Inaugural dissertation

for obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Sebastian Pirmann

born in: Bad Dürkheim, Germany

Oral examination: June 10th, 2024

Pharmacogenomic Analyses of Next-Generation Sequencing Data from Cancer Patients

Referees: Prof. Dr. Benedikt Brors

Dr. Dr. med. Daniel Hübschmann

Abstract

Pharmacogenomics (PGx) investigates how genetic factors influence the effects and side effects of drugs. Recently, however, PGx has moved beyond the genome and aims at also accounting for epigenetic and transcriptomic variation to further study differences in drug response between patients. Typically, PGx is mainly concerned with germline variation in pharmacogenes, encoding drug-metabolizing enzymes and transporters, that influence pharmacokinetics and drug metabolism. In contrast, somatic variation has so far been studied almost exclusively in drug target genes. In this thesis, the genomic, epigenetic, and transcriptomic variation of 60 selected pharmacogenes was analyzed based on germline DNA sequencing, tumor DNA and RNA sequencing, and tumor methylation profiling. The data was derived from matched tumor and germline control samples of 2,371 cancer patients suffering from rare or advanced cases of cancers that have already undergone all standard lines of treatment. The focus of this thesis was especially on somatic variation and its effects on pharmacokinetics in the tumor, which has so far been neglected in pharmacogenomic research but has already been hypothesized repeatedly as a potential mechanism for the development of drug resistance in tumors.

First, a comprehensive and efficient in-silico PGx analysis pipeline was developed. Germline samples were analyzed for star-allele genotypes and phenotypes based on known functional single nucleotide polymorphisms and copy number variants. In addition, rare variants in the germline of these patients, and their functional effect were assessed using variant effect prediction tools. These rare variants were integrated into the germline PGx profiles which showed that they can superimpose on the phenotypes derived solely from star alleles. The pipeline was also integrated into a molecular tumor board workflow providing PGx recommendations.

Comprehensive PGx analyses were also carried out for the tumor samples. The results showed that in rare cases somatic variants at star-allele positions can change the genotype between tumor and matched control sample. However, a large part of somatic variation of the pharmacogenes included copy number aberrations. Analyses of the expression in the tumor samples revealed that some (especially phase II genes and transporters) are expressed in multiple tumor types. In addition, the expression of some genes was strongly associated with the copy number aberrations while for others methylation seems to be the major regulating factor. Finally, a combined multivariate analysis of all the aforementioned data levels was done to assess the proportion of variance explained in tumor gene expression.

Zusammenfassung

Die Pharmakogenomik (PGx) untersucht, wie genetische Faktoren die Wirkungen und Nebenwirkungen von Medikamenten beeinflussen. In letzter Zeit hat sich die Pharmakogenomik jedoch über das Genom hinaus entwickelt und hat das Ziel, auch epigenetische und transkriptomische Variation zu berücksichtigen, um Unterschiede in der Arzneimittelreaktion zwischen Patienten noch genauer untersuchen und beschreiben zu können. Typischerweise befasst sich die Pharmakogenomik hauptsächlich mit Keimbahnvarianten in Pharmakogenen, die für Enzyme und Transporter codieren, welche die Pharmakokinetik und den Arzneimittelstoffwechsel beeinflussen. Im Gegensatz dazu wurde somatische Variation bisher fast ausschließlich in Genen, die Arzneimittel-Targets darstellen, untersucht. In dieser Arbeit wurde deshalb die genomische, epigenetische und transkriptomische Variation von 60 ausgewählten Pharmakogenen anhand von Keimbahn-DNA-Sequenzierung, Tumor-DNA- und RNA-Sequenzierung sowie Tumor-Methylierungsprofilierung analysiert. Die Daten stammen von gepaarten Tumor- und Keimbahnproben von 2.371 Krebspatienten mit seltenen oder fortgeschrittenen Krebserkrankungen, die bereits alle Standardtherapien durchlaufen haben. Der Schwerpunkt dieser Arbeit lag insbesondere auf somatischer Variation und deren Auswirkung auf die Pharmakokinetik im Tumor, die bisher in der pharmakogenomischen Forschung vernachlässigt wurden, aber in der Theorie bereits mehrfach als möglicher Mechanismus für die Entwicklung von Arzneimittelresistenzen in Tumoren diskutiert wurde.

Zunächst wurde eine umfassende und effiziente in-silico Pharmakogenomik-Pipeline entwickelt. Keimbahnproben wurden auf Sternallel-Genotypen und -Phänotypen basierend auf bekannten funktionellen Einzelnukleotidpolymorphismen und Kopienzahlvarianten analysiert. Darüber hinaus wurden seltene Varianten in der Keimbahn dieser Patienten gefunden, und ihre funktionelle Wirkung wurde mithilfe von computerbasierten Prädiktionstools untersucht. Diese seltenen Varianten wurden in die Keimbahn-PGx-Profile der Patienten integriert, und es zeigte sich, dass sie die Phänotypen, welche ausschließlich aus Sternallelen abgeleitet wurden, überlagern können. Die Pipeline wurde auch in den Workflow eines moleularen Tumorboards integriert, der den Onkologen Pharmakogenomik-Empfehlungen für die Behandlung bereitstellt.

Umfassende pharmakogenomische Analysen wurden auch für die Tumorproben durchgeführt. Die Ergebnisse zeigten, dass in seltenen Fällen somatische Varianten an bekannten Sternallel-Positionen zu einer Veränderung des Genotyps zwischen Tumor und Kontrollprobe eines Patienten führen können. Ein großer Teil der somatischen Variation der ausgewählten Pharmakogene umfasste jedoch Kopienzahlaberrationen. Analysen der Expression

in den Tumorproben zeigten, dass einige (insbesondere Phase-II Gene und Transporter) in mehreren Tumorarten exprimiert werden. Darüber hinaus war die Expression einiger Gene stark mit den Kopienzahlaberrationen assoziiert, während für andere die Methylierung den hauptsächlichen Regulationsfaktor darstellte. Schließlich wurde eine kombinierte Analyse aller oben genannten Datenebenen mithilfe eines multivariaten Modells durchgeführt, um deren Einfluss auf die Expression der Pharmakogene im Tumor zu untersuchen.

Contents

List of Figures

List of Tables

List of Abbreviations

PGx	Pharmacogenomics
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
NGS	Next Generation Sequencing
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
ADME	Absorption Distribution Metabolism Excretion
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
CNV	Copy Number Variation
CNA	Copy Number Aberration
VEP	Variant Effect Prediction
MASTER	Molecularly Aided Stratification for Tumor Eradication Research

Chapter 1

Introduction

1.1 Next-Generation-Sequencing

The ability to determine nucleotide sequences of ribonucleic acids such as DNA or RNA, the genetic basis of all living organisms, has made significant progress in recent decades and has revolutionized the life sciences. From the initial development of low-throughput methods such as Sanger sequencing [1], and the advances made in the Human Genome Project [2] to high-throughput techniques like those employed in the 1000 Genomes Project [3], there has been a continuous technological evolution for generation and analysis of sequencing data. Next-generation sequencing (NGS) refers to sequencing methods that are characterized by massive parallelization enabling a significant increase in speed compared to previous methods. NGS has been a breakthrough offering unparalleled precision and efficiency for decoding an individual's genetic information [4], as the entire sequence of a human genome can be determined within a single day.

NGS is comprised of the following methods: The DNA to be sequenced is divided into smaller pieces (fragments), which are amplified (e.g. by polymerase chain reaction) and can then be sequenced in parallel. In NGS the sequencing-by-synthesis method is often used, a term introduced by one of the vendors. Here all DNA fragments (library) are bound to a flow cell and read out by the continuous addition of complementary bases marked with fluorescent dye, requiring no chain termination [5, 6]. The respective color stands for one of the 4 nucleobases (adenine, guanine, cytosine, thymine) and can be recorded using an imaging system (base calling). This generates short reads (35-700 base pairs [6]) of the sequenced DNA, usually stored in FASTQ file format [7], which can subsequently be used to align the reads to a genome reference. This allows to determine the original region of the genome the reads belong to and therefore enables a reconstruction of the complete sequence. The amount of reads covering a genomic position is referred to as coverage (often one aims at >30 [8]). The higher the coverage, the higher the probability to correctly discriminate

sequencing errors (with a close to random and uniform distribution) from actual mutations or variants (all at the same position). Reads together with their coordinates obtained via alignment are stored in a compressed BAM format (binary alignment map) [7] and can be used for a wide variety of computer-based analyses.

One of the main applications of NGS is research in diseases with genetic alterations, like mendelian inherited diseases with germline alterations or cancer with mostly somatic, but also sometimes germline alterations [9–11]. To identify the genetic causes of a disease, it is necessary to determine in which way the genetic information of affected individuals differs from that of healthy ones. In NGS there are different approaches, notably whole exome sequencing (WES) and whole genome sequencing (WGS), targeting only protein-coding regions or the entire genome, respectively. Additionally, RNA sequencing (RNA-Seq) provides detailed transcriptomic information (via sequencing a cDNA library obtained after reverse transcription) [12]. This vast amount of biological data has led to the development of increasingly specialized algorithms for subsequent analyses. An example is the detection of genetic variants that the genome of a patient carries in comparison to a standardized reference genome. These can be detected by variant calling [13], which uses aligned reads and determines the positions at which the investigated sequence differs from the reference [14]. This enables, for example, the detection of mutations in a tumor and allows subsequent biological interpretation. Such mutations in a tumor can either be inherited (in the case of germline mutations) or acquired in the course of a lifetime through somatic mutation processes. In personalized oncology, NGS facilitates the discovery of diagnostic and prognostic molecular biomarkers. This understanding has led to the design of targeted therapies that selectively intervene in molecular pathways driving tumorigenesis and progression [15, 16].

Since the possibilities of modern omics technologies are no longer restricted to only sequencing DNA or RNA, but instead also enable the generation of a wealth of other biological data such as methylation or abundance of proteins and metabolites through other high-throughput technologies, more and more methods have been developed to integrate and analyze these data layers together in order to obtain the most comprehensive picture of the underlying biology of diseases [17]. These multi-omics approaches involve the simultaneous analysis and integration of diverse molecular datasets such as genomics, epigenomics, transcriptomics, proteomics, or metabolomics for classification, clustering, or correlation tasks [18]. In cancer research, multi-omics allows researchers to decipher complex oncogenic pathways and comprehend disease heterogeneity [19]. This systems biology perspective enabled by multi-omics goes beyond isolated molecular findings, offering a holistic view of cancer as a complex, dynamic system. This depth of information not only refines our understanding of cancer types but also aids in the identification of robust biomarkers critical for early diagnosis, prognosis, and predicting treatment responses [16]. Moreover, the integrative analysis

of multi-omics datasets extends to personalized medicine, where tailored treatment strategies are developed based on the unique molecular profiles of individual patients [17, 20].

1.2 Pharmacogenomics and Drug Metabolism

Pharmacogenomics, often referred to as PGx, is a field of research that tries to investigate how genetic variation influences an individual's response to drugs and includes pharmacokinetic and pharmacodynamic processes [21–23]. Genetic factors are thought to contribute to 15%–30% of the variance in drug response between individuals [24]. However, it is still common nowadays for patients to receive standard doses of medication that have been developed for a standard, average patient based on regular clinical trials. Pharmacogenomics, as part of personalized medicine, tries to address the limitations of this one-size-fits-all approach and leverages genetic information to optimize drug therapy for individual patients [25]. Based on the individual genetic make-up of each patient, an optimal dosage can be determined for many drugs. PGx has gained major importance and attention due to its high potential to improve the practice of drug-based medicine, making it safer by reducing side effects and simultaneously increasing efficacy [26]. Despite these advances, comprehensive clinical implementation of PGx is progressing only slowly and is mainly limited to individual well-proven associations between single genes and drugs for which commercial tests are available. The genes of interest for PGx are the so-called pharmacogenes which are genes coding for proteins that interact with drugs including drug transporters, drug-metabolizing enzymes, and drug targets. Drug metabolism is part of xenobiotic metabolism, which generally also includes other exogenous substances like environmental toxins. The genes that regulate xenobiotic metabolism are summarized as ADME genes, being a subset of pharmacogenes. ADME refers to the individual steps of the path drugs or other xenobiotics take through the body (pharmacokinetics). These include absorption (A) into the bloodstream, distribution (D) throughout the body, metabolism (M) including the breakdown and chemical modification/detoxification, and excretion (E) from the body. Some ADME genes are additionally involved in the metabolism of endogenous substances like hormones or fatty acids [27]. Besides ADME, the effect of drugs on their targets and thus on the body (pharmacodynamics) is also a major component of pharmacogenomic research. Drug metabolism can conceptually be divided into distinct phases in which a substance undergoes different chemical modifications. This involves phase I, phase II, and phase III reactions, as well as drug transport across cell membranes [28]. Usually, drugs have lipophilic properties, which is why they have to be metabolized to become more water-soluble to be excreted. In phase I metabolism, enzymes such as cytochrome P450 initiate oxidation, reduction, and hydrolysis reactions. The primary objective is to introduce functional groups,

such as hydroxyl, rendering the drug more polar and reactive. This prepares the drug for subsequent reactions in phase II which involves transferase enzymes. These catalyze the addition of large functional groups to phase I metabolites, increasing their water solubility, and usually reducing their pharmacological activity (except for pro-drugs). In phase III, further modifications can take place before efflux drug transporters finally eliminate the modified drug from the cell and ultimately the metabolites leave the body through urine or bile. Together, these phases ensure the efficient processing and elimination of drugs, preventing their accumulation to toxic levels and facilitating their safe removal from the body.

In the following, the genes and enzymes involved in drug metabolism are described in more detail. Phase I is mainly comprised of reactions involving cytochrome P450 enzymes. Cytochromes contain heme as a prosthetic group and are involved in various biological processes that include the transfer of electrons such as oxidation and reduction reactions (e.g. electron transport chain). Cytochrome P450 enzymes, often abbreviated as CYPs, are a specific family of cytochromes that play a crucial role in the biotransformation of xenobiotics and endogenous substances [29]. These enzymes are primarily located in the endoplasmic reticulum of hepatocytes and, to a lesser extent, in other tissues. The ”P450” designation comes from the fact that these enzymes absorb light at a wavelength of 450 nm. Cytochrome P450 enzymes are categorized into different subfamilies and isoforms based on their genetic and structural characteristics [30]. The major subfamilies of CYP genes involved in drug metabolism are CYP1, CYP2, and CYP3, which account for the metabolism of about 70-80% of clinically used drugs [31]. A comprehensive understanding of the role and function of cytochrome P450 enzymes is critical in the fields of pharmacology and personalized medicine. For example, tacrolimus, an immunosuppressive drug mostly used after organ transplantation, is mostly metabolized by *CYP3A5*, and depending on genetic variation the achieved dosage can vary considerably between patients [32]. Additionally, drug-drug interactions, as well as dietary compounds, can also influence the activity of cytochrome P450 enzymes. Therefore, these enzymes are important factors to consider in drug development, prescribing medications, and optimizing therapeutic outcomes.

Examples of phase II drug-metabolizing enzymes include several transferases such as glutathione S-transferases (GSTs), UDP-glucuronosyltransferases (UGTs), sulfotransferases (SULTs), N-acetyltransferases (NATs), and methyltransferases. These enzymes catalyze reactions including glutathione conjugation, glucuronidation, sulfation, acetylation, and methylation of a broad range of substrates, including environmental toxins, carcinogens, and certain types of drugs. For example, *UGT1A1* metabolizes the active metabolite SN-38 of the prodrug irinotecan, a topoisomerase inhibitor used to treat several forms of cancer. Patients harboring certain variants in this gene are more susceptible to accumulation of SN-38 resulting in toxicity and guidelines with recommendations about dose reduction have been

developed accordingly [33].

The combined action of both phase I and phase II drug-metabolizing enzymes contributes to the overall clearance of xenobiotics. Additionally, drug transporters are vital for the influx and efflux transport of drugs across cell membranes, playing a crucial role in pharmacokinetics including drug disposition and elimination. Generally, there are 3 superfamilies of drug transporters: solute-linked carrier (SLC), solute-carrier organic anion (SLCO), and ATP-binding cassette (ABC) transporters [34]. These transporters facilitate the passive and active (ATP-requiring) transport of drugs across cell membranes. Variants in drug transporters can impact intracellular drug levels and dosage has to be adjusted accordingly. For example, cellular uptake of statins, a class of drugs commonly used to lower cholesterol levels, is facilitated by *SLCO1B1* and patients carrying increased function alleles have a higher risk of muscle pain and myopathy during treatment [35].

Drug metabolism takes place primarily in the hepatocytes of the liver, which is the major organ responsible for the biotransformation of exogenous substances, including drugs. The liver contains a high concentration of enzymes involved in both phase I and phase II reactions. However, the expression of some pharmacogenes is also present in other tissues and also in cancers to some extent [36–38]. It has been hypothesized that the expression of ADME genes in tumors can contribute to drug resistance by affecting local transport and biotransformation in the tumor. Most evidence to date has been provided about changes in drug transport mechanisms but also the altered activity of drug-metabolizing enzymes in tumors is part of current research efforts. The *ABCB1* and *ABCG2* transporters, also known as multidrug-resistance-protein (MDR1) and breast-cancer-resistance-protein (BCRP) respectively, contribute to cancer drug resistance [39, 40]. These transporters are responsible for efflux transport of anticancer drugs from cancer cells, reducing their intracellular concentration and, consequently, their effectiveness. This process, known as multidrug resistance, can lead to the failure of chemotherapy [41, 42]. Therefore, understanding and potentially inhibiting the function of these transporters is a crucial aspect in overcoming drug resistance in cancer therapy [43]. Other forms of cancer resistance are the mutation of drug targets in the tumor and the activation of compensatory pathways. Genetic variation in drug targets can cause cancer cells to be susceptible or resistant to certain drugs, which is the basis of targeted therapies [44]. Especially in precision oncology, targeting or inhibiting specific proteins or pathways that are driving the growth and survival of a tumor has gained importance in recent years.

Variations in pharmacogenes can range from small variants (SNV, InDels) to large structural events (CNVs, fusions) including duplications and deletions of parts or even whole genes [45]. These can have various functional consequences on the protein level (loss or gain of function), resulting from missense, stop-loss, or splicing variants, deletions, or am-

plifications. The star allele nomenclature has been established to clearly identify and report these functional variants of pharmacogenes [46]. Star alleles are a set of genetic variants that determine a certain functional genotype and can affect how an individual responds to drugs. The star allele naming convention uses standardized combinations of numbers, letters, and a star symbol. For example, the variant rs776746 (6981A>G, splice defect) in the *CYP3A5* gene can be identified as *CYP3A5*3*. This naming convention helps non-specialists to identify and understand these variants more easily and avoid reporting errors. Definitions of star alleles are curated by consortia like the Pharmacogene Variation Consortium[1]. In germline PGx, these functional variants can be translated into several phenotypes of the resulting enzymes or transporters, ranging e.g. from poor to ultrarapid metabolism or decreased to increased transport of drugs. These germline metabolizer profiles have been shown to have a high impact on systemic drug metabolism and thereby also affect response and toxicity [23]. For these phenotypes, there are established guidelines on drugs that are affected, which are developed, for example, by the CPIC (Clinical Pharmacogenetics Implementation Consortium) and DPWG (Dutch Pharmacogenetics Working Group) consortia. Integrating such PGx data into personalized oncology is leading to more effective treatment with fewer adverse effects since many cancer drugs have a narrow therapeutic index [47–49]. ADME pharmacogenomic profiling, when integrated into a molecular tumor board, holds great potential to benefit cancer patients [50–52]. In the context of cancer treatment, where therapeutic options must be increasingly personalized, pharmacogenomic profiling can aid in drug selection and dosage optimization. In this way, oncologists can tailor treatment strategies to optimize drug response and minimize adverse effects. Individuals who may be more susceptible to specific drugs or those who require dosage adjustments based on their genetic profile can be determined ahead of therapy.

In addition to germline variation, there is also somatic tumor-specific variation arising from various endogenous and exogenous causes and mutational processes [53–55]. These somatic variants can generally also occur in pharmacogenes in tumor cells. The focus of pharmacogenomics has so far mostly been on germline variants in ADME genes affecting systemic drug metabolism. However, to comprehensively describe altered drug metabolism as a resistance mechanism in tumors, the tumor-specific somatically acquired variants also have to be taken into account. The relevance of the activity of ADME genes in tumors has often been discussed [41, 56–63], but the exact tumor-specific variations on different omics layers have not yet been comprehensively described. Somatic variations could affect the function and expression of pharmacogenes, thereby influencing drug metabolism and transport in the tumor cells and ultimately influence anti-cancer drug efficacy. For example, the activity of ABC transporters has been shown to increase the efflux of drugs from tumor

[1] `https://www.pharmvar.org/`

cells [42, 58, 64].

Recently, as in many areas of life sciences, pharmacogenomics has been expanded to include further omics layers (usually referring to the entirety of data of certain biological molecules or processes), leading to the emergence of the term pharmaco-omics [65, 66]. Here, not only genetic variation but also the expression and methylation of pharmacogenes, as well as the abundance of drug-interacting proteins and their effect on drug response are being investigated. The insights from pharmacogenomics and ADME research have fundamentally reshaped the pharmaceutical landscape. It has not only improved the efficacy and safety of existing medications but also influenced drug discovery and development. For instance, pharmacogenomic insights can guide the selection of the most promising drug candidates during the early stages of drug development, increasing the likelihood of success in clinical trials.

1.3 The NCT/DKTK MASTER Program

NCT/DKTK MASTER[2] (Molecularly Aided Stratification for Tumor Eradication Research) is a precision oncology program, headed by the National Center for Tumor Diseases (NCT) Heidelberg and The German Cancer Consortium (DKTK), that provides comprehensive molecular analysis and personalized treatment recommendations for patients suffering from rare cancers of any age and for young adults with advanced cancers that exhausted available standard therapies. Rare cancers pose unique challenges due to their limited prevalence and the resulting scarcity of research and treatment options. Advanced cancers are often resistant to established standard therapies, requiring targeted treatment approaches. Molecular diagnostics in MASTER include biomarkers derived from matched control and tumor DNA sequencing (whole exome or genome), tumor RNA sequencing, and array-based methylation profiling. By leveraging such multi-omics data, the program seeks to identify targetable lesions and other molecular features that can inform personalized treatment strategies [67]. Based on molecular biomarkers recommendations are assigned to different treatment baskets including ”tyrosine kinase signaling, PI3K-AKT-mTOR signaling, RAF-MEK-ERK signaling, developmental pathways (e.g. Hedgehog signaling), DNA damage response signaling, cell cycle regulation, and immune evasion” [68]. The complete workflow[3] of MASTER is depicted in Figure 1.1. After patient registration and enrollment, samples of the tumor tissue and peripheral blood are taken and subjected to sequencing. This is followed by molecular profiling using bioinformatics pipelines. Results are then curated and interpreted by clinical

[2] https://www.nct-heidelberg.de/en/research/molecular-stratification/master.html

[3] https://www.nct-heidelberg.de/forschung/molecular-stratification/master/master-workflow.html

bioinformaticians and provided to the oncologists. The multidisciplinary approach of the program, which includes a molecular tumor board, has led to the identification of targetable lesions in more than 80% of patients [69]. Additionally, the program is involved in the development of molecularly stratified clinical trials, further contributing to the advancement of precision oncology. This advanced omics-based research has the potential to uncover the complexities of rare and advanced cancers, which could lead to the development of more personalized treatments and better outcomes.

Molecular data in MASTER is stored in an R data structure called dataMASTER which includes patient metadata and several biological data layers (e.g. germline and somatic small and structural variants, fusions, and expression data) in the form of MultiAssayExperiment or RaggedExperiment R objects [70]. This data structure is highly protected and accessible only from inside the DKFZ network. Currently, it contains molecular data from more than 4600 patients and is regularly updated with new data as patients are continuously enrolled in MASTER.

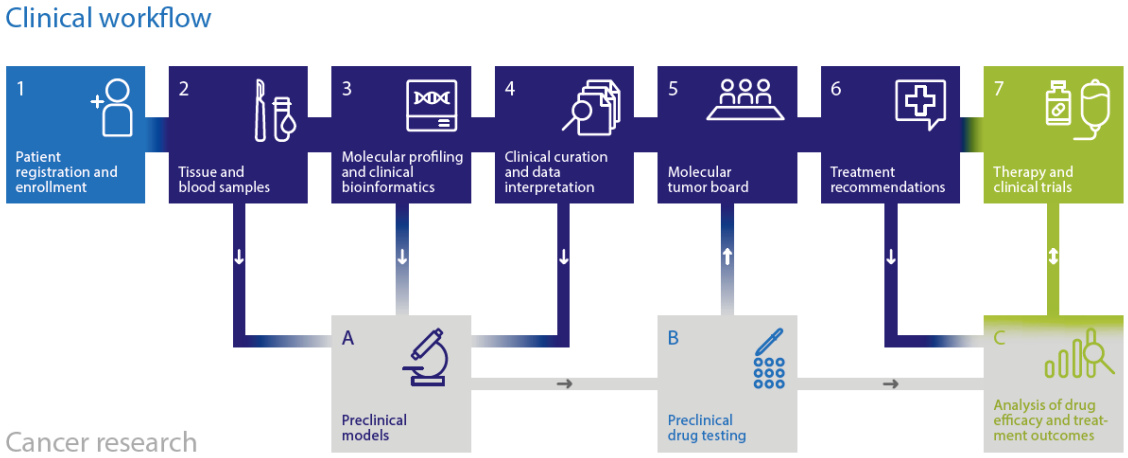


Figure 1.1: Clinical workflow of NCT/DKTK MASTER.
Image taken from NCT website[3].

1.4 Aims of this Thesis

As illustrated in the previous sections, pharmacogenes are essential for the metabolism of endogenous and exogenous substances including xenobiotics like drugs or carcinogens, and therefore play an important role in the development and treatment of cancer. By gaining a comprehensive understanding of germline and tumor-specific pharmacogenomic processes, therapies can be further personalized and tailored to individual cancer patients. For example, known associations between germline PGx variation and drug metabolism can be used to avoid ineffective doses and serious side effects. Furthermore, understanding somatic PGx variation in the tumor can be used to develop more effective therapies that exploit weaknesses or circumvent resistance mechanisms. In this thesis, the NCT/DKTK MASTER cohort is used as an exemplary cohort of a broad range of cancer patients for various germline and somatic PGx analyses. An overview of this thesis is shown in Figure 1.2.
For germline PGx, this thesis aims to establish a computational pipeline for variant identification, genotyping, translation into star alleles, and assignment of phenotypes. This pipeline is applied to the MASTER cohort for a retrospective analysis of known and novel germline variants, their distributions, descriptive statistics across the different tumor entities, and their functional consequences. Furthermore, the integration of germline PGx profiling into the molecular tumor board, providing actionable recommendations for treatment decisions, is a major objective.
For somatic PGx, the research objectives encompass a thorough examination of differences between germline and tumor pharmacogenomic profiles, as well as the comprehensive assessment of somatic variants in pharmacogenes and their functional effects. Further aims are the analysis of methylation and expression in tumors, as well as the association of variants and methylation patterns with expression profiles. Additionally, a major aim is the integration and combined analysis of all data layers (including genomic, epigenetic, and transcriptomic information) to be able to assess the effect of variants and methylation on tumor expression. Through these research objectives, this thesis wants to advance current knowledge of PGx factors, especially tumor-specific effects, influencing the outcomes of cancer treatment.

In summary this thesis includes the following aims:
Germline Pharmacogenomics:

- The development and implementation of a computational pipeline for pharmacogenomic analysis from NGS data, including variant identification, and translation into star alleles, genotypes, and phenotypes. This comprises the evaluation, integration, and harmonization of available algorithms.

- A retrospective pharmacogenomic analysis of cancer patients of the NCT/DKTK MASTER cohort using the implemented pipeline to identify known pharmacogenomic germline variants (including SNV, CNV, and translated star allele genotypes and phenotypes).
- Assessment of the distributions and descriptive statistics of these variants. Examination of the actionable genotypes identified through germline pharmacogenomic profiling and the derived clinical implications from existing PGx guidelines.
- Identification of differences in germline PGx profiles between the cancer entities included in MASTER.
- Integration of germline PGx profiling based on NGS data into the molecular tumor board of the MASTER program. The pipeline has to be adapted to report relevant pharmacogenomic variants of patients. This includes mapping genotyping results to PGx guidelines relevant to administered cancer drugs and supportive medication, as well as deriving recommendations for dose adjustments and warnings about possible toxicities. This also requires the generation of a comprehensive structured report for physicians aiding pharmacogenomics-informed treatment decisions.
- Investigation of rare and novel germline PGx variants that are not covered by current standards or known star allele genotypes. Assessment of the distributions, descriptive statistics, and functional consequences of these variants. For the latter, computational tools for predicting the potential impact on enzyme or transporter are required (Variant Effect Prediction).

Somatic Pharmacogenomics:

- Identification of differences between the pharmacogenomic profiles of the matched germline and tumor samples. This includes analysis of somatic variants (SNVs, Indels, and sCNAs) that are exclusively present in the tumor and are resulting in genotype changes (affecting star alleles). This also shows the limitations when applying germline PGx profiling to tumor tissue and tests the applicability of dedicated somatic variant calling tools, which are however not specifically developed for the PGx domain, to PGx genotyping in tumor samples.
- Comprehensive assessment of somatic variants present in pharmacogenes in the tumor samples (SNC, Indel, sCNAs) and their distributions in the different cancer entities.
- Investigation of the functional effect of these somatic variants in the tumor.
- Investigation of the origin of pharmacogenomic sCNAs and their potential causes.

- Analysis of pharmacogene expression in the tumor entities. This includes analyzing how a subset of the somatic variants affect expression in the tumor.
- Assessment of promoter and intragenic methylation of the pharmacogenes in the tumors and whether there are entity-specific differences. Testing the influence of methylation on expression in the tumor.
- Integration of the different data layers (genomic, epigenetic, transcriptomic) to assess the respective proportion of variance explained in tumor gene expression. Assessment of entity-specific patterns.

In addition to my main project described above, I was involved in several side projects during my time as a doctoral researcher. These included the genomic and transcriptomic analysis of some specific rare cancer entities within MASTER (parathyroid carcinoma, adrenocortical carcinoma, and chordoma). These results are also presented in this thesis in separate chapters. The focus of these projects was the identification of recurrent germline and somatic mutations (SNVs, InDels, CNVs/CNAs, fusions), mutational signatures, and quantification of immune cell admixture in order to expand the current knowledge about these diseases, contribute to their molecular characterization, and find potential rationales for targeted treatments.

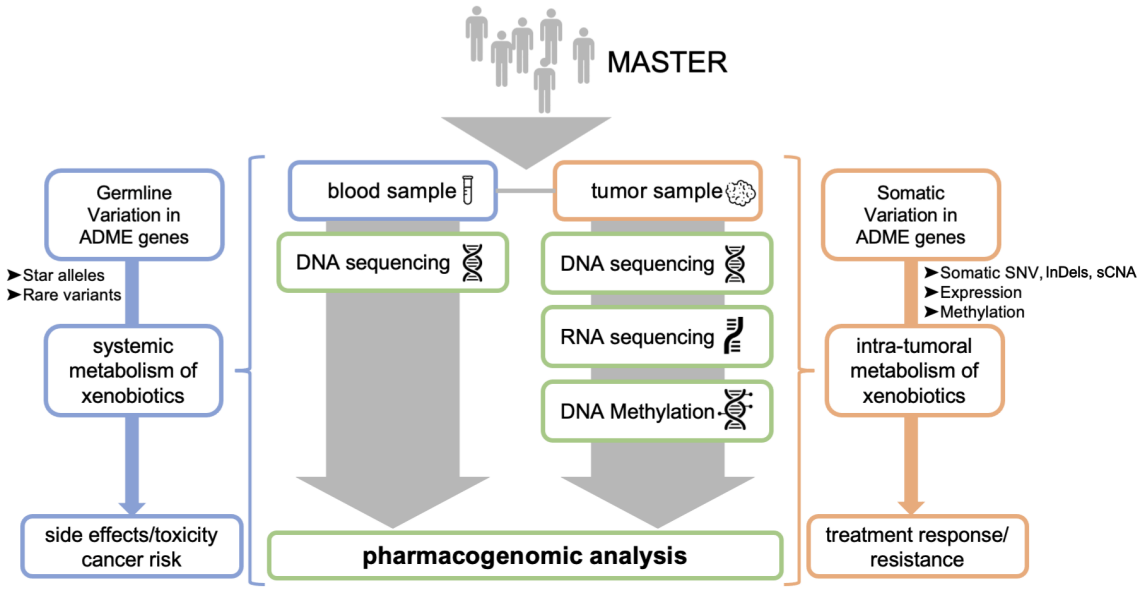


Figure 1.2: Overview of this thesis. Pharmacogenomic analyses were performed based on next-generation sequencing data from matched control and tumor samples of the NCT/DKTK MASTER program (including DNA sequencing, RNA sequencing, and DNA methylation profiling).

Chapter 2

Results

2.1 Overview of the MASTER Cohort

I analyzed the germline and somatic pharmacogenomic profile of 2,371 cancer patients of the MASTER cohort. This is the subset of the total of over 4500 cases for which the DNA sequencing was done with WGS (the rest being WES). The patients in MASTER represent a wide spectrum of cancers, including mainly young adults suffering from advanced stages of common cancers or adult patients of any age with rare cancers [68, 69]. An overview of cancer types and corresponding case numbers is shown in Figure 2.1. A large fraction of the cohort consisted of soft tissue sarcomas (291, 12.5%), neuroendocrine neoplasms (254, 10.8%), and hepatopancreaticobiliary cancers (169, 7.2%). Also, other forms of sarcomas and rare cancers (1.35-5%), as well as rare subgroups of more common cancers such as colorectal (159, 6.7%) or breast are represented (79, 3.3%). The median age of the patients at the time of the molecular tumor board was 48 years with a minimum of 16 and a maximum of 86 years (Figure 2.1). The gender ratio was approximately 50/50 (1178 females, 1193 males). The availability of the various omics datasets of MASTER that I used for this thesis is shown in Figure 2.2.

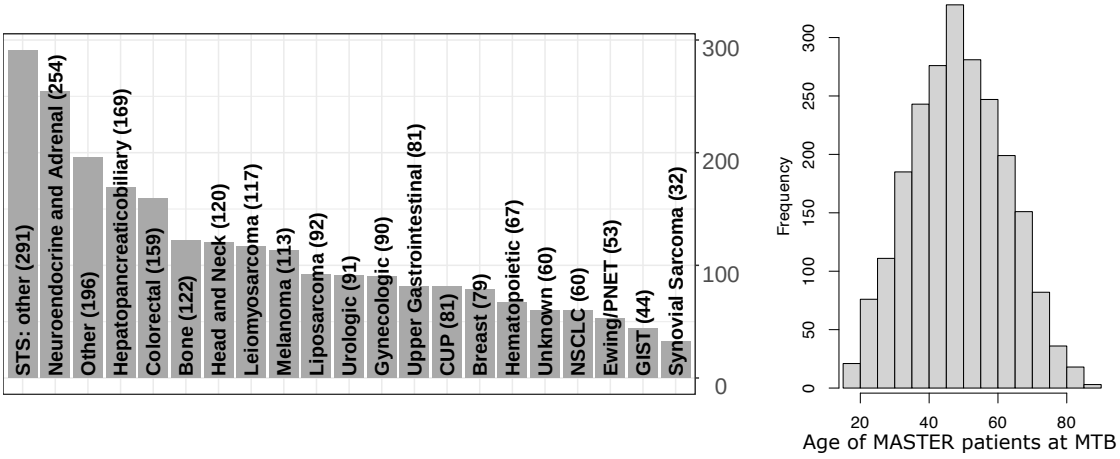


Figure 2.1: Case numbers for cancer entities in the MASTER cohort and age distribution. STS: Soft Tissue Sarcoma, CUP: Cancer of Unknown Primary, NSCLC: Non-Small Cell Lung Cancer, GIST: Gastrointestinal Stromal Tumor.

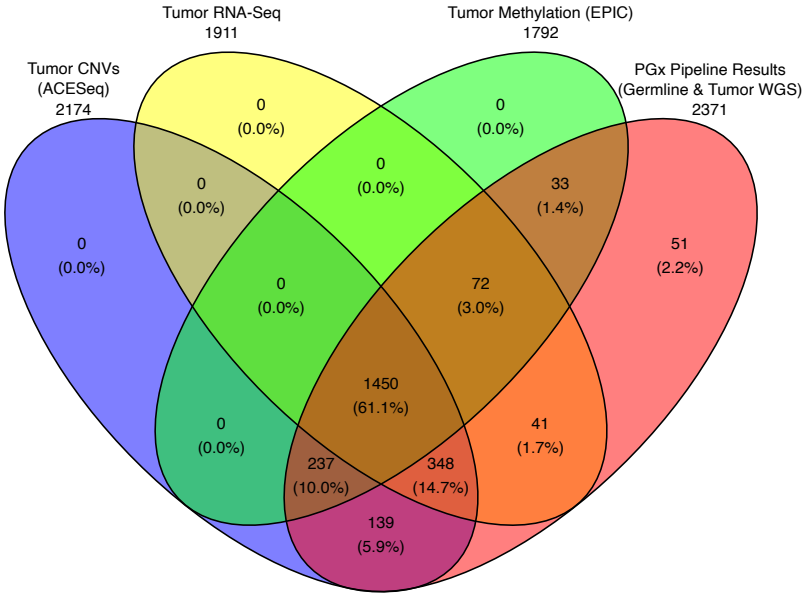


Figure 2.2: Availability of sequencing and PGx data in MASTER

2.2 Pharmacogenomics Analysis Pipeline

To enable the comprehensive and automated analyses of pharmacogenomic variation in large NGS datasets, I developed an in-silico pharmacogenomics pipeline (PGx pipeline) as shown in Figure 2.3 using the workflow management system Nextflow [71]. This pipeline integrates the PGx genotyping tools Aldy [72], Cyrius [73], PyPGx [74] and Stargazer [75, 76]. Details on pipeline implementation, tools, and supported genes can be found in method section 6.1.1 and Table 6.2 in the appendix. To derive PGx results from matched germline and tumor samples of the MASTER cohort, the pipeline was run at several time points until data freeze if a sufficient number of new patients were enrolled in MASTER. The reproducibility of the pipeline was demonstrated by the fact that the results for already included samples

remained stable across runs. In addition, parallelization of the pipeline allowed good scalability for the analysis of larger cohorts. With a selected set of 18 patients for which both WGS and WES of blood samples were available, it was investigated whether the coverage of important functional star allele variants was sufficient with exome sequencing. For many variants, mostly regulatory and intronic variants e.g. in *CYP3A4/5*, *CYP2D6*, and *CYP2C19*, no sufficient coverage could be ensured in the WES samples, as shown in Figure 2.4 highlighted with red rectangles. Therefore, only WGS was used throughout this thesis.

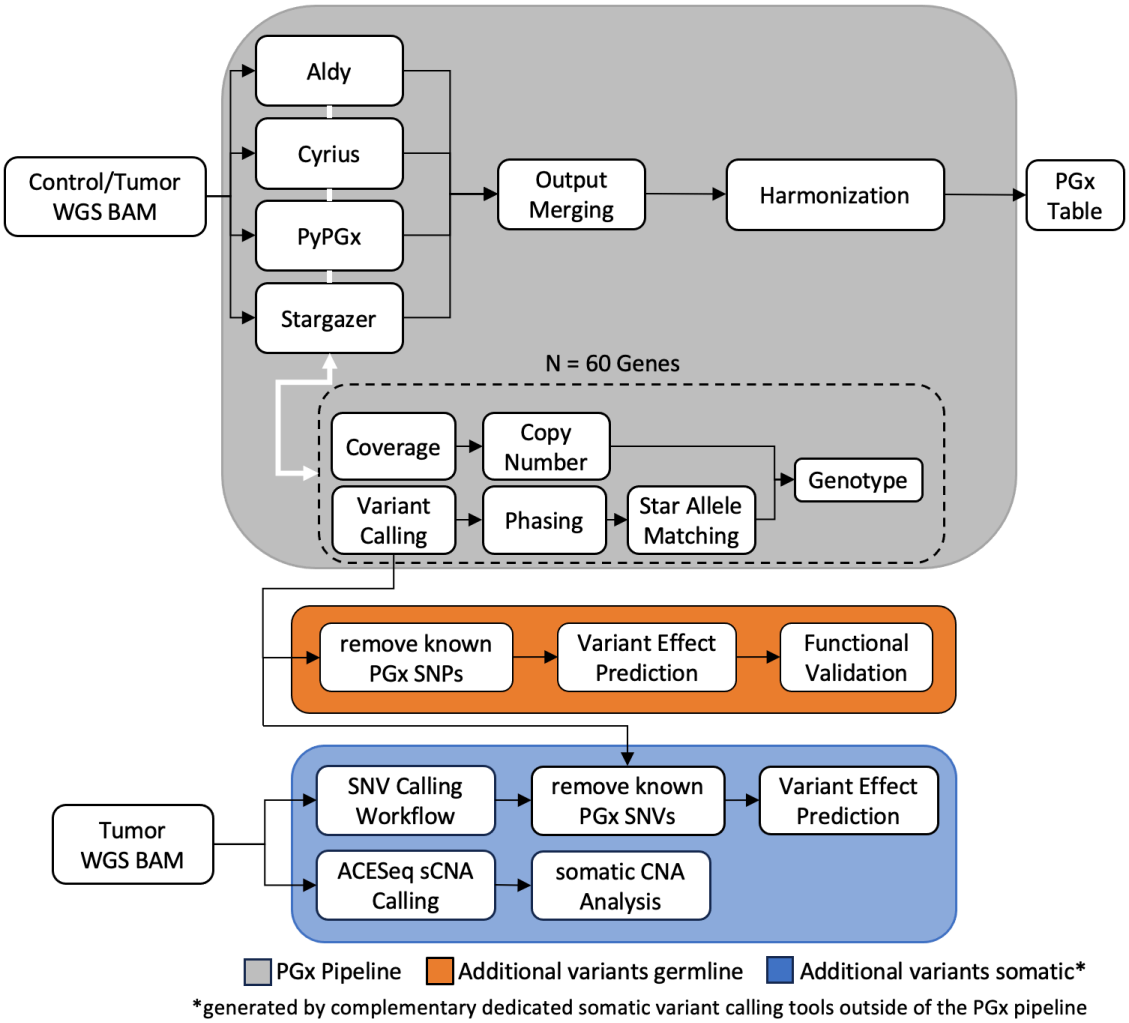


Figure 2.3: PGx pipeline for detection and analysis of pharmacogenomic variants. The core part (grey) includes the PGx genotyping tools and consensus harmonization of results. The pipeline also enables the detection and inclusion of additional germline variants (orange) and somatic SNVs and CNAs (blue).

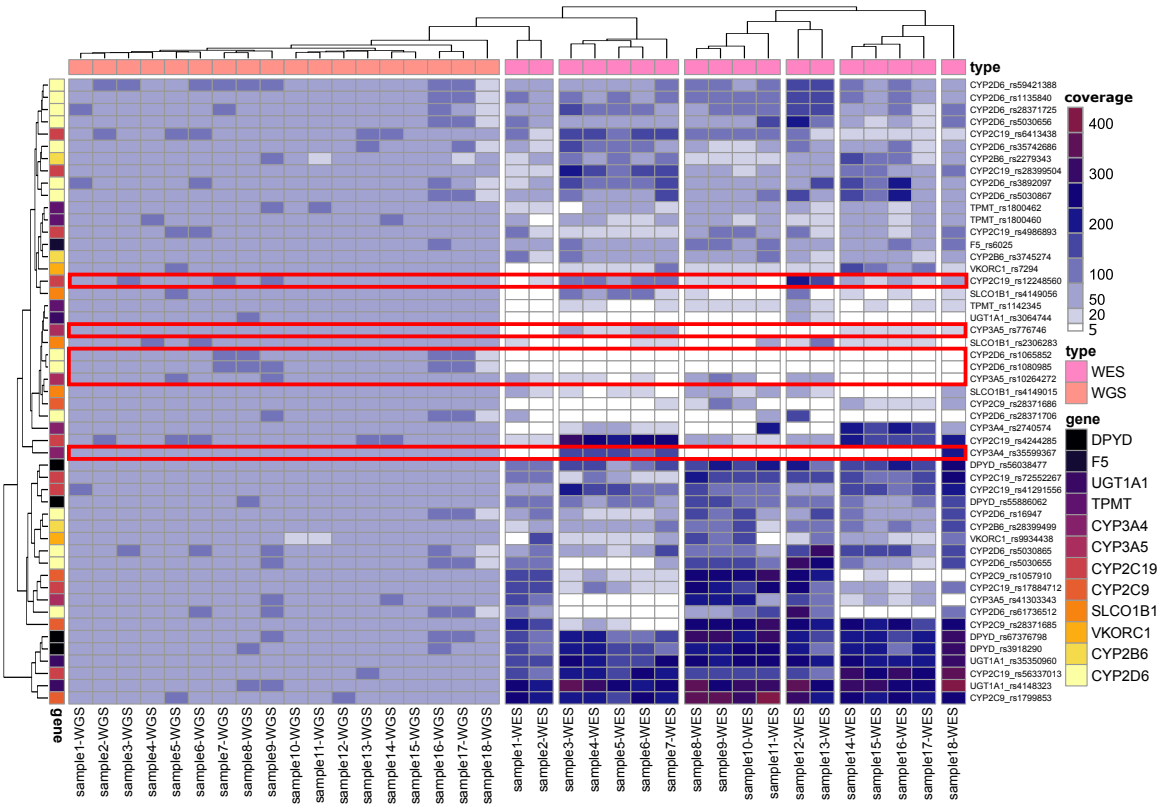


Figure 2.4: Comparison of how many important genomic loci are covered between WGS (left) and WES (right) of 18 patients for important functional PGx variants. Regulatory and intronic variants, like in *CYP3A4/5*, *CYP2D6*, and *CYP2C19* were not sufficiently covered in the WES samples as highlighted by the red rectangles.

Contributions: Sebastian Pirmann assessed the suitability of the genotyping tools and developed the PGx pipeline using Nextflow and additional R and Python scripts. Roman Tremmel and Sebastian Pirmann curated tables for the mapping of variants and star alleles between genotyping tools (see method section 6.1.1).

2.3 Germline Pharmacogenomics in MASTER

Many studies have demonstrated that germline variation in ADME genes impacts the overall metabolism of cancer drugs, affecting both the obtained effective dosage throughout the body and the probability of experiencing side effects [47,77–79]. Therefore, one of the aims of this was to establish a comprehensive picture of pharmacogenomic germline variation in cancer patients by analyzing 2,371 WGS cases of the MASTER cohort. Section 2.3.1 describes the output of the genotyping tools included in the PGx pipeline, and shows their concordance, limitations, and the approach for harmonizing results into a consensus. The consensus germline genotype and phenotype results of the 60 pharmacogenes are described in section 2.3.2. This section mainly deals with already known functional variants (star al-

leles) and their distribution across the different cancer entities in MASTER. Since it was shown that, in addition to small variants, copy number variants also have a major influence on the function of pharmacogenes [80–82], in section 2.3.3 a comprehensive overview of pharmacogenomic germline CNVs detected in the MASTER cohort is given. Section 2.3.4 contains additional rare or novel germline small variants found in the MASTER cohort beyond known star alleles and their predicted functional effect, as it has been reported that rare individual variants account for about 10% of functional variants and can influence individual drug response [83]. Finally, section 2.3.5 describes how pharmacogenomics profiling with the developed PGx pipeline was integrated into the molecular tumor board workflow of the MASTER program for providing treatment recommendations and improving personalized cancer therapy.

2.3.1 PGx Pipeline Results and Harmonization of Consensus Genotypes

First, I have compared the results of the individual tools. The raw results of the four tools were concordant in 93% of the cases on average (Table 2.1). Genes with more than 99% overlap were *CACNA1S*, *CYP1A1*, *CYP1A2*, *CYP2A13*, *CYP2C9*, *CYP2J2*, *CYP2R1*, *CYP2S1*, *CYP3A43*, *CYP4B1*, *CYP19A1*, *CYP26A1*, *G6PD*, *GSTP1*, *IFNL3*, *NUDT15*, *RYR1*, *SLC15A2*, *SLCO1B3*, *SLCO2B1*, *TBXAS1*, *TPMT*, *UGT1A4*, and *VKORC1*. In general, lower concordance was observed for genes with a large number of known variants (e.g. *CYP2D6*, *DPYD*) compared to those for which only very few variants are known and implemented in the tools (*CACNA1S*, *CYP26A1*, *VKORC1*). For a small number of genes, the concordance between tools was generally low including *CYP2D6* (60.1%), *CYP4F2* (73.5%), *GSTM1* (79.4%), *UGT1A1* (87.7%), *DPYD* (88.8%). Among the reasons for these discrepancies were nonidentical naming conventions and different sets of supported variants of the genotyping tools. A major problem of the tools was that some variants could not be phased based on the short-read sequencing input. This led to contradictory star allele results (e.g. *CYP4F2*2, *3, *4*). Additionally, some tools have problems with calling certain variants like *CFTR F508del*, *SULT1A1* heterozygous deletions (only reliably detected by PyPGx), *GSTM1* deletions, or *UGT1A1*28 & *80*. Differences in naming conventions and supported variant sets were partially resolved by defined rules and manual curation. The harmonization process increased the overall concordance to 98%. For most genes, a higher concordance was achieved as shown in Table 2.1. The largest improvement was made for *CYP2D6* from 60.1% to 98%. Not for all genes 100% concordance was reached after curation because in 1.6% of genotype calls, all tools reported discrepant results that could not be resolved. Approximately one-third of these discrepancies are different results in *CYP4F2*

where the tools provide different phasing results and the genotype cannot be reliably determined by short-read sequencing. For one sample, the genotype in *CYP2D6* could not be determined as none of the tools provided an output. In a small subset of patients (n=25), consensus star allele genotypes of 10 selected pharmacogenes were confirmed by Roman Tremmel using a custom Openarray TaqMan panel, as used in [84]. Unfortunately, this small number of samples did not allow a sound evaluation of the accuracy of the pipeline compared to this orthogonal method.

Table 2.1: Concordance of genotyping results between tools before and after harmonization. Concordance was only calculated if at least 2 tools supported the genotyping of the gene.

Gene	Concordance	Concordance (harmonized)	Comment
ABCB1	-	-	
ABCG2	-	-	
CACNA1S	100.0%	100.0%	
CFTR	96.0%	100.0%	Stargazer does not detect all F508del
COMT	-	-	
CYP17A1	-	-	
CYP19A1	99.7%	99.9%	Stargazer does not detect all *4
CYP1A1	99.2%	100.0%	
CYP1A2	99.5%	100.0%	
CYP1B1	98.7%	98.7%	
CYP26A1	100.0%	100.0%	
CYP2A13	99.7%	100.0%	
CYP2A6	93.7%	98.2%	
CYP2B6	94.2%	99.5%	
CYP2C19	85.1%	99.9%	
CYP2C8	97.3%	100.0%	
CYP2C9	99.0%	100.0%	
CYP2D6	60.1%	98.0%	Large number of alleles and differences between implemented variants in tools
CYP2E1	74.3%	95.0%	
CYP2F1	94.0%	99.7%	
CYP2J2	99.3%	100.0%	
CYP2R1	99.6%	100.0%	
CYP2S1	99.7%	100.0%	
CYP2W1	98.5%	100.0%	
CYP3A4	79.7%	100.0%	
CYP3A43	99.4%	100.0%	
CYP3A5	93.7%	100.0%	
CYP3A7	98.7%	99.8%	
CYP4A11	-	-	
CYP4A22	-	-	
CYP4B1	99.4%	99.7%	
CYP4F2	69.9%	73.5%	Due to phasing differences of the tools concerning 2, *3, *4 (unresolvable with short read sequencing)
DPYD	88.8%	88.8%	Discrepant results between all tools
F5	-	-	
G6PD	99.6%	99.6%	
GSTM1	60.6%	79.4%	Discrepancies in deletion calls
GSTP1	99.0%	99.9%	

GSTT1	97.7%	98.4%	
IFNL3	99.8%	99.8%	
NAT1	57.5%	99.0%	
NAT2	97.9%	99.5%	
NUDT15	99.7%	100.0%	
POR	98.5%	99.6%	
PTGIS	-	-	
RYR1	99.8%	100.0%	
SLC15A2	99.4%	99.5%	
SLC22A2	98.1%	98.7%	
SLCO1B1	72.6%	99.4%	
SLCO1B3	99.3%	99.3%	
SLCO2B1	99.8%	99.8%	
SULT1A1	91.0%	97.5%	Only pypgx was able to call deletions
TBXAS1	99.7%	99.7%	
TPMT	99.7%	100.0%	
UGT1A1	85.9%	87.7%	Tools have problems calling *28 and *80
UGT1A4	99.2%	99.5%	
UGT2B15	93.5%	95.5%	
UGT2B17	97.4%	97.6%	
UGT2B7	98.3%	100.0%	
VKORC1	100.0%	100.0%	
XPC	-	-	

2.3.2 Germline Genotypes and Phenotypes of Pharmacogenes

The germline consensus genotypes for the 60 pharmacogenes were determined from WGS data of peripheral blood samples of each patient using the PGx pipeline. While the complete genotyping results and allele frequencies of all genes for the whole MASTER cohort can be found in Table 6.3 in the appendix, the distributions of star alleles for selected pharmacogenes, that are affecting the metabolism of anti-cancer drugs and align with available guidelines (e.g. CPIC, DPWG), are shown in Figure 2.5. The distribution of genotypes of CYPs matched expected population frequencies that have been previously described [85]. 96.4% of patients possessed at least one actionable genotype, allowing for treatment adjustments based on the guidelines. This is in line with previously reported results [86, 87]. The amount of genes with actionable genotypes per patient is shown in Figure 2.6. For most patients, 2 or more genes carried actionable genotypes.

The germline consensus genotyping results were translated into phenotypes for applicable genes (14/60). A complete overview of these results can be found in the appendix (Table 6.4). Depending on the gene, these phenotypes range from poor to ultrarapid metabolism in the case of drug-metabolizing enzymes or poor to increased function for transporters, as well as more specific phenotypes for individual genes like F5 (Coagulation Factor V; favorable/unfavorable response). The distribution of translated phenotypes for the individual

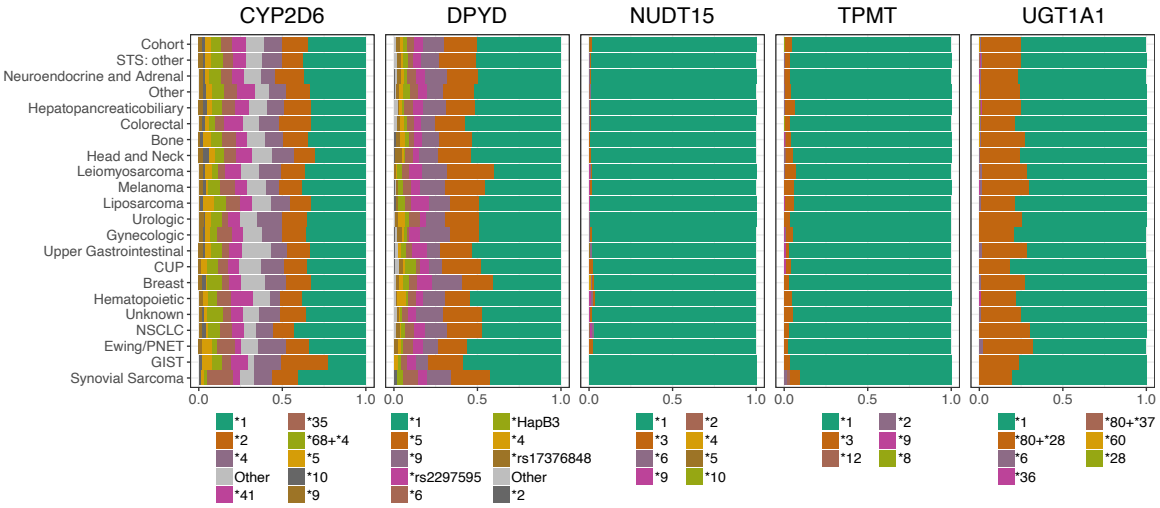


Figure 2.5: Frequencies of alleles for cancer entities in the MASTER cohort. Only selected genes, that are directly related to anti-cancer drug metabolism, are shown here. Genotyping results of the remaining genes can be found in the appendix in Table 6.3.
Relevant cancer drugs for each gene are:
CYP2D6: Tamoxifen; *DPYD*: 5-FU, Capecitabine, Tegafur; *NUDT15*/*TPMT*: Azathioprine, Mercaptopurine; *UGT1A1*: Irinotecan.

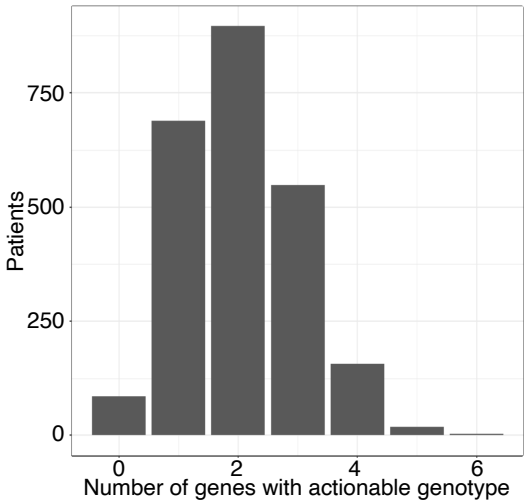


Figure 2.6: Number of pharmacogenes with actionable genotypes per patient. For most patients, 2 or more genes carried an actionable genotype.

cancer entities of the MASTER cohort is shown in Figure 2.7 for selected genes. Small differences in the frequencies were observed between entities and were tested for significance using Fisher tests. Hepatopancreaticobiliary cancers were enriched for normal metabolizers of *CYP3A5* ($p = 0.003$). In contrast, *CYP3A5* poor metabolizers were least frequent in colorectal cancers ($p = 0.005$). Ultrarapid metabolizers of *CYP2D6* were more common in upper gastrointestinal cancers ($p = 0.006$). Intermediate metabolizers of *DPYD* were more frequent in cancers of unknown primary (CUP) ($p = 0.001$). Only the frequency differences of intermediate metabolizers of *DPYD* in CUP remained significant after adjustment for multiple testing.

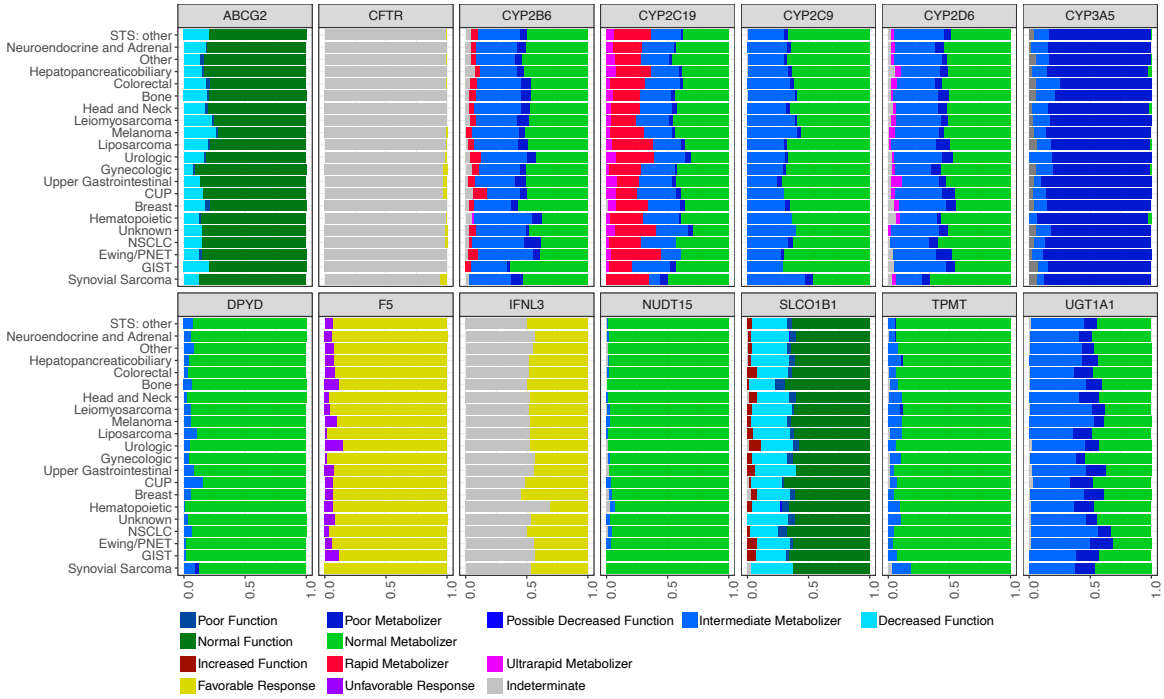


Figure 2.7: Metabolizer and transporter phenotypes of selected pharmacogenes in the MASTER cohort. The colors reflect enzyme phenotypes e.g. poor to ultrarapid metabolizer, transporter activity (increase, normal, and decreased function), and other phenotypes such as F5 (favorable/unfavorable response). On the x-axis, the frequency of the phenotype is shown as percentages in (n=2371 patients). The y-axis shows the 21 different entity baskets of the MASTER cohort.

Contributions: Sebastian Pirmann ran the pipeline, analyzed the data, and created the figures. Roman Tremmel and Sebastian Pirmann curated genotype and phenotype results, computed concordance, and created harmonization and mapping tables. Roman Tremmel ran the genotyping experiments with the Openarray TaqMan panel.

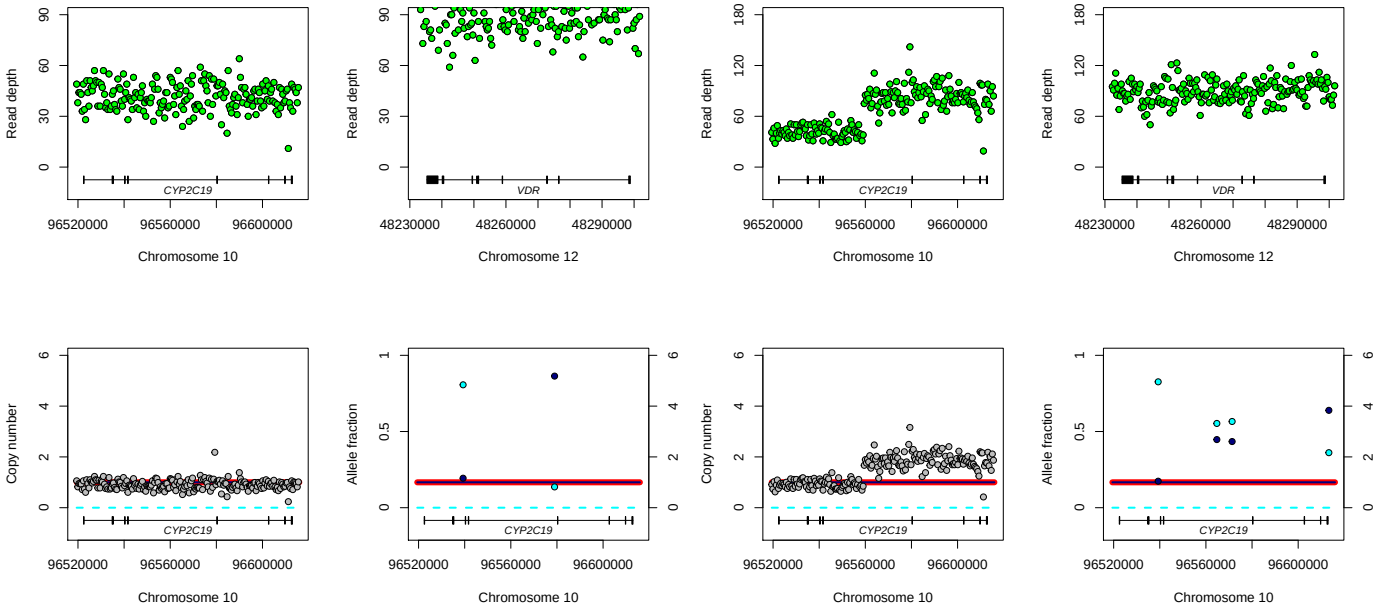
2.3.3 Germline Copy Number Variants in Pharmacogenes

Based on the PGx pipeline results I analyzed germline CNVs for 17 pharmacogenes (*CYP1B1, CYP2A6, CYP2B6, CYP2C19, CYP2D6, CYP2E1, CYP4F2, G6PD, GSTM1, GSTT1, IFNL3,*

SLC22A2, SLCO1B1, SLCO1B3, SULT1A1, UGT1A4, UGT2B15, UGT2B17) including different types of CNVs such as duplications, deletions, and pseudogene hybrids. The frequency results are shown in Table 2.2. Of note, Stargazer can call CNVs for all its supported 51 pharmacogenes, while the remaining tools are restricted to known CNVs. However, only in the reported 17 genes, CNVs of considerable quality were detected by Stargazer. The results for *G6PD*, located on the X chromosome, correctly re-identified male patients by calling heterozygous deletions of the gene. Furthermore, the pipeline identified the most common deletions for the genes *GSTM1*, *GSTT1*, and *UGT2B17* in frequencies as published. Other well-documented frequent deletions were found in *CYP2D6*, *CYP2A6*, and *SULT1A1*, which were in accordance with published data [88]. The deletions in *SULT1A1* were only sufficiently called by PyPGx. In contrast, Aldy could not detect homozygous deletions of *GSTM1*, occurring in 709 patients(29.9%), whereas the other two tools consistently identified these deletions in all cases. Well-known duplications of *SULT1A1*, *CYP2D6*, and *CYP2E1* were detected also with expected frequencies. Hybrids with pseudogenes were called for *CYP2A6/7*, *CYP2B6/7*, and *CYP2D6/7*. Additionally, rare whole-gene and partial deletions were detected including the *CYP2C19* locus, the end of *SLCO1B3* and the whole *SLCO1B1* gene, the whole *CYP1B1* gene, a partial deletion of *IFNL3*, and intronic deletions of *UGT1A4* and *SLC22A2* (Figure 2.8).

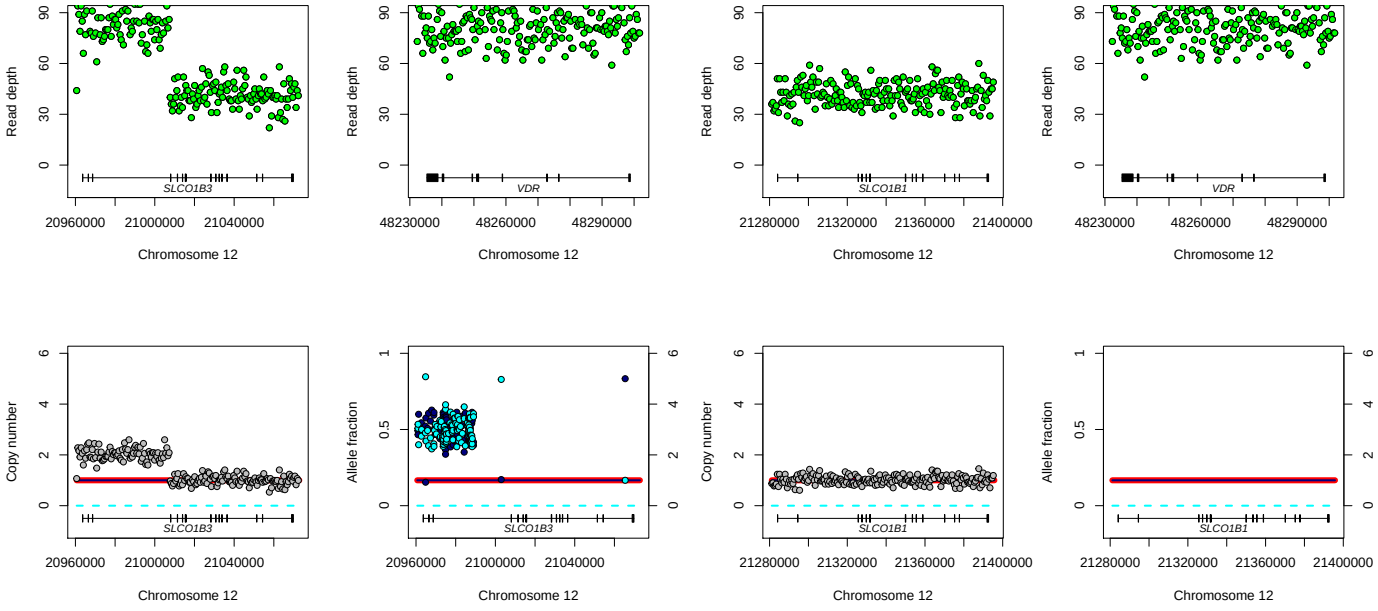
Table 2.2: Germline CNVs in pharmacogenes and their frequencies in the MASTER cohort.
CNVs were detected by genotyping tools in the PGx pipeline.

Gene	Tool	Deletions	Duplications	Hybrids with pseudogenes
CYP1B1	Stargazer	0,04%		
CYP2A6	Aldy, PyPGx, Stargazer	1,7%	0,6%	5,2%
CYP2B6	PyPGx, Stargazer			0,1%
CYP2C19#	Stargazer	0,1%		
CYP2D6	all	6,2%	5,8%	11,9%
CYP2E1	PyPGx, Stargazer	0,2%	4,2%	
G6PD	Aldy, PyPGx, Stargazer	51%		
GSTM1*	Aldy, PyPGx, Stargazer	89,9%		
GSTT1	PyPGx, Stargazer	64,9%		
IFNL3#	Stargazer	0,04%		
SLC22A2#	PyPGx, Stargazer	0,3%		
SLCO1B1	Stargazer	0,08%		
SLCO1B3#	Stargazer	0,04%		
SULT1A1	PyPGx, Stargazer	4,8% ##	31,8%	
UGT1A4#	PyPGx, Stargazer	0,1%		
UGT2B15	PyPGx, Stargazer	0,7%	1,0%	
UGT2B17	PyPGx, Stargazer	54,5%		
*for calculation the sample size was lower (n=1903) due to the exclusion of samples with indeterminate calls				
#including partial deletions of exonic or intronic sequence				
##deletions were only sufficiently called by PyPGx				



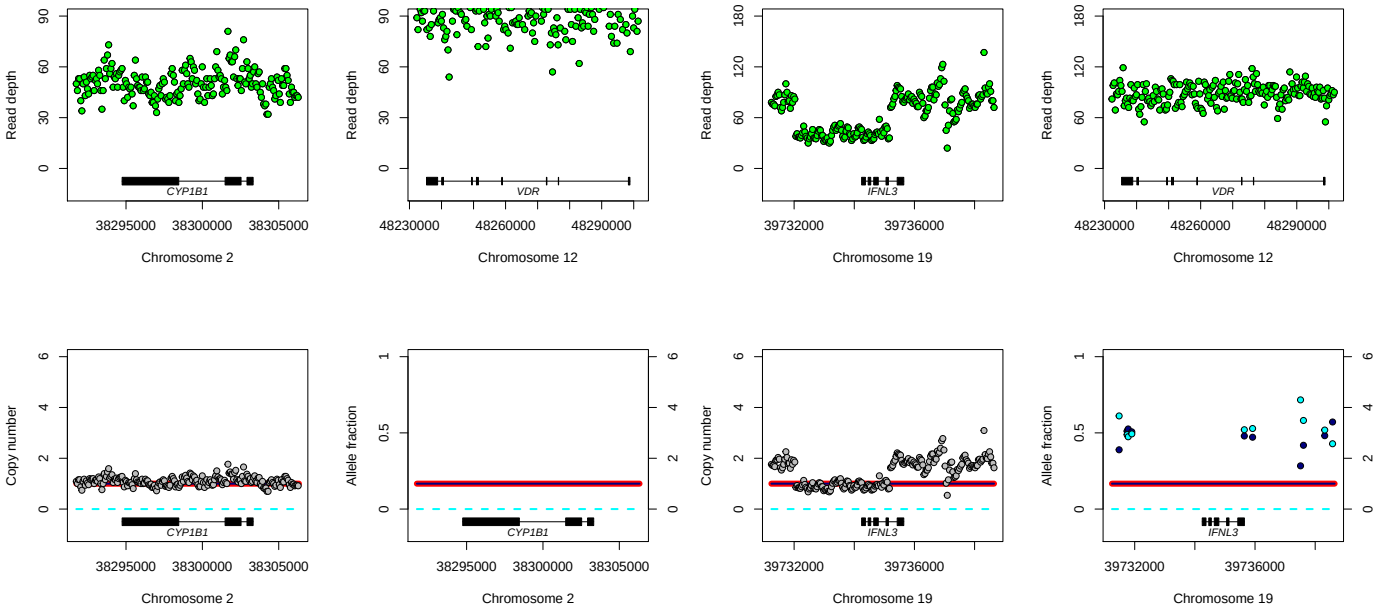
(a) *CYP2C19* whole deletion leading to a copy number of one compared to the reference locus.

(b) *CYP2C19* partial deletion where the beginning of the gene has a copy number of one.



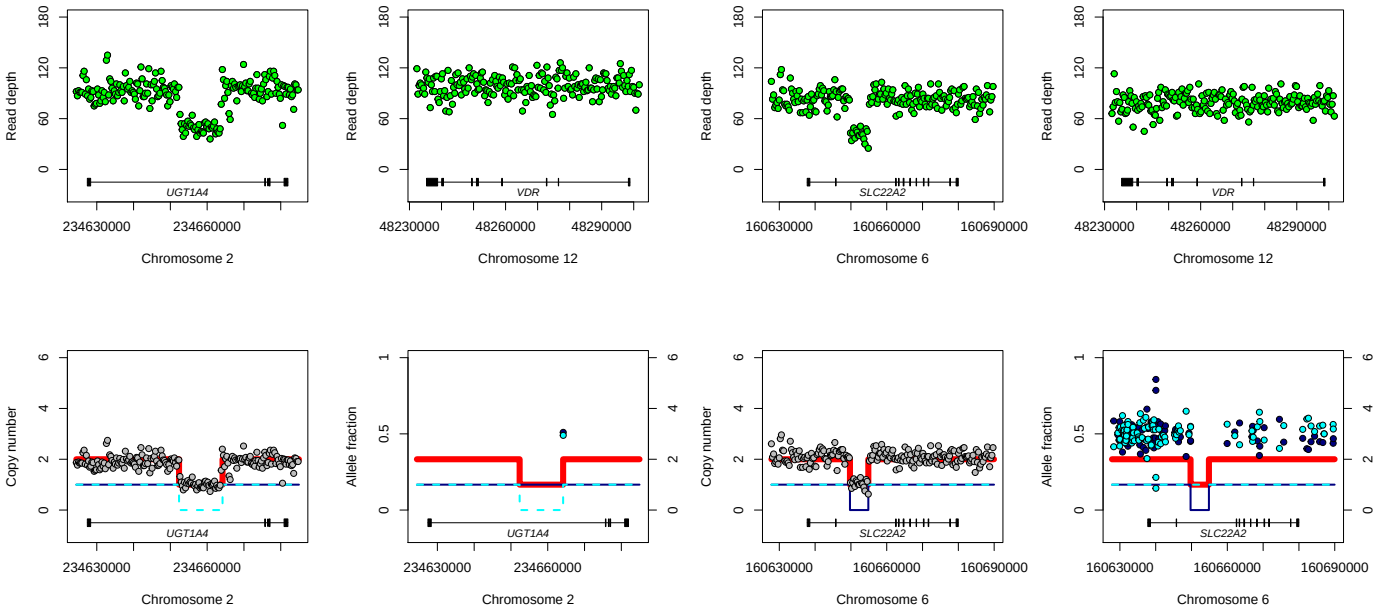
(c) *SLCO1B3* partial deletion leading to a copy number of one in the rear end of the gene.

(d) *SLCO1B1* whole deletion leading to a copy number of one compared to the reference locus. This is the same sample as (c), showing that this deleted segment spans from the rear end of *SLCO1B3* over the whole *SLCO1B1* locus.



(e) *CYP1B1* whole deletion leading to a copy number of one compared to the reference locus.

(f) *IFNL3* partial deletion. The deleted segment starts before the gene and affects about half of it, where the copy number changes to one.



(g) *UGT1A4* partial deletion leading to a copy number of one in the first intron.

(h) *SLC22A2* partial deletion leading to a copy number of one in the second intron.

Figure 2.8: Stargazer CNV plots of rare whole gene and partial deletions affecting *CYP1B1*, *CYP2C19*, *IFNL3*, *SLC22A2*, *SLCO1B1*, *SLCO1B3*, and *UGT1A4*. Per CNV (a-h), 4 panels are displayed. The upper left panel shows read depth (coverage, indicated by green dots) over the genomic position, for the gene of interest (including a gene track with introns and exons). The upper right panel always shows read depth for the reference locus *VDR*, to which the coverage of the gene of interest is compared for CNV calling. The lower left panel shows absolute copy numbers over the genomic position for the gene of interest (red line). The lower right panel displays the allele fractions of PGx SNVs over genomic position for the gene of interest (blue dots), in addition to the aforementioned copy number information. All samples were diploid and the copy number of the reference locus *VDR* was always two.

Contributions: Sebastian Pirmann ran the pipeline and curated and analyzed the CNV data. Roman Tremmel and Sebastian Pirmann assessed the CNV types, frequencies, and stargazer plots of selected variants.

2.3.4 Additional Germline Variants in Pharmacogenes beyond Star Alleles and their Functional Effect

The germline WGS data allowed me to analyze the occurrence of additional germline SNVs in the 60 pharmacogenes beyond the well-known star allele variants. All SNVs in the MASTER germline samples that were not part of common star allele definitions were identified through GATK variant calling and filtered accordingly. In total, 66,813 unique variants were found. Generally, additional variants were detected in all individuals and across all 60 analyzed pharmacogenes, but at varying rates. Genes with the highest total numbers of additional variants in all patients (containing duplicates) included *DPYD* (13,369), *TBXAS1* (4,746), and *RYR1* (4,410) while *UGT1A1* (7), *CYP3A7* (11), and *GSTT1* (18) harbored the least variants.

The 66,813 unique variants were annotated using ANNOVAR. The annotations showed that 60,213 (90%) of the additional variants were located in introns. Descriptive statistics for the remaining 6,582 (10%) variants are depicted in Figure 2.9 and included exonic non-synonymous/missense (1,624; 24.7%), 3' UTR (1,335; 20.3%), downstream (1,132; 17.2%), upstream (1,116; 17%), exonic synonymous (929; 14.1%), 5' UTR (366; 5.6%), exonic start-loss/stop-loss/stop-gain (46; <1%) and splicing variants (34; <1%). The distribution of these results per gene is shown in Figure 2.10. Generally, the variability of numbers and types of additional variants between genes was high. *RYR1* had by far the most unique variants, mainly non-synonymous SNVs. In contrast, *UGT1A1* carried almost no variants. Interestingly, a large fraction of genes also harbored a high amount of 3' UTR variants. On average a patient carried 3,092 additional variants across all 60 investigated genes, with 4,620 variants in the most affected and 760 variants in the least affected patient.

For the additional missense variants (n=1,624), I used an ADME-optimized variant effect prediction (VEP) framework (APF) [91] in addition to established and validated VEP models to determine the impact of these variants on the functionality of the resulting protein in silico. The consensus of the predictions, using optimized thresholds for APF and standard thresholds for all other models, is displayed in Figure 2.11. The standard tools that had the highest overlap with the optimized APF were CADD [92] (86%), PolyPhen2 [93] (80%), MutationAssessor [94] (80%), and FATHMM_MKL [95] (80%). The least consensus with APF was observed for PROVEAN [96] (38%) and FATHMM [97] (47%). In general, there were considerable differences between the predictions of the tools, and the level of agree-

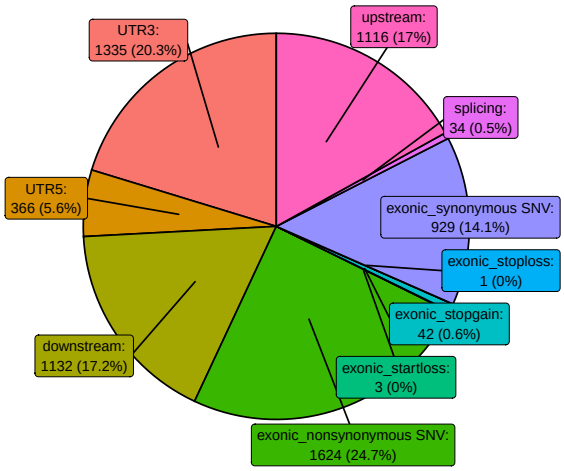


Figure 2.9: Types and frequencies of additional non-intronic germline small variants found in the MASTER cohort. Variants were called using GATK [89] and annotated with ANNOVAR [90]. The missense variants were functionally assessed using the ADME-optimized Variant Effect Prediction Framework (APF) [91]. UTR3=3' UTR, UTR5=5' UTR.

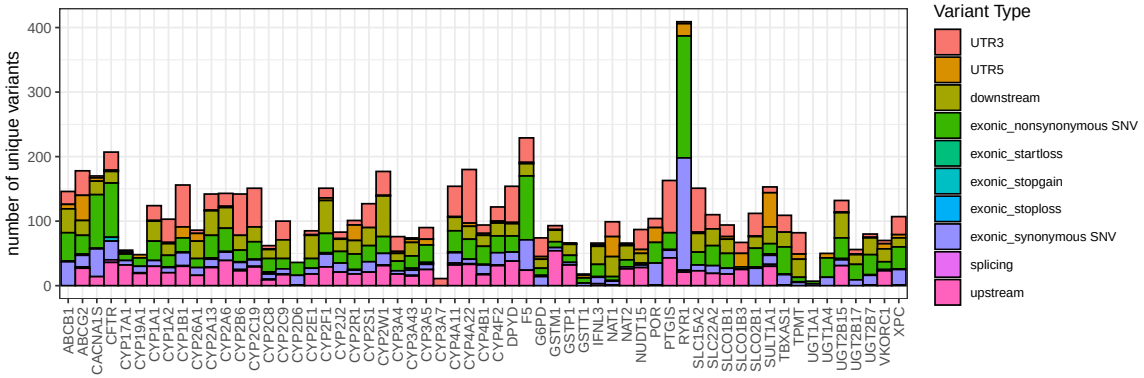


Figure 2.10: Types and frequencies of additional non-intronic germline small variants found in the MASTER cohort per gene. UTR3=3' UTR, UTR5=5' UTR.

ment was rather moderate. About half of the missense variants (856, 53%) were predicted as damaging to the protein function by APF (~1% of all additional germline variants). The complete list of germline variants predicted to be damaging can be found in Table 6.5 in the appendix. The remaining variants (768) were predicted as neutral, i.e. having no impact on the resulting protein. The average population allele frequency for additional non-synonymous variants (both damaging and neutral) across all genes was 0.4% showing that, as expected from previous studies, most of these variants were rare, with 99.9% having minor allele frequency <1% (Figure 2.12).

While the number of predicted damaging variants was modestly, but significantly correlated with the gene length ($R^2 = 0.38, p = 0.0044$), the number of neutral variants was not

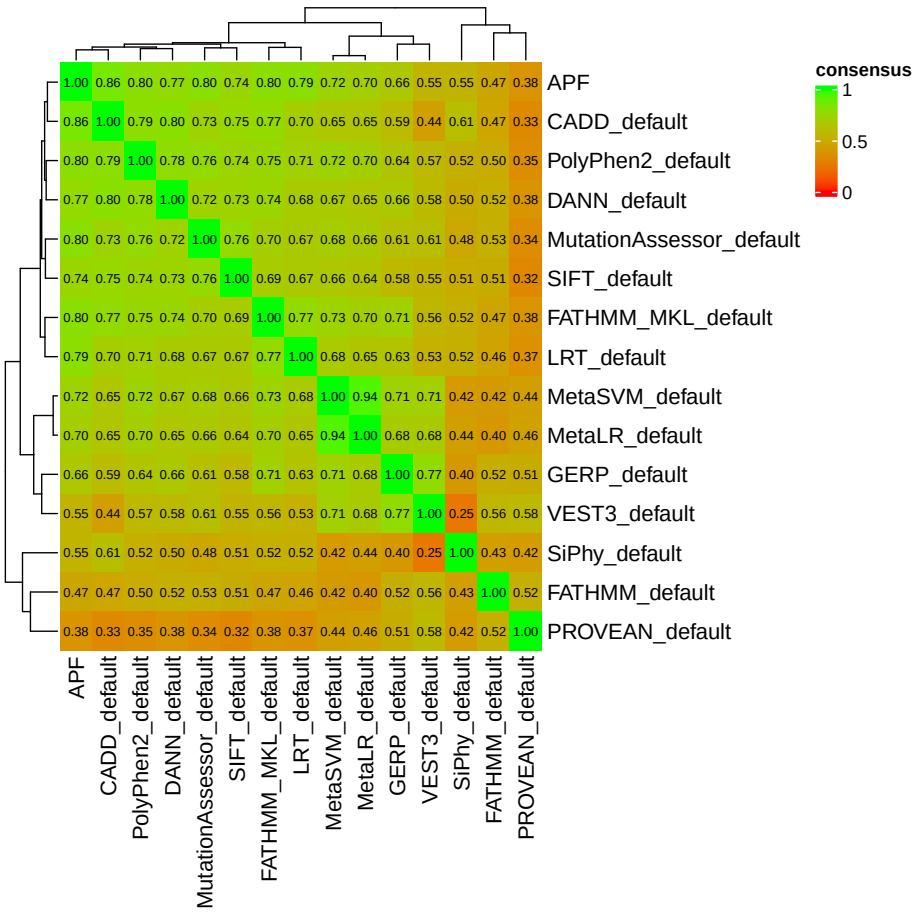


Figure 2.11: Consensus heatmap showing prediction overlap of classical and ADME optimized VEP methods for all germline missense variants. Class prediction only included damaging or neutral. Consensus was calculated based on the Jaccard index of matching predictions.

length-dependent ($R^2 = 0.19, p = 0.15$, Figure 2.13). Most (unique) additional damaging variants were detected in the genes *RYR1* (138), *CFTR* (66), and *CACNA1S* (59). The least affected genes included *CYP3A4* (2), *UGT1A1* (2) and *TPMT* (2). However, relative to gene length, *IFNL3*, *CYP1A1*, *CYP2A13*, and *GSTP1* harbored most damaging variants. It is noteworthy that although *UGT1A1* carried the least amount of additional variants (7), still 2 of them were damaging. Linkage analysis using LDlink [98] showed no linkage between the predicted damaging and known star allele variants which suggests their independent impact on the protein function. Overall 2,066 (87%) of all patients carried at least one damaging variant in one of the 60 investigated genes. On average a patient carried 2.4 additional variants across all these genes, with 12 variants in the most affected and 1 variant in the least affected patients. In total, 2.3% of all damaging variants in all patients and genes occurred homozygous (affecting both copies), while the remaining were heterozygous. In the homozygous case, the damaging variant can lead to a complete loss of function for a gene with potentially strong consequences. To see at which positions the damaging variants are located in the gene and which functional groups of the resulting protein are affected, I created lollipop plots (Figure 2.14). Most interesting was one damaging variant, 3157G>T in exon 7 of *CYP2D6* which results in R329L amino acid substitution, and which occurred in 627 patients (26% of the cohort).

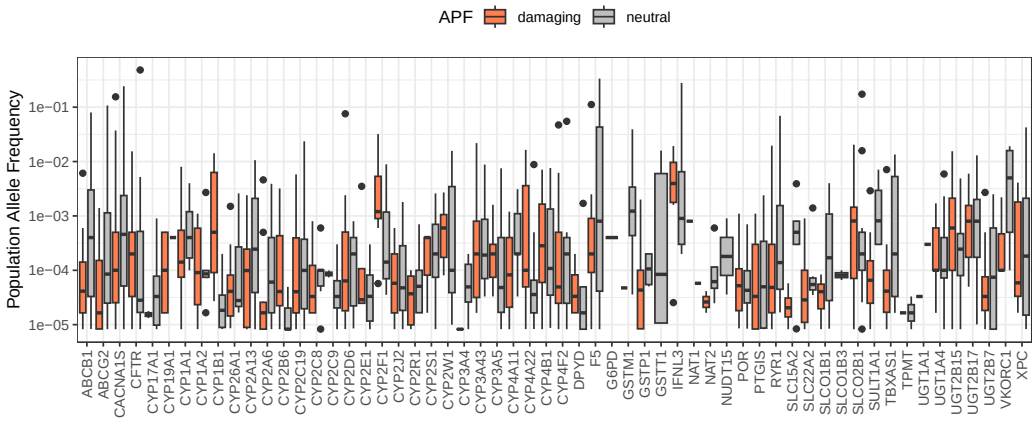


Figure 2.12: Population allele frequencies (ExAC All) for additional non-synonymous germline variants. Frequencies were annotated using ANNOVAR [90]. Colors show prediction of APF [91] (red: damaging variants, grey: neutral variants). The y-axis is logarithmically scaled.

Finally, the APF results for the 856 predicted damaging variants were compared with the predictions of AlphaMissense [99], a recently published method to predict the influence of missense variants on protein fitness incorporating AlphaFold protein structure information. AlphaMissense results[1] were available for 44 of the 60 genes, resulting in data for only 50% of the variants. 35% of these variants were predicted as damaging by both APF and

[1] `https://storage.googleapis.com/dm_alphamissense/AlphaMissense_hg19.tsv.gz`

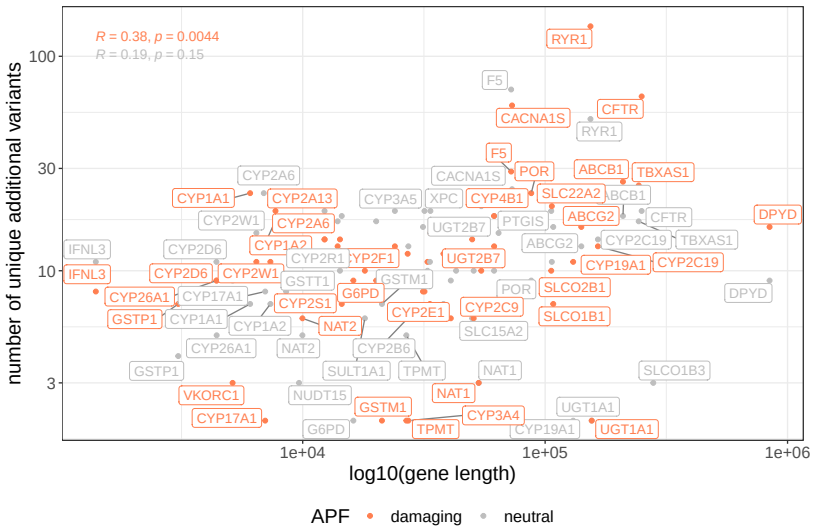


Figure 2.13: Correlation of gene length and number of additional germline small variants. Colors show the prediction of APF [91] (red: damaging variants, grey: neutral variants). The number of variants predicted to be damaging correlated with the length of the genes (Spearman $R^2 = 0.38, p = 0.0044$). The y-axis is logarithmically scaled.

AlphaMissense. AlphaMissense further classified 46% as benign and 19% as ambiguous (Figure 2.15). A subset of carefully selected variants for *SLCO1B1* were subjected to additional functional validation in vitro using transporter assays. The experiments are ongoing and the results are still pending at the time of submission of this thesis.

The predicted damaging non-synonymous variants of APF were then combined with the germline genotyping results of section 2.3.2, to see how these affect the star allele results and metabolizer phenotypes of the patients. When considering patients with extensive, normal, or increased metabolizer phenotype (i.e. having two or more functional alleles of a gene), 109 patients (4.6%) carried at least one of the predicted damaging variants in one pharmacogene which could reduce the actual enzyme activity or transporter function. The most affected gene was *CYP2D6*, in which 49 extensive/ultrarapid metabolizer patients carried one of three functional variants (rs3915951 *R329L*, rs200229206 *N82T*, rs141739595 *R380C*). Further less affected genes included *ABCG2* (22 patients, 13 SNVs), *CYP2B6* (14 patients, 6 SNVs), *SLCO1B1* (10 patients, 5 SNVs), *CYP2C19* (7 patients, 7 SNVs), *UGT1A1* (3 patients, 2 SNVs), and *CYP2C9* and *TPMT* (2 patients, 2 SNVs). In contrast, 629 patients (26.5%) with a decreased or poor phenotype carried additional variants, primarily in *CYP2D6*, *CYP3A5*, and *CYP2B6*. Here the additional variant could either have minimal to no effect if it hits the less functional copy or result in an even stronger loss of function in case the remaining functional copy is affected. Due to short-read sequencing-based limitations in phasing, these two scenarios may not be disentangled in most cases. As a result, the metabolizer phenotype in all these patients might be wrongly assigned with potential clinical implications during drug therapy.

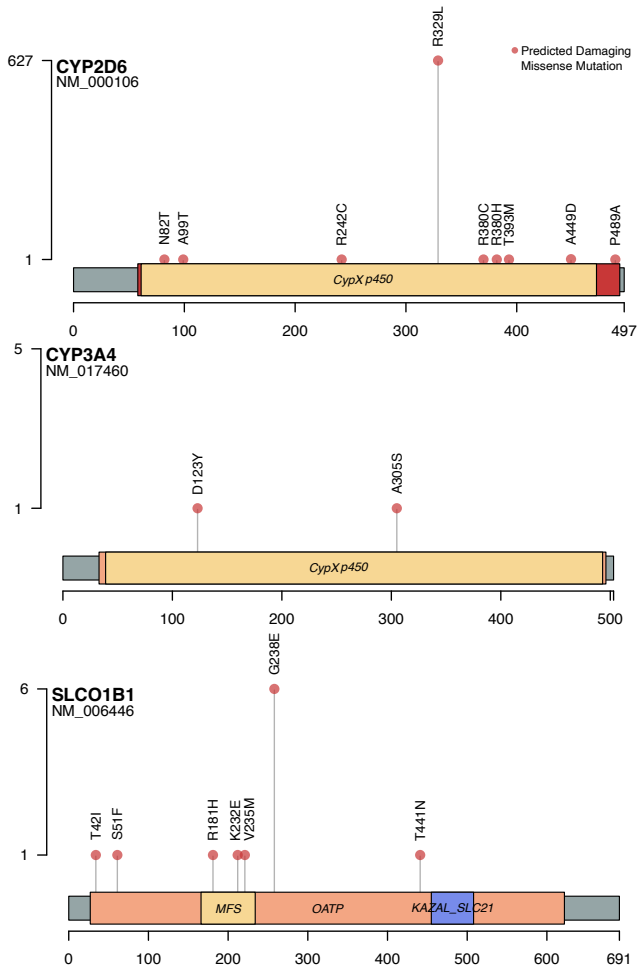


Figure 2.14: Lollipop plots showing positions of predicted damaging variants in *CYP2D6*, *CYP3A4*, and *SLCO1B1*.

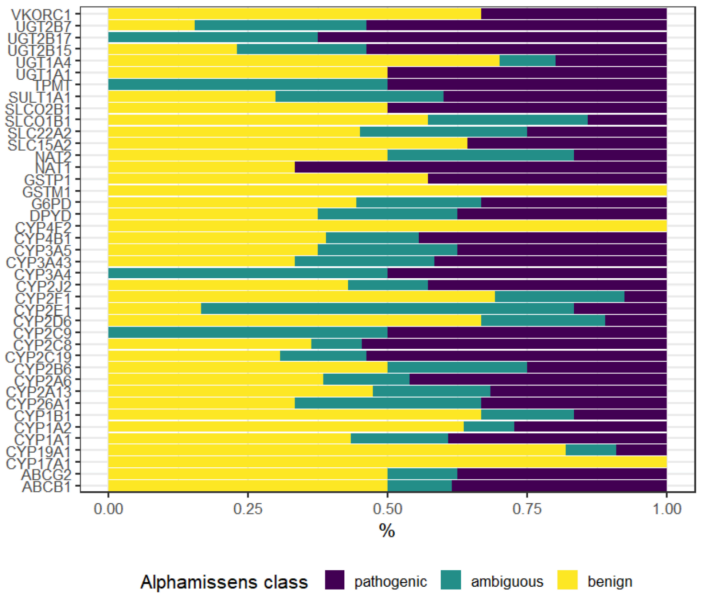


Figure 2.15: Predicted functional effects of AlphaMissense for the 856 variants that were predicted as damaging by APF. Figure created by Roman Tremmel.

Contributions: Sebastian Pirmann analyzed and curated the data on additional germline variants, ran the experiments with ANNOVAR and APF, and created the figures. Yitian Zhou helped with the setup and preparations for APF. Roman Tremmel and Sebastian Pirmann analyzed the AlphaMissense results.

2.3.5 Implementation of Germline Pharmacogenomics Reporting in a Molecular Tumor Board

The PGx pipeline, which was developed as part of his thesis and used for the retrospective analysis of the MASTER cohort, was integrated into the prospective clinical workflow of the molecular tumor board of the NCT/DKTK MASTER precision oncology program. For this purpose, the pipeline is applied weekly to incoming germline WGS samples of patients, and genotypes for a core panel of selected genes (*CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *CYP3A5*, *DPYD*, *F5*, *SLCO1B1*, *TPMT*, *UGT1A1*, *VKORC1*) are are being determined. These are then annotated with CPIC guideline recommendations[2] and uploaded to the knowledge management and decision support system KnowledgeConnector[3] (KC) [100], where they can be assessed by the treating physicians. Figure 2.16 shows a screenshot of the pharmacogenomics page in the KC for an exemplary patient including consensus genotypes and phenotypes, as well as associated recommendations for matching drugs. More than 280 patients have been analyzed so far, however, since this project is in the implementation phase, no statement regarding the application, utility, or any other endpoint can be made yet.

[2] `https://cpicpgx.org/guidelines/`

[3] `https://www.dkfz.de/de/clinical-trial-office/knowledgeconnector.html`

Figure 2.16: Example of PGx tab in Knowledge Connector for molecular tumor board. Personal details have been redacted as this is a real patient. Some of the displayed text is in German since the MASTER tumor board is held in Germany.

Contributions: Sebastian Pirmann adapted the PGx pipeline for integration in the tumor board and wrote a Python script for adding CPIC annotations. Alexander Knurr and Benjamin Roth provided interfaces for the CPIC database lookup and Knowledge Connector upload. Peter Horak provided guidance on the selected set of relevant genes and drugs.

2.4 Somatic Pharmacogenomics in MASTER

This section provides an overview of the somatic variation of drug metabolizing enzymes and transporters in tumors, that could contribute to altered intra-tumor metabolism and drug resistance. First, I assessed the differences in genotyping results of the PGx pipeline between matched germline and tumor samples of the MASTER patients (section 2.4.1). However, since the tools could only detect a small part of the somatic variation, I subsequently used variant calling pipelines optimized for tumor samples to describe the somatic pharmacogenomic profile of the tumors even more comprehensively. Therefore, I have investigated all somatic SNVs and CNAs in the 60 selected pharmacogenes for all tumor samples of the MASTER cohort (sections 2.4.2 and 2.4.3). After showing the complete spectrum of somatic PGx variants, I return to the resulting problems and limitations of the application of PGx genotyping tools to tumor samples, which showed that the genotyping differences described in section 2.4.1 only represent a small fraction of functional changes (section 2.4.4).

2.4.1 PGx Pipeline Results of Tumor Samples

The PGx tools were developed for germline and have not been applied to tumor samples previously. Therefore, I first compared the consensus star allele genotypes between matched control and tumor samples of the MASTER patients. This comparison allowed me to see how much the results between matched samples differ due to somatic variation and how much of this is detectable by the PGx tools. For this, the BAM files of the tumor samples were used as input to the PGx pipeline following the same workflow as in the germline analysis in section 2.3.1. Based on the pipeline outputs, 95.3% of genotypes for all genes were identical between matched germline and tumor samples. However, for 66% of the patients, at least one differing genotype was observed for at least one gene, whereas in 34% of patients, all results were in concordance. Genes with the most differences between tumor and control samples, in more than 15% of the samples, were *UGT2B15*, *SULT1A1*, *G6PD*, and *SLC22A2*, *GSTT1*, and *UGT2B17* (Figure 2.17). In these genes, the genotype differences mainly stem from the fact that the tools detect in the tumor sample a copy number that is different from the germline sample. For example, an altered copy number in the tumor was detected, e.g. in 510 patients in *SULT1A1*. Other interesting differences were

found in *CYP2C9* and *CYP2C19*. Examples of such genotype changes include 46 patients who carried heterozygous star allele *CYP2C19*1/*17* in the germline. In these patients, the genotype changed to either homozygous increased function alleles (*17/*17) or *1/*1 in the tumor. Similar results were observed in 27 patients with germline *1/*2 or *1/*3 germline genotypes of *CYP2C9* changing to homozygous *1/*1. Additionally, in 28 patients, the heterozygous *1/*28 germline genotype of *UGT1A1* changed to wildtype *1/*1. It is important to note that all these genotype differences were solely assessed from the output of the PGx pipeline, and are based on the implemented set of variants. This set includes known germline pharmacogenomic SNVs and CNVs, and possibly neglected superimposed somatic variation. Therefore, I used optimized pipelines to further investigate and comprehensively describe the somatic variation in the pharmacogenes and to assess the limitations of the implemented tools regarding somatic variations. These analyses are shown in the following sections.

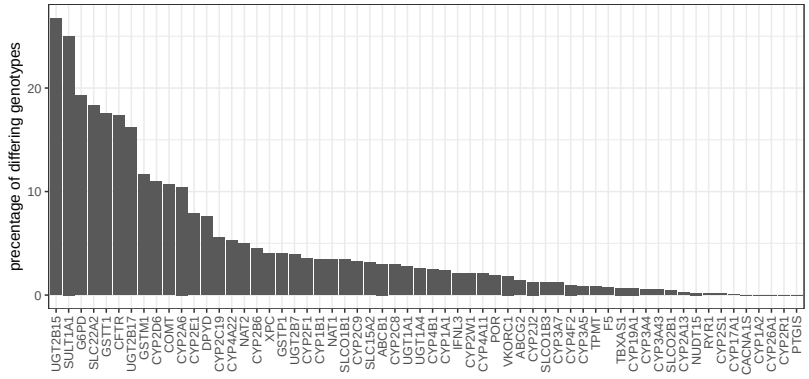


Figure 2.17: Number of differing genotypes (star allele calls) per gene between matched germline and tumor samples.

Contributions: Sebastian Pirmann performed all analyses and created all of the figures.

2.4.2 Somatic pharmacogenomic SNVs

All somatic SNVs in the tumor samples of the MASTER patients were extracted from an established DKFZ in-house small variant calling pipeline [101, 102]. The pipeline removes germline variants in the tumor by subtracting the variants found in the matched control sample. In the whole cohort, 175 unique somatic SNVs were occurring at known germline SNV positions that are part of star allele definitions (hereafter called PGx SNVs), across the 60 analyzed genes in 1% of patients (n=24). The 175 somatic PGx SNVs were detected most frequently in *CYP2D6*, *CYP2B6*, *CYP4A22*, and *SLC15A2*. In 6 of the 24 patients, these variants lead to a star allele diplotype change in the tumor compared to the germline. The reason why not all variants lead to a change is because I did not distinguish between

suballeles that, besides the main functional variant, carried a different set of variants (e.g., *CYP2D6**4.001 or *4.002). In summary, 0.3% of all patients had a metabolizer or transporter phenotype change in their tumor compared to their germline sample based on somatic SNVs occurring at known PGx SNV positions. This shows that these somatic SNVs at known PGx SNV positions explain only a very small proportion of the differences in observed genotypes between the matched samples described above.

In addition to the somatic PGx SNVs, there were 22,271 additional exclusive somatic SNVs in the 60 analyzed genes in total. As illustrated in Figure 2.18, 20620 (93%) were intronic, and the remaining 1561 (7%) were non-intronic variants including exonic non-synonymous (635, 38.5%), exonic synonymous (375, 22.7%), upstream (283, 17.1%), downstream (255, 15.4%), 3' UTR (54, 3.3%), 5' UTR (19, 1.2%), exonic start-loss/stop-loss/stop-gain (28, 1.7%) and splicing variants (2, <1%) as annotated using ANNOVAR. In comparison to the corresponding distribution in germline variants, considerable differences were found for exonic non-synonymous SNVs (40% vs. 23%) and 3' UTR (4% vs. 20%). Distributions of variant types per gene are displayed in Figure 2.19.

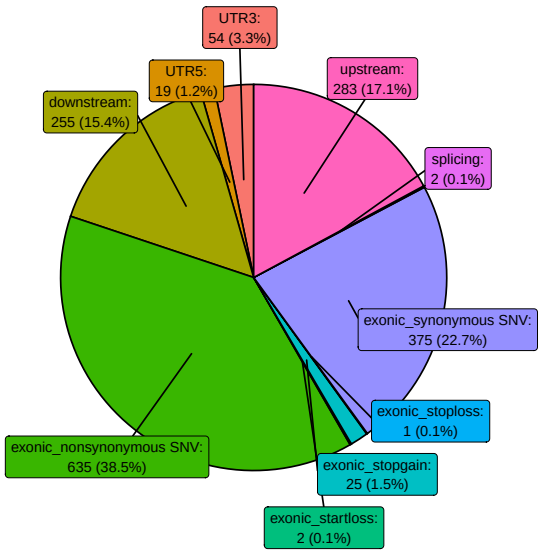


Figure 2.18: Types and frequencies of additional non-intronic somatic variants found in the MASTER cohort. Variants were extracted from OTP SNV Calling Workflow and annotated using ANNOVAR [90]. Missense variants were further functionally assessed using the ADME Variant Effect Prediction Framework [91]. UTR3=3' UTR, UTR5=5' UTR.

Similar to the germline, 54% of the additional non-synonymous/missense variants were predicted as functional by APF [91]. In contrast to the germline, the additional somatic SNVs were correlated to gene length for both neutral ($R^2 = 0.27, p = 0.047$) and damaging variants ($R^2 = 0.47 p = 0.00048$); however, the correlation for damaging variants was stronger. The number of damaging variants found in the germline was higher than in the tumor for most genes, as shown in Figure 2.20. Figure 2.21 summarizes the comparison of germline

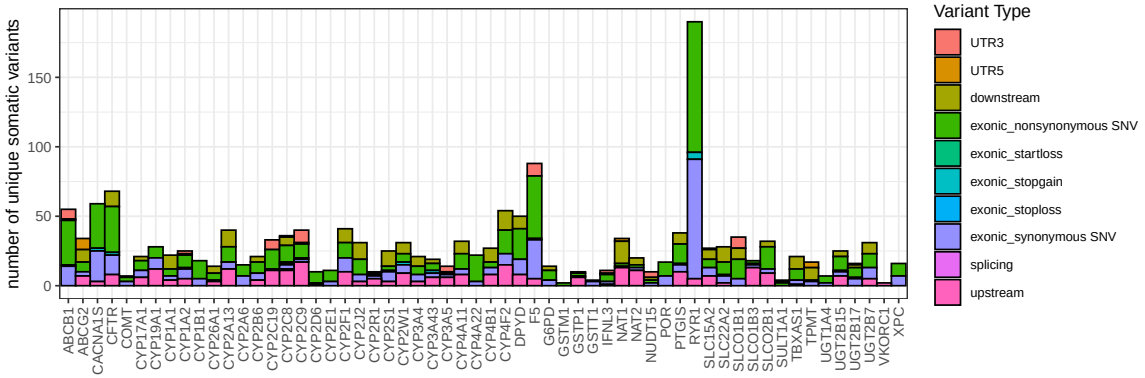


Figure 2.19: Types and frequencies of additional non-intronic somatic small variants found in the MASTER cohort per gene. *RYR1* harbors most somatic variants, both synonymous and non-synonymous SNVs. UTR3=3' UTR, UTR5=5' UTR.

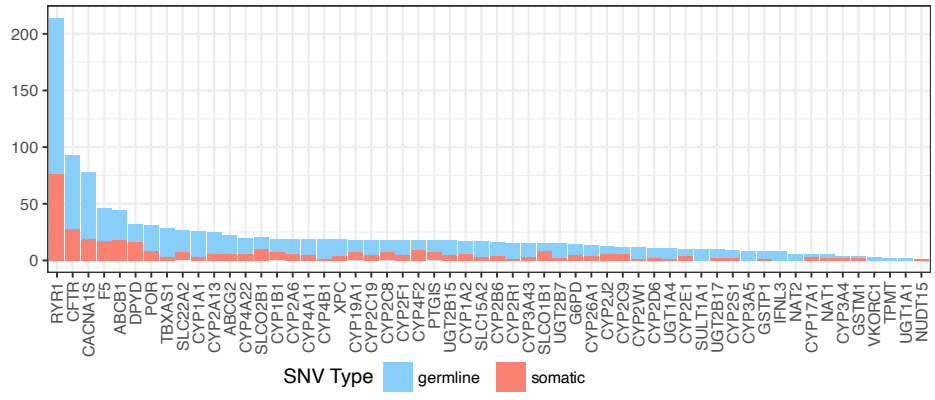


Figure 2.20: Numbers of predicted damaging germline and somatic SNVs per gene. For most genes the fraction of damaging germline variants was higher.

and somatic small variants in the MASTER cohort for the 60 analyzed pharmacogenes. It shows the amount of known and additional SNVs found in germline and tumor samples, their type, functional prediction (for exonic-nonsynonymous variants), and their distribution across gene regions. Out of the 2,603 PGx variants that are supported by the genotyping tools and are part of star alleles, 1007 were found in the germline and 175 additionally in the tumor samples. Compared to the implemented variants, a large number of additional variants were found. While for these additional variants, the ratio between intronic and exonic variants was approximately equal, germline variants were more numerous in absolute terms. In the tumor, an enrichment of non-synonymous variants was observed while the ratio of predicted damaging variants stayed similar.

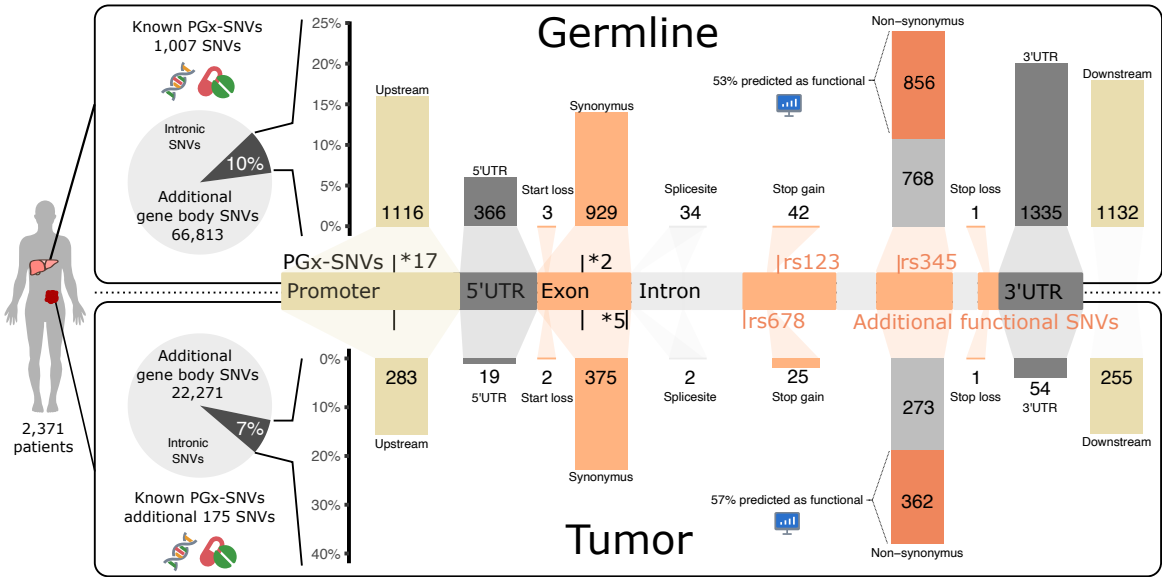


Figure 2.21: Comparison of germline and somatic SNVs found in the pharmacogenes for the MASTER cohort. Includes known PGx variants and additional variants with functional prediction in germline and tumor. Figure created by Roman Tremmel.

Contributions: Sebastian Pirmann performed all analyses and created most of the figures. Roman Tremmel created Figure 2.21.

2.4.3 Somatic Pharmacogenomic Copy Number Aberrations

Somatic copy number aberrations (sCNAs), defined as chromosomal gains and losses of varying sizes, are known to play a major role in the origin and progression of cancer [103, 104] and have been suggested to promote drug resistance in the context of pharmacogenes [105, 106]. Therefore, I investigated the sCNA profiles in the 60 pharmacogenes. sCNA calling results were available for 2,174 patients from the DKFZ in-house pipeline ACESeq [107]. This pipeline calls sCNAs by segmenting the WGS based on coverage as well as B-allele frequency and subsequently merges neighboring segments with the same number

of allele-specific copies. In addition, an estimate of the combination of base ploidy and purity (tumor cell content) of the tumor sample is obtained by fitting copy number states to integer numbers. The average tumor ploidy was 3 (ranging from 1.4 to 6.5), showing some extent of cancer-specific aneuploidy [108–110]. Only sCNA segments overlapping with pharmacogenes were included in the following analyses.

In total, I found 63,536 sCNA events across 25,971 unique segments, including amplifications, duplications, deletions, and loss of heterozygosity (LOH) affecting pharmacogenes. The segments of these sCNA events covered a wide range of sizes from 1 kilobase to 138 megabases (Figure 2.22). 95.5% of sCNA segments were relatively large in the size ranges of chromosome arm-level events ($> 10^6$ base pairs). Figure 2.23 shows the total number of sCNAs of the 60 pharmacogenes in the MASTER cohort per sCNA category. The 63,536 CNVs included 25,689 duplications, 20,604 deletions, 15,929 LOH events, and 1,314 amplifications. This number of sCNAs is considerably higher than the SNVs presented in the previous section and involves far more base pairs, indicating that sCNAs contribute substantially to somatic variation in pharmacogenes.

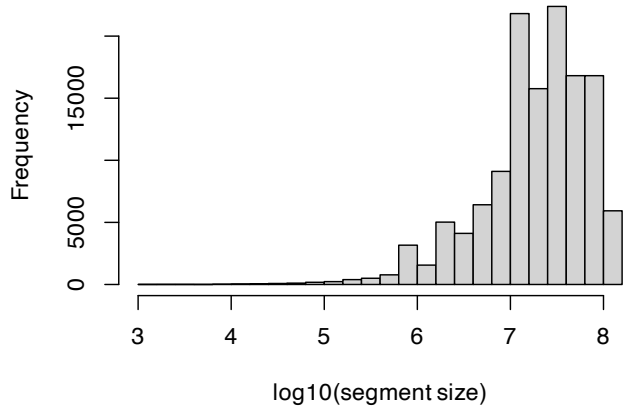


Figure 2.22: Histogram of segment sizes of sCNAs affecting pharmacogenes.

As shown in Figure 2.24, the genes could be separated into 3 groups which were enriched for deletions, duplications, or equally affected by both at the cohort-wide level. Additionally, loss of heterozygosity (LOH) occurred very frequently across deleted genes but was also found in duplicated genes to a lesser extent. Generally, the top three most sCNA-affected genes were *NAT1/2* and *NUDT15* which were mostly deleted and affected by LOH. The least sCNA-affected genes were *SULT1A1*, *GSTP1*, *CYP1B1*, and *VKORC1*. Most deletions were found in *NUDT15* (706), *NAT1* (670), and *NAT2* (658), whereas *PTGIS* (92) and *F5* (99) were the least affected by deletions. In contrast, most duplications were found in *PTGIS* (841), *CYP2W1* (808), *F5* (785), and *POR* (782). Genes with the lowest number

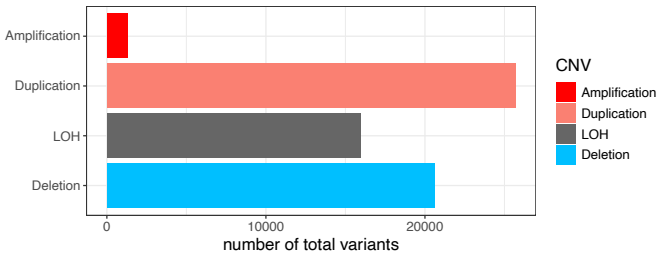


Figure 2.23: Total numbers of sCNA events in the 60 pharmacogenes in the MASTER cohort. SCNAs in all 60 pharmacogenes for the whole MASTER cohort were extracted from the ACESeq pipeline [107].

of duplications included *CYP26A1* and *CYP2C8/9/19* (165 each). The highest number of amplifications was detected in *SLCO1B1/3* (72). These findings are in line with pan-cancer-wide observations of recurrent gains and losses of chromosome segments and explain that various pharmacogenes are susceptible to either duplications or deletions [104]. Examples include deletions on chromosomes 1p (*CYP2J2*, *CYP4B1/4A11/4A22*, *DPYD*, and *GSTM1*), 3p (*XPC*), 4q (*UGT2B7/15/17* and *ABCG2*), 8p (*NAT1/2*), 10q (*CYP2C8/9/19*, *CYP2E1*, *CYP17A1*, and *CYP26A1*), 13q (*NUDT15*), 22q (*COMT*, *GSTT1*, and *CYP2D6*), or duplications on chromosome 1q (*CACNA1S* and *F5*), 7pq (*CYP2W1*, *CYP3A4/5/7/43*, *ABCB1*, *CFTR*, *POR*, and *TBXAS1*), 12p (*SLCO1B1/3*), 20q (*PTGIS*). Pharmacogenes that were least affected by sCNAs (*CYP1A1/2* and *UGT1A1/4*) were located in regions 2q and 15q where only very few pan-cancer-wide gains and losses were observed [104].

The analysis of sCNA events was also stratified by cancer entities, results of which are shown in the heatmaps in Figure 2.25. For each sCNA type, cancer entity, and gene, the fraction of affected patients was calculated (scaled percentages in the heatmaps). The fraction of sCNA-affected patients per cancer entity and gene was up to 80% for duplications and deletions, 20% for amplifications, and almost 100% for LOH in some entities and genes. The genes that were highly duplicated across many entities included *F5* and *CACNA1S* (1q), *PTGIS* (20q), and mainly genes located on chromosome 7 including *CYP2W1*, *POR*, *ABCB1*, the *CYP3A* family, *TBXAS1*, and *CFTR*. Colorectal and neuroendocrine cancers were among the entities with the highest fractions of duplicated genes, whereas GIST, hematopoietic cancers, and synovial sarcomas harbored the least duplications. Notable cases in which individual entities and genes differed from the rest of the cohort were an enrichment of duplications of *NUDT15* in colorectal cancer, *TPMT* in melanoma, and *NAT1/2* in Ewing sarcoma/PNET. In line with the fact that they represent a stronger alteration, amplifications were generally less common than duplications; entities mostly affected were breast and urologic cancers. Breast cancers had the most frequent amplifications across several genes, but especially in *SLCO2B1*, *XPC*, and *GSTP1*. *SLCO1B1/3* were frequently amplified in urologic cancers. In addition to the frequent duplications of *NUDT15* in colorectal cancer, am-

plifications of *NUDT15* as well as *PTGIS* were also enriched in this entity. In summary, the sCNAs in *NUDT15* for colorectal cancers differed greatly from the other entities in which *NUDT15* was generally deleted. As shown below in section 2.5.1 (Figure 2.32), *NUDT15* was also expressed the highest in colorectal cancers. For deletions, the most affected genes were *NUDT15*, genes located on chromosome 10q including *CYP2E1*, the *CYP2C* family, *CYP17A1*, *CYP26A1*, and *NAT1/2* (8p). In total, leiomyosarcomas, GIST, colorectal, and breast cancers had the most deletions across many genes. In contrast, hematopoietic cancers showed the least deletions. Some entity-specific enrichments were deletions of *NAT1/2* in colorectal and breast cancers and *SLC22A2* in melanomas. LOH was mainly observed in the same genes and entities that were enriched for deletions. Examples included GIST and leiomyosarcoma, as well as breast and bone cancers. Interestingly, LOH was also found very frequently in many genes in the neuroendocrine cancers, which, however, were not as strongly affected by deletions suggesting copy number-neutral LOH.

For many tumor entities, the whole genome-wide sCNA profile has already been described and matches some of the entity-specific patterns observed here. For example, frequent sCNAs, including deletions and LOH of many chromosomal segments were previously found in leiomyosarcoma [111], especially on chromosome 10q which includes the genes *CYP2Cs*, *CYP2E1*, *CYP17A1*, and *CYP26A1*, which were also found to be frequently deleted in this entity in this work. However, it has to be noted that the cohorts described in [111] and here to some extent contain the same samples. Also, frequent deletions of the p arm of chromosome 1 have been reported in GIST [112–115], which includes the frequently deleted *CYP4B1*, *CYP4A11*, *CYP4A22*, *CYP2J2*, *DPYD*, and *GSTM1* found in this entity. The frequent duplications located on chromosome 7q (*POR*, *CYP3A4/5/7/43*, *CFTR*, and *TBXAS1*) in colorectal cancer match known chromosomal gains of this region [116–118]. The role of LOH of *CYP2D6* in breast cancer has been described repeatedly by previous studies [119–122] and additionally, LOH was found in many other pharmacogenes where deletions were observed in breast cancer like chromosome 8p which includes *NAT1/2* [123]. These are only some non-exhaustive examples where the observed sCNAs of pharmacogenes follow the reported genomic sCNA profile of the respective cancers. In summary, some of the pharmacogenomic sCNA events appear to be common at pan-cancer level, while others are related to entity-specific genomic patterns.

It is known that tumor cells carry complex structural aberrations in their genome. In particular, oncogenes and tumor suppressor genes that promote tumor development [124] and may thus have characteristics of driver genes are often affected by sCNAs. As already shown in the previous analyses (Figure 2.25), for some cancer entities the sCNA profile of pharmacogenes follows the same chromosomal patterns that have been described in these individual entities, which could be linked to specific driver genes. Furthermore, as 95.5%

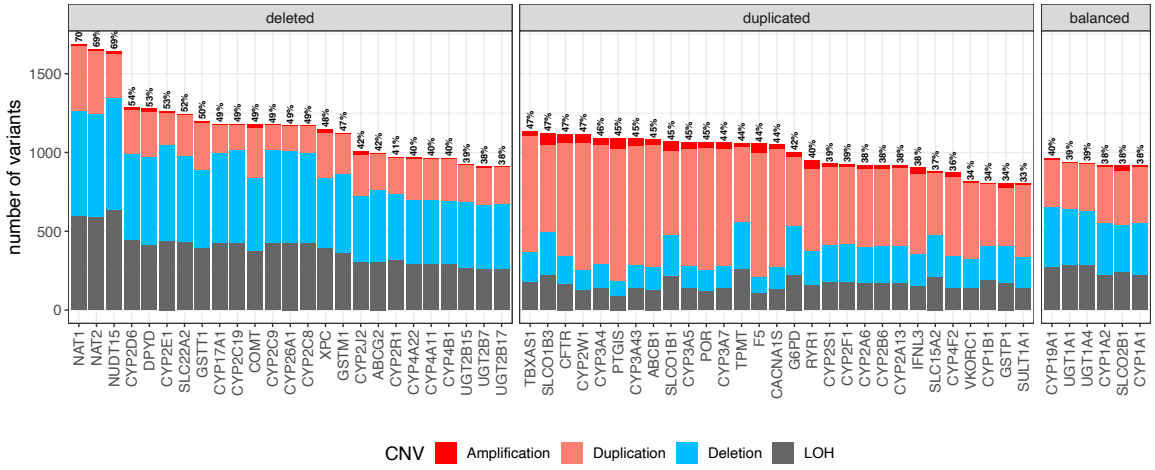


Figure 2.24: Number of somatic CNAs of pharmacogenes in the MASTER cohort. 3 groups could be separated on a pan-MASTER level.

of the analyzed sCNA segments were large (>1000kb) in size, they contain many genes in addition to the analyzed pharmacogenes, including driver genes such as oncogenes and tumor suppressors. I thus wanted to investigate whether these driver genes could be the cause of the observed PGx-sCNA profiles since pharmacogenes are currently not suspected of being cancer drivers themselves. Therefore, I analyzed whether well-known relevant cancer genes, as listed in the cancer gene census [125], including oncogenes (n=106), and tumor suppressors (n=183) are commonly among the co-affected genes in the sCNA segments (genes described as both oncogene and tumor suppressor in the cancer gene census list were excluded). The hypothesis was that deletions of pharmacogenes can be observed more frequently in loci containing tumor suppressors in the neighborhood and vice versa, that duplications are more commonly detected when oncogenes are located in the proximity of the pharmacogene.

For each sCNA event of each patient affecting a pharmacogene, I summarized the number of oncogenes and tumor suppressors located on the same sCNA segment. In this data, I found 19,787 sCNA segments (76.2% of all sCNA segments) that included pharmacogenes coaffected with either oncogenes and/or tumor suppressors. As shown in figure 2.26, globally, the fraction of co-duplications of pharmacogenes with oncogenes (60%) was higher than with tumor suppressors (40%), and co-deletions occurred more frequently with tumor suppressors than with oncogenes (56% vs. 44%). There were gene-specific differences, which are shown in Figure 2.26. For *POR* and *VKORC1* no sCNA events including oncogenes or tumor suppressors could be found, however, these genes were close to centromeres and the evaluation of these segments showed that ACEseq has problems merging neighboring segments here, leading to small isolated regions for these genes. Consequently, these were removed for subsequent analyses. As illustrated in Figure 2.26, genes

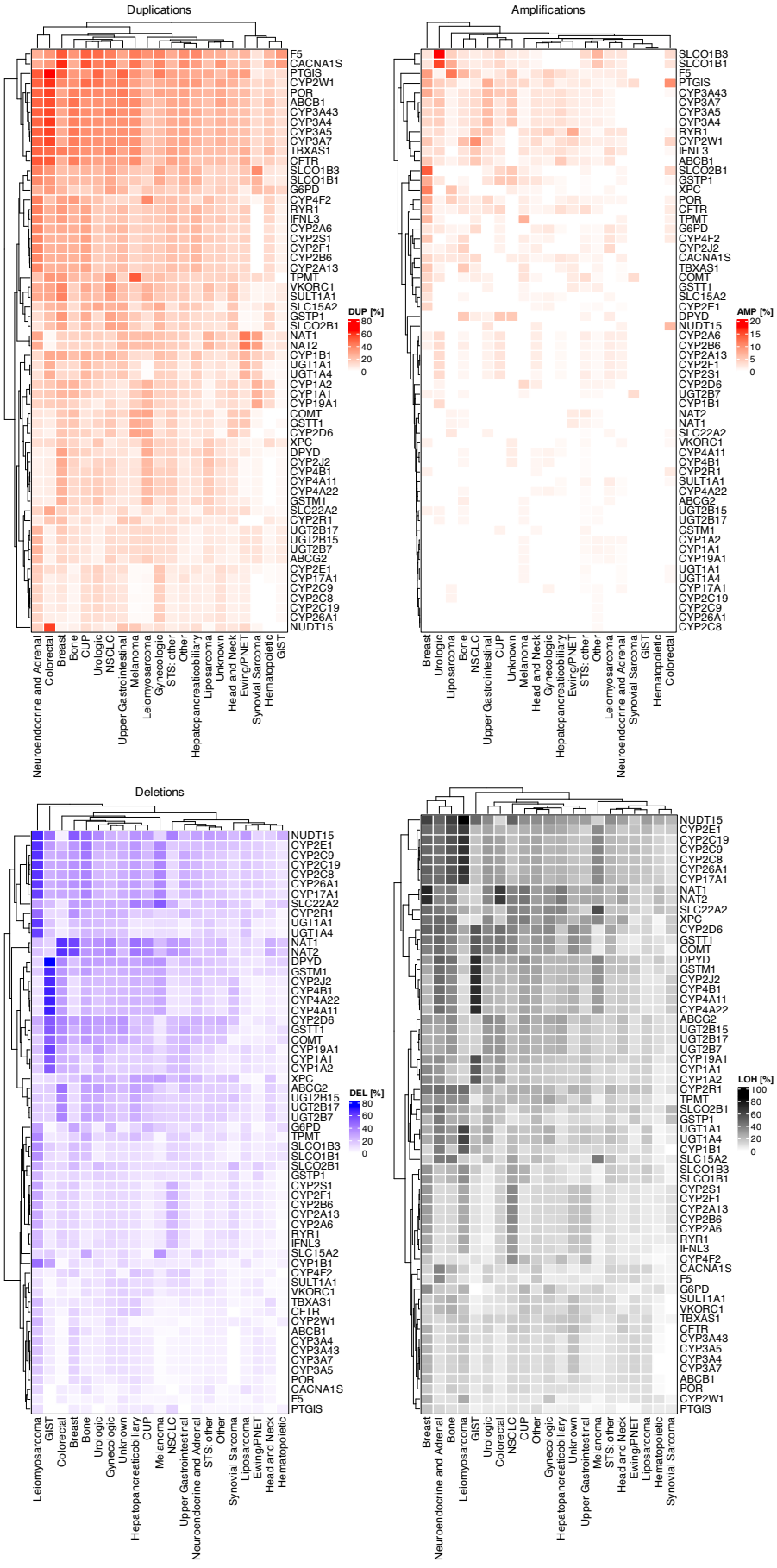


Figure 2.25: Heatmaps with somatic CNAs of pharmacogenes per cancer entity in MASTER. Each heatmap shows one sCNA type (Amplification, Duplication, Deletion, and LOH) with cancer entities as columns and pharmacogenes as rows. The intensity of the heatmap shows the fraction of patients affected by this sCNA type in the respective pharmacogene and entity.

commonly (>60%) deleted together with tumor suppressors were *CYP26A1*, *CYP2C8/9/19*, and *CYP2E1*. Common duplications (>60%) with oncogenes were found for *ABCB1* and *CYP3A4/5/7/43*. Interestingly, I observed unexpected patterns for some genes including *SULT1A1*, *G6PD*, *GSTP1*, and *SLCO1B2* where duplications occurred despite none of the included oncogenes being present. These genes had only tumor suppressors in their proximity and had co-duplications with these in 46-70% of events. The sCNA events affecting pharmacogenes together with oncogenes and tumor suppressors are summarized per cancer entity in Figure 2.27.

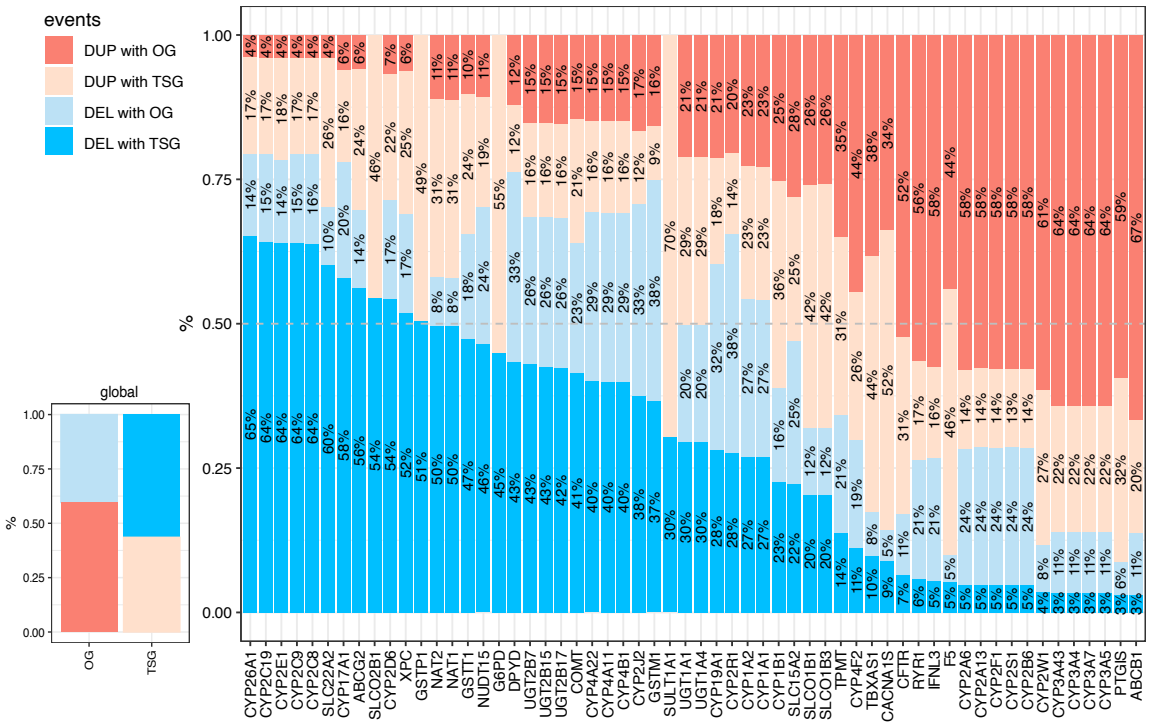


Figure 2.26: Somatic sCNAs of pharmacogenes including co-affected oncogenes and tumor suppressors ordered by event type. (DUP: Duplication, DEL: Deletion, OG: oncogene, TSG: tumor suppressor gene).

Next, I examined chromosomal regions more closely for all pharmacogenes and their neighboring oncogenes and tumor suppressors. Figures 2.28 and 2.29 depict the ideograms of sCNA events (duplications in red, deletions in blue) across chromosomes and show neighboring oncogenes (red text) and tumor suppressors (blue text) for each pharmacogene (black text). The group of commonly duplicated pharmacogenes on the long arm of chromosome 7 including *ABCB1*, *CYP3A4/5/7/43*, overlapped with locations of the oncogenes *TRRAP* and *GRM3*. Particularly colorectal and neuroendocrine cancers showed this enrichment pattern. Also *SLCO1B1/3* which were most frequently amplified (26%) in urologic cancers are closely located next to the *KRAS* oncogene on chromosome 12p. *PTGIS*, mostly amplified in colorectal cancers, is located in the proximity of several oncogenes (*SRC*, *SALL4*, *GNAS*). The role of *SRC*, a well-known proto-oncogene, in colorectal cancers has been previously

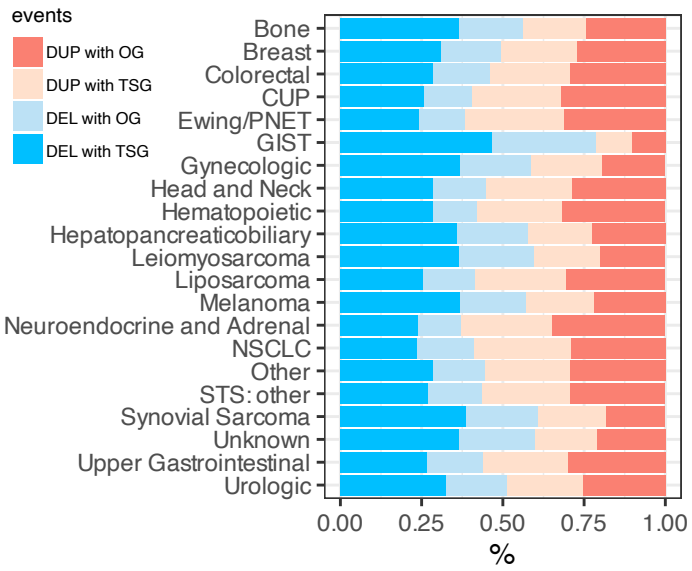


Figure 2.27: sCNAs co-affecting pharmacogenes with oncogenes and tumor suppressors per entity. (DUP: Duplication, DEL: Deletion, OG: oncogene, TSG: tumor suppressor gene).

discussed [126, 127], and gains of the long arm of chromosome 20 including *GNAS* [128] have been reported. This is reflected by the high amount of duplications and amplifications of *PTGIS* in the colorectal cancer group found here. Furthermore, *PTGIS* itself has also recently also been studied in the context of this entity [129, 130] and amplifications were found to be the most frequent alteration. Additionally, high expression of *PTGIS* in colorectal cancer was associated with worse overall survival [130]. In contrast, on chromosome 13 *NUDT15* was frequently co-deleted with tumor suppressor *RB1*, which is a well-known cancer driver. For example, frequent deletions in Leiomyosarcomas included segments on chromosomes 10 and 13, and the respective tumor suppressors in these regions like *PTEN*, *BRCA2*, and *RB1* are commonly observed features in this cancer [111]. The *CYP2C* gene family on chromosome 10 was among the most commonly deleted genes and is close to several tumor suppressors like *PTEN*, *FAS*, *CPEB3*, and *SUFU*.

In some regions where CNAs with oncogenes and tumor suppressors were equally frequent, no enrichment of duplication or deletions was found (chromosome 15q24 with *CYP1A* genes and chromosome 2q37 with UGT1A genes). In contrast, the genes showing unexpected patterns like *SULT1A1*, *G6PD*, *GSTP1*, and *SLCO2B1* as mentioned earlier, only had tumor suppressors in their proximity but were still frequently duplicated. For *SULT1A1* and *G6PD* duplications were even more frequent than deletions, while for *GSTP1* and *SLCO2B1* duplications and deletions occurred equally frequent. Figure 2.30 shows the correlation between the number of sCNA events affecting a pharmacogene (duplications and deletions) and the number of oncogenes and tumor suppressors in proximity (located on the same sCNA segment). With increasing numbers of driver genes in proximity, the amount of sCNAs increases for both categories. Several somatic pharmacogenomic CNAs could therefore be

passenger events originating during the development of cancer through oncogenic mutational processes.

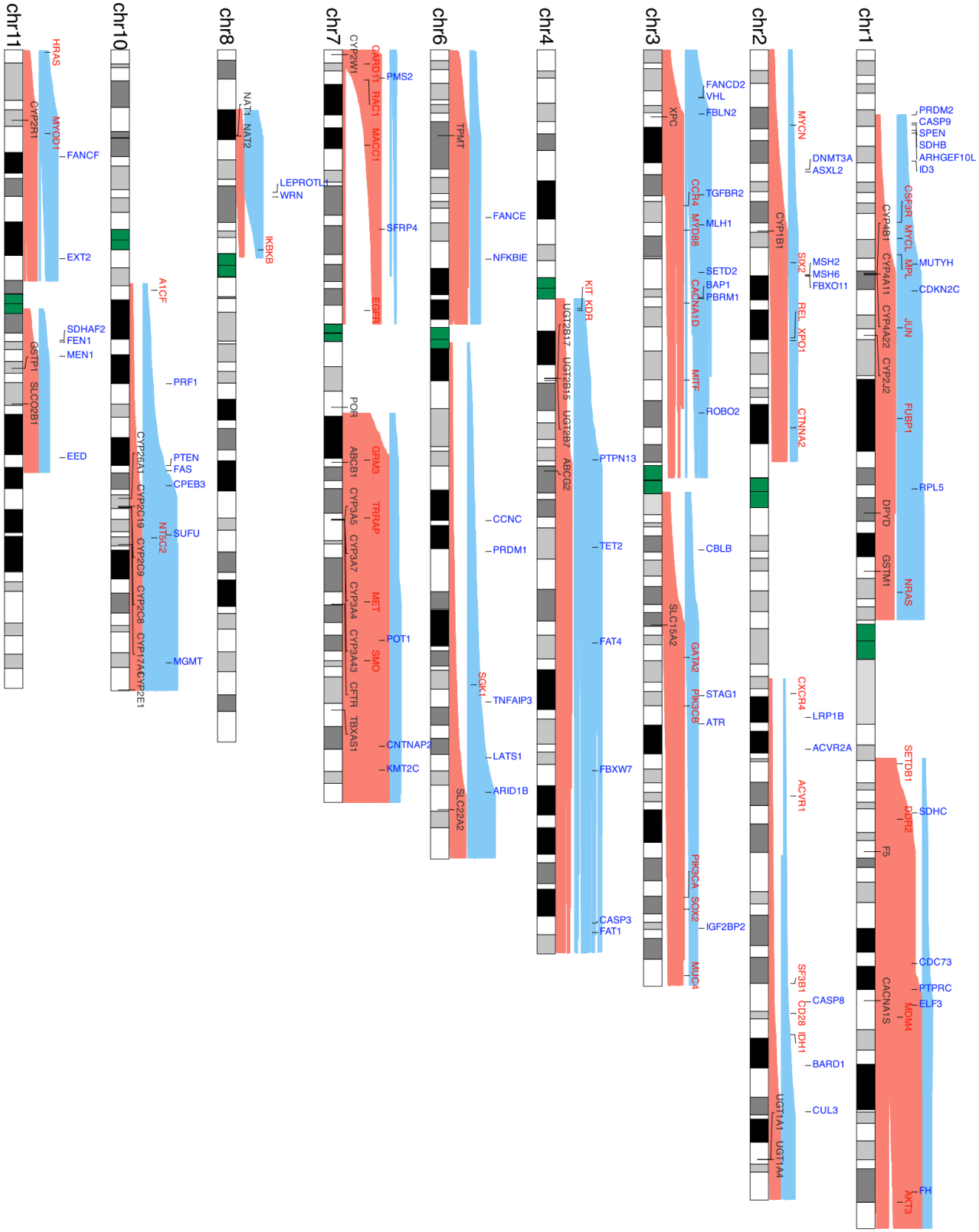


Figure 2.28: Ideogram displaying co-localization of sCNAs in pharmacogenes with oncogenes and tumor suppressors on chromosomes 1-11. Deleted segments are marked in blue and duplicated segments in red. Pharmacogenes are labeled in black text, tumor suppressors in blue text, and oncogenes in red text.

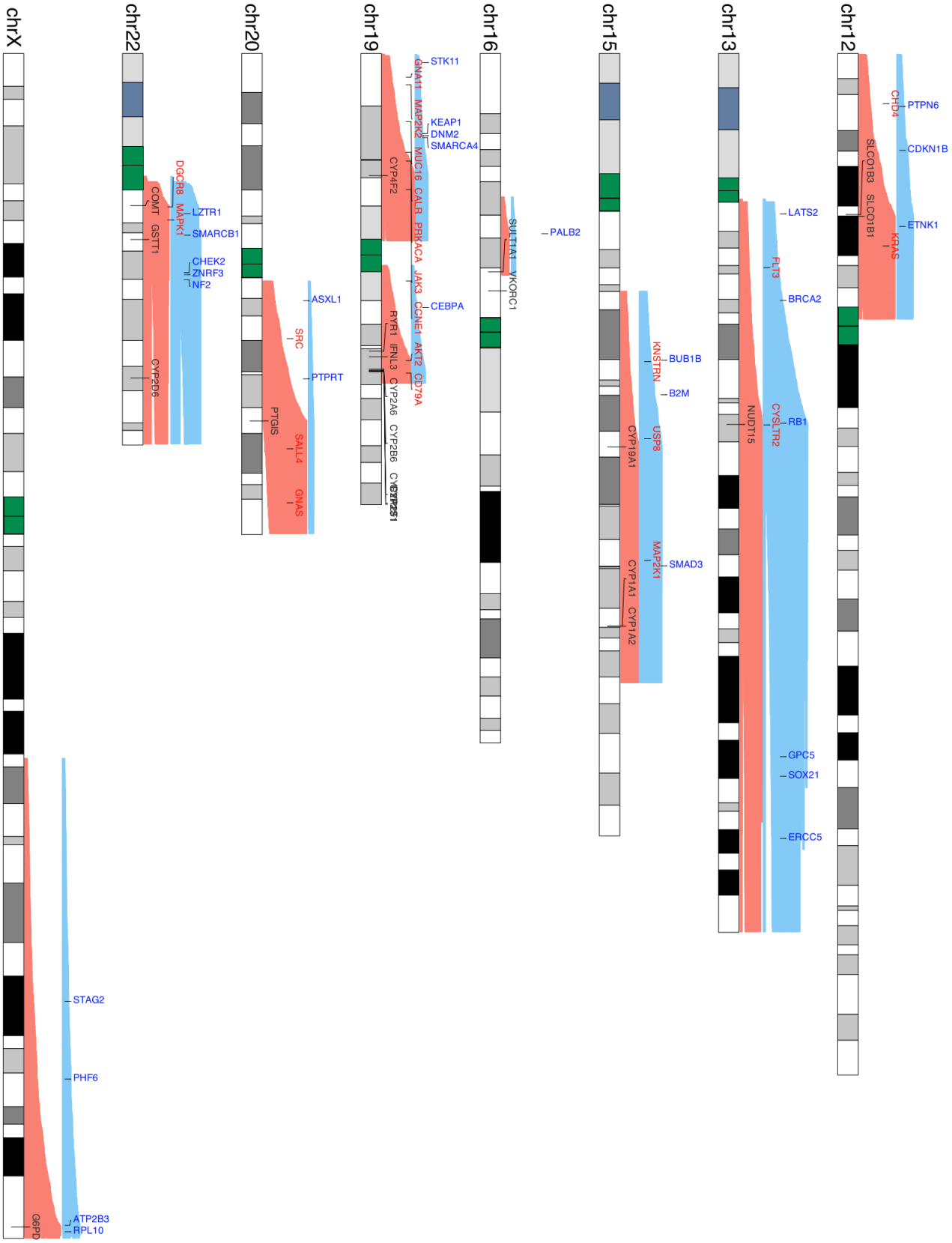


Figure 2.29: Ideogram displaying co-localization of sCNAs in pharmacogenes with oncogenes and tumor suppressors on chromosomes 12-X. Deleted segments are marked in blue, and duplicated segments in red. Pharmacogenes are labeled in black text, tumor suppressors in blue text, and oncogenes in red text.

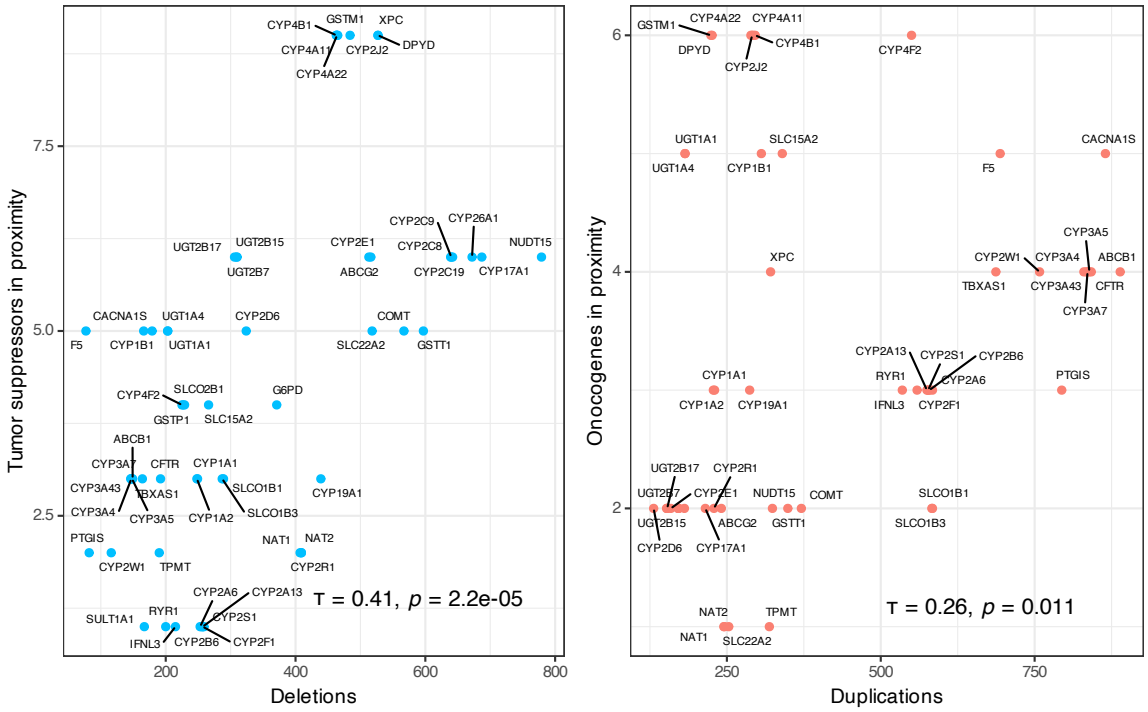


Figure 2.30: Correlation of the number of sCNAs and the number of co-affected oncogenes and tumor suppressors. The left panel shows the number of deletions per gene found in MASTER and the number of tumor suppressors in proximity. The right panel shows the same for duplications and oncogenes in proximity. Correlation was calculated using Kendall's tau coefficient.

Contributions: Sebastian Pirmann performed all analyses and created all of the figures. Małgorzata Oleś provided R code for extraction of sCNA data from the dataMASTER object.

2.5 SCNAs, Epigenetics, and their Effect on Tumor Expression of Pharmacogenes

In the following sections, I present data from other omics layers beyond DNA for the pharmacogenes. In section 2.5.1 I show results of analyses of the somatic gene expression profiles of pharmacogenes in the MASTER entities from tumor bulk RNA-sequencing data. Furthermore, I assessed the effect of sCNAs on gene expression. The methylation of pharmacogenes in the tumor and its effect on expression is described in section 2.5.2. Lastly, a combined multivariate model of the genomic, epigenomic, and transcriptomic data, for estimating the contribution of each layer to pharmacogene activity in the tumor, is presented in section 2.5.3.

2.5.1 Pharmacogene Expression in Tumors

The expression of ADME genes and the effect of somatic variants on the transcriptome will be assessed in the following section, as they may contribute to drug response and resistance [37]. RNA expression of pharmacogenes in tumor samples has been reported for several cancers [62, 131–133]. I analyzed the expression of the selected 60 pharmacogenes in the MASTER cohort from available bulk RNA-sequencing of the tumor samples (n=1936). TPM normalized values were calculated from raw read counts per gene for each patient. A general tendency of higher expression of phase II genes and transporters compared to phase I genes was observed cohort-wide, as shown in Figure 2.31 (Kruskal-Wallis test, $p < 2 * 10^{-16}$). Phase I enzymes had a few high expression outliers, mainly from hepatopancreaticobiliary, gastrointestinal, and colorectal cancers.

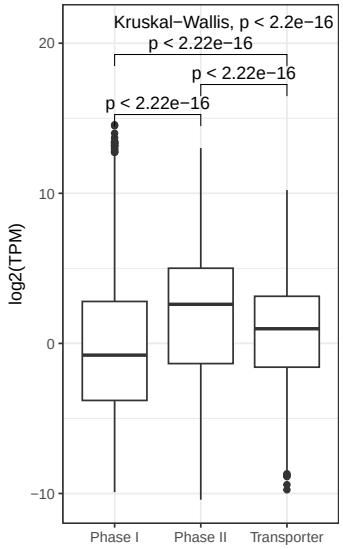


Figure 2.31: Comparison of expression per gene class (Phase I, Phase II, Transporter).
TPM values were grouped by gene class and tested with a Kruskal-Wallis test. Pairwise post-hoc-testing was done with Wilcoxon tests.

Next, I investigated whether there are entity-specific differences in the expression of pharmacogenes across the different tumor tissues. The heatmap in Figure 2.32 shows the mean log2-transformed TPM values per cancer entity for all 60 pharmacogenes. As illustrated, the expression patterns of pharmacogenes in the different cancers could be classified into five groups, using hierarchic clustering as implemented in the ComplexHeatmap package [134]. The first cluster contains *GSTP1*, *POR*, *COMT*, *VKORC1*, *G6PD*, *XPC*, *NUDT15*, and *TPMT* with expression in and only minor variation between tumor entities. This group comprises mainly phase II genes and genes involved in metabolic processing of substances other than drugs. *GSTP1* was the highest ubiquitously expressed gene and in accordance with this finding, high expression of *GSTP1* was already reported in several cancers [135–

138]. The second cluster contained several phase II pharmacogenes like *SULT1A1*, *DPYD*, *GSTM1/GSTT1*, as well as the organic anion transporter *SLCO2B1*, *CYP1B1*, and *CYP2R1*, which were all expressed, but less high than the genes in the first cluster. This is consistent with the increased expression found for phase II genes across the whole pan-rare cancer MASTER cohort. Of note, this cluster includes *PTGIS* and *TBXAS1* which catalyze the modification of prostaglandin H2 [139,140]. Two other gene clusters were characterized by low or moderate tissue-specific expression patterns of mainly phase I genes like *CYP2/3/4* families and some *ABC* and *SLC* transporters, as well as most *UGTs*. The tissue-specific patterns in these two clusters mostly distinguish cancers of drug-metabolizing organs, where generally higher expression of these genes was observed, from the remaining cohort. For example, cancers where phase I genes and transporters had rather low expression were bone cancers, lipo- and synovial sarcoma, and hematopoietic cancers. The remaining cluster is comprised of genes with generally low expression across all cancers except for some tissue-specific outliers like *CYP17A1* in neuroendocrine cancer tissues.

The majority of analyzed genes code for drug-metabolizing enzymes and drug transporters which are usually highly expressed in the main drug-metabolizing tissues like the liver, kidney, and intestines. As expected, the expression of these genes in the hepatopancreaticobiliary, gastrointestinal, and colorectal cancer entities was significantly higher compared to the remaining cancer entities (Kruskal-Wallis test, $p = 1.7 * 10^{-6}$). This indicates that the intra-tumor metabolism of drugs could be particularly relevant for these tumors. Additionally, for some genes and entities outliers of expression levels were observed. These include *PTGIS* in GIST, *CYP2S1* in upper gastrointestinal and colorectal cancers, and *NUDT15*, *NAT2*, *CYP2B6*, *CYP2W1*, and *CFTR* in colorectal cancer. Tumor-specific expression of *CYP2W1* in colorectal cancer has been reported previously [141–143].

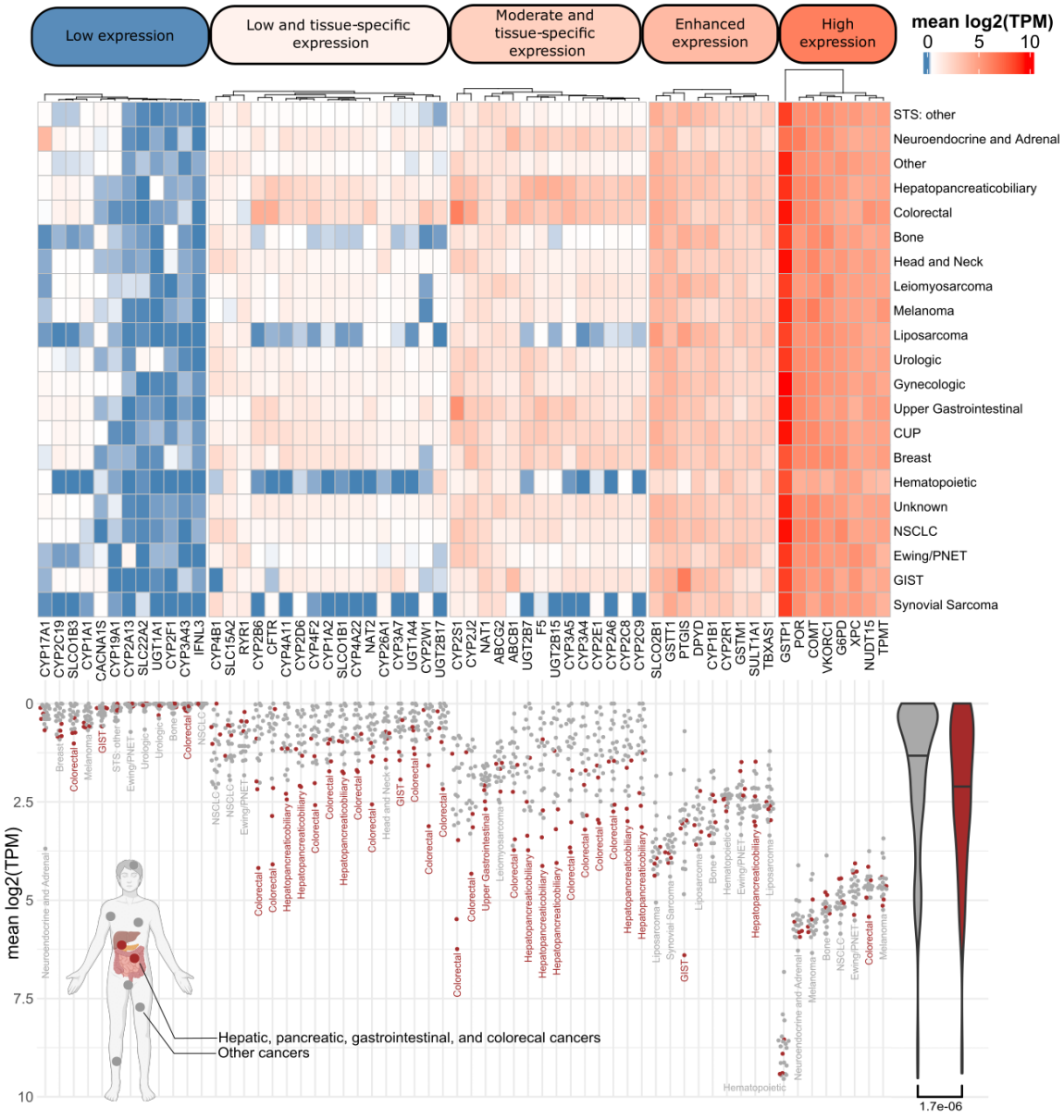


Figure 2.32: Somatic RNA expression of pharmacogenes in the cancer entities of the MASTER cohort. The heatmap shows the average expression of baskets (for visualization log2(TPM+1) was used and averaged for each cancer). The plot below indicates the distribution of expression across cancer entities (with inverted y-axis, i.e., having increasing values towards the bottom). The entity with the highest expression is labeled and hepatic, pancreatic, gastrointestinal, and colorectal cancers are marked in red. These cancers had significantly higher expression across all genes (Wilcoxon test $p = 1.7 * 10^{-6}$).

Contributions: Sebastian Pirmann performed all analyses, and created all of the figures. Figure 2.32 was created together with Roman Tremmel.

2.5.2 Association of Pharmacogenomic sCNAs with Tumor Gene Expression

It is known that sCNAs play a role in the origin and progression of cancer and that changes in the copy number of a gene can affect its expression [144]. This association has been reported for several important oncogenes and tumor suppressors like *MYC*, *KRAS*, *RB1*, and *TP53* [144], but has also been shown for pharmacogenes in germline [145]. Therefore, I performed a cohort-wide analysis of expression levels and sCNA status for the pharmacogenes to see if comparable effects could be observed in the tumor tissues. For every pharmacogene, patients were stratified into three sCNA groups (deleted, neutral, duplicated/amplified), the association of which with the corresponding TPM values were then tested using Kruskal–Wallis tests (Figure 2.33). Among the highly significant results after Benjamini-Hochberg correction were *NUDT15* ($p_{adj} = 2.84 * 10^{-56}$), *POR* ($p_{adj} = 2.85 * 10^{-39}$), *XPC* ($p_{adj} = 1.24 * 10^{-29}$), *COMT* ($p_{adj} = 7.2 * 10^{-29}$), *TPMT* ($p_{adj} = 2.42 * 10^{-25}$), *CYP2R1* ($p = 2.33 * 10^{-21}$), *SLC15A2* ($3.96 * 10^{-17}$), and *GSTP1* ($p_{adj} = 8.99 * 10^{-14}$).

Interestingly, almost all of these genes were in the clusters of high or enhanced expression (Figure 2.32). *NUDT15* was also among the genes most heavily affected by sCNAs and in particular had the highest number of deletions across most cancer entities. The highly expressed *GSTP1* had a less strong but still significant association, however, it was among genes with the lowest number of sCNAs and mostly duplicated/amplified in breast cancer. This suggests that especially for *GSTP1* there might be other mechanisms (other than sCNAs) that regulate the expression in the tumor. Association for *POR* was also highly significant and it was one of the most frequently duplicated genes across many entities. *POR* plays a major role in drug metabolism as it codes for the enzyme NADPH-cytochrome P450 oxidoreductase and is essential for the functionality of CYP enzymes by transfer of electrons from NADPH [146].

Similar association analyses of sCNA status and expression levels were also performed for each cancer entity separately. Results are shown in Figure 2.34. Significant results after Benjamini-Hochberg correction included *NUDT15*, *COMT*, *DPYD*, *POR* in soft tissue sarcoma, *XPC*, *RYR1*, and *CYP2R1* in neuroendocrine and adrenal cancers, *NAT1* in colorectal and upper gastrointestinal cancers, and *TPMT* in liposarcoma. Many of the significant genes of the cohort-wide analysis were confirmed and I was able to assess in which tissues the highest association signal was present. Of note, this sub-group analysis was dependent on sample sizes. *NUDT15* was highly duplicated, amplified, and expressed in

colorectal cancer whereas for the remaining entities it was mainly deleted; however, it remained below the significance threshold after adjustment for multiple testing, most likely due to the low case numbers in the deletion group. Still, the effect of *NUDT15* sCNAs on gene expression was strong in colorectal cancers and significant if tested separately as shown in Figure 2.35. Entity-specific effects where sCNAs were significantly associated with expression were observed for *NAT1* in upper gastrointestinal and colorectal cancers and *RYR1* in neuroendocrine and adrenal cancer. The potential of expression of *NAT1* as a prognostic biomarker for colorectal cancer has been investigated [147]. Also, several studies have found *NAT1* polymorphisms and hypermethylation as risk factors for colorectal cancer [148–150]. *RYR1* appears in several studies on somatic variants of neuroendocrine cancers and pituitary adenomas [151, 152], but there is no pathophysiological connection yet. In conclusion, it can be said that there is a high association between sCNAs and the expression levels of several pharmacogenes in tumor tissue, especially for the more highly expressed genes.

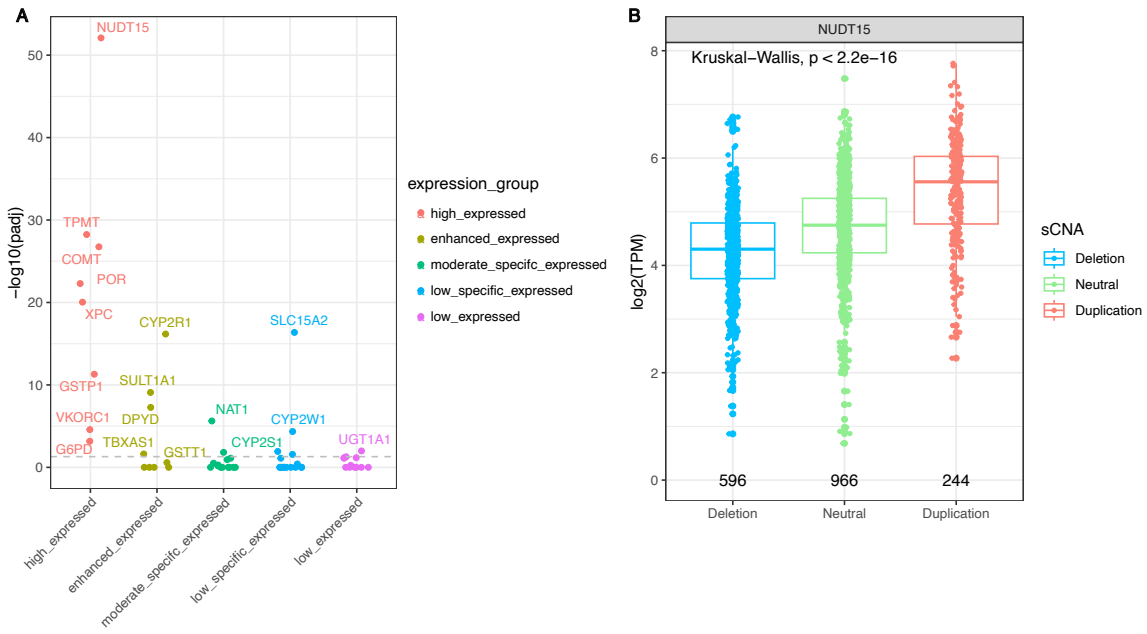


Figure 2.33: Cohort wide association of sCNA status and expression levels. A: The Y-axis shows the significance of the Kruskal–Wallis tests between sCNAs and expression levels per gene. Points are jittered across the x-axis for readability. B: The boxplots show this association exemplarily for the most significant result (*NUDT15*). The y-axis shows the log2-transformed TPM expression values.

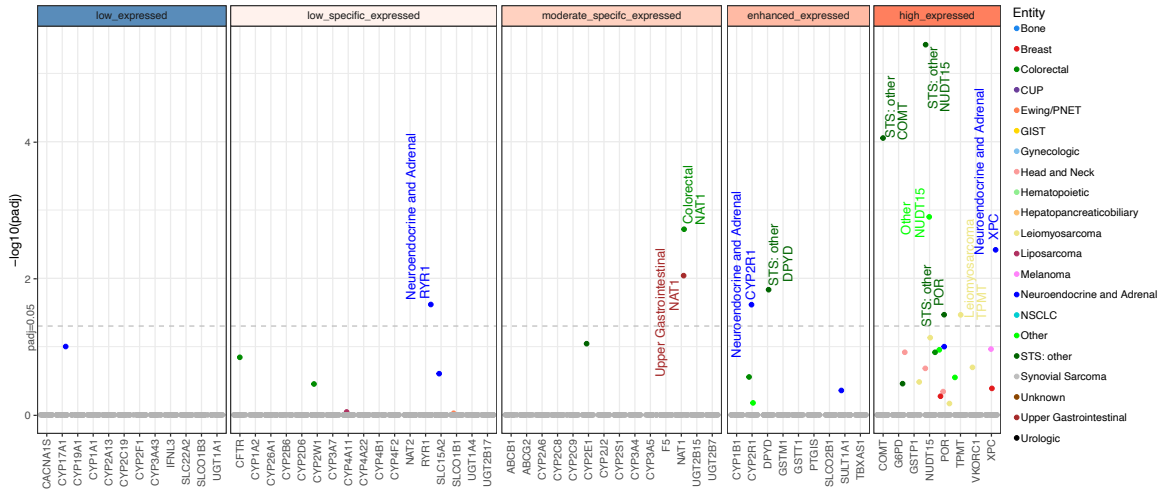


Figure 2.34: Entity-wise correlation of sCNA status and RNA expression levels. The Y-axis shows the significance of the Kruskal–Wallis tests.

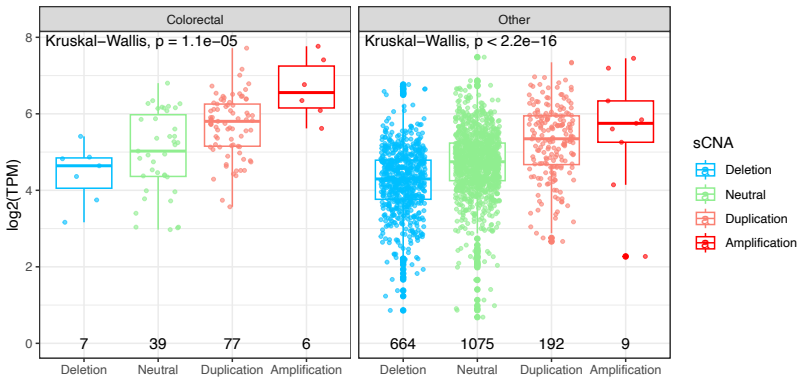


Figure 2.35: Correlation of sCNA status and RNA expression levels of *NUDT15* in colorectal cancers Compared to the remaining entities (right panel) the sample size of the deletion group was rather small, as especially in colorectal cancers *NUDT15* was more often duplicated in contrast to the rest of the cohort. The effect of sCNAs on expression is strong, but in the cohort-wide analysis it remained below significance.

Contributions: Sebastian Pirmann performed all analyses and created all of the figures.

2.5.3 Methylation of Pharmacogenes in Tumors

In the following, I give an overview of methylation of pharmacogenes in tumor samples of the MASTER program. Methylation plays a major role in regulating the expression of genes [153] and is known to promote cancer through activation of oncogenes by hypomethylation or silencing of tumor suppressors by hypermethylation [154]. For ADME genes, epigenetic mechanisms of regulation have already been described for normal and tumor tissues; the current state of the art indicates that individual differences in drug response cannot be explained by genetic variation alone [155].

In this work, DNA methylation (5-Methylcytosin) was measured with Illumina 850k EPIC Arrays as previously described [156] and was available for 1,792 of the 2,371 patients sequenced with WGS (75%). Beta values of 1579 CpG sites in the 60 pharmacogenes were extracted. These included 1226 intragenic CpG sites and 353 promoter CpGs up to 5000 base pairs upstream of the transcription start site (TSS). Numbers of CpGs per gene that were measured and extracted from the EPIC array are displayed in Figure 2.36. For *CYP3A7* and *UGT2B7* no CpG sites in the 5000 base pair region upstream of the TSS were available, also no intragenic CpGs were available for *CYP2A6* and *IFNL3*. Genes with the most analyzed CpGs included e.g. *RYR1*, *TBXAS1*, *POR*, *DPYD*.

For all statistical analyses, beta values were transformed into M values since they have been shown to be more valid for statistical analysis due to their homoscedasticity [157, 158]. In contrast, beta-values have high heteroscedasticity in the strongly methylated and unmethylated regions and are problematic for models that assume normally distributed data. Still, beta values were used for visualization since they are more easily interpretable. The distributions of beta and M values for all analyzed CpG sites are displayed in Figure 2.37.

First, I averaged the beta values of all CpG sites per gene and cancer entity to examine global tendencies of methylation as shown in the heatmap in Figure 2.38(A). Based on the clustering of the methylation values in the heatmap, a rough grouping of carcinomas vs. sarcomas can be seen, with overall methylation values being slightly higher in the carcinomas. The genes with the highest methylation values across entities included *UGT2B15*, *CYP2C8*, and *SLC15A2*. The least methylated genes were *CYP26A1*, *NUDT15*, and *GSTT1*. Some genes showed general entity-specific differences such as *CYP2A13* and *CYP2E1* or individual outliers of single entities like *CYP2D6* and *CYP2W1* which, e.g., were less methylated in synovial sarcoma.

Additionally, average beta values of CpG sites separated by chromatin state (promoter vs. intragenic) are shown in the lower panel (B) of Figure 2.38. Beta value cutoffs were defined as

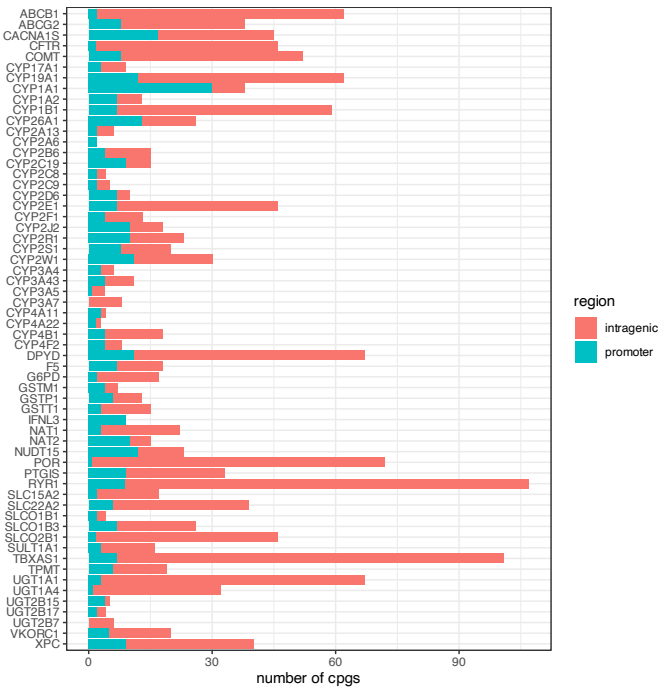


Figure 2.36: Number of CpG sites per pharmacogene that were extracted and included in the methylation analysis. Methylation data was derived from Illumina 850k EPIC Arrays.

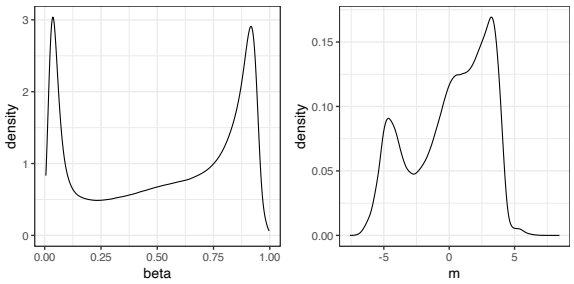


Figure 2.37: Distributions of beta and M values for all CpG sites in pharmacogenes. Beta values were used for visualization and M values for statistical analysis.

hypermethylated (beta > 0.75) and hypomethylated (beta < 0.25). Genes showing promoter hypermethylation in several cancer entities included *POR*, *G6PD*, *CYP2C8*, *UGT2B15*, and *SLC15A2*. Hypomethylation of promoters was found for genes *XPC*, *NUDT15*, *TPMT*, *DPYD*, *CYP2S1*, *CYP2J2*, *CYP16A1*, and *CYP1A1*. Intragenic hypermethylation of genes in several entities included *CYP2C8*, *UGT2B15*, *NAT2*, and *CYP17A1*. In contrast, intragenic hypomethylation affected *GSTP1*, *VKORC1 NUDT15*, *GSTT1*, *CYP2R1*, *SLCO1B1*, and *CYP26A1*. For some genes, promoter and intragenic CpGs showed the same direction of methylation like *NUDT15*, *CYP2C8*, *UGT2B15*, and *CYP26A1*. For other genes, promoter and intragenic methylation generally differed in direction like *DPYD* and *XPC*.

Most genes in the highly expressed group had a tendency towards lower promoter and intragenic methylation values while the methylation values in the tissue-specific expression groups were most diverse. The group of low-expressed genes tended to have a higher

promoter and intragenic methylation. *NUDT15*, which was most notably hypomethylated across all entities, also showed high expression as well as a high association of expression with sCNA status. In contrast, the expression of *POR* also had a high association with sCNA status, but methylation, especially in the promoter, was generally high (average beta value > 0.75).

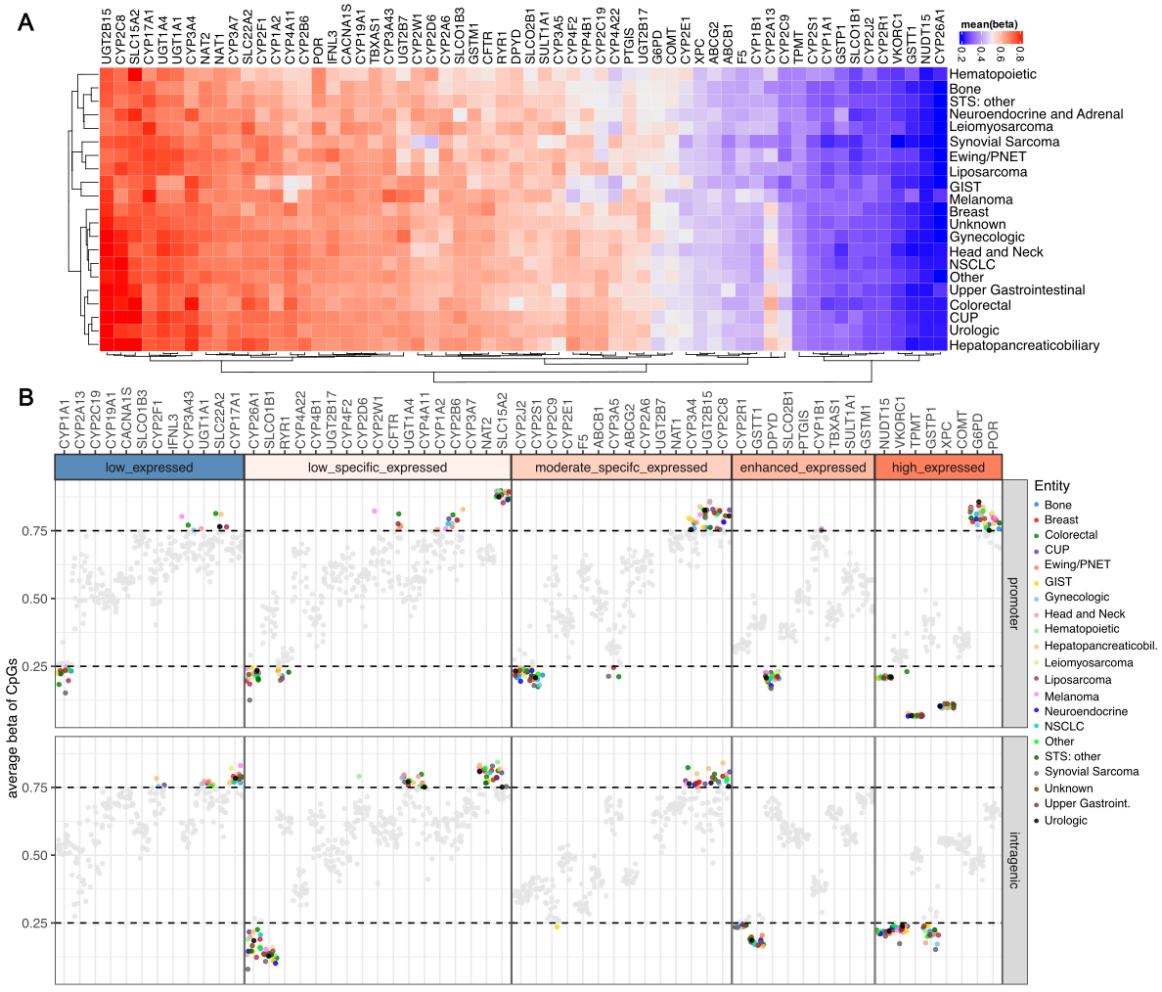


Figure 2.38: Overview of entity-specific methylation of pharmacogenes. The upper panel (A) shows entity-specific promoter and intragenic methylation for all pharmacogenes separated by expression groups. The lower panel (B) shows a heatmap of beta methylation values per gene and cancer entity describing global patterns of methylation across cancer entities and pharmacogenes.

Next, I performed an entity-specific correlation analysis between methylation and expression levels using Spearman correlation tests between M and TPM values per CpG site and gene within each cancer entity. *GSTT1* and *GSTM1* were removed from the analysis since their frequent germline deletions resulted in outliers of very low expression. Figure 2.39 shows the results of the correlation tests separated by promoter and intragenic CpGs and the direction of correlation using a significance cutoff of adjusted p-values < 0.001. The top 200 significant results can additionally be found in Table 6.7 in the appendix.

In principle, among all 1,579 CpG sites, there were 1,200 (427 unique CpGs, 27%) negative

and 1,540 (524 unique CpGs, 33%) positive significant correlations across all entities and genes. The number of significant p-values was higher for the intragenic CpGs (798 unique intragenic and 313 unique promoter CpGs). However, the analysis included 3.5 times more intragenic ones. Overall, p-values for the intragenic CpGs were considerably lower, showing a stronger association with expression. For intragenic CpGs more positive (1,419 with 466 unique CpGs) than negative (1,008 with 344 unique CpGs) correlations were found while for promoter CpGs the opposite was observed (121 positive correlations with 58 unique CpGs, and 192 negative with 58 unique CpGs).

Most significant results (lowest p-values) were found in the high and enhanced expression groups for negative correlation and intragenic CpG sites. In the group of highly expressed genes, 312 (120 unique CpGs) negative and 26 (20 unique CpGs) positive correlations were found. However, except for one promoter CpG, the positive correlations were exclusively intragenic. Also, in the high expression group, the negatively correlated intragenic CpGs had lower p-values (lowest p=$1.22 * 10^{-37}$) compared to positively correlated ones (lowest p=$5.37 * 10-09$). The significant correlations in the high and enhanced expression groups were predominantly observed in the three entities neuroendocrine tumors, soft tissue sarcomas, and colorectal cancers.

For *GSTP1* several intragenic CpG sites were negatively correlated with expression in neuroendocrine cancers and melanoma, but also in many other entities with lower significance. Also, many intragenic CpGs in POR were negatively correlated with expression in neuroendocrine cancers. In soft tissue sarcomas, intragenic CpGs in PTGIS were significantly correlated with expression. In the enhanced expression group, *DPYD* showed a significant positive correlation for intragenic CpGs. , Top hits for negatively correlated promoter CpGs included *CYP2W1* in colorectal and *F5* in neuroendocrine cancers. Expression of *CYP2W1* in colorectal cancers has been reported to be regulated by methylation [159]. For positively correlated CpGs, the intragenic results included neuroendocrine cancer with *ABCB1*, and soft tissue sarcomas with *DPYD*. For positively correlated promoter CpGs also *ABCB1* in neuroendocrine cancers was significant. Selected examples of the most significant correlation tests are shown in the lower panel in Figure 2.39.

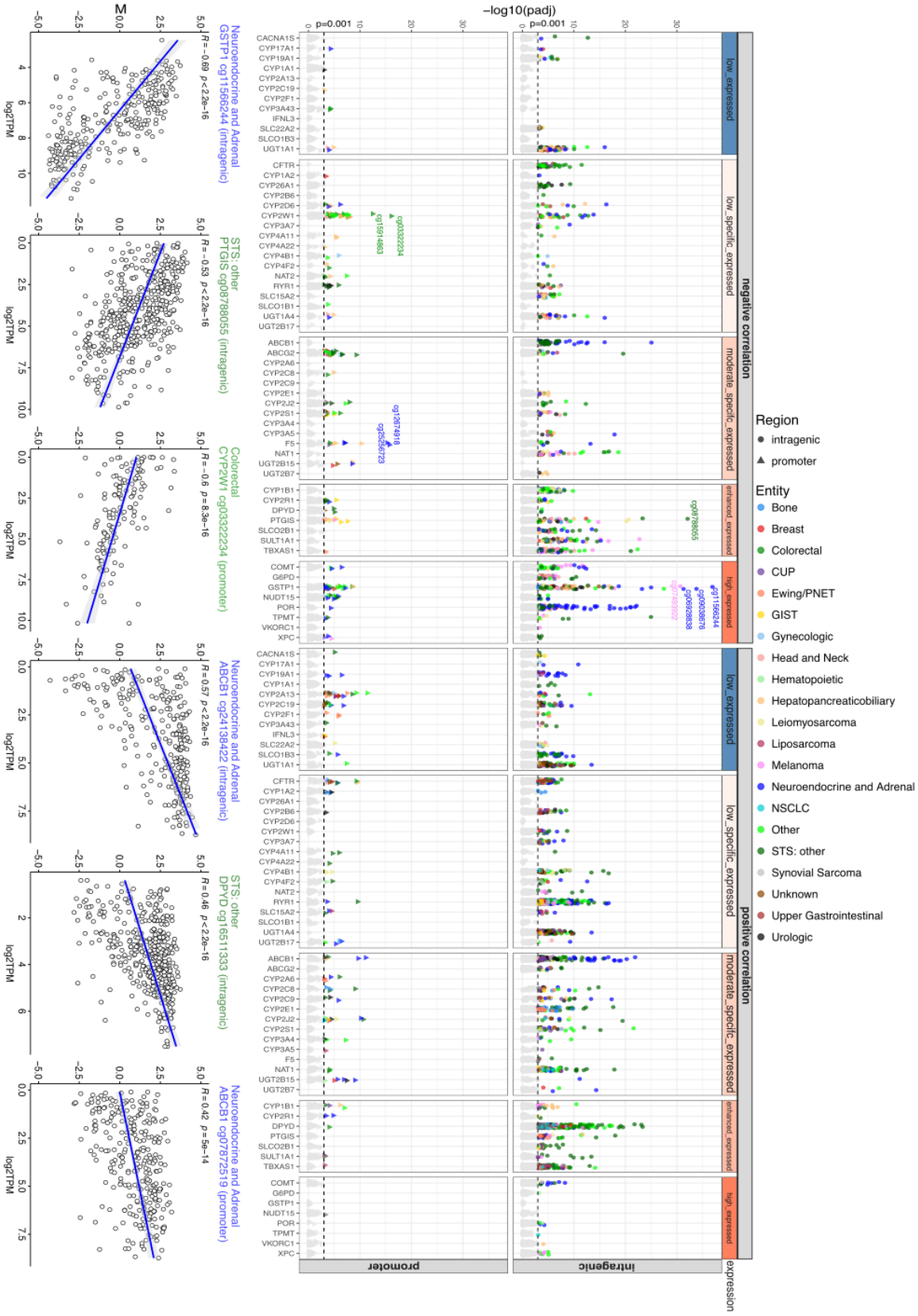


Figure 2.39: Correlation of methylation values and expression per CpG site grouped by gene and cancer entity. Selected scatterplots of highly correlated CpG sites are displayed below.

Contributions: Sebastian Pirmann performed all analyses and created all of the figures. Daniel Lipka provided the methylation data from the EPIC arrays.

2.5.4 Multivariate Models of Somatic Pharmacogene Expression

The influence of germline variants on expression, protein activity, and drug response is evident for many pharmacogenes including CYPs, UGTs, GSTs, SULTs, and transporters. With the next analysis, I aimed to investigate the combined influence of germline variants, somatic variants, and epigenetics on the intra-tumor gene expression of pharmacogenes. The univariate effect of somatic genetic and epigenetic variation on the somatic ADME expression was described thoroughly in the previous sections. As demonstrated in the previous analyses the activity of some ADME genes seems to be mainly influenced by sCNAs (*NUDT15*) while for others expression seems to be more strongly regulated by methylation (*GSTP1*). In this section, I aimed to understand the amount of variance that is explained by the genetic and epigenetic factors on ADME RNA expression. Therefore I integrated these data layers and used multivariate linear regression models, as the interpretability of such models is high. The curated germline genotype results of the pipeline were combined with sCNAs and methylation data in order to model RNA expression. In total, complete data was available for 1,450 patients.

The choice of input data for the regression models was motivated by the following reasoning: (i) the functional effect of germline star alleles has been extensively demonstrated and the germline variants are also present in the tumor; (ii) only very rarely (in 1% of patients), somatic SNVs in the tumors matched both the very position and the exact exchange of thoroughly described germline PGx SNVs; (iii) due to complex convolution of small variants and sCNAs in tumor genomes, an exact determination of the zygosity of the small variants may remain unreliable making an exact star allele genotyping for tumors difficult. The combination of genotyping results from the germline with somatic CNAs was therefore considered as a good approximation for the genetic component influencing intra-tumor gene expression.

GSTT1 and *GSTM1* were excluded from the analysis because of their frequent germline deletions. For methylation, CpG sites were further restricted to the ones that were significantly correlated to expression based on the previous univariate analysis (padj < 0.0001), to reduce the amount of features in this data layer. First, for each gene, models were fitted pan-MASTER cohort-wide, and then further models were fitted per cancer entity separately using the following formula for multivariate linear regression:

$$TPM \sim consensus\_genotype\_germline + sCNA\_type + significant\_cpgs.$$

For each model, the selected features were extracted and grouped by data layer (germline genotype, somatic CNAs, and methylation). The significance of features for the cohort-wide models per gene is shown in figure 2.40; results were grouped by expression gene classes as described above. In the group of highly expressed genes sCNA status was mostly the dominating feature while for the tissue-specific expression groups, methylation was most significant, confirming the results of the previous univariate analyses. The most significant results include methylation in *CYP3A5*, *CYP2S1*, *CYP2C9*, and *UGT2B15* in the tissue-specific expression group. Also, findings for *NUDT15*, for which sCNAs had the most significant association with expression and in which methylation was generally low, were confirmed in this analysis. In contrast, *GSTP1* was also highly expressed but the association with sCNA status was lower. As expected from the previous methylation analysis, the multivariate models for this gene mainly contain CpG sites as a dominant feature.

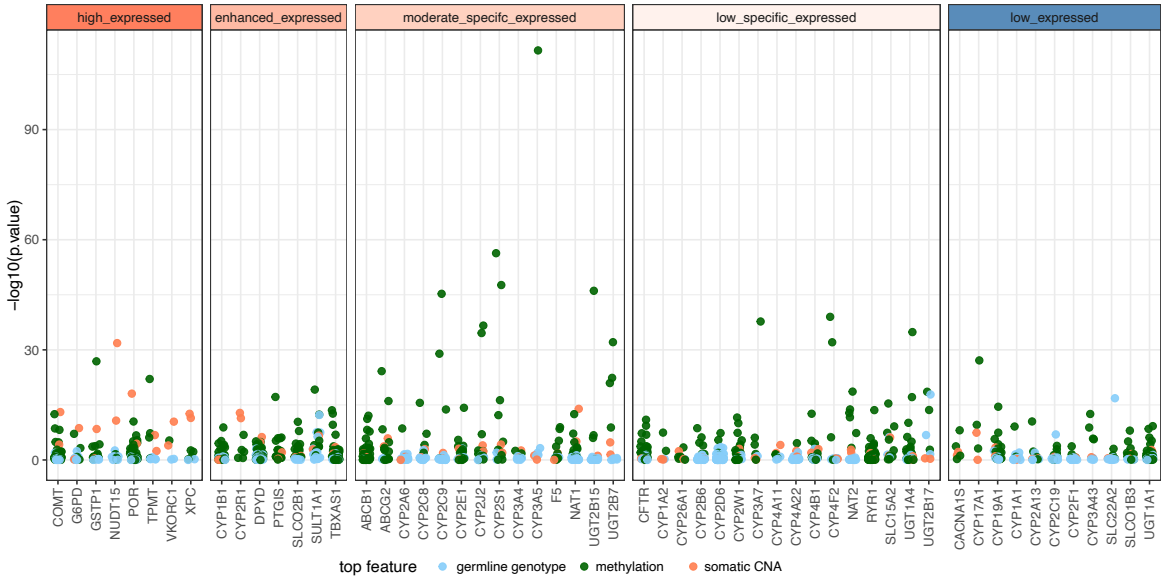


Figure 2.40: Most significantly associated features of linear models for somatic RNA expression. For every gene a multivariate linear model was constructed taking genetic and epigenetic data as input. For each model, the most significantly associated feature was extracted. Genes are grouped by expression category as described above and the y-axis shows the significance of the feature in the regression models.

The results of the entity-specific models per gene are displayed in Figure 2.41. For some entities, no model could be created due to limited cohort size and the resulting lack of feature levels (gray tiles in Figure 2.41). In summary, methylation was the most strongly associated factor for 44 (75.8%) pharmacogenes, followed by star allele genotypes for 5 (8.6%), and sCNAs for 4 (6.8%) pharmacogenes as shown in the right row annotation barplot. The remaining 5 genes had equal numbers of associations of at least 2 data layers. For all entities, methylation was the dominant feature across most genes as shown in the top column annotation barplot. Entities with the most methylation-associated genes included soft tissue

sarcomas, hepatopancreaticobiliary cancers, and neuroendocrine tumors. The proportions of genes for which sCNAs and germline genotypes were the top features were approximately equal between the entities. Genes predominantly influenced by star alleles were genes from the CYP family including *CYP2A6*, *CYP2B6*, *CYP2D6*, *CYP2F1*, and *CYP4A22*. This seems plausible for *CYP2D6*, as here the largest number of genotypes is known and functionally studied and there are around 40 non-functional alleles, where methylation status and additional somatic copies have no effect on expression in the tumor. Genes for which sCNAs play the major regulative role were *NUDT15*, *XPC*, *G6PD*, and *VKORC1*. For the remaining genes, methylation was most strongly associated. In the high-expression group, sCNAs and methylation were the dominating factors while germline genotype associations were rare. The enhanced expression group showed methylation as a dominating feature for all genes except *CYP2R1*, the expression of which was associated with sCNAs. In the remaining 3 groups, methylation was also the most significant feature with some exceptions, such as *CYP3A4*, which also has a proportion of sCNAs.

The analysis confirmed some results from the univariate entity-wide association analyses. For instance, the regulation of *GSTP1* expression was mainly associated with methylation in neuroendocrine cancers. Another example is the influence of methylation on *CYP2W1* on expression in colorectal cancers. Also, the dependency of *NAT1* expression in colorectal cancer and *NUDT15* expression in soft tissue sarcoma on sCNA status was confirmed.

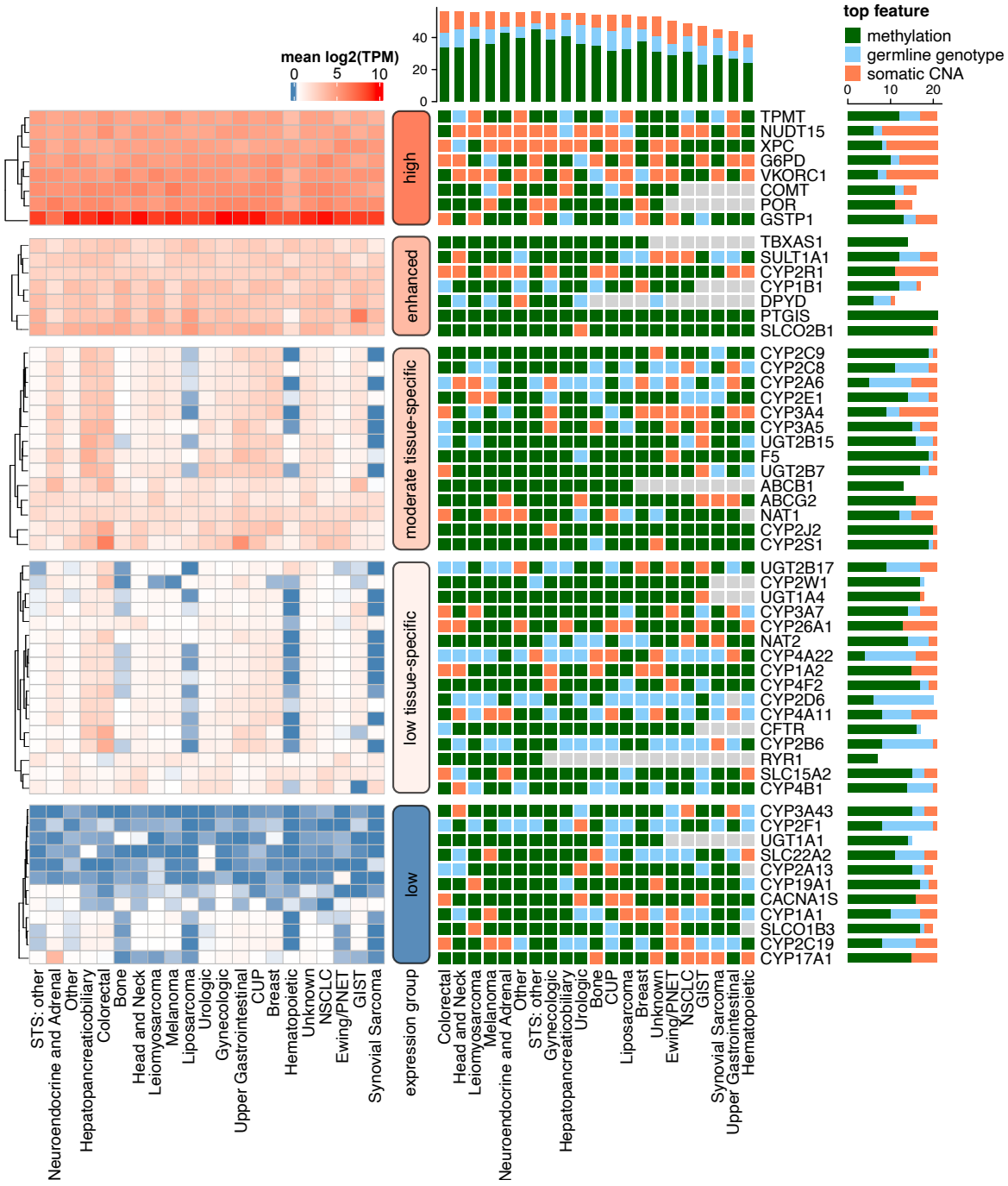


Figure 2.41: Top features of linear models for somatic expression.
Entity-specific multivariate linear models for all pharmacogenes were constructed based on genetic and epigenetic data. For each model, the top significant feature was extracted.

Contributions:
Sebastian Pirmann performed all analyses and created all of the figures.

2.6 Genomic and Transcriptomic Analyses of Rare Cancers

The following sections show the results of three side projects about cancer genomics cohort analyses for different rare cancers (parathyroid carcinomas, adrenocortical carcinomas, and chordomas) that were part of larger collaborations. Being part of these collaborations, I carried out genomic and transcriptomic analyses based on germline and tumor whole genome DNA and tumor bulk RNA sequencing.

2.6.1 Parathyroid Carcinoma

Parathyroid carcinoma (PC) is an extremely rare form of cancer with a yearly incidence of about 3-5 per 10 million [160]. Currently, there are no established systemic treatments or known actionable alterations, and the risk of recurrence after surgical removal is high [161]. In this study, the genomic and transcriptomic profiles of 4 advanced (metastatic) PC patients (2 female, 2 male) were analyzed, with the aim of identifying molecular alterations that drive the disease and providing recommendations for personalized experimental treatment. The complete publication of the study can be found here [162]. This section will only focus on the computational analyses of the genome and transcriptome that were performed as part of the larger study.
The genomic analyses included the detection of germline and somatic variants from WGS. This included small variants (SNV, Indels) somatic copy number aberrations (sCNAs), and structural variants (translocations, inversions). Gene fusions were detected and gene expression was quantified from RNA-seq of the tumor samples. The genomic landscape, including a selected set of recurrently mutated genes (present in >1 patient) and genes that are known drivers for PC, is depicted as an oncoprint in Figure 2.42. Germline SNVs (non-synonymous) were only found in one patient in *MUTYH* and *MSH6*. Somatic SNVs in one patient affected *CDC73* (stop gain) and *MSH6* (non-synonymous). Indels were present in CDC73 (frameshift deletion) and MEN1 (non-frameshift deletion). Gene level sCNAs included deletions in 3 patients in *CDC73*, *DICER1*, *MEN1*, *HRAS*, and amplifications in *CCND1* and *HRAS*. One fusion was found in one patient in *DICER1*. Arm-level events were found for chromosome regions chr3q (3 deletions, one copy number-neutral LOH) and chr13q (3 deletions). Two of the patients had a ploidy larger than two. The total number of variants for each patient including all genes (beyond the listed genes) are shown in

the bar plots in the column annotation at the top of the oncoprint. One case had a very high mutational burden (>40000 SNVs) and fulfilled the criteria of a hypermutated tumor.

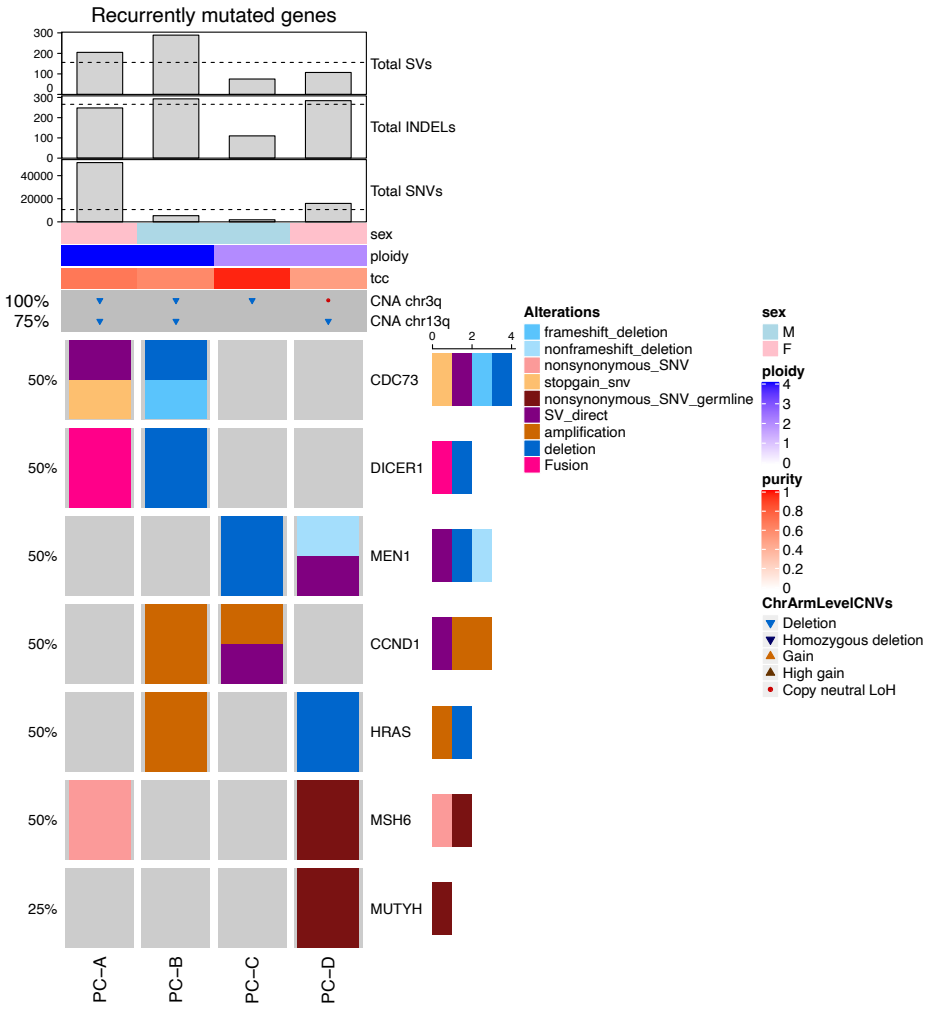


Figure 2.42: Oncoprint showing the genomic landscape of 4 parathyroid carcinoma cases. The listed genes are a subset of recurrently mutated or PC-relevant genes. The top annotations show the total numbers of SVs, InDels, and SNVs per patient in all genes (also beyond the displayed ones). Additional top annotations include sex, ploidy, tumor cell content (TCC), and recurrently CNA-affected genomic regions (chromosome arm-level events).

Additionally, I performed a mutational signature analysis based on somatic SNVs with YAPSA [163] to identify underlying mutational processes. The signatures found in the patients and their respective exposure values with error bars representing 95% confidence intervals are shown in Figure 2.43. The detected signatures [54] included single-base substitution signature 3 found in 3 patients (SBS3), which is associated with homologous recombination repair deficiency (HRD), SBS2, and SBS13 resulting from overactivation of APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide) in all patients, and SBS18 resulting from damage by reactive oxygen species in one patient.

Based on TPM expression values calculated from available bulk RNA-Sequencing of the

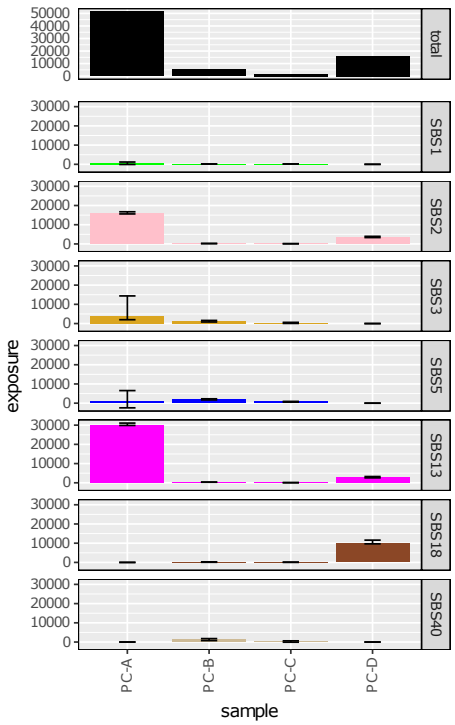


Figure 2.43: Exposures barplot showing the exposure to different mutational processes that underlie the mutational signatures in the tumors of the patients.

tumor samples I performed an in silico immune cell type quantification with the immunedeconv package [164]. The package includes seven different algorithms with five deconvolution-based approaches and 2 methods based on marker genes. Marker-gene methods quantify cell types independently based on signature gene expression values, either directly [165, 166] or through a statistical test for enrichment [167]. Deconvolution methods formulate the problem as linear equations using signature and gene expression matrices [168–174]. Various regression techniques, such as support vector regression, constrained least square regression, or linear least square regression, are employed for this purpose. Results of the deconvolution analysis with Cibersort [170] are shown in figure 2.44. 3 of 4 patients mainly had macrophages (M1, M2, M3) and T cells present in their tumor microenvironment. The amount of B cells in all samples was rather low. For one patient (PC-C) there were fewer admixed immune cells in the tumor compared to the others. Also, one patient (PC-A) had a higher fraction of T cells present in the tumor, with a high number of CD8+ T cells and T-follicular helper cells. In contrast, PC-D had high numbers of regulatory T cells and M2 macrophages, while fewer M1 macrophages, CD8+ T cells, and T-follicular helper cells. This indicated a potential response of PC-D to immune checkpoint inhibitor treatment.

Figure 2.44: Results of immune cell deconvolution of parathyroid carcinoma patients with Cibersort [170]. Based on RNA-Sequencing from tumor samples (top: absolute values, bottom: percentages).

Contributions: Matthias Kroiß and Veronica Teleanu designed the study, selected patients, collected clinical data, and drafted the manuscript. Sebastian Pirmann and Nagarajan Paramasivam performed genomic variant analyses and generated the oncoprint. Sebastian Pirmann did the mutational signature analysis and immune cell deconvolution and created the respective figures.

2.6.2 Adrenocortical Carcinoma

Adrenocortical carcinoma (ACC) is a rare (0.7–2 per million) and aggressive form of cancer originating in the adrenal cortex [175, 176], which is the outer layer of the adrenal glands. Prognosis varies by stage, with generally poor outcomes for advanced stages. Current treatments involve mostly surgery and chemotherapy. Therefore, comprehensive genomic analyses are still needed to better understand the drivers and find potential targeted treatment options. Although several studies have already been carried out, the entire picture of the origin and progression of ACC has not yet been fully deciphered. The aim of this project was to investigate the genomic and transcriptomic landscape of this uniquely large cohort of ACC patients to gain more detailed insights into this rare disease.

Whole genome or exome sequencing data of peripheral blood and tumor samples was available for 113 ACC patients, with additional tumor bulk RNA sequencing in 89 cases. Figure 2.45 shows the genomic landscape of recurrently mutated genes and chromosomal segments

in the cohort. The tumor mutational burden was rather low in most cases, with a few exceptions with up to 40 mutations per megabase. Among the most frequently mutated genes (20-40%) were previously reported ACC drivers like *TP53*, *TERT*, *ZNRF3*, *CTNNB1* [176]. The single most frequently mutated gene (38%) was *TP53*, affected by somatic SNVs, indels, and a few somatic homozygous deletions and germline SNVs. Frequent amplifications were found in *TERT* located on chromosome 5 in cytoband 5p15.33. Another mainly sCNA-affected gene was *ZNRF3*, with a high amount of homozygous deletions resulting from deletions of chromosomal segment 22q12.1. Interestingly, *CYP17A1* was also altered in 16% of cases, mostly affected by gene fusions, and the pharmacogenomic analysis of the MASTER cohort has shown that this gene was highly expressed in neuroendocrine and adrenal cancers.

Chromosomal regions recurrently affected by sCNAs (significantly amplified or deleted across cohort) were identified using GISTIC [177] and confirmed previously reported patterns [175, 176]. These included frequent (50-70%) amplifications of regions on chromosomes 12 (*CDK4*) and 19 (*CCNE1*), and deletions of regions on chromosome 22 (*ZNRF3*, 40-50%). The complete GISTIC profile of the cohort is shown in Figure 2.46.

Mutational signatures were analyzed with YAPSA [163] and significantly enriched signatures included AC1, AC2, AC3, AC13, and AC23. AC1 is a clocklike signature from spontaneous deamination of 5-methylcytosine to thymine. AC2 and AC13 are resulting from APOBEC activity, while AC3 is related to defective homologous repair. The origin of AC23 is currently still unknown [53, 54].

Additionally, immune cell deconvolution was performed with immunedeconv [164], as described earlier, for the 89 cases where tumor bulk RNA sequencing was available. The amount of infiltrating immune cells was rather low for most samples with a few exceptions of high T-cell, and one case with high B-cell contributions.

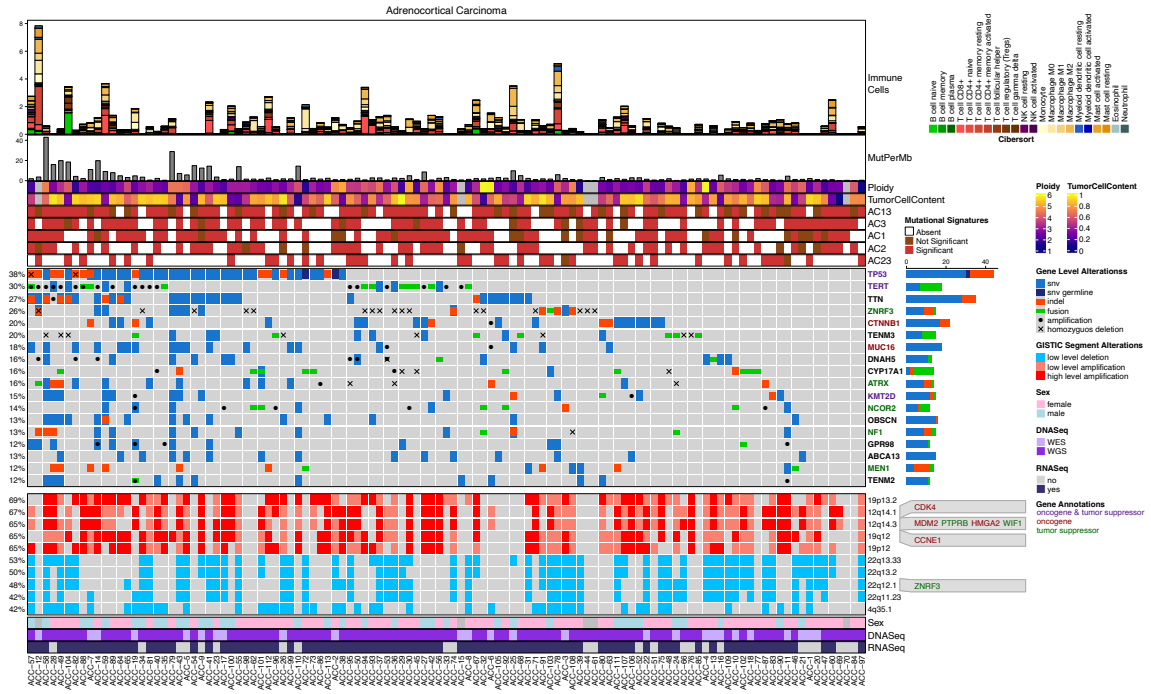


Figure 2.45: Oncoprint showing the genomic landscape of the adrenocortical carcinoma cohort. From top to bottom this integrated oncoprint shows deconvolution of admixed immune cells, mutational burden, mutational signatures, gene level alteration, and chromosome segment alterations. Each column represents one sample.

One of the few drugs used for adjuvant therapy of ACC is mitotane, a cytostatic agent that selectively inhibits cell division in the adrenal cortex [178]. By increasing free cholesterol mitotane leads to cell death [179, 180]. However, its mechanism of action is yet to be fully described. There have been studies investigating the effect of several pharmacogenes on mitotane concentrations achieved in ACC patients, including variants in *CYP2B6*, *CYP2W1*, *CYP2C19*, *SLCO1B1/3* [181–183]. For example, the variant rs3745274 in *CYP2B6* was previously reported to influence mitotane concentrations [181]. For 43 patients measurements of maximum achieved mitotane concentrations were available from two time points, prior to biopsy and at the last follow-up. The genotypes of rs3745274 for these patients (GG=25, GT=13, TT=5) were determined using the PGx pipeline. Figure 2.47 shows the distribution of concentration for the three genotype groups at the two time points. In both cases, no significant difference in concentrations could be found between the genotypes. Unfortunately, it was not possible to carry out a more detailed analysis as no concentration measurements over time, such as concentration curves, were available.

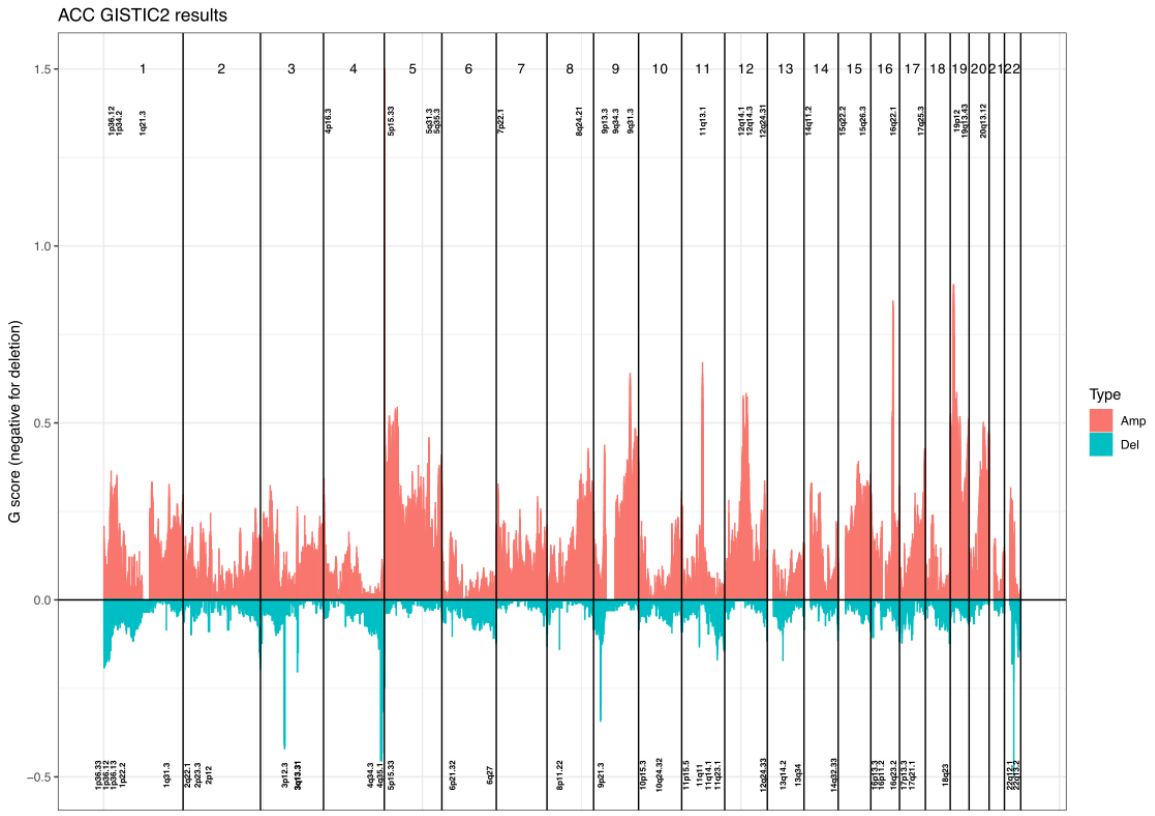


Figure 2.46: All recurrently amplified and deleted chromosome segments identified in the ACC cohort by GISTIC [177]. The G-score (y-axis) shows the significance of recurrently amplified and deleted regions in the whole cohort. The most frequently affected chromosome regions are labeled and were also integrated into the oncoprint above.

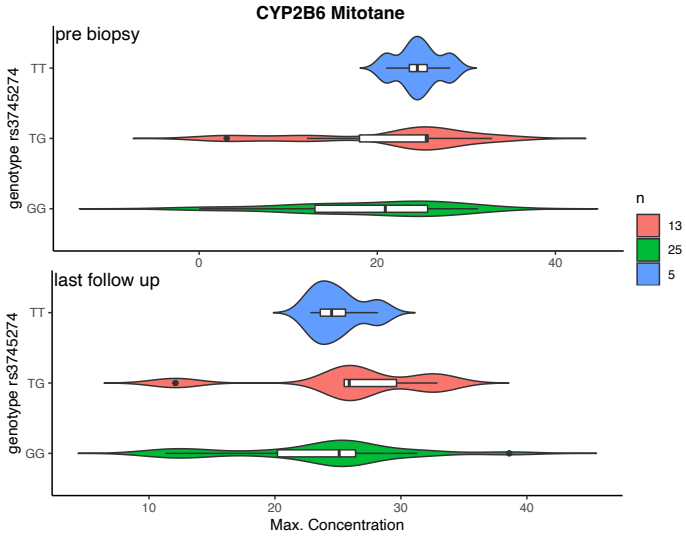


Figure 2.47: Mitotane concentration in different *CYP2B6* rs3745274 genotypes.

Contributions: Matthias Kroiß and Veronica Teleanu designed the study, selected patients, and collected clinical data. Sebastian Pirmann performed genomic analyses and created the oncoprint, did the mutational signature analysis and immune cell deconvolution, and created

the respective figures.

2.6.3 Chordoma

Chordoma is a rare cancer (incidence of 0.18-0.84 per million per year [184]) that typically originates in the vicinity of bones of the skull, spine, and sacrum arising from residual notochord tissue, a structure formed during early embryonic development [185]. These tumors are slow-growing but locally aggressive [186] and despite surgery being the primary treatment, complete removal can be challenging due to the tumor's invasive nature and nearby critical structures [185, 187]. Ongoing research and clinical trials aim to enhance our understanding and treatment options for chordoma, and organizations like the Chordoma Foundation [4] provide valuable resources and awareness for researchers and those affected by this rare cancer.

As part of a larger collaboration investigating multi-omics data of chordomas in the NCT/DKTK MASTER program, I performed genomic and transcriptomic analyses of these chordoma patients (n=103). Figure 2.48 shows the oncoprint with the genomic landscape of the chordoma cohort including small (SNVs, InDels) variants, sCNAs, structural variants, and gene fusion, both germline and somatic. Additionally, the oncoprint includes immune cell deconvolution results, from 68 available tumor RNA sequencing samples and the localization of the tumor to distinguish chordoma subtypes. For most recurrently mutated genes, the number of patients in which they were mutated was rather low. Variants in the most frequently mutated gene *CDKN2A*, mainly homozygous deletions, affected 32% of the cohort. Due to the frequent homozygous deletions of *CDKN2A*, a subcohort of chordomas was treated with palbociclib as part of a clinical trial (PMO1601:CDK4/6 inhibition in locally advanced/metastatic chordoma) which is currently still being evaluated (ClinicalTrials.gov Identifier NCT03110744[5]). The next most frequently mutated gene *FN1* already affected only 16% of the cohort. The number of germline variants was rather low and the proportion of small variants in the recurrently mutated genes generally only covers a small part of the cohort. Compared to the few small variants, many genes were affected by fusions like *FN1*, *ACAN*, *SASH1*, *SSP1*, *COL1A2*, and *EEF1A1*.

The cohort also showed frequent amplifications and deletions of several chromosomal regions. Most frequently affected by deletions was 9p21.3 (67%) followed by several regions on chromosomes 1, 3, 14, and 22. The complete profile of chromosomal aberrations analyzed with GISTIC [177] is shown in Figure 2.49. Deletions of chromosomal regions (chromosomes 1, 2, 4, 9, 10, 13, 14, 18, 21, and 22) were far more frequent than amplifications (chromosomes 1, 2, 7, and 19). Some of these alterations are known across many cancer

[4] https://www.chordomafoundation.org/

[5] `https://classic.clinicaltrials.gov/ct2/show/NCT03110744`

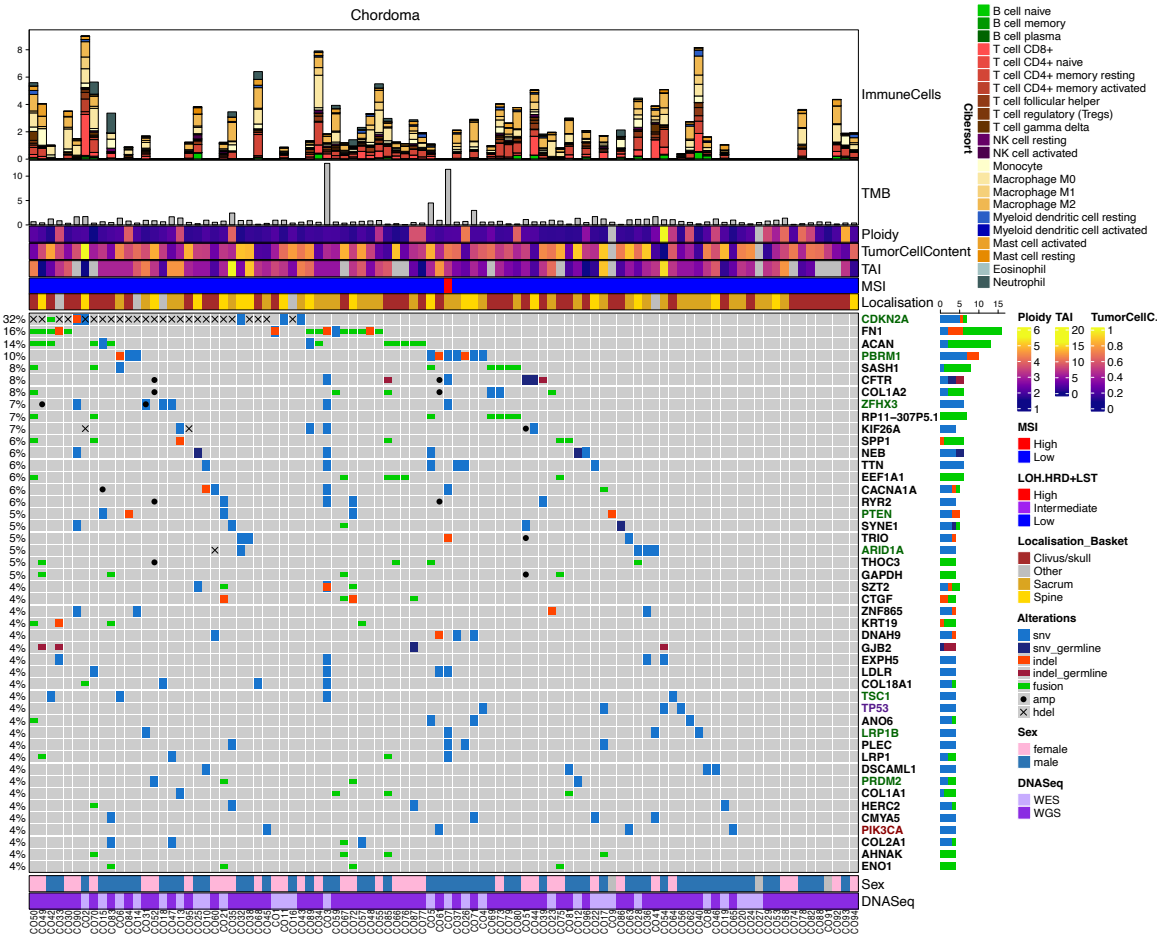


Figure 2.48: Oncoprint showing the genomic landscape of the chordoma cohort. The structure is similar to the oncoprints shown in the previous sections. The top annotation contains immune cell deconvolution results, tumor mutational burden (TMB), ploidy and purity of the tumor samples, telomeric allelic imbalance (TAI) score, microsatellite instability (MSI) score, and the tumor localization. The matrix in the center shows samples as columns and mutated genes (sorted by frequency of mutations in the cohort) as columns. The barplot annotation on the right shows the aggregated number of all types of mutations per gene in the cohort.

types like 9p (*MTAP*, *CDKN2A/B*), 10q (*PTEN*), 13q (*RB1*), and 17p (*TP53*). In general, considerably more deleted than amplified genome regions were detected.

Based on the available bulk RNA-seq data of the tumor samples, I assessed the proportions of admixed and infiltrating immune cells in 68 cases by applying deconvolution algorithms as described earlier [164]. Results generated with cibersort [170] are shown in the top annotation of Figure 2.48. The total number of admixed immune cells differs considerably between patients. The immune cell deconvolution mainly showed an admixture of macrophages, in a few cases a very high amount of T cells, and a small amount of B cells in some cases.

I then also used the bulk RNA-seq data to assess differentially expressed genes between localization subgroups, comparing spine/sacrum (n=33) and skull-based (n=27) chordomas. Figure 2.50 shows a volcano plot with differentially expressed genes between the groups. Genes on the left-hand side were upregulated in spine/sacrum-based chordomas while genes

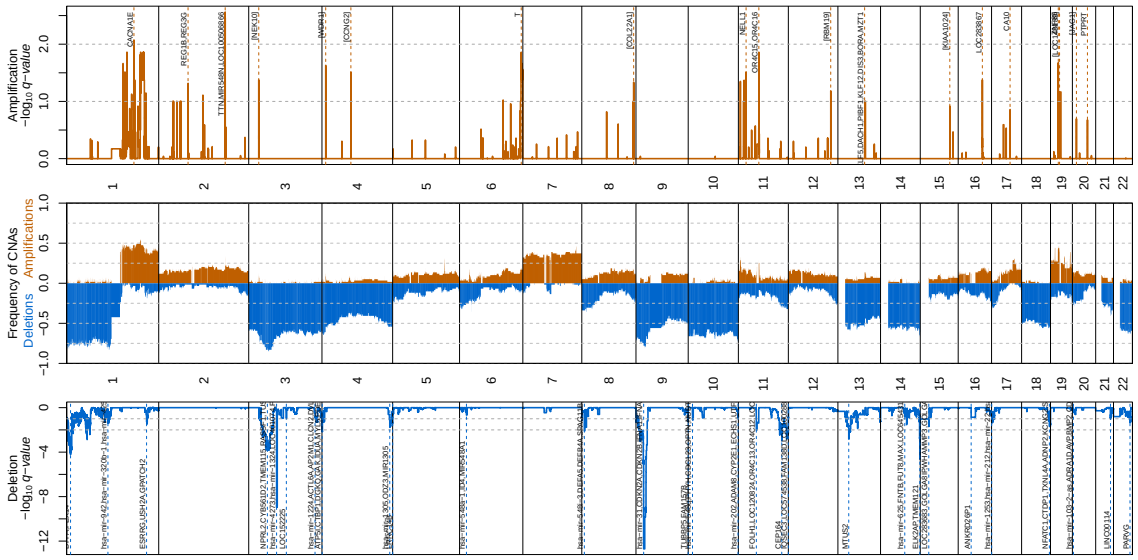


Figure 2.49: Recurrently amplified and deleted genomic regions in chordomas as detected by GISTIC. The top track shows the most significantly amplified segments and affected genes, with significance (q-values) on the y-axis. The track in the center displays aggregated frequencies of amplifications and deletions in the whole cohort, distributed over chromosomes, showing the whole genome pattern of frequent sCNAs. The bottom track shows the most significantly deleted segments and affected genes, with significance (q-values) on the y-axis (mirrored).

on the right-hand side were upregulated in skull-based chordomas. The most differentially expressed genes that were highly expressed in skull-based chordomas include *SLN*, *COX6A2*, *MYL2*, whereas genes expressed in spine/sacrum chordomas included *PTCHD2*. Additionally, gene set enrichment analysis of the list of differentially expressed genes between spine/sacrum and skull-based chordomas was performed with the R packages cola [188] and simplifyEnrichment [189]. The enrichment analysis was run with gene Ontology (GO) terms. These enrichment terms represent the rows and columns of the matrices in Figures 2.51 and 2.52. The terms are then clustered by semantic similarity in the GO tree, allowing for recurrence assessment and display in word clouds, with words displayed larger being more frequently represented in the gene list. Common terms enriched in both chordoma types are from genes related to development, organization, proliferation, transport, and regulation. Terms exclusively present in the skull-based chordomas are cytokinesis, and viral. Terms exclusively present in the spine/sacrum-based chordomas are apoptotic, adhesion, and repolarization.

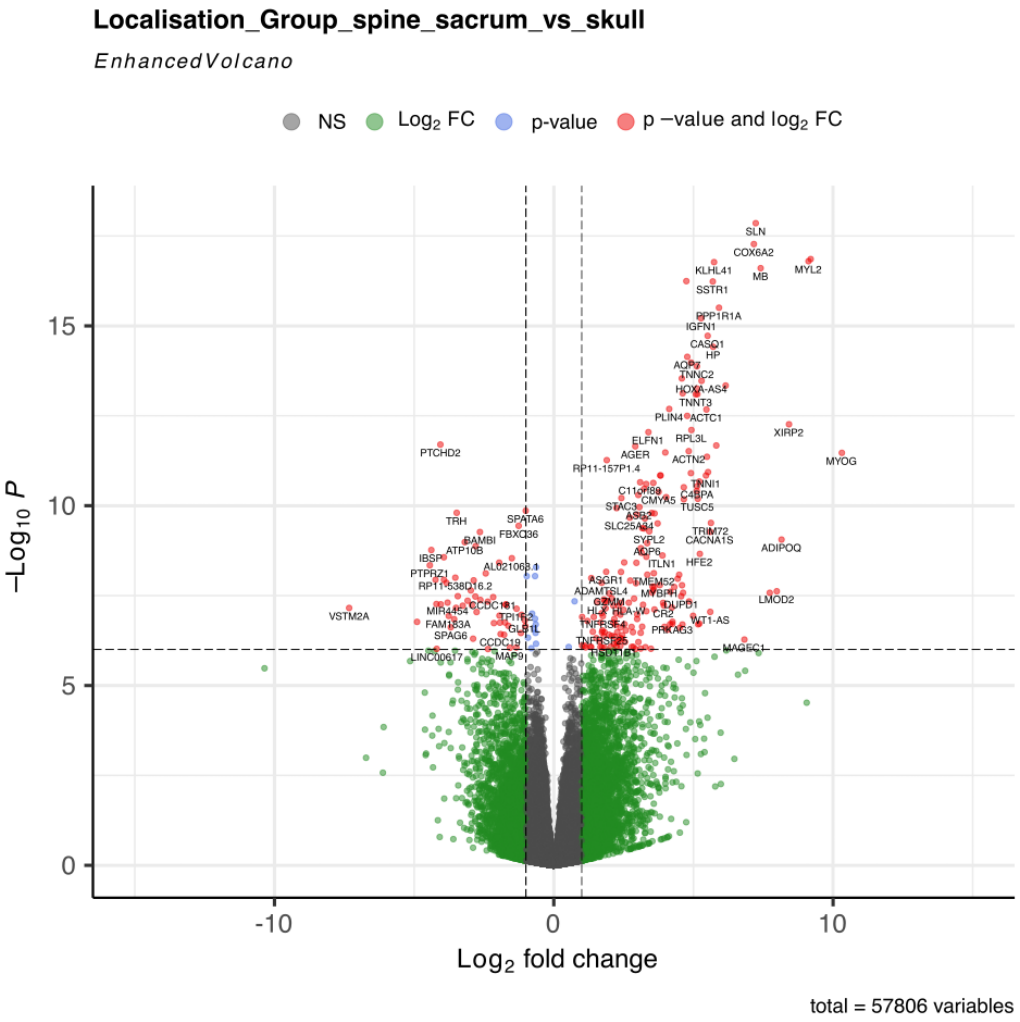


Figure 2.50: Differentially expressed genes between spine/sacrum and skull-based chordomas. The y-axis shows the significance and the x-axis shows the log2 fold change of expression values between the groups. Genes upregulated in spine/sacrum-based chordomas are shown on the left, and the ones upregulated in skull-based chordomas on the right.

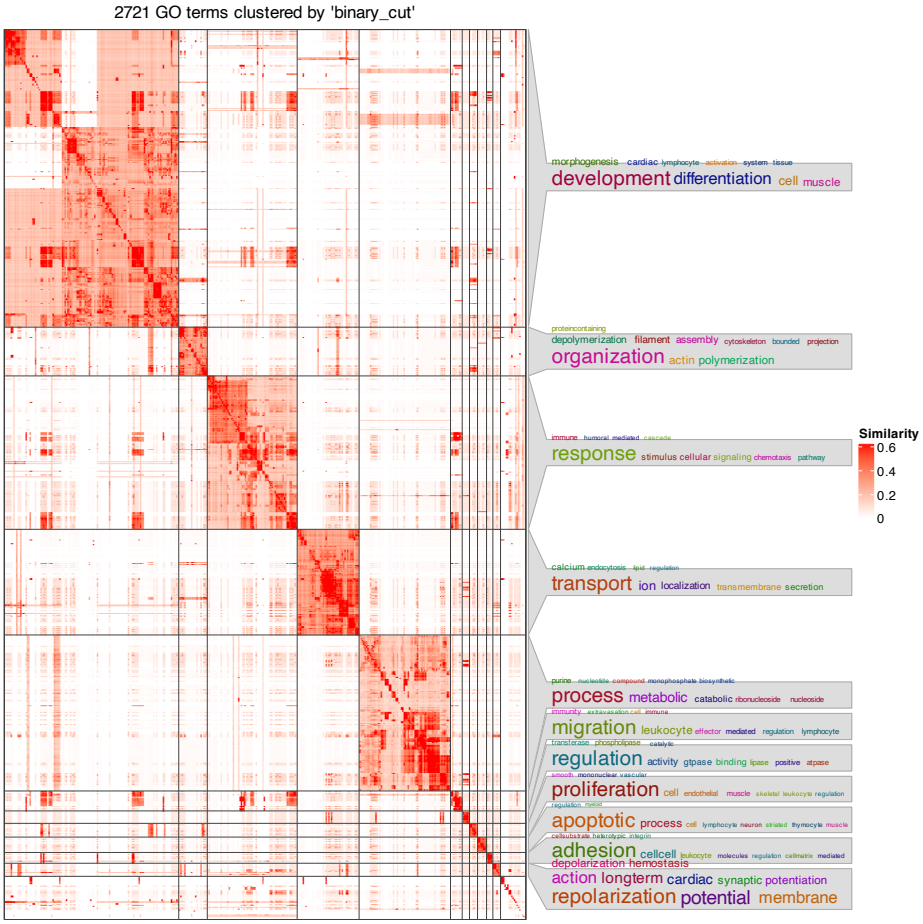


Figure 2.51: Gene set enrichment analysis of genes upregulated in spine/sacrum-based chordomas from differential gene expression analysis. The word clouds show GO terms associated with the list of upregulated genes, with font size according to the significance of the terms.

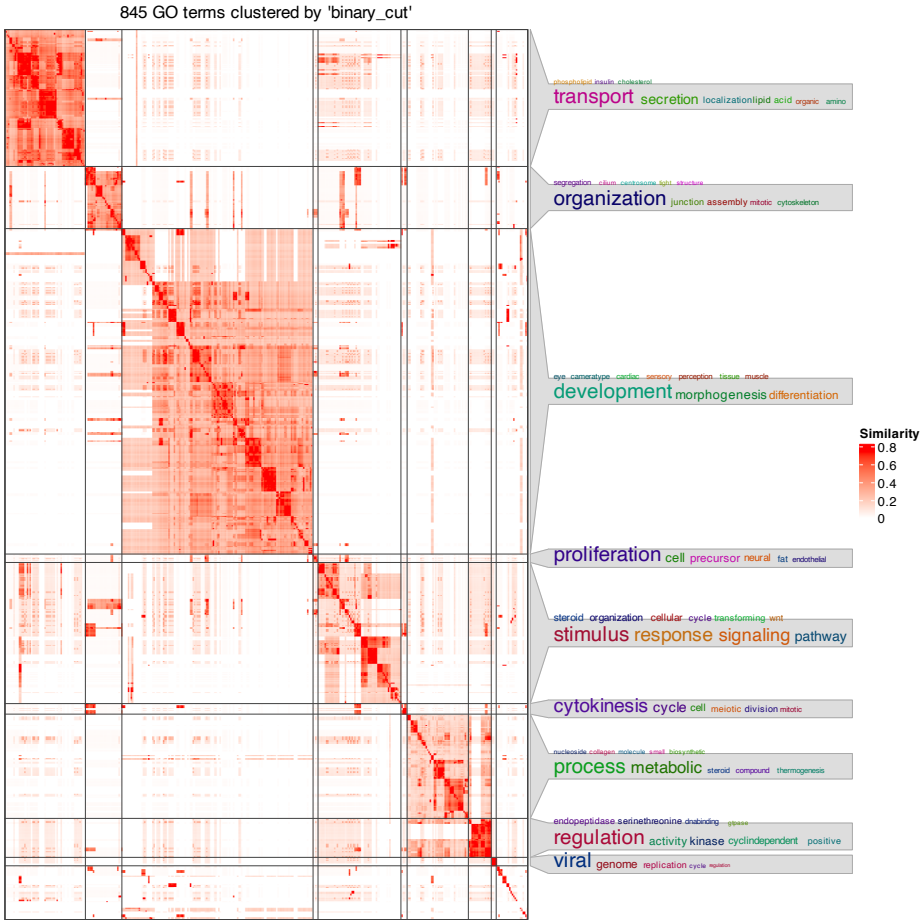


Figure 2.52: Gene set enrichment analysis of genes upregulated in skull-based chordomas from differential gene expression analysis. The word clouds show GO terms associated with the list of upregulated genes, with font size according to the significance of the terms.

PGx analysis of all matched control samples (peripheral blood) of the chordoma patients was performed using the PGx pipeline, described in section 2.2, analogous to the PGx analysis of the whole MASTER cohort. I analyzed variants of pharmacogenes associated with effects related to toxicity, efficacy, or metabolism on drugs important for anticancer/chordoma therapies and supportive medications. Genotype and resulting phenotype information on key pharmacogenes such as *CYP2D6*, *CYP3A4/5*, *TPMT*, and *UGT1A1* was obtained from the pipeline results. Results for 10 selected pharmacogenes (*CYP2B6*, *CYP2C19*, *CYP2C9*, *CYP2D6*, *CYP3A5*, *DPYD*, *SLCO1B1*, *TPMT*, *UGT1A1*, *VKORC1*) in 61 samples were validated using a quantitative PCR (qPCR) assay covering 60 of the most frequent variants. There was a >99% concordance between the results inferred from WGS and the genotypes generated by qPCR showing the accuracy of the ensemble genotyping approach of the PGx pipeline. The detected variants allowed the investigation of functional phenotypes. 16 patients with functional *CYP3A5*1/*3* or *CYP3A4*1/*22* diplotypes were found that may correlate with toxicity due to altered metabolism and drug concentrations compared with non-carriers for several drugs, including tyrosine kinase inhibitors. Furthermore, four homozygous carriers of *UGT1A1* genotype *28 susceptible to adverse effects during pazopanib treatment were identified among the 61 patients. Moreover, approximately one-third of the patients carried a *CYP2D6* genotype for which guideline recommendations are available.

Contributions: Sebastian Pirmann performed the genomic and transcriptomic analyses and created all Figures. Roman Tremmel performed the PGx genotype validation with quantitative PCR (qPCR) assays.

Chapter 3

Discussion

In drug-based anti-cancer therapies, the effectiveness of the therapy and the minimization of undesirable side effects are of paramount importance. Both pharmacokinetic (pertaining to how the body processes the drug) and pharmacodynamic (concerning what the drug does to the body) factors determine drug response. Genes encoding drug metabolizing enzymes and drug transporters play an important role and influence drug dose, adverse drug events, and the efficacy of drugs [26]. Regarding to cancer therapies, ADME processes can be separated into two categories based on anatomic localization: processes in healthy tissue, mainly in the liver of cancer patients, which are encoded by the germline genome and variation of which can be attributed to germline variation, vs. processes in cancer tissue, the regulation of which is additionally modulated by somatic variation. Germline variation in ADME genes influences the systemic metabolism of cancer drugs and thus the achieved effective dose. This germline variation thus influences potential side effects or may even lead to non-response [23, 47, 77–79], given that most cancer drugs have a narrow therapeutic index [47, 48]. Examples include variants leading to *DPYD* deficiency, which in turn can result in severe toxicity during 5-FU treatment [190], or variants in *UGT1A1* that are associated with a higher likelihood of side effects like diarrhea and neutropenia from irinotecan [33]. Moreover, in order to achieve the desired effect, most cancer drugs must reach a certain dose in the tumor cells. This means that in addition to the systemic, mainly hepatic metabolism, intratumor processes can affect the efficacy of a drug. These may be increased efflux from the cancer cell, increased drug inactivation, or decreased active uptake [41, 42, 56–58, 58–64].

Therefore, the primary aim of this thesis was to comprehensively describe the germline and somatic ADME profiles of cancer patients from the NCT/DKTK MASTER precision oncology program using NGS data including genomic, epigenetic, and transcriptomic data. A set of 60 pharmacogenes was curated for the analysis based on the following criteria: genes for which star alleles have been defined and which are supported by genotyping tools, as well

as the most important genes for which guidelines exist. The breadth of the DNA sequencing data (WGS) allowed the detection of complete genomic profiles. In the germline, this included rare individual variants that were predicted to be functional by in-silico tools beyond known functional PGx alleles. In the tumor samples, a complex interplay of somatic genetic variation (SNV, sCNA), epigenetics, and expression of pharmacogenes was observed. These combinations and alterations may have various recently discussed [57, 60, 62, 63] and yet-to-be-deciphered effects on therapeutic outcomes by driving cancer drug resistance or increasing targetability. All the data layers assessed in this thesis helped to further investigate the pharmacogenomic landscape of cancer and to obtain an as accurate as possible picture of the pharmacogenomic profiles of cancer patients.

The objective of the second part of this thesis was to investigate the genomic and transcriptomic landscape of several rare cancer entities based on NGS data. In three projects, parathyroid carcinomas, adrenocortical carcinomas, and chordomas were investigated. These cohort analyses showed that a comprehensive molecular analysis and description of these diseases can decipher disease-specific mechanisms on a molecular level, give a rationale for implementing available targeted therapies, or even point to new therapeutic options.

3.1 Pharmacogenomic Analysis Pipeline

The analysis of pharmacogenomic variation for clinical and research purposes has progressed in recent years from array-based genetic testing of individual genes for specific variants to a comprehensive pharmacogenomic characterization based on NGS [191, 192]. However, comprehensive direct comparisons of the clinical benefits of NGS vs. panel-based PGx genotyping are still lacking and the number of studies to date is small [193]. Although WES data is commonly available and clinically used, in this thesis only WGS was considered as it is inherently more comprehensive, given that some star alleles (e.g. *CYP2C19*17*, *CYP3A4*22*, or *CYP3A5*3*) include regulatory upstream, intronic, or splicing variants. These variants were not sufficiently covered by WES in the analyzed MASTER samples as shown in section 2.2 and in particular Figure 2.4. These limitations have also been observed by others [194, 195].

Several computational tools have been developed that facilitate PGx analyses directly from NGS data [72–76, 196, 197]. A comprehensive overview by colleagues and myself can also be found in [198]. My aim was to develop a PGx analysis pipeline that incorporates many of these tools to determine consensus genotypes for 60 pharmacogenes based on WGS data. The applicability of this pipeline was tested in the MASTER cohort for germline as well as matching tumor samples. Although the concordance of some of the tools was previously tested by others and high accuracy with orthogonal methods was reported [199], I observed

significant gene-dependent differences between the tools integrated in the pipeline (Table 2.1). Therefore, I developed a consensus approach that compensates for the differences of the individual tools wherever possible. The evaluation of the PGx pipeline indicated that this consensus approach, i.e. an ensemble of several algorithms, represents an advantage over the use of a single tool and increases the reliability and concordance of the results. This was also observed in previous efforts to combine PGx results of several computational tools. A study by Tafazoli et al. [200] reported similar problems in combining PGx results which differed between tools due to their implemented variant sets. To resolve these issues they have implemented a majority rule and in the case that this does not apply, the Stargazer results are used as default. In my pipeline, however, if the discrepancy between the tools cannot be resolved, a more conservative approach was chosen and no final result is provided. In addition, their study is only based on 100 exomes, which, as I described at the beginning, do not cover some important PGx variants compared to the genomes used here. The number of patients used in their study was also substantially lower than in this work.

For instance, I observed discrepancies due to different naming conventions between the used PGx tools. By resolving this issue the concordance of several genes was increased (e.g. for *CYP2D6* from 60 to 98%). However, some discrepancies remain, like for *CYP4F2* only a minor increase in concordance was achieved (69 to 73%). The reason is the discrepancy in phasing results for the alleles *2, *3, and *4, highlighting one of the limitations of short-read NGS. Besides phasing of SNVs, further limitations include errors of read alignment in repetitive/homologous regions as ADME genes have many pseudogenes, and a reliable calling of CNVs from coverage data. Novel technologies such as long-read sequencing could resolve these issues [201]. Of note, most of the tools have also recently incorporated long-read NGS analysis as an option, however such data was not available in MASTER. In addition, the genotyping tools showed inherent problems when applied to tumor samples, which will be discussed in more detail below.

Of note, another high-quality PGx tool has recently been developed called PharmCAT[1] (Pharmacogenomics Clinical Annotation Tool) [202–204]. Since the prerequisite input for this tool are reads aligned to the reference genome GRCh38, I have not implemented it into the PGx pipeline. A realignment or likely error-prone liftover was not feasible for the large number of MASTER samples. However, since the new reference genome will also be established in the MASTER program in the near future, PharmCAT could also be integrated thereafter.

Finally, the implementation of the pipeline in Nextflow allows easy portability and use on various computational systems. In addition to retrospective pharmacogenomic analyses of large patient cohorts like MASTER, which can be performed rapidly with the pipeline, as

[1] `https://pharmcat.org/`

the parallelization approach enables fast processing, there is also the possibility to apply this pipeline in an ongoing clinical setting and to derive PGx recommendations from databases like CPIC. This was demonstrated by integrating the pipeline into the molecular tumor board workflow of NCT/DKTK MASTER. In summary, the results have shown that in silico genotyping methods are very useful and accurate for germline samples, even though there remains some effort needed for standardization, inclusion of rare variants, and confirmation by adding orthogonal methods. Nevertheless, the pipeline results for the MASTER patients showed that the repurposing of NGS data with respect to pharmacogenes is comprehensively possible.

3.2 Germline Pharmacogenomics in MASTER

Germline pharmacogenomic variation has been linked to inter-individual variability of drug response and side effects of many anti-cancer drugs [44, 205]. Furthermore, there is some evidence that altered metabolism of carcinogens may promote the development of cancer [206]. The MASTER cohort represents a very diverse group of cancers and is well suited to study pan-cancer germline pharmacogenomic variation.

PGx genotyping results and actionable variants

The consensus genotyping results of the PGx pipeline showed that 96.4% of patients had at least one gene that carried an actionable genotype. On average the MASTER patients carried two actionable genotypes across the 60 analyzed genes (with up to 6 genes maximum in a few patients (Figure 2.6)). These numbers were in line with frequencies observed in other cohorts with samples of European ancestry [26] and demonstrate the importance of pharmacogenomic genotyping of cancer patients in a clinical setting. The PGx pipeline analyses several genes that are relevant for anti-cancer drug therapies including *CYP2D6*, *TPMT & NUDT15*, *UGT1A1* and *DPYD* for which germline genotype results could be determined in MASTER. For these genes, guidelines with treatment recommendations for cancer drugs have already been established by international consortia like the CPIC (Clinical Pharmacogenetics Implementation Consortium) or the DPWG (Dutch Pharmacogenetics Working Group).

CYP2D6 influences the effective concentration of the prodrug tamoxifen by metabolizing it into its active form endoxifen. Tamoxifen is commonly used for the treatment of hormone receptor-positive breast cancers and is a type of selective estrogen receptor modulator that works by blocking the growth effects of estrogen in breast tissue and *CYP2D6* poor/intermediate metabolizers are anticipated to exhibit reduced endoxifen concentrations [122, 207–210]. 46.2% of breast cancer patients in MASTER had poor/intermediate metab-

olizer status for *CYP2D6*.

Variants in *TPMT* and *NUDT15* have been shown to affect the metabolism and toxicity of thiopurine drugs, such as mercaptopurine (6-MP) and azathioprine [211–215]. These are commonly used in cancer therapy, particularly in the treatment of hematologic malignancies, such as leukemia, due to their antimetabolite and immunosuppressive effects. In MASTER, about 8% of patients were *TPMT* intermediate or poor metabolizers being susceptible to thiopurine-related toxicities. The proportion in hematopoietic cancers was 8.75%.

DPYD plays an important role in the breakdown of fluoropyrimidines into inactive metabolites. Fluoropyrimidine drugs, such as 5-fluorouracil (5-FU) and capecitabine, are used in cancer chemotherapy, particularly in the treatment of colorectal and breast cancer and other solid tumors. However, in patients with decreased *DPYD* functionality, there is an increased risk of elevated levels of active 5-FU, which can result in severe and potentially life-threatening toxicities [190, 216–218]. Therefore, current guidelines recommend a 25-50% dose reduction for *DPYD* intermediate metabolizers [217]. In the MASTER cohort, about 6% of patients were DPYD intermediate metabolizers and these comprised 6.3% of breast and 3.7% of colorectal cancer patients.

UGT1A1 is involved in irinotecan metabolism and variants have been associated with the risk of severe toxicity, such as (febrile) neutropenia or diarrhea [33]. For poor metabolizers (12% in MASTER) it is recommended to reduce the starting dose to 70%. Additional genes with available CPIC or DPWG guideline recommendations that were analyzed by the PGx pipeline and are related to anti-cancer therapies or supportive medications include *ABCG2*, *CACNA1S*, *CFTR*, *CYP2B6*, *CYP2C9*, *CYP2C19*, *SLCO1B1*, and *VKORC1*.

An implementation strategy based on the aforementioned germline results was pursued in the context of this thesis with the integration of the PGX pipeline in the prospective workflow of the molecular tumor board (MTB) of the NCT/DKTK MASTER precision oncology program. The resulting benefit and outcome evaluation of this new functionality in the MTB have to be investigated in the future. As of now, no statement can be made regarding the application, utility, or any other endpoint. The retrospective results from the MASTER cohort have shown that there is generally great potential for comprehensive pharmacogenomic analysis of NGS data with in silico methods, which can provide valuable information about known actionable germline variants in a therapeutic setting.

Among the recommended interventions in MASTER, therapies related to tyrosin kinase signaling made up 35% [68]. Such tyrosine kinase inhibitors (TKIs) are extensively metabolized by isozymes of the *CYP3A* family and *UGT1A1* and are actively transported by *ABCB1* [219]. However, a guideline is not yet available. Nevertheless, variants of *CYP3A4/5* and *ABCB1* membrane drug transporters have been reported to affect resistance and toxicity to TKIs like, e.g., sunitinib, imatinib, or pazopanib [220]. My analysis showed that 8% of

MASTER patients carry at least one reduced functional *CYP3A4* allele (*22) and that 13.6% are expressors of *CYP3A5* (at least one *1) with potential effect for the drug therapy. Further studies are warranted to elucidate the impact of these genetic variants on the plasma levels of TKIs as well as metabolites and the ADR profile.

Since the available clinical data in MASTER is sparse and it is a last-line setting where the patients have had many pre-treatments and where a very diverse set of targeted therapies is applied, there was no opportunity to go into a deeper analysis of germline-related drug response and side effect mechanism in these cancer entities. Another limitation of this cohort is the current classification of entity baskets in the MASTER program, as some groups (like soft tissue sarcoma) combine various diseases and it would be better to subdivide them even more precisely. Such efforts are currently ongoing based on OncoTree [221] which means that subgroups of individual entities will be categorized even more precisely in the future and additional analyses can be carried out thereafter.

PGx variants and cancer susceptibility

The distribution of germline metabolizer and transporter phenotypes between the cancer entities in MASTER was quite similar, however, previous studies reported associations between certain alleles and susceptibility to cancer such as for *CYP2A6* and lung cancers [222]. Furthermore, genetic variations in *CYP1A1* and *CYP1B1* genes can reduce the detoxifying metabolism of polycyclic aromatic hydrocarbons (PAHs) and therefore contribute to higher susceptibility to cancers associated with PAH exposure, such as lung cancer. Hence, the efficiency of detoxification processes is influenced by genetic variation in ADME enzymes, which can affect an individual's ability to eliminate environmental substances, including carcinogens. Certain environmental substances are pro-carcinogens, which can be converted into carcinogenic forms through metabolic processes. Gene-environment interactions can contribute to the variability in cancer risk observed among individuals with similar environmental exposures. In summary, many mechanisms for the metabolism of exogenous and endogenous substances suspected of causing cancer have already been described but study findings are inconsistent and have been debated [150, 223–229]. Nevertheless, my analysis showed an enrichment of normal metabolizers of *CYP3A5* in hepatopancreaticobiliary cancers, ultrarapid metabolizers of *CYP2D6* in upper gastrointestinal cancers, and intermediate metabolizers of *DPYD* in cancers of unknown primary (CUP).

Germline copy number variants

The analysis of germline CNVs has shown that very rare variants like the *CYP1B1* whole gene deletion can be found if a sufficiently large cohort is used for analysis. However, these CNVs could only be detected by Stargazer (of the tools in the pipeline), as it is the only tool

that has implemented a generic CNV detection for all genes based on coverage information. The rare deletion of the whole *CYP1B1* gene and the resulting loss of function has been associated with primary congenital glaucoma [230, 231]. Also, some rare CNVs that have been previously described, like partial deletions of *CYP2C19* and *SLCO1B3* [232], were found in MASTER. These results indicate that it is not sufficient to include only known common CNVs in the genotyping tools, as CNVs contribute to pharmacogenomic variability to a high degree and are well studied in only a subset of ADME genes [81, 88].

Additional germline variants and variant effect prediction

A large part of the inheritable variation in drug response isn't explained by common variants, suggesting other genetic factors are significant. In the last years, rare genetic variants, which make up over 90% of genetic diversity in pharmacogenes, were suggested as a main contributor [83]. Even though the variant effect prediction tools used in this work only allowed the prediction of non-functionality, i.e., damaging effects, and I therefore was not able to analyze the potential effect of rare variants on increased protein function, I did analyze the occurrence of additional SNVs beyond the known star alleles in the WGS data and assessed their impact on protein function. Furthermore, I combined the predicted damaging rare variants with the functional haplotypes/star alleles to better assess their superimposing non-functional effect and their relevance for drug response or ADR risks, which, to my knowledge, is an innovative approach and has not been published so far.

The results in MASTER showed that there is an enormous number of rare variants (99.9% with MAF <1%). 24.7% of them were exonic missense variants and half of these were predicted to be damaging to protein function showing that they can have functional effects. The analysis of rare variants in relation to common known alleles showed that in 27.6% of the patients affected by damaging rare variants, these occur in combination with at least one reduced or non-functional allele (e.g. *CYP2D6*4*, *CYP3A5*3*). In contrast, I observed 94.5% of the patients with at least one damaging variant in combination with normal or increased function alleles. Furthermore, in 109 patients (5.28%), at least one of the damaging variants was homozygous, affecting both copies of the respective gene. In these individuals, the actual underlying genotype and therefore the resulting protein function may have been incorrectly determined due to the existence of additional non-functional rare variants. These findings are extending published work that has illustrated the importance of taking rare individual variants in pharmacogenes into account [83, 233–235]; however, underlying star alleles had not been taken into account there. Generally, I observed that none of the rare damaging variants seem to be linked to known star alleles suggesting individual interfering effects on protein function. A complete list of predicted damaging variants in the germline samples of MASTER can be found in Table 6.5 in the appendix.

The contribution of rare variants was particularly evident in 49 *CYP2D6* ultrarapid metabolizers in which 3 different damaging variants were found that could interfere with the actual phenotype (rs3915951, rs200229206, rs141739595). One of these three SNVs, rs3915951, was recently found in denisovan genomes and a papuan individual and was predicted to likely impact enzyme function [236]. The other variants rs200229206 and rs141739595 were present in gnomAD but have not been described or functionally analysed elsewhere.

Due to the continuous decrease in sequencing costs, NGS data is increasingly used in routine clinical settings. Consequently, the functional prediction of rare variants, whether through in silico, in vitro, or in vivo methods, is becoming more important. A lot of work is still needed to thoroughly assess the function of these variants in order to include them in future guidelines so that patients can benefit from them. Methods for a comprehensive functional analysis of pharmacogenomic variants of unknown significance have been previously reviewed by colleagues and myself [198]). The comparison of the variant effect prediction tools on the SNVs in MASTER (Figure 2.11) has shown that optimizing such tools for the respective area of application like pharmacogenomics offers advantages over more generalized prediction models, as the general concordance was moderate. This has also been demonstrated in previous studies [91, 237]. The functional analysis of pharmacogenomic variants in silico has made great progress in recent years with a plethora of available tools and will continue to improve as the capabilities of artificial intelligence increase and VEP methods are further optimized. However, these results must always be validated in the laboratory with orthogonal methods before these functional variants are used in clinical practice, a necessity also highlighted by the moderate overlap of predictions between AlphaMissense [99], the current state-of-the-art, and the domain-specific APF framework. Generally, the functional validation of the selected *SLCO1B1* variants has shown that in silico methods are suitable for discovering new relevant variants and predicting their effect. This allows to follow a rational step-wise approach: the enormous number of variants that are potentially interesting and relevant for further studies can initially be limited and filtered, to keep the subsequent laboratory work manageable. The analysis of the abundance of additional rare germline SNVs has clearly shown how important a comprehensive pharmacogenomic characterization of the patient is and that the current collection of star alleles has to be extended to account for such variants.

Some pharmacogenes were more frequently affected by germline SNVs than others. The number of additional variants between the genes was highly variable, which is related to both the gene length and the degree of research into the genes to date (e.g. for *CYP2D6*, the most studied pharmacogene, a large number of variants have already been described and incorporated into star alleles, which is why only very few rare variants can still be discovered here).

3.3 Somatic Pharmacogenomics in MASTER

PGx studies in cancer to date investigated mostly germline variants in pharmacogenes, focusing on systemic pharmacokinetic and pharmacodynamic effects [26]. Despite FDA approval of some pharmacogenes as biomarkers [238, 239], notably excluding ABC transporters and most CYPs, the emphasis remains on germline variants rather than tumor-specific mutations, which are likely contributing to treatment resistance and poor response. However, the importance of considering somatic mutations has gained attention recently [57, 240]. Ensuring sufficient drug exposure within tumors is vital for the efficacy of systemic therapies with chemotherapeutic drugs or small molecules. Membrane transporters, primarily from ABC or SLC families, facilitate the transport of such drugs across cell membranes [240–242]. Germline variation in these genes is acknowledged to affect transporter activity and impact intracellular drug concentrations. In contrast, the potential effects of somatic genetic and epigenetic alterations in these genes during cancer progression due to somatic events were only recently recognized. These somatic variants could further influence drug exposure within tumors and it is conceivable that treatments with systemic cancer drugs, inducing somatic mutations, might result in the proliferation and outgrowth of resistant tumor subclones.

In this thesis, the somatic genomic (SNV, CNA), epigenomic, and transcriptomic variation in tumors was systematically investigated for the 60 selected pharmacogenes. As mentioned above, despite many current advances, the focus of pharmacogenomics in oncological clinical practice and research is still mainly on germline variants in ADME genes, and somatic variants are only studied as drug targets in cancer [47, 48]. Although a couple of studies investigated the association of somatic expression levels of ADME genes and clinical outcomes in cancers [37], a comprehensive description of somatic variation in ADME genes in tumor samples from a large cohort has not been carried out in this way before, as somatic variants have so far mainly been studied in drug targets and not in ADME genes. However, the importance of ADME variation in tumors has recently been recognized and discussed [57, 60, 62, 63]. This chapter discusses the findings of these somatic analyses.

PGx genotyping in tumors and limitations of genotyping tools

I assessed the extent to which the genotyping results of the PGx pipeline for the tumor samples differed from those of the matched germline due to somatic variation. These differences exceeded 15% for some genes, affecting 66% of the patients overall, as shown in section 2.4.1. Among the reasons leading to the captured star allele changes in tumors, I identi-

fied somatic SNVs as a minor effect, and sCNAs, especially LOH, as major contributors. Generally, there were limitations of the genotyping tools with respect to tumor WGS data. Subtotal tumor cell content (purity), a substantial portion of aneuploidy, as well as a missing stable reference locus for CNV calling (commonly *VDR* is used in the genotyping tools, but if itself affected by a sCNA, this choice is suboptimal) limited the applicability of the tools and even made them unsuitable for a reliable genotyping of tumor samples. This observation is based on my comparative analysis with somatic genomic profiles extracted from specially optimized pipelines for tumor samples covering varying tumor cell content and complex genomic alterations such as aneuploidy, sCNAs, and LOH. In the next sections, I will discuss these findings in more detail.

Most importantly, the observed genomic differences between matched germline and tumor samples underline the fact that tumor samples should not be used as a proxy for pharmacogenomic germline genotyping - a finding which is also mentioned in the CPIC guideline on tamoxifen [209]. This is based on studies that showed that the *CYP2D6* locus is frequently (15-41%) affected by LOH in breast cancer tissue [119, 121], which led to 19% of *CYP2D6*4* calls being discordant between tumor and matched normal control samples (buccal cells). However, as sCNAs including LOH were found to be very abundant in pharmacogenes in MASTER, this recommendation seems to be generalizable to many other ADME genes, also given that in 66% of MASTER patients, there were differing genotyping results between matched samples.

Another challenge for genotyping arises from the phasing of variants in the tumor. As shown in the germline results, PGx genotyping tools showed some problems with phasing variants from short-read sequencing. In the tumor, however, there can be arbitrary numbers of copies of a gene (due to aneuploidy or sCNAs) making accurate phasing even more challenging. Current efforts are undertaken to come closer to bioinformatic solutions to this problem and to assess how many copies of a gene in the tumor are affected by a variant (ZygosityPredictor [243]). Unfortunately, for a considerable fraction of the variants, the problem of phasing from short reads remains. If long-read sequencing or also a hybrid approach with both short and long reads were used here, phasing could be very much enhanced and a determination of the exact star allele configuration in tumor samples would be possible in most cases.

Somatic SNVs in pharmacogenes

I thoroughly investigated whether somatic SNVs affect known loci of star allele variants of 60 pharmacogenes using SNV data from the pipelines optimized for somatic SNV calling in tumor samples. I showed for the first time that mutations occur at known PGx SNV loci (star allele defining variants), however, compared to the number of sCNAs found in the tumors, these seem to only be minor contributors to somatic pharmacogenomic variation.

The proportion of these star allele changes in the MASTER patients was quite low (1% of patients carried such somatic variants). Although this was rarely observed in the MASTER cohort, the impact on the few affected patients could still be considerable, as in 27% of these patients a somatic mutation caused a metabolizer phenotype change compared to the germline. For non-PGx-SNVs, three times more germline than somatic SNVs were found in total. However, the ratio of predicted damaging variants was roughly the same in both cases (approx. 50%). Interestingly, the proportion of non-synonymous variants was higher in the somatic than in the germline SNVs (40% and 23%), which could indicate an accumulation of such variants in the tumor possibly due to a reduced selective pressure [244]. This finding also confirms previous results from WGS data on *ABC* and *SLC* drug transporters in a large cohort of metastatic cancer patients [57]. The study showed that somatic SNVs found in the coding regions of the transporters were mainly non-synonymous variants and also correlated with gene length. Another study analyzed the occurrence of somatic variants in *ABC* transporters and *CYP* genes in the TCGA and COSMIC datasets and discussed the potential role of these variants in drug resistance [240]. Among the top 30 mutated transporters and genes, they also found *ABCB1*, *ABCG2*, *CFTR* which underlines the results found in data from the MASTER program in this work.

Regrettably, the aforementioned studies missed any prediction on variant functionality. However, I showed that the ratio of predicted damaging variants in the tumor was roughly the same as in the germline, with half of the variants having damaging effects. This finding is of importance to refine further assessments on drug resistance. I observed also differences between the drug transporter genes. While for *ABCB1* about 55% of variants were damaging, the fraction of damaging variants in *ABCG2* was 83%. This finding is important since depending on which transporters are affected, the tumor cell could be more or less susceptible to a specific therapy [241]. The damaging variants could lead to reduced activity of the drug metabolizing enzymes and transporters in the tumor. For transporters, this could mean both a reduction in the uptake of the drug into the tumor cell and a reduction in its elimination. Depending on which transporters are affected, the tumor cell could be more or less susceptible to the therapy [241]. For the phase 1 and 2 enzymes, the damaging variants could lead to an increased effect of the drug, as it is degraded more slowly in the tumor cell and more dose remains at the site of action [41,56]. In contrast for prodrugs opposite effects could be plausible.

Somatic CNAs in pharmacogenes, entity-specific patterns and relation to cancer drivers

In this thesis, I observed that somatic CNAs are abundant and very frequently affecting pharmacogenes in tumors. In contrast, only little has been described in other studies about the amount of contribution of sCNAs to somatic pharmacogenomic variation [57]. However,

this is modulated by different, not standardized definitions of sCNAs. The chosen definition in this thesis is quite broad and includes both small focal events and large segments of the size of chromosome arms. This definition is based on the ACESeq pipeline [107], which merges neighboring regions of the same copy number into larger segments when segmenting the whole genome. The set of investigated genes could be separated into 3 groups and showed that for most genes there is a cohort-wide tendency for either duplications (e.g. *ABCB1*, *CFTR*, *CYP3A* family, or *PTGIS*) or deletions (*ABCG2*, *CYP2C* family, *NAT1/2*, *NUDT15*) reflecting the genome-wide pan-cancer pattern of sCNAs previously reported [104].

In addition to the pan-rare-cancer analysis, the analysis of sCNA events in the different cancer entities revealed entity-specific patterns of somatic pharmacogenomic CNAs that mainly follow the reported genomic/chromosomal sCNA profile of these cancers. This precise description of somatic pharmacogenomic CNAs in the various cancer entities has not been done before and is a major finding of this thesis. Examples include deletions on the q arms of chromosomes 10 and 13 with resulting LOH in leiomyosarcomas, deletion on chromosome 1p with LOH in GIST, amplifications of chromosome 7q in colorectal cancers, and deletions and LOH of regions on chromosomes 8 and 22 including *CYP2D6* in breast cancer [111, 116, 119–122]. The frequency of LOH of *CYP2D6* in breast cancer in MASTER was about 50% and previous studies reported frequencies of 15% for *HER2*-positive, 35% for *ER*-positive, and 41% for *ER*-negative breast cancers in TCGA [121]. The frequently amplified regions q21-22 on chromosome 7, including the *ABCB1* efflux transporter and *CYP3A* genes found in MASTER, had already been reported to contribute to drug resistance in cancer cells [245].

Due to the observed abundance of sCNAs in the pharmacogenes, I analyzed whether these might be related to sCNAs in neighboring oncogenes and tumor suppressors, which are common cancer driver events. The hypothesis was that deletions of pharmacogenes co-occur more frequently with tumor suppressors and duplications with oncogenes in their vicinity. This analysis has shown that the somatic pharmacogenomic CNAs could at least partially be passenger events, arising due to other oncogenic processes. Regions dense in oncogenes or with single very impactful oncogenes were more frequently duplicated affecting the closely located pharmacogenes. This was demonstrated for chromosome 7q including *ABCB1*, *CYP3A* genes, and oncogenes *GRM3*, and *TRRAP*, as well as *SLCO1B1/3* which are close to *KRAS* on chromosome 12. Accordingly, *ABCB1* and *SLCO1B3* have also recently been reported to be most frequently affected by somatic sCNAs among drug transporters [57]. The results of the oncogene analyses presented here thus provide a potential explanation for this phenomenon. Furthermore, the role of *SLCO1B3* in resistance to cancer treatment and in precision oncology has been discussed, suggesting exploitation of the overactivity of influx transporters in tumors as a therapeutic option to increase intracellu-

lar drug levels, or demonstrating resistance by deactivation of such transporters [246, 247]. *SLCO1B3* was duplicated and most expressed in colorectal cancers, however, it was demonstrated in colon cancer cells that they harbor a specific variant of *SLCO1B3* that has reduced transport ability resulting in reduced uptake of substrates [248]. In colorectal cancers, *PTGIS* was amplified together with *SRC*, for which increased expression in colorectal cancer was suggested to increase metastatic activity and lead to chemotherapy resistance through various signaling pathways. Therefore, blocking *SRC* could potentially be beneficial in treating colon cancer [127]. High expression and amplification of *PTGIS* have also been associated with poor prognosis in colorectal cancer patients through various cancer-promoting effects [130]. Complementarily to co-amplifications with oncogenes, the relationship between deleted pharmacogenes and deleted tumor suppressors was also impressively demonstrated in the MASTER results, for example for *NUDT15*, which was frequently deleted in leiomyosarcoma, where deletions of neighboring *RB1* and *BRCA2* are common events [111]. Cohort-wide *XPC*, which is important in nucleotide excision repair [249,250], was rather affected by deletions. It could be possible that the deletions in the tumors lead to an increased mutation rate and thus to tumor progression. In contrast, *XPC* was frequently amplified in breast cancers in MASTER. An increased activity of this gene could indicate a coping mechanism of the tumor against chromosomal instability or chemotherapeutic drugs.

Tumor expression of pharmacogenes

Regarding the expression of ADME genes in tumors there have only been a few studies which mainly compared tumors with corresponding healthy tissue. The analysis of ADME gene expression from tumor RNA-Seq data in this work revealed that a large proportion of them seem to be active in tumor tissues. Generally, phase II genes and transporters were more expressed than phase I genes. While some genes were highly expressed across all cancer entities like *GSTP1*, other genes showed a more tissue-specific pattern. Overexpression of *GSTP1* has been previously reported in several cancers [135, 136]. GSTs are involved in drug resistance either via increased detoxification or by inhibiting the MAPK pathway and resulting apoptosis [43]. *POR*, which was also among the highly expressed genes, had been described to influence tamoxifen-resistance in breast cancer via the STAT1/c-Myc pathway [251]. Other more specific genes in the group of highly expressed genes included *XPC* which is important for the DNA nucleotide excision repair pathway and the cause of Xeroderma pigmentosum [249, 250] and, as described above, it has been suggested that expression of *XPC* could provide a coping mechanism against the intended chromosomal instability resulting from chemotherapeutic drugs.

An analysis of correlation between DNA alterations and gene expression showed a significant influence of sCNAs on expression for 32% of the 60 genes in an entity-dependent

manner. Although at the level of pharmacogenes, these sCNAs are likely to be passenger events, they do affect RNA expression of ADME genes and could thus have an impact on intratumor drug transport and metabolism. Correlation of expression and sCNAs was recently also reported for drug transporters [57]. As described earlier, drugs and xenobiotics are mostly metabolized in a small set of drug-metabolizing tissues, inlcuding the liver and the colon. In the gene expression analysis carried out here, the expression of pharmacogenes was generally higher in those cancers whose cells of origin were in drug-metabolizing tissues. The fact that in addition, ADME genes were also expressed in tumors in which one would not expect it based on tissue of origin may have various causes. Theoretically, e.g., this may be explained the same way as for other genes in cancer, i.e. by loss of genomic imprinting through epigenetics [154]. However, no comprehensive comparison could be made with the healthy tissue of the respective tumor entities, as MASTER has such rare tumor entities that for many the tissue of origin is not even known. It is therefore not possible to say whether the expression shown here is increased or decreased compared to normal tissue, meaning that this analysis only allows a comparison between the different tumor entities.

The analysis of promoter and intragenic methylation of ADME genes in the tumors (see below) showed that there are strong differences both at the levels of different genes and of different entities. Generally, the differences in methylation between the entities were largest for the group of predominantly tissue-specifically expressed genes. In contrast, entity-specific differences were smallest for the most strongly expressed genes. For example, *NUDT15* was hypomethylated in all entities and expressed at similar levels.

Tumor methylation of pharmacogenes and entity-specific patterns

Many studies have shown that one important factor in the regulation of ADME genes in cancer is DNA methylation as reviewed by Fisel et al. [155]. However, previous work has mainly investigated hypo- or hypermethylation in comparison to the respective healthy tissue. The analyses shown here assessed the differences between the various tumor types and therefore do not allow any relative conclusions to be drawn about the difference from healthy tissue. Nevertheless, the methylation analyses showed that there is a difference in DNA methylation of some ADME genes between cancer entities. The correlation of expression and methylation in MASTER has shown significant results for a few promoter and a large amount of intragenic CpGs across several entities. Many of these associations followed the expected pattern of anti-correlation of methylation and expression in the promoter. However, a surprising number of anti-correlations were also found for CpGs in the gene body, which is much less expected. This may be due to the fact that in several PGx genes, DNA methylation may not be the main source of regulation for gene expression, in particular in the highly expressed genes.

Much has been reported, for example, about methylation and expression of *CYP2W1* in colorectal cancer. My analyses showed that two CpG sites in the promoter of *CYP2W1* were found to be significantly negatively correlated with expression in colorectal cancers in MASTER and this entity was an outlier with the highest expression of this gene. Expression of *CYP2W1* can be found in fetal and cancerous tissues of the colon but not in healthy tissue and it has been shown to be a prognostic marker associated with worse survival [141–143, 252]. Furthermore, *CYP2W1* has been suggested as a target for the treatment of colorectal cancer [253]. In accordance with previous work, the results of the methylation analysis suggest epigenetic mechanisms for the activation of *CYP2W1* in colorectal cancers [159]. Although it showed only low expression in most cancers in MASTER, *UGT1A1* expression was previously reported to be regulated by methylation in colon cancer and it has been hypothesized that this results in differential efficacy of irinotecan [254, 255].

The prognostic and diagnostic role of *GSTP1* in cancer has been thoroughly discussed [256] and the results presented here showed ubiquitous expression across all cancers. The correlation analysis of expression and sCNAs as well as methylation suggests copy number status and DNA methylation in the gene body as the main factor influencing tumor expression of *GSTP1*. Hypermethylation of *GSTP1* has already been shown several times for various types of cancer compared to healthy tissue. However, the analyses carried out here show that *GSTP1* is hypomethylated and strongly expressed relative to the other ADME genes examined.

Multivariate models of somatic pharmacogene expression

I used multivariate models to integrate all data layers of the previous analyses. This confirmed several results of the univariate analyses including the sCNA-dependent expression of *NUDT15* and *POR* in soft tissue sarcoma, *TPMT* in leiomyosarcoma, *XPC* and *CYP2R1* in neuroendocrine cancers, and *NAT1* in colorectal cancers. The germline genotype data appears to be particularly relevant for genes where germline CNVs are frequent, which seem to translate through to the tumor and affect their expression. This was true for *CYP2D6*, *CYP2A6*, and *UGT2B17*. Furthermore, the multivariate analyses confirmed the dependency of expression on methylation for *GSTP1*, *POR*, and *ABCB1* in neuroendocrine cancers, *DPYD* and *PTGIS* in soft tissue sarcoma, and *CYP2W1* in colorectal cancers.

However, there are some more detailed differences in the two individual factors (sCNAs, methylation). In particular, this can be shown for the group of ubiquitously highly expressed genes. *NUDT15* shows almost exclusive dependence on sCNA status, while it is also dominant in *POR*, *VKORC*, and *XPC*, where methylation shows some additional influence. For the remaining genes from the ubiquitously highly expressed group, *COMT*, *G6PD*, *GSTP1*, and *TPMT*, the methylation effects are stronger. For *GSTP1* and *TPMT* in particular, methy-

lation is by far the most expression-influencing factor. While *NUDT15*, *POR*, and *XPC* showed the most significant results in the association of sCNAs and expression, the association for the remaining genes in this group was not as strong, which can now be attributed to the additional influence of methylation. *GSTP1* expression was still strongly correlated with sCNAs but was the weakest among the genes of this group. Accordingly, the multivariate analysis shows the greatest influence of methylation. For *NUDT15*, where expression is almost exclusively associated with sCNAs, the single methylation analysis showed that it is hypomethylated cohort-wide, both in the intragenic regions and in the promoter. Therefore, no influence of methylation is to be expected here. In the group of enhanced expression genes, *CYP2R1* was, as expected, mainly dependent on sCNAs, while it still plays a more subordinate role in *DPYD* and *SULT1A1*. For *CYP1B1*, *PTGIS*, *SLCO2B1*, and *TBXAS1*, methylation has the greatest influence, which explains why these genes did not emerge from the correlation between sCNAs and expression. In the genes that were expressed in a tissue-specific manner, methylation is by far the dominant feature.

3.4 Genomic and Transcriptomic Analyses of Rare Cancers

As mentioned above, MASTER is a precision oncology program which aims at identifying targetable lesions through broad molecular profiling in either young adult patients with relapsed/refractory cancer, or adult patients with rare cancers at any age. A posteriori, the samples collected this way can be assembled to cohorts, and cancer genomics strategies can be applied to characterize the entities. Due to the focus on rare cancers, for several rare cancer entities, the MASTER program has generated comparably large cohorts. In the framework of this thesis, I was involved in genomically characterizing three of them. In general, the whole genome and transcriptome-based analyses provided valuable insights that allowed to describe the diseases even more precisely and provided clinically relevant information.

Parathyroid Carcinoma

In the case of parathyroid carcinoma (PC), targeted treatment based on the molecular data was applied in two of four patients and resulted in stable disease. This was achieved using immune checkpoint inhibitors in one patient and muti-receptor tyrosine kinase inhibitors followed by PARP (Poly(ADP-Ribose) Polymerase) inhibitors in the second patient. In the genomic data, previously reported mutations relevant to PC were confirmed including deactivating *CDC73* (Cell Division Cycle 73) and over-activating *CCND1* (Cyclin D1) vari-

ants [257, 258]. Research on *DICER1* mutations in parathyroid carcinoma is ongoing. In two of the four patients, a fusion and deletion of the gene were detected. *DICER1* mutations have been found to have a significant impact on the development and behavior of various endocrine tumors, including those that occur in the thyroid, parathyroid, pituitary, and adrenal cortex. If present in the germline, these mutations are associated with a rare genetic disorder known as *DICER1* syndrome, which predisposes individuals to the development of different tumors, both benign and malignant [259]. The *MEN1* gene, which encodes the protein menin, if mutated in the germline, is associated with Multiple Endocrine Neoplasia Type 1, a hereditary condition that is linked to the development of various endocrine tumors [257]. However, the variants found in *MEN1* here were somatic. Germline mutations were found in one patient in *MSH6* and *MUTYH*. The *MSH6* gene is associated with Lynch syndrome, which increases the risk of developing certain types of cancers, particularly colorectal cancer. *MSH6* is part of the DNA mismatch repair pathway and is responsible for producing a protein that plays a crucial role in repairing DNA. Defects in the *MSH6* gene can lead to the accumulation of unrepaired DNA errors, increasing the risk of tumor formation, especially in the colon [260]. Yet, links to parathyroid carcinoma for *MSH6* have not been established. *MUTYH* is involved in the base excision repair pathway, which corrects DNA damage that can result from exposure to certain chemicals or ROS (reactive oxygen species) In accordance, in this patient the mutational signature SBS18 was present, resulting from damage by reactive oxygen species. Other mutational signatures that were found in the PC patients were APOBEC-related (SBS2/13). APOBEC activity was found to be enriched in *CDC73* mutated PC by others [258], and in accordance the signature was present in the *CDC73* mutated patient in this work. *CDC73* (Cell Division Cycle 73) is a tumor suppressor gene that codes for the nuclear protein parafibromin, which regulates transcription as a subunit in the PAF1 (RNA Polymerase II Associated Factor) complex [261]. Mutations of this gene in the germline lead to a predisposition to diseases such as Hyperparathyroidism-Jaw Tumor (HPT-JT) syndrome and parathyroid carcinoma [261]. For *CDC73* mutated PC also a lower fraction of immune cells had been reported, which could not be replicated here since the patient with the lowest amount of immune cells did not have a *CDC73* mutation. The analysis of the small PC cohort nevertheless provided important findings for genomic characterization and subsequent clinical treatment decisions. In addition, new potentially important genetic variants not previously described for PC, such as *MSH6* and *MUTYH* as a predisposition element, were uncovered.

Adrenocortical Carcinoma

The analysis of adrenocortical carcinomas (ACC) offered a unique opportunity for genomic characterization of this disease due to the large cohort size (n=113), which is very special for

such a rare disease. Previous efforts to characterize the genomic landscape of ACC included several studies with cohort sizes ranging from 33 to 91 cases [175, 176, 262, 263], however, compared to the cohort analyzed here these were not pretreated or advanced stages. The mutational landscape of ACCs identified in this work confirmed many of the previously reported driver genes including inactivating *TP53*, *CTNNB1*, and *MEN1* mutations, amplifications of *TERT*, and deletions of *ZNRF3*. The frequent mutations in *TTN* (27% of the ACC cohort) were previously reported in metastatic cancer patients [262]. 66 samples of the ACC cohort were metastasic, but there seems to be no biological link between *TTN* and ACCs; mutations in *TTN* in cancer seem to result from a generally high mutational burden [264] and the fact that the gene is relatively long, passenger mutations are thus likely. However, the mutational burden in the ACC cohort was rather low in most patients. Another frequent event reported in ACCs is whole genome doubling [176] and the analyzed cohort here showed a tumor ploidy larger than 2 in 87% of patients. GISTIC analysis by two other studies showed similar genomic patterns of amplified and deleted chromosomal segments as shown here [175, 176, 265]. Confirmed were e.g. gains on chromosomes 5 (*TERT*), 12, and 19, and losses on chromosomes 2 and 22 (*ZNRF3*) among many others. Interestingly, with regard to pharmacogenomics, *CYP17A1* fusions were frequently detected in the ACC cohort and may be related to the alteration of steroidogenic signaling in these tumors. However there was no recurrent common fusion partner, therefore the importance of these findings is questionable. Generally, excess of secreted steroid hormones affects around 60% of patients and was linked to suppression of infiltrating immune cells resulting in tumor promotion and poorer survival [266, 267]. This is consistent with the fact that the immune cell deconvolution from the RNA-Seq data of the ACC cohort revealed only a small proportion of tumor-infiltrating immune cells for most patients. In addition, for several pharmacogenes (*CYP2B6*, *CYP2W1*, *CYP2C19*, *SLCO1B1/3*) previous studies have reported multiple associations between PGx variants and achieved mitotane concentrations. Mitotane is a selective cytotoxic drug for the adrenal cortex and a steroidogenesis inhibitor. The correlation of available mitotane concentration measurements with the PGx variants determined via the PGx pipeline was investigated. Unfortunately, due to the limited availability of the drug concentration measurements, no robust results were achieved there.

Chordoma

The mutation landscape has shown that generally, only a few mutations per sample occur in the chordoma cohort. This is consistent with previous studies that have shown that chordomas are not highly mutated. In the cohort analyzed, 13% of the chordomas did not show any mutations in the recurrently mutated genes (mutated in at least 4% of the samples). The frequent homozygous deletions of *CDKN2A* have been reported several times as

a chordoma-specific trait [268]. This loss makes the patients suitable for therapy with CDK inhibitors such as palbociclib. This therapeutic option is currently being tested with a small sub-cohort of chordoma patients in a clinical trial "NCT-PMO-1601: CDK4/6 inhibition in locally advanced/metastatic chordoma" (ClinicalTrials.gov Identifier NCT03110744[2]). GITSIC analysis revealed frequent amplifications of chromosome segment 6p27 including the *TBXT* gene coding for the brachyury transcription factor (responsible for the axial polarization of the embryo during development), which is expressed in the majority of chordomas and contributing to cell proliferation [269].

[2] `https://classic.clinicaltrials.gov/ct2/show/NCT03110744`

Chapter 4

Conclusion

The results presented in this thesis make contributions to the field of cancer pharmacogenomics by investigating the complex landscape of germline and tumor variation of pharmacogenes in a large cohort of cancer patients. PGx profiles of the MASTER patients have been thoroughly assessed including germline and somatic variation as well as their impact on the activity of pharmacogenes in tumors, potentially contributing to drug resistance.

The developed pharmacogenomics pipeline has proven to be suitable for the fast and reliable determination of star allele genotypes and resulting phenotypes as well as retrieving guideline recommendations for germline samples. During the development of the pipeline, it became apparent that a great deal of effort is required to harmonize the complex, historically evolved nomenclature of star allele variants between the various genotyping tools. However, this harmonization is essential and enables the determination of a consensus result of an ensemble approach, which increases the reliability.

When applying the developed PGx pipeline to rare cancers in the MASTER cohort, I was able to show that 96.4% of patients carry at least one gene with an action-necessitating phenotype. Furthermore, the analysis of additional rare and functional variants beyond star allele definitions has shown that it is equally important to integrate these into the pharmacogenomic profile of a patient and to take them into account. The integration of the pipeline results into the MASTER molecular tumor board has translated the gained knowledge directly into practice in order to quickly generate a benefit for future patients.

Somatic variants in pharmacogenes in tumors have not yet been comprehensively described. This thesis has shown that somatic SNVs only play a minor role in the PGx profile of the tumor, but suggests that sCNAs have a major contribution to somatic PGx variation. Adding analyses of tumor RNA expression and tumor DNA methylation highlighted the complexity of intra-tumor PGx. Specialized tools were identified as essential for the detection and analysis of somatic variations, distinct from germline PGx genotyping methods. SCNAs affecting pharmacogenes, partially as passenger events in vicinity to drivers, emerged as a

major contributor to somatic pharmacogenomic variation. Pharmacogenes can be grouped based on gene expression in the tumor. In general, the expression of phase II genes was highest, followed by transporters and phase I genes. Tissue-specific expression patterns were observed for a large proportion of the genes, with tumors in organs involved in drug metabolism (like liver or colorectal cancer) showing the highest relative expression. A correlation analysis of sCNAs and expression showed that especially for a group of highly expressed genes like e.g. *NUDT15* and *POR* the correlation was very strong. Specifically for *NUDT15*, sCNAs seem to be the major determinant of expression in the tumor. The methylation analysis revealed high diversity across genes, entities, and chromatin states (promoters vs. gene body). In general, methylation correlated negatively with expression. Examples are strong associations for intragenic CpGs that were negatively correlated with expression in highly expressed genes. Particularly expression of *GSTP1* in neuroendocrine cancers and *CYP2W1* in colorectal cancers seem to be regulated by several intragenic CpGs.

A multivariate analysis modeled gene expression taking sCNAs, methylation, and germline star alleles as input, both pan-MASTER and in entity-specific settings. Results of the univariate analyses were confirmed and extended and in particular, general principles which data layers have the main contribution to the regulation of gene expression of which gene class were derived.

Chapter 5

Outlook

Thorough pharmaco-omics characterization of cancer patients can not only prevent side effects, but it could also pre-emptively reveal resistance mechanisms, optimize identification of (co-)targets on the tumor, and therefore still has untapped potential for personalized cancer therapies.

The results of this thesis provide the basis for a wealth of ideas for subsequent projects. First, similar analyses could be carried out and compared with other large NGS data sets of cancer patients to investigate recurring patterns or possible differences in more detail. In addition, deeper analyses of the entity-specific mechanisms could be pursued. MASTER also includes a few longitudinal and multi-regional tumor samples taken from the same patients at different times or from different regions of the body (from both primary tumors and metastases). These samples could be used to investigate a possible evolution of PGx variation in the tumor over time (such as acquired resistance mutations in pharmacogenes) or local differences in pharmacogene activity between different metastases of the patient. An additional aspect that was not considered in this thesis is intratumoral heterogeneity. Not every cell in a tumor is identical and therefore the somatic PGx variants shown here may only affect a subpopulation of tumor cells or subclone of the tumor. Future projects could aim to investigate whether PGx resistance mechanisms offer a possible selective advantage and thus lead to the outgrowing of resistant subclones.

Another follow-up project could be the analysis of allele-specific expression of pharmacogenes in tumors. It could be investigated whether some alleles (e.g. with increased protein activity) are preferentially expressed in the tumor. For this purpose, the coverage of heterozygous SNPs from RNA and DNA sequencing could be compared. Since it was shown that some pharmacogenes are ubiquitously expressed in all tumors while others show tissue-specific patterns, another idea would be to study the activity of transcription factors to better understand the regulation of pharmacogene expression in the tumor. In addition, the expression in the tumor entities should be compared relative to the respective healthy tissue to

identify up or down-regulations wherever the tissue of origin is known.
Additional possibilities for a multitude of additional analyses would arise if sufficient clinical data on the drugs administered, observed side effects, and the success of the therapy would become available. Association analyses between the genetic factors shown here and clinical phenotypes could then be carried out. This would contribute important knowledge for the future development of predictive models that provide personalized pharmacogenomic recommendations for treating cancer patients. This could be realized by the implementation of supervised models based on state-of-the-art machine learning algorithms for the prediction of outcomes and adverse events from pharmaco-omics data.
In summary, it can be said that the role of pharmacogenes in cancer is very diverse and that this topic still offers many opportunities for deeper research into specific mechanisms.

Chapter 6

Methods

6.1 Methods for Pharmacogenomic Analyses

6.1.1 PGx Pipeline

I developed an in silico PGx pipeline, for comprehensive and automated analyses of pharmacogenomic variation in large NGS datasets, using Nextflow (22.07.1-edge, DSL version 2) [71], a structured Unix-style environment for creating software-based data processing and analysis workflows. The pipeline can be applied to short reads aligned to the GRCh37 (hg19) reference genome (BAM files) from whole genome sequencing of peripheral blood or tumor samples for the genotyping of 60 pharmacogenes as illustrated in code block 6.1. The pipeline combines multiple star allele calling or genotyping tools, including Aldy v4.3 [72], Cyrius v1.1.1 [73], PyPGX v0.19.0 [74] and Stargazer v1.0.8 [75, 76], to derive genotypes and phenotypes for the 60 supported pharmacogenes, mainly including genes coding for drug transporters as well as phase I and phase II drug-metabolizing enzymes. These tools have been shown to have comparable performance to orthogonal PGx testing methods [199]. PyPGx includes 59 genes, Stargazer 58, Aldy 35, and Cyrius is specifically designed to only call *CYP2D6* variants. Stargazer can call CNVs for all its supported 51 pharmacogenes, while the remaining tools are restricted to known CNVs. A Complete overview of the genotyping tools can be found in Table 6.1 and Table 6.2 in the appendix. Most genotyping tools allow usage of whole exome sequencing data, however, only whole genome sequencing was used in this thesis since it is inherently more comprehensive for PGx analyses as some star alleles (e.g. *CYP2C19*17*, *CYP3A4*22*, or *CYP3A5*3*) include regulatory upstream, intronic, or splicing variants. A comparison of coverage of these variants in WES and WGS is shown in Figure 2.4 in the results section. The genotyping tools and their post-processing were implemented as encapsulated Nextflow processes (code block 6.2). Each tool internally performs variant calling, phasing, and star allele matching, as well as

coverage analysis for copy number estimation, to determine resulting diplotypes. All software packages were integrated via a conda environment (Anaconda 2019.07, conda 4.10.1, python 3.7.9, R 4.0.0) and the pipeline's runtime was optimized through parallel processing for each sample and gene. Nextflow ensures reproducibility through tracking of all executed processes. In total, the pipeline supports 2,603 known pharmacogenomic SNVs of 60 genes which can be matched to the star allele nomenclature [46]. Because The output of each tool is slightly different due to differences in naming conventions, covered variants, and the final calling results, I implemented a harmonization workflow to call a consensus result (including mapping tables of variants and star alleles). For each sample, this workflow merged the results of the tools by a combination of majority voting, regular expression-based rules, and a look-up table. This table was manually curated for each pharmacogene if the automatic rules were not sufficient (the table is confidential and part of a patent application; it is not shown here). Furthermore, the pipeline reports additional germline SNVs using the GATK HaplotypeCaller and GenotypeGVCF Workflow (GATK 4.2.0.0). Moreover, the pipeline also integrates the results of somatic SNVs and sCNAs of tumor samples based on tumor-specific in-house variant calling workflows for SNVs [101, 102] and for sCNAs, ACESeq[1] [107] (allele-specific copy number estimation from whole genome sequencing).

For the integration into the molecular tumor board of the NCT/DKTK MASTER program for the prospective analysis of regularly incoming new patient samples, additional Bash and Python scripts were developed that apply the pipeline to the new patient samples every week and for a selected set of genes: (*CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *CYP3A5*, *DPYD*, *F5*, *SLCO1B1*, *TPMT*, *UGT1A1*, and *VKORC1*). The specific genotypes are annotated with CPIC recommendations. For this purpose, a local copy of the CPIC database was set up on the DKFZ cluster which can be queried with SQL (Structured Query Language) queries.

Table 6.1: Technical overview of genotyping tools (adapted from [198]).

Tool	Nr. genes	NGS technology	Input file	Genome reference	Accessibility	Output
Aldy	35	WGS, WES, long-read, array	BAM	hg19, hg38	Command line	Diplotype, additional variants
Cyrius	1	WGS	BAM	hg19, hg38	Command line	Diplotype
PyPgx	59	WGS, WES, long-read, array	BAM	hg19, hg38	Command line	Diplotype, phenotype, recommendation
Stargazer	58	WGS, WES, long-read, array	BAM (coverage), VCF	hg19, hg38	Command line	Diplotype, phenotype

[1] `https://aceseq.readthedocs.io/en/latest`

```
//Workflow Definition for PGx Pipeline
workflow pgx_pipeline {

   take:
     input_bams

   main:
     // Run all PGx tools and GATK on input bam files
     input_bams | (cyrius_genotyping &
                 pypgx_genotyping &
                 aldy_genotyping &
                 gatk_haplotypecaller &
                 gatk_depthofcoverage_germline)
     gatk_genotype_gvcf(gatk_haplotypecaller.out)
     stargazer_input = gatk_genotype_gvcf.out
                                  .join( gatk_depthofcoverage_germline.out, by: 0 )
     stargazer_genotyping_germline(stargazer_input)

     // Collect ouptut files of all PGx tools
     aldy_results=aldy_genotyping.out.result.collect()
     cyrius_results=cyrius_genotyping.out.result.collect()
     pypgx_results=pypgx_genotyping.out.result.collect()
     stargazer_results=stargazer_genotyping_germline.out.result.collect()
     results=aldy_results.concat(cyrius_results, pypgx_results, stargazer_results)
                      .flatten()
                      .toList()

     // Combine results into one table and apply consensus and harmonization rules
     merge_pgx_tools_results(results)
     postprocess_results (merge_pgx_tools_results.out.merged_file)
}

workflow {
   bam_files = Channel.fromPath(params.input_bam_files_path, checkIfExists: true)
                   .map { file -> [ file.getName()
                                    .replaceFirst(/_merged.mdup.bam/, ""),
                            file ]}
   bai_files = Channel.fromPath(params.input_bai_files_path, checkIfExists: true)
                   .map { file -> [ file.getName()
                                     .replaceFirst(/_merged.mdup.bam.bai/, ""),
                            file ]}
   input = bam_files.join(bai_files, by: 0)
   pgx_pipeline(input)
}
```

Code Block 6.1: Definition of PGx pipeline as Nextflow workflow

```
//Process Definition for Cyrius
process cyrius_genotyping {

    errorStrategy { task.attempt <= maxRetries ? 'retry' : 'ignore' }
    maxRetries 2
    cpus 2
    memory '8 GB'
    time { 2.hour }
    conda '/omics/groups/OE0246/internal/s754n/conda_envs/pharmacogenomics'
    tag "${sample_id}"
    publishDir "${params.cyrius_results_dir}${sample_id}/", mode: 'copy'

    input:
    tuple val(sample_id), path(cyrius_bam), path(cyrius_bai)

    output:
    path ("*.tsv"), emit: result
    path ("*.json"), emit: json

    script:
    """
    echo $cyrius_bam > sample_list.txt
    python3 ${params.cyrius_dir}star_caller.py -m sample_list.txt -g 37 -p ${sample_id}_cyrius -o
        \$PWD --threads $task.cpus
    sed -e s/_merged.mdup//g -i ${sample_id}_cyrius.tsv
    rm sample_list.txt
    """
}
```

Code Block 6.2: Cyrius genotyping as Nextflow process

6.1.2 NCT/DKTK MASTER Data

For the PGx analysis, I used a subcohort of the NCT/DKTK MASTER program that included 2371 patients with whole genome sequencing data of matched control and tumor samples. In brief, nucleic acids were isolated from blood and somatic tissue. The library preparation and sequencing of DNA and RNA, alignment and mapping of sequencing data, somatic variant calling (SNV, InDel, sCNA), and quality control measures for MASTER samples were performed as previously described [68, 69]. Methylation (5-methylcytosine) was measured with Illumina Infinium MethylationEPIC 850K microarrays as described in [156]. All tumor samples were assessed for sufficient tumor cell content (>20%) by a board-certified pathologist. SCNA calling results were available for 2174 tumor samples, RNA expression for 1911, and methylation data for 1792. Sequencing data until November 2021 as used in [69] has been deposited in the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/datasets) under accession EGAS00001004813. The data cutoff

date for the analyses performed in this thesis was July 7, 2023. Samples included in MASTER between the publication of Horak et al. [69] and the data cutoff date of this thesis were processed by the same methods. Patients included in MASTER provided written informed consent for the banking of tumor and control tissue, molecular analysis, and the collection of clinical data under a protocol (S-206/2011) as approved by the Ethics Committee of the Medical Faculty of Heidelberg University. All studies were conducted in accordance with the Declaration of Helsinki.

6.1.3 Germline Pharmacogenomics

Matched control samples (peripheral blood) of 2371 MASTER patients were used for whole genome sequencing and short read alignment to reference genome GRCh37/hg19 (build 37, version hs37d5) using workflows as described previously [69]. The alignment files were used as input to the PGx pipeline to detect pharmacogenomic SNVs and CNVs and call consensus genotypes and phenotypes of 60 pharmacogenes as described in section 4.1.1. Germline CNVs were extracted from the pipeline results and manually assessed using coverage and allele frequency plots created by Stargazer [75, 76] and the Integrative Genomics Viewer (version 2.8.10) [270]. Analysis of frequencies and types of star alleles and CNVs was done using custom R scripts (version 4.3.0) with the tidyverse and ggplot2 packages. Specific information about selected variants was obtained from PharmVar (`https://www.pharmvar.org/`) and PharamGKB (`https://www.pharmgkb.org/`). Within the PGx pipeline, additional germline SNVs including rare or novel variants were detected using GATK HaplotypeCaller and GenotypeGVCF workflow [89] with standard parameters using the following commands:

```
gatk --java-options -Xmx2g HaplotypeCaller
    -R $reference_genome
    -I $bam
    -O $output_vcf_file
    -L $pgx_gene_interval_list
    -ERC GVCF
    --output-mode EMIT_ALL_ACTIVE_SITES

gatk --java-options -Xmx4g GenotypeGVCFs
    -R $reference_genome
    -V $output_vcf_file -O $output_vcf
```

Code Block 6.3: GATK variant calling commands used in the PGx pipeline

6.1.4 Variant Effect Prediction and Validation

Germline and somatic SNVs that were not part of any star allele definition used in the PGx pipeline (unlike the 2,603 known pharmacogenomic SNVs) were extracted from the VCF files produced by the GATK HaplotypeCaller and GenotypeGVCF workflow using bcftools (version 1.12). The VCFs with the remaining variants (not part of star alleles) of all samples were then merged with bcftools and converted into a TSV table format. The variants were then annotated with ANNOVAR [90] using standard parameters and the following Perl command (Perl version 5.24.1):

```
perl table_annovar.pl additional_variants_MASTER.tsv humandb/
    -buildver hg19
    -out additional_variants_MASTER_annovar.tsv
    -remove
    -protocol refGene,cytoBand,exac03,avsnp147,dbnsfp30a
    -operation gx,r,f,f,f
    -nastring .
    -polish
    -xref example/gene_xref.txt
```

Code Block 6.4: ANNOVAR command used to annotate SNVs

These annotations included variant type, gene region, and ExAC population frequencies. The ANNOVAR results were filtered for exonic missense variants which were further used as input for an ADME-optimized functional prediction framework APF [91] to predict damaging effects to the resulting ADME protein function. The framework was re-implemented as an R function with the help of Yitian Zhou. The optimized framework was chosen since standard VEP tools were trained on pathogenic variants and have been shown to have poorer performance on pharmacogenomic variants [91, 233]. The APF framework includes an ensemble of established and validated VEP models but uses prediction thresholds that were optimized on datasets of known pharmacogenomic variants. Additionally, I compared the APF framework predictions with the ones of the standard VEP methods. Prediction results were created with the optimized and standard thresholds, as described in [91], for each integrated VEP method. The concordance of predictions was calculated as the fraction of matching classifications (Jaccard index). The concordance heatmap (2.11) was created using the R package ComplexHeatmap (version 2.16.0) [134]. AlphaMissense [99] variant effect prediction results were downloaded from `https://storage.googleapis.com/dm_alphamissense/AlphaMissense_hg19.tsv.gz` and filtered for the variants that were predicted as damaging by APF. The AlphaMissense data only included about 50% of the APF variants. For these 50% the overlap of AlphaMissense and APF was deter-

mined. Spearman Correlation of the number of variants and gene length per pharmacogene was performed using the R function `stat_cor(method ="sp")` from ggpubr (version 0.6.0). Information on gene length was obtained from GENCODE release 19 (`https://www.gencodegenes.org/human/release_19.html`). Lollipop plots of damaging variants and affected gene regions were created with the maftools package (version 2.18.0) [271].

6.1.5 Somatic Pharmacogenomic SNVs and sCNAs

Tumor samples of 2371 MASTER patients were used for whole genome sequencing and short read alignment to reference genome GRCh37/hg19 (build 37, version hs37d5) as described previously [69]. The alignment bam files were used as input to the PGx pipeline to detect somatic pharmacogenomic SNVs and CNAs and call consensus genotypes and phenotypes of 60 pharmacogenes as described in section 4.1.1. The differences in pipeline results (star alleles) between matched control and tumor samples were assessed using custom R scripts. All additional somatic SNVs in the tumor samples of the MASTER patients were extracted from an established DKFZ in-house analysis pipeline (SNVCalling Workflow) [101, 102]. The pipeline removes germline variants in the tumor by subtracting the variants found in the matched control sample. SNVs at known star allele loci were extracted using the list of variants implemented in the PGx pipeline. Similar to the germline analysis described above, all additional SNVs were annotated with ANNOVAR using standard parameters, and subsequently exonic non-synonymous SNVs were functionally assessed with the ADME-optimized functional prediction framework APF [91].

The sCNA calling results were extracted from the DKFZ in-house pipeline ACESeq [107]. This pipeline calls sCNAs by segmenting the WGS based on coverage and B-allele frequency (BAF) as well as a coverage ratio of tumor and matched control sample. By fitting segment copy number states to integer numbers, the pipeline tries to find the optimal base ploidy and purity (tumor cell content) of the sample. ACESeq results from the output files of all samples were merged into one data frame and descriptive statistics were computed using R.

The list of oncogenes (106) and tumor suppressor genes (183) for the analysis of sCNA segments was obtained from the cancer gene census [125] (`https://cancer.sanger.ac.uk/census#`). For each sCNA segment affecting a pharmacogene, the co-affected oncogenes and tumor suppressors were extracted. Numbers of co-affected genes were counted for each overlapping segment using the `join_overlap_inner()` function of the plyranges R package (1.20.0) [272]. Plots of chromosomal regions recurrently affected by sCNAs were created with the karyoploteR package (version 1.28.0) [273].

6.1.6 Somatic Pharmacogenomic Expression Anaylses

Bulk RNA-Seq data from tumor samples was available for 1911 patients of the MASTER cohort. Sample processing, RNA-sequencing, and alignment were done as previously described [69]. The expression of the 60 pharmacogenes in the cancer entities of the MASTER cohort was analyzed based on TPM (transcript per million reads) values. For statistical and downstream analyses, TPM values were log2 transformed. The expression heatmap was created using the R package ComplexHeatmap (version 2.16.0) [134]. For the analysis of the association of sCNAs and expression levels, patients were categorized into 3 sCNA groups (deleted, neutral, duplicated/amplified). The association of sCNA category and TPM values was computed globally for the whole MASTER cohort and per cancer entity and pharmacogene using Kruskal-Wallis Rank Sum Tests. Plots were created in R using ggplot2. P-values were adjusted by the Benjamini-Hochberg method and p-values below 0.05 were considered statistically significant.

6.1.7 Somatic Pharmacogenomic Methylation Analyses

Illumina Infinium MethylationEPIC 850k microarrays were used to measure beta and M values for 1792 tumor samples as previously described [156]. From the list of all available CpG sites, I filtered for intragenic CpGs in the gene body and promoter CpGs up to 5000 base pairs upstream of the transcription start site (TSS) of the 60 ADME genes. This included 1226 intragenic and 353 promoter CpGs. Statistical analyses were based on M values and for visualization beta values were used as recommended [157, 158]. The methylation heatmap (Figure 2.38) was created using the R package ComplexHeatmap (version 2.16.0) [134]. The remaining plots were created with ggplot2. Correlation analysis of methylation and expression per CpG site grouped by gene and cancer entity was done using Spearman correlation tests between the M values of each CpG and the corresponding TPM expression values. P values were adjusted by Benjamini-Hochberg correction and p-values below 0.05 were considered statistically significant.

6.1.8 Multivariate Analyses of Somatic Pharmacogene Expression

For the development of the multivariate models, the separate data layers (germline genotype, sCNA, tumor methylation, and tumor expression) were merged into one dataframe containing complete data of 1450 patients. The included CpG sites from the methylation data were further restricted to the ones that were significantly correlated to expression based on the previous analysis using a cutoff by adjusted p-value (padj < 0.0001) to reduce the number of features. The multivariate models were constructed with the function `lm()` of the stats R

package and the following formula:

$$TPM \sim consensus\_genotype\_germline + sCNA\_type + significant\_cpgs.$$

For each of the genes, a model was fitted cohort-wide, and model parameters including coefficients and p-values were extracted. Additional models were fitted per cancer entity and gene separately using the same formula and preprocessing. For visualization features were categorized into 3 categories (germline genotype, somatic CNAs, and methylation), and the most significant/predictive feature per model was extracted. Plots were created using the R packages ComplexHeatmap (version 2.16.0) [134] and ggplot2.

6.2 Methods for Genomic and Transcriptomic Cohort Analyses

Patient inclusion, sample collection, and processing were part of the NCT/DKTK MASTER program as described earlier and in [69]. Germline and somatic genomic data (SNVs, InDels, fusions, sCNAs) were derived from the DKFZ OTP pipelines and the integrated RObject dataMASTER which encapsulates all genomic data and patient metadata of the MASTER cohort. This is an object of the type `MultiAssayExperiment`, and there is a pipeline that fills this object with content from variant calling pipelines at regular intervals. Variants for the oncoprints were filtered by recurrence depending on cohort size and total amount of variants (25% for parathyroid carcinoma, 10% for adrenocortical carcinoma, and 4% for chordoma). Oncoprints were plotted using a custom R script merging all included data and the oncoPrint function of the R package ComplexHeatmap [134]. Mutational signature analysis was performed based on the aforementioned somatic SNVs with the R package YAPSA [163] and standard parameters, including signature-specific cutoffs. Quantification of immune cell admixture was done with the immunedeconv R package (version 2.0.2) which includes several established algorithms [164]. Results for display were obtained from the Cibersort results of immunedeconv. GISTIC [177] analysis of regions recurrently affected by sCNAs was done based on sCNAs extracted from the previously mentioned ACESeq pipeline results which are also integrated into dataMASTER. The GISTIC command line tool was integrated into a custom R script. Differential gene expression (DGE) for the chordoma subgroups was analyzed with the DESeq2 package (1.42.0) based on raw read counts from bulk tumor RNA-sequencing, as it performs internal normalization of read counts [274]. Volcano plots of DGE results were plotted with EnhancedVolcano (version 1.20.0). Gene set enrichment analysis and assignment of Gene Ontology

terms [275] was done with the `functional_enrichment()` method of the cola R package using standard parameters [188]. Similarity measures of the GO terms and representation of the matrix were computed with the simplifyEnrichment R package [189] using the `simplifyGOFromMultipleLists()` function and a p-value cutoff of 0.001.
PGx genotyping of chordoma samples was done with the PGX pipeline as described in the PGx analysis of the MASTER cohort. Validation of the PGx genotypes of chordoma samples was conducted by Roman Tremmel at IKP Stuttgart using selected TaqMan assays and real-time PCR as described in [84].

Own Publications

Co-Authorship:

Tremmel R, **Pirmann S**, Zhou Y, Lauschke VM. *Translating pharmacogenomic sequencing data into drug response predictions—How to interpret variants of unknown significance.* Br J Clin Pharmacol. 2023; 1-12. doi:10.1111/bcp.15915

Teleanu MV, Fuss CT, Paramasivam N, **Pirmann S**, Mock A, Terkamp C, Kircher S, Landwehr LS, Lenschow C, Schlegel N, Stenzinger A, Jahn A, Fassnacht M, Glimm H, Hübschmann D, Fröhling S, Kroiss M. *Targeted therapy of advanced parathyroid carcinoma guided by genomic and transcriptomic profiling.* Mol Oncol. 2023 Jul;17(7):1343-1355. doi: 10.1002/1878-0261.13398. Epub 2023 Apr 11. PMID: 36808802; PMCID: PMC10323885.

Heilig CE, Laßmann A, Mughal SS, Mock A, **Pirmann S**, Teleanu V, Renner M, Andresen C, Köhler BC, Aybey B, Bauer S, Siveke JT, Hamacher R, Folprecht G, Richter S, Schröck E, Brandts CH, Ahrens M, Hohenberger P, Egerer G, Kindler T, Boerries M, Illert AL, von Bubnoff N, Apostolidis L, Jost PJ, Westphalen CB, Weichert W, Keilholz U, Klauschen F, Beck K, Winter U, Richter D, Möhrmann L, Bitzer M, Schulze-Osthoff K, Brors B, Mechtersheimer G, Kreutzfeldt S, Heining C, Lipka DB, Stenzinger A, Schlenk RF, Horak P, Glimm H, Hübschmann D, Fröhling S. *Gene expression-based prediction of pazopanib efficacy in sarcoma.* Eur J Cancer. 2022 Sep;172:107-118. doi: 10.1016/j.ejca.2022.05.025. Epub 2022 Jun 25. PMID: 35763870.

Manuscripts in preparation:

First-Authorship:

Pirmann S, Tremmel R, Schäffeler E, Fröhling S, Schwab M, Hübschmann D. *The Germline and Tumor ADME Pharmaco-omics Landscape of Cancer Patients Revealed by Next Generation Sequencing.*

Co-Authorship:

Zhou Y, **Pirmann S**, Lauschke V. *APF2: an improved ensemble method for pharmacogenomic variant effect prediction.*

Tremmel R, Hübschmann D, Schaeffeler E, **Pirmann S**, Fröhling S, Schwab M. *Innovation in cancer pharmacotherapy through integrative consideration of germline and tumor genomes.*

In addition, together with collaborators, several manuscripts for some of the work shown here are prepared at the time of submission of this thesis. These include the multi-omics analysis of chordomas, the genomic analysis in the NCT-PMO1601 study (CDK4/6 inhibition in locally advanced/metastatic chordoma), as well as the genomic and transcriptomic analyses of adrenocortical carcinomas.

Acknowledgements

As with most achievements in life, the completion of this work was only made possible through collaborative efforts with various great people. I am grateful to my family and friends, whose consistent support and encouragement served as a great source of motivation. First and foremost I want to thank **Daniel Hübschmann** for his exceptional guidance and supervision, both in the completion of this thesis and in my involvement in other diverse projects within his Computational Oncology group at DKFZ. Considering that I was quite new to the field of bioinformatics and precision oncology, he taught me a wealth of knowledge and useful skills. Additionally, I thank all members of the Computational Oncology group, and in particular **Nagarajan Paramsivam**, **Małgorzata Oleś**, **Zuguang Gu**, and **Jude Alsabah**. Despite the challenges of remote collaboration, our exchange of ideas remained consistently fun and enriching. Acknowledgments are also due to **Stefan Fröhling** and the NCT Heidelberg for the initiation, support, and guidance of this project. I also want to acknowledge the valuable contributions of the whole Translational Medical Oncology group. I would like to thank the Institute of Clinical Pharmacology in Stuttgart, especially **Roman Tremmel**, **Elke Schäffeler**, and **Matthias Schwab** for the close collaboration that was established even beyond this thesis, their insightful suggestions, and extensive support. Furthermore, I thank **Benedikt Brors** for being the first examiner of this thesis and providing valuable input. I also would like to thank **Matthias Kroiß** and his team for the great collaboration on the parathyroid and adrenocortical carcinoma studies. Additionally, I want to thank **Alexander Knurr** and the whole SUDO team, the DKFZ IT Core Facility (ITCF) and cluster management, and the DKFZ Omics IT and Data Management Core Facility (ODCF). I thank **Volker Lauschke** and his Personalized Medicine and Drug Development group at the Karolinska Institute in Stockholm for hosting me during a wonderful 3-month internship. Special thanks also to **Yitian Zhou** for our collaborative efforts on pharmacogenomic variant effect prediction. I also thank **Jan Lohmann** as an additional member of the examination board. Lastly, I thank the DKFZ PhD Program and the HIDSS4Health Graduate School for their support and for providing numerous opportunities for social events and continuing education offers throughout my doctoral journey.

Bibliography

[1] F. Sanger and A. R. Coulson. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3):441–448, 5 1975.

[2] E. S. Lander, L. M. Linton, B. Birren, et al. Initial sequencing and analysis of the human genome. *Nature 2001 409:6822*, 409(6822):860–921, 2 2001.

[3] D. L. Altshuler, R. M. Durbin, G. R. Abecasis, et al. A map of human genome variation from population scale sequencing. *Nature*, 467(7319):1061, 10 2010.

[4] C. Thermes. Ten years of next-generation sequencing technology. *Trends in genetics : TIG*, 30(9), 2014.

[5] M. L. Metzker. Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11(1):31–46, 1 2010.

[6] S. Goodwin, J. D. McPherson, and W. R. McCombie. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6):333–351, 6 2016.

[7] H. Zhang. Overview of sequence data formats. In E. Mathé and S. Davis, editors, *Statistical Genomics (Methods in Molecular Biology Series)*, volume 1418, chapter 1, pages 3–18. Humana Press - Springer Nature, 2016.

[8] A. Petrackova, M. Vasinek, L. Sedlarikova, et al. Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. *Frontiers in Oncology*, 9, 2019.

[9] M. Meyerson, S. Gabriel, and G. Getz. Advances in understanding cancer genomes through second-generation sequencing. *Nature Reviews Genetics*, 11(10), 2010.

[10] S. S. Jamuar and E. C. Tan. Clinical application of next-generation sequencing for Mendelian diseases. *Human genomics*, 9, 2015.

[11] T. Shen, S. H. Pajaro-Van de Stadt, N. C. Yeat, and J. C. Lin. Clinical applications of next generation sequencing in cancer: From panels, to exomes, to genomes. *Frontiers in Genetics*, 6(JUN), 2015.

[12] Z. Wang, M. Gerstein, and M. Snyder. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 2009.

[13] S. Zverinova and V. Guryev. Variant calling: Considerations, practices, and developments. *Human Mutation*, 43(8), 2022.

[14] D. C. Koboldt. Best practices for variant calling in clinical sequencing. *Genome Medicine*, 12(1), 2020.

[15] J. Gagan and E. M. Van Allen. Next-generation sequencing to guide cancer therapy. *Genome Medicine*, 7(1), 2015.

[16] M. Kalia. Biomarkers for personalized oncology: Recent advances and future challenges. *Metabolism: Clinical and Experimental*, 64(3), 2015.

[17] C. Chen, J. Wang, D. Pan, et al. Applications of multi-omics analysis in human diseases. *MedComm*, 4(4):e315, 8 2023.

[18] M. Krassowski, V. Das, S. K. Sahu, and B. B. Misra. State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Frontiers in Genetics*, 11:610798, 12 2020.

[19] Y. J. Heo, C. Hwa, G. H. Lee, et al. Integrative multi-omics approaches in cancer research: From biological networks to clinical subtypes. *Molecules and Cells*, 44(7), 2021.

[20] Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome Biology*, 18(1), 2017.

[21] W. E. Evans and M. V. Relling. Pharmacogenomics: Translating functional genomics into rational therapeutics. *Science*, 286(5439):487–491, 10 1999.

[22] D. W. Nebert. Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist? *Clinical Genetics*, 56(4):247–258, 10 1999.

[23] D. M. Roden, H. L. McLeod, M. V. Relling, et al. Pharmacogenomics. *The Lancet*, 394(10197):521–532, 2019.

[24] M. Eichelbaum, M. Ingelman-Sundberg, and W. E. Evans. Pharmacogenomics and Individualized Drug Therapy. *https://doi.org/10.1146/annurev.med.56.082103.104724*, 57(7):119–137, 1 2006.

[25] W. E. Evans and M. V. Relling. Moving towards individualized medicine with pharmacogenomics. *Nature*, 429(6990):464–468, 5 2004.

[26] M. Pirmohamed. Pharmacogenomics: current status and future perspectives. *Nature Reviews Genetics 2023 24:6*, 24(6):350–362, 1 2023.

[27] J. C. Granados, J. D. Watrous, T. Long, et al. Regulation of Human Endogenous Metabolites by Drug Transporters and Drug Metabolizing Enzymes: An Analysis of Targeted SNP-Metabolite Associations. *Metabolites*, 13(2), 2 2023.

[28] P. Anzenbacher and U. M. Zanger. Metabolism of Drugs and Other Xenobiotics. *Metabolism of Drugs and Other Xenobiotics*, 6 2012.

[29] D. W. Nebert and D. W. Russell. Clinical importance of the cytochromes P450. *Lancet (London, England)*, 360(9340):1155–1162, 10 2002.

[30] U. M. Zanger, M. Turpeinen, K. Klein, and M. Schwab. Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Analytical and bioanalytical chemistry*, 392(6):1093–1108, 11 2008.

[31] U. M. Zanger and M. Schwab. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & therapeutics*, 138(1):103–141, 4 2013.

[32] K. A. Birdwell, B. Decker, J. M. Barbarino, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for CYP3A5 Genotype and Tacrolimus Dosing. *Clinical pharmacology and therapeutics*, 98(1):19–24, 7 2015.

[33] E. C. Hulshof, M. J. Deenen, M. Nijenhuis, et al. Dutch pharmacogenetics working group (DPWG) guideline for the gene-drug interaction between UGT1A1 and irinotecan. *European journal of human genetics : EJHG*, 31(9):982–987, 9 2023.

[34] A. G. Roberts. The Structure and Mechanism of Drug Transporters. *Methods in molecular biology (Clifton, N.J.)*, 2342:193, 2021.

[35] R. M. Cooper-DeHoff, M. Niemi, L. B. Ramsey, et al. The Clinical Pharmacogenetics Implementation Consortium Guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and Statin-Associated Musculoskeletal Symptoms. *Clinical pharmacology and therapeutics*, 111(5):1007–1021, 5 2022.

[36] A. Chhibber, C. E. French, S. W. Yee, et al. Transcriptomic variation of pharmacogenes in multiple human tissues and lymphoblastoid cell lines. *The Pharmacogenomics Journal 2016 17:2*, 17(2):137–145, 2 2016.

[37] D. G. Hu, P. I. Mackenzie, P. C. Nair, et al. The Expression Profiles of ADME Genes in Human Cancers and Their Associations with Clinical Outcomes. *Cancers 2020, Vol. 12, Page 3369*, 12(11):3369, 11 2020.

[38] A. M. Wheeler, C. D. Eberhard, E. P. Mosher, et al. Achieving a Deeper Understanding of Drug Metabolism and Responses Using Single-Cell Technologies. *Drug Metabolism and Disposition*, 51(3):350, 3 2023.

[39] Q. Mao and J. D. Unadkat. Role of the breast cancer resistance protein (ABCG2) in drug transport. *AAPS Journal*, 7(1), 5 2005.

[40] Y. Choi and A.-M. Yu. ABC Transporters in Multidrug Resistance and Pharmacokinetics, and Strategies for Drug Development. *Current Pharmaceutical Design*, 20(5):793–807, 2 2014.

[41] R. Vadlapatla, A. Vadlapudi, D. Pal, and A. Mitra. Mechanisms of drug resistance in cancer chemotherapy: coordinated role and regulation of efflux transporters and metabolizing enzymes. *Current pharmaceutical design*, 19(40):7126–7140, 12 2013.

[42] Y. K. Zhang, Y. J. Wang, P. Gupta, and Z. S. Chen. Multidrug Resistance Proteins (MRPs) and Cancer Therapy. *AAPS Journal*, 17(4):802–812, 7 2015.

[43] S. Pathania, R. Bhatia, A. Baldi, et al. Drug metabolizing enzymes and their inhibitors' role in cancer resistance. *Biomed Pharmacother*, 105:53–65, 2018.

[44] H. T. Chan, Y. M. Chin, and S.-K. Low. The Roles of Common Variation and Somatic Mutation in Cancer Pharmacogenomics. *Oncology and Therapy 2018 7:1*, 7(1):1–32, 1 2019.

[45] K. A. Frazer, S. S. Murray, N. J. Schork, and E. J. Topol. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics 2009 10:4*, 10(4):241–251, 4 2009.

[46] L. V. Kalman, J. A. Agúndez, M. L. Appell, et al. Pharmacogenetic Allele Nomenclature: International Workgroup Recommendations for Test Result Reporting. *Clinical pharmacology and therapeutics*, 99(2):172, 2 2016.

[47] H. E. Wheeler, M. L. Maitland, M. E. Dolan, et al. Cancer pharmacogenomics: strategies and challenges. *Nature reviews. Genetics*, 14(1):23, 1 2013.

[48] H. L. McLeod. Cancer pharmacogenomics: early promise, but concerted effort needed. *Science*, 339(6127):1563–1566, 3 2013.

[49] F. Feng, B. Shen, X. Mou, et al. Large-scale pharmacogenomic studies and drug response prediction for personalized cancer medicine. *Journal of Genetics and Genomics*, 48(7):540–551, 7 2021.

[50] D. L. Hertz, A. Glatz, A. L. Pasternak, et al. Integration of Germline Pharmacogenetics Into a Tumor Sequencing Program. *JCO Precision Oncology*, 2:1–15, 7 2018.

[51] D. Nichols, S. Arnold, H. L. Weiss, et al. Pharmacogenomic potential in advanced cancer patients. *American Journal of Health-System Pharmacy*, 76(7), 2019.

[52] N. D. Seligson, J. M. Kolesar, B. Alam, et al. Integrating pharmacogenomic testing into paired germline and somatic genomic testing in patients with cancer. *https://doi.org/10.2217/pgs-2023-0125*, 24(13):731–738, 9 2023.

[53] L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, et al. Signatures of mutational processes in human cancer. *Nature*, 500(7463):415–421, 2013.

[54] L. B. Alexandrov, J. Kim, N. J. Haradhvala, et al. The repertoire of mutational signatures in human cancer. *Nature*, 578(7793):94–101, 2020.

[55] C. Greenman, P. Stephens, R. Smith, et al. Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132):153, 3 2007.

[56] K. O. Alfarouk, C. M. Stock, S. Taylor, et al. Resistance to cancer chemotherapy: failure in drug response from ADME to P-gp. *Cancer Cell International*, 15(1):71, 7 2015.

[57] W. S. van de Geer, R. H. Mathijssen, J. van Riet, et al. Identifying somatic changes in drug transporters using whole genome and transcriptome sequencing data of advanced tumors. *Biomedicine & Pharmacotherapy*, 159:114210, 3 2023.

[58] H. Joyce, A. McCann, M. Clynes, and A. Larkin. Influence of multidrug resistance and drug transport proteins on chemotherapy drug metabolism. *Expert Opinion on Drug Metabolism and Toxicology*, 11(5):795–809, 5 2015.

[59] M. Michael and M. M. Doherty. Drug metabolism by tumours: its nature, relevance and therapeutic implications. *Expert opinion on drug metabolism & toxicology*, 3(6):783–803, 12 2007.

[60] G. Kaur, S. K. Gupta, P. Singh, et al. Drug-metabolizing enzymes: role in drug resistance in cancer. *Clinical and Translational Oncology*, 22(10):1667–1680, 10 2020.

[61] M. Michael and M. M. Doherty. Tumoral drug metabolism: Overview and its implications for cancer therapy. *Journal of Clinical Oncology*, 23(1):205–229, 1 2005.

[62] Y. Li, A. Steppi, Y. Zhou, et al. Tumoral expression of drug and xenobiotic metabolizing enzymes in breast cancer patients of different ethnicities with implications to personalized medicine. *Scientific Reports*, 7(1), 12 2017.

[63] S. Feng, A. Li, Y.-C. Zheng, and H.-M. Liu. Role of Drug-metabolizing Enzymes in Cancer and Cancer Therapy. *Current Drug Metabolism*, 21(1):67–76, 1 2020.

[64] E. R. Lepper, K. Nooter, J. Verweij, et al. Mechanisms of resistance to anticancer drugs: The role of the polymorphic ABC transporters ABC1 and ABCG2. *Pharmacogenomics*, 6(2):115–138, 3 2005.

[65] U. A. Meyer, U. M. Zanger, and M. Schwab. Omics and Drug Response. *https://doi.org/10.1146/annurev-pharmtox-010510-100502*, 53:475–502, 1 2013.

[66] C. Auwerx, M. C. Sadler, A. Reymond, and Z. Kutalik. From pharmacogenetics to pharmaco-omics: Milestones and future directions. *HGG advances*, 3(2), 4 2022.

[67] A. Mock, M. V. Teleanu, S. Kreutzfeldt, et al. NCT/DKFZ MASTER handbook of interpreting whole-genome, transcriptome, and methylome data for precision oncology. *npj Precision Oncology 2023 7:1*, 7(1):1–15, 10 2023.

[68] P. Horak, B. Klink, C. Heining, et al. Precision oncology based on omics data: The NCT Heidelberg experience. *International journal of cancer*, 141(5):877–886, 9 2017.

[69] P. Horak, C. Heining, S. Kreutzfeldt, et al. Comprehensive Genomic and Transcriptomic Analysis for Guiding Therapeutic Decisions in Patients with Rare Cancers. *Cancer Discovery*, 11(11):2780–2795, 11 2021.

[70] M. Ramos, L. Schiffer, A. Re, et al. Software for the Integration of Multiomics Experiments in Bioconductor. *Cancer research*, 77(21):e39–e42, 2017.

[71] P. DI Tommaso, M. Chatzou, E. W. Floden, et al. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319, 4 2017.

[72] I. Numanagić, S. Malikić, M. Ford, et al. Allelic decomposition and exact genotyping of highly polymorphic and structurally variant genes. *Nature communications*, 9(1):828, 2018.

[73] X. Chen, F. Shen, N. Gonzaludo, et al. Cyrius: accurate CYP2D6 genotyping using whole-genome sequencing data. *Pharmacogenomics Journal*, 2021.

[74] S.-b. Lee, J.-Y. Shin, N.-J. Kwon, et al. ClinPharmSeq: A targeted sequencing panel for clinical pharmacogenetics implementation. *PLOS ONE*, 17(7):1–19, 2022.

[75] S.-b. Lee, M. M. Wheeler, K. Patterson, et al. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6 as a model. *Genetics in medicine : official journal of the American College of Medical Genetics*, 21(2):361–372, 2019.

[76] S.-b. Lee, M. M. Wheeler, K. E. Thummel, and D. A. Nickerson. Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clinical pharmacology and therapeutics*, 106(6):1328–1337, 2019.

[77] E. L. Moen, L. A. Godley, W. Zhang, and M. E. Dolan. Pharmacogenomics of chemotherapeutic susceptibility and toxicity. *Genome Medicine*, 4(11):90, 11 2012.

[78] N. N. Miteva-Marcheva, H. Y. Ivanov, D. K. Dimitrov, and V. K. Stoyanova. Application of pharmacogenetics in oncology. *Biomarker Research*, 8(1):1–10, 8 2020.

[79] D. F. Carr, R. M. Turner, and M. Pirmohamed. Pharmacogenomics of anticancer drugs: Personalising the choice and dose to manage drug response. *British Journal of Clinical Pharmacology*, 87(2):237–255, 2 2021.

[80] Y. He, J. M. Hoskins, and H. L. McLeod. Copy Number Variants in pharmacogenetic genes. *Trends in molecular medicine*, 17(5):244, 5 2011.

[81] I. Johansson and M. Ingelman-Sundberg. CNVs of human genes and their implication in pharmacogenetics. *Cytogenetic and genome research*, 123(1-4):195–204, 3 2008.

[82] M. Ingelman-Sundberg, S. C. Sim, A. Gomez, and C. Rodriguez-Antona. Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacology and Therapeutics*, 116(3):496–526, 12 2007.

[83] M. Ingelman-Sundberg, S. Mkrtchian, Y. Zhou, and V. M. Lauschke. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum Genomics*, 12(1), 1 2018.

[84] J. P. Müller, J. Sarömba, P. Ziegler, et al. Nutrimetric Validation of Solanidine as Dietary-Derived CYP2D6 Activity Marker In Vivo. *Clinical pharmacology and therapeutics*, 115(2):309–317, 2 2024.

[85] Y. Zhou, M. Ingelman-Sundberg, and V. M. Lauschke. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. *Clinical pharmacology and therapeutics*, 102(4):688–700, 10 2017.

[86] H. M. Dunnenberger, K. R. Crews, J. M. Hoffman, et al. Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annual review of pharmacology and toxicology*, 55:89–106, 1 2015.

[87] C. P. I. Schärfe, R. Tremmel, M. Schwab, et al. Genetic variation in human drug-related genes. *Genome medicine*, 9(1):117, 12 2017.

[88] R. Tremmel, Y. Zhou, M. Schwab, and V. M. Lauschke. Structural variation of the coding and non-coding human pharmacogenome. *npj Genomic Medicine 2023 8:1*, 8(1):1–11, 9 2023.

[89] R. Poplin, V. Ruano-Rubio, M. A. DePristo, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, page 201178, 7 2018.

[90] K. Wang, M. Li, and H. Hakonarson. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16), 7 2010.

[91] Y. Zhou, S. Mkrtchian, M. Kumondai, et al. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics Journal*, 19(2):115–126, 4 2019.

[92] M. Kircher, D. M. Witten, P. Jain, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, 46(3):310–315, 2014.

[93] I. A. Adzhubei, S. Schmidt, L. Peshkin, et al. A method and server for predicting damaging missense mutations. *Nature methods*, 7(4):248, 4 2010.

[94] B. Reva, Y. Antipin, and C. Sander. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, 39(17):e118–e118, 9 2011.

[95] H. A. Shihab, M. F. Rogers, J. Gough, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics (Oxford, England)*, 31(10):1536–1543, 5 2015.

[96] Y. Choi, G. E. Sims, S. Murphy, et al. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, 7(10), 10 2012.

[97] M. F. Rogers, H. A. Shihab, M. Mort, et al. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*, 34(3):511, 2 2018.

[98] T. A. Myers, S. J. Chanock, and M. J. Machiela. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Frontiers in Genetics*, 11:513535, 2 2020.

[99] J. Cheng, G. Novati, J. Pan, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664), 9 2023.

[100] GMS | 65th Annual Meeting of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Meeting of the Central European Network (CEN: German Region, Austro-Swiss Region and Polish Region) of the International Biometric Society (IBS) | The Knowledge Connector: Workflow Support for Molecular Tumor Boards.

[101] E. Reisinger, L. Genthner, J. Kerssemakers, et al. OTP: An automatized system for managing and processing NGS data. *Journal of Biotechnology*, 261:53–62, 11 2017.

[102] P. J. Campbell, G. Getz, J. O. Korbel, et al. Pan-cancer analysis of whole genomes. *Nature 2020 578:7793*, 578(7793):82–93, 2 2020.

[103] M. R. Stratton, P. J. Campbell, and P. A. Futreal. The cancer genome. *Nature 2009 458:7239*, 458(7239):719–724, 4 2009.

[104] R. Beroukhim, C. H. Mermel, D. Porter, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283):899–905, 2 2010.

[105] J. Rueff and A. S. Rodrigues. Cancer drug resistance: A brief overview from a genetic viewpoint. *Methods in Molecular Biology*, 1395:1–18, 2016.

[106] E. Incze, K. Mangó, F. Fekete, et al. Potential Association of Cytochrome P450 Copy Number Alteration in Tumour with Chemotherapy Resistance in Lung Adenocarcinoma Patients. *International Journal of Molecular Sciences*, 24(17), 9 2023.

[107] K. Kleinheinz, I. Bludau, D. Hübschmann, et al. ACEseq – allele specific copy number estimation from whole genome sequencing. *bioRxiv*, page 210807, 10 2017.

[108] D. J. Gordon, B. Resio, and D. Pellman. Causes and consequences of aneuploidy in cancer. *Nat. Rev. Genet.*, 13(3):189–203, 3 2012.

[109] U. Ben-David and A. Amon. Context is everything: aneuploidy in cancer. *Nature Reviews Genetics 2019 21:1*, 21(1):44–62, 9 2019.

[110] A. M. Taylor, J. Shih, G. Ha, et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer cell*, 33(4):676–689, 4 2018.

[111] P. Chudasama, S. S. Mughal, M. A. Sanders, et al. Integrative genomic and transcriptomic analysis of leiomyosarcoma. *Nature communications*, 9(1):144, 1 2018.

[112] A. Astolfi, M. Nannini, M. A. Pantaleo, et al. A molecular portrait of gastrointestinal stromal tumors: an integrative analysis of gene expression profiling and high-resolution genomic copy number. *Laboratory Investigation 2010 90:9*, 90(9):1285–1294, 6 2010.

[113] M. C. Heinrich, J. Patterson, C. Beadling, et al. Genomic aberrations in cell cycle genes predict progression of KIT-mutant gastrointestinal stromal tumors (GISTs). *Clinical Sarcoma Research*, 9(1):3, 12 2019.

[114] H. M. Namløs, K. Khelik, S. Nakken, et al. Chromosomal instability and a deregulated cell cycle are intrinsic features of high□risk gastrointestinal stromal tumours with a metastatic potential. *Molecular Oncology*, 17(11):2432, 11 2023.

[115] L. Gorunova, K. Boye, I. Panagopoulos, et al. Cytogenetic and molecular analyses of 291 gastrointestinal stromal tumors: site-specific cytogenetic evolution as evidence of pathogenetic heterogeneity. *Oncotarget*, 13(1):508–517, 1 2022.

[116] T. Ried, R. Knutzen, R. Steinbeck, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes and Cancer*, 15(4), 1996.

[117] D. M. Oliveira, G. Santamaria, C. Laudanna, et al. Identification of copy number alterations in colon cancer from analysis of amplicon-based next generation sequencing data. *Oncotarget*, 9(29):20409, 4 2018.

[118] T. Ried, G. A. Meijer, D. J. Harrison, et al. The landscape of genomic copy number alterations in colorectal cancer and their consequences on gene expression levels and disease outcome. *Molecular Aspects of Medicine*, 69:48–61, 10 2019.

[119] J. A. Johnson, I. S. Hamadeh, and T. Y. Langaee. Loss of Heterozygosity at the CYP2D6 Locus in Breast Cancer: Implications for Tamoxifen Pharmacogenetic Studies. *JNCI Journal of the National Cancer Institute*, 107(2):dju437–dju437, 1 2015.

[120] High-Throughput Loss of Heterozygosity Mapping in 26 Commonly Deleted Regions in Breast Cancer1 | Cancer Epidemiology, Biomarkers & Prevention | American Association for Cancer Research.

[121] M. P. Goetz, J. X. Sun, V. J. Suman, et al. Loss of Heterozygosity at the CYP2D6 Locus in Breast Cancer: Implications for Germline Pharmacogenetic Studies. *J Natl Cancer Inst*, 107(2), 2 2015.

[122] T. P. Ahern, D. L. Hertz, P. Damkier, et al. Cytochrome P-450 2D6 (CYP2D6) Genotype and Breast Cancer Recurrence in Tamoxifen-Treated Patients: Evaluating the Importance of Loss of Heterozygosity. *American Journal of Epidemiology*, 185(2):75–85, 1 2017.

[123] F. Kaveh, L. O. Baumbusch, D. Nebdal, et al. A systematic comparison of copy number alterations in four types of female cancer. *BMC Cancer*, 16(1), 11 2016.

[124] T. I. Zack, S. E. Schumacher, S. L. Carter, et al. Pan-cancer patterns of somatic copy number alteration. *Nature Genetics 2013 45:10*, 45(10):1134–1140, 9 2013.

[125] Z. Sondka, S. Bamford, C. G. Cole, et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature reviews. Cancer*, 18(11):696–705, 11 2018.

[126] A. Sirvent, C. Benistant, and S. Roche. Oncogenic signaling by tyrosine kinases of the SRC family in advanced colorectal cancer. *American Journal of Cancer Research*, 2(4):357, 2012.

[127] J. Chen, A. Elfiky, M. Han, et al. The role of Src in colon cancer and its therapeutic implications. *Clinical colorectal cancer*, 13(1):5–13, 3 2014.

[128] S. Mamlouk, L. H. Childs, D. Aust, et al. DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer. *Nature Communications 2017 8:1*, 8(1):1–12, 1 2017.

[129] J. Frigola, M. Mũoz, S. J. Clark, et al. Hypermethylation of the prostacyclin synthase (PTGIS) promoter is a frequent event in colorectal cancer and associated with aneuploidy. *Oncogene*, 24(49):7320–7326, 11 2005.

[130] H. Ding, K. y. Wang, S. y. Chen, et al. Validating the role of PTGIS gene in colorectal cancer by bioinformatics analysis and in vitro experiments. *Scientific Reports 2023 13:1*, 13(1):1–13, 10 2023.

[131] Z. Huang, M. J. Fasco, H. L. Figge, et al. Expression of cytochromes P450 in human breast tissue and tumors. *Drug Metabolism and Disposition*, 24(8):899–905, 8 1996.

[132] D. Downie, M. C. McFadyen, P. H. Rooney, et al. Profiling cytochrome P450 expression in ovarian cancer: Identification of prognostic markers. *Clinical Cancer Research*, 11(20):7369–7375, 10 2005.

[133] G. I. Murray, M. C. McFadyen, R. T. Mitchell, et al. Cytochrome P450 CYP3A in human renal cell cancer. *British journal of cancer*, 79(11-12):1836–1842, 1999.

[134] Z. Gu, R. Eils, and M. Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, 9 2016.

[135] J. Cui, G. Li, J. Yin, et al. GSTP1 and cancer: Expression, methylation, polymorphisms and signaling (Review). *International journal of oncology*, 56(4):867–878, 2020.

[136] L. Sawers, M. J. Ferguson, B. R. Ihrig, et al. Glutathione S-transferase P1 (GSTP1) directly influences platinum drug chemosensitivity in ovarian tumour cell lines. *Br J Cancer*, 111(6):1150–1158, 9 2014.

[137] S. M. Louie, E. A. Grossman, L. A. Crawford, et al. GSTP1 Is a Driver of Triple-Negative Breast Cancer Cell Metabolism and Pathogenicity. *Cell Chemical Biology*, 23(5):567–578, 5 2016.

[138] S. Q. Wang, J. J. Chen, Y. Jiang, et al. Targeting GSTP1 as Therapeutic Strategy against Lung Adenocarcinoma Stemness and Resistance to Tyrosine Kinase Inhibitors. *Advanced Science*, 10(7):2205262, 3 2023.

[139] P. V. Ershov, E. O. Yablokov, L. A. Kaluzhskiy, et al. Prostanoid Signaling in Cancers: Expression and Regulation Patterns of Enzymes and Receptors. *Biology*, 11(4), 4 2022.

[140] B. Wang, L. Wu, J. Chen, et al. Metabolism pathways of arachidonic acids: mechanisms and potential therapeutic targets. *Signal Transduction and Targeted Therapy 2021 6:1*, 6(1):1–30, 2 2021.

[141] M. Karlgren, A. Gomez, K. Stark, et al. Tumor-specific expression of the novel cytochrome P450 enzyme, CYP2W1. *Biochemical and Biophysical Research Communications*, 341(2):451–458, 3 2006.

[142] M. Karlgen and M. Ingelman-Sundberg. Tumour-specific expression of CYP2W1: Its potential as a drug target in cancer therapy. *Expert Opinion on Therapeutic Targets*, 11(1):61–67, 1 2007.

[143] D. Edler, K. Stenstedt, K. Öhrling, et al. The expression of the novel CYP2W1 enzyme is an independent prognostic factor in colorectal cancer - a pilot study. *European journal of cancer (Oxford, England : 1990)*, 45(4):705–712, 3 2009.

[144] X. Shao, N. Lv, J. Liao, et al. Copy number variation is highly correlated with differential gene expression: A pan-cancer study. *BMC Medical Genetics*, 20(1):1–14, 11 2019.

[145] R. Tremmel, K. Klein, F. Battke, et al. Copy number variation profiling in pharmacogenes using panel-based exome resequencing and correlation to human liver expression. *Human genetics*, 139(2):137–149, 2 2020.

[146] D. Hamdane, C. Xia, S. C. Im, et al. Structure and Function of an NADPH-Cytochrome P450 Oxidoreductase in an Open Conformation Capable of Reducing Cytochrome P450. *The Journal of Biological Chemistry*, 284(17):11374, 4 2009.

[147] J. Cai, H. Sun, L. Chen, et al. NAT1 is a critical prognostic biomarker and inhibits proliferation of colorectal cancer through modulation of PI3K/Akt/mTOR. *Future oncology (London, England)*, 17(19):2489–2498, 7 2021.

[148] A. L. Hubbard, C. Moyes, A. H. Wyllie, et al. N-acetyl transferase 1: two polymorphisms in coding sequence identified in colorectal cancer patients. *British Journal of Cancer 1998 77:6*, 77(6):913–916, 1998.

[149] C. Shi, L. Y. Xie, Y. P. Tang, et al. Hypermethylation of N-Acetyltransferase 1 Is a Prognostic Biomarker in Colon Adenocarcinoma. *Frontiers in Genetics*, 10:485192, 11 2019.

[150] C. Lilla, E. Verla-Tebit, A. Risch, et al. Effect of NAT1 and NAT2 Genetic Polymorphisms on Colorectal Cancer Risk Associated with Exposure to Tobacco Smoke and Meat Consumption. *Cancer Epidemiology, Biomarkers & Prevention*, 15(1):99–107, 1 2006.

[151] J. Chen, M. Shi, S. Y. Chuen Choi, et al. Genomic alterations in neuroendocrine prostate cancer: A systematic review and meta□analysis. *BJUI Compass*, 4(3):256, 5 2023.

[152] R. Peculis, I. Mandrika, R. Petrovska, et al. Pituispheres Contain Genetic Variants Characteristic to Pituitary Adenoma Tumor Tissue. *Frontiers in Endocrinology*, 11:313, 5 2020.

[153] R. Jaenisch and A. Bird. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33(3S), 2003.

[154] A. P. Feinberg. Phenotypic plasticity and the epigenetics of human disease. *Nature*, 447(7143):433–440, 5 2007.

[155] P. Fisel, E. Schaeffeler, and M. Schwab. DNA Methylation of ADME Genes. *Clinical Pharmacology & Therapeutics*, 99(5):512–527, 5 2016.

[156] L. Möhrmann, M. Werner, M. Oleś, et al. Comprehensive genomic and epigenomic analysis in cancer of unknown primary guides molecularly-informed therapies despite heterogeneity. *Nature Communications 2022 13:1*, 13(1):1–15, 8 2022.

[157] C. Xie, Y. K. Leung, A. Chen, et al. Differential methylation values in differential methylation analysis. *Bioinformatics*, 35(7):1094, 4 2019.

[158] P. Du, X. Zhang, C. C. Huang, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11, 2010.

[159] A. Gomez, M. Karlgren, D. Edler, et al. Expression of CYP2W1 in colon tumors: regulation by gene methylation. *Pharmacogenomics*, 8(10):1315–1325, 10 2007.

[160] A. Ullah, J. Khan, A. Waheed, et al. Parathyroid Carcinoma: Incidence, Survival Analysis, and Management: A Study from the SEER Database and Insights into Future Therapeutic Perspectives. *Cancers*, 14(6):1426, 3 2022.

[161] E. C. Holmes, D. L. Morton, and A. S. Ketcham. Parathyroid carcinoma: a collective review. *Annals of surgery*, 169(4):631–640, 1969.

[162] M. V. Teleanu, C. T. Fuss, N. Paramasivam, et al. Targeted therapy of advanced parathyroid carcinoma guided by genomic and transcriptomic profiling. *Mol Oncol*, 17(7), 7 2023.

[163] D. Hübschmann, L. Jopp-Saile, C. Andresen, et al. Analysis of mutational signatures with yet another package for signature analysis. *Genes Chromosomes and Cancer*, 60(5):314–331, 5 2021.

[164] G. Sturm, F. Finotello, F. Petitprez, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics (Oxford, England)*, 35(14):i436–i445, 2019.

[165] E. Becht, N. A. Giraldo, L. Lacroix, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 17(1):218, 10 2016.

[166] F. Petitprez, Y. A. Vano, E. Becht, et al. Transcriptomic analysis of the tumor microenvironment to guide prognosis and immunotherapies. *Cancer Immunology, Immunotherapy*, 67(6), 2018.

[167] D. Aran, Z. Hu, and A. J. Butte. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology*, 18(1):220, 11 2017.

[168] A. M. Newman, C. B. Steen, C. L. Liu, et al. CIBERSORT. *Nature Biotechnology*, 37(7), 2019.

[169] B. Chen, M. S. Khodadoust, C. L. Liu, et al. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods in Molecular Biology*, 1711, 2018.

[170] A. M. Newman, C. L. Liu, M. R. Green, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods 2015 12:5*, 12(5):453–457, 3 2015.

[171] F. Finotello, C. Mayer, C. Plattner, et al. quanTIseq: quantifying immune contexture of human tumors. *bioRxiv*, 2017.

[172] C. Plattner, F. Finotello, and D. Rieder. Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. *Methods in Enzymology*, 636, 2020.

[173] T. Li, J. Fan, B. Wang, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Research*, 77(21), 2017.

[174] B. Li, E. Severson, J. C. Pignon, et al. Comprehensive analyses of tumor immunity: Implications for cancer immunotherapy. *Genome Biology*, 17(1):1–16, 8 2016.

[175] G. Assié, E. Letouzé, M. Fassnacht, et al. Integrated genomic characterization of adrenocortical carcinoma. *Nature Genetics*, 46(6):607–612, 4 2014.

[176] S. Zheng, A. D. Cherniack, N. Dewal, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*, 29(5):723–736, 5 2016.

[177] C. H. Mermel, S. E. Schumacher, B. Hill, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology*, 12(4), 4 2011.

[178] R. M. Paragliola, F. Torino, G. Papi, et al. Role of Mitotane in Adrenocortical Carcinoma – Review and State of the art. *European Endocrinology*, 14(2):62, 2018.

[179] M. Lo Iacono, S. Puglisi, P. Perotti, et al. Molecular Mechanisms of Mitotane Action in Adrenocortical Cancer Based on In Vitro Studies. *Cancers*, 13(21), 11 2021.

[180] C. R. Corso, A. Acco, C. Bach, et al. Pharmacological profile and effects of mitotane in adrenocortical carcinoma. *British Journal of Clinical Pharmacology*, 87(7):2698–2710, 7 2021.

[181] A. D'avolio, S. De Francia, V. Basile, et al. Influence of the CYP2B6 polymorphism on the pharmacokinetics of mitotane. *Pharmacogenetics and Genomics*, 23(6):293–300, 2013.

[182] B. Altieri, S. Sbiera, S. Herterich, et al. Effects of germline CYP2W1*6 and CYP2B6*6 single nucleotide polymorphisms on mitotane treatment in adrenocortical carcinoma: A multicenter ENSAT study. *Cancers*, 12(2), 2 2020.

[183] A. Yin, M. H. Ettaieb, J. J. Swen, et al. Population Pharmacokinetic and Pharmacogenetic Analysis of Mitotane in Patients with Adrenocortical Carcinoma: Towards Individualized Dosing. *Clinical Pharmacokinetics*, 60(1):89, 1 2021.

[184] S. H. Bakker, W. C. Jacobs, W. Pondaag, et al. Chordoma: a systematic review of the epidemiology and clinical prognostic factors predicting progression-free and overall survival. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*, 27(12):3043–3058, 12 2018.

[185] S. M. Barber, S. S. Sadrameli, J. J. Lee, et al. Chordoma—Current Understanding and Modern Treatment Paradigms. *Journal of Clinical Medicine*, 10(5):1–18, 3 2021.

[186] P. P. Varga, Z. Szövérfi, C. G. Fisher, et al. Surgical treatment of sacral chordoma: prognostic variables for local recurrence and overall survival. *European Spine Journal*, 24(5):1092–1101, 5 2015.

[187] T. Ailon, R. Torabi, C. G. Fisher, et al. Management of locally recurrent chordoma of the mobile spine and sacrum: A systematic review. *Spine*, 41:S193–S198, 10 2016.

[188] Z. Gu, M. Schlesner, and D. Hübschmann. cola: an R/Bioconductor package for consensus partitioning through a general framework. *Nucleic acids research*, 49(3):e15, 2 2021.

[189] Z. Gu and D. Hübschmann. simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results. *Genomics, proteomics & bioinformatics*, 21(1):190–202, 2 2023.

[190] C. A. Lunenburg, C. H. van der Wouden, M. Nijenhuis, et al. Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene–drug interaction of DPYD and fluoropyrimidines. *European Journal of Human Genetics*, 28(4):508–517, 4 2020.

[191] M. van der Lee, M. Kriek, H.-J. Guchelaar, and J. J. Swen. Technologies for Pharmacogenomics: A Review. *Genes*, 11(12), 2020.

[192] J. J. Swen, C. H. van der Wouden, L. E. Manson, et al. A 12-gene pharmacogenetic panel to prevent adverse drug reactions: an open-label, multicentre, controlled, cluster-randomised crossover implementation study. *The Lancet*, 401(10374):347–356, 2 2023.

[193] T. Huebner, M. Steffens, and C. Scholl. Current status of the analytical validation of next generation sequencing applications for pharmacogenetic profiling. *Molecular Biology Reports*, 50(11):9587–9599, 11 2023.

[194] M. van der Lee, W. G. Allard, S. Bollen, et al. Repurposing of Diagnostic Whole Exome Sequencing Data of 1,583 Individuals for Clinical Pharmacogenetics. *Clinical Pharmacology and Therapeutics*, 107(3), 2020.

[195] S. Reisberg, K. Krebs, M. Lepamets, et al. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. *Genetics in medicine : official journal of the American College of Medical Genetics*, 21(6):1345–1354, 6 2019.

[196] I. Numanagić, S. Malikić, V. M. Pratt, et al. Cypiripi: exact genotyping of CYP2D6 using high-throughput sequencing data. *Bioinformatics*, 31(12):27–34, 2015.

[197] D. Twesigomwe, B. I. Drögemöller, G. E. Wright, et al. StellarPGx: A Nextflow pipeline for calling star alleles in cytochrome P450 genes. *Clinical Pharmacology & Therapeutics*, 2021.

[198] R. Tremmel, S. Pirmann, Y. Zhou, and V. M. Lauschke. Translating pharmacogenomic sequencing data into drug response predictions-How to interpret variants of unknown significance. *British journal of clinical pharmacology*, 2023.

[199] D. Twesigomwe, G. E. B. Wright, B. I. Drögemöller, et al. A systematic comparison of pharmacogene star allele calling bioinformatics algorithms: a focus on CYP2D6 genotyping. *NPJ genomic medicine*, 5:30, 2020.

[200] A. Tafazoli, M. van der Lee, J. J. Swen, et al. Development of an extensive workflow for comprehensive clinical pharmacogenomic profiling: lessons from a pilot study on 100 whole exome sequencing data. *The Pharmacogenomics Journal*, 22(5-6):276, 12 2022.

[201] S. M. Caspar, T. Schneider, J. Meienberg, and G. Matyas. Added value of clinical sequencing: WGS-based profiling of pharmacogenes. *International Journal of Molecular Sciences*, 21(7), 4 2020.

[202] T. E. Klein and M. D. Ritchie. PharmCAT: A Pharmacogenomics Clinical Annotation Tool. *Clinical Pharmacology & Therapeutics*, 104(1):19–22, 7 2018.

[203] K. Sangkuhl, M. Whirl-Carrillo, R. M. Whaley, et al. Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clinical pharmacology and therapeutics*, 107(1):203–210, 2020.

[204] B. Li, K. Sangkuhl, K. Keat, et al. How to Run the Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clinical Pharmacology & Therapeutics*, 113(5):1036–1047, 5 2023.

[205] N. M. Reizine and P. H. O'Donnell. Modern developments in germline pharmacogenomics for oncology prescribing. *CA: A Cancer Journal for Clinicians*, 72(4):315–332, 7 2022.

[206] B. Mittal, S. Tulsyan, S. Kumar, et al. Cytochrome P450 in Cancer Susceptibility and Treatment. *Advances in clinical chemistry*, 71:77–139, 2015.

[207] W. Schroth, L. Antoniadou, P. Fritz, et al. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(33):5187–5193, 11 2007.

[208] M. P. Goetz, S. K. Knox, V. J. Suman, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Research and Treatment*, 101(1):113–121, 1 2007.

[209] M. P. Goetz, K. Sangkuhl, H. J. Guchelaar, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and Tamoxifen Therapy. *Clinical pharmacology and therapeutics*, 103(5):770–777, 5 2018.

[210] J. M. Nardin, W. Schroth, T. A. Almeida, et al. The Influences of Adherence to Tamoxifen and CYP2D6 Pharmacogenetics on Plasma Concentrations of the Active Metabolite (Z)-Endoxifen in Breast Cancer. *Clinical and translational science*, 13(2):284–292, 3 2020.

[211] M. V. Relling, E. E. Gardner, W. J. Sandborn, et al. Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clinical Pharmacology and Therapeutics*, 89(3):387–391, 2011.

[212] J. J. Yang, W. Landier, W. Yang, et al. Inherited NUDT15 variant is a genetic determinant of mercaptopurine intolerance in children with acute lymphoblastic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(11):1235–1242, 4 2015.

[213] T. Moriyama, R. Nishii, V. Perez-Andreu, et al. NUDT15 polymorphisms alter thiopurine metabolism and hematopoietic toxicity. *Nature genetics*, 48(4):367–373, 3 2016.

[214] E. Schaeffeler, S. U. Jaeger, V. Klumpp, et al. Impact of NUDT15 genetics on severe thiopurine-related hematotoxicity in patients with European ancestry. *Genetics in medicine : official journal of the American College of Medical Genetics*, 21(9):2145–2150, 9 2019.

[215] D. S. Wang, C. H. Yu, C. Y. Lin, et al. Childhood acute lymphoblastic leukemia mercaptopurine intolerance is associated with NUDT15 variants. *Pediatric Research 2020 89:1*, 89(1):217–222, 3 2020.

[216] K. E. Caudle, C. F. Thorn, T. E. Klein, et al. Clinical pharmacogenetics implementation consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clinical Pharmacology and Therapeutics*, 94(6):640–645, 12 2013.

[217] U. Amstutz, L. M. Henricks, S. M. Offer, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Dihydropyrimidine Dehydrogenase Genotype and Fluoropyrimidine Dosing: 2017 Update. *Clinical pharmacology and therapeutics*, 103(2):210–216, 2 2018.

[218] F. Innocenti, S. C. Mills, H. Sanoff, et al. All You Need to Know About DPYD Genetic Testing for Patients Treated With Fluorouracil and Capecitabine: A Practitioner-Friendly Guide . *JCO Oncology Practice*, 16(12):793–798, 12 2020.

[219] M. H. Diekstra, J. J. Swen, H. Gelderblom, and H. J. Guchelaar. A decade of pharmacogenomics research on tyrosine kinase inhibitors in metastatic renal cell cancer: a systematic review. *Expert review of molecular diagnostics*, 16(5):605–618, 5 2016.

[220] C. Neul, E. Schaeffeler, A. Sparreboom, et al. Impact of Membrane Drug Transporters on Resistance to Small-Molecule Tyrosine Kinase Inhibitors. *Trends in pharmacological sciences*, 37(11):904–932, 11 2016.

[221] R. Kundra, H. Zhang, R. Sheridan, et al. OncoTree: A Cancer Classification System for Precision Oncology. *JCO clinical cancer informatics*, 5(5):221–230, 12 2021.

[222] A. W. Langlois, J. G. Pouget, J. Knight, et al. Associating CYP2A6 structural variants with ovarian and lung cancer risk in the UK Biobank: replication and extension. *European Journal of Human Genetics 2023*, pages 1–4, 12 2023.

[223] A. Y. Angstadt, A. Berg, J. Zhu, et al. The effect of copy number variation (CNV) in the phase II detoxification genes, UGT2B17 and UGT2B28, on colorectal cancer risk. *Cancer*, 119(13):2477, 7 2013.

[224] S. Arslan, Y. Silig, and H. Pinarbasi. Sulfotransferase 1A1 Arg213His polymorphism and prostate cancer risk. *Experimental and Therapeutic Medicine*, 2(6):1159, 11 2011.

[225] Y. Lu, X. Zhu, C. Zhang, et al. Role of CYP2E1 polymorphisms in breast cancer: A systematic review and meta-analysis. *Cancer Cell International*, 17(1), 2017.

[226] Z. Fang, Y. Wu, and N. Zhang. Association between CYP2E1 genetic polymorphisms and urinary cancer risk: A meta-analysis. *Oncotarget*, 8(49), 2017.

[227] M. A. K. Sharzehan, H. Sito, N. Abdullah, et al. Association between CYP2E1 polymorphisms and colorectal cancer risk: a systematic review and meta-analysis. *Scientific Reports*, 12(1), 2022.

[228] C. Rodriguez-Antona, A. Gomez, M. Karlgren, et al. Molecular genetics and epigenetics of the cytochrome P450 gene family and its relevance for cancer risk and treatment. *Human Genetics*, 127(1), 2010.

[229] E. G. Hrycay and S. M. Bandiera. Involvement of Cytochrome P450 in Reactive Oxygen Species Formation and Cancer. *Advances in Pharmacology*, 74, 2015.

[230] K. Damjanovich, E. E. Baldwin, T. Lewis, and P. Bayrak-Toydemir. Novel homozygous CYP1B1 deletion in siblings with primary congenital glaucoma. *Ophthalmic genetics*, 34(3):180–181, 9 2013.

[231] E. Souzeau, A. Dubowsky, J. B. Ruddle, and J. E. Craig. Primary congenital glaucoma due to paternal uniparental isodisomy of chromosome 2 and CYP1B1 deletion. *Molecular Genetics & Genomic Medicine*, 7(8):e774, 8 2019.

[232] M. Santos, M. Niemi, M. Hiratsuka, et al. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genetics in Medicine*, 20(6):622–629, 6 2018.

[233] Y. Zhou, K. Fujikura, S. Mkrtchian, and V. M. Lauschke. Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. *Frontiers in Pharmacology*, 0:1437, 2018.

[234] Y. Zhou, S. Koutsilieri, E. Eliasson, and V. M. Lauschke. A paradigm shift in pharmacogenomics: From candidate polymorphisms to comprehensive sequencing. *Basic & Clinical Pharmacology & Toxicology*, 131(6):452–464, 12 2022.

[235] Y. Zhou, R. Tremmel, E. Schaeffeler, et al. Challenges and opportunities associated with rare-variant pharmacogenomics. *Trends in pharmacological sciences*, 43(10):852–865, 10 2022.

[236] T. H. Wroblewski, K. E. Witt, S.-b. Lee, et al. Pharmacogenetic Variation in Neanderthals and Denisovans and Implications for Human Health and Response to Medications. *Genome Biology and Evolution*, 15(12), 12 2023.

[237] Y. Zhou and V. M. Lauschke. Computational Tools to Assess the Functional Consequences of Rare and Noncoding Pharmacogenetic Variability. *Clinical Pharmacology and Therapeutics*, 110(3):626–636, 9 2021.

[238] V. M. Lauschke, L. Milani, and M. Ingelman-Sundberg. Pharmacogenomic Biomarkers for Improved Drug Therapy—Recent Progress and Future Developments. *AAPS Journal*, 20(1), 1 2018.

[239] J. A. Kim, R. Ceccarelli, and C. Y. Lu. Pharmacogenomic Biomarkers in US FDA-Approved Drug Labels (2000–2020). *Journal of Personalized Medicine*, 11(3):1–13, 2021.

[240] H. V, H. P, and S. P. Pharmacogenomics to Predict Tumor Therapy Response: A Focus on ATP-Binding Cassette Transporters and Cytochromes P450. *Journal of personalized medicine*, 10(3):1–18, 9 2020.

[241] R. W. Robey, K. M. Pluchino, M. D. Hall, et al. Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nature Reviews Cancer 2018 18:7*, 18(7):452–464, 4 2018.

[242] H. Xiao, Y. Zheng, L. Ma, et al. Clinically-Relevant ABC Transporter for Anti-Cancer Drug Resistance. *Frontiers in Pharmacology*, 12, 4 2021.

[243] M. Rheinnecker, M. Fröhlich, M. Rübsam, et al. ZygosityPredictor. *bioRxiv*, page 2023.03.09.531877, 3 2023.

[244] K. Ovens and C. Naugler. Preliminary evidence of different selection pressures on cancer cells as compared to normal tissues. *Theoretical Biology & Medical Modelling*, 9(1):44, 2012.

[245] I. Genovese, A. Ilari, Y. G. Assaraf, et al. Not only P-glycoprotein: Amplification of the ABCB1-containing chromosome region 7q21 confers multidrug resistance upon cancer cells by coordinated overexpression of an assortment of resistance-related proteins. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy*, 32:23–46, 5 2017.

[246] M. D. Nyquist, B. Prasad, and E. A. Mostaghel. Harnessing Solute Carrier Transporters for Precision Oncology. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, 22(4), 4 2017.

[247] R. Sun, Y. Ying, Z. Tang, et al. The Emerging Role of the SLCO1B3 Protein in Cancer Resistance. *Protein and Peptide Letters*, 27(1):17, 9 2020.

[248] N. Thakkar, K. Kim, E. R. Jang, et al. A cancer-specific variant of the slco1b3 gene encodes a novel human organic anion transporting polypeptide 1B3 (OATP1B3) localized mainly in the cytoplasm of colon and pancreatic cancer cells. *Molecular Pharmaceutics*, 10(1):406–416, 1 2013.

[249] S. G. Khan, V. Muniz-Medina, T. Shahlavi, et al. The human XPC DNA repair gene: arrangement, splice site information content and influence of a single nucleotide polymorphism in a splice acceptor site on alternative splicing and function. *Nucleic Acids Research*, 30(16):3624, 8 2002.

[250] N. A. Nasrallah, B. M. Wiese, and C. R. Sears. Xeroderma Pigmentosum Complementation Group C (XPC): Emerging Roles in Non-Dermatologic Malignancies. *Frontiers in Oncology*, 12, 4 2022.

[251] S. Chen, D. Wu, Q. Liu, et al. POR overexpression induces tamoxifen-resistance in breast cancer through the STAT1/c-Myc pathway. *Molecular carcinogenesis*, 62(2):249–260, 2 2023.

[252] E. Choong, J. Guo, A. Persson, et al. Developmental regulation and induction of cytochrome P450 2W1, an enzyme expressed in colon tumors. *PLoS ONE*, 10(4), 4 2015.

[253] S. Travica, K. Pors, P. M. Loadman, et al. Colon cancer-specific cytochrome P450 2W1 converts duocarmycin analogues into potent tumor cytotoxins. *Clinical Cancer Research*, 19(11):2952–2961, 6 2013.

[254] J. F. Gagnon, O. Bernard, L. Villeneuve, et al. Irinotecan inactivation is modulated by epigenetic silencing of UGT1A1 in colon cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12(6):1850–1858, 3 2006.

[255] A. S. Bélanger, J. Tojcic, M. Harvey, and C. Guillemette. Regulation of UGT1A1 and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells. *BMC Molecular Biology*, 11, 1 2010.

[256] M. Schnekenburger, T. Karius, and M. Diederich. Regulation of epigenetic traits of the glutathione S-transferase P1 gene: from detoxification toward cancer prevention and diagnosis. *Frontiers in Pharmacology*, 5, 2014.

[257] F. Marini, F. Giusti, G. Palmini, et al. Genetics and Epigenetics of Parathyroid Carcinoma. *Frontiers in Endocrinology*, 13, 2 2022.

[258] S. Y. Jo, N. Hong, S. Lee, et al. Genomic and transcriptomic profiling reveal molecular characteristics of parathyroid carcinoma. *Experimental & Molecular Medicine*, 55(5):886, 5 2023.

[259] M. Solarski, F. Rotondo, W. D. Foulkes, et al. DICER1 gene mutations in endocrine tumors. *Endocrine-Related Cancer*, 25(3):R197–R208, 3 2018.

[260] L. Baglietto, N. M. Lindor, J. G. Dowty, et al. Risks of Lynch Syndrome Cancers for MSH6 Mutation Carriers. *JNCI Journal of the National Cancer Institute*, 102(3):193, 2 2010.

[261] P. J. Newey, M. R. Bowl, and R. V. Thakker. Parafibromin–functional insights. *Journal of internal medicine*, 266(1):84–98, 7 2009.

[262] S. K. Gara, J. Lack, L. Zhang, et al. Metastatic adrenocortical carcinoma displays higher mutation rate and tumor heterogeneity than primary tumors. *Nature communications*, 9(1):4172, 12 2018.

[263] T. Fojo, L. Huff, T. Litman, et al. Metastatic and recurrent adrenocortical cancer is not defined by its genomic landscape. *BMC Medical Genomics*, 13(1):165, 12 2020.

[264] J. H. Oh, S. J. Jang, J. Kim, et al. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *npj Genomic Medicine 2020 5:1*, 5(1):1–11, 1 2020.

[265] C. C. Juhlin, G. Goh, J. M. Healy, et al. Whole-exome sequencing characterizes the landscape of somatic mutations and copy number alterations in adrenocortical carcinoma. *Journal of Clinical Endocrinology and Metabolism*, 100(3):E493–E502, 3 2015.

[266] L.-S. S. Landwehr, B. Altieri, J. Schreiner, et al. Interplay between glucocorticoids and tumor-infiltrating lymphocytes on the prognosis of adrenocortical carcinoma. *Journal for immunotherapy of cancer*, 8(1), 5 2020.

[267] N. Georgantzoglou, S. Kokkali, G. Tsouroufl is, and S. Theocharis. Tumor microenvironment in adrenocortical carcinoma: Barrier to immunotherapy success? *Cancers*, 13(8), 4 2021.

[268] T. Sharifnia, M. J. Wawer, A. Goodale, et al. Mapping the landscape of genetic dependencies in chordoma. *Nature Communications*, 14(1):1–17, 4 2023.

[269] N. Presneau, A. Shalaby, H. Ye, et al. Role of the transcription factor T (brachyury) in the pathogenesis of sporadic chordoma: a genetic and functional-based study. *The Journal of pathology*, 223(3):327–335, 2 2011.

[270] J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, et al. Integrative Genomics Viewer. *Nature biotechnology*, 29(1):24, 1 2011.

[271] A. Mayakonda, D. C. Lin, Y. Assenov, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Research*, 28(11):1747–1756, 11 2018.

[272] S. Lee, D. Cook, and M. Lawrence. Plyranges: A grammar of genomic data transformation. *Genome Biology*, 20(1):1–10, 1 2019.

[273] B. Gel and E. Serra. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics (Oxford, England)*, 33(19):3088–3090, 10 2017.

[274] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):1–21, 12 2014.

[275] S. Carbon, E. Douglass, N. Dunn, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1):D330–D338, 1 2019.

Appendix

Table 6.2: Overview of integrated genotyping tools and supported genes.
G=Genotype, P=Phenotype, SV=Structural Variants

Gene	Gene Family	Aldy(4.3)	Cyrius(1.1.1)	PyPGX(0.19.0)	Stargazer(1.0.8)
ABCB1	Transporter			G	
ABCG2	Transporter			G, P	
CACNA1S	Transporter			G, P	G, P, SV
CFTR	Other	G		G, P	G, P, SV
COMT	Other	G			
CYP17A1	Phase I	G		G	G, P, SV
CYP19A1	Phase I	G		G	G, P, SV
CYP1A1	Phase I			G	G, P, SV
CYP1A2	Phase I	G, SV		G, SV	G, P, SV
CYP1B1	Phase I	G		G	G, P, SV
CYP26A1	Phase I	G		G, P, SV	G, P, SV
CYP2A13	Phase I	G		G	G, P, SV
CYP2A6	Phase I	G		G, P	G, P, SV
CYP2B6	Phase I	G		G, P	G, P, SV
CYP2C19	Phase I	G, SV	G, SV	G, P, SV	G, P, SV
CYP2C8	Phase I	G		G, SV	G, P, SV
CYP2C9	Phase I	G		G	G, P, SV
CYP2D6	Phase I	G		G	G, P, SV
CYP2E1	Phase I	G		G	G, P, SV
CYP2F1	Phase I	G		G	G, P, SV
CYP2J2	Phase I	G		G	G, P, SV
CYP2R1	Phase I	G		G	G, P, SV
CYP2S1	Phase I	G		G, P	G, P, SV
CYP2W1	Phase I	G		G	G, P, SV
CYP3A4	Phase I	G		G	G, P, SV
CYP3A43	Phase I			G	
CYP3A5	Phase I			G	
CYP3A7	Phase I			G	G, P, SV
CYP4A11	Phase I	G		G, SV	G, P, SV
CYP4A22	Phase I			G	
CYP4B1	Phase I			G	G, P, SV
CYP4F2	Phase I			G	G, P, SV
DPYD	Phase I	G		G, P	G, P, SV
F5	Other			G, P	
G6PD	Other	G		G, SV	G, P, SV
GSTM1	Phase II	G, SV		G, SV	G, P, SV
GSTP1	Phase II	G		G	G, P, SV
GSTT1	Phase II			G, SV	G, P, SV
IFNL3	Other	G		G, P	G, P, SV
NAT1	Phase II	G		G	G, P, SV
NAT2	Phase II	G		G	G, P, SV
NUDT15	Phase II	G		G, P	G, P, SV
POR	Other			G	G, P, SV
PTGIS	Other			G	
RYR1	Other			G, P	G, P, SV
SLC15A2	Transporter			G	G, P, SV
SLC22A2	Transporter			G, SV	G, P, SV
SLCO1B1	Transporter	G		G, P	G, P, SV
SLCO1B3	Transporter			G	G, P, SV
SLCO2B1	Transporter			G	G, P, SV
SULT1A1	Phase II			G, SV	G, P, SV
TBXAS1	Other			G	G, P, SV
TPMT	Phase II	G		G, P	G, P, SV
UGT1A1	Phase II	G		G, P	G, P, SV
UGT1A4	Phase II			G, SV	G, P, SV
UGT2B15	Phase II	G		G	G, P, SV
UGT2B17	Phase II			G, SV	G, P, SV
UGT2B7	Phase II			G, SV	G, P, SV
VKORC1	Other	G		G	G, P, SV
XPC	Other			G	

Table 6.3: Genotyping results from PGx pipeline for MASTER cohort. For some samples and genes no genotyping tool could provide a result (None)

Gene	Consensus Genotype	n	Frequency
ABCB1	*1/*2	1192	49.75%
ABCB1	*2/*2	788	32.89%
ABCB1	*1/*1	416	17.36%
ABCG2	Reference/Reference	1955	81.59%
ABCG2	Reference/rs2231142	416	17.36%
ABCG2	rs2231142/rs2231142	25	1.04%
CACNA1S	Reference/Reference	2396	100.00%
CFTR	*WT/*WT	2303	96.12%
CFTR	Reference/F508del	53	2.21%
CFTR	Reference/R117H	13	0.54%
CFTR	Reference/Reference	13	0.54%
CFTR	Reference/F1052V	5	0.21%
CFTR	Reference/D1152H	2	0.08%
CFTR	Reference/G1069R	2	0.08%
CFTR	None	1	0.04%
CFTR	R117H/F508del	1	0.04%
CFTR	Reference/F1074L	1	0.04%
CFTR	Reference/G551D	1	0.04%
CFTR	Reference/R74W	1	0.04%
COMT	*Met/*ValA	935	39.02%
COMT	*Met/*Met	620	25.88%
COMT	*ValA/*ValA	359	14.98%
COMT	*Met/*ValB	209	8.72%
COMT	*ValA/*ValB	154	6.43%
COMT	*ValB/*ValB	62	2.59%
COMT	*Met/*ValC	13	0.54%
COMT	*ValA/*ValC	8	0.33%
COMT	*ValA/*ValE	6	0.25%
COMT	*ValB/*ValC	5	0.21%
COMT	*A72S/*ValA	4	0.17%
COMT	*ValD/*ValD	3	0.13%
COMT	*ValE/*ValE	3	0.13%
COMT	*Met/*MetB	2	0.08%
COMT	*ValC/*ValC	2	0.08%
COMT	*A72S/*Met	1	0.04%
COMT	*A72S/*ValB	1	0.04%
COMT	*Met/*ValD	1	0.04%
COMT	*Met/*ValE	1	0.04%
COMT	*MetB/*ValA	1	0.04%
COMT	*ValB/*ValE	1	0.04%
CYP17A1	Reference/Reference	2394	99.92%
CYP17A1	G90D/G90D	1	0.04%
CYP17A1	Reference/R347H	1	0.04%
CYP19A1	*1/*1	2068	86.31%
CYP19A1	*1/*4	178	7.43%
CYP19A1	*1/*3	135	5.63%
CYP19A1	*3/*4	6	0.25%
CYP19A1	*4/*4	3	0.13%
CYP19A1	*1/*2	2	0.08%
CYP19A1	None	2	0.08%
CYP19A1	*2/*4	1	0.04%
CYP19A1	*3/*3	1	0.04%
CYP1A1	*1/*1	1738	72.54%
CYP1A1	*1/*2	351	14.65%
CYP1A1	*1/*4	191	7.97%
CYP1A1	*2/*2	35	1.46%
CYP1A1	*1/*5	32	1.34%
CYP1A1	*2/*4	27	1.13%
CYP1A1	*4/*4	10	0.42%
CYP1A1	*1/*13	4	0.17%
CYP1A1	*2/*5	4	0.17%
CYP1A1	*2/*13	2	0.08%
CYP1A1	*13/*13	1	0.04%
CYP1A1	None	1	0.04%
CYP1A2	*1/*1	2394	99.92%
CYP1A2	*1/*3	1	0.04%
CYP1A2	*1/*6	1	0.04%
CYP1B1	*2/*3	583	24.33%
CYP1B1	*3/*3	420	17.53%
CYP1B1	*3/*4	379	15.82%
CYP1B1	*2/*4	247	10.31%
CYP1B1	*2/*2	224	9.35%
CYP1B1	*1/*3	177	7.39%
CYP1B1	*1/*2	132	5.51%
CYP1B1	*1/*4	85	3.55%
CYP1B1	*4/*4	70	2.92%
CYP1B1	None	30	1.25%
CYP1B1	*1/*1	27	1.13%
CYP1B1	*3/*6	7	0.29%
CYP1B1	*2/*6	5	0.21%
CYP1B1	*1/*6	2	0.08%
CYP1B1	*4/*6	2	0.08%
CYP1B1	*4/*7	2	0.08%
CYP1B1	*3/*5	1	0.04%

CYP1B1	*3/*7	1	0.04%
CYP1B1	*6/*6	1	0.04%
CYP1B1	*6/*7	1	0.04%
CYP26A1	*1/*1	2395	99.96%
CYP26A1	*1/*4	1	0.04%
CYP2A13	*1/*1	2294	95.74%
CYP2A13	*1/*2	42	1.75%
CYP2A13	*1/*7	42	1.75%
CYP2A13	*1/*8	13	0.54%
CYP2A13	*2/*2	3	0.13%
CYP2A13	*1/*3	2	0.08%
CYP2A6	*1/*1	1514	63.19%
CYP2A6	*1/*9	254	10.60%
CYP2A6	*1/*14	139	5.80%
CYP2A6	*1/*12	98	4.09%
CYP2A6	*1/*2	83	3.46%
CYP2A6	*1/*18	63	2.63%
CYP2A6	None	44	1.84%
CYP2A6	*1/*21	34	1.42%
CYP2A6	*1/*4	23	0.96%
CYP2A6	*9/*9	19	0.79%
CYP2A6	*9/*14	13	0.54%
CYP2A6	*1/*7	10	0.42%
CYP2A6	*1/*35	9	0.38%
CYP2A6	*1/*22	7	0.29%
CYP2A6	*9/*12	7	0.29%
CYP2A6	*2/*12	6	0.25%
CYP2A6	*1/*1x2	5	0.21%
CYP2A6	*18/*18	5	0.21%
CYP2A6	*9/*18	5	0.21%
CYP2A6	*1/*17	4	0.17%
CYP2A6	*1/*34	4	0.17%
CYP2A6	*2/*9	4	0.17%
CYP2A6	*9/*21	4	0.17%
CYP2A6	*1/*28	3	0.13%
CYP2A6	*12/*14	3	0.13%
CYP2A6	*12/*21	3	0.13%
CYP2A6	*14/*18	3	0.13%
CYP2A6	*1x2/*2	3	0.13%
CYP2A6	*12/*12	2	0.08%
CYP2A6	*12/*18	2	0.08%
CYP2A6	*1x2/*21	2	0.08%
CYP2A6	*2/*2	2	0.08%
CYP2A6	*9/*17	2	0.08%
CYP2A6	*1/*10	1	0.04%
CYP2A6	*1/*15	1	0.04%
CYP2A6	*1/*39	1	0.04%
CYP2A6	*1/*9x2	1	0.04%
CYP2A6	*14/*14	1	0.04%
CYP2A6	*14/*21	1	0.04%
CYP2A6	*18/*21	1	0.04%
CYP2A6	*1x2/*14	1	0.04%
CYP2A6	*1x2/*17	1	0.04%
CYP2A6	*1x2/*39	1	0.04%
CYP2A6	*1x2/*9	1	0.04%
CYP2A6	*2/*18	1	0.04%
CYP2A6	*2/*4	1	0.04%
CYP2A6	*4/*14	1	0.04%
CYP2A6	*4/*18	1	0.04%
CYP2A6	*4/*4	1	0.04%
CYP2A6	*4/*9	1	0.04%
CYP2B6	*1/*1	684	28.55%
CYP2B6	*1/*6	580	24.21%
CYP2B6	*1/*5	316	13.19%
CYP2B6	*6/*6	138	5.76%
CYP2B6	*1/*2	132	5.51%
CYP2B6	*5/*6	124	5.18%
CYP2B6	*2/*6	68	2.84%
CYP2B6	*1/*4	66	2.75%
CYP2B6	*2/*5	33	1.38%
CYP2B6	*4/*6	28	1.17%
CYP2B6	*1/*22	27	1.13%
CYP2B6	*5/*5	27	1.13%
CYP2B6	*1/*15	18	0.75%
CYP2B6	*6/*22	18	0.75%
CYP2B6	*1/*10	15	0.63%
CYP2B6	*6/*15	13	0.54%
CYP2B6	None	13	0.54%
CYP2B6	*1/*11	8	0.33%
CYP2B6	*2/*2	7	0.29%
CYP2B6	*2/*4	7	0.29%
CYP2B6	*4/*5	6	0.25%
CYP2B6	*5/*15	6	0.25%
CYP2B6	*6/*10	6	0.25%
CYP2B6	*1/*13	5	0.21%
CYP2B6	*2/*22	4	0.17%
CYP2B6	*1/*9	3	0.13%
CYP2B6	*4/*10	3	0.13%

CYP2B6	*5/*11	3	0.13%
CYP2B6	*2/*10	2	0.08%
CYP2B6	*2/*15	2	0.08%
CYP2B6	*3/*22	2	0.08%
CYP2B6	*4/*15	2	0.08%
CYP2B6	*4/*9	2	0.08%
CYP2B6	*5/*10	2	0.08%
CYP2B6	*5/*13	2	0.08%
CYP2B6	*5/*22	2	0.08%
CYP2B6	*6/*11	2	0.08%
CYP2B6	*6/*13	2	0.08%
CYP2B6	*6/*7	2	0.08%
CYP2B6	*1/*18	1	0.04%
CYP2B6	*1/*26	1	0.04%
CYP2B6	*1/*29	1	0.04%
CYP2B6	*1/*3	1	0.04%
CYP2B6	*1/*38	1	0.04%
CYP2B6	*1/*7	1	0.04%
CYP2B6	*10/*29	1	0.04%
CYP2B6	*11/*22	1	0.04%
CYP2B6	*15/*22	1	0.04%
CYP2B6	*2/*11	1	0.04%
CYP2B6	*2/*12	1	0.04%
CYP2B6	*2/*13	1	0.04%
CYP2B6	*2/*36	1	0.04%
CYP2B6	*4/*13	1	0.04%
CYP2B6	*6/*18	1	0.04%
CYP2B6	*9/*9	1	0.04%
CYP2C19	*1/*1	977	40.78%
CYP2C19	*1/*17	622	25.96%
CYP2C19	*1/*2	414	17.28%
CYP2C19	*2/*17	171	7.14%
CYP2C19	*17/*17	120	5.01%
CYP2C19	*2/*2	49	2.05%
CYP2C19	*1/*8	11	0.46%
CYP2C19	*1/*4	6	0.25%
CYP2C19	*1/*3	4	0.17%
CYP2C19	*1/*35	4	0.17%
CYP2C19	*8/*17	4	0.17%
CYP2C19	None	3	0.13%
CYP2C19	*1/*6	2	0.08%
CYP2C19	*2/*8	2	0.08%
CYP2C19	*3/*17	2	0.08%
CYP2C19	*1/*33	1	0.04%
CYP2C19	*15/*17	1	0.04%
CYP2C19	*17/*35	1	0.04%
CYP2C19	*2/*3	1	0.04%
CYP2C19	*35/*35	1	0.04%
CYP2C8	*1/*1	1656	69.12%
CYP2C8	*1/*3	408	17.03%
CYP2C8	*1/*4	205	8.56%
CYP2C8	*1/*15	34	1.42%
CYP2C8	*3/*3	33	1.38%
CYP2C8	*3/*4	26	1.09%
CYP2C8	*1/*2	10	0.42%
CYP2C8	*4/*4	8	0.33%
CYP2C8	*3/*15	5	0.21%
CYP2C8	*4/*15	4	0.17%
CYP2C8	*1/*7	2	0.08%
CYP2C8	*2/*3	2	0.08%
CYP2C8	*15/*15	1	0.04%
CYP2C8	*2/*2	1	0.04%
CYP2C8	*2/*4	1	0.04%
CYP2C9	*1/*1	1555	64.90%
CYP2C9	*1/*2	453	18.91%
CYP2C9	*1/*3	258	10.77%
CYP2C9	*2/*3	45	1.88%
CYP2C9	*2/*2	40	1.67%
CYP2C9	*1/*12	10	0.42%
CYP2C9	*3/*3	9	0.38%
CYP2C9	*1/*11	5	0.21%
CYP2C9	*1/*8	5	0.21%
CYP2C9	*2/*12	4	0.17%
CYP2C9	*1/*14	1	0.04%
CYP2C9	*1/*29	1	0.04%
CYP2C9	*1/*34	1	0.04%
CYP2C9	*1/*36	1	0.04%
CYP2C9	*1/*5	1	0.04%
CYP2C9	*1/*62	1	0.04%
CYP2C9	*1/*9	1	0.04%
CYP2C9	*2/*11	1	0.04%
CYP2C9	*2/*20	1	0.04%
CYP2C9	*2/*43	1	0.04%
CYP2C9	*3/*12	1	0.04%
CYP2C9	None	1	0.04%
CYP2D6	*1/*1	291	12.15%
CYP2D6	*1/*2	236	9.85%
CYP2D6	*1/*4	177	7.39%

CYP2D6	*1/*41	151	6.30%
CYP2D6	*1/*35	114	4.76%
CYP2D6	*1/*68+*4	103	4.30%
CYP2D6	*2/*2	69	2.88%
CYP2D6	*2/*4	65	2.71%
CYP2D6	*2/*41	60	2.50%
CYP2D6	*1/*5	56	2.34%
CYP2D6	None	49	2.05%
CYP2D6	*2/*68+*4	45	1.88%
CYP2D6	*4/*41	43	1.79%
CYP2D6	*2/*35	42	1.75%
CYP2D6	*4/*35	39	1.63%
CYP2D6	*4/*4	38	1.59%
CYP2D6	*1/*9	34	1.42%
CYP2D6	*4/*68+*4	32	1.34%
CYP2D6	*1/*10	25	1.04%
CYP2D6	*2/*5	25	1.04%
CYP2D6	*35/*41	23	0.96%
CYP2D6	*4/*5	23	0.96%
CYP2D6	*1/*3	22	0.92%
CYP2D6	*1/*33	22	0.92%
CYP2D6	*41/*68+*4	20	0.83%
CYP2D6	*35/*68+*4	19	0.79%
CYP2D6	*1/*6	18	0.75%
CYP2D6	*1/*4+*4	15	0.63%
CYP2D6	*35/*35	15	0.63%
CYP2D6	*2/*6	13	0.54%
CYP2D6	*2/*9	13	0.54%
CYP2D6	*2/*10	12	0.50%
CYP2D6	*9/*35	12	0.50%
CYP2D6	*1/*1x2	11	0.46%
CYP2D6	*41/*41	11	0.46%
CYP2D6	*9/*41	11	0.46%
CYP2D6	*10/*41	10	0.42%
CYP2D6	*5/*35	10	0.42%
CYP2D6	*4/*10	9	0.38%
CYP2D6	*5/*41	9	0.38%
CYP2D6	*1/*32	8	0.33%
CYP2D6	*1/*36+*10	8	0.33%
CYP2D6	*1/*59	8	0.33%
CYP2D6	*1x2/*2	8	0.33%
CYP2D6	*1x2/*41	8	0.33%
CYP2D6	*2/*3	8	0.33%
CYP2D6	*3/*4	8	0.33%
CYP2D6	*4/*9	8	0.33%
CYP2D6	*1/*2x2	7	0.29%
CYP2D6	*10/*68+*4	7	0.29%
CYP2D6	*1x2/*4	7	0.29%
CYP2D6	*3/*35	7	0.29%
CYP2D6	*3/*41	7	0.29%
CYP2D6	*10/*35	6	0.25%
CYP2D6	*2/*2x2	6	0.25%
CYP2D6	*2/*33	6	0.25%
CYP2D6	*2x2/*35	6	0.25%
CYP2D6	*5/*68+*4	6	0.25%
CYP2D6	*68+*4/*68+*4	6	0.25%
CYP2D6	*3/*5	5	0.21%
CYP2D6	*33/*41	5	0.21%
CYP2D6	*4/*13	5	0.21%
CYP2D6	*4/*6	5	0.21%
CYP2D6	*6/*68+*4	5	0.21%
CYP2D6	*9/*68+*4	5	0.21%
CYP2D6	*1/*116	4	0.17%
CYP2D6	*1/*22	4	0.17%
CYP2D6	*1/*28	4	0.17%
CYP2D6	*1/*2x3	4	0.17%
CYP2D6	*1/*4x2	4	0.17%
CYP2D6	*1x2/*68+*4	4	0.17%
CYP2D6	*2/*32	4	0.17%
CYP2D6	*2/*59	4	0.17%
CYP2D6	*3/*68+*4	4	0.17%
CYP2D6	*33/*68+*4	4	0.17%
CYP2D6	*4/*33	4	0.17%
CYP2D6	*4/*59	4	0.17%
CYP2D6	*6/*35	4	0.17%
CYP2D6	*1/*15	3	0.13%
CYP2D6	*1/*27	3	0.13%
CYP2D6	*13/*41	3	0.13%
CYP2D6	*1x2/*35	3	0.13%
CYP2D6	*2/*4+*4	3	0.13%
CYP2D6	*2x2/*4	3	0.13%
CYP2D6	*2x2/*41	3	0.13%
CYP2D6	*3/*10	3	0.13%
CYP2D6	*4+*4/*35	3	0.13%
CYP2D6	*4+*4/*41	3	0.13%
CYP2D6	*4/*4+*4	3	0.13%
CYP2D6	*41+*68/*68+*4	3	0.13%
CYP2D6	*5/*10	3	0.13%

CYP2D6	*5/*5	3	0.13%
CYP2D6	*5/*59	3	0.13%
CYP2D6	*5/*9	3	0.13%
CYP2D6	*6/*41	3	0.13%
CYP2D6	*1/*11	2	0.08%
CYP2D6	*1/*13	2	0.08%
CYP2D6	*1/*17	2	0.08%
CYP2D6	*1/*68x5+*4	2	0.08%
CYP2D6	*10/*36+*10	2	0.08%
CYP2D6	*13+*2/*6	2	0.08%
CYP2D6	*1x2/*9	2	0.08%
CYP2D6	*2/*13	2	0.08%
CYP2D6	*2/*13+*1	2	0.08%
CYP2D6	*2/*22	2	0.08%
CYP2D6	*2/*29	2	0.08%
CYP2D6	*2/*36+*10	2	0.08%
CYP2D6	*2/*41x2	2	0.08%
CYP2D6	*2/*43	2	0.08%
CYP2D6	*2/*4x2	2	0.08%
CYP2D6	*22/*68+*4	2	0.08%
CYP2D6	*28/*41	2	0.08%
CYP2D6	*2x2/*13	2	0.08%
CYP2D6	*2x2/*68+*4	2	0.08%
CYP2D6	*3/*13	2	0.08%
CYP2D6	*3/*59	2	0.08%
CYP2D6	*3/*9	2	0.08%
CYP2D6	*33/*33	2	0.08%
CYP2D6	*33/*35	2	0.08%
CYP2D6	*4/*108	2	0.08%
CYP2D6	*4/*28	2	0.08%
CYP2D6	*4/*7	2	0.08%
CYP2D6	*4x2/*33	2	0.08%
CYP2D6	*4x2/*41	2	0.08%
CYP2D6	*4x2/*68+*4	2	0.08%
CYP2D6	*59/*68+*4	2	0.08%
CYP2D6	*6/*33	2	0.08%
CYP2D6	*9x2/*10	2	0.08%
CYP2D6	*1/*122	1	0.04%
CYP2D6	*1/*124	1	0.04%
CYP2D6	*1/*125	1	0.04%
CYP2D6	*1/*1x3	1	0.04%
CYP2D6	*1/*21	1	0.04%
CYP2D6	*1/*29	1	0.04%
CYP2D6	*1/*2x4	1	0.04%
CYP2D6	*1/*31	1	0.04%
CYP2D6	*1/*35x2	1	0.04%
CYP2D6	*1/*41x3	1	0.04%
CYP2D6	*1/*43	1	0.04%
CYP2D6	*1/*45	1	0.04%
CYP2D6	*1/*68x2+*4	1	0.04%
CYP2D6	*1/*7	1	0.04%
CYP2D6	*1/*71	1	0.04%
CYP2D6	*1/*9x2	1	0.04%
CYP2D6	*10/*10	1	0.04%
CYP2D6	*10/*13	1	0.04%
CYP2D6	*10/*39	1	0.04%
CYP2D6	*13+*2/*35	1	0.04%
CYP2D6	*13+*2/*4	1	0.04%
CYP2D6	*13+*2/*9	1	0.04%
CYP2D6	*13/*22	1	0.04%
CYP2D6	*13/*35	1	0.04%
CYP2D6	*13/*68+*4	1	0.04%
CYP2D6	*17/*29	1	0.04%
CYP2D6	*17/*35	1	0.04%
CYP2D6	*17/*41	1	0.04%
CYP2D6	*1x2/*10	1	0.04%
CYP2D6	*1x2/*1x3	1	0.04%
CYP2D6	*1x4/*2	1	0.04%
CYP2D6	*2/*11	1	0.04%
CYP2D6	*2/*116	1	0.04%
CYP2D6	*2/*127	1	0.04%
CYP2D6	*2/*2x3	1	0.04%
CYP2D6	*2/*7	1	0.04%
CYP2D6	*2/*74	1	0.04%
CYP2D6	*22x2/*35	1	0.04%
CYP2D6	*27/*27	1	0.04%
CYP2D6	*28/*35	1	0.04%
CYP2D6	*2x2/*10	1	0.04%
CYP2D6	*2x2/*9	1	0.04%
CYP2D6	*2x4/*13	1	0.04%
CYP2D6	*32/*33	1	0.04%
CYP2D6	*32/*35	1	0.04%
CYP2D6	*35+*68/*68+*4	1	0.04%
CYP2D6	*35/*115	1	0.04%
CYP2D6	*35/*39	1	0.04%
CYP2D6	*35/*59	1	0.04%
CYP2D6	*36+*10/*36+*10	1	0.04%
CYP2D6	*4+*4/*22	1	0.04%

CYP2D6	*4+*4/*68+*4	1	0.04%
CYP2D6	*4/*19	1	0.04%
CYP2D6	*4/*22	1	0.04%
CYP2D6	*4/*35x2	1	0.04%
CYP2D6	*4/*36+*10	1	0.04%
CYP2D6	*4/*41x2	1	0.04%
CYP2D6	*4/*41x3	1	0.04%
CYP2D6	*4/*43	1	0.04%
CYP2D6	*4/*4x2	1	0.04%
CYP2D6	*41/*108x2	1	0.04%
CYP2D6	*41/*119	1	0.04%
CYP2D6	*41/*124	1	0.04%
CYP2D6	*41/*68x5+*4	1	0.04%
CYP2D6	*5/*28	1	0.04%
CYP2D6	*5/*6	1	0.04%
CYP2D6	*68+*4/*116	1	0.04%
CYP2D6	*82/*68+*68+*4	1	0.04%
CYP2D6	*9/*10	1	0.04%
CYP2D6	*9/*33	1	0.04%
CYP2E1	*1/*1	1645	68.66%
CYP2E1	*1/*7	459	19.16%
CYP2E1	None	119	4.97%
CYP2E1	*1x2/*7	73	3.05%
CYP2E1	*7/*7	67	2.80%
CYP2E1	*1/*7x2	18	0.75%
CYP2E1	*1/*1x2	7	0.29%
CYP2E1	*1/*3	4	0.17%
CYP2E1	*7/*7x2	3	0.13%
CYP2E1	*1/*4	1	0.04%
CYP2F1	*1/*1	862	35.98%
CYP2F1	*1/*2	532	22.20%
CYP2F1	*1/*5	262	10.93%
CYP2F1	*1/*3	168	7.01%
CYP2F1	*1/*4	139	5.80%
CYP2F1	*2/*2	75	3.13%
CYP2F1	*2/*5	75	3.13%
CYP2F1	*2/*3	64	2.67%
CYP2F1	*1/*6	41	1.71%
CYP2F1	*2/*4	33	1.38%
CYP2F1	*3/*5	32	1.34%
CYP2F1	*4/*5	24	1.00%
CYP2F1	*3/*3	17	0.71%
CYP2F1	*3/*4	13	0.54%
CYP2F1	*5/*5	13	0.54%
CYP2F1	*2/*6	9	0.38%
CYP2F1	*4/*6	8	0.33%
CYP2F1	*5/*6	8	0.33%
CYP2F1	*4/*4	7	0.29%
CYP2F1	None	7	0.29%
CYP2F1	*3/*6	5	0.21%
CYP2F1	*6/*6	2	0.08%
CYP2J2	*1/*1	2063	86.10%
CYP2J2	*1/*7	310	12.94%
CYP2J2	*7/*7	19	0.79%
CYP2J2	*1/*3	3	0.13%
CYP2J2	*1/*9	1	0.04%
CYP2R1	*1/*1	2396	100.00%
CYP2S1	*1/*1	2274	94.91%
CYP2S1	*1/*3	121	5.05%
CYP2S1	*1/*2	1	0.04%
CYP2W1	*1/*1	1413	58.97%
CYP2W1	*1/*6	614	25.63%
CYP2W1	*1/*2	214	8.93%
CYP2W1	*2/*6	72	3.01%
CYP2W1	*6/*6	60	2.50%
CYP2W1	*2/*2	23	0.96%
CYP3A4	*1/*1	2179	90.94%
CYP3A4	*1/*22	186	7.76%
CYP3A4	*1/*3	9	0.38%
CYP3A4	*1/*7	6	0.25%
CYP3A4	*22/*22	6	0.25%
CYP3A4	*1/*15	4	0.17%
CYP3A4	*1/*10	3	0.13%
CYP3A4	*1/*16	2	0.08%
CYP3A4	*10/*22	1	0.04%
CYP3A43	*1/*1	2075	86.60%
CYP3A43	*1/*2	213	8.89%
CYP3A43	*1/*3	95	3.96%
CYP3A43	*2/*2	7	0.29%
CYP3A43	*2/*3	3	0.13%
CYP3A43	*3/*3	2	0.08%
CYP3A43	None	1	0.04%
CYP3A5	*3/*3	2064	86.14%
CYP3A5	*1/*3	310	12.94%
CYP3A5	*1/*1	15	0.63%
CYP3A5	*3/*6	2	0.08%
CYP3A5	*1/*6	1	0.04%
CYP3A5	*3/*7	1	0.04%

CYP3A5	*6/*6	1	0.04%
CYP3A5	*6/*7	1	0.04%
CYP3A5	None	1	0.04%
CYP3A7	*1/*1	1954	81.55%
CYP3A7	*1/*2	404	16.86%
CYP3A7	*2/*2	34	1.42%
CYP3A7	None	4	0.17%
CYP4A11	*1/*1	1779	74.25%
CYP4A11	*1/F434S	568	23.71%
CYP4A11	F434S/F434S	47	1.96%
CYP4A11	*1/S353G	2	0.08%
CYP4A22	*1/*1	1055	44.03%
CYP4A22	*1/*5	421	17.57%
CYP4A22	*1/*12	376	15.69%
CYP4A22	*1/*15	244	10.18%
CYP4A22	*5/*12	60	2.50%
CYP4A22	*5/*15	49	2.05%
CYP4A22	*12/*12	44	1.84%
CYP4A22	*12/*15	44	1.84%
CYP4A22	*5/*5	36	1.50%
CYP4A22	*1/*9	23	0.96%
CYP4A22	*15/*15	16	0.67%
CYP4A22	*5/*9	6	0.25%
CYP4A22	*9/*12	5	0.21%
CYP4A22	*1/*3	3	0.13%
CYP4A22	*3/*12	3	0.13%
CYP4A22	*3/*15	2	0.08%
CYP4A22	*1/*13	1	0.04%
CYP4A22	*1/*4	1	0.04%
CYP4A22	*3/*5	1	0.04%
CYP4A22	*3/*9	1	0.04%
CYP4A22	*4/*4	1	0.04%
CYP4A22	*5/*13	1	0.04%
CYP4A22	*6/*6	1	0.04%
CYP4A22	*9/*15	1	0.04%
CYP4A22	*9/*9	1	0.04%
CYP4B1	*1/*1	1225	51.13%
CYP4B1	*1/*2	479	19.99%
CYP4B1	*1/*3	423	17.65%
CYP4B1	*2/*3	95	3.96%
CYP4B1	*2/*2	44	1.84%
CYP4B1	*1/*4	42	1.75%
CYP4B1	*3/*3	32	1.34%
CYP4B1	*1/*5	18	0.75%
CYP4B1	None	8	0.33%
CYP4B1	*3/*4	7	0.29%
CYP4B1	*2/*4	6	0.25%
CYP4B1	*2/*5	6	0.25%
CYP4B1	*3/*5	5	0.21%
CYP4B1	*5/*5	4	0.17%
CYP4B1	*4/*4	1	0.04%
CYP4B1	*4/*5	1	0.04%
CYP4F2	*1/*1	1302	54.34%
CYP4F2	None	635	26.50%
CYP4F2	*1/*3	345	14.40%
CYP4F2	*2/*3	71	2.96%
CYP4F2	*3/*3	31	1.29%
CYP4F2	*1/*2	10	0.42%
CYP4F2	*2/*2	2	0.08%
DPYD	*1/*1	646	26.96%
DPYD	*1/*5	393	16.40%
DPYD	None	268	11.19%
DPYD	*1/*9	169	7.05%
DPYD	*5/*9	100	4.17%
DPYD	*9/*rs2297595	98	4.09%
DPYD	*5/*5	90	3.76%
DPYD	*1/*6	84	3.51%
DPYD	*5/*rs2297595	58	2.42%
DPYD	*1/*HapB3	51	2.13%
DPYD	*5/*6	48	2.00%
DPYD	*1/*rs2297595	47	1.96%
DPYD	*1/*4	46	1.92%
DPYD	*9/*9	41	1.71%
DPYD	*1/*rs17376848	36	1.50%
DPYD	*4/*5	24	1.00%
DPYD	*5/*HapB3	21	0.88%
DPYD	*6/*9	20	0.83%
DPYD	*5/*rs17376848	19	0.79%
DPYD	*6/*rs2297595	16	0.67%
DPYD	*1/*2	15	0.63%
DPYD	*9/*rs17376848	15	0.63%
DPYD	*9/*HapB3	11	0.46%
DPYD	*1/*rs67376798	10	0.42%
DPYD	*4/*9	8	0.33%
DPYD	*6/*6	7	0.29%
DPYD	*rs2297595/*HapB3	6	0.25%
DPYD	*6/*rs17376848	5	0.21%
DPYD	*5/*rs67376798	4	0.17%

DPYD	*9/*rs45589337	3	0.13%
DPYD	*9/*rs67376798	3	0.13%
DPYD	*1/*rs145112791	2	0.08%
DPYD	*1/*rs61622928	2	0.08%
DPYD	*4/*6	2	0.08%
DPYD	*4/*rs2297595	2	0.08%
DPYD	*5/*rs45589337	2	0.08%
DPYD	*6/*HapB3	2	0.08%
DPYD	*9/*rs61622928	2	0.08%
DPYD	*HapB3/*HapB3	2	0.08%
DPYD	*rs2297595/*rs17376848	2	0.08%
DPYD	*1/*13	1	0.04%
DPYD	*1/*rs114096998	1	0.04%
DPYD	*2/*4	1	0.04%
DPYD	*2/*5	1	0.04%
DPYD	*2/*6	1	0.04%
DPYD	*2/*9	1	0.04%
DPYD	*2/*rs17376848	1	0.04%
DPYD	*4/*4	1	0.04%
DPYD	*4/*HapB3	1	0.04%
DPYD	*4/*rs67376798	1	0.04%
DPYD	*5/*7	1	0.04%
DPYD	*5/*rs59086055	1	0.04%
DPYD	*5/*rs72549308	1	0.04%
DPYD	*6/*rs148799944	1	0.04%
DPYD	*rs45589337/*HapB3	1	0.04%
DPYD	*rs67376798/*HapB3	1	0.04%
F5	Reference/Reference	2238	93.41%
F5	Reference/Leiden	150	6.26%
F5	Leiden/Leiden	8	0.33%
G6PD	*B/*B	1186	49.50%
G6PD	*B/*DEL	1181	49.29%
G6PD	None	9	0.38%
G6PD	*B/*seattle	6	0.25%
G6PD	*A-/*DEL	3	0.13%
G6PD	*A/*DEL	2	0.08%
G6PD	*B/*Gond	2	0.08%
G6PD	*B/*mediterranean	2	0.08%
G6PD	*mediterranean/*DEL	2	0.08%
G6PD	*A/*B	1	0.04%
G6PD	*Surabaya/*DEL	1	0.04%
G6PD	*Union,Maewo,Chinese-2,Kalo/*DEL	1	0.04%
GSTM1	*0/*0	722	30.13%
GSTM1	*0/*A	662	27.63%
GSTM1	None	493	20.58%
GSTM1	*0/*B	318	13.27%
GSTM1	*A/*B	96	4.01%
GSTM1	*A/*A	76	3.17%
GSTM1	*3/*3	20	0.83%
GSTM1	*1/*2	8	0.33%
GSTM1	*A/*Ax2	1	0.04%
GSTP1	*A/*A	1087	45.37%
GSTP1	*A/*B	764	31.89%
GSTP1	*A/*C	257	10.73%
GSTP1	*B/*B	151	6.30%
GSTP1	*B/*C	114	4.76%
GSTP1	*C/*C	20	0.83%
GSTP1	None	2	0.08%
GSTP1	*B/*D	1	0.04%
GSTT1	*0/*A	1141	47.62%
GSTT1	*A/*A	801	33.43%
GSTT1	*0/*0	415	17.32%
GSTT1	None	39	1.63%
IFNL3	*1/*1	1130	47.16%
IFNL3	*1/*rs12980275	1010	42.15%
IFNL3	*rs12980275/*rs12980275	251	10.48%
IFNL3	None	5	0.21%
NAT1	*4/*4	1266	52.84%
NAT1	*4/*10	605	25.25%
NAT1	*4/*11	107	4.47%
NAT1	*10/*10	95	3.96%
NAT1	*3/*4	80	3.34%
NAT1	*4/*14	66	2.75%
NAT1	*10/*11	30	1.25%
NAT1	None	25	1.04%
NAT1	*4/*15	22	0.92%
NAT1	*4/*22	19	0.79%
NAT1	*10/*14	15	0.63%
NAT1	*4/*17	15	0.63%
NAT1	*3/*10	13	0.54%
NAT1	*10/*15	6	0.25%
NAT1	*11/*14	6	0.25%
NAT1	*10/*17	4	0.17%
NAT1	*3/*11	4	0.17%
NAT1	*3/*15	3	0.13%
NAT1	*3/*3	3	0.13%
NAT1	*10/*22	2	0.08%
NAT1	*10/*27	2	0.08%

NAT1	*11/*11	2	0.08%
NAT1	*14/*14	2	0.08%
NAT1	*11/*17	1	0.04%
NAT1	*14/*15	1	0.04%
NAT1	*3/*14	1	0.04%
NAT1	*4/*27	1	0.04%
NAT2	*5/*6	557	23.25%
NAT2	*4/*5	493	20.58%
NAT2	*5/*5	474	19.78%
NAT2	*4/*6	357	14.90%
NAT2	*6/*6	210	8.76%
NAT2	*4/*4	155	6.47%
NAT2	*5/*7	40	1.67%
NAT2	*6/*7	36	1.50%
NAT2	*4/*7	21	0.88%
NAT2	*5/*12	12	0.50%
NAT2	None	12	0.50%
NAT2	*13/*13	5	0.21%
NAT2	*6/*12	5	0.21%
NAT2	*4/*12	3	0.13%
NAT2	*4/*13	3	0.13%
NAT2	*5/*13	2	0.08%
NAT2	*7/*7	2	0.08%
NAT2	*11/*11	1	0.04%
NAT2	*12/*12	1	0.04%
NAT2	*4/*14	1	0.04%
NAT2	*4/*19	1	0.04%
NAT2	*5/*11	1	0.04%
NAT2	*5/*14	1	0.04%
NAT2	*6/*13	1	0.04%
NAT2	*6/*14	1	0.04%
NAT2	*7/*12	1	0.04%
NUDT15	*1/*1	2338	97.58%
NUDT15	*1/*3	31	1.29%
NUDT15	*1/*6	15	0.63%
NUDT15	*1/*9	7	0.29%
NUDT15	*1/*2	2	0.08%
NUDT15	*1/*10	1	0.04%
NUDT15	*1/*4	1	0.04%
NUDT15	*1/*5	1	0.04%
POR	*1/*1	1217	50.79%
POR	*1/*28	967	40.36%
POR	*28/*28	189	7.89%
POR	None	9	0.38%
POR	*1/*29	5	0.21%
POR	*1/*45	2	0.08%
POR	*1/*5	2	0.08%
POR	*28/*45	2	0.08%
POR	*1/*11	1	0.04%
POR	*27/*28	1	0.04%
POR	*28/*46	1	0.04%
PTGIS	*1/*1	2396	100.00%
RYR1	*1/*1	2393	99.87%
RYR1	*1/*S15	1	0.04%
RYR1	*1/*S17	1	0.04%
RYR1	*1/*S29	1	0.04%
SLC15A2	*1/*2	1169	48.79%
SLC15A2	*1/*1	753	31.43%
SLC15A2	*2/*2	463	19.32%
SLC15A2	None	11	0.46%
SLC22A2	*1/*2	753	31.43%
SLC22A2	*2/*3	404	16.86%
SLC22A2	*2/*2	389	16.24%
SLC22A2	*1/*1	384	16.03%
SLC22A2	*1/*3	318	13.27%
SLC22A2	*3/*3	100	4.17%
SLC22A2	None	32	1.34%
SLC22A2	*1/*6	4	0.17%
SLC22A2	*2/*6	2	0.08%
SLC22A2	*2/*S1	2	0.08%
SLC22A2	*3/*6	2	0.08%
SLC22A2	*1/*4	1	0.04%
SLC22A2	*2/*4	1	0.04%
SLC22A2	*2/*DEL	1	0.04%
SLC22A2	*3/*4	1	0.04%
SLC22A2	*3/*S1	1	0.04%
SLC22A2	*6/*6	1	0.04%
SLCO1B1	*1/*1	758	31.64%
SLCO1B1	*1/*14	382	15.94%
SLCO1B1	*1/*15	374	15.61%
SLCO1B1	*1/*37	178	7.43%
SLCO1B1	*1/*20	127	5.30%
SLCO1B1	*14/*15	115	4.80%
SLCO1B1	*1/*5	81	3.38%
SLCO1B1	*14/*14	56	2.34%
SLCO1B1	*15/*15	54	2.25%
SLCO1B1	*14/*37	46	1.92%
SLCO1B1	*15/*37	44	1.84%

SLCO1B1	*14/*20	29	1.21%
SLCO1B1	*15/*20	28	1.17%
SLCO1B1	*5/*15	28	1.17%
SLCO1B1	*5/*14	22	0.92%
SLCO1B1	*37/*37	18	0.75%
SLCO1B1	None	14	0.58%
SLCO1B1	*20/*37	13	0.54%
SLCO1B1	*1/*46	5	0.21%
SLCO1B1	*5/*37	5	0.21%
SLCO1B1	*5/*5	4	0.17%
SLCO1B1	*20/*20	3	0.13%
SLCO1B1	*37/*46	2	0.08%
SLCO1B1	*1/*19	1	0.04%
SLCO1B1	*1/*26	1	0.04%
SLCO1B1	*1/*31	1	0.04%
SLCO1B1	*1/*45	1	0.04%
SLCO1B1	*14/*27	1	0.04%
SLCO1B1	*14/*46	1	0.04%
SLCO1B1	*15/*19	1	0.04%
SLCO1B1	*20/*46	1	0.04%
SLCO1B1	*31/*37	1	0.04%
SLCO1B1	*4/*14	1	0.04%
SLCO1B3	*rs7311358/*rs7311358	1655	69.07%
SLCO1B3	*Reference/*rs7311358	647	27.00%
SLCO1B3	*1/*1	78	3.26%
SLCO1B3	None	16	0.67%
SLCO2B1	*1/*1	2188	91.32%
SLCO2B1	*1/*S464F	111	4.63%
SLCO2B1	*1/*S1	88	3.67%
SLCO2B1	None	4	0.17%
SLCO2B1	*S464F/*S464F	3	0.13%
SLCO2B1	*S1/*S1	1	0.04%
SLCO2B1	*S1/*S464F	1	0.04%
SULT1A1	*1/*2	705	29.42%
SULT1A1	*1/*1	594	24.79%
SULT1A1	*1/*1x2	343	14.32%
SULT1A1	*2/*2	270	11.27%
SULT1A1	*1x2/*2	248	10.35%
SULT1A1	*1/*1x3	70	2.92%
SULT1A1	None	60	2.50%
SULT1A1	*1x3/*2	31	1.29%
SULT1A1	*1/*2x2	26	1.09%
SULT1A1	*2/*2x2	19	0.79%
SULT1A1	*3/*3	6	0.25%
SULT1A1	*1/*1x4	5	0.21%
SULT1A1	*1x2/*1x3	5	0.21%
SULT1A1	*1x3/*2x2	3	0.13%
SULT1A1	*1x4/*2	3	0.13%
SULT1A1	*2/*2x3	3	0.13%
SULT1A1	*1x2/*1x4	2	0.08%
SULT1A1	*1/*2x4	1	0.04%
SULT1A1	*1x2/*2x3	1	0.04%
SULT1A1	*1x4/*3	1	0.04%
TBXAS1	*1/*1	2092	87.31%
TBXAS1	*1/*8	115	4.80%
TBXAS1	*1/*7	93	3.88%
TBXAS1	*1/*3	44	1.84%
TBXAS1	*1/*2	11	0.46%
TBXAS1	*3/*3	10	0.42%
TBXAS1	*1/*9	7	0.29%
TBXAS1	None	7	0.29%
TBXAS1	*1/*5	5	0.21%
TBXAS1	*7/*8	4	0.17%
TBXAS1	*8/*8	3	0.13%
TBXAS1	*1/*4	2	0.08%
TBXAS1	*3/*7	1	0.04%
TBXAS1	*3/*9	1	0.04%
TBXAS1	*7/*7	1	0.04%
TPMT	*1/*1	2198	91.74%
TPMT	*1/*3	172	7.18%
TPMT	*1/*2	9	0.38%
TPMT	*3/*3	6	0.25%
TPMT	*1/*9	5	0.21%
TPMT	*1/*12	4	0.17%
TPMT	*1/*8	1	0.04%
TPMT	*2/*3	1	0.04%
UGT1A1	*1/*1	1075	44.87%
UGT1A1	*1/*80+*28	990	41.32%
UGT1A1	None	295	12.31%
UGT1A1	*1/*6	14	0.58%
UGT1A1	*1/*36	7	0.29%
UGT1A1	*1/*80+*37	4	0.17%
UGT1A1	*1/*28	2	0.08%
UGT1A1	*1/*60	2	0.08%
UGT1A1	*6/*6	2	0.08%
UGT1A1	*6/*80+*28	2	0.08%
UGT1A1	*36/*80+*28	1	0.04%
UGT1A1	*36/*80+*37	1	0.04%

UGT1A1	*80+*28/*80+*28	1	0.04%
UGT1A4	*1/*1	1722	71.87%
UGT1A4	*1/*3	406	16.94%
UGT1A4	*1/*2	197	8.22%
UGT1A4	*3/*3	22	0.92%
UGT1A4	*2/*3	19	0.79%
UGT1A4	None	12	0.50%
UGT1A4	*2/*2	10	0.42%
UGT1A4	*1/*4	4	0.17%
UGT1A4	*1/*S2	2	0.08%
UGT1A4	*2/*S2	1	0.04%
UGT1A4	*3/*4	1	0.04%
UGT2B15	*4/*5	436	18.20%
UGT2B15	*2/*4	334	13.94%
UGT2B15	*2/*5	310	12.94%
UGT2B15	*4/*4	277	11.56%
UGT2B15	*1/*4	200	8.35%
UGT2B15	*5/*5	185	7.72%
UGT2B15	*2/*2	164	6.84%
UGT2B15	*1/*2	160	6.68%
UGT2B15	*1/*5	135	5.63%
UGT2B15	None	108	4.51%
UGT2B15	*1/*1	41	1.71%
UGT2B15	*2x2/*5	8	0.33%
UGT2B15	*2/*S1	5	0.21%
UGT2B15	*2x2/*4	5	0.21%
UGT2B15	*4/*S1	5	0.21%
UGT2B15	*1/*S1	3	0.13%
UGT2B15	*2/*2x2	2	0.08%
UGT2B15	*2/*4x2	2	0.08%
UGT2B15	*4x2/*5	2	0.08%
UGT2B15	*5/*6	2	0.08%
UGT2B15	*5/*S1	2	0.08%
UGT2B15	*1/*2x2	1	0.04%
UGT2B15	*1/*4x2	1	0.04%
UGT2B15	*1/*5x2	1	0.04%
UGT2B15	*1x2/*5	1	0.04%
UGT2B15	*2/*5x2	1	0.04%
UGT2B15	*2/*6	1	0.04%
UGT2B15	*2/*DEL	1	0.04%
UGT2B15	*4/*DEL	1	0.04%
UGT2B15	*5/*5x2	1	0.04%
UGT2B15	*5/*DEL	1	0.04%
UGT2B17	*1/*1	1031	43.03%
UGT2B17	*1/*2	1006	41.99%
UGT2B17	*2/*2	301	12.56%
UGT2B17	None	58	2.42%
UGT2B7	*1/*2	1141	47.62%
UGT2B7	*2/*2	693	28.92%
UGT2B7	*1/*1	548	22.87%
UGT2B7	*1/*3	8	0.33%
UGT2B7	*2/*3	4	0.17%
UGT2B7	*2/*4	1	0.04%
UGT2B7	*3/*3	1	0.04%
VKORC1	Reference/rs9923231	1157	48.29%
VKORC1	Reference/Reference	843	35.18%
VKORC1	rs9923231/rs9923231	396	16.53%
XPC	Reference/rs2228001	1126	46.99%
XPC	rs2228001/rs2228001	845	35.27%
XPC	Reference/Reference	364	15.19%
XPC	rs2228000/rs2228000	1	0.04%
XPC	rs2228000/rs2228001	1	0.04%

Table 6.4: Translated phenotypes for applicable pharmacogenes from PGx pipeline for MASTER cohort. Phenotypes are not available for all alleles (Indeterminate).

Gene	Consensus Phenotype	n	Frequency
ABCG2	Normal Function	1955	81.59%
ABCG2	Decreased Function	416	17.36%
ABCG2	Poor Function	25	1.04%
CACNA1S	Uncertain Susceptibility	2396	100.00%
CFTR	Indeterminate	2367	98.79%
CFTR	Favorable Response	29	1.21%
COMT	Indeterminate	2391	99.79%
CYP2B6	Normal Metabolizer	1184	49.42%
CYP2B6	Intermediate Metabolizer	853	35.60%
CYP2B6	Poor Metabolizer	153	6.39%
CYP2B6	Rapid Metabolizer	112	4.67%
CYP2B6	Indeterminate	93	3.88%
CYP2B6	Ultrarapid Metabolizer	1	0.04%
CYP2C19	Normal Metabolizer	974	40.65%
CYP2C19	Rapid Metabolizer	625	26.09%

CYP2C19	Intermediate Metabolizer	621	25.92%
CYP2C19	Ultrarapid Metabolizer	121	5.05%
CYP2C19	Poor Metabolizer	54	2.25%
CYP2C19	Indeterminate	1	0.04%
CYP2C9	Normal Metabolizer	1553	64.82%
CYP2C9	Intermediate Metabolizer	783	32.68%
CYP2C9	Poor Metabolizer	56	2.34%
CYP2C9	Indeterminate	4	0.17%
CYP2D6	Normal Metabolizer	1246	52.00%
CYP2D6	Intermediate Metabolizer	861	35.93%
CYP2D6	Poor Metabolizer	154	6.43%
CYP2D6	Indeterminate	69	2.88%
CYP2D6	Ultrarapid Metabolizer	66	2.75%
CYP3A5	Poor Metabolizer	1987	82.93%
CYP3A5	Intermediate Metabolizer	311	12.98%
CYP3A5	Indeterminate	83	3.46%
CYP3A5	Normal Metabolizer	15	0.63%
DPYD	Normal Metabolizer	2246	93.74%
DPYD	Intermediate Metabolizer	149	6.22%
DPYD	Poor Metabolizer	1	0.04%
F5	Favorable Response	2238	93.41%
F5	Unfavorable Response	158	6.59%
IFNL3	Indeterminate	1268	52.92%
IFNL3	Favorable Response	1128	47.08%
NUDT15	Normal Metabolizer	2338	97.58%
NUDT15	Intermediate Metabolizer	40	1.67%
NUDT15	Indeterminate	18	0.75%
RYR1	Uncertain Susceptibility	2394	99.92%
RYR1	Malignant Hyperthermia Susceptibility	2	0.08%
SLCO1B1	Normal Function	1517	63.31%
SLCO1B1	Decreased Function	687	28.67%
SLCO1B1	Increased Function	93	3.88%
SLCO1B1	Poor Function	88	3.67%
SLCO1B1	Indeterminate	9	0.38%
SLCO1B1	Possible Decreased Function	2	0.08%
TPMT	Normal Metabolizer	2198	91.74%
TPMT	Intermediate Metabolizer	181	7.55%
TPMT	Indeterminate	10	0.42%
TPMT	Poor Metabolizer	7	0.29%
UGT1A1	Normal Metabolizer	1063	44.37%
UGT1A1	Intermediate Metabolizer	1016	42.40%
UGT1A1	Poor Metabolizer	293	12.23%
UGT1A1	Indeterminate	24	1.00%

Table 6.5: Predicted damaging non-synonymous SNVs in germline. In silico prediction was done with the APF framework [91].

Chr	Pos	Ref	Alt	rsID	Gene	Region	Type	FunctionalPrediction
1	47264908	A	G	rs772338414	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47276502	A	G	rs753724766	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47276532	C	T	rs56059446	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47278174	G	A	rs148753850	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47278243	A	G	rs144157811	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47279612	G	A	rs139750942	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47279636	C	T	rs144659997	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47279696	C	T	rs200200785	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47279697	G	A	rs372884535	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47279898	C	T	rs45446505	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47280765	T	G	rs746996053	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47280785	G	A	rs144531409	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47280852	A	C	rs12094024	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47280875	C	T	.	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47282755	G	C	rs59694031	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47282816	G	C	.	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47283850	T	A	.	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47283878	G	A	rs141281141	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47395917	C	T	rs148423796	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47395918	G	A	rs150500700	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47395969	T	C	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47398439	C	T	rs771932669	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47398493	C	T	rs199678286	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47399915	G	A	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47399944	G	A	rs141672858	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47399986	G	A	rs755248704	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47400714	G	A	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47400803	G	T	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47401231	C	T	rs143503396	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47402373	T	C	rs143639289	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47402453	C	G	rs144085677	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47406943	G	A	rs375391044	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47603326	C	T	rs148805480	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47603338	G	A	rs112604161	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47606567	A	T	rs61507155	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47607808	G	C	rs752724599	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47607812	T	C	rs371221965	CYP4A22	exonic	nonsynonymous SNV	damaging

1	47607825	G	A	rs138940178	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47610029	A	C	rs778465891	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47610314	C	G	rs369645508	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47610574	C	T	rs61736431	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47610627	C	T	rs371439568	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47611756	A	T	.	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47611765	G	A	rs150794228	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47614422	C	T	rs148057835	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47614425	C	T	rs138009089	CYP4A22	exonic	nonsynonymous SNV	damaging
1	60370563	C	A	rs748001282	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60370667	C	T	rs142713068	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60373495	C	G	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60373523	G	T	rs767380029	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60377365	C	T	rs115453547	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60377366	G	A	rs201070738	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60377940	C	A	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	97658683	G	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97700472	G	A	rs547099198	DPYD	exonic	nonsynonymous SNV	damaging
1	97700520	G	T	rs374825099	DPYD	exonic	nonsynonymous SNV	damaging
1	97700521	C	A	rs672601276	DPYD	exonic	nonsynonymous SNV	damaging
1	97771825	C	T	rs778298325	DPYD	exonic	nonsynonymous SNV	damaging
1	97771841	C	A	rs202212118	DPYD	exonic	nonsynonymous SNV	damaging
1	97839117	C	G	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98015214	C	A	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98015252	T	C	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98039375	A	G	rs200693895	DPYD	exonic	nonsynonymous SNV	damaging
1	98157329	G	C	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98164964	C	A	rs376073289	DPYD	exonic	nonsynonymous SNV	damaging
1	98187101	T	C	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98205979	A	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98293716	T	C	rs367619008	DPYD	exonic	nonsynonymous SNV	damaging
1	98386442	C	T	rs769820114	DPYD	exonic	nonsynonymous SNV	damaging
1	110231863	G	C	rs572826828	GSTM1	exonic	nonsynonymous SNV	damaging
1	110231874	G	T	rs199816990	GSTM1	exonic	nonsynonymous SNV	damaging
1	169483582	C	T	.	F5	exonic	nonsynonymous SNV	damaging
1	169489822	G	C	.	F5	exonic	nonsynonymous SNV	damaging
1	169495167	C	T	rs774639785	F5	exonic	nonsynonymous SNV	damaging
1	169495169	T	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169495234	A	G	rs377129476	F5	exonic	nonsynonymous SNV	damaging
1	169497292	C	T	rs6026	F5	exonic	nonsynonymous SNV	damaging
1	169497306	G	A	rs141977229	F5	exonic	nonsynonymous SNV	damaging
1	169499000	T	C	rs41272455	F5	exonic	nonsynonymous SNV	damaging
1	169499020	G	C	rs6034	F5	exonic	nonsynonymous SNV	damaging
1	169500173	C	T	rs201556325	F5	exonic	nonsynonymous SNV	damaging
1	169509629	T	A	.	F5	exonic	nonsynonymous SNV	damaging
1	169509650	A	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169509698	C	A	rs139288793	F5	exonic	nonsynonymous SNV	damaging
1	169510849	T	A	.	F5	exonic	nonsynonymous SNV	damaging
1	169511464	C	A	rs199507543	F5	exonic	nonsynonymous SNV	damaging
1	169511830	T	C	.	F5	exonic	nonsynonymous SNV	damaging
1	169512106	T	C	rs144979314	F5	exonic	nonsynonymous SNV	damaging
1	169513573	T	C	.	F5	exonic	nonsynonymous SNV	damaging
1	169513743	A	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169519112	C	T	rs6020	F5	exonic	nonsynonymous SNV	damaging
1	169519934	G	A	rs368387623	F5	exonic	nonsynonymous SNV	damaging
1	169524438	A	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169524505	G	A	rs746260106	F5	exonic	nonsynonymous SNV	damaging
1	169524537	C	G	rs118203906	F5	exonic	nonsynonymous SNV	damaging
1	169524573	G	T	.	F5	exonic	nonsynonymous SNV	damaging
1	169525926	C	T	rs747353298	F5	exonic	nonsynonymous SNV	damaging
1	169528493	G	T	rs144937515	F5	exonic	nonsynonymous SNV	damaging
1	169541513	C	A	.	F5	exonic	nonsynonymous SNV	damaging
1	169541561	A	G	rs367901835	F5	exonic	nonsynonymous SNV	damaging
1	201009011	C	T	rs72749169	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201009182	A	G	rs12139527	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201012622	A	G	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201013535	G	A	rs183195890	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201013548	C	T	rs138768414	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201013574	C	T	rs372436488	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201016671	C	T	rs775885648	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201016695	T	C	rs373248127	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201019588	G	C	rs371849585	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201019610	T	C	rs748210869	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201020150	A	G	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201021695	G	C	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201021733	C	T	rs200042281	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201021734	G	A	rs780390034	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201022344	C	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201022387	C	T	rs138144724	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201022621	C	T	rs530655602	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201022657	C	T	rs750637537	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201023671	C	T	rs148870919	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201027599	C	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201028331	C	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201030563	C	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201031099	G	A	rs200224590	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201031217	C	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201031636	A	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging

1	201034983	C	T	rs569324688	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201034997	A	G	rs575247457	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201035025	C	T	rs373701906	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201035034	C	T	rs146903750	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201035070	C	G	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201035428	C	T	rs146823170	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201036117	G	A	rs200334886	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201038312	A	C	rs752178238	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201038650	C	T	rs139956524	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201039487	G	A	rs759887262	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201044667	A	T	rs144590408	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201046058	C	T	rs142356235	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201046128	C	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201046205	C	T	rs4915212	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201047034	C	T	rs748711395	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201047133	C	A	rs150590855	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201047133	C	T	rs150590855	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201047187	A	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201052335	C	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201052382	A	G	rs146136274	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201052398	C	T	rs750807406	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201054623	C	T	rs763360081	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058501	C	T	rs776311349	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058513	C	T	rs35534614	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058529	G	A	rs555596737	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058543	G	A	rs200665694	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058579	G	A	rs767790285	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201060837	T	C	rs566565917	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201060844	C	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201061111	G	A	rs141204958	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201063146	T	C	rs140330831	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201079344	G	C	rs12406479	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201079372	T	C	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201079384	T	G	rs373778743	CACNA1S	exonic	nonsynonymous SNV	damaging
2	38298080	C	T	rs138388190	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38298394	C	T	rs79204362	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38301585	T	A	rs749521942	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38301756	C	T	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38301847	C	T	rs57865060	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38301879	T	A	rs72549383	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38301919	T	C	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38301924	T	C	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38302291	A	T	rs9282671	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38302297	G	A	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38302332	C	G	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38302377	G	A	rs201824781	CYP1B1	exonic	nonsynonymous SNV	damaging
2	234627616	G	C	rs45510694	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234627634	C	A	rs144275831	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234627647	C	T	rs199607987	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234627827	T	C	.	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234627932	C	T	.	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234627939	G	C	rs149433426	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234628073	G	A	.	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234628179	T	C	.	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234628292	A	C	rs147342917	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234628319	A	C	.	UGT1A4	exonic	nonsynonymous SNV	damaging
2	234669100	A	C	rs140365717	UGT1A1	exonic	nonsynonymous SNV	damaging
2	234669569	C	A	.	UGT1A1	exonic	nonsynonymous SNV	damaging
3	14187609	A	T	rs776266193	XPC	exonic	nonsynonymous SNV	damaging
3	14189464	C	T	rs775486844	XPC	exonic	nonsynonymous SNV	damaging
3	14190078	C	T	rs200148127	XPC	exonic	nonsynonymous SNV	damaging
3	14193906	G	A	.	XPC	exonic	nonsynonymous SNV	damaging
3	14197964	G	A	.	XPC	exonic	nonsynonymous SNV	damaging
3	14199593	C	T	rs763740883	XPC	exonic	nonsynonymous SNV	damaging
3	14199594	G	A	rs753379728	XPC	exonic	nonsynonymous SNV	damaging
3	14199642	C	T	.	XPC	exonic	nonsynonymous SNV	damaging
3	14199940	C	A	rs182616621	XPC	exonic	nonsynonymous SNV	damaging
3	14200115	C	T	rs376808339	XPC	exonic	nonsynonymous SNV	damaging
3	14201260	A	G	.	XPC	exonic	nonsynonymous SNV	damaging
3	14206331	A	C	.	XPC	exonic	nonsynonymous SNV	damaging
3	14206341	G	C	rs184879571	XPC	exonic	nonsynonymous SNV	damaging
3	14206351	C	G	rs778281904	XPC	exonic	nonsynonymous SNV	damaging
3	14206353	A	C	rs35629274	XPC	exonic	nonsynonymous SNV	damaging
3	121615340	G	T	rs151029304	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121616330	C	T	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121616364	T	C	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121630432	A	G	rs747718082	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121631884	G	A	rs759936043	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121634092	T	C	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121634123	C	A	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121641638	G	A	rs778421141	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121641965	T	G	rs761781029	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121643217	C	T	rs765554353	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121643850	G	A	rs778014741	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121643880	C	T	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121658304	A	G	rs370025673	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121658320	G	A	.	SLC15A2	exonic	nonsynonymous SNV	damaging
4	69403599	A	G	rs138121512	UGT2B17	exonic	nonsynonymous SNV	damaging

4	69416515	G	T	rs377204498	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69417580	G	A	rs148430260	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69426346	A	G	rs748669369	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69431385	G	C	.	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69433479	C	G	rs145791375	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69433505	C	G	rs148958723	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69433763	T	A	rs143522336	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69512917	C	T	rs72551390	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69512937	T	A	rs199547744	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69513006	C	T	rs147866157	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69513007	G	A	rs147164238	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69519875	G	A	rs138762595	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69533786	T	G	rs147882612	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69533852	C	T	rs371697004	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69533886	C	T	rs758424244	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69535705	C	A	.	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69535823	T	G	rs200638397	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69535835	G	T	rs747378153	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69536234	G	T	rs529876617	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69536261	C	T	.	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69962375	T	C	rs61361928	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69962737	C	T	rs747704916	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69962821	C	A	.	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69964336	C	T	rs758222821	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69964341	C	T	.	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69968588	A	G	rs767539882	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973863	A	T	.	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973866	G	A	rs771987274	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973877	G	A	rs771444554	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973890	G	A	rs753133394	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973920	C	A	rs563256432	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973974	T	A	rs144232904	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69978184	G	T	rs145217059	UGT2B7	exonic	nonsynonymous SNV	damaging
4	89016685	C	T	rs748169857	ABCG2	exonic	nonsynonymous SNV	damaging
4	89016695	T	G	rs200894058	ABCG2	exonic	nonsynonymous SNV	damaging
4	89016707	G	A	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89016744	C	A	rs759323853	ABCG2	exonic	nonsynonymous SNV	damaging
4	89018670	C	T	rs45605536	ABCG2	exonic	nonsynonymous SNV	damaging
4	89020542	A	G	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89020572	C	T	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89022416	C	T	rs765641486	ABCG2	exonic	nonsynonymous SNV	damaging
4	89022448	G	A	rs769734146	ABCG2	exonic	nonsynonymous SNV	damaging
4	89039261	C	T	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89039300	C	T	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89042860	T	G	rs12721643	ABCG2	exonic	nonsynonymous SNV	damaging
4	89042886	C	G	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89052340	G	A	rs770985871	ABCG2	exonic	nonsynonymous SNV	damaging
4	89053780	T	C	rs148475733	ABCG2	exonic	nonsynonymous SNV	damaging
4	89060966	T	C	rs769486810	ABCG2	exonic	nonsynonymous SNV	damaging
6	18132390	G	A	rs755899157	TPMT	exonic	nonsynonymous SNV	damaging
6	18139230	C	A	.	TPMT	exonic	nonsynonymous SNV	damaging
6	160645750	C	T	rs372435157	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160645766	C	G	rs144511904	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160662580	A	G	rs754967456	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160663405	A	G	rs762784077	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160664672	C	T	rs567153149	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160664679	C	T	rs535926721	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160664726	A	G	rs779624954	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160664733	C	A	rs150866933	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160664793	C	A	rs759789804	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160668223	G	A	rs868711069	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160668329	A	C	rs144729356	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160670336	G	C	rs767713938	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160671633	C	T	rs372563664	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160677647	T	C	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160677662	C	T	rs370177229	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679387	T	G	rs748283994	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679473	G	A	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679570	T	G	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679600	C	A	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679772	G	C	rs139039970	SLC22A2	exonic	nonsynonymous SNV	damaging
7	1022945	C	T	rs544975130	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1024180	G	A	rs200427519	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1024861	C	G	rs117826462	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1024867	G	A	rs750767254	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1026828	C	T	.	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1027109	C	T	.	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1027114	G	A	rs549964637	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1027150	G	A	rs201612311	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1027922	G	A	rs752381534	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1028294	C	T	rs200715910	CYP2W1	exonic	nonsynonymous SNV	damaging
7	1028337	T	G	rs201684384	CYP2W1	exonic	nonsynonymous SNV	damaging
7	75608776	A	T	rs782037392	POR	exonic	nonsynonymous SNV	damaging
7	75608819	G	C	rs782226844	POR	exonic	nonsynonymous SNV	damaging
7	75608844	G	A	rs375997962	POR	exonic	nonsynonymous SNV	damaging
7	75610404	T	A	.	POR	exonic	nonsynonymous SNV	damaging
7	75610418	A	G	rs782681272	POR	exonic	nonsynonymous SNV	damaging
7	75610912	G	A	rs782746344	POR	exonic	nonsynonymous SNV	damaging

7	75611612	C	T	rs200471958	POR	exonic	nonsynonymous SNV	damaging
7	75612919	G	T	.	POR	exonic	nonsynonymous SNV	damaging
7	75612936	T	G	.	POR	exonic	nonsynonymous SNV	damaging
7	75613075	G	A	rs540924885	POR	exonic	nonsynonymous SNV	damaging
7	75613084	T	C	rs782419727	POR	exonic	nonsynonymous SNV	damaging
7	75613108	G	A	rs562241770	POR	exonic	nonsynonymous SNV	damaging
7	75614128	C	G	.	POR	exonic	nonsynonymous SNV	damaging
7	75614482	C	T	rs72557935	POR	exonic	nonsynonymous SNV	damaging
7	75614485	G	A	rs375535318	POR	exonic	nonsynonymous SNV	damaging
7	75614511	G	A	rs72557936	POR	exonic	nonsynonymous SNV	damaging
7	75614928	T	G	rs781929293	POR	exonic	nonsynonymous SNV	damaging
7	75614954	G	A	rs373347327	POR	exonic	nonsynonymous SNV	damaging
7	75615084	C	T	rs782248163	POR	exonic	nonsynonymous SNV	damaging
7	75615273	G	A	rs779082897	POR	exonic	nonsynonymous SNV	damaging
7	75615367	C	T	.	POR	exonic	nonsynonymous SNV	damaging
7	75615678	A	T	.	POR	exonic	nonsynonymous SNV	damaging
7	75615698	G	A	rs372930296	POR	exonic	nonsynonymous SNV	damaging
7	87133704	C	T	rs201578293	ABCB1	exonic	nonsynonymous SNV	damaging
7	87133729	G	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87135236	C	T	rs202030954	ABCB1	exonic	nonsynonymous SNV	damaging
7	87135302	G	A	rs199676098	ABCB1	exonic	nonsynonymous SNV	damaging
7	87138710	C	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87138736	A	G	rs199931681	ABCB1	exonic	nonsynonymous SNV	damaging
7	87138760	T	G	rs55852620	ABCB1	exonic	nonsynonymous SNV	damaging
7	87150164	C	T	rs774299788	ABCB1	exonic	nonsynonymous SNV	damaging
7	87160644	A	G	rs375296280	ABCB1	exonic	nonsynonymous SNV	damaging
7	87174210	T	G	rs145840638	ABCB1	exonic	nonsynonymous SNV	damaging
7	87175195	T	C	rs141018820	ABCB1	exonic	nonsynonymous SNV	damaging
7	87175288	C	T	rs56107566	ABCB1	exonic	nonsynonymous SNV	damaging
7	87175289	G	A	rs28381914	ABCB1	exonic	nonsynonymous SNV	damaging
7	87175318	A	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87175330	C	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87178749	C	T	rs763454753	ABCB1	exonic	nonsynonymous SNV	damaging
7	87178750	G	A	rs199852575	ABCB1	exonic	nonsynonymous SNV	damaging
7	87179256	G	A	rs142600685	ABCB1	exonic	nonsynonymous SNV	damaging
7	87179329	C	G	rs201178758	ABCB1	exonic	nonsynonymous SNV	damaging
7	87179815	T	A	rs144933300	ABCB1	exonic	nonsynonymous SNV	damaging
7	87180049	T	C	rs199766539	ABCB1	exonic	nonsynonymous SNV	damaging
7	87180094	C	T	rs200966236	ABCB1	exonic	nonsynonymous SNV	damaging
7	87183138	G	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87196129	C	T	rs61122623	ABCB1	exonic	nonsynonymous SNV	damaging
7	87196272	C	T	rs201352004	ABCB1	exonic	nonsynonymous SNV	damaging
7	87214993	G	A	rs761584848	ABCB1	exonic	nonsynonymous SNV	damaging
7	99250194	T	C	.	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99250288	C	G	.	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99250365	T	C	rs149888520	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99261707	A	G	rs142004817	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99262851	A	C	rs147489136	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99264600	G	T	rs539204136	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99270222	G	T	rs41279857	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99270262	G	T	.	CYP3A5	exonic	nonsynonymous SNV	damaging
7	99361591	C	A	rs71581996	CYP3A4	exonic	nonsynonymous SNV	damaging
7	99367810	C	A	rs752473076	CYP3A4	exonic	nonsynonymous SNV	damaging
7	99425774	G	T	.	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99441825	T	G	.	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99445188	G	T	rs140041607	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99445207	A	G	rs149853346	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99453321	C	T	rs749902724	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99454459	C	G	rs746142784	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99454482	G	A	rs45621431	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99459342	A	G	rs139401065	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99459381	G	A	.	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99461160	G	T	rs143991326	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99461219	G	A	rs145743239	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99461225	G	A	rs142155405	CYP3A43	exonic	nonsynonymous SNV	damaging
7	117120162	C	T	rs193922501	CFTR	exonic	nonsynonymous SNV	damaging
7	117120186	C	G	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117144344	C	T	rs1800073	CFTR	exonic	nonsynonymous SNV	damaging
7	117144378	C	T	rs143456784	CFTR	exonic	nonsynonymous SNV	damaging
7	117144402	C	A	rs397508220	CFTR	exonic	nonsynonymous SNV	damaging
7	117149147	G	A	rs1800076	CFTR	exonic	nonsynonymous SNV	damaging
7	117171037	G	A	rs201958172	CFTR	exonic	nonsynonymous SNV	damaging
7	117171038	C	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117171120	C	G	rs759310470	CFTR	exonic	nonsynonymous SNV	damaging
7	117171122	T	C	rs35516286	CFTR	exonic	nonsynonymous SNV	damaging
7	117171155	T	C	rs397508727	CFTR	exonic	nonsynonymous SNV	damaging
7	117174348	C	T	rs578029902	CFTR	exonic	nonsynonymous SNV	damaging
7	117174349	G	A	rs1800079	CFTR	exonic	nonsynonymous SNV	damaging
7	117174401	C	A	rs397508754	CFTR	exonic	nonsynonymous SNV	damaging
7	117174403	T	C	rs766640075	CFTR	exonic	nonsynonymous SNV	damaging
7	117175323	G	A	rs138338446	CFTR	exonic	nonsynonymous SNV	damaging
7	117175372	A	G	rs121909046	CFTR	exonic	nonsynonymous SNV	damaging
7	117176630	A	G	rs191456345	CFTR	exonic	nonsynonymous SNV	damaging
7	117180173	C	T	rs397508814	CFTR	exonic	nonsynonymous SNV	damaging
7	117180174	G	A	rs143486492	CFTR	exonic	nonsynonymous SNV	damaging
7	117180186	A	G	rs150691494	CFTR	exonic	nonsynonymous SNV	damaging
7	117180198	T	G	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117180217	C	G	rs121909016	CFTR	exonic	nonsynonymous SNV	damaging

7	117180285	G	A	rs397508137	CFTR	exonic	nonsynonymous SNV	damaging
7	117180297	C	T	rs77409459	CFTR	exonic	nonsynonymous SNV	damaging
7	117180336	C	G	rs1800086	CFTR	exonic	nonsynonymous SNV	damaging
7	117180363	C	T	rs75053309	CFTR	exonic	nonsynonymous SNV	damaging
7	117182078	A	C	rs73215912	CFTR	exonic	nonsynonymous SNV	damaging
7	117182116	C	T	rs143860237	CFTR	exonic	nonsynonymous SNV	damaging
7	117188795	G	A	rs765791986	CFTR	exonic	nonsynonymous SNV	damaging
7	117199641	A	G	rs1800091	CFTR	exonic	nonsynonymous SNV	damaging
7	117199648	T	G	rs74571530	CFTR	exonic	nonsynonymous SNV	damaging
7	117199683	G	A	rs77646904	CFTR	exonic	nonsynonymous SNV	damaging
7	117199705	A	G	rs374453187	CFTR	exonic	nonsynonymous SNV	damaging
7	117227874	A	G	rs75789129	CFTR	exonic	nonsynonymous SNV	damaging
7	117230408	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117230411	G	A	rs1800097	CFTR	exonic	nonsynonymous SNV	damaging
7	117232223	C	T	rs1800100	CFTR	exonic	nonsynonymous SNV	damaging
7	117232329	A	T	rs748664864	CFTR	exonic	nonsynonymous SNV	damaging
7	117232470	C	T	rs140455771	CFTR	exonic	nonsynonymous SNV	damaging
7	117232631	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117232638	A	G	rs397508375	CFTR	exonic	nonsynonymous SNV	damaging
7	117232642	A	G	rs1800103	CFTR	exonic	nonsynonymous SNV	damaging
7	117232662	A	C	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117234999	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117243741	T	G	rs193922511	CFTR	exonic	nonsynonymous SNV	damaging
7	117243828	T	C	rs1800110	CFTR	exonic	nonsynonymous SNV	damaging
7	117246736	C	T	rs747139295	CFTR	exonic	nonsynonymous SNV	damaging
7	117250575	G	C	rs1800111	CFTR	exonic	nonsynonymous SNV	damaging
7	117250609	G	A	rs184724618	CFTR	exonic	nonsynonymous SNV	damaging
7	117250625	A	G	rs149279509	CFTR	exonic	nonsynonymous SNV	damaging
7	117251817	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117254688	G	C	rs397508550	CFTR	exonic	nonsynonymous SNV	damaging
7	117254714	A	G	rs397508556	CFTR	exonic	nonsynonymous SNV	damaging
7	117254743	C	A	rs397508565	CFTR	exonic	nonsynonymous SNV	damaging
7	117267592	G	T	rs1800120	CFTR	exonic	nonsynonymous SNV	damaging
7	117267610	A	G	rs150326506	CFTR	exonic	nonsynonymous SNV	damaging
7	117267720	C	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117267812	T	G	rs34911792	CFTR	exonic	nonsynonymous SNV	damaging
7	117292899	G	A	rs769931559	CFTR	exonic	nonsynonymous SNV	damaging
7	117292931	C	G	rs80034486	CFTR	exonic	nonsynonymous SNV	damaging
7	117304834	G	T	rs113857788	CFTR	exonic	nonsynonymous SNV	damaging
7	117305584	G	A	rs780396890	CFTR	exonic	nonsynonymous SNV	damaging
7	117306992	G	A	rs867990936	CFTR	exonic	nonsynonymous SNV	damaging
7	117307154	A	G	.	CFTR	exonic	nonsynonymous SNV	damaging
7	139529239	C	G	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139572060	T	C	rs757418800	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139572102	T	C	rs771838691	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139611050	T	C	rs184269562	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139611089	T	A	rs370871916	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139636023	A	G	rs774271298	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139636038	C	T	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139655298	G	A	rs769131779	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139655341	C	G	rs137946697	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139655362	G	A	rs748986878	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139657466	G	A	rs568328354	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139657540	C	T	rs140774405	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139706981	A	G	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139715558	T	C	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139715564	A	G	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139715597	C	T	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139715599	G	A	rs759019222	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139715603	G	A	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139717481	G	A	rs756633372	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139717485	C	A	rs753974178	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139717523	G	A	rs149988492	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139717523	G	T	rs149988492	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139717626	A	T	rs200663004	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139719829	C	T	rs13306050	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139719877	A	G	rs780924927	TBXAS1	exonic	nonsynonymous SNV	damaging
8	18079675	T	C	.	NAT1	exonic	nonsynonymous SNV	damaging
8	18080005	T	C	.	NAT1	exonic	nonsynonymous SNV	damaging
8	18080128	C	T	rs141552883	NAT1	exonic	nonsynonymous SNV	damaging
8	18257522	T	G	.	NAT2	exonic	nonsynonymous SNV	damaging
8	18257529	T	C	.	NAT2	exonic	nonsynonymous SNV	damaging
8	18257647	G	T	.	NAT2	exonic	nonsynonymous SNV	damaging
8	18258046	C	A	.	NAT2	exonic	nonsynonymous SNV	damaging
8	18258099	C	A	rs771698130	NAT2	exonic	nonsynonymous SNV	damaging
8	18258202	G	T	rs749948990	NAT2	exonic	nonsynonymous SNV	damaging
10	94834543	T	C	rs762272547	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94834578	G	A	rs140213678	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94834596	G	A	rs757177182	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94834621	G	T	rs150571738	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94834668	C	T	rs368680474	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94834937	G	T	rs763644436	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94835632	C	T	rs142962735	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94836750	A	C	rs200904706	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94837005	G	A	rs750770595	CYP26A1	exonic	nonsynonymous SNV	damaging
10	96522522	G	C	rs769429735	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96522551	C	T	rs750669985	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96534903	C	G	.	CYP2C19	exonic	nonsynonymous SNV	damaging

10	96535189	G	A	rs141774245	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96535224	G	C	rs374950950	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96535225	G	A	rs766813172	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96535296	G	C	rs181297724	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96540292	C	T	rs61311738	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96541719	G	A	rs577255883	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96580394	G	T	rs778852965	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96602666	T	A	rs201132803	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96602680	G	A	rs201509150	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96612490	G	A	rs770375701	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96701616	T	A	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96708905	G	A	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96740966	G	A	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96741073	C	A	rs139532088	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96748616	G	C	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96748751	C	T	rs530950257	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96796999	T	A	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96798663	A	G	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96798666	G	T	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96798758	T	G	rs201301235	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96798795	C	T	rs143386810	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96818263	G	C	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96824609	A	C	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96827024	A	G	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96827075	C	T	rs369591911	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96827409	C	G	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96829125	G	T	rs756655248	CYP2C8	exonic	nonsynonymous SNV	damaging
10	104590528	G	C	rs147557447	CYP17A1	exonic	nonsynonymous SNV	damaging
10	104590596	C	T	rs763457719	CYP17A1	exonic	nonsynonymous SNV	damaging
10	135341012	C	T	.	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135341012	C	G	rs76271067	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135345131	G	A	rs773718909	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135345657	G	A	rs60452492	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135350645	A	G	rs753186824	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135350716	A	C	.	CYP2E1	exonic	nonsynonymous SNV	damaging
11	14899807	A	G	rs545401539	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14899823	C	T	rs782741911	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14900710	T	C	rs201004240	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14900766	G	T	.	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14900824	A	T	rs782535484	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14900891	C	G	rs781954755	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14901708	A	G	rs145224817	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14901831	A	G	rs200183599	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14902021	C	T	rs143448859	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14902131	G	A	rs781970760	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14902179	G	C	.	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14907326	C	T	.	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14907417	C	T	rs782115359	CYP2R1	exonic	nonsynonymous SNV	damaging
11	14913535	A	G	.	CYP2R1	exonic	nonsynonymous SNV	damaging
11	67351964	G	T	rs192307201	GSTP1	exonic	nonsynonymous SNV	damaging
11	67351981	C	A	rs752215721	GSTP1	exonic	nonsynonymous SNV	damaging
11	67351992	A	T	rs45506591	GSTP1	exonic	nonsynonymous SNV	damaging
11	67352636	C	T	.	GSTP1	exonic	nonsynonymous SNV	damaging
11	67352698	A	T	rs199833944	GSTP1	exonic	nonsynonymous SNV	damaging
11	67353861	T	C	.	GSTP1	exonic	nonsynonymous SNV	damaging
11	67353950	G	C	rs753365034	GSTP1	exonic	nonsynonymous SNV	damaging
11	74875027	G	A	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74876878	G	A	rs142693902	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74876889	C	T	rs148248368	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74880370	G	A	rs35199625	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74883531	T	C	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74904401	C	T	rs747713084	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74904589	C	T	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74904608	C	T	rs764144223	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74911293	G	A	rs143480565	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74914358	T	A	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
12	21325624	C	T	rs576786579	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21325651	C	T	rs769900186	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21331570	G	A	rs142101690	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21331921	A	G	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21331930	G	A	rs147421160	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21331940	G	A	rs374113543	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21355611	C	A	rs141779296	SLCO1B1	exonic	nonsynonymous SNV	damaging
15	51503075	G	A	rs763421146	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51503226	C	A	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51504546	C	T	rs199887515	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51504587	A	G	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51507273	C	G	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51507335	A	G	rs143839949	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51510760	A	T	rs143562020	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51514627	C	T	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51514699	G	A	rs201842322	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51520054	T	C	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51520054	T	A	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	75012837	C	A	rs56343424	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75012838	G	A	rs148638069	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75013005	C	T	rs180744198	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75013098	G	A	rs769227467	CYP1A1	exonic	nonsynonymous SNV	damaging

15	75013344	C	T	rs750520977	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75013576	C	T	rs201174966	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75013932	C	G	.	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014027	A	G	rs4987133	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014049	G	A	rs34260157	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014718	G	A	rs149687459	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014727	G	A	rs61747605	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014847	T	C	.	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014861	G	A	.	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014884	G	T	rs140459785	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75014957	T	G	rs202221673	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015036	G	A	rs45442501	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015045	C	G	rs35196245	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015063	C	T	rs375219443	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015138	C	T	rs141173079	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015147	G	A	rs754416936	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015206	A	G	rs17861094	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015243	T	C	rs35035798	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015293	A	G	rs565370983	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75042122	C	T	rs60086777	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042389	G	A	rs34067076	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042488	C	T	rs758124536	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042767	G	C	rs765435682	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042776	G	A	rs201537008	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75043539	C	T	rs45468096	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75044485	C	T	rs148157092	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75044519	C	T	.	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75045593	G	C	.	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75045602	A	G	.	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75047346	G	A	rs763531887	CYP1A2	exonic	nonsynonymous SNV	damaging
16	28617243	C	T	.	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28617437	T	C	rs140288278	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28617472	G	A	rs150459557	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28618170	T	C	rs759716692	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28618278	C	T	rs141878102	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28619647	G	C	rs144188544	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28619676	C	T	rs760293838	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28619838	C	G	rs375616347	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28619841	G	A	rs201320226	SULT1A1	exonic	nonsynonymous SNV	damaging
16	28620112	T	C	rs143603811	SULT1A1	exonic	nonsynonymous SNV	damaging
16	31102655	G	A	rs72547528	VKORC1	exonic	nonsynonymous SNV	damaging
16	31105894	G	T	rs781304132	VKORC1	exonic	nonsynonymous SNV	damaging
16	31105945	C	A	rs61742245	VKORC1	exonic	nonsynonymous SNV	damaging
19	15989688	G	C	rs762784709	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15990222	C	T	rs138971789	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15990223	G	A	rs142113670	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15990703	C	T	rs772862283	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15996756	G	A	rs200373927	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15996828	G	C	rs145174239	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15996851	G	A	rs757642625	CYP4F2	exonic	nonsynonymous SNV	damaging
19	16001215	C	A	rs3093153	CYP4F2	exonic	nonsynonymous SNV	damaging
19	16003189	G	A	rs200629062	CYP4F2	exonic	nonsynonymous SNV	damaging
19	38931405	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38933004	C	G	rs769890047	RYR1	exonic	nonsynonymous SNV	damaging
19	38934252	C	T	rs118192173	RYR1	exonic	nonsynonymous SNV	damaging
19	38935253	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38937121	C	T	rs727504129	RYR1	exonic	nonsynonymous SNV	damaging
19	38937148	A	G	rs766836202	RYR1	exonic	nonsynonymous SNV	damaging
19	38937375	G	A	rs772767943	RYR1	exonic	nonsynonymous SNV	damaging
19	38942430	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38942437	G	A	rs532700459	RYR1	exonic	nonsynonymous SNV	damaging
19	38945887	A	G	rs147723844	RYR1	exonic	nonsynonymous SNV	damaging
19	38945972	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38946109	A	G	rs780036569	RYR1	exonic	nonsynonymous SNV	damaging
19	38946141	T	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38946315	A	C	rs779551357	RYR1	exonic	nonsynonymous SNV	damaging
19	38948261	A	G	rs747459771	RYR1	exonic	nonsynonymous SNV	damaging
19	38948702	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38948720	C	T	rs757908433	RYR1	exonic	nonsynonymous SNV	damaging
19	38948803	C	T	rs772751128	RYR1	exonic	nonsynonymous SNV	damaging
19	38948887	G	A	rs138874610	RYR1	exonic	nonsynonymous SNV	damaging
19	38949893	A	G	rs147320363	RYR1	exonic	nonsynonymous SNV	damaging
19	38949911	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38949938	G	A	rs147918857	RYR1	exonic	nonsynonymous SNV	damaging
19	38951044	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38951086	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38951142	C	T	rs142548565	RYR1	exonic	nonsynonymous SNV	damaging
19	38954066	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38954130	C	T	rs143701391	RYR1	exonic	nonsynonymous SNV	damaging
19	38954139	G	A	rs370634440	RYR1	exonic	nonsynonymous SNV	damaging
19	38954391	G	A	rs147515913	RYR1	exonic	nonsynonymous SNV	damaging
19	38954490	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38955362	C	T	rs201827275	RYR1	exonic	nonsynonymous SNV	damaging
19	38956757	C	T	rs143179371	RYR1	exonic	nonsynonymous SNV	damaging
19	38956780	G	A	rs748676912	RYR1	exonic	nonsynonymous SNV	damaging
19	38956955	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38958270	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38958369	C	T	rs750429900	RYR1	exonic	nonsynonymous SNV	damaging

19	38960055	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964369	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964370	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964371	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964372	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964376	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964377	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38964382	T	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38965975	A	G	rs137933390	RYR1	exonic	nonsynonymous SNV	damaging
19	38965983	G	A	rs772111760	RYR1	exonic	nonsynonymous SNV	damaging
19	38968461	C	T	rs200546266	RYR1	exonic	nonsynonymous SNV	damaging
19	38969139	C	T	rs752736122	RYR1	exonic	nonsynonymous SNV	damaging
19	38974074	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38974155	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976316	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976331	G	A	rs146504767	RYR1	exonic	nonsynonymous SNV	damaging
19	38976610	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976655	C	T	rs34934920	RYR1	exonic	nonsynonymous SNV	damaging
19	38980786	C	T	rs751621150	RYR1	exonic	nonsynonymous SNV	damaging
19	38983186	C	T	rs758631725	RYR1	exonic	nonsynonymous SNV	damaging
19	38985040	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38985070	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38985135	C	T	rs750275456	RYR1	exonic	nonsynonymous SNV	damaging
19	38985195	G	C	rs143398211	RYR1	exonic	nonsynonymous SNV	damaging
19	38987056	G	A	rs537994744	RYR1	exonic	nonsynonymous SNV	damaging
19	38987105	C	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38987133	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38987509	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38989817	A	G	rs34390345	RYR1	exonic	nonsynonymous SNV	damaging
19	38990346	G	A	rs146306934	RYR1	exonic	nonsynonymous SNV	damaging
19	38990409	G	A	rs777021627	RYR1	exonic	nonsynonymous SNV	damaging
19	38990446	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38990455	G	A	rs749136416	RYR1	exonic	nonsynonymous SNV	damaging
19	38990562	C	T	rs145787667	RYR1	exonic	nonsynonymous SNV	damaging
19	38990594	G	T	rs193922808	RYR1	exonic	nonsynonymous SNV	damaging
19	38990601	T	A	rs118192174	RYR1	exonic	nonsynonymous SNV	damaging
19	38991271	C	T	rs141959437	RYR1	exonic	nonsynonymous SNV	damaging
19	38991355	C	A	rs141298868	RYR1	exonic	nonsynonymous SNV	damaging
19	38991605	T	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38993310	G	A	rs751180702	RYR1	exonic	nonsynonymous SNV	damaging
19	38993537	T	C	rs769326916	RYR1	exonic	nonsynonymous SNV	damaging
19	38993543	A	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38993572	G	C	rs193922825	RYR1	exonic	nonsynonymous SNV	damaging
19	38995499	G	A	rs545688934	RYR1	exonic	nonsynonymous SNV	damaging
19	38995509	A	G	rs112196644	RYR1	exonic	nonsynonymous SNV	damaging
19	38995965	C	T	rs147707463	RYR1	exonic	nonsynonymous SNV	damaging
19	38995998	C	G	rs35180584	RYR1	exonic	nonsynonymous SNV	damaging
19	38998362	G	A	rs79294840	RYR1	exonic	nonsynonymous SNV	damaging
19	39001369	G	A	rs759007399	RYR1	exonic	nonsynonymous SNV	damaging
19	39002726	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39002913	G	A	rs145044872	RYR1	exonic	nonsynonymous SNV	damaging
19	39006751	C	G	rs587784379	RYR1	exonic	nonsynonymous SNV	damaging
19	39006759	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39006807	A	G	rs199738299	RYR1	exonic	nonsynonymous SNV	damaging
19	39008026	A	G	rs200950673	RYR1	exonic	nonsynonymous SNV	damaging
19	39008109	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39008109	A	C	rs201588259	RYR1	exonic	nonsynonymous SNV	damaging
19	39008161	G	A	rs757753317	RYR1	exonic	nonsynonymous SNV	damaging
19	39008173	G	A	rs756487708	RYR1	exonic	nonsynonymous SNV	damaging
19	39008276	A	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39009877	C	T	rs118204421	RYR1	exonic	nonsynonymous SNV	damaging
19	39009932	G	A	rs137932199	RYR1	exonic	nonsynonymous SNV	damaging
19	39009974	G	A	rs754760055	RYR1	exonic	nonsynonymous SNV	damaging
19	39010076	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39013682	C	T	rs150977342	RYR1	exonic	nonsynonymous SNV	damaging
19	39016006	A	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39016020	C	T	rs375127981	RYR1	exonic	nonsynonymous SNV	damaging
19	39016059	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39016132	G	A	rs143987857	RYR1	exonic	nonsynonymous SNV	damaging
19	39017670	A	T	rs199689862	RYR1	exonic	nonsynonymous SNV	damaging
19	39018366	C	G	rs748082431	RYR1	exonic	nonsynonymous SNV	damaging
19	39019010	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39019242	C	G	rs114351116	RYR1	exonic	nonsynonymous SNV	damaging
19	39026677	G	A	rs145087576	RYR1	exonic	nonsynonymous SNV	damaging
19	39027398	C	T	rs138593495	RYR1	exonic	nonsynonymous SNV	damaging
19	39028586	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39028598	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39034191	A	G	rs147136339	RYR1	exonic	nonsynonymous SNV	damaging
19	39034446	C	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39038927	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39039016	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39051820	A	T	rs765132716	RYR1	exonic	nonsynonymous SNV	damaging
19	39051876	C	A	rs193922849	RYR1	exonic	nonsynonymous SNV	damaging
19	39051876	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39052023	G	A	rs151119428	RYR1	exonic	nonsynonymous SNV	damaging
19	39052029	C	A	rs761486041	RYR1	exonic	nonsynonymous SNV	damaging
19	39055897	T	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39055930	G	A	rs539194350	RYR1	exonic	nonsynonymous SNV	damaging

19	39056409	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39057618	A	G	rs139647387	RYR1	exonic	nonsynonymous SNV	damaging
19	39057626	G	C	rs150396398	RYR1	exonic	nonsynonymous SNV	damaging
19	39058426	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39058480	C	T	rs763661413	RYR1	exonic	nonsynonymous SNV	damaging
19	39062797	G	A	rs752668333	RYR1	exonic	nonsynonymous SNV	damaging
19	39062801	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39066601	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39068588	C	T	rs766887342	RYR1	exonic	nonsynonymous SNV	damaging
19	39068670	C	T	rs146793368	RYR1	exonic	nonsynonymous SNV	damaging
19	39070731	G	A	rs193922875	RYR1	exonic	nonsynonymous SNV	damaging
19	39076790	C	G	rs368874586	RYR1	exonic	nonsynonymous SNV	damaging
19	39078031	C	T	rs374070555	RYR1	exonic	nonsynonymous SNV	damaging
19	39078052	C	T	rs140584202	RYR1	exonic	nonsynonymous SNV	damaging
19	39734325	G	A	rs150748693	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734352	C	T	rs62120527	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734490	G	A	rs139076671	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734507	C	T	.	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734531	G	A	rs779256274	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734721	T	G	.	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734754	G	A	rs145428712	IFNL3	exonic	nonsynonymous SNV	damaging
19	39734773	C	G	rs147679979	IFNL3	exonic	nonsynonymous SNV	damaging
19	41349763	G	A	rs561053481	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351288	C	T	rs757428419	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351288	C	G	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351867	C	T	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351924	C	G	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351935	A	T	rs148693084	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41352791	G	A	rs780129128	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41354126	G	C	rs771986786	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41354593	A	G	rs777098658	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41354651	C	G	rs61562160	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41355777	C	T	rs145308399	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41355802	C	A	rs368359507	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41355839	C	T	rs752300065	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41497314	G	A	rs781365650	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41509919	A	G	.	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41509934	C	T	rs138264188	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41509987	C	T	.	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41510301	G	A	rs139173201	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41512872	G	A	rs58871670	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41512901	T	A	rs767612288	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41512910	C	A	.	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41515959	C	T	.	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41515998	C	T	rs373559488	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41518204	G	C	rs187378204	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41522557	G	A	rs764288403	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41594486	C	G	rs781487437	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41594522	T	C	rs551458619	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41594916	A	G	rs138627841	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41596044	G	A	rs201720562	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41596059	G	C	rs764008365	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41596322	T	G	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41599593	A	G	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600182	G	A	rs200636194	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600191	C	T	rs138941528	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600212	C	T	rs149632806	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600240	T	C	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600254	C	A	rs116368403	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600317	C	T	rs147797134	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600925	C	T	rs202218822	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600976	A	T	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41601694	G	A	rs772441450	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41601710	T	C	rs201481142	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41601727	A	G	rs762462392	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41601785	C	T	rs781684223	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41627401	G	C	rs2287941	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41627402	T	G	rs2287942	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41627408	A	C	rs140643766	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41627432	G	A	.	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41628001	G	A	rs200744662	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41630629	G	C	rs182353952	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41630642	T	A	rs372824170	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41630665	C	T	rs376080668	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41630767	C	T	.	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41630783	C	T	rs138507242	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41633809	G	A	rs139951793	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41633841	A	C	rs146029724	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41633853	C	G	rs139597756	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41699287	C	T	.	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41699344	C	T	.	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41700468	C	T	rs62119652	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41704373	C	T	.	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41707230	C	T	rs145747863	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41711887	G	T	rs148468532	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41711977	C	T	.	CYP2S1	exonic	nonsynonymous SNV	damaging
20	48124465	G	A	rs372248049	PTGIS	exonic	nonsynonymous SNV	damaging
20	48124477	G	A	rs774641163	PTGIS	exonic	nonsynonymous SNV	damaging

20	48124489	C	T	rs373851981	PTGIS	exonic	nonsynonymous SNV	damaging
20	48127583	G	A	rs747383414	PTGIS	exonic	nonsynonymous SNV	damaging
20	48129759	G	C	rs767323052	PTGIS	exonic	nonsynonymous SNV	damaging
20	48130913	G	T	.	PTGIS	exonic	nonsynonymous SNV	damaging
20	48140626	C	T	rs61734270	PTGIS	exonic	nonsynonymous SNV	damaging
20	48156124	C	T	rs759208880	PTGIS	exonic	nonsynonymous SNV	damaging
20	48156149	C	T	rs143394422	PTGIS	exonic	nonsynonymous SNV	damaging
20	48156236	C	T	rs148768155	PTGIS	exonic	nonsynonymous SNV	damaging
20	48184620	C	T	.	PTGIS	exonic	nonsynonymous SNV	damaging
22	42522605	G	C	.	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42522724	G	T	rs79392742	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42522990	G	A	rs757030056	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42523483	C	T	rs776076897	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42523484	G	A	rs141739595	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42523636	C	A	rs3915951	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42524295	G	A	rs146819268	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42525797	C	T	rs746115614	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42525847	T	G	rs200229206	CYP2D6	exonic	nonsynonymous SNV	damaging
X	153760404	T	A	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762288	C	G	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762577	A	G	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762607	T	A	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762618	C	G	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762628	A	C	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762640	T	G	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153762641	G	A	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153774294	A	T	.	G6PD	exonic	nonsynonymous SNV	damaging

Table 6.6: Predicted damaging non-synonymous SNVs in tumor. In silico prediction was done with the APF framework [91].

Chr	Pos	Ref	Alt	rsID	Gene	Region	Type	FunctionalPrediction
1	47283819	G	A	rs773003953	CYP4B1	exonic	nonsynonymous SNV	damaging
1	47395831	G	T	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47398439	C	T	rs771932669	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47399965	T	C	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47402458	C	T	rs62621075	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47403738	C	G	.	CYP4A11	exonic	nonsynonymous SNV	damaging
1	47606542	C	T	rs780190981	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47607785	G	A	rs2056900	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47609034	T	C	.	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47610115	G	A	.	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47611765	G	A	rs150794228	CYP4A22	exonic	nonsynonymous SNV	damaging
1	47614284	G	A	rs554623281	CYP4A22	exonic	nonsynonymous SNV	damaging
1	60366644	G	C	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60370634	T	C	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60373542	C	T	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60377426	G	T	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60377849	T	C	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	60381688	C	T	.	CYP2J2	exonic	nonsynonymous SNV	damaging
1	97544536	C	G	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97547984	C	T	rs776236081	DPYD	exonic	nonsynonymous SNV	damaging
1	97658670	G	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97658729	G	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97700466	C	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97700550	C	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97771741	G	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97847992	C	T	rs548783838	DPYD	exonic	nonsynonymous SNV	damaging
1	97915679	C	A	.	DPYD	exonic	nonsynonymous SNV	damaging
1	97981388	G	A	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98015180	T	G	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98039478	G	A	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98058866	C	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98144732	A	C	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98164985	C	T	.	DPYD	exonic	nonsynonymous SNV	damaging
1	98386466	G	A	rs772097379	DPYD	exonic	nonsynonymous SNV	damaging
1	110230503	T	C	.	GSTM1	exonic	nonsynonymous SNV	damaging
1	110230813	C	T	.	GSTM1	exonic	nonsynonymous SNV	damaging
1	169484765	A	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169484842	A	T	.	F5	exonic	nonsynonymous SNV	damaging
1	169487723	C	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169494143	C	T	.	F5	exonic	nonsynonymous SNV	damaging
1	169497162	G	A	.	F5	exonic	nonsynonymous SNV	damaging
1	169497240	C	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169500107	C	T	rs775890784	F5	exonic	nonsynonymous SNV	damaging
1	169505812	C	A	.	F5	exonic	nonsynonymous SNV	damaging
1	169511471	T	C	.	F5	exonic	nonsynonymous SNV	damaging
1	169512106	T	C	rs144979314	F5	exonic	nonsynonymous SNV	damaging
1	169512188	C	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169519112	C	T	rs6020	F5	exonic	nonsynonymous SNV	damaging
1	169519977	T	C	.	F5	exonic	nonsynonymous SNV	damaging
1	169528433	G	T	.	F5	exonic	nonsynonymous SNV	damaging
1	169528463	C	T	.	F5	exonic	nonsynonymous SNV	damaging
1	169541549	C	G	.	F5	exonic	nonsynonymous SNV	damaging
1	169541566	G	T	.	F5	exonic	nonsynonymous SNV	damaging

1	201016295	C	T	rs774256022	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201016734	C	T	rs533353353	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201021748	C	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201022618	G	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201022621	C	T	rs530655602	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201023649	C	G	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201027563	G	C	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201029910	G	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201030485	G	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201031162	C	T	rs747618077	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201036049	C	T	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201036093	T	G	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201036117	G	C	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201046158	A	G	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201047050	C	T	rs557195329	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201047107	C	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058513	C	T	rs35534614	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201058532	G	A	rs151005797	CACNA1S	exonic	nonsynonymous SNV	damaging
1	201060852	G	A	.	CACNA1S	exonic	nonsynonymous SNV	damaging
2	38297923	G	A	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38298152	C	A	rs149049138	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38298250	T	C	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38298332	T	A	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38298365	G	C	.	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38298394	C	T	rs79204362	CYP1B1	exonic	nonsynonymous SNV	damaging
2	38302075	G	A	rs368041729	CYP1B1	exonic	nonsynonymous SNV	damaging
2	234627720	C	T	.	UGT1A4	exonic	nonsynonymous SNV	damaging
3	14190078	C	T	rs200148127	XPC	exonic	nonsynonymous SNV	damaging
3	14206318	C	T	.	XPC	exonic	nonsynonymous SNV	damaging
3	14207048	C	T	rs778771038	XPC	exonic	nonsynonymous SNV	damaging
3	14220055	C	T	.	XPC	exonic	nonsynonymous SNV	damaging
3	121616366	G	A	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121641697	G	A	.	SLC15A2	exonic	nonsynonymous SNV	damaging
3	121659234	C	T	rs775409002	SLC15A2	exonic	nonsynonymous SNV	damaging
4	69416406	A	C	.	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69417601	G	A	.	UGT2B17	exonic	nonsynonymous SNV	damaging
4	69513067	G	A	.	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69513067	G	T	.	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69533825	C	T	rs367595613	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69535897	T	C	.	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69536234	G	A	rs529876617	UGT2B15	exonic	nonsynonymous SNV	damaging
4	69973886	C	A	.	UGT2B7	exonic	nonsynonymous SNV	damaging
4	69973895	C	A	.	UGT2B7	exonic	nonsynonymous SNV	damaging
4	89015796	C	T	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89016739	T	C	rs762421964	ABCG2	exonic	nonsynonymous SNV	damaging
4	89020507	A	C	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89039280	C	A	.	ABCG2	exonic	nonsynonymous SNV	damaging
4	89052323	G	T	rs2231142	ABCG2	exonic	nonsynonymous SNV	damaging
4	89053778	C	T	.	ABCG2	exonic	nonsynonymous SNV	damaging
6	160663365	A	T	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160664719	G	C	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160668320	C	G	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160670282	A	C	rs316019	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160677728	A	T	.	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679669	C	T	rs548362661	SLC22A2	exonic	nonsynonymous SNV	damaging
6	160679734	T	C	.	SLC22A2	exonic	nonsynonymous SNV	damaging
7	1027090	C	T	rs759809358	CYP2W1	exonic	nonsynonymous SNV	damaging
7	75608886	G	A	.	POR	exonic	nonsynonymous SNV	damaging
7	75611546	C	T	.	POR	exonic	nonsynonymous SNV	damaging
7	75613105	C	T	.	POR	exonic	nonsynonymous SNV	damaging
7	75613154	T	C	.	POR	exonic	nonsynonymous SNV	damaging
7	75614167	T	C	.	POR	exonic	nonsynonymous SNV	damaging
7	75614484	C	T	rs781999828	POR	exonic	nonsynonymous SNV	damaging
7	75614990	C	T	rs781994682	POR	exonic	nonsynonymous SNV	damaging
7	75615699	A	G	.	POR	exonic	nonsynonymous SNV	damaging
7	87133683	G	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87135257	G	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87138706	C	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87145862	G	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87145935	T	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87148705	C	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87160680	C	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87168621	C	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87168639	C	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87170695	G	A	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87175318	A	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87178735	G	T	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87179855	G	C	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87180124	T	G	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87183082	G	C	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87190630	G	C	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87214897	C	G	.	ABCB1	exonic	nonsynonymous SNV	damaging
7	87214992	C	T	rs199551851	ABCB1	exonic	nonsynonymous SNV	damaging
7	99361560	G	A	rs770129614	CYP3A4	exonic	nonsynonymous SNV	damaging
7	99381677	C	T	.	CYP3A4	exonic	nonsynonymous SNV	damaging
7	99434157	G	T	.	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99447233	C	G	.	CYP3A43	exonic	nonsynonymous SNV	damaging
7	99461219	G	A	rs145743239	CYP3A43	exonic	nonsynonymous SNV	damaging

7	117149147	G	A	rs1800076	CFTR	exonic	nonsynonymous SNV	damaging
7	117171037	G	A	rs201958172	CFTR	exonic	nonsynonymous SNV	damaging
7	117174373	G	C	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117175333	C	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117180180	C	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117180297	C	T	rs77409459	CFTR	exonic	nonsynonymous SNV	damaging
7	117180323	C	T	rs397508147	CFTR	exonic	nonsynonymous SNV	damaging
7	117180347	C	T	rs144720913	CFTR	exonic	nonsynonymous SNV	damaging
7	117182148	G	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117182155	G	T	rs397508174	CFTR	exonic	nonsynonymous SNV	damaging
7	117188801	C	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117199651	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117199683	G	A	rs77646904	CFTR	exonic	nonsynonymous SNV	damaging
7	117199688	C	G	rs140552874	CFTR	exonic	nonsynonymous SNV	damaging
7	117227878	C	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117232158	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117232223	C	T	rs1800100	CFTR	exonic	nonsynonymous SNV	damaging
7	117232347	G	A	rs397508342	CFTR	exonic	nonsynonymous SNV	damaging
7	117232427	C	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117232611	C	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117242881	T	C	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117246769	G	A	rs764644021	CFTR	exonic	nonsynonymous SNV	damaging
7	117251646	A	G	rs374403559	CFTR	exonic	nonsynonymous SNV	damaging
7	117282492	G	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117282508	G	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117304776	G	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117304781	C	A	.	CFTR	exonic	nonsynonymous SNV	damaging
7	117307043	A	T	.	CFTR	exonic	nonsynonymous SNV	damaging
7	139611094	A	T	.	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139653218	C	T	rs373695119	TBXAS1	exonic	nonsynonymous SNV	damaging
7	139717481	G	A	rs756633372	TBXAS1	exonic	nonsynonymous SNV	damaging
8	18079663	C	T	.	NAT1	exonic	nonsynonymous SNV	damaging
8	18079929	T	G	.	NAT1	exonic	nonsynonymous SNV	damaging
10	94834659	G	A	.	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94835050	C	T	.	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94836729	A	T	.	CYP26A1	exonic	nonsynonymous SNV	damaging
10	94836903	G	A	.	CYP26A1	exonic	nonsynonymous SNV	damaging
10	96534947	C	T	.	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96541686	C	T	.	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96580341	G	T	.	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96580341	G	A	.	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96602689	C	T	.	CYP2C19	exonic	nonsynonymous SNV	damaging
10	96698527	C	T	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96708970	G	A	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96731954	A	G	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96741123	C	T	rs767815842	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96745901	A	G	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96748604	G	A	.	CYP2C9	exonic	nonsynonymous SNV	damaging
10	96797060	C	T	rs748167187	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96798659	G	A	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96798774	G	C	rs74454169	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96805672	T	A	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96824696	A	T	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96827038	C	A	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	96829041	C	T	.	CYP2C8	exonic	nonsynonymous SNV	damaging
10	104591315	G	A	.	CYP17A1	exonic	nonsynonymous SNV	damaging
10	104592323	G	A	rs104894142	CYP17A1	exonic	nonsynonymous SNV	damaging
10	104594546	A	G	.	CYP17A1	exonic	nonsynonymous SNV	damaging
10	135342087	A	G	.	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135345664	C	T	rs548262477	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135350675	G	A	rs745528149	CYP2E1	exonic	nonsynonymous SNV	damaging
10	135352430	C	T	.	CYP2E1	exonic	nonsynonymous SNV	damaging
11	14900665	G	A	.	CYP2R1	exonic	nonsynonymous SNV	damaging
11	67353646	G	T	.	GSTP1	exonic	nonsynonymous SNV	damaging
11	74876847	T	G	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74876893	G	A	rs764735701	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74876908	G	A	rs763847000	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74880370	G	A	rs35199625	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74880729	C	G	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74880749	G	A	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74904463	G	A	rs757854988	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74907630	C	A	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74915588	C	T	rs144746239	SLCO2B1	exonic	nonsynonymous SNV	damaging
11	74915590	G	T	.	SLCO2B1	exonic	nonsynonymous SNV	damaging
12	21325589	C	G	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21327634	T	C	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21331882	G	T	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21349974	A	G	rs192911820	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21353467	C	G	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21358894	G	A	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21370053	G	A	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
12	21370068	A	T	.	SLCO1B1	exonic	nonsynonymous SNV	damaging
13	48615244	A	G	.	NUDT15	exonic	nonsynonymous SNV	damaging
15	51503190	C	T	rs201638381	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51504584	C	A	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51504711	C	T	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51507327	G	A	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51520051	C	T	rs745845217	CYP19A1	exonic	nonsynonymous SNV	damaging

15	51529063	T	A	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	51529141	C	G	.	CYP19A1	exonic	nonsynonymous SNV	damaging
15	75014889	A	G	.	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015204	C	T	.	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75015395	G	T	.	CYP1A1	exonic	nonsynonymous SNV	damaging
15	75042329	C	T	.	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042771	C	T	.	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042776	G	A	rs201537008	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75042820	C	A	rs542879792	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75045550	G	A	.	CYP1A2	exonic	nonsynonymous SNV	damaging
15	75047279	G	C	rs763859277	CYP1A2	exonic	nonsynonymous SNV	damaging
19	15989736	C	T	rs146148233	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15990211	C	T	rs764508037	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15990222	C	T	rs138971789	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15990451	C	A	.	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15996785	C	A	.	CYP4F2	exonic	nonsynonymous SNV	damaging
19	15996840	G	A	.	CYP4F2	exonic	nonsynonymous SNV	damaging
19	16001215	C	A	rs3093153	CYP4F2	exonic	nonsynonymous SNV	damaging
19	16006346	C	T	rs372270252	CYP4F2	exonic	nonsynonymous SNV	damaging
19	16006348	G	A	.	CYP4F2	exonic	nonsynonymous SNV	damaging
19	38931388	G	A	rs755878800	RYR1	exonic	nonsynonymous SNV	damaging
19	38931457	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38934430	G	A	rs142474192	RYR1	exonic	nonsynonymous SNV	damaging
19	38935236	G	A	rs766256366	RYR1	exonic	nonsynonymous SNV	damaging
19	38939322	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38946139	A	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38946372	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38948796	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38948884	G	A	rs376526576	RYR1	exonic	nonsynonymous SNV	damaging
19	38949807	C	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38949827	C	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38951043	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38951097	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38951143	G	A	rs777191617	RYR1	exonic	nonsynonymous SNV	damaging
19	38954096	G	A	rs267605463	RYR1	exonic	nonsynonymous SNV	damaging
19	38955331	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38956903	C	T	rs139006437	RYR1	exonic	nonsynonymous SNV	damaging
19	38956949	C	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38958302	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38958345	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38959663	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38959954	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38960007	G	A	rs760235443	RYR1	exonic	nonsynonymous SNV	damaging
19	38964255	G	A	rs750256869	RYR1	exonic	nonsynonymous SNV	damaging
19	38964278	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38966023	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38966089	C	T	rs191656849	RYR1	exonic	nonsynonymous SNV	damaging
19	38969105	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976313	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976331	G	A	rs146504767	RYR1	exonic	nonsynonymous SNV	damaging
19	38976415	G	A	rs371566475	RYR1	exonic	nonsynonymous SNV	damaging
19	38976498	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976585	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976649	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976654	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38976759	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38979883	A	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38985124	G	A	rs530885842	RYR1	exonic	nonsynonymous SNV	damaging
19	38986931	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38987107	G	A	rs773947484	RYR1	exonic	nonsynonymous SNV	damaging
19	38989854	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38990452	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38990457	G	A	rs111364296	RYR1	exonic	nonsynonymous SNV	damaging
19	38991285	G	A	rs765019465	RYR1	exonic	nonsynonymous SNV	damaging
19	38991494	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38991613	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38991616	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38993204	G	A	rs771523641	RYR1	exonic	nonsynonymous SNV	damaging
19	38993271	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38993295	G	A	rs753507343	RYR1	exonic	nonsynonymous SNV	damaging
19	38994881	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38997544	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	38998429	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39001201	C	G	rs374272827	RYR1	exonic	nonsynonymous SNV	damaging
19	39002229	C	T	rs761187396	RYR1	exonic	nonsynonymous SNV	damaging
19	39002998	C	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39003006	C	T	rs61739911	RYR1	exonic	nonsynonymous SNV	damaging
19	39006806	G	A	rs185371036	RYR1	exonic	nonsynonymous SNV	damaging
19	39008241	G	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39015975	G	A	rs371645169	RYR1	exonic	nonsynonymous SNV	damaging
19	39016131	C	T	rs759605800	RYR1	exonic	nonsynonymous SNV	damaging
19	39016132	G	A	rs143987857	RYR1	exonic	nonsynonymous SNV	damaging
19	39017669	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39018300	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39019292	C	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39019295	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39025961	G	T	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39034427	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging

19	39034489	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39034492	C	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39051943	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39052011	G	A	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39071088	T	C	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39075640	A	G	.	RYR1	exonic	nonsynonymous SNV	damaging
19	39075715	G	T	rs193922892	RYR1	exonic	nonsynonymous SNV	damaging
19	39075728	C	T	rs546322293	RYR1	exonic	nonsynonymous SNV	damaging
19	41349877	G	A	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351864	C	T	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351933	C	T	rs771319834	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351950	G	A	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41351981	C	T	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41356267	G	C	.	CYP2A6	exonic	nonsynonymous SNV	damaging
19	41497314	G	A	rs781365650	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41509972	G	A	.	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41510301	G	A	rs139173201	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41518385	C	T	.	CYP2B6	exonic	nonsynonymous SNV	damaging
19	41594853	C	T	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41594951	G	T	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41596420	T	C	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41597670	C	A	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41600230	A	G	.	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41601727	A	G	rs762462392	CYP2A13	exonic	nonsynonymous SNV	damaging
19	41622188	C	T	rs200700685	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41626317	A	T	.	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41627414	T	C	.	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41630764	C	T	.	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41631519	C	A	.	CYP2F1	exonic	nonsynonymous SNV	damaging
19	41709525	G	T	.	CYP2S1	exonic	nonsynonymous SNV	damaging
19	41711916	C	G	.	CYP2S1	exonic	nonsynonymous SNV	damaging
20	48127625	G	C	.	PTGIS	exonic	nonsynonymous SNV	damaging
20	48127647	G	A	rs377540375	PTGIS	exonic	nonsynonymous SNV	damaging
20	48156118	G	A	.	PTGIS	exonic	nonsynonymous SNV	damaging
20	48156125	G	T	.	PTGIS	exonic	nonsynonymous SNV	damaging
20	48164460	A	G	.	PTGIS	exonic	nonsynonymous SNV	damaging
20	48166656	C	T	.	PTGIS	exonic	nonsynonymous SNV	damaging
20	48166689	G	A	.	PTGIS	exonic	nonsynonymous SNV	damaging
22	42522927	C	T	rs574629217	CYP2D6	exonic	nonsynonymous SNV	damaging
22	42523636	C	A	rs3915951	CYP2D6	exonic	nonsynonymous SNV	damaging
X	153760303	C	T	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153760885	T	G	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153761819	G	A	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153763477	C	T	.	G6PD	exonic	nonsynonymous SNV	damaging
X	153775028	C	T	.	G6PD	exonic	nonsynonymous SNV	damaging

Table 6.7: Top 200 significant CpGs from correlation analysis of methylation and expression.

Gene	CpG	Region	Cancer Entity	Correlation	Direction	Adjusted P Value
GSTP1	cg11566244	intragenic	Neuroendocrine and Adrenal	-0.6863	neg	1.23E-37
GSTP1	cg09038676	intragenic	Neuroendocrine and Adrenal	-0.6647	neg	1.55E-34
PTGIS	cg08788055	intragenic	STS: other	-0.5364	neg	8.25E-33
GSTP1	cg06928838	intragenic	Neuroendocrine and Adrenal	-0.6454	neg	5.05E-32
GSTP1	cg07493922	intragenic	Melanoma	-0.8213	neg	2.10E-31
GSTP1	cg07493922	intragenic	Other	-0.6397	neg	3.45E-28
ABCB1	cg00862116	intragenic	Neuroendocrine and Adrenal	-0.5948	neg	8.87E-26
GSTP1	cg22224704	intragenic	Neuroendocrine and Adrenal	-0.5938	neg	1.02E-25
TPMT	cg16056511	intragenic	STS: other	-0.4834	neg	1.06E-25
PTGIS	cg06357305	intragenic	STS: other	-0.4701	neg	4.51E-24
DPYD	cg16511333	intragenic	STS: other	0.4699	pos	4.51E-24
GSTP1	cg26250609	intragenic	Neuroendocrine and Adrenal	-0.5766	neg	5.98E-24
DPYD	cg00080253	intragenic	STS: other	0.4679	pos	7.82E-24
SULT1A1	cg01332815	intragenic	Melanoma	-0.754	neg	1.99E-23
TBXAS1	cg12104698	intragenic	Other	-0.5916	neg	4.31E-23
POR	cg02138834	intragenic	Neuroendocrine and Adrenal	-0.5658	neg	6.63E-23
ABCB1	cg24138422	intragenic	Neuroendocrine and Adrenal	0.5628	pos	1.30E-22
POR	cg02727959	intragenic	Neuroendocrine and Adrenal	-0.5625	neg	1.32E-22
DPYD	cg25969802	intragenic	STS: other	0.4551	pos	1.70E-22
CYP2S1	cg05312704	intragenic	Other	0.5827	pos	2.62E-22
TBXAS1	cg12104698	intragenic	STS: other	-0.4486	neg	8.94E-22
POR	cg20220522	intragenic	Neuroendocrine and Adrenal	-0.5534	neg	9.34E-22
POR	cg02742533	intragenic	Neuroendocrine and Adrenal	-0.5527	neg	1.04E-21
PTGIS	cg06357305	intragenic	Leiomyosarcoma	-0.6676	neg	2.42E-21
ABCB1	cg07469128	intragenic	Neuroendocrine and Adrenal	0.545	pos	5.61E-21
PTGIS	cg08788055	intragenic	Leiomyosarcoma	-0.6624	neg	6.78E-21
POR	cg18630265	intragenic	Neuroendocrine and Adrenal	-0.543	neg	8.28E-21
ABCG2	cg23706819	intragenic	STS: other	-0.4345	neg	2.68E-20
DPYD	cg10890168	intragenic	STS: other	0.4341	pos	2.90E-20
CYP2J2	cg02089480	intragenic	STS: other	0.4331	pos	3.59E-20
ABCB1	cg26551025	intragenic	Neuroendocrine and Adrenal	0.5322	pos	7.83E-20
DPYD	cg17752576	intragenic	STS: other	0.4295	pos	8.37E-20
POR	cg16182457	intragenic	Neuroendocrine and Adrenal	-0.5311	neg	9.44E-20
DPYD	cg11226378	intragenic	STS: other	0.4263	pos	1.78E-19
ABCB1	cg05496710	intragenic	Neuroendocrine and Adrenal	0.5278	pos	1.81E-19
GSTP1	cg04920951	intragenic	Neuroendocrine and Adrenal	-0.5273	neg	1.97E-19
NAT1	cg22116708	intragenic	Other	-0.5493	neg	2.29E-19

PTGIS	cg09062977	intragenic	STS: other	0.419	pos	9.46E-19
CYP3A5	cg05867406	intragenic	Neuroendocrine and Adrenal	-0.518	neg	1.29E-18
ABCB1	cg22899422	intragenic	Neuroendocrine and Adrenal	0.5176	pos	1.37E-18
NAT1	cg18509990	intragenic	STS: other	0.4171	pos	1.39E-18
ABCB1	cg00001224	intragenic	Neuroendocrine and Adrenal	0.5172	pos	1.42E-18
GSTP1	cg22224704	intragenic	Melanoma	-0.6958	neg	1.44E-18
POR	cg21748691	intragenic	Neuroendocrine and Adrenal	-0.5167	neg	1.49E-18
CYP2E1	cg05417377	intragenic	STS: other	0.4146	pos	2.31E-18
DPYD	cg11055641	intragenic	Other	0.5339	pos	4.26E-18
GSTP1	cg07493922	intragenic	STS: other	-0.4118	neg	4.29E-18
CYP2C9	cg23202385	intragenic	Neuroendocrine and Adrenal	0.5107	pos	4.73E-18
GSTP1	cg22224704	intragenic	Hepatopancreaticobiliary	-0.6605	neg	7.39E-18
DPYD	cg08024495	intragenic	STS: other	0.4081	pos	9.73E-18
POR	cg15177211	intragenic	Neuroendocrine and Adrenal	-0.5065	neg	1.04E-17
POR	cg27342333	intragenic	Neuroendocrine and Adrenal	-0.506	neg	1.11E-17
DPYD	cg00080253	intragenic	Other	0.5271	pos	1.38E-17
GSTP1	cg07493922	intragenic	Hepatopancreaticobiliary	-0.6559	neg	1.53E-17
NAT1	cg22116708	intragenic	STS: other	-0.4055	neg	1.63E-17
CYP4B1	cg14222679	intragenic	STS: other	0.405	pos	1.79E-17
RYR1	cg01781084	intragenic	Neuroendocrine and Adrenal	0.5031	pos	1.82E-17
SULT1A1	cg14944435	intragenic	Melanoma	-0.6796	neg	1.95E-17
ABCB1	cg07479137	intragenic	Neuroendocrine and Adrenal	0.5022	pos	2.10E-17
RYR1	cg20880196	intragenic	Neuroendocrine and Adrenal	0.4982	pos	4.53E-17
NAT1	cg21172319	intragenic	Other	-0.5196	neg	5.02E-17
CYP2D6	cg12720083	intragenic	Neuroendocrine and Adrenal	-0.4975	neg	5.02E-17
CYP2W1	cg03322234	promoter	Colorectal	-0.6586	neg	8.15E-17
RYR1	cg08761223	intragenic	STS: other	0.3974	pos	9.00E-17
UGT1A1	cg23714547	intragenic	Neuroendocrine and Adrenal	-0.3601	neg	1.03E-16
UGT1A4	cg23714547	intragenic	Neuroendocrine and Adrenal	-0.3601	neg	1.03E-16
ABCB1	cg27507700	intragenic	Neuroendocrine and Adrenal	-0.4932	neg	1.06E-16
ABCB1	cg00373554	intragenic	Neuroendocrine and Adrenal	0.4931	pos	1.06E-16
DPYD	cg16320208	intragenic	Other	0.5146	pos	1.16E-16
GSTP1	cg07493922	intragenic	Neuroendocrine and Adrenal	-0.492	neg	1.27E-16
NAT1	cg21172319	intragenic	STS: other	-0.3946	neg	1.51E-16
F5	cg12674918	promoter	Neuroendocrine and Adrenal	-0.4898	neg	1.88E-16
CYP4B1	cg00503298	intragenic	STS: other	0.3933	pos	1.97E-16
ABCB1	cg24138422	intragenic	STS: other	0.3932	pos	1.99E-16
GSTP1	cg22224704	intragenic	STS: other	-0.393	neg	2.04E-16
DPYD	cg17140186	intragenic	STS: other	0.3906	pos	3.48E-16
F5	cg25256723	promoter	Neuroendocrine and Adrenal	-0.4843	neg	4.95E-16
SULT1A1	cg01332815	intragenic	Neuroendocrine and Adrenal	-0.483	neg	6.27E-16
CYP2S1	cg23532138	intragenic	STS: other	0.3871	pos	7.24E-16
SULT1A1	cg22042399	intragenic	Melanoma	-0.6558	neg	7.47E-16
GSTP1	cg07493922	intragenic	Unknown	-0.7441	neg	1.42E-15
DPYD	cg25633983	intragenic	STS: other	0.3894	pos	1.71E-15
CYP2C8	cg04164578	intragenic	STS: other	0.3828	pos	1.73E-15
PTGIS	cg08788055	intragenic	Melanoma	-0.6496	neg	1.88E-15
SULT1A1	cg14944435	intragenic	STS: other	-0.3801	neg	2.99E-15
POR	cg05153729	intragenic	Neuroendocrine and Adrenal	-0.4733	neg	3.30E-15
DPYD	cg00260775	intragenic	STS: other	0.3786	pos	4.02E-15
UGT2B7	cg10961486	intragenic	Neuroendocrine and Adrenal	0.4712	pos	4.70E-15
PTGIS	cg25962137	intragenic	STS: other	0.3776	pos	4.86E-15
SLCO2B1	cg18589858	intragenic	STS: other	-0.3773	neg	5.21E-15
POR	cg14500655	intragenic	Neuroendocrine and Adrenal	-0.469	neg	6.69E-15
GSTP1	cg22224704	intragenic	Other	-0.4902	neg	6.76E-15
TBXAS1	cg06239618	intragenic	Other	0.4896	pos	7.38E-15
POR	cg08262464	intragenic	Neuroendocrine and Adrenal	-0.4674	neg	8.58E-15
ABCB1	cg22403253	intragenic	Neuroendocrine and Adrenal	0.4672	pos	8.83E-15
ABCB1	cg06977014	intragenic	Neuroendocrine and Adrenal	0.4665	pos	9.94E-15
RYR1	cg06891424	intragenic	STS: other	0.3726	pos	1.29E-14
ABCB1	cg07611207	intragenic	Neuroendocrine and Adrenal	0.4646	pos	1.35E-14
CYP2W1	cg23332328	intragenic	Neuroendocrine and Adrenal	-0.4638	neg	1.54E-14
DPYD	cg25969802	intragenic	Other	0.4856	pos	1.55E-14
DPYD	cg16511333	intragenic	Other	0.4832	pos	2.02E-14
TBXAS1	cg12104698	intragenic	Melanoma	-0.6316	neg	2.34E-14
SLCO2B1	cg23577865	intragenic	Neuroendocrine and Adrenal	-0.4592	neg	3.27E-14
DPYD	cg25633983	intragenic	Colorectal	0.6214	pos	3.65E-14
UGT1A1	cg00764099	intragenic	Neuroendocrine and Adrenal	0.3328	pos	3.74E-14
UGT1A4	cg00764099	intragenic	Neuroendocrine and Adrenal	0.3328	pos	3.74E-14
SULT1A1	cg22375718	intragenic	STS: other	0.3665	pos	4.04E-14
CYP4B1	cg19092343	intragenic	STS: other	0.3665	pos	4.04E-14
RYR1	cg24830036	intragenic	Neuroendocrine and Adrenal	0.4576	pos	4.08E-14
DPYD	cg17752576	intragenic	Other	0.4777	pos	4.62E-14
CYP2C9	cg14191040	intragenic	Neuroendocrine and Adrenal	0.4567	pos	4.64E-14
POR	cg16684958	intragenic	Neuroendocrine and Adrenal	-0.4555	neg	5.66E-14
ABCB1	cg19199866	intragenic	Neuroendocrine and Adrenal	0.4555	pos	5.66E-14
F5	cg09891761	intragenic	Neuroendocrine and Adrenal	-0.4549	neg	6.13E-14
DPYD	cg07313437	intragenic	STS: other	0.3632	pos	7.38E-14
ABCB1	cg07469128	intragenic	STS: other	0.3635	pos	7.46E-14
CYP2A13	cg11988807	intragenic	Other	0.4743	pos	7.52E-14
RYR1	cg22393277	intragenic	Neuroendocrine and Adrenal	0.4529	pos	8.38E-14
TBXAS1	cg06239618	intragenic	STS: other	0.3622	pos	8.67E-14
DPYD	cg08024495	intragenic	Other	0.4721	pos	1.06E-13
DPYD	cg07717191	intragenic	Colorectal	0.6078	pos	1.06E-13
CYP2C8	cg12759420	intragenic	STS: other	0.361	pos	1.08E-13
RYR1	cg13982590	intragenic	STS: other	0.3605	pos	1.19E-13
DPYD	cg11227979	intragenic	STS: other	0.361	pos	1.23E-13
PTGIS	cg08788055	intragenic	Other	-0.4707	neg	1.28E-13

RYR1	cg01781084	intragenic	STS: other	0.3598	pos	1.31E-13
RYR1	cg20880196	intragenic	STS: other	0.3597	pos	1.34E-13
RYR1	cg01781084	intragenic	Other	0.4702	pos	1.34E-13
GSTP1	cg11566244	intragenic	Melanoma	-0.6175	neg	1.35E-13
ABCB1	cg14319793	intragenic	Neuroendocrine and Adrenal	0.4484	pos	1.61E-13
ABCB1	cg02364454	intragenic	Neuroendocrine and Adrenal	0.4478	pos	1.75E-13
DPYD	cg10890168	intragenic	Other	0.4684	pos	1.75E-13
CYP2W1	cg22025233	intragenic	Colorectal	-0.6035	neg	1.75E-13
GSTP1	cg11566244	intragenic	STS: other	-0.3575	neg	1.98E-13
CACNA1S	cg00095526	intragenic	STS: other	-0.3574	neg	1.99E-13
ABCB1	cg25438493	intragenic	Neuroendocrine and Adrenal	-0.4461	neg	2.25E-13
GSTP1	cg11566244	intragenic	Hepatopancreaticobiliary	-0.5883	neg	2.35E-13
SULT1A1	cg01332815	intragenic	STS: other	-0.3559	neg	2.62E-13
ABCB1	cg13287933	intragenic	Neuroendocrine and Adrenal	0.445	pos	2.62E-13
UGT1A1	cg01478198	intragenic	Other	-0.3382	neg	2.63E-13
UGT1A4	cg01478198	intragenic	Other	-0.3382	neg	2.63E-13
CYP4F2	cg01231180	intragenic	Neuroendocrine and Adrenal	0.4449	pos	2.63E-13
DPYD	cg02805559	intragenic	STS: other	0.3552	pos	2.87E-13
DPYD	cg24960006	intragenic	Colorectal	0.5991	pos	2.96E-13
DPYD	cg08024495	intragenic	Colorectal	0.5987	pos	3.10E-13
NAT1	cg20726908	intragenic	STS: other	0.3544	pos	3.29E-13
CYP2W1	cg15914863	promoter	Colorectal	-0.5979	neg	3.39E-13
DPYD	cg24960006	intragenic	STS: other	0.3541	pos	3.48E-13
ABCB1	cg00141548	intragenic	Neuroendocrine and Adrenal	-0.4427	neg	3.55E-13
COMT	cg08825848	intragenic	Neuroendocrine and Adrenal	-0.4418	neg	4.09E-13
PTGIS	cg08788055	intragenic	Breast	-0.6635	neg	4.33E-13
SULT1A1	cg14944435	intragenic	Hepatopancreaticobiliary	-0.5825	neg	4.66E-13
CYP2E1	cg16538390	intragenic	STS: other	0.3523	pos	4.74E-13
CYP2J2	cg11540204	intragenic	Colorectal	-0.595	neg	4.77E-13
DPYD	cg11226378	intragenic	Colorectal	0.5938	pos	5.63E-13
CYP2D6	cg12720083	intragenic	Hepatopancreaticobiliary	-0.5793	neg	7.01E-13
CYP2W1	cg08911208	intragenic	Neuroendocrine and Adrenal	-0.4379	neg	7.38E-13
RYR1	cg18684755	intragenic	Other	0.4577	pos	7.80E-13
POR	cg22567591	intragenic	Neuroendocrine and Adrenal	-0.4372	neg	8.15E-13
TBXAS1	cg05711445	intragenic	Other	-0.4572	neg	8.30E-13
COMT	cg08730070	intragenic	Melanoma	-0.6019	neg	8.87E-13
DPYD	cg16320208	intragenic	STS: other	0.3485	pos	9.24E-13
POR	cg03135313	intragenic	Neuroendocrine and Adrenal	-0.4359	neg	9.69E-13
CYP2S1	cg05312704	intragenic	STS: other	0.3479	pos	1.03E-12
NAT1	cg07470176	intragenic	STS: other	0.3473	pos	1.16E-12
COMT	cg16834011	intragenic	Neuroendocrine and Adrenal	-0.4344	neg	1.21E-12
CYP2E1	cg13092589	intragenic	STS: other	0.3469	pos	1.21E-12
DPYD	cg11055641	intragenic	STS: other	0.3465	pos	1.32E-12
DPYD	cg07313437	intragenic	Other	0.4539	pos	1.32E-12
SULT1A1	cg14944435	intragenic	Neuroendocrine and Adrenal	-0.432	neg	1.73E-12
NAT1	cg17579232	intragenic	Other	-0.4518	neg	1.79E-12
DPYD	cg13959721	intragenic	Melanoma	0.5953	pos	1.96E-12
PTGIS	cg01346423	intragenic	Hepatopancreaticobiliary	0.5704	pos	2.04E-12
SULT1A1	cg02266268	intragenic	STS: other	0.3439	pos	2.21E-12
GSTP1	cg07493922	intragenic	Urologic	-0.5973	neg	2.32E-12
RYR1	cg02226644	intragenic	Other	0.4498	pos	2.32E-12
COMT	cg10122187	intragenic	Melanoma	-0.5937	neg	2.34E-12
CYP2W1	cg07131210	intragenic	Neuroendocrine and Adrenal	-0.4297	neg	2.35E-12
ABCB1	cg05496710	intragenic	STS: other	0.343	pos	2.37E-12
NAT2	cg17522953	intragenic	STS: other	0.3422	pos	2.77E-12
CFTR	cg26635219	intragenic	Colorectal	-0.5794	neg	3.14E-12
RYR1	cg22393277	intragenic	STS: other	0.3414	pos	3.16E-12
ABCB1	cg00001224	intragenic	STS: other	0.3412	pos	3.25E-12
DPYD	cg14072140	intragenic	Colorectal	0.5789	pos	3.30E-12
ABCB1	cg09105881	intragenic	Neuroendocrine and Adrenal	-0.4269	neg	3.52E-12
CYP2A13	cg04017324	promoter	Other	0.4462	pos	3.76E-12
PTGIS	cg09062977	intragenic	Other	0.4461	pos	3.84E-12
DPYD	cg01750053	intragenic	Other	0.4455	pos	4.18E-12
DPYD	cg10890168	intragenic	Colorectal	0.5768	pos	4.20E-12
NAT1	cg15138846	intragenic	Other	-0.4446	neg	4.69E-12
DPYD	cg25633983	intragenic	Other	0.4491	pos	5.32E-12
GSTP1	cg22224704	intragenic	Urologic	-0.5895	neg	5.60E-12
CYP2W1	cg04141948	intragenic	Colorectal	-0.5743	neg	5.60E-12
DPYD	cg16511333	intragenic	Colorectal	0.5733	pos	6.32E-12
DPYD	cg11226378	intragenic	Other	0.4422	pos	6.49E-12
DPYD	cg17140186	intragenic	Other	0.4421	pos	6.55E-12
SLCO2B1	cg05706446	intragenic	STS: other	-0.3362	neg	7.52E-12
UGT2B15	cg01714676	intragenic	Neuroendocrine and Adrenal	0.4214	pos	7.52E-12
ABCB1	cg07872519	promoter	Neuroendocrine and Adrenal	0.4219	pos	7.71E-12
TBXAS1	cg06365890	intragenic	Other	-0.4401	neg	8.66E-12