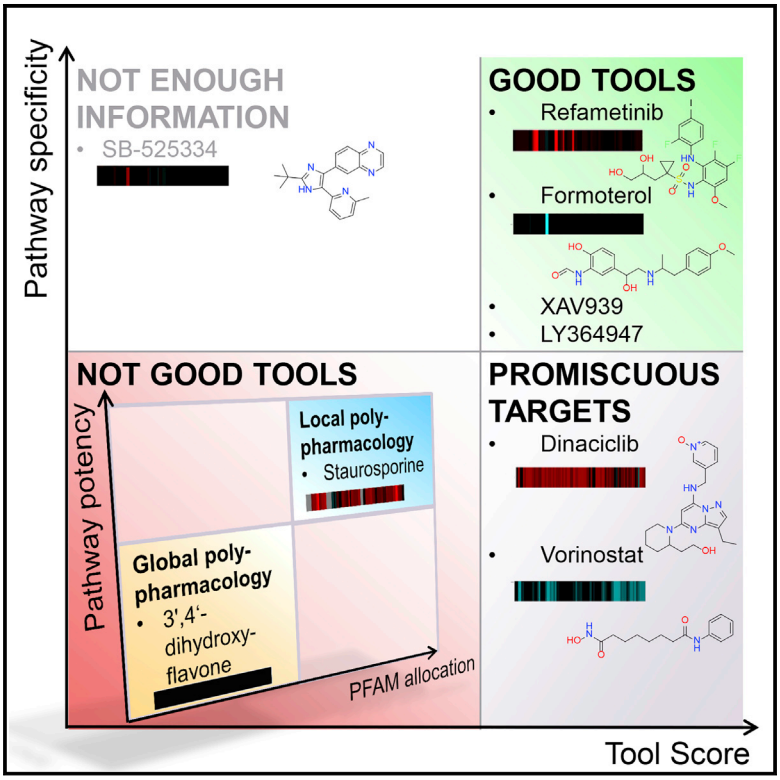
Resource

Cell Chemical Biology

Evidence-Based and Quantitative Prioritization of Tool Compounds in Phenotypic Drug Discovery

Graphical Abstract



Highlights

- Assertions about tool compound confidence strength and selectivity are automated
- A tool score (TS) inferred from assertions prioritizes probes from large databases
- TS-prioritized compounds display phenotypic selectivity across pathway assays
- Code for computing TS is available online, www.github.com/novartis

Authors

Yuan Wang, Allen Cornett, Fred J. King, ..., C. Gregory Paris, Gregory McAllister, Jeremy L. Jenkins

Correspondence

yuan-2.wang@novartis.com (Y.W.), jeremy.jenkins@novartis.com (J.L.J.)

In Brief

Wang et al. create a metric to quantify tool compound confidence from integrated, large-scale heterogeneous data sources to automate the prioritization of tool compounds for screening with higher phenotypic selectivity.

Wang et al., 2016, Cell Chemical Biology *23*, 862–874
July 21, 2016
http://dx.doi.org/10.1016/j.chembiol.2016.05.016

CellPress

Please cite this article as: Wang et al., Evidence-Based and Quantitative Prioritization of Tool Compounds in Phenotypic Drug Discovery, Cell Chemical Biology (2016), http://dx.doi.org/10.1016/j.chembiol.2016.05.016

Evidence-Based and Quantitative Prioritization of Tool Compounds in Phenotypic Drug Discovery

Yuan Wang,[1,*] Allen Cornett,[1] Fred J. King,[3] Yi Mao,[2] Florian Nigsch,[4] C. Gregory Paris,[1] Gregory McAllister,[1] and Jeremy L. Jenkins[1,*]

[1]Novartis Institutes for BioMedical Research Inc., 250 Massachusetts Avenue, Cambridge, MA 02139, USA
[2]Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
[3]Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA
[4]Novartis Institutes for BioMedical Research, Novartis Pharma AG, Novartis Campus, Basel 4056, Switzerland
*Correspondence: yuan-2.wang@novartis.com (Y.W.), jeremy.jenkins@novartis.com (J.L.J.)
http://dx.doi.org/10.1016/j.chembiol.2016.05.016

SUMMARY

The use of potent and selective chemical tools with well-defined targets can help elucidate biological processes driving phenotypes in phenotypic screens. However, identification of selective compounds en masse to create targeted screening sets is non-trivial. A systematic approach is needed to prioritize probes, which prevents the repeated use of published but unselective compounds. Here we performed a meta-analysis of integrated large-scale, heterogeneous bioactivity data to create an evidence-based, quantitative metric to systematically rank tool compounds for targets. Our tool score (TS) was then tested on hundreds of compounds by assessing their activity profiles in a panel of 41 cell-based pathway assays. We demonstrate that high-TS tools show more reliably selective phenotypic profiles than lower-TS compounds. Additionally we highlight frequently tested compounds that are non-selective tools and distinguish target family polypharmacology from cross-family promiscuity. TS can therefore be used to prioritize compounds from heterogeneous databases for phenotypic screening.

INTRODUCTION

Phenotypic screening, as an alternative to target-based approaches, has proved its value in the discovery of over half of small-molecule first-in-class new molecular entities (Moffat et al., 2014; Swinney, 2013). Phenotypic assays are widely used for target and lead discovery, as well as to profile for compound mechanism of action (MOA) (Feng et al., 2009; Hart, 2005; Hopkins, 2008; Jones and Diamond, 2007; King et al., 2009; Lee et al., 2012; Moffat et al., 2014; Schirle and Jenkins, 2015; Swinney, 2013). For example, reporter gene assays (RGAs) are a representative type of high-throughput, low-cost phenotypic screening technology whereby pathway response to compound treatment can be observed in reporter-transfected cells typically driving expression of a fluorescent protein (King et al., 2009). Phenotypic screening is powerful in connecting chemistry to pathway modulation, but presents a challenge of following up hits with target identification. For example, the compound XAV939 was first identified in a Wnt pathway screen, and its mechanism of tankyrase inhibition and prevention of axin degradation was later elucidated in a separate workflow using chemical proteomics (Figure 1A) (Huang et al., 2009). Similarly, a phenotypic screen identified an inhibitor of the Smoothened-dependent Hedgehog pathway, which later led to the identification of GPR39 as the target (Bassilana et al., 2014). To make phenotypic drug discovery more efficient and systematic in the follow-up phase, one can use small molecules with well-understood MOAs as chemical probes to observe the phenotypic effect of target modulation (Arrowsmith et al., 2015; Bunnage et al., 2013; Eggert, 2013; Frye, 2010; Uitdehaag et al., 2012; Workman and Collins, 2010), either individually or in designed sets. A high-quality tool compound can also play a critical role in a drug discovery project as a positive control molecule, from optimizing signal-to-noise in screening assay development to preclinical in vivo target validation.

Criteria that define the credibility of compounds as target probes has been proposed previously: (1) target engagement, i.e., potency and selectivity on target both in cell-free and cell-based assays; (2) cell permeability, i.e., exposure at site of action; (3) proven utility as a probe, i.e., phenotypic relevance via a demonstrated proximal biomarker; and (4) identity of the active species and availability (Arrowsmith et al., 2015; Bunnage et al., 2013; Eggert, 2013; Frye, 2010; Uitdehaag et al., 2012; Workman and Collins, 2010). While these qualities are ideal for probes, the requisite data are difficult to obtain and available for very few compounds relative to all possible tool compounds in large databases that could be employed in a screening setting. More often probes are obtained by researchers from vendor catalogs, even though many well-known purchasable compounds may not be as selective as they appear based on the catalog description. Although recently an international collaboration was initiated to collect qualifying tool compounds from researchers in the chemical biology community (Arrowsmith et al., 2015), the practical aspect of prioritizing existing tool compounds from large, diverse bioactivity databases has not been sufficiently addressed.

One major difficulty for tool compound selection is compound polypharmacology. It is well established that drugs often bind to more than one molecular target (Gregori-Puigjane et al., 2012; Hopkins, 2009; Roth et al., 2004), a property that emerges prominently in large-scale, integrated bioactivity databases.

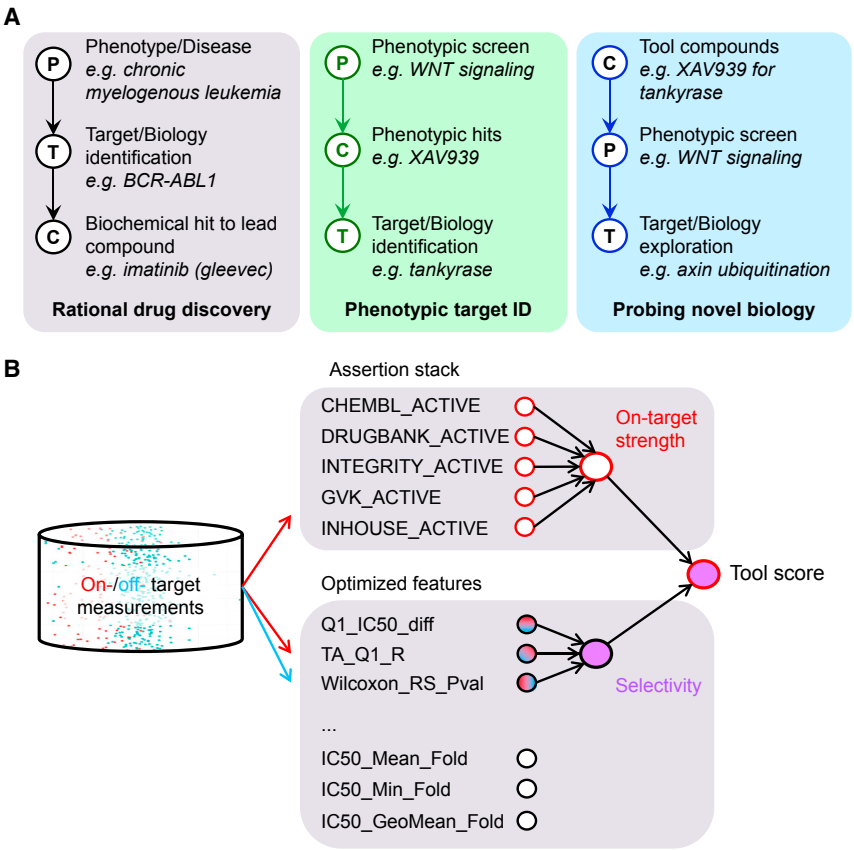


Figure 1. Validating Target Hypotheses in Phenotypic Screens Using Tool Compounds

(A) Rational drug design (Phenotype→Target→Compound, left), phenotypic hits to target identification (P→C→T, middle) and target validation/biological exploration with tool compounds (C→P→T, right). Rational drug design workflow can be exemplified by the discovery of Gleevec (Druker et al., 1996; Gambacorti-Passerini, 2008), whereas the work of Huang et al. (2009) demonstrated the power of phenotypic screens. The authors found XAV939 in a WNT signaling phenotypic screen, identified its target tankryase, then used it as a tool compound to probe tankryase biology, such as axin ubiquitination inhibition. Phenotypic hits (green "C") and tool compounds (blue "C") may be the same, but not necessarily: the latter have more strict selection criteria.

(B) Schematic of the workflow for TS calculation for prioritization of tool compounds.

Extensive polypharmacology limits a compound's utility as a targeted tool. The known polypharmacology in bioactivity databases is unfortunately a function of both the compound's target selectivity and the extent to which it has been profiled. Many studies have investigated the selectivity of small molecules against a panel of targets and the quantification thereof (Graczyk, 2007; Jester et al., 2012; Uitdehaag et al., 2012; Uitdehaag and Zaman, 2011). In one study that applied a selectivity entropy method, Uitdehaag et al. (2012) collected representative tool compounds for well-profiled kinase targets and evaluated their selectivity across panels of more than 200 targets. However, it is impractical to profile many thousands of compounds in this manner, and published profiling panels are so far all limited to a biased, subgenomic set of targets. In addition, the current selectivity quantification methods are usually implemented on a single-result data type (e.g., EC_{50}, K_i) that describes the potency of each compound for each target in the panel (Graczyk, 2007; Jester et al., 2012; Uitdehaag et al., 2012; Uitdehaag and Zaman, 2011), but it is not uncommon to have evidence from more than one laboratory on the activity of a certain compound-target pair. To bring data from multiple potency measurements together, Kalliokoski and colleagues have statistically analyzed the SD and correlation of heterogeneous K_i and IC_{50} data and have proposed a numerical conversion to unity (Kalliokoski et al., 2013; Kramer et al., 2012), while Tang et al. (2014) have designed the KIBA score to integrate multiple bioactivity types and predict interactions between ligand and kinase targets. Yet if any source of the data should contain errors due to assay conditions, target annotation, or result type (Tiikkainen et al., 2013), the mistake might be propagated to the final quantification. Using such quantification to prioritize tool compounds will lead to suboptimal candidates.

Computational sciences have increasingly affected the discipline of chemical biology. Machine learning methods such as naive Bayesian models and Similarity Ensemble Approach have been proved as valuable and efficient methods for virtual screening of compounds, in silico target prediction, and side-effect prediction (Hopkins, 2008; Keiser et al., 2009; Lounkine et al., 2012; Nidhi et al., 2006; Scheiber et al., 2009). For example, random forest models have been successfully used for quantitative structure-activity relationships and compound activity predictions (Cumming et al., 2013; Riniker et al., 2013, 2014; Sakiyama, 2009). In addition, the semantic web and triple stores are commonly used to assess and infer the associations of drugs and their targets, and/or biological and clinical phenotypes from existing connections (Azzaoui et al., 2013; Bolling et al., 2014; Chen et al., 2012a, 2012b; Riazanov et al., 2013; Ruttenberg et al., 2007; Samwald et al., 2011; Shadbolt et al., 2006; Wild et al., 2012; Williams et al., 2012). However, the accuracy of all of these computational approaches relies on the quality of reference compound data, and not all target assertions should be treated equally: First, the number of bioactivity measurements is highly variable. Bioactivity between a compound and a target is not sometimes composed of one IC_{50}/EC_{50}/K_i number value measured by one data source, but also quite often composed of many heterogeneous measurements from diverse sources (shown in Figure 2A). Second, for the same target, experimental results using different assay technologies, in different biological systems, in different laboratories, can vary dramatically, sometimes often more than one to three orders of magnitude. Third, sources such as DrugBank (Law et al., 2014) or Thomson Reuters's Integrity

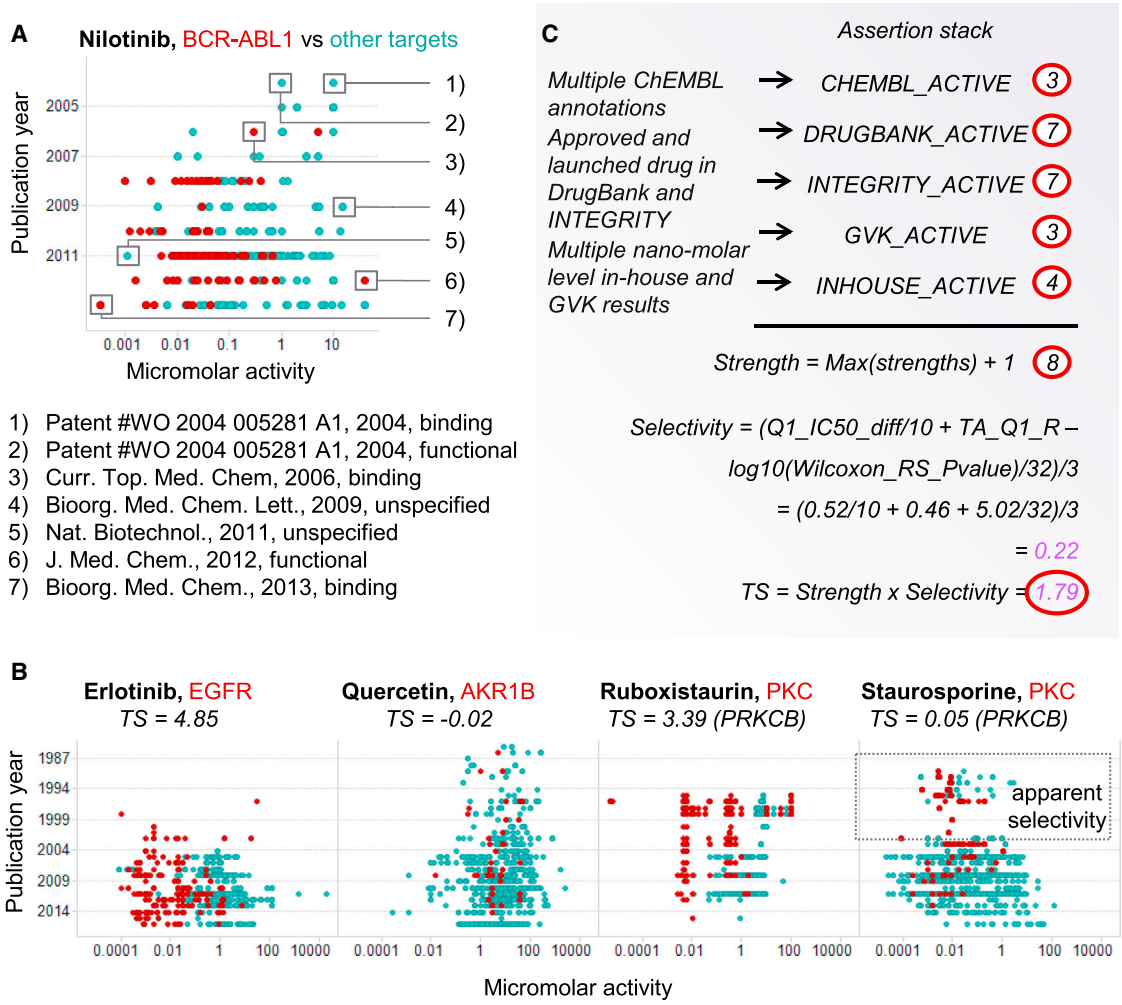


Figure 2. Integration of Compound Activity and Selectivity Knowledge Accumulated Over Time

Mining data from all available experiments can provide insights into the concentration-dependent activity and selectivity footprint of the compound in different types of assays from multiple laboratories, which can be more stable than one experiment over a single panel of targets.

(A and B) Visualizing compound activity data against on-targets (red) and off-targets (blue). Compounds such as nilotinib and erlotinib have a larger separation between on-target and off-target potency measurements than non-selective compounds such as quercetin. Selective tool compounds can help elucidate on-target biology in phenotypic assays; however, high doses in phenotypic screens (e.g., 10 μM) can render good chemical probes ineffective because multiple off-targets can be affected as well as the on-targets. Another pair of examples are two protein kinase C (PKC) inhibitors, staurosporine and ruboxistaurin, with different selectivity profiles. With limited profiling over time (pre-2003), staurosporine was often used as a probe to target PKC. However, accumulated testing in publications has proved it to be unselective with respect to multiple kinases. On the other hand, ruboxistaurin has demonstrated selectivity for PKC, especially the β isoform.

(C) TS calculation example on nilotinib. As described in Table S2, strength assertion depends on individual assertions describing activity from each data source, "CHEMBL_ACTIVE," "DRUGBANK_ACTIVE," "INTEGRITY_ACTIVE," "GVK_ACTIVE," and "INHOUSE_ACTIVE." Selectivity depends on three prioritized features calculated using quantitative data. Public-domain data for this calculation are included in Data S1 (example_inputdata_toolscore_calculation.csv).

(integrity.thomson-pharma.com) have curated qualitative target annotations and should also be weighted as evidence for the compound-target activity.

In this study we integrated large-scale heterogeneous data from many data sources and investigated critical features that drive the tool utility of a compound for a target. We quantified this utility as the tool score (TS), which is a computational summary of diverse activity and selectivity information that does not require data from a target panel. Next we assessed the TS with specific phenotypic observations for 384 compounds tested in a panel of 41 RGAs representing different cellular pathways and identified selective tool compounds for eight targets. We also identified 68 non-selective compounds that were frequently published and whose polypharmacology could be computationally classified as within a target family (local polypharmacology) or general promiscuity (global polypharmacology). We found that TS-prioritized tool compounds had not only on-target potency and selectivity, but also exhibited phenotypic specificity, which facilitates the molecular interpretability of phenotypic screening.

Table 1. Number of Compounds Analyzed for Each Target Gene and Tool Compound Recommendations

Target	Description	No. of Compounds	Recommended Tool Compounds	Not Recommended Tool Compounds
ADRB2	β2-adrenergic receptor	40	formoterol, salmeterol	terbutaline, ritodrine
CDK9	cyclin-dependent kinase 9	26	inhibiting target inhibits multiple RGA pathways	
GSK3B	glycogen synthase kinase 3 β	34	CHIR-99021	
HDAC1	histone deacetylase 1	34	inhibiting target activates multiple RGA pathways	
JAK2	Janus kinase 2	65	ruxolitinib	
MAP2K1	mitogen-activated protein kinase kinase 1 (MEK1)	25	refametinib, SCHEMBL346145	CHEMBL95002, SCHEMBL1228956
PRKCA	protein kinase C α	79	sotrastaurin, ruboxistaurin	staurosporine
PRKCB	protein kinase C β	7		
SRC	proto-oncogene tyrosine-protein kinase	34	SCHEMBL8108556	
TGFBR1	transforming growth factor β receptor 1	18	LY364974, SCHEMBL3799018, SCHEMBL3798498	SCHEMBL2239687
TNKS2	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase 2	22	XAV939, HMS1899N20, CHEMBL2419697	3′,4′-dihydroxyflavone
Total	8 kinases, 3 other enzymes	384		

For each target we show the number of compounds analyzed in the reporter gene assay (RGA). Using this analysis we recommend representative tool compounds for targets that our RGA panel can distinguish. In addition to well-tested compounds, we have also identified novel tools (underlined) for MAP2K1, SRC, TGFBR1, and TNKS2. We have also shown that commonly used compounds such as terbutaline, staurosporine, and 3′,4′-dihydroxyflavone (more such compounds in Table S5) were not good tool compounds.

RESULTS

Meta-Analysis of Heterogeneous Bioactivity Data to Calculate a Tool Score

To obtain a global perspective of compound bioactivity, we integrated several data sources (Figure 2A and Supplemental Experimental Procedures). From these many sources we wanted to understand generally the potency, selectivity, and other evidence supporting a compound for a given target across qualitative target assertions and diverse quantitative in vitro assay data types. We first defined a metric called strength, to represent a combination of potency (where available), amount of evidence, and confidence in data sources (as judged by the authors' experience), which is intended to automate human expertise. Strength can be used to quantify confidence in the in vitro target of a compound, but not to quantify confidence in its selectivity as a cellular probe.

With regard to selectivity, the issue of multiple inconsistent measurements was a substantial challenge. Therefore we launched a systematic evaluation of candidate features that describe compound selectivity. Supervised machine learning prioritized three selectivity features that distinguished selective compounds more accurately than those comparing fold IC_{50}s (Table S4). We therefore implemented these features in the selectivity component of the TS.

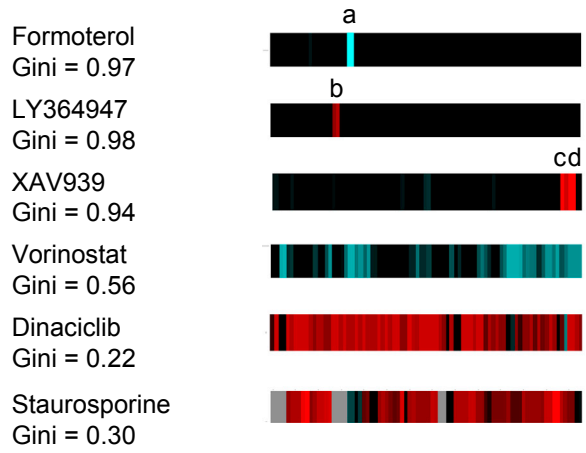
The TS is a multilayered information model between a compound and a target, whose layers are tuned with empirical knowledge, heuristics, and machine learning (Figure 1B). In Figure 2C we show an example TS calculation workflow for nilotinib and ABL1. We have shown that TS performed better at distinguishing selective and non-selective tool compounds compared with conventional fold comparisons using various numeric aggregations on IC_{50} measurements (Figure S4).

Corresponding Phenotypic Selectivity in RGA Panel and Tool Compound Utility

TS can be computed for a large number of compounds, and we believe it can serve as a guideline in prioritizing multiple chemical tool compounds in a phenotypic experiment for the target to be probed. To confirm this, we next investigated how the TS metric reflects utility in our RGAs. In total, we selected 384 compounds annotated with 11 targets with a range of TS values (Table 1). Some compounds triggered agonistic or inhibitory responses in one pathway reporter, others in multiple pathway reporters (see Figure 3A, example compounds).

As an example, formoterol, a long-acting selective β2-adrenoreceptor agonist, is a strong cyclic AMP (cAMP) signaling agonist with a DR_{50} of 1.0 (i.e., active at all doses tested starting from the lowest dose) at both 12 and 24 hr (Figure 3A). This observation is consistent with previous experiments on β2 agonists (Tsvetanova and von Zastrow, 2014; Violin et al., 2008). Similarly, LY364947, a short-acting transforming growth factor β (TGF-β) receptor inhibitor (Kano et al., 2007), inhibited TGF-β signaling at high doses with DR_{50} of −0.42 at 12 hr and −0.55 at 24 hr. XAV939, a selective tankyrase inhibitor (Huang et al., 2009), specifically inhibited the Wnt signaling pathway but not cAMP signaling, TGF-β signaling, or any of the other pathways. To quantify phenotypic selectivity we calculated the Gini coefficient for compound activity for the entire panel of RGAs (Graczyk, 2007): the higher the pathway selectivity, the higher the Gini coefficient. The above compounds each had high Gini coefficients because they only triggered pathway responses specific to their biochemical targets: in other words, ideal tool compounds. By contrast,

A



	Reporters	Pathways
a	CRE-luc, 12h, 24h	cyclic AMP signaling
b	CAGA12/luc (+TGFb cDNA), 12h, 24h	TGFβ signaling
c	STF-luc (+Wnt3a condition media), 12h, 24h	WNT signaling
d	STF-luc (+Rspo2 cDNA), 12h, 24h	WNT signaling

B

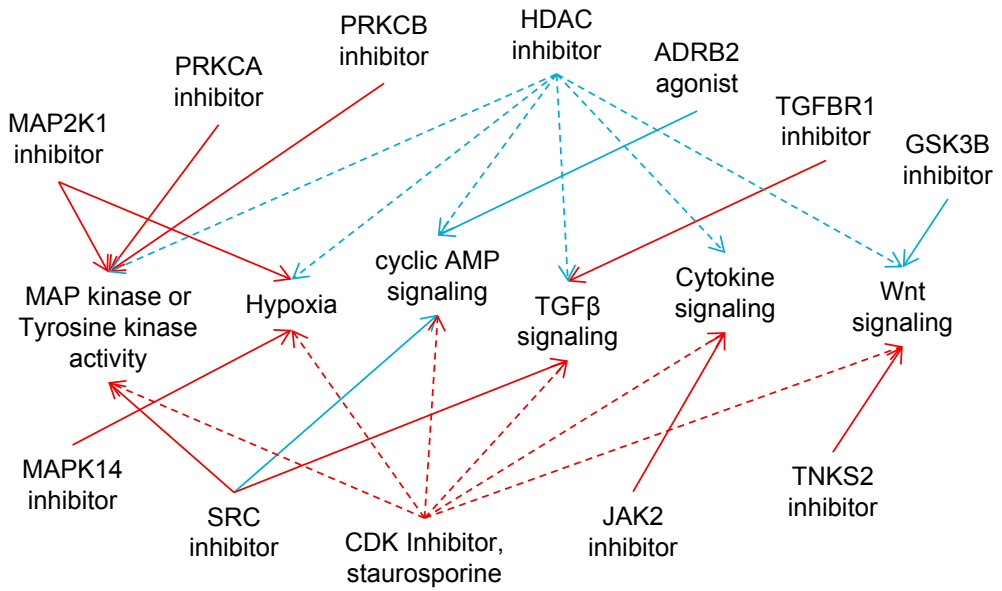


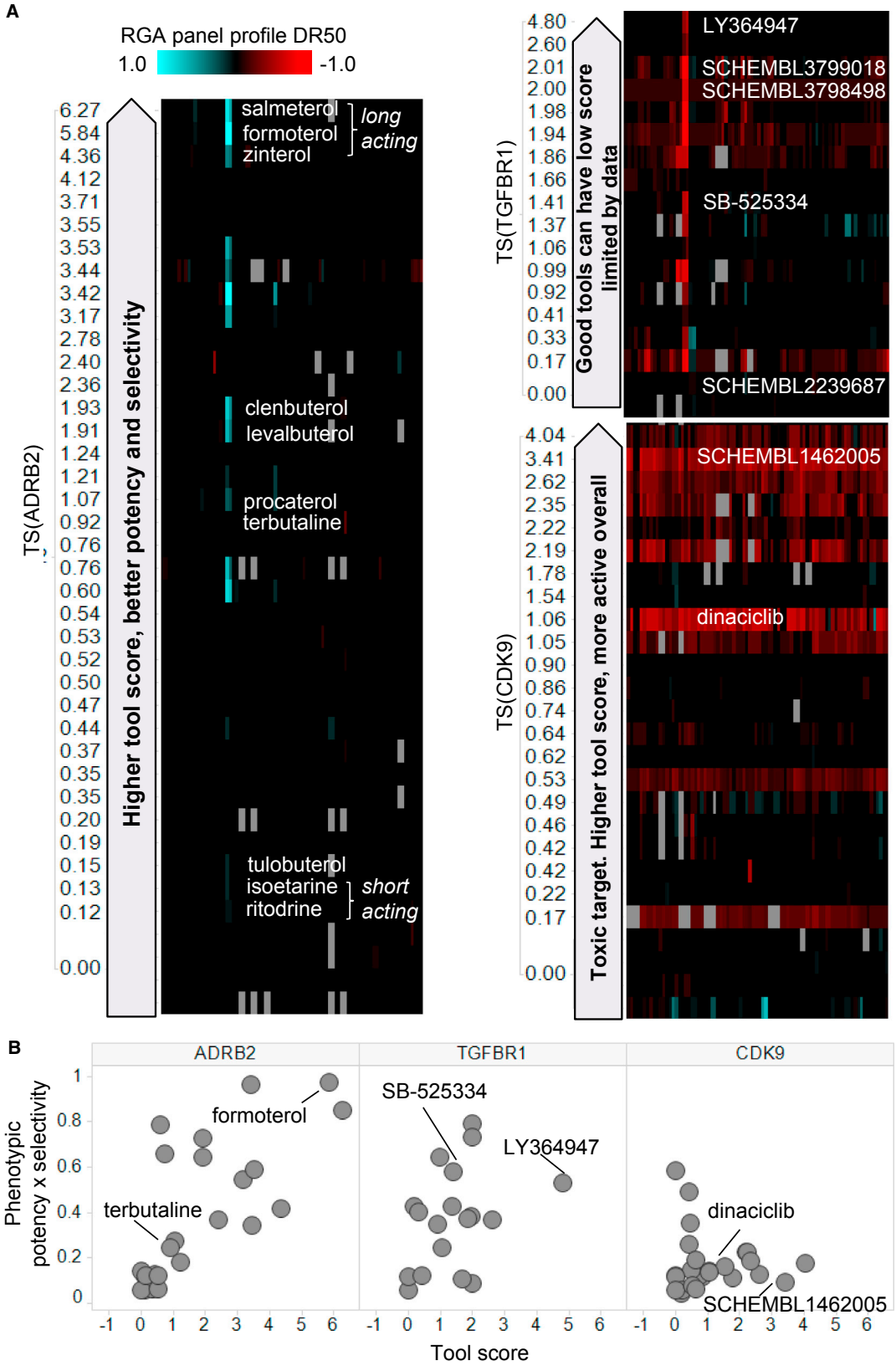
Figure 3. Connecting On-Target Biology with Specific Phenotypic Profiles

(A) RGA panel profiles of seven compounds and their Gini coefficients. For each compound, 12-hr and 24-hr DR_{50}s of the 41 RGAs are plotted as a heatmap (DR_{50} = 1.0, strong agonist; DR_{50} = 0, inactive; DR_{50} = −1.0, strong antagonist; gray, no data). Formoterol is a selective β2-agonist (TS = 5.84) and specifically induces cyclic AMP pathway reporter. LY364947 is a selective TGF-β receptor inhibitor and inhibits TGF-β signaling. Likewise the selective tankyrase inhibitor, XAV939 (TS = 5.27), selectively inhibits Wnt signaling. By contrast, the HDAC inhibitor vorinostat, CDK inhibitor dinaciclib, and pan-kinase inhibitor staurosporine regulate the upstream biology of multiple RGAs in the panel and have low phenotypic selectivity, which can be reflected by their Gini coefficients.

(B) Summary of targets discussed in this study and pathways found to be regulated by these targets in the RGA panel.

vorinostat (pan-HDAC inhibitor) (Dokmanovic et al., 2007), dinaciclib (pan-CDK inhibitor) (Pei and Xiong, 2005), and staurosporine (pan-kinase inhibitor) (Ruegg and Burgess, 1989) activated or inhibited multiple pathways, leading to low Gini coefficients. They did not have pathway-specific profiles, suggesting target polypharmacology and/or phenotypic pleiotropy (represented by dotted lines in Figure 3B), a property that would lead to frequent hitting across phenotypic screens.

By applying the TS to represent compound-target potency and selectivity, DR_{50} as phenotypic potency metric, and Gini coefficient as phenotypic specificity metric, we were able to systematically compare the potency and selectivity in both target dimension and pathway dimension and to study their correlation. We observed that overall compounds having high TS were more likely to have high phenotypic potency and selectivity if the target was directly upstream of the pathway readout



(legend on next page)

(Figures 3B and 4). For example, compounds with higher TS on ADRB2 were more likely to be phenotypically potent, selective, and long acting (Figure 4A). By contrast, if the target was upstream of multiple pathways then even compounds having high TS could not correlate to high pathway selectivity. For example, higher TS for CDK9 would likely induce apoptosis (Dai and Grant, 2003; Pei and Xiong, 2005; Shapiro, 2006) and an increased inhibitory effect on almost all our RGAs (Figure 4A). Compounds having low TS were less likely to be potent and selective in pathways reflecting on-target biology, but there were exceptions (dots at upper-left quadrant of plots in Figure 4B). For example, some compounds having low TS for TGFBR1 also had high potency and selectivity against the TGF-β signaling pathway (Figure 4). This is because we evaluate TS based on known bioactivity data and poorly profiled compounds cannot by definition have high TS, even if they are potentially good tools. In other words, a data-driven prioritization of probes will be more adept at penalizing bad tool compounds than detecting good tool compounds with little historical data, a trade-off that may be acceptable in the context of designing sets with interpretable outcomes in phenotypic screening. This is evident in the lack of high-scoring tool compounds with low Gini coefficients (Figure S5).

The same analysis was then extended to a total of 1,457 previously tested in-house compounds with high potency (absolute $DR_{50} \geq 0.5$) to study their selectivity. RGA profiles were consistent with what we observed for our handpicked 384 compounds in their relationship to TS (Figure S6). We have also included CHIR-99021, which was nominated in the Chemical Probes portal (Arrowsmith et al., 2015) as a GSK3B tool compound, in our analysis. It has been confirmed to be a good tool compound with TS of 1.94 and was selectively upregulating Wnt signaling in our RGA panel (Gini: 0.70). Overall, these results support the ability to use integrated but disparate bioactivity to help prioritize tool compounds where data is sufficient.

Comparison of Selective and Non-selective Tool Compounds

In Table 1 we list our recommendations regarding fit and unfit tool compounds for eight targets. In addition to well-profiled drugs, compounds that were less measured but exhibited decent selectivity could also be useful tool compounds. We have also tested compounds having similar structures but different TS (Figure 5). Interestingly, well-tested 3′,4′-dihydroxyflavone did not appear to have detectable cellular activity in any RGA assay, although it was shown to bind TNKS2 (Narwal et al., 2013). XAV939 and CHEMBL562310 were very different in structure but both were good TNKS2 tool compounds and selectively inhibited Wnt signaling. Similarly, SCHEMBL346145 and refametinib were both decent MAP2K1 tool compounds. However, CHEMBL95002, a close analog of SCHEMBL346145, was more promiscuous. Using TS we can identify selective and inactive/non-selective analogs from the same scaffold to better elucidate target hypotheses.

Importantly, many non-selective compounds are used improperly as target probes in published studies that the TS metric would help to avoid. Having applied the TS to all compounds in our database tested against at least 100 targets, we propose a list of compounds that are unsuitable for phenotypic tools (Table S5). We observed that these 68 compounds can be divided into two categories, target family-specific (such as pan-kinase inhibitors) and generally promiscuous compounds. To distinguish the two categories we have defined PFAM allocation similarly to our definition of target allocation (Supplemental Experimental Procedures). Compounds having high primary PFAM allocation are more specific to certain conserved PFAM domains (Bateman et al., 2004; Finn et al., 2014), such as the kinase domain, whereas compounds having low primary PFAM allocation were reported to be active against multiple types of targets. Examples of the former category include staurosporine (active against many kinases) and chlorpromazine (active against many GPCRs), while the latter include quercetin and resveratrol. In Figure 6 we show the phenotypic potency of these less attractive tool compounds: Pan-kinase inhibitors tend to have high potency and low selectivity in the RGA panel, whereas pan-GPCR binders and general promiscuous compounds have weak phenotypic activity in the panel overall. We posit that while PFAM-selective compounds may have utility in investigating target class polypharmacology (Gujral et al., 2014), PFAM-unselective compounds have little explanatory value in cellular assays.

DISCUSSION

In this study we have systematically studied important criteria of good tool compounds: potency, selectivity, and phenotypic correspondence. We derived the TS to systematically evaluate integrated heterogeneous compound bioactivity data and to assess the utility of a compound as chemical probe for a given target in terms of potency and selectivity. Using TS we have collected and categorized fit and unfit tool compounds, and also prospectively prioritized tool compounds and validated their pathway specificity.

Understanding Heterogeneous Potency and Selectivity Information is Non-trivial

Contrary to the common belief that it is possible to use mean, geometric mean, or minimum to aggregate multiple measurements and derive good understanding of the potency and selectivity of a compound, the complexity and heterogeneity of

Figure 4. Compound-Target Tool Scores and Corresponding Phenotypic RGA Panel Profiles

(A) For each compound, 12-hr and 24-hr DR_{50}s of the 41 RGAs are plotted next to each other as a heatmap next to its TS ($DR_{50} = 1.0$, strong agonist; $DR_{50} = 0$, inactive; $DR_{50} = -1.0$, strong antagonist; gray, no data) Compounds having high TS on ADRB2 and TGFBR1 were more likely to be potent and selective against the corresponding pathway reporter assay, while compounds having higher TS on CDK9 were more potent on almost all tested assays. Among ADRB2 agonists, long-acting compounds such as formoterol had higher TS and stronger phenotypic selectivity than short-acting ones such as ritodrine. Some TGFBR1 inhibitors having low TS (due to incomplete knowledge) also had high phenotypic potency/specificity (e.g., SB-525334, 33 measurements in ChEMBL versus LY364947, 600 measurements in ChEMBL).

(B) Correlation between tool score and phenotypic potency/selectivity for ADRB2, CDK9, and TGFBR1 (more in Figure S5).

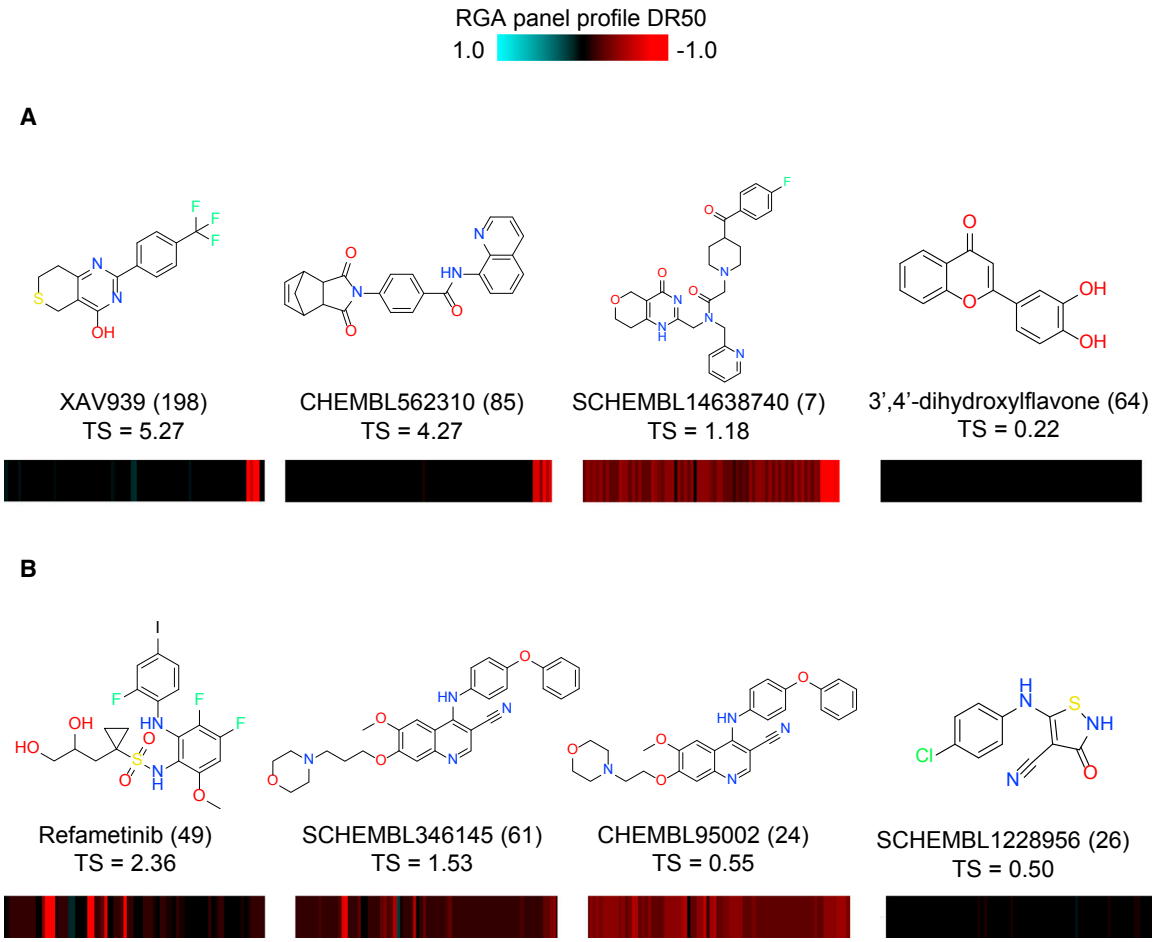


Figure 5. Structure, Bioactivity Measurements, TS, and Phenotypic Selectivity

For TNKS2 (A) and MAP2K1 (B), tool compounds having different activity/selectivity are shown. Structurally similar compounds such as SCHEMBL346145 and CHEMBL95002 can have different selectivity and behave differently in pathway RGAs. Using TS we can identify selective tool compounds as well as their inactive/non-selective analogs as control compounds in a phenotypic assay to better elucidate the target hypothesis. The number of measured activities for each compound is shown in parentheses. The well-tested 3′,4′-dihydroxyflavone did not appear to have detectable cellular activity in the RGA panel. TS does not favor popular compounds, but thorough profiling data helps confirm selectivity. In addition, TS avoids the effect of one-off measurements, and captures the overall selectivity despite the apparent lack of IC_{50} fold difference (Figure S7).

bioactivity data, as shown in Figures 2A and 2B (plus qualitative annotation not shown in the figure), is beyond simple arithmetic aggregation. The use of assertion stack helps us bring together diverse data and infer knowledge of the compound-target activity.

Similarly, the selectivity cannot be treated as ratio or fold of aggregated data because data from different origins may not agree with each other. Our supervised machine learning approach and follow-up comparison of selectivity features have shown that fold IC_{50}s are not the best features to use for classifying selective tool compounds. For example, if we had used fold IC_{50} to pick MAP2K1 tool compounds, we would not have identified SCHEMBL346145 because one-off measurements gave it negative log fold IC_{50} values (Figure S7). Instead, one needs to study the distribution of on-target and off-target measurements and rely on reproducible evidence of their potency differences. The TS is able to collect such information via layers of potency assertion and multiple features of selectivity to produce a relative ranking of compounds among a library of tool compound candidates.

Avoiding the Bias of Popularity and "Target-at-First-Sight"

It is important not to be biased by the first known target of a compound, even if the compound comes from a project designing compounds for this target, and/or has very high potency on this target. Although we did not place a hard cutoff on the number of targets profiled, more profiling would clearly enlarge the knowledge space of the compound and facilitate better decision making using the TS: all other conditions being equal (sufficient evidence of confirmed activity and selectivity), the more measurements a compound has been through, the higher TS it will be assigned. For example, with more than 900 measurements and still maintaining PKC selectivity, ruboxistaurin (Ishii et al., 1996) has a high TS of 3.39 (Figure 2B).

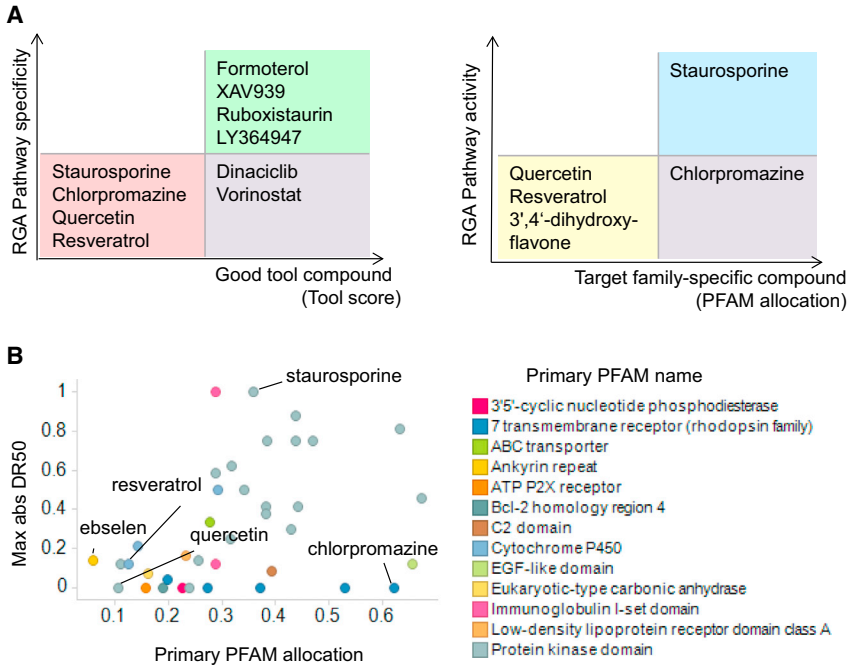


Figure 6. Identification and Classification of Selective and Non-selective Tool Compounds

(A) Example classification of compounds using their TS as well as their PFAM allocation. The former can be used to rank good tool compounds (green box) against undesirable ones (red box) while the latter can be used to distinguish target family-specific (locally promiscuous, blue box) and globally promiscuous compounds (yellow box) among the undesirable tool compounds. Good tool compounds can be used as target validation probes while locally promiscuous compounds can be used to evaluate the phenotypic assay system.

(B) Local promiscuity versus overall RGA potency for less selective tool compounds. Locally promiscuous compounds such as pan-kinase inhibitors have high potency against all reporter gene assays (upper right), whereas compound targeting GPCR and globally promiscuous compounds usually have lower potency over all assays (lower left).

On the other hand, being "famous" and well tested alone will not inflate the compound TS: staurosporine has a low TS on PRKCB (TS = 0.05) because it is well profiled and has many highly potent off-targets (Karaman et al., 2008; Ruegg and Burgess, 1989; Tamaoki et al., 1986) (Figure 2B), which are properly picked up by the TS calculation. The lack of selectivity of staurosporine is also reflected in the RGA panel, with a low Gini of 0.3 (Figure 3A). Another example is terbutaline, an ADRB2 agonist and known drug (Ahlquist, 1976; Jones and Scott, 2011). Terbutaline inhibited high-affinity choline transporter 1 (SLC5A7) activity at the nanomolar level (http://pubchem.ncbi.nlm.nih.gov/bioassay/588401), and affected cell confluence in a functional cAMP assay (Kaya et al., 2012). Additionally it also failed to exhibit a clear cAMP signaling pattern in our RGA panel. As a result, we recommend formoterol instead of terbutaline as chemical probe for ADRB2's cAMP signaling mechanism (Table 1). Indeed many low-TS compounds are popular and well tested (Table S5), and some are marketed drugs, but do not meet the criteria of a good tool compound. This confirms the observation that sometimes there are more strict criteria for selective chemical tool compounds than drugs (Arrowsmith et al., 2015). Many of the low-quality probes mentioned by Arrowsmith et al. (2015) were also found using our computational approach and were given low TS.

During our evaluation of these compounds and to which on-/off-targets they have been measured, we found that investigation around a compound could be biased toward what was known to be the primary target. For example, erlotinib had more than 200 total bioactivity measurements in ChEMBL, 23% of which were against epidermal growth factor receptor (EGFR) (i.e., target allocation = 23%). With apomorphine and DRD2 this number was higher, at 36% (Figure S1A). While this multiple evidence improved the on-target strength, it did not directly contribute to the selectivity of these compounds. However, once the measurements were binned into different potency ranges, the high-potency target allocations were reflective of a compound's selectivity. For example, with higher potency of erlotinib the binned target allocation of EGFR increases from 41% at 100 nM, to 86% at 10 nM, and 100% at $\leq$1 nM (Figure S1B). By contrast, the allocation of DRD2 in apomorphine's measurements did not increase at high potency and there were other sub-nM targets such as DRD1/4/5. For the well-used quercetin (Gryglewski et al., 1987; Sankari et al., 2014; Shin et al., 2015; Vita, 2005), more than 300 quantitative investigations were distributed among a diverse set of targets, none of which had high-potency allocation (Figures 2B and S1B). In other words, quercetin is not a potent or selective tool compound to the best of our knowledge. Strikingly, quercetin appears as a bound ligand in 20 different published structures in the RCSB PDB bound to diverse proteins and even G-quadruplex DNA.

Handling Incomplete Profiling Data

It is unfortunately not realistic to have every target profiled for every compound before picking the best tool compound for a target. Uneven target panels for individual candidate compounds are often the best we can get, despite our effort in large-scale information integration. The TS would not give a high score to poorly tested compounds for which too little was known, and we can only be as good as data available at any time point. For example, even staurosporine would have appeared selective if we had calculated its TS back in the early 2000s (Figure 2B, highlighted in dotted box). As a result, as new data become available, updating TS is important for the best prioritization of tool compounds as time goes on. The completeness issue is also one of the reasons why we have designed the TS to be independent of target panels because they are inherently incomplete. Unlike target panel-based Gini coefficients, the TS is equally computable whether the compound has been tested against one, ten, or 100 targets, and the scores computed with unequal numbers of measurements can still be compared with each other (Figure 5).

Importantly, the score is not a de novo target prediction, but rather an attempt to infer the complex property of "tool-ness" by organizing prior knowledge with a combination of expert-derived rules and machine learning. Therefore, in terms of validating the TS, there would be minimal value in simply reconfirming target binding experimentally. Moreover, there are no large-scale methods to test tool compound potency and selectivity across the human proteome. Instead we chose to validate our TS in terms of how it translates to selectivity profiles in a panel of cellular pathway assays, which better reflects the goal of prioritizing compounds as probes for phenotypic drug discovery. Admittedly, the RGA panel with its 41 reporters has limited power of dissecting all possible pathway phenotypes and underlying targets. However, this is an extremely useful model to connect target specificity and pathway specificity and confirm prioritized tool compounds within its coverage.

Application of Tool Compounds in Phenotypic Assays Requires Data Awareness and Caution

There are other criteria for tool compounds that were mentioned in previous studies, such as cell permeability or physicochemical properties, which are not explicitly investigated here. However, bioactivity data collected from cell-based dose-response assays is often a reliable proxy for such properties.

Another important factor is the expression levels of the intended targets in the phenotypic experimental system (cell line/mutant cell line, primary tissue, etc.). Compounds cannot act on targets that are not expressed in the tissue. As a result, apparent "dirty" drugs, i.e., compounds with polypharmacology, can still help deconvolute the biology when the tissue expression of their individual targets is taken into consideration (Emig et al., 2011). For example, Gujral et al. (2014) selected multiple kinase inhibitors with a broad range of polypharmacology targets and applied regularized regression to kinase expression levels to validate cell-type-specific kinases that regulate cancer cell migration. Furthermore, for stress testing a phenotypic system and determining whether the readout is specific enough to desired biology, target family-specific compounds such as pan-kinase inhibitors can also be used to evaluate the readout under multiple perturbations, and thus computational methods to identify them is also important. In such cases, less selective tool compounds could still become good tools in probing targets for phenotypes or in developing more specific assays.

The inverse is true, unfortunately, for good tool compounds that might become bad tools in an unfit phenotypic assay (Arrowsmith et al., 2015). A potent and selective compound cannot have cellular activity if its target is absent in the phenotypic assay. In addition, as shown in Figure 2A, many experiments have confirmed nilotinib's selectivity on BCR-ABL at below 10 nM, whereas at above 1 μM nilotinib will bind to other targets as well. Thus the experimental design should take into consideration what the effective concentration within the phenotypic system would be. Testing nilotinib at an effective concentration of 10 nM or lower will likely reveal the on-target biology of BCR-ABL, whereas if the intracellular compound concentration reaches beyond the micromolar range, multiple kinases will be inhibited and the observed phenotype, such as proliferation inhibition, can be the result of the on-target, or any of the off-targets, or the combination thereof. Furthermore, targets affecting multiple downstream signaling pathways, or induction of apoptosis, are less likely to be specifically probed in a phenotypic assay setup, as shown in the example of vorinostat (Dokmanovic et al., 2007), and dinaciclib (Pei and Xiong, 2005) (Figure 3). Using better and more specific tool compounds would not have improved the fishing of these targets from our RGA experiments due to the design of the reporters and the corresponding pathways (King et al., 2009). Therefore, careful evaluation of historical experimental data and setting up ongoing phenotypic assays is as important as the tool compound itself for the success of target hypothesis validation.

SIGNIFICANCE

One of the key interfaces between chemistry and biology is chemical tool compounds and how they provide insight into biological function and mechanism in phenotypic assays. In this study we systematically evaluated heterogeneous bioactivity data and phenotypic effect, designed a computational method for the prioritization of tool compounds, identified public-domain compounds with high and low selectivity, and profiled their cellular phenotype. This is the first report of a large-scale, data-driven method for tool compound selection not tailored to a specific target panel. Rather it uses a maximum of qualitative and quantitative bioactivity data and is robust, reproducible, and recomputable as the data evolve. Our analyses can be widely used by chemical biologists for prioritizing among existing libraries of tool compounds (e.g., vendor catalogs) from experiments, literature, and patents, and for designing phenotypic experiments that can better utilize the power of chemical tool compounds and avoid biases of well-tested compounds.

EXPERIMENTAL PROCEDURES

Defining Primary Target and Primary PFAM Domain

Target allocation, defined as the proportion of measurements on each target for a given compound, was used to rank all targets of the compound. The target with the highest allocation was assigned the primary target (on-target). If two targets had the same allocation for a compound, they were compared with the lowest micromolar activity measured and the lower one was assigned the primary target.

Similar to the definition of target allocation, we have defined the PFAM allocation to be the proportion of the PFAM domain represented by all targets of a compound. For example, if a compound is active against two kinase targets with each m and n measurements, the kinase domain allocation is m + n/all measurements on all domains. The PFAM domain that had the highest allocation was assigned the primary PFAM domain of the compound.

Potency Strength

We have designed an "assertion stack" consisting of assertions for each data source. Different from conventional Boolean assertions (for example, in graph approaches), each of our compound-target assertions is a qualitative predicate with a weight ("strength") evaluated empirically from its supporting evidence. Higher strength represents higher reliability of the assertion. Each high-level assertion is supported by multiple low-level assertions and the strength increments with multiple orthogonal sources of evidence (Figure 1B; Tables S1 and S2; Supplemental Experimental Procedures).

Selectivity Features

We trained supervised machine learning models on a training set of 507 compounds from DrugBank (Law et al., 2014) (Supplemental Experimental Procedures). From the many possible combinations of 36 selectivity features

(Figure S3 and Table S3), our models prioritized the combination of high-potency target allocation (TA_Q1_R, Figure S1), on-/off-target potency difference (Q1_IC50_diff, Figure S2), and on-/off-target measurement distribution difference (Wilcoxon_RS_Pval, Table S4 and Supplemental Experimental Procedures). Respectively, these three features represent a historical preponderance of high-activity data points obtained for the primary target, the spread of difference between on- and off-targets, and the significance of that spread. TS is defined as the product of potency strength and combined selectivity.

Evaluation of TS

To evaluate the performance of TS, we used it to rank selective and non-selective compounds in the 507 compound modeling set, and compared it with fold-IC_{50} features (Figure S4). Additionally we conducted RGA experiments on compounds with high and low TS to confirm their pathway potency and selectivity.

Identification of Non-selective Compounds Using TS

From our integrated database (Supplemental Experimental Procedures) we have filtered for compounds, which have been profiled against over 100 targets. Then for each of these compounds we identified the primary targets as described earlier. TS was calculated for each compound-primary target pair and we have identified 68 public compounds having TS <1.0. Then for each of the 68 compound we identified the primary PFAM domain and calculated PFAM allocation (Table S5).

Reporter Gene Assay

Compounds shown in Table 1 and Figure S6 were tested in duplicate in eight-point dose response (half-log dilution from top concentration 2 μM) in 41 RGAs representing 18 pathways with multiple stimuli (King et al., 2009). The dose-response potency was measured using DR_{50} (averaged over duplicates), defined as the difference in number of active doses and inactive doses divided by total number of doses. A potent agonist has a DR_{50} value of 1.0, a potent antagonist has a DR_{50} of −1.0, and an inactive compound has a DR_{50} of 0 in the RGA (King et al., 2009).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, five tables, and two data files and can be found with this article online at http://dx.doi.org/10.1016/j.chembiol.2016.05.016.

AUTHOR CONTRIBUTIONS

A.C. and F.N. collected and maintained internal and external compound bioactivity data; A.C. designed the assertion stack.; Y.W., C.G.P., and J.L.J. designed research; F.K. performed RGA experiments; G.M. analyzed RGA data; Y.M. and Y.W. performed modeling; Y.W. and J.L.J. wrote the manuscript.

ACKNOWLEDGMENT

Y.M. is funded with the summer scholar fellowship by the Novartis Institutes for BioMedical Research education office. The authors thank Iain Wallace, Eugen Lounkine, and Anne Mai Wassermann for useful discussions.

Received: June 1, 2015
Revised: April 29, 2016
Accepted: May 13, 2016
Published: July 14, 2016

REFERENCES

Ahlquist, R.P. (1976). Present state of alpha- and beta-adrenergic drugs I. The adrenergic receptor. Am Heart J *92*, 661–664.

Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., et al. (2015). The promise and peril of chemical probes. Nat. Chem. Biol. *11*, 536–541.

Azzaoui, K., Jacoby, E., Senger, S., Rodriguez, E.C., Loza, M., Zdrazil, B., Pinto, M., Williams, A.J., de la Torre, V., Mestres, J., et al. (2013). Scientific competency questions as the basis for semantically enriched open pharmacological space development. Drug Discov. Today *18*, 843–852.

Bassilana, F., Carlson, A., DaSilva, J.A., Grosshans, B., Vidal, S., Beck, V., Wilmeringwetter, B., Llamas, L.A., Showalter, T.B., Rigollier, P., et al. (2014). Target identification for a Hedgehog pathway inhibitor reveals the receptor GPR39. Nat. Chem. Biol. *10*, 343–349.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. (2004). The Pfam protein families database. Nucleic Acids Res. *32*, D138–D141.

Bolling, C., Weidlich, M., and Holzhutter, H.G. (2014). SEE: structured representation of scientific evidence in the biomedical domain using Semantic Web techniques. J. Biomed. Semantics *5*, S1.

Bunnage, M.E., Chekler, E.L., and Jones, L.H. (2013). Target validation using chemical probes. Nat. Chem. Biol. *9*, 195–199.

Chen, B., Ding, Y., and Wild, D.J. (2012a). Assessing drug target association using semantic linked data. PLoS Comput. Biol. *8*, e1002574.

Chen, B., Ding, Y., and Wild, D.J. (2012b). Improving integrative searching of systems chemical biology data using semantic annotation. J. Cheminform. *4*, 6.

Cumming, J.G., Davis, A.M., Muresan, S., Haeberlein, M., and Chen, H. (2013). Chemical predictive modelling to improve compound quality. Nat. Rev. Drug Discov. *12*, 948–962.

Dai, Y., and Grant, S. (2003). Cyclin-dependent kinase inhibitors. Curr. Opin. Pharmacol. *3*, 362–370.

Dokmanovic, M., Clarke, C., and Marks, P.A. (2007). Histone deacetylase inhibitors: overview and perspectives. Mol. Cancer Res. *5*, 981–989.

Druker, B.J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G.M., Fanning, S., Zimmermann, J., and Lydon, N.B. (1996). Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat. Med. *2*, 561–566.

Eggert, U.S. (2013). The why and how of phenotypic small-molecule screens. Nat. Chem. Biol. *9*, 206–209.

Emig, D., Kacprowski, T., and Albrecht, M. (2011). Measuring and analyzing tissue specificity of human genes and protein complexes. EURASIP J. Bioinform Syst. Biol. *2011*, 5.

Feng, Y., Mitchison, T.J., Bender, A., Young, D.W., and Tallarico, J.A. (2009). Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nat. Rev. Drug Discov. *8*, 567–578.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230.

Frye, S.V. (2010). The art of the chemical probe. Nat. Chem. Biol. *6*, 159–161.

Gambacorti-Passerini, C. (2008). Part I: milestones in personalised medicine—imatinib. Lancet Oncol. *9*, 600.

Graczyk, P.P. (2007). Gini coefficient: a new way to express selectivity of kinase inhibitors against a family of kinases. J. Med. Chem. *50*, 5773–5779.

Gregori-Puigjane, E., Setola, V., Hert, J., Crews, B.A., Irwin, J.J., Lounkine, E., Marnett, L., Roth, B.L., and Shoichet, B.K. (2012). Identifying mechanism-of-action targets for drugs and probes. Proc. Natl. Acad. Sci. USA *109*, 11178–11183.

Gryglewski, R.J., Korbut, R., Robak, J., and Swies, J. (1987). On the mechanism of antithrombotic action of flavonoids. Biochem. Pharmacol. *36*, 317–322.

Gujral, T.S., Peshkin, L., and Kirschner, M.W. (2014). Exploiting polypharmacology for drug target deconvolution. Proc. Natl. Acad. Sci. USA *111*, 5048–5053.

Hart, C.P. (2005). Finding the target after screening the phenotype. Drug Discov. Today *10*, 513–519.

Hopkins, A.L. (2008). Network pharmacology: the next paradigm in drug discovery. Nat. Chem. Biol. *4*, 682–690.

Hopkins, A.L. (2009). Drug discovery: predicting promiscuity. Nature *462*, 167–168.

Huang, S.M., Mishina, Y.M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G.A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., et al. (2009). Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. Nature *461*, 614–620.

Ishii, H., Jirousek, M.R., Koya, D., Takagi, C., Xia, P., Clermont, A., Bursell, S.E., Kern, T.S., Ballas, L.M., Heath, W.F., et al. (1996). Amelioration of vascular dysfunctions in diabetic rats by an oral PKC beta inhibitor. Science *272*, 728–731.

Jester, B.W., Gaj, A., Shomin, C.D., Cox, K.J., and Ghosh, I. (2012). Testing the promiscuity of commercial kinase inhibitors against the AGC kinase group using a split-luciferase screen. J. Med. Chem. *55*, 1526–1537.

Jones, J.O., and Diamond, M.I. (2007). Design and implementation of cell-based assays to model human disease. ACS Chem. Biol. *2*, 718–724.

Jones, G.H., and Scott, S.J. (2011). Continuous infusions of terbutaline in asthma—a review. J. Asthma *48*, 753–756.

Kalliokoski, T., Kramer, C., Vulpetti, A., and Gedeck, P. (2013). Comparability of mixed IC(5)(0) data—a statistical analysis. PLoS One *8*, e61007.

Kano, M.R., Bae, Y., Iwata, C., Morishita, Y., Yashiro, M., Oka, M., Fujii, T., Komuro, A., Kiyono, K., Kaminishi, M., et al. (2007). Improvement of cancer-targeting therapy, using nanocarriers for intractable solid tumors by inhibition of TGF-beta signaling. Proc. Natl. Acad. Sci. USA *104*, 3460–3465.

Karaman, M.W., Herrgard, S., Treiber, D.K., Gallant, P., Atteridge, C.E., Campbell, B.T., Chan, K.W., Ciceri, P., Davis, M.I., Edeen, P.T., et al. (2008). A quantitative analysis of kinase inhibitor selectivity. Nat. Biotechnol. *26*, 127–132.

Kaya, A.I., Onaran, H.O., Ozcan, G., Ambrosio, C., Costa, T., Balli, S., and Ugur, O. (2012). Cell contact-dependent functional selectivity of beta2-adrenergic receptor ligands in stimulating cAMP accumulation and extracellular signal-regulated kinase phosphorylation. J. Biol. Chem. *287*, 6362–6374.

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

King, F.J., Selinger, D.W., Mapa, F.A., Janes, J., Wu, H., Smith, T.R., Wang, Q.-Y., Niyomrattanakitand, P., Sipes, D.G., and Brinker, A. (2009). Pathway reporter assays reveal small molecule mechanisms of action. J. Assoc. Lab. Autom. *14*, 374–382.

Kramer, C., Kalliokoski, T., Gedeck, P., and Vulpetti, A. (2012). The experimental uncertainty of heterogeneous public K(i) data. J. Med. Chem. *55*, 5165–5173.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. *42*, D1091–D1097.

Lee, J.A., Uhlik, M.T., Moxham, C.M., Tomandl, D., and Sall, D.J. (2012). Modern phenotypic drug discovery is a viable, neoclassic pharma strategy. J. Med. Chem. *55*, 4527–4538.

Lounkine, E., Keiser, M.J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J.L., Lavan, P., Weber, E., Doak, A.K., Cote, S., et al. (2012). Large-scale prediction and testing of drug activity on side-effect targets. Nature *486*, 361–367.

Moffat, J.G., Rudolph, J., and Bailey, D. (2014). Phenotypic screening in cancer drug discovery - past, present and future. Nat. Rev. Drug Discov. *13*, 588–602.

Narwal, M., Haikarainen, T., Fallarero, A., Vuorela, P.M., and Lehtio, L. (2013). Screening and structural analysis of flavones inhibiting tankyrases. J. Med. Chem. *56*, 3507–3517.

Nidhi, Glick, M., Davies, J.W., and Jenkins, J.L. (2006). Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J. Chem. Inf. Model. *46*, 1124–1133.

Pei, X.H., and Xiong, Y. (2005). Biochemical and cellular mechanisms of mammalian CDK inhibitors: a few unresolved issues. Oncogene *24*, 2787–2795.

Riazanov, A., Klein, A., Shaban-Nejad, A., Rose, G.W., Forster, A.J., Buckeridge, D.L., and Baker, C.J. (2013). Semantic querying of relational data for clinical intelligence: a semantic web services-based approach. J. Biomed. Semantics *4*, 9.

Riniker, S., Fechner, N., and Landrum, G.A. (2013). Heterogeneous classifier fusion for ligand-based virtual screening: or, how decision making by committee can be a good thing. J. Chem. Inf. Model. *53*, 2829–2836.

Riniker, S., Wang, Y., Jenkins, J.L., and Landrum, G.A. (2014). Using information from historical high-throughput screens to predict active compounds. J. Chem. Inf. Model. *54*, 1880–1891.

Roth, B.L., Sheffler, D.J., and Kroeze, W.K. (2004). Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat. Rev. Drug Discov. *3*, 353–359.

Ruegg, U.T., and Burgess, G.M. (1989). Staurosporine, K-252 and UCN-01: potent but nonspecific inhibitors of protein kinases. Trends Pharmacol. Sci. *10*, 218–220.

Ruttenberg, A., Clark, T., Bug, W., Samwald, M., Bodenreider, O., Chen, H., Doherty, D., Forsberg, K., Gao, Y., Kashyap, V., et al. (2007). Advancing translational research with the semantic web. BMC Bioinformatics *8* (*Suppl 3*), S2.

Sakiyama, Y. (2009). The use of machine learning and nonlinear statistical tools for ADME prediction. Expert Opin. Drug Metab. Toxicol. *5*, 149–169.

Samwald, M., Jentzsch, A., Bouton, C., Kallesoe, C.S., Willighagen, E., Hajagos, J., Marshall, M.S., Prud'hommeaux, E., Hassenzadeh, O., Pichler, E., et al. (2011). Linked open drug data for pharmaceutical research and development. J. Cheminform. *3*, 19.

Sankari, S.L., Babu, N.A., Rani, V., Priyadharsini, C., and Masthan, K.M. (2014). Flavonoids - clinical effects and applications in dentistry: a review. J. Pharm. Bioallied Sci. *6*, S26–S29.

Scheiber, J., Chen, B., Milik, M., Sukuru, S.C., Bender, A., Mikhailov, D., Whitebread, S., Hamon, J., Azzaoui, K., Urban, L., et al. (2009). Gaining insight into off-target mediated effects of drug candidates with a comprehensive systems chemical biology analysis. J. Chem. Inf. Model *49*, 308–317.

Schirle, M., and Jenkins, J.L. (2015). Identifying compound efficacy targets in phenotypic drug discovery. Drug Discov. Today *21*, 82–89.

Shadbolt, N., Hall, W., and Berners-Lee, T. (2006). The semantic web revisited. Ieee Intell. Syst. *21*, 96–101.

Shapiro, G.I. (2006). Cyclin-dependent kinase pathways as targets for cancer treatment. J. Clin. Oncol. *24*, 1770–1783.

Shin, Y.J., Kim, J.J., Kim, Y.J., Kim, W.H., Park, E.Y., Kim, I.Y., Shin, H.S., Kim, K.S., Lee, E.K., Chung, K.H., et al. (2015). Protective effects of quercetin against HgCl(2)-induced nephrotoxicity in Sprague-Dawley rats. J. Med. Food *18*, 524–534.

Swinney, D.C. (2013). Phenotypic vs. target-based drug discovery for first-in-class medicines. Clin. Pharmacol. Ther. *93*, 299–301.

Tamaoki, T., Nomoto, H., Takahashi, I., Kato, Y., Morimoto, M., and Tomita, F. (1986). Staurosporine, a potent inhibitor of phospholipid/Ca^{++} dependent protein kinase. Biochem. Biophys. Res. Commun. *135*, 397–402.

Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K., and Aittokallio, T. (2014). Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J. Chem. Inf. Model. *54*, 735–743.

Tiikkainen, P., Bellis, L., Light, Y., and Franke, L. (2013). Estimating error rates in bioactivity databases. J. Chem. Inf. Model. *53*, 2499–2505.

Tsvetanova, N.G., and von Zastrow, M. (2014). Spatial encoding of cyclic AMP signaling specificity by GPCR endocytosis. Nat. Chem. Biol. *10*, 1061–1065.

Uitdehaag, J.C., and Zaman, G.J. (2011). A theoretical entropy score as a single value to express inhibitor selectivity. BMC Bioinformatics *12*, 94.

Uitdehaag, J.C., Verkaar, F., Alwan, H., de Man, J., Buijsman, R.C., and Zaman, G.J. (2012). A guide to picking the most selective kinase inhibitor tool compounds for pharmacological validation of drug targets. Br. J. Pharmacol. *166*, 858–876.

Violin, J.D., DiPilato, L.M., Yildirim, N., Elston, T.C., Zhang, J., and Lefkowitz, R.J. (2008). beta(2)-adrenergic receptor signaling and desensitization elucidated by quantitative modeling of real time cAMP dynamics. J. Biol. Chem. *283*, 2949–2961.

Vita, J.A. (2005). Polyphenols and cardiovascular disease: effects on endothelial and platelet function. Am. J. Clin. Nutr. *81*, 292S–297S.

Wild, D.J., Ding, Y., Sheth, A.P., Harland, L., Gifford, E.M., and Lajiness, M.S. (2012). Systems chemical biology and the Semantic Web: what they mean for the future of drug discovery research. Drug Discov. Today *17*, 469–474.

Williams, A.J., Harland, L., Groth, P., Pettifer, S., Chichester, C., Willighagen, E.L., Evelo, C.T., Blomberg, N., Ecker, G., Goble, C., et al. (2012). Open PHACTS: semantic interoperability for drug discovery. Drug Discov. Today *17*, 1188–1198.

Workman, P., and Collins, I. (2010). Probing the probes: fitness factors for small molecule tools. Chem. Biol. *17*, 561–577.

Cell Chemical Biology, Volume *23*

Supplemental Information

Evidence-Based and Quantitative Prioritization of Tool Compounds in Phenotypic Drug Discovery

Yuan Wang, Allen Cornett, Fred J. King, Yi Mao, Florian Nigsch, C. Gregory Paris, Gregory McAllister, and Jeremy L. Jenkins

Supplemental Information

Supplemental Figures and Legends

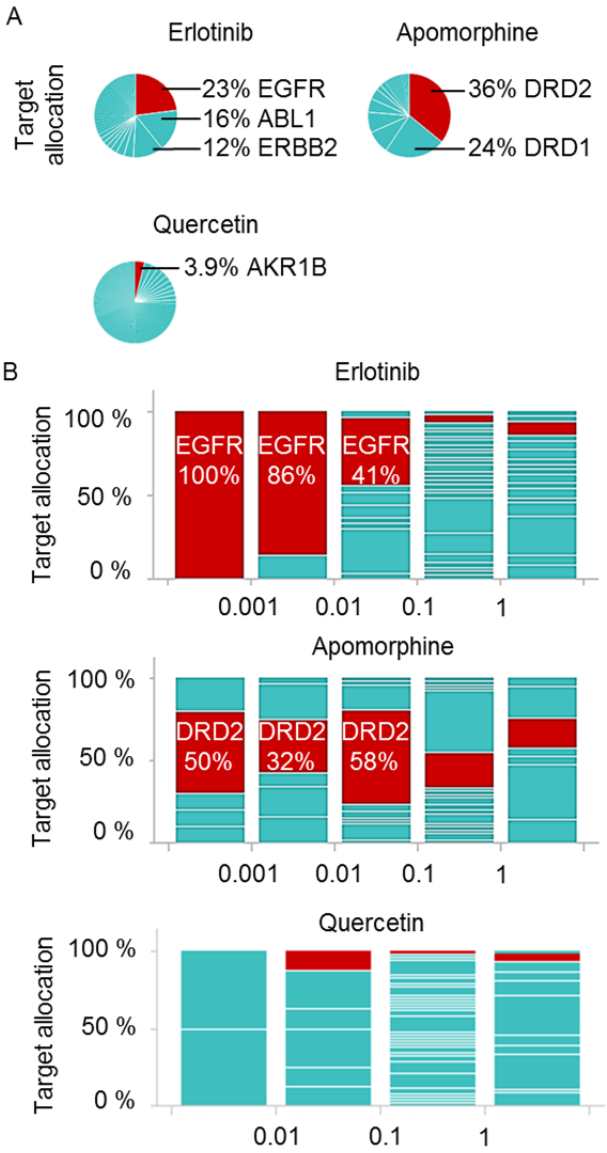


Figure S1. (Related to Figure 1 and Figure 2) Target allocation (proportion of measurements) for each target of erlotinib, apomorphine and quercetin.
A) The total target allocation reveals preference of targets. B) Target allocation binned into different potency ranges. Erlotinib has high allocation for EGFR at high potency, reflective of frequently confirmed selectivity, while apomorphine and quercetin (primary target DRD2/AKR1B highlighted in red) have less selectivity even at high potency.

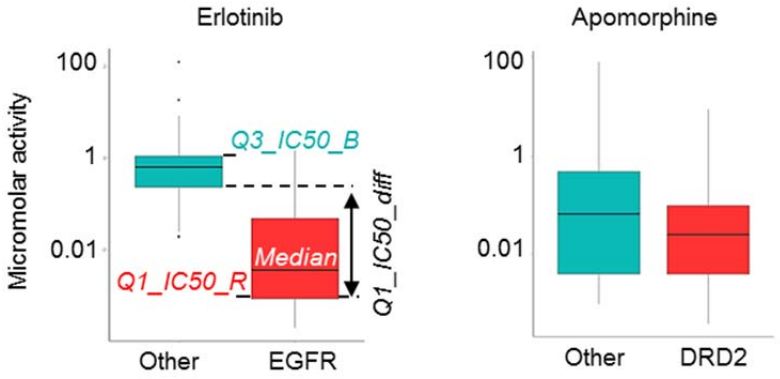


Figure S2. (Related to Figure 1 and Figure 2) Definition of Q1 and Q3 of on-/off-target micromolar activities, and Q1_IC50_diff.

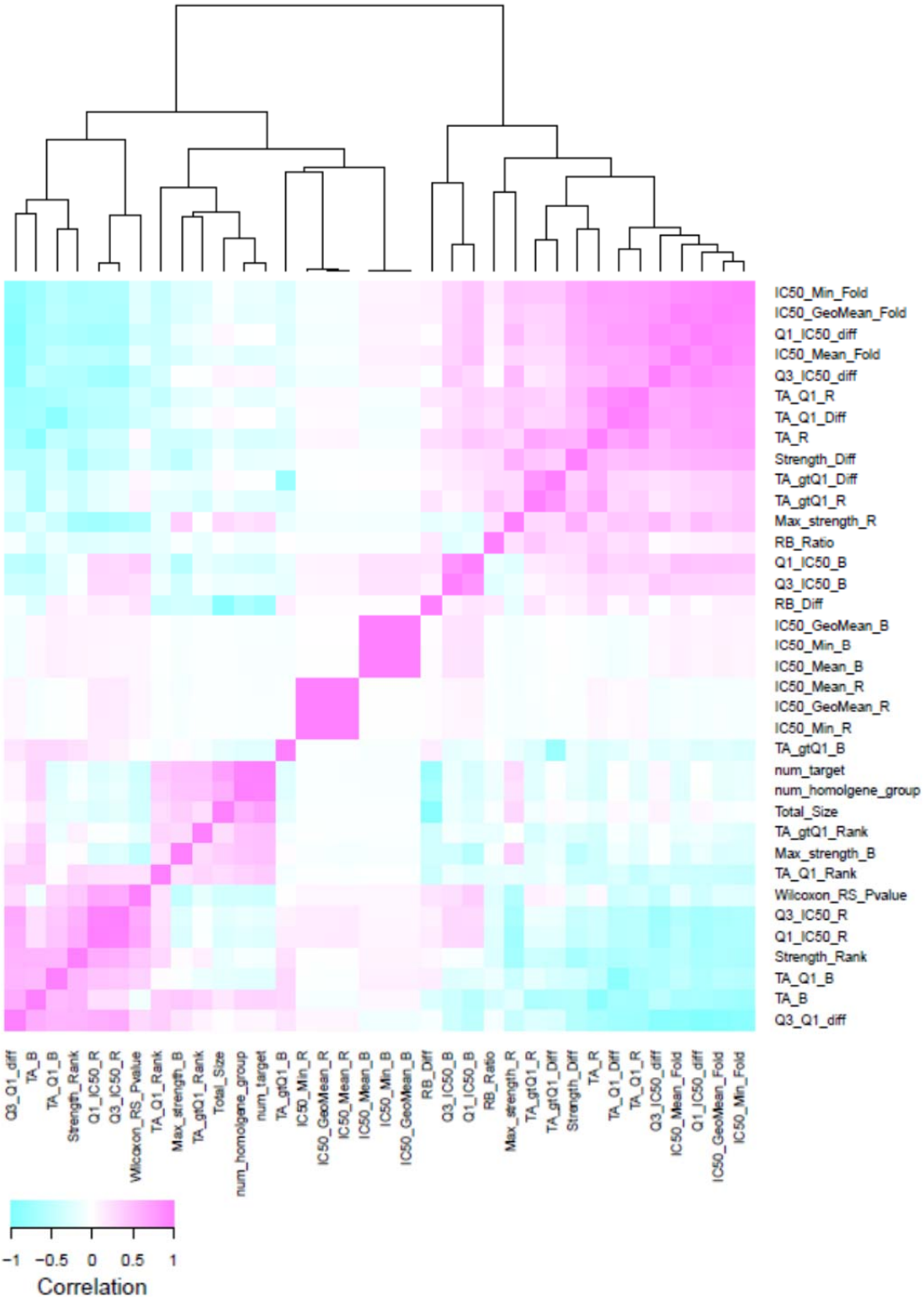


Figure S3. (Related to Experimental Procedures) Feature correlation matrix.
Some selectivity features were highly correlated, others not.

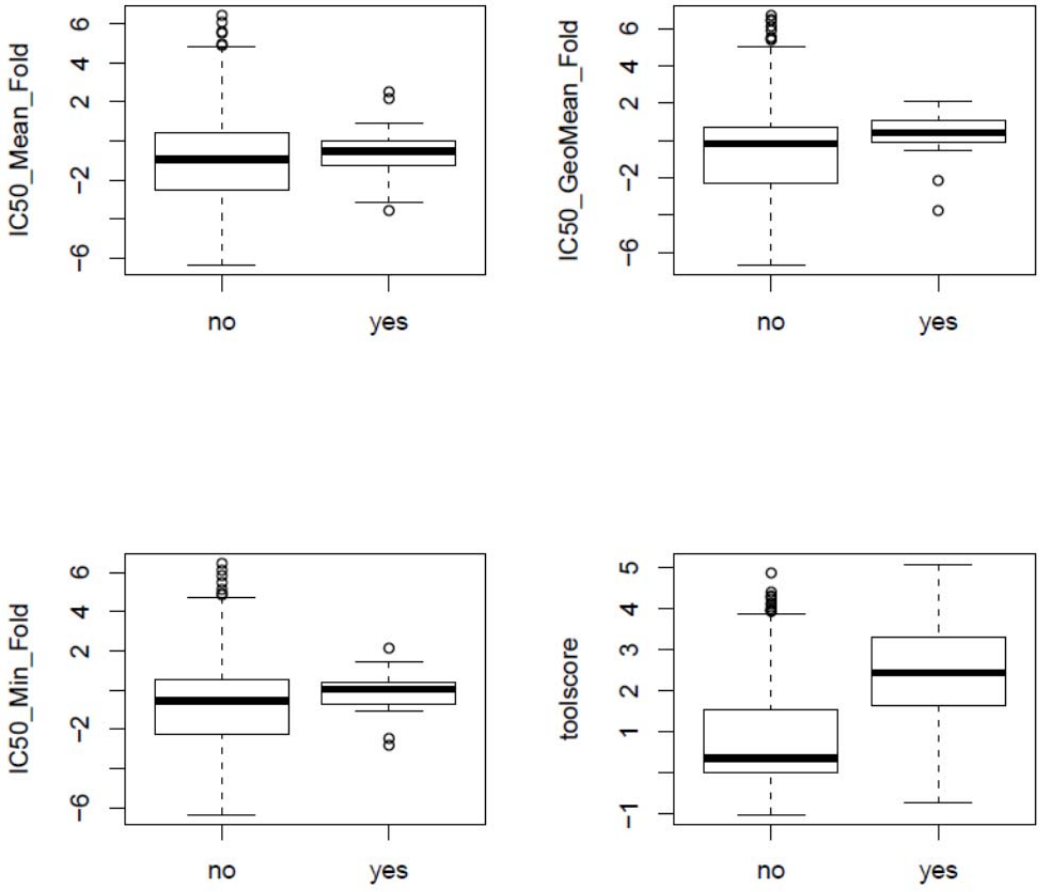


Figure S4. (Related to Experimental Procedures) Comparing value distributions for fold IC50 features and TS.

To evaluate the performance of TS, we used it to rank selective and non-selective compounds in the 507 compound modeling set, and compared it with fold-IC50 features. (Y axis is the corresponding log fold IC50 using various aggregation methods). The TS was better at separating the manually annotated good ("yes") and less selective ("no") tool compounds in the DrugBank dataset.

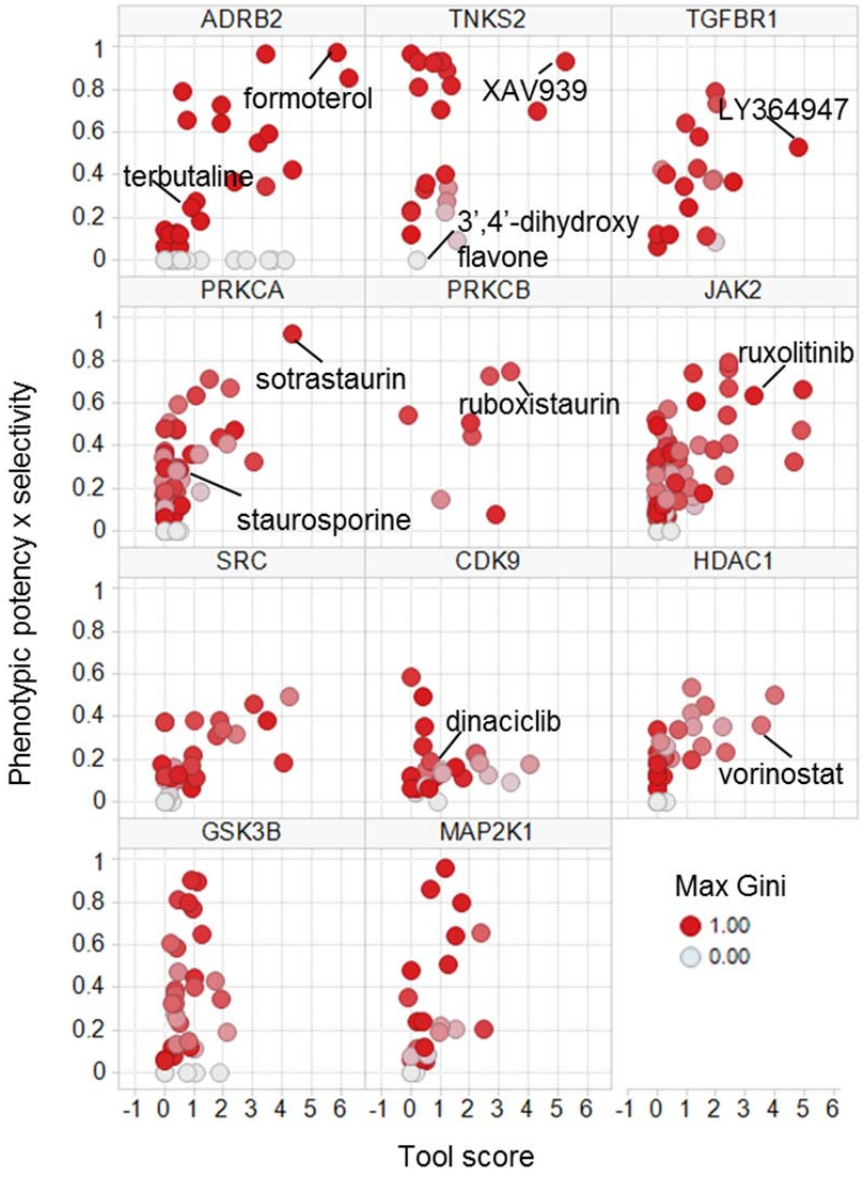


Figure S5. (Related to Figure 4) Correlation between compound target TS and phenotypic potency and selectivity.

Each dot represents a compound. Its TS (on-target potency × selectivity) is shown on the x-axis, and the product of phenotypic potency (maximum absolute DR50) and phenotypic selectivity (Gini coefficient) on the y-axis. Compounds having high TS were more likely to be more potent in the phenotypic assay, and if their targets were upstream of the reported pathway (Figure 3B), then their selectivity will also be higher.

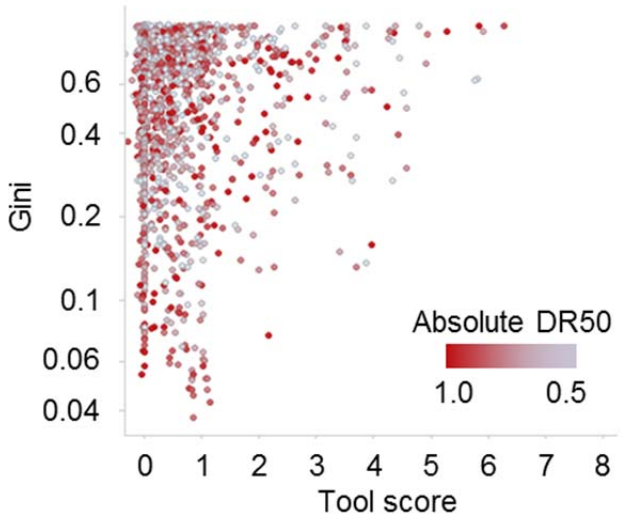


Figure S6. (Related to Figure 4) Tool utility and phenotypic specificity.
For 1457 compounds the activity and selectivity in target dimension (TS, x-axis) and selectivity on pathway dimension (Gini coefficient, y-axis) are plotted against each other. Each dot represents one compound, colored by DR50 (red: high potency). Compounds with high TS tend to be potent and selective in only one or a few of the 41 pathway RGAs (upper-right quadrant), whereas compounds with low TS can be selective due to lack of data/knowledge (upper left) or non-selective as indicated by low TS (lower left). Importantly, few compounds with high TS exhibit a low Gini score (lower-right quandrant).

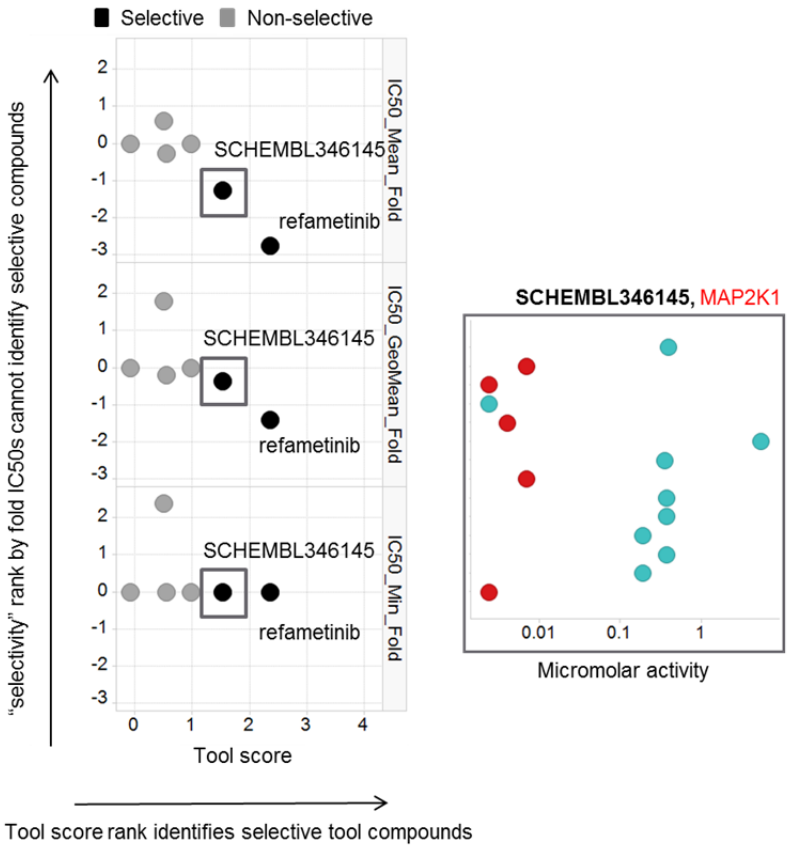


Figure S7. (Related to Table1 and Figure 5) TS identifies selective tool compounds while IC50s alone not.
X axis is TS and Y axis is the corresponding log fold IC50. Ranking by TS allowed us to prospectively select MAP2K1 tool compounds, which were later validated to be selective in the RGA panel, while high fold IC50 differences alone could not have distinguished these compounds due to heterogeneity in the measurements. For example, refametinib would not be chosen as tool compound using any of the three fold IC50 metrics. For SCHEMBL346145 (highlighted) the measurements were shown in the box. Red dots are MAP2K1 measurements, and blue dots multiple other targets. Depending on the number of data points per target the IC50 aggregations are different. One off target having only one data point (leftmost blue dot among the red ones in the box) has lower aggregated IC50s than MAP2K1's. TS captures the overall selectivity and gives this compound a high rank despite the apparent lack of IC50 fold difference. Only data from public domain were shown.

Supplemental Tables

Table S1. (Related to Experimental Procedures) Attributes of an assertion.

Attribute	Description
assertion_id	Unique assertion ID
assertion	Short string identifying the kind of assertion. The meaning of each assertion is listed in **Table S2**.
inchi_key	The InChI key for the compound this assertion is applied to.
gene_id	NCBI (Entrez) Gene ID
gene_symbol	NCBI (Entrez) Gene Symbol
tax_id	NCBI taxonomy ID
Strength	Strength of the assertion on a scale from 1 (weakest) to 8 (strongest)
Attrs	Key-Value list of assertion-specific information (e.g. data source identifiers)

Table S2. (Related to Experimental Procedures) Assertions and corresponding strength contributions by evidence type, source and potency range for compound-target activity.

Type	Assertion	Description	Strength	Evidenced By
	ACTIVE	Compound is active against a target.	Maximum of evidence strengths from below sources, plus 1 if evidence count > 2	CHEMBL_ACTIVE DRUGBANK_ACTIVE GVK_ACTIVE INTEGRITY_ACTIVE INHOUSE_ACTIVE
Qualitative, curated data	DRUGBANK_ACTIVE	Compound has an activity association with a target in DrugBank	7 for approved drugs 2 for approved nutraceuticals else 1	DRUGBANK_IDS
	INTEGRITY_ACTIVE	Compound has an activity association with a target in Integrity	7 for approved drugs 5 if Phase III or higher 3 if IND filed or higher 2 if Preclinical else 1	INTEGRITY_ENTRY_NUMBERS
Quantitative assay data	CHEMBL_ACTIVE	Compound is active against a target in ChEMBL.	3 if more than one low nanomolar value for target (<=50 nM) 2 if more than four activities (< 1 micromolar) for target 1 if it has a micromolar value <= 1 for target	CHEMBL_ASSAY_IDS CHEMBL_TARGET_IDS CHEMBL_MOLECULE_IDS
	GVK_ACTIVE	Compound is active against a target in GVK	3 if more than one low nanomolar value (<=50 nM) for target 2 if more than four activities (<1 micromolar) for target 1 if it has a micromolar value <= 1 for target	GVK_ACTIVITY_IDS MICROMOLAR_VALUES
Internally validated quantitative assay data	INHOUSE_ACTIVE	Compound is active against a target in internal assays	4 if more than one low nanomolar value (<=50 nM) for target 3 if more than four activities (<1 micromolar) for target 1 if it has a micromolar value <= 1 for target	INHOUSE_ASSAY_IDS MICROMOLAR_VALUES

Table S3. (Related to Experimental Procedures) Calculated features describing on-/off-target measurements (features selected for TS selectivity in bold)

Feature	Description
num_target	Number of different targets (each target identified by its gene id) tested for each compound.
num_homologene_group	Number of different targets (each target identified by its homologene group id, same for the remaining descriptors below) tested for each compound.
IC50_GeoMean_R	Compound-target dose response value (aggregated by geometric mean per target) of on-target.
IC50_GeoMean_B	Minimum of compound-target dose response value (aggregated by geometric mean per target) over all off-targets
IC50_GeoMean_Fold	IC50_GeoMean_B/IC50_GeoMean_R, log scale
IC50_Mean_R	Compound-target dose response value (aggregated by mean per target) of on-target.
IC50_Mean_B	Minimum of compound-target dose response value (aggregated by mean per target) over all off-targets
IC50_Mean_Fold	IC50_Mean_B/IC50_Mean_R, log scale
IC50_Min_R	Compound-target dose response value (aggregated by min per target) of on-target.
IC50_Min_B	Minimum of compound-target dose response value over all off-targets
IC50_Min_Fold	IC50_Min_B/IC50_Min_R, log scale
Max_strength_R	Maximum of compound-target strength of on-target group.
Max_strength_B	Maximum of compound-target strength of off-target over all off-targets.
Strength_Diff	Max_Strength_R – Max_Strength_B.
Strength_Rank	Ranking the Max_Strength_R and strength of off-target group in a descending order. Strength_Rank is the rank of the Max_Strength_R.
Q1_IC50_R	The lowest 25 percentile of micromolar activities (Q1_IC50) of on-target group, log scale.
Q1_IC50_B	The lowest 25 percentile of micromolar activities (Q1_IC50) of off-target group, log scale.
Q1_IC50_diff	Difference of on target and off target Q1_IC50, log scale.
Q3_IC50_R	75-percentile of the micromolar activity of on-target measurements.
Q3_IC50_B	75-percentile of the micromolar activity of off-target measurements.
Q3_IC50_diff	Difference of on target and off target Q3_IC50, log scale.
Q3_Q1_diff	Q3_IC50_R - Q1_IC50_B
Total_Size	The total number of measurements of each compound.
TA_R	On-target allocation
TA_B	Maximum off-target allocation over all off-targets
RB_Diff	TA_R - TA_B
RB_Ratio	TA_R/TA_B, or 0 if TA_B undefined
TA_Q1_R	High-potency (less than 25 percentile of all measurements) on-target allocation.
TA_Q1_B	Maximum high-potency (less than 25 percentile of all measurements) off-target allocation over all off-targets.
TA_Q1_Diff	TA_Q1_R - TA_Q1_B.
TA_Q1_Rank	All targets' allocations at high-potency are ranked from high to low and TA_HP_Rank is the rank of the on-target allocation within this list.
TA_gtQ1_R	Low-potency (above 25 percentile) on-target allocation.
TA_gtQ1_B	Maximum low-potency (above 25 percentile) off-target allocation over all off-

	targets.
TA_gtQ1_Diff	TA_gtQ1_R – TA_gtQ1_B.
TA_gtQ1_Rank	All targets' allocations at low-potency are ranked from high to low and TA_gtQ1_Rank is the rank of the on-target allocation within this list.
Wilcoxon_RS_Pvalue	P-value of Wilcoxon rank-sum test comparing micromolar activity between on-target and all off-target measurements.

Table S4. (Related to Experimental Procedures) Performance (AUC) of models trained with different feature sets.

Feature set	5-fold cross validation	DT	RF	NB	Average
All 36 features	1	0.78	0.89	0.84	0.83
	2	0.62	0.90	0.80	0.77
	3	0.64	0.97	0.82	0.81
	4	0.84	0.89	0.79	0.84
	5	0.70	0.90	0.63	0.75
	Average	**0.72**	**0.91**	**0.78**	**0.80**
3 features using fold IC50	1	0.77	0.77	0.72	0.75
	2	0.63	0.79	0.80	0.74
	3	0.71	0.82	0.75	0.76
	4	0.75	0.87	0.74	0.78
	5	0.47	0.69	0.71	0.62
	Average	**0.66**	**0.79**	**0.74**	**0.73**
3 features using iterative optimization	1	0.78	0.76	0.78	0.78
	2	0.85	0.86	0.90	0.87
	3	0.63	0.92	0.92	0.82
	4	0.83	0.83	0.84	0.83
	5	0.63	0.85	0.83	0.77
	Average	**0.74**	**0.84**	**0.85**	**0.81**

Table S5. (Related to Figure 6, separate file) 68 compounds with over 100 targets and not suitable as tool compounds for the corresponding targets, selected using TS (less than 1). Their primary targets as well as primary PFAM domains were reported. Primary PFAM allocation is the proportion of measurements on the primary PFAM domain. The higher the PFAM allocation, the more target family-specific the compound is likely to be. 38 out of the 68 less selective tool compounds were tested in the RGA panel and in **Figure 6B** the correlation between PFAM allocation and RGA potency is shown.

Supplemental Experimental Procedures

Collection of bioactivity data

Data from both internal and external sources are collected and normalized within Novartis. These data consist of two pillars: One is bioinformatics knowledge from sources like NCBI Entrez Gene [1], GeneOntology [2, 3], UniProt [4]. The other is cheminformatic knowledge from sources like Drugbank [5], ChEMBL [6] and BindingDB [7, 8]. Compound-related data are standardized (CACTVS toolkit, Xemistry) and connected using the IUPAC InChIKey [9, 10] representation of the structures, whereas target-related data are standardized and connected using NCBI Entrez Gene IDs. Multiple genes in the same homolog gene group are merged using HomoloGene (http://www.ncbi.nlm.nih.gov/homologene). Compound bioactivity data in dose response are normalized to micro-molar as the standard unit.

One limitation of the data normalization process, where targets were converted to Entrez Gene IDs and HomoloGene IDs, was the identification of subunits of complexes and target isoforms. It was possible that one subunit or isoform was assigned on-target and the rest off-targets. The comparable potency values on both on- and off- targets led to apparent lack of selectivity for a compound hitting a complex or multiple isoforms.

Calculating compound-target activity strength

Conventionally, computational inferences are done using binary assertions (true-false predicates). They can be described in a form like "if $\vdash X$ and $\vdash Y$ then $\vdash Z$", meaning if assertions of the form X and Y have been proven true, then Z is also proven [11]. For example, "if 'a inhibits b' can be proven, and 'b activates c' can be proven, then 'inhibiting a will activate b and cause induction of c' is proven". However, the bioactivities of compounds are rarely binary in real world. Extending this key logical foundation to make new assertions with quantitative attributes is necessary to describe knowledge of compound-target network.

Here we define assertions as statements about compounds that are derived from data integrated from multiple sources. Each assertion is an entity with attributes (**Table S1)**. Some assertions are derived from other assertions, while others are derived directly from data sources. The relationship between derived assertions and the evidences they are derived from is captured in an evidence table. Assertion types and their supporting evidence assertions are provided (**Table S2**). In practice, however, these assertions will vary depending on the reader's available data sources. Both ChEMBL and DrugBank are publicly available resources that may be used for a TS calculation. The particular numeric values for lower-level assertions were developed by the authors on the basis of personal expertise and experience working with the named data sources. For example, we propose that the advancement of a drug in a clinical pipeline is a type of proxy for in vivo validation and we expect that a drug that progresses to a launched status is better validated for its primary target than a tool from preclinical studies; we have therefore weighted Integrity compound-target pairs accordingly. We advise the reader to similarly implement personal domain-expertise to tune strength values for additional content that may be added. We have programmed these assertion logic rules so that all assertions derived from all data we collected from in-house and external sources are pre-calculated and stored in our database tables. Both the integrated data and the dependent assertions are maintained and updated regularly.

Calculating descriptors of selectivity

Estimating compound selectivity from disparate integrated data is a previously unexplored meta-analysis wherein we assume that we do not know *a priori* the most informative parameters. We therefore created and calculated thirty-six features that describe various aspects of the distribution of on-/off-target bioactivity measurements. Definitions of all 36 features are listed in **Table S3**. In this list we included features commonly used as selectivity descriptors such as the fold differences of IC50 (and other dose-response results such as Ki, EC50 etc., abbreviated as IC50 for simplicity), as well as descriptors describing the size and distribution of dose response measurements. Missing IC50 values were filled with 10000 micro-molar for the calculation of dependent features such as conventional IC50 aggregations and fold differences, and the lack of compound-target activity assertion was filled by strength=0. Wilcoxon rank-sum p-values which were not computable were filled with 1, and which exceeded the significance value of 1E-32 were replaced with 1E-32 for ease of computation. In **Figure S3** we show the correlation of these features. While some features were closely related, others were describing very different information.

Prioritizing descriptors of selectivity

We used a dataset consisting of 507 drugs from DrugBank and manually annotated 35 of them as selective tool compounds on the basis of their apparent selectivity in bioactivity profiles (as visible in **Figure 2A-B**). Then we built three supervised machine learning models using this dataset: decision trees (DT [12]), random forest (RF [13]) and naive Bayesian (NB [14]). We iteratively used subsets of the 36 features to train the models, and used five-fold cross validation to evaluate the performance of the models. Area under the curve (AUC), defined as the area under the receiver operating characteristic curve (ROC curve), was compared among the models. In **Table S4** we report the AUC of models trained with all 36 features, fold IC50 features, and a final optimized subset consisting of three features. This latter set of optimized features allowed us to train models that had comparable or sometimes better performance than the complete feature set, and consistently outperform the models trained with only fold IC50 features.

Defining tool score (TS)

Generally speaking, good chemical probes should have high numbers of measurements, high strength, without sacrificing the on-target potency, separation of on-target and off-target measurements, or the high-potency on-target allocation in their measurements. While it was possible to predict compounds' selectivity using machine learning models and the optimized feature set, for convenience and interpretability we incorporated the key selectivity features from machine learning into a computable metric that could be used independent of a model. We defined selectivity as

$$Selectivity = (Q1\_IC50\_diff/10 + TA\_Q1\_R - log10(Wilcoxon\_RS\_Pvalue))/3.$$

The TS is then defined as

$$TS = Strength \times Selectivity.$$

In **Figure S4** we compare four features of good and bad tool compounds among the training set, IC50_Mean_Fold, IC50_GeoMean_Fold, IC50_Min_Fold, and TS. TS is better at distinguishing the good and bad tool compounds.

Table S5, Figure S5-S7 are discussed in main manuscript.

Code for computing the TS is made available at www.github.com/novartis.

Supplemental References

1. Maglott, D., et al., *Entrez Gene: gene-centered information at NCBI.* Nucleic Acids Res, 2011. **39**(Database issue): p. D52-7.
2. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
3. Gene Ontology, C., *Gene Ontology Consortium: going forward.* Nucleic Acids Res, 2015. **43**(Database issue): p. D1049-56.
4. UniProt, C., *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue): p. D204-12.
5. Law, V., et al., *DrugBank 4.0: shedding new light on drug metabolism.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1091-7.
6. Bento, A.P., et al., *The ChEMBL bioactivity database: an update.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1083-90.
7. Liu, T., et al., *BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities.* Nucleic Acids Res, 2007. **35**(Database issue): p. D198-201.
8. Gilson, M.K., et al., *BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology.* Nucleic Acids Res, 2015.
9. Heller, S., et al., *InChI - the worldwide chemical structure identifier standard.* J Cheminform, 2013. **5**(1): p. 7.
10. Heller, S.R., et al., *InChI, the IUPAC International Chemical Identifier.* J Cheminform, 2015. **7**: p. 23.
11. Hoare, C.A.R., *An Axiomatic Basis for Computer Programming.* Communications of the Acm, 1969. **12**(10): p. 576-&.
12. Therneau, T., B. Atkinson, and B. Ripley, *Recursive Partitioning and Regression Trees.* cran.r-project.org.

13. Liaw, A. and M. Wiener, *Breiman and Cutler's Random Forests for Classification and Regression*. cran.r-project.org.
14. Meyer, D., et al., *Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. cran.r-project.org.