DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
372

KRISTI KREBS

Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
372

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
372

KRISTI KREBS

Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records

UNIVERSITY OF TARTU
Press

Institute of Molecular and Cell Biology, University of Tartu, Estonia

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Molecular Biomedicine on 25.06.2020 by the Council of the Institute of Molecular Cell Biology, University of Tartu.

Supervisors: Prof. Lili Milani, PhD
Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Prof. Andres Metspalu, MD, PhD
Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia
Chair of Biotechnology, Institute of Molecular Cell Biology, University of Tartu, Tartu, Estonia

Reviewer: Prof. Maido Remm, PhD
Chair of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Opponent: Prof. Sir Munir Pirmohamed, PhD
Institute of Systems, Molecular and Integrative Biology, The University of Liverpool, Liverpool, United Kingdom

Commencement: Room No. 105, 23B Riia St., Tartu, on 27. 08. 2020 at 11:15 am.

The publication of this dissertation is granted by the Institute of Molecular and Cell Biology at the University of Tartu.

This research was funded by EU H2020 grant 692145, Archimedes Foundation student mobility program Dora Plus T1.2, Estonian Research Council Grants IUT20-60, IUT24-6: Estonian Centre for Genomics, PRG 184, ETF9293 and European Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012 GENTRANSMED, Project No. 2014-2020.4.01.16-0125 and the scholarships in smart specialisation growth areas; the Graduate School in Biomedicine and Biotechnology ASTRA project of the University of Tartu, PER ASPERA, Broad Institute project MP1GV16307 and Uppsala University Strategic Research Grant Science for Life Laboratory fellowship program. Data analyses were carried out in part in the High-Performance Computing Center of the University of Tartu.

ISSN 1024-6479
ISBN 978-9949-03-385-0 (print)
ISBN 978-9949-03-388-1 (pdf)

University of Tartu Press
www.tyk.ee

To my dear family, colleagues
and fellow PhD students ♥

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, referred to in the text by Roman numerals (Ref. I to Ref. IV):

I Tasa, T.*, **Krebs, K**.*, Kals, M., Mägi, R., Lauschke, V.M., Haller, T., Puurand, T., Remm, M., Esko, T., Metspalu, A., Vilo, J., Milani, L. (2019). Genetic variation in the Estonian population: pharmacogenomics study of adverse drug effects using electronic health records. Eur. J. Hum. Genet. 27, 442–454.

II **Krebs, K.*,** Bovijn, J.*, Lepamets, M., Censin, J.C., Jürgenson, T., Särg, D., Luo, Y., Skotte, L., Geller, F., Feenstra, B., et al. (2020). Genome-wide study identifies association between HLA-B*55:01 and penicillin allergy. BioRxiv 2020.02.27.967497.

III Bovijn, J.*, **Krebs, K.***, Chen, C.-Y.*, Boxall, R.*, Censin, J.C., Ferreira, T., Pulit, S.L., Glastonbury, C.A., Laber, S., Millwood, I.Y., et al. (2020). Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics. Sci. Transl. Med. 12, 549.

IV Reisberg, S., **Krebs, K.**, Lepamets, M., Kals, M., Mägi, R., Metsalu, K., Lauschke, V.M., Vilo, J., and Milani, L. (2019). Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. Genet. Med. *21*, 1345–1354.

***** These authors contributed equally

My contributions to the listed publications were as follows:

Ref. I Participated in the study design, data analysis, interpretation of results, prepared figures and tables, drafted the manuscript, and participated in its revision.

Ref. II Participated in the study design, performed most of the analyses, interpretation of the results, prepared the figures and tables, wrote the first manuscript and participated in its revision.

Ref. III Performed analyses in Estonian Biobank and revised the manuscript.

Ref. IV Participated in the study design, coordinated the design of the analysis pipeline and interpretation of the results. Participated in writing and revising the manuscript.

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
ADE	Adverse event
ADME	Absorption, distribution, metabolism and elimination
ADR	Adverse drug reactions
ATC	Anatomical Therapeutic Chemical
BMD	Bone mineral density
CAD	Coronary artery disease
CDS	Clinical decision support
CE	HumanCoreExome array
CPIC	Clinical Pharmacogenetics Implementation Consortium
CYP	Cytochrome P-450
DDD	Defined daily doses
DPWG	Dutch Pharmacogenetics Working Group
EHR	Electronic health records
eQTL	Expression quantitative trait loci
ES	Exome sequencing
EstBB	Estonian Biobank
GS	Genome sequencing
GSA	Global screening array
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
ICD	International Classification of Disease
IM	Intermediate metabolizers
LoF	Loss of function
MAF	Minor allele frequency
MHC	Major histocompatibility complex
NLP	Natural language processing
OMNI	HumanOmniExpress array
OR	Odds ratio
PC	Principal components
PGx	Pharmacogenomics
PharmGKB	The Pharmacogenomics Knowledge Base
PM	Poor metabolizers
PRS	Polygenic risk scores
SLC	Solute carrier
SNP	Single nucleotide polymorphism
UKBB	UK Biobank
UM	Ultrarapid metabolizers
VIP	Very important pharmacogenes

INTRODUCTION

Genetic variation between individuals can affect the efficacy and safety of a drug. The field of pharmacogenomics (PGx) aims to determine these genetic variants and poses as a cornerstone of personalized medicine providing guidance for drug treatment based on the genetic factors influencing drug response. The first PGx candidate gene studies started more than 30 years ago, but the sequencing of the human genome, advances in next generation sequencing technologies and the wave of genome wide association studies started a new era of PGx research in the past decade.

Several studies of drug efficacy have indicated the role of genetic variants, but there is less research on the genetic factors influencing the incidence of adverse drug reactions. One of the reasons for that is the rare nature of these reactions that makes the collection of samples more difficult. Further, the issue with small sample sizes in many pharmacogenomics studies also makes the replication of discovered associations more challenging. The combination of electronic health records (EHRs) with genetic data has brought the research of complex diseases into the next stage by providing a way to gather sufficient sample sizes to identify novel genetic associations with a plethora of phenotypes extracted from EHRs. Population-based biobanks can serve as the potential untapped resource for different routes of PGx research as well.

This thesis focuses on characterizing the value and results of using EHRs linked with genetic information of Estonian biobank (EstBB) participants for different pharmacogenomics research directions. In the first part of the thesis I will cover the elements of pharmacogenomics based on the literature: The onset of drug response and different factors influencing it; describe adverse drug reactions; the wave of pharmacogenomic research and the potential of electronic health records; the translation of pharmacogenomics into the clinic and future outlooks for unlocking the full potential of genetics for a more guided pharmacotherapy. In the second part, I will describe the first proof of concept search for genetic associations with adverse drug reactions among individuals with prescriptions of specific drugs by using the information embedded in EHRs and linked with population-based Biobanks. Further, I will demonstrate the potential of human genetics for the study of drug target mediated adverse events and finally, I will discuss the translation of genetic data into pharmacogenomic recommendations.

1. REVIEW OF THE LITERATURE

1.1. Drug response – pharmacokinetics- and dynamics

The main focus of pharmacogenomics (PGx) is to determine the genetic variants that influence drug efficacy, drug safety, or both and to use this information for a more guided pharmacotherapy, aiming to improve the clinical outcome of treatments (Lonergan et al. 2017). After administration, the drug is absorbed and distributed to its site of action where it interacts with the target molecules, receptors, and as a final step, undergoes metabolism followed by elimination (Figure 1). Genetic variants that influence any of these processes by altering drug pharmacokinetics – drug absorption, distribution, metabolism and elimination (ADME) – or pharmacodynamics – modifying targets – may pose a clinically significant effect on the treatment outcome (Weinshilboum 2003; Mary V. Relling and Evans 2015).

1.1.1. Pharmacokinetics – what the body does to a drug

The first studies in PGx mainly focused on uncovering relevant variants in drug metabolism (Weinshilboum 2003). Usually the objective of drug metabolism is to convert drugs into metabolites that are more water soluble and, thus can be more easily eliminated. An exception of this is the metabolism of prodrugs – partially inactive modified drug molecules (Testa 2004), that are first converted through metabolism into therapeutically active compounds (Buxton and Benet 2015). The chemical processes that are relevant in metabolism pathways are generally divided into two, phase I reactions (e.g oxidation, hydrolysis, reduction) and phase II reactions (like glucuronidation, sulfation, acetylation, methylation), that all involve different families of drug metabolizing enzymes (Weinshilboum 2003; Wilkinson 2005) The major source of variability in drug pharmacokinetics are the polymorphic cytochrome P-450 (CYP) enzymes, the most important family of enzymes catalyzing phase I reactions (Nebert and Russell 2002; Zanger and Schwab 2013). Humans are known to have 57 putatively functional CYP enzymes (D. R. Nelson et al. 2004) of which around a dozen are involved in the biotransformation of most drugs and other lipophilic xenobiotics (Zanger and Schwab 2013). It has been estimated that six of the CYP enzymes are jointly accountable for 80–90% of the metabolism of all drugs (Wilkinson 2005; Evans and Relling 1999; Lynch and Price 2007). Although CYP enzymes have a substrate specificity to a certain region of a drug molecule, there may also be a considerable overlap. This means that for a given drug, an individual CYP may be the enzyme primarily responsible for the metabolism, or there may be a contribution of numerous CYP enzymes (Wilkinson 2005). Between different populations there is a large variability in the distribution and frequency of genetic variants in genes encoding CYP enzymes (Zanger and Schwab 2013). If the metabolism of a given drug is predominantly dependent on a single CYP, a polymorphism that is impairing its functionality may have a serious effect on the clinical outcomes

(Wilkinson 2005). In the case of multiple pathways by which drugs are metabolised, abolished functionality of one CYP due to different polymorphisms is likely to exert a minor effect on drug response (Roden et al. 2011).

In general, pharmacokinetic phenotypes of drug metabolizing enzymes can be divided into four categories (Figure 1):

1) Poor metabolizers (PM) are carriers of homozygous or compound heterozygous alleles that are lacking functionality;
2) Extensive or normal metabolizers are carriers of alleles conferring the "normal" phenotype, in most cases referring to the majority of the population;
3) Intermediate metabolizers (IM) are carriers of one normal and one non-functional or decreased function allele, or carriers of two alleles with decreased function;
4) Ultrarapid metabolizers (UM) are carriers of gain-of-function alleles (Zanger and Schwab 2013).

When considering these phenotypes in the clinical context, variants decreasing or abolishing function might reduce the elimination of a drug and thus as a consequence increase its plasma concentrations, while variants leading to gain in function might increase drug elimination and therefore diminish drug concentrations. In the case of an active drug, increased concentrations may lead to drug toxicity resulting in adverse drug reactions (ADR), while decreased concentrations often lead to low or absent drug efficacy. This situation is especially relevant in case of drugs with a narrow therapeutic range, i.e. the margin between therapeutic and toxic doses (Figure 1). In the case of a metabolically activated prodrug, one can expect a contrary effect compared to an active drug (Zanger and Schwab 2013) (Figure 1).

Drug transporters are relevant determinants of drug distribution and elimination (Petzinger and Geyer 2006). For a desired drug effect, it is necessary that the drug would reach its target at an adequate concentration, and this can be resolved by the balanced activity of multiple uptake and efflux transporters. Uptake transporters enable the translocation of drugs into cells, and efflux transporters facilitate their export from the intracellular to extracellular environment, which is often carried out against high concentration gradients (Ho and Kim 2005). Transporters are expressed in tissues throughout the body as membrane-bound proteins and for a drug response the ones expressed in liver, intestine and kidney are particularly of interest due to their relevance in drug disposition (DeGorter et al. 2011). Two superfamilies of transporters with relevant roles in drug efficacy and toxicity are solute carrier (SLC) uptake transporters and ATP-binding cassette (ABC) efflux transporters. Members of the SLC family translocate their substrates across membranes through various mechanisms, while members of the ABC superfamily use energy from the hydrolysis of ATP for transport (Ho and Kim 2005; DeGorter et al. 2011). Similarly, to drug metabolism, several studies have also indicated the relevance of different genetic polymorphism in transporter genes for the observed therapeutic efficacy or toxicities of certain drugs (Yee, Chen, and Giacomini 2010; Schaller and Lauschke 2019).

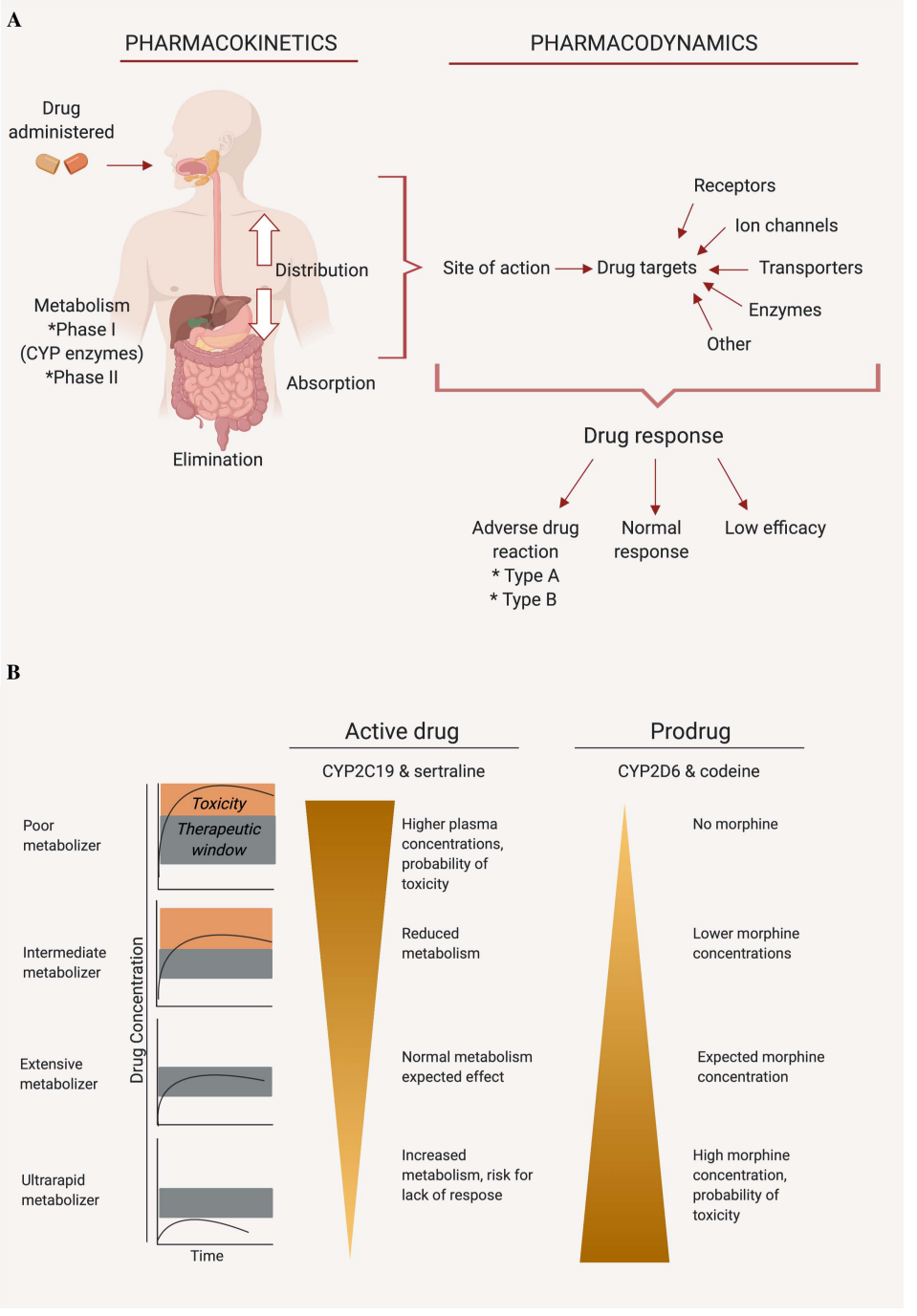

Figure 1. Drug response. (A) Illustration of drug pharmacokinetics and pharmacodynamics (B) Pharmacokinetic phenotypes of drug metabolizing enzymes. Examples of the contrary response phenotypes in the cases of active and prodrugs.

One of the most prominent examples is studies of simvastatin pharmacokinetics where several works have demonstrated the increased risk for myopathy caused by polymorphisms reducing Solute Carrier Organic Anion Transporter Family Member 1B1 (SLCO1B1) function (Pasanen et al. 2006; Meade et al. 2008).

1.1.2. Pharmacodynamics – what a drug does to the body

In addition to pharmacokinetic sources of variation, genetic variability in genes encoding drug effector molecules may have a major effect on drug efficacy and toxicity (Evans and Relling 1999). In order to have therapeutic action, drugs may target and act on ion channels, receptors, enzymes and nucleic acids. Polymorphisms that modify or inactivate these targets may influence drug therapeutic action. For example, in case of warfarin treatment, polymorphisms in the gene encoding vitamin K epoxide reductase complex 1 (VKORC1) may affect the variability of the necessary warfarin dose up to 25% (Rieder et al. 2005). Another example is ivacaftor treatment for cystic fibrosis where treatment is not recommended in case of specific deletions in Cystic fibrosis transmembrane conductance regulator gene (*CFTR*) (Clancy et al. 2014). Germline genetic variants in drug target genes is also relevant for the success of cancer treatment. Functional variants in drug targets carry over to the cancer genome and thus influence the treatment outcome (Schärfe et al. 2017).

Importantly, genetic variability of drug targets has a beneficial side for drug development by guiding the prediction of efficacy and adverse effects as well as for understanding the associative relationship of a target and outcome. There are several examples of when genetic associations studies have led to the identification of potential drug targets and development of drugs (Plenge, Scolnick, and Altshuler 2013). One of the most prominent examples is Proprotein convertase subtilisin/kexin type 9 (*PCSK9*). Gain of function variants in the gene were first associated with elevated low-density lipoproteins (LDL) levels (Abifadel et al. 2003), and later another study demonstrated that loss of function variants reduce LDL levels and thereby lifetime risk for coronary artery disease (CAD)(Cohen et al. 2006). This led to the development of PCSK9 inhibitors for the treatment of elevated LDL levels and prevention of CAD. In addition to first providing insights for drug development, genetic variants have the potential to shorten the entire process and increase the probability for the success of a drug candidate. Studies have indicated that 50% of drugs fail due to efficacy during phase II of the drug development process and 25% due to toxicity (Arrowsmith 2011; Harrison 2016). If the aspects for these failures would be known before the developmental phases, there is a higher potential for trials success. Human genetics can be used to estimate the on-target side effects associated with the drug effect. As in case of *PCSK9*, naturally occurring loss of function variants of the gene can add insight as to what phenotypes occur with the complete inhibition of PCSK9 (Plenge, Scolnick, and Altshuler 2013). Strong evidence has already emerged that drug targets with support by human genetic studies are more likely

to be successful when compared to those without such genetic validation (M. R. Nelson et al. 2015; King, Wade Davis, and Degner 2019).

PGx studies enable the prediction of pharmacodynamics effects in the early stages of drug development processes, which can be complemented with data from cohorts with retrospective information (Liou, Stringer, and Hirayama 2012). Considering this, individual level clinical data followed over long periods of time together with genetic information, e.g. population-based biobanks coupled with electronic health records, serve as a valuable resource for studies on drug targets. Once known genetic variants in drug targets have been proven to be safe, these and previously uncharacterized genetic variants may still influence drug response. This is the area that pharmacodynamics studies focus on.

1.2. Adverse drug reactions

Unfortunately, besides the therapeutic action of drugs, there is always a risk for unwanted adverse drug reactions (ADRs). A meta-analysis of prospective studies in the US indicated the incidence of serious ADRs to be 6.7% among hospitalized patients and the cause of more than 100,000 deaths annually (Lazarou, Pomeranz, and Corey 2003). In Europe it has been estimated that 3.5% of hospital admissions are due to ADRs and 10.1% of patients experienced ADRs during hospitalization (Bouvy, De Bruin, and Koopmanschap 2015). Evaluation of the direct costs of ADRs range from €702.21 to €40,273.08 in ambulatory care, and from €943.40 to €7,192.36 in the hospital (Marques et al. 2016). It is clear that the development of efficient prediction methods of ADRs is a crucial step for better healthcare.

The European Medicines Agency defines an adverse drug reaction as an unintended and noxious response to a medicinal product (Medicines Agency 2017). When the clear causality of the adverse outcome is not yet linked to a specific drug, the term ‘adverse event’ (ADE) should be used instead, acknowledging this way that that respective adverse outcome is not necessarily always drug related (Edwards and Aronson 2000). ADRs are typically classified into two types. Type A reactions are more predictable, based on the pharmacological action of a drug by change in drug pharmacokinetics, -dynamics and indicating also a clear dose–response relationship. These types of reactions are, for example, bleeding due to warfarin or higher risk for toxicity due to impaired metabolism of amitriptyline by cytochrome P450 family 2 subfamily D member 6 (CYP2D6) (Osanlou, Pirmohamed, and Daly 2018). Type B reactions are less predictable, considered dose-independent and caused by allergic or non-allergic mechanisms additionally involving the immune system of the host (Böhm and Cascorbi 2016; Iasella, Johnson, and Dunn 2017; Ana Alfirevic and Pirmohamed 2017). Although type B ADRs are less frequent than type A reactions, they are commonly more serious than and more frequently lead to the withdrawal of drugs from the market (Iasella, Johnson, and Dunn 2017). Type B reactions are often missed in clinical trials and

sometimes discovered yeas later during post-marketing surveillance (Charlton and Thompson 2017).

Genetic variants have a role in the occurrence of both types of ADRs and the degree of the contribution depends on a specific drug and patient. One of the challenges in predicting ADRs is that the mechanism of their occurrence is often unknown (Charlton and Thompson 2017). This complexity of not knowing the exact targets and pharmacokinetic aspects adds further difficulty in estimating the contribution of genetic factors in susceptibility to ADRs (Ana Alfirevic and Pirmohamed 2017). Thus, further pharmacogenetic studies are needed not only to validate and replicate previously reported genetic association with ADRs or to find clinically actionable variants, but also to better understand the mechanisms of ADRs.

1.1.2. Drug induced hypersensitivity

Since the majority of type B reactions involve the immune system, they are also named hypersensitivity reactions (Negrini and Becquemont 2017). Type B reactions are caused by allergic or non-allergic mechanisms. Typically, hypersensitivity reactions are further divided into four types, and as non-allergic reaction cannot be fitted into these categories, they are described separately. The mechanism of non-allergic hypersensitivity, formerly also known as pseudo-allergy, is not well understood, but the involvement of complement activation or direct effects on mast cells have been proposed (Zhang et al. 2018). Based on the timing of their onset, allergic hypersensitivity reactions are also divided into immediate (type I) and delayed reactions (type IV). Type II and III are uncommon, involve IgG and IgM antibodies, and their time of onset is variable – 1 to 3 weeks after drug exposure (Riedl and Casillas 2003). Type I reactions are IgE-mediated, symptoms usually occur within an hour after drug exposure. Delayed type IV hypersensitivity reactions can occur days or weeks after drug exposure and are T-cell mediated through three proposed immunopathogenic models: the hapten/prohapten model, pharmacological interaction (p-i) and altered peptide repertoire hypothesis(Negrini and Becquemont 2017; Pavlos et al. 2015). These theories are not mutually exclusive and mechanism valid for a given drug, may not be true for another. In case of all of these three mechanism, type IV ADRs indicate a strong association with a plethora of the human leukocyte antigen (HLA) alleles (Böhm and Cascorbi 2016) (Table1).

Table1. Well-defined association of HLA-drug hypersensitivity reactions.

Drug	HLA allele	Reaction	OR
Abacavir	B∗ 57:01	HSR	>950
Allopurinol	B∗ 58:01	SJS/TEN and DRESS/DIHS	>800
Carbamazepine	B∗ 15:02	SJS/TEN	>1000
Dapsone	B∗ 13:01	DRESS/DIHS	20
Flucloxacillin	B∗ 57:01	DILI	81

Abbreviations: DIHS, drug-induced hypersensitivity syndrome; DILI, drug-induced liver disease; DRESS, drug reaction with eosinophilia and systemic symptoms; HLA, human leukocyte antigen; HSR, hypersensitivity reaction; OR, odds ratio. Table adapted from Pavlos et al. (2015).

The HLA system, also named the human major histocompatibility complex (MHC), is responsible for T-cell stimulation to create an immune response, and is the most polymorphic region in the human genome (Vandiedonck and Knight 2009). This variability determines the different shape of the peptide-binding pocket in HLA molecules, which in turn ensures the huge repertoire of peptides that can bind to a specific HLA molecule and thereafter can be presented to a T-cell, a key step in adaptive immune response (Negrini and Becquemont 2017). However, this variety may additionally increase the risk for off-target binding of small drugs or haptens by HLA molecules, that can generate immune system-mediated ADRs (Illing, Purcell, and McCluskey 2017). The exact mechanism of most of the HLA-drug interactions is not understood yet, except for association between abacavir hypersensitivity syndrome and HLAB*57:01 that is resolved at a mechanistic level (Negrini and Becquemont 2017; Illing, Purcell, and McCluskey 2017). Furthermore, drugs can induce hypersensitivity through several mechanism, for example, penicillin can cause both immediate and delayed types of ADRs (Blanca et al. 2009). Despite the increasing evidence and studies showing the role of the HLA system in drug induced hypersensitivity, much remains to be learned by conducting further studies, and HLA molecules are only part of the story.

1.3. Pharmacogenomic research

Studies of associations between genetic variants and drug response are crucial for highlighting actionable variants that could one day be clinically implemented to improve treatment outcome. The recent decade has provided us with a number of associations ready for implementation into the clinic and a plethora of associations that require further validation studies.

PGx research started out with candidate gene studies where a genetic variant was linked to the studied phenotype in a group of subjects (Maggo, Savage, and Kennedy 2016). These types of studies were “educated guesses” where the presumed association was based on the known metabolism pathway or function

of a certain receptor, enzyme etc. Genes encoding these relevant enzymes or receptors were then examined for variants that may contribute to the variability in drug response (Collins, Carr, and Pirmohamed 2016). Although in general the clinical implementation of these associations from candidate gene studies has been slow, there are still various examples already proven to be beneficial in clinical practice, and these studies also laid a solid foundation for further PGx research (Collins, Carr, and Pirmohamed 2016; Daly 2010). Limitations apply for a candidate gene approach in a sense that it requires a degree of insight into the mechanism of a studied drug, thus prior pharmacological research would be relevant (Collins, Carr, and Pirmohamed 2016). The launch of genome-wide association studies (GWAS) helped to bring PGx research into a new era with the hypothesis-free approach broadening the horizon to contributions of less obvious genes and leading to the discoveries of novel relevant mechanisms.

In a GWAS, hundreds of thousands to millions of genetic variants across the genome of a number of individuals with a given phenotype is tested against controls to identify genotype-phenotype associations. Over the past decade, GWASes have provided a range of discoveries for human complex traits and diseases, demonstrating the important role of genetics in these phenotypes (Bush and Moore 2012; Visscher et al. 2017; Tam et al. 2019). Due to the challenges in pharmacogenomic GWAS, only 5% of all the GWASes represent drug response phenotypes (Osanlou, Pirmohamed, and Daly 2018). However, since 2007 these published PGx GWASes have demonstrated the benefit of using a genome-wide approach for the study of drug response by revealing several novel drug-phenotype associations. A survey of the published literature reported that around 70% of all the PGx GWASes are studies of drug response and the remaining are of ADRs (Daly 2010). The main reason for this is the rare nature of ADR phenotypes, which makes the sample collection for well-powered GWAS more difficult. However, with the first GWASes of both ADRs and drug response in general, it became clear, that even with a small sample size, one can see large genetic effects driven by few risk alleles in PGx studies (SEARCH Collaborative Group et al. 2008; Daly et al. 2009). With some exceptions, usually in case of common traits, small to moderate effects are detected in GWASes and the odds ratios (OR) for the effect are typically < 1.5 (K. Zhou and Pearson 2012). The effect sizes of PGx phenotypes are much greater and therefore even with small sample sizes novel associations have been identified (Daly et al. 2009; SEARCH Collaborative Group et al. 2008; Harper and Topol 2012). Overall, large effect sizes are seven times more frequent in PGx GWAS when compared with human disease risk or complex traits (K. M. Giacomini et al. 2012). A comparison of allele frequency (AF) distributions between common diseases and drug response also indicated that an excess of low AF variants exist in the signals of drug response (K. Zhou and Pearson 2012). Thus, the studies of rare ADRs are indicating a plausibility that genetic contribution to rare ADRs is more similar to monogenic than complex diseases. However, PGx GWASes with larger sample sizes have also identified variants with moderate effects, thereby indicating that drug response phenotypes are similar to common diseases (Daly 2010; K. Zhou and Pearson 2012). For

example, the existing knowledge of warfarin response and the role of variants in the *VKORC1* and (Cytochrome P450 Family 2 Subfamily C Member 9 (*CYP2C9*) genes (Cooper et al. 2019) was broadened with the role of an additional gene Cytochrome P450 Family 4 Subfamily F Member 2 (*CYP4F2*) by gaining a larger sample size for the GWAS (Takeuchi et al. 2009). Evidently, common variants with small effects have a contributing role in the variation of drug response, but to properly study the influence of rare variants in a given population, further studies using whole genome sequencing is needed. Thus, performing well-powered GWAS in large patient cohorts with the potential to reveal the complete spectrum of alleles is necessary to elucidate the extent of genetic variants that is relevant in drug response.

Since the biological mechanism of treatment outcome of many commonly used drugs is poorly understood (K. Zhou and Pearson 2012), studies performed without the advantage of hypothesis-free design, has been a limiting factor for studies of pharmacodynamics. The era of GWAS also enabled the detection of new associations of drug target genetics (Cao and Moult 2014). For example, the contribution of genetic factors in the occurrence of drug induced prolongation of QT intervals was further examined by GWAS studies and new candidate genes were identified (Niemeijer et al. 2015). New pharmacodynamics genetic variants have been identified for cisplatin-induced deafness (Xu et al. 2015) and variants in HLA genes associated with type B ADRs provide a number of examples of pharmacodynamics variants (Pavlos, Mallal, and Phillips 2012; A Alfirevic and Pirmohamed 2009; Becquemont 2010). Further, Metformin has been used to decrease blood glucose levels in patients with type 2 diabetes for some time, and recently a GWAS of 3,920 patients clarified the genetic basis for its mechanism (GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group et al. 2011). In addition to holding promise in the context of drug development as means to identify novel drug targets, GWAS efforts can retrospectively help to identify the genetic basis for drugs already in use. GWASes can also be used for drug repurposing, finding new indications for an already existing therapeutic drug and thereby potentially speeding up the approval and marketing process of new drugs (Robinson et al. 2018).

One of the main challenges in applying GWAS to PGx has been the collection of cases due to the rare nature of ADR phenotypes, especially in case of very rare ADRs (Maggo, Savage, and Kennedy 2016). Additionally, there is also evidence of significant global under-reporting of ADRs, including severe ADRs (Hazell and Shakir 2006). Further, since there are a number of drugs available for the treatment of most common diseases, it is sometimes difficult to obtain adequate numbers of patients who received the same drug (Kathleen M. Giacomini et al. 2017). Gathering a substantial number of well phenotyped samples is an obstacle for PGx GWAS and for further research of relevant variants in the occurrence of ADRs it is necessary to find new approaches for identifying and collecting data on ADR phenotypes.

1.3.1. The value of electronic health records

Electronic health records (EHRs) have been growingly appreciated for genomic studies of diseases and have provided several opportunities for in-depth research.

EHRs are real-time longitudinal records of the health information of individuals, which is generated at the point of care by healthcare providers. Systematic collection of patient information enables sharing and accessing it across healthcare systems to obtain more comprehensive clinical care (Abul-Husn and Kenny 2019). The information that is stored in the EHR is a combination of structured and unstructured data. Structured data uses a uniform format for recording information on the patient's diagnoses, dispensed medications, procedures and additional analyses of biochemistry, which can all be made available through EHRs, thus enable us to construct a detailed clinical picture of each individual (Abul-Husn and Kenny 2019; Kohane 2011). An important aspect of structured data is the use of controlled vocabularies like International Classification of Disease codes (ICD) for patient diagnoses, procedures, complications and The Anatomical Therapeutic Chemical (ATC) Classification System for classification of drugs. In contrast, unstructured data does not follow any particular format and enables the healthcare providers to enter all health information as free text (Abul-Husn and Kenny 2019). While structured data is consistent, unstructured data requires specific tools like natural language processing (NLP) for extraction of information (Pendergrass and Crawford 2018).

Nationwide biobanks are emerging in several countries and provide a rich resource for discovery studies and can also be stepping stones for translation of genomics into clinical practice (Abul-Husn and Kenny 2019). Linking the aforementioned comprehensive medical records with the biological material of individuals poses as a powerful resource for further genomic research. Since the EHRs supply longitudinal information on medication exposures and diagnoses, coupling these with the genotype data of biobanks also makes it an important platform for the study of drug effects (Robinson et al. 2018). One of the key advantages of EHRs over clinical trials is the opportunity to collect data on a large number of individuals in a timely manner, not to mention at a low expense and faster completion (Robinson et al. 2018). This is particularly relevant in PGx research where the rapid collection of ADR cases for clinical trials can be limited, especially for cases of rare ADRs. However, these can be systematically and retrospectively retrieved from large population-based EHRs (K. Zhou and Pearson 2012). One of the possible solutions for the identification of ADR is to use the ICD-10 coded diagnoses, that are routinely used in the hospital (Hodgkinson, Dirnbauer, and Larmour 2009; Stausberg and Hasford 2010). Furthermore, through EHRs, a great deal of information in free-text clinical notes can be made available with NLP algorithms, thus providing a further source to obtain information on ADRs. In PGx studies, the use of EHRs also enables to control for polypharmacy that is not always captured in clinical trials (K. Zhou and Pearson 2012). In addition to the use of EHRs for the PGx discovery studies, it has the

potential to be easily used for the validation of previously documented PGx associations across different health systems, populations, and clinical contexts (Abul-Husn and Kenny 2019; Wilke et al. 2011; D M Roden et al. 2012).

EHRs linked with biorepositories also have limitations that need to be accounted for and understood in research. Since the primary purpose of EHR data is clinical use and billing, not research, there may be particular challenges of inaccuracy and missingness (Hersh et al. 2013). The quantity of available information can vary greatly, and there can be some inaccuracies caused by clinical uncertainty or billing errors (Robinson et al. 2018). For example, clinicians bill for the evaluation of a diagnosis, and this is recorded, even if the evaluation does not reveal the disease (Kohane 2011). Furthermore, in countries that lack national health identification systems, it can be a challenge to combine EHR data for large population studies (Kohane 2011). Thus, data may be partial due to the various healthcare providers and centers a patient may visit (Robinson et al. 2018).

In general, however, EHRs combined with genotype information has introduced a new wave in genetic studies of diseases and provide sufficient sample sizes for association studies, despite the noise they contain. Furthermore, EHRs are also a key resource for the implementation of genomic data into healthcare. Integrating clinical decision support (CDS) tools together with EHRs have the ability to support other care-related activities to provide necessary knowledge at appropriate times to improve the quality of clinical care (Abul-Husn and Kenny 2019).

1.4. From pharmacogenomic associations to treatment recommendations

The main intention of comprehensive PGx studies is to find actionable genetic information that can be used in everyday clinical practice to guide the drug treatment of patients. Despite several decades of PGx research, implementation into the clinic has been lagging behind due to several barriers. Nevertheless, within recent years the number of implementation initiatives has increased and for overcoming the obstacles, these efforts have given rise to constantly improving solutions and resources for more straightforward PGx implementation (Krebs and Milani 2019).

One of the challenges on the path towards implementation of PGx has been the question of platform choice for genetic testing that further includes the choice of which variants or genes to test. Different implementation initiatives have selected among various PGx arrays that enable simultaneous analysis of a number of SNPs in several genes (Arbitrio et al. 2016; J. M. Pulley et al. 2012; van der Wouden et al. 2017). Limitations such as capturing newly identified but potentially clinically relevant alleles, or the case of differences in the designs of assays that might pose some difficulties for the comparison of results from several assays (Kalman et al. 2016), have tilted the decision of some initiatives towards more comprehensive approaches. Despite the fact that the cost of whole-genome sequencing is continuing to decline, it still remains too expensive for wider

clinical use and there are currently other barriers as well (Krebs and Milani 2019). One possible solutions is the use of capture libraries to sequence only the genes of interest, e.g. all the currently known relevant pharmacogenes, thereby balancing cost, throughput, and deep coverage(Gordon et al. 2016; Rasmussen-Torvik et al. 2014). Genome-wide genotyping arrays also have the potential as a method that balances well between comprehensiveness and cost. Using genotyping together with phasing and imputation, the genotypes for the relevant alleles can be estimated. When considering the precision of imputation, a higher accuracy can be achieved with a population-specific reference panel for imputation (Mitt et al. 2017). The comparison of the different opportunities and barriers in the use of genome-wide arrays against sequencing-based methods for the detection of currently defined pharmacogenetically relevant alleles is valuable since it would be a highly cost-effective tool for identifying individuals who need altered dosing recommendations.

1.4.1. Opportunities for translation of genetic data into recommendations

With the first implementation initiatives of pharmacogenetics, several challenges were revealed in the translation of PGx test results into clinical action. This resulted in the development of several resources to aid the translation of acquired information on pharmacogenetic genotypes into treatment recommendations.

As a solution to the challenge of interpreting genetic test results into clinical action, two consortia, the Dutch Pharmacogenetics Working Group (DPWG) (Jesse J. Swen et al. 2018; J J Swen et al. 2011) and the Clinical Pharmacogenetics Implementation Consortium (CPIC) (M V Relling and Klein 2009) developed a list of therapeutic recommendations to help facilitate the translation of pharmacogenetics into clinical care. These drug guidelines offer direct guidance to clinicians regarding the dosing recommendations or options for alternative medication based on drug-gene pairs that are selected through careful curation and have a clear evidence-based impact on the outcome of pharmacotherapy. Consequently, these guidelines additionally solve the question of which pharmacogenes that are relevant for testing.

As a next challenge after having available PGx guidelines is the question of translating genotype data at hand into phenotype information or prescription recommendations more specifically. The Pharmacogenomics Knowledge Base (PharmGKB) was created 20 years ago with the purpose of collecting and curating knowledge about the impact of human genetic variants on drug response (Whirl-Carrillo et al. 2012; Barbarino et al. 2018). Through well-defined criteria that are based on the careful curation of the literature, variant–drug associations are assigned with a level of evidence. Under the summaries of ‘very important pharmacogenes’ (VIP), the database lists all the critical genes in drug response accompanied with detailed overviews (Whirl-Carrillo et al. 2012). Further, PharmGKB and CPIC provide specific translation tables on how to define pharmacogenetic

alleles on the basis of genetic variants for gene-drug pairs covered in these guidelines (Caudle et al. 2014). They additionally provide knowledge on how to assign diplotypes against the interpreted phenotypes. However, some challenges still remain concerning the assignment of actionable alleles when multiple variants, e.g. gain and loss of function, occur on the same allele. Also, the PGx nomenclature can be confusing.

To determine diplotypes based on microarray and sequencing data knowing the correct phase is crucial for actionable pharmacogenetic alleles that are assigned by more than one variant. Computational phasing of genotypes provides the easiest and fastest solution for identifying variants co-located on the same chromosome and several well-known algorithms have been designed for that purpose (Browning and Browning 2011; Choi et al. 2018).

The translation tables for interpretation that are provided by PharmGKB are based on the star (*) allele nomenclature – the most widely used nomenclature in pharmacogenomics. It indicates haplotype patterns defined at the level of the gene (also termed as pharmacogene). Usually, the allele that is most frequent in a population and coding a functional protein product, is designated with *1 (Robarge et al. 2007). Haplotypes that contain one or more variants are designated with other numeric labels. Often the assignment of a reference allele is done in the absence of variants defining other alleles, thus *1 designation depends on the variants interrogated.

Therefore, despite the available resources, the assignment of PGx diplotypes still remains somewhat challenging. Straightforward guidelines for adapting these tables with input data coming from microarrays, sequencing, or any other platform, is needed for the implementation of PGx.

1.5. Future directions of pharmacogenomics

PGx is in the forefront of personalized medicine with its wide implementation into the clinic in several countries. With the current ongoing trials and initiatives, the following years will probably bring even more knowledge and evidence of its value in specific contexts.

This year, the PREemptive Pharmacogenomic Testing for Preventing Adverse Drug Reactions (PREPARE) study of the Ubiquitous Pharmacogenomics consortium funded by the European Comission is scheduled to report the results of its pre-emptive PGx testing strategy for the reduction of ADRs (van der Wouden et al. 2017). For half of the patients, treatment is guided based on the results of genetic testing and the outcome is compared with conventional treatment and dosing. This will potentially bring direct evidence of the value of preemptive PGx testing and will help to increase its world-wide adoption in the clinic. Moreover, because questions have been raised regarding the amount of evidence for the cost-effectiveness of PGx testing, which is relevant for the key decision makers when it comes to reimbursement U-PGx, are evaluating this area to add more proof.

Despite the extensive knowledge base for the implementation of PGx, further research of the effects of rare variants in pharmacogenes is necessary to explain more of the variability in drug response. With the continuing decrease in the cost of whole genome sequencing there will be more studies describing the complete spectrum of variants in drug response. Current and future methods for obtaining more information from EHRs will help to design more genetic association studies with larger sample sizes and for different drugs. One of the barriers in PGx GWASes and in genomic research in general has been the failure or difficulties in finding opportunities for replication of discovered associations (Daly 2010; McCarthy et al. 2008). The increasing use of EHRs in different countries and growth in population based genetic information will enable more collaboration between different research groups to validate discovered associations in similar settings. The growing numbers of detected rare variants in pharmacogenes and GWAS studies of drug response will likely reveal more knowledge about the potential polygenicity in the variability of a specific drug response. This, together with the numerous protein-altering genetic variants in each gene, will make the basis for the development of PGx polygenic risk scores (PRS). As both common and rare variants in different pharmacogenes might influence drug PK and PD and thereby drug response, developing scores that consider the entire spectrum of variants would be relevant for better guidance of treatment.

Since the occurrence of type B ADRs (ADRs considered as dose-independent and caused by allergic or non-allergic mechanisms) is rare the collection of enough cases for genetics studies has been challenging, and therefore, less is known about the nature of hypersensitivity. Due to their severity, more studies of immuno-pharmacogenetics are crucial to understand the mechanism of their occurrence. Using EHRs together with genetic data will help brake the barrier of sample size to gain more knowledge of genetic risk factors in the occurrence of hypersensitivity.

The increase in the use of genetics in drug development and repurposing is driven by the clear supporting evidence of the success in using genetics for drug target validation. Potential drug repurposing opportunities were covered extensively in a study where 92 genes were mapped to a GWAS trait different than their drug indications (Sanseau et al. 2012). Furthermore, a recent study indicated that phenotypes that have been associated with genes encoding drug targets, can predict side effects in clinical trials (Nguyen et al. 2019). Thus, human genetics data can not only help in selecting effective drug targets for development, but also aid the development of safer drugs.

It is clear that, in addition to previously known physiological parameters, an individual's genetic variants have a substantial role in drug response, but there are other aspects to consider as well. Recent studies have indicated the relevance of the gut microbiome in the development of drug response (Spanogiannopoulos et al. 2016). One study reported that many drugs are chemically modified by the microorganisms in the gut, which can add more variability in individuals' responses to medicines (Zimmermann et al. 2019). Moreover, the connection can occur in the reverse manner, a drug affects the microbiome that then, in turn,

facilitates adverse phenotypes. This is illustrated by a study with metformin, which showed that altered gut microbiota mediated some of metformin's anti-diabetic effects (Wu et al. 2017). Thus, more studies are needed that not only account for genetic variants in human cells, but also consider the role of the meta-genome for monitoring of treatment response or understanding the adverse side effects of prescribed drugs.

Knowledge of pharmacogenomic variants is rapidly increasing and several ongoing important frontiers and advancements will pave the way for more personalized treatment, improved health and drug development (Lauschke, Zhou, and Ingelman-Sundberg 2019).

2. AIMS OF THE STUDY

The aim of this thesis was to study the genetics of adverse events in pharmacotherapy and explore the translation of genotype data into clinical recommendations using population-based biobanks linked with electronic health records.

The specific objectives of this thesis were following:

- To characterize genetic variants associated with adverse drug events using the information imbedded in EHRs and genetic information from population-based Biobanks
- To evaluate possible drug target-mediated adverse events by analyzing associations between cardiovascular phenotypes from EHRs and genetic variants mimicking the inhibiting effect of a drug
- To develop and test algorithms that systematically translate the genotype or sequencing data of Biobank participants into treatment recommendations based on existing genotype–phenotype associations

3. RESULTS AND DISCUSSION

3.1. Population-based biobanks in combination of electronic health records for research on adverse drug events (Ref. I, II)

Despite the recent decades of PGx studies, which have provided a considerable amount of knowledge about the role of genetic variants in drug response, for most drugs the relevance of genetic variants is yet to be discovered. The proportion of studies that focus on ascertaining relevant genetic variants in ADEs have been minimal and, of the studies that have been performed, most have not been replicated. There is a need for well-powered genome-wide studies to reveal both common and low frequency variants relevant in the onset of ADEs and to find new ways to validate associations that have already been identified. However, gathering enough cases for the study of ADEs has been a challenge, especially for rare ADEs.

The first part of the thesis focuses on studying the role of genetic variants in the occurrences of ADEs on a population-scale by utilizing the large amount of information on drug dispense data and diagnoses imbedded in the EHRs.

3.1.1. Description of cohorts and methods

The basis of discovery studies for this part of the thesis was the Estonian Biobank cohort of ~52,000 participants at the time of the study, which is approximately 5% of the adult population of Estonia (Leitsalu et al. 2015).

For the detection of individuals with variants in pharmacogenes we studied the whole genome sequencing (GS) data of 2,240 participants and, at the time of the first study (Ref. I), 13,986 biobank participants had been genotyped with Illumina microarrays. A population-specific reference panel of $16.5*10^6$ SNV (Mitt et al. 2017) was used to impute missing variants into the genotype data. We set a separate focus on the variants of 64 pharmacogenes with previous evidence of relevance in drug response (Supplementary Table 2, Ref. I). The list was assembled based on the core list of genes from the PharmaADME database and very important pharmacogenes (VIP) listed in PharmGKB.

For the identification of the phenotype of interest, we studied the use of medications based on data in EHRs, which store the longitudinal collection of medication exposures, thus making it a potentially relevant platform for studies of drug effects. The participants of the biobank have signed a broad informed consent form that allows their records to be repeatedly updated by linking to central EHR databases. We have drug dispensing data for every participant, including drug Anatomical Therapeutic Chemical (ATC) Classification System codes, prescription status, and purchase date (if available). Diagnosed diseases are classified based on the ICD10 coding system. We used this resource to

retrospectively identify participants with a prescription of drugs that have been previously designated as high-risk for an unwanted drug response among carriers of a specific genetic variants. To assess the incidence of ADEs among the aforementioned biobank participants we assembled a list of 79 ICD10 codes indicating a possible drug-induced reaction (Supplementary Table 1 in Ref. I). The ICD10 codes, which made no direct mention of the relationship with the drug in the name of the diagnosis were further manually curated by reviewing the free-text fields in medical histories for affirmative comments from the treating physician about the link between the diagnosis and drug. Finally, all the individuals who reported ADEs in the Biobank questionnaire were included in the list of cases. This information on drug prescriptions and ADE phenotypes, together with a wealth of genetic information, set the basis for a hypothesis-free population study to find relevant associations in the occurrences of ADEs, focusing first solely on pharmacogenes and then more broadly by performing a GWAS.

By the time of the second study (Ref. II), all of the first 52,000 Biobank participants had been genotyped using Illumina's Global screening array (n=33,157) or earlier arrays (HumanOmniExpress and HumanCoreExome arrays). In this study, the focus was set on type B ADRs by assessing the occurrence of hypersensitivity related adverse events caused by penicillin. Here, a second source of information was harnessed by gathering the same information on drug sensitivity from the EHRs of the UK Biobank (UKBB). UKBB is a cohort of approximately 500,000 participants with both genome-wide genotype data and a rich variety of phenotypic and health-related information collected for each participant (Bycroft et al. 2018). In the second study, cases were defined as participants with a Z88.0 ICD10 code, reporting a history of allergy status due to penicillin. Due to the few cases with Z88.0 codes in EstBB, all the participants that had reported drug allergy at the recruitment interview were categorized by drug class, using the ATC code J01C* (beta-lactam antibacterials, penicillins) to match this to the respective Z88.0 ICD10 code. We performed separate genome-wide association studies in both cohorts and combined the result in a meta-analysis. To fine-map the association with HLA alleles, we used imputed SNP to HLA typing data for up to 22,554 and 488,377 individuals from the Estonian and UK cohorts, respectively. The results were replicated in two additional cohorts involving a total of 1.14 million individuals.

3.1.2. A plethora of rare variants

We studied the genome sequencing data of 2,240 participants from EstBB and identified 1,314 variants distributed across 64 candidate pharmacogenes that have a role in drug pharmacokinetics or pharmacodynamics (Table 2). Notably, 80.3% of these variants were rare (MAF < 1%) and 42.6% were singletons. Furthermore, we found 41 predicted loss of function variants of which 58.5% were singletons or doubletons and 32.5% of the participants carried at least one LoF variant in an ADMET gene (Table 2).

Table2. Numbers of detected variants and proportion of loss-of-function (LoF) variants in targeted pharmacogenes (n = 64) from whole-genome sequences (n = 2,240).

Variants in 64 targeted pharmacogenes	n	%
Unique variants	1,314	
Novel variants	267	79.7
Known variants	1047	12.5
MAF < 1%	1055	80.3
Singletons	560	42.6
Loss-of-function variants in 64 targeted pharmacogenes		
Unique variants	41	3.1
Unique genes with LoF	25	39.1
Individuals (n= 2,240) with at least 1 LoF	727	32.5
Novel variants	10	24.4
Known variants	31	75.6
MAF < 0.05 %	24	58.5

With the growing number of sequencing-based studies in PGx, the proportion of rare variants discovered in pharmacogenes is increasing (Ingelman-Sundberg et al. 2018; Wright et al. 2018; Kozyra, Ingelman-Sundberg, and Lauschke 2016). This part of the thesis demonstrates this as well: that sequencing can potentially yield a significant amount of additional information for pharmacogenomic predictions in the form of functional rare variants. However, the potential actionability and effect of these rare variants on drug response need to be determined – variants need to pass through functional validation studies before clinical implementation – a prerequisite in the translation of PGx research.

Computational prediction of variant function is the fastest solution, but the basis of algorithms of most of the better-known prediction methods are not modified for pharmacogenetic variants as they are calibrated on disease data sets (Lauschke and Ingelman-Sundberg 2016). However, recently an optimized framework for prediction was developed especially for pharmacogenetic assessments that outperforms the previous computational algorithms (Y. Zhou et al. 2019) thus making the functionality predictions of PGx variants more feasible. More importantly, the past decade brought a revolutionizing tool for biological research – the bacterial clustered, regularly interspaced, short palindromic repeats (CRISPR)–Cas9 system – which can be used for the experimental validation of novel variants (Shalem, Sanjana, and Zhang 2015; Adli 2018). Gathering information on rare variants discovered in pharmacogenes within different populations is valuable for future studies and depositing this information in public databases is necessary to enable open access for different research groups. This is one of the

goals of the eMERGE initiative in which targeted PGx sequencing is used and rare variants with unknown significance are linked to a repository (Rasmussen-Torvik et al. 2014). With this study, we added resources on population-scale PGx variability that can be used for further validation of variants and their functional outcomes.

3.1.3. First pass discovery and replication of genetic variants associated with ADEs (Ref. I)

To examine the role of genetic variants in the occurrence of ADEs, we coupled phenotypic information of medication exposures and incidence of ADEs with the genotype data of the participants of EstBB. From the discovery set of biobank participants we identified 1,187 (7.1%) individuals with possible ADE diagnoses, of which the most prevalent were allergic skin reactions or muscle inflammation (Extended Table 3 in Ref. I).

The main objective of Ref. I was to validate to some extent the EHR-based approach for studying ADEs. For this we first aimed to replicate the PharmGKB variant-drug associations ranked as 'high-confidence' or 'moderate evidence' in our discovery setting. We managed to replicate two of the variant-drug associations ranked by the PharmGKB as high-confidence and four ranked as moderate (Table 2 in Ref. I, Figure 2 A). Next, we explored whether we could identify novel associations by testing new putative high-impact variants in the genes in the same manner. After conditional analysis with the previously known gene-drug variants as co-variates we identified nine novel independent signals (Figure 2, C).

To further discover novel genes potentially relevant in the occurrences ADEs, we performed an association analysis at the whole-genome level among participants who have prescriptions for a specific drug. To have the power to study a whole spectrum of variants, including potential variants with relatively low effects, we focused on 43 different drug groups that had each been prescribed to at least 1000 Biobank participants. Although we discovered 63 associations that will all need further replication, we restricted the first-pass replication stage to five associations. The choice was based on a literature survey and functional and pathway analyses, but also on the descriptive summaries of significant associations (Ref. I. Supplementary Table 4). We managed to replicate one association in an independent sample set from the EstBB using Taqman assays. The association is between a non-coding variant rs75495219 (replication P-value = 6×10^{-4}; meta-analysis p = 2.47×10^{-7}) in the seventh intron of the catenin alpha 3 (*CTNNA3*) gene and the occurrence of myopathy-related ADEs among individuals taking oxicams (Figure 2; B, D). However, further attempts to gain biological insight into this non-coding variant remained challenging. There are some hints of the relevance of *CTNNA3* or other catenins in drug response (Biernacka et al. 2015; Menke et al. 2012; Hamada et al. 2014; Asthma et al. 2012), but whether this variant truly has a role in Oxicam related myositis needs further functional investigation. Although we did not find any previously related eQTL associations

of this variant, the current hypothesis is that rs75495219 might influence the occurrence of ADE by influencing the expression of relevant genes in the oxicam pathway.

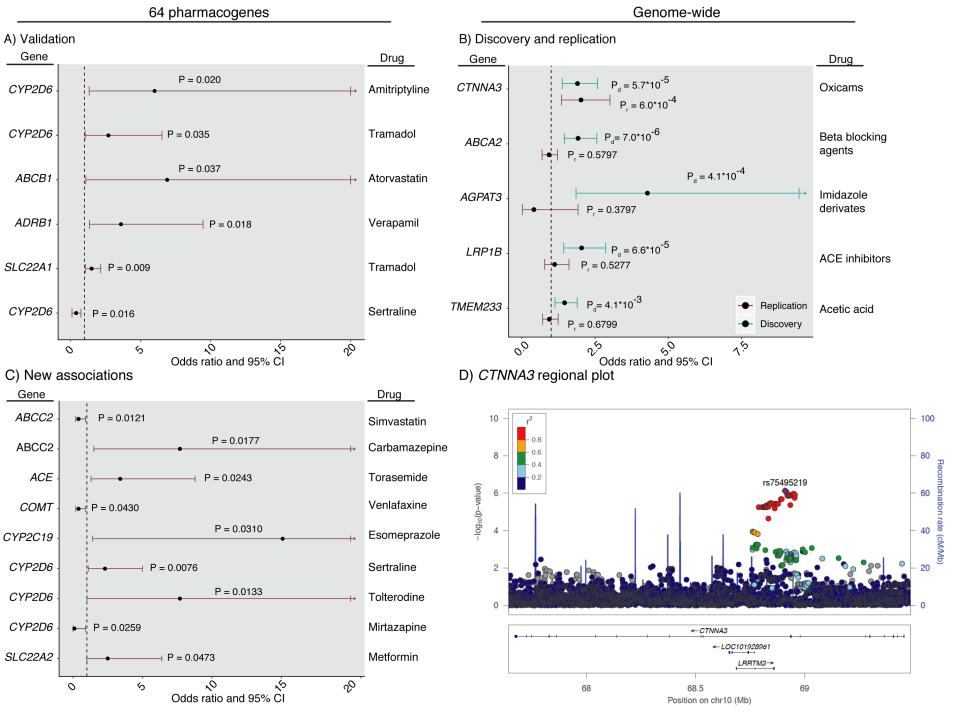

Figure 2. Discovery of genetic variants associated with putative ADEs obtained from the electronic health records of Estonian biobank participants. Validation (A) and new associations (C) of variants in 64 pharmacogenes. Odds ratios (dots) and 95% confidence intervals (horizontal lines) are displayed; the arrow indicates upper CI-s exceeding the limits of the x-axis. (B) Top five significant findings from genome-wide analysis. Discovery associations are displayed in green and replication in purple (D) Regional plot of replicated loci. Colored dots show linkage disequilibrium values for surrounding single-nucleotide variants that are calculated from the 1000 Genomes Project release of 2012 (EUR population) and human hg19 assembly.

3.1.4. Genetic variants in penicillin-induced hypersensitivity (Ref.II, unpublished)

After experiencing the challenges of the broad approach used in Ref I, we decided to pursue a more focused study in Ref II. We found that although penicillin allergy, a type B ADR, is one of the most well-known hypersensitivity reactions to drugs, little was known about its genetic causes. We searched for penicillin-induced allergy using the Z88.0 ICD10 code in UKBB and EstBB, and we identified 15,690 unrelated individuals (4.2%) in UKBB and only 7 (0.02%)

individuals in EstBB. Therefore, we expanded the list of cases in EstBB by including participants who had reported allergy in the recruitment questionnaire. As a result, we then had 961 (2.9%) unrelated cases with penicillin allergy for further GWAS analysis in EstBB.

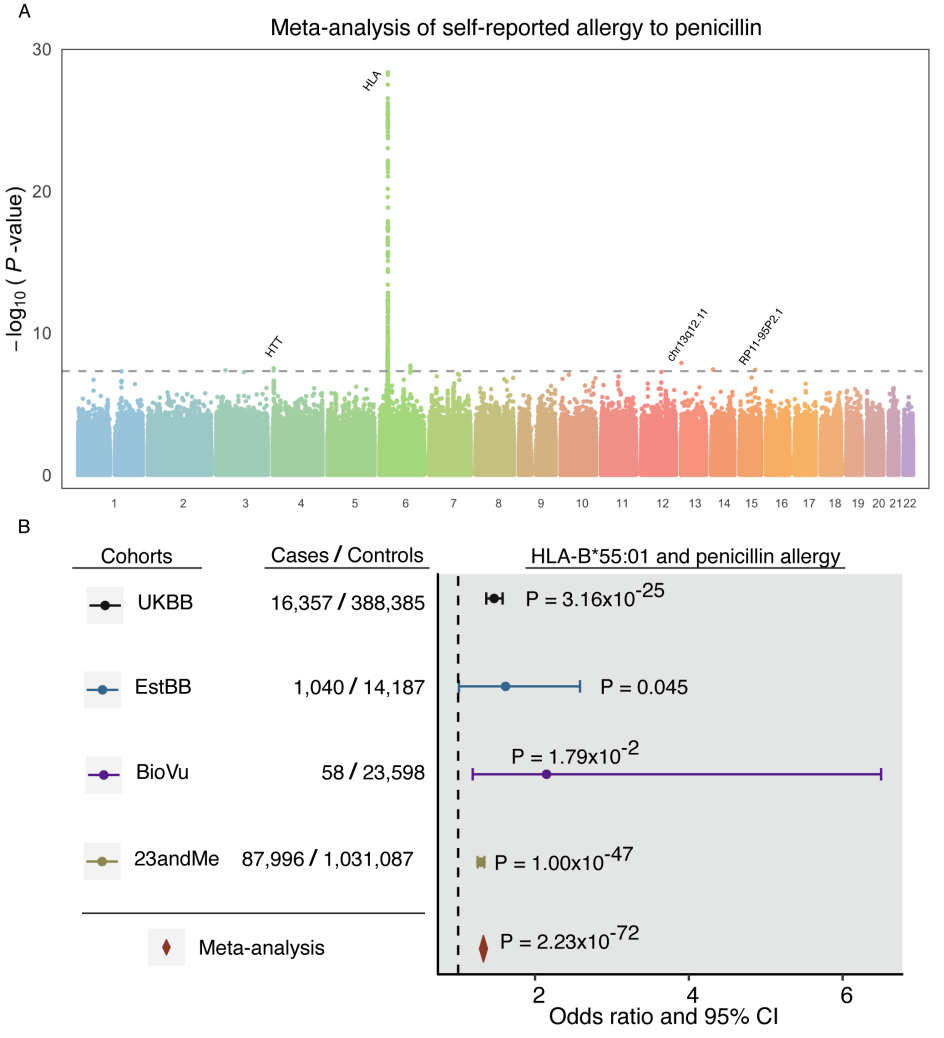

Figure 3. (A) Manhattan plot of the genome-wide association study of self-reported allergy status to penicillin. The X-axes indicate chromosomal positions and Y-axes – log10 of the P-values. Individual dots represent a single nucleotide polymorphism (SNP). The genome-wide P-value threshold (P-value<5.0×10–8) is represented with a dotted line. (B) HLA-B*55:01 allele association with penicillin allergy. Dots are odds ratios with 95% confidence intervals (CI, horizontal lines) and plot is annotated with P-values and case-control numbers. Color coding blue and black indicates the results for discovery cohorts Estonian UK biobank and green and purple are replication results of the HLA*B-55:01 allele in 23andMe research cohort and Vanderbilt University's biobank BioVU Odds ratio of meta-analysis with all four cohorts is indicated with a red diamond.

To ascertain the relevant genetic variants associated with penicillin-induced allergy, we ran a genome-wide study with the allergy reported cases and undiagnosed controls in both the Estonian and UK biobanks (Ref. II). With the meta-analysis we demonstrate a strong genome-wide significant signal in the HLA locus (lead variant rs114892859, MAF(EstBB) = 0.7%, MAF(UKBB) = 2%, $P = 4.59\times10^{-29}$, OR 1.59 95% CI 1.47–1.73; Figure 3 A). To fine-map this signal to a specific HLA allele, we used an additive logistic regression model to test for associations between different four-digit HLA alleles and penicillin allergy in UKBB and EstBB (Supplementary table 5 in Ref. II), and with the meta-analysis we see the strongest association for penicillin allergy with the HLA-B*55:01 allele (P-value 4.63×10^{-26}; OR 1.47 95% CI 1.37–1.58), which is also tagged ($r^2>0.95$) by the GWAS lead variant rs114892859. We obtained further confirmation for this association of the HLA-B*55:01 allele with self-reported penicillin allergy by replicating it among 87,996 cases and 1,031,087 controls from the 23andMe research cohort and among 58 cases and 23,598 controls from the published dataset of Vanderbilt University's biobank BioVU (Figure 3 B). With the meta-analysis of results in the discovery and replication cohorts, this part of the thesis demonstrates a novel, strong association of the HLA-B*55:01 allele with penicillin allergy (P-value 2.23×10^{-72}; OR 1.33 95% CI 1.29–1.37; Figure 3 B). There is one previous GWAS with beta-lactam induced hypersensitivity, where patients with an immediate type of allergic reactions to β-lactams were studied and variants of HLA-DRA, MHC class II alleles, predicted a 1.6-fold reduction in allergy risk (Guéant et al. 2015). There are also studies with beta-lactams indicating strong associations of MHC class I molecules with drug-induced liver-injury, another idiosyncratic ADE (Daly et al. 2009; Clare, Miller, and Dillon 2017).. Almost all cells express MHC I molecules, which are needed to present peptides for cytotoxic CD8+ T cells (Negrini and Becquemont 2017; Chaplin 2010). Several MHC I alleles have previously been associated with drug-induced hypersensitivity (Negrini and Becquemont 2017; Pavlos, Mallal, and Phillips 2012; Sousa-Pinto et al. 2016) and cytotoxic CD8+ T cells have been shown to be relevant especially in allergic skin reactions (Kalish and Askenase 1999; Romano et al. 2004; Adam, Pichler, and Yerly 2011). Here we see an association with the MHC class I molecule and HLA-B*55:01 allele as well as further indications of the tag-SNP overlapping with T cell regulatory annotations, and an association with lymphocyte levels (Ref. II), which hints that we are detecting a predisposition to a T cell-mediated, delayed type of penicillin allergy.

This part of the thesis draws the conclusion that EHRs can be a valuable resource, not only for confirming previously reported gene-drug associations of ADEs, but also to carry their potential over to discovery studies. Furthermore, when considering the extensive work of collecting cases and building a research cohort, the longitudinal nature of EHRs enables retrospective identification of potential ADE cases, including reactions that are considered rare.

3.2. The potential of human genetics studies in the prediction of drug target mediated adverse events (Ref. III)

Evaluating drug target-mediated adverse effects is a crucial step in the drug development process. Only a very small fraction of drug candidates achieve approval due to failures in showing safety and efficacy, despite all the extensive efforts made in the prediction of safety (Waring et al. 2015). Research has demonstrated that naturally occurring genetic variants in humans can be used to approximate the drug's likely efficacy and adverse drug reactions and demonstrate the association between targets and outcomes (Dan M Roden et al. 2019; Plenge, Scolnick, and Altshuler 2013). Several pieces of evidence support the indication that drug targets confirmed by human genetic studies will have increased success in reaching the market (Sanseau et al. 2012; M. R. Nelson et al. 2015).

Sclerostin is a protein encoded by the *SOST* gene, secreted by osteocytes and negatively regulates bone formation. The discovery that a loss of function mutation in *SOST* leads to high bone mass, formed the basis for the development of romosozumab (Figure 4 A), a drug for the treatment of osteoporosis that acts through SOST inhibition (Brunkow et al. 2001). Despite the phase II and III randomized control trial (RCT) showing romosozumab to be effective in increasing heel-bone bone mineral density (BMD), data on adverse events reported in the phase III trials have proposed that romosozumab has a role in an increased risk of cardiovascular events (Lewiecki et al. 2018; Saag et al. 2017).

Using genetic variants as a proxy for the inhibiting effect of romosozumab and EHRs for finding their association with an increased risk for cardiovascular adverse events, this part of the thesis characterizes the valuable opportunity to improve drug development processes with human genetics studies.

3.2.1. Description of cohorts and methods

As one of the participating cohorts of the meta-analysis study, we used the phenotypic and genotypic data of 36,073 unrelated Biobank participants with an average age of 45 years. Participants were either genotyped with one of the Illumina microarrays (Global Screening Array, Infinium CoreExome-24 BeadChips, HumanCNV370-Duo BeadChips or HumanOmniExpress Beadchips) and, for 2420 participants, genome sequencing data was available for the analysis. To study SOST inhibition we selected two independent genetic variants in the SOST locus and to investigate the effects of SOST inhibition on the occurrence of cardiovascular adverse events (Figure 4 B), we used the EHRs to identify individuals with prevalent coronary artery disease (ICD-10 codes I20, I21, I22, I23, I24, I25), infarction (ICD-10 codes I21, I22, I25.2) and utilized the information on systolic blood pressure measured at recruitment to the Biobank. To

find an association of two SOST SNPs with the cardiovascular phenotypes, we obtained the effect estimates using a logistic regression and adjusting for age, sex, and 15 principal components calculated based on the participants genotype data. All the associations were tested under an additive model using the glm function with R software (3.3.2).

3.2.2. Assessing the risk for adverse events with genetic variants in drug target

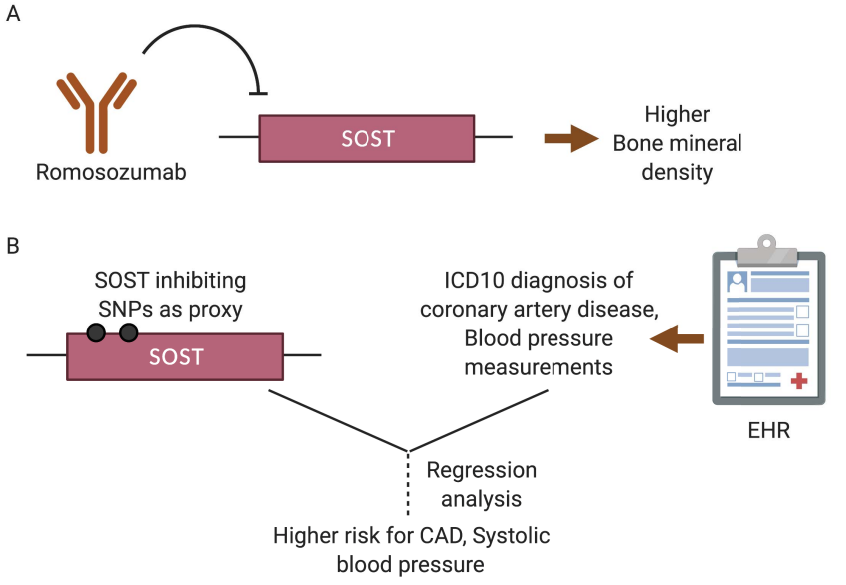

Figure 4. (A) Effect of romosozumab drug for treatment of osteoporosis. (B) Using naturally occurring two SOST variants as proxy to study the risk for drug target mediated adverse events.

As a first step to study SOST inhibition, two independent genetic variants, rs7209826 (A>G, AF(G) in EstBB = 48%, in UKBB = 40%) and rs188810925 (G>A, AF(A) in EstBB = 10%, in UKBB = 8%), were selected for testing in the SOST locus based on the latest GWAS for estimated heel-bone bone mineral density (BMD) (Kemp et al. 2017).

First, to confirm whether inhibition of sclerostin leads to a lower risk of osteoporosis and fracture, we tested the association of the SNPs with osteoporosis and fracture phenotype. Meta-analysis of both SNPs indicated a 57% lower risk of osteoporosis (OR, 0.43; 95% CI, 0.36–0.52; P-value=2.4×10^{-18}) and a 41% lower risk of sustaining a bone fracture (OR, 0.59; 95% CI, 0.54–0.66; P-value=1.4×10^{-24}).

As a next step, we analyzed the association of these SNPs with cardiovascular events (Figure 4 B). In a meta-analysis of 69,649 cases, these variants were associated with an 18% higher risk of myocardial infarction (OR, 1.18; 95% CI, 1.06–1.32; P-value=0.003), and using a wider definition of coronary heart disease yielded a 10% increased risk of disease (OR, 1.10; 95% CI, 1.00–1.20; P-value=0.04). Further, we saw that the SOST variants were significantly associated with 1.3 mmHg higher systolic blood pressure (95% CI, 0.76–1.91; P-value=5.9×10^{-6}), but observed no effect on diastolic blood pressure.

The value of using human genetics to estimate the effects or phenotypes mediated by drug targets has been assessed by others before (Nguyen et al. 2019; Plenge, Scolnick, and Altshuler 2013; Jill M. Pulley et al. 2017; Sanseau et al. 2012; Holmes 2019). With the results of this study showing the increased risk of coronary heart disease from sclerostin inhibition, we validate and support the Food and Drug Administration boxed warning of cardiovascular events with romosozumab (Moritz, Knezevich Emily, and Spangler Miayla 2019). We further hypothesize that, at least partially, this risk might be driven by an increase of cardiometabolic risk factors, like hypertension. It is noteworthy that with these analyses an increased risk of cardiovascular events could have been recognized before the clinical development step, and perhaps enabled a better trial design or, as indicated before, sped up the development process (M. R. Nelson et al. 2015).

This part of the thesis emphasizes the tremendous value of human genetic data from biobanks linked with EHRs to evaluate the inhibiting/stimulating effects of drug targets and their potential impact on drug development. Further, with retrospective studies, more insight can be gained for the post-marketing surveillance of drugs. Using this valuable resource systematically in the future during the process of drug development and also for the discovery of new safe targets should be considered.

3.3. Translating the various sources of genetic data into pharmacogenetic recommendations (Ref. IV)

The various ongoing implementation studies of PGx have applied different approaches for testing variants and clinical implementation (Krebs and Milani 2019). Methods used for genotype assessment include specific panels of candidate pharmacogenes (Hoffman et al. 2014; J. M. Pulley et al. 2012; O'Donnell et al. 2012; Blagec et al. 2018), but also more comprehensive approaches like targeted sequencing (Rasmussen-Torvik et al. 2014). Since the key decision-makers in implementation request cost-saving solutions (Patrinos and Mitropoulou 2017), the underlying question is finding a method with a balance between cost-effectiveness and comprehensiveness of a test. When considering that genome-wide genotyping arrays allow the detection of detailed information about the individuals' genetic variants, the assessment of the accuracy in calling already known high-risk PGx association based on these arrays would be practical.

By harnessing the preexisting genotype information in the Estonian Biobank, this part of the thesis covers ways of translating the data of over 44,000 participants acquired by genome sequencing, exome sequencing, and genotyping together with imputation into pharmacogenetic recommendations and evaluates the distribution of identified phenotypes requiring dose adjustments.

3.3.1. Description of cohorts and methods

We compared the translation of genetic data to pharmacogenetic recommendations among 44,448 Estonian Biobank participants of which 8132 have been genotyped using the HumanOmniExpress beadchip (OMNI) and 33,157 using the Global Screening Array (GSA) from Illumina, 2445 have exome (ES) and 2420 genome sequencing data (GS). For calling pharmacogenetic star alleles from microarray data we used the genotyped variants together with imputed variants. All the genotypes from different platforms were phased for further star allele calling.

The next step was to prepare the publicly available information for calling the star alleles. We set our focus on 11 pharmacogenes that are covered in CPIC guidelines, i.e. they have a validated clinically important impact on drug response. PharmGKB and CPIC provide gene-specific definition tables on how to call star alleles based on the SNP-based genotype information, indicate what the function of the star alleles are, show how to further assess the phenotypes and what the allele frequency is in different populations. Before setting up a decision tree for variant calling we first pruned the allele definition tables based on functional effects of the variants and removed duplicates as well as proxy alleles. The decision pipeline was designed by first examining only the existence of non-functional star alleles, allowing these to over-rule other variants, and then testing the remaining star alleles (Figure 1 in Ref. IV). In ideal cases, only a single matching star allele was detected. Finally, for all the participants, a subsequent phenotype of an allele diplotype was called based on PharmGKB's diplotype-to-phenotype mapping tables.

3.3.2. Comparison of predictions obtained by the different platforms

Our aim was to compare the PGx predictions obtained from any of the different microarrays or sequencing platforms. The most important conclusion based on the comparison was that the microarray-based methods combined with imputation produce results that are very similar to genome sequencing (GS) (Figures 2 and 3 in Ref. IV). However, as seen in Ref. I, when assessing the proportion of rare variants based on sequencing data, we see that 89% of putatively LoF or missense variants detected by GS and ES in these 11 highly relevant pharmacogenes, are rare with MAF < 1% (Table 1 in Ref. IV). For the detection of both rare and common variants, GS is the most precise and comprehensive technology. Nonetheless, considering the striking similarity in the PGx results obtained by

sequencing and microarray data, we proposed that genotyping with phasing and imputation represents a cost-effective and comprehensive alternative when the focus is set on predefined alleles. Moreover, computational phasing enables precise haplotype calls, which is particularly relevant in cases where actionable star alleles are defined by multiple variants. However, for wider use, some of the challenges emerging with phasing and imputation need to be overcome, such as the need for a specific pipeline to first perform imputation and later evaluate its accuracy.

Another aspect is that rare variants will be discovered most accurately with GS and, when considering the proportion of rare variants seen in PGx studies and in this thesis as well, it is clear that their role in drug response needs to be further investigated and accounted for. Yet, since pharmacogenetic reports should only include actionable alleles with clear effects, the function and relevance of this plethora of rare variants need to be validated before they can be included in the allele assessments. Bearing this in mind, it is fair to say that genotyping arrays are clearly a more cost-effective and reliably comprehensive alternative. Admittedly, the collection of information on rare variants and adding them into testing panels is still highly beneficial for further functional research and development projects to one day add them to the allele assessment as well.

3.3.3. Proportion of PGx high-risk phenotypes

When we focused on alleles that define phenotypes that require dose adjustments (poor, rapid, or ultrarapid metabolizers), – commonly referred to as high-risk phenotypes – we found that nonstandard dosing information is necessary based on at least one gene for 99.8% of the participants (Figure 5). A similar proportion has been reported by other studies as well. For example, the studies of the Vanderbilt Pharmacogenomic program indicated that 91% of the genotyped patients had more than one actionable PGx variant (Van Driest et al. 2014). The PG4KDS PGx study at St. Jude Children's Research Hospital reported that around 98.5% of whites and 99.1% of blacks in the US carry at least one high-risk PGx diplotype (Dunnenberger et al. 2015). Further, the PGx implementation programs, the Mayo RIGHT and eMERGE-PGx, outline that 99% and >96% of individuals, respectively, carry high priority PGx actionable variants (Dan M. Roden et al. 2018; Ji et al. 2016). Considering these studies together with this part of the thesis, it highlights the large portion of individuals who would potentially benefit from testing, even when using a conservative panel of well-established pharmacogenes.

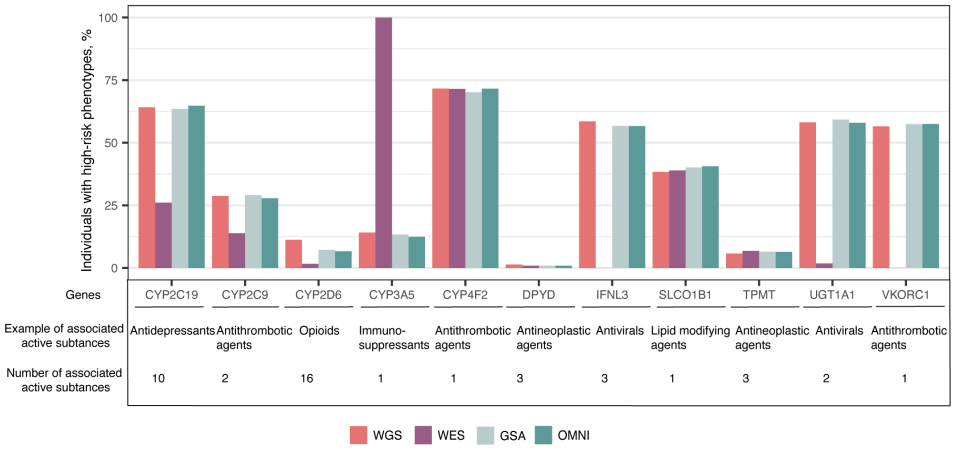

Figure 5. Proportion of phenotypic predictions requiring drug therapy adjustments. High-risk phenotypes are defined as phenotypes requiring a different drug dosing or recommendation.

To further characterize the relevance of the detected phenotypes, we analyzed the frequency of the prescription of medications covered in the CPIC guidelines. We compared the drug consumption in defined daily doses (DDD) based on the annual statistics of five Nordic countries. Notably, some of the active substances covered in the CPIC guidelines are more commonly used, with DDDs for 1000 inhabitants ranging up to 40, which corresponds to a daily use of this drug by 4% of the population ("Statistics on Medicines" n.d.). We further examined the dispensing of medications covered in the CPIC guidelines among EstBB participants and, as an example, 12,254 individuals have had a prescription of at least one drug linked to *CYP2C19*, and 40.7% of them carry the *CYPC219* high-risk phenotype. Thus, they may have needed dosing adjustments to improve the treatment outcome. Further, in a review that covered an analysis of the dispensing of the active substances of 46 PGx drugs reported in the CPIC guidelines, we revealed that 37% (19198/52062) of the participants of the Estonian biobank have purchased at least one prescription for a PGx high-risk drug (Krebs and Milani 2019). Other studies have characterized the proportion of individuals taking PGx-guided medication as well. The Vanderbilt study established that 65% of the 52,000 surveyed individuals had consumed medications with PGx guidance (Schildcrout et al. 2012). A US study of insurance claims of >55 million individuals revealed that around one-quarter had been prescribed a drug with a PGx recommendation label (Frueh et al. 2008). Furthermore, a study of pediatric patients demonstrated that during a 1-year period, 48% of the patients had received a PGx drug (Dunnenberger et al. 2015).

Thus, this part of the thesis confirms that although some drugs with PGx recommendation labels are rarely prescribed, in general, PGx medications are frequently used. With further PGx research, more and more associations will be ascertained and the necessity or benefit of PGx testing will no longer be a question.

CONCLUSIONS

Knowing the predisposing genetic risk factors of drug efficacy and adverse reactions has the potential to help personalize treatment towards the desired outcome. Biobanked genetic data coupled with Electronic health records (EHRs) is a valuable resource for different pharmacogenomics research directions that help to determine the entire spectrum of genetic variants influencing drug response. Discovered and validated actionable genetic markers can be implemented in the clinical setting to detect the individuals in need of dosing adjustments for safer, more effective, cost-saving treatments.

The main conclusions drawn from this thesis are as follows:

- Studies of adverse events (ADE) that are extracted from EHRs and linked with the genetic information of individuals sets a framework for the validation and discovery of potential genetic risk factors for the incidence of ADEs.
- EHRs enable the rapid collection of a large number of cases with ADEs, such as hypersensitivity due to penicillin. A strongly significant and replicated association of penicillin allergy with the MHC class I molecule HLA-B*55:01 suggest its role in creating predisposition a hypersensitivity reaction. Further signals of the tag-SNP for the association, which overlapped with regulatory annotations in T cells and correlated with lymphocyte levels, proposes the HLA-B*55:01 alllele's role in T-cell-mediated, delayed type of penicillin allergy.
- Genetic variants in the *SOST* gene that encodes sclerostin have been associated with a higher bone mineral density and lower fracture risk, which lead to the development of romosozumab, a first-in-class sclerostin inhibitor for the treatment of osteoporosis. However, variants mimicking the therapeutic inhibition of sclerostin indicate an association with the increased risk of cardiovascular disease in large-scale biobanks, supporting the evidence of adverse cardiovascular effects reported in clinical trials of romosozumab. This demonstrates the potential value that could be gained if genomic guidance were embedded in drug development programs to avoid serious side-effects.
- Microarray-based methods combined with imputation produce pharmaco-genetic recommendation results that are very similar to genome sequencing (GS) in their accuracy. Although GS is the most precise and comprehensive technology for the detection of both rare and common variants, when the focus is set on predefined validated PGx alleles, genotyping with phasing and imputation presents as a potentially cost-effective and comprehensive alternative. Furthermore, the high percentage of individuals with at least one genotype associated with non-standard drug dosing of a medication, illustrates the large proportion of individuals who would potentially benefit from PGx testing. Finally, algorithms that enable the semi-automated analysis of genetic data of individuals form the basis of decision support software that can be implemented in clinical care.

SUMMARY IN ESTONIAN

Farmakoteraapias esinevate kõrvaltoimete geneetika uurimine biopankade ja elektrooniliste terviseandmete põhjal

Inimeste geneetiline varieeruvus võib mõjutada ravimi efektiivsust ja ka kõrvaltoimete tekkimist. Farmakogenoomika kui teadusharu eesmärk on tuvastada sellised geneetilised markerid, mis mõjutavad ravimi efektiivsust ja toksilisust ning kasutada seda informatsiooni ennetavalt, et aidata suunata ravi soovitud tulemuse saavutamiseks.

Inimese genoomi sekveneerimine koos uue põlvkonna DNA analüüsimeetodite arenguga ning ülegenoomsete assotsiatsioonianalüüside lainega tõi küll viimaste aastakümnete jooksul palju edasiarenguid komplekshaiguste geneetikas, kuid farmakogenoomikas on areng olnud mõnevõrra aeglasem. Üheks põhjuseks on ravimivastuse kui fenotüübi määramise keerukus, mis on eriti probleemne ravimi kõrvaltoimete korral, kuna juhtumid on harvad ja alaraporteeritud. Nii struktureeritud kui ka mittestruktureeritud elektroonilised terviseandmed on väga mahukas ressurss, mis võimaldab erinevate analüüside läbiviimist ning retroperspektiivset indiviidide ravimikasutuse ja kõrvaltoimete esinemise jälgimist. Sidudes need andmed biopankades olemasolevatega on võimalik uurida ka geneetilise varieeruvuse mõju. Varasemate teadusuuringute käigus tuvastatud ravimi tõhusust ja ohutust mõjutavaid geneetilisi seoseid saab lisaks võimalikule kasutusele ravi suunamisel potentsiaalselt ära kasutada ka ravimiarenduse valdkonnas.

Käesoleva doktoritöö käigus uuriti andmemahukaid elektroonilisi terviseandmeid koos populatsiooni põhiste biopankade andmetega selleks, et läbi viia erinevaid farmakogenoomilisi uuringud. Töö esimeses pooles antakse teaduskirjandusele tuginev ülevaade farmakogenoomika kui teadusharu olemusest ja sellest, kuidas geneetiline varieeruvus ravimivastust mõjutab. Eraldi keskendutakse geneetika rollile ravimite kõrvaltoimete, sealhulgas ravimite ülitundlikkusreaktsioonide tekkes. Lisaks kirjeldatakse erinevaid, juba varasemalt teostatud farmakogenoomika-alaseid uuringuid ning nende läbiviimisel kasutatud metoodikaid. Viimasena antakse ülevaade farmakogeneetika rakendamisest meditsiinivaldkonnas ja võimalikest rakendussuundadest tulevikus. Töö eksperimentaalse osa eesmärgiks oli analüüsida elektrooniliste terviseandmete kasutamist koos biopankade geeniandmetega, et tuvastada ravimite kõrvaltoimete teket mõjutavaid geneetilisi põhjuseid ning kajastada geenidoonorite kohta juba varasemalt kogutud geneetiliste andmete alusel farmakogeneetiliste ravimisoovituste loomist ja selle rakendamise võimalikkust.

Ravimi kõrvaltoimete tekkimise geneetiliste seoste uurimisel on olnud üheks väljakutseks küllalt suure hulga juhtude leidmine nende analüüsimiseks, eriti harvaesinevate kõrvaltoimete korral. Piisavalt suur valim on vajalik, et kirjeldada geneetilist varieeruvust kõigis olulistes aspektides ja tagada analüüsi statistiline usaldusväärsus. Me analüüsisime ravimite kõrvaltoimetega seotud geneetiliste

variantide leidmist kasutades Eesti geenivaramu geenidoonorite elektroonilisi terviseandmeid. Antud teadustöö on esimene uuring, kus püütakse kasutada elektroonilisi tervise- ja biopankades leiduvaid andmeid selleks, et uurida ravimi kõrvaltoimeid ja mille tulemusena õnnestus kinnitada varem näidatud farmakogeneetilisi seoseid ning tuvastada uusi. Lisaks sellele näitas see uuring harvade geneetiliste variantide suurt hulka populatsioonis, mis vajaksid edasisi analüüse, et tuvastada nende täpsemat rolli ravivastuse kujunemisel.

Ravimitest tingitud ülitundlikkusreaktsioonid esinevad küll harvemini, kuid on oma olemuselt ühed kõige tõsisemad kõrvaltoimed, mis tingivad sageli tõsiseid meetmeid kuni ravimite turult tagasikutsumiseni või viivad patsiendi äärmiselt kriitilisse tervislikku seisundisse. Penitsilliiniallergia on üks tihedamini esinev ravimi poolt põhjustatud ülitundlikkusreaktsioonidest. Selleks, et tuvastada penitsilliiniallergiat raporteerinud isikud, kasutasime nii Eesti kui Inglismaa biopanga terviseandmeid ning teostasime ülegenoomse assotsiatsioonianalüüsi, et leida selle teket mõjutavaid võimalikke geneetilisi variatsioone. Metaanalüüsi tulemusena nägime tugevat signaali inimese koesobivusantigeeni (HLA) piirkonnas ning seose täpsemaks kaardistamiseks leidsime edasise HLA tüpiseerimisandmete analüüsimisel mõlemas biopangas olulise seose HLA-B*55:01 markeriga, mida ei ole varem kirjeldatud. Leitud HLA markeri ja penitsilliiniallergia vahelise seose tugevust kinnitas kahes täiendavas kohordis (*23andMe* ja Vanderbilti Ülikooli biopank *BioVU*) läbi viidud analüüs, mis hõlmas kokku 1,14 miljonit inimest. Antud tulemus on paljulubavaks aluseks edasisteks uuringuteks, et mõista mehhanismi HLA-B*55:01 markeri ja penitsilliiniallergia vahel. Lisaks on edasine fenotüübi täpsustamine konkreetsete penitsilliinipõhiste ravimite ja ravimireaktsioonide suhtes vajalik selleks, et anda alus kliiniliselt kasutatava ja ravi suunava seose kujunemiseks.

Ravimiarendus on väga kulukas ja ajamahukas protsess ning väga vähesed ravimikandidaadid jõuavad turule, kuna nende puhul ei õnnestu veenvalt näidata ravimi piisavat ohutust ja tõhusust. Mitmed uuringud on näidanud, et inimeste geeniuuringute abil kinnitatud ravimikandidaadid on turunduse seisukohalt edukamad. Võttes analüüsis arvesse sklerostiinivalku (SOST) kodeeriva geeni geneetilisi variatsioone uurisime selliselt arendusjärgus oleva osteoporoosi kanditaatravimi romosozumab terapeutilist SOSTi pärssivat toimet. Elektroonilistest terviseandmetest saadud diagnoosiandmete seoseanalüüsi kaudu tuvastasime erinevate biopankade indiviididel SOSTi inhibeerimise järgselt suurenenud riski kardiovaskulaarsete kõrvaltoimete tekkeks. Seega kinnitavad meie tulemused ka juba kolmes romosozumabi kliinilises uuringus nähtud seost riski suurenemisega. Antud analüüs illustreerib biopankade ja elektrooniliste terviseandmete kasutamise suurt potentsiaali, mis võimaldaks ennustada juba enne ravimikandidaadi kliinilisi uuringuid võimalike kõrvaltoimete tekkimist, aidates sellega vähendada ravimite väljatöötamise protsessis suuri kulutusi ja tõstes ravimiohutust.

Farmakogeneetiliste uuringute peamine siht on tuvastada olulised geneetilised markerid, mida saaks igapäevases kliinilises praktikas patsiendi ravi määramisel abiks kasutada. Viimaste aastakümnete jooksul avaldatud teadusuuringud on tuvastanud mitmeid tõenduspõhiseid geen-ravim seoseid, mille puhul on välja

töötatud ka arstidele suunatud ravijuhised, mis annavad soovitusi patsientide ravi muutmiseks lähtuvalt nende geneetilisest eripärast. Võttes aluseks 11 geen-ravim seost võrdlesime 44,000 Eesti geenivaramu geenidoonori erinevate meetodite abil saadud geeniandmete kasutamist farmakogeneetiliste ravijuhiste määramiseks. Olulise tulemusena nägime, et genotüpiseerimisandmed koos järgneva imputeerimisega annavad täpsuselt sarnase tulemuse kui kogu genoomi sekveneerimise andmete abil. Kuigi harvade variantide leidmiseks on kogu genoomi sekveneerimine endiselt kõige täpsem meetod, siis kui eesmärgiks on määrata varieeruvus eelnevalt kinnitatud, olulistes farmakogeneetilistes markerites, on mikrokiibi analüüs koos järgneva imputeerimisega selleks piisavalt põhjalik ja samas väga kulutõhus alternatiiv. Lisaks nägime, keskenduses ainult kõrge riskiga farmakogeneetilistele fenotüüpidele, et inimeste hulk, kellel vähemalt ühe geeni varieeruvuse alusel oleks vajalik tavapärasest erinev ravimiannus, on 99,8%. See tähendab, et praktiliselt kõik inimesed saaksid potentsiaalselt kasu selle paneeli farmakogeneetiliste geenide ennetavast testimisest.

Kokkuvõtteks anname antud tööga panuse farmakogenoomika valdkonna arengusse, tuues uusi teadmisi farmakoteraapiasse ravimiohutuse suurendamiseks ja efektiivsema ravi tagamiseks.

REFERENCES

Abifadel, Marianne, Mathilde Varret, Jean Pierre Rabès, Delphine Allard, Khadija Ouguerram, Martine Devillers, Corinne Cruaud, et al. 2003. "Mutations in PCSK9 Cause Autosomal Dominant Hypercholesterolemia." *Nature Genetics* 34 (2): 154–56.

Abul-Husn, Noura S., and Eimear E. Kenny. 2019. "Personalized Medicine and the Power of Electronic Health Records." *Cell* 177 (1): 58–69.

Adam, Jacqueline, Werner J. Pichler, and Daniel Yerly. 2011. "Delayed Drug Hypersensitivity: Models of T-Cell Stimulation." *British Journal of Clinical Pharmacology* 71 (5): 701–7.

Adli, Mazhar. 2018. "The CRISPR Tool Kit for Genome Editing and Beyond." *Nature Communications* 9 (1): 1911.

Alfirevic, A, and M Pirmohamed. 2009. "Predictive Genetic Testing for Drug-Induced Liver Injury : Considerations of Clinical Utility." *Clinical Pharmacology & Therapeutics* 92 (3): 376–80.

Alfirevic, Ana, and Munir Pirmohamed. 2017. "Genomics of Adverse Drug Reactions." *Trends in Pharmacological Sciences* 38 (1): 100–109.

Arbitrio, Mariamena, Maria Teresa Di Martino, Francesca Scionti, Giuseppe Agapito, Pietro Hiram Guzzi, Mario Cannataro, Pierfrancesco Tassone, and Pierosandro Tagliaferri. 2016. "DMET™ (Drug Metabolism Enzymes and Transporters): A Pharmacogenomic Platform for Precision Medicine." *Oncotarget* 7 (33).

Arrowsmith, John. 2011. "Trial Watch: Phase II Failures: 2008–2010." *Drug News Perspect*. Vol. 23. Nature Publishing Group.

Asthma, Allergy, Immunol Res, Mahmood Al Kindi, Vidya Limaye, and Pravin Hissaria. 2012. "Meloxicam-Induced Rhabdomyolysis in the Context of an Acute Ross River Viral Infection" 4 (1): 52–54.

Barbarino, Julia M., Michelle Whirl-Carrillo, Russ B. Altman, and Teri E. Klein. 2018. "PharmGKB: A Worldwide Resource for Pharmacogenomic Information." *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*.

Becquemont, Laurent. 2010. "HLA: A Pharmacogenomics Success Story." *Pharmacogenomics* 11 (3): 277–81.

Biernacka, J M, K Sangkuhl, G Jenkins, R M Whaley, P Barman, A Batzler, R B Altman, V Arolt, J Brockmöller, and C H Chen. 2015. "The International SSRI Pharmacogenomics Consortium (ISPC): A Genome-Wide Association Study of Antidepressant Treatment Response," no. February: 1–9.

Blagec, Kathrin, Rudolf Koopmann, Mandy Crommentuijn-Van Rhenen, Inge Holsappel, Cathelijne H. Van Der Wouden, Lidija Konta, Hong Xu, et al. 2018. "Implementing Pharmacogenomics Decision Support across Seven European Countries: The Ubiquitous Pharmacogenomics (U-PGx) Project." *Journal of the American Medical Informatics Association* 25 (7): 893–98.

Blanca, M, A Romano, M J Torres, C Mayorga, J Rodriguez, P Demoly, P J Bousquet, et al. 2009. "Update on the Evaluation of Hypersensitivity Reactions to Betalactams," no. 3: 183–93.

Böhm, Ruwen, and Ingolf Cascorbi. 2016. "Pharmacogenetics and Predictive Testing of Drug Hypersensitivity Reactions." *Frontiers in Pharmacology*.

Bouvy, Jacoline C., Marie L. De Bruin, and Marc A. Koopmanschap. 2015. "Epidemiology of Adverse Drug Reactions in Europe: A Review of Recent Observational Studies." *Drug Safety* 38 (5): 437–53.

Browning, Sharon R., and Brian L. Browning. 2011. “Haplotype Phasing: Existing Methods and New Developments.” *Nature Reviews Genetics* 12 (10): 703–14.

Brunkow, Mary E., Jessica C. Gardner, Jeff Van Ness, Bryan W. Paeper, Brian R. Kovacevich, Sean Proll, John E. Skonier, et al. 2001. “Bone Dysplasia Sclerosteosis Results from Loss of the SOST Gene Product, a Novel Cystine Knot–Containing Protein.” *The American Journal of Human Genetics* 68 (3): 577–89.

Bush, William S., and Jason H. Moore. 2012. “Chapter 11: Genome-Wide Association Studies.” *PLoS Computational Biology* 8 (12).

Buxton, Iain L O, and Leslie Z Benet. 2015. “Pharmacokinetics: The Dynamics of Drug Absorption, Distribution, Metabolism, and Elimination.” In *Goodman & Gilman’s: The Pharmacological Basis of Therapeutics, 12e*, edited by Laurence L Brunton, Bruce A Chabner, and Björn C Knollmann. New York, NY: McGraw-Hill Education.

Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, et al. 2018. “The UK Biobank Resource with Deep Phenotyping and Genomic Data.” *Nature* 562 (7726): 203–9.

Cao, Chen, and John Moult. 2014. “GWAS and Drug Targets” 15 (Suppl 4): 1–14.

Caudle, Kelly, Teri Klein, James Hoffman, Daniel Muller, Michelle Whirl-Carrillo, Li Gong, Ellen McDonagh, et al. 2014. “Incorporation of Pharmacogenomics into Routine Clinical Practice: The Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline Development Process.” *Current Drug Metabolism* 15 (2): 209–17.

Chaplin, David D. 2010. “Overview of the Immune Response.” *The Journal of Allergy and Clinical Immunology* 125 (2 Suppl 2): S3–23.

Charlton, Matthew, and Jonathan Thompson. 2017. “Adverse Drug Reactions,” 205–9.

Choi, Yongwook, Agnes P. Chan, Ewen Kirkness, Amalio Telenti, and Nicholas J. Schork. 2018. “Comparison of Phasing Strategies for Whole Human Genomes.” *PLoS Genetics* 14 (4): 1–26.

Clancy, J. P., S. G. Johnson, S. W. Yee, E. M. Mcdonagh, K. E. Caudle, T. E. Klein, M. Cannavo, and K. M. Giacomini. 2014. “Clinical Pharmacogenetics Implementation Consortium (CPIC) Guidelines for Ivacaftor Therapy in the Context of CFTR Genotype.” *Clinical Pharmacology and Therapeutics* 95 (6): 592–97.

Clare, Kathleen E., Michael H. Miller, and John F. Dillon. 2017. “Genetic Factors Influencing Drug-Induced Liver Injury: Do They Have a Role in Prevention and Diagnosis?” *Current Hepatology Reports* 16 (3): 258–64.

Cohen, Jonathan C., Eric Boerwinkle, Thomas H. Mosley, and Helen H. Hobbs. 2006. “Sequence Variations in *PCSK9,* Low LDL, and Protection against Coronary Heart Disease.” *New England Journal of Medicine* 354 (12): 1264–72.

Collins, Susannah L., Daniel F. Carr, and Munir Pirmohamed. 2016. “Advances in the Pharmacogenomics of Adverse Drug Reactions.” *Drug Safety* 39 (1): 15–27.

Cooper, Gregory M, Julie A Johnson, Taimour Y Langaee, Hua Feng, Ian B Stanaway, Ute I Schwarz, Marylyn D Ritchie, et al. 2019. “A Genome-Wide Scan for Common Genetic Variants with a Large Influence on Warfarin Maintenance Dose” 112 (4): 1022–28.

Daly, Ann K. 2010. “Genome-Wide Association Studies in Pharmacogenomics.” *Nature Reviews Genetics*. Nat Rev Genet.

Daly, Ann K., Peter T. Donaldson, Pallav Bhatnagar, Yufeng Shen, Itsik Pe’Er, Aris Floratos, Mark J. Daly, et al. 2009. “HLA-B5701 Genotype Is a Major Determinant of Drug-Induced Liver Injury Due to Flucloxacillin.” *Nature Genetics* 41 (7): 816–19.

DeGorter, M.K., C.Q. Xia, J.J. Yang, and R.B. Kim. 2011. “Drug Transporters in Drug Efficacy and Toxicity.” *Annual Review of Pharmacology and Toxicology* 52 (1): 249–73.

Driest, S. L. Van, Y. Shi, E. Bowton, J. Schildcrout, J. Peterson, J. Pulley, J. Denny, and D. Roden. 2014. “Clinically Actionable Genotypes among 10,000 Patients with Preemptive Pharmacogenomic Testing.” *Clinical Pharmacology and Therapeutics* 95 (4): 423–31.

Dunnenberger, Henry M, Kristine R Crews, James M Hoffman, Kelly E Caudle, Scott C Howard, Robert J Hunkler, Teri E Klein, William E Evans, and Mary V Relling. 2015. “Preemptive Clinical Pharmacogenetics Implementation: Currentprograms in Five United States Medical Centers.” *Annu Rev Pharmacol Toxicol* 55: 89–106.

Edwards, I Ralph, and Jeffrey K Aronson. 2000. “Adverse Drug Reactions: Definitions, Diagnosis, and Management.” *The Lancet* 356 (9237): 1255–59.

Evans, W E, and M V Relling. 1999. “Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics.” *Science (New York, N.Y.)* 286 (5439): 487–91.

Frueh, Felix W, Shashi Amur, Padmaja Mummaneni, Robert S Epstein, Ronald E Aubert, Teresa M DeLuca, Robert R Verbrugge, Gilbert J Burckart, and Lawrence J Lesko. 2008. “Pharmacogenomic Biomarker Information in Drug Labels Approved by the United States Food and Drug Administration: Prevalence of Related Drug Use.” *Pharmacotherapy* 28 (8): 992–98.

Giacomini, K. M., S. W. Yee, M. J. Ratain, R. M. Weinshilboum, N. Kamatani, and Y. Nakamura. 2012. “Pharmacogenomics and Patient Care: One Size Does Not Fit All.” *Science Translational Medicine* 4 (153): 153ps18–153ps18.

Giacomini, Kathleen M., Sook Wah Yee, Taisei Mushiroda, Richard M. Weinshilboum, Mark J. Ratain, and Michiaki Kubo. 2017. “Genome-Wide Association Studies of Drug Response and Toxicity: An Opportunity for Genome Medicine.” *Nature Reviews Drug Discovery* 16 (1): 70–70.

GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group, Wellcome Trust Case Control Consortium 2, Kaixin Zhou, Celine Bellenguez, Chris C A Spencer, Amanda J Bennett, Ruth L Coleman, et al. 2011. “Common Variants near ATM Are Associated with Glycemic Response to Metformin in Type 2 Diabetes.” *Nature Genetics* 43 (2): 117–20.

Gordon, Adam S., Robert S. Fulton, Xiang Qin, Elaine R. Mardis, Deborah A. Nickerson, and Steve Scherer. 2016. “PGRNseq: A Targeted Capture Sequencing Panel for Pharmacogenetic Research and Implementation.” *Pharmacogenetics and Genomics* 26 (4): 161–68.

Guéant, Jean Louis, Antonino Romano, Jose Antonio Cornejo-Garcia, Abderrahim Oussalah, Celine Chery, Natalia Blanca-López, Rosa Maria Guéant-Rodriguez, et al. 2015. “HLA-DRA Variants Predict Penicillin Allergy in Genome-Wide Fine-Mapping Genotyping.” *Journal of Allergy and Clinical Immunology* 135 (1): 253–259.e10.

Hamada, Shunsuke, Naohisa Futamura, Kunihiro Ikuta, Hiroshi Urakawa, Eiji Kozawa, and Naoki Ishiguro. 2014. “CTNNB1 S45F Mutation Predicts Poor Efficacy of Meloxicam Treatment for Desmoid Tumors : A Pilot Study” 9 (5): 5–10.

Harper, Andrew R, and Eric J Topol. 2012. “Pharmacogenomics in Clinical Practice and Drug Development.” *Nature Biotechnology* 30 (11): 1117–24.

Harrison, Richard K. 2016. “Phase II and Phase III Failures: 2013–2015.” *Nature Reviews Drug Discovery* 15 (12): 817–18.

Hazell, Lorna, and Saad A W Shakir. 2006. “Under-Reporting of Adverse Drug Reactions A Systematic Review.” *Drug Safety*. Vol. 29.

Hersh, William R., Mark G. Weiner, Peter J. Embi, Judith R. Logan, Philip R.O. Payne, Elmer V. Bernstam, Harold P. Lehmann, et al. 2013. "Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research." *Medical Care* 51 (8 Suppl 3): S30–37.

Ho, Richard H., and Richard B. Kim. 2005. "Transporters and Drug Therapy: Implications for Drug Disposition and Disease." *Clinical Pharmacology and Therapeutics* 78 (3): 260–77.

Hodgkinson, Marisa R, Nicole J Dirnbauer, and Ian Larmour. 2009. "Identification of Adverse Drug Reactions Using the ICD-10 Australian Modification Clinical Coding Surveillance." *Journal of Pharmacy Practice and Research* 39 (1): 19–23.

Hoffman, James M., Cyrine E. Haidar, Mark R. Wilkinson, Kristine R. Crews, Donald K. Baker, Nancy M. Kornegay, Wenjian Yang, et al. 2014. "PG4KDS: A Model for the Clinical Implementation of Pre-Emptive Pharmacogenetics." *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics* 166 (1): 45–55.

Holmes, Michael V. 2019. "Human Genetics and Drug Development." *New England Journal of Medicine* 380 (11): 1076–79.

Iasella, Carlo J., Heather J. Johnson, and Michael A. Dunn. 2017. "Adverse Drug Reactions." *Clinics in Liver Disease* 21 (1): 73–87.

Illing, Patricia T., Anthony W. Purcell, and James McCluskey. 2017. "The Role of HLA Genes in Pharmacogenomics: Unravelling HLA Associated Adverse Drug Reactions." *Immunogenetics*.

Ingelman-Sundberg, Magnus, Souren Mkrtchian, Yitian Zhou, and Volker M. Lauschke. 2018. "Integrating Rare Genetic Variants into Pharmacogenetic Drug Response Predictions." *Human Genomics*.

Ji, Yuan, Jennifer M. Skierka, Joseph H. Blommel, Brenda E. Moore, Douglas L. Vancuyk, Jamie K. Bruflat, Lisa M. Peterson, et al. 2016. "Preemptive Pharmacogenomic Testing for Precision Medicine: A Comprehensive Analysis of Five Actionable Pharmacogenomic Genes Using Next-Generation DNA Sequencing and a Customized CYP2D6 Genotyping Cascade." *Journal of Molecular Diagnostics* 18 (3): 438–45.

Kalish, Richard S., and Philip W. Askenase. 1999. "Molecular Mechanisms of CD8+T Cell-Mediated Delayed Hypersensitivity: Implications for Allergies, Asthma, and Autoimmunity." *Journal of Allergy and Clinical Immunology* 103 (2 II): 192–99.

Kalman, L. V., J. A.G. Agúndez, M. Lindqvist Appell, J. L. Black, G. C. Bell, S. Boukouvala, C. Bruckner, et al. 2016. "Pharmacogenetic Allele Nomenclature: International Workgroup Recommendations for Test Result Reporting." *Clinical Pharmacology and Therapeutics*.

Kemp, John P, John A Morris, Carolina Medina-Gomez, Vincenzo Forgetta, Nicole M Warrington, Scott E Youlten, Jie Zheng, et al. 2017. "Identification of 153 New Loci Associated with Heel Bone Mineral Density and Functional Involvement of GPC6 in Osteoporosis." *Nature Genetics* 49 (10): 1468–75.

King, Emily A., J. Wade Davis, and Jacob F. Degner. 2019. "Are Drug Targets with Genetic Support Twice as Likely to Be Approved? Revised Estimates of the Impact of Genetic Support for Drug Mechanisms on the Probability of Drug Approval." *PLoS Genetics* 15 (12): e1008489.

Kohane, Isaac S. 2011. "Using Electronic Health Records to Drive Discovery in Disease Genomics." *Nature Reviews Genetics*. Nat Rev Genet.

Kozyra, Mikael, Magnus Ingelman-Sundberg, and Volker M Lauschke. 2016. "Rare Genetic Variants in Cellular Transporters, Metabolic Enzymes, and Nuclear Receptors Can Be Important Determinants of Interindividual Differences in Drug Response."

Genetics in Medicine : Official Journal of the American College of Medical Genetics, no. November 2015.

Krebs, Kristi, and Lili Milani. 2019. "Translating Pharmacogenomics into Clinical Decisions: Do Not Let the Perfect Be the Enemy of the Good." *Human Genomics*. NLM (Medline).

Lauschke, Volker M., and Magnus Ingelman-Sundberg. 2016. "Requirements for Comprehensive Pharmacogenetic Genotyping Platforms." *Pharmacogenomics*.

Lauschke, Volker M., Yitian Zhou, and Magnus Ingelman-Sundberg. 2019. "Novel Genetic an Epigenetic Factors of Importance for Inter-Individual Differences in Drug Disposition, Response and Toxicity." *Pharmacology and Therapeutics*.

Lazarou, Jason, Bruce H. Pomeranz, and Paul N. Corey. 2003. "Incidence of Adverse Drug Reactions in Hospitalized Patients." *Jama* 279 (15): 1200.

Leitsalu, Liis, Helene Alavere, Mari-liis Tammesoo, Erkki Leego, and Andres Metspalu. 2015. "Linking a Population Biobank with National Health Registries — The Estonian Experience." *Journal of Personalized Medicine*, 96–106.

Lewiecki, E Michael, Tomasz Blicharski, Stefan Goemaere, Kurt Lippuner, Paul D Meisner, Paul D Miller, Akimitsu Miyauchi, Judy Maddox, Li Chen, and Stephane Horlait. 2018. "A Phase III Randomized Placebo-Controlled Trial to Evaluate Efficacy and Safety of Romosozumab in Men With Osteoporosis." *The Journal of Clinical Endocrinology & Metabolism* 103 (9): 3183–93.

Liou, Shyh-Yuh, Fran Stringer, and Masashi Hirayama. 2012. "The Impact of Pharmacogenomics Research on Drug Development." *Drug Metabolism and Pharmacokinetics* 27 (1): 2–8.

Lonergan, Mike, Stephen J. Senn, Christine McNamee, Ann K. Daly, Robert Sutton, Andrew Hattersley, Ewan Pearson, and Munir Pirmohamed. 2017. "Defining Drug Response for Stratified Medicine." *Drug Discovery Today* 22 (1): 173–79. https://doi.org/10.1016/j.drudis.2016.10.016.

Lynch, Tom, and Amy Price. 2007. "The Effect of Cytochrome P450 Metabolism on Drug Response, Interactions, and Adverse Effects." *American Family Physician* 76 (3): 391–96.

Maggo, Simran D.S., Ruth L. Savage, and Martin A. Kennedy. 2016. "Impact of New Genomic Technologies on Understanding Adverse Drug Reactions." *Clinical Pharmacokinetics*. Springer International Publishing.

Marques, Francisco Batel, Ana Penedones, Diogo Mendes, and Carlos Alves. 2016. "A Systematic Review of Observational Studies Evaluating Costs of Adverse Drug Reactions." *ClinicoEconomics and Outcomes Research*.

McCarthy, Mark I., Gonçalo R. Abecasis, Lon R. Cardon, David B. Goldstein, Julian Little, John P.A. Ioannidis, and Joel N. Hirschhorn. 2008. "Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges." *Nature Reviews Genetics*. Nature Publishing Group.

Meade, T., P. Sleight, R. Collins, J. Armitage, S. Parish, J. Barton, C. Bray, et al. 2008. "SLCO1B1 Variants and Statin-Induced Myopathy – A Genomewide Study." *New England Journal of Medicine* 359 (8): 789–99.

Medicines Agency, European. 2017. "Guideline on Good Pharmacovigilance Practices (GVP) – Annex I – Definitions (Rev 4)." www.ema.europa.eu.

Menke, Andreas, Katharina Domschke, Darina Czamara, Torsten Klengel, Johannes Hennings, Florian Holsboer, Susanne Lucae, Bernhard T Baune, Volker Arolt, and Bertram Mu. 2012. "Genome-Wide Association Study of Antidepressant Treatment-Emergent Suicidal Ideation," 797–807.

Mitt, Mario, Mart Kals, Kalle Pärn, Stacey B Gabriel, Eric S Lander, Aarno Palotie, Samuli Ripatti, et al. 2017. “Improved Imputation Accuracy of Rare and Low-Frequency Variants Using Population-Speci Fi c High-Coverage WGS-Based Imputation Reference Panel,” no. December 2016: 1–8.

Moritz, Mackenzie, Knezevich Emily, and Spangler Miayla. 2019. “Updates in the Treatment of Postmenopausal Osteoporosis.” *Orthopedics* 44 (9): 32–32.

Nebert, Daniel W., and David W. Russell. 2002. “Clinical Importance of the Cytochromes P450.” *Lancet* 360 (9340): 1155–62.

Negrini, Simone, and Laurent Becquemont. 2017. “HLA-Associated Drug Hypersensitivity and the Prediction of Adverse Drug Reactions.” *Pharmacogenomics*. Future Medicine Ltd.

Nelson, David R, Darryl C Zeldin, Susan M G Hoffman, Lois J Maltais, Hester M Wain, and Daniel W Nebert. 2004. “Comparison of Cytochrome P450 (CYP) Genes from the Mouse and Human Genomes, Including Nomenclature Recommendations for Genes, Pseudogenes and Alternative-Splice Variants.” *Pharmacogenetics* 14 (1): 1–18.

Nelson, Matthew R, Hannah Tipney, Jeffery L Painter, Judong Shen, Paola Nicoletti, Yufeng Shen, Aris Floratos, et al. 2015. “The Support of Human Genetic Evidence for Approved Drug Indications.” *Nature Genetics* 47 (8): 856–60.

Nguyen, Phuong A., David A. Born, Aimee M. Deaton, Paul Nioi, and Lucas D. Ward. 2019. “Phenotypes Associated with Genes Encoding Drug Targets Are Predictive of Clinical Trial Side Effects.” *Nature Communications* 10 (1): 1579.

Niemeijer, Maartje N, Marten E van den Berg, Mark Eijgelsheim, Peter R Rijnbeek, and Bruno H Stricker. 2015. “Pharmacogenetics of Drug-Induced QT Interval Prolongation: An Update.” *Drug Safety* 38 (10): 855–67.

O’Donnell, P. H., A. Bush, J. Spitz, K. Danahey, D. Saner, S. Das, N. J. Cox, and M. J. Ratain. 2012. “The 1200 Patients Project: Creating a New Medical Model System for Clinical Implementation of Pharmacogenomics.” *Clinical Pharmacology and Therapeutics* 92 (4): 446–49.

Osanlou, Orod, Munir Pirmohamed, and Ann K. Daly. 2018. “Pharmacogenetics of Adverse Drug Reactions.” In *Advances in Pharmacology*.

Pasanen, Marja K., Mikko Neuvonen, Pertti J. Neuvonen, and Mikko Niemi. 2006. “SLCO1B1 Polymorphism Markedly Affects the Pharmacokinetics of Simvastatin Acid.” *Pharmacogenetics and Genomics* 16 (12): 873–79.

Patrinos, G. P., and C. Mitropoulou. 2017. “Measuring the Value of Pharmacogenomics Evidence.” *Clinical Pharmacology and Therapeutics* 102 (5): 739–41.

Pavlos, Rebecca, Simon Mallal, David Ostrov, Soren Buus, Imir Metushi, Bjoern Peters, and Elizabeth Phillips. 2015. “T Cell–Mediated Hypersensitivity Reactions to Drugs.” *Annual Review of Medicine* 66 (1): 439–54.

Pavlos, Rebecca, Simon Mallal, and Elizabeth Phillips. 2012. “HLA and Pharmacogenetics of Drug Hypersensitivity.” *Pharmacogenomics* 13 (11): 1285–1306.

Pendergrass, Sarah A., and Dana C. Crawford. 2018. “Using Electronic Health Records To Generate Phenotypes For Research.” *Current Protocols in Human Genetics* 100 (1): e80.

Petzinger, Ernst, and Joachim Geyer. 2006. “Drug Transporters in Pharmacokinetics.” *Naunyn-Schmiedeberg’s Archives of Pharmacology* 372 (6): 465–75.

Plenge, Robert M., Edward M. Scolnick, and David Altshuler. 2013. “Validating Therapeutic Targets through Human Genetics.” *Nature Reviews Drug Discovery* 12 (8): 581–94.

Pulley, J. M., J. C. Denny, J. F. Peterson, G. R. Bernard, C. L. Vnencak-Jones, A. H. Ramirez, J. T. Delaney, et al. 2012. "Operational Implementation of Prospective Genotyping for Personalized Medicine: The Design of the Vanderbilt PREDICT Project." *Clinical Pharmacology and Therapeutics* 92 (1): 87–95.

Pulley, Jill M., Jana K. Shirey-Rice, Robert R. Lavieri, Rebecca N. Jerome, Nicole M. Zaleski, David M. Aronoff, Lisa Bastarache, et al. 2017. "Accelerating Precision Drug Development and Drug Repurposing by Leveraging Human Genetics." *ASSAY and Drug Development Technologies* 15 (3): 113–19.

Rasmussen-Torvik, L. J., S. C. Stallings, A. S. Gordon, B. Almoguera, M. A. Basford, S. J. Bielinski, A. Brautbar, et al. 2014. "Design and Anticipated Outcomes of the EMERGE-PGx Project: A Multicenter Pilot for Preemptive Pharmacogenomics in Electronic Health Record Systems." *Clinical Pharmacology and Therapeutics* 96 (4): 482–89.

Relling, M V, and T E Klein. 2009. "CPIC : Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network." *Clinical Pharmacology & Therapeutics* 89 (3): 464–67.

Relling, Mary V., and William E. Evans. 2015. "Pharmacogenomics in the Clinic." *Nature* 526 (7573): 343–50.

Rieder, Mark J., Alexander P. Reiner, Brian F. Gage, Deborah A. Nickerson, Charles S. Eby, Howard L. McLeod, David K. Blough, Kenneth E. Thummel, David L. Veenstra, and Allan E. Rettie. 2005. "Effect of *VKORC1* Haplotypes on Transcriptional Regulation and Warfarin Dose." *New England Journal of Medicine* 352 (22): 2285–93.

Riedl, Marc A, and Adrian M Casillas. 2003. "Adverse Drug Reactions :," no. Table 3.

Robarge, J. D., L. Li, Z. Desta, A. Nguyen, and D. A. Flockhart. 2007. "The Star-Allele Nomenclature: Retooling for Translational Genomics." *Clinical Pharmacology and Therapeutics*.

Robinson, Jamie R, Joshua C Denny, Dan M Roden, and Sara L Van Driest. 2018. "Genome-Wide and Phenome-Wide Approaches to Understand Variable Drug Actions in Electronic Health Records," no. October 2017: 112–22. https://doi.org/10.1111/cts.12522.

Roden, D M, H Xu, J C Denny, and R A Wilke. 2012. "Electronic Medical Records as a Tool in Clinical Pharmacology: Opportunities and Challenges." *Clinical Pharmacology and Therapeutics* 91 (6): 1083–86.

Roden, Dan M., Sara L. Van Driest, Jonathan D. Mosley, Quinn S. Wells, Jamie R. Robinson, Joshua C. Denny, and Josh F. Peterson. 2018. "Benefit of Preemptive Pharmacogenetic Information on Clinical Outcome." *Clinical Pharmacology and Therapeutics* 103 (5): 787–94.

Roden, Dan M, Howard L McLeod, Mary V Relling, Marc S Williams, George A Mensah, Josh F Peterson, and Sara L Van Driest. 2019. "Pharmacogenomics." *Lancet (London, England)* 394 (10197): 521–32.

Roden, Dan M, Russell A Wilke, Heyo K Kroemer, and C Michael Stein. 2011. "Pharmacogenomics: The Genetics of Variable Drug Responses." *Circulation* 123 (15): 1661–70.

Romano, A., M. Blanca, M. J. Torres, A. Bircher, W. Aberer, K. Brockow, W. J. Pichler, and P. Demoly. 2004. "Diagnosis of Nonimmediate Reactions to Beta-Lactam Antibiotics." *Allergy* 59 (11): 1153–60.

Saag, Kenneth G., Jeffrey Petersen, Maria Luisa Brandi, Andrew C. Karaplis, Mattias Lorentzon, Thierry Thomas, Judy Maddox, Michelle Fan, Paul D. Meisner, and Andreas Grauer. 2017. "Romosozumab or Alendronate for Fracture Prevention in Women with Osteoporosis." *New England Journal of Medicine* 377 (15): 1417–27.

Sanseau, Philippe, Pankaj Agarwal, Michael R Barnes, Tomi Pastinen, J Brent Richards, Lon R Cardon, and Vincent Mooser. 2012. "Use of Genome-Wide Association Studies for Drug Repositioning." *Nature Biotechnology* 30 (4): 317–20.

Schaller, Lena, and Volker M. Lauschke. 2019. "The Genetic Landscape of the Human Solute Carrier (SLC) Transporter Superfamily." *Human Genetics* 138 (11–12): 1359–77.

Schärfe, Charlotta Pauline Irmgard, Roman Tremmel, Matthias Schwab, Oliver Kohlbacher, and Debora Susan Marks. 2017. "Genetic Variation in Human Drug-Related Genes." *Genome Medicine* 9 (1).

Schildcrout, J S, J C Denny, E Bowton, W Gregg, J M Pulley, M A Basford, J D Cowan, et al. 2012. "Optimizing Drug Outcomes Through Pharmacogenetics: A Case for Preemptive Genotyping." *Clinical Pharmacology & Therapeutics* 92 (2): 235–42.

SEARCH Collaborative Group, E Link, S Parish, J Armitage, L Bowman, S Heath, F Matsuda, I Gut, M Lathrop, and R Collins. 2008. "*SLCO1B1* Variants and Statin-Induced Myopathy – A Genomewide Study." *New England Journal of Medicine* 359 (8): 789–99.

Shalem, Ophir, Neville E Sanjana, and Feng Zhang. 2015. "High-Throughput Functional Genomics Using CRISPR-Cas9." *Nature Reviews. Genetics* 16 (5): 299–311.

Sousa-Pinto, Bernardo, Cláudia Correia, Lídia Gomes, Sara Gil-Mata, Luís Araújo, Osvaldo Correia, and Luís Delgado. 2016. "HLA and Delayed Drug-Induced Hypersensitivity." *International Archives of Allergy and Immunology* 170 (3): 163–79.

Spanogiannopoulos, Peter, Elizabeth N. Bess, Rachel N. Carmody, and Peter J. Turnbaugh. 2016. "The Microbial Pharmacists within Us: A Metagenomic View of Xenobiotic Metabolism." *Nature Reviews Microbiology*. Nature Publishing Group.

"Statistics on Medicines." n.d. Accessed August 21, 2019. http://pxweb.tai.ee/esf/pxweb2008/Database_en/Medicines/Medicines/ATCinfo_en.htm.

Stausberg, Jürgen, and Joerg Hasford. 2010. "Identification of Adverse Drug Events: The Use of ICD-10 Coded Diagnoses in Routine Hospital Data." *Deutsches Arzteblatt International* 107 (3): 23–29.

Swen, J J, M Nijenhuis, A de Boer, L Grandia, A H Maitland-van der Zee, H Mulder, G A P J M Rongen, et al. 2011. "Pharmacogenetics: From Bench to Byte – an Update of Guidelines." *Clinical Pharmacology and Therapeutics* 89 (5): 662–73.

Swen, Jesse J., Marga Nijenhuis, Mandy van Rhenen, Nienke J. de Boer-Veger, Anne Marie Buunk, Elisa J.F. Houwink, Hans Mulder, et al. 2018. "Pharmacogenetic Information in Clinical Guidelines: The European Perspective." *Clinical Pharmacology and Therapeutics* 103 (5): 795–801.

Takeuchi, Fumihiko, Ralph Mcginnis, Stephane Bourgeois, Chris Barnes, Niclas Eriksson, Pamela Whittaker, Venkatesh Ranganath, et al. 2009. "A Genome-Wide Association Study Confirms VKORC1 , CYP2C9 , and CYP4F2 as Principal Genetic Determinants of Warfarin Dose" 5 (3).

Tam, Vivian, Nikunj Patel, Michelle Turcotte, Yohan Bossé, Guillaume Paré, and David Meyre. 2019. "Benefits and Limitations of Genome-Wide Association Studies." *Nature Reviews Genetics*. Nature Publishing Group.

Testa, Bernard. 2004. "Prodrug Research: Futile or Fertile?" *Biochemical Pharmacology* 68 (11): 2097–2106.

Vandiedonck, C., and J. C. Knight. 2009. "The Human Major Histocompatibility Complex as a Paradigm in Genomics Research." *Briefings in Functional Genomics and Proteomics* 8 (5): 379–94.

Visscher, Peter M., Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *The American Journal of Human Genetics* 101 (1): 5–22.

Waring, Michael J., John Arrowsmith, Andrew R. Leach, Paul D. Leeson, Sam Mandrell, Robert M. Owen, Garry Pairaudeau, et al. 2015. "An Analysis of the Attrition of Drug Candidates from Four Major Pharmaceutical Companies." *Nature Reviews Drug Discovery* 14 (7): 475–86.

Weinshilboum, Richard. 2003. "Inheritance and Drug Response." *New England Journal of Medicine* 348 (6): 529–37.

Whirl-Carrillo, M, E M McDonagh, J M Hebert, L Gong, K Sangkuhl, C F Thorn, R B Altman, and T E Klein. 2012. "Pharmacogenomics Knowledge for Personalized Medicine." *Clinical Pharmacology & Therapeutics* 92 (4): 414–17.

Wilke, R A, H Xu, J C Denny, D M Roden, R M Krauss, C A Mccarty, R L Davis, T Skaar, J Lamba, and G Savova. 2011. "The Emerging Role of Electronic Medical Records in Pharmacogenomics" 89 (3): 379–86.

Wilkinson, Grant R. 2005. "Drug Metabolism and Variability among Patients in Drug Response." *New England Journal of Medicine* 352 (21): 2211–21.

Wouden, C. H. van der, A. Cambon-Thomsen, E. Cecchin, K. C. Cheung, C. L. Dávila-Fajardo, V. H. Deneer, V. Dolžan, et al. 2017. "Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium." *Clinical Pharmacology and Therapeutics* 101 (3): 341–58.

Wright, G. E.B., B. Carleton, M. R. Hayden, and C. J.D. Ross. 2018. "The Global Spectrum of Protein-Coding Pharmacogenomic Diversity." *Pharmacogenomics Journal* 18 (1): 187–95.

Wu, Hao, Eduardo Esteve, Valentina Tremaroli, Muhammad Tanweer Khan, Robert Caesar, Louise Mannerås-Holm, Marcus Ståhlman, et al. 2017. "Metformin Alters the Gut Microbiome of Individuals with Treatment-Naive Type 2 Diabetes, Contributing to the Therapeutic Effects of the Drug." *Nature Medicine* 23 (7): 850–58.

Xu, Heng, Giles W Robinson, Jie Huang, Joshua Yew-suang Lim, Hui Zhang, Johnnie K Bass, Alberto Broniscer, et al. 2015. "Common Variants in ACYP2 Influence Susceptibility to Cisplatin-Induced Hearing Loss" 47 (3).

Yee, Sook Wah, Ligong Chen, and Kathleen M. Giacomini. 2010. "Pharmacogenomics of Membrane Transporters: Past, Present and Future." *Pharmacogenomics* 11 (4): 475–79.

Zanger, Ulrich M., and Matthias Schwab. 2013. "Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation." *Pharmacology & Therapeutics* 138 (1): 103–41.

Zhang, Bo, Qin Li, Chenyang Shi, and Xinyue Zhang. 2018. "Drug-Induced Pseudo-allergy : A Review of the Causes and Mechanisms" 310013: 104–10.

Zhou, Kaixin, and Ewan R. Pearson. 2012. "Insights from Genome-Wide Association Studies of Drug Response." *Annual Review of Pharmacology and Toxicology* 53 (1): 299–310.

Zhou, Yitian, Souren Mkrtchian, Masaki Kumondai, Masahiro Hiratsuka, and Volker M. Lauschke. 2019. "An Optimized Prediction Framework to Assess the Functional Impact of Pharmacogenetic Variants." *Pharmacogenomics Journal* 19 (2): 115–26.

Zimmermann, Michael, Maria Zimmermann-Kogadeeva, Rebekka Wegmann, and Andrew L. Goodman. 2019. "Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes." *Nature* 570 (7762): 462–67.

ACKNOWLEDGMENTS

My sincerest and deepest gratitude goes to my dear supervisor Prof. Lili Milani. You are the best supervisor in the world and I believe it's not at all a biased opinion, but a true (statistically) significant statement;)! It honestly feels impossible to try to express in words on how thankful I am for all of your wisdom, continuous support and warm care. I'll always remain grateful that I had you as my supervisor and during my PhD I have learned so much from you on how to be better both in science and in life in general. You have been my mentor and forever inspire me as scientist and are always my role model as a person. I'm beyond grateful for your belief in me and finally all I can say is, thank you so much for being you, dearest Lili♥!

My heartfelt thank you to my second supervisor, Prof. Andres Metspalu, for the warm welcome to the Estonian Genome Center family and for providing this amazing environment and opportunities for doing science! I'll always be thankful that you gave me this chance to study and work here and that you continuously supported, protected and motivated with kind words and dared us PhD students to dream bigger.

I'm profoundly grateful and indebted to you, dear Prof. Reedik Mägi. Thank you for all the help and knowledge you were willing to share with me and for teaching me so much. As I've said before, you are like a fairy godfather to us PhD students, always ready to support and with a bag full of solutions to every problem. Just know that I believe you are one of the best persons I know and I'm so grateful I had this opportunity to learn from you.

My thanks to all the wonderful co-authors, it has been a pleasure to do science with you all and I'm grateful for everything that I've learned from each of you. I particularly would like to thank Sulev Reisberg and Jonas Bovijn. Sulev, I liked that you were always so determined and hardworking during our project and at the same time all your snarky comments and humour made working with you so fun. Know that I truly appreciate and always be grateful for all the kind words and support I got from you! Jonas, I'm grateful that I had the opportunity to work with such an inspiring co-author and I greatly appreciate all the knowledge you provided.

My warm gratitude goes to all the colleagues in the family of Estonian genome Center. It has been truly a delight to work with you all! I would like to thank Elin Org for being such a wonderful and supportive office mate for the first years, I truly miss all your fun stories. Dear Kelli Lehto, thank you for being so sweet and caring and for bringing in Fika tradition! I would especially like to thank our sweet Ene Mölder, Krista Liiv and Helene Alavere for all of their support and care for us PhD students. I would like to extend warm thanks to all the people in our dear IT-room – I truly appreciate all the essential help we got for doing science. Finally, thank you sincerely dearest Liis Karo-Astover! I truly value all the support you gave me and thank you greatly for always listening, caring and helping!

I'm most grateful that I'm surrounded by you all amazing fellow PhD students at the Estonian Genome Center. I truly think the world of you! I'm thankful that we always support and are there for each other and I'm glad that I had the opportunity to share this journey with you. I especially would like to thank you, sweetest Marili, for being my support person at work and often my rock. You always knew what I needed to hear and I'm eternally grateful for your friendship. Dearest Kreete, Anette, Nele and Kertu – you are all so wonderful and motivating individuals. I admire you all and I'm very grateful for all of you! Also thank you Anette for treating me with all the chocolate and cola when I was writing my thesis and Kreete for knitting the best socks in the world and thank you both for being so sweet and dear friends!

My utmost gratitude goes to my two favourite statisticians Dr. Silva Kasela and Dr. Kristi Läll, for all of your support, wisdom and friendship. I'm forever grateful that you sort of took me under your wing during the first years of my PhD and were always willing to help. My dear Swedish Estonians- I'm so glad that I got to know you dear Kristel and Loora, you guys are wonderful. Cheers to all my Captain Obvious friends in Tallinn, thank you guys for all of your belief in me, you are the greatest friends! My dear Heli, Merka and Triin, it makes me feel warm when I think about all the continuous support you give me, thank you sweeties.

Finally, I would not be here where I am today without my dear family! Dear brother, my favourite person in the world, thanks for always having my back and forever believing in me and continuously lifting my spirit – you are my hero! Dearest Kerli, thank you so much for all the kind and motivating words. I'm very grateful for all of your support. Sweet Liisi– my borrowed sister – thank you for being there for me no matter what, for all the confidence boosts and for helping me through my fears and worries. Sweetest Kai and Peep, my sincerest gratitude for all of your support and all the belief in me. Dear dad, thank you for believing that I can accomplish anything I work for and for always trying to convince mom to send all the food from the refrigerator with me to Tartu. I'm thankful for all of your support. Dear mom, you are the greatest mom in the world! I'm forever grateful and I cannot put it into words, what your support means to me – thank you for your unconditional love!

Lastly, my dear Ryan, thank you for coming into my life and making me smile.

PUBLICATIONS

CURRICULUM VITAE

Name:	Kristi Krebs
Date of birth:	January 9, 1990
Contact:	Estonian Genome Center, Institute of Genomics, University of Tartu Riia 23b, 51010, Tartu
E-mail:	kristi.krebs@ut.ee

Education:

2015–…	PhD in Molecular Biomedicine, Institute of Molecular and Cell Biology, University of Tartu, Estonia
01/2018–12/2018	Guest doctoral student at Department of Medical Sciences, Uppsala University, Sweden
2012–2014	Master's Degree in Gene Technology (*cum laude*), Faculty of Science, Tallinn university of technology
2009–2012	Bachelor's Degree in Gene Technology, Faculty of Science, Tallinn university of technology

Professional employment:

2017–…	Estonian Genome Center, Institute of Genomics, University of Tartu, Specialist
2014–2015	Repforce, Health consultant/ Drug representative
2012–2014	Tallinn University of Technology Faculty of Science, Department of Gene Technology, Chair of Molecular Biology, Specialist

Administrative work:

2012–…	Estonian Society of Human genetics member

Publications:

Krebs, K.*, Bovijn, J.*, Lepamets, M., et al. (2020). Genome-wide study identifies association between HLA-B*55:01 and penicillin allergy. *BioRxiv* 2020.02.27.967497.

Bovijn, J.*, **Krebs, K**.*, Chen, C.-Y.*, Boxall, R.*, et al. (2020). Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics. *Sci. Transl. Med*. 12 (549).

Zhou, Y., **Krebs, K.,** Milani L., Lauschke V.M, (2020). Global frequencies of clinically important HLA alleles and their implications for the cost-effectiveness of preemptive pharmacogenetic testing. *Clinical Pharmacology & Therapeutics,* June, cpt.1944.

Krebs, K, Milani, L (2019). Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good. *Human Genomics*, 13 (1), 39–39.

Reisberg, S; **Krebs, K**; Lepamets, M; et al.; (2019). Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. *Genet Med* 21, 1345–1354.

Tasa, T,*; **Krebs, K***.; Kals, M.; Mägi, R.; Lauschke, V.M.; Haller, T.; Puurand, T.; Remm, M.; Esko, T.; Metspalu, A.; Vilo, J.; Milani, L. (2019). Genetic variation in the Estonian population: pharmacogenomics study of adverse drug effects using electronic health records. *European Journal of Human Genetics*, 442–454.

Palover, M; **Krebs, K**; Milani, L; Tõnisson, N. (2019) Kuidas geenivaramu saab toetada rahvastikupõhist terviseedendust, haiguste ennetust ja varast avastamist. *Eesti Arst*.

Krebs, K; Ruusmann, A; Simonlatser, G; Velling, T (2015). Expression of FLNa in human melanoma cells regulates the function of integrin α1β1 and phosphorylation and localisation of PKB/AKT/ERK1/2 kinases. *European Journal of Cell Biology*, 564–575.

* These authors contributed equally

Supervised dissertations:

Pille-Riin Vare	Master's Degree, Institute of Mathematics and Statistics, The Impact of Genetic Variations for Antidepressant Use Based on Estonian Genome Center's Data
Pauliina Ranne	Bachelor's Degree, Institute of Molecular and Cell Biology, University of Tartu, Estonia. The benefits and potential of using pre-emptive pharmacogenetical testing in Estonian medical system
Karin Kõrgesaar	Bachelor's Degree, Institute of Molecular and Cell Biology, University of Tartu, Estonia. Identification of genetic variation in drug metabolizing cytochrome P450 genes in Estonian population

Awards and stipends:

2019 ESPT congress "Precision medicine and personalised health", Best poster award

2019 Scholarship of Graduate School in Biomedicine and Technology; Poster presentation in the ESPT congress "Precision medicine and personalised health",

2018 Dora Plus, Scholarship for study Mobility of Doctoral Degree Students

2017 Kristjan Jaak Scholarship for foreign visits

2017 ESHG conference fellowship for European countries

2016 ECNP seminar award

2016 Finalist and nominee of the lecture competition "Science in 3 minutes" of Estonian Academy of Sciences (topic- "Drug prescription- signed with genetics")

2016 Scholarships in smart specialisation growth areas

ELULOOKIRJELDUS

Nimi:	Kristi Krebs
Sünniaeg:	9.jaanuar.1990
Aadress:	Eesti Geenivaramu, Genoomikainstituut, Tartu Ülikool Riia 23b, 51010, Tartu
E-post:	kristi.krebs@ut.ee

Haridus:

2015–…	Doktoriõpe, molekulaar- ja rakubioloogia, loodus- ja täppisteaduste valdkond, Tartu Ülikool
01/2018–12/2018	Külalisdoktorant, Molekulaarse meditsiini õppetool, Uppsala Ülikool, Rootsi
2012–2014	Magistriõpe (*cum laude*) Geenitehnoloogia, Loodusteaduskond, Tallinna Tehnikaülikool
2009–2012	Bakalaureuseõpe, Geenitehnoloogia, Loodusteaduskond, Tallinna Tehnikaülikool

Teenistuskäik:

2017–…	Eesti Geenivaramu, Genoomika instituut, Tartu Ülikool, Spetsialist
2014–2015	Repforce, Tervisenõustaja/ravimiesitleja
2012–2014	Tallinna Tehnikaülikool, Loodusteaduskond, Geenitehnoloogia instituut, Spetsialist

Teadusorganisatsiooniline ja- administratiivne tegevus:

2012–…	Eesti Inimesegeneetikaühingu liige

Teaduspublikatsioonid:

Krebs, K.*, Bovijn, J.*, Lepamets, M., et al. (2020). Genome-wide study identifies association between HLA-B*55:01 and penicillin allergy. *BioRxiv* 2020.02.27.967497.

Bovijn, J.*, **Krebs, K**.*, Chen, C.-Y.*, Boxall, R.*, et al. (2020). Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics. *Sci. Transl. Med*. 12 (549).

Zhou, Y., **Krebs, K.,** Milani L., Lauschke V.M, (2020). Global frequencies of clinically important HLA alleles and their implications for the cost-effectiveness of preemptive pharmacogenetic testing. *Clinical Pharmacology & Therapeutics,* June, cpt.1944.

Krebs, K, Milani, L (2019). Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good. *Human Genomics*, 13 (1), 39–39.

Reisberg,S; **Krebs, K**; Lepamets, M; et al.; (2019). Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. *Genet Med* 21, 1345–1354.

Tasa, T,; **Krebs, K**.; Kals, M.; Mägi, R.; Lauschke, V.M.; Haller, T.; Puurand, T.; Remm, M.; Esko, T.; Metspalu, A.; Vilo, J.; Milani, L. (2019). Genetic variation in the Estonian population: pharmacogenomics study of adverse drug effects using electronic health records. *European Journal of Human Genetics*, 442–454.

Palover, M; **Krebs, K**; Milani, L; Tõnisson, N. (2019) Kuidas geenivaramu saab toetada rahvastikupõhist terviseedendust, haiguste ennetust ja varast avastamist . *Eesti Arst*.

Krebs, K; Ruusmann, A; Simonlatser, G; Velling, T (2015). Expression of FLNa in human melanoma cells regulates the function of integrin α1β1 and phosphorylation and localisation of PKB/AKT/ERK1/2 kinases. *European Journal of Cell Biology*, 564–575.

* These authors contributed equally

Juhendatud väitekirjad:

Pille-Riin Vare	Magistritöö, 2019, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, matemaatika ja statistika instituut, "Geneetilise varieeruvuse mõju antidepressantide tarvitamisele TÜ Eesti Geenivaramu andmete põhjal"
Pauliina Ranne	Bakalaureusetöö, Geenitehnoloogia õppekaval, "Ennetava farmakogeneetilise testimise vajalikkus ning potentsiaal Eesti tervishoiusüsteemis"
Karin Kõrgesaar	Bakalaureusetöö, "Ravimivastuses oluliste tsütokroom P450 geenide varieeruvuse määramine Eesti populatsioonis"

Stipendiumid:

2019 ESPT kongress "Precision medicine and personalised health", parima posteri auhind
2019 Biomeditsiini ja biotehnoloogia doktorikool lähetustoetus; Posteri ettekanne ESPT kongress "Precision medicine and personalised health"
2017 Dora Pluss doktorantide õpirände stipendium
2017 Kristjan Jaagu välissõidu stipendium: ettekandega esinemine konverentsil
2017 ESHG konverentsi stipendium Euroopa riikidele
2016 ECNP seminari auhind
2016 Eesti Teaduste Akadeemia "Teadus 3 minutiga" loengute konkursi finalist ja laureaat (loengu teema "Ravimiretsept – allkirjastatud geneetikaga")
2016 Kõrghariduse erialastipendium nutika spetsialiseerumise kasvuvaldkondades

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.
62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.

63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.
83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.

84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli**. The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.

106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Krista Kaasik**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in India: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.
125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.

126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.

147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida***.** Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.

166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli*. Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.

187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.

207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.

225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae*. Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.
245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.

246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreo-nemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and inter-annual patterns. Tartu, 2014, 115 p.
263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.

265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p
282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.
284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.

285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd**. Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro**. Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus**. Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.
293. **Urmo Võsa**. MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation. Tartu, 2016, 163 p.
294. **Kristina Mäemets-Allas**. Studies on cell growth promoting AKT signaling pathway – a promising anti-cancer drug target. Tartu, 2016, 146 p.
295. **Janeli Viil**. Studies on cellular and molecular mechanisms that drive normal and regenerative processes in the liver and pathological processes in Dupuytren's contracture. Tartu, 2016, 175 p.
296. **Ene Kook**. Genetic diversity and evolution of *Pulmonaria angustifolia* L. and *Myosotis laxa sensu lato* (Boraginaceae). Tartu, 2016, 106 p.
297. **Kadri Peil**. RNA polymerase II-dependent transcription elongation in *Saccharomyces cerevisiae*. Tartu, 2016, 113 p.
298. **Katrin Ruisu**. The role of RIC8A in mouse development and its function in cell-matrix adhesion and actin cytoskeletal organisation. Tartu, 2016, 129 p.
299. **Janely Pae**. Translocation of cell-penetrating peptides across biological membranes and interactions with plasma membrane constituents. Tartu, 2016, 126 p.
300. **Argo Ronk**. Plant diversity patterns across Europe: observed and dark diversity. Tartu, 2016, 153 p.
301. **Kristiina Mark**. Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.
302. **Jaak-Albert Metsoja**. Vegetation dynamics in floodplain meadows: influence of mowing and sediment application. Tartu, 2016, 140 p.
303. **Hedvig Tamman**. The GraTA toxin-antitoxin system of *Pseudomonas putida*: regulation and role in stress tolerance. Tartu, 2016, 154 p.

304. **Kadri Pärtel**. Application of ultrastructural and molecular data in the taxonomy of helotialean fungi. Tartu, 2016, 183 p.
305. **Maris Hindrikson**. Grey wolf (*Canis lupus*) populations in Estonia and Europe: genetic diversity, population structure and -processes, and hybridization between wolves and dogs. Tartu, 2016, 121 p.
306. **Polina Degtjarenko**. Impacts of alkaline dust pollution on biodiversity of plants and lichens: from communities to genetic diversity. Tartu, 2016, 126 p.
307. **Liina Pajusalu**. The effect of CO_2 enrichment on net photosynthesis of macrophytes in a brackish water environment. Tartu, 2016, 126 p.
308. **Stoyan Tankov**. Random walks in the stringent response. Tartu, 2016, 94 p.
309. **Liis Leitsalu**. Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.
310. **Richard Meitern**. Redox physiology of wild birds: validation and application of techniques for detecting oxidative stress. Tartu, 2016, 134 p.
311. **Kaie Lokk**. Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue. Tartu, 2016, 127 p.
312. **Mihhail Kurašin**. Processivity of cellulases and chitinases. Tartu, 2017, 132 p.
313. **Carmen Tali**. Scavenger receptors as a target for nucleic acid delivery with peptide vectors. Tartu, 2017, 155 p.
314. **Katarina Oganjan**. Distribution, feeding and habitat of benthic suspension feeders in a shallow coastal sea. Tartu, 2017, 132 p.
315. **Taavi Paal**. Immigration limitation of forest plants into wooded landscape corridors. Tartu, 2017, 145 p.
316. **Kadri Õunap**. The Williams-Beuren syndrome chromosome region protein WBSCR22 is a ribosome biogenesis factor. Tartu, 2017, 135 p.
317. **Riin Tamm**. In-depth analysis of factors affecting variability in thiopurine methyltransferase activity. Tartu, 2017, 170 p.
318. **Keiu Kask**. The role of RIC8A in the development and regulation of mouse nervous system. Tartu, 2017, 184 p.
319. **Tiia Möller**. Mapping and modelling of the spatial distribution of benthic macrovegetation in the NE Baltic Sea with a special focus on the eelgrass *Zostera marina* Linnaeus, 1753. Tartu, 2017, 162 p.
320. **Silva Kasela**. Genetic regulation of gene expression: detection of tissue- and cell type-specific effects. Tartu, 2017, 150 p.
321. **Karmen Süld**. Food habits, parasites and space use of the raccoon dog *Nyctereutes procyonoides*: the role of an alien species as a predator and vector of zoonotic diseases in Estonia. Tartu, 2017, p.
322. **Ragne Oja**. Consequences of supplementary feeding of wild boar – concern for ground-nesting birds and endoparasite infection. Tartu, 2017, 141 p.
323. **Riin Kont**. The acquisition of cellulose chain by a processive cellobiohydrolase. Tartu, 2017, 117 p.
324. **Liis Kasari**. Plant diversity of semi-natural grasslands: drivers, current status and conservation challenges. Tartu, 2017, 141 p.

325. **Sirgi Saar**. Belowground interactions: the roles of plant genetic relatedness, root exudation and soil legacies. Tartu, 2017, 113 p.
326. **Sten Anslan**. Molecular identification of Collembola and their fungal associates. Tartu, 2017, 125 p.
327. **Imre Taal**. Causes of variation in littoral fish communities of the Eastern Baltic Sea: from community structure to individual life histories. Tartu, 2017, 118 p.
328. **Jürgen Jalak**. Dissecting the Mechanism of Enzymatic Degradation of Cellulose Using Low Molecular Weight Model Substrates. Tartu, 2017, 137 p.
329. **Kairi Kiik**. Reproduction and behaviour of the endangered European mink (*Mustela lutreola*) in captivity. Tartu, 2018, 112 p.
330. **Ivan Kuprijanov**. Habitat use and trophic interactions of native and invasive predatory macroinvertebrates in the northern Baltic Sea. Tartu, 2018, 117 p.
331. **Hendrik Meister**. Evolutionary ecology of insect growth: from geographic patterns to biochemical trade-offs. Tartu, 2018, 147 p.
332. **Ilja Gaidutšik**. Irc3 is a mitochondrial branch migration enzyme in *Saccharomyces cerevisiae*. Tartu, 2018, 161 p.
333. **Lena Neuenkamp**. The dynamics of plant and arbuscular mycorrhizal fungal communities in grasslands under changing land use. Tartu, 2018, 241 p.
334. **Laura Kasak**. Genome structural variation modulating the placenta and pregnancy maintenance. Tartu, 2018, 181 p.
335. **Kersti Riibak**. Importance of dispersal limitation in determining dark diversity of plants across spatial scales. Tartu, 2018, 133 p.
336. **Liina Saar**. Dynamics of grassland plant diversity in changing landscapes. Tartu, 2018, 206 p.
337. **Hanna Ainelo**. Fis regulates *Pseudomonas putida* biofilm formation by controlling the expression of *lapA*. Tartu, 2018, 143 p.
338. **Natalia Pervjakova**. Genomic imprinting in complex traits. Tartu, 2018, 176 p.
339. **Andrio Lahesaare**. The role of global regulator Fis in regulating the expression of *lapF* and the hydrophobicity of soil bacterium *Pseudomonas putida*. Tartu, 2018, 124 p.
340. **Märt Roosaare**. *K*-mer based methods for the identification of bacteria and plasmids. Tartu, 2018, 117 p.
341. **Maria Abakumova**. The relationship between competitive behaviour and the frequency and identity of neighbours in temperate grassland plants. Tartu, 2018, 104 p.
342. **Margus Vilbas**. Biotic interactions affecting habitat use of myrmecophilous butterflies in Northern Europe. Tartu, 2018, 142 p.
343. **Liina Kinkar**. Global patterns of genetic diversity and phylogeography of *Echinococcus granulosus* sensu stricto – a tapeworm species of significant public health concern. Tartu, 2018, 147 p.

344. **Teivi Laurimäe**. Taxonomy and genetic diversity of zoonotic tapeworms in the species complex of *Echinococcus granulosus* sensu lato. Tartu, 2018, 143 p.
345. **Tatjana Jatsenko**. Role of translesion DNA polymerases in mutagenesis and DNA damage tolerance in Pseudomonads. Tartu, 2018, 216 p.
346. **Katrin Viigand**. Utilization of α-glucosidic sugars by *Ogataea* (*Hansenula*) *polymorpha*. Tartu, 2018, 148 p.
347. **Andres Ainelo**. Physiological effects of the *Pseudomonas putida* toxin grat. Tartu, 2018, 146 p.
348. **Killu Timm**. Effects of two genes (DRD4 and SERT) on great tit (*Parus major*) behaviour and reproductive traits. Tartu, 2018, 117 p.
349. **Petr Kohout**. Ecology of ericoid mycorrhizal fungi. Tartu, 2018, 184 p.
350. **Gristin Rohula-Okunev**. Effects of endogenous and environmental factors on night-time water flux in deciduous woody tree species. Tartu, 2018, 184 p.
351. **Jane Oja**. Temporal and spatial patterns of orchid mycorrhizal fungi in forest and grassland ecosystems. Tartu, 2018, 102 p.
352. **Janek Urvik**. Multidimensionality of aging in a long-lived seabird. Tartu, 2018, 135 p.
353. **Lisanna Schmidt**. Phenotypic and genetic differentiation in the hybridizing species pair *Carex flava* and *C. viridula* in geographically different regions. Tartu, 2018, 133 p.
354. **Monika Karmin**. Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver**. Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly**. The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.
360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.
361. **Erwan Pennarun**. Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.
362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.

364. **Anne-Mai Ilumäe**. Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte**. Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts**. The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.
368. **Viktorija Kukuškina**. Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers. Tartu, 2020, 169 p.
369. **Martti Vasar**. Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur**. Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.
371. **Helen Post**. Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks. Tartu, 2020, 143 p.