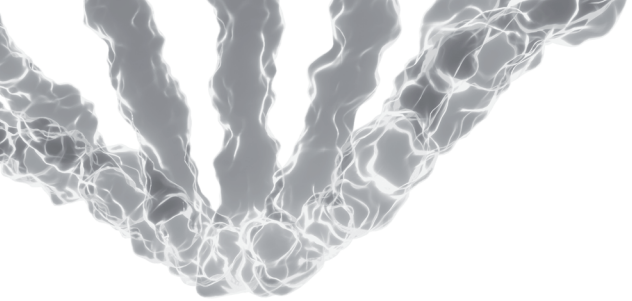
Towards solving the missing heritability in pharmacogenomics

Lee, M. van der



CHAPTER 1

General introduction

Pharmacogenomics

Personalized medicine is slowly becoming the new standard in healthcare and winning terrain over the one size fits all approach. One of the corner stones of personalized medicine is pharmacogenomics (PGx), which studies the influence of genetics biomarkers on drug response [1-3].

Studies regarding the link between drug response and genetics were published as early as the first half of the 20th century [4]. At that point PGx was described as the concept of "chemical individuality". From the 1950s onwards, PGx started to develop as its own distinct discipline. Many studies investigating PGx followed, resulting in an abundance of evidence for gene-drug interactions [4,5], many of which are still relevant today. Since then, the field has advanced significantly both in regards to technological developments as well as clinical implementation.

Technologies for pharmacogenomics

The application of PGx in clinical practice requires identification of genetic variants in the pharmacogenes of interest. The wide range of available technologies for genotyping can be roughly divided into four major categories: single nucleotide variant (SNV)-panels or arrays, first generation sequencing (e.g. Sanger sequencing), next generation sequencing (NGS) (e.g. Illumina short-read sequencing) and third generation or long-read sequencing (e.g. Pacific Bioscience Single molecule real time sequencing). In the field of clinical PGx, SNV panels are most commonly used [6-8]. These panels have been developed to identify a predefined list of variants. Sequencing approaches, on the other hand, in general allow for the characterization of all SNVs and small indels in the covered region. This includes rare variants as well. Sequencing can be divided into short-read sequencing (up to ~200bp) and long-read sequencing (>10kbp). While for PGx the SNV panels are still the golden standard, the sequencing based approaches might show valuable in the future [9]. A more detailed overview of available technologies and their application is presented in chapter 2.

Clinical pharmacogenomics

The variants identified in the pharmacogenes are translated into haplotypes and predicted phenotypes in order to use them in clinical practice, Figure 1.1. In 2000, Ingelman-Sundberg et al. proposed the *-nomenclature in anticipation of the abundance of variants which was expected to be discovered [10]. Within this system *-haplotypes are assigned to combinations of variants which are seen in a gene, Figure 1.1. The two *-haplotypes

(one per allele) are combined into a diplotype which is subsequently translated into a predicted phenotype (Figure 1.1). For most metabolic enzymes, four predicted phenotype categories are recognized: Poor Metabolizer (PM), Intermediate Metabolizers (IM), Normal Metabolizers (NM) (previously called extensive metabolizers (EM)) and ultra-rapid metabolizers (UM). These phenotypes are used in the dosing guidelines of the Dutch Pharmacogenetics Working Group (DPWG) and the American Clinical Pharmacogenetics Implementation Consortium (CPIC) [11-13]. These guidelines help clinicians to adjust a patients' therapy based on the patients' genetic profile. Currently, the DPWG and CPIC combined have developed guidelines containing clinically actionable advice for over 50 drugs and 12 different genes [14,15].

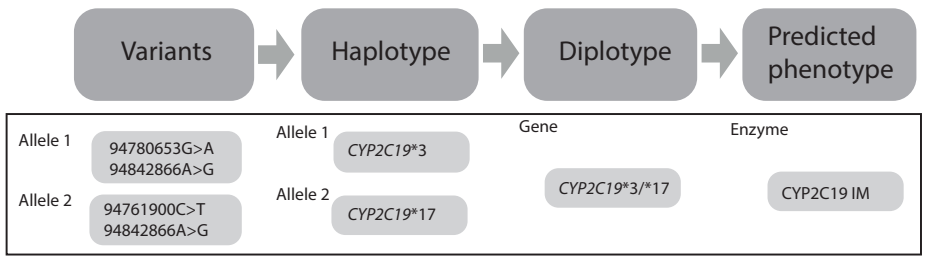


Figure 1.1: Pharmacogenomic genotype and phenotype assignment using the *CYP2C19* gene as an example
Variants per allele are assigned a haplotype according to the star(*)-allele nomenclature. The two haplotypes together form the diplotype which is subsequently translated into a predicted phenotype. IM: intermediate metabolizer.

The *-haplotype system has many benefits. It allows for standardization of the interpretation of common PGx variants and the *-haplotypes are relatively easy to understand and recognize for clinicians who have to work with PGx but are not specialized in genetics. However, the limitations are becoming more apparent as the field advances. The activity attributed to a specific *-allele is dependent on the gene. For example, a *CYP2D6**2 allele is considered to have a normal function while a *CYP2C9**2 allele is predicted to result in a non-functional enzyme. Moreover, variants which are not (yet) a part of the *-nomenclature cannot be taken into account in phenotype predictions [16-18]. With the adoption of sequencing for PGx, the promise of a 'SNP explosion' as stated by Ingelman-Sundberg et al. is becoming reality [10]. The scale on which sequencing is currently performed causes a substantial rise in newly identified variants. For example, in 2019 an overview of the number of alleles in the PharmVar database was published. At that point there were 122 *-haplotypes (*1-*122) recognized for *CYP2D6*. Currently, 2021, there are already 23 additional *-haplotypes (*1-*145) for the same gene. The same type of increase is observed

for other genes [19,20]. Similar trends are observed for the number of variants reported, with 324 million human reference SNPs in 2017 to 730 million in 2020 [21,22]. This vast increase of variant and haplotype information forced a shift towards approaches that are better capable at processing these large amounts of genetic data.

Missing heritability of drug response

Most clinical applications of PGx focus on the analysis of a small panel of 50–100 well-known variants and haplotypes to predict the phenotypes. These approaches are projected to result in less adverse drug reactions (ADRs), better efficacy and overall lower health costs [23,24]. For a limited number of drug-gene interactions, gold-standard evidence proving the clinical benefit of PGx is available [25]. For other genes, lower levels of evidence are available and used in the clinical guidelines [15,26]. Nonetheless, not all genetic variability can be explained with current clinical PGx. A phenomenon known as the missing heritability [27-29]. For example, for metoprolol, monozygotic twin studies have shown that approximately 90% of its pharmacokinetics is hereditary. Nonetheless, current clinical PGx is only able to explain 41% of the variation of its metabolism, indicating that a large amount of the genetic variability is yet unexplained [27]. Metoprolol is predominantly metabolized by CYP2D6, leading to the hypothesis that not all genetic variability in *CYP2D6* is currently used in clinical practice. The same effect is observed for *CYP3A4*, with a missing heritability of approximately 80%; Up till 90% is genetic but only ~10% can be explained with current *CYP3A4* PGx [28,30]. Many potential causes for this missing heritability have been suggested.

Firstly, rare variants might play a crucial role (Figure 1.2A). Individual rare variants are by definition uncommon (minor allele frequency (MAF) <1%), however, in total they make up the largest group of variants in pharmacogenes [18]. Moreover, rare variants are expected to be highly deleterious. Due to their low frequency it is difficult to assess their impact.

Secondly, the majority of pharmacogenes is at least partially complex and several are extremely complex (Figure 1.2B) [31]. With complex defined as regions that overlap with (low copy) repeats. These complexities can be small repeats, such as the *UGT1A1**28 TATA-box they can indicate complete gene duplications or deletions. Another category of complex variants are the pseudogenes. Several pharmacogenes have a (neighboring) pseudogene, which is highly similar in regard to genetic make-up but often unfunctional [32,33]. These pseudogenes often share a significant amount of their sequence with the pharmacogene, complicating the characterization of the pharmacogene. This sequence homology can be as high as 95% [32,34,35]. Moreover, pseudogenes can form hybrid structures with their neighboring pharmacogene resulting in an non-functional hybrid formation [32,34].

Besides being highly similar to pseudogenes, pharmacogenes can also have high sequence homology with each other. The genes in the *CYP3A* locus (*CYP3A4*, *CYP3A5* and *CYP3A7*) share as much as 80% of their sequence [28,36]. An inability to accurately characterize these complex variants can lead to miss-classification of a phenotype.

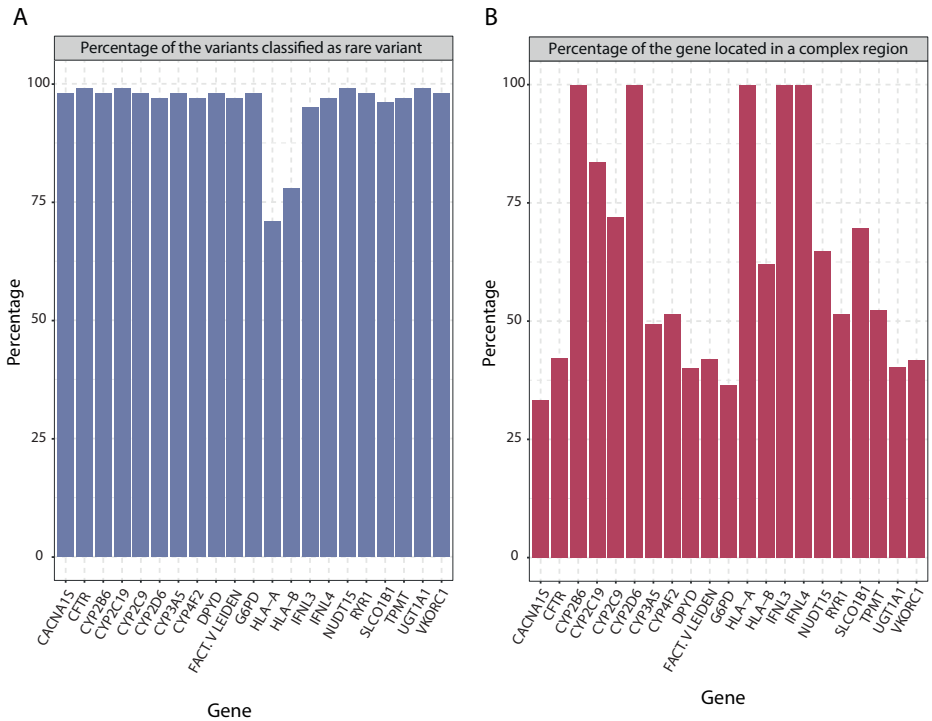


Figure 1.2: Rare variants and genetic complexity of pharmacogenes
Genes included are the genes with a clinically actionable guideline in either the Dutch Pharmacogenetic Working Group guidelines or in the guidelines from the Clinical Pharmacogenetics Implementation Consortium. Not for of the all variants within these genes it is known what their functional effect is, the variants used in the clinical guidelines represent only a small portion of all variants. These figures include all variants within the pharmacogenes, independent of their effect. **A:** percentage of variants within the pharmacogene which is classified as rare (a minor allele frequency <1% in an aggregated population). Data is extracted from Gnomad (GRCh38). **B:** Part of the locus defined as complex, this is the percentage of the locus defined as a (low copy) repeat on the UCSC browser (University of California Santa Cruz).

Thirdly, enzyme function is not a categorical and is not a matter of 'on' or 'off'. Nonetheless, variants are grouped into *-haplotypes and predicted phenotype. This leads to the clustering of a wide range of variants into only 4 categories, Figure 1.3. This approach uses a 0%, 50% or 100% activity for *-haplotypes and a doubling of this activity in the case of a gene duplication. This artificial clustering is likely to lead to a loss in information and detail in phenotype predictions.

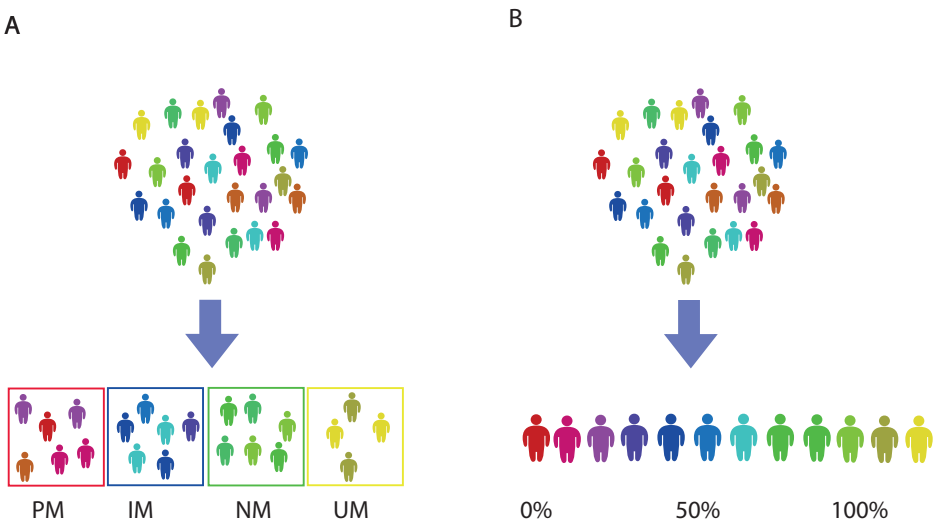


Figure 1.3: Phenotype prediction models
A: Conventional methods to assign haplotypes and phenotypes result in four categories. **B:** A continuous scale might better reflect differences in drug response. PM: poor metabolizer, IM: intermediate metabolizer, NM: normal metabolizer, UM: ultra-rapid metabolizer.

Finally, not every substrate is affected in the same way by a specific *-haplotype, indicating the effect of substrate specific interactions. The clinical guidelines provide advice for gene-drug interactions assuming similar effect of a gene on every substrate [37,38].

All these factors might play a role in the missing heritability. Accounting for these factors in phenotype predictions might help improve the explained variability in PGx.

Aim and outline of this thesis

The general aim of this thesis is to investigate the missing heritability in PGx with the use of novel technologies.

In **chapter 2**, we provide an overview of different genetic technologies which can be used in PGx. The technologies discussed range from SNV (single nucleotide variant) panels to short- and long-read sequencing. For each technology a short description will be provide as well as examples of the use of the technologies for PGx. In the following chapters we apply several of these technologies for PGx problems. **Chapter 3** merges the fields of human genetics and PGx by extracting a clinically meaningful PGx profile from existing diagnostic whole exome sequencing (WES) data. For many individuals, genetic data is readily available as it has previously been obtained for diagnostic purposes. By extracting data based on the Ubiquitous pharmacogenomics variant panel, we ensure that

the results are clinically useful for the patients [6]. Moreover, the clinical data obtained is based on child-parents trio sequencing. This allows us to phase variants to the paternal and maternal allele and to investigate the added value of haplotype phasing in clinical PGx. However, WES cannot cover all important regions and is limited in characterizing the more complex pharmacogenes [31,39]. In **chapter 4** we, therefore, use long-read whole genome data of one well characterized individual to determine the added value of long-read sequencing in regards to solving the complexities of pharmacogenetic regions. In this proof of concept study it is shown that with long-read sequencing one is capable to resolve the majority of complex pharmacogenes. Moreover, long-read sequencing allows us to phase variants to their alleles of origin without the need for trio sequencing or for computational phasing. Next, in **chapter 5** we combine long-read sequencing with haplotype independent phenotype predictions. Here, we use long-read Pacific Bioscience (PacBio) technology to sequence the entire *CYP2D6* locus in one fully phased read. We then combine the sequencing data with pharmacokinetic data of individuals who have used tamoxifen. The combination of sequencing data and kinetic data allows us to develop a neural network to predict the CYP2D6 metabolism directly from the variant information, thereby omitting the commonly used (inadequate) phenotype summarization. The neural network is validated in two independent clinical cohorts and with in vitro data.

In the final chapters, we zoom in on the variability in the activity of the CYP2D6 enzyme. In *CYP2D6* gene-drug interactions, the phenotype assignments are generalized for the entire enzyme. However, variant can affect different substrates to a different extend, thereby causing substrate specific effects [37,38]. In **chapter 6**, we explore these substrate specific effects by systematically looking at in vitro results found in literature. While in vitro data is routinely used to assign or confirm the effect of a specific haplotype, the data is highly variable and cannot easily be generalized to the in vivo situation and to other in vitro studies. Indications for substrate specific effects were found which can be explained by alterations in the substrate binding pocket. Nonetheless, the number of substrates investigated in regards to substrate specificity is limited and conclusion can only be drawn for a small number of substrates. In **chapter 7** we are taking a step towards clinical practice and focus on variability within clinically used CYP2D6 phenotype categories. We compare the variance in different metabolic groups to assess which group contains the most variability and therefore the most potentially miss-classified phenotypes. Moreover, we compare the observed activity of different gene activity score groups to find the source of the variability in the phenotype groups.

This thesis concludes with a general discussion including future perspectives (**chapter 8**) followed by English and Dutch summaries (**chapter 9**).

References

1. Roden, D.M.; McLeod, H.L.; Relling, M.V.; Williams, M.S.; Mensah, G.A.; Peterson, J.F.; Van Driest, S.L. Pharmacogenomics. *Lancet (London, England)* **2019**, *394*, 521-532, doi:10.1016/s0140-6736(19)31276-0.
2. Relling, M.V.; Evans, W.E. Pharmacogenomics in the clinic. *Nature* **2015**, *526*, 343-350, doi:10.1038/nature15817.
3. Weinshilboum, R.; Wang, L. Pharmacogenomics: bench to bedside. *Nat Rev Drug Discov* **2004**, *3*, 739-748, doi:10.1038/nrd1497.
4. Meyer, U.A. Pharmacogenetics - five decades of therapeutic lessons from genetic diversity. *Nature reviews. Genetics* **2004**, *5*, 669-676, doi:10.1038/nrg1428.
5. Charlab, R.; Zhang, L. Pharmacogenomics: historical perspective and current status. *Methods in molecular biology (Clifton, N.J.)* **2013**, *1015*, 3-22, doi:10.1007/978-1-62703-435-7_1.
6. van der Wouden, C.H.; van Rhenen, M.H.; Jama, W.O.M.; Ingelman-Sundberg, M.; Lauschke, V.M.; Konta, L.; Schwab, M.; Swen, J.J.; Guchelaar, H.J. Development of the PGx-Passport: A Panel of Actionable Germline Genetic Variants for Pre-emptive Pharmacogenetic Testing. *Clinical pharmacology and therapeutics* **2019**, 10.1002/cpt.1489, doi:10.1002/cpt.1489.
7. Jannetto, P.J.; Laleli-Sahin, E.; Wong, S.H. Pharmacogenomic genotyping methodologies. *Clinical chemistry and laboratory medicine* **2004**, *42*, 1256-1264, doi:10.1515/cclm.2004.246.
8. Mukerjee, G.; Huston, A.; Kabakchiev, B.; Piquette-Miller, M.; van Schaik, R.; Dorfman, R. User considerations in assessing pharmacogenomic tests and their clinical support tools. *NPJ genomic medicine* **2018**, *3*, 26, doi:10.1038/s41525-018-0065-4.
9. van der Lee, M.; Kriek, M.; Guchelaar, H.J.; Swen, J.J. Technologies for Pharmacogenomics: A Review. *Genes (Basel)* **2020**, *11*, doi:10.3390/genes11121456.
10. Ingelman-Sundberg, M.; Daly, A.K.; Oscarson, M.; Nebert, D.W. Human cytochrome P450 (CYP) genes: recommendations for the nomenclature of alleles. *Pharmacogenetics* **2000**, *10*, 91-93, doi:10.1097/00008571-200002000-00012.
11. Swen, J.J.; Wilting, I.; de Goede, A.L.; Grandia, L.; Mulder, H.; Touw, D.J.; de Boer, A.; Conemans, J.M.; Egberts, T.C.; Klungel, O.H., et al. Pharmacogenetics: from bench to byte. *Clinical pharmacology and therapeutics* **2008**, *83*, 781-787, doi:10.1038/sj.clpt.6100507.
12. Relling, M.V.; Klein, T.E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clinical pharmacology and therapeutics* **2011**, *89*, 464-467, doi:10.1038/clpt.2010.279.
13. Caudle, K.E.; Sangkuhl, K.; Whirl-Carrillo, M.; Swen, J.J.; Haidar, C.E.; Klein, T.E.; Gammal, R.S.; Relling, M.V.; Scott, S.A.; Hertz, D.L., et al. Standardizing CYP2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clinical and translational science* **2020**, *13*, 116-124, doi:10.1111/cts.12692.
14. Royal Dutch Pharmacist Association. Pharmacogenetics. Availabe online: www.knmp.nl/farmacogenetica
15. Clinical pharmacogenetics Implementation ConsortiumCPIC-guidelines. Availabe online: https://cpicpgx.org/
16. Drogemoller, B.I.; Wright, G.E.; Warnich, L. Considerations for rare variants in drug metabolism genes and the clinical implications. *Expert opinion on drug metabolism & toxicology* **2014**, *10*, 873-884, doi:10.1517/17425255.2014.903239.
17. Kozyra, M.; Ingelman-Sundberg, M.; Lauschke, V.M. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genetics in medicine : official journal of the American College of Medical Genetics* **2017**, *19*, 20-29, doi:10.1038/gim.2016.33.

18. Ingelman-Sundberg, M.; Mkrtchian, S.; Zhou, Y.; Lauschke, V.M. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Human genomics* **2018**, *12*, 26, doi:10.1186/s40246-018-0157-3.
19. Gaedigk, A.; Sangkuhl, K.; Whirl-Carrillo, M.; Twist, G.P.; Klein, T.E.; Miller, N.A. The Evolution of PharmVar. *Clinical pharmacology and therapeutics* **2019**, *105*, 29-32, doi:10.1002/cpt.1275.
20. Pharmacogene Variation Consortium. PharmVar. Availabe online: https://www.pharmvar.org
21. National Library of Medicine. NCBI insights. Availabe online: https://ncbiinsights.ncbi.nlm.nih.gov/2017/05/08/dbsnps-human-build-150-has-doubled-the-amount-of-refsnp-records/
22. National Library of Medicine. Available online: https://ncbiinsights.ncbi.nlm.nih.gov/2020/06/11/dbsnp-human-build-154-release-alfa-data/
23. Wu, A.C.; Fuhlbrigge, A.L. Economic evaluation of pharmacogenetic tests. *Clinical pharmacology and therapeutics* **2008**, *84*, 272-274, doi:10.1038/clpt.2008.127.
24. Pirmohamed, M. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. *Annu Rev Genomics Hum Genet* **2014**, *15*, 349-370, doi:10.1146/annurev-genom-090413-025419.
25. van der Wouden, C.H.; Cambon-Thomsen, A.; Cecchin, E.; Cheung, K.C.; Davila-Fajardo, C.L.; Deneer, V.H.; Dolzan, V.; Ingelman-Sundberg, M.; Jonsson, S.; Karlsson, M.O., et al. Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium. *Clinical pharmacology and therapeutics* **2017**, *101*, 341-358, doi:10.1002/cpt.602.
26. Dutch Pharmacogenomics Working Group. Pharmacogenetics Guidelines. Royal Dutch Pharmacists Association: KNMP Kennisbank, 2020.
27. Matthaei, J.; Brockmoller, J.; Tzvetkov, M.V.; Sehrt, D.; Sachse-Seeboth, C.; Hjelmborg, J.B.; Moller, S.; Halekoh, U.; Hofmann, U.; Schwab, M., et al. Heritability of metoprolol and torsemide pharmacokinetics. *Clinical pharmacology and therapeutics* **2015**, *98*, 611-621, doi:10.1002/cpt.258.
28. Klein, K.; Zanger, U.M. Pharmacogenomics of Cytochrome P450 3A4: Recent Progress Toward the "Missing Heritability" Problem. *Frontiers in genetics* **2013**, *4*, 12, doi:10.3389/fgene.2013.00012.
29. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A., et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747-753, doi:10.1038/nature08494.
30. Ozdemir, V.; Kalow, W.; Tang, B.K.; Paterson, A.D.; Walker, S.E.; Endrenyi, L.; Kashuba, A.D. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. *Pharmacogenetics* **2000**, *10*, 373-388, doi:10.1097/00008571-200007000-00001.
31. Lauschke, V.M.; Milani, L.; Ingelman-Sundberg, M. Pharmacogenomic Biomarkers for Improved Drug Therapy-Recent Progress and Future Developments. *The AAPS journal* **2017**, *20*, 4, doi:10.1208/s12248-017-0161-x.
32. Gaedigk, A.; Jaime, L.K.; Bertino, J.S., Jr.; Berard, A.; Pratt, V.M.; Bradfordand, L.D.; Leeder, J.S. Identification of Novel CYP2D7-2D6 Hybrids: Non-Functional and Functional Variants. *Frontiers in pharmacology* **2010**, *1*, 121, doi:10.3389/fphar.2010.00121.
33. Jacob, R.M.; Johnstone, E.C.; Neville, M.J.; Walton, R.T. Identification of CYP2B6 sequence variants by use of multiplex PCR with allele-specific genotyping. *Clinical chemistry* **2004**, *50*, 1372-1377, doi:10.1373/clinchem.2004.031708.
34. Gaedigk, A. Complexities of CYP2D6 gene analysis and interpretation. *Int Rev Psychiatry* **2013**, *25*, 534-553, doi:10.3109/09540261.2013.825581.

35. Qiao, W.; Yang, Y.; Sebra, R.; Mendiratta, G.; Gaedigk, A.; Desnick, R.J.; Scott, S.A. Long-Read Single Molecule Real-Time Full Gene Sequencing of Cytochrome P450-2D6. *Human mutation* **2016**, *37*, 315-323, doi:10.1002/humu.22936.
36. Williams, J.A.; Ring, B.J.; Cantrell, V.E.; Jones, D.R.; Eckstein, J.; Ruterbories, K.; Hamman, M.A.; Hall, S.D.; Wrighton, S.A. Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7. *Drug metabolism and disposition: the biological fate of chemicals* **2002**, *30*, 883-891, doi:10.1124/dmd.30.8.883.
37. Muroi, Y.; Saito, T.; Takahashi, M.; Sakuyama, K.; Niinuma, Y.; Ito, M.; Tsukada, C.; Ohta, K.; Endo, Y.; Oda, A., et al. Functional characterization of wild-type and 49 CYP2D6 allelic variants for N-desmethyltamoxifen 4-hydroxylation activity. *Drug metabolism and pharmacokinetics* **2014**, *29*, 360-366.
38. Bogni, A.; Monshouwer, M.; Moscone, A.; Hidestrand, M.; Ingelman-Sundberg, M.; Hartung, T.; Coecke, S. Substrate specific metabolism by polymorphic cytochrome P450 2D6 alleles. *Toxicol In Vitro* **2005**, *19*, 621-629, doi:10.1016/j.tiv.2005.04.001.
39. Ingelman-Sundberg, M.; Sim, S.C. Intronic polymorphisms of cytochromes P450. *Human genomics* **2010**, *4*, 402-405.