RESEARCH ARTICLE

REVISED

Monitoring drug promiscuity over time [v2; ref status: indexed, http://f1000r.es/4oa]

Ye Hu, Jürgen Bajorath

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Bonn, D-53113, Germany

v2 **First published:** 11 Sep 2014, **3**:218 (doi: 10.12688/f1000research.5250.1)
Latest published: 04 Nov 2014, **3**:218 (doi: 10.12688/f1000research.5250.2)

Abstract

Drug promiscuity and polypharmacology are much discussed topics in pharmaceutical research. Experimentally, promiscuity can be studied by profiling of compounds on arrays of targets. Computationally, promiscuity rates can be estimated by mining of compound activity data. In this study, we have assessed drug promiscuity over time by systematically collecting activity records for approved drugs. For 518 diverse drugs, promiscuity rates were determined over different time intervals. Significant differences between the number of reported drug targets and the promiscuity rates derived from activity records were frequently observed. On the basis of high-confidence activity data, an increase in average promiscuity rates from 1.5 to 3.2 targets per drug was detected between 2000 and 2014. These promiscuity rates are lower than often assumed. When the stringency of data selection criteria was reduced in subsequent steps, non-realistic increases in promiscuity rates from ~6 targets per drug in 2000 to more than 28 targets were obtained. Hence, estimates of drug promiscuity significantly differ depending on the stringency with which target annotations and activity data are considered.

Open Peer Review

Referee Status: ✓ ✓ ✓

	Invited Referees 1	2	3
REVISED **version 2** published 04 Nov 2014	*AWAITING PEER REVIEW*		
version 1 published 11 Sep 2014	✓ report	✓ report	✓ report

1 **Steffen Renner**, Novartis Institute for Biomedical Research Switzerland

2 **Stefan Laufer**, Eberhard Karls University Tubingen Germany

3 **Michael Walters**, University of Minnesota USA, **Kathy Nelson**, University of Minnesota USA

Discuss this article

Comments (0)

Corresponding author: Jürgen Bajorath (bajorath@bit.uni-bonn.de)

How to cite this article: Hu Y and Bajorath J. **Monitoring drug promiscuity over time [v2; ref status: indexed, http://f1000r.es/4oa]** *F1000Research* 2014, **3**:218 (doi: 10.12688/f1000research.5250.2)

Grant information: The author(s) declared that no grants were involved in supporting this work.

Competing interests: No competing interests were disclosed.

First published: 11 Sep 2014, **3**:218 (doi: 10.12688/f1000research.5250.1)
First indexed: 16 Oct 2014, **3**:218 (doi: 10.12688/f1000research.5250.1)

REVISED **Amendments from Version 1**

In this version, all comments by the third reviewer have been addressed (please, also see our comments in response to Reviewer 1).

1. It is "order of magnitude".
2. The majority of these 282 drugs were only annotated with a single target, which is stated in the revision.
3. Target family profiles have been provided for top five of 14 drugs with the largest changes in promiscuity in a new Table 4. Targets of these drugs belong to a wide range of related or unrelated families.
4. *F1000Research* has been informed to re-install the ZENODO link.
5. No pre-defined potency cut-off was applied in this study. As suggested, potency distributions are reported for compounds in three data sets with varying confidence levels in a new Figure 10.
6. The set of 26 substructures indicative of PAINS liability have been searched against all drugs. The promiscuity rates of PAINS-positive drugs over time are reported for three data sets in a new Figure 11.

See referee reports

Introduction

Promiscuous compounds specifically interact with multiple biological targets[1]. As such, they are distinct from compounds that exhibit assay liabilities or engage in various non-specific interactions. Compound promiscuity is often functionally relevant and represents the molecular origin of polypharmacology[2], a concept that experiences increasing interest in drug discovery. Drugs are often, but not always, found to act on multiple targets and modulate multiple cellular pathways and/or signaling cascades. Such effects might often substantially contribute to therapeutic efficacy, for example, in cancer treatment[3]. The potentially far reaching consequences of drug polypharmacology for therapy, the frequency of these effects, and likely pros and cons are just beginning to be understood.

Experimentally, promiscuity can be assessed by profiling of compounds or drugs on arrays of biological targets[1,2], although such studies might often only provide an incomplete picture of *in vivo* effects. The same applies to computational estimates of promiscuity. Given the increasingly large amounts of compound activity data that are becoming available, the promiscuity of drugs and bioactive compounds can be explored through data mining by systematically evaluating activity annotations[1]. For the assessment of compound and drug promiscuity, public databases such as ChEMBL[4], the major repository of compounds and activity data from medicinal chemistry, the PubChem BioAssay database[5], the major repository of screening data, and DrugBank[6], which collects approved and experimental drugs, have become indispensible resources.

Computational analyses reported thus far have suggested different degrees of promiscuity among bioactive compounds and drugs, dependent on the compound sources used and the methods applied. For example, drug-target network analysis has indicated that a drug might on average act on two targets[7]. Other computational studies have suggested that drugs might on average interact with two to seven targets depending on the target classes the drugs are active against[8]. In addition to varying compound sources and analysis concepts, taking activity measurement characteristics and data confidence criteria into account is also of critical importance for compound promiscuity analysis. For example, it has been shown that the increase in the number of compounds with activity against targets from different families in ChEMBL has mostly resulted from assay-dependent IC_{50} but not (assay-independent) K_i measurements (equilibrium constants)[9]. In addition, by exclusively considering high-confidence activity data, it has been found that the majority of promiscuous bioactive compounds interact with two to five targets from the same target family, are predominantly active in sub-μM range, and display potency differences within one or two orders of magnitude against their targets[10]. This represents a prevalent promiscuity profile among bioactive compounds. On the basis of high-confidence activity data, it has also been calculated that compounds from ChEMBL interact on average with one to two targets and compounds from PubChem confirmatory assays with two to three targets[11]. By contrast, target annotation analysis has suggested that approved drugs interact on average with close to six targets, whereas experimental drugs (including candidates in clinical trials) interact with one to two targets[11]. The reasons for this apparent discrepancy in target numbers between drugs at different development stages are currently unknown. As increasing amounts of activity data become available, it is likely that recently detected promiscuity rates might further increase. However, the magnitude of such increases as a consequence of data incompleteness[12] is difficult to predict, especially considering the low promiscuity rates that can currently be confirmed on the basis of high-confidence data[1,11].

In this study, we further extend the computational analysis of promiscuity by evaluating the progression of drug promiscuity rates over time, which required a systematic assessment of activity records with release dates. Different data selection criteria were applied and the calculated promiscuity rates were compared to available drug target annotations. Small to moderate increases in drug promiscuity over time were detected when high-confidence activity data were considered. Lowering the stringency of data selection criteria led to unrealistic estimates of promiscuity rates and their progression.

Materials and methods

Data collection

From ChEMBL (release 18)[4], compounds with direct interactions (i.e., assay relationship type "D") with human targets at the highest confidence level (i.e., assay confidence score 9) were collected. The two ChEMBL parameters 'assay relationship type' and 'assay confidence score' qualify and quantify the level of confidence that the activity against a given target is evaluated in a relevant assay system, respectively. Accordingly, type "D" and score 9 represent the highest level of confidence for activity data. In addition, two types of activity measurements were considered including assay-independent equilibrium constants (K_i values) and assay-dependent IC_{50} values. To ensure a high level of data integrity, only compounds with explicitly defined K_i or IC_{50} values were selected. Hence, approximate measurements such as ">", "<", and "~" were disregarded. Compounds with multiple K_i or IC_{50} measurements for the same target were retained if all these values fell within the same order of magnitude. Otherwise, the target activity was omitted from further consideration. Structures of all qualifying bioactive compounds were standardized using the Molecular Operating Environment (MOE)[13]

and transformed into canonical SMILES strings[14]. The so assembled compound set exclusively utilized high-confidence activity data (high-confidence data set).

Approved small molecule drugs with available structure and activity information were collected from the latest release of DrugBank (version 4.1)[6]. To synchronize the activity analysis in ChEMBL and DrugBank, all reported 'drug action' targets, metabolizing enzymes, transporters, and carriers were assembled for approved drugs. In some instances, drug target activity might refer to a group of related proteins. For example, atomoxetine was annotated with N-methyl-D-aspartate (NMDA) receptor including seven subtypes. Accordingly, seven UniProt[15] accession IDs (UniProtIDs) were associated with NMDA receptor. Thus, the maximal number of target annotations was collected for approved drugs on the basis of UniProtIDs. Drug structures were also standardized using MOE and transformed into canonical SMILES strings.

Monitoring drug activity records over time

Most compound activity data in ChEMBL are extracted from medicinal chemistry literature and patent sources[4]. Therefore, the release dates of activity data are frequently recorded in this database. However, DrugBank does not report dates for individual target annotations. To systematically monitor drug promiscuity over time, all approved drugs from DrugBank were mapped to ChEMBL by comparing canonical SMILES strings. If a drug (D) and a bioactive compound (B) shared the same SMILES string, a match was obtained. It should be noted that the name of a drug in DrugBank and ChEMBL might differ (i.e., matching by drug/compound name is not reliable). For each match, activity data release dates of compound B were recorded and assigned to drug D. Each activity record represented a target annotation (the terms target activity and target annotation are synonymously used). For instance, if compound B was reported to be active against target I in 2001, target II in 2005, and target III in 2009, the cumulative activity records for drug D consisted of target I in 2001, targets I and II in 2005, and targets I, II, and III in 2009. Thus, the promiscuity rate of D increased over time from 1 to 3. All activity records were organized into 14 time intervals, as illustrated in Figure 1. All activity records reported before 2000 were assigned to 2000, the starting point of our analysis, and all activity data released after 2012 were assigned to the last period ">2012". For each time interval, the cumulative activity profile was recorded. Hence, changes in the promiscuity rate of a drug were successively determined over the years. Cumulative activity profiles were compared to target annotations available in DrugBank.

Low-confidence data sets

In order to investigate the effect of activity data confidence levels on drug promiscuity, two data sets with lower confidence were assembled from ChEMBL (release 18). For the generation of low-confidence data sets, two criteria that influence the compound data integrity, i.e., the confidence level of activity and the type of activity measurements were disregarded in subsequent steps. In low-confidence set 1, the criterion of activity measurement type was not considered. Hence, in addition to K_i and IC_{50} values, all other potency annotations were equally considered (including "%max", "Efficacy", "EC_{50}", "K_d", and "Residual Activity") for all compounds with 'direct interactions' with human targets and assay confidence score 9. In addition, the consistency and quality of potency measurements was not considered. In low-confidence set 2, the confidence level of activity (assay relationship type and assay confidence score) was not considered, in addition to the type of activity measurements. Therefore, the stringency of activity data and compound selection decreased from the high-confidence set over low-confidence set 1 to low-confidence set 2.

Progression of drug promiscuity over time was systematically evaluated on the basis of all three data sets.

Results and discussion

Bioactive compounds and approved drugs

On the basis of the selection criteria described above, a total of 143,424 bioactive compounds with high-confidence activity data were obtained from ChEMBL. These compounds were active against 1376 different targets and yielded 219,602 compound-target

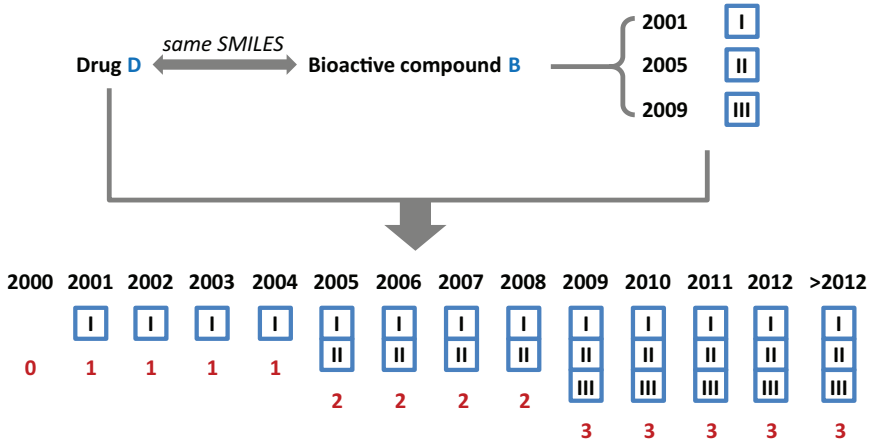


Figure 1. Organization of activity records. The organization of the activity records for a drug over different years is schematically illustrated. Drug D and a bioactive compound B share the same SMILES string (D is mapped to B). The activity records of compound B are extracted from ChEMBL. B is reported to be active against target I in 2001, II in 2005, and III in 2009. These activity records are then assigned to drug D and organized into 14 time intervals (12 of which represent individual years, except 2000 (see text) and >2012). For each interval, a cumulative activity profile is generated for D and recorded. The total number of activity annotations is given in red.

interactions, as reported in Table 1. Furthermore, from DrugBank 4.1, 1429 approved drugs were obtained that were annotated with 1657 target proteins corresponding to 10,679 drug-target interactions (Table 1). Thus, there were nearly 100 times more bioactive compounds than approved drugs. However, with 1657 targets, drugs covered a larger target space than bioactive compounds (1376 targets). On average, a bioactive compound was active against 1.5 targets, whereas an approved drug was annotated with 7.5 targets. Compared to a recent analysis of promiscuity rates[11], which also included a previous release of DrugBank, the average promiscuity rate of approved drugs further increased from 5.9 to 7.5, while the degree of promiscuity among bioactive compounds remained essentially constant.

Table 1. Data sets.

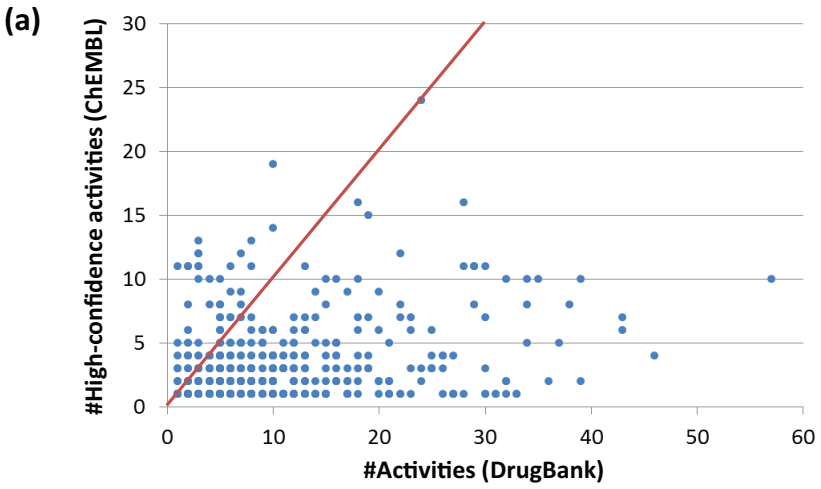
Number of	DrugBank 4.1	ChEMBL release 18
Drugs/compounds	1429	143,424
Targets	1657	1376
Interactions	10,679	219,602

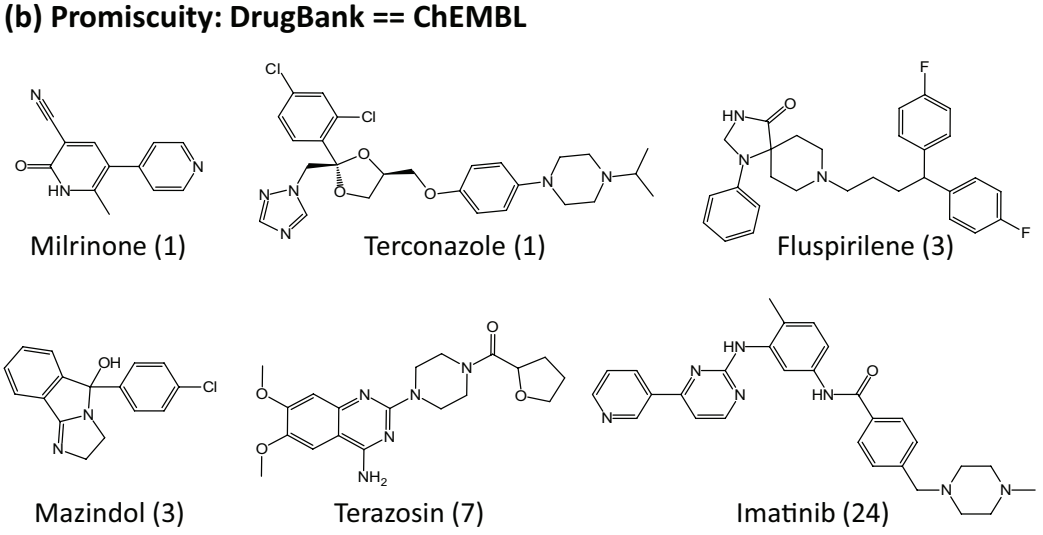
For DrugBank 4.1 (drugs) and ChEMBL release 18 (compounds), the number of drugs/compounds, targets the drugs/compounds were active against, and the total number of interactions is reported.

To monitor drug promiscuity over time, all approved drugs were mapped to bioactive compounds in ChEMBL for which release dates of activity records were reported (as detailed in the Methods section). For 518 of the 1429 approved drugs taken from DrugBank, high-confidence activity data released over different years were found in ChEMBL. These 518 drugs provided the basis for our time-dependent promiscuity analysis.

Data inconsistency

For the 518 qualifying drugs, we first compared their target annotations in DrugBank and the total number of targets derived from high-confidence activity records in ChEMBL. As reported in Figure 2a, most of the drugs had different numbers of targets in





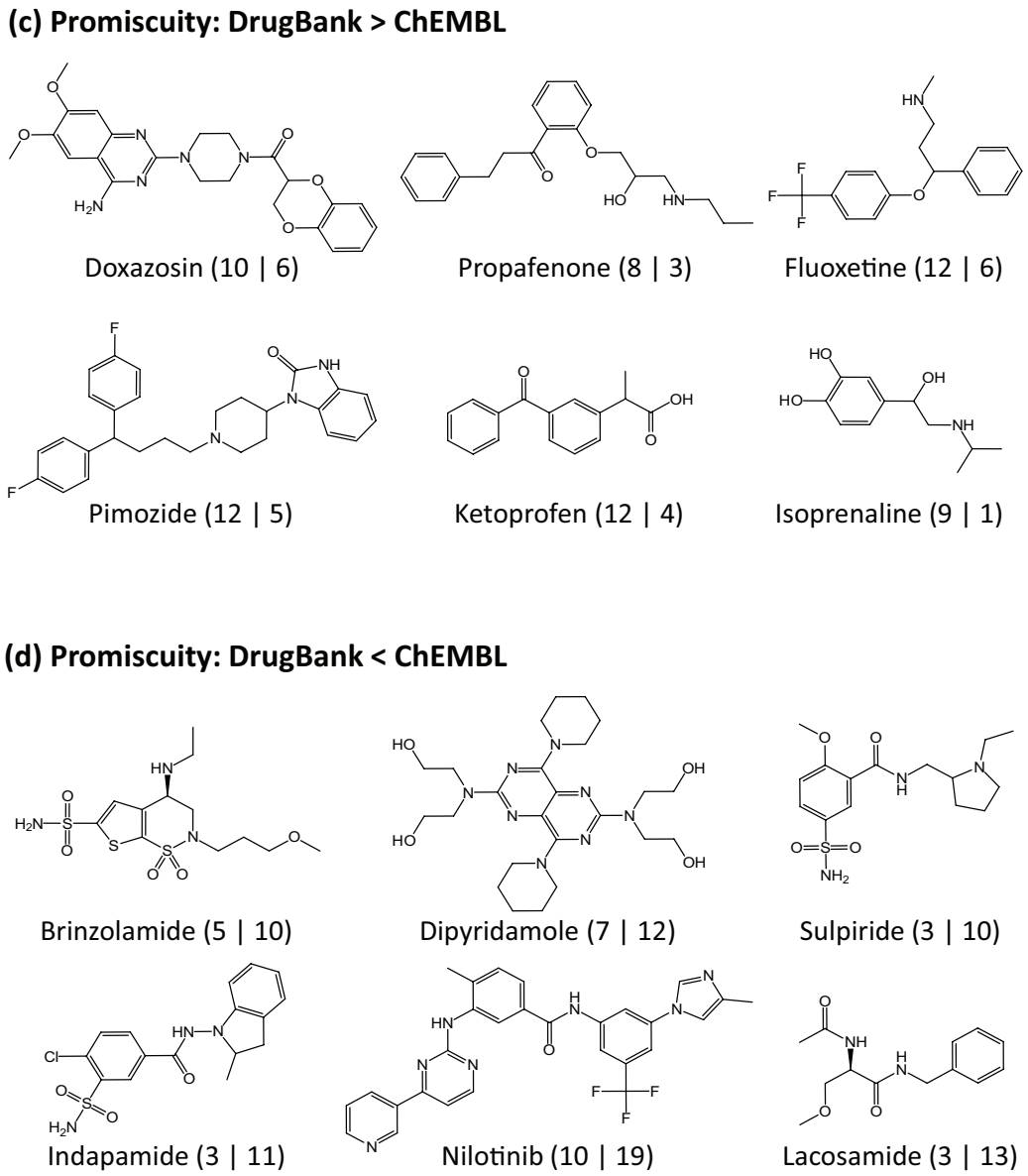


Figure 2. Drug promiscuity in DrugBank vs ChEMBL. (**a**) For 518 qualifying drugs, the number of targets reported in DrugBank and the number of high-confidence activity annotations in ChEMBL are compared in a scatter plot. Each dot represents a drug. The diagonal (indicating perfect correlation) is drawn in red. In (**b**), (**c**), and (**d**), exemplary drugs are shown that had the same number of targets in DrugBank and ChEMBL (i.e., the same promiscuity rate), a higher promiscuity rate in DrugBank, and a higher rate in ChEMBL, respectively. For each drug in (**b**), the number of targets is given. For each drug in (**c**) and (**d**), the numbers of targets reported in DrugBank and ChEMBL are compared. For example, "10 | 6" indicates that the drug was annotated with 10 targets in DrugBank and with six in ChEMBL.

the two databases. Only 32 drugs (~6%) were found to have the same number of target annotations in DrugBank and ChEMBL. The total number of target annotations of a drug represented its promiscuity rate. A total of 439 drugs had higher promiscuity rates in DrugBank than in ChEMBL. Opposite observations were only made for 47 drugs. On average, the 518 drugs were annotated with ~10.1 targets in DrugBank and ~3.2 targets derived from high-confidence ChEMBL activity records. Hence, promiscuity rates in DrugBank were much higher than in ChEMBL. Exemplary drugs having the same or different degrees of promiscuity in DrugBank and ChEMBL are shown in Figure 2b–Figure 2d.

Differences in promiscuity rates were quantified, as reported in Figure 3a. Among the 486 drugs (~94%) with varying degrees of promiscuity in DrugBank and ChEMBL, 48 and 58 drugs differed by one and two targets, respectively. By contrast, the promiscuity

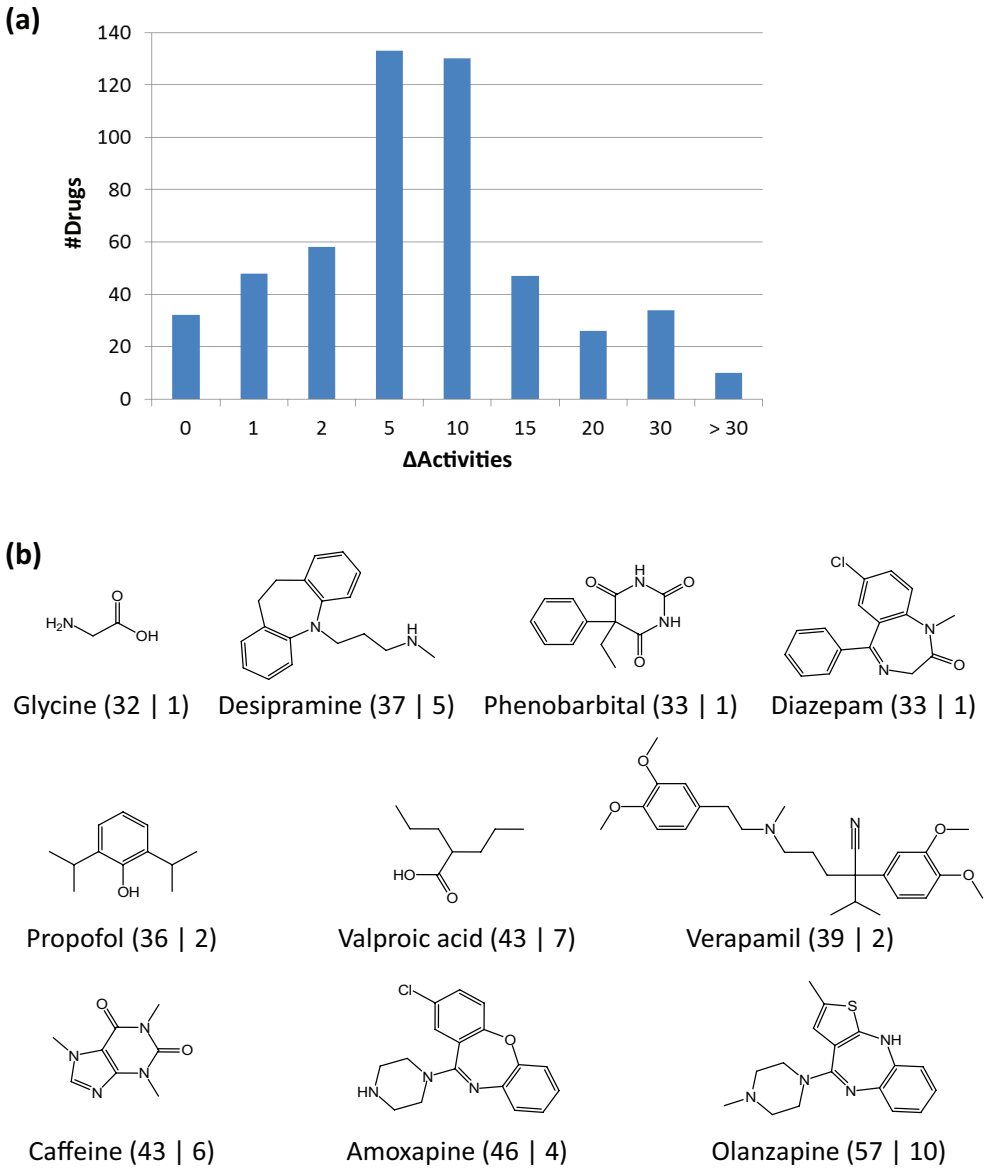


Figure 3. Promiscuity variation in DrugBank vs ChEMBL. (**a**) Reported is the distribution of promiscuity differences (ΔActivities) between DrugBank and ChEMBL for 518 drugs. (**b**) Shown are 10 drugs with the largest difference in promiscuity (ΔActivities > 30). Target annotations are represented according to Figure 2c and Figure 2d.

rates of nearly half of the drugs (247; ~48%) varied by more than five targets. Moreover, for the 10 drugs shown in Figure 3b, the promiscuity rates differed by more than 30 targets, which reflected a particularly high degree of data inconsistency. All of these drugs were annotated with many more targets in DrugBank than targets derived from high-confidence activity records in ChEMBL. The extreme case was olanzapine the promiscuity rate of which differed by 47 targets between the two databases.

In addition to comparing the number of target annotations, the activity profiles of drugs were further examined to determine the consistency of the annotations. As reported in Figure 4, 175 drugs (~34%) had non-overlapping sets of targets in these two databases, which was another surprising finding. The remaining 343 drugs had overlapping yet distinct target sets. However, the majority of these drugs shared only one or two targets, reflecting substantial discrepancies between target annotations.

For the study of changes in drug promiscuity over time, accessing original activity records and their release dates was an essential requirement, as rationalized above. Such information is not available in DrugBank.

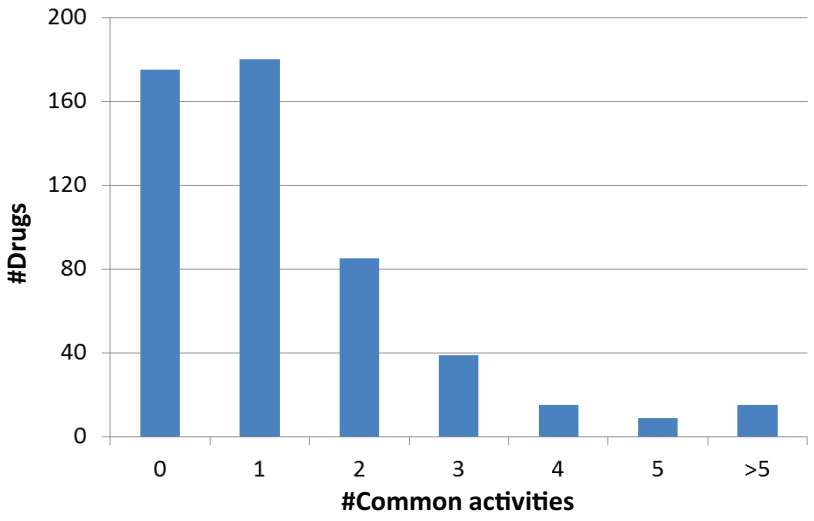


Figure 4. Comparison of activity profiles in DrugBank and ChEMBL. The activity profiles of 518 drugs in DrugBank and ChEMBL are compared. Reported is the number of drugs sharing increasing numbers of activities (#Common activities) in the two databases.

Drugs on a time course

Next, we organized the 518 drugs on the basis of activity record release dates. Drugs were assigned to the individual time intervals in which high-confidence activity data were first published. For example, if the first activity record of a given drug was detected in 2005, the drug was assigned to the 2005 interval and traced during all subsequent years. The cumulative number of drugs in different time intervals is reported in Figure 5a. By 2000, high-confidence activity data were publicly available for 78 drugs. From 2000 to 2001, activity data became available for 26 additional drugs. The number of drugs for which qualifying activity records were available in subsequent years ranged from 20 to 64, with an average of ~34 drugs per interval. The largest increase was detected for 2007/2008. The time period for which the activity records were assembled spanned a maximum of 24 years (for captopril, from 1981 to 2005), with an average of 3.3 years per drug. Exemplary drugs for which activity records were first reported before 2000 and after 2008 are shown in Figure 5b and Figure 5c, respectively.

Changes in drug promiscuity over time

For individual time intervals, the distribution of drug promiscuity rates was determined, as reported in Figure 6a. The box plots reveal an increase in drug promiscuity rates over time, with a maximal rate of six targets per drug in 2000 and 24 targets per drug in interval >2012. However, median promiscuity rates only slightly increased from one (until 2005) to two (beginning in 2006) targets per drug. The distribution of average promiscuity rates is shown in Figure 6b, which slightly but steadily increased over time from 1.5 to 3.2 targets per drug. The larger relative increase of average than median promiscuity rates indicated that the average values were influenced by small numbers of drugs with large numbers of targets, i.e., a small subset of highly promiscuous drugs, consistent with earlier observations[11]. On the basis of median values, detectable increases in drug promiscuity over time were limited.

Changes in promiscuity over time were also monitored for individual drugs. For each drug, the increase in the cumulative promiscuity rates from its first to its most recent activity records was determined (for the hypothetical example in Figure 1, the increase in promiscuity rates is 2). For the 518 drugs, increases are reported in Table 2. Surprisingly, for 282 drugs (~54%), no increase in promiscuity was detected on the basis of high-confidence activity records. This indicated that the majority of these drugs did not receive additional high-confidence activity annotations since their first records were released. Of 282 drugs, 203 drugs were only annotated with a single

Table 2. Increasing promiscuity.

Increase in promiscuity rates	#Drugs (%)
0	282 (54.4%)
1	84 (16.2%)
2	44 (8.5%)
3	39 (7.5%)
4	15 (2.9%)
5	15 (2.9%)
6	5 (1.0%)
7	8 (1.5%)
8	6 (1.2%)
9	6 (1.2%)
10	8 (1.5%)
> 10	6 (1.2%)

The number (percentage) of drugs with increasing promiscuity rates (i.e., number of targets) is reported.

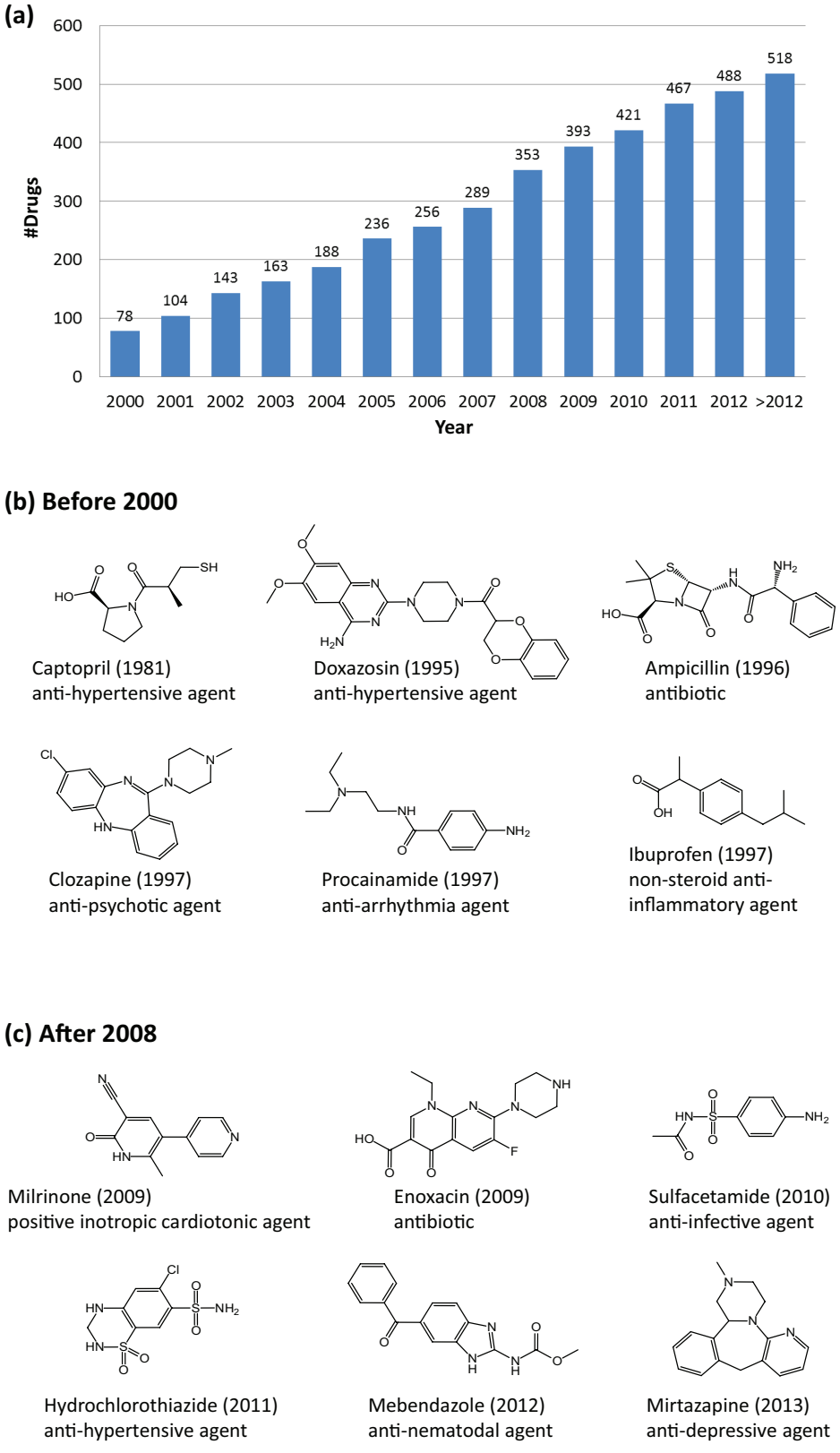


Figure 5. Monitoring high-confidence drug activity data over time. (**a**) Reported is the cumulative number of drugs for which high-confidence activity data became available in different years. In (**b**) and (**c**), six exemplary approved drugs are shown for which high-confidence activity data were first recorded before 2000 or after 2008, respectively. For each drug, its name, year of first data report, and therapeutic indication are provided.

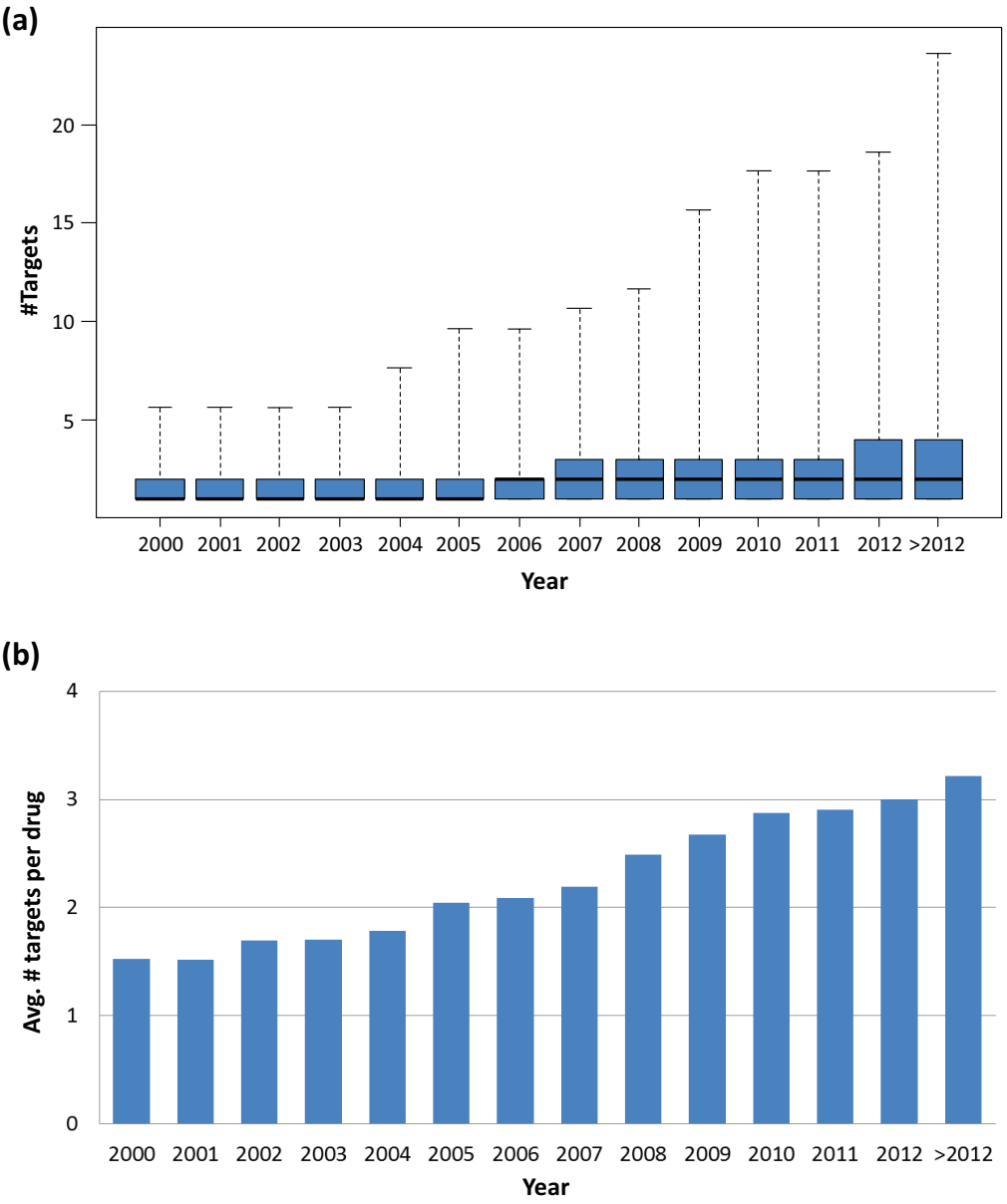


Figure 6. Monitoring drug promiscuity over time. (**a**) Box plots capture the distribution of the number of targets per drug in different years. Each box plot reports the smallest value (bottom line), lower quartile (lower boundary of the box), median (thick horizontal line), upper quartile (upper boundary of the box), and the largest value (top line). (**b**) Reported are average numbers of targets per drug in different years.

target. Exemplary drugs with constant promiscuity rates are shown in Figure 7. For the remaining 236 drugs, increasing numbers of targets were detected. However, in most cases, the increase in target numbers was limited, i.e., the promiscuity rates of 197 drugs increased by one to five targets (Table 2). There were only 14 drugs with an increase in promiscuity rates by 10 or more targets. Five drugs with largest increase in promiscuity rates are shown in Figure 8. For example, the promiscuity rate of imatinib increased from one in 2002 to 24 (>2012), with 11 new targets reported between 2008 and 2009. The drugs in Figure 8 belonged to the subset of highly promiscuous drugs that statistically influenced the calculation of average promiscuity rates, as discussed above.

Drug promiscuity across different target families was also assessed. For the 236 drugs with increases in promiscuity rates over time, their targets were assigned to families and the number of target

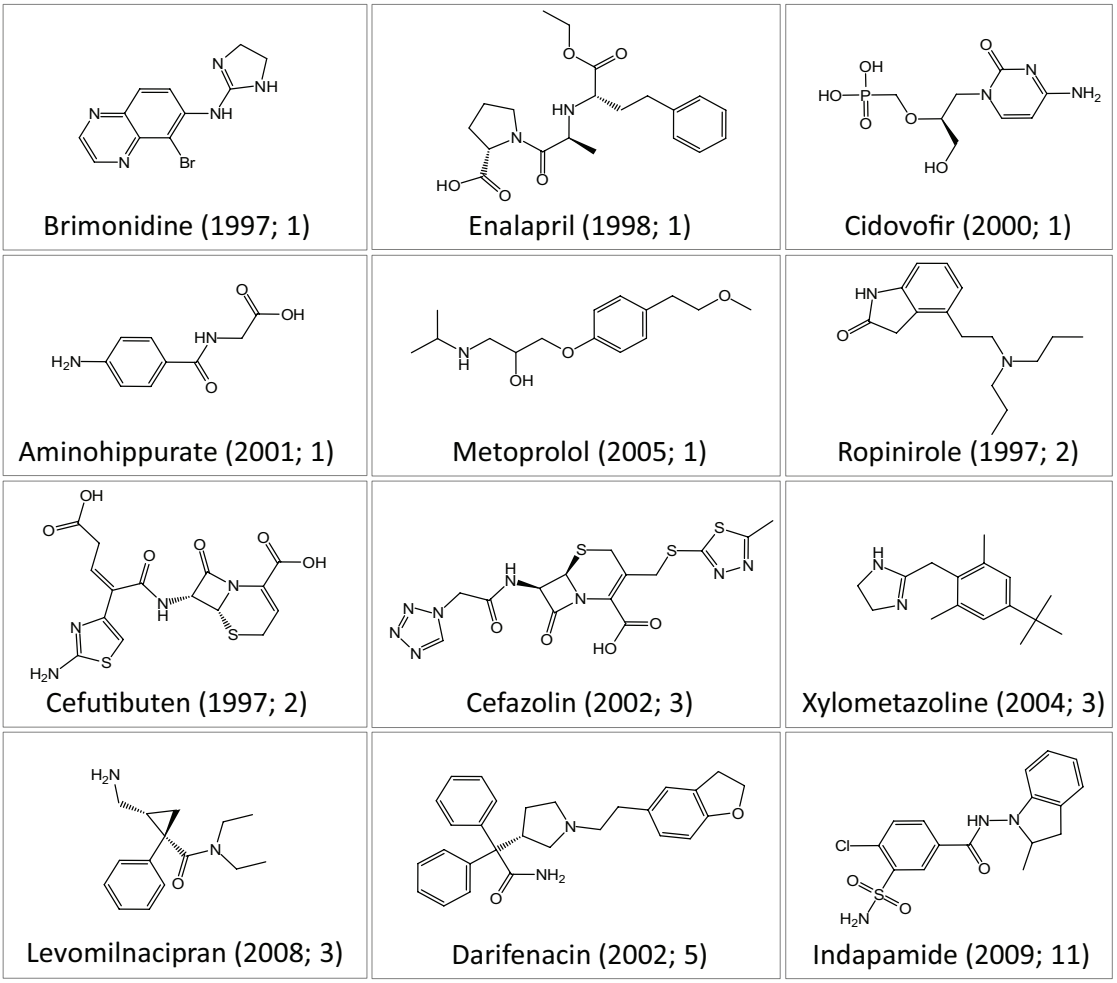


Figure 7. Drugs with constant promiscuity over time. Shown are 12 exemplary drugs having a constant promiscuity rate on the basis of high-confidence activity data. For each drug, the year of its first activity report and the number of targets it was active against are given. For example, brimonidine was first reported to be active against a single target in 1997.

families was determined and followed over time. Table 3 reports the number of drugs with increasing target family annotations. For the majority of drugs, the number of target families increased by one or two. For top five drugs with largest changes in promiscuity (Figure 8), their target family profiles are provided in Table 4. The first activity records of all these five drugs belonged to only one target family including protein kinase family, GPCR subfamily, and transporter subfamilies. Compared to their most recent activity records, the number of target families increased by three to nine, spanning a wide range of related or unrelated target families. It indicated that these drugs might have been tested against a large panel of targets over time and that a number of activities have been confirmed at a high level of confidence. For 47 drugs, the number of target families remained constant.

Drug promiscuity on the basis of low-confidence data sets

Two compound sets with lower activity data confidence were also assembled from ChEMBL, as described above. The composition of

Table 3. Promiscuity increase across different target families.

Increase in promiscuity across target families	#Drugs (%)
0	47 (19.9%)
1	105 (44.5%)
2	47 (19.9%)
3	21 (8.9%)
4	9 (3.8%)
5	4 (1.7%)
> 5	3 (1.3%)

The number (percentage) of drugs with increasing target family promiscuity (i.e., increasing number of protein families the drug targets belong to) is reported.

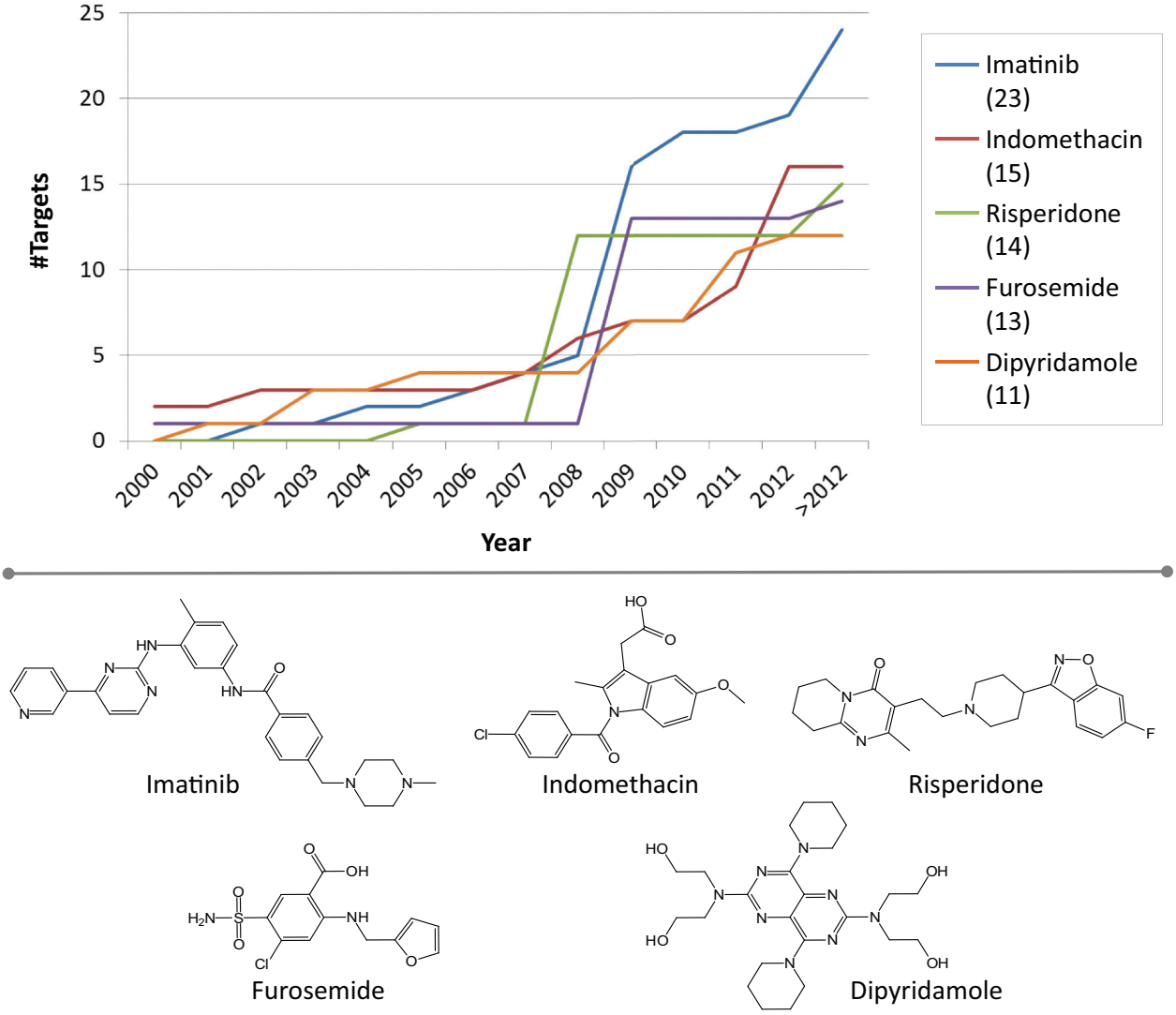


Figure 8. Top five drugs with largest changes in promiscuity. For the five drugs with largest changes in promiscuity over time, cumulative numbers of targets are reported for different years (top). For each drug, the overall difference in target annotations is given in parentheses. The structures of these drugs are shown at the bottom.

these sets is summarized in Table 5. Low-confidence set 1 in which the types of activity measurements were not specified contained a total of 605,206 compounds active against 2144 targets, yielding more than 2,600,000 interactions. Low-confidence set 2 in which, in addition, the confidence level of activity was undefined consisted of a larger number of 936,924 compounds active against 3934 targets, yielding more than 6,000,000 interactions. All 518 drugs were mapped to these two low-confidence data sets. The cumulative distribution of these drugs over time is reported in Figure 9a. The number of drugs with low-confidence activity annotations in 2000 increased from 78 (high-confidence set) to 194 (low-confidence set 1) and 335 (low-confidence set 2). On average, ~26 and ~15 drugs became available during each year for low-confidence set 1 and 2, respectively.

Figure 9b compares the distribution of average drug promiscuity rates for the three data sets over time. In contrast to the high-confidence data set in which drug promiscuity only slightly increased over the years, the average promiscuity rates of drugs in both low-confidence sets were higher and significantly increased. In low-confidence set 2, the average promiscuity rate was 6.3 targets per drug in 2000 and further increased to 28.2 targets (>2012). Thus, by reducing the stringency of selection criteria for activity records, high average promiscuity rates were obtained. The large increases in average promiscuity rates seen in Figure 9b ultimately resulted in 18 (low-confidence set 1) or nearly 30 (set 2) targets per drug are most likely artificial in nature. The comparison reveals how the choice of different activity data selection criteria, or the lack of well-defined criteria, might bias promiscuity analysis.

Table 4. Target family profiles for top five drugs with largest changes in promiscuity.

Drug name	#Targets	#Families	Family list
Imatinib	24	7	ATP binding cassette transporters; Carbonic anhydrases; Multi antimicrobial extrusion (MATE) transporters; NAD(P)H dehydrogenases (quinone); Organic cation transporters; Ser_Thr protein kinases; **Tyr protein kinases**
Indomethacin	16	10	ATP binding cassette transporters; Aldo/keto reductases; Glyoxalases I; Intercrines; **Lipoxygenases**; MAPEGs; Organic cation transporters; Organo anion transporters; Potassium ion channels; Short-chain dehydrogenases/reductases (SDR)
Risperidone	15	4	**Monoamine GPCRs**; Multi antimicrobial extrusion (MATE) transporters; Organic cation transporters; Sodium:neurotransmitter symporters (SNF)
Furosemide	14	4	**Bile acid:sodium symporters (BASS)**; Carbonic anhydrases; Carboxylic acid GPCRs; Short-chain dehydrogenases/reductases (SDR)
Dipyridamole	12	5	**ATP binding cassette transporters**; Multi antimicrobial extrusion (MATE) transporters; Organic cation transporters; Class C PPases; Phosphodiesterases

For the top five drugs with largest changes in promiscuity over time, the total number of targets and families is reported. The family with the first target annotation of a drug is shown in bold. Target family abbreviation: MAPEG, membrane-associated proteins in eicosanoid and glutathione metabolism.

Table 5. Low-confidence data sets.

Number of	Set 1	Set 2
Compounds	605,206	936,924
Targets	2144	3934
Interactions	2,639,767	6,295,086

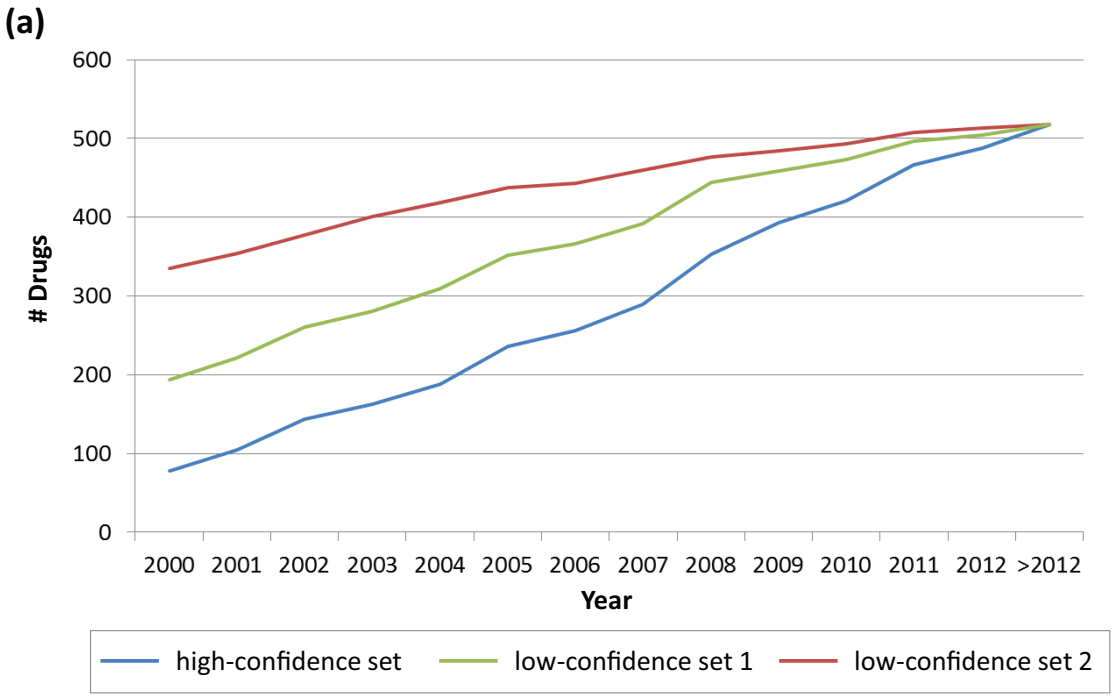
For both low-confidence sets (see text for details), the number of compounds, targets, and interactions is reported.

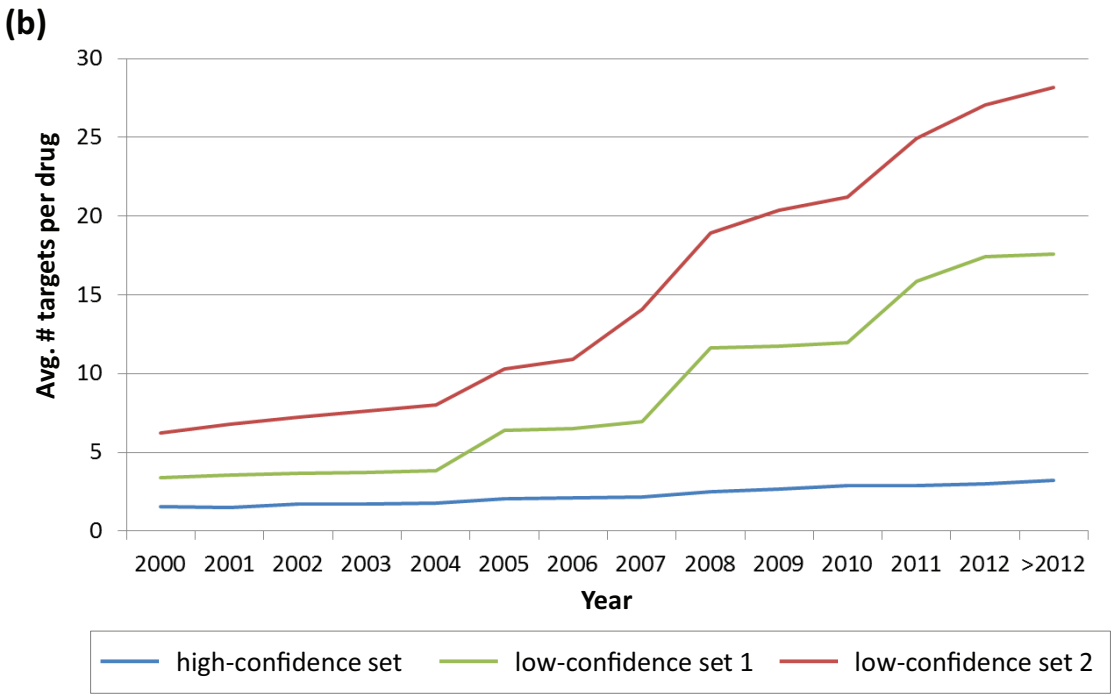
Imatinib represented a striking example for the presence of unreliable target annotations under non-stringent data selection criteria (Figure 9c). In both low-confidence sets 1 and especially 2 dramatic increases were observed between 2005 and 2008, ultimately leading to 406 and 689 targets for imatinib, respectively (hence exceeding the total number of targets in the human kinome). By contrast, on the basis of high-confidence activity data, the final (>2012) promiscuity rate of imatinib was 24.

In addition, the distributions of potency values were compared across different data sets, as reported in Figure 10. It should be noted that only K_i and IC_{50} values were considered here, although all other types of potency annotations were included in low-confidence sets 1 and 2. In general, the distribution of the high-confidence set was comparable to the low-confidence set 1. The majority of negative logarithmic potency values ranged from ~4.5 (i.e., ~32 μM) to 7.0 (i.e., 100 nM). By contrast, the majority of potency values in low-confidence set 2 were confined to a narrow range.

PAINS substructures

Compounds that are reactive or cause other non-specific effects in a variety of assays are typically false positives and have been termed pan assay interference compounds (PAINS)[16]. Baell and Holloway described a set of 26 substructures that are indicative of PAINS liability[16]. This set of substructures was utilized as a filter to identify drugs that contain PAINS substructures in our three data sets with varying activity confidence levels. A total of 23 drugs (i.e., ~4.4%) were found to contain PAINS substructures. Figure 11 reports the average promiscuity rates of PAINS-positive drugs compared to all available drugs over time. It can be seen that drugs with potential PAINS liability in two low-confidence sets displayed much higher degrees of promiscuity than the global rates. For example, the latest average promiscuity rate (i.e., >2012) of drugs containing PAINS substructures in low-confidence set 2 increased from ~28.2


(a)
600
500
400
300
200
100
0
Drugs
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 >2012
Year
high-confidence set
low-confidence set 1
low-confidence set 2


(b)
30
25
20
15
10
5
0
Avg. # targets per drug
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 >2012
Year
high-confidence set
low-confidence set 1
low-confidence set 2

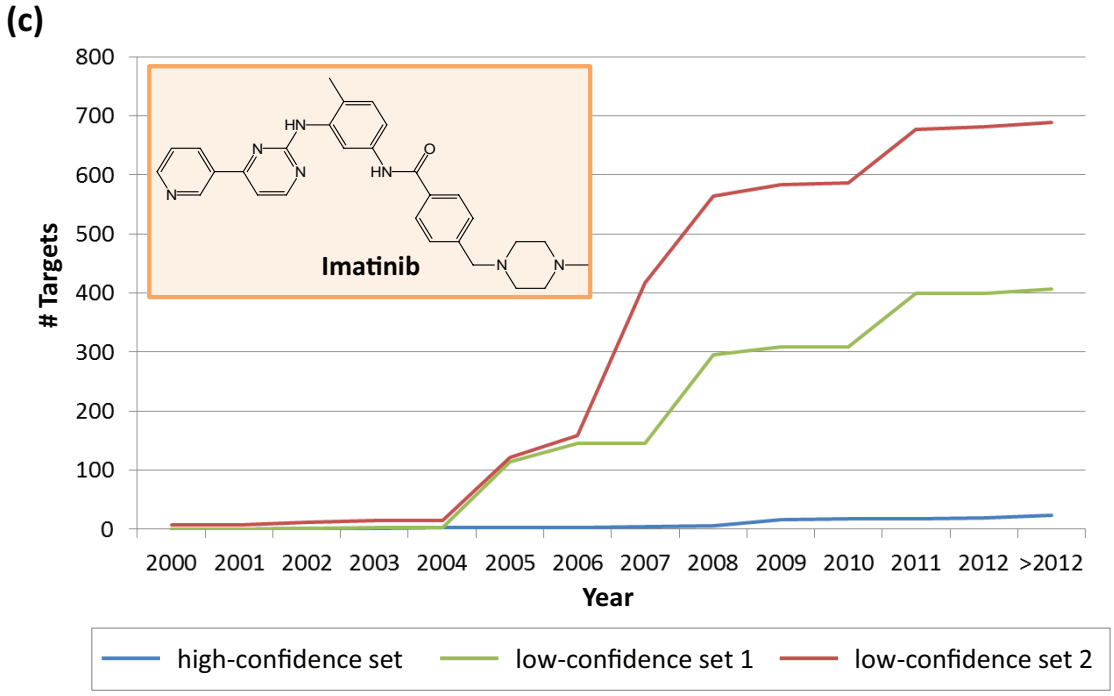


Figure 9. Drugs in sets with varying activity confidence levels. (**a**) Reported is the cumulative number of drugs in three data sets of varying confidence levels over time. (**b**) Shown is the distribution of average promiscuity rates for drugs in these three data sets. (**c**) For imatinib, the cumulative number of targets is reported for different years.

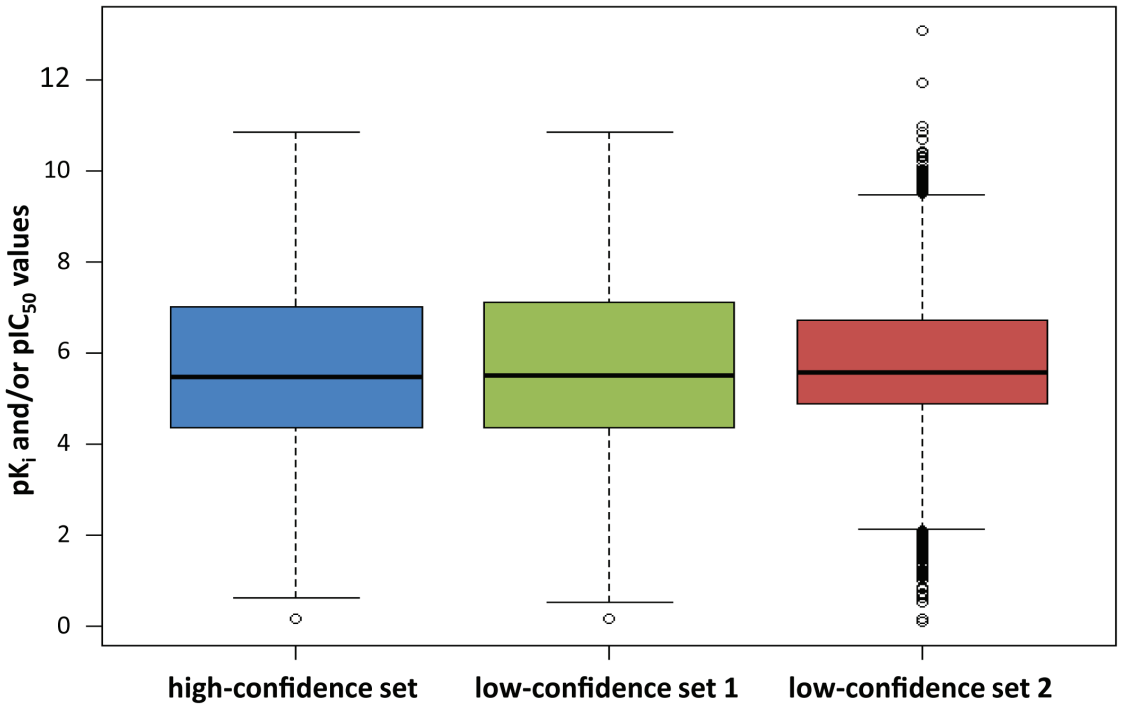


Figure 10. Potency distribution. The distribution of pK_i and/or pIC_{50} values in three data sets with varying confidence levels is reported in box plots. Each box plot provides the lowest potency value within the 1.5 interquartile range of the lower quartile (bottom line), lower quartile (lower boundary of the box), median value (thick line), upper quartile (upper boundary of the box), and the highest value within the 1.5 interquartile range of the upper quartile (top line). Potency values falling outside these ranges are indicated by empty circles.

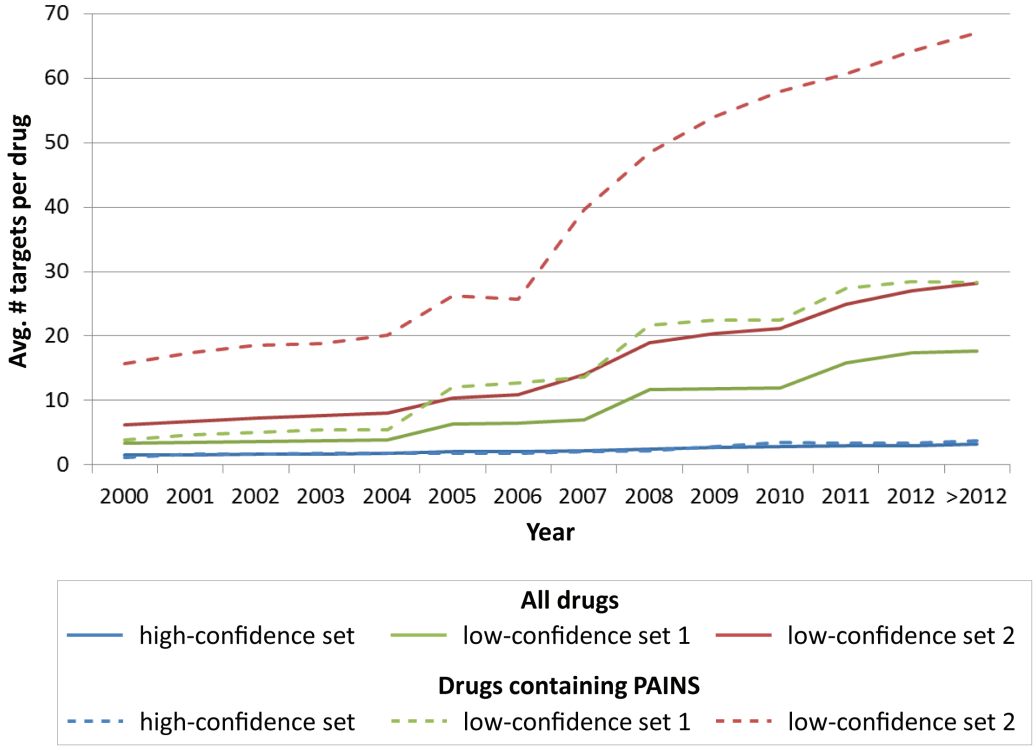


Figure 11. Drugs containing PAINS substructures. Shown is the distribution of average promiscuity rates over time for all drugs (solid lines) and for drugs that contain PAINS substructures (dashed lines) in three data sets with varying activity confidence levels, respectively.

to ~67.0. By contrast, PAINS-positive drugs in the high-confidence set displayed a comparable degree of promiscuity. These findings suggest that PAINS-related effects might also be controlled by applying rigorous data confidence criteria.

Conclusions

The analysis reported herein was designed to monitor drug promiscuity over time through computational data mining. It was facilitated by systematically collecting available activity records with release dates for approved drugs from the ChEMBL database. For more than 500 drugs, it was possible to assess promiscuity rates over a time course. Current promiscuity rates derived from high-confidence ChEMBL activity records are typically much lower than those calculated from target annotations available in DrugBank, which should merit further consideration. Data selection criteria for the assignment of drug targets might at least in part be responsible for the observed differences. On the basis of high-confidence activity data, an increase in the average drug promiscuity rates from only 1.5 to 3.2 targets per drug was observed. The magnitude of average promiscuity rates was influenced by a small subset of highly promiscuous drugs. Thus, increases in average drug promiscuity over time were generally small. However, they frequently involved targets from at least two families. By contrast, for low-confidence data sets, calculated promiscuity rates were much higher and dramatic increases in apparent drug promiscuity were observed over the years. From our point of view, such trends are unreliable. These observations further emphasize the need for well-defined and stringent data selection criteria for promiscuity analysis. Taken together, the findings reported herein reveal a small to moderate increase in detectable drug promiscuity over time while the volumes of compound activity data rapidly grow.

Data availability

The high-confidence and the two low-confidence drug data sets are made available in ZENODO. For each drug in each set, the ChEMBL activity records are provided for individual time intervals.

ZENODO: Drug activity data, doi: 10.5281/zenodo.11576[17]

Author contributions

JB conceived the study, YH planned and performed the analysis, YH and JB wrote the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

References

1. Hu Y, Bajorath J: **Compound promiscuity: what can we learn from current data?** *Drug Discov Today.* 2013; **18**(13–14): 644–650.
PubMed Abstract | Publisher Full Text
2. Boran AD, Iyengar R: **Systems approaches to polypharmacology and drug discovery.** *Curr Opin Drug Discov Devel.* 2010; **13**(3): 297–309.
PubMed Abstract | Free Full Text
3. Knight ZA, Lin H, Shokat KM: **Targeting the cancer kinome through polypharmacology.** *Nat Rev Cancer.* 2010; **10**(2): 130–137.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Gaulton A, Bellis LJ, Bento AP, *et al.*: **ChEMBL: a large-scale bioactivity database for drug discovery.** *Nucleic Acids Res.* 2012; **40**(Database issue): D1100–D1107.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Wang Y, Xiao J, Suzek TO, *et al.*: **PubChem's BioAssay Database.** *Nucleic Acids Res.* 2012; **40**(Database issue): D400–D412.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Law V, Knox C, Djoumbou Y, *et al.*: **DrugBank 4.0: shedding new light on drug metabolism.** *Nucleic Acids Res.* 2014; **42**(Database issue): D1091–D1097.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Yildirim MA, Goh KI, Cusick ME, *et al.*: **Drug-target network.** *Nat Biotechnol.* 2007; **25**(10): 1119–1126.
PubMed Abstract | Publisher Full Text
8. Jalencas X, Mestres J: **On the origins of drug polypharmacology.** *Med Chem Comm.* 2013; **4**(1): 80–87.
Publisher Full Text
9. Hu Y, Bajorath J: **Growth of ligand-target interaction data in ChEMBL is associated with increasing and measurement-dependent compound promiscuity.** *J Chem Inf Model.* 2012; **52**(10): 2550–2558.
PubMed Abstract | Publisher Full Text
10. Hu Y, Bajorath J: **Activity profile relationships between structurally similar promiscuous compounds.** *Eur J Med Chem.* 2013; **69**: 393–398.
PubMed Abstract | Publisher Full Text
11. Hu Y, Bajorath J: **High-resolution view of compound promiscuity. [v2; ref status: indexed, http://f1000r.es/1ig].** *F1000Res.* 2013; **2**: 144.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Mestres J, Gregori-Puigjané E, Valverde S, *et al.*: **Data completeness--the Achilles heel of drug-target networks.** *Nat Biotechnol.* 2008; **26**(9): 983–984.
PubMed Abstract | Publisher Full Text
13. **Molecular Operating Environment (MOE), 2011.10, Chemical Computing Group Inc.**, 1010 Sherbooke St. West, Suite#910, Montreal, QC, Canada, H3A 2R7, 2011.
Reference Source
14. Weininger D: **SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules.** *J Chem Inf Comput Sci.* 1988; **28**(1): 31–36.
Publisher Full Text
15. UniProt Consortium. **The Universal Protein Resource (UniProt) in 2010.** *Nucleic Acids Res.* 2010; **38**(Database issue): D142–D148.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Baell JB, Holloway GA: **New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays.** *J Med Chem.* 2010; **53**(7): 2719–2740.
PubMed Abstract | Publisher Full Text
17. Hu Y, Bajorath J: **Drug activity data.** *Zenodo.* 2014.
Data Source

Open Peer Review

Current Referee Status: ✓ ✓ ✓

Version 1

Referee Report 24 October 2014

doi:10.5256/f1000research.5597.r6405

Michael Walters[1], **Kathy Nelson**[2]

[1] Department of Medicinal Chemistry, University of Minnesota, Minneapolis, USA

[2] University of Minnesota, Minneapolis, USA

The authors have presented an interesting and careful study on drug promiscuity as analyzed through available activity databases. The authors also illustrate the different conclusions that can be drawn based on the quality of the data analyzed. The discrepancies they point out highlight the importance of high-quality data when evaluating the promiscuity of drugs.

The manuscript is clear and well-written, the methods are scientifically sound, and the analysis is thorough and clearly presented.

Comments:

"Compounds with multiple K_i or IC_{50} measurements for the same target were retained if all these values fell within the same order." Should this read "…the same order of magnitude."?

"Surprisingly, for 282 drugs (~54%), no increase in promiscuity was detected on the basis of high-confidence activity records." It would increase the impact of this article if the authors explained why this finding is surprising. For example, do these compounds represent drugs that were highly optimized for a single target?

"There were only 14 drugs with an increase in promiscuity rates by 10 or more targets." As above, the authors might want to add their thoughts as to why these drugs have such high promiscuity rates. For example, are they kinase inhibitors which had known broad kinase panel activity and the complete scope of their kinase activity is only now coming to light?

Suggestions for minor modifications:

The dataset has been made publicly available through a DOI link. The link was broken when I attempted to follow it and look at the data. This link should be fixed.

The authors have not provided any summary of the criteria used to determine what constitutes activity of a compound at a target (e.g. 10 μM cut-off? 100 μM?). If a strict cut-off cannot be given for each dataset, it is suggested that authors reported the range of activities reported in each dataset (one for the high-confidence data set, and one for each of the two low-confidence data sets).

The authors mention that all the compounds were converted into SMILES strings during data analysis. It is suggested that the authors screen the SMILES list against a REOS/PAINS/etc. filters. It would be interesting to see if any of the high promiscuity compounds could be flagged as having substructures that could be responsible for non-specific activity. Or would have impeded the development of these drugs had they been prepared after cheminformatic filtering came into vogue.

These suggested modifications would help increase the impact of this manuscript and could be of great interest to the drug discovery community.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 16 October 2014

doi:10.5256/f1000research.5597.r6398

Stefan Laufer
Pharmaceutical Chemistry, Eberhard Karls University Tubingen, Tubingen, Germany

Key points:

1. 500+ drugs successfully mapped to ChEMBL to follow a time course of promiscuity.

2. When high-confidence activity data were considered, only small average increase in drug promiscuity over time were observed (from ca 1.5-3).

3. When the stringency of data selection criteria was gradually relaxed and low-confidence data were considered, unrealistic increases in promiscuity over time were detected.

4. As an illustration, please, see the imatinib example in Figure 9 (in comparison to Figure 8a) that the reader might find interesting.

In addition, please, see the conclusions, especially the second part beginning with " On the basis of high-confidence activity data ..."

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: I have a few publications together with J.B. but no common grants etc.

Referee Report 15 October 2014

doi:10.5256/f1000research.5597.r6235

Steffen Renner
Novartis Institute for Biomedical Research, Basel, Switzerland

The authors provide and interesting study on drug promiscuity. Increasing our understanding of this important topic is of high value for the advancement of drug discovery. The current study investigates drug promiscuity over time, i.e. the number of targets reported for drugs over time. The authors conclude that this number is surprisingly low, if only high quality interactions are considered.

The manuscript is well written, scientifically sound, and the analyzed dataset is made publicly available.

Some points for minor modifications:

- When I read the title my first impression was that the manuscript compares the promiscuity of drugs with respect to their release data, e.g. "do more recent drugs show more or less promiscuity than older drugs". I suggest to change the title to better reflect that the release date of the target information is analyzed rather than the release date of the drugs.

- Page 4, second column: The only definition of the promiscuity rate I found in the manuscript is the following: "The total number of target annotations of a drug represented its promiscuity rate". The Wikipedia definition of a rate is "Rate (mathematics), a specific kind of ratio, in which two measurements are related to each other", therefore it appears to me that another name might better reflect the meaning of the promiscuity rate, e.g. just "promiscuity". On the other hand it is not the promiscuity which is changing: it is the number of identified interactions. Therefore another name might be even better.

- I did not find an activity cut-off (e.g. 10µM?) used for the activities in ChEMBL. Please add this information.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response (*Member of the F1000 Faculty* and *F1000Research Advisory Board Member*) 20 Oct 2014

Jürgen Bajorath, Department of Life Science Informatics, B-IT and LIMES Institutes, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany

We thank the reviewer for his comments and the points raised.

1. "Title": We would prefer retaining the current concise title. In our view, it reflects the time course of the promiscuity analysis.

2. "Rate": From a mathematical/statistical point of view, the reviewer is correct ("rate" reflects a ratio). However, outside statistics, the term "rate" is often used synonymously with "degree", which is the intended meaning here. In previous publications, we have used both terms in the context of promiscuity analysis. We agree that the use of the term "degree" would avoid a potential inconsistency and should be preferred.

3. "Activity cut-off": A cut-off value has not been applied in this analysis because the promiscuity time course was compared on the basis of low- vs. high-confidence activity data. The preferred use of high-confidence data would make the application of activity cut-off values a matter of debate. In a previous study (reference 11), it was shown that representative promiscuous compounds identified on the basis of high-confidence activity

data are rarely weakly potent against their targets (which we consider an interesting observation).

Many thanks again to both referees for taking the time to review this manuscript

Competing Interests: None