



## Research Article

# Workflow for Defining Reference Chemicals for Assessing Performance of *In Vitro* Assays

Richard S. Judson<sup>1</sup>, Russell S. Thomas<sup>1</sup>, Nancy Baker<sup>2</sup>, Anita Simha<sup>3</sup>, Xia Meng Howey<sup>3</sup>, Carmen Marable<sup>3</sup>, Nicole C. Kleinstreuer<sup>4</sup> and Keith A. Houck<sup>1</sup>

<sup>1</sup>US EPA, National Center for Computational Toxicology, Research Triangle Park, NC, USA; <sup>2</sup>Leidos, Inc., Research Triangle Park, NC, USA;

<sup>3</sup>ORAU, contractor to U.S. Environmental Protection Agency through the National Student Services Contract, Research Triangle Park, NC, USA;

<sup>4</sup>National Toxicology Program, Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC, USA

### Abstract

Instilling confidence in use of *in vitro* assays for predictive toxicology requires evaluation of assay performance. Performance is typically assessed using reference chemicals – compounds with defined activity against the test system target. However, developing reference chemical lists has historically been very resource-intensive. We developed a semi-automated process for selecting and annotating reference chemicals across many targets in a standardized format and demonstrate the workflow here. A series of required fields defines the potential reference chemical: the *in vitro* molecular target, pathway, or phenotype affected; and the chemical's mode (e.g. agonist, antagonist, inhibitor). Activity information was computationally extracted into a database from multiple public sources including non-curated scientific literature and curated chemical-biological databases, resulting in the identification of chemical activity in 2995 biological targets. Sample data from literature sources covering 54 molecular targets ranging from data-poor to data-rich was manually checked for accuracy. Precision rates were 82.7% from curated data sources and 39.5% from automated literature extraction. We applied the final reference chemical lists to evaluating performance of EPA's ToxCast program *in vitro* bioassays. The level of support, i.e., the number of independent reports in the database linking a chemical to a target, was found to strongly correlate with likelihood of positive results in the ToxCast assays, although individual assay performance had considerable variation. This overall approach allows rapid development of candidate reference chemical lists for a wide variety of targets that can facilitate performance evaluation of *in vitro* assays as a critical step in imparting confidence in alternative approaches.

## 1 Introduction

There is increased interest in using high-throughput, *in vitro* bioassays to generate data for predictive toxicology models. Uses of these models range from prioritization of chemicals for more detailed studies to the replacement of current *in vivo* tests. *In vitro* assays have long been used for genotoxicity testing, including regulatory purposes, but are now being employed in numerous other areas. A recent example of regulatory use is screening compounds for endocrine effects<sup>1,2</sup> (U.S. EPA 2013; Judson et al., 2015). Assays for pharmacological targets are widely used

to identify candidate drugs, but also to test compounds for off-target toxicity (Whitebread et al., 2005). Recently, the International Agency for Research on Cancer (IARC) has started using diverse, high-throughput *in vitro* data to provide evidence for certain cancer mechanisms (Chiu et al., 2018). In the environmental chemical toxicity space, these efforts are exemplified by the U.S. ToxCast and Tox21 programs (Judson et al., 2010; Kavlock et al., 2012; Tice et al., 2013; Attene-Ramos et al., 2015), and follow ideas discussed in the U.S. National Research Council's report Toxicity Testing in the 21<sup>st</sup> Century (NRC, 2007). The growing use of the adverse outcome pathway (AOP)

<sup>1</sup> EPA Endocrine Disruptor Screening Program (EDSP). <http://www.epa.gov/endo/> (accessed 08.08.2008).

<sup>2</sup> FIFRA SAP Meeting held January 29-31, 2013 on the Scientific Issues Associated with Prioritizing the Universe of Endocrine Disruptor Screening Program (EDSP) Chemicals Using Computational Toxicology Tools. [http://ntp.niehs.nih.gov/NTP/About\\_NTP/SACATM/2013/September/SAPMtgRpt\\_Jan2013\\_508BE.pdf](http://ntp.niehs.nih.gov/NTP/About_NTP/SACATM/2013/September/SAPMtgRpt_Jan2013_508BE.pdf) (accessed 28.10.2013).

Disclaimer: The views expressed in this paper are those of the authors and do not necessarily reflect the statements, opinions, views, conclusions, or policies of the Environmental Protection Agency.

Received September 28, 2018; Accepted December 6, 2018;  
Epub December 17, 2018; © The Authors, 2018.

ALTEX 36(2), 261-276. doi:10.14573/altex.1809281

Correspondence: Richard Judson, Ph.D.,  
U.S. EPA, Office of Research and Development  
National Center for Computational Toxicology, Research Triangle Park, NC, USA  
(judson.richard@epa.gov)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



concept to represent the mechanistic events leading to an adverse outcome in an animal or population emphasizes the importance of *in vitro* approaches targeting molecular initiating events (MIEs) or key events (KEs). As the goal of this strategy is to develop alternative methods to replace animal testing, validation of the performance of the assays targeting the MIEs and KEs will be critical to regulatory acceptance (Vinken, 2013).

Appropriate use of *in vitro* data generated from high-throughput screening assays requires both statistical and biological confidence in the results. Methods for appropriate statistical evaluation and metrics to ensure robust operational performance of high-throughput screening approaches in support of drug discovery have been developed (Zhang et al., 1999; Bleicher et al., 2003; Malo et al., 2006). Typically, these rely on the use of a single substance as a positive control and a corresponding neutral (solvent) control. While this is important for assay development efforts, a single chemical is rarely representative of the diversity of chemical compounds to be screened in the assay. Large and diverse compound libraries are likely to include active compounds with a range of potency and, possibly, efficacy values; compounds that may modulate the target through both direct and indirect mechanisms; compounds that interfere with the assay yielding false positive or false negative results; and truly inactive compounds (Feng and Shoichet, 2006; Thorne et al., 2010; Bruns and Watson, 2012). By testing a larger set of chemicals with known expected outcomes on biological targets, i.e., reference chemicals, the sensitivity and specificity of the assay can be assessed.

With a large and diverse set of reference chemicals, one can also assess the domain of applicability (DOA) of an assay. DOA is a concept from quantitative structure-activity relationship (QSAR) modeling (Jaworska et al., 2005) that defines the chemical universe for which predictions can be considered relevant by comparing the structural similarity of a new chemical to those used in the development of the model. For *in vitro* assay evaluation, assessing the DOA involves testing reference chemicals that cover the range of possible mechanisms capable of modulating the biological target. The ability to measure modulation of the biological target will define the DOA rather than the chemical structure as in a QSAR. Reference chemicals that target the mechanism(s) being interrogated will be needed to prove that a given assay endpoint is responsive when particular mechanisms are affected and establish the limits of an assay's DOA. In addition, the physicochemical properties of a chemical will also define if it can be successfully tested in an assay, e.g., the solvent and aqueous solubility of the chemical, its volatility, stability under assay conditions, potential to interfere with assay detection technology, etc., and hence contribute to defining the assay DOA.

Here we use the following operational definition of a reference chemical for a specific biological target and activity mode (e.g., agonist, antagonist, inhibitor): a reference chemical is one that

gives consistent results (active vs. inactive) across multiple different assays (usually run in different laboratories) that measure activity against a target or molecular mechanism, for a specified mode. For detailed analysis, we focused on assays that we could readily link to Entrez Gene IDs and we refer to generically as biological targets or just "targets". This definition allows one to start with a collection of chemicals that have been tested in a suite of assays against a specified biological target and extract reference chemicals from that data set. For example, we could evaluate chemicals tested in assays measuring agonist-mode binding to the estrogen receptor. All chemicals that were consistently active would then be denoted as positive reference chemicals, and those consistently inactive would be considered negative reference compounds. Ideally, the assays would vary in cell background, detection technology and other parameters, so that the consistent activity would clearly reflect the underlying biology and not be an artifact of a specific assay type.

This operational definition raises two questions: (1) Does "consistent" activity mean 100% of observations? We say no for two reasons. First, a chemical that is active, but only weakly so, may be negative in an assay because it was not tested to a high enough concentration, as different assays can have different sensitivities. Second, a true inactive chemical might be active in one assay but inactive in a second (or the converse) because of interference with the detection technology in one of the assays. (2) Can a novel target have reference chemicals? In our definition, until there are multiple assays available against which chemicals can be tested, there can be no reference chemicals for a novel target because there is no way to test consistency. Apparent exceptions would be a single laboratory that, through a variety of tests, proves that a chemical acts on a specific biological target. However, this is just a special case of using multiple assays that give consistent results.

In the *in vitro* toxicology field, there have been few targets for which large sets of reference chemicals have been compiled. The estrogen receptor (ER) is one such target where a large set of reference chemicals has been developed through extensive efforts over years of work with input from various international bodies<sup>3</sup>. Despite many years of effort, this list still generates controversy, for instance about whether a chemical that is a positive reference for agonism can automatically serve as a positive control for binding assays (binding is a requisite precursor to agonism) or whether effects observed at relatively high concentrations or in specific cell/tissue types are broadly relevant. Given that there are many biological targets and mechanisms being explored for use in *in vitro* screening of chemicals for potential toxicity, it is impractical to take a decade to generate the reference chemicals for each target to evaluate assay performance. Here we explored a semi-automated approach to the development of reference chemical lists. In our approach, we automatically extracted chemical-target-mode-activity call data from several public online sources to find chemicals with multiple lines of evidence

<sup>3</sup> NICEATM DRAFT ED BRD: BG1Luc ER TA Test Method – Section 3.0. <http://iccvam.niehs.nih.gov/methods/endocrine/BG1Luc/Section3-24Jan2011.pdf> (accessed 21.03.2011).

of interaction with a given target. We populated a database we call RefChemDB with these data and manually reviewed a subset of the source documents to assess the reliability of the data (i.e., chemical, target, mode, activity) extracted from the online sources and check for correctness of the original data. The manual review process was carried out for 54 targets that ranged from data-poor to data-rich. Next, from our resulting reference chemical database, we proposed lists of candidate reference chemicals for 50 molecular targets, based on their having sufficient experimental supporting data. Additionally, we evaluated a subset of these targets and candidate reference chemicals against a set of *in vitro* assays run as part of the ToxCast and Tox21 screening programs.

## 2 Methods

### Initial database construction

To provide reference chemical lists for use in evaluating *in vitro* assay performance, we developed a system that captures relevant information describing the link between chemicals and potential biological targets *in vitro*. Information is extracted from the data sources and compiled into two key database tables in RefChemDB: *source\_chemical* and *target\_summary* (Fig. 1). The *source\_chemical* table contains the name and Chemical Abstracts Registry Number (CASRN) from the source database record along with chemical structure information (SMILES, inchi\_key, pubchem\_cid), while the *target* table contains information on data source, the biological target (gene name, symbol, Entrez gene identifier (geneid)), the mode (e.g., agonist, antagonist; full list of modes available in Tab. S1<sup>4</sup>), whether the chemical was active or inactive, and reference information such as a PubMed ID (pmid). Data in these tables are then summarized into a *chemical* table with one unique record per chemical, linked to the EPA DSSTox<sup>5</sup> (Richard et al., 2006) record which provides curated chemical identity and chemical structure information. The *target\_summary* table contains the individual records from the literature linking a chemical to a target. It also contains mode information (e.g., agonist, antagonist, inhibitor), the source of the record, and whether the chemical was active or inactive (activity\_class) for the target. For the current work, we focus primarily on gene-based biological targets (e.g., specific receptors, enzymes, ion channels, etc.), although the database also contains information on chemicals with other types of targets (e.g., mitochondria, cell membranes) as well as broader

target or mode of action classes (e.g., antibacterial, antiviral) without specific target information. These target types are not further described in the present work. The contents of the database, linking 40,897 chemicals and 1,395,852 target summaries (*assigned to gene ID's*), are provided in Table S2<sup>4</sup>.

Information is compiled from the following sources:

- ChEMBL<sup>6</sup> (Gaulton et al., 2012): This is a large repository of *in vitro* assay activity manually curated from the open literature. In our database, ChEMBL was divided into two sets called ChEMBL and ChEMBL Drug. The first is taken from the literature records in ChEMBL, while the second is from the drug target annotation in that database. We used ChEMBL version 23. Modes and PMIDs are provided.
- CTD – Comparative Toxicogenomics Database<sup>7</sup> (Mattingly et al., 2006; Davis et al., 2009; Wiegers et al., 2009): CTD contains information on chemical activity including target interactions and gene expression data, manually curated from the open literature. Modes of action from this source were not standard compared to those of other sources (e.g., affects^activity/affects^binding). A dictionary mapping from CTD to RefChemDB modes was created by checking a set of PMIDs with both sets of terms. The mapping is available in Table S3<sup>4</sup>. PMIDs are provided.
- DrugBank<sup>8</sup> (Wishart et al., 2006, 2008): DrugBank contains information on drugs (both approved and experimental), including the intended targets. Modes of action were inferred using a search for certain keywords, first from drug category and, if not provided there, from the drug description, pharmacology, or MOA fields. PMIDs are provided.
- Eurofins<sup>9</sup>: Eurofins is a company that provides *in vitro* screening contract services, and has compiled a set of reference chemicals for their assays. These reference chemicals and assays are divided into biochemical and functional classes. Modes are provided but PMIDs are not.
- Iuphar/BPS<sup>10</sup>: Iuphar/BPS is a resource developed by the International Union of Basic and Clinical Pharmacology and provides information on drugs and their targets. Modes are provided but PMIDs are not.
- KEGG Drug<sup>11</sup> (Kanehisa et al., 2006, 2008): KEGG Drug is another source of drug/target interactions. Neither modes nor PMIDs are provided.
- KIDB – Ki Database<sup>12,13</sup>: KIDB is a database of affinity information on drugs and drug candidates, focused on GPCRs. Neither modes nor PMIDs are provided.

<sup>4</sup> doi:10.14573/altex.1809281s1

<sup>5</sup> <https://comptox.epa.gov>

<sup>6</sup> <https://www.ebi.ac.uk/chembl/>

<sup>7</sup> <http://ctdbase.org/>

<sup>8</sup> <https://www.drugbank.ca/>

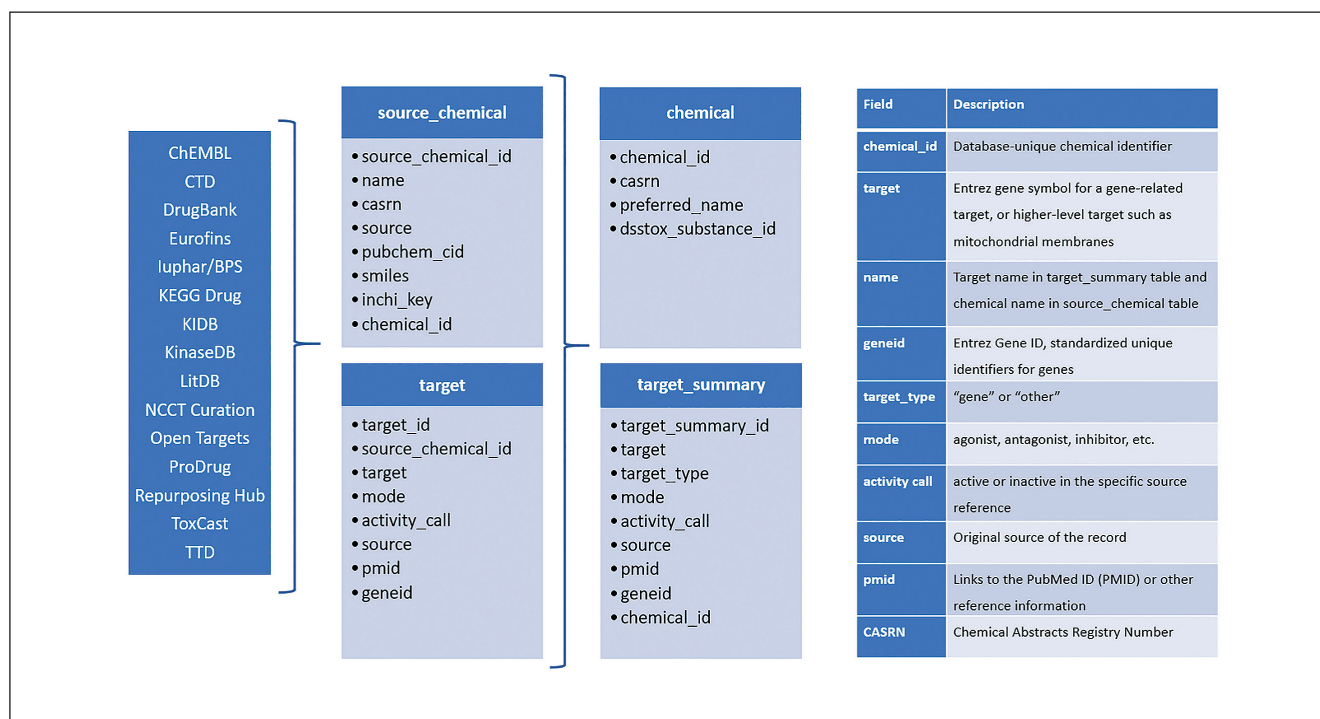
<sup>9</sup> <https://www.eurofinsdiscoveryservices.com/cms/cms-content/resources/>

<sup>10</sup> <http://www.guidetopharmacology.org/>

<sup>11</sup> <https://www.genome.jp/kegg/drug/>

<sup>12</sup> <https://kiddbdev.med.unc.edu/databases/kiddb.php>

<sup>13</sup> PDSP Ki Database. <http://pdsp.med.unc.edu/kiddb.php> (accessed 08.08.2008).



**Fig. 1: RefChemDB Database Schema**

Data sources and fields in RefChemDB. Data from various sources are collected and categorized into chemical information and target information, which are then reorganized into *target\_summary* and *chemical*. Each row in the *target\_summary* table summarizes a record from the literature – source, PMID if given, target, chemical ID, mode, etc. Each row in the *chemical* table contains information on a chemical mentioned in the literature – name, CASRN, and DSSTox Substance ID. The table at the right defines key terms.

- KinaseDB<sup>14</sup>: This is a dataset of reference kinase inhibitors provided by the commercial assay vendor Eidogen Sertanty (Sharma et al., 2016). Neither modes nor PMIDs are provided.
- LitDB: LitDB is the EPA's database of MeSH term annotations extracted from Medline (Baker and Hemminger, 2010). For RefChemDB, records were extracted from LitDB that contained annotations for chemicals and activity on targets (antagonism or agonism). Modes and PMIDs are provided. A more detailed description of methods used to create LitDB and extract from it for inclusion in RefChemDB can be found in supplementary data S4<sup>15</sup>.
- NCCT Web Curation: This is a collection of chemical, target and mode information manually curated from public web resources other than the open literature by the authors, focused on compounds in the ToxCast library (see below). Neither modes nor PMIDs are provided.
- Open Targets<sup>16</sup>: This is a public database of drugs and their targets. Modes are provided but PMIDs are not.
- ProDrug: This is a collection of prodrugs (drugs that require metabolic activation) taken from Casida (2017). Neither modes nor PMIDs are provided.
- Repurposing Hub<sup>17</sup> (Corsello et al., 2017): This is a database of drugs and potential new targets they could be active against. Modes are provided but PMIDs are not.
- ToxCast<sup>18</sup> (Dix et al., 2007; Kavlock et al., 2012; Judson et al., 2014): The ToxCast/Tox21 project is a collaboration between the US EPA, NIH and FDA to screen thousands of chemicals against hundreds of *in vitro* assays. The ToxCast database contains detailed information on the assays, chemicals and activity. Data was extracted from invitrodb V2 (August 2018) in the form of chemical-assay activity calls (activity=1 if the chemical assay pair was considered active, =0 otherwise). The only assays considered were those with a specific biological target, i.e., excluding assays such as those for mitochondrial disruption. Modes are provided but PMIDs are not.

<sup>14</sup> <http://www.eidogen-sertanty.com/kinasednld.php>

<sup>15</sup> doi:10.14573/altex.1809281s2

<sup>16</sup> <https://www.targetvalidation.org/>

<sup>17</sup> <https://clue.io/repurposing>

<sup>18</sup> <https://actor.epa.gov/dashboard>



- TTD<sup>19</sup> (Therapeutic Target Database) (Li et al., 2017): TTD contains information on known therapeutic targets of drugs. Modes are provided but PMIDs are not.

For nonstandard target symbols, a dictionary was created to standardize them and associate a gene ID. For example, 5-HT5B became symbol HTR5B and gene ID of 15564. This dictionary file is available in Tab. S5<sup>4</sup>. When a given gene target was ambiguous (e.g., HTR5) or unable to be associated with a gene ID, the information was excluded. Non-gene targets such as mitochondria were included without IDs in RefChemDB, but were excluded from subsequent analyses which focused on gene-based biological targets.

#### *Comparison with ToxCast/Tox21 data*

Candidate reference chemical activity was evaluated against a large collection of in-house *in vitro* data from the ToxCast and Tox21 projects (Judson et al., 2010; Kavlock et al., 2012; Attene-Ramos et al., 2013; Tice et al., 2013). These programs have generated data on hundreds of *in vitro* assays for up to 9000 chemicals, and produce both activity calls (i.e., active/inactive) as well as potency estimates (i.e., AC50 values). For this evaluation, the ToxCast and Tox21 data in the RefChemDB database for the intended biological target was excluded. Data from the sources LitDB, ProDrug, and Repurposing Hub were also excluded (see “Creation of Candidate Reference Chemical Collection”). We compared the activity (active/inactive) for every pair of reference chemicals and assays that matched on target and mode. Candidate reference chemicals could have literature support for more than one mode per target (e.g., some references might indicate agonist activity and others antagonist activity). For comparison with the ToxCast/Tox21 data, a single mode was selected, usually the one with the largest support. Support is defined as the number of unique references (papers) documenting that a chemical interacts with a given target through a given mode (e.g., agonist, antagonist) and activity call (active, inactive). The exceptions were cases where the default mode was “binder”, in which case the second most supported mode was selected. For the case of receptors, binding is a prerequisite for both agonists and antagonists. Because most receptor assays have a specific mode (agonism/antagonism), more specific reference chemical modes are more informative. Assay modes were agonist, antagonist, binder, and inhibitor. For this comparison, several uncommon modes were mapped to more common variants: agonist, activator, stimulator, positive allosteric modulator, positive modulator, partial agonist, inducer, enhancer were mapped to “agonist”; antagonist, allosteric antagonist were mapped to “antagonist”; negative modulator, inactivator, inhibitor covalent, inhibitor reversible, blockade, blocker, channel blocker, negative allosteric modulator, gating inhibitor, suppressor, uptake inhibitor, inhibitor were mapped to “inhibitor”. After mapping original modes to the standard modes, the support was

calculated as the sum of the individual support values. The rules for matching reference chemical modes and assay modes were: reference chemical mode (binder) matches assay mode (binder); reference chemical agonist mode (agonist) matches assay modes (agonist, binder); reference chemical modes (antagonist, inhibitor) match assay modes (antagonist, inhibitor, binder).

#### *Targets selected for manual analysis*

A total of 54 molecular targets, listed in Table S6<sup>4</sup>, were selected for the manual curation analysis with a total of 1375 chemicals associated with them. These are either nuclear receptors, GPCRs, transporters or enzymes. Four of the 54 targets lacked chemicals with a sufficient level of support to be included resulting in the 50 targets included in Table S6<sup>4</sup>. Sufficient support was considered to be at least five references listing a target-chemical interaction with a particular mode.

#### *Manual curation process*

Three curators were assigned the manual curation process. Each had a degree in biological science and familiarity with chemical and gene nomenclature. To start the curation process, a query was run to extract target-chemical pairs that potentially had sufficient support (see above) to be considered reference chemicals. For each target in Table S6<sup>4</sup>, candidate chemicals were exported as Microsoft Excel files, and were manually curated to assess the accuracy of the raw database information. The following information was extracted from the paper: chemical identity, target, mode, activity, potency and units, quality control status, and additional notes. Up to ten records were analyzed for each chemical-target pair for the quality control protocol. In developing a final reference chemical set for any specific target, such a limit would not be used, but in this case, the limit allowed us to estimate the scope of quality issues across many targets while limiting the resources required. Only records with PMIDs were searched. Some database sources (e.g., TTD) do not provide PMIDs, so they were not included in this step. The paper for each record was downloaded based on the PMID. The mode was recorded as agonist, pan agonist (active against most/all targets in a protein family), antagonist, inverse agonist, inhibitor, modulator or binder, based on the paper. To identify the mode, the chemical for each record was searched in the paper in relation to its gene target. When the preferred chemical or gene name was not listed in the paper, synonymous names were then searched on the EPA CompTox Chemicals Dashboard database<sup>20</sup> and on NCBI's Gene database<sup>21</sup>, respectively. The activity of the chemical was recorded as active or inactive based on whether the paper confirmed the action of the chemical. To determine the potency, the quantitative potency values (e.g., AC50, IC50, ED50, Ki or Kd) were extracted and the units were recorded as molar concentrations or units of mass/volume. The quality control status was determined as yes or no based on whether the chemical was

<sup>19</sup> <http://bidd.nus.edu.sg/group/cjttd/>

<sup>20</sup> <https://comptox.epa.gov>

<sup>21</sup> <https://www.ncbi.nlm.nih.gov/gene>



reported in the paper to act on the target listed in the database source. Information not available in the literature was marked as unspecified.

A blind cross-validation was also performed to compare judgments of curators. Two of the three curators took part in this effort. Previously completed records across the targets were compiled for review. Data entered by the original curator were removed, while the second curator extracted and entered information for each record as listed above. A total of either 33 or 34 records initially completed by one of the original three curators were re-reviewed by a second curator. The reviewing curator's quality control status entries were then compared to the original curator's quality control determination. Inter-rater reliability between the two reviewing curators was assessed via Cohen's Kappa test.

#### *Creation of candidate reference chemical collection*

For the subsequent analyses, we defined four categories of information: (1) Records manually curated by the authors (all have PMIDs); (2) other curated records with PMIDs (ChEMBL, CTD, DrugBank, LitDB); (3) other records without PMIDs (Eurofins, Iuphar/BPS, KEGG Drug, KIDB, KinaseDB, NCCT Web Curation, Open Targets, ProDrug, Repurposing Hub, TTD); (4) data points from ToxCast/Tox21 assays. A key parameter determining reference chemical identification is the "support" value, which is the number of independent records across the subset of these categories being analyzed. For sources with PMIDs, it is the number of unique PMIDs across all sources matching a specified chemical-target-mode-activity call combination. For the non-PMID sources, each source added one to the support count. For ToxCast/Tox21 data, each assay with matching chemical-target-mode-activity call added one to the support.

Results of the manual curation step were used to inform additional curation of the raw data extracted in the initial phase of database construction using a series of automated scripts. To create the final list of reference chemicals, data without a target name or without a DSSTox substance ID were removed. Data from LitDB were excluded due to weak precision rates in the manual curation process. ProDrug was excluded because it contains chemicals requiring metabolic activation, which is not a property of most of the assays to be validated although may be a valuable resource for applications requiring knowledge of assay xenobiotic metabolism capacity. Repurposing Hub was excluded because most of the activities reported are hypothesized based on target similarity rather than on experimental evidence. Independent records were curated in the following manner to account for potential duplicates: If the same record was repeated across multiple sources (same chemical, target, mode, activity status, and PMID), only the first instance was recorded. The information in records without PMIDs was counted once per data source (same chemical, target, mode, activity status). Finally, support was determined by tallying the number of records for the

chemical-target-mode-activity call combination. Modes were consolidated into "positive" (e.g., "activator", "agonist", "stimulator"), "negative" (e.g., "inhibitor", "antagonist", "blocker"), or "unspecified" (e.g., "binder", "modulator", "stabilizer"), and the amount of support for each chemical-target-mode-activity call combination were re-tallied. The mode consolidation decisions are given in Table S1<sup>4</sup>. The contents of the database after mapping all modes along with target symbols and gene IDs are available in Table S7<sup>4</sup>.

#### *Software*

All software for this project is written in R (Ihaka and Gentleman, 1996). Data is stored in a custom MySQL database. Input files, code and the database are available<sup>22</sup>.

### **3 Results**

#### **3.1 Summary of RefChemDB database**

Table 1 summarizes the size and diversity of the database and information from the different sources. ChEMBL has the largest number of unique chemicals, while CTD has the largest number of individual targets. ToxCast/Tox21 has the largest number of unique chemical-target-mode-activity call combinations. Most chemical-target-mode-activity call combinations show up in the database only once, with an average support (average number of times the combination appears in the database) of only 1.2. The median support for each source was 1, except for DrugBank, which had median support of 3 records. LitDB and DrugBank had the highest median support. Although most combinations have only a few references, there are combinations with support of up to 286.

#### **3.2 Analysis of ToxCast/Tox21 data activity calls vs. other sources**

A comparison of candidate reference chemicals curated from public sources and ToxCast/Tox21 *in vitro* data was performed to help quantify reliability of both the reference chemicals and the ToxCast/Tox21 assays. The first comparison examines the performance of the assays, assuming the reference chemicals are correct, as a function of the reference chemical support, calculated excluding ToxCast/Tox21 data. See Methods for the rules on matching chemical and assay modes. Full results by assay and chemical are given in Table S8<sup>4</sup>. The results are summarized in Figure 2A (data from Tab. S9<sup>4</sup>). One can see that the assay performance (fraction of correct matches between the assay result and the putative reference chemical activity) increases with support, indicating that chemicals with greater support are more likely to be correctly assigned their target and mode. There were 92 assays with at least one candidate reference chemical with support  $\geq 5$ . Of those, 63 assays demonstrated expected activity for  $\geq 80\%$  of the reference chemicals. The lower performing assays are listed in Table 2. There are

<sup>22</sup> [ftp://newftp.epa.gov/COMPTOX/NCCT\\_Publication\\_Data/Judson/Reference\\_Chemicals\\_for\\_In\\_Vitro\\_Assays/](http://newftp.epa.gov/COMPTOX/NCCT_Publication_Data/Judson/Reference_Chemicals_for_In_Vitro_Assays/)

**Tab. 1: Summary statistics for the RefChemDB database**

ChEMBL provides the largest number of unique chemicals, while CTD has the largest number of individual targets. ToxCast has the largest number of unique chemical-target-mode-activity combinations. Most combinations show up in the database only once from each source, with an average support (number of times in the database) per source of only 1.2. LitDB and DrugBank have higher multiplicities.

Source	Chemicals	Targets	Chemical-target-mode-activity combinations	Mean multiplicity	PMIDs
ChEMBL	28,832	2,238	310,984	1.16	11,520
ChEMBL Drug	1,187	738	4,099	1	0
CTD	2,317	7,904	25,606	1.22	5,280
DrugBank	1,630	1,169	3,623	3.41	6,274
Eurofins Biochemical	206	570	925	1	0
Eurofins Functional	211	239	706	1	0
Iuphar BPS	1,860	941	5,081	1	0
KEGG Drug	661	263	1,201	1	0
KIDB	535	450	6,532	1	0
KinaseDB	133	168	676	1	1
LitDB	2,654	88	8,348	4.94	27,909
Open Targets	1,031	820	3,973	1	0
Prodrug	41	33	41	1	1
Repurposing Hub	2,279	2,172	10,209	1	0
ToxCast	9,136	343	852,470	1.03	0
TTD	3,916	1,575	11,557	1.00	0
Web Curation	3,940	1,059	5,617	1.01	0
<b>Total</b>	<b>37,301</b>	<b>11,055</b>	<b>123,4580</b>	<b>1.02</b>	<b>49,883</b>

several targets where multiple assays show this performance, including HDAC inhibitors (3/3 assays  $\leq$  80%), NR1I3 / CAR (3/3 assays  $\leq$  80%), PPARG agonist (4/5 assays  $\leq$  80%), PPARG agonist (2/2 assays  $\leq$  80%). In a number of these cases, there are only 1 or 2 reference chemicals. In the case of the HDAC assays, there were three targets, HDAC1, HDAC3 and HDAC6, tested with 41, 2 and 3 reference chemicals, respectively. For HDAC1, 32 chemicals had only one reference and four had two. Suberoylanilide hydroxamic acid, a pharmaceutical with nM potency against multiple HDAC's, had 206 references and was active against HDAC1. However, valproic acid, with at least 6 references for each of the three HDAC's, was inactive in each. Its potency (IC<sub>50</sub>) is in the hundreds-of-micromolar range and thus expected to be inactive in the ToxCast biochemical HDAC assays where the upper testing concentration was less than 100  $\mu$ M (Huber et al., 2011). Butanoic acid, with four references for HDAC1, was inactive against HDAC1 in ToxCast (not tested in the others) and, like valproate, had an IC<sub>50</sub> in the hundreds of micromolar range (Huber et al., 2011). N-hydroxybenzamide, with four references, was active against HDAC1 (not tested against HDAC3 or HDAC6) and is also a

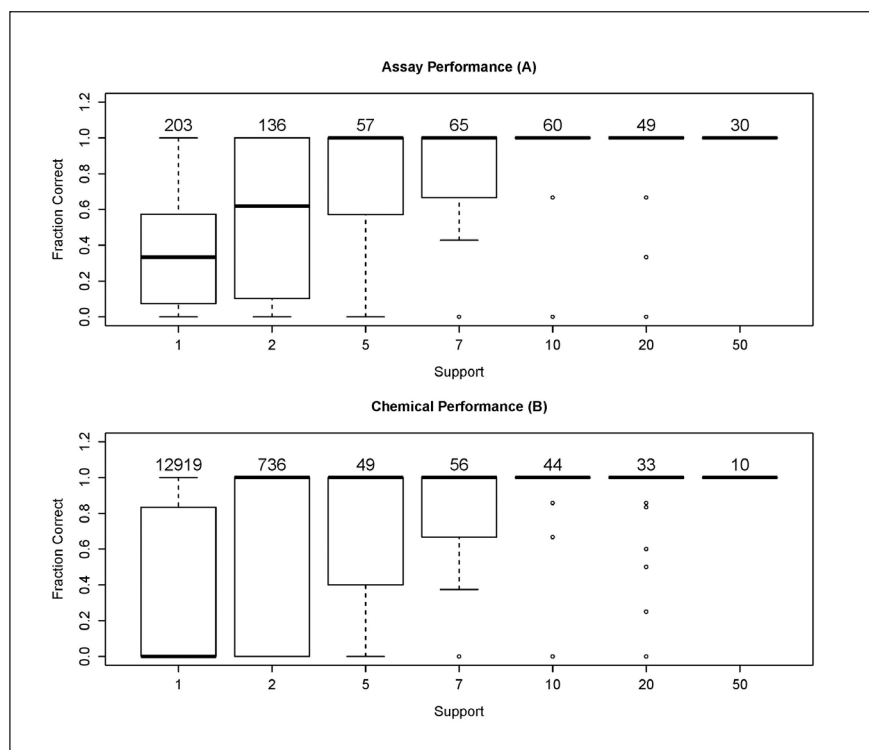
structure used as a drug scaffold with sub-micromolar potency (Huong et al., 2017). Decitabine, the only other inhibitor with more than two supporting references, was inactive against HDAC1, but review of the literature showed it was a synergist with HDAC inhibitors but acting as a hypomethylating agent independent of direct effects on the HDAC enzyme (Kalac et al., 2011). This underscores a need for careful examination of potency values as well as subject matter expert review before finalizing candidate reference chemical lists for assay performance validation.

The second comparison examines the performance of the reference chemicals, assuming that the assays perform well, again as a function of support. This is summarized in Figure 2B (from Tab. S10<sup>4</sup>). Here one can also see that the performance increases with support. Strikingly, the candidate reference chemicals with a single report are inactive in the corresponding assays most of the time, indicating that chemicals with limited support should generally not be considered as candidate reference chemicals. For chemicals with high support, there are several cases where the automated assignment of mode is incorrect. These include raloxifene and tamoxifen for ESR1 and nilutamide for AR, each



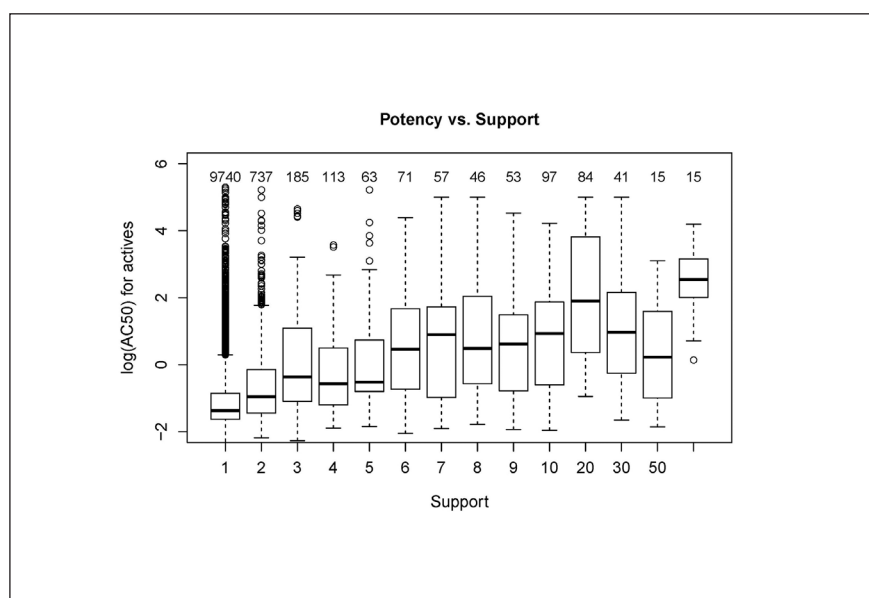
of which were assigned agonist, when the true mode is most often considered to be antagonist. Raloxifene and tamoxifen are both examples of complicated pharmacology as they are defined as selective estrogen receptor modulators (SERMs) (Shang and Brown, 2002). SERMs can behave as either agonists or antagonists depending on the details of the assay such as the cell type or assay mode (agonist *versus* antagonist design). Again,

subject matter expertise is beneficial during the curation step. Also, for ESR1 and AR, there are several correct reference chemicals known to have weak potency, which causes less potent assays to report negative results. These chemicals include kepone, o,p'-DDT and butylparaben (ESR1 agonist), and linuron and methoxychlor (AR antagonist) (Judson et al., 2015; Kleinstreuer et al., 2017).



**Fig. 2: Performance of the candidate reference chemicals against the ToxCast/Tox21 assays**

In all cases, support is calculated excluding ToxCast/Tox21 data. (A) Performance at the assay level, matching on both target and mode. The boxplots show the fraction of reference chemicals that are active in an assay for the matching target and mode. The X-axis gives the minimum support, or independent references linking a chemical to a target, for the bin. The numbers over the boxplot are the numbers of chemical-assay pairs in the bin. Note that a given assay might contribute to multiple bins. (B) Performance at the chemical level, matching on both target and mode. The boxplots show the fraction of assays that the candidate reference chemical was active in for assays matching the chemical's target and mode. The numbers over the boxplot are the numbers of chemicals in the bin. A chemical will only contribute to a single bin.



**Fig. 3: Relationship between potency and support in ToxCast assays**

Each boxplot summarizes  $-\log(\text{AC}_{50})$  values, which measure potency (larger values are more potent, i.e., chemicals interact with the target at lower concentrations), for chemical-target-mode combinations with the specified support range. The X-axis gives the minimum level of support for the data in each boxplot. Support is defined as the number of unique literature records linking a target to a chemical. Inactive compounds were excluded. The number of chemical-target-mode combinations within each support range are shown above the bar.



We also observe a trend that chemical-target pairs with greater support tend to be more potent. These data are summarized in Figure 3. A linear model fit between the continuous support and  $-\log(\text{AC}_{50})$  for ToxCast assay-chemical pairs that are active gives an adjusted  $R^2$  value of 0.18 with  $p$ -value  $< 2 \times 10^{-16}$ . A similar trend was seen with a large data set of estrogen receptor actives extracted from public data (Mansouri et al., 2016).

The basis of this observation is that chemicals with low potency (only active at relatively high concentrations) will only be active in the most sensitive assays, and in situations where they are tested to high concentrations. Since typical screening campaigns limit testing to 10 or 100  $\mu\text{M}$ , these weaker compounds will more often be missed.

**Tab. 2: ToxCast/Tox21 assays showing activity in  $\leq 80\%$  of reference chemicals with support  $\geq 5$ , where support is the number of independent reports linking a target to a chemical**

The assay name, biological target in the assay, and the mode of action of the reference chemical(s) on the target are shown. The pass rate, or the rate with which the reference chemicals showed the same activity as seen in the literature is calculated and displayed along with the number of reference chemicals tested for each assay. For example, in the NVS\_NR\_hPXR assay, the assay had a 0.33 pass rate because only one of the three reference chemicals was observed to have an agonist mode of action.

Assay	Target	Mode	Pass rate	Reference chemicals
ATG_CAR_TRANS_up	NR1I3	agonist	0	2
ATG_RARb_TRANS_up	RARB	agonist	0	2
NVS_ADME_hCYP2C9	CYP2C9	inhibitor	0	1
NVS_ENZ_hHDAC3	HDAC3	inhibitor	0	1
NVS_ENZ_hHDAC6	HDAC6	inhibitor	0	1
NVS_IC_hKhERGCh	KCNH2	inhibitor	0	1
OT_NURR1_NURR1RXRa_0480	RXRA	agonist	0	1
TOX21_CAR_Antagonist	NR1I3	inhibitor	0	1
NVS_NR_hPXR	NR1I2	agonist	0.33	3
TOX21_CAR_Agonist	NR1I3	agonist	0.33	3
NVS_NR_hPPARa	PPARA	agonist	0.4	5
ATG_PBBREM_CIS_up	NR1I3	agonist	0.5	2
ATG_RARg_TRANS_up	RARG	agonist	0.5	2
NVS_ADME_hCYP3A4	CYP3A4	inhibitor	0.5	2
NVS_IC_rNaCh_site2	SCN1A	inhibitor	0.5	2
OT_AR_ARELUC_AG_1440	AR	agonist	0.5	16
TOX21_FXR_BLA_agonist_ratio	NR1H4	agonist	0.5	2
TOX21_HDAC_Inhibition	HDAC1	inhibitor	0.5	2
NVS_NR_hPPARg	PPARG	agonist	0.6	5
OT_PPARG_PPARGSRC1_0480	PPARG	agonist	0.6	5
OT_PPARG_PPARGSRC1_1440	PPARG	agonist	0.6	5
TOX21_PPARD_BLA_agonist_ratio	PPARD	agonist	0.6	5
TOX21_Aromatase_Inhibition	CYP19A1	inhibitor	0.66	6
ATG_GRE_CIS_up	NR3C1	agonist	0.75	8
ATG_PPARD_TRANS_up	PPARD	agonist	0.75	4
TOX21_AChE_Inhibition_Fluor	ACHE	inhibitor	0.77	9
TOX21_ERb_BLA_Agonist_ratio	ESR2	agonist	0.77	9
TOX21_ERa_BLA_Agonist_ratio	ESR1	agonist	0.79	24
TOX21_PPARG_BLA_Agonist_ratio	PPARG	agonist	0.8	10



### 3.3 Analysis of manually curated target set

Manual curation is a time-consuming process. Because it was not feasible to manually curate the entire database of ~50,000 PMIDs plus other source data, we selected a sample of data from which to estimate the rate of correct and incorrect annotations of chemical-target-mode-activity combinations. For this effort, we selected 54 gene targets including nuclear receptors (NRs), G-protein couple receptors (GPCRs), transporters and enzymes, listed in Table S6<sup>4</sup>. Some of these are well-studied (e.g., dopamine and histamine receptors) while others are targets that are relatively poorly explored, but for which *in vitro* assays are currently being developed for chemical toxicology screening at the EPA. For each target, we queried RefChemDB for the number of chemicals that had 1, 2, 3, ... 10 or more reports in the *target\_summary* table. For the current effort, we then selected chemicals with 5 or more reports as candidate reference chemicals for manual curation. Counts for all targets are given in Table S11.1<sup>4</sup>. Out of a larger initial set of 106 targets, 72 had at least 5 chemicals with at least 5 reports. At the high end, several targets had more than a hundred candidate reference chemicals (AR, ESR1, HRH1, DRD2, OPRM1, HRH2), while at the low end, 5 targets had no reports in any of the sources (DIO, DIO3, NR2A2, SLC16A2, SRD5A3). For the current exercise, we selected targets from the middle range of data size. Note that this process was limited to the sources that provided PMIDs (i.e., ChEMBL, CTD, DrugBank, and LitDB).

Each of the three individual curators read a set of papers. Most papers were read by a single curator, but a subset was read by two (see below). The different data sources providing links to the literature (PMIDs) were not assumed to be of equal quality. Data in some of the sources have already passed through a level of manual curation. DrugBank, for instance, is a curated source of chemical target information and we therefore assumed it to be definitive. On the other hand, LitDB is a collection of MeSH keywords assigned to articles to aid lookup, not to delineate a specific relationship between a chemical and target; it is likely to have a higher rate of records not appropriate for RefChemDB (i.e., false positives). Here we describe some of the specific findings of the manual curation effort and provide overall statistics below. Common issues that led to QC failure were:

1. The target or chemical of interest was not listed in the paper.
2. Only a CASRN and not the chemical name was listed, and the CASRN appeared to be invalid based on an EPA Comp-Tox Dashboard database search.
3. Both the chemical and the gene target were described in the paper, but the specific chemical-target interaction was not.
4. The specific interaction between chemical and target was mentioned in the paper, but the paper lacked sufficient evidence to confirm details of the interaction.
5. The Entrez Gene ID number from the source database was incorrect.
6. The chemical only indirectly interacted with the target
7. The target indirectly regulated the level of the chemical, e.g., the chemical was a second messenger (e.g., cAMP increases as a result of ADRB2 activation)
8. The gene target was a different subtype from the one listed in the source database (e.g., ADRB2 vs. ADRB1).
9. The paper was not in English.
10. The provided version of the paper was illegible, i.e., the paper was so old that only a poorly photocopied version was available.
11. Multiple chemicals were tested together, but not individually.
12. Other issues that were seen, but which did not necessarily lead to a QC fail include:
  - a. Records from multiple sources pointed to the same PMID, but the sources recorded different modes or activities. In these cases, the paper was reexamined for alternative modes
  - b. The chemical had a different form or chirality (e.g., enantiomers, salts, hydrates) than was listed in the source database

Data from the manual curation step were analyzed to determine the reliability of the data from each source (Tab. S11.2<sup>4</sup>). Two-sample tests for equality of proportions were performed across all source-averaged pass rates. The rates of positive QC results were compared in pairwise proportion tests across sources (Tab. S11.3<sup>4</sup>). There were statistical differences between all sources except ChEMBL and CTD, which also had the highest rates of agreement. Two trends are immediately clear: (1) sources ChEMBL, CTD and DrugBank had significantly higher precision rates than LitDB (~40% higher); and (2) curator B had significantly lower accuracy rates than curators A and C. For further analyses, the results of curator B were excluded. Upon excluding results from curator B, statistical differences between the curated sources ChEMBL/CTD and CTD/DrugBank were no longer observed (Tab. S11.4<sup>4</sup>), though precision rates of all curated sources maintained statistical differences from LitDB, the source that had undergone no previous manual curation. Because of this relatively low rate of precision, LitDB references were excluded from further analyses. Next, assuming that our curators were no more or less perfect than the initial curators for the source databases, we had a second reviewing curator check a subset of records. From the references evaluated in Tab. S11.2<sup>4</sup>, reviewing curator A checked 33 of curator C's records and 17 of curator B's records. Reviewing curator C checked 33 of curator A's records and 17 of curator B's records. These records were randomly chosen to be about half QC passes and half QC fails. The matching percentage was calculated by comparing the reviewing curator's QC status with the original curator's. Reviewing curators were not statistically different in their judgement ( $p = 0.8$ ). The results in Table S11.4<sup>4</sup> show a matching percentage of approximately 75-80% for the curators A and C, which was marginally below the precision rates in Table S11.2<sup>4</sup> (excluding curator B and source LitDB) but not statistically significant ( $p = 0.227$ ). Inter-rater agreement between curator A and C was moderate ( $\kappa = 0.54$ ). From these results, we conclude that any given result in the RefChemDB database that was curated from the literature is correct with a frequency between approximately 75-90% of the time. One consequence of this is that a single report of a chemical-target-mode-activity combination is not immediately trustworthy, but multiple independent reports of the same combination are more reliable.

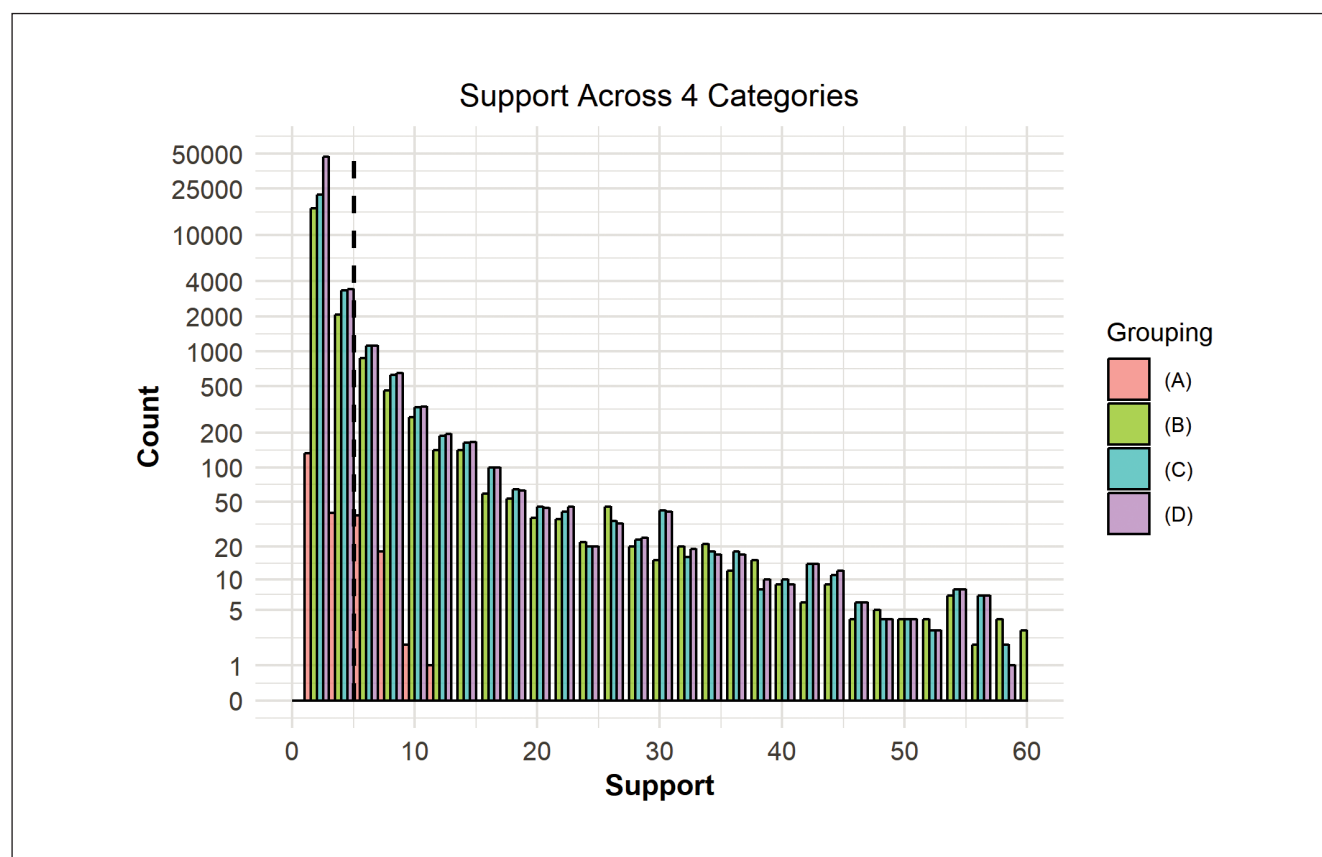
### 3.4 Candidate reference chemicals

After extracting from the database, consolidating modes, and tallying support for chemical-target-mode-activity combinations, those chemical-target-mode combinations with support values  $\geq 5$  were chosen as candidate reference chemicals. The list of support tallies is available in Table S12<sup>4</sup>. Because of the large number of alternate mode terms used across different types of assays, we have consolidated the different terms to “positive”, “negative” and “unspecified” for the mode mapping. The modes mapped to the positive class are activator, agonist, enhancer, inducer, opener, partial agonist, positive allosteric modulator, positive modulator, releasing agent and stimulator. The modes mapped to the negative class are allosteric antagonist, antagonist, blockade, blocker, channel blocker, gating inhibitor, inactivator, inhibitor, inhibitor covalent, inhibitor reversible, inverse agonist, negative allosteric modulator, negative modulator, suppressor

and uptake inhibitor. The complete set of modes and mappings is given in Table S14<sup>4</sup>. Table 3 shows the number of candidate reference chemicals from the final RefChemDB database for the 20 targets with the largest number of candidate reference chemicals and their corresponding cumulative support totals. Only chemicals having support values of  $\geq 5$  are shown. The final candidate reference chemicals by gene target are included in Table S12<sup>4</sup>. The distribution of support for each chemical-target-mode-activity combination is summarized in Figure 4.

### 4 Discussion

The use of *in vitro* assays to evaluate chemical safety is becoming more prominent due to potential regulatory uses, increased relevance to human biology, cost and testing efficiencies, and an-



**Fig. 4: Histogram of support where “count” is the number of target-chemical-mode-activity combinations**

The x-axis shows the amount of support, defined as literature records linking a chemical to a target, and the y-axis shows the frequency, or number of target-chemical combinations, with that support level. Data is shown across four categories: Category 1 (manually curated records), Category 2 (non-curated records with PMIDs), Category 3 (records without PMIDs), and Category 4 (ToxCast/Tox21 records). A “record” is a data point represented by one database record. Groups (A), (B), (C), and (D) each include different categories of data: (A) Category 1, or just curated data; (B) Category 1 + Category 2, or just data with PMIDs; (C) Category 1 + Category 2 + Category 3, or all data except *in vitro* ToxCast/Tox21 data; (D) Category 1 + Category 2 + Category 3 + Category 4, or all data. The vertical dashed line shows the cutoff at a support value of 5, the minimum threshold for candidate reference chemicals for a target with a given mode and activity call. Only chemicals with support value  $> 1$  and  $< 61$  are shown for ease of visualization.



**Tab. 3: The 20 most data-rich targets based on counts of Category 1, 2, and 3 data (all data excluding ToxCast/Tox21) for chemicals with support  $\geq 5$ , where support is the number of independent reports in the database linking a chemical to a target. The total support for those chemicals is also shown; sorted in descending order of support value.**

Target symbol	Target name	Chemicals	Support
CA2	Carbonic Anhydrase 2	106	2453
CA1	Carbonic Anhydrase 1	105	1981
ESR1	Estrogen Receptor 1	85	1371
DRD2	Dopamine Receptor D2	81	952
AR	Androgen Receptor	63	750
ESR2	Estrogen Receptor 2	41	664
PTGS2	Prostaglandin-Endoperoxide Synthase 2	46	646
OPRM1	Opioid Receptor Mu 1	48	638
CA9	Carbonic Anhydrase 9	23	512
PPARA	Peroxisome Proliferator-Activated Receptor Alpha	27	511
PPARG	Peroxisome Proliferator-Activated Receptor Gamma	26	486
NR3C1	Nuclear Receptor Subfamily 3 Group C Member 1	37	482
HTR2A	5-Hydroxytryptamine Receptor 2A (Serotonin Receptor 2A)	43	476
ADRB2	Beta-2 Adrenergic Receptor	44	472
ACHE	Acetylcholinesterase	28	470
SLC6A4	Solute Carrier Family 6 Member 4	40	423
ABCB1	ATP Binding Cassette Subfamily B Member 1	35	418
KCNH2	Potassium Voltage-Gated Channel Subfamily H Member 2	48	412
HRH1	Histamine Receptor H1	45	399
HDAC1	Histone Deacetylase 1	12	387

imal welfare concerns. To increase scientific confidence in the use of alternatives for regulatory decisions, evaluation frameworks propose that the reliability and performance of the *in vitro* assays be evaluated using reference chemicals (National Toxicology Program, 2018). The goal of this current work is to assess how well a semi-automated approach works for identifying candidate reference chemicals for use in evaluating the performance of *in vitro* assays, and how many reference chemicals one can find for a range of biological targets of interest. In this paper, we introduced a method for identifying candidate reference chemicals and annotating them in a standardized way. To work through our proposed methodology, we created a database by extracting information from multiple data sources and parsing it into relevant tables. A first evaluation compared the candidate reference chemicals' targets and modes against a large database of *in vitro* assay data from the ToxCast database. This analysis showed that for chemical-target-mode-activity combinations with  $\geq 5$  non-ToxCast references (support), the activity mostly matched with the *in vitro* ToxCast data, indicating that these candidate reference chemicals are largely trustworthy. In a second evaluation of this method, a subset of the data containing PMIDs

was manually checked. Reliability of expert-curated data sources (ChEMBL, CTD, and DrugBank) ranges from  $\sim 75\%$  to  $\sim 90\%$  according to hand curation of active/inactive assignments undertaken in this effort. This is consistent with bioactivity database assessment values reported in the literature, such as for ChEMBL version 14, which had an error rate estimate of approximately 13.6% (Tiikkainen et al., 2013).

The manual validation step revealed that sources varied in their reliability. For example, sources like ChEMBL that already comprise manually curated data have higher accuracy than LitDB, a source with no previous expert curation. Given that extracting from LitDB is purely automated, easy, and yields many candidate chemical-target pairs, it could function as a supporting or exploratory step in identifying candidate reference chemicals. A data file with LitDB's contents is provided in Table S13<sup>4</sup> and a document detailing the processing of the automated data is given in supplementary data S4<sup>15</sup>. Manual curation of the entire current database is unrealistic given time and resource constraints (413,248 records, excluding ToxCast/Tox21). However, our recommendations for developing reference chemical lists for specific targets (see below) do involve a manual curation step. We



saw variation between reviewer designations of the same source documents. A second reviewer agreed with an initial quality control status designation a mean of 79% of the time. This indicates that an element of human judgement is involved in interpretation of information found in papers. In addition, some QC failures occurred due to an inability to find the chemical listed or interpreting different names for the same chemical. For instance, there were papers that did not state the chemical name in text and only contained an image of the chemical structure or a chart with assigned R-groups linked to an image of a chemical substructure. To minimize differences in judgement, curators should adhere to the check-list developed from experience with manually curating large numbers of documents. The issues observed here with respect to the accuracy of online chemical-target databases have been documented previously (Tiikkainen et al., 2013). These authors assessed 3 sources: ChEMBL (version 14), Linceptor, and WOMBAT, and looked at inconsistencies of overlapping data extracted from same source documents. They hypothesized when 2 of 3 had identical data, the discrepant 3<sup>rd</sup> was assumed to be incorrect, but this proved false (82.2% of the time the 2 matching data values were correct, in 6.7% of instances the 3<sup>rd</sup> discrepant data value was correct, and in 11.1% the source article was ambiguous). These authors speculated that estimated error rates are likely to be even higher than reported because only errors from ligand (chemical), target, activity value, and activity type were taken into account.

With only a few exceptions, the publicly available sources (excluding ToxCast/Tox21) provide only positive reference chemicals. Ideally, during assay validation, one would also test a set of negative reference compounds to assess assay specificity (positive reference compounds help assess sensitivity) and help define the chemical domain of applicability. The ToxCast and Tox21 databases provide a large number of negative compounds, because for each assay, between 1000 and 8000 unique chemicals were tested. As a result, for targets where ToxCast and Tox21 contain multiple assays (e.g., nuclear receptors), chemicals that are negative in all corresponding assays or within different technology platforms (e.g., cellular reporter assays, biochemical binding assays) are good candidates to be negative reference chemicals. Another important aspect of assay evaluation is seeing what kinds of chemical treatment (chemical, dose) will lead to potential false positive activity. We have documented the degree of false activity that occurs, especially at high concentration, due to a variety of cell stress mechanisms (Judson et al., 2016). This is just an extension of the idea of pan-active assay interference compounds or PAINs (Baell and Holloway, 2010; Bruns and Watson, 2012). Some of this false positive activity is specific to the assay technology (cell type, readout). An active area of research is the development of assay interference reference chemicals specific to assay technology types. The databases we use will contain false positive and false negative results for reasons beyond what has already been mentioned. Different laboratories may run assays to different maximum concentrations, so weakly potent compounds may be active in one but not the other (one explanation for the trend seen in Fig. 3). Different assays for the same target can also have different chemical kinet-

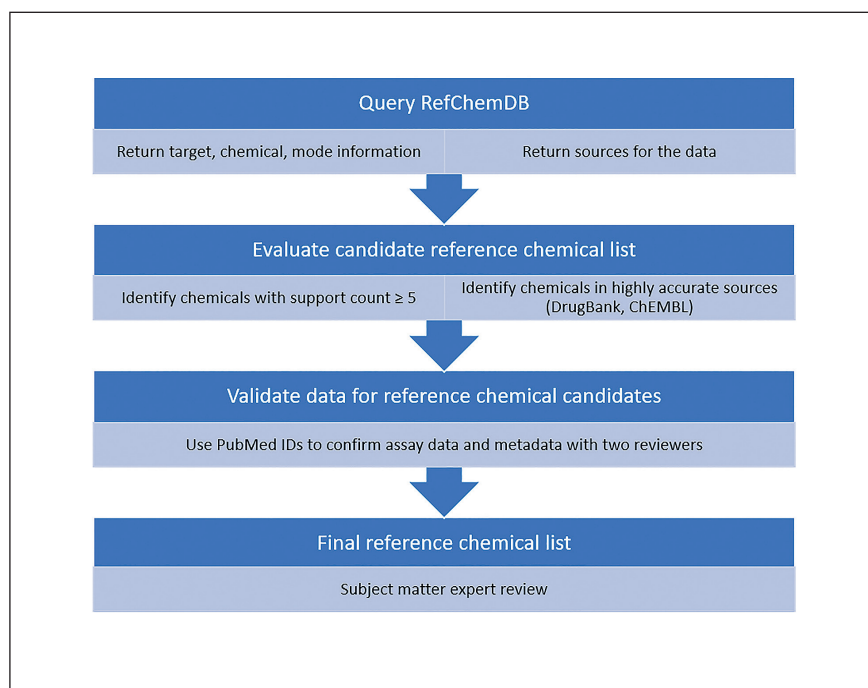
ics, affecting how much chemical gets into the cells. Factors that can drive this are the cell type and number, plate format (e.g., 96- vs. 384-well), the concentration of protein in the medium, the type of plastic used, etc. Recent work has demonstrated that incorporating such factors into reference chemical comparisons allows for quantitative *in vitro* to *in vivo* extrapolation and establishes scientific confidence in using *in vitro* approaches for risk assessment (Casey et al., 2018).

The current database provides an indication of the limit of availability of possible reference chemical data across many targets. We have incorporated data from all known public sources except for the large screening libraries in PubChem. These were excluded for two reasons. First, there are few cases where a given target has been probed with multiple assays from different labs. Second, these large libraries are primarily experimental compounds (drug screening libraries), where the compounds would not be widely available from commercial vendors. From a practical standpoint, a reference chemical should be widely available at a reasonable cost. With our current database, there are 451 gene targets with at least 2 candidate reference chemicals with support  $\geq 5$ . This analysis excluded information from the automated LitDB process. Note that, despite the relatively low precision rates from the manual reviewers for LitDB (~40%), it is still high enough to be of practical use in compiling candidate reference chemical information for specific new targets. This is especially true given that most targets will have only a small number of good reference chemicals. From Figure 4, we see that the modal number of reference chemicals for a target where support  $\geq 5$  is a single reference chemical, as illustrated in Figure 5.

Given all this information, we recommend the following tiered approach for developing a set of reference chemicals for a new target/assay:

- Using the gene identifier of the target of interest, query RefChemDB (including ToxCast/Tox21) to select records for that target and the desired mode or modes (antagonist, agonist). The chemicals that are returned are candidate reference chemicals. This candidate set can be evaluated in two ways: 1) The total count of records for each chemical provides one measure of support. 2) The occurrence of a chemical in a highly accurate source like DrugBank is another measure. The chemicals should be roughly ranked by their level of support. Chemicals with at least 5 supporting records, for instance, could be prioritized for follow-up.
- Next, the actual data for each chemical should be manually investigated for its appropriateness to the specific assay in question. This step requires reading the publications. Use the PubMed IDs provided to look up the article and extract the chemical, mode, potency metrics, and specific assay information, potentially from the cited literature as well. Assay metadata would include cell line and assay technology at a minimum. We recommend manual curation with at least two reviewers per reference and a focus on the consensus data between them.

A second phase of literature examination may be useful to identify more definitive information about the interaction of a chem-



**Fig. 5: Summary of overall process of developing a final reference chemical list**

ical and a target. For this deeper dive, the articles identified by LitDB may be useful. One tool we use is the EPA's PubMed Abstract Sifter either in the Excel version or the implementation on the EPA Comptox Chemicals Dashboard (Baker et al., 2017; Williams et al., 2017).

Reference chemicals are critical for evaluating the performance of *in vitro* assays, predictive models, and other next-generation methods to generate confidence in these approaches in the eyes of the regulatory community and other end-users. The diverse and large number of targets and pathways required to implement Toxicity Testing in the 21<sup>st</sup> Century makes reliance on past methods for defining reference chemicals impractical. The approach described here provides a rapid means of producing initial lists of reference chemicals on large numbers of targets based on the accumulated knowledge available in public databases and the scientific literature.

## References

- Attene-Ramos, M. S., Miller, N., Huang, R. et al. (2013). The Tox21 robotic platform for the assessment of environmental chemicals – From vision to reality. *Drug Discov Today* 18, 716-723. doi:10.1016/j.drudis.2013.05.015
- Attene-Ramos, M. S., Huang, R., Michael, S. et al. (2015). Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. *Environ Health Perspect* 123, 49-56. doi:10.1289/ehp.1408642
- Baell, J. B. and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53, 2719-2740. doi:10.1021/jm901137j
- Baker, N. C. and Hemminger, B. M. (2010). Mining connections between chemicals, proteins, and diseases extracted from Medline annotations. *J Biomed Inform* 43, 510-519. doi:10.1016/j.jbi.2010.03.008
- Baker, N., Knudsen, T. and Williams, A. (2017). Abstract sifter: A comprehensive front-end system to PubMed. *F1000Res* 6. doi:10.12688/f1000research.12865.1
- Bleicher, K. H., Bohm, H. J. Muller, K. and Alanine, A. I. (2003). Hit and lead generation: Beyond high-throughput screening. *Nat Rev Drug Discov* 2, 369-378. doi:10.1038/nrd1086
- Bruns, R. F. and Watson, I. A. (2012). Rules for identifying potentially reactive or promiscuous compounds. *J Med Chem* 55, 9763-9772. doi:10.1021/jm301008n
- Casey, W. M., Chang, X., Allen, D. G. et al. (2018). Evaluation and optimization of pharmacokinetic models for *in vitro* to *in vivo* extrapolation of estrogenic activity for environmental chemicals. *Environ Health Perspect* 126, 97001. doi:10.1289/EHP1655
- Casida, J. E. (2017). Why prodrugs and propesticides succeed. *Chem Res Toxicol* 30, 1117-1126. doi:10.1021/acs.chemrestox.7b00030
- Chiu, W. A., Guyton, K. Z., Martin, M. T. et al. (2018). Use of high-throughput *in vitro* toxicity screening data in cancer hazard evaluations by IARC Monograph Working Groups. *ALTEX* 35, 51-64. doi:10.14573/altex.1703231
- Corsello, S. M., Bittker, J. A., Liu, Z. et al. (2017). The drug repurposing hub: A next-generation drug library and information resource. *Nat Med* 23, 405-408. doi:10.1038/nm.4306

- Davis, A. P., Murphy, C. G., Saraceni-Richards, C. A. et al. (2009). Comparative toxicogenomics database: A knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res* 37 (Database issue), D786-792. doi:10.1093/nar/gkn580
- Dix, D. J., Houck, K. A., Martin, M. T. et al. (2007). The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* 95, 5-12. doi:10.1093/toxsci/kfl103
- Feng, B. Y. and Shoichet, B. K. (2006). A detergent-based assay for the detection of promiscuous inhibitors. *Nat Protoc* 1, 550-553. doi:10.1038/nprot.2006.77
- Gaulton, A., Bellis, L. J., Bento, A. P. et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40 (Database issue), D1100-1107. doi:10.1093/nar/gkr777
- Huber, K., Doyon, G., Plaks, J. et al. (2011). Inhibitors of histone deacetylases: Correlation between isoform specificity and reactivation of HIV type 1 (HIV-1) from latently infected cells. *J Biol Chem* 286, 22211-22218. doi:10.1074/jbc.M110.180224
- Huong, T. T., Dung, D. T., Huan, N. V. et al. (2017). Novel N-hydroxybenzamides incorporating 2-oxoindoline with unexpected potent histone deacetylase inhibitory effects and anti-tumor cytotoxicity. *Bioorg Chem* 71, 160-169. doi:10.1016/j.bioorg.2017.02.002
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *J Comput and Graphical Statistics* 5, 299-314.
- Jaworska, J., Nikolova-Jeliazkova, N. and Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set descriptor space: A review. *Altern Lab Anim* 33, 445-459.
- Judson, R. S., Houck, K. A., Kavlock, R. J. et al. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ Health Perspect* 118, 485-492. doi:10.1289/ehp.0901392
- Judson, R., Houck, K., Martin, M. et al. (2014). In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast programme. *Basic Clin Pharmacol Toxicol* 115, 69-76. doi:10.1111/bcpt.12239
- Judson, R. S., Magpantay, F. M., Chickarmane, V. et al. (2015). Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol Sci* 148, 137-154. doi:10.1093/toxsci/kfv168
- Judson, R., Houck, K., Martin, M. et al. (2016). Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol Sci* 152, 323-339. doi:10.1093/toxsci/kfw092
- Kalac, M., Scotto, L., Marchi, E. et al. (2011). HDAC inhibitors and decitabine are highly synergistic and associated with unique gene-expression and epigenetic profiles in models of DLBCL. *Blood* 118, 5506-5516. doi:10.1182/blood-2011-02-336891
- Kanehisa, M., Goto, S., Hattori, M. et al. (2006). From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res* 34 (Database issue), D354-357. doi:10.1093/nar/gkj102
- Kanehisa, M., Araki, M., Goto, S. et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36 (Database issue), D480-484.
- Kavlock, R., Chandler, K., Houck, K. et al. (2012). Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem Res Toxicol* 25, 1287-1302. doi:10.1021/tx3000939
- Kleinstreuer, N. C., Ceger, P., Watt, E. D. et al. (2017). Development and validation of a computational model for androgen receptor activity. *Chem Res Toxicol* 30, 946-964. doi:10.1021/acs.chemrestox.6b00347
- Li, Y. H., Yu, C. Y., Li, X. X. et al. (2017). Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res* 46, D1121-1127.
- Malo, N., Hanley, J. A., Cerquozzi, S. et al. (2006). Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 24, 167-175. doi:10.1038/nbt1186
- Mansouri, K., Abdelaziz, A., Rybacka, A. et al. (2016). CERAPP: Collaborative estrogen receptor activity prediction project. *Environ Health Perspect* 124, 1023-1033. doi:10.1289/ehp.1510267
- Mattingly, C. J., Rosenstein, M. C., Davis, A. P. et al. (2006). The comparative toxicogenomics database: A cross-species resource for building chemical-gene interaction networks. *Toxicol Sci* 92, 587-595. doi:10.1093/toxsci/kfl008
- National Toxicology Program (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States.
- NRC (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington D.C., USA: National Academies Press.
- Richard, A. M., Gold, L. S. and Nicklaus, M. C. (2006). Chemical structure indexing of toxicity data on the internet: Moving toward a flat world. *Curr Opin Drug Discov Devel* 9, 314-325.
- Shang, Y. and Brown, M. (2002). Molecular determinants for the tissue specificity of SERMs. *Science* 295, 2465-2468. doi:10.1126/science.1068537
- Sharma, R., Schurer, S. C. and Muskal, S. M. (2016). High quality, small molecule-activity datasets for kinase research. *F1000Res* 5. doi:10.12688/f1000research.8950.1
- Thorne, N., Auld, D. S. and Inglese, J. (2010). Apparent activity in high-throughput screening: Origins of compound-dependent assay interference. *Curr Opin Chem Biol* 14, 315-324. doi:10.1016/j.cbpa.2010.03.020
- Tice, R. R., Austin, C. P., Kavlock, R. J. and Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: A tox21 update. *Environ Health Perspect* 121, 756-765. doi:10.1289/ehp.1205784
- Tiikkainen, P., Bellis, L., Light, Y. and Franke, L. (2013). Estimating error rates in bioactivity databases. *J Chem Inf Model* 53, 2499-2505. doi:10.1021/ci400099q
- Vinken, M. (2013). The adverse outcome pathway concept:



- A pragmatic tool in toxicology. *Toxicology* 312, 158-165. doi:10.1016/j.tox.2013.08.011
- Whitebread, S., Hamon, J., Bojanic, D. and Urban, L. (2005). Keynote review: In vitro safety pharmacology profiling: An essential tool for successful drug development. *Drug Discov Today* 10, 1421-1433. doi:10.1016/S1359-6446(05)03632-9
- Wieggers, T. C., Davis, A. P., Cohen, K. B. et al. (2009). Text mining and manual curation of chemical-gene-disease networks for the comparative toxicogenomics database (CTD). *BMC Bioinformatics* 10, 326. doi:10.1186/1471-2105-10-326
- Williams, A. J., Grulke, C. M., Edwards, J. et al. (2017). The CompTox Chemistry Dashboard: A community data resource for environmental chemistry. *J Cheminform* 9, 61. doi:10.1186/s13321-017-0247-6
- Wishart, D. S., Knox, C., Guo, A. C. et al. (2006). DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 34 (Database issue), D668-672. doi:10.1093/nar/gkj067
- Wishart, D. S., Knox, C., Guo, A. C. et al. (2008). DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 36 (Database issue), D901-906. doi:10.1093/nar/gkm958
- Zhang, J. H., Chung, T. D. and Oldenburg, K. R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 4, 67-73. doi:10.1177/108705719900400206

### Conflict of interest

The authors certify that they do not have any conflicts of interest to declare.

### Acknowledgement

All funding for this research was provided by the US Environmental Protection Agency.

## 3R und Ersatzmethoden - bessere Forschung, weniger Tierleid

12. Tierversuchstagung des Schweizer Tierschutz STS

Donnerstag, 20. Juni 2019

Beginn: 09:00 Uhr

Kongresszentrum Hotel Arte

Riggenbachstrasse 10

4600 Olten



Die Tagung wird von der Vereinigung der Schweizer Kantonstierärztinnen und Kantonstierärzte (VSKT) den zuständigen kantonalen Behörden zur Anerkennung im Rahmen der Aus- und Weiterbildung von Fachpersonal für Tierversuche empfohlen sowie von der Gesellschaft Schweizer Tierärztinnen und Tierärzte GST mit zwei Bildungspunkten anerkannt.

### Tagungsgebühr

(inkl. Verpflegung und Tagungsunterlagen)

Vollzahler(in) CHF 190.-

Student(in) CHF 95.-

Tagungssprache: Hochdeutsch, Französisch, Englisch

Simultanübersetzung: Deutsch-Französisch-Englisch und Englisch-Französisch-Deutsch

### Anmeldung

Schweizer Tierschutz STS

Geschäftsstelle

Dornacherstrasse 101, Postfach  
4018 Basel

Tel. 0041-(0)61-365 99 99

www.tierschutz.com

