Supplementary Note 3: Experimental screening results and statistics

The experimental data sets were drawn from a wide variety of therapeutic areas. CombinatoRx has systematically tested over a million combinations in more than ten disease areas, using cell-based assays and ~3,000 approved pharmaceutical ingredients (API), emerging therapeutics, and research probes. These screening efforts focus on whole-cell assays using phenotypes that integrate across cellular functions to enable the efficient detection of synergistic interactions between targets in different pathways[1]. Each screen usually starts with single agent testing to determine which library compounds are active and to facilitate optimal dose sampling for combination screening. Most combination data are collected as dose matrices using full or sparse formats (Fig. S8), in some cases matched in two or more phenotypic assays.

Each of our combination screens tested combinations of a diverse set of chemical agents (Fig. S9; Suppl. Data 5). The CombinatoRx chemical library has been classified into mechanistic classes using information drawn from the World Drug Index[2], DrugBank[3,4], and other public sources. Mechanistic annotations were manually curated to ensure consistency and grouped into categories to aid with assigning multi-target mechanisms for observed synergies. For agents with multiple targets, all major activities were recorded in the descriptor, but only one broad mechanistic class was chosen based on the primary mechanism underlying that drug's principal therapeutic use. Those drugs and probes without a known mechanism of action were classified as "Unknown". Each screens covered most of the 15 broad mechanistic categories, the least diverse drug sets corresponding to small experiments with a narrow focus (eg, Huntington's disease), or gathered from heterogeneous experiments (eg, the Bacterial set). Also, most of our screens excluded drugs with specific, non-human targets.

The chemicals used in the screens themselves show differing levels of activity and selectivity, which can strongly affect both synergy and selectivity measurements (Methods). To illustrate the level of these biases, we show the distribution of test assay activity (maximum inhibition at high concentration) and SI50 (the single agent ***SI*** at 50% effect) across all agents for each pair of assays in our screens (Fig. S10). Some of the screens incorporated single agent selectivity explicitly in their design, while most others required only that the agent be active in one or more of the assays under consideration. These priorities, as well as pragmatic factors like differential sensitivities arising from growth rate variations between cancer cell lines, led to substantial differences between assay responses in some of the screens.

Screen descriptions

The "Viral RSV 2008" screen used a survival assay with the A-2 respiratory syncytial virus (RSV) strain and HEp-2 lung cells. To monitor host cell survival, we compared the viability of infected and uninfected cells in the presence of drug treatments. After testing the library as single agents, we selected 60 drugs based on selective activity, mechanistic diversity, and approval status. Agents were chosen to be both active (>20% inhibition) in the viral assay and strongly selective over the viability assay. All pairwise combinations were then screened as full 6x6 dose matrices in both assays, and 1,480 combinations had been tested in both assays at the time of this analysis.

"Viral HepC 2005" screen. The hepatitis C experiments compared viral load in Huh7 liver cells, using a constitutive luciferase expressing replicon, to uninfected host cell viability. These combinations were selected from an earlier screen that used sparsely-sampled 6x6 dose matrices to test all pairs of 126 drugs in only the replicon assay. Drugs had been chosen for antiviral selectivity in these assays. This analysis includes the 392 combinations tested in both assays with full 9x9 dose matrices.

The "Bacterial 2002-2005" data set was assembled from disparate experiments testing proliferation either of methicillin-resistant *Staphylococcus aureus* or human colon-derived cell lines (HCT116). Because the original combination screens had been separately designed to include approved drug ingredients with strong activity in each of their own assays, there was little overlap between the bacterial and HCT116 screens, resulting in only 122 combinations with matched assays, mostly in partially-overlapping 6x6 dose matrices. There was also tendency for single agents to be selective between the two assays. However, as many of these data resulted from follow-on experiments from antibacterial synergies, there was a bias towards both single agent activity and to a certain extent synergy for combinations in this data set.

The "Cancer 180x180 2005" screen. This screen extended our combination screening methodology to a set of 180 chemical probes with a focus on diverse mechanistic coverage. The 180 chemical probes covered ~120 targets (based on published activities), and single agent activity was used for prioritization rather than inclusion in the screen. All pairwise combinations were tested against cell lines from colon (HCT116) or lung (A549) tumors, as well as immortalized fibroblasts (MRC9) as a "normal" control, to identify selective modulators of cancer proliferation. Using sparsely-sampled (12 combination points) dose matrices, ~14,700 agent pairs were tested in each of the cancer lines, and only 7,095 pairs for the more slowly growing MRC9 cells. As sparse matrices had been assembled from partial blocks on separate plates, they could not be readily separated into independent replicates.

The "Cancer 6x180 2005" screen focused on cotherapies, comparing responses across lung, ovarian, and melanoma-derived cell lines. In this screen, six drug candidates were combined with a diverse set of 200 approved pharmaceutical ingredients (API) and mechanistic probes, where all the APIs were components of standard cancer therapies and the probes targeted oncology-relevant pathways. Single agent activity in our assays was not a criterion for inclusion in the screen. There were substantial differences in activity between the cell lines, largely resulting from all three cell lines having been run using a 72h incubation time despite their different growth rates (A549 having the shortest and SKOV3 having the longest doubling time, longer doubling times providing less opportunity for growth during incubation, and correspondingly lower possible levels of inhibition for cytostatic agents). Combinations were sampled in full 6x6 dose matrices, and 1,101 pairs were tested in all three cancer lines.

The "Cancer 44x90 2006" screen crossed a sample of 44 emerging therapies and new chemical agents relevant to cancer with a diverse set of 90 approved drug ingredients. The 44 enhancees were selected solely on their therapeutic potential and mechanistic diversity, while the 90 APIs were chosen to have activity in one or both cell lines while covering most mechanistic classes in the approved drug fraction of the CombinatoRx compound library. As with the 6x180 screen above, the differential single agent activities can mostly be attributed to growth rate differences between the cell lines. The combinations were tested as full 6x6 dose matrices with only one replicate, and 3,494 had measurements in both assays.

The "Cancer MM 2007" experiments tested four multiple myeloma (MM) lines (MM-1R, being resistant to glucocorticoids, a standard-of-care treatment), in several sections. The first tested six standard-of-care MM therapies with 128 diverse APIs and emerging therapies, with an emphasis on cancer-relevant pathways and agents that synergize with corticosteroids in our inflammation assays. The second section combined 9 standard and emerging therapies with only 50 agents, with the aim of identifying potentially therapeutic synergies. Selectivity between the lines was not an explicit aim, but synergistic combinations were characterized as either broad spectrum (affecting most lines) or selective (affecting only one), with no preferential ranking between cell lines. In general, agents showed less activity in the steroid resistant line (MM-1R) than in the others, and the selectivity between sensitive lines reflected growth rates as for in the other oncology screens. Overall, ~1,200 combinations were tested n each of the four cell lines as full 6x6 dose matrices, usually with 2 or more replicates.

The "Inflam 2002-2005" data set was assembled from disparate experiments testing cytokine induction in buffy coat lymphocytes from blood donation samples. Inflammatory responses were stimulated using either phorbol ester calcium ionophores (PI) or bacterial lipopolysachharides (LPS), and the response was monitored using ELISA assays for tumor necrosis alpha (TNF-α) or interleukin 2 (IL-2). Because the original combination screens had been separately designed to include approved drug ingredients with strong activity in each of their own assays, there was little overlap between the screens, resulting in ~200-300 combinations in each assay with matched data, mostly in partially-overlapping 6x6 dose matrices. Given the heterogeneity of this set, independent replicates were rarely available.

The "Inflam 48x48 2006" screen focused on cytokine expression versus cell viability compared interleukin 1 (IL-1) induction in buffy-coat lymphocytes extracted from human donor blood samples to cell viability for NHDF dermal fibroblasts. Forty-eight drugs were chosen based with demonstrated inhibitory activity in the IL-1 assay, and most showed stronger inhibition in the viability assay, suggesting no therapeutic window. The objective for this screen was to identify synergies between the agents that would overcome this toxicity and give rise to anti-inflammatory combinations with little corresponding toxicity. All pairwise combinations were tested as fully-sampled 6x6 dose matrices, with 1,163 drug pairs tested in both assays.

The "Inflam 12x100 2006" screen tested twelve agents 200 diverse drug ingredients with proven activity in at least one of the assays tested. Two inflammatory cytokine (TNF-alpha and IL-1) induction assays were compared with two viability assays for primary endothelial and smooth muscle cardiovascular cells. As with the 48x48 screen, the therapeutic objective was to identify non-toxic anti-inflammatory synergies, and most single agents with TNF-alpha or IL-1 activity were substantially toxic in the viability assays. Frozen blood samples from a limited set of donors were used to control variability in the inflammation assay. Of the three primary cell lines used for viability, differences between activities reflected relative growth rates (HUVEC being the most sensitive to drug treatment). Combinations were sampled as full 6x6 matrices, with 746 tests in TNF-alpha and AoSMC, but only 397 for the IL1 and HUVEC assays.

The "Vascul 90x200 2007" cardiovascular screen compared monocyte chemoattracting protein 1 (MCP-1) induction in human coronary artery endothelial cells (HCAEC) to metabolic viability measurements on the same samples. Twenty-three APIs, chosen to have selective activity in these assays or in corresponding tests involving coronary artery smooth muscle cells (CASMC), were combined with 253 chemicals from the CombinatoRx library that had been selected to cover a diverse set of mechanistic targets without considering their activity in these assays. Again, the therapeutic objective was to identify anti-inflammatory synergies with low levels of corresponding toxicity. The compounds had been explicitly chosen to be selective for MCP-1 activity, and consequently the SI50 distribution of the agents shows a strong positive bias. Combinations were sampled using 3x6 dose matrices (with the short axis for the 253 probes), and 5,885 combinations were tested in both assays.

The "Vascul 90x90 2007" cardiovascular screen tested endothelial and smooth muscle cells, both for MCP-1 induction and cell viability. The objectives and methods were identical to those for the previously described cardiovascular screen. Ninety-three single agents were chosen based on selective responses in these assays (favoring MCP-1) either in CASMC or HCAEC cells, and pairwise combinations (~4,300 in endothelial and ~3900 in smooth muscle cells) were tested as sparsely sampled 6x6 dose matrices. These sparse matrices, being assembled from separated sub-blocks, are not easily separated into independent copies.

Finally, the "Huntington Disease 30x30 2007" screen covers neurodegeneration experiments that used microscopic measurements of neuronal cells deficient in huntingtin protein (Htt) to compare Htt translocation to viability as determined by the total cell counts. These combinations were collected as part of an ongoing screen using full 6x6 dose matrices to test all pairs of 30 compounds that had been chosen based on selectivity in these assays favoring Htt translocation, augmented by some other combinations. The single agents were

chosen to be selective for inhibiting Htt translocation, and this bias is strongly reflected in their SI50 distribution. All 455 combinations from the screen are included in this study.

Experimental selectivity bias results

Synergy and selectivity values were computed for all combinations tested, across all pairs of assays in the screen (Suppl. Data 6), and were used to calculate selectivity biases ***B***, comparing the ***SI*** distribution of the top 5% synergies to that for all of the combinations (Fig. S11; Suppl. Data 7). For each pair of assays in a screen, we compared the ***SI*** distribution for the $S > S_{cut}$ synergies to that of the entire set of combinations. For small data sets, we automatically decreased S_{cut} to ensure at ~10 combinations with replicates among the synergistic set. The selectivity bias was measured as the difference of the mean ***SI*** between the top synergies and all combinations. There were generally positive biases for all assay pairs, in agreement with the expectation that synergies between drug targets are more context-specific than single target activities. The strongest ***B*** measurements tended occur between very different assays (eg, RSV survival and host viability, MRSA proliferation and HCT116 viability, or Htt translocation and cell count), and usually were not recapitulated in the corresponding "reverse" assay pairing.

Null selectivity tests

To serve as a null test for selectivity, we also compared each assay to its self for those screens having true replicate matrices (Fig. S12). As seen in our simulated screens (Sup. 1), when agent activity or synergy are strongly linked between the assays the selectivity bias disappears. Thus, any deviations from $SI = 0$ represent noise variations between the replicates, giving a good estimate of the stochastic error in our ***B*** measurements. Most of the screens had insignificant biases, as expected, with the weighted consensus $B_{null} = 0.023 \pm 0.007$.

Meta-analyses

We used multiple regression analyses to explore the dependence of ***B*** results on the properties of the screening data sets (Fig. S13). All assay pairs within each screen were characterized by variables that describe screen design parameters and single agent activities, and each such set was analyzed across five levels of S_{cut} (returning >2%, >3%, >5%, >10%, >20% of the synergies). The partial correlations of each variable with ***B*** across these 380 data sets showed that screen size and the agent selectivity had a strong influence. It is notable that S_{cut}, the most arbitrary parameter in our analysis, only weakly affected the selectivity bias.

To quantify the selectivity bias more generally, we performed a meta-analysis across all of the data sets, weighting each assay pair by the inverse square of its 95% confidence interval (Fig. 3). The confidence interval for each pair was estimated from twice the standard error of the mean ***B***, with a sequential Bonferroni adjustment[5] for multiple hypothesis testing across assay pairs in each screen. The consensus $B = 0.104 \pm 0.010$ (95% confidence) was highly significant. The strength of the effect increases to 0.22 ± 0.02 if only the 16 assay pairs that are aligned with a clear therapeutic objective are included in the weighted average.

Yield of therapeutically selective synergies

The fractional yield rates of selective treatments are greater among synergies than for combinations in general or single agents in our screens. For a pair of assays, we can count the fraction f_{SEL} of single agents with more than threefold selectivity ($SI > \log_{10} 3$) favoring the test over the control assay. This fraction can be compared to the corresponding fraction of combinations with the same level of selectivity (Suppl. Data 7). Across all our experiments, weighted by the number of combinations tested, the average f_{SEL} for the single agents is 0.05 (0.02 for therapeutically aligned assay pairs). For combinations the average f_{SEL} was 0.32 (0.14 for aligned assays), which increased to 0.42 for the top 5% synergies (0.18 for aligned assays).

The number of agents useful for modulating targets relevant to a disease is limited (eg, for cancer there are ~100 probes with distinct targets[4]), and the number of therapeutically selective

agents will be correspondingly small. Given typical performance for our screens (Fig. 3), 100 chemical probes are expected to have ~2-5 agents with >3x selectivity between disease and control assays. Pairwise combination testing would be expected to find ~250 synergies (top 5% of scores), of which ~40-80 would be expected to have >3x selectivity between those assays. Thus, because they are so numerous and statistically more selective, synergistic combinations greatly increase expected yield of selective treatments from a limited set of agents.

References

1. Zimmermann, G.R., Lehár, J. & Keith, C.T. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov Today* **12**, 34-42 (2007).
2. Daylight Chemical Information Systems, I.n.c. World Drug Index, Authoritative index for marketed and development drugs. *Derwent Publications* (2008).
3. Wishart, D.S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668-672 (2006).
4. Wishart, D.S. et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**, D901-906 (2008).
5. Holm, S. A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70 (1979).

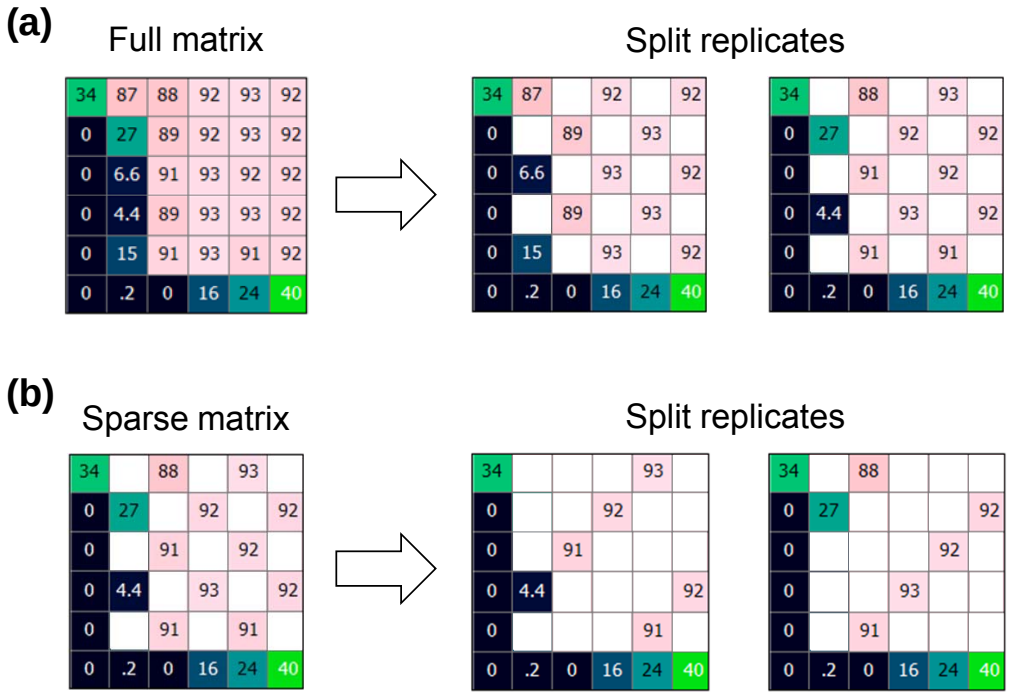

Figure S8. Dose matrix densities used in our screens and how they were split into replicates. The dose matrices are shown as in Fig. 1, with single agent curves along the left and bottom edges and doses increasing to the right and towards the top. (a) Most of the experiments collected full matrices, which can be split into two independent replicates containing alternate data samples. (b) For screens that used sparse matrices, samples were alternated as shown. For screens containing true replicates, the consensus matrix was used to generate split matrices, and the analysis was repeated using the true single replicates to validate this methodology.

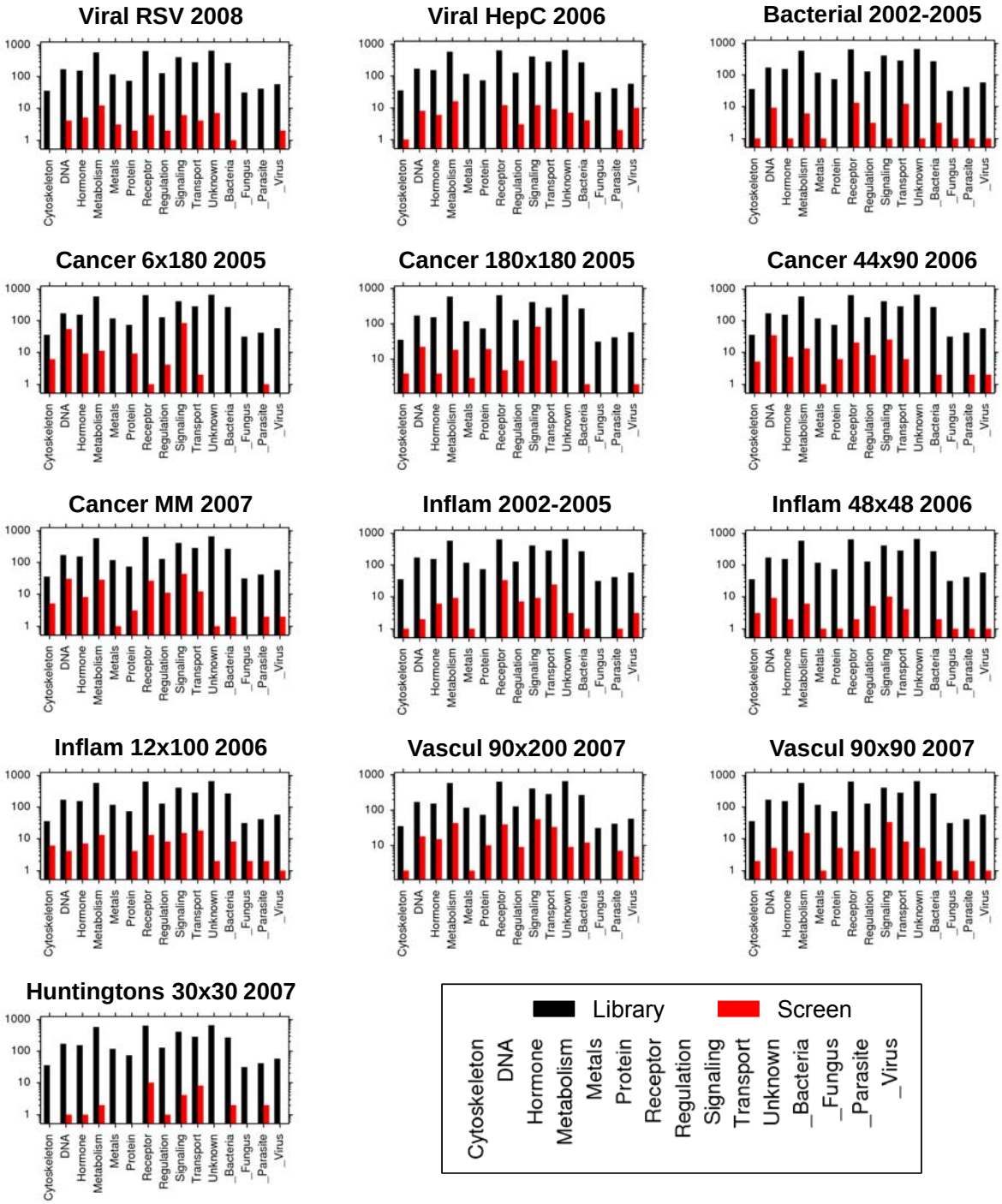

Figure S9. Mechanistic distributions for the chemical agents in each screen. Each plot shows the distribution of agents in the CombinatoRx library divided into broad mechanistic classes (black histograms), along with the mechanistic distribution for those agents (red) included in one of the combination screens analyzed in Fig. 3. The differences in mechanistic coverage reflect the therapeutic objectives underlying each screen's design.

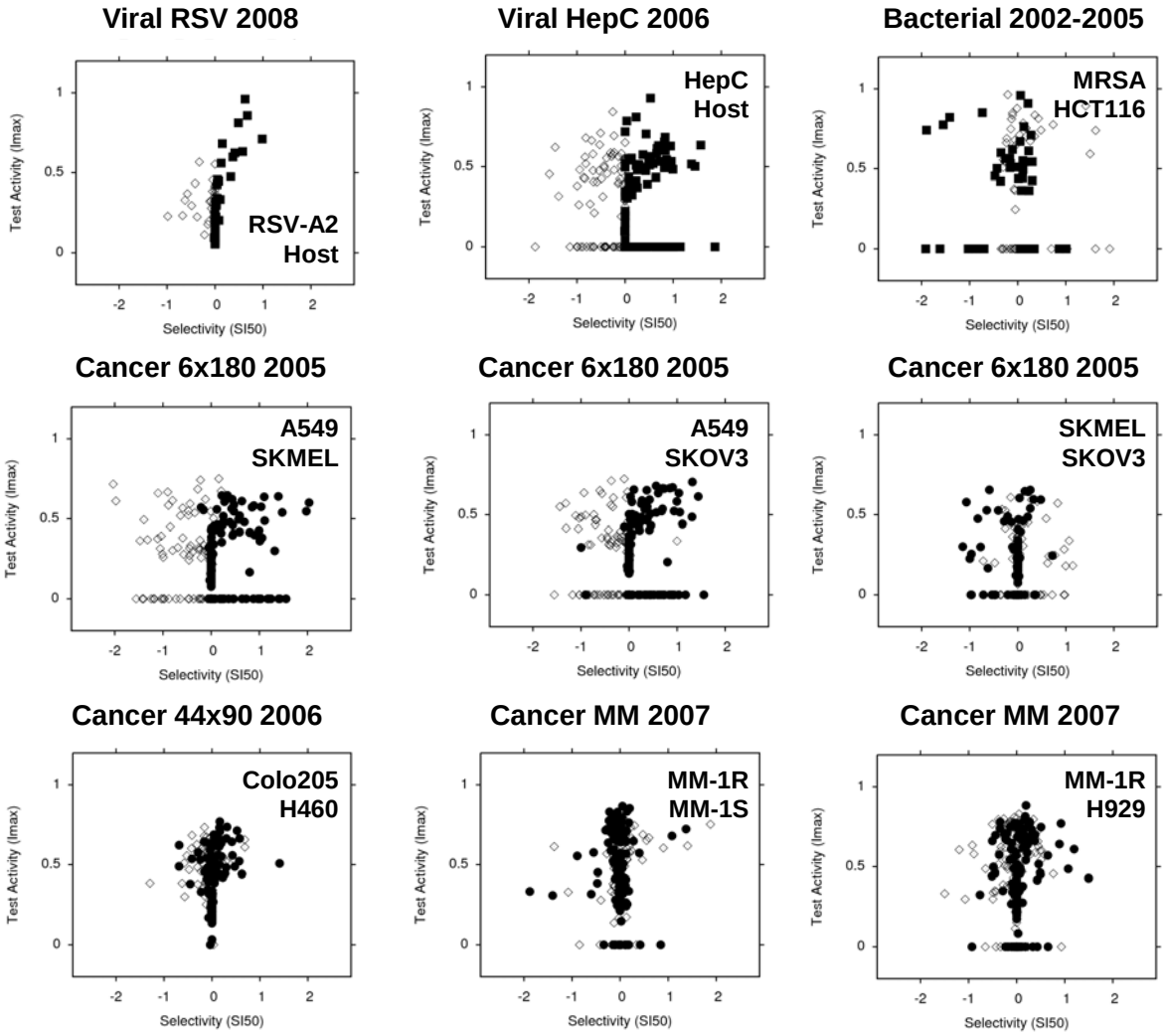

Figure S10a. Single agent activity (maximum inhibition) and selectivity (***SI*** at 50% effect) for each screen. For each assay pair, "forward" distributions relative to the assay listings in Fig. 3 are marked with filled and "reverse" with open symbols. For screens with selectivity as a clear therapeutic objective, forward comparisons (aligned with that objective) are marked as boxes. Some screens (eg, RSV, HepC) clearly show clear single agent selectivity built into their design. In screens with a clear therapeutic selectivity objective, single agent selectivity was not always required in the design for all assay comparisons, resulting in cases (eg, IL1-NHDF, IL1-HUVEC, TNF-HUVEC) where the predominant selectivity favors the "reversed" assay relative to the screen's objective.

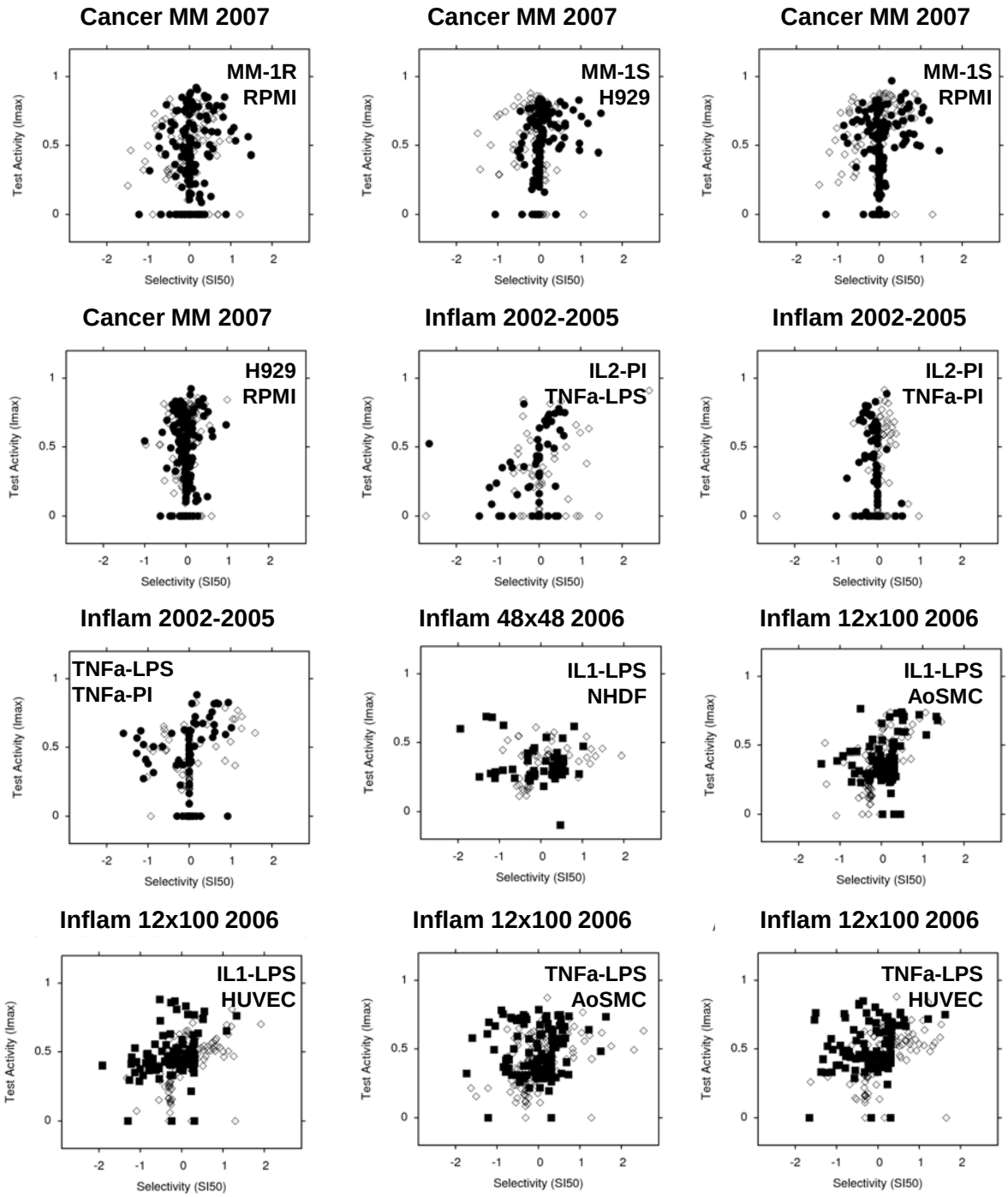

Figure S10b. Single agent activity and selectivity, continued.

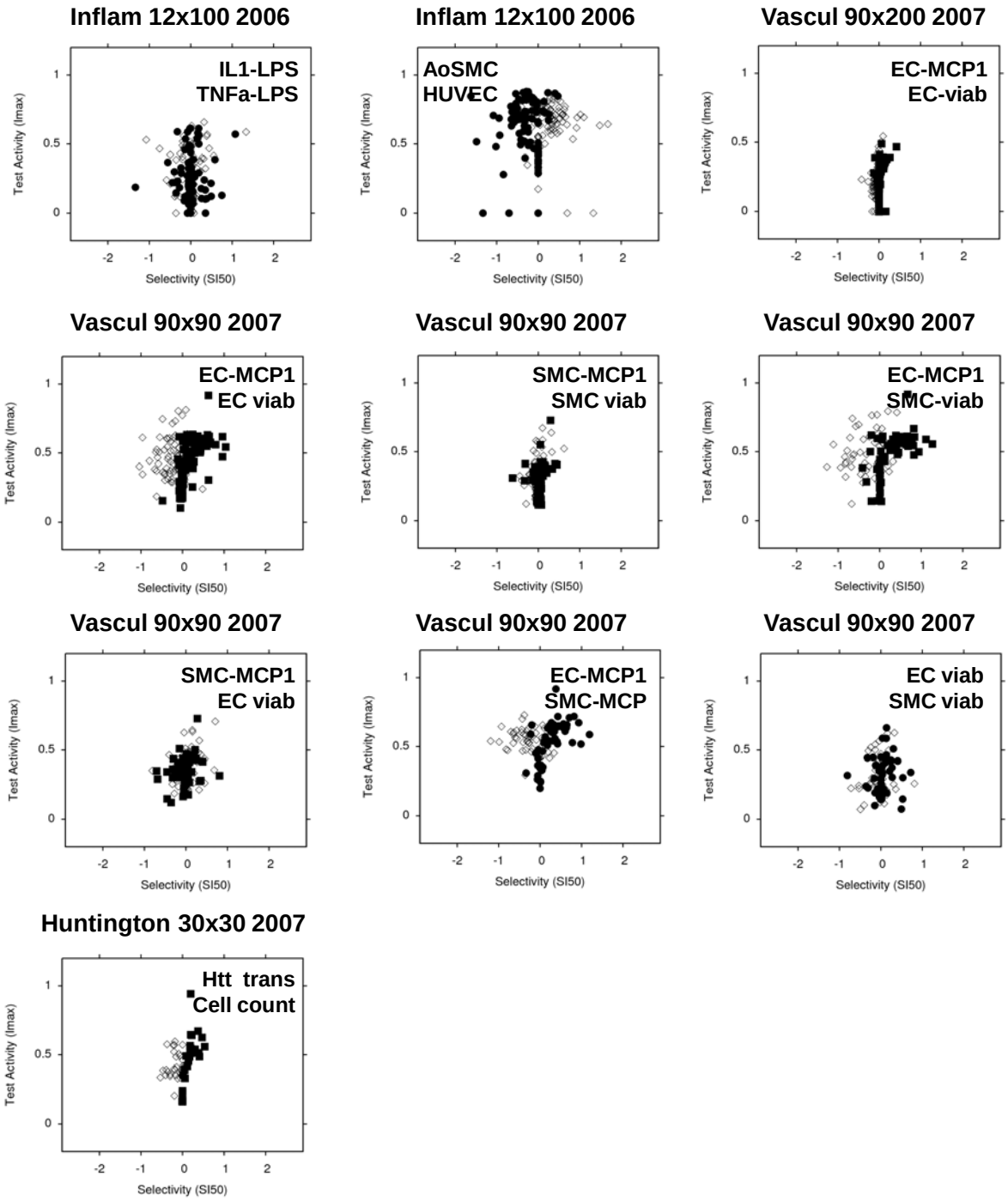

Figure S10c. Single agent activity and selectivity, continued.

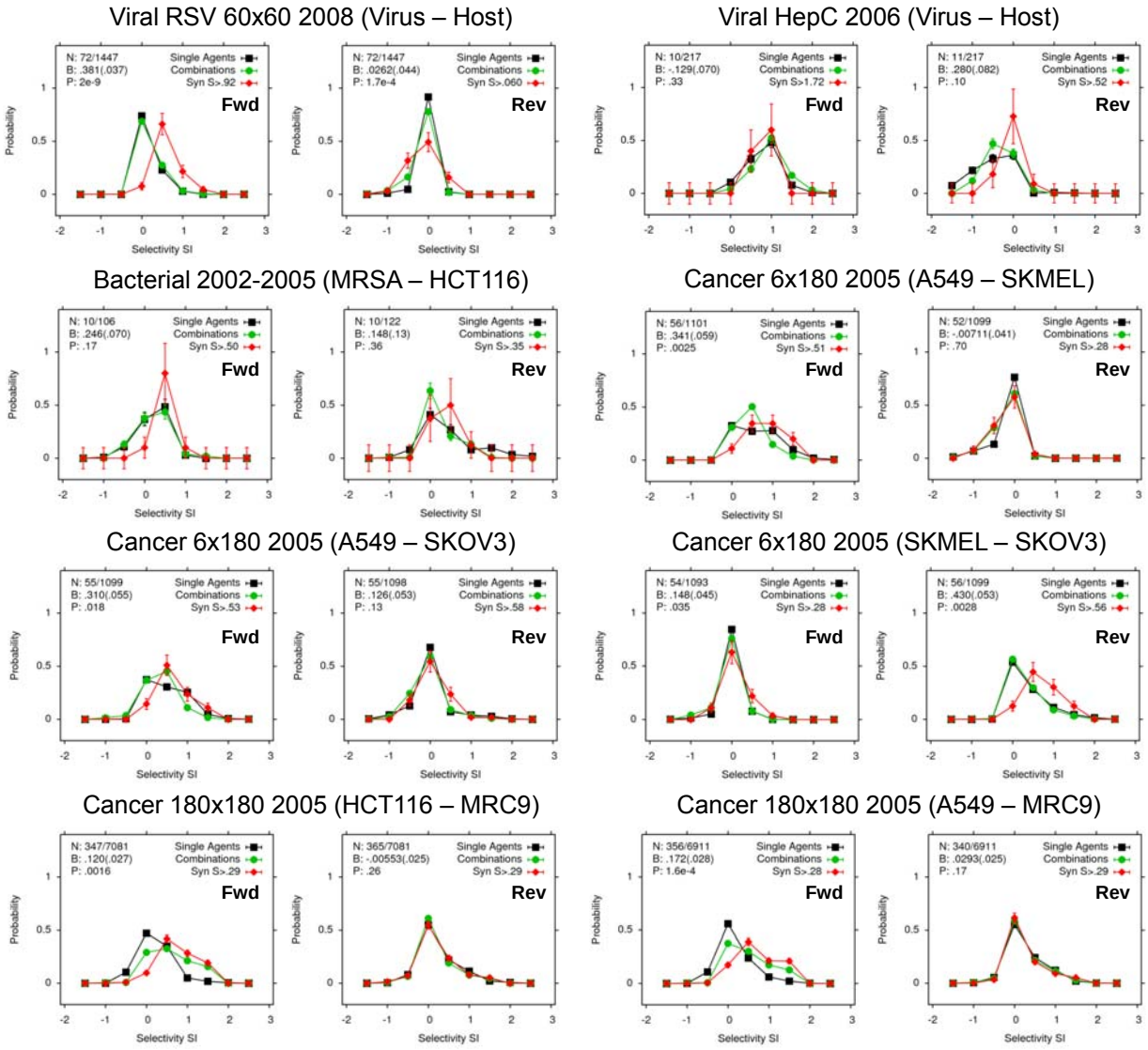

Figure S11a. Selectivity bias for the experimental screens. Each panel shows a comparison between the ***SI*** distributions for the top 5% synergies, all combinations, and the more selective single agent in each combination. Each assay comparison in a screen has panels for both the therapeutically "forward" and "reverse" direction relative to the assay listings in Fig. 3. Each combination's synergy ***S*** and ***SI*** value was calculated as described in Fig. 1, using $\boldsymbol{Z}_{cut} = \max(\boldsymbol{Z}_{test})/2$. The number of combinations (top 5%/all), the selectivity bias ***B*** (with its standard error), and the probability of obtaining this level of distribution difference based on Poisson counting statistics are shown in the top left of each panel. Most of the assay comparisons produced positive biases for the synergistic combinations, and observed shifts were larger than any seen for random noise (Fig. S2) but comparable to those seen in simulations with introduced context-dependent synergy (Fig. S3). This confirms that the selectivity shift arises from biological selectivity rather than measurement errors.

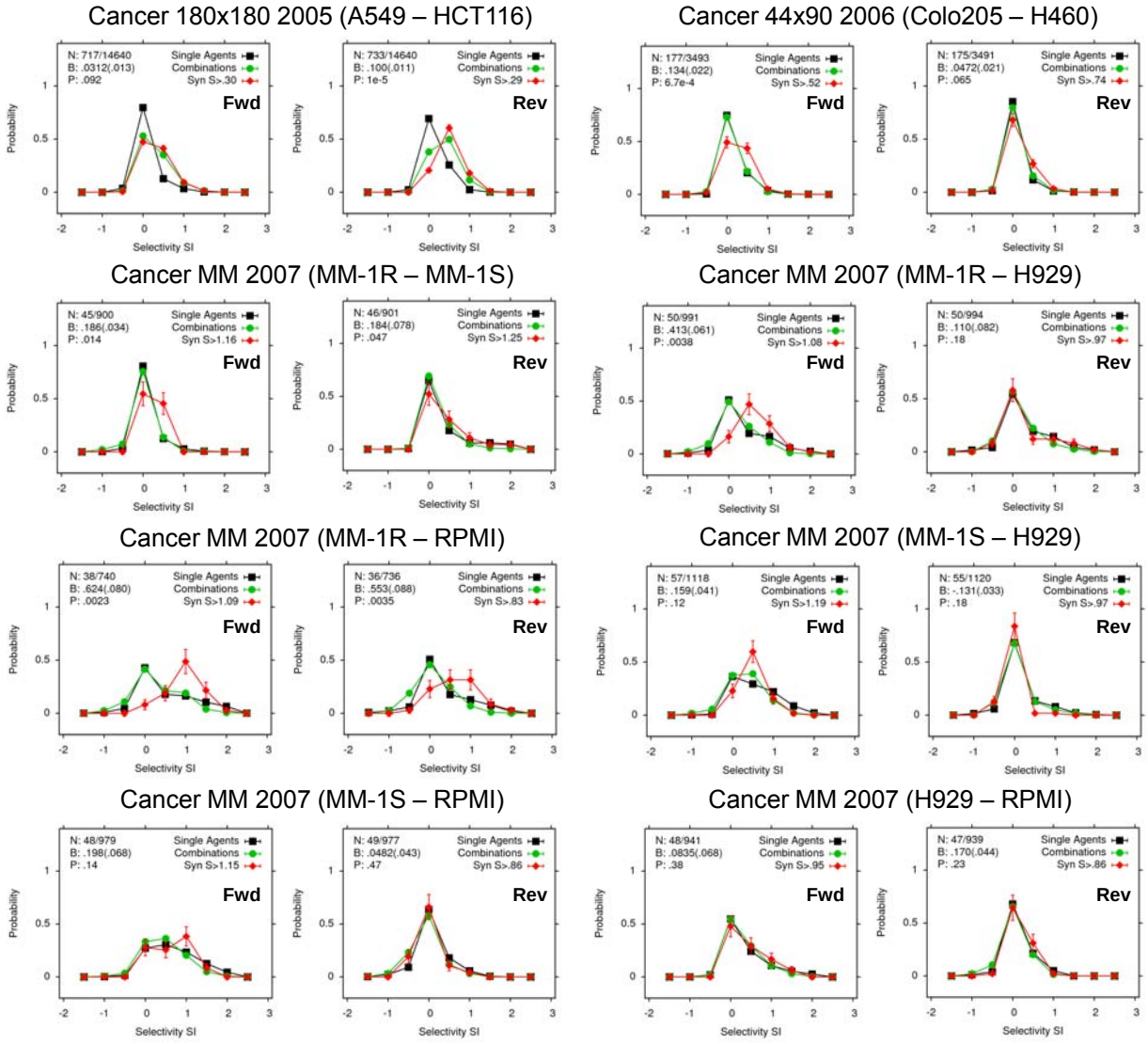

Figure S11b. Selectivity bias for the experimental screens, continued.

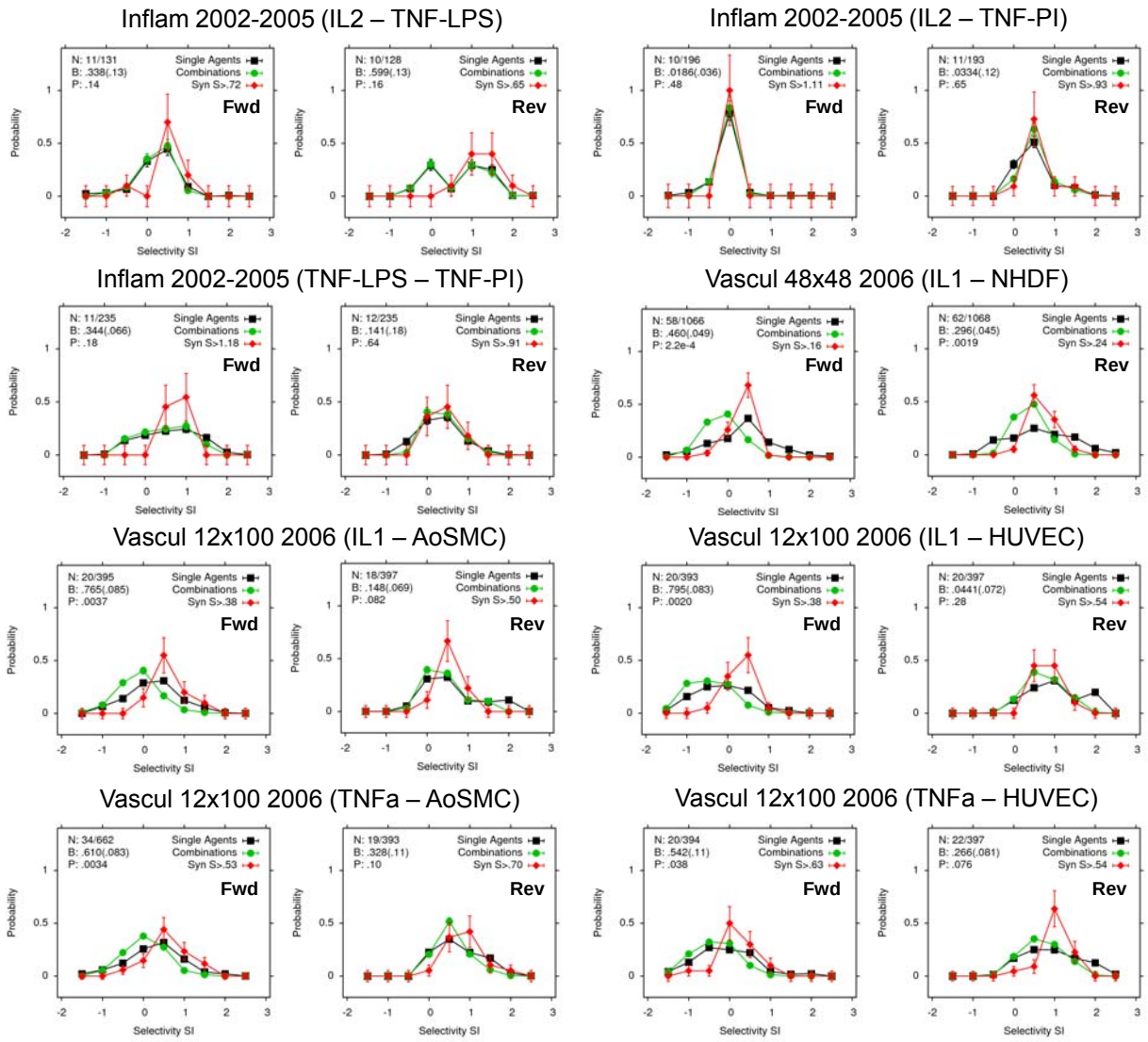

Figure S11c. Selectivity bias for the experimental screens, continued.

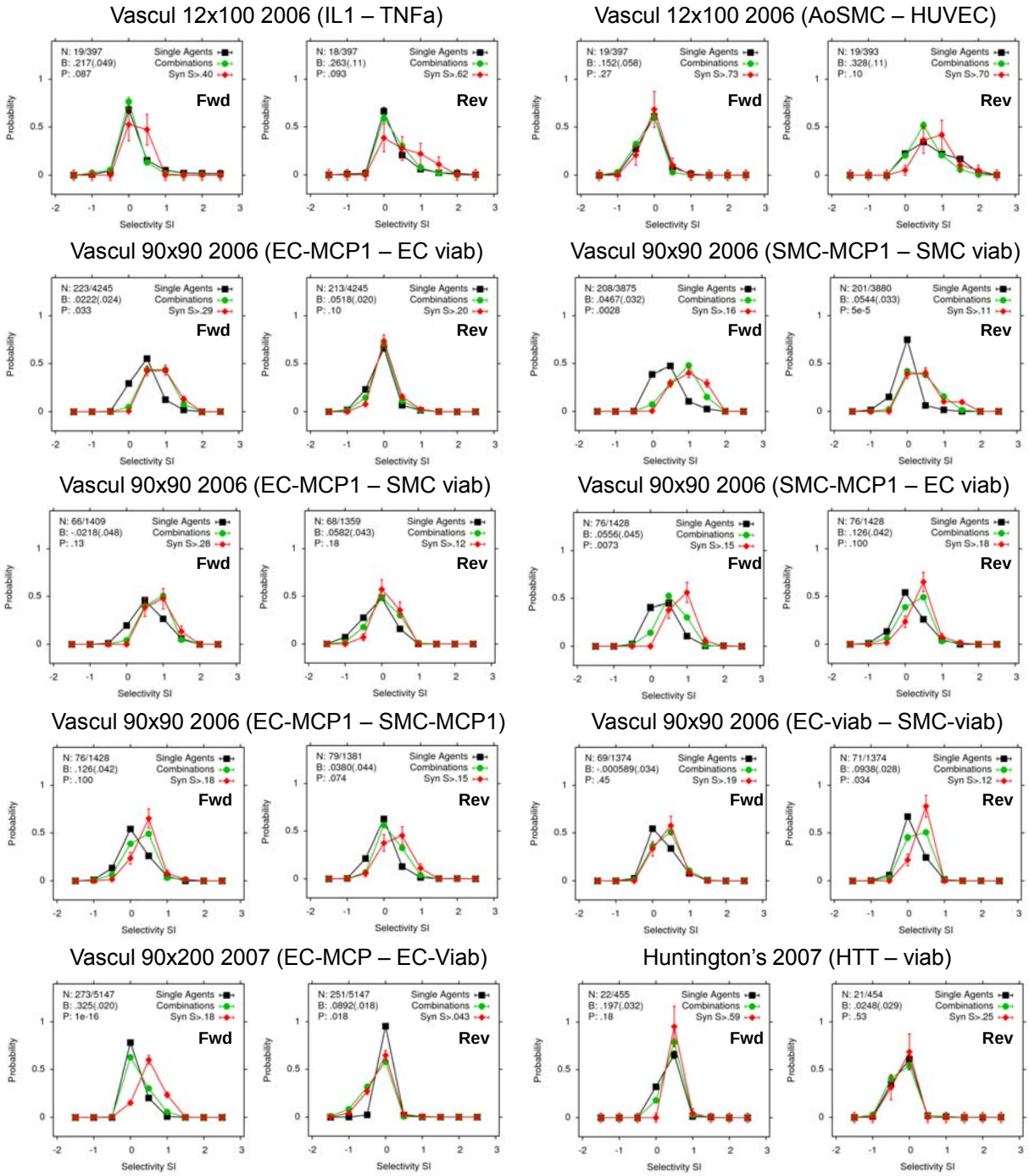

Figure S11d. Selectivity bias for the experimental screens, continued.

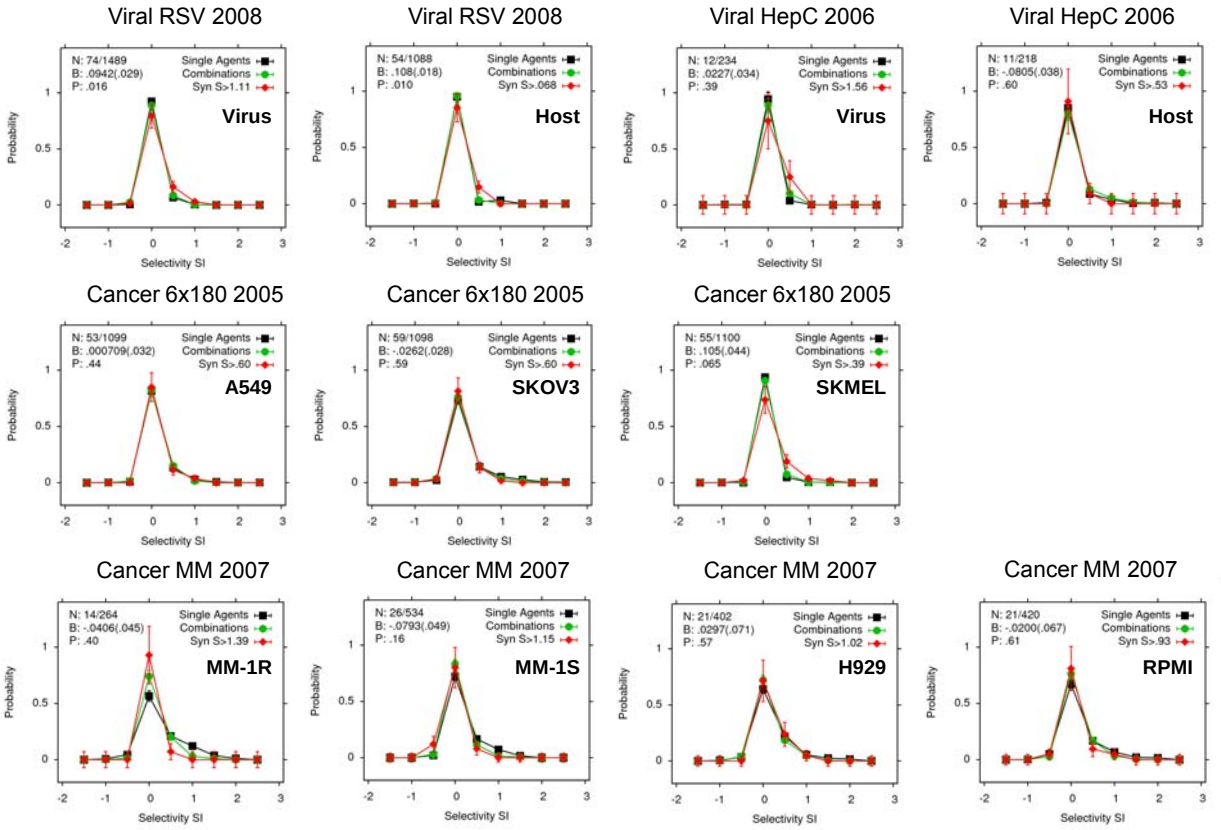

Figure S12a. Null selectivity bias for the experimental screens with true replicate matrices. Each panel shows the same analysis as for Fig. S11, except that each assay is compared to itself, using one copy for the control and the other copy for a split-matrix ***S*** vs. ***SI*** analysis. Because both sides of the selectivity calculation are the same assay, any variations from ***SI*** = 0 are due to experimental noise. Each combination's ***S*** and ***SI*** value was calculated as in Fig. 1, using $Z_{cut} = \max(Z_{test})/2$. The number of combinations (top 5%/all), the selectivity bias ***B*** (with its standard error), and the probability of obtaining this level of distribution difference based on Poisson counting statistics are shown in the top left of each panel. This analysis confirms that the shifts observed in Fig. S11 do not arise from stochastic errors, even using the non-Gaussian noise in our actual data.

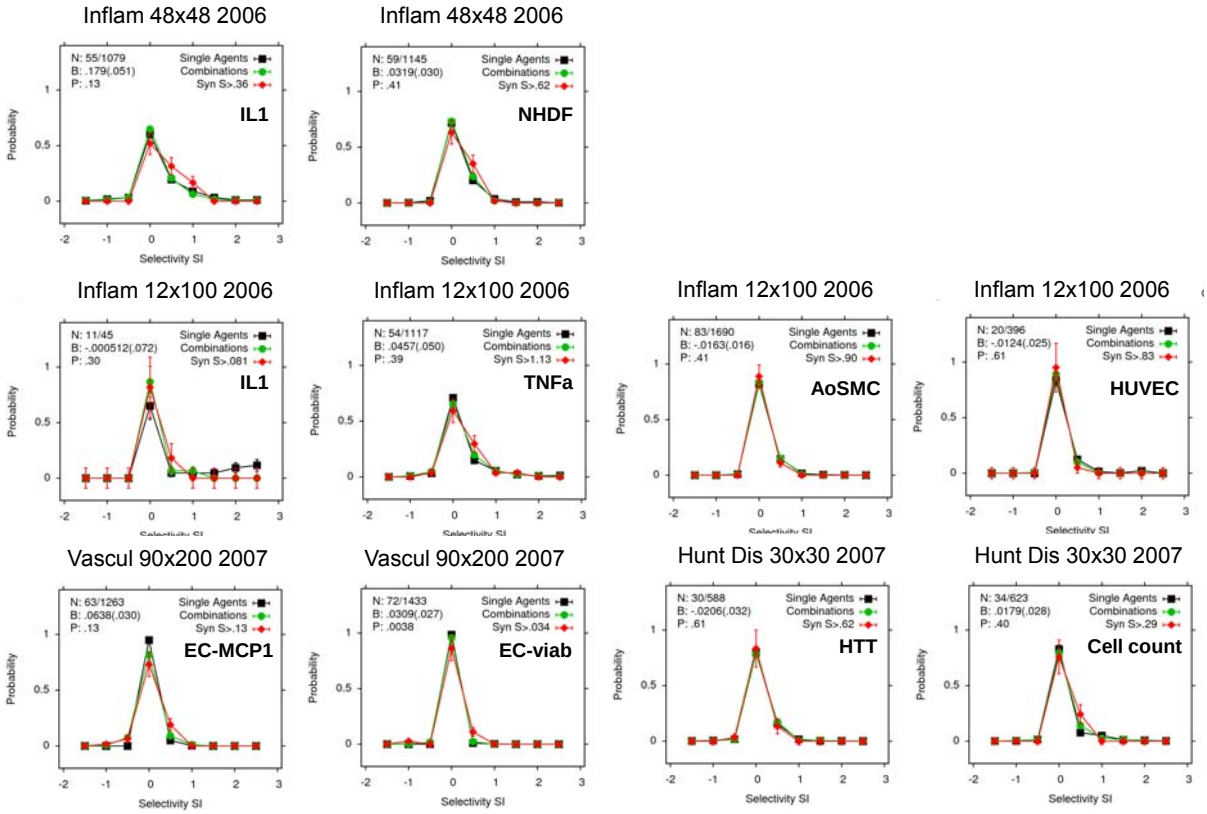

Figure S12b. Null selectivity bias for the experimental screens with replicates, continued.

(a)

Model	R^2	X^2	DOF	F-test	P-val
Linear	0.26	282	370	1.313	0.004
Quad	0.35	247	361	1.114	0.15
Cubic	0.47	202	352	1.194	0.047
Poly4	0.54	185	343	1.060	0.294

(b)

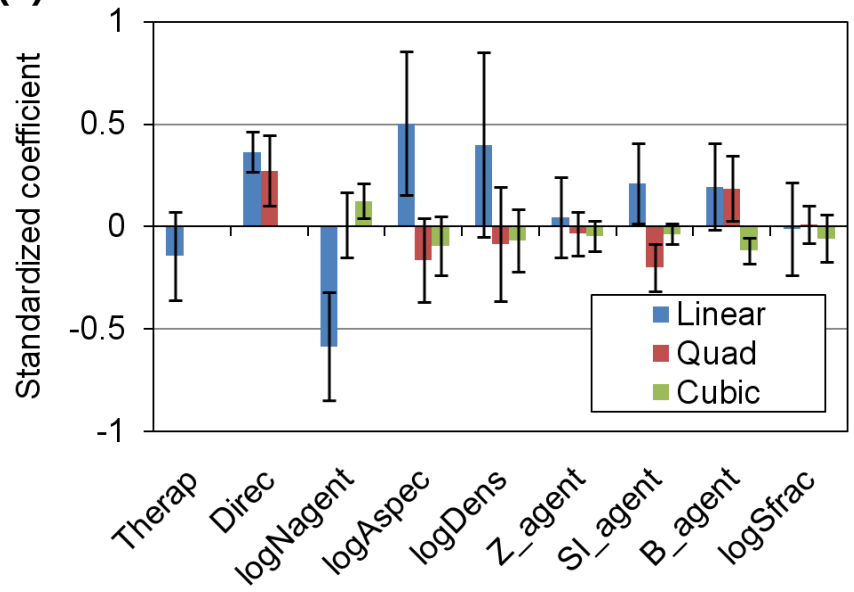

Figure S13. Multiple regression analysis of the experimental screen results. All assay pairs within each screen were characterized by 9 design variables: "Therap", noting whether both assays are in different therapeutic areas (ie, inflam vs. prolif); "Direc", representing the assay ordering relative to the therapeutic objective; "logAspect", logarithm of the aspect ratio of the number of agents used in each dimension of the combination screen; "logNagent", logarithm of the number of agents used in the screen; "logDens", logarithm of the number of combination samples in each dose matrix; "Z_agent", the average maximum inhibition in the test assay for all agents used in the screen; "SI_agent", the average ***SI*** at 50% effect for the single agents in the screen; "B_agent", the single agent ***SI*** averaged over all combinations in the screen; and "logSfrac", logarithm of the fraction of combinations passing $\boldsymbol{S} > \boldsymbol{S}_{cut}$. Each screen was analyzed across all pairs of assays at five levels of logSfrac (2%, 3%, 5%, 10%, 20%), resulting in 380 data sets. Multiple regression analysis on the selectivity bias ***B*** using polynomial functions shows the relative contributions of each variable. (a) The model fits improved significantly for increasing polynomial order up to cubic. (b) Examining the standardized coefficients for the cubic fit (with 95% confidence error bars), the screen's size and shape and the dose matrix density had the greatest effect on ***B***, as expected. The alignment of the assay pair with the screen's therapeutic objectives and the single agent ***SI*** levels also had significant predictive power, but the choice of $\boldsymbol{S}_{cut}$ had a relatively small effect.