Efficacy of computational predictions of the functional effect of idiosyncratic pharmacogenetic variants

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

Abstract Background. Pharmacogenetic variation is important to drug responses through diverse and complex mechanisms. Predictions of the functional impact of missense pharmacogenetic variants primarily rely on the degree of sequence conservation between species as a primary discriminator. However, idiosyncratic or off-target drug-variant interactions sometimes involve effects that are peripheral or accessory to the central systems in which a gene functions. Given the importance of sequence conservation to functional prediction tools – these idiosyncratic pharmacogenetic variants very likely violate the assumptions of predictive software commonly used to infer their effect.

Methods. Here we exhaustively assess the effectiveness of eleven missense mutation functional inference tools on all known pharmacogenetic missense variants contained in the Pharmacogenomics Knowledgebase (PharmGKB) repository. We categorize PharmGKB entries into sub-classes to catalog likely off-target interactions , such that we may compare predictions across different variant annotations.

Results. As previously demonstrated, 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 off-target, type B adverse drug reactions, that are largely unrelated to the main pharmacological action of the drug. Specifically, variants associated with off-target effects (hence referred to as off-target variants) were most often incorrectly classified as ‘benign’. These results highlight the importance of understanding the underlying mechanism of pharmacogenetic variants and how variants associated with off-target effects will ultimately require new predictive algorithms. We describe how to identify variants associated with 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.

Efficacy of computational predictions of the functional effect of idiosyncratic pharmacogenetic variants

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 is important to drug responses through diverse and complex mechanisms. Predictions of the functional impact of missense pharmacogenetic variants primarily rely on the degree of sequence conservation between species as a primary discriminator. However, idiosyncratic or off-target drug-variant interactions sometimes involve effects that are peripheral or accessory to the central systems in which a gene functions. Given the importance of sequence conservation to functional prediction tools – these idiosyncratic pharmacogenetic variants very likely violate the assumptions of predictive software commonly used to infer their effect.

Methods. Here we exhaustively assess the effectiveness of eleven missense mutation functional inference tools on all known pharmacogenetic missense variants contained in the Pharmacogenomics Knowledgebase (PharmGKB) repository. We categorize PharmGKB entries into sub-classes to catalog likely off-target interactions , such that we may compare predictions across different variant annotations.

Results. As previously demonstrated, 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 off-target, type B adverse drug reactions, that are largely unrelated to the main pharmacological action of the drug. Specifically, variants associated with off-target effects (hence referred to as off-target variants) were most often incorrectly classified as 'benign'. These results highlight the importance of understanding the underlying mechanism of pharmacogenetic variants and how variants associated with off-target effects will ultimately require new predictive algorithms. We describe how to identify variants associated with 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 safe and effective patient treatment (Meyer 2000; Pirmohamed 2001). Many oncology drugs have be shown to be ineffective in subsets of patients, 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 between 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 been 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 mechanism of such genetic factors is often little understood. Numerous other factors play a role in variable drug response, including age, ethnicity, gender and differences in alcohol intake.

There are a growing number of databases that aggregate, curate and annotate the increasing body of identified genetic variants that occur in genes that interact with a pharmaceutical (pharmacogenes). (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 nonsynonymous and synonymous single nucleotide variants (SNVs) to non-coding, intergenic and copy number variants.

Predicting the potential functional impact of a missense mutation is necessary, due to the disparity between the number of identified variants and the number that have experimentally-derived 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 are data tools that 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). These tools are known to suffer from high false positive rates with previous work estimating >40% of all variants predicted to be deleterious had little measurable impact on protein function (Miosge et al. 2015). Generally, the algorithms work by deriving a multi-factorial score with higher values (with the exception of SIFT) representative of variants more likely to be damaging to the structure and function of the protein. Several algorithms bin their values into discrete named categories with PolyPhen2

applying labels of 'benign' for scores from 0 and <0.15, 'possibly damaging' for scores from 0.15 and <0.85 and 'probably damaging' for scores of 0.85-1. The algorithms are often trained on distinct sets of variants with CADD employing a machine learning model trained on a binary distinction between fixed variants arisen since the human/chimp split and simulated *de novo* variants. The types of evidence employed by the algorithms are numerous with CADD considered 60 annotation sets based largely based on conservation (e.g. phastCons, GERP), epigenetic modifications (e.g. DNase-Seq, H3K9Ac), functional prediction (e.g. TF motif, amino acid change), and genetic content (e.g. GC content, CpG content) amongst others. Despite the diversity of evidence types however, 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 mutations, to our knowledge all tools incorporate sequence conservation data in their algorithms. Hence, sequence conservation 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. 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). Specific missense mutations have been shown to cause adverse off-target effects with rs1050828 causing glucose-6-phosphate dehydrogenase (G6PD) deficiency which induces haemolytic anaemia from anti-malarial drugs such as primaquine (Gampio Gueye et al. 2019). 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. 2018b). Such studies led us to examine pharmacogenetic variants in order to identify subclasses that are likely to perform poorly with such tools such as variants associated with adverse drug reactions (ADRs).

ADRs are broadly classified according to general mechanistic distinctions (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 reactions (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 anesthetics 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 variant sub-classes associated with Type A and Type B reactions to determine whether the functional inferences differ. We find that most PharmGKB entries incorrectly classified as benign are generally off-target, idiosyncratic variants. As such, using current functional-effect prediction tools may produce poor inferences on idiosyncratic pharmacogenetic variants. Identifying lists of such variants generates a training set suitable to develop and calibrate new algorithms designed for this specific sub-class of variant.

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. To obtain functional inference scores, variants were further annotated with Variant Effect Predictor (VEP) (McLaren et al. 2016).

Classification of Off-Target Pharmacogenetic Variants

A simple classification scheme was devised to identify and confirm likely off-target variants (**Table 1**). Input variants for this classification were the complete set of PharmGKB variants (Whirl-Carrillo et al. 2012). First, of these variants, all synonymous and non-coding variants were excluded, leaving just missense variants. All clinical variants were then 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 absorption, distribution, metabolism, and excretion process genes (ADME; 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'. With this filtered list, the cited literature for each variant was then appraised to discern whether a mechanism of action was known and that the variant resulted in an off-target or idiosyncratic effect. The classification scheme and variant counts at each step is summarized in **Table 1** and the list of filtered variants in **Supplemental Table S1.**

Validation of classification scheme

To validate how effective this classification system was at capturing off-target variants, we randomly sampled high-confidence, missense variants from the PharmGKB (Whirl-Carrillo et al. 2012) until we derived 30 variants with a known pharmacogenetic mechanism (**Supplemental Table S2**). For each of these 30 variants we conducted a literature review to manually classify each variant as Type A or Type B. Manual classification was geared toward being stringent and followed a checklist where the variant was assigned to Type B if it satisfied all of the following criteria: 1) was not a metabolic process gene associated with normal metabolism of the drug, 2) was not in a gene associated with the system the drug is prescribed to target, and 3) did not have a dose effect. These variants were used as a truth set to measure the performance of the classification scheme.

Functional Effect Prediction

For all PharmGKB missense mutations, 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 dbNSFP v2.0 (Liu et al. 2013)~~, queried with Variant Effect Predictor (McLaren et al. 2016)~~ (**Supplemental Table S3**).

Receiver Operator Curves (ROC)

ROC curves were obtained for the most-widely used subset of these tools (CADD, PolyPhen2, SIFT, Mutation Assessor, MutPred and REVEL) using the R package ROCR (Sing et al. 2005) (**Figure 1**). All high-confident PharmGKB Category 1 variants were input as the positive set while a set of randomly selected common variants (MAF > 0.1) were input as the negative set using the Perl function rand() across the entire set of dbSNP missense variants. Labels were inverted for SIFT due to lower scores representing likely damaging mutations and CADD and Mutation Assessor scores were scaled into the range of 0-1. Area under the curve and Matthew Correlation Coefficient were calculated using the R-package ROCR performance function (**Table 2**).

Results

Distributions of pharmacogenetic variant functional inferences

Functional inference scores were obtained for 561 missense single nucleotide variants (SNVs) present in PharmGKB, that also had dbSNP cluster identifiers. Predictions were made for each SNV with eleven different prediction tools (SIFT, PolyPhen2, CADD, DANN, FATHMM, GERP++, MutPred, Mutation Assessor, Mutation Taster, REVEL and PhastCons) (**Supplemental Table S3**). Instances where a tool produced no value for a given SNV were recorded as an NA

value. The distributions of scores from the most widely-used subset of these tools (CADD, PolyPhen2, SIFT, Mutation Assessor, Mutation Predictor and REVEL) are plotted with variants grouped by major PharmGKB category 1-4 (**Figure 2**). The predictions calculated for these functional variants ranged widely from benign to deleterious. While four of the tools generate scores in the range of 0-1 (with 1 being most damaging ~~expect~~ for SIFT ~~with~~ 0 ~~being~~ most damaging). Mutation Assessor and CADD employ a range of positive values with CADD calculating a Phred-quality score. For example, a CADD score of 20 implies the variant is ranked in the top 0.1% of all possible variant scores based on all possible changes in the human genome (CADD score of 10 is top 1%, CADD score of 20 is top 0.1%, CADD score of 30 is top 0.01%, etc.). For comparison to expected background levels, we also selected a random set of 2155 common human missense SNVs with assigned RS cluster identifiers. Overall, the distribution of random variants mirrors our previous work with Polyphen2 exhibiting a characteristic hourglass curve with very few intermediate values (Andrews et al. 2012). 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 2 as seq (sequence conservation), struct (protein structural metrics), and ens (ensemble tool that integrates individual tools). The results demonstrate how some 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 are predicted to be deleterious by PolyPhen2 (median score 0.996), however 6 variants were classified benign (rs116855232, rs1057910, rs121909041, rs3745274, rs1050828 and rs2228001). 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). Lower-confidence PharmGKB variants in category 3 and 4 are expected to contain a mix of real and false positive pharmacogenetic variants making it difficult to gauge whether these categories are enriched for type B variants.

To better assess the performance of the individual algorithms on high quality pharmacogenetic variants, ROC plots were generated for CADD, PolyPhen2, SIFT, Mutation Assessor, Mutation Predictor and REVEL (**Figure 1**). Area under the curve (AUC) and Matthews Correlation Coefficient were calculated (**Table 2**). Overall, Mutation Predictor had the highest AUC at 0.974 followed by REVEL at 0.94, PolyPhen2 at 0.852 and the remaining algorithms raging from 0.727-0.774. PolyPhen2 had the highest Matthews Correlation Coefficient at 0.986, followed by SIFT at 0.97, Mutation Predictor at 0.828 and the remaining algorithms raging from 0.643-0.786. While there are significant differences, no single best tool performs optimally across all variants.

In addition to missense mutations, tools such as CADD are able to generate scores for other variant types such as non-coding SNVs. While detailed analysis of this type of variant is beyond the scope

of this study, we identified 14 PharmGKB high-quality category 1 and 2 non-coding variants and generated CADD scores. The median CADD score was 14.0, well below the average of 27.2 for category 1 PharmGKB and even less than the 22.1 for the random 2155 dbSNP variants.

Classification of pharmacogenetic variation to detect off-target effects

A prior study (Zhou et al. 2018b) 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 associated with off-target effects. To investigate the possibility that type B pharmacogenetic variants are predominantly predicted to be benign, we devised a simple classification system to enrich these from PharmGKB (described in *Materials and Methods* and summarized in **Table 1**). Our starting data was ~~a possible all~~ PharmGKB clinical variants found to cause missense mutations. In order to discern how effective this classification system was at capturing Type B variants, we randomly sampled from the starting set of PharmGKB variants until we derived 30 variants from distinct genes with a published pharmacogenetic mechanism (**Supplemental Table S2**). From a literature review we manually classified each variant as being of Type A or Type B (see Methods), finding nine of the 30 variants (30%) to be Type B variants. With this validation data as a truth set, we counted true and false positives and negatives resulting from our classification scheme (**Supplementary Table S3**). Of the starting set of 30 variants, the filtering scheme retained eight putative Type B variants. Of these eight, five were true positive Type B variants (with no false negatives) and three were false positives. Given this retention of Type B variants from the unfiltered variant pool (9/30=0.3) to the enriched pool (5/8=0.63), we estimate from this data that this classification scheme yields a 2.1-fold (0.63/0.3) enrichment of Type B missense variants, with a sensitivity of 63% (5/8) and a specificity of 100% (21/21). Subsequently, with this classification scheme, we performed the classification on the full set of 561 PharmGKB variants causing missense mutations. Of these, this filtering system retained 142 missense variants and generated a median PolyPhen2 score of 0.061 (**Supplemental Table S1**) substantially lower than the median value of 0.245 for the randomly selected variants. Further, these 142 variants included all nine Type B variants identified in the validation set.

Discussion

In this work we have appraised whether pharmacogenetic variants associated with Type B, off-target effects are consistently predicted to be less deleterious than other functionally-important variants. We find this to be the case and postulate that this results arises from the reliance of the current generation of missense mutation inference tools on sequence conservation information. Generally, when a deeply conserved genetic element is found to be mutated, this will result in this mutation being predicted to be deleterious. However, should a nucleotide not be conserved across deep evolutionary distances, but be the site of interaction with a recently developed (in an evolutionary sense) drug molecule, mutations at this site are likely to be predicted 'benign'. ~~Yet,~~

this is clearly not the case, should this drug confer a life-saving benefit. The current cohort of predictive tools that utilize sequence conservation assume that nucleotide sites that vary over evolutionary timescales are necessarily benign. However, there are several reasons why this assumption may not hold for some pharmacogenetic variants.

One explanation is the time required for purifying selection to act. Purifying selection acts to remove damaging mutations over an evolutionary time scale. Xenobiotic drugs being prescribed to human patients, however, is a very recent occurrence in an evolutionary context. Unless a pharmacogenetic variant is related to the evolved functions of a gene, then no information is present in the ancestral sequence record from which to detect functional importance. 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 genes that interact with drugs (pharmacogenes) contain variants which generate Type A ADRs. For these Type A pharmacovariants, the drug most often just another xenobiotic compound which the target-protein acts upon. Variants which adversely affect the function of the pharmacogene should be correctly classified as 'deleterious' by the current generation of functional inference tools. This is supported by previous work showing an association between the residual evidence intolerance score (RVIS which measure the tolerance of a gene to mutations) and targets of approved drugs (Nelson et al. 2015) however it is unknown whether this holds for off-target variants. Variants that cause a type B or off-target effect are much less likely to be subject to the same selection pressures as those of type A, meaning such variants will may 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. Given the importance of pharmacogenetic variation and the numerous nature of Type B pharmacovariants, new methods are urgently needed to capture this important class of variation.

Another possible explanation proposed by Zhou et al (Zhou et al. 2018a) is that the genes containing many pharmacogenetic variants are often poorly conserved, making the reliance of the algorithms on sequence conservation alone problematic. The quality of the multiple sequence alignments is also important with the class of multiple sequence alignment algorithm selected shown to substantially impact downstream analyses (Blackburne & Whelan 2013). An additional consideration is the constraints imposed by domain structure on missense mutations across the human genome (MacGowan et al. 2017). This work identified regions of the genome depleted of missense mutations and while most such regions were conserved across species, they identified regions that are not conserved yet were enriched for pathogenic variant, ligands, and DNA and protein binding interactions. Such variants are also unsuitable for the sequence based tools and similar to pharmacogenetic variants require non-sequence based tools to accurately predict their functional impact. A final consideration is the possibility that a SNV could disrupt the interaction between the protein and the drug however it would exhibit no impact on the

protein function in isolation. Such pharmacogenetic variants would be expected to be invisible to the current generation of functional inference prediction software.

While sequence conservation is a useful metric for predicting the impact of many variants, we have shown for certain subclasses of variants it is not suitable. However, 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 (**Table 2:** average AUC for tools using structural information is 0.849 vs 0.825 for tools that do not; p-value=0.79 2-sample t-test). 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. 2018b). Regardless of the eventual outcome, the ability to accurately predict pharmacogenetic variants associated with 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. Our analysis confirms that in many cases, the assumptions of functional inference tools are invalid, particularly for variants associated with 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 accurately detect pharmacogenetic variation.

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.

Andrews TD, Whittle B, Field MA, Balakishnan B, Zhang Y, Shao Y, Cho V, Kirk M, Singh M, Xia Y, Hager J, Winslade S, Sjollema G, Beutler B, Enders A, and Goodnow CC. 2012. Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an immediate source for thousands of new mouse models. *Open Biol* 2:120061.

Blackburne BP, and Whelan S. 2013. Class of multiple sequence alignment algorithm affects genomic analysis. *Mol Biol Evol* 30:642-653.

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

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, and Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 6:e1001025.

Eichelbaum M, Ingelman-Sundberg M, and Evans WE. 2006. Pharmacogenomics and individualized drug therapy. *Annu Rev Med* 57:119-137.

Field MA, Cho V, Andrews TD, and Goodnow CC. 2015. Reliably Detecting Clinically Important Variants Requires Both Combined Variant Calls and Optimized Filtering Strategies. *PLoS One* 10:e0143199.

Gabb GM, Vitry A, Condon J, Limaye V, and Alhami G. 2013. Serious statin-associated myotoxicity and rhabdomyolysis in Aboriginal and Torres Strait Islanders: a case series. *Intern Med J* 43:987-992.

Gampio Gueye NS, Peko SM, Nderu D, Koukouikila-Koussounda F, Vouvoungui C, Kobawila SC, Velavan TP, and Ntoumi F. 2019. An update on glucose-6-phosphate dehydrogenase deficiency in children from Brazzaville, Republic of Congo. *Malar J* 18:57.

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, and Sieh W. 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 99:877-885.

Kalinin AA, Higgins GA, Reamaroon N, Soroushmehr S, Allyn-Feuer A, Dinov ID, Najarian K, and Athey BD. 2018. Deep learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics* 19:629-650.

Khan S, and Vihinen M. 2010. Performance of protein stability predictors. *Hum Mutat* 31:675-684.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, and Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46:310-315.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, and Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25:2744-2750.

Liu X, Jian X, and Boerwinkle E. 2013. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 34:E2393-2402.

MacGowan SA, Madeira F, Britto-Borges T, Schmittner MS, Cole C, and Barton GJ. 2017. Human Missense Variation is Constrained by Domain Structure and Highlights Functional and Pathogenic Residues. *bioRxiv*:127050.
McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, and Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* 17:122.
Meyer UA. 2000. Pharmacogenetics and adverse drug reactions. *Lancet* 356:1667-1671.
Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, Balakishnan B, Liang R, Zhang Y, Lyon S, Beutler B, Whittle B, Bertram EM, Enders A, Goodnow CC, and Andrews TD. 2015. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci U S A*.
Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR, Whittaker JC, and Sanseau P. 2015. The support of human genetic evidence for approved drug indications. *Nat Genet* 47:856-860.
Pang GS, Wang J, Wang Z, and Lee CG. 2009. Predicting potentially functional SNPs in drug-response genes. *Pharmacogenomics* 10:639-653.
Patton K, and Borshoff DC. 2018. Adverse drug reactions. *Anaesthesia* 73 Suppl 1:76-84.
Pirmohamed M. 2001. Pharmacogenetics and pharmacogenomics. *Br J Clin Pharmacol* 52:345-347.
Quang D, Chen Y, and Xie X. 2015. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 31:761-763.
Reva B, Antipin Y, and Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 39:e118.
Robinson R, Carpenter D, Shaw MA, Halsall J, and Hopkins P. 2006. Mutations in RYR1 in malignant hyperthermia and central core disease. *Hum Mutat* 27:977-989.
Schiemann AH, and Stowell KM. 2016. Comparison of pathogenicity prediction tools on missense variants in RYR1 and CACNA1S associated with malignant hyperthermia. *Br J Anaesth* 117:124-128.
Schwarz JM, Cooper DN, Schuelke M, and Seelow D. 2014. MutationTaster2: mutation 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.
Sing T, Sander O, Beerenwinkel N, and Lengauer T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics* 21:3940-3941.
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, Fujikura K, Mkrtchian S, and Lauschke VM. 2018a. Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. *Front Pharmacol* 9:1437.

Zhou Y, Mkrtchian S, Kumondai M, Hiratsuka M, and Lauschke VM. 2018b. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J*.

Figure 1 – Receiver operator curves (ROC) for CADD, Mutation Assessor, MutPred, PolyPhen2, REVEL, and SIFT generated with ROCR. All high-confident PharmGKB Category 1 missense variants were input as the positive set while a set of randomly selected common variants (MAF > 0.1) were input as the negative set using the Perl function rand() across the entire set of dbSNP missense variants. Labels were inverted for SIFT due to lower scores representing likely damaging mutations and CADD and Mutation Assessor scores were scaled into the range of 0-1.

Figure 2 - 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. Each tool employs categorical cutoffs with CADD suggesting >15, PolyPhen2 using >0.85, SIFT using < 0.05, MutPred using >0.5, Mutation Assessor using >0.65, and REVEL suggesting > 0.5.

Figure 1

Receiver operator curves (ROC) for six functional inference tools.

Receiver operator curves (ROC) for CADD, Mutation Assessor, MutPred, PolyPhen2, REVEL, and SIFT generated with ROCR. All high-confident PharmGKB Category 1 missense variants were input as the positive set while a set of randomly selected common variants (MAF > 0.1) were input as the negative set using the Perl function rand() across the entire set of dbSNP missense variants. Labels were inverted for SIFT due to lower scores representing likely damaging mutations and CADD and Mutation Assessor scores were scaled into the range of 0-1.

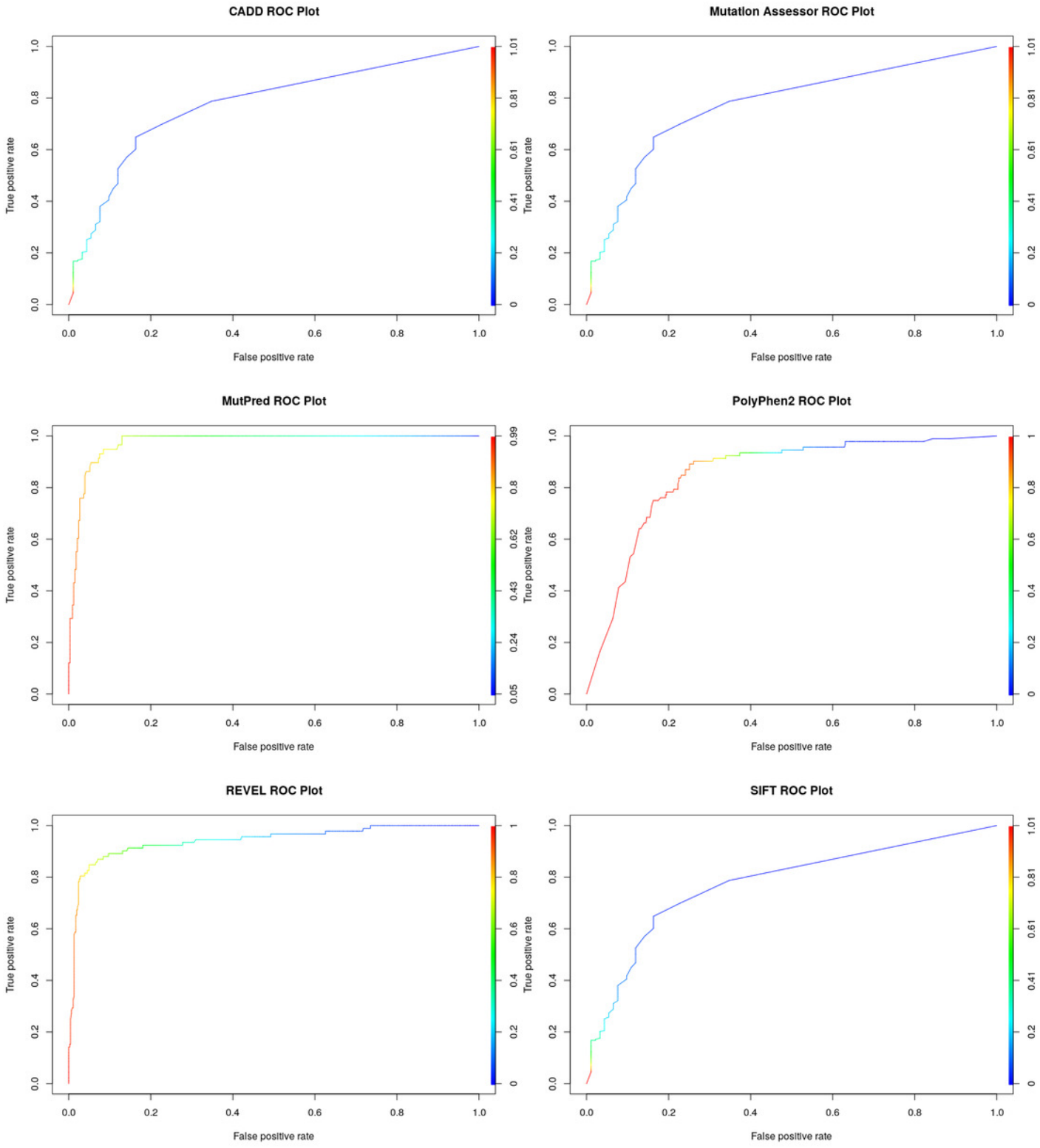
CADD ROC Plot
Mutation Assessor ROC Plot
MutPred ROC Plot
PolyPhen2 ROC Plot
REVEL ROC Plot
SIFT ROC Plot
True positive rate
False positive rate

Figure 2

Functional inference scores across all PharmGKB variants grouped by confidence level.

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. Each tool employs categorical cutoffs with CADD suggesting >15, PolyPhen2 using >0.85, SIFT using < 0.05, MutPred using >0.5, Mutation Assessor using >0.65, and REVEL suggesting > 0.5.

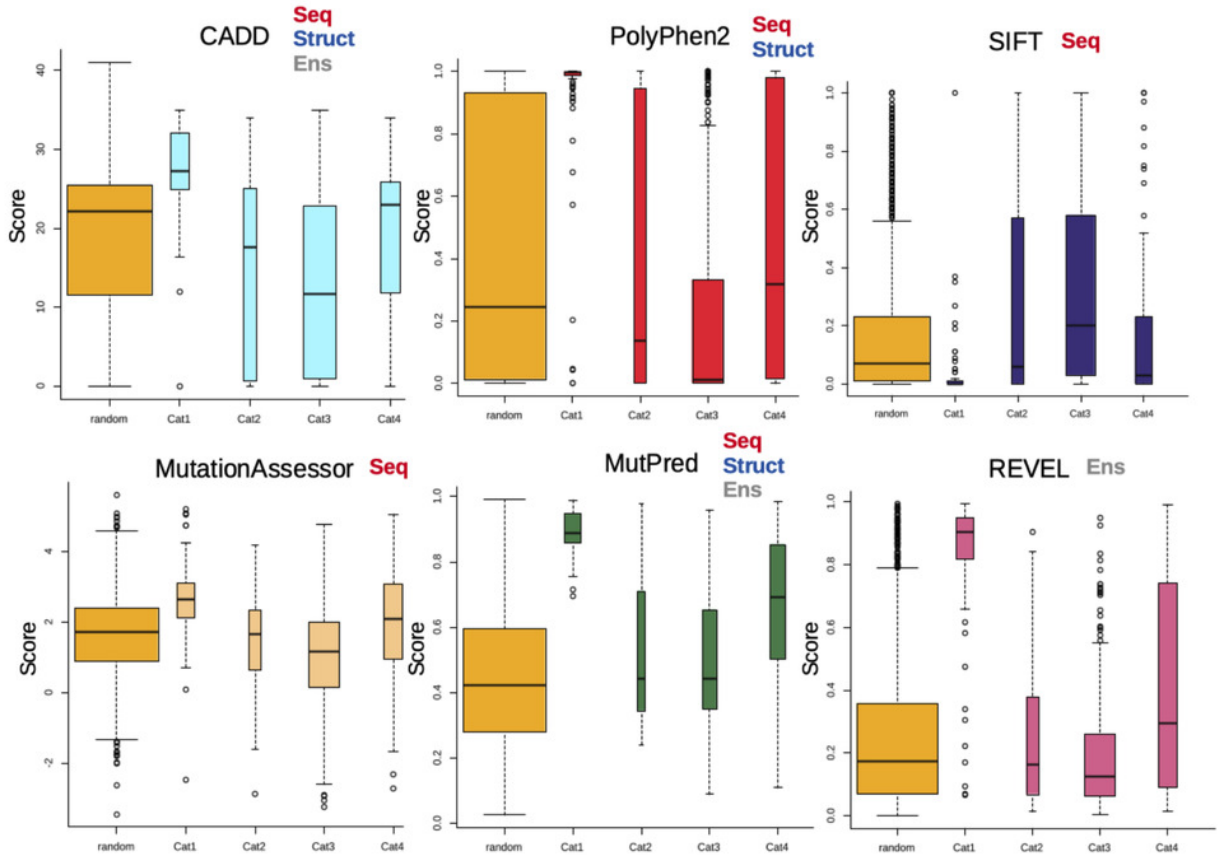
CADD
Seq
Struct
Ens
PolyPhen2
Seq
Struct
SIFT
Seq
MutationAssessor
Seq
MutPred
Seq
Struct
Ens
REVEL
Ens
Score
random
Cat1
Cat2
Cat3
Cat4

Table 1(on next page)

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

Step	Filter	Number ~~Varints~~
1	Exclude synonymous and non-coding variants	561
2	Include variants that have type:toxicity/ADR	339
3	For drug and gene pairs, exclude variants with additional effect types other than Toxicity/ADR	273
4	Gene containing variant is NOT an ADME process gene OR annotated in GO with 'xenobiotic metabolic process' OR 'transporter'	196
5	Literature review indicates a known mechanism AND mechanism indicated an idiosyncratic drug/protein interaction	142

1
2

Table 2 (on next page)

AUC and MCC from PolyPhen2, MutPred, REVEL, SIFT, CADD, and Mutation Assessor using category 1 and 2 PharmGKB variants versus common dbSNP variants.

Algorithm	Area Under Curve (AUC)	Matthews Correlation Coefficient (MCC)
PolyPhen2	0.8521586	0.986
MutPred	0.9745274	0.828
REVEL	0.9416667	0.786
SIFT	0.7739619	0.97
CADD	0.7279054	0.643
Mutation Assessor	0.7632535	0.6932515

1