DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
404

KATRI PÄRNA

Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
404

DOCTORAL DISSERTATION
UNIVERSITY OF GRONINGEN

KATRI PÄRNA

Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes

UNIVERSITY OF TARTU
Press
1632

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

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on July 4, 2022 by the Council of the Institute of Molecular and Cell Biology, Faculty of Sciences and Technology, University of Tartu.

Supervisors: Luca Pagani, PhD; Visiting Professor, Department of Biology, University of Padova, Italy; Senior Research Fellow of Population Genetics, Institute of Genomics, University of Tartu, Estonia

Davide Marnetto, PhD; Researcher, Department of Neurosciences "Rita Levi Montalcini", University of Turin, Italy

Harold Snieder, PhD; Professor, Department of Epidemiology, Unit of Genetic Epidemiology and Bioinformatics, University of Groningen, University Medical Center Groningen, the Netherlands

Ilja M. Nolte, PhD; Senior Researcher, Department of Epidemiology, Unit of Genetic Epidemiology and Bioinformatics, University of Groningen, University Medical Center Groningen, the Netherlands

Krista Fischer, PhD; Professor of Mathematical Statistics, Institute of Mathematics and Statistics, University of Tartu, Estonia

Reedik Mägi, PhD; Professor of Bioinformatics, Institute of Genomics, University of Tartu, Estonia

Opponent: Ryan Daniel Hernandez, PhD; Professor, Bioengineering & Therapeutic Sciences, School of Pharmacy, University of California, San Francisco, United States

Commencement: Aula of University of Groningen, Broerstraat 5, Groningen, the Netherlands, on 7th of September 2022, at 12.45 pm

Publication of this thesis is granted by the Institute of Molecular and Cell Biology and the Institute of Genomics, University of Tartu.

ISSN 1024-6479
ISBN 978-9949-03-979-1 (print)
ISBN 978-9949-03-980-7 (pdf)

Copyright: Katri Pärna, 2022

University of Tartu Press
www.tyk.ee

university of groningen

UNIVERSITY OF TARTU
1632

Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans

and

to obtain the degree of PhD at the
University of Tartu
on the authority of the
Rector Magnificus Prof. T. Asser
and in accordance with
the decision by the Council of the Institute of Molecular and Cell Biology.

Double PhD degree

This thesis will be defended in public on

Wednesday 7 September 2022 at 12.45 hours

by

Katri Pärna

born on 23 April 1989
in Tartu, Estonia

Supervisors
Prof. H. Snieder

Dr. L. Pagani

Co-supervisors
Dr. I.M. Nolte

Dr. D. Marnetto

Assessment committee
Prof. N. Barban

Prof. R.D. Hernandez

Prof. J.A. Kuivenhoven

Paranymphs

O. Minaeva

T. Xie

TABLE OF CONTENTS

GLOSSARY

Complex trait/disease	Trait or a disease that is influenced by multiple (genetic and non-genetic) factors.
Heritability	The proportion of variation in a trait or a disease, which is explained by genetic variation between individuals.
Genome-wide association study	Study design testing millions of genetic variants over the whole genome without an *a priori* hypothesis to detect associations between these variants and phenotype.
Single nucleotide polymorphism	Substitution of one nucleotide with another one in a DNA sequence, which occurs with at least 1% frequency in a population.
Polygenic Risk Score	A measure summarizing person's estimated genetic risk for a trait/disease based on GWAS effect sizes and person's genetic data.
Effect size	A statistical measure showing the strength of an association between, e.g., a genetic variant and an outcome.
Population structure	The occurrence of systematic allele frequency differences between populations due to evolutionary processes (e.g. migration, non-random maiting).
Principal Component Analysis	A statistical dimensionality reduction method to better summarize the dataset.
Genetic admixture	Exchange of genes between two previously isolated populations.
Genetic ancestry	Origin of genetic material from specific descendants.
Epigenetics	A field of study, which involves modifications on top of DNA, which are involved in gene expression, but do not change the DNA sequence.
Personalized medicine	A field in medicine, which aims to improve stratification and timing of health care by using individual's genetic and non-genetic information to result in better prevention, prediction or treatment of a disease.

ABBREVIATIONS

aspPS	Ancestry-specfic partial polygenic score
BMI	Body mass index
casPS	Combined ancestry-specific polygenic score
DNA	Desoxyribonucleic acid
dwGRS	Doubly-weighted genetic risk score
EstBB	Estonian Biobank
EWAS	Epigenome-wide association study
FPG	Fasting plasma glucose
GRS	Genetic risk score
GWAS	Genome-wide association study
HbA1c	Glycated hemoglobin
LAD	Local ancestry deconvolution
LD	Linkage disequilibrium
Lifelines	Lifelines Cohort Study and Biobank
Meta-GWAS	Meta-analysis of genome-wide association study
MS	Methylation score
PC	Principal component
PCA	Principal component analysis
pPS	Partial polygenic score
PRS	Polygenic risk score
SNP	Single nucleotide polymorphism
swGRS	Single-weighted GRS
T2D	Type 2 diabetes

NOTES FOR THE READER

* As the science is always a result of a good team work not an individual effort, the author of this thesis chose to use ‘We’ instead of ‘I’ in the parts describing included chapters. However, the author can be considered responsible for the research design, realization, analyses, and interpretation of results.

** There is no absolute agreement, which scientific term to use for polygenic risk score (PRS). Therefore, throughout this thesis the term varies according to the corresponding publications. We have used terms as doubly-weighted GRS and polygenic score (PS), which both indicate the PRS calculation via applying several p-value thresholds as described in the paragraph ‘polygenic risk scores’.

GENERAL INTRODUCTION

Type 2 diabetes

Type 2 Diabetes (T2D) is a chronic metabolic disease characterized by elevated blood glucose levels due to the body's ineffective use of insulin, which is responsible for glucose uptake in liver, fat, and muscle[1,2]. The less responsive these tissues become to insulin, i.e., insulin resistance, the more insulin is produced by pancreatic beta cells till these cells are exhausted by the high production of insulin leading to their progressive deterioration[3]. Thus, β-cell deterioration and insulin resistance are the main causes of T2D. T2D predisposes to co-morbidities such as cardiovascular and renal diseases and other long-term complications such as retinopathy and neuropathy or even limb amputation if appropriate and timely treatment is not administered[4]. Along with these complications T2D leads to lower quality of life and it may result in premature mortality with a 5–10 years lower life expectancy[4].

Prevalence and diagnosis of type 2 diabetes

Currently there are approximately 537 million adults between 20 and 79 years old diagnosed with diabetes (T2D accounts for approximately 90% of the diabetes cases) and this number is projected to rise to 643 million by the year 2030 and 783 million by the year 2045 (Figure 1)[5].

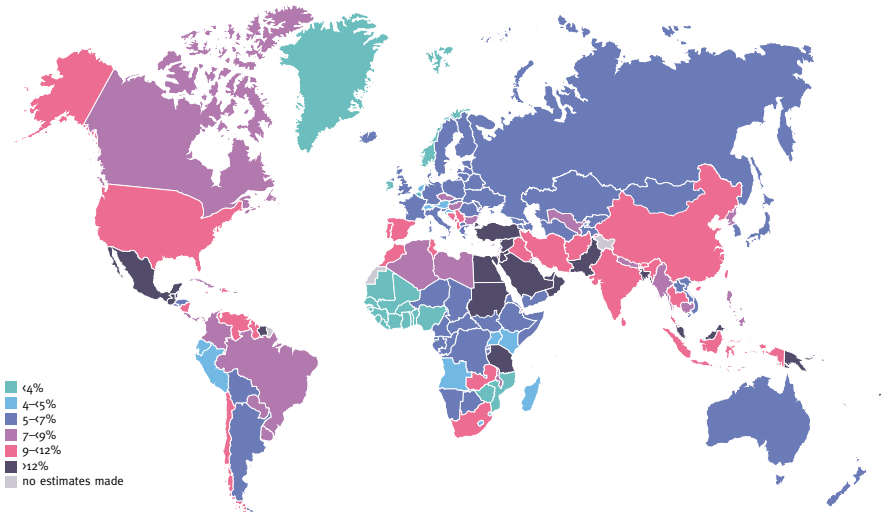

Figure 1. Estimated age-adjusted comparative prevalence of diabetes in adults (20–79 years) in 2021 (IDF, 2021).

Furthermore, due to slow progression of hyperglycemia, i.e., a condition with excessive levels of glucose in blood, the clinical symptoms of T2D are often mild or absent. Therefore, depending on the country, it is estimated that 30–80% of the T2D cases remain undiagnosed[6]. In the year 2021, diabetes caused approximately 6.7 million deaths often accompanied by other comorbidities[5].

T2D is typically diagnosed if one of the following criteria is met: fasting plasma glucose (FPG) > 7.0 mmol/L and/or glycated hemoglobin (HbA1c) ≥ 6.5% or 2-hour plasma glucose ≥ 11.1 mmol/L[4]. FPG is the blood glucose level measured after at least eight hours of fasting. HbA1c is a form of hemoglobin that is chemically linked to glucose molecules. Since glucose molecules are not prone to form a chemical bond with hemoglobin, elevated levels of HbA1c indicate increased blood glucose levels. Hba1c represents the average plasma glucose of approximately the last three months[7]. 2-hour plasma glucose is the measure of blood glucose level 2 hours after taking the 75-gram oral glucose tolerance test[6].

Complex traits

T2D is a common complex disease caused by genetic and non-genetic (e.g. environment and lifestyle) risk factors, and by the interactions between them[1]. Its complex nature is similar to that of height and body mass index (BMI), which allows us to draw parallels between the findings from the studies for BMI and height and those for T2D. Therefore, in the current thesis, BMI and height were used as complex model traits, which have several advantages over T2D. First, based on both twin and family studies, the heritability estimates for height reach to 90%[8,9], while for BMI a larger environmental contribution is observed with heritability estimates ranging from 30 to 90% depending on the study design[10,11]. Second, height and BMI are based on standard measures that are relatively easy to collect, therefore it is more certain that databases (biobanks and other data repositories) have these measures available. Third, in genetic studies complex continuous model traits are often preferred over the dichotomous ones since these require much smaller sample size to reach the same statistical power to detect new genetic loci[12,13], which makes these studies more feasible to conduct.

Risk factors for T2D

In the current thesis risk factors of T2D are divided into '*non-genetic*' and '*genetic*' risk factors. Historically, '*non-genetic*' risk factors are the most explored and established ones. Examples include BMI, age, and lifestyle habits such as smoking, alcohol consumption, and unhealthy diet[4]. Here these are called "*non-genetic*" although it is well known that most of these factors also have a genetic component as described above for BMI. Identifying additional risk factors, including specific genetic loci or variants, would improve the understanding of etiology of T2D and help to ease its societal and individual burden.

Non-genetic, established risk factors

For T2D, the main risk factor is obesity, which is often caused by urbanization, less active lifestyle, and higher intake of unhealthy food[4]. The fact that there has been a rapid, simultaneous increase in the number of obese individuals and T2D cases is an indicator of their intertwined nature. In research, obesity is generally represented by a BMI (units: weight(kg)/height(m)^2) equal to 30 or higher[14].

Besides obesity, risk of having T2D increases with age due to the simultaneous decrease in insulin sensitivity and it has been shown that the pancreas is not able to renew beta cells beyond the age of 30[3,15]. Therefore, it could be that the increasing numbers of T2D cases are partly explained by the world's aging population[16]. For example, there is a consistent increase in T2D prevalence by age reaching to its highest value for the age group of 50 to 59 years[5]. Nevertheless despite that the abovementioned non-genetic risk factors are well established, they do have variable effects on different individuals, e.g. there are many obese individuals who doesn't get T2D, while some non-obese people do. This could be explained by differences in genetic susceptibility[1].

Genetic risk factors

It has been shown that genetic factors play a major role in T2D with heritability estimates ranging from 30–69% depending on the study design[17,18]. In large proportion these estimates also contain the heritability of obesity, since around 90% of the individuals with T2D are overweight (defined as BMI≥25kg/m^2) or obese[1,19]. Due to rapid methodological advancements and high heritability of T2D, genetic risk factors are currently thoroughly investigated[20,21]. Their inclusion in disease prediction models is becoming more common since genetic factors are fixed from the birth onwards and are seen as longer-term predictors compared to the non-genetic risk factors, which often occur later in life such as weight gain or rise in blood sugar levels.

Genetic variation and their discovery

The most common type of genetic variants are single nucleotide polymorphisms (SNPs), the replacement of one deoxyribonucleic acid (DNA) base pair (nucleotide) with another one in a specific location in the genome occurring with a frequency of at least 1% in the population[22]. On average one human genome differs from the reference genome on approximately 4 to 5 million SNPs[23]. SNPs are an important source of differences in genetic disease susceptibility[24]. Therefore, in a clinical context, SNPs are often used to represent our genetic risk for a certain phenotype. To detect SNPs involved in the pathogenesis of complex traits and diseases, the primary method is a genome-wide association study (GWAS), which aims to detect genotype-phenotype associations by testing millions of genetic variants over the whole genome without an *a priori* hypothesis[25]. In GWAS each SNP association with the complex trait or disease is independently tested. Therefore, there is a high multiple testing burden and the significance of

the SNP needs to be very low ($p<5\times10^{-8}$) for it to be regarded as a true positive association. Thus, to improve the statistical power, often the GWASs are combined into a meta-analysis of GWASs (meta-GWAS)[26].

The first GWAS for T2D was conducted in 2007 in a French cohort of 661 T2D cases and 614 controls with information available for approximately 400,000 SNPs. Back then only five significantly associated genetic loci were detected[27]. In 2020, so far the largest meta-GWAS for T2D was published including five ancestral groups with over 1.4 million individuals from which approximately 16% were T2D cases. Millions of variants were tested and 568 genomic regions associated with T2D were detected[28]. Such increases in GWAS sample size and improvements in genotyping techniques have resulted in a rapid escalation of the number of SNPs identified for T2D, although each associated SNP individually only has a small effect on a polygenic disease such as T2D[29,30]. Therefore, often these variants are combined into a single measure called genetic or polygenic risk score (GRS or PRS, respectively).

Polygenic risk scores

A PRS is a measure combining genetic risk across the genome and therefore representing each person's genetic susceptibility for a certain trait or a disease[31]. It is calculated by summing up the copies of genetic risk variants weighted by their effect sizes obtained from earlier GWASs or meta-GWASs[31]. Initially the risk score included only genome-wide significant SNPs ($p<5\times10^{-8}$) and it was called a Genetic Risk Score (GRS). The PRS allows more lenient p-value thresholds to include more SNPs not reaching the genome-wide significance level due to insufficient statistical power. Such a PRS improves the variance explained for the outcome trait[32]. Although the PRS has demonstrated its high potential for the future application in clinical practice by detecting individuals in different risk categories (Figure 2)[33–35], it still has some limitations. For example, it has been recognized that if GWAS summary statistics are used for a PRS calculation in a population with a genetic population structure different from that of the discovery cohort, the PRS has much lower predictive value (also called *transferability or generalizability issue*)[36,37].

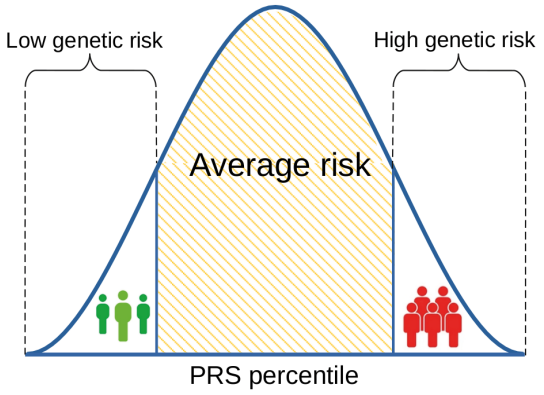

Figure 2. Risk stratification by PRS.

Population structure

Population structure is the presence of systematic allele frequency differences between (sub)populations due to genetic drift, non-random mating and recent migration processes[38]. Such variation in allele frequencies has been detected within and between different populations due to their unique history[39]. Therefore, population structure has remained a main confounder for genetic association studies and it is still under-explored[40]. There are several methods to account for population structure such as Principal Component Analysis (PCA)[41], Genomic Control (GC)[42], Linear Mixed Models[43] or Linkage Disequilibrium Score Regression (LDSC)[44]. However some argue that these methods do not correct for it completely[45–47]. One explanation for this could be that individuals included in GWASs are often assumed to originate from genetically homogeneous populations, which means that only individuals belonging to the largest ancestry group are included[48]. So far, most of the GWASs (~80%) have been conducted in European populations, but when using these European-based GWAS effect sizes for calculation and application of PRSs in non-Europeans, the predictability becomes much lower or even inaccurate[36,48]. Although, population structure has been demonstrated at the continental level[36,37,49], several studies have shown its existence also on a finer scale[50–54]. This indicates that populations are genetically more heterogeneous than expected, including the GWAS discovery cohorts, due to evolutionary processes such as admixture, selection, and non-random mating[47]. Hence, it is clear that the population structure is a confounder for genetic studies, but the challenge is to remove it and not to make wrong conclusions due to existing population structure. Therefore, two of the chapters in this thesis focused on how to account for population structure on a finer-scale, among Europeans and for the admixed individuals in the PRS construction in order to improve the transferability or generalizability issue.

Genetic admixture

Genetic admixture occurs when individuals from previously separated populations intermix and their offspring will carry the genetic information of both populations[55]. Due to past events in human history genetic admixture has resulted in ancestry differences between populations, between individuals from one population, and even within one human genome (Figure 3).

As a result of admixture events each person's genome is like a mosaic of segments originating from different ancestries ('*genetic ancestry*')[55]. Especially modern human populations are becoming more mixed, for example in large metropolises, which are major melting pots for people originating from different ancestries. Such mosaic genomes result in a wide range of genetic and phenotypic variation, which are important to understand from an epidemiological perspective to more accurately predict and explain the differences in health outcomes[56].

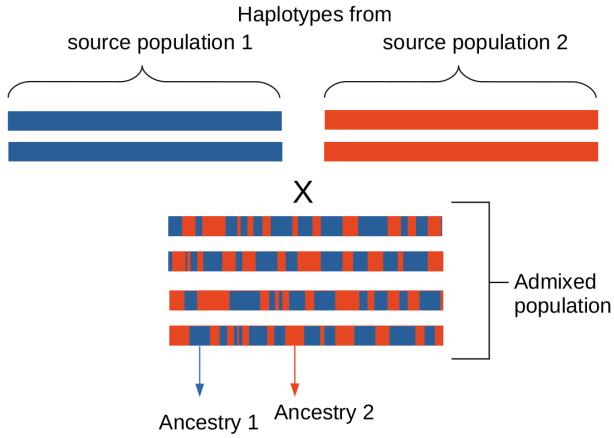

Figure 3. Admixture of two source populations, which after generations of recombinations result in admixed genomes in the following population containing part of the genetic info from source 1 (ancestry 1) and part from source 2 (ancestry 2).

Besides, current GWASs effect sizes might be population dependent due to the differences in linkage disequilibrium patterns, allele frequencies, rare variants and environmental effects[29,57,58], which makes studying admixed individuals' genomes with several ancestral backgrounds especially complicated. However, since admixture is one of the fastest evolutionary processes, it is a great mechanism to reveal differences in ancestral genetic variation related to disease[59]. For example, leveraging ancestral inference, a method to detect the genetic ancestry of a locus (*local ancestry inference*) or relative proportions of ancestry in a genome (*global ancestry inference*), may help overcoming confounding effects introduced by population specific LD patterns, hence pinpointing the true causative variants. Furthermore, such an ancestry informed approach may improve prevention and treatment especially for complex traits, where incorporating local ancestry inference has already shown promising results in increasing detection of more genetic associations and in improving the genetic prediction for admixed individuals[59,60].

Epigenetics

The recent rapid increases in worldwide prevalence of T2D cannot be explained by genetic components, since the population structure only changes minimally from generation to generation. Therefore, the scientists are exploring potential molecular mechanisms triggered by the environmental exposures and gene-environment interactions involved in T2D. Whereas genetic factors are fixed for life, epigenetic factors (those responsible for gene expression without altering the DNA sequence) are partly reversible by environmental and lifestyle factors[61,62]. Epigenetic mechanisms such as DNA methylation (the most commonly investigated epigenetic process, Figure 4) and histone modification are used by cells to regulate gene expression in response to environmental triggers[16].

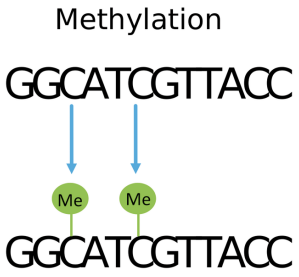

Figure 4. Methylation – addition of a methyl group on top of the DNA strand. It typically takes place at CpG sites, where cytosine is followed by guanine nucleotide.

Many studies have shown that epigenetic factors have a high potential to explain part of the T2D pathogenesis[61,63–65]. For example, common methylation patterns associated with T2D have been detected by epigenome-wide association studies (EWAS)[66,67]. These could lead to a better understanding of inter-individual differences in disease susceptibility due to the environmental and lifestyle factors involved in T2D pathogenesis[68]. Wahl and colleagues showed in 2017 that 62 methylation markers out of 187 that were associated with BMI, were also associated with incident T2D, offering promisine for using epigenetics markers in disease prediction[63]. To reduce the global burden of T2D and to better understand environmental factors and gene-environment interactions involved in the development of T2D, the inclusion of epigenetics in disease prediction should be further investigated.

Personalized prediction

Many studies have demonstrated the high potential of a PRS to stratify individuals into risk categories according to their genetics[33,35,69]. For example, individuals in the highest PRS risk quantile for incident T2D have been demonstrated to have 3 times higher risk of T2D than the individuals in the lowest PRS quantile[69]. Also, for breast and prostate cancer PRS has shown its ability to clearly distinguish individuals belonging to different risk categories[35,70]. Moreover, Khera and colleagues (2018) concluded that their PRS for coronary artery disease could even detect individuals at risk comparable to rare monogenic mutations with large effects[33]. Such predictions based on genetic information have only been made possible by major recent advances in the genomics field such as increasing GWAS sample sizes, improved coverage of the genome and the initiation of biobanks. It has been suggested that such PRS-based personalized prediction could lead to personalized medicine with the ultimate aim to postpone the onset of complex diseases such as T2D or even to prevent them via more frequent screening or better preventive strategies for high-risk individuals[33,69,71,72].

AIMS OF THIS THESIS

The main aim of this thesis was to improve the prediction of T2D by combining approaches from genetic epidemiology, population genetics, and epigenetics. The sub-aims were to improve the prediction by refining the PRS calculation, by addressing the PRS transferability issue and by adding an epigenetic component to the prediction of T2D. In addition we reviewed the latest advancements in the genomics field that pave the way towards personalized medicine.

OUTLINE OF THIS THESIS

Chapter 1 validates and evaluates the performance of the doubly-weighted GRS (dwGRS) in the Estonian Biobank and in the Lifelines Cohort Study and Biobank[69]. The dwGRS applies an additional weight for each included SNP to correct for the '*Winner's curse*' phenomenon compared to the traditional, single-weighted GRS (swGRS) for the prediction of incident T2D.

Chapter 2 explores the performance of the local ancestry deconvolution (LAD) software[73] in detecting the unique genetic tiling of admixed individuals via inferring their genetic ancestries. Next, these ancestral estimations are used to improve the calculations of PRSs resulting in the development of novel methods of the ancestry-specific partial PRS (aspPS) and the combined ancestry specific Polygenic Score (casPS). These PRSs aim to improve the personalized prediction for admixed individuals via combining the knowledge from ancestral estimates and publicly available GWAS summary statistics, which have been obtained from more homogeneous genomes. These methods are applied to the example traits of height and BMI and diseases such as T2D and breast cancer.

Chapter 3 focuses on Principal Component Analysis and its limitations while applied in GWAS to account for population structure. The effect sizes of SNPs estimated in one population are population dependent and cause a lower predictability when used in the calculation of the PRS for validation in a different population. Traditionally Principal Components (PCs) are calculated using population-specific genotype data. In this study it was tested whether calculating the PCs using projection onto those from a reference population for both the discovery and validation sample would mitigate the PRS transferability issues, and whether the adjustment for PCs in the PRS validation model would be necessary.

Chapter 4 investigates the associations between the Methylation Scores (MSs) and prevalent T2D and its glycemic endophenotypes, i.e., FPG and HbA1c. Besides the MSs, GRSs were calculated and their individual and combined effects on the outcomes were tested in order to evaluate their independent additive effect on the outcome.

Chapter 5 provides an overview of the latest advancements in the field of genomics and how these advancements result in more genetic discoveries leading the way towards higher genetic prediction accuracy and eventually the implementation of personalized medicine. EstBB was used as a prime example for which we described the challenges of implementing personalized medicine on a national level.

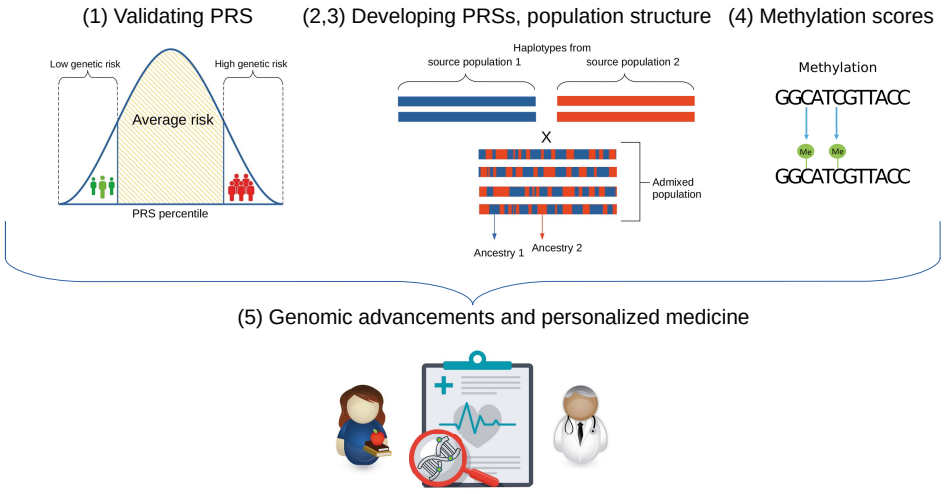

Figure 5. Illustrative outline of this thesis. Numbers indicate the chapters.

METHODS

Study sample/cohorts

Below is an overview of three large prospective European Biobanks, which data were used in this thesis. These biobanks share the goal of investigating genetic and non-genetic risk factors to improve the prediction, diagnosis, and treatment of diseases with a focus on common complex diseases. All three biobanks have been approved by their ethical committees and all the participants have signed informed consent[74,76,79].

The Lifelines Cohort Study

The Lifelines Cohort Study (Lifelines) is a multidisciplinary prospective population-based cohort study and biobank with a unique three-generation design examining the health and health-related behaviors of over 167,000 persons living in the North-Netherlands. Individuals between the ages 0 to 93 years were invited for participation in the study between 2006–2013 with the aim to follow them up for at least 30 years. Starting from baseline, every five years biomaterials are collected, a physical examination is done, and extensive questionnaires are completed. In between, participants fill in questionnaires approximately every 1.5–2.5 years[74,75]. Besides the information on sociodemographic, behavioral, mental, and psychosocial factors collected with the questionnaires and the exposome data, also genome-wide genetic data are currently available for 51,000 participants, but it is planned to have these data for all participants in the near future[75]. In the current thesis the Lifelines data from only the adult participants (≥18 years) have been used in Chapters 1 and 4.

Estonian Biobank

The Estonian Biobank (EstBB) is a prospective population-based biobank with the first wave of data collection from approximately 52,000 volunteers conducted between the years 2002–2011[76]. Participants were 18 years or older[77]. At baseline, extensive questionnaires, physical measures, and biomarkers were collected. Follow-up data are made available via linkage with the national health registries and via new examination of individuals. Besides, electronic health records containing phenotypic information are updated every six months[76,77]. Currently the EstBB has recruited more than 200,000 gene donors, which represent approximately 20% of the whole Estonian population[78]. EstBB data were used in Chapters 1, 2 and 3.

UK Biobank

The UK Biobank (UKBB) Project is a prospective population-based cohort study with data collected from approximately 500,000 individuals aged 40 to 69 years from across the United Kingdom at their recruitment visit between the years 2006 to 2010[79]. At baseline, a broad range of phenotypic, genetic, and health-related measures and information were collected. Genome-wide genotype data are available for all participants. Follow-up data are collected by web-based questionnaires and by data linkage to health and medical records[79,80]. UKBB data were used in Chapters 2 and 3.

Statistical analyses

Here are briefly introduced the core methods used and/or developed to improve the prediction for T2D. See each chapter for more details. A sidenote about genetic (GRS) and polygenic risk score (PRS) terms. There is no absolute agreement how to use these terms[24]. Similarly in chapters of this thesis I have used terms as doubly-weighted GRS and polygenic score (PS), which both indicate the PRS calculation via applying several p-value thresholds as described in the paragraph 'polygenic risk scores'.

Doubly-weighted GRS

In Chapter 1, the novel method of doubly-weighted GRS (dwGRS) was internally replicated in the EstBB and externally validated in Lifelines. The usual method for GRS calculation involves summing up independent genome-wide significant ($p<5\times10^{-8}$) genetic variants and weighing these by their effect sizes from an independent meta-GWAS. In this manner genetic variants, which are truly associated with the disease, but did not reach the genome-wide significance level due to low power of the meta-GWAS, are left out from the GRS resulting in suboptimal prediction performance. Therefore, the EstBB statistical research team developed a new method called the 'doubly-weighted GRS' (dwGRS)[69]. The dwGRS aims to overcome the *Winner's curse* (a phenomenon stating that the genome-wide significant SNP effects are overestimated by chance), by weighing each genome-wide significant SNP with an extra weight. This weight is the estimated probability ($\hat{\pi}_i$) that each specific genetic variant belongs to the set of top SNPs of pre-defined size showing the true association with the outcome. The probability estimate is obtained via a simulation approach, where a simulated effect size for each SNP is drawn from a normal distribution with the mean and standard deviation equal to the original effect size and standard error from the meta-GWAS, respectively. The $\hat{\pi}_i$ is then computed as the average score over 1000 repetitions. Equation 1 for dwGRS:

$$dwGRS = \sum_{i=1}^{N} \widehat{\pi}_i(1000)\widehat{\beta}_i X_i$$

Equation 1. Calculation of doubly-weighted GRS

$\widehat{\pi}_i(1000)$ – estimated probability for the *i*-th marker to belong to the set of 1000 top SNPs with the strongest effect on T2D received from simulation studies

$\widehat{\beta}_i$ – estimated logistic regression parameter for SNP *i* obtained from meta-analysis

X_i – allele dosage of *i*-th SNP

N – total number of SNPs included in the score

Ancestry-specific partial PS

In Chapter 2, to test the hypothesis that the GWAS summary statistics are partly population dependent, different formulas for PS calculations were developed for admixed individuals so that each part of the genome originating from a specific ancestry would receive a corresponding population GWAS effect sizes if available. Firstly, the formula of partial PS (pPS) was developed, which uses only part of the genome's genetic variants instead of the genetic variants across the whole genome. Equations 2 and 3 for the partial PS:

$$\bar{x}'_j = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{\beta}_i X_{ij}$$

Equation 2. Calculation of raw pPS.

$$\mathrm{pPS}_j = \frac{\bar{x}'_j - \mu_{\bar{x}'}}{\sigma_{\bar{x}'}}$$

Equation 3. Standardization of pPS.

$\bar{x}'_j$ – raw pPS for individual *j*

N_s – subset of the variants of the genome used for the pPS calculation

$\hat{\beta}_i$ – estimated effect size from the GWAS for SNP *i*

X_{ij} – allelic state at site *i* for individual *j*

$\sigma_{\bar{x}}$ – standard deviation of the $\bar{x}'$ statistic computed across all N_I individuals of a reference population while using only subset N_V of variants of the genome.

$\mu_{\bar{x}}$ – mean of the $\bar{x}'$ statistic computed across all N_I individuals of a reference population while using only subset N_V of variants of the genome.

pPSj – standardized partial polygenic score for individual *j*

Following, the Efficient Local Ancestry Inference (ELAI)[81] (a software package learning the structure of haplotypes) for local ancestry deconvolution (LAD), was used at first to infer the genetic ancestries of admixed study individuals: Egyptians, Ethiopians, African-American. For all these listed populations it is known that

they have both African (ancestry A) and West-Eurasian (ancestry B) background. The results from the LAD analysis helped to identify the proportions of the genomes that come from these ancestries A and B. Similarly, the LAD was applied to UKBB admixed individuals to detect proportions of their genomes coming from 'European', 'African', or 'East Asian' ancestries. Such a LAD allowed the improvement from the pPS into an ancestry-specific PS (aspPS), where only the detected genomic subsets related to the correct ancestry were used for pPS calculations. Detection of the correct ancestry for the genomic subset of SNPs allows applying the GWAS summary statistics based on the population most similar to its ancestry. For the individuals, where at least two aspPSs could be calculated, the main advantage is to combine these aspPS into combined ancestry-specific PS (casPS) weighted by their ancestry proportions.

Correcting GWAS and PS validation with projected PCs

In Chapter 3 GWASs were conducted for height and BMI in a subset of the UKBB. A projection approach for PC adjustment in a GWAS discovery set (UKBB subset) and in PRS target sets (independent UKBB and EstBB subsets) was tested. The PCs used for adjustment in GWAS and in PRS target sets were computed via projecting the GWAS discovery samples and the PRS target set samples onto the PC spaces of the reference dataset of 1000 Genomes and an external subset of the UKBB or EstBB sample, respective to the target set. The core of the projection approach is that only the external sample set is used to infer the eigenvectors of the PC space and the discovery/target set individuals are projected onto the generated PC space to obtain their PC coordinates. The hypothesis was that such an approach will better account for the population dependent nature of the GWAS effect sizes and lead to an improvement in PRS transferability between two populations.

Methylation Scores

In Chapter 4, MSs were calculated by 1) regressing out the methylation plate and position on each epigenome-wide significant ($p<1\times10^{-7}$) CpG site and then, 2) summing up these epigenome-wide significant CpG residuals weighted by their effect sizes from EWASs. Next to the MSs GRSs were calculated by summing up the weighted genome-wide significant ($p<5\times10^{-8}$) SNPs to represent the person's genetic risk for a disease. Equation 4 for methylation score and equation 5 for genetic risk score:

$$MS_j = \sum_{k}^{K} \hat{\beta}_k \, cpg_{kj}$$

Equation 4. Calculation of methylation score

cpg_{kj} – standardized residualized methylation level for individual *j* and cpg site *k*
$\hat{\beta}_k$ – effect size estimated for the *k*-th cpg site from the EWAS
K – number of CpG sites included in the methylation score

$$GRS_j = \sum_{k}^{K} \hat{\beta}_k \, X_{kj}$$

Equation 5. Calculation of genetic risk score

X_{kj} – allele dosage for *k*-th SNP and *j*-th individual
$\hat{\beta}_k$ – estimated effect size from GWAS for SNP *k*

Author contribution to the current thesis chapters

Chapter 1 I ran all the required analyses, interpreted the results, designed the figures, and drafted the manuscript.

Chapter 2 I helped to run part of the PS analyses, prepared data and figures and contributed in the revision of the manuscript.

Chapter 3 I ran all the required analyses, interpreted the results, designed most of the figures, drafted the manuscript.

Chapter 4 I provided the analysis plan and scripts for all the sub-cohorts, ran the analyses in the LL pT2D sub-cohort, interpreted and summarized the results from all the sub-cohorts, designed the figures, and drafted the manuscript.

Chapter 5 I helped to prepare figures and co-wrote the manuscript.

REFERENCES

1. Ali, O. Genetics of type 2 diabetes. *Current Science* **4**, 114–123 (2013).
2. Stumvoll, M., Goldstein, B. J. & Haeften, T. W. Van. Type 2 diabetes : principles of pathogenesis and therapy. *Lancet* **365**, 1333–1346 (2005).
3. Kahn, S., Cooper, M. & Del Prato, S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet (London, England)* **383**, 1068–83 (2014).
4. WHO. Global Report on Diabetes. *Isbn* **978**, 88 (2016).
5. International Diabetes Federation. *IDF Diabetes Atlas*. *IDF* **10**, (2021).
6. WHO. *Classification of diabetes mellitus*. *WHO* **21**, (2019).
7. Nathan, D. M., Turgeon, H. & Regan, S. Relationship between glycated haemoglobin levels and mean glucose levels over time. *Diabetologia* **50**, 2239–2244 (2007).
8. Silventoinen, K. *et al.* Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Research.* **6**, 399–408 (2003).
9. Macgregor, S., Cornes, B. K., Martin, N. G. & Visscher, P. M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Human Genetics* **120**, 571–580 (2006).
10. Elks, C. E. *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Frontiers in Endocrinology* **3**, 1–16 (2012).
11. Min, J., Chiu, D. T. & Wang, Y. Variation in the heritability of body mass index based on diverse twin studies: A systematic review. *Obesity Reviews* **14**, 871–882 (2013).
12. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics* **9**, 1003348 (2013).
13. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *British Medical Journal* **332**, 1080 (2006).
14. Akpinar, E., Bashan, I., Bozdemir, N. & Saatci, E. Which is the best anthropometric technique to identify obesity: Body mass index, waist circumference or waist-hip ratio? *Collegium Antropologicum* **31**, 387–393 (2007).
15. Perl, S. Y. *et al.* Significant human β-cell turnover is limited to the first three decades of life as determined by in vivo thymidine analog incorporation and radiocarbon dating. *Journal of Clinical Endocrinology and Metabolism* **95**, 95 (2010).
16. Capparelli, R. & Iannelli, D. Role of epigenetics in type 2 diabetes and obesity. *Biomedicines* **9**, (2021).
17. Almgren, P. *et al.* Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811–2819 (2011).
18. Meigs, J. B., Cupples, L. A. & Wilson, P. W. F. *Parental Transmission of Type 2 Diabetes The Framingham Offspring Study*. Diabetes **49**, (2000).
19. Bramante, C. T., Lee, C. J. & Gudzune, K. A. Treatment of Obesity in Patients With Diabetes. *Diabetes Spectrum* **30**, 237–243 (2017).
20. Cole, J. B. & Florez, J. C. Genetics of diabetes mellitus and diabetes complications. *Nature Reviews Nephrology* **16**, 377–390 (2020).
21. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics* **50**, 1505–1513 (2018).
22. Sherry, S. T., Ward, M. & Sirotkin, K. dbSNP – database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Research* **9**, 677–679 (1999).
23. 1000G. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

24. Wray, N. R. *et al.* From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry* **78**, 101–109 (2021).
25. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research* **27**, (2018).
26. de Bakker, P. I. W. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human Molecular Genetics* **17**, R122 (2008).
27. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
28. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature Genetics* **52**, 680–691 (2020).
29. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics* **101**, 5–22 (2017).
30. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk of complex disease. *Current Opinion in Genetics and Development* **18**, 257–263 (2008).
31. Wray, N. R. *et al.* Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **55**, 1068–1087 (2014).
32. Choi, S. W., Mak, T. S. H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
33. Khera, A. V *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219–1224 (2018).
34. Lecarpentier, J. *et al.* Prediction of breast and prostate cancer risks in male BRCA1 and BRCA2 mutation carriers using polygenic risk scores. *Journal of Clinical Oncology* **35**, 2240–2250 (2017).
35. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics* **50**, 928–936 (2018).
36. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics* **51**, 584–591 (2019).
37. Reisberg, S., Iljasenko, T., Läll, K., Fischer, K. & Vilo, J. Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One* **12**, (2017).
38. Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).
39. Cavalli-Sforza, L. L., Cavalli-Sforza, L. & Piazza, A. *The History and Geography of Human Genes*. (1994).
40. Lawson, D. J. *et al.* Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* **139**, 23–41 (2020).
41. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–909 (2006).
42. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
43. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* **47**, 284–290 (2015).
44. Bulik-Sullivan, B. K. & Neale, B. M. LD Score Regression Distinguishes Confounding from Polygenicity in GWAS. *Nature Genetics* **47**, 291–295 (2015).

45. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK biobank. *Elife* **8**, 1–47 (2019).
46. Sohail, M. *et al.* Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* **8**, 1–17 (2019).
47. Zaidi, A. A. & Mathieson, I. Demographic history mediates the effect of stratification on polygenic scores. *Elife* **9**, 1–30 (2020).
48. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human Genetic Studies. *Cell* **177**, 26–31 (2019).
49. Peter, B. M., Petkova, D. & Novembre, J. Genetic landscapes reveal how human genetic diversity aligns with geography. *Molecular Biology and Evolution* **37**, 943–951 (2020).
50. Abdellaoui, A. *et al.* Genetic correlates of social stratification in Great Britain. *Nature Human Behaviour* **3**, 1332–1342 (2019).
51. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications* **10**, (2019).
52. Kerminen, S. *et al.* Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. *American Journal of Human Genetics* **104**, 1169–1181 (2019).
53. Pankratov, V. *et al.* Differences in local population history at the finest level: the case of the Estonian population. *Eur. J. Hum. Genet.* **28**, 1580–1591 (2020).
54. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
55. Korunes, K. L. & Goldberg, A. Human genetic admixture. *PLoS Genetics* **17**, (2021).
56. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American Journal of Human Genetics* **100**, 635–649 (2017).
57. Skotte, L. *et al.* Ancestry-specific association mapping in admixed populations. *Genetic Epidemiology* **43**, 506–521 (2019).
58. Scutari, M., Mackay, I. & Balding, D. Using Genetic Distance to Infer the Accuracy of Genomic Prediction. *PLoS Genetics* **38**, 879–887 (2016).
59. Mazandu, G. K., Geza, E., Seuneu, M. & Chimusa, E. R. Orienting Future Trends in Local Ancestry Deconvolution Models to Optimally Decipher Admixed Individual Genome Variations. *IntechOpen* 225–240 (2016).
60. Schubert, R., Andaleon, A. & Wheeler, H. E. Comparing local ancestry inference models in populations of two- and three-way admixture. *PeerJ* **8**, (2020).
61. Shah, S. *et al.* Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *American Journal of Human Genetics* **97**, 75–85 (2015).
62. Walaszczyk, E. *et al.* DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA 1c levels: a systematic review and replication in a case–control sample of the Lifelines study. *Diabetologia* **61**, 354–368 (2018).
63. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81–86 (2017).
64. Gillberg, L. & Ling, C. The potential use of DNA methylation biomarkers to identify risk and progression of type 2 diabetes. *Frontiers in Endocrinology* **6**, 43 (2015).
65. Dayeh, T. *et al.* Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion. *PLoS Genetics* **10**, e1004160 (2014).
66. Kulkarni, H. *et al.* Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Human Molecular Genetics* **24**, 5330–5344 (2015).

67. Chambers, J., Loh, M. & Lehne, B. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes : a nested case-control study. *Lancet Diabetes Endocrinology* **3**, 526–534 (2016).
68. Xu, X. *et al.* A genome-wide methylation study on obesity. *Epigenetics* **8**, 522–533 (2013).
69. Läll, K., Mägi, R., Morris, A., Metspalu, A. & Fischer, K. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. *Genetics in Medicine* **19**, 322–329 (2017).
70. Läll, K. *et al.* Polygenic prediction of breast cancer: Comparison of genetic predictors and implications for risk stratification. *BMC Cancer* **19**, 1–9 (2019).
71. Inouye, M. *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *Journal of the American College of Cardiology* **72**, 1883–1893 (2018).
72. Kuchenbaecker, K. B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *JNCI Journal of the National Cancer Institute* **109**, (2017).
73. Guan, Y. Detecting structure of haplotypes and local ancestry. *Genetics* **196**, 625–642 (2014).
74. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology* **44**, 1172–1180 (2015).
75. Sijtsma, A. *et al.* Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. *International Journal of Epidemiology* 1–8 (2021).
76. Leitsalu, L. *et al.* Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *International Journal of Epidemiology* **44**, 1137–1147 (2015).
77. Leitsalu, L., Alavere, H., Tammesoo, M. L., Leego, E. & Metspalu, A. Linking a population biobank with national health registries — The Estonian experience. *Journal of Personalized Medicine* **5**, 96–106 (2015).
78. UT. Estonian Biobank. (2021). Available at: https://genomics.ut.ee/en/content/estonian-biobank. (Accessed: 1st April 2022)
79. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
80. Conroy, M. *et al.* The advantages of UK Biobank's open-access strategy for health research. *Journal of Internal Medicine* **286**, 389–397 (2019).
81. Xu, H. & Guan, Y. Detecting local haplotype sharing and haplotype association. *Genetics* **197**, 823–838 (2014).

GENERAL DISCUSSION

Summary of the main findings

Common complex diseases, with T2D being a prime example, have the highest health-burden worldwide since there is still a lack of knowledge of all the risk factors involved and how to best apply these for disease prediction and prevention. However, it is known that there is great variability between individuals in the extent to which the complex disease is explained by genetics and by non-genetic (lifestyle and environment) risk factors. Therefore, personalized prediction based on genetic and non-genetic determinants, is seen as a way to tailor prevention and treatment to the high risk groups for T2D. Nonetheless, these personalized approaches are not yet widely practiced and one of the reasons for that is the existence of several methodological caveats such as being able to explain only a small part of the estimated heritability and PRS transferability issues due to population structure. Therefore, my thesis focused mainly on improving the personalized prediction by refining the PRS calculation, by addressing the PRS transferability issue and by adding an epigenetic component to the prediction of T2D. Finally, we reviewed the latest advancements in the genomics field, which could pave the way towards personalized medicine.

Chapter 1 internally and externally validated a novel method of polygenic risk score (PRS) calculation called the '*doubly-weighted genetic risk score*' (dwGRS) in the Estonian Biobank (EstBB) and the Lifelines cohort, respectively. This method applies additional weighing of the included single nucleotide polymorphisms (SNPs) based on their probability to belong to the top associated variants with the aim to correct for the '*Winner's curse*'. In both biobanks the dwGRS for prevalent T2D demonstrated stronger association with incident T2D than the traditional GRS, the latter consisting of previously identified genome-wide significant SNPs only[1]. In addition, when measuring the 5-year predicted risks based on the model with and without the dwGRS, the model including the dwGRS demonstrated its ability to predict the incident T2D cases better than the model without, which was a clear indication for the clinical relevance of the dwGRS.

Chapter 2 introduced novel methods to overcome the current polygenic score (PS) applicability issues for recently admixed (admixture event less than 100 generations ago) individuals from the UK Biobank (UKBB) and compared these with the traditional PSs for height and BMI. First, the results confirmed that traditional PSs using European-based GWAS effect sizes have much lower predictive value among recently admixed individuals than among Europeans. Second, when local ancestry deconvolution was performed on the UKBB admixed individuals and the specific ancestry genomic segment was matched with the corresponding ancestry GWAS summary statistics to calculate ancestry-specific partial PS (aspPS), unbiased PS distributions were achieved. Third, when for the UKBB admixed individuals with partial European background the aspPSs were combined (called *combined ancestry-specific PS*) for the parts of the genome for which the

corresponding ancestry GWAS summary statistics were available, the trait predictability improved and in most of the cases outperformed the total traditional PSs based on UKBB or Japanese Biobank summary statistics.

Chapter 3 aimed to minimize the PRS transferability issue through a Principal Component (PC) projection approach in two European cohorts: UKBB and the EstBB for model traits of height and BMI. The hypothesis was that using a reference population to define the PCs used in correcting for population stratification in GWAS, would minimize the population dependency of GWAS effect sizes. Chapter 3 showed that such a projection approach for PCs did not improve the transferability of PRS calculation from UKBB to EstBB. Out of four projection sets (European, Non-European, the full 1000 Genomes Project cohort, and a subsample from the same large dataset as used in the GWAS), the latter one still performed the best together with dataset-specific PC adjustment in the PRS prediction model. However, some population structure still remained in the PRS even in the best conditions, warranting the cautionary inclusion of PC covariates when validating a PRS.

Chapter 4 studied the effects of Methylation Scores (MSs), epigenetic risk profiles likely reflecting gene-environment interaction and environmental effects, on prevalent T2D and its underlying endophenotypes of fasting plasma glucose (FPG) and glycated hemoglobin (HbA1c). By using the data from three Dutch sub-cohorts (LL pT2D, LL COPD and LL DEEP), all originating from the large North-Netherlands Lifelines Cohort Study and Biobank, Chapter 4 showed that depending on the outcome trait or disease, MSs for prevalent T2D, incident T2D, and FPG had mostly significant effects on the outcome and that their effects were mainly independent of the effects of GRSs. Finding such a trend towards MS independent effect indicates that MSs mostly reflect environmental risk factors or gene-environment effects. However, future studies with larger datasets are warranted to confirm this pattern.

Chapter 5 reviewed how further genetic discoveries are improving personalized prediction and advance functional insights into the link between genetics and disease. Some examples of important developments are increasing efforts of whole genome sequencing, ever larger datasets and meta-analyses, creation of biobanks, better computational and storage resources, and exploration of the neglected parts of the genome. In addition to these developments, this review showed that highly supportive conditions are necessary to implement and use such advancements in favor of personalized prediction and medicine with the EstBB as a prime example.

Discussion of the main findings

The increasing interest in including genomic information in disease risk prediction and recent advancements in the genomics field have been successfully translated into personalized prediction for common complex diseases, highlighted by a number of studies, in which the substantial clinical potential of PRSs was demonstrated[1–7]. However, the PRSs still only explain a fraction of total heritability estimates based on twin and family studies[8,9] and PRSs are still limited by their low transferability[10–18]. Therefore, my thesis focused on potential solutions to overcome the PRS limitations and to improve the amount of explained variance by (epi)genetic risk factors for T2D.

Improvement in polygenic risk score performance

Improving PRSs is highly relevant since complex diseases such as T2D impose a high burden on the medical system[19]. The ultimate goal of personalized prediction involving PRSs would be avoiding or at least postponing the onset of T2D. The dwGRS applied in **Chapter 1** showed a slight improvement in incident T2D prediction compared to the traditional GRS. Although, the added explained variance by the dwGRS was small, especially when compared to parts explained by the established phenotypic risk factors such as BMI and age for T2D, Wray et al. (2021) have emphasized that the real clinical potential of PRSs should be evaluated by their ability to differentiate between disease risk categories[8]. Similarly in **Chapter 1**, we compared dwGRS quintiles while adjusting for the environmental and clinical risk factors and demonstrated its ability to detect 2.8 and 2.3 times higher risk of incident T2D for individuals in the highest quintile compared to the lowest in EstBB and Lifelines, respectively. Similarly, other PRS using new methodological approaches have also shown their ability to differentiate high-risk from lower-risk individuals for complex diseases. For example, Vujkovic et al. showed that individuals in the upper 10^{th} decile of traditional PRS have a 5.21 higher risk of incident T2D compared to the ones in the lowest[20]. Läll and colleagues showed that women in the top 5% of the metaGRS (a method using weighted averages of previously selected top two predicting GRSs, which both use GWAS weights from different sources) had 4.2 times higher risk for breast cancer compared to the lowest 50% in EstBB[5]. Furthermore, after the successful performance of the dwGRS in the original study in the EstBB, where the dwGRS was developed[1], EstBB decided to apply dwGRS in providing personalized feedback for the participants (shown in Chapter 5, Figure 4).

As shown by the examples above, incorporating PRS in personalized prediction is quite promising. Nevertheless the genomics field is still in search of better methods for PRS construction to optimize explained variance for complex diseases[21–27]. Many new methods for PRS computation have been developed in recent years each with their pros and cons. The most promising PRS methods (determined by the percentage of times this PRS ended up among the top two methods with the best prediction performance in the review article by Ma and

Zhou[28]) were PRS-CS, BSLMM, AnnoPred, BayesR, SbayesR, lassosum, multi-BLUP, LDpred, and MTGBLUP. All the listed PRS methods include some advanced methodological steps such as more flexible modeling assumptions, accounting for the strength of linkage disequilibrium (LD) between all SNPs instead of selecting an LD pruned SNPs set, incorporating functional SNP annotations or incorporating computationally more efficient algorithms compared to traditional PRS. All these PRS methods demonstrate clinical potential for detecting high-risk individuals, which could lead to more frequent screening of high-risk individuals and to more cost-efficient prevention programs[29–31].

Nevertheless, developing and validating new PRS methods is just one piece of the puzzle towards reaching the goal of personalized medicine in my opinion, because the explained variance for none of these methods gets anywhere near the total heritability estimates based on twin and family studies. One explanation why PRS do not reach these estimates is over-estimation of the heritability from the twin and family studies due to the high chance that estimates include the shared environmental component[9]. Other pieces of the puzzle, which have been also described in **Chapter 5**, could be increase in GWAS sample size (and thus power) to reach more accurate PRS[32], which may be easier for continuous than for dichotomous outcomes such as T2D[33]; inclusion of rare variants, which are believed to constitute an important part of the genetic component of complex diseases[34,35]; inclusion of structural DNA variants[36] and increased attempts to also model interactions within the genetic loci (dominance effects) and between presumably independent loci (epistasis). The first steps towards inclusion of rare variants are taken by increasing efforts of whole genome sequencing (WGS) with some great examples of large population level WGS initiatives highlighted in **Chapter 5**. One of these, the UK Biobank initiative, just reached the 200,000 samples WGS milestone in November, 2021[37]. Additionally, increasing the GWASs sample size would lead to higher accuracy of PRS[38,39], as shown for example by Hirschhorn and colleagues, who found that the PRS for height based on a GWAS including approximately 5 million individuals finally reached the estimate for common SNP-based heritability[40]. Such findings have only become feasible due to the creation of large biobanks and building large international consortia. All in all, based on **Chapter 1** and other previous literature introducing and validating new PRS methods, there has been an improvement in the performance of the PRS in disease risk prediction. I believe that it is just a matter of time before improved genomic resolution combined with improved PRS methods allows us to target high-risk individuals for complex disease such as T2D in clinical settings. Nevertheless, despite these promising perspectives for PRS, there still remains the question of the validity of the PRS (‘*Are we measuring what we think we are?’)* and whether it is valid for all individuals.

PRS transferability: problems and possible solutions

In recent years the number of studies aiming to tackle the problem of PRS transferability between different populations has increased[10,15,41–43] since currently personalized prediction is not equally applicable for everyone. The PRS constructed from GWAS summary statistics based on European cohorts has much lower performance in other populations, most probably due to differences in allele frequencies, rare variants and linkage disequilibrium patterns between populations[10,18,44,45]. For example, Martin and colleagues showed that a PRS calculated on European summary statistics across 17 quantitative traits provided prediction accuracies that were on average 4.9 times lower in Africans, 2.5 times lower in East Asians and approximately 1.7 times lower in South-Asians and Latino Americans when compared to Europeans[11]. However, the problem does not occur only on the level of global populations. Each human genome has its unique tiling with parts originating from specific ancestries, especially in modern societies, where different cultures come together and there are more genetically mixed individuals. For example, it is estimated that more than one-third of the US population stems from more than one ancestral population[46]. Until recently the common approach was to systematically remove admixed individuals from large-scale genetic studies to avoid possible bias resulting from insufficient correction for population structure[47]. Therefore, in **Chapter 2** via applying local ancestry deconvolution the new PRS methods were developed to extend personalized medicine on admixed individuals resulting in more accurate PRS prediction for them. I believe that aspPS and casPS are currently among the most advanced methods to calculate as accurate PRSs as possible for admixed individuals when matching their genetic ancestry proportions with the corresponding ancestry GWAS. However, these advanced methods are applicable only for the individuals with part of their genome originating from Europeans or any other population, for which there are powerful enough (meta)-GWASs available. For non-European GWASs, available sample sizes are typically smaller, which implies reduction in prediction accuracy[39]. Therefore, these new methods of aspPS and casPS could become even more useful in the personalized prediction for admixed individuals when genomic data resolution improves through increased sample size and through inclusion of more diverse populations. In fact, there are already initiatives, which include more diverse populations such as Pan-UK, which has included six continental ancestry groups and large-enough admixed groups for whom there are more than 16,000 GWASs for different phenotypes available[48]. Also other initiatives such as the African Genome Variation Project[49], the GenomeAsia 100K Project[50] and Human Heredity and Health in Africa (H3Africa)[51], which are all aiming to include, introduce, and develop precision medicine in non-European populations. However, expanding genetic studies to non-Europeans also requires development of customized genotyping arrays and more diverse WGS reference panels[43,49] to better capture specific risk variants, which could differ between the populations.

Until very recently there was a trend towards using uniform GWAS discovery sets to minimize the confounding by population structure, rather than exploring

the diversity and admixture to receive biologically more relevant universal effect sizes arising from different LD patterns in diverse datasets. Therefore, besides the need for more genetically diverse datasets, there is also a need for better resolution of the genome achieved via detecting the genetic ancestry for genome parts for example via ancestry deconvolution. In addition to the PRS methods developed in **Chapter 2,** there are other approaches developed to include admixed individuals in genetic studies such as the recently developed software package called '*Tractor*', which enables GWAS in admixed individuals while applying local ancestry-aware regression[47]. Other than that, there have also been attempts to calculate '*polyethnic scores*' by the software package XP-BLUP, which combines transethnic and ancestry-specific information to improve the PRS prediction[52] or the multiethnic PRS by Márquez-Luna and colleagues, which takes advantage of weights based on GWAS among Europeans (accuracy by large sample size) and weights based on GWAS among sub-population from the target population (accuracy by the same LD patterns)[53]. Also fine-mapping, a method to identify real causal variants in genomic regions resulting in disease risk, has been seen as a promising alternative to the GWAS population-specific tagging variant summary statistics to expand PRS also to other populations[54]. For example, PolyPred is a novel cross-population PRS method incorporating fine-mapping to solve the LD differences and it demonstrated significant increase in prediction accuracy for UKBB Africans and in Biobank Japan[55]. Importantly, all these studies concluded that expanding genetic studies on non-European populations should continue to enlarge sample sizes in order to provide enough statistical power. Only via increasing diversity and more accurately accounting for the origin of the genome, is it possible to make PRS prediction globally feasible.

Now that I have addressed the transferability issue of PRS for admixed and non-European individuals, a next question is PRS transferability among European populations, which was investigated in **Chapter 3**. Although it has been confirmed that the PRS transferability problem increases with the genetic distance between populations[56] and it is caused by differences in population genetic structure, several recent studies have highlighted the fine-scale population genetic structure among Europeans (or even inside a single country) causing biased PRS prediction performance[12,17,57–59]. If this population structure is not correctly accounted for, it results in spurious disease associations in GWAS and lower predictive power of the PRS even in another European population[56]. Correction for the population structure can be done by adding PCs as covariates in the statistical analysis[60]. Such PCs can be obtained by PC analysis using only the genetic data of the individuals analyzed or they can be determined by projection onto a PC space created using a reference set[61]. In **Chapter 3** the hypothesis was that by using a reference dataset to receive the PCs, the PRS transferability problems possibly arising from the use of discovery cohort specific effect sizes could be mitigated. It was shown that population-specific PCs still resulted in better performing PRSs in an independent cohort than the PCs calculated by projecting the study samples into the 1000G reference set. Besides, regardless of the PC approach taken, the PRS performance was always lower when applied to

another cohort than when applied in an independent sample from the same cohort. In other words the transferability issue remained. This could be explained by the fact that PCs based on common variants (traditional approach of calculating PCs) do not capture the recent population structure as well as PCs that also include rare variants[62]. Thus, findings from Chapter 3 highlight that the traditional way of conducting GWAS through incorporating PCs does not entirely remove the existing population structure.

The contribution of epigenetics

Although T2D is a highly heritable disease, the recent rapid increase in diabetes prevalence cannot be explained by the genetic component. To a large extent it is explained by environmental, especially lifestyle factors, and by the interactions between genetics and environment[63]. Methylation as the most common and also reversible epigenetic process is believed to be a molecular link between the environment and disease[64]. Adding to or removing a methylation group from the DNA could switch certain genes on or off[65,66]. It has been shown that diet, physical activity, and smoking can influence methylation patterns in the human genome[67–71], which makes it a promising mechanism for disease prevention and treatment monitoring. Therefore, **Chapter 4** investigated the added value of a methylation score (MS) in explaining the variation in T2D and its endophenotypes (FPG and HbA1c). Results of this chapter were mostly confirming the hypothesis that MS could represent environmental effects as MSs explained a small proportion of the inter-individual variation in T2D in addition to the GRS and that the effects of the MSs and GRSs were largely independent. Nevertheless, the contribution of the MSs was not as large as that of the GRS. It could be explained by the small sample size of the EWAS used to weight the CpG sites included in the MS, because the predictive power of the MS seemed to increase with increasing sample size of the discovery EWAS. Therefore, initiatives for large EWASs or even meta-EWASs are urgently needed. Another downside is that the methylation chips only cover a small part of the CpG sites in the entire genome. For example only 1.5% of all CpG sites mostly from CpG-dense genomic regions are covered by the 450K chip, leaving quite a large proportion of the epigenome still to investigate. Although the more recent studies are already using the HumanMethylation850 (EPIC) microarray, which includes 850,000 CpG sites and around half of them are located in CpG-sparse regions, which have been shown to have effect on gene expression as well[72,73]. As a result, there are few other studies confirming the hypothesis that methylation markers represent environmental risk components[74–76], and only one other that also incorporated genetic predictors[64].

Till now, due to the lower costs and better feasibility, most of the EWASs have a cross-sectional study design meaning that methylation levels and outcomes are measured at the same time point[77,78]. Therefore, future longitudinal studies are warranted to investigate the predictive effect of MSs on incident T2D. Furthermore, similar as genetic studies, epigenetic studies could be more diverse and

should be expanded to include other epigenetic processes such as histone modification[79] , including more ancestries[80–82] and to have a better coverage of epigenome[72], in order to increase the amount of variance explained.

Future research

Based on the results of the current thesis, different PRSs and MS show great promise as screening tools, to detect individuals at high (epi)genetic risk for complex diseases. However, both methods could be further improved. One way to do this would be by using better (epi)genomic resolution, also described in Chapter 5 and by further methodological developments (described below).

Improvement of T2D classification

T2D is a very heterogeneous disease with patients presenting a broad range of characteristics. The current definition of T2D may be an *umbrella-term* including many different subtypes of T2D. For example, a study conducted in a large diabetes cohort in southern Sweden[83] demonstrated that it is possible to dissect adult-onset diabetes into five different subtypes (four subtypes for T2D) based on age at diabetes onset, HbA1c, BMI, measures of insulin resistance and secretion, and glutamic acid decarboxylase antibodies (GADA) (see Figure below).

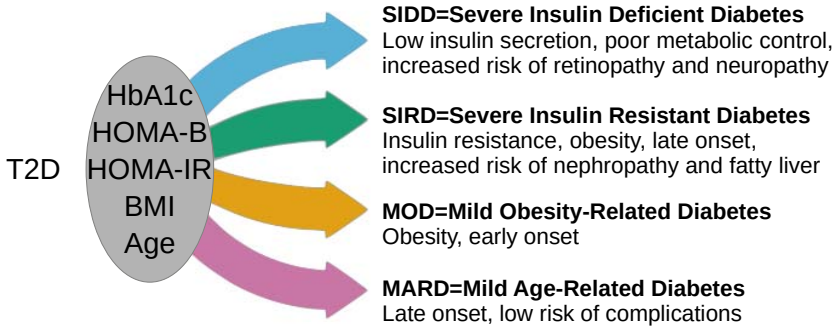

Figure 1. Novel type 2 diabetes subtype characteristics. Adapted from Ahlqvist et al. 2020.

These subtypes showed differences in clinical characteristics, complication severity, drug response, and disease progression and were also replicated among other European cohorts[84–86], and cohorts from India[85], China and the United States[87] showing the generalizability of such a classification. More importantly, these subtypes revealed partial but strong distinctions in their genetic etiology[88]. Therefore, future studies should not only focus on the predictors of T2D in general, but should apply such a refined diabetes definition to improve the PRSs accuracy resulting in a more tailored medical treatment for each individual with a specific subtype of T2D.

Multidisciplinary research

As also demonstrated in the current thesis, bringing together different research fields – genetics, epigenetics and population genetics – could improve our understanding of the genetic and environmental effects on T2D and complex traits in general. However, under the notion that '*the whole is greater than the sum of its parts*' these research fields could be even more intertwined in the future. One great example of this would be evolutionary medicine. Understanding the human past is important to better understand the genetic and environmental risk factors involved in disease progress, as also shown in **Chapter 2**. Another excellent example of merging the research fields is a study by Schradel et al (2022), where they showed that the individuals belonging to different T2D sub-types described in previous paragraph, also differed by the methylation patterns measured in blood[89]. Therefore, as a next step I would suggest to combine the approaches from the separate chapters of this thesis and to build prediction models that include PRSs that consider the genetic ancestry for parts of the genome and a MS, while using improved, more homogeneous subtypes of T2D as outcome.

Towards multi-omics

In addition to the rapid advancements in the genetics field, we should zoom in on other molecular levels to get a more detailed understanding of the complexity of T2D. That could be done via epigenomics, transcriptomics, proteomics, metabolomics and pharmacogenomics revealing new biomarkers and disease mechanisms, which would result in more precise personalized interventions and treatment approaches. However, similar to genetics, also the multi-omics field is in need of more data and data diversity before reliable results can be produced[90–92]. Even if the technology for the multi-omics is available, limitations of motivation, time and costs preclude its application and integration in clinical settings.

CONCLUSIONS

The findings of this thesis aimed to remove methodological hurdles along the way towards accelerating personalized medicine for complex diseases in general while using T2D as a specific example. Here were tested existing and developed new approaches to reveal more of T2D's complex nature and indicating ways towards more personalized prediction and medicine accessible and feasible for everyone. These findings indicate that the scientists should continue unraveling the mechanisms leading to complex diseases with practicing more multi-disciplinary approaches, which could lead to novel methods with improved accuracy to target high-risk individuals. In this way personalized prediction becomes more feasible and inseparable from the medical field.

REFERENCES

1. Läll, K., Mägi, R., Morris, A., Metspalu, A. & Fischer, K. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. *Genet. Med.* **19**, 322–329 (2017).
2. Khera, A. V *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219–1224 (2018).
3. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.* **10**, 1–10 (2019).
4. Inouye, M. *et al.* Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).
5. Läll, K. *et al.* Polygenic prediction of breast cancer: Comparison of genetic predictors and implications for risk stratification. *BMC Cancer* **19**, 1–9 (2019).
6. Kuchenbaecker, K. B. *et al.* Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in BRCA1 and BRCA2 mutation carriers. *J. Natl. Cancer Inst.* **109**, (2017).
7. Mavaddat, N. *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* **104**, 21–34 (2019).
8. Wray, N. R. *et al.* From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry* **78**, 101–109 (2021).
9. Nolte, I. M. *et al.* Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *Eur. J. Hum. Genet.* **25**, 877–885 (2017).
10. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
11. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
12. Kerminen, S. *et al.* Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. *Am. J. Hum. Genet.* **104**, 1169–1181 (2019).
13. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* **10**, (2019).
14. Marnetto, D. *et al.* Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat. Commun.* **11**, 1–9 (2020).
15. Reisberg, S., Iljasenko, T., Läll, K., Fischer, K. & Vilo, J. Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One* **12**, (2017).
16. Bitarello, B. D. & Mathieson, I. Polygenic Scores for Height in Admixed Populations. *G3: Genes|Genomes|Genetics* g3.401658.2020 (2020). doi:10.1534/g3.120.401658
17. Sakaue, S. *et al.* Dimensionality reduction reveals fine-scale structure in the Japanese population with consequences for polygenic risk prediction. *Nat. Commun.* **11**, 1–11 (2020).
18. Wang, Y. *et al.* Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* **11**, (2020).
19. International Diabetes Federation. *IDF Diabetes Atlas*. *IDF* **10**, (2021).

20. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680–691 (2020).
21. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41**, 469–480 (2017).
22. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
23. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, (2019).
24. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, (2019).
25. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLoS Genet* **9**, 1003264 (2013).
26. Hu, Y. *et al.* Leveraging functional annotations in genetic risk prediction for human complex diseases. (2017). doi:10.1371/journal.pcbi.1005589
27. Speed, D. & Balding, D. J. MultiBLUP: Improved SNP-based prediction for complex traits. *Genome Res.* **24**, 1550–1557 (2014).
28. Ma, Y. & Zhou, X. Genetic prediction of complex traits with polygenic scores: a statistical review. *Trends in Genetics* **37**, 995–1011 (2021).
29. Hynninen, Y., Linna, M. & Vilkkumaa, E. Value of genetic testing in the prevention of coronary heart disease events. *PLoS One* **14**, (2019).
30. Pashayan, N., Morris, S., Gilbert, F. J. & Pharoah, P. D. P. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer A Life-Table Model. *JAMA Oncol.* **4**, 1504–1510 (2018).
31. Gibson, G. On the utilization of polygenic risk scores for therapeutic targeting. *PLoS Genet.* **15**, (2019).
32. Spencer, C. C. A., Su, Z., Donnelly, P. & Marchini, J. Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* **5**, 1000477 (2009).
33. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *Br. Med. J.* **332**, 1080 (2006).
34. Panoutsopoulou, K., Tachmazidou, I. & Zeggini, E. In search of low-frequency and rare variants affecting complex traits. *Hum. Mol. Genet.* **22**, (2013).
35. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–89 (2015).
36. Connolly, J. J. *et al.* Copy number variation analysis in the context of electronic medical records and large-scale genomics consortium efforts. *Frontiers in Genetics* **5**, (2014).
37. UKRI. UK Biobank sequences whole genomes of 200,000 participants. UK Research and Innovation. (2021). Available at: https://www.ukri.org/news/uk-biobank-sequences-whole-genomes-of-200000-participants/. (Accessed: 1st April 2022)
38. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genet.* **9**, 1003348 (2013).
39. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. (2018). doi:10.1002/mpr.1608
40. Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J. & Bartell, E. A saturated map of common genetic variants associated with human height from 5 . 4 million individuals of diverse ancestries. 1–38

41. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
42. De La Vega, F. M. & Bustamante, C. D. Polygenic risk scores: A biased prediction? *Genome Med.* **10**, 95–97 (2018).
43. Kim, M. S., Patel, K. P., Teng, A. K., Berens, A. J. & Lachance, J. Genetic disease risks can be misestimated across global populations. *Genome Biol.* **19**, 1–14 (2018).
44. Privé, F. *et al.* High-resolution portability of 240 polygenic scores when derived and applied in the same cohort. 1–18 (2021).
45. Bien, S. A. *et al.* The Future of Genomic Studies Must Be Globally Representative: Perspectives from PAGE. *Annual Review of Genomics and Human Genetics* **20**, 181–200 (2019).
46. Parker, K., Menasche Horowitz, J., Morin, R. & Lopez, M. H. Multiracial in America: Proud, Diverse and Growing in Numbers | Pew Research Center. (2015). Available at: https://www.pewresearch.org/social-trends/2015/06/11/multiracial-in-america/. (Accessed: 21st May 2022)
47. Atkinson, E. G. *et al.* Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).
48. Pan UKBB Team. Pan UKBB | Pan UKBB. Available at: https://pan.ukbb.broadinstitute.org/. (Accessed: 13th June 2022)
49. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).
50. Wall, J. D. *et al.* The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106–111 (2019).
51. Mulder, N. *et al.* H3Africa: Current perspectives. *Pharmacogenomics and Personalized Medicine* **11**, 59–66 (2018).
52. Coram, M. A., Fang, H., Candille, S. I., Assimes, T. L. & Tang, H. Leveraging Multi-ethnic Evidence for Risk Assessment of Quantitative Traits in Minority Populations. *Am. J. Hum. Genet.* **101**, 218–226 (2017).
53. Márquez-Luna, C. *et al.* Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
54. Spain, S. L. & Barrett, J. C. Strategies for fine-mapping complex traits. *Human Molecular Genetics* **24**, R111–R119 (2015).
55. Weissbrod, O. *et al.* Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* | **54**, 450–458 (2022).
56. Privé, F. *et al.* Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort (The American Journal of Human Genetics (2022) 109(1) (12–23), (S0002929721004201), (10.1016/j.ajhg.2021.11.008)). *Am. J. Hum. Genet.* **109**, 373 (2022).
57. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* **10**, 1–9 (2019).
58. Sohail, M. *et al.* Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* **8**, 1–17 (2019).
59. Pankratov, V. *et al.* Differences in local population history at the finest level: the case of the Estonian population. *Eur. J. Hum. Genet.* **28**, 1580–1591 (2020).
60. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **27**, 1–10 (2018).
61. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

62. Zaidi, A. A. & Mathieson, I. Demographic history mediates the effect of stratification on polygenic scores. *Elife* **9**, 1–30 (2020).
63. Ali, O. Genetics of type 2 diabetes. *Curr. Sci.* **4**, 114–123 (2013).
64. Shah, S. *et al.* Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am. J. Hum. Genet.* **97**, 75–85 (2015).
65. Ling, C. & Rönn, T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metab.* **29**, 1028–1044 (2019).
66. Wolffe, A. P. & Guschin, D. Review: Chromatin structural features and targets that regulate transcription. *J. Struct. Biol.* **129**, 102–122 (2000).
67. Santos, J. M., Tewari, S. & Benite-Ribeiro, S. A. The effect of exercise on epigenetic modifications of PGC1: The impact on type 2 diabetes. *Med. Hypotheses* **82**, 748–753 (2014).
68. Langdon, R. J., Yousefi, P., Relton, C. L. & Suderman, M. J. Epigenetic modelling of former, current and never smokers. *Clin. Epigenetics* **13**, 1–13 (2021).
69. Rönn, T. & Ling, C. Effect of exercise on DNA methylation and metabolism in human adipose tissue and skeletal muscle. *Epigenomics* **5**, 603–605 (2013).
70. Gillberg, L. *et al.* Adipose tissue transcriptomics and epigenomics in low birthweight men and controls: role of high-fat overfeeding. *Diabetologia* **59**, 799–812 (2016).
71. Xu, X. *et al.* A genome-wide methylation study on obesity: Differential variability and differential methylation. *Epigenetics* **8**, 522–533 (2013).
72. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**, 389–399 (2016).
73. Gutierrez-Arcelus, M. *et al.* Tissue-Specific Effects of Genetic and Epigenetic Variation on Gene Regulation and Splicing. *PLoS Genet.* **11**, 1004958 (2015).
74. Richmond, R. C., Suderman, M., Langdon, R., Relton, C. L. & Smith, G. D. DNA methylation as a marker for prenatal smoke exposure in adults. *Int. J. Epidemiol.* **47**, 1120–1130 (2018).
75. Amador, C. *et al.* Genome-wide methylation data improves dissection of the effect of smoking on body mass index. *PLoS Genet.* **17**, (2021).
76. Samblas, M., Milagro, F. I. & Martínez, A. DNA methylation markers in obesity, metabolic syndrome, and weight loss. *Epigenetics* **14**, 421–444 (2019).
77. Walaszczyk, E. *et al.* DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA 1c levels: a systematic review and replication in a case–control sample of the Lifelines study. *Diabetologia* **61**, 354–368 (2018).
78. Kulkarni, H. *et al.* Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum. Mol. Genet.* **24**, 5330–5344 (2015).
79. Nakato, R. & Sakata, T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* **187**, 44–53 (2021).
80. Hüls, A. & Czamara, D. Methodological challenges in constructing DNA methylation risk scores. *Epigenetics* **15**, (2020).
81. Heyn, H. *et al.* DNA methylation contributes to natural human variation. doi:10.1101/gr.154187.112
82. Elliott, H. R. *et al.* Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin. Epigenetics* **6**, 1–10 (2014).
83. Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).

84. Slieker, R. C. *et al.* Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* **64**, 1982–1989 (2021).
85. Ahlqvist, E., Prasad, R. B. & Groop, L. Subtypes of type 2 diabetes determined from clinical parameters. *Diabetes* **69**, 2086–2093 (2020).
86. Ahlqvist, E., van Zuydam, N. R., Groop, L. C. & McCarthy, M. I. The genetics of diabetic complications. *Nat. Rev. Nephrol.* **11**, 277–287 (2015).
87. Zou, X., Zhou, X., Zhu, Z. & Ji, L. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *The Lancet Diabetes and Endocrinology* **7**, 9–11 (2019).
88. Mansour Aly, D. *et al.* Genome-wide association analyses highlight etiological differences underlying newly defined subtypes of diabetes. *Nat. Genet.* **53**, 1534–1542 (2021).
89. Schrader, S. *et al.* Novel Subgroups of Type 2 Diabetes Display Different Epigenetic Patterns, Which Associate With Future Diabetic Complications. *Diabetes Care* 1–10 (2022). doi:10.2337/dc21-2489
90. Diamanti, K. *Integrating multi-omics for type 2 diabetes: Data science and big data towards personalized medicine*. (2019). doi:10.13140/RG.2.2.23730.84162
91. Ahmed, Z. Multi-omics strategies for personalized and predictive medicine: past, current, and future translational opportunities. (2022). doi:10.1042/ETLS20210244
92. Reel, P. S., Reel, S., Pearson, E., Trucco, E. & Jefferson, E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* **49**, 107739 (2021).

EESTIKEELNE KOKKUVÕTE (Summary in Estonian)

Komplekstunnuste ja -haiguste personaalse ennetamise edendamine teist tüüpi diabeedi näitel

Levinud komplekshaigused, näiteks teist tüüpi diabeet (*type 2 diabetes*, T2D), on ühed juhtivad haigestumuse ja suremuse põhjused kogu maailmas, kuna endiselt puuduvad teadmised kõigi nendega seotud riskitegurite kohta ja selle kohta, kuidas olemasolevaid riskitegureid kõige paremini rakendada haiguste prognoosimiseks ja ennetamiseks. Siiski on teada, et komplekshaigustega seotud geneetilised ja mitte-geneetilised (sh elustiil ja keskkond) riskitegurid varieeruvad indiviidide vahel suurel määral. Seetõttu on geneetilistel ja mitte-geneetilistel andmetel põhinevate algoritmide väljatöötamine oluline, et suunata ennetus ja ravi eelkõige nendele indiviididele, kes kuuluvad kõrgesse T2D riskirühma. Sellele vaatamata ei kasutata personaliseeritud lähenemisviise veel laialdaselt, kuna esinevad erinevad metoodilised piirangud. Näiteks on praegused geneetilised meetodid võimelised seletama ainult väikese osa hinnangulisest päritavusest ja polügeensed riskiskoorid (PRS-d), mis on huvipakkuva tunnusega seotud geneetiliste variantide kaalutud alleelide summad, ei ole populatsioonistruktuuri tõttu otseselt ülekantavad ühest populatsioonist teise.

Käesolev doktoritöö keskendus peamiselt T2D geneetilistele ja epigeneetilistele riskiteguritele eesmärgiga parendada haiguse personaliseeritud ennetusvõimet, et nende meetodite laialdasema kasutusega inimeste tervist edendada. Lisaks anti töös ülevaade genoomika valdkonna uutest arengutest ja nendega seotud tulevikuvisioonist personaalmeditsiini rakendamisel. Järgnevad viis peatükki põhinevad väitekirja teadusartiklitel.

1. peatükis valideeriti uudne PRS-i arvutamise meetod, mida nimetatakse "topeltkaalutud geneetiliseks riskiskooriks" (doubly-weighted genetic risk score, dwGRS) Eesti Geenivaramu ja Lifelines'i biopankade andmestikes. dwGRS puhul kaaluti kaasatud ühenukleotiidilised polümorfismid (*single nucleotide polymorphisms,* SNPs) vastavalt empiiriliselt hinnatud tõenäosusele kuuluda SNP-de hulka, millel on uuritava tunnusega tegelik seos. dwGRS eeliseks on see, et korrigeeritakse juhuslikku SNP-de seose ülehindamist uuritava tunnusega. Mõlemas biopanga andmestikus näitas dwGRS tugevamat seost T2D haigestumusega kui traditsiooniline GRS, mis koosneb ainult eelnevalt tuvastatud kogu genoomi hõlmavatest olulistest SNP-dest. Lisaks näitas viie aasta haigestumustõenäosuse hindamisel dwGRS-i sisaldav mudel paremat T2D haigestumuse ennustusvõimet, mis näitab selgelt dwGRS-i sobivust kliiniliseks rakenduseks.

2. peatükis analüüsiti uusi meetodeid polügeensete skooride (*polygenic scores*, PS) kasutamiseks hiljuti segunenud (segunemissündmus vähem kui 100 põlvkonda tagasi) indiviididel, kelle esivanemad pärinevad erinevatest populatsioonidest, Ühendkuningriigi biopangas (*UK Biobank,* UKBB) ja võrreldi neid traditsiooniliste PS-dega pikkuse ja kehamassiindeksi jaoks. Esiteks näitasid tulemused, et traditsioonilistel PS-del, mis kasutavad skooris kaasatud SNP-de jaoks kaalusid Euroopa-põhistest genoomiülestest assotsiatsiooniuuringutest (*Genome-*

Wide Association Study, GWAS), on hiljuti segunenud isikute puhul palju madalama ennustusvõimega kui eurooplaste puhul. Teiseks, UKBB segunenud indiviidide genoomiandmetel kasutati kohaliku põlvnemise lahtiharutamise (*Local Ancestry Deconvolution*, LAD) meetodit. Selle tulemusena sai kokku sobitada kindla põlvnemisega genoomse segmendi ja vastava päritoluga GWAS-i kaalud põlvnemis-spetsiifilise osalise PS (*ancestry-specific partial polygenic score,* aspPS) arvutamiseks, siis saavutas see tõepärase PS-i jaotuse. Kolmandaks, UKBB-st pärit segunenud põlvnemisega indiviididel, kel oli osaliselt Euroopa taust, sai arvutada ning omavahel kombineerida mitu aspPS-i – kombineeritud põlvnemis-spetsiifiline PS (*combined ancestry specific polygenic score,* casPS). CasPS-i sai arvutada just nende genoomi osade põhjal, mille kohta olid olemas vastava päritoluga GWAS-i kaalud. Selline uudne PS saavutas parema komplekstunnuste ennustusvõime, mis enamustel juhtudel ületas traditsiooniliste PS-de tulemuse, mis põhinesid kas ainult UKBB või Jaapani biopanga kaaludel.

3. peatüki eesmärk oli vähendada PRS-i ülekantavuse probleemi kahe Euroopa kohordi (*Estonian Biobank* – EstBB ja UKBB) vahel, kus uuritavateks tunnusteks olid pikkus ja kehamassiindeks. Selleks kasutati peakomponentide (*Principal Components*, PC) projektsioonil põhinevat lähenemisviisi. Kui tavapäraselt kohandatakse GWAS peakomponentidele, mis on arvutatud sama uuringukohordi andmete põhjal, et limiteerida populatsiooni geneetilise struktuuri mõju, siis selles uuringus oli GWAS korrigeeritud peakomponentidele, mis olid arvutatud referentspopulatsiooni põhjal. Eelduseks oli, et selline lähenemine vähendab GWAS-ist pärinevate kaalude sõltuvust uuringupopulatsiooni geneetilisest struktuurist ning vähendab PRS-i ülekantavuse probleemi teise populatsiooni, kus on omakorda erinevused geneetilises struktuuris. Tulemused näitasid, et selline PC-de projektsioonimeetod ei parandanud PRS-i ülekantavust UKBB-st EstBB-sse. Neljast projektsioonikogumist (eurooplased, mitte-eurooplased, kogu 1000 Genoomi Projekti kohort ja alamvalim samast suurest andmekogumist, mida kasutati GWASs) oli viimane siiski kõige parem koos andmekogumi-spetsiifilise PC kohandamisega PRS-i valideerimismudelis. Siiski sisaldas PRS isegi parima PC korrektsiooni korral teatavat populatsioonistruktuuri, mis rõhutab PRS-i arvutamisel ja seejärel valideerimisel kasutatavate populatsiooni struktuuri korrigeerivate meetodite tähtsust ja kriitilist suhtumist vajaliku populatsioonistruktuuri korrigeeriva meetodi osas.

4. peatükis arvutati metülatsiooniskoorid (*Methylation Score*, MS) ja uuriti nende rolli T2D esinemise korral ning nende glükeemiliste endofenotüüpide nagu paastuplasma glükoosi (*Fasting plasma glycose*, FPG) ja glükohemoglobiini (*glycosated hemoglobin*, HbA1c) tasemetes. Metülatsioon on molekulaarne mehhanism, mis leiab aset geenide ning keskkonna koosmõjul ja/või ainult keskkonna mõjul desoksüribonukleiinhappe (DNA) pinnal. Põhja-Hollandi rahvastikupõhise kohordi (Lifelines) kolme alamkohordi (LL pT2D, LL COPD ja LL DEEP) andmete põhjal leiti, et sõltuvalt uuritavast fenotüübilisest tunnusest või haigusest oli MS-pT2D, MS-iT2D ja MS-FPG-l (kolm MS-i, mis olid vastavalt T2D levimust, haigestumust ja FPG kaalusid kasutades arvutatud) enamasti statistiliselt oluline mõju uuritavale tunnusele või haigusele ning nende mõju oli

enamasti sõltumatu geneetilise riskiskoori mõjust. Selline tulemus näitab, et MS võib olla molekulaarne mehhanism, mis peegeldab keskkonna mõjusid haiguse levimuses.

5. peatükis anti ülevaade, kuidas hiljutised arengud geneetiliste andmetega uuringutes võimaldavad geneetiliste meetodite paremat ennustusvõimet ja edendavad teadmisi funktsionaalsetest seostest geneetika ja haiguste vahel. Mõned näited olulistest arengusuundadest on üha suurenevad jõupingutused kogugenoomide sekveneerimisel, suurenevad andmekogumid ja nende meta-analüüsid, biopankade loomine, paremad arvutus- ja salvestusressursid ning tähelepanuta jäetud genoomi osade uurimine. Lisaks nendele arengutele näitas käesolev ülevaateuuring, et selliste teadmiste kasutamiseks personaliseeritud ennetuses ja meditsiinis on vaja soosivaid tingimusi. Nii on Eesti Geenivaramu heaks mudelnäiteks sellest, kuidas arengud ja avastused geneetiliste andmetega, turvaliseks geenidoonorluseks vajaliku seadusandluse sätestamine ning rahvastiku kõrge osalushuvi avavad uusi võimalusi personaliseeritud ennetuse ja personaalmeditsiini arendamiseks.

NEDERLANDSE SAMENVATTING (Summary in Dutch)

Verbetering van de persoonlijke predictie van complexe eigenschappen en ziektes: een toepassing op Type 2 Diabetes

Veel voorkomende complexe ziekten, waarvan Type 2 Diabetes (T2D) een uitstekend voorbeeld is, hebben wereldwijd de hoogste gezondheidslasten, omdat er nog steeds een gebrek is aan kennis van alle betrokken risicofactoren. Daarnaast weet men nog niet hoe deze het best kunnen worden toegepast voor ziektevoorspelling en -preventie. Het is echter bekend dat er tussen individuen grote variabiliteit bestaat in de mate waarin de complexe ziekte wordt verklaard door genetische en door niet-genetische (leefstijl en omgeving) risicofactoren. Daarom wordt gepersonaliseerde voorspelling, gebaseerd op genetische en niet-genetische informatie, gezien als een manier om preventie en behandeling op maat te maken en te richten op hoog risicogroepen voor T2D. Toch worden deze gepersonaliseerde benaderingen nog niet op grote schaal toegepast en één van de redenen daarvoor is het bestaan van verschillende methodologische limitaties, zoals het feit dat huidige genetische risicoprofielen slechts een klein deel van de geschatte erfelijkheidsgraad kunnen verklaren en dat ze niet overdraagbaar zijn naar niet-Europese individuen ten gevolge van populatiestructuren. Daarom richt mijn proefschrift zich vooral op het verbeteren van de gepersonaliseerde voorspelling door het verfijnen van de PRS-berekening, door het aanpakken van het PRS overdraagbaarheidsprobleem, door het toevoegen van een epi-genetische component aan de voorspelling van T2D en door het samenvatten van de laatste ontwikkelingen op het gebied van genomics, zodat uiteindelijk een weg gebaand zou kunnen worden naar gepersonaliseerde geneeskunde.

Hoofdstuk 1 valideerde intern en extern een nieuwe methode voor de berekening van polygene risicoscores, de zogenaamde ‘dubbel gewogen genetische risicoscore’ (dwGRS) in respectievelijk de Estonian Biobank en het Lifelines cohort. Deze methode past een extra weging toe van de opgenomen enkel-nucleotide-polymorfismen (SNP’s) op basis van hun waarschijnlijkheid om tot de sterkst geassocieerde varianten te behoren met als doel te corrigeren voor de “Winner‘s curse”. In beide biobanken toonde de dwGRS voor prevalente T2D een sterkere associatie met incidente T2D dan de traditionele GRS, die alleen bestaat uit eerder geïdentificeerde genoom-breed significante SNP’s. Bovendien, bij het meten van de vijfjaars voorspelde risico’s op basis van het model met en zonder de dwGRS, kon het model met de dwGRS beter incidente T2D voorspellen dan het model zonder, wat een duidelijke indicatie was voor de klinische relevantie van de dwGRS.

Hoofdstuk 2 introduceerde nieuwe methoden om de huidige problemen met de overdraagbaarheid van de polygene risicoscore (PS) voor individuen van recentelijk gemengde (minder dan 100 generaties geleden) afkomst uit de UK Biobank (UKBB) op te lossen en vergeleek deze methoden met de traditionele PS’en voor lengte en BMI. Ten eerste bevestigden de resultaten dat traditionele PS’en, die gebruik maken van effectgrootte schattingen uit Europese genoom-

brede associatie studies (GWAS), een veel lagere voorspellende waarde hebben onder individuen van recentelijk gemengde afkomst dan onder Europeanen. Ten tweede, wanneer lokale voorouderlijke deconvolutie werd toegepast op de UKBB individuen van gemengde afkomst en het specifieke voorouderlijke genomische segment werd afgestemd op de samenvattende statistieken uit overeenkomstige voorouderlijke GWAS om voorouderlijk-specifieke partiële PS (aspPS) te berekenen, werden onvertekende PS distributies gevonden. Ten derde, wanneer voor de UKBB individuen van een gemengde, maar gedeeltelijke Europese afkomst de aspPSs werden gecombineerd (gecombineerde voorouderlijk-specifieke PS genoemd) met de delen van het genoom waarvoor samenvattende statistieken uit GWASs gebaseerd op corresponderende afkomst beschikbaar waren, dan verbeterde dat de voorspelbaarheid van de uitkomstmaten en presteerde het in sommige gevallen beter dan de totale traditionele PS'en gebaseerd op ofwel de samenvattende statistieken uit de UKBB of Japanse Biobank GWAS.

Hoofdstuk 3 richtte zich op het minimaliseren van het PRS overdraagbaarheidsprobleem door middel van een Principal Component (PC) projectie benadering in twee Europese cohorten, UKBB en de Estse Biobank (EstBB), voor de modeleigenschappen lichaamslengte en BMI. De hypothese was dat het gebruik van een referentiepopulatie om de PCs te definiëren die gebruikt worden bij het corrigeren voor populatiestratificatie in GWAS, de populatieafhankelijkheid van GWAS effectgroottes zou minimaliseren. Hoofdstuk 3 toonde aan dat een dergelijke projectiebenadering voor PCs de overdraagbaarheid van de PRS van UKBB naar EstBB niet verbeterde. Van de vier projectiesets (het Europese, het niet-Europese deel en het volledige 1000-genoom project cohort, en een deelsteekproef uit dezelfde grote dataset als gebruikt in de GWAS), presteerde de laatste nog steeds het beste, samen met dataset-specifieke PC-correctie in het PRS-voorspellingsmodel. Er bleef echter nog steeds enige populatiestructuur in de PRS aanwezig, zelfs onder de beste omstandigheden, wat de opname van PC-covariaten bij de validatie van een PRS rechtvaardigt.

Hoofdstuk 4 introduceerde de Methylation Score (MS) die mogelijk gen-omgeving interactie en omgevingseffecten weergeeft die van invloed zijn op prevalente T2D en de onderliggende endofenotypes van nuchtere plasma glucose (FPG) en geglyceerd hemoglobine (HbA1c). Door gebruik te maken van de gegevens van drie Nederlandse subcohorten (LL pT2D, LL COPD en LL DEEP), allen afkomstig uit de grote Noord-Nederlandse Lifelines Cohort Study en Biobank, toonde hoofdstuk 4 aan dat afhankelijk van het kenmerk of de ziekte, de MS voor prevalente T2D, incidentele T2D, en FPG meestal significante effecten hadden op de uitkomst en dat hun effecten meestal onafhankelijk waren van de effecten van de GRS. De bevindingen van deze trend aangaande een MS-onafhankelijk effect geeft aan dat de MS gezien kan worden als een mogelijk moleculair mechanisme dat de omgevingsrisicofactoren of gen-omgeving interactie-effecten weerspiegelt, maar toekomstige studies met grotere datasets moeten worden gedaan om een dergelijk patroon te bevestigen.

In **hoofdstuk 5** wordt besproken hoe verdere genetische ontdekkingen de voorspelling van persoonlijke aandoeningen kunnen verbeteren en functionele

inzichten in het verband tussen genetica en ziekte bevorderen. Enkele voorbeelden van belangrijke ontwikkelingen zijn de toenemende focus op whole genome sequencing, steeds grotere datasets en meta-analyses, het creëren van biobanken, betere computationele en opslagmiddelen voor data, en de verkenning van de genegeerde delen van het genoom. Naast deze ontwikkelingen is uit dit overzicht gebleken dat er veel ondersteunende voorwaarden nodig zijn om een dergelijke vooruitgang te implementeren en te gebruiken ten behoeve van gepersonaliseerde ziektevoorspelling en genezing, met de Estse Biobank als een uitstekend voorbeeld.

ACKNOWLEDGEMENTS

When starting a PhD, four years seems a very long time. When finalizing my PhD, it felt as how come this time went so fast. During a usual PhD track, a PhD candidate has two families, the one, who is unfortunately shifted to the background for the busy moments of the PhD and the academic one, who is the closest support through the PhD track. However, I have been very lucky having even two of the additional academic families by following a Joint PhD track.

First, I want to thank my family at the University of Tartu, Institute of Genomics, Estonian Biocentre, PhD room ('*pigeon room*'), especially my deskmates :) Of course the best moments to recall out of many are: Firstly, Anne-Mai, who was always there to listen and to share thought especially when some scientific texts needed a 'quick' translation from English to Estonian. Second, the other deskmate in the beginning of my PhD was one quiet Mexican guy Rodrigo, who turned out to be the coolest Mexican-Estonian guy I know, who has managed to be there thoughout my PhD no matter am I in Tartu or Groningen, still catching up. Besides them, there were Tina, Mathilde, Linda, Ludovica, Stefania, Ajai, Hovic, Mayukh, Vasili, Fransesco, Bayazit, Monika, Tuuli, Helja, Freddi, Lena, Marcel, Merjam, Lehti, Anu, Kristiina, Mait, Jüri, Richard, Helen, Mariza, Merilin, Merit, Ingrid, Erwan, Helen, Siiri, Ene, Maere, who all made my work-related return to Estonia fun and welcoming. Especially the first listed ones with endless board game evenings, fun dinners or PhD talks. I will never forget the celebration moments of Estonian Epee Team winning the gold medal on the Olympics 2021 or those endless chats and cakes in the EBC kitchen. Müts maha, Mait, sa oled suutnud kokku panna ühe kõige toredama töötajaskonna, keda siiani olen kohanud. It was just amazing to experience how supportive and creative this team is. Although I was not always present, I always felt as at home!

Second, I am grateful of having my academic family in the Netherlands. First and foremost, I would like to thank my first officemates: Tian and Jing. It was the best time to share our happy and sad moments with each other and to be there providing support! I do miss the good old Triade building times.

After moving to an opened office, my greatest appreciation goes to the second floor cool team without whom the working days would definitely be less bright and fun. Thank you: Nigus, Peter, Bale, Zekai, Wenbo, Elnaz, Tian, Zhen, Pato, Melisa, Kebede, Carel-Peter (actually the first floor), Rujia, Rima, Simon, Sitsi, Qihua, Getaneh, Sisay, Nigussie, Tigist, Eliza, Chris, Anna and the others. Special thanks go to Kebede, thank you, it's always easier to be in the PhD submission process together with another friend.

Also I am grateful for all the Genetic Epidemiology unit members for your valuable feedback and discussion during our meeting and for the Department of Epidemiology for nice working atmosphere.

To my supervisors I like to address my special gratitude. It has definitely not been an easy journey, but we have made it! While some PhD students do an internship or research visit to have different research experience, then for me it

was an unreachable goal due to one long-lasting pandemic. However, combining two PhDs gave me an even better experience seeing how two very different yet both professional and super cool teams work.

Davide, the insight, perfection and skills learned from you are invaluable. It is difficult to express it in words, but I will try. What I regret the most about my PhD? Nothing, except the fact that I could not learn four years in a row the bioinformatics skills from you. Concerning these skills, I considered myself "*tabula rasa*" when I started my PhD. I guess I was not skilled enough since my educational track goes from green biology (*brown bears*) to wet lab (*antibotic resistance in humans*) to epidemiology and genetic epidemiology, however, in the end what I love the most is bioinformatics and working on nice figures, still long way to go, right ;) So let's see where the next career steps take me, but I do wish you success and patience as you had with me, with your next students and colleagues.

Ilja, thank you for the first insights to genetic epidemiology and data handling already during my research master. Without you, I wouldn't have had such a great start! You were always like a cheerful guardian angel keeping the hand on the pulse of my project, always there if needed and always up to date, where and what am I doing!

Luca, I admire the speed of your work and creativity. From you I have learned, there are no impassable obstacles. You were always well reachable with average waiting time of five minutes in Skype to have my answer. Although, we didn't have too long meetings, what I do appreciate the most is how you manage to see the situation, which needs a solution, via student's perspective, thank you.

Harold, thank you for these many years I have had the great opportunity to learn from you how to polish my ongoing projects till the top notch, how to communicate the science and how to make all my scientific results easily reachable for the audience. I remember how in the beginning of research master you showed me the way to entrance 24 (the old Triade building) to introduce me the Epidemiology Department. Since the UMCG is as enormous labyrinth, you said: 'I believe you won't find your way back to this entrance?' However, I did the opposite, now, six years later, I am still here defending my PhD. Thank you!

Krista and Reedik, thank you both for your support and input when requested. It was my honor and pleasure to learn from the greatest experts on the genetics field and to be supervised by all of you.

Although having two academic families has been great, there were also some bureaucratic challenges – lots of paperwork, emails and discussion on a joint PhD procedure between all the parties involved. The ones, who have made this part as smooth as possible are Merilin, Aukje, Lisette, Lilian, Erwin, Mariza and both institution higher forces, who have made it happen. Thank you for all your help and assistance! Especially, Toivo Maimets, who did not hesitate to let me defend in the University of Groningen.

Thank you to Prof. Triin Laisk for reviewing my thesis and for all the opponents here today testing my knowledge!

Last but not least, from the academic family, I am grateful to my research master family without whom I would not have had such a motivation and

inspiration to start my PhD track. I do hope that one day I will have an opportunity to meet and collaborate with all of you mentioned above.

Now getting back to the real family. Thank you for the support and patients with me Mana, Emps, Reinder, Kaur, Liis, Rain, Lennu and the support teams from Norway, US and the Netherlands. Although, some of you might ask this question: 'How long are you planning to study?' I would say, pursuing a PhD is a job as every other, perhaps even more fun since you are the boss of your own time and tasks and the most important – you never stop learning. These four years have provided me a life lesson and I have grown the wings to say *per aspera ad astra*. *When there is something you can dream of, you can also achieve it*!

CURRICULUM VITAE

Name: Katri Pärna
Date of birth: 23rd of April, 1989
Nationality: Estonian
Addresses: University of Tartu, Institute of Genomics, Riia 23b, 51010, Estonia

University Medical Center Groningen, University of Groningen, Department of Epidemiology, Hanzeplein 1, 9713 GZ Groningen, the Netherlands
E-mail: **katri.parna@ut.ee; k.parna@umcg.nl**

EDUCATION

2018–2022 **Doctoral studies in Gene Technology**, Estonia, University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology
Doctoral Studies in Life Course Epidemiology, the Netherlands, University Medical Center of Groningen, University of Groningen, Department of Epidemiology, Unit of Genetic Epidemiology and Bioinformatics

2018 MSc Research Master of Clinical and Psychosocial Epidemiology, the Netherlands, University Medical Center Groningen, University of Groningen, Faculty of Medical Sciences, Department of Epidemiology
Master thesis project "The potential of doubly-weighted genetic risk scores and gene-environment interactions for the prediction of type 2 diabetes."

2015 MSc Master in Molecular and Cell Biology, Master in Zoology and Hydrobiology, Estonia, University of Tartu, Faculty of Science and Technology, Institute of Molecular and Cell Biology; Institute of Ecology and Earth Sciences
Thesis I: "Integrons in *Enterobacteriaceae* isolated from human in Baltic Sea countries."
Thesis II: "Brown bear *(Ursus arctos)* denning areas in Estonia: preferences and the spatial model."
Study duration: 2012–2015

2011 BSc Bachelor in Zoology and Hydrobiology, Estonia, University of Tartu, Faculty of Science and Technology, Institute of Ecology and Earth Sciences
Thesis: "Hibernation and den site selection of brown bear (*Ursus arctos*)"
Study duration: 2008–2011

PROFESSIONAL EMPLOYMENT

2019–2022 University of Tartu, Institute of Genomics, Junior Researcher

TEACHING

2022; 2018 Assisting the course Study Design in Clinical Epidemiology
2021 Supervision of a bachelor thesis

VOLUNTARY WORK

2018– International Alumni Ambassador of University of Groningen https://www.rug.nl/alumni/stay-active/abroad/ambassadors/2018-2019/testimonial-katri-parna-estonia

2017–2018 Student mentor**,** Faculty of Medical Sciences, University of Groningen

INTERNATIONAL COURSES AND CONFERENCES

2021 International Genetic Epidemiology Society, PRS workshop
2021 26th Summer Institute in Statistical Genetics (SISG), University of Washington, Seattle, WA
- Applications of Population Genetics
- Association Mapping: GWAS and Sequencing Data
- Computational Pipeline for WGS Data

2020 Online attendance on the European Society of Human Genetics conference, American Society of Human Genetics conference
2019 Statistical Practice in Epidemiology Using R
2019 Poster presentation at European Society of Human Genetics, Gothenburg, Sweden
2019 HealhtyR Notebooks: a training course for healthcare data analysis
2019 CodeRefinery workshop by Nordic-Infrastructure Collaboration
2018 Doctoral summer course 'Analyses of genotyping and sequencing data in medical and population genetics', Copenhagen, Denmark

AWARDS

2019 Dora Pluss Scholarship
2018 PhD scholarship, the graduate school of medical sciences, the University of Groningen

PUBLICATIONS

In this thesis

Pärna, Katri; Snieder, Harold; Läll, Kristi; Fischer, Krista; Nolte, Ilja (2020). Validating the doubly weighted genetic risk score for the prediction of type 2 diabetes in the Lifelines and Estonian Biobank cohorts. Genetic Epidemiology, 44 (6), 589–600. DOI: 10.1002/gepi.22327.

Marnetto, Davide; **Pärna, Katri**; Läll, Kristi; Molinaro, Ludovica; Montinaro, Francesco; Haller, Toomas; Metspalu, Mait; Mägi, Reedik; Fischer, Krista; Pagani, Luca (2020). Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. Nature Communications, 11 (1), 1628–1628. DOI: 10.1038/s41467-020-15464-w.

Prins, Bram Peter; Leitsalu, Liis; **Pärna, Katri**; Fischer, Krista; Metspalu, Andres; Haller, Toomas; Snieder, Harold (2021). Advances in Genomic Discovery and Implications for Personalized Prevention and Medicine: Estonia as Example. Journal of Personalized Medicine, 11 (5). DOI: 10.3390/jpm11050358.

Pärna Katri, Nolte Ilja M., Fischer Krista, Snieder Harold, Estonian Biobank Research Team ,Marnetto Davide, Pagani Luca (2022). A principal component informed approach to address polygenic risk score transferability across European cohorts. Frontiers in Genetics (accepted 26/05/2022).

Pärna Katri, Lu Xueling, de Vries Maaike, Fraszczyk Eliza, Vonk M. Judith, Boezen Marike, Frank Lude, Zhernakova Alexandra, Fu Jingyuan, van der Most J. Peter, Nolte M. Ilja, Snieder Harold. Effect of methylation and genetic risk scores on prevalent type 2 diabetes and its glycemic endophenotypes. In preparation.

Other publications

Zhang, J., Chen, Z., **Pärna, K.,** van Zon, S.K.R., Snieder, H., Thio, C.H.L. (2022). Mediators of the association between educational attainment and type 2 diabetes mellitus: a two-step multivariable Mendelian randomisation study. Diabetologia. https://doi.org/10.1007/s00125-022-05705-6

Marnetto, D, Pankratov, V., Mondal, M., Montinaro, F., **Pärna, K**., Vallini, L., Molinaro, L., Saag, L., Loog, L., Montagnese, S., Costa, R., Estonian Biobank Research Team, Metspalu, M., Eriksson, A., Pagani, L. (2022). Acestral genomic contributions to complex traits in contemporary Europeans. Current Biology.

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., … **Pärna, K**.,…Hirschhorn. A Saturated Map of Common Genetic Variants Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries. bioRxiv January 2022.

Tammeleht, E., Kull, A., **Pärna, K.** (2020). Assessing the importance of protected areas in human-dominated lowland for brown bear (Ursus arctos) winter denning. Mammal Research, 65 (1), 105–115. DOI: 10.1007/s13364-019-00447-0.

Pavelkovich, A., Balode, A., Edquist, P., Egorova, S., Ivanova, M., Kaftyreva, L., Konovalenko, I., Kõljalg, S., Lillo, J., Lipskaya, L., Miciuleviciene, J., Pai, K., Parv, K., **Pärna, K.,** Rööp, T., Sepp, E., Štšepetova, J., Naaber, P. (2014). Detection of Carbapenemase-Producing Enterobacteriaceae in the Baltic Countries and St. Petersburg Area. BioMed Research International, 2014, Article 548960. DOI: 10.1155/2014/548960.

ELULOOKIRJELDUS

Nimi: Katri Pärna
Sünniaeg: 23. aprill, 1989
Rahvus: Eesti
Kontaktid: Tartu Ülikool, Genoomika Instituut, Riia 23b, 51010 Tartu, Eesti

Groningeni Ülikool, Groningeni Ülikooli Meditsiiniline Keskus, Epidemioloogia osakond, Hanzeplein 1, 9713 GZ Groningen, Holland
E-post: **katri.parna@ut.ee; k.parna@umcg.nl**

HARIDUSKÄIK

2018–2022 **Doktoriõpe geenitehnoloogias**, Eesti, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, Molekulaar- ja rakubioloogia instituut
Doktoriõpe epidemioloogias, Holland, Groningeni Ülikooli Meditsiiniline Keskus, Groningeni Ülikool, Epidemioloogia osakond, Geneetilise epidemioloogia ja bioinfromaatika allüksus

2018 MSc Teadusmagister Kliiniline ja psühhosotsiaalne epidemioloogia, Holland, Groningeni Ülikooli Meditsiiniline Keskus, Groningeni Ülikool, Epidemioloogia osakond, Geneetilise epidemioloogia ja bioinformaatika allüksus.
Magistritöö pealkiri "Topeltkaalutud geneetilise riskiskoori ning geeni-keskkonna koosmõjude potensiaal teist tüüpi diabeedi haigestumuse ennetamisel."

2015 MSc Loodusteaduste magister molekulaar- ja rakubioloogia ning zooloogia ja hürdobioloogia erialadel. Eesti, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, Molekulaar- ja rakubioloogia instituut; Ökoloogia ja maateaduste Instituut.
Magistritöö I: "Integronid Läänemere piirkonna bioloogilistest materjalidest isoleeritud enterobakterites."
Magistritöö II: "Pruunkaru (*Ursus arctos*) talvitusalad Eestis: eelistused ja ruumiline mudel."
Õppekestus: 2012–2015

2011 BSc Loodusteaduste bakalaureus, Eesti, Tartu Ülikool, Loodus- ja täppisteaduste valdkond, Zooloogia ja hüdrobioloogia, Ökoloogia ja maateaduste instituut.
Bakalaureusetöö: "Pruunkaru (*Ursus arctos*) taliuinak ja talvitumiskoha valik"
Õppekestus: 2008–2011

TÖÖKOGEMUS

2019–2022 Tartu Ülikool, Genoomika instituut, nooremteadur

ÕPETAMISKOGEMUS

2022; 2018 Õppeülesannete täitja kursusel 'Uuringudisain Kliinilises Epidemioolgias'

2021 Bakalaureusetöö juhendamine

VABATAHTLIKU TÖÖ

2018– Rahvusvaheline vilistlassaadik Groningeni ülikoolis https://www.rug.nl/alumni/stay-active/abroad/ambassadors/2018-2019/testimonial-katri-parna-estonia

2017–2018 Üliõpilaste mentor, Meditsiinteaduste kõrgkool, Groningeni ülikool

RAHVUSVAHELISED KURSUSED JA KONVERENTSID

2021 Rahvusvaheline geneetilise epidemioloogia ühing, PRS töötuba

2021 26th suveinstituut statistilises geneetikas (SISG), Washingtoni ülikool, Seattle, WA
- Populatsioonigeneetika rakendused
- Assotsiatsioonide kaardistamine: genoomiülesed assotsiatsiooni-uuringud ja sekveneerimisandmed
- Arvutuslik töövoog ülegenoomsetel sekveneermisandmetel

2020 Veebipõhine osalus Euroopa inimgeneetika ühingu konverentsil ja Ameerika inimgeneetika ühingu konverentsil

2019 Statistilised rakendused epidemioloogias R-i tarkvara programmis

2019 Postri esitlus Euroopa inimgeneetika ühingu konverentsil, Göteborg, Rootsi

2019 HealhtyR Notebooks: tervishoiuandmete analüüsi koolituskursus R-I tarkvaras

2019 CodeRefinery, automaattestimise, tarkvaraarenduse ja moodulikoodiarenduse töötuba, mille korraldas The Nordic e-Infrastructure, Tartu, Eesti

2018 Doktorantide suvekursus "Genotüpiseerimis- ja sekveneerimisandmete analüüsid meditsiinilises ja populatsioonigeneetikas", Kopenhaagen, Taani.

AUHINNAD

2019 Dora Pluss Stipendium

2018 PhD stipendium, meditsiinteaduste kõrgkool, Groningeni ülikool

PUBLIKATSIOONID

Loetletud ingliskeelese CV rubriigis publikatsioonid ('Publications')

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 boreonemoral 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 interannual 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.
372. **Kristi Krebs**. Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records. Tartu, 2020, 151 p.
373. **Kärt Ukkivi**. Mutagenic effect of transcription and transcription-coupled repair factors in *Pseudomonas putida.* Tartu, 2020, 154 p.
374. **Elin Soomets**. Focal species in wetland restoration. Tartu, 2020, 137 p.
375. **Kadi Tilk**. Signals and responses of ColRS two-component system in *Pseudomonas putida.* Tartu, 2020, 133 p.
376. **Indrek Teino**. Studies on aryl hydrocarbon receptor in the mouse granulosa cell model. Tartu, 2020, 139 p.
377. **Maarja Vaikre**. The impact of forest drainage on macroinvertebrates and amphibians in small waterbodies and opportunities for cost-effective mitigation. Tartu, 2020, 132 p.
378. **Siim-Kaarel Sepp**. Soil eukaryotic community responses to land use and host identity. Tartu, 2020, 222 p.
379. **Eveli Otsing**. Tree species effects on fungal richness and community structure. Tartu, 2020, 152 p.
380. **Mari Pent**. Bacterial communities associated with fungal fruitbodies. Tartu, 2020, 144 p.

381. **Einar Kärgenberg**. Movement patterns of lithophilous migratory fish in free-flowing and fragmented rivers. Tartu, 2020, 167 p.
382. **Antti Matvere**. The studies on aryl hydrocarbon receptor in murine granulosa cells and human embryonic stem cells. Tartu, 2021, 163 p.
383. **Jhonny Capichoni Massante**. Phylogenetic structure of plant communities along environmental gradients: a macroecological and evolutionary approach. Tartu, 2021, 144 p.
384. **Ajai Kumar Pathak**. Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe. Tartu, 2021, 197 p.
385. **Tanel Vahter**. Arbuscular mycorrhizal fungal biodiversity for sustainable agroecosystems. Tartu, 2021, 191 p.
386. **Burak Yelmen**. Characterization of ancient Eurasian influences within modern human genomes. Tartu, 2021, 134 p.
387. **Linda Ongaro**. A genomic portrait of American populations. Tartu, 2021, 182 p.
388. **Kairi Raime.** The identification of plant DNA in metagenomic samples. Tartu, 2021, 108 p.
389. **Heli Einberg**. Non-linear and non-stationary relationships in the pelagic ecosystem of the Gulf of Riga (Baltic Sea). Tartu, 2021, 119 p.
390. **Mickaël Mathieu Pihain**. The evolutionary effect of phylogenetic neighbourhoods of trees on their resistance to herbivores and climatic stress. Tartu, 2022, 145 p.
391. **Annika Joy Meitern**. Impact of potassium ion content of xylem sap and of light conditions on the hydraulic properties of trees. Tartu, 2022, 132 p.
392. **Elise Joonas**. Evaluation of metal contaminant hazard on microalgae with environmentally relevant testing strategies. Tartu, 2022, 118 p.
393. **Kreete Lüll**. Investigating the relationships between human microbiome, host factors and female health. Tartu, 2022, 141 p.
394. **Triin Kaasiku**. A wader perspective to Boreal Baltic coastal grasslands: from habitat availability to breeding site selection and nest survival. Tartu, 2022, 141 p.
395. **Meeli Alber**. Impact of elevated atmospheric humidity on the structure of the water transport pathway in deciduous trees. Tartu, 2022, 170 p.
396. **Ludovica Molinaro**. Ancestry deconvolution of Estonian, European and Worldwide genomic layers: a human population genomics excavation. Tartu, 2022, 138 p.
397. **Tina Saupe**. The genetic history of the Mediterranean before the common era: a focus on the Italian Peninsula. Tartu, 2022, 165 p.
398. **Mari-Ann Lind**. Internal constraints on energy processing and their consequences: an integrative study of behaviour, ornaments and digestive health in greenfinches. Tartu, 2022, 137 p.
399. **Markus Valge**. Testing the predictions of life history theory on anthropometric data. Tartu, 2022, 171 p.
400. **Ants Tull**. Domesticated and wild mammals as reservoirs for zoonotic helminth parasites in Estonia. Tartu, 2022, 152 p.

401. **Saleh Rahimlouye Barabi**. Investigation of diazotrophic bacteria association with plants. Tartu, 2022, 137 p.
402. **Farzad Aslani**. Towards revealing the biogeography of belowground diversity. Tartu, 2022, 124 p.
403. **Nele Taba**. Diet, blood metabolites, and health. Tartu, 2022, 163 p.

Research Institute SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen.
Further information regarding the institute and its research can be obtained from our internet site: https://umcgresearch.org/w/share

More recent theses can be found in the list below.
(supervisors are between brackets)

Argillander TE
Preoperative risk assessment and optimization of older patients undergoing oncological abdomincal surgery
(prof BC van Munster, dr P van Duijvendijk, dr HJ van der Zaag-Loonen)

Smit AC
The prologue to depression; a tale about complex dynamics and simple trends
(prof M Wichers, prof AJ Oldehinkel, dr E Snippe, dr L Bringmann)

Wildeboer AT
Focus on functioning in person-centered nurse-led diabetes care
(prof PF Roodbol, prof EJ Finnema, dr HA Stallinga)

Barzeva SA
Time alone will tell; longitudinal links between social withdrawal and social relationships across adolescence and early adulthood
(prof AJ Oldehinkel, prof WHJ Meeus, dr JS Klop-Richards)

Dutmer AL
Groningen Spine Cohort; impact of chronic low back pain in patients referred to multidisciplinary spine care
(prof MF Reneman, prof AP Wolff, dr HR Schiphorst-Preuper, dr R Soer)

Banierink H
Pelvic ring injuries; recovery of patient-perceived physical functioning and quality of life
(prof E Heineman, dr JHF Reininga, dr FFA Ijpma)

Olthuis RA
Words have power: talking yourself towards changes in visual control and movement execution
(prof KAPM Lemmink, dr SR Caljouw, dr J van der Kamp)

Summeren JJGT van
Management of childhood functional constipation in primary care
(prof MY Berger, dr JH Dekker, dr GA Holtman)

Ansuategui Echeita J
Central sensitization and physical functioning in patients with chronic low back pain
(prof MF Reneman, prof R Dekker, dr HR Schiphorst Preuper)

Zhang X
Frailty among older adults in the community: insight in the complexity of frailty
(prof CP van der Schans, dr JSM Hobbelen, dr WP Krijnen)

Cheung SL
Family, health, and wellbeing: the lives of Chinese immigrants in The Netherlands
(prof CP van der Schans, dr JSM Hobbelen, dr WP Krijnen)

Overwijk A
Lifestyle opportunities; supporting a healthy lifestyle of people with moderate to profound intellectual disabilities
(prof CP van der Schans, dr AAJ van der Putten, dr A Waninge, dr T Hilgenkamp)

Welling W
Return to sport after an anterior cruciate ligament reconstruction
(prof KAPM Lemmink, dr A Benjaminse, dr A Gokeler)

Terlouw G
The role of boundary objects in health innovation; design for alignment
(prof JPEN Pierie, dr DA Kuipers, dr JTB van 't Veer, dr JT Prins)

Schey-Schulmann CF
Multi-criteria decision analysis for assessing orphan drugs
(prof MJ Postma, prof PF Krabbe)

For earlier theses please visit our website

TARTU ÜLIKOOL

Tartu 2022

ISSN 1024-6479
ISBN 978-9949-03-979-1