Science Advances
AAAS

Supplementary Materials for

Accelerating drug discovery and repurposing by combining transcriptional signature connectivity with docking

Alexander W. Thorman *et al.*

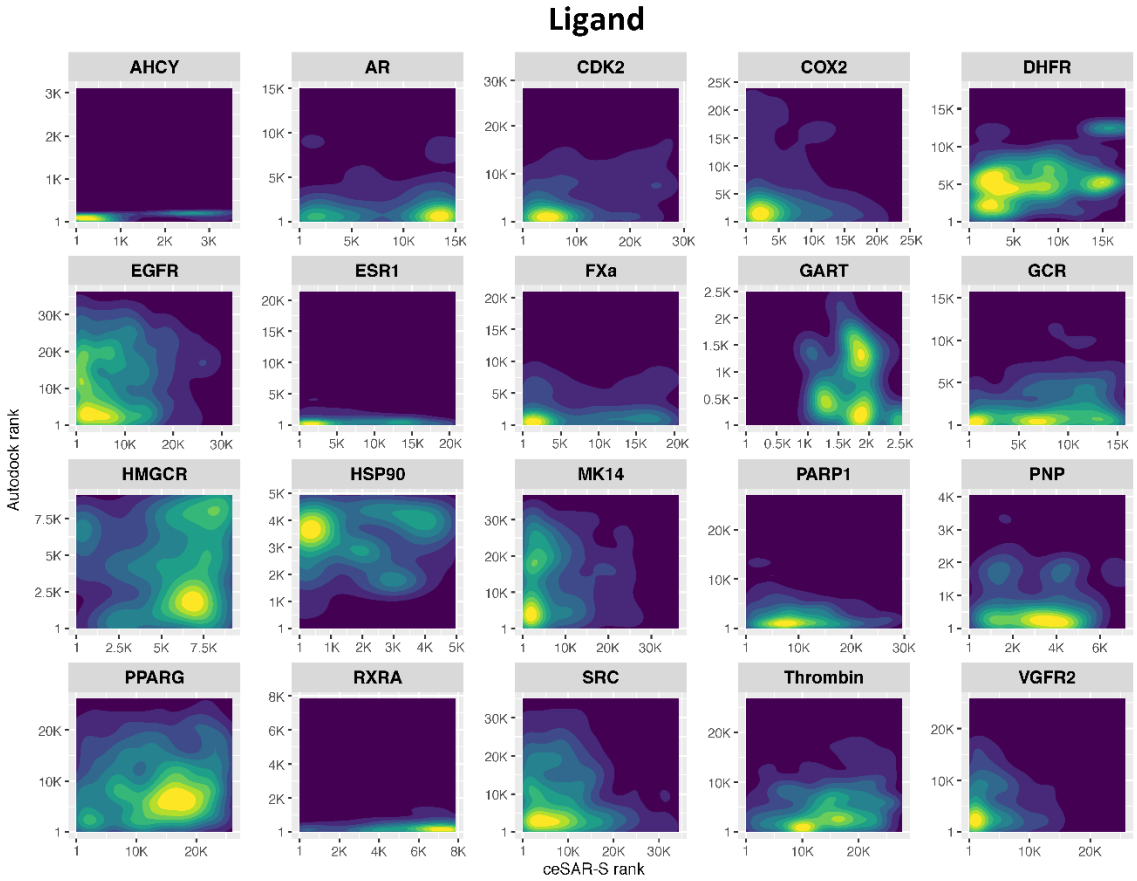
Corresponding author: Andrew B. Herr, andrew.herr@cchmc.org; Jarek Meller, mellerj@ucmail.uc.edu

Sci. Adv. **10**, eadj3010 (2024)
DOI: 10.1126/sciadv.adj3010

This PDF file includes:

Complementarity of signature connectivity and docking-based scores (S1)

The new approach for drug discovery, dubbed *connectivity enhanced* Structure Activity Relationship (*ceSAR*), combines drug and target transcriptional signature connectivity analysis with molecular docking. The method effectively combines two very different types of signals. The signature concordance analysis generates scores that indicate the level of similarity between downstream transcriptional responses induced by the loss of function of the target, either due to a chemical inhibitor or genetic knock-down. On the other hand, molecular docking simulations rank candidate molecules by predicted binding affinities or other scores (generated here by using *Autodock* or another suitable docking program) that capture their biophysical compatibility with the target protein structure. In this section, we assess the complementarity of these two signals by comparing ranking of true positives (ligands) and decoys for 20 DUD-E targets used also in the main manuscript.



Supplemental Figure 1: *ceSAR*-S (X-axis ranks) vs. *Autodock* (Y-axis rank) ranking of true positives (ligands) for 20 DUD-E target libraries. Yellow: high density; Dark Blue: low density.

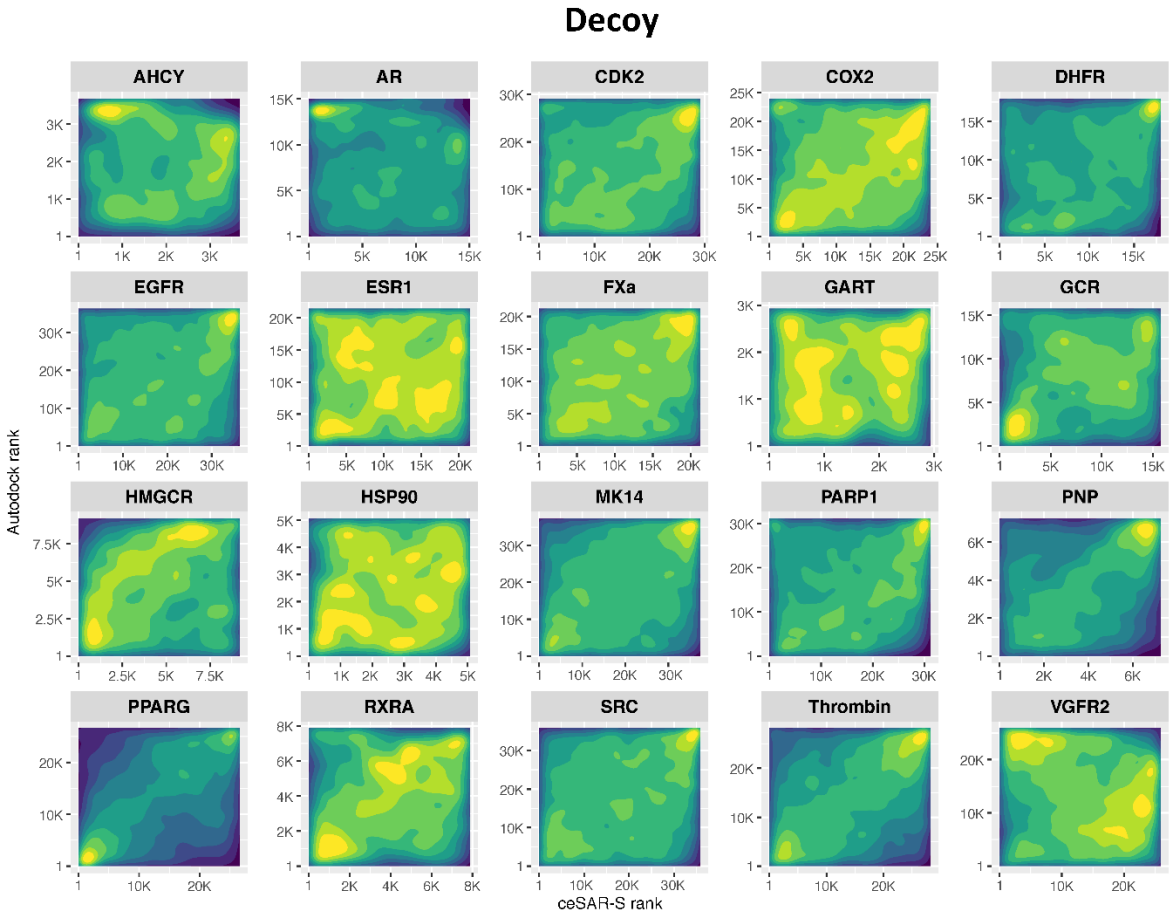
In the simplest form of the method, which is target structure independent and referred to as *ceSAR*-S (for signature-based) throughout the main manuscript, candidate molecules are ranked based on the chemical similarity to their 'concordant' LINCS analogs. Namely, for a target gene *t*

with at least one knock-down transcriptional signature available in LINCS, $t \in L$, and for a library of small molecules, Q, the following similarity score is computed for each $q \in Q$ as a basis for *ceSAR*-S ranking:

$$s(q) = \max_{k \in L,\, c^*(k,t) \geq c_0} \{\sigma(q,k)\}$$

where $\sigma(q,k)$ is the Tanimoto coefficient (Jaccard similarity measure) (*21*) between compounds q and $k \in L$ represented as binary fingerprints, while $c^*(k,t)$ is the signature concordance score for k, defined as the maximum concordance (over all cell lines for t, and cell line, concentration, exposure time tuples for k) between the signatures of chemical perturbagen k and genetic knock-downs of t:

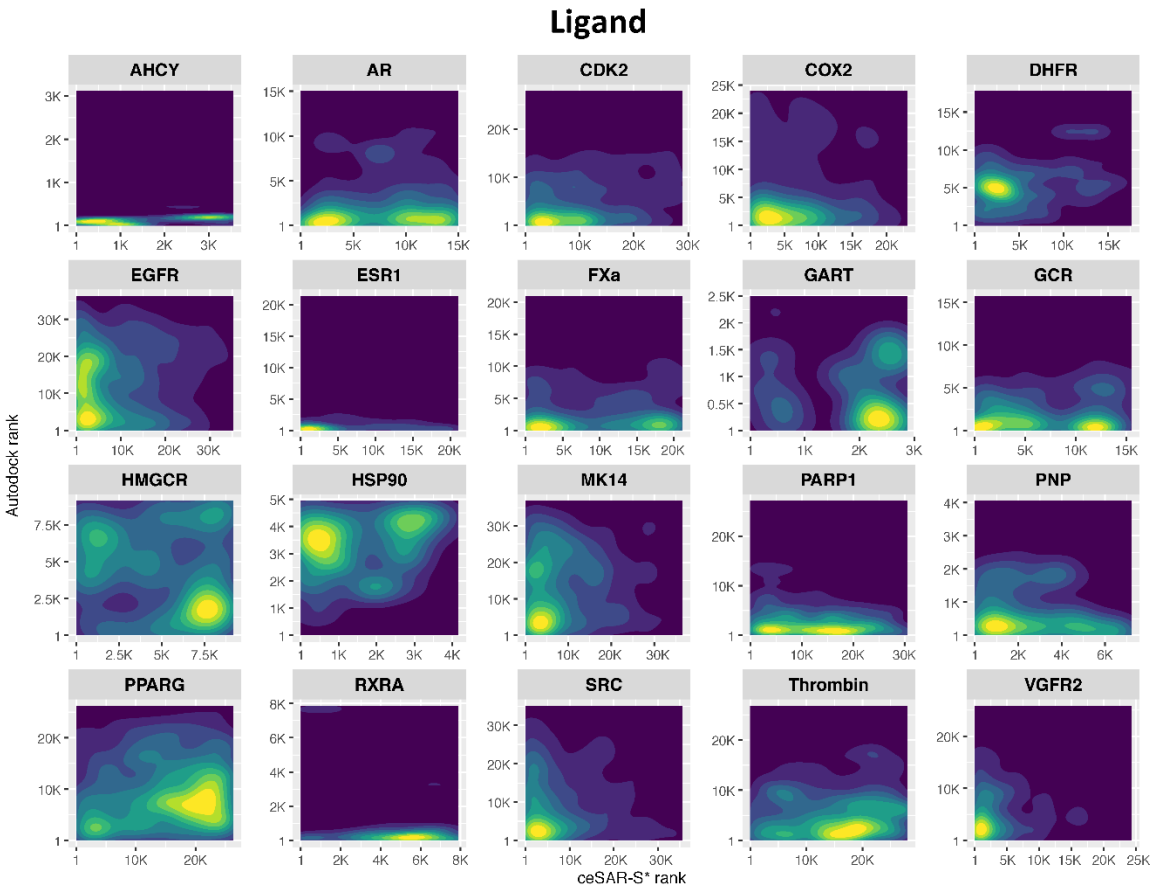
$$c^*(k,t) = \max_{k,t \in L} \{c(k,t)\}.$$



Supplemental Figure 2: *ceSAR*-S (X-axis ranks) vs. *Autodock* (Y-axis rank) ranking of true negatives (decoys) for 20 DUD-E target libraries. Yellow: high density; Dark Blue: low density.

Thus, *ceSAR*-S ranks candidate molecules by using a similarity score, $s(q)$, which is in fact the Tanimoto coefficient for the closest 'concordant' LINCS analog of q. By increasing the

similarity threshold, $s_0 \in [0,1]$, one can reduce the initial library to a subset by taking only those compounds q that receive a score larger than s_0. When reducing the library to a small subset for further re-ranking and validation using S, increasingly close analogs are being considered. The value of the signature concordance score, transferred from the closest concordant analog in LINCS, is used to break ties, resulting in concordance-based ranking of direct analogs of candidate molecules with $s(q) = 1$.

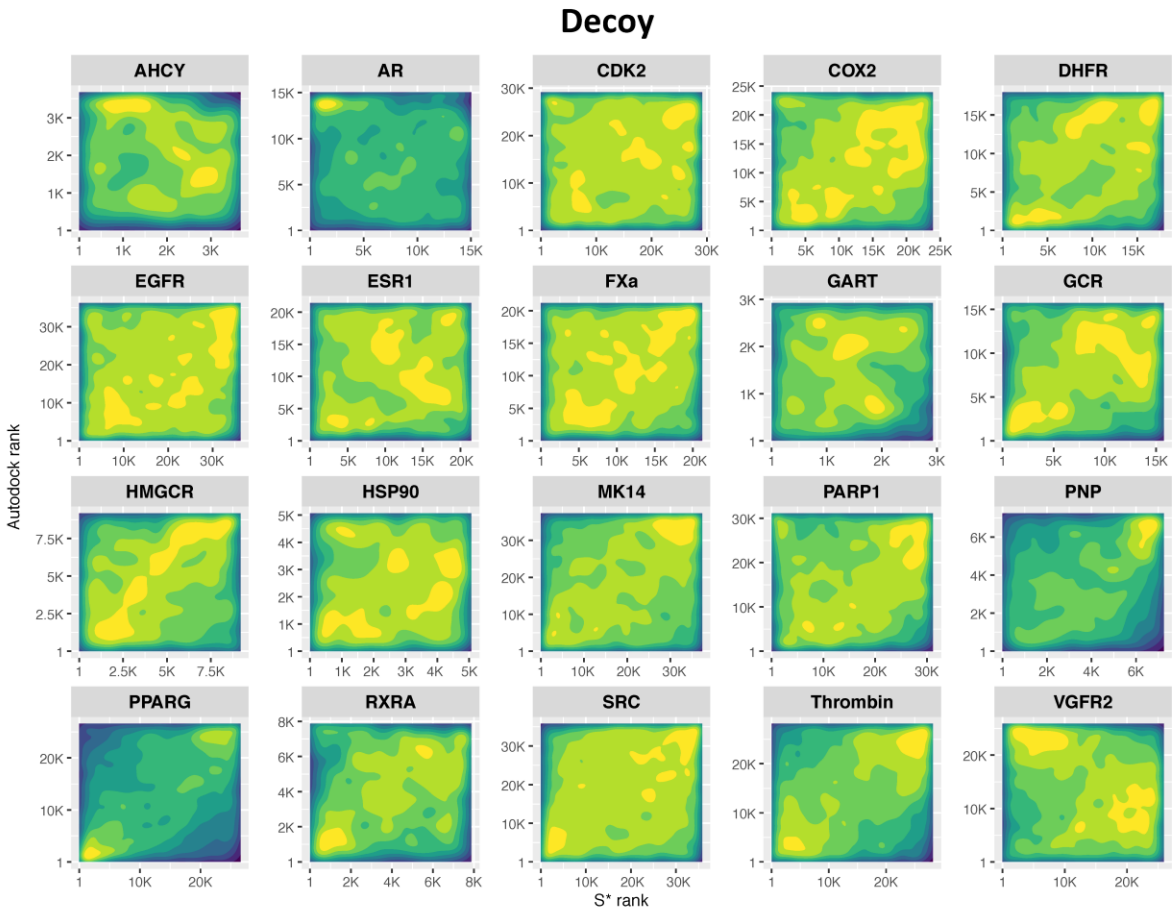


Supplemental Figure 3: *ceSAR*-S* (X-axis ranks) vs. *Autodock* (Y-axis rank) ranking of true positives (ligands) for 20 DUD-E target libraries. Yellow: high density; Dark Blue: low density.

The distributions of true ligands (**Supplemental Figure 1**) and decoys (**Supplemental Figure 2**) are shown using blue (low density) to yellow (high density) color scale. In general, there is little correlation between the two rankings. Even when the bulk of true positives ranks highly by both methods, i.e., achieve ranks close to 1 on both X and Y-axes, corresponding to yellow density peaks near the left bottom corner, which is observed for AHCY, COX2, CDK2, ESR1, FXa SRC and VGFR2, there are important differences remaining. For example, compared to *Autodock*, *ceSAR*-S does better for VGFR2 and worse for FXa in terms of median true positive rank. Differences between the two signals are even more pronounced for the remaining targets, with docking ranking the bulk of true positives higher for AR, GART, GCR, DHFR, HMGCR,

PARP1, RXRA and Thrombin and signature connectivity for EGFR, HSP90, MK14. At the same time, different subsets of decoys are often pushed towards lower ranks, and thus separated from true positives, by either *ceSAR*-S or *Autodock*, highlighting the complementarity of both approaches. This is particularly pronounced in the case of AHCY, ESR1, GART1, HSP90 and VGFR2.

We also consider an alternative form of the method, referred to as *ceSAR*-S*, that finds the closest concordant LINCS analog for each candidate compound q, and then ranks the compounds by using Fisher's consensus to combine chemical similarity and signature concordance scores (see Methods section in the main manuscript). In this form of the target structure independent method, candidate compounds with more distant closest analogs can get elevated in ranking if their signature concordance with the target gene KD is high. The ranking of ligands and decoys using *ceSAR*-S* vs. *Autodock* is shown in **Supplementary Figures 3** and **4**.



Supplemental Figure 4: *ceSAR*-S* (X-axis ranks) vs. *Autodock* (Y-axis rank) ranking of true negatives (decoys) for 20 DUD-E target libraries. Yellow: high density; Dark Blue: low density.

While the overall results are qualitatively similar, we note several differences. Marked improvements over *ceSAR*-S in terms of ligand ranks are observed for several targets, including

AR, DHFR, GCR and PNP. However, this accompanied by subsets of decoys also scoring near the top rank, making the discrimination within the top-ranking small library subsets challenging. Overall, *ceSAR*-S* improves somewhat the performance in terms of separation of true positives and decoys, while achieving a lower precision at the extreme library reduction, as compounds with strongly concordant distant analogs, including potential pathway inhibitors, are also retained in this form of the method. These differences further motivated using machine learning to combine concordance, chemical similarity, and docking-based scores into a consensus method.

Fast exact chemical similarity search using *minSim* (S2)

The LINCS library of drug-like molecules comprises over 40,000 compounds, of which a subset of over 15,000 transcriptionally profiled molecules are available through *iLINCS*. Likewise, the user defined library of small molecules Q to be ranked can comprise tens to hundreds of thousands of molecules in the context of virtual screening. Therefore, retrieval of closest LINCS analogs of candidate compounds, which requires that the chemical similarity measure, $\sigma\ (q, k)$, is computed for all pairs (q, k) between candidate compounds, $q \in Q$, and LINCS compounds, $k \in L$, represented as binary fingerprints, can become computationally expensive. To address this computational bottleneck, an efficient solution for computing the Jaccard similarity measure (Tanimoto coefficient) and retrieving the closest matches for the case of sparse binary vectors (fingerprints) is introduced here.

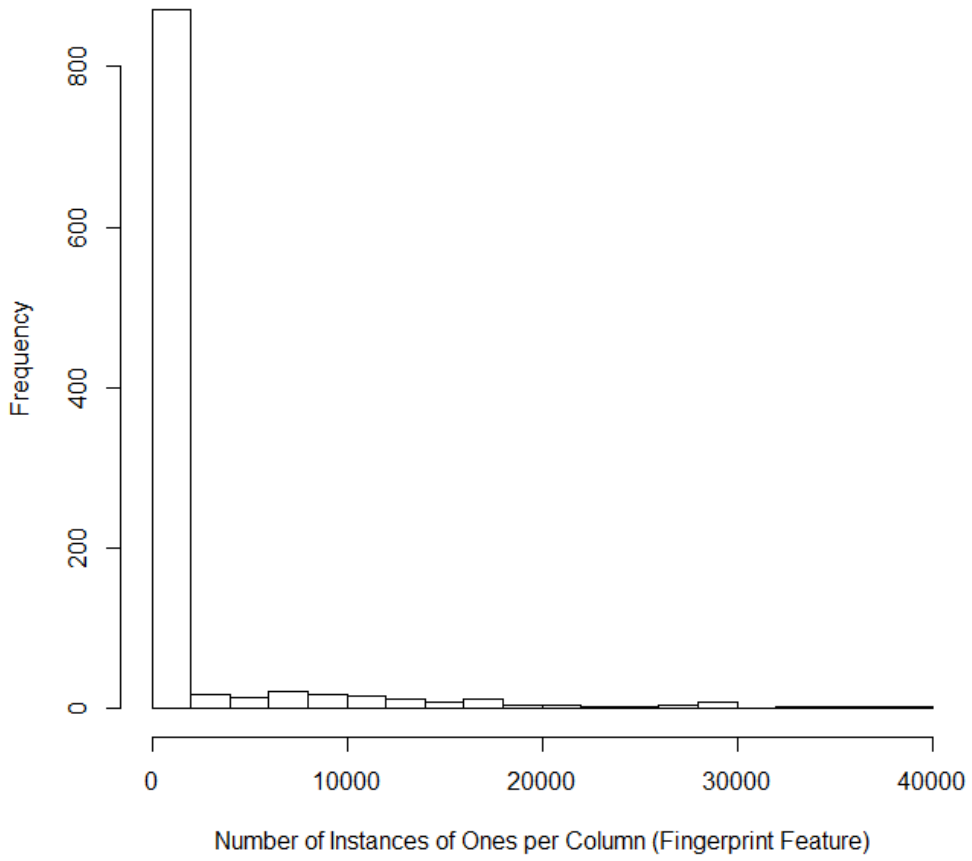
Binary fingerprints are widely used in cheminformatics for efficient chemical similarity search and SAR analyses (*45–48*). In this approximation, small molecules are represented as binary vectors indicating the presence of substructures (subgraphs), pharmacophores or chemical groups (*45*, *46*). Here, we use the 1024-bit atom-pair fingerprint representation (*45*, *49*), as generated by the *ChemmineR* package (*38*, *53*). This representation of a compound allows one to capitalize on the sparsity of the binary vector.

Supplemental Figure 5 shows the distribution of ones for the LINCS compound database for each of the 1024 fingerprint features (columns of the fingerprint matrix). Note that most columns are indeed very sparse, with only few columns that have between 10,000 and 30,000 ones and therefore have a balanced split between the two classes (presence vs. absence of a substructure or atom-pairs). On the other hand, **Supplemental Figure 6** shows the distribution of ones for compounds in the LINCS database. As can be seen from the figure a majority of LINCS compounds have under 100 ones in their respective fingerprint which highlights the sparsity of the database fingerprint vectors used here.

Consider now a search for a query compound $q \in Q$ against database compounds $k \in L$ using such defined binary fingerprint. Note that the formula for the Tanimoto coefficient, $\sigma(q, k)$, which is defined for two binary fingerprints q and k as the ratio of the number of positions with ones in both q and k and the number of positions with ones in either q or k, can be written in the following form:

$$\sigma(q, k) = \frac{sim\ (q, k)}{m(q) + m(k) - sim\ (q, k)}$$

where $m(q)$ and $m(k)$ are the number of ones that can be pre-computed for all database molecules k, while $sim\ (q,k)$ is the number of ones in common for q and k.

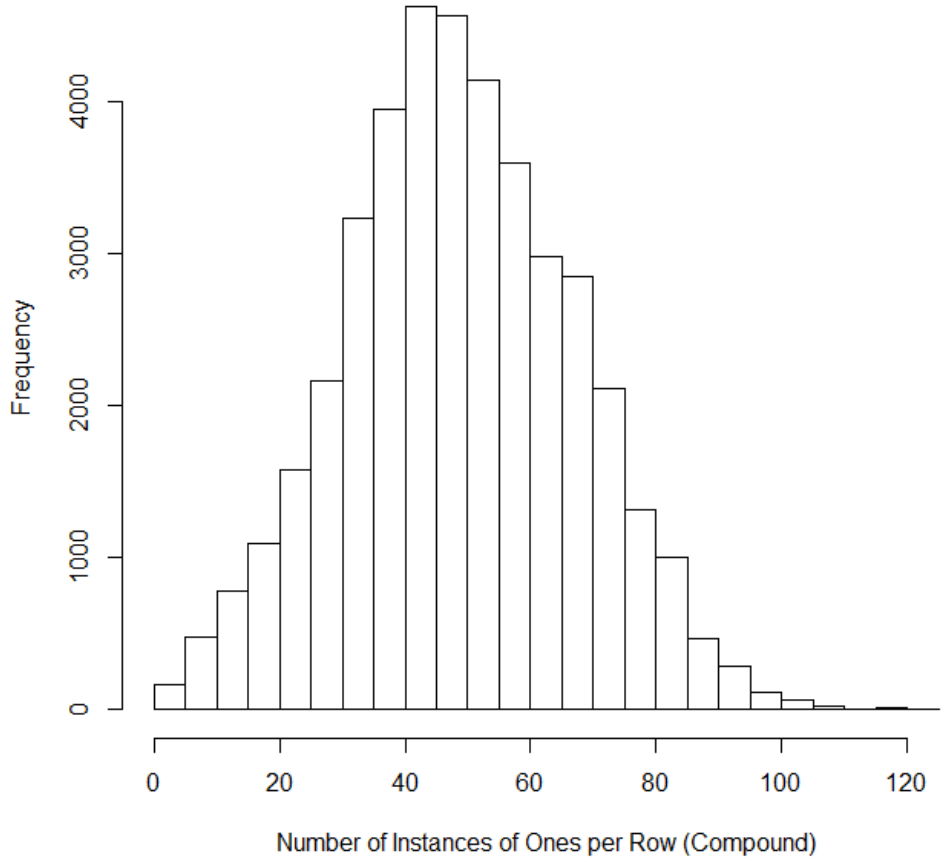


Supplemental Figure 5: Histogram showing the distribution of ones for LINCS compounds of each fingerprint feature (column in the binary fingerprint).

Note that the computation of $sim\ (q,k)$ can be limited to only those columns in the binary fingerprint where q is in the minority state, which is assumed to be 1. Note also that a simple re-coding of states for columns where this is not the case can be used to guarantee that ones correspond to minority states. Furthermore, by using pre-processing of the reference data set of compounds (here LINCS library) one can optimally exploit the sparsity in each column by pre-computing indexes of database compounds in the minority state at each column, as illustrated in **Supplemental Figure 7**.

To optimally exploit the sparse nature of binary fingerprints used to represent chemical moieties, the following list of database vectors k_i is pre-computed for each column j in the fingerprint:

$$ones(j) = \{k_i \mid k_i\,(j) = 1\}.$$



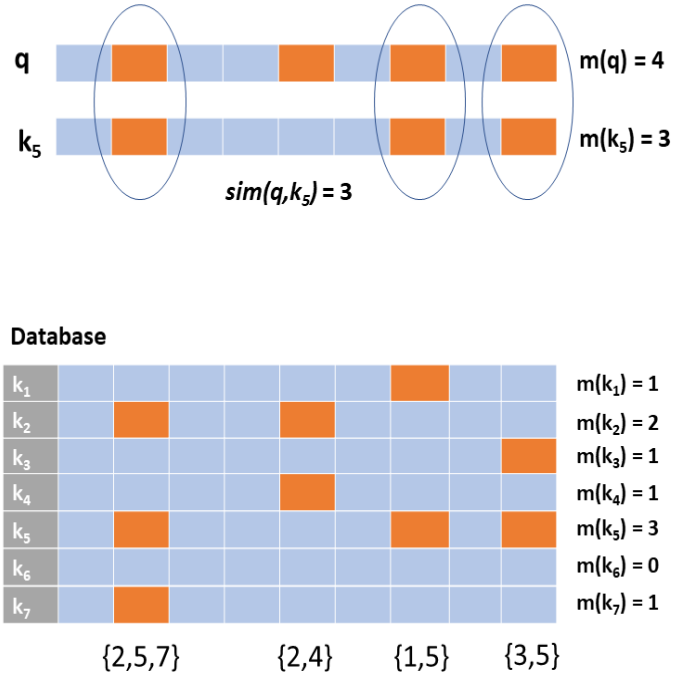
Supplemental Figure 6: Histogram portraying the distribution of ones in each row or compound for the LINCS library of compounds. Note that none of the compounds in the LINCS library has more than 120 ones out of 1024 positions in the binary fingerprint used here.

The *minSim* (for *minority Sim*) algorithm, computes all Tanimoto coefficients for a query molecule q by updating integer counters $sim\ (q, k)$, which are set to zero for all k at the beginning of the search, in a simple loop over minority columns in q and minority lists in each minority column:

for all minority columns j **in** q

for all k_i **in** $ones(j)$

$$sim(q, k_i) = sim(q, k_i) + 1$$



Supplemental Figure 7: The principle of the *minSim* algorithm for fast exact similarity search and retrieval of chemical analogs from a pre-processed dataset of small molecules represented as sparse binary fingerprints. Ones, or minority states, are represented by yellow boxes, whereas blue boxes represent zeros. The pre-computed lists of molecules in minority state at each column are shown in brackets below each fingerprint column.

By combining such obtained integer counters (numbers of ones for each (q, k) pair) with the precomputed numbers of minority states (ones) in each database vector, $m\,(k)$, *minSim* efficiently computes the Tanimoto coefficient by only considering those pairs of columns j in the fingerprint and rows k_i in the database that are both in the minority state, and thus contribute to the Jaccard similarity. The implementation of the algorithm in R is included in **Supplemental Figure 8**.

We posit that *minSim* optimally exploits the sparse nature of binary fingerprints by considering only those fingerprint columns (positions) where the query molecule q is in the minority state, and by using precomputed lists of all database compounds k that are in the minority states at these positions. Note also that *minSim* results in exact Jaccard similarity indexes for fast chemical similarity search, without using approximate techniques, such as those based on hashing (*36*, *37*, *53*).

```
minSim <- function(concordant_lincs_compounds, candidate_set, m, ones)
{
    tanimoto_query <- list()
    lincs_compound_max_tanimoto <- list()

    for (k in 1:dim(candidate_set)[1]){
        ones_query <- which(candidate_set[k,] == 1)
        m_query <- length(ones_query)
        sim <- vector("integer", length=41572)

        if (m_query == 0){
            next
        }

        for (j in 1:length(ones_query)){
                sim[ ones[[ ones_query[j] ]] ] <- sim[ ones[[ ones_query[j] ]] ] + 1
        }
        tanimoto = sim / (m + m_query - sim)

        tanimoto_query[[k]] <<- max(tanimoto[concordant_lincs_compounds])
        lincs_compound_max_tanimoto[[k]]  <<- which.max(tanimoto[concordant_lincs_compounds])

    }
}
```

Supplemental Figure 8: An R implementation of the *min-Sim* algorithm. Note that the search is performed against the entire (pre-processed) LINCS dataset, rather than a subset concordant with the target KD.

As can be seen from **Supplemental Table 1**, for the retrieval from the LINCS library for different DUD-E datasets, *minSim* provides between 60 and 150-fold speed-up compared to fpSim function, which represents current approaches for exact chemical similarity search (*53*). Note that

these speed-ups are consistent with the observed levels of sparsity in the LINCS dataset, while also reflecting the varying degree of sparsity in DUD-E datasets of query molecules.

Supplemental Table 1. A comparison of times to run similarity search for all compound for all DUD-E targets. *fpSim*, as a representative state-of-the-art method is compared with the *minSim* algorithm optimized for binary sparse vectors. The final column shows the increase in speed of *minSim* vs. *fpSim*.

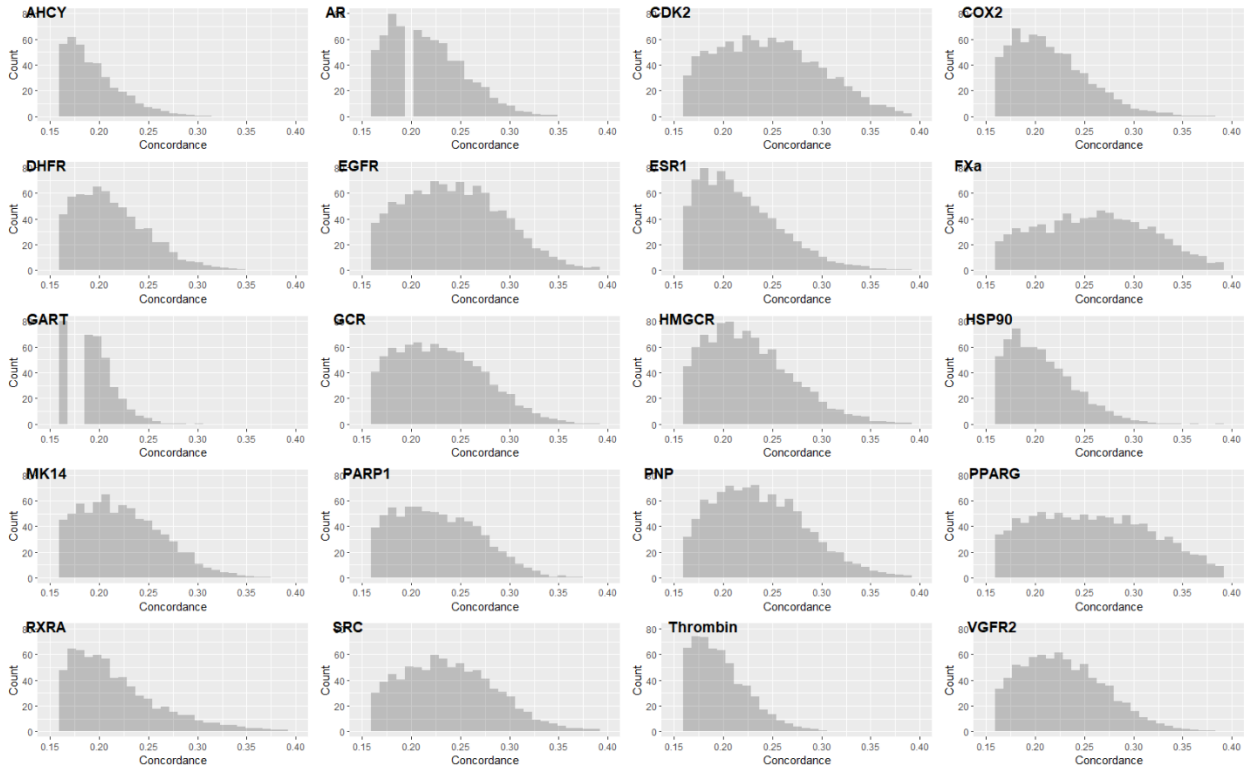
Target	Library Size	fpSim (min)	minSim (min)	Increase in Speed
AHCY	3673	46.059	0.335	137.39
AR	15026	223.313	2.386	93.60
CDK2	29126	414.321	4.568	90.70
COX2	23936	323.638	3.870	83.64
DHFR	17950	232.602	2.428	95.81
EGFR	36274	541.216	7.034	76.95
ESR1	21445	296.228	3.739	79.22
FXa	21209	278.580	3.991	69.79
GART	2926	35.932	0.462	77.76
GCR	15748	225.486	3.062	73.64
HMGCR	9183	122.989	1.892	65.00
HSP90	5067	64.866	0.993	65.35
MK14	37347	539.328	6.823	79.04
PARP1	31171	438.801	4.440	98.83
PNP	7249	95.761	0.640	149.62
PPARG	26590	367.176	5.261	69.79
RXRA	7869	99.169	1.629	60.89
SRC	35790	523.830	7.123	73.54
Thrombin	28182	404.822	4.941	81.93
VGFR2	25900	408.805	4.655	87.82

LINCS library of transcriptional signatures (S3)

ceSAR makes use of the LINCS library of transcriptional signatures to identify candidate compounds that induce signatures concordant with the loss (or gain) of function of the target gene. The L1000 high-throughput gene expression assay that measures the mRNA transcript abundance of 978 "landmark" genes from human cells is used as the primary transcriptional readout in LINCS. Computational analysis of gene expression compendia by the Broad Institute's CMap group indicated that these 978 genes capture a large fraction (~80%) of information of the transcriptome at a fraction of the cost (*4*, *12*).

For most genes, multiple shRNA knock-down signatures are generated to define consensus signatures and minimize off-target effects (*12*, *54*). *ceSAR* uses L1000 signatures to identify small molecules that are 'concordant' with a target gene knock-down. Small molecules are considered

concordant to a genetic perturbation if they have an extreme Pearson correlation-based concordance score adopted by *i*LINCS, of 0.162 or higher with a consensus shRNA signature for a gene target.



Supplemental Figure 9: Tail distributions of *iLINCS* concordance values between LINCS target KD and small molecule signatures for 20 DUD-E targets. All LINCS compounds are considered, i.e., no similarity to DUD-E library compounds filter is applied, and the counts on the y axis are normalized by the number of available KDs per target (for most targets there are between 5 and 10 KD consensus signatures from different cell lines).

Concordance scores are retrieved using API methods of *iLINCS* (*54*) and a library of LINCS signatures accessible therein that comprises 143,374 signatures of 15,193 small molecules (a subset of the entire LINCS library) and 37,275 consensus gene KD signatures of 4,348 genes in 242 cell lines. The distributions of concordance scores retrieved from *iLINCS* for DUD-E datasets (discussed in the next section) are summarized in **Supplemental Table 2** and **Supplemental Figure 9**.

As can be seen from the figure, the maximum concordance score for three targets, i.e., AHCY, GART and thrombin is less than 0.35, indicating overall weak transcriptional signatures, or the lack of correlation between chemically vs. genetically induced loss of function. This does not bode well for the signature connectivity-based ranking of candidate compounds. Indeed, these three targets proved to be difficult for *ce*SAR, although the consensus approach works well for AHCY. The shape of the distribution of concordance values, and the maximum value observed, can serve as a measure of the likelihood of success for the signature connectivity-based and target

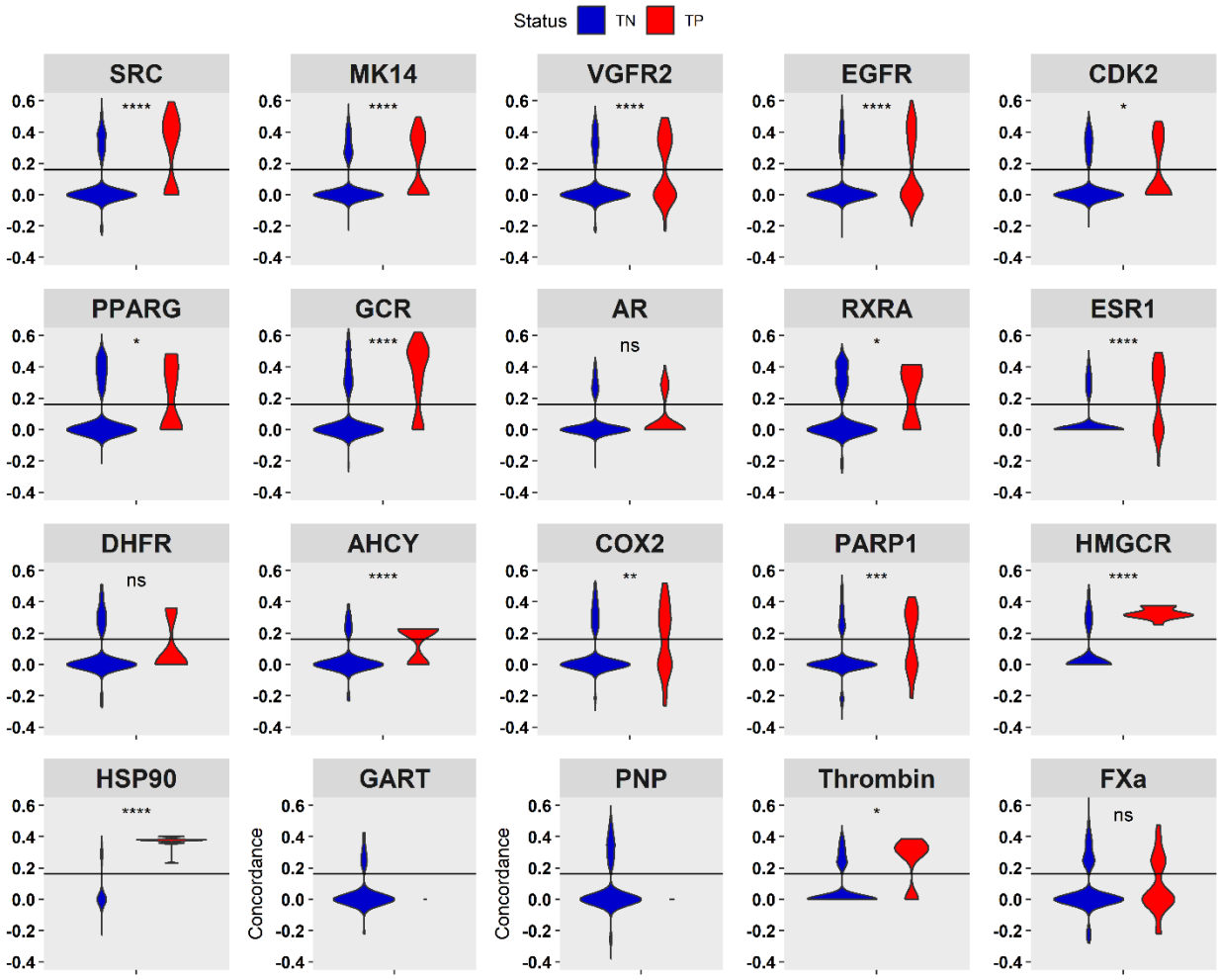
structure-independent *ceSAR*-S, and has informed the choice of features for the machine learning-based consensus approach.

Supplemental Table 2. Summary of the concordance and chemical similarity distributions for all targets of the DUD-E library. The target, library size, number of true positives (ligands) and number of true negatives (decoys) are shown in the first four columns with the concordance and chemical similarity with 'concordant' LINCS analogs distributions following. The distributions are displayed with the range followed by the median in parenthesis for both true positives and true negatives.

Target	Library Size	TP	TN	Max Concordance (Tail Median)		Tanimoto Range (Median)	
				TP	TN	TP	TN
AHCY	3673	190	3483	0.281 (0.178)	0.309 (0.191)	0.36 - 0.941 (0.619)	0.125 - 1 (0.591)
AR	15025	523	14502	0.334 (0.216)	0.357 (0.219)	0.385 - 1 (0.694)	0.188 - 1 (0.727)
CDK2	29119	798	28321	0.403 (0.268)	0.416 (0.261)	0.429 - 1 (0.764)	0.222 - 1 (0.705)
COX2	23929	531	23398	0.349 (0.221)	0.41 (0.216)	0.417 - 1 (0.787)	0.143 - 1 (0.714)
DHFR	17949	566	17383	0.338 (0.236)	0.38 (0.213)	0.472 - 1 (0.721)	0.2 - 1 (0.681)
EGFR	36273	832	35441	0.393 (0.282)	0.459 (0.253)	0.619 - 1 (0.789)	0.125 - 1 (0.714)
ESR1	21444	626	20818	0.414 (0.246)	0.434 (0.216)	0.541 - 1 (0.772)	0.25 - 1 (0.722)
FXa	21209	792	20417	0.444 (0.252)	0.438 (0.271)	0.532 - 1 (0.754)	0.286 - 1 (0.708)
GART	2926	201	2725	0.26 (0.164)	0.3 (0.185)	0.537 - 0.852 (0.627)	0.24 - 1 (0.634)
GCR	15746	563	15183	0.385 (0.254)	0.388 (0.237)	0.465 - 1 (0.75)	0.3 - 1 (0.75)
HMGCR	9183	299	8884	0.399 (0.224)	0.406 (0.234)	0.286 - 1 (0.705)	0.25 - 1 (0.736)
HSP90	5067	125	4942	0.301 (0.227)	0.392 (0.204)	0.536 - 1 (0.746)	0.364 - 1 (0.706)
MK14	37345	915	36430	0.379 (0.241)	0.401 (0.228)	0.553 - 1 (0.815)	0.111 - 1 (0.719)
PARP1	31151	742	30409	0.365 (0.238)	0.374 (0.233)	0.515 - 1 (0.744)	0.111 - 1 (0.688)
PNP	7245	233	7012	0.388 (0.267)	0.452 (0.238)	0.364 - 1 (0.714)	0.2 - 1 (0.68)
PPARG	26590	723	25867	0.421 (0.26)	0.474 (0.277)	0.568 - 1 (0.74)	0.318 - 1 (0.745)
RXRA	7869	162	7707	0.407 (0.22)	0.515 (0.22)	0.489 - 1 (0.711)	0.3 - 1 (0.78)
SRC	35790	831	34959	0.446 (0.288)	0.446 (0.246)	0.583 - 1 (0.793)	0.25 - 1 (0.725)
Thrombin	28181	861	27320	0.324 (0.191)	0.349 (0.195)	0.456 - 0.942 (0.65)	0.67 - 1 (0.655)
VGFR2	25900	620	25280	0.394 (0.275)	0.419 (0.231)	0.45 - 1 (0.817)	0.292 - 1 (0.706)

On the other hand, of those targets that are characterized by long tails of the distribution (the maximum concordance greater than 0.6), only PNP results in no enrichment by *ce*SAR, as discussed in the main paper and in the next section. However, as can be seen from **Figure 8** and **Supplemental Table 2**, PNP binders are not well represented in the LINCS database of transcriptionally profiled compounds, pointing out the reason for failure in this case. Note that this is also true for AHCY and thrombin.

As shown in Supplemental Figure 5, these two targets have no true binders (TP) included in LINCS (no LINCS analogs with Tanimoto coefficient of 1), while PNP has just several, making these 3 targets challenging in this regard. However, *ceSAR* effectively transfers signature connectivity-based signal from LINCS analogs via chemical similarity mapping, thus expanding the application of the method. This is further illustrated in in terms of distributions of concordance values for true positives vs. true negatives while identifying LINCS analogs with decreasing Tanimoto coefficients from 1.0 (direct counterparts in LINCS, **Figure 8**, panel B), 0.9 (very close analogs, **Supplemental Figure 10**), and 0.8 (close analogs, **Figure 8**, panel C).



Supplemental Figure 10: Distributions of concordance scores for true positives (red) vs. true negatives (blue) with concordant LINCS analogs mapped using Tanimoto coefficient of 0.9. Note generally higher concordance values for true positives, except for 2 targets (GART and PNP) for which the number of true positives with concordant LINCS analogs at Tanimoto of 0.9 is less than 3; this can be contrasted with Tanimoto of 0.8 (**Figure 8**, panel C), which allows for robust analysis of the performance for all targets as the number of analogs to be considered increases.

DUD-E Benchmark (S4)

Twenty target structures from the Database of Useful Decoys Enhanced (DUD-E) were downloaded from Protein Data Bank (PDB). Each structure was prepared by using *AutodockTools* v.1.5.6. The PDBQT file was generated after removing the bound ligand from the binding pocket, removing water molecules, and adding hydrogens as a pre-processing step. The grid box center was identified as the center of the bound ligand with the dimension of 20x20x20Å. Details for 20

targets with grid box centers are provided in **Supplemental Table 3**. Active ligands and drug-liked decoys for each target were downloaded from the DUD-E library in the SDF format, which were then converted to the PDBQT format before screening.

Supplemental Table 3. DUD-E targets and grid centers used for *AutoDock* simulation.

Target	Protein name	PDB ID	Grid center
AHCY	Adenosylhomocysteinase	1li4	47, -20, 103
AR	Androgen receptor	2am9	27, 2, 5
CDK2	Cyclin-dependent kinase 2	1h00	2, 28, 9
COX2	Cyclooxygenase-2	3ln1	31, -22, -17
DHFR	Dihydrofolate reductase	3nxo	15, 6, 0
EGFR	Epidermal growth factor receptor erbB1	2rgp	16, 35, 92
ESR1	Estrogen receptor alpha	1sj0	31, -2, 25
FXa	Coagulation factor X	3kl6	2, -8, -13
GART	GAR transformylase	1njs	75, 23, 32
GCR	Glucocorticoid receptor	3bqd	40, 30, 9
HMGCR	HMG-CoA reductase	3ccw	-7, -1, 32
HSP90	Heat shock protein HSP 90-alpha	1uyg	3, 12, 25
MK14	MAP kinase p38 alpha	2qd9	-3, -1, 24
PARP1	Poly ADP-ribose polymerase-1	3l3m	26, 11, 27
PNP	Purine nucleoside phosphorylase	3bgs	15, 11, 58
PPARG	Peroxisome proliferator-activated receptor gamma	2gtk	4, 25, 17
RXRA	Retinoid X receptor alpha	1mv9	61, 46, 31
SRC	Tyrosine-protein kinase SRC	3el8	2, 5, 8
Thrombin	Thrombin	1ype	17, -12, 22
VGFR2	Vascular endothelial growth factor receptor 2	2p2i	38, 36, 12

Virtual screening with docking was performed using *Autodock* v. 4.2.6 on a Linux-based computational cluster, consisting mostly of 16 Intel (R) Xeon (R) CPU E5-2667 v3 @ 3.20GHz core nodes (simulations for each target were performed distributing the computation over 50 nodes). Genetic algorithm was applied to search the optimal ligand binding conformation using the number of energy evaluations of 1,000,000, the population size of 150, and the number of repeated docking runs of 50, which represents a relatively deep level of sampling (*55*).

In order to test the effect of using a different docking method, virtual screening using *MTiOpenScreen*, a web-based implementation of *Autodock Vina*, was performed on eight out of 20 targets, including 5 kinases (CDK2, EGFR, MK14, SRC, VGFR2), 2 nuclear receptors (ESR1, PPARG) and 1 enzyme (FXa). Since *MTiOpenScreen* accepts the protein structure in the PDB format, each structure was prepared by removing the bound ligand from the binding pocket. Active ligands and drug-liked decoys for each target were downloaded from the DUD-E library in the SDF format.

Since the *MTiOpenScreen* only accepts up to 5,000 compounds per run, multiple subsets of compounds were used for each compound database. In each run, a subset consisting of 100 active ligands and 3,900 decoys (or less than 100 and 3,900, respectively, in the last subset), was uploaded as the SDF file to the server. The default parameters were applied for each analysis except the grid box calculation was set to the custom mode. The grid box center was identified as the center of the bound ligand and the dimension of 20x20x20Å. Details for these targets with PDB ID and grid box centers were provided in **Supplemental Table 4**. We have also used the predecessor of DUD-E, i.e., the original DUD benchmark for further benchmarking observing similar overall trends.

Supplemental Table 4. Subset of DUD-E targets and grid centers used for *MTiOpenScreen* simulations.

Target	Protein name	PDB ID	Grid center
CDK2	Cyclin-dependent kinase 2	3ezr	-2, 38, 3
EGFR	Epidermal growth factor receptor erbB1	2rgp	16, 35, 92
ESR1	Estrogen receptor alpha	1sj0	31, -2, 25
FXa	Coagulation factor X	3kl6	2, -8, -13
MK14	MAP kinase p38 alpha	2qd9	-3, -1, 24
PPARG	Peroxisome proliferator-activated receptor gamma	2gtk	4, 25, 17
SRC	Tyrosine-protein kinase SRC	2oiq	-3, -33, 8
VGFR2	Vascular endothelial growth factor receptor 2	2p2i	38, 36, 12

Machine Learning-based Consensus Approach (S5)

The simple consensus-based *ceSAR* approaches, referred to as *ceSAR*-C_{100}, *ceSAR*-C_5 or *ceSAR*-C_1 in the main manuscript, combine initial filtering and ranking using similarity to 'concordant' LINCS analogs with subsequent consensus re-ranking based on docking results. The method is very simple and takes the form of a ranking-based filter, with just 2 meta-parameters, namely the concordance threshold (see Section S1) and the fraction of the library initially selected by using the signature connectivity-based *ceSAR*-S ranking to be subsequently re-ranked using the consensus with docking results (A).

The success of such a simple method led us to explore if machine learning techniques could be used to learn a more complex model that improves upon the simple consensus. We hypothesized that by learning from data, concordance, chemical similarity, and docking-based features can be optimally combined into an accurate and robust classification model to discriminate true positives from true negatives and provide improved consensus ranking of candidate molecules.

To test that hypothesis, several random forest (RF) and neural network (NN) classifiers were trained, combined into an ensemble classifier, and tested using the set of 20 DUD-E target compound libraries. *Leave-one-out* target cross-validation was employed to successively train the

model on 19 targets and test on the remaining target, using R packages *RandomForest* (*56*) and *neuralnet* (*57*), respectively.

Multiple sets of features were evaluated, and the final models included in the consensus were trained either with the set of 11 or 13 features listed in **Supplemental Table 5** that are designed to capture the complementary signals from signature connectivity analysis and docking simulations. These features include: the Tanimoto similarity of the DUD-E compound to its LINCS analog, the (scaled) docking energy of compound and target, concordance of the LINCS analog, as well as these values normalized to the median, plus rank-based features encoded as ratio of the compound rank and the total number of compounds for *ceSAR*-S, *Autodock*, and concordance separately, followed by the 2- and 3-feature consensus ranks (Fisher_2 and Fisher_3). For docking-based features, the median of the docking energy of compound and target across 50 runs were used. The fractional rank-based features were motivated by the Fisher method for combining probabilities or p-values derived from individual models or statistical tests that sums logs of probabilities to obtain the consensus score:

$$consensus\_score \sim \sum_{i=1}^{k} \log\ (p_i)$$

Two additional features, which are highlighted in yellow and expand the set of features from 11 to 13, were included to allow for discrimination of the strong concordance-on-strong distribution vs. weak concordance-on-strong distribution of concordances between the signature of the candidate compound and the target knockdown. These features are defined as (1) the maximum concordance value of all LINCS analogs and the gene knockdown target (c_{max}) over 1 minus the concordance $c^*(k,t)$ of the compound and gene knockdown target and (2) the maximum concordance value of all LINCS analogs and the gene knockdown target (c_{max}) minus the concordance $c^*(k,t)$ of the compound and gene knockdown target divided by the maximum concordance of all LINCS analogs and the gene knockdown target (c_{max}). An example of distributions of individual feature values for true positives (ligand) and negatives (decoys) for EGFR are shown in **Supplemental Figure 11**.

The training was performed with a balanced approach in which all true positives were utilized for each DUD-E library, while the negative examples were sampled from two subsets of true negatives: one from the top 5 % (*ceSAR*-S ranking) and the other from the remaining 95% of the library. For NNs, 2 hidden layers with 5 neurons each were used for both 11 and 13 feature models. The sigmoid activation function, sum-of-squares error function, and rprop+ resilient back-propagation algorithm was used for the training. The threshold for convergence was determined by assessing the error function using threshold values of 0.1 to 0.5. The error did not vary substantially using these thresholds, and the threshold of 0.5 was used for further analysis to facilitate speed of convergence with little loss in error.

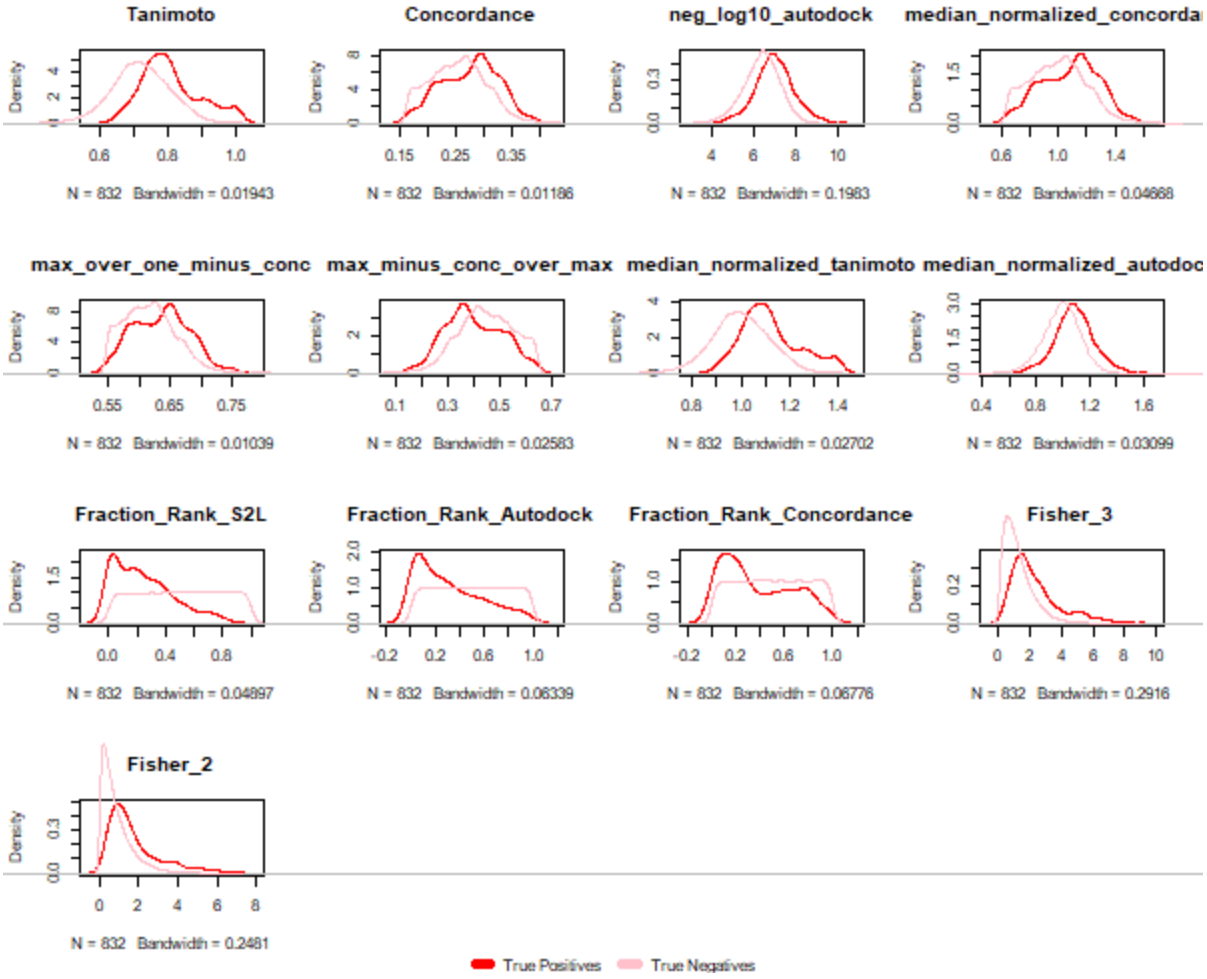
To assess the contribution of the individual features to the RF classifiers, the *importance* of individual features was assessed using the median decrease in the Gini Index for 20 models optimized using *Leave-one-out* target cross-validation. For both 11 and 13 feature models, the fraction of the library of *ceSAR*-S and the fraction of the library of *Autodock* were found to be the most dominant features. Thus, *ceSAR*-S and *Autodock* scores both play an important role and

complement each other, underlying the success of the ML-based consensus of the signature connectivity and docking-based scores.

Supplemental Table 5. Definition of features employed to train machine learning models.

Feature	Definition
Tanimoto	Chemical similarity to closest LINCS analog
Concordance	Concordance of closest LINCS analog
neg_log10_*Autodock*	-LOG10 (docking binding energy)
median_normalized_concordance	Concordance of closest LINCS analog normalized to the median of the concordance distribution
median_normalized_tanimoto	Chemical similarity to closest LINCS analog normalized to the median of the distribution
median_normalized_*Autodock*	Docking binding energy normalized to the median of the distribution
Fraction_Rank_ *ceSAR*-S	Rank of *ceSAR*-S adjusted by the number of compounds in the library (also referred to as Fraction_Rank_S2L)
Fraction_Rank_*Autodock*	Rank of *Autodock* adjusted by the number of compounds in the library
Fraction_Rank_Concordance	Rank of concordance adjusted by number of compounds in library
Fisher_3	Fisher's consensus of *ceSAR*-S, *Autodock*, and Concordance
Fisher_2	Fisher's consensus of *ceSAR*-S and *Autodock*
max conc over / 1- Conc	Maximum Concordance of the distribution divided by 1 minus the concordance of the closest LINCS analog
max_conc_minus_conc / max_conc	Maximum Concordance of the distribution minus the Concordance of the closest LINCS analog divided by the Maximum Concordance of the distribution

In order to further compare the performance of different machine learning methods, several ensemble support vector machine classifiers, using both Gaussian and linear kernels, as well as logistic regression baseline classifiers were trained using the same *leave-one-out* target cross-validation approach, and the R *caret* package (*58*), yielding results worse than the NN and RF ensemble model and therefore not included in the final consensus.



Supplemental Figure 11: An example of distributions of individual feature values for true positives (ligand) and negatives (decoys) for EGFR.

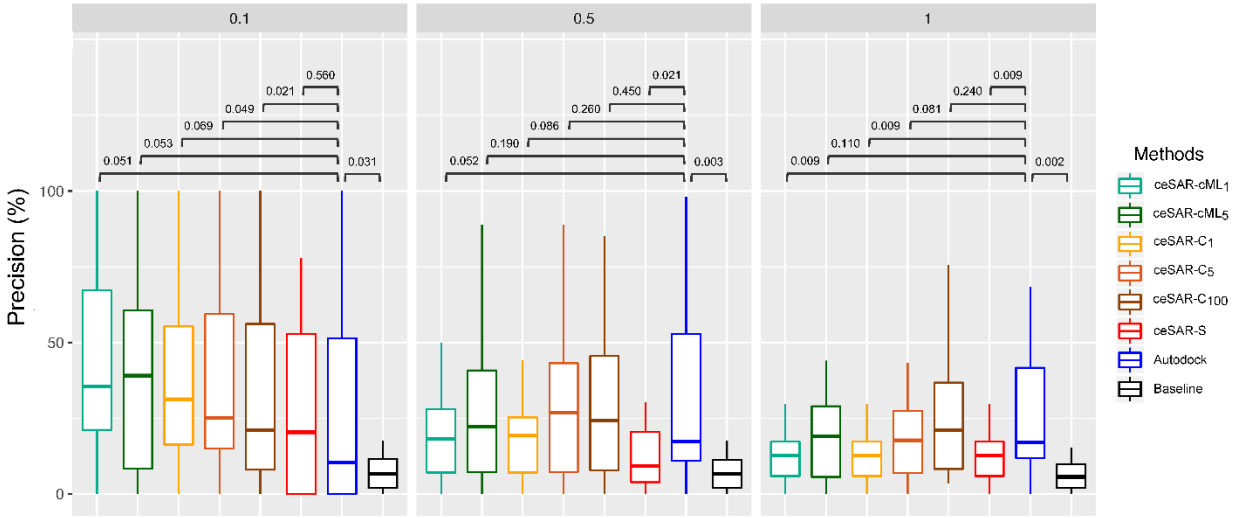
While further feature selection, aggregation and optimization of the model could well improve the results of machine learning-based predictors, the risk of overfitting on a limited set of targets and the success of simple, filter-based approaches argue against using more complex models without further extensive benchmarking on independent data sets.

Supplemental Results (S6)

The performance of docking simulation, referred to as A, the simple signature connectivity-based and target structure-independent *ceSAR*-S search, consensus methods referred to as C or cML for machine learning based consensus, and a simple baseline approach (B) that ignores signature connectivity while retrieving the closest LINCS analogs using the chemical similarity search, was compared on a set of 20 DUD-E targets. These targets spanned several classes of

proteins, including kinases (CDK2, EGFR, MK14, SRC, VGFR2), other enzymes (AHCY, COX2, DHFR, FXa, GART, HMGCR, PARP1, PNP, Thrombin), a chaperone (HSP90), nuclear receptors and transcription factors (AR, ESR1, GCR, PPARG, RXRA).

As discussed in the main paper, precision curves for the active ligand prediction over a range of library size of 0.1% to 100% were used to evaluate and compare the performance of both docking and connectivity enhanced approaches, providing a direct assessment of the fraction of true positives selected as the library is reduced in size by either signature connectivity analysis, docking binding affinity, or consensus ranking. In addition, boxplots representing the distributions of precision values at different library sizes were used to better assess the robustness and likelihood of failure for a specific target for each method tested here, as shown in **Supplemental Figure 12**.



Supplemental Figure 12: Distribution of precision values over 20 DUD-E targets at different library sizes. *Autodock* (A), baseline (B), *ceSAR*-S, and consensus approaches *ceSAR*-c[] with p-values obtained using pairwise Wilcoxon signed rank test shown on top to assess statistical significance.

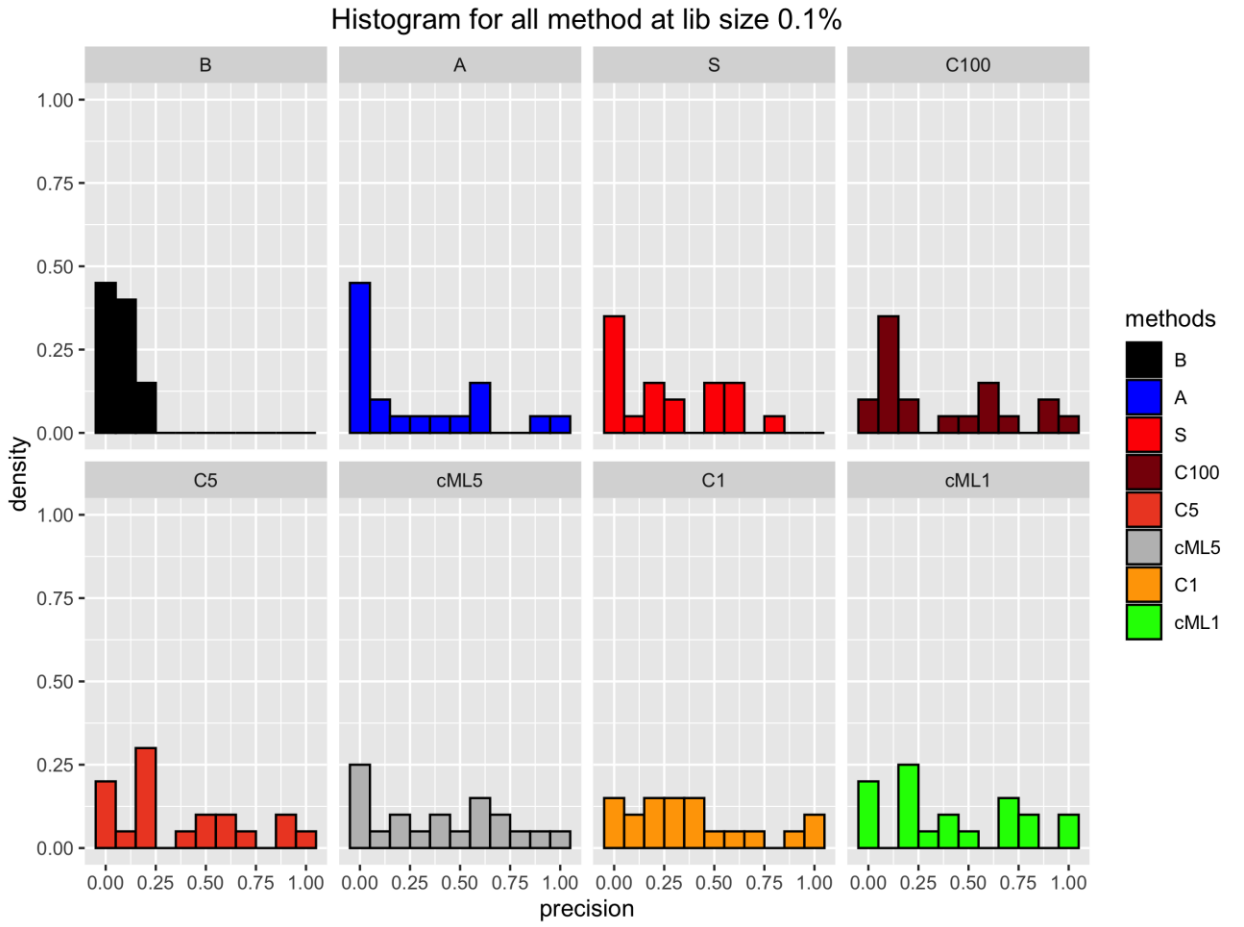
As can be seen from the figure, the performance of both docking and *ce*SAR approaches varies considerably over 20 targets. Importantly, all methods tested here significantly outperform the baseline (*Baseline*) at all three library reduction levels. The performance of *ceSAR*-S, which has a negligible computational cost compared to docking (*Autodock*), is better than that of docking at the furthest library reduction, yielding median precision of about 20% compared to about 10% for docking, although that shift is deemed statistically insignificant due to wide distributions of precision values (Wilcoxon signed rank test p-value of 0.56).

The consensus approaches improve upon both *ceSAR*-S and *Autodock* at 0.1% library size, achieving statistical significance or near significance with respect to docking (Wilcoxon test p-

values of 0.02 to 0.07), while performing on par with docking at 0.5%. Note that *ceSAR*-C_{100}, -C_5 (-cML_5), and -C_1 (-cML_1) use the consensus of the *ceSAR*-S and *AutoDock* ranking for the 100%, top 5% and top 1% of the library selected initially by *ceSAR*-S.

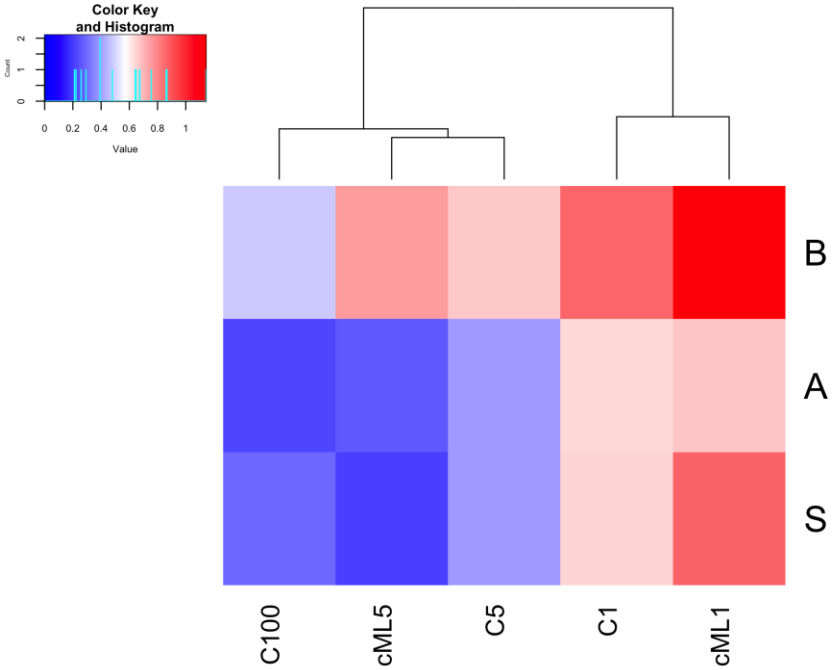
The distribution of precision values for the methods considered here at the furthest library reduction is dissected in detail in **Supplemental Figures 13** and **14**. As can be seen from the figure, the baseline method (B) fails to achieve precision of more 25% for any target, yielding less than 5% of true positives for 9 targets. The median precision of the baseline method over all targets is equal to 7.5% and this value is subsequently used as a reference to measure improvements on B. While both *Autodock* and *ceSAR*-S perform better than B, it should be noted that they still fail to yield more than 7.5% of true positives for 10 and 7 targets, respectively.

On the other hand, consensus methods improve upon both *Autodock* and *ceSAR*-S. In particular, *ceSAR*-C_1 and *ceSAR*-cML_1 consensus methods yield precision better than baseline median precision of 7.5% for 17 and 16 targets, respectively, compared with 10 for *Autodock*, resulting in chi-square p-values of less than 0.02 and 0.05, respectively. Thus, the distributions of precision values at the furthest library reduction are statistically significantly different for *ceSAR*-C_1 and *ceSAR*-cML_1 consensus methods compared to *Autodock*.



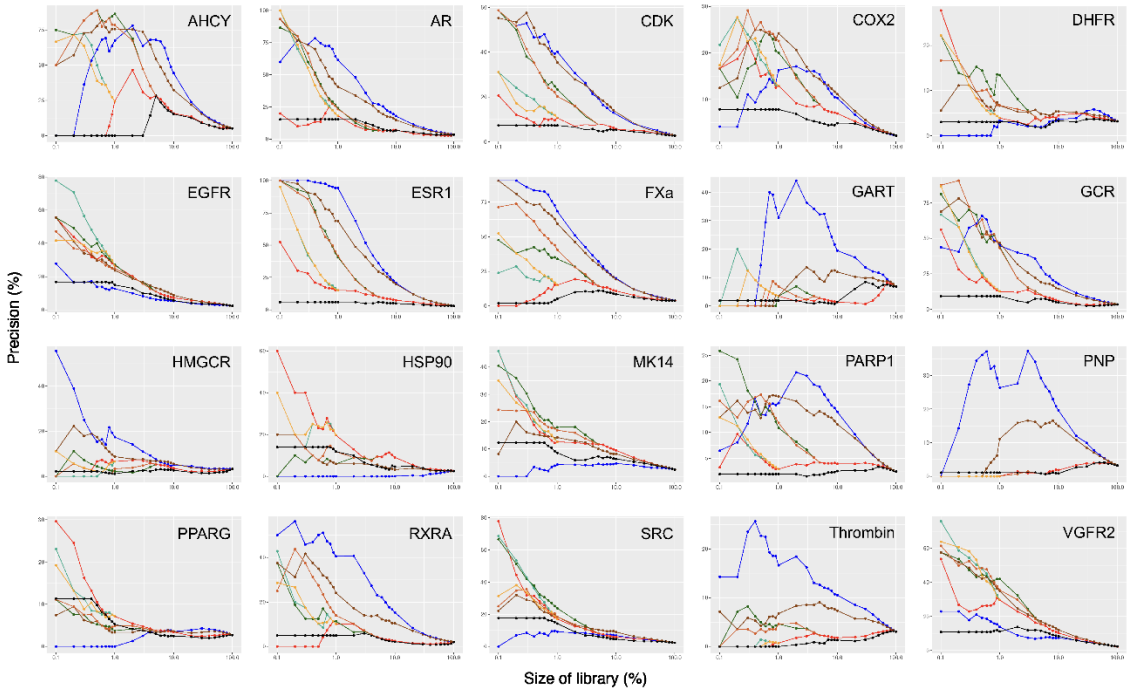
Supplemental Figure 13: Distribution of precision values over 20 DUD-E targets at the furthest library reduction for *Autodock* (A), baseline (B), *ceSAR*-S (S), and consensus approaches (c/C).

These two consensus methods, which combine signature connectivity analysis with docking for the top 1% of the library ranked by *ceSAR*-S, also have the most significantly different distributions over 20 targets with respect to the baseline (B), and individual *Autodock (*A) and *ceSAR*-S (S) methods, as measured by the Kullback-Leibler divergence measure, which is shown in **Supplemental Figure 14**. These observations led to the conclusion that *ceSAR*-C_1 and *ceSAR*-cML_1 consensus methods provide the best trade-off between speed and accuracy.

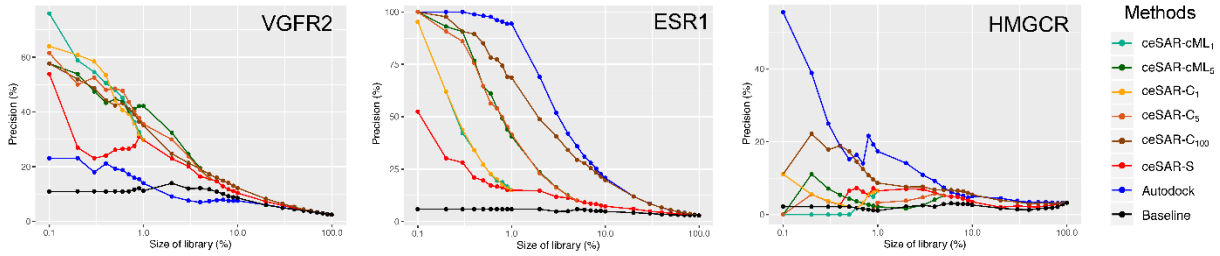


Supplemental Figure 14: Clustering of *ceSAR* consensus approaches (c/C) based on their Kullback-Leibler divergence from the distributions of precision values over 20 DUD-E targets at the furthest library reduction. *Autodock* (A), baseline (B), *ceSAR*-S (S). High vs. low Kullback-Leibler divergence measure values are indicated in red vs. blue scale.

Individual precision curves for 20 DUD-E targets and separately for three targets that are illustrative of varying degree of success of *ceSAR*, are included in **Supplemental Figures 15** and **16**. In a highly successful case (VGFR2), the *ceSAR*-S and consensus approaches show a better performance than *Autodock* even at the large library sizes (10%, 5%). In a moderately successful case (ESR1), *ceSAR*-S and/or consensus approaches provide a comparable performance to that of docking at small library sizes, while having an advantage of faster computational time. These two scenarios are observed for the majority of DUD-E datasets.



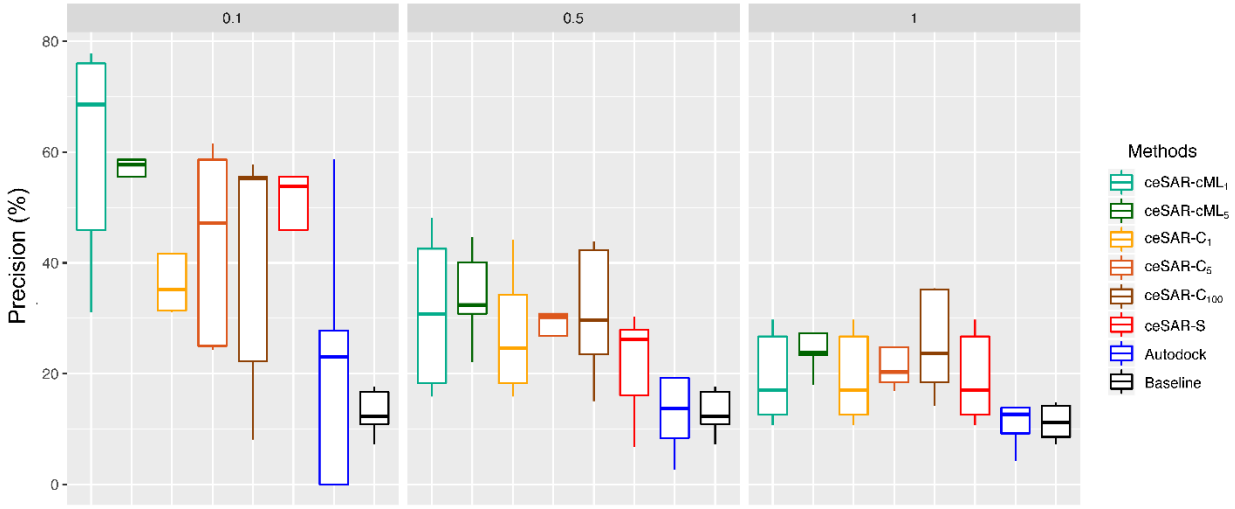
Supplemental Figure 15: The performance of *ceSAR*-S and consensus methods *ceSAR*-c/cML relative to *Autodock* and a simple baseline method (B). Results shown using the precision curve as the function of the library size for 20 DUD-E targets. This is the same as Figure 5B, included here for direct comparison with the results in the next 7 figures.



Supplemental Figure 16: The performance of *ceSAR*-S and consensus methods *ceSAR*-c/cML relative to *Autodock* and a simple baseline method (B). Results shown using the precision curve as the function of the library size for three targets representing kinases, nuclear receptors, and miscellaneous DUD-E classes of targets.

However, in a few cases, including HMGCR shown here, *ceSAR* performs poorly compared to docking, even if some degree of enrichment over the baseline is observed, except for the furthest library reduction. Note that the failure on HMGCR, which encodes a 3-Hydroxy-3-Methylglutaryl-CoA Reductase, cannot be rationalized by pointing out a short tail of the distribution of concordances in **Supplemental Figure 9**, or the total lack of close analogs in LINCS, even if their number is relatively low, as highlighted in **Figure 8** and **Supplemental Table 2**. Rather, it is the strength of concordance for the true negatives (decoys) relative to true positives (active compounds) – see **Supplemental Table 2**.

As discussed in the main manuscript, and in Supplemental Section S3, *ceSAR* is dependent on the availability of a representative set of transcriptionally profiled molecules that broadly cover the drug-like universe. While this is largely true about the overall LINCS library that comprises over 40,000 compounds (not all of which have been profiled transcriptionally), not all classes of drugs are well represented, or may not induce sufficiently strong signatures to be considered for the connectivity analysis.



Supplemental Figure 17: Performance on kinases using boxplots to compare the distributions of precision values at different library sizes.

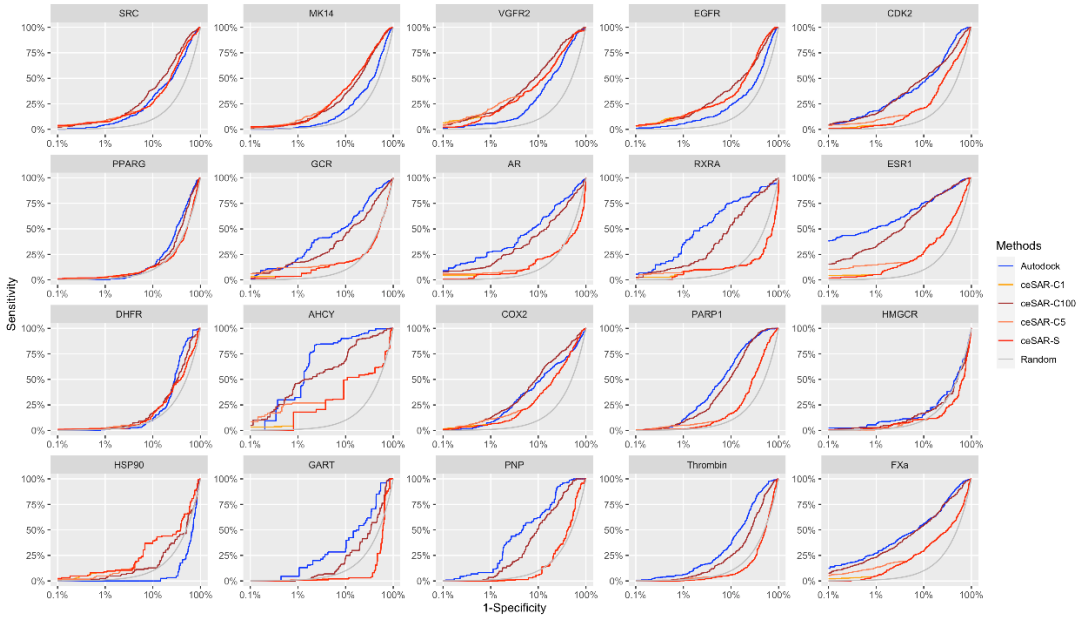
On the other hand, some classes of targets and their antagonists, including kinase inhibitors, are well represented in LINCS, contributing to the high accuracy of *ceSAR* on the 5 kinases included in our evaluation. For these kinases, *ceSAR*-S and all forms of consensus yield improvements over docking already at 1% library size, and for the most reduced library size even the simplest form of the method, i.e., *ceSAR*-S) yields 2-fold increase in median precision compared to *Autodock*, even though docking performs well on these targets achieving over 20% precision at 0.1% library size (see **Supplemental Figure 17**).

Similar trends are also observed in terms of ROC and precision-recall curves (**Supplemental Figure 18**). that demonstrate the success of *ceSAR*-S and consensus methods relative to docking on kinases (except for CDK2), generally better performance of consensus methods and docking relative to *ceSAR*-S on nuclear receptors (except for PPARG), and better performance of *Autodock* on GART, PNP, Thrombin, and HMGCR, on which *ceSAR* methods fail for reasons discussed before.

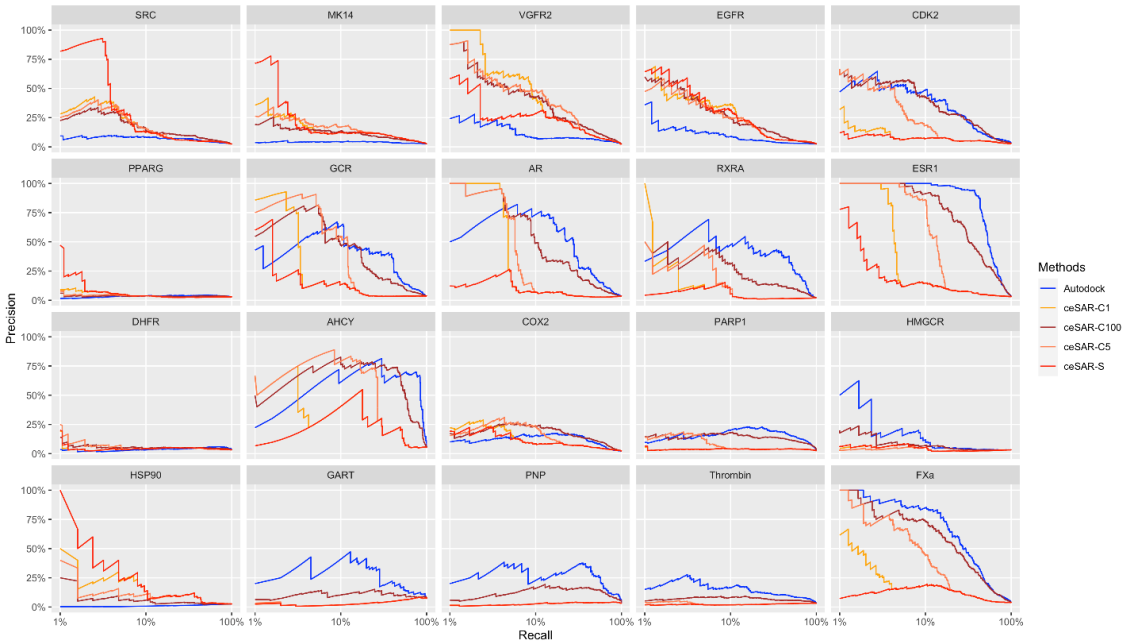
As emphasized before, for successful identification of true binders while nominating top candidates for validation, achieving high recall with a relatively large subset of the library is less important than higher precision at the low recall values corresponding to a small subset of the

library amenable to experimental validation. Note that to emphasize this point, the recall values and false positive rates (% decoys selected) are shown in log scale in **Supplemental Figure 18**.

A



B

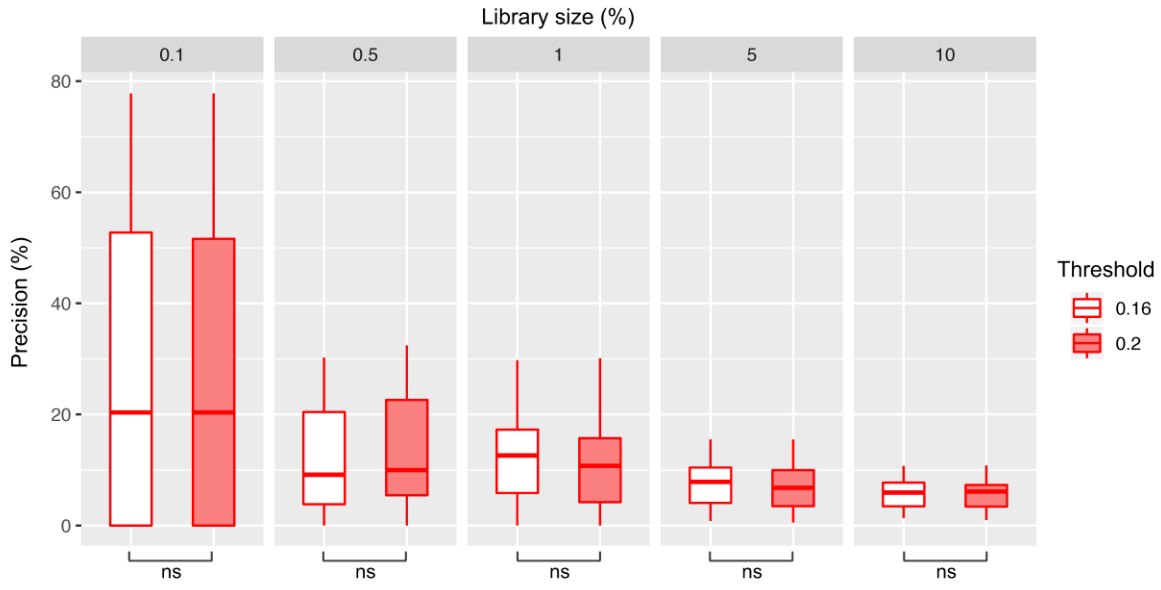


Supplemental Figure 18: The performance of *ceSAR*-S and consensus methods *ceSAR*-C relative to *Autodock*, using the ROC (panel A) and precision-recall curves (panel B) for 20 DUD-E targets.

The results for EGFR, which represents an important class of tyrosine kinase growth factor receptors, commonly targeted in cancer, are analyzed in detail in **Supplemental Table 6**. The compounds in top 0.1% of the DUD-E library for EGFR, ordered by using the consensus *ceSAR*-cML$_1$ method ranking, with true positives and true negatives (decoys) denoted using CHEMBL and ZINC identifiers, respectively. Note that 36 corresponds to the top 0.1% of the EGFR DUD-E library.

As can be seen from MoA annotations that are either obtained from LINCS for the closest analogs used for scoring, or directly from PubChem for the actual molecule in the DUD-E data set, essentially all DUD-E true positives are in fact annotated as EGFR inhibitors. There are 8 decoys predicted as ligands, for which there are no directly transferable annotations and thus are nominated here as potential inhibitors to be tested. Note that this is consistent with the precision of about 78% for the *ceSAR*-cML$_1$ method on the EGFR dataset at 0.1% library size reported in Figure 3.

All results of *ceSAR* reported in the main manuscript were obtained using the concordance threshold of 0.162 to identify significantly 'concordant' LINCS small molecules. This value corresponds to Bonferroni corrected p-value of 0.05 to adjust for multiple testing given the size of LINCS signature libraries. The sensitivity to this arbitrary threshold (even if guided by the analysis of distribution of concordance scores used by *iLINCS*) was assessed by systematically comparing results obtained using different thresholds, indicating that the results are robust. For example, the performance of *ceSAR* performance had no statistically significant differences between the threshold of 0.162 and 0.2, as shown by the box plots of precision values over 20 targets at different library sizes (**Supplemental Figure 19**).

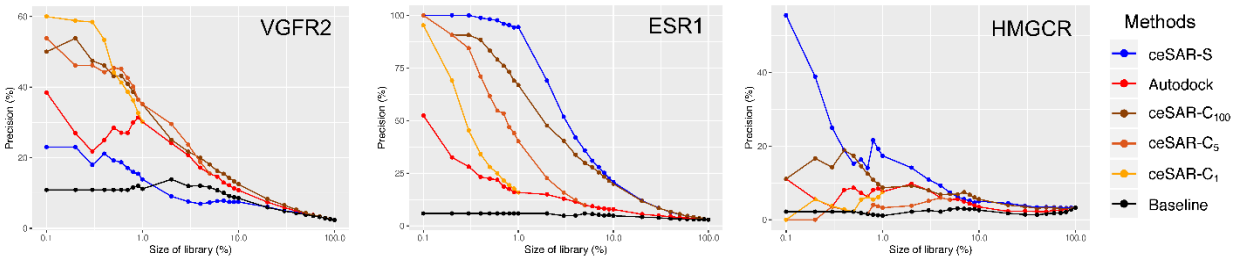


Supplemental Figure 19: **Comparing the performance of *ceSAR*-S over 20 targets when using the primary filter of concordance thresholds of 0.162 and 0.2.**

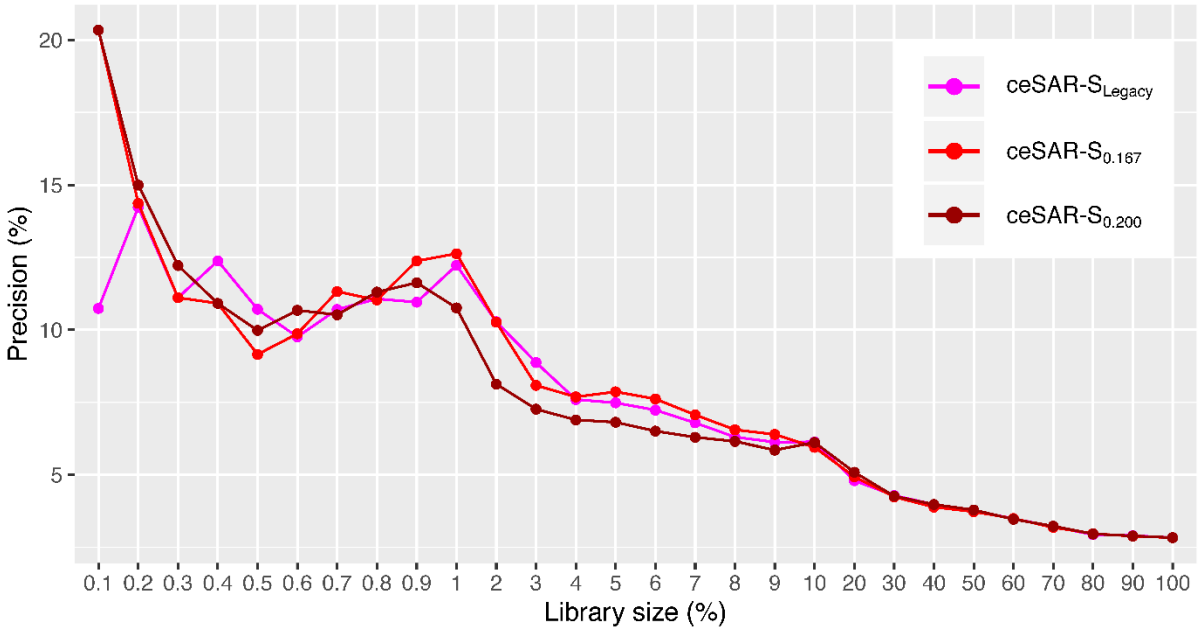
Supplemental Table 6. The top 36 candidate EGFR inhibitors were selected by using the machine-learning-based consensus *ceSAR*-cML$_1$ approach. DUD-E true positives (known active compounds) are indicated using green background, whereas DUD-E true negatives (decoys) are indicated using yellow background. Mode of Action (MoA) annotations from LINCS and PubChem are shown in the second and third columns, respectively.

Top Compounds	LINCS_MoA	PubChem_MoA
CHEMBL1421	KIT, SRC, BCR-ABL, Ephrin, PDGFR	Cell Growth Inhibition, ABL1, EPH, BCR, Lck, MAP4K4, FYN, SIK1, PKMYT1, FGFR2, FGFR1, MAP2K2, SRMS, MAP3K20, MAPK14, STK26, CDC42BPG, RET
ZINC03984008	NA	NA
CHEMBL1421.1	KIT, SRC, BCR-ABL, Ephrin, PDGFR	Cell Growth Inhibition, ABL1, EPH, BCR, Lck, MAP4K4, FYN, SIK1, PKMYT1, FGFR2, FGFR1, MAP2K2, SRMS, MAP3K20, MAPK14, STK26, CDC42BPG, RET
CHEMBL1173655.1	EGFR	EGFR, ERBB2
CHEMBL1173655	EGFR	EGFR, ERBB3
CHEMBL1421.2	KIT, SRC, BCR-ABL, Ephrin, PDGFR	Cell Growth Inhibition, ABL1, EPH, BCR, Lck, MAP4K4, FYN, SIK1, PKMYT1, FGFR2, FGFR1, MAP2K2, SRMS, MAP3K20, MAPK14, STK26, CDC42BPG, RET
ZINC04047855	NA	NA
CHEMBL56543	EGFR, ErbB2, JAK	EGFR, GLS, JAK1, HIF1A, GMNN, CBX1
CHEMBL554	EGFR, ErbB2	EGFR, ErbB2, Erk-1, Erk-2, AKT
ZINC04047854	NA	NA
CHEMBL1421.3	KIT, SRC, BCR-ABL, Ephrin, PDGFR	Cell Growth Inhibition, ABL1, EPH, BCR, Lck, MAP4K4, FYN, SIK1, PKMYT1, FGFR2, FGFR1, MAP2K2, SRMS, MAP3K20, MAPK14, STK26, CDC42BPG, RET
CHEMBL483321	Tyrosine Kinase, EGFR	CDK19, MKNK2, ERBB2, ERBB4, CDK8, EGFR
CHEMBL1421.4	KIT, SRC, BCR-ABL, Ephrin, PDGFR	Cell Growth Inhibition, ABL1, EPH, BCR, Lck, MAP4K4, FYN, SIK1, PKMYT1, FGFR2, FGFR1, MAP2K2, SRMS, MAP3K20, MAPK14, STK26, CDC42BPG, RET
CHEMBL939	EGFR	EGFR
CHEMBL939.1	EGFR	EGFR
CHEMBL180022	EGFR	HER-2, EGFR
CHEMBL180022.1	EGFR	HER-2, EGFR
CHEMBL572881	KIT, PDGFR, VEGFR	VEGFR, PDGFR
CHEMBL180227	NA	ERBB2, EGFR
CHEMBL337026	NA	PRKD3, PRKCG, PRKCB, EGFR, SRC
CHEMBL180227.1	NA	ERBB2, EGFR
CHEMBL337026.1	NA	PRKD3, PRKCG, PRKCB, EGFR, SRC
CHEMBL472581	NA	EGFR
ZINC03234670	NA	NA
CHEMBL472581.1	NA	EGFR
ZINC63627870	NA	NA
CHEMBL515082	NA	EGFR
CHEMBL515082.1	NA	EGFR
ZINC00631023	NA	NA
CHEMBL168915	NA	EGFR
ZINC05276009	NA	NA
ZINC01822324	NA	NA
CHEMBL1090357	NA	EGFR, ERBB2, ERBB3, ERBB4, IGF1R
CHEMBL179652	NA	ERBB2, EGFR
CHEMBL56142	NA	EGFR
CHEMBL179652.1	NA	ERBB2, EGFR
CHEMBL291514	NA	EGFR

To assess the dependence on the concordance threshold further, precision curves for three different targets with the concordance threshold of 0.2 are included in **Supplemental Figure 20**, showing very similar performance to that obtained using the threshold of 0.162 in **Supplemental Figure 15**. This result implies that the performance of *ceSAR* is largely insensitive to the changes of the concordance threshold, which is used to filter out compounds without signatures concordant to a target KD signature.



Supplemental Figure 20: **Examples of three representative targets using *ceSAR*-S, -C_{100}, -C_5 and -C_1 while applying the primary filter at the concordance threshold of 0.2 (as opposed to the default value of 0.162).** For comparison: *Autodock* (A) and a simple baseline method (B).



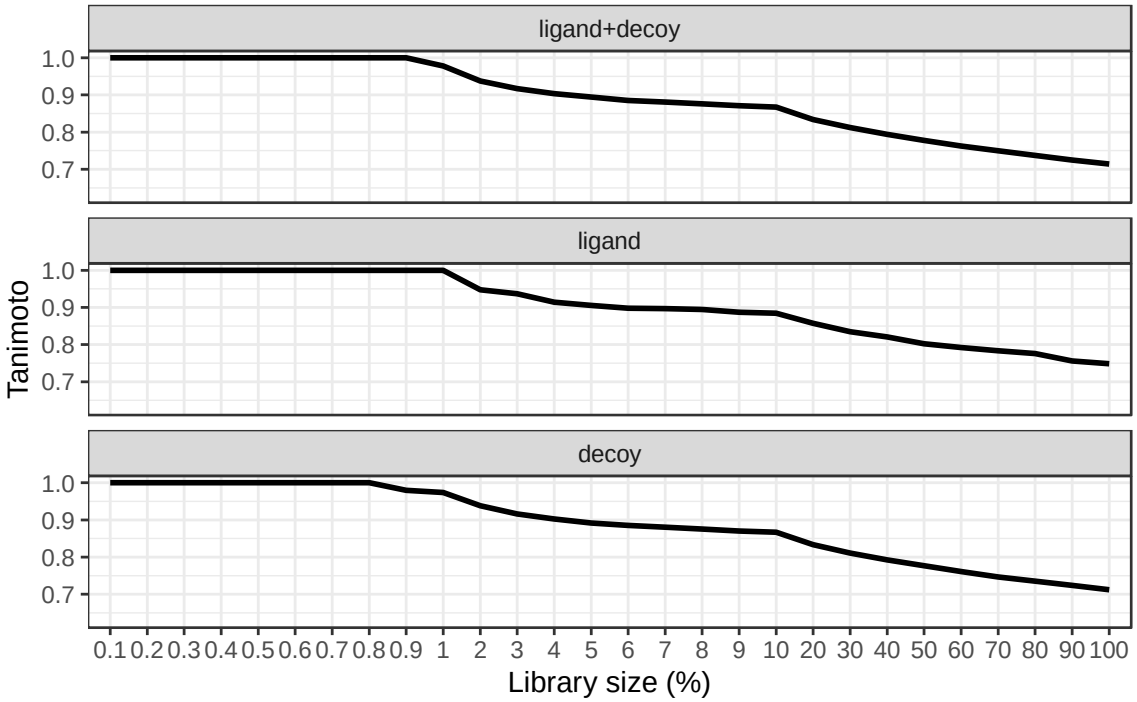
Supplemental Figure 21: **Comparing *ceSAR*-S median precision curves over 20 targets for legacy vs. new version of *iLINCS*, using the primary filter of concordance thresholds of 0.162 and 0.2.**

To further assess the robustness of the method, while evaluating the importance of careful signature library curation and the choice of concordance measures, we also compared the results with those obtained using the *iLINCS* 2018 legacy version. Precision curves at two different values of the concordance threshold for the new *iLINCS* library used for this contribution are compared with those of the legacy *iLINCS* version (*54*).

As can be seen from **Supplemental Figure 21**, while the precision curves are overall very similar, there is a substantial improvement in the median precision at the furthest library reduction. Thus, filtering out spurious concordances due to additional curation in the new version of *iLINCS*

library to remove low quality signatures leads to further reduction of false positive rates and higher chances of success when nominating leads for further validation. The results are very similar when using e.g., a threshold of 0.321, which corresponds to the highest 0.5% of all concordance values between chemical and genetic perturbation signatures in the legacy *iLINCS* version.

We next analyze potential biases due to uneven coverage of different classes of compounds in LINCS. *ceSAR*-S ranks candidate compounds based on the Jaccard similarity to the closest LINCS analog. In the case of ties, the compounds with the equal similarity are re-ranked by the concordance score. The degeneracy of similarity scores is becoming more pronounced as the library is reduced to a small subset, with all or nearly all candidate compounds included at that stage having direct counterparts in the LINCS library with Tanimoto coefficient of 1.



Supplemental Figure 22: The average Tanimoto coefficient at various library sizes over 20 targets. Upper, middle, and lower panels show the distribution of Tanimoto similarity for various library sizes of all compounds, only true ligands, and only decoys, respectively. The black line represents the median over 20 targets.

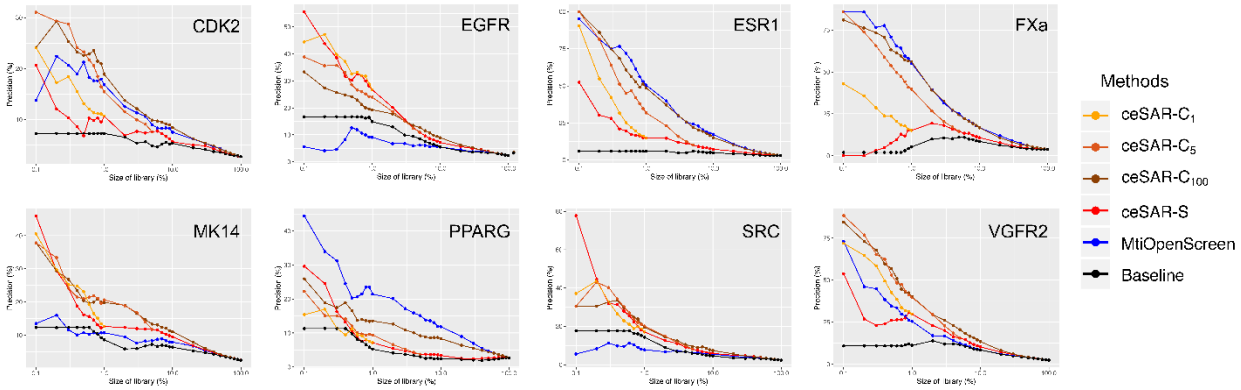
This could potentially lead to biases in the selection of compounds based on their inclusion in the LINCS library. As shown in **Supplemental Figure 22**, however, the distributions of Tanimoto coefficients over 20 targets are very similar for both active ligands and decoys at the small library sizes of 0.1%-1%, with average Tanimoto coefficients in the range of [0.95, 1]. Moreover, as shown in the main manuscript and **Supplemental Figures 12** through **17**, the baseline approach (B) that ignores signature connectivity performs consistently poorly and significantly worse than any other method discussed. Taken together with the results for compounds that have direct counterparts in LINCS shown in **Figure 8**, one can conclude that in fact the concordance filter is required to achieve enrichment into true positives.

Next, we evaluate the choice of the docking method to be combined into a consensus *ceSAR* approach. While we primarily used *Autodock* because of its well benchmarked performance and wide adoption in the field, there are certainly multiple other docking methods that could be used. In **Supplemental Table 7**, we compare *Autodock* and *Autodock*-based *ceSAR* consensus methods with two widely used docking programs, namely *DOCK 3.7* and *Autodock Vina*.

Supplemental Table 7: Comparison of *ceSAR* and docking methods on DUD-E benchmark. Results for individual targets and median values from this work for *Autodock* (A), *ceSAR*-S (S), *ceSAR*-C_{100} (C100), *ceSAR*-C_5 (C5) and ceSAR-C_1 (C1) are compared with those of *DOCK 3.7* (Dock), and *Autodock Vina* (Vina) from a recent benchmark of docking methods (*23*), using adjusted log AUC that captures overall performance in terms of area under the ROC curve relative to a random classifier (see Supplemental Figure 18), and Enrichment Factor at 1% and 0.1% library size, denoted as EF1 and EF.1, respectively.

Target	A AUC	S AUC	C100 AUC	Dock AUC	Vina AUC	A EF1	S EF1	C100 EF1	Dock EF1	Vina EF1	A EF.1	S EF.1	C100 EF.1	C5 EF.1	C1 EF.1
SRC	11.29	13.42	17.39	10.77	10.12	3.98	7.36	7.96	4.02	5.36	0	34.45	9.84	9.84	13.54
MK14	5.3	16.69	15.03	10.92	17.18	1.75	5.14	5.8	6.59	9.19	0	18.75	3.31	9.93	14.34
VGFR2	13.85	21.79	26.36	24.35	23.6	5.81	12.42	14.68	15.18	17.14	10.03	21.72	25.06	25.06	26.74
EGFR	8.35	17.94	18.95	33.82	8.81	5.42	11.68	10.24	27.17	5.36	12.11	24.22	24.22	20.59	18.17
CDK2	24.5	8.02	23.85	15.52	17.56	14.67	3.89	12.92	4.85	10.76	21.39	7.55	20.13	21.39	11.32
PPARG	4.63	1.29	3.51	9.19	19.35	0	2.64	1.11	4.97	6.62	0	11.32	2.83	2.83	7.07
GCR	27.76	3.55	21.13	-2.44	16.94	12.65	3.56	13	2.72	13.61	13.05	16.78	18.65	24.24	24.24
AR	29.41	1.64	21.64	2.78	21.78	17.62	5.17	11.68	4.84	21.58	17.24	5.75	26.81	26.81	28.73
RXRA	37.55	-1.7	23.74	1.39	38.08	19.93	4.98	11.83	5.39	32.36	20.82	0	20.82	13.88	13.88
ESR1	50.05	10.22	39.73	10.41	28.89	32.33	5.12	23.53	8.9	19.36	34.26	17.94	34.26	34.26	32.62
DHFR	5.47	3.97	6.05	18.4	16.48	1.06	1.24	2.13	19.05	5.2	0	9.33	1.87	5.6	5.6
AHCY	48.17	14.63	41.59	43.82	34.78	12.89	4.3	14.5	36.64	23.9	0	0	6.44	12.89	12.89
COX2	18.59	13.09	22.64	11.66	27.43	7.35	5.85	10.94	12.68	19.83	1.96	7.84	5.88	7.84	7.84
PARP1	26.98	6.52	23.3	20.57	24.96	6.61	1.08	7.15	16.56	12.42	2.71	1.35	5.42	6.77	5.42
HMGCR	4.44	-2.06	2.33	28.78	15.74	5.4	2.02	2.7	25.92	5.3	17.06	3.41	3.41	0	3.41
HSP90	-7.13	12.37	3.62	16.22	-9.5	0	8.11	2.43	10.3	0	0	24.32	8.11	16.21	16.21
GART	16.1	-5.62	5.85	10.21	32.85	4.52	0.5	0.5	2.03	6.1	0	0	0	0	0
PNP	29.64	0.98	20.01	46.38	28.46	8.21	0	3.45	44.02	13.7	0	0	0	0	0
Thrombin	17.54	-1.84	9.69	20.1	16.74	5.47	0.23	2.21	15.62	4.2	4.68	0	2.34	0	0
FXa	31.87	9.75	28.94	37.14	28.68	18.32	4.04	15.92	27.27	20.54	24.23	0	24.23	19.13	14.03
Median	**18.06**	**7.27**	**20.57**	**15.87**	**20.56**	**6.21**	**4.17**	**9.1**	**11.49**	**11.59**	**3.69**	**7.69**	**7.27**	**11.41**	**13.21**

Results from a recent benchmark of these docking methods (*23*), indicate that *Autodock Vina* performs somewhat better than *Autodock* in terms of adjusted logAUC, with median values of about 20 vs. 18. On the other hand, while the results at 0.1% library reduction are not available, the Enrichment Factor at 1% (EF1) is substantially higher for both *DOCK* and *Autodock Vina* relative to *Autodock* on the set of targets used here. Based on these observations, *Autodock Vina* could potentially provide additional boost in performance if used instead of *Autodock* for *ceSAR* consensus methods, which is investigated next.



Supplemental Figure 23: Precision curves on a subset of targets from the DUD-E benchmark using *MTiOpenScreen* docking server.

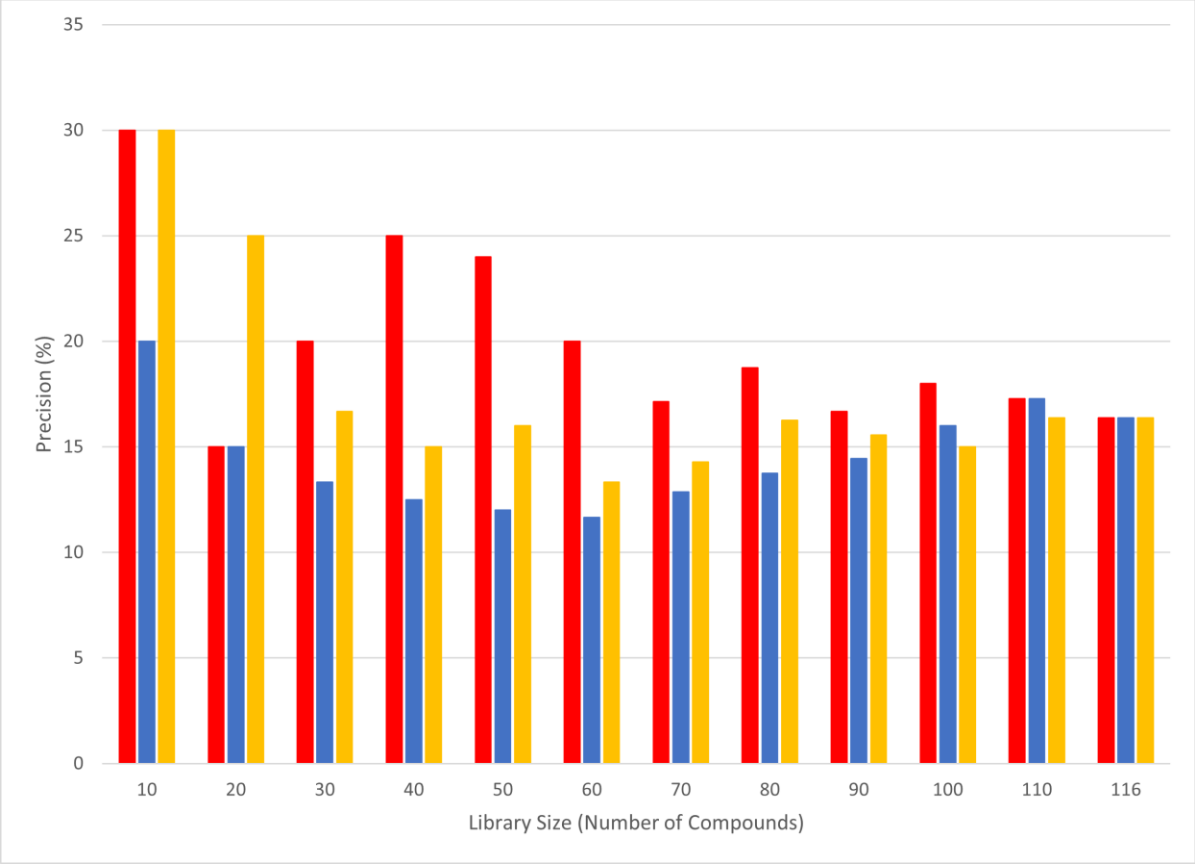
To test the impact of using *Autodock Vina*, we investigated the performance of *ce*SAR in combination with *MTiOpenScreen*, the *Autodock Vina*-based web server for docking and virtual screening (**Supplemental Figure 23**). *ceSAR*-S, -C_{100}, -C_5 or -C_1 yielded better performance in five out of eight targets screened, which were kinases (CDK2, EGFR, MK14, SRC, and VGFR2). *MTiOpenScreen* provided better results on FXa and two nuclear receptors, ESR1 and PPARG, on the latter also notably better than *Autodock*, although *ceSAR* achieved a relatively high enrichment over baseline in the latter two cases as well. Overall, the results are qualitatively similar to those obtained using *Autodock*, indicating that *ceSAR* consensus rescoring of a small library subset obtained after applying the initial *ceSAR*-S filter is insensitive to the choice of the docking method used at this stage.

Identification, *ceSAR* rescoring and Validation of Putative BCL2A1 Inhibitors (S7)

BCL2A1 has been implicated in a wide array of diseases, ranging from autoimmunity resulting in pre-term birth to chemotherapeutic resistance in cancer. The BCL2-family of proteins interact with BH3 peptides to produce either a pro-apoptotic or anti-apoptotic response. Due to the similarity across the family, this has been a difficult target with which to achieve specificity. To that end, an inhibitor virtual screen was performed using an NCI compound library of 90,087 drug-like small molecules and *Autodock* v. 4.2.6.

The top 300 compounds by *Autodock* were clustered and representatives of each cluster were tested *in vitro* using differential scanning fluorimetry thermal shift and fluorescence polarization competition assays described in detail below. Compounds with substantial reduction in polarization through high dose experiments were additionally tested for a dose response via FP. High dose FP was normalized using unbound Noxa as the 0% bound baseline and A1 with Noxa but no compound as the 100% value.

ceSAR-S was applied retroactively to re-score the tested compounds and test the hypothesis that a set of putative weak binders identified experimentally (guided by the initial virtual screening) can be successfully used to seed *ceSAR* search, enrich the top ranked compounds by using *ceSAR*-S rescoring in conjunction with docking, and thus reduce the number of compounds for further validation, as shown in **Supplemental Figure 24** and discussed in the main manuscript.



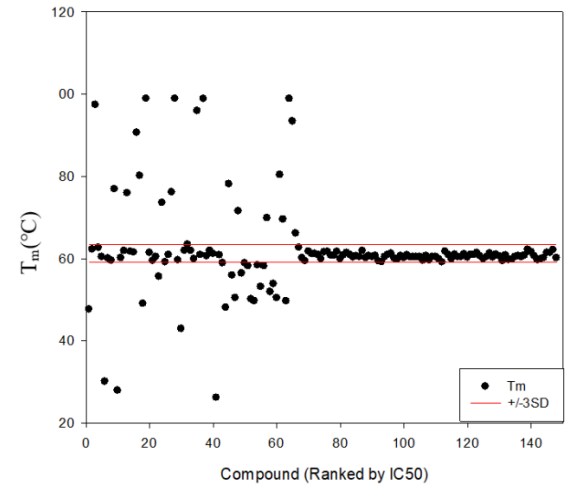
Supplemental Figure 24: *ceSAR* re-scoring of candidate BCLA1 inhibitors identified using a drug-discovery pipeline that combined docking simulation using *Autodock, SAR* analysis and multiple experimental validation methods. Note that *ceSAR*-S (Red) and consensus *ceSAR*-C (Yellow), which combines *Autodock* with *ceSAR*-S using geometric mean of individual ranks, yield a higher enrichment into experimentally validated inhibitors relative to *Autodock* alone (Blue).

In this section, experimental validation of such identified candidate molecules is discussed further and summarized in **Supplemental Tables 8 and 9**, as well as **Supplemental Figures 25 and 26,** with detailed results at each stage of validation included below.

Supplemental Table 8. Summary of differential scanning fluorimetry and fluorescence polarization results of top tested compounds. Compounds highlighted in blue have dose responses shown in **Supplemental Figure 26.**

ZINC ID	Average Tm Shift (℃)	High Dose FP %Bound Noxa
ZINC263599489	1.92	0
ZINC04974300	-30.75	1.2
ZINC05930871	38.82	4.5
ZINC04822049	0.17	23.84
ZINC04803984	-1.5	36.11
NSC114566	0.33	36.67
ZINC01717014	38.83	37.77
ZINC01650974	1.37	39.73
ZINC03954010	14.75	39.74
ZINC04934733	33.32	40.1
ZINC05840007	0.49	41.48
ZINC03194765	-1.42	42.04
ZINC04896374	11.41	42.72
ZINC01600320	-2.11	43.15
ZINC04804129	29.16	43.51
ZINC04217277	-33.6	45.98
ZINC04523161	-0.59	46.69
ZINC05839997	8.49	48.34
ZINC04366919	-1.43	51.13
ZINC04758328	11.62	56.8
ZINC03953829	1.16	60.6
ZINC04804154	35.86	62.06
ZINC01673413	-12.92	62.06
ZINC01872824	0.24	62.31
ZINC01569504	-0.89	66.56
ZINC04803990	0.43	68.17
ZINC01593456	19.25	70.39
ZINC03953823	37.41	70.74
ZINC01040738	0.33	70.86
ZINC01592019	19.82	73.95
ZINC16968863	-0.41	75.56
ZINC03953869	-2.26	77.7
ZINC16957598	-0.38	77.83
ZINC04212058	-0.85	100.64

Thermal Shift Assay. Compounds were first tested for impacting thermal stability of BCL2A1 as a test for binding. 100 µM compounds were applied to purified A1 at 4.4 µM in triplicate. Sypro Orange dye was added at a final dilution of 1:1000 to protein- and compound- containing wells. An Applied Biosystems StepOnePlus was used to perform Differential Scanning Fluorimetry (DSF). DSF was performed with 100 µM of each compound for binding or Noxa binding inhibition. By DSF, all compounds that drove an increase in Tm that corresponded to at least three standard deviations from the mean were included in further assays (see **Supplemental Figure 25**).

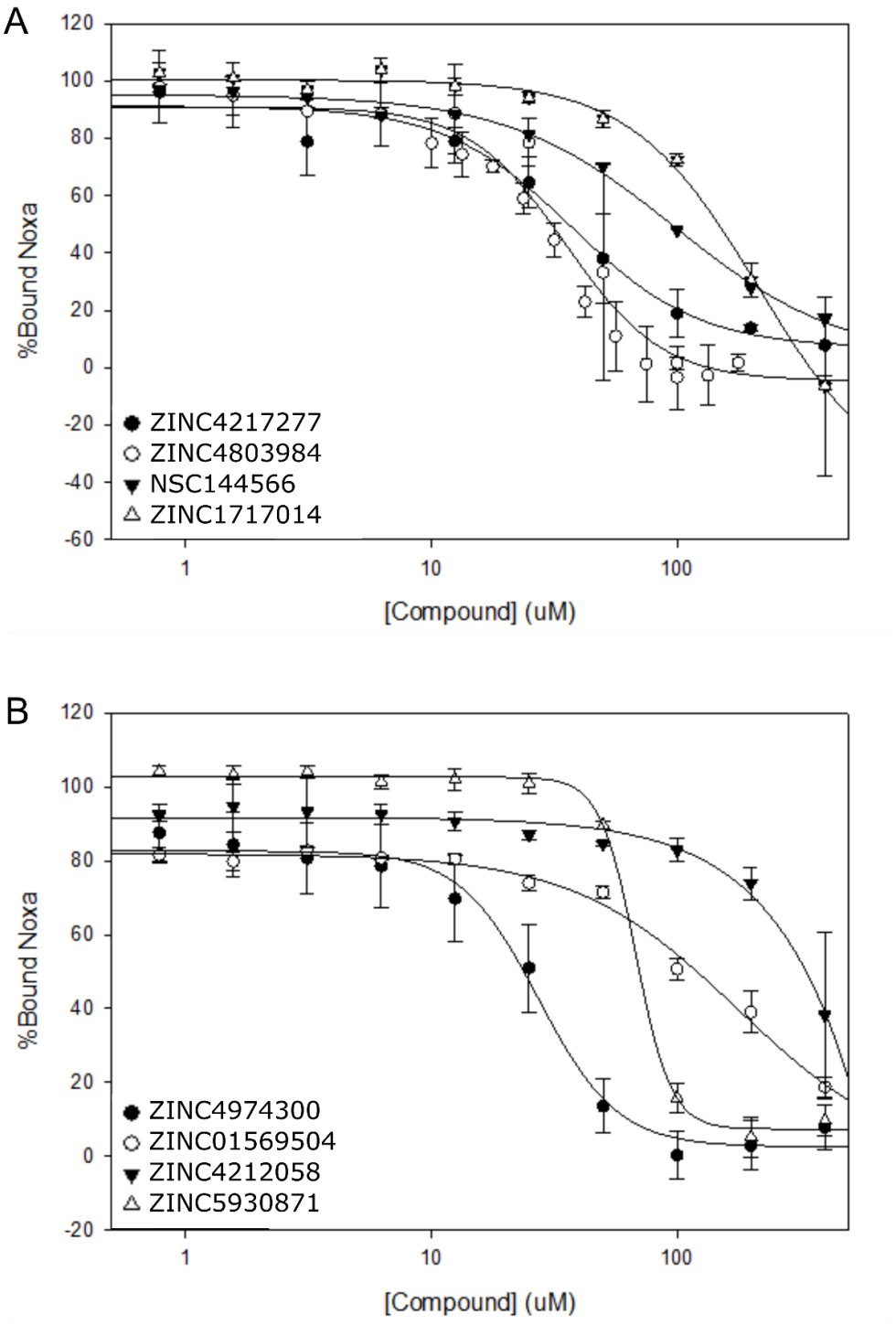


Supplemental Figure 25: High Dose DSF experiments were performed to further reduce the overall number of compounds to be tested.

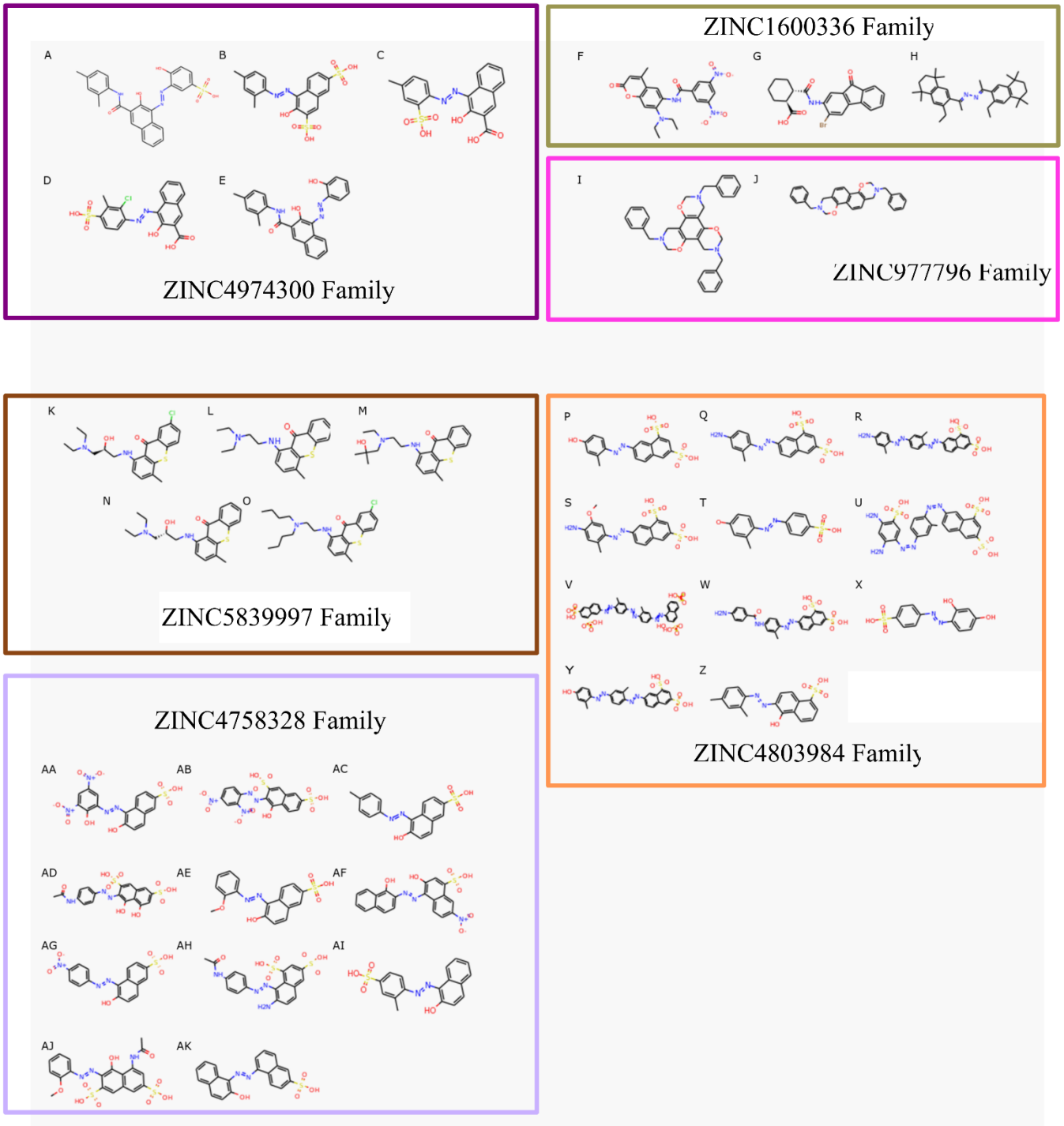
Fluorescence Polarization Assay. Compounds were additionally tested for specificity of binding by displacement of FITC-labeled mouse Noxa (mNoxa) (Peptide2.0) peptide via Fluorescence Polarization (FP). FP was performed as a two-step process, first identifying any compounds with polarization shifts at a single, high-dose, followed by a dose-response of any compounds that showed a substantial shift in polarization, or shift in thermal stability had a dose response measured via FP (**Supplementary Figure 26**).

Dose response curves. Compounds that displayed decreased polarization or increased Tm values at least three standard deviations from the control were tested in a dose response FP experiment and their IC50s were determined. Dose response curves were measured by adding 3 µM BCL2A1 to a serial two-fold dilution series of each compound ranging from 400 µM to 781 nM in 20 mM Tris pH 7, 500 mM NaCl, 0.005% Tween-20 buffer. All dose responses were performed in triplicate.

Cell Death Assay. Compounds that demonstrated effective inhibition of BCL2A1-Noxa binding through FP were further tested in an activated T-cell death assay. Single cell suspensions from spleen were generated by maceration through a 100 µm nylon mesh followed by LympholyteM ficoll gradient separation (Cedarlane Labs, Burlington, NC). Purified cells were then cultured on anti-CD3 coated (3 µg/mL, coated overnight, Biolegend, San Diego, CA) six-well plates in the presence of soluble anti-CD28 (2 µg/mL, Bio X Cell, West Lebanon, NH) and IL-2 (10 ng/mL, R&D Systems, Inc., Minneapolis, MN) in RPMI media (Life Technologies, Carlsbad, CA) for 24 hours at 37°C.



Supplemental Figure 26: Fluorescence Polarization-based identification of compounds that inhibit Noxa binding. Representatives of the tested compounds with fitted inhibition curves are displayed in groups of 4 in panels A and B, respectively.



Supplemental Figure 27: Chemical structures of compounds structurally related to initial hits grouped by compound family.

Supplemental Table 9. List of compounds included in SAR-like study for identification of important functional groups on lead compounds. The compound column corresponds to the structures displayed in **Supplemental Figure 27** above. Bolded compound names are initial hits with their analogs listed below. IC_{50} from FP dose response curves and K_i values are displayed.

Compound	FP IC50 (μM)	K_i (μM)
A ZINC04974300	27.3	**1.38**
B ZINC03861456	>400	>20.3
C ZINC04409853	387	19.6
D ZINC03953869	211	10.7
E ZINC100080130	13.8	0.70
F ZINC01600336	397	20.1
G ZINC01678715	>400	>20.3
H ZINC08642221	>400	>20.3
I ZINC00977796	389.5	19.7
J ZINC90751664	>400	>20.3
K ZINC05839997	>400	>20.3
L ZINC03831012	299.6	15.2
M ZINC04212058	333.6	16.9
N ZINC05840007	49.1	2.49
O ZINC04366916	>400	>20.3
P ZINC04803984	31.8	1.61
Q ZINC04707580	>400	>20.3
R ZINC04822049	15.8	0.80
S ZINC04523128	234.9	11.9
T ZINC04804120	>400	>20.3
U ZINC150367617	234.9	11.9
V ZINC263599489	19.4	0.98
W ZINC04523161	104.5	5.29
X ZINC03882219	79.8	4.04
Y ZINC04896374	100	5.06
Z ZINC04804151	>400	>20.3
AA ZINC04758328	68.9	3.49
AB ZINC04217277	45.8	2.32
AC ZINC03953829	119.6	6.06
AD ZINC04803990	>400	>20.3
AE ZINC04804142	>400	>20.3
AF ZINC03953823	320	16.2
AG ZINC04804129	59.2	3.00
AH ZINC04523127	>400	>20.3
AI ZINC04528652	55	2.78
AJ ZINC04705990	189.3	9.58
AK ZINC03954010	32	1.62

Cells were then washed and cultured again in IL-2 (10 ng/mL) for 24 hours at 37°C. Cells were harvested and cultured for 24 hours on anti-CD3 coated 96-well plates at 500,000 cells per well with 2 µg/mL soluble anti-CD28, 10 ng/mL IL-2, 0.125 µg purified anti-mouse FasL (Biolegend, San Diego, CA), and varying concentrations of BCL2A1 inhibitor compounds +/- polybrene (2 µg/mL, EMD Millipore, Burlington, MA). Cells were then harvested, and live and dead cells enumerated by trypan blue staining using the TC20 automated cell counter (Bio-Rad Laboratories, Des Plaines, IL).

Putative inhibitors identified in the initial screening by FP and DSF were expanded to include close structural analogs that may increase effectiveness of this inhibition, see **Supplemental Figure 27**, **Supplemental Table 9**. Analogs to initial hits from early screens were ordered for further testing. Some of these compounds show close similarity in structure, but deviate in IC50 value, illustrating the importance of distinct functional groups. ZINC4974300 Family (purple) and ZINC593997 Family (brown) represent the structural analogs from **Figure 7** of the manuscript. ZINC5840007 was initially an analog of a molecule that showed initial promise through thermal shift and differs only by a single -Cl in the chemical structure.

Sig2Lead workflows and implementation details (S8)

The transcriptional signature concordance-based and target structure-independent *ceSAR*-S search has been implemented as an R Shiny package, dubbed *Sig2Lead*, which is available as a docker container and can be also accessed as a web server at http://sig2lead.net. *Sig2Lead* workflow starts with a target of interest, specified by the user by its gene name. The *iLINCS* web server and its APIs (*54*) are used by *Sig2Lead* for retrieval of chemical perturbagens that have a similar (positively correlated) or dissimilar (negatively correlated) signatures with those of a targeted gene knockdown.

As an alternative, a user provided loss of function signature for a target of interest, which can be readily generated using standard pipelines for the identification of differentially expressed genes, can be used. The search against LINCS KD or user provided loss of function signature generates an initial list of candidate compounds that comprises ‘concordant’ LINCS drug-like molecules. These LINCS molecules have concordant signatures to the genetic knockdowns as a set of potential inhibitors of the target pathway.

Sig2Lead can also be run with an arbitrary (non-LINCS) library of user defined compounds to be ranked, supplied as canonical SMILES or SDF. User defined libraries of candidate molecules are screened out based on similarity to concordant LINCS molecules. APFP fingerprint library (*45*, *49*) of ChemmineR (*53*, *59*, *60*) is used to convert small molecules in SMILES or SDF format to binary fingerprints. Chemical similarity is measured using the Tanimoto coefficient (Jaccard similarity) which is computed using the *minSim* algorithm optimized for sparse binary fingerprints (see Section S2).

Sig2Lead can also perform hierarchical clustering of candidate molecules by chemical similarity to determine clusters of chemically similar compounds that also have substantial concordance. These data are depicted as a heatmap of the dissimilarity matrix, along with an MDS plot with pie charts demonstrating the contribution by user provided vs. LINCS candidates. These clusters can be further submitted to a STITCH network analysis pipeline that will identify literature reported associations with any compounds identified via *Sig2Lead*. Users can download top ranking molecules in the form of canonical SMILES or SDF for further analyses, including docking simulations or experimental validation.

Sig2Lead is a public domain package available at https://github.com/sig2lead, including a set of scripts to combine ranking by docking with those by *ceSAR*-S using machine learning-based consensus with an ensemble of classifiers developed as a part of this contribution. *Sig2Lead* is also available as a web server at http://sig2lead.net.

REFERENCES AND NOTES

1. K. P. Seiler, G. A. George, M. P. Happ, N. E. Bodycombe, H. A. Carrinski, S. Norton, S. Brudz, J. P. Sullivan, J. Muhlich, M. Serrano, P. Ferraiolo, N. J. Tolliday, S. L. Schreiber, P. A. Clemons, ChemBank: A small-molecule screening and cheminformatics resource database. *Nucleic Acids Res.* **36**, D351–359 (2007).

2. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–1107 (2012).

3. J. Eder, R. Sedrani, C. Wiesmann, The discovery of first-in-class drugs: Origins and evolution. *Nat. Rev. Drug Discov.* **13**, 577–587 (2014).

4. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).

5. R. H. Shoemaker, A. Monks, M. C. Alley, D. A. Scudiero, D. L. Fine, T. L. McLemore, B. J. Abbott, K. D. Paull, J. G. Mayo, M. R. Boyd, Development of human tumor cell line panels for use in disease-oriented drug screening. *Prog. Clin. Biol. Res.* **276**, 265–286 (1988).

6. J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, L. A. Garraway, The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

7. W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, S. Ramaswamy, P. A. Futreal, D. A. Haber, M. R. Stratton, C. Benes, U. McDermott, M. J. Garnett, Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–961 (2012).

8. N. Krishnamurthy, A. A. Grimshaw, S. A. Axson, S. H. Choe, J. E. Miller, Drug repurposing: A systematic review on root causes, barriers and facilitators. *BMC Health Serv. Res.* **22**, 970 (2022).

9. P. M. Haverty, E. Lin, J. Tan, Y. Yu, B. Lam, S. Lianoglou, R. M. Neve, S. Martin, J. Settleman, R. L. Yauch, R. Bourgon, Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).

10. B. Karaman, W. Sippl, Computational drug repurposing: Current trends. *Curr. Med. Chem.* **26**, 5389–5409 (2019).

11. N. El-Hachem, W. Ba-Alawi, I. Smith, A. S. Mer, B. Haibe-Kains, Integrative cancer pharmacogenomics to establish drug mechanism of action: Drug repurposing. *Pharmacogenomics* **18**, 1469–1472 (2017).

12. A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong, T. R. Golub, A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017).

13. A. B. Keenan, S. L. Jenkins, K. M. Jagodnik, S. Koplev, E. He, D. Torre, Z. Wang, A. B. Dohlman, M. C. Silverstein, A. Lachmann, M. V. Kuleshov, A. Ma'ayan, V. Stathias, R. Terryn, D. Cooper, M. Forlin, A. Koleti, D. Vidovic, C. Chung, S. C. Schürer, J. Vasiliauskas, M. Pilarczyk, B. Shamsaei, M. Fazel, Y. Ren, W. Niu, N. A. Clark, S. White, N. Mahi, L. Zhang, M. Kouril, J. F. Reichard, S.

Sivaganesan, M. Medvedovic, J. Meller, R. J. Koch, M. R. Birtwistle, R. Iyengar, E. A. Sobie, E. U. Azeloglu, J. Kaye, J. Osterloh, K. Haston, J. Kalra, S. Finkbiener, J. Li, P. Milani, M. Adam, R. Escalante-Chong, K. Sachs, A. Lenail, D. Ramamoorthy, E. Fraenkel, G. Daigle, U. Hussain, A. Coye, J. Rothstein, D. Sareen, L. Ornelas, M. Banuelos, B. Mandefro, R. Ho, C. N. Svendsen, R. G. Lim, J. Stocksdale, M. S. Casale, T. G. Thompson, J. Wu, L. M. Thompson, V. Dardov, V. Venkatraman, A. Matlock, J. E. Van Eyk, J. D. Jaffe, M. Papanastasiou, A. Subramanian, T. R. Golub, S. D. Erickson, M. Fallahi-Sichani, M. Hafner, N. S. Gray, J.-R. Lin, C. E. Mills, J. L. Muhlich, M. Niepel, C. E. Shamu, E. H. Williams, D. Wrobel, P. K. Sorger, L. M. Heiser, J. W. Gray, J. E. Korkola, G. B. Mills, M. LaBarge, H. S. Feiler, M. A. Dane, E. Bucher, M. Nederlof, D. Sudar, S. Gross, D. F. Kilburn, R. Smith, K. Devlin, R. Margolis, L. Derr, A. Lee, A. Pillai, The library of integrated network-based cellular signatures NIH program: System-level cataloging of human cells response to perturbations. *Cell Syst.* **6**, 13–24 (2018).

14. B. S. Glicksberg, L. Li, R. Chen, J. Dudley, B. Chen, Leveraging big data to transform drug discovery. *Methods Mol. Biol.* **1939**, 91–118 (2019).

15. A. Musa, S. Tripathi, M. Dehmer, O. Yli-Harja, S. A. Kauffman, F. Emmert-Streib, Systems pharmacogenomic landscape of drug similarities from LINCS data: Drug association networks. *Sci. Rep.* **9**, 7849 (2019).

16. D. Vidović, A. Koleti, S. C. Schürer, Large-scale integration of small molecule-induced genome-wide transcriptional responses, kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action. *Front. Genet.* **5**, 10.3389/fgene.2014.00342 (2014).

17. E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, A. G. Taranto, Structure-based virtual screening: From classical to artificial intelligence. *Front. Chem.* **8**, 343 (2020).

18. S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell, A. J. Olson, Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).

19. C. M. Labbé, J. Rey, D. Lagorce, M. Vavruša, J. Becot, O. Sperandio, B. O. Villoutreix, P. Tufféry, M. A. Miteva, MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Res.* **43**, W448–W454 (2015).

20. M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).

21. D. Bajusz, A. Rácz, K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Chem.* **7**, 20 (2015).

22. M. Pilarczyk, M. Fazel-Najafabadi, M. Kouril, B. Shamsaei, J. Vasiliauskas, W. Niu, N. Mahi, L. Zhang, N. A. Clark, Y. Ren, S. White, R. Karim, H. Xu, J. Biesiada, M. F. Bennett, S. E. Davidson, J. F. Reichard, K. Roberts, V. Stathias, A. Koleti, D. Vidovic, D. J. B. Clarke, S. C. Schürer, A. Ma'ayan, J. Meller, M. Medvedovic, Connecting omics signatures and revealing biological mechanisms with iLINCS. *Nat. Commun.* **13**, 4678 (2022).

23. M. Xu, C. Shen, J. Yang, Q. Wang, N. Huang, Systematic investigation of docking failures in large-scale structure-based virtual screening. *ACS Omega* **7**, 39417–39428 (2022).

24. P. Presicce, C.-W. Park, P. Senthamaraikannan, S. Bhattacharyya, C. Jackson, F. Kong, C. M. Rueda, E. DeFranco, L. A. Miller, D. A. Hildeman, N. Salomonis, C. A. Chougnet, A. H. Jobe, S. G. Kallapur, IL-1 signaling mediates intrauterine inflammation and chorio-decidua neutrophil recruitment and activation. *JCI Insight* **3**, **e98306** (2018).

25. C. K. Hind, M. J. Carter, C. L. Harris, H. T. C. Chan, S. James, M. S. Cragg, Role of the pro-survival molecule Bfl-1 in melanoma. *Int. J. Biochem. Cell Biol.* **59**, 94–102 (2015).

26. K.-P. Li, S. Shanmuganad, K. Carroll, J. D. Katz, M. B. Jordan, D. A. Hildeman, Dying to protect: Cell death and the control of T cell homeostasis. *Immunol. Rev.* **277**, 21–43 (2017).

27. T. Oltersdorf, S. W. Elmore, A. R. Shoemaker, R. C. Armstrong, D. J. Augeri, B. A. Belli, M. Bruncko, T. L. Deckwerth, J. Dinges, P. J. Hajduk, M. K. Joseph, S. Kitada, S. J. Korsmeyer, A. R. Kunzer, A. Letai, C. Li, M. J. Mitten, D. G. Nettesheim, S. Ng, P. M. Nimmer, J. M. O'Connor, A. Oleksijew, A. M. Petros, J. C. Reed, W. Shen, S. K. Tahir, C. B. Thompson, K. J. Tomaselli, B. Wang, M. D. Wendt,

H. Zhang, S. W. Fesik, S. H. Rosenberg, An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **435**, 677–681 (2005).

28. W. H. Wilson, O. A. O'Connor, M. S. Czuczman, A. S. LaCasce, J. F. Gerecitano, J. P. Leonard, A. Tulpule, K. Dunleavy, H. Xiong, Y.-L. Chiu, Y. Cui, T. Busman, S. W. Elmore, S. H. Rosenberg, A. P. Krivoshik, S. H. Enschede, R. A. Humerickhouse, Navitoclax, a targeted high-affinity inhibitor of BCL-2, in lymphoid malignancies: A phase 1 dose-escalation study of safety, pharmacokinetics, pharmacodynamics, and antitumour activity. *Lancet Oncol.* **11**, 1149–1159 (2010).

29. A. J. Souers, J. D. Leverson, E. R. Boghaert, S. L. Ackler, N. D. Catron, J. Chen, B. D. Dayton, H. Ding, S. H. Enschede, W. J. Fairbrother, D. C. S. Huang, S. G. Hymowitz, S. Jin, S. L. Khaw, P. J. Kovar, L. T. Lam, J. Lee, H. L. Maecker, K. C. Marsh, K. D. Mason, M. J. Mitten, P. M. Nimmer, A. Oleksijew, C. H. Park, C.-M. Park, D. C. Phillips, A. W. Roberts, D. Sampath, J. F. Seymour, M. L. Smith, G. M. Sullivan, S. K. Tahir, C. Tse, M. D. Wendt, Y. Xiao, J. C. Xue, H. Zhang, R. A. Humerickhouse, S. H. Rosenberg, S. W. Elmore, ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **19**, 202–208 (2013).

30. H. Sasaki, T. Hirose, T. Oura, R. Otsuka, I. Rosales, D. Ma, G. Lassiter, A. Karadagi, T. Tomosugi, A. Dehnadi, M. Matsunami, S. Raju Paul, P. M. Reeves, I. Hanekamp, S. Schwartz, R. B. Colvin, H. Lee, T. R. Spitzer, A. B. Cosimi, P. E. Cippà, T. Fehr, T. Kawai, Selective Bcl-2 inhibition promotes hematopoietic chimerism and allograft tolerance without myelosuppression in nonhuman primates. *Sci. Transl. Med.***15**, eadd5318 (2023).

31. H.-W. Lee, S.-J. Park, B. K. Choi, H. H. Kim, K.-O. Nam, B. S. Kwon, 4-1BB promotes the survival of $CD8^+$ T lymphocytes by increasing expression of Bcl-x_L and Bfl-1. *J. Immunol.* **169**, 4882–4888 (2002).

32. P. Tripathi, B. Koss, J. T. Opferman, D. A. Hildeman, Mcl-1 antagonizes Bax/Bak to promote effector $CD4^+$ and $CD8^+$ T-cell responses. *Cell Death Differ.* **20**, 998–1007 (2013).

33. T. Cheng, Q. Li, Y. Wang, S. H. Bryant, Identifying compound-target associations by combining bioactivity profile similarity search and public databases mining. *J. Chem. Inf. Model.* **51**, 2440–2448 (2011).

34. H. Matthews, J. Hanison, N. Nirmalan, “Omics”-informed drug and biomarker discovery: Opportunities, challenges and future perspectives. *Proteomes* **4**, 28 (2016).

35. K. B. Dar, A. Bhat, S. Amin, R. Hamid, S. Anees, S. Anjum, B. A. Reshi, M. A. Zargar, A. Masood, S. A. Ganie, Modern computational strategies for designing drugs to curb human diseases: A prospect. *Curr. Top. Med. Chem.* **18**, 2702–2719 (2018).

36. D. A. Rachkovskij, Index structures for fast similarity search for binary vectors. *Cybern. Syst. Anal.* **53**, 799–820 (2017).

37. J. Wang, H. T. Shen, J. Song, J. Ji, Hashing for similarity search: A survey. arXiv:1408.2927 [cs.DS] (13 August 2014).

38. M. González-Medina, J. Jesús Naveja, N. Sánchez-Cruz, J. L. Medina-Franco, Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Adv.* **7**, 54153–54163 (2017).

39. J. C. Pereira, E. R. Caffarena, C. N. Dos Santos, Boosting docking-based virtual screening with deep learning. *J. Chem. Inf. Model.* **56**, 2495–2506 (2016).

40. F. Gentile, J. C. Yaacoub, J. Gleave, M. Fernandez, A.-T. Ton, F. Ban, A. Stern, A. Cherkasov, Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **17**, 672–697 (2022).

41. F. Gentile, V. Agrawal, M. Hsing, A.-T. Ton, F. Ban, U. Norinder, M. E. Gleave, A. Cherkasov, Deep docking: A deep learning platform for augmentation of structure based drug discovery. *ACS Cent. Sci.* **6**, 939–949 (2020).

42. J. Sieg, F. Flachsenberg, M. Rarey, In need of bias control: Evaluating chemical data for machine learning in structure-based virtual screening. *J. Chem. Inf. Model.* **59**, 947–961 (2019).

43. Y. Li, Z. Fan, J. Rao, Z. Chen, Q. Chu, M. Zheng, X. Li, An overview of recent advances and challenges in predicting compound-protein interaction (CPI). *Med. Rev.* **3**, 465–486 (2023).

44. J. Lyu, J. J. Irwin, B. K. Shoichet, Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* **19**, 712–718 (2023).

45. I. Muegge, P. Mukherjee, An overview of molecular fingerprint similarity search in virtual screening. *Expert Opin. Drug Discovery* **11**, 137–148 (2016).

46. M. Floris, A. Manganaro, O. Nicolotti, R. Medda, G. F. Mangiatordi, E. Benfenati, A generalizable definition of chemical similarity for read-across. *J. Chem.* **6**, 39 (2014).

47. P. Willett, Similarity searching using 2D structural fingerprints. *Methods Mol. Biol.* **672**, 133–158 (2011).

48. P. Willett, Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today* **11**, 1046–1053 (2006).

49. A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallvé, G. Pujadas, Molecular fingerprint similarity search in virtual screening. *Methods* **71**, 58–63 (2015).

50. A. B. Shapiro, G. K. Walkup, T. A. Keating, Correction for interference by test samples in high-throughput assays. *J. Biomol. Screen.* **14**, 1008–1016 (2009).

51. C. Yung-Chi, W. H. Prusoff, Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).

52. C. Smits, P. E. Czabotar, M. G. Hinds, C. L. Day, Structural plasticity underpins promiscuous binding of the prosurvival protein A1. *Structure* **16**, 818–829 (2008).

53. Y. Cao, A. Charisi, L.-C. Cheng, T. Jiang, T. Girke, ChemmineR: A compound mining framework for R. *Bioinformatics* **24**, 1733–1734 (2008).

54. M. Pilarczyk, M. F. Najafabadi, M. Kouril, J. Vasiliauskas, W. Niu, B. Shamsaei, N. Mahi, L. Zhang, N. Clark, Y. Ren, S. White, R. Karim, H. Xu, J. Biesiada, M. F. Bennet, S. Davidson, J. F. Reichard, V. Stathias, A. Koleti, D. Vidovic, D. J. B. Clark, S. Schurer, A. Ma'ayan, J. Meller, M. Medvedovic,

Connecting omics signatures of diseases, drugs, and mechanisms of actions with iLINCS. bioRxiv 826271 [Preprint] (2019). https://doi.org/10.1101/826271.

55. J. Biesiada, A. Porollo, J. Meller, “On setting up and assessing docking simulations for virtual screening” in *Rational Drug Design*, Methods in Molecular Biology (Humana Press, 2012), pp. 1–16; http://link.springer.com/protocol/10.1007/978-1-62703-008-3_1.

56. R Core Team, R: A language and environment for statistical computing (2017); www.R-project.org/.

57. D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C. Chang, C.-C. Lin, e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien (2019).

58. M. Kuhn, Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).

59. N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: An open chemical toolbox. *J. Chem.* **3**, 33 (2011).

60. K. Horan, T. Girke, ChemmineOB: R interface to a subset of OpenBabel functionalities, version 1.18.0 (2017).