Supplemental Online Information for:

Prediction of compound synergism from chemical-genetic interactions by machine learning

Jan Wildenhain[1], Michaela Spitzer[1,2], Sonam Dolma[3*], David Bellows[3†], Nick Jarvik[3#], Rachel White[1^], Marcia Roy[1&], Emma Griffiths[2§], Gerry Wright[2] and Mike Tyers[4]

Supplemental Experimental Procedures

Yeast strains and Media. All *S. cerevisiae* deletion strains were obtained from the Euroscarf deletion collection (see **Table S3**). The wild type parental strain for this collection is BY 4741 *MAT***a** *his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0. An isogenic *pdr1*Δ *pdr3*Δ strain (MT2481) was created using *PDR1*::*nat* and *PDR*::*URA3* deletion cassettes. All strains were grown and screened in synthetic complete (SC) medium with 2% glucose (Spitzer et al., 2011).

Chemical libraries. Compounds used in this study were from the MicroSource Spectrum library (MicroSource Discovery), LOPAC library (Sigma), Maybridge Hitskit 1000 (Ryan Scientific) and a custom Yeast Bioactive Collection derived from a 53,000 compound synthetic library (Ryan Scientific) (Ishizaki et al., 2010; Zhou et al., 2012). 10 mM compound library stocks were diluted to working stocks of 1 mM in DMSO in 96 well plates. The Microsource Spectrum library was re-purchased twice over the course of this study; variances in the composition of each version resulted in 2300 unique molecules instead of the standard 2000 compounds in each set. The 128 compounds used for the cryptagen matrix were resupplied from MicroSource Discovery. Individual compounds were purchased from the original supplier when available, or from Sigma-Aldrich. All stock solutions were dissolved in DMSO and care was taken to ensure that all compounds remained in solution after dilution in aqueous medium.

Abbreviations for compounds described in the figures and supplementary figures are as follows: Ast - astemizole, Ber – berberine, But – butamben, Cam – camptothecin, Ced – cedrelone, Chr – chrysin, Cic - cyclosporine, Cle - clemistine, Dac – dactinomycin, Dan –

danazol, Deh – dehydroabietamide, Dmf – 3,8-dimethoxyflavone , Ech – echinocystic acid, Eno - enoxolone, Fbz - fenbendazole, Fen - fenamisal, Flu - flufenazine, Flx - fluxetine, Gar – garcinolic acid, Hal - haloperidol, Leo - leoidin, Mec - meclofenamate, Nal – nalidixic acid, Nif - nifedipine, Ost - osthol, Osa – osajin, Oxy - oxybenzone, Par - parthenolide, Peu - peucedanin, Pip - piperine, Rha - rhamnetin, Sal - salinomycin, Sem - semustine, Tox - toxicarol, Xan – xanthyletin.

Chemical-genetic matrix screens. Strains were seeded at 50,000 cells per well in a volume of 100 μL in 96 well plates followed by addition of 2 μL of 1 mM compound stock for final concentration of 20 μM. Screens were conducted in technical duplicate using a Biomek FX liquid handling workstation with an integrated stacker carousel. DMSO solvent only controls and 10 μM cycloheximide positive controls were seeded in columns 1 and 12 of each assay plate. Plates were incubated at 30 °C without shaking for approximately 18 h or until culture saturation was achieved for the solvent controls. Cultures were resuspended by shaking on the robotic platform prior to reading OD600 values on either Tecan M1000 or Tecan Sunrise plate readers (Wong et al., 2013).

All data was subjected to the following analysis workflow: 1. Consistent spatial effects on growth across plates were corrected by Lowess regression using an empirically estimated sliding window of 1/3 and normalized based on plate median. 2. If more than 30% of compounds were active within a plate, data was not Lowess corrected but was normalized to DMSO controls. 3. Median normalization was applied to all plates and experiments. 4. Z-factors for growth inhibition were calculated using the median and the interquartile range (IQR) by fitting a normal distribution with N(1,IQR) to the experimental data. In addition, we calculated the Z-factor, percent inhibition and normalized OD values for manual validation. 5. Data points with high variation between replicates (> 3 MAD) were removed as inconsistent outliers. Each of the 4 source libraries yielded cryptagen compounds: Maybridge Hitskit (27%), the Yeast Bioactive Library (10%), the LOPAC (5%) and the MicroSource Spectrum (18%). All raw and processed data is available online at http://chemgrid.org/cgm.

Cryptagen matrix screen. The 128 cryptagen compounds were selected from chemical structural clusters derived from the Microsource Spectrum library molecules based on hits against 186 of

the 195 sentinel strains and resupplied from MicroSource (Groton, CT). A 128x128 chemical-chemical interaction matrix was generated using a *pdr1*Δ *pdr3*Δ strain (MT2481). Strains were seeded at 50,000 cells per well in a volume of 100 μL in 96 well plates followed by addition of 2 μL of 0.5 mM compound stock for final concentration of 10 μM. Plates were incubated and read as above. All experiments were conducted with either a technical or biological replicate. Several biological repeats were taken over the course of the study for both *pdr1*Δ*pdr3*Δ strain and a wild-type strain (BY4743) to ensure data reproducibility. Measurements were normalized to DMSO controls and data averaged between replicates.

The single concentration Bliss independence model (Bliss, 1939) does not account for possible non-linear concentration effects of either drug, which must be measured over a two dimensional dose-response surface in order to determine Loewe additivity (Loewe, 1928), sometimes calculated as the fractional inhibitory concentration index or FICI (Greco et al., 1995). As the Loewe additivity model requires extensive single and combination drug inhibition measurements, it is not practical for large surveys. Our initial estimates for synergy therefore relied on the Bliss independence model, as calculated from single concentrations for each drug and the combination, as determined using the standard equation $E_{xy} = E_x + E_y - (E_x E_y)$ (Greco et al., 1995). Bliss independence values within 90% density kernel fit represented at best additive effects; based on the density kernel density estimation, values above 0.25 represented synergism and values below -0.18, antagonism.

Derivation of structural correlates of chemical-genetic interactions. The Naïve Bayes classifier was chosen to build a predictor for the primary CGM data in part due to size of the dataset. With approximately 713,000 chemical-genetic relationships, 4900 different compounds, and up to 16,384 bit-segmented chemical structural features per molecule, the CGM dataset is computationally challenging. In general terms, the NBL algorithm performs a special case of a regression analysis given the observation of active compound features *f* for each sentinel strain *s*. Assuming that compound features are independent p(*f*|*s*), we compute the conditional probability for each strain *p*(*s*|*f*). The Naïve Bayes approach is well suited for multi-class learning for complex data and has been proven to be efficient with respect to complexity and predictive performance (Keiser et al., 2009; Besnard et al., 2012; Fernandez-Delgado et al., 2014). For each deletion strain, compounds were classified as active (Z-score < - 4) or non-active. For each

compound, ECFP4 (long) fingerprints based on a circular topological connectivity traverse algorithm were used as implemented in Pipeline Pilot 6.0 (Accelrys). All molecules were stripped of any embedded solvents. Structural features were learned for each deletion strain using the Pipeline Pilot implementation of a multinomial Laplacian-Modified Naïve Bayes learner. For each sentinel Δs a subset of a compounds will show activity out of a total of k, this defines the baseline probability of compounds being active per sentinel:

$$p(\Delta s) = a/k \quad (1)$$

Each compound contains a number of structural features that are either active or non-active in a particular sentinel. The conditional probability for each feature f_i that is present in compounds to be active in Δs, would then be:

$$p(f_i|\Delta s) = a_{fi}/k_{fi} \quad (2)$$

The 'naivety' of this approach treats all features as independent probabilities, so Bayes' rule can be written as:

$$p(s|f) = \frac{1}{\prod_{i=1}^{m} p(f_i)} p(s) \prod_{i=1}^{m} p(f_i|s) \quad (3)$$

where $p(s) = a/k$ and m is the total number of features in the training data. A correction parameter is introduced to avoid over confidence of $p(s|f) = 1$ (if a subset of features that are unique and all found active in the dataset) and to avoid $p(s|f) = 0$ (a single feature f_i may not be present). The probability estimate can be approximated and the formula becomes:

$$p\hat{}(s|f_i) = (F_s + L * p(s))/(F + L) \quad (4)$$

where $L = p(s)^{-1}$ is the Laplace correction and F_s is the active feature count in Δs and F is the total feature count for f_i. The likelihood score E_s for each sentinel strain given a compound is therefore:

$$E_s = \sum_i \log\left(\frac{F_s+1}{F+L}\right) \quad (5)$$

Where i is the sum of all feature log probabilities *1* to m. The validation for each class Δs given a compound c^* was performed using leave-one-out cross-validation as illustrated in **Fig. S5A**. All Δs models and AUC performance values can be found in **Table S6**.

Prediction of chemical synergies from genetic interaction network data (SONAR^G and SONAR^{GN}). The SONAR^G algorithm utilizes bipartite graphs and a second-degree measure to describe the underlying network between two compounds based on genetic interactions. The

algorithm first builds a bipartite graph to narrow in on the most likely targets of two compounds c_i and c_j, and then ranks the genetic interaction edges presumed to mediate a synergistic interaction between c_i and c_j. The target space and target genes for single compounds were identified through the following steps: (1) Sensitive deletion strains V_S for each cryptagen were identified based on CGM data (Z-score ≤ -4 for SONAR^{G}) or the Naïve Bayes likelihoods (using the third quartile, Q3, for SONAR^{GN}). (2) Genetic interactions for the set of sensitive deletion strains V_S were used to identify all neighbors V_T that formed the target space. V_S and V_T are sets of nodes of a bipartite graph that are connected by an edge if a genetic interaction between two genes has been reported (based on BioGRID release number 3.076) (Chatr-Aryamontri et al., 2015). No edge in *E* connects vertices within the same set of nodes. The graph is defined as $G = \langle V_S \cup V_T, E \rangle$, where $V_S = \{s_i | 1 \leq i \leq k\}$ and $V_T = \{t_j | 1 \leq j \leq n\}$. To represent edges within V_S, we introduce the connected nodes into V_T represented as $V_T = \{s_i' \wedge t_j | 1 \leq i \leq m, 1 \leq j \leq n\}$. Each s_i is either sensitive $s_i^+ = 1$ or non-active $s_i^- = 0$ ($s_i^+ = 1$ and $s_i^- = -1$ were also tested but the algorithm did not perform as well). The weight of edge E_{ij} for a pair of nodes s_i, t_j is defined as s_i*m where m is the number of inferred interactions between these nodes based on genetic interaction data. To rank the nodes in V_T, each node t_j is assigned a score that is the sum of the weights of all edges that link to t_j. (3) The sum of the n=35 highest t_j ($hs_V = \sum_1^n t_j$) is used to characterize the target space of each cryptagen. P-values are calculated from 1,000 permutations to estimate a background distribution using as many s_i^- as there are s_i^+ (Spitzer et al., 2011; Dittmar et al., 2013). The value of n was chosen empirically to balance target space size with computational costs. (4) If there is no significant enrichment in hs_V over the background permutations (p-value >= 0.05), return to step 2, remove the weakest bioactive hit and re-run analysis. Continue to loop through steps 2-4 until SONAR hs_V is significant or terminate calculation if total number of s_i^+ is reduced to 4. If a significant gene set s_i^+ is found, this defines target space set t_i^s. (5) The highest ranked genes in t_i^s are most likely to represent actual targets of molecule c_i (see **Table S6**). (6) For each compound pair, a score indicative of an interaction between compounds c_i and c_j is calculated using the target space sets t_i^s and t_j^s. Edges are added between nodes in t_i^s and t_j^s if a genetic interaction has been documented between genes in the two sets. For example, the weight on edge $t_i^s + t_j^s$ is $t_i + t_j$. The score for the target space sets t_i^s and t_j^s is the sum of all edge weights between the two target space sets ($hs_{Exy} =$

$\sum_1^n E_{ij}$, with n=35). A p-value estimate was derived as in step 3 above.

Prediction of chemical synergies using a random forest ensemble learner (SONAR^{GNR} and SONAR^{NR}). Random forest learners are ensembles of decision trees that enable robust predictive performance in machine learning applications and, importantly, require little parameter tuning (Breiman, 2001; Calle and Urrea, 2011; Fernandez-Delgado et al., 2014). The chemical-genetic relationships of the 195 deletion strains in conjunction with the 8128 pairwise compound interactions are computationally less complex than the entire CGM, which invokes a large number of genetic relationships that may or may not be informative. The chemical-genetic response is thus complex and it is not straightforward to link putative targets to the compound action. The non-parametric nature of the random forest approach is preferable given the derived dataset. The data was analysed using R and the packages *FactoMineR* and *corrplot*. Parameter tuning on the training and test data were carried out with the R packages *caret*, *randomForest* and *ROCR*. The SONAR^{GNR} descriptor space used for machine learning contained the Pearson correlation p_{ij} of strain sensitivity between all pairs of compounds, the shared genetic interactions (*sgi*) between target spaces (Spitzer et al., 2011), the sum values hs_{Vx}, hs_{Vy} hs_{Exy} and the p-values for *Vx* and *Vy*. All antagonistic compound pairs were removed and pairs with Bliss independence > 0.25 were considered synergistic. To obtain a balanced data set, 700 molecule pairs each for synergistic and non-synergistic class were selected at random from the CM. The dataset was randomly split into 1/3 training and 2/3 test data. In order to build a predictor and visualize the data, 5-fold cross validation, using 512 trees with 3 variables at each split were used (see **Fig. S3E**). The synergy prediction dataset for SONAR^{NR} was built as illustrated in **Fig. S7A**. The RF algorithm uses 512 trees, randomly tests 17 sentinels at each split and the node size limit is set to a minimum of 14 outcomes per leaf node (**Fig. S7B,C**). Forest assemblies with different numbers of sentinel strains were used to check how many sentinel model parameters are needed to stabilize and minimize the out of bag error rate (**Fig. S7D**).

Dose-response surface assays. Dose-response surfaces for growth inhibition by compound pairs were assessed in wild type (BY4743) and *pdr1*Δ*pdr3*Δ (MT2481) *S. cerevisiae* strains in 96-well flat bottom plates. Strains were seeded at 50,000 cells per well and treated with 2-fold serial dilutions for each compound (1 μM to 128 μM) for all possible combinations in an 8x8 matrix.

Plates were incubated at 30 °C and OD_{600} measurements taken when DMSO controls reached saturation, between 16 h and 24 h after inoculation.

For the pathogenic fungi *Cryptococcus neoformans* (H99), *Cryptococcus gattii* (R265), *Candida albicans* (ATCC#90028), *Candida parapsilosis* (ATCC#90018) and *Aspergillus fumigatus* (Af293) 4x4 mini-checkerboards were performed with compound concentrations adjusted around minimal inhibitory concentrations (MICs) for each species. Strains were grown in 100 μL SC medium in round bottom (*Candida* and *Cryptococcus*) or flat bottom (*A. fumigatus*) 96-well plates. For *Candida* and *Cryptococcus* species, overnight cultures were adjusted to OD_{600} = 0.14, then diluted 1:500 and grown in the presence of compounds or controls for 48 h (*Candida*) and 72 h (*Cryptococcus*) at 30 °C without shaking, followed by OD_{600} determinations of re-suspended cultures. *A. fumigatus* was plated at a density of $1x10^4$ conidia per well and growth was assessed visually after 24 h incubation at 37 °C in the presence of 5% CO_2.

Compound dilution series (1 μM to 128 μM) were prepared in Dulbecco's phosphate-buffered saline (PBS) at pH 7.4 for assay of HeLa cells. Final DMSO concentrations in cell cultures were kept at or below 0.03% for all dilutions. HeLa cells were grown in Phenol Red-free medium using dialyzed fetal bovine serum to avoid potential interactions with estrogen and other hormones. Cells were initially plated at 7000 cells per well in 100 μL MEM (Gibco) in flat bottom 96-well plates and incubated at 37 °C with 5% CO_2 for 72 h prior to addition of diluted compounds. To avoid confounding effects of evaporation, medium containing compounds or solvent control was replaced every 24 h. For HEK293 cells, compounds were diluted in 100 μL MEM for a final DMSO concentration of less than 1%. Cells were initially plated at 5000 cells per well in 100 μL MEM (Gibco), and incubated at 37 °C with 5% CO_2 for 48 h before addition of diluted compounds. After 48 h incubation at 37°C, cell viability was assessed with PrestoBlue (Invitrogen) after 2 h incubation at 37 °C with 5% CO_2 as per instructions of the manufacturer.

General statistical methods. Heatmap representations of CGM, CM and NBL data shown in Figures 1, 2 and 3 were clustered using average linkage hierarchical clustering. Heatmaps generated for Figures S6 and S9 used single linkage hierarchical clustering. Networks in Figures 4D, S6A and S9A were generated using Cytoscape (Su et al., 2014), as well as the Cytoscape

plug-ins BINGO (Maere et al., 2005) and Golorize (Garcia et al., 2007). Graphs were built by obtaining all sentinel strains that were among top 3 predicted compound targets for synergistic compound pairs (SONAR^{GNR} network) or top 3 sensitive sentinel strains (SONAR^{NR} network). Edge weights were based on SONAR^{GNR}-derived parameter hs_{Exy} or NBL likelihood of sensitivity to compounds. Only the top 100 edges are shown. Graphs were corrected by subtraction of average weights for the same graph sampled from 730 non-synergistic pairs. All genetic interactions were drawn from BioGRID release 3.076 (Chatr-Aryamontri et al., 2015) and annotated protein complexes in Figure S1D were derived from a previous study (Pu et al., 2009).

Supplemental Tables

All supplemental tables are available at http://chemgrid.org/cgm/index.php.

Table S1. CGM Library Composition, Related to Figure 1

Table S2. Unique Compounds in CGM, Related to Figure 1

Table S3. *S. cerevisiae* Sentinel Strains Used in this Study, Related to Figure 1

Table S4. Information on the 128 Cryptagens in the CM, Related to Figure 2

Table S5. Literature Curation of Yeast Targets for 27 Characterized Compounds and SONAR^{G} Target Predictions, Related to Figure 3

Table S6. Area Under the Curve (AUC) for Each Deletion Strain Class (Sentinel) using a Naïve Bayes Multinomial Classifier, Related to Figure 3

Table S7. Bliss Independence and SONAR^{NR} Synergy Scores for Pairwise Combinations in the CM, Related to Figure 4

Table S8. Confusion Matrices and Prediction Statistics for SONAR^{NR}, Related to Figure 4

Table S9. Dose-Response Surface Verification Data for *S. cerevisiae*, Related to Figures 5, 6, S10. Drug pump deficient strain, 163 pairs; wild type strain, 83 pairs.

Table S10. Bliss Independence Scores from Dose-Response Surfaces in *S. cerevisiae*, Related to Figures 5, 6 and S10

Supplemental References

Besnard, J., Ruda, G.F., Setola, V., Abecassis, K., Rodriguiz, R.M., Huang, X.P., Norval, S., Sassano, M.F., Shin, A.I., Webster, L.A., et al. (2012). Automated design of ligands to polypharmacological profiles. Nature *492*, 215-220.

Bliss, C.I. (1939). The toxicity of poisons applied jointly. Ann Appl Biol *26*, 585-615.

Breiman, L. (2001). Random forests. Machine Learn *45*, 5–32.

Calle, M.L., and Urrea, V. (2011). Letter to the editor: Stability of Random Forest importance measures. Brief Bioinform *12*, 86-89.

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., et al. (2015). The BioGRID interaction database: 2015 update. Nucleic Acids Res *43*, D470-478.

Dittmar, J.C., Pierce, S., Rothstein, R., and Reid, R.J. (2013). Physical and genetic-interaction density reveals functional organization and informs significance cutoffs in genome-wide screens. Proc Natl Acad Sci U S A *110*, 7389-7394.

Fernandez-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classfiers to solve real world classification problems? J Mach Learn Res *15*, 3133-3181.

Garcia, O., Saveanu, C., Cline, M., Fromont-Racine, M., Jacquier, A., Schwikowski, B., and Aittokallio, T. (2007). GOlorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. Bioinformatics *23*, 394-396.

Greco, W.R., Bravo, G., and Parsons, J.C. (1995). The search for synergy: a critical review from a response surface perspective. Pharmacol Rev *47*, 331-385.

Ishizaki, H., Spitzer, M., Wildenhain, J., Anastasaki, C., Zeng, Z., Dolma, S., Shaw, M., Madsen, E., Gitlin, J., Marais, R., et al. (2010). Combined zebrafish-yeast chemical-genetic screens reveal gene-copper-nutrition interactions that modulate melanocyte pigmentation. Dis Model Mech *3*, 639-651.

Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al. (2009). Predicting new molecular targets for known drugs. Nature *462*, 175-181.

Loewe, S. (1928). Die quantitativen problems der pharmakologie. Ergbnisse der Physiologie *27*, 47-187.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448-3449.

Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S.J. (2009). Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res *37*, 825-831.

Spitzer, M., Griffiths, E., Blakely, K.M., Wildenhain, J., Ejim, L., Rossi, L., De Pascale, G., Curak, J., Brown, E., Tyers, M., and Wright, G.D. (2011). Cross-species discovery of syncretic drug combinations that potentiate the antifungal fluconazole. Mol Syst Biol *7*, 499.

Su, G., Morris, J.H., Demchak, B., and Bader, G.D. (2014). Biological network exploration with cytoscape 3. Curr Protoc Bioinformatics *47*, 8.13.11-18.13.24.

Wong, L.H., Unciti-Broceta, A., Spitzer, M., White, R., Tyers, M., and Harrington, L. (2013). A yeast chemical genetic screen identifies inhibitors of human telomerase. Chem Biol *20*, 333-340.

Zhou, L., Ishizaki, H., Spitzer, M., Taylor, K.L., Temperley, N.D., Johnson, S.L., Brear, P., Gautier, P., Zeng, Z., Mitchell, A., et al. (2012). ALDH2 mediates 5-nitrofuran activity in multiple species. Chem Biol *19*, 883-892.

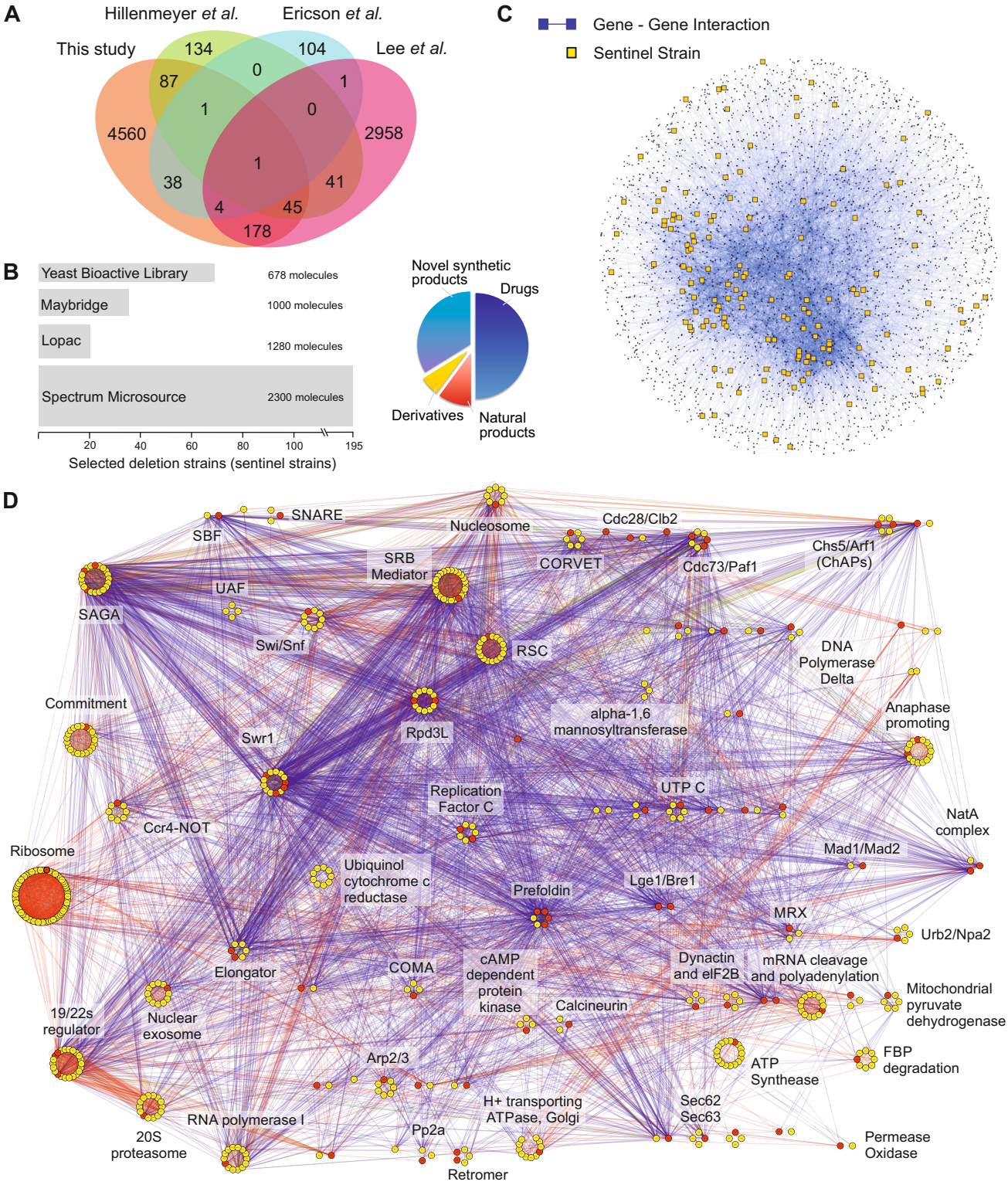

Figure S1. Sentinel Strains and Compound Libraries Screened in this Study, Related to Figure 1. **(A)** Overlap of compounds used in this study and previously published large-scale studies that have explored chemical-genetic interactions in *S. cerevisiae* (Ericson et al., 2008; Hillenmeyer et al., 2008; Lee et al., 2014). **(B)** Distribution of compound classes in libraries used in this study and number of sentinel strains screened against the four different chemical libraries. **(C)** Distribution of sentinel strains in genetic synthetic lethal interaction space based on a single systematic survey (Costanzo et al., 2010). Edges indicate genetic interactions and yellow squares sentinel strains used in this study. **(D)** Protein complexes targeted with sentinel deletion strains. Red lines indicate physical interactions, blue lines synthetic lethal genetic interactions. Red nodes represent deletion strains used in this study. Protein complexes were taken from (Pu et al., 2009).

Figure S1

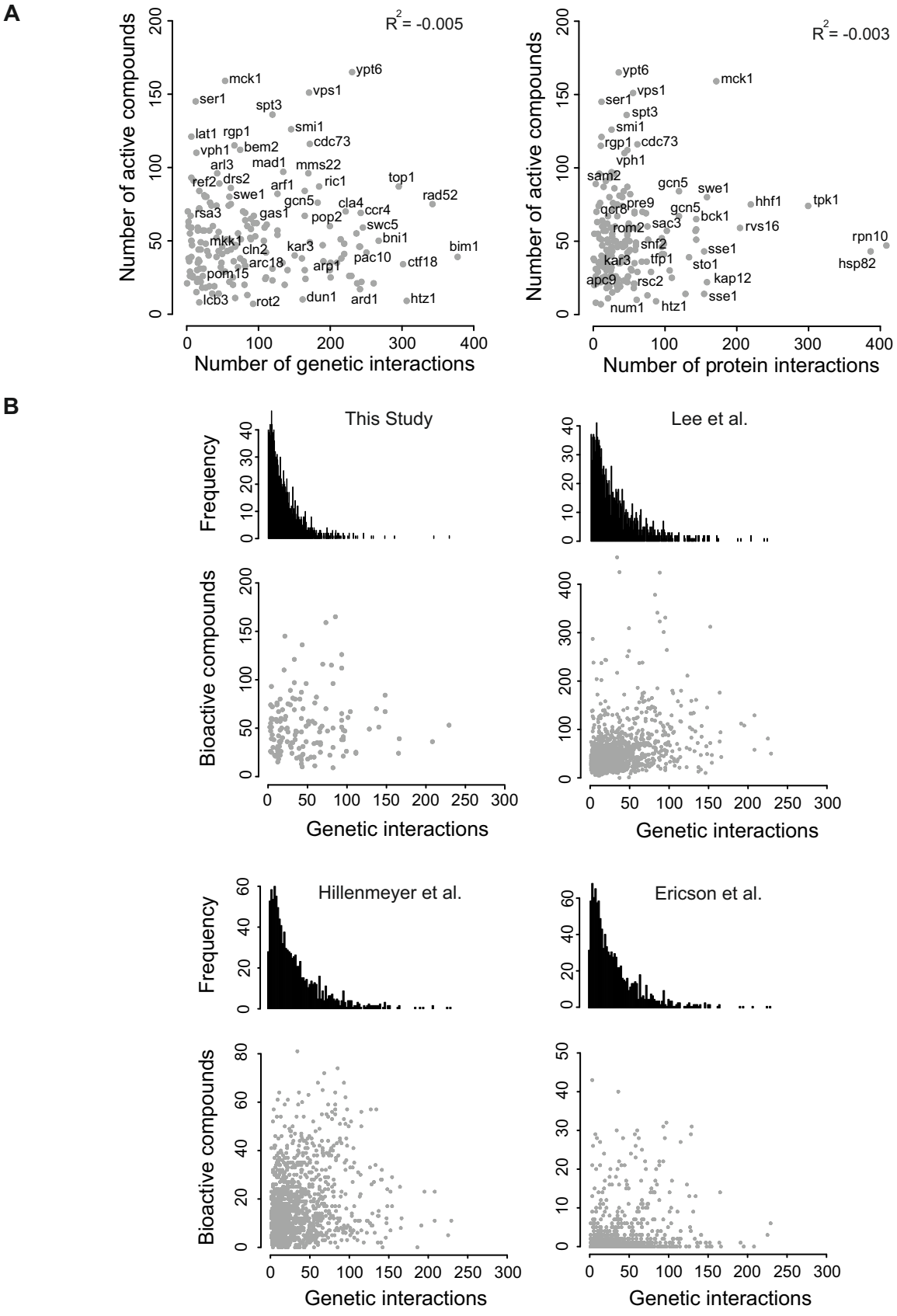

Figure S2. Comparison of Genetic and Chemical-Genetic Interactions of Sentinel Strains, Related to Figure 1. **(A)** Plots of chemical-genetic interactions versus genetic or protein interactions for each sentinel strain. **(B)** Scatterplot and Histograms showing gene-gene interactions (Costanzo et al., 2010) compared to chemical-genetic interactions reported in this study and previously published large-scale studies (Ericson et al., 2008; Hillenmeyer et al., 2008; Lee et al., 2014). Histograms indicate the gene-gene connectedness.

Figure S2

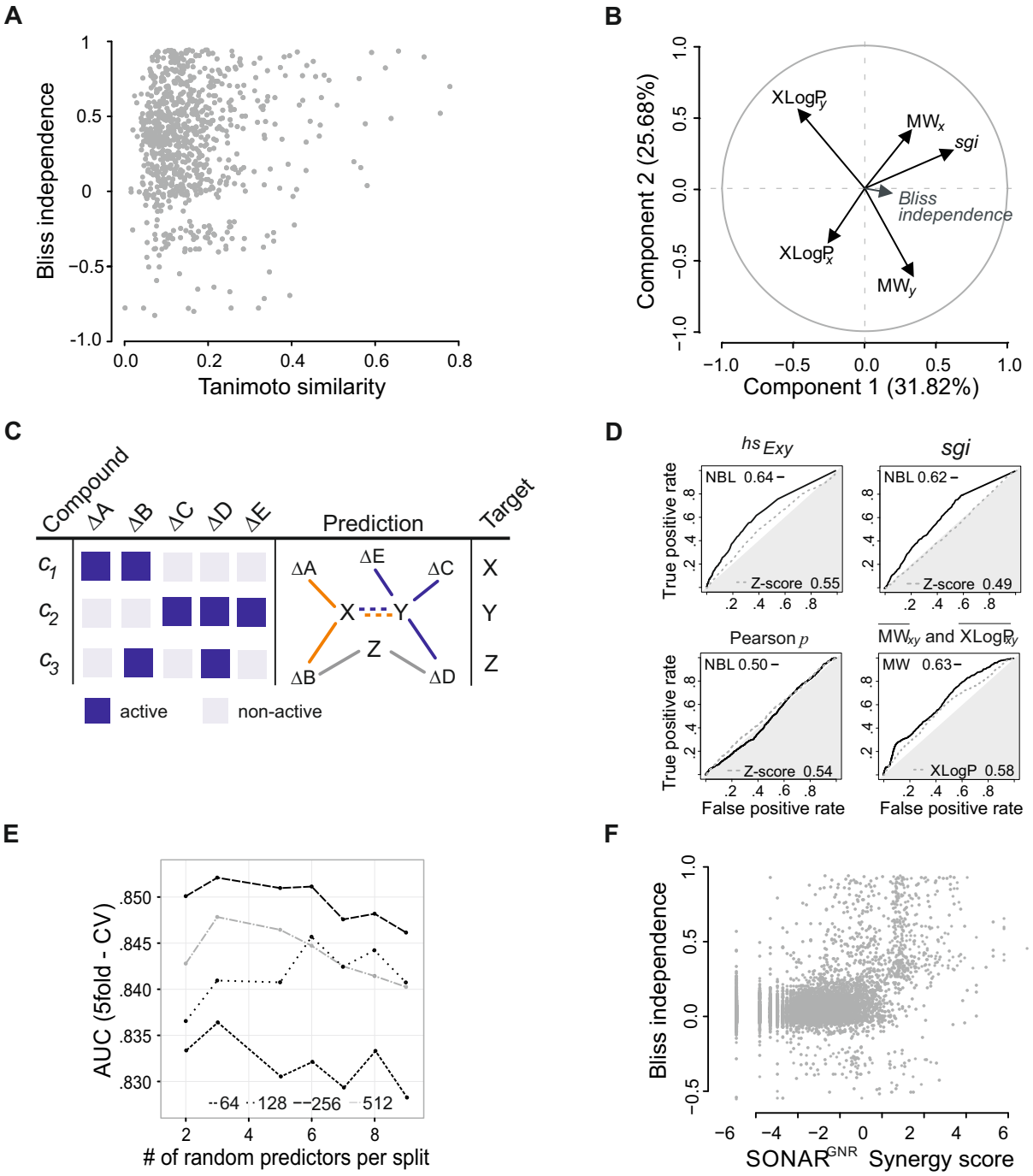

Figure S3. Physico-Chemical Properties of Synergistic Compound Pairs, Related to Figures 2, 3. **(A)** Structural similarity (Tanimoto) of compound pairs versus compound pair synergism assessed by Bliss independence. **(B)** Loading vectors for lipophilicity (*XLogP*), molecular weight (*MW*) and shared genetic interactions (*sgi*) compared to Bliss independence. **(C)** Small molecule target prediction based on chemical-genetic and genetic interactions shown in Fig. 3A and B. **(D)** AUC values for hs_{Exy}, *sgi* and Pearson correlation parameters for $SONAR^{G}$ (dashed grey line) and $SONAR^{GNR}$ (solid black line). The average molecular weight MW_{xy} and partition coefficient $XLogP_{xy}$ for all compound pairs in the CM are provided as reference. **(E)** Training data tested for different random splits and forest sizes. 512 trees with 3 random splits were chosen as parameters for the final algorithm. **(F)** Scatterplot of experimental Bliss independence values and $SONAR^{GNR}$ synergy scores.

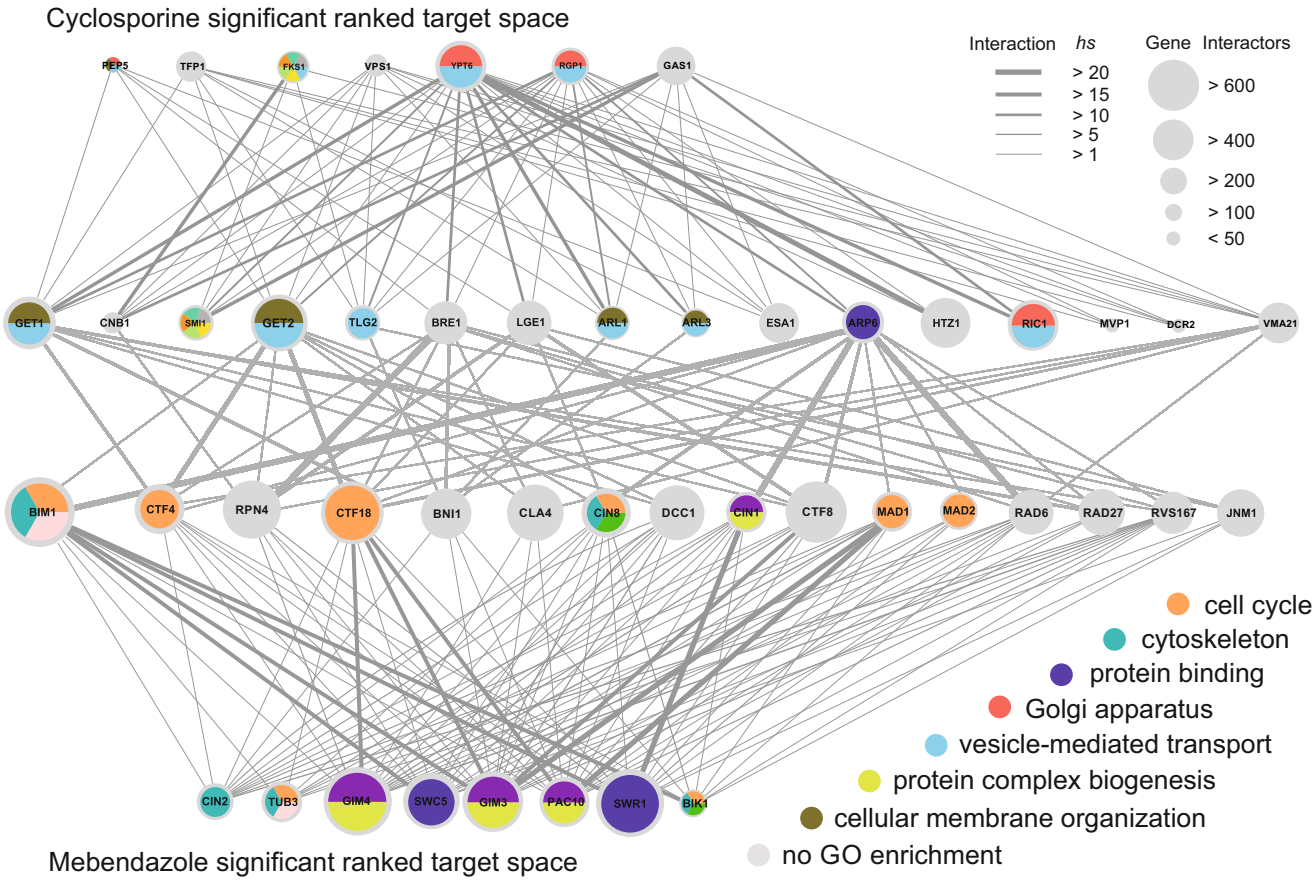

Figure S4. Example of a SONAR^{G} Network Graph, Related to Figure 3. The SONAR^{G} algorithm draws on chemical-genetic interaction target space of each compound and the shared genetic interaction space between the two compounds, as illustrated for the non-synergistic pair cyclosporine and mebendazole for a restricted target space of 16 nodes.

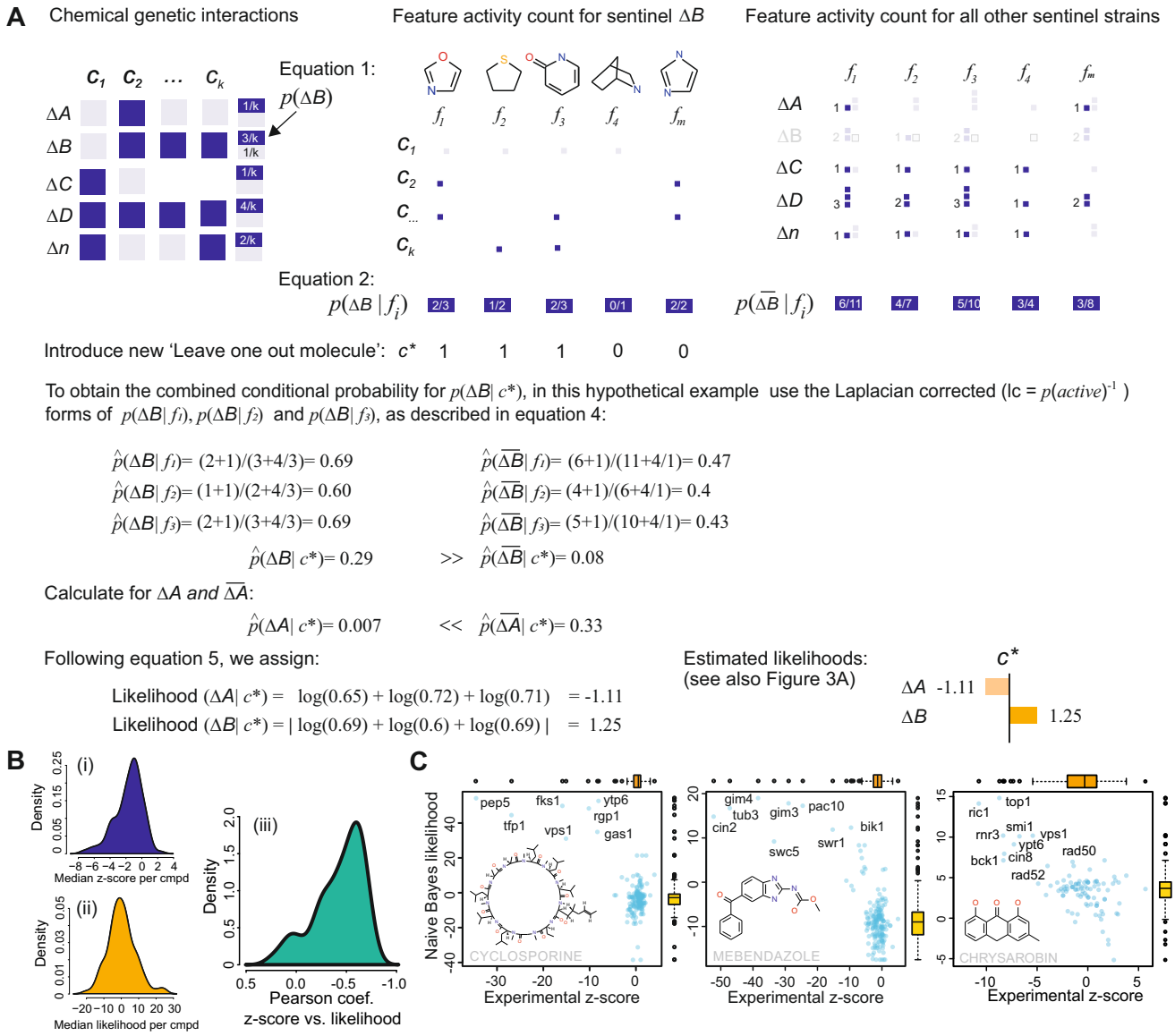

Figure S5. A Structure-Based Naïve Bayes Multi-Nomial Learner (NBL) to Predict Chemical-Genetic Interactions, Related to Figure 3. **(A)** Design scheme for Naïve Bayes multi-class learner used in this study. Further details are provided in the Supplemental Methods. **(B)** (i) Distribution of median experimental Z-scores for growth inhibition by cryptagens over all sentinel strains. (ii) Distribution of median likelihoods for each cryptagen; activity was defined as median Z-score < -4. (iii) Distribution of Pearson correlation coefficients between the likelihood scores and Z-scores across all sentinel strains for each compound. **(C)** Examples of sentinel strain sensitivity likelihood scores versus experimental Z-scores for cyclosporine, mebendazole and chrysarobin. Cyclosporine inhibits calcineurin, which is required for survival of cell wall and cationic stress, and as expected conferred sensitivity to strains deleted for genes implicated in cell wall biosynthesis (*FKS1*, *GAS1*), vacuolar function (*PEP5*, *VMA1*), and vesicle trafficking (*VPS1*, *YPT6*, *RGP1*). Mebendazole targets microtubules and sensitized strains defective in different aspects of microtubule and spindle function (*BIK1*, *CIN2*, *TUB3*, *GIM3*, *GIM4*, *PAC10*) as well as strains deleted for *SWR1* or *SWC5*, which encode components of the SWR chromatin remodeling complex. Chrysarobin generates reactive oxygen species and elicited a complex chemical-genetic profile that included strains disrupted for DNA damage and repair (*TOP1*, *RNR3*, *RAD50*, *RAD52*), vesicle trafficking (*YPT6*, *RIC1*, *VPS1*), cell wall integrity (*BCK1*, *SMI1*) and spindle assembly (*CIN8*). **(D)** Plots of sentinel strain sensitivity Naïve Bayes likelihood scores versus experimental Z-scores from the CGM for 56 compounds (next two pages).

Figure S5 A-C

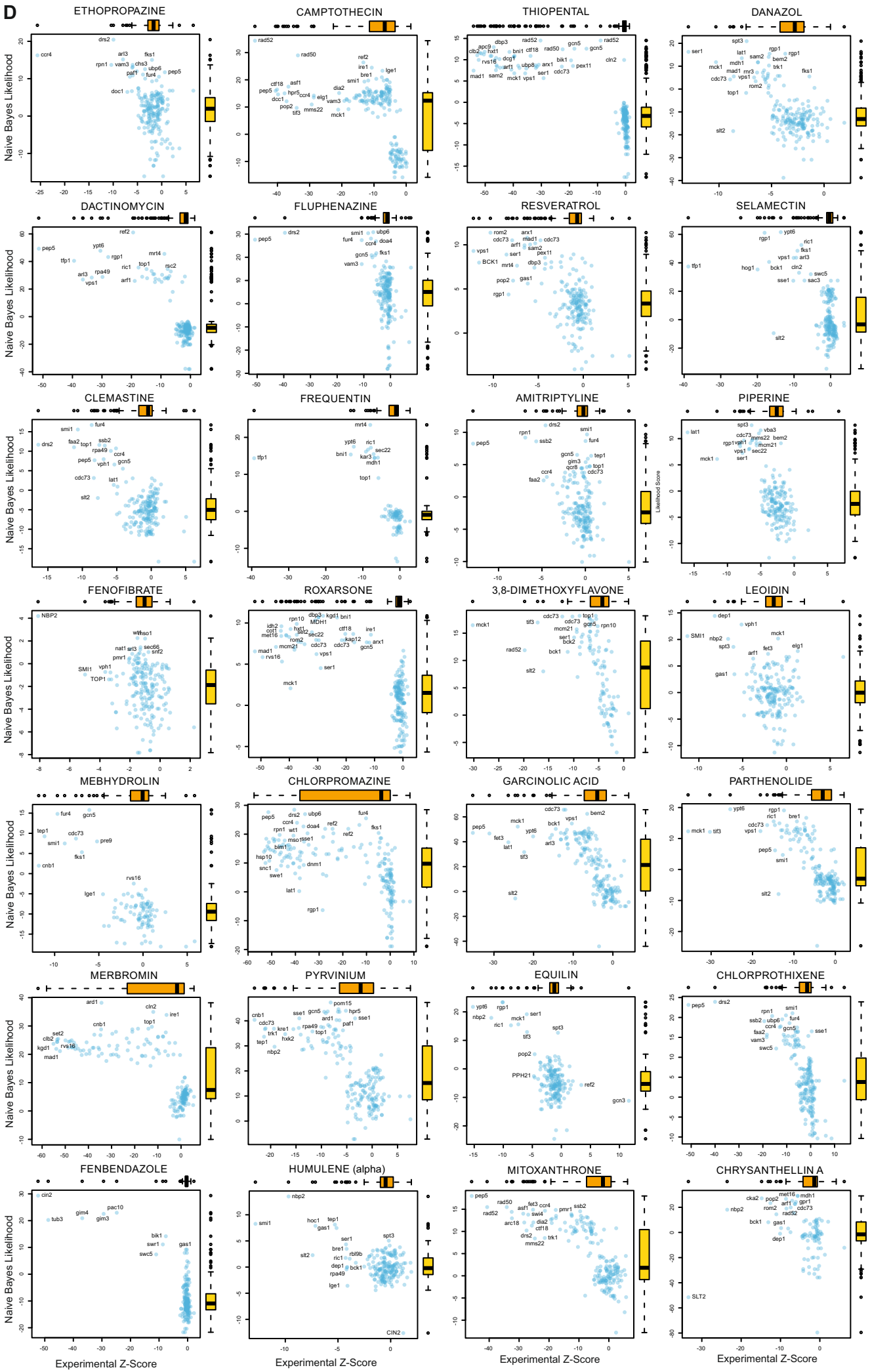
D
ETHOPROPAZINE
CAMPTOTHECIN
THIOPENTAL
DANAZOL
DACTINOMYCIN
FLUPHENAZINE
RESVERATROL
SELAMECTIN
CLEMASTINE
FREQUENTIN
AMITRIPTYLINE
PIPERINE
FENOFIBRATE
ROXARSONE
3,8-DIMETHOXYFLAVONE
LEOIDIN
MEBHYDROLIN
CHLORPROMAZINE
GARCINOLIC ACID
PARTHENOLIDE
MERBROMIN
PYRVINIUM
EQUILIN
CHLORPROTHIXENE
FENBENDAZOLE
HUMULENE (alpha)
MITOXANTHRONE
CHRYSANTHELLIN A
Naive Bayes Likelihood
Likelihood Score
Experimental Z-Score

Figure S5D

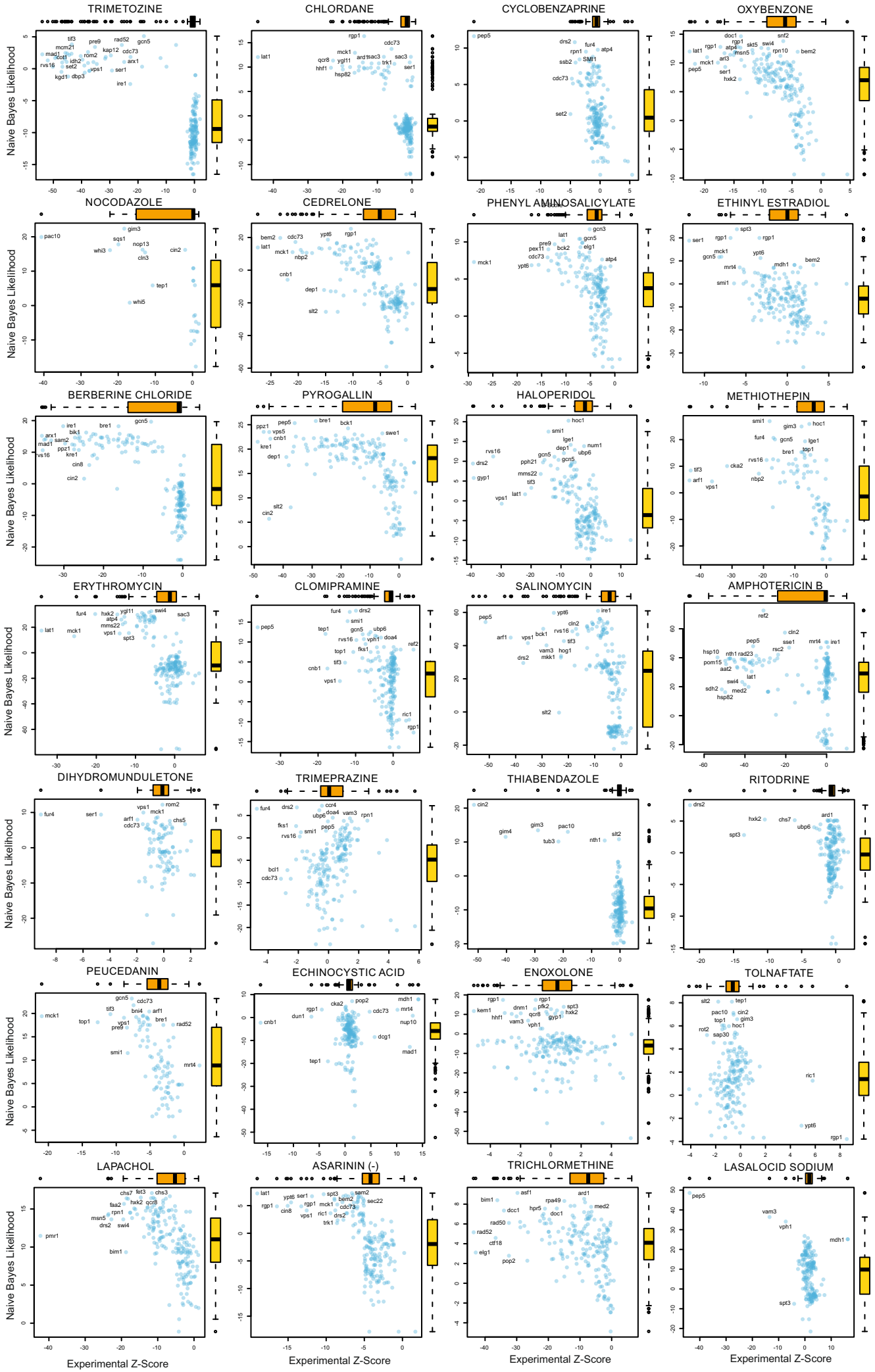

Figure S5D (continued)

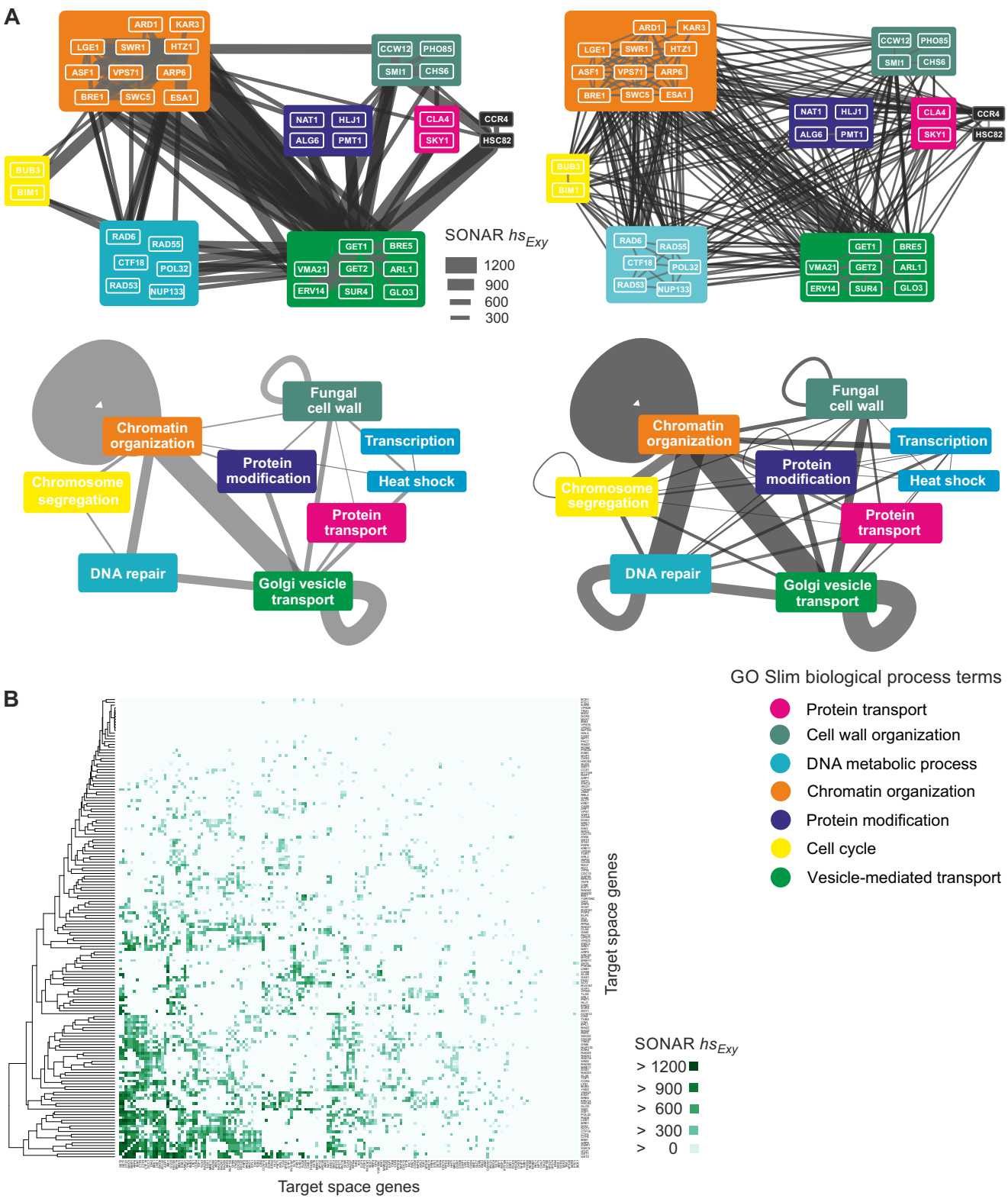

Figure S6. SONAR^{GNR}-Derived Networks Underpinning Small Molecule Synergies, Related to Figure 3. **(A)** Top left: Connectivity between SONAR^{GNR}-predicted compound target genes (top 3 genes per compound) is shown for synergistic compound pairs. Edge weights are based on the SONAR^{GNR}-derived parameter hs_{Exy}. Only the top 100 edges are shown. The data was normalised against target spaces from non-synergistic compound pairs in the CM. Top right: Corresponding genetic interaction network for the genes in left panel based on data from BioGRID (Chatr-Aryamontri et al., 2015). Bottom left: Same as top left, but genes and edges were grouped by biological processes. Bottom right: Same as top right, but genes and edges were grouped by biological processes. **(B)** Heatmap of hs_{Exy} values between the top 15 target space genes present in all synergistic combinations.

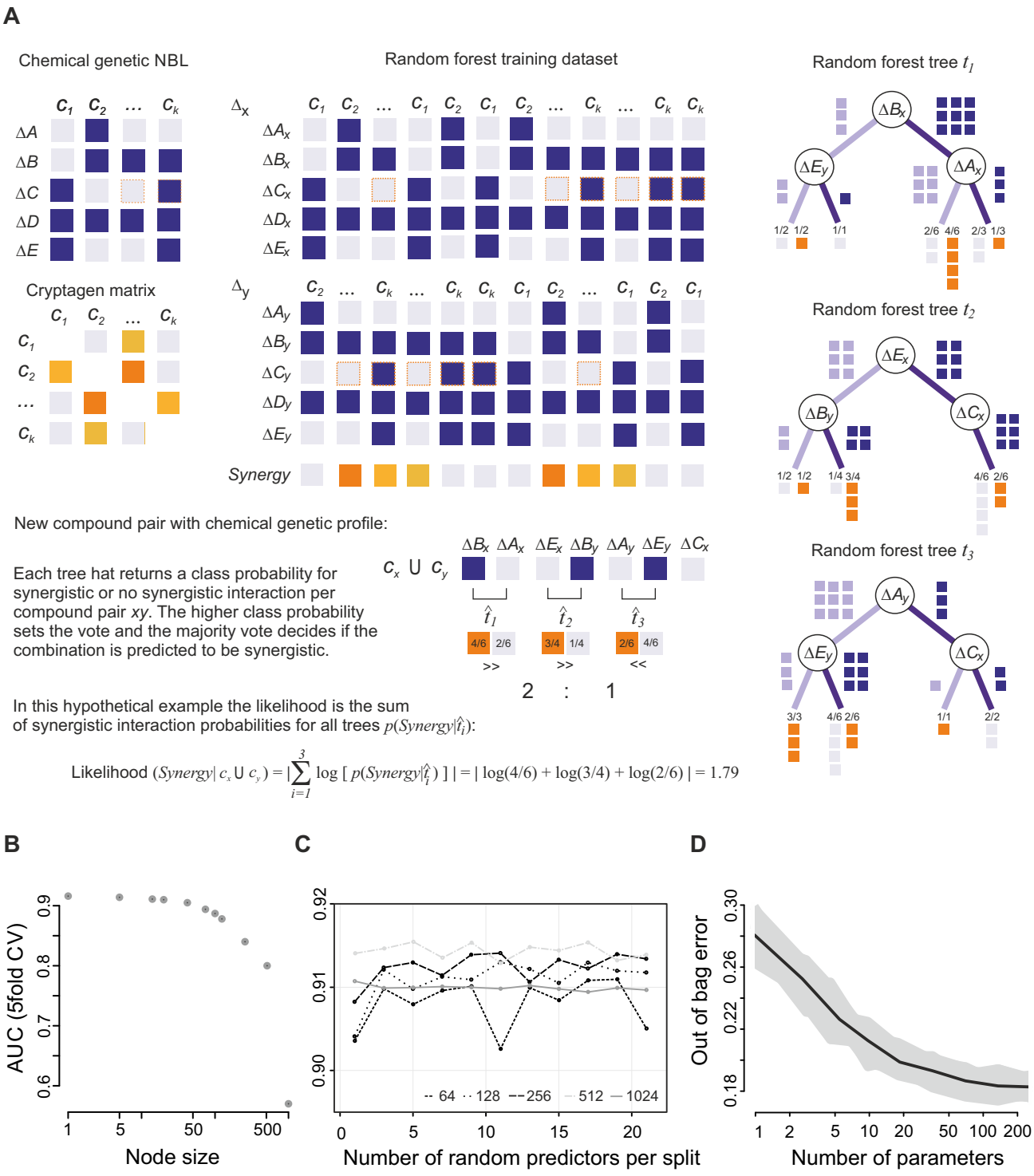

Figure S7. Construction of Random Forest Classifiers, Related to Methods. **(A)** Overview of random forest classifier procedures used in this study. Further details are provided in the Supplemental Methods. **(B)** Assessment of random forest performance (AUC) with restricted minimum data points in each terminal node. To balance tree depth and accuracy, a node size of 14 was chosen for further processing. **(C)** Evaluation of different split parameters for forest sizes between 64 and 1024 on training data. 512 trees with 17 random splits were chosen as parameters for further processing. **(D)** Performance of bootstrapped random forest classifiers with different numbers of input parameters (sentinel genes). Black line indicates the average out of bag error.

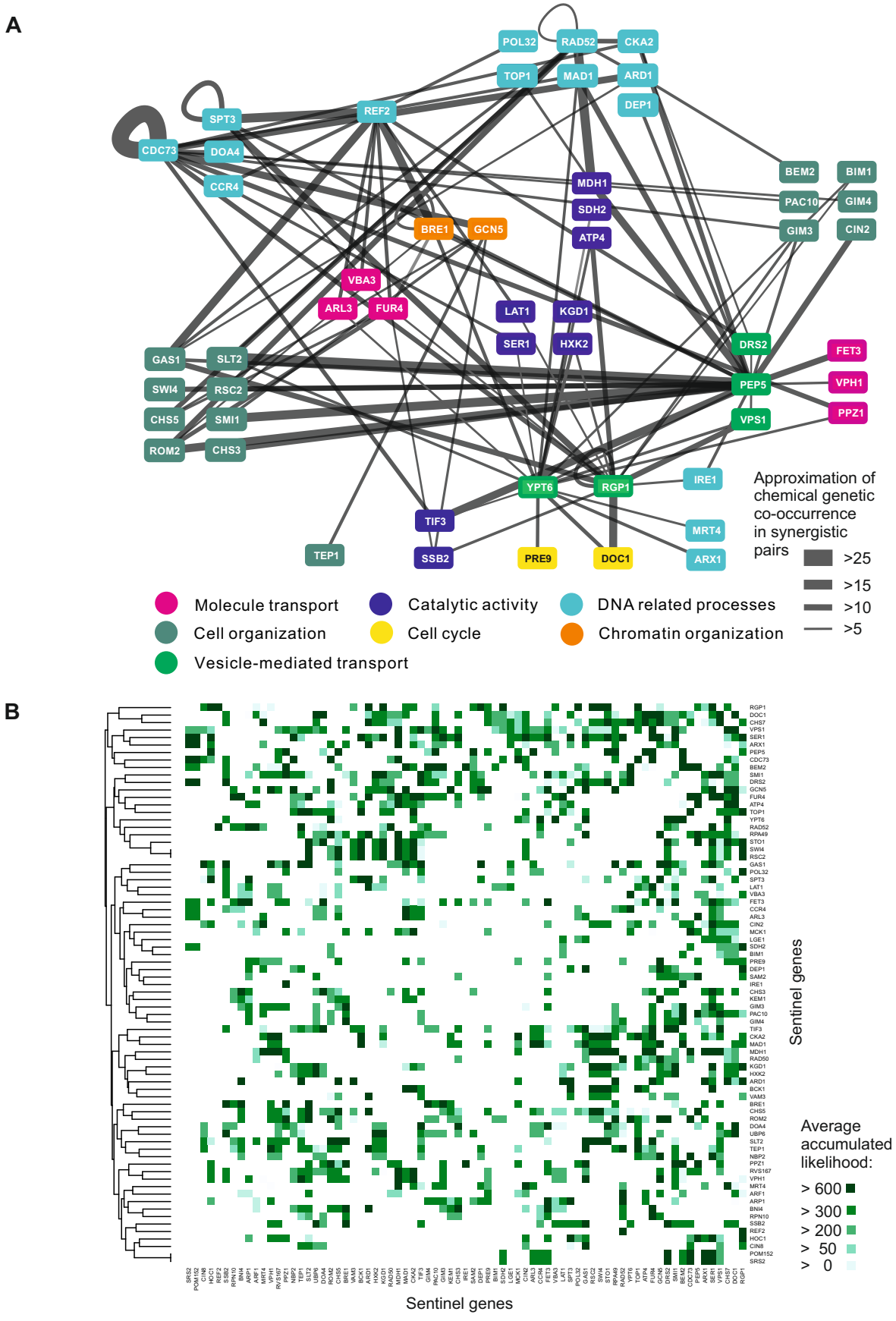

Figure S8. Sentinel Strains Associated with Synergistic Compound Interactions, Related to Figure 4. **(A)** Edge weights were determined as the frequency of two genes observed among the top 3 sensitive genes for synergistic compounds pairs and were background corrected against the same graph sampled from 730 non-synergistic compound pairs. Only the top 100 edges are shown in the graph. **(B)** Heatmap of all gene-gene associations bridged by synergistic compound pairs.

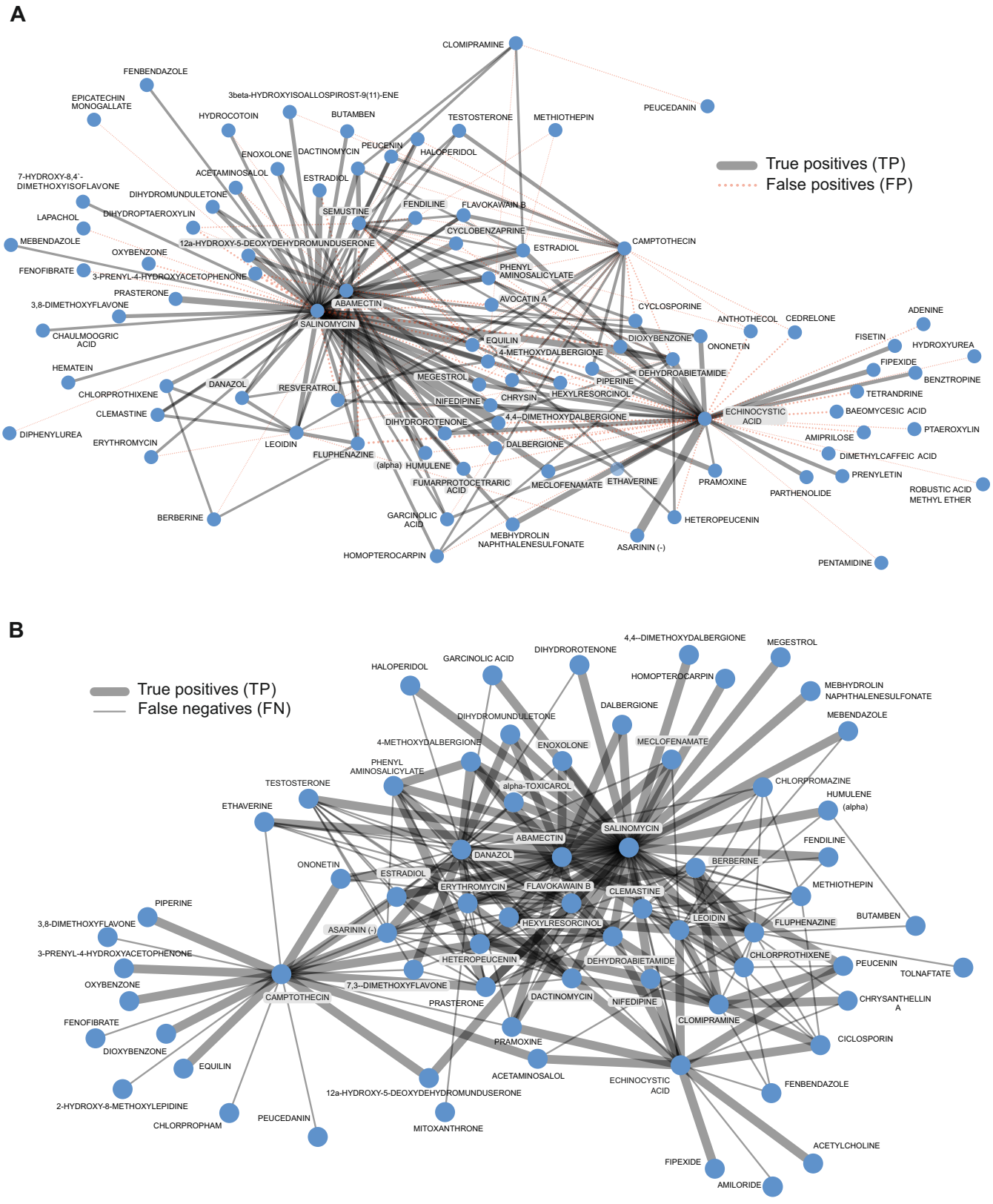

Figure S9. Top 200 SONAR^{NR}-Predicted and Experimental Synergistic Compound Pairs, Related to Figure 4. **(A)** Top 200 ranked SONAR^{NR} predicted synergy scores overlaid with true positives (TP) and false positives (FP) based on the CM. **(B)** Top 200 ranked synergistic Bliss independence pairs overlaid with the predicted true positives (TP) and false negatives (FN) based on SONAR^{NR} synergy score > 1.

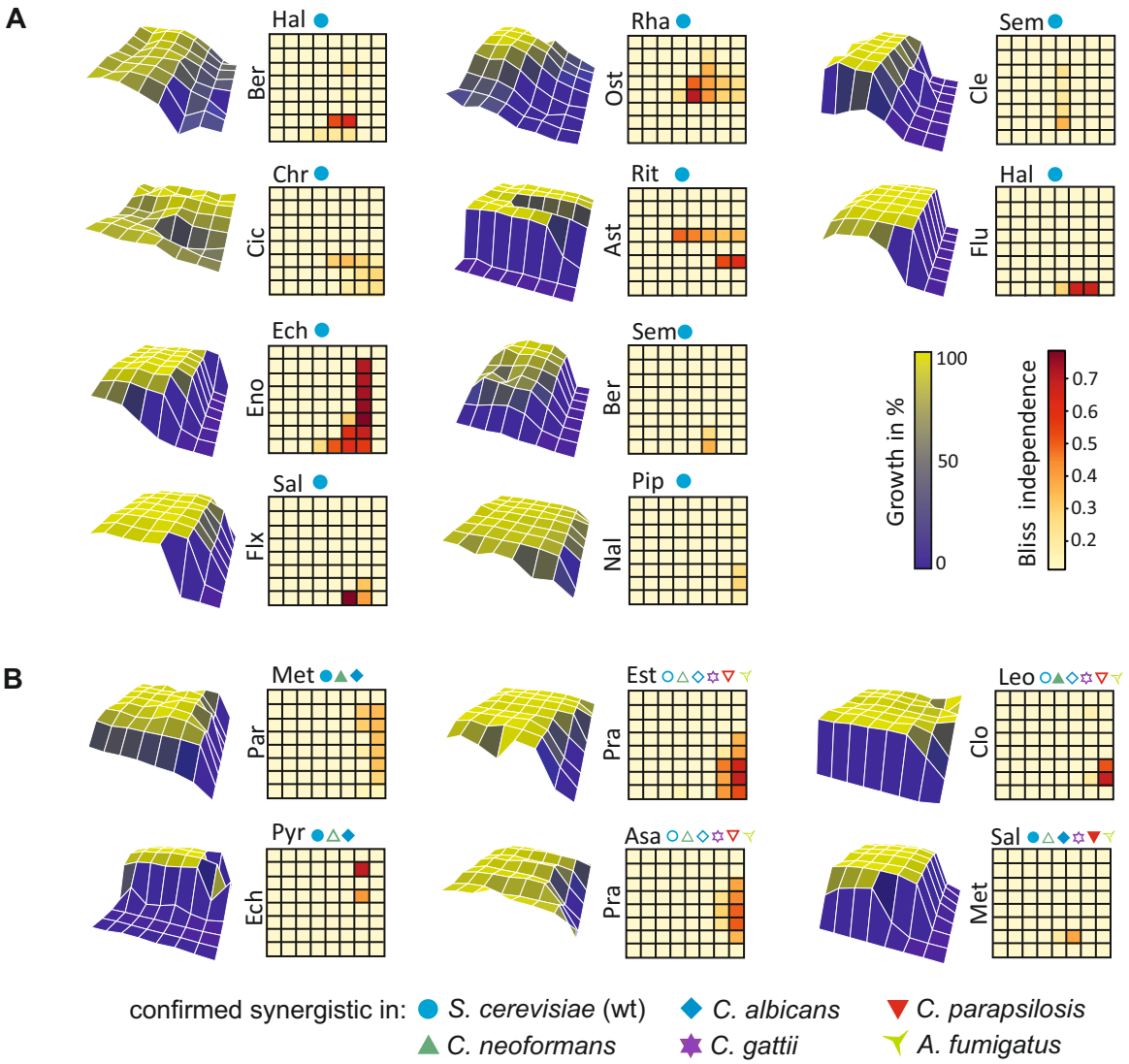

Figure S10. Synergistic Combinations in *S. cerevisiae* and Pathogenic Fungi, Related to Figure 6. **(A)** Additional synergistic combinations that were verified in a wild type *S. cerevisiae* strain. Growth OD_{600} measurements are shown as surface graphs at control saturation time points together with Bliss independence heatmaps. Concentrations were 0, 1, 2, 4, 8, 16, 32 and 64 μM for each compound. **(B)** Additional synergistic dose-response matrices from Fig. 6A that were not shown in Fig. 6B. Plots are as described in Fig. 6B.