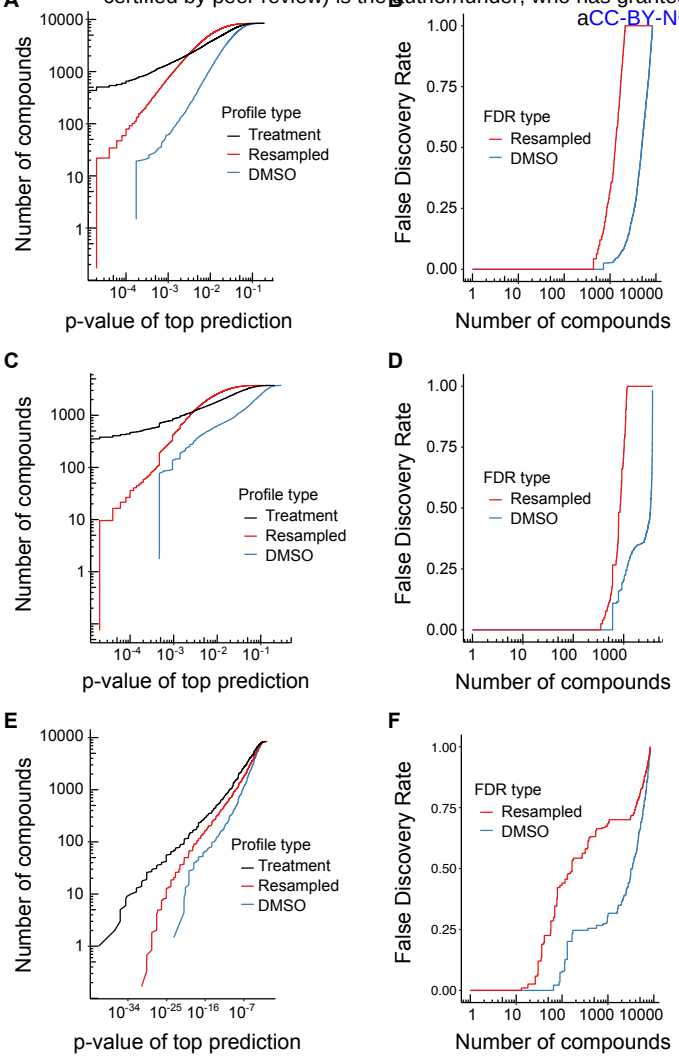


Figure 2

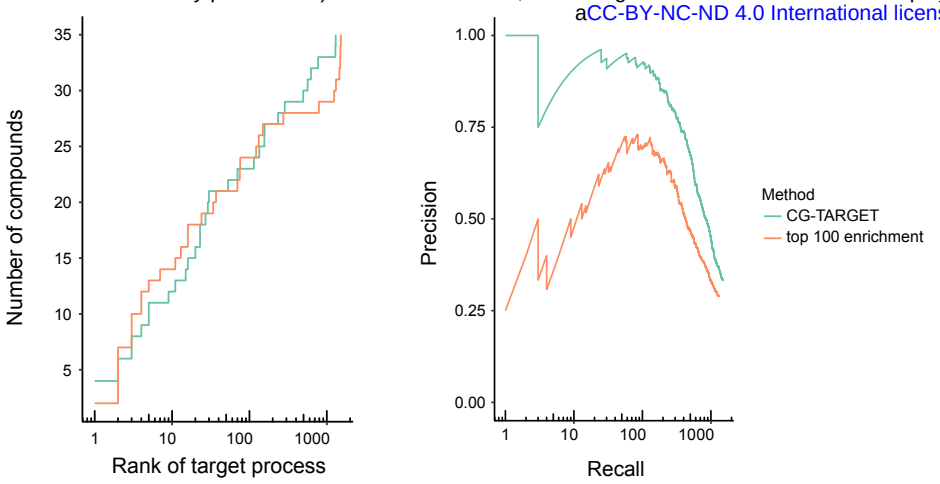
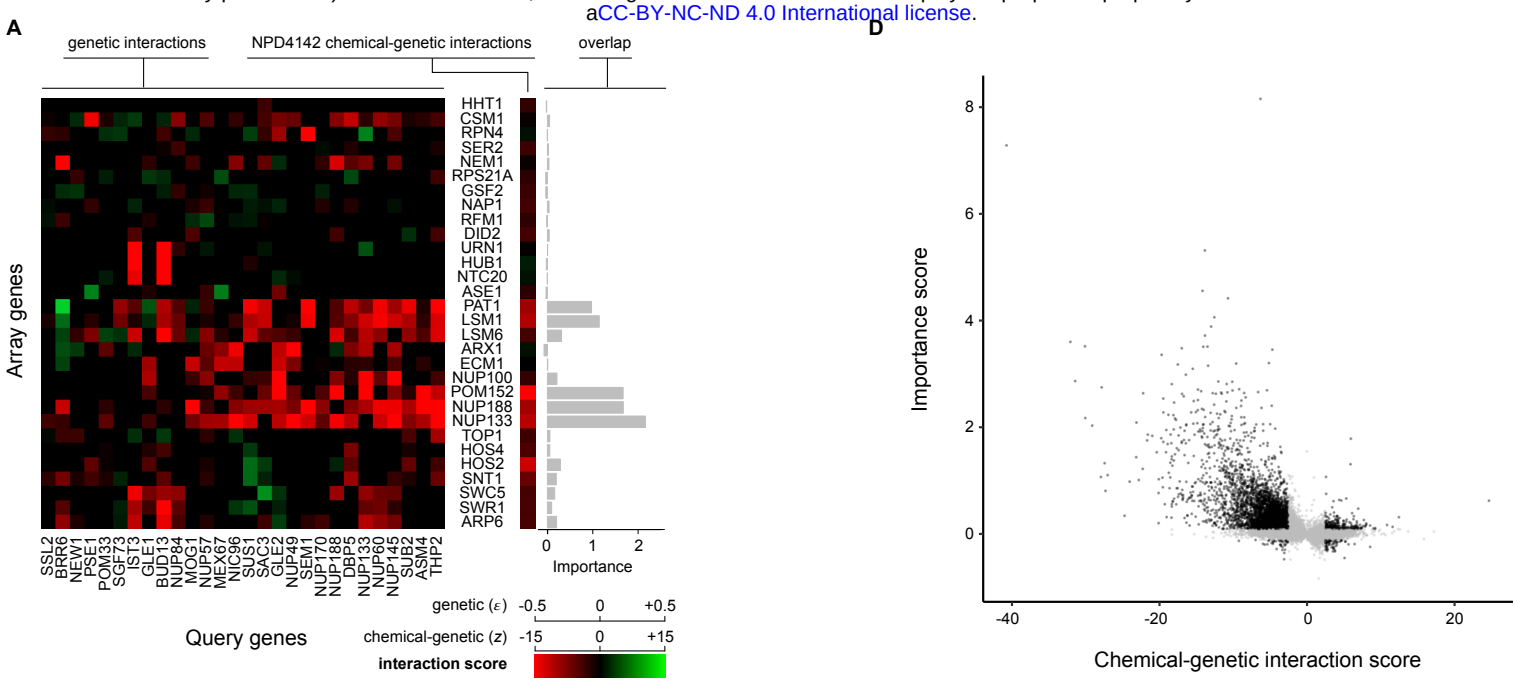
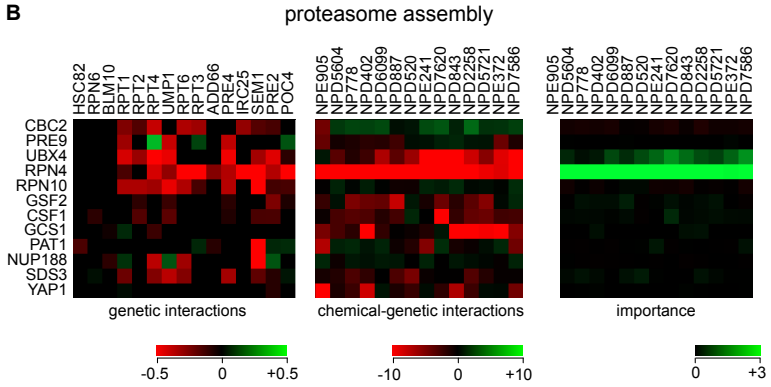


Figure 3





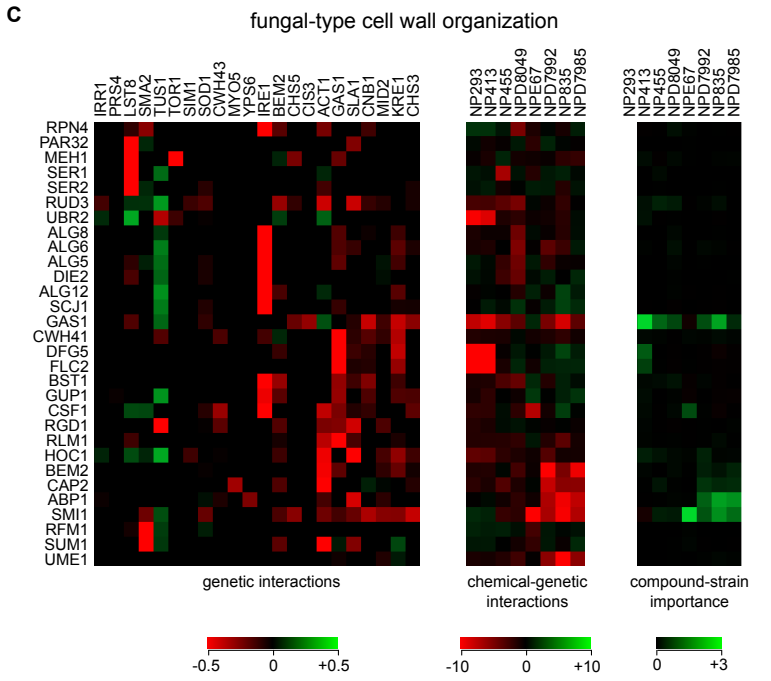
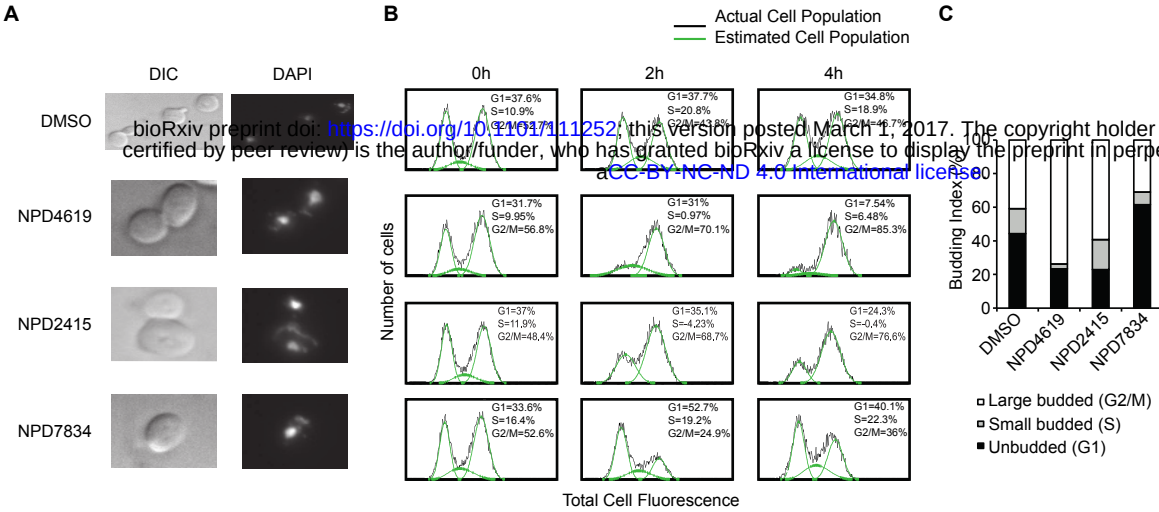


Figure 4



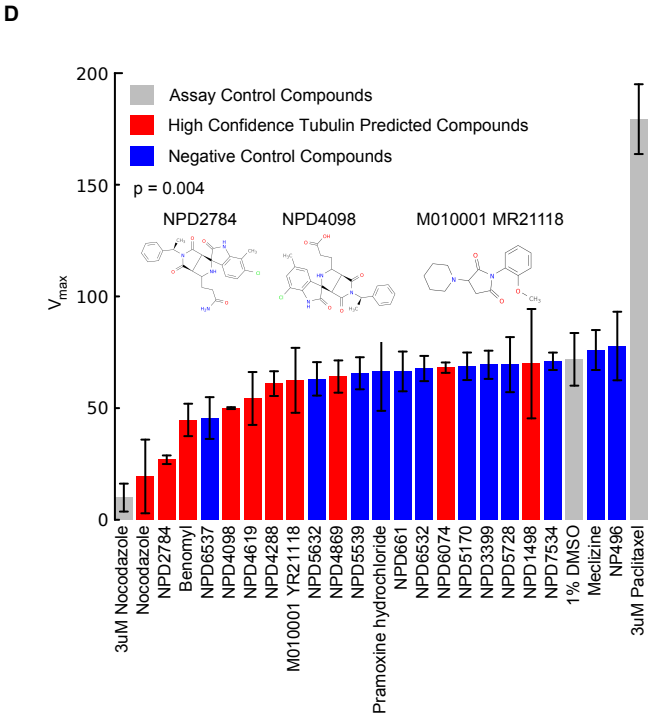


Figure 5

Table 1. Comparison of number of compounds discovered at selected false discovery rate thresholds for CG-TARGET vs. the best-performing enrichment method (top 100 gene target candidates). The CG-TARGET method for predicting chemical-process targets was applied to two large-scale chemical-genetic interaction screens, one of compounds from the RIKEN Natural Product Depository (RIKEN) and the other consisting of 6 chemical compound collections from the National Cancer Institute, National Institutes of Health, and GlaxoSmithKline (NCI/NIH/GSK).

Dataset	RIKEN				NCI/NIH/GSK	
Prediction method	top-100 enrichment		CG-TARGET			
FDR cutoff	p-value	number of compounds	p-value	number of compounds	p-value	number of compounds
0.00	7.4E-32	13	0.00E+00	586	0.00E+00	352
0.05	9.3E-30	26	2.00E-05	610	4.00E-05	405
0.10	5.2E-29	30	8.00E-05	649	1.60E-04	494
0.25*	3.5E-25	57	2.80E-04	848	4.70E-04	705

*This cutoff is 0.27 for the NCI/NIH/GSK dataset

Table 2. Evaluation against literature-derived, gold standard compound-process annotations. The target process rank was determined by its position in the list of all process predictions for each gold standard compound, with the significance computed empirically by shuffling the processes and re-computing the rank (bold p-values indicate significance, $p < 0.05$). Grey boxes indicate cases in which the false discovery rate of the gold standard compound-process prediction was less than 25%. The "effective rank" represents the rank of the process prediction after considering its similarity to processes predicted with higher confidence.

		top-100 enrichment			CG-TARGET		
Chemical	Target GO process	Target process rank	Rank significance	Effective rank	Target process rank	Rank significance	Effective rank
Aclacinomycin A	DNA conformation change (GO:0071103)	11	**0.0076**	1	1	**6.00E-04**	1
Caffeine	TOR signaling cascade (GO:0031929)	16	**0.0113**	4	1	**6.00E-04**	1
Mycophenolic acid	DNA metabolic process (GO:0006259)	1	**6.00E-04**	1	1	**6.00E-04**	1
Tunicamycin	glycosylation (GO:0070085)	4	**0.004**	2	1	**6.00E-04**	1
Benomyl	microtubule-based process (GO:0007017)	2	**0.0015**	1	2	**0.0021**	2
Nocodazole	microtubule-based process (GO:0007017)	16	**0.0099**	4	2	**0.0021**	2
Bortezomib	protein catabolic process (GO:0030163)	4	**0.0033**	1	3	**0.0025**	1
MMS	DNA repair (GO:0006281)	2	**0.0015**	1	3	**0.0025**	1
Hedamycin	DNA repair (GO:0006281)	3	**0.0022**	1	4	**0.0034**	1
Tyrocidine B	cell wall organization or biogenesis (GO:0071554)	24	**0.0165**	5	5	**0.0035**	1
Haloperidol	steroid metabolic process (GO:0008202)	13	**0.0097**	2	5	**0.0043**	2
Oligomycin A	response to pH (GO:0009268)	3	**0.0022**	2	9	**0.0045**	2
Latrunculin B	cytoskeleton organization (GO:0007010)	37	**0.0243**	2	11	**0.0083**	1
Mitomycin C	DNA replication (GO:0006260)	3	**0.0022**	1	15	**0.0093**	4
Camptothecin	DNA conformation change (GO:0071103)	274	0.181	1	16	**0.0142**	4
Furazolidone	DNA replication (GO:0006260)	7	**0.0054**	1	20	**0.014**	4
FK228	chromatin organization (GO:0006325)	2	**0.0015**	1	23	**0.0168**	2
Trichostatin A	chromatin organization (GO:0006325)	5	**0.003**	2	23	**0.0168**	3
5-Fluorocytosine	RNA biosynthetic process (GO:0032774)	2	**0.0015**	2	27	**0.0188**	2
Hydroxyurea	DNA replication (GO:0006260)	76	**0.0478**	2	29	**0.02**	6
Acriflavine	DNA metabolic process (GO:0006259)	1	**6.00E-04**	1	30	**0.0214**	1
Podophyllotoxin	microtubule-based process (GO:0007017)	149	0.0965	8	53	**0.0401**	6
Daunorubicin	DNA replication (GO:0006260)	1245	0.8145	139	70	0.0551	21

Fluconazole	steroid metabolic process (GO:0008202)	70	**0.0481**	6	114	0.0877	12
Cisplatin	DNA replication (GO:0006260)	34	**0.0208**	1	134	0.1025	23
Rapamycin	TOR signaling cascade (GO:0031929)	790	0.5843	7	156	0.1164	8
Nigericin	Golgi vesicle transport (GO:0048193)	2	**0.0015**	1	157	0.1226	13
Itraconazole	steroid metabolic process (GO:0008202)	122	0.08	18	234	0.1748	29
Gramicidin S	cell wall organization or biogenesis (GO:0071554)	75	0.0506	11	286	0.2142	39
Micafungin	cell wall organization or biogenesis (GO:0071554)	132	0.0891	21	495	0.3809	47
Brefeldin A	ER to Golgi vesicle-mediated transport (GO:0006888)	1452	0.9541	196	565	0.4198	32
Calcofluor White	cell wall organization or biogenesis (GO:0071554)	1498	0.9815	177	624	0.4577	90
Blasticidin S	translation (GO:0006412)	1519	0.9962	258	772	0.5742	57
Griseofulvin	microtubule-based process (GO:0007017)	1313	0.8642	244	1291	0.9723	227
Polyoxin D	cell wall organization or biogenesis (GO:0071554)	1490	0.976	127	1302	0.9773	225
	Number with significant rank		23			22	
	Number with significant rank and prediction FDR <= 25%		2			16	