Predictive functional classification of pharmacogenetic variants with off-target effects

Hannah McConnell Equal first author, [1], **T. Daniel Andrews** Corresp., Equal first author, [1], **Matt A Field** Corresp. [1, 2]

[1] John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia

[2] Australian Institute of Tropical Health and Medicine, Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Smithfield, Australia

Corresponding Authors: T. Daniel Andrews, Matt A Field
Email address: dan.andrews@anu.edu.au, matt.field@anu.edu.au

Background. Pharmacogenetic variation becomes important to drug responses through diverse and highly complex mechanisms. To help predict the functional impact for the increasing number of newly discovered pharmacogenetic variants, software is employed that largely relies on sequence conservation across species as the primary discriminator. However, sequence conservation in a species is largely maintained by evolutionary purifying selection which operates on a much longer time scale than recently developed pharmaceutical drugs. Further, drugs often cause off-target effects, pharmacological activity that affect cellular processes unrelated to the primary cellular function, which are unlikely to be subject to purifying selection and violate the assumptions of the predicative software.

Methods. Here we exhaustively assess the effectiveness of eleven missense mutation functional inference tools on all known pharmacogenetic missense variants contained in the Pharmacogenomics Knowledge base (PharmGKB) repository. We categorize PharmGKB entries into sub-classes and compare results across a variety of variant annotations.

Results. Overall functional inference tools perform poorly on the complete set of PharmGKB variants with large numbers of variants incorrectly classified as benign. However, we find substantial differences amongst PharmGKB variant sub-classes, particularly in variants known to cause type B adverse drug reactions that are largely unrelated to the main pharmacological action of the drug. Variants causing off-target effects are unlikely to have been subject to purifying selection and as such were most often incorrectly classified as benign. These results highlight the importance of understanding the underlying mechanism of pharmacogenetic variants and how variants causing off-target effects will ultimately require new predicative algorithms. We describe how to identify variants causing off-target effects within PharmGKB in order to generate a training set of variants that is needed to develop new algorithms specifically for this class of variant. Development of such tools will lead to more accurate functional predictions and pave the way for the increased wide-spread adoption of pharmacogenetics in clinical practice.

Predictive functional classification of pharmacogenetic variants with off-target effects

Hannah McConnell[1], and T. Daniel Andrews[1,#], Matt A. Field[1,2,#]

[1] Genome Informatics Laboratory, Department of Immunology and Infectious Disease, JCSMR, Australian National University, Canberra, ACT, Australia, 0200
[2] Australian Institute of Tropical Health and Medicine and the Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Smithfield, QLD, Australia, 4878
[#] Contributed equally to this work

Corresponding Author:
Dan Andrews
Garran Road, Canberra, ACT, Australia
dan.andrews@anu.edu.au

Abstract

Background. Pharmacogenetic variation becomes important to drug responses through diverse and highly complex mechanisms. To help predict the functional impact for the increasing number of newly discovered pharmacogenetic variants, software is employed that largely relies on sequence conservation across species as the primary discriminator. However, sequence conservation in a species is largely maintained by evolutionary purifying selection which operates on a much longer time scale than recently developed pharmaceutical drugs. Further, drugs often cause off-target effects, pharmacological activity that affect cellular processes unrelated to the primary cellular function, which are unlikely to be subject to purifying selection and violate the assumptions of the predicative software.

Methods. Here we exhaustively assess the effectiveness of eleven missense mutation functional inference tools on all known pharmacogenetic missense variants contained in the Pharmacogenomics Knowledge base (PharmGKB) repository. We categorize PharmGKB entries into sub-classes and compare results across a variety of variant annotations.

Results. Overall functional inference tools perform poorly on the complete set of PharmGKB variants with large numbers of variants incorrectly classified as benign. However, we find substantial differences amongst PharmGKB variant sub-classes, particularly in variants known to cause type B adverse drug reactions that are largely unrelated to the main pharmacological action of the drug. Variants causing off-target effects are unlikely to have been subject to purifying selection and as such were most often incorrectly classified as benign. These results highlight the importance of understanding the underlying mechanism of pharmacogenetic variants and how variants causing off-target effects will ultimately require new predicative algorithms. We describe how to identify variants causing off-target effects within PharmGKB in order to

generate a training set of variants that is needed to develop new algorithms specifically for this class of variant. Development of such tools will lead to more accurate functional predictions and pave the way for the increased wide-spread adoption of pharmacogenetics in clinical practice.

Introduction

Individual variability in drug response poses a large challenge to the efficient and effective treatment of patients (Meyer 2000; Pirmohamed 2001). Many oncology drugs have be shown to be ineffective in a subset of individuals, meaning that often multiple drugs must be tried before an effective treatment is found (Dancey et al. 2012). For example, it is not understood why statins (a class of drugs commonly prescribed for cardiovascular disease) behave differently in individuals (Silva et al. 2006), and can even cause a very severe toxic reaction in a small number of patients (Gabb et al. 2013). It is estimated that 15-30% of this variability in drug response is due to genetic factors (Eichelbaum et al. 2006; Pang et al. 2009) however the precise role of such genetic factors is often little understood . There is a growing number of databases that aggregate, curate and annotate the increasing body of identified genetic factors (Sim et al. 2011). The Pharmacogenomic Knowledgebase (PharmGKB) (Whirl-Carrillo et al. 2012) is the largest, open database of pharmacogenetic data, and at time of publication, includes information on nearly 150 pathways and over 23,000 individual variant annotations. Variants within PharmGKB are also annotated with effect types (dosage, efficacy, toxicity) and the level of confidence (categories 1-4) of the pharmacogenetic association, with category 1 being the highest. The pharmacogenetic variants included in PharmGKB cover a wide range of mutation types, from missense and synonymous single nucleotide variants (SNVs) to non-coding, intergenic and copy number variants.

Classifying newly predicted pharmacogenetic variants is a challenge as the number of actual variants is several orders of magnitude larger than the cases that are presently backed by experimentally-verified functional data. For missense mutations, this interpretation gap is presently filled by mutation function inference tools, such as PolyPhen2 (Adzhubei et al. 2010), CADD (Kircher et al. 2014) and SIFT (Sim et al. 2012). These inferential tools integrate sequence conservation and, often, structural information to predict whether alterations to the amino-acid sequence are likely to alter the function of a protein (Khan & Vihinen 2010). In almost all cases, if a variant or mutation lies in a highly conserved region in a multi-species alignment of orthologous gene sequences, the variant will very likely be considered deleterious or damaging. Conversely, should the variant be broadly similar to existing sequence variation in this same alignment, the variant will be considered benign or functionally homologous. While some tools do consider structural information pertaining to the missense change, most of these tools also incorporate sequence conservation data in their algorithms. Hence, the strength of purifying selection to remove non-functional or poorly functional variants from the population is the strongest evidence presently used to classify a variant as either benign or deleterious.

The reliance of such tools on sequence conservation is critical when considering pharmacogenetic variation as in many instances such variants will not have been subject to purifying selection on an evolutionary timescale. A recent study assessed the effectiveness of eight tools on variants in the RYR1 gene, which is linked to pharmacogenetic disorder malignant hyperthermia (MH) (Schiemann & Stowell 2016). They compared MH-causative variants and common variants and found none of the prediction programmes could classify all variants correctly as either 'damaging' or as 'benign' respectively (84% - 100% range for sensitivity and 25% - 83% range of specificity). A broader study appraised mutation functional inference methods across a variety of pharmacogenetic missense variants and also found them to perform poorly with the effect attributed to the ill-suited training sets used to build the models on which the algorithms rely (Zhou et al. 2018). Such studies led us to examine pharmacogenetic variants in order to identify subclasses that were unlikely to be subject to purifying selection such as variants causing adverse drug reactions (ADRs).

ADRs are broadly classified based on a general mechanistic distinction (Patton & Borshoff 2018). Type A reactions are common and their effects are predictable and mostly dose-dependent. Type A reactions relate to interactions of a drug with its primary drug target. Conversely, type B reactions are less common and are mostly unrelated to the main pharmacological action of the drug. Type B reactions, sometimes also called idiosyncratic drug reaction (Uetrecht & Naisbitt 2013), can be dose-dependent or dose-independent, may be immunologically-mediated and/or ~~may~~ involve off-target drug interactions (Patton & Borshoff 2018). Immune-mediated type B reactions involve the drug inducing a specific immune response, such as the development of a skin rash commonly caused by administration of penicillin (Weiss & Adkinson 1988). Off-target drug effects can also occur without an immunological component, such as the interactions of anaesthetics with the ryanodine receptor 1 (RYR1) protein causing malignant hyperthermia (Robinson et al. 2006).

We extracted all PharmGKB variants causing missense mutations and assessed the effectiveness of eleven functional inference tools. PharmGKB contains substantial numbers of variants, across all variant evidence levels, that are computationally predicted to be benign. We independently analyzed variants causing type A and type B reactions to determine whether the functional inferences for these variant sub-classes differ. We find that most PharmGKB entries incorrectly classified as benign are overwhelmingly off-target and idiosyncratic variants.

Materials & Methods

Pharmacogenetic Variant Datasets

A set of pharmacogenetic variants with DBSNP reference cluster identifiers (RS) (Sherry et al. 2001) were obtained from PharmGKB (Whirl-Carrillo et al. 2012) and custom overlap code used to combine variant annotations (Field et al. 2015). Variants within PharmGKB are classified by gene, type of effect, level of evidence, specific drug, chemical, disease and phenotype. Variants

were further annotated with using Variant Effect Predictor (VEP) (McLaren et al. 2010).

Classification of Off-Target Pharmacogenetic Variants

A simple classification scheme was devised to identify and confirm likely off-target variants. All clinical variants with evidence category of 1A, 1B, 2A and 2B from the PharmGKB database (Whirl-Carrillo et al. 2012) were first filtered for PharmGKB annotations of effect type 'Toxicity/ADR' for any particular chemical and/or drug. Variants were removed if they also had an additional effect type (other than 'Toxicity/ADR') for the same drug. Next, variants were removed if they were present in ADME process genes (categorized as such in the PharmaADME database; www.pharmaadme.org) or were annotated with Gene Ontology (24) categories of 'xenobiotic metabolism process' or with 'transporter'. Of these remaining variants, synonymous and non-coding variants were next excluded, leaving just missense variants. With this filtered list, the cited literature for each variant was appraised to discern whether the variant resulted in an off-target effect. Variants were retained where either a molecular mechanism for the adverse drug effect was known or when a protein or cellular system related to the intended drug effect was involved. The classification scheme is summarized in **Table 1**.

Functional Effect Prediction

The predicted functional effect of mutations was predicted with SIFT (Sim et al. 2012), PolyPhen2 (Adzhubei et al. 2010), CADD (Kircher et al. 2014), DANN (Quang et al. 2015), FATHMM (Shihab et al. 2013), GERP++ (Davydov et al. 2010), MutPred (Li et al. 2009), Mutation Assessor (Reva et al. 2011), Mutation Taster (Schwarz et al. 2014), REVEL (Ioannidis et al. 2016) and PhastCons (Siepel et al. 2005), relative to EnsEMBL canonical transcripts annotated with VEP (McLaren et al. 2010) .

Results

Distributions of pharmacogenetic variant functional inferences

Functional inference scores were obtained for 561 missense single nucleotide variants (SNVs) with DBSNP (RS) cluster identifiers from the PharmGKB database using eleven different prediction tools (SIFT, PolyPhen2, CADD, DANN, FATHMM, GERP++, MutPred, Mutation Assessor, Mutation Taster, REVEL and PhastCons) (Supplementary Table S1). The distributions of scores from six of these tools (CADD, PolyPhen2, SIFT, MutationAssessor, MutPred and REVEL) are plotted (**Figure 1**). The predictions calculated for these functional variants ranged widely from benign to deleterious. For comparison to expected background levels, we also selected a random set of 2155 human missense SNVs with assigned RS cluster identifiers. Overall, the pattern of variant inferences is broadly similar across all tools and no single tool produces inferences that are qualitatively different (note that the scores calculated by SIFT run in the opposite direction to the other tools). These tools represent a broad range of methodologies available for mutation functional prediction and the categories of information used by each tool are annotated in Figure 1 as seq (sequence conservation), struct (protein structural metrics), and ens (ensemble tool that integrates individual tools). The results demonstrate how many of the highest confidence PharmGKB variants annotated as functionally important are predicted to be benign. Of the 119

highest confidence category 1 variants, the majority of predictions to be deleterious by PolyPhen2 (median score 0.996), however 6 variants were classified benign. The 183 variants in category 2 had a much broader range of predicted functional effects with 33 variants predicted as benign and an overall median score of 0.138, even less than the median score of 0.245 for the randomly selected variants. Similarly, the distribution of functional effect predictions in category 3 was strongly skewed towards benign variants (PolyPhen2 median score 0.012) and category 4 had a distribution very similar to the random variant set (PolyPhen2 median score 0.319). This pattern of inferred functional effects for PharmGKB variants was very similar between tools.

Classification of pharmacogenetic variation to detect off-target effects

A prior study (Zhou et al. 2018) demonstrated that functional prediction tools do not perform well across all pharmacogenetic variation. While overall our results support this conclusion, we hypothesized that the majority of pharmacogenetic variants predicted to be benign were type B variants causing off-target effects. To investigate the possibility that type B pharmacogenetic variants are predominantly predicted to be benign, we devised a simple classification system of PharmGKB variants (described in *Materials and Methods* and Table 1) that would be selective for those of type B. From a possible 140 high confidence PharmGKB variants (categories 1A, 1B, 2A and 2B) this classification system identified 24 variants as potentially type B or off-target variants (**Table 2**). Half of these (12) were missense variants and the remainder being at non-coding or synonymous sites. Of the 12 missense variants, six were predicted as benign by PolyPhen2 or had a CADD Phred quality score of less than 20 (a conservative threshold for deleterious variants). All variants, including non-coding variants were scored with CADD, of these only five had a Phred quality score greater than 20. Hence, even within these high confidence pharmacogenetic variants, only a minority are predicted to be functionally important by existing tools.

For each variant predicted to cause off-target effects in Table 2, a manual literature survey was conducted to confirm our prediction. Of the 24 variants, the mechanism of action was known in 15 cases and every one of these had an off-target (or type B) effect, including four cases where the variant was not a missense variant. This included all five high-confidence category 1 pharmacogenetic variants. Treating this result as strong validation of the classification system we applied this classification system to lower evidence variants from PharmGKB categories 3 and 4.

Given that category 3 variants were strongly skewed towards being predicted as benign (Figure 1), we asked the question whether this category of variants is overwhelmingly off-target, type B pharmacogenetic variants. Where the mechanism of pharmacogenomic effect was known, 12 out of a total 17 variants we manually appraised caused off-target effects (Supplementary Table S2). This result provides a measure of the efficacy 70% (12/17) of this method to detect off-target pharmacogenetic variants purely from data analysis without associated mechanistic studies.

Further precision may be obtained through excluding potential off-target variants with a deleterious functional prediction.

Discussion

In this work we have appraised whether pharmacogenetic variants causing off-target effects are consistently predicted to be less deleterious than other functionally-important variants. The current generation of missense mutation inference tools make distinctions on the basis of sequence conservation with a variant occurring in a conserved region almost always classified as deleterious. Protein sequence conserved between species implies that the function of the encoded protein was intolerant to mutations with any changes removed by purifying selection over evolutionary timescales. Drugs prescribed to human patients however is a very recent occurrence in an evolutionary context and unless a pharmacogenetic variant is related to the primary function of the gene, no information is present in the sequence record from which to detect functional importance. Variants causing off-target and/or idiosyncratic reactions in particular are too recent a selective condition for evolutionary processes to impact the sequence content, even within the human genome. This means these variants are mostly invisible to current inferential tools and will require new methods to estimate functional impact. We confirm this by extracting all pharmacogenetic variants with an off-target mechanism in PharmGKB and show that almost all such variants are incorrectly classified as benign.

Functional predictions of the likely impact of a given variant are predominately driven based on sequence conservation levels despite algorithms incorporating other types of evidence. Unlike pathogenic variants causing genetic diseases however, the action of drugs is a recent event on an evolutionary time scale and further has only applied to a very limited range of species. Many pharmacogenes contain variants which generate type A ADRs with the drug most often just another xenobiotic compound which the target-protein acts upon. Such variants will likely have been subject to purifying selection and should be correctly classified as deleterious by the current generation of functional inference tools. However, variants that cause a type B or off-target effect are much less likely to be subject to the same selection pressures meaning such variants will likely be incorrectly classified as benign due to the lack of observed sequence conservation. Indeed, we show most off-target pharmacogenetic variants of this type are predicted to be functionally unimportant and will be missed using current tools. New methods that can detect pharmacogenetic variation that has not been subjected to purifying selection are urgently needed to capture this important type of pharmacogenetic variation.

Given the reliance of variant functional inference tools on sequence conservation, it is unsurprising that the tools perform poorly on pharmacogenetic variation that ~~have~~ not ~~been~~ subject~~ed~~ to purifying selection. Without using sequence conservation information as a primary discriminator, what methods and datasets are available to differentiate between truly benign and functionally important variation causing off-target effects? Tools that incorporate protein

structural information would be expected to work better on such variants, yet our investigation showed little difference between tools which use structural features and those which do not. In hopes of finding new ways to predict damaging missense mutations, researchers are increasingly applying machine learning techniques to improve functional prediction algorithms particularly for identifying disease causing variants (Kalinin et al. 2018). However, for pharmacogenetic variants options are limited. A recent study reported improved sensitivity and specificity using a functionality prediction framework optimized for pharmacogenetic variants however no code has yet been released to independently assess this claim (Zhou et al. 2018). Regardless of the eventual outcome, the ability to accurately predict pharmacogenetic variants causing off-target effects is critical for the increased adoption of pharmacogenetics in clinical practice.

Conclusions

Pharmacogenetic missense variants represent a complex set of genetic factors with highly diverse functional mechanisms that influence drug efficacy. Functional predictions of the likely impact of a given missense variant are driven by measures of sequence conservation over deep evolutionary timescales including mammals, invertebrates and even yeast. However, unlike pathogenic variants identified in rare genetic diseases, the drug's actions have not necessarily been subject to evolutionary selective forces and purifying selection meaning the assumptions of the tools are violated. Our analysis confirms that in many cases, the assumptions of functional inference tools are invalid, particularly for variants causing off-target, type B adverse drug reactions. We describe a simple method to identify such variants and note that the majority are predicted to be benign and functionally unimportant. Generating a subset of such variants will enable the development of urgently needed new methods that can detect pharmacogenetic variation that has not been subjected to purifying selection.

Acknowledgements

We thank the National Computational Infrastructure (Australia) for continued access to significant computation resources and technical expertise.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, and Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248-249.

Dancey JE, Bedard PL, Onetto N, and Hudson TJ. 2012. The genetic basis for cancer treatment decisions. *Cell* 148:409-420.

278 Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, and Batzoglou S. 2010. Identifying a
279 high fraction of the human genome to be under selective constraint using GERP++.
280 *PLoS Comput Biol* 6:e1001025.
281 Eichelbaum M, Ingelman-Sundberg M, and Evans WE. 2006. Pharmacogenomics and
282 individualized drug therapy. *Annu Rev Med* 57:119-137.
283 Field MA, Cho V, Andrews TD, and Goodnow CC. 2015. Reliably Detecting Clinically Important
284 Variants Requires Both Combined Variant Calls and Optimized Filtering Strategies.
285 *PLoS One* 10:e0143199.
286 Gabb GM, Vitry A, Condon J, Limaye V, and Alhami G. 2013. Serious statin-associated
287 myotoxicity and rhabdomyolysis in Aboriginal and Torres Strait Islanders: a case series.
288 *Intern Med J* 43:987-992.
289 Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q,
290 Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J,
291 Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin
292 G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD,
293 Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA,
294 Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, and Sieh W. 2016.
295 REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense
296 Variants. *Am J Hum Genet* 99:877-885.
297 Kalinin AA, Higgins GA, Reamaroon N, Soroushmehr S, Allyn-Feuer A, Dinov ID, Najarian K,
298 and Athey BD. 2018. Deep learning in pharmacogenomics: from gene regulation to
299 patient stratification. *Pharmacogenomics* 19:629-650.
300 Khan S, and Vihinen M. 2010. Performance of protein stability predictors. *Hum Mutat* 31:675-
301 684.
302 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, and Shendure J. 2014. A general
303 framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*
304 46:310-315.
305 Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, and Radivojac P.
306 2009. Automated inference of molecular mechanisms of disease from amino acid
307 substitutions. *Bioinformatics* 25:2744-2750.
308 McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, and Cunningham F. 2010. Deriving the
309 consequences of genomic variants with the Ensembl API and SNP Effect Predictor.
310 *Bioinformatics* 26:2069-2070.
311 Meyer UA. 2000. Pharmacogenetics and adverse drug reactions. *Lancet* 356:1667-1671.
312 Pang GS, Wang J, Wang Z, and Lee CG. 2009. Predicting potentially functional SNPs in drug-
313 response genes. *Pharmacogenomics* 10:639-653.
314 Patton K, and Borshoff DC. 2018. Adverse drug reactions. *Anaesthesia* 73 Suppl 1:76-84.
315 Pirmohamed M. 2001. Pharmacogenetics and pharmacogenomics. *Br J Clin Pharmacol* 52:345-
316 347.
317 Quang D, Chen Y, and Xie X. 2015. DANN: a deep learning approach for annotating the
318 pathogenicity of genetic variants. *Bioinformatics* 31:761-763.
319 Reva B, Antipin Y, and Sander C. 2011. Predicting the functional impact of protein mutations:
320 application to cancer genomics. *Nucleic Acids Res* 39:e118.
321 Robinson R, Carpenter D, Shaw MA, Halsall J, and Hopkins P. 2006. Mutations in RYR1 in
322 malignant hyperthermia and central core disease. *Hum Mutat* 27:977-989.
323 Schiemann AH, and Stowell KM. 2016. Comparison of pathogenicity prediction tools on
324 missense variants in RYR1 and CACNA1S associated with malignant hyperthermia. *Br J*
325 *Anaesth* 117:124-128.
326 Schwarz JM, Cooper DN, Schuelke M, and Seelow D. 2014. MutationTaster2: mutation
327 prediction for the deep-sequencing age. *Nat Methods* 11:361-362.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, and Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29:308-311.
Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, and Gaunt TR. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 34:57-65.
Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, and Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15:1034-1050.
Silva MA, Swanson AC, Gandhi PJ, and Tataronis GR. 2006. Statin-related adverse events: a meta-analysis. *Clin Ther* 28:26-35.
Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, and Ng PC. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 40:W452-457.
Sim SC, Altman RB, and Ingelman-Sundberg M. 2011. Databases in the area of pharmacogenetics. *Hum Mutat* 32:526-531.
Uetrecht J, and Naisbitt DJ. 2013. Idiosyncratic adverse drug reactions: current concepts. *Pharmacol Rev* 65:779-808.
Weiss ME, and Adkinson NF. 1988. Immediate hypersensitivity reactions to penicillin and related antibiotics. *Clin Allergy* 18:515-540.
Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, and Klein TE. 2012. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 92:414-417.
Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, and Lauschke VM. 2018. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J*.

Figure 1 – Distribution of functional effect scores of PharmGKB variants predicted by six mutation effect inference tools. Boxplots shown are of **a)** CADD Phred score, **b)** PolyPhen2 score, **c)** SIFT score, **d)** Mutation Assessor score, **e)** MutPred score and **f)** REVEL score. Scores are plotted for each tool in variant confidence categories (from 1 (highest) to 4 (lowest)) assigned by the PharmGKB annotation. Each tool is annotated with the information types it employs to make predictions – **Seq**: sequence conservation, **Struct**: protein structural metrics, **Ens**: an ensemble tool that integrates results of several individual tools.

Figure 1

Distribution of functional effect scores of PharmGKB variants predicted by six mutation effect inference tool.

Boxplots shown are of **a)** CADD Phred score, **b)** PolyPhen2 score, **c)** SIFT score, **d)** Mutation Assessor score, **e)** MutPred score and **f)** REVEL score. Scores are plotted for each tool in variant confidence categories (from 1 (highest) to 4 (lowest)) assigned by the PharmGKB annotation. Each tool is annotated with the information types it employs to make predictions – **Seq**: sequence conservation, **Struct**: protein structural metrics, **Ens**: an ensemble tool that integrates results of several individual tools.

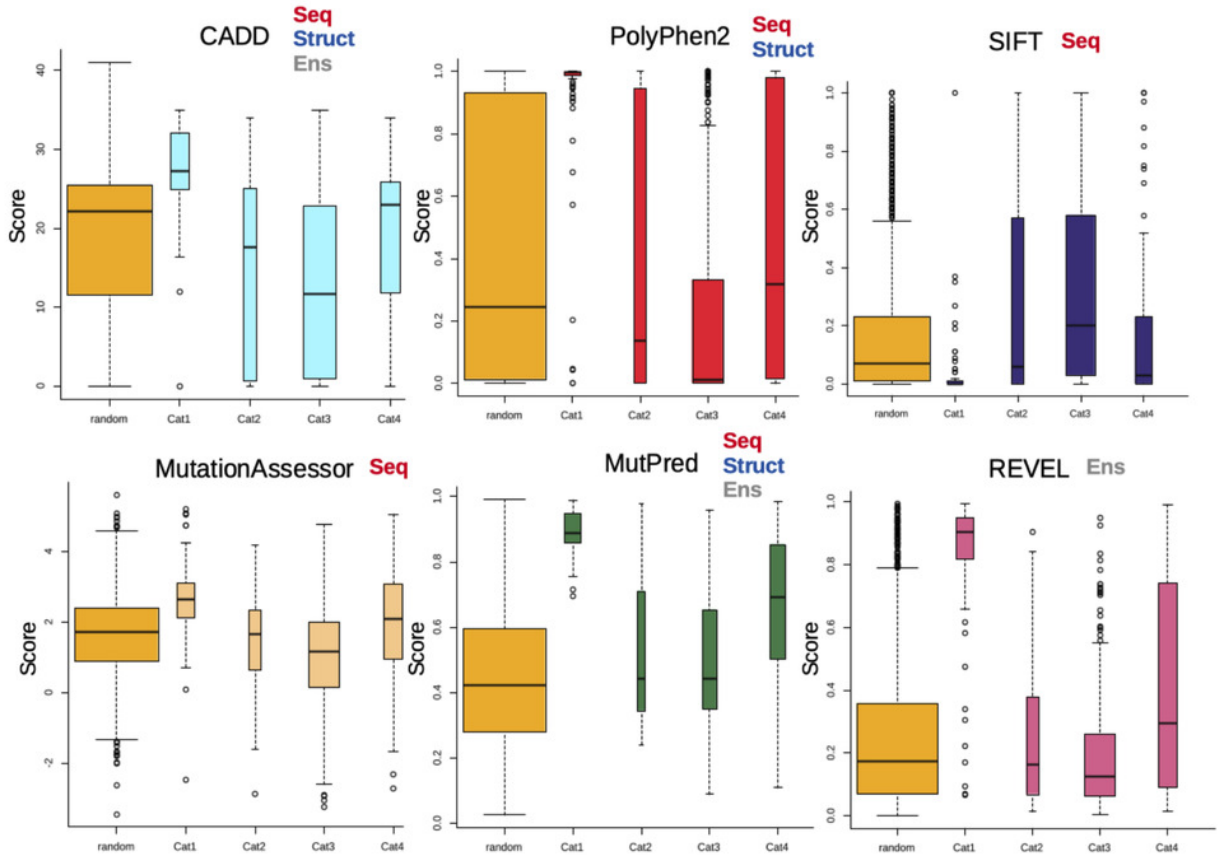

Table 1(on next page)

Classification criteria used to identify off-target pharmacogenetic variants from the PharmGKB databas.

Step	Filter
1	Variant only has type:toxicity/ADR
2	For drug and gene pairs, there are no other variants with effect type other than ‘Toxicity/ADR’
3	Gene containing variant is NOT an ADME process gene OR annotated in GO with 'xenobiotic metab process' OR 'transporter')
4	Exclude synonymous and non-coding variants
5	Literature review indicates a known mechanism, which is an off-target drug/protein interaction

Table 1 – Classification criteria used to identify off-target pharmacogenetic variants from the PharmGKB database.

Table 2(on next page)

High confidence PharmGKB variants sorted with criteria to identify possible off-target effects.

Variant	Gene	PharmGKB level of evidence	Example PharmGKB Chemicals	PharmGKB Phenotypes	CADD Phred score	PolyPhen category (score)	Mechanism Known?	Category	Variant type
rs1050828	G6PD	1B	chlorproguanil, dapsone	Malaria	-	Benign (0.202)	Yes (25)	Off-target	Missense
rs267606617	MT-RNR1	1B	aminoglycoside antibacterials	Ototoxicity	-	Benign (0)	Yes (26)	Off-target	Non-coding
rs111888148	RYR1	1A	desflurane, enflurane, halothane	Malignant Hyperthermia	27.5	Probably Damaging (0.998)	Limited (27)	Off-target	Missense
rs1800559	CACNA1S	1A	desflurane, enflurane, halothane	Malignant Hyperthermia	24.4	Probably Damaging (0.998)	Yes(28)	Off-target	Missense
rs1799971	OPRM1	2B	ethanol	Alcoholism	24	Possibly Damaging (0.775)	Yes (29)	Off-target	Missense
rs1127354	ITPA	2B	peginterferon alpha-2b, ribavirin	Hepatitis C	22.7	-	Yes (30)	Off-target	Missense
rs1933437	FLT3	2B	sunitinib	Carcinoma, Leukopenia	21.4	Possibly Damaging (0.69)			Missense
rs1801394	MTRR	2B	methotrexate	Lymphoma	20.9	Probably Damaging (0.99)	Proposed (31)	Off-target	Missense
rs2228001	XPC	1B	cisplatin	Neoplasms	18.95	Benign (0)	Yes (32)	Off-target	Missense
rs6025	F5	2A	hormonal contraceptives	Thrombosis	18.92	Benign (0)	Yes (33)	Off-target	Missense
rs738409	PNPLA3	2B	prednisolone, vincristine	Lymphoma	15.7	Possibly Damaging (0.906)	Yes (34)	Off-target	Missense
rs2232228	HAS3	2B	anthracyclines	Cardiomyopathies	15.61	-	Proposed (35)	Off-target	Synonymous
rs16969968	CHRNA5	2B	nicotine	Tobacco Use Disorder	14.97	Benign (0.011)	Yes (36)	Off-target	Missense
rs1800497	ANKK1,DRD2	2B	ethanol	Alcoholism	10.6	Benign (0)	Yes (37)	Off-target	Missense
rs1051730	CHRNA3	2B	nicotine		10.1	-	No		Synonymous
rs1076560	DRD2	2B	cocaine	Cocaine-Related Disorders	8.805	-	No		Intronic
rs746647	CCHCR1	2B	nevirapine	HIV	5.575	-	No		Intronic
rs4693075	COQ2	2B	statins	Muscular Diseases	2.092	-	No		Intronic
rs10497203	TANC1	2B	radiotherapy	Prostatic Neoplasms	1.974	-	Proposed (38)	Off-target	Intronic
rs730012	LTC4S	2B	aspirin	Urticaria	1.936	-	No		5'-flanking
rs1872328	ACYP2	2B	cisplatin	Brain Neoplasms, Ototoxicity	0.897	-	Limited (39)	Off-target	Intronic
rs716274	DYNC2H1	2B	etoposide, platinum compounds		0.621	-	No		Intergenic
rs489693	MC4R	2B	clozapine, risperidone	Autism, Schizophrenia	0.325	-	No		Intergenic
rs7779029	SEMA3C	2B	irinotecan	Carcinoma	0.165	-	No		Intronic
rs1517114	C8orf34	2B	irinotecan	Carcinoma	0.081	-	No		Intronic

Table 2 – High confidence PharmGKB variants sorted with criteria to identify possible off-target effects.