Cytotoxic Profiling of Annotated and Diverse Chemical Libraries Using Quantitative High-Throughput Screening

Olivia W. Lee, Shelley M. Austin, Madison R. Gamma, Dorian M. Cheff, Tobie D. Lee, Kelli M. Wilson, Joseph M. Johnson, Jameson C. Travers, John C. Braisted, Rajarshi Guha, Carleen Klumpp-Thomas, Min Shen*, Matthew D. Hall*

National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA

* To whom correspondence should be addressed:
Min Shen: Tel: 301- 480-9794; E-mail: shenmin@mail.nih.gov
Matthew D. Hall: Tel: 301- 480-9928; E-mail: hallma@mail.nih.gov

Abstract

Cell-based phenotypic screening is a commonly used approach to discover biological pathways, novel drug targets, chemical probes and high-quality hit-to-lead molecules. Many hits identified from high-throughput screening campaign are ruled out through a series of follow-up potency, selectivity/specificity, and cytotoxicity assays. Prioritization of molecules with little or no cytotoxicity for a more in-depth downstream evaluation influence the future direction of projects, so cytotoxicity profiling of the screening libraries at early stage is essential for increasing the likelihood of candidate success. In this study, we analyzed results from a cell-based cytotoxicity screening campaign, comprising nearly 10,000 compounds in NCATS annotated libraries, and over 100,000 compounds in a diversity library, evaluated in four "normal" cell lines (HEK 293, NIH 3T3, CRL-7250 and HaCat) and a cancer cell line (KB 3-1, a HeLa subline) with overall screening outcomes, hit rates, pan-activity and selectivity. For the annotated library, we also examined the primary targets and mechanistic pathways regularly associated with cell death. To our knowledge, this is the first study to use high-throughput screening to profile large libraries for cytotoxicity in both normal and cancer cell lines. In summary, the results generated here constitutes a valuable resource for the scientific community and provides insight on identifying cytotoxic compounds with particular mechanisms of action, and prioritizing compounds with suitable cytotoxicity profile for further evaluation.

Introduction

The development of new chemical probes enables therapeutic target validation, hypothesis testing, and new insight into the biological role of genes and proteins (1). At the NIH National Center for Advancing Translational Sciences (NCATS), the development of small molecule probes for the scientific community allows understanding of rare and neglected diseases, novel targets, and enables basic biological understanding of the "undrugged" genome. This is accomplished through a team science approach that begins with assay development and automated quantitative high-throughput screening (qHTS) with a small molecule library to identify active hits (2). One or more chemotypes that emerge in a qHTS campaign progress to medicinal chemistry to develop a small molecule probe with strong biochemical and/or cell-based activity, specificity, and optimized properties to enable use in *in vivo* models. To accommodate unbiased qHTS discovery for medicinal chemistry, large libraries of small molecules are created and curated, containing molecules that capture diverse chemical space that are ideally synthetically tractable. These large libraries of small molecules are generally referred to diversity libraries or collections (3).

A second significant discovery strategy utilizes libraries of annotated small molecules. Annotated libraries contain drugs, probes and tool molecules with one or more known mechanisms of action (4). They have emerged as information-rich databases to integrate both biological and chemical data. These can be screened in (primarily) cell or organism-based assays to identify targets relevant to a phenotype, or for potential drug repurposing (5). While compounds in a diversity library are expected to demonstrate weak biological activity, annotated libraries by definition are medicinal chemistry-optimized products with known activity and in many cases, known mechanism of actions. Both diversity collection and annotated library screening are important components of the NCATS Chemical Genomics Center (NCGC) program (6).

Throughout a qHTS campaign, the activities of hits are confirmed in a re-test, and a number of orthogonal and counter-assays are performed to confirm that the observed modulatory activity of is on-target. This also to ensure that compounds demonstrating artefactual activity in an assay are triaged. For example, biochemical and cell-based assays that utilize firefly luciferase (fLuc) are sensitive to compounds that modulate luciferase activity (7). To ensure this is not the case, the screening libraries at NCATS are profiled for inhibitory activity of fLuc, allowing compounds that interfere with luciferase to be automatically triaged from hit lists without the need for additional screening. In one study it was shown that ~5% of compounds in a qHTS library inhibit fLuc (8). Perhaps the most notorious example of the need to perform counter-assays is the drug ataluren (PTC124) approved for treatment of patients with Duchenne muscular dystrophy, that may have been discovered due to fLuc inhibition rather than on-target activity (9).

A resurgence in cell-based screening, both target-based and phenotypic, means that an increasing number of cell-based qHTS campaigns are performed. Yet the profiling of libraries against cell-based assay read-outs is rarely reported, and only one report of assessing compound cytotoxicity across a library of 1,408 compounds appears in the literature (10). A significant collection of disease-agnostic annotated small molecules exist in libraries at NCATS, including a collection of drugs approved by American, European and Japanese therapeutic regulatory agencies (the NCATS Pharmaceutical Collection, NPC (11)), and many small molecules reported as tools, probes, or clinical/pre-clinical candidates. Understandably, a significant proportion of these molecules arose from oncology programs, and a portion of NCATS' collaborations are oncology-related phenotypic screens. We were motivated to profile the activity of our annotated libraries for two reasons: first, to allow scientists to have a reference dataset for discerning compounds whose activity is selective for cancer cell lines versus a set of non-cancer "normal" cell lines; and second, to enable scientists to discriminate promiscuous/cytotoxic compounds when reviewing data from cell-based/phenotypic assays and provide valuable input for prioritizing compounds for further evaluation.

To this end, we assessed the cytotoxicity of NCATS annotated libraries of nearly 10,000 compounds against four “normal” cell lines (HEK 293, NIH 3T3, CRL-7250 and HaCat) and a cancer cell line (KB 3-1, a HeLa subline), and examined the hit rates and mechanistic pathways regularly associated with cell death. Furthermore, we assessed that activity of a diversity library (>100,000 compounds) against two “normal” cell lines (HEK 293 and NIH 3T3), assessed hit rate, and compared active compounds against a cancer cell line (KB 3-1). This study provides insight into the extent of cell-based killing activity in annotated and diversity libraries, and the importance of confirming that active compounds in phenotypic screens are not cytotoxic.

Results

Profiling of an annotated library

Annotated libraries were profiled for cell viability by performing a primary screen against four normal cell lines using CellTiter-Glo (CTG) as the assay read-out (screening assay protocol displayed in Table 1). Three of the cell lines are immortalized (HEK 293, NIH 3T3 and HaCat), while CRL-7250 is a primary cell line. A total of 9,893 compounds were tested in either an 8-pt concentration-response ranging from 0.6 nM to 46 µM (1:5 dilution) or an 11-pt concentration-response ranging from 0.8 nM to 46 µM (1:3 dilution) according to different compound plating mechanisms.

Variation of sensitivity was observed across all four cell lines at both time-points (48 and 72 hr, Table 2). The HEK 293 cell line was more sensitive to compound treatment (Figure 1a), with 41% of compounds eliciting a dose-response cell killing effect (16% high-quality actives, 25% low-quality actives), while the three remaining cell lines demonstrated reduced sensitivity (7.9 – 10.6% high-quality actives, 14.2-18.1% low-quality actives), where the high-quality actives were defined as compounds in curve classes 1 and 2 with efficacy higher than 50% and EC_{50} values less than 10 µM. Comparison of compounds active against the four cell lines (Figure 1b, Venn diagram) revealed 853 compounds active only against HEK 293 cells (consistent with its greater sensitivity to compounds). 574 compounds demonstrated pan-activity against all four cell lines, which corresponds to a consensus hit rate of 5.8% for the entire annotated library.

Comparison of the AUC (Area Under the dose-response Curve) values of each compound using unbiased hierarchical clustering across four cell lines revealed some cross-correlation of activity, with the HEK 293 cells clustering away from the other cell lines (Figure 2a). Given the long-term intention of utilizing profiling data to make comparisons with activity against cancer cell lines, we also tested the library against KB 3-1 human adenocarcinoma cells (KB 3-1 cells are a HeLa sub-clone). Intriguingly, the KB 3-1 cell line did not demonstrate remarkably different sensitivity compared to the normal cells, but surprisingly its sensitivity clustered close with the

NIH 3T3 murine fibroblast line. The pair-wise AUC correlation with R^2 for all actives among five cell lines were calculated and shown in Figure 2b, the R^2 ranges from 0.31 to 0.77. NIH 3T3 and KB 3-1 showed the strongest correlation with $R^2 = 0.77$. Violin plots of the distribution of sensitivity to compounds across the libraries reinforce that the majority of compounds were not cytotoxic in cell lines (Figure 2c, pink area), while focused plots of only high-quality actives (curve class 1 or 2, Figure 2c, cyan area) show that only a small number of compounds are highly cytotoxic (AUC < −300) against cell lines.

One strength of annotated compound libraries is the ability to integrate compound mechanism of action (MOA) and targets into downstream analysis. Each compound that met the criteria for high-quality active in each cell line were aggregated by mechanism of action. The treemap shown in Figure 3 contains 1 box for each MOA containing three or more compounds. The size of each box is representative of the number of compounds with that MOA. The MOAs with the most compounds were antibacterial agents and PI3 Kinase inhibitors with 42 compounds present in each. We also assessed average AUC for each MOA which is represented in Figure 3 by the color gradient for each box. Kinesin like Spindle Protein (KIF11) inhibitors and proteasome inhibitors had the lowest average AUC of all MOAs.

There were several instances where multiple molecules with the same target were screened. To examine which MOAs were enriched among molecules active against cell lines, the proportion of active drugs to total drugs for each MOA target were used to derive an enrichment ratio (Figure 4a, where 1 = 100% active, 0 = none active). All clusters of MOA targets with at least five compounds presented in the annotated library were included in the enrichment analysis. The top five target clusters that resulted in consensus cell killing against the normal cell lines were proteasome, heat shock protein 90 (HSP90), anaplastic lymphoma kinase (ALK), mammalian target of rapamycin (mTOR), and cyclin-dependent kinase (CDK), with the enrichment ratio greater than 50%. To examine the significance of the association between consensus activity outcomes with the primary MOA, Fisher's exact test was applied on the entire data set and the statistical significance was reported in Supplementary Table 1. Eight out of eleven highly

enriched MOA clusters have P-value < 0.05, and seven clusters with P-value < 0.001, indicating that compounds with similar activity profiles as determined in normal cell line screens tend to share similar annotated MOAs. These highly enriched MOAs (along with many other targets shown in Figure 4a) are generally associated with anti-cancer cell killing by targeting essential cellular processes. For example, nine of ten proteasome inhibitors tested resulted in acute cytotoxicity across all four cell lines (see proteasome inhibitor delanzomib, Figure 4b). As a reference point, the cancer cell line KB 3-1 did not demonstrate hypersensitivity compared with the normal cell lines, this is also consistent to the results we discussed in the heatmap clustering analysis (Figure 2a). The global cytotoxicity of proteasome inhibitors is displayed by radar plot (Figure 4c), and is extended to other mechanisms of actions (Supplementary Figure 1).

As outlined, one use of cell profiling data for mechanistically-annotated screening libraries is to enable the identification of targeted agents selectively killing cancer cells. Two such examples of selective killing are displayed in Figures 4d and e. First, the dual specificity tyrosine phosphorylation regulated kinase 1A and 1B (DYRK1A and DYR1B) inhibitor TC-S7004 demonstrated complete killing of KB 3-1 cells (EC_{50} = 3.2 µM), with no effect on any of the four normal cell lines. Second, the potent and selective BMP receptor inhibitor LDN-212854 demonstrated complete killing of KB 3-1 cells (EC_{50} = 4.1 µM), and partial killing of one normal line (HEK 293, EC_{50} = 16.4 µM).

Profiling of a diversity library

The diversity library was profiled for cell viability by performing a primary screen against HEK 293 and NIH 3T3 cell lines, using CTG as the assay read-out (screening assay protocol displayed in Table 1). A total of 102,726 compounds were tested in a 4-pt dose response manner at varied concentrations with nearly 80% of the compounds tested at 115, 58, 12 and 2.3 µM. 30.0% and 34.6% of the library was cytotoxic to cell lines NIH 3T3 and HEK 293 respectively, with 1.2% of the library classified as high-quality actives in each of the cell lines (Fig 5 a, b). Of these high-

quality actives, there were 285 compounds overlapped and shared similar activity in both NIH 3T3 and HEK 293 (Fig 5 c).

After cherry-picking 588 hits based on their potency, maximal response and structural features, confirmation of activity was performed in 12 dose-point testing for most of the compounds against HEK 293, NIH 3T3, along with an orthogonal test against the KB 3-1 cell line. The final concentration of the compounds in the 5 µl assay volume ranged from 0.3 nM to 46 µM. To analyze the data, we generated a heat map with hierarchical clustering analysis (dendrogram) of compound activities based on their activity outcomes, which shows that NIH 3T3 clusters away from HEK 293 and KB 3-1 (Fig 5d).

The Promega CTG reagent utilizes an engineered version of firefly luciferase to measure ATP concentrations. As such, we tested compounds that appeared cytotoxic against recombinant fLuc in a biochemical assay to triage any compounds that were inhibiting fLuc rather than reducing cell viability. A number of compounds inhibited fLuc, and two examples from a novel fLuc inhibitor chemotype are shown in Figures 6a and 6b. In both cases, the compounds appeared equally active against all three cell lines in the CTG assay and demonstrated potent fLuc inhibition. The Promega CTG reagent contains a proprietary engineered fluc that claims to be resistant to inhibition compared with wild-type fLuc (12). This indeed appears to be the case, as a counter-assay testing hits in cell-free media with CTG and ATP added as substrate demonstrated at least a 10-fold lower sensitivity to inhibition that overlaid with the primary assay data in the presence of cell lines. Bright-field microscopy of cells grown in the presence of NCGC00413522 and NCGC00413607 (Figure 6c) demonstrated unaffected cell growth at 48 hr at concentrations suggested to be cytotoxic by CTG dose-response data (DMSO and 500 nM bortezomib as negative and positive control respectively).

Moreover, comparing the AUC values of high-quality actives in the three cell lines, we found that they moderately correlated: NIH 3T3 vs HEK 293, R^2 = 0.77; NIH 3T3 vs KB 3-1, R^2 = 0.59; KB 3-1 vs HEK 293, R^2 = 0.61 (Fig 7 a, b, c). Sixty-one of the top-ranking hits were active across all

cell lines (pan-active), whereas others were more selective toward a specific cell line or cell lines. If only taking the pan-actives that showed strong cytotoxicity against all three cell lines in the analysis, it was clearly shown that most of the compounds have very good EC_{50} agreement with low standard deviation, except for few outlier compounds that showed differential cell killing effect in three cell lines (Fig 7d). These outlier compounds had greater cell killing against two normal cell lines (HEK 293 and NIH 3T3) although the selectivity window to the cancer cell line KB-3-1 was not significant, individual dose-response curves were shown in Supplementary Fig 2. Among these five compounds, NCGC00420512, NCGC00420455 and NCGC00420435 belonged to the same structural chemotype with a tetrahydropyrazolo-pyrimidine core. To examine if the observed selectivity is a chemotype-specific outcome, we further analyzed the screening data through substructure search and found 172 compounds sharing the same tetrahydropyrazolo-pyrimidine scaffold presented in our diverse library, but none of the remaining 169 compounds showed similar differential cell killing effect, indicating that the selectivity is irrelevant to the specific structural features.

The remaining 56 pan-actives were cytotoxic across all three cell lines lacking any selectivity toward either normal or cancer lines. Representatives of the top-ranking pan-actives were shown in Figure 7e. Through structural clustering analysis of the pan-actives, we identified three enriched clusters based on their structural similarity using MACCS fingerprint and Tanimoto coefficient (Table 3). The most significant enriched scaffolds are in Clusters 1 and 2. Members in Cluster 1 shared the same pyrimidine-quinuclidine core, with comparable EC_{50} values across all three cell lines ranging from 7 to 12 µM, except one compound NCGC00421344 which was > 10-fold more potent in HEK 293 cell than the other two cell lines. Cluster 2 had the spiro structural feature and it was the most cytotoxic cluster among all four clusters discussed in Table 3. Compounds in this cluster showed almost identical activity across three cell lines, with EC_{50} values ranging from 0.5 to 3 µM for most of the compounds. For example, NCGC00419015 was the most cytotoxic agent in Cluster 2 with EC_{50} values for HEK 293, NIH 3T3 and KB 3-1 being 0.51, 0.72 and 0.64 µM, respectively. Furthermore, Cluster 3 with tetrahydro-imidazo-diazepine core was a less enriched cluster but also showed on-

selective clusters in which compounds were lacking selectivity against a specific cell line. We also ran the pan-actives against multiple toxicity datasets including Bursi Mutagenicity dataset (13) (4,337 compounds with mutagenicity data) and National Toxicology Program dataset (14) (503 structures with carcinogenicity data for male/female mouse and male/female rat), but found no overlapping structures which indicated that the chemotypes identified from our diverse library profiling are new additions to the well-known toxic collections.

Discussion

Here, we describe our effort to profile the cytotoxicity of two distinct libraries: 1. our annotated libraries (nearly 10,000 compounds of known MOAs or therapeutic indications) against four "normal" cell lines (HEK 293, NIH 3T3, CRL-7250 and HaCat) and a cancer cell line (KB 3-1, a HeLa sub-line); and 2. a diversity library (>100,000 compounds) against two "normal" cell lines (HEK 293 and NIH 3T3) and a cancer cell line (KB 3-1, a HeLa sub-line). The assays were all performed in quantitative HTS format. Hit rates and mechanistic pathways regularly associated with cell death are described. As one might reasonably anticipate, annotated (or mechanistic) libraries containing small molecules developed against targets or known to possess phenotypic activity have a high hit rate (10-20%) against the cell lines, especially for many annotated molecules were developed against oncology targets.

The large diversity collection revealed a low rate of cytotoxicity, and this is consistent with the fact that the majority are not optimized against targets, but reinforces the need to confirm toxicity of such compounds through counter-assays as part of screening for compound prioritization and triage. In our case, this profiling data is now used at NCATS to allow for rapid cross-referencing of active compounds in a screen against our dataset of general cytotoxicity in normal cell lines. To facilitate the use of these data by the broader research community, we have also made a significant amount of this data available through PubChem.

What is a "normal" cell line, and what are their limitations? The four "normal" cell lines used in this study were selected for a number of reasons. HEK 293 and NIH 3T3 cell lines are very

commonly employed in the scientific literature as control, normal, or comparator cell lines, along with conventional research uses. A common criticism of these lines is how normal they are given their capacity for uncontrolled cell growth. The HEK 293 cell line is one of the most utilized tool cell lines. It is a neuronal lineage line, generated in 1973 from embryonic kidney cells from an aborted fetus immortalized by transformation with adenovirus (15). The NIH 3T3 murine fibroblast cell line was generated in 1962 from Swiss albino mouse embryo, and spontaneously immortalized after multiple passages (16). The HaCat human keratinocyte cell line spontaneously immortalized in culture and was first reported in 1988 (17). All three of the aforementioned immortalized cell lines are adherent, and all are non-diploid despite being non-cancer non-tumorigenic cells. The CRL-7250 cell line differs from the others in that it is a primary, non-immortalized fibroblast cell line generated from human foreskin (18). The HEK 293 cell line demonstrated greater sensitivity than the other three "normal" cell lines in the annotated library screen.

One potential use for normal cell line profiling data is to enable the identification of compounds selectively active against cancer cell lines based on their target biology. A well-known example are inhibitors of MAPK/ERK kinase (MEK), that elicit acute cell killing in cells harboring somatic activating mutations of Ras, but not in cells expressing wild-type Ras (such as the "normal" cells discussed here (19)). As a proof-of-concept comparison, we tested the KB 3-1 adenocarcinoma cells against the annotated libraries. KB 3-1 cells are a sub-clone of HeLa cells (20), originally called KB squamous cell carcinoma before it was identified as a HeLa contaminant (21) and has subsequently been studied and acknowledged as such. KB cells do not possess activating Ras mutations, but selective activity was seen by a DYRK1A/B inhibitor that killed KB 3-1 cells without affecting the normal cells, demonstrating the utility. This profiling data has already been applied to analysis of multiple oncology-related screens across NCATS.

Perhaps the key result (reassuring from a screening perspective) from cytotoxicity profiling of >100,000 diversity library compounds was the very low rate of cytotoxicity observed, though some chemotype-related activity was observed. Cell lines like HEK 293 (along with cell lines

such as Chinese hamster ovary, CHO, cell line) are commonly used to engineer reporter cell lines for cell target-based and high-content assays, and the relative insensitivity to diversity libraries supports their utilization. The only prior study we have identified (also from NCATS but not involving any of the current authors) tested 1,408 compounds against thirteen cell lines, including a number of cancer cell lines along with HEK 293 and NIH 3T3 cells (10). While the compound number was relatively modest in scale, the normal HEK 293 and NIH 3T3 cells were reported to be among the least sensitive, supporting the basis for utilization of normal cell lines rather than cancer cell lines for designing cell-based assays for discovery screening.

The data reported here is made available through PubChem, and it is hoped this data acts as a general guide for normal cell line sensitivity to killing and assists in guiding others in the design of counter-assays for high-throughput screening of cell-based assays.

Acknowledgements

We thank Dr. Aleksandra Michaloa for advice regarding data analysis. This work was supported by the NCATS Division of Pre-Clinical Innovation Intramural Program.

Methods

Profiling annotated/diversity libraries and cherry-picked compounds. HEK 293, NIH 3T3, CRL-7250, HACAT and KB 3-1 cells were seeded into white 1536-well plates using a Multidrop Combi peristaltic dispenser (ThermoFisher, Waltham, MA) at a density of 250, 400, 500, 500, 500 cells/well in 5 µL of medium respectively. A pintool (Kalypsys) was used to transfer 23 nL of compound solution to the 1536-well assay plates. After 48 or 72 hr incubation at 37 °C, 5% CO and 85% humidity, 2.5 µL of CellTiter-Glo (Promega) was dispensed into each well using a dispenser (Aspect Automation, St. Paul, MN) with solenoid valves (Lee Valves, Westbrook CT). Plates were left at room temperature for 10 min before imaging the ATP-coupled luminescence using a ViewLux microplate imager (PerkinElmer, Waltham, MA).

Luciferase assay protocol. Assays determining firefly luciferase inhibition were performed as previously described (22). Briefly, 3 µL of luciferase substrate solution (10 µM ATP, 10 µM D-Luciferin, 10 mM Mg-Acetate, 0.01 % Tween-20, 0.05 % BSA, 50 mM Tris Acetate, pH 7.6, in final 4 µL volume) was dispensed into each well of white, solid bottom, 1536-well plates using a dispenser. A pintool (Kalypsis) was used to transfer 23 nL of compound solution to the assay plates. Following a 15 minute incubation at room temperature protected from light, 1 µL of purified luciferase enzyme solution was added to a final concentration of 10 nM *Photinus pyralis* luciferase (Sigma). Luminescence was detected by a Viewlux (Perkin-Elmer, Waltham, MA) by using a 10 s exposure time and 2X binning.

Data analysis and clustering of compounds by activity outcomes. To determine compound activity in the qHTS assay, the concentration-response data for each sample was plotted and modeled by a four parameter logistic fit yielding IC_{50} and efficacy (maximal response) values as previously described (23). Raw plate reads for each titration point were first normalized relative to positive control (2 mM Bortezomib, 0% activity, full inhibition) and DMSO only wells (basal, 100% activity). Data normalization and curve fitting were performed using in-house informatics tools. Compounds were designated as Class 1–4 according to the type of concentration–response curve (CRC) observed. In brief, Class 1.1 and 1.2 were the highest-confidence

complete CRCs containing upper and lower asymptotes with efficacies ≥ 80% and < 80%, respectively. Class 2.1 and 2.2 were incomplete CRCs having only one asymptote with efficacy ≥ 80% and < 80%, respectively. Class 3 CRCs showed activity at only the highest concentration or were poorly fit. Class 4 CRCs were inactive having a curve-fit of insufficient efficacy or lacking a fit altogether.

Compounds were further clustered hierarchically using TIBCO Spotfire 6.0.0 (Spotfire Inc., Cambridge, MA. https://spotfire.tibco.com/) based on their activity outcomes from the primary or follow up screen across different cell lines. Compound AUCs calculated based on the qHTS data analysis and curve fittings were utilized for clustering. In the heatmap, darker color indicates compounds that are more potent and efficacious, i.e. high-quality actives, and lighter color indicates less potent and efficacious compounds. If a compound didn't show any activity in an assay, it was highlighted as white in the heatmap.

Statistical analysis. In order to determine whether the hits predominantly identified from the specific therapeutic categories were over-represented in the chemical library, an enrichment analysis was implemented against the drug library. 9,893 compounds in the annotated library were broken down to different therapeutic categories based on their primary mechanisms of action and pharmaceutical indications, the enrichment was calculated from the following formula: $E = a/n$, given a is the number of actives and n is the total number of drugs in each therapeutic category. Fisher's exact tested was used as a measure of the consensus cytotoxicity potential of compounds in each MOA; all calculations were performed in R statistical computing software (https://www.r-project.org/).

Table 1. CellTiter-Glo assay protocol

Step		Value	Description
1	Cell Addition	5 µL	Cells in culture medium
2	Compound Addition	23 nL	Dilution series
3	Incubation	48 hr or 72 hr	37 °C, 5% CO_2, 85% RH
4	CellTiter-Glo Reagent Addition	2.5 µL	CellTiter-Glo Reagent Addition
5	Incubation	10 – 15 mins	Room temperature
6	Detection	Luminescence	ViewLux, 1 sec exposure

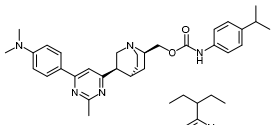
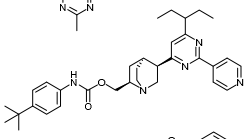
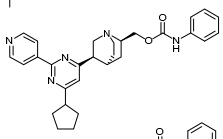
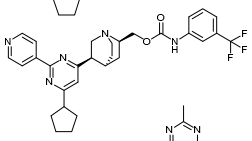
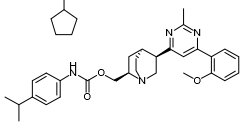
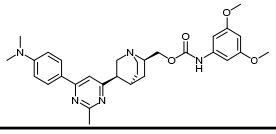
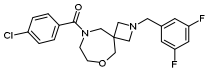
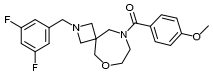
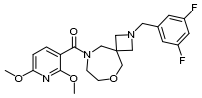
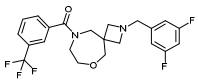
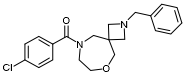
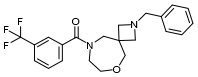
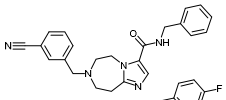
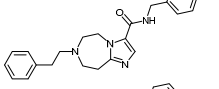
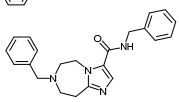
Notes	
1	Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies) supplemented with 10% fetal bovine serum (FBS; HyClone), 100 U/ml penicillin and 100 µg/ml streptomycin.
2	Pintool transfer
3	37 °C, 5% CO_2, 85% RH incubator incubation for 48 hr or 72 hr treatment conditions
4	CellTiter-Glo (Promega) reagent quantifies cellular ATP levels as a proxy for viability

Table 2. Summary table of qHTS screen against the annotated library in normal cell lines.

Cell line	Activity category	Number of compounds	
		48 hours	72 hours
HEK 293	High-quality actives*	1588	1820
	Low-quality actives*	2504	2600
	Inactives*	5801	5473
NIH 3T3	High-quality actives	1055	1020
	Low-quality actives	1788	1637
	Inactives	7050	7236
CRL-7250	High-quality actives	783	1146
	Low-quality actives	1408	1814
	Inactives	7702	6933
HaCat	High-quality actives	1050	887
	Low-quality actives	1624	1511
	Inactives	7219	7495
Total number of compounds in Annotated library		9893	9893

*High-quality active: compounds in class -1.1, -1.2, -2.1, -2.2, -2.3, and -2.4, maximum response ≤ -50%, and EC_{50} value ≤ 10 µM; inactives: compounds in curve class 4; low-quality actives: others.

Table 3. Representative clustered hits identified from qHTS against diversity chemical library.

Cluster	Sample ID	Structure	EC_{50} HEK 293 (μM)	EC_{50} NIH 3T3 (μM)	EC_{50} KB 3-1 (μM)	Avg. EC_{50} (μM)	SD
Cluster 1: pyrimidine-quinuclidine	NCGC00395730		7.91	7.91	7.05	7.62	0.50
	NCGC00394559		8.87	7.05	7.05	7.65	1.05
	NCGC00421344		0.11	11.17	12.53	7.94	6.81
	NCGC00421062		8.87	7.91	7.91	8.23	0.56
	NCGC00395320		8.87	8.87	11.17	9.64	1.33
	NCGC00395322		8.87	8.87	11.17	9.64	1.33
Cluster 2: spiro	NCGC00419015		0.51	0.72	0.64	0.63	0.11
	NCGC00418983		1.02	1.62	1.29	1.31	0.30
	NCGC00418997		1.15	2.29	1.62	1.68	0.57
	NCGC00419003		1.82	2.29	2.04	2.05	0.24
	NCGC00418995		2.57	5.12	4.56	4.08	1.34
	NCGC00419005		8.11	11.46	12.86	10.81	2.44
Cluster 3: tetrahydro-imidazo-diazepine	NCGC00409420		4.56	5.12	4.56	4.75	0.32
	NCGC00409515		5.74	7.23	7.23	6.74	0.86
	NCGC00409536		10.22	11.46	10.22	10.63	0.72

Supplementary Table 1. Fisher's exact test for enriched MOA clusters.

MOA Cluster	Number of actives in each MOA	Total number of actives	Total number of compounds in each MOA	Total number of compounds with MOA	P-value
Proteasome	9	344	10	2165	5.08E-07
hsp 90	15	344	19	2165	1.67E-09
ALK	4	344	7	2165	1.47E-02
mTOR	14	344	25	2165	4.26E-06
CDK	15	344	30	2165	1.22E-05
PI3K	23	344	49	2165	2.30E-07
HDAC	12	344	27	2165	3.86E-04
DNA Top	14	344	35	2165	4.78E-04
Aurora Kinase	6	344	24	2165	1.69E-01
EGFR	5	344	32	2165	5.92E-01
Tubulin	5	344	40	2165	7.86E-01

Figure legends

Figure 1. a) Pie chart distribution of high-quality actives (orange), low-quality actives (yellow), and inactives (green) identified from the annotated compound library qHTS against the four normal cell lines for 48 hr incubation condition. High-quality actives: compounds in class -1.1, -1.2, -2.1, -2.2, -2.3, and -2.4, maximum response ≤ -50%, and EC_{50} value ≤ 10 µM. Inactives: compounds in curve class 4; low-quality actives: others. **b)** Compound overlapping Venn diagram for HEK 293, HaCat, NIH 3T3, and CRL-7250 cell lines. Number of high-quality actives in each cell line and number of compounds overlapped were calculated.

Figure 2. a) The comparison of AUC values of each compound in four normal cell lines and KB 3-1 human adenocarcinoma cells. In the heat map each row corresponds to a compound and each column to a cell line. Darker red color indicates more potent and efficacious compound. **b)** Pair-wise R^2 correlation matrix among five cell lines. Only high-quality actives were included in this analysis and the Pearson correlation was based on AUC values. **c)** Split violin plot showing the distribution of AUC values for all compounds screened and high-quality actives across four normal cell lines. The lines within each distribution area represent the 0.25, 0.5 and 0.75 quantiles.

Figure 3. A treemap representation of the mechanisms of action (MOA) of all high-quality actives from the annotated library compounds screened. Box size represents the total number of compounds representing each MOA (bigger box size indicates more compounds present in the high-quality actives). Color represents the average AUC from the cytotoxicity screen in four normal cell lines (darker red indicates a lower AUC meaning a more potent and efficacious hits). 1 = Lineage specific differentiation; 2 = RNA polymerase; 3 = Immuno-suppressant.

Figure 4. **a)** Enrichment analysis of active agents in each drug category. Enrichment ratio = the number of actives/the total number of drugs in each drug category. **b)** Dose-response curves for delanzomib, a representative proteasome inhibitor, in cytotoxicity screens. **c)** A radar plot

displaying Log AC_{50} of all active proteasome inhibitors in four normal cell lines. **d)** Dose-response curves for TC-S7004, a potent and selective DYRK1A/B inhibitor, in cytotoxicity screens, including KB 3-1 cells. **e)** Does-response curves for LDN-212854, an ALK2 inhibitor, in cytotoxicity screens, including KB 3-1 cells.

Figure 5. **a-b)** Pie chart distribution of high-quality actives (orange), low-quality actives (yellow), and inactives (green) identified from the qHTS against diversity collection compound library in NIH 3T3 and HEK 293 cell line, respectively. **c)** Compound overlapping Venn diagram for NIH 3T3 and HEK 293 cell lines. Number of high-quality actives in each cell line was calculated and number of compounds overlapped were labeled. **d)** The comparison of AUC values of 588 cherry-picked compounds in HEK 293, NIH 3T3 and KB 3-1 cell lines. In the heat map each row corresponds to a compound and a column to a cell line. Darker red color indicates more potent and efficacious compound.

Figure 6. **a-b)** Dose-response curves for NCGC00413522 and NCGC00413607 in cytotoxicity screens, luciferase inhibition assay and CellTiter-Glo luciferase assay. **c)** Representative bright field images of HEK293 cells after 48 hr treatment.

Figure 7. **a-c)** Activity correlation plot between HEK 293, NIH 3T3 and KB 3-1 cell lines. Only high-quality actives were included in this analysis and the correlation was based on AUC values. **d)** Log EC_{50} distribution of the pan-active compounds in all three cell line screens. Error bars represent the standard deviation of Log EC_{50}. **e)** Dose-response curves for representative compounds in diversity library showing pan-activity across three cell lines.

Supplementary Figure 1. Radar plots displaying the Log AC_{50} for each compound in four normal cell lines. Radar plots are separated by MOAs that were overrepresented in the high-quality actives compounds such as HDAC, PI3K, HSP90, Proteasome, CDK, ALK, mTOR, EGFR, Tubulin, Aurora, and Topoisomerase.

Supplementary Figure 2. Dose-response curves for representative compounds in diversity library screen showing differential cell killing effect across three cell lines.

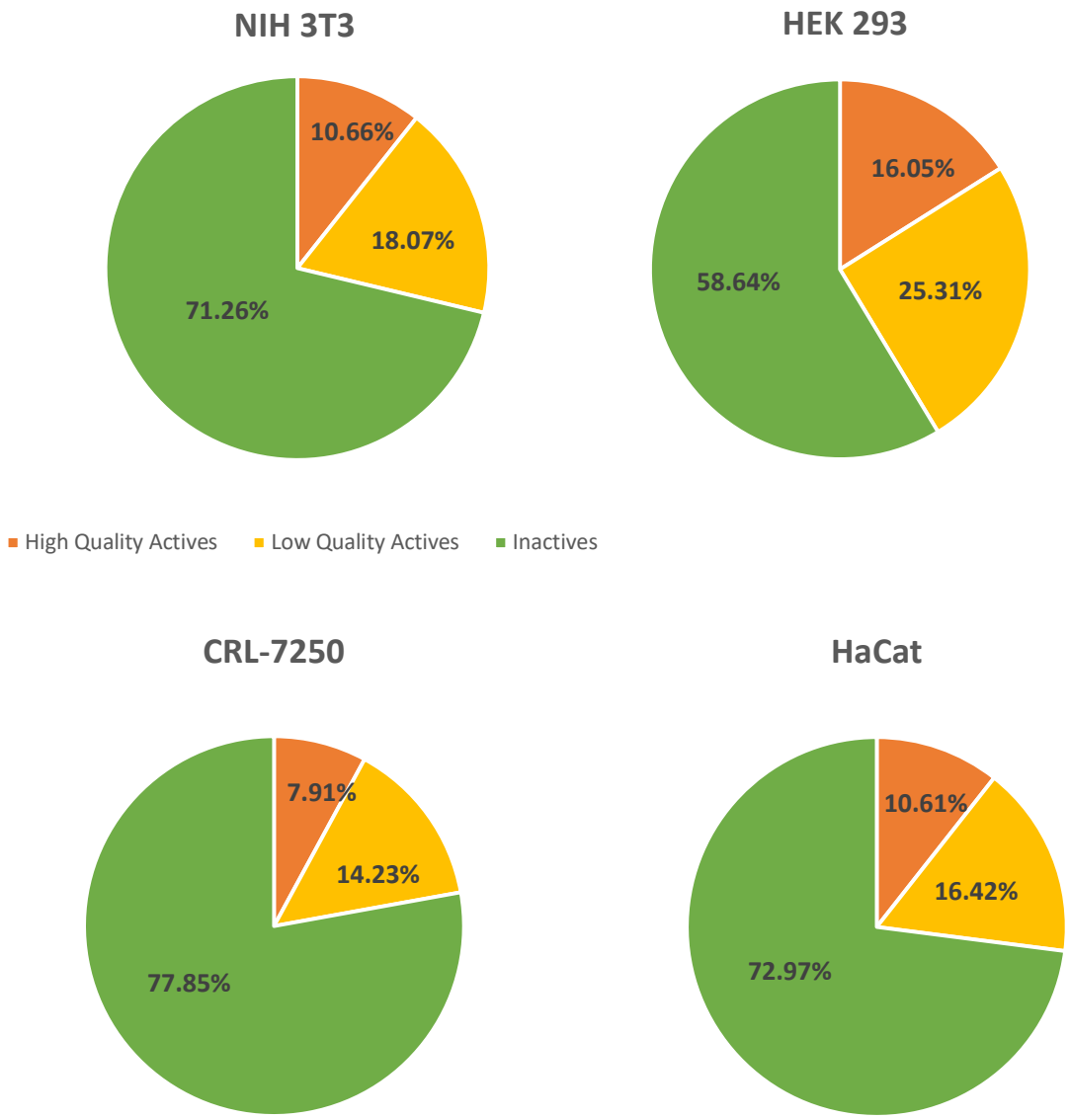
References

1. Schreiber SL, Kotz JD, Li M, Aube J, Austin CP, Reed JC, et al. Advancing Biological Understanding and Therapeutics Discovery with Small-Molecule Probes. Cell. 2015;161(6):1252-65.
2. Michael S, Auld D, Klumpp C, Jadhav A, Zheng W, Thorne N, et al. A robotic platform for quantitative high-throughput screening. Assay Drug Dev Technol. 2008;6(5):637-57.
3. Renner S, Popov M, Schuffenhauer A, Roth HJ, Breitenstein W, Marzinzik A, et al. Recent trends and observations in the design of high-quality screening collections. Future Med Chem. 2011;3(6):751-66.
4. Wassermann AM, Camargo LM, Auld DS. Composition and applications of focus libraries to phenotypic assays. Front Pharmacol. 2014;5:164.
5. Coussens NP, Braisted JC, Peryea T, Sittampalam GS, Simeonov A, Hall MD. Small-Molecule Screens: A Gateway to Cancer Therapeutic Agents with Case Studies of Food and Drug Administration-Approved Drugs. Pharmacol Rev. 2017;69(4):479-96.
6. Thomas CJ, Auld DS, Huang R, Huang W, Jadhav A, Johnson RL, et al. The pilot phase of the NIH Chemical Genomics Center. Curr Top Med Chem. 2009;9(13):1181-93.
7. Auld DS, Lovell S, Thorne N, Lea WA, Maloney DJ, Shen M, et al. Molecular basis for the high-affinity binding and stabilization of firefly luciferase by PTC124. Proc Natl Acad Sci U S A. 2010;107(11):4878-83.
8. Auld DS, Thorne N, Nguyen DT, Inglese J. A specific mechanism for nonspecific activation in reporter-gene assays. ACS Chem Biol. 2008;3(8):463-70.
9. Auld DS, Thorne N, Maguire WF, Inglese J. Mechanism of PTC124 activity in cell-based luciferase assays of nonsense codon suppression. Proc Natl Acad Sci U S A. 2009;106(9):3585-90.
10. Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, et al. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect. 2008;116(3):284-91.
11. Huang R, Southall N, Wang Y, Yasgar A, Shinn P, Jadhav A, et al. The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. Sci Transl Med. 2011;3(80):80ps16.
12. Hook B, Schagat T. Compound Interference of CellTiter-Glo® vs PE ATPlite™ 1 step [Poster]. Promega Corporation; [Available from: https://www.promega.com/-/media/files/resources/posters/compound-interference-of-celltiter-glo-vs-pe-atplite-1-step-poster.pdf.
13. Kazius J, McGuire R, Bursi R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. Journal of Medicinal Chemistry. 2005;48(1):312-20.
14. US National Toxicology Program data set [Internet]. 2001 [cited 2018 Aug 23]. Available from: https://www.predictive-toxicology.org/data/ntp/.
15. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol. 1977;36(1):59-74.
16. Todaro GJ, Green H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. J Cell Biol. 1963;17:299-313.

17. Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig NE. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J Cell Biol. 1988;106(3):761-71.
18. ATCC® CRL7250™ Product Sheet [PDF]. 2015 [cited 2018 Aug 24]. https://www.atcc.org/~/ps/CRL-7250.ashx: American Type Culture Collection.
19. Zhao Y, Adjei AA. The clinical development of MEK inhibitors. Nat Rev Clin Oncol. 2014;11(7):385-400.
20. Cornwell MM, Gottesman MM, Pastan IH. Increased vinblastine binding to membrane vesicles from multidrug-resistant KB cells. J Biol Chem. 1986;261(17):7921-8.
21. Vaughan L, Glanzel W, Korch C, Capes-Davis A. Widespread Use of Misidentified Cell Line KB (HeLa): Incorrect Attribution and Its Impact Revealed through Mining the Scientific Literature. Cancer Res. 2017;77(11):2784-8.
22. Auld DS, Southall NT, Jadhav A, Johnson RL, Diller DJ, Simeonov A, et al. Characterization of chemical libraries for luciferase inhibitory activity. J Med Chem. 2008;51(8):2372-86.
23. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, et al. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci U S A. 2006;103(31):11473-8.

Figure 1

a


NIH 3T3
10.66%
18.07%
71.26%
HEK 293
16.05%
25.31%
58.64%
High Quality Actives
Low Quality Actives
Inactives
CRL-7250
7.91%
14.23%
77.85%
HaCat
10.61%
16.42%
72.97%

b

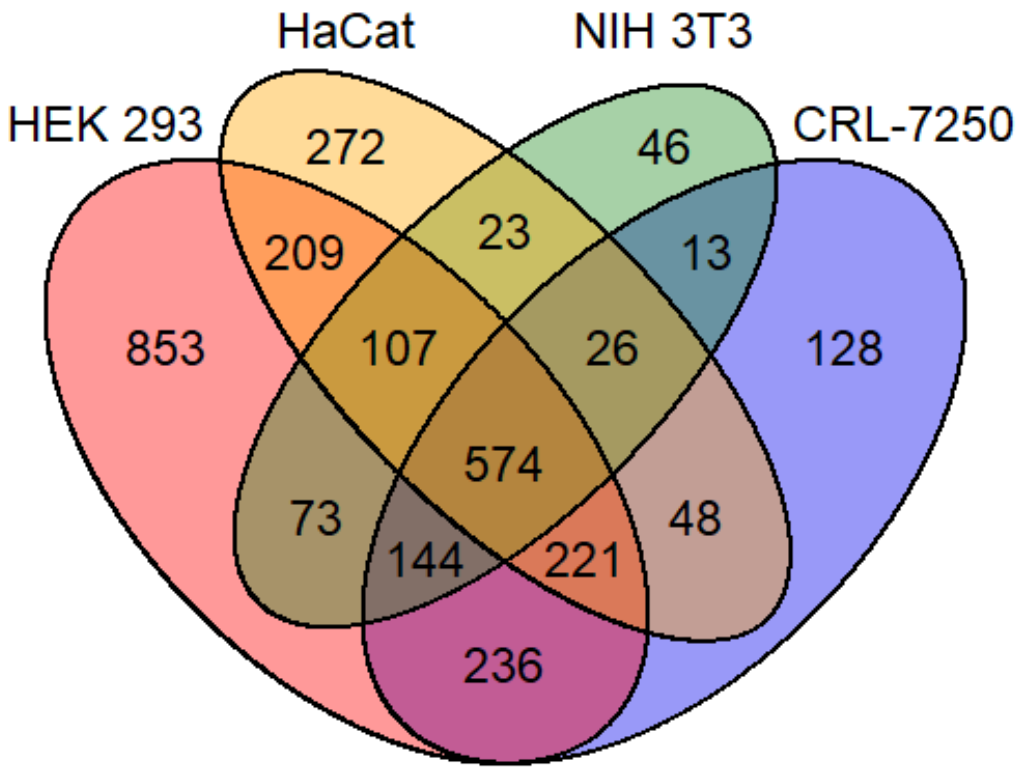
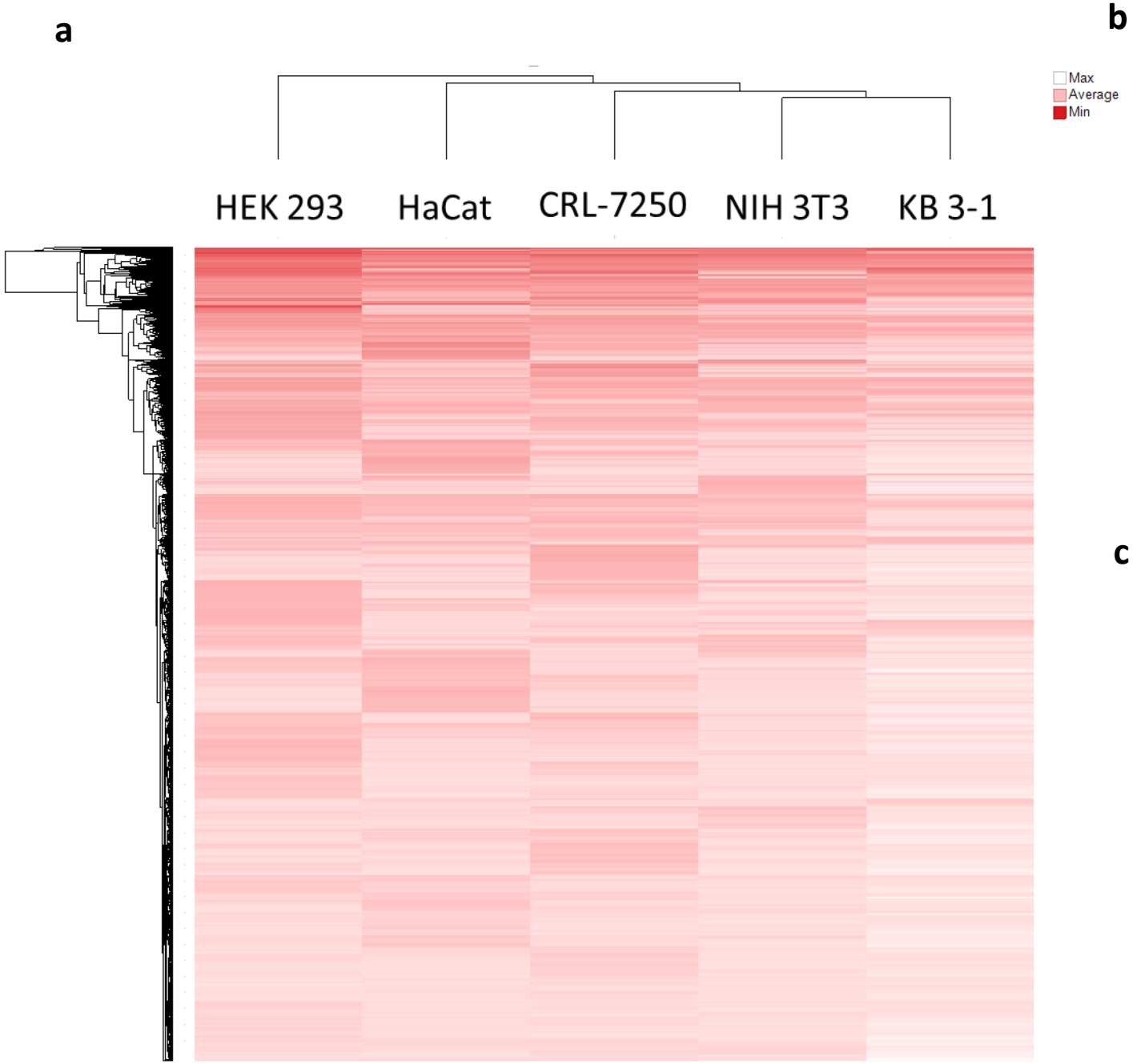
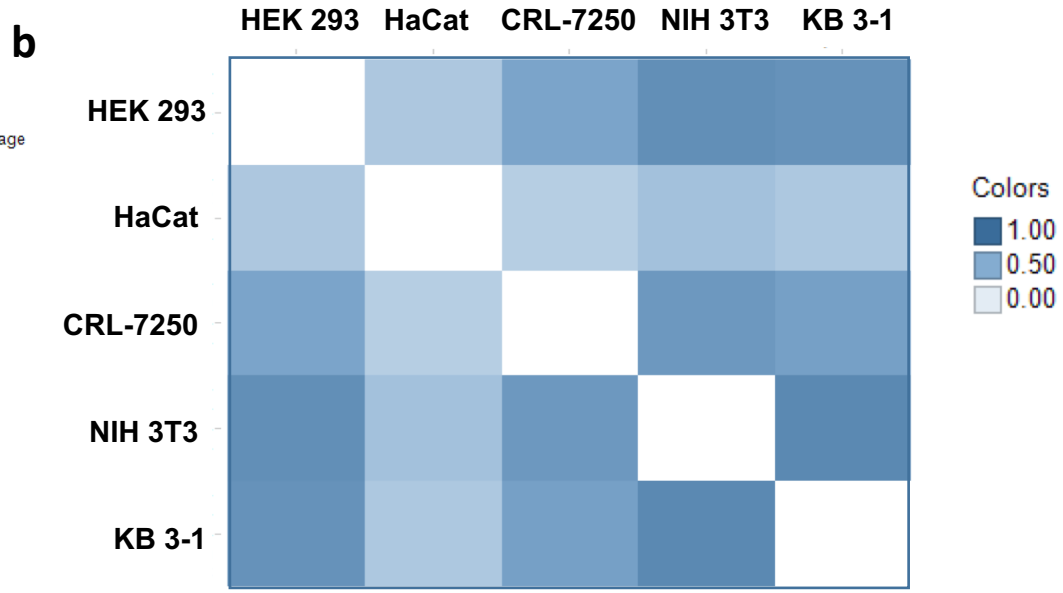

HaCat
NIH 3T3
HEK 293
CRL-7250
272
46
209
23
13
853
107
26
128
574
73
48
144
221
236

Figure 2


a
HEK 293
HaCat
CRL-7250
NIH 3T3
KB 3-1
Max
Average
Min


b
HEK 293
HaCat
CRL-7250
NIH 3T3
KB 3-1
Colors
1.00
0.50
0.00

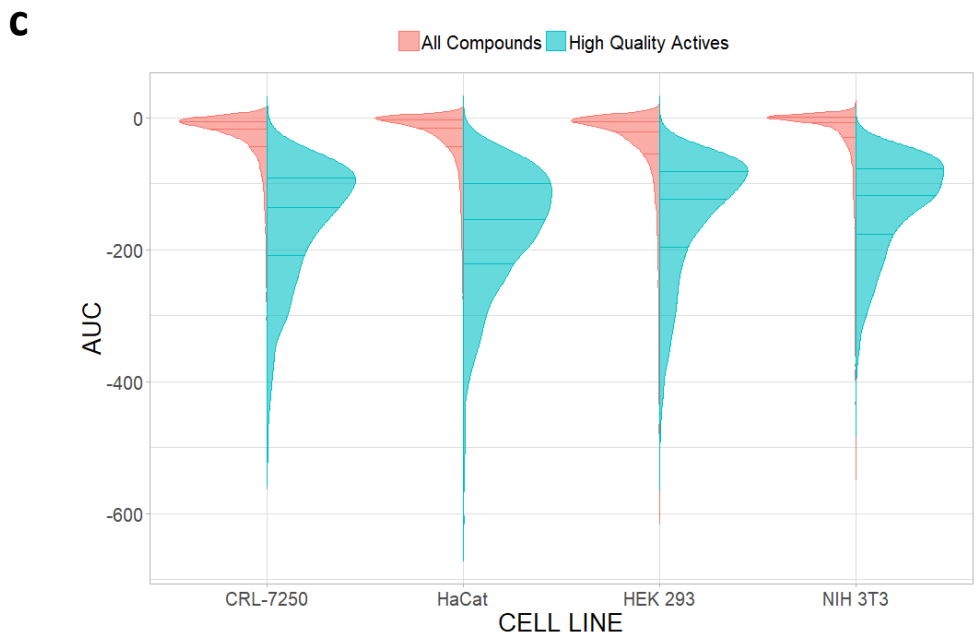

c
All Compounds
High Quality Actives
0
-200
-400
-600
AUC
CRL-7250
HaCat
HEK 293
NIH 3T3
CELL LINE

Figure 3

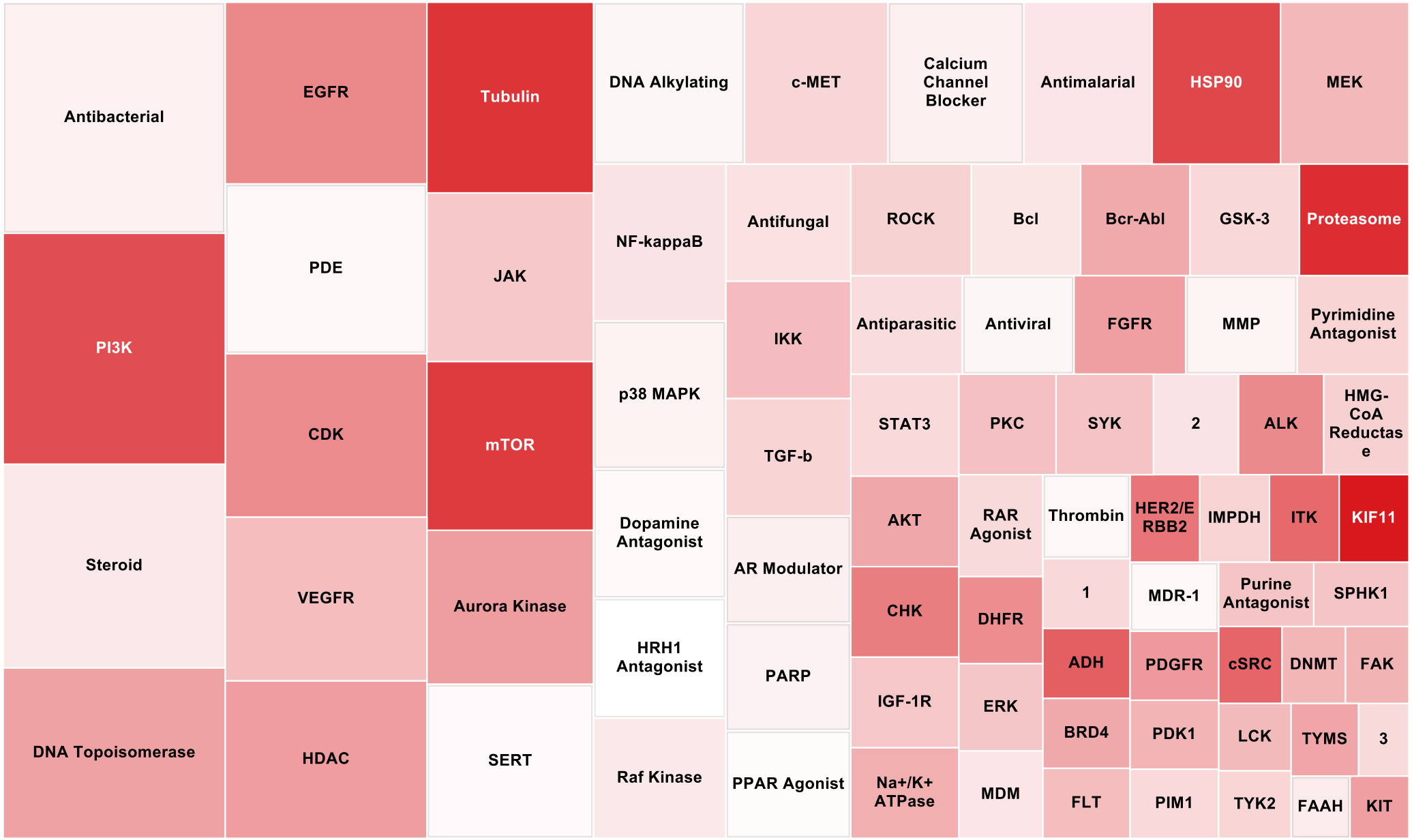
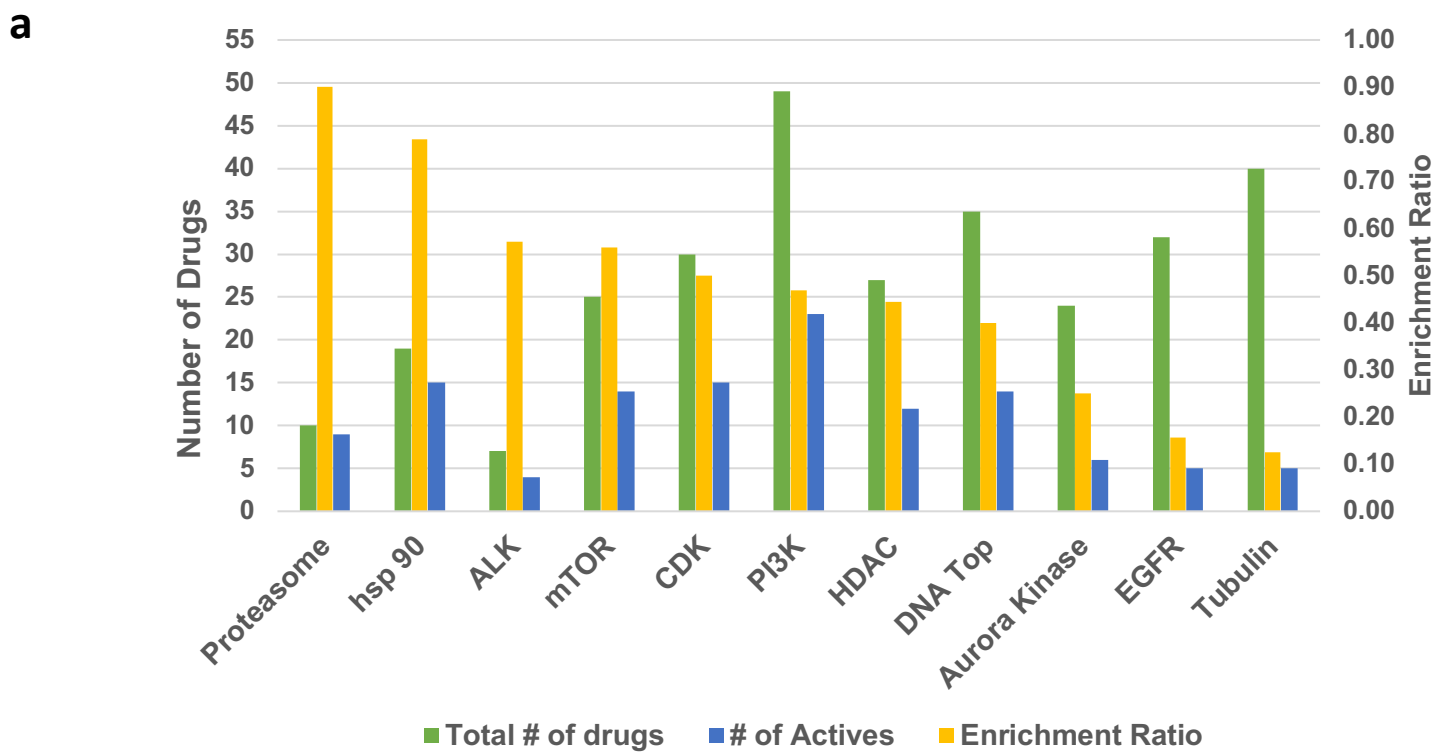
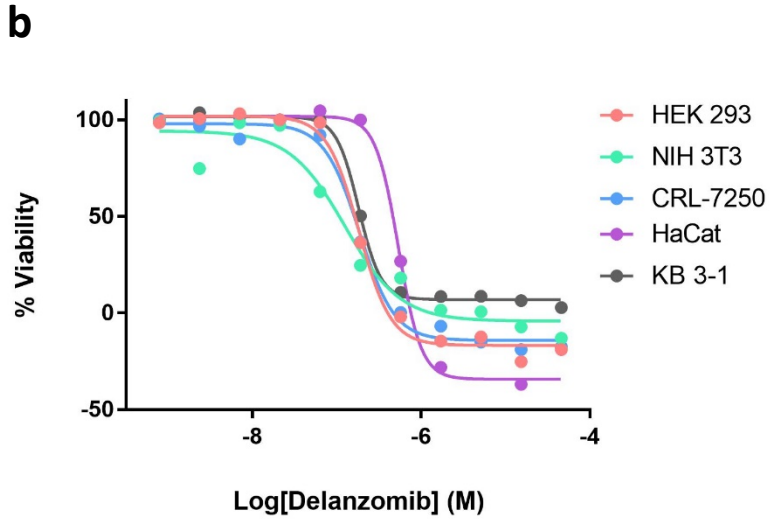
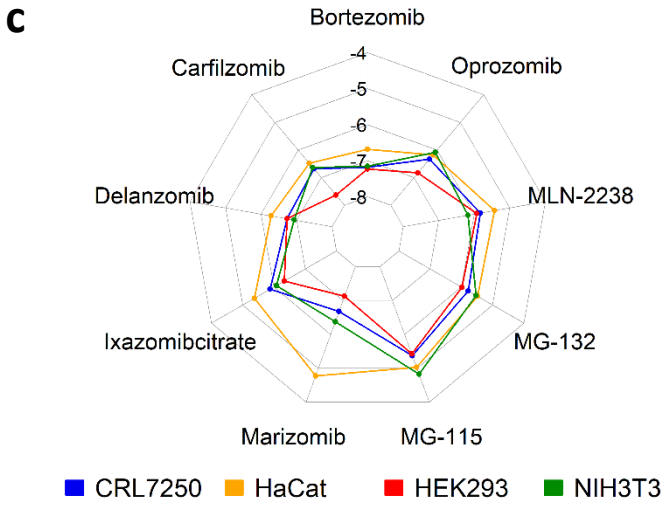
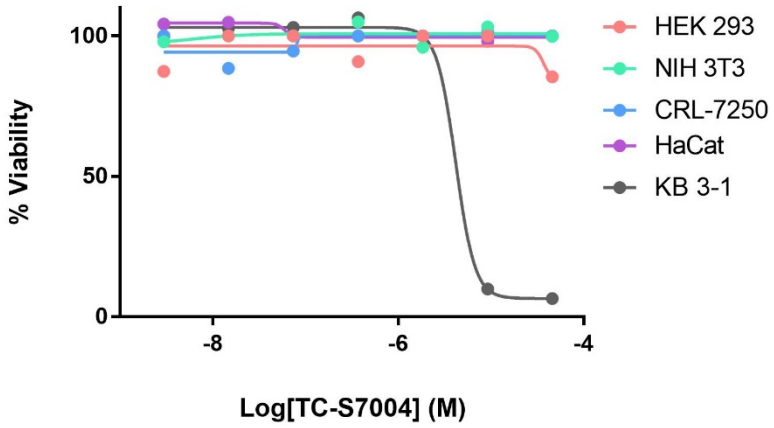

Antibacterial
EGFR
Tubulin
DNA Alkylating
c-MET
Calcium Channel Blocker
Antimalarial
HSP90
MEK
PI3K
PDE
JAK
NF-kappaB
Antifungal
ROCK
Bcl
Bcr-Abl
GSK-3
Proteasome
CDK
mTOR
p38 MAPK
IKK
Antiparasitic
Antiviral
FGFR
MMP
Pyrimidine Antagonist
Steroid
VEGFR
Aurora Kinase
Dopamine Antagonist
TGF-b
STAT3
PKC
SYK
2
ALK
HMG-CoA Reductase
DNA Topoisomerase
HDAC
SERT
HRH1 Antagonist
AR Modulator
AKT
RAR Agonist
Thrombin
HER2/ERBB2
IMPDH
ITK
KIF11
PARP
CHK
DHFR
1
MDR-1
Purine Antagonist
SPHK1
Raf Kinase
PPAR Agonist
IGF-1R
ERK
ADH
PDGFR
cSRC
DNMT
FAK
BRD4
PDK1
LCK
TYMS
3
Na+/K+ ATPase
MDM
FLT
PIM1
TYK2
FAAH
KIT

Figure 4


a
Number of Drugs
Enrichment Ratio
55
50
45
40
35
30
25
20
15
10
5
0
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Proteasome
hsp 90
ALK
mTOR
CDK
PI3K
HDAC
DNA Top
Aurora Kinase
EGFR
Tubulin
Total # of drugs
# of Actives
Enrichment Ratio


b
% Viability
100
50
0
-50
-8
-6
-4
Log[Delanzomib] (M)
HEK 293
NIH 3T3
CRL-7250
HaCat
KB 3-1


c
Bortezomib
Oprozomib
MLN-2238
MG-132
MG-115
Marizomib
Ixazomibcitrate
Delanzomib
Carfilzomib
-4
-5
-6
-7
-8
CRL7250
HaCat
HEK293
NIH3T3


d
% Viability
100
50
0
-8
-6
-4
Log[TC-S7004] (M)
HEK 293
NIH 3T3
CRL-7250
HaCat
KB 3-1

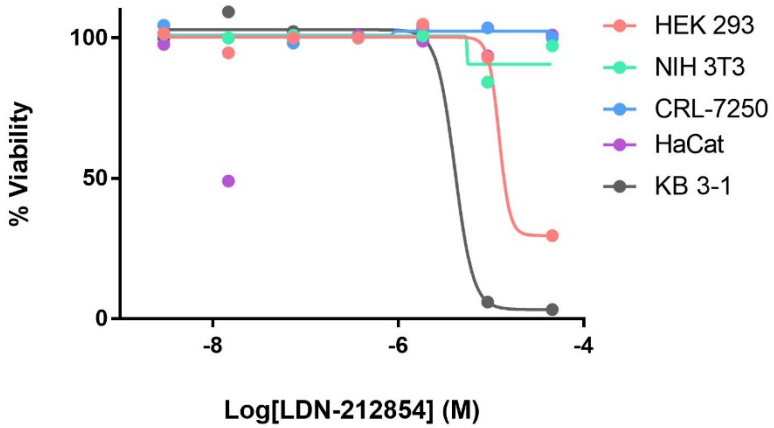

e
% Viability
100
50
0
-8
-6
-4
Log[LDN-212854] (M)
HEK 293
NIH 3T3
CRL-7250
HaCat
KB 3-1

Figure 5

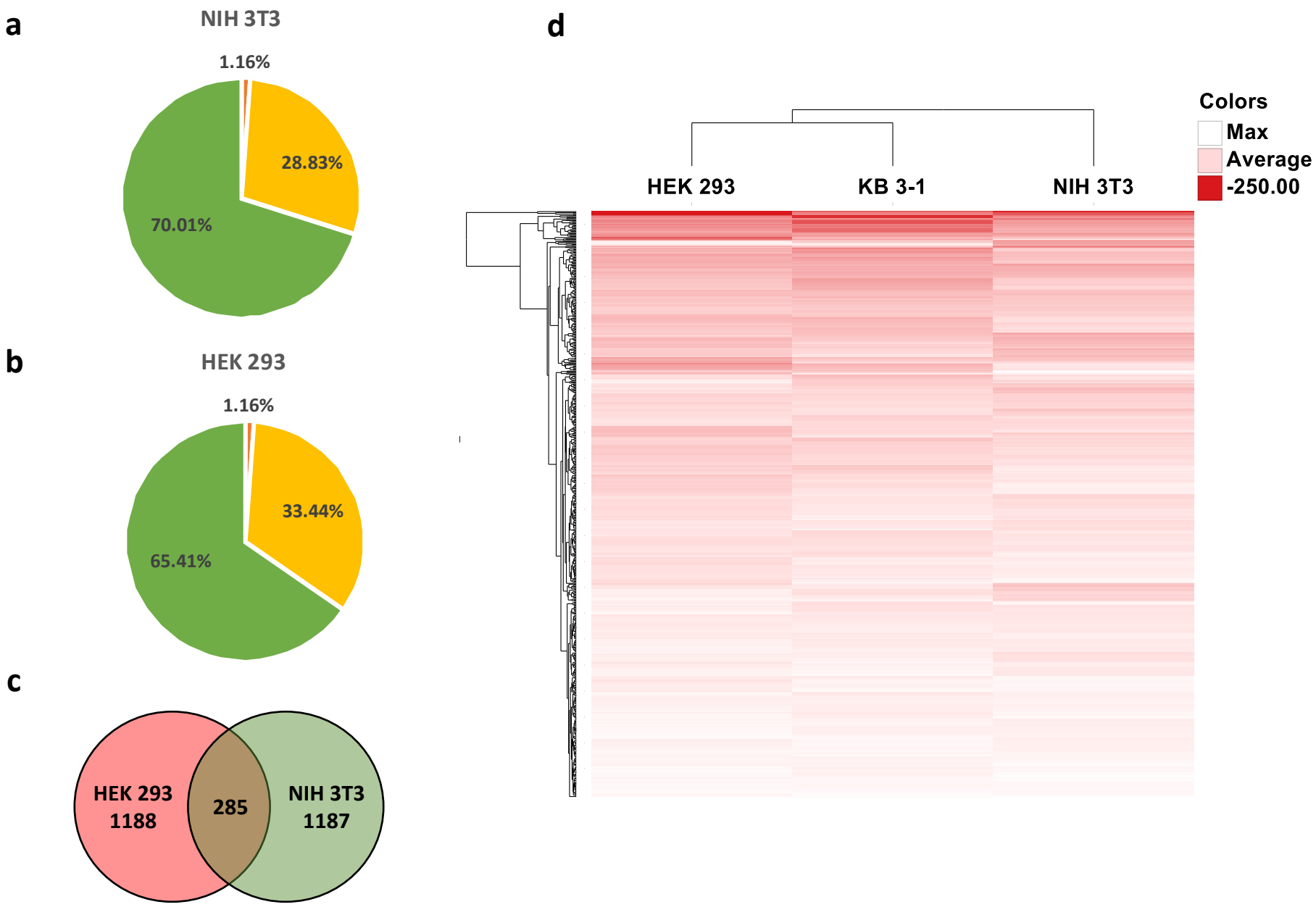
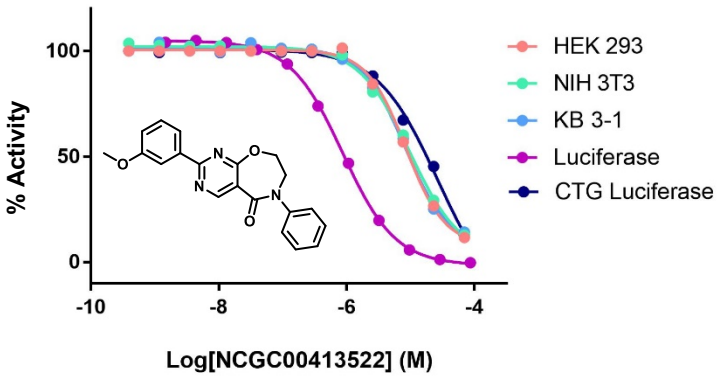
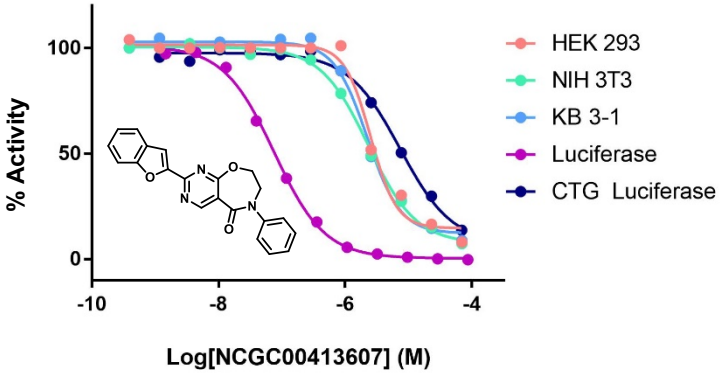

a
NIH 3T3
1.16%
28.83%
70.01%
b
HEK 293
1.16%
33.44%
65.41%
c
HEK 293
1188
285
NIH 3T3
1187
d
HEK 293
KB 3-1
NIH 3T3
Colors
Max
Average
-250.00

Figure 6

a


100
50
0
% Activity
-10
-8
-6
-4
Log[NCGC00413522] (M)
HEK 293
NIH 3T3
KB 3-1
Luciferase
CTG Luciferase

b


100
50
0
% Activity
-10
-8
-6
-4
Log[NCGC00413607] (M)
HEK 293
NIH 3T3
KB 3-1
Luciferase
CTG Luciferase

c

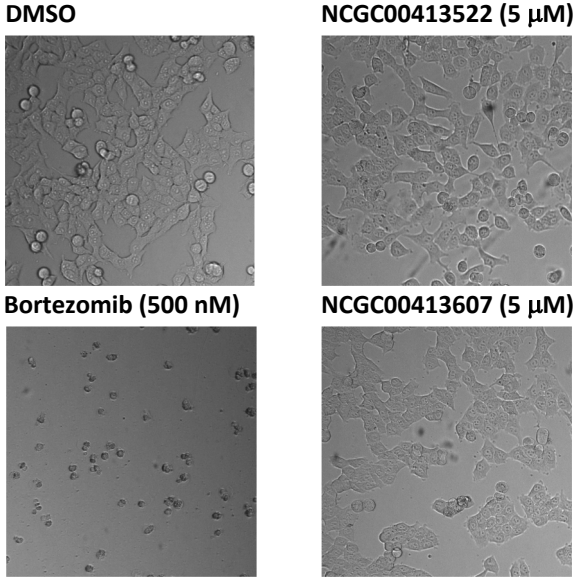
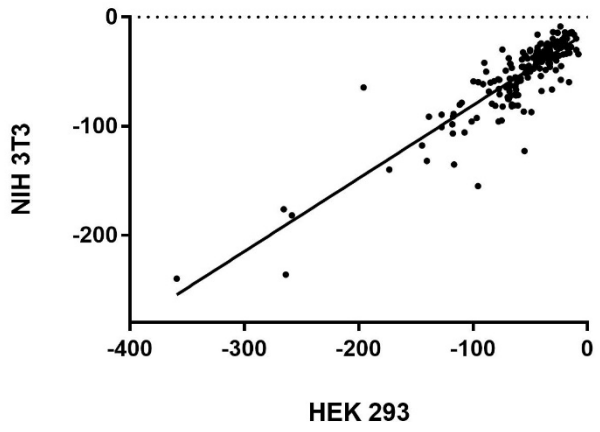
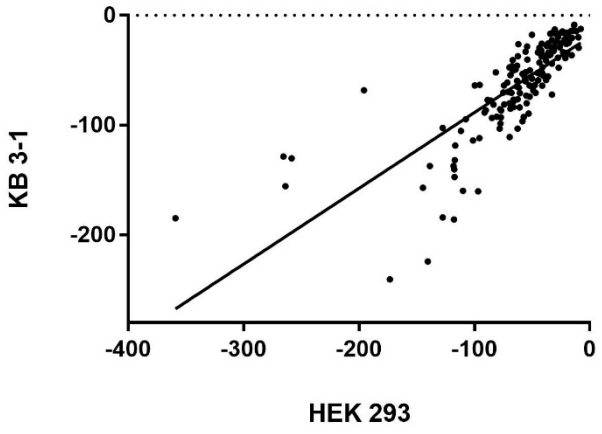

DMSO
NCGC00413522 (5 µM)
Bortezomib (500 nM)
NCGC00413607 (5 µM)

Figure 7

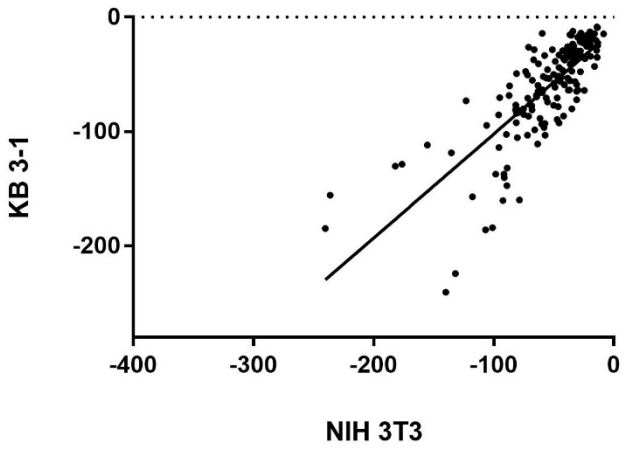
a



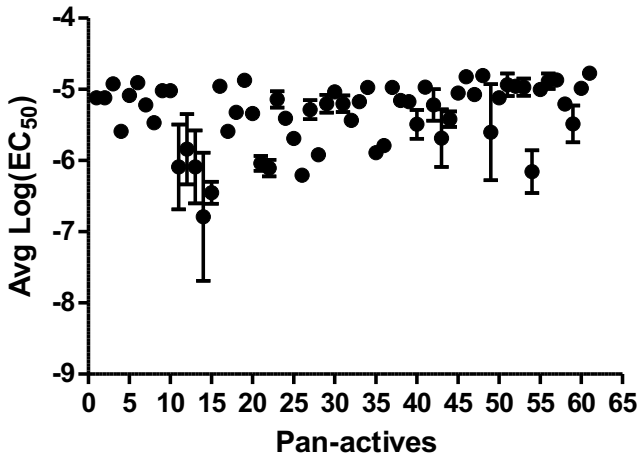
b



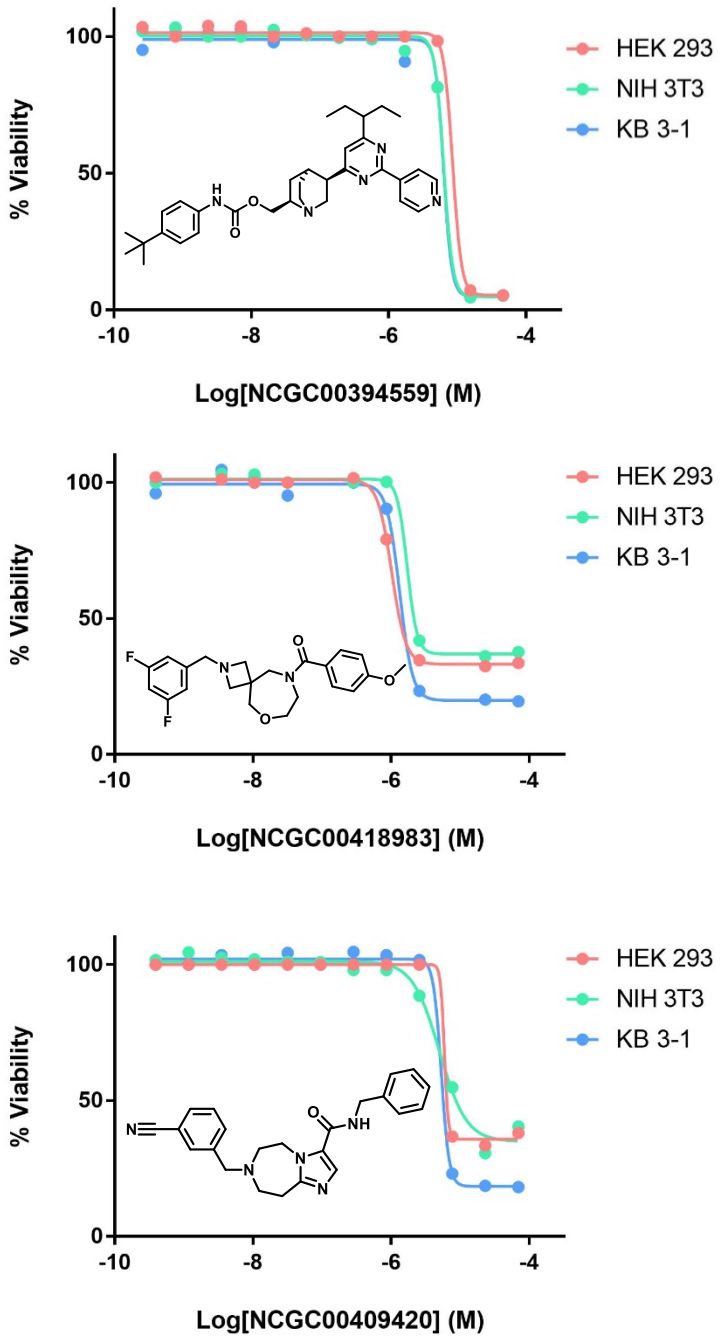
c



d

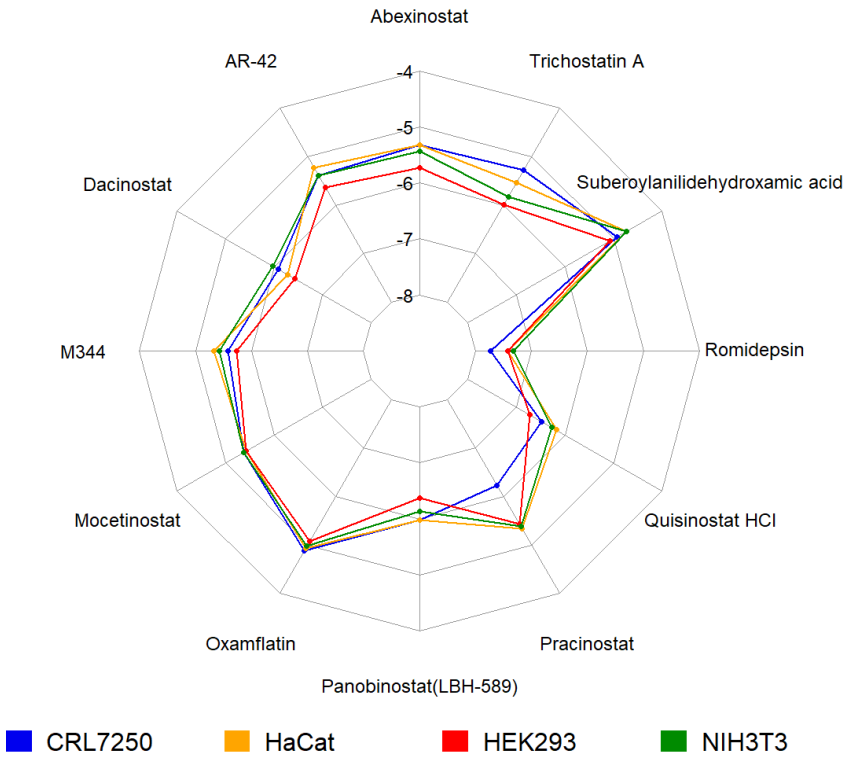


e

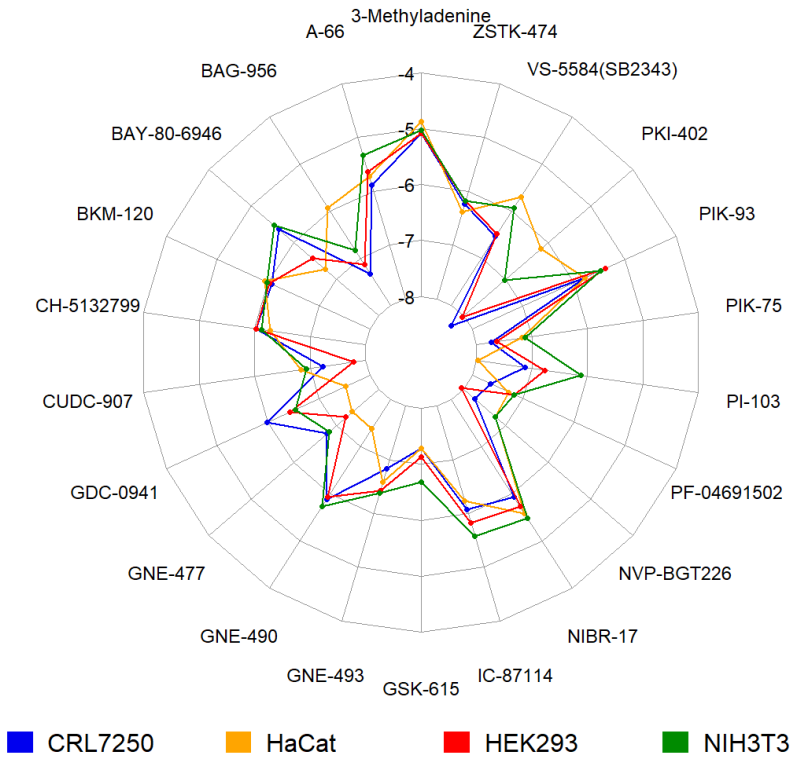


Supplementary Figure 1.

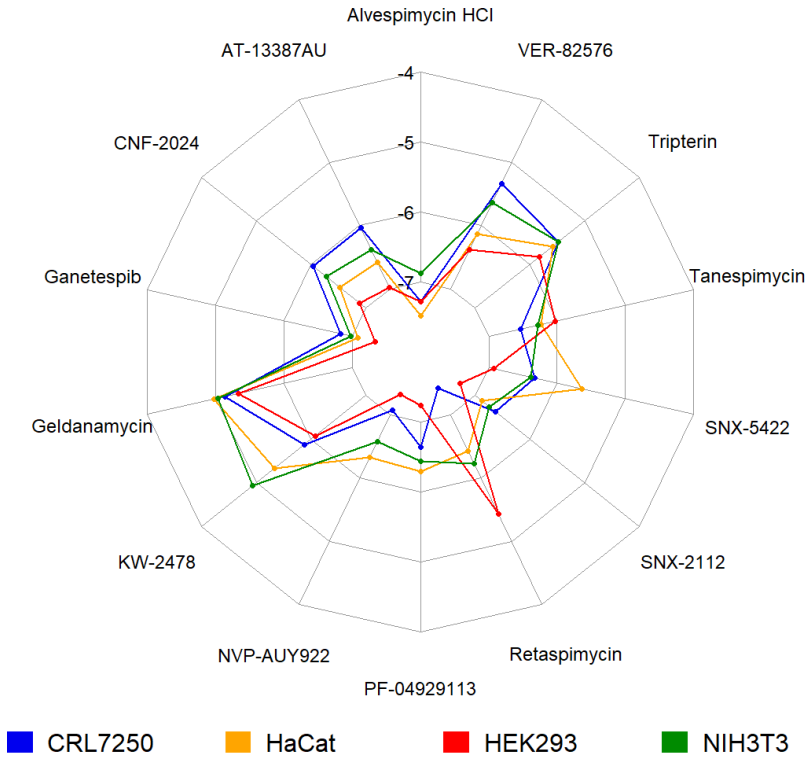
HDAC



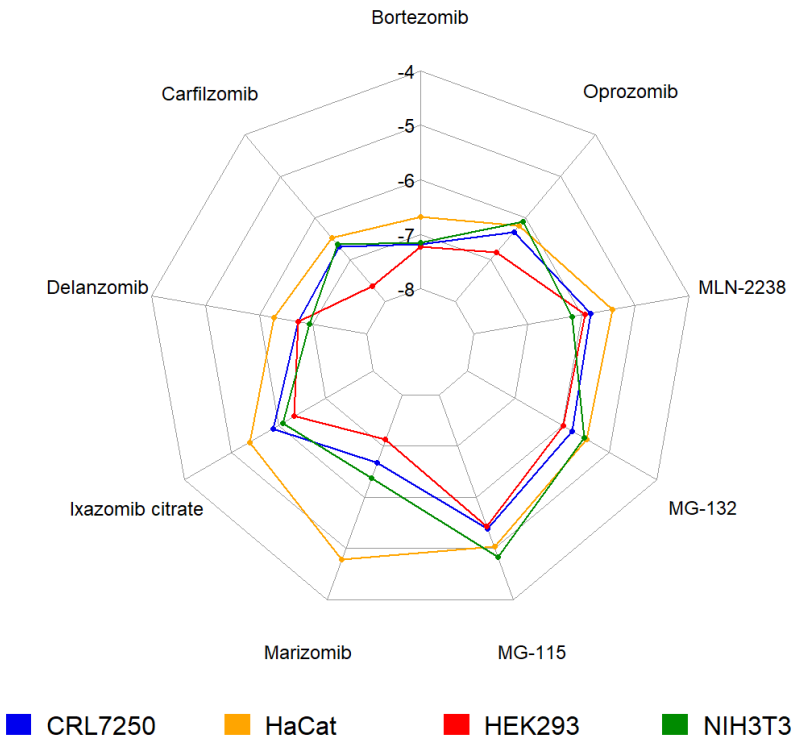
PI3K



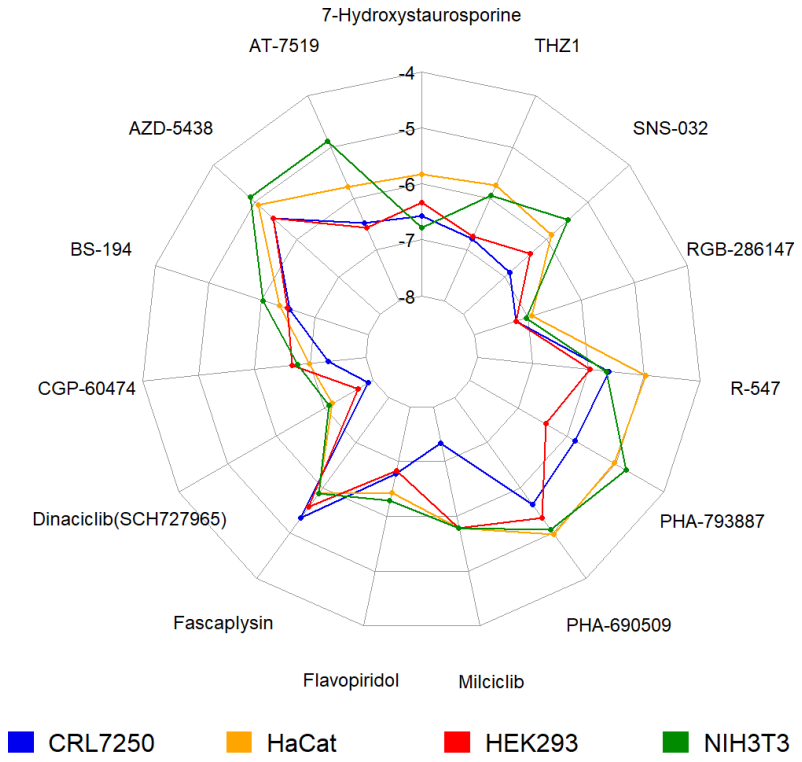
HSP90


Alvespimycin HCl
VER-82576
Tripterin
Tanespimycin
SNX-5422
SNX-2112
Retaspimycin
PF-04929113
NVP-AUY922
KW-2478
Geldanamycin
Ganetespib
CNF-2024
AT-13387AU
-4
-5
-6
-7
CRL7250
HaCat
HEK293
NIH3T3

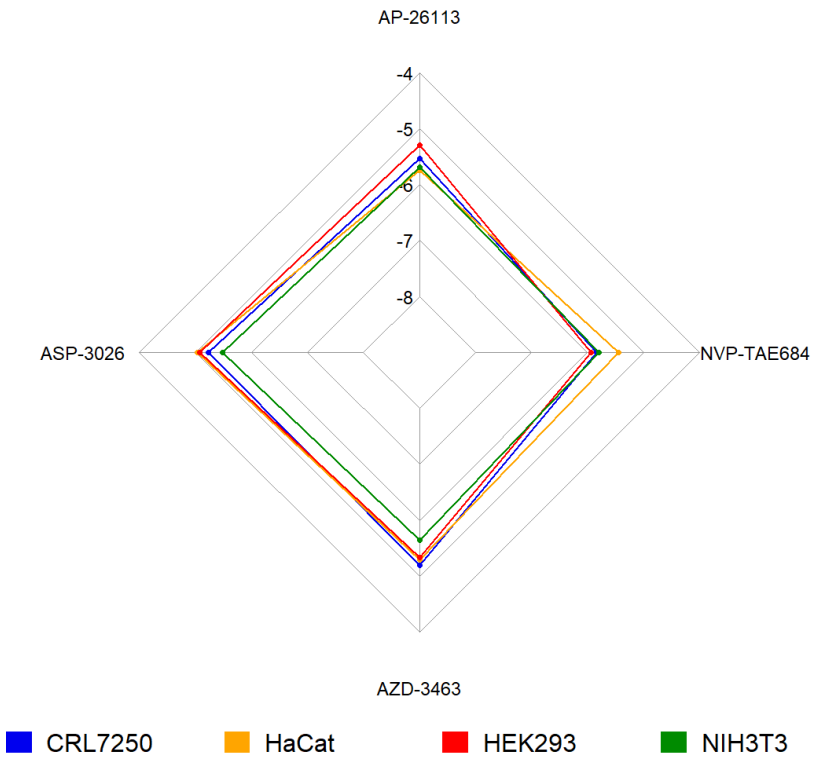
Proteasome



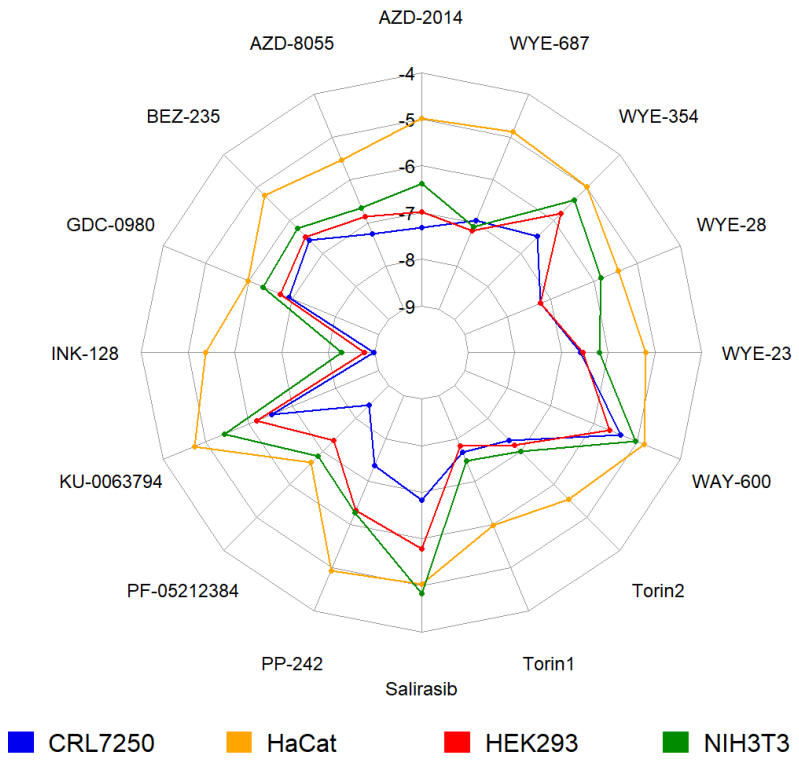
CDK



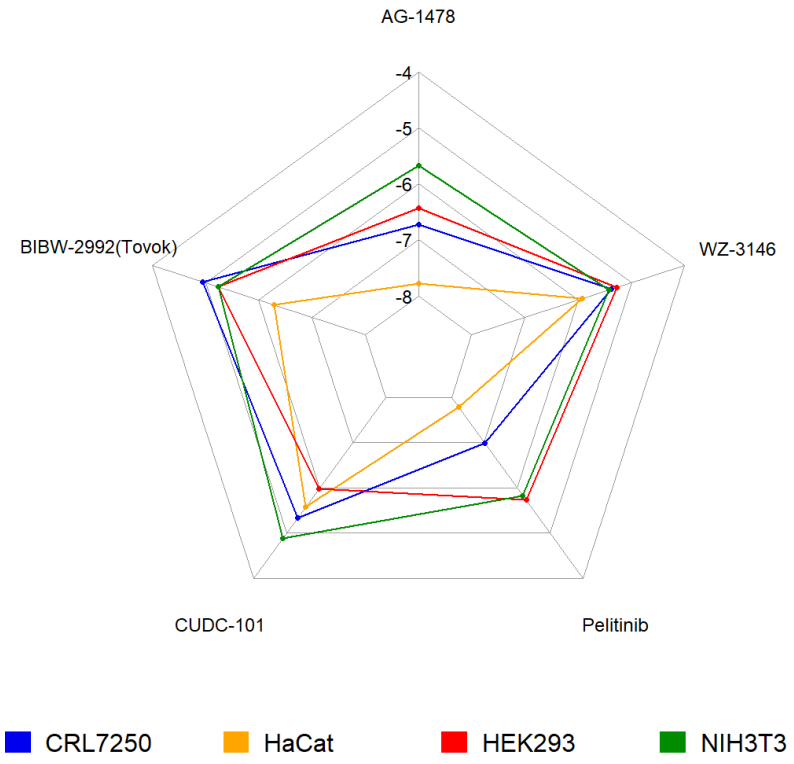
ALK


AP-26113
-4
-5
-6
-7
-8
ASP-3026
NVP-TAE684
AZD-3463
CRL7250
HaCat
HEK293
NIH3T3

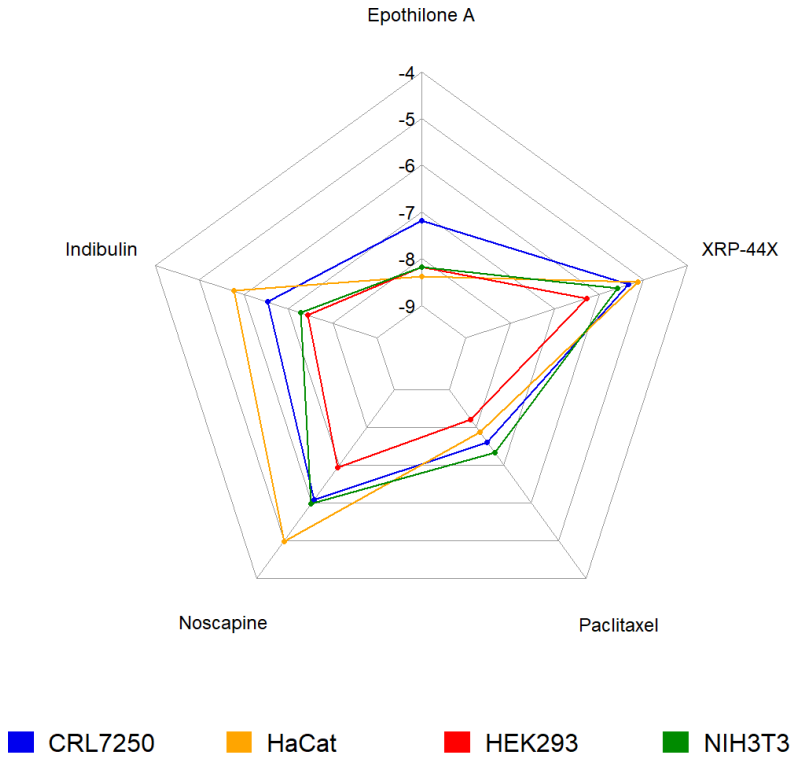
mTOR


AZD-2014
WYE-687
WYE-354
WYE-28
WYE-23
WAY-600
Torin2
Torin1
Salirasib
PP-242
PF-05212384
KU-0063794
INK-128
GDC-0980
BEZ-235
AZD-8055
-4
-5
-6
-7
-8
-9
CRL7250
HaCat
HEK293
NIH3T3

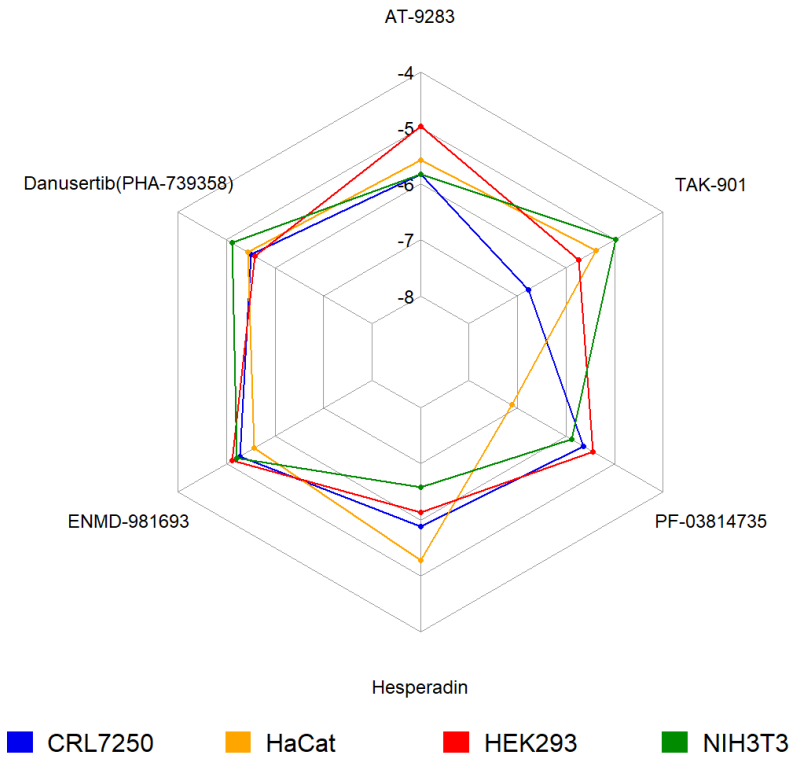
EGFR



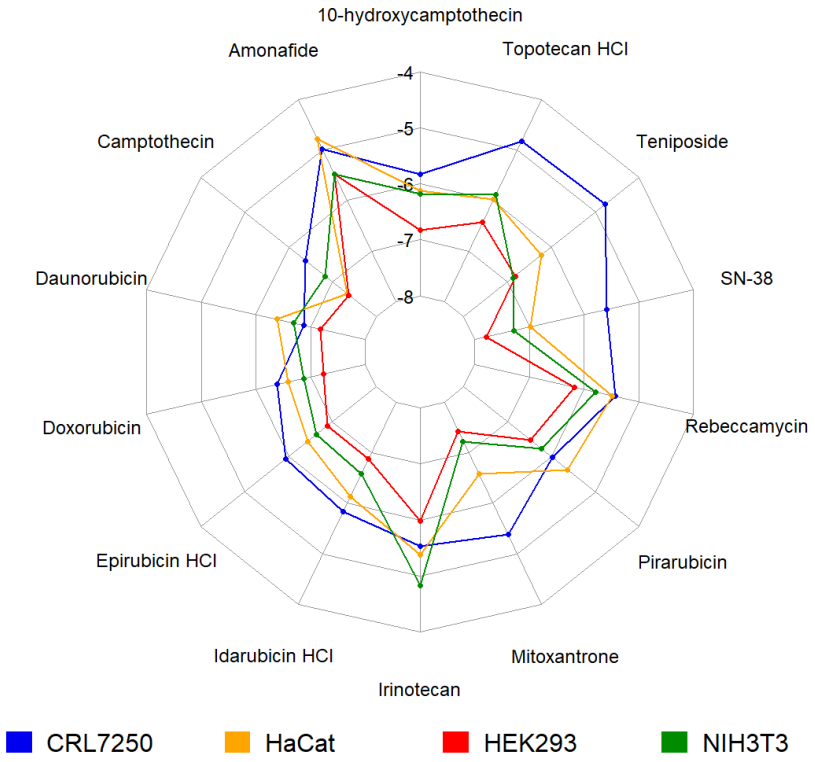
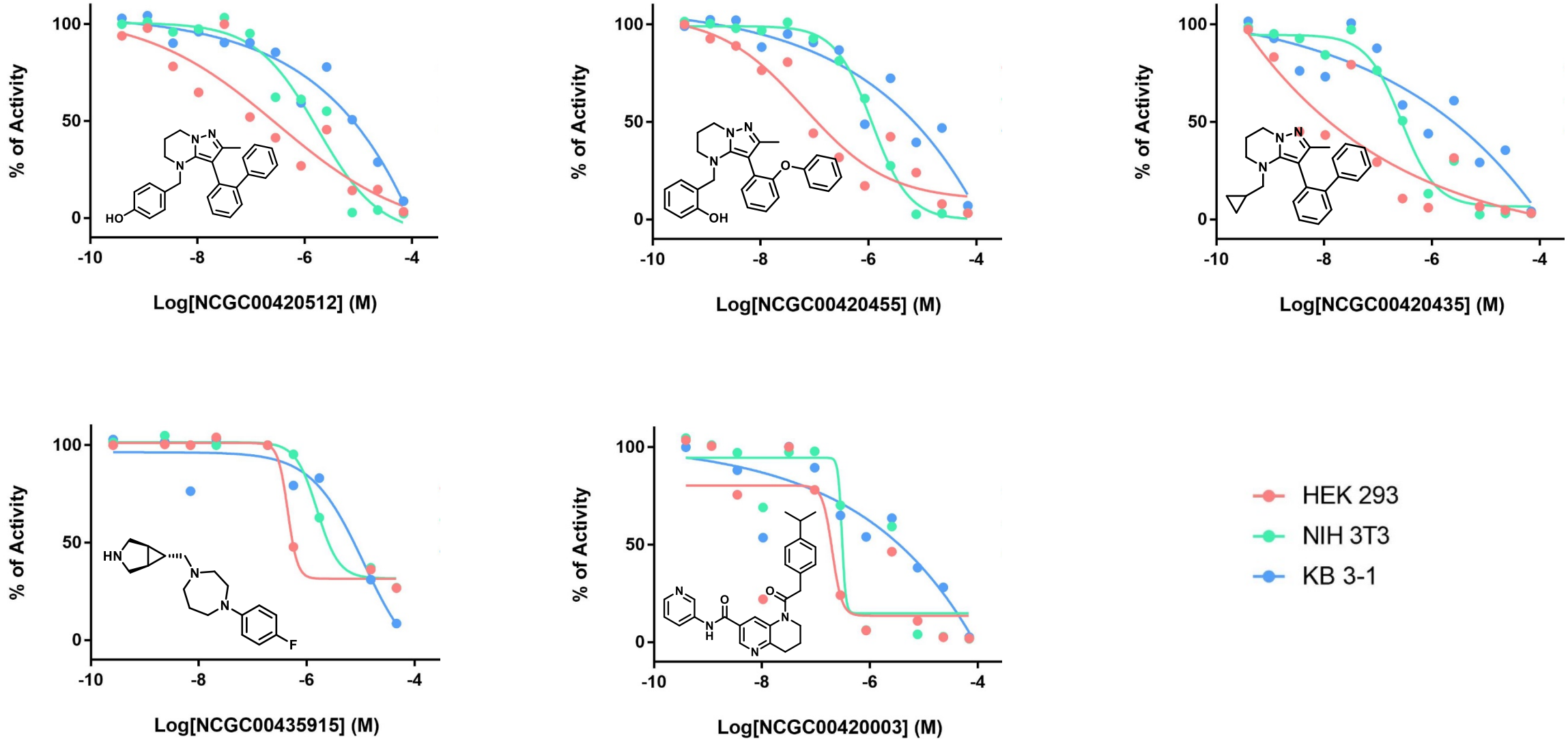
Tubulin



Aurora


AT-9283
TAK-901
PF-03814735
Hesperadin
ENMD-981693
Danusertib(PHA-739358)
-4
-5
-6
-7
-8
CRL7250
HaCat
HEK293
NIH3T3

Topo


10-hydroxycamptothecin
Topotecan HCl
Teniposide
SN-38
Rebeccamycin
Pirarubicin
Mitoxantrone
Irinotecan
Idarubicin HCl
Epirubicin HCl
Doxorubicin
Daunorubicin
Camptothecin
Amonafide
-4
-5
-6
-7
-8
CRL7250
HaCat
HEK293
NIH3T3

Supplementary Figure 2. Hits selective to only normal cell lines