

NEXT-GENERATION SEQUENCING IN PHARMACOGENETICS/GENOMICS

EDITED BY: Ulrich M. Zanger, José A. G. Agúndez and Amit V. Pandey
PUBLISHED IN: Frontiers in Pharmacology and Frontiers in Genetics





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-891-8

DOI 10.3389/978-2-88963-891-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

NEXT-GENERATION SEQUENCING IN PHARMACOGENETICS/GENOMICS

Topic Editors:

Ulrich M. Zanger, Dr. Margarete Fischer-Bosch Institut für Klinische Pharmakologie (IKP), Germany

José A. G. Agúndez, University of Extremadura, Spain

Amit V. Pandey, University of Bern, Switzerland

Citation: Zanger, U. M., Agúndez, J. A. G., Pandey, A. V., eds. (2020). Next-Generation Sequencing in Pharmacogenetics/genomics. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88963-891-8

Table of Contents

- 05 *Pharmacogenetic Variation in Over 100 Genes in Patients Receiving Acenocumarol***
Vanessa Gonzalez-Covarrubias, Javier Urena-Carrion, Beatriz Villegas-Torres, J. Eduardo Cossio-Aranda, Sergio Trevethan-Cravioto, Raul Izaguirre-Avila, O. Javier Fiscal-López and Xavier Soberon
- 14 *Multiplexed Nanopore Sequencing of HLA-B Locus in Māori and Pacific Island Samples***
Kim N. T. Ton, Simone L. Cree, Sabine J. Gronert-Sum, Tony R. Merriman, Lisa K. Stamp and Martin A. Kennedy
- 26 *PASSPORT-seq: A Novel High-Throughput Bioassay to Functionally Test Polymorphisms in Micro-RNA Target Sites***
Joseph Ipe, Kimberly S. Collins, Yangyang Hao, Hongyu Gao, Puja Bhatia, Andrea Gaedigk, Yunlong Liu and Todd C. Skaar
- 36 *Development of an AmpliSeq™ Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain***
Dario Kringel, Mari A. Kaunisto, Catharina Lippmann, Eija Kalso and Jörn Lötsch
- 58 *Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data***
Yitian Zhou, Kohei Fujikura, Souren Mkrtchian and Volker M. Lauschke
- 75 *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***
Kathrin Klein, Roman Tremmel, Stefan Winter, Sarah Fehr, Florian Battke, Tim Scheurenbrand, Elke Schaeffeler, Saskia Biskup, Matthias Schwab and Ulrich M. Zanger
- 90 *Next-Generation Sequencing of PTGS Genes Reveals an Increased Frequency of Non-synonymous Variants Among Patients With NSAID-Induced Liver Injury***
María Isabel Lucena, Elena García-Martín, Ann K. Daly, Miguel Blanca, Raúl J. Andrade and José A. G. Agúndez
- 98 *Genetic Association of Olanzapine Treatment Response in Han Chinese Schizophrenia Patients***
Wei Zhou, Yong Xu, Qinyu Lv, Yong-hui Sheng, Luan Chen, Mo Li, Lu Shen, Cong Huai, Zhenghui Yi, Donghong Cui and Shengying Qin
- 105 *Actionable Pharmacogenetic Variation in the Slovenian Genomic Database***
Keli Hočevár, Aleš Maver and Borut Peterlin
- 116 *Integrating Next-Generation Sequencing in the Clinical Pharmacogenomics Workflow***
Efsthia Giannopoulou, Theodora Katsila, Christina Mitropoulou, Evangelia-Eirini Tsermpini and George P. Patrinos

122 *Star Allele-Based Haplotyping versus Gene-Wise Variant Burden Scoring for Predicting 6-Mercaptopurine Intolerance in Pediatric Acute Lymphoblastic Leukemia Patients*

Yoomi Park, Hyery Kim, Jung Yoon Choi, Sunmin Yun, Byung-Joo Min, Myung-Eui Seo, Ho Joon Im, Hyoung Jin Kang and Ju Han Kim

132 *Identification of Novel Biomarkers for Drug Hypersensitivity After Sequencing of the Promoter Area in 16 Genes of the Vitamin D Pathway and the High-Affinity IgE Receptor*

Gemma Amo, Manuel Martí, Jesús M. García-Menaya, Concepción Cordobés, José A. Cornejo-García, Natalia Blanca-López, Gabriela Canto, Inmaculada Doña, Miguel Blanca, María José Torres, José A. G. Agúndez and Elena García-Martín



Pharmacogenetic Variation in Over 100 Genes in Patients Receiving Acenocumamol

Vanessa Gonzalez-Covarrubias^{1*}, Javier Urena-Carrion¹, Beatriz Villegas-Torres¹, J. Eduardo Cossío-Aranda², Sergio Trevethan-Cravioto², Raul Izaguirre-Avila², O. Javier Fiscal-López² and Xavier Soberon¹

¹ Instituto Nacional de Medicina Genómica, Mexico City, Mexico, ² Instituto Nacional de Cardiología, Mexico City, Mexico

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Marcelo Rizzatti Luizon,
Universidade Federal de Minas Gerais,
Brazil
Wendy Hernandez,
University of Chicago, United States

*Correspondence:

Vanessa Gonzalez-Covarrubias
vgonzalez@inmegen.gob.mx

Specialty section:

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 14 September 2017

Accepted: 09 November 2017

Published: 23 November 2017

Citation:

Gonzalez-Covarrubias V,
Urena-Carrion J, Villegas-Torres B,
Cossío-Aranda JE,
Trevethan-Cravioto S,
Izaguirre-Avila R, Fiscal-López OJ and
Soberon X (2017) Pharmacogenetic
Variation in Over 100 Genes in
Patients Receiving Acenocumamol.
Front. Pharmacol. 8:863.
doi: 10.3389/fphar.2017.00863

Coumarins are widely prescribed worldwide, and in Mexico acenocumamol is the preferred form. It is well known that despite its efficacy, coumarins show a high variability for dose requirements. We investigated the pharmacogenetic variation of 110 genes in patients receiving acenocumamol using a targeted NGS approach. We report relevant population differentiation for variants on *CYP2C8*, *CYP2C19*, *CYP4F11*, *CYP4F2*, *PROS*, and *GGCX*, *VKORC1*, *CYP2C18*, *NQO1*. A higher proportion of novel-to-known variants for 10 genes was identified on 41 core pharmacogenomics genes related to the PK (29), PD (3), of coumarins, and coagulation proteins (9) including, *CYP1A1*, *CYP3A4*, *CYP3A5*, and *F8*, and a low proportion of novel-to-known variants on *CYP2E1*, *VKORC1*, and *SULT1A1/2*. Using a Bayesian approach, we identified variants influencing acenocumamol dosing on, *VKORC1* (2), *SULT1A1* (1), and *CYP2D8P* (1) explaining 40–55% of dose variability. A collection of pharmacogenetic variation on 110 genes related to the PK/PD of coumarins is also presented. Our results offer an initial insight into the use of a targeted NGS approach in the pharmacogenomics of coumarins in Mexican Mestizos.

Keywords: coumarins, acenocumamol, pharmacogenomics, Mexican Mestizo, population differences, targeted sequencing

INTRODUCTION

Anticoagulants such as warfarin, acenocumamol, and phenprocoumon act as inhibitors of vitamin K reducing enzymes, which regenerate vitamin K, a cofactor for several clotting proteins. Acenocumamol is the coumarin of choice in Europe and Latin America (Ufer, 2005). The heterogeneity in coumarins efficacy, safety, and dosing has been partly explained by clinical, demographic, and genetic parameters. Several polymorphisms on *CYP2C9*, *VKORC1*, and *CYP4F2* can account for about 40–50% on coumarin dose differences (Scott et al., 2014). However, most studies have been performed for warfarin and in populations other than Mexican. Acenocumamol (4-nitrowarfarin) is the most commonly prescribed oral coumarin in the public health care system in Mexico. In contrast to warfarin, the more potent isomer, S-acenocumamol, is rapidly eliminated and the drug's therapeutic effect is most likely due to R-acenocumamol. The R isomer is metabolized by several members of the cytochrome P-450 family including, *CYP2C9*, *CYP2C19*, *CYP2C8*, *CYP2C18*, *CYP3A4*, *CYP1A1*, and *CYP1A2* (Tassies et al., 2002; Ufer, 2005). Hence, genetic variation on genes coding for these proteins should putatively influence acenocumamol dosing.

The collection of pharmacogenetic variation in Mexican populations is still scarce (Fricke-Galindo et al., 2016). Reports indicate that some of the actionable markers on *VKORC1* and *UGT1A1* present significant population differences (Bonifaz-Peña et al., 2014), suggesting the existence of variants with distinctive allele frequency in these populations potentially influencing drug response.

Endeavors are currently ongoing to amass a more comprehensive picture of the pharmacogenetic variation in Mexican Mestizos. Here, we investigated genetic variation in over 100 genes by targeted NGS in patients receiving acenocumarol, including genes involved in general pharmacokinetics and pharmacodynamics, vitamin K recycling, and coagulation proteins, these latter also potentially affecting acenocumarol response (Allan et al., 2005; Harrington et al., 2008; Carcao et al., 2015; Tong et al., 2016). Clinical and genetic data were used to develop an algorithm to explain dose variability in this group of patients.

Genomic data analyses provided with a collection of pharmacogenetic variation for this population. This approach hints to toward the consideration of multiple variants to assess acenocumarol dosing for an individualized dose assessment.

MATERIALS AND METHODS

Participants and DNA extraction. The National Institute of Cardiology in Mexico City prescribes acenocumarol on a regular basis mostly after stroke, stent implants, or for thrombosis. A hundred and fifty patients treated with acenocumarol between 2006 and 2010 were surveyed and monitored for acenocumarol efficacy through at least three consecutive INR measurements. Of these, 103 blood samples were available for DNA extraction using the DNeasy Blood & Tissue kit (Qiagen, Valencia CA, USA) from a routine blood sample in EDTA-Vacutainer collection tubes, sample characteristics are depicted in **Table 1**. All participants gave written informed consent according with the Declaration of Helsinki. The project was reviewed and approved by The Research and Ethics Committees at The National Institute of Cardiology and The National Institute of Genomic Medicine (INMEGEN) Mexico City, project approval 25/2016/I.

Next Generation Sequencing

We investigated coding, 25 bp of adjacent introns, and 5' and 3' UTR regions of 110 genes related to general pharmacogenomics including core pharmacokinetics and pharmacodynamics targets in 100 DNA samples. We selected these genes according to the general PGx and the PK/PD of acenocumarol by searching the available literature using the keywords, pharmacogenetics, pharmacogenomics, acenocumarol, coumarin pharmacokinetics,

Abbreviations: AIC, Akaike information criteria; BIC, Bayesian information criterion; DIC, deviance information criterion; FDR, false discovery rate; Fst, fixation index as a measure of genetic variance; GLM, generalized linear model; BMI, body mass index; INR, international normalized ratio based on prothrombin time; LD, linkage disequilibrium; MAF, minor allele frequency; NGS, next generation sequencing; PD, pharmacodynamics; PharmGKB, pharmacogenomics knowledge base; PK, pharmacokinetics; SNV, single nucleotide variant; SNP, single nucleotide polymorphism.

TABLE 1 | Patient demographics.

	Males	Females	P-value
N	46	54	
Age (y)	54 (17–84)	56 (26–85)	>0.05
Weight (kg)	76 (54–121)	64 (37–90)	<0.05
Height (m)	1.7 (1.5–1.8)	1.5 (1.4–1.7)	<0.05
INR	2.8 (2.1–3.4)	2.8 (2.0–3.9)	>0.05
Dose (mg/wk)	15.4 (3.2–56)	15.0 (4.0–52)	>0.05
Mo under treatment	12 (4–60)	13 (6–56)	>0.05

and pharmacodynamics (van Leeuwen et al., 2008; Soria et al., 2009; Whirl-Carrillo et al., 2012; Tong et al., 2016). Regions of interest were captured using a Haloplex custom Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) defined for 150 × 2 paired-end reads, in a panel size of 1.1Mbp. In addition, we included a set of 360 ancestry informative markers (AIMS) to assess genetic admixture using SNPs from the HapMap database for CEU and YRI populations, and Natives from Mexico (Galanter et al., 2012). Sequencing libraries were generated according to the manufacturer's protocol (version D.5, May 2013). Briefly, all 100 DNA samples (225 ng) were digested with 8-paired restriction enzymes, fragmentation pattern was analyzed in a 2100 Biolalyzer (Agilent Technologies). DNA fragments were hybridized with the Haloplex synthetic probes, adapters were ligated followed by PCR amplification for library enrichment. Library quality for fragment size and molarity was also performed using 2100 Biolalyzer information. Samples were pooled and sequenced in a Genome Analyzer II (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Targeted genes are listed in Supplemental Table ST1.

Bioinformatic and Statistical Analyses

Sequence reads were processed according to the Broad Institute recommended best practices workflow and the Genome Analysis ToolKit (GATK) (Acland et al., 2013; Van der Auwera et al., 2013). Briefly, paired-end reads were trimmed to remove adapters and low quality regions using Trimmomatic (Bolger et al., 2014), reads with an average Q ≤ 30 were discarded, followed by elimination of reads shorter than 36 nucleotides. Mapping and alignment of sequencing reads were performed with BWA, Samtools, and Picard using the hg19 human genome reference (dbSNP build 137) (McKenna et al., 2010). Base quality score calibration and single nucleotide variant (SNV) calling were assessed using GATK v3.3. Variants were confirmed visually in the integrative genomic viewer, IGV (Robinson et al., 2011), and their functional impact was annotated using SnpEff, and ranked as low, moderate, modifier, or high (Sherry et al., 2001; Cingolani et al., 2012; Exome Variant Server¹).

The data analysis toolset, PLINK was used to determine descriptive statistics, allele frequencies, Hardy-Weinberg, and

¹Exome Variant Server. [updated 2014/01/02/10:04:37]. Available online at: <http://evs.gs.washington.edu/EVS>

population differentiation, this latter was assessed by determining the measure of genetic variance in this subpopulation relative to population variance in other continental groups using the F_{ST} statistic. A threshold value of $P < 0.05$ after FDR was considered as statistically significant (Purcell et al., 2007).

Pharmacogenetic Model

We utilized a Bayesian statistical approach to incorporate genetic variants to an algorithm for acenocumaryl dose estimation (Sebastiani et al., 2009; Chen et al., 2012). The rejection of the null hypothesis (lack of association between acenocumaryl dose and genetic variants) was based on probabilities of stochastic computations of Markov Chain Monte Carlo methods (MCMC). Also, we tested the association between dose and all SNV alone or in combination including those previously identified via single-SNP analysis ($P > 0.05$, FDR). We considered 4614 SNPs and 815 variant interactions for model development. This strategy allows for the identification of independent genetic variants or those that depend on each to influence dose variation. Variants were considered with a MAF < 0.95 and > 0.05 , and interactions between variants within the same gene with frequencies < 0.9 and > 0.10 , i.e., that two or more variants in a gene may have adding or balancing effects on the dose. First, we used a Bayesian Generalized Linear model for variable selection to obtain the posterior probability of a gene variant affecting acenocumaryl dose, then we used Bayes Factors, a form of Bayesian hypothesis tests, to prioritize a set of models, and then we evaluated the selected models through Deviance Information Criterion (DIC), a measure of model selection related to AIC and BIC criteria, commonly used in Bayesian hierarchical models. Briefly, for the former, we used a gamma likelihood function with logarithmic link function, variance τ and mean conforming to Equation (1), where $v_{1,j}$ and $v_{2,j}$ represent binary variables for each genotype of a SNP j , refers to interactions between variants, G is a set of genes g , and n_g the number of SNPs in g ; v_{r,j_g} represents genotype r of SNP j in gene g , and $\beta_{j_g,k_g}^{r_1,r_2}$ represents the effect size of genotypes v_{r_1,j_g} and v_{r_2,k_g} ; I_{j_g,k_g} and $I_{j,k}$ are binary variables for the inclusion or exclusion of SNPs and SNP interactions, and x_i and α_i represent m non-genetic covariates including age, sex, BMI, and height.

$$\begin{aligned} \log(E[y|\theta]) = & c + \sum_i^m \alpha_i x_i + \sum_j^n I_{j_g}(\beta_{j_g,k_g}^{r_1,r_2} v_{r_1,j_g} + \beta_{j_g,k_g}^{r_2,r_1} v_{r_2,k_g} \\ & + \sum_{g \in G} \sum_{j_g < k_g}^{n_g} I_{j_g,k_g}(\beta_{j_g,k_g}^{1,1} v_{1,j_g} v_{1,k_g} + \beta_{j_g,k_g}^{1,2} v_{1,j_g} v_{2,k_g} \\ & + \beta_{j_g,k_g}^{2,1} v_{2,j_g} v_{1,k_g} + \beta_{j_g,k_g}^{2,2} v_{2,j_g} v_{2,k_g}) \end{aligned} \quad (1)$$

Where, $\tau \sim \text{Gamma}(\lambda, \kappa)$, $c \sim \text{Normal}(0, \tau_c^2)$, $c \sim \text{Normal}(0, \tau_c^2)$, $\beta_{j,k} \sim \text{Laplace}(0, \tau_\beta^2)$, $\beta_{j_g,k_g}^{r_1,r_2} \sim \text{Laplace}(0, \tau_\beta^2)$, $I_j \sim \text{Bernoulli}(\pi)$, $I_{j_g,k_g} \sim \text{Bernoulli}(\pi)$, $\pi \sim U(a, b)$, and $\alpha_i \sim \text{Normal}(0, \tau_\alpha^2)$

Next, we standardized clinical variables for mean zero and unitary variance, and using JAGS 4.1.0 and R 3.2.0 we obtained MCMC from the posterior distribution. We ran five chains of

110,000 iterations each, including a burn-in period of 10,000 iterations and random initial values, convergence was verified via the Gelman-Rubin statistic $\hat{R} < 1.2$, followed by a series of Bayes Factors to condition on the presence or absence of variants, branching them into a decision tree, as part of the pharmacogenetic model development. Further details on the model development were included in Supplemental Table ST2.

RESULTS

Demographic characteristics stratified by sex are presented in Table 1. Bioinformatic analyses revealed 5108 variants in 110 genes in 100 DNA samples with an average depth of 250x and $> 98\%$ coverage, but a wide range was registered depending on the gene (30x–600x, 80–100%). These 110 genes represent less than 1% of the coding genome, approximately 25% of a pharmacogenome, more than half of the Coriell reference list for pharmacogenomics, and include $> 20\%$ of actionable pharmacogenetic markers listed by CPIC (Pratt et al., 2010; Relling and Klein, 2011). After quality control, variant calling, and annotation, 4290 SNVs were utilized for statistical analyses (ST3). There was a complete agreement between genotypes of variants assessed by NGS and allele discrimination performed for *CYP2C9**2,*3, and *5, *CYP4F2* rs2108622, and *VKORC1* rs9934438. Admixture analysis with 314 ancestry informative markers showed an average population structure of 50–92% Mexican Native and 6–54% Caucasian (CEU), all individuals showed less than 5% of Sub-Saharan African (YRI) admixture.

Of these 4290 SNVs, 28% have not been reported before (1237 without an rs identifier) and 274 of these novel variants had a minor allele frequency (MAF $> 1\%$). On average, each individual showed 908 SNVs, 534 heterozygous and 374 homozygous, of which 258 were present per individual (Table 2). Four-hundred and seven variants in 65 genes did not suffice the equilibrium of Hardy-Weinberg (9.8%, Supplemental Table ST4).

Variants were classified by SNPEff according to their to their *in-silico* functional impact as high, moderate, modifier, or low (Cingolani et al., 2012). We listed a total of 36 known SNVs (27 heterozygous and 9 homozygous) with a high functional impact (Table 2 and ST3). These resequencing descriptive statistics seem to compare to other reports (Waldron, 2016).

Pharmacogenetic Variation

The F_{ST} statistic was assessed to evaluate genetic differentiation between Mexican Mestizos and three major continental populations, Chinese Han from Beijing (CHB), Yoruba from Ibadan, Nigeria (YRI), and Europeans from Utah, USA (CEU) utilizing the 1000 genomes database. YRI showed the largest differentiation with 377 variants with a F_{ST} value above 0.25, followed by CEU (51 variants with $F_{ST} > 0.25$), and CHB (32 variants with $F_{ST} > 0.25$, ST5). $F_{ST} > 0.25$ values were identified for variants on several genes related to the coagulation cascade or coumarin metabolism. For example, when comparing to Caucasians we found high population differentiation for

TABLE 2 | Summary of NGS genetic variation.

	N	Observations
Total SNVs	4,290	in 100 individuals
SNVs per individual	908	534 heterozygous, 374 homozygous
Coding	3,335	
UTRs	471	
Non coding	484	
Known SNVs	3,053	650 per individual (384 heterozygous and 266 homozygous)
Novel SNVs	1,237	258 per individual (150 heterozygous and 108 homozygous)
Novel SNVs MAF > 1%	274	Present in 56 genes including 19CYPs, 3 UGTs, <i>SULT1A2</i>
Heterozygous SNVs	2,463	1660 known and 803 novel
Heterozygous, high impact	27	8 novel variants on 8 genes including <i>CYP3A4</i> and <i>CYP2B6</i> , and 19 known variants in 9 CYPs, 2 UGTs, <i>SULT1A1</i> , and 5 coagulation proteins
Heterozygous, moderate impact	325	106 novel in 56 genes including 12 CYPs, 5 UGTs, and <i>SULT1A2</i> , 219 known variants in 69 genes, including, 14 CYPs, 6 UGTs, and <i>SULT1A2</i>
Total homozygous	1,110	Known and novel variants
Homozygous, high impact	9	Present on 4 CYPs, including, <i>CYP2B6</i> POS.4151213, <i>CYP2C9</i> rs114071557, and on <i>UGT2B7</i>
Homozygous, moderate impact	129	36 novel variants in 25 genes (7 CYPs, 3 UGTs) including <i>CYP4F2</i> POS15996832, <i>CYP2D6</i> POS42522724, and 93 known variants in 33 genes (12 CYPs, 5 UGTs, and 5 drug targets)

HGVS, Human Genome variation society nomenclature; C, coding region; P, protein. The 1000 Genomes Project Consortium et al. (2015). POS indicates the position on the chromosome for novel variants.

variants on *CYP2C8*, *CYP2C19*, *CYP4F11*, *CYP4F2*, *PROS*, and *GGCX*. Comparing to CHB, differences arose on *CYP2C8*, *VKORC1*, *2C18*, *NQO1*, and for YRI differences were observed on major CYPs, FMOs, *F13B*, *F8*, *PROS*, and *SERPINA10* among others (ST5). Allele frequency comparisons between Mexicans from the 1000 genome project (MXL) and those in this study showed similar F_{ST} values for most variants, except for *CYP2C18* rs2281889, *CYP2C8* rs1891071, *CYP4F2* rs309319, and *CYP4F11* rs11086013, for which we observed F_{ST} values between 0.15 and 0.33.

We analyzed allele frequency variation for 30 major pharmacogenes and 10 genes related to the coagulation cascade. The largest number of variants per gene was observed on *SULT1A1*, *CYP2E1*, *CYP1B1*, *CYP3A4*, *CYP3A5*, *F5*, and *F11*. Interestingly, *CYP1A2*, *CYP3A4*, *CYP3A5*, *CYP1A6*, *F11*, *F13B*, and *F8* showed a large proportion of novel variants compared to known variants. Genes with significantly fewer variants were, *CYP1A3*, *CYP1A5*, *CYP1A9*, and *F9*, the three former did not show any novel variants (ST3). Next, we assessed the presence of known and novel variant considering a MAF >5% and a predicted functional impact as high or moderate in these genes. For known variants, we list 7 with a high predicted functional impact, 2 on UGTs and one on *SULT1A1*, *CYP2C9*, *CYP2C19*, and *CYP2C8* (Table 3).

Novel variants were observed on 37 of these 40 genes in counts from 1 on *UGT1A1*, to 23 on *CYP3A5*, 25 in *F5*, and 26 in *F8*. Of these, 2 showed a high functional impact predicting a stop codon on *CYP3A4* and *CYP2B6*, a moderate impact was reported for 16 variants on 14 genes (Table 3). The proportion of novel-to-known variants and its functional impact for these pharmacogenes is represented in Figure 1.

Acenocumarol Pharmacogenetic Model

We developed a pharmacogenomic model to predict acenocumarol dose, using a Bayesian approach that included all SNP variants and the interaction among those on the same gene. We fitted the Bayesian GLM through five MCMC chains where genetic variants were prioritized by their posterior inclusion probability. The higher the posterior probability of a variant, the larger its influence on dose. Figure 2 shows a hierarchical tree indicating an ordered relevance of variants from *VKORC1*, *CYP2D8P*, and *SULT1A1*, followed by those on *CYP4F12*, *F13B*, and *F8*. Values of posterior probability for all variants are listed in ST7 and ST8.

The dosing algorithm accounted for age, sex, weight, and height. The interaction between *VKORC1* variants rs8050894 and rs9934438 which are in LD ($R = 0.492$), showed the highest posterior inclusion probability (mean, 0.96), followed by a novel variant, on *CYP2D8P* (POS42547668), and variants on *SULT1A1* rs11648192, and *CYP2C8* rs1058932, and rs2275620. Unfortunately, pharmacogenetic variant *VKORC1* rs9923231 did not pass NGS quality controls thus, it was not modeled, but it is in complete LD with rs9934438, which was included in these analyses (Rieder et al., 2005).

Final model evaluation, we used R to implement the series of Bayes Factors as described in Supplemental Table ST2 with a cut-off value, $c_j = 3 + m_j$, where m_j is the number of variants conditioned to be absent from the model, and 3 as a minimum cut-off value based on Harold Jeffrey scale of interpretation for Bayes Factors (Baldi and Long, 2001). We selected two models according to the lowest DIC values, the first one included SNVs interactions, *VKORC1* rs8050894 and rs9934438 and variants on *SULT1A1* and *CYP2D8P* (ST7). The second model excluded variant interactions. Modeled variants and clinical parameters (age, sex, weight, and height) explain up-to 55.9% of dose

TABLE 3 | Novel and known variants on relevant pharmacogenes at MAF > 5%.

Gene	Chr	ID	Predicted functional impact	HGVS_C	HGVS_P
<i>UGT1A10</i>	2	rs200144439	High	c.1533C>A	p.Tyr511*
<i>UGT2B7</i>	4	rs11935951	High	c.–159+1G>T	Intron/splicing
<i>CYP2C19</i>	10	rs28399504	High	c.1A>G	p.Met1?
<i>CYP2C9</i>	10	rs114071557	High	c.1A>G	p.Met1?
<i>CYP2C8</i>	10	rs181982392	High	c.1198G>T	p.Glu400*
<i>SULT1A1</i>	16	rs144422872	High	c.307C>T	p.Arg103*
NOVEL VARIANTS					
Gene	Chr	Position	Predicted functional impact	HGVS_C	HGVS_P
<i>F5</i>	1	169530002	Moderate	c.376G>T	p.Ala126Ser
<i>F10</i>	13	113803697	Moderate	c.1333C>T	p.Arg445Cys
<i>UGT1A10</i>	2	234676866	Moderate	c.1076G>T	p.Gly359Val
<i>UGT1A10</i>	2	234676884	Moderate	c.1094C>A	p.Ala365Asp
<i>SULT1A2</i>	16	28603787	Moderate	c.473C>A	p.Thr158Asn
<i>CYP3A5</i>	7	99250198	Moderate	c.1231C>T	p.Pro411Ser
<i>CYP3A4</i>	7	99358497	Moderate	c.1361T>C	p.Leu454Pro
<i>CYP3A4</i>	7	99365982	High	c.665C>A	p.Ser222*
<i>CYP3A4</i>	7	99377662	Moderate	c.118G>T	p.Gly40Trp
<i>CYP2C19</i>	10	96522531	Moderate	c.69C>A	p.Ser23Arg
<i>CYP2C8</i>	10	96800713	Moderate	c.896T>A	p.Phe299Tyr
<i>CYP2C8</i>	10	96802829	Moderate	c.967G>T	p.Val323Phe
<i>CYP2E1</i>	10	135352449	Moderate	c.1463G>T	p.Cys488Phe
<i>CYP1A2</i>	15	75042614	Moderate	c.535C>A	p.Leu179Met
<i>CYP4F2</i>	19	15996832	Moderate	c.1017G>T	p.Trp339Cys
<i>CYP2B6</i>	19	41518213	High	c.975C>A	p.Tyr325*
<i>CYP2D6</i>	22	42522724	Moderate	c.1346C>A	p.Ala449Asp
<i>NQO1</i>	16	69745184	Moderate	c.520A>C	p.Ser174Arg

POS, position on the chromosome for novel variants. p.Met1?, change in the translation of the initiation codon with unknown effect. *Insertion of a termination codon.

variation for this study group. Values of high density intervals (HDI) are presented considering for 95% posterior probability (ST7 and ST8). The addition of additional variants to the model increased DIC values significantly which translates into a decreasing impact on acenocumarol dose (Table 4).

Pharmacogenetic Model Considering Variant Interactions

$\ln \text{Dose} = -0.6935 - 0.0071 \cdot \text{age (y)} + 0.0035 \cdot \text{weight (kg)} - 0.1136 \text{ (if male)} + 1.0709 \cdot \text{height (m)} - 0.213 \text{ (if VKORC1 rs8050894 is C/G and rs9934438 is G/A)} - 0.719 \text{ (if VKORC1 rs8050894 is G/G and rs9934438 is A/A)} + 0.899 \text{ (if CYP2D8P POS.42547668 is T/C)} + 0.203 \text{ (if SULT1A1 rs11648192 is C/T)}$. Variance explained 55.9%.

Pharmacogenetic Model without Variant Interactions

$\ln \text{Dose} = -0.5846 - 0.0069 \cdot \text{age (y)} + 0.0045 \cdot \text{weight (kg)} - 0.0945 \text{ (if male)} + 0.9795 \cdot \text{height (m)} - 0.239 \text{ (if VKORC1 rs9934438 is G/A)} - 0.529 \text{ (if VKORC1 rs9934438 is A/A)} + 1.092 \text{ (if CYP2D8P POS. 42547668 is T/C)} + 0.188 \text{ (if SULT1A1 rs11648192 is C/T)}$. Variance explained 40.0%.

Finally, we used these models to recalculate acenocumarol dose in patients with an INR 2–3 receiving a stable dose.

Pharmacogenetic dose calculations approached given acenocumarol doses for all INR-sable patients ($P > 0.05$) except for one patient, R4 who needed 5.4 mg/day, and models estimated 2.47 mg/day. All individual dose estimations were listed in ST9.

DISCUSSION

Here, we investigated pharmacogenetic variation in 110 genes by targeted NGS in patients treated with acenocumarol.

Novel Variants

Genetic variation analyses showed that the presence of novel variants varied widely among genes. For example, the largest number of novel variants (≥ 20) was observed on *CYP3A4*, *CYP3A5*, *SULT1A1*, *GGCX*, *F11*, *F5*, *F8*, and *F9*. High functional impact variants were present on *CYP2B6* (MAF, 50%), *CYP2C8* (MAF 1%), *CYP3A4* (MAF 8%), *F5* (MAF 2%), and *VKORC1* (MAF 1%). These are relevant for its allele frequency, the dozens of drugs they metabolize, and because their impact predicts a stop codon. *CYP2B6* and *CYP3A4* are among the most polymorphic genes thus, it is not surprising the presence of relevant novel variants. Similarly, for *VKORC1* population differentiation has been previously reported and the presence of a novel variant

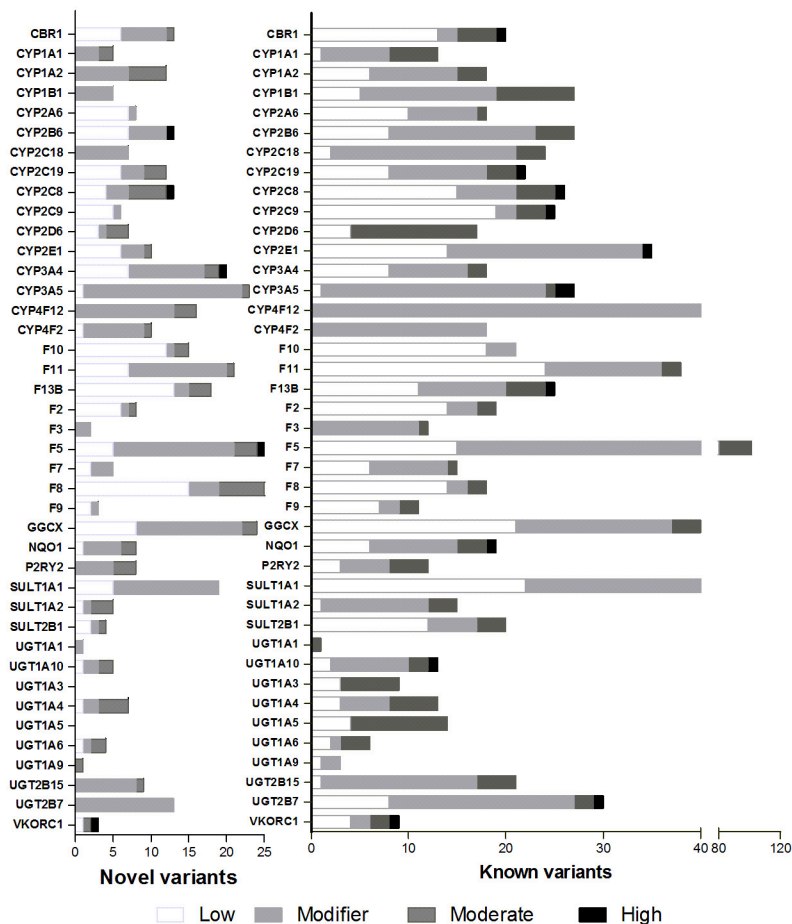


FIGURE 1 | Proportion of novel and known variants. Functional impact determined *in-silico* is depicted in white for low impact, light gray for modifier, dark gray for moderate impact, and black for high impact.

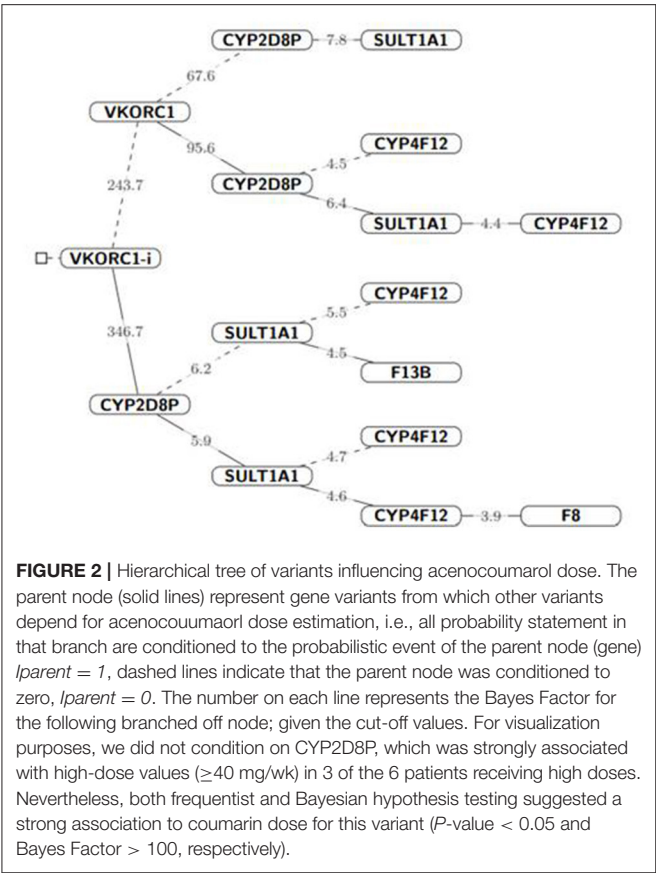
with a high functional impact may be in part, a consequence of this stratification (Bonifaz-Peña et al., 2014). Novel variants on major metabolizing genes, *CBR1*, *CYP1A2*, *CYP3A4*, *CYP3A5*, *P2RY2*, and *UGT1A6* represented 40–50% of novel and known variants, suggesting that the collection of variation on these genes is probably not yet complete in Mestizos. Other metabolizing genes, *CYP1B1*, *CYP2C18*, *CYP2C19*, *UGT1A* members, *CYP2E1*, *SULT1A2*, and *SULT1A1* showed a low proportion of novel-to-know variants (Figure 1). This may not necessarily mean that genetic variation is complete for these genes. For example, *SULT1A1* presented 63 known and 19 novel variants ranking this gene as second with the largest number of variants.

Also, we confirmed population differences previously reported with an $F_{st} > 0.19$, on *VKORC1* rs9934438 and four variants on *UGT2B15* when comparing to YRI and CHB (Bonifaz-Peña et al., 2014). Differences between Mestizos and CEU were observed for *UGT2B15*, *CYP2E1*, *CYP1A2*, *CYP4F2*, *UGT2B7*, *F12*, and *F12* (ST5). Interestingly, a few variants showed an $F_{st} > 0.20$ between Mestizos from this study and Mexicans from Los Angeles (MXL) from the 1000 genome project, on *CYP2C8*, *CYP2C18*, and *CYP4F2* relevant for the pharmacokinetics of coumarins, phenytoin, vitamin K,

and lipids. Allele frequencies of all variants are listed on ST6. Observations on these relevant pharmacogenes highlight the need to for a cautious implementation of pharmacogenomics in Mexican Mestizos.

We developed a pharmacogenetic model to estimate acenocumarol dose testing over four thousand variants. The model considered relevant variants on, *SULT1A1*, *CYP2D8P*, and *VKORC1*. For the latter gene, variants, rs8050894 and rs9934438, are well-known pharmacogenetic markers of coumarin dosing with the highest PharmGKB level of evidence. The interaction of these SNPs has already been reported as part of a haplotype (CG vs. TA) that aids to classify patients into high and low dose requirements (Rieder et al., 2005).

Interestingly, the model did not associate variants on *CYP4F2* or *CYP2C9* with acenocumarol dose (Table 4). Maybe because *CYP4F2* (rs2108622) has a lower impact (Danese et al., 2012), and R-acenocumarol is metabolized by several CYPs other than *CYP2C9*. Moreover, reported variants that impair *CYP2C9* activity are present in low frequency in Mexican Mestizos (Villegas-Torres et al., 2015). Instead, we observed dose association with variants on *SULT1A1* and *CYP2DP8*. *CYP2D8P* is a pseudogene in the *CYP2D* cluster comprising *CYP2D6*,



CYP2D7P, and *CYP2D8P* the former known to metabolize around 25% of all prescribed drugs. This cluster seems to have rapidly evolved due to environmental adversity with ethnic differences (Heim and Meyer, 1992). Wang et al identified a *CYP2D6* transcription enhancer in the *CYP2D* cluster supporting the consideration of variants outside the *CYP2D6* loci for functional genotyping (Wang et al., 2015). We identified a new variant on *CYP2D8P* POS.42547668, 26 Kbp upstream *CYP2D6*, and although there is no xenobiotic metabolism reported for this pseudogene, we can speculate that this variant is in LD with another one affecting gene expression or drug metabolism. The inclusion of many variants to dissect a pharmacogenetic phenotype is becoming more common as it increases our knowledge in paths and network interactions not previously considered (Cruz-Correa et al., 2017; Oliveira-Paula et al., 2017).

Our model is similar to others in that it includes typical clinical variants (age, sex, weight, and height), a dose prediction around 50% confidence, and the inclusion of *VKORC1* as the primary determinant of acenocoumarol dose. And even though we report *VKORC1* rs9934438 vs. *VKORC1* 9923231 this latter, most commonly studied (Zhang et al., 2015; Tong et al., 2016) these are in complete LD. Finally, dose assessment using this model closely approached the dose received to achieve an INR 2–3, except for patient R4. Therefore, we delved into the genetic variability of this sample observing 20 heterozygous and 9 homozygous variants, these latter on *SULT1A1*, *FGB*, *CDH12*, *KCNJ6*, *CBR3*, and *CYP2E1*. However, this variation does not necessarily explain a lack of dose prediction. We can speculate that it is the presence

TABLE 4 | Pharmacogenomic model parameters.

	Genotype	μ	HDI	Bayes Factor ^a
MODEL 1				
Intercept	–	0.884	0.72 to 1.03	>900
<i>SULT1A1</i> , rs11648192	C/T	0.203	0.06 to 0.33	91.3
<i>CYP2D8P</i> , POS42547668	T/C	0.899	0.54 to 1.27	>900
<i>VKORC1</i> , rs8050894, rs9934438	C/G, G/A	–0.213	–0.37 to –0.06	>900
<i>VKORC1</i> , rs8050894, rs9934438	G/G, A/A	–0.719	–0.97 to –0.48	>900
Age	–	–0.102	–0.17 to –0.04	23.48
Weight	–	0.047	–0.05 to 0.15	2.61
Height	–	0.110	0.01 to 0.22	22.08
Sex	–	–0.124	–0.33 to 0.05	6.41
MODEL 2				
Intercept	–	0.929	0.752 to 1.107	>900
<i>SULT1A1</i> , rs11648192	C/T	0.188	0.042 to 0.335	46.5
<i>CYP2D8P</i> , POS42547668	T/C	1.092	0.709 to 1.481	>900
<i>VKORC1</i> , rs9934438	G/A	–0.235	–0.399 to –0.061	>900
<i>VKORC1</i> , rs9934438	A/A	–0.529	–0.784 to –0.246	>900
Age	–	–0.099	–0.171 to –0.018	13.63
Weight	–	0.061	–0.037 to 0.17	2.37
Height	–	0.097	–0.021 to 0.211	10.54
Sex	–	–0.098	–0.313 to 0.092	3.43

^aThe Bayes Factor corresponds to a hypothesis test with $H_0: \beta = 0$ and $H_1: \beta = \mu$, where β is the coefficient and μ is the posterior mean. HDI is the high-density interval for μ .

of multiple variants on certain genes that affects several steps of the pharmacodynamics or pharmacokinetics and thus, drug efficacy.

We acknowledge the size and closed patient group studied retrospectively in individuals that were already assigned a dose by trial and error, not allowing for a prospective use of the genetic information. These observations will require confirmation and replication. We provide a list of 20 variants in 18 genes ordered by its impact on acenocoumarol dose around the PK/PD of coumarins and the biochemistry of the coagulation cascade.

Our results offer an initial insight to the use of a genomic approach in pharmacogenetics showing that the advent of next generation sequencing may offer an alternative to identify and utilize individual variation to potentially explain a pharmacological relevant phenotype (Cheng et al., 2015). Future endeavors should focus on confirming these observations in a larger population.

As of June 2017, there are under a dozen reported studies of NGS in Mexican populations (NBCI), here, we present a list of variants in 110 pharmacogenes in Mexican Mestizos providing population information for allele frequency, differentiation from other continental groups and phenotype associations, which may complement the current catalog of pharmacogenomic variation in different populations.

AUTHOR CONTRIBUTIONS

VG-C: Performed genomic experiments, interpreted results, and wrote the manuscript, JU-C: Analyzed the data

and developed the Bayesian model, BV-T: Coordinated demographics and clinical data, performed genotyping experiments. JC-A: Conceived, planned, and performed the clinical study, ST-C: Conceived, planned, and performed the clinical study, RI-A: Conceived, planned, and performed the clinical study, OF-L: Conceived, planned, and performed the clinical study, XS: Conceived and coordinated genomic analyses.

ACKNOWLEDGMENTS

This work was supported by Conacyt CB-2015-01 No.252952 to the project “Diversidad Farmacogenetica en Mexicanos, coleccion e interpretacion,” and by INMEGEN to the Pharmacogenomics Laboratory, project No.22/2014/I. The technical expertise of Roberto Galindo from Winter

Genomics during bioinformatic analyses is gratefully acknowledged. M.Sc. Alfredo Mendoza Vargas from the NGS unit, kindly provided support and assistance during DNA sequencing. We also thank the staff at the National Institute of Cardiology, Marisol Serna Galarza, Sandra J. Rodriguez Duarte, MD. Jose Luis Lopez, and at INMEGEN Itzel A. Alva-Velazquez, and Anayelli Munoz-Rivas for their clinical and technical assistance during patient treatment, and sample management.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2017.00863/full#supplementary-material>

REFERENCES

- Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bolln, C., et al. (2013). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 41, D8–D20. doi: 10.1093/nar/gks1189
- Allan, P., Rettie, E., and Jones, J. P. (2005). clinical and toxicological relevance of cyp2c9: drug-drug interactions and pharmacogenetics. *Annu. Rev. Pharmacol. Toxicol.* 45, 477–494. doi: 10.1146/annurev.pharmtox.45.120403.095821
- Baldi, P., and Long, A. D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* 17, 509–519. doi: 10.1093/bioinformatics/17.6.509
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bonifaz-Peña, V., Contreras, A. V., Struchiner, C. J., Roela, R. A., Furuya-Mazzotti, T. K., Chammas, R., et al. (2014). Exploring the distribution of genetic markers of pharmacogenomics relevance in brazilian and mexican populations. *PLoS ONE* 9:e112640. doi: 10.1371/journal.pone.0112640
- Carcao, M., Re, W., and Ewenstein, B. (2015). The role of previously untreated patient studies in understanding the development of FVIII inhibitors. *Haemophilia* 22, 22–31. doi: 10.1111/hae.12790
- Chen, C., Mengersen, K. L., Ickstadt, K., Keith, J., Clair, L., Alston Anthony, N., et al. (2012). *Case Studies in Bayesian Statistical Modelling and Analysis, 1st Edn.* Brisbane, QLD: John Wiley.
- Cheng, R., Leung, R. K., Chen, Y., Pan, Y., Tong, Y., Li, Z., et al. (2015). Virtual pharmacist: a platform for pharmacogenomics. *PLoS ONE* 10:e0141105. doi: 10.1371/journal.pone.0141105
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Cruz-Correa, O. F., León-Cachón, R. B. R., Barrera-Saldaña, H. A., and Soberón, X. (2017). Prediction of atorvastatin plasmatic concentrations in healthy volunteers using integrated pharmacogenetics sequencing. *Pharmacogenomics* 18, 121–131. doi: 10.2217/pgs-2016-0072
- Danese, E., Montagnana, M., Johnson, J. A., Rettie, A. E., Zambon, C. F., Lubitz, S. A., et al. (2012). Impact of the CYP4F2 p.V433M polymorphism on coumarin dose requirement: systematic review and meta-analysis. *Clin. Pharmacol. Ther.* 92, 746–756. doi: 10.1038/clpt.2012.184
- Fricke-Galindo, I., Jung-Cook, H., LLerena, A., López-López, M., Scott, S., and Stein, C. (2016). Interethnic variability of pharmacogenetic biomarkers in Mexican healthy volunteers: a report from the RIBEF (Ibero-American Network of Pharmacogenetics and Pharmacogenomics). *Drug Metab. Pers. Ther.* 31, 625–629. doi: 10.1515/dmpt-2015-0030
- Galanter, J. M., Fernandez-Lopez, J. C., Gignoux, C. R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., et al. (2012). Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet.* 8:e1002554. doi: 10.1371/journal.pgen.1002554
- Harrington, D. J., Gorska, R., Wheeler, R., Davidson, S., Murden, S., Morse, C., et al. (2008). Pharmacodynamic resistance to warfarin is associated with nucleotide substitutions in VKORC1. *J. Thromb. Haemost.* 6, 1663–1670. doi: 10.1111/j.1538-7836.2008.03116.x
- Heim, M. H., and Meyer, U. A. (1992). Evolution of a highly polymorphic human cytochrome P450 gene cluster: CYP2D6. *Genomics* 14, 49–58. doi: 10.1016/S0888-7543(05)80282-4
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Oliveira-Paula, G. H., Luizon, M. R., Lacchini, R., Fontana, V., Silva, P. S., Biagi, C., et al. (2017). Gene-gene interactions among PRKCA, NOS3 and BDKRB2 polymorphisms affect the antihypertensive effects of enalapril. *Basic Clin. Pharmacol. Toxicol.* 120, 284–291. doi: 10.1111/bcpt.12682
- Pratt, V. M., Zehnbauser, B., Wilson, J. A., Baak, R., Babic, N., Bettinotti, M., et al. (2010). Characterization of 107 genomic DNA reference materials for CYP2D6, CYP2C19, CYP2C9, VKORC1, and UGT1A1: A GeT-RM and association for molecular pathology collaborative project. *J. Mol. Diagnost.* 12, 835–846. doi: 10.2353/jmoldx.2010.100090
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Relling, M. V., and Klein, T. E. (2011). CPIC: Clinical Pharmacogenetics Implementation Consortium of the pharmacogenomics research network. *Clin. Pharmacol. Ther.* 89, 464–467. doi: 10.1038/clpt.2010.279
- Rieder, M. J., Reiner, A. P., Gage, B. F., Nickerson, D. A., Eby, C. S., McLeod, H. L., et al. (2005). Effect of VKORC1 Haplotypes on transcriptional regulation and warfarin dose. *N. Engl. J. Med.* 352, 2285–2293. doi: 10.1056/NEJMoa044503
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotech.* 29, 24–26. doi: 10.1038/nbt.1754
- Scott, S. A., Khasawneh, R., Peter, I., Kornreich, R., and Desnick, R. J. (2014). Combined CYP2C9, VKORC1 and CYP4F2 frequencies among racial and ethnic groups. *Pharmacogenomics* 11, 781–791. doi: 10.2217/pgs.10.49
- Sebastiani, P., Timofeev, N., Dworkis, D. A., Perls, T. T., and Steinberg, M. H. (2009). Genome-wide association studies and the genetic dissection of complex traits. *Am. J. Hematol.* 84, 504–515. doi: 10.1002/ajh.21440
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., and Smigielski, E. M. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi: 10.1093/nar/29.1.308
- Soria, J. M., Almasy, L., Souto, J. C., Sabaterlleal, M., Fontcuberta, J., and Blangero, J. (2009). The F7 gene and clotting factor VII levels: dissection of a human quantitative trait locus. 2005. *Hum. Biol.* 81, 853–867. doi: 10.3378/027.081.0627

- Tàssies, D., Freire, C., Pijoan, J., Maragall, S., Monteagudo, J., Ordinas, A., et al. (2002). Pharmacogenetics of acenocoumarol: cytochrome P450 CYP2C9 polymorphisms influence dose requirements and stability of anticoagulation. *Haematologica* 87, 1185–1191.
- The 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Tong, H. Y., Dávila-Fajardo, C. L., Borobia, A. M., Martínez-González, L. J., Lubomirov, R., Perea León, L. M., et al. (2016). A new pharmacogenetic algorithm to predict the most appropriate dosage of acenocoumarol for stable anticoagulation in a mixed spanish population. *PLoS ONE* 11:e0150456. doi: 10.1371/journal.pone.0150456
- Ufer, M. (2005). Comparative pharmacokinetics of vitamin K antagonists. *Clin. Pharmacokinet.* 44, 1227–1246. doi: 10.2165/00003088-200544120-00003
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From fastq data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–33. doi: 10.1002/0471250953.bi1110s43
- van Leeuwen, Y., Rosendaal, F. R., and van der Meer, F. J. M. (2008). The relationship between maintenance dosages of three vitamin K antagonists: acenocoumarol, warfarin and phenprocoumon. *Thromb. Res.* 123, 225–230. doi: 10.1016/j.thromres.2008.01.020
- Villegas-Torres, B., Sánchez-Girón, F., Jaramillo-Villafuerte, K., Soberón, X., and Gonzalez-Covarrubias, V. (2015). Genotype frequencies of VKORC1 and CYP2C9 in Native and Mestizo populations from Mexico, potential impact for coumarin dosing. *Gene* 558, 235–240. doi: 10.1016/j.gene.2014.12.068
- Waldron, D. (2016). Genetic variation: diving deep into the genome. *Nat. Rev. Genet.* 17, 716–717. doi: 10.1038/nrg.2016.144
- Wang, D., Papp, A. C., and Sun, X. (2015). Functional characterization of CYP2D6 enhancer polymorphisms. *Hum. Mol. Genet.* 24, 1556–1562. doi: 10.1093/hmg/ddu566
- Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 92, 414–417. doi: 10.1038/clpt.2012.96
- Zhang, Y., de Boer, A., Verhoef, T. I., van der Meer, F. J., Le Cessie, S., Maitland-van der Zee, A. H., et al. (2015). Comparison of dosing algorithms for acenocoumarol and phenprocoumon using clinical factors with the standard care in the Netherlands. *Thromb. Res.* 136, 94–100. doi: 10.1016/j.thromres.2015.04.034

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Gonzalez-Covarrubias, Urena-Carrion, Villegas-Torres, Cossio-Aranda, Trevethan-Cravioto, Izaguirre-Avila, Fiscal-López and Soberon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multiplexed Nanopore Sequencing of HLA-B Locus in Māori and Pacific Island Samples

Kim N. T. Ton¹, Simone L. Cree¹, Sabine J. Gronert-Sum², Tony R. Merriman³,
Lisa K. Stamp⁴ and Martin A. Kennedy^{1*}

¹ Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand, ² JSI Medical Systems GmbH, Ettenheim, Germany, ³ Biochemistry Department, University of Otago, Dunedin, New Zealand, ⁴ Department of Medicine, University of Otago, Christchurch, New Zealand

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Federica Sangiuolo,
Università degli Studi di Roma Tor
Vergata, Italy
Louise Warrich,
Stellenbosch University, South Africa

*Correspondence:

Martin A. Kennedy
martin.kennedy@otago.ac.nz

Specialty section:

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Genetics

Received: 21 December 2017

Accepted: 12 April 2018

Published: 30 April 2018

Citation:

Ton KNT, Cree SL, Gronert-Sum SJ,
Merriman TR, Stamp LK and
Kennedy MA (2018) Multiplexed
Nanopore Sequencing of HLA-B
Locus in Māori and Pacific Island
Samples. *Front. Genet.* 9:152.
doi: 10.3389/fgene.2018.00152

The human leukocyte antigen (HLA) system encodes the human major histocompatibility complex (MHC). HLA-B is the most polymorphic gene in the MHC class I region and many HLA-B alleles have been associated with adverse drug reactions (ADRs) and disease susceptibility. The frequency of such HLA-B alleles varies by ethnicity, and therefore it is important to understand the prevalence of such alleles in different population groups. Research into HLA involvement in ADRs would be facilitated by improved methods for genotyping key HLA-B alleles. Here, we describe an approach to HLA-B typing using next generation sequencing (NGS) on the MinION™ nanopore sequencer, combined with data analysis with the SeqNext-HLA software package. The nanopore sequencer offers the advantages of long-read capability and single molecule reads, which can facilitate effective haplotyping. We developed this method using reference samples as well as individuals of New Zealand Māori or Pacific Island descent, because HLA-B diversity in these populations is not well understood. We demonstrate here that nanopore sequencing of barcoded, pooled, 943 bp polymerase chain reaction (PCR) amplicons of 49 DNA samples generated ample read depth for all samples. HLA-B alleles were assigned to all samples at high-resolution with very little ambiguity. Our method is a scaleable and efficient approach for genotyping HLA-B and potentially any other HLA locus. Finally, we report our findings on HLA-B genotypes of this cohort, which adds to our understanding of HLA-B allele frequencies among Māori and Pacific Island people.

Keywords: HLA-B, nanopore sequencing, Māori, Pacific Island, pharmacogenetics, Polynesian

INTRODUCTION

The human leukocyte antigen (HLA) locus contains a large family of genes encoding the human major histocompatibility complex (MHC) proteins. It is located on chromosome 6p21 and divided into three classes: class I, class II, and class III. HLA molecules are extremely variable due to their peptide-binding function and are associated with autoimmune diseases and adverse drug reactions (ADRs) (Tiwari and Terasaki, 1985; Bharadwaj et al., 2010). HLA-B is the most polymorphic gene, with over 4,600 known alleles encoding 3,408 unique proteins (IMGT/HLA Database release 3.27 in January 2017) (Robinson et al., 2014). Previous studies have identified particular HLA-B alleles as risk factors for drug-induced hypersensitivity reactions (Alfirevic and Pirmohamed, 2010; Sukasem et al., 2014). For example, screening for the HLA-B*57:01

A preprint of this manuscript is in bioRxiv with doi: <https://doi.org/10.1101/169078>

allele is recommended prior to abacavir treatment to decrease risk of a hypersensitivity reaction (Martin et al., 2012). Strong association between HLA-B*58:01 and allopurinol-induced severe cutaneous adverse reactions such as Stevens–Johnson syndrome (SJS) or toxic epidermal necrolysis (TEN) have been reported across different populations (Somkruea et al., 2011). However, some HLA alleles associated with drug-induced hypersensitivity can be ethnic group-specific. For example, HLA-B*15:02 allele is a risk factor for carbamazepine-induced SJS/TEN found in several Asian populations but not in Caucasian and Japanese populations (Tassaneeyakul et al., 2010; Phillips et al., 2018). Given the correlation of ethnic-specific risk alleles with ADRs, a good knowledge of HLA allele frequencies, and the prevalence of susceptibility alleles in particular, is important for the study of pharmacogenetics.

The high level of polymorphism in the MHC family means HLA genotyping is complex. HLA alleles are mostly determined by the sequences of exons 2 and 3 in HLA class I genes and exon 2 in HLA class II genes (Shiina et al., 2012). Present DNA-based methods for HLA typing are polymerase chain reaction (PCR) -sequence-specific priming (PCR-SSP), PCR-sequence-specific oligo hybridization (PCR-SSO), PCR-restriction fragment length polymorphism (PCR-RFLP), and sequence-based typing (SBT) (Tait et al., 2009; Bontadini, 2012; Erlich, 2012). SBT is currently considered the gold standard method applied in high-resolution HLA typing (Erlich, 2012). However, this approach may generate ambiguous HLA typing due to haplotype phase issues and incomplete sequencing. Other approaches have various limitations in resolution, workflow complexity, probe design and testing requirements, as new HLA alleles are submitted to the IMGT/HLA sequence database (Erlich, 2012; Shiina et al., 2012).

Recently, next-generation sequencing (NGS) methods have become widely established for HLA typing (Abbott et al., 2006; Bentley et al., 2009; Erlich et al., 2011; Erlich, 2012; Shiina et al., 2012; Hosomichi et al., 2013, 2015; Schöfl et al., 2017). Such approaches reduce the risk of phase ambiguity and allow high-throughput, high-resolution HLA typing. Various approaches to HLA typing using NGS have been developed, including PCR-based HLA sequencing (Erlich et al., 2011; Boegel et al., 2012; Liu et al., 2012; Shiina et al., 2012; Hosomichi et al., 2013; Schöfl et al., 2017), whole exome sequencing (WES) or whole genome sequencing (WGS) data-derived typing (Liu et al., 2012; Major et al., 2013). However, these methods are of limited value for research studies of ADRs, where a more targeted screening for specific HLA alleles may be all that is required.

Here we describe the development of high-throughput HLA typing from next-generation DNA sequencing data, focusing on the HLA-B locus, and its application to identifying HLA-B alleles within the Māori and Pacific Island population of New Zealand. Our strategy took advantage of a recent iteration of the novel NGS platform, the MinION™ nanopore sequencer (Oxford Nanopore Technologies), and barcode sequences for labeling and simultaneously analyzing HLA-B amplicons from multiple samples. The MinION is a tiny, portable nanopore sequencer powered by a USB 3.0 port (Quick et al., 2014). It allows analysis of sequencing data in

real time and generation of very long reads (Urban et al., 2015). A small pilot study used the device to examine HLA-A and HLA-B alleles from a single sample, using the earlier, quite error-prone R7.0 flow-cell chemistry (Ammar et al., 2015). Oxford Technologies released a new chemistry (R9.4) in May 2016, which has proven to be more accurate and with higher throughput. This major update motivated us to examine the performance of this pocket-sized device on one of the most polymorphic genes in the human body, HLA-B.

New Zealand is a multi-ethnic country with people from many different nations. The Māori are the indigenous Polynesian people, who first settled in New Zealand. New Zealand is also home to many people from the Pacific Islands, with its main city of Auckland referred to as the “Polynesian capital of the world” (Anae, 2005). Polynesian people in New Zealand include people from Samoa, Cook Islands, Tongan, Fiji, Niue and Tokelau, which together account for 7.6% of New Zealand population (Geck, 2017). To date, there is a paucity of studies providing prevalence data of HLA-B alleles in Māori and Pacific Island population in New Zealand (Abbott et al., 2006; Edinur et al., 2013). Given that HLA-B alleles are so relevant to disease predisposition and ADRs, it is important to establish a prevalence dataset for HLA-B for these population groups. Therefore, we sought to develop an assay for HLA-B screening in this population, using the MinION.

MATERIALS AND METHODS

Participants

Forty unrelated Māori and Pacific Island individuals with no history of inflammatory disorders were recruited from the Otago and Auckland regions of New Zealand. The proportion of ancestry for each participant was estimated by recording the ethnicity of all four grandparents. This study was carried out in accordance with the recommendations of the Standard Operating Procedures for Health and Disability Ethics Committees (New Zealand), as reviewed by the Lower South Ethics Committee (New Zealand), with written informed consent from all subjects. Additionally, five individuals were included from a local study on ADRs called Understanding ADRs or responses Using Genome Sequencing (UDRUGS) (Maggo et al., 2017), which was approved by the Southern Health and Disability Ethics Committee (New Zealand), with written informed consent from all subjects. A further four samples of known HLA-B genotype were obtained from the Coriell Institute for Medical Research (Camden, NJ, USA). These nine individuals were used as a reference set for the MinION analysis, after confirmation by either Sanger sequence based typing (SBT) or data retrieved from the 1000 Genomes project, or both.

HLA-B Genotyping by Sanger Sequencing

We selected a subset of our participants (four Polynesian, five UDRUGS and two Coriell samples) to analyze by Sanger sequencing, as additional references. Nested PCR was used to amplify a 1,710 bp region spanning exon 2 and exon 3 of HLA-B. PCR products were diluted and used

as templates in second round PCR to amplify a 943 bp amplicon. These amplicons were then directly sequenced in both forward and reverse directions using a set of six sequencing primers (Table 1). The primers used for amplification included some nucleotide redundancies at sites of HLA-B variation, to prevent allelic drop-out during the PCR step. All PCR primers and sequencing primers were derived from published work (Abbott et al., 2006; Cotton et al., 2012). The HLA-B genotypes of these 11 samples were generated from the Sanger sequence data using SBTengine v.3.10.0.2610 (GenDX, Utrecht, Holland).

Minion Library Construction

The primer used to amplify a fragment of 943 bp HLA-B exon 2 and 3 included a specific sequence (Table 1) at the 5' end, which is compatible with barcode sequences (Oxford Nanopore Technologies). A standard protocol of the Kapa LongRange Hotstart DNA PCR (Kapa Biosystems) was applied, consisting of 1X Kapa LongRange Buffer (without Mg^{2+}), 2.0 mM $MgCl_2$, 0.3 mM dNTPs (2.5 mM each dNTP), 0.5 μ M of each primer, 1.25 U/50 μ l Kapa LongRange HotStart DNA Polymerase, 50 ng genomic DNA, and water up to 50 μ l. Thermal cycling conditions were 94°C for 3 min, 25 cycles at 94°C for 15 s, 68°C for 15 s, and 72°C for 1 min, with a final extension at 72°C for 1 min. The PCR products were visualized by electrophoresis on 2% agarose gel stained with SYBRTM Safe DNA Gel Stain (Invitrogen), and then purified using 1x Agencourt AMPure XP beads (Beckman Coulter).

PCR products were quantified by Qubit[®] 2.0 Fluorometer (ThermoFisher Scientific) and were diluted to 2 nM in water. A second PCR was performed to incorporate barcode sequences using Oxford Nanopore PCR Barcoding kit (EXP-PBC096). Each 100 μ l reaction contained 1X Kapa LongRange Buffer (without Mg^{2+}), 2.0 mM $MgCl_2$, 0.3 mM dNTPs (10 mM each dNTP), 0.2 μ M PCR Barcode primers (from BC01 to BC49), 2.0 U Kapa LongRange HotStart DNA Polymerase, and 0.5 nM of first-round PCR product. The cycling parameters were an initial

denaturation 95°C for 3 min, followed by 15 cycles at 95°C for 15 s, 62°C for 15 s, and 65°C for 1 min, with a final extension at 65°C for 1 min. All 49 barcoded products were cleaned up with 1x Agencourt AMPure XP beads, then quantified. Purified PCR products were normalized by concentration before being pooled for library preparation.

The pooled library was prepared using the Oxford Nanopore Sequencing protocol (SQK-NSK007). We used 5 μ g of library as an input, instead of the recommended 1 μ g, to improve yield for downstream steps. Briefly, 5 μ g of purified amplicon library was prepared with the NEBNext end repair module (New England Biolabs), then dA-tailed using the NEBNext dA-tailing module (New England Biolabs). The end-prepared, dA-tailed library was subsequently ligated with leader and hairpin adapters, followed by purification using Dynabeads[®] MyOneTM Streptavidin C1 beads (Invitrogen).

The final prepared library from 49 participants was loaded into the MinION R9.4 flowcell (Oxford Nanopore Technologies). The flowcell was run for 48 h using the MinKNOW software (0.51.1.39).

Data Analysis

Raw sequence data were uploaded for base-calling using Metrichor software (2D Basecalling for SQKMAP007 - v1.107). Sequences in FASTA format were extracted from the raw FAST5 files using poretools v.0.6.0 (Loman and Quinlan, 2014). Statistical analysis of the MinION sequencing data were generated and visualized by De Coster et al. (2017). In order to determine error rates, base-calls in FASTQ format were extracted using poretools v.0.6.0 (Loman and Quinlan, 2014) and then aligned against Sanger sequenced reference using BWA-MEM (version 0.7.12-41044), parameter “-x ont2d”. Additional statistical analyses were performed with Python and R scripts available at <https://github.com/camilla-ip/marcp2>. We only used two-dimensional (2D) reads, which are consensus calls of the combined template and complement strands, to perform HLA-B locus high-resolution typing with SeqPilot

TABLE 1 | Primers used for amplifying and sequencing.

Assay	Primer	Sequence	Source
SBT	HLA-B_1710_F	TGTCGGGTCCTCTCTCCAGG	Abbott et al., 2006
	HLA-B_1710_R	GAAATTCAGGCGCTTT	
	HLA-B_943_F	GCAGGCGGGGCGCAGGACC	Cotton et al., 2012
	HLA-B_943_R	GGAGATGGGGAAGGCTCCCCACT	
	HLA-B_Seq1	GGAGCCGCGCCGGAGGAGGGTC	
	HLA-B_Seq2	GGATGGGGAGTCGTGACCT	
	HLA-B_Seq3	ACKGKGCTGACCGCGGGG	
	HLA-B_Seq4	CGGGGTCACTCACCGKCCTC	
	HLA-B_Seq5	GGSCKGGGCCAGGGTCTCAC	
	HLA-B_Seq6	ACTGCCCTGGTACCMGCGC	
MinION	HLA-B_MinION_F	TTTCTGTTGGTGCTGATATTCGGGAGGAGMRAGGGGACCSCAG	
	HLA-B_MinION_R	ACTTGCTGTGCTCTATCTTCGGAGGCCATCCCCGGCGACCTAT	

Underlined letters are IUPAC codes indicating base redundancies at positions corresponding to known HLA-B variants.

v.4.3.1 using default settings (JSI medical systems). HLA-B ambiguities were designated as G group nomenclature (http://hla.alleles.org/alleles/g_groups.html). Samples that could not be assigned genotypes due to mismatches were re-analyzed for error correction. Nanopolish pipeline was applied on input reads of these samples to check improvement on the base-level accuracy (Loman et al., 2015). After polishing, consensus sequences were re-processed to assign HLA-B genotypes of these samples.

RESULTS

A total of 40 Maori and Pacific Island, four Coriell and five UDRUGS individuals were selected for library construction. We successfully amplified a region of 943 bp spanning exon 2 and 3 of HLA-B in a single PCR. After purifying, all 49 PCR products were diluted to reach the desired concentration (2.0 nM). Three gave insufficient yield, ranging from 0.19 to 1.47 nM. However, these three were still included to test whether they could be effectively amplified in the barcoding step.

Two primers used in the first PCR were tailed with the adapter sequences, which were compatible with Oxford Nanopore barcodes. Each PCR product was then amplified with barcode primers, at which point all 49 PCR products were tagged with barcodes, increasing their length to 1,063 bp (**Figure 1**).

PCR products were subjected to normalization prior to pooling and sequencing on the MinION (~368 ng/each). Five samples had significantly lower concentrations than other samples (range: 20.4–66.8 ng), but these were included in the pool (**Table 2**). The total DNA quantity in the pool was 7.5 µg and 5 µg was used for downstream steps. After end-repair, adapter ligation and purification steps, 585 ng of prepared library remained and was loaded into the MinION flow cell.

Given that for this version of the MinION chemistry, 2D reads were more accurate and had greater length than 1D reads, we extracted only the 2D reads for downstream analysis. After conversion, all 49 FASTQ files were imported into SeqPilot

software for HLA-B allele assignment. The mean read depth was 5,807x and the mean read length was 1,029 bp, close to the expected size of all amplicons (**Figure 2**). An average of 5,854 sequence reads per barcoded sample was obtained from a total of 289,095 2D pass reads. The proportion of reads with a Q-value threshold of 15 was 83.3% (**Figure 3**). There were 286,852 reads with uniquely identified barcodes, of which 199,297 reads passed the quality filters and were aligned to the assigned allele sequences. The distribution histogram of both assigned and aligned reads for each barcoded sample is shown in **Figure 4**.

The five samples that amplified poorly still produced ample reads for SeqPilot to generate HLA-B typing calls (**Table 2**). Notably, individual PI_C3, which had the lowest number of mapped reads (80) that aligned to the reference sequence, was

TABLE 2 | DNA quantity and number of mapped reads of poorly amplified samples.

Individual	Index	DNA quantity (ng)	Mapped 2D read (number)
PL_G2	BC15	66.84	473
PI_C3	BC19	20.4	80
PI_D3	BC20	57.96	869
PI_H3	BC22	64.44	333
PI_C6	BC42	33.36	158

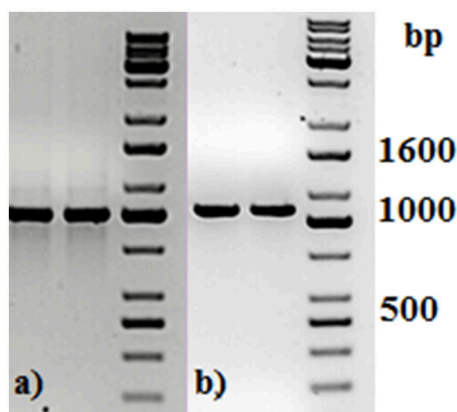


FIGURE 1 | Agarose gel electrophoresis of two representative PCR amplicons (a) before and (b) after PCR barcoding step. DNA molecular size marker obtained from KAPA Universal Ladder (KAPA Biosystems, Boston, MA).

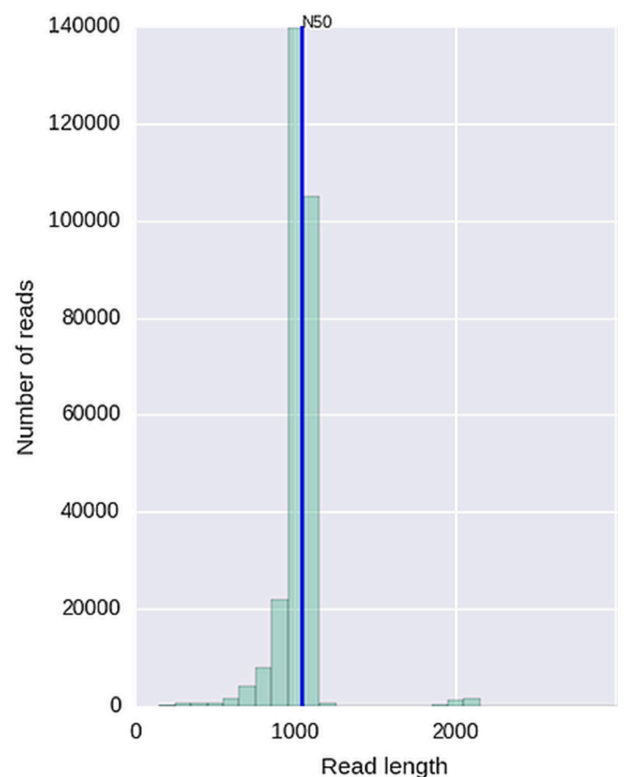


FIGURE 2 | Histogram of read length with read N50 metric.

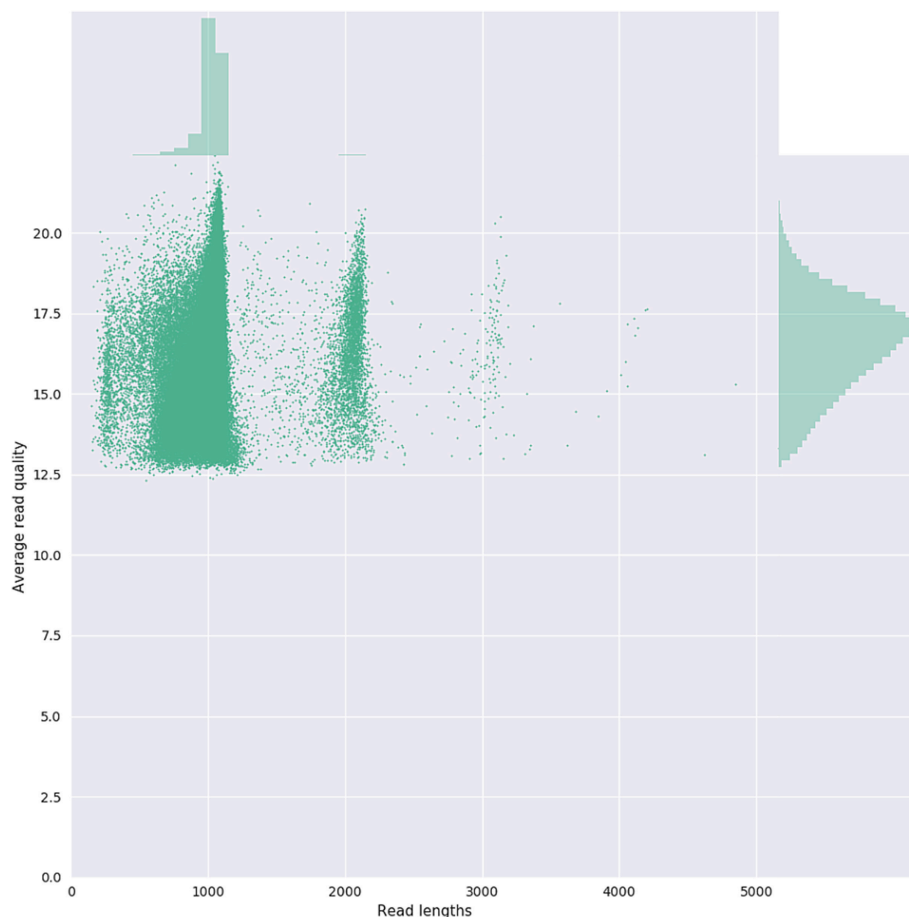


FIGURE 3 | Bivariate plot shows with a kernel density estimate the read length compared to the average read basecall Phred quality.

assigned the alleles HLA-B*44:04, 56:02:01. We are confident these alleles are correct as no mismatches occurred at key polymorphic sites in the reads.

Using FASTQ files as inputs, SeqPilot effectively assigned HLA-B alleles of 49 individuals. There were 4 homozygotes and 45 heterozygotes, resulting in 38 alleles called at the third field (formerly 6-digit) resolution and five alleles at second field (4-digit) resolution (Table 3). There were six that could not be automatically called by the software due to mismatches with reference sequences. Notably, the phasing bias mostly happened at nucleotide position 130–136 of exon 2. We realized that there was A/G heterozygous at nucleotide 133 and the sequence around this position was a repeat of G and A (GAGAGRGGAG). We also observed that if nucleotide 133 was G, there was usually a deletion of two to four nucleotides (Figure 5). Figure 6 illustrates the ambiguous results of sample PI_A2 with several possible alleles and their corresponding mismatched locations. HLA-B*27:05:02G was not able to be confidently called due to two mismatches at the area mentioned above (Figure 6). Sanger sequencing these six samples confirmed that these mismatches were not PCR artifacts, suggesting these

errors might happen during the nanopore sequencing step. To explore whether such errors could be corrected, we ran the Nanopolish pipeline to compute consensus sequences with improved base quality. After comparing these polished sequences with sequences obtained from Sanger sequencing, we found that Nanopolish did not improve the accuracy for these error regions further (Figure 7). In this case, we manually assigned allele pairs from the suggested list, choosing those with the least mismatch. Although other NGS approaches have been applied to HLA analysis, Sanger sequencing is regarded as the gold standard for HLA-typing. Consequently, we selected 11 individuals for HLA-B genotyping using the Sanger SBT method for validation. These were four Polynesian, five UDRUGS and two Coriell samples. Six amplicons covering exon 2 and 3 of each individual were directly sequenced. HLA-B alleles were defined based on these polymorphic sequences. Using SBTengine (GenDX, Utrecht, Holland) for analysis, we found that there was a high consistency of variant calls with calls derived from the MinION data. Our genotyping data for all four Coriell individuals were concordant with the data from these samples generated by the 1000 Genomes project. These

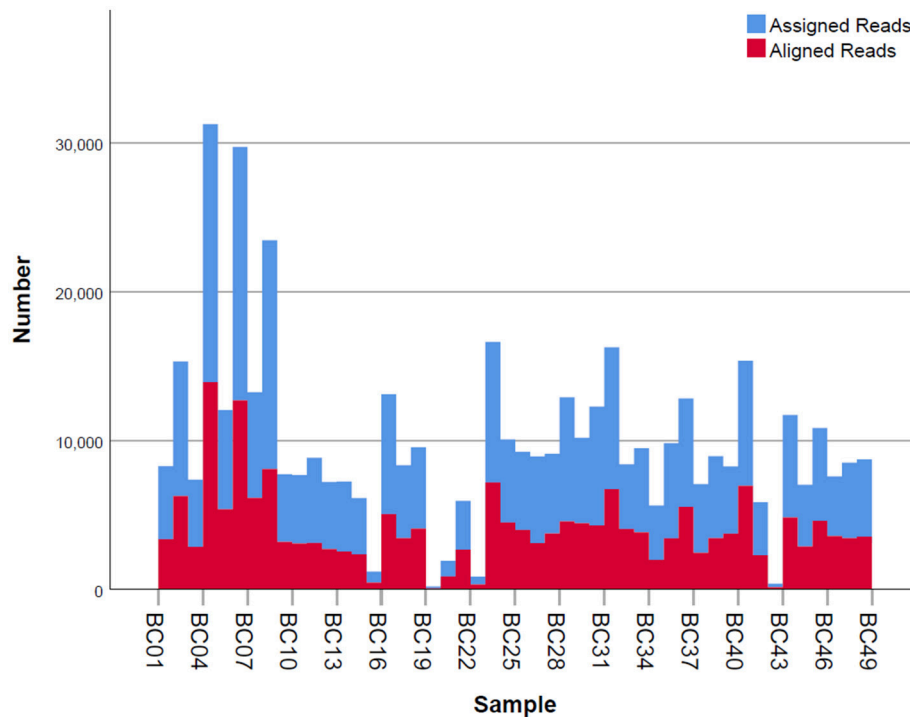


FIGURE 4 | Number of mapped and unmapped reads per individual. Blue bars indicate the assigned reads that are mapped to the regions of interest. Red bars are the aligned reads that pass the filter and are used for allele assignment.

results indicate that this MinION sequencing method is able to generate consensus sequences for high-resolution HLA-B typing with considerable accuracy.

There were 38 HLA-B alleles identified in the 40 Māori and Pacific Island individuals examined. Among these alleles, HLA-B*40:01:01 had the highest frequency (28.95%), followed by HLA-B*44:02:01 (21.05%) and HLA-B*07:02:01 (18.42%). According to the HLA allele frequency database (<http://www.allelefrequencies.net>), HLA-B*40:01:01 is the most prevalent allele in the Han Chinese population (allele frequency of 0.155). It has been reported that the Polynesian people are ancestrally related to Micronesia, Taiwanese Aborigines and East Asia (Kayser et al., 2008; Edinur et al., 2013). This is consistent with our observation on the frequency of HLA-B*40:01:01 in this population. On the other hand, the HLA-B*44:02:01 and HLA-B*07:02:01 alleles are present in multiple populations. When comparing with other studies on Polynesians, we found there were four previously observed alleles not represented in our data (Edinur et al., 2013). Presumably this is because our sample size is insufficient to reflect the full range of HLA-B alleles in Pacific Island or Māori populations. Nevertheless, the most common allele (HLA-B*40:01:01) in our study is similar to the Edinur findings, suggesting that this is a common HLA-B allele in this population. Interestingly, there were no HLA-B*15:02:01 nor B*58:01 observed in the Polynesian individuals. The lack of HLA-B*58:01 in Polynesians has previously been reported in other studies (Abbott et al., 2006; Roberts et al., 2015). These alleles have been implicated in

carbamazepine and allopurinol-induced severe cutaneous adverse reactions, respectively. However, HLA-B*57:01:01, important for abacavir-associated hypersensitivity reactions, was apparent in two samples.

DISCUSSION

The primary goal of our study was to develop methods for HLA-B class 1 typing on the MinION nanopore sequencer. This study applied PCR across HLA-B exon 2 and exon 3, followed by barcoding and nanopore sequencing of 49 samples simultaneously. The high quality and good depth of coverage of our sequencing data for all samples, including several that were present at low concentration, enabled accurate assignment of HLA-B alleles. With ongoing improvements to the speed, throughput and workflow of MinION flowcells and the associated chemistry, it is likely that multiplexing could be extended to much greater numbers without compromising the ability for accurate typing.

The workflow we employed was straightforward and solely PCR-based. Though the ONT protocol requires 2.0 nM of input amplicon prior to the barcoding step, sample PI_G2 was successfully amplified and produced sequencing data at 0.19 nM, less than 10-fold the recommended amount. That said, even these poorly represented samples generated sufficient data for HLA-B typing. Of all the reads generated from the sequencing device, 199,297 reads had good quality and passed the filters for alignment. Reads were ignored if they did not map to the

TABLE 3 | Assignment result of HLA-B obtained from MinION sequencing and SBT.

Sample	Ethnicity#	MinION		SBT	
NA16688 [§]	Ch	B*15:07:01G	*35:01:01G		
NA17019 [¥]	Ch	B*15:02:01G	*15:11:01G	B*15:02:01G	*15:11:01G
NA17240 ^{¥§}	C	B*07:02:01G	*57:01:01G	B*07:02:01G	*57:01:01G
NA19834 [¥]	A	B*35:01:01G	*39:10:01		
UDRUGS01 [¥]	C	B*40:02:01G	*44:02:01G	B*40:02:01G	*44:02:01G
UDRUGS02 [¥]	C	B*15:01:01G	*44:02:01G	B*15:01:01G	*44:02:01G
UDRUGS29 ^{¥§}	C	B*18:01:01G	*27:09	B*18:01:01G	*27:09
UDRUGS41 [¥]	C	B*15:01:01G	*51:01:01G	B*15:01:01G	*51:01:01G
UDRUGS44 [¥]	C	B*15:01:01G	*44:03:01G	B*15:01:01G	*44:03:01G
PI_A1 [¥]	0.25 NM 0.75 C	B*37:01:01G	*39:01:01G	B*37:01:01G	*39:01:01G
PI_A2 ^{¥§}	0.5 NM 0.5 C	B*27:05:02G	*57:01:01G	B*27:05:02G	*57:01:01G
PI_A3 [¥]	0.5 CM 0.5 P	B*39:05:01	*52:01:01G	B*39:05:01	*52:01:01G
PI_A5	0.125 NM 0.25 P 0.625 C	B*08:01:01G	*35:01:01G		
PI_B1	1.0 S	B*46:01:01G	*55:02:01G		
PI_B2	0.5 NM 0.5 C	B*07:02:01G	*57:01:01G		
PI_B3	U	B*27:07:01G	*35:03:01G		
PI_B4 [¥]	0.333 NM 0.667 P	B*40:01:01G	*44:02:01G	B*40:01:01G	*44:02:01G
PI_B5	0.5 NM 0.5 C	B*07:02:01G	*55:02:01G		
PI_B6	1.0 S	B*40:01:01G	*56:02:01		
PI_C1	0.5 NM	B*44:02:01G	*55:02:01G		
PI_C2	U	B*52:02:01	*52:02:01		
PI_C3	0.25 NM 0.25 C	B*44:04	*56:02:01		
PI_C4	0.25 NM 0.75 C	B*08:156	*42:01:01		
PI_C5	0.25 NM 0.75 N	B*55:01:01G	*55:02:01G		
PI_C6	0.25 NM 0.75 C	B*44:02:01G	*55:01:01G		
PI_D1	0.5 NM 0.5 C	B*15:01:01G	*49:01:01G		
PI_D2	0.125 NM 0.875 C	B*08:01:01G	*44:03:01G		
PI_D3	0.25 NM 0.75 C	B*40:01:01G	*51:01:01G		
PI_D4 [§]	0.5 NM 0.5 C	B*13:02:01G	*39:01:01G		
PI_D5 [§]	0.25 NM 0.5 N 0.25 C	B*40:01:01G	*40:01:01G		
PI_D6 [§]	0.375 NM 0.125 C	B*35:60	*56:09		
PI_E1	0.5 NM 0.5 C	B*35:03:01G	*40:01:01G		
PI_E2	0.5 S 0.5 N	B*56:02:01	*56:02:01		
PI_E4	0.5 NM 0.5 C	B*14:02:01G	*56:01:01G		
PI_E5	0.5 NM 0.5 S	B*39:01:01G	*55:01:01G		
PI_F1	0.5 NM 0.5 C	B*44:02:01G	*50:01:01G		
PI_F2	0.75 NM 0.25 C	B*07:02:01G	*40:01:01G		
PI_F3	0.125 CM 0.875 C	B*18:01:01G	*44:02:01G		
PI_F4	0.5 NM 0.5 C	B*44:02:01G	*43:03:01G		
PI_F5	0.125 NM 0.875 C	B*08:01:01G	*40:01:01G		
PI_F6	0.25 NM 0.75 C	B*35:01:01G	*48:01:01G		
PI_G1	0.5 NM 0.5 C	B*15:17:01G	*37:01:01G		
PI_G2	1.0 U	B*07:02:01G	*53:17:02		
PI_G4	0.125 NM 0.875 C	B*07:02:01G	*07:02:01G		
PI_G5	0.25 NM 0.25 C	B*35:01:01G	*40:01:01G		
PI_G6	0.333 NM 0.583 N 0.083 C	B*40:01:01G	*44:03:01G		
PI_H1	0.5 NM 0.5 C	B*27:10	*55:01:01G		
PI_H2	0.5 NM 0.5 C	B*07:02:01G	*40:10:01G		
PI_H3	1.0 NM	B*40:01:01G	*48:01:01G		

A, African; Ch, Chinese; C, Caucasian; NM, NZ Maori; CM, Cook Island Maori; S, Samoan; N, Niuean; P, others; U, unknown. ¥ Samples selected for validation. § Samples required manual allele assignment.

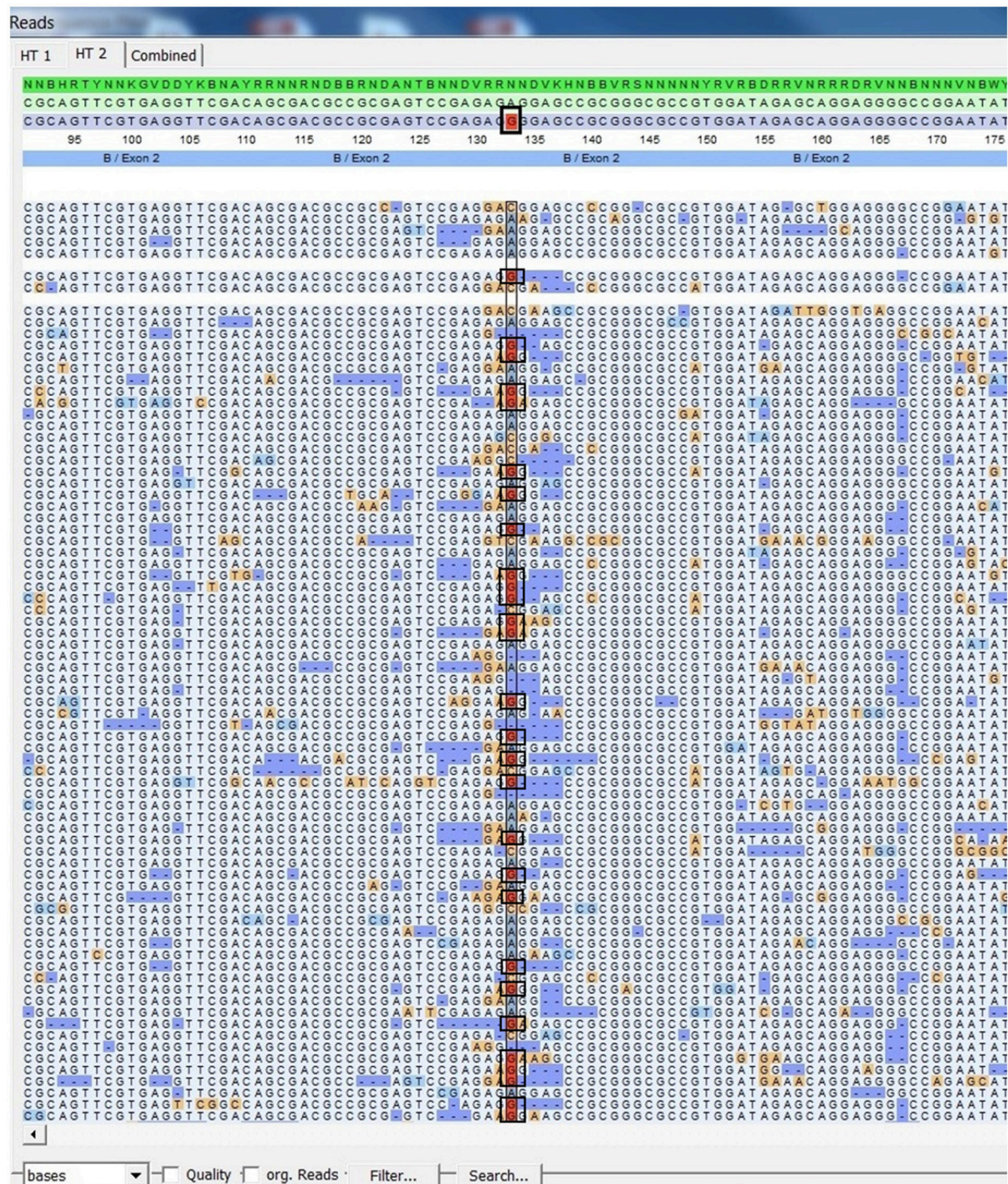


FIGURE 5 | SeqPilot screenshot from mapping sequencing reads with reference sequences. Bias phasing position (black box) followed by deletion errors.

region of interest. Among those usable reads, varied numbers of reads between samples were observed, ranging from 80 to 17,329 (Figure 4). Regardless of this read-depth variability between samples, adequate coverage was achieved for all samples and allowed for confident HLA-B phasing. It is also worth noting that individual PI_C3 only had 80 reads that were aligned with

reference sequences, but alleles could be assigned confidently. Our protocol takes 3 days to complete, comprising 1–1.5 days for library construction, 1 day for sequencing and base calling, and a half-day for data analysis. The sequencing step takes up to 2 days if aiming for more reads; however, it is possible to analyze data prior to the end of the run if desired. It can be



FIGURE 6 | SeqPilot screenshot from analysis of sample PI_A2. This illustrates the ambiguous results of sample PI_A2 with several possible alleles and their corresponding mismatched locations. HLA-B*27:05:02G was not confidently called due to two mismatches at nucleotide position 133 and 136 of exon 2. The software provides possible allele pairs with number of mismatches and their corresponding positions (red box). The mismatched nucleotides from the assigned allele are shown in each sequence read (ellipses).

argued that the turnaround time of our method is still longer than that of the gold standard SBT method. Other NGS workflows for MinION library construction may take even longer (3–4 days) with the employment of biotinylated probes (Karamitros and Magiorkinis, 2015). Moreover, given the fact that the capacity of the MinION can be enlarged to analyze other HLA loci at high-resolution and more sample input (up to 96 samples at present), MinION-based HLA typing can overcome this limitation. The ability to confidently call the genotype with such a small number of reads suggests that it would be reasonable to increase throughput by sequencing additional HLA loci, or indexing up to 96 individuals per sequencing run.

Our average read length was 1,028 bp indicating that our reads were long enough to cover exon 2 and exon 3 of HLA-B. We are aware that ambiguous typing of HLA-B might occur due to variants outside the region analyzed. This resulted in the assignment of G Codes in several samples (Table 3). In HLA typing, the letter “G” is used to report ambiguous alleles which have identical exon sequences encoding the peptide binding domains (exon 2 and exon 3 for HLA class I). A whole gene sequencing is required to resolve this ambiguity and obtain a full resolution of HLA allele. However, as not all alleles have been completely sequenced, only exon sequences can be mapped in some cases. For example, there are 4,356 HLA-B alleles but only 384 alleles have complete sequence information (Robinson et al., 2014). By this, we mean it is currently more practical to focus on

exons solely than to sequence the entire gene for HLA-B typing with minimal ambiguity. Obviously, we are able to increase the read length for complete sequencing of the HLA-B if necessary, as the MinION is capable of very long sequence reads (Carter and Hussain, 2017). Therefore, once all HLA-B alleles in the IMGT/HLA database have full information, our method can be adapted to sequence a full-length HLA-B with greater specificity and sensitivity.

Our results showed that SeqPilot was able to identify HLA-B alleles accurately using the MinION reads. Although the software could not automatically assign HLA-B alleles of all participants (6 cases), it listed the most likely genotype combinations in rank of number of mismatch sites. This enabled us to manually assign HLA-B alleles based on this order. We found that these errors only occurred at nucleotide position 130–136 of exon 2 and on samples which had a (GA)₃ repeat on one of the allele sequence. Deletion rates of the MinION using R7.3 and R9.0 chemistry are 4.1 and 3.5%, respectively (Jain et al., 2017). Therefore, it may be that deletion errors produced by the MinION device combined with the complex nature of the HLA-B, make accurate interpretation of genotype particularly difficult around this region. Of the six individuals that were manually phased, alleles from three had been identified either by the SBT method or the 1000 Genomes project or both. These alleles were all consistent with our manual assignments, suggesting this approach can be applicable in such circumstances.

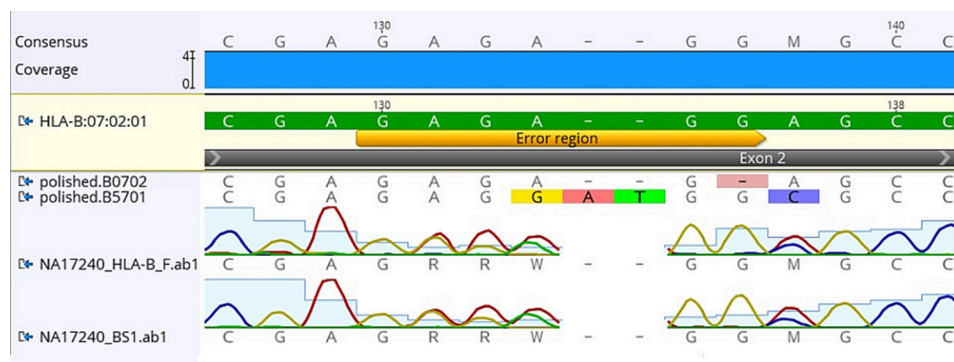


FIGURE 7 | Comparison of consensus sequences of NA17240 (BC47) obtained from Sanger sequencing with those from Nanopolish. Yellow bar indicates the region where mismatched alignment happens (nt 130–136). HLA-B*07:02:01 is the reference sequence; polished.B0702 and polished.B5701 are consensus sequences generated by Nanopolish. Data was visualized on Geneious v.9.1.5 (Biomatters Ltd., New Zealand).

The second aim of our study was to examine HLA-B alleles in individuals of Māori and Pacific Island descent living in New Zealand. Previous studies used allele-specific primer PCR for HLA-B genotyping, which provides typing at first field resolution (Edinur et al., 2013; Roberts et al., 2013). Here, we report a feasible method for high-throughput and high-resolution HLA-B typing using NGS. Though ours is a relatively small sample, this initial finding can be used as a reference for future studies on the prevalence of HLA-B in these ethnic groups.

In recent years, various high-throughput HLA typing studies have been conducted using different NGS technologies (Carapito et al., 2016). Though NGS-based HLA typing can be time-consuming (Chua and Ng, 2016), it offers high-resolution, unambiguous, phase-defined HLA alleles, overcoming some of the limitations of traditional approaches. For instance, the current gold standard method (SBT) may generate ambiguous typing due to genotype phase issues and incomplete sequencing. Modern NGS-based HLA typing methods were mostly developed on Illumina MiSeq/HiSeq or PGM Ion Torrent platforms, which employed short-to-medium sequencing read data as an input for HLA allele assignment (Hosomichi et al., 2015). At present, the Illumina platform has been widely adopted due to its high accuracy and high precision for HLA typing. However, the advent of long read single-molecule sequencing holds the promise of achieving full-length phase-defined HLA genes as well as detecting novel and rare variants. An early, very preliminary study suggested the MinION nanopore sequencer has potential for HLA analysis (Ammar et al., 2015). The device and associated chemistry has been continuously improving (Jain et al., 2017). For example, the total error of 2D reads reduced from 9.1% in R7.3 chemistry to 7.3% in R9.0 chemistry (now R9.4). Another advantage of the MinION device is its portability, which raises the possibility of using the device for HLA analysis in field situations or point-of-care settings.

Before this assay could be applied clinically, at least two things would need to occur. First, the nanopore technology is still in a state of relatively rapid development, and the MinION platform would need to stabilize before clinical implementation would be

possible. For example, since the work described in this report was completed there have been various further iterations of chemistry and library preparation procedures for the MinION. In addition, newer versions of the nanopore sequencing equipment with higher throughput (such as the GridION and PromethION) have been recently released to the market. Second, a much large study would be required to assess the sensitivity and specificity of the nanopore sequencing and allele calling procedure described in this paper. This would need to be carried out in a cohort which had undergone HLA typing using the current gold standard HLA typing approach of SBT (Erlich, 2012).

CONCLUSION

We have described here the development and evaluation of a PCR-based HLA-B sequencing method using MinION Nanopore Technology on R9.4 flow cell. We demonstrated that our method is relatively straightforward and can generate accurate sequencing data from many barcoded samples in a single run. We also reported that precise HLA-B alleles could be obtained from the MinION reads with minimal phase ambiguity. Our protocol can be easily adapted for other HLA loci, or for full gene sequencing, or to employ greater levels of multiplexing. The method could be particularly valuable for research studies examining the role of HLA alleles in ADRs.

DATA AVAILABILITY

The complete sequencing data can be accessed at the NCBI Sequence Read Archive database with the accession number SRP138979.

AUTHOR CONTRIBUTIONS

KT carried out the laboratory work, data analysis and drafted the manuscript. SC advised on nanopore sequencing

procedures and bioinformatic analyses. SG-S contributed to the analysis and assignment of HLA alleles. TM and LS recruited subjects and provided DNA for this analysis. MK supervised the work and contributed to preparation of the manuscript.

REFERENCES

- Abbott, W. G., Tukuitonga, C., Ofanoa, M., Munn, S., and Gane, E. (2006). Low-cost, simultaneous, single-sequence genotyping of the HLA-A, HLA-B and HLA-C loci. *Tissue Antigens* 68, 28–37. doi: 10.1111/j.1399-0039.2006.00620.x
- Alfirevic, A., and Pirmohamed, M. (2010). Drug induced hypersensitivity and the HLA complex. *Pharmaceuticals* 4, 69–90. doi: 10.3390/ph4010069
- Ammar, R., Paton, T. A., Torti, D., Shlien, A., and Bader, G. D. (2015). Long read nanopore sequencing for detection of *HLA* and *CYP2D6* variants and haplotypes. *F1000Research* 4:17. doi: 10.12688/f1000research.6037.2
- Anae, M. M. (2005). “Samoans”. *Te Ara - the Encyclopedia of New Zealand*. Available online at: <http://www.TeAra.govt.nz/en/samoans>
- Bentley, G., Higuchi, R., Hoglund, B., Goodridge, D., Sayer, D., Trachtenberg, E., et al. (2009). High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens* 74, 393–403. doi: 10.1111/j.1399-0039.2009.01345.x
- Bharadwaj, M., Illing, P., and Kostenko, L. (2010). Personalized medicine for HLA-associated drug-hypersensitivity reactions. *Per. Med.* 7, 495–516. doi: 10.2217/pme.10.46
- Boegel, S., Löwer, M., Schäfer, M., Bukur, T., de Graaf, J., Boissguérin, V., et al. (2012). HLA typing from RNA-Seq sequence reads. *Genome Med.* 4:102. doi: 10.1186/gm403
- Bontadini, A. (2012). HLA techniques: typing and antibody detection in the laboratory of immunogenetics. *Methods* 56, 471–476. doi: 10.1016/j.ymeth.2012.03.025
- Carapito, R., Radosavljevic, M., and Bahram, S. (2016). Next-generation sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies. *Hum. Immunol.* 77, 1016–1023. doi: 10.1016/j.humimm.2016.04.002
- Carter, J. M., and Hussain, S. (2017). Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome Open Res.* 2:23. doi: 10.12688/wellcomeopenres.11246.2
- Chua, E. W., and Ng, P. Y. (2016). MinION: a novel tool for predicting drug hypersensitivity? *Front. Pharmacol.* 7:156. doi: 10.3389/fphar.2016.00156
- Cotton, L. A., Abdur Rahman, M., Ng, C., Le, A. Q., Milloy, M., Mo, T., et al. (2012). HLA class I sequence-based typing using DNA recovered from frozen plasma. *J. Immunol. Methods* 382, 40–47. doi: 10.1016/j.jim.2012.05.003
- De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2017). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. doi: 10.1093/bioinformatics/bty149
- Edinur, H. A., Dunn, P. P., Hammond, L., Selwyn, C., Brescia, P., Askar, M., et al. (2013). HLA and MICA polymorphism in polynesians and New Zealand Maori: implications for ancestry and health. *Hum. Immunol.* 74, 1119–1129. doi: 10.1016/j.humimm.2013.06.011
- Erlich, H. (2012). HLA DNA typing: past, present, and future. *Tissue Antigens* 80, 1–11. doi: 10.1111/j.1399-0039.2012.01881.x
- Erlich, R. L., Jia, X., Anderson, S., Banks, E., Gao, X., Carrington, M., et al. (2011). Next-generation sequencing for HLA typing of class I loci. *BMC Genomics* 12:42. doi: 10.1186/1471-2164-12-42
- Geck, C. (2017). The world factbook. *Charlest. Adv.* 19, 58–60. doi: 10.5260/chara.19.1.58
- Hosomichi, K., Jinam, T. A., Mitsunaga, S., Nakaoka, H., and Inoue, I. (2013). Phase-defined complete sequencing of the HLA genes by next-generation sequencing. *BMC Genomics* 14:355. doi: 10.1186/1471-2164-14-355
- Hosomichi, K., Shiina, T., Tajima, A., and Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. *J. Hum. Genet.* 60, 665–673. doi: 10.1038/jhg.2015.102
- Jain, M., Tyson, J. R., Loose, M., Ip, C. L. C., Eccles, D. A., O'Grady, J., et al. (2017). MinION analysis and reference consortium: phase 2 data release and analysis of R9.0 chemistry. *F1000Research* 6:670. doi: 10.12688/f1000research.11354.1
- Karamitros, T., and Magiorkinis, G. (2015). A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Res.* 43:e152. doi: 10.1093/nar/gkv773
- Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nürnberg, P., et al. (2008). Genome-wide analysis indicates more Asian than melanesian ancestry of polynesians. *Am. J. Hum. Genetics* 82, 194–198. doi: 10.1016/j.ajhg.2007.09.010
- Liu, C., Xiao, Y., Duffy, B., Zody, M., Tycksen, E., Shrivastava, S., et al. (2012). High resolution HLA typing by next generation exome sequencing. *Blood* 120, 4166–4166. Available online at: <http://www.bloodjournal.org/content/120/21/4166>
- Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444
- Loman, N. J., and Quinlan, A. R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* 30, 3399–3401. doi: 10.1093/bioinformatics/btu555
- Maggo, S. D., Chua, E. W., Chin, P., Cree, S., Pearson, J., Doogue, M., et al. (2017). A New Zealand platform to enable genetic investigation of adverse drug reactions. *N. Z. Med. J.* 130, 62–69. Available online at: <http://www.nzma.org.nz/journal/read-the-journal/all-issues/2010-2019/2017/vol-130-no-1466-1-december-2017/7429>
- Major, E., Rigó, K., Hague, T., Bérces, A., and Juhos, S. (2013). HLA typing from 1000 genomes whole genome and whole exome illumina data. *PLoS ONE* 8:e78410. doi: 10.1371/journal.pone.0078410
- Martin, M. A., Klein, T. E., Dong, B. J., Pirmohamed, M., Haas, D. W., and Kroetz, D. L., (2012). Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. *Clin. Pharmacol. Therapeut.* 91, 734–738. doi: 10.1038/clpt.2011.355
- Phillips, E. J., Sukasem, C., Whirl-Carrillo, M., Müller, D. J., Dunnenberger, H. M., Chantaratita, W., et al. (2018). Clinical pharmacogenetics implementation consortium guideline for HLA genotype and use of carbamazepine and oxcarbazepine: 2017 update. *Clin. Pharmacol. Ther.* 103, 574–581. doi: 10.1002/cpt.1004
- Quick, J., Quinlan, A. R., and Loman, N. J. (2014). A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience* 3:22. doi: 10.1186/2047-217X-3-22
- Roberts, R. L., Wallace, M. C., Harrison, A., Dalbeth, N., Merriman, T. R., and Stamp, L. K. (2015). A human leukocyte antigen locus haplotype confers risk for allopurinol-related adverse effects in Caucasian patients with gout. *Pharmacogenet. Genomics* 25, 412–415. doi: 10.1097/FPC.0000000000000147
- Roberts, R. L., Wallace, M. C., Jones, G. T., van Rij, A. M., Merriman, T. R., Harrison, A., et al. (2013). Prevalence of HLA-B*27 in the New Zealand population: effect of age and ethnicity. *Arthritis Res. Ther.* 15:R158. doi: 10.1186/ar4341
- Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., and Marsh, S. G. (2014). The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* 43(Database issue), D423–D431. doi: 10.1093/nar/gku1161
- Schöfl, G., Lang, K., Quenzel, P., Böhme, I., Sauter, J., Hofmann, J. A., et al. (2017). 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. *BMC Genomics* 18:161. doi: 10.1186/s12864-017-3575-z
- Shiina, T., Suzuki, S., Ozaki, Y., Taira, H., Kikkawa, E., Shigenari, A., et al. (2012). Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue Antigens* 80, 305–316. doi: 10.1111/j.1399-0039.2012.01941.x

ACKNOWLEDGMENTS

KT was supported by a University of Otago PhD scholarship. This work was also supported by funding from the Jim and Mary Carney Charitable Trust.

- Somkruea, R., Eickman, E. E., Saokaew, S., Lohitnavy, M., and Chaiyakunapruk, N. (2011). Association of HLA-B* 5801 allele and allopurinol-induced Stevens Johnson syndrome and toxic epidermal necrolysis: a systematic review and meta-analysis. *BMC Med. Genet.* 12:118. doi: 10.1186/1471-2350-12-118
- Sukasem, C., Puangpetch, A., Medhasi, S., and Tassaneeyakul, W. (2014). Pharmacogenomics of drug-induced hypersensitivity reactions: challenges, opportunities and clinical implementation. *Asian Pac. J. Allergy Immunol.* 32, 111–123. Available online at: <https://search.proquest.com/docview/1614082458?accountid=14700>
- Tait, B. D., Hudson, F., Cantwell, L., Brewin, G., Holdsworth, R., Bennett, G., et al. (2009). Luminex technology for HLA antibody detection in organ transplantation. *Nephrology* 14, 247–254. doi: 10.1111/j.1440-1797.2008.01074.x
- Tassaneeyakul, W., Tiamkao, S., Jantararoungtong, T., Chen, P., Lin, S. Y., Chen, W. H., et al. (2010). Association between HLA-B* 1502 and carbamazepine-induced severe cutaneous adverse drug reactions in a Thai population. *Epilepsia* 51, 926–930. doi: 10.1111/j.1528-1167.2010.02533.x
- Tiwari, J. L., and Terasaki, P. I. (1985). *HLA and Disease Associations*. New York, NY: Springer Science & Business Media. doi: 10.1007/978-1-4613-8545-5
- Urban, J. M., Bliss, J., Lawrence, C. E., and Gerbi, S. A. (2015). Sequencing Ultra-Long DNA Molecules with the Oxford Nanopore MinION. *bioRxiv*:019281.
- Conflict of Interest Statement:** Author SG-S was employed by JSI medical systems GmbH. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ton, Cree, Gronert-Sum, Merriman, Stamp and Kennedy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



PASSPORT-seq: A Novel High-Throughput Bioassay to Functionally Test Polymorphisms in Micro-RNA Target Sites

Joseph Ipe¹, Kimberly S. Collins^{1,2}, Yangyang Hao^{3,4}, Hongyu Gao^{3,4}, Puja Bhatia¹, Andrea Gaedigk⁵, Yunlong Liu^{3,4} and Todd C. Skaar^{1*}

¹ Division of Clinical Pharmacology, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States, ² Department of Pharmacology and Toxicology, Indiana University School of Medicine, Indianapolis, IN, United States, ³ Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, United States, ⁴ Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, United States, ⁵ Division of Clinical Pharmacology, Toxicology and Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO, United States

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch-Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Dylan Glubb,
QIMR Berghofer Medical Research
Institute, Australia
Eric R. Gamazon,
The University of Chicago,
United States

*Correspondence:

Todd C. Skaar
tskaar@iu.edu

Specialty section:

This article was submitted to
Pharmacogenetics
and Pharmacogenomics,
a section of the journal
Frontiers in Genetics

Received: 30 November 2017

Accepted: 29 May 2018

Published: 15 June 2018

Citation:

Ipe J, Collins KS, Hao Y, Gao H,
Bhatia P, Gaedigk A, Liu Y and
Skaar TC (2018) PASSPORT-seq:
A Novel High-Throughput Bioassay
to Functionally Test Polymorphisms
in Micro-RNA Target Sites.
Front. Genet. 9:219.
doi: 10.3389/fgene.2018.00219

Next-generation sequencing (NGS) studies have identified large numbers of genetic variants that are predicted to alter miRNA-mRNA interactions. We developed a novel high-throughput bioassay, PASSPORT-seq, that can functionally test in parallel 100s of these variants in miRNA binding sites (mirSNPs). The results are highly reproducible across both technical and biological replicates. The utility of the bioassay was demonstrated by testing 100 mirSNPs in HEK293, HepG2, and HeLa cells. The results of several of the variants were validated in all three cell lines using traditional individual luciferase assays. Fifty-five mirSNPs were functional in at least one of three cell lines (FDR ≤ 0.05); 11, 36, and 27 of them were functional in HEK293, HepG2, and HeLa cells, respectively. Only four of the variants were functional in all three cell lines, which demonstrates the cell-type specific effects of mirSNPs and the importance of testing the mirSNPs in multiple cell lines. Using PASSPORT-seq, we functionally tested 111 variants in the 3' UTR of 17 pharmacogenes that are predicted to alter miRNA regulation. Thirty-three of the variants tested were functional in at least one cell line.

Keywords: SNP, functional testing, genetic variants, miRNA, high-throughput screening assays, 3' UTR

INTRODUCTION

Large scale sequencing studies and genome-wide association studies (GWASs) have identified 1000s of genotype-phenotype associations (Welter et al., 2014). Some of the phenotype-associated variants alter gene function and many of them are in linkage disequilibrium with the functional variants. The functional impacts of variants can be predicted using bioinformatic algorithms, but the *in silico* predictions are often incorrect and need experimental validation. While there are several experimental methods to functionally test variants, most do not have the capacity to simultaneously test the large number of variants.

Nearly 90% of genetic variants associated with phenotypes have been described to be located in non-coding regions such as the untranslated regions (UTRs) (Hindorff et al., 2009). Variants, including single nucleotide polymorphisms (SNPs) within non-coding regions, can impact gene

expression in several ways; one example is by altering the interaction between mRNAs and micro-RNAs (miRNAs). Polymorphisms within miRNA-binding sites have been implicated in diseases such as cancer (Pelletier and Weidhaas, 2010; Iuliano et al., 2013), Alzheimer's disease (Liu et al., 2017), and diabetes (Elek et al., 2015).

miRNAs are 21–23 nucleotide long RNAs that post-transcriptionally silence genes or reduce their expression levels by complementarily binding to target sites within mRNAs. More than 29,000 human mRNAs are collectively targeted by are over 2500 miRNAs (Kozomara and Griffiths-Jones, 2014). Several different miRNA binding sites may be present on one mRNA and many contain genetic variations. To date, over 400,000 SNPs have been identified in miRNA binding sites (Liu et al., 2012). Interestingly, only about 32,000 have a minor allele frequency greater than 1% classifying most of them as rare variants (Liu et al., 2012). Thus, tests to identify functional SNPs affecting miRNA binding, here referred to as mirSNPs, will involve screening a large number of variants.

Testing these large number of mirSNPs using GWAS requires statistical correction for multiple testing, such as the Bonferroni correction. The low minor allele frequency of many causal variants, and routine multiple comparisons corrections make it very difficult, or impossible, to statistically identify functionally relevant variants in genome-wide studies. Consequently, in GWAS, impractically large numbers of subjects from diverse populations would be required to identify rare functional variants that are statistically significant. Lowering the statistical threshold or not correcting for multiple comparisons increases the sensitivity to detect rare variant associations, but results in the detection of many false positives signals. Despite some technical challenges, high-throughput *in vitro* approaches have been implemented that are specific to variants in certain non-coding regions, such as splice-junctions (Soemedi et al., 2014) and promoters (Kwasniewski et al., 2012; Melnikov et al., 2012). However, we are not aware of any high-throughput assays available to functionally test variants in miRNA binding sites (Ipe et al., 2017).

We developed PASSPORT-seq (parallel assessment of polymorphisms in miRNA target-sites by sequencing), a high-throughput bioassay that involves pooled synthesis, parallel cloning and single-well transfection followed by next-generation sequencing (NGS) to functionally test 100s of mirSNPs at once. This assay produced results that are reproducible and consistent with luciferase reporter assays, a gold-standard platform widely used to assess gene expression *in vitro*. We also demonstrate the application of this assay to test 111 genetic variants that are predicted to alter miRNA regulation of 17 pharmacogenes.

MATERIALS AND METHODS

Selection of mirSNPs

RNA samples from thirty human livers were sequenced using SOLiD® technology (Thermo Fisher Scientific, Waltham, MA, United States). SNPs in the 3' UTRs were identified (O'Leary et al., 2016) and an 8-base pair region on either side of the

reference and variant alleles was analyzed using TargetScan (Lewis et al., 2005) to identify SNPs that were in miRNA seed binding regions. SNPs that altered the predicted miRNA seed binding sites were considered for further analysis. For assay development, 84 SNPs that were associated with allele-specific expression in the sequencing dataset were selected. A flowchart representing the selection process of the 84 test mirSNPs is shown in Supplementary Figure 1. In addition, we selected 16 mirSNPs from the SomamiR database (Bhattacharya et al., 2013) that have been linked with cancer. The list of 100 SNPs used for assay development are listed in Supplementary Table 1. Similarly, 111 mirSNPs located in the 3' UTR regions of 17 pharmacogenes—the core absorption, distribution, metabolism, and excretion (ADME) genes¹, PXR, CAR, and HNF4α which showed allele specific expression in the sequencing dataset were selected to demonstrate the application of the assay. The list of these 111 SNPs are listed in Supplementary Table 4. The RNA analysis and genotyping was approved by the Indiana University Institutional Review Board.

Test Sequence Design

The 5' and 3' flanking regions for each SNP were obtained from dbSNP. A 32-nucleotide region which contained either the variant or reference nucleotide flanked by nine nucleotides on the 3' end and 22 nucleotides on the 5' end was selected as the test sequence. Two-hundred such regions (100 reference and 100 variant) were selected to test 100 SNPs. Universal primer binding regions were added on the 5' (GTAATTCTAGGAGCTC) and 3' (CGTTCTAGAGTCGGG) end of each test region. The final test fragment was 63 nucleotides in length (see Supplementary Figure 2). The 200 test fragments were commercially synthesized as pooled single-stranded DNA oligonucleotides (Oligomix®, LC Sciences, Houston, TX, United States). The pool contained 10–50 attomoles of each sequence. The oligonucleotides were synthesized as single-stranded DNA and was diluted 1:5. One µL of the diluted Oligomix® was amplified in a 50 µL PCR reaction using 0.3 µM universal primers and 25 µL 2X CloneAmp™ HiFi PCR premix (Takara, Mountain View, CA, United States). PCR conditions used were: 98°C (10 s), 53°C (5 s), and 72°C (5 s) for 35 cycles.

Seven SNPs were also tested in individual luciferase assays using 63-nucleotide long single stranded oligonucleotides that were individually synthesized (reference and variant); (Integrated DNA Technologies, Coralville, IA, United States), made double stranded as described for pooled oligonucleotides, and cloned into the pIS-0 plasmid.

Plasmid Library Preparation

The pIS-0 vector (plasmid 12178; Addgene, Cambridge, MA, United States) (Yekta et al., 2004) (see Supplementary Figure 3) was linearized with *SacI*-HF® and *BmtI*-HF® restriction endonucleases (New England Biolabs, Ipswich, MA, United States) and purified using QIAquick® PCR spin columns (Qiagen, Germantown, MD, United States). Plasmid assembly was performed using 40 ng of linearized plasmid and

¹<http://www.pharmaadme.org/joomla/>

2 μL of unpurified PCR product containing double-stranded test oligonucleotides using the NEBuilder[®] HiFi DNA assembly kit (NEB, Ipswich, MA, United States) per manufacturer's instructions. The universal primers used to amplify the test-oligonucleotide pool also served as the flanking homology regions for the NEBuilder[®] assembly. Two μL of the NEBuilder[®] assembly product were transformed into 60 μL chemically competent *E. coli* (transformation efficiency $> 5 \times 10^8$ cfu/ μg) (Takara, Mountain View, CA, United States) and plated on six standard 100 mm LB-agar plates containing 100 $\mu\text{g}/\text{ml}$ ampicillin. After overnight incubation, all colonies were dislodged from the plates by adding 2 ml LB-broth containing 100 $\mu\text{g}/\text{ml}$ ampicillin and agitation using 10–20 ColiRollers[™] glass beads (EMD Millipore, Billerica, MA, United States). The colonies harvested from the six plates in LB-broth were pooled together. The liquid culture was incubated at 37°C for 5 h after which plasmids were isolated using 10 QIAprep[®] Spin miniprep columns (Qiagen, Germantown, MD, United States) as per manufacturer's instructions. Column elutions were combined to create the plasmid library that was used for downstream experiments. The plasmid DNA concentration was determined using a Quant-iT[™] DNA Broad Range kit (Thermo Fisher Scientific, Waltham, MA, United States).

Sanger Sequencing

To determine the representation of the test constructs in the plasmid library, 28 individual colonies were grown in LB-broth containing 100 $\mu\text{g}/\text{ml}$ ampicillin. Plasmids were isolated using QIAprep[®] Spin miniprep columns (Qiagen, Germantown, MD, United States) as per manufacturer's instructions and Sanger sequenced using a primer (GTGGTTTGTCCAACTCATC) near the test insert (ACGT, Inc., Wheeling, IL, United States).

Transfection of Cells in Culture

The plasmid library was used to transfect three different human cell lines: HEK293 (embryonic kidney), HepG2 (liver carcinoma), and HeLa (cervical cancer). Cells were seeded at a density of 0.9×10^5 cells per well into 24-well plates. The cells were transfected 24 h after plating with 500 ng/well of the pIS-0 plasmids. Ten ng of pGL4.74, a Renilla luciferase reporter plasmid, was added to each well as a transfection control. Transfection was performed using 50 μL transfection-mix in Opti-MEM[®] (Life Technologies, Carlsbad, CA, United States) containing 1.5 μL Lipofectamine[®] 3000 (Life Technologies, Carlsbad, CA, United States) per the manufacturer's instructions. Opti-MEM[®] and culture media were used with no antibiotics.

RNA Isolation and cDNA Synthesis

Transfected cells were incubated for 48 h, lysed *in situ* and total RNA isolated using a RNeasy[®] purification kit with the optional DNase treatment (Qiagen, Germantown, MD, United States). RNA was quantified using the Quant-iT[™] RNA Broad Range kit (Thermo Fisher Scientific, Waltham, MA, United States) and cDNA synthesized from 800 ng of total RNA using the QuantiTect[®] Reverse Transcription kit (Qiagen, Germantown, MD, United States).

Molecular Barcoding

Using the cDNAs from the transfected cells, the miRNA binding sites within the 3' UTR of the luciferase genes were amplified in 50 μL PCR reactions using 0.3 μM flanking universal primer and 25 μL 2X CloneAmp[™] HiFi PCR premix (Takara, Mountain View, CA, United States). In a separate reaction for each sample, 2 μL of cDNA and 1 pg of the input plasmid pool was used as PCR template. PCR conditions used were: 98°C (10 s), 54°C (5 s), and 72°C (5 s) for 25 cycles. A 6-nucleotide unique molecular barcode was added to the 5'-end of both the forward and reverse primer (see Supplementary Table 2). The input pools ($n = 4$ replicates) and the five biological replicates in the three different cell lines were each 'barcoded' by a unique pair of sequences. The barcoded PCR products were purified using a MinElute[®] PCR Purification kit (Qiagen, Germantown, MD, United States). The barcoded libraries were combined in equimolar concentrations to create a sequencing pool with 19 different molecular barcodes. A schematic representation of the steps involved in creating this sequencing pool is shown in Figure 1.

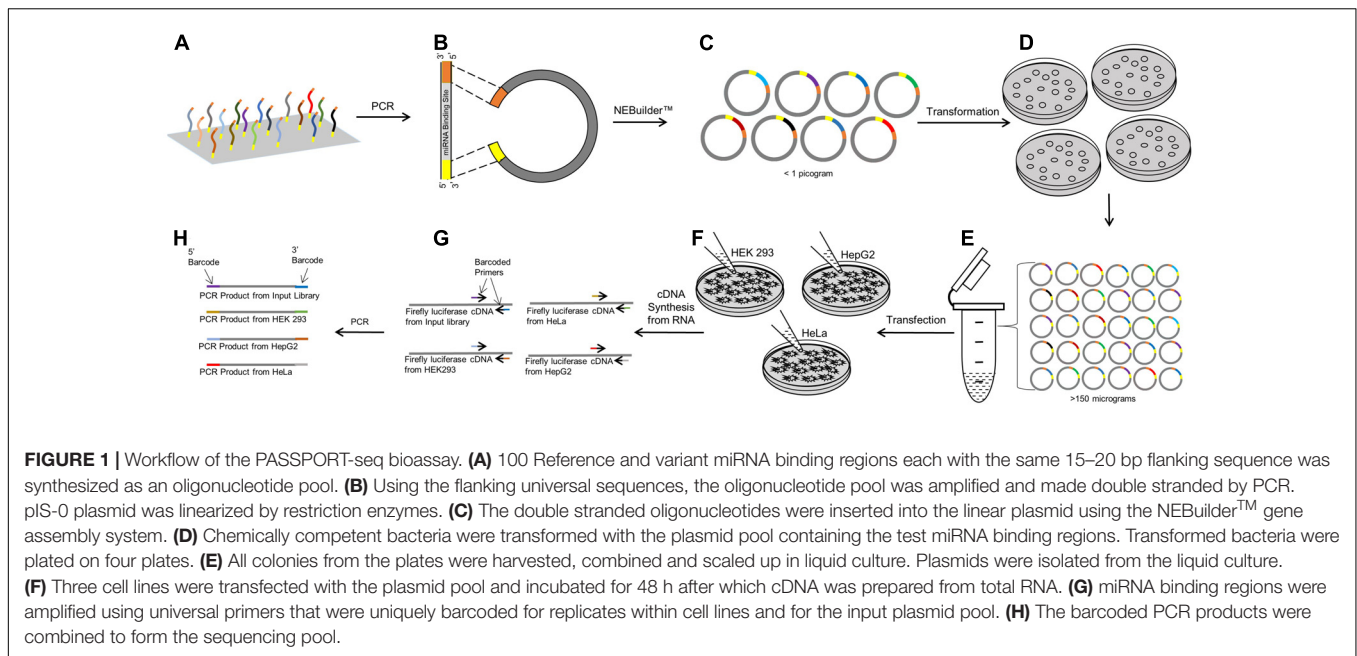
Next-Generation Sequencing

The pooled PCR products were sequenced using a modified protocol for the Ion Proton[™] system (Thermo Fisher Scientific, Waltham, MA, United States). Briefly, the sequencing library was created by end-polishing the barcoded PCR products, followed by adapter ligation and amplification. The resulting library was quantified and its quality assessed with the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). Eight microliters of the 100 pM library were then applied to Ion Sphere Particles to prepare the sequencing template. The template was amplified using Ion OneTouch 2. The Ion Sphere Particles with the amplified template were loaded onto an IonPI[®] chip and sequenced on the Ion Proton system per manufacturer's instructions. Approximately, 41 million reads were generated from each sequencing run. Raw reads were generated as fastq files for bioinformatic analysis. Sequencing data has been made publicly available through GEO (Accession No. GSE111845).

Bioinformatic Analysis

The raw reads were aligned to the reference library containing the 200 test sequences (TMAP- Ion Torrent Suite[®], Thermo Fisher Scientific, Waltham, MA, United States). The reads that aligned to the reference library were filtered to retain reads with a mapping quality greater than 20 and further filtered to include only those sequences with perfect barcodes at both ends.

Differential expression analysis compared the expression of each variant to its respective reference allele for all SNPs. To account for differences in the concentrations of the variant and reference plasmids that were used for the transfections, a plasmid input correction factor for each target site was calculated as the average of the number of reads from the variant plasmid divided by the number of reads from the reference plasmid across four



replicates of the plasmids. The reads from the variant alleles for all biological replicates were divided by the input correction factor. The corrected read counts were fit into a generalized linear model using EdgeR (Robinson et al., 2010) assuming a negative binomial distribution. Biological replicates and the genotype were used as covariates. *p*-Values and \log_2 fold-change of the variant alleles compared to the respective reference alleles were derived using a likelihood ratio test on the genotype variable in the generalized linear model. The *p*-values were corrected for a false discovery rate (FDR) using the Benjamini and Hochberg algorithm (Benjamini and Hochberg, 1995). The EdgeR script can be found in Supplementary File 1.

The two sequencing runs, each with five biological replicates, were analyzed together by fitting the number of reads for 10 pairs of variant and reference alleles (five from each experiment). The different sequencing runs were included as an additional covariate. The statistical analysis was performed as described above.

Luciferase Reporter Assay

Genetic variants, including mirSNPs have been functionally tested using a reporter plasmid such as the pIS-0 vector (Yekta et al., 2004; Adams et al., 2007; Ramamoorthy et al., 2012). This plasmid contains the firefly luciferase gene whose expression can be quantified either by qPCR or by the luciferase reporter assay. The reference or variant allele version of the predicted miRNA binding sites were cloned into the 3' UTR of the luciferase gene within the plasmid. The plasmids were then transfected into cells as described above. Forty-eight hours after transfection, cells were lysed *in situ* and Dual-Luciferase® assays were performed per manufacturer's instructions (Promega, Madison, WI, United States). The luciferase reporter activity was measured using a 96-well plate-reader (BioTek, Winooski, VT, United States). The firefly luciferase activity was normalized

to that of Renilla luciferase in each well. The ratio of the normalized luciferase activity from the variant and reference plasmid provides a relative measure of SNP-mediated differential mRNA expression.

RESULTS

The traditional luciferase reporter assay is useful in low-throughput experiments, but is not a practical and cost-effective method to test the 1000s of mirSNPs identified at a genome-wide scale. As a novel approach, we modified the luciferase reporter assay to develop PASSPORT-seq that can functionally test 100s of mirSNPs in parallel. Since one of the mechanisms of miRNA regulation is by degrading mRNA, this assay was specifically designed to evaluate the impact of genetic variation in miRNA binding sites on mRNA expression. We recognize that miRNAs also alter mRNA translation, however measuring protein levels does not distinguish between the impact on mRNA vs. translation and thus would not provide the same mechanistic insights. We identified 100 variants in predicted seed sequences of miRNA binding sites, and cloned the binding sites into the pIS-0 luciferase plasmid; each contained either the reference or variant nucleotide of 100 selected mirSNPs. The pool of the resulting 200 plasmids was then transfected into three cell lines and the luciferase gene expression measured by NGS. A difference in mRNA expression between the reference and the variant plasmids indicated a functional mirSNP.

Cloning Efficiency and Plasmid Representation

To test the efficiency of the parallel cloning, plasmids from 28 individual colonies were isolated and the sequence of the inserts were determined by Sanger sequencing. Out of the 28

colonies, 25 contained inserts without errors; of those, 24 were unique sequences suggesting that the cloning efficiency was high and there was negligible sequence-bias in the plasmid pool. Furthermore, as described below, all 200 sequences were observed in the NGS of the entire pool.

Reproducibility Across Biological Replicates

To test the reproducibility of the PASSPORT-seq bioassay, we first performed the assay with five biological replicates in each of the three cell lines and compared the number of reads from two of the five biological replicates. The input plasmid libraries had representation of all 100 allelic pairs. The R^2 value for the comparison of sets of two of the five input normalized-biological replicates within the same sequencing run was between 0.68 and 0.98 ($p < 0.05$; Supplementary Figure 4) demonstrating highly reproducible results within a run.

Reproducibility Across Runs

Next, we repeated the PASSPORT-seq assay again with another five biological replicates in each of the three cell lines to validate the observed results. A separate cDNA and sequencing library was created for the experiments. A strong correlation ($R^2 = 0.98$; $p < 0.05$) was observed in the results from the plasmid pool from the first sequencing run and those from the second sequencing run (see Supplementary Figure 5). There was a high correlation ($R^2 = 0.74$; $p < 0.05$) in the results between the first and second set of biological replicates (Figure 2 and Supplementary Figure 6). This strong correlation between results of the first sequencing run with those of the second sequencing run across the three cell lines demonstrates the high reproducibility of the observed results.

Identification of Functional mirSNPs

Of the 100 mirSNPs tested, 69 showed statistically significant ($p < 0.05$) differences in expression between the variant and its respective reference allele in at least one cell line (Figure 2B).

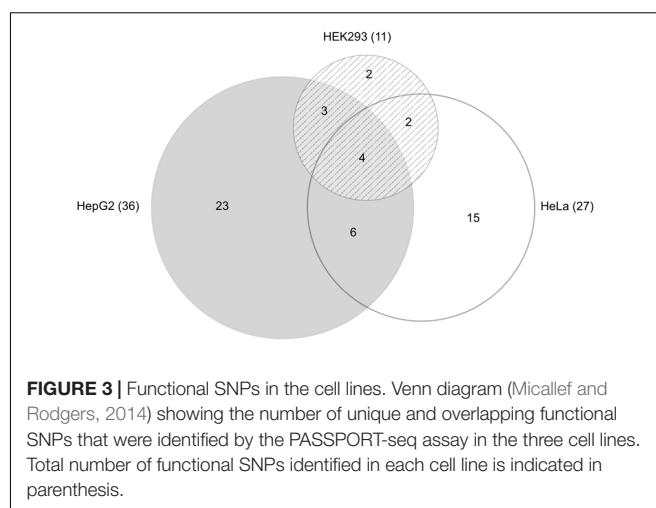


FIGURE 3 | Functional SNPs in the cell lines. Venn diagram (Micallef and Rodgers, 2014) showing the number of unique and overlapping functional SNPs that were identified by the PASSPORT-seq assay in the three cell lines. Total number of functional SNPs identified in each cell line is indicated in parenthesis.

In HEK293, HepG2, and HeLa a significant effect was seen in 27, 44, and 39 mirSNPs, respectively (see Supplementary Figure 7). Due to the large number mirSNPs tested, the results were corrected using the Benjamini and Hochberg procedure across cell lines. This conservative threshold ($FDR \leq 0.05$) was met by 55 mirSNPs in at least one cell line with 11, 36, and 27 in HEK293, HepG2, and HeLa cells, respectively. Because these variants were informatically predicted to be functional, this may be an overly conservative statistical correction. The effect of the mirSNPs was cell line- specific; four SNPs were functional across all three cell lines, while others were functional in either two cell lines or unique to one cell line (Figure 3).

Validation With Traditional Luciferase Assays

Twenty-one of the results were validated using traditional individual luciferase transfection experiments. The variant and

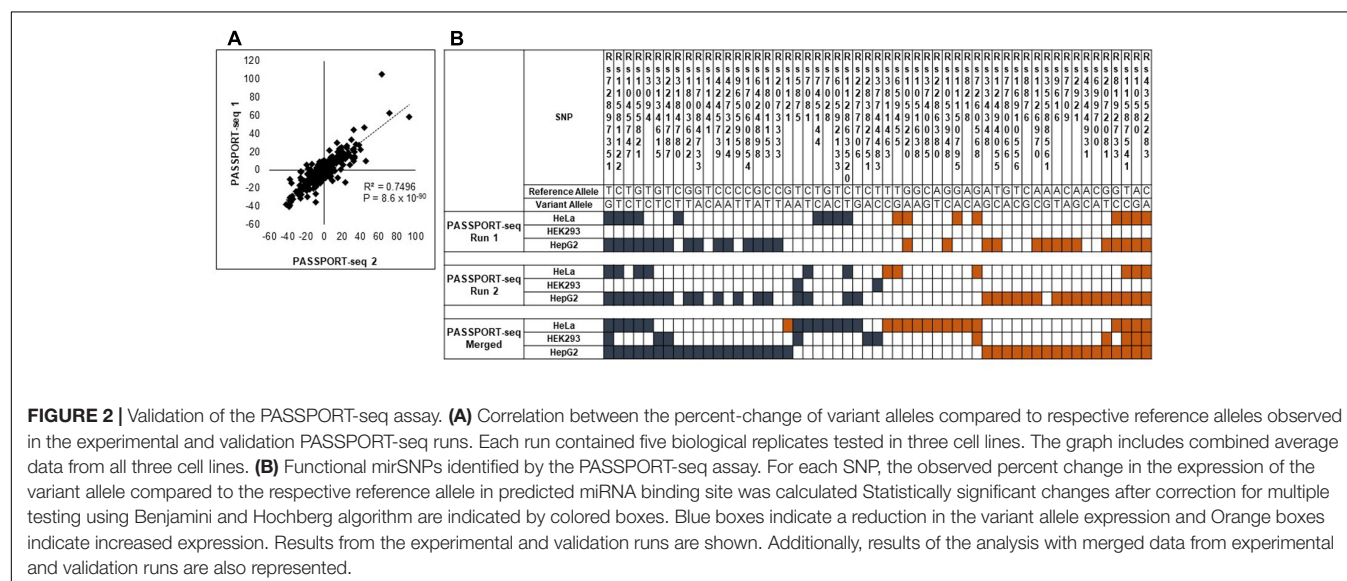
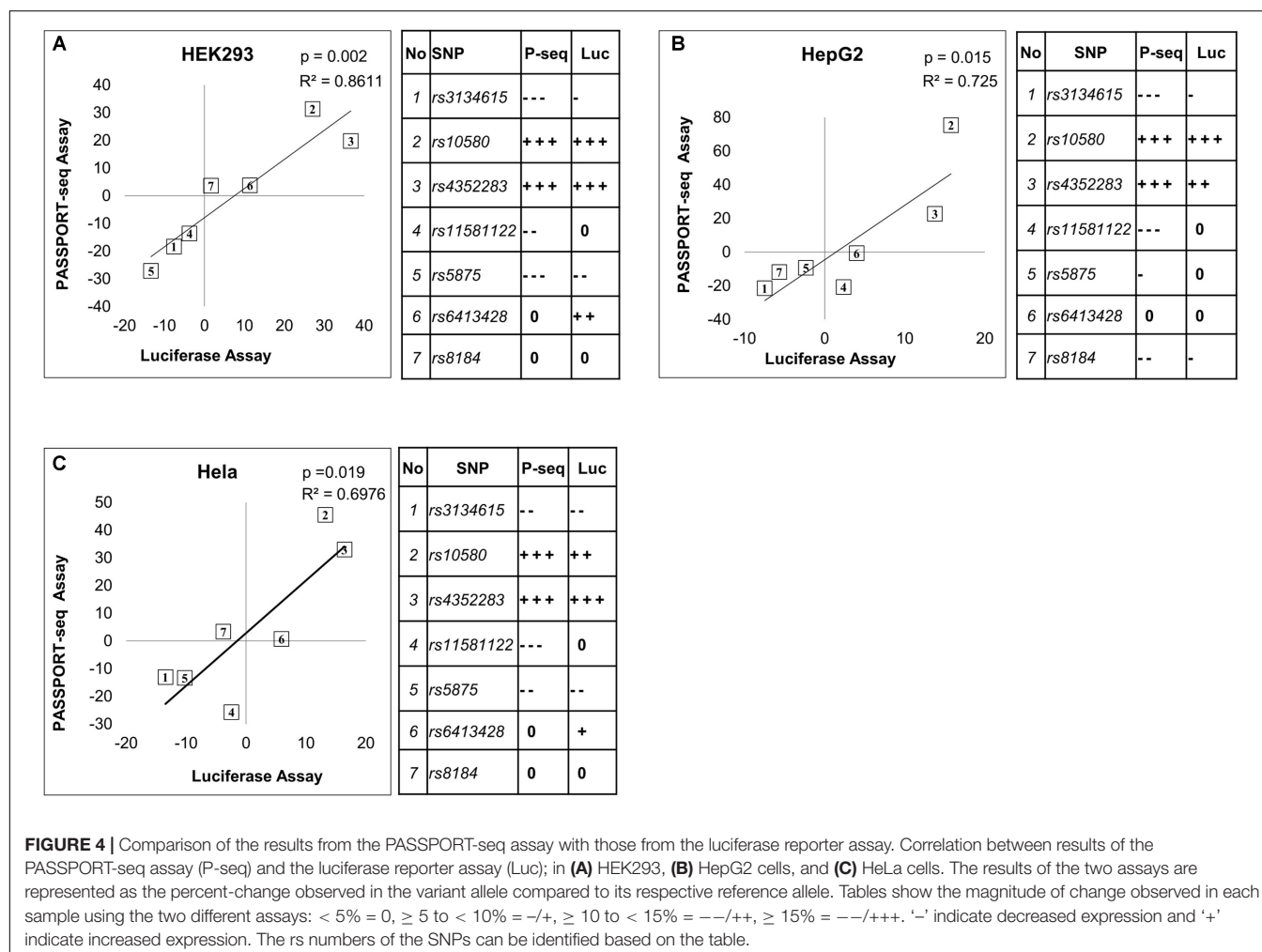


FIGURE 2 | Validation of the PASSPORT-seq assay. **(A)** Correlation between the percent-change of variant alleles compared to respective reference alleles observed in the experimental and validation PASSPORT-seq runs. Each run contained five biological replicates tested in three cell lines. The graph includes combined average data from all three cell lines. **(B)** Functional mirSNPs identified by the PASSPORT-seq assay. For each SNP, the observed percent change in the expression of the variant allele compared to the respective reference allele in predicted miRNA binding site was calculated. Statistically significant changes after correction for multiple testing using Benjamini and Hochberg algorithm are indicated by colored boxes. Blue boxes indicate a reduction in the variant allele expression and Orange boxes indicate increased expression. Results from the experimental and validation runs are shown. Additionally, results of the analysis with merged data from experimental and validation runs are also represented.

reference allele binding sites of the selected mirSNPs were individually cloned into the 3' UTR of the luciferase gene within the pIS-0 plasmid and transfected into HEK293, HepG2, and HeLa cells. This included seven of the miRNA target sequences, each with reference and variant sequences, tested in three cell lines for a total of 21 validations. Within each cell line, the luciferase activity in the cells transfected with the reference plasmid was compared to the activity in the cells transfected with the variant plasmids. The effect of the variants in these individual luciferase assays were compared with the results from the PASSPORT-seq assay (**Figure 4**). In 17 of the 21 comparisons, the statistical significance of the results and the direction of the effect of the variant matched the PASSPORT-seq results (Supplementary Figure 8 and Supplementary Table 3). In an additional two comparisons (rs3134615 in HeLa and HEK293 cells), the results were statistically significant in one assay, but not the other, but the direction and magnitude of effect of the variant in the PASSPORT-seq was similar to the luciferase assay in both cell lines (Supplementary Figure 8). Thus, the results were very similar in 19 of the 21 comparisons (>90%).

Application of PASSPORT-seq

To demonstrate the utility of this assay, mirSNPs predicted to alter miRNA regulation of 17 pharmacogenes were selected for functional testing from the RNA-seq dataset described earlier. These variants were functionally tested using the PASSPORT-seq assay in HeLa, HepG2, HEK293, and HepaRG (hepatic cells that retain characteristics of primary human hepatocytes) cells. Out of the 111 genetic variants tested, the effect of 33 variants were statistically significant in at least one cell line, including 6, 13, 12, and 27 in HeLa, HepG2, HEK293, and HepaRG cells, respectively (**Figure 5** and Supplementary Table 4). The effects of several mirSNPs were shown to be cell line-specific. Only four mirSNPs had significant effects in all the four cell lines (**Figure 6**). The effect of a genetic variant (rs12979270), located in the 3' UTR of the pharmacogene- CYP2B6, was shown to be statistically significant in HepaRG cells. A recent study shows that this variant, could explain part of the interindividual variability seen in the activity of this critical pharmacogene (Burgess et al., 2017). These results demonstrate the potential of this assay to identify clinically relevant functional genetic variants.



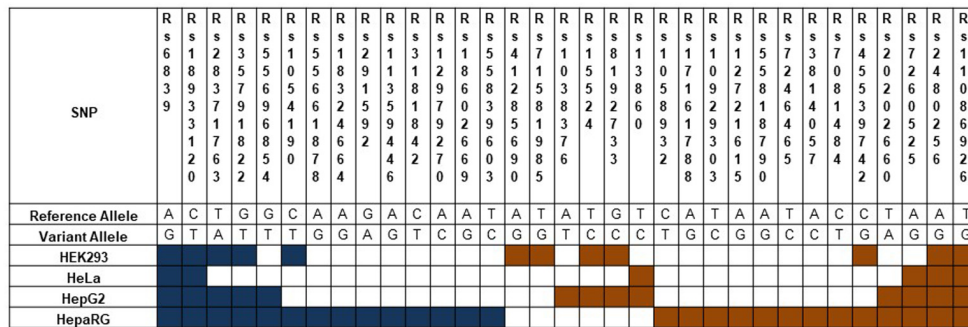


FIGURE 5 | Functional miRSNPs in pharmacogenes identified by the PASSPORT-seq assay. For each SNP, the observed percent change in the expression of the variant allele compared to the respective reference allele in predicted miRNA binding site was calculated. Statistically significant changes after correction for multiple testing using Benjamini and Hochberg algorithm are indicated by colored boxes. Blue boxes indicate a reduction in the variant allele expression and Orange boxes indicate increased expression.

DISCUSSION

We developed PASSPORT-seq to screen 100s of SNPs that are predicted to alter miRNA-mRNA interactions. The availability of pooled oligonucleotide synthesis and the NEBuilder® gene assembly system have made this assay possible. Our assay builds and substantially improves the technologies that have had only limited success (Reid et al., 2009). For example, in previous high-throughput splicing assays, short inserts were underrepresented during the library construction using traditional cloning methods (Chen and Chasin, 1994; Ke et al., 2011; Soemedi et al., 2014). In contrast, our library had representation of all allele pairs. This may be explained by the cloning method we utilized, i.e., the NEBuilder® gene assembly system that produces covalently closed circular plasmids as opposed to traditional cloning methods, which yield a nicked-relaxed plasmid topology. This change in topology may result in a more efficient plasmid uptake by chemically competent bacteria (Hanahan, 1983; Xie et al., 1992; Kobori and Nojima, 1993). Better transformation efficiency increases the probability of both the variant and reference plasmids being represented in the resulting pool, which is a critical prerequisite for studying the allele-specific activity of miRNAs.

The activity of miRNAs on target sequences have been studied using reporter assays where the target sites of interest are cloned into the 3' UTR of a reporter gene followed by quantification of the reporter activity. Typical reporter assays also overexpress the miRNAs that are predicted to regulate the target site (Cloonan et al., 2008; Loya et al., 2009; Baccarini and Brown, 2010). Such overexpression, however, may not provide a physiological context to the miRNA-mRNA interaction. For example, high miRNA expression levels may force interactions with mRNAs that do not normally occur. They can also compete with the miRNA processing machinery or binding sites and alter normal miRNA function. In contrast, our assay was performed in the endogenous miRNA expression background, which provides a more physiologically relevant context of the results. In addition, we used multiple different cell lines to

allow parallel testing to identify cell line-specific effects of the mirSNPs.

miRNAs regulate gene expression by either degrading the target mRNAs or by binding to mRNAs and blocking translation. Traditional luciferase assays test the effect of miRNA by measuring differences in its target protein activity. However, the differences in protein activity due to mRNA degradation and those from translational blockage will be indistinguishable using

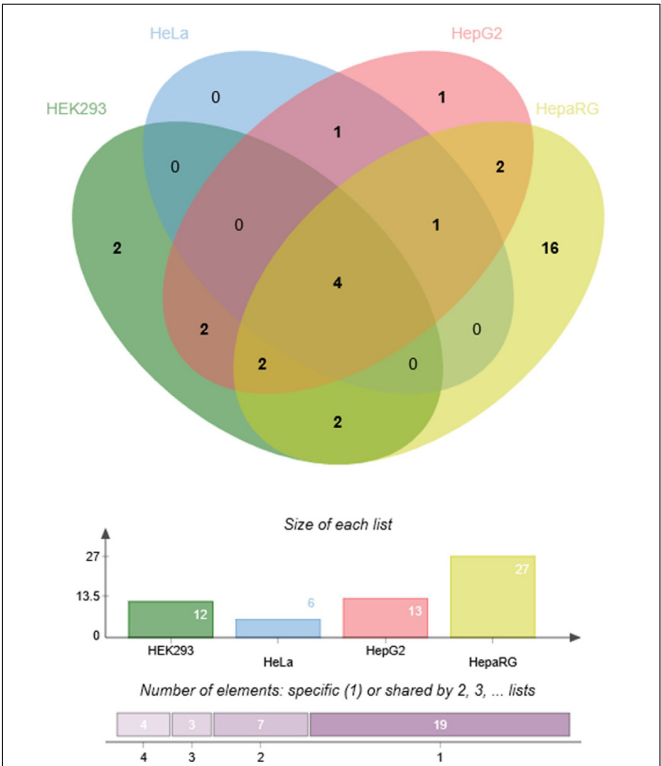


FIGURE 6 | Cell line-specificity of functional mirSNPs in pharmacogenes. Venn diagram (Bardou et al., 2014) showing the number of unique and overlapping functional mirSNPs in the four cell lines tested.

luciferase activity assays. The PASSPORT-seq assay provides additional evidence of the mechanism underlying the effect of the variant by specifically detecting only the changes in mRNA transcript levels. We demonstrated that the results obtained with our PASSPORT-seq assay did reflect those obtained with the traditional luciferase reporter assay set-up. As described above, the PASSPORT-seq assay quantifies the relative expression of luciferase mRNAs, whereas the luciferase assay measures the luciferase enzyme activity. Thus, it is not surprising that there may be differences in the magnitude of effects between the different assays. For example, the effect on the luciferase activity could be larger due to both the degradation of the mRNA and the blockade of translation. In contrast, the effect on the protein levels and activity could be smaller due to delays from the time of changes in mRNA levels until the changes in protein levels and activities are observed. Typically, changes in mRNA expression due to endogenous miRNA-mediated regulation are subtle (<30%) (Baek et al., 2008; Bartel, 2009; Denzler et al., 2016). Consequently one would expect relatively small effect sizes of the variants, which was what was seen in many of these variants. However, there are many examples demonstrating the clinical impact of these types of variants (Bhattacharya et al., 2014).

One of the key findings of these studies is the cell line-specific function of mirSNPs. This is likely in part due to the cell type specific variation in miRNA expression profiles resulting in the effect of a mirSNP being evident in one cell line and not in another (Landgraf et al., 2007; Ludwig et al., 2016). We observed that the identity and number of functional mirSNPs reproducibly varied across the different cell types. This demonstrates one of the strengths of PASSPORT-seq in that it can identify cell line-specific effects of mirSNPs. These differences were validated using a second PASSPORT-seq run that reproduced the cell line effect. Additionally, the cell line-specific effect was also observed in the application of the assay to test 111 mirSNPs in pharmacogenes. The tissue/cell specificity of mirSNP function could also explain why the effects of mirSNPs are not always consistent across studies. This further complicates the bioinformatics predictions of the functional impact of the mirSNPs. Thus, when using this assay, the cell line must be carefully chosen to reflect the cell type of interest regarding the central biological hypothesis of the study. Since studies have shown that mirSNPs affect a wide variety of biological processes such as cancer, neurodegenerative disorders, infectious diseases, cardiovascular disease, and metabolic disorders (Bhattacharya et al., 2014), the *in vitro* model for testing the mirSNPs is an important consideration.

As with any *in vitro* assay, PASSPORT-seq has some limitations. First, it detects only changes in mRNA, rather than protein levels. miRNA-mRNA interactions are known to cause mRNA destabilization, but can also lead to translational repression (Lim et al., 2005; Bartel, 2009). Since this assay does not detect changes in translation, it may underestimate the functional impact of some of the mirSNPs. Second, the variants could be affecting mRNA stability by mechanisms other

than by altering miRNA targeting. For example, it could be altering RNA binding protein function that could alter the mRNA stability. Although this would need additional validation experiments to determine the mechanism of action, it would still be of biological value. Last, like other studies using the pISO plasmid, the miRNA binding site is tested in the context of the luciferase mRNA, rather than the endogenous mRNA.

In summary, the PASSPORT-seq assay is a powerful tool that bridges bioinformatic predictions and high-throughput mechanistic investigation of functional genetic variants that affect miRNA-mRNA interactions. Future efforts will be aimed toward further increasing the capacity of the assay and identifying translational effects. This assay also has the potential to be modified to be applicable to genetic variants in other functional genomic regions such as promoters and splice junctions. Collectively, these assays will be key to elucidating the mechanisms underlying the genetic contribution to the inter-individual phenotypic variability.

AUTHOR CONTRIBUTIONS

Jl, PB, AG, and TS designed the assay. Jl, KC, and PB performed the *in vitro* experiments. YH, HG, and YL performed the bioinformatic analysis. Jl, KC, and TS performed the data analysis. Jl, KC, YH, YL, AG, and TS wrote and/or edited the manuscript.

FUNDING

This work was supported by National Institutes of Health [NIH/NIGMS R01-GM088076 (TS), NIH/NIGMS F31-GM119401 (Burgess), NIH/NCRR RR025761]; and Vera Bradley Foundation for Breast Cancer (Jl). Funding for open access charge: National Institutes of Health.

ACKNOWLEDGMENTS

We thank the Center for Medical Genomics at Indiana University School of Medicine for their support on DNA sequencing. Biospecimens were stored in the CTSI Specimen Storage Facility which is supported, in part, by grant NIH/NCRR RR025761. We also thank, Drs. Michael Eadon, Eric Benson, Thomas De Luca, and Marelize Swart for their assistance via intellectual discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00219/full#supplementary-material>

REFERENCES

- Adams, B. D., Furneaux, H., and White, B. A. (2007). The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor- α (ER α) and represses ER α messenger RNA and protein expression in breast cancer cell lines. *Mol. Endocrinol.* 21, 1132–1147. doi: 10.1210/me.2007-0022
- Baccarini, A., and Brown, B. D. (2010). Monitoring microRNA activity and validating microRNA targets by reporter-based approaches. *Methods Mol. Biol.* 667, 215–233. doi: 10.1007/978-1-60761-811-9_15
- Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71. doi: 10.1038/nature07242
- Bardou, P., Mariette, J., Escudie, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15:293. doi: 10.1186/1471-2105-15-293
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Bhattacharya, A., Ziebarth, J. D., and Cui, Y. (2013). SomamiR: a database for somatic mutations impacting microRNA function in cancer. *Nucleic Acids Res.* 41, D977–D982. doi: 10.1093/nar/gks1138
- Bhattacharya, A., Ziebarth, J. D., and Cui, Y. (2014). PolymiRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 42, D86–D91. doi: 10.1093/nar/gkt1028
- Burgess, K. S., Ipe, J., Swart, M., Metzger, I. F., Lu, J., Gufford, B. T., et al. (2017). Variants in the CYP2B6 3'UTR alter in vitro and in vivo CYP2B6 activity: potential role of microRNAs. *Clin. Pharmacol. Ther.* doi: 10.1002/cpt.892 [Epub ahead of print].
- Chen, I. T., and Chasin, L. A. (1994). Large exon size does not limit splicing in vivo. *Mol. Cell. Biol.* 14, 2140–2146. doi: 10.1128/MCB.14.3.2140
- Cloonan, N., Brown, M. K., Steptoe, A. L., Wani, S., Chan, W. L., Forrest, A. R., et al. (2008). The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. *Genome Biol.* 9:R127. doi: 10.1186/gb-2008-9-8-r127
- Denzler, R., Mcgeary, S. E., Title, A. C., Agarwal, V., Bartel, D. P., and Stoffel, M. (2016). Impact of microRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. *Mol. Cell.* 64, 565–579. doi: 10.1016/j.molcel.2016.09.027
- Elek, Z., Nemeth, N., Nagy, G., Nemeth, H., Somogyi, A., Hosszufalusi, N., et al. (2015). Micro-RNA binding site polymorphisms in the WFS1 gene are risk factors of diabetes mellitus. *PLoS One* 10:e0139519. doi: 10.1371/journal.pone.0139519
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580. doi: 10.1016/S0022-2836(83)80284-8
- Hindorf, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9362–9367. doi: 10.1073/pnas.0903103106
- Ipe, J., Swart, M., Burgess, K. S., and Skaar, T. C. (2017). High-throughput assays to assess the functional impact of genetic variants: a road towards genomic-driven medicine. *Clin. Transl. Sci.* 10, 67–77. doi: 10.1111/cts.12440
- Iuliano, R., Vismara, M. F., Dattilo, V., Trapasso, F., Baudi, F., and Perrotti, N. (2013). The role of microRNAs in cancer susceptibility. *Biomed Res. Int.* 2013:591931. doi: 10.1155/2013/591931
- Ke, S., Shang, S., Kalachikov, S. M., Morozova, I., Yu, L., Russo, J. J., et al. (2011). Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.* 21, 1360–1374. doi: 10.1101/gr.119628.110
- Kobori, M., and Nojima, H. (1993). A simple treatment of DNA in a ligation mixture prior to electroporation improves transformation frequency. *Nucleic Acids Res.* 21:2782. doi: 10.1093/nar/21.11.2782
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Kwasniewski, J. C., Mogno, I., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2012). Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19498–19503. doi: 10.1073/pnas.1210678109
- Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., et al. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129, 1401–1414. doi: 10.1016/j.cell.2007.04.040
- Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20. doi: 10.1016/j.cell.2004.12.035
- Lim, L. P., Lau, N. C., Garrett-Engle, P., Grimson, A., Schelter, J. M., Castle, J., et al. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773. doi: 10.1038/nature03315
- Liu, C., Zhang, F., Li, T., Lu, M., Wang, L., Yue, W., et al. (2012). MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics* 13:661. doi: 10.1186/1471-2164-13-661
- Liu, C. X., Tan, L., Sun, F. R., Zhang, W., Miao, D., Tan, M. S., et al. (2017). The association of MME microRNA Binding Site Polymorphism with risk of late onset Alzheimer's Disease in Northern Han Chinese. *Curr. Neurovasc. Res.* 14, 90–95. doi: 10.2174/156720261466617031310301
- Loya, C. M., Lu, C. S., Van Vactor, D., and Fulga, T. A. (2009). Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. *Nat. Methods* 6, 897–903. doi: 10.1038/nmeth.1402
- Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., et al. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* 44, 3865–3877. doi: 10.1093/nar/gkw116
- Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30, 271–277. doi: 10.1038/nbt.2137
- Micallef, L., and Rodgers, P. (2014). eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* 9:e01717. doi: 10.1371/journal.pone.0101717
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., Mcveigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- Pelletier, C., and Weidhaas, J. B. (2010). MicroRNA binding site polymorphisms as biomarkers of cancer risk. *Expert Rev. Mol. Diagn.* 10, 817–829. doi: 10.1586/erm.10.59
- Ramamoorthy, A., Li, L., Gaedigk, A., Bradford, L. D., Benson, E. A., Flockhart, D. A., et al. (2012). In silico and in vitro identification of microRNAs that regulate hepatic nuclear factor 4 α expression. *Drug Metab. Dispos.* 40, 726–733. doi: 10.1124/dmd.111.040329
- Reid, D. C., Chang, B. L., Gunderson, S. I., Alpert, L., Thompson, W. A., and Fairbrother, W. G. (2009). Next-generation SELEX identifies sequence and structural determinants of splicing factor binding in human pre-mRNA sequence. *RNA* 15, 2385–2397. doi: 10.1261/rna.1821809
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Soemedi, R., Vega, H., Belmont, J. M., Ramachandran, S., and Fairbrother, W. G. (2014). Genetic variation and RNA binding proteins: tools and techniques to detect functional polymorphisms. *Adv. Exp. Med. Biol.* 825, 227–266. doi: 10.1007/978-1-4939-1221-6_7

- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006. doi: 10.1093/nar/gkt1229
- Xie, T. D., Sun, L., Zhao, H. G., Fuchs, J. A., and Tsong, T. Y. (1992). Study of mechanisms of electric field-induced DNA transfection. IV. Effects of DNA topology on cell uptake and transfection efficiency. *Biophys. J.* 63, 1026–1031. doi: 10.1016/S0006-3495(92)81675-2
- Yekta, S., Shih, I. H., and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304, 594–596. doi: 10.1126/science.1097434

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ipe, Collins, Hao, Gao, Bhatia, Gaedigk, Liu and Skaar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Development of an AmpliSeq™ Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain

Dario Kringel¹, Mari A. Kaunisto², Catharina Lippmann³, Eija Kalso⁴ and Jörn Lötsch^{1,3*}

¹ Institute of Clinical Pharmacology, Goethe-University, Frankfurt, Germany, ² Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland, ³ Fraunhofer Institute for Molecular Biology and Applied Ecology – Project Group Translational Medicine and Pharmacology, Frankfurt, Germany, ⁴ Division of Pain Medicine, Department of Anesthesiology, Intensive Care and Pain Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch-Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Theodora Katsila,
University of Patras, Greece
Cheryl D. Cropp,
Samford University, United States

*Correspondence:

Jörn Lötsch
j.loetsch@em.uni-frankfurt.de

Specialty section:

This article was submitted to
Pharmacogenetics
and Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 24 May 2018

Accepted: 17 August 2018

Published: 19 September 2018

Citation:

Kringel D, Kaunisto MA, Lippmann C,
Kalso E and Lötsch J (2018)
Development of an AmpliSeq™ Panel
for Next-Generation Sequencing of a
Set of Genetic Predictors of Persisting
Pain. *Front. Pharmacol.* 9:1008.
doi: 10.3389/fphar.2018.01008

Background: Many gene variants modulate the individual perception of pain and possibly also its persistence. The limited selection of single functional variants is increasingly being replaced by analyses of the full coding and regulatory sequences of pain-relevant genes accessible by means of next generation sequencing (NGS).

Methods: An NGS panel was created for a set of 77 human genes selected following different lines of evidence supporting their role in persisting pain. To address the role of these candidate genes, we established a sequencing assay based on a custom AmpliSeq™ panel to assess the exomic sequences in 72 subjects of Caucasian ethnicity. To identify the systems biology of the genes, the biological functions associated with these genes were assessed by means of a computational over-representation analysis.

Results: Sequencing generated a median of $2.85 \cdot 10^6$ reads per run with a mean depth close to 200 reads, mean read length of 205 called bases and an average chip loading of 71%. A total of 3,185 genetic variants were called. A computational functional genomics analysis indicated that the proposed NGS gene panel covers biological processes identified previously as characterizing the functional genomics of persisting pain.

Conclusion: Results of the NGS assay suggested that the produced nucleotide sequences are comparable to those earned with the classical Sanger sequencing technique. The assay is applicable for small to large-scale experimental setups to target the accessing of information about any nucleotide within the addressed genes in a study cohort.

Keywords: pain, data science, knowledge discovery, functional genomics, next generation sequencing (NGS)

INTRODUCTION

Persisting pain has been proposed to result from a gene environment interaction where nerve injuries or inflammatory processes act as triggers while the clinical symptoms develop only in a minority of subjects (Lee and Tracey, 2013). A role of the genetic background in pain is supported by evidence of many variants modulating the individual perception of pain and the development of its persistence (Diatchenko et al., 2005; Lötsch et al., 2009b; Mogil, 2012). Genetic variants have been reported to confer protection against pain such as the rs1799971 variant in the μ -opioid receptor gene (*OPRM1*) (Lötsch et al., 2006), or to increase the risk for persisting pain such as the rs12584920 variant of the 5-hydroxytryptamine receptor 2A gene (*HTR2A*) (Nicholl et al., 2011) or the rs734784 polymorphism in the voltage-gated potassium ion channel modifier, subfamily S member 1, gene (*KCNS1*) (Costigan et al., 2010). Nevertheless, the genetic background of persisting pain is still incompletely understood (Mogil, 2009; Lötsch and Geisslinger, 2010) and under intense discussion.

Until recently, research focused on the role of selected functional genetic variants as protective or risk factors of persisting pain. This has changed with the broader availability of next generation sequencing (NGS) (Metzker, 2010). To make use of these technical advancements, we developed a custom AmpliSeqTM library and sequencing assay for efficient detection of genetic variants possibly associated with persisting pain. We propose an assay of a set of 77 genes supported by evidence of an involvement in pain and its development toward persistence. The set size fully uses the technical specifications of the AmpliSeqTM gene sequencing library technique.

MATERIALS AND METHODS

Selection of Genes Relevant for Persisting Pain

A set of candidate genes with shown or biologically plausible relevance to persisting pain was created by applying a combination of criteria, which provided three different genetic subsets. **Subset 1** was chosen exclusively on the basis of computational functional genomics based on a recently published analysis of persisting pain regarded as displaying systemic features of learning and neuronal plasticity (Mansour et al., 2014). As discussed previously (Ultsch et al., 2016), the view of chronic pain as a dysregulation in biological processes of learning and neuronal plasticity (Alvarado et al., 2013) seems to be captured by the controlled vocabulary (Camon et al., 2004) of the Gene Ontology (GO) knowledge base by the GO terms “learning or memory” (GO:0007611)¹ and “nervous system development” (GO:0007399)². An intersection of the genes annotated to these GO terms with a set of 539 “pain genes” identified empirically as relevant to pain provided the first subset of 34 genes described in detail previously (Ultsch et al., 2016). Briefly, the intersecting

set of so-called “pain genes” consists of a combination of (i) genes listed in the PainGenes database (Lacroix-Fralish et al., 2007)³, (ii) genes causally involved in human hereditary diseases associated with extreme pain phenotypes, (iii) genes found to be associated with chronic pain in at least three human studies, and (iv) genes coding for targets of novel analgesics under clinical development (Lötsch et al., 2013).

Subset 2 consisted of genes that were reported to carry variants modulating the risk or the phenotypic symptoms in at least two different clinical settings of persisting pain. They were obtained using (i) a PubMed database search for the string “(chronic OR persisting OR neuropathic OR back OR inflammatory OR musculoskeletal OR visceral OR widespread OR idiopathic OR fibromyalgia) AND pain AND (polymorphism OR variant) NOT review,” to which genes highlighted in overviews on pain genetics (e.g., Edwards, 2006) were added. The intersection of the queried genes with the set of 539 “pain genes” (see above) provided a subset of 13 genes (**Table 1**).

Finally, **subset 3** comprised genes that have consistently been included in human pain research projects over the last several years. One of them is the *OPRM1* gene that codes for the human μ -opioid receptor and which has been shown to modulate the time course of persisting cancer pain by delaying the necessity of opioid treatment (Lötsch et al., 2010). However, further genes were added such as the *GDNF* gene coding for the glial cell derived neurotrophic factor, which has been shown to be involved in a glia-dependent mechanism of neuropathic pain (Wang et al., 2014) although no modulating human genetic variants have been reported so far. Following expert counseling within the EU-funded “glial-opioid interface in chronic pain, GLORIA” research consortium (Kringel and Lötsch, 2015)⁴, a subset of 30 genes (**Table 1**) was identified. Thus, the complete set as the union of the three subsets comprised $43 + 13 + 30 = 77$ genes that are proposed to be included in an NGS panel of human persisting pain.

DNA Sample Origin

Due to the costs of assay development (for details, see second paragraph of the Discussion), the AmpliSeqTM panel was established in a limited number of $n = 72$ DNA samples. This corresponds to the number of samples used in comparable recent studies for NGS assay establishment and validation (Bruera et al., 2018; De Luca et al., 2018; Mustafa et al., 2018; Shah et al., 2018). To further limit the project costs, the AmpliSeqTM panel was established in a subset of samples originating from a clinical cohort of 1,000 women who had undergone breast cancer surgery (Kaunisto et al., 2013; Lötsch et al., 2018). The study followed the Declaration of Helsinki and was approved by the Coordinating Ethics Committee of the Helsinki University Hospital. Each participating subject had provided a written informed consent including genetic studies.

Specifically, for the presently reported method establishment, a subsample of 72 women (age 58.4 ± 8 years, mean \pm standard deviation, weight 69.3 ± 11 kg), was drawn from the clinical

¹<http://amigo.geneontology.org/amigo/term/GO:0007611>

²<http://amigo.geneontology.org/amigo/term/GO:0007399>

³<http://www.jbldesign.com/jmogil/enter.html>

⁴<http://gloria.helsinki.fi>

TABLE 1 | Genes included in the proposed NGS panel of persisting pain, combined from three subsets included on different bases.

Gene symbol	NCBI	Gene description	Reference
Subset #1			
<i>ADCY1</i>	107	Adenylate cyclase 1	Vadakkan et al., 2006
<i>BDNF</i>	627	Brain-derived neurotrophic factor	Obata and Noguchi, 2006
<i>CDK5</i>	1020	Cyclin-dependent kinase 5	Yang et al., 2014
<i>CHRNA2</i>	1141	Cholinergic receptor, nicotinic, beta 2	Dineley et al., 2015
<i>CNR1</i>	1268	Cannabinoid receptor 1 (brain)	Smith et al., 1998
<i>DLG4</i>	1742	Disks, large homolog 4 (<i>Drosophila</i>)	Florio et al., 2009
<i>DRD1</i>	1812	Dopamine receptor D1	Onojigihofia et al., 2014
<i>DRD2</i>	1813	Dopamine receptor D2	Onojigihofia et al., 2014
<i>DRD3</i>	1814	Dopamine receptor D3	Potvin et al., 2009
<i>EGR1</i>	1958	Early growth response 1	Ko et al., 2005
<i>FOS</i>	2353	Cellular oncogene FOS	Abbadie et al., 1994
<i>FYN</i>	2534	Src family tyrosine kinase	Liu et al., 2014
<i>GABRA5</i>	2558	GABA A receptor, alpha 5	Bravo-Hernández et al., 2016
<i>GALR2</i>	8811	Galanin receptor 2	Hulse et al., 2012
<i>GRIN1</i>	2902	Glutamate receptor, NMDA 1	Petrenko et al., 2003
<i>GRIN2A</i>	2903	Glutamate receptor, NMDA 2A	Petrenko et al., 2003
<i>GRIN2B</i>	2904	Glutamate receptor, NMDA 2B	Petrenko et al., 2003
<i>GRM5</i>	2915	Glutamate receptor, metabotropic 5	Walker et al., 2001
<i>HRH3</i>	11255	Histamine receptor H3	Huang et al., 2007
<i>KIT</i>	3815	Tyrosine kinase KIT	Sun et al., 2009
<i>NF1</i>	4763	Neurofibromin 1	Wolters et al., 2015
<i>NGF</i>	4803	Nerve growth factor	Kumar and Mahal, 2012
<i>NTF4</i>	4909	Neurotrophin 4	Kumar and Mahal, 2012
<i>NTRK1</i>	4914	Neurotrophic tyrosine kinase 1	Kumar and Mahal, 2012
<i>OXT</i>	5020	Oxytocin prepropeptide	Goodin et al., 2015
<i>PLCB1</i>	23236	Phospholipase C, beta 1	Shi T.-J.S. et al., 2008
<i>PRKCG</i>	5582	Protein kinase C, gamma	Sluka and Audette, 2006
<i>PRNP</i>	5621	Prion protein	Gadotti and Zamponi, 2011
<i>PTN</i>	5764	Pleiotrophin	Gramage and Herradon, 2010
<i>PTPRZ1</i>	5803	Protein tyrosine phosphatase Z 1	Ultsch et al., 2016
<i>RELN</i>	5649	Reelin	Buchheit et al., 2012
<i>S100B</i>	6285	S100 calcium binding protein B	Zanette et al., 2014
<i>SLC6A4</i>	6532	Serotonin transporter	Offenbaecher et al., 1999
<i>TH</i>	7054	Tyrosine hydroxylase	Bravo et al., 2014
Subset #2			
<i>ADRB2</i>	154	Adrenoceptor beta 2	Hocking et al., 2010
<i>COMT</i>	1312	Catechol-O-methyltransferase	Feng et al., 2013
<i>ESR1</i>	2099	Estrogen Receptor 1	Ribeiro-Dasilva et al., 2009
<i>GCH1</i>	2643	GTP cyclohydrolase 1	Tegeder et al., 2006
<i>IL1B</i>	3553	Interleukin 1B	Loncar et al., 2013
<i>IL4</i>	3565	Interleukin 4	Sugaya et al., 2002
<i>IL6</i>	3569	Interleukin 6	Shoskes et al., 2002
<i>IL10</i>	3586	Interleukin 10	Stephens et al., 2014
<i>P2RX7</i>	5027	Purinergic Receptor P2X7	Sorge et al., 2012
<i>SCN9A</i>	6335	Sodium voltage-gated alpha subunit 9	Reimann et al., 2010
<i>SOD2</i>	6648	Superoxide dismutase 2	Schwartz et al., 2009
<i>TNF</i>	7124	Tumor necrosis factor	Leung and Cahill, 2010
<i>TRPV1</i>	7442	Transient receptor potential cation channel, subfamily V, member 1	Bourinet et al., 2014
Subset #3			
<i>ABHD12</i>	26090	Abhydrolase domain containing 12	Kim, 2015
<i>ABHD16A</i>	7920	Abhydrolase domain containing 16A	Kim, 2015
<i>ABHD6</i>	57406	Abhydrolase domain containing 6	Kim, 2015

(Continued)

TABLE 1 | Continued

Gene symbol	NCBI	Gene description	Reference
CACNG2	10369	Calcium voltage-gated channel auxiliary subunit gamma 2	Nissenbaum et al., 2010
CSF1	1435	Colony stimulating factor 1	Thuaud, 2016
DRD4	1815	Dopamine receptor D4	Buskila et al., 2004
FAAH	2166	Fatty acid amide hydrolase	Jayamanne et al., 2006
FKBP5	2289	Fk506 binding protein 5	Fujii et al., 2014
GDNF	2668	Glial cell derived neurotrophic factor	Sah et al., 2005
GFRA1	2674	GDNF family receptor alpha 1	Yamamoto et al., 2003
GPR132	29933	G protein-coupled receptor 132	Hohmann et al., 2017
HCN2	610	Hyperpolarization-activated cyclic nucleotide-gated	Tsantoulas et al., 2016
HLA-DQB1	3119	Major histocompatibility complex, class II, DQ beta 1	Dominguez et al., 2013
HLA-DRB1	3123	Major histocompatibility complex, class II, DR beta 1	Dominguez et al., 2013
HTR1A	3350	5-hydroxytryptamine (serotonin) receptor 1A	Lindstedt et al., 2012
HTR2A	3356	5-hydroxytryptamine (serotonin) receptor 2A	Nicholl et al., 2011
IL1R2	7850	Interleukin 1 receptor type 2	Stephens et al., 2014
KCNS1	3787	Potassium voltage-gated channel, modifier subfamily S, member 1	Costigan et al., 2010
LTB4R	1241	Leukotriene b4 receptor	Zinn et al., 2017
LTB4R2	56413	Leukotriene b4 receptor 2	Zinn et al., 2017
OPRD1	4985	Opioid receptor delta 1	Law et al., 2013
OPRK1	4986	Opioid receptor kappa 1	Guerrero et al., 2010
OPRM1	4988	Opioid receptor mu 1	Lötsch and Geisslinger, 2005
RET	5979	RET receptor tyrosine kinase	Snider and McMahon, 1998
RUNX1	861	Runt related transcription factor 1	Chen et al., 2006
TLR4	7099	Toll like Receptor 4	Hutchinson et al., 2010
TRPA1	8989	Transient receptor potential cation channel, subfamily A, member 1	Bourinet et al., 2014
TRPM8	79054	Transient receptor potential cation channel, subfamily M, member 8	Bourinet et al., 2014
TRPV4	59341	Transient receptor potential cation channel, subfamily V, member 4	Bourinet et al., 2014
TSPO	706	Translocator protein	Loggia et al., 2015

Subset #1 comprises $d = 34$ genes that had resulted from a computational functional genomics analysis (Ultsch et al., 2016) pursuing the hypothesis that persisting pain displays systemic features of learning and neuronal plasticity (Mansour et al., 2014). Hence, from a set of genes identified empirically as relevant to pain and listed in the PainGenes database (<http://www.jbldesign.com/jmogil/enter.html>, Lacroix-Fralish et al., 2007), those were selected that are annotated to the Gene Ontology (Ashburner et al., 2000) terms “learning or memory” and “nervous system development.” The references are those found to provide evidence for an association with pain, except for PTPRZ1 that was a novel finding in (Ultsch et al., 2016). Subset #2 comprises $d = 13$ genes identified empirically as relevant to pain and listed in the PainGenes database (<http://www.jbldesign.com/jmogil/enter.html>, Lacroix-Fralish et al., 2007) and reported to carry variants that modulated the risk or the symptomatology in at least two different clinical settings of persisting pain. Subset #3 comprises $d = 30$ genes repeatedly shown during the last several years to play a role in the human genetics of persisting pain or recently reported as novel players.

subgroup not having developed persisting pain during the observation period. This was believed to come closer to a random sample than a mixture of patients with persisting and without persisting pain. This limitation of the sample selection has probably affected which and how many variants were identified. However, it is unlikely to have jeopardized the general applicability of the gene selection heuristics, assay establishment and validation, and of the functional analysis of the selected subset of genes.

DNA Template Preparation and Amplification

A multiplex PCR amplification strategy for the coding gene sequences was accomplished online (Ion AmpliseqTM Designer)⁵ to amplify the target region specified above (for primer sequences, see **Supplementary Table 1**) with 25 base pair exon padding. After a comparison of several primer design options,

⁵<http://www.ampliseq.com>

the design providing the maximum target sequence coverage was chosen. The ordered 1,953 amplicons covered approximately 97.5% of the target sequence (**Supplementary Table 2**). A total of 10 ng DNA per sample was used for the target enrichment by a multiplex PCR and each DNA pool was amplified with the Ion AmpliseqTM Library Kit in conjunction with the Ion AmpliseqTM “custom Primer Pool”-protocols according to the manufacturer’s procedures (Life Technologies, Darmstadt, Germany).

After each pool had undergone 18 PCR cycles, the PCR primers were removed with FuPa Reagent and the amplicons were ligated to the sequencing adaptors with short stretches of index sequences (barcodes) that enabled sample multiplexing for subsequent steps (Ion XpressTM Barcode Adapters Kit; Life Technologies). After purification with AMPure XP beads (Beckman Coulter, Krefeld, Germany), the barcoded libraries were quantified with a Qubit[®] 2.0 Fluorimeter (Life Technologies, Darmstadt, Germany) and normalized for DNA concentration to a final concentration of 20 pmol/l using the Ion Library EqualizerTM Kit (Life Technologies, Darmstadt, Germany).

Equalized barcoded libraries from seven to eight samples at a time were pooled. To clonally amplify the library DNA onto the Ion Sphere Particles (ISPs; Life Technologies, Darmstadt, Germany), the library pool was subjected to emulsion PCR by using an Ion PGM HI-Q View Template Kit on an PGM OneTouch system (Life Technologies, Darmstadt, Germany) following the manufacturer's protocol.

Sequencing

Enriched ISPs which carried many copies of the same DNA fragment were subjected to sequencing on an Ion 318 Chip to sequence pooled libraries with seven to eight samples. During this process, bases are inferred from light intensity signals, a process commonly referred to as base-calling (Ledergerber and Dessimoz, 2011). The number of combined libraries that can be accommodated in a single sequencing run depends on the size of the chip, the balance of barcoded library concentration, and the coverage required. The high-capacity 318 chip was chosen (instead of the low-capacity 314 or the medium-capacity 316 chip) to obtain a high sequencing depth of coverage for a genomic DNA library with >95% of bases at 30x. Sequencing was performed using the sequencing kit (Ion PGM Hi-Q Sequencing Kit; Life Technologies, Darmstadt, Germany) as per the manufacturer's instructions with the 200 bp single-end run configuration. This kit contained the most advanced sequencing chemistry available to users of the Ion PGM System (Life Technologies, Darmstadt, Germany).

Data Analysis

Bioinformatics Generation of Sequence Information

The raw data (unmapped BAM-files) from the sequencing runs were processed using Torrent Suite Software (Version 5.2.2, Life Technologies, Darmstadt, Germany) to generate read alignments which were filtered by the software into mapped BAM-files using the reference genomic sequence (hg19) of target genes. Variant calling was performed with the Torrent Variant Caller Plugin using as key parameters: minimum allele frequency = 0.15, minimum quality = 10, minimum coverage = 20 and minimum coverage on either strand = 3.

The annotation of called variants was done using the Ion Reporter Software (Version 4.4; Life Technologies, Darmstadt, Germany) for the VCF files that contained the nucleotide reads and the GenomeBrowse® software (Version 2.0.4, Golden Helix, Bozeman, MT, United States) to map the sequences to the reference sequences GRCh37 hg19 (dated February 2009). The SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, United States) was used for the analysis of sequence quality, coverage and for variant identification.

Based on the observed allelic frequency, the expected number of homozygous and heterozygous carriers of the respective SNP (single nucleotide polymorphism) was calculated using the Hardy-Weinberg equation. Only variants within the Hardy-Weinberg equilibrium as assessed using Fisher's exact test (Emigh, 1980) were retained. The SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, United States) was used for the analysis of sequence quality, coverage and for variant identification.

Assay Validation

Method validation was accomplished by means of Sanger sequencing (Sanger and Coulson, 1975; Sanger et al., 1977) in an independent external laboratory (Eurofins Genomics, Ebersberg, Germany). As performed previously with different AmpliSeq™ panels (Kringel et al., 2017) and other genotyping assays (Skarke et al., 2004, 2005), four DNA samples have been chosen randomly from an independent cohort of healthy subjects and sequenced with the current NGS panel. For the detected variant type, single nucleotide polymorphisms from five different genomic regions for which clinical associations have been reported (Table 2), i.e., rs324420 (FAAH), rs333970 (CSF1), rs4986790 (TLR4), rs4633 (COMT), and rs17151558 (RELN) were chosen for external sequencing. Amplification of the respective DNA segments was done using PCR primer pairs (forward, reverse) of (i) 5'-TTTCTTAAAAAGGCCAGCCTCCT-3' and 5'-AATGACCCAAGATGCAGAGCA-3' (ii) 5'-GCCTTCAACCCCGGGATGG-3' and 5'-CTCCGATCCCTGGTGC TCCTC-3' (iii) 5'-TTTATTGCACAGACTTGCGGGTTC-3' and 5'-AGCCTTTTGAGAGATTTGAGTTTCA-3' (iv) 5'-CC TTATCGGCTGGAACGAGTT-3' and 5'-GTAAGGGCTTT GATGCCTGGT-3' (v) 5'-GTTATTCCTCTGTAAGCAGCTGCC T-3' and 5'-TGTTTGTTTTAGATTGTGGTGGGTT-3'. Results of Sanger sequencing were aligned with the genomic sequence and analyzed using Chromas Lite® (Version 2.1.1, Technelysium Pty Ltd, South Brisbane, QLD, Australia) and the GenomeBrowse® (Version 2.0.4, Golden Helix, Bozeman, MT, United States) was used to compare the sequences obtained with NGS or Sanger techniques.

RESULTS

The NGS assay of the proposed set of 77 human genes relevant to persisting pain was established in 72 genomic DNA samples. As applied previously (Kringel et al., 2017), only exons including 25 bases of padding around all targeted coding regions for which the realized read-depths for each nucleotide was higher than 20 were contemplated as successfully analyzed. With this acceptance criterion the whole or almost whole coverage of the relevant sequences was obtained (Table 1; for details on missing variants, see Supplementary Table 3). The NGS sequencing process of the whole patient cohort required ten separate runs, each with samples of $n = 7$ or $n = 8$ patients. Coverage statistics were analogous between all runs and matched the scope of accepted quality levels [20–22]. A median of $2.85 \cdot 10^6$ reads per run was produced. The mean depth was close to 200 reads, the mean read length of called bases resulted in 205 bases and average chip loading was 71% (Figure 1A). To establish a sequencing output with a high density of ISPs on a sequencing chip, the chip loading value should exceed 60% (Life Technologies, Carlsbad, United States). The generated results of all NGS runs matched with the results obtained with Sanger sequencing of random samples (Figure 1B), meaning the accordance of nucleotide sequences between NGS and Sanger sequencing was 100% in all validated samples.

TABLE 2 | A list of coding human variants in the 77 putative chronic pain genes, found in the present random sample of 72 subjects of Caucasian ethnicity, for which clinical associations have been reported.

Gene	Variant	dbSNP# accession number	Known clinical association	Reference
Pain context				
<i>FAAH</i>	1:46870761-SNV	rs324420	Effect of endocannabinoid degradation on pain	Cajanus et al., 2016
<i>FAAH</i>	1:46870761-SNV	rs324420	Cold and heat pain sensitivity	Kim et al., 2006b
<i>CSF1</i>	1:110466338-SNV	rs333970	Rheumatoid arthritis	Solus et al., 2015
<i>NGF</i>	1:115829313-SNV	rs6330	Procedural pain	Ersig et al., 2017
<i>NGF</i>	1:115829313-SNV	rs6330	Susceptibility to migraine	Coskun et al., 2016
<i>IL1B</i>	2:113590966-SNV	rs1143634	Adverse effects in postoperative pain	Somogyi et al., 2016
<i>IL1B</i>	2:113590966-SNV	rs1143634	Low back pain	Feng et al., 2016
<i>SCN9A</i>	2:167099158-SNV	rs6746030	Pain susceptibility in Parkinson disease	Greenbaum et al., 2012
<i>SCN9A</i>	2:167099158-SNV	rs6746030	Congenital insensitivity to pain	Klein et al., 2013
<i>SCN9A</i>	2:167099158-SNV	rs6746030	Basal Pain Sensitivity	Duan et al., 2015
<i>SCN9A</i>	2:167145122-SNV	rs188798505	Altered pain perception	Reimann et al., 2010
<i>DRD3</i>	3:113890815-SNV	rs6280	Acute pain in sickle cell disease	Jhun et al., 2014
<i>DRD3</i>	3:113890815-SNV	rs6280	Higher prevalence of migraine	Hu et al., 2014
<i>ADRB2</i>	5:148206646-SNV	rs1042717	Musculoskeletal pain	Diatchenko et al., 2006
<i>ADRB2</i>	5:148206885-SNV	rs1800888	Migraine	Schurks et al., 2009
<i>ESR1</i>	6:152129077-SNV	rs2077647	Migraine	Schürks et al., 2010
<i>ESR1</i>	6:152129077-SNV	rs2077647	Musculoskeletal pain	Wise et al., 2009
<i>OPRM1</i>	6:154360797-SNV	rs1799971	Pain of various origins	Lötsch et al., 2009c
<i>SOD2</i>	6:160113872-SNV	rs4880	Migraine	Palmirotta et al., 2015
<i>IL6</i>	7:22771039-SNV	rs13306435	Low back pain	Eskola et al., 2010
<i>OPRK1</i>	8:54142157-SNV	rs702764	Neuropathic pain	Garassino et al., 2013
<i>TLR4</i>	9:120475302-SNV	rs4986790	Musculoskeletal pain	Gębura et al., 2017
<i>TH</i>	11:2188238-SNV	rs6357	Widespread Pain	Jhun et al., 2015
<i>TH</i>	11:2190951-SNV	rs6356	Migraine	Corominas et al., 2009
<i>BDNF</i>	11:27679916-SNV	rs6265	Widespread Pain	Ersig et al., 2017
<i>DRD2</i>	11:113283459-SNV	rs6277	Post-surgical pain	Kim et al., 2006a
<i>DRD2</i>	11:113283477-SNV	rs6275	Migraine	Onaya et al., 2013
<i>P2RX7</i>	12:121600253-SNV	rs208294	Cold pain sensitivity	Idé et al., 2014
<i>P2RX7</i>	12:121605355-SNV	rs7958311	Neuropathic pain	Ursu et al., 2014
<i>HTR2A</i>	13:47409034-SNV	rs6314	Migraine susceptibility	Yücel et al., 2016
<i>TRPV1</i>	17:3480447-SNV	rs8065080	Neuropathic pain	Doehring et al., 2011
<i>KCNS1</i>	20:43723627-SNV	rs734784	Neuropathic pain	Doehring et al., 2011
<i>COMT</i>	22:19950235-SNV	rs4633	Postoperative pain	Khalil et al., 2017
<i>COMT</i>	22:19950263-SNV	rs6267	Widespread Pain	Lin et al., 2017
<i>COMT</i>	22:19951271-SNV	rs4680	Altered pain perception	Wang et al., 2015
Other context				
<i>CSF1</i>	1:110466466-SNV	rs1058885	Periodontitis	Chen et al., 2014
<i>CSF1</i>	1:110466555-SNV	rs2229165	Carcinogenesis/breast cancer	Savas et al., 2006
<i>NTRK1</i>	1:156846233-SNV	rs6334	Nephropathy	Hahn et al., 2011
<i>NTRK1</i>	1:156848946-SNV	rs6339	Acute myeloid leukemia	Schweinhardt et al., 2008
<i>SCN9A</i>	2:167143050-SNV	rs41268673	Erythromelalgia	Klein et al., 2013
<i>TRPM8</i>	2:234854550-SNV	rs11562975	Hyperresponsiveness in bronchial asthma	Naumov et al., 2015
<i>TRPM8</i>	2:234905078-SNV	rs11563208	Anthropometric parameters	Potapova et al., 2014
<i>DRD3</i>	3:113890789-SNV	rs3732783	Phenotypic traits relevant to anorexia nervosa	Root et al., 2011
<i>KIT</i>	4:55593464-SNV	rs3822214	Cancer risk	Pelletier and Weidhaas, 2010
<i>KIT</i>	4:55602765-SNV	rs3733542	Glandular odontogenic cyst	Siqueira et al., 2017
<i>HTR1A</i>	5:63257483-SNV	rs1799921	Bipolar disorders	Goodyer et al., 2010
<i>ADRB2</i>	5:148206646-SNV	rs1042717	Cognitive dysfunction in opioid-treated patients with cancer	Kurita et al., 2016
<i>DRD1</i>	5:174868840-SNV	rs155417	Alcohol dependence	Hack et al., 2011
<i>HLA-DQB1</i>	6:32629920-SNV	rs41544112	Ulcerative colitis	Achkar et al., 2012
<i>FKBP5</i>	6:35544942-SNV	rs34866878	Clinical response in pediatric acute myeloid leukemia	Mitra et al., 2011

(Continued)

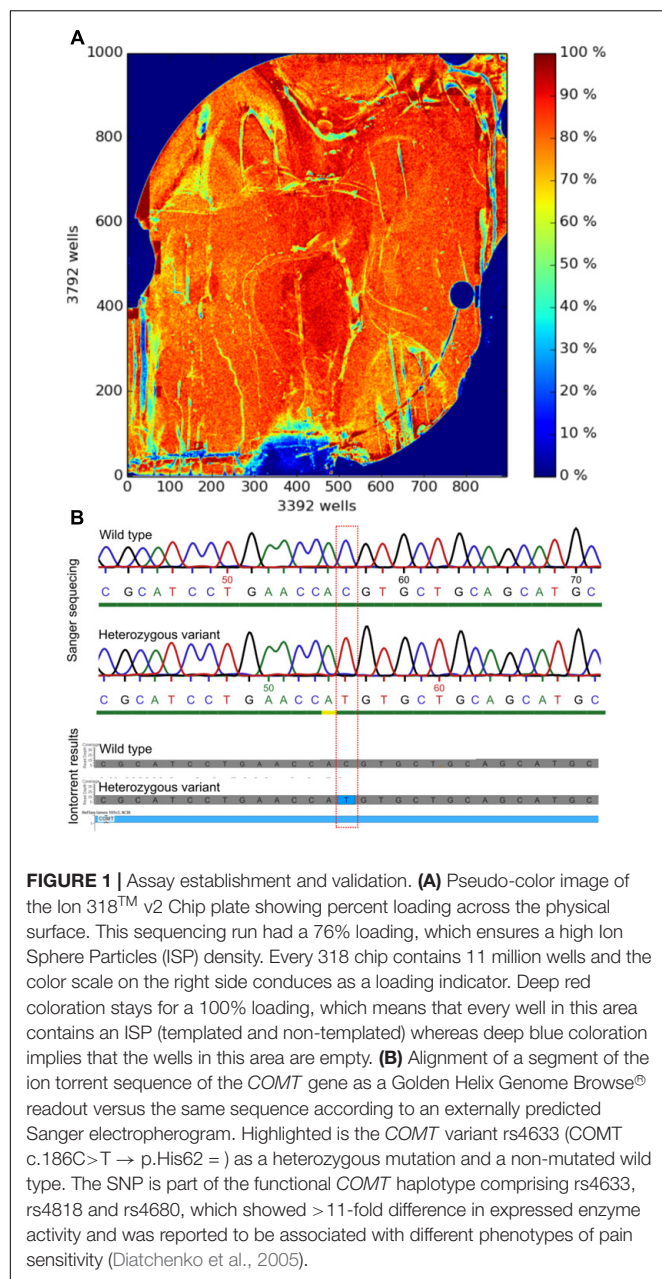
TABLE 2 | Continued

Gene	Variant	dbSNP# accession number	Known clinical association	Reference
CNR1	6:88853635-SNV	rs1049353	Bone mineral density	Woo et al., 2015
CNR1	6:88853635-SNV	rs1049353	Alcohol dependence	Marcos et al., 2012
CNR1	6:88853635-SNV	rs1049353	Nicotine dependence	Chen et al., 2008
CNR1	6:88853635-SNV	rs1049353	Obesity	Schleinitz et al., 2010
CNR1	6:88853635-SNV	rs1049353	Psychiatric disorders	Hillard et al., 2012
ESR1	6:152129077-SNV	rs2077647	Breast cancer susceptibility	Li et al., 2016
ESR1	6:152129077-SNV	rs2077647	Prostate cancer development	Jurečková et al., 2015
ESR1	6:152129077-SNV	rs2077647	Osteoporosis	Sonoda et al., 2012
ESR1	6:152129308-SNV	rs746432	Mood disorders	Mill et al., 2008
ESR1	6:152201875-SNV	rs4986934	Endometrial cancer risk	Wedrén et al., 2008
OPRM1	6:154360508-SNV	rs6912029	Irritable bowel syndrome	Camilleri et al., 2014
OPRM1	6:154360797-SNV	rs1799971	Schizophrenia	Serý et al., 2010
OPRM1	6:154414573-SNV	rs562859	Depressive disorder	Garriock et al., 2010
OPRM1	6:154414563-SNV	rs675026	Treatment response for opiate dependence	Al-Eitan et al., 2012
SOD2	6:160113872-SNV	rs4880	Development of type 2 diabetes mellitus	Li et al., 2015
SOD2	6:160113872-SNV	rs4880	Breast cancer susceptibility	Rodrigues et al., 2014
SOD2	6:160113872-SNV	rs4880	Asthma	Yucesoy et al., 2012
ADCY1	7:45703971-SNV	rs1042009	Bipolar disorder	Shi J. et al., 2008
RELN	7:103124207-SNV	rs1062831	Attention deficit hyperactivity disorder	Kwon et al., 2016
RELN	7:103251161-SNV	rs362691	Childhood epilepsy	Dutta et al., 2011
OPRK1	8:54142154-SNV	rs16918875	Susceptibility to addiction	Kumar et al., 2012
TRPV1	8:72948588-SNV	rs13280644	Perception olfactory stimuli	Schütz et al., 2014
TLR4	9:120475602-SNV	rs4986791	Breast cancer susceptibility	Milne et al., 2014
GRIN1	9:140051238-SNV	rs6293	Schizophrenia	Georgi et al., 2007
RET	10:43610119-SNV	rs1799939	Hirschsprung's disease	Vaclavikova et al., 2014
RET	10:43615094-SNV	rs1800862	Medullary thyroid carcinoma	Ceolin et al., 2012
GFRA1	10:117884950-SNV	rs2245020	Age-related macular degeneration	Schmidt et al., 2006
DRD4	11:637537-Del	rs587776842	Acousticous neurinoma	Nöthen et al., 1994
BDNF	11:27720937-SNV	rs66866077	Irritable bowel syndrome-diarrhea	Camilleri et al., 2014
DRD2	11:113283484-SNV	rs1801028	Neurologic disorders	Doehring et al., 2009
GRIN2B	12:13717508-SNV	rs1806201	Alzheimer's disease	Andreoli et al., 2014
TRPV4	12:110252547-SNV	rs3742030	Hyponatremia	Tian et al., 2009
P2RX7	12:121592689-SNV	rs17525809	Multiple sclerosis	Oyanguren-Desez et al., 2011
HTR2A	13:47466622-SNV	rs6305	Susceptibility to substance abuse	Herman and Balogh, 2012
LTB4R	14:24785092-SNV	rs34645221	Asthma susceptibility	Tulah et al., 2012
GABRA5	15:27182357-SNV	rs140682	Autism-spectrum disorders	Hogart et al., 2007
GRIN2A	16:9943666-SNV	rs2229193	Hyperactivity disorder	Kim et al., 2017
DLG4	17:7099811-SNV	rs17203281	Schizophrenia	Tsai et al., 2007
SLC6A4	17:28530193-SNV	rs6352	Autism-spectrum disorders	Prasad et al., 2009
NF1	17:29553485-SNV	rs2285892	Neurofibromatosis	Maertens et al., 2007
HCN2	19:607984-SNV	rs3752158	Risk of depression	McIntosh et al., 2012
PRKCG	19:54394965-SNV	rs3745396	Osteosarcoma susceptibility	Lu et al., 2015
PRNP	20:4680251-SNV	rs1799990	Creutzfeldt-Jakob disease	Mead et al., 2009
HRH3	20:60791422-SNV	rs3787430	Risk of chronic heart failure	He et al., 2016
S100B	21:48022230-SNV	rs1051169	Schizophrenia	Liu et al., 2005

The selection is restricted to one or two publications per variant, and it focuses on a pain context corresponding to the main aim of the present NGS gene panel; however, functional variants highlighted in another clinical context are additionally provided in the lower part of the table. #Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD, United States): National Center for Biotechnology Information, National Library of Medicine. Available from: <http://www.ncbi.nlm.nih.gov/SNP/> (Sherry et al., 2001).

Following elimination of nucleotides agreeing with the standard human genome sequence GRCh37 g1k (dated February 2009), the result of the NGS consisted of a vector of nucleotide information about the $d = 77$ genes for each individual DNA

sample (Figure 2). This vector had a length equaling the set union of the number of chromosomal positions in which a non-reference nucleotide had been found in any probe of the actual cohort. Specifically, a total of 3,185 genetic variants was found, of



which 659 were located in coding parts of the genes, 1,241 were located in introns and 1,285 in the 3'-UTR, 5'-UTR, upstream or downstream regions. The coding variants for which a clinical or phenotypic association have been reported are listed in **Table 2** together with an example of each variant. Most of the observed variants were single nucleotide polymorphisms ($d = 571$) whereas mixed polymorphisms ($d = 26$), nucleotide insertions ($d = 18$) or nucleotide deletions ($d = 44$) were more rarely found.

DISCUSSION

In this report, development and validation of a novel Ampliseq™ NGS assay for the coding regions and boundary parts of $d = 77$

genes qualifying as candidate modulators of persisting pain is described. The NGS assay produced nucleotide sequences that corresponded, with respect to the selected validation probes, to the results of classical Sanger sequencing. However, the NGS assay substantially reduced the laboratory effort to obtain the genetic information and provides the prerequisites to be used in high throughput environments. In particular, the presented NGS assay is convenient for small up to large-scale setups. As mentioned in the methods section, a limitation of the present results applies to the identified genetic variants as only samples from Caucasian women were included. By contrast, the validity of gene selection and assay establishment is unlikely to be reduced by this selection chosen to remain within the financial limits of the present project.

Specifically, as observed previously (Kringel et al., 2017), the comprehensive genetic information and the high throughput are reflected in the assay costs. Specifically, sequencing of the 77 genes in 72 DNA samples required approximately € 18,000 for the AmpliSeq™ custom panel, € 5,500 for library preparation, € 700 for template preparation and € 700 for sequencing. Ten 318 sequencing chips cost around € 7,000 and in addition and basic consumables and laboratory supplies issued approximately € 800. With 7–8 barcoded samples loaded on ten chips, the expense to analyses the gene sequence for a single patient were around € 325. While NGS costs are likely to decrease in the near future (Lohmann and Klein, 2014), present assay establishment was therefore applied in DNA samples planned for future genotype versus phenotype association analysis, which required using DNA from patients of a pain-relevant cohort instead from a true random sample of healthy subjects.

As a result of the present assay development, a set of $d = 77$ genes was chosen as potentially relevant to persisting pain. The chosen set of genes differs from alternative proposals aiming at similar phenotypes (Mogil, 2012; Zorina-Lichtenwalter et al., 2016). However, when analyzing these alternatives for mutual agreement, only limited overlap could be observed (**Figure 3**). This emphasizes that the genetic architecture of persisting pain is incompletely understood, and several independent lines of research can be pursued. Of note, the present set showed the largest agreement with a set of $d = 539$ genes identified empirically as relevant to pain and listed in the PainGenes database (Lacroix-Fralish et al., 2007)⁶ or recognized as causing human hereditary diseases associated with extreme pain phenotypes (Lötsch et al., 2013; Ultsch et al., 2016). Combining all proposals into a large panel was not an option due to the technical limitations of the IonTorrent restricting the panel size to 500 kb (pipeline version 5.6.2); therefore, further genes would need to be addressed in separate panels.

In the present study sample, selected with a certain bias by using, as explained above for cost saving, clinical samples from only women and only Caucasians, a total of 659 genetic coding variants were found. Regardless of the sample preselection, 105 clinical associations (**Table 2**) could be queried for the observed variants from openly obtainable data sources comprising (i) the

⁶<http://www.jbldesign.com/jmogil/enter.html>

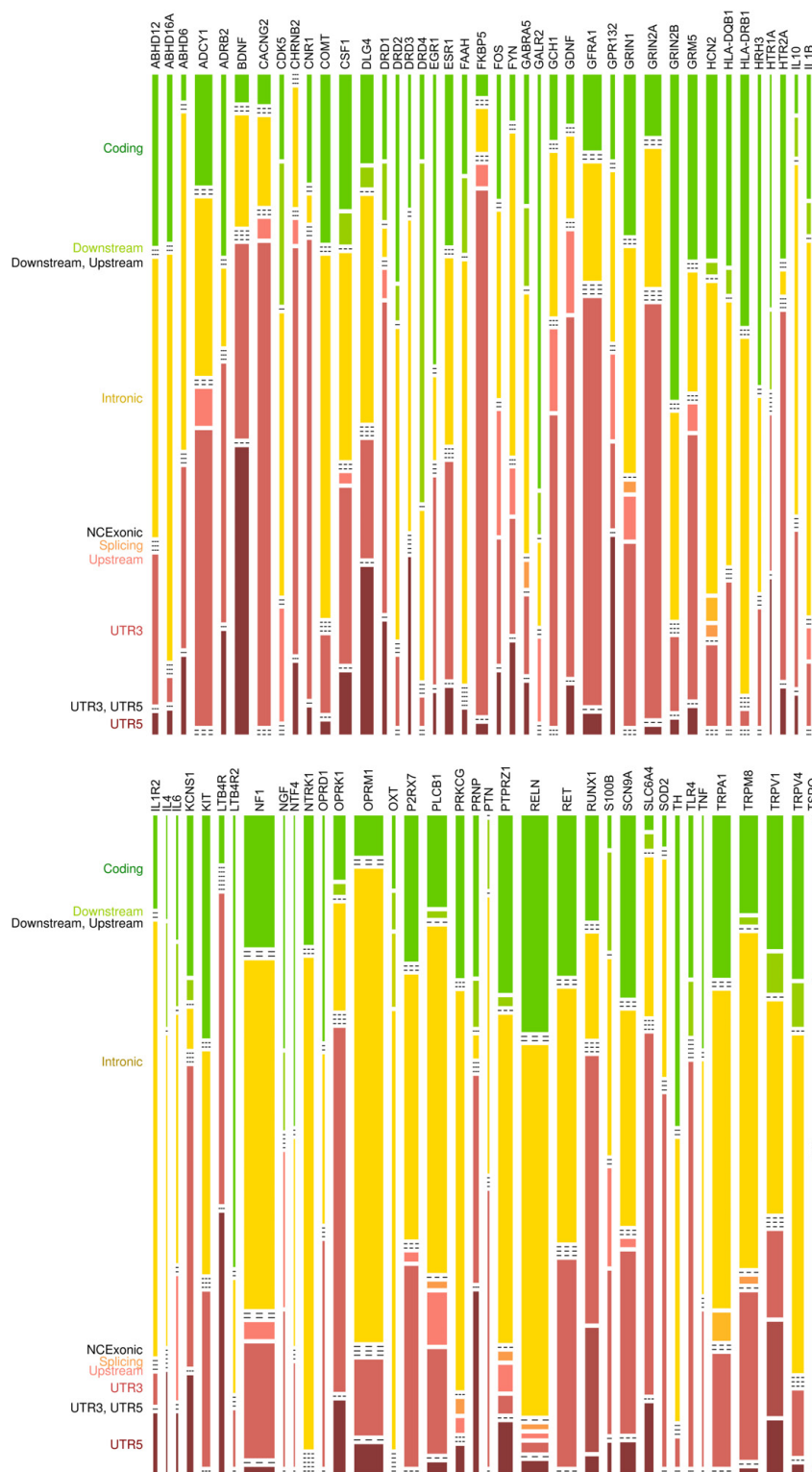
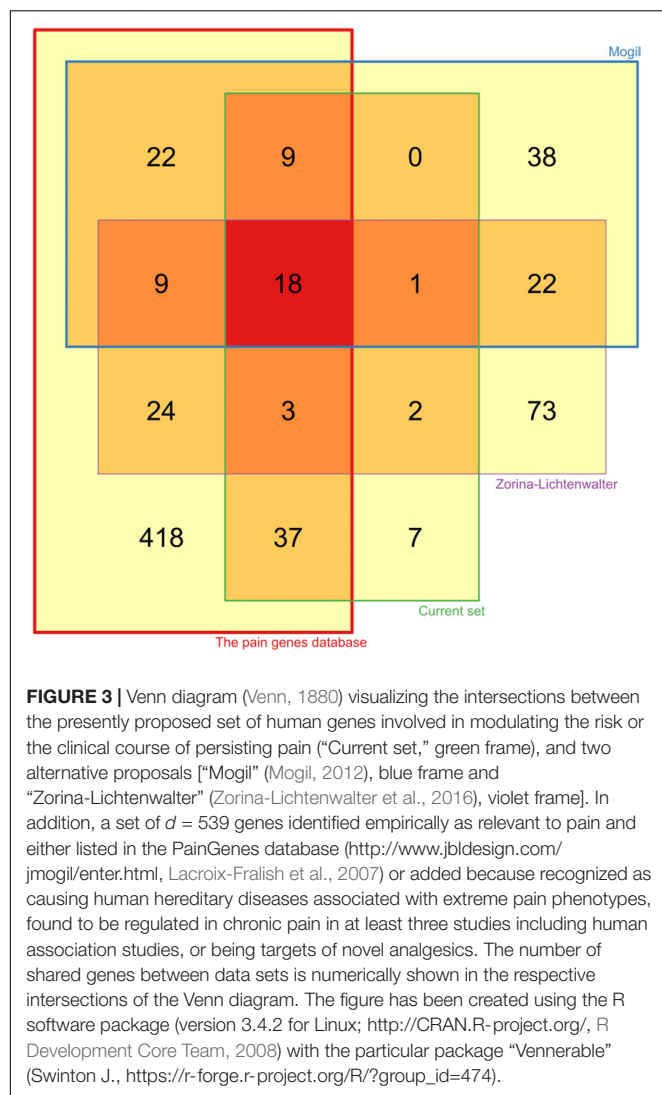


FIGURE 2 | Mosaic plot representing a contingency table of the types of genetic variants detected by means of the present AmpliSeq™ panel versus the genes included in the assay. The vertical size of the cells is proportional to the number of variants of a particular type; the horizontal size of the cells is proportional to the number of variants found in the respective gene. The location of the variants is indicated at the left of the mosaic plot in letters colored similarly to the respective bars in the mosaic plot. Variants were not found at all possible locations of each gene, which causes the reduction of several bars to dashed lines drawn as placeholders and indicating that at the particular location no variant has been found in the respective gene. The figure has been created using the R software package (version 3.4.2 for Linux; <http://CRAN.R-project.org/>, R Development Core Team, 2008). UTR: untranslated region. NCExonic: Non-coding exonic.



Online Mendelian Inheritance in Man (OMIM®) database⁷, (ii) the NCBI gene index database⁸, the GeneCards database⁹ [27] and the "1000 Genomes Browser"¹⁰ (all accessed in December 2017). The observation of functional variants in the present cohort preselected for the absence of pain persistence is plausible as (i) variants can exert protective effects against chronic pain and (ii) most genetic variants identified so far exert only small effects on pain and the individual result of their functional modulations depends on their combined effects or from the sum of positive and negative effects on pain perception (Lötsch et al., 2009a).

The selection of genes (Table 1) relied on empirical evidence of their involvement in pain. For subset #1 ($d = 34$), this had been shown for 33 genes in the original paper (Ultsch et al., 2016). As the hypothesis that persisting pain displays systemic features of

learning and of neuronal plasticity (Mansour et al., 2014) could be substantiated at a computational functional genomics level, the further gene (*PTPRZ1*, protein tyrosine phosphatase Z 1) can also be regarded as supported by prior knowledge to be included in the present set. The subset comprised, for example, genes associated with the mesolimbic dopaminergic system, i.e., *DRD1*, *DRD2*, *DRD3*, which code for dopamine receptors, and *TH*, which is the coding gene for the tyrosine hydroxylase, a metabolic restricting enzyme in dopaminergic pathways, which have been implicated in promoting chronic back pain (Hagelberg et al., 2003, 2004; Jaaskelainen et al., 2014; Martikainen et al., 2015). Further 14 genes were involved in the circadian rhythm recognized as a modulatory factor in various pain conditions such as arthritis (Haus et al., 2012; Gibbs and Ray, 2013) and neuropathic pain (Gilron and Ghasemlou, 2014). The subset further included three NMDA receptor genes (*GRIN1*, *GRIN2A*, and *GRIN2B*) known to be major players in a number of essential physiological functions including neuroplasticity (Coyle and Tsai, 2004). In addition, metabotropic glutamate receptors (mGluR) have been implemented in several chronic pain conditions. One subtype, mGluR5, coded by *GRM5*, is of particular interest in the context of pain conditions as recent studies showed a pro-nociceptive role of mGluR5 in models of chronic pain (Walker et al., 2001; Crock et al., 2012). Furthermore, genes associated with histaminergic signaling such as *HRH3* have been implicated in pain transmission (Hough and Rice, 2011) and analgesia (Huang et al., 2007).

The second subset of genes relied on a new PubMed search rather than on a previously published and hypothesis-based selection of candidate genes. A computational functional genomics analysis of this subset (details not shown) suggested its involvement in (i) immune processes and (ii) nitric oxide signaling. The genes annotated to the GO term "immune system process" included interleukin (*IL1B*, *IL4*, *IL6*, *IL10*) (Dinarello, 1994; Choi and Reiser, 1998; Mocellin et al., 2004; Nemeth et al., 2004) and histocompatibility complex related (*HLA-B*) genes (Dupont and Ceppellini, 1989), which have been shown to be involved in immunological mechanisms of pain (Sato et al., 2002; de Rooij et al., 2009). This is also supported by published evidence for the further genes in this list, such as, *TNF* (Vassalli, 1992; Franchimont et al., 1999), *GCH1* (Schott et al., 1993) and *P2RX7* (Chen and Brosnan, 2006). The second major process group emerging from the functional genomics analysis of the key evidence for genetic modulation of clinical chronic pain was nitric oxide signaling, in particular metabolic processes, summarized in this context under the GO term "reactive oxygen species metabolic process" which includes the genes *IL6* (Deakin et al., 1995), *TNF* (Deakin et al., 1995; Katusic et al., 1998), *ESR1* (Clapauch et al., 2014), *IL10* (Cattaruzza et al., 2003), *GCH1* (Katusic et al., 1998; Zhang et al., 2007), *IL1B* (Katusic et al., 1998), *IL4* (Coccia et al., 2000), *P2RX7* (Gendron et al., 2003), *SOD2* (Fridovich, 1978). Furthermore, catecholamines including noradrenaline, adrenaline and dopamine have multiple functions in the brain and spinal cord including pain perception and processing (D'Mello and Dickenson, 2008). Catechol-*O*-methyltransferase, encoded by the *COMT* gene, is one of several enzymes that degrade dopamine, noradrenaline and adrenaline

⁷<http://www.ncbi.nlm.nih.gov/omim>

⁸<http://www.ncbi.nlm.nih.gov/gene>

⁹<http://www.genecards.org>

¹⁰<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes>

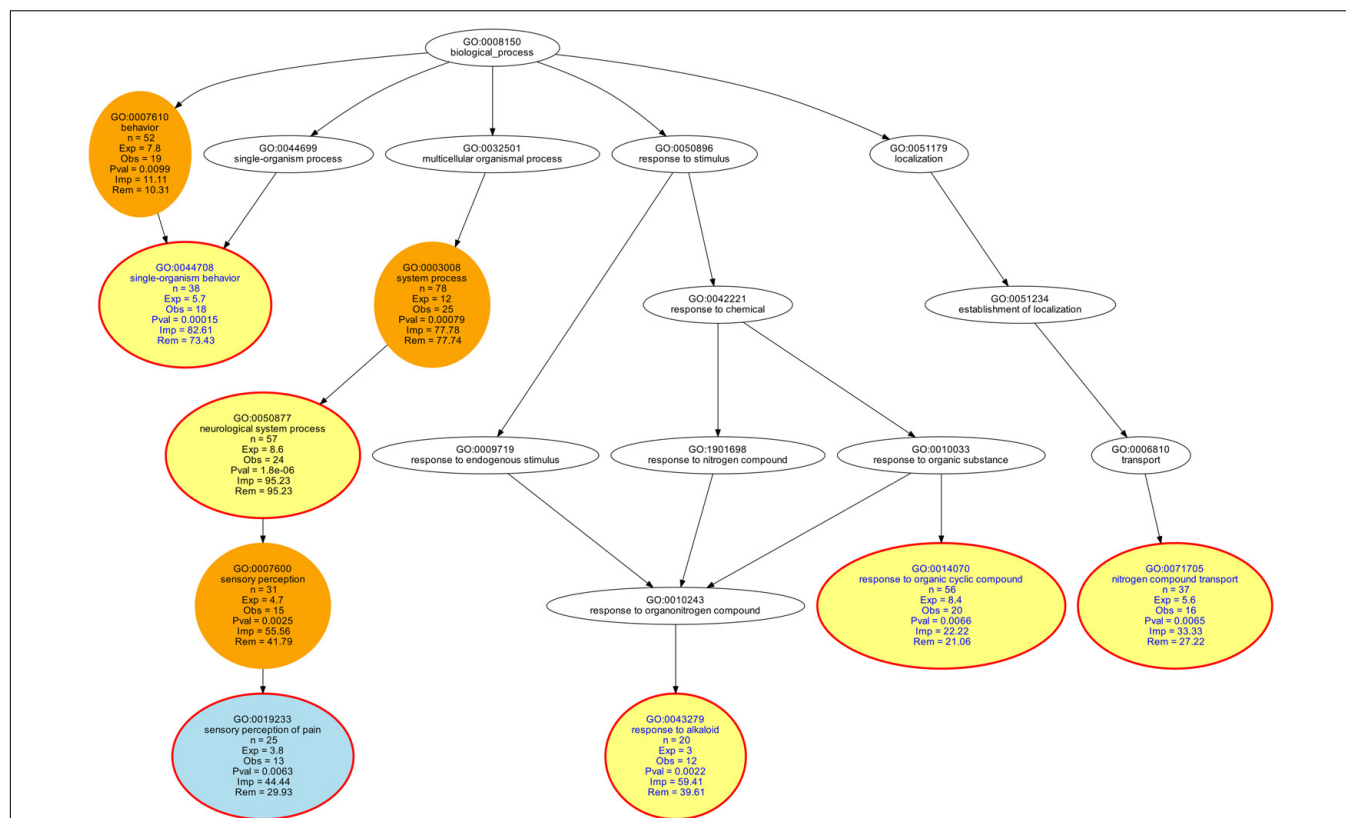


FIGURE 4 | Top-down representation of the annotations (GO terms) representing the taxonomy of the functional differences between the set of $d = 77$ genes included in the proposed NGS panel of persisting pain and two alternative proposals of genes modulating persisting pain in humans (Mogil, 2012; Zorina-Lichtenwalter et al., 2016). The figure represents the results of an over-representation analysis of the present set of $d = 77$ genes against the reference comprising the set intersection of the alternative gene lists. A p -value threshold of 0.01 and Bonferroni α -correction were applied. Significant terms are shown as colored circles with the number of member genes, the number of expected genes by chance and the significance of the deviation of the observed from the expected number of genes indicated (yellow = headline, red = significant term, blue = significant term located as a leaf at the end of a taxonomy in the polyhierarchy). The graphical representation follows the standard of the GO knowledgebase, where GO terms are related to each other by “is-a,” “part-of,” and “regulates” relationships forming a polyhierarchy organized in a directed acyclic graph (DAG, Thulasiraman and Swamy, 1992). The figure has been created using our R library “dbtORA” (<https://github.com/IME-TMP-FFM/dbtORA>, Lippmann et al., 2018) on the R software package (version 3.4.2 for Linux; <http://CRAN.R-project.org/>, R Development Core Team, 2008) and the freely available graph visualization software GraphViz (<http://www.graphviz.org>, Gansner and North, 2000).

and has become one of the most frequently addressed genes in pain research (Nackley et al., 2006).

Finally, subset #3 ($d = 30$) consists of genes repeatedly shown to play a role in the genetic modulation of persisting pain in humans or, by contrast, included a few novel items only recently published in the context of pain. This included members of the transient receptor potential (TRP) family (*TRPA1*, *TRPM8*, *TRPV4*) that are expressed at nociceptors and which are well established players in the perception of pain via their excitation by chemical, thermal or mechanical stimuli (Clapham, 2003). This similarly applies to the opioidergic system represented by the inclusion of the genes coding for the major opioid receptors (*OPRM1*, *OPRK1*, *OPRD1*), which have been associated with variations in pain or opioid response in various settings (Löttsch and Geisslinger, 2005). The most important of this group, the μ -opioid receptor encoded by the *OPRM1* gene, carries several variants of which the 118 A>G (rs1799971) has been studied most extensively since the early description of its association with a functional phenotype in humans (Löttsch et al., 2002).

Almost half of the present sets of genes were chosen based on a computational functional genomics analysis that attributed persisting pain to GO processes of “learning or memory” and “nervous system development” (Ultsch et al., 2016) as likely to reflect systemic features of persisting pain. This implied a functional bias and therefore, the present set of $d = 77$ genes (Figure 4) was analyzed whether this bias prevailed when comparing it with the alternative sets of human genes proposed to modulate persisting pain (Mogil, 2012; Zorina-Lichtenwalter et al., 2016). As applied previously (Lippmann et al., 2018), the biological roles of the set of $d = 77$ genes were queried from the Gene Ontology knowledgebase (GO)¹¹ (Ashburner et al., 2000) where the knowledge about the biological processes, the molecular functions and the cellular components of genes is formulated using a controlled and clearly defined vocabulary of GO terms. Particular biological roles of the set of $d = 77$ genes, among all human genes, were analyzed by

¹¹<http://www.geneontology.org/>

TABLE 3 | Current targeting of the genes included in the proposed NGS panel of persisting pain by novel drugs that are currently under active clinical development and include analgesia as the main clinical target or at least as one of the intended clinical indication.

Gene	Status	Drug	Action	Company
<i>ABHD12</i>	–	–	–	–
<i>ABHD16A</i>	–	–	–	–
<i>ABHD6</i>	Preclinical	Benzylpiperidin methanone	Acyamino-Acid-Releasing Enzyme	Scripps Research Institute
<i>ADCY1</i>	Under Active Development	NB-001	Adenylate Cyclase Inhibitors	Forever Cheer International
<i>ADRB2</i>	Phase II/III	Gencaro	Signal Transduction Modulators	ARCA
<i>BDNF</i>	Phase I	CXB-909	Nerve Growth Factor (NGF) Enhancers	Krenitsky
<i>CACNG2</i>	Preclinical	Hanfangchin	Calcium Channel Blockers	Millenia Hope Kaken
<i>CDK5</i>	Biological Testing	Litvinolin	CDK5/p25 Inhibitors	Hong Kong University
<i>CHRNA2</i>	Biological Testing	Epiboxidine	Nicotinic alpha4beta2 Receptor Agonists	Pfizer
<i>CNR1</i>	Registered	Epidiolex	Cannabinoid Receptor Agonists	InSys Therapeutics
<i>COMT</i>	Clinical	Nitecapone	Catechol-O-Methyl Transferase (COMT) Inhibitors	Orion
<i>CSF1</i>	–	–	–	–
<i>DLG4</i>	Preclinical	AB-125	Protein Inhibitors	Lundbeck University of Copenhagen
<i>DRD1</i>	Phase II/III	Ecopipam	Dopamine D1 Receptor (DRD1) Antagonists	Merck & Co.
<i>DRD2</i>	Phase II/III	Sarizotan hydrochloride	Dopamine D2 Receptor (DRD2) Antagonists	Newron
<i>DRD3</i>	Phase II	Brilaroxazine	D3 Receptor (DRD3) Agonists	Reviva Pharmaceuticals
<i>DRD4</i>	Biological Testing	Mesulergine hydrochloride	Dopamine Receptor Agonists	Novartis
<i>EGR1</i>	Phase II	Brivolidide	EGR1 Expression Inhibitors	Adynxx
<i>ESR1</i>	Phase II	Zindoxifene	Selective Estrogen Receptor Modulators	Evonik
<i>FAAH</i>	Phase I/II	Minerval	Fatty Acid Amide Hydrolase (FAAH) Inhibitors	Scripps Research Institute
<i>FKBP5</i>	Phase II	Barusiban	Oxytocin Receptor Antagonist	Ferring
<i>FOS</i>	Registered	Macrilen	FOS Expression Enhancers	Strongbridge Biopharma
<i>FYN</i>	Phase II	Bafetinib	Fyn Kinase Inhibitors	Nippon Shinyaku
<i>GABRA5</i>	Phase III	Ganaxolone	GABA(A) Receptor Modulators	Marinus Pharmaceuticals
<i>GALR2</i>	Preclinical	NAX-810-2	GAL2 Receptor Ligands	NeuroAdjuvants
<i>GCH1</i>	–	–	–	–
<i>GDNF</i>	Phase II	Edonerpic maleate	Signal Transduction Modulators	Toyama
<i>GFRA1</i>	–	–	–	–
<i>GPR132</i>	–	–	–	–
<i>GRIN1</i>	Phase II	Dimiracetam	Signal Transduction Modulators	Metys Pharmaceuticals
<i>GRIN2A</i>	Phase I	Dexanabinol	NMDA Receptor Antagonists	e-Therapeutics Pharms
<i>GRIN2B</i>	Phase I	Gacyclidine	NMDA Receptor Antagonists	INSERM
<i>GRM5</i>	Phase II	Mavoglurant	Signal Transduction Modulators	Novartis
<i>HCN2</i>	Clinical	Ivabradine	Adrenoceptor Antagonists	Servier
<i>HLA-DQB1</i>	–	–	–	–
<i>HLA-DRB1</i>	–	–	–	–
<i>HRH3</i>	Phase I	Immethridine	Histalean	Abbott
<i>HTR1A</i>	Phase II	Eltoprazine hydrochloride	5-HT1A Receptor Agonists	Elto Pharma
<i>HTR2A</i>	Phase II	Midomafetamine	5-HT2 Receptor Agonists	Assoc
<i>IL10</i>	Phase II	BT-063	Signal Transduction Modulators Anti-IL-10	Biotest AG
<i>IL1B</i>	Phase III	Resunab	IL-1beta Inhibitors	Corbus
<i>IL1R2</i>	–	–	–	–
<i>IL4</i>	–	–	–	–
<i>IL6</i>	Preclinical	Azintrel	Signal Transduction Modulators Anti-IL-6	Jazz Pharmaceuticals
<i>KCNS1</i>	Preclinical	Crotamine	Voltage-Gated K(V) Channel Blockers	Celtic Biotech
<i>KIT</i>	Phase II	Vatalanib succinate	KIT (C-KIT) Inhibitors	Novartis
<i>LTB4R</i>	Phase II	Coversin	Signal Transduction Modulators	Akari Therapeutics
<i>LTB4R2</i>	Phase II	Coversin	Signal Transduction Modulators	Akari Therapeutics
<i>NF1</i>	–	–	–	–
<i>NGF</i>	Phase III	Tanezumab	Anti-Nerve Growth Factor (NGF)	Pfizer
<i>NTF4</i>	–	–	–	–

(Continued)

TABLE 3 | Continued

Gene	Status	Drug	Action	Company
<i>NTRK1</i>	Phase II	Danuseritib	NTRK1 Inhibitors	Pfizer
<i>OPRD1</i>	Preclinical	Metenkephalin	Delta-Opioid Receptor Agonists	TNI Pharmaceuticals
<i>OPRK1</i>	Phase III	Morphine glucuronide	Opioid Receptor Agonists	PAION
<i>OPRM1</i>	Registered	Naltrexone	mu-Opioid Receptor Antagonists	Pfizer
<i>OXT</i>	Phase II	Barusiban	Oxytocin Receptor Antagonist	Ferring
<i>P2RX7</i>	Preclinical	BIL-06v	Anti-P2RX7	Biosceptre International
<i>PLCB1</i>	Biological Testing	Vinaxanthone	Signal Transduction Modulators	Roche
<i>PRKCG</i>	Phase III	Rydapt	Protein Kinase C (PKC) Inhibitors	Yeda
<i>PRNP</i>	–	–	–	–
<i>PTN</i>	–	–	–	–
<i>PTPRZ1</i>	–	–	–	–
<i>RELN</i>	Preclinical	IAIPs	Serine Protease Inhibitors	ProThera Biologics
<i>RET</i>	Phase II	Danuseritib	Ret (RET) Inhibitors	Pfizer
<i>RUNX1</i>	–	–	–	–
<i>S100B</i>	–	–	–	–
<i>SCN9A</i>	Phase III	Priralfinamide	Voltage-Gated Sodium Channel Blockers	Newron
<i>SLC6A4</i>	Phase II	Litoxetine	Signal Transduction Modulators	Sanofi
<i>SOD2</i>	Phase II	Avasopasem manganese	Superoxide Dismutase (SOD) Mimetics	MetaPhore
<i>TH</i>	–	–	–	–
<i>TLR4</i>	Phase II	Eritoran tetrasodium	Toll-Like Receptor 4 (TLR4) Antagonists	Eisai
<i>TNF</i>	Phase III	Givinostat hydrochloride	TNF-alpha Release Inhibitors	Italfarmaco
<i>TRPA1</i>	Phase II	Cannabidiol	TRPA1 Agonists	GW Pharmaceuticals
<i>TRPM8</i>	Phase II	Cannabidiol	TRPM8 Antagonists	GW Pharmaceuticals
<i>TRPV1</i>	Phase I/II	Resiniferatoxin	TRPV1 (Vanilloid VR1 Receptor) Agonists	Icos
<i>TRPV4</i>	Phase II	GSK-2798745	TRPV4 Antagonists	GlaxoSmithKline
<i>TSPO</i>	Clinical	[11C]CB-184	Translocator Protein (TSPO) Ligands	Tokyo Metrop Geriatr Hosp Inst Gerontol

The information was queried from the Thomson Reuters Integrity database at <https://integrity.thomson-pharma.com> on July 11, 2018.

means of over-representation analysis (ORA). This compared the occurrence of the particular GO terms associated with the present set of genes with their expected occurrence by chance (Backes et al., 2007). In contrast to enrichment analysis, any quantitative criteria such as gene expression values are disregarded (Backes et al., 2007). The analyses were performed using our R library “dbtORA” (Lippmann et al., 2018)¹² on the R software environment (version 3.4.2 for Linux; R Development Core Team, 2008)¹³.

Surprisingly, the results of this analysis indicated that the functional bias of the present gene set toward “learning or memory” (GO:0007611) and “nervous system development” (GO:0007399) was not maintained against the alternative gene sets. Instead, a few more general GO terms such as “behavior” (“single organism behavior,” GO:0044708), or “response to organic cyclic compound” (GO:0014070) and response to alkaloid (GO:0043279), which could be identified as morphine and cocaine when repeating the analysis with a less conservative α -correction (further details not shown), were overrepresented, as well as the pain specific term “sensory perception of pain” (GO:0019233). A possible explanation that the selection bias of

the present gene set was not maintained when comparing it with alternative proposals is that the two biological processes, “learning or memory” and “nervous system development,” reflect indeed an important biological function of persisting pain and even when choosing candidate genes without having these processes in mind as for the alternative gene sets, they are nevertheless included. This may be regarded as support for the present gene set as suitable candidates for future association studies with persisting pain phenotypes.

Although the present gene set has been assembled with a focus of a relevance to pain, many of its members have pharmacological implications. Specifically, 58 of the 77 genes (75%) have been chosen as targets of analgesics, approved or under current clinical development (Table 3). Moreover, several of the genes in the present NGS panel have been implicated in pharmacogenetic modulations of drug effects (Table 4). Possibly the most widely studied gene in analgesic research is *OPRM1* because coding for the primary target of opioids (Peiro et al., 2016). Several polymorphisms have been described in *OPRM1*, among which the best characterized may be rs1799971 (*OPRM1* 118A>G) that leads to an asparagine to aspartate substitution at the extracellular terminal of the receptor protein (Bond et al., 1998). May studies have addressed this variant (for reviews, see Walter et al., 2013; Somogyi et al., 2015).

¹²<https://github.com/IME-TMP-FFM/dbtORA>

¹³<http://CRAN.R-project.org/>

TABLE 4 | Summary of variants in genes included in the proposed NGS panel of persisting pain, that have been implicated in a pharmacogenetic context to modulate the effects of drugs administered for the treatment of pain or as disease modifying therapeutics in painful disease.

Modulated process	Gene	Variant	Affected drug	Findings	Reference
G protein coupled signaling	<i>COMT</i>	rs4680 (Val158Met)	Morphine	Carriers of val/val and val/met genotype required higher morphine dose compared to carriers of met/met genotype	Reyes-Gibby et al., 2007
	<i>DRD2</i>	rs6275	Heroin	Polymorphism is associated with decreased likelihood of headache disorders	Cargnin et al., 2014
	<i>DRD4</i>	rs1800955	Heroin	Polymorphism had lower pain threshold versus CC/CT controls	Ho et al., 2008
	<i>OPRM1</i>	rs1799971 (A118G)	Various opioids	Tendency toward increased pain in dose-dependent manner with the μ -opioid receptor variant 118G	Lötsch et al., 2009c
	<i>OPRK1</i>	rs1051660	Morphine	Patients with the polymorphism and cancer-related pain may require a reduced dose escalation of morphine	Chatti et al., 2017
Neurotransmitters	<i>BDNF</i>	rs6265	Various opioids	Polymorphism is associated with decreased likelihood of headache disorders	Cargnin et al., 2014
	<i>HTR2A</i>	rs12584920	Various opioids	Increased likelihood of having chronic widespread pain	Nicholl et al., 2011
Ion Channels	<i>TRPV1</i>	7 intronic SNPs	Capsaicin	TRPV1 polymorphisms had only 50% of the mRNA and protein expression levels of normally sensing subjects	Park et al., 2007
Proinflammatory Cytokines	<i>IL6</i>	rs1800795	Etanercept	Polymorphism is associated with increased response to adalimumab, etanercept or infliximab in people with painful Arthritis	Davila-Fajardo et al., 2014
Other	<i>ESR1</i>	rs2234693	Leflunomide	Polymorphism is associated with increased response to leflunomide in women with painful Arthritis	Dziedziejko et al., 2011
	<i>FAAH</i>	rs2295632	Various opioids	Polymorphism is associated with increased risk of Respiratory Insufficiency	Biesiada et al., 2014
	<i>TLR4</i>	rs4986790	Methotrexate	Polymorphism associated with increased risk of adverse drug events when treated with folic acid and methotrexate in people with Arthritis	Kooloos et al., 2010
	<i>TNF</i>	rs361525	Infliximab	Polymorphism is associated with increased response to infliximab in people with painful Arthritis	Maxwell et al., 2008

The information was derived by literature search and by querying the Pharmacogenetics Research Network/Knowledge base at <http://www.pharmgkb.org> (accessed in July 2018). Only key or example references are given.

Summarizing its effects, the variant is associated with decreased receptor expression and signaling efficiency (Oertel et al., 2012) which leads to reproducibly reduced pharmacodynamic effects in human experimental settings while the effect size seems insufficient to be a major factor of opioid response in clinical settings, despite several reports of modulations of opioid demands or side effects. For example, subjects carrying the 118A>G variant were found to have a reduced response to morphine treatment (Hwang et al., 2014), reduced analgesic response to alfentanil (Oertel et al., 2006) and demanded higher doses of morphine for pain relief (Klepstad et al., 2004; Hwang et al., 2014). However, the importance of this variant seems to be comparatively high in patients with an Asian ethnic background,

which might be related to the higher allelic frequency as compared to other ethnicities. *COMT* is a key modulator of dopaminergic neurotransmission and in the signaling response to opioids The Val158Met polymorphism (rs4680) causes an amino acid substitution in the enzyme, which reduced the enzyme active to a forth (Peiro et al., 2016). Carriers of the homozygous Met/Met variant had lower morphine requirements than those with a the wild type *COMT* (Rakvag et al., 2005). Furthermore, a modulation of the effects of *TRPV1* targeting analgesics is supported by observations that intronic *TRPV1* variants were associated with insensitivity to capsaicin (Park et al., 2007) while the coding *TRPV1* variant rs8065080 was associated with altered responses to experimentally induced pain

(Kim et al., 2004). Moreover, gain-of-function mutations in *TRPV1* have been associated with increased pain sensitivity (Boukalova et al., 2014), for which *TRPV1* antagonists would enable a specific pharmacogenetics-based personalized cure.

CONCLUSION

The breakthrough in mapping the whole human genome (Lander et al., 2001; Venter et al., 2001) along with genome wide association studies (GWAS) has led to rapid advances in the knowledge of the genetic bases of human diseases (Wellcome Trust Case Control and Consortium, 2007). Genetic research in pain medicine has directed to the recognition of genes in which variants influence pain behavior, post-operative drug requirements, and the temporal developments of pain toward persistence (James, 2013). While many candidate gene association studies have identified multiple genes relevant for pain phenotypes (Fillingim et al., 2008), pain related genetic studies have so far been owned by investigations of a limited number of genes. Roughly ten genes or gene complexes account for over half of the extant findings and several of these candidate gene associations have held up in replication (Mogil, 2012). The selection of variants has been limited and they have been addressed in most studies repeatedly, leading to the perception that genetic research in pain produces often unsatisfactory results (Mogil, 2009). However, this may soon change with the arise of new technologies. In this manuscript, we present a validated NGS assay for a set of 77 genes supported by empirical evidence and computational functional genomics analyses as relevant

factors modulating the risk for persisting pain or its clinical picture.

AUTHOR CONTRIBUTIONS

JL, DK, and EK conceived and designed the experiments. DK performed the experiments. JL and DK analyzed the data and wrote the paper. CL provided methodological expertise and bioinformatical tools. DK and JL interpreted the results. EK and MK provided DNA samples.

FUNDING

This work has been funded by the European Union Seventh Framework Programme (FP7/2007 – 2013) under grant agreement no. 602919 (“GLORIA”, EK and JL) and the LandesOffensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für Translationale Medizin und Pharmakologie (JL). These public funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.01008/full#supplementary-material>

REFERENCES

- Abbadie, C., Besson, J. M., and Calvino, B. (1994). c-Fos expression in the spinal cord and pain-related symptoms induced by chronic arthritis in the rat are prevented by pretreatment with *Freund adjuvant*. *J. Neurosci.* 14, 5865–5871. doi: 10.1523/JNEUROSCI.14-10-05865.1994
- Achkar, J.-P., Klei, L., De Bakker, P. I. W., Bellone, G., Rebert, N., Scott, R., et al. (2012). Amino acid position 11 of HLA-DRβ1 is a major determinant of chromosome 6p association with ulcerative colitis. *Genes Immun.* 13, 245–252. doi: 10.1038/gene.2011.79
- Al-Eitan, L. N., Jaradat, S. A., Su, S. Y., Tay, G. K., and Hulse, G. K. (2012). Mu opioid receptor (OPRM1) as a predictor of treatment outcome in opiate-dependent individuals of Arab descent. *Pharmgen. Pers. Med.* 5, 99–111. doi: 10.2147/PGPM.S33351
- Alvarado, S., Tajerian, M., Millicamps, M., Suderman, M., Stone, L. S., and Szyf, M. (2013). Peripheral nerve injury is accompanied by chronic transcriptome-wide changes in the mouse prefrontal cortex. *Mol. Pain* 9:21. doi: 10.1186/1744-8069-9-21
- Andreoli, V., De Marco, E. V., Trecroci, F., Cittadella, R., Di Palma, G., and Gambardella, A. (2014). Potential involvement of GRIN2B encoding the NMDA receptor subunit NR2B in the spectrum of Alzheimer's disease. *J. Neural Transm. (Vienna)* 121, 533–542.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Gene Ontol. Consortium. Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y. A., et al. (2007). GeneTrail-advanced gene set enrichment analysis. *Nucleic Acids Res.* 35, W186–W192. doi: 10.1093/nar/gkm323
- Biesiada, J., Chidambaran, V., Wagner, M., Zhang, X., Martin, L. J., Meller, J., et al. (2014). Genetic risk signatures of opioid-induced respiratory depression following pediatric tonsillectomy. *Pharmacogenomics* 15, 1749–1762. doi: 10.2217/pgs.14.137
- Bond, C., Laforge, K. S., Tian, M., Melia, D., Zhang, S., and Borg, L. (1998). Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction. *Proc. Natl. Acad. Sci. U.S.A.* 95, 9608–9613. doi: 10.1073/pnas.95.16.9608
- Boukalova, S., Touska, F., Marsakova, L., Hynkova, A., Sura, L., Chvojka, S., et al. (2014). Gain-of-function mutations in the transient receptor potential channels TRPV1 and TRPA1: how painful? *Physiol. Res.* 63(Suppl. 1), S205–S213.
- Bourinet, E., Altier, C., Hildebrand, M. E., Trang, T., Salter, M. W., and Zamponi, G. W. (2014). Calcium-permeable ion channels in pain signaling. *Physiol. Rev.* 94, 81–140. doi: 10.1152/physrev.00023.2013
- Bravo, L., Torres-Sanchez, S., Alba-Delgado, C., Mico, J. A., and Berrocoso, E. (2014). Pain exacerbates chronic mild stress-induced changes in noradrenergic transmission in rats. *Eur. Neuropsychopharmacol.* 24, 996–1003. doi: 10.1016/j.euroneuro.2014.01.011
- Bravo-Hernández, M., Corleto, J. A., Barragán-Iglesias, P., González-Ramírez, R., Pineda-Farías, J. B., Félix, R., et al. (2016). The α5 subunit containing GABAA receptors contribute to chronic pain. *Pain* 157, 613–626. doi: 10.1097/j.pain.0000000000000410
- Bruera, G., Pepe, F., Malapelle, U., Pisapia, P., Mas, A. D., Di Giacomo, D., et al. (2018). KRAS, NRAS and BRAF mutations detected by next generation sequencing, and differential clinical outcome in metastatic colorectal cancer (MCRC) patients treated with first line FIr-B/FOX adding bevacizumab (BEV) to triplet chemotherapy. *Oncotarget* 9, 26279–26290. doi: 10.18632/oncotarget.25180
- Buchheit, T., Van De Ven, T., and Shaw, A. (2012). Epigenetics and the transition from acute to chronic pain. *Pain. Med.* 13, 1474–1490. doi: 10.1111/j.1526-4637.2012.01488.x

- Buskila, D., Cohen, H., Neumann, L., and Ebstein, R. P. (2004). An association between fibromyalgia and the dopamine D4 receptor exon III repeat polymorphism and relationship to novelty seeking personality traits. *Mol. Psychiatry* 9, 730–731. doi: 10.1038/sj.mp.4001568
- Cajanus, K., Holmström, E. J., Wessman, M., Anttila, V., Kaunisto, M. A., and Kalso, E. (2016). Effect of endocannabinoid degradation on pain: role of FAAH polymorphisms in experimental and postoperative pain in women treated for breast cancer. *Pain* 157, 361–369. doi: 10.1097/j.pain.0000000000000398
- Camilleri, M., Klee, E. W., Shin, A., Carlson, P., Li, Y., Grover, M., et al. (2014). Irritable bowel syndrome-diarrhea: characterization of genotype by exome sequencing, and phenotypes of bile acid synthesis and colonic transit. *Am. J. Physiol. Gastrointest. Liver Physiol.* 306, G13–G26. doi: 10.1152/ajpgi.00294.2013
- Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., et al. (2004). The gene ontology annotation (GOA) database: sharing knowledge in uniprot with Gene Ontology. *Nucleic Acids Res.* 32, D262–D266. doi: 10.1093/nar/gkh021
- Cargnin, S., Viana, M., Sances, G., Bianchi, M., Ghiotto, N., Tassorelli, C., et al. (2014). Combined effect of common gene variants on response to drug withdrawal therapy in medication overuse headache. *Eur. J. Clin. Pharmacol.* 70, 1195–1202. doi: 10.1007/s00228-014-1726-6
- Cattaruzza, M., Słodowski, W., Stojakovic, M., Krzesz, R., and Hecker, M. (2003). Interleukin-10 induction of nitric-oxide synthase expression attenuates CD40-mediated interleukin-12 synthesis in human endothelial cells. *J. Biol. Chem.* 278, 37874–37880. doi: 10.1074/jbc.M301670200
- Ceolin, L., Siqueira, D. R., Romitti, M., Ferreira, C. V., and Maia, A. L. (2012). Molecular basis of medullary thyroid carcinoma: the role of RET polymorphisms. *Int. J. Mol. Sci.* 13, 221–239. doi: 10.3390/ijms13010221
- Chatti, I., Woillard, J. B., Mili, A., Creveaux, I., Ben Charfeddine, I., Feki, J., et al. (2017). Genetic analysis of mu and kappa opioid receptor and COMT enzyme in cancer pain tunisian patients under opioid treatment. *Iran J. Public Health* 46, 1704–1711.
- Chen, C. L., Broom, D. C., Liu, Y., De Nooij, J. C., Li, Z., Cen, C., et al. (2006). Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. *Neuron* 49, 365–377. doi: 10.1016/j.neuron.2005.10.036
- Chen, D., Zhang, T.-L., and Wang, L.-M. (2014). The association of CSF-1 gene polymorphism with chronic periodontitis in the Han Chinese population. *J. Periodontol.* 85, e304–e312. doi: 10.1902/jop.2014.130688
- Chen, L., and Brosnan, C. F. (2006). Regulation of immune response by P2X7 receptor. *Crit. Rev. Immunol.* 26, 499–513. doi: 10.1615/CritRevImmunol.v26.i6.30
- Chen, X., Williamson, V. S., An, S.-S., Hetteima, J. M., Aggen, S. H., Neale, M. C., et al. (2008). Cannabinoid receptor 1 gene association with nicotine dependence. *Arch. Gen. Psychiatry* 65, 816–824. doi: 10.1001/archpsyc.65.7.816
- Choi, P., and Reiser, H. (1998). IL-4: role in disease and regulation of production. *Clin. Exp. Immunol.* 113, 317–319. doi: 10.1046/j.1365-2249.1998.00690.x
- Clapauch, R., Mourao, A. F., Mecnas, A. S., Maranhao, P. A., Rossini, A., and Bouskela, E. (2014). Endothelial function and insulin resistance in early postmenopausal women with cardiovascular risk factors: importance of ESR1 and NOS3 polymorphisms. *PLoS One* 9:e103444. doi: 10.1371/journal.pone.0103444
- Clapham, D. E. (2003). TRP channels as cellular sensors. *Nature* 426, 517–524. doi: 10.1038/nature02196
- Coccia, E. M., Stellacci, E., Marziali, G., Weiss, G., and Battistini, A. (2000). IFN-gamma and IL-4 differently regulate inducible NO synthase gene expression through IRF-1 modulation. *Int. Immunol.* 12, 977–985. doi: 10.1093/intimm/12.7.977
- Corominas, R., Ribases, M., Camiña, M., Cuenca-León, E., Pardo, J., Boronat, S., et al. (2009). Two-stage case-control association study of dopamine-related genes and migraine. *BMC Med. Genet.* 10:95. doi: 10.1186/1471-2350-10-95
- Coskun, S., Varol, S., Ozdemir, H. H., Agacayak, E., Aydın, B., Kapan, O., et al. (2016). Association of brain-derived neurotrophic factor and nerve growth factor gene polymorphisms with susceptibility to migraine. *Neuropsychiatr. Dis. Treat.* 12, 1779–1785. doi: 10.2147/NDT.S108814
- Costigan, M., Belfer, I., Griffin, R. S., Dai, F., Barrett, L. B., Coppola, G., et al. (2010). Multiple chronic pain states are associated with a common amino acid-changing allele in KCNS1. *Brain* 133, 2519–2527. doi: 10.1093/brain/awq195
- Coyle, J. T., and Tsai, G. (2004). NMDA receptor function, neuroplasticity, and the pathophysiology of schizophrenia. *Int. Rev. Neurobiol.* 59, 491–515. doi: 10.1016/S0074-7742(04)59019-0
- Crock, L. W., Stemler, K. M., Song, D. G., Abbosh, P., Vogt, S. K., Qiu, C. S., et al. (2012). Metabotropic glutamate receptor 5 (mGluR5) regulates bladder nociception. *Mol. Pain* 8:20. doi: 10.1186/1744-8069-8-20
- Davila-Fajardo, C. L., Marquez, A., Pascual-Salcedo, D., Moreno Ramos, M. J., Garcia-Portales, R., Magro, C., et al. (2014). Confirmation of -174G/C interleukin-6 gene promoter polymorphism as a genetic marker predicting antitumor necrosis factor treatment outcome. *Pharmacogenet. Genom.* 24, 1–5. doi: 10.1097/FPC.0000000000000013
- De Luca, C., Rappa, A. G., Gragnano, G., Malapelle, U., Troncone, G., and Barberis, M. (2018). Idylla assay and next generation sequencing: an integrated EGFR mutational testing algorithm. *J. Clin. Pathol.* 71, 745–750. doi: 10.1136/jclinpath-2018-205197
- de Rooij, A. M., Florencia Gosso, M., Haasnoot, G. W., Marinus, J., Verduijn, W., Claas, F. H., et al. (2009). HLA-B62 and HLA-DQ8 are associated with complex regional pain syndrome with fixed dystonia. *Pain* 145, 82–85. doi: 10.1016/j.pain.2009.05.015
- Deakin, A. M., Payne, A. N., Whittle, B. J., and Moncada, S. (1995). The modulation of IL-6 and TNF-alpha release by nitric oxide following stimulation of J774 cells with LPS and IFN-gamma. *Cytokine* 7, 408–416. doi: 10.1006/cyto.1995.0056
- Diatchenko, L., Anderson, A. D., Slade, G. D., Fillingim, R. B., Shabalina, S. A., Higgins, T. J., et al. (2006). Three major haplotypes of the beta2 adrenergic receptor define psychological profile, blood pressure, and the risk for development of a common musculoskeletal pain disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 141B, 449–462. doi: 10.1002/ajmg.b.30324
- Diatchenko, L., Slade, G. D., Nackley, A. G., Bhalang, K., Sigurdsson, A., Belfer, I., et al. (2005). Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum. Mol. Genet.* 14, 135–143. doi: 10.1093/hmg/ddi013
- Dinarello, C. A. (1994). The biological properties of interleukin-1. *Eur. Cytokine Netw.* 5, 517–531.
- Dineley, K. T., Pandya, A. A., and Yakel, J. L. (2015). Nicotinic ACh receptors as therapeutic targets in CNS disorders. *Trends Pharmacol. Sci.* 36, 96–108. doi: 10.1016/j.tips.2014.12.002
- D'Mello, R., and Dickenson, A. H. (2008). Spinal cord mechanisms of pain. *Br. J. Anaesth.* 101, 8–16. doi: 10.1093/bja/aen088
- Doehring, A., Kirchhof, A., and Lötsch, J. (2009). Genetic diagnostics of functional variants of the human dopamine D2 receptor gene. *Psychiatr. Genet.* 19, 259–268. doi: 10.1097/YPG.0b013e32832d0941
- Doehring, A., Küsener, N., Flühr, K., Neddermeyer, T. J., Schneider, G., and Lötsch, J. (2011). Effect sizes in experimental pain produced by gender, genetic variants and sensitization procedures. *PLoS One* 6:e17724. doi: 10.1371/journal.pone.0017724
- Dominguez, C. A., Kalliomaki, M., Gunnarsson, U., Moen, A., Sandblom, G., Kockum, I., et al. (2013). The DQB1 *03:02 HLA haplotype is associated with increased risk of chronic pain after inguinal hernia surgery and lumbar disc herniation. *Pain* 154, 427–433. doi: 10.1016/j.pain.2012.12.003
- Duan, G., Guo, S., Zhang, Y., Ying, Y., Huang, P., Wang, Q., et al. (2015). The effect of SCN9A variation on basal pain sensitivity in the general population: an experimental study in young women. *J. Pain* 16, 971–980. doi: 10.1016/j.jpain.2015.06.011
- Dupont, B., and Ceppellini, R. (1989). *Immunobiology of HLA*. New York, NY: Springer-Verlag.
- Dutta, S., Gangopadhyay, P. K., Sinha, S., Chatterjee, A., Ghosh, S., and Rajamma, U. (2011). An association analysis of reelin gene (RELN) polymorphisms with childhood epilepsy in eastern Indian population from West Bengal. *Cell. Mol. Neurobiol.* 31, 45–56. doi: 10.1007/s10571-010-9551-7
- Dziedzic, V., Kurawski, M., Safranow, K., Chlubek, D., and Pawlik, A. (2011). The effect of ESR1 and ESR2 gene polymorphisms on the outcome of rheumatoid arthritis treatment with leflunomide. *Pharmacogenomics* 12, 41–47. doi: 10.2217/pgs.10.164
- Edwards, R. R. (2006). Genetic predictors of acute and chronic pain. *Curr. Rheumatol. Rep.* 8, 411–417. doi: 10.1007/s11926-006-0034-2
- Emigh, T. H. (1980). A comparison of tests for hardy-weinberg equilibrium. *Biometrics* 36, 627–642. doi: 10.2307/2556115

- Ersig, A. L., Schutte, D. L., Standley, J., Leslie, E., Zimmerman, B., Kleiber, C., et al. (2017). Relationship of genetic variants with procedural pain, anxiety, and distress in children. *Biol. Res. Nurs.* 19, 339–349. doi: 10.1177/1099800417692878
- Eskola, P. J., Kjaer, P., Daavittila, I. M., Solovieva, S., Okuloff, A., Sorensen, J. S., et al. (2010). Genetic risk factors of disc degeneration among 12–14-year-old danish children: a population study. *Int. J. Mol. Epidemiol. Genet.* 1, 158–165.
- Feng, Y., Egan, B., and Wang, J. (2016). Genetic factors in intervertebral disc degeneration. *Genes Dis.* 3, 178–185. doi: 10.1016/j.gendis.2016.04.005
- Feng, Y., Zhao, X., Zhou, C., Yang, L., Liu, Y., Bian, C., et al. (2013). The associations between the Val158Met in the catechol-O-methyltransferase (COMT) gene and the risk of uterine leiomyoma (ULM). *Gene* 529, 296–299. doi: 10.1016/j.gene.2013.07.019
- Fillingim, R. B., Wallace, M. R., Herbstman, D. M., Ribeiro-Dasilva, M., and Staud, R. (2008). Genetic contributions to pain: a review of findings in humans. *Oral Dis.* 14, 673–682. doi: 10.1111/j.1601-0825.2008.01458.x
- Florio, S. K., Loh, C., Huang, S. M., Iwamaye, A. E., Kitto, K. F., Fowler, K. W., et al. (2009). Disruption of nNOS-PSD95 protein-protein interaction inhibits acute thermal hyperalgesia and chronic mechanical allodynia in rodents. *Br. J. Pharmacol.* 158, 494–506. doi: 10.1111/j.1476-5381.2009.00300.x
- Franchimont, D., Martens, H., Hagelstein, M. T., Louis, E., Dewe, W., Chrousos, G. P., et al. (1999). Tumor necrosis factor alpha decreases, and interleukin-10 increases, the sensitivity of human monocytes to dexamethasone: potential regulation of the glucocorticoid receptor. *J. Clin. Endocrinol. Metab.* 84, 2834–2839.
- Fridovich, I. (1978). The biology of oxygen radicals. *Science* 201, 875–880. doi: 10.1126/science.210504
- Fujii, T., Ota, M., Hori, H., Hattori, K., Teraishi, T., Matsuo, J., et al. (2014). The common functional FKBP5 variant rs1360780 is associated with altered cognitive function in aged individuals. *Sci. Rep.* 4:6696. doi: 10.1038/srep06696
- Gadotti, V. M., and Zamponi, G. W. (2011). Cellular prion protein protects from inflammatory and neuropathic pain. *Mol. Pain* 7:59. doi: 10.1186/1744-8069-7-59
- Gansner, E. R., and North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Softw. Pract. Exp.* 30, 1203–1233. doi: 10.1002/1097-024X(200009)30:11<1203::AID-SPE338>3.0.CO;2-N
- Garassino, M. C., Piva, S., La Verde, N., Spagnoletti, I., Iorno, V., Carbone, C., et al. (2013). Randomised phase II trial (NCT00637975) evaluating activity and toxicity of two different escalating strategies for pregabalin and oxycodone combination therapy for neuropathic pain in cancer patients. *PLoS One* 8:e59981. doi: 10.1371/journal.pone.0059981
- Garriock, H. A., Tanowitz, M., Kraft, J. B., Dang, V. C., Peters, E. J., Jenkins, G. D., et al. (2010). Association of mu-opioid receptor variants and response to citalopram treatment in major depressive disorder. *Am. J. Psychiatry* 167, 565–573. doi: 10.1176/appi.ajp.2009.08081167
- Gębura, K., Świerkot, J., Wysoczańska, B., Korman, L., Nowak, B., Wiland, P., et al. (2017). Polymorphisms within genes involved in regulation of the NF-κB pathway in patients with rheumatoid arthritis. *Int. J. Mol. Sci.* 18, E1432. doi: 10.3390/ijms18071432
- Gendron, F. P., Chalimoniuk, M., Strosznajder, J., Shen, S., Gonzalez, F. A., Weisman, G. A., et al. (2003). P2X7 nucleotide receptor activation enhances IFN gamma-induced type II nitric oxide synthase activity in BV-2 microglial cells. *J. Neurochem.* 87, 344–352. doi: 10.1046/j.1471-4159.2003.01995.x
- Georgi, A., Jamra, R. A., Klein, K., Vilella, A. W., Schumacher, J., Becker, T., et al. (2007). Possible association between genetic variants at the GRIN1 gene and schizophrenia with lifetime history of depressive symptoms in a German sample. *Psychiatr. Genet.* 17, 308–310. doi: 10.1097/YPG.0b013e3280c1e5fb
- Gibbs, J. E., and Ray, D. W. (2013). The role of the circadian clock in rheumatoid arthritis. *Arthritis Res. Ther.* 15:205. doi: 10.1186/ar4146
- Gilon, I., and Ghasemlou, N. (2014). Chronobiology of chronic pain: focus on diurnal rhythmicity of neuropathic pain. *Curr. Opin. Support. Palliat. Care* 8, 429–436. doi: 10.1097/SPC.0000000000000085
- Goodin, B. R., Ness, T. J., and Robbins, M. T. (2015). Oxytocin – A multifunctional analgesic for chronic deep tissue pain. *Curr. Pharm. Des.* 21, 906–913. doi: 10.2174/1381612820666141027111843
- Goodyer, I. M., Croudace, T., Dunn, V., Herbert, J., and Jones, P. B. (2010). Cohort profile: risk patterns and processes for psychopathology emerging during adolescence: the ROOTS project. *Int. J. Epidemiol.* 39, 361–369. doi: 10.1093/ije/dyp173
- Gramage, E., and Herradon, G. (2010). Genetic deletion of pleiotrophin leads to disruption of spinal nociceptive transmission: evidence for pleiotrophin modulation of morphine-induced analgesia. *Eur. J. Pharmacol.* 647, 97–102. doi: 10.1016/j.ejphar.2010.08.029
- Greenbaum, L., Tegeder, I., Barhum, Y., Melamed, E., Roditi, Y., and Djaldetti, R. (2012). Contribution of genetic variants to pain susceptibility in Parkinson disease. *Eur. J. Pain* 16, 1243–1250. doi: 10.1002/j.1532-2149.2012.00134.x
- Guerrero, M., Urbano, M., Brown, S. J., Cayanan, C., Ferguson, J., Cameron, M., et al. (2010). “Optimization and characterization of an opioid kappa receptor (OPRK1) antagonist,” in *Probe Reports from the NIH Molecular Libraries Program*. Bethesda (MD): National Center for Biotechnology Information (US).
- Hack, L. M., Kalsi, G., Aliev, F., Kuo, P.-H., Prescott, C. A., Patterson, D. G., et al. (2011). Limited associations of dopamine system genes with alcohol dependence and related traits in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD). *Alcohol. Clin. Exp. Res.* 35, 376–385. doi: 10.1111/j.1530-0277.2010.01353.x
- Hagelberg, N., Forssell, H., Aalto, S., Rinne, J. O., Scheinin, H., Taiminen, T., et al. (2003). Altered dopamine D2 receptor binding in atypical facial pain. *Pain* 106, 43–48. doi: 10.1016/S0304-3959(03)00275-6
- Hagelberg, N., Jaaskelainen, S. K., Martikainen, I. K., Mansikka, H., Forssell, H., Scheinin, H., et al. (2004). Striatal dopamine D2 receptors in modulation of pain in humans: a review. *Eur. J. Pharmacol.* 500, 187–192. doi: 10.1016/j.ejphar.2004.07.024
- Hahn, W.-H., Suh, J.-S., and Cho, B.-S. (2011). Linkage and association study of neurotrophins and their receptors as novel susceptibility genes for childhood IgA nephropathy. *Pediatr. Res.* 69, 299–305. doi: 10.1203/PDR.0b013e31820b9365
- Haus, E., Sackett-Lundeen, L., and Smolensky, M. H. (2012). Rheumatoid arthritis: circadian rhythms in disease activity, signs and symptoms, and rationale for chronotherapy with corticosteroids and other medications. *Bull. NYU Hosp. Jt. Dis.* 70(Suppl. 1), 3–10.
- He, G.-H., Cai, W.-K., Zhang, J.-B., Ma, C.-Y., Yan, F., Lu, J., et al. (2016). Associations of polymorphisms in HRH2, HRH3, DAO, and HNMT genes with risk of chronic heart failure. *Biomed. Res. Int.* 2016:1208476. doi: 10.1155/2016/1208476
- Herman, A. I., and Balogh, K. N. (2012). Polymorphisms of the serotonin transporter and receptor genes: susceptibility to substance abuse. *Subst. Abuse Rehabil.* 3, 49–57. doi: 10.2147/SAR.S25864
- Hillard, C. J., Weinlander, K. M., and Stuhr, K. L. (2012). Contributions of endocannabinoid signaling to psychiatric disorders in humans: genetic and biochemical evidence. *Neuroscience* 204, 207–229. doi: 10.1016/j.neuroscience.2011.11.020
- Ho, A. M., Tang, N. L., Cheung, B. K., and Stadlin, A. (2008). Dopamine receptor D4 gene -521C/T polymorphism is associated with opioid dependence through cold-pain responses. *Ann. N. Y. Acad. Sci.* 1139, 20–26. doi: 10.1196/annals.1432.054
- Hocking, L. J., Smith, B. H., Jones, G. T., Reid, D. M., Strachan, D. P., and Macfarlane, G. J. (2010). Genetic variation in the beta2-adrenergic receptor but not catecholamine-O-methyltransferase predisposes to chronic pain: results from the 1958 British Birth Cohort study. *Pain* 149, 143–151. doi: 10.1016/j.pain.2010.01.023
- Hogart, A., Nagarajan, R. P., Patzel, K. A., Yasui, D. H., and Lasalle, J. M. (2007). 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. *Hum. Mol. Genet.* 16, 691–703. doi: 10.1093/hmg/ddm014
- Hohmann, S. W., Angioni, C., Tunaru, S., Lee, S., Woolf, C. J., Offermanns, S., et al. (2017). The G2A receptor (GPR132) contributes to oxaliplatin-induced mechanical pain hypersensitivity. *Sci. Rep.* 7:446. doi: 10.1038/s41598-017-00591-0
- Hough, L. B., and Rice, F. L. (2011). H3 receptors and pain modulation: peripheral, spinal, and brain interactions. *J. Pharmacol. Exp. Ther.* 336, 30–37. doi: 10.1124/jpet.110.171264
- Hu, Y., Tang, W., Liu, R., Dong, Z., Chen, X., Pan, M., et al. (2014). Higher prevalence of migraine in essential tremor: a case-control study. *Cephalalgia* 34, 1142–1149. doi: 10.1177/0333102414531153

- Huang, L., Adachi, N., Nagaro, T., Liu, K., and Arai, T. (2007). Histaminergic involvement in neuropathic pain produced by partial ligation of the sciatic nerve in rats. *Reg. Anesth Pain. Med.* 32, 124–129. doi: 10.1097/00115550-200703000-00006
- Hulse, R. P., Donaldson, L. F., and Wynick, D. (2012). Peripheral galanin receptor 2 as a target for the modulation of pain. *Pain Res. Treat.* 2012, 545386. doi: 10.1155/2012/545386
- Hutchinson, M. R., Zhang, Y., Shridhar, M., Evans, J. H., Buchanan, M. M., Zhao, T. X., et al. (2010). Evidence that opioids may have toll-like receptor 4 and MD-2 effects. *Brain Behav. Immun.* 24, 83–95. doi: 10.1016/j.bbi.2009.08.004
- Hwang, I. C., Park, J. Y., Myung, S. K., Ahn, H. Y., Fukuda, K., and Liao, Q. (2014). OPRM1 A118G gene variant and postoperative opioid requirement: a systematic review and meta-analysis. *Anesthesiology* 121, 825–834. doi: 10.1097/ALN.0000000000000405
- Ide, S., Nishizawa, D., Fukuda, K.-I., Kasai, S., Hasegawa, J., Hayashida, M., et al. (2014). Haplotypes of P2RX7 gene polymorphisms are associated with both cold pain sensitivity and analgesic effect of fentanyl. *Mol. Pain* 10:75. doi: 10.1186/1744-8069-10-75
- Jaaskelainen, S. K., Lindholm, P., Valmunen, T., Pesonen, U., Taiminen, T., Virtanen, A., et al. (2014). Variation in the dopamine D2 receptor gene plays a key role in human pain and its modulation by transcranial magnetic stimulation. *Pain* 155, 2180–2187. doi: 10.1016/j.pain.2014.08.029
- James, S. (2013). Human pain and genetics: some basics. *Br. J. Pain* 7, 171–178. doi: 10.1177/2049463713506408
- Jayamanne, A., Greenwood, R., Mitchell, V. A., Aslan, S., Piomelli, D., and Vaughan, C. W. (2006). Actions of the FAAH inhibitor URB597 in neuropathic and inflammatory chronic pain models. *Br. J. Pharmacol.* 147, 281–288. doi: 10.1038/sj.bjp.0706510
- Jhun, E., He, Y., Yao, Y., Molokie, R. E., Wilkie, D. J., and Wang, Z. J. (2014). Dopamine D3 receptor Ser9Gly and catechol-o-methyltransferase Val158Met polymorphisms and acute pain in sickle cell disease. *Anesth. Analg.* 119, 1201–1207. doi: 10.1213/ANE.0000000000000382
- Jhun, E. H., Yao, Y., He, Y., Mack, A. K., Wilkie, D. J., Molokie, R. E., et al. (2015). Prevalence of pain-related single nucleotide polymorphisms in patients of African origin with sickle cell disease. *Pharmacogenomics* 16, 1795–1806. doi: 10.2217/pgs.15.126
- Jurečková, J., Babušková, E., Kmeťová, M., Kliment, J., and Dobrota, D. (2015). Estrogen receptor alpha polymorphisms and the risk of prostate cancer development. *J. Cancer Res. Clin. Oncol.* 141, 1963–1971. doi: 10.1007/s00432-015-1966-6
- Katusic, Z. S., Stelter, A., and Milstien, S. (1998). Cytokines stimulate GTP cyclohydrolase I gene expression in cultured human umbilical vein endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* 18, 27–32. doi: 10.1161/01.ATV.18.1.27
- Kaunisto, M. A., Jokela, R., Tallgren, M., Kambur, O., Tikkanen, E., Tasmuth, T., et al. (2013). Pain in 1000 women treated for breast cancer: a prospective study of pain sensitivity and postoperative pain. *Anesthesiology* 119, 1410–1421. doi: 10.1097/ALN.0000000000000012
- Khalil, H., Sereika, S. M., Dai, F., Alexander, S., Conley, Y., Gruen, G., et al. (2017). OPRM1 and COMT gene-gene interaction is associated with postoperative pain and opioid consumption after orthopedic trauma. *Biol. Res. Nurs.* 19, 170–179. doi: 10.1177/1099800416680474
- Kim, H., Lee, H., Rowan, J., Brahim, J., and Dionne, R. A. (2006a). Genetic polymorphisms in monoamine neurotransmitter systems show only weak association with acute post-surgical pain in humans. *Mol. Pain* 2:24. doi: 10.1186/1744-8069-2-24
- Kim, H., Mittal, D. P., Iadarola, M. J., and Dionne, R. A. (2006b). Genetic predictors for acute experimental cold and heat pain sensitivity in humans. *J. Med. Genet.* 43:e40. doi: 10.1136/jmg.2005.036079
- Kim, H., Neubert, J. K., San Miguel, A., Xu, K., Krishnaraju, R. K., Iadarola, M. J., et al. (2004). Genetic influence on variability in human acute experimental pain sensitivity associated with gender, ethnicity and psychological temperament. *Pain* 109, 488–496. doi: 10.1016/j.pain.2004.02.027
- Kim, H. Y. (2015). Phospholipids: a neuroinflammation emerging target. *Nat. Chem. Biol.* 11, 99–100. doi: 10.1038/nchembio.1740
- Kim, J. I., Kim, J.-W., Park, J.-E., Park, S., Hong, S.-B., Han, D. H., et al. (2017). Association of the GRIN2B rs2284411 polymorphism with methylphenidate response in attention-deficit/hyperactivity disorder. *J. Psychopharmacol.* 31, 1070–1077. doi: 10.1177/0269881116667707
- Klein, C. J., Wu, Y., Kilfoyle, D. H., Sandroni, P., Davis, M. D., Gavrilova, R. H., et al. (2013). Infrequent SCN9A mutations in congenital insensitivity to pain and erythromelalgia. *J. Neurol. Neurosurg. Psychiatry* 84, 386–391. doi: 10.1136/jnnp-2012-303719
- Klepstad, P., Rakvag, T. T., Kaasa, S., Holthe, M., Dale, O., Borchgrevink, P. C., et al. (2004). The 118 A > G polymorphism in the human micro-opioid receptor gene may increase morphine requirements in patients with pain caused by malignant disease. *Acta Anaesthesiol. Scand.* 48, 1232–1239. doi: 10.1111/j.1399-6576.2004.00517.x
- Ko, S. W., Vadakkan, K. I., Ao, H., Gallitano-Mendel, A., Wei, F., Milbrandt, J., et al. (2005). Selective contribution of Egr1 (zif/268) to persistent inflammatory pain. *J. Pain* 6, 12–20. doi: 10.1016/j.jpain.2004.10.001
- Kooloos, W. M., Wessels, J. A., Van Der Straaten, T., Allaart, C. F., Huizinga, T. W., and Guchelaar, H. J. (2010). Functional polymorphisms and methotrexate treatment outcome in recent-onset rheumatoid arthritis. *Pharmacogenomics* 11, 163–175. doi: 10.2217/pgs.09.139
- Kringel, D., and Lötsch, J. (2015). Pain research funding by the european union seventh framework programme. *Eur. J. Pain* 19, 595–600. doi: 10.1002/ejp.690
- Kringel, D., Sisignano, M., Zinn, S., and Lötsch, J. (2017). Next-generation sequencing of the human TRPV1 gene and the regulating co-players LTB4R and LTB4R2 based on a custom AmpliSeq panel. *PLoS One* 12:e0180116. doi: 10.1371/journal.pone.0180116
- Kumar, D., Chakraborty, J., and Das, S. (2012). Epistatic effects between variants of kappa-opioid receptor gene and A118G of mu-opioid receptor gene increase susceptibility to addiction in Indian population. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 36, 225–230. doi: 10.1016/j.pnpbp.2011.10.018
- Kumar, V., and Mahal, B. A. (2012). NGF – The TrkA to successful pain treatment. *J. Pain Res.* 5, 279–287. doi: 10.2147/JPR.S33408
- Kurita, G. P., Ekholm, O., Kaasa, S., Klepstad, P., Skorpen, F., and Sjøgren, P. (2016). Genetic variation and cognitive dysfunction in opioid-treated patients with cancer. *Brain Behav.* 6:e00471. doi: 10.1002/brb3.471
- Kwon, H. J., Jang, W.-C., and Lim, M. H. (2016). Association between RELN gene polymorphisms and attention deficit hyperactivity disorder in Korean children. *Psychiatry Investig.* 13, 210–216. doi: 10.4306/pi.2016.13.2.210
- Lacroix-Fralish, M. L., Ledoux, J. B., and Mogil, J. S. (2007). The pain genes database: an interactive web browser of pain-related transgenic knockout studies. *Pain* 131, e1–e4. doi: 10.1016/j.pain.2007.04.041
- Lander, E. S., Linton, L. M., Birren, B., Nussbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Law, P. Y., Reggio, P. H., and Loh, H. H. (2013). Opioid receptors: toward separation of analgesic from undesirable effects. *Trends Biochem. Sci.* 38, 275–282. doi: 10.1016/j.tibs.2013.03.003
- Ledergerber, C., and Dessimoz, C. (2011). Base-calling for next-generation sequencing platforms. *Brief. Bioinform.* 12, 489–497. doi: 10.1093/bib/bbq077
- Lee, M., and Tracey, I. (2013). Neuro-genetics of persistent pain. *Curr. Opin. Neurobiol.* 23, 127–132. doi: 10.1016/j.conb.2012.11.007
- Leung, L., and Cahill, C. M. (2010). TNF-alpha and neuropathic pain—a review. *J. Neuroinflamm.* 7:27. doi: 10.1186/1742-2094-7-27
- Li, J. Y., Tao, F., Wu, X. X., Tan, Y. Z., He, L., and Lu, H. (2015). Polymorphic variations in manganese superoxide dismutase (MnSOD) and endothelial nitric oxide synthase (eNOS) genes contribute to the development of type 2 diabetes mellitus in the Chinese Han population. *Genet. Mol. Res.* 14, 12993–13002. doi: 10.4238/2015.October.21.20
- Li, T., Zhao, J., Yang, J., Ma, X., Dai, Q., Huang, H., et al. (2016). A Meta-Analysis of the association between ESR1 genetic variants and the risk of breast cancer. *PLoS One* 11:e0153314. doi: 10.1371/journal.pone.0153314
- Lin, C.-H., Chaudhuri, K. R., Fan, J.-Y., Ko, C.-I., Rizos, A., Chang, C.-W., et al. (2017). Depression and Catechol-O-methyltransferase (COMT) genetic variants are associated with pain in Parkinson's disease. *Sci. Rep.* 7:6306. doi: 10.1038/s41598-017-06782-z
- Lindstedt, F., Karshikoff, B., Schalling, M., Olgart Hoglund, C., Ingvar, M., Lekander, M., et al. (2012). Serotonin-1A receptor polymorphism (rs6295) associated with thermal pain perception. *PLoS One* 7:e43221. doi: 10.1371/journal.pone.0043221

- Lippmann, C., Kringel, D., Ultsch, A., and Lötsch, J. (2018). Computational functional genomics-based approaches in analgesic drug discovery and repurposing. *Pharmacogenomics* 19, 783–797. doi: 10.2217/pgs-2018-0036
- Liu, J., Shi, Y., Tang, J., Guo, T., Li, X., Yang, Y., et al. (2005). SNPs and haplotypes in the S100B gene reveal association with schizophrenia. *Biochem. Biophys. Res. Commun.* 328, 335–341. doi: 10.1016/j.bbrc.2004.12.175
- Liu, Y. N., Yang, X., Suo, Z. W., Xu, Y. M., and Hu, X. D. (2014). Fyn kinase-regulated NMDA receptor- and AMPA receptor-dependent pain sensitization in spinal dorsal horn of mice. *Eur. J. Pain* 18, 1120–1128. doi: 10.1002/j.1532-2149.2014.00455.x
- Loggia, M. L., Chonde, D. B., Akeju, O., Arabasz, G., Catana, C., Edwards, R. R., et al. (2015). Evidence for brain glial activation in chronic pain patients. *Brain* 138, 604–615. doi: 10.1093/brain/awu377
- Lohmann, K., and Klein, C. (2014). Next generation sequencing and the future of genetic diagnosis. *Neurotherapeutics* 11, 699–707. doi: 10.1007/s13311-014-0288-8
- Loncar, Z., Curic, G., Mestrovic, A. H., Mickovic, V., and Bilic, M. (2013). Do IL-1B and IL-1RN modulate chronic low back pain in patients with post-traumatic stress disorder? *Collegium Antropol.* 37, 1237–1244.
- Lötsch, J., Doehring, A., Mogil, J. S., Arndt, T., Geisslinger, G., and Ultsch, A. (2013). Functional genomics of pain in analgesic drug development and therapy. *Pharmacol. Ther.* 139, 60–70. doi: 10.1016/j.pharmthera.2013.04.004
- Lötsch, J., Fluhr, K., Neddermayer, T., Doehring, A., and Geisslinger, G. (2009a). The consequence of concomitantly present functional genetic variants for the identification of functional genotype-phenotype associations in pain. *Clin. Pharmacol. Ther.* 85, 25–30. doi: 10.1038/clpt.2008.103
- Lötsch, J., Geisslinger, G., and Tegeder, I. (2009b). Genetic modulation of the pharmacological treatment of pain. *Pharmacol. Ther.* 124, 168–184. doi: 10.1016/j.pharmthera.2009.06.010
- Lötsch, J., Von Hentig, N., Freynhagen, R., Griessinger, N., Zimmermann, M., Doehring, A., et al. (2009c). Cross-sectional analysis of the influence of currently known pharmacogenetic modulators on opioid therapy in outpatient pain centers. *Pharmacogenet. Genom.* 19, 429–436. doi: 10.1097/FPC.0b013e32832b89da
- Lötsch, J., and Geisslinger, G. (2005). Are mu-opioid receptor polymorphisms important for clinical opioid therapy? *Trends Mol. Med.* 11, 82–89. doi: 10.1016/j.molmed.2004.12.006
- Lötsch, J., and Geisslinger, G. (2010). A critical appraisal of human genotyping for pain therapy. *Trends Pharmacol. Sci.* 31, 312–317. doi: 10.1016/j.tips.2010.04.002
- Lötsch, J., Klepstad, P., Doehring, A., and Dale, O. (2010). A GTP cyclohydrolase 1 genetic variant delays cancer pain. *Pain* 148, 103–106. doi: 10.1016/j.pain.2009.10.021
- Lötsch, J., Sipilä, R., Tasmuth, T., Kringel, D., Estlander, A. M., Meretoja, T., et al. (2018). Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy. *Breast Cancer Res./Treatment*. [Epub ahead of print]. doi: 10.1007/s10549-018-4841-8
- Lötsch, J., Skarke, C., Grosch, S., Darimont, J., Schmidt, H., and Geisslinger, G. (2002). The polymorphism A118G of the human mu-opioid receptor gene decreases the pupil constrictory effect of morphine-6-glucuronide but not that of morphine. *Pharmacogenetics* 12, 3–9. doi: 10.1097/00008571-200201000-00002
- Lötsch, J., Stuck, B., and Hummel, T. (2006). The human mu-opioid receptor gene polymorphism 118A > G decreases cortical activation in response to specific nociceptive stimulation. *Behav. Neurosci.* 120, 1218–1224. doi: 10.1037/0735-7044.120.6.1218
- Lu, H., Zhu, L., Lian, L., Chen, M., Shi, D., and Wang, K. (2015). Genetic variations in the PRKCG gene and osteosarcoma risk in a Chinese population: a case-control study. *Tumour Biol.* 36, 5241–5247. doi: 10.1007/s13277-015-3182-z
- Maertens, O., De Schepper, S., Vandesompele, J., Brems, H., Heyns, I., Janssens, S., et al. (2007). Molecular dissection of isolated disease features in mosaic neurofibromatosis type 1. *Am. J. Hum. Genet.* 81, 243–251. doi: 10.1086/519562
- Mansour, A. R., Farmer, M. A., Baliki, M. N., and Apkarian, A. V. (2014). Chronic pain: the role of learning and brain plasticity. *Restor. Neurol. Neurosci.* 32, 129–139.
- Marcos, M., Pastor, I., De La Calle, C., Barrio-Real, L., Laso, F.-J., and González-Sarmiento, R. (2012). Cannabinoid receptor 1 gene is associated with alcohol dependence. *Alcohol. Clin. Exp. Res.* 36, 267–271. doi: 10.1111/j.1530-0277.2011.01623.x
- Martikainen, I. K., Nuechterlein, E. B., Pecina, M., Love, T. M., Cummmiford, C. M., Green, C. R., et al. (2015). Chronic back pain is associated with alterations in dopamine neurotransmission in the ventral striatum. *J. Neurosci.* 35, 9957–9965. doi: 10.1523/JNEUROSCI.4605-14.2015
- Maxwell, J. R., Potter, C., Hyrich, K. L., Barton, A., Worthington, J., Isaacs, J. D., et al. (2008). Association of the tumour necrosis factor-308 variant with differential response to anti-TNF agents in the treatment of rheumatoid arthritis. *Hum. Mol. Genet.* 17, 3532–3538. doi: 10.1093/hmg/ddn245
- McIntosh, A. M., Simen, A. A., Evans, K. L., Hall, J., Macintyre, D. J., Blackwood, D., et al. (2012). Genetic variation in hyperpolarization-activated cyclic nucleotide-gated channels and its relationship with neuroticism, cognition and risk of depression. *Front. Genet.* 3:116. doi: 10.3389/fgene.2012.00116
- Mead, S., Poulter, M., Uphill, J., Beck, J., Whitfield, J., Webb, T. E. F., et al. (2009). Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *Lancet Neurol.* 8, 57–66. doi: 10.1016/S1474-4422(08)70265-5
- Metzker, M. L. (2010). Sequencing technologies – The next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626
- Mill, J., Kiss, E., Baji, I., Kapornai, K., Daróczy, G., Vetró, A., et al. (2008). Association study of the estrogen receptor alpha gene (ESR1) and childhood-onset mood disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 1323–1326. doi: 10.1002/ajmg.b.30751
- Milne, R. L., Burwinkel, B., Michailidou, K., Arias-Perez, J.-I., Zamora, M. P., Menéndez-Rodríguez, P., et al. (2014). Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer association consortium. *Hum. Mol. Genet.* 23, 6096–6111. doi: 10.1093/hmg/ddu311
- Mitra, A. K., Crews, K., Pounds, S., Cao, X., Downing, J. R., Raimondi, S., et al. (2011). Impact of genetic variation in FKBP5 on clinical response in pediatric acute myeloid leukemia patients: a pilot study. *Leukemia* 25, 1354–1356. doi: 10.1038/leu.2011.74
- Mocellin, S., Marincola, F., Rossi, C. R., Nitti, D., and Lise, M. (2004). The multifaceted relationship between IL-10 and adaptive immunity: putting together the pieces of a puzzle. *Cytokine Growth. Factor. Rev.* 15, 61–76. doi: 10.1016/j.cytogfr.2003.11.001
- Mogil, J. S. (2009). Are we getting anywhere in human pain genetics? *Pain* 146, 231–232. doi: 10.1016/j.pain.2009.07.023
- Mogil, J. S. (2012). Pain genetics: past, present and future. *Trends Genet.* 28, 258–266. doi: 10.1016/j.tig.2012.02.004
- Mustafa, A. E., Faquih, T., Baz, B., Kattan, R., Al-Issa, A., Tahir, A. I., et al. (2018). Validation of ion torrent(TM) inherited disease panel with the PGM(TM) sequencing platform for rapid and comprehensive mutation detection. *Genes (Basel.)* 9, 267. doi: 10.3390/genes9050267
- Nackley, A. G., Shabalina, S. A., Tchivileva, I. E., Satterfield, K., Korchynskyi, O., Makarov, S. S., et al. (2006). Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 314, 1930–1933. doi: 10.1126/science.1131262
- Naumov, D. E., Perelman, J. M., Kolosov, V. P., Potapova, T. A., Maksimov, V. N., and Zhou, X. (2015). Transient receptor potential melastatin 8 gene polymorphism is associated with cold-induced airway hyperresponsiveness in bronchial asthma. *Respirology* 20, 1192–1197. doi: 10.1111/resp.12605
- Nemeth, E., Rivera, S., Gabayan, V., Keller, C., Taudorf, S., Pedersen, B. K., et al. (2004). IL-6 mediates hypoferrremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J. Clin. Invest.* 113, 1271–1276. doi: 10.1172/JCI200420945
- Nicholl, B. I., Holliday, K. L., Macfarlane, G. J., Thomson, W., Davies, K. A., O'Neill, T. W., et al. (2011). Association of HTR2A polymorphisms with chronic widespread pain and the extent of musculoskeletal pain: results from two population-based cohorts. *Arthritis Rheum.* 63, 810–818. doi: 10.1002/art.30185
- Nissenbaum, J., Devor, M., Seltzer, Z., Gebauer, M., Michaelis, M., Tal, M., et al. (2010). Susceptibility to chronic pain following nerve injury is genetically affected by CACNG2. *Genome Res.* 20, 1180–1190. doi: 10.1101/gr.104976.110

- Nöthen, M. M., Cichon, S., Hemmer, S., Hebebrand, J., Remschmidt, H., Lehmkuhl, G., et al. (1994). Human dopamine D4 receptor gene: frequent occurrence of a null allele and observation of homozygosity. *Hum. Mol. Genet.* 3, 2207–2212. doi: 10.1093/hmg/3.12.2207
- Obata, K., and Noguchi, K. (2006). BDNF in sensory neurons and chronic pain. *Neurosci. Res.* 55, 1–10. doi: 10.1016/j.neures.2006.01.005
- Oertel, B. G., Doebering, A., Roskam, B., Kettner, M., Hackmann, N., Ferreiros, N., et al. (2012). Genetic-epigenetic interaction modulates mu-opioid receptor regulation. *Hum. Mol. Genet.* 21, 4751–4760. doi: 10.1093/hmg/dds314
- Oertel, B. G., Schmidt, R., Schneider, A., Geisslinger, G., and Lötsch, J. (2006). The mu-opioid receptor gene polymorphism 118A>G depletes alfentanil-induced analgesia and protects against respiratory depression in homozygous carriers. *Pharmacogenet. Genom.* 16, 625–636. doi: 10.1097/01.fpc.0000220566.90466.a2
- Offenbaeher, M., Bondy, B., De Jonge, S., Glatzeder, K., Kruger, M., Schoeps, P., et al. (1999). Possible association of fibromyalgia with a polymorphism in the serotonin transporter gene regulatory region. *Arthritis Rheum.* 42, 2482–2488. doi: 10.1002/1529-0131(199911)42:11<2482::AID-ANR27>3.0.CO;2-B
- Onaya, T., Ishii, M., Katoh, H., Shimizu, S., Kasai, H., Kawamura, M., et al. (2013). Predictive index for the onset of medication overuse headache in migraine patients. *Neurol. Sci.* 34, 85–92. doi: 10.1007/s10072-012-0955-7
- Onojighofia, T., Meshkin, B., Nguyen, S. V., Schwartz, D., and Akindele, B. (2014). Perception of analgesia in narcotic users with chronic pain: a multi-center cross-sectional study comparing genotype to pain VAS (P.A.I.N. Study). *Neurology* 82:E39.
- Oyanguren-Desez, O., Rodríguez-Antigüedad, A., Villoslada, P., Domercq, M., Alberdi, E., and Matute, C. (2011). Gain-of-function of P2X7 receptor gene variants in multiple sclerosis. *Cell Calcium* 50, 468–472. doi: 10.1016/j.ceca.2011.08.002
- Palmirotta, R., Barbanti, P., De Marchis, M. L., Egeo, G., Aurilia, C., Fofi, L., et al. (2015). Is SOD2 Ala16Val polymorphism associated with migraine with aura phenotype? *Antioxid. Redox Signal.* 22, 275–279. doi: 10.1089/ars.2014.6069
- Park, J. J., Lee, J., Kim, M. A., Back, S. K., Hong, S. K., and Na, H. S. (2007). Induction of total insensitivity to capsaicin and hypersensitivity to garlic extract in human by decreased expression of TRPV1. *Neurosci. Lett.* 411, 87–91. doi: 10.1016/j.neulet.2006.10.046
- Peiro, A. M., Planelles, B., Juhasz, G., Bagdy, G., Libert, F., Eschaliere, A., et al. (2016). Pharmacogenomics in pain treatment. *Drug Metab. Pers. Ther.* 31, 131–142. doi: 10.1515/dmpt-2016-0005
- Pelletier, C., and Weidhaas, J. B. (2010). MicroRNA binding site polymorphisms as biomarkers of cancer risk. *Expert. Rev. Mol. Diagn.* 10, 817–829. doi: 10.1586/erm.10.59
- Petrenko, A. B., Yamakura, T., Baba, H., and Shimoji, K. (2003). The role of N-methyl-D-aspartate (NMDA) receptors in pain: a review. *Anesth. Analg.* 97, 1108–1116. doi: 10.1213/01.ANE.0000081061.12235.55
- Potapova, T. A., Babenko, V. N., Kobzev, V. F., Romashchenko, A. G., Maksimov, V. N., and Voevoda, M. I. (2014). Associations of cold receptor TRPM8 gene single nucleotide polymorphism with blood lipids and anthropometric parameters in Russian population. *Bull. Exp. Biol. Med.* 157, 757–761. doi: 10.1007/s10517-014-2660-4
- Potvin, S., Larouche, A., Normand, E., De Souza, J. B., Gaumond, I., Grignon, S., et al. (2009). DRD3 Ser9Gly polymorphism is related to thermal pain perception and modulation in chronic widespread pain patients and healthy controls. *J. Pain* 10, 969–975. doi: 10.1016/j.jpain.2009.03.013
- Prasad, H. C., Steiner, J. A., Sutcliffe, J. S., and Blakely, R. D. (2009). Enhanced activity of human serotonin transporter variants associated with autism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 163–173. doi: 10.1098/rstb.2008.0143
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Development Core Team.
- Rakvag, T. T., Klepstad, P., Baar, C., Kvam, T. M., Dale, O., Kaasa, S., et al. (2005). The Val158Met polymorphism of the human catechol-O-methyltransferase (COMT) gene may influence morphine requirements in cancer pain patients. *Pain* 116, 73–78. doi: 10.1016/j.pain.2005.03.032
- Reimann, F., Cox, J. J., Belfer, I., Diatchenko, L., Zaykin, D. V., Mchale, D. P., et al. (2010). Pain perception is altered by a nucleotide polymorphism in SCN9A. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5148–5153. doi: 10.1073/pnas.09131811107
- Reyes-Gibby, C. C., Shete, S., Rakvag, T., Bhat, S. V., Skorpen, F., Bruera, E., et al. (2007). Exploring joint effects of genes and the clinical efficacy of morphine for cancer pain: OPRM1 and COMT gene. *Pain* 130, 25–30. doi: 10.1016/j.pain.2006.10.023
- Ribeiro-Dasilva, M. C., Peres Line, S. R., Leme Godoy, Dos Santos, M. C., Arthuri, M. T., Hou, W., et al. (2009). Estrogen receptor-alpha polymorphisms and predisposition to TMJ disorder. *J. Pain* 10, 527–533. doi: 10.1016/j.jpain.2008.11.012
- Rodrigues, P., De Marco, G., Furriol, J., Mansego, M. L., Pineda-Alonso, M., Gonzalez-Neira, A., et al. (2014). Oxidative stress in susceptibility to breast cancer: study in Spanish population. *BMC Cancer* 14:861. doi: 10.1186/1471-2407-14-861
- Root, T. L., Szatkiewicz, J. P., Jonassaint, C. R., Thornton, L. M., Pinheiro, A. P., Strober, M., et al. (2011). Association of candidate genes with phenotypic traits relevant to anorexia nervosa. *Eur. Eat Disord. Rev.* 19, 487–493. doi: 10.1002/erv.1138
- Sah, D. W., Ossipov, M. H., Rossomando, A., Silvian, L., and Porreca, F. (2005). New approaches for the treatment of pain: the GDNF family of neurotrophic growth factors. *Curr. Top. Med. Chem.* 5, 577–583. doi: 10.2174/1568026054367593
- Sanger, F., and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94, 441–448. doi: 10.1016/0022-2836(75)90213-2
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463
- Sato, M., Ohashi, J., Tsuchiya, N., Kashiwase, K., Ishikawa, Y., Arita, H., et al. (2002). Association of HLA-A*3303-B*4403-DRB1*1302 haplotype, but not of TNFA promoter and NKp30 polymorphism, with postherpetic neuralgia (PHN) in the Japanese population. *Genes Immun.* 3, 477–481. doi: 10.1038/sj.gene.6363890
- Savas, S., Schmidt, S., Jarjanazi, H., and Ozcelik, H. (2006). Functional nsSNPs from carcinogenesis-related genes expressed in breast tissue: potential breast cancer risk alleles and their distribution across human populations. *Hum. Genomics* 2, 287–296. doi: 10.1186/1479-7364-2-5-287
- Schleinitz, D., Carmienke, S., Böttcher, Y., Tönjes, A., Berndt, J., Klötting, N., et al. (2010). Role of genetic variation in the cannabinoid type 1 receptor gene (CNRI) in the pathophysiology of human obesity. *Pharmacogenomics* 11, 693–702. doi: 10.2217/pgs.10.42
- Schmidt, S., Hauser, M. A., Scott, W. K., Postel, E. A., Agarwal, A., Gallins, P., et al. (2006). Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am. J. Hum. Genet.* 78, 852–864. doi: 10.1086/503822
- Schott, K., Gutlich, M., and Ziegler, I. (1993). Induction of GTP-cyclohydrolase I mRNA expression by lectin activation and interferon-gamma treatment in human cells associated with the immune response. *J. Cell. Physiol.* 156, 12–16. doi: 10.1002/jcp.1041560103
- Schurks, M., Kurth, T., Buring, J. E., and Zee, R. Y. (2009). A candidate gene association study of 77 polymorphisms in migraine. *J. Pain* 10, 759–766. doi: 10.1016/j.jpain.2009.01.326
- Schürks, M., Rist, P. M., and Kurth, T. (2010). Sex hormone receptor gene polymorphisms and migraine: a systematic review and meta-analysis. *Cephalalgia* 30, 1306–1328. doi: 10.1177/0333102410364155
- Schütz, M., Oertel, B. G., Heimann, D., Doebering, A., Walter, C., Dimova, V., et al. (2014). Consequences of a human TRPA1 genetic variant on the perception of nociceptive and olfactory stimuli. *PLoS One* 9:e95592. doi: 10.1371/journal.pone.0095592
- Schwartz, E. S., Kim, H. Y., Wang, J., Lee, I., Klann, E., Chung, J. M., et al. (2009). Persistent pain is dependent on spinal mitochondrial antioxidant levels. *J. Neurosci.* 29, 159–168. doi: 10.1523/JNEUROSCI.3792-08.2009
- Schweinhardt, P., Sauro, K. M., and Bushnell, M. C. (2008). Fibromyalgia: a disorder of the brain? *Neuroscientist* 14, 415–421. doi: 10.1177/1073858407312521
- Serý, O., Příkrýl, R., Castulík, L., and Státný, F. (2010). A118G polymorphism of OPRM1 gene is associated with schizophrenia. *J. Mol. Neurosci.* 41, 219–222. doi: 10.1007/s12031-010-9327-z
- Shah, N. D., Shah, P. S., Panchal, Y. Y., Katudia, K. H., Khatri, N. B., Ray, H. S. P., et al. (2018). Mutation analysis of BRCA1/2 mutations with special reference to polymorphic SNPs in Indian breast cancer patients. *Appl. Clin. Genet.* 11, 59–67. doi: 10.2147/TACG.S155955

- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi: 10.1093/nar/29.1.308
- Shi, J., Badner, J. A., Hattori, E., Potash, J. B., Willour, V. L., McMahon, F. J., et al. (2008). Neurotransmission and bipolar disorder: a systematic family-based association study. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 1270–1277. doi: 10.1002/ajmg.b.30769
- Shi, T.-J. S., Liu, S.-X. L., Hammarberg, H., Watanabe, M., Xu, Z.-Q. D., and Hökfelt, T. (2008). Phospholipase C[β 3] in mouse and human dorsal root ganglia and spinal cord is a possible target for treatment of neuropathic pain. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20004–20008. doi: 10.1073/pnas.0810899105
- Shoskes, D. A., Albakri, Q., Thomas, K., and Cook, D. (2002). Cytokine polymorphisms in men with chronic prostatitis/chronic pelvic pain syndrome: association with diagnosis and treatment response. *J. Urol.* 168, 331–335. doi: 10.1016/S0022-5347(05)64916-6
- Siqueira, E. C. D., De Sousa, S. F., França, J. A., Diniz, M. G., Pereira, T. D. S. F., Moreira, R. G., et al. (2017). Targeted next-generation sequencing of glandular odontogenic cyst: a preliminary study. *Oral. Surg. Oral Med. Oral Pathol. Oral Radiol.* 124, 490–494. doi: 10.1016/j.oooo.2017.07.001
- Skarke, C., Kirchhof, A., Geisslinger, G., and Lötsch, J. (2004). Comprehensive mu-opioid-receptor genotyping by pyrosequencing. *Clin. Chem.* 50, 640–644. doi: 10.1373/clinchem.2003.027607
- Skarke, C., Kirchhof, A., Geisslinger, G., and Lötsch, J. (2005). Rapid genotyping for relevant CYP1A2 alleles by pyrosequencing. *Eur. J. Clin. Pharmacol.* 61, 887–892. doi: 10.1007/s00228-005-0029-3
- Sluka, K. A., and Audette, K. M. (2006). Activation of protein kinase C in the spinal cord produces mechanical hyperalgesia by activating glutamate receptors, but does not mediate chronic muscle-induced hyperalgesia. *Mol. Pain* 2:13. doi: 10.1186/1744-8069-2-13
- Smith, F. L., Fujimori, K., Lowe, J., and Welch, S. P. (1998). Characterization of delta-9-tetrahydrocannabinol and anandamide antinociception in nonarthritic and arthritic rats. *Pharmacol. Biochem. Behav.* 60, 183–191. doi: 10.1016/S0091-3057(97)00583-2
- Snider, W. D., and McMahon, S. B. (1998). Tackling pain at the source: new ideas about nociceptors. *Neuron* 20, 629–632. doi: 10.1016/S0896-6273(00)81003-X
- Solus, J. F., Chung, C. P., Oeser, A., Li, C., Rho, Y. H., Bradley, K. M., et al. (2015). Genetics of serum concentration of IL-6 and TNF α in systemic lupus erythematosus and rheumatoid arthritis: a candidate gene analysis. *Clin. Rheumatol.* 34, 1375–1382. doi: 10.1007/s10067-015-2881-6
- Somogyi, A. A., Collier, J. K., and Barratt, D. T. (2015). Pharmacogenetics of opioid response. *Clin. Pharmacol. Ther.* 97, 125–127. doi: 10.1002/cpt.23
- Somogyi, A. A., Sia, A. T., Tan, E.-C., Collier, J. K., Hutchinson, M. R., and Barratt, D. T. (2016). Ethnicity-dependent influence of innate immune genetic markers on morphine PCA requirements and adverse effects in postoperative pain. *Pain* 157, 2458–2466. doi: 10.1097/j.pain.0000000000000661
- Sonoda, T., Takada, J., Iba, K., Asakura, S., Yamashita, T., and Mori, M. (2012). Interaction between ESR α polymorphisms and environmental factors in osteoporosis. *J. Orthop. Res.* 30, 1529–1534. doi: 10.1002/jor.22083
- Sorge, R. E., Trang, T., Dorfman, R., Smith, S. B., Beggs, S., Ritchie, J., et al. (2012). Genetically determined P2X7 receptor pore formation regulates variability in chronic pain sensitivity. *Nat. Med.* 18, 595–599. doi: 10.1038/nm.2710
- Stephens, K., Cooper, B. A., West, C., Paul, S. M., Baggott, C. R., Merriman, J. D., et al. (2014). Associations between cytokine gene variations and severe persistent breast pain in women following breast cancer surgery. *J. Pain* 15, 169–180. doi: 10.1016/j.jpain.2013.09.015
- Sugaya, K., Nishijima, S., Yamada, T., Miyazato, M., Hatano, T., and Ogawa, Y. (2002). Molecular analysis of adrenergic receptor genes and interleukin-4/interleukin-4 receptor genes in patients with interstitial cystitis. *J. Urol.* 168, 2668–2671. doi: 10.1016/S0022-5347(05)64241-3
- Sun, Y.-G., Gracias, N. G., Drobish, J. K., Vasko, M. R., Gereau, R. W., and Chen, Z.-F. (2009). The c-kit signaling pathway is involved in the development of persistent pain. *Pain* 144, 178–186. doi: 10.1016/j.pain.2009.04.011
- Tegeder, I., Costigan, M., Griffin, R. S., Abele, A., Belfer, I., Schmidt, H., et al. (2006). GTP cyclohydrolase and tetrahydrobiopterin regulate pain sensitivity and persistence. *Nat. Med.* 12, 1269–1277. doi: 10.1038/nm1490
- Thuaud, S. (2016). A peripheral messenger for chronic pain. *Nat. Neurosci.* 19:9. doi: 10.1038/nn.4217
- Thulasiraman, K., and Swamy, M. N. S. (1992). *Graphs : Theory and Algorithms*. New York, NY: Wiley. doi: 10.1002/9781118033104
- Tian, W., Fu, Y., Garcia-Elias, A., Fernández-Fernández, J. M., Vicente, R., Kramer, P. L., et al. (2009). A loss-of-function nonsynonymous polymorphism in the osmoregulatory TRPV4 gene is associated with human hyponatremia. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14034–14039. doi: 10.1073/pnas.0904084106
- Tsai, S.-J., Hong, C.-J., Cheng, C.-Y., Liao, D.-L., and Liou, Y.-J. (2007). Association study of polymorphisms in post-synaptic density protein 95 (PSD-95) with schizophrenia. *J. Neural Transm. (Vienna)* 114, 423–426. doi: 10.1007/s00702-006-0587-2
- Tsantoulas, C., Mooney, E. R., and McNaughton, P. A. (2016). HCN2 ion channels: basic science opens up possibilities for therapeutic intervention in neuropathic pain. *Biochem. J.* 473, 2717–2736. doi: 10.1042/BCJ20160287
- Tulah, A. S., Beghe, B., Barton, S. J., Holloway, J. W., and Sayers, I. (2012). Leukotriene B4 receptor locus gene characterisation and association studies in asthma. *BMC Med. Genet.* 13:110. doi: 10.1186/1471-2350-13-110
- Ultsch, A., Kringel, D., Kalso, E., Mogil, J. S., and Lötsch, J. (2016). A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. *Pain* 157, 2747–2757. doi: 10.1097/j.pain.0000000000000694
- Ursu, D., Ebert, P., Langron, E., Ruble, C., Munsie, L., Zou, W., et al. (2014). Gain and loss of function of P2X7 receptors: mechanisms, pharmacology and relevance to diabetic neuropathic pain. *Mol. Pain* 10:37. doi: 10.1186/1744-8069-10-37
- Vaclavikova, E., Dvorakova, S., Skaba, R., Pos, L., Sykorova, V., Halkova, T., et al. (2014). RET variants and haplotype analysis in a cohort of Czech patients with Hirschsprung disease. *PLoS One* 9:e98957. doi: 10.1371/journal.pone.0098957
- Vadakkan, K. I., Wang, H., Ko, S. W., Zastepa, E., Petrovic, M. J., Sluka, K. A., et al. (2006). Genetic reduction of chronic muscle pain in mice lacking calcium/calmodulin-stimulated adenylyl cyclases. *Mol. Pain* 2:7. doi: 10.1186/1744-8069-2-7
- Vassalli, P. (1992). The pathophysiology of tumor necrosis factors. *Annu. Rev. Immunol.* 10, 411–452. doi: 10.1146/annurev.iy.10.040192.002211
- Venn, J. (1880). On the diagrammatic and mechanical representation of propositions and reasonings. *Dublin Philos. Mag. J. Sci.* 9, 1–18. doi: 10.1080/14786448008626877
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040
- Walker, K., Bowes, M., Panesar, M., Davis, A., Gentry, C., Kesingland, A., et al. (2001). Metabotropic glutamate receptor subtype 5 (mGlu5) and nociceptive function. I. Selective blockade of mGlu5 receptors in models of acute, persistent and chronic pain. *Neuropharmacology* 40, 1–9. doi: 10.1016/S0028-3908(00)00113-1
- Walter, C., Doebering, A., Oertel, B. G., and Lötsch, J. (2013). μ -opioid receptor gene variant OPRM1 118 A>G: a summary of its molecular and clinical consequences for pain. *Pharmacogenomics* 14, 1915–1925. doi: 10.2217/pgs.13.187
- Wang, C., Wang, H., Pang, J., Li, L., Zhang, S., Song, G., et al. (2014). Glial cell-derived neurotrophic factor attenuates neuropathic pain in a mouse model of chronic constriction injury: possible involvement of E-cadherin/p120ctn signaling. *J. Mol. Neurosci.* 54, 156–163. doi: 10.1007/s12031-014-0266-y
- Wang, X.-S., Song, H.-B., Chen, S., Zhang, W., Liu, J.-Q., Huang, C., et al. (2015). Association of single nucleotide polymorphisms of ABCB1, OPRM1 and COMT with pain perception in cancer patients. *J. Huazhong Univ. Sci. Technol. Med. Sci.* 35, 752–758. doi: 10.1007/s11596-015-1502-6
- Wedrén, S., Lovmar, L., Humphreys, K., Magnusson, C., Melhus, H., Syvänen, A.-C., et al. (2008). Estrogen receptor alpha gene polymorphism and endometrial cancer risk—a case-control study. *BMC Cancer* 8:322. doi: 10.1186/1471-2407-8-322
- Wellcome Trust, Case Control, and Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678. doi: 10.1038/nature05911
- Wise, B. L., Demissie, S., Cupples, L. A., Felson, D. T., Yang, M., Shearman, A. M., et al. (2009). The relationship of estrogen receptor-alpha and -beta genes with osteoarthritis of the hand. *J. Rheumatol.* 36, 2772–2779. doi: 10.3899/jrheum.081208

- Wolters, P. L., Burns, K. M., Martin, S., Baldwin, A., Dombi, E., Toledo-Tamula, M. A., et al. (2015). Pain interference in youth with neurofibromatosis type 1 and plexiform neurofibromas and relation to disease severity, social-emotional functioning, and quality of life. *Am. J. Med. Genet. A* 167A, 2103–2113. doi: 10.1002/ajmg.a.37123
- Woo, J. H., Kim, H., Kim, J. H., and Kim, J. G. (2015). Cannabinoid receptor gene polymorphisms and bone mineral density in Korean postmenopausal women. *Menopause* 22, 512–519. doi: 10.1097/GME.0000000000000339
- Yamamoto, M., Ito, Y., Mitsuma, N., Hattori, N., and Sobue, G. (2003). Pain-related differential expression of NGF, GDNF, IL-6, and their receptors in human vasculitic neuropathies. *Intern. Med.* 42, 1100–1103. doi: 10.2169/internalmedicine.42.1100
- Yang, L., Gu, X., Zhang, W., Zhang, J., and Ma, Z. (2014). Cdk5 inhibitor roscovitine alleviates neuropathic pain in the dorsal root ganglia by downregulating N-methyl-D-aspartate receptor subunit 2A. *Neurol. Sci.* 35, 1365–1371. doi: 10.1007/s10072-014-1713-9
- Yücel, Y., Coşkun, S., Cengiz, B., Özdemir, H. H., Uzar, E., Çim, A., et al. (2016). Association of polymorphisms within the serotonin receptor genes 5-HTR1A, 5-HTR1B, 5-HTR2A and 5-HTR2C and migraine susceptibility in a Turkish population. *Clin. Psychopharmacol. Neurosci.* 14, 250–255. doi: 10.9758/cpn.2016.14.3.250
- Yucesoy, B., Johnson, V. J., Lummus, Z. L., Kissling, G. E., Fluharty, K., Gautrin, D., et al. (2012). Genetic variants in antioxidant genes are associated with diisocyanate-induced asthma. *Toxicol. Sci.* 129, 166–173. doi: 10.1093/toxsci/kfs183
- Zanette, S. A., Dussan-Sarria, J. A., Souza, A., Deitos, A., Torres, I. L. S., and Caumo, W. (2014). Higher serum S100B and BDNF levels are correlated with a lower pressure-pain threshold in fibromyalgia. *Mol. Pain* 10:46. doi: 10.1186/1744-8069-10-46
- Zhang, L., Rao, F., Zhang, K., Khandrika, S., Das, M., Vaingankar, S. M., et al. (2007). Discovery of common human genetic variants of GTP cyclohydrolase 1 (GCH1) governing nitric oxide, autonomic activity, and cardiovascular risk. *J. Clin. Invest.* 117, 2658–2671. doi: 10.1172/JCI31093
- Zinn, S., Sisignano M., Kern K., Pierre S., Tunaru S., Jordan H., et al. (2017). The leukotriene B4 receptors BLT1 and BLT2 form an antagonistic sensitizing system in peripheral sensory neurons. *J. Biol. Chem.* 292, 6123–6134. doi: 10.1074/jbc.M116.769125
- Zorina-Lichtenwalter, K., Meloto, C. B., Khoury, S., and Diatchenko, L. (2016). Genetic predictors of human chronic pain conditions. *Neuroscience* 338, 36–62. doi: 10.1016/j.neuroscience.2016.04.041

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kringel, Kaunisto, Lippmann, Kalso and Lötsch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data

Yitian Zhou¹, Kohei Fujikura², Souren Mkrtchian¹ and Volker M. Lauschke^{1*}

¹ Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden,

² Department of Diagnostic Pathology, Kobe University Graduate School of Medicine, Kobe, Japan

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Theodora Katsila,
University of Patras, Greece
Greg Slodkowitz,
MRC Laboratory of Molecular Biology
(MRC), United Kingdom

*Correspondence:

Volker M. Lauschke
volker.lauschke@ki.se

Specialty section:

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 03 August 2018

Accepted: 20 November 2018

Published: 04 December 2018

Citation:

Zhou Y, Fujikura K, Mkrtchian S and
Lauschke VM (2018) Computational
Methods for the Pharmacogenetic
Interpretation of Next Generation
Sequencing Data.
Front. Pharmacol. 9:1437.
doi: 10.3389/fphar.2018.01437

Up to half of all patients do not respond to pharmacological treatment as intended. A substantial fraction of these inter-individual differences is due to heritable factors and a growing number of associations between genetic variations and drug response phenotypes have been identified. Importantly, the rapid progress in Next Generation Sequencing technologies in recent years unveiled the true complexity of the genetic landscape in pharmacogenes with tens of thousands of rare genetic variants. As each individual was found to harbor numerous such rare variants they are anticipated to be important contributors to the genetically encoded inter-individual variability in drug effects. The fundamental challenge however is their functional interpretation due to the sheer scale of the problem that renders systematic experimental characterization of these variants currently unfeasible. Here, we review concepts and important progress in the development of computational prediction methods that allow to evaluate the effect of amino acid sequence alterations in drug metabolizing enzymes and transporters. In addition, we discuss recent advances in the interpretation of functional effects of non-coding variants, such as variations in splice sites, regulatory regions and miRNA binding sites. We anticipate that these methodologies will provide a useful toolkit to facilitate the integration of the vast extent of rare genetic variability into drug response predictions in a precision medicine framework.

Keywords: precision medicine, personalized medicine, variant effect prediction, ADME, NGS, rare variant analysis, noncoding variation, pharmacogenomics

INTRODUCTION

Inter-individual differences in drug response are clinically important phenomena that result in reduced efficacy or adverse reactions in 25–50% of all patients and genetic factors have been estimated to account for around 20–30% of these (Spear et al., 2001; Sim et al., 2013). Fueled by technological advances in Next-Generation Sequencing (NGS) technologies, the application of comprehensive sequencing approaches is on the rise for various applications, including studies of biodiversity, population genetics and biomedical research (Levy and Myers, 2016). Furthermore, plummeting costs to <1,000 USD per human genome and increasing worldwide sequencing capacities that we estimate to exceed 100 petabases per year (10^{15} bases corresponding to the size of around 100,000 human genomes) open tremendous possibilities for NGS to revolutionize precision medicine.

Strikingly, these massive NGS data sets revealed that individuals harbored on average more than 3.7 million single nucleotide variants (SNVs) and more than 350,000 insertions and deletions across different populations, emphasizing the substantial variability of the human genome (The 1000 Genomes Project Consortium, 2012). Particularly genes involved in drug absorption, distribution, metabolism and excretion (ADME) proved to be highly diverse and genetically complex (Fujikura et al., 2015; Bush et al., 2016; Kozyra et al., 2017). Across 208 ADME genes more than 69,000 SNVs have been described, 98.5% of these being rare with minor allele frequencies (MAF) <1% (Ingelman-Sundberg et al., 2018). The overall pharmacogenetic variability was highly population specific, particularly for isolated populations, such as Ashkenazi Jews (Ahn and Park, 2017; Kozyra et al., 2017; Zhou and Lauschke, 2018). Given this enormous pharmacogenetic variability, one of the key frontiers of contemporary pharmacogenomics is the translation of these comprehensive genomic data into clinically actionable treatment recommendations (Lauschke and Ingelman-Sundberg, 2016a, 2018).

Heterologous expression in cell lines followed by quantitative determination of gene product functionality using appropriate end points is considered as the gold standard strategy to characterize the functional impact of pharmacogenetic variants. Furthermore, epidemiological association studies can provide additional indications about the consequences of genetic variants on drug metabolism related phenotypes *in vivo*. However, for the functional interpretation of rare variants these approaches suffer from multiple shortcomings:

- i) These methods are generally low throughput and are not compatible with the interrogation of tens of thousands of variants.
- ii) Experimental characterizations are time consuming, expensive and require specially trained technical staff, which renders them unsuitable for the rapid functional interpretation of the pharmacogenotype of an individual patient at the point of care.
- iii) Epidemiological analyses require a sufficient number of patients who carry the allele, which drastically limits their feasibility for rare genetic variant studies.

Thus, in the absence of viable experimental strategies, computational prediction methodologies are routinely used to predict the functional impact of genetic variants. Most of these algorithms focus on predicting the functional consequences of variants that result in amino acid substitutions. However, recently much progress has also been made regarding the interpretation of non-coding variants that affect splice sites, promoters, enhancers or miRNA binding sites (Figure 1).

Prediction algorithms are generally trained on pathogenic variant sets and most tools base their conclusions, at least in part, on the evolutionary conservation of the respective sequence. Importantly however, pharmacogenes are hallmarked by low evolutionary conservation and are generally not associated with human disease. These peculiarities result in specific problems for the interpretation of pharmacogenetic variants. Here, we provide an updated overview of computational approaches

for the functional interpretation of genetic variants, specifically focusing on their suitability for pharmacogenetic predictions. We describe the underlying statistical frameworks and discuss their different bases for decision-making. Furthermore, we highlight important progress particularly in the interpretation of non-coding genetic variability. We conclude that computational tools are essential for the functional interpretation of an individual's pharmacogenotype and that their further improvement constitutes one of the most important frontiers for the clinical implementation of NGS-based genotyping.

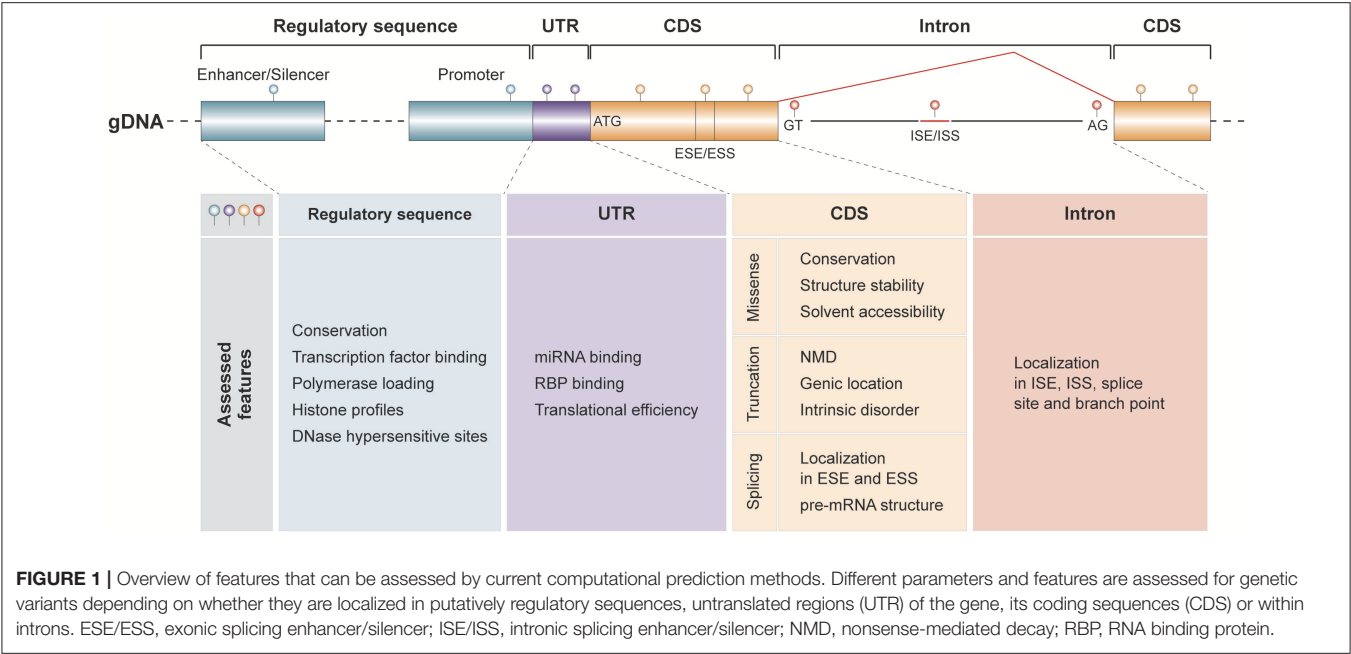
INTERPRETATION OF VARIANTS RESULTING IN AMINO ACID EXCHANGES

Genetic variants that result in amino acid substitution, henceforth termed missense variants, can impact the functionality of the respective protein by various mechanisms, including alterations in active sites, structural destabilization due to protein misfolding, perturbations in solvent accessibility or modification of post-translational processing. Each individual harbors 10,000–12,000 missense variants, many of which are rare (The 1000 Genomes Project Consortium, 2015). These rare variants have been suggested as important modulators of complex disease risk (Kryukov et al., 2007) and inter-individual differences in drug response (Kozyra et al., 2017). Among all variant classes, missense variants are the most extensively studied and a plethora of computational methods is available for their functional interpretation. Conceptually, these algorithms predict the functional impact of missense variants based on sequence information, primarily evolutionary conservation of the respective residues, and/or structural information of the corresponding gene product. In the following, we highlight recent progress, provide an overview of available tools and discuss their utility for pharmacogenetic predictions. For methodological details we refer the interested reader to excellent recent reviews (Ng and Henikoff, 2006; Peterson et al., 2013; Tang and Thomas, 2016).

Predictions Based on Sequence Information

Evolutionary conservation scores are calculated by analyzing the evolutionary variation dynamics of DNA or amino acid sequences among homologs with the hypothesis that the extent of conservation is a strong predictor of the importance of the respective sequence for structure and function of the corresponding gene product. Thus, positions with a high evolutionary rate are thought to be dispensable, whereas slowly evolving, i.e., conserved sequences indicate a selective pressure against variation in these regions and thus deleterious effects if mutated.

Evolutionary conservation as a metric to distinguish deleterious from neutral variants is considered by most computational prediction algorithms. The majority of approaches that focus on the functional interpretation of missense variants utilize amino acid sequence alignment, whereas others utilize nucleotide sequence alignments or a combination



of both methods (Table 1). While alignment of amino acid sequence proved to be effective for the analysis of missense variants, genomic sequence alignments provide additional versatility and allow to extend functional interpretations to variant classes that do not alter the amino acid sequence, such as synonymous and regulatory variants. Notably, commonly used conservation-based functionality predictors do not consider sequence interdependencies. Explicit integration of residue dependency information obtained from multiple sequence alignments was however recently shown to improve predictive performance (Hopf et al., 2017), emphasizing the added value of complementing conservation based functionality predictions with variant interaction data.

On the basis of multiple sequence alignments, algorithms derive their functionality predictions either based on direct theoretical models, or by various machine-learning approaches. The former methods predict the functional impact of variants based on phenomenological scores derived from theoretical models that are known *a priori*. In contrast, machine learning methods search for patterns in multi-dimensional training data sets consisting of labeled deleterious and benign variations, which will then be used as the basis to generate predictions on new unlabeled data. Machine learning approaches include support vector machines, random forests, artificial neural networks, naive Bayes approaches, gradient tree boosting and regression models. With increasing wealth of large-scale data sets to learn from, machine learning methods become increasingly popular as versatile tools to generate predictive models in many areas of biomedicine (Camacho et al., 2018).

Commonly used algorithms are generally designed to flag deleterious variants, which are mostly assumed to result in a reduced gene product function, and their performance of gain-of-function variants is substantially worse (Flanagan et al., 2010).

Notably, the algorithm B-SIFT, a modified version of the widely used SIFT tool (Ng and Henikoff, 2001), was developed to overcome this limitation (Lee et al., 2009). Conceptually, B-SIFT identifies increased functionality variants based on protein sequence alignments by scoring whether a given mutation results in a change commonly present in protein homologs and the tool successfully identified experimentally validated gain-of-function variants in cancer.

While computational missense variant predictors are generally reported to achieve high predictive accuracies with areas under the receiver operating characteristic curve (AUC_{ROC}) that often pivot around 0.9, drastic drops in performance to AUC_{ROC} of 0.5–0.75 have been reported on independent, functionally determined human variant datasets (Mahmood et al., 2017). These findings were corroborated by a recent cross-comparison of 23 methods based on three independent pathogenicity datasets in which the authors found that REVEL and VEST3 performed overall best, whereas the most commonly used methods SIFT and PolyPhen-2 performed only medially (Li et al., 2018). Furthermore, no functional consequences could be detected using various *in vitro* or *in vivo* tools for 40% of variants predicted to be deleterious by common functionality prediction tools (Miosge et al., 2015). Thus, while current tools have proven powerful in clinical diagnostics to prioritize potentially causative mutations in genetic diseases for further analyses (Boycott et al., 2013), their predictive power is not yet sufficient to predict functional variant effects without substantial subsequent validations.

Importantly, the quality of prediction models critically relies on accurate training data sets. For instance, models are commonly generated using training sets of pathogenic variants as positive controls and polymorphisms identified to be common in large-scale sequencing projects as negative, i.e., functionally

TABLE 1 | Methods to predict the functional effect of missense variants based on sequence information.

Algorithm	Model	Basis of decision	Model training or evaluation	References
SIFT	Direct	Prediction of functionality based on sequence conservation metrics that make use of Dirichlet priors	Variants from protein specific studies (LacI, HIV-1 Protease and Bacteriophage T4 Lysozyme)	Ng and Henikoff, 2001
PANTHER	HMM	Sequence conservation analysis using HMM	Variants from HGMD and dbSNP as deleterious and functionally neutral variants, respectively	Thomas et al., 2003
MAPP	Direct	Quantification of the physicochemical characteristics at each position of the amino acid sequence based on observed evolutionary variation	Protein specific studies (LacI, HIV-1 Protease, HIV reverse transcriptase and Bacteriophage T4 Lysozyme)	Stone and Sidow, 2005
PhastCons	HMM	Identification of conserved elements using a two-state phylogenetic HMM	Calibration on genomes from four model species (human, <i>D. melanogaster</i> , <i>C. elegans</i> , and <i>S. cerevisiae</i>)	Siepel et al., 2005
SNPs3D	SVM	Variant effect prediction based on amino acid sequence conservation metrics and folded state stability of protein structure	Variants from HGMD and dbSNP as deleterious and functionally neutral variants, respectively	Yue et al., 2006
PhD-SNP	SVM	Prediction of variant pathogenicity based on sequence profiles	Variants from HumVar and HumVarProf datasets	Capriotti et al., 2006
SiPhy	HMM	Sequence conservation analysis using HMM	ENCODE Phase I regions	Garber et al., 2009
LRT	Direct	Evolutionary conservation model across 32 vertebrates	Variants in three sequenced human genomes	Chun and Fay, 2009
SNPs&GO	SVM	Variant effect prediction based on sequence information, evolutionary conservation and defined gene ontology score	Variants from SwissProt	Calabrese et al., 2009
B-SIFT	Direct	Sequence conservation metrics that calculate the difference between wild-type and mutant allele	Variants from SwissProt database and protein specific study (Dnase I)	Lee et al., 2009
PolyPhen-2	NB	Considering sequence conservation, Structure parameters such as hydrophobic propensity and B factor	Variants from HumDiv and HumVar from UniProt Database	Adzhubei et al., 2010
MutationTaster	NB	Prediction of mutation pathogenicity based on evolutionary conservation, splice-site changes, loss of protein features and changes that affect expression levels	Variants from OMIM database, HGMD and the literature as pathogenic set and neutral variants from dbSNP as controls	Schwarz et al., 2014
MutationAssessor	Direct	Evolutionary conservation patterns within protein families and across species using combinatorial entropy	Variants from UniProt database (HumSaVar)	Reva et al., 2011
Condel	Direct	Integration of five algorithms (SIFT, PolyPhen-2, MAPP, MutationAssessor, and Log R Pfam E-value) into single output score	Variants from HumVar, HumDiv, Cosmic database, IARC TP53 database	González-Pérez and López-Bigas, 2011
PROVEAN	Direct	Alignment-based score that can also assess in-frame insertions, deletions, and multiple amino acid substitutions	Missense variants and indels, replacements from UniProt database	Choi et al., 2012
FATHMM	HMM	Identification of pathogenic variants based on sequence conservation, protein domain-based information and species-specific pathogenicity weights. Also suitable for prediction of non-coding variations.	Variants from the HGMD and Uniprot databases	Shihab et al., 2013, 2015; Rogers et al., 2018
VEST	RF	Prioritization of variants underlying Mendelian diseases	Rare variants from HGMD database as pathogenic set and variants from ESP	Carter et al., 2013
Evolutionary Action	Direct	Prediction of variant effects on evolutionary fitness using a formal genotype-phenotype perturbation equation	Variants from 1000 Genomes Project	Katsonis and Lichtarge, 2014
MetaSVM	SVM	Ensemble score integrating nine functionality predictors (SIFT, PolyPhen-2, GERP++, MutationTaster, MutationAssessor, FATHMM, LRT, SiPhy and PhyloP)	Variants causing Mendelian diseases as pathogenic set and variants that are not associated with any phenotypes as controls, all from Uniprot database	Dong et al., 2015
MetaLR	RM	Same as MetaSVM but using logistic regression instead of SVM		Dong et al., 2015

(Continued)

TABLE 1 | Continued

Algorithm	Model	Basis of decision	Model training or evaluation	References
SuSPect	SVM	Sequence conservation metrics, structure features and additional network information	Variants from Humsavar database	Yates et al., 2014
PredictSNP	EL	Ensemble score integrating six functionality predictors (MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT and SNAP)	Variants mainly from SwissProt, HGMD, dbSNP and Humsavar database	Bendl et al., 2014
SNAP2	NN	Prediction of amino acid variations based on amino acid properties, predicted binding residues, predicted disordered and low-complexity regions, proximity to N- and C-terminus, statistical contact potentials, co-evolving positions, secondary structure and solvent accessibility	Variants from PMD, Swiss-Prot, OMIM, HumVar and protein specific data sets (LacI)	Hecht et al., 2015
REVEL	RF	Ensemble method tailored specifically for the prediction of rare genetic variant effects integrating MutPred, FATHMM, VEST, PolyPhen, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP, SiPhy, phyloP, and phastCons	Variants from HGMD as pathogenic set and neutral variants from ESP as controls	Ioannidis et al., 2016
ConSurf	Empirical Bayesian method and maximum likelihood estimation	Mapping of evolutionarily conserved residues on protein surfaces by estimating the evolutionary rates of each nucleic acid and amino acid sequence position using multiple sequence alignments. Also offers RNA secondary structure predictions.	Protein with at least five known 3D structure homologs and precise annotation of their functional sites (with different nature)	Ashkenazy et al., 2016
VIPUR	RM	Combination of sequence- and structure-based features to identify and functionally interpret deleterious variants	Variants from HumDiv and UniProt with clear evidence of protein disruption	Baugh et al., 2016
Envision	GTB	Decision tree ensemble-based tool using a stochastic gradient boosting learning algorithm	Variants from nine large-scale experimental mutagenesis datasets in eight proteins	Gray et al., 2018
EVmutation	Direct	Unsupervised method exploiting sequence conservation by incorporating interaction information between all pairs of residues in protein	34 data sets from 21 proteins and a tRNA gene extracted from 27 publications	Hopf et al., 2017
PredSAV	GTB	Identification of pathogenic variants based on sequence, structure, residue-contact networks as well as structural neighborhood features	Human variants from Uniprot and OMIM as pathogenic set and Ensemble variants as neutral controls	Pan et al., 2017
SNPMuSiC	NN	Structure stability based, implement PoPMuSiC and HoTMuSiC on the basis of 13 statistical potentials (distance potentials, solvent accessibility potentials and torsion potentials) and 2 biophysical characteristics (solvent accessibility of mutated residue and difference in volume)	Variants from dbSNP, SwissVar and HumSaVar datasets	Ancien et al., 2018
DEOGEN2	RF	Integration of 11 scores and metrics into one meta-score, considering evolutionary features, folding predictions, domain information as well as gene features to identify deleterious variants	Training and test on variants from the UniProt Humsavar16 dataset	Raimondi et al., 2017
ADME prediction framework	Direct	Integration of prediction scores from five orthogonal algorithms (LRT, MutationAssessor, PROVEAN, VEST3 and CADD) using parameters optimized for pharmacogenes	Training and validation specifically on experimentally characterized pharmacogenetic data sets from 43 ADME genes	Zhou et al., 2018

HMM, hidden Markov model; SVM, support vector machine; NB, naive Bayes classifier; EL, ensemble learning; RF, random forest; RM, regression model; NN, neural networks; GTB, gradient tree boosting; HGMD, Human Gene Mutation Database; OMIM, Online Mendelian Inheritance in Man; ESP, Exome Sequencing Project; PMD, Protein Mutant Database.

neutral variants. For pharmacogenetic predictions such a strategy is associated with multiple problems: Firstly, training on disease-associated data sets will, in the best case, result in prediction models that accurately predict the pathogenicity of variants. However, only very few ADME genes are directly associated with disease, suggesting that pathogenicity is not the right endpoint to inform about variant effects in the pharmacogenetic arena. Secondly, while evolutionary conservation constitutes a useful metric to predict functional consequences in genes under

purifying selection, evolutionary conservation in pharmacogenes is generally much lower (Fujikura, 2016), indicating that conservation cannot reliably inform about functional impacts of variations in pharmacogenes. Finally, the choice of common polymorphisms as neutral training sets is problematic. Genetic variants that occur with high frequencies are not necessarily functionally neutral, particularly in pharmacogenetic loci, as evidenced by a multitude of high-frequency loss of function variants in *CYP* genes, such as *CYP3A5**3 (MAF = 95% in

Europeans), *CYP2C19**2 (MAF = 34% in South Asians) and *CYP2D6**4 (MAF = 16% in Latinos) (Zhou et al., 2017).

The indicated problems incentivized us to develop a prediction framework tailored specifically toward pharmacogenetic functionality assessment (Zhou et al., 2018). Specifically, the model was devised using a two-step procedure: Firstly, functionality classification threshold of 18 commonly used functional prediction algorithms were optimized by leveraging a dataset of 337 experimentally characterized pharmacogenetic variants using 5-fold cross validations. In a second step, we integrated the best performing orthogonal algorithms following a strategy that had been shown to further improve predictive accuracy (Martelotto et al., 2014). The resulting method achieved 93% for both sensitivity and specificity for both loss-of-function and functionally neutral variants. Moreover, the returned score can provide quantitative estimates of the effect of the variant in question on gene function, thus facilitating the functional and personalized interpretation of an individual's NGS-based pharmacogenome.

Recent progress in large-scale experimental mutagenesis screens provides a promising approach to further expand the development of powerful training resources for missense variant effect predictors. While such a strategy has already been used to develop a prediction method based on 10 proteins from different species with disparate structures (Gray et al., 2018), we propose that deep mutational scanning data from ADME proteins is likely to substantially refine the resulting model for pharmacogenetic predictions. For such an endeavor, we recommend to use multiple substrates for each protein, as correlations between prediction and experiments improved with more comprehensive interrogation of protein function (Gallion et al., 2017). Combined with ADME-optimized prediction models, we envision that such an approach can further enhance the predictive accuracy of *in silico* methods and yield sufficiently accurate tools to allow for the clinical implementation of computational pharmacogenetic predictions.

Utilization of Structural Data

While evolutionary conservation scores can provide useful metrics to assess the pathogenicity of missense variants, they have limitations when applied to the less conserved genes, such as most ADME genes, which prompted the search for additional orthogonal *in silico* methods. To this end, the analysis of predicted or experimental structural data provides an appealing concept, as the correct folding of polypeptide chains into three-dimensional tertiary structures is of paramount importance for their biological functions. Structure-based approaches either directly use known crystal or NMR structures, preferably at high resolution <2–3 Å (Wlodawer et al., 2008) or, should such data not be available, leverage knowledge of the experimental 3D structures of homologous sequences (Table 2).

The effect of variants is predicted by how the folding free energy difference between the unfolded and folded states (ΔG°) is modified upon point mutations ($\Delta\Delta G^\circ$) with negative and positive values of $\Delta\Delta G^\circ$ indicating destabilizing and stabilizing mutations, respectively. In recent years a large number of mechanistically diverse approaches have been presented,

with machine learning-based strategies being most prevalent. SDM constitutes a statistical potential energy function that can estimate variant effects on protein stability (Topham et al., 1997). This approach pioneered the knowledge-based prediction of mutation effects on protein stability and has also been successfully used in combination with machine learning techniques (Pires et al., 2014a). An updated version of the tool, SDM2 (Pandurangan et al., 2017), with a 5-fold increase in underlying structural information as well as extensions for interaction modeling can be accessed through a free, publically available web server interface. Similarly, the algorithm HOPE (Venselaar et al., 2010) can calculate structural and functional effects of amino acid exchanges based on homology modeling. It should be however noted that most of the current tools are strongly biased toward the detection of destabilizing effects (Pucci et al., 2018).

Approximately 70% of the human proteome can be structurally modeled by homology (Somody et al., 2017). Yet, the number of resolved 3D structures for genes involved in drug ADME remains relatively low, at least in part due to the membrane bound nature of many of these proteins. Furthermore, as many metabolic enzymes, such as cytochrome p450s (CYPs) exhibit marked active-site flexibility, which often results in ligand-induced conformational changes, prediction of variant effects based on direct structural data is difficult for these proteins and substrate-specific effects have to be considered. Thus, while the prediction of amino acid exchanges on substrate metabolism remain difficult, folding stability of variant proteins of interest can be estimated using existing computational tools based on sequence homology modeling (Kulshreshtha et al., 2016).

EVALUATION OF TRUNCATION VARIANTS

Drug metabolizing enzymes and transporters have been found to harbor a multitude of truncation variants, such as micro-insertions and micro-deletions (indels) causing frameshifts, stop-gain and start-lost variants. Some of these variants are clinically relevant and occur with high frequencies in specific populations, including the stop-gain variant *CYP2C19**3 in East Asians and the frameshift variants *CYP2D6**3 and *CYP2D6**6 in Europeans (Zhou et al., 2017). As most pharmacogenes have only minor endogenous functions, they are under low evolutionary pressure and, consequently, such loss-of-function variants are often not selected against (Lauschke et al., 2017). Moreover, it has been speculated that pharmacogenetic loss-of-function alleles can even be selected for in modern humans, possibly due to reduced bioactivation of dietary toxicants (Fujikura, 2016). Truncation variants are commonly assumed to have deleterious effects and only few studies have been presented that provide approaches to quantitatively assess the functional consequences of such mutations (Cline and Karchin, 2011).

Early bioinformatic tools, such as LOFTEE, prioritize truncation variants based on a set of empirical rules, including whether the variant of interest occurs in the last 5% of transcript or whether the truncating allele is the ancestral

TABLE 2 | Methods to predict the functional effect of missense variants based primarily on structural features.

Algorithm	Model	Basis of decision	Model training or evaluation	References
SDM	Direct	Predicts variant effects on thermal protein stability using conformationally constrained environment-specific substitution tables derived from 2,054 protein family sequence and structure alignments from the TOCCATA database	Validated on 2,690 SNVs from 132 different protein structures.	Topham et al., 1997; Pandurangan et al., 2017
I-Mutant	SVM	Protein structure or sequence-based prediction of point mutation effects on protein stability	Training and testing on thermodynamic experimental data of free energy changes of protein stability upon mutation from the ProTherm database	Capriotti et al., 2005
HOPE	Direct	Analyzes the structural and functional effects of point mutations based on available crystal structures, homology modeling and sequence information.	Evaluated using case studies.	Venselaar et al., 2010
mCSM	RM	Translation of distance patterns between atoms into graph-based signatures providing data that is complementary to potential energy based approaches	Prediction of protein stability changes, protein-protein and protein-nucleic acid interactions and pathogenicity based on an array of preexisting experimental data sets	Pires et al., 2014b
DUET	SVM	SVM predictor that integrates mCSM and SDM in a consensus prediction	Benchmarking again mCSM and SDM alone on p53 data set.	Pires et al., 2014a
STRUM	GTB	Predicts variant effects on protein stability based on 3D models constructed by iterative threading assembly refinement simulations	Evaluated on 3,421 experimentally determined mutations distributed across 150 proteins.	Quan et al., 2016
ELASPIC	GTB	Predicts effects of mutations on protein folding and protein–protein interactions using homology modeling of domains and domain–domain interactions	Performance analysis via case study using EP300 mutations found in COSMIC	Witvliet et al., 2016
SAAFEC	RM	Prediction of effects of amino acid changes on folding free energy using a Molecular Mechanics Poisson-Boltzmann approach	Training and testing on thermodynamic experimental data of free energy changes of protein stability upon mutation from the ProTherm database	Getov et al., 2016

SVM, support vector machine; RM, regression model; GTB, gradient tree boosting.

state (MacArthur et al., 2012). Other approaches, such as Likelihood-ratio scoring (Zia and Moses, 2011), SIFT Indel (Hu and Ng, 2012) and NutVar (Rausell et al., 2014), primarily utilize the evolutionary conservation of amino acid residues. However, predictive performance of these tools for loss-of-function mutations is limited when trained on only missense mutations. Moreover, these methods are trained on genes that have high-quality annotations, which poses problems for the functional interpretation of truncation variants in genes for which such annotations are not readily available.

To overcome these shortcomings, CADD was developed by integrating many diverse functional genomics annotations into a single score for each variant, which allows to estimate the impact of all classes of genetic variation, including truncating variants (Kircher et al., 2014). Newer approaches, such as DDIG-in (Folkman et al., 2015) and VEST-Indel (Douville et al., 2016) supplement conservation-based features with information about sequence and structural properties at nucleotide and protein levels as well as intrinsic disorder predictions from the region affected by stop gain and frameshift variants. Notably, the recently developed tool ALoFT (Annotation of Loss-of-Function Transcripts) can categorize the pathogenic importance of putative loss-of-function mutations by integrating variant

information with redundancy and haplosufficiency data of the corresponding gene (Balasubramanian et al., 2017). However, aforementioned methods are primarily focused on distinguishing benign and disease-causing mutations. Thus, future studies are needed to evaluate whether this emphasis on the pathogenicity of variants might affect the performance of these methods regarding the functionality prediction of truncating variants in genes not associated with disease, such as many ADME genes.

In addition to impacts on functional and structural properties of proteins, truncating variants can affect nonsense-mediated mRNA decay (NMD). NMD is a conserved translation-dependent mechanism that is responsible for recognizing and eliminating aberrant mRNA transcripts to prevent the production of truncated peptides, thereby playing a critical role in preventing the accumulation of misfolded protein and subsequent initiation of the unfolded protein response (UPR) (Kervestin and Jacobson, 2012; Schoenberg and Maquat, 2012). Recently, Hsu et al. presented NMD Classifier, a tool for the systematic classification of NMD events, which was reported to correctly identify 99.3% of the NMD-causing transcript structural changes (Hsu et al., 2017). The incorporation of this information alongside functional estimates is expected to not only increase discriminative power but also to suggest the nature

of the functional impact of a given variant. Interestingly, there is evidence that NMD efficiency varies between individuals and that these differences correlate with response to NMD inhibitors in cystic fibrosis patients (Linde et al., 2007; Kerem et al., 2008). While this phenomenon has to the best of our knowledge not been explicitly tested in the context of pharmacogenomics, inter-individual differences in NMD magnitude could, at least in part, explain the large differences in drug response between patients with loss-of-function genotypes (Jukić et al., 2018) and thus have important implications for therapy.

In summary, much progress has been made regarding the functional interpretation of variants causing truncations of the corresponding gene product and current computational tools are able to incorporate a variety of features into their predictions, including evolutionary conservation, sequence and structural information as well as putative effects on NMD. However, it remains to be demonstrated whether these available tools will also be suitable for the prediction of effects of truncation variants in poorly conserved pharmacogenetic loci.

PREDICTION OF ABERRANT SPLICING EVENTS

Splicing of pre-mRNA is a critical step during mRNA maturation in which introns are excised and exons are ligated. This process necessitates the presence of 5' and 3' splicing signals and branch point sequence and is further regulated by exonic and intronic splicing enhancer/silencer (ESE/ESS and ISE/ISS, respectively) (Lee and Rio, 2015; Shi, 2017). Mutations in these regions can disrupt the splicing process and result in aberrantly processed transcripts, which can trigger NMD or result in the production of dysfunctional proteins. The functional importance of genetic variants in splice sites is emphasized by estimates that around 15% of human pathogenic mutations cause dysregulation of splicing (Baralle et al., 2009).

Variants located in canonical splice sites are considered having the largest effect on splicing events. Therefore, a multitude of computational algorithms were developed to handle the prediction of 5' and 3' splice site, such as NNSplice (Reese et al., 1997), MaxEntScan (Yeo and Burge, 2004), GeneSplicer (Pertea et al., 2001), and SplicePort (Dogan et al., 2007; **Table 3**). Moreover, variants outside splice sites can have substantial effects on splicing (Soukarieh et al., 2016) and a variety of computational methods have been developed to predict the effect of such regulatory sequences. Examples are sequence the conservation-based algorithm Skippy (Woolfe et al., 2010) and the machine learning tools MutPred Splice (Mort et al., 2014), scSNVEL (Jian et al., 2014b), SPANR (Xiong et al., 2015), and CryptSplice (Lee et al., 2017). Further tools are available for the identification of branch point sequences (Corvelo et al., 2010; Zhang et al., 2017). Lastly, the secondary structure of pre-mRNAs can interfere with splice-site recognition, modulate spliceosome binding or can facilitate splicing efficiency by bringing splice donors and acceptors into close proximity (Warf and Berglund, 2010). Consequently, genetic variants that alter pre-mRNA structure were found to promote alternative splicing (Wan et al.,

2014), incentivizing the incorporation of structural information provided by tools, such as TurboFold (Harmanici et al., 2011) or CentroidFold (Sato et al., 2009), into variant effect predictions. For a more detailed description of structural RNA analyses we refer the interested reader to excellent recent reviews (Jian et al., 2014a; Lorenz et al., 2016; Ohno et al., 2018).

In ADME genes, dysregulation of splicing has long been recognized as a cause for inter-individual variability drug metabolism (Hanioka et al., 1990) and toxicity (Raida et al., 2001) and the liver was found to be among the tissues with highest levels of alternative splicing activity (Yeo et al., 2004). As splicing is highly tissue specific, these data indicate that algorithms for the prediction of variant splice effects in pharmacogenetics should ideally be trained on positive control sets for which aberrant splicing is confirmed in the tissue of interest, i.e., primarily liver. To this end, the GTEx project (GTEx Consortium, 2017) provides a rich resource that has already been successfully utilized for the identification of tissue-specific splice events in pharmacogenes (Chhibber et al., 2017).

In summary, the toolkit of available computational algorithms for the prediction of variant effects on splicing has rapidly grown and by now allows not only to evaluate direct impact on splice sites, but also to assess mutations in regulatory splice enhancers and silencers, as well as branch points. For the application of these methods for pharmacogenomics there is a need to benchmark available tools on splice variants in ADME genes. Moreover, we anticipate that the utilization of tissue-specific expression data will further refine splice site predictions.

FUNCTIONAL IMPACT OF VARIANTS IN UNTRANSLATED REGIONS

miRNAs play important roles in the regulation of mRNA stability and translation. miRNA-mRNA interaction occurs through conserved miRNA binding sites in the 3'-UTRs and at least 10% of all SNPs are located in 3'-UTRs and might affect complementary miRNA-mRNA pairing (Xiao et al., 2009). Furthermore, miRNAs have been shown to be important modulators of ADME gene expression profiles (Rieger et al., 2013). Therefore, functional interpretation of genetic variations within miRNA target sites constitutes an important factor for the prediction of the fate of corresponding transcript. Thus, to evaluate the potential relevance of genetic polymorphisms in UTRs various databases, such as the polymiRTS Database 3.0 (Bhattacharya et al., 2014) or MirSNP (Liu et al., 2012), provide useful resources that contains a collection of experimentally confirmed SNPs and indels not only in miRNA target sites but also in miRNA seed regions responsible for mRNA binding. Furthermore, a variety of other SNP effect prediction servers are publically available (Fehlmann et al., 2017).

In case no experimental data is available, various computational tools can be used to predict possible disruption of the miRNA-mRNA pairing for a given variant (**Table 3**). MicroSNiPer (Barenboim et al., 2010) and ImiRP (Ryan et al., 2016) identify and predict such disruptions by comparing the mutant 3'-UTR sequences with major variant databases.

TABLE 3 | Tools for the prediction of variant effects on splicing, transcript levels or translation.

Algorithm	Application	Basis of decision	Model training or evaluation	References
NMD Classifier	NMD	Prediction of NMD for a given transcript based on comparison to most similar coding transcript	Simulation-based evaluation based on screening artificial transcript structure-altering events	Hsu et al., 2017
NNSplice	Splicing (splice sites)	Sequence splice site analysis using HMM	Distinguish splice site sequences from sequences in the neighborhood of real splice sites	Reese et al., 1997
MaxEntScan	Splicing (splice sites)	Splice site analysis by modeling short sequence motifs using the maximum entropy principle with constraints estimated from available data.	1,821 transcripts unambiguously aligned across the entire coding region, spanning a total of 12,715 introns	Yeo and Burge, 2004
GeneSplicer	Splicing (splice sites)	Splice site prediction using maximal dependence decomposition with the addition of markov model to capture dependencies among neighboring bases	Annotated genes from the Exon-Intron Database	Pertea et al., 2001
SplicePort	Splicing (splice sites)	Splice site prediction using C-modified least squares learning based on positional and compositional sequence features	Training on 4,000 pre-mRNA human RefSeq sequences and test on B2Hum data set	Dogan et al., 2007
Skippy	Splicing (regulatory sequences)	Prediction of variants causing exon skipping, exon inclusion or ectopic splice site activation based on sequence information, proximity to splice junctions and evolutionary constraint of the peri-variant region	Multiple exonic splicing regulatory elements datasets as positive data and HapMap variants as splicing-neutral variants	Woolfe et al., 2010
MutPred Splice	Splicing (regulatory sequences)	Prediction of auxiliary splice sequences using multiple variant-, flanking exon- and gene-based features	Splicing variants from HGMD as pathogenic set and non-splicing variants from both HGMD and 1000G as neutral controls	Mort et al., 2014
scSNVEL	Splicing (splice sites)	Ensemble prediction using 8 algorithms using random forest learning	Splice variants from HGMD, SpliceDisease database and DBASS as pathogenic set and variants not implicated in splicing from both HGMD and 1000G as controls	Jian et al., 2014b
SPANR	Splicing (splice sites and splice regulatory sequences)	Integrating 1,393 sequence features from each exon and its neighboring introns and exons to identify splice sites as well as intronic and exonic splice regulators	10,689 exons that displayed evidence of alternative splicing	Xiong et al., 2015
CryptSplice	Splicing (splice sites)	Prediction of cryptic splice-site activation using an SVM model	Sequences from the annotated NN269 and HS3D splice datasets with positive sequence in splice sites and control sequence outside splice sites	Lee et al., 2017
Corvelo <i>et al.</i>	Splicing (branch points)	Analysis of splice site sequence conservation and position bias using SVM	A set of 8,156 conserved putative branch point sequences from 7 mammalian species	Corvelo et al., 2010
BPP	Splicing (branch points)	Identification of branch point motifs by integrating information on the branch point sequence and the polypyrimidine tract	Intron sequences longer than 300 nucleotides	Zhang et al., 2017
TurboFold	Splicing (pre-mRNA structure)	Probabilistic method that integrates comparative sequence analyses with thermodynamic folding models	Thorough benchmarking against three methods that estimate base pairing probabilities and eight tools for structural predictions based on known RNA structures	Harmanci et al., 2011
CentroidFold	Splicing (pre-mRNA structure)	RNA secondary structure prediction using the γ -centroid estimator	Validation based on 151 RNA experimentally determined RNA structures	Sato et al., 2009
mrSNP	miRNA binding	miRNA binding energy calculations for reference and variant containing sequence and report of binding difference	Evaluation based on variants that map to miRNA targets predicted by TargetScan	Deveci et al., 2014
PinPor	RBP binding	Bayesian network approach that incorporates information about sequence features, stabilization of RNA secondary structure and evolutionary conservation	Inframe indels from HGMD as pathogenic and common indels from 1000G as neutral controls	Zhang et al., 2014

HGMD, Human Gene Mutation Database; 1000G=1000 Genomes Project; DBASS, Database for Aberrant Splice Sites; NMD, nonsense-mediated decay; HMM, hidden Markov model; RBP, RNA binding protein.

Similarly, mrSNP can predict the effect of any variant identified in NGS-based projects on miRNA-target transcript interaction (Deveci et al., 2014). However, it is important to note that miRNA target predictions seem to have a high false-positive rate (Pinzón et al., 2017), suggesting that these problems might be lingering for studies utilizing miRNA-target databases without stringent experimental validations. Besides predicting the effect of genetic variants in putative miRNA target sites, multiple online tools are available for inverse approaches, analyzing variants in miRNAs or pre-miRNAs for possible deleterious effects. For more comprehensive collection of miRNA related variant interpretation tools the reader is referred to the recent reviews and online resources (Akhtar et al., 2016; Moszynska et al., 2017).

In addition, recent approaches expanded the methodological portfolio beyond miRNA binding site prediction to include effects of UTR variants on binding of RNA-binding proteins (RBPs), translational efficacy and ribosomal loading. Effects of indels on RBP binding can be evaluated using PinPor, which has been demonstrated to have some success in distinguishing disease-causing and neutral indels (Zhang et al., 2014). Furthermore, Sample *et al.* presented the preprint of a deep learning approach based on experimental polysome profiling to predict the impact of UTR sequence on translation (Sample et al., 2018). These developments nicely indicate the diversification of parameters that can be incorporated into variant effect predictions, thus further refining biological interpretation of NGS data sets.

ANALYSIS OF REGULATORY VARIANTS

Non-coding regions account for more than 99% of the human genome and, consequently, their consideration substantially expands the analysis space of computational predictions. Variants in non-coding regions can affect regulatory elements, such as promoters, enhancers, silencers, and insulators, which, in turn, may alter their affinity to transcription factor or remodel the local chromatin structure (Zhang and Lupski, 2015; Deplancke et al., 2016). Accurate prediction of the functional consequences of such variants constitutes one of the major challenges in human genetics.

To interpret noncoding variants, a variety of different strategies have been presented. The first approaches, such as SiPhy (Garber et al., 2009), PhyloP (Pollard et al., 2010), PhastCons (Siepel et al., 2005), GERP++ (Davydov et al., 2010), or SCONE (Asthana et al., 2007), were based on evolutionary constraint using sequence alignments. However, the observation that no enhanced constraints were identified in regulatory elements at the level of DNA sequence despite conserved transcription factor binding led to the realization that conservation of regulatory regions can only be a weak indicator of the functional effects of SNVs in regulatory regions (Schmidt et al., 2010; Arbiza et al., 2013). Consequently, conservation metrics were complemented with additional functional genomics features, such as the sequence and genic context, transcription factor binding profiles (Johnson et al., 2007), histone modification data (Zhang et al., 2010) and DNase

I hypersensitive sites (Boyle et al., 2008) in an attempt to improve prediction quality. Based on these rich data sets, a variety of ensemble classifiers were developed using various machine learning approaches that aim to distinguish neutral from pathogenic variants, including GWAVA (Ritchie et al., 2014), CADD (Kircher et al., 2014), FATHMM (Shihab et al., 2013, 2015; Rogers et al., 2018), DANN (Quang et al., 2015), DIVAN (Chen et al., 2016), and Genomiser (Smedley et al., 2016) (Table 4).

In contrast, other methods, such as gkm-SVM (Lee et al., 2015) and DeepSEA (Zhou and Troyanskaya, 2015) have been developed to predict regulatory elements based on primary sequence alone. Trained on publically available cell type-specific chromatin data provided by ENCODE (The ENCODE Project Consortium, 2012) and the Roadmap Epigenomics Project (Roadmap Epigenomics Consortium et al., 2015) as well as transcription factor binding patterns accessible via JASPAR (Khan et al., 2018), these algorithms predict to what extent a genetic variant will cause changes to the local chromatin profiles and how these effects translate into functional consequences. The resulting data demonstrate that inferring consequences from functional genomics data is highly cell type and context specific and relies on biologically appropriate training sets. These convincing findings incentivize the generation of functional genomics data from carefully phenotyped human tissues involved in drug ADME to derive tissue-specific regulatory lexica and we envision that training machine learning approaches on these data sets will substantially increase the power of regulatory pharmacogenetic prediction classifiers.

As with coding variants, the use of potentially biased training sets and multi-dimensional circularity between training and test data constitutes an inherent problem for current variant prediction tools (Grimm et al., 2015). For instance, a variety of algorithms consider common variants from the 1000 Genomes project as functionally neutral control sets for model training. However, while these variants are likely to be depleted of pathogenic variants in haploinsufficient genes, many common variants entail functional consequences in their respective gene product, particularly if the gene is rapidly evolving, such as many CYP genes. Similar problems arise when the model is trained using phenotype associated GWAS polymorphisms as functional variant sets, as only 5.5% of GWAS index SNPs are estimated to be causal whereas the remainder is only in linkage disequilibrium with the true functional variant in the locus (Farh et al., 2015).

To overcome these problems, unsupervised approaches have been developed that do not rely on the labeling of training data, thereby reducing the dependence on preexisting variant classifications and existing models of mutation. These unsupervised models, such as GenoCanyon (Lu et al., 2015) and Eigen (Ionita-Laza et al., 2016), represent powerful tools for the genome-wide interpretation of variants. However, as they are calibrated on genome-wide data, it remains to be determined whether gene class-specific peculiarities, such as low evolutionary conservation in ADME genes, might affect the predictive accuracy of these approaches for pharmacogenetic applications.

TABLE 4 | Algorithms for the functional interpretation of regulatory variants.

Algorithm	Model	Application	Model training	Features	References
FATHMM	HMM	Pathogenic variants	HGMD regulatory variants as pathogenic set and common 1000G variants as controls	Evolutionary conservation data (PhastCons and PhyloP), chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data, genome segmentation, frequency data (1000G and ESP) and information about genic and sequence context	Shihab et al., 2013, 2015; Rogers et al., 2018
GWAVA	RF	Pathogenic variants	HGMD regulatory variants as pathogenic set and common 1000G variants as controls	Evolutionary conservation data (GERP), chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data, genome segmentation, frequency data (1000G) and information about genic and sequence context	Ritchie et al., 2014
CADD	SVM	Deleterious variants	Sites with MAF<5% where for which the human genome differed from the inferred human-chimp ancestral genome and equal number of simulated variants	Evolutionary conservation data (GERP++, PhastCons and PhyloP), chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data, genome segmentation, frequency data (1000G and ESP) and information about genic and sequence context	Kircher et al., 2014
DANN	NN	Deleterious variants	Same as CADD but using deep neural networks instead of linear SVM.		Quang et al., 2015
DeepSEA	NN	Variants that affect gene expression	HGMD regulatory variants, eQTLs and NHGRI GWAS phenotype-associated SNPs	Evolutionary conservation data (GERP++, PhastCons and PhyloP), chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data	Zhou and Troyanskaya, 2015
gkm-SVM	SVM	Variants that affect gene expression	Tissue-specific enhancer sequences marked by H3K4me1 from length-, GC content- and repeat-matched random control	Definition of tissue-specific regulatory dictionary based on chromatin accessibility (DNase-HSS) and H3K4me1 ChIP-Seq data	Lee et al., 2015
fitCons	INSIGHT	Prediction of <i>cis</i> -regulatory elements	Unsupervised classifier that clusters genomic regions on the basis of functional genomic data and then estimates a probability of fitness consequences for each group from associated patterns of genetic polymorphism and divergence.	Evolutionary conservation data (GERP, PhastCons and PhyloP), chromatin accessibility (DNase-HSS), TF binding and histone modification ChIP-Seq data, genome segmentation and RNA-Seq data	Gulko et al., 2015
GenoCanyon	US	Identification of functional regions	Unsupervised classifier based on the estimated proportion of functional regions in the human genome.	Evolutionary conservation data (GERP and PhyloP), chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data	Lu et al., 2015
DIVAN	EL	Disease-specific risk variants	Disease-specific regulatory NHGRI GWAS SNPs and common 1000G variants or benign GWAS SNPs as controls	Chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data	Chen et al., 2016
Genomiser	RF	Mendelian disease	Sites with MAF<5% where for which the human genome differed from the inferred human-chimp ancestral genome as functionally neutral variation and 453 positive variants based on literature review	Evolutionary conservation data (GERP++, PhastCons and PhyloP), chromatin accessibility (DNase-HSS), TF binding and histone modification ChIP-Seq data, frequency data (1000G and ESP) and information about enhancer context from FANTOM5	Smedley et al., 2016
Eigen	US	Effect of variants on gene expression and disease risk	Unsupervised classifier based on the blockwise conditional independence between annotations given the functional impact of the variant.	Evolutionary conservation data (GERP, PhastCons and PhyloP), chromatin accessibility (DNase-HSS and FAIRE-Seq), TF binding and histone modification ChIP-Seq data and frequency data (1000G)	Ionita-Laza et al., 2016

RF, random forest; SVM, support vector machine; HMM, hidden Markov model; EL, ensemble learning; NN, neural networks; INSIGHT, Inference of Natural Selection from Interspersed Genomically Coherent Elements Gronau et al., 2011; US, unsupervised; HGMD, Human Gene Mutation Database; 1000G, 1000 Genomes Project; ESP, Exome Sequencing Project; TF, transcription factor; HSS, hypersensitive site; FAIRE, Formaldehyde-Assisted Isolation of Regulatory Elements Giresi et al., 2007; NHGRI, National Human Genome Research Institute.

CONCLUSIONS

Technical progress in NGS technology has resulted in its routine application in medical genetics and clinical diagnostics. In contrast, clinical implementation of NGS-based

pharmacogenomics is largely lagging behind (Lauschke and Ingelman-Sundberg, 2016b; Ji et al., 2018). Most importantly, in order to utilize the major advantage of NGS-based genotyping, which is the discovery of the entire panorama of the individual's genetic portfolio, tools have to be in place, which allow to

translate these variability data into functional consequences and clinical recommendations. Whereas, the identification of rare putatively deleterious mutations in congenital diseases is aided by clear phenotypic alterations of the affected patient and the possibility to perform comparative genomic analyses of unaffected family members, pharmacogenomic phenotypes are generally more difficult to detect as they only present in a given context, such as exposure to specific medications. In the absence of drug response associations or experimental characterizations that support the functional interpretation of rare variants, there is thus an urgent need for reliable computational prediction tools to fill this space.

Importantly, recent developments in computational variant effect prediction methods promise to narrow the gap to meet the exacting demands on genomics applications in the clinics. Machine learning constitutes an important tool kit to fully harness the power of large data sets provided by NGS. However, these approaches rely on accurate labeling of input variants, i.e., training data need to be correctly classified into deleterious and functionally neutral variants. Thus, we advocate

for approaches that leverage smaller data sets of variants for which comprehensive experimental or functional genomic data is available instead of training algorithms on large but functionally poorly annotated data, such as treating all common polymorphisms identified in the 1000 Genomes Project as functionally neutral. In addition, we endorse previous appeals for the sharing of codes and data sets, which will enable comparative benchmarking of newly developed tools and algorithms and will accelerate research progress within the area of computational pharmacogenomics and beyond (Kalinin et al., 2018).

The functional consequences of missense variants have been most extensively studied. Respective methods base their predictions on evolutionary conservation and structural information of the polypeptide encoded by the respective gene. Importantly, while evolutionary conservation is a suitable measure to inform about the deleteriousness of a variant, i.e., its effect on organismal fitness, it is not suitable for the prediction of variant effects in genes under low selective pressure, such as most pharmacogenes. Recognition of these conceptual problems resulted in the development of computational predictors trained









<div><div><div></div><div>Common variant</div></div><div><div></div><div>Rare variant</div></div></div>	A	B	C
	Past	Current	Future
Primary genotyping method	SNP array (10s -100s variants) 	WES (many 1000s variants) 	WGS (All variants) 
Major source of training data	—	Pathogenic variants database 	Literature  ADME specific datasets - Deep mutational scanning - Mutagenesis studies in MPS
Assessed features	—	Missense and nonsense variants	All (missense, nonsense, non-coding and regulatory variations)
Functional interpretation	 Literature	Computational:  Pathogenicity prediction	Computational:  ADME-specific functionality prediction
Predictive accuracy	Very high	Low	High
Extent of personalization	Very low	High	Very high

FIGURE 2 | The past, present and future of pharmacogenetic phenotype predictions. **(A)** Conventionally, pharmacogenetic predictions were based on the interrogation of few common candidate SNPs, whose functional effects were predicted based on extensive literature evidence, resulting in high predictive accuracy but only few considered variations. **(B)** With increasing prevalence of whole exome sequencing (WES), a multitude of pharmacogenetic variants with unknown functional relevance are identified. These variants can be interpreted using computational methods. However, current algorithms are generally trained to detect the pathogenicity rather than the functionality of queried variants, resulting in overall relatively low predictive accuracy. Furthermore, only effects of missense and nonsense variants are evaluated. **(C)** In the near future, whole genome sequencing (WGS) will become the predominant genotyping methodology, revealing not only coding variants but also variants in regulatory regions and introns. To facilitate interpretation of this data, we envision that pharmacogenetic predictors will be directly trained on functionally annotated ADME data sets. Emerging technologies, such as deep mutational scanning for the systematic interrogation of missense variants or mutagenesis screens in microphysiological systems (MPS) for the characterization of variants in regulatory regions, provide powerful tools to generate these data, boosting the predictive performance of data hungry machine learning tools. These advances allow to go beyond the interpretation of missense and nonsense variants and to include also non-coding and regulatory variations into pharmacogenetic assessments.

specifically on ADME missense variants (Zhou et al., 2018). We envision that these approaches will become more powerful with increasing functionally annotated pharmacogenetic variant data.

Furthermore, multiple strategies have been developed to analyze the functional impact of variants in non-coding regions of the genome, which are increasingly recognized as a substantial contributor to inter-individual variability. An increasing number of algorithms is by now available that base their predictions on a multitude of different parameters, including effects on miRNA binding or translational efficiency, modulation of splicing and impacts on transcriptional events by disruption of transcription factor binding sites or polymerase loading (Figure 1). While these developments provide a methodological arsenal to comprehensively characterize all different classes of genetic variants, these methods are generally trained on pathogenic variant sets and have not been benchmarked on independent data sets. Thus, their predictive power for pharmacogenetic assessments remains to be evaluated.

The prediction of drug metabolism phenotypes based on the genotype of the individual has made tremendous progress over the last decades (Figure 2). Conventional approaches use data from few candidate variants for which substantial *in vitro* or *in vivo* characterization data was available to predict drug response. While this strategy has been successful in incorporating common pharmacogenetic variability into clinical decision-making, they fail to address functional effects of the vast extent of rare genetic variants. To also include rare variants, pilot programs were initiated in which WES was used to comprehensively interrogate the genetic landscape of pharmacogenomic loci (Bielinski et al., 2014). However, analyses were restricted to pharmacogenetic missense variants and the effects of SNVs with unknown functional relevance were interpreted using computational models trained on pathogenic data sets with negative impacts on the accuracy of phenotype predictions, as discussed above. Thus, while these strategies constitute an important step toward the further personalization of genotype-guided treatment decisions their predictive accuracy is rather low.

We expect that technological, methodological and analytical progress will contribute to a further refinement of NGS-guided drug treatment in the near future. Firstly, technological advances will result in an increasing dissemination of WGS, which facilitates the incorporation of the entire profile of an individual's genetic variability, including regulatory variants, into pharmacogenetic predictions. Secondly, we envision that novel high-throughput methodologies for functional characterizations, such as deep mutational scanning, will provide powerful approaches to generate large functionally annotated pharmacogenetic variant data sets. In addition, recent advances in the development of microphysiological systems (MPS) that

allow to model key target tissues associated with drug metabolism or safety provide (Ewart et al., 2018) provide promising tools to generate tissue-specific and human-relevant data sets for studies of gene-drug interactions (Ingelman-Sundberg and Lauschke, 2018). Using this integrated wealth of functional pharmacogenetic data to train machine learning models aspires to provide high-accuracy predictions based on the entire genetic variability landscape of the respective patient.

Importantly, leveraging this information as guidance for clinical decision-making promises to increase treatment efficacy and reduce the risks of adverse events in carriers of pharmacogenetic variants whose effects have not been experimentally evaluated. Current market analysis estimates suggests that implementation of artificial intelligence into the clinical decision support toolbox might increase average life expectancy in the Western World by 0.2–1.3 years and reduce total health care expenditures by 5–9%, corresponding to 2 trillion to 10 trillion USD globally per year (Bughin et al., 2017). However, in order to realize these exciting prospects, there is a need for prospective, randomized controlled trials that evaluate patient outcomes and cost-effectiveness of such preemptive advice across genes, drugs and health care systems.

In summary, computational prediction methods are essential for the implementation of NGS into clinical decision-making. While much progress has been made and a plethora of conceptually diverse tools is already available, there is a need to develop specialized methods that are optimized for the prediction of variant functionality rather than pathogenicity and are calibrated specifically on pharmacogenetic data. We envision that technological, methodological and analytical advances will soon allow to comprehensively predict variant effects with sufficient accuracy to justify the design of trials in which the clinical value of NGS-guided treatment decisions can be tested in a prospective setting.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The work in the authors' laboratory is supported by the Swedish Research Council [grant agreement numbers: 2016-01153 and 2016-01154], by the Strategic Research Programme in Diabetes at Karolinska Institutet, by the European Union's Horizon 2020 research and innovation program U-PGx [grant agreement No. 668353], and by the Lennart Philipson and Harald och Greta Jeansson Foundations.

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Ahn, E., and Park, T. (2017). Analysis of population-specific pharmacogenomic variants using next-generation sequencing data. *Sci. Rep.* 7: 8416. doi: 10.1038/s41598-017-08468-y
- Akhtar, M. M., Micolucci, L., Islam, M. S., Olivieri, F., and Procopio, A. D. (2016). Bioinformatic tools for microRNA dissection. *Nucleic Acids Res.* 44, 24–44. doi: 10.1093/nar/gkv1221

- Ancien, F., Pucci, F., Godfroid, M., and Rooman, M. (2018). Prediction and interpretation of deleterious coding variants in terms of protein structural stability. *Sci. Rep.* 8: 4480. doi: 10.1038/s41598-018-22531-2
- Arbiza, L., Gronau, I., Aksoy, B. A., Hubisz, M. J., Gulko, B., Keinan, A., et al. (2013). Genome-wide inference of natural selection on human transcription factor binding sites. *Nat. Genet.* 45, 723–729. doi: 10.1038/ng.2658
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. doi: 10.1093/nar/gkw408
- Asthana, S., Roytberg, M., Stamatoyannopoulos, J., and Sunyaev, S. (2007). Analysis of sequence conservation at nucleotide resolution. *PLoS Comput. Biol.* 3:e254. doi: 10.1371/journal.pcbi.0030254
- Balasubramanian, S., Fu, Y., Pawashe, M., McGillivray, P., Jin, M., Liu, J., et al. (2017). Using ALOFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat. Commun.* 8: 382. doi: 10.1038/s41467-017-00443-5
- Baralle, D., Lucassen, A., and Buratti, E. (2009). Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep.* 10, 810–816. doi: 10.1038/embor.2009.170
- Barenboim, M., Zoltick, B. J., Guo, Y., and Weinberger, D. R. (2010). MicroSNiPer: a web tool for prediction of SNP effects on putative microRNA targets. *Hum. Mutat.* 31, 1223–1232. doi: 10.1002/humu.21349
- Baugh, E. H., Simmons-Edler, R., Müller, C. L., Alford, R. F., Volfovsky, N., Lash, A. E., et al. (2016). Robust classification of protein variation using structural modelling and large-scale data integration. *Nucleic Acids Res.* 44, 2501–2513. doi: 10.1093/nar/gkw120
- Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., et al. (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput. Biol.* 10:e1003440. doi: 10.1371/journal.pcbi.1003440
- Bhattacharya, A., Ziebarth, J. D., and Cui, Y. (2014). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 42, D86–D91. doi: 10.1093/nar/gkt1028
- Bielinski, S. J., Olson, J. E., Pathak, J., Weinshilboum, R. M., Wang, L., Lyke, K. J., et al. (2014). Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time—using genomic data to individualize treatment protocol. *Mayo Clin. Proc.* 89, 25–33. doi: 10.1016/j.mayocp.2013.10.021
- Boycott, K. M., Vanstone, M. R., Bulman, D. E., and MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* 14, 681–691. doi: 10.1038/nrg3555
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., et al. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* 132, 311–322. doi: 10.1016/j.cell.2007.12.014
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, S., Allas, T., Dahlström, P., et al. (2017). *Artificial Intelligence the Next Digital Frontier?* McKinsey and Company Global Institute.
- Bush, W. S., Crosslin, D. R., Owusu-Obeng, A., Wallace, J., Almoguera, B., Basford, M. A., et al. (2016). Genetic variation among 82 pharmacogenes: the PGRNseq data from the eMERGE network. *Clin. Pharmacol. Therapeut.* 100, 160–169. doi: 10.1002/cpt.350
- Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., and Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* 30, 1237–1244. doi: 10.1002/humu.21047
- Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., and Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell* 173, 1581–1592. doi: 10.1016/j.cell.2018.05.015
- Capriotti, E., Calabrese, R., and Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 22, 2729–2734. doi: 10.1093/bioinformatics/btl423
- Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 33, W306–W310. doi: 10.1093/nar/gki375
- Carter, H., Douville, C., Stenson, P. D., Cooper, D. N., and Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 14(Suppl. 3), S3. doi: 10.1186/1471-2164-14-S3-S3
- Chen, L., Jin, P., and Qin, Z. S. (2016). DIVAN: accurate identification of non-coding disease-specific risk variants using multi-omics profiles. *Genome Biol.* 17: 252. doi: 10.1186/s13059-016-1112-z
- Chhibber, A., French, C. E., Yee, S. W., Gamazon, E. R., Theusch, E., Qin, X., et al. (2017). Transcriptomic variation of pharmacogenes in multiple human tissues and lymphoblastoid cell lines. *Pharmacogenomics J.* 17, 137–145. doi: 10.1038/tpj.2015.93
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* 7:e46688. doi: 10.1371/journal.pone.0046688
- Chun, S., and Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561. doi: 10.1101/gr.092619.109
- Cline, M. S., and Karchin, R. (2011). Using bioinformatics to predict the functional impact of SNVs. *Bioinformatics* 27, 441–448. doi: 10.1093/bioinformatics/btq695
- Corvelo, A., Hallegger, M., Smith, C. W., and Eyra, E. (2010). Genome-wide association between branch point properties and alternative splicing. *PLoS Comput. Biol.* 6:e1001016. doi: 10.1371/journal.pcbi.1001016
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., 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–e1001013. doi: 10.1371/journal.pcbi.1001025
- Deplancke, B., Alpern, D., and Gardeux, V. (2016). The genetics of transcription factor DNA binding variation. *Cell* 166, 538–554. doi: 10.1016/j.cell.2016.07.012
- Deveci, M., Catalyürek, U. V., and Toland, A. E. (2014). mrSNP: software to detect SNP effects on microRNA binding. *BMC Bioinformatics* 15:73. doi: 10.1186/1471-2105-15-73
- Dogan, R. I., Getoor, L., Wilbur, W. J., and Mount, S. M. (2007). SplicePort—an interactive splice-site analysis tool. *Nucleic Acids Res.* 35, W285–W291. doi: 10.1093/nar/gkm407
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24, 2125–2137. doi: 10.1093/hmg/ddu733
- Douville, C., Masica, D. L., Stenson, P. D., Cooper, D. N., Gyax, D. M., Kim, R., et al. (2016). Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). *Hum. Mutat.* 37, 28–35. doi: 10.1002/humu.22911
- Ewart, L., Dehne, E.-M., Fabre, K., Gibbs, S., Hickman, J., Hornberg, E., et al. (2018). Application of microphysiological systems to enhance safety assessment in drug discovery. *Annu. Rev. Pharmacol. Toxicol.* 58, 65–82. doi: 10.1146/annurev-pharmtox-010617-052722
- Farh, K. K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337–343. doi: 10.1038/nature13835
- Fehlmann, T., Sahay, S., Keller, A., and Backes, C. (2017). A review of databases predicting the effects of SNPs in miRNA genes or miRNA-binding sites. *Brief. Bioinformatics*. doi: 10.1093/bib/bbx155. [Epub ahead of print].
- Flanagan, S. E., Patch, A.-M., and Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet. Test. Mol. Biomarkers* 14, 533–537. doi: 10.1089/gtmb.2010.0036
- Folkman, L., Yang, Y., Li, Z., Stantic, B., Sattar, A., Mort, M., et al. (2015). DDIG-in: detecting disease-causing genetic variations due to frameshifting indels and nonsense mutations employing sequence and structural properties at nucleotide and protein levels. *Bioinformatics* 31, 1599–1606. doi: 10.1093/bioinformatics/btu862
- Fujikura, K. (2016). Premature termination codons in modern human genomes. *Sci. Rep.* 6:22468. doi: 10.1038/srep22468
- Fujikura, K., Ingelman-Sundberg, M., and Lauschke, V. M. (2015). Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenet. Genomics* 25, 584–594. doi: 10.1097/FPC.0000000000000172
- Gallion, J., Koire, A., Katsonis, P., Schoenegge, A.-M., Bouvier, M., and Lichtarge, O. (2017). Predicting phenotype from genotype: improving accuracy through more robust experimental and computational modeling. *Hum. Mutat.* 38, 569–580. doi: 10.1002/humu.23193
- Garber, M., Guttman, M., Clamp, M., Zody, M. C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by

- exploiting biased substitution patterns. *Bioinformatics* 25, i54–i62. doi: 10.1093/bioinformatics/btp190
- Getov, I., Petukh, M., and Alexov, E. (2016). SAAFE: predicting the effect of single point mutations on protein folding free energy using a knowledge-modified MM/PBSA approach. *Int. J. Mol. Sci.* 17, 512–514. doi: 10.3390/ijms17040512
- Giresi, P. G., Kim, J., McDaniel, R. M., Iyer, V. R., and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885. doi: 10.1101/gr.5533506
- González-Pérez, A., and López-Bigas, N. (2011). Improving the assessment of the outcome of Nonsynonymous SNVs with a consensus deleteriousness score, condel. *Am. J. Hum. Genet.* 88, 440–449. doi: 10.1016/j.ajhg.2011.03.004
- Gray, V. E., Hause, R. J., Luebeck, J., Shendure, J., and Fowler, D. M. (2018). Quantitative missense variant effect prediction using large-scale mutagenesis data. *Cell Syst.* 6, 116–124.e3. doi: 10.1016/j.cels.2017.11.003
- Grimm, D. G., Azencott, C.-A., Aicheler, F., Gieraths, U., MacArthur, D. G., Samocha, K. E., et al. (2015). The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Hum. Mutat.* 36, 513–523. doi: 10.1002/humu.22768
- Gronau, I., Arbiza, L., Mohammed, J., and Siepel, A. (2011). Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. *Mol. Biol. Evol.* 30, 1159–1171. doi: 10.1093/molbev/mst019
- GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. doi: 10.1038/nature24277
- Gulko, B., Hubisz, M. J., Gronau, I., and Siepel, A. (2015). A method for calculating probabilities of fitness consequences for point mutations across the human genome. *Nat. Genet.* 47, 276–283. doi: 10.1038/ng.3196
- Hanioka, N., Kimura, S., Meyer, U. A., and Gonzalez, F. J. (1990). The Human Cyp2d locus associated with a common genetic-defect in drug oxidation—a G1934-Ja base change in intron-3 of a mutant Cyp2d6 allele results in an Aberrant-3' Splice Recognition site. *Am. J. Hum. Genet.* 47, 994–1001.
- Harmanci, A. O., Sharma, G., and Mathews, D. H. (2011). TurboFold: iterative probabilistic estimation of secondary structures for multiple RNA sequences. *BMC Bioinformatics* 12:108. doi: 10.1186/1471-2105-12-108
- Hecht, M., Bromberg, Y., and Rost, B. (2015). Better prediction of functional effects for sequence variants. *BMC Genomics* 16(Suppl 8):S1. doi: 10.1186/1471-2164-16-S8-S1
- Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P., Springer, M., Sander, C., et al. (2017). Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* 35, 128–135. doi: 10.1038/nbt.3769
- Hsu, M.-K., Lin, H.-Y., and Chen, F.-C. (2017). NMD Classifier: A reliable and systematic classification tool for nonsense-mediated decay events. *PLoS ONE* 12:e0174798. doi: 10.1371/journal.pone.0174798
- Hu, J., and Ng, P. C. (2012). Predicting the effects of frameshifting indels. *Genome Biol.* 13, R9. doi: 10.1186/gb-2012-13-2-r9
- Ingelman-Sundberg, M., and Lauschke, V. M. (2018). Human liver spheroids in chemically defined conditions for studies of gene–drug, drug–drug and disease–drug interactions. *Pharmacogenomics* 19, 1133–1138. doi: 10.2217/pgs-2018-0096
- Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y., and Lauschke, V. M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* 12, 26. doi: 10.1186/s40246-018-0157-3
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., et al. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* 99, 877–885. doi: 10.1016/j.ajhg.2016.08.016
- Ionita-Laza, I., McCallum, K., Xu, B., and Buxbaum, J. D. (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Rev. Clin. Oncol.* 48, 214–220. doi: 10.1038/ng.3477
- Ji, Y., Si, Y., McMillin, G. A., and Lyon, E. (2018). Clinical pharmacogenomics testing in the era of next generation sequencing: challenges and opportunities for precision medicine. *Expert Rev. Mol. Diagn.* 18, 411–421. doi: 10.1080/14737159.2018.1461561
- Jian, X., Boerwinkle, E., and Liu, X. (2014a). In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genetics in Medicine* 16, 497–503. doi: 10.1038/gim.2013.176
- Jian, X., Boerwinkle, E., and Liu, X. (2014b). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* 42, 13534–13544. doi: 10.1093/nar/gku1206
- Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502. doi: 10.1126/science.1141319
- Jukić, M. M., Haslemo, T., Molden, E., and Ingelman-Sundberg, M. (2018). Impact of CYP2C19 genotype on escitalopram exposure and therapeutic failure: a retrospective study based on 2,087 patients. *Am. J. Psychiatry* 175, 463–470. doi: 10.1176/appi.ajp.2017.17050550
- Kalinin, A. A., Higgins, G. A., Reamaroon, N., Soroushmehr, S., Allyn-Feuer, A., Dinov, I. D., et al. (2018). Deep learning in pharmacogenomics: from gene regulation to patient stratification. *Pharmacogenomics* 19, 629–650. doi: 10.2217/pgs-2018-0008
- Katoni, P., and Lichtarge, O. (2014). A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. *Genome Res.* 24, 2050–2058. doi: 10.1101/gr.176214.114
- Kerem, E., Hirawat, S., Armoni, S., Yaakov, Y., Shoseyov, D., Cohen, M., et al. (2008). Effectiveness of PTC124 treatment of cystic fibrosis caused by nonsense mutations: a prospective phase II trial. *Lancet* 372, 719–727. doi: 10.1016/S0140-6736(08)61168-X
- Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* 13, 700–712. doi: 10.1038/nrm.3454
- Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., R., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 46, D260–D266. doi: 10.1093/nar/gkx1188
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Rev. Clin. Oncol.* 46, 310–315. doi: 10.1038/ng.2892
- Kozyra, M., Ingelman-Sundberg, M., and Lauschke, V. M. (2017). 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. doi: 10.1038/gim.2016.33
- Kryukov, G. V., Pennacchio, L. A., and Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* 80, 727–739. doi: 10.1086/513473
- Kulshreshtha, S., Chaudhary, V., Goswami, G. K., and Mathur, N. (2016). Computational approaches for predicting mutant protein stability. *J. Comput. Aided Mol. Des.* 30, 401–412. doi: 10.1007/s10822-016-9914-3
- Lauschke, V. M., and Ingelman-Sundberg, M. (2016a). Precision medicine and rare genetic variants. *Trends Pharmacol. Sci.* 37, 85–86. doi: 10.1016/j.tips.2015.10.006
- Lauschke, V. M., and Ingelman-Sundberg, M. (2016b). Requirements for comprehensive pharmacogenetic genotyping platforms. *Pharmacogenomics* 17, 917–924. doi: 10.2217/pgs-2016-0023
- Lauschke, V. M., and Ingelman-Sundberg, M. (2018). How to consider rare genetic variants in personalized drug therapy. *Clin. Pharmacol. Therapeut.* 19, 20. doi: 10.1002/cpt.976
- Lauschke, V. M., Milani, L., and Ingelman-Sundberg, M. (2017). Pharmacogenomic biomarkers for improved drug therapy—recent progress and future developments. *AAPS J.* 20, 4. doi: 10.1208/s12248-017-0161-x
- Lee, D., Gorkin, D. U., Baker, M., Strober, B. J., Asoni, A. L., McCallion, A. S., et al. (2015). A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* 47, 955–961. doi: 10.1038/ng.3331
- Lee, M., Roos, P., Sharma, N., Atalar, M., Evans, T. A., Pellicore, M. J., et al. (2017). Systematic computational identification of variants that activate exonic and intronic cryptic splice sites. *Am. J. Hum. Genet.* 100, 751–765. doi: 10.1016/j.ajhg.2017.04.001
- Lee, W., Zhang, Y., Mukhyala, K., Lazarus, R. A., and Zhang, Z. (2009). Bi-directional SIFT predicts a subset of activating mutations. *PLoS ONE* 4:e8311. doi: 10.1371/journal.pone.0008311
- Lee, Y., and Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323. doi: 10.1146/annurev-biochem-060614-034316

- Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-022413
- Li, J., Zhao, T., Zhang, Y., Zhang, K., Shi, L., Chen, Y., et al. (2018). Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res.* 46, 7793–7804. doi: 10.1093/nar/gky678
- Linde, L., Boelz, S., Nissim-Rafinia, M., Oren, Y. S., Wilschanski, M., Yaacov, Y., et al. (2007). Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin. *J. Clin. Invest.* 117, 683–692. doi: 10.1172/JCI28523
- Liu, C., Zhang, F., Li, T., Lu, M., Wang, L., Yue, W., et al. (2012). MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics* 13:661. doi: 10.1186/1471-2164-13-661
- Lorenz, R., Wolfinger, M. T., Tanzer, A., and Hofacker, I. L. (2016). Predicting RNA secondary structures from sequence and probing data. *Methods* 103, 86–98. doi: 10.1016/j.ymeth.2016.04.004
- Lu, Q., Hu, Y., Sun, J., Cheng, Y., Cheung, K.-H., and Zhao, H. (2015). A statistical framework to predict functional non-coding regions in the human genome through integrated analysis of annotation data. *Sci. Rep.* 5:10576. doi: 10.1038/srep10576
- MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–828. doi: 10.1126/science.1215040
- Mahmood, K., Jung, C.-H., Philip, G., Georgeson, P., Chung, J., Pope, B. J., et al. (2017). Variant effect prediction tools assessed using independent, functional assay-based datasets: implications for discovery and diagnostics. *Hum. Genomics* 11: 10. doi: 10.1186/s40246-017-0104-8
- Martelotto, L. G., Ng, C. K., De Filippo, M. R., Zhang, Y., Piscuoglio, S., Lim, R. S., et al. (2014). Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations. *Genome Biol.* 15: 484. doi: 10.1186/s13059-014-0484-1
- Miosge, L. A., Field, M. A., Sontani, Y., Cho, V., Johnson, S., Palkova, A., et al. (2015). Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5189–E5198. doi: 10.1073/pnas.1511585112
- Mort, M., Sterne-Weiler, T., Li, B., Ball, E. V., Cooper, D. N., Radivojac, P., et al. (2014). MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol.* 15:R19. doi: 10.1186/gb-2014-15-1-r19
- Moszynska, A., Gebert, M., Collawn, J. F., and Bartoszewski, R. (2017). SNPs in microRNA target sites and their potential role in human disease. *Open Biol.* 7:170019. doi: 10.1098/rsob.170019
- Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res.* 11, 863–874. doi: 10.1101/gr.176601
- Ng, P. C., and Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* 7, 61–80. doi: 10.1146/annurev.genom.7.080505.115630
- Ohno, K., Takeda, J.-I., and Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip. Rev.* 9:e1451. doi: 10.1002/wrna.1451
- Pan, Y., Liu, D., and Deng, L. (2017). Accurate prediction of functional effects for variants by combining gradient tree boosting with optimal neighborhood properties. *PLoS ONE* 12:e0179314. doi: 10.1371/journal.pone.0179314
- Pandurangan, A. P., Ochoa-Montaña, B., Ascher, D. B., and Blundell, T. L. (2017). SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 45, W229–W235. doi: 10.1093/nar/gkx439
- Perte, M., Lin, X., and Salzberg, S. L. (2001). GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* 29, 1185–1190. doi: 10.1093/nar/29.5.1185
- Peterson, T. A., Doughty, E., and Kann, M. G. (2013). Towards precision medicine: advances in computational approaches for the analysis of human variants. *J. Mol. Biol.* 425, 4047–4063. doi: 10.1016/j.jmb.2013.08.008
- Pinzón, N., Li, B., Martínez, L., Sergeeva, A., Presumey, J., Apparailly, F., et al. (2017). microRNA target prediction programs predict many false positives. *Genome Res.* 27, 234–245. doi: 10.1101/gr.205146.116
- Pires, D. E., Ascher, D. B., and Blundell, T. L. (2014a). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* 42, W314–W319. doi: 10.1093/nar/gku411
- Pires, D. E., Ascher, D. B., and Blundell, T. L. (2014b). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30, 335–342. doi: 10.1093/bioinformatics/btt691
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121. doi: 10.1101/gr.097857.109
- Pucci, F., Bernaerts, K., Kwasigroch, J. M., and Rooman, M. (2018). Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* 13: 3031. doi: 10.1093/bioinformatics/bty348
- Quan, L., Lv, Q., and Zhang, Y. (2016). STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics* 32, 2936–2946. doi: 10.1093/bioinformatics/btw361
- Quang, D., Chen, Y., and Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 31, 761–763. doi: 10.1093/bioinformatics/btu703
- Raida, M., Schwabe, W., Häusler, P., Van Kuilenburg, A. B., Van Gennip, A. H., Behnke, D., et al. (2001). Prevalence of a common point mutation in the Dihydropyrimidine dehydrogenase (DPD) gene within the 5'-splice donor site of intron 14 in patients with severe 5-fluorouracil (5-FU)-related toxicity compared with controls. *Clin. Cancer Res.* 7, 2832–2839.
- Raimondi, D., Tanyalcin, I., Ferte, J., Gazzo, A., Orlando, G., Lenaerts, T., et al. (2017). DEOGEN2: prediction and interactive visualization of single amino acid variant deleteriousness in human proteins. *Nucleic Acids Res.* 45, W201–W206. doi: 10.1093/nar/gkx390
- Rausell, A., Mohammadi, P., McLaren, P. J., Bartha, I., Xenarios, I., Fellay, J., et al. (2014). Analysis of stop-gain and frameshift variants in human innate immunity genes. *PLoS Comput. Biol.* 10:e1003757. doi: 10.1371/journal.pcbi.1003757
- Reese, M. G., Eeckman, F. H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311–323. doi: 10.1089/cmb.1997.4.311
- Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39: e118. doi: 10.1093/nar/gkr407
- Rieger, J. K., Klein, K., Winter, S., and Zanger, U. M. (2013). Expression variability of absorption, distribution, metabolism, excretion-related microRNAs in human liver: influence of nongenetic factors and association with gene expression. *Drug Metab. Dispos.* 41, 1752–1762. doi: 10.1124/dmd.113.052126
- Ritchie, G. R., Dunham, I., Zeggini, E., and Flicek, P. (2014). Functional annotation of noncoding sequence variants. *Nat. Methods* 11, 294–296. doi: 10.1038/nmeth.2832
- Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., A. Heravi-Moussavi. (2015). P, Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. doi: 10.1038/nature14248
- Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., and Campbell, C. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* 34, 511–513. doi: 10.1093/bioinformatics/btx536
- Ryan, B. C., Werner, T. S., Howard, P. L., and Chow, R. L. (2016). ImiRP: a computational approach to microRNA target site mutation. *BMC Bioinformatics* 17:190. doi: 10.1186/s12859-016-1057-y
- Sample, P. J., Wang, B., Reid, D. W., Presnyak, V., McFadyen, I., Morris, D. R., et al. (2018). Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *bioRxiv*. doi: 10.1101/310375
- Sato, K., Hamada, M., Asai, K., and Mituyama, T. (2009). CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic Acids Res.* 37, W277–W280. doi: 10.1093/nar/gkp367
- Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328, 1036–1040. doi: 10.1126/science.1186176
- Schoenberg, D. R., and Maquat, L. E. (2012). Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* 13, 246–259. doi: 10.1038/nrg3160

- Schwarz, J. M., Cooper, D. N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* 11, 361–362. doi: 10.1038/nmeth.2890
- Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670. doi: 10.1038/nrm.2017.86
- Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., et al. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* 34, 57–65. doi: 10.1002/humu.22225
- Shihab, H. A., Rogers, M. F., Gough, J., Mort, M., Cooper, D. N., Day, I. N., et al. (2015). An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* 31, 1536–1543. doi: 10.1093/bioinformatics/btv009
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050. doi: 10.1101/gr.3715005
- Sim, S. C., Kacevska, M., and Ingelman-Sundberg, M. (2013). Pharmacogenomics of drug-metabolizing enzymes: a recent update on clinical implications and endogenous effects. *Pharmacogenomics J.* 13, 1–11. doi: 10.1038/tpj.2012.45
- Smedley, D., Schubach, M., Jacobsen, J. O. B., Köhler, S., Zemojtel, T., Spielmann, M., et al. (2016). A whole-genome analysis framework for effective identification of pathogenic regulatory variants in mendelian disease. *Am. J. Hum. Genet.* 99, 595–606. doi: 10.1016/j.ajhg.2016.07.005
- Somody, J. C., MacKinnon, S. S., and Windemuth, A. (2017). Structural coverage of the proteome for pharmaceutical applications. *Drug Discov. Today* 22, 1792–1799. doi: 10.1016/j.drudis.2017.08.004
- Soukari, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., et al. (2016). Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using *in silico* tools. *PLoS Genet.* 12:e1005756. doi: 10.1371/journal.pgen.1005756
- Spear, B. B., Heath-Chiozzi, M., and Huff, J. (2001). Clinical application of pharmacogenetics. *Trends Mol. Med.* 7, 201–204. doi: 10.1016/S1471-4914(01)01986-4
- Stone, E. A., and Sidow, A. (2005). Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* 15, 978–986. doi: 10.1101/gr.3804205
- Tang, H., and Thomas, P. D. (2016). Tools for predicting the functional impact of nonsynonymous genetic variation. *Genetics* 203, 635–647. doi: 10.1534/genetics.116.190033
- The 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141. doi: 10.1101/gr.772403
- Topham, C. M., Srinivasan, N., and Blundell, T. L. (1997). Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng.* 10, 7–21. doi: 10.1093/protein/10.1.7
- Venselaar, H. Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L., and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 11:548. doi: 10.1186/1471-2105-11-548
- Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709. doi: 10.1038/nature12946
- Warf, M. B., and Berglund, J. A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* 35, 169–178. doi: 10.1016/j.tibs.2009.10.004
- Witvliet, D. K., Stokach, A., Giraldo-Forero, A. F., Teyra, J., Colak, R., and Kim, P. M. (2016). ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity. *Bioinformatics* 32, 1589–1591. doi: 10.1093/bioinformatics/btw031
- Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* 275, 1–21. doi: 10.1111/j.1742-4658.2007.06178.x
- Woolfe, A., Mullikin, J. C., and Elnitski, L. (2010). Genomic features defining exonic variants that modulate splicing. *Genome Biol.* 11:R20. doi: 10.1186/gb-2010-11-2-r20
- Xiao, Y., Wigneshweraraj, S. R., Weinzierl, R., Wang, Y.-P., and Buck, M. (2009). Construction and functional analyses of a comprehensive sigma54 site-directed mutant library using alanine-cysteine mutagenesis. *Nucleic Acids Res.* 37, 4482–4497. doi: 10.1093/nar/gkp419
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D. R., Yuen, R. K., et al. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 1254806. doi: 10.1126/science.1254806
- Yates, C. M., Filippis, I., Kelley, L. A., and Sternberg, M. J. (2014). SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J. Mol. Biol.* 426, 2692–2701. doi: 10.1016/j.jmb.2014.04.026
- Yeo, G., and Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* 11, 377–394. doi: 10.1089/1066527041410418
- Yeo, G., Holste, D., Kreiman, G., and Burge, C. B. (2004). Variation in alternative splicing across human tissues. *Genome Biol.* 5, R74–R15. doi: 10.1186/gb-2004-5-10-r74
- Yue, P., Melamud, E., and Moul, J. (2006). SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* 7:166. doi: 10.1186/1471-2105-7-166
- Zhang, F., and Lupski, J. R. (2015). Non-coding genetic variants in human disease. *Hum. Mol. Genet.* 24, R102–R110. doi: 10.1093/hmg/ddv259
- Zhang, Q., Fan, X., Wang, Y., Sun, M.-A., Shao, J., and Guo, D. (2017). BPP: a sequence-based algorithm for branch point prediction. *Bioinformatics* 33, 3166–3172. doi: 10.1093/bioinformatics/btx401
- Zhang, X., Lin, H., Zhao, H., Hao, Y., Mort, M., Cooper, D. N., et al. (2014). Impact of human pathogenic micro-insertions and micro-deletions on post-transcriptional regulation. *Hum. Mol. Genet.* 23, 3024–3034. doi: 10.1093/hmg/ddu019
- Zhang, Y., Lv, J., Liu, H., Zhu, J., Su, J., Wu, Q., et al. (2010). HHMD: the human histone modification database. *Nucleic Acids Res.* 38, D149–D154. doi: 10.1093/nar/gkp968
- Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* 12, 931–934. doi: 10.1038/nmeth.3547
- Zhou, Y., Ingelman-Sundberg, M., and Lauschke, V. M. (2017). Worldwide distribution of cytochrome P450 Alleles: a meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Therapeut.* 102, 688–700. doi: 10.1002/cpt.690
- Zhou, Y., and Lauschke, V. M. (2018). Comprehensive overview of the pharmacogenetic diversity in Ashkenazi Jews. *J. Med. Genet.* 55, 617–627. doi: 10.1136/jmedgenet-2018-105429
- Zhou, Y., Mkrtchian, S., Kumondai, M., Hiratsuka, M., and Lauschke, V. M. (2018). An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* 28: 1. doi: 10.1038/s41397-018-0044-2
- Zia, A., and Moses, A. M. (2011). Ranking insertion, deletion and nonsense mutations based on their effect on genetic information. *BMC Bioinformatics* 12:299. doi: 10.1186/1471-2105-12-299

Conflict of Interest Statement: VL is co-founder and owner of HepaPredict AB.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhou, Fujikura, Mkrtchian and Lauschke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



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

OPEN ACCESS

Edited by:

Rick Kittles,

Irell & Manella Graduate School
of Biological Sciences, United States

Reviewed by:

Jeannine S. McCune,
Beckman Research Institute,
United States

Wendy Hernandez,
The University of Chicago,
United States

Jason Hansen Karnes,
The University of Arizona,
United States

*Correspondence:

Ulrich M. Zanger
uli.zanger@ikp-stuttgart.de

Specialty section:

This article was submitted to
Pharmacogenetics
and Pharmacogenomics,
a section of the journal
Frontiers in Genetics

Received: 26 October 2018

Accepted: 09 January 2019

Published: 31 January 2019

Citation:

Klein K, Tremmel R, Winter S,
Fehr S, Battke F, Scheurenbrand T,
Schaeffeler E, Biskup S, Schwab M
and Zanger UM (2019) 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. 10:7.
doi: 10.3389/fgene.2019.00007

Kathrin Klein^{1,2}, Roman Tremmel^{1,2}, Stefan Winter^{1,2}, Sarah Fehr^{3,4}, Florian Battke^{3,4},
Tim Scheurenbrand^{3,4}, Elke Schaeffeler^{1,2}, Saskia Biskup^{3,4}, Matthias Schwab^{1,2,5,6} and
Ulrich M. Zanger^{1,2*}

¹ Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany, ² Medical School, University of
Tübingen, Tübingen, Germany, ³ CeGaT GmbH, Tübingen, Germany, ⁴ Praxis für Humangenetik Tübingen, Tübingen,
Germany, ⁵ Department of Clinical Pharmacology, University Hospital Tübingen, Tübingen, Germany, ⁶ Department
of Pharmacy and Biochemistry, University of Tübingen, Tübingen, Germany

We developed a panel-based NGS pipeline for comprehensive analysis of 340 genes involved in absorption, distribution, metabolism and excretion (ADME) of drugs, other xenobiotics, and endogenous substances. The 340 genes comprised phase I and II enzymes, drug transporters and regulator/modifier genes within their entire coding regions, adjacent intron regions and 5' and 3'UTR regions, resulting in a total panel size of 1,382 kbp. We applied the ADME NGS panel to sequence genomic DNA from 150 Caucasian liver donors with available comprehensive gene expression data. This revealed an average read-depth of 343 (range 27–811), while 99% of the 340 genes were covered on average at least 100-fold. Direct comparison of variant annotation with 363 available genotypes determined independently by other methods revealed an overall accuracy of >99%. Of 15,727 SNV and small INDEL variants, 12,022 had a minor allele frequency (MAF) below 2%, including 8,937 singletons. In total we found 7,273 novel variants. Functional predictions were computed for coding variants ($n = 4,017$) by three algorithms (Polyphen 2, Provean, and SIFT), resulting in 1,466 variants (36.5%) concordantly predicted to be damaging, while 1,019 variants (25.4%) were predicted to be tolerable. In agreement with other studies we found that less common variants were enriched for deleterious variants. *Cis*-eQTL analysis of variants with ($MAF \geq 2\%$) revealed significant associations for 90 variants in 31 genes after Bonferroni correction, most of which were located in non-coding regions. For less common variants ($MAF < 2\%$), we applied the SKAT-O test and identified significant associations to gene expression for *ADH1C* and *GSTO1*. Moreover, our data allow comparison of functional predictions with additional phenotypic data to prioritize variants for further analysis.

Keywords: ADME, next generation sequencing, pharmacogenomics, eQTL analysis, rare variants

INTRODUCTION

Genetic variation in genes that function in the absorption, distribution, metabolism, and elimination (ADME) of drugs contributes significantly to the interindividual variability in efficacy and toxicity of numerous drugs from practically all therapeutic categories. In the past half century, pharmacogenetic research has unraveled many clinically meaningful associations between germline genetic variants and pharmacokinetic or drug response phenotypes (Meyer, 2004; Zanger and Schwab, 2013; Alfrevic and Pirmohamed, 2017). Clinical implementation of this knowledge is currently being pursued worldwide by several consortia (Caudle et al., 2013; Dunnenberger et al., 2015; Relling and Evans, 2015; Cecchin et al., 2017; Swen et al., 2018). For example, the Clinical Pharmacogenetics Implementation Consortium (CPIC) has so far issued 65 dosing guidelines for 38 drugs and 15 relevant genes (October 2018¹). Until recently, pharmacogenetics has mainly focused on common genetic variants, which can be relatively easily assessed for association with pharmacokinetic or drug response phenotypes. However, a considerable proportion of genetic variability remains unexplained even for well-studied genes like CYP2D6, as recently shown by twin studies (Matthaei et al., 2015). Currently, it is widely assumed that rare deleterious variants fill this gap and contribute significantly to functional variability, which is further supported by the fact that rare variants are enriched for deleterious alleles due to purifying selection (1000 Genomes Project Consortium et al., 2012; Lek et al., 2016; Ingelman-Sundberg et al., 2018). Indeed, with the increasing availability of next-generation-sequencing (NGS) technology, several studies explored genetic variability of pharmacologically relevant “pharmacogenes” and revealed large numbers of rare variants, most of which were previously unknown (Tennesen et al., 2012; Fujikura et al., 2015; Han et al., 2016; Kozyra et al., 2016; Hovelson et al., 2017; Schärfe et al., 2017). For statistical reasons it is intrinsically more difficult to investigate the functional significance of rare variants as compared to common variants, especially regarding pharmacogenetic phenotypes, for which studies including relevant phenotypic data are essentially lacking. On the other hand, *in vitro* testing of thousands of variants is currently prohibitive for time and financial reasons. Current hopes to integrate rare variants into clinical pharmacogenomics therefore rely mainly on computational prediction tools, many of which are publically available (Ingelman-Sundberg et al., 2018; Zhou et al., 2018a). Computational predictions of “damaging” or “loss-of-function” (LOF) versus “tolerable” (TOL) functionality performed on ADME rare variants detected in genetic screens indicated that up to 30% of drug response

variability could be due to rare variants and that likely every patient carries at least one “actionable” pharmacogenetic variant (Crosslin et al., 2015; Ji et al., 2016). However, data on the validity of functional prediction are scarce and their performance as well as the true contribution of rare variants to pharmacogenetics variability remains unclear, especially since current predictive algorithms rely largely on principles of evolutionary conservation, which may be more appropriate in the context of disease than for drug metabolism and response.

In this study we have developed a panel-based NGS pipeline for comprehensive sequence analysis of 340 ADME genes comprising all major genes known to be involved in phase 1 and phase 2 drug metabolism, drug transport and its regulation, as well as numerous additional genes of potential interest in this context. We applied our ADME NGS panel on genomic DNA from 150 human liver samples that we have previously genotyped by other methods and for which comprehensive mRNA expression data and some additional ADME phenotypes are available. This allowed us to directly compare genotype with expression for common and rare variants, unraveling numerous novel associations and potential candidates. In addition, we performed functional prediction for subsets of variants and exemplarily compared these with hepatic phenotype. This type of analysis, which has rarely been done, should be helpful to improve functional prediction and allow to prioritization of interesting rare variants for further analysis.

MATERIALS AND METHODS

Patient DNA and Liver Samples

Liver tissues and corresponding blood samples were previously collected from patients of White European descent undergoing liver surgery at the Department of General, Visceral, and Transplantation Surgery (A. K. Nuessler, P. Neuhaus, Campus Virchow, University Medical Center Charité, Humboldt University Berlin, Germany) (Klein et al., 2012). The study protocol was approved by the ethics committees of the medical faculties of the Charité, Humboldt University, and the University of Tübingen. The study was conducted in accordance with the Declaration of Helsinki, and written informed consent was obtained from each patient. Only non-tumorous tissue was collected, as confirmed by histological examination, and stored at -80°C . Available patient documentation includes sex, age, smoking habits, alcohol consumption, presurgery medication, diagnosis leading to liver resection, and serological liver function parameters. Samples from patients with hepatitis, cirrhosis, or chronic alcohol abuse were excluded. A summary of the data is presented in **Supplementary Table S1**.

Phenotypic data were available from previous studies. Genome-wide mRNA expression profiling was previously performed using Illumina Human-WG6v2 Expression BeadChip (see below). For selected genes quantitative mRNA levels were determined by real-time PCR, protein levels by Western blot, and enzyme activity levels by mass spectrometry (**Supplementary Table S2**).

Abbreviations: ADME, Absorption Distribution Metabolism Excretion; bp, basepair; CNV, copy number variant; eQTL, expression quantitative trait loci; HWE, Hardy-Weinberg equilibrium; INDEL, insertion/deletion; Kbp, kilo basepair; LOF, loss of function; MAF, minor allele frequency; NGS, next generation sequencing; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; TOL, tolerated; UTR, untranslated region.

¹ www.pharmgkb.org/guidelines

Genomic DNA was isolated from corresponding blood samples as described previously (Gomes et al., 2009). Quality and concentration of gDNA were determined using both, the Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Dreieich, Germany) and Nanodrop ND-8000 (Thermo Fisher Scientific, Dreieich, Germany). Gene expression and genotyping data assessed by Human-WG6v2 Expression BeadChip and HumanHap300 Genotyping BeadChip (Illumina, Eindhoven, Netherlands) were preprocessed as previously described (Schröder et al., 2013) and the data are accessible through GEO Series accession numbers GSE32504 and GSE39036, respectively.

Targeted ADME NGS Panel Sequencing

Genomic DNA was enriched using a custom design Agilent SureSelect XT in-solution kit (Agilent Technologies, Santa Clara, CA, United States). The design of the PGX panel for all relevant ADME classified and ADME related genes (340 genes in total) included publically available gene lists of PharmaADME.org² (CORE/EXTEND, $n = 236$), pharmGKB³ (Whirl-Carrillo et al., 2012); [very important pharmacogenes (VIP), $n = 36$], as well as additional genes with confirmed or putative ADME-related function according to literature search ($n = 104$; **Supplementary Table S2**). For analysis, the genes were assorted into functional groups as follows: ATP-binding cassette transporters (ABC; $n = 45$), solute carrier transporters, solute carrier organic anion transporters, and ion channels (SLC/SLCO; $n = 64$), members of phase I metabolism excluding cytochrome P450 and other modifying enzymes (Phase1; $n = 36$), members of phase II metabolism (Phase 2; $n = 53$), cytochrome P450s/modifying enzymes (CYP/modifiers; $n = 53$), nuclear receptors/transcription regulators (NR/TR; $n = 46$), and genes of other background and potentially related to ADME (others; $n = 43$) (**Figure 1B** and **Supplementary Table S2**). Positions of exon regions, 3' and 5' UTR (untranslated regions) were based on RefSeq major transcripts sequences (GRCh37; hg19; UCSC genome browser). Exon sizes were extended by 20 nucleotides on each side. Sequence of very short exons was symmetrically increased to at least 160 nucleotides. For selected genes 5' regions were extended to cover 2 kbp ($n = 29$). The total number of exons was 4,210 and total target size reached 1,382 kbp (**Supplementary Table S2**). Panel details are available on demand.

Target capturing was specifically designed for NGS of selected regions and DNA libraries were generated using Agilent in-solution target capture technology from up to 1 µg high quality genomic DNA for each sample. NGS was carried out on the Illumina HiSeq2500 system (Illumina Inc., San Diego, CA, United States) at high depth with 2×100 bps paired-end reads. Raw sequencing reads generated by the Illumina platform were demultiplexed using Illumina bcl2fastq (1.8.2) (Illumina, San Diego, CA, United States). Adapter sequences were removed with cutadapt and the trimmed reads mapped to the human reference genome (GRCh37 hg19) using the Burrows Wheeler Aligner (BWA-mem 0.7.2; Li and Durbin, 2010). Reads mapping to more

than one location with identical mapping scores were discarded (in house software). Read duplicates likely resulting from PCR amplification were removed (samtools 0.1.18). Variants were called using samtools and varscan (2.3.5)⁴. Technical artifacts were removed (in-house software) and the remaining variants were annotated based on several internal and external databases. We created a read count matrix for sequenced targets and 150 samples using the R package cn.mops.1.12.0 and the BAM files to assess the quality of coverage per gene and per target region. Approximately 5.9 million on target reads were generated per sample with a mean mapping quality of 58.2 and a mean coverage of 343 per target site. A Frequentist or a Bayesian algorithm was applied to call SNVs and small insertions/deletions (INDELs). Detection of insertions is limited by read length and no insertions above 50 bp were observed. Variant annotations were retrieved from UCSC genome data browser⁵, dbSNP build151 (March 22, 2018), and Sequence Ontology (SO) terms to describe the effect of each variant on genes in terms of transcript structure. Enrichment and sequencing procedure were established, validated, and provided by CeGaT GmbH, Tübingen, Germany. CeGaT is accredited by DAkkS according to DIN EN ISO 15189:2014, by the College of American Pathologists (CAP) and CLIA-certified (Dohrn et al., 2017). Sequence variant data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003426. Further information about EGA can be found on <https://ega-archive.org> (Lappalainen et al., 2015).

High Quality Variants

Only variants within the predefined target regions were selected and further analyzed ($n = 16,928$). Variant calls with sequencing coverage below $20\times$ were regarded as invalid. Moreover, heterozygous calls were regarded as invalid when variant allele ratios were $<5\%$. Invariant positions and variants with less than 70% valid values in all samples were excluded. Furthermore, 696 variants with HWE p -values $< 10^{-5}$ were considered suspicious and consequently excluded from all subsequent analyses. Finally, 13,838 SNVs and 1,889 INDELs were further investigated in this work. Genedata Profiler Analyst Module (V12.0.2; Genedata AG, Basel, Switzerland) and GraphPad Prism (V5.04; GraphPad Software Inc., La Jolla, CA, United States) were used for data filtering, visualization, and basic statistical calculations.

Global Validation

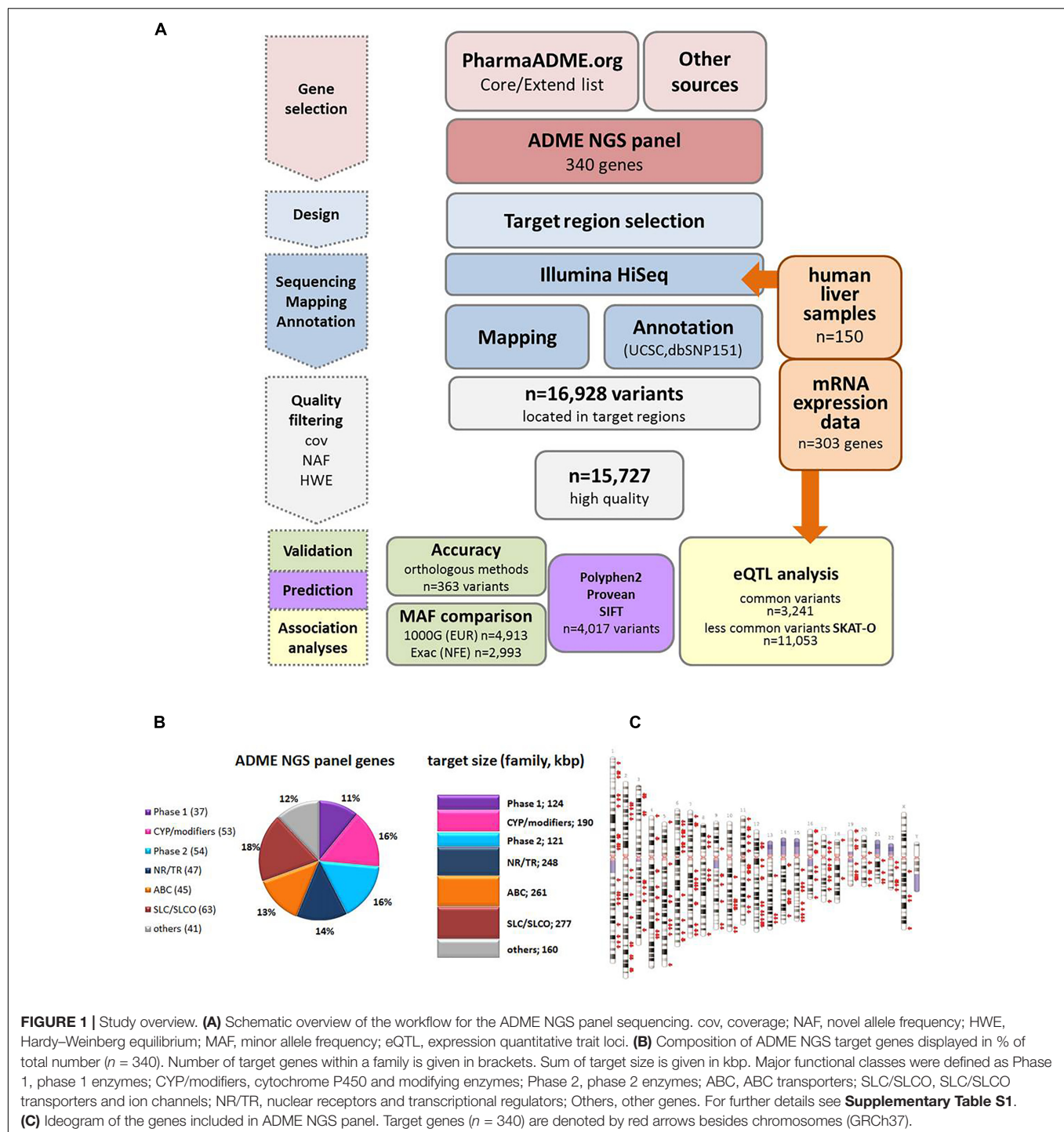
Evaluation of ADME panel sequencing data was performed by direct comparison of sample genotypes to available genome wide SNP data (Illumina HumanHAP300 SNP; GEO Series accession number GSE39036; Schröder et al., 2013) as well as genotype data of 87 individual SNVs determinations obtained with several other genotyping methods in former studies (RFLP, Sanger sequencing, TaqMan allelic discrimination, MALDI-TOF, and other arrays) from the same sample set. Array variant data

²<http://pharmaadme.org>

³<https://www.pharmgkb.org/>

⁴<http://dkoboldt.github.io/varscan>

⁵<https://genome.ucsc.edu/cgi-bin/hgVai>



were “lifted” to GRCh37 (hg19), and only SNVs within the target regions defined above and with HWE p -value $> 10^{-5}$ were extracted ($n = 276$). Finally, genotype data for 363 variants were available for validation. Concordance of genotype data from ADME NGS and results from orthogonal methods was evaluated by computing percentage of identical genotype calls over all variants and samples. Variant positions within the above defined target boundaries were extracted from publically

available databases from the Exome Aggregation Consortium ExAC⁶ (Lek et al., 2016) and 1000 Genomes project⁷ (1000 Genomes Project Consortium et al., 2015). In total, 11,558 and 68,918 variants were retrieved in the demanded genomic regions from 1000G and ExAC, respectively. Chromosomal

⁶<http://exac.broadinstitute.org/>

⁷<http://www.internationalgenome.org/>

position and nucleotide change (reference/alternative) were used to identify corresponding variants in the ADME NGS panel data. After adjusting frequency data to MAF numbers ranging between 0 and 50%, MAF from European (EUR, 1000G) or non-Finnish European (NFE, ExAC) were compared to observed MAF from our cohort. In addition, several well known variants in *CYP2D6*, *CYP2C9*, *CYP2C19* and *CYP2B6*, *NAT2* and *DPYD* were confirmed by Sanger sequencing. A concordance of 100% was observed covering 57 SNVs in 19 samples.

In silico Prediction

The impact of coding variants on protein function was predicted using Polyphen 2 (PP2⁸ Adzhubei et al., 2013) as well as the Proven Human Genome Variants tool [Protein Variation Effect Analyzer (PROVEAN)⁹; Choi et al., 2012], providing Proven and in addition SIFT (Sorting Intolerant from Tolerant; Sim et al., 2012) scores. All algorithms are based, among other features, on sequence conservation and were used with default settings. For a total of 4,017 coding variants including missense ($n = 3,893$), frameshift ($n = 37$), initiator codon ($n = 7$), stop codon ($n = 46$) and other coding variants (Table 1), prediction was performed using chromosomal genomic positions, reference and variant nucleotide. Functional predictions of the type LOF versus tolerated (TOL) was retrieved from Proven (cutoff 2.5; deleterious/neutral), SIFT (cutoff 0.05; damaging/tolerated) and Polyphen2 (probably and possibly damaging/benign). It must be pointed out that frameshift variants ($n = 37$) as well as mutations of stop codons (gain/loss; $n = 46$) are not predictable by these tools.

Cis-eQTL Analyses

Cis-eQTL analysis between the 15,727 variants (13,838 SNVs and 1,889 INDELs) and their corresponding gene were performed

with statistical software R-3.5.0 (R Core Team, 2018) and additional packages SNPassoc (v1.9-2; González et al., 2014), SKAT (v1.3.2.1; Lee, 2017), and illuminaHumanv2.db (v1.26.0; Dunning et al., 2015).

mRNA expression levels were assessed by Human-WG6v2 Expression BeadChip (Illumina, Eindhoven, Netherlands) and preprocessed as described (Schröder et al., 2013). Probe sets were re-annotated using the R package illuminaHumanv2.db (Dunning et al., 2015). Only probe sets with “good” or “perfect” probe quality as defined by illuminaHumanv2fullReannotation were considered for the eQTL analyses. Of the 340 ADME and ADME related genes described above, 303 genes (89%) were represented on the Human-WG6v2 Expression BeadChip with at least one “good” or “perfect” probe set. If several “good” or “perfect” probe sets were annotated to a gene, data of these entire probe sets (i.e., log2 normalized expression signals) were averaged, finally resulting in an expression matrix of size 303 genes \times 150 samples for the eQTL analyses. Of the 15,727 variants, 14,294 (90.9%) were annotated to one of the 303 genes.

For individual eQTL analyses, only variants with MAF $\geq 2\%$ and annotated to one of the 303 genes ($n = 3,241$) were considered, in order to avoid testing variants with very few minor allele carriers (a MAF $\geq 2\%$ in 150 patients corresponds to at least 3 minor allele carriers; in our dataset, all variants with MAF $\geq 2\%$ actually comprised at least 4 minor allele carriers). For 8 of the 303 genes, only variants with MAF $< 2\%$ were annotated in the ADME NGS panel (*ABCB9*, *ALDH2*, *CYP11A1*, *GSTK1*, *GSTM1*, *GSTT1*, *PRMT1*, and *SULT1A4*), leaving 295 genes and 3,241 variants for individual cis-eQTL analyses. These analyses were performed using the generalized linear model framework of R-package SNPassoc (González et al., 2014), considering four different genetic models: codominant, dominant, recessive, and additive. Only the minimal p -value of the four genetic models for each SNP was reported. Besides univariate analyses, cis-effects of variants on mRNA expression were analyzed controlling for 10 covariates [sex, age, smoking, alcohol consumption, diagnosis, C-reactive protein (CRP) level, cholestatic liver disease, presurgical medication (no drugs, P450

⁸<http://genetics.bwh.harvard.edu/pph2/>

⁹<http://proven.jcvi.org/index.php>

TABLE 1 | Structural classification of ADME panel variants ($n = 15,727$).

Coding ($n = 6,058$; 38.5%)			Non-coding ($n = 9,669$; 61.5%)		
Class ^a	Variant (known ^b)	Variant (novel ^b)	Class ^a	Variant (known ^b)	Variant (novel ^b)
Initiator_cod	3	4	Upstream	476	359
Missense	1,610	2,283	5'UTR	501	499
Stop_gained	22	22	Non-coding exon	95	34
Stop_lost	2		Intron	1,166	1,441
Synonymous	1,219	764	Splice	296	520
Inframe	29	28	3'UTR	2,922	1,261
Frameshift	19	18	Downstream	68	31
Other coding ^c	26	9			
Total	2,930 (48%)	3138 (52%)	Total	5,524 (57%)	4,145 (43%)

^aClassification nomenclature according to ENSEMBLE variation sequence ontology terms. ^bKnown/novel: with/without dbSNP database identifier. ^cIncluding: coding-exon-variant, stop-retained.

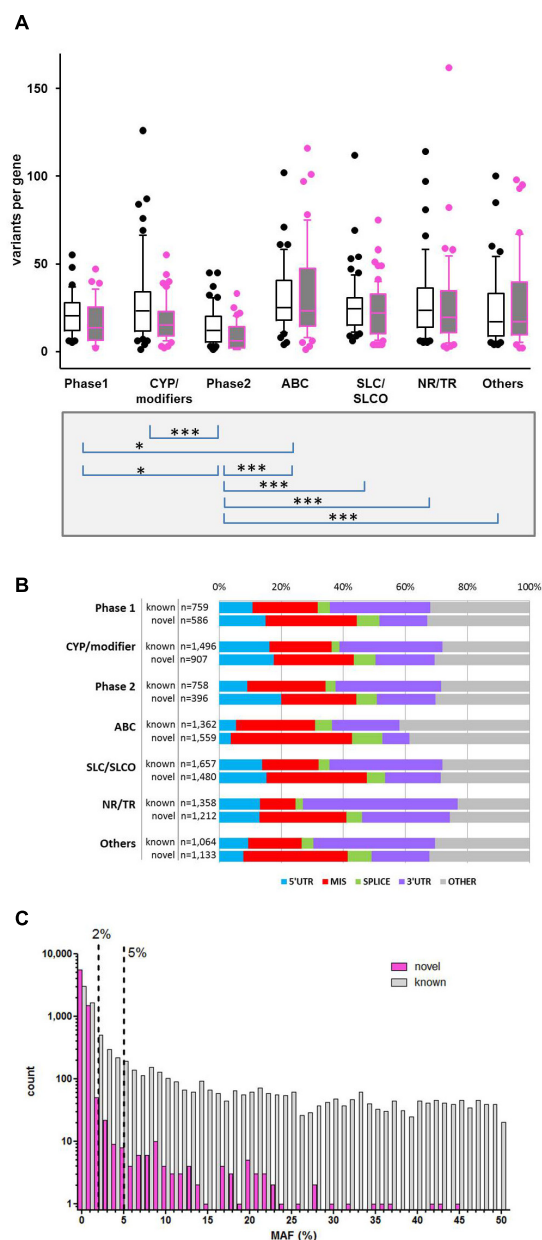


FIGURE 2 | Variability of gene families. **(A)** Distribution of known and novel variants in ADME gene families. The numbers of observed known and novel variants (including SNVs and INDELs) per gene are shown for the seven major functional classes of ADME genes defined in **Figure 1**. Open boxes, known variants; filled boxes, novel variants; boxes show median with 75th and 25th percentiles and whiskers represent 10th and 90th percentiles. Lower part: statistical significance calculated by Kruskal–Wallis with Dunn’s multiple comparison test of total number of variants per genes between family groups: * $P \leq 0.05$, *** $P \leq 0.001$. **(B)** Functional categorization of variants. Total number and proportion of variants observed in each functional class is shown separately for known and novel variants. Functional classes are defined as follows: 5’UTR, upstream and 5’ untranslated region; MIS, initiator codon, missense and stop codon variants; SPLICE, variants in consensus splice site acceptor and donor regions; 3’UTR, downstream and 3’ untranslated region; OTHER, other functional classes (intronic, frameshift, synonymous, other coding and non-coding variants). **(C)** Comparison of minor allele frequencies (Continued)

FIGURE 2 | Continued

(MAF) between novel and known observations. Total number of known observations with dbSNP identifier (open white bars; $n = 8,454$), novel observations (filled purple bars; $n = 7,273$); dotted line marks MAF = 2 and 5%.

inducer and other drugs), serum total bilirubin (TBILI) level, and serum gamma glutamyl transferase (GGT) level; see further details in **Supplementary Table S1**]. We used the Bonferroni method for multiple testing correction and set the significance level at $0.05/3,241 = 1.54E-05$.

Moreover, we performed combined *cis*-eQTL analyses of the rare variants ($MAF < 2\%$; $n = 11,053$) using the optimal unified association test framework for sets of variants (SKAT-O; Lee et al., 2012) implemented in R-package SKAT. To be more precise, for each of the 303 genes, the association of the set of all rare variants annotated to this gene and the corresponding mRNA expression data was investigated applying the SKAT-O test with standard weights. The same 10 covariates as in the eQTL analysis of common variants were used for an analogous multivariate SKAT-O analysis. For combined *cis*-eQTL analysis of rare variants, the Bonferroni-corrected significance level was set to $0.05/303 = 1.65E-04$.

RESULTS

Development and Performance of the Targeted ADME NGS Panel

Figure 1A gives an overview of the project workflow. The selection of genes was based on the PharmaADME.org gene lists “core” and “extend” and the PharmGKB VIP genes and was complemented with numerous additional genes of potential relation to drug metabolism (**Figure 1B**). All 340 genes finally included were targeted for all exons, exon/intron boundaries, as well as 5’ and 3’UTRs. An extended 5’ region of 2 kb was included for a group of 29 selected genes. The total panel size comprised 1,382 kbp distributed over all chromosomes except the Y chromosome (**Figure 1C** and **Supplementary Table S2**). In our cohort of 150 liver samples, the gene target regions were covered to a mean read-depth of $343\times$ (25th percentile = 265; 75th percentile = 398; **Supplementary Figure S1A**). More than 98% of the target regions were covered at more than $30\times$. The highest coverage was obtained for *UGT2B11* (average 811), while *GSTT2B* showed the lowest average coverage of 27. These discrepancies did not hinder our analysis and can be resolved in a further iteration of design. Overall, 99% of the genes were covered on average at least 100-fold. Direct comparison of variant annotation with 363 available genotypes determined independently by other methods revealed an overall concordance of $>99\%$ (**Supplementary Figure S2**). The accuracy obtained with data derived from the Illumina HumanHap300 genotyping platform (99.3%) was slightly lower compared to data from other genotyping methods (99.6%), which may be due to inaccurate genotype

TABLE 2 | eQTL analysis: Significant associations from multivariate regression models after Bonferroni correction (only minimal *p*-values of four genetic models used are listed).

Gene	Variant ^a	dbSNP151	Functional class	Minimal <i>p</i> -value ^b
ABCA5	17_67242245_G_A	rs12942867	3'UTR	1.20E-08 A
	17_67242551_AG_A	rs321469	3'UTR	1.20E-08 A
	17_67242756_G_A	rs1990248	3'UTR	9.00E-08 A
	17_67243289_A_T	rs15886	3'UTR	1.20E-08 A
	17_67260926_A_G	rs12449649	Synonymous	1.00E-08 A
	17_67267317_T_C	rs557491	Missense	1.10E-07 A
	17_67282332_T_C	rs1550828	Intron	1.00E-08 A
ABCC11	16_48250011_G_A	rs11863233	Intron	6.90E-11 A
	16_48250026_G_T	rs11863236	Missense	6.90E-11 A
	16_48250218_T_C	rs28654935	Intron	6.90E-11 A
	16_48256602_T_C	rs16945974	Synonymous	6.90E-11 A
	16_48265777_C_T	rs16945988	Missense	6.90E-11 A
	16_48269120_TAGAGATGCAA_T	rs398088092	Upstream	3.20E-10 C
	16_48269140_AAGAGATGCAA_A		Upstream	1.80E-09 A
	16_48269561_A_G	rs10521167	Upstream	6.90E-11 A
	16_48269918_T_C	rs16946006	Upstream	6.90E-11 A
	16_48270429_C_T	rs9926206	Upstream	6.90E-11 A
	16_48270508_T_C	rs9934833	Upstream	6.90E-11 A
	16_48270574_A_G	rs9932328	Upstream	6.90E-11 A
	7_150553605_C_T	rs10156191	Missense	8.50E-06 A
	7_150555915_A_G	rs10893	Synonymous	3.40E-08 R
AOC1	7_150557622_G_A	rs12179	Synonymous	6.50E-09 R
	7_150557665_C_G	rs1049793	Missense	6.50E-09 R
	7_150558366_C_T	rs12539	3'UTR	6.30E-09 D
ALDH6A1	14_74527190_A_G	rs8204	3'UTR	1.40E-05 C
ARNT	1_150783934_G_GCACA	rs71580328	3'UTR	1.90E-10 D
	1_150783934_G_GCACACA	rs71580328	3'UTR	5.80E-10 C
	1_150783985_T_C	rs11552229	3'UTR	3.10E-12 A
	1_150804401_G_GA	rs200891935	Intron	4.60E-10 C
	1_150808889_C_G	rs2228099	Synonymous	6.70E-12 A
ARSA	1_150850904_CA_C	rs10305645	Upstream	1.50E-11 D
	22_51062832_G_A	rs8142033	3'UTR	9.10E-15 A
	22_51063477_T_C	rs6151429	3'UTR	3.70E-18 A
	22_51064039_G_C	rs743616	Missense	6.40E-07 A
	22_51064416_T_C	rs2071421	Missense	7.30E-12 A
CAV1	7_116200587_C_T	rs1049337	3'UTR	2.50E-16 A
CYP2D6	22_42528382_C_G	rs1080985	Upstream	2.40E-06 D
CYP2R1	11_14900931_G_A	rs117913124	Synonymous	1.10E-05 D
CYP3A5	7_99245914_A_G	rs15524	3'UTR	2.10E-13 D
CYP4F11	19_16023318_C_G	rs61175303	3'UTR	3.10E-06 C
	19_16023378_G_A	rs58046343	3'UTR	3.10E-06 C
	19_16023619_T_C	rs58153611	3'UTR	3.10E-06 C
CYP4F12	19_15791132_T_A	rs2074568	Intron	1.80E-09 A
	19_15793235_T_C	rs2285888	Missense	8.30E-06 A
	19_15807884_A_G	rs593818	Missense	1.80E-06 A
EPHX2	8_27373923_T_C	rs4149243	Splice_region	2.20E-08 D
	8_27396208_G_A	rs4149253	Synonymous	5.80E-06 C
	8_27401964_A_C	rs1126452	Synonymous	3.60E-12 D
	8_27402074_A_G	rs1042032	3'UTR	3.60E-12 D
	8_27402132_T_C	rs1042064	3'UTR	1.20E-12 D
FMO4	1_171311003_A_C	rs1042772	3'UTR	1.70E-06 A
GPX4	19_1106477_G_C	rs8178977	Intron	1.80E-09 A
	19_1106615_T_C	rs713041	3'UTR	6.90E-06 A

(Continued)

TABLE 2 | Continued

Gene	Variant ^a	dbSNP151	Functional class	Minimal <i>p</i> -value ^b
GPX7	1_53074532_C_A	rs1047635	3'UTR	2.50E-23 C
GSR	8_30535660_C_A	rs3594	3'UTR	1.10E-15 A
	8_30536581_A_G	rs1138054	3'UTR	2.90E-07 A
GSTM2	1_110210780_C_G	rs530021	Splice_region	6.10E-07 A
GSTO2	10_106034491_A_G	rs2297235	5'UTR	1.00E-12 A
	10_106037894_T_C	rs157077	Intron	2.40E-08 A
GSTZ1	14_77788908_G_A	rs2363643	Intron	1.40E-06 A
	14_77793207_G_A	rs7975	Missense	8.70E-06 A
NUDT8	11_67395714_C_T	rs7124513	Synonymous	2.50E-07 A
PIAS2	18_44390536_T_C	rs17472	3'UTR	3.50E-07 A
	18_44391566_T_TAG	rs149022619	3'UTR	3.50E-07 A
PON1	7_94927924_C_T	rs854552	3'UTR	1.20E-06 A
	7_94953895_G_A	rs705379	Upstream	1.90E-09 A
SLC22A10	11_63057925_G_A	rs1790218	Stop_gained	2.50E-21 C
	11_63064823_T_C	rs576641	Synonymous	5.10E-19 A
	11_63072310_C_T	rs1201559	Missense	2.50E-21 C
	11_63078986_T_C	rs1404608	3'UTR	2.50E-21 C
	11_63079101_AT_A	rs5792282	3'UTR	4.40E-20 A
SLC29A4	7_5338714_T_C	rs6950111	Synonymous	8.20E-09 A
	7_5342413_T_C	rs11979775	Intron	4.10E-07 D
	7_5342980_C_T	rs56166050	3'UTR	2.80E-06 D
SQSTM1	5_179260153_C_T	rs4935	Synonymous	1.40E-13 A
	5_179260213_G_A	rs4797	Synonymous	2.80E-13 A
	5_179264731_T_C	rs10277	3'UTR	7.60E-17 A
	5_179264915_G_T	rs1065154	3'UTR	1.30E-15 A
SULT2A1	19_48374306_G_A	rs112468411	3'UTR	3.50E-10 C
	19_48374320_C_T	rs112285002	3'UTR	1.50E-61 C
	19_48374538_T_C	rs296366	3'UTR	1.90E-64 C
	19_48374551_C_G	rs296365	3'UTR	5.60E-21 R
	19_48389363_G_A	rs296361	Intron	3.30E-34 C
UGT2A1	4_70454289_A_G	rs4148312	3'UTR	5.20E-08 A
UROC1	3_126200146_A_T	rs777513	3'UTR	1.40E-10 A
	3_126200291_C_T	rs800950	3'UTR	7.70E-07 R
	3_126200403_A_C	rs1799398	3'UTR	3.10E-12 A
	3_126202257_G_A	rs1687477	Synonymous	8.50E-06 R
VKORC1	16_31102321_C_T	rs7294	3'UTR	5.50E-07 A
XRCC5	2_217012901_A_G	rs207906	Synonymous	7.60E-06 D

^aVariant identifier "chromosome _ position _ reference nucleotide _ variant nucleotide". ^bGenetic model with minimal *p*-value: A, additive; R, recessive; D, dominant; C, codominant.

calling by the array method. Further details on performance and validation of the ADME NGS panel are presented in the Sections "Materials and Methods" and **Supplementary Material**.

Analysis of DNA Variants

A total of 16,928 genetic variants were detected within the defined target regions. Of these, 1,201 were excluded from further analysis because of low genotype quality ($n = 505$) or due to HWE *p*-values below 10^{-5} ($n = 696$). The remaining 15,727 variants comprised 13,838 SNV and 1,889 variants classified as small insertions or deletions (INDELs). The length changes of these ranged from deletion of 33 nucleotides up to insertion of 20 nucleotides, with 1 bp

deletions or insertions being the most frequent. Larger structural variants including copy number variations (CNVs) are currently under investigation using other methods.

As expected, most SNVs were biallelic, only 62 were triallelic and no tetraallelic variants were found. Among triallelic variants, transversions were more common ($n = 80$) than transitions, and G to T and G to A were the most common observations ($n = 26$ and $n = 25$, respectively).

None of the sequenced regions was invariant. On average, we observed 10.5 variants/kbp, corresponding to a mean distance of variants of 95 bp. Based on SNV density, the least variable genes were *UGT1A9* and *UGT1A10* with <2 SNVs/kbp and the genes with highest observed variant densities were *CYP4F11*

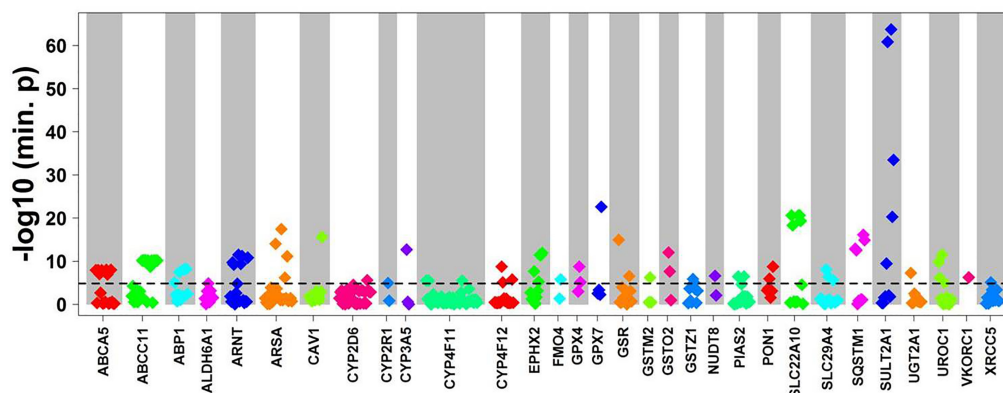


FIGURE 3 | *Cis*-eQTL analysis of common variants. Top *cis*-associations of common variants to mRNA expression. Manhattan plot presenting top results from multivariate *cis* association analysis between mRNA expression and common variants (MAF $\geq 2\%$) investigated for the corresponding gene. Displayed are minimal *p*-values (min. *p*) from four genetic models (codominant, dominant, recessive, and additive). In total, $n = 3,241$ common variants in $n = 295$ genes were analyzed. Only genes with at least one significant *cis*-association after Bonferroni correction ($p < 0.05/3,241 = 1.54\text{E-}05$; dotted line) are shown with all minimal *p*-values. The significant *p*-values are presented in **Table 2**.

(42 SNVs/kbp) and *CYP2D6* (31 SNVs/kbp) (**Supplementary Figure S1C**).

Variant annotation revealed that 7,273 (46.2%) of the variants were not yet annotated in the NCBI dbSNP database (dbSNP build 151, March 2018) and thus considered as novel observations. **Figure 2A** displays the number of variants per gene for known and unknown variants in the different ADME gene groups while **Figure 2B** depicts the fraction of variants according to functional annotation. The number of variants per gene was highest in the ABC and SLC/SLCO transporters and lowest in phase II genes. As reported in several recent studies the number of novel observations was substantial in all gene and functional groups (Fujikura et al., 2015; Gordon et al., 2016; Han et al., 2016). Of 15,727 SNV and small INDEL variants, 12,022 had a MAF below 2%, including 8,937 singletons. Of the 7,273 novel variants, 7,139 (>98%) had MAFs below 2% (**Figure 2C**), while 80 (1.1%) had MAFs $\geq 5\%$. Most of these were located in non-coding regions.

Functional classification based on major transcripts for each gene according to UCSC database revealed 6,058 variants in coding regions (including 3,893 missense and 46 stop gain variants; **Table 1** and **Figure 2B**) and 9,669 variants in various non-coding regions (e.g., 1,000 in 5'UTR and 4,138 in 3'UTR; **Table 1** and **Figure 2B**). We also analyzed 36 VIP genes, derived from PharmaGKB and PharmaADME websites separately for novel SNVs. In total we observed 502 unannotated variants in these genes (dbSNP151), 120 of them representing missense variants (**Supplementary Table S3**).

For comparison with publicly available population data, we extracted small variants from the 1000 Genomes (EUR population) and ExAC (NFE, non-Finnish European) databases for the ADME NGS panel target regions, resulting in 11,558 and 68,918 variants, respectively (**Supplementary Figure S3A**). The MAFs of the matching variants in our sample set (ExAC/NFE: $n = 2,993$; 1000G/EUR: $n = 4,913$) were in good correlation with published population frequency data (Pearson $r = 0.96$ and

$r = 0.98$ for both EUR and NFE populations, respectively). The median MAF of these SNVs was 1.16% for NFE and 2.98% for EUR. We did not detect another 6,645 (EUR) and 65,925 (NFE) known variants with median MAFs of 0.1% (EUR) and 0.002% (NFE) (**Supplementary Figures S3A,B**). Together these data indicate that mainly very rare variants with allele frequencies below 0.1% were missed in our cohort.

Association With Expression Levels

To directly evaluate the functional impact of variants, we assessed liver mRNA expression in an existing dataset (Schröder et al., 2013). To ensure high data quality only mRNA expression data of genes with “perfect” or “good” probes (see section “Materials and Methods”) were considered (available for $n = 303$ genes). Due to sample size and statistical power considerations, we performed separate analyses for less common (MAF < 2%) and more common (MAF $\geq 2\%$) variants.

To evaluate the impact of more common variants ($n = 3,241$) on expression of the corresponding genes we performed *cis*-eQTL analysis using univariate regression models. This analysis revealed significant associations for 94 variants after Bonferroni correction. In multivariate analysis with correction for 10 covariates (see section “Materials and Methods”) 90 variants in 31 genes remained significant after Bonferroni correction (minimal *p*-value of the four genetic models < $1.54\text{E-}05$; **Figure 3** and **Table 2**). Interestingly, 62 (70%) of these were located in non-coding regions, and most of these ($n = 40$) in 3'UTR regions. Of note, three eQTLs represented PharmaGKB VIP genes (*CYP2D6*: rs1080985; *CYP3A5*: rs15524; *VCORC1*: rs7294).

Association analysis of rare variants is challenging. To overcome the problem of limited sample size/statistical power, various methods have been developed to test sets of rare variants. Here we used the SKAT-O approach (Lee et al., 2012) for group-wise association of all rare variants in a gene with mRNA expression data. These variants are incorporated into a gene-wise test statistic via a weighted sum. Thus, *p*-values relate to genes,

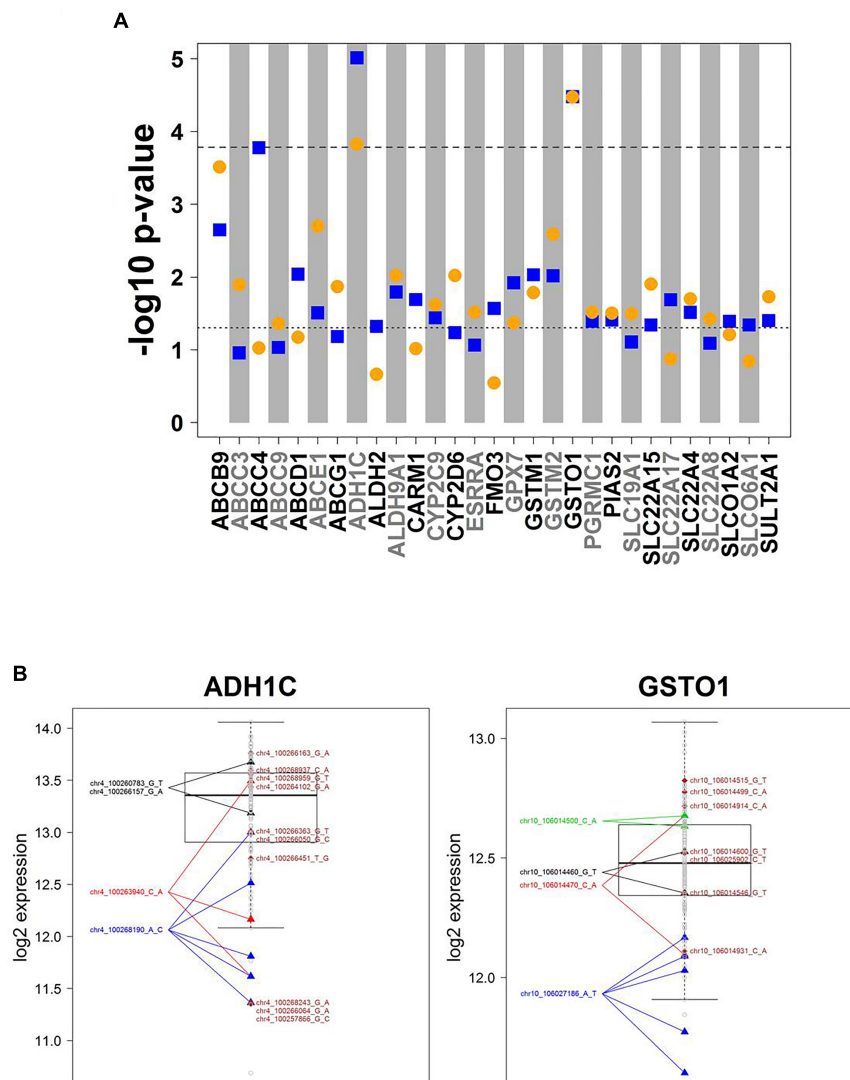


FIGURE 4 | *Cis*-associations of rare variants and mRNA expression (SKAT-O analysis). **(A)** Manhattan plot displaying SKAT-O test p -values from uni- and multivariate *cis*-association analysis between mRNA expression and the set of all rare variants (MAF < 2%) investigated for the corresponding gene. In total, $n = 11,053$ rare variants in 303 genes were analyzed. Only genes with a minimal association p -value < 0.05 are shown. Horizontal dotted lines indicate significance level at 0.05 (lower) and Bonferroni corrected significance level at 0.05/303 = 1.65E-04 (upper). Blue squares: univariate analysis, orange circles: multivariate analysis.

(B) Boxplots of *ADH1C* and *GSTO1* gene expression, the two genes with SKAT-O test p -values < 1.65E-05 in both uni- and multivariate analysis. All variants are heterozygous. Patients with rare variants (MAF < 2%) in *ADH1C* or *GSTO1* are marked by triangles if several patients are carrying a rare mutation or diamonds if a rare mutation is only present in one patient. Gray dots represent patients without rare variants for the gene displayed. Colors differentiate variants.

not to variants. SKAT-O combines the strengths of burden tests thereby being powerful in different scenarios, i.e., when many variants of a gene are associated with expression levels and have the same effect direction, or when there are only few associated variants or variants that differ in effect direction. **Figure 4A** summarizes the results for univariate and multivariate SKAT-O analyses. After correction for multiple testing, two associations, for *ADH1C* and *GSTO1*, remained statistically significant. Further details showing expression levels of individual carriers are presented in **Figure 4B**. For example, five samples with a rather low expression were heterozygous carriers of the SNP chr10_106027186 A > T (3'UTR; rs17885600), including the two

individuals with the lowest *GSTO1* levels (**Figure 4B**). Hence, SKAT-O analysis resulted in identification of at least two genes with plausible genotype–phenotype correlations for variants with MAF < 2%.

Prediction of Functional Effects

We concentrated on coding variants resulting in amino acid change (missense), frameshift, or affecting initiator and stop codons, together accounting for 66% of coding variants and one fourth of all variants (**Figure 2B**). We used the common tools Polyphen 2 (PP2), Provan, and SIFT, that make dichotomous functional predictions of the type “loss of function” (LOF) versus

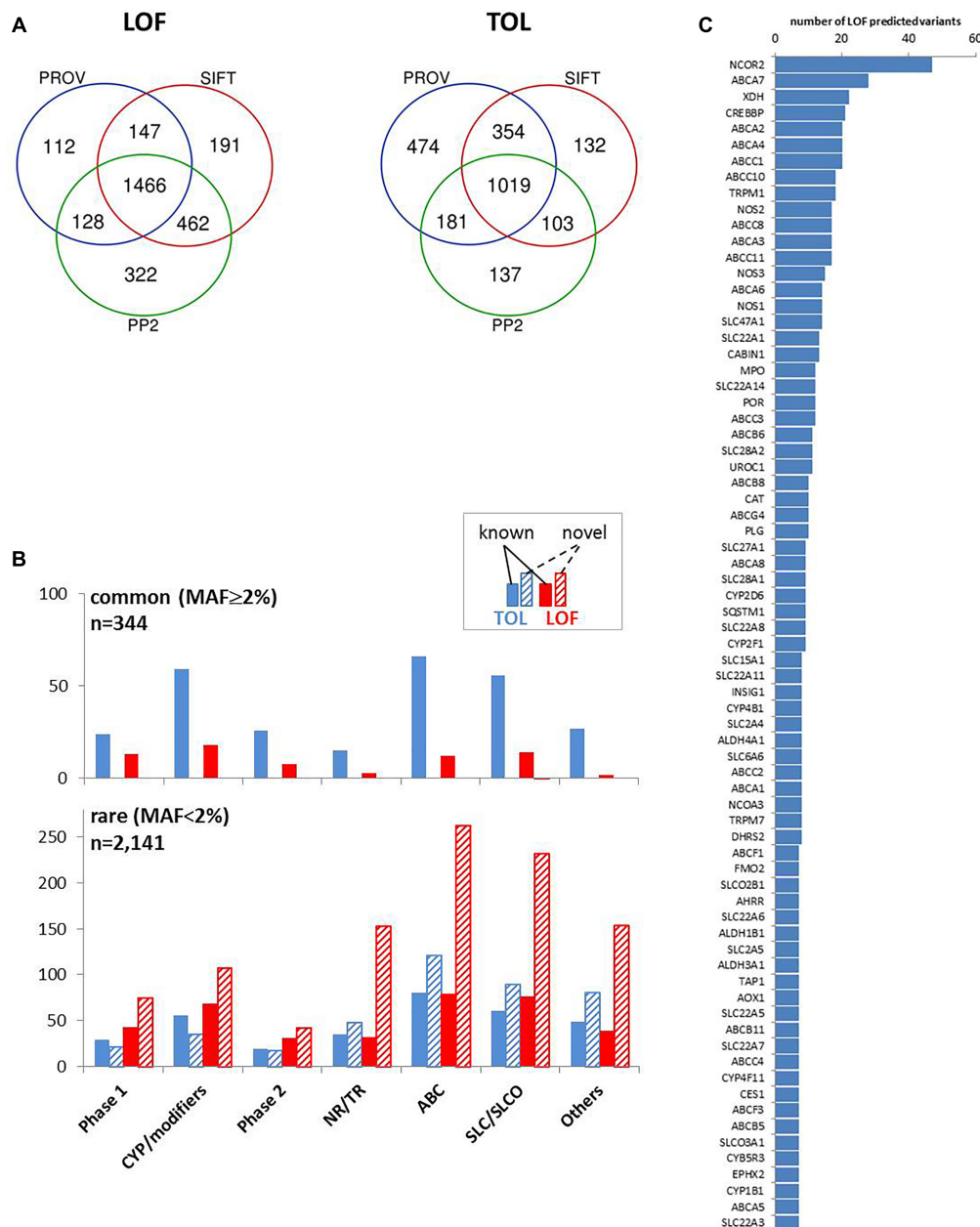


FIGURE 5 | Prediction of coding variant effects. **(A)** Comparison of loss-of-function (LOF) and tolerable (TOL) predictions obtained by three different prediction tools. Venn diagrams are shown for “LOF” and “TOL” predictions for $n = 4,017$ coding variants from Provean (“deleterious”), SIFT (“damaging”), Polyphen2 (PP2; “probably/possibly damaging”). **(B)** Occurrence of TOL and LOF variants in gene family groups. The distribution of the number of concordant TOL ($n = 1,019$; blue colored) and LOF ($n = 1,466$; red colored) predictions is shown for the indicated gene groups for known (filled bars) and novel (hatched bars) variants. Upper chart: variants with $\text{MAF} \geq 2\%$; lower chart: variants with $\text{MAF} < 2\%$. **(C)** Top LOF-variant carrier genes. Shown are genes with at least seven predicted LOF-variants.

“tolerated” (TOL) (Zhou et al., 2018a). Of the analyzed subset of 4,017 coding variants, more than 95% were predictable by these algorithms (PP2, $n = 3,818$; PROV, $n = 3,874$; SIFT, $n = 3,881$). LOF prediction was retrieved concordantly by all three algorithms for 1,466 variants (36.5%) and TOL was concordantly calculated for 1,019 variants (25.4%; **Figure 5A**). In agreement with other studies (Bush et al., 2016; Han et al., 2016; Hovelson et al., 2017) we found that the proportion of LOF- versus

TOL-predicted variants was significantly higher among the less common ($\text{MAF} < 2\%$) compared to more common variants (Chi-square test, $p < 0.0001$). With one exception (*SLC28A1* G254V, $\text{MAF} = 2.3\%$) all novel LOF-predicted variants were less common with $\text{MAF} < 2\%$ (**Figure 5B**).

Interestingly, transporters and nuclear receptors/transcriptional regulators had large proportions of predicted LOF variants that had not yet been listed in the

dbSNP database. The highest number of predicted LOF variants in one gene was observed in *NCOR2* ($n = 47$), and nine ABC transporters (*A7*, *A2*, *A4*, *C1*, *C10*, *C8*, *A3*, and *C11*) are found among the genes with the highest LOF-predicted variants (Figure 5C).

Integrating Prediction and Association

While the SKAT-O test identified only two significant associations, functional prediction indicated a much larger number of predicted LOF variants, as also reported by others (Han et al., 2016; Hovelson et al., 2017). In contrast to former studies, our data allow inspection of genotype-phenotype correlations individually for each variant and for several available phenotypes. While these excessive data are currently being analyzed, we illustrate here a typical example. Of particular interest are protein levels, as functionally damaging ADME gene variants are frequently associated with lower protein levels. Figure 6 shows exemplarily the correlation of all detected amino acid variants of *ABCC11*, encoding the drug transporter MRP8, with MRP8 protein levels obtained for the same liver cohort in a previous study (Magdy et al., 2013). Interestingly, carriers of concordantly LOF-predicted variants ($n = 73$) showed highly variable protein levels (23-fold; coefficient of variation 81%), essentially covering the entire range of MRP8 variability, while carriers of only TOL-predicted variants ($n = 30$) were spread across a smaller protein range (ninefold; coefficient of variation 53%). Of note, the median protein levels of carriers of LOF-predicted and TOL-only-predicted variants were similar ($P = 0.73$; Figure 6). Thus, our phenotypic data allow identification of several MRP8 low and high expressors in relation to genotype. While there does not seem to be a simple relation between functional prediction and phenotypic expression, our data should be helpful to prioritize variants for further investigation and to improve prediction tools.

DISCUSSION

In this study we designed a new panel to target 340 ADME genes for NGS. We tested and validated our ADME NGS panel on a cohort of 150 human liver specimens with comprehensive genetic, functional, and medical characterization. This allowed us not only to perform extensive genotype-phenotype correlations to identify novel relationships for common and rare variants but also to compare computational predictions of functional effects with real phenotypes, which should be useful to further develop and optimize prediction algorithms for variant effects.

We designed our ADME NGS panel to comprise 340 genes including most phase I and phase II enzymes, drug transporters and numerous transcriptional regulators and other modifiers of xenobiotics and endogenous substances. We used Agilent in-solution target capture technology to allow informed selection of relevant regions and optimization of coverage on targets. Only four genes, *SULT1A3*, *SULT1A4*, *MIF*, and *CYP26C1*, were covered below 100-fold. Low coverage of some genes was also observed by others who speculated that common null functional alleles, high sequence homology as well as

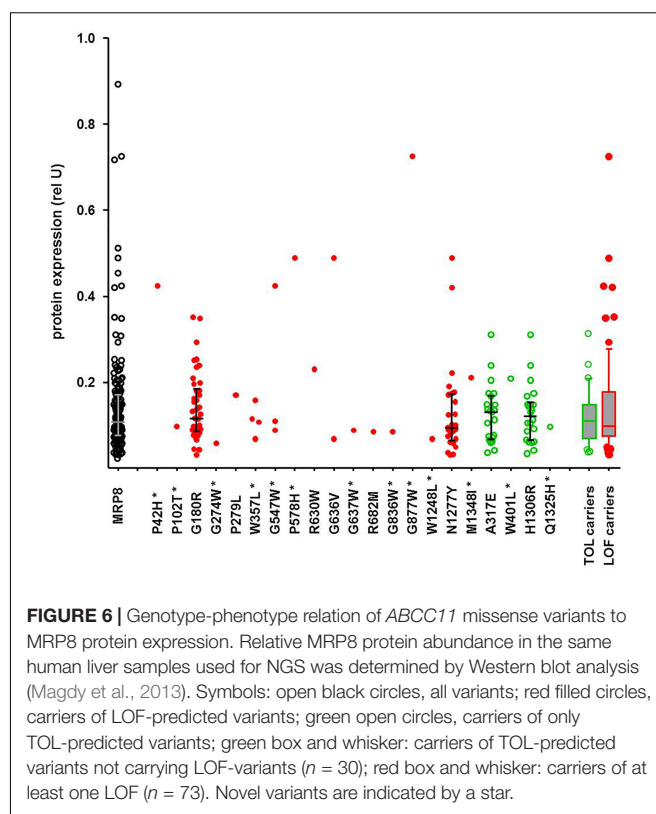


FIGURE 6 | Genotype-phenotype relation of *ABCC11* missense variants to MRP8 protein expression. Relative MRP8 protein abundance in the same human liver samples used for NGS was determined by Western blot analysis (Magdy et al., 2013). Symbols: open black circles, all variants; red filled circles, carriers of LOF-predicted variants; green open circles, carriers of only TOL-predicted variants; green box and whisker: carriers of TOL-predicted variants not carrying LOF-variants ($n = 30$); red box and whisker: carriers of at least one LOF ($n = 73$). Novel variants are indicated by a star.

pseudogenes may disturb capture of such regions (Han et al., 2016). Direct comparison of 363 genotype data available from previous pharmacogenetic studies in the liver cohort revealed an overall accuracy of the ADME NGS panel of $>99\%$. The overall performance of our ADME NGS panel was comparable to other targeted capture sequencing panels (Bush et al., 2016; Gordon et al., 2016; Han et al., 2016; Hovelson et al., 2017). Compared to these other platforms we included a greater number of genes with the intention to investigate not only established ADME genes but also less well known ADME candidate genes.

While several NGS studies of different types recently explored genetic variation in ADME genes (Fujikura et al., 2015; Bush et al., 2016; Han et al., 2016; Kozyra et al., 2016; Hovelson et al., 2017; Schärfe et al., 2017), our study is, to our knowledge, the only one that provides phenotypic measurements in human samples. In this study we analyzed only SNVs and small INDELs, while larger structural variations will be analyzed separately (Tremmel et al., in preparation). For the more common variants ($MAF \geq 2\%$) multivariate eQTL analysis revealed 90 significantly associated variants, most of them located in non-coding regions. Six of these loci had already been described in our previous genome wide association study, e.g., rs7294 in *VKORC1* 3'UTR, or rs1201559 (P516L) in *SLC22A10* (Schröder et al., 2013). Interestingly, several of the SNVs located in 3'UTRs (*ARNT* rs11552229, *CYP3A5*10* rs15524, *EPHX2* rs1042032 and rs1042064, *UGT2A1* rs4148312 and *VKORC1* rs7294) are discussed as potential micro-RNA binding sites, partially proven by tissue eQTL (Wei et al., 2012). Furthermore, our data confirm

predicted eQTL effects on expression in liver tissue in the Genotype-Tissue Expression portal (GTEx¹⁰; Lonsdale et al., 2013) e.g., for the *EPHX2* variant rs1042032 and *VKORC1* rs7294. Some other eQTLs we found had also been reported previously in the context of phenotype/genotype correlations. For example, rs1080985 in *CYP2D6* corresponds to the $-1584C > G$ variant that is linked to the low-expression *CYP2D6**41 allele (Raimundo et al., 2000; Raimundo et al., 2004); the *PON1* rs854552 variant had been found in a nutrigenetic approach on markers of cardiovascular disease (Rizzi et al., 2016); and the *AOC1* (diamine oxidase) variant rs10156191 was associated with hypersensitivity response to non-steroidal anti-inflammatory drugs (Agúndez et al., 2012).

In contrast to common variants, association of individual rare variants is greatly limited by sample size and thus presents a special challenge. The problem is aggravated by the fact that by far most rare variants occur in heterozygous condition, where any effect could be masked by the variability of the “normal” allele. Furthermore, rare variants can be damaging in many ways, affecting expression, protein abundance, or catalytic function. A single phenotype such as expression may thus not reveal the deleterious nature of a particular variant. Nevertheless we assume that analysis of gene or protein expression should be most promising, because damaging variants often affect expression negatively. This is the case, for example, for most low-activity CYP variants (e.g., CYPs 2B6, 2C19, 2D6, 3A4, 3A5 mostly due to aberrant splicing; Zanger and Schwab, 2013), and many established variants of clinical relevance like *UGT1A1**28 and Gilberts syndrome (Ehmer et al., 2012) and *VKORC1* variants in warfarin metabolism (Li et al., 2009). Our statistical approach to relate rare variants to gene expression data by SKAT-O test revealed two significant associations for rare variants of *ADH1C* and *GSTO1*, both of which appear highly plausible and would not have been detected by the *cis*-eQTL analysis. The variant rs283413 in *ADH1C*, a stop gain mutation at protein position G78, is discussed as risk factor for Parkinson’s disease (Buervenich et al., 2005) and alcohol biodisposition (Martínez et al., 2010; Way et al., 2015). The *GSTO1* rare variants have so far not been reported to be associated with expression to our knowledge, but a significant genotype influence of the 3’UTR SNP rs17885600 on expression of the adjacent *GSTO2* in liver tissue supports a potential eQTL effect of this variant (Lonsdale et al., 2013).

As a further approach to identify deleterious ADME rare variants, we used computational prediction, which has recently been used in several studies (Bush et al., 2016; Han et al., 2016; Hovelson et al., 2017). However, in none of these studies, phenotypic information was provided to compare prediction with a phenotypic parameter. Similar to other studies we found a considerable fraction of all variants (36.5%) to be predicted as damaging by all three prediction tools used. Somewhat unexpectedly, preliminary analyses did not reveal statistically significant associations between LOF-predicted variants and lower expression. As exemplarily illustrated for *ABCC11* and MRP8 protein abundance, LOF predicted variants were not more frequently associated with lower protein levels as compared to

TOL predicted variants. Thorough analyses of these data are currently in progress. A recent advanced approach integrated prediction and functional activity data available from diverse sources to develop an improved prediction framework adopted to pharmacogenetic assessments (Zhou et al., 2018b). Our data should be highly valuable to test and further improve such approaches.

CONCLUSION

We designed a new targeted NGS pipeline to determine SNVs and small INDELs for 340 ADME genes and used it to analyze 150 well characterized human liver samples. In addition to common known variants we confirmed the existence of large numbers of rare and previously unknown germline variants. Available phenotypic information on the samples allowed us to elucidate numerous novel eQTLs for common variants and to identify novel relationships between rare variants and expression. Furthermore our data allow direct comparison of computationally predicted functional effects for coding variants with actual phenotypes. Using data for the transporter *ABCC11*/MRP8, we showed that variants predicted as deleterious are present in both high and low expressors of MRP8. While this emphasizes challenges and current limitations of computational prediction approaches to integrate rare variants into pharmacogenomics, such data are important to assess and improve the current strategies.

AUTHOR CONTRIBUTIONS

KK, ES, MS, UZ, SF, and SB designed the study. KK, UZ, and MS provided DNA samples. SF, FB, TS, and SB designed the panel and generated sequencing data. KK, RT, SW, and SF analyzed the data. KK, RT, and UZ wrote the manuscript. All authors contributed to editing and final proofreading the manuscript.

FUNDING

This study was supported by the Robert Bosch Foundation, Stuttgart, Germany and the European Commission Horizon 2020-PHC-2015 grant U-PGx 668353.

ACKNOWLEDGMENTS

We thank Dr. Florian Büttner for help with large datasets (1000G, ExAC). The excellent technical assistance of Igor Liebermann is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00007/full#supplementary-material>

¹⁰<https://commonfund.nih.gov/gtex>

REFERENCES

- 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632
- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7:Unit7.20. doi: 10.1002/0471142905.hg0720s76
- Agúndez, J. A. G., Ayuso, P., Cornejo-García, J. A., Blanca, M., Torres, M. J., Doña, I., et al. (2012). The diamine oxidase gene is associated with hypersensitivity response to non-steroidal anti-inflammatory drugs. *PLoS One* 7:e47571. doi: 10.1371/journal.pone.0047571
- Alfirevic, A., and Pirmohamed, M. (2017). Genomics of adverse drug reactions. *Trends Pharmacol. Sci.* 38, 100–109. doi: 10.1016/j.tips.2016.11.003
- Buervenich, S., Carmine, A., Galter, D., Shahabi, H. N., Johnels, B., Holmberg, B., et al. (2005). A rare truncating mutation in ADH1C (G78Stop) shows significant association with Parkinson disease in a large international sample. *Arch. Neurol.* 62, 74–78. doi: 10.1001/archneur.62.1.74
- Bush, W. S., Crosslin, D. R., Owusu-Obeng, A., Wallace, J., Almoguera, B., Basford, M. A., et al. (2016). Genetic variation among 82 pharmacogenes: the PGRNseq data from the eMERGE network. *Clin. Pharmacol. Ther.* 100, 160–169. doi: 10.1002/cpt.350
- Caudle, K. E., Thorn, C. F., Klein, T. E., Swen, J. J., McLeod, H. L., Diasio, R. B., et al. (2013). Clinical pharmacogenetics implementation consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clin. Pharmacol. Ther.* 94, 640–645. doi: 10.1038/clpt.2013.172
- Cecchin, E., Roncato, R., Guchelaar, H. J., and Toffoli, G. (2017). Ubiquitous pharmacogenomics (U-PGx): the time for implementation is now. *An Horizon2020 program to drive pharmacogenomics into clinical practice. Curr. Pharm. Biotechnol.* 18, 204–209. doi: 10.2174/1389201018666170103103619
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7:e46688. doi: 10.1371/journal.pone.0046688
- Crosslin, D. R., Robertson, P. D., Carrell, D. S., Gordon, A. S., Hanna, D. S., Burt, A., et al. (2015). Prospective participant selection and ranking to maximize actionable pharmacogenetic variants and discovery in the eMERGE network. *Genome Med.* 7:67. doi: 10.1186/s13073-015-0181-z
- Dohrn, M. F., Glöckle, N., Mulahasanovic, L., Heller, C., Mohr, J., Bauer, C., et al. (2017). Frequent genes in rare diseases: panel-based next generation sequencing to disclose causal mutations in hereditary neuropathies. *J. Neurochem.* 143, 507–522. doi: 10.1111/jnc.14217
- Dunnenberger, H. M., Crews, K. R., Hoffman, J. M., Caudle, K. E., Broeckel, U., Howard, S. C., et al. (2015). Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annu. Rev. Pharmacol. Toxicol.* 55, 89–106. doi: 10.1146/annurev-pharmtox-010814-124835
- Dunning, M., Lynch, A., and Eldridge, M. (2015). *illuminaHumanv2.db: Illumina HumanWG6v2 Annotation data (chip illuminaHumanv2). R package version 1.26.0.*
- Ehmer, U., Kalthoff, S., Fakundiny, B., Pabst, B., Freiberg, N., Naumann, R., et al. (2012). Gilbert syndrome redefined: a complex genetic haplotype influences the regulation of glucuronidation. *Hepatology* 55, 1912–1921. doi: 10.1002/hep.25561
- Fujikura, K., Ingelman-Sundberg, M., and Lauschke, V. M. (2015). Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenet. Genomics* 25, 584–594. doi: 10.1097/FPC.0000000000000172
- Gomes, A. M., Winter, S., Klein, K., Turpeinen, M., Schaeffeler, E., Schwab, M., et al. (2009). Pharmacogenomics of human liver cytochrome P450 oxidoreductase: multifactorial analysis and impact on microsomal drug oxidation. *Pharmacogenomics* 10, 579–599. doi: 10.2217/pgs.09.7
- González, J. R., Armengol, L., Guinó, E., Solé, X., and Moreno, V. (2014). *SNPassoc: SNPs-based Whole Genome Association Studies. R package version 1.9-2.* Available at: <https://CRAN.R-project.org/package=SNPassoc>
- Gordon, A. S., Fulton, R. S., Qin, X., Mardis, E. R., Nickerson, D. A., and Scherer, S. (2016). PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. *Pharmacogenet. Genomics* doi: 10.1097/FPC.0000000000000202 [Epub ahead of print].
- Han, S. M., Park, J., Lee, J. H., Lee, S. S., Kim, H., Han, H., et al. (2016). Targeted next-generation sequencing for comprehensive genetic profiling of pharmacogenes. *Clin. Pharmacol. Ther.* 101, 396–405. doi: 10.1002/cpt.532
- Hovelson, D. H., Xue, Z., Zawistowski, M., Ehm, M. G., Harris, E. C., Stocker, S. L., et al. (2017). Characterization of ADME gene variation in 21 populations by exome sequencing. *Pharmacogenet. Genomics* 27, 89–100. doi: 10.1097/FPC.0000000000000260
- Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y., and Lauschke, V. M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* 12:1714. doi: 10.1186/s40246-018-0157-3
- Ji, Y., Skierka, J. M., Blommel, J. H., Moore, B. E., VanCuyk, D. L., Bruflat, J. K., et al. (2016). Preemptive pharmacogenomic testing for precision medicine. *J. Mol. Diagn.* 18, 438–445. doi: 10.1016/j.jmoldx.2016.01.003
- Klein, K., Thomas, M., Winter, S., Nussler, A. K., Niemi, M., Schwab, M., et al. (2012). PPARA: a novel genetic determinant of CYP3A4 in vitro and in vivo. *Clin. Pharmacol. Ther.* 91, 1044–1052. doi: 10.1038/clpt.2011.336
- Kozyra, M., Ingelman-Sundberg, M., and Lauschke, V. M. (2016). 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. doi: 10.1038/gim.2016.33
- Lappalainen, I., Almeida-King, J., Kumanduri, V., Senf, A., Spalding, J. D., Ur-Rehman, S., et al. (2015). The european genome-phenome archive of human data consented for biomedical research. *Nat. Genet.* 47, 692–695. doi: 10.1038/ng.3312
- Lee, S. (2017). *SKAT: SNP-Set (Sequence) Kernel Association Test. R Package Version 1.3.2.1.* Available at: <https://CRAN.R-project.org/package=SKAT>
- Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., et al. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* 91, 224–237. doi: 10.1016/j.ajhg.2012.06.007
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. doi: 10.1038/nature19057
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, J., Wang, S., Barone, J., and Malone, B. (2009). Warfarin pharmacogenomics. *P T* 34, 422–427.
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., and Shad, S. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45, 580–585. doi: 10.1038/ng.2653
- Magdy, T., Arlanov, R., Winter, S., Lang, T., Klein, K., Toyoda, Y., et al. (2013). ABCC11/MRP8 polymorphisms affect 5-fluorouracil-induced severe toxicity and hepatic expression. *Pharmacogenomics* 14, 1433–1448. doi: 10.2217/pgs.13.139
- Martínez, C., Galván, S., García-Martín, E., Ramos, M. I., Gutiérrez-Martín, Y., and Agúndez, J. A. G. (2010). Variability in ethanol biodisposition in whites is modulated by polymorphisms in the ADH1B and ADH1C genes. *Hepatology* 51, 491–500. doi: 10.1002/hep.23341
- Matthaei, J., Brockmöller, J., Tzvetkov, M. V., Seht, D., Sachse-Seeboth, C., Hjelmborg, J. B., et al. (2015). Heritability of metoprolol and torsemide pharmacokinetics. *Clin. Pharmacol. Ther.* 98, 611–621. doi: 10.1002/cpt.258
- Meyer, U. A. (2004). Pharmacogenetics - five decades of therapeutic lessons from genetic diversity. *Nat. Rev. Genet.* 5, 669–676. doi: 10.1038/nrg1428
- R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>
- Raimundo, S., Fischer, J., Eichelbaum, M., Gries, E. U., Schwab, M., and Zanger, U. M. (2000). Elucidation of the genetic basis of the common 'intermediate metabolizer' phenotype for drug oxidation by CYP2D6. *Pharmacogenetics* 10, 577–581. doi: 10.1097/00008571-200010000-00001
- Raimundo, S., Toscano, C., Klein, K., Fischer, J., Gries, E.-U., Eichelbaum, M., et al. (2004). A novel intronic mutation, 2988G > A, with high predictivity for

- impaired function of cytochrome P450 2D6 in white subjects. *Clin. Pharmacol. Ther.* 76, 128–138. doi: 10.1016/j.clpt.2004.04.009
- Relling, M. V., and Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature* 526, 343–350. doi: 10.1038/nature15817
- Rizzi, F., Conti, C., Dogliotti, E., Terranegra, A., Salvi, E., Braga, D., et al. (2016). Interaction between polyphenols intake and PON1 gene variants on markers of cardiovascular disease: a nutrigenetic observational study. *J. Transl. Med.* 14:186. doi: 10.1186/s12967-016-0941-6
- Schärfe, C. P. I., Tremmel, R., Schwab, M., Kohlbacher, O., and Marks, D. S. (2017). Genetic variation in human drug-related genes. *Genome Med.* 9:117. doi: 10.1186/s13073-017-0502-5
- Schröder, A., Klein, K., Winter, S., Schwab, M., Bonin, M., Zell, A., et al. (2013). Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. *Pharmacogenomics J.* 13, 12–20. doi: 10.1038/tpj.2011.44
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40, W452–W457. doi: 10.1093/nar/gks539
- Swen, J. J., Nijenhuis, M., van Rhenen, M., Boer-Veger, N. J., de Buunk, A.-M., Houwink, E. J. F., et al. (2018). Pharmacogenetic information in clinical guidelines: the european perspective. *Clin. Pharmacol. Ther.* 103, 795–801. doi: 10.1002/cpt.1049
- Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64–69. doi: 10.1126/science.1219240
- Way, M., McQuillin, A., Saini, J., Ruparel, K., Lydall, G. J., Guerrini, I., et al. (2015). Genetic variants in or near ADH1B and ADH1C affect susceptibility to alcohol dependence in a British and Irish population. *Addict. Biol.* 20, 594–604. doi: 10.1111/adb.12141
- Wei, R., Yang, F., Urban, T. J., Li, L., Chalasani, N., Flockhart, D. A., et al. (2012). Impact of the interaction between 3'-UTR SNPs and microRNA on the expression of human xenobiotic metabolism enzyme and transporter genes. *Front. Genet.* 3:248. doi: 10.3389/fgene.2012.00248
- Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 92, 414–417. doi: 10.1038/clpt.2012.96
- Zanger, U. M., and Schwab, M. (2013). Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* 138, 103–141. doi: 10.1016/j.pharmthera.2012.12.007
- Zhou, Y., Fujikura, K., Mkrtchian, S., and Lauschke, V. M. (2018a). Computational methods for the pharmacogenetic interpretation of next generation sequencing data. *Front. Pharmacol.* 9:1437. doi: 10.3389/fphar.2018.01437
- Zhou, Y., Mkrtchian, S., Kumondai, M., Hiratsuka, M., and Lauschke, V. M. (2018b). An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* doi: 10.1038/s41397-018-0044-2

Conflict of Interest Statement: SF, FB, TS, and SB were employed by CeGaT GmbH, Tübingen.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Klein, Tremmel, Winter, Fehr, Battke, Scheurenbrand, Schaeffeler, Biskup, Schwab and Zanger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Next-Generation Sequencing of *PTGS* Genes Reveals an Increased Frequency of Non-synonymous Variants Among Patients With NSAID-Induced Liver Injury

María Isabel Lucena^{1†}, Elena García-Martín^{2†}, Ann K. Daly³, Miguel Blanca⁴, Raúl J. Andrade¹ and José A. G. Agúndez^{2*}

¹ Unidad de Gestión Clínica de Aparato Digestivo, Servicio de Farmacología Clínica, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas Málaga, Instituto de Investigación Biomédica de Málaga, Hospital Universitario Virgen de la Victoria, Universidad de Málaga, Málaga, Spain, ² Instituto de Salud Carlos III, University Institute of Molecular Pathology Biomarkers, UNEx, ARADyAL, Cáceres, Spain, ³ Liver Research Group, Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle upon Tyne, United Kingdom, ⁴ Servicio de Alergología, Hospital Infanta Leonor, ARADyAL, Madrid, Spain

OPEN ACCESS

Edited by:

George P. Patrinos,
University of Patras, Greece

Reviewed by:

Volker Martin Lauschke,
Karolinska Institute (KI), Sweden
Su-Jun Lee,
Inje University, South Korea

*Correspondence:

José A. G. Agúndez
jagundez@unex.es

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Genetics

Received: 19 November 2018

Accepted: 08 February 2019

Published: 28 February 2019

Citation:

Lucena MI, García-Martín E, Daly AK,
Blanca M, Andrade RJ and
Agúndez JAG (2019) Next-Generation
Sequencing of *PTGS* Genes Reveals
an Increased Frequency of
Non-synonymous Variants Among
Patients With NSAID-Induced Liver
Injury. *Front. Genet.* 10:134.
doi: 10.3389/fgene.2019.00134

Purpose: The etiopathogenesis of drug-induced liver injury (DILI) is still far from being elucidated. This study aims to the study of genetic variations in DILI, related to the drug target, and specifically in the genes coding for the cyclooxygenase enzymes.

Methods: By using Next-generation Sequencing we analyzed the genes coding for COX enzymes (*PTGS1* and *PTGS2*) in 113 individuals, 13 of which were patients with DILI caused by COX-inhibitors.

Results: The key findings of the study are the increased frequency, among DILI patients, of SNPs causing alterations in transcription factor binding sites and non-synonymous *PTGS* gene variants, as compared to control subjects. Moreover, the association with non-synonymous SNPs was exclusive of DILI patients with late-onset (50 days or more) $P_c < 0.001$ as compared to DILI patients with early onset, or with control subjects.

Conclusions: Our findings suggest an interaction of long-term exposure to COX inhibitors combined with functional variants of the COX enzymes in the risk of developing DILI. This is a novel observation that might have been overlooked by previous genetic studies on DILI because of the limited coverage of *PTGS* genes in exome chips.

Keywords: *PTGS1*, *PTGS2*, next generation sequencing, drug-induced liver injury, *COX1*, *COX2*

BACKGROUND

Although drug-induced liver injury (DILI) is a rare adverse drug event, it is often life-threatening because of the risk of developing acute liver failure. The mechanisms underlying DILI risk are not well understood and hence, the search for biomarkers of DILI risk is a major research field that aims to identify markers that could be used as both proof of the mechanisms involved and of the risk factors that can be used for DILI prediction, as has already been done with many pharmacogenomics biomarkers (Lucena et al., 2008, 2010; Agúndez, 2009; Agundez et al., 2009, 2011; Andrade et al., 2009; Robles-Díaz et al., 2016; Nicoletti et al., 2017). There are presently several

independent hypotheses to explain idiosyncratic DILI, but none of these is able to explain all the circumstances in which DILI occurs.

Some genetic biomarkers for DILI either mechanistically-based using a case-control strategy or with a GWAs/exome sequencing approaches have been identified [for a review, see (Robles-Diaz et al., 2016)]. However, the involvement of genetic changes in DILI risk (for instance HLA risk alleles) has been documented for only a few drugs (Kaliyaperumal et al., 2018). On the other hand, case-control genotyping studies, GWAS and exome sequencing have important limitations because only some SNPs are tested, and most of the target sequence is not checked. To overcome this problem, deep sequencing comprising whole genes is necessary.

In this study, we analyzed the potential effect of mutations in cyclooxygenase genes (*PTGS1* and *PTGS2*) on DILI risk related to NSAIDs. From a mechanistic point of view, such a risk could be related to genetic alteration in the arachidonic acid pathways, which are closely related to inflammation. On the other hand, adverse drug events for drugs acting on the COX enzymes (that is, COX inhibitors) may be more likely if COX activity is altered because of genetic variations. For this reason, we analyzed patients who developed DILI after the administration of COX-inhibitors and healthy individuals who tolerated COX-inhibitors.

CASE PRESENTATION

Thirteen patients (8 women and 5 men) who experienced DILI caused by COX inhibitors and 100 individuals who tolerated COX-inhibitors at standard doses were included in this study. The culprit drug for DILI and clinical details of patients are shown in **Table 1**. Gender-matched control individuals who tolerated COX-inhibitors (62 women and 38 men) individuals were recruited among staff and medical students of the Hospitals and the Universities participating in this study. Individuals which were considered as healthy after medical examination, to exclude pre-existing disorders and history of adverse events after the use of COX-inhibitors, were asked to participate and over 95% of these agreed to do so. We selected consecutive control subjects matched with patients for drug exposure: Fifty control subjects who have received ibuprofen within the previous month to sample collection, 20 who received diclofenac, 10 indomethacin, 10 naproxen, and 10 rofecoxib. These frequencies match with the frequencies for the DILI patients, except that no control subject received nimesulide since this drug was discontinued from the Spanish market due to liver safety. Both patients and controls were Caucasian Spanish individuals. Written informed consent for participation in this case report was obtained from all

participants. The protocol for this study was in accordance with the Declaration of Helsinki and its subsequent revisions and was approved by the respective Ethics Committees of the participating Hospitals.

DESCRIPTION OF LABORATORY INVESTIGATIONS AND DIAGNOSTIC TESTS

To achieve complete gene capture, we sequenced all exons, intron-exon boundaries as well as the 5' and 3' flanking regions for both genes. Referred to the GRCh37 assembly of the human genome, the sequences studied were the following: *PTGS1*: Chromosome 9:125.131.159 to 125.158.017; *PTGS2*: Chromosome 1:186.640.825 to 186.651.605. Partially overlapping amplicons with a size lower than 400 bp were designed. A total of 62 CS1/CS2 tagged primer pairs were synthesized and used to amplify 113 DNA samples using the Access Array platform (Fluidigm). During amplification, samples were labeled with standard MID barcodes designed for the FLX454 sequencing system. After amplification and MID-labeling, individual amplicon libraries were analyzed using a Bioanalyzer 2100 (Agilent) and bioanalyzer traces were used to estimate the amplicon concentration for each sample. Samples were then pooled, and libraries were purified by SPRI using Ampure beads to remove all possible traces of small molecules, primers, primer-dimers, or any other contaminants. The pooled library was again quantified and titrated so that a final amount of 1.95×10^{10} molecules with an enrichment percentage of 7% was loaded on a Pico Titer Plate (Roche) for a 200-cycle titanium-based sequencing run, made on FLX-454 equipment. Reads were processed using an amplicon processing pipeline and sff files were used for further analyses. Coverage averaged around 50x for the whole project. Coverage for the SNPs identified (shown in **Supplemental Table 1**) was always over 50x. Sequencing reads were de-multiplexed and aligned using the Amplicon Variant Analyzer software v2.8 (Roche) so that reads for each particular sample- target region combination were analyzed in search of variants. Details of the amplification and sequencing primers are available in **Supplemental Table 1**.

The putative effect on the non-synonymous variants identified *in silico* was assessed by using the Sorting Tolerant form Intolerant (SIFT) and Polymorphism Phenotyping (PolyPhen) scores as shown in the 1,000 genomes website for every SNP, as well as the online application MutationAssessor (<http://mutationassessor.org/r3/>).

RESULTS

The sequencing results (summarized in **Table 2**) reveal that *PTGS* genes are well conserved. Although dozens of *PTGS1* and *PTGS2* single nucleotide polymorphisms (SNPs) have been described to occur in Caucasian populations (see Agúndez et al., 2015), our findings show that most of these SNPs were not identified, or were extremely rare, in this cohort.

Abbreviations: COX, Cyclooxygenase, prostaglandin-endoperoxide synthase; NSAID, Non-steroidal anti-inflammatory drug; DILI, Drug-induced liver injury; GWAS, Genome-wide association study; SNP, Single nucleotide polymorphism; HLA, Human leukocyte antigen; *PTGS1*, Prostaglandin-Endoperoxide Synthase 1; *PTGS2*, Prostaglandin-Endoperoxide Synthase 2.

TABLE 1 | Demographic and clinical characteristics of 13 patients with NSAIDs-induced idiosyncratic liver injury.

N°	Culprit drug	Age/ Sex	Indication	Daily dose (mg)	Duration treatment (days)	Time to Associated onset (days)	Clinical presentation	Type of liver injury	TB (xULN)	ALT (xULN)	ALP (xULN)	Severity index	Outcome	CIOMS score
1	Diclofenac	76/F	Lumbar Pain	50	93	30	Jaundice ASMA 1/80	HC	2.9	24	2.6	Moderate	Resolution 221 days	Highly probable
2	Nimesulide	62/F	Osteoarthritis	200	30	30	Jaundice Eosinophilia	HC	24	98	2.5	Moderate	Resolution 90 days	Probable
3	Nimesulide	61/F	Osteoarthritis	200	62	62	Fosinopril Torasemide	HC	15	30	1.7	Moderate, hospitalized	Resolution 126 days	Highly probable
4	Indometacin	54/F	Osteoarthritis	100	11	8	Tramadol Tetrazepam Dexamethasone	Mix	1	16	3.6	Mild	Resolution 44 days	Highly probable
5	Ibuprofen	64/F	Pain	600	17	10	Enalapril Insulin	Mix	9	5.9	1.7	Moderate, hospitalized	Resolution 180 days	Highly probable
6	Naproxen	65/M	Pain	1,000	44	48	Tetrazepam Ranitidine Heparin Bisoprolol Metamizole	Mix	4.5	9.7	3.4	Moderate, hospitalized	Lost of follow-up	Highly probable
7	Ibuprofen	18/M	Pain	600	1	2	None	Mix	7.9	5.5	2.6	Moderate, hospitalized	Lost of follow-up	Highly probable
8	Rofecoxib	82/M	Osteoarthritis	12.5	60	59	Nimodipine Omeprazole Paracetamol Troloxerutin Insulin Vitamins Folic Acid	Chol	2.6	4.9	4.4	Moderate	Resolution 94 days	Highly probable
9	Ibuprofen	57/M	Pain	1,200	31	50	Gilbencamide	HC	1	24	0.9	Mild	Lost of follow-up	Highly probable
10	Nimesulide	59/F	Osteoarthritis	200	25	466	Atenolol Captopril Hydrochlorothiazide Insulin Metformin	Chol	1.2	1.2	5	Moderate, hospitalized	Resolution 284 days	Highly probable
11	Ibuprofen	43/M	Pain	1,200	8	8	Metronidazole	HC	1.9	8.2	1.3	Mild, hospitalized	Resolution 308	Probable
12	Diclofenac	80/F	Pain	75	171	171	Telmisartan Allopurinol Gilbencamide Ebastine Troloxerutin	HC	4.9	38	1.9	Moderate, hospitalized	Resolution 348	Highly probable
13	Ibuprofen	41/F	Rheumatoid arthritis	600	3	7	Diclofenac Paracetamol Diazepam Fluoxetine	HC	1.6	14	1.8	Mild	Resolution 31	Probable

F, Female; M, Male; ASMA, anti-smooth muscle autoantibody; HC, Hepatocellular; Mix, mixed; Chol, cholestatic; TB, total bilirubin; ALT, alanine aminotransferase; ALP, alkaline phosphatase; ULN, upper limit of normal. Severity index, Mild, elevated ALT/ALP meeting DILI criteria with total bilirubin <2 mg/dL; Moderate, elevated ALT/ALP with total bilirubin ≥ 2 g/dL; Severe, elevated ALT/ALP and one of the following, ascites, encephalopathy, international normalization ratio > 1.5 and/or other organ failure considered to be due to DILI; Fatal, death or transplantation due to DILI.

TABLE 2 | PTGS1 and PTGS2 variant sequences identified in the study group.

Coordinate GRCCh37.p13 (GCA_000001405.14)	rs ID	N° case	1	2	3	4	5	6	7	8	9	10	11	12	13	Effect	MAF DILI	MAF control	MAF IBS	MAF AFR	MAF AMR	MAF EAS	MAF EUR	MAF SAS
PTGS1 (COX-1)																								
9:125131480	rs10306108	0	0	0	0	0	0	0	1	0	1	0	0	0	0	UGV	0.077	0.055	0.06	0.13	0.03	0.00	0.07	0.01
9:125131631	rs10306109	0	0	0	0	0	0	0	1	0	1	0	0	0	0	UGV	0.077	0.055	0.06	0.13	0.03	0.00	0.07	0.01
9:125131688	rs1330344	0	0	0	0	0	1	1	1	0	1	0	1	0	0	UGV	0.154	0.210	0.24	0.52	0.21	0.42	0.20	0.39
9:125131832	rs10306225	0	0	0	0	0	0	0	0	2	0	2	0	0	0	UGV	0.154	0.000	0.00	0.00	0.00	0.00	0.01	0.00
9:125132027	rs115693689	0	0	0	0	0	0	0	1	0	0	0	0	0	0	UGV	0.038	0.055	0.07	0.13	0.04	0.05	0.08	0.02
9:125132028	rs114079139	0	0	0	0	0	0	0	1	0	0	0	0	0	0	UGV	0.038	0.055	0.07	0.13	0.04	0.05	0.08	0.02
9:125132069	rs77676149	0	0	0	0	0	0	0	0	1	0	1	0	0	0	UGV	0.077	0.000	0.00	0.01	0.01	0.00	0.00	0.00
9:125132223	rs75993350	0	0	0	0	0	0	0	0	0	1	0	0	0	0	UGV	0.038	0.055	0.06	0.13	0.03	0.00	0.07	0.01
9:125132311	rs10306110	0	0	0	0	0	0	0	0	0	1	0	0	0	0	UGV	0.038	0.055	0.06	0.13	0.03	0.00	0.07	0.01
9:125132522	rs10306114	0	0	0	0	0	0	0	0	0	1	0	0	0	0	UGV	0.038	0.055	0.06	0.13	0.03	0.00	0.07	0.01
9:125132909	rs10306115	0	0	0	0	0	0	0	0	0	0	0	0	0	0	UGV	0.038	0.055	0.06	0.13	0.03	0.00	0.07	0.01
9:125133479	rs1236913	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5'RRV	0.077	0.000	0.00	0.07	0.00	0.00	0.00	0.00
9:125133507	rs3842787	0	0	1	0	0	0	0	0	0	1	0	0	1	0	MSV ^a	0.115	0.065	0.07	0.01	0.07	0.01	0.06	0.18
9:125140206	rs3842788	0	0	0	0	0	0	0	0	0	0	0	0	1	0	MSV ^b	0.077	0.060	0.06	0.15	0.03	0.00	0.07	0.01
9:125140287	rs3842790	0	0	0	0	0	0	0	1	0	0	0	0	0	0	SV	0.038	0.025	0.02	0.32	0.03	0.07	0.04	0.04
9:125140696	rs2282169	0	1	0	1	0	0	0	2	0	0	0	1	0	0	IV	0.192	0.140	0.15	0.60	0.28	0.11	0.19	0.23
9:125140823	rs5787	0	0	0	0	0	0	0	0	0	0	2	0	0	0	MSV ^c	0.077	0.000	0.00	0.00	0.00	0.00	0.00	0.00
9:125141239	rs12555242	0	0	0	0	0	0	0	0	0	0	0	2	0	0	IV	0.077	0.070	0.02	0.00	0.03	0.05	0.02	0.08
9:125143707	rs3842792	0	0	0	0	0	0	0	0	0	0	0	0	1	0	MSV ^d	0.038	0.000	0.00	0.04	0.00	0.00	0.00	0.00
9:125143792	rs5788	0	1	0	1	0	0	0	1	0	0	0	0	0	0	SV	0.115	0.110	0.11	0.69	0.24	0.04	0.14	0.10
9:125143882	rs3842794	0	0	0	0	0	0	0	0	1	0	1	0	0	0	IV	0.077	0.000	0.00	0.02	0.00	0.00	0.00	0.00
9:125144040	rs3215925	0	1	0	1	0	0	0	1	0	0	0	0	0	0	IV	0.115	0.095	0.11	0.67	0.24	0.04	0.14	0.10
9:125145743	rs3842798	0	1	0	1	0	0	2	0	0	0	0	2	0	0	IV	0.231	0.170	0.15	0.76	0.30	0.11	0.20	0.23
9:125155408	rs8046	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3'UTRV	0.038	0.000	0.04	0.51	0.10	0.09	0.07	0.14
9:125155930	rs10306192	0	0	0	0	0	0	0	0	0	0	0	2	0	0	3'UTRV	0.077	0.000	0.04	0.01	0.05	0.09	0.06	0.13
9:125156374	rs199981440	0	0	0	0	0	0	0	0	1	0	0	0	1	0	3'UTRV	0.077	0.000	-	-	-	-	-	-
9:125157198	rs10306194	0	0	1	0	1	0	0	2	0	0	1	2	0	0	3'UTRV	0.269	0.175	0.20	0.02	0.11	0.04	0.15	0.14
9:125157316	rs10306196	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3'UTRV	0.038	0.065	0.04	0.01	0.05	0.09	0.06	0.13
9:125157357	rs10306197	0	0	0	0	0	0	0	1	0	0	0	0	0	1	3'UTRV	0.077	0.000	0.00	0.02	0.00	0.00	0.00	0.00
9:125157672	rs10306199	0	0	0	0	1	1	0	0	0	0	0	0	0	0	3'UTRV	0.077	0.000	0.00	0.02	0.00	0.00	0.00	0.00
9:125157718	rs9233	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3'UTRV	0.038	0.065	0.04	0.01	0.05	0.09	0.06	0.14
PTGS2 (COX-2)																								
1:186640853	rs4648304	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3'UTRV	0.038	0.000	0.00	0.10	0.01	0.00	0.00	0.00
1:186641058	rs689470	1	0	1	0	0	0	0	0	0	0	0	0	0	0	3'UTRV	0.077	0.000	0.05	0.50	0.05	0.01	0.03	0.07
1:186641265	rs689468	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3'UTRV	0.038	0.005	0.03	0.00	0.01	0.01	0.02	0.03
1:186641273	rs689467	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3'UTRV	0.038	0.025	0.07	0.00	0.03	0.00	0.05	0.03

(Continued)

TABLE 2 | Continued

Coordinate GRCh37.p13 (GCA_000001405.14)	N° case	1	2	3	4	5	6	7	8	9	10	11	12	13	Effect	MAF DILI	MAF control	MAF IBS	MAF AFR	MAF AMR	MAF EAS	MAF EUR	MAF SAS
1:186641577		0	0	0	0	0	0	0	0	1	0	0	0	0	3'UTRV	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.00
1:186641682		0	0	1	0	0	0	0	0	0	0	0	0	0	3'UTRV	0.038	0.005	0.03	0.00	0.01	0.00	0.02	0.03
1:186642059		0	0	0	0	0	0	0	1	0	0	0	0	0	3'UTRV	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.00
1:186642429		0	0	0	0	0	0	0	0	1	0	0	0	0	3'UTRV	0.038	0.050	0.06	0.00	0.05	0.00	0.09	0.06
1:186642856		0	0	0	0	0	0	0	1	1	0	0	0	0	3'UTRV	0.077	0.000	0.00	0.08	0.00	0.00	0.00	0.00
1:186642987		1	0	0	0	0	0	0	0	0	0	0	0	0	3'UTRV	0.038	0.015	0.00	0.18	0.01	0.00	0.00	0.00
1:186643058		1	0	1	0	0	1	1	1	0	0	0	1	1	3'UTRV	0.269	0.205	0.33	0.64	0.37	0.20	0.31	0.40
1:186643204		1	0	0	0	0	0	0	1	0	0	0	0	0	3'UTRV	0.077	0.000	0.02	0.47	0.04	0.00	0.01	0.04
1:186643238		0	0	0	0	0	0	0	0	1	0	0	0	0	3'UTRV	0.038	0.010	0.00	0.00	0.01	0.00	0.01	0.00
1:186643803		0	0	0	0	0	0	0	0	0	0	0	1	0	SV	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.00
1:186645078		0	0	0	0	0	0	0	0	1	0	0	0	0	SV	0.038	0.005	0.00	0.08	0.00	0.00	0.00	0.00
1:186645669		0	0	0	0	0	0	0	0	1	0	0	0	0	SV	0.038	0.000	0.00	0.08	0.00	0.00	0.00	0.00
1:186645927		0	0	0	0	0	0	0	0	0	1	0	0	0	IV	0.038	0.105	0.12	0.36	0.19	0.04	0.12	0.16
1:186646005		0	0	0	0	0	0	0	0	1	0	0	0	0	MSV ^e	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.00
1:186647418		0	0	0	0	0	0	0	0	0	0	1	0	0	SV	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.00
1:186648157		0	0	0	0	0	0	0	0	0	1	0	0	0	IV	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.00
1:186648197		0	1	1	0	0	0	0	2	0	0	2	2	1	IV	0.346	0.180	0.15	0.01	0.11	0.04	0.18	0.06
1:186650163		1	0	0	0	0	0	0	0	0	0	0	0	0	RRV	0.038	0.000	0.00	0.10	0.01	0.00	0.00	0.00
1:186650214		0	0	0	0	0	0	0	0	0	0	0	0	0	RRV	0.077	0.000	0.15	0.08	0.00	0.00	0.00	0.00
1:186650321		0	0	1	0	0	0	0	0	0	2	0	0	1	RRV	0.154	0.190	0.00	0.35	0.21	0.04	0.15	0.19
1:186650688		1	0	0	0	0	0	0	0	0	0	0	0	0	RRV	0.038	0.000	0.15	0.10	0.01	0.00	0.00	0.00
1:186650751		1	0	0	1	0	0	0	0	0	0	0	0	0	RRV	0.077	0.215	0.15	0.08	0.26	0.48	0.19	0.13
1:186650846		0	0	0	0	0	0	0	0	0	0	2	0	0	RRV	0.154	0.235	0.11	0.16	0.18	0.05	0.13	0.16
1:186650857		0	1	0	0	0	0	0	0	0	0	0	0	0	RRV	0.038	0.000	0.00	0.00	0.00	0.00	0.00	0.01
1:186650877		0	0	1	0	0	0	0	0	0	0	0	0	0	RRV	0.038	0.005	0.05	0.27	0.04	0.00	0.03	0.03
1:186651296		0	0	0	0	0	0	0	0	0	2	0	0	0	RRV	0.077	0.015	0.00	0.00	0.07	0.03	0.01	0.07
1:186651571	unknown	0	1	0	0	0	0	0	0	0	0	0	0	0	RRV	0.038	0.000	-	-	-	-	-	-

MAF, Minor allele frequency referred to the GRCh37 assembly as shown in the 1,000 genomes website http://phases3browser.1000genomes.org/Homo_sapiens/Info/Index. 0, Non-mutated; 1, heterozygous; 2, Homozygous for the minor allele; UGV, Upstream gene variant; RRV, Regulatory region variant.; MSV, Missense variant; SV, Synonymous variant; IV, Intron variant; UTRV, Untranslated region variant; Populations, IBS correspond to the Iberian Populations in Spain (a subpopulation of Europeans); AFR, Africans; AMR, Admixed americans; EAS, East asians; EUR, Europeans; SAS, South asians, as described in detail in the website <http://phases3browser.1000genomes.org/Help/Faq?id=328>. Predicted consequences for missense variants:

^aNonsynonymous (W8R), SIFT score = 0.85 (tolerated, low confidence), PolyPhen score = 0 (unknown), Mutation Assessor = neutral.

^bNonsynonymous (P17L), SIFT score = 1.00 (tolerated, low confidence), PolyPhen score = 0 (unknown), Mutation Assessor = low impact.

^cNonsynonymous (R108Q), SIFT score = 0.12 (tolerated), PolyPhen score = 0.21 (benign), Mutation Assessor = medium impact.

^dNonsynonymous (K185T), SIFT score = 0.36 (tolerated), PolyPhen score = 0.007 (benign), Mutation Assessor = neutral.

^eNonsynonymous (R228H), SIFT score = 1.00 (tolerated), PolyPhen score = 0.002 (benign), Mutation Assessor = neutral.

TABLE 3 | Detailed genotype distribution for relevant SNPs.

Coordinate GRCh37.p13 (GCA_000001405.14)	rs ID	Effect	Patients with late-onset DILI N = 5 (Non carriers / heterozygous/ homozygous); MAF	Rest of DILI patients N = 8 (Non carriers / heterozygous/ homozygous); MAF	Control individuals N = 100 (Non carriers / heterozygous/ homozygous); MAF
9:125131832	rs10306225	UGV	3/0/2; 0.400	8/0/0; 0.000	100/0/0; 0.000
9:125133479	rs1236913	MSV	2/3/0; 0.300	8/0/0; 0.000	88/11/1; 0.065
9:125133507	rs3842787	MSV	3/2/0; 0.200	8/0/0; 0.000	89/10/1; 0.060
9:125140823	rs5787	MSV	4/0/1; 0.200	8/0/0; 0.000	100/0/0; 0.000
9:125143707	rs3842792	MSV	4/1/0; 0.100	8/0/0; 0.000	100/0/0; 0.000
1:186646005	rs3218622	MSV	4/1/0; 0.100	8/0/0; 0.000	100/0/0; 0.000

MAF, Minor allele frequency; UGV, Upstream gene variant; MSV, Missense variant.

TABLE 4 | Haplotype analysis.

Haplotype frequencies	rs10306225	rs1236913	rs3842787	rs5787	rs3842792	rs3218622	Frequency (Total)	Frequency (late-onset DILI cases)	Frequency (controls)
1	A	T	C	G	A	C	0.9048	0.300	0.935
2	A	C	T	G	A	C	0.0619	0.100	0.060
3	T	T	C	A	A	C	0.0095	0.200	NA
4	A	C	C	G	A	C	0.0048	NA	0.005
5	A	T	C	G	A	T	0.0048	0.100	NA
6	T	C	C	G	A	C	0.0048	0.100	NA
7	T	T	C	G	A	C	0.0048	0.100	NA
8	T	T	C	G	A	C	0.0048	0.100	NA
9	T	T	C	G	A	C	0.0048	0.100	NA
Haplotype association with late-onset DILI	rs10306225	rs1236913	rs3842787	rs5787	rs3842792	rs3218622	Frequency (Total)	OR (95% CI)	P-value
1	A	T	C	G	A	C	0.9048	1.00	—
2	A	C	T	G	A	C	0.0619	0.09 (0.01–1.43)	0.0910
Rare haplotypes	*	*	*	*	*	*	0.0333	0.00 (0.00–0.09)	0.0024

Global haplotype association $p < 0.0001$.

NA, not applicable; *any nucleotide.

Interestingly, most of the *PTGS1* and *PTGS2* SNPs included in the Illumina human exome chip or human core exome chip (Urban et al., 2012) are also absent in this study group. This raises doubts about the coverage of exome chips to identify genetic associations related to *PTGS1* and *PTGS2* genes.

In the whole population study, we identified 31 single nucleotide polymorphisms (SNPs) for *PTGS1*, including four non-synonymous SNPs. For *PTGS2* we identified 31 SNPs including one non-synonymous. We observed an increased frequency of *PTGS1* and *PTGS2* mutations among DILI patients, as compared to that observed in control individuals. Most of the SNPs identified in patients were rare among control individuals and were rare also according to the 1,000 genomes database (as shown in **Table 2**). All patients but one (case 1 in **Table 2**) had mutations at the *PTGS1* gene and all patients but one (case 5 in **Table 2**) had mutations at the *PTGS2* gene. **Table 3** summarizes the comparison of relevant SNPs across patients with late-onset DILI, the rest of DILI patients and control individuals.

DISCUSSION OF THE UNDERLYING PATHOPHYSIOLOGY AND THE NOVELTY OR SIGNIFICANCE OF THE CASE

The most remarkable findings in this study are the presence among DILI patients of SNPs causing alterations in transcription factor binding sites such as the *PTGS1* SNP rs10306225 (Agúndez et al., 2014), and the *PTGS2* SNPs rs4648253, rs689466, and rs20417, as well as non-synonymous SNPs such as *PTGS1* rs1236913 (W 8 R), rs3842787 (P 17 L), rs5787 (R 108 Q), rs3842792 (K 185 T), and *PTGS2* rs3218622 (R 228 H). These missense variants are extremely rare among European individuals (Agúndez et al., 2015). The putative effects of the most relevant SNPs shown in **Table 3** have been revised elsewhere (Agúndez et al., 2015). In brief, besides the rs10306225 SNP, which is a promoter variant that causes a modification in a CDX1 binding site (Agúndez et al., 2014), the rest of SNPs are non-synonymous. According to functional predictions and functional analyses (reviewed in Agúndez et al., 2015) the SNPs rs1236913, rs3842787 have a little functional effect,

although clinical associations for these SNPs with urticaria induced by NSAIDs (Cornejo-Garcia et al., 2012) and myocardial infarction/stroke (Lee et al., 2008; Lemaitre et al., 2009; Gao et al., 2014), respectively, have been proposed. The functional effect of the rs5787 SNP is unknown, although functional prediction suggests a mild functional impact (see **Table 2**), rs3842792 SNP is predicted as functional (**Table 2**), but *in vitro* findings suggest reduced functionality (Lee et al., 2007), and no functional impact for the PTGS2 SNP rs3218622 has been described.

No particular association of missense SNPs with culprit drug, age, gender, clinical presentation, type of liver injury, and severity of the disease was identified. However, as shown in **Table 1**, there is heterogeneity in the duration of treatment before DILI onset. This heterogeneity, rather than being a weakness, is a strong point in this study because it allowed discriminating the frequencies of PTGS gene variations in DILI patients with late and short-term onset. All the five DILI patients with the longest times to DILI onset (50 or more days; patients n° 3, 8, 9, 10, 12 in **Table 1**) had missense variants, and no patient with shorter time to DILI onset had such missense variants. The intergroup comparison values for carriers of any non-synonymous PTGS variants were as follows: Patients with late DILI onset (50 or more days) vs. the rest of DILI patients ($P < 0.001$). Patients with late DILI onset vs. control individuals ($P < 0.001$). By turn, no significant differences for carriers of non-synonymous PTGS variants were observed among patients with DILI onset shorter than 50 days and control subjects ($P = 0.325$). Haplotype analyses (**Table 4**), and linkage disequilibrium (LD) analyses (**Supplemental Table 2**), show that the risk is due to the presence of rare haplotypes (containing missense variants) in the group of patients with late-onset DILI, but it is not due to LD variations for these variants. The strong association observed in this report, although it is based in five cases only, suggests a relationship of non-synonymous PTGS gene variations with DILI onset after long-term NSAID therapy. This is a novel observation that has not been raised by previous studies. Although the putative role of PTGS gene variations has been explored using the Illumina human exome chip or human core exome chip, it is of note that chip coverage was very limited for PTGS genes (Urban et al., 2012). By turn, this study has complete

coverage thus allowing the identification of, as yet, disregarded SNPs. Another relevant difference with most DILI genetic studies is that in this report we stratified patients according to the time to onset. It cannot be ruled out heterogeneity in the etiopathogenesis of DILI, and it is conceivable that the mechanisms involved in DILI with a late onset might be different from those involved in immediate or short-latency reactions. This study, albeit with the inherent limitations of statistical power that case reports have, reinforces the view that a complete gene coverage and a detailed phenotype stratification of DILI patients could be essential to gain strength in further genetic association studies.

AUTHOR CONTRIBUTIONS

ML and EG-M participated in the design of the study, in data acquisition, and in critical revision for important intellectual content. AD, MB, and RA participated in the analysis and interpretation of the data and critical revision for important intellectual content. JA participated in the conception, design, data analysis and interpretation, the drafting of the manuscript and critical revision for important intellectual content. All authors approved the final version of the manuscript and all agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

FUNDING

Financed in part by grants PI12/00241, PI12/00378, PI12/00324, PI15/00303, and RETICS RD16/0006/0004 from Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spain, and IB16170, GR18145 from Junta de Extremadura, Spain. Financed in part with FEDER funds from the European Union.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00134/full#supplementary-material>

REFERENCES

- Agúndez, J. A. (2009). Recent advances in drug intolerance. *Curr. Drug Metab.* 10:946. doi: 10.2174/138920009790711869
- Agúndez, J. A., Blanca, M., Cornejo-Garcia, J. A., and Garcia-Martin, E. (2015). Pharmacogenomics of cyclooxygenases. *Pharmacogenomics* 16, 501–522. doi: 10.2217/pgs.15.6
- Agúndez, J. A., Gonzalez-Alvarez, D. L., Vega-Rodriguez, M. A., Botello, E., and Garcia-Martin, E. (2014). Gene variants and haplotypes modifying transcription factor binding sites in the human cyclooxygenase 1 and 2 (PTGS1 and PTGS2) genes. *Curr. Drug Metab.* 15, 182–195. doi: 10.2174/138920021502140327180336
- Agúndez, J. A., Lucena, M. I., and Martinez, C. (2011). Assessment of nonsteroidal anti-inflammatory drug-induced hepatotoxicity. *Expert Opin. Drug Metab. Toxicol.* 7, 817–828. doi: 10.1517/17425255.2011.574613
- Agúndez, J. A., Martinez, C., Perez-Sala, D., Carballo, M., Torres, M. J., and Garcia-Martin, E. (2009). Pharmacogenomics in aspirin intolerance. *Curr. Drug Metab.* 10, 998–1008. doi: 10.2174/138920009790711814
- Andrade, R. J., Agúndez, J. A., Lucena, M. I., Martinez, C., Cueto, R., and Garcia-Martin, E. (2009). Pharmacogenomics in drug induced liver injury. *Curr. Drug Metab.* 10, 956–970. doi: 10.2174/138920009790711805
- Cornejo-Garcia, J. A., Jagemann, L. R., and Blanca-Lopez, N. (2012). Genetic variants of the arachidonic acid pathway in non-steroidal anti-inflammatory drug-induced acute urticaria. *Clin. Exp. Allergy* 42, 1772–1781. doi: 10.1111/j.1365-2222.2012.04078.x
- Gao, L. C., Wang, D., and Liu, F. Q. (2014). Influence of PTGS1, PTGFR, and MRP4 genetic variants on intraocular pressure response to latanoprost in Chinese primary open-angle glaucoma patients. *Euro. J. Clin. Pharmacol.* 71, 43–50. doi: 10.1007/s00228-014-1769-8

- Kaliyaperumal, K., Grove, J. I., Delahay, R. M., Griffiths, W. J. H., Duckworth, A., and Aithal, G. P. (2018). Pharmacogenomics of drug-induced liver injury (DILI): molecular biology to clinical applications. *J. Hepatol.* 69, 948–957. doi: 10.1016/j.jhep.2018.05.013
- Lee, C. R., Bottone, F. G. Jr., and Krahn, J. M. (2007). Identification and functional characterization of polymorphisms in human cyclooxygenase-1 (PTGS1). *Pharmacogenet. Genomics* 17, 145–160. doi: 10.1097/01.fpc.0000236340.87540.e3
- Lee, C. R., North, K. E., Bray, M. S., Couper, D. J., Heiss, G., and Zeldin, D. C. (2008). Cyclooxygenase polymorphisms and risk of cardiovascular events: the Atherosclerosis Risk in Communities (ARIC) study. *Clin. Pharmacol. Ther.* 83, 52–60. doi: 10.1038/sj.clpt.6100221
- Lemaitre, R. N., Rice, K., and Marciano, K. (2009). Variation in eicosanoid genes, non-fatal myocardial infarction and ischemic stroke. *Atherosclerosis* 204, e58–63. doi: 10.1016/j.atherosclerosis.2008.10.011
- Lucena, M. I., Andrade, R. J., and Martinez, C. (2008). Glutathione S-transferase m1 and t1 null genotypes increase susceptibility to idiosyncratic drug-induced liver injury. *Hepatology* 48, 588–596. doi: 10.1002/hep.22370
- Lucena, M. I., Garcia-Martin, E., and Andrade, R. J. (2010). Mitochondrial superoxide dismutase and glutathione peroxidase in idiosyncratic drug-induced liver injury. *Hepatology* 52, 303–312. doi: 10.1002/hep.23668
- Nicoletti, P., Aithal, G. P., and Bjornsson, E. S. (2017). Association of liver injury from specific drugs, or groups of drugs, with polymorphisms in HLA and other genes in a genome-wide association study. *Gastroenterology* 152, 1078–1089. doi: 10.1053/j.gastro.2016.12.016
- Robles-Diaz, M., Medina-Caliz, I., Stephens, C., Andrade, R. J., and Lucena, M. I. (2016). Biomarkers in DILI: one more step forward. *Front. Pharmacol.* 7:267. doi: 10.3389/fphar.2016.00267
- Urban, T. J., Shen, Y., and Stolz, A. (2012). Limited contribution of common genetic variants to risk for liver injury due to a variety of drugs. *Pharmacogenet. Genomics* 22, 784–795. doi: 10.1097/FPC.0b013e3283589a76

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lucena, García-Martín, Daly, Blanca, Andrade and Agúndez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Association of Olanzapine Treatment Response in Han Chinese Schizophrenia Patients

Wei Zhou^{1†}, Yong Xu^{2†}, Qinyu Lv³, Yong-hui Sheng⁴, Luan Chen¹, Mo Li¹, Lu Shen¹, Cong Huai¹, Zhenghui Yi^{3*}, Donghong Cui^{3,5*} and Shengying Qin^{1,6*}

OPEN ACCESS

Edited by:

Amit V. Pandey,
University of Bern, Switzerland

Reviewed by:

Ming Ta Michael Lee,
Geisinger Health System,
United States
Vindhya Udhane,
Medical College of Wisconsin,
United States
Julio Benitez,
Universidad de Extremadura, Spain

*Correspondence:

Zhenghui Yi
yizhenghui1971@163.com
Donghong Cui
manyucc@126.com
Shengying Qin
chinsir@sjtu.edu.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Pharmacogenetics
and Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 03 September 2018

Accepted: 11 February 2019

Published: 04 March 2019

Citation:

Zhou W, Xu Y, Lv Q, Sheng Y-h,
Chen L, Li M, Shen L, Huai C, Yi Z,
Cui D and Qin S (2019) Genetic
Association of Olanzapine Treatment
Response in Han Chinese
Schizophrenia Patients.
Front. Pharmacol. 10:177.
doi: 10.3389/fphar.2019.00177

¹ Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Bio-X Institutes, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China, ² Department of Psychiatry, First Hospital of Shanxi Medical University, Taiyuan, China, ³ Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China, ⁴ Shandong Mental Health Center, Jinan, China, ⁵ Shanghai Key Laboratory of Psychotic Disorders, Shanghai, China, ⁶ The Third Affiliated Hospital, Guangzhou Medical University, Guangzhou, China

Olanzapine, a second-generation antipsychotic medication, plays a critical role in current treatment of schizophrenia (SCZ). It has been observed that the olanzapine responses in schizophrenia treatment are different across individuals. However, prediction of this individual-specific olanzapine response requires in-depth knowledge of biomarkers of drug response. Here, we performed an integrative investigation on 238 Han Chinese SCZ patients to identify predictive biomarkers that were associated with the efficacy of olanzapine treatment. This study applied HaloPlex technology to sequence 143 genes from 79 Han Chinese SCZ patients. Our result suggested that there were 12 single nucleotide polymorphisms (SNPs) had significant association with olanzapine response in Han Chinese SCZ patients. Using MassARRAY platform, we tested that if these 12 SNPs were also statistically significant in 159 other SCZ patients (independent cohort) and the combined 238 SCZ patients (composed of two tested cohorts). The result of this analysis showed that 2 SNPs were significantly associated with the olanzapine response in both independent cohorts (rs324026, $P = 0.023$; rs12610827, $P = 0.043$) and combined SCZ patient population (rs324026, adjusted $P = 0.014$; rs12610827, adjusted $P = 0.012$). Our study provides systematic analyses of genetic variants associated with olanzapine responses of Han Chinese SCZ patients. The discovery of these novel biomarkers of olanzapine-response will facilitate to advance future olanzapine treatment specific for Han Chinese SCZ patients.

Keywords: olanzapine, polymorphism, schizophrenia, pharmacogenetics, biomarker, association study

INTRODUCTION

Schizophrenia (SCZ) is a severe chronic neuropsychiatric illness. According to a survey carried out in 33 countries, about 15 out of 100,000 individuals were suffering from SCZ globally (McGrath et al., 2004). These SCZ patients were estimated to have higher risk of death (about 2.5 times more) compared to healthy individuals (McGrath et al., 2004; Saha et al., 2007). Patients with SCZ disorder require long-term treatments to prevent themselves from illness progression or symptom relapse

(Howes et al., 2015; Chong et al., 2016). Although the development of SCZ has long been regarded as caused by a combination of genetic and environmental factors, detailed pathophysiological mechanism of SCZ still remains unclear.

Second-generation antipsychotic (SGA) medications are widely considered as the most advanced and effective treatment for SCZ patients nowadays. These SGA includes olanzapine (OLA), risperidone and quetiapine (Owen et al., 2016). However, SCZ patients who received SGA treatments often experienced severe adverse drug reactions (ADRs) (Zhang and Malhotra, 2011). In fact, current SGA therapies could be seen as a subjective “trial and error” process. For some of the SCZ patients, these SGA treatments did not exert any therapeutic effects on their symptoms. This individual-specific response to SGA may be caused by the association between individual-specific genetic variation and the efficacy of SGAs (Zhang and Malhotra, 2011). In order to address such specificity of SGA response across different patients, scientists in pharmacogenomics field is now exploring the possibility of predicting drug response using individual-specific genetic signatures (Arranz and de Leon, 2007).

Olanzapine is one of the most commonly used SGAs. It has been shown to have relatively superior efficacy in various clinical trials compared to other SGAs (Lieberman et al., 2005; Leucht et al., 2009). It has been reported that OLA treatment had relatively low extrapyramidal side effects and better efficacy to minimize negative symptoms of SCZ patients when it is used in clinically effective doses (Meltzer, 1999). OLA binds to serotonin type 2 (5-HT₂) and dopamine (D₂) receptors with high affinity in patients' body. Diphosphate glucuronosyltransferases (UGT), a member of cytochrome P450 family and flavin-containing monooxygenase 1 (FMO1), catalyze the oxidative hepatic metabolism process of OLA (Ring et al., 1996; Kassahun et al., 1997; Linnet, 2002). On the other hand, due to the heterogeneity in different SCZ patients, not all patients respond to OLA treatment adequately well as we expected. Some patients who received OLA therapy even experienced severe adverse side effects that resulted in non-compliance with drug treatment (Zhang and Malhotra, 2011; Musil et al., 2015). If these SCZ patients with no other effective therapies specific for them they would have to face the coming disease progression, relapses and potential long-term hospitalizations (Robinson et al., 1999; King et al., 2014).

Although numerous studies have been performed on the factors that influence the therapeutic efficacy of OLA, there were very few of them focused on the individual-specific genetic biomarkers of OLA response (Söderberg and Dahl, 2013). In comparison, earlier attempts to search for biomarkers of OLA response focused mainly on the relationship between OLA response and its metabolic pathways, including glucuronidation, hydroxylation, *N*-demethylation and *N*-oxidation pathways (Laika et al., 2010; Haslemo et al., 2012; Mao et al., 2012; Söderberg et al., 2013; Brandl et al., 2015). A number of genetic variants, including *UGT2B10* rs61750900 (*UGT2B10**2) (Erickson-Ridout et al., 2011), *CYP1A2* rs762551 (*CYP1A2**1F) (Laika et al., 2010), *DRD3* rs6280 (Adams et al., 2008), *AHR* rs4410790 (Söderberg et al., 2013), *FMO3* K158–G308,

FMO1 rs12720462 (*FMO1**6) and *FMO1* rs7877 have been reported to play important roles to influence OLA metabolism (Soderberg et al., 2013). In addition, the drug response and the pharmacokinetics of OLA have also been found to associate with genetic elements that are not directly involved in the metabolic pathway of OLA (Lin et al., 2006; Meary et al., 2008; Cabaleiro et al., 2013; Yu et al., 2018). For example, *P*-glycoprotein, a membrane protein that pumps foreign substances out of cells and is regarded as element that is not directly related to OLA metabolic pathway, affects the penetration of OLA into the central nervous system (Lin et al., 2006). These discoveries suggested that a comprehensive study of OLA response requires clear understanding of the complicated biological network that is composed of enzymes involved in drug metabolism, drug transportation and drugs targeted receptors.

In this study, we investigated the associations between SNPs in 143 genes and the OLA response of 79 Han Chinese SCZ patients using target-sequencing technology. The newly found biomarkers was considered as genetic signature of drug responses to 8-week treatment with OLA and were validated in the other independent Han Chinese SCZ patient cohort.

MATERIALS AND METHODS

Subjects

In this study, we collected 2 independent sets of OLA response data from Han Chinese SCZ patients in order to validate our discoveries. We named the first set as ‘discovery cohort’ and other one as ‘independent cohort.’ The demographics and clinical details of the both sets of patients are demonstrated in **Table 1**.

The discovery cohort was composed of 79 recruited Han Chinese SCZ patients who had been treated with OLA from the Shanghai Mental Health Center of China. It comprised 37 males and 42 females. The mean age of them was 43.1 ± 18.3 years old (**Table 1**). SCZ of the patients was diagnosed according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) and confirmed by at least two experienced psychiatrists. Patients who had physical complications or other substance abuse were excluded from this investigation. Our analysis only considered patients who had not been previously treated with atypical antipsychotics and had not received any medication for more than 4 weeks before their enrollment in this study.

The independent cohort contained 159 recruited Han Chinese SCZ patients who were undergoing OLA monotherapy from the Shanghai Mental Health Center of China and the First Hospital of Shanxi Medical University. They were composed of 75 males and 84 females. The mean age of independent cohort was 38.5 ± 16.4 years (**Table 1**). This cohort of patients was used to validate the novel biomarkers found by the analysis of discovery cohort. The selection criteria were similar to the discovery cohort. Finally, patient data from discovery cohort and independent cohort were combined in order to gain a greater power in statistical analysis. Therefore, a total of 238 patients were used to perform another validation of the newly found biomarkers.

TABLE 1 | Demographic and clinical details of patients suffering from schizophrenia in three cohorts.

Baseline characteristics ^a	Discovery cohort		P-value	Independent cohort		P-value	Total cohort		P-value
	Good responders	Poor responders		Good responders	Poor responders		Good responders	Poor responders	
Number	38	41	—	119	40	—	157	81	—
Age (years)	45.4 ± 19.7	39.9 ± 16.2	0.15	39.5 ± 16.77	36.68 ± 15.4	0.35	41.1 ± 17.6	39.1 ± 16.2	0.41
Gender (male/female)	17/21	20/21	0.72	56/63	19/21	0.96	73/84	39/42	0.78
BMI, kg m ⁻²	21.6 ± 3.5	21.9 ± 3.6	0.84	22.8 ± 3.8	22.5 ± 3.4	0.71	22.5 ± 3.8	22.1 ± 3.4	0.46
PANSS total score T0	78.4 ± 18.0	73.8 ± 12.2	0.32	82.6 ± 23.3	78.0 ± 21.1	0.28	81.5 ± 22.1	75.9 ± 17.4	0.051
PANSS total score T1	45.5 ± 6.9	58.3 ± 8.4	<0.001	44.2 ± 10.1	63.3 ± 14.4	<0.001	44.5 ± 9.4	60.9 ± 12.1	<0.001

BMI, body mass index; T0 – baseline measurement, T1 – follow-up measurement (8 weeks). Bold font indicates statistically significant values ($P < 0.05$).

Values are shown as means ± standard deviation.

^aGender was analyzed by χ^2 test and other characteristics were analyzed by Student's *t*-test.

Clinical Assessment

Clinical effects of patients were evaluated using the Positive and Negative Syndrome Scale (PANSS) by two fully qualified psychiatrists during the 8 weeks of OLA treatment. The inter-rater reliability between the two raters was found to be high since the intraclass correlation coefficients (ICCs) was larger than 0.8. Based to Obermeier's method, patients were classified as good responders (reduction of PNASS score $\geq 50\%$) and poor responders (reduction of PNASS score $< 50\%$) for analysis (Obermeier et al., 2010). The initial daily dose of OLA was 10 mg per day and then it gradually increased to 15 mg per day within the 1st week. After that, the dosage was adjusted based on individual tolerance to the treatment. During the medication period, nursing staff closely monitored any medication compliance occurred in patients. No other drugs were administered when OLA monotherapy was performed, except for sennoside for constipation, flunitrazepam or lorazepam for acute insomnia and biperiden for any extrapyramidal side effects.

Ethics Statement

The study was approved by the Ethical Committee of Human Genetic Resources in Shanghai, China. All subjects or their legal guardians understood the procedure and had given written informed consent to their participation in this study according to the Declaration of Helsinki (Human, 1999).

Targeted Genes and Capture Design

One hundred and forty-three genes were selected for targeted sequencing were based on their involvement in drug metabolism (including genes encode for drug-metabolizing enzymes, drug-transporting enzymes and the receptors mediating drug responses) from PharmGKB database¹ and the relevant literatures (Arranz et al., 2011; Li and Bluth, 2011; Arranz et al., 2016) that reported potential genes related to SGAs efficacy. We aimed to study these genes in order to investigate novel biomarkers of OLA response. The details of these 143 genes are listed in **Supplementary Table S1**.

¹<https://www.pharmgkb.org/>

Sequencing probes for the 143 targeted genes were designed using Agilent's SureDesign tool². Targeted regions of these genes of interest included their coding regions ± 10 bp and untranslated regions (UTR) according to information from RefSeq, Ensembl, CCDS, and GENCODE databases (Harrow et al., 2012; Pruitt et al., 2014; Cunningham et al., 2015).

Library Preparation and Next Generation Sequencing

Genomic DNA was extracted from whole blood using a QIAamp DNA Blood Mini Kit (Qiagen GmbH, Hilden, Germany). The quantity and quality of the genomic DNA were measured by Nanodrop 2000 (Thermo Scientific, United States). Then we adjusted the genomic DNA to a final concentration of 100 ng/ μ l with high-purity water and stored at -20°C . Libraries were prepared with a HaloPlex Target Enrichment System Kit (Agilent Technologies, Santa Clara, CA, United States) following the manufacturer's instructions. Libraries were then quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies). Sequencing was performed with the HiSeq 2500 platform (Illumina, San Diego, CA, United States) using paired-end libraries (2×101 -bp).

Raw data were processed following standard protocols used in earlier reports (Gaynor et al., 2016). In short, raw image files were first converted to the FASTQ format and the reads were aligned to the human reference genome (hg19, GRCh37). SNPs were identified according to GATK standard hard filtering parameters (DePristo et al., 2011). On average, 99% of reads covered $>80\times$ and 81% $>200\times$, which suggested that the coverage was sufficiently high to detect variants with appropriate sensitivity. The program ANNOVAR was used to annotate SNVs that covered $>20\times$ according to the information from Ensembl Variation, dbSNP, and 1000genome database (Sherry et al., 2001; Abecasis et al., 2010; Wang et al., 2010; Flicek et al., 2012). Subsequently, individual and SNP-level quality controls were performed using PLINK (v1.07) software (Purcell et al., 2007). Data cleaning was performed according to the following criteria: genotypic call rate $< 95\%$, Hardy-Weinberg equilibrium (HWE) < 0.001 , and minor allele frequency (MAF) < 0.01 .

²www.agilent.com/genomics/suredesign

After we conducted quality controls of the sequences, there were 77 individuals (38 good responders and 39 poor responders) and 807 SNPs remained in our data for later analysis.

Validation Trial

We identified 12 SNPs that were significantly associated with OLA response using discovery cohort. Similar analysis was performed using both independent cohort and patient samples with two cohort combined in order to validate these newly found SNPs. These 12 SNPs were genotyped using the Sequenom MassARRAY platform (Agena Bioscience, San Diego, CA, United States) following manufacturer's instructions. MassARRAY primers were designed using a semi-automated software Assay Design Suite v2.0³. The primer sequences are listed in **Supplementary Table S2**. Data cleaning was performed according to the following criteria: genotypic call rate < 95%, Hardy-Weinberg equilibrium (HWE) < 0.001, and minor allele frequency (MAF) < 0.01. Similarly, to the earlier processing of discovery cohort, quality control was carried on and only the filtered data was used for our analysis.

Statistical Analyses

The demographic characteristics of the both 'good responder' and 'poor responder' groups were examined to confirm the homogeneity of the data used in our analysis. The data was found to have normal distribution, allowing student's *t*-tests to be performed on the obtained data (age, PANSS score, etc.). Gender differences were analyzed using the Chi-square test. SPSS software (version 11.0, Chicago, IL, United States) was used for all the statistical analyses in this study. The association between genotype and OLA response was assessed using logistic regression model by PLINK v1.07 software (Purcell et al., 2007). *P*-values were corrected using Bonferroni method for multiple testing adjustments. Two-tailed *P*-values of 0.05 were considered to be statistically significant. Power analysis was performed by the software GPower 3.1.

RESULTS

Patients Characteristics and Sequencing Profile

The demographic and clinical details of patient subjects included in this study are shown in **Table 1**. Among the 79 patients in the discovery cohort, 38 patients were defined to be good responders to OLA while 41 of them were poor responders. On the other hand, among 159 patients in independent cohort, 119 patients were good responders and 40 as poor responders to OLA. In the total cohort that comprised both sets, there were 157 good responders and 81 poor responders out of a total of 238 patient subjects. There was no statistically significant difference in the baseline characteristics between good responders and poor responders, except in the case of the PANSS total scores at the 8-week endpoint, meaning that the population was homogeneous (**Table 1**).

³<https://agenacx.com/>

Effects of Individual Polymorphisms on the OLA Response in the Discovery Cohort

Twelve out of 807 tested SNPs were found to be significantly associated with OLA response of 79 Han Chinese SCZ patients in discovery cohort. **Table 2** lists the results of the SNP association analysis of pharmacogenetic impact on OLA treatment response ($P < 0.05$). Two newly found variants were located on the exon's region (rs6280, $P = 0.026$, OR = 3.0, 95% CI = 1.14–7.87; rs2011404, $P = 0.04$, OR = 5.4, 95% CI = 1.08–26.93). The other 10 variants, which were not located in exons regions, were also found to be significantly associated with OLA response. However, there were no variants remained statistically significant after multiple-testing corrections (data not shown).

Verification of the Genetic Variants Associated With the Response to OLA in the Independent and Total Cohort

We used independent cohort and total sample population composed of both discovery cohort and independent cohort to validate the 12 SNP signatures found from the analysis of discovery cohort. The relevant clinical information of the data is shown in **Table 1**. A total of 12 SNPs was genotyped from patients in these two sets. In particular, SNP rs324026 and rs12610827 were found to be significantly associated with OLA treatment response in the independent cohort ($P = 0.023$ and $P = 0.023$). In the combined cohort, 4 SNPs displayed significant difference in OLA response between good responders and poor responders. We obtained strong evidence to conclude that these 2 variants (rs324026 and rs6280) in the dopamine receptor D3 (*DRD3*) gene were significantly associated with OLA response in Han Chinese SCZ patients ($P = 0.001$ and 0.0047). In addition, SNP rs12610827 (near to *PLK5*) and rs1543494 (located in *SUPT16H*) were also shown to be significantly associated with the OLA treatment response ($P = 0.001$ and 0.038). Detailed information of these significant SNPs is shown in **Table 3**.

Power Analysis

Post hoc power analysis revealed that the statistical power of the discovery cohort size ($n = 79$) in detecting a significant association ($P < 0.05$) was 0.76 with a medium effect size (Odds ratio = 2.0). The power of independent cohort size ($n = 159$) was 0.96 with the same effect size. These results indicated that the sample size in our study was sufficient to achieve a considerably low risk of a type II error.

DISCUSSION

To date, most pharmacogenomic studies on the OLA response focused on a few genes that are known to be relevant to OLA metabolism. Our study represents a more systematic survey of genetic biomarkers, including drug metabolic enzyme genes, receptor genes and other related genes. 143 genes of interest were sequenced using Next-generation sequencing technology for our association analysis. This study is one of the most comprehensive

TABLE 2 | 12 SNPs that were found to be significantly associated with responses to olanzapine treatment, by targeted sequencing.

Chr	SNP	Chr. Pos	MA	R_freq	NR_freq	OR (95% CI)	P	Gene	Annotation
2	rs4325692	211539363	A	0.07	0.00	NA	0.023	<i>CPS1</i>	Intronic
3	rs6280	113890815	C	0.31	0.18	3.0 (1.14–7.87)	0.026	<i>DRD3</i>	Non-synonymous
19	rs12610827	1538138	T	0.14	0.03	6.1 (1.23–30.30)	0.027	<i>PLK5</i> (neighboring gene)	Intergenic
14	rs7157205	43129681	C	0.20	0.07	3.9 (1.17–12.81)	0.027	<i>CTD-2307P3.1</i>	Intronic
7	rs1547470	151432869	T	0.14	0.04	4.9 (1.13–17.71)	0.032	<i>PRKAG2</i>	Intronic
3	rs324026	113891042	C	0.31	0.18	2.9 (1.08–7.54)	0.034	<i>DRD3</i>	Intronic
14	rs1543494	21829404	T	0.10	0.01	7.4 (0.86–63.22)	0.035	<i>SUPT16H</i>	Intronic
2	rs2011404	234627937	T	0.12	0.03	5.4 (1.08–26.93)	0.040	<i>UGT1A4</i>	Synonymous
19	rs2232778	8370120	T	0.27	0.39	0.3 (0.11–0.95)	0.040	<i>CD320</i>	3'UTR
8	rs4737005	39900682	T	0.08	0.20	0.3 (0.09–0.97)	0.045	<i>IDO2</i> (neighboring gene)	Intergenic
10	rs1831019	120691814	A	0.18	0.08	3.0 (1.01–9.02)	0.048	<i>LDHAP5</i> (neighboring gene)	Intergenic
10	rs9422807	126093991	T	0.03	0.11	0.23 (0.05–1.12)	0.049	<i>OAT</i>	Intronic

Chr, chromosome; SNP, single-nucleotide polymorphism; Chr. Pos, chromosome position; MA, minor allele; R_freq, responder frequency; NR_freq, non-responder frequency; OR, odds ratio; CI, confidence interval.

TABLE 3 | Validation of SNPs associated with the olanzapine response.

Chr	SNP	Chr. Pos	MA	Total cohort			Independent cohort		Gene	Functional consequence
				OR (95% CI)	P	Adjust P ^a	OR (95% CI)	P		
3	rs324026	113891042	C	2.35 (1.41–3.93)	0.001	0.014	2.11 (1.1–4.0)	0.023	<i>DRD3</i>	Intronic
19	rs12610827	1538138	T	3.9 (1.81–8.45)	0.00098	0.012	2.49 (1.0–6.0)	0.043	<i>PLK5</i> (neighboring gene)	–
3	rs6280	113890815	C	2.1 (1.25–3.46)	0.0047	0.057	1.85 (0.98–3.51)	0.058	<i>DRD3</i>	Missense
14	rs1543494	21829404	T	2.48 (1.05–5.86)	0.038	0.46	1.42 (0.55–3.67)	0.47	<i>SUPT16H</i>	Intronic

Chr, chromosome; SNP, single-nucleotide polymorphism; Chr. Pos, chromosome position; MA, minor allele; OR, odds ratio; CI, confidence interval; ^aP were adjusted by Bonferroni method. Bold font indicates statistically significant values.

pharmacogenetic analyses of association between SNP variants and OLA response.

Our result suggested that SNP rs324026 in *DRD3* gene had significant association with OLA response using independent cohort. This difference still remained significant in the total cohort comprised 2 cohorts even after Bonferroni correction. However, the other variant rs6280 was only found to be significantly associated with an 8-week treatment of OLA response in the combined cohort population and did not have evidence to have significant associations with OLA response in independent cohort. This inconsistent result may be caused by the small sample set we tested. Notably, these 2 SNPs both exhibited strong linkage disequilibrium ($r^2 > 0.9$) in the HaploReg database (Ward and Kellis, 2012). Therefore, both rs6280 and rs324026 may serve as biomarkers of OLA treatment response.

It is known that rs6280 mutation leads to a glycine for serine substitution and is associated with altered dopamine binding affinity. This glycine variant had been suggested to be able to increase the densities of the dopamine receptor D3 (DRD-3) in some areas in human brain (Jeanneteau et al., 2006). Adams et al. (2008) reported that DRD-3 gly/gly genotype and other polymorphisms in linkage disequilibrium with ser-9-gly variant were significantly associated with an increase in PANSS total score. Therefore, we concluded that our result was consistent with the earlier discoveries of the association between ser-9-gly variant and clozapine, which is the most similar receptor binding profile to OLA. Cerrato et al. (2017) surveyed 65 papers and found that

rs6280 was successfully replicated as prognostic biomarkers of clozapine efficacy. In contrast, rs324026 variant had never been reported to affect the therapeutic efficacy of OLA. Rs324026 is located next to exon 5. Our analysis results suggested that we could only find evidence of significant association between SNP rs324026 and OLA efficacy after Bonferroni correction in the combined sample with both discovery and independent cohorts. Additionally, individuals with C alleles of rs324026 generally experience significantly better efficacy of OLA treatment.

In this study, SNP rs12610827 was validated in patients from independent cohort and its association with OLA response remained significant for multiple testing in the total cohort after Bonferroni correction. Rs12610827 variant is located near the *PLK5* gene. Polo-like kinases (Plks) family, consisted of 5 members (Plk1–Plk5), is traditionally regarded to play an important part only in cell cycle progression. However, mounting evidence showed that Plk2 and Plk5 are also closely involved in neuron biology (de Carcer et al., 2011a). It had been suggested that Plk2 modulates neurite formation in response to activities of brain-derived growth factor (BDGF) (Inglis et al., 2009). Additionally, Plk5 was highly expressed in the central nervous system and it serve as a Plk2-like role in the cerebellum according earlier report (de Carcer et al., 2011b). It had been suggested that Plk5 was regulated by CpG methylation of the promoter region on the transcriptional level (de Carcer et al., 2011a). The level of *PLK5* gene expression may be influenced by the methylation status of this variant. In this study, Han Chinese SCZ patients who

carried allele T in SNP rs12610827 showed more good response in OLA treatment. However, no previous reports have found such association between *PLK5* variants and drug response. Therefore, we believe rs12610827 variant is worthy of further investigations in order to verify its influence on OLA response in the future.

This study has several limitations. First, our analysis did not consider some other genes that may have significant association with antipsychotic. Therefore, our results may have neglected some important biomarkers of OLA response due to this incomplete gene collection. Secondly, a number of identified SNP associations failed to stay statistically significant after Bonferroni corrections. This may be caused by over-correction because sample size was relatively small. Therefore, employing strict multiple corrections such as Bonferroni to the data may be too harsh for this specific study.

CONCLUSION

In sum, we performed a comprehensive study on 238 Han Chinese SCZ patients in order to identify potential biomarkers of Han Chinese-specific OLA responses. The result showed that 143 genes were significantly associated with OLA. In addition, 2 variants (rs324026 and rs12610827) were found to have significant association with the OLA response. Future investigations with larger sample sizes and high-throughput methods such as high-density SNP arrays and whole exome sequencing are warranted to find more biomarkers to predict the efficacy of OLA in the Han Chinese population.

DATA AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

- Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., and Gibbs, R. A. (2010). A map of human genome variation from population-scale sequencing. *Nature* 7319, 1061–1073. doi: 10.1038/nature09534
- Adams, D. H., Close, S., Farmen, M., Downing, A. M., Breier, A., and Houston, J. P. (2008). Dopamine receptor D3 genotype association with greater acute positive symptom remission with olanzapine therapy in predominately caucasian patients with chronic schizophrenia or schizoaffective disorder. *Hum. Psychopharmacol.* 4, 267–274. doi: 10.1002/hup.930
- Arranz, M. J., and de Leon, J. (2007). Pharmacogenetics and pharmacogenomics of schizophrenia: a review of last decade of research. *Mol. Psychiatry* 8, 707–747. doi: 10.1038/sj.mp.4002009
- Arranz, M. J., Gallego, C., Salazar, J., and Arias, B. (2016). Pharmacogenetic studies of drug response in schizophrenia. *Expert Rev. Precis. Med. Drug Dev.* 1, 79–91. doi: 10.1080/23808993.2016.1140554
- Arranz, M. J., Rivera, M., and Munro, J. C. (2011). Pharmacogenetics of response to antipsychotics in patients with schizophrenia. *CNS Drugs* 25, 933–969. doi: 10.2165/11595380-000000000-00000
- Brandl, E. J., Chowdhury, N. I., Tiwari, A. K., Lett, T. A. P., Meltzer, H. Y., and Kennedy, J. L. (2015). Genetic variation in CYP3A43 is associated with response to antipsychotic medication. *J. Neural Transm.* 1, 29–34. doi: 10.1007/s00702-014-1298-8
- Cabaleiro, T., Lopez-Rodriguez, R., Ochoa, D., Roman, M., Novalbos, J., and Abad-Santos, F. (2013). Polymorphisms influencing olanzapine metabolism and adverse effects in healthy subjects. *Hum. Psychopharmacol.* 3, 205–214. doi: 10.1002/hup.2308
- Cerrato, F., Guizzaro, L., Scudellari, P., Atti, A. R., and De Ronchi, D. (2017). Predictive biomarkers in clozapine-treated patients: assessment of the evidences and suggestion for research methodology. *Eur. Psychiatry* 41:S378. doi: 10.1016/j.eurpsy.2017.02.403
- Chong, H. Y., Teoh, S. L., Wu, D. B., Kotirum, S., Chiou, C. F., and Chaiyakunapruk, N. (2016). Global economic burden of schizophrenia: a systematic review. *Neuropsychiatr. Dis. Treat.* 12, 357–373. doi: 10.2147/NDT.S96649
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., and Brent, S. (2015). Ensembl 2015. *Nucleic Acids Res.* 43, D662–D669. doi: 10.1093/nar/gku1010
- de Carcer, G., Escobar, B., Higuero, A. M., Garcia, L., Anson, A., and Perez, G. (2011a). Plk5, a polo box domain-only protein with specific roles in neuron differentiation and glioblastoma suppression. *Mol. Cell. Biol.* 6, 1225–1239. doi: 10.1128/MCB.00607-10
- de Carcer, G., Manning, G., and Malumbres, M. (2011b). From Plk1 to Plk5: functional evolution of polo-like kinases. *Cell Cycle* 14, 2255–2262.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., and Hartl, C. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 5, 491–498. doi: 10.1038/ng.806

AUTHOR CONTRIBUTIONS

WZ and YX performed the experiments and wrote the manuscript. CH and ML aided in processing the data. QL, Y-hS, and YX aided in the collection of the materials. LC and LS helped in revising the manuscript. ZY, DC, and SQ designed and revised the manuscript.

FUNDING

This work was supported by grants from the 863 Program (Grant Numbers 2012AA02A515 and 2012AA021802), the National Natural Science Foundation of China (Grant Numbers 81421061, 81273596, J1210047, 30900799, 81361120389, 30972823, 81671326, and 81671336), National Key Research and Development Program (Grant Numbers 2016YFC0905000, 2016YFC0905002, 2016YFC1200200, 2016YFC0906400, and 2017YFC0909200), The 4th Three-Year Action Plan for Public Health of Shanghai (the Project No. 15GWZK0101), Shanghai Key Laboratory of Psychotic Disorders (Grant Number 13dz2260500), Public Science and Technology Research Funds (Grant Number 201210056), The Fourth Round of Shanghai Three-Year Action Plan on Public Health Discipline and Talent Program: Women and Children's Health (Grant Number 15GWZK0401), the Shanghai Jiao Tong University Interdisciplinary Research fund, and the Shanghai Leading Academic Discipline Project (Grant Number B205).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00177/full#supplementary-material>

- Erickson-Ridout, K. K., Zhu, J., and Lazarus, P. (2011). Olanzapine metabolism and the significance of UGT1A448V and UGT2B1067Y variants. *Pharmacogenet. Genomics* 9, 539–551. doi: 10.1097/FPC.0b013e328348c76b
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., and Carvalho-Silva, D. (2012). Ensembl 2012. *Nucleic Acids Res.* 40, D84–D90. doi: 10.1093/nar/gkr991
- Gaynor, S. C., Breen, M. E., Monson, E. T., de Klerk, K., Parsons, M., and DeLuca, A. P. (2016). A targeted sequencing study of glutamatergic candidate genes in suicide attempters with bipolar disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 8, 1080–1087. doi: 10.1002/ajmg.b.32479
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., and Kokocinski, F. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 9, 1760–1774. doi: 10.1101/gr.135350.111
- Haslemo, T., Loryan, I., Ueda, N., Mannheimer, B., Bertilsson, L., and Ingelman-Sundberg, M. (2012). UGT1A4*3 encodes significantly increased glucuronidation of olanzapine in patients on maintenance treatment and in recombinant systems. *Clin. Pharmacol. Ther.* 2, 221–227. doi: 10.1038/clpt.2012.46
- Howes, O., McCutcheon, R., and Stone, J. (2015). Glutamate and dopamine in schizophrenia: an update for the 21st century. *J. Psychopharmacol.* 2, 97–115. doi: 10.1177/0269881114563634
- Human, D. (1999). Declaration of Helsinki. *Lancet* 353:1888. doi: 10.1016/S0140-6736(05)75101-1
- Inglis, K. J., Chereau, D., Brigham, E. F., Chiou, S. S., Schobel, S., and Frigon, N. L. (2009). Polo-like kinase 2 (PLK2) phosphorylates alpha-synuclein at serine 129 in central nervous system. *J. Biol. Chem.* 5, 2598–2602. doi: 10.1074/jbc.C800206200
- Jeanneteau, F., Funalot, B., Jankovic, J., Deng, H., Lagarde, J. P., and Lucotte, G. (2006). A functional variant of the dopamine D3 receptor is associated with risk and age-at-onset of essential tremor. *Proc. Natl. Acad. Sci. U.S.A.* 28, 10753–10758. doi: 10.1073/pnas.0508189103
- Kassahun, K., Mattiuz, E., and Nyhart, E. (1997). Disposition and biotransformation of the antipsychotic agent olanzapine in humans. *Drug Metab. Dispos.* 25, 81–93.
- King, D., Knapp, M., Patel, A., Amadeo, F., Tansella, M., and Schene, A. (2014). The impact of non-adherence to medication in patients with schizophrenia on health, social care and societal costs: Analysis of the QUATRO study. *Epidemiol. Psychiatr. Sci.* 1, 61–70. doi: 10.1017/S2045796013000097
- Laika, B., Leucht, S., and Heres, S. (2010). Pharmacogenetics and olanzapine treatment: CYP1A2*1F and serotonergic polymorphisms influence therapeutic outcome. *Pharmacogenomics J.* 10, 20–29. doi: 10.1038/tpj.2009.32
- Leucht, S., Corves, C., Arbter, D., Engel, R., Li, C., and Davis, J. (2009). Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet* 373, 31–41. doi: 10.1016/S0140-6736(08)61764-X
- Li, J., and Bluth, M. H. (2011). Pharmacogenomics of drug metabolizing enzymes and transporters: implications for cancer therapy. *Pharmacogenomics Pers. Med.* 4, 11–33. doi: 10.2147/PGPM.S18861
- Lieberman, J., Stroup, T., and McEvoy, J. (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* 353, 1209–1223. doi: 10.1056/NEJMoa051688
- Lin, Y., Ellingrod, V. L., Bishop, J. R., and Miller, D. D. (2006). The relationship between P-glycoprotein (PGP) polymorphisms and response to olanzapine treatment in schizophrenia. *Ther. Drug Monit.* 5, 668–672. doi: 10.1097/01.fdt.0000246761.82377.a6
- Linnet, K. (2002). Glucuronidation of olanzapine by cDNA-expressed human UDP-glucuronosyltransferases and human liver microsomes. *Hum. Psychopharmacol.* 5, 233–238. doi: 10.1002/hup.403
- Mao, M., Skogh, E., and Scordo, M. (2012). Interindividual variation in olanzapine concentration influenced by UGT1A4 L48V polymorphism in serum and upstream FMO polymorphisms in cerebrospinal fluid. *J. Clin. Psychopharmacol.* 32, 287–289. doi: 10.1097/JCP.0b013e31824997a8
- McGrath, J., Saha, S., Welham, J., El, S. O., MacCauley, C., and Chant, D. (2004). A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med.* 2:13. doi: 10.1186/1741-7015-2-13
- Meary, A., Brousse, G., Jamain, S., Schmitt, A., Szoke, A., and Schurhoff, F. (2008). Pharmacogenetic study of atypical antipsychotic drug response: involvement of the norepinephrine transporter gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 4, 491–494. doi: 10.1002/ajmg.b.30635
- Meltzer, H. (1999). The role of serotonin in antipsychotic drug action. *Neuropsychopharmacology* 21, 106S–115S. doi: 10.1016/S0893-133X(99)00046-9
- Musil, R., Obermeier, M., Russ, P., and Hamerle, M. (2015). Weight gain and antipsychotics: a drug safety review. *Expert Opin. Drug Saf.* 1, 73–96. doi: 10.1517/14740338.2015.974549
- Obermeier, M., Mayr, A., and Schennach-Wolff, R. (2010). Should the PANSS be rescaled? *Schizophr. Bull.* 36, 455–460. doi: 10.1093/schbul/sbp124
- Owen, M. J., Sawa, A., and Mortensen, P. B. (2016). Schizophrenia. *Lancet* 388, 86–97. doi: 10.1016/S0140-6736(15)01121-6
- Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., and Ermolaeva, O. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 42, D756–D763. doi: 10.1093/nar/gkt1114
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., and Bender, D. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 3, 559–575. doi: 10.1086/519795
- Ring, B. J., Catlow, J., Lindsay, T. J., Gillespie, T., Roskos, L. K., and Cerimele, B. J. (1996). Identification of the human cytochromes P450 responsible for the in vitro formation of the major oxidative metabolites of the antipsychotic agent olanzapine. *J. Pharmacol. Exp. Ther.* 2, 658–666.
- Robinson, D., Woerner, M., Alvir, J., Bilder, R., Goldman, R., and Geisler, S. (1999). Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. *Arch. Gen. Psychiatry* 56, 241–247. doi: 10.1001/archpsyc.56.3.241
- Saha, S., Chant, D., and McGrath, J. (2007). A systematic review of mortality in schizophrenia - Is the differential mortality gap worsening over time? *Arch. Gen. Psychiatry* 10, 1123–1131. doi: 10.1001/archpsyc.64.10.1123
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., and Smigielski, E. M. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 1, 308–311. doi: 10.1093/nar/29.1.308
- Söderberg, M. M., and Dahl, M. (2013). Pharmacogenetics of olanzapine metabolism. *Pharmacogenomics* 11, 1319–1336. doi: 10.2217/pgs.13.120
- Söderberg, M. M., Haslemo, T., Molden, E., and Dahl, M. (2013). Influence of CYP1A1/CYP1A2 and AHR polymorphisms on systemic olanzapine exposure. *Pharmacogenet. Genomics* 5, 279–285. doi: 10.1097/FPC.0b013e3283602876
- Soderberg, M. M., Haslemo, T., Molden, E., and Dahl, M. L. (2013). Influence of FMO1 and 3 polymorphisms on serum olanzapine and its N-oxide metabolite in psychiatric patients. *Pharmacogenomics J.* 6, 544–550. doi: 10.1038/tpj.2012.47
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 16:e164. doi: 10.1093/nar/gkq603
- Ward, L. D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934. doi: 10.1093/nar/gkr917
- Yu, H., Yan, H., Wang, L., Li, J., Tan, L., and Deng, W. (2018). Five novel loci associated with antipsychotic treatment response in patients with schizophrenia: a genome-wide association study. *Lancet Psychiatry* 5, 327–338. doi: 10.1016/S2215-0366(18)30049-X
- Zhang, J., and Malhotra, A. (2011). Pharmacogenetics and antipsychotics: therapeutic efficacy and side effects prediction. *Expert Opin. Drug Metab. Toxicol.* 1, 9–37. doi: 10.1517/17425255.2011.532787

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhou, Xu, Lv, Sheng, Chen, Li, Shen, Huai, Yi, Cui and Qin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Actionable Pharmacogenetic Variation in the Slovenian Genomic Database

Keli Hočevar*, Aleš Maver and Borut Peterlin*

Clinical Institute of Medical Genetics, University Medical Centre Ljubljana, Ljubljana, Slovenia

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Vanessa Gonzalez-Covarrubias,
Instituto Nacional de Medicina
Genómica (INMEGEN), Mexico
Collet Dandara,
University of Cape Town, South Africa

*Correspondence:

Keli Hočevar
kelihoevar@gmail.com
Borut Peterlin
borut.peterlin@guest.arnes.si

Specialty section:

This article was submitted to
Pharmacogenetics
and Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 19 September 2018

Accepted: 26 February 2019

Published: 14 March 2019

Citation:

Hočevar K, Maver A and
Peterlin B (2019) Actionable
Pharmacogenetic Variation
in the Slovenian Genomic Database.
Front. Pharmacol. 10:240.
doi: 10.3389/fphar.2019.00240

Background: Genetic variability in some of the genes that affect absorption, distribution, metabolism, and elimination (“pharmacogenes”) can significantly influence an individual’s response to the drug and consequently the effectiveness of treatment and possible adverse drug events. The rapid development of sequencing methods in recent years and consequently the increased integration of next-generation sequencing technologies into the clinical settings has enabled extensive genotyping of pharmacogenes for personalized treatment. The aim of the present study was to investigate the frequency and variety of potentially actionable pharmacogenetic findings in the Slovenian population.

Methods: De-identified data from diagnostic exome sequencing in 1904 cases submitted to our institution were analyzed for variants within 293 genes associated with drug response. Filtered variants were classified according to population frequency, variant type, the functional impact of the variant, pathogenicity predictions and characterization in the Pharmacogenomics Knowledgebase (PharmGKB) and ClinVar.

Results: We observed a total of 24 known actionable pharmacogenetic variants (PharmGKB 1A or 1B level of evidence), comprising approximately 26 drugs, of which, 12 were rare, with the population frequency below 1%. Furthermore, we identified an additional 61 variants with PharmGKB 2A or 2B clinical annotations. We detected 308 novel/rare potentially actionable variants: 177 protein-truncating variants and 131 missense variants predicted to be pathogenic based on several pathogenicity predictions.

Conclusion: In the present study, we estimated the burden of pharmacogenetic variants in nationally based exome sequencing data and investigated the potential clinical usefulness of detected findings for personalized treatment. We provide the first comprehensive overview of known pharmacogenetic variants in the Slovenian population, as well as reveal a great proportion of novel/rare variants with a potential to influence drug response.

Keywords: next-generation sequencing, pharmacogenomics, personalized medicine, Slovenian population, PharmGKB

INTRODUCTION

Genetic variation in genes associated with drug pharmacokinetics (i.e., absorption, distribution, metabolism, elimination) or pharmacodynamics (e.g., alerting a drug's target or perturbing biological pathways) can significantly contribute to individual responsiveness to drugs, and thus on the therapeutic efficacy and toxicity (Dunnenberger et al., 2015; Relling and Evans, 2015). As an integral part of personalized medicine, pharmacogenomics has great potential to enhance clinical benefit, decrease adverse drug reactions and cost of treatment by optimizing drug selection and dosing for an individual. The rapidly dropping cost of next-generation sequencing within recent years, and consequently, the increased integration of these technologies into the clinical settings offered an unprecedented opportunity for extensive genotyping of pharmacogenes. Exome sequencing is a powerful tool for gaining insight into both common and rare coding variation. However, presently, there are no comprehensive studies regarding the application of exome sequencing data for reporting of pharmacogenetic variants. Therefore, there are two challenges that arise while analyzing exome-sequencing data. First, how and which variants with established pharmacogenomic effects should be reported, and second, how to evaluate novel putatively functional variants or variants in genes with less established pharmacogenetic function.

The lack of a comprehensive overview of the distribution of rare variants in pharmacogenes among different ethnic groups and the lack of knowledge about their functional consequences and clinical actionability additionally limit the fully integrated use of pharmacogenomics in a routine clinical practice. Therefore, pharmacogenomics usually remains focused on a limited number of common variants in a small number of genes, the detection of which is primarily based on targeted gene panels or genotyping arrays, such as AmpliChip CYP450 test (Roche) or Affymetrix DMET Plus Assay (Potamias et al., 2014). However, these approaches do not consider the complete heterogeneity of the variation within pharmacogenes and do not address the issue of rare and private variants with potentially large effects. Moreover, previous studies have shown that the vast majority of protein-coding variation is rare, previously unknown, population-specific and enriched for deleterious alleles (Nelson et al., 2012; Tennessen et al., 2012; Gordon et al., 2014; Fujikura et al., 2015). Thus, it is likely that rare variation importantly contributes to some currently unexplained differences in pharmacological responsiveness and metabolism. Consistent with this notion, recent research highlighted that rare variants account for 30–40% of the functional variability in the pharmacogenes (Kozyra et al., 2016). In the study by Ramsey et al. (2012), authors

showed that rare variants account for 17.8% of the variability attributed to *SLCO1B1*, a gene associated with methotrexate clearance and disposition of many other medications including statins and irinotecan.

With an objective to enable the clinical use of pharmacogenomics, projects like eMERGE are systematically documenting and evaluating both common and rare variants in pharmacogenes, thus creating clinically useful electronic networks of pharmacogenetic variation (Rasmussen-Torvik et al., 2014). Additionally, the Clinical Pharmacogenetics Implementation Consortium (CPIC¹) (Caudle et al., 2014), a shared project between the Pharmacogenomics Knowledgebase (PharmGKB²) (Whirl-Carrillo et al., 2012) and the Pharmacogenomics Research Network (PGRN) (Shuldiner et al., 2013), started to develop peer-reviewed, evidence-based guidelines for specific gene/drug combinations. By September 2018 CPIC published 65 dosing guidelines covering 15 genes and 38 drugs³. The efforts to facilitate implementation have also been undertaken by other nationwide networks such as the Royal Dutch Association for the Advancement of Pharmacy and Canadian Pharmacogenomics Network for Drug Safety (Ross et al., 2010). Currently, 23 different genes have described actionable variants (corresponding to PharmGKB level 1A or 1B of evidence) for germline pharmacogenomics (last accessed on September 9th, 2018).

However, there are presently no consensus recommendations on which pharmacogenetic findings should be actively sought and reported back to patients when analyzing exome or genome sequencing data. Nevertheless, the potential usefulness of pharmacogenomic findings in the exome sequencing data has recently been implicated. In the study of Lee et al. (2016), 21 potentially useful PharmGKB actionable variants (1A and 1B) were identified in 645 individuals who have undergone clinical exome sequencing. In a related study by Cousin et al. (2017), secondary pharmacogenetic findings from clinical whole exome sequencing (WES) testing were reported in a cohort of 94 primarily pediatric patients referred for a suspected genetic disorder. The study results showed that 91% of patients had at least one pharmacogenetic variant allele in *CYP2C19*, *CYP2C9*, and *VKORC1* genes and that 20% of them had potential immediate implications on current medication use. A study on 60,706 human exomes from ExAC population dataset further estimated the prevalence of common as well as rare functional variants in 806 drug-related genes and its implications for 1236 FDA approved drugs. The extended exome data analysis revealed that four in five patients are likely to carry a variant with possibly functional effects (Schärfe et al., 2017).

As population data are specific and cannot be generalized, even within closely related European populations (Mizzi et al., 2017), we used a genomic database of 1904 Slovenian individuals to comprehensively assess the population burden of pharmacogenetic variants. We conducted a nationally based survey of genetic variation within 293 genes, known to influence

Abbreviations: ABC, ATP-binding cassette; ACMG, American College of Medical Genetics and Genomics; BWA, Burrows-Wheeler algorithm; CADD, Combined Annotation-Dependent Depletion Score; CNV, copy number variation; CPIC, Clinical Pharmacogenetics Implementation Consortium; CYP, cytochrome P450 superfamily; FDA, Food and Drug Administration; MAF, minor allele frequency; MAF_{Slo}, minor allele frequencies for the Slovenian population; gnomAD, the Genome Aggregation Database; PGRN, Pharmacogenomics Research Network; PharmGKB, Pharmacogenomics Knowledgebase; SLC, solute carrier; TCAs, tricyclic antidepressants; UGTs, UDP-glucuronosyltransferases; UCSC, University of California Santa Cruz.

¹www.eu-pic.net

²www.pharmgkb.org

³www.pharmgkb.org/guidelines

drug response. Our additional motivation was to assess the applicability of the pharmacogenetic reporting as a part of the routine analysis of exomes and to gain insight into the opportunities and challenges that arise. Accordingly, we analyzed (1) which pharmacogenomic variants could be covered with exome sequencing data, (2) the frequency of known actionable variants in the Slovenian population, (3) and the frequency of rare variants with putative functional impacts and the possibilities for their interpretation.

MATERIALS AND METHODS

Participants

Exome datasets from 1904 patients who were referred to the Clinical Institute of Medical Genetics, University Medical Centre, Ljubljana, Slovenia from July 2014 to October 2017, and have undergone clinical (Illumina TruSight One panel, targeting 4813 genes associated with Mendelian disorders) or whole exome sequencing (Agilent SureSelect All Exon V5 or Illumina Nextera coding exome capture), were recruited for this analysis. All patients gave informed consent for participation in accordance with the Declaration of Helsinki. The study was approved by the Slovenian National Medical Ethics Committee (0120-561/2016). Data were de-identified and phenotype data of the patients were not available.

Panel Design

The panel of genes analyzed in the present study was selected based to include genes that were captured with all of the used protocols (Illumina TruSight One protocol, Nextera Coding Exome, and the Agilent SureSelect All Exon v5 protocol). We have established the pharmacogenetic list of 293 genes associated with pharmacological impacts, which is based on 33 genes from VeraCode® ADME Core Panel Assay (Illumina) and supplemented with 260 additional genes from PharmaADME (198 genes), PharmGKB (37 genes), and eMERGE-PGx (Sphinx) (25 genes) websites (**Supplementary Table S1**). Combining clinically relevant genes from these sources ensures that our gene set covers the majority of the key genes currently reviewed in pharmacogenomics research that are also captured with both – the clinical and whole exome sequencing.

Exome Sequencing

Of the 1904 samples sequenced, the majority (1,582 samples, 83.1%) of the samples were enriched using the Illumina TruSight One protocol, followed by Nextera Coding Exome (188 samples, 9.9%) and the remaining were analyzed using the Agilent SureSelect All Exon v5 (134 samples, 7.0%) protocol. Raw sequence files were processed using a custom exome analysis pipeline, based upon GATK best practices backbone. Reads were aligned to UCSC hg19 human reference genome assembly using Burrows-Wheeler (BWA) algorithm and duplicate sequences were removed using Picard MarkDuplicates, followed by base quality score recalibration, variant calling, variant quality score recalibration, and variant filtering using elements of the GATK toolset (Depristo et al., 2011). In all cases, we attained a minimum

median exome coverage of 60x, with over 95% of targets covered with at least 10× sequencing depth. Although the cytochrome genes are characterized by a high degree of homology, we were able to uniquely map over 90% of the reads in these regions, while the non-uniquely mapped reads were attributed mapping quality of 0 by the BWA. GATK variant caller did not emit sequence variants in these regions, thereby reducing the rate of low-quality variants in regions of high homology.

Variant Analysis

Variants were stored and annotated in our in-house variant collection and annotation system, which is based on vTools software. Variant effect predictions were made using snpEff (Cingolani et al., 2012) and ANNOVAR tools (Wang et al., 2010) and were based on RefSeq gene models (O'Leary et al., 2016), whereas annotations from dbSNP v141 were used for single nucleotide polymorphism (SNP) annotation. Genome Aggregation Database (gnomAD) (Lek et al., 2016) was employed as a source of variant frequencies in worldwide populations. The consensus calls of dbNSFP v2 (Liu et al., 2013) precomputed pathogenicity predictions were used to predict functional effect for missense variants, including SIFT (Sim et al., 2012), Polyphen-2 (Adzhubei et al., 2010), MutationTaster (Schwarz et al., 2014), CADD (Combined Annotation-Dependent Depletion Score) (Kircher et al., 2014), and MetaSVM (Dong et al., 2015). GERP++ rejected substitution (RS) scores were used as the source of information for evolutionary sequence conservation applicable to all types of variants (Davydov et al., 2010). Our pipeline included ClinVar as a source of known disease or drug response association of identified variants. Variants that reached coverage less than 20 and quality less than 300 were excluded from the subsequent analysis.

Variant Filtration and Characterization

Firstly, we applied a 293-gene panel for the filtration of exome data. Next, variants were characterized according to PharmGKB levels of evidence for variant-drug associations⁴ (Whirl-Carrillo et al., 2012) (accessed on 9th September 2018). Level 1A category includes variant-drug pairs with a CPIC pharmacogenetic guideline or variants implemented at a PGRN site or another major health system. Level 1B annotations comprise variant-drug combinations in which the preponderance of evidence shows an association that has been replicated in more than one cohort, with significant *p*-values and preferably with a strong effect size. Clinical annotation of Level 2A refers to variants within known pharmacogenes that are more likely to have a functional significance. Level 2B annotation refers to variant-drug pairs with moderate evidence of an association that has been replicated, but the results might not be statistically significant or the effect size may be small. Initially, we searched for actionable variants, defined as variants with PharmGKB 1A and 1B levels of evidence. The search was based on dbSNP accession numbers. Star allele assignments (ec. CYP2C9*3) were searched for corresponding rs numbers where possible, using

⁴www.pharmgkb.org/downloads

star-allele nomenclature from PharmVar Database⁵ and TPMT Nomenclature Committee websites⁶. Additionally, we extracted variant-drug pairs with PharmGKB annotations of 2A or 2B. We also inspected how many variants with ClinVar accession “drug response” were detected in our dataset. The dosing algorithms were obtained from CPIC guidelines and variant-drug-phenotype associations from PharmGKB website⁷.

Next, we filtered variants on the basis of their minor allele frequencies (MAFs), with an exclusion of variants with MAF > 0.01 in the gnomAD database. We also excluded variants that were detected in more than 19 (1%) individuals as heterozygous and variants detected in more than 15 individuals as homozygous in the Slovenian genomic database. We rated the variants according to variant functional impact, variant type, and theoretical pathogenicity predictions (PolyPhen-2, SIFT, Mutation Tester, MetaSVM, CADD). Additionally, median Phred normalized CADD annotation scores for each gene were calculated (Kircher et al., 2014). Subsequently, we examined the distribution of rare exonic and splicing

variation across major pharmacogenetic gene groups, including cytochrome P450 (CYP) superfamily, ATP-binding cassette (ABC) superfamily, solute carrier (SLC) superfamily, and UDP-glucuronosyltransferases (UGTs).

RESULTS

Overview

Using exome-sequencing data from 1904 individuals we detected a total of 72,293 high-quality variants in 293 pharmacogenes. Our data revealed that most of the variants in pharmacogenes were rare ($n = 65,059$, $MAF_{\text{gnomAD}} < 0.01$), comprising about 90.0% of all variants. Of these rare variants 4360 were annotated as missense, 2239 as synonymous, 174 as frameshifts, and 127 as stop gained. Among the rare non-coding variants, 48,822 were classified as intronic, 1142 as upstream variants, 700 as downstream variants, 1914 as 3'UTR variants, and 735 as 5'UTR variants. Of the rare variants, 9229 (14.2%) were previously reported in the gnomAD database. The number of variants by type is summarized in **Table 1**. The counts by variant annotation impact and variant annotation type for rare exonic and splicing variation are presented in **Figure 1**. The distribution

⁵ www.pharmvar.org/genes

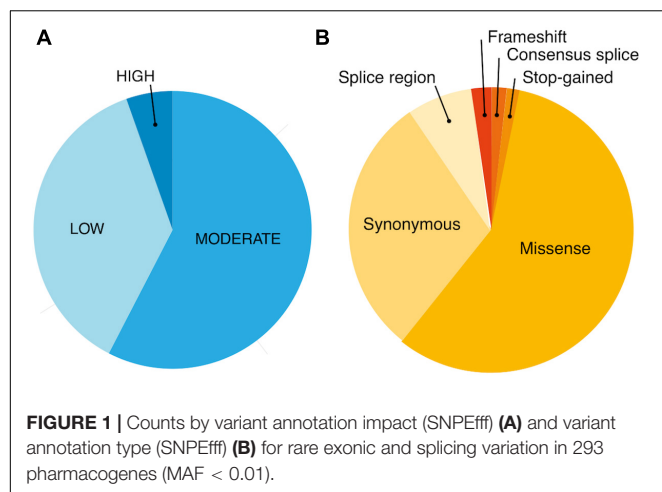
⁶ www.imh.liu.se/tpmtalleles

⁷ www.pharmgkb.org

TABLE 1 | SNPEff Variant types.

	In Slovenian genomic dataset	In gnomAD	In dbSNP 141	Novel (not in gnomAD, dbSNP or ClinVar)
All variants	72293	10646	3083	59428
Rare variants ($MAF_{\text{gnomAD}} < 0.01$), number of heterozygotes < 19, number of homozygotes < 15	61060	9229	2269	51393
Intronic rare	48822	2703	20	46087
Upstream rare	1142	70	0	1070
Downstream rare	700	58	0	641
3'UTR rare	1914	286	2	1354
5'UTR rare	735	155	3	528
Exonic + splicing rare	7571	5831	2225	1669
Missense rare	4360	3394	2079	920
Frameshift rare	174	81	0	86
Synonymous rare	2239	1786	71	448
Splice Acceptor rare	48	24	9	23
Splice Donor rare	60	38	13	21
Stop Gain rare	127	80	48	40
Splice region rare	563	428	6	135
Rare CADD > 20	2101	1615	990	437
Missense rare 4xD + CADD > 20	565	420	260	131
HIGH impact rare (truncating variants)	429	234	76	177
ClinVar at least one 6 (all)	89	70		
ClinVar at least one 6 (rare)	16	9		
PharmGKB Level 1A/1B (all)	24	23		
PharmGKB Level 1A/ 1B (rare)	12	12		
PharmGKB Level 2A/2B (all)	68	50		
PharmGKB Level 2A/ 2B (rare)	12	4		

gnomAD, Genome Aggregation Database; MAF, minor allele frequency; PharmGKB, Pharmacogenomics Knowledgebase.



of rare variation across major pharmacogenetic gene groups is presented in **Figure 2**.

Known Actionable Pharmacogenetic Variants

Firstly, we focused on variants featured in the PharmGKB or/and in ClinVar database. We looked for PharmGKB annotated variant-drug combinations with four highest levels of evidence 1A, 1B, 2A, and 2B. Within the exome sequencing data we identified 24 PharmGKB unique variants with the highest levels of evidence, 1A or 1B, associated with response to about 26 drugs. Twelve of them were rare, with the MAF not exceeding 1% in gnomAD and the Slovenian genomic database. Rare actionable variants (PharmGKB 1A or 1B) located in the exonic or splicing regions were observed in the following genes: *CYP2D6* (frameshift variant), *CYP2C19* (one start lost and another missense variant), *CFTR* (disruptive inframe deletion and five missense variants), *DYPD* (one missense variant and another splice donor variant), and *TPMT* (one missense variant).

We estimated minor allele frequencies for the Slovenian population (MAF_{Slo}) for each potentially actionable finding and the results are presented in **Table 2** and **Supplementary Tables S2–S4**. The most prevalent actionable variant (PharmGKB 1A or 1B level) in our database was a missense variant in *CYP4F2* gene (Val433Met, rs2108622) with MAF_{Slo} of 27.4%, associated with warfarin dosage (PharmGKB level 1A of evidence), followed by missense variants in *CYP2B6* (Gln172His, rs3745274, MAF_{Slo} = 22.4%, PharmGKB 1B), *CYP2D6* (Pro34Ser, rs1065852, MAF_{Slo} = 19.5%, PharmGKB 1A), and *SLCO1B1* (Val174Ala, rs4149056, MAF_{Slo} = 19.2%, PharmGKB 1A) genes. Further most prevalent variants in PharmGKB 1A or 1B category were splice acceptor variant (c.506-1G > A, rs3892097) in *CYP2D6* gene (MAF_{Slo} = 16.7%) and synonymous variant (Pro227Pro, rs4244285) in *CYP2D6* gene (MAF_{Slo} = 12.6%).

Additionally, we identified 68 variants with PharmGKB 2A or 2B levels of evidence, with seven of them also in the 1st categories (PharmGKB 1A and 1B), but presented a different type of an association or different drug-variant pair and were for that reason listed twice. Altogether, the most

common pharmacogenetic variant detected in the Slovenian genomic database was a missense variant in the *F5* gene (Gln534Arg, rs6025, PharmGKB 2A) with MAF_{Slo} of 88.4% (MAF_{gnomAD} = 98.0%). This was followed by a synonymous variant in *ABCC4* gene (Lys1116Lys, rs1751034, PharmGKB 2B), associated with a response to tenofovir and MAF_{Slo} of 77.0% (MAF_{gnomAD} = 81.0%). A missense variant in the *TP53* gene (Pro72Arg, rs1042522, PharmGKB 2B) with MAF in the Slovenian population of 71.8% (MAF_{gnomAD} = 66.9%) was the third most frequently detected pharmacogenetic variant in Slovenian individuals. The variant is associated with the efficacy and toxicity/ADR of antineoplastic agents, such as cisplatin, cyclophosphamide, fluorouracil, and paclitaxel. A start lost variant in *VDR* gene (Met1? rs2228570) was the second most prevalent variant in the PharmGKB 2A category, with MAF_{Slo} of 52.8% (MAF_{gnomAD} = 62.9%). The variant is associated with the efficacy in response to peginterferon alfa-2b in patients suffering from chronic hepatitis C. A missense variant in *COMT* gene (Val158Met, rs4680, PharmGKB 2A) reached the MAF_{Slo} of 48.6%, which is in line with gnomAD MAF frequency of 46.3%. Variant-drug pairs along with corresponding frequencies are summarized in **Table 2** and **Supplementary Tables S2–S4**.

We detected 89 ClinVar variants with at least one accession number 6 (“drug response”), 16 of them with MAF of less than 1% (ClinVar version 02.10.2017).

When we compared MAFs for each risk variant in the Slovenian genomic database to MAFs in the gnomAD database, we generally got consistent results for detected exonic and splicing variation. However, some inconsistencies in the MAFs among databases were apparent. For example, the variant in *ANKK1* gene (Glu713Lys, rs1800497) associated with the toxicity and ADR of antipsychotics, had a MAF of 26.4% in the gnomAD database ($MAF_{European(Non-Finnish)}$ = 19.2%), but had a MAF of only 17.3% in the Slovenian genomic database. A synonymous variant in the *CYP2C19* gene (Pro227Pro, rs4244285) influencing the efficacy of several drugs including amitriptyline, clopidogrel, citalopram, and clomipramine, had a MAF of 17.6% in the gnomAD database ($MAF_{European(Non-Finnish)}$ = 14.7%) and 12.6% in the Slovenian database. The variant in *F5* gene (Gln534Arg, rs6025), associated with the adverse event of thrombosis in systemic hormonal contraceptives use, had a MAF of 98.0% in gnomAD ($MAF_{European(Non-Finnish)}$ = 97.0%) and only of 88.4% in Slovenian population. In contrast, missense variant in the *SLCO1B1* gene (Val174Ala, rs4149056) associated with the adverse drug reaction and toxicity of simvastatin, has MAF of 13.3% in gnomAD ($MAF_{European(Non-Finnish)}$ = 15.6%) and 19.2% in the Slovenian database. It is important to note that some of the detected variants were intronic, therefore, their MAF_{Slo} may be unreliable due to the lack of sequence coverage for these regions in exome sequencing data.

Functional Impacts of Variants

Next, we characterized the rare variants on the basis of predicted functional impacts. We detected 2101 variants that reached CADD score above the cut-off value of 20 and thus ranked into the 1st percentile of the most deleterious variants (Kircher et al., 2014). Several *in silico* prediction algorithms

TABLE 2 | Variant-drug pairs with 1A or 1B clinical annotation according to Pharmacogenomics Knowledgebase (PharmGKB).

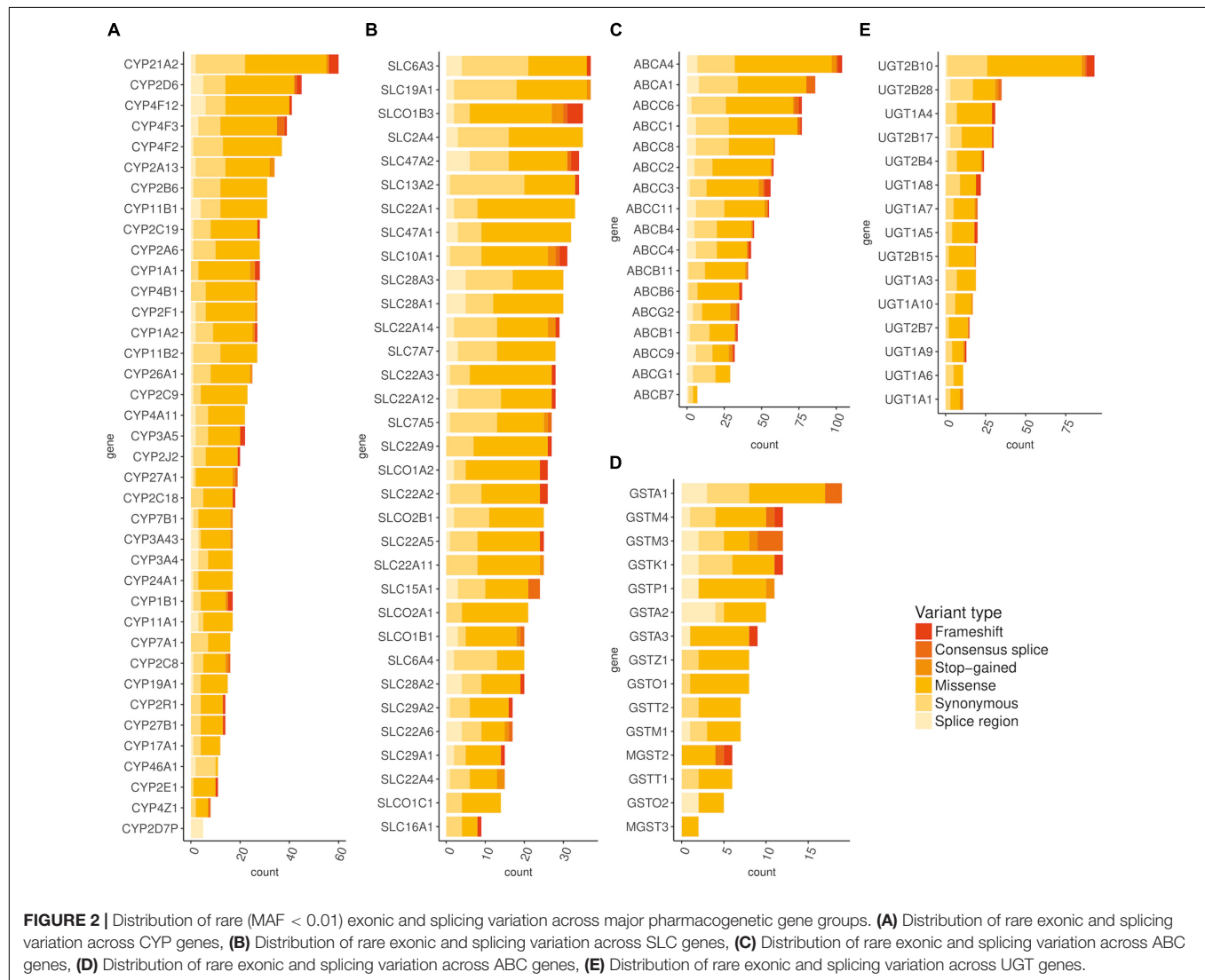
Gene	Variant type	Transcript level variant	Functional effect	dbSNP	MAF _{gnomAD}	Het _{Slo}	MAF _{Slo}	Hom _{Slo}	PharmGKB	Drugs
<i>CYP2C9</i> (*3)	Missense	c.1075A > C	Ile359Leu	1057910	0.0636	247	0.0691	8	1A	Warfarin, phenytoin
<i>CYP2C9</i> (*2)	Missense	c.430C > T	Arg144Cys	1799853	0.0926	390	0.121	35	1A	Warfarin, phenytoin
<i>CYP2D6</i> (*4)	Splice acceptor	c.506-1G > A	–	3892097	0.138	487	0.167	74	1A	Amitriptyline, clomipramine, desipramine, doxepin, imipramine, nortriptyline, trimipramine
<i>CYP2D6</i> (*6) (non-functional)	Frameshift	c.454delT	Trp152fs	5030655	0.00791	42	0.0110	0	1A	Paroxetine, fluvoxamine, amitriptyline, nortriptyline, codeine
<i>CYP2D6</i> (all the variants in which it appears have reduced or no <i>CYP2D6</i> activity)	Missense	c.100C > T	Pro34Ser	1065852	0.207	511	0.195	115	1A	Paroxetine, nortriptyline, codeine, amitriptyline
<i>SLCO1B1</i> (*5)	Missense	c.521T > C	Val174Ala	4149056	0.133	585	0.192	73	1A	Simvastatin
<i>CYP2D6</i> (*3)	Frameshift	c.775delA	Arg259fs	35742686	0.0124	67	0.0186	2	1A	Amitriptyline, nortriptyline, trimipramine, clomipramine, tamoxifen, codeine, paroxetine, doxepin, fluvoxamine
<i>VKORC1</i>	3'-UTR	c.*134G > A	–	7294	.	64	0.0273	20	1B	Warfarin
<i>CYP4F2</i>	Missense	c.1297G > A	Val433Met	2108622	0.274	759	0.274	142	1A	Warfarin
<i>CYP2B6</i>	Missense	c.516G > T	Gln172His	3745274	0.272	651	0.224	101	1B	Efavirenz
<i>CYP2C19</i> (*2)	Synonymous	c.681G > A	Pro227Pro	4244285	0.176	411	0.126	35	1A	Amitriptyline, clopidogrel
<i>CYP2C19</i> (*4)	Start lost	c.1A > G	Met1?	28399504	0.00231	11	0.00289	0	1A	Clopidogrel
<i>CYP2C19</i> (*8) (non-functional)	Missense	c.358T > C	Trp120Arg	41291556	0.00152	11	0.00289	0	1A	Clopidogrel
<i>CFTR</i>	Disruptive inframe deletion	c.1521_1523 delCTT	Phe508del	113993960	0.00696	39	0.0102	0	1A	Ivacaftor
<i>CFTR</i>	Missense	c.220C > T	Arg74Trp	115545701	0.00142	1	0.000263	0	1A	Ivacaftor
<i>CFTR</i>	Missense	c.328G > C	Asp110His	113993958	0.0000203	1	0.000263	0	1A	Ivacaftor
<i>CFTR</i>	Missense	c.3154T > G	Phe1052Val	150212784	0.000632	4	0.00105	0	1A	Ivacaftor
<i>CFTR</i>	Missense	c.3209G > A	Arg1070Gln	78769542	0.000623	1	0.000263	0	1A	Ivacaftor
<i>CFTR</i>	Missense	c.3454G > C	Asp1152His	75541969	0.000407	2	0.000526	0	1A	Ivacaftor
<i>DYPD</i>	Missense	c.2846A > T	Asp949Val	67376798	0.00284	9	0.00236	0	1A	Capecitabine, fluorouracil, pyrimidine analogs, tegafur
<i>DPYD</i> (*2A)	Splice donor	c.1905+1G > A	–	3918290	0.00574	10	0.00263	0	1A	Capecitabine, fluorouracil, pyrimidine analogs, tegafur
<i>TPMT</i> (*2)	Missense	c.238G > C	Ala80Pro	1800462	0.00172	6	0.00158	0	1A	Azathioprine, mercaptopurine, purine analogs, thioguanine

(Continued)

TABLE 2 | Continued

Gene	Variant type	Transcript level variant	Functional effect	dbSNP	MAF _{gnomAD}	Het _{Slo}	MAF _{Slo}	Hom _{Slo}	PharmGKB	Drugs
<i>TPMT</i> (*3B/potentially encoding *3A allele)	Missense	c.460G > A	Ala154Thr	1800460	0.0280	93	0.0249	1	1A	Azathioprine, mercaptopurine, purine analogs, thioguanine
<i>TPMT</i> (potentially encoding *3C allele / *3A)	Missense	c.719A > G	Tyr240Cys	1142345	0.0366	98	0.0268	2	1A	Azathioprine, mercaptopurine, purine analogs, thioguanine

Het_{Slo}, number of heterozygotes in the Slovenian genomic database; *Hom_{Slo}*, number of homozygotes in Slovenian genomic database; *MAF*, minor allele frequency; *1A*, variant-drug pairs with a CPIC pharmacogenetic guideline or variants implemented at a PGRN site or another major health system; *1B*, variant-drug pairs in which the preponderance of evidence shows an association that has been replicated in more than one cohort, with significant *p*-values and preferably with a strong effect size.



(Mutation Tester, Polyphen-2, SIFT, MetaSVM) predicted in consensus as pathogenic 565 missense variants that also reached CADD score above 20. These included 131 novel variants- not previously reported in gnomAD, dbSNP or ClinVar database (Supplementary Table S5). We further analyzed rare variants with protein-truncating effects, including frameshift variants,

stop-gain variants, and variants affecting splicing, which resulted in additional 429 variants predicted to be highly pathogenic based on their functional impact, of which 177 were novel (Supplementary Table S6). Most of the rare putatively functional variants were thus missense, followed by frameshift ($n = 174$) and stop-gain variants ($n = 127$).

Additionally, we examined CADD annotation scores separately for each gene and further compared median CADD scores of rare variants with common variants. We observed the highest median CADD scores for rare variants in the following genes: *CDA*, *SLC10A2*, *TPMT*, and *SULF1*. Among CYP genes, the highest median score was detected in the *CYP1A1* gene, followed by *CYP24A1*, and *CYP2R1*. In the SLC group, *SLC10A2*, *SLC22A2*, and *SLC22A6* genes ranked the highest. Altogether, the highest CADD score of 54 was detected for the known pathogenic variant in *ABCA4* gene (c.6445C > T, Arg2149*, rs61750654), followed by the stop-gain variant in *EPHX2* gene (Arg467*, CADD = 51). Expectedly, we identified significantly more damaging variants among the rare variants than among common variants with the population frequency exceeding 1%.

DISCUSSION

The implementation of exome sequencing technologies into daily clinical practice makes the prospect of a personalized treatment increasingly available. Here, we showed that by using clinical and whole exome sequencing technologies it is possible to identify not only variants that are causative for patients' clinical presentation but also a considerable proportion of pharmacogenetics findings with established evidence and potential clinical utility.

Within the study population, we identified a high frequency of well-established examples of common genetic polymorphisms, as well as known rare actionable variants. We detected 24 variants with compelling evidence of pharmacogenetic significance (PharmGKB level 1A or 1B variants) associated with about 26 drugs, where 12 of them were rare, and 61 additional variants with a level 2A or 2B PharmGKB evidence. Our results are consistent with those of the previously published study by Lee et al. (2016), in which authors used combined SNP chip and exome sequence data of 1101 individuals. In their study, 29 variants were detected that ranked in the PharmGKB 1A and 1B categories; 21 of them were detected by exome sequencing technology. Similarly, 22 actionable clinical variants (PharmGKB 1A/1B) were found in 120 pharmacogenes when analyzing 1000 Genomes Phase 3 data of 2540 individuals (Wright et al., 2018).

Furthermore, our results are correlated with the already known fact that rare variants are enriched for deleterious variation. We identified 308 novel variants of potential functional significance, including 131 missense variants (predicted in consensus as pathogenic by functional prediction algorithms: Mutation Tester, Polyphen-2, SIFT, MetaSVM, CADD) and 177 protein-truncating variants. We observed that especially when testing an expanded set of genes, novel putatively functional variants and variants in genes with less established effects represent a considerable challenge in result interpretation and reporting. To date, very few studies have conducted a systematic overview of the distribution and frequency of genetic variation with potentially high impact over a large set of pharmacogenes (Kozrya et al., 2016; Schärfe et al., 2017; Wright et al., 2018). So far, studies examining rare genetic variation have been limited to small sets of genes or on gene groups, such as largely studied cytochrome P450 (CYP) gene family (Gordon et al.,

2014; Fujikura et al., 2015). Further evaluation of functional consequences and clinical effects is required to extend our understanding of rare variants. Therefore, a considerable part of the variation in response to treatment still remains unclear and has not yet been integrated into routine clinical practice.

Moreover, we have observed that MAFs of some known variants differ significantly in the Slovenian dataset when compared to gnomAD MAFs. This raises the importance of establishing population specific databases of pharmacogenomics variation. With growing pharmacogenetic databases and increased integration of sequencing technologies into clinical practice, we will also gain additional insight into the rare pharmacogenetic variation. This will make publicly accessible and easily updatable data repositories such as CPIC, PharmGKB, ClinVar, Pharmacogene Variation (PharmVar) Consortium, as well as population-specific databases, essential for the accurate interpretation of pharmacogenomics results along with the subsequent integration of dosing recommendations and guidelines into electronic healthcare record systems.

Also, the identification and reporting of pharmacogenetic findings are in many aspects distinct from reporting of the disease causative variants. The proposed American College of Medical Genetics and Genomics (ACMG) criteria for interpretation of sequence variants are not intended for pharmacogenomic findings (Richards et al., 2015). While the comprehensive phenotyping data could be of particular value when interpreting the putative disease-causative variants, the genotype-phenotype correlation for pharmacogenomic findings is apparent only when the patient is exposed to a specific drug. Furthermore, the results may not be useful at the time of reporting, but only when the particular drug is prescribed to the patient. However, by potential reporting or storing of actionable variants from sequencing data preemptively, they may be available prior the prescription and for a wide range of medications, subsequently influencing decisions about treatment, which could significantly medically benefit the patients (Dunnenberger et al., 2015; Ji et al., 2016).

Compared to approaches targeting only known pharmacogenetics variants, sequencing technologies are beneficial for a number of additional aspects. SNP genotyping assays may be unable to detect low-frequency variants with potential deleterious functional effects. Besides, the response to a majority of the drugs is influenced by several genes, including genes encoding drug metabolizing enzymes, transporters, drug targets, and disease-modifying genes, or by various variants within the same gene, which may not be detected using approaches targeting known pharmacogenetics variants (Relling and Evans, 2015). We have demonstrated that exome sequencing is an effective method for the detection of both rare and common pharmacogenetic variants in a large set of genes under one investigation.

The present study identifies a high number of clinically relevant highly actionable variant-drug associations, with already established dosing guidelines and recommendations applicable for the use in personalized treatment. Here we highlight the potential clinical utility for a selection of variants detected in the Slovenian database.

A decreased function missense variant in the *SLCO1B1* gene (Val174Ala, rs4149056, allele *5) has a MAF of 19.2%

in the Slovenian population. It was identified as heterozygous in 585/1904 (31%) individuals and as homozygous in 73/1904 (4%) individuals, who, therefore, have an intermediate and high myopathy risk, respectively, when receiving simvastatin treatment. Consequently, a lower dose or alternative statin (e.g., pravastatin or rosuvastatin) and routine creatine kinase (CK) surveillance are recommended for these individuals (Wilke et al., 2012; Ramsey et al., 2014).

Variants of *CYP2D6* and *CYP2C19* genes affect the exposure, efficacy, and safety of tricyclic antidepressants (TCAs) (Hicks et al., 2013). A synonymous variant (Pro227Pro, rs4244285) in the *CYP2C19* gene represents no function allele (*2) and thus greatly decreases the conversion of tertiary amines to secondary amines, which may cause a sub-optimal response. The MAF of the variant in the Slovenian population was estimated at 12.6%; 35/1904 (1.8%) individuals carried the variant in the homozygous state, who should avoid the use of tertiary amine and alternative drugs that are not metabolized by *CYP2C19* are recommended. Moreover, we detected c.506-1G > A variant (allele *4, rs3892097), with anticipated effect on splicing in *CYP2D6* gene, resulting in a greatly reduced metabolism of TCAs to less active compounds. The variant was found in 487/1904 (25.6%) of Slovenian individuals as heterozygous, and in 74/1904 (3.9%) as homozygous. Additionally, the variant also has a major role in the activation of prodrugs such as codeine and tramadol.

Cytochrome P450 *CYP2C19* also catalyzes the bioactivation of the antiplatelet prodrug clopidogrel that inhibits the ADP-dependent P2Y₁₂ receptor. *CYP2C19* (*2) loss-of-function allele impairs formation of active metabolites (Scott et al., 2013). Both heterozygous 411/1904 (21.6%) and homozygous 35/1904 (1.8%) clopidogrel-treated patients with acute coronary syndromes have significantly reduced platelet inhibition and thus an increased risk for serious adverse cardiovascular events. Alternative antiplatelet medication, such as prasugrel or ticagrelor is strongly recommended in individuals with this variant.

Furthermore, we detected two rare variants with PharmGKB level 1A of evidence, one variant with the effect on splicing (c.1905+1G > A, allele *2A, rs3918290, MAF_{Slo} = 0.263%) and another missense variant (c.2846A > T, Asp949Val, rs67376798, MAF_{Slo} = 0.236%) in the *DPYD* gene. Heterozygotes for one of the detected variants in the *DPYD* gene have reduced leukocyte dihydropyrimidine dehydrogenase (DPD) activity (at 30–70% that of the normal population) and an increased risk of severe or even lethal drug toxicity when treated with fluoropyrimidine drugs. At least 50% reduction in starting dose is recommended, followed by titration of dose based on toxicity or pharmacokinetic test (Caudle et al., 2013).

Our study also has some limitations. Firstly, the application of exome sequencing for the detection of pharmacogenomics variants is limited (Londin et al., 2014). Because of the lack of coverage of the exome test, we were not able to accurately detect the majority of the intronic variation (e.g., the rs9923231 variant of the *VKORC1* gene). Due to technical limitations, structural variants and repetitive regions were not sufficiently assessed. We also recognize the limitation of the sensitivity of exome sequencing in highly homologous regions of the human genome, including the cytochrome genes (e.g., part of

the known actionable variability of the *CYP2D6* gene). We recognize the possibility that we failed to detect a minor part of pharmacogenomic variation due to the limited detection of variants in these regions. Furthermore, in the present study, we did not extend the exome analysis on copy number variation (CNVs). Also, with de-identified data, we could not identify the compound heterozygous states or assess the polygenic effects of variants. Nevertheless, when analyzing each patient's data separately, it will be possible to include multigenic effects, compound heterozygous states and some of the risk haplotypes with established pharmacogenetic effects in future patient's records, which will add the considerable value to the exome sequencing results. With such valuable data, we could significantly benefit future patients by increasing the efficacy and decreasing adverse drug responses of pharmacologic treatment.

CONCLUSION

In conclusion, our results demonstrate that nationally based exome sequencing data represents a valuable source for identification of pharmacogenetic variants. The direct inclusion of actionable pharmacogenetics findings in patient's records could significantly improve the outcome in patients who underwent diagnostic exome and genome sequencing. Furthermore, our data provide the first comprehensive overview of the distribution of both rare and common variants within several pharmacogenes and provides first estimates on their prevalence for the Slovenian population. We have shown that testing beyond known polymorphisms is warranted to gain further insight into rare variation and to facilitate more reliable future interpretation and reporting of pharmacogenetic findings. We anticipate that the present dataset will be of great importance for future research and validation of pharmacogenetics variation in the Slovenian population. Based on our results we propose that known pharmacogenetic variants with well-established effects should be a part of every genetic report.

AUTHOR CONTRIBUTIONS

BP, KH, and AM contributed to conception and design of the study. KH and AM performed the statistical analysis. KH wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This study was funded by the Slovenian Research Agency (ARRS), grant no. P3-0326.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00240/full#supplementary-material>

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Caudle, K. E., Klein, T. E., Hoffman, J. M., Muller, D. J., Whirl-Carrillo, M., Gong, L., et al. (2014). Incorporation of pharmacogenomics into routine clinical practice: the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline development process. *Curr. Drug Metab.* 15, 209–217. doi: 10.2174/1389200215666140130124910
- Caudle, K. E., Thorn, C. F., Klein, T. E., Swen, J. J., McLeod, H. L., Diasio, R. B., et al. (2013). Clinical pharmacogenetics implementation consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clin. Pharmacol. Ther.* 94, 640–645. doi: 10.1038/clpt.2013.172
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Cousin, M. A., Matey, E. T., Blackburn, P. R., Boczek, N. J., McAllister, T. M., Kruijselbrink, T. M., et al. (2017). Pharmacogenomic findings from clinical whole exome sequencing of diagnostic odyssey patients. *Mol. Genet. Genomic Med.* 5, 269–279. doi: 10.1002/mggg.3.283
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., 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. doi: 10.1371/journal.pcbi.1001025
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–501. doi: 10.1038/ng.806
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24, 2125–2137. doi: 10.1093/hmg/ddu733
- Dunnenberger, H. M., Crews, K. R., Hoffman, J. M., Caudle, K. E., Howard, S. C., Hunkler, R. J., et al. (2015). Preemptive clinical pharmacogenetics implementation: current programs in five United States medical centers. *Annu. Rev. Pharmacol. Toxicol.* 55, 89–106. doi: 10.1146/annurev-pharmtox-010814-124835.Preemptive
- Fujikura, K., Ingelman-Sundberg, M., and Lauschke, V. M. (2015). Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenet. Genomics* 25, 584–594. doi: 10.1097/FPC.0000000000000172
- Gordon, A. S., Tabor, H. K., Johnson, A. D., Snively, B. M., Assimes, T. L., Auer, P. L., et al. (2014). Quantifying rare, deleterious variation in 12 human cytochrome P450 drug-metabolism genes in a large-scale exome dataset. *Hum. Mol. Genet.* 23, 1957–1963. doi: 10.1093/hmg/ddt588
- Hicks, J. K., Swen, J. J., Thorn, C. F., Sangkuhl, K., Kharasch, E. D., Ellingrod, V. L., et al. (2013). Clinical pharmacogenetics implementation consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. *Clin. Pharmacol. Ther.* 93, 402–408. doi: 10.1038/clpt.2013.2
- Ji, Y., Skierka, J. M., Blommel, J. H., Moore, B. E., Vancuyk, D. L., Bruflat, J. K., et al. (2016). Preemptive pharmacogenomic testing for precision medicine: a comprehensive analysis of five actionable pharmacogenomic genes using next-generation DNA sequencing and a customized CYP2D6 genotyping cascade. *J. Mol. Diagn.* 18, 438–445. doi: 10.1016/j.jmoldx.2016.01.003
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., Shendure, J., et al. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi: 10.1038/ng.2892
- Kozyra, M., Ingelman-Sundberg, M., and Lauschke, V. M. (2016). 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. doi: 10.1038/gim.2016.33
- Lee, E. M. J., Xu, K., Mosbrook, E., Links, A., Guzman, J., Adams, D. R., et al. (2016). Pharmacogenomic incidental findings in 308 families: the NIH Undiagnosed Diseases Program experience. *Genet. Med.* 18, 1303–1307. doi: 10.1038/gim.2016.47
- Lek, M., Karczewski, K. J., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* 536:030338. doi: 10.1101/030338
- 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–E2402. doi: 10.1002/humu.22376
- London, E. R., Clark, P., Sponziello, M., Kricka, L. J., Fortina, P., and Park, J. Y. (2014). Performance of exome sequencing for pharmacogenomics HHS public access. *Per. Med.* 12, 109–115. doi: 10.2217/PME.14.77
- Mizzi, C., Dalabira, E., Kumuthini, J., Dzimir, N., Balogh, I., Başak, N., et al. (2017). Correction: a european spectrum of pharmacogenomic biomarkers: implications for clinical pharmacogenomics. *PLoS One* 12:e0162866. doi: 10.1371/journal.pone.0172595
- Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., St. Jean, P., Verzilli, C., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337, 100–104. doi: 10.1126/science.1217876
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufio, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- Potamias, G., Lakiotaki, K., Katsila, T., Lee, M. T. M., Topouzis, S., Cooper, D. N., et al. (2014). Deciphering next-generation pharmacogenomics: an information technology perspective. *Open Biol.* 4, 421–432. doi: 10.1098/rsob.140071
- Ramsey, L. B., Bruun, G. H., Yang, W., Treviño, L. R., Vattathil, S., Scheet, P., et al. (2012). Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res.* 22, 1–8. doi: 10.1101/gr.129668.111
- Ramsey, L. B., Johnson, S. G., Caudle, K. E., Haidar, C. E., Voora, D., Wilke, R. A., et al. (2014). The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update supplement. *Clin. Pharmacol. Ther.* 96, 423–428. doi: 10.1038/clpt.2014.125
- Rasmussen-Torvik, L. J., Stallings, S. C., Gordon, A. S., Almoguera, B., Basford, M. A., Bielinski, S. J., et al. (2014). Design and anticipated outcomes of the eMERGE-PGX project: a multicenter pilot for preemptive pharmacogenomics in electronic health record systems. *Clin. Pharmacol. Ther.* 96, 482–489. doi: 10.1038/clpt.2014.137
- Relling, M. V., and Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature* 526, 343–350. doi: 10.1038/nature15817
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17, 405–423. doi: 10.1038/gim.2015.30
- Ross, C. J. D., Visscher, H., Sistonen, J., Brunham, L. R., Pussegoda, K., Loo, T. T., et al. (2010). The canadian pharmacogenomics network for drug safety: a model for safety pharmacology. *Thyroid* 20, 681–687. doi: 10.1089/thy.2010.1642
- Schärfe, C. P. I., Tremmel, R., Schwab, M., Kohlbacher, O., and Marks, D. S. (2017). Genetic variation in human drug-related genes. *Genome Med.* 9, 1–15. doi: 10.1186/s13073-017-0502-5
- Schwarz, J. M., Cooper, D. N., Schuelke, M., and Seelow, D. (2014). Mutationtaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* 11, 361–362. doi: 10.1038/nmeth.2890
- Scott, S. A., Sangkuhl, K., Stein, C. M., Hulot, J.-S., Mega, J. L., Roden, D. M., et al. (2013). Clinical pharmacogenetics implementation consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin. Pharmacol. Ther.* 94, 317–323. doi: 10.1038/clpt.2013.105
- Shuldiner, A. R., Relling, M. V., Peterson, J. F., Hicks, K., Freimuth, R. R., Sadee, W., et al. (2013). The pharmacogenomics research network translational pharmacogenetics program: overcoming challenges of real-world implementation. *Clin. Pharmacol. Ther.* 94, 207–210. doi: 10.1038/clpt.2013.59
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40, W452–W457. doi: 10.1093/nar/gks539
- Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64–69. doi: 10.1126/science.1219240
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, 1–7. doi: 10.1093/nar/gkq603

- Whirl-Carrillo, M., McDonogh, E., Herbert, J., Gong, L., Sangkuhl, K., Thotn, C., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 92, 414–417. doi: 10.1038/clpt.2012.96.
- Pharmacogenomics Consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin. Pharmacol. Ther.* 92, 112–117. doi: 10.1038/clpt.2012.57
- Wright, G. E. B., Carleton, B., Hayden, M. R., and Ross, C. J. D. (2018). The global spectrum of protein-coding pharmacogenomic diversity. *Pharmacogenomics J.* 18, 187–195. doi: 10.1038/tpj.2016.77

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hočevár, Maver and Peterlin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Integrating Next-Generation Sequencing in the Clinical Pharmacogenomics Workflow

Efstathia Giannopoulou¹, Theodora Katsila¹, Christina Mitropoulou², Evangelia-Eirini Tsermpini¹ and George P. Patrinos^{1,3,4*}

¹ Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece, ² The Golden Helix Foundation, London, United Kingdom, ³ Department of Pathology, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates, ⁴ Zayed Center of Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

OPEN ACCESS

Edited by:

Ulrich M. Zanger,
Dr. Margarete Fischer-Bosch-Institut
für Klinische Pharmakologie (IKP),
Germany

Reviewed by:

Joseph Borg,
University of Malta, Malta
Volker Martin Lauschke,
Karolinska Institute (KI), Sweden

*Correspondence:

George P. Patrinos
gpatrinos@upatras.gr

Specialty section:

This article was submitted to
Pharmacogenetics
and Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 09 January 2019

Accepted: 27 March 2019

Published: 05 April 2019

Citation:

Giannopoulou E, Katsila T,
Mitropoulou C, Tsermpini E-E and
Patrinos GP (2019) Integrating
Next-Generation Sequencing
in the Clinical Pharmacogenomics
Workflow. *Front. Pharmacol.* 10:384.
doi: 10.3389/fphar.2019.00384

Pharmacogenomics has been recognized as a fundamental tool in the era of personalized medicine with up to 266 drug labels, approved by major regulatory bodies, currently containing pharmacogenomics information. Next-generation sequencing analysis assumes a critical role in personalized medicine, providing a comprehensive profile of an individual's variome, particularly that of clinical relevance, comprising of pathogenic variants and pharmacogenomic biomarkers. Here, we propose a strategy to integrate next-generation sequencing into the current clinical pharmacogenomics workflow from deep resequencing to pharmacogenomics consultation, according to the existing guidelines and recommendations.

Keywords: clinical pharmacogenomics, workflow, implementation, next-generation sequencing, clinical decision support tools

INTRODUCTION

Since the 1950s, many pioneers in the biomedicine field have reported individual variability in disease management and envisioned personalized medicine in health care (Evans and Relling, 1999). Notwithstanding, the realistic application of genomic findings and technologies in the clinic goes beyond the discovery of gene variants and their validation in clinical trials. Lam (2013) has suggested a series of stages regarding the development and implementation pathways for pharmacogenomic tests, namely: (i) discovery of pharmacogenomic biomarkers and validation in well-controlled studies with independent populations; (ii) replication of drug-gene(s) association and demonstration of utility in at-risk patients; (iii) development and regulatory approval of companion diagnostic test; (iv), assessing the clinical impact and cost-effectiveness of the pharmacogenomic biomarkers; (v), involvement of all stakeholders in clinical implementation (Lam, 2013).

Noteworthy, the scientific challenges and implementation barriers existing within the abovementioned stages are still rather unmet. Pharmacogenomic testing occurs by genotyping or sequencing and is mostly outsourced from hospitals to private companies, being a time-consuming and costly process (Harper and Topol, 2012). Unfortunately, there is still a profound lack of understanding within the medical community regarding genomics and the impact of genomic variants in rationalizing drug prescription (Stanek et al., 2012; Mitropoulou et al., 2014).

On the other hand, the pharmacy benefit managers (involved with authorizing of fulfilling most prescriptions in the United States) have been particularly interested in the use of pharmacogenomic testing to save employers (their customers) the cost of a drug through genotyping, making the pharmacy benefit managers in question more competitive (Topol, 2010).

In 2013, the United States Food and Drug Administration (FDA) announced a guidance for industry entitled “Clinical Pharmacogenomics: Premarket Evaluation in Early-Phase Clinical Studies and Recommendations for Labeling¹” in an effort to address the challenges that need to be met. The FDA has also established the Genomics and Targeted Therapy Group² toward the advancement of the application of genomic technologies in the discovery, development, regulation, and use of medications. In the same context, the United States National Cancer Institute has announced a rather similar research and development workflow toward treatment strategies in cancer, including: (i) the support of the routine collection of germline and tumor biospecimens from clinical trials or population-based studies, (ii) the support in efficacy/toxicity biomarker development, (iii) the incorporation of pharmacogenomic markers into clinical trials, and (iv) the consideration of ethical, legal, social, biospecimen, and data-sharing implications of pharmacogenomics research (Freedman et al., 2010). Today, FDA has approved 266 drugs that include genetic information in their labels (Drozda et al., 2018) and the same is true for the European Medicines Agency (Ehmann et al., 2015). The distribution of these drugs between various target diseases indicates that oncology, cardiology, psychiatry, and neurology are among the most common ones in which pharmacogenomics are readily applicable for routine clinical care (Potamias et al., 2014).

NEXT-GENERATION SEQUENCING GENOTYPING IN PHARMACOGENOMICS

Considering the plummeting cost of genotyping, particularly in a high-throughput format, such as panel-based genotyping and/or next-generation sequencing as well as data accuracy improvements, one would envisage that comprehensive pharmacogenomic testing using these approaches could be readily applicable in a clinical setting (Kitzmiller et al., 2011). Indeed, major academic institutions, government-sponsored as well as private organizations and research consortia are engaged into collaborative programs that focus on next generation sequencing of the cancer genome, aiming to describe the architecture of cancer-specific somatic alterations and as such, aid clinicians toward disease management (Simon and Roychowdhury, 2013), while others, such as the SEAPharm Consortium³ are currently exploring the use of targeted pharmacogene resequencing in 100 pharmacogenes to explore the pharmacogenomic variants allelic architecture and the

most prevalent pharmacogenomic biomarkers in Southeast Asian populations.

Recently, by investigating the exome sequences of over 60000 individuals, Ingelman-Sundberg et al. (2018) demonstrated that each individual harbors, on average, approximately 41 putatively functional pharmacogenomic variants from which 10.8% are rare and found to be highly gene- and drug-specific, accounting for a substantial part of the unexplained inter-individual differences in drug metabolism phenotypes.

Still, and contrary to identifying the genetic basis of disorders characterized by a high degree of phenotypic and clinical variability and/or genetic heterogeneity (Ku et al., 2016), in case of pharmacogenomic testing, where the role of several pharmacogenes is well established, targeted gene resequencing seems to be perhaps more relevant compared to whole exome sequencing, as it also captures rare pharmacovariants that are present in other genomic positions than the gene exons, such as promoters, intronic and untranslated sequences, which have been shown to lead to drastic reduction of drug metabolizing enzyme activity. This is further highlighted in a recent study comparing the results obtained by whole genome sequencing, whole exome sequencing, and microarray-based genotyping, indicating that the performance of genotyping arrays is similar to that of whole genome sequencing, whereas whole exome sequencing is not suitable for pharmacogenomics predictions (Reisberg et al., 2019). In any case, novel and rare pharmacovariants that can only be identified by next-generation sequencing approaches are of utmost importance in personalized drug therapy to provide information of use to avoid adverse drug reactions and lack of response (Lauschke and Ingelman-Sundberg, 2018).

Tumor samples are known to contain both acquired and inherited alterations, along with somatic DNA. Thus, cancer sequencing efforts also capture germline information. This germline information plays a crucial role in optimizing the dose and selection of therapy. A unique benefit to next generation sequencing is the ability to discover rare variants (in cancer patients, germline DNA is also analyzed as a means to identify variants in the tumor) in the genome and then, delineate their impact on drug response (Gillis et al., 2014). This has been previously demonstrated by Mizzi et al. (2014), indicating that novel and rare variants can exert a deleterious effect in drug metabolizing enzymes, such as CYP2D6, TPMT, CYP2C19, involved in anti-cancer, psychiatric and cardiology drug treatment, among others, by introducing premature stop-codons or out-of-frame frameshifts very close to the N-terminus of the enzyme. These authors also demonstrated that whole genome sequencing could identify novel CYP2C9 variants relevant to anticoagulation treatment, which could not have been identified using microarray-based genotyping approaches, which could potentially guide toward alternative anticoagulation treatment modalities in two patients suffering from atrial fibrillation (Mizzi et al., 2014). Furthermore, rather than Sanger sequencing, next generation sequencing technology yields more accurate quantitative results, when somatic variation is considered and can be achieved at a higher throughput scale (Simon and Roychowdhury, 2013). Indeed, findings in genes involved in the metabolism of anti-cancer drugs

¹<https://federalregister.gov/a/2013-01638>

²<https://www.fda.gov/drugs/scienceresearch/ucm572617.htm>

³<http://www.pharmagtc.org/seapharm>

further demonstrate the potential applicability of whole genome sequencing for pharmacogenomic testing in a clinical setting in the not too distant future (McCarty et al., 2011; Mizzi et al., 2014; Karageorgos et al., 2015).

INFORMATION TECHNOLOGIES AND DATA INTERPRETATION

Currently, difficulties in pharmacogenomics data interpretation are claimed responsible for the slow clinical uptake of pharmacogenomics. Two main aspects of data interpretation have been identified to affect pharmacogenomics translation into clinical practice: (i) the interpretation of reported genetic results by clinicians and (ii) the interpretation of published research results. It has become evident that the vast majority of health professionals even though acknowledges that genetic variations may influence drug response, only a limited number of those feel adequately informed about pharmacogenomic testing and data interpretation (Stanek et al., 2012; Mitropoulou et al., 2014). So far, standardization in conducting pharmacogenomics studies is lacking, mainly due to inconsistencies in results reporting (O'Donnell and Ratain, 2012). These inconsistencies make data interpretation challenging or even chaotic to researchers, professional organizations, consortia and clinicians alike and international efforts are currently ongoing to standardize pharmacogenomics testing reporting (Kalman et al., 2016).

With the advent of next generation sequencing, collaboration toward data accumulation would help maximize its clinical benefit, as large sample sizes would provide the means to retrospectively analyze large patient cohorts for (i) discovery of common and rare variants, (ii) validation, and (iii) pharmacogenomics outcomes toward decision-making. Today, the Electronic Medical Records and Genomics (eMERGE) Network, attempts to maximize the benefit from next generation sequencing analyses, focusing on the combination of DNA biorepositories with electronic medical records to facilitate large-scale, high-throughput genetic research and return genetic testing results to patients in a clinical setting (McCarty et al., 2011). Such efforts would be beneficial to be exploited, including somatic and germline variation discovery and implementation as well as clinical and uptake outcomes.

To this end, in the big data era, biomedicine scientists need to critically appraise data, collaborate in an efficient and effective way and make decisions. For this, large-scale volumes of complex multi-faceted data need to be meaningfully assembled, mined, analyzed and provided in a user-friendly manner. An innovative web-based collaboration support platform that adopts a hybrid approach on the basis of the synergy between machine and human intelligence was previously reported, aiming to facilitate the underlying sense-making and decision making processes (Tsiliki et al., 2014). Clinical decision support (CDS) tools have been also proved valuable in the context of clinical pharmacogenomics, as they provide guidance on clinical decisions, through electronic medical records (Bell et al., 2014). Again, these tools demand clear and precise algorithms based on scientifically robust findings, ideally synergizing among different

variant prediction tools to take novel and rare pharmacogenomic variants into consideration to determine their pathogenicity.

VALIDATION AND ACCREDITATION OF SERVICES

The application of pharmacogenomics in personalized medicine is very challenging and influence medicine and biomedical research in many areas, namely clinical medicine, drug development, drug regulation, pharmacology, and toxicology (Tremblay and Hamet, 2013; Drozda et al., 2018). However, many issues have to be addressed including genomic data quality and assays' accreditation.

According to the European Medicines Agency (EMA) guidelines, there is a regulatory framework defined by Good Clinical Practice (DCP) compliance (European Medicines Agency, 2001/2005), Good Laboratory Practice (GLP) compliance (European Medicines Agency, 2015), Good Manufacturing Practice (GMP), and Good Distribution Practice (GDP) (European Medicines Agency, 2001), while recently a guideline for Good Pharmacogenomics Practice has been produced (European Medicines Agency, 2018). In particular, this guideline stresses the importance of all steps included in any next-generation sequencing protocol from DNA extraction, DNA processing, preparation of libraries, generation of sequence reads and base calling, sequence mapping, variant annotation and filtering, variant classification, and interpretation. According to this guideline, a crucial parameter for next-generation sequencing analysis is the minimum sequencing coverage, which in case of germline pharmacovariants should be at least 30×, while in case of rare variants, a higher coverage is needed in order to ensure that also the rarer variants are detected by the sequencing. Also, in case of highly homologous genes and pseudogenes, that can contribute to miscalled variants due to sequencing artifacts, it is recommended to include methods that use substantially longer read lengths, i.e., fragments longer than 1000 base pairs.

This guideline portfolio has been developed to ensure the quality of medical products and services. The transfer of this policy to pharmacogenomics assays is critical, since numerous studies have pointed sources of inter- and intra-laboratory error and variability in experimental results (Ji and Davis, 2006). The quality issues of pharmacogenomics rely on the genomic complexity of the region of interest that can impact accuracy and precision of an assay. Consequently, it is important to understand and give due consideration to assay design (Pant et al., 2014), especially when it comes to next-generation sequencing.

Additionally, the validation of the discovery findings coming from pharmacogenomics studies in large randomized clinical trials is often difficult, due to high costs and ethical considerations (Wheeler et al., 2013). In the case of prospective clinical trials, specific drug-dosing schedules are used, providing consistent and well-maintained drug data for pharmacogenomics studies. To increase the sample size for a particular phenotype, it may be useful to combine data from the treatment arms of a clinical trial and then, control for potential confounding,

owing to treatment differences during the statistical analyses. In this context, cancer pharmacogenomics studies have shown promising results, although replication may still be an issue (Hyman et al., 2015). Currently, there are not enough well phenotypic patient data sets for most cancer drugs under investigation to make replication studies feasible, especially when effect sizes are small (Spencer et al., 2009; Daly, 2010). Despite the limitations and difficulties with samples' size, cancer pharmacogenomics studies have demonstrated the potential to make therapy safer and more effective for patients (Spencer et al., 2009; Daly, 2010; Wheeler et al., 2013).

CONSULTATION

There is uncertainty about the ways that the results of pharmacogenomics can be translated into clinical care decisions by the government agencies. This reflects the complex genetic interactions, the paucity of evidence (in some cases) as well as the legal constraints by the regulatory bodies. As a consequence, health professionals are in a vulnerable position (Maliepaard et al., 2013; Trent et al., 2013). This status is imprinted by the United States FDA policy that orders every pharmacogenomics product to provide any relative information available, but without any use recommendation (Maliepaard et al., 2013; Trent et al., 2013). Uncertainty and lack of information ask for additional pressure on professional societies to develop the appropriate clinical practice guidelines to ensure that patient care is not compromised or unnecessary genetic testing is avoided (Maliepaard et al., 2013; Trent et al., 2013). No doubt, multiple sources of information on pharmacogenomics tests can create confusion in clinical decision-making. To overcome this, PharmGKB⁴ was established to consolidate datasets into one curated database, where users can query for drug, gene, disease or metabolic pathway to obtain information such as drug properties, pathway diagrams as well as related publications in a centralized manner. Also, the

Clinical Pharmacogenetics Implementation Consortium (CPIC⁵) and the Dutch Pharmacogenetics Working Group (Dutch Pharmacogenetics Working Group, 2005) have issued guidelines per gene-drug combination assisting healthcare professional to interpret pharmacogenomic testing results and reciprocally adjust the dose or select an alternative drug.

CONCLUDING REMARKS

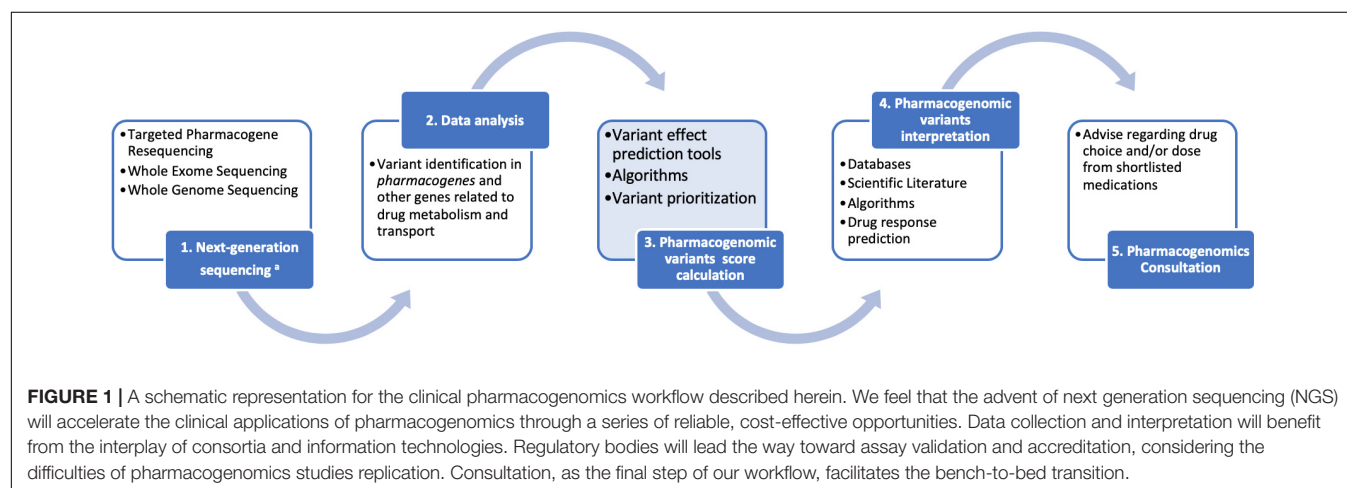
In the era of big data and -omics technologies, the translation of pharmacogenomics in the clinic has yet to be met. This does not only refer to next-generation sequencing-based genotyping but also the more easily applicable low-to-medium throughput (single variant to panel-based) genotyping. Nevertheless, next-generation sequencing will soon be part of the clinical reality and as such, one of the first areas that will be readily applicable is the rationalization of drug use.

Depending on the available resources and infrastructure, application of next-generation sequencing in pharmacogenomics will vary from targeted pharmacogene resequencing in low resource settings, be it either in a panel-based format per drug categories (e.g., cardiovascular diseases, oncology, psychiatric diseases, etc.) or in a more comprehensive pre-emptive pharmacogenomics format including as many pharmacogenes as possible. In those settings, where whole exome, or – ideally – whole genome, sequencing is available, then pharmacogenomic variant identification will be performed simultaneously with the disease genetic diagnosis, focusing only on those variants in the pharmacogenes. As such, the following workflow is recommended for clinical pharmacogenomics (outlined in **Figure 1**):

- (1) Next generation sequencing (targeted pharmacogene resequencing, whole exome and/or whole genome sequencing) will be performed in duly accredited laboratories, following the established guidelines for good pharmacogenomics and other practices,

⁴www.pharmgkb.org

⁵www.cpicpgx.org



- (2) Data analysis will follow for the identification of the common, rare and even novel, genomic variants in the pharmacogenes and other genes involved and/or related to drug metabolism and transport, using dedicated data analysis software packages,
- (3) An integral part of the pharmacogenomic variant annotation will be the calculation and assignment, to each pharmacovariants, of a specific score. This score will be calculated based on certain criteria, such as: (i) the variant itself [known (well established function) or novel (not functionally validated)], (ii) in case of novel variants or variants of unknown significance, the nature of the variant itself (nonsense, frameshift, non-synonymous, other, etc.), the effect of which will be determined *in silico* by a battery of variant prediction tools, (iii) the variant's frequency in the population (common or rare), (iv) the existing evidence of the pharmacogene role, in which the variant is identified, in drug metabolism and transport, deducted from the various databases [e.g., Level A–D (for CPIC) or Level 1–4 (for PharmGKB), etc]. Subsequently, variant prioritization will be performed, based on these scores,
- (4) After pharmacogenomic variants are prioritized, their interpretation will follow, based on the scientific literature, databases, algorithms, from which the corresponding drug response predictions will be derived, also in conjunction with recommendation resources, such as the CPIC, PharmGKB, or the Dutch Pharmacogenomics Working Group, and lastly,
- (5) Pharmacogenomics consultation, performed by a qualified pharmacogenomics expert or clinical geneticist, that will include the provision of advice regarding the drug choice from a shortlist of suggested medications to avoid adverse drug reactions and/or to ensure the optimal drug treatment.

Such a pharmacogenomics scoring system is currently being developed (Patrinos GP, unpublished) to facilitate integration of next-generation sequencing for pharmacogenomics into the routine clinical care. In addition, there are further opportunities for omics-related disciplines, beyond genomics, to be employed for personalized drug response predictions, namely pharmacoeigenomics (Lauschke et al., 2018), pharmacometagenomics (Balasopoulou et al., 2016) and/or pharmacometabolomics (Balasopoulou et al., 2016; Balashova et al., 2018).

We feel that proper implementation of the proposed workflow for next-generation sequencing-based pharmacogenomic testing can occur only via the synergy of all stakeholders and their will to implement the current technological advances, in this case, next generation sequencing and information technologies. In cancer, particularly, such a synergy would be greatly beneficial toward the enigmatic complexity of the disease and great individual variability.

AUTHOR CONTRIBUTIONS

All authors have compiled and approved the manuscript.

FUNDING

This work has been partly funded by a European Commission grant (Ubiquitous Pharmacogenomics (U-PGx); H2020-668353) to GP and encouraged by the Genomic Medicine Alliance Pharmacogenomics Working Group. GP is a Full Member of the European Medicines Agency, CHMP-Pharmacogenomics Working Party.

REFERENCES

- Balashova, E. E., Maslov, D. L., and Lokhov, P. G. (2018). A metabolomics approach to pharmacotherapy personalization. *J. Pers. Med.* 8:E28. doi: 10.3390/jpm8030028
- Balasopoulou, A., Patrinos, G. P., and Katsila, T. (2016). Pharmacometabolomics informs viromics toward precision medicine. *Front. Pharmacol.* 7:411. doi: 10.3389/fphar.2016.00411
- Bell, G. C., Crews, K. R., Wilkinson, M. R., Haidar, C. E., Hicks, J. K., Baker, D. K., et al. (2014). Development and use of active clinical decision support for preemptive pharmacogenomics. *J. Am. Med. Inform. Assoc.* 21, e93–e99. doi: 10.1136/amiajnl-2013-001993
- Daly, A. K. (2010). Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet.* 11, 241–246. doi: 10.1038/nrg2751
- Drozda, K., Pacanowski, M. A., Grimstein, C., and Zineh, I. (2018). Pharmacogenetic labeling of FDA-approved drugs: a regulatory retrospective. *JACC Basic Transl. Sci.* 3, 545–549. doi: 10.1016/j.jacbs.2018.06.001
- Dutch Pharmacogenetics Working Group (2005). *Pharmacogenetics*. Available at: <https://www.knmp.nl/patientenzorg/medicatiebewaking/farmacogenetica/pharmacogenetics-1> (accessed February 21, 2019).
- Ehmann, F., Caneva, L., Prasad, K., Paulmichl, M., Maliepaard, M., Llerena, A., et al. (2015). Pharmacogenomic information in drug labels: European medicines agency perspective. *Pharmacogenomics J.* 15, 201–210. doi: 10.1038/tj.2014.86
- European Medicines Agency (2001). *Good Distribution Practice*. Available at: <https://www.ema.europa.eu/en/human-regulatory/post-authorisation/compliance/good-distribution-practice> (accessed February 21, 2019).
- European Medicines Agency (2015). *Good Laboratory Practice Compliance*. Available at: <https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-laboratory-practice-compliance> (accessed February 21, 2019).
- European Medicines Agency (2018). *Good pharmacogenomic practice*. Available at: <https://www.ema.europa.eu/en/good-pharmacogenomic-practice> (accessed February 21, 2019).
- European Medicines Agency (2001/2005). *Good Clinical Practice*. Available at: <https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-clinical-practice-compliance> (accessed February 21, 2019).
- Evans, W. E., and Relling, M. V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286, 487–491. doi: 10.1126/science.286.5439.487
- Freedman, A. N., Sansbury, L. B., Figg, W. D., Potosky, A. L., Weiss Smith, S. R., Khoury, M. J., et al. (2010). Cancer pharmacogenomics and pharmacoepidemiology: setting a research agenda to accelerate translation. *J. Natl. Cancer Inst.* 102, 1698–1705. doi: 10.1093/jnci/djq390
- Gillis, N. K., Patel, J. N., and Innocenti, F. (2014). Clinical Implementation of germline cancer pharmacogenetic variants during the next-generation sequencing era. *Clin. Pharmacol. Ther.* 95, 269–280. doi: 10.1038/clpt.2013.214
- Harper, A. R., and Topol, E. J. (2012). Pharmacogenomics in clinical practice and drug development. *Nat. Biotechnol.* 30, 1117–1124. doi: 10.1038/nbt.2424

- Hyman, D. M., Puzanov, I., Subbiah, V., Faris, J. E., Chau, I., Blay, J.-Y., et al. (2015). Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *New Engl. J. Med.* 373, 726–736. doi: 10.1056/NEJMoa1502309
- Ingelman-Sundberg, M., Mkrtychian, S., Zhou, Y., and Lauschke, V. M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* 12:26. doi: 10.1186/s40246-018-0157-3
- Ji, H., and Davis, R. W. (2006). Data quality in genomics and microarrays. *Nat. Biotechnol.* 24, 1112–1113. doi: 10.1038/nbt0906-1112
- Kalman, L. V., Agúndez, J., Appell, M. L., Black, J. L., Bell, G. C., Boukouvala, S., et al. (2015). Pharmacogenetic allele nomenclature: international workgroup recommendations for test result reporting. *Clin. Pharmacol. Ther.* 99, 172–185. doi: 10.1002/cpt.280
- Karageorgos, I., Mizzi, C., Giannopoulou, E., Pavlidis, C., Peters, B. A., Zagoriti, Z., et al. (2015). Identification of cancer predisposition variants in apparently healthy individuals using a next-generation sequencing-based family genomics approach. *Hum. Genomics* 9:12. doi: 10.1186/s40246-015-0034-2
- Kitzmillier, J. P., Groen, D. K., Phelps, M. A., and Sadee, W. (2011). Pharmacogenomic testing: relevance in medical practice: why drugs work in some patients but not in others. *Cleveland Clin. J. Med.* 78, 243–257. doi: 10.3949/ccjm.78a.10145
- Ku, C. S., Cooper, D. N., and Patrinos, G. P. (2016). The Rise and rise of exome sequencing. *Publ. Health Genomics* 19, 315–324. doi: 10.1159/000450991
- Lam, Y. W. F. (2013). Scientific challenges and implementation barriers to translation of pharmacogenomics in clinical practice. *ISRN Pharmacol.* 2013:641089. doi: 10.1155/2013/641089
- Lauschke, V. M., Barragan, I., and Ingelman-Sundberg, M. (2018). Pharmacoeugenetics and toxicoeugenetics: novel mechanistic insights and therapeutic opportunities. *Annu. Rev. Pharmacol. Toxicol.* 58, 161–185. doi: 10.1146/annurev-pharmtox-010617-053021
- Lauschke, V. M., and Ingelman-Sundberg, M. (2018). How to consider rare genetic variants in personalized drug therapy. *Clin. Pharmacol. Ther.* 103, 745–748. doi: 10.1002/cpt.976
- Maliepaard, M., Nofziger, C., Papaluca, M., Zineh, I., Uyama, Y., Prasad, K., et al. (2013). Pharmacogenetics in the evaluation of new drugs: a multiregional regulatory perspective. *Nat. Rev. Drug Discov.* 12, 103–115. doi: 10.1038/nrd3931
- McCarty, C. A., Chisholm, R. L., Chute, C. G., Kullo, I. J., Jarvik, G. P., Larson, E. B., et al. (2011). The eMERGE network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* 4:13. doi: 10.1186/1755-8794-4-13
- Mitropoulou, C., Mai, Y., van Schaik, R. H., Vozikis, A., and Patrinos, G. P. (2014). Stakeholder analysis in pharmacogenomics and genomic medicine in greece. *Publ. Health Genomics* 17, 280–286. doi: 10.1159/000365896
- Mizzi, C., Peters, B., Mitropoulou, C., Mitropoulos, K., Katsila, T., Agarwal, M. R., et al. (2014). Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* 15, 1223–1234. doi: 10.2217/pgs.14.102
- O'Donnell, P. H., and Ratain, M. J. (2012). Germline pharmacogenomics in oncology: decoding the patient for targeting therapy. *Mol. Oncol.* 6, 251–259. doi: 10.1016/j.molonc.2012.01.005
- Pant, S., Weiner, R., and Marton, M. J. (2014). Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays and companion diagnostics. *Front. Oncol.* 4:78. doi: 10.3389/fonc.2014.00078
- Potamias, G., Lakiotaki, K., Katsila, T., Lee, M. T. M., Topouzis, S., Cooper, D. N., et al. (2014). Deciphering next-generation pharmacogenomics: an information technology perspective. *Open Biol.* 4:140071. doi: 10.1098/rsob.140071
- Reisberg, S., Krebs, K., Lepamets, M., Kals, M., Mägi, R., Metsalu, K., et al. (2019). Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. *Genet. Med.* doi: 10.1038/s41436-018-0337-5 (Epub ahead of print).
- Simon, R., and Roychowdhury, S. (2013). Implementing personalized cancer genomics in clinical trials. *Nat. Rev. Drug Discov.* 12, 358–369. doi: 10.1038/nrd3979
- Spencer, C. C., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 5:e1000477. doi: 10.1371/journal.pgen.100477
- Stanek, E. J., Sanders, C. L., Taber, K. A. J., Khalid, M., Patel, A., Verbrugge, R. R., et al. (2012). Adoption of pharmacogenomic testing by US physicians: results of a nationwide survey. *Clin. Pharmacol. Therapeut.* 91, 450–458. doi: 10.1038/clpt.2011.306
- Topol, E. J. (2010). Pharmacy benefit managers, pharmacies, and pharmacogenomic testing: prescription for progress?. *Sci. Transl. Med.* 2:44cm22. doi: 10.1126/scitranslmed.3001067
- Tremblay, J., and Hamet, P. (2013). Role of genomics on the path to personalized medicine. *Metabolism* 62(Suppl. 1), S2–S5. doi: 10.1016/j.metabol.2012.08.023
- Trent, R. J., Cheong, P. L., Chua, E. W., and Kennedy, M. A. (2013). Progressing the utilisation of pharmacogenetics and pharmacogenomics into clinical care. *Pathology* 45, 357–370. doi: 10.1097/PAT.0b013e328360b66e
- Tsiliki, G., Karacapilidis, N., Christodoulou, S., and Tzagarakis, M. (2014). Collaborative mining and interpretation of large-scale data for biomedical research insights. *PLoS One* 9:e108600. doi: 10.1371/journal.pone.0108600
- Wheeler, H. E., Maitland, M. L., Dolan, M. E., Cox, N. J., and Ratain, M. J. (2013). Cancer pharmacogenomics: strategies and challenges. *Nat. Rev. Genet.* 14, 23–34. doi: 10.1038/nrg3352

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Giannopoulou, Katsila, Mitropoulou, Tsermpini and Patrinos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Star Allele-Based Haplotyping versus Gene-Wise Variant Burden Scoring for Predicting 6-Mercaptopurine Intolerance in Pediatric Acute Lymphoblastic Leukemia Patients

Yoomi Park¹, Hyery Kim², Jung Yoon Choi^{3,4}, Sunmin Yun¹, Byung-Joo Min¹, Myung-Eui Seo¹, Ho Joon Im², Hyoung Jin Kang^{3,4} and Ju Han Kim^{1,5*}

OPEN ACCESS

Edited by:

Martin A. Kennedy,
University of Otago, New Zealand

Reviewed by:

Chakradhara Rao
Satyanarayana Uppugunduri,
Université de Genève,
Switzerland
Maria J. Prata,
University of Porto, Portugal
William Newman,
University of Manchester,
United Kingdom

*Correspondence:

Ju Han Kim
juhan@snu.ac.kr

Specialty section:

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Pharmacology

Received: 14 January 2019

Accepted: 20 May 2019

Published: 11 June 2019

Citation:

Park Y, Kim H, Choi JY, Yun S,
Min B-J, Seo M-E, Im HJ, Kang HJ
and Kim JH (2019) Star Allele-Based
Haplotyping versus Gene-Wise
Variant Burden Scoring for Predicting
6-Mercaptopurine Intolerance in
Pediatric Acute Lymphoblastic
Leukemia Patients.
Front. Pharmacol. 10:654.
doi: 10.3389/fphar.2019.00654

¹ Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, South Korea, ² Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea, ³ Department of Pediatrics, Seoul National University College of Medicine, Seoul, South Korea, ⁴ Seoul National University Cancer Research Institute, Seoul, South Korea, ⁵ Center for Precision Medicine, Seoul National University Hospital, Seoul, South Korea

Nudix Hydrolase 15 (NUDT15) and *Thiopurine S-Methyltransferase (TPMT)* are strong genetic determinants of thiopurine toxicity in pediatric acute lymphoblastic leukemia (ALL) patients. Since patients with *NUDT15* or *TPMT* deficiency suffer severe adverse drug reactions, star (*) allele-based haplotypes have been used to predict an optimal 6-mercaptopurine (6-MP) dosing. However, star allele haplotyping suffers from insufficient, inconsistent, and even conflicting designations with uncertain and/or unknown functional alleles. Gene-wise variant burden (GVB) scoring enables us to utilize next-generation sequencing (NGS) data to predict 6-MP intolerance in children with ALL. Whole exome sequencing was performed for 244 pediatric ALL patients under 6-MP treatments. We assigned star alleles with PharmGKB haplotype set translational table. GVB for *NUDT15* and *TPMT* was computed by aggregating *in silico* deleteriousness scores of multiple coding variants for each gene. Poor last-cycle dose intensity percent (DIP < 25%) was considered as 6-MP intolerance, resulting therapeutic failure of ALL. DIPs showed significant differences ($p < 0.05$) among *NUDT15* poor (PM, $n = 1$), intermediate (IM, $n = 48$), and normal (NM, $n = 195$) metabolizers. *TPMT* exhibited no PM and only seven IMs. GVB showed significant differences among the different haplotype groups of both *NUDT15* and *TPMT* ($p < 0.05$). Kruskal–Wallis test for DIP values showed statistical significances for the seven different GVB score bins of *NUDT15*. GVB^{*NUDT15*} outperformed the star allele-based haplotypes in predicting patients with reduced last-cycle DIPs at all DIP threshold levels (i.e., 5%, 10%, 15%, and 25%). In *NUDT15*-and-*TPMT* combined interaction analyses, GVB^{*NUDT15,TPMT*} outperformed star alleles [area under the receiver operating curve (AUROC) = 0.677 vs. 0.645] in specificity (0.813 vs. 0.796), sensitivity (0.526 vs. 0.474), and positive (0.192 vs. 0.164) and negative (0.953 vs. 0.947) predictive values. Overall, GVB correctly classified five more patients (i.e., one into *below* and four into *above* 25% DIP groups) than did star allele haplotypes. GVB analysis demonstrated that 6-MP intolerance in pediatric ALL

can be reliably predicted by aggregating NGS-based common, rare, and novel variants together without hampering the predictive power of the conventional haplotype analysis.

Keywords: 6-mercaptopurine, drug toxicity, variant burden, pharmacogenetics, pharmacogenomics, next-generation sequencing, Nudix Hydrolase 15 (NUDT15), Thiopurine S-Methyltransferase (TPMT)

INTRODUCTION

6-Mercaptopurine (6-MP) is a commonly used drug in the maintenance therapy of pediatric acute lymphoblastic leukemia (ALL). Since patients have a potential to experience medication-induced life-threatening side effects including bone marrow suppression and hepatotoxicity, providing a tailored drug dosing regimen is essential in clinical practice (Vogenberg et al., 2010).

One of the strongest ways to determine initial 6-MP dose is an experimental assessment of potential for drug adverse reactions, such as severe neutropenia by monitoring 6-MP metabolite concentration or using *in vitro* activity profiles (Dubinsky et al., 2000; Ansari et al., 2002; Cuffari, 2005; Bradford, 2011; Supandi et al., 2018). However, applying such methods into routine clinical practice for predicting drug-induced toxicity is still challenging because it is extremely time-consuming, expensive, and inefficient (González-Lama and Gisbert, 2016).

As recent studies have demonstrated the strong association between genetic polymorphisms and inter-individual variability in 6-MP dose intensity, approaches to predict drug tolerance on the basis of individual genomic profiles have arisen. The primary genetic determinant of thiopurine toxicity is *TPMT*, which plays a crucial role in identifying patients in need of treatment modification with reduced enzyme activity (Lennard, 2014). However, this has not been applicable to East Asian populations since the frequency of *TPMT* polymorphisms varies by ethnicity (Relling et al., 2013). Recently, a novel pharmacogenetic marker, *NUDT15*, has clarified its role in predicting thiopurine toxicity in Asian populations (Yang et al., 2014; Yang et al., 2015; Zgheib et al., 2016; Kakuta et al., 2017). Clinical Pharmacogenetics Implementation Consortium (CPIC) published an updated guideline for thiopurine dosing based on both *TPMT* and *NUDT15* genotypes using the star allele-based dose prediction method (Relling and Klein, 2009; Relling et al., 2018). This prevailing method provides therapeutic recommendations for dosing based on star allele genotypes. However, the utilization of star alleles in clinical practice has many obstacles that occur mainly due to 1) the extremely complex nomenclature system, 2) the limited resolution of phenotype prediction due to many unknown and uncertain function alleles, 3) ignorance of functional impacts of rare and/or novel variants, and 4) limited use in previously studied populations only (Robarge et al., 2007). Next-generation sequencing (NGS) challenges the conventional star alleles on the basis of genotyping technologies and clinical studies in case-control settings.

In the era of NGS, the comprehensive genotyping capabilities of NGS platform have enabled us to capture the true diversity of gene variation, and researchers propose alternative ways to predict individual intolerance towards a drug. One promising method is a gene-wise variant burden (GVB) scoring approach that can

calculate gene-wise cumulative variant deleteriousness scores including common, rare, and even novel genetic variants for each gene (Lee et al., 2016). Here, we assessed the utility of GVB scoring method in quantifying the potential contributing effect of variants on enzymatic activity. By combining the clinically proven and well-established associations between the two genes, i.e., *NUDT15* and *TPMT*, and 6-MP dose intensity percent (DIP, actual/planned dose) as a clinical endpoint, we performed a comparison study of the conventional star allele-based haplotyping and GVB scoring methods for predicting the last-cycle 6-MP DIP as an indicator for 6-MP intolerance of ALL patients with *NUDT15* and/or *TPMT* deficiency. Overall, both star alleles and GVB showed significant correlations with 6-MP DIP values. Star allele-based haplotype groups showed significant correlation with GVB score groups. For predicting reduced last-cycle DIP values, GVB analysis outperformed the conventional star allele method for *NUDT15* and showed comparable result for *TPMT*. In *NUDT15*-and-*TPMT* combined interaction analyses, GVB^{*NUDT15,TPMT*} outperformed star allele-based predictions [area under the receiver operating curve (AUROC) = 0.677 vs. 0.645] in specificity (0.813 vs. 0.796), sensitivity (0.526 vs. 0.474), and positive (PPV; 0.192 vs. 0.164) and negative (NPV; 0.953 vs. 0.947) predictive values. It is demonstrated that gene-wise evaluation of *in silico* deleterious variant score burden can be a useful method for predicting 6-MP intolerance in pediatric ALL patients, considering NGS-based common, rare, and novel variants concurrently while not hampering the predictive power of the conventional haplotype analysis.

MATERIALS AND METHODS

Patients and Clinical Data Collection

A total of 298 Korean pediatric ALL patients with 6-MP treatment during maintenance therapy were recruited in the present study from two major teaching hospitals, i.e., Asan Medical Center (AMC) and Seoul National University Hospital (SNUH). Of the 298 subjects, 244 individuals who did not meet the exclusion criteria (i.e., relapse of the disease, stem cell transplantation, Burkitt's lymphoma, mixed phenotype acute leukemia, infant ALL, or very high risk) were selected. All participants provided written informed consent. The study was approved by the AMC Review Boards and the SNUH Review Boards. The 6-MP dose per meter body surface area over a 12-week cycle was recorded. The maximum tolerated dose of 6-MP was defined as the dose at the last maintenance cycle for each patient. Patients from two hospitals had received treatment under the same treatment protocol and dose adjustment guidelines to maintain the ANC levels within target levels (500–1,500/ μ L). Genotype-guided dose modification was not conducted. Additional demographic data are shown in **Table 1**.

TABLE 1 | Clinical characteristics of study subjects.

Characteristics	Study cohorts		
	AMC	SNUH	Total
No. of subjects	95	149	244
Age at diagnosis (year) , mean \pm SD [†]	5.23 \pm 1.8	8.57 \pm 4.6	7.26 \pm 4.1
Sex			
Male	52	93	145
Female	43	56	99
Last-cycle 6-MP dose (mg/m²/day) , mean \pm SD (N)			
6-MP < 12.5	8.14 \pm 1.7 (2)	6.25 \pm 2.9 (4)	6.88 \pm 2.6 (6)
12.5 \leq 6-MP < 25	17.39 \pm 3.4 (4)	19.40 \pm 3.6 (9)	18.78 \pm 3.7 (13)
25 \leq 6-MP < 37.5	32.19 \pm 3.4 (10)	30.72 \pm 4.0 (16)	31.28 \pm 3.8 (26)
37.5 \leq 6-MP < 50	44.52 \pm 3.7 (13)	45.80 \pm 3.5 (14)	45.18 \pm 3.6 (27)
6-MP \geq 50	79.15 \pm 18.1 (66)	78.84 \pm 23.1 (106)	78.96 \pm 21.3 (172)
Total	65.37 \pm 26.6 (95)	65.03 \pm 30.0 (95)	65.16 \pm 28.7 (244)

[†]Data for age at diagnosis were not available for one subject. 6-MP, 6-mercaptopurine; AMC, Asan Medical Center; SNUH, Seoul National University Hospital.

Data Generation and Sequencing

Exome sequencing was performed using Ion AmpliSeq™ Exome panel to screen coding sequence region of entire genome. This panel included the exome of 19,072 genes and the size of the total targeted region was 57.7 Mb. The panel contained 293,903 primer pairs that were multiplexed into 12 pools to avoid primer-dimer formation and interference during PCR. The range of amplicons amplified by these oligo primer pairs ranged from 125 to 275 bp, and the rate of “on target” coverage for this panel was 95.69%. PCR assays were performed directly to amplify 100 ng of genomic DNA samples extracted from normal blood cells in bone marrow aspirates or peripheral blood so as to collect the target regions using the oligo primer pairs of the panel. Reaction parameters were as follows: 99°C for 2 min, followed by 10 cycles of 99°C for 15 s, 60°C for 16 min, and 10°C for 1 min. After amplification, library construction was performed by using the Ion AmpliSeq library kit plus as described in the manufacturer’s instructions (Thermo Scientific, Waltham, MA). Libraries were quantified using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) and then diluted to ~10 pM. Subsequently, 33.3 μ L of the barcoded libraries was combined in sets of three barcodes. The combined libraries were sequenced using the Ion Proton platform with PI chip V3, following the manufacturer’s instructions (Thermo Scientific, Waltham, MA). Reads were mapped to the human reference genome build (hg19) with a mapping alignment program from Thermo Fisher (version 4.4, Torrent Suite Software) on germ-line and low stringency settings (minimum observed allele frequency required for a non-reference variant call is 0.18 for single-nucleotide variant (SNV) and 0.23 for InDel, minimum phred scales call quality is 14 for SNV and 19 for InDel, minimum coverage for called variants is 35 for SNV and 40 for InDel, and maximum strand bias is 0.95 for SNV and 0.75 for InDel). Single-nucleotide variants (SNVs) and short insertions/deletions (InDels) were identified *via* Genome Analysis Toolkit (GATK) 2.8-1 Unified Genotyper (DePristo et al., 2011). To estimate the pathogenicity of variants, two *in silico* variant deleteriousness prediction scores were annotated: sorting intolerant from tolerant (SIFT) (Ng, 2003) and combined annotation dependent depletion (CADD) (Kircher et al., 2014). The protein-coding gene region was defined using ANNOVAR

(<http://annovar.openbioinformatics.org/>) (Wang et al., 2010). All the variants identified in 244 ALL samples are described in **Supplementary Table S1** and **S2**.

Calculation of Gene-Wise Variant Burden Score

Gene-wise deleterious variant burden was computed for *NUDT15* and *TPMT* as described by Lee et al. (2016) and Seo et al. (2018). Under the hypothesis that variants that have potential effects to change protein function not necessarily guarantee but have power to cause harmful phenotypes, only variants with SIFT scores less than 0.7 were further considered.

$$G_i = \{v \mid v \text{ with a SIFT score less than } 0.7\}$$

As SIFT does not provide functional scores for InDels, $_{adj}v$ for all InDel variants were assigned as 1e-8 under the hypothesis that InDels are more deleterious than single-nucleotide substitutions. Considering the dosage effects, adjusted SIFT score $_{adj}v$ was calculated for each variant according to their genotype.

$$_{adj}v_j = \begin{cases} (\text{SIFT score})^{0.5}, & \text{if } v_j \in G_i \text{ and heterozygote} \\ \text{SIFT score} & , \text{if } v_j \in G_i \text{ and homozygote} \end{cases}$$

For each gene G_i with n deleterious variants, we calculated $GVB(G_i)$, the cumulative genic effect for all coding variants of the gene, by calculating the geometric mean of $_{adj}v$ (Equation). GVB_g is considered as 1 if the count n of variant j with scores less than 0.7 in a gene is 0, indicating that the gene is not displaying any deleterious variant.

$$GVB(G_i) = \begin{cases} 1 & , \text{if } n(G_i) = 0 \\ \left(\prod_{j=1}^n _{adj}v_j \right)^{\frac{1}{n}} & , \text{if } n(G_i) > 0 \end{cases}$$

We obtained GVB_g values for each individual ranging from 0 to 1. To predict 6-MP sensitivity, $GVB^{NUDT15,TPMT}$ was generated by calculating the geometric mean of GVB^{NUDT15} and GVB^{TPMT} .

$$GVB^{NUDT15,TPMT} = (GVB^{NUDT15} \times GVB^{TPMT})^{\frac{1}{2}}$$

Prediction of Star Allele Diplotypes for 244 Acute Lymphoblastic Leukemia Samples

To classify 244 ALL samples into three metabolizer groups, we inferred haplotypes using the PHASE 2.1.1 software (Stephens et al., 2001; Stephens and Scheet, 2005) (Supplementary Figure S3). On the basis of the inferred haplotype information, we extracted star alleles that matched the haplotype set translational table from PharmGKB (<https://www.pharmgkb.org/>) (Whirl-Carrillo et al., 2012). Predicted genotypes were translated into molecular phenotypes on the basis of the coded genotype-phenotype translation tables from Moriyama et al. (2016) for *NUDT15* and from PharmGKB tables for *TPMT*.

Estimation of Diagnostic Accuracies by Receiver Operating Curve Analyses

To assess prediction accuracies, we calculated DIP, the percentage of the actual administered dose to the planned dose, as an index for 6-MP drug toxicity. Dose in the last maintenance cycle was used, since the doses of 6-MP in the final maintenance cycle were supposed to be the maximum tolerated doses for patients (Kim et al., 2012). DIP prediction accuracies of GVB (GVB^{NUDT15} , GVB^{TPMT} , and $GVB^{NUDT15,TPMT}$) and star allele-based predictions were compared using AUROC analysis with the R language pROC package (Robin et al., 2011). We computed specificity, sensitivity, PPV, and NPV under the binary classification model with nine different cutoff levels (i.e., 5%, 10%, 15%, 25%, 35%, 45%, 60%, 80%, and 100%) for defining high-risk DIP groups. All statistical analyses were performed using R version 3.5.1.

RESULTS

Relation of Gene-Wise Variant Burden and Star Allele-Based Molecular Phenotypes

NUDT15 and *TPMT* haplotypes of each subject were first inferred from whole exome sequencing (WXS) data by using the PHASE tool, and matched star allele genotypes were assigned for each subject. The star allele genotypes were then translated into three molecular phenotype groups according to their allele combinations; poor (PM, No function|No function), intermediate (IM, Normal|No function or Normal|Decreased), and normal (NM, Normal|Normal) metabolizers. Six and four star alleles were identified for *NUDT15* and *TPMT* genes, respectively, from the 244 ALL patients with their frequencies (Table 2). Table 3 shows the distribution of subsequently predicted enzymatic metabolizer phenotypes for *NUDT15* and *TPMT* among the 244 ALL patients.

While 49 (20.1%) of 244 ALL patients were classified into non-NM (one PM and 48 IMs) phenotype for *NUDT15*, only

TABLE 2 | Alleles identified in 244 ALL samples with known allele functions.

Gene	Number of identified alleles	Alleles identified in 244 ALL samples	Frequencies (%)
<i>NUDT15</i>	6	*1	438 (89.75)
		*2	6 (1.23)
		*3	35 (7.17)
		*4	4 (0.82)
		*5	4 (0.82)
		*6	1 (0.20)
<i>TPMT</i>	4	*1	127 (26.02)
		*1S	354 (72.54)
		*3C	6 (1.23)
		*6	1 (0.20)

Haplotypes were inferred via PHASE 2. Star alleles were assigned by the PharmGKB haplotype set translational table. ALL, acute lymphoblastic leukemia.

TABLE 3 | Distribution of predicted enzymatic metabolizer phenotypes.

Molecular phenotype	Function	<i>NUDT15</i>	<i>TPMT</i>
Poor (%)	No function No function	1 (0.41)	NA
Intermediate (%)	Normal No function	48 (19.67)	6 (2.46)
	Normal Decreased	NA	1 (0.41)
Normal (%)	Normal Normal	195 (79.92)	237 (97.13)
Total (%)		244 (100)	244 (100)

Molecular phenotypes were assigned using the PharmGKB haplotype set translational table. Star (*) allele genotype-to-phenotype correlation was adapted from information available at the Moriyama et al. (*NUDT15*) and the Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline (*TPMT*); NA, not available.

seven (2.9%) IMs were identified for *TPMT*, reflecting ethnic variation of *NUDT15* and *TPMT* variants, in a consistent manner (Table 3). Since individuals with *TPMT* homozygous mutant alleles are rarely observed in East Asian population, none of the patients were classified into the poor metabolizer group. IMs were stratified into two groups: 1) individuals carrying one copy of a normal function allele and one copy of a decreased function allele and 2) individuals carrying one copy of a normal function allele and one copy of no function allele. Carriers of non-functional allele, compared with carriers of decreased function allele, are considered to be at an increased risk for functional decline.

Patients with *NUDT15* normal metabolizing alleles (DIP = 67.608 ± 28.2 , $n = 195$) tolerated significantly higher DIPs of 6-MP than did slow metabolizers [5.712 (PM, $n = 1$), 56.452 ± 28.2 (IM, $n = 48$)] (Figure 1A). Clinical usefulness of the conventional star allele-based classification was successfully demonstrated for *NUDT15* variants in the present study. Due to the small number of non-NM subjects for *TPMT* in Korean ALL patients, the difference of DIPs between NM (65.702 ± 28.4 , $n = 237$) and IM (46.805 ± 35.7 , $n = 7$) did not reach statistical significance ($p = 0.10$, Figure 1B).

GVB scores among different molecular phenotype groups for *NUDT15* (PM = 0.09, IM = 0.248 ± 0.1 , and NM = 0.995 ± 0.1 , Figure 2A) and for *TPMT* (IM = 0.229 ± 0.3 , NM = 1 ± 0.0 , Figure 2B) showed statistically significant differences. The observed positive correlation between our GVB score and the conventional enzymatic metabolizer phenotypes for both

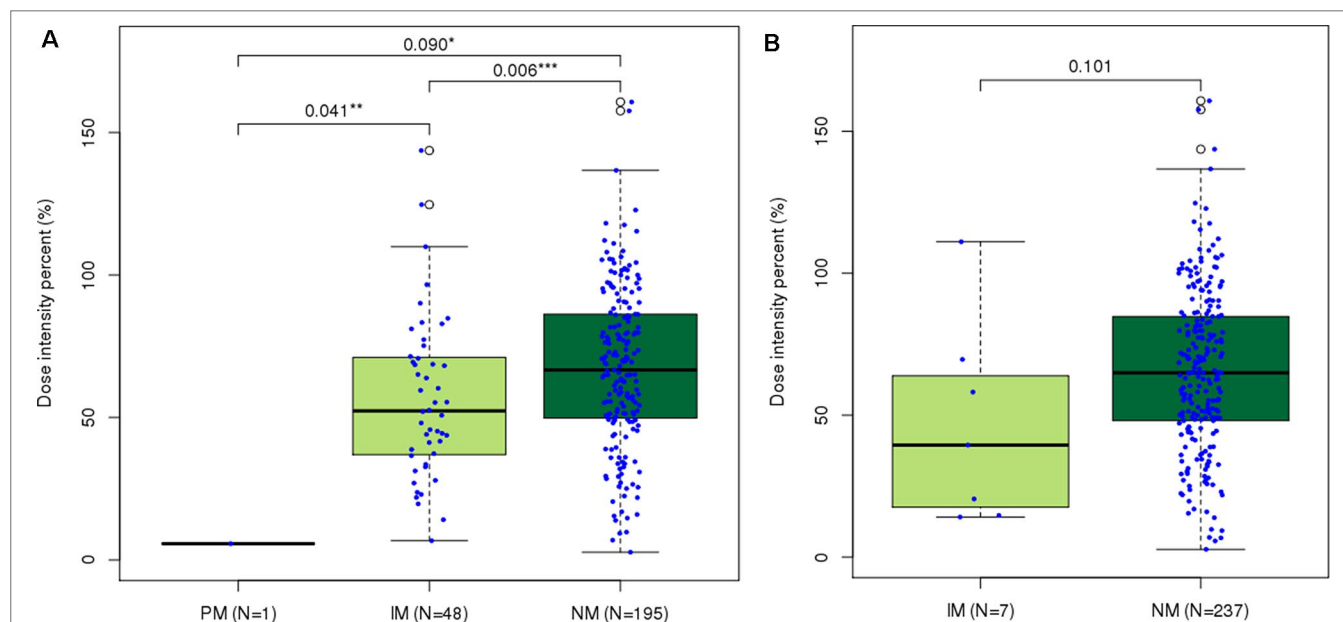


FIGURE 1 | Distribution of last-cycle dose intensity percent of 6-mercaptopurine according to star allele-based molecular phenotype groups in ALL. Dose intensity percent distribution across (A) *Nudix Hydrolase 15 (NUDT15)* and (B) *Thiopurine S-Methyltransferase (TPMT)* molecular phenotype groups. Normal metabolizers of *NUDT15* showed significantly higher dose intensity percent than did intermediate ($p = 0.006$) and poor ($p = 0.090$) metabolizers. * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ by Mann–Whitney U test.

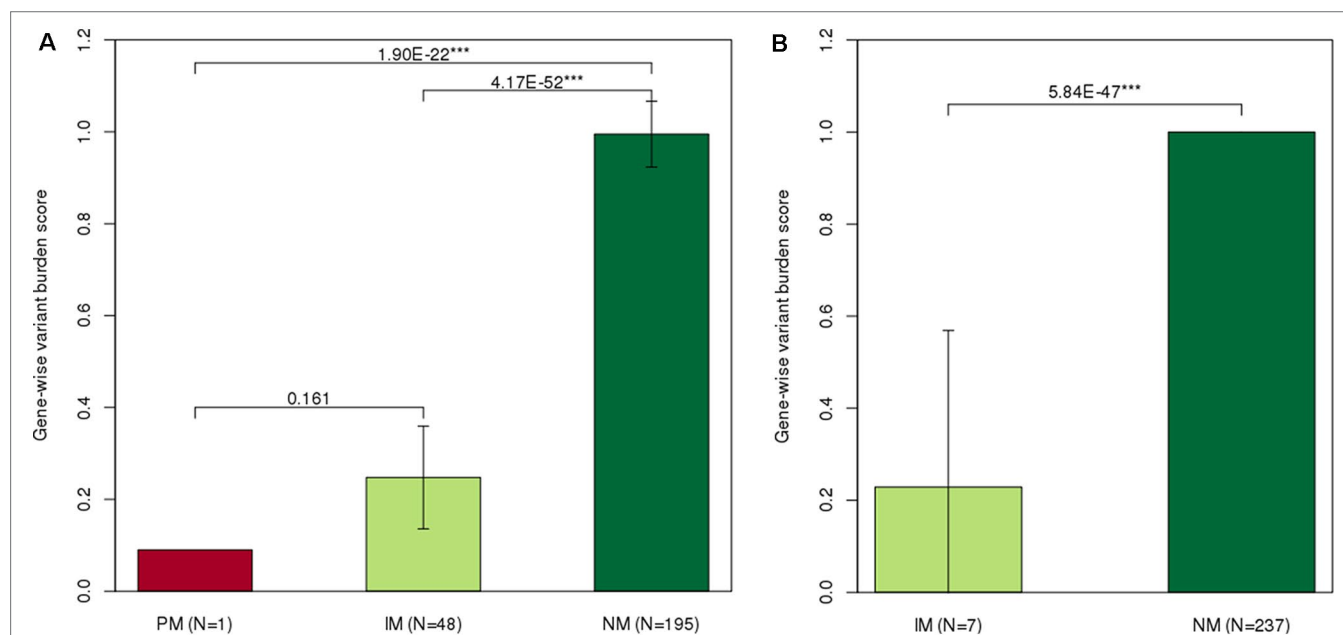


FIGURE 2 | Distribution of gene-wise variant burden (GVB) scores according to the star allele-based molecular phenotype groups. Gene-wise variant burden (GVB) scores across (A) *NUDT15* and (B) *TPMT* molecular phenotype groups. Normal metabolizers showed significantly higher dose intensity percent than did intermediate (*NUDT15*, $p = 4.17E-52$; *TPMT* $p = 5.84E-47$) and poor (*NUDT15*, $p = 1.9E-22$) metabolizers. * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ by Mann–Whitney U test.

NUDT15 and *TPMT* variants strongly supported our further analysis. Note that those pharmacogenetic star alleles have long been empirically developed by clinical case–control studies and/or animal and molecular studies. In contrast, the GVB analysis is based on purely theoretical *ab initio* and *in silico* methods without

requiring empirical studies that are prohibitively costly considering the numerous drugs and genetic variants discovered by NGS technologies and the interactions. In the following sections, we explore the potential of the GVB scoring method for predicting DIPs as an indicator of 6-MP intolerance in pediatric ALL patients.

Gene-Wise Variant Burden Scores for Predicting Last-Cycle 6-Mercaptopurine Dose Intensity Percent

Since both *NUDT15* and *TPMT* genes are not highly variable, only seven and two GVB value groups for *NUDT15* and *TPMT*, respectively, were identified in the 244 ALL patients. $\text{GVB}^{\text{NUDT15}}$ demonstrated statistically significant positive correlation with DIP ($p = 0.016$ by Kruskal–Wallis test, $p = 0.001$ ($\rho = 0.21$) by Spearman's rank correlation, $p = 0.001$ ($\tau = 0.17$) by Kendall's rank correlation) (Figure 3A). Exclusion of the two patients having both *NUDT15* and *TPMT* variants slightly improved statistical significance (Supplementary Figure S4). Due to the low frequency of *TPMT* alleles in East Asian population, 97.5% ($n = 238$) of all ALL patients were classified into wild type ($\text{GVB}^{\text{TPMT}} = 1.00 \pm 0.00$) and only six (2.50%) were classified into variant type ($\text{GVB}^{\text{TPMT}} = 0.10 \pm 0.00$) groups, resulting in poor statistical significance ($p = 0.408$ by *T*-test, $p = 0.272$ ($\rho = 0.07$) by Spearman's rank correlation, $p = 0.271$ ($\tau = 0.06$) by Kendall's rank correlation) (Figure 3B).

Performance Comparisons Between Gene-Wise Variant Burden and Star Allele-Based Molecular Phenotypes Across Different Risk Group Decision Thresholds

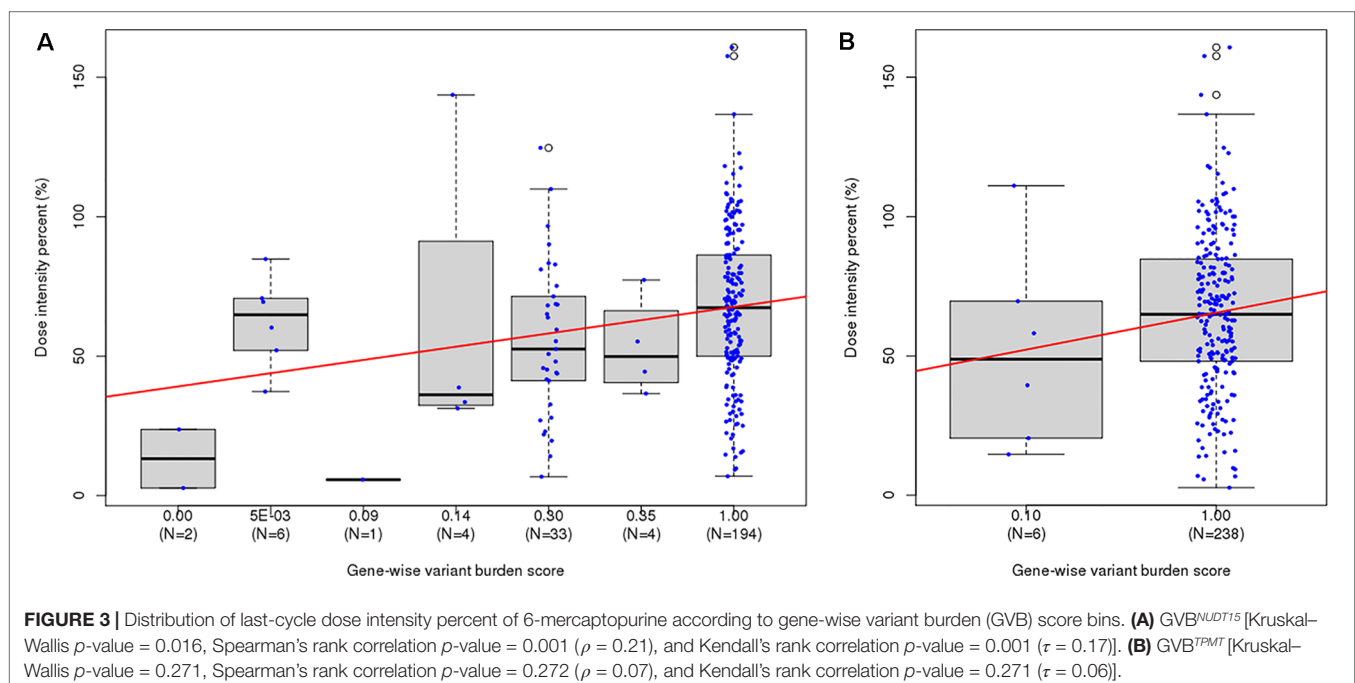
Using ROC analysis, we evaluated the performances of GVB at nine cutoff levels (i.e., DIP < 5%, 10%, 15%, 25%, 35%, 45%, 60%, 80%, and 100%) for defining the 6-MP high-risk groups. Star allele-based classification was also applied for systematic comparison across different DIP threshold levels. DIP below 25% of planned dose of 6-MP is a generally accepted threshold for predicting 6-MP intolerance. Figure 4A demonstrates that $\text{GVB}^{\text{NUDT15}}$ showed better AUCs at all threshold DIP levels below 25% (0.998 (DIP < 5%), 0.676 (DIP < 10%), 0.669 (DIP < 15%),

and 0.653 (DIP < 25%)) than did the conventional star allele-based molecular phenotypes (AUC = 0.618). Moreover, exclusion of the two confounding patients with both *NUDT15* and *TPMT* variant alleles slightly improved performances than did both before-exclusion $\text{GVB}^{\text{NUDT15}}$ at all threshold DIP levels below 25% [AUC = 0.998 (DIP < 5%), 0.676 (DIP < 10%), 0.639 (DIP < 15%), and 0.627 (DIP < 25%)] and the star allele-based (AUC = 0.596) analyses (Figure 4B). Mainly due to the low frequency of *TPMT* variant alleles in East Asian population, both GVB^{TPMT} and star allele-based predictions using *TPMT* seem to show poor AUCs for predicting DIP at all threshold levels (Figure 4C and D).

More importantly, we performed ROC analysis by aggregating the genetic effects of these two genes, *NUDT15* for East Asian and *TPMT* for European heritages. We computed and evaluated $\text{GVB}^{\text{NUDT15,TPMT}}$, which outperformed $\text{GVB}^{\text{NUDT15}}$ or GVB^{TPMT} alone as well as the combined molecular phenotypes of both *NUDT15* and *TPMT* at all DIP threshold levels (Figure 5). In summary, at the clinically important DIP level of below or above 25%, the best AUC values for $\text{GVB}^{\text{NUDT15,TPMT}}$, $\text{GVB}^{\text{NUDT15}}$, GVB^{TPMT} , and combined star alleles were 0.677, 0.653, 0.574, and 0.645, respectively. $\text{GVB}^{\text{NUDT15,TPMT}}$ not only showed the best performance but also successfully included the two confounding patients with both *NUDT15* and *TPMT* variant alleles. While combining GVB scores of multiple genes is simple and straightforward, it is not the case for star alleles, which do not provide a uniform way of combining method for multiple genes.

Comparison of Prediction Accuracies Between Gene-Wise Variant Burden and Star Allele-Based Methods

To test the clinical utility of GVB method for guiding 6-MP dosing and/or for providing systematic framework for clinical studies of



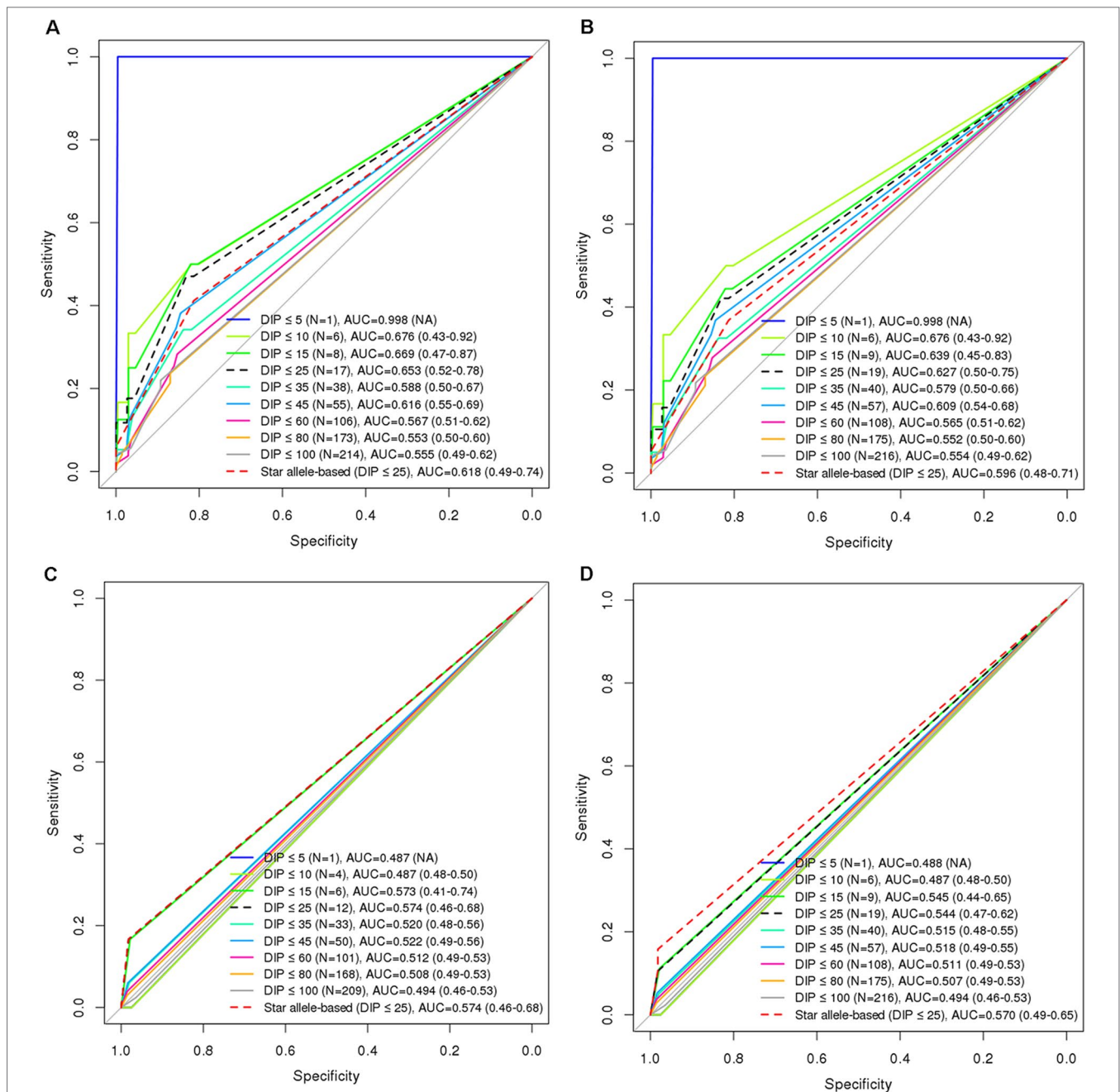
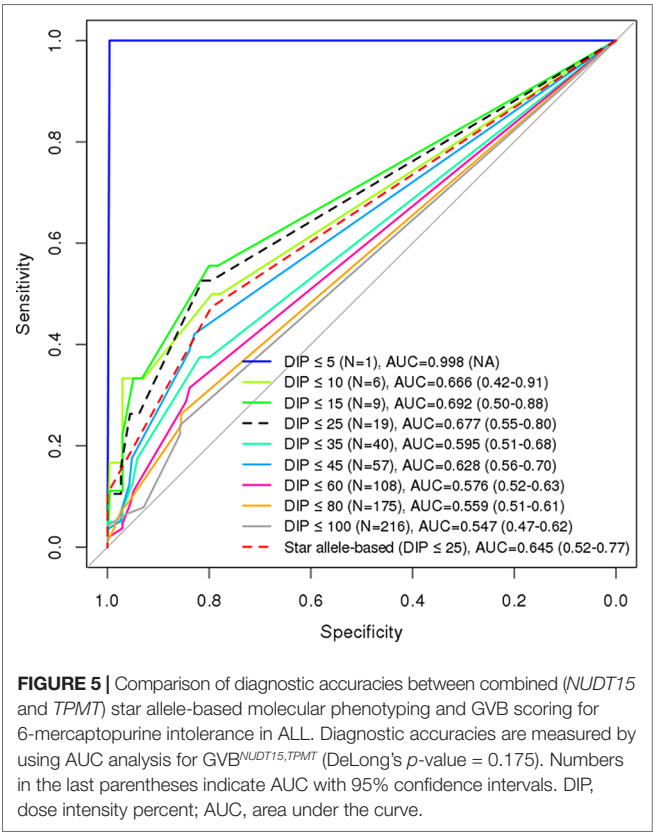


FIGURE 4 | Comparison of diagnostic accuracies between star allele-based molecular phenotyping and GVB scoring for 6-mercaptopurine intolerance in ALL. Diagnostic accuracies are measured by using AUROC analysis for (A) GVB^{NUDT15} excluding two subjects with *TPMT* variants (DeLong's *p*-value = 0.163), (B) GVB^{NUDT15} (DeLong's *p*-value = 0.163), (C) GVB^{TPMT} excluding seven subjects with *NUDT15* variants (DeLong's *p*-value = 0.5), and (D) GVB^{TPMT} (DeLong's *p*-value = 0.841). Numbers in the last parentheses indicate area under the curve (AUC) with 95% confidence intervals. DIP, dose intensity percent; AUC, area under the curve.

6-MP intolerance and its genetic determinants of *NUDT15* and *TPMT* for predicting DIP groups, we evaluated the diagnostic characteristics of the conventional star allele-based and GVB scoring methods in a simulated clinical setting. Table 4A and 4B exhibits diagnostic accuracies for star allele-based molecular phenotype groups and gene-wise variant burden score groups, respectively, for 6-MP intolerance among 244 pediatric ALL patients by the last-cycle DIP of 6-MP. Of the 244 ALL patients, 189 (84.4%) exhibited no *NUDT15* or *TPMT*

variant and hence was classified into NMs for both genes (Table 4A). Of the rest 55 non-NM patients, nine (16.4%) showed DIP below 25%, while 10 of 189 (5.3%) NM patients showed low DIP values.

Although one can choose many threshold levels of GVB, because star alleles can just provide a small number of categories, we chose the most reliable binning threshold of GVB^{NUDT15,TPMT} ≤ 0.3, the cut-point that maximizes the Youden index (Supplementary Figure S5), for classifying the patients into the below and above 25% DIP



groups as shown in **Table 4B**. It is a coincidence that Lee et al. (2016) also suggested GVB^{Pharmacogenes} ≤ 0.3 as the threshold for predicting pharmaceutical market withdrawals in general. GVB^{*NUDT15,TPMT*} correctly classified one more high-risk (DIP ≤ 25%) and four more low-risk (DIP > 25%) patients into the correct-risk groups (**Table 4B**) than did the traditional haplotype-based method (**Table 4A**), with an improved sensitivity from 47.36% to 52.63% and an improved specificity from 79.56% to 81.33%, though the difference did not reach statistical significance (*p*-value for sensitivity = 1 and *p*-value for specificity = 0.134, as determined using a McNemar test). Both PPV and NPV increase from 16.36% to 19.23% and from 94.70% to 95.31%, respectively. Overall, it is suggested that the “computational” GVB^{*NUDT15,TPMT*} is an improved or at least comparable predictor than the “empirical” star allele-based haplotypes for determining subjects with increased risk of 6-MP intolerance in pediatric ALL patients measured by the last-cycle 6-MP DIP.

DISCUSSION

An enduring challenge in precision medicine is to predict adequate drug responses for individual patients (Shah and Shah, 2012). Recent discoveries have revealed a few highly functional and clinically relevant novel variants associated with 6-MP intolerance. However, since implicating drug toxicity based on a single variant is notoriously unreliable as shown in **Supplementary Figure S1** for SIFT and **Supplementary Figure S2** for CADD, developing strategies to aggregate the key effects over a range of genomic

TABLE 4 | Comparison of star allele-based haplotyping versus gene-wise variant burden (GVB^{*NUDT15,TPMT*}) analyses for 6-mercaptopurine intolerance measured by last-cycle dose intensity percent in ALL. Diagnostic accuracy table of **(A)** star allele-based haplotypes and dose intensity percent groups and **(B)** gene-wise variant burden score and dose intensity percent groups.

(A)

NUDT15 and TPMT metabolizer	Dose intensity percent groups		Total	
	≤25	>25		
PM + IM	9	46	55	PPV 16.36% (9/55)
NM	10	179	189	NPV 94.70% (179/189)
Total	19	225	244	
	Sensitivity 47.36% (9/19)	Specificity 79.56% (179/225)		Accuracy 77.05% (188/244)

(B)

Gene-wise variant burden score	Dose intensity percent groups		Total	
	≤25	>25		
GVB ^{NUDT15,TPMT} ≤ 0.3	10	42	52	PPV 19.23% (10/52)
GVB ^{NUDT15,TPMT} > 0.3	9	183	192	NPV 95.31% (183/192)
Total	19	225	224	
	Sensitivity 52.63% (10/19)	Specificity 81.33% (183/225)		Accuracy 79.10% (193/244)

PM, poor metabolizer; IM, intermediate metabolizer; NM, normal metabolizer; PPV, positive predictive value; NPV, negative predictive value.

region is highly required. In the present study, we evaluated the utility of gene-wise deleterious variant burden scoring method, as a sequencing-based, simple, reliable, quantitative, and easy-to-compare score for predicting 6-MP intolerance of 244 pediatric ALL patients. In addition to DIP, GVB showed a statistically significant negative correlation with the incidence of grade 4 neutropenia ($p = 1.43\text{E}-04$ by Kruskal–Wallis test, $p = 3.89\text{E}-07$ ($\rho = -0.32$) by Spearman's rank correlation, and $p = 8.06\text{E}-07$ ($\tau = -0.27$) by Kendall's rank correlation (Supplementary Figure S6). This implies that GVB is a reliable score that can predict hematological toxicity in pediatric ALL patients. When beginning treatment, NGS-based drug intolerance prediction is useful because it is practical to detect patients at high risk of toxicity. For example, patients with low GVB have a high probability of 6-MP toxicity at the initial recommended dose range; thus, clinicians may attempt to reduce the initial target dose of 6-MP. After an initial target dose is determined, a close therapeutic drug monitoring could help to avoid potential causes for toxicity, such as clinically relevant drug–drug interactions, reduced drug clearance due to liver and/or renal impairment, and altered drug utilization due to physiological conditions, as a complementary type of practice during the treatment (Ju-Seop Kang, 2009).

GVB analysis has several benefits over conventional star allele-based approaches. GVB 1) quantitates gene-wise variant burden with a single score; 2) provides a measure of inter-individual genetic variability for each gene; 3) considers common, rare, and novel genetic variants together; 4) provides an ethnic variability-neutral method for studying pharmacogenomics; 5) provides a systematic and reliable framework for designing further pharmacogenomics studies considering many gene interactions for clinical endpoints; and 6) adopts the contributing effect of novel low-frequency variants with potentially reduced function in predicting individual drug toxicity.

Based upon the very recent CPIC updates on *NUDT15*, three newly enrolled alleles were characterized (Moriyama et al., 2017). Since new haplotype designation is highly dependent on the characteristics of the study population, there will be restrictions in incorporating new or even as-yet-unidentified evidences in predicting future drug intensity. GVB can be used to develop a model to determine optimal doses without requiring a multi-ethnic population study, especially for under-studied subpopulations.

The following limitations are inherent in the present study. To evaluate the validity of GVB, independent replication studies for an expanded gene–drug set with sufficient sample sizes in diverse ethnic groups are required as no novel variant was identified in the current study. A conventional single variant-based association test of rare variants requires infeasible magnitude of sample sizes (Bansal et al., 2010), but approaches that aggregate common, rare, and novel variants jointly will substantially reduce a required effective sample sizes (Witte, 2012). The robustness of the analysis framework shall further be improved as novel prognostic markers on 6-MP DIP are acquired. The limitations in interpreting the score includes that all InDels are treated as highly damaging as SIFT provides scores for only single-nucleotide variants. As there are many *in silico* variant deleteriousness scoring method based on different principles, comprehensive evaluation of different method is required (Supplementary Figure S7). We also performed CADD-based computation of GVB values, resulting in similar

results (Supplementary Figures S8 and S9). It has been reported that CADD tends to evaluate in-frame InDels as relatively benign (Kircher et al., 2014). However, recent *in vitro* activity assay of *NUDT15* (Moriyama et al., 2017) proved that in-frame InDel carriers are more likely to be in states with severely diminished response to 6-MP. It is strongly recommended that for clinical applications, potential clinical impacts of genetic variants on drug sensitivity should be further examined to improve estimation accuracy, as *in silico* prediction scores can exhibit incorrect predictions. Producing a custom capture panel for clinically actionable genes could be more cost-effective than an exome-based approach.

One subject who was correctly classified by GVB carried a low-frequency novel deletion and predicted to belong to the high-risk group by GVB, whereas star allele-based prediction classified this patient into the NM group for both *NUDT15* and *TPMT*. The patient required reduced dose than recommended (DIP = 23.7%), supporting that GVB analysis resulted in 6-MP dose-related adverse drug reactions. The patient's variant was heterozygous p.Gly17_Val18del, which was very recently assigned as *NUDT15**9 with uncertain functionality. The other four who were correctly classified by GVB had p.Arg139His on one allele, which has assigned them to the IM (*NUDT15* *1/*4) group. GVB classified them as relatively safe for drug toxicity, and none of them required a 25% reduction from the starting dose. Additionally, one patient who was classified as high risk by GVB was assigned to IM for both *NUDT15* and *TPMT* and required a severely reduced dose (14%), suggesting that GVB^{*NUDT15,TPMT*} exhibits benefits in aggregating effects of many moderate genetic determinants into a single quantitative value.

ETHICS STATEMENT

The study was approved by the Asan Medical Center (AMC) Review Boards and the Institutional Review Board of Seoul National University Hospital (SNUH). Informed written consents for blood sampling and analyses were obtained from all participants.

AUTHOR CONTRIBUTIONS

YP and JK designed the model and the framework. HK, JC, HI, and HK collected samples and clinical data. BM and MS carried out the experiment. YP and SY analyzed the data and carried out the implementation. YP performed the calculations. YP and JK wrote the manuscript. JK conceived the study and was in charge of overall direction and planning. All authors read and approved the final manuscript.

FUNDING

This research was supported by a grant (16183MFDS541) from the Ministry of Food and Drug Safety in 2019.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00654/full#supplementary-material>

REFERENCES

- Ansari, A., Hassan, C., Duley, J., Marinaki, A., Shobowale-Bakre, E. M., Seed, P., et al. (2002). Thiopurine methyltransferase activity and the use of azathioprine in inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 16, 1743–1750. doi: 10.1046/j.1365-2036.2002.01353.x
- Bansal, V., Libiger, O., Torkamani, A., and Schork, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. *Nat. Rev. Genet.* 11, 773–785. doi: 10.1038/nrg2867
- Bradford, K. (2011). Optimizing 6-mercaptopurine and azathioprine therapy in the management of inflammatory bowel disease. *WJG.* 17, 4166–4168. doi: 10.3748/wjg.v17.i37.4166
- Cuffari, C. (2005). A physician's guide to azathioprine metabolite testing. *Gastroenterol. Hepatol.* 2, 58–63. Retrieved from <http://www.gastroenterologyandhepatology.net/>.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Dubinsky, M. C., Lamothe, S., Yang, H. Y., Targan, S. R., Sinnett, D., Théorêt, Y., et al. (2000). Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. *Gastroenterology* 118, 705–713. doi: 10.1016/S0016-5085(00)70140-5
- González-Lama, Y., and Gisbert, J. P. (2016). Monitoring thiopurine metabolites in inflammatory bowel disease. *Frontline Gastroenterol.* 7, 301–307. doi: 10.1136/flgastro-2015-100681
- Ju-Seop Kang, M.-H. L. (2009). Overview of therapeutic drug monitoring. *Korean J. Inter. Med.* 24, 1. doi: 10.3904/kjim.2009.24.1.1
- Kakuta, Y., Kinouchi, Y., and Shimosegawa, T. (2017). Pharmacogenetics of thiopurines for inflammatory bowel disease in East Asia: prospects for clinical application of NUDT15 genotyping. *J. Gastroenterol.* 53, 172–180. doi: 10.1007/s00535-017-1416-0
- Kim, H., Kang, H. J., Kim, H. J., Jang, M. K., Kim, N. H., Oh, Y., et al. (2012). Pharmacogenetic analysis of pediatric patients with acute lymphoblastic leukemia: a possible association between survival rate and ITPA polymorphism. *PLoS ONE* 7, e45558–45510. doi: 10.1371/journal.pone.0045558
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi: 10.1038/ng.2892
- Lee, K. H., Baik, S. Y., Lee, S. Y., Park, C. H., Park, P. J., and Kim, J. H. (2016). Genome sequence variability predicts drug precautions and withdrawals from the market. *PLoS ONE* 11, e0162135–0162115. doi: 10.1371/journal.pone.0162135
- Lennard, L. (2014). Implementation of TPMT testing. *Br. J. Clin. Pharmacol.* 77, 704–714. doi: 10.1111/bcp.12226
- Moriyama, T., Nishii, R., Perez-Andreu, V., Yang, W., Klusmann, F. A., Zhao, X., et al. (2016). NUDT15 polymorphisms alter thiopurine metabolism and hematopoietic toxicity. *Nat. Genet.* 48, 367–373. doi: 10.1038/ng.3508
- Moriyama, T., Yang, Y.-L., Nishii, R., Ariffin, H., Liu, C., Lin, T.-N., et al. (2017). Novel variants in NUDT15 and thiopurine intolerance in children with acute lymphoblastic leukemia from diverse ancestry. *Blood* 130, 1209–1212. doi: 10.1182/blood-2017-05-782383
- Ng, P. C. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509
- Relling, M. V., and Klein, T. E. (2009). CPIC: Clinical Pharmacogenetics Implementation Consortium of the pharmacogenomics research network. *Clin. Pharmacol. Ther.* 89, 464–467. doi: 10.1038/clpt.2010.279
- Relling, M. V., Gardner, E. E., Sandborn, W. J., Schmiegelow, K., Pui, C.-H., Yee, S. W., et al. (2013). Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. *Clin. Pharmacol. Ther.* 93, 324–325. doi: 10.1038/clpt.2013.4
- Relling, M. V., Schwab, M., Whirl-Carrillo, M., Suarez-Kurtz, G., Pui, C.-H., Stein, C. M., et al. (2018). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for thiopurine dosing based on TPMT and NUDT15 genotypes: 2018 update. *Clin. Pharmacol. Ther.* 105, 1095–1105. doi: 10.1002/cpt.1304
- Robarge, J. D., Li, L., Desta, Z., Nguyen, A., and Flockhart, D. A. (2007). The star-allele nomenclature: retooling for translational genomics. *Clin. Pharmacol. Ther.* 82, 244–248. doi: 10.1038/sj.clpt.6100284
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* 12, 77. doi: 10.1186/1471-2105-12-77
- Seo, H., Kwon, E. J., You, Y.-A., Park, Y., Min, B. J., Yoo, K., et al. (2018). Deleterious genetic variants in ciliopathy genes increase risk of ritodrine-induced cardiac and pulmonary side effects. *BMC Med. Genomics* 11, 31. doi: 10.1186/s12920-018-0323-4
- Shah, R. R., and Shah, D. R. (2012). Personalized medicine: is it a pharmacogenetic mirage? *Br J Clin. Pharmacol.* 74, 698–721. doi: 10.1111/j.1365-2125.2012.04328.x
- Stephens, M., and Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76, 449–462. doi: 10.1086/428594
- Stephens, M., Smith, N. J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989. doi: 10.1086/319501
- Supandi, S., Harahap, Y., Harmita, H., and Andalusia, R. (2018). Quantification of 6-mercaptopurine and its metabolites in patients with acute lymphoblastic leukemia using dried blood spots and UPLC-MS/MS. *Sci. Pharm.* 86, 18–18. doi: 10.3390/scipharm86020018
- Vogenberg, F. R., Barash, C. I., and Pursel, M. (2010). Personalized medicine, part 1: evolution and development into therapeutics. *Pharm. Ther.* 35, 560–576. Retrieved from <https://www.ptcommunity.com/>.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164–e164. doi: 10.1093/nar/gkq603
- Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., et al. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* 92, 414–417. doi: 10.1038/clpt.2012.96
- Witte, J. S. (2012). Rare genetic variants and treatment response: sample size and analysis issues. *Stat. Med.* 31, 3041–3050. doi: 10.1002/sim.5428
- Yang, J. J., Landier, W., Yang, W., Liu, C., Hageman, L., Cheng, C., et al. (2015). Inherited NUDT15 variant is a genetic determinant of mercaptopurine intolerance in children with acute lymphoblastic leukemia. *JCO* 33, 1235–1242. doi: 10.1200/JCO.2014.59.4671
- Yang, S.-K., Hong, M., Baek, J., Choi, H., Zhao, W., Jung, Y., et al. (2014). A common missense variant in NUDT15 confers susceptibility to thiopurine-induced leukopenia. *Nat. Genet.* 46, 1017–1020. doi: 10.1038/ng.3060
- Zgheib, N. K., Akika, R., Mahfouz, R., Aridi, C. A., Ghanem, K. M., Saab, R., et al. (2016). NUDT15 and TPMT genetic polymorphisms are related to 6-mercaptopurine intolerance in children treated for acute lymphoblastic leukemia at the Children's Cancer Center of Lebanon. *Pediatr. Blood Cancer* 64, 146–150. doi: 10.1002/pbc.26189

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Park, Kim, Choi, Yun, Min, Seo, Im, Kang and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of Novel Biomarkers for Drug Hypersensitivity After Sequencing of the Promoter Area in 16 Genes of the Vitamin D Pathway and the High-Affinity IgE Receptor

Gemma Amo^{1,2}, Manuel Martí^{1,2}, Jesús M. García-Menaya^{3,4}, Concepción Cordobés^{5,6}, José A. Cornejo-García^{7,8}, Natalia Blanca-López^{9,10}, Gabriela Canto^{9,10}, Inmaculada Doña^{11,12}, Miguel Blanca^{9,10}, María José Torres^{11,12}, José A. G. Agúndez^{1,2} and Elena García-Martín^{1,2*}

OPEN ACCESS

Edited by:

Marcelo Rizzatti Luizon,
Federal University of Minas
Gerais, Brazil

Reviewed by:

Ken Batai,
University of Arizona, United States
Meenal Gupta,
The University of Utah, United States

*Correspondence:

Elena García-Martín
elenag@unex.es

Specialty section:

This article was submitted to
Pharmacogenetics and
Pharmacogenomics,
a section of the journal
Frontiers in Genetics

Received: 12 November 2018

Accepted: 04 June 2019

Published: 25 June 2019

Citation:

Amo G, Martí M, García-Menaya JM,
Cordobés C, Cornejo-García JA,
Blanca-López N, Canto G, Doña I,
Blanca M, Torres MJ, Agúndez JAG
and García-Martín E (2019)
Identification of Novel Biomarkers for
Drug Hypersensitivity After
Sequencing of the Promoter Area in
16 Genes of the Vitamin D Pathway
and the High-Affinity IgE Receptor.
Front. Genet. 10:582.
doi: 10.3389/fgene.2019.00582

¹ University Institute of Molecular Pathology Biomarkers, UEx, Cáceres, Spain, ² ARADyAL Instituto de Salud Carlos III, Cáceres, Spain, ³ Allergy Service, Badajoz University Hospital, Badajoz, Spain, ⁴ ARADyAL Instituto de Salud Carlos III, Badajoz, Spain, ⁵ Allergy Service, Mérida Hospital, Badajoz, Spain, ⁶ ARADyAL Instituto de Salud Carlos III, Cáceres, Spain, ⁷ Research Laboratory, IBIMA, Regional University Hospital of Málaga, UMA, Málaga, Spain, ⁸ ARADyAL Instituto de Salud Carlos III, Cáceres, Spain, ⁹ Allergy Service, Infanta Leonor University Hospital, Madrid, Spain, ¹⁰ ARADyAL Instituto de Salud Carlos III, Madrid, Spain, ¹¹ Allergy Unit, IBIMA, Regional University Hospital of Málaga, UMA, Málaga, Spain, ¹² ARADyAL Instituto de Salud Carlos III, Málaga, Spain

The prevalence of allergic diseases and drug hypersensitivity reactions (DHRs) during recent years is increasing. Both, allergic diseases and DHRs seem to be related to an interplay between environmental factors and genetic susceptibility. In recent years, a large effort in the elucidation of the genetic mechanisms involved in these disorders has been made, mostly based on case-control studies, and typically focusing on isolated SNPs. These studies provide a limited amount of information, which now can be greatly expanded by the complete coverage that Next Generation Sequencing techniques offer. In this study, we analyzed the promoters of sixteen genes related to the Vitamin D pathway and the high-affinity IgE receptor, including *FCER1A*, *MS4A2*, *FCER1G*, *VDR*, *GC*, *CYP2R1*, *CYP27A1*, *CYP27B1*, *CYP24A1*, *RXRA*, *RXRB*, *RXRG*, *IL4*, *IL4R*, *IL13*, and *IL13RA1*. The study group was composed of patients with allergic rhinitis plus asthma (AR+A), patients with hypersensitivity to beta-lactams (BLs), to NSAIDs including selective hypersensitivity (SH) and cross-reactivity (CR), and healthy controls without antecedents of atopy or adverse drug reactions. We identified 148 gene variations, 43 of which were novel. Multinomial analyses revealed that three SNPs corresponding to the genes *FCER1G* (rs36233990 and rs2070901), and *GC* (rs3733359), displayed significant associations and, therefore, were selected for a combined dataset study in a cohort of 2,476 individuals. The strongest association was found with the promoter *FCER1G* rs36233990 SNP that alters a transcription factor binding site. This SNP was over-represented among AR+A patients and among patients with IgE-mediated diseases, as compared with control individuals or with the rest of patients in this study. Classification

models based on the above-mentioned SNPs were able to predict correct clinical group allocations in patients with DHRs, and patients with IgE-mediated DHRs. Our findings reveal gene promoter SNPs that are significant predictors of drug hypersensitivity, thus reinforcing the hypothesis of a genetic predisposition for these diseases.

Keywords: Next-Generation Sequencing (NGS), vitamin D, high-affinity IgE receptor (FC ϵ RI), NSAIDs (non-steroidal anti-inflammatory drugs), beta-lactam antibiotic, drugs hypersensitivity reactions, allergic rhinitis, asthma

INTRODUCTION

The prevalence of atopy, allergic diseases, and drug hypersensitivity reactions (DHRs) is increasing worldwide. In Europe, studies have estimated a prevalence of 20–25% allergic diseases in adults, with many young people being unaware of their disease (Linneberg, 2011; Kruse and Vanijcharoenkarn, 2018), which means an important economic impact for healthcare (European Commission, 2008; Bouvy et al., 2015) reaching an amount from €55 to €151 billion per year in European Union, including indirect costs related to the absence or reduced productivity at work (Zuberbier et al., 2014; Kruse and Vanijcharoenkarn, 2018). Due to their complexity, it is difficult to understand the specific mechanisms and molecules involved in the development of these diseases or to establish a way to prevent or reduce them. Allergic rhinitis (AR) reduces the quality of life by affecting sleep, school, work productivity, and social life. AR is an immunoglobulin E (IgE) mediated inflammatory disease, which is associated with other inflammatory diseases such as asthma. It has been estimated that around 20% of people in the USA and Europe suffer from allergic rhinitis (Durham et al., 2012; Ozdoganoglu and Songu, 2012; Rondon et al., 2017). Taking this into consideration, AR has been classified as a major chronic respiratory disease (Brozek et al., 2017). Drug hypersensitivity reactions (DHR) account for, ~3 to 6% of all hospital admissions. These reactions occur in 10 to 15% of hospitalized patients (Gomes and Demoly, 2005; Szczeklik and Nizankowska-Mogilnicka, 2009; Doña et al., 2014). Beta-lactam antibiotics (BLs) are the most common cause of DHRs mediated by specific immunological mechanisms (Antúnez et al., 2006; Doña et al., 2012, 2014) and, although the mechanisms of how the immune system recognizes these drugs are not fully determined, BLs are considered the classical model of this type of reactions (Blanca et al., 2009). Together with BLs, non-steroidal anti-inflammatory drugs (NSAIDs) are account for the vast majority of DHRs (Gomes et al., 2004; Messaad et al., 2004; Chen et al., 2012; Doña et al., 2012), but, in this case, these DHRs are not exclusively mediated by specific immunological mechanisms (selective hypersensitivity), involving a response to a single drug and good tolerance to other chemically unrelated NSAIDs (Canto et al., 2009; Cornejo-Garcia et al., 2009); but also by nonspecific immunological mechanisms (cross-reactions), which can be caused by more than one chemically unrelated NSAIDs (Agúndez et al., 2012; Kowalski et al., 2013).

Recent investigation proposes the vitamin D pathway among putative factors linked to allergic diseases, because of its important role in immune system (Veldman et al., 2000;

Cantorna et al., 2015) and its direct relation with allergic diseases (Black and Scragg, 2005; Camargo et al., 2007; Benson et al., 2012; Suaini et al., 2015). There are many molecules involved in the vitamin D pathway: hydroxylases from CYP450 family, such as CYP27A1, CYP27B1, CYP2R1, and CYP24A1; the vitamin D binding protein (GC) that acts like a transporter, the vitamin D receptor (VDR), the retinoid receptor X (RXR) and interleukins which participate in downstream pathway (IL4 and IL13). In addition, there are other target molecules and signaling pathways, which could be involved in allergic mechanisms, such as the high-affinity IgE Receptor (FC ϵ RI), which plays a key role in allergic reactions. This receptor is stimulated by IgE, triggering mast cells and basophils activation, and the consequent release of inflammatory mediators. In human mast cells and basophils, FC ϵ RI consist of a heterotetramer composed by three subunits: FC ϵ RI α , the ligand-binding subunit which is encoded by *FCER1A* gene; FC ϵ RI β , a signal-augmenting subunit encoded by *MS4A2*; and FC ϵ RI γ , a signal-transducing subunit that is presented like a dimer and it is encoded by *FCER1G* (Kinet, 1999; Potaczek and Kabesch, 2012). Elevated levels of IgE have been detected in atopic conditions like allergic rhinitis, asthma, atopic dermatitis, anaphylaxis (Platts-Mills, 2001; Wallace et al., 2008) thus making FC ϵ RI a plausible target molecule in the study of the mechanisms involved in the development and in the clinical presentation of allergy.

It could be hypothesized that variations related to expression and/or function in genes of the vitamin D signaling pathways or FC ϵ RI might modify the risk of developing rhinitis or DHRs, and/or the presentation of clinical manifestations of these reactions. As a matter of fact, several studies demonstrated an association between different allergic diseases, including DHRs, and polymorphisms in these genes (Poon et al., 2004; Raby et al., 2004; Donfack et al., 2005; Bossé et al., 2009; Saadi et al., 2009; Pillai et al., 2011; Micheal et al., 2013; Berenguer et al., 2014; Amo et al., 2016a; Narozna et al., 2016). Several studies addressed the putative impact of exonic and intronic SNPs within the above-mentioned genes and the risk of allergic diseases and/or DHR (Wjst, 2005; Wjst et al., 2006; Battle et al., 2007; Arshad et al., 2008; Sadeghnejad et al., 2008; Weidinger et al., 2008; Black et al., 2009; Bossé et al., 2009; Ferreira et al., 2009; Knutsen et al., 2010; Li et al., 2010, 2012, 2014, 2016; Michel et al., 2010; Moffatt et al., 2010; Cooper et al., 2011; Joubert et al., 2011; Liu et al., 2011; Lu et al., 2011; Park et al., 2011; Paternoster et al., 2011; Pillai et al., 2011; Burkhardt et al., 2012; Choi et al., 2012; Granada et al., 2012; Lasky-Su et al., 2012; Ramasamy et al., 2012; Robinson et al., 2012; Zhou et al., 2012; Anderson et al., 2013; Hur et al., 2013; Ismail et al., 2013; Movahedi et al., 2013; Potaczek et al.,

2013; Sharma et al., 2014; Yang et al., 2014; Kumar et al., 2015; Papadopoulou et al., 2015; Pino-Yanes et al., 2015; Tian et al., 2015; Amo et al., 2016a,b; Han et al., 2016; Karaca et al., 2016; Narozna et al., 2016; Overton et al., 2016; Ådjers et al., 2017; Ashley et al., 2017; Park and Tantisira, 2017; Sun et al., 2017; Xu et al., 2017; Zhang et al., 2017; Zhao et al., 2017). However, there is little information about SNPs located in the promoters of these genes, which might have functional consequences.

In an attempt to identifying genetic susceptibility factors associated with allergy and/or DHRs, that may provide novel information to gain a better understanding of these pathologies, we carried out an exhaustive analysis of genetic variations situated in the promoter region of the mentioned genes by using Next Generation Sequencing (NGS) in patients with allergic rhinitis plus asthma (AR+A), BLs hypersensitivity, selective NSAIDs hypersensitivity (SH) and cross-reactions to NSAIDs (CR), as well as in healthy control individuals. The genes included in the study were *FCER1A*, *MS4A2*, *FCER1G*, *VDR*, *GC*, *CYP2R1*, *CYP27A1*, *CYP27B1*, *CYP24A1*, *RXRA*, *RXRB*, *RXRG*, *IL4*, *IL4R*, *IL13*, and *IL13RA1*. In addition, we also analyzed the interaction of genetic and non-genetic factors, such as age, gender, and antecedents of atopy, in the risk of developing these diseases.

PATIENTS AND METHODS

Study Population

A total cohort of 2,476 individuals participated in this study. All were Caucasian Spanish individuals. These included 406 healthy controls without antecedents of atopy or adverse drug reactions, 528 patients with AR+A, 561 individuals with BLs hypersensitivity, 668 patients with NSAIDs cross-reactivity (CR), and 313 selective hypersensitivity patients (SH) which were single-NSAIDs responders. Written consent for participation was obtained for all participants. Patients were recruited at Hospitals participating in the study. All the patients who were invited to participate in the study agreed to do so. Control individuals were selected among students and staff in the University and Hospitals participating in the study. Characteristics of the study groups are summarized in **Table 1**. The diagnosis was carried out as described elsewhere (García-Martín et al., 2007; Doña et al., 2011; Amo et al., 2016a; Lacombe-Barrios, 2018). The protocol for this study was in accordance with the Declaration of Helsinki and its subsequent revisions and was approved by the respective Ethics Committees of the participating Hospitals.

To get a further analysis of the sample, we put together some of the groups of patients which share a specific characteristic. Thus, we defined three new groups of study: “DHR group,” were we included all the patients with DHR: namely, patients with hypersensitivity to BLs and NSAIDs (both, CR and SH); “DHR-IgE Group,” which comprises selective hypersensitivity to BLs and SH; and “IgE Mediated Group,” where we included all the IgE-mediated reactions (AR+A, BLs and SH).

Identification of Novel Variants Using NGS

A subset of participants were selected for this phase. A total cohort of 175 individuals participated in this NGS analysis. These included 22 healthy controls without antecedents of

atopy or adverse drug reactions, 22 patients with AR+A, 43 individuals with BLs hypersensitivity, 41 patients with NSAIDs cross-reactivity (CR), and 46 selective hypersensitivity patients (SH) which were single-NSAIDs responders. Characteristics of the participants are summarized in **Table S1**. Genomic DNA was obtained from leukocytes and purified according to standard procedures. DNA samples were analyzed by NGS after specific enrichment based on the Haloplex design. Details of the areas sequenced are shown in **Table S2**. DNA was digested with restriction enzymes specific for this design (Haloplex, Agilent, Santa Clara, CA, USA), followed by hybridization with specific probes, DNA circularization and selection of the target areas, according to the protocol supplied. Sequencing was carried out in a MiSeq sequencer (Illumina, San Diego, CA, USA) using the pair end format. The coverage was always higher than that recommended by the manufacturer (23.7 Mb per sample). All variants identified had at least a 50X coverage and more than 95% of these had more than 100x coverage. The sequencing results were analyzed by using the application SureCall 4.0 (Agilent, Santa Clara, CA, USA), adapted to the analysis of enriched Haloplex sequences, and MiseqReporter V04 (Illumina, San Diego, CA). Sequence revision against human genome was carried out by using the Integrative Genomes Viewer (Broad Institute, Cambridge, MA, USA).

Combined Dataset Analyses

All patients and controls participated in this phase. Analyses were carried out by using TaqMan genotyping focused on the SNVs raised after multiple comparison analyses of the NGS phase (see the results section for further details). The SNPs were analyzed in triplicate, by using SNP TaqMan assays (Life Technologies S.A., Alcobendas, Madrid, Spain), and following the conditions specified by the manufacturer. Assay details are as follows: *FCER1G*-rs36233990, Custom TaqMan® Assay; *FCER1G*-rs2070901, (C__15867981_20); and *GC*-rs3733359, (C__25652813_40).

Statistical Analysis

The R package SNPasoc (Gonzalez et al., 2014) was used to calculate allele and genotypic frequencies, to determine the Hardy-Weinberg equilibrium using exact test (Wigginton, 2005) and to analyse differences between groups (González et al., 2007). The comparison between groups was performed with the Fisher's Exact Test (FET) and Likelihood Ratio Test (LRT) with an initial crude analysis followed by an adjusted analysis including gender as the categorical covariate when it was possible. False Discovery Rate (FDR) correction was used for the multiple comparison adjustments (Benjamini et al., 2001). The results were considered statistically significant when *P*-values were under 0.05. The association between SNPs and traits was estimated by odds ratio (OR) with a 95% confidence interval (CI) or by Relative Risk (RR) when the variation was not found in the control group. The Relative Risk was calculated by using EpiBasic, a tool for statistical analysis of tabular information, performing a stratified analysis, using the inverse variance ($1/SE^2$) as weigh. This tool was developed as a companion to a Danish textbook on epidemiology (Juhl, 2012), and the

TABLE 1 | Characteristics of the participants.

	Healthy controls (<i>n</i> = 406)	Patients with AR+A (<i>n</i> = 528)	Patients with BLs hypersensitivity (<i>n</i> = 561)	Patients with NSAIDs cross-reactions (CR) (<i>n</i> = 668)	Patients with NSAIDs selective hypersensitivity (SH) (<i>n</i> = 313)
Women. <i>n</i> (%)	253 (62.3%)	292 (55.3%)	318 (56.7%)	390 (58.4%)	204 (65.2%)
Age + sd (range)	22.1 ± 4.7 (20–58)	32.4 ± 14.2 (14–79)	46.7 ± 14.5 (4–91)	41.8 ± 15.3 (5–92)	45.5 ± 16.0 (5–82)
Antecedents of atopy	0	100%	23.3%	20.8%	23.3%

spreadsheet could be download from the following link (Juul and Frydenberg, 2016): <http://ph.medarbejdere.au.dk/undervisning-og-uddannelse/software/>.

Association between each SNP and each clinical phenotype was assessed by using binary logistic regression. Then, predictive Models based on Multinomial Logistic Regression (MLR) (Agresti, 2003) were performed for SNPs showing association in the binary regression analyses by using SPSS (IBM SPSS Statistics for Windows, Version 22.0). The *p*-values associated with every SNP were calculated using the Chi-Square test. Each model has associated several pseudo-*R*² coefficients as indicators of the strength of the association between the response and the predictor variables. Cox and Snell is based on the loglikelihood for the model compared to the log likelihood for a baseline model and it has a theoretical maximum value of <1, even for a “perfect” model (Cox and Snell, 1989) and Nagelkerke is an adjusted version of the Cox & Snell *R*-square that adjusts the scale of the statistic to cover the full range from 0 to 1 (Nagelkerke, 1991). McFadden is another version, based on the log-likelihood kernels for the intercept-only model and the full estimated model. This is the pseudo-*R*² coefficient most frequently used and the correlation between variables is good when the values are comprised between 0.2 and 0.4, and better up to 0.4 (McFadden, 1974, 1977). The first model includes all the groups separately, that is, AR+A, BLs, CR and SH. Model 2 considered two groups of patients: AR+A and a group combining all DRH patients. Model 3 considered three groups of patients: AR+A, patients with IgE-Mediated DHR, and patients with DHR not related to IgE (that is, CR patients). For all models the control group was always the reference group. Coefficients were calculated by dropping samples with missing data in explanatory variables, which have been selected using stepwise regression method. The statistical power was calculated from variant allele frequencies with a genetic model analysing the frequency for carriers of the disease gene with a RR value = 2 (*p* = 0.05) for the genetic associations identified in the combined dataset model as described elsewhere (Pértegas Díaz and Pita Fernández, 2003). These values are shown in Table S3. The functional impact of the gene variants was analyzed by using TRANSFAC (Matys et al., 2003, 2006).

RESULTS

Identification of Novel Variants Using NGS

In this phase we identified 148 variations situated in the promoter region of genes related with vitamin D and FCεRI genes.

The information about the variations, their frequency in the whole sample and the Hardy-Weinberg equilibrium values is summarized in Table 2. It is to be noted that 84 out of the 148 (56.7%) of the SNPs identified in this study were found in cases only and not in control individuals.

Among the 148 gene variations identified, 43 were novel. Within the 105 already described SNPs, 25 have not been described or studied earlier in European individuals, although they show marginal MAF in our study (only three SNPs show MAF above 0.010). Regarding known SNPs, the frequencies are concordant with the results previously described in the 1,000 Genomes public database (<http://grch37.ensembl.org/index.html>) for individuals with European descent for all the variations except for the rs4020369 SNP in the GC gene, where the described frequency for Europeans is equal to 0, but in our population it shows a MAF close to 0.040, that is in agreement with the global frequency described in 1,000 Genomes for overall individuals.

One hundred and three out of the 148 variants identified were at Hardy-Weinberg equilibrium (HWE) in the overall study population (see Table S4), which is to be expected given the high number of SNPs analyzed and the limited sample size in the NGS analyses. Within the 45 variants that were not in HWE, only 7 showed a MAF > 0.050 in agreement with frequencies described in literature. We carried out binary logistic regression analyses excluding those variants with MAF < 0.02 (see Table 2). Results of the regression analyses are shown in Table S5. Among the 44 variants we selected those with adjusted *P* ≤ 0.10 for multinomial analyses. Therefore, 25 variants were included in the multinomial analysis as well as gender, antecedents of atopy and clinical group (Allergic rhinitis + Asthma; BLs, CR and SH). It is to be noted that some SNPs with a high significance after logistic binary regression analyses (See Table S5), such as rs1467664 (RXRG), rs3733359 (GC), rs2070874 (IL4), rs4303288, and rs4307775 (VDR) and rs2259735 (CYP24A1), were not significant after multinomial analysis. The statistically significant variables raised after this analysis were three SNPs (FCER1G rs36233990, FCER1G rs2070901, and GC rs3733359), as well as antecedents of atopy.

Combined Dataset Analyses

The three SNPs mentioned above were analyzed in the whole study group. The FCER1G rs36233990 SNP was monomorphic in the control group, whereas heterozygous subjects were identified in all subgroups of patients and homozygous individuals were identified in the AR+A and CR groups. Statistically significant

TABLE 2 | SNPs with MAF ≥ 0.02 observed in the NGS study.

Gene	dbSNP	Chromosomal Location	Alleles	MAF	HWE	MAF 1,000 Genomes ALL individuals	MAF 1,000 Genomes European individuals	MAF genomAD ALL individuals	MAF genomAD European individuals
FCER1A	rs2427837	1:159258545	G/A	0.213 A	0.501	0.151 A	0.304 A	0.217 A	0.289 A
FCER1A	rs61828219	1:159258641	C/A	0.141 A	0.752	0.085 A	0.160 A	0.129 A	0.175 A
FCER1A	rs12135235	1:159259029	T/G	0.043 G	1.000	0.033 G	0.033 G	0.034 G	0.037 G
FCER1G	rs36233990	1:161184658	C/T	0.020 T	0.134	0.011 T	0.006 T	0.010 T	0.013 T
FCER1G	None	1:161184792	GTCTCAAAAA/G	0.023 G	1.000	–	–	–	–
FCER1G	None	1:161184793	TCTCAAAAA/T	0.026 T	1.000	–	–	–	–
FCER1G	None	1:161184794	CTCAAAAAA/C	0.052 C	1.000	–	–	–	–
FCER1G	None	1:161184795	TCAAAAA/T	0.121 T	0.474	–	–	–	–
FCER1G	None	1:161184796	C/A	0.020 A	0.000	–	–	–	–
FCER1G	rs11587213	1:161184875	A/G	0.171 G	0.284	0.146 G	0.179 G	0.142 G	0.171 G
FCER1G	rs41270847	1:161184976	G/A	0.023 A	1.000	0.007 A	0.020 A	0.010 A	0.015 A
FCER1G	rs2070901	1:161185058	G/T	0.276 T	0.328	0.391 T	0.269 T	0.314 T	0.248 T
RXRG	rs3753897	1:165414511	C/A	0.152 A	0.082	0.215 A	0.195 A	0.211 A	0.193 A
RXRG	rs1467664	1:165414933	T/C	0.178 C	0.073	0.203 C	0.142 C	0.181 C	0.138 C
GC	rs3733359	4:72649774	G/A	0.069 A	0.894	0.206 A	0.055 A	0.122 A	0.060 A
GC	rs76781122	4:72669661	C/A	0.034 A	1.000	0.013 A	0.034 A	0.019 A	0.028 A
GC	rs6843222	4:72669944	C/T	0.029 T	1.000	0.004 T	0.016 T	0.007 T	0.011 T
GC	rs35096193	4:72670093	C/A	0.236 A	0.059	0.165 A	0.284 A	0.203 A	0.272 A
GC	rs1565572	4:72670191	A/C	0.210 C	0.499	0.432 C	0.196 C	0.363 C	0.192 C
GC	rs4020369	4:72670448	G/A	0.037 A	1.000	0.033 A	0	0.037 A	0.008 A
IL13	None	5:131992098	G/C	0.023 C	0.000	–	–	–	–
IL13	rs1800925	5:131992809	C/T	0.223 T	0.001	0.255 T	0.178 T	0.270 T	0.226 T
IL13	rs2066960	5:131994435	C/A	0.082 A	1.000	0.199 A	0.115 A	0.176 A	0.125 A
IL13	rs1295687	5:131994462	G/C	0.072 C	0.045	–	–	0.131 C	0.064 C
IL4	rs2243250	5:132009154	C/T	0.141 T	0.752	0.470 T	0.168 T	0.371 T	0.176 T
IL4	rs2070874	5:132009710	C/T	0.132 T	0.507	0.401 T	0.168 T	0.279 T	0.148 T
RXRB	rs76929655	6:33169182	T/C	0.020 C	1.000	0.002 C	0.005 C	0.007 C	0.011 C
CYP2R1	rs12794714	11:14913575	G/A	0.445 A	0.647	0.349 A	0.447 A	0.406 A	0.433 A
MS4A2	rs573790	11:59855385	C/T	0.382 T	1.000	0.441 T	0.356 T	0.460 T	0.411 T
MS4A2	rs574700	11:59855483	C/T	0.032 T	1.000	0.124 T	0.039 T	0.084 T	0.018 T
MS4A2	rs1441585	11:59855711	T/C	0.032 C	1.000	0.112 C	0.038 C	–	–
MS4A2	rs1441586	11:59856028	T/C	0.414 C	0.755	0.460 C	0.456 C	0.425 C	0.418 C
VDR	rs117397914	12:48276613	A/G	0.031 G	0.150	0.009 G	0.018 G	0.011 G	0.016 G
VDR	rs11168293	12:48293716	G/T	0.284 T	0.712	0.166 T	0.321 T	0.282 T	0.355 T
VDR	rs4303288	12:48336619	A/C	0.467 C	0.035	0.404 A	0.397 A	0.402 A	0.389 A
VDR	rs4307775	12:48336623	C/G	0.139 G	0.000	0.171 G	0.209 G	0.185 G	0.249 G
IL4R	rs12927172	16:27325021	G/A	0.424 A	0.013	0.405 A	0.372 A	0.353 A	0.367 A
IL4R	rs12927543	16:27325023	A/G	0.109 G	0.125	0.082 G	0.088 G	0.066 G	0.080 G
CYP24A1	rs35873579	20:52788190	G/A	0.023 A	1.000	0.001 A	0.004 A	0.002 A	0.003 A
CYP24A1	rs36106327	20:52788294	C/A	0.026 A	0.100	0.007 A	0.020 A	0.010 A	0.016 A
CYP24A1	rs2259735	20:52788314	T/C	0.455 C	0.029	0.566 C	0.420 C	0.515 C	0.408 C
CYP24A1	rs2762943	20:52790786	G/T	0.095 T	0.652	0.034 T	0.085 T	0.037 T	0.081 T
CYP24A1	rs2585427	20:52790976	G/C	0.376 C	0.873	0.444 C	0.390 C	0.447 C	0.373 C
IL13RA1	rs6603441	X:117861321	T/G	0.356 G	0.000	0.452 G	0.315 G	0.407 G	0.322 G

–, not described; MAF, Minor Allele Frequency.

differences were identified for this SNP in all subgroups of patients (Table 3), although after FDR correction differences remained significant for the AR+A and BLs groups. The *FCER1G*

rs2070901 SNP had a marginal trend toward higher frequency of the variant allele among CR patients, which was not statistically significant after FDR analysis. These two *FCER1G* SNPs are not

TABLE 3 | Combined dataset analyses.

SNP ID	Genotype	Control frequency	Rhinits+Asthma (AR+A) vs. controls comparison values (N = 528 vs. 406)			BIs vs. Controls comparison values (N = 561 vs. 406)			CR vs. Controls comparison values (N = 668 vs. 406)			SH vs. Controls comparison values (N = 313 vs. 406)		
			AR+A frequency	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)	BLs frequency	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)	CR Frequency	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)	SH Frequency	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)
FCER1G-rs36233990	C/C	1	0.953	1.00	0.000 (0.000)	0.991	1.00	0.025 (0.041)	0.990	1.00	0.043 (0.071)	0.993	1.00	0.077 (0.231)
	C/T	0	0.045	1.76 (1.66–1.86)		0.009	1.67 (1.53–1.76)		0.008	*		0.007	*	
	T/T	0	0.002	*		–	–		0.002	*		0	–	
FCER1G-rs2070901	G/G	0.482	0.531	1.00	0.065 (0.097)	0.545	1.00	0.073 (0.073)	0.524	1.00	0.047 (0.071)	0.521	1.00	0.328 (0.492)
	G/T	0.453	0.381	0.75 (0.56–0.99)		0.380	0.73 (0.55–0.96)		0.382	0.77 (0.59–1.01)		0.399	0.81 (0.59–1.11)	
	T/T	0.065	0.088	1.18 (0.69–2.00)		0.075	1.00 (0.58–1.72)		0.094	1.32 (0.80–2.20)		0.080	1.13 (0.62–2.06)	
GC-rs3733359	G/G	0.874	0.892	1.00	0.646 (0.646)	0.910	1.00	0.027 (0.041)	0.901	1.00	0.212 (0.212)	0.886	1.00	0.632 (0.632)
	G/A	0.118	0.104	0.88 (0.57–1.36)		0.090	0.73 (0.47–1.13)		0.098	0.79 (0.52–1.22)		0.110	0.92 (0.57–1.49)	
	A/A	0.008	0.004	0.50 (0.08–3.03)		0	0.00 (0.00–)		0.002	0.22 (0.02–2.10)		0.003	0.37 (0.04–3.63)	

Comparison between groups of patients and controls. (1) Relative Risk was calculated when the reference group (control) shows only non-mutated frequency. (2) OR and P-value were adjusted for gender. (*) OR adjusted by gender cannot be calculated due to lack of cases.

at linkage disequilibrium, being the D' value = 0.3415 and the r value = 0.0504. The minor allele frequency for the GC rs3733359 SNP was lower in the BLs group as compared to that of healthy individuals. This comparison remained statistically significant after FDR analysis.

The patients were grouped according to the underlying mechanism of the reaction. The first group was composed of all patients with drug hypersensitivity (DHR group), that is, (BLs + CR + SH). The second group of patients was composed of all patients with IgE-mediated drug hypersensitivity (D-IgE) (BLs + SH). The third group of patients was composed of all patients with IgE-mediated reactions (BLs + SH + AR+A), which were compared vs. healthy controls (Table 4). The two FCER1G SNPs displayed statistically significant differences in DHR and IgE-mediated reactions as compared to control individuals, although the only difference that remained significant after FDR correction was that of the SNP FCER1G rs2070901 in patients with IgE-mediated diseases (Table 4).

Classification Models

We built models including the three SNPs with significant associations in the combined dataset analyses phase, as well as the antecedents of atopy and gender, and the pseudo R-square values for each model (those which provide the best classification of each patient in its correct group) are shown in Table 5. All the models selected the FCER1G rs36233990 as a good variable. The variable “Antecedents of atopy” was also selected, although this was expected because none of the control individuals had antecedents of atopy. Model 1 was made by including the three SNPs with a significant P-value and all the clinical groups separately, as compared to control subjects. Model 2 included all patients with DHR compared to control individuals, and model 3 included all patients with IgE-mediated diseases vs. control individuals. The classifications per group are shown in Table 6. It is to be noted the high percentage of correct allocations using the three SNPs only (that is, without considering antecedents of atopy and gender). For comparison, we show in Table 6 the results of the same models including covariables such as antecedents of atopy, age and gender. The fact that antecedents of atopy predicted 100% of AR+A patients has little value because control individuals had no antecedents. Age and gender, however, are not good predictors either (Table 6). Therefore, these covariables did not improve the predictive capacity of the models based on the SNPs.

DISCUSSION

Genetic variation is a major cause of interindividual differences in the susceptibility to a number of disorders. In this regard, a huge number of genetic association studies related to allergic disorders and drug hypersensitivity events have been carried out. Most of these studies have a case-control design interrogating only a few polymorphisms, typically, a few SNPs located within the coding region. The use of NGS techniques allows for a complete coverage of large areas thus revealing novel SNPs or analysing SNPs that are not included in most studies. In a previous NGS study in the promoter area of the genes encoding the COX-1 and COX-2 enzymes, we identified several novel SNPs. More

TABLE 4 | Combined dataset analyses.

SNP ID	Genotype	Control Freq.	DHR vs. Control comparison values (N = 1,542 vs. 406)			DHR-IgE- vs. Control comparison values (N = 874 vs. 406)			All IgE-Mediated vs. Control comparison values (N = 1402 vs. 406)		
			Drugs (BLs, CR, SH) Freq	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)	Drugs IgE (BLs, SH) Freq.	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)	IgE Mediated (AR+A, SH, BLs) Freq.	OR (95% CI) with covariates (1) (2)	P _{cov} (FDR) (2)
FCER1G-rs36233990	C/C	1	0.991	1.00	0.047 (0.071)	0.992	1.00	0.025 (0.075)	0.977	1.00	0.001 (0.003)
	C/T	0	0.008	1.26 (1.23–1.29)		0.008	1.43 (1.38–1.49)		0.022	1.28 (1.25–1.31)	
	T/T	0	0.001	*		0	–		0.001	*	
FCER1G-rs2070901	G/G	0.482	0.531	1.00	0.044 (0.071)	0.536	1.00	0.089 (0.094)	0.534	1.00	0.047 (0.071)
	G/T	0.453	0.385	0.77 (0.61–0.97)		0.387	0.76 (0.59–0.98)		0.385	0.76 (0.60–0.96)	
GC-rs3733359	T/T	0.065	0.084	1.17 (0.74–1.87)		0.077	1.06 (0.64–1.74)		0.081	1.11 (0.69–1.77)	
	G/G	0.874	0.901	1.00	0.076 (0.076)	0.902	1.00	0.094 (0.094)	0.898	1.00	0.181 (0.181)
	G/A	0.118	0.097	0.80 (0.55–1.15)		0.097	0.80 (0.54–1.18)		0.100	0.83 (0.57–1.20)	
	A/A	0.008	0.001	0.17 (0.03–1.00)		0.001	0.14 (0.01–1.32)		0.002	0.26 (0.05–1.30)	

Comparisons depending on the mechanism. (1) Relative Risk was calculated when the reference group (control) shows only non-mutated frequency. (2) OR and P-value were adjusted for gender. (*) OR adjusted by gender cannot be calculated due to lack of cases.

TABLE 5 | Statistically significant variables in each model and goodness-of-fit measures.

Variables	Model 1: Single groups model (<i>p</i> -value)*	Model 2: DHR Model (BLs, CR, SH)/Rhinitis+ asthma/Control) (<i>p</i> -value)*	Model 3: IgE Mediated Drugs Model (BLs, SH)/Rhinitis+ asthma/CR/Control) (<i>p</i> -value)*
<i>FCER1G</i> rs36233990	0.001	0.000	0.000
<i>FCER1G</i> rs2070901	0.520	0.169	0.337
<i>GC</i> rs3733359	0.810	0.517	0.766
Antecedents of atopy	0.030	0.000	0.000
Gender	0.000	0.106	0.137
Pseudo R-square			
Cox and Snell	0.522	0.520	0.521
Nagelkerke	0.546	0.606	0.557
McFadden	0.237	0.376	0.268

*The overall effectiveness of the model was assessed by using the Chi-square statistic.

TABLE 6 | Prediction models.

Observed (clinical classification)	Predicted (Group allocation according to the model) SNPs only (%) correct	Predicted (Group allocation according to the model) SNPs + Antecedents of atopy + gender+ age (%) correct
AR+A (<i>n</i> = 481)	4.1	98.2
BLs (<i>n</i> = 230)	49.0	7.0
CR (<i>n</i> = 553)	53.4	68.6
SH (<i>n</i> = 300)	0	12.3
DHR (<i>n</i> = 1083)	99.0	81.8
DHR IgE-Mediated (<i>n</i> = 530)	99.2	49.6
All IgE-Mediated (<i>n</i> = 1,326)	99.9	79.6

than 70 SNPs modified transcription factor binding sites, either by disrupting existing sequences or by creating new binding sites (Agundez et al., 2014).

The present study is aimed to analyse the promoter areas of 16 genes related to allergic diseases and drug hypersensitivity reactions (DHRs). We have focused on the promoter gene region due to its crucial role in transcriptional activity and expression of the gene, as it has been observed for the SNPs located in the promoter of *FCERIA* (Potaczek et al., 2009), or *IL13* (Cameron et al., 2006; Kiesler et al., 2009; Li et al., 2014). The rationale for the selection of the 16 genes included in this study is based on putative mechanisms involved this type of reactions. *FCERI* plays an essential role in IgE-mediated mechanisms and variants in *FCERI* genes have been previously described as genetic factors related to asthma (Cui et al., 2003; Kim et al., 2006; Palikhe et al., 2008; Joubert et al., 2011; Ramphul et al., 2014; Yang et al., 2014), allergy (Hasegawa et al., 2003; de Guia et al., 2015; Liao et al.,

2015; Amo et al., 2016a,b), and food sensitization (Liu et al., 2011; Hong and Wang, 2012). It has been also described that some variants in genes involved in the vitamin D pathway are related with asthma (Poon et al., 2004; Raby et al., 2004; Wjst, 2005; Wjst et al., 2006; Bossé et al., 2009; Saadi et al., 2009; Li et al., 2011; Pillai et al., 2011; Maalmi et al., 2013; Leung et al., 2015; Hutchinson et al., 2017), especially variations in genes regulated by vitamin D, such as *IL4* and its receptor (Burchard et al., 1999; Donfack et al., 2005; Ober and Hoffjan, 2006; Battle et al., 2007; Michel et al., 2010; Baye et al., 2011; Hesselmars et al., 2012; Liu et al., 2012; Micheal et al., 2013; Nie et al., 2013; Zhu et al., 2013; Al-Muhsen et al., 2014; Berenguer et al., 2014; Klaassen et al., 2015; Zhang et al., 2015; Narozna et al., 2016) and *IL13* (Black et al., 2009; Bottema et al., 2010; Palikhe et al., 2010; Cui et al., 2012; Accordini et al., 2016; Xu et al., 2017), which are also related to IgE (Marsh et al., 1994; Kabisch et al., 2006). According to previous research, the mechanisms involved in cross-reactions and selective ones are different, (Doña et al., 2012; Ayuso et al., 2013; Torres et al., 2014; Nissen et al., 2015; Amo et al., 2016a) and previously published results show that some variations, either related to *FCERI* or to vitamin D, are strongly associated with IgE-mediated pathologies, like rs12135235 in *FCERIA*, rs144205117 in *CYP2R1*, rs1467664 in *RXRG* or rs4303288 in *VDR*. Association between the rs2070874 in *IL4* and atopy and hypersensitivity, has been described in previously published works (Burchard et al., 1999; Donfack et al., 2005; Kabisch et al., 2006; Ober and Hoffjan, 2006; Kim et al., 2010; Madore and Laprise, 2010; Baye et al., 2011; Lu et al., 2011; Liu et al., 2012; Andiappan et al., 2013; Hsu et al., 2013; Micheal et al., 2013; Movahedi et al., 2013; Zhu et al., 2013; Berenguer et al., 2014; Caniatti et al., 2014; Li et al., 2014; de Guia et al., 2015; Klaassen et al., 2015; Zhang et al., 2015; Hua et al., 2016; Narozna et al., 2016).

Although binary logistic regression analyses pointed to six SNPs corresponding to *RXRG*, *GC*, *IL4*, *VDR*, and *CYP24A1* (see the Results section), statistical significance for these SNPs was not supported after multinomial analysis, except for the *GC* SNP. By turn, two additional *FCER1G* SNPs, as well as the *GC* SNP, were statistically significant after multinomial analyses. It is important to note that the major findings obtained in the present study are novel, since only one of the three SNPs that remained after the multinomial analysis have been previously related with atopy or drug hypersensitivity. Among these, one of the *FCER1G* SNPs is novel and hence have not been studied before, and the other one has been related with food sensitization (Liu et al., 2011) and has been previously studied in patients with selective hypersensitivity to NSAIDs and allergic rhinitis without significant association (Amo et al., 2016a,b). After the NGS and combined dataset analyses phases, prediction models revealed that one of these SNPs, designated as *FCER1G* rs36233990 was correct in all models and it allowed an excellent prediction for patients with DHR, IgE-mediated DHR and all IgE-mediated diseases analyzed. It should be taken into consideration that the significant *p*-values observed for rs36233990 in case-control association analyses might be inflated because this SNP was not observed in controls. However, this is a commonly observed SNP in European populations,

which underscores the need for large control sets. This is a limitation in this study. The rs36233990 variation is located in a regulatory region where multiple transcription factor binding sites exist. The variant allele T triggers the appearance of E2F-3:Prx1 complex and GSKF (KLF4). On the other hand, the variant allele T leads the disappearance of a binding site for the transcription factor p300. Our own previous findings supported a role of *FCER* gene variations in patients with AR+A, but not in patients with SH (Park et al., 2011; Amo et al., 2016a) which are consistent with those raised in this study. The minor allele of rs2070901 in *FCERIG* triggers the disappearance of a transcription factor binding site for ELK-1: OC-2. The GC variation designated as rs3733359 is located in a splice region for transcripts 2 and 3 of GC, and in the 5' untranslated region for transcripts 1 and X1. This variant has been previously related to immune and other disorders (Jung et al., 2011; Wang et al., 2015; Xie et al., 2018). Our findings regarding the GC polymorphism support the hypothesis of a relevant role of vitamin D in allergy (Hall and Agrawal, 2017; Tian and Cheng, 2017).

In summary, our findings show that the analysis of the gene promoters is useful for the identification of genetic biomarkers of risk for DHRs and AR+A. Models using these gene variations allow a high degree of prediction, that is, correct group allocations (Table 6) based on these SNPs only. It should be kept in mind that the variant allele frequencies for these SNPs are relatively low, that is, the frequency of carriers of the risk variants is relatively low, specially for the most significant SNP *FCERIG* rs36233990 (<2% of patients). The frequencies of carriers for other two SNPs are 47 and 10.5% for *FCERIG* rs2070901 and *GC* rs3733359, respectively. Therefore, the presence of these gene variations cannot explain, by itself, the development of most cases of DHR. However, the SNPs raised in this study, point to mechanisms involved in

DHR and add novel information that can be used as a proof of mechanism.

AUTHOR CONTRIBUTIONS

EG-M and JA contributed conception and design of the study. JG-M, CC, JC-G, NB-L, GC, ID, MB, and MT recruited and characterized patients. MM performed the statistical analysis. GA wrote the first draft of the manuscript. MM, EG-M, and JA wrote sections of the manuscript. All authors contributed to manuscript critical revision with important intellectual contribution, read, and approved the submitted version.

FUNDING

This work was supported in part by Grants PI15/00303, PI15/00726, PI17/01593, PI18/00540 and ARADyAL RD16/0006/0001, RD16/0006/0004 and RD16/0006/0024 from Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spain, IB16170 and GR18145 from Junta de Extremadura, Spain. Financed in part with FEDER funds from the European Union.

ACKNOWLEDGMENTS

JC-G is a researcher from the Miguel Servet Program (Ref CP14/00034), and ID from the Juan Rodés Program (Ref JR15/0036), both from the Carlos III National Health Institute, Spanish Ministry of Economy and Competitiveness).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00582/full#supplementary-material>

REFERENCES

- Adjers, K., Luukkainen, A., Pekkanen, J., Hurme, M., and Huhtala, H., Renkonen, R., et al. (2017). Self-reported allergic rhinitis and/or allergic conjunctivitis associate with IL13 rs20541 polymorphism in finnish adult asthma patients. *Int. Arch. Allergy Immunol.* 172, 123–128. doi: 10.1159/000456009
- Accordini, S., Calcianno, L., Bombieri, C., Malerba, G., Belpinati, F., Lo Presti, A. R., et al. (2016). An interleukin 13 polymorphism is associated with symptom severity in adult subjects with ever asthma. *PLoS ONE* 11:e0151292. doi: 10.1371/journal.pone.0151292
- Agresti, A. (2003). Logit models for multinomial responses. *Categ Data Anal.* doi: 10.1002/0471249688.ch7
- Agúndez, J. A., Ayuso, P., Cornejo-García, J. A., Blanca, M., Torres, M. J., Doña, I., et al. (2012). The diamine oxidase gene is associated with hypersensitivity response to non-steroidal anti-inflammatory drugs. *PLoS ONE* 7:e47571. doi: 10.1371/journal.pone.0047571
- Agundez, J. A., Gonzalez-Alvarez, D. L., Vega-Rodriguez, M. A., Botello, E., and Garcia-Martin, E. (2014). Gene variants and haplotypes modifying transcription factor binding sites in the human cyclooxygenase 1 and 2 (PTGS1 and PTGS2) genes. *Curr. Drug Metab.* 15, 182–195. doi: 10.2174/138920021502140327180336
- Al-Muhsen, S., Vazquez-Tello, A., Alzaabi, A., Al-Hajjaj, M. S., Al-Jahdali, H. H., and Halwani, R. (2014). IL-4 receptor alpha single-nucleotide polymorphisms rs1805010 and rs1801275 are associated with increased risk of asthma in a Saudi Arabian population. *Ann. Thorac. Med.* 9, 81–86. doi: 10.4103/1817-1737.128849
- Amo, G., Cornejo-García, J. A., García-Menaya, J. M., Cordobes, C., Torres, M. J., Esguevillas, G., et al. (2016b). FCER1 and histamine metabolism gene variability in selective responders to NSAIDs. *Front. Pharmacol.* 7:353. doi: 10.3389/fphar.2016.00353
- Amo, G., García-Menaya, J., Campo, P., Cordobés, C., Plaza Serón, M. C., Ayuso, P., et al. (2016a). A Nonsynonymous FCER1B SNP is associated with risk of developing allergic rhinitis and with IgE levels. *Sci. Rep.* 6:19724. doi: 10.1038/srep19724
- Anderson, W. H., Koshy, B. T., Huang, L., Mosteller, M., Stinnett, S. W., Condeary, L. D., et al. (2013). Genetic analysis of asthma exacerbations. *Ann. Allergy Asthma Immunol.* 110, 416–422.e2. doi: 10.1016/j.anai.2013.04.002
- Andiappan, A. K., Nilsson, D., Halldén, C., Yun, W. D., Säll, T., Cardell, L. O., et al. (2013). Investigating highly replicated asthma genes as candidate genes for allergic rhinitis. *BMC Med. Genet.* 14:51. doi: 10.1186/1471-2350-14-51
- Antúnez, C., Martín, E., Cornejo-García, J. A., Blanca-Lopez, N. R., Pena, R., Mayorga, C., et al. (2006). Immediate hypersensitivity reactions to penicillins and other betalactams. *Curr. Pharm. Des.* 12, 3327–3333. doi: 10.2174/138161206778194042
- Arshad, S. H., Karmaus, W., Kurukulaaratchy, R., Sadeghnejad, A., Huebner, M., and Ewart, S. (2008). Polymorphisms in the interleukin 13 and GATA

- binding protein 3 genes and the development of eczema during childhood. *Br. J. Dermatol.* 158, 1315–1322. doi: 10.1111/j.1365-2133.2008.08565.x
- Ashley, S. E., Tan, H. T., Peters, R., Allen, K. J., Vuillermin, P., Dharmage, S. C., et al. (2017). Genetic variation at the Th2 immune gene IL13 is associated with IgE-mediated paediatric food allergy. *Clin. Exp. Allergy* 47, 1032–1037. doi: 10.1111/cea.12942
- Ayuso, P., Blanca-López, N., Doña, I., Torres, M. J., Guéant-Rodríguez, R. M., Canto, G., et al. (2013). Advanced phenotyping in hypersensitivity drug reactions to NSAIDs. *Clin. Exp. Allergy* 43, 1097–1109. doi: 10.1111/cea.12140
- Battle, N. C., Choudhry, S., Tsai, H. J., Eng, C., Kumar, G., Beckman, K. B., et al. (2007). Ethnicity-specific gene-gene interaction between IL-13 and IL-4 α among African Americans with asthma. *Am. J. Respir. Crit. Care Med.* 175, 881–887. doi: 10.1164/rccm.200607-992OC
- Baye, T. M., Butsch Kovacic, M., Biagini Myers, J. M., Martin, L. J., Lindsey, M., Patterson, T. L., et al. (2011). Differences in candidate gene association between European ancestry and African American asthmatic children. *PLoS ONE* 6:e16522. doi: 10.1371/journal.pone.0016522
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125, 279–284. doi: 10.1016/S0166-4328(01)00297-2
- Benson, A. A., Toh, J. A., Vernon, N., and Jariwala, S. P. (2012). The role of vitamin D in the immunopathogenesis of allergic skin diseases. *Allergy* 67, 296–301. doi: 10.1111/j.1398-9995.2011.02755.x
- Berenguer, A. G., Fernandes, A. T., Oliveira, S., Rodrigues, M., Ornelas, P., Romeira, D., et al. (2014). Genetic polymorphisms and asthma: findings from a case-control study in the Madeira island population. *Biol. Res.* 47:40. doi: 10.1186/0717-6287-47-40
- Black, P. N., and Scragg, R. (2005). Relationship between serum 25-hydroxyvitamin D and pulmonary function in the third national health and nutrition examination survey. *Chest* 128, 3792–3798. doi: 10.1378/chest.128.6.3792
- Black, S., Teixeira, A. S., Loh, A. X., Vinall, L., Holloway, J. W., Hardy, R., et al. (2009). Contribution of functional variation in the IL13 gene to allergy, hay fever and asthma in the NSHD longitudinal 1946 birth cohort. *Allergy* 64, 1172–1178. doi: 10.1111/j.1398-9995.2009.01988.x
- Blanca, M., Romano, A., Torres, M. J., Fernández, J., Mayorga, C., Rodriguez, J., et al. (2009). Update on the evaluation of hypersensitivity reactions to betalactams. *Allergy* 64, 183–193. doi: 10.1111/j.1398-9995.2008.01924.x
- Bossé, Y., Lemire, M., Poon, A. H., Daley, D., He, J. Q., Sandford, A., et al. (2009). Asthma and genes encoding components of the vitamin D pathway. *Respir. Res.* 10:98. doi: 10.1186/1465-9921-10-98
- Bottema, R. W., Nolte, I. M., Howard, T. D., Koppelman, G. H., Dubois, A. E., de Meer, G., et al. (2010). Interleukin 13 and interleukin 4 receptor- α polymorphisms in rhinitis and asthma. *Int. Arch. Allergy Immunol.* 153, 259–267. doi: 10.1159/000314366
- Bouvry, J. C., De Bruin, M. L., and Koopmanschap, M. A. (2015). Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. *Drug Saf.* 38, 437–453. doi: 10.1007/s40264-015-0281-0
- Brozek, J. L., Bousquet, J., Agache, I., Agarwal, A., Bachert, C., Bosnic-Anticevich, S., et al. (2017). Allergic rhinitis and its impact on asthma (ARIA) guidelines-2016 revision. *J. Allergy Clin. Immunol.* 140, 950–958. doi: 10.1016/j.jaci.2017.03.050
- Burchard, E. G., Silverman, E. K., Rosenwasser, L. J., Borish, L., Yandava, C., Pillari, A., et al. (1999). Association between a sequence variant in the IL-4 gene promoter and FEV(1) in asthma. *Am. J. Respir. Crit. Care Med.* 160, 919–922. doi: 10.1164/ajrccm.160.3.9812024
- Burkhardt, J., Kirsten, H., Wolfram, G., Quente, E., and Ahnert, P. (2012). Differential allelic expression of IL13 and CSF2 genes associated with asthma. *Genet. Mol. Biol.* 35, 567–574. doi: 10.1590/S1415-47572012005000055
- Camargo, C. A., Rifas-Shiman, S. L., Litonjua, A. A., Rich-Edwards, J. W., Weiss, S. T., Gold, D. R., et al. (2007). Maternal intake of vitamin D during pregnancy and risk of recurrent wheeze in children at 3 y of age. *Am. J. Clin. Nutr.* 85, 788–795. doi: 10.1093/ajcn/85.3.788
- Cameron, L., Webster, R. B., Strempel, J. M., Kiesler, P., Kabesch, M., Ramachandran, H., et al. (2006). Th2 cell-selective enhancement of human IL13 transcription by IL13-1112C>T, a polymorphism associated with allergic inflammation. *J. Immunol.* 177, 8633–8642. doi: 10.4049/jimmunol.177.12.8633
- Caniatti, M. C., Marchioro, A. A., Guilherme, A. L., and Tsuneto, L. T. (2014). Association of cytokines in individuals sensitive and insensitive to dust mites in a Brazilian population. *PLoS ONE* 9:e107921. doi: 10.1371/journal.pone.0107921
- Canto, M. G., Andreu, I., Fernandez, J., and Blanca, M. (2009). Selective immediate hypersensitivity reactions to NSAIDs. *Curr. Opin. Allergy Clin. Immunol.* 9, 293–297. doi: 10.1097/ACI.0b013e32832db943
- Cantorna, M. T., Snyder, L., Lin, Y. D., and Yang, L. (2015). Vitamin D and 1,25(OH) $_2$ D regulation of T cells. *Nutrients* 7, 3011–3021. doi: 10.3390/nu7043011
- Chen, C. J., Cheng, C. F., Lin, H. Y., Hung, S. P., Chen, W. C., and Lin, M. S. (2012). A comprehensive 4-year survey of adverse drug reactions using a network-based hospital system. *J. Clin. Pharm. Ther.* 37, 647–651. doi: 10.1111/j.1365-2710.2012.01359.x
- Choi, W. A., Kang, M. J., Kim, Y. J., Seo, J. H., Kim, H. Y., Kwon, J. W., et al. (2012). Gene-gene interactions between candidate gene polymorphisms are associated with total IgE levels in Korean children with asthma. *J. Asthma* 49, 243–252. doi: 10.3109/02770903.2012.660294
- Cooper, J. D., Smyth, D. J., Walker, N. M., Stevens, H., Burren, O. S., Wallace, C., et al. (2011). Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. *Diabetes* 60, 1624–1631. doi: 10.2337/db10-1656
- Cornejo-García, J. A., Blanca-López, N., Doña, I., Andreu, I., Agúndez, J. A., Carballo, M., et al. (2009). Hypersensitivity reactions to non-steroidal anti-inflammatory drugs. *Curr. Drug Metab.* 10, 971–980. doi: 10.2174/138920009790711841
- Cox, D. R., and Snell, E. J. (1989). *Analysis of Binary Data, 2nd Edn.* London: Taylor & Francis.
- Cui, L., Jia, J., Ma, C. F., Li, S. Y., Wang, Y. P., Guo, X. M., et al. (2012). IL-13 polymorphisms contribute to the risk of asthma: a meta-analysis. *Clin. Biochem.* 45, 285–288. doi: 10.1016/j.clinbiochem.2011.12.012
- Cui, T. P., Wang, L., Xie, J. G., Hu, L. H., and Wu, J. M. (2003). Study on the high-affinity IgE receptor beta gene polymorphism and its association with the susceptibility to allergic asthma in Han nationality of Hubei province. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 20, 307–310.
- de Guia, R. M., Echavez, M. D., Gaw, E. L., Gomez, M. R., Lopez, K. A., Mendoza, R. C., et al. (2015). Multifactor-dimensionality reduction reveals interaction of important gene variants involved in allergy. *Int. J. Immunogenet.* 42, 182–189. doi: 10.1111/iji.12200
- Doña, I., Barrionuevo, E., Blanca-Lopez, N., Torres, M. J., Fernandez, T. D., Mayorga, C., et al. (2014). Trends in hypersensitivity drug reactions: more drugs, more response patterns, more heterogeneity. *J. Invest. Allergol. Clin. Immunol.* 24, 143–153. doi: 10.1007/s10077-014-0153-1
- Doña, I., Blanca-López, N., Cornejo-García, J. A., Torres, M. J., Laguna, J. J., Fernández, J., et al. (2011). Characteristics of subjects experiencing hypersensitivity to non-steroidal anti-inflammatory drugs: patterns of response. *Clin. Exp. Allergy* 41, 86–95. doi: 10.1111/j.1365-2222.2010.03651.x
- Doña, I., Blanca-López, N., Torres, M. J., García-Campos, J., García-Núñez, I., Gómez, F., et al. (2012). Drug hypersensitivity reactions: response patterns, drug involved, and temporal variations in a large series of patients. *J. Invest. Allergol. Clin. Immunol.* 22, 363–371.
- Donack, J., Schneider, D. H., Tan, Z., Kurz, T., Dubchak, I., Frazer, K. A., et al. (2005). Variation in conserved non-coding sequences on chromosome 5q and susceptibility to asthma and atopy. *Respir. Res.* 6:145. doi: 10.1186/1465-9921-6-145
- Durham, S. R., Emminger, W., Kapp, A., de Monchy, J. G., Rak, S., Scadding, G. K., et al. (2012). SQ-standardized sublingual grass immunotherapy: confirmation of disease modification 2 years after 3 years of treatment in a randomized trial. *J. Allergy Clin. Immunol.* 129, 717–725.e5. doi: 10.1016/j.jaci.2011.12.973
- European Commission (2008). *Proposal for a Regulation Amending, as Regards Pharmacovigilance of Medicinal Products for Human Use.* Regulation (EC) No 726/2004. Impact assessment. Available online at: http://ec.europa.eu/health/files/pharmacovigilance/pharmapsacks_12_2008/pha
- Ferreira, M. A., Zhao, Z. Z., Thomsen, S. F., James, M., Evans, D. M., Postmus, P. E., et al. (2009). Association and interaction analyses of eight genes under asthma linkage peaks. *Allergy* 64, 1623–1628. doi: 10.1111/j.1398-9995.2009.02091.x
- García-Martín, E., García-Menaya, J., Sánchez, B., Martínez, C., Rosendo, R., and Agúndez, J. A. (2007). Polymorphisms of histamine-metabolizing enzymes and

- clinical manifestations of asthma and allergic rhinitis. *Clin. Exp. Allergy* 37, 1175–1182. doi: 10.1111/j.1365-2222.2007.02769.x
- Gomes, E., Cardoso, M. F., Praça, F., Gomes, L., Mariño, E., and Demoly, P. (2004). Self-reported drug allergy in a general adult Portuguese population. *Clin. Exp. Allergy* 34, 1597–1601. doi: 10.1111/j.1365-2222.2004.02070.x
- Gomes, E. R., and Demoly, P. (2005). Epidemiology of hypersensitivity drug reactions. *Curr. Opin. Allergy Clin. Immunol.* 5, 309–316. doi: 10.1097/01.all.0000173785.81024.33
- González, J. R., Armengol, L., Solé, X., Guinó, E., Mercader, J. M., Estivill, X., et al. (2007). SNPpass: an R package to perform whole genome association studies. *Bioinformatics* 23, 644–645. doi: 10.1093/bioinformatics/btm025
- Gonzalez, J. R., Lluís-Armengol, E. G., Sole, X., and Moreno, V. (2014). *SNPs-Based Whole Genome Association Studies*. Available online at: <https://CRAN.R-project.org/package=SNPassoc>
- Granada, M., Wilk, J.B., Tuzova, M., Strachan, D. P., Weidinger, S., Albrecht, E., et al. (2012). A genome wide Association study of plasma total IgE concentration in the framingham heart study. *J. Allergy Clin. Immunol.* 129, 840–845.e21. doi: 10.1016/j.jaci.2011.09.029
- Hall, S. C., and Agrawal, D. K. (2017). Vitamin D and bronchial asthma: an overview of the last five years. *Clin. Ther.* 39, 917–929. doi: 10.1016/j.clinthera.2017.04.002
- Han, J. C., Du, J., Zhang, Y. J., Qi, G. B., Li, H. B., Zhang, Y. J., et al. (2016). Vitamin D receptor polymorphisms may contribute to asthma risk. *J. Asthma* 53, 790–800. doi: 10.3109/02770903.2016.1158267
- Hasegawa, M., Nishiyama, C., Nishiyama, M., Akizawa, Y., Mitsuishi, K., Ito, T., et al. (2003). A novel–66T/C polymorphism in Fc epsilon RI alpha-chain promoter affecting the transcription activity: possible relationship to allergic diseases. *J. Immunol.* 171, 1927–1933. doi: 10.4049/jimmunol.171.4.1927
- Hesselmar, B., Enelund, A. C., Eriksson, B., Padyukov, L., Hanson, L. Å., and Aberg, N. (2012). The heterogeneity of asthma phenotypes in children and young adults. *J. Allergy* 2012:163089. doi: 10.1155/2012/163089
- Hong, X., and Wang, X. (2012). Early life precursors, epigenetics, and the development of food allergy. *Semin. Immunopathol.* 34, 655–669. doi: 10.1007/s00281-012-0323-y
- Hsu, J., Avila, P. C., Kern, R. C., Hayes, M. G., Schleimer, R. P., and Pinto, J. M. (2013). Genetics of chronic rhinosinusitis: state of the field and directions forward. *J. Allergy Clin. Immunol.* 131, 977–993.e5. doi: 10.1016/j.jaci.2013.01.028
- Hua, L., Zuo, X. B., Bao, Y. X., Liu, Q. H., Li, J. Y., Lv, J., et al. (2016). Four-locus gene interaction between IL13, IL4, FCER1B, and ADRB2 for asthma in Chinese Han children. *Pediatr. Pulmonol.* 51, 364–371. doi: 10.1002/ppul.23322
- Hur, G. Y., Ye, Y. M., Koh, D. H., Kim, S. H., and Park, H. S. (2013). IL-4 receptor α polymorphisms may be a susceptible factor for work-related respiratory symptoms in bakery workers. *Allergy Asthma Immunol. Res.* 5, 371–376. doi: 10.4168/aa.2013.5.6.371
- Hutchinson, K., Kerley, C.P., Faul, J., Greally, P., Coghlan, D., Louw, M., et al. (2017). Vitamin D receptor variants and uncontrolled asthma. *Eur. Ann. Allergy Clin. Immunol.* 50:108–116. doi: 10.23822/EurAnnACI.1764-1489.46
- Ismail, M. F., Elnady, H. G., and Fouda, E. M. (2013). Genetic variants in vitamin D pathway in Egyptian asthmatic children: a pilot study. *Hum. Immunol.* 74, 1659–1664. doi: 10.1016/j.humimm.2013.08.284
- Joubert, B. R., Reif, D. M., Edwards, S. W., Leiner, K. A., Hudgens, E. E., Eggeghy, P., et al. (2011). Evaluation of genetic susceptibility to childhood allergy and asthma in an African American urban population. *BMC Med. Genet.* 12:25. doi: 10.1186/1471-2350-12-25
- Jung, K. H., Kim, T. H., Sheen, D. H., Lim, M. K., Lee, S. K., Kim, J. Y., et al. (2011). Associations of vitamin D binding protein gene polymorphisms with the development of peripheral arthritis and uveitis in ankylosing spondylitis. *J. Rheumatol.* 38, 2224–2229. doi: 10.3899/jrheum.101244
- Juul, S. (2012). *Epidemiologi Og Evidens. 2nd Edn*. København: Munksgaard.
- Juul, S., and Frydenberg, M. (2016). EpiBasic. Available online at: <http://ph.medarbejdere.au.dk/undervisning-og-uddannelse/software/> (accessed February 06, 2019).
- Kabesch, M., Schedel, M., Carr, D., Woitsch, B., Fritzsche, C., Weiland, S. K., et al. (2006). IL-4/IL-13 pathway genetics strongly influence serum IgE levels and childhood asthma. *J. Allergy Clin. Immunol.* 117, 269–274. doi: 10.1016/j.jaci.2005.10.024
- Karaca, S., Civelek, E., Karaca, M., Sahiner, U. M., Ozgul, R. K., Kocabas, C. N., et al. (2016). Allergy-specific phenome-wide association study for immunogenes in turkish children. *Sci. Rep.* 6:33152. doi: 10.1038/srep33152
- Kiesler, P., Shakya, A., Tantin, D., and Vercelli, D. (2009). An allergy-associated polymorphism in a novel regulatory element enhances IL13 expression. *Hum. Mol. Genet.* 18, 4513–4520. doi: 10.1093/hmg/ddp411
- Kim, B. S., Park, S. M., Uhm, T. G., Kang, J. H., Park, J. S., Jang, A. S., et al. (2010). Effect of single nucleotide polymorphisms within the interleukin-4 promoter on aspirin intolerance in asthmatics and interleukin-4 promoter activity. *Pharmacogenet. Genomics* 20, 748–758. doi: 10.1097/FPC.0b013e3283402155
- Kim, S. H., Bae, J. S., Holloway, J. W., Lee, J. T., Suh, C. H., Nahm, D. H., et al. (2006). A polymorphism of MS4A2 (-109T > C) encoding the beta-chain of the high-affinity immunoglobulin E receptor (Fc epsilon R1beta) is associated with a susceptibility to aspirin-intolerant asthma. *Clin. Exp. Allergy* 36, 877–883. doi: 10.1111/j.1365-2222.2006.02443.x
- Kinet, J. P. (1999). The high-affinity IgE receptor (Fc epsilon RI): from physiology to pathology. *Annu. Rev. Immunol.* 17, 931–972. doi: 10.1146/annurev.immunol.17.1.931
- Klaassen, E. M., Penders, J., Jöbsis, Q., van de Kant, K. D., Thijs, C., Mommers, M., et al. (2015). An ADAM33 polymorphism associates with progression of preschool wheeze into childhood asthma: a prospective case-control study with replication in a birth cohort study. *PLoS ONE* 10:e0119349. doi: 10.1371/journal.pone.0119349
- Knutsen, A. P., Vijay, H. M., Kariuki, B., Santiago, L. A., Graff, R., Wofford, J. D., et al. (2010). Association of IL-4RA single nucleotide polymorphisms, HLA-DR and HLA-DQ in children with Alternaria-sensitive moderate-severe asthma. *Clin. Mol. Allergy* 8:5. doi: 10.1186/1476-7961-8-5
- Kowalski, M. L., Asero, R., Baybek, S., Blanca, M., Blanca-Lopez, N., Bochenek, G., et al. (2013). Classification and practical approach to the diagnosis and management of hypersensitivity to nonsteroidal anti-inflammatory drugs. *Allergy* 68, 1219–1232. doi: 10.1111/all.12260
- Kruse, R. L., and Vanijcharoenkarn, K. (2018). Drug repurposing to treat asthma and allergic disorders: progress and prospects. *Allergy* 73, 313–322. doi: 10.1111/all.13305
- Kumar, A., Das, S., Agrawal, A., Mukhopadhyay, I., and Ghosh, B. (2015). Genetic association of key Th1/Th2 pathway candidate genes, IRF2, IL6, IFNGR2, STAT4 and IL4RA, with atopic asthma in the Indian population. *J. Hum. Genet.* 60, 443–448. doi: 10.1038/jhg.2015.45
- Lacombe-Barrios, J., Gomez, F., Perez, N., et al. (2018). Accurate of the diagnosis of allergy reactions at the Emergency Department. *J. Invest. Allergol. Clin. Immunol.* doi: 10.18176/jiaci.0313. [Epub ahead of print].
- Lasky-Su, J., Lange, N., Brehm, J. M., Damask, A., Soto-Quiros, M., Avila, L., et al. (2012). Genome-wide association analysis of circulating vitamin D levels in children with asthma. *Hum. Genet.* 131, 1495–1505. doi: 10.1007/s00439-012-1185-z
- Leung, T. F., Wang, S. S., Tang, M. F., Kong, A. P., Sy, H. Y., Hon, K. L., et al. (2015). Childhood asthma and spirometric indices are associated with polymorphic markers of two vitamin D 25-hydroxylase genes. *Pediatr. Allergy Immunol.* 26, 375–382. doi: 10.1111/pai.12392
- Li, F., Jiang, L., Willis-Owen, S. A., Zhang, Y., and Gao, J. (2011). Vitamin D binding protein variants associate with asthma susceptibility in the Chinese han population. *BMC Med. Genet.* 12:103. doi: 10.1186/1471-2350-12-103
- Li, J., Lin, L. H., Wang, J., Peng, X., Dai, H. R., Xiao, H., et al. (2014). Interleukin-4 and interleukin-13 pathway genetics affect disease susceptibility, serum immunoglobulin E levels, and gene expression in asthma. *Ann. Allergy Asthma Immunol.* 113, 173–179.e1. doi: 10.1016/j.anai.2014.05.004
- Li, T., Ren, Z., Deng, Y., Wang, Y., and Zhou, H. (2016). Lack of association between RAD50-IL13 polymorphisms and pediatric asthma susceptibility in Northeastern Han Chinese. *J. Asthma* 53, 114–118. doi: 10.3109/02770903.2015.1067322
- Li, X., Ampleford, E. J., Howard, T. D., Moore, W. C., Torgerson, D. G., Li, H., et al. (2012). Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. *J. Allergy Clin. Immunol.* 130, 861–8.e7. doi: 10.1016/j.jaci.2012.04.041

- Li, X., Howard, T. D., Zheng, S. L., Haselkorn, T., Peters, S. P., Meyers, D. A., et al. (2010). Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J. Allergy Clin. Immunol.* 125, 328–335.e11. doi: 10.1016/j.jaci.2009.11.018
- Liao, E. C., Chang, C. Y., Hsieh, C. W., Yu, S. J., Yin, S. C., and Tsai, J. J. (2015). An exploratory pilot study of genetic marker for ige-mediated allergic diseases with expressions of fcepsilon1alpha and cepsilon. *Int. J. Mol. Sci.* 16, 9504–9519. doi: 10.3390/ijms16059504
- Linneberg, A. (2011). The increase in allergy and extended challenges. *Allergy* 66 (Suppl. 9), 1–3. doi: 10.1111/j.1398-9995.2011.02619.x
- Liu, S., Li, T., and Liu, J. (2012). Interleukin-4 rs2243250 polymorphism is associated with asthma among Caucasians and related to atopic asthma. *Cytokine* 59, 364–369. doi: 10.1016/j.cyto.2012.05.006
- Liu, X., Wang, G., Hong, X., Wang, D., Tsai, H. J., Zhang, S., et al. (2011). Gene-vitamin D interactions on food sensitization: a prospective birth cohort study. *Allergy* 66, 1442–1448. doi: 10.1111/j.1398-9995.2011.02681.x
- Lu, M. P., Chen, R. X., Wang, M. L., Zhu, X. J., Zhu, L. P., Yin, M., et al. (2011). Association Study on IL4, IL13 and IL4RA polymorphisms in mite-sensitized persistent allergic rhinitis in a chinese population. *PLoS ONE* 6:e27363. doi: 10.1371/journal.pone.0027363
- Maalmi, H., Sassi, F. H., Berraies, A., Ammar, J., Hamzaoui, K., and Hamzaoui, A. (2013). Association of vitamin D receptor gene polymorphisms with susceptibility to asthma in Tunisian children: a case control study. *Hum. Immunol.* 74, 234–240. doi: 10.1016/j.humimm.2012.11.005
- Madore, A. M., and Laprise, C. (2010). Immunological and genetic aspects of asthma and allergy. *J. Asthma Allergy* 3, 107–121. doi: 10.2147/JAA.S8970
- Marsh, D. G., Neely, J. D., Breazeale, D. R., Ghosh, B., Freidhoff, L. R., Ehrlich-Kautzky, E., et al. (1994). Linkage analysis of IL4 and other chromosome 5q31.1 markers and total serum immunoglobulin E concentrations. *Science* 264, 1152–1156. doi: 10.1126/science.8178175
- Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucl. Acids Res.* 31, 374–378. doi: 10.1093/nar/gkg108
- Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., et al. (2006). TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucl. Acids Res.* 34:D108–D110. doi: 10.1093/nar/gkj143
- McFadden, D. (1974). “Conditional logit: analysis of qualitative choice behavior,” in *Frontiers in Econometrics*, ed P. Zarembka (New York, NY: Academic Press), 105–142.
- McFadden, D. (1977). *Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments*. Cowles Foundation Discussion Papers 474, Cowles Foundation for Research in Economics, Yale University.
- Messaad, D., Sahla, H., Benahmed, S., Godard, P., Bousquet, J., and Demoly, P. (2004). Drug provocation tests in patients with a history suggesting an immediate drug hypersensitivity reaction. *Ann. Intern. Med.* 140, 1001–1006. doi: 10.7326/0003-4819-140-12-200406150-00009
- Micheal, S., Minhas, K., Ishaque, M., Ahmed, F., and Ahmed, A. (2013). IL-4 gene polymorphisms and their association with atopic asthma and allergic rhinitis in Pakistani patients. *J. Invest. Allergol. Clin. Immunol.* 23, 107–111.
- Michel, S., Liang, L., Depner, M., Klopp, N., Ruether, A., Kumar, A., et al. (2010). Unifying candidate gene and GWAS approaches in asthma. *PLoS ONE* 5:e13894. doi: 10.1371/journal.pone.0013894
- Moffatt, M. F., Gut, I. G., Demenais, F., Strachan, D. P., Bouzigon, E., Heath, S., et al. (2010). A large-scale, consortium-based genome-wide association study of asthma. *N. Engl. J. Med.* 363, 1211–1221. doi: 10.1056/NEJMoa0906312
- Movahedi, M., Amirzargar, A. A., Nasiri, R., Hirbod-Mobarakeh, A., Farhadi, E., Tavakol, M., et al. (2013). Gene polymorphisms of Interleukin-4 in allergic rhinitis and its association with clinical phenotypes. *Am. J. Otolaryngol.* 34, 676–681. doi: 10.1016/j.amjoto.2013.05.002
- Nagelkerke, N.J.D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692. doi: 10.1093/biomet/78.3.691
- Narozna, B., Hoffmann, A., Sobkowiak, P., Schoneich, N., Breborowicz, A., and Szczepankiewicz, A. (2016). Polymorphisms in the interleukin 4, interleukin 4 receptor and interleukin 13 genes and allergic phenotype: a case control study. *Adv. Med. Sci.* 61, 40–45. doi: 10.1016/j.advms.2015.07.003
- Nie, W., Zang, Y., Chen, J., and Xiu, Q. (2013). Association between Interleukin-4 Receptor α Chain (IL4RA) I50V and Q551R Polymorphisms and asthma risk: an update meta-analysis. *PLoS ONE* 8:e69120. doi: 10.1371/journal.pone.0069120
- Nissen, C. V., Bindslev-Jensen, C., and Mortz, C. G. (2015). Hypersensitivity to non-steroidal anti-inflammatory drugs (NSAIDs): classification of a Danish patient cohort according to EAACI/ENDA guidelines. *Clin. Transl. Allergy* 5:10. doi: 10.1186/s13601-015-0052-0
- Ober, C., and Hoffjan, S. (2006). Asthma genetics the long and winding road to gene discovery. *Genes Immun.* 7:6364284. doi: 10.1038/sj.gene.6364284
- Overton, N. L., Denning, D. W., Bowyer, P., and Simpson, A. (2016). Genetic susceptibility to allergic bronchopulmonary aspergillosis in asthma: a genetic association study. *Allergy Asthma Clin. Immunol.* 12:47. doi: 10.1186/s13223-016-0152-y
- Ozdoganoglu, T., and Songu, M. (2012). The burden of allergic rhinitis and asthma. *Ther. Adv. Respir. Dis.* 6, 11–23. doi: 10.1177/1753465811431975
- Palikhe, N. S., Kim, S. H., Cho, B. Y., Choi, G. S., Kim, J. H., Ye, Y. M., et al. (2010). IL-13 Gene Polymorphisms are Associated With Rhinosinusitis and Eosinophilic Inflammation in Aspirin Intolerant Asthma. *Allergy Asthma Immunol. Res.* 2, 134–140. doi: 10.4168/aa.2010.2.2.134
- Palikhe, N. S., Kim, S. H., Cho, B. Y., Ye, Y. M., Hur, G. Y., and Park, H. S. (2008). Association of three sets of high-affinity IgE receptor (FcepsilonR1) polymorphisms with aspirin-intolerant asthma. *Respir. Med.* 102, 1132–1139. doi: 10.1016/j.rmed.2008.03.017
- Papadopoulou, A., Kouis, P., Middleton, N., Kolokotroni, O., Karpathios, T., Nicolaidou, P., et al. (2015). Association of vitamin D receptor gene polymorphisms and vitamin D levels with asthma and atopy in Cypriot adolescents: a case-control study. *Multidiscip. Respir. Med.* 10:26. doi: 10.1186/s40248-015-0025-0
- Park, H. W., and Tantisira, K. G. (2017). Genetic signatures of asthma exacerbation. *Allergy Asthma Immunol. Res.* 9, 191–199. doi: 10.4168/aa.2017.9.3.191
- Park, K. Y., Park, M. K., Kim, E. J., Lee, M. K., and Seo, S. J. (2011). FCεRI gene promoter polymorphisms and total IgE levels in susceptibility to atopic dermatitis in Korea. *J. Korean Med. Sci.* 26, 870–874. doi: 10.3346/jkms.2011.26.7.870
- Paternoster, L., Standl, M., Chen, C. M., Ramasamy, A., Bønnelykke, K., Duijts, L., et al. (2011). Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat. Genet.* 44, 187–192. doi: 10.1038/ng.1017
- Pértegas Díaz, S., and Pita Fernández, S. (2003). Cálculo del poder estadístico de un estudio. *Cad. Atención Primaria.* 10, 59–63.
- Pillai, D. K., Iqbal, S. F., Benton, A. S., Lerner, J., Wiles, A., Foerster, M., et al. (2011). Associations between genetic variants in vitamin D metabolism and asthma characteristics in young African Americans: a pilot study. *J. Invest. Med.* 59, 938–946. doi: 10.2310/JIM.0b013e318220df41
- Pino-Yanes, M., Gignoux, C. R., Galanter, J. M., Levin, A. M., Campbell, C. D., Eng, C., et al. (2015). Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. *J. Allergy Clin. Immunol.* 135, 1502–1510. doi: 10.1016/j.jaci.2014.10.033
- Platts-Mills, T. A. (2001). The role of immunoglobulin E in allergy and asthma. *Am. J. Respir. Crit. Care Med.* 164:S1–5. doi: 10.1164/ajrccm.164.supplement_1.2103024
- Poon, A. H., Laprise, C., Lemire, M., Montpetit, A., Sinnett, D., Schurr, E., et al. (2004). Association of vitamin D receptor genetic variants with susceptibility to asthma and atopy. *Am. J. Respir. Crit. Care Med.* 170, 967–973. doi: 10.1164/rccm.200403-412OC
- Potaczek, D. P., and Kabesch, M. (2012). Current concepts of IgE regulation and impact of genetic determinants. *Clin. Exp. Allergy* 42, 852–871. doi: 10.1111/j.1365-2222.2011.03953.x
- Potaczek, D. P., Maeda, K., Wang, Q. H., Nakano, N., Kanada, S., Stepień, E., et al. (2009). FcepsilonR1alpha gene-18483A>C polymorphism affects transcriptional activity through YY1 binding. *Immunogenetics* 61, 649–655. doi: 10.1007/s00251-009-0391-x
- Potaczek, D. P., Michel, S., Sharma, V., Zeilinger, S., Vogelberg, C., von Berg, A., et al. (2013). Different FCER1A polymorphisms influence IgE levels in asthmatics and non-asthmatics. *Pediatr. Allergy Immunol.* 24, 441–449. doi: 10.1111/pai.12083
- Raby, B. A., Lazarus, R., Silverman, E. K., Lake, S., Lange, C., Wjst, M., et al. (2004). Association of vitamin D receptor gene polymorphisms with

- childhood and adult asthma. *Am. J. Respir. Crit. Care Med.* 170, 1057–1065. doi: 10.1164/rccm.200404-447OC
- Ramasamy, A., Kuokkanen, M., Vedantam, S., Gajdos, Z. K., Couto Alves, A., Lyon, H. N., et al. (2012). Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. *PLoS ONE* 7:e44008. doi: 10.1371/journal.pone.0044008
- Ramphul, K., Lv, J., Hua, L., Liu, Q. H., Fang, D. Z., Ji, R. X., et al. (2014). Single nucleotide polymorphisms predisposing to asthma in children of Mauritian Indian and Chinese Han ethnicity. *Braz. J. Med. Biol. Res.* 47, 394–397. doi: 10.1590/1414-431X20143751
- Robinson, C. L., Baumann, L. M., Gilman, R. H., Romero, K., Combe, J. M., Cabrera, L., et al. (2012). The Peru Urban versus Rural Asthma (PURA) Study: methods and baseline quality control data from a cross-sectional investigation into the prevalence, severity, genetics, immunology and environmental factors affecting asthma in adolescence in Peru. *BMJ Open* 2:e000421. doi: 10.1136/bmjopen-2011-000421
- Rondon, C., Blanca-Lopez, N., Campo, P., Mayorga, C., Jurado-Escobar, R., Torres, M. J., et al. (2017). Specific immunotherapy in local allergic rhinitis: A randomized, double-blind placebo-controlled trial with Phleum pratense subcutaneous allergen immunotherapy. *Allergy* 73, 905–915. doi: 10.1111/all.13350
- Saadi, A., Gao, G., Li, H., Wei, C., Gong, Y., and Liu, Q. (2009). Association study between vitamin D receptor gene polymorphisms and asthma in the Chinese Han population: a case-control study. *BMC Med. Genet.* 10:71. doi: 10.1186/1471-2350-10-71
- Sadeghnejad, A., Karmaus, W., Arshad, S. H., Kurukulaaratchy, R., Huebner, M., and Ewart, S. (2008). IL13 gene polymorphisms modify the effect of exposure to tobacco smoke on persistent wheeze and asthma in childhood, a longitudinal study. *Respir. Res.* 9:2. doi: 10.1186/1465-9921-9-2
- Sharma, V., Michel, S., Gaertner, V., Franke, A., Vogelberg, C., von Berg, A., et al. (2014). A role of FCER1A and FCER2 polymorphisms in IgE regulation. *Allergy* 69, 231–236. doi: 10.1111/all.12336
- Suaini, N. H., Zhang, Y., Vuillermin, P. J., Allen, K. J., and Harrison, L. C. (2015). Immune modulation by vitamin D and its relevance to food allergy. *Nutrients* 7, 6088–6108. doi: 10.3390/nu7085271
- Sun, L., Xue, W., Li, J., Zhou, Z., and Han, W. (2017). Palm dermatoglyphs and interleukin-4 receptor polymorphisms in asthma. *Biomed. Rep.* 6, 21–26. doi: 10.3892/br.2016.803
- Szczeklik, A., and Nizankowska-Mogilnicka, E. S. M. (2009). *Middleton's Allergy Principles and Practice*, eds N. Adkinson, W. Busse, B. Bochner, S. Holgate, and R. Lemanske (London: Mosby Inc; Elsevier), 1227–1243.
- Tian, H. Q., Chen, X. Y., Lu, Y., Lu, W. M., Wang, M. L., Zhao, H. L., et al. (2015). Association of VDR and CYP2R1 Polymorphisms with Mite-Sensitized Persistent Allergic Rhinitis in a Chinese Population. *PLoS ONE* 10:e0133162. doi: 10.1371/journal.pone.0133162
- Tian, H. Q., and Cheng, L. (2017). The role of vitamin D in allergic rhinitis. *Asia Pac. Allergy* 7, 65–73. doi: 10.5415/apallergy.2017.7.2.65
- Torres, M. J., Mayorga, C., Blanca-López, N., and Blanca, M. (2014). Hypersensitivity reactions to beta-lactams. *EXS* 104, 165–184. doi: 10.1007/978-3-0348-0726-5_11
- Veldman, C. M., Cantorna, M. T., and DeLuca, H. F. (2000). Expression of 1,25-Dihydroxyvitamin D3 receptor in the immune system. *Arch. Biochem. Biophys.* 374, 334–338. doi: 10.1006/abbi.1999.1605
- Wallace, D. V., Dykewicz, M. S., Bernstein, D. I., Blessing-Moore, J., Cox, L., Khan, D. A., et al. (2008). The diagnosis and management of rhinitis: an updated practice parameter. *J. Allergy Clin. Immunol.* 122 (Suppl. 2), S1–S4. doi: 10.1016/j.jaci.2008.06.003
- Wang, Y., Wang, O., Li, W., Ma, L., Ping, F., Chen, L., et al. (2015). Variants in vitamin d binding protein gene are associated with gestational diabetes mellitus. *Medicine* 94:e1693. doi: 10.1097/MD.0000000000001693
- Weidinger, S., Gieger, C., Rodriguez, E., Baurecht, H., Mempel, M., Klopp, N., et al. (2008). Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. *PLoS Genet.* 4:e1000166. doi: 10.1371/journal.pgen.1000166
- Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. (2005). A note on exact tests of hardy-weinberg equilibrium. *Am. J. Hum. Genet.* 76, 887–893. doi: 10.1086/429864
- Wjst, M. (2005). Variants in the vitamin D receptor gene and asthma. *BMC Genet.* 6:2. doi: 10.1186/1471-2156-6-2
- Wjst, M., Altmüller, J., Faus-Kessler, T., Braig, C., Bahnweg, M., and André, E. (2006). Asthma families show transmission disequilibrium of gene variants in the vitamin D metabolism and signalling pathway. *Respir. Res.* 7:60. doi: 10.1186/1465-9921-7-60
- Xie, C. N., Yue, M., Huang, P., Tian, T., Fan, H. Z., Wu, M. P., et al. (2018). Vitamin D binding protein polymorphisms influence susceptibility to hepatitis C virus infection in a high-risk Chinese population. *Gene* 679, 405–411. doi: 10.1016/j.gene.2018.09.021
- Xu, Y., Li, J., Ding, Z., Li, J., Li, B., Yu, Z., et al. (2017). Association between IL-13 +1923C/T polymorphism and asthma risk: a meta-analysis based on 26 case-control studies. *Biosci. Rep.* 37, 1–13. doi: 10.1042/BSR20160505
- Yang, H. J., Zheng, L., Zhang, X. F., Yang, M., and Huang, X. (2014). Association of the MS4A2 gene promoter C-109T or the 7th exon E237G polymorphisms with asthma risk: a meta-analysis. *Clin. Biochem.* 47, 605–611. doi: 10.1016/j.clinbiochem.2014.01.022
- Zhang, S., Li, Y., and Liu, Y. (2015). Interleukin-4–589C/T polymorphism is associated with increased pediatric asthma risk: a meta-analysis. *Inflammation* 38, 1207–1212. doi: 10.1007/s10753-014-0086-9
- Zhang, Y., Wang, Z., and Ma, T. (2017). Associations of Genetic Polymorphisms relevant to metabolic pathway of vitamin d3 with development and prognosis of childhood bronchial asthma. *DNA Cell Biol.* 36, 682–692. doi: 10.1089/dna.2017.3730
- Zhao, D. D., Yu, D. D., Ren, Q. Q., Dong, B., Zhao, F., and Sun, Y. H. (2017). Association of vitamin D receptor gene polymorphisms with susceptibility to childhood asthma: a meta-analysis. *Pediatr. Pulmonol.* 52, 423–429. doi: 10.1002/ppul.23548
- Zhou, J., Zhou, Y., Lin, L. H., Wang, J., Peng, X., Li, J., et al. (2012). Association of polymorphisms in the promoter region of FCER1A gene with atopic dermatitis, chronic urticaria, asthma, and serum immunoglobulin E levels in a Han Chinese population. *Hum. Immunol.* 73, 301–305. doi: 10.1016/j.humimm.2011.12.001
- Zhu, N., Gong, Y., Chen, X. D., Zhang, J., Long, F., He, J., et al. (2013). Association between the polymorphisms of interleukin-4, the interleukin-4 receptor gene and asthma. *Chin. Med. J.* 126, 2943–2951.
- Zuberbier, T., Lötval, J., Simoens, S., Subramanian, S. V., and Church, M. K. (2014). Economic burden of inadequate management of allergic diseases in the European Union: a GA(2) LEN review. *Allergy* 69, 1275–1279. doi: 10.1111/all.12470

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Amo, Martí, García-Menaya, Cordobés, Cornejo-García, Blanca-López, Canto, Doña, Blanca, Torres, Agúndez and García-Martín. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership