DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
**355**

MARIS ALVER

# Value of genomics for atherosclerotic cardiovascular disease risk prediction

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
**355**

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
**355**

**MARIS ALVER**

# Value of genomics for atherosclerotic cardiovascular disease risk prediction

UNIVERSITY OF TARTU
Press

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

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Molecular and Cell Biology on April 8th, 2019 by the Council of the Institute of Molecular Cell Biology, University of Tartu.

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

Tõnu Esko, PhD
Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

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

Opponent: Prof. Martina C. Cornel, MD, PhD
Clinical Genetics/Amsterdam Public Health research institute
Amsterdam UMC, Amsterdam, The Netherlands

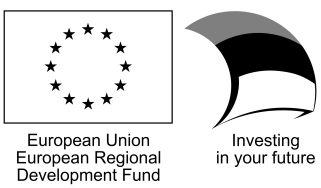
Commencement: Room No. 105, 23B Riia St., Tartu, on June 4th, 2019, at 14:15.

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

This study was funded by EU H2020 grant 692145, Estonian Research Council Grant IUT20-60, IUT24-6, PUT1660, and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), GENTRANSMED end Project No. 2014-2020.4.01.16-0125, the Graduate School in Biomedicine and Biotechnology ASTRA project of the University of Tartu, PER ASPERA, EAS grant no EU49305, NIH – GIANT, ERA-CVD grant Detectin-HF and NIH-BMI Grant 2R01DK075787-06A1.

Data analyses were carried out in part in the High-Performance Computing Center of University of Tartu.

ISSN 1024-6479
ISBN 978-9949-03-024-8 (print)
ISBN 978-9949-03-025-5 (pdf)






University of Tartu Press
www.tyk.ee

# TABLE OF CONTENTS


LIST OF ORIGINAL PUBLICATIONS ..... 7

LIST OF ABBREVATIONS ..... 9

INTRODUCTION ..... 10

1. REVIEW OF THE LITERATURE ..... 12
   1.1. Atherosclerotic cardiovascular disease ..... 12
      1.1.1. Genetics of atherosclerotic cardiovascular disease ..... 13
   1.2. Atherosclerotic cardiovascular disease risk prediction in clinical practice ..... 19
      1.2.1. Commonly used risk prediction tools in clinical practice ..... 20
      1.2.2. Diagnosis of familial hypercholesterolemia in clinical practice ..... 23
   1.3. Advances in improving atherosclerotic cardiovascular disease risk prediction ..... 25
      1.3.1. Lipoprotein(a) ..... 25
      1.3.2. Polygenic risk scores ..... 29
   1.4. Next steps for refining atherosclerotic cardiovascular disease risk prediction ..... 31
      1.4.1. Harnessing biobank information for personalizing atherosclerotic cardiovascular disease risk prediction ..... 31
      1.4.2. Multi-omics approach ..... 35

2. AIMS OF THE STUDY ..... 36

3. RESULTS AND DISCUSSION ..... 37
   3.1. Performance of phenotypic risk scores in a high ASCVD risk setting (Ref. I) ..... 37
      3.1.1. Description of cohort and materials ..... 37
      3.1.2. Performance accuracy of phenotypic risk scores ..... 38
   3.2. Value of high-coverage whole-genome sequencing for dissecting the genetic underpinnings of ASCVD-associated lipid parameters (Ref. II, III) ..... 40
      3.2.1. Description of cohort and materials ..... 41
      3.2.2. Genetic interrogation of Lp(a) ..... 42
      3.2.3. Inter-ethnic differences in genetic determinants of Lp(a) and heritability estimates ..... 43
      3.2.4. Phenotypic consequences of the genetic determinants of *LPA* on ASCVD ..... 44
      3.2.5 Dissection of the genetic basis of LDL-C variation across ancestrally distinct populations ..... 45
   3.3. Genomics-guided approach for identifying individuals at high ASCVD risk for clinical management (Ref. IV) ..... 46
      3.3.1. Description of cohort and materials ..... 47

3.3.2. Overview of the recall-by-genotype study design and clinical management of FH-associated variant carriers ..... 47
3.3.3. Clinical evaluation of FH-associated variant carriers ..... 49
3.3.4. Clinical management of FH-associated variant carriers ..... 51
4. CONCLUSIONS ..... 53
SUMMARY IN ESTONIAN ..... 54
REFERENCES ..... 56
ACKNOWLEDGEMENTS ..... 70
PUBLICATIONS ..... 73
CURRICULUM VITAE ..... 123
ELULOOKIRJELDUS ..... 127

# LIST OF ORIGINAL PUBLICATIONS

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

**I** Saar A*, Läll K*, **Alver M**, Marandi T, Ainla T, Eha J, Metspalu A, Fischer K. 2019. Estimating the performance of three cardiovascular risk scores: the Estonian Biobank cohort study. *J Epidemiol Community Health.* 73(3):272–277.

**II** Zekavat SM, Ruotsalainen S, Handsaker RE, **Alver M**, Bloom J, Poterba T, Seed C, Ernst J, Chaffin M, Engreitz J, Peloso GM, Manichaikul A, Yang C, Ryan KA, Fu M, Johnson WC, Tsai M, Budoff M, Vasan RS, Cupples LA, Rotter JI, Rich SS, Post W, Mitchell BD, Correa A, Metspalu A, Wilson JG, Salomaa V, Kellis M, Daly MJ, Neale BM, McCarroll S, Surakka I, Esko T, Ganna A, Ripatti S, Kathiresan S, Natarajan P; NHLBI TOPMed Lipids Working Group. 2018. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. *Nat Commun*. 9(1):3493.

**III** Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, Bloom JM, Engreitz JM, Ernst J, O'Connell JR, Ruotsalainen SE, **Alver M**, Manichaikul A, Johnson WC, Perry JA, Poterba T, Seed C, Surakka IL, Esko T, Ripatti S, Salomaa V, Correa A, Vasan RS, Kellis M, Neale BM, Lander ES, Abecasis G, Mitchell B, Rich SS, Wilson JG, Cupples LA, Rotter JI, Willer CJ, Kathiresan S; NHLBI TOPMed Lipids Working Group. 2018. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat Commun*. 9(1):3391.

**IV** **Alver M**, Palover M, Saar A, Läll K, Zekavat SM, Tõnisson N, Leitsalu L, Reigo A, Nikopensius T, Ainla T, Kals M, Mägi R, Gabriel SB, Eha J, Lander ES, Irs A, Philippakis A, Marandi T, Natarajan P, Metspalu A, Kathiresan S, Esko T. 2018. Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. *Genet Med*. Oct 1.

My contributions to the listed publications were as follows:

**Ref. I** Participated in the study design, interpreted the results, participated in writing the manuscript.

**Ref. II** Participated in the study design and data analysis, interpreted the results of the association analyses and the results of the mendelian randomization study based on nuclear magnet resonance-measured biomarkers in the biological context (latter not included in the publication), participated in writing the manuscript.

**Ref. III** Analysed the data, interpreted the results, revised the manuscript.

**Ref. IV** Participated in the study design, conducted the analyses, interpreted the results and wrote the manuscript.

# LIST OF ABBREVATIONS

| | |
|---|---|
| ACC/AHA | American College of Cardiology/American Heart Association |
| apo(a) | apolipoprotein(a) |
| apoB | apolipoprotein B |
| ASCVD | atherosclerotic cardiovascular disease |
| bp | base pair |
| CAC | coronary artery calcium |
| CHD | coronary heart disease |
| CI | confidence interval |
| CNV | copy number variation |
| EHR | electronic health records/repositories |
| ESC/EAS | European Society of Cardiology/European Atherosclerosis Society |
| FH | familial hypercholesterolemia |
| GWAS | genome-wide association analysis |
| HDL-C | high-density lipoprotein cholesterol |
| HR | hazard ratio |
| iPSC | induced pluripotent stem cells |
| kb | kilobases |
| KIV | kringle IV |
| KIV2-CN | kringle IV domain type 2 copy number |
| LD | linkage disequilibrium |
| LDL-C | low-density lipoprotein cholesterol |
| LoF | loss of function |
| Lp(a) | lipoprotein(a) |
| Lp(a)-C | Lp(a)-cholesterol |
| MAF | minor allele frequency |
| MEDPED | Make Early Diagnosis to Prevent Early Death |
| MESA | Multi-Ethnic Study of Atherosclerosis |
| MR | mendelian randomization |
| NICE | National Institute for Health and Care Excellence |
| OR | odds ratio |
| oxPL | oxidized phospholipids |
| PCE | Pooled Cohort Equation |
| PGS | polygenic risk score |
| QALY | quality-adjusted life year |
| RbG | recall-by-genotype |
| SCORE | Systemic COronary Risk Evaluation |
| SD | standard deviation |
| SIR | standardized incidence ratios |
| SNV | single nucleotide variant |
| VAP | Vertical Auto Profile |
| VLDL | very low-density lipoprotein |
| WES | whole-exome sequencing |
| WGS | whole-genome sequencing |

# INTRODUCTION

Atherosclerotic cardiovascular disease (ASCVD) is the world's number one killer, markedly affecting healthy life years globally, and demanding urgent and well-planned strategies to avert and combat the detrimental effects it marks upon societies at large. One of the practical approaches to the latter purpose would be to enhance preventive strategies by improving risk prediction and risk stratification. The aspect that ASCVD is a complex multifactorial disease, however, impedes straight-forward solutions and requires sturdy endeavours to enhance the understanding of the phenotype. The underlying pathology of ASCVD is atherosclerosis, an inflammatory condition that develops over many years and is often advanced by the time symptoms occur. Extensive research on ASCVD has accelerated our knowledge on the epidemiological, molecular and genetic aspects of the phenotype, yet the full comprehension of the disease is far from complete.

Risk prediction guidelines accompanied by risk prediction tools endorsed by cardiological societies have been the cornerstone of ASCVD prevention strategies for decades and are widely used in today's clinical setting. While the risk prediction tools generally do identify individuals at the highest risk for therapeutic interventions, these lack in precision and can misestimate disease risk, thereby inaccurately targeting individuals for preventative therapies. More importantly, as the risk estimation of these models relies on the assessment of traditional phenotypic risk factors, these in reality foretell ASCVD risk based on already perturbed molecular pathways and hence lack in providing accurate estimation for primordial prevention, that is estimating risk before risk factors emerge.

One of the current, hotly debated topics regarding the precision and improvement of ASCVD risk prediction has centred on the incorporation of genetic information into the daily clinical setting. The scrutiny on the genetic architecture of ASCVD conducted in recent decades has now resulted in estimates that can be of clinical utility and value. Research has hitherto revealed numerous rare and common genetic loci associated with the phenotype, and unveiled molecular pathways and trajectories leading to the disease. Moreover, it has incited large-scale collaboration projects and propelled the development of novel approaches to utilize and combine vast datasets, and new strategies to address hypotheses and research questions.

Today, we have arrived at the dawn of a paradigm shift that will bring molecular and genetic research at the doorstep of routine clinical practice. This is brought about by the plummeting costs in array and sequencing technologies, the acknowledgment of the value of integrating different information levels and sources in data analysis, and the creation of novel research avenues triggered by the detailing of the unravelled genetic findings in different molecular systems. In this doctoral thesis, I have aimed to give an overview of the endeavours that have set the stage for this transition. Furthermore, I will bring examples on how

high-coverage sequencing can advance the genetic interrogation of ASCVD-related phenotypes and how the utilization of the information contained within a population-based biobank can truly progress precision prevention of ASCVD.

# 1. REVIEW OF THE LITERATURE

## 1.1. Atherosclerotic cardiovascular disease

Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of morbidity and mortality worldwide, accounting for more than 15 million deaths annually (WHO, 2016). While the prevention and treatment strategies applied in routine clinical care, and the attention gained for attaining a healthy lifestyle have substantially contributed to the declining ASCVD mortality rates in Europe, Estonia is still among the top countries with the highest ASCVD incidence rates and considered a high-risk country in terms of cardiovascular mortality by the European Society of Cardiology (Piepoli et al., 2016).

ASCVD is a complex multifactorial disease that results from the interplay of genetic susceptibility and environmental factors. The root cause of the disease is atherosclerosis, a slowly progressive chronic inflammatory condition of the large and medium size arteries. It is driven by the infiltration of lipoprotein particles into the coronary and carotid arteries due to the dysfunction of the endothelial cells that line these vessels. A subsequent inflammatory response leads to the formation of atherosclerotic plaques, that consist of macrophage- and vascular smooth muscle cell-derived foam cells, apoptotic and necrotic cells and cellular debris, and are covered by a collagen, proteoglycans and smooth muscle cells containing fibrous cap. Besides obstructing the blood flow, plaques can rupture due to intrinsic lytic processes and lack of repair that drive the cap weakening, thus causing a thrombotic occlusion of the artery and hence manifesting as myocardial infarction or stroke (Erdmann et al. 2018; Schankman et al., 2015; Khera and Kathiresan, 2017; Falk, 2006; Lafont, 2003) (Fig. 1).

Extensive epidemiological studies conducted in the past decades have established a number of risk factors that contribute to the progression of the disease. These include age, sex, smoking status, elevated blood pressure, diabetes, obesity and sedentary lifestyle. Increased levels of circulating low-density lipoprotein cholesterol (LDL-C) and triglyceride-rich lipoproteins, and decreased concentrations of high-density lipoprotein cholesterol (HDL-C) have also been defined as key clinically attainable biomarkers associated with the disease (Khera and Kathiresan, 2017). Knowledge of these risk factors have led to the development of phenotypic risk prediction models that are now commonly used in routine clinical practice to target those at high risk for lifestyle modifications and/or preventative therapies.

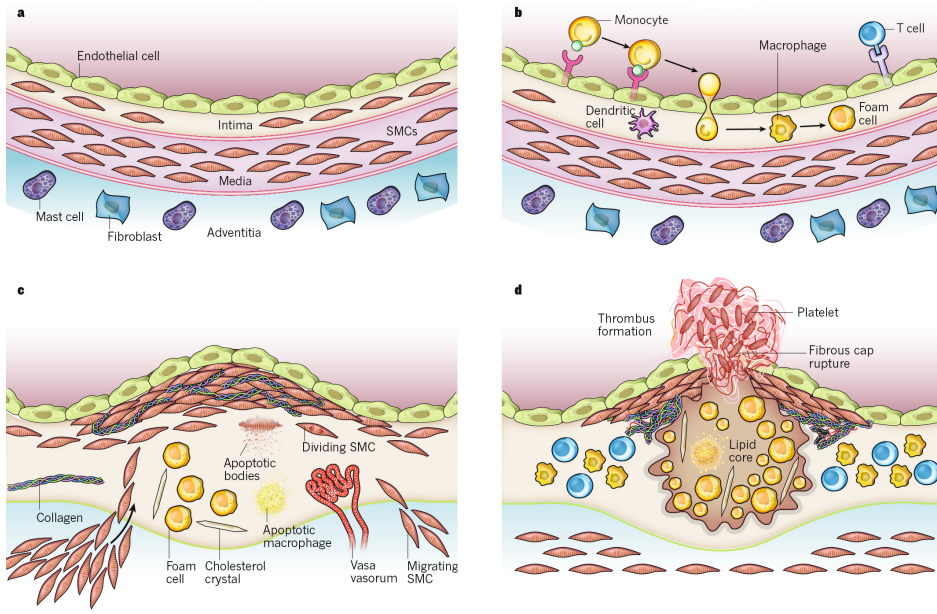



**Figure 1.** Overview of the progression of atherosclerosis. a) An artery contains three layers: tunica intima, lined by a monolayer of endothelial cells; tunica media, containing smooth muscle cells embedded in extracellular matrix; and adventitia, consisting of immune cells, nerve endings and micro-vessels. b) The progression of atherosclerosis starts with the adhesion of leukocytes to the activated endothelial cells, migration of the bound leukocytes into intima, and maturation of monocytes into macrophages and subsequently into foam cells via the uptake of lipoproteins. c) Atherosclerotic plaque progression involves the migration and proliferation of smooth muscle cells, and synthesis of extracellular matrix macromolecules such as collagen, elastin and proteoglycans that form a fibrous cap that covers the plaque. Extracellular lipids derived from apoptotic and necrotic cells accumulate as a necrotic lipid core. d) Thrombosis can result from the fracture of the fibrous cap (usually thin and collagen-poor with abundant macrophages), leading to the exposure of the pro-coagulant material in the core to coagulation proteins in the blood. SMC – smooth muscle cell. (Libby et al., 2011).

### 1.1.1. Genetics of atherosclerotic cardiovascular disease

Large-scale twin and family studies have estimated the heritability of ASCVD to be approximately 50–60% and indicated that family history of the disease in a parent or sibling is a strong predictor of the incidence of ASCVD (Murabito et al., 2005; Lloyd-Jones et al., 2004; Zdravkovic et al., 2002). This knowledge has propelled extensive research efforts for understanding the genetic underpinnings of atherosclerosis-driven manifestations, driven largely by the dissection of the genetic determinants in families predisposed to premature ASCVD and the assessment of the polygenic architecture in large-scale genomic analyses. These endeavours have subsequently motivated functional studies

for causality inference and broad collaboration efforts for deepening the comprehension of ASCVD pathophysiology.

#### 1.1.1.1. Family-based studies

The initial clues for the genetic component of ASCVD were first reported in the late 1930s when the family clustering of increased cholesterol levels and premature coronary heart disease (CHD) were identified (Müller, 1938). Subsequent studies using linkage analysis narrowed the genetic cause to the *LDLR* gene (encodes the LDL receptor responsible for the hepatic cholesterol-containing LDL particle uptake), and confirmed that the pedigree structures were consistent with the autosomal dominant inheritance of a single defective gene (Goldmann et al., 2010; Lehrman et al., 1985). Today, familial hypercholesterolemia (FH), the monogenic disorder of ASCVD, is considered as one of the most common single-gene disorders with the prevalence of 1 in 217 (Benn et al., 2016). The main cause of the disease is the dysfunction of the LDL receptor, that results in increased LDL-C and cholesterol levels in plasma, and premature atherosclerotic progression (Khera and Kathiresan, 2017). Over 1,700 genetic variants in the *LDLR* gene have been linked to FH (Leigh et al., 2017) with additional deleterious variants identified in the *APOB* (hampered binding of LDL particles to LDL receptor for uptake (Soria et al., 1989)) and gain-of-function variants in the *PCSK9* gene (increased LDL receptor catabolism (Abifadel et al., 2003)).

A separate pathway leading to a monogenic form of ASCVD, that is independent from the hampered clearance of cholesterol, has also been acknowledged (Erdmann et al., 2018). A family-based study with functional follow-up experiments in mice noted rare deleterious variants in genes related to nitric oxide signalling in multiple affected family members with premature myocardial infarction. The genetic variants in *GUCY1a3* and *CCT2* were functionally associated with diminished ability to generate nitric oxide from endothelial cells, increased platelet activation and arterial thrombosis in mouse models (Erdmann et al., 2013). This study underscores the value of conducting family-based studies with the potential to pinpoint yet unknown disease trajectories to be functionally tested for disease causation.

#### 1.1.1.2. Common variant association studies

While ample evidence supports the co-segregation of a genetic determinant with premature ASCVD within families, the monogenic disorders account for only a minority of all ASCVD cases and do not explain the high prevalence of the disease in the population at large. Extensive research on the aetiology of ASCVD has established that the disease is far more complex with multiple pleiotropic pathways contributing to the disease progression. Significant

advances in molecular technology and computational capacities have led to tremendous progress in understanding the genomic architecture of complex diseases. Use of genotyping arrays that capture common genetic variation and the advances of haplotype imputation have laid the foundation for genome-wide association studies (GWAS). While commercially available genotyping arrays enable to ascertain genomic variation in pre-selected sites, usually chosen to represent the most informative single nucleotide variants (SNV) within a haplotype based on linkage disequilibrium (LD), imputation facilitates the determination of untyped SNVs based on a common haplotype reference panel built using whole-genome sequencing (WGS) data (e.g. 1000 Genomes project (Auton et al., 2015)). Essentially, in a genome-wide scan, the relationships between common genome sequence variations and disease predisposition or trait variation in unrelated individuals are systematically surveyed on a genome-wide scale. Each variant is tested individually and those, that are found to be significantly more frequent in individuals with a trait of interest, indicate genomic regions that likely affect the phenotype. The statistical power to detect the associations largely depend on the sample size, the distribution of effect sizes and frequency of putative causal genetic variants segregating within the population, and the LD between the genotyped or imputed genetic variants and true causal variants (Visscher et al., 2017; Marchini and Howie, 2010; McCarthy et al., 2008).

The common variant search for the phenotype started in 2007, when four independent groups with sample sizes ranging from 5,000 to 23,000 individuals identified a common variant at the 9p21 locus to be associated with a 30% increased risk for ASCVD per copy of the risk allele (Wellcome Trust Case Control Consortium, 2007; Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). Subsequent genome-wide scans have consistently replicated this finding demonstrating the strongest signal and effect (Nikpay et al., 2015; CardioGRAMplusC4D Consortium, 2012; Schunkert et al., 2011), yet the causal mechanism, despite ten years of scrutiny, remains uncertain. Rigorous interrogation of this association has proposed a number of intertwined hypotheses. It is suggested that 9p21 risk variants could alter the activity of two nearby cyclin-dependent kinase inhibitors (CDKN2A and CDKN2B), involved in regulating cell cycle and cellular proliferation, via modulating the expression of non-coding RNA *ANRIL* that resides in this locus (Holdt et al., 2010; Jarinova et al., 2009), or that the expression of these kinase inhibitors is affected by interferon-γ-mediated inflammatory signalling (Harismendy et al., 2011). Moreover, the effect has also been endowed to the ratio of circular to linear ANRIL, whereby the circular form confers protection from atherosclerosis by controlling ribosomal biogenesis in vascular smooth muscle cells and macrophages (Holdt et al., 2016). Lastly, the latest study utilizing the generation of vascular smooth muscle cells from genome edited induced pluripotent stem cells (iPSC) of individuals with and without the risk haplotype proposed that the risk haplotype might induce a cellular phenotypic switch predisposing vascular smooth muscle cells to increased proliferation as well as reduced contraction and adhesion (Lo Sardo et al., 2018). It is highly plausible that the causal

mechanism in this locus is manifested by the intricate interactions between different regulatory elements in the non-coding genome and regulatory molecules in different cell types. Future studies attempting to resolve this mechanistic association in fine detail will prove a major step forward in revealing the vast, pleiotropic and highly complex interplay within the noncoding genome and across different molecular levels that can lead to the concomitant perturbations of various molecular pathways.

Fast-forwarding to this day, markedly larger sample sizes have been used in the GWA analyses of ASCVD, substantially broadening the spectrum of the genetic associations and linking approximately 160 common genetic loci with the disease (van der Harst and Verweij, 2018; Klarin et al., 2017; Nelson et al., 2017; Nikpay et al., 2015; CardioGRAMplusC4D Consortium, 2012; Schunkert et al., 2011; Coronary Artery Disease Genetics Consortium, 2011). The vast majority of the genetic variants, however, are common with minor allele frequency (MAF) of >5% in the population and are associated with modest increases in ASCVD risk (<10% change in risk per allele). While these loci cumulatively explain roughly 30–40% of the disease heritability, these account for a far greater proportion of ASCVD heritability than identified rare variants combined. Even though the rare variant effects are larger compared to the effects of common variants, these are present in a much smaller number of individuals and therefore explain less of the overall heritability of the disease. Furthermore, GWAS results have highlighted different trajectories of the disease with pathways linked to lipid metabolism, blood pressure, inflammation, cellular proliferation, vascular remodelling, vascular tone and nitric oxide signalling (Fig. 2). Yet for the majority of the uncovered loci, the exact mechanism leading to the disease remains elusive (Erdmann et al., 2018; Khera and Kathiresan, 2017).

Due to the essence of the GWA study design, the variants identified as most significant oftentimes reside within loci spanning several kilobases (kb) encompassing multiple genes and are generally in LD with true causal variants. Despite sturdy efforts of untangling the causal effects for the identified loci, mechanistic associations have been unveiled only for a handful. One of the first functionally resolved locus associated with decreased LDL-C concentrations (Willer et al., 2008) and CHD risk (Kathiresan et al., 2009) is the non-coding region on chromosome 1p31 encompassing genes *CELSR2*, *PCRC1* and *SORT1*. In a series of functional studies, the causal variant was pinpointed to rs12740374, where the minor allele creates a new binding site for CCAAT-enhancer-binding transcription factor (C/EBPα) and thereby alters the hepatic expression of the *SORT1* gene (Musunuru et al., 2010). The overexpression of *SORT1* results in the decreased secretion of very low-density lipoprotein (VLDL) and apolipoprotein B (apoB), and reduced total plasma cholesterol and LDL-C levels (Strong et al., 2012). Detailed elucidation of the *ADAMTS7* locus on chromosome 15 similarly clarified the protective association with ASCVD (Schunkert et al., 2011), whereby the minor allele of top variant rs3825807 affected gene maturation and smooth muscle cell migration *in vitro* (Pu et al.,

2013). Moreover, studies on *Adamsts7−/−* knockout mice highlighted decreased atherosclerosis burden and cellular proliferation, and enhanced cell repair in response to injury (Bauer et al., 2015; Kessler et al., 2015), making ADAMTS7 a potential therapeutic target. Lastly, while a 6p24 locus variant rs9349379 within the intron of *PHACTR1* was linked with the expression of the given gene in vascular tissues, albeit without a plausible biological mechanism on atherogenesis progression (Wang and Musunuru, 2018; Beaudoin et al., 2015), a study using fine-mapping approach, epigenome profiling and CRISPR/Cas9 genome editing using iPSC-derived endothelial cells identified the rs9349379 variant to regulate the expression of endothelin 1 gene 600 kb upstream (Gupta et al., 2017), previously shown to promote atherosclerosis plaque development (Amiri et al., 2004).

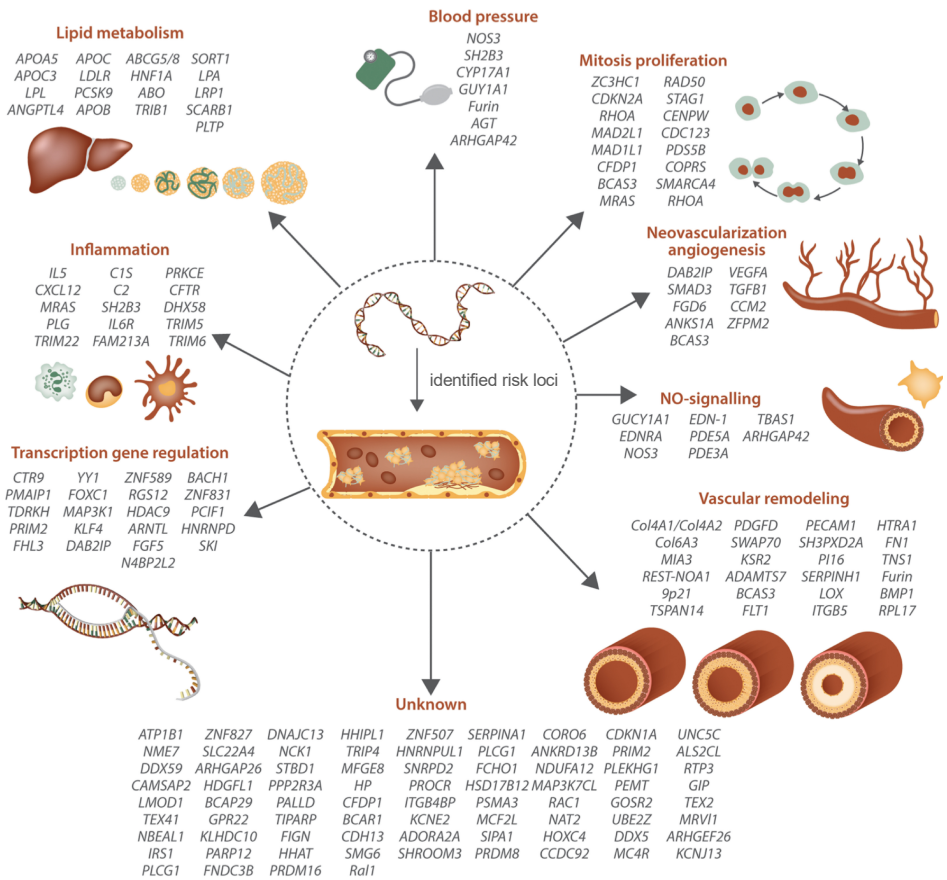



**Figure 2.** Genes and molecular pathways linked to the genetic loci associated with ASCVD through GWA studies (Erdmann et al., 2018).

Importantly, given that ASCVD is a multifactorial disease, multiple GWA analyses have additionally been conducted for the established risk factors, including lipid parameters (Surakka et al., 2015), diabetes mellitus type II (Mahajan et al., 2018), smoking (Liu et al., 2019), body mass index (Speliotes et al., 2010) and blood pressure (Evangelou et al., 2018). Attention is now being centred on unravelling the underlying shared genetic determinants and polygenic overlap across these phenotypes. With the development of bioinformatics-based approaches (e.g. LD Score regression for estimating genetic correlation between traits (Bulik-Sullivan et al., 2015), or two-sample mendelian randomization for causal effect estimation (Hartwig et al., 2016)), summary statistics of multiple genomic studies can be leveraged for identifying shared genetic components among phenotypes driving atherosclerosis progression and to determine their causal effect on ASCVD (Zhu et al., 2018; LeBlanc et al., 2016; van Iperen et al., 2016; Bulik-Sullivan et al., 2015).

#### 1.1.1.3. Rare variant association studies

The advent of next-generation sequencing has allowed to assess, whether rare variants (MAF <1%) not typically included in genotyping arrays could additionally be ascertained in population-based studies, and facilitated bioinformatics and statistical approaches, e.g. gene-based rare variant testing, for novel gene discoveries. While in genome-wide scans, a single rare variant is usually underpowered to yield a statistically significant signal when comparing individuals affected by disease to control subjects, aggregating rare variants across a gene can improve power. In the latter case, rare variants within genes are combined into a genomic unit, thereby enabling the comparison of the aggregate frequencies between affected and unaffected individuals (Guo et al., 2016; Lee et al., 2014). Unsurprisingly, in the first gene-based rare variant analysis with approximately 5,000 premature CHD cases and 5,000 control subjects using whole-exome sequencing (WES) data, Do et al. identified the *LDLR* gene as the strongest signal with four-fold increased risk for CHD in 2% of the cases. However, they additionally identified that carriers of rare non-synonymous variants in *APOA5* (protein product is a component of VLDL, HDL and chylomicron particles (Guardiola and Ribalta, 2017)) were at two-fold increased risk for myocardial infarction (Do et al., 2014). Another hypothesis-driven analysis on the *LPL* gene (protein product modulates the clearance of dietary fat from the circulation (Mead et al., 2002)) identified that 0.4% of the studied individuals with a rare deleterious variant in *LPL* were at two-fold increased risk for CHD and displayed higher triglyceride levels (Khera et al., 2017).

Importantly, rare variant analyses have additionally unveiled genetic markers that confer protection from the disease. As opposed to gain-of-function variants identified in FH cases, two inactivating genetic variants in *PCSK9* in individuals of African ancestry have been linked to substantially lower LDL-C levels and up to 88% reduced risk for CHD (Cohen et al., 2005; Cohen et al.,

2006). Protective low-frequency variants have also been found in *ANGPTL4* and *APOC3* (protein products regulate plasma triglyceride levels by inhibiting lipoprotein lipase (Lafferty et al., 2013; Zheng et al., 2010) with 11% to 40% decreased risk (Stitziel et al., 2016; Dewey et al., 2016; Crosby et al., 2014; Jørgensen et al., 2014). A comprehensive study based on imputed array data of 119,146 Icelanders with reference haplotypes derived from the whole-genome sequences of 2,636 Icelanders revealed that a 12-base-pair (bp) deletion in *ASGR1* leads to the inactivation of the asialo-glycoprotein receptor (transmembrane protein that plays a role in the homeostasis of circulating glycoproteins) in heterozygous carriers (1% of the study population) and is associated with 34% lower CHD risk and lower non-HDL levels (Nioi et al., 2016). This study was among the first highlighting the value of deriving a population-specific reference panel from whole-genome sequences of a representative set of a population. This approach benefits the imputation of missing genotypes into the genomes of a larger proportion of a specific population with array-derived genotypes and thus enhances the accuracy of identifying population-specific low-frequency variants that can putatively lead to the ascertainment of novel disease pathways.

The discovery of protective variants has instigated intense interest in developing potential new drug targets to combat ASCVD. For instance, the scrutiny on the effect of inactive variants in *PCSK9* resulted in the development of monoclonal antibody-based inhibitors. These inhibit PCSK9 by binding to the protein with high affinity and thereby demolish the ability of the protein to bind to and catabolize the LDL receptor. Clinical studies have shown that the PCSK9 inhibitors decrease circulation LDL-C levels by >50%, that is 30% to 40% more compared to current medical treatment (Schmidt, 2017). Similarly, antisense oligonucleotides that mimic the protective variant in *APOC3* have shown to reduce 70% of triglyceride levels in early-phase studies (Burdett, 2015). These examples underscore the value of using WES and WGS datasets for rare variant discovery, and for unearthing novel putative drug targets that can ultimately benefit populations at large.

## 1.2. Atherosclerotic cardiovascular disease risk prediction in clinical practice

Early and accurate identification of individuals at high ASCVD risk is critical for effective clinical intervention before these individuals develop overt cardiovascular events. Estimating the probabilistic susceptibility of an individual to the disease is central to today's clinical decision-making (Khambhati et al., 2018; Torkamani et al., 2018; Abraham et al., 2016). To this end, several risk prediction tools that rely on demographic characteristics, basic health parameters, and lifestyle and clinical factors have been constructed in order to target risk groups for preventative measures and treatment initiation. Consideration of different criteria for risk estimation can, however, lead to differences in the

proportion of individuals categorized for preventative therapies at a population level (Gray et al., 2014). Therefore, to accurately estimate disease risk, effective and reliable tools that are rigorously validated before clinical applicability for the target population are required.

### 1.2.1. Commonly used risk prediction tools in clinical practice

The ASCVD prevention guidelines endorsed by the American College of Cardiology/American Heart Association (ACC/AHA) (Goff et al., 2014), by the European Society of Cardiology jointly with the European Atherosclerosis Society (ESC/EAS) (Piepoli et al., 2016) and by the UK National Institute for Health and Care Excellence (NICE) (NICE, 2018) are the three major guidelines considered in routine clinical practice. While their common aim is to stratify individuals based on risk factors into risk categories, the guidelines differ in terms of prediction algorithms, predicted endpoints and risk factors considered in companion risk prediction tools, as well as the risk thresholds and LDL-C cut points for assignment of therapeutic treatment (Greenland and RO, 2016).

The Pooled Cohort Equation (PCE) is the companion tool to the 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk to estimate a 10-year risk for hard ASCVD (defined as coronary death or non-fatal myocardial infarction, or fatal or non-fatal stroke) in individuals of 40–79 years of age. The risk factors considered in the score include age, sex, ethnicity, total cholesterol, HDL-C, systolic blood pressure, antihypertensive medication use, diabetes and smoking status (Table 1). The risk estimates were derived from several large, racially and geographically diverse cohorts, including the Atherosclerosis Risk in Communities study, the Cardiovascular Health Study, the Coronary Artery Risk Development in Young Adults study and the Framingham/ Framingham Offspring Study cohorts, and were externally validated on six independent study cohorts. The ACC/AHA guidelines advise to consider statin therapy for individuals, whose estimated 10-year risk for ASCVD is above ≥7.5% according to the prediction tool. Individuals at intermediate risk (10-year absolute risk of 5–7.5%) could be considered for statin therapy for primary prevention based on patient preferences or perceived benefit based on additional risk factors. Importantly, the ACC/AHA guideline does not endorse a treat-to-target strategy, i.e. aiming specific HDL-C and LDL-C levels, but instead specifies the appropriate intensity of statin for each risk category (Nayor and Vasan, 2016; Goff et al., 2014; Stone et al., 2013).

NICE recommends the QRISK2 algorithm to estimate a 10-year risk of developing a fatal or non-fatal ASCVD event (defined as angina pectoris, myocardial infarction, coronary heart disease, stroke and transient ischaemic attack) for individuals aged 25–84 years without diabetes mellitus type I. The considered risk factors include age, sex, ethnicity, smoking status, systolic blood pressure, ratio of total cholesterol and HDL-C, body mass index, family history of CHD in a first degree relative <60 years of age, social deprivation index (UK

postcode), antihypertensive medication use, and status of rheumatoid arthritis, chronic kidney disease, diabetes and atrial fibrillation (Table 1). The score was developed on 563 general practices in England and Wales with 2/3 of the participants randomly allocated to the derivation dataset and 1/3 assigned to the validation dataset. NICE Guidelines recommend statin therapy for primary prevention in individuals with diabetes mellitus type II, or in those without diabetes, but with an estimated 10-year absolute ASCVD risk of ≥10%. Similarly to ACC/AHA guidelines, treat-to-target strategy is not endorsed (NICE, 2018; Nayor and Vasan, 2016; Hippisley-Cox et al., 2010).

**Table 1.** Overview of the considered risk factors, predicted endpoints and target population of three commonly used risk prediction tools: QRISK2 used in the UK, Pooled Cohort Equation (PCE) applied in the US, and Systemic COronary Risk Evaluation recommended in Europe.

| **QRISK2 (UK)**<br>**fatal and non-fatal ASCVD**<br>**40–79 years of age** | **PCE (US)**<br>**hard ASCVD**<br>**25–84 years of age without DM type I** | **SCORE (Europe)**<br>**fatal CVD**<br>**>40 years of age without CKD** |
|---|---|---|
| age | age | age |
| sex | sex | sex |
| systolic blood pressure | systolic blood pressure | systolic blood pressure |
| smoking status | smoking status | smoking status |
| total cholesterol/HDL-C ratio | total cholesterol | total cholesterol or total cholesterol/HDL-C ratio |
| | HDL-C | |
| ethnicity | ethnicity | |
| diabetes | diabetes | |
| antihypertensive treatment | antihypertensive treatment | |
| body mass index | | |
| atrial fibrillation | | |
| rheumatoid arthritis | | |
| chronic kidney disease | | |
| family history of ASCVD | | |
| UK postcode (proxy for social deprivation) | | |

DM – diabetes mellitus; CKD – chronic kidney disease

The Systemic COronary Risk Evaluation (SCORE) was initiated to develop a risk scoring system for clinical management of cardiovascular risk in the European clinical practice. It targets individuals above 40 years of age without chronic kidney disease and estimates 10-year cardiovascular mortality (opposed

to combined fatal and non-fatal events). Sex, age, systolic blood pressure, total cholesterol or ratio of total cholesterol and HDL-C, and smoking status are considered in the algorithm (Table 1). The risk estimation was derived from a pooled dataset of population-based and occupational cohort studies from 12 European countries, and externally validated on 11 European cohorts. Opposite to the US and the UK guidelines, a treat-to-target strategy based on estimates extrapolated from clinical trials is applied. Drug intervention is recommended for those, whose SCORE-based 10-year fatal ASCVD risk is 5–9.9% and LDL-C level is ≥2.6 mmol/L (100 mg/dL), or risk ≥10% and LDL-C level ≥1.8 mmol/L (≥70 mg/dL). Therapy is considered for those, whose SCORE-based 10-year fatal ASCVD risk is <1% and LDL-C level is ≥4.9 mmol/L (190 mg/dL), risk ≥1–4.9% and LDL-C level ≥2.6 mmol/L (≥100 mg/dL), risk ≥5–9.9% and LDL-C level ≥1.8–2.6 mmol/L (≥70–100 mg/dL), or risk ≥10% and LDL-C level <1.8 mmol/L (<70 mg/dL) (Catapano et al., 2016; Nayor and Vasan, 2016; Piepoli et al., 2016).

#### 1.2.1.1. Efficacy of commonly used risk prediction tools

Due to the different strategies considered, a number of studies have been conducted to evaluate the predictive ability of these guidelines and companion tools on independent cohorts. Interestingly, the results consistently demonstrate overestimation of ASCVD risk, thereby predisposing a great proportion of individuals for therapeutic intervention (DeFilippis et al., 2017; Defilippis et al., 2015; Demir et al., 2015; Kavousi et al., 2014; Mortensen and Falk, 2014; Van Staa et al., 2014). For instance, a comparison of ACC/AHA and ESC guidelines on 7279 ASCVD-free individuals aged 40 to 75 years showed that the recommendations by these guidelines highly overlapped (95.8%), yet the US guidelines advocated statin treatment for substantially more individuals (58.9%) compared to the ESC guidelines (33.0%) (Pavlović et al., 2016). Two studies on more than >40,000 individuals classified 42% of the individuals for statin treatment according to the ACC/AHA guidelines while <15% by the ESC/EAS guidelines (Mortensen and Nordestgaard, 2018; Mortensen et al., 2017).

The disproportional overestimation of disease risk can partly be attributed to the discordance of baseline risk between the scores' derivation and target cohorts. The clinically applicable risk scores were developed decades ago on cohorts that have hitherto followed the premise of improving lifestyle habits and incorporating therapeutic interventions into risk reduction, and thus do not translate into similar incidence rates today. With overestimation, a great number of individuals can be targeted for unnecessary statin treatment (Kavousi et al., 2014). On the other hand, setting too high of a threshold for treatment recommendations for primary prevention can miss a great proportion of individuals, who later do develop an ASCVD event, thus undermining the main aim of clinical risk prediction (Mortensen and Nordestgaard, 2018). Additionally, it needs to be considered that treatment benefits are not constant for all indi-

viduals and may vary markedly among them (Thanassoulis et al., 2016). Therefore, for improved risk prediction, further steps are required to accurately ascertain individuals for targeted therapy.

### 1.2.2. Diagnosis of familial hypercholesterolemia in clinical practice

Despite the clear genetic cause for FH, the disease is most often diagnosed clinically. The phenotype is characterized by clinical and/or family history of premature ASCVD, severely elevated plasma LDL-C levels (≥4.9 mmol/L or ≥190 mg/dl) and presence of physical features such as xanthomata, xanthelasma and/or corneal arcus (Reiner, 2015). While characteristics required for clinical FH diagnosis is generally agreed upon, a universal consensus on the weighting of these features is thus far lacking (Berberich and Hegele, 2018). There are three formal diagnostic scores that have been proposed and used in the clinical setting. These are the US Make Early Diagnosis to Prevent Early Death (MEDPED) criteria (Williams et al., 1993), the Simon Broome Register Group criteria in the UK (Simon Broome Register Group, 1991) and the Dutch Lipid Clinic Network diagnostic score (Defesche et al., 2004). While the latter two are similar in terms of considered risk factors and inclusion of genetic testing, the MEDPED criteria rely on age, family history and total cholesterol only.

Although FH is clinically and genetically well-described, it remains significantly under-diagnosed and inadequately treated. Epidemiological studies highlight that FH patients are diagnosed late in life, do not receive timely and adequate therapy, and display premature subclinical atherosclerotic changes (Degoma et al., 2016; Krogh et al., 2016; Perak et al., 2016). Poor performance of the diagnostic tools can in part be attributed to the lack of physical features, absence of or difficulty in ascertaining family history, or identification of LDL-C levels lower than the set thresholds (Nordestgaard et al., 2013; Civeira et al., 2008; Kee et al., 1993). Khera et al. described that among individuals with LDL-C levels above the clinically defined threshold (≥4.9 mmol/L or ≥190 mg/dl), less than 2% harboured an FH-associated variant. However, the risk for ASCVD in these individuals was three-fold greater compared to non-carriers with similarly high LDL-C levels, and two-fold greater in carriers compared to non-carriers displaying borderline high LDL-C levels (130 mg/dl or 3.4 mmol/L) (Khera et al., 2016) (Fig. 3). These results clearly illustrate that lifelong exposure to increased LDL-C levels due to a genetic defect drives atherosclerosis progression and that the improvement for targeting FH-associated variant carriers, especially those whose LDL-C levels are below clinically set thresholds, is demanded.

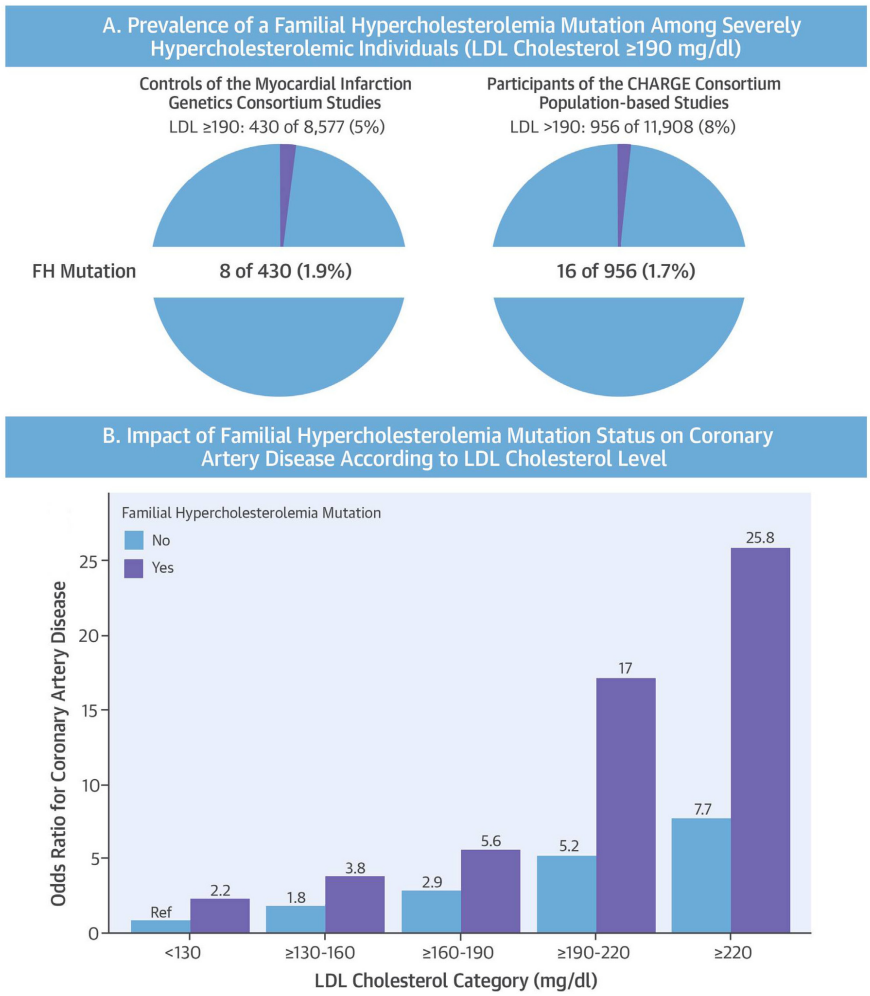



**Figure 3.** Prevalence of FH-associated genetic variants and impact on coronary heart disease risk. A. Among severely hypercholesterolemic (LDL-C ≥190 mg/dl or ≥4.9 mmol/L) individuals (5% and 8% in 2 separate study cohorts) less than 2% harboured an FH-associated variant. B. Risk of coronary heart disease across LDL-C categories in FH-associated variant carriers and non-carriers. Adapted from (Khera et al., 2016).

In light of the shortcomings in diagnosis, high prevalence of the disease, and high mortality and morbidity if left untreated, screening for FH is undoubtedly warranted (Wonderling et al., 2004). One approach to identify individuals affected with FH is genetic testing-based cascade screening. Targeting relatives based on genetic testing has shown to not only reduce the average age at which individuals with FH are diagnosed and receive timely treatment, but has also been confirmed to be cost-effective (Kerr et al., 2017; Lázaro et al., 2017; Wonderling et al., 2004). One of the success stories of cascade screening is the targeted screening program for FH established in the Netherlands in 1994. As

the participation rate in a five-year period yielded 90%, the program resulted in the identification of eight family members with the disease-causing variant per proband and significantly increased the proportion of carriers receiving treatment (Umans-Eckenhausen et al., 2001). The cost-effectiveness of the Dutch targeted screening program was shown subsequently (Wonderling et al., 2004). NICE published guidelines for identification and management of FH and recommendations for cascade screening in 2008 (DeMott et al., 2008). An analysis based on the UK screening services data was undertaken in 2017 by Kerr et al. It was shown that with testing 1.33 relatives per proband, the cost per quality-adjusted life year (QALY) of DNA cascade testing per tested relative was estimated to be four times lower compared to the cost-effectiveness threshold commonly used in the UK's National Health Service. Furthermore, it was speculated that if the number of relatives tested per proband rose to three, the cost per QALY would be reduced tenfold (Kerr et al., 2017). Cascade screening has also proven to be effective in Australia and Brazil, where each index case yielded a further two cases (Bell et al., 2015; Jannes et al., 2015).

## 1.3. Advances in improving atherosclerotic cardiovascular disease risk prediction

In addition to the wide applicability of prediction tools in routine clinical practice, a number of biomarkers have been proposed for improving ASCVD risk prediction. Incorporating non-traditional risk factors, clinically measured biomarkers for atherosclerosis, or imaging tools results may help to refine estimation for high-risk as well as truly low-risk individuals. Furthermore, these may help to tailor intervention for those initially classified as at intermediate risk or for whom risk remained uncertain (Khambhati et al., 2018).

The proposed non-traditional risk factors and biomarkers include levels of high-sensitivity C-reactive protein, haemoglobin A1c, lipoprotein(a), lipoprotein-associated phospholipase and instrumental measurements of ankle-branchial index, coronary artery calcium (CAC) score and carotid intima-media thickness. While studies have shown their potential additive predictive value for ASCVD, the incremental value beyond known risk factors for the most part is relatively modest (Folsom, 2013). Furthermore, the genetic associations for the majority of these remain to be determined.

Lastly, the plethora of GWA studies are paving the way for calculating genetic risk scores that are now yielding power for clinical utility.

### 1.3.1. Lipoprotein(a)

Lipoprotein(a) (Lp(a)) has received wide clinical, molecular and genetic scrutiny in the last years, and has emerged as a promising therapeutic target and valuable biomarker that could putatively improve ASCVD risk estimation for those

targeted at intermediate risk category. Lp(a) is secreted by the liver and resembles an LDL particle, but has additionally a very large glycoprotein, apolipoprotein(a) (Apo(a)), covalently linked to the apoB-100 moiety of the LDL particle via a disulfide bond (Tsimikas and Hall, 2012; Berglund and Ramakrishnan, 2004; Utermann, 1989). Apo(a) is encoded by the *LPA* gene that consists of ten different types of kringle IV domains (KIV-1 to KIV-10), one kringle V domain and one protease-like domain. The KIV-1 and KIV-3 to KIV-10 domains are each present in a single copy, whereas KIV-2 shows an extensive copy number variation (CNV), repeating between 5 to >40 times per chromosome. Each KIV-2 copy is 5.5 kb in size and consists of two exons of 160 bp and 182 bp in length, linked by a large intron (4 kb) and a short intron (1.2 kb) that joins it to the next copy (Kraft et al., 1992; Noureen et al., 2015) (Fig. 4).

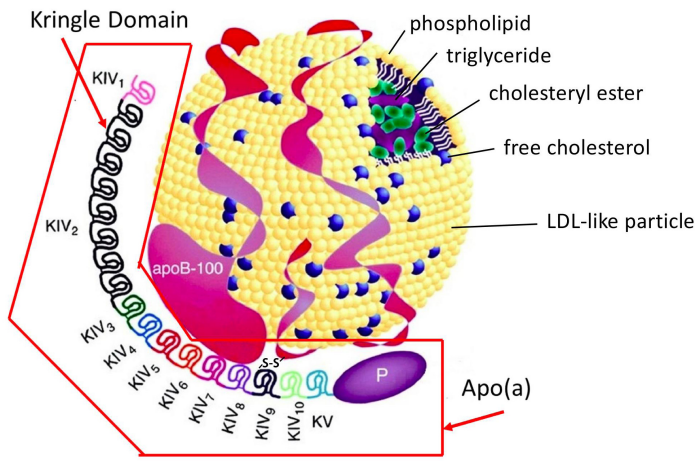



**Figure 4.** Structure of lipoprotein(a). Lipoprotein(a) is an LDL-like particle that has an ApoA molecule (denoted in red) covalently linked to the apoB-100 moiety of the lipoprotein particle. ApoA consists of a protease domain, one kringle V domain and ten different types of kringle IV domains, of which kringle IV type 2 is repetitive. Adapted from Albers et al., 2007.

The plasma concentrations of Lp(a) are determined mainly by the rate of hepatic synthesis of apo(a) (Koschinsky and Marcovina, 2004). Most individuals are heterozygous for two different apo(a) isoforms. As the hepatic secretion is lower for larger isoforms, the smallest one typically prevails in plasma (Nordestgaard et al., 2010). Human Lp(a) concentrations range from <0.1 mg/dl to more than 200 mg/dl, thus exhibiting up to three orders of magnitude difference among individuals. On average, Africans have two- to three-fold higher Lp(a) plasma concentrations than Europeans and most Asian populations

(Schmidt et al., 2016). Lp(a) levels have been shown to be genetically determined by the number of KIV-2 repeats, which correlate with apo(a) size and inversely with Lp(a) concentration (Lanktree et al., 2010). The effects from diet or lifestyle seem to be rather weak (Kettunen et al., 2016). While the concentrations of Lp(a) are generally resistant to statins (Leebmann et al., 2013), PCSK9 inhibitor evolocumab has been shown to reduce circulating Lp(a) by approximately 25–30% (Raal et al., 2014).

The main method for measuring Lp(a) in clinical practice is an immunoassay-based method that is sensitive to the entire mass of the Lp(a) particle, including apo(a) isoform and lipid content (McConnell et al., 2014). As of now there are no commercially available assays that are completely insensitive to the variability in particle mass. The lipid parameter can also be quantified as Lp(a)-cholesterol (Lp(a)-C) by the Vertical Auto Profile (VAP) method. This method is based on density gradient ultracentrifugation, whereby the cholesterol concentrations of lipoprotein classes are measured simultaneously. After vertical rotor-based centrifugation, the content of a centrifuge tube, where the lipoproteins have been separated based on their density across the horizontal axis, is analysed for cholesterol proportion via the spectrophotometric detection of an enzymatic reaction specific for cholesterol (Kulkarni, 2006; Kulkarni et al., 1994).

#### 1.3.1.1. Genetic underpinnings of lipoprotein(a) and association with atherosclerotic cardiovascular disease

Twin studies have estimated the heritability of Lp(a) to be around 90% (Mooser et al., 1997; Austin et al., 1992; Lamon-Fava et al., 1991), whereas the GWAS explain approximately half of that (Mack et al., 2017). KIV-2 repeats and two common SNVs in the *LPA* gene have been shown to be the main determinants of Lp(a) levels (Lanktree et al., 2010). These 2 SNVs that are inversely correlated with KIV-2 CNV (Clarke et al., 2009), are rs3798220, located in the protease-like domain of *LPA*, and rs10455872, residing in the long intron of the KIV-7 domain, and together explain about 36% of the variation in Lp(a) levels in Europeans (Saleheen et al., 2017; Lanktree et al., 2010; Kamstrup et al., 2009; Clarke et al., 2009). Furthermore, these two variants are associated with up to two-fold increased risk for coronary heart disease (Clarke et al., 2009). A meta-analysis on five GWA studies (n=13,781) identified 31 SNVs (30 in *LPA*, 1 in the *APOE* gene region), after adjusting for apo(a) isoform size, to be significantly associated with Lp(a) concentrations (Mack et al., 2017). A splice site variant, rs41272114, linked to decreased Lp(a) levels seems to, however, confer protection from ASCVD (odds ratio (OR) 0.84) (Lim et al., 2014).

Lp(a) has been established as an independent risk factor for ASCVD, including myocardial infarction, cerebral infarction, thrombosis and calcific aortic valve stenosis, irrespective of co-existing lipid concentrations, including LDL-C (Ergou et al., 2016; Alonso et al., 2014; Thanassoulis et al., 2013; Kamstrup et

al., 2009). A mendelian randomization (MR) study on approximately 40,000 individuals demonstrated that genetically elevated Lp(a) by KIV-2 CNV was causally linked with increased risk of myocardial infarction (Kamstrup et al., 2009). Moreover, elevated Lp(a) levels have been associated with sub-clinical disease (higher CAC scores) in individuals with family history of premature ASCVD (OR 1.79) (Verweij et al., 2018). A study based on seven randomized, placebo-controlled, statin outcome trials of 29,069 individuals showed that despite receiving statin treatment, the risk for an ASCVD event persisted in those with elevated Lp(a). Furthermore, the Lp(a) concentrations were associated more strongly with residual ASCVD risk in those assigned statins than in individuals allocated placebo (Willeit al., 2018). Moreover, an independent GWA study on approximately 10,000 individuals on statin therapy, of which 3,099 where CHD cases, highlighted the *LPA* locus as the strongest signal associated with adverse events (Wei et al., 2018). These results emphasize the requisite for Lp(a)-targeted therapeutics for overall and residual ASCVD risk reduction (Willeit al., 2018).

In light of the above, it has been hypothesized that Lp(a) measurement could provide value for prognostic risk prediction. Importantly, it has been shown that adding Lp(a) to established prediction tools reclassified up to 40% of the individuals initially stratified as intermediate risk group into either higher or lower risk categories (Verbeek et al., 2017; Willeit et al., 2014). Furthermore, measurement of Lp(a) has now been added to the ESC/EAS guideline for individuals at moderate risk or for those with family history of premature ASCVD (Piepoli et al., 2016).

A number of mechanisms have been proposed explaining the independent association between Lp(a) and ASCVD. First, Lp(a) has been shown to be the preferential carrier of proinflammatory oxidized phospholipids (oxPL) on apoB-100 particles in human plasma (Bergmark et al., 2008). The entrapment of oxPL-containing particles into the arterial wall could accelerate atherosclerosis by inducing an inflammatory response via activating macrophages, and contribute to foam cell formation and smooth muscle cell proliferation (van der Valk et al., 2016; Podrez et al., 2002; Zhao and Xu, 2000). Moreover, Lp(a) particles could be more avidly retained in the arterial wall by binding to the extracellular matrix via the apoB as well as the apo(a) component (Nielsen, 1999). Secondly, apo(a) can impair fibrinolysis due to its structural similarity to fibrinolytic proenzyme plasminogen. It is suggested that Lp(a) can induce thrombosis by competing with plasminogen for binding sites on fibrin, thereby reducing the conversion of plasminogen to plasmin and impairing clot degradation, and through inactivation of tissue factor pathway inhibitor that mediates the coagulation cascade (Boffa et al., 2004; Deb and Caplice, 2006; Feric et al., 2008). Elucidation of the molecular mechanisms of the Lp(a) particle contributing to atherosclerosis and the potential drug targets that could reduce the levels of Lp(a) are under intense scrutiny, and will undoubtedly enhance the value of Lp(a) measure in the clinical setting.

### 1.3.2. Polygenic risk scores

One of the drawbacks of phenotypic risk scores applied in the clinical setting is that they do not provide sufficient discrimination at a younger age when the implementation of preventative measures would probably result in the greatest long-term benefit. Furthermore, presence of diabetes, smoking history, abnormal lipid concentrations or blood pressure levels reflect already altered molecular trajectories and perturbations that could be curbed beforehand. One of the propositions to circumvent this limitation is to use genomic information obtained from GWA studies by calculating polygenic risk scores (PRS). As genomic information does not change over the course of a lifetime, one could estimate the genetic determinants already at birth, thereby providing value for risk prediction far before the clinical risk scores attain predictive capacity at later ages. Principally, a PRS is calculated per individual by aggregating the estimates obtained for a trait of interest in a GWA analysis into a single number. This number is calculated as a weighted sum of the number of disease risk variant alleles ($X$) in each individual, where the risk allele weights ($\beta$; variant effect sizes log(OR)) are retrieved from the initial GWA analysis:

$$PRS_x = \sum_{i=1}^{n} X_i \beta_i,$$

where $n$ is the number of SNVs included in the score. The PRS values tallied for each individual are then stratified within a distribution obtained for a sample set into distinct risk categories based on percentile rank cut-off values, such that each individual will probabilistically be assigned into a specific risk category (Torkamani et al., 2018; Khera and Kathiresan, 2017; Chatterjee et al., 2016).

The first CHD PRS was composed in 2010 by Ripatti et al. based on 13 independent ASCVD-associated SNVs. While 20% of the individuals were ascertained to be at a 70% increased risk for a CHD event, the PRS did not provide sufficient value for clinical utility (Ripatti et al., 2010). A subsequent study, where a 46 SNV PRS was calculated using the results from a largest to date ASCVD GWAS, yielded similar results, displaying marginal improvement of risk prediction beyond traditional risk factors (Ganna et al., 2013). Tada et al. demonstrated that a 50 SNV PRS improved CHD risk prediction beyond traditional risk factors and family history, and that the individuals in the top quintile were at a two-fold increased risk for CHD compared to those in the bottom quintile (Tada et al., 2016). Importantly, a UK biobank-based study revealed using a 182 SNV PRS that premature ASCVD (≤40 years of age for men, ≤45 years of age for women) was also polygenically driven. Although the number of cases analysed was small, individuals with premature ASCVD but without a rare deleterious dyslipidaemia-associated genetic variant belonged to the top PRS quartile associated with a two-fold increase risk for premature ASCVD compared to the general population (Thériault et al., 2018).

Mega et al. moved beyond studying the PRS on a representative population-based cohort and estimated the predictive value of a PRS composed of 27 SNVs in the randomized, placebo-controlled studies of statin therapy. They identified that individuals in the top quintile of the risk score distribution, being at a 70% increased risk for a CHD event, appeared to benefit considerably more from statin therapy compared to those in the bottom quintile (Mega et al., 2015). It was speculated that those with the highest genomic burden had more plaques that statins could stabilize (Thanassoulis et al., 2012). This finding and hypothesis was confirmed by Natarajan et al. using a 57 SNV PRS showing that those at the top quintile of risk score distribution had a greater burden of subclinical atherosclerosis and derived greater benefit from statin treatment, compared to all other risk subgroups, despite similar levels of LDL-C lowering by statin therapy (Natarajan et al., 2017). These studies clearly show the benefit of statin treatment for CHD prevention as well as demonstrate that pinpointing genetic markers that are specifically linked to sub-clinical atherosclerosis could further help in refining a risk group that would benefit from statin therapy initiation the most.

Substantial improvement in estimating the predictive value of a PRS on ASCVD risk arose from combining tens of thousands to millions SNVs into a PRS. This was based on a premise that the consideration of all ASCVD-associated risk alleles, regardless of their significance level in a GWAS, could capture a greater proportion of the genomic burden and hence multiple mechanistic pathways that lead to the disease. Three studies using the latest ASCVD GWA analyses (Nikpay et al., 2015; CardioGRAMplusC4D Consortium, 2012) and different statistical approaches for risk score calculation, resulted in the composition of 49,310, 1.7 million and 6.6 million SNV PRS (Abraham et al., 2016; Inouye et al., 2018; Khera et al., 2018). Abraham et al. demonstrated that the 49,310 SNV PRS was associated with incident CHD events (hazard ratio (HR) of 1.5 per 1 standard deviation (SD) increase) independently of established phenotypic risk scores and individual CHD risk factors, including family history (Abraham et al., 2016). Inouye et al. evidenced that a meta-score of 1.7 million SNVs predicted two-fold increased CHD risk for individuals already on therapeutic interventions when comparing top and bottom quintiles, addressing the requirement for continuous targeting of residual disease risk (Inouye et al., 2018). Khera et al. showed that a PRS aggregating 6.6 million common variants in the UK Biobank-based sample set of 290,000 individuals can identify twenty-fold more individuals at comparable three-fold increased risk identified in FH-associated variant carriers (Khera et al., 2018). These estimates provide a solid support for determining genomic risk at an early age that can be followed up by assessing established risk factors further in life. Assurance for such a strategy stems from the analysis on 55,685 individuals demonstrating that among individuals with a high genomic burden of ASCVD-associated variants calculated on 50 SNVs, those who led a healthy lifestyle had a 50% lower 10-year risk of developing a coronary event compared to those with unfavourable lifestyle habits (Khera et al., 2016). Furthermore, incorpo-

rating PRS into clinical practice could benefit initiation of and/or adherence to lifestyle modifications and preventative therapies. For instance, disclosure of CHD risk estimates incorporating also genetic info on 28 SNVs led to significantly lower LDL-C levels at six months follow-up than disclosure of risk based on conventional risk factors alone. However, no differences in dietary fat intake nor physical activity were noted (Kullo et al., 2016).

Altogether, it is fair to say that we are at the dawn of integrating PRS into clinical practice and of conveying clinically meaningful risk estimation to those whose genomes are enriched in risk alleles. While the prediction tools incorporating PRS can facilitate identifying subgroups of individuals who would benefit from the prioritization of preventive actions (Torkamani et al., 2018), future efforts will help refining these groups more specifically and accurately. With knowledge on risk factor-specific and gene-environment interactions (Wijmenga and Zhernakova, 2018), and delineation of sub-risk scores based on molecular pathways, it is highly likely that therapeutic-specific clinical protocols will eventually start to herald clinically applied precision medicine.

## 1.4. Next steps for refining atherosclerotic cardiovascular disease risk prediction

Assessment of the power of utilizing polygenic risk scores, clinical validation of novel biomarkers and refinement of risk prediction tools have and will contribute to the efforts of reducing the global ASCVD burden by targeting primary prevention. However, with the use of large-scale cohorts, sophisticated bioinformatics methodologies and data management approaches, and utilization of WGS and WES data, a paradigm shift in terms of tackling the complexity of ASCVD is transpiring. This will usher the change of focus from cohort-specific research endeavours to state-of-the-art advancements, whereby population-based research will essentially guide the integration of refined biological underpinnings and genetic associations into clinical practice.

### 1.4.1. Harnessing biobank information for personalizing atherosclerotic cardiovascular disease risk prediction

Biobank-contained data that hold genetic, molecular and phenotypic information on a large number of individuals is now considered as a key resource that has the potential of laying the groundwork for precision medicine. Major efforts in founding, maintaining and upgrading large-scale and population-based biobanks are being undertaken, e.g. the Estonian Biobank (Leitsalu et al., 2015), UK Biobank (Bycroft et al., 2018), Geisinger MyCode Community Health (Carey et al., 2016), Lifelines (Scholtens et al., 2015), deCODE Genetics (Gudbjartsson et al., 2015), Electronic Medical Records and Genomics Network (Gottesman et al., 2013), Kaiser Permanente Research Program on Genes,

Environment and Health (Kaiser, 2018), and the Million Veteran Program (Gaziano et al., 2016). These resources integrate high-quality genomic data with electronic health repositories (EHR) with the central aim of tackling genotype-phenotype associations on a larger scale (Dewey et al., 2016). One of the crucial advantages that these initiatives hold is the built of population-specific haplotype reference panels from whole-genome sequencing a set of a cohort, that enables the imputation of rare and low-frequency variants not captured on genome-wide arrays (Mitt et al., 2017; Huang et al., 2015; Deelen et al., 2014) and not identified on common haplotype reference panels (1000 Genomes project (Auton et al., 2015).

The UK Biobank is one of the largest projects that has demonstrated the successful collection, and sharing of genetic and clinical information of 500,000 individuals on a large scale. Furthermore, they now venture to increase its content by sequencing the genomes of one million individuals. The initiative holds the objective to make its complete datasets as well as results from conducted studies fully available (Bycroft et al., 2018). Contrary to capturing the resources on multi-ethnic individuals, small countries, such as Estonia and the Netherlands, contain information on genetically more homogeneous individuals within a more uniform national infrastructure, allowing simpler follow-up on individuals and more consistent retrieval of information from established EHR (Wijmenga and Zhernakova, 2018). Furthermore, an individual-level advantage that stems from these large-scale resources is the return of clinically actionable genomic findings to research participants, bridging the transition of research information to clinical care. Geisinger MyCode Community Health Initiative has started to undertake such a genotype-first approach based on the pathogenic and likely pathogenic variants from a set list of genes associated with monogenic conditions. They have thus far disclosed the results to 542 participants and outlined a detailed model for clinical reporting of genomic findings (Schwartz et al., 2018).

#### 1.4.1.1 Value of electronic health records

Linking EHR to biobank-contained resources allows to address a wide variety of research questions on the full spectrum of health- and disease-related traits, enables fine-grained longitudinal analyses for health over time, and offers resolution of clusters of diseases and endpoints of interest (Hemingway et al., 2018). Valuable phenotypic and clinical data can be retrieved from national health insurance and regional hospital databases, causes of death and disease-specific registries. The most prominent use of EHR data is the utilization of structured and unstructured data from hospital resources that are usually generated during routine clinical care (such as physiological measurements, demographic information, clinical blood analyses, imaging results, clinically used device data etc.). Structured EHR data are recordings based on fixed clinical terminology (Systematized Nomenclature of Medicine Clinical Terms

(Shahpori and Doig, 2010)), medical classification for diseases, signs, symptoms, abnormal findings, complaints and external causes of injury or diseases (International Statistical Classification of Diseases and Related Health Problems (ICD, 2016)), surgical procedures (Nordic Medico-Statistical Committee Classification of Surgical Procedures (NOMESCO, 2011)) and classification of active ingredients of drugs (Anatomical Therapeutic Chemical Classification System (ATC, 2018)). Patient's medical history, discharge summaries and imaging reports are usually recorded as unstructured text (Hemingway et al., 2018).

Besides enabling large-scale genome-wide scans in a hypothesis-free manner for traits of interest, coupling of EHR with genomic data facilitates also fine-graining of different phenotypes. For instance, one can ascertain carriers of rare deleterious protein truncating variants and detail the phenotype or sub-phenotypes, and subsequently recall the carriers for additional testing in a clinical setting. Second, gene-based testing by aggregating intra-genic rare variants into a genomic unit allows to conduct phenome-wide association studies for interrogating novel phenotypes or unidentified biological mechanisms for genes of interest. Third, with the use of large-scale population data, true population prevalence of a condition can be estimated. And fourth, speedy recruitment of individuals harbouring deleterious variants for clinical trials of new therapeutics could directly benefit precision medicine (Denny et al., 2013; Rader and Damrauer, 2016) (Fig. 5).

One of the examples of implementing such a strategy is the DiscovEHR collaboration, where WES data of 50,726 adult participants of European ancestry were coupled to longitudinal EHR of up to 14 years. Each participant was found to harbour on average 21 rare predicted loss of function (LoF) SNVs, whereas linkage with EHR enabled to confirm the phenotypic consequences for these variants. Next, gene-based testing on lipid traits identified novel rare variants, for instance in the *G6PC* gene that encodes glucose-6 phosphatase catalytic subunit and has been associated with glycogen storage disease type 1a (Chou et al., 2002). Association testing with triglyceride levels revealed that heterozygotes for protein-disrupting variants in the *G6PC* gene displayed moderately increased hypertriglyceridemia. Furthermore, interrogation of 76 clinically actionable disease-associated genes linked with 27 medical conditions unveiled that 3.5% of the study participants harbour a pathogenic or likely pathogenic variant that can be acted upon via therapeutic interventions (Dewey et al., 2016). A targeted scan for FH in the same sample set revealed that clinical diagnosis of FH based on established criteria identifies only a quarter of FH-associated variant carriers (Abul-Husn et al., 2016). Lastly, in a comprehensive phenome-wide association analyses-based study leveraging EHR data and over 600,000 common variants, disease pairs were constructed on the basis of shared genetic components. In addition to confirming the clustering of autoimmune diseases (e.g. diabetes type I, rheumatoid arthritis, psoriasis and multiple sclerosis), genetic-based networking linked immune-mediated diseases also with hyperlipidaemia (Verma et al., 2019). This finding highlights the value of

phenome-wide analysis across a large-scale EHR-linked biobank for identifying novel associations.

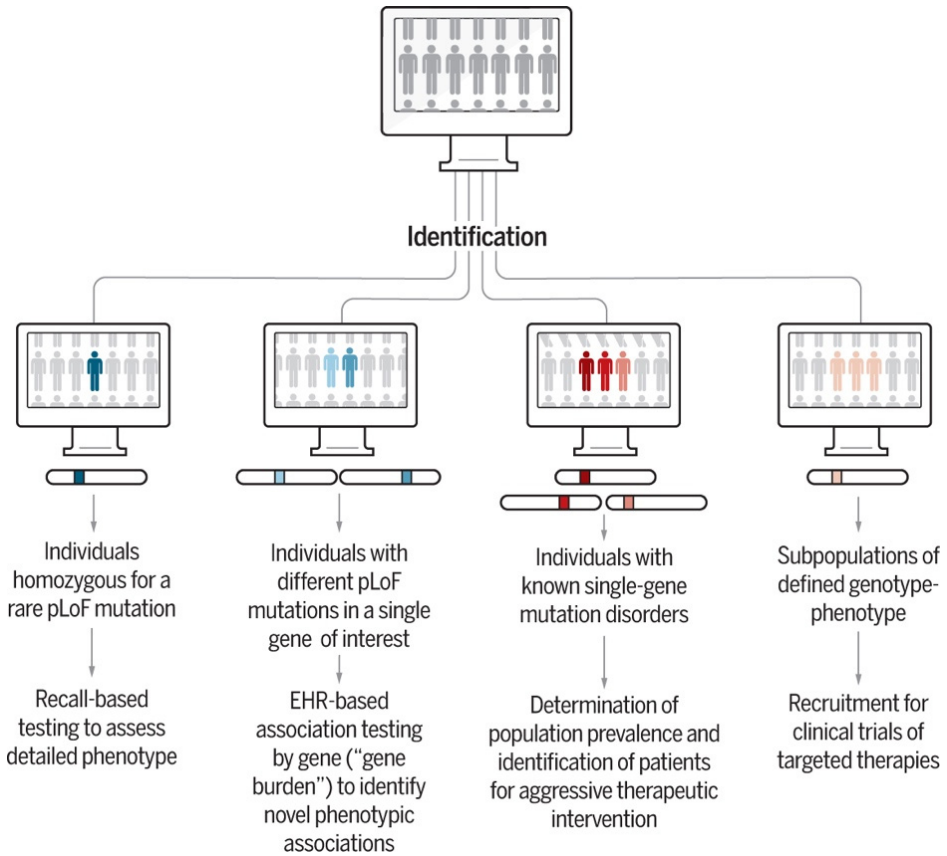



**Figure 5.** Research approaches of linking EHR information with genomic data (Rader and Damrauer, 2016). pLoF - putative loss-of-function.

The successful implementation of EHR data for genetic analysis, however, faces a number of caveats. As EHR are obtained as a by-product of routine clinical care aimed mainly for clinical use, these contain missing values, incorrect entries and vast quantities of unstructured free text that contain valuable information on pathological findings. Furthermore, absence of data points in EHR does not specifically translate into absence of phenotype, but could indicate unidentified disease diagnosis. Text mining, natural language processing methods and more sophisticated extraction and curation of information are being developed to counteract these errors. Altogether, incorporation of EHR data in genomic-based research holds the potential for precision medicine on a population scale (Denny, 2012; He et al., 2017).

### 1.4.2. Multi-omics approach

Combination of thousands to millions genomic markers into a PRS to maximise the converging effect of multiple pathways driving atherosclerosis progression holds great potential for timely identification of individuals at high ASCVD risk. However, for deciphering the complex interplay of the disease triggers in detail, and delineating the molecular cascades that lead from a potential to actual disease manifestation, more elaborated approaches are required. While the genomic information constitutes a powerful anchor point for mechanistic studies of the pathophysiology of the disease, the single layer of information does not capture the full spectrum of molecular effects that jointly and in intricate combination manifest as an outcome. Based on current knowledge, it is fair to conclude that non-linear and complex interactions and molecular cascades occur in atherosclerosis involving simultaneous pathological changes in diverse cell types, tissues and organs at multiple molecular levels. The technological advances that have broadened the availability, cost-effectiveness and quality of genome, epigenome, transcriptome, proteome, metabolome and microbiome data are now paving way for amalgamating multiple layers of information across different omics domains. Multi-dimensional data integration can empower the search for mechanistic drivers underlying ASCVD, help in delineating its subtypes, understand the phenotypic variations on different molecular levels and identify biomarkers with diagnostic and prognostic value.

One of the examples of such a strategy was demonstrated by Björkegren et al. who analysed the genomic information and transcriptome data from seven disease-relevant vascular and metabolic tissues collected from 600 ASCVD cases during coronary artery bypass surgery in the STARNET study. By performing a number of integrative network analyses, key drivers causal for regulatory gene networks of CHD interconnected in vascular and metabolic tissues were distinguished (Talukdar et al., 2016). Furthermore, regulatory SNVs identified in GWA studies for cardiometabolic diseases were more extensively found tissue-specifically or across tissues of CHD-affected individuals than in tissue- and disease-unspecific expression studies (Franzén et al., 2016).

Despite some notable efforts, the multi-omics field is still in its infancy. The main limitations stem from the high complexity within and between individual datasets, technical and computational shortcomings in modelling to accurately capture true signals, and lack of appropriate data in large enough cohorts from multiple time points and different tissues. However, with the joint collaborations between national and international biobanks and clinicians, these data will undoubtedly start to emerge in the coming years (Lau and Wu, 2018; Vilne and Schunkert, 2018; Arneson et al., 2017; Hasin et al., 2017).

# 2. AIMS OF THE STUDY

The general aim of the doctoral thesis was twofold. First, to study the predictive ability of phenotype-based ASCVD risk estimation. And secondly, to interrogate the value of genomics for improving ASCVD risk prediction.

The objectives of the thesis were as follows:

1) Determine and compare the performance accuracy of three commonly used and clinically applied ASCVD risk scores in the Estonian population.
2) Ascertain the value of high-coverage whole-genome sequencing data for broadening the knowledge of clinically measured and clinically meaningful ASCVD biomarkers, namely lipoprotein(a) and LDL-cholesterol.
3) Evaluate the implications of genomics-guided identification of individuals at high ASCVD risk (genetically predisposed to familial hypercholesterolemia) for clinical management.

# 3. RESULTS AND DISCUSSION

## 3.1. Performance of phenotypic risk scores in a high ASCVD risk setting (Ref. I)

Today, the identification of individuals at increased ASCVD risk relies mainly on clinically applied phenotypic risk scores that facilitate the estimation of one's absolute 10-year risk for developing an ASCVD event based on established risk factors. Those identified at the highest risk are targeted for therapeutic interventions to prevent the onset of an adverse event. There are three main commonly used risk scores: Pooled Cohort Estimation endorsed by the ACC/AHA in the US (Goff et al., 2014), SCORE by the ESC/EAS in Europe (Piepoli et al., 2016) and QRISK2 by the NICE guidelines in the UK (NICE, 2018). Despite their uniform aim, these vary in terms of predicted outcomes, algorithms used for risk calculation and risk factors considered, and furthermore, have been shown to overestimate the actual risk for developing an adverse ASCVD event, thereby predisposing a proportion of individuals for unnecessary treatment (DeFilippis et al., 2017; Defilippis et al., 2015; Demir et al., 2015; Kavousi et al., 2014; Mortensen and Falk, 2014; Van Staa et al., 2014). This inaccuracy is mainly attributed to the difference in incidence rates between the risk score derivation cohorts (usually sample sets from developed Western populations comprised decades ago) and the risk score target cohort. As Estonia is considered a high-risk country in Europe in terms of ASCVD incidence rates, the first objective was to assess the predictive ability of the three mainly used phenotypic risk scores and compare their performance on targeting individuals at high ASCVD risk for therapeutic interventions.

### 3.1.1. Description of cohort and materials

The performance accuracy of three commonly applied risk scores, PCE, QRISK2 and SCORE, was evaluated on the Estonian Biobank cohort of 51,603 individuals (Leitsalu et al., 2015). Follow-up data on ASCVD incidence and mortality were obtained based on linked information to the Estonian Health Insurance Fund database (EHIF, 2018) and the Estonian Causes of Death Registry (Causes of Death Registry, 2018). As the median follow-up time was 7.8 years and more than 70% of the individuals had seven-year follow-up data available, the score-specific events were censored at 7 years and the original scores of 10-year risk were modified to obtain 7-year risk estimates. Based on the guideline-specific criteria, 4,356, 7,191 and 3,987 individuals were eligible for calculating the risk estimates of PCE, QRISK2 and SCORE, respectively (Fig. 1 in Ref. I). The mean age of the study sample was 40.8 years and 64.1% were women (S.Table 2 in Ref. I). Cox proportional hazard models were fitted for each outcome with the 7-year risk score estimate as the only covariate. Calibration was assessed by standardized incidence ratios (SIR, expected

number of events divided by observed number of events) and discrimination by Harrell's C-statistic. To illustrate the implications of risk prediction, three categories of recommendations for statin treatment were comprised based on the ACC/AHA (Goff et al., 2014), NICE (NICE, 2018) and ESC (Piepoli et al., 2016) guidelines, and applied on a uniform subset of individuals (n=3,729, between the age range of 40–70 years and without diabetes mellitus type I or II or chronic kidney disease).

### 3.1.2. Performance accuracy of phenotypic risk scores

During the seven-year follow-up, 220 hard ASCVD events (PCE outcome), 671 ASCVD events (QRISK2 outcome) and 94 ASCVD events (SCORE outcome) occurred in the score-specific subsets (Table 2 in Ref. I). Based on SIRs calculated for evaluating calibration, PCE (SIR 1.03, 95% confidence interval (CI) 0.90 – 1.18) and SCORE (SIR 0.99, 95% CI 0.81 – 1.21) performed accurately, predicting 3% more hard ASCVD events and 1% less ASCVD deaths than observed, respectively. QRISK2, however, severely underestimated the risk (SIR 0.52, 95% CI 0.48 – 0.56), predicting approximately 50% less events than observed (Fig. 6; Table 2 in Ref. I). The latter result could partly stem from the inaccuracy of considering soft endpoints (e.g. angina pectoris), that lack strict diagnostic criteria opposite to hard endpoints (e.g. myocardial infarction or ASCVD death). In regards to discrimination, PCE (C-statistic 0.778) was inferior to QRISK2 (C-statistic 0.812) and SCORE (C-statistic 0.865) (Table 1 in Ref. I), demonstrating that the latter two risk scores performed more accurately in terms of predicting higher risk for those who developed an endpoint than those who did not.

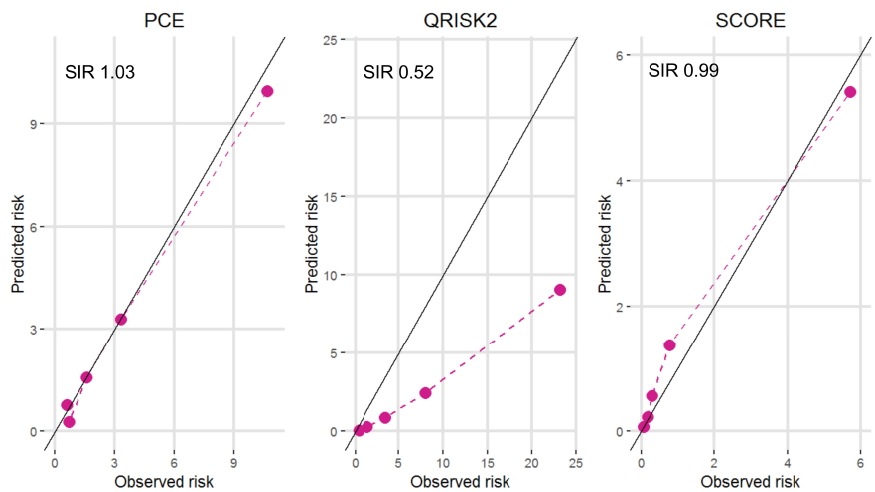



**Figure 6.** Calibration plots with SIRs demonstrating the performance of risk scores in different risk quintiles over the 7-year follow-up.

Notably, the risk estimates for individuals with and without events greatly overlapped, such that among those who experienced an adverse event, approximately a quarter of men and third of women were categorized into low or moderate risk category (generally considered cut-off of 5%) (S.Fig. 1 in Ref. I). While statin treatment recommendations were relatively similar between the guidelines, varying from 57% (ACC/AHA) to 36% (SCORE) for men and 27% (ACC/AHA) to 10% (SCORE) for women, target groups for whom statins could be considered varied greatly. While 14% of men and 7% of women could be considered for statin treatment based on the ACC/AHA guidelines, 49% of men and 35% of women were targeted for treatment consideration based on the ESC guidelines (Fig. 7; Fig. 3 in Ref. I).

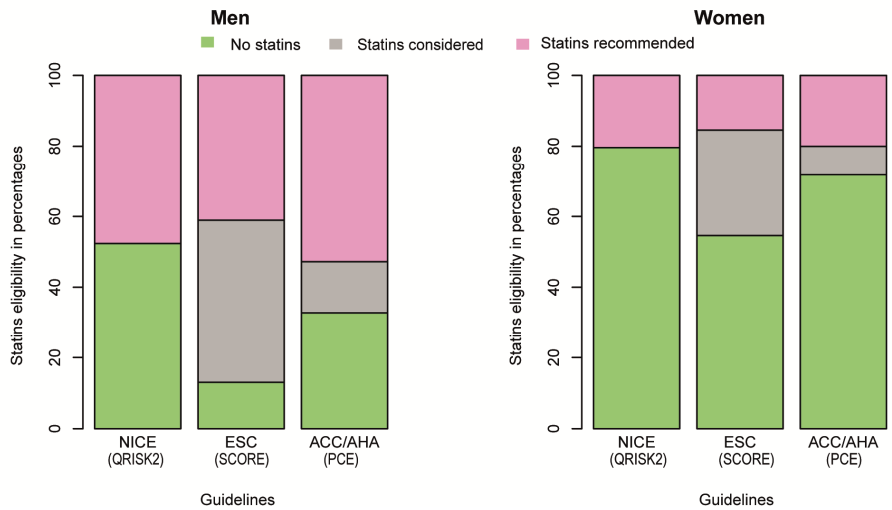



**Figure 7.** Statin treatment recommendations according to the ASCVD prevention guidelines on a uniform subset of individuals (n=3,729, between the age range of 40–70 years, and without diabetes mellitus type I or II or chronic kidney disease).

These observations permit several important conclusions. First, phenotypic risk scores do perform sufficiently well in a high-risk population, however, up to a third of individuals who will develop an adverse event are misclassified as moderate or low-risk. Secondly, while individuals were targeted similarly for therapeutic intervention across the guidelines, statins could be considered additionally for 49% of the individuals by the ESC guidelines. This marks a great percentage of the population for whom risk remains ambiguous. Thirdly, consideration of soft outcomes for and incorporation of a wide variety of risk factors into risk prediction does not translate into more efficacious predictive ability of a risk score, as seen for the QRISK2 estimates. Altogether, these results highlight the need for further improvement of ASCVD risk prediction.

This holds especially true to a high-risk population setting, where improved strategies and policies for decreasing ASCVD incidence rates are anticipated.

## 3.2. Value of high-coverage whole-genome sequencing for dissecting the genetic underpinnings of ASCVD-associated lipid parameters (Ref. II, III)

An important aspect that needs to be borne in mind is that the discrimination of ASCVD risk according to phenotypic risk scores is based on already actualized perturbations (abnormal lipid parameter or blood pressure levels, or presence of concomitant diseases). In order to improve the ascertainment of individuals at high ASCVD risk before they actually develop overt events and moreover, before they develop changes in clinically measurable quantifiable biomarkers, further advances are needed. Knowing that ASCVD has its genetic basis, one can scrutinize the genetic foundation of the disease and its risk factors, and identify associations that can additionally carry prognostic value. While extensive research has been conducted in large-scale cohorts based on imputed genotyping arrays for ASCVD and ASCVD-related traits, these have generally resulted in the identification of putative associations without revealing true causality. High-coverage sequencing information, however, does offer a reliable platform for dissecting the genomic landscape of ASCVD-related traits with the potential to precisely pinpoint clinically meaningful genetic markers.

Lipoprotein(a), Lp(a), serves as a perfect example for this kind of endeavour due to its high heritability noted in twin studies (Austin et al., 1992; Lamon-Fava et al., 1991) and solid associations with ASCVD incidence (Ergou et al., 2016; Kamstrup et al., 2009). Lp(a) is an LDL particle-like plasma lipoprotein that has a large glycoprotein apo(a) attached to it. The latter is encoded by the *LPA* gene, containing a large highly polymorphic CNV of the kringle IV type 2 domain (KIV2-CN). The CNV and SNVs in the *LPA* locus have been associated with Lp(a) concentration in plasma and ASCVD risk (Mack et al., 2017; Lanktree et al., 2010; Kamstrup et al., 2009). Due to the complex nature of the gene and the strong established basis for genetic causality, WGS data for ancestrally distinct cohorts was exploited to investigate the full spectrum of the genomic variation affecting Lp(a) concentrations and furthermore, to interrogate the causality for ASCVD.

A comprehensively studied (Surakka et al., 2015; Willer et al., 2013; Teslovich et al., 2010; Kathiresan et al., 2008) strong predictor of ASCVD risk, namely LDL-C, was similarly studies based on WGS-derived genotypes in ancestrally distinct cohorts for locus discovery.

### 3.2.1. Description of cohort and materials

To dissect the genetic landscape of Lp(a), deep-coverage (30X) WGS data of 8,392 individuals (2,284 Estonians from the Estonian Biobank (Leitsalu et al., 2015), 2,690 Finns from the Finland FINRISK study (Vartiainen et al., 2000) and 3,418 African Americans from the Jackson Heart study from the NIH/ NHLBI Trans-Omics for Precision Medicine program (Taylor et al., 2005) were used (Fig. 1 in Ref. II). The sample genotypes were called separately for the African American cohort and jointly for the Estonian and the Finnish cohort according to Best Practices (Tan et al., 2015; van der Auwera et al., 2013). KIV2-CN in the *LPA* gene was estimated with Genome STRiP (Handsaker et al., 2011) and quantified for each individual as the sum of the KIV2 allelic copy number across both chromosomes.

While the Lp(a) parameter was not quantified for the Finnish sample set with WGS, an imputation model was developed based on their WGS and WES data for imputing the KIV2-CN into the array data of 27,344 individuals. This was done based on least absolute shrinkage and selection operator in a 4MB window around *LPA* and across high-quality (imputation quality >0.8) variants with MAF >0.1% available in the Finnish imputation dataset. This resulted in a 61-variant model, explaining 60% of the variation in genotyped KIV2-CN.

Lp(a) parameter measurements were available either as Lp(a), measured for Finns and African Americans with an immunoassay-based method, or as Lp(a)-C, quantified for Estonians and African Americans with the VAP method. Association analyses were carried out with both or either measurement where applicable. Both Lp(a) measurements were available for the African Americans, displaying moderate correlation (r=0.46) (S.Fig. 5 in Ref. II). Median Lp(a) levels in African Americans (median (interquartile range) 46 (24–79) mg/dL) were approximately ten times higher than in Finns (5 (2–10) mg/dL), while the Lp(a)-C distributions were similar between Estonians (7 (5–9) mg/dL) and African Americans (7 (5–11) mg/dL) (S.Table 3 in Ref. II). All study participants had provided written and informed consent in accordance with the respective institutional ethics review board for each of the participating study cohort (Fig. 1 in Ref. II).

For LDL-C analyses, jointly called genotypes of 30X WGS data according to Best Practices (Tan et al., 2015; van der Auwera et al., 2013) from the Framingham Heart Study (n=4,064), the Jackson Heart Study (n=3,247) and Old Order Amish (n=1,083), as well as from the Finland FINRISK Study (n=1,165) and the Estonian Biobank (n=2,255) were studied together with the individual WGS dataset of the Multi-Ethnic Study of Atherosclerosis (MESA; n=4,510) (Fig. 1 in Ref. III). Conventionally measured LDL-C values were used, with statin effect taken into account by dividing by 0.7 as implemented previously (Peloso et al., 2014). The mean (SD) age was 51 (15 years) and 8669 (53%) were women. Thirty-six per cent of the participants were of non-European ancestry.

### 3.2.2. Genetic interrogation of Lp(a)

The strongest genetic signal for Lp(a) has been shown to lie in the *LPA* gene locus and specifically as the CNV of the KIV2 repeat (Lanktree et al., 2010). In this analysis, the KIV2-CN solely explained 18% to 26% of Lp(a) variation and 14% of Lp(a)-C variation across ethnicities. The distribution of KIV2-CN differed slightly between African Americans (mean 38.5 (SD 7.4)) and Europeans (mean 43.7 (SD 6.2)), ranging between 12–85 copies, and were overall negatively associated with Lp(a)-C (-0.05 SD/CN, $p = <1\times10^{-61}$) and Lp(a) (-0.07 SD/CN, $p = <1\times10^{-190}$) (Fig. 3 in Ref. II). When centring only on the *LPA* locus (1MB window around the gene), 770 variants for either of the Lp(a) parameter reached the genome-wide significance level, of which five showed independent and ethnic-specific association for Lp(a)-C, and 43 for Lp(a) (S.Fig. 15a,b in Ref. II). On average, *LPA* locus genetic variants yielding a 1 SD increase in Lp(a) resulted in a 0.48 SD increase in Lp(a)-C. Furthermore, two known *LPA* LoF variants (rs41272114 ($p = 8\times10^{-77}$), rs143431368 ($p = 2\times10^{-26}$) (Kyriakou et al., 2014; Lim et al., 2014) and one novel LoF variant (rs199583644 in exon 28 only observed in African Americans (MAF 0.28%, $p = 3\times10^{-13}$) showed an association with the Lp(a) parameter.

When widening the search to common variants (MAF >0.1%) genome-wide, three loci for Lp(a)-C at *LPA* (rs140570886, $p = 3.3\times10^{-30}$), *CETP* (rs247616, $p = 6.1\times10^{-10}$) and *SORT1* (rs12740374, $p = 1.0\times10^{-21}$), and two for Lp(a) at *LPA* (rs6938647, $p = 4.7\times10^{-129}$) and *APOE* (rs7412, $p = 1.3\times10^{-23}$) were identified when conditioned on KIV2-CN. As the lead variants in the *SORT1* and *APOE* locus, and in the *CETP* locus have previously been associated with LDL-C and HDL-C respectively, conditional analyses on these lipid parameters were performed. Whereas the association for the *SORT1* and the *APOE* locus were not substantially altered, the association in *CETP* disappeared (Fig. 8; S.Fig. 9 in Ref. II and Fig. 4a,b in Ref. II). Gene-based burden testing on rare (MAF <1%) coding or non-coding variants did not pinpoint significant KIV2-CN-independent associations.

Next, to ascertain whether any genetic variants affect the relationship between KIV2-CN and Lp(a)-C or Lp(a) concentrations, variant-by-KIV2-CN interaction analyses at a 4MB window around *LPA* was performed. Three independent variants yielded genome-wide significance across ethnicities (rs13192132, $p = 1.73\times10^{-15}$, rs1810126, $p = 6.84\times10^{-14}$, rs1740445, $p = 6.35\times10^{-9}$). The variant rs13192132 tags the top hit in the Estonian-specific interaction analysis, a 3-base deletion ($r^2 = 0.88$), that resides 7,508 bases downstream of the *LPA* transcription start site and overlies H3K4me3 and H3K27ac peaks associated with active transcription based on the adult liver regulatory annotations from the Roadmap Epigenome Project (Kundaje et al., 2015) (S.Fig. 21A in Ref. II).

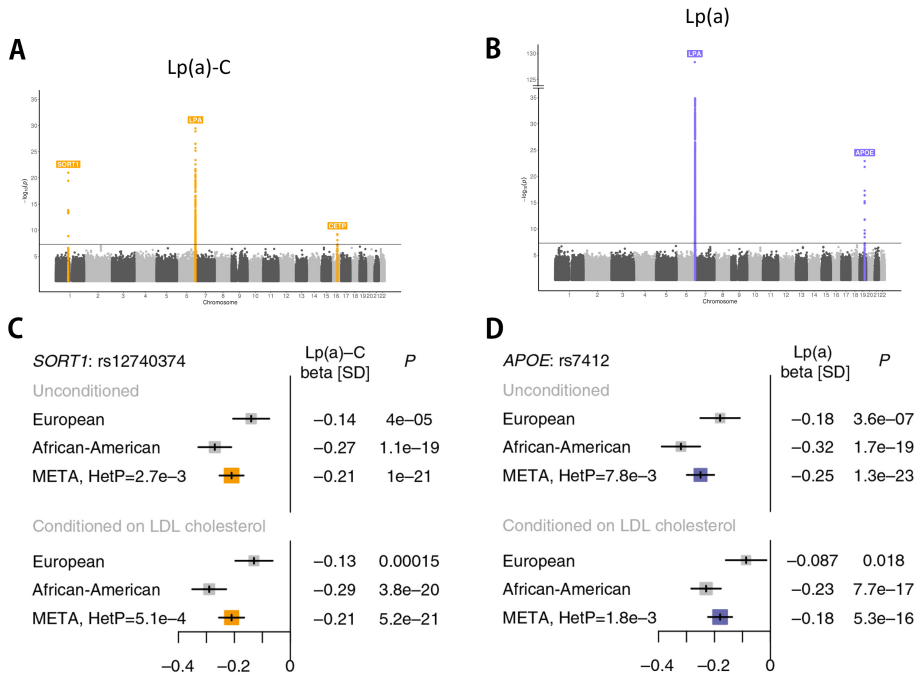



**Figure 8.** Manhattan plots of meta-analysed single variant associations of A) Lp(a)-C and B) Lp(a). Associations (betas in SD and 95% CI) for top variants at C) the *SORT1* and D) the *APOE* loci by ethnicity, and before and after conditioning on LDL-cholesterol.

### 3.2.3. Inter-ethnic differences in genetic determinants of Lp(a) and heritability estimates

While the latest array-based GWAS estimated the heritability of Lp(a) to be 49% (Mack et al., 2017), WGS-based estimation yielded 85% for Lp(a) and 52% for Lp(a)-C in African Americans, and 75% for Lp(a) in Finns and 75% for Lp(a)-C in Estonians (Fig. 4d in Ref. II).

To dissect any inter-ethnic genetic differences, effects of *LPA* locus variants attaining sub-threshold significance ($p = <1\times10^{-4}$) in either ethnicity for Lp(a) and Lp(a)-C were compared, yielding correlation for Lp(a)-C to be 0.38 and for Lp(a) to be 0.16 (S.Fig. 16 in Ref. II). Moderately associated ($p = <1\times10^{-2}$) *LPA* locus variants largely private in African Americans (MAF <0.1% in Finns) had larger absolute effects across MAFs compared to such variants observed in both ethnicities ($p = 3\times10^{-32}$). Furthermore, an *LPAL2* intronic variant at the *LPA* locus (rs192873801) yielded significant yet opposing inter-ethnic effect (+0.80 SD with MAF 2.8% in African Americans and -0.61 SD with MAF 2.7% in Finns), likely indicating influences from haplotype structure or gene-environment interactions (Fig. 9; Fig. 4d in Ref. II).

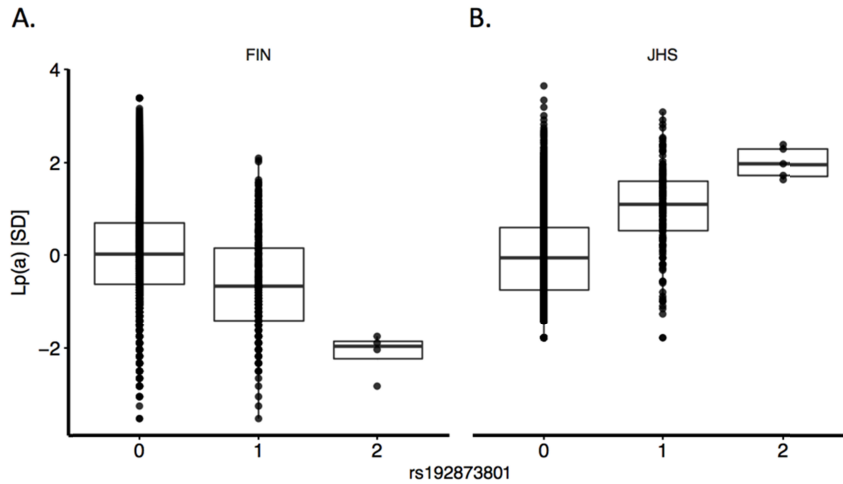



**Figure 9**. *LPAL2* intronic variant rs192873801 with significant heterogeneity across ethnicities. Shown are boxplots by cohort for A) Finns (FIN) and B) African Americans (JHS) showing the association of 0, 1 and 2 alternate alleles of rs192873801 with Lp(a) [SD].

### 3.2.4. Phenotypic consequences of the genetic determinants of *LPA* on ASCVD

To assess causality for ASCVD, MR analysis based on the Finnish imputed data of 1056 incident CHD cases and 21,207 controls were conducted. In a MR study, genetic variants associated with a modifiable exposure, risk factor, biomarker or biological intermediate (in this case Lp(a)) are used to estimate the causal relationship between the variable and a clinically relevant outcome. It is based on the premise that alleles of genetic determinants are randomly allocated during human gamete formation and are nonmodifiable, and therefore ensure lifelong exposure and mitigate concerns about reverse causation. Hence, if a variable has a causal association with a disease, the genetic determinants of the variable should also associate with the disease to the extent predicted by the size of the effect of the genetic variant on the intermediate risk factor and the size of the effect of the risk factor on the disease (Musunuru and Kathiresan, 2019; Emdin et al., 2017). In this study, three different genetic instruments were composed: 1) PRS, comprised of the sum of the KIV2-CN-adjusted variant effects from LD-pruned variants in a 4MB window around *LPA* with sub-threshold significance ($P = <1\times10^{-4}$; 339 variants); 2) KIV2-CN; 3) PRS + KIV2-CN combined. Each genetic instrument was normalized such that 1 unit increase in the score was equal to 1 SD increase in Lp(a). The PRS, KIV2-CN and PRS + KIV2-CN score explained 30%, 18% and 47% of Lp(a) variance, respectively. Interestingly, PRS had the largest effect on incident CHD risk (HR

1.36 per Lp(a) SD, p = $7.6\times10^{-8}$) compared to PRS + KIV2-CN (HR 1.25 per Lp(a) SD, P = $7\times10^{-7}$) or KIV2-CN alone (HR 1.13 per Lp(a) SD, p = 0.076) and compared to the effect of measured Lp(a) on incident CHD risk (HR 1.16, p = $3.71\times10^{-3}$) (Fig. 10; Fig. 7a in Ref. II). These results suggest that knowledge of *LPA* variant class genotypes may provide valuable information on cardiovascular risk beyond circulating Lp(a) levels.

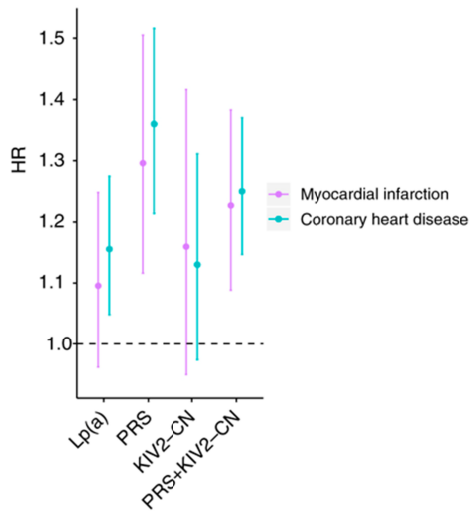



**Figure 10**. Associations (HR and 95% CI) of incident coronary heart disease and myocardial infarction with the Lp(a) measurement and with genetic instruments among the genotyped and imputed FIN individuals.

Altogether, the comprehensive interrogation on the full genomic spectrum of Lp(a) parameters highlights the value of analysing high-coverage WGS data in different ethnicities. WGS allowed to capture a great proportion of Lp(a) heritability, identify common and unique genetic determinants across and specific to different ethnicities, and ascertain the added value of estimating trait-associated SNVs for ASCVD causality. Importantly, knowledge of these genetic determinants provides valuable information for ASCVD risk prediction in addition to the biomarker level alone.

### 3.2.5 Dissection of the genetic basis of LDL-C variation across ancestrally distinct populations

The common variant search for LDL-C based on high-coverage sequencing data resulted in the ascertainment of seven previously identified (Surakka et al., 2015; Willer et al., 2013; Teslovich et al., 2010; Kathiresan et al., 2008) and

three novel loci, covering 697 variants at genome-wide significance. Importantly, conditional analyses in ancestrally distinct cohorts revealed a low-frequency haplotype specific to African Americans in the *LDLR* gene locus with variants in the first intron of *LDLR*, in the *LDLR* promoter and within the enhancer 4 kb upstream from the *LDLR* transcription start site in strong LD ($r^2 = 0.8$). This haplotype was associated with 28 mg/dl or 0.72 mmol/L lower LDL-C levels. While analyses of rare coding variants pinpointed genes that have previously been associated with monogenic dyslipidaemia disorders (*LDLR*, *APOB*, *PCSK9*, *APOE*), non-coding variant investigation did not detect any signals.

The incremental value of using WGS information for a genome-wide scan of a trait of interest does not yet facilitate the identification of novel putatively causal associations due to small sample sizes. However, it does allow to confidently unveil ethnically distinct genomic determinants, that are usually missed or identified with low confidence in genotyping array-based analyses that utilize a common haplotype reference panel for imputation. With the increasing availability of WGS data for different cohorts and built of sample set-specific haplotype reference panels for imputing whole-genome array datasets, WGS-based genome-wide scans will undoubtedly start to unravel distinct common and rare putatively causal variants of ASCVD-related phenotypes.

## 3.3. Genomics-guided approach for identifying individuals at high ASCVD risk for clinical management (Ref. IV)

High-coverage sequencing information lends more confidence in genome-wide scans to identifying true genetic underpinnings of a trait of interest compared to imputed array-derived datasets, which merely unravel causal variant containing genomic chunks. However, the phenotypic-driven analyses hold an intrinsic bias that stems from the definition of selection criteria for ascertaining the trait of interest and can therefore fundamentally dilute the signal of association. This is especially true for qualitative traits (e.g. disease outcomes), where the affected status could be difficult to ascertain and usually relies on clinically meaningful yet biologically less coherent classification systems. To overcome this issue, one can apply the reverse approach and start the selection process from genetic information. The subsequent determination of genotype-associated phenotypes or sub-phenotypes can facilitate causal inference analysis, study of biological heterogeneity and assessment of health outcomes on an individual level. With acknowledging the valuable resources contained within the Estonian Biobank, the latter approach (termed “recall-by-genotype” or RbG) on a clinically actionable singe-gene disorder, namely FH, was applied.

While FH is genetically and clinically well-described, it remains significantly underdiagnosed and inadequately treated (Nordestgaard et al., 2013). Individuals affected by this disorder have significantly increased LDL-C levels and a substantially greater risk for experiencing myocardial infarction at a young age (Khera et al., 2016). As the disease arises mainly due to a single rare

deleterious variant that hampers the hepatic clearance of LDL particles, high-coverage sequencing information provides an optimal platform for determining the individuals harbouring a causal variant. In light of the above, the strategy was to identify rare deleterious variants in well-known FH-associated genes (*LDLR*, *APOB*, *PCSK9*) using WGS and WES datasets, re-contact individuals who harbour those variants, and detail their phenotypic expression and clinical management via cascade screening and clinical evaluation.

### 3.3.1. Description of cohort and materials

FH-associated variants were ascertained in high-coverage sequencing data available for 10% of the Estonian Biobank cohort (4,776 individuals of 52,274 participants) (Leitsalu et al., 2015). WGS data was available for 2,420 individuals and WES data for 2,356 individuals. Rare (MAF <0.5%) LoF or deleterious missense variants in *LDLR* or *PCSK9,* or in the LDL receptor binding domain of *APOB* were ascertained in individuals whose untreated baseline LDL-C level was ≥4.0 mmol/L. Baseline lipid parameters (i.e., available already at the biobank prior to conducting the study) for the WGS subset had been measured with either the conventional enzymatic colorimetric or with the VAP method, and for the WES subset with the conventional method only. The effect of statin treatment in individuals, who had self-reported use at baseline lipid measurement, was taken into account by dividing LDL-C value by 0.7, as implemented previously (Peloso et al., 2014) (termed "statin-adjusted"). The project was approved by the Research Ethics Committee of the University of Tartu and complies with the Declaration of Helsinki.

### 3.3.2. Overview of the recall-by-genotype study design and clinical management of FH-associated variant carriers

A total of 11 distinct FH-associated variants in three FH-associated genes (9 in *LDLR*, 1 in *APOB* and 1 in *PCSK9*) were identified in 27 probands in heterozygous state (S.Table 4 in Ref. IV). While an invitation to join the study was sent to all 27 individuals, 21 (mean age 47.1 (SD 15.9), 9 were female) responded and were scheduled for an appointment with a clinical cardiologist and a clinical genetics specialist. At the initial appointment, probands were explained the details of the project and signed an informed consent form. Next, probands' clinical and family history (up to three generations) was ascertained, presence of physical features common for FH (tendon xanthomata, xanthelasma and corneal arcus) were assessed, and 50 ml of fasting blood from a peripheral vein for biochemical measurements and for a DNA-based confirmation of the genetic finding was drawn. If the proband had not experienced an ASCVD event prior to the study, investigations for subclinical atherosclerosis were performed (computed tomography for CAC score, carotid ultrasound for intima-

media thickness assessment and exercise electrocardiogram). During the second visit, all probands received feedback on their genetic and clinical results, and were provided guidance regarding treatment and lifestyle modifications. Furthermore, all carriers were guided to engage their first- and second-degree relatives in cascade screening. The clinical management of relatives followed the same approach as for the probands (S.Fig. 1 in Ref. IV).

Of 112 invited family members, 64 relatives (57%, mean age 46.6 (SD 17.1), 52% were female) joined the study. Only those family members, who carried the variants (n=20) and had not experienced a clinical ASCVD event (n=17), were subjected to instrumental investigations. Altogether, the RbG approach with cascade screening resulted in the identification of 41 carriers, highlighting great interest of the subjects to participate in a genomics-guided study (78% of the probands and 76% of the invited families engaged) (Fig. 11; Fig. 1 in Ref. IV, Table 1 in Ref. IV).

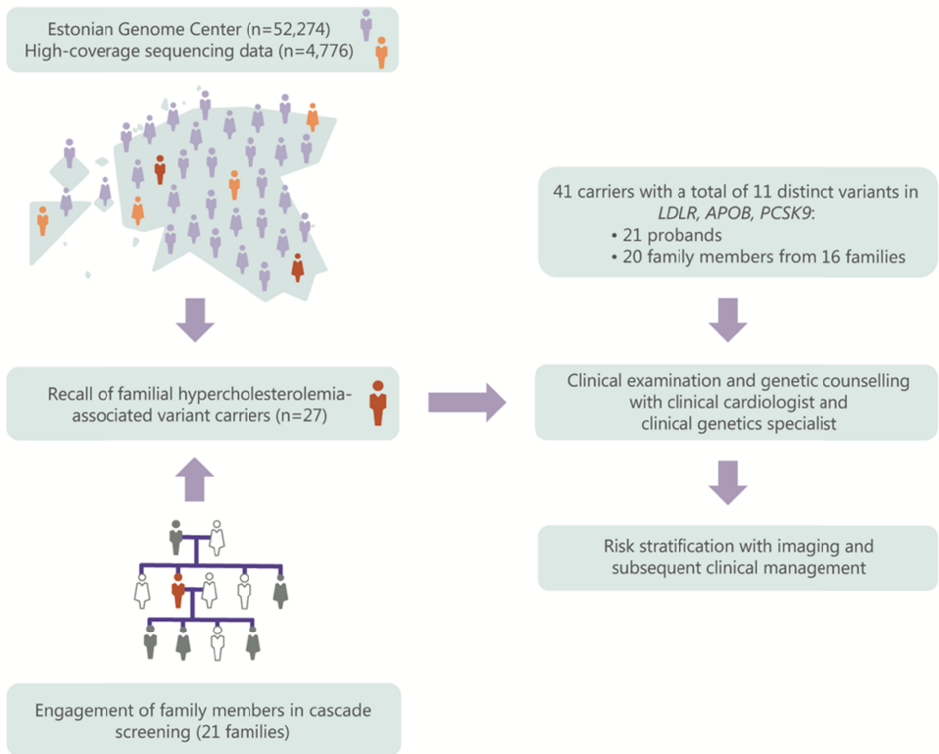



**Figure 11.** Overview of the recall-by-genotype approach for familial hypercholesterolemia within the Estonian Genome Center and subsequent clinical management.

### 3.3.3. Clinical evaluation of FH-associated variant carriers

LDL-C concentrations varied from 3.22 mmol/L to 11.56 mmol/L (mean 5.78 mmol/L (SD 2.05)) among carriers not on treatment (n=28). Statins had been prescribed for 19 carriers (46%), however, 13 (32%) of the 41 adhered to treatment, while one was taking over-the-counter lipid-lowering supplements. Importantly, LDL-C levels had not been lowered to the target level of 2.6 mmol/L recommended for FH cases without ASCVD according to the ESC guidelines (Catapano et al., 2016) in any of the 13 on treatment. The LDL-C values ranged from 3.89 mmol/L to 9.10 mmol/L (mean 5.79 mmol/L (SD 1.41), statin-adjusted). The effect of FH-associated genetic variants on LDL-C levels was interrogated in 41 FH-associated variant carriers and 978 non-carriers for whom WGS data and conventionally quantified lipid measurements were available using a linear mixed model. LDL-C value was increased on average by 2.33 mmol/L (SD 0.18, $p = 1.55\times10^{-21}$) in FH-associated variant carriers (n=41), compared to those without an FH-associated variant (n=978). Despite the significantly increased LDL-C levels in carriers, pronounced overlap in LDL-C distributions was noted between the two groups (Fig. 12; Fig. 2. in Ref. IV).

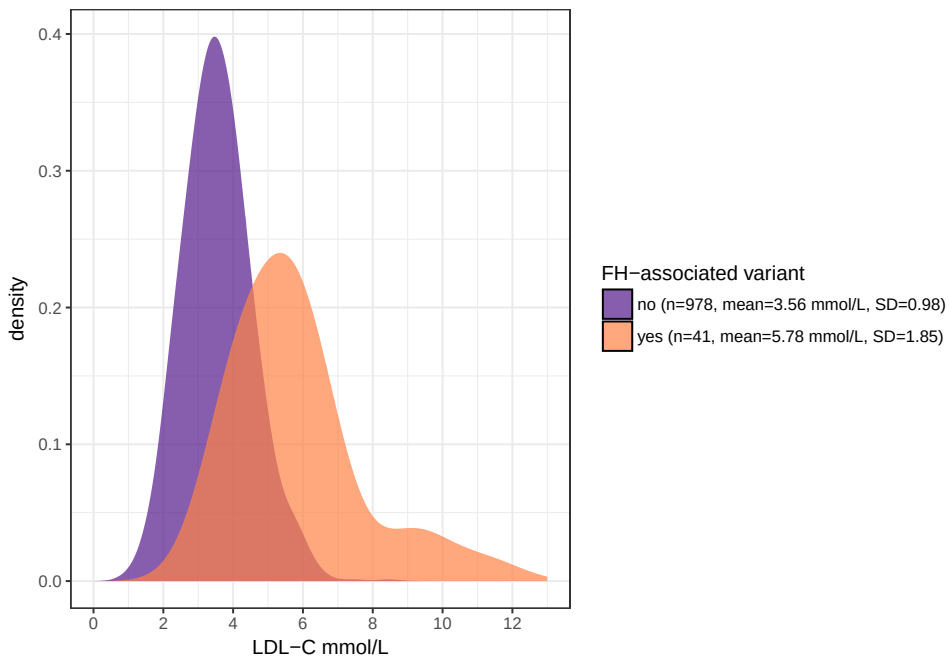



**Figure 12**. Distribution of statin-adjusted LDL-C values in FH-associated variant carriers (n=41, mean LDL-C 5.78 mmol/L, SD 1.85) and non-carriers (n=978, mean LDL-C 3.56 mmol/L, SD 0.98). The LDL-C level was increased by 2.33 mmol/L (SD 0.18) in individuals harbouring an FH-associated variant compared to non-carriers.

In regards to diagnostic criteria commonly considered in clinical FH diagnosis, 6 individuals (15%) reported positive family history of premature ASCVD in a first-degree family member, and only two carriers (5%) displayed a visible physical symptom (corneal arcus). The presence of sub-clinical ASCVD was ascertained via imaging-based phenotyping. Four FH-associated variant carriers (10%) had experienced an ASCVD event prior to the study, (mean age 73.5 (SD 11.6), 2 were female), twenty (49%) displayed sub-clinical disease (mean age 50.5 (SD 12.8), 9 were female, mean statin-adjusted LDL-C 6.31 (SD 1.76)), and fourteen (34%) did not (mean age 36.2 (SD 8.9), 6 were female, mean statin-adjusted LDL-C 4.55 (SD 1.06)). Burden of atherosclerosis could not be ascertained in three individuals (7%) as they declined the instrumental investigation procedures (S.Table 5 in Ref. V).

Of 20 individuals with sub-clinical atherosclerosis, five had plaques in carotid arteries and 19 had CAC >0. To assess, whether the atherosclerotic burden was pronounced in those with sub-clinical disease, the CAC scores of the 19 individuals were compared with the expected CAC scores distribution among those without symptomatic clinical ASCVD (2,503 men and women of Caucasian ethnicity from the MESA sub-cohort (McClelland et al., 2006)). Sixteen carriers (84%) were complete outliers in terms of the expected age- and sex-specific CAC scores determined in ASCVD-free individuals, while three participants fell within the expected distribution range. However, among the latter three, two had demonstrable plaques in carotid arteries, indicating the presence of sub-clinical disease (Fig. 13; Fig. 3 in Ref. IV).

These results permit several clinically valuable conclusions. First, lack of physical symptoms in variant carriers and substantial overlap in LDL-C levels between carriers and non-carriers denote under-performance of clinically applied diagnostic criteria. Secondly, low statin use and LDL-C levels not lowered with treatment to the target level commended by the ESC guidelines (Catapano et al., 2016) highlight unmet clinical management of FH cases. Third, pronounced clinical and sub-clinical disease among 59% of the variant carriers conforms with the disease pathology, whereby a genetic defect predisposes to the premature progression of atherosclerosis due to the lifelong exposure to increased LDL-C levels. And fourth, high heterogeneity in clinical expression, even among individuals with the same FH-associated variant, and lack of sub-clinical disease among younger participants (34%) necessitate further study for pinpointing potential modifying genetic variants and assessment of the effect of global genomic burden, and the refinement of a sub-group who requires further cardiological surveillance.

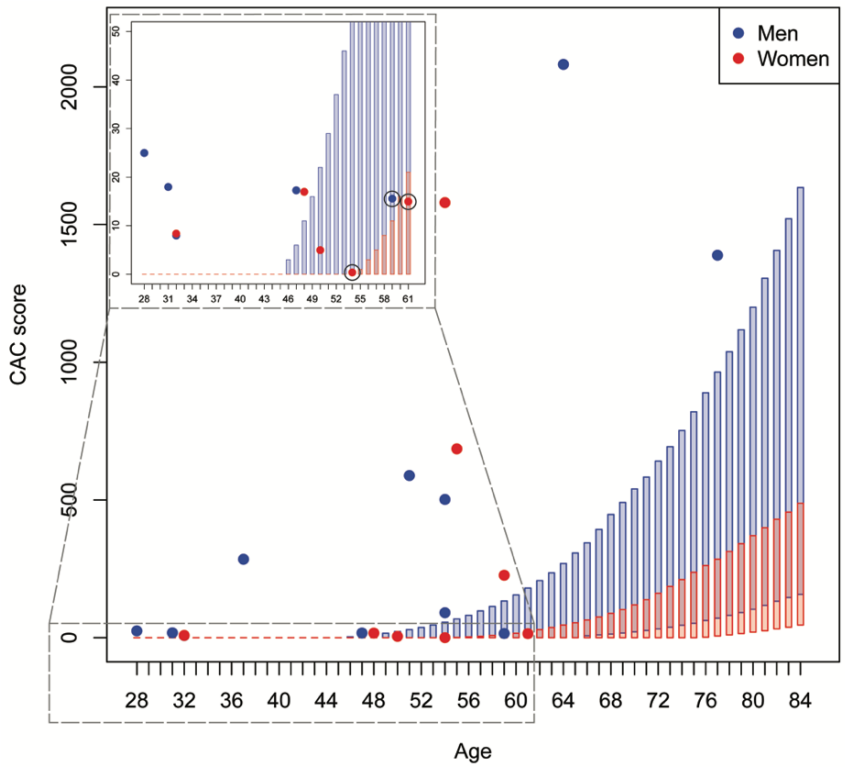



**Figure 13.** The CAC scores of FH-associated variant carriers with sub-clinical disease and CAC >0 (n=19) (filled coloured circles) in comparison with the distribution of CAC scores in the MESA sub-cohort of Caucasian ethnicity and without symptomatic clinical ASCVD and treated diabetes. The rectangles represent the expected CAC score distribution between the 25$^{th}$ and 75$^{th}$ percentile in the MESA sub-cohort for every age and for men (blue) and women (red) separately, with age on the x-axis and CAC score on the y-axis. While 16 individuals were not expected to have subclinical ASCVD, three individuals did (indicated with black circles). However, the 61-year-old female and the 54-year-old female displayed plaques in carotid arteries.

### 3.3.4. Clinical management of FH-associated variant carriers

After clinical and imaging-based phenotyping, 37 individuals (90%) were diagnosed with FH. While three (7%) carried an FH diagnosis before participating and twenty-one (51%) were reclassified from having non-specific hypercholesterolemia to having FH, thirteen (32%) had completely gone unrecognized by the medical system (Fig. 14A; Fig. 4A in Ref. IV). Four (10%) need further assessment for confident disease diagnosis. Moderate-intensity statin treatment was initiated for fourteen (34%), treatment was up-titrated for eleven (27%), left unchanged for two (5%) and not prescribed for six (14%). Eight participants (20%) either had contraindications or declined (Fig. 14B; Fig. 4B in Ref. IV). The pronounced clinical yield demonstrates the power of genomics-

guided disease management and the value of initiating the search for affected individuals within a healthcare-orientated population-based biobank with the potential to impact clinical management of affected individuals.

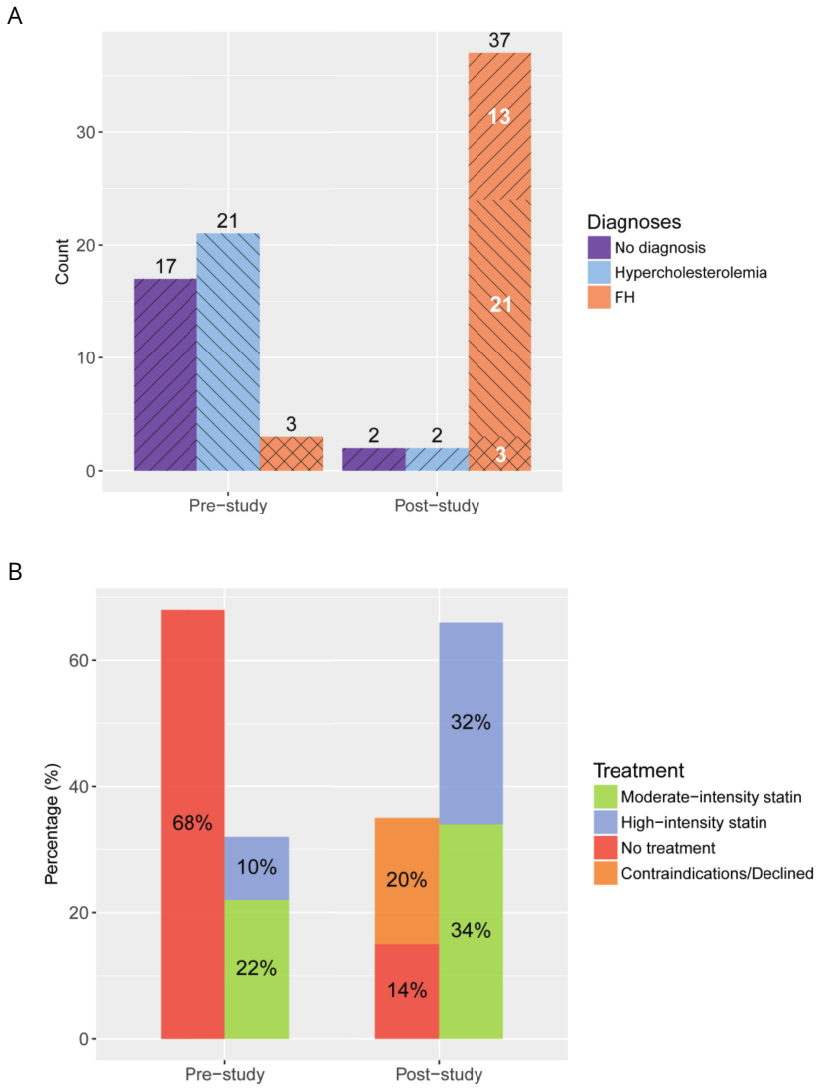



**Figure 14.** (A) Disease diagnoses and (B) statin treatment in FH-associated variant carriers before and after the study.

# 4. CONCLUSIONS

This doctoral thesis aimed to provide an overview of the status quo of the genomic research on ASCVD. With acknowledging the burden of the disease on human health, substantial endeavours have been made for understanding the disease aetiology, by identifying epidemiological factors that accelerate the disease progression, and by unravelling the genetic and molecular underpinnings that influence the disease initiation and outcome. Today, we recognize the complexity of the phenotype and acknowledge the requisite for fine detail analyses for piecing the phenotype together. These undertakings facilitate the improvement of ASCVD risk prediction strategies, such that more elaborate approaches for clinical utility can be developed.

The current approach applied in today's clinical setting to target those at increased ASCVD risk is based on the assessment of traditional risk factors. The evaluation of the predictive ability of the commonly used prediction models revealed that the companion tools of ASCVD prevention guidelines do work sufficiently well in Estonia, however, do not target high risk individuals for treatment uniformly. While the commonly applied clinical strategy is beneficial for avoiding disease incidence, it does not always result in definite prevention, since the assessment of traditional risk factors reflects the identification of already perturbed disease trajectories. To this end, the proposition to utilize genetic information has gained wide attention. Detailed analysis based on high resolution genomic data allows to unravel valuable information that can be assessed before molecular perturbations occur. The comprehensive interrogation of the genetic underpinnings of a highly heritable ASCVD biomarker, lipoprotein(a), in African-Americans and Europeans based on high-coverage sequencing data facilitated to capture a great proportion of lipoprotein(a) heritability, identify common and population-specific genetic determinants, and assess the causal impact of the parameter on ASCVD incidence. This study serves as an example demonstrating that the knowledge of the genetic determinants of an ASCVD biomarker provides additional value to the mere clinical measurement.

The final study outlined in the thesis represents the example of bridging genetic information with routine clinical practice. With identifying individuals harbouring rare deleterious variants predisposing to familial hypercholesterolemia (monogenic form of premature ASCVD) within a population-based biobank and detailing their phenotype in the clinical setting, the genotype-first approach did not only prove to be feasible and to significantly improved the clinical management of high-risk individuals, but additionally highlighted diagnostic and treatment gaps as well as underscored the heterogeneity of the phenotype. This study represents one of the first implementations of the genomics-guided strategy within a national healthcare system and demonstrates that the integration of genomic data contained in a population-based biobank into routine clinical practice has the potential to transform ASCVD risk prediction and enhance precision prevention strategies.

# SUMMARY IN ESTONIAN

## Geeniinfo väärtus südame-veresoonkonnahaiguste riski hindamisel

Ateroskleroотiline südame-veresoonkonnahaigus (ASVH) on peamiseks suremuse põhjuseks maailmas. Et tagada haiguse eelsoodumusega ja haigust põdevatele indiviididele õigeaegne ja tõhus ennetusmeede ja/või ravi, tuleb praktilisest seisukohast suunata fookus eeskätt efektiivsete ja täpsete ennetus- ja ennustusstrateegiate arendamisele. Lihtsate ja kohe kliinilises praktikas kasutust leidvate meetmete rakendamist raskendab aga tõik, et tegemist on kompleksse geneetilistest ja keskkonnateguritest mõjutatud haigusega. Selle peamiseks tekkepõhjuseks on ateroskleroos – krooniline aeglase kuluga põletikuline haigus, mille sümptomid avalduvad tihtipeale alles siis, kui haigus on molekulaarselt juba oluliselt progresseerunud.

Tänases kliinilises praktikas on kardioloogiliste eeskirjade raames väljatöötatud riskimudelid ASVH riski hindamise ja kliinilise käsitluse nurgakiviks. Need mudelid hindavad fenotüübiliste riskitegurite alusel ASVH kliiniliste avaldumisvormide tekke või nendesse suremise kümne aasta riski. Kuna algoritmid on töötatud välja aga aastakümneid tagasi kogutud kohortide andmetel, ei pruugi mudelites kasutatud statistikud vastata tänastele haiguse riskifaktorite tunnustele ega haigusjuhtude esinemissagedusele. Seda eeskätt seetõttu, et nendel samadel kohortidel põhinevad epidemioloogilised uuringud on suuresti ajendanud viimaste aastakümnete ASVH ennetusstrateegiaid ning soodustanud elustiili muutuste olulisuse rõhutamise ning ravistrateegiate edendamisega kliiniliste avaldumisvormide langustendentsi. Nimetatud aspekt on põhjuseks, miks riskimudelite omavahelised võrdlused näitavad lääne riikide kohortidel haigusriski märkimisväärset ülehindamist ning ravi määramist oluliselt rohkematele indiviididele kui vaja. Kuna Eesti on ASVH-sse suremuse osas Euroopas esimeste hulgas, võrreldi enim kasutatud riskialgoritmide ennustustäpsust Eesti Geenivaramu kui kõrge ASVH esinemissagedusega kohordis. Kuigi analüüsitulemuste kohaselt hindasid riskiskoorid kliiniliste avaldumisvormide riski üldplaanis hästi, klassifitseeriti kolmandik haigusjuhuga indiviide aga madalamasse riskikategooriasse ning pea pooltele jäi ravi määramine ebatäpseks. Oluline kitsaskoht fenotüübiliste riskiskooride rakendamisel ilmneb selles, et need mudelid hindavad riskitegurite üleslugemisega tegelikkuses molekulaarsel tasandil juba toimunud muutusi, see tähendab metaboolsete radade ja vereringesüsteemi häirumist (näiteks kolesterooli ja/või vererõhu kõrgenenud tase, suitsetamise kestus, diabeedi kaasesinemine). Seega leevendatakse praeguse strateegia kasutamisel pigem patoloogia progresseerunud kulgu, kui pärsitakse või ennetatakse molekulaarsete mehhanismide häirumist varases staadiumis.

Üheks võimalikuks riskiskooride täpsemaks muutmise lahenduseks pakutakse tunnuse geneetilise eripära arvestamist. ASVH geneetilisi seoseid on viimastel kümnenditel uuritud peamiselt detailsete perekonnauuringutega, kus

on tuvastatud haiguse monogeenseid vorme (näiteks perekondlikku hüperkolesteroleemiat) põhjustavaid kõrge efektisuurusega mutatsioone, ning genoomiülestel kiipidel põhinevate seoseuuringutega, millega on hinnatud fenotüübi polügeenset komponenti. Kuigi genoomiülesed assotsiatsiooniuuringud võimaldavad haigusseoselisi genoomipiirkondi efektiivselt tuvastada, jääb konkreetsete põhjuslike geenivariantide kindlaksmääramine kiipide tagasihoidliku resolutsiooni taha. Kõrge kattuvuse ja resolutsiooniga kogu genoomi järjestusandmestik võimaldab selles vallas olulist edasiarengut. Nimetatud lähenemist kasutati kliinilises praktikas mõõdetava ASVH biomarkeri lipoproteiin(a) geneetiliste seoste analüüsimiseks. Kolme eri populatsiooni põhine analüüs lubas hinnata biomarkeri päritavuse määra, tuvastada nii populatsioonide üleseid kui ka spetsiifilisi geneetilisi assotsiatsioone ning määrata biomarkeri ja ASVH põhjuslikku seost. Analüüsitulemused tugevdavad parameetri geneetilise komponendi hindamise vajalikkust, ennustades ASVH kliinilisi avaldumisvorme kliinikumis mõõdetud väärtustest paremini.

Näiteks, kuidas geneetilise informatsiooni kasutamine kliinilises praktikas rajab teed personaalsele meditsiinile, on Eesti Geenivaramu genotüübi-põhisel tagasikutsumisel baseeruv perekondliku hüperkolesteroleemia kliinilise käsitluse uuring. Nimelt tuvastati populatsiooni-põhise biopanga kõrge resolutsiooniga järjestusandmestikus harva geneetilise variandi, kuid kõrge haigusriskiga (perekondlikku hüperkolesteroleemiat põdevad) indiviidid ja kaskaaduuringu kaudu samu variante kandvad lähisugulased ning hinnati geneetilise nõustamise ja detailise kardioloogilise profiilimisega nende kliinilist käsitlust. Kui pooltel haigusvariandi kandjatest tuvastati varasemalt kindlaksmääratud kõrgenenud kolesterooli taseme põhjuseks geeniviga perekondliku hüperkolesteroleemiaga seotud geenis, siis oli kolmandik nendest jäänud meditsiinisüsteemis täielikult märkamata. Lisaks haiguse aladiagnoosimise ja alaravitavuse kinnitamisele, näitas töö veel olemasolevate diagnostiliste kriteeriumide ebatäpsust ning haiguse fenotüübilise pildi heterogeensust.

Käesoleva doktoritöö peamine eesmärk oli anda ülevaade, kuidas tänane ASVH riski hindamine kliinilises praktikas on muutumas ja kuidas geneetilise informatsiooni kaasamine igapäeva kliinilistesse otsustesse seda edendada võiks. Jätkuv uurimistöö, seda nii metodoloogiliste edasiarenduste, fenotüübi etioloogia täpsema mõistmise, erinevate molekulaarsete tasemete lõimimise ning suuremahuliste kohortide ja biopankade andmete integreerimisega meditsiinisüsteemi, lubab selles vallas olulist edasiarengut ning vähendada haigusest põhjustatud tüsistusi ja sellesse suremust.

# REFERENCES


Abifadel, M, Varret, M, Rabès, JP, Allard, D, Ouguerram, K, Devillers, M, et al. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature Genetics, 34:154.

Abraham, G, Havulinna, AS, Bhalala, OG, Byars, SG, De Livera, AM, Yetukuri, L, et al. (2016). Genomic prediction of coronary heart disease. European Heart Journal, 37(43): 3267–3278.

Abul-Husn, NS, Manickam, K, Jones, LK, Wright, EA, Hartzel, DN, Gonzaga-Jauregui, C, et al. (2016). Genetic identification of familial hypercholesterolemia within a single U.S. health care system. Science, 354(6319):aaf7000.

Albers, JJ, Koschinsky, ML, Marcovina, SM. (2007). Evidence mounts for a role of the kidney in lipoprotein(a) catabolism. Kidney International, 71(10):961–962.

Alonso, R, Andres, E, Mata, N, Fuentes-Jiménez, F, Badimón, L, López-Miranda, J, et al. (2014). Lipoprotein(a) levels in familial hypercholesterolemia: an important predictor of cardiovascular disease independent of the type of LDL receptor mutation. Journal of the American College of Cardiology, 63(19):1982–1989.

Amiri, F, Virdis, A, Fritsch Neves, M, Iglarz, M, Seidah, NG, Touyz, RM, et al. (2004). Endothelium-restricted overexpression of human endothelin-1 causes vascular remodeling and endothelial dysfunction. Circulation, 110(15):2233–2240.

Arneson, D, Shu, L, Tsai, B, Barrere-Cain, R, Sun, C, and Yang, X. (2017). Multi-dimensional integrative genomics approaches to dissecting cardiovascular disease. Frontiers in Cardiovascular Medicine, 4:8.

ATC. (2018). Retrieved from: https://www.whocc.no/atc_ddd_index/

Austin, MA, Sandholzer, C, Selby, JV, Newman, B, Krauss, RM, and Utermann, G. (1992). Lipoprotein(a) in women twins: heritability and relationship to apolipo-protein(a) phenotypes. American Journal of Human Genetics, 51(4):829–840.

Auton, A, Abecasis, GR, Brooks, LD, Durbin, RM, Garrison EP, Korbel, JO, et al. (2015). A global reference for human genetic variation. Nature, 526(7571):68–74.

Bauer, RC, Tohyama, J, Cui, J, Cheng, L, Yang, J, Zhang, X, et al. (2015). Knockout of Adamts7, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. Circulation, 131(13):1202–1213.

Beaudoin, M, Gupta, RM, Won, HH, Lo, KS, Do, R, Henderson, CA, et al. (2015). Myocardial infarction-associated SNP at 6p24 interferes with MEF2 binding and associates with PHACTR1 expression levels in human coronary arteries. Arterio-sclerosis, Thrombosis, and Vascular Biology, 35:1472–1479.

Bell, DA, Pang, J, Burrows, S, Bates, TR, van Bockxmeer, FM, Hooper, AJ, et al. (2015). Effectiveness of genetic cascade screening for familial hypercholeste-rolaemia using a centrally co-ordinated clinical service: an Australian experience. Atherosclerosis, 239(1):93–100.

Benn, M, Watts, GF, Tybjærg-Hansen, A, Nordestgaard, BG. (2016). Mutations causative of familial hypercholesterolaemia: screening of 98 098 individuals from the Copenhagen General Population Study estimated a prevalence of 1 in 217. European Heart Journal, 37(17):1384–1394.

Berberich, AJ, Hegele, RA. (2019). The complex molecular genetics of familial hyper-cholesterolemia. Nature Reviews Cardiology, 16(1):9–20.

Berglund, L, Ramakrishnan, R. (2004). Lipoprotein(a): an elusive cardiovascular risk factor. Arteriosclerosis, Thrombosis, and Vascular Biology, 24(12):2219–2226.

Bergmark, C, Dewan, A, Orsoni, A, Merki, E, Miller, ER, Shin, MJ, et al. (2008). A novel function of lipoprotein[a] as a preferential carrier of oxidized phospholipids in human plasma. Journal of Lipid Research, 49(10):2230–2239.

Boffa, MB, Marcovina, SM, and Koschinsky, ML. (2004). Lipoprotein(a) as a risk factor for atherosclerosis and thrombosis: mechanistic insights from animal models. Clinical Biochemistry, 37(5):333–343.

Bulik-Sullivan, B, Finucane, HK, Anttila, V, Gusev, A, Day, FR, Loh, PR, et al. (2015). An atlas of genetic correlations across human diseases and traits. Nature Genetics, 47(11):1236–1241.

Burdett, H. (2015). Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. Annals of Clinical Biochemistry, 53(3):415.

Bycroft, C, Freeman, C, Petkova, D, Band, G, Elliott, LT, Sharp, K, et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562 (7726):203–209.

Carey, DJ, Fetterolf, SN, Davis, FD, Faucett, WA, Kirchner, HL, Mirshahi, U, et al. (2016). The Geisinger MyCode Community Health Initiative: an electronic health record-linked biobank for Precision Medicine research. Genetics in Medicine, 18(9):906–913.

CardioGRAMplusC4D Consortium, Deloukas, P, Kanoni, S, Willenborg, C, Farrall, M, Assimes, TL, Thompson, JR, et al. (2012). Large-scale association analysis identifies new risk loci for coronary artery disease. Nature Genetics, 45(1):25–33.

Catapano, AL, Graham, I, De Backer, G, Wiklund, O, Chapman, MJ, Drexel, H, et al. (2016). 2016 ESC/EAS guidelines for the management of dyslipidaemias. European Heart Journal, 37(39):2999–3058l.

Causes of Death Registry (2018). Retrieved from: http://www.tai.ee/en/r-and-d/registers/estonian-causes-of-death-registry

Chatterjee, N, Shi, J, and García-Closas, M. (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nature Reviews Genetics, 17(7);392–406.

Chou, JY, Matern, D, Chen, Mansfield, BC, Chen, YT. (2002). Type I glycogen storage diseases: disorders of the glucose-6-phosphatase complex. Current Molecular Medicine, 2(2):121–43.

Civeira, F, Ros, E, Jarauta, E, Plana, N, Zambon, D, Puzo, J, et al. (2008). Comparison of genetic versus clinical diagnosis in familial hypercholesterolemia. American Journal of Cardiology, 102(9).

Clarke, R, Peden, JF, Hopewell, JC, Kyriakou, T, Goel, A, Heath, SC, et al. (2009). Genetic variants associated with Lp(a) lipoprotein level and coronary disease. New England Journal of Medicine, 361(26):2518–2528.

Cohen, J, Pertsemlidis, A, Kotowski, IK, Graham, R, Garcia, CK, Hobbs, HH. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nature Genetics, 37:161.

Cohen, JC, Boerwinkle, E, Mosley, TH, Hobbs, HH. (2006). Sequence Variations in PCSK9, low LDL, and protection against coronary heart disease. New England Journal of Medicine, 354(12):1264–1272.

Coronary Artery Disease Genetics Consortium. (2011). A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nature Genetics, 43(4):339–44.

Crosby, J, Peloso, GM, Auer, PL, Crosslin, DR, Stitziel, NO, Lange LA, et al. (2014). Loss-of-function mutations in APOC3, triglycerides, and coronary disease. New England Journal of Medicine, 371(1):22–31.

Deb, A, Caplice, NM. (2006). Lipoprotein(a): new insights into mechanisms of atherogenesis and thrombosis. Clinical Cardiology, 27(5):258–264.

Deelen, P, Menelaou, A, van Leeuwen, EM, Kanterakis, A, van Dijk, F, Medina-Gomez, C, et al. (2014). Improved imputation quality of low-frequency and rare variants in European samples using the "Genome of The Netherlands." European Journal Of Human Genetics, 22:1321.

Defesche, JC, Lansberg, PJ, Umans-Eckenhausen, MA, Kastelein, JJ. (2004). Advanced method for the identification of patients with inherited hypercholesterolemia. Seminars in Vascular Medicine, 4(1):59–65.

Defilippis, AP, Young, R, Carrubba, CJ, McEvoy, JW, Budoff, MJ, Blumenthal, RS. (2015). An analysis of calibration and discrimination among multiple cardiocavascular risk scores in a modern multiethnic cohort. Annals of Internal Medicine, 162:266–275.

DeFilippis, AP, Young, R, McEvoy, JW, Michos, ED, Sandfort, V, Kronmal, RA, et al. (2017). Risk score overestimation: the impact of individual cardiovascular risk factors and preventive therapies on the performance of the American Heart Association-American College of Cardiology-Atherosclerotic Cardiovascular Disease risk score in a modern multi-ethnic cohort. European Heart Journal, 38(8):598–608.

Degoma, EM, Ahmad, ZS, O'Brien, EC, Kindt, I, Shrader, P, Newman, CB, et al. (2016). Treatment gaps in adults with heterozygous familial hypercholesterolemia in the United States. Circulation: Cardiovascular Genetics, 9(3):240–249.

Demir, OM, Dobson, P, Papamichael, ND, Byrne, J, Plein, S, Alfakih, K. (2015). Comparison of ESC and NICE guidelines for patients with suspected coronary artery disease: evaluation of the pre-test probability risk scores in clinical practice. Clinical Medicine, 15(3):234–238.

DeMott K, Nherera L, Shaw EJ, Minhas R, Humphries SE, Kathoria M, RG, et al. (2008). Clinical guidelines and evidence review for familial hypercholesterolaemia: the identification and management of adults and children with familial hypercholesterolaemia. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners.

Denny, JC. (2012). Chapter 13: Mining Electronic Health Records in the Genomics Era. PLoS Computational Biology, 8(12).

Denny, JC, Bastarache, L, Ritchie, MD, Carroll, RJ, Zink, R, Mosley, JD, et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature Biotechnology, 31:1102.

Dewey, FE, Gusarova, V, O'Dushlaine, C, Gottesman, O, Trejos, J, Hunt, C, et al. (2016). Inactivating variants in ANGPTL4 and risk of coronary artery disease. New England Journal of Medicine, 374(12):1123–1133.

Dewey, FE, Murray, MF, Overton, JD, Habegger, L, Leader, JB, Fetterolf, SN, et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science, 354(6319).

Do, R, Stitziel, NO, Won, HH, Jørgensen, AB, Duga, S, Angelica Merlini, P, et al. (2014). Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature, 518:102.

EHIF. (2018). Retrieved from: https://www.haigekassa.ee/en
Emdin, CA, Khera, AV, Kathiresan, S. (2017). Mendelian randomization. JAMA, 318(19):1925–1926.
Erdmann, J, Stark, K, Esslinger, UB, Rumpf, PM, Koesling, D, de Wit, C, et al. (2013). Dysfunctional nitric oxide signalling increases risk of myocardial infarction. Nature, 504: 432.
Erdmann, J, Kessler, T, Venegas, LM, Schunkert, H. (2018). A decade of genome-wide association studies for coronary artery disease: the challenges ahead. Cardiovascular Research, 114(9):1241–1257.
Ergou, S, Kaptoge, S, Perry, PL, Di Angelantonio, E, Thompson, A, White, IR, et al. (2016). Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvacular mortality. JAMA, 302(4):412–423.
Evangelou, E, Warren, HR, Mosen-Ansorena, D, Mifsud, B, Pazoki, R, Gao, H, et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nature Genetics, 50(10):1412–1425.
Falk, E. (2006). Pathogenesis of Atherosclerosis. Journal of the American College of Cardiology, 47(8 Suppl):C7–12.
Feric, NT, Boffa, MB, Johnston, SM, and Koschinsky, ML. (2008). Apolipoprotein(a) inhibits the conversion of Glu-plasminogen to Lys-plasminogen: a novel mechanism for lipoprotein(a)-mediated inhibition of plasminogen activation. Journal of Thrombosis and Haemostasis, 6(12):2113–2120.
Folsom, AR. (2013). Classical and novel biomarkers for cardiovascular risk prediction in the United States. Journal of Epidemiology, 23(3):158–162.
Franzén, O, Ermel, R, Cohain, A, Akers, NK, Di Narzo, A, Talukdar, HA, et al. (2016). Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science, 353(6301):827 LP–830.
Ganna, A, Magnusson, PKE, Pedersen, NL, de Faire, U, Reilly, M, Arnlov, J, et al. (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, Thrombosis, and Vascular Biology, 33(9):2267–2272.
Gaziano, JM, Concato, J, Brophy, M, Fiore, L, Pyarajan, S, Breeling, J, et al. (2016). Million Veteran Program: a mega-biobank to study genetic influences on health and disease. Journal of Clinical Epidemiology, 70:214–223.
Goff, DC, Lloyed-Jones, Bennet, G, Coady, S, D'Agostino, RB, Gibbons, R, et al. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. Journal of the American College of Cardiology, 63:2935–2959.
Goldmann, R, Tichý, L, Freiberger, T, Zapletalová, P, Letocha, O, Soska, V, et al. (2010). Genomic characterization of large rearrangements of the LDLR gene in Czech patients with familial hypercholesterolemia. BMC Medical Genetics, 11:115.
Gottesman, O, Kuivaniemi, H, Tromp, G, Faucett, WA, Li, R, Manolio, TA, et al. (2013). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genetics In Medicine, 15:761.
Gray, BJ, Bracken, RM, Turner, D, Morgan, K, Mellalieu, SD, Thomas, M, et al. (2014). Predicted 10-year risk of cardiovascular disease is influenced by the risk equation adopted: a cross-sectional analysis. British Journal of General Practice, 64(627):e634–e640.
Greenland, P, Robert, OB. (2016). Interpretation and use of another statin guideline. JAMA, 316(19):1977–1979.

Guardiola, M, Ribalta, J. (2017). Update on APOA5 genetics: toward a better understanding of its physiological impact. Current Atherosclerosis Reports, 19(7):30.
Gudbjartsson, DF, Sulem, P, Helgason, H, Gylfason, A, Gudjonsson, SA, Zink, F, et al. (2015). Sequence variants from whole genome sequencing a large group of Icelanders. Scientific Data, 2, 150011.
Guo, M, Dauber, A, Lippincott, M, Chan, YM, Salem, R, Hirschhorn, J. (2016). Determinants of power in gene-based burden testing for monogenic disorders. American Journal of Human Genetics, 99(3):527–539.
Gupta, RM, Hadaya, J, Trehan, A, Zekavat, SM, Roselli, C, Klarin, D, et al. (2017). A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell, 170(3):522–533.
Handsaker, RE, Korn, JM, Nemesh, J, McCarroll, SA. (2011). Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. Nature Genetics, 43:269.
Harismendy, O, Notani, D, Song, X, Rahim, NG, Tanasa, B, Heintzman, N, et al. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature, 470:264.
Hartwig, FP, Davies, NM, Hemani, G, Davey Smith, G. (2016). Two-sample mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. International Journal of Epidemiology, 45(6):1717–1726.
Hasin, Y, Seldin, M, Lusis, A. (2017). Multi-omics approaches to disease. Genome Biology, 18:83.
He, KY, Ge, D, and He, MM. (2017). Big data analytics for genomic medicine. International Journal of Molecular Sciences, 18(2):412.
Helgadottir, A, Thorleifsson, G, Manolescu, A, Gretarsdottir, S, Blondal, T, Jonasdottir, A, et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science, 316(5830):1491 LP–1493.
Hemingway, H, Asselbergs, FW, Danesh, J, Dobson, R, Maniadakis, N, Maggioni, A, et al. (2018). Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. European Heart Journal, 39(16):1481–1495.
Hippisley-Cox, J, Coupland, C, Robson, J, and Brindle, P. (2010). Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ, 341.
Holdt, LM, Beutner, F, Scholz, M, Gielen, S, Gäbel, G, Bergert, H, et al. (2010). ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. Arteriosclerosis, Thrombosis, and Vascular Biology, 30(3):620–627.
Holdt, LM, Stahringer, A, Sass, K, Pichler, G, Kulak, NA, Wilfert, W, et al. (2016). Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. Nature Communications, 7:12429.
Huang, J, Howie, B, McCarthy, S, Memari, Y, Walter, K, Min, JL, et al. (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nature Communications, 6:8111.
ICD-10. (2016). Retrieved from: https://icd.who.int/browse10/2016/en
Inouye, M, Abraham, G, Nelson, CP, Wood, AM, Sweeting, MJ, Dudbridge, F, et al. (2018). Genomic risk prediction of coronary artery disease in 480,000 adults. Journal of the American College of Cardiology, 72(16):1883 LP–1893.

Jannes, CE, Santos, RD, de Souza Silva, PR, Turolla, L, Gagliardi, ACM, Marsiglia, JDC, et al. (2015). Familial hypercholesterolemia in Brazil: cascade screening program, clinical and genetic aspects. Atherosclerosis, 238(1):101–107.

Jarinova, O, Stewart, AFR, Roberts, R, Wells, G, Lau, P, Naing, T, et al. (2009). Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. Arteriosclerosis, Thrombosis, and Vascular Biology, 29(10):1671–1677.

Jørgensen, AB, Frikke-Schmidt, R, Nordestgaard, BG, and Tybjærg-Hansen, A. (2014). Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. New England Journal of Medicine, 371(1):32–41.

Kaiser Permanente Research Program on Genes, Environment and Health. (2018). Retrieved from: https://divisionofresearch.kaiserpermanente.org/genetics/rpgeh

Kamstrup, PR, Tybjærg-Hansen, A, Steffensen, R, and Nordestgaard BG. (2009). Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. JAMA, 301(22):2331–2339.

Kathiresan, S, Willer, CJ, Peloso, GM, Demissie, S, Musunuru, K, Schadt, EE. (2008). Common variants at 30 loci contribute to polygenic dyslipidemia. Nature Genetics, 41:56.

Kavousi, M, Leening, MJG, Nanchen, D, Greenland, P, Graham, IM, Steyerberg, EW, et al. (2014). Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European cohort. JAMA, 311(14):1416.

Kee, F, Tiret, L, Robo, JY, Nicaud, V, McCrum, E, Evans, A, and Cambien, F. (1993). Reliability of reported family history of myocardial infarction. British Medical Journal, 307(6918):1528–1530.

Kerr, M, Pears, R, Miedzybrodzka, Z, Haralambos, K, Cather, M, Watson, M, Humphries, SE. (2017). Cost effectiveness of cascade testing for familial hypercholesterolaemia, based on data from familial hypercholesterolaemia services in the UK. European Heart Journal, 38(23):1832–1839.

Kessler, T, Zhang, L, Liu, Z, Yin, X, Huang, Y, Wang, Y, et al. (2015). ADAMTS-7 inhibits re-endothelialization of injured arteries and promotes vascular remodeling through cleavage of thrombospondin-1. Circulation, 131(13):1191–1201.

Kettunen, J, Demirkan, A, Würtz, P, Draisma, HHM, Haller, T, Rawal, R, et al. (2016). Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nature Communications, 7:1–9.

Khambhati, J, Allard-Ratick, M, Dhindsa, D, Lee, S, Chen, J, Sandesara, PB, et al. (2018). The art of cardiovascular risk assessment. Clinical Cardiology, 41(5):677–684.

Kathiresan, S, Voight, BF, Purcell, S, Musunuru, K, Ardissino, D, Mannucci, PM, et al. (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nature Genetics, 41(3):334–41.

Khera, AV, Emdin, CA, Drake, I, Natarajan, P, Bick, AG, Cook, NR, et al. (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. New England Journal of Medicine, 375(24):2349–2358.

Khera, AV, Won, HH, Peloso, GM, Lawson, KS, Bartz, TM, Deng, X, et al. (2016). Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. Journal of the American College of Cardiology, 67(22):2578–2589.

Khera, AV, Won, H, Peloso, GM, O'Dushlaine, C, Liu, D, Stitziel, NO, et al. (2017). Association of rare and common variation in the lipoprotein lipase gene with coronary artery disease. JAMA, 317(9):937–946.

Khera, AV, Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. Nature Reviews Genetics, 18(6):331–344.

Khera, AV, Chaffin, M, Aragam, KG, Haas, ME, Roselli, C, Choi, SH, et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics, 50(9):f1219–1224.

Klarin, D, Zhu, QM, Emdin, CA, Chaffin, M, Horner, S, McMillan, BJ, et al. (2017). Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nature Genetics, 49:1392.

Verma, A, Bang, L, Miller, JE, Zhang, Y, Lee, M, Zhang, Y, et al. (2019). Human-disease phenotype map derived from PheWAS across 38,682 individuals. American Journal of Human Genetics, 104(1):55–64.

Koschinsky, ML, Marcovina, SM. (2004). Structure–function relationships in apolipoprotein(a): insights into lipoprotein(a) assembly and pathogenicity. Genetics and Molecular Biology, 15(2):167–174.

Kraft, HG, Köchl, S, Menzel, HJ, Sandholzer, C, and Utermann, G. (1992). The apolipoprotein(a) gene: a transcribed hypervariable locus controlling plasma lipoprotein(a) concentration. Human Genetics, 90(3):220–230.

Krogh, HW, Mundal, L, Holven, KB, and Retterstøl, K. (2016). Patients with familial hypercholesterolaemia are characterized by presence of cardiovascular disease at the time of death. European Heart Journal, 37(17):1398–1405.

Kulkarni, KR. (2006). Cholesterol profile measurement by vertical auto profile method. Clinics in Laboratory Medicine, 26(4):787–802.

Kulkarni, KR, Garber, DW, Marcovina, SM, Segrest, JP. (1994). Quantification of cholesterol in all lipoprotein classes by the VAP-II method. Journal of Lipid Research, 35(1):159–168.

Kullo, IJ, Jouni, H, Austin, EE, Brown, SA, Kruisselbrink, TM, Isseh, IN, et al. (2016). Incorporating a genetic risk score into coronary heart disease risk estimates. Circulation, 133(12):1181–1188.

Kundaje, A, Meuleman, W, Ernst, J, Bilenky, M, Yen, A, Heravi-Moussavi, A, et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature, 518:317.

Kyriakou, T, Seedorf, U, Goel, A, Hopewell, JC, Clarke, R, Watkins, H, and Farrall, M. (2014). A common LPA null allele associates with lower lipoprotein(a) levels and coronary artery disease risk. Arteriosclerosis, Thrombosis, and Vascular Biology, 34(9):2095–2099.

Lafferty, MJ, Bradford, KC, Erie, DA, Neher, SB. (2013). Angiopoietin-like protein 4 inhibition of lipoprotein lipase: evidence for reversible complex formation. Journal of Biological Chemistry, 288(40):28524–28534.

Lafont, A. (2003). Basic aspects of plaque vulnerability. Heart, 89(10):1262–1267.

Lamon-Fava, S, Jimenez, D, Christian, JC, Fabsitz, RR, Reed, T, Carmelli, D, et al. (1991). The NHLBI Twin Study: heritability of apolipoprotein A-I, B, and low density lipoprotein subclasses and concordance for lipoprotein(a). Atherosclerosis, 91(1):97–106.

Lanktree, MB, Anand, SS, Yusuf, S, Hegele, RA. (2010). Comprehensive analysis of genomic variation in the LPA locus and its relationship to plasma lipoprotein(a) in South Asians, Chinese, and European Caucasians. Circulation: Cardiovascular Genetics, 3(1):39–46.

Lau, E, Wu, JC. (2018). Omics, big data, and precision medicine in cardiovascular sciences. Circulation Research, 122(9):1165–1168.
Lázaro, P, Pérez de Isla, L, Watts, GF, Alonso, R, Norman, R, Muñiz, O, et al. (2017). Cost-effectiveness of a cascade screening program for the early detection of familial hypercholesterolemia. Journal of Clinical Lipidology, 11(1):260–271.
LeBlanc, M, Zuber, V, Andreassen, BK, Witoelar, A, Zeng, L, Bettella, F, et al. (2016). Identifying novel gene variants in coronary artery disease and shared genes with several cardiovascular risk factors. Circulation Research, 118(1):83–94.
Lee, S, Abecasis, GR, Boehnke, M, and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. American Journal of Human Genetics, 95(1):5–23.
Leebmann, J, Roeseler, E, Julius, U, Heigl, F, Spitthoever, R, Heutling, D, et al. (2013). Lipoprotein apheresis in patients with maximally tolerated lipid-lowering therapy, lipoprotein(a)-hyperlipoproteinemia, and progressive cardiovascular disease prospective observational multicenter study. Circulation, 128(24):2567–2576.
Lehrman, MA, Schneider, WJ, Sudhof, TC, Brown, MS, Goldstein, JL, Russell, DW. (1985). Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science, 227(4683):140 LP–146.
Leigh, S, Futema, M, Whittall, R, Taylor-Beadling, A, Williams, M, den Dunnen, JT, Humphries, SE. (2017). The UCL low-density lipoprotein receptor gene variant database: pathogenicity update. Journal of Medical Genetics, 54(4):217 LP–223.
Leitsalu, L, Haller, T, Esko, T, Tammesoo, ML, Alavere, H, Snieder, H, et al. (2015). Cohort profile: Estonian biobank of the Estonian Genome Center, University of Tartu. International Journal of Epidemiology, 44(4):1137–1147.
Libby, P, Ridker, PM, Hansson, GK. (2011). Progress and challenges in translating the biology of atherosclerosis. Nature, 473(7347):317–325.
Lim, ET, Würtz, P, Havulinna, AS, Palta, P, Tukiainen, T, Rehnström, K, et al. (2014). Distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genetics, 10(7).
Li, M, Jiang Yu, Wedow, R, Li, Y, Brazel, DM, Chen, F, et al. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nature Genetics, Jan 14.
Lloyd-Jones, DM, Nam, BH, D'Agostino, RB, Levy D, Murabito JM, et al. (2004). Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA, 291(18):2204–2211.
Lo Sardo, V, Chubukov, P, Ferguson, W, Kumar, A, Teng, EL, Duran, M, et al. (2018). Unveiling the role of the most impactful cardiovascular risk locus through haplotype editing. Cell, 175(7):1796–1810.
Mack, S, Coassin, S, Rueedi, R, Yousri, NA, Seppälä, I, Gieger, C, et al. (2017). A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. Journal of Lipid Research, 58(9):1834–1844.
Mahajan, A, Taliun, D, Thurner, M, Robertson, NR, Torres, JM, Rayner, NW, et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nature Genetics, 50(11): 1505–1513.
Marchini, J, Howie, B. (2010). Genotype imputation for genome-wide association studies. Nature Reviews Genetics,11:499.

McCarthy, MI, Abecasis, GR, Cardon, LR, Goldstein, DB, Little, J, Ioannidis, JPA, Hirschhorn, JN. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics, 9:356.

McClelland, RL, Chung, H, Detrano, R, Post, W, and Kronmal, RA. (2006). Distribution of coronary artery calcium by race, gender, and age. Circulation, 113(1):30 LP–37.

McConnell, JP, Guadagno, PA, Dayspring, TD, Hoefner, DM, Thiselton, DL, Warnick, GR, and Harris, WS. (2014). Lipoprotein(a) mass: a massively misunderstood metric. Journal of Clinical Lipidology, 8(6):550–553.

McPherson, R, Pertsemlidis, A, Kavaslar, N, Stewart, A, Roberts, R, Cox, DR, et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. Science, 316(5830):1488 LP–1491.

Mead, JR, Irvine, SA, Ramji, DP. (2002). Lipoprotein lipase: structure, function, regulation, and role in disease. Journal of Molecular Medicine, 80(12):753–769.

Mega, JL, Stitziel, NO, Smith, JG, Chasman, DI, Caulfield, M, Devlin, JJ, et al. (2015). Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy. The Lancet, 385(9984):2264–2271.

Mitt, M, Kals, M, Pärn, K, Gabriel, SB, Lander, ES, Palotie, A, et al. (2017). Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. European Journal Of Human Genetics, 25:869.

Mooser, V, Scheer, D, Marcovina, SM, Wang, J, Guerra, R, Cohen, J, and Hobbs, HH. (1997). The Apo(a) gene is the major determinant of variation in plasma Lp(a) levels in African Americans. American Journal of Human Genetics, 61(2):402–417.

Mortensen, MB, Falk, E. (2014). Real-life evaluation of European and American high-risk strategies for primary prevention of cardiovascular disease in patients with first myocardial infarction. BMJ Open, 4(10):1–8.

Mortensen, MB, Nordestgaard, BG, Afzal, S, Falk, E. (2017). ACC/AHA guidelines superior to ESC/EAS guidelines for primary prevention with statins in non-diabetic Europeans: the Copenhagen General Population Study. European Heart Journal, 38(8):586–594.

Mortensen, M, Nordestgaard, B. (2018). Comparison of five major guidelines for statin use in primary prevention in a contemporary general population. Annals of Internal Medicine, 168(2):85–92.

Müller, C. (1938). Xanthomata, hypercholesterolemia, angina pectoris. Journal of Internal Medicine, 95(89):75–84.

Murabito JM, Pencina MJ, Nam BH, D'Agostino RB, Wang TJ, Lloyd-Jones D, et al. (2005). Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA, 294(24):3117–3123.

Musunuru, K, Kathiresan S. (2019). Genetics of common, complex coronary artery disease. Cell, 177(1):132–145.

Musunuru, K, Strong, A, Frank-Kamenetsky, M, Lee, NE, Ahfeldt, T, Sachs, KV, et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature, 466:714.

Natarajan, P, Young, R, Stitziel, NO, Padmanabhan, S, Baber, U, Mehran, R, et al. (2017). Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. Circulation, 135(22):2091–2101.

Nayor, M, Vasan, RS. (2016). Recent update to the US cholesterol treatment guidelines: a comparison with international guidelines. Circulation, 133(18):1795–1806.
Nelson, CP, Goel, A, Butterworth, AS, Kanoni, S, Webb, TR, Marouli, E, et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nature Genetics, 49:1385.
NICE. Cardiovascular disease: risk assessment and reduction, including lipid modification. (2018). Retrieved from: https://www.nice.org.uk/guidance/cg181
Nikpay, M, Goel, A, Won, HH, Hall, LM, Willenborg, C, Kanoni S, et al. (2015). A comprehensive 1000 genomes–based genome-wide association meta-analysis of coronary artery disease. Nature Genetics, 47(10):1121–1130.
Nielsen, LB. (1999). Atherogenecity of lipoprotein(a) and oxidized low density lipoprotein: insight from in vivo studies of arterial wall influx, degradation and efflux. Atherosclerosis, 143(2):229–243.
Nioi, P, Sigurdsson, A, Thorleifsson, G, Helgason, H, Agustsdottir, AB, Norddahl, GL, et al. (2016). Variant ASGR1 associated with a reduced risk of coronary artery disease. New England Journal of Medicine, 374(22):2131–2141.
NOMESCO. Classification of Surgical Procedures. Version 1.15, no.93:2010 Copenhagen. (2011). Retrieved from: https://norden.divaportal.org/smash/get/ diva2: 970547/FULLTE XT 01.pdf
Nordestgaard, BG, Chapman, MJ, Ray, K, Borén, J, Andreotti, F, Watts, GF, et al. (2010). Lipoprotein(a) as a cardiovascular risk factor: current status. European Heart Journal, 31(23):2844–2853.
Nordestgaard, BG, Chapman, MJ, Humphries, SE, Ginsberg, HN, Masana, L, Descamps, OS, et al. (2013). Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: Guidance for clinicians to prevent coronary heart disease. European Heart Journal, 34(45):3478–3490.
Noureen, A, Fresser, F, Utermann, G, and Schmidt, K. (2015). Sequence variation within the KIV-2 copy number polymorphism of the human LPA gene in African, Asian, and European populations. PLoS ONE, 10(3):1–23.
Pavlović, J, Greenland, P, Deckers, JW, Brugts, JJ, Kavousi, M, Dhana, K, et al. (2016). Comparison of ACC/AHA and ESC guideline recommendations following trial evidence for statin use in primary prevention of cardiovascular disease: results from the population-based rotterdam study. JAMA Cardiology, 1(6), 708–713.
Peloso, GM, Auer, PL, Bis, JC, Voorman, A, Morrison, AC, Stitziel, NO, et al. (2014). Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. American Journal of Human Genetics, 94(2):223–232.
Perak, AM, Ning, H, De Ferranti, SD, Gooding, HC, Wilkins, JT, and Lloyd-Jones, DM. (2016). Long-term risk of atherosclerotic cardiovascular disease in US adults with the familial hypercholesterolemia phenotype. Circulation, 134(1):9–19.
Podrez, EA, Poliakov, E, Shen, Z, Zhang, R, Deng, Y, Sun, M, et al. (2002). A novel family of atherogenic oxidized phospholipids promotes macrophage foam cell formation via the scavenger receptor CD36 and is enriched in atherosclerotic lesions. Journal of Biological Chemistry, 277(41):38517–38523.
Piepoli MF, Hoes, AW, Agewall S, Albus C,Brotons C, Catapano AL, et al. (2016). 2016 European guidelines on cardiovascular disease prevention in clinical practice. The Sixth Joint Task Force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice. European Heart Journal 37(29): 2315–2381.

Pu, X, Xiao, Q, Kiechl, S, Chan, K, Ng, FL, Gor, S, et al. (2013). ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. American Journal of Human Genetics, 92(3):366–374.

Raal, FJ, Giugliano, RP, Sabatine, MS, Koren, MJ, Langslet, G, Bays, H, et al. (2014). Reduction in lipoprotein(a) with PCSK9 monoclonal antibody evolocumab (AMG 145): A pooled analysis of more than 1,300 patients in 4 phase II trials. Journal of the American College of Cardiology, 63(13):1278–1288.

Rader, DJ, Damrauer, SM. (2016). "Pheno" menal value for human health. Science, 354(6319):1534 LP–1536.

Reiner, Ž. (2015). Management of patients with familial hypercholesterolaemia. Nature Reviews Cardiology, 12(16):565–575.

Ripatti, S, Tikkanen, E, Orho-Melander, M, Havulinna, AS, Silander, K, Sharma, A, et al. (2010). A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. The Lancet, 376(9750):1393–1400.

Saleheen, D, Haycock, PC, Zhao, W, Rasheed, A, Taleb, A, Imran, A, et al. (2017). Apolipoprotein(a) isoform size, lipoprotein(a) concentration, and coronary artery disease: a mendelian randomisation analysis. The Lancet Diabetes and Endocrinology, 5(7):524–533.

Samani, NJ, Erdmann, J, Hall, AS, Hengstenberg, C, Mangino, M, Mayer, B, et al. (2007). Genomewide association analysis of coronary artery disease. The New England Journal of Medicine, 357(5), 443–453.

Schmidt, K, Noureen, A, Kronenberg, F, and Utermann, G. (2016). Structure, function, and genetics of lipoprotein(a). Journal of Lipid Research, 57(8):1339–1359.

Schmidt, AF, Pearce, LS, Wilkins, JT, Overington, JP, Hingorani AD, Casas JP. (2017). PCSK9 monoclonal antibodies for the primary and secondary prevention of cardiovascular disease. Cochrane Database of Systemic Reviews, 4:CD011748.

Scholtens, S, Smidt, N, Swertz, MA, Bakker, SJL, Dotinga, A, Vonk, JM, et al. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. International Journal of Epidemiology, 44(4):1172–1180.

Schunkert, H, König, IR, Kathiresan, S, Reilly, MP, Assimes, TL, Holm, H, et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nature Genetics, 43:333.

Schwartz, MLB, McCormick, CZ, Lazzeri, AL, Lindbuchler, DM, Hallquist, MLG, Manickam, K, et al. (2018). A model for genome-first care: returning secondary genomic findings to participants and their healthcare providers in a large research cohort. The American Journal of Human Genetics, 103(3):328–337.

Shahpori, R, Doig, C. (2010). Systematized nomenclature of medicine-clinical terms direction and its implications on critical care. Journal of Critical Care, 25(2):364.e1–364.e9.

Shankman, LS, Gomez, D, Cherepanova, OA, Salmon, M, Alencar, GF, Haskins, RM, et al. (2015). KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. Nature Medicine, 21(6):628–37.

Simon Broome Register Group. Risk of fatal coronary heart disease in familial hypercholesterolaemia. Scientific Steering Committee on behalf of the Simon Broome Register Group. (1991). British Medical Journal, 303(6807):893–896.

Soria, LF, Ludwig, EH, Clarke, HR, Vega, GL, Grundy, SM, McCarthy, BJ. (1989). Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. Proceedings of the National Academy of Sciences of the United States of America, 86(2):587–591.

Speliotes, EK, Willer, CJ, Berndt, SI, Monda, KL, Thorleifsson, G, Jackson, AU, et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics, 42:937.

Stitziel NO, Stirrups KE, Masca NG, Erdmann J, Ferrario PG, König IR, et al. (2016). Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. New England Journal of Medicine, 374(12):1134–1144.

Stone, NJ, Robinson, J, Lichtenstein, AH, Merz, CNB, Blum, CB, Eckel, RH, et al. (2013). 2013 ACC/AHA Guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults. Circulation, 129:S1–S45.

Strong, A, Ding, Q, Edmondson, AC, Millar, JS, Sachs, K V, Li, X, et al. (2012). Hepatic sortilin regulates both apolipoprotein B secretion and LDL catabolism. The Journal of Clinical Investigation, 122(8):2807–2816.

Surakka, I, Horikoshi, M, Mägi, R, Sarin, AP, Mahajan, A, Lagou, V, et al. (2015). The impact of low-frequency and rare variants on lipid levels. Nature Genetics, 47(6): 589–597.

Tada, H, Melander, O, Louie, JZ, Catanese, JJ, Rowland, CM, Devlin, JJ, et al. (2016). Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. European Heart Journal, 37(6):561–567.

Talukdar, HA, Asl, HF, Jain, RK, Ermel, R, Ruusalepp, A, Franzén, O, et al. (2016). Cross-tissue regulatory gene networks in coronary artery disease. Cell Systems, 2(3):196–208.

Tan, A, Abecasis, GR, Kang, HM. (2015). Unified representation of genetic variants. Bioinformatics, 31(13):2202–2204.

Taylor, HA, Wilson, JG, Jones, DW, Sarpong, DF, Srinivasan, A, Garrison, RJ. (2005). Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethnicity & Disease, 15(4 Suppl 6):S6–4–17.

Teslovich, TM, Musunuru, K, Smith, AV, Edmondson, AC, Stylianou, IM, Koseki, M, et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature, 466:707.

Thanassoulis, G, Peloso, GM, Pencina, MJ, Hoffmann, U, Fox, CS, Cupples, LA, et al. (2012). A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium - The Framingham Heart Study. Circulation: Cardiovascular Genetics, 5(1):113–121.

Thanassoulis, G, Campbell, CY, Owens, DS, Smith, JG, Smith, A V, Peloso, GM, et al. (2013). Genetic associations with valvular calcification and aortic stenosis. The New England Journal of Medicine, 368(6):503–512.

Thanassoulis, G, Williams, K, Altobelli, KK, Pencina, MJ, Cannon, CP, Sniderman, AD. (2016). Individualized statin benefit for determining statin eligibility in the primary prevention of cardiovascular disease. Circulation, 133(16):1574–1581.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447(7145):661–78.

Thériault, S, Lali, R, Chong, M, Velianou, JL, Natarajan, MK, and Paré, G. (2018). Polygenic contribution in individuals with early-onset coronary artery disease. Circulation: Genomic Precision Medicine, 11(1):e001849.

Torkamani, A, Wineinger, NE, Topol, EJ. (2018). The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics, 19(9):581–590.

Tsimikas, S, Hall, JL. (2012). Lipoprotein(a) as a potential causal genetic risk factor of cardiovascular disease: a rationale for increased efforts to understand its pathophysiology and develop targeted therapies. Journal of the American College of Cardiology, 60(8):716–721.

Umans-Eckenhausen, MA, Defesche, JC, Sijbrands, EJG, Scheerder, RLJM, Kastelein, JJP. (2001). Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands. The Lancet, 357(9251):165–168.

Utermann, G. (1989). The mysteries of lipoprotein(a). Science, 246(4932):904 LP–910.

van der Auwera, GA, Carneiro, MO, Hartl, C, Poplin, R, Del Angel, G, Levy-Moonshine, A, et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics, 43(1110):11.10.1–11.10.33.

van der Harst P and Verweij N. (2018). Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circulation Research, 122(3):433–443.

van der Valk, FM, Bekkering, S, Kroon, J, Yeang, C, Van Den Bossche, J, Van Buul, JD, et al. (2016). Oxidized phospholipids on Lipoprotein(a) elicit arterial wall inflammation and an inflammatory monocyte response in humans. Circulation, 134(8):611–624.

van Iperen, EPA, Sivapalaratnam, S, Holmes, M V, Hovingh, GK, Zwinderman, AH, Asselbergs, FW. (2016). Genetic analysis of emerging risk factors in coronary artery disease. Atherosclerosis, 254:35–41.

Van Staa, TP, Gulliford, M, Ng, ESW, Goldacre, B, and Smeeth, L. (2014). Prediction of cardiovascular risk using framingham, ASSIGN and QRISK2: how well do they predict individual rather than population risk? PLoS ONE, 9(10).

Vartiainen, E, Jousilahti, P, Alfthan, G, Sundvall, J, Pietinen, P, and Puska, P. (2000). Cardiovascular risk factor changes in Finland, 1972–1997. International Journal of Epidemiology, 29(1):49–56.

Verbeek, R, Sandhu, MS, Hovingh, GK, Sjouke, B, Wareham, NJ, Zwinderman, AH, et al. (2017). Lipoprotein(a) improves cardiovascular risk prediction based on established risk algorithms. Journal of the American College of Cardiology, 69(11):1513–1515.

Verweij, SL, de Ronde, MWJ, Verbeek, R, Boekholdt, SM, Planken, RN, Stroes, ESG, Pinto-Sietsma, SJ. (2018). Elevated lipoprotein(a) levels are associated with coronary artery calcium scores in asymptomatic individuals with a family history of premature atherosclerotic cardiovascular disease. Journal of Clinical Lipidology, 12(3):597–603.e1.

Vilne, B, Schunkert, H. (2018). Integrating genes affecting coronary artery disease in functional networks by multi-omics approach. Frontiers in Cardiovascular Medicine, 5:89.

Visscher, PM, Wray, NR, Zhang, Q, Sklar, P, McCarthy, MI, Brown, MA, Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. American Journal of Human Genetics, 101(1):5–22.

Wang X, Musunuru K. (2018). Confirmation of causal rs9349379-PHACTR1 expression quantitative trait locus in human-induced pluripotent stem cell endothelial cells. Circulation: Genomic and Precision Medicine, 11.

Wei, WQ, Li, X, Feng, Q, Kubo, M, Kullo, IJ, Peissig, PL. (2018). LPA variants are associated with residual cardiovascular risk in patients receiving statins. Circulation, 138(17):1839–1849.

WHO. (2016). Retrieved from: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Wijmenga, C, Zhernakova, A. (2018). The importance of cohort studies in the post-GWAS era. Nature Genetics, 50(3):322–328.

Willeit, P, Kiechl, S, Kronenberg, F, Witztum, JL, Santer, P, Mayr, M, et al. (2014). Discrimination and net reclassification of cardiovascular risk with lipoprotein(a): prospective 15-year outcomes in the Bruneck study. Journal of the American College of Cardiology, 64(9):851–860.

Willeit, P, Ridker, PM, Nestel, PJ, Simes, J, Tonkin, AM, Pedersen, TR, et al. (2018). Baseline and on-statin treatment lipoprotein(a) levels for prediction of cardiovascular events: individual patient-data meta-analysis of statin outcome trials. The Lancet, 392(10155):1311–1320.

Willer, CJ, Sanna, S, Jackson, AU, Scuteri, A, Bonnycastle, LL, Clarke, R, et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature Genetics, 40:161.

Willer, CJ, Schmidt, EM, Sengupta, S, Peloso, GM, Gustafsson, S, Kanoni, S et al. (2013). Discovery and refinement of loci associated with lipid levels. Nature Genetics, 45(11):1274–1283.

Williams, RR, Hunt, SC, Schumacher, MC, Hegele, RA, Leppert, MF, Ludwig, EH, Hopkins, PN. (1993). Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. The American Journal of Cardiology, 72(2):171–176.

Wonderling, D, Umans-Eckenhausen, MAW, Marks, D, Defesche, JC, Kastelein, JJP, Thorogood, M. (2004). Cost-Effectiveness Analysis of the genetic screening program for familial hypercholesterolemia in the Netherlands. Seminars in Vascular Medicine, 4(1):97–104.

Zdravkovic, S, Wienke, A, Pedersen, NL, Marenberg, ME, Yashin, AI, De Faire, U. (2002). Heritability of death from coronary heart disease: a 36-year follow-up of 20966 Swedish twins. Journal of Internal Medicine, 252(3):247–254.

Zhao, S, Xu, D. (2000). Oxidized lipoprotein(a) enhanced the expression of P-selectin in cultured human umbilical vein endothelial cells. Thrombosis Research, 100(6): 501–510.

Zheng, C, Khoo, C, Furtado, J, Sacks, FM. (2010). Apolipoprotein C-III and the metabolic basis for hypertriglyceridemia and the dense low-density lipoprotein phenotype. Circulation, 121(15):1722–1734.

Zhu, Z, Zheng, Z, Zhang, F, Wu, Y, Trzaskowski, M, Maier, R, et al. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nature Communications, 9(1):224.

# ACKNOWLEDGEMENTS

To put it honestly, I have to say that I am in awe of all the people whose paths I have crossed during this eventful PhD journey. You have all marked a stamp on my heart and illuminated that the route to magic stems when genuine effort, laughter and serendipity are at play.

First and foremost, I would like to express my sincere gratitude to my supervisors for trusting me to trust myself and for reigniting my curiosity for science whenever the light was dimmed or self-doubt prevailed. My deepest and heartfelt appreciation goes to professor Andres Metspalu. While it comes without saying that you are the embodiment of tenacity, showing how things get done with a clear goal in mind, good sense of humour to rely on, and an encyclopaedia-full of knowledge and loads of stories in the back pocket, you have had a profound influence on me, not only for illuminating the vastness of research and making me dig deeper in scientific reasoning, but for shining rays of light with subtle wisdom on all corners of life whenever I have run into a standstill, augmenting my determination and self-confidence. I clearly did not imagine the gracious impact you will have on me the first time I sat at your Riia 23 office, quiet as a mouse, just nodding and totally unaware of what this thing called science is. I am equally indebted to Tõnu Esko, for hopping on along the way to be my supervisor and pushing me into cold water to learn to manoeuvre my way through and learn more than a ton in the process. This I did not expect, but can only say that it was totally worth it. Thank you for all the discussions, arguments, friendly nudges and kind words – I hope this has only been the beginning of a fruitful journey to follow.

Lili Milani, you have been God-sent to me. Knowing that you are two doors down or a quick lunch date away has been so comforting that I have started to take it for granted. You are such an awesome and beautiful person. I highly value our conversations on science and beauty of life, and thank you for always pulling me back to my self-reliance and calm zone, for gently enabling me to face myself, and for providing me a cosy safe place to land whenever needed.

I am profoundly grateful to Kristi Läll and Reedik Mägi. May I stow you in my pocket and take you along the next endeavours so I can unpack you whenever I could use a splash of knowledge on all the details I will never know or when your warm glance to push ahead would be much appreciated? I am indebted to your patience and persistence for explaining however minor yet important aspect the hundredth of time that I gullably did not write down the first ninety-nine.

Elin Org, Kirsta Fischer and Katrin Männik, I look up to you! You are the embodiment of Mothers in science to me, perpetually radiating the glow peculiar to women whose heads tackle thoroughly rational questions, while hearts explore further and embrace gentleness, warmth and kindness in whatever at hand.

Heartfelt thanks to you, Urmo Võsa, Triin Laisk and Mait Metspalu. Urmo, your curiosity for research and the scientific discussions we've had have been highly thought-provoking and inspirational to me. Triin, your black humour coupled with the awesome science you seemingly effortlessly produce along the way make me question why I did not know you before? And Mait, your "just start writing" was a needed push for me to really get the ball rolling, and thanks for the conversations, jokes and laughter.

My fellow PhD students – awesome to have been on this journey with you to push through the tedious yet inescapable aspects that come with the PhD package with humour and support.

And of course, I am deeply grateful to every single person at the Estonian Genome Center, for creating an atmosphere full of sincerity, love and enthusiasm, and for more than often listening to my straightforward ranting that you always gracefully eased away.

My family's support has been invaluable. My father, for always believing in me and pushing me especially at times when I have felt the lowest. I cannot really put into words what it means to have you as an imperturbable rock in my life, allowing me to learn my own lessons, while always being there with clear wisdom, humour, gentleness and ease. My mother, for perpetually providing me motherly nourishment that I so often take for granted. My brother, for showing me how persistence and love towards what one does matters the most. My grandma, for giving me a homey pitstop in Tartu, and for always being curious of what I do and for nodding with interest yet with mild incomprehensiveness whenever I tried to give a simple overview but yet again failed. My beautiful daughter, Moon, my little fairy, for always guiding me to be more patient, more gentle, more accepting and more and more in love with life. Thank you for choosing me to be your mother. And Taavi, for enabling me to attend every conference, course and seminar, knowing that Moon is safe and sound doing all the things I will never find the persistence for or have the know-how of. And for re-fixing every extra half millimetre on my figures I find yet again need modification.

I owe my deepest gratitude to the beautiful women in my life – Gerttu Juhkam, Saskia Lillepuu, Grete Veskiväli, Liis Pihl, Este Leidmaa, Kadri Sikk, Laura Ojala, Sandra Vokk – for relentlessly and full-heartedly walking your paths of life, radiating enthusiasm and authenticity, and providing endless comfort and laughter on the peculiarities of life. You keep me sane and are my biggest inspiration!

And last but not least, Evelyn Tuul for introducing me to the world of yoga that has deeply been seeded in me by your gentle guidance, Maarja Värv, for always being up for a pleasurable and sometimes just simply lazy pastime with our daughters, and Mattias Rennel, Liis Nigul and Kaisa Thomson, for enabling to rid the routine whenever needed and for enabling the mere simplicity to just be.

I thank you and I love you all!

# PUBLICATIONS

# CURRICULUM VITAE

**Name**: Maris Alver
**Date of birth**: July 28, 1986
**Contact**: Estonian Genome Center, Institute of Genomics, University of Tartu
Riia 23b, 51010, Tartu
**E-mail**: maris.alver@ut.ee

**Education**

| | |
|---|---|
| 2012–… | PhD in Molecular Biomedicine, Institute of Molecular and Cell Biology, University of Tartu, Estonia |
| 2010–2012 | Master's Degree in Biomedicine (*cum laude*), Institute of Molecular and Cell Biology, University of Tartu, Estonia |
| 2005–2009 | Bachelor's Degree in Gene Technology, Institute of Molecular and Cell Biology, University of Tartu, Estonia |

**Work Experience**

| | |
|---|---|
| 2012–… | Specialist, Estonian Genome Center, Institute of Genomics, University of Tartu, |
| Estonia | |
| 2009–2009 | Lab Technician, Synthetic Biology Lab, Institute of Technology, University of Tartu, Estonia |
| 2008–2009 | Lab Technician, Biotechnology department, Institute of Molecular and Cell Biology, University of Tartu, Estonia |

**Publications**

Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, ..., **Alver M**, et al. (2019). Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identify new loci associated with serum lipids. Nat Genet. 51(4):636–648.

Patel RS, Schmidt AF, Tragante V, McCubrey RO, Holmes MV, Howe LJ, …, **Alver M**, et al. (2019). Association of chromosome 9p21 with subsequent coronary heart disease events: a GENIUS-CHD study of individual participant data. Circ Genom Precis Med. Mar 21.

Patel RS, Tragante V, Schmidt AF, McCubrey RO, Holmes MV, Howe LJ, …, **Alver M**, et al. (2019). Subsequent event risk in individuals with established coronary heart disease: design and rationale of the GENIUS-CHD Consortium. Circ Genom Precis Med. Mar 21.

de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, …, **Alver M**, et al. (2019). Multi-ancestry genome-wide association study of lipid levels incorporating gene-alcohol intractions. Am J Epidemiol. Jan 29.

Kilpeläinen TO, Bentley AR, Noordam R, Sung YJ, Schwander K, Winkler TW, Jakupovic H, Chasman DI, …, **Alver M**, et al. (2019). Multi-ancestry

study of blood lipid levels identifies four loci interacting with physical activity. Nat Commun. Jan 22;10(1):376.

Saar A, Läll K, **Alver M**, Marandi T, Ainla T, Eha J, Metspalu A, Fischer K. (2019). Estimating the performance of three cardiovascular risk scores in a high risk European country. J Epidemiol Community Health. Jan 11.

**Alver M**, Palover M, Saar A, Läll K, Zekavat SM, Tõnisson N, Leitsalu L, Reigo A, et al. (2018). Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. Genet Med. Oct 1.

Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, …, **Alver M**, et al. (2018). Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nat Commun. Aug 23;9(1):3493.

Lee J, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, …, **Alver M**, et al. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. Nat Genet. Aug;50(8):1112–1121.

Zekavat SM, Ruotsalainen S, Handsaker RE, **Alver M**, Bloom J, Poterba T, Seed C, Ernst J, et al. (2018). Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. Nat Commun. Jul 4;9(1):2606.

Feitosa M, Kraja AT, Chasman DI, Sung YI, Winkler TW, Ntalla I, Guo X, Franceshini N, …, **Alver M**, et al. (2018). Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. Jun 18;13(6):e0198166.

Sung YJ, Winkler TW, de Las Fuentes L, Bentley AR, Brown MR, Kraja AT, Schwander K, Ntalla I, …, **Alver M**, et al. (2018). A Large-Scale Multi-Ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. Am J Hum Genet. Mar 1;102(3):375–400.

Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NG, Jansen H, Kanoni S, Nelson CP, …, **Alver M**, et al. (2017). Systemic Evaluation of Pleiotropy Identifies 6 Further Loci Associated with Coronary disease. J Am Coll Cardiol. Feb 21; 69(7):823–836.

Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, Ntalla I, Surendran P, …, **Alver M**, et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. Mar;49(3):403–415.

Loley C, **Alver M**, Assimes TL, Bjonnes A, Goel A, Gustafsson S, Heresniemi J, Hopewell JC, …, et al. (2016). No association of coronary artery disease with X-chromosomal variants in comprehensive international meta-analysis. Sci Rep. Oct 12;6:35278.

Elosua R, Lluis-Ganella C, Subirana I, Havulinna A, Läll K, Lucas G, Saylos-Baizeras S, Pietilä A, **Alver M**, et al. (2016). Cardiovascular risk factors and

ischemic heart disease: is the confluence of risk factors greater than the parts. A genetic approach. Circ Cardiovasc Genet. Jun;9(3):279–86.
Stitziel, N, Stirrups KE, Masca NG, Erdmann J, Ferrario PG, König IR, Weeke PE, Webb TR, …, **Alver M**, et al. (2016). Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. N Engl J Med. Mar 24;374(12):1134–44.
Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, …, **Alver M**, et al. (2015). A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. Oct;47(10):1121–30.
Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, …, **Alver M**, et al. (2014). Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Hum Mol Genet. Aug 15;23(16):4420–32.

**Supervision**

Anett Tähiste — Master's Degree, Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Estonia. Identifying familial hypercholesterolemia associated regulatory variants based on the whole-genome sequencing data of the Estonian Genome Center of the University of Tartu

Anett Tähiste — Bachelor's Degree, Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Estonia. Risk estimation of coronary artery disease based on genetic markers in the Estonian Genome Center cohort

**Fellowship and awards**

2019 Graduate School in Biomedicine and Biotechnology Fellowship Biology of Genomes, May 6–11, Cold Spring Harbor Laboratory, New York, United States

2019 ePerMed Mobility Fellowship to Center for Integrative Genomics, University of Lausanne, funding from H2020 grant, European Commission, February 17–23, Lausanne, Switzerland

2018 ePerMed Mobility Fellowship to Center for Integrative Genomics, University of Lausanne, funding from H2020 grant, European Commission, September 21 - October 3, Lausanne, Switzerland

2018 Graduate School in Biomedicine and Biotechnology Fellowship WGC Advanced Course: Genetic Analysis of Mendelian and Complex Diseases, July 18–24, Hinxton, United Kingdom

2018 Graduate School in Biomedicine and Biotechnology Fellowship Biology of Genomes, May 7–12, Cold Spring Harbor Laboratory, New York, United States

2017 Dora Plus Mobility Fellowship, Archimedes Foundation, American Society of Human Genetics, October 17–21, Orlando, Florida, United States

2017 Dora Plus Mobility Fellowship, Archimedes Foundation, Genomics of Rare Disease, April 5–7, Hinxton, United Kingdom

2016 ePerMed Mobility Fellowship to Institute for Molecular Medicine Finland, funding from H2020 grant, European Commission, January 28 - February 26, Helsinki, Finland

2013 Dora Plus Mobility Fellowship, Archimedes Foundation, BiomarCaRE Statistics Workshop, June 2–5, Hamburg, Germany

2013 Graduate School in Biomedicine and Technology Mobility Fellowship, Logical Reasoning in Human Genetics 2013, January 13–18, Helsinki, Finland

2011 Kristjan Jaak Mobility Fellowship, Estonian Research Council, European Society of Human Genetics conference, May 28–31, Amsterdam, The Netherlands

# ELULOOKIRJELDUS

**Nimi**: Maris Alver
**Sünniaeg**: 28. juuli, 1986
**Aadress**: Eesti Geenivaramu, Genoomika Instituut, Tartu Ülikool
Riia 23b, 51010, Tartu
**E-post**: maris.alver@ut.ee

**Haridus**

| | |
|---|---|
| 2012–… | Doktoriõpe, molekulaar- ja rakubioloogia, loodus- ja täppisteaduste valdkond, Tartu Ülikool |
| 2010–2012 | Magistriõpe (*cum laude*), biomeditsiin, loodus- ja täppisteaduste valdkond, Tartu Ülikool |
| 2005–2009 | Bakalaureuseõpe, geenitehnoloogia, loodus- ja täppisteaduse valdkond, Tartu Ülikool |

**Teenistuskäik**

| | |
|---|---|
| 2012–… | spetsialist, Eesti Geenivaramu, Genoomika instituut, Tartu Ülikool |
| 2009–2009 | laborant, sünteetilise bioloogia labor, Tehnoloogiainstituut, Tartu Ülikool |
| 2008–2009 | laborant, biotehnoloogia õppetool, Molekulaar- ja rakubioloogia instituut, Tartu Ülikool |

**Teaduspublikatsioonid**

Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, ..., **Alver M**, et al. (2019). Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identify new loci associated with serum lipids. Nat Genet. 51(4):636–648.

Patel RS, Schmidt AF, Tragante V, McCubrey RO, Holmes MV, Howe LJ, …, **Alver M**, et al. (2019). Association of chromosome 9p21 with subsequent coronary heart disease events: a GENIUS-CHD study of individual participant data. Circ Genom Precis Med. Mar 21.

Patel RS, Tragante V, Schmidt AF, McCubrey RO, Holmes MV, Howe LJ, …, **Alver M**, et al. (2019). Subsequent event risk in individuals with established coronary heart disease: design and rationale of the GENIUS-CHD Consortium. Circ Genom Precis Med. Mar 21.

de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, …, **Alver M**, et al. (2019). Multi-ancestry genome-wide association study of lipid levels incorporating gene-alcohol intractions. Am J Epidemiol. Jan 29.

Kilpeläinen TO, Bentley AR, Noordam R, Sung YJ, Schwander K, Winkler TW, Jakupovic H, Chasman DI, …, **Alver M**, et al. (2019). Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. Nat Commun. Jan 22;10(1):376

Saar A, Läll K, **Alver M**, Marandi T, Ainla T, Eha J, Metspalu A, Fischer K. (2019). Estimating the performance of three cardiovascular risk scores in a high risk European country. J Epidemiol Community Health. Jan 11.

**Alver M**, Palover M, Saar A, Läll K, Zekavat SM, Tõnisson N, Leitsalu L, Reigo A, et al. (2018). Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. Genet Med. Oct 1.

Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, …, **Alver M**, et al. (2018). Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nat Commun. Aug 23;9(1):3493.

Lee J, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, …, **Alver M**, et al. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. Nat Genet. Aug;50(8):1112–1121.

Zekavat SM, Ruotsalainen S, Handsaker RE, **Alver M**, Bloom J, Poterba T, Seed C, Ernst J, et al. (2018). Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. Nat Commun. Jul 4;9(1):2606.

Feitosa M, Kraja AT, Chasman DI, Sung YI, Winkler TW, Ntalla I, Guo X, Franceshini N, …, **Alver M**, et al. (2018). Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. Jun 18;13(6):e0198166.

Sung YJ, Winkler TW, de Las Fuentes L, Bentley AR, Brown MR, Kraja AT, Schwander K, Ntalla I, …, **Alver M**, et al. (2018). A Large-Scale Multi-Ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. Am J Hum Genet. Mar 1;102(3):375–400.

Webb TR, Erdmann J, Stirrups KE, Stitziel NO, Masca NG, Jansen H, Kanoni S, Nelson CP, …, **Alver M**, et al. (2017). Systemic Evaluation of Pleiotropy Identifies 6 Further Loci Associated with Coronary disease. J Am Coll Cardiol. Feb 21; 69(7):823–836.

Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, Ntalla I, Surendran P, …, **Alver M**, et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. Mar;49(3):403–415.

Loley C, **Alver M**, Assimes TL, Bjonnes A, Goel A, Gustafsson S, Heresniemi J, Hopewell JC, …, et al. (2016). No association of coronary artery disease with X-chromosomal variants in comprehensive international meta-analysis. Sci Rep. Oct 12;6:35278.

Elosua R, Lluis-Ganella C, Subirana I, Havulinna A, Läll K, Lucas G, Saylos-Baizeras S, Pietilä A, **Alver M**, et al. (2016). Cardiovascular risk factors and ischemic heart disease: is the confluence of risk factors greater than the parts. A genetic approach**.** Circ Cardiovasc Genet. Jun;9(3):279–86.

Stitziel, N, Stirrups KE, Masca NG, Erdmann J, Ferrario PG, König IR, Weeke PE, Webb TR, …, **Alver M**, et al. (2016). Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. N Engl J Med. Mar 24;374(12):1134–44.

Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, …, **Alver M**, et al. (2015). A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. Oct;47(10):1121–30.

Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, …, **Alver M**, et al. (2014). Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Hum Mol Genet. Aug 15;23(16):4420–32.

**Juhendatud väitekirjad**

Anett Tähiste magistritöö, biomeditsiini õppekava "Perekondliku hüperkolesteroleemia seotud regulatoorsete variantide identifitseerimine"

Anett Tähiste bakalaureusetöö, geenitehnoloogia õppekava "Geneetilistel markeritel põhinev südame isheemiatõve riski hindamine Eesti Geenivaramu kohordis"

**Stipendiumid**

2019 Biomeditsiini ja biotehnoloogia doktorikool, konverents *Biology of Genomes*, 6.–11. mai, Cold Spring Harbor Laboratory, New York, Ameerika Ühendriigid

2019 ePerMed õpirände toetus Lausanne ülikooli Integratiivse genoomika keskuse külastuseks, 17.–23. veebruar, Lausanne, Šveits

2018 ePerMed õpirände toetus Lausanne ülikooli Integratiivse genoomika keskuse külastuseks, 21. September – 3. Oktoober, Lausanne, Šveits

2018 Biomeditsiini ja biotehnoloogia doktorikool, kursus *Genetic Analysis of Mendelian and Complex Diseases*, 18.–24. juuli, Hinxton, Ühendkuningriigid

2018 Biomeditsiini ja biotehnoloogia doktorikool, konverents *Biology of Genomes*, 7.–12. mai, Cold Spring Harbor Laboratory, New York, Ameerika Ühendriigid

2017 Dora Pluss lühiajalise õpiränne, Ameerika Inimgeneetika ühingu aastakonverents, 17.–21. oktoober, Orlando, Florida, Ameerika Ühendriigid

2017 Dora Pluss lühiajalise õpiränne, konverents *Genomics of Rare Disease*, 5.–7. aprill, Hinxton, Ühendkuningriigid

2016 ePerMed õpirände toetus Soome molekulaarse meditsiini instituudi külastuseks, 28. jaanuar – 26. veebruar, Helsingi, Soome

2013 Dora Pluss lühiajalise õpiränne kursus *BiomarCaRE Statistics Workshop*, 2.–5. juuni, Hamburg, Germany

2013 Biomeditsiini ja biotehnoloogia doktorikool, kursus *Logical Reasoning in Human Genetics 2013*, 13.–18. jaanuar, Helsingi, Soome

2011 Kristjan Jaagu välissõidu stipendium, Euroopa Inimgeneetika Ühingu aastakonverents, 28.–31. mai, Amsterdam, Holland

# 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. **Kaasik, Krista**. 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 Usküla**. 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.