

## REVIEW ARTICLE OPEN

# Emerging strategies to bridge the gap between pharmacogenomic research and its clinical implementation

Volker M. Lauschke<sup>1\*</sup> and Magnus Ingelman-Sundberg<sup>1</sup>

The genomic inter-individual heterogeneity remains a significant challenge for both clinical decision-making and the design of clinical trials. Although next-generation sequencing (NGS) is increasingly implemented in drug development and clinical trials, translation of the obtained genomic information into actionable clinical advice lags behind. Major reasons are the paucity of sufficiently powered trials that can quantify the added value of pharmacogenetic testing, and the considerable pharmacogenetic complexity with millions of rare variants with unclear functional consequences. The resulting uncertainty is reflected in inconsistencies of pharmacogenomic drug labels in Europe and the United States. In this review, we discuss how the knowledge gap for bridging pharmacogenomics into the clinics can be reduced. First, emerging methods that allow the high-throughput experimental characterization of pharmacogenomic variants combined with novel computational tools hold promise to improve the accuracy of drug response predictions. Second, tapping of large biobanks of therapeutic drug monitoring data allows to conduct high-powered retrospective studies that can validate the clinical importance of genetic variants, which are currently incompletely characterized. Combined, we are confident that these methods will improve the accuracy of drug response predictions and will narrow the gap between variant identification and its utilization for clinical decision-support.

*npj Genomic Medicine* (2020)5:9; <https://doi.org/10.1038/s41525-020-0119-2>

## INTRODUCTION

The past several decades of research have seen a dramatic increase in our understanding of how inter-individual genetic variation can impact drug pharmacokinetics, response, and toxicity. However, the clinical application of this knowledge has progressed only slowly. A variety of reasons have been identified to contribute to this latency in clinical implementation.<sup>1–3</sup> Arguably among the most important factors is the substantial amount of underpowered preliminary studies that report interactions between specific variants and pharmacological phenotypes without replication in independent cohorts or corroborative mechanistic data. Such associations are not useful to guide prescribing and fuel the frequently stated concerns with clinical utility as a key barrier for the clinical use of pharmacogenomics. Partly as a consequence of this vast body of literature, clinicians and service providers feel insecure and insufficiently educated in pharmacogenomics to promote further clinical implementation.

In addition, pharmacogenomic biomarker discovery is complicated by the remarkable pharmacogenetic complexity. Specifically, recent next-generation sequencing (NGS) have identified thousands of novel pharmacogenetic variants that are overlooked when using conventional analysis methods for single nucleotide polymorphism and copy number variation (CNV) profiling.<sup>4,5</sup> Importantly, many of these rare variants putatively affect the function of the corresponding gene product and are likely to contribute to inter-individual differences in drug disposition and response. Although statistical tests such as burden or variant-component tests provide important frameworks for the analysis of rare variant associations, their suitability for pharmacogenomic applications is limited due to the large number of causal variants whose effects can differ in directionality.<sup>6</sup> Interpretation of NGS-based data is thus suggested to use a two-pronged approach. Experimental or clinical information is used for variants for which

such data are available, whereas the interpretation of uncharacterized variants has to rely on computational predictions. Importantly, it is evident that the success of such a strategy strongly depends on both solid experimental data as well as accurate computational prediction tools. Finally, there is a lack of clear consensus guidelines in product labels and recommendations from different pharmacogenomic expert groups for utilizing genomic data to guide personalized prescribing.

The field of pharmacogenomic biomarker-guided drug therapy has mainly advanced in the field of oncology where drugs are commonly targeted to specific genetic variations in the somatic tumor genome. However, also germline variants can guide the choice and optimal dosing strategy for various chemotherapeutic agents, including fluoropyrimidines or thiopurines. By contrast, the implementation of pharmacogenomics in other therapeutic areas is lagging behind. In order to enhance the reliability and value of predictive pharmacogenomic information in the clinics, there is a need to expand the genetic information retrieved by considering the entire genetic landscape into clinical decision-making and to directly assess the outcome and cost benefit of sequencing-guided therapy. In this matter, retrospective studies based on previous available therapeutic drug monitoring (TDM) data have provided important contributions.

In the following sections, we summarize the state-of-the-art of how to include the entire pharmacogenetic makeup of an individual, including rare variants, into personalized pharmacogenomic advice. Furthermore, we highlight novel strategies and tools to improve the quality of such drug response predictions. Lastly, we provide examples how retrospective analyses of large existing pharmacokinetic TDM data sets can improve the reliability of conclusions regarding the clinical importance of genetic variations.

<sup>1</sup>Department of Physiology and Pharmacology, Karolinska Institutet, SE-171 77 Stockholm, Sweden. \*email: volker.lauschke@ki.se

## HOW TO IMPROVE THE FUNCTIONAL INTERPRETATION OF RARE PHARMACOGENETIC VARIABILITY

Inter-individual differences in drug response and toxicity are partly due to heritable factors, with common estimates suggesting that the genetic component might explain 20–30% of this variability. Each individual genome contains around 25,000 genetic variations in exons, including 10,000–12,000 missense variants and 100 putative loss-of-function variants.<sup>7,8</sup> Genes involved in drug absorption, distribution, metabolism, and excretion (ADME) are particularly complex, and analyses of whole exome and whole genome sequencing data from more than 100,000 individuals indicated that *CYPs*<sup>9,10</sup> and drug transporters of the *SLC*<sup>11</sup> and *SLCO*<sup>12</sup> gene families harbor tens of thousands of different single nucleotide variants (SNVs) and indels. Importantly, the overwhelming majority of these variants (>98%) are rare with minor allele frequencies in the general population below 1%.<sup>13–16</sup> Moreover, analyses of CNVs in ADME genes revealed that of the 208 genes analyzed, 201 (97%) carried deletions and duplications, most of which were rare, spanning one to multiple exons.<sup>17</sup> Similarly, rare SNVs and CNVs were highly prevalent in genes encoding G-protein-coupled receptors, one of the most important families of drug targets.<sup>18</sup>

The identification of rare variants requires the use of methods that can comprehensively profile genes of importance for drug pharmacokinetics and response. Genetically mediated modulation of pharmacokinetic profiles can furthermore translate into altered risks for adverse drug reactions, particularly for chemotherapeutic drugs. Specific examples include variants in *DPYD* that increase fluorouracil toxicity and variations in *TPMT* and *UGT1A1*, which predispose to thiopurine- and irinotecan-induced myelosuppression, respectively. The increase in speed and accuracy of NGS, combined with decreasing sequencing costs, facilitates the use of sequencing for routine applications. Importantly, however, a considerable number of genes with importance for drug response and toxicity (up to 50, including *CYP2D6*, *CYP2A6*, and *HLA* genes) are located in complex loci that feature low complexity regions, segmental duplications, and variable numbers of tandem repeats. Analysis of the genetic variability within such genes requires the use of specialized sequencing technologies and variant calling pipelines. Promising emerging methods in this space include single-molecule real-time sequencing and Nanopore sequencing, which have both been successfully used for whole genome sequencing applications,<sup>19,20</sup> as well as for targeted haplotyping and CNV profiling of the complex *CYP2D6*<sup>21–24</sup> and *HLA* loci.<sup>21,25</sup>

To characterize the functional effects of common genetic variants, a variety of methods and tools are available, including genetic association studies, in vitro characterizations, and computational predictions. However, for multiple reasons this toolbox is not directly applicable for rare variant assessments:

- i. Genetic association studies require a substantial number of variant carriers to be sufficiently powered to identify effects of a variant of interest on a clinical phenotype, which, depending on the experimental design and phenotypic heterogeneity, can require tens to hundreds or thousands of individuals. However, recruitment of these numbers of carriers for rare alleles is most often unfeasible or impractical.
- ii. Functional characterization of tens of thousands of variants using heterologous expression systems is not feasible using conventional protocols.
- iii. Most computational algorithms to predict the functional consequences of genetic variants have been trained on pathogenic data sets and use, at least in part, evolutionary conservation as the basis for decision-making. However, for genes involved in drug disposition, evolutionary constraint is generally low and thus conservation at the nucleotide

level constitutes only a poor proxy for functional consequences of the encoded gene product.

Importantly, in the last few years a variety of methodological and technological advances have opened exciting avenues to overcome these challenges. In the following two sections, we will provide an overview of computational (“Computational analysis of pharmacogenomic variation”) and experimental (“Experimental high-throughput characterization of pharmacogenomic variants”) strategies to improve the functional interpretation of rare pharmacogenetic variants.

### Computational analysis of pharmacogenomic variation

The translation of personal genomic information into actionable advice requires predictions about the functional consequences of the identified genetic variability for which no experimental or epidemiological characterization data are available. In the past decades, more than 60 different prediction tools have been developed.<sup>26</sup> SIFT,<sup>27</sup> PolyPhen2,<sup>28</sup> PROVEAN,<sup>29</sup> and MutationTaster2<sup>30</sup> constitute the most widely used models. However, they are restricted to the interpretation of variants that cause amino acid exchanges. Prediction tools that are suitable for the interpretation beyond missense variants include FATHMM,<sup>31,32</sup> CADD,<sup>33</sup> and Eigen.<sup>34</sup>

The vast majority of tools are based on machine learning, often used synonymously with artificial intelligence, to optimize predictions. Training of such models relies on data sets of deleterious and neutral variants in which variations are annotated regarding their functional consequence. As deleterious data sets, most studies use variants implicated in Mendelian disorders, mined from publically available collations such as OMIM, HGMD, or UniProtKB. For neutral variants, the most common approach is to consider all variants as neutral whose frequency is above a certain threshold (e.g., 1% or 5%) in the general population, as revealed by large-scale sequencing projects, such as the 1000 Genomes Project. This selection is based on the rationale that common variants that are present in a considerable fraction of the population are apparently not sufficient to cause clinically manifest heritable disease. From these training sets, the algorithm aims to identify patterns of statistical associations among features, including information about the sequence context of the affected nucleotide, evolutionary conservation, or physicochemical consequences of the amino acid exchange on protein level, and, based on these rules, predict the functional consequence of unannotated variants.

Importantly, although this approach is promising for the prediction of a variant’s pathogenicity, this approach is linked to multiple problems when the functionality of pharmacogenomic variants should be analyzed. First, pharmacogenes harbor a multitude of variants that alter function without being pathogenic. Common examples of variations with minor allele frequencies > 10% that reduce gene function and have established effects on drug pharmacokinetics or toxicity include *CYP2C9*\*2 (rs1799853), *CYP2C19*\*2 (rs4244285), *CYP2D6*\*4 (rs3892097), *CYP3A5*\*3 (rs776746), *UGT1A1*\*28 (rs34983651), and *SLCO1B1*\*5 (rs4149056). Using the approach illustrated above, all these variants are included in the neutral training set. Second, the statistical associations between features are different between pathogenic and functional but non-pathogenic variants. In particular, evolutionary conservation is different between variants within genes associated with congenital diseases that are under purifying selection and those which are not disease associated, such as most pharmacogenes.<sup>35</sup> As a consequence, application of the relationships between conservation scores and functional impacts that were established on pathogenic vs. non-pathogenic training sets is not suitable to provide reliable predictions in the context of pharmacogenes.

To overcome these limitations, we recently developed a computational tool specifically tailored towards the functional interpretation of pharmacogenetic variations. To this end, we

leveraged quantitative, experimentally determined high-quality activity data from 337 variants distributed across 43 genes involved in drug metabolism and transport as training data.<sup>36</sup> In a first step, we optimized the parameterization of 18 current prediction methods and then integrated the best performing methods into one ensemble score. Using cross-validations, the resulting score achieved sensitivity and specificity of 93% for loss-of-function and functionally neutral variants, respectively, and, importantly, can provide quantitative estimates of the functional consequences of a pharmacogenetic variant. Notably, this predictive power was achieved despite the low number of pharmacogenomic variants available for model training. We expect that model accuracy will further improve as the number of functionally characterized variants increases and in the following section we highlight methods that could be instrumental in this process.

#### Experimental high-throughput characterization of pharmacogenomic variants

Due to the sheer number of variants, experimental characterization, the gold standard for functionality assessment, of the entire portfolio of rare pharmacogenetic variants has not been feasible. As a result, the impact of rare variants encountered in the genome of a given patient can currently only be estimated using computational tools, which impairs personalized medicine. Deep mutational scanning constitutes a state-of-the-art technology that combines large-scale parallel mutagenesis with phenotypic selection and deep sequencing to assess the activities of more than 100,000 mutant versions of a protein in a single experiment.<sup>37</sup>

Principally, deep mutational scanning requires a cell-based selection assay as well a diversity library of plasmids in which every amino acid is substituted individually to every other amino acid (Fig. 1). Although the selection assay is highly protein specific, metabolic activation of a substrate into a toxic metabolite or transport of toxins into or out of cells is likely applicable in the context of pharmacogenes. Diversity libraries can be constructed using parallelized oligonucleotide-directed mutagenesis, which avoids the bias in the introduction of mutations that occur when using alternative mutagenesis methods, such as error-prone PCR.<sup>38</sup> One-pot saturation mutagenesis offers an elegant and efficient method for the construction of comprehensive mutagenesis libraries within few days with minimal hands-on time.<sup>39</sup> The assembled library is then transfected into cells in amounts that result in transformants only harboring a single variant. Subsequently, the selection assay is applied, resulting in the specific selection of transfectants with neutral or deleterious variant, depending on the selection assay configuration. For instance, if deep mutational scanning is used to characterize a drug transporter and the selection assay comprises the export of cytotoxic compounds, variants that reduce the transport activity will be depleted after selection, whereas the frequency of neutral variants will increase accordingly. Variants detected by NGS in the initial diversity library that are over- or underrepresented following selection indicate increased and reduced function variants, respectively. Moreover, the extent of variant representation in the pool directly correlates with resistance to the applied agent, thus providing a quantitative readout of variant function.

This method has been successfully used for the comprehensive evaluation of variant effects on organismal fitness as well as for domain characterizations and the mapping of protein interaction sites, primarily in bacteria, yeast, and rodent.<sup>40–43</sup> Recently, these analyses were also extended to disease-associated human proteins, including BRCA,<sup>44</sup> YAP,<sup>45</sup> and PPARG,<sup>46</sup> thus providing an appealing methodology to experimentally characterize the entire catalog of possible genetic variability affecting a protein's amino acid sequence for a given human gene. Notably, however,

the workflow has to be tailored to each gene individually, resulting in the protocol to be work- and cost-intensive.

To overcome this difficulty, Matreyek et al.<sup>47</sup> developed VAMP-Seq, which systematically probes the effects of all variants on protein abundance rather than function per se. Using the clinically actionable genes *PTEN* and *TPMT* as targets, the authors identified 1138 and 777 variants that alter protein levels, respectively. For those *PTEN* variants for which abundance scores could be obtained, VAMP-Seq correctly flagged 64% of pathogenic variants as low abundance and 12% as possibly low abundance. However, the specificity of VAMP-Seq could not be accurately quantified due to a lack of *PTEN* variants with confirmed functional neutrality. For *TPMT*, the well-characterized alleles *TPMT\*2* and *TPMT\*3* account for 95% of *TPMT* loss-of-function variants and, thus, the immediate clinical impact of the additional variants affecting *TPMT* abundance is likely limited.

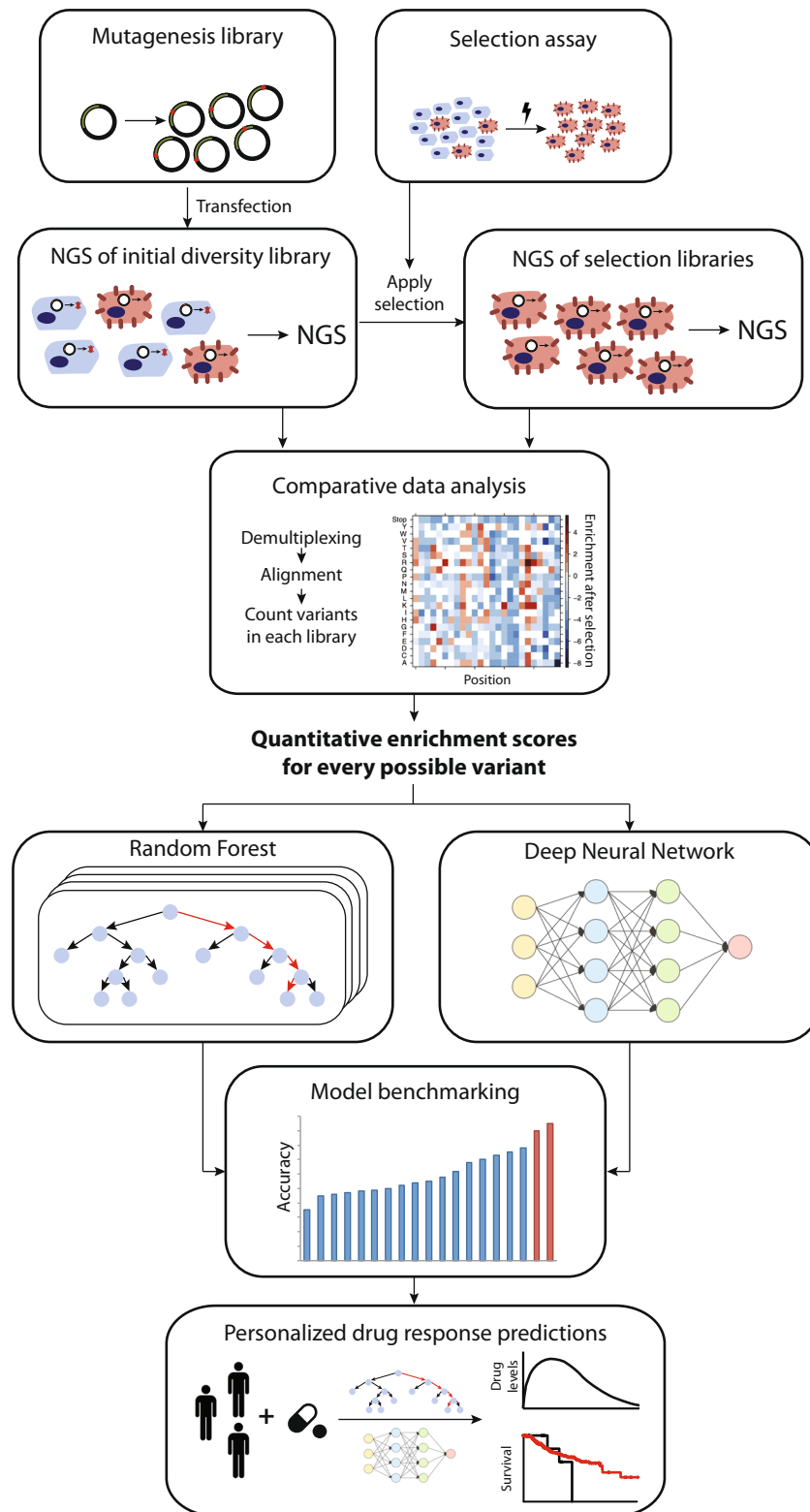
Importantly, such comprehensive functionally annotated data sets provide powerful tools that can be leveraged to improve computational predictions of genes not yet covered by deep mutational scanning. An interesting recent seminal study provided proof-of-concept that training of a decision tree-based model using available large-scale mutagenesis data resulted in superior performance compared with pathogenicity-trained variant predictors.<sup>48</sup> Specifically, the authors used a total of 21,026 experimentally determined variant effect scores from eight proteins to train a supervised, stochastic gradient boosting learning algorithm. The resulting tool, termed Envision, could predict quantitative variant effects and outperformed the previous predictors SIFT, PolyPhen2, SNAP2, and EVmutation also on protein data that were not used for model fitting. Interestingly, structural features were found to contribute more to model performance compared with evolutionary features, indicating the importance of feature completeness. As expected, inclusion of data from more proteins increased Envision's predictive performance. However, the tool has not yet been used for systematic pharmacogenetic predictions.

These developments are paralleled by recent advances in the variant effect prediction using deep learning. By leveraging variation data from non-human primate genomes for model training, deep neural networks have been successfully used to predict pathogenic variation in rare disease patients.<sup>49</sup> Furthermore, training of deep neural networks using available genome-wide association study data or phenotype similarity metrics can facilitate the identification of causative pathogenic variation within and outside of the coding regions.<sup>50,51</sup> Although the abovementioned tools focus on the identification and interpretation of disease-associated variation, the application of deep learning for pharmacogenomic variant effect prediction is currently limited to specific applications, such as the prediction of metabolic activity of *CYP2D6* genotypes and haplotypes, primarily due to the scarcity of available training data.<sup>52</sup>

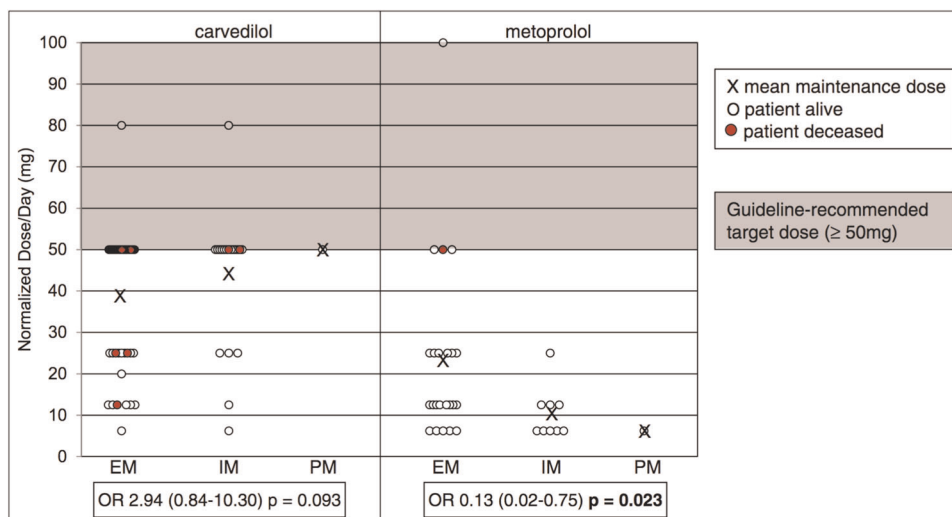
For the near future, we foresee that the integration of pharmacogene-specific experimental characterizations with model training based on the resulting quantitative data provides a promising strategy to drastically improve the quality of functional predictions, which constitutes one of the key challenges of personalized pharmacogenomics (Fig. 1).

#### HOW TO INCREASE THE RELIABILITY OF PHARMACOGENOMIC STUDIES

The low power and reliability of many studies in the published pharmacogenomic literature constitutes a key bottleneck for the translation of pharmacogenomic markers into clinical practice. The most important shortcomings are the limited power, often caused by a low number of patients in relation to the frequencies of the genetic variants studied, unbalanced inclusion and erroneous definition of the patients cohorts studied, analytical bias due to a



**Fig. 1 Schematic depiction of how deep mutational scanning aspires to improve the translation of pharmacogenomic information into actionable advice.** Cells are transfected with a diversity library of expression plasmids covering many (ideally all possible) amino acid substitutions for a given pharmacogene of interest and a protein-specific selection assay is applied. Deep sequencing of the transfected cells before and after selection, and comparative analysis of the obtained data can provide enrichment scores for each variant, thus enabling the massively parallelized experimental characterization of thousands of variations. Importantly, this data provides a powerful resource for the training and optimization of computational algorithms, including but not restricted to random forests and deep neural networks, which in turn can be applied to the integrated analysis of a patient's entire pharmacome.



**Fig. 2 Associations of *CYP2D6* genotype with the achievement of guideline-recommended target doses for carvedilol and metoprolol.** Analysis comprises 98 systolic heart failure patients. Only *CYP2D6*\*4 was considered for the definition of *CYP2D6* metabolizer status. Patients were 7.7 times more likely to be treated with lower maintenance doses of metoprolol for each *CYP2D6*\*4 allele. For carvedilol, a trend was observed between *CYP2D6*\*4 and higher maintenance dose. Gray shaded box indicates the range of guideline-recommended target doses ( $\geq 50$  mg). EM extensive metabolizer (no \*4 alleles), IM intermediate metabolizer (one \*4 allele), OR odds ratio, PM poor metabolizer (two \*4 alleles). Figure modified with permission from ref. <sup>56</sup>.

nonrandomized selection of the patient groups, and lack of result validation in independent cohorts. The resulting uncertainty is reflected in the limited consensus between pharmacogenomic labels as considered by different regulatory agencies.<sup>53,54</sup>

Large prospective randomized clinical trials would be most informative to cope with these problems. However, there is very limited public funding for such pharmacogenomic trials and only few examples without corporate sponsorship have been reported in the literature, such as the EU-PACT study for warfarin gene–dose relationships.<sup>55</sup> Thus, the field has tapped into the vast resource of available TDM data to analyze the impact of pharmacogenomic variation. Specifically, patients who were treated in the absence of genotypic information and whose drug or metabolite levels were monitored can be retrospectively genotyped, given that appropriate ethical approval is available or can be obtained.

Interesting examples that demonstrate the power of this design include the relationship of metoprolol maintenance dose in heart failure patients with *CYP2D6* genotype<sup>56,57</sup> (Fig. 2). Similarly, in a study of 993 patients it was concluded that the dose-adjusted phenytoin blood concentrations were higher and the risk for neurologic side effects was elevated by 2.4-fold among carriers of reduced function *CYP2C9* alleles.<sup>58</sup> A similar retrospective approach has been applied in many studies from different countries regarding the relative importance of *VKORC1* and *CYP2C9* genotypes for warfarin maintenance doses.<sup>59</sup> Indeed, a multitude of dosing algorithms have been developed in different countries for such predictions.

Large retrospective studies have been furthermore useful for evaluating gene–dose effects of antidepressant and antipsychotic treatment. In a study encompassing 2087 patients, a correlation between the *CYP2C19* genotype and TDM-based concentrations of escitalopram and metabolites were found, showing that poor and ultrarapid *CYP2C19* metabolizers received escitalopram doses above or below the therapeutic interval, respectively.<sup>60</sup> As a consequence, the rate of drug switching, as a proxy for lack of efficacy or adverse reactions, among these phenotype groups, which affected a total of 10% of patients, was 2.3- to 3-fold higher than in extensive metabolizers.

Similar associations were observed among patients receiving the antipsychotics risperidone or aripiprazole with both poor and ultrarapid *CYP2D6* metabolizers being more prone to drug switching

than extensive metabolizers.<sup>61</sup> Interestingly, it was observed that patients carrying reduced activity variants in *CYP2D6* were indeed prescribed lower doses than those with extensive metabolism, despite the fact that the physicians were not aware of the genotype of the patients, indicating that phenotypic peculiarities linked to the drug concentration were noticed by the physician and doses were adjusted accordingly. Interestingly, however, for risperidone these dose adjustments only resulted in 50% of the optimal dose reduction predicted by the *CYP2D6* genotype (Fig. 3).

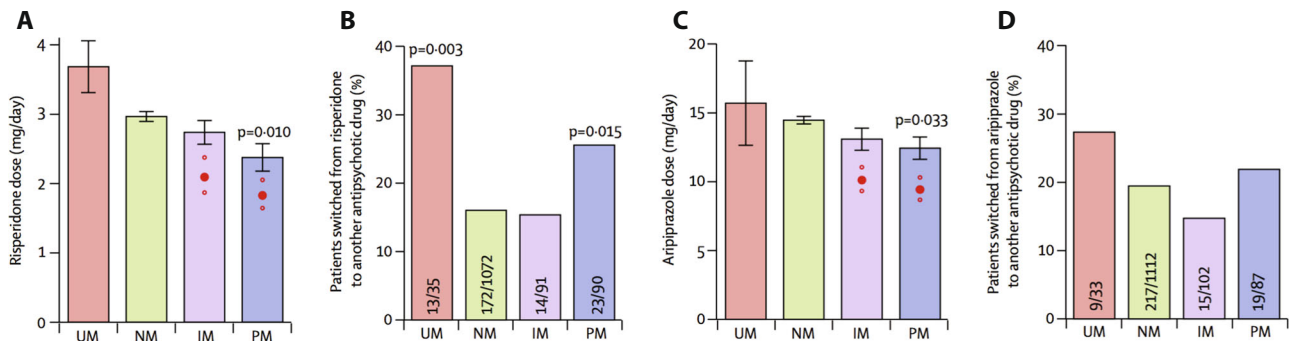
Overall, these results indicate that the retrospective genotyping of large TDM cohorts can indeed provide with clinically useful information regarding gene drug interactions, which in turn can inform genotype-guided dosing in clinical routine.

## CONCLUSIONS

The limited number of sufficiently powered and controlled pharmacogenomic association studies in combination with a lack of functional characterization data for many thousands of genetic variations impair our ability to fully harness a patient's pharmacogenomic information for genotype-guided prescribing. Thus, although progress in NGS technologies has unlocked exciting possibilities to comprehensively profile pharmacogenomic variability, the resulting additional information is currently not clinically actionable. Novel experimental and computational methods hold promise to improve the functional interpretation of pharmacogenomic complexity. In addition, various companies develop high-throughput expression methods for functional characterizations of missense mutations.

Due to impracticalities involved in recruitment of a sufficient number of rare variant carriers and the associated high cost, we do not envision that assessments of the functional importance of rare genetic variants can be conducted using conventional clinical trial designs. Rather, we envision that retrospective analysis of the numerous large TDM-based data sets with thousands of individuals can increase the possibility to validate potential pharmacokinetic phenotype–genotype relationships and to discover novel associations.

Importantly, pharmacogenomic variability differs drastically across ethnicities. Clinically relevant examples include common reduced function variants in *CYP2B6*, *CYP2C8*, and *CYP2C9*, which are up to 20-fold less prevalent in East Asian populations



**Fig. 3 Effect of CYP2D6 genotype status on risperidone and aripiprazole dose and treatment failure rates.** Plots show the associations between CYP2D6 genotype and daily dose and switching frequency for 725 patients treated with risperidone (a, b) and 890 patients treated with aripiprazole (c, d). Considered genotypes include CYP2D6\*3, \*4, \*5, \*6, \*9, \*10, \*41, and functional duplicated alleles \*1xN or \*2xN. In a and c, the dose reductions needed to compensate for the increase of risperidone exposure in patients who were intermediate metabolizers (IM) and poor metabolizers (PM; 95% CI) are indicated by red dots. Error bars indicate SEM. Statistical tests of subgroups are performed using the normal metabolizers (NM) as reference. IM intermediate metabolizers, PM poor metabolizers, UM ultrarapid metabolizers. Figure modified with permission from ref. <sup>61</sup>.

compared with Africans, whereas East Asians carry the highest frequency of decreased function variants in CYP2D6 and CYP1A2.<sup>62</sup> Besides these common well-characterized alleles, rare pharmacogenomic variants show similar patterns of inter-ethnic variability. By analyzing exome data from 6503 individuals with European or African ancestry, we previously found that <9% of all rare variants were found in both populations.<sup>13</sup> Similarly, in an analysis of 141,456 individuals from 7 major human populations, up to 83% of variants in SLC and SLCO genes were restricted to a single population.<sup>11,12</sup> These inter-ethnic differences have important clinical implications, as exemplified by the inter-ethnic differences in anticoagulant and antihypertensive dose requirements,<sup>63</sup> improved gefitinib response rates in East Asians,<sup>64</sup> increased rates of cutaneous ADRs upon carbamazepine in South and East Asians,<sup>53</sup> as well as efavirenz-related ADRs in Zimbabwe.<sup>65</sup> Furthermore, 26 drugs approved by the Food and Drug Administration from 2008 to 2013 carry labels for racial or ethnic differences.<sup>66</sup> These examples demonstrate that consideration of the specific genetic makeup of the population in question in the framework of a precision public health strategy provides appealing opportunities to improve clinical outcomes. African and admixed Latin American populations are especially heterogeneous and are genetically variable, and, as a result, members of these racial and ethnic groups might benefit most from high-resolution pharmacogenomic characterizations and genetically informed prescribing, particularly in those regions of the world where access to individual genomic profiling is limited.

In conclusion, we think that the gap between identification and utilization of genetic variants for truly individualized treatment recommendations will diminish in the future. This is due to more frequent use of large patient cohorts, which allow firm conclusions regarding the clinical importance of genetic variations, in parallel to the impressive development in analytical techniques and computational functionality predictions of novel variations.

Received: 25 September 2019; Accepted: 15 January 2020;  
Published online: 05 March 2020

## REFERENCES

- Klein, M. E., Parvez, M. M. & Shin, J.-G. Clinical implementation of pharmacogenomics for personalized precision medicine: barriers and solutions. *J. Pharm. Sci.* **106**, 2368–2379 (2017).
- Borden, B. A. et al. Assessment of provider-perceived barriers to clinical use of pharmacogenomics during participation in an institutional implementation study. *Pharmacogenet. Genomics* **29**, 31–38 (2019).

- Owusu Obeng, A. et al. Physician-reported benefits and barriers to clinical implementation of genomic medicine: a multi-site IGNITE-network survey. *J. Pers. Med.* **8**, pii: E24 (2018).
- Bush, W. S. et al. Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. *Clin. Pharmacol. Ther.* **100**, 160–169 (2016).
- Klein, K. et al. A new panel-based next-generation sequencing method for ADME genes reveals novel associations of common and rare variants with expression in a human liver cohort. *Front. Genet.* **10**, 56 (2019).
- Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
- Consortium, G. P. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
- Gordon, A. S. et al. Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Hum. Mol. Genet.* **23**, 1957–1963 (2014).
- Fujikura, K., Ingelman-Sundberg, M. & Lauschke, V. M. Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenet. Genomics* **25**, 584–594 (2015).
- Schaller, L. & Lauschke, V. M. The genetic landscape of the human solute carrier (SLC) transporter superfamily. *Hum. Genet.* **7**(Suppl 3), 248–219 (2019).
- Zhang, B. & Lauschke, V. M. Genetic variability and population diversity of the human SLCO (OATP) transporter family. *Pharmacol. Res.* **139**, 550–559 (2019).
- Kozyra, M., Ingelman-Sundberg, M. & Lauschke, V. M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med.* **19**, 20–29 (2017).
- Ahn, E. & Park, T. Analysis of population-specific pharmacogenomic variants using next-generation sequencing data. *Sci. Rep.* **7**, 8416 (2017).
- Wright, G. E. B., Carleton, B., Hayden, M. R. & Ross, C. J. D. The global spectrum of protein-coding pharmacogenomic diversity. *Pharmacogenomics J.* **18**, 187–195 (2018).
- Ingelman-Sundberg, M., Mkrтчian, S., Zhou, Y. & Lauschke, V. M. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* **12**, 26 (2018).
- Santos, M. et al. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* **20**, 622–629 (2018).
- Hauser, A. S. et al. Pharmacogenomics of GPCR drug targets. *Cell* **172**, 41–54.e19 (2018).
- Pendleton, M. et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods* **12**, 780–786 (2015).
- Bowden, R. et al. Sequencing of human genomes with nanopore technology. *Nat. Commun.* **10**, 1869 (2019).
- Ammar, R., Paton, T. A., Torti, D., Shlien, A. & Bader, G. D. Long read nanopore sequencing for detection of HLA and CYP2D6 variants and haplotypes. *F1000Research* **4**, 17 (2015).
- Qiao, W. et al. Long-read single molecule real-time full gene sequencing of cytochrome P450-2D6. *Hum. Mutat.* **37**, 315–323 (2016).
- Buermans, S. J. et al. Flexible and scalable full-length CYP2D6 long amplicon PacBio sequencing. *Hum. Mutat.* **38**, 310–316 (2017).
- Liau, Y. et al. Nanopore sequencing of the pharmacogene CYP2D6 allows simultaneous haplotyping and detection of duplications. *bioRxiv* **116**, 496 (2019).

25. Turner, T. R. et al. Single molecule real-time DNA sequencing of HLA genes at ultra-high resolution from 126 International HLA and Immunogenetics Workshop cell lines. *HLA* **91**, 88–101 (2018).
26. Zhou, Y., Fujikura, K., Mkrtchian, S. & Lauschke, V. M. Computational methods for the pharmacogenetic interpretation of next generation sequencing data. *Front. Pharmacol.* **9**, 1437 (2018).
27. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001).
28. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
29. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688 (2012).
30. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
31. Shihab, H. A. et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2012).
32. Rogers, M. F. et al. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* **34**, 511–513 (2018).
33. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
34. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016).
35. Fujikura, K. Premature termination codons in modern human genomes. *Sci. Rep.* **6**, 22468 (2016).
36. Zhou, Y., Mkrtchian, S., Kumondai, M., Hiratsuka, M. & Lauschke, V. M. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* **19**, 115–126 (2019).
37. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
38. Jain, P. C. & Varadarajan, R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Anal. Biochem.* **449**, 90–98 (2014).
39. Wrenbeck, E. E. et al. Plasmid-based one-pot saturation mutagenesis. *Nat. Methods* **13**, 928–930 (2016).
40. McLaughlin, R. N. Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 (2012).
41. Roscoe, B. P., Thayer, K. M., Zeldovich, K. B., Fushman, D. & Bolon, D. N. A. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* **425**, 1363–1377 (2013).
42. Firnberg, E., Labonte, J. W., Gray, J. J. & Ostermeier, M. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* **31**, 1581–1592 (2014).
43. Melamed, D., Young, D. L., Miller, C. R. & Fields, S. Combining natural sequence variation with high throughput mutational data to reveal protein interaction sites. *PLoS Genet.* **11**, e1004918 (2015).
44. Starita, L. M. et al. Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics* **200**, 413–422 (2015).
45. Fowler, D. M. et al. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).
46. Majithia, A. R. et al. Prospective functional classification of all possible missense variants in PPAR $\gamma$ . *Nat. Genet.* **48**, 1570–1575 (2016).
47. Matreyek, K. A. et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).
48. Gray, V. E., Hause, R. J., Luebeck, J., Shendure, J. & Fowler, D. M. Quantitative missense variant effect prediction using large-scale mutagenesis data. *Cell Syst.* **6**, 116–124.e113 (2018).
49. Sundaram, L. et al. Predicting the clinical impact of human mutation with deep neural networks. *Nat. Genet.* **50**, 1161–1170 (2018).
50. Zhou, J. et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).
51. Boudellioua, I., Kulmanov, M., Schofield, P. N., Gkoutos, G. V. & Hoehndorf, R. DeepPVP: phenotype-based prioritization of causative variants using deep learning. *BMC Bioinformatics* **20**, 65 (2019).
52. McInnes, G. et al. Hubble2D6: a deep learning approach for predicting drug metabolic activity. *bioRxiv* **2**, e195345 (2019).
53. Lauschke, V. M., Zhou, Y. & Ingelman-Sundberg, M. Novel genetic and epigenetic factors of importance for inter-individual differences in drug disposition, response and toxicity. *Pharmacol. Ther.* **197**, 122–152 (2019).
54. Shekhani, R., Steinacher, L., Swen, J. J. & Ingelman-Sundberg, M. Evaluation of current regulation and guidelines of pharmacogenomic drug labels; opportunities for improvements. *Clin. Pharmacol. Ther.* <https://doi.org/10.1002/cpt.1720> (2019). [Epub ahead of print].
55. Pirmohamed, M. et al. A randomized trial of genotype-guided dosing of warfarin. *N. Engl. J. Med.* **369**, 2294–2303 (2013).
56. Luzum, J. A. et al. CYP2D6 genetic variation and beta-blocker maintenance dose in patients with heart failure. *Pharm. Res.* **34**, 1615–1625 (2017).
57. Poulussen, F. C. P. et al. The effect of the CYP2D6 genotype on the maintenance dose of metoprolol in a chronic Dutch patient population. *Pharmacogenet. Genomics* **29**, 179–182 (2019).
58. Fohner, A. E. et al. Assessing the clinical impact of CYP2C9 pharmacogenetic variation on phenytoin prescribing practice and patient response in an integrated health system. *Pharmacogenet. Genomics* **29**, 192–199 (2019).
59. de Oliveira Magalhães Mourão, A. et al. Algorithm for predicting low maintenance doses of warfarin using age and polymorphisms in genes CYP2C9 and VKORC1 in Brazilian subjects. *Pharmacogenomics J.* **5**, 1 (2019).
60. Jukić, M. M., Haslemo, T., Molden, E. & Ingelman-Sundberg, M. Impact of CYP2C19 genotype on Escitalopram exposure and therapeutic failure: a retrospective study based on 2,087 patients. *Am. J. Psychiatry* **175**, 463–470 (2018).
61. Jukić, M. M., Smith, R. L., Haslemo, T., Molden, E. & Ingelman-Sundberg, M. Effect of CYP2D6 genotype on exposure and efficacy of risperidone and aripiprazole: a retrospective, cohort study. *Lancet Psychiatry* **6**, 418–426 (2019).
62. Zhou, Y., Ingelman-Sundberg, M. & Lauschke, V. M. Worldwide distribution of cytochrome P450 alleles: a meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Ther.* **102**, 688–700 (2017).
63. Johnson, J. A. Ethnic differences in cardiovascular drug response: potential contribution of pharmacogenetics. *Circulation* **118**, 1383–1393 (2008).
64. Thomas, S. K. et al. Asian ethnicity as a predictor of response in patients with non-small-cell lung cancer treated with gefitinib on an expanded access program. *Clin. Lung Cancer* **7**, 326–331 (2006).
65. Lauschke, V. M., Milani, L. & Ingelman-Sundberg, M. Pharmacogenomic biomarkers for improved drug therapy—recent progress and future developments. *AAPS J.* **20**, 4 (2017).
66. Ramamoorthy, A., Pacanowski, M. A., Bull, J. & Zhang, L. Racial/ethnic differences in drug disposition and response: review of recently approved drugs. *Clin. Pharmacol. Ther.* **97**, 263–273 (2015).

## ACKNOWLEDGEMENTS

The work in the authors' laboratories is supported by the Swedish Research Council [grant agreement numbers: 2015-02760, 2016-01153, and 2016-01154], as well as by the European Union's Horizon 2020 research and innovation program U-PGx [grant agreement number 668353]. Open access funding provided by Karolinska Institutet.

## AUTHOR CONTRIBUTIONS

Both authors actively contributed to the writing of this manuscript and approved submission of the final version.

## COMPETING INTERESTS

V.M.L. and M.I.-S. are co-founders and shareholders of HepaPredict AB.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to V.M.L.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020