OXFORD

SOT | Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 167(2), 2019, 593–603

doi: 10.1093/toxsci/kfy265
Advance Access Publication Date: October 22, 2018
Research Article

Rationalizing Secondary Pharmacology Screening Using Human Genetic and Pharmacological Evidence

Aimee M. Deaton,[1] Fan Fan, Wei Zhang, Phuong A. Nguyen, Lucas D. Ward, and Paul Nioi

Comparative Biology and Safety Sciences, Amgen Inc, Cambridge, Massachusetts 02142

[1]To whom correspondence should be addressed at Amgen, Inc., 360 Binney St, Cambridge, MA 02142; FAX 805-447-1010. E-mail: adeaton@amgen.com.

ABSTRACT

Safety-related drug failures remain a major challenge for the pharmaceutical industry. One approach to ensuring drug safety involves assessing small molecule drug specificity by examining the ability of a drug candidate to interact with a panel of "off-target" proteins, referred to as secondary pharmacology screening. Information from human genetics and pharmacology can be used to select proteins associated with adverse effects for such screening. In an analysis of marketed drugs, we found a clear relationship between the genetic and pharmacological phenotypes of a drug's off-target proteins and the observed drug side effects. In addition to using this phenotypic information for the selection of secondary pharmacology screens, we also show that it can be used to help identify drug off-target protein interactions responsible for drug-related adverse events. We anticipate that this phenotype-driven approach to secondary pharmacology screening will help to reduce safety-related drug failures due to drug off-target protein interactions.

Key words: drug safety; secondary pharmacology screening; off-targets; genetics.

Unintended adverse side effects remain a major reason for clinical trial failure and postmarketing drug withdrawal (Hamon *et al.*, 2009; Kola and Landis, 2004; Roberts *et al.*, 2014). One approach used to improve drug safety is to ensure specificity of the drug candidate against the intended target or, at the very least, to eliminate off-target interactions likely to cause adverse events. To help ensure small molecule drug specificity, pharmaceutical companies often perform *in vitro* secondary pharmacology screening whereby a compound is assessed for its ability to bind to and/or modulate a variety of off-target proteins (Bowes *et al.*, 2012; Whitebread *et al.*, 2016).

Currently, regulatory authorities do not mandate any specific secondary pharmacology screening, except for testing whether drugs inhibit the Kv11.1 or "hERG" potassium channel (encoded by *KCNH2*) (ICH, 2005), a known mechanism through which drugs can cause cardiac arrhythmias (Hamon *et al.*, 2009). Knowing what proteins to select for a counter screen is difficult and requires knowledge of a given protein's function, and the consequences of it being inadvertently "drugged" in humans. In many ways this challenge is similar to that faced by drug developers when trying to determine what proteins to target to treat disease; understanding the outcome of drugging a certain target is difficult and often is not determined until the completion of lengthy clinical trials. Indeed, the FDA has highlighted the fact that proteins included on counter screens often lack a clear link to human safety (Papoian *et al.*, 2015) although when four pharmaceutical companies compared their secondary pharmacology screens there was some overlap in terms of the proteins included (Bowes *et al.*, 2012). To tackle these issues, there are two sources of data that provide insight into the function and consequence of drugging specific proteins in humans; pharmacology; and human genetics.

Pharmacology provides knowledge of drug indications which reveal the phenotypic effects of functionally modulating a drug target, but this is limited by the small number of human proteins that have been successfully drugged (672 out of approximately 20 000) (Ponomarenko *et al.*, 2016). Genetic evidence can also be used to predict the effects of drugging a protein and is increasingly being used in the pharmaceutical industry to select therapeutic targets (Kamb *et al.*, 2013; Plenge *et al.*, 2013). Prior

work has established genetics as a way of identifying drug targets both in large systematic analyses (Nelson *et al.*, 2015) and for individual drug targets such as PCSK9 (Abifadel *et al.*, 2003; Ference *et al.*, 2018; Hess *et al.*, 2018). Genetics could be used in a similar manner to select "off-target" proteins that we do not want drugs to modulate by helping us identify proteins whose modulation would have safety concerns. For example, loss-of-function variants in *KCNH2* (encoding the hERG channel) cause long QT syndrome (Curran *et al.*, 1995) consistent with the fact that pharmacological inhibition of hERG can result in cardiac arrhythmia. Similarly, loss-of-function variants in *CTSD* (encoding cathepsin D) cause neuronal ceroid lipofuscinosis, a retinal disease, mirroring retinal phenotypes observed in animals administered drugs that inadvertently inhibited cathepsin D (Siintola *et al.*, 2006; Steinfeld *et al.*, 2006; Zuhl *et al.*, 2016). For these examples, the relationship between the genetics of the off-target protein and drug side effects is well established. In this study, we examined whether genetics, as well as pharmacology, can be used more broadly to help predict drug side effects resulting from drug off-target protein interactions.

To do this, we systematically compared the off-target drug-protein interactions and the observed side effects of a set of marketed drugs. Layering in known genetic associations and drug indications for these off-target proteins allowed us demonstrate a clear relationship between these phenotypes and drug side effects. Our findings have implications for the selection of small molecule secondary pharmacology screens as well as for the retrospective identification of off-target proteins contributing to drug side effects.

MATERIALS AND METHODS

Data Download and Processing

Drug Side Effect and Target Information

The following pharmacology databases were downloaded: DrugBank (v5.0.6, r2017-04-01; http://www.drugbank.ca; last accessed April 10, 2017) (Knox *et al.*, 2011; Law *et al.*, 2014; Wishart *et al.*, 2006, 2008), SIDER (v4.1, r2015-10-21; http://sideeffects.embl.de/; last accessed January 14, 2016) (Kuhn *et al.*, 2010, 2016), OFFSIDES (r2012-03-14; http://tatonettilab.org/resources/tatonetti-stm.html; last accessed June 23, 2016) (Tatonetti *et al.*, 2012), and Pharmaprojects, Pharma Intelligence 2016 (d2016-11-22; https://pharmaintelligence.informa.com/; last accessed November 22, 2016).

Side effects of marketed drugs were obtained from the SIDER database that lists side effects reported on drug labels (Kuhn *et al.*, 2010, 2016). Side effect phenotypes were mapped to MedDRA terminology to allow systematic comparison between phenotype terms from different sources (see below). Side effects seen in patients treated with a placebo were removed. To get high-confidence side effects we only considered side effects also present in postmarketing physicians' reports (within the FDA Adverse Event Reporting System, or FAERS) that were statistically significant, as determined from the OFFSIDES database (Tatonetti *et al.*, 2012); p values of drug-side effect associations were used to impose a 5% false discovery rate (Benjamini and Hochberg, 1995). Side effects belonging to the "General disorders and administration site conditions" MedDRA category were removed as these were likely to be common side effects associated with drug treatment generally rather than side effects due to specific off-target interactions. The indications of these drugs were obtained from Pharmaprojects.

All proteins that interact with the set of drugs extracted from SIDER (including both intended targets and "off-targets") were identified using Prous Institute Symmetry and Chemotargets CLARITY (http://www.chemotargets.com), which integrate carefully selected data on compound-target interactions from literature, patent applications, and both publically accessible and commercial databases (Excelra GOSTAR). Bioinfogate's safety intelligence portal, OFF-X (http://www.targetsafety.info), was also used in the process. From this set of drug-protein interaction pairs, the therapeutic drug-target pairs were identified using Drugbank (Knox *et al.*, 2011) with the remainder of interactions classified as "off-target." Interactions with a number of cytochrome P450 enzymes were excluded to rule out interactions related to drug metabolism. Altogether, we examined 618 drugs with side effect and target information in the enrichment analysis. For the logistic regression we limited ourselves to small molecule drugs using Drugbank as the source of modality information and analyzed 587 drugs.

Drug target phenotype information. The phenotypic associations of these drugs' targets were obtained from both human genetics and pharmacology. To get genetic information, genes involved in Mendelian traits curated in Online Mendelian Inheritance in Man (OMIM) were obtained from the Human Phenotype Ontology (HPO) (r2017-04-13; http://human-phenotype-ontology.github.io; last accessed June 7, 2017) (Kohler *et al.*, 2017). For each OMIM syndrome, we considered only phenotype terms defined as frequently associated with the disease by HPO ("frequent frequencies") giving 3159 genes associated with 4056 OMIM syndromes, each of which was described by a list of HPO phenotype terms. If the protein was the intended target of an approved drug, pharmacological phenotypes were assigned using the indication of that drug obtained from Pharmaprojects. Where there were different variants in the same gene with varying disease associations or drugs with different indications modulating the same protein we took the union of all phenotypes to make sure we considered all potential safety risks that could be caused by modulating a particular target.

Phenotype Mapping and Comparison

To allow comparison between side effect terms and phenotypes from pharmacology and genetics we mapped all phenotypes to MedDRA terminology using the Unified Medical Language System (UMLS) Metathesaurus (2015AB release), MetaMap natural language processing (NLP) tool (MetaMap2016, r2016-01), and the UMLS-Interface software (https://www.nlm.nih.gov/research/umls/; last accessed May 9, 2016) (McInnes *et al.*, 2009). All phenotype terms were mapped to the most specific MedDRA term within the UMLS which was in turn mapped to MedDRA Higher Level Group Terms (HLGT) and System Organ Class (SOC) terms. Similarity scores for all MedDRA terms were calculated using the UMLS-Similarity modules from the UMLS-Interface software. For the enrichment analysis we defined a similarity score between two terms of greater than or equal to 0.7 as a "phenotype match."

Enrichment Analysis

We annotated phenotype matches (as defined above) between drug side effects and off-target phenotypes obtained from genetics, pharmacology, or both. To test for which phenotypes off-target genetics/pharmacology tended to reflect drug side effects we performed an enrichment analysis using 230 MedDRA HLGT

terms (ie, all HLGT terms for which there were side effects in our dataset). First, we removed all side effect terms that mapped to the same SOC term as the drug indication to avoid side effects likely to represent exaggerated pharmacology (eg, drugs with cardiac indications were removed when we considered cardiac arrhythmia as a side effect). Second, we excluded drugs where the intended target had genetic or pharmacological evidence for the phenotype reasoning that such effects were more likely to be modulated by the intended target rather than an off-target interaction. We then recorded the number of drugs that had a matching side effect and off-target phenotype (ie, the drug side effect matched the phenotype of at least one of the drug's off-targets), absence of the side effect and absence of a matching off-target phenotype, presence of side effect but absence of a matching off-target phenotype, and absence of a side effect but presence of an off-target phenotype. Drugs with phenotype matches with the intended target were removed from the analysis. We constructed 2×2 contingency tables containing the number of drugs fulfilling each of these criteria and calculated an odds ratio and p value for each HLGT term using a two-sided Fisher's exact test (Agresti, 2002; Fisher, 1935) "fisher.test" in the R "stats" package (R version 3.4.2). Fisher's exact test was chosen to be robust to small sample sizes in certain contingency tables (Kim, 2017; Ludbrook, 2008). For instances where there were zero values in the contingency table (ie, when no drugs matched the criteria) these were assigned a pseudocount of one to avoid infinite or zero odds ratio values. We corrected our significance threshold for multiple testing using the Bonferroni method which adjusts the p value based on the number of tests performed (Bland and Altman, 1995). In this instance, we examined 618 drugs over each of 230 phenotypes giving a total of 1.4×10^5 tests performed. We considered a p value of $< 3.5 \times 10^{-7}$ as significant, which is equivalent to an adjusted p value $< .05$.

Logistic Regression

To assess the correlation between off-target phenotypes (from genetics and pharmacology) and the side effect profile of a drug, we performed a multivariate logistic regression (using the "glm" function in the R "stats" package) (R version 3.4.2). Out of the 46 MedDRA HLGT phenotype terms significant in the enrichment analysis, 44 had a sufficient number of drugs with that side effect to build a model. The logistic regression model for each of these phenotypes used disease indication (21 MedDRA SOC or organ system level terms), whether the intended targets have genetic evidence matching that phenotype, and whether the off-targets have evidence for the phenotype as predictors of drug side effect. All predictors were encoded as binary variables.

Deep Neural Network Modeling of ADRA2B Activity

The R "deepnet" package version 2.0 (Warr, 2012) was used to generate a categorical deep neural network (DNN) model to predict whether a compound can bind to ADRA2B. This DNN model was trained using compounds derived from CHEMBL database (version 23, last accessed 2017-09-22) with known activities against ADRA2B (Bento *et al.*, 2014). Compounds marked as "not active" in CHEMBL were collected as "inactives" for training set. To avoid the potential discrepancy between different assays, only compounds with confirmed IC_{50} values (between 0.52 nM and 31.88 μM) in radioligand ADRA2B binding assays were used for the "active" set. A total of 824 structurally unique CHEMBL compounds, including 720 "inactives" and 104 "actives," were used to build the model. The 824 CHEMBL compounds were randomly separated into a training set (75%, 641 compounds) and a testing set (25%, 183 compounds). The model was first generated based on the training set, and then validated using the testing set. The default neural net parameters and multiple compound properties, including FCFP_4 fingerprints, AlogP, molecular weight, number of fragment, molecular polar surface area, molecular solubility, number of H donors, number of H acceptors, number of aromatic rings, etc. were used for model training. The generated DNN model reached a high predictive accuracy of 98.5% (632 of 641) for the training set, with 91.6% sensitivity of (76 of 83) and 99.6% specificity (556 of 558). Cross-validation using the testing set resulted a comparable accuracy of 95.6% (175 of 183), with 76.2% sensitivity (16 of 21) and 98.1% specificity (159 of 162).

After the DNN model was validated using the test set of compounds, it was applied to predict the activities of additional compounds not known to bind ADRA2B. From SIDER, 59 marketed drugs with seizure as a side effect whose intended targets did not include ADRA2B and that had no ADRA2B interaction documented in Symmetry were collected. Drugs containing mixtures of compounds were removed leaving 55 drugs. Out of the 55 drugs, 16 already had ADRA2B binding data in CHEMBL with 5 of these drugs showing binding to the receptor. For the 39 remaining drugs, their activity against ADRA2B was predicted using the DNN model.

Selection of Targets for Secondary Pharmacology Screen

Phenotypes were assigned to all human proteins using genetic and pharmacological evidence as described above. Using MedDRA terms, proteins with phenotypes affecting the cardiovascular, respiratory, and nervous systems were selected (SOC terms "Cardiac disorders," "Vascular disorders," "Respiratory, thoracic, and mediastinal disorders," "Nervous system disorders") giving 2542 human proteins. To construct a selectivity screen with immediate application in the pharmaceutical industry we identified which of these proteins currently have *in vitro* assays available from major suppliers (CEREP, Panlabs, DiscoveRx). We excluded DNA methyltransferases, histone methyltransferases and transcription factors (with the exception of nuclear receptors). To reduce redundancy on the panel representative members were selected. Protein families were defined using HUGO gene nomenclature committee gene family assignation. Representative proteins from families were selected by aligning all members of a family against each other using Clustal Omega (Goujon *et al.*, 2010; Larkin *et al.*, 2007). For each family member, the percent identity against all other family members was summed, and the family member with the highest value was chosen as the most representative. When strong evidence for association with a concerning cardiovascular, respiratory, or nervous system phenotype was present, this "representative" protein was chosen for inclusion in the selectivity panel. If phenotypic evidence was less convincing the next most "representative" family member was chosen.

RESULTS

The Genetic and Pharmacological Phenotypes of a Drug's Off-Targets Match Its Side Effects

We performed a retrospective analysis of marketed compounds to assess whether the genetics and intended pharmacology of a drug's off-target proteins correlate with its side effects. Drug side effects were obtained from the drug label (SIDER database)

Table 1. Phenotypes Most Significantly Enriched for Matches Between Side Effect and Off-Target Genetics/Pharmacology

Phenotype Term (MedDRA)	p Value	OR	Most Frequently Hit Off-Target
Blood platelet disorders	1.45×10^{-40}	353.97	TBXAS1
Seizures	1.82×10^{-37}	541.61	ADRA2B
Vision disorders	1.24×10^{-34}	80.94	KCNJ13
Glucose metabolism disorders	1.81×10^{-25}	66.24	KCNJ11
Movement disorders	1.10×10^{-23}	58.98	KCNA1, KCNC1, KCNA2, KCND3, KCNJ10, ADRA2B
Heart failures	3.29×10^{-23}	218.55	KCNJ5, KCNJ8
Nonhemolytic anemias	3.05×10^{-22}	125.69	TBXAS1
Cardiac and vascular investigations	6.91×10^{-22}	11.80	ADRA1B, ADRA1A, ADRA1D
Skin appendage conditions	9.96×10^{-22}	51.19	KCNJ8
Bone disorders	9.42×10^{-21}	100.83	TBXAS1
Disease coronary artery	1.83×10^{-20}	323.05	ADRA1A
Vascular hemorrhagic disorders	2.37×10^{-20}	40.39	TBXAS1
Acid-base disorders	3.53×10^{-20}	82.54	KCNJ1, KCNJ10
Cardiac arrhythmia NOS	3.10×10^{-19}	9.44	KCNQ1, KCNJ2, KCNH2

A p value $< 3.5 \times 10^{-7}$ from Fisher's exact test was considered significant based on Bonferroni correction for the 1.4×10^{-5} tests performed. This is equivalent to an adjusted p value $< .05$.

(Kuhn *et al.*, 2016) and these were intersected with side effects from postmarketing reports (OFFSIDES database) (Tatonetti *et al.*, 2012) to get high-confidence side effects for each compound. To exclude the possibility that some side effects may represent exaggerated pharmacology rather than being the result of off-target interactions, we removed side effects in the same organ system as the drug indication (eg, cardiac arrhythmia side effects were excluded for any drug treating a cardiac disorder). We annotated both the drugs' intended targets and off-targets using various databases as described in the methods. Altogether 618 individual drugs with side effect and target information were analyzed.

We then examined phenotypic associations for each "off-target" protein with which these drugs interact. Phenotypes were assigned from genetics (Mendelian diseases caused by variants in the gene encoding the protein) and pharmacology (disease indications of approved drugs targeting that protein) or both and mapped to MedDRA terminology. To test for correspondence between these genetic and pharmacological phenotypes and drug side effects, we performed an enrichment test, excluding drugs where the intended target had genetic or pharmacological evidence for involvement in the side effect phenotype.

For 230 MedDRA terms, we recorded the number of drugs that had a matching side effect and off-target phenotype (ie, the drug side effect matched the phenotype of at least one of the drug's off-targets), absence of the side effect and absence of a matching off-target phenotype, presence of side effect but absence of a matching off-target phenotype, and absence of a side effect but presence of an off-target phenotype. We constructed 2×2 contingency tables containing the number of drugs fulfilling each of these criteria and calculated an odds ratio and p value for each phenotype term using Fisher's exact test.

When considering genetic and pharmacological phenotypes combined we found that for 46 side effect categories, drugs with off-target phenotypes predicted by these data were more likely have that particular side effect. Drugs associated with blood platelet disorders ($p = 1.44 \times 10^{-40}$, OR $\approx$ 354) and seizure ($p = 5.38 \times 10^{-39}$, OR $\approx$ 541.61) side effects showed the most significant enrichment followed by drugs with side effects affecting vision and glucose metabolism (Table 1). Other phenotypes of high safety concern where drugs with off-target evidence were overrepresented were movement disorders, heart failure, vascular hemorrhage, and cardiac arrhythmia (Table 1 and Supplementary Table 1). To further explore the biology driving these enrichment results, for each significant phenotype we identified the most frequently hit off-target protein with genetic and/or pharmacological support (Table 1).

We then repeated our enrichment analysis using genetic and pharmacological sources of phenotypes separately. When only genetic associations were used, results were similar to those from the combined analysis, suggesting that genetic data drove most of these predictions (Supplementary Table 2). This might reflect the fact that fewer off-targets have been drugged than have genetic evidence; 315 off-target proteins in the analysis had genetic information whereas 202 had pharmacological information (of these, 94 proteins had both genetics and pharmacology). When considering only the pharmacological phenotypes of drug off-targets, 14 out of 230 phenotypes were significant and drugs associated with heart failure, coronary artery disease and seizure showed the most significant enrichment (Supplementary Table 3).

These results show that the phenotypes of a drug's off-target proteins inferred from genetics or pharmacology can be used to help predict its side effects and that combining both sources of phenotypic evidence is likely to be the most powerful approach to identify the most problematic proteins that a drug could unintentionally modulate.

Regression Modeling Shows That the Correlation Between Off-Target Phenotypes and Drug Side Effects Persists When Controlling for Confounders

Whereas a causal relationship between the phenotypes of a drug's off-target proteins and its side effects is mechanistically plausible, there are potential confounding factors that could contribute to the enrichment we observe, including side effects that correlate with the disease being treated. In the enrichment analysis we removed side effects where there was a likely contribution from drug indication or the genetics of the drug's intended target but may not have controlled adequately for all confounders. To address this, we tested the contribution of drug indication, on-target genetics and off-target phenotypes (from genetics and pharmacology) to the side effects of 587 small molecule drugs in a logistic regression model. We tested the 46

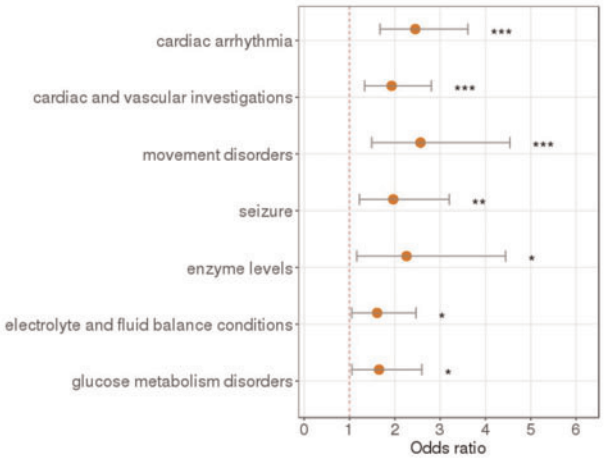

Figure 1. Results for off-target phenotypes in logistic regression models. Showing all phenotypes where off-target phenotypes had $p < .05$. Circular points indicate odds ratios from the models; error bars are 95% confidence intervals. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

phenotypes that were significant in the enrichment analysis and, out of 44 with enough data to run the model, we found that off-target phenotypes were a significant predictor of drug side effects for 7 phenotypes tested ($p < .05$). Three of these phenotypes passed a more conservative threshold adjusted for the number of side effect classes tested ($p < .001$) (Figure 1 and Supplementary Table 4). The vast majority of side effect phenotypes tested showed a positive association between off-target phenotypes and drug side effects (38/44). The side effect that had the most significant contribution from off-target genetics was cardiac arrhythmia ($p = 5 \times 10^{-6}$, OR = 2.45) (Figure 1). Thus, the contribution of off-target phenotypes to drug side effects persists even when controlling for well-established contributors to drug side effects.

Genetics and Pharmacology Can Help to Retrospectively Identify Off-Target Proteins Contributing to Drug Side Effects

Secondary pharmacology screening is ideally performed as part of the process of selecting which drug should continue into clinical trials. However, in reality, only a subset of human proteins can practically be screened against during development. This means that there is a risk that a drug could have off-target activity that is not detected until human exposure to the molecule. When such instances arise, it is critical to be able to determine which off-target the drug is interacting with so that a back-up molecule without these effects can be advanced into trials. We wanted to test whether our approach of applying phenotypic associations from human genetics and pharmacology could be useful for "issue resolution" in such instances.

In our analysis of marketed compounds, 29% of drugs causing seizure had off-target interactions with the alpha-2B adrenergic receptor (ADRA2B) in the dataset used. Recent reports have implicated gain of function mutations in *ADRA2B* in familial adult myoclonic epilepsy (De Fusco *et al.*, 2014) (OMIM:607876). We asked whether 55 drugs that have seizure as a side effect but do not have ADRA2B binding data in our off-target protein dataset might have interactions with ADRA2B that could contribute seizure. To assess this, we extracted ADRA2B binding data from ChEMBL (Bento *et al.*, 2014) and found that five of the 55 drugs with seizure as a side effect had known ADRA2B interactions (Table 2). Because most of the drugs (39/55) did not have ADRA2B binding data in ChEMBL, we generated a ligand-based DNN model to predict whether any of these 39 drugs might also bind ADRA2B. This DNN model had a high prediction accuracy of 95.6% as demonstrated by 2-fold cross validation (see Materials and Methods section).

The DNN model was then applied to predict ADRA2B binding for the 39 drugs for which seizure was documented as a side effect, but without known binding to ADRA2B. The predictions indicated that five out of these 39 drugs are likely to bind ADRA2B. Interestingly, eight of the drugs with confirmed or predicted ADRA2B binding were antipsychotics or anti-histamines and six out of these 10 drugs had structurally similar tricyclic ring scaffolds (Table 2). Notably, one of the drugs identified as an ADRA2B interactor, clozapine, is known for its risk of seizure although little was previously known about the underlying mechanism (Varma *et al.*, 2011; Williams and Park, 2015). The confirmation that 10 out of 55 drugs with seizure as a side effect have known or predicted off-target activity against ADRA2B demonstrates that human genetics can be used more broadly to prioritize proteins that may be responsible for a given side effect.

Systemic Selection of Proteins for Secondary Pharmacology Screening Using Genetics and Pharmacology

In addition to the use of *in vitro* assays to explain adverse events related to drug treatment in a preclinical or clinical setting, secondary pharmacology screens are used routinely during drug development to improve specificity and help select the best clinical candidates. We used phenotypes from genetics and pharmacology to systematically select proteins to include on such a screen (Figure 2A). Proteins associated with phenotypes affecting the cardiovascular, respiratory and central nervous systems were prioritized for inclusion, as disrupting the function of such vital organ systems has the potential to result in serious life-threatening conditions. This is in line with international regulatory guidance emphasizing that these organ systems are the most critical for safety pharmacology (ICH, 2001). To design a panel that could immediately be used by the pharmaceutical industry we identified which of those proteins have commercially available *in vitro* assays. Furthermore, to reduce redundancy on the panel, representative family members were selected in cases where more than three related proteins were under consideration (see Materials and Methods section). This resulted in a secondary pharmacology screen consisting of 70 targets from diverse protein classes (Figure 2B) associated with cardiovascular, respiratory and nervous system phenotypes (Figure 2C, Table 3 and Supplementary Table 5).

A number of proteins on our panel have phenotypic associations that come solely from human genetics including the hyperpolarization-activated ion channel HCN4 and transient receptor potential cation channel TRPM4, which are included based on links between mutations in the genes encoding these proteins and cardiac conduction defects (Kruse *et al.*, 2009; Liu *et al.*, 2010; Milanesi *et al.*, 2006; Milano *et al.*, 2014; Nof *et al.*, 2007; Schulze-Bahr *et al.*, 2003; Schweizer *et al.*, 2014; Stallmeyer *et al.*, 2012; Ueda *et al.*, 2009). Some proteins on the screen have evidence coming from both pharmacology and genetics. One such example is the purinergic receptor P2RY12 which is a target of the anti-clotting drugs clopidogrel and ticlopidine (Boeynaems *et al.*, 2005; Dorsam *et al.*, 2003; Herbert and Savi, 2003). Consistent with a biological role in blood clotting, loss-of-function mutations in the *P2RY12* gene cause a platelet-type bleeding disorder (Cattaneo *et al.*, 2003; Nurden *et al.*, 1995).

By systematically choosing proteins with known phenotypic associations for screening, we can eliminate interactions

Table 2. Drugs With Seizure As A Side Effect: Drugs With Seizure As A Side Effect That Have Known ADRA2B Binding From ChEMBL ("Known Active") or Predicted Binding From the ADRA2B DNN Model ("Predicted Active", Score from Model Is Shown)

Drug	Activity Against ADRA2B	ADRA2B Activity Score	Intended Target	Drug Indication
Chlorpromazine	Known active	1 (known)	DRD2, DRD1, HTR2A, HTR1A, ADRA1A, ADRA1B, HRH1	Schizophrenia, nausea, vomiting, agitation
Cyproheptadine	Known active	1 (known)	HRH1, HTR2A, HTR2C	Allergic rhinitis, angioedema, urticaria, anaphylaxis
Maprotiline	Known active	1 (known)	SLC6A2	Depression
Quetiapine	Known active	1 (known)	HTR2A, DRD2	Bipolar disorder, schizophrenia, depression
Desloratadine	Known active	1 (known)	HRH1	Rhinitis, urticaria, eczema, pruritus
Clozapine	Predicted active	0.98	DRD2, HTR2A	Schizophrenia
Hydroxyzine	Predicted active	0.88	HRH1	Anxiety, pruritus, allergy
Duloxetine	Predicted active	0.71	SLC6A4, SLC6A2	Depression, pain, anxiety disorder
Eletriptan	Predicted active	0.74	HTR1D, HTR1B, HTR1F	Migraine
Aprepitant	Predicted active	0.96	TACR1	Nausea, vomiting

The intended target of these drugs (from Drugbank) along with the drug indication (from Drugbank/Pharmaprojects) is shown.

between drug candidates and the most concerning off-target proteins. Furthermore, if certain drug off-target protein interactions cannot be avoided, knowing the potential consequences of such interactions allows us to monitor for them in preclinical and clinical studies.

DISCUSSION

We show that drug side effect can be predicted, in part, from the genetics and pharmacology of its "off-target" protein interactions. We examined the contribution made by the phenotypes of a drug's off-target proteins to its side effects by examining a set of marketed drugs using two different methods—enrichment analysis and logistic regression.

The enrichment analysis showed that for 46 side effect phenotypes, including many of high safety concern such as seizure and platelet disorders, drugs with off-target evidence for the same phenotype were more likely to have that side effect. We then attempted to more comprehensively model the effect of off-target phenotype, drug indication, and on-target genetics on drug side effects using logistic regression. For seven side effect phenotypes we found that off-target genetics and/or pharmacology were a significant predictor of drug side effects, with cardiac arrhythmia having the most significant effect. The reasons why the phenotypes significant in the enrichment analysis are not all significant in the logistic regression may be due to an inability to separate the contribution of drug off-target protein interactions from the contribution of other predictors such as drug indication or due to limited power of the regression models due to the relatively small number of drugs examined. Nonetheless these results, taken together with the enrichment analysis, suggest that using phenotypes from genetics and pharmacology can help to identify drug off-target protein interactions contributing to side effects.

A challenge with this type of analysis is having a comprehensive set of "off-target" interactions available for each drug. Here we used a curated database of *in vitro* drug-target interactions, a limitation of which is a lack of negative experimental results (ie, a set of targets with which a drug does not interact). Furthermore, available *in vitro* data may be skewed towards highly studied proteins. In the future, as such data sources expand and methods of *in silico* target prediction improve, more comprehensive information on drug off-target interactions may strengthen such analyses. In addition, if a larger set of well-curated drug side effect information were available this would increase statistical power and perhaps allow us detect effects for additional phenotypes.

For some phenotypes certain off-targets predominate, for example drugs causing platelet side effect frequently modulated thromboxane synthase (TBXAS1) and drugs with seizure side effects often interacted with ADRA2B as well as a number of potassium channels. This could be due to a number of factors. One of these may be the propensity of small molecules to interact with these proteins due to structural features or other protein properties. In the case of TBXAS1, although its structure remains undermined, a detailed fluorescence spectroscopy study demonstrated that the hydrophobic active site of the enzyme is large so that TBXAS1 could accommodate multiple ligands simultaneously. The ligand binding is further facilitated by a cluster of phenylalanine residues near the ligand binding pocket, which increases the chance of unspecific entropic binding (Chao *et al.*, 2013). In the case of ADRA2B and many other GPCR targets, their flexible and half buried binding pockets tend to attract compounds with varied chemical structures containing hydrophobic fragment(s) (Maudsley *et al.*, 2005). Another explanation is that proteins similar to a drug's intended target are more likely to have unintended interactions with that molecule. As approximately 34% of marketed drugs modulate GPCRs to exert their therapeutic effects (Hauser *et al.*, 2017; Rask-Andersen *et al.*, 2014; Santos *et al.*, 2017) this may partially explain the propensity of ADRA2B to have off-target interactions with some of these molecules. Such frequently occurring proteins from our analysis could be prioritized for off-target screening when investigating drug-protein interactions that are responsible for particular side effects.

During the drug development process secondary pharmacology screening is used for two main applications and prioritizing proteins using genetic and pharmacological evidence can be useful for both. First, assessment of drug off-target protein interactions can be used to explain which proteins may be mediating particular adverse events seen in preclinical or clinical studies. Mimicking this situation, we examined whether drugs causing seizure interacted with alpha-2B adrenergic receptor (ADRA2B), a protein with genetic evidence for involvement in epilepsy. We found that 10 out of 55 drugs examined had known

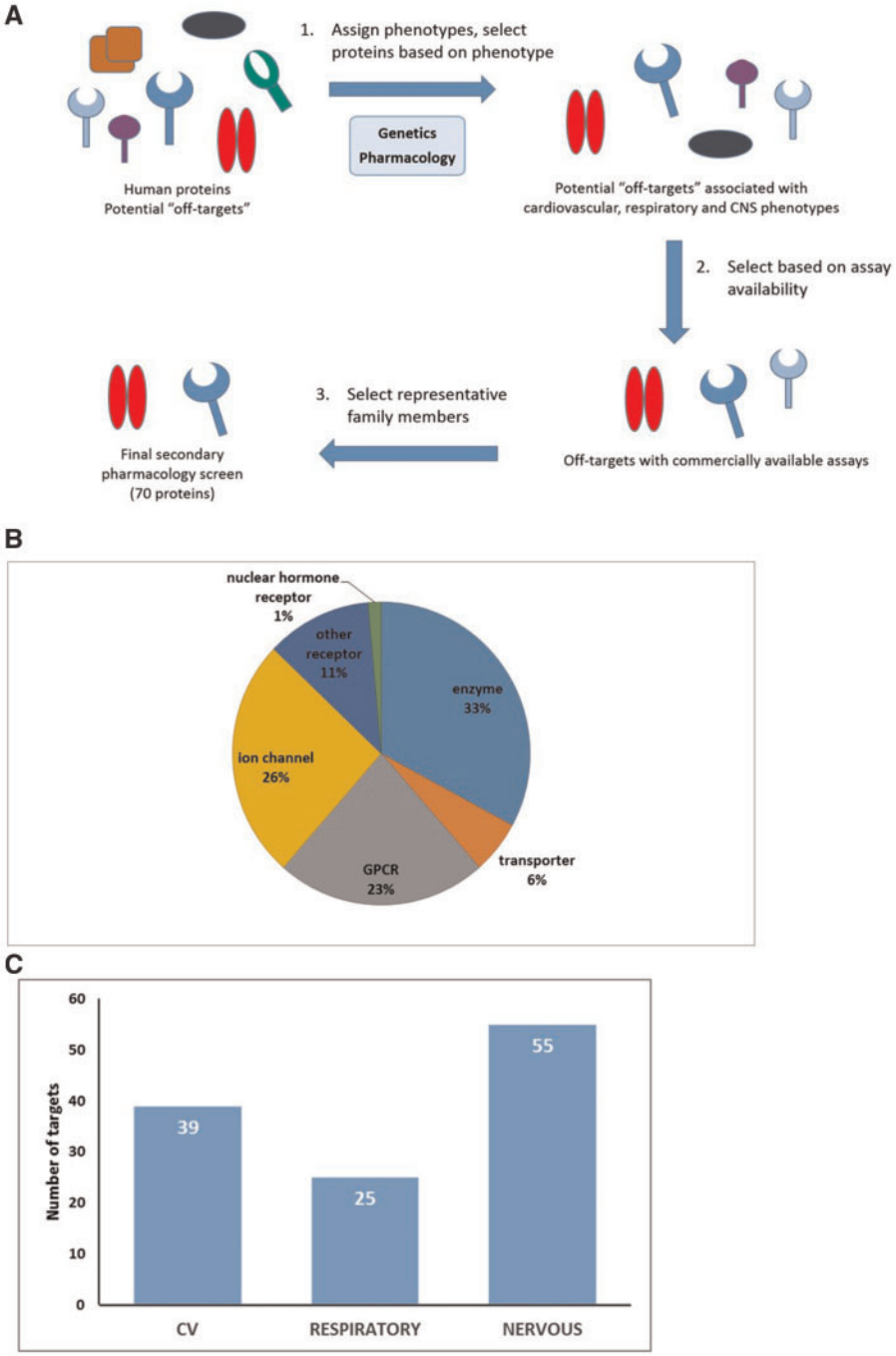

Figure 2. Selection and characteristics of phenotype-focused secondary pharmacology screen. A, Selection of targets to include on secondary pharmacology screen, using phenotypes from human genetics and drug indications. B, Protein class distribution. C, Distribution of key safety phenotypes for the 70 proteins included on our proposed secondary pharmacology panel.

or predicted binding to ADRA2B. This may indicate that, for these drugs, off-target interactions with ADRA2B could contribute to seizure. For the drugs predicted to be inactive against ADRA2B, their observed seizure side effect may result from activities against other off-target proteins. Indeed, 26% of drugs with seizure side effects analyzed interacted with the potassium channels KCNQ2, KCNT1, and/or KCNA2 all of which have genetic evidence for involvement in epilepsy and seizure (Barcia *et al.*, 2012; Singh *et al.*, 1998; Syrbe *et al.*, 2015). Our enrichment analysis does not distinguish between whether a single off-target is the likely culprit for causing a side effect or if, in certain situations, interaction between a drug and multiple off-target proteins is needed to elicit a particular side effect. This would be an interesting avenue for further study.

Second, a routine part of lead optimization is screening drugs against a secondary pharmacology panel to ensure specificity against the proteins most likely to cause serious safety concerns. Achieving complete selectivity for the intended drug target is challenging for small molecule drugs so it is important to avoid unintended interactions with the proteins most likely to have serious effects in humans. Using genetic and pharmacological evidence we designed a panel consisting of proteins

Table 3. Secondary Pharmacology Screen for Use in Small Molecule Drug Development: Selected Proteins Out of 70 Proteins on the Screen and Their Associated Phenotypes

Protein	Gene	Phenotype Pharm	Example Drugs	Phenotype Genetics	Genetics Reference	Category
Potassium channel subunit SUR2	ABCC9	Angina	Nicorandil	Atrial fibrillation	OMIM:614050	CV
Adenosine receptor A2A	ADORA2A	Asthma and bronchodilation	Theophylline, theobromine	None	None	Respiratory
Adrenergic receptor A2B	ADRA2B	Hypertension	Clonidine, prazosin	Epilepsy	OMIM:607876	CV, nervous
Na+/K+-ATPase 1A1	ATP1A1	Heart failure	Digoxin	None	None	CV
Cathepsin F	CTSF	None	None	Ceroid lipofuscinosis	OMIM:615362	Nervous
Dopamine receptor 2	DRD2	Parksinson's disease	Cabergoline	Dystonia, myoclonic	OMIM:159900	Nervous
Coagulation factor II receptor	F2R	Thrombosis	Vorapaxar	None	None	CV
Gamma-aminobutyric acid type A receptor alpha1 subunit	GABRA1	Epilepsy	Estazolam, topiramate	Epilepsy	OMIM:615744	Nervous
Glutamate metabotropic receptor 1	GRM1	None	None	Spinocerebellar ataxia	OMIM:614831	Nervous
Hyperpolarization-activated cyclic nucleotide gated potassium channel 4	HCN4	None	None	Sick sinus syndrome, Brugada syndrome	OMIM:163800 OMIM:613123	CV
Purinergic receptor P2RY12	P2RY12	Thrombosis	Clopidogrel, ticlopidine	Platelet bleeding disorder	OMIM:609821	CV
Nav1.5 channel	SCN5A	Cardiac arrhythmias	Disopyramide, flecainide	Brugada syndrome, ventricular fibrillation, long QT syndrome, sick sinus syndrome, atrial fibrillation	OMIM:601144 OMIM:601154 OMIM:603829 OMIM:603830 OMIM:608567 OMIM:614022	CV
High-affinity choline transporter	SLC5A7	None	None	Neuronopathy	OMIM:158580	Nervous
Thromboxane A synthase 1	TBXAS1	None	None	Ghosal hematodiaphyseal syndrome	OMIM:231095	CV
Transient receptor potential cation channel subfamily M member 4	TRPM4	None	None	Progressive familial heart block	OMIM:604559	CV

associated with cardiovascular, respiratory and nervous system phenotypes, consistent with regulatory guidance. A key consideration in drug safety is the dose at which adverse events occur. The potency of the drug candidate against off-target proteins can be determined by titration of the compound in the secondary pharmacology assay and this information can be incorporated into the model of exposure used preclinically or clinically. During *in vivo* studies, positive hits on the *in vitro* secondary pharmacology screen could be followed up in a number of ways: (1) Appropriate monitoring for the phenotypes expected from off-target hits can be added to preclinical or clinical studies. For example, for hits against Nav1.5/SCN5A we would monitor for cardiac arrhythmia as genetic and pharmacological evidence shows strong associations between Nav1.5 and arrhythmia (see Table 3 and [Chen *et al.*, 1998; Wang *et al.*, 1995]). (2) We can also integrate biomarkers or assays establishing off-target activity into preclinical or clinical studies. For example for inhibitors of the purinergic receptor P2RY12, ADP-induced platelet aggregation assays are commonly used to assess efficacy (Storey *et al.*, 2009), the same assays could be used to assess whether off-target activity against P2RY12 is detectable *in vivo*. As data are generated using this new secondary pharmacology screen of 70 proteins, the relationship between off-target *in vitro* activity and *in vivo* drug side effects will become more clear, and the screen can be further refined.

Our secondary pharmacology screen is limited based on available assays. Given the changing nature of pharmaceutical companies' portfolios, a broader range of potential off-target proteins may need to be considered for future screening. Proteins for which no assays are currently available could be prioritized for assay development based on genetic evidence linking them to adverse phenotypes in humans.

In some cases, the phenotypes resulting from genetic perturbation and drug indications gleaned from pharmacology may be different. For example, ADRA2B is a target of the drug Clonidine used to treat hypertension but gain-of-function mutations in *ADRA2B* are associated with epilepsy (De Fusco *et al.*, 2014). This may be due to a number of factors. Notably, genetic information usually results from perturbing a single gene whereas drugs may have multiple intended targets that contribute to their therapeutic phenotypes as is the case for Clonidine. Other differences between genetic and pharmacological phenotypes could be due to differences in effects on protein function (eg, pharmacological agonism but genetic loss-of-function). There may also be differences in the phenotypes manifesting as a result of the lifelong perturbation caused by genetic variants and those caused by acute pharmacological modulation. Nonetheless, our enrichment analysis results support using both genetic and pharmacological to get a comprehensive view of the potential consequences of drugging a particular protein.

An additional consideration for assessing the impact of drug off-target interactions is the effect of genetic polymorphisms in the gene encoding the off-target protein. Individuals harboring variants in the genes encoding drug off-target proteins may be more susceptible to side effects caused by such drug off-target interactions. One example of this is people with nonpathogenic variants in genes encoding ion channels such as hERG/KCNH2 and KCNE1 who are more susceptible to drug-induced arrhythmias caused by off-target interactions with these channels (Kannankeril *et al.*, 2010; Paulussen *et al.*, 2004).

Beyond the applications examined here, which focus on small molecule therapeutics, there are a number of other uses for this kind of phenotype-centered approach in off-target screening. For example, when developing antibody therapeutics although specificity is not as big an issue as it is for small molecules, a counter screen against a protein related to the drug target is usually employed. Genetics could be used in this situation to select the target family member associated with the most concerning phenotype for counter screening.

We anticipate that integrating phenotypic information from genetics and pharmacology into secondary pharmacology screening will be relevant to current and future drug development programs and will help to reduce safety-related failures and drug side effects.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

ACKNOWLEDGMENTS

We thank Cynthia Afshari and Jing Yuan for comments on the manuscript. All authors are current or former employees of Amgen, Inc.

FUNDING

This study was funded by Amgen, Inc.

REFERENCES

Abifadel, M., Varret, M., Rabes, J. P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., and Erlich, D. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156.

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed. Wiley, New York, NY.

Barcia, G., Fleming, M. R., Deligniere, A., Gazula, V. R., Brown, M. R., Langouet, M., Chen, H., Kronengold, J., Abhyankar, A., Cilio, R., *et al.* (2012). *De novo* gain-of-function KCNT1 channel mutations cause malignant migrating partial seizures of infancy. *Nat. Genet.* **44**, 1255–1259.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Kruger, F. A., Light, Y., Mak, L., McGlinchey, S., *et al.* (2014). The ChEMBL bioactivity database: An update. *Nucleic Acids Res.* **42**(Database issue), D1083–D1090.

Bland, J. M., and Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ* **310**, 170.

Boeynaems, J. M., van Giezen, H., Savi, P., and Herbert, J. M. (2005). P2Y receptor antagonists in thrombosis. *Curr. Opin. Investig. Drugs* **6**, 275–282.

Bowes, J., Brown, A. J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., and Whitebread, S. (2012). Reducing safety-related drug attrition: The use of *in vitro* pharmacological profiling. *Nat. Rev. Drug Discov.* **11**, 909–922.

Cattaneo, M., Zighetti, M. L., Lombardi, R., Martinez, C., Lecchi, A., Conley, P. B., Ware, J., and Ruggeri, Z. M. (2003). Molecular bases of defective signal transduction in the platelet P2Y12 receptor of a patient with congenital bleeding. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1978–1983.

Chao, W. C., Lu, J. F., Wang, J. S., Yang, H. C., Pan, T. A., Chou, S. C., Wang, L. H., and Chou, P. T. (2013). Probing ligand binding to thromboxane synthase. *Biochemistry* **52**, 1113–1121.

Chen, Q., Kirsch, G. E., Zhang, D., Brugada, R., Brugada, J., Brugada, P., Potenza, D., Moya, A., Borggrefe, M., Breithardt, G., *et al.* (1998). Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature* **392**, 293–296.

Curran, M. E., Splawski, I., Timothy, K. W., Vincen, G. M., Green, E. D., and Keating, M. T. (1995). A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* **80**, 795–803.

De Fusco, M., Vago, R., Striano, P., Di Bonaventura, C., Zara, F., Mei, D., Kim, M. S., Muallem, S., Chen, Y., and Wang, Q. (2014). The alpha2B-adrenergic receptor is mutant in cortical myoclonus and epilepsy. *Ann. Neurol.* **75**, 77–87.

Dorsam, R. T., Murugappan, S., Ding, Z., and Kunapuli, S. P. (2003). Clopidogrel: Interactions with the P2Y12 receptor and clinical relevance. *Hematology* **8**, 359–365.

Ference, B. A., Cannon, C. P., Landmesser, U., Luscher, T. F., Catapano, A. L., and Ray, K. K. (2018). Reduction of low density lipoprotein-cholesterol and cardiovascular events with proprotein convertase subtilisin-kexin type 9 (PCSK9) inhibitors and statins: An analysis of FOURIER, SPIRE, and the Cholesterol Treatment Trialists Collaboration. *Eur. Heart J.* **39**, 2540–2545.

Fisher, R. A. (1935). The logic of inductive inference. *J. R. Stat. Soc. Ser. A* **98**, 39–54.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**(Web Server issue), W695–W699.

Hamon, J., Whitebread, S., Techer-Etienne, V., Le Coq, H., Azzaoui, K., and Urban, L. (2009). *In vitro* safety pharmacology profiling: What else beyond hERG? *Future Med. Chem.* **1**, 645–665.

Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schioth, H. B., and Gloriam, D. E. (2017). Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842.

Herbert, J. M., and Savi, P. (2003). P2Y12, a new platelet ADP receptor, target of clopidogrel. *Semin. Vasc. Med.* **3**, 113–122.

Hess, C. N., Low Wang, C. C., and Hiatt, W. R. (2018). PCSK9 inhibitors: Mechanisms of action, metabolic effects, and clinical outcomes. *Ann. Rev. Med.* **69**, 133–145.

ICH. (2001). *Guidance for Industry: S7A Safety Pharmacology Studies for Human Pharmaceuticals*. https://www.ich.org/products/guidelines/safety/article/safety-guidelines.html

ICH. (2005). *Guidance for Industry: S7B Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals*. https://www.ich.org/products/guidelines/safety/article/safety-guidelines.html

Kamb, A., Harper, S., and Stefansson, K. (2013). Human genetics as a foundation for innovative drug development. *Nat. Biotechnol.* **31**, 975–978.

Kannankeril, P., Roden, D. M., and Darbar, D. (2010). Drug-induced long QT syndrome. *Pharmacol. Rev.* **62**, 760–781.

Kim, H. Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor. Dent. Endod.* **42**, 152–155.

Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., *et al.* (2011). DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**(Database issue), D1035–D1041.

Kohler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Ayme, S., Baynam, G., Bello, S. M., Boerkoel, C. F., Boycott, K. M., *et al.* (2017). The human phenotype ontology in 2017. *Nucleic Acids Res.* **45**, D865–D876.

Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**, 711–715.

Kruse, M., Schulze-Bahr, E., Corfield, V., Beckmann, A., Stallmeyer, B., Kurtbay, G., Ohmert, I., Schulze-Bahr, E., Brink, P., and Pongs, O. (2009). Impaired endocytosis of the ion channel TRPM4 is associated with human progressive familial heart block type I. *J. Clin. Invest.* **119**, 2737–2744.

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., and Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **6**, 343.

Kuhn, M., Letunic, I., Jensen, L. J., and Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Res.* **44**, D1075–D1079.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

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

Liu, H., El Zein, L., Kruse, M., Guinamard, R., Beckmann, A., Bozio, A., Kurtbay, G., Megarbane, A., Ohmert, I., Blaysat, G., *et al.* (2010). Gain-of-function mutations in TRPM4 cause autosomal dominant isolated cardiac conduction disease. *Circ. Cardiovasc. Genet.* **3**, 374–385.

Ludbrook, J. (2008). Analysis of 2 × 2 tables of frequencies: Matching test to experimental design. *Int. J. Epidemiol.* **37**, 1430–1435.

Maudsley, S., Martin, B., and Luttrell, L. M. (2005). The origins of diversity and specificity in g protein-coupled receptor signaling. *J. Pharmacol. Exp. Ther.* **314**, 485–494.

McInnes, B. T., Pedersen, T., Pakhomov, S. V. (2009). UMLS-interface and UMLS-similarity: Open source software for measuring paths and semantic similarity. *AMIA Annu. Symp. Proc.* **2009**, 431–435.

Milanesi, R., Baruscotti, M., Gnecchi-Ruscone, T., and DiFrancesco, D. (2006). Familial sinus bradycardia associated with a mutation in the cardiac pacemaker channel. *N. Engl. J. Med.* **354**, 151–157.

Milano, A., Vermeer, A. M., Lodder, E. M., Barc, J., Verkerk, A. O., Postma, A. V., van der Bilt, I. A., Baars, M. J., van Haelst, P. L., Caliskan, K., *et al.* (2014). HCN4 mutations in multiple families with bradycardia and left ventricular noncompaction cardiomyopathy. *J. Am. Coll. Cardiol.* **64**, 745–756.

Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J., *et al.* (2015). The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860.

Nof, E., Luria, D., Brass, D., Marek, D., Lahat, H., Reznik-Wolf, H., Pras, E., Dascal, N., Eldar, M., and Glikson, M. (2007). Point mutation in the HCN4 cardiac ion channel pore affecting synthesis, trafficking, and functional expression is associated with familial asymptomatic sinus bradycardia. *Circulation* **116**, 463–470.

Nurden, P., Savi, P., Heilmann, E., Bihour, C., Herbert, J. M., Maffrand, J. P., and Nurden, A. (1995). An inherited bleeding disorder linked to a defective interaction between ADP and its receptor on platelets. Its influence on glycoprotein IIb-IIIa complex function. *J. Clin. Investig.* **95**, 1612–1622.

Papoian, T., Chiu, H. J., Elayan, I., Jagadeesh, G., Khan, I., Laniyonu, A. A., Li, C. X., Saulnier, M., Simpson, N., and Yang, B. (2015). Secondary pharmacology data to assess potential

off-target activity of new drugs: A regulatory perspective. *Nat. Rev. Drug Discov.* **14**, 294.

Paulussen, A. D., Gilissen, R. A., Armstrong, M., Doevendans, P. A., Verhasselt, P., Smeets, H. J., Schulze-Bahr, E., Haverkamp, W., Breithardt, G., Cohen, N., *et al.* (2004). Genetic variations of KCNQ1, KCNH2, SCN5A, KCNE1, and KCNE2 in drug-induced long QT syndrome patients. *J. Mol. Med.* **82**, 182–188.

Plenge, R. M., Scolnick, E. M., and Altshuler, D. (2013). Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581 (Review Article).

Ponomarenko, E. A., Poverennaya, E. V., Ilgisonis, E. V., Pyatnitskiy, M. A., Kopylov, A. T., Zgoda, V. G., Lisitsa, A. V., and Archakov, A. I. (2016). The size of the human proteome: The width and depth. *Int. J. Anal. Chem.* **2016**, 1.

Rask-Andersen, M., Masuram, S., and Schioth, H. B. (2014). The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. *Annu. Rev. Pharmacol. Toxicol.* **54**, 9–26.

Roberts, R. A., Kavanagh, S. L., Mellor, H. R., Pollard, C. E., Robinson, S., and Platz, S. J. (2014). Reducing attrition in drug development: Smart loading preclinical safety assessment. *Drug Discov. Today* **19**, 341–347.

Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T. I., *et al.* (2017). A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34.

Schulze-Bahr, E., Neu, A., Friederich, P., Kaupp, U. B., Breithardt, G., Pongs, O., and Isbrandt, D. (2003). Pacemaker channel dysfunction in a patient with sinus node disease. *J. Clin. Invest.* **111**, 1537–1545.

Schweizer, P. A., Schroter, J., Greiner, S., Haas, J., Yampolsky, P., Mereles, D., Buss, S. J., Seyler, C., Bruehl, C., Draguhn, A., *et al.* (2014). The symptom complex of familial sinus node dysfunction and myocardial noncompaction is associated with mutations in the HCN4 channel. *J. Am. Coll. Cardiol.* **64**, 757–767.

Siintola, E., Partanen, S., Stromme, P., Haapanen, A., Haltia, M., Maehlen, J., Lehesjoki, A. E., and Tyynela, J. (2006). Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. *Brain* **129**, 1438–1445.

Singh, N. A., Charlier, C., Stauffer, D., DuPont, B. R., Leach, R. J., Melis, R., Ronen, G. M., Bjerre, I., Quattlebaum, T., Murphy, J. V., *et al.* (1998). A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. *Nat. Genet.* **18**, 25–29.

Stallmeyer, B., Zumhagen, S., Denjoy, I., Duthoit, G., Hebert, J. L., Ferrer, X., Maugenre, S., Schmitz, W., Kirchhefer, U., Schulze-Bahr, E., *et al.* (2012). Mutational spectrum in the Ca(2+)-activated cation channel gene TRPM4 in patients with cardiac conductance disturbances. *Hum. Mutat.* **33**, 109–117.

Steinfeld, R., Reinhardt, K., Schreiber, K., Hillebrand, M., Kraetzner, R., Bruck, W., Saftig, P., and Gartner, J. (2006). Cathepsin D deficiency is associated with a human neurodegenerative disorder. *Am. J. Hum. Genet.* **78**, 988–998.

Storey, R. F., Melissa Thornton, S., Lawrance, R., Husted, S., Wickens, M., Emanuelsson, H., Cannon, C. P., Heptinstall, S., and Armstrong, M. (2009). Ticagrelor yields consistent dose-dependent inhibition of ADP-induced platelet aggregation in patients with atherosclerotic disease regardless of genotypic variations in P2RY12, P2RY1, and ITGB3. *Platelets* **20**, 341–348.

Syrbe, S., Hedrich, U. B. S., Riesch, E., Djemie, T., Muller, S., Moller, R. S., Maher, B., Hernandez-Hernandez, L., Synofzik, M., Caglayan, H. S., *et al.* (2015). *De novo* loss- or gain-of-function mutations in KCNA2 cause epileptic encephalopathy. *Nat. Genet.* **47**, 393–399.

Tatonetti, N. P., Ye, P. P., Daneshjou, R., and Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Sci. Transl. Med.* **4**, 125ra31.

Ueda, K., Hirano, Y., Higashiuesato, Y., Aizawa, Y., Hayashi, T., Inagaki, N., Tana, T., Ohya, Y., Takishita, S., Muratani, H., *et al.* (2009). Role of HCN4 channel in preventing ventricular arrhythmia. *J. Hum. Genet.* **54**, 115–121.

Varma, S., Bishara, D., Besag, F. M., and Taylor, D. (2011). Clozapine-related EEG changes and seizures: Dose and plasma-level relationships. *Ther. Adv. Psychopharmacol.* **1**, 47–66.

Wang, Q., Shen, J., Splawski, I., Atkinson, D., Li, Z., Robinson, J. L., Moss, A. J., Towbin, J. A., and Keating, M. T. (1995). SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* **80**, 805–811.

Warr, W. A. (2012). Scientific workflow systems: Pipeline Pilot and KNIME. *J. Comput. Aided Mol. Des.* **26**, 801–804.

Whitebread, S., Dumotier, B., Armstrong, D., Fekete, A., Chen, S., Hartmann, A., Muller, P. Y., and Urban, L. (2016). Secondary pharmacology: Screening and interpretation of off-target activities—Focus on translation. *Drug Discov. Today* **21**, 1232–1242.

Williams, A. M., and Park, S. H. (2015). Seizure associated with clozapine: Incidence, etiology, and management. *CNS Drugs* **29**, 101–111.

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008). DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**(Database issue), D901–D906.

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: A comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* **34**(Database issue), D668–D672.

Zuhl, A. M., Nolan, C. E., Brodney, M. A., Niessen, S., Atchison, K., Houle, C., Karanian, D. A., Ambroise, C., Brulet, J. W., Beck, E. M., *et al.* (2016). Chemoproteomic profiling reveals that cathepsin D off-target activity drives ocular toxicity of beta-secretase inhibitors. *Nat. Commun.* **7**, 13042.