Evaluation of Web Service Based Querying of Pharmacogenomics (PGx) Clinical Guidelines Using MyVariant.info, PharmGKB and HGVS Nomenclature

Moritz Juchler

MASTERARBEIT

eingereicht am
Masterstudiengang

Medizinische Informatik

in Heilbronn

im Januar 2016

Declaration

I hereby affirm that I composed this Master thesis by myself, that the work contained herein is my own except where explicitely stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified; nor has it been published.

Heidelberg, January 20, 2016

Moritz Juchler

Contents

Preface

This work is the documentation of my master's thesis project at the Department of Biomedical Informatics and Medical Education, University of Washington, in Seattle, Washington, USA. It was written for the Institute of Medical Biometry and Informatics at Heidelberg University and the Faculty of Informatics at Heilbronn University.

Acknowledgements

First of all, I would like to thank Peter Tarczy-Hornoch for inviting me to Seattle to conduct my masters research at the BIME department and Thomas Wetter for offering me the possibility to take this opportunity and help establish contact.

Special thanks to Sean D. Mooney for being a great supervisor during my visit in Seattle and for supporting me with plenty of suggestions, great ideas and explanations, and for teaching me valuable lessons about the work as a researcher.

I am incredibly grateful for the persistent support and everlasting motivation by my parents and my sister, who were always there for me.

Kurzfassung

Hunderttausende Patienten leiden jährlich an Therapieversagen oder unerwünschten Arzneimittelwirkung und können deswegen von pharmakogenomischen Tests profitieren. Für die Anwendung der Pharmakogenomik im klinischen Alltag wird automatisierte Entscheidungsunterstützung benötigt. Momentant liegt das Wissen lediglich als Leitlinien vor, in Textform und unstrukturiert.

Die vorliegende Arbeit evaluiert, ob ein hier entworfener Webservice klinische relevante Genvarianten mit Informationen aus Leitlinien annotieren kann. Das vorgestellte Programm gibt patientenspezifische Informationen pharmakogenomischer Leitlinien in formalisierter Darstellung wider und ermöglicht die Annotierung von Genomen im Variant Call Format (VCF) mit Informationen der Pharmacogenomic Knowledge Base (PharmGKB) und Leitlinien der Clinical Pharmacogenetics Implementation Consortium (CPIC).

Die Eignung des Webservices, klinische relevante Genvarianten mit Informationen pharmakogenomischer Leitlinien zu annotieren, wird evaluiert, indem fünf Leitlinien in den Webservice integriert werden und das Programm auf öffentlich verfügbaren Genomen getestet wird. Der Workflow findet genetische Varianten, die in CPIC Leitlinien beschrieben werden und durch diese Genvarianten beeinflusste Medikamente.

Die Ergebnisse zeigen, dass der Webservice benutzt werden kann, um schnell klinisch relevante Genvarianten mit aktuellen Informationen aus pharmakogenomischen Leitlinien zu annotieren, wobei Hürden wie die Übersetzung von Genvarianten in die Star Allele Nomenklatur oder das Fehlen einer einzigen Haplotyp Nomenklatur die Anwendungen dieser Herangehensweise an anderen Medikamenten und in der Klinik schwierig machen.

Abstract

Every year, hundreds of thousands of patients are affected by treatment failure or adverse drug reactions, many of which could be prevented by pharmacogenomic testing. To address these deficiencies in care, clinics require automated clinical decision support through computer based systems, which provide clinicians with patient-specific recommendations. The primary knowledge needed for clinical pharmacogneomics is currently being developed through textual and unstructured guidelines.

In this thesis, it is evaluated whether a web service can annotate clinically relevant genetic variants with guideline information using web services and identify areas of challenge. The proposed tool displays a formal representation of pharmacogenomic guideline information through a web service and existing resources. It enables the annotation of variant call format (VCF) files with clinical guideline information from the Pharmacogenomic Knowledge Base (PharmGKB) and Clinical Pharmacogenetics Implementation Consortium (CPIC).

The applicability of the web service to annotate clinically relevant variants with pharmacogenomics guideline information is evaluated by translating five guidelines to a web service workflow and executing the process to annotate publically available genomes. The workflow finds genetic variants covered in CPIC guidelines and influenced drugs.

The results show that the web service could be used to annotate in real time clinically relevant variants with up-to-date pharmacogenomics guideline information, although several challenges such as translating variants into star allele nomenclature and the absence of a unique haplotype nomenclature remain before the clinical implementation of this approach and the use on other drugs.

Acronyms

ADR Adverse Drug Reaction

CDS Clinical Decision Support

CNV Copy Number Variation

CPIC Clinical Pharmacogenetics Implementation Consortium

dbSNP Single Nucleotide Polymorphism Database

EHR Electronic Health Record

HGVS Human Genome Variation Society

INDEL INsertion DELetion

JSON JavaScript Notation Format

PGP Personal Genome Project

PGx Pharmacogenomics

PharmGKB Pharmacogenomic Knowledge Base

POC Point-Of-Care

SNP Single Nucleotide Polymorphism

VCF Variant Call Format

Chapter 1

Background

It has been known for decades that the clinical response to drug treatments can vary significantly between individuals [1]. This results in hundreds of thousands of treatment failures or other adverse drug reactions (ADRs) every year. A therapy can be effective and still cause serious adverse events in one subgroup of patients, while delivering no response in terms of toxicity or therapeutic effect in others [2].

The response rates of treatment in 14 therapeutic areas like Alzheimer's (30%), diabetes (57%), and cancer (25%) have varied from 25-80%, with an average response rate of just over 50% [3]. This implies that the remaining patient population does not receive optimal medication or suffers from adverse events. To further illustrate this, approximately 0.24% of the population receive treatment in emergency departments with adverse drug events every year in the US [4].

Serious ADRs occur in 6.7% of hospitalized patients, while fatal ADRs occur in 0.32% of hospitalized patients and represent a frequent event estimated to be between the fourth and sixth leading cause of death in the USA, clearly posing challenges to the healthcare system in terms of patient wellbeing and medical costs [5, 6]. Under the simplifying assumption that the results of the US studies are applicable to Germany, the 16.5 million hospital cases of the year 2001 resulted in 31.600 - 83.000 casualties due to the undesired effects of medical interventions in hospitals. A recent study estimated the costs related to ADRs in Germany at about €816 million[7].

Historically, the efficacy and safety of therapies in the "average patient" have been obtained through randomized clinical trials on large cohorts. Recent progress in genomics has lead to a much better understanding of specific molecular influence to variability in phenotypic response [8]. Part of the variability in drug response can be explained by genetic variation between patients, which can influence how each drugs is metabolized [9] - the focus of the discipline called pharmacogenomics (PGx) [10]. The understanding of PGx responses, which accounts for an estimated 24%-95% of the variability

in drug disposition and effects [11], made possible the optimization of outcomes by individualizing therapy through dosing recommendations or a different selection of medication. Although many nongenetic factors influence drug response, including age, drug-drug interactions, add-on treatments and organ functions, there are many cases in which interindividual differences in medication effects are attributable to sequence variants in drug-metabolizing genes [1, 12, 13].

The objective of PGx is to develop ways to individualize therapies for patients and thereby optimize outcomes through knowledge of the variability of the human genome and its influence on drug response.

1.1 Variants & Haplotypes

This chapter describes details of interindividual genetic differences mentioned in the last section and their significance on drug response.

Our knowledge of the genetic influence on individual drug responses has progressed through research on genes and medications relevant to a wide range of diseases. A polymorphism is defined as a DNA sequence variant present in >1% of the general population [14]. If a DNA sequence is present in <1% of the population, it is defined as a mutation.

An allele refers to two different versions of the same gene at a certain position on a chromosome [15] . Within a gene, variations of an individual nucleotide can be considered alleles. Alleles include the wild type: The usual sequence, mutations and polymorphisms of each gene. Two alleles of each gene at a locus is defined as the genotype. The phenotype is the set of characteristics or the clinical presentation of an individual, which results from the particular genotype.

The genetic variations that have been examined include single-nucleotide polymorphisms (SNPs), genomic insertions and deletions (INDELs), and genetic copy-number variations (CNVs). SNPs are at 90% the most frequent sequence variation [16]. SNPs that appear within the same region of the DNA are often statistically associated in haplotypes (group of alleles). In general, variants in haplotypes are <55 kilobases apart [17].

Drug efficacy is influenced by haplotypes in drug-metabolizing genes and genes that encode for drug receptors, transport protein, and drug targets. For example, the dose levels required by individual patients are strongly influenced by a common promoter variant in the molecular target of warfarin (VKORC1). The haplotype VKORC1 encodes the vitamin K-epoxide reductase protein, which is the target enzyme of warfarin. Variants in VKORC1 alter a transcription factor binding site, which leads to lower protein expression, and thus altered warfarin sensitivity and reduced dose requirements [18]. This makes it a candidate gene for the variability in warfarin response [19].

1.2 Variant and Haplotype Nomenclature

Even if comprehensive genetic test results are brought to clinicians that include variants and haplotypes as described in the last chapter, it remains challenging to interpret the results consistently [20]. Reasons include the rapid discovery and publication of new variants and the non-enforcement and non-compliance of one single nomenclature [21], as well as variable annotation pipelines across institutions [22]. Therefore, and due to increasingly complex and thus error-prone descriptions, unambiguous variant description is of great importance. Often journals require compliance with specific recommendations, but rarely enforce them. In the following the three most important description methods are presented.

1.2.1 dbSNP

Reference SNP cluster rs#'s are created by the National Center for Biotechnology Information (NCBI) during periodic builds of the Single Nucleotide Polymorphism Database (dbSNP). A reference SNP cluster record has the format rs[NCBI SNP ID] where 'rs' is always lower case [23]. The SNP nomenclature of dbSNP is widely adopted and referenced in the literature.

However, some historical variants may never be reported in dbSNP by the investigator, thus some PGx variants may not appear in the database. Another challenge is abandoned rs#'s due to regular clustering, which could lead to ambiguity in publications [24].

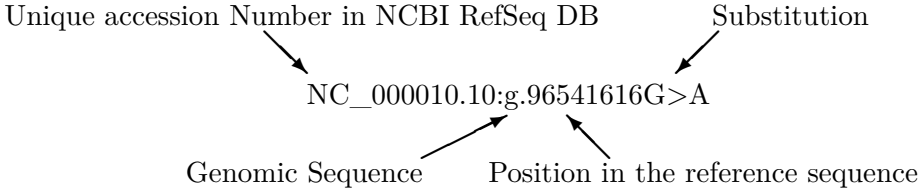
1.2.2 HGVS nomenclature

A generic syntax has been recommended by the Human Genome Variation Society[1] (HGVS) [25] and has been adopted broadly. According to the proposed standard, variants should be described by the following expression:

> <Reference sequence><Type of concerned sequence><Position of the variant in the reference sequence><Observed variation>

Conjunction of those four fields yields an unambiguous representation of the variant.

Example:



[1] The latest version can be found at the homepage of HGVS http://www.hgvs.org/mutnomen

However, this nomenclature has not been universally adopted yet [24] although it is widely used. One reason is the possibility of ambiguity of names due to changes in the reference sequence as well as changes to the syntax itself.

1.2.3 Star allele nomenclature

The star-allele nomenclature was created in order to standardize annotation for haplotypes in the cytochrome P450 genes [26] and is used for many genes. A star allele represents either a single genetic variant or a haplotype.

PGx studies require genetic testing of individuals for multiple variants in drug metabolism enzyme and transporter genes. In order to understand the phenotype, genotyping results must often be translated to the star (*) allele nomenclature. Star alleles are haplotype patterns that have been defined at the gene level and often have been connected with protein activity levels. Typically *1 is the most commonly occurring wild-type allele. A patient who carries two wild-type alleles for a gene would be described as having a *1/*1 genotype, which is associated with normal gene activity. Every other allelic variant such as *2 etc. is in some way benign, nonfunctional, hyperactive, or partially active.

It is crucial to know the combination of variants inside a given haplotype and the diploid content in an individual to study drug metabolism, drug response, and adverse drug reactions . For example, carriers of two reduced-function CYP2C19 alleles, such as *2/*2, are associated with low gene activity. The drug clopidogrel requires activation by CYP2C19, which means that patients with low gene activity have reduced active clopidogrel metabolites compared to patients with normal metabolism [27].

It is often challenging to translate SNP combinations into star allele nomenclature after finding novel or rare combinations of variants that do not exactly match existing star alleles. Furthermore, since variants are shared by multiple star alleles, finding a unique star allele combination can be problematic. Also, due to the use of drug metabolizing assays in contrast to whole genome sequencing (WGS) some variant markers could be missing, which could lead to failure of the haplotype mapping and false negatives, incorrectly identifying patients at risk of an ADR as not-at-risk.

1.3 Clinical Decision Support

The increasing number of clinically relevant genes and their corresponding variants will soon make it impossible for a clinician to perform without decision support. The challenges of reporting, organizing, and interpreting PGx test results can be reduced by the increasing adoption of electronic health records (EHR) [28]. Ideally clinicians are supported by alerts from the clinical decision support (CDS) system, which allow for point-of-care

(POC) interventions [29], and enable gene-based drug dosing years after the genetic test result is reported [30]. CDS tools should warn users when attempting to prescribe, dispense or administer high-risk affected drugs. Effective CDS facilitates the use of PGx results like drugs with decreased efficacy or increased toxicity over a patient's lifetime [31]. Further possible applications of CDS are dosing support, POC alerts, and displaying relevant information at the right time in the right form to the right people. These scenarios illustrate the invaluable importance of integration of medication use with lab testing via CDS.

Shirts BH et al. [32] proposed an alert in the EHR at the time the prescription is written, which would prompt the clinician to order a specific genetic test to help predict response to drug therapy.

1.4 PharmGKB & CPIC Guidelines

In the presented method, PGx information is retrieved from a centralized repository called Pharmacogenomic Knowledge Base (PharmGKB), funded by the NIH[33]. PharmGKB was initiated in 2000 and uses peer-reviewed published articles. It manually summarizes the clinical information into dosing guidelines and drug labels; clinically actionable gene-drug associations and genotype-phenotype relationships. The genetic variant annotations comprise population details of individual studies and use statistics like P-values, ontologies such as DrugBank [34], and the HUGO Gene Nomenclature Committee (HGNC) [35] as well as standardized sentences, which allow for easy comparison of results and quick acquisition of key information [33, 36].

It also offers a summary of all variant annotations between variant and drug response, which includes the level evidence. There are four levels of evidence ranging from preliminary (level 4) to high (level 1) with the majority of evidence showing a variant-drug association, which must be replicated in more than one cohort with significant P-value and strong effect size. Additionally, PharmGKB provides haplotype – star allele mapping through translation tables that can be downloaded for some genes.

CPIC was founded in 2009 to address a barrier to the clinical implementation of PGx [37]. CPIC guidelines are designed to help clinicians understand how available genetic test results should be used to optimize drug therapy. Between 2011 and 2015 24 CPIC guidelines have been published[2]. A consortium of PGx and domain experts write drug dosing guidelines assuming that the clinician has the relevant genotypes at hand. The CPIC guidelines contain information regarding both the drugs and genes of interest, with emphasis on tables mapping genotype to phenotype, and phenotype to dosing/prescribing information [38–40].

[2] https://www.pharmgkb.org/page/cpic

1.5 MyVariant.info, JSON and VCF syntax

JavaScript Object Notation (JSON) is a compact data-interchange format to transmit hierachically structured data objects, while maintaining the readability for both human and computers. It was first introduced at the JSON.org website in 2001. The MyVariant.info[3] platform aggregates multiple variant annotation sources like dbSNP and ClinVar (the complete data sources are available on the metadata website of MyVariant.info[4]) into a single web service API by merging all annotations relevant to a variant into a single JSON object using HGVS nomenclature as key. Through community efforts it succeeded in amassing annotations for more than 300 million variants.

Genetic variants, such as SNPs, INDELs or larger structural variants, are commonly stored in variant call format (VCF) files [41]. Besides the reference base(s) and the alternate base(s), VCF files contain tab-delimited fields to describe the call quality and additional information.

1.6 Using web services to annotate clinical guidelines

Below a potential mechanism shall be presented providing automated annotation of genetic variants using clinical guidelines through web services and available databases and websites. The objective is also to identify the challenges and difficulties in annotating PGx guidelines automatically with available services and demonstrate utility. Despite the current limitations described above, putting this workflow into clinical practice is still evaluated.

[3] http://myvariant.info
[4] http://myvariant.info/metadata

Chapter 2

Methods

2.1 Systematic literature review

In order to assess the presence of such services being developed or already in use, a literature review was conducted.To this end, a Pubmed search was conducted covering the years from 2005 to 2016 and using a search strategy adapted from previous systematic reviews of CDS [42, 43] and genetic health services [44]. The final literature search was conducted on January 3, 2016. The inclusion criteria for the review were as follows:

1. English or German article
2. human focus
3. manuscript in peer-reviewed journal
4. primary focus on the use of computers to deliver genetically guided, patient-specific recommendations to guide decision making on drug therapy

The final search query was:

```
("Decision Making, Computer-Assisted"[Mesh] OR "Decision Support Systems, Clinical"[Mesh] OR "Expert Systems"[Mesh] OR "Decision Making"[Mesh:NoExp] )
AND (
    "Drug Therapy"[Mesh]
        AND (
            "genetics"[Mesh] OR "genetic variation"[Mesh] OR "Genomics"[Mesh] OR "genetic predisposition to disease"[Mesh]
            )
    OR "Precision Medicine"[Mesh]
    OR "Pharmacogenetics"[Mesh]
    )
AND ("2005"[dp] : "2016"[dp])
AND (english[la] OR german[la])
```

For all identified references the title, index terms, and available abstracts were reviewed to determine if the articles have met all inclusion criteria. If

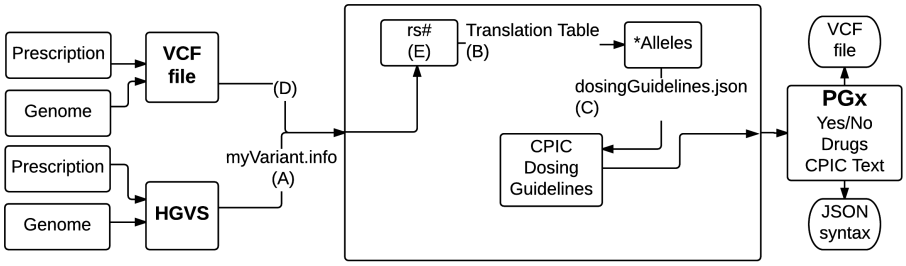


Figure 2.1: General Workflow of the approach showing data input, annotation and analysis and result presentation.
(E) rs# not available for all PGx variants

at this stage insufficient information made it impossible to make a decision, the article was included for full-text retrieval. Each full-text article was then reviewed to determine its final status.

For each of the articles that met the inclusion criteria above, the clinical application area, article type, study location, and notable informatics aspects were abstracted.

2.2 General workflow

The overall architecture of the proposed method is shown in Fig. 2.1. It illustrates the translation of file types and variant descriptors to biological entities like alleles and recommendations related to pairs of actionable variant-drug associations. The general workflow consists of three basic steps:

1. The input on the left originating from a user or a sequencing instrument
2. The computational workflow in the middle is written in Python, which translates genetic information into PGx recommendations
3. The output on the right with the two result format possibilities: An edited VCF file or JSON format

The remainder of this section describes the methods used to implement these components in more detail. All resources and scripts described in this section are freely available [1].

2.2.1 Translating HGVS identifiers to rs#'s

Users can enter the workflow using two different id keys, the genomic HGVS nomenclature and the rs# (Figure 2.1 (A)). The reason for using rs# as internal reference ID is the common use as reference for variants in the

[1] https://github.com/mjuchler/VCF-annotation-pharmacogenomics

	A	B	D	E	F	G	H
1	**Haplotype Set Id**	PA165980496					
2	**Haplotype Set Name**	Haplotypes	19154G>A*	80160C>T*	12460G>C*	-98T>C*	12662A>G*
3	**Haplotype Id**	**CYP2C19**	**rs4244285**	**rs3758580**	**rs17878459**	**rs4986894**	**rs12769205**
4	PA165980634	*1	G	C	G	T	A
5	PA165816510	*1A	G	C	G	T	A
6	PA165816511	*1B					
7	PA165816512	*1C					
8	PA165980635	*2	A				
9	PA165816513	*2A	A	T			
10	PA165816514	*2B	A	T	C		
11	PA165945747	*2C	A	T		C	G
12	PA165816516	*2D	A	T		C	G
13	PA166123286	*2E	A	T			
14	PA166123287	*2F	A	T			
15	PA166123288	*2G	A	T			
16	PA166123289	*2H	A	T			
17	PA166123290	*2J	A	T			
18	PA165980636	*3					
19	PA165945748	*3A					

Figure 2.2: Translation Table for CYP2C19 derived from the Human Cytochrome P450 (CYP) Allele Nomenclature Database and downloaded from PharmGKB.

context of our system, specifically PharmGKB. The specifics of entering the workflow with a VCF file are described below. Since HGVS names are unique and are required to be always mapped to a standard genome build (currently hg19), they can be used as keys for describing variants. By using MyVariant.info, the HGVS IDs are mapped to the rs# which will be used as internal reference ID as seen in Figure 2.1 (E).

2.2.2 Mapping rs# to Star Alleles and Other Haplotypes

The haplotype translation tables from PharmGKB are used to map rs#'s to star alleles (Figure 2.1 (B)). Each line in one of the Excel tables (Figure 2.2) describes one star allele as a list of rs#'s. The Python script downloads the translation tables from PharmGKB and performs a search for the rs# through all available translation Excel tables, and, if the rs# was found, it returns a list of haplotypes. This list of haplotypes contains either one or multiple haplotypes. Figure 2.2 shows an example for a translation table for the gene CYP2C19. In this instance, a patient carries the variant rs4244285, which is a substitution of guanine to adenine (G>A) at chr10:94781859.

A search for this variant is conducted. The results show a match in the translation table. After checking that the patient has this specific SNP, the haplotype *2 gets returned (see black arrows in Figure 2.2). The translation tables, i.e. the translations from haplotypes to variants for all of the genes on PharmGKB, come from human pharmacogene variant nomenclature committees or other resources. For example, the resource for the translation table in Figure 2.2 is from the Human Cytochrome P450 (CYP) Allele Nomen-

Listing 2.1: An example of returned dosing guidelines for a genetic variant, in JSON syntax, highlighting the challenge of allele assignment from single variants.

```
{
    "_id": "rs4244825",
    "pharmgkb": {
        "rsid": "rs4244825",
        "gene": "CYP2C19",
        "haplotypes": [
            "CYP2C19*2"
        ],
        "drugrecommendations": [
            {
                "drug": "clopidogrel",
                "haplotypes": "*1*2",
                "levelOfEvidence": "Level of Evidence: Moderate",
                "recommendation": "Alternative antiplatelet therapy (if no contraindication); e.g., prasugrel, ticagrelor"
            },
            {
                "drug": "amitriptyline",
                "haplotypes": "*1*2",
                "levelOfEvidence": "Level of Evidence: Strong",
                "recommendation": "Initiate therapy with recommended starting dose"
            }
        ]
    }
}
```

clature Committee [2].

2.2.3 Linking rs# to Relevant Guidelines

Now, the rs# match needs to be linked to guideline recommendations (if present for this variant). For some gene-drug combinations PharmGKB provides downloadable genotype based dosing guidelines in JSON format on their homepage [3]. Most of these guidelines are based on star allele nomenclature as a key for the drug recommendations, which were just received from 2.2.2. By using the Python script the phenotype, the metabolizer status, implications and dosing recommendation texts, plus the level of evidence can be then automatically retrieved when given a haplotype as an input parameter (see Figure 2.1 C).

Listing 2.1 shows an example result that the system returns when a user

[2] http://www.cypalleles.ki.se/
[3] https://www.pharmgkb.org/downloads/

enters the workflow with the variant rs4244285. This includes a drug recommendation for clopidogrel and another for amitriptyline, both of which are inherited through the *1/*2 haplotype. Either include the CPIC recommendation text (“Alternative antiplatelet therapy...” or “Initiate therapy with recommended starting dose”) and the level of evidence (“Moderate” or “Strong”). Also included is the list of haplotypes where the variant is present (“CYP2C19*2”).

2.2.4 Linking VCF files to relevant clinical guidelines

To insert PGx guideline information about a variant into a VCF file (Figure 2.1 (D)) the corresponding user extensible annotation (INFO) field is edited, which is one of the fields in a variant-line. A meta-information line in the header section has to be added to provide a description of tags and annotations.

The INFO fields should be described as follows (all keys are required):

```
##INFO=<ID=DRUG,Number=.,Type=String,Description="Influences response of mentioned drugs">
```

As ID of the PGx recommendation, ‘DRUG’ is proposed. The Number entry is an Integer that describes the number of values that can be included with the INFO field. When the INFO field contains only a single number, then this value should be 1; if the INFO field describes a pair of numbers, then this value should be 2, and so on. If the number of possible values varies this value should be ‘.’ Since variants typically influence more than one drug, a ‘.’ i.e. ‘unknown’ is used.

Possible Types for INFO fields are: ’Integer’, ’Float’, ’Flag’, ’Character’, and ’String’. In this case the type is ‘String’ because influenced drugs should be added.

Eventually, the meta-information line about the newly added PGx information should look like this:

```
DRUG=clopidogrel,amitriptyline
```

The INFO fields are encoded as a semicolon-separated series of short keys with optional values in the format: <key>=<data>[,data]. They typically contain information about membership of the variant in databases like dbSNP, HAPMAP or information about the ancestral allele. If there is more than one influenced drug the separation gets added by a comma. Adding PGx information to the INFO fields could therefore look like this:

In the future, more detailed PGx information like the dosage recommendations could be added. The following line demonstrates how it was

#CHROM	POS	ID	REF	ALT	QUAL	INFO	FORMAT	SAMPLE1	SAMPLE2	SAMPLE3
10	9...	rs...	G	A	.	A...	GT:GQ:DP	1|1:94:35	0|0:54:18	1|0:100:47

originally found in the VCF file. Before that comes the header line that names the eight fixed, mandatory columns.

Before adding PGx recommendation, which has an influence on two drugs to this variant, first the newly created meta-information line has to be added to provide the necessary knowledge how to process and read the newly added annotation. The line for the sequence variant itself, which now contains information about the influenced drugs, needs to follow the newly constructed meta-information line and gets information added into the INFO field separated by a semicolon. This is how the VCF file could look like after the proposed service (the newly added annotation is marked red):

##INFO=<ID=DRUG,Number=.,Type=String,Description="Influences response of mentioned drugs">							
#CHROM	POS	ID	REF	ALT	INFO	FORMAT	SAMPLE1
10	9...	r...	G	A	DP=105;DRUG=clopidogrel,amitriptyline;HM2	GT:GQ:DP	1|1:94:35

2.3 Application to existing genomes: Evaluation

To test the workflow this approach was applied to individuals in the personal genome project (PGP[4]) [45]. The PGP is a public database of whole-genome sequence data with phenotypic information from voluntary participants [46].

In order to do so, all available VCF files from participants were downloaded from the PGP website and analysed via the proposed workflow.

Three figures were calculated to quantify the prevalence of PGx variants and to estimate the usefuleness of this proposal.

1. The number of rs#'s per participant is defined by how many of the variants in the VCF files are found in the PharmGKB translation tables. This number is just an indicator to see if there are any rs#'s found in the translation tables, as not all of the rs#'s influence drug response. Since only some of these variants are subject to CPIC guidelines, the following data was collected:
2. The number of clinically actionable PGx variants per participant that are covered in the CPIC guidelines and have an actual impact on drug response.
3. The number of drugs, whose response is influence by the genetic variation found in a patient and that have drug dosing guidelines available.

After these numbers were calculated, they were averaged over all found participants.

[4] https://my.pgp-hms.org/public_genetic_data

Chapter 3

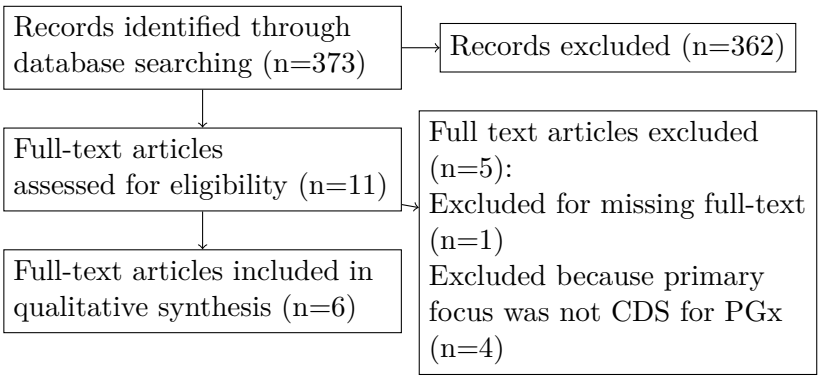
Results

3.1 Systematic literature review

The initial Pubmed search identified 373 potentially relevant articles. During the title and abstract review, one article was rejected because it did not focus on humans, 17 articles were rejected for not being peer-reviewed, and 344 articles were rejected because the primary focus of the work was not on the use of computers to deliver genetically guided, patient-specific drug therapy recommendations. The remaining eleven articles underwent full-text review: At this stage another article was rejected because its full-text was not retrievable; and four more articles were rejected because they primarily did not focus on the use of computers to deliver genetically guided, patient-specific medication guidance (Figure 3.1). The final set of included manuscripts, all published between the years 2012 to 2014, consisted of 6 primary research articles.

Table 3.1 summarizes the six research articles identified. These studies include a feasability study comparing warfarin dose prediction algorithms

Figure 3.1: Manuscript selection process



[47]. Other investigators focused on how CDS for PGx could be integrated with primary clinical information systems such as provider order entry systems [48–50], which would allow users to electronically write orders, track online medication adminstration records, and review changes made to an order [51]. The evaluated studies refer to web resources at POC in order to gain information on the relevance of specific PGx results [31], make use of the CPIC guidelines [31, 47, 49, 52], and use star allele nomenclature [47, 48].

Table 3.1: Summary of research on CDS systems for PGx

Citation and name of system (if applicable)	Manuscript summary and trial details	Users and study location	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Bielinski, 2014 [49]; RIGHT protocol	This study proposed a multivariable prediction model to identify patients with a high risk of initiating statin therapy	Clinicians in USA	Yes	CDS is integrated in the EMR and flags patient-specific drug-gene interactions and provides therapeutic guidance	System description	"Alert fatigue" is considered in the design and exclusion criteria are included in the rules to avoid unnecessary repetitive alerts.
Liu, 2012 [47]	This study compared the performance of 8 PGx algorithms to predict warfarin dose	Clinicians in China	No	Therapeutic dose guidance for warfarin	Feasability study	Single population algorithms tended to perform better than mixed ones
Goldspiel, 2014 [50]	System description of how the National Institutes of Health Clinical Center implemented a CDS logic for human leukocyte antigen variants to predict severe hypersensitivity reactions	Clinicians in USA	Yes	Providing PGx CDS in the EHR during order entry. The system works for three drugs and HLA variants.	System description	The CDS is programmed using Medical Logic Modules with Arden Syntax programming language

Bell, 2014 [31]	This study presents how St. Jude Children's Research Hospital implemented active CDS for PGx test results	Clinicians in USA	Yes	Active CDS utilizes automated alerts to intercept the clinician at the POC and includes both pre- and post-test alerts.	System description	In an effort to prevent alert fatigue, they targeted alerts only for the relatively rare event of prescribing a high-risk drug to a patient with a high-priority phenotype.
Gottesman, 2013 [52]; CLIP-MERGE	Description of how the Icahn School of Medicine is preemptively genotyping a panel of germline PGx variants, storing data in an external data-management platform that interfaces with the EHR, and delivering CDS at the POC through the EHR.	Clinicians in USA	Yes	CDS presents genetic variants with established clinical significance (e.g., gene/drug pairs with CPIC guidelines) to internal medicine physicians through the EHR	System description	CDS rules are based on actionable variants extracted from each patient's genotype data, and are combined with relevant phenotypic data in the project database, which includes longitudinal clinical data extracted from the EHR
Pulley; 2012 [48]; PREDICT	Program description how Vanderbilt University MC implemented prospective PGx testing for antiplatelet therapy and cardiovascular stents	Clinicians in USA	Yes	Active CDS deploying POC decision when clopidogrel is prescribed for those with variant genotypes	System description	Genotype data is stored in a separate DB from the EHR

3.2 Design

A method is demonstrated that can annotate genetic variants using up to date CPIC guidelines and other current web resources. The data from PharmGKB, namely the haplotype translation tables and the CPIC guidelines in JSON format, provides powerful curated resources for translational research. The workflow written in Python uses these resources to annotate a patients VCF file, and results in personalized drug dosage recommendations.

3.3 Evaluation

To evaluate the accuracy and relevance of the workflow this approach was applied to individuals in PGP. 14 participants had VCF files available. These VCF files were downloaded from the PGP website and the workflow was executed on this data. Table 3.2 shows the number of SNPs per participant, the number of rs#'s found in translation tables per participant, the number of clinically actionable PGx variants per participant that are covered in the CPIC guidelines, and the number of drugs whose response is influenced by the genetic variation found in a patient and that have drug dosing guidelines available.

Table 3.2: Number of SNPs, PGx variants and influenced drugs per participant

part ID	SNPs	rs#'s	CPIC	drugs
hu0B13B7	109085	17	7	31
hu250634	2391739	27	9	25
hu34D5B9	145012	28	12	31
hu4040B8	114863	14	8	21
hu448C4B	120481	22	10	31
hu4963A1	105835	10	4	19
hu555913	5194660	37	13	32
hu6ABACE	103986	20	6	36
hu80855C	3478763	18	5	13
hu97DB4A	112669	10	4	19
huA3A815	112881	8	5	21
huAA16BD	176044	21	7	31
huD7960A	104001	11	6	19

Shown in table 3.3 are the mean figures for these values. The mean number of clinically actionable PGx variants per participant is 18.1. This indicates that there are rs#'s in the participants VCF files that are found in the PharmGKB translation table. Since only some of these variants are

Table 3.3: Mean and median figures of SNPs, PGx variants and influenced drugs per participant

part ID	SNPs	rs#'s	CPIC	drugs
mean	883837.4	18.1	7.1	24.9
median	113872.0	17.5	6.5	23.0
mean w/o 3 big	118960.1	15.5	6.6	25.3
median w/o 3 big	112669.0	14.0	6.0	21.0

subject to CPIC guidelines and therefore mostly have an actual impact on drug response the next number is of higher importance.

The mean number of clinically actionable PGx variants per participant that are covered in the CPIC guidelines and have an actual impact on drug response is 7.1 with a median of 6.5.

The number of drugs with different generic names whose response is influenced by the genetic variation found in a patient and that have drug dosing guidelines available is 24.9.

Due to the big variation in SNPs in the three participants with IDs hu250634, hu555913, and hu80855C, these participants were later removed from the mean and median numbers, which decreased the standard deviation from 1.559.098 to 21.285.

Table 3.4 describes all the gene-drug pairs with computable CPIC guidelines and translation tables available in PharmGKB. As of today there are 157 unique drugs and 65 unique genes with PharmGKB guidelines. There are still hundreds of gene-drug pairs that will be transformed into CPIC guidelines. Many of the CPIC guidelines cover multiple gene-drug pairs in one guideline, which results in 168 alleles with CPIC guidelines.

Table 3.4: Table of Gene-Drug Pairs with computable CPIC guidelines and available translation tables

Gene	Drugs
CYP2C19	Amitriptyline, clomipramine, clopidogrel, doxepin, imipramine, selective serotonin reuptake inhibitors, sertraline, trimipramine, citalopram, escitalopram
CYP2D6	Amitriptyline, clomipramine, codeine, desipramine, doxepin, fluvoxamine, imipramine, nortriptyline, paroxetine, selective serotonine reuptake inhibitors, trimipramine
DPYD	Capecitabine, fluorouracil, tegafur
TPMT	Azathioprine, mercaptopurine, thioguanine
UGT1A1	Irinotecan, nilotinib

Chapter 4

Discussion

4.1 Discussion of literature review

The systematic review has provided an overview of the existing range of CDS, delivering genetically-guided, patient-specific assessments or recommendations to clinicians in order to guide decision-making on drug therapy. The results enabled us to effectively design a web service, which provides CDS with choosing drug dosing after finding genetic variants, and to overcome drawbacks found in the studies. In conducting a literature search that extends from the year 2005 to 2016, 373 manuscript were screened and six primary research articles included. All of the articles describe genotype-driven CDS.

The limitation of this study is that it only included manuscripts written in either English or German, which may have led to the exclusion of other manuscripts published in a different language. Second, some articles from the years 2015 and 2016 may not have been indexed during the time of the research and therefore possibly falsely excluded. Third, some relevant studies may not have been found because the corresponding journals may not be indexed by MEDLINE and thus do currently not include MeSH subject terms. Finally, there is the issue of a potential publication bias concerning clinical trials, i.e. studies with unsuccessful outcomes are less likely to be published than studies with successful outcomes. The high success rate (6 out of 6) indicates such a bias, whereas the expected rate of success rate would typically be in the range of around 60% [53]. The limited sample size could explain this discrepancy. The high success might also stem from study protocols, which required the use of these systems and thus, increases the probability of its use.

So far, no study has evaluated the integration of means to keep the PGx knowledge base, and therefore the CDS rules up-to-date, making the proposed system unique among the included articles. Another aspect that is not discussed in the presented studies is the direct use of the patients VCF file

as a source of genetic variants at POC. Most of the investigators only store the clinically actionable PGx variants in the patient EHR to drive POC CDS. This involves a range of disadvantages. First, knowledge on newly discovered clinically actionable PGx variants are not automatically integrated into the saved actionable PGx variants in the EHR. Instead, they need to be integrated manually, which could take a long time or never happen if the development of PGx projects is discontinued. Second, the PGx knowledge base about one previuosly known PGx variant could change through further research, making its continued use potentially dangerous. A study conducted in 2012 has shown that over the span of six years there were 214 classification changes to previously reported variants for genetic testing on hypertrophic cardiomyopathy alone. These changes in variant classification, if not quickly translated into the CDS, could lead to misclassifications and ultimately the prescription of a drug or drug dosage that could lead to an ADR.

As demonstrated in this study, both of these drawbacks are eliminated with the proposed workflow by a) using topical PGx guideline information like the CPIC guidelines, and b) using a patient VCF file as the source of clinically actionable PGx variants instead of the variants stored in the EHR.

4.2 Discussion of web service

The web service fulfills key requirements to successful clinical implementation of PGx, such as consistent interpretation of test results and availability of guidelines for drug prescription based on test results. A major advantage is the topicality of the annotation results by querying online sources while the system performs the search. Most similar attempts query data sources once and then store the PGx data in databases where no new knowledge becomes integrated.

Challenges for implementation exist. First and foremost, star allele nomenclature is not centralized and highly difficult to implement. Second, rule based decision workflows would tremendously improve CPIC guidelines for implementation.

Further development would require the integration of the remaining CPIC guidelines. Beyond that, the service should consolidate more PGx knowledge sources and thus be expanded to a greater variant coverage. Speed optimization is also planned.

4.3 Discussion of evaluation

The first observation is that all participants have actionable PGx variants, which comes as no surprise since studies have shown that many people carry actionable PGx variants [54–58] .

Second, one of the most common observed variant was rs16947 in the CYP2D6 gene, seen in 12 of the 14 (85,7%) participants. This variant has an allele frequency of 65.67% in the population (EXaC [59]). Genotype quality is 93.24 in one case and 99 in the other cases (Phred quality scores). Seven out of 12 participants carrying the variant rs16947 are homzygous for this allele, while five are heterozygous. This common variant in CYP2D6 is among others found in the CYP2D6*3 haplotype among others. Part of the result output would be the recommendation that patients with CYP2D6*3 should avoid codeine use due to the lack of efficacy.

The successful application of the workflow to the PGP data set confirmed the ability of the proposed tool to annotate VCF files with clinical guideline information from the PharmGKB and CPIC.

Chapter 5

Conclusion

In this project a workflow was built that can help clinical decision support in determing drug dosing and selecint drug after finding a genetic variant using the tools and resources MyVariant.info, PharmGKB and CPIC guidelines as well as standards like HGVS nomenclature, VCF syntax and JSON. The workflow was successfully validated on genomes from the PGP.

These are the most important conclusions during the development of this project:

1. Automated real-time web service annotation of PGx guidelines is possible
2. Assignment of haplotypes remains a significant challenge
3. Structuring CPIC guidelines is recommended
4. Web services enable direct querying from the systems of record
5. HGVS nomenclature is close but imperfect for use as a unique key

This workflow could in the future help to find clinically actionable variants and annotate them with PGx information, such as drug recommendations and haplotypes, where the variant is present. This is achieved by either editing VCF files or printing the information in JSON format. The workflow assists in the translation of PGx knowledge to clinical care. This enables clinicians to use existing and future data to personalize treatment, and identify patients at high risk of treatment failure due to excessive toxicity or inferior efficacy. It would lead to a paradigm shift in the practices of clinicians if the EHR were to include PGx test results. Instead of, "I want to prescribe drug X; to optimize outcome I should order a test for genes Y and Z to test for genetic variants" the new normal would be "I want to prescribe drug X; I should check PGx profile for this patient to see if there are genetic variants influencing drug response for this patient" [37].

References

Literature

[1] William E Evans and Mary V Relling. "Pharmacogenomics: translating functional genomics into rational therapeutics". In: *science* 286.5439 (1999), pp. 487–491 (cit. on pp. 1, 2).

[2] I. Ralph Edwards and Jeffrey K. Aronson. "Adverse drug reactions: definitions, diagnosis, and management". In: *The Lancet* 356.9237 (2000), pp. 1255–1259 (cit. on p. 1).

[3] Brian B Spear, Margo Heath-Chiozzi, and Jeffrey Huff. "Clinical application of pharmacogenetics". In: *Trends in molecular medicine* 7.5 (2001), pp. 201–204 (cit. on p. 1).

[4] Daniel S Budnitz, Daniel A Pollock, Kelly N Weidenbach, Aaron B Mendelsohn, Thomas J Schroeder, and Joseph L Annest. "National surveillance of emergency department visits for outpatient adverse drug events". In: *Jama* 296.15 (2006), pp. 1858–1866 (cit. on p. 1).

[5] Giovanni Severino and Maria Del Zompo. "Adverse drug reactions: role of pharmacogenomics". In: *Pharmacological Research* 49.4 (2004), pp. 363–373 (cit. on p. 1).

[6] Emma C Davies, Christophe F Green, David R Mottram, and Munir Pirmohamed. "Adverse drug reactions in hospitals: a narrative review". In: *Current drug safety* 2.1 (2007), pp. 79–87 (cit. on p. 1).

[7] Renee G Stark, Jürgen John, and Reiner Leidl. "Health care use and costs of adverse drug events emerging from outpatient treatment in Germany: a modelling approach". In: *BMC health services research* 11.1 (2011), p. 9 (cit. on p. 1).

[8] Francis S Collins and Harold Varmus. "A new initiative on precision medicine". In: *New England Journal of Medicine* 372.9 (2015), pp. 793–795 (cit. on p. 1).

[9] Grant R Wilkinson. "Drug metabolism and variability among patients in drug response". In: *New England Journal of Medicine* 352.21 (2005), pp. 2211–2221 (cit. on p. 1).

[10] Daniel W Nebert. “Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist?” In: *Clinical genetics* 56.4 (1999), pp. 247–258 (cit. on p. 1).

[11] W. Kalow B. K. Tang and L. Endrenyi. “Hypothesis: Comparisons of inter- and intra-individual variations can substitute for twin studies in drug research”. In: *Pharmacogenetics* 8.4 (1998), pp. 283–289 (cit. on p. 2).

[12] William E Evans and Julie A Johnson. “Pharmacogenomics: the inherited basis for interindividual differences in drug response”. In: *Annual review of genomics and human genetics* 2.1 (2001), pp. 9–39 (cit. on p. 2).

[13] Howard L McLeod and William E Evans. “Pharmacogenomics: unlocking the human genome for better drug therapy”. In: *Annual review of pharmacology and toxicology* 41.1 (2001), pp. 101–121 (cit. on p. 2).

[14] Philip Hedrick. *Genetics of populations*. Jones & Bartlett Learning, 2011 (cit. on p. 2).

[15] Christian P Schaaf and Johannes Zschocke. *Basiswissen Humangenetik*. Springer-Verlag, 2012 (cit. on p. 2).

[16] Anthony J Brookes. “The essence of SNPs”. In: *Gene* 234.2 (1999), pp. 177–186 (cit. on p. 2).

[17] International HapMap Consortium. “A haplotype map of the human genome”. In: *Nature* 437.7063 (2005), pp. 1299–1320 (cit. on p. 2).

[18] Hsiang-Yu Yuan, Jin-Jer Chen, MT Michael Lee, et al. “A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity”. In: *Human molecular genetics* 14.13 (2005), pp. 1745–1751 (cit. on p. 2).

[19] Dean Cain, Susan M Hutson, and Reidar Wallin. “Warfarin resistance is associated with a protein component of the vitamin K 2, 3-epoxide reductase enzyme complex in rat liver”. In: *Thromb Haemost* 80.1 (1998), pp. 128–133 (cit. on p. 2).

[20] Sue Richards, Nazneen Aziz, Sherri Bale, et al. “Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology”. In: *Genetics in Medicine* (2015) (cit. on p. 3).

[21] M. Wildeman, E. van Ophuizen, J. T. den Dunnen, and P. E. Taschner. “Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker”. In: *Hum Mutat* 29.1 (2008), pp. 6–13 (cit. on p. 3).

[22] Peter Tarczy-Hornoch, Laura Amendola, Samuel J Aronson, et al. “A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record”. In: *Genetics in Medicine* 15.10 (2013), pp. 824–832 (cit. on p. 3).

[23] Stephen T Sherry, Minghong Ward, and Karl Sirotkin. “dbSNP—database for single nucleotide polymorphisms and other classes of minor genetic variation”. In: *Genome research* 9.8 (1999), pp. 677–679 (cit. on p. 3).

[24] D. C. Poo, S. Cai, and J. T. Mah. “UASIS: Universal Automatic SNP Identification System”. In: *BMC Genomics* 12 Suppl 3 (2011), S9 (cit. on pp. 3, 4).

[25] Johan T Den Dunnen and Stylianos E Antonarakis. “Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion”. In: *Human mutation* 15.1 (2000), pp. 7–12 (cit. on p. 3).

[26] JD Robarge, L Li, Z Desta, A Nguyen, and DA Flockhart. “The star-allele nomenclature: retooling for translational genomics”. In: *Clinical Pharmacology & Therapeutics* 82.3 (2007), pp. 244–248 (cit. on p. 4).

[27] Ku-Lang Chang, Kristin Weitzel, and Siegfried Schmidt. “Pharmacogenetics: Using Genetic Information to Guide Drug Therapy”. In: *American family physician* 92.7 (2015) (cit. on p. 4).

[28] Joseph L Kannry and Marc S Williams. “Integration of genomics into the electronic health record: mapping terra incognita”. In: *Genetics in Medicine* 15.10 (2013), pp. 757–760 (cit. on p. 4).

[29] Mark Ebell. “Information at the point of care: answering clinical questions”. In: *The Journal of the American Board of Family Practice* 12.3 (1999), pp. 225–235 (cit. on p. 5).

[30] Josh F Peterson, Erica Bowton, Julie R Field, et al. “Electronic health record design and implementation for pharmacogenomics: a local perspective”. In: *Genetics in Medicine* 15.10 (2013), pp. 833–841 (cit. on p. 5).

[31] Gillian C Bell, Kristine R Crews, Mark R Wilkinson, et al. “Development and use of active clinical decision support for preemptive pharmacogenomics”. In: *Journal of the American Medical Informatics Association* 21.e1 (2014), pp. 93–99 (cit. on pp. 5, 14, 16).

[32] B. H. Shirts, J. S. Salama, S. J. Aronson, et al. “CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record”. In: *J Am Med Inform Assoc* (2015) (cit. on p. 5).

[33] Ellen M McDonagh, Michelle Whirl-Carrillo, Yael Garten, Russ B Altman, and Teri E Klein. "From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource". In: *Biomarkers in medicine* 5.6 (2011), pp. 795–806 (cit. on p. 5).

[34] David S Wishart, Craig Knox, An Chi Guo, et al. "DrugBank: a comprehensive resource for in silico drug discovery and exploration". In: *Nucleic acids research* 34.suppl 1 (2006), pp. 668–672 (cit. on p. 5).

[35] Kristian A Gray, Louise C Daugherty, Susan M Gordon, Ruth L Seal, Mathew W Wright, and Elspeth A Bruford. "Genenames. org: the HGNC resources in 2013". In: *Nucleic acids research* (2012), gks1066 (cit. on p. 5).

[36] Robert SA Hayward, Mark C Wilson, Sean R Tunis, Eric B Bass, Haya R Rubin, and R Brian Haynes. "More informative abstracts of articles describing clinical practice guidelines". In: *Annals of Internal Medicine* 118.9 (1993), pp. 731–737 (cit. on p. 5).

[37] M. V. Relling and T. E. Klein. "CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network". In: *Clin Pharmacol Ther* 89.3 (2011), pp. 464–7 (cit. on pp. 5, 22).

[38] Julie A Johnson, Li Gong, Michelle Whirl-Carrillo, et al. "Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing". In: *Clinical Pharmacology & Therapeutics* 90.4 (2011), pp. 625–629 (cit. on p. 5).

[39] SA Scott, K Sangkuhl, EE Gardner, et al. "Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450–2C19 (CYP2C19) genotype and clopidogrel therapy". In: *CliniCal pharmaCology & TherapeuTiCs* 90.2 (2011), pp. 328–332 (cit. on p. 5).

[40] KR Crews, A Gaedigk, HM Dunnenberger, et al. "Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype". In: *CliniCal pharmaCology & TherapeuTiCs* 91.2 (2012), pp. 321–326 (cit. on p. 5).

[41] Petr Danecek, Adam Auton, Goncalo Abecasis, et al. "The variant call format and VCFtools". In: *Bioinformatics* 27.15 (2011), pp. 2156–2158 (cit. on p. 6).

[42] Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas, and David F Lobach. "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success". In: *Bmj* 330.7494 (2005), p. 765 (cit. on p. 7).

[43] Brandon M Welch and Kensaku Kawamoto. "Clinical decision support for genetically guided personalized medicine: a systematic review". In: *Journal of the American Medical Informatics Association* 20.2 (2013), pp. 388–400 (cit. on p. 7).

[44] Maren T Scheuner, Pauline Sieverding, and Paul G Shekelle. "Delivery of genomic medicine for common chronic adult diseases: a systematic review". In: *Jama* 299.11 (2008), pp. 1320–1334 (cit. on p. 7).

[45] George M Church. "The personal genome project". In: *Molecular Systems Biology* 1.1 (2005) (cit. on p. 12).

[46] Latanya Sweeney, Akua Abu, and Julia Winn. "Identifying participants in the personal genome project by name". In: *Available at SSRN 2257732* (2013) (cit. on p. 12).

[47] Yu Liu, Jie Yang, Qiang Xu, et al. "Comparative performance of warfarin pharmacogenetic algorithms in Chinese patients". In: *Thrombosis research* 130.3 (2012), pp. 435–440 (cit. on pp. 14, 15).

[48] Jill M Pulley, Joshua C Denny, Josh F Peterson, et al. "Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project". In: *Clinical Pharmacology & Therapeutics* 92.1 (2012), pp. 87–95 (cit. on pp. 14, 16).

[49] Suzette J Bielinski, Janet E Olson, Jyotishman Pathak, et al. "Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time—using genomic data to individualize treatment protocol". In: *Mayo Clin. Proc.* 89.1 (2014), pp. 25–33 (cit. on pp. 14, 15).

[50] Barry R Goldspiel, Willy A Flegel, Gary DiPatrizio, et al. "Integrating pharmacogenetic information and clinical decision support into the electronic health record". In: *Journal of the American Medical Informatics Association* 21.3 (2014), pp. 522–528 (cit. on pp. 14, 15).

[51] Gilad J Kuperman, Anne Bobb, Thomas H Payne, et al. "Medication-related clinical decision support in computerized provider order entry systems: a review". In: *Journal of the American Medical Informatics Association* 14.1 (2007), pp. 29–40 (cit. on p. 14).

[52] Omri Gottesman, Stuart A Scott, Stephen B Ellis, et al. "The CLIPMERGE PGx program: clinical implementation of personalized medicine through electronic health records and genomics–pharmacogenomics". In: *Clinical Pharmacology & Therapeutics* 94.2 (2013), pp. 214–217 (cit. on pp. 14, 16).

[53] Monique WM Jaspers, Marian Smeulers, Hester Vermeulen, and Linda W Peute. "Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings". In: *Journal of the American Medical Informatics Association* 18.3 (2011), pp. 327–334 (cit. on p. 19).

[54] Jacob N Beverage, Tristan M Sissung, Amy M Sion, Romano Danesi, and William D Figg. "CYP2D6 polymorphisms and the impact on tamoxifen therapy". In: *Journal of pharmaceutical sciences* 96.9 (2007), pp. 2224–2231 (cit. on p. 20).

[55] E Link, S Parish, J Armitage, et al. "SLCO1B1 variants and statin-induced myopathy–a genomewide study". In: *The New England journal of medicine* 359.8 (2008), pp. 789–799 (cit. on p. 20).

[56] Jean-Philippe Collet, Jean-Sébastien Hulot, Anna Pena, et al. "Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: a cohort study". In: *The Lancet* 373.9660 (2009), pp. 309–317 (cit. on p. 20).

[57] Alan R Shuldiner, Jeffrey R O'Connell, Kevin P Bliden, et al. "Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy". In: *Jama* 302.8 (2009), pp. 849–857 (cit. on p. 20).

[58] Robert S Epstein, Thomas P Moyer, Ronald E Aubert, et al. "Warfarin genotyping reduces hospitalization rates: results from the MM-WES (Medco-Mayo Warfarin Effectiveness Study)". In: *Journal of the American College of Cardiology* 55.25 (2010), pp. 2804–2812 (cit. on p. 20).

[59] Monkol Lek, Konrad Karczewski, Eric Minikel, et al. "Analysis of protein-coding genetic variation in 60,706 humans". In: *bioRxiv* (2015), p. 030338 (cit. on p. 21).