DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
320

SILVA KASELA

Genetic regulation of gene expression: detection of tissue- and cell type-specific effects

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
320

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
320

SILVA KASELA

Genetic regulation of gene expression: detection of tissue- and cell type-specific effects

UNIVERSITY OF TARTU
Press

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

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Molecular Biomedicine on June 12, 2017 by the Council of the Institute of Molecular Cell Biology, University of Tartu.

Supervisors: Lili Milani, PhD
Estonian Genome Center, University of Tartu, Tartu, Estonia

Krista Fischer, PhD
Estonian Genome Center, University of Tartu, Tartu, Estonia

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

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

Opponent: Tuuli Lappalainen, PhD
New York Genome Center, New York, United States of America
Department of Systems Biology, Columbia University, New York, United States of America

Commencement:
Room No. 105, 23B Riia St., Tartu, on August 22, 2017, at 2:15 pm.

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

This research was funded by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), Centre of Excellence for Genomics and Translational Medicine (GENTRANSMED), the Estonian Research Council Grant IUT20-60, EU H2020 grant no. 692145 (TWINNING), the Ministry of Education and Research, and the Archimedes Foundation.

ISSN 1024-6479
ISBN 978-9949-77-482-1 (print)
ISBN 978-9949-77-483-8 (pdf)

European Union
European Regional
Development Fund

Investing
in your future

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

I Bonder MJ*, **Kasela S***, Kals M, Tamm R, Lokk K, Barragan I, Buurman WA, Deelen P, Greve JW, Ivanov M, Rensen SS, van Vliet-Ostaptchouk JV, Wolfs MG, Fu J, Hofker MH, Wijmenga C, Zhernakova A, Ingelman-Sundberg M, Franke L*, Milani L*. 2014. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genomics*. 15:860.

II **Kasela S**, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K, Esko E, Westra HJ, Fairfax BP, Makino S, Knight JC, Franke L, Metspalu A, Peterson P, Milani L. 2017. Pathogenic Implications for Autoimmune Mechanisms Derived by Comparative eQTL Analysis of $CD4^+$ Versus $CD8^+$ T cell. *PLoS Genet.* 13(3):e1006643.

III Võsa U, Esko T, **Kasela S**, Annilo T. 2015. Altered gene expression associated with microRNA binding site polymorphisms. *PLoS One.* 10(10):e0141351.

My contributions to the listed publications were as follows:

Ref. I Performed differential expression and differential methylation analyses, performed correlation analysis between expression and methylation values, prepared the figures and participated in writing the manuscript.

Ref. II Mined the data, performed all of the analyses, prepared the figures, and drafted the manuscript.

Ref. III Contributed to the data analysis and revised the manuscript.

LIST OF ABBREVIATIONS

GWAS	Genome-wide association study
eQTL	Expression quantitative trait loci
GTEx	Genotype-Tissue Expression project
SNP	Single-nucleotide polymorphism
LD	Linkage disequilibrium
ChIP-seq	Chromatin immunoprecipitation sequencing
RNA-seq	RNA-sequencing
ENCODE	Encyclopedia of DNA Elements
miRNA	microRNA
UTR	Untranslated region
AU	Adenylate-uridylate
TSS	Transcription start site
LDL-C	Low-density lipoprotein cholesterol
MI	Myocardial infarction
Kb	Kilobase, 1,000 base pairs
VLDL	Very low-density lipoprotein
CI	Confidence interval
Mb	Megabase, 1,000,000 base pairs
PC	Principal component
FDR	False discovery rate
EM	Expectation-maximization
BF	Bayes factor
GRS	Genetic risk score
TRS	Transcriptional risk scores
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
SAT	Subcutaneous adipose tissue
VAT	Visceral adipose tissue
eQTM	Expression quantitative trait methylation
meQTL	Methylation quantitative trait loci
ADME	Absorption, distribution, metabolism and excretion of drugs
T1D	Type 1 diabetes
RA	Rheumatoid arthritis
PBMC	Peripheral blood mononuclear cells
MRE	miRNA response element

INTRODUCTION

Tremendous progress in the field of genomics has created a basis for large-scale characterization of the role of genetics in human complex traits and diseases. Linking genetic variants with the variability in traits using genome-wide association studies (GWAS) has enhanced our understanding of the heritability and genetic architecture of human traits. However, a majority of the trait-associated variants fall in noncoding regions of the human genome, providing no direct information about the functional consequences of the genetic variants. This limits our ability to discern key genes and molecular mechanisms affected by the associated variants.

Gene expression gives rise to observable phenotypes latent in the underlying genetic code. The regulation of gene expression is a dynamic process involving a diverse set of mechanisms in time- and context-specific manners. Thereby, genetic variants play an important role in various events, such as transcription factor binding efficiency at promoter or enhancer regions, alternative splicing, microRNA activity in post-transcriptional regulation, and post-translational modifications. Considering genetic variants as modulators of gene expression levels, the intermediate quantitative trait between DNA variants and human traits or diseases, offers new instruments to leverage the knowledge gathered by genome-wide association studies by pinpointing the crucial links and refining our understanding about gene regulation in general. By learning how genetic variants change gene expression in healthy individuals, we can gain insights into how diseases develop in people. Therefore, expression quantitative trait loci (eQTL) mapping holds promise to improve our understanding of how to read the genetic code and translate it to the benefit of human health.

As gene expression is differentially regulated in different tissues and environmental or cellular contexts, the comprehensive profiling of the effects of genetic variants on gene expression in a broad range of tissues and contexts is necessary to obtain a complete picture of the complex pattern of gene regulation. Over the last years, the maps of eQTLs in humans have been extensively composed and improved. As an example of one ambitious initiative, the National Institutes of Health launched the Genotype-Tissue Expression (GTEx) project in September 2010, which aims to profile the association between genetic variants and gene expression levels in more than 50 tissue types from 900 post-mortem donors.

This thesis focuses on characterizing the different aspects of genetic regulation of gene expression. In the first part of the thesis, I give an overview of eQTLs from different perspectives based on the literature: the rationale behind eQTLs; the nature, properties and importance of eQTLs; the analytical framework for eQTL mapping; and the future prospects of eQTLs to improve human health. In the second part, I describe the investigation of the genetic regulation of gene expression in liver tissue and purified T cells, aiming to study normal variation in gene expression and to detect tissue- and cell type-specific effects. Finally, I discuss and analyze the impact of polymorphisms in microRNA binding sites in regulating gene expression levels.

1. REVIEW OF THE LITERATURE

1.1. Genome-wide association studies

Dissecting the role of common variation in human traits and complex diseases have been the key questions in genomics for the last decade. GWAS have proven to be a well-suited experimental design to investigate the genetic architecture and heritability of complex human traits and to provide novel biological mechanisms (Bush and Moore, 2012; Visscher et al., 2012). In GWAS, the distribution of the genotypes at each common single-nucleotide polymorphism (SNP) in individuals with a phenotype of interest (cases) and in the general population (controls) are compared, resulting in an estimated effect size of the main additive effect of a SNP for the given phenotype.

Since the first GWAS in 2005, over two thousand studies have been conducted and tens of thousands of genetic variants linked to complex diseases and traits have been found (Welter et al., 2014, GWAS Catalog by the National Human Genome Research Institute and the European Bioinformatics Institute accessed on 27/02/2017), highlighting the important role of genetics in common diseases. However, GWAS hits explain a relatively moderate portion of the phenotypic variance (Visscher et al., 2012) and it remains challenging to pinpoint the precise disease-causal variants and genes (Edwards et al., 2013).

In general, the numerous identified associations mark the region of SNPs in strong linkage disequilibrium (LD) that influences the risk for a certain disease or phenotypic expression. Typically, the susceptibility region, which is expected to include the biologically relevant variant, is substantial in length (Schaub et al., 2012). Moreover, a majority of the trait-associated variants are either localized in intergenic regions or introns (Hindorff et al., 2009). Thus, instead of directly affecting the sequence of proteins via non-synonymous substitutions, the functional effects of SNPs on phenotypes may be mediated by other mechanisms, such as regulation of gene expression levels. Indeed, the success of mapping functional elements in the human genome have revealed that GWAS SNPs are enriched in or near regulatory regions outside of protein-coding genes (Dunham et al., 2012; Schaub et al., 2012), such as chromatin immunoprecipitation sequencing (ChIP-seq) peaks and DNase I hypersensitive sites, and are more likely to influence gene expression (Nicolae et al., 2010). Furthermore, their molecular consequences are likely to be restricted to specific tissue or cell type and to particular periods of development (Fu et al., 2012; Maurano et al., 2012; Trynka et al., 2013).

To translate the information gathered by GWAS into clinical advances, the next challenge beyond finding association signals is defining the molecular mechanisms whereby the risk SNPs modulate traits and diseases (McCarthy et al., 2008). Thus, to unravel the causal chain, eQTL studies hold promise to characterize the downstream targets of risk SNPs through their effect on gene expression in specific tissues or cell types most relevant to the phenotype of interest.

1.2. Genetic regulation of gene expression

Gene expression forms a fundamental bridge between genotypes and observable traits through the synthesis of functional gene products (proteins, functional RNA) from the genetic code stored in the DNA. The variety of elements and mechanisms that induce or repress the expression of a gene act in a concerted manner (Phillips, 2008; White and Sharrocks, 2010). These processes include epigenetic modifications, such as DNA methylation, histone acetylation, and other histone modifications, to the genome to modulate the accessibility of DNA to transcription factors, binding of transcription factors to specific DNA elements to activate transcription, and mechanisms that modulate the translation of mRNA transcripts into proteins (Cooper, 2000; Lelli et al., 2012). Of note, genetic variation plays a critical role in various events regulating gene expression (Knight, 2014).

Importantly, the control of gene expression is a dynamic process altered in response to the changing cellular context during given stage of development and environmental stimulus (Jaenisch and Bird, 2003; White and Sharrocks, 2010). In general, gene regulation defines the function of a cell and in case of mis-regulation, if improper transcripts are expressed at a wrong time or location, can lead to diseases, such as cancer, autoimmune and inflammatory diseases, developmental disorders, diabetes and cardiovascular diseases (Lee and Young, 2013).

Gene expression levels can be considered as a quantitative trait like height and lipid levels. As a heritable intermediate phenotype between DNA variants and complex traits (Price et al., 2011; Wright et al., 2014; Lloyd-Jones et al., 2017), dissecting the genetic basis of gene expression in tissues and cell types with biological relevance to the trait of interest provides characterization of the mechanistic link between the genetic variant and the trait (Dermitzakis, 2008). For this reason, it is appealing to map functional genetic variants (both coding and regulatory) that affect gene expression, to advance the understanding and interpretation of the genetic architecture of natural and disease-associated variation in gene expression.

Large scale eQTL mapping studies were proposed in 2001 (Jansen and Nap, 2001). The first genome-wide mapping of expression levels measured with microarrays was carried out in 2002 in a cross between two strains of yeast (Brem et al., 2002), and the first large eQTL study in humans was reported in 2004 by linkage analysis in 14 large families (Morley et al., 2004). Starting from 2010, the advances in RNA-sequencing (RNA-seq) technologies have enabled higher-resolution analyses, including transcripts not present on commercial microarrays and additionally, allowing for analyses of alternative splicing and allele-specific expression to determine *cis*-acting variants *per se* (Montgomery et al., 2010; Pickrell et al., 2010). RNA-seq, combined with the advances in whole-genome sequencing, provides a deeper understanding of the functional variants (Lappalainen et al., 2013). As the effect of the functional variants depends on the dynamic epigenomic landscape in which the variants

exert their effect (Knight, 2014), it is highly informative to accompany eQTLs with epigenomic maps in different human cell and tissue types. Thus, initiatives such as the Encyclopedia of DNA Elements (ENCODE Project Consortium, 2004, 2012; Birney et al., 2007) and National Institutes of Health Roadmap Epigenomics Mapping Consortium (Bernstein et al., 2010; Roadmap Epigenomics Consortium et al., 2015) are essential to characterize the genomic properties of eQTLs.

Today, the analysis of eQTLs has become a standard approach to provide functional annotation to GWAS hits. Also, it extends the conception of the genetics of gene expression and comprehension of the nature, properties, and importance of eQTLs in general.

1.2.1. The nature of eQTLs

eQTLs are classified into two groups based on either the distance between the genetic variant and the gene (local or distal eQTLs) or the underlying molecular nature through which expression is affected (*cis*- or *trans*-eQTLs) (Rockman and Kruglyak, 2006).

Usually, local eQTLs act in *cis* and influence the expression levels of a gene in an allele-specific manner, resulting in an allelic imbalance (Fig. 1A; Albert and Kruglyak, 2015), i.e. differential expression of the two alleles of a heterozygous individual. On the other hand, local eQTLs that act in *trans* modify both of the alleles of a close by gene (Albert and Kruglyak, 2015). Two typical scenarios of the *trans* association are a polymorphism in a nearby gene and an autoregulative feedback loop (Rockman and Kruglyak, 2006). Firstly, a polymorphism in a nearby gene regulates the gene of interest due to the close location of regulators and their targets in the genome. Secondly, a gene is regulated by its own product triggered by a polymorphism in the coding sequence.

Distal eQTLs typically act in *trans* through the downstream effects of coding variants in the regulatory genes or local eQTLs of such genes (Fig. 1B; Albert and Kruglyak, 2015). The phenomena of *trans*-eQTL hotspots, loci that can affect the expression of hundreds of distal genes, have been reported both in model organisms and human studies (Breitling et al., 2008). *Cis*-mediated mechanisms and co-regulatory landscapes are considered as two possible explanations for *trans*-eQTL hotspots. According to the first proposal, the SNP has both *cis* and *trans* effects, therefore *trans* regulation is mediated by the expression of a gene with a *cis*-eQTL effect serving as a master regulator for a number of distal genes (Pierce et al., 2014; Yao et al., 2017). The second proposal suggests that *trans*-eQTL hotspots regulate genes of functionally related groups, e.g. genes that encode proteins that interact directly, genes from the same biological pathways, or genes regulated by the same transcription factors (Brynedal et al., 2017).

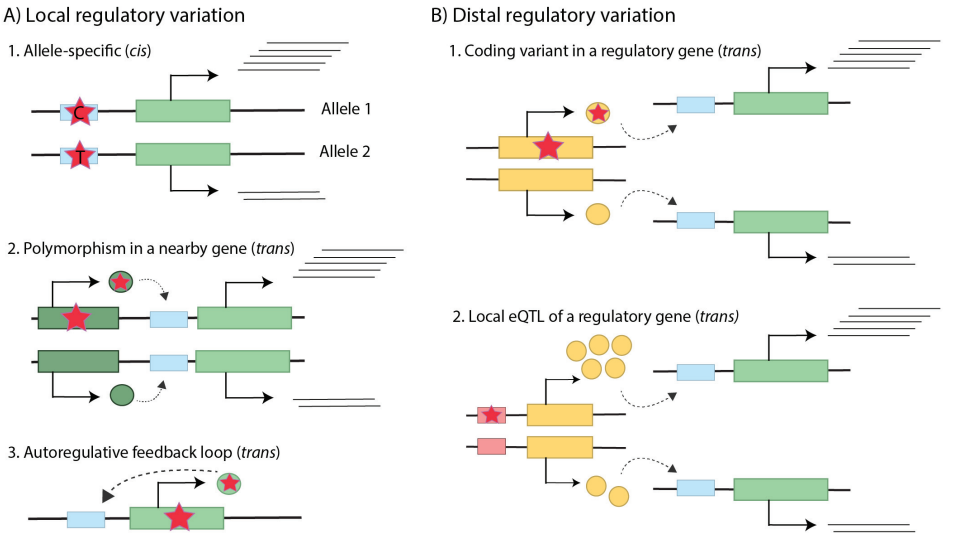

Figure 1. Local and distal eQTLs. Local and distal effects of regulatory variants (red stars) are illustrated for different scenarios highlighting the roles of coding regions (bars with arrows), promoters (smaller bars), mRNA transcripts (black lines), gene products/proteins (circles). Affected genes and promoter regions are denoted by pale green and pale blue bars, respectively. (A) Local regulatory variants are found in the proximity of the target gene. Local eQTLs can act both in *cis* and *trans*. *Cis*-regulatory variants affect gene expression in an allele-specific manner (A1). For example, a regulatory mutation in the promoter sequence can influence the binding of a transcription factor and hence the abundance of mRNA transcripts. In the illustrated example, a transcription factor binds with stronger affinity to allele 1 than allele 2, resulting in differential expression of the two alleles of a heterozygous individual. A c*is*-regulatory variant that acts in *trans* can be a SNP in a nearby gene that alter the encoded protein, which in turn affects the expression of a target gene (A2), or a SNP in the coding sequence of a gene, which regulates its own expression by binding to its promoter sequence (A3). (B) Distal regulatory variants are found farther from the target gene or on a different chromosome. Distal eQTLs act in *trans*. A distal eQTL can be a SNP in the coding region of a gene that encodes a product that acts as a transcription factor for the target gene (B1). In the illustrated example, an altered transcription factor binds with stronger affinity to the promotor region on both alleles, hence resulting in more mRNA transcripts compared to the original transcription factor, which binds poorly to both alleles of the target gene. A distal eQTL can also be a local eQTL of a regulatory gene (B2), influencing the abundance of the gene's product which affects a downstream target gene. In the illustrated example, the abundance of the protein positively affects the expression of the target gene. Adapted from Albert and Kruglyak (2015).

1.2.1.1. The molecular basis of eQTLs

The growing availability of whole genome sequences has provided a deeper characterization of regulatory variation causing changes in the mechanism regulating gene expression levels. *Cis*-eQTLs have been linked to different molecular mechanisms and processes prior to, during, and after transcription, which can, in turn, affect the expression of other genes as *trans*-eQTLs.

Pre-transcriptional regulatory variation alters transcription rate uniformly over all isoforms of a gene (Gaffney, 2013). The key processes prior to transcription are transcription factor binding, histone modifications, enhancer activity, and DNA methylation. Of these, most support has been gained for genetically driven variation in transcription factor binding as likely the strongest contributor to gene expression regulation, interacting with or even affecting other regulatory mechanisms, such as DNA methylation and chromatin accessibility (Pai et al., 2015).

In contrast to the pre-transcriptional regulatory variation, post-transcriptional variants alter the relative abundances of a set of isoforms (Gaffney, 2013). Splicing, the removal of transcribed introns from the pre-mRNA, is one of the several important mechanisms taking place co-transcriptionally and after transcription. The genetic basis of mRNA splicing variation is mainly localized in the proximity of or within the targeted exon (Fraser and Xie, 2009; Pickrell et al., 2010). Interestingly, most of the genetic variants that affect splicing have no effect on gene expression levels, on the contrary, they possibly influence the protein function by affecting the coding sequence (Li et al., 2016).

Following RNA processing, steady-state mRNA levels are determined by the rate of mRNA degradation modulated by mRNA stability and small regulatory RNAs, including microRNAs (miRNA) (Garneau et al., 2007). Genetic variation affects post-transcriptional mechanisms by influencing the general RNA decay, alternative polyadenylation and miRNA binding (Pai et al., 2015). The variants associated with the aforementioned processes are found predominantly in the untranslated regions (UTRs) at the 3' end of an mRNA transcript, specifically at the regulatory elements associated with mRNA-stability determinants, such as adenylate-uridylate (AU)-rich elements and miRNA binding sites (Lu and Clark, 2012; Pai et al., 2012; Yoon et al., 2012).

All these mechanisms illustrate the molecular complexity and combinatorial nature of gene regulation (Fig. 2). However, the exact extent of contributions from specific pre-, co- and post-transcriptional processes currently remains unclear.

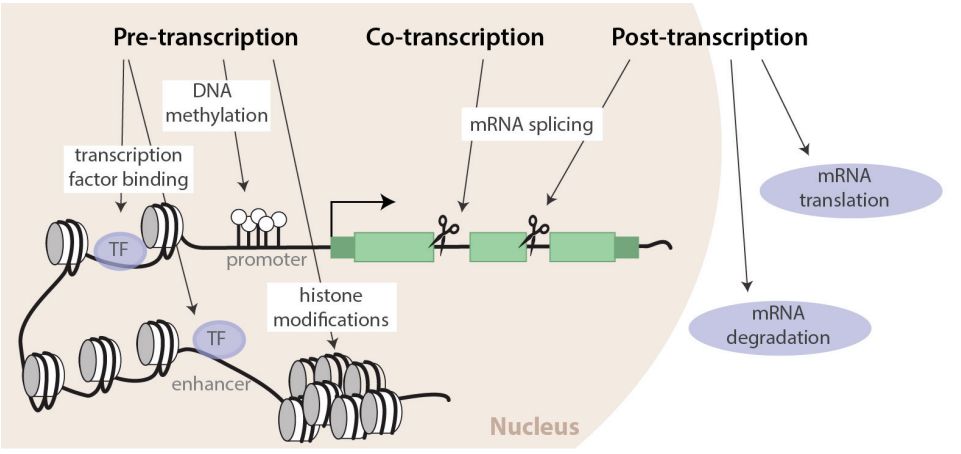

Figure 2. Mechanisms by which an eQTL can regulate gene expression from pre-transcription to post-transcription. Prior to transcription, an eQTL can impact transcription factor binding to promotor or enhancer regions and epigenetic alterations, such as DNA methylation and histone modifications. For example, heavy methylation in gene promoter regions is known to repress gene transcription and histone modifications influence the accessibility of DNA regions, thus activating or repressing transcription. During and after transcription, an eQTL localized in the vicinity of the targeted exon can impact the inclusion of the exon in the processed mRNA transcript, thus resulting in alternatively spliced isoforms. After transcription, eQTLs have a role both in mRNA degradation and translation. For example, by affecting miRNA binding efficiency, eQTLs may indirectly repress the expression of its target gene via the destabilization of mRNA or translation blockage of mRNA. TF – transcription factor. Adapted from Pai et al. (2015).

1.2.2. Properties of eQTLs

Tremendous progress in the field of eQTL mapping has led to characterizing the effect of genetic variants on gene expression in a diverse range of human populations and context. eQTLs are extremely widespread throughout the human genome and highly conserved across populations of different ancestry with similar allelic direction and effect size (Raj et al., 2014; Stranger et al., 2012). It is expected that a majority of the eQTLs do not dramatically impact organism-level phenotype, however, they are essential for understanding the potential function of the non-coding variants on gene expression (Battle and Montgomery, 2014; Lappalainen, 2015).

To date, eQTL maps have been composed for several easily accessible tissues such as whole blood (Westra et al., 2013; Yao et al., 2017), liver (Schadt et al., 2008; Innocenti et al., 2011; Fu et al., 2012), subcutaneous adipose tissue (Fu et al., 2012; Grundberg et al., 2012), visceral adipose tissue (Fu et al., 2012), and cell types such as lymphoblastoid cell-lines (Grundberg et al., 2012; Lappalainen et al., 2013; Bryois et al., 2014), B cells (Fairfax et al., 2012), T cells (Ferraro et

al., 2014; Raj et al., 2014; Chen et al., 2016), monocytes (Zeller et al., 2010; Fairfax et al., 2012; Raj et al., 2014; Chen et al., 2016), neutrophils (Andiappan et al., 2015; Naranbhai et al., 2015; Chen et al., 2016). As a part of the GTEx project, 44 tissues and cell lines from post-mortem samples have been assessed (GTEx Consortium, 2015). Each of these studies has made a notable contribution to describing the complex landscape of genetic regulation of gene expression and determining the extent of tissue- and condition-specific effects.

1.2.2.1. Effect sizes and functional architecture

Cis-eQTLs are widespread in the human genome. By the most recent estimate based on a broad range of tissues, 88% of all annotated protein-coding genes are regulated by local genetic variants and the number of discoveries has not reached the plateau yet (Aguet et al., 2016). Moreover, it is consistently estimated that one-third of the genes have conditionally independent effects in at least one tissue type (Aguet et al., 2016; Brown et al., 2013). The number of discovered and replicable *trans*-eQTLs has been low compared to *cis*-eQTLs, mainly reflecting the multiple testing burden with generally smaller effect sizes in *trans* and limited number of samples in comparable cohorts and tissues (Westra and Franke, 2014). Growing sample sizes of the datasets hold great promise to find more *cis*- and *trans*-eQTLs with even smaller effect sizes as the statistical power is influenced by the sample size and the number of highly expressed genes (Joehanes et al., 2017).

Generally, the effect size of a local eQTL increases as the distance to the transcription start site (TSS) of a gene decreases (Fairfax and Knight, 2014; Westra and Franke, 2014), with a stronger effect on gene expression demonstrated by upstream variants compared to downstream variants (Aguet et al., 2016). Also, the average effect size is negatively correlated with the number of tissues the gene is expressed in. Therefore, at the same significance level, tissue-specific local eQTLs have significantly larger effect sizes than tissue-shared eQTLs (Aguet et al., 2016). Altogether, genetic variants that are likely to affect transcription (by altering transcription factor binding sites or other *cis*-regulatory elements) rather than post-transcriptional regulation of mRNA levels have been shown to have stronger impact on gene expression levels (Aguet et al., 2016). For example, variants at canonical splice sites have the strongest effect on gene expression, inversely, variants in 3'UTRs have the weakest effect with median absolute normalized effect sizes ranging from 0.3 to 1.6-fold between the two eQTL alleles.

A comprehensive study of the functional architecture of local and distal eQTLs in peripheral blood explored the enrichment of eQTLs depending on their location in the genome as well as their "epigenomic context", altogether referred to as 57 functional categories (Liu et al., 2017). Of note, the authors made use of stratified LD-score regression (Bulik-Sullivan et al., 2015) to be able to include those genetic variants below the stringent significance threshold, and estimated the enrichment of a functional category as a proportion of gene

expression heritability in that category divided by the proportion of SNPs in the given category. The most enriched categories for both local eQTLs and distal eQTLs included coding regions, conserved regions, and histone marks H3K27ac (linked with active enhancers) and H3K4me3 (linked with transcriptionally active promoters). The importance of variants in 5'UTRs, promoter regions, and transcription factor binding sites were observed for local regulatory control, while the coding regions of expressed genes (such as transcription factors) were important contributors in the distal regulatory control of gene expression, well in line with the known nature of local and distal eQTLs. Importantly, the functional architecture of the regulation of gene expression was consistent across 15 different tissues (Liu et al., 2017).

1.2.2.2. Cell type- and tissue-specificity

Numerous analyses of different tissues and cell types has extended the knowledge about tissue/cell type-specificity of eQTLs. In general, *cis*-eQTLs tend to have stronger effect sizes and are mainly shared across tissues, while *trans*-eQTLs have weaker effects sizes and are tissue-specific (Grundberg et al., 2012; Price et al., 2011). For example, to quantify the tissue-sharing of *cis*-eQTLs, a high degree of shared *cis*-eQTLs have been found between brain tissues (mean estimated overlap of 0.864), arterial tissues (mean estimated overlap of 0.854), skeletal muscle and heart tissues (mean estimated overlap of 0.819) from the GTEx V6 study (Aguet et al., 2016), reflecting the functional similarities of the tissues. Among primary immune cells, expectedly, sharing of local genetic variants is marginally greater between myeloid cells (monocytes and neutrophils) than with lymphoid cells ($CD4^+$ T cells), profiled as part of the BLUEPRINT epigenome (Chen et al., 2016) and Immune Variation (Raj et al., 2014) projects, with estimated proportions of shared effects being 0.85 for profiled myeloid cells and 0.62–0.71 between myeloid and lymphoid cells. In general, tissue sharing of local eQTLs is linked with cell type-specific implications in complex genetic diseases (Aguet et al., 2016). The genes affected by an eQTL in a specific tissue are overrepresented in different disease gene datasets annotated by several databases, such as the GWAS Catalog (Welter et al., 2014) and the list of loss-of-function intolerant genes from the Exome Aggregation Consortium (Lek et al., 2016).

While the sharing patterns of *trans*-eQTLs between tissues in the GTEx V6 study are in concordance with the patterns among *cis*-eQTLs, the regulation of gene expression in *trans* demonstrates greater tissue-specificity than regulation in *cis* (Jo et al., 2016). This is also confirmed by the patterns of genetic correlation of gene expression across tissues (Liu et al., 2017). Moreover, *trans*-eQTLs reveal greater enrichment for enhancer regions than *cis*-eQTLs (Jo et al., 2016), well in line with the greater tissue-specificity linked with enhancer activity. Also, similar promoter *versus* enhancer enrichment and specificity relatedness is observed among primary and secondary *cis*-eQTLs (conditionally independent from the primary *cis*-eQTLs), in accordance with the increased

distance of the secondary variants from the TSS and their greater tissue-specific nature (Aguet et al., 2016).

1.2.2.3. Condition-specificity

In addition to tissue- or cell-dependent regulation of gene expression, the extent and the strength of the effect of eQTLs on gene expression levels may vary with different cellular and environmental conditions.

In 2014, the first comprehensive studies described stimulus-induced eQTLs in monocytes (Fairfax et al., 2014) and human dendritic cells (Lee et al., 2014). For example, Fairfax and colleagues showed that in monocytes, a majority of the identified *cis*-eQTLs were observed only in cells activated by the exposure to early (2 hours) or late (24 hours) bacterial lipopolysaccharide or interferon-gamma treatment, which trigger an inflammatory response in cells. In addition to the treatment- and time-specific eQTLs, a notable proportion (54.1%) of *cis*-eQTLs found before stimulation were specific to the unstimulated state in resting cells, showing no association after the treatment. In general, c*is*-eQTLs were found to act through well-annotated pathways of monocyte signaling, while *trans*-eQTLs were putatively driven by various mechanisms, such as coding variants in genes known to modulate enzymatic activity and a *cis*-eQTL for a gene that modulates cytokine networks over time.

Recent studies have highlighted the interplay between environmental modifiers and genetic predisposition to complex diseases. Interestingly, a substantial fraction of disease-associated eQTLs remain hidden in resting cells, while about a half of them already alter the chromatin accessibility in unstimulated cells, waiting for the right cellular state or condition to modulate gene expression as a response to immune activation (Alasoo et al., 2017). Moreover, genetic predisposition to autoimmune diseases may also be driven by faulty response to an environmental stimuli rather than static and uniform malfunction (Kim-Hellmuth et al., 2017). Altogether, context-specific eQTLs present a highly dynamic role of response to stimulus in modulating genetic regulation of gene expression and the need to study this phenomenon in depth.

1.2.3. Importance of eQTLs

The most notable benefit of mapping eQTLs is the identification of genes and pathways that contribute to disease susceptibility or trait formation and are the targets of GWAS loci (Gibson et al., 2015). As the disease genes tend to be overexpressed in specific tissues (Goh et al., 2007; Lage et al., 2008), eQTLs are also promising in narrowing down the tissue or cell type essential for disease etiology. Thereby, eQTLs provide means for improved understanding of disease mechanisms, identification of promising new targets for therapeutic interventions, and discoveries of drug mechanisms.

1.2.3.1. eQTLs as means for understanding disease mechanisms

A striking example is the unravelling of causal connections between the chromosome 1p13 locus, plasma low-density lipoprotein cholesterol (LDL-C), and myocardial infarction (MI) in humans (Musunuru et al., 2010). LDL-C is a well-known risk factor for MI (Brown and Goldstein, 1996; Rader et al., 2003). A GWAS for plasma lipoprotein traits identified the locus on 1p13 as the strongest for the associations with LDL-C across the genome (Teslovich et al., 2010). Meanwhile, the same SNPs in the locus have been linked to coronary artery disease and MI (Kathiresan et al., 2009; Samani et al., 2007). Notably, the SNPs lie in the noncoding region of 6.1 kilobases (kb) in size between the 3'UTR of *CELSR2* and 3'UTR of *PSRC1* (oriented in the opposite direction). eQTL mapping harboring the 1p13 locus in liver, subcutaneous fat, and omental fat, suggests liver-specific *cis*-regulation of the expression of *PSRC1* and the nearby *SORT1* gene, whereas the 1p13 minor haplotype associates with increased expression of the genes as well as decreased LDL-C. The SNP rs12740374 was established to be the putative causal SNP. Its minor allele creates a binding site for the C/EBP transcription factor that consequently upregulates the *SORT1* gene. The functional effects of altered *SORT1* expression on lipoprotein metabolism were confirmed by knockdown and overexpression studies of *Sort1* in the livers of mice. Concordant with the findings in humans, the overexpression of *Sort1* lowered LDL-C and very low-density lipoprotein (VLDL) particle levels in the mouse liver, with the effects arising from altered hepatic VLDL secretion.

Another example involves the fat mass and obesity-associated (*FTO*) gene. Common variants in introns 1 and 2 of the *FTO* gene are strongly associated with childhood and adult obesity (Dina et al., 2007; Frayling et al., 2007; Scuteri et al., 2007). In the original study by Frayling et al. (2007), the 16% of the adults carrying both of the risk alleles were about 3 kilograms heavier than those with no risk alleles. However, no clear candidate functional variants for the association in the coding region of the *FTO* gene were revealed, neither in minimal splice sites nor 3'UTRs. The mouse studies that followed demonstrated that *Fto*-deficiency resulted in leaner mice (Fischer et al., 2009) and the overexpression of *Fto* lead to increased body mass due to the increased food intake (Church et al., 2010). Meta-analysis of 45 studies (218,166 adults in total) highlighted the benefits of physical activity in regulation of body weight in adults, despite their genetic predisposition to obesity (Kilpeläinen et al., 2011). The authors found that each additional risk allele increased the odds of obesity by 1.23-fold (95% confidence interval (CI): 1.20–1.26), while the odds of obesity was 1.22-fold/per risk allele (95% CI 1.19–1.25) for physically active individuals and 1.30-fold/per risk allele (95% CI: 1.24–1.36) for the physically inactive individuals. By utilizing chromatin looping, a physical contact over a range of 500 kb between the noncoding sequence of the *FTO* and the homeobox gene *IRX3* was established (Smemo et al., 2014). Namely, the obesity-associated *FTO* region interacts with the promoters of both *FTO* and *IRX3* in

the human, mouse and zebrafish genomes. *IRX3* encodes a transcription factor highly expressed in the brain. Concordantly, by utilizing eQTL mapping, the intronic variants were shown to influence the expression of *IRX3* in the human brain, but not *FTO* itself, suggesting that the obesity-associated interval belongs to the regulatory region of *IRX3*. Notably, *Irx3*-deficient mice reduced body weight by 25–30%, mostly through the loss of fat mass and elevated basal metabolic rate, and were protected against diet-induced obesity. Still, more work needs to be done to elucidate the causal variants and precise molecular mechanisms involving *IRX3* as an important metabolic regulator.

1.3. Analytical framework for eQTL mapping

Mapping of eQTLs aims to find genetic variants that influence gene expression levels in a given cell or tissue under specific conditions in a cohort of unrelated individuals. The trait of interest is a quantitative measure of mRNA abundance, measured by a high-throughput and low-cost genome-wide expression array or RNA-seq, which enables a finer-scale resolution. Determination of genetic variants in each individual is done by genotyping using microarrays followed by imputation or whole-genome sequencing. For each gene (or probe or exon), a set of SNPs is used to test for significant correlation between allele dosages and expression levels, using linear regression or non-parametric Spearman correlation.

Associations between SNPs and gene expression levels are categorized into two classes based on the SNP-gene proximity. As a golden standard in mapping studies, SNPs with the physical distance of +/– 1 megabase (Mb, the distance of the longest human enhancer from the TSS (Raj et al., 2013)) from the gene are considered as local or *cis*-eQTLs, and all other SNPs with physical distances greater than 5 Mb from the gene, or where the SNP is located on a different chromosome than the gene, as distal or *trans*-eQTLs. Thus, the terminology used does not differentiate the relative position from the underlying mode of action of the eQTL. True *cis*-eQTLs can only be distinguished from local eQTLs by using RNA-seq data allowing to assess the imbalance in the expression levels of the two alleles, which is impossible with microarrays. Of note, testing for *trans* associations is usually constrained to SNPs that have been found to be significant in the genome-wide association studies to limit the number of tests performed. For example, a *cis*-eQTL mapping with close to 6 million autosomal SNPs and 39,000 probes results in approximately 155 million SNP-probe pairs to test, but for *trans*-eQTL mapping the number of SNP-probe pairs to test would be 234 billion.

Further, to formally link the disease-associated loci to their target genes (in specific tissue or cell types), a colocalization analysis is needed to assess whether both the GWAS and eQTL signals are driven by the same causal variant. Due to the extensive LD structure in the genome, it is possible that a particular genetic variant is associated with both signals, but it is more likely that two distinct underlying causal SNPs exist in LD with the given one

(Wallace et al., 2012). Currently, there are different colocalization approaches available, such as COLOC (Giambartolomei et al., 2014) and eCAVIAR (Hormozdiari et al., 2016), based on summary statistics, or regulatory trait concordance (Nica et al., 2010) that uses individual genotype level data. As an advantage over the other methods, eCAVIAR allows for more than one causal variant in a given locus, which is biologically relevant phenomena named as allelic heterogeneity.

1.3.1. Statistical approaches

The success of eQTL mapping depends on the sample size and the effect size of the genetic variant on gene expression levels. Linear regression analysis or non-parametric Spearman correlation are efficient approaches for detecting eQTL effects in a single tissue (tissue-by-tissue analysis). Due to multiple testing burden and arbitrary cut-off levels for significance, counting the overlap of eQTLs in different tissues or cell types in a tissue-by-tissue analysis may lead to higher estimates of tissue-specificity than actually present (Flutre et al., 2013). Therefore, multi-tissue eQTL analysis is a favorable choice to formally estimate the proportion of shared eQTLs among tissues and gain power to detect eQTLs by modelling the sharing patterns across tissues. Additionally, gene-by-environment interaction models allow to search for cell type- or context-dependent eQTLs in whole blood samples without the need to sort or stimulate the cells (Westra et al., 2015; Zhernakova et al., 2016).

Details of tissue-by-tissue and multi-tissue analyses of microarray data are described in a framework of a non-parametric eQTL model based on the Spearman correlation developed in Prof. Lude Franke's lab (Westra et al., 2013) and the Bayesian model averaging developed in Prof. Matthew Stephens' lab (Flutre et al., 2013), respectively, in the next subchapters.

1.3.1.1. Tissue-by-tissue analysis

The eQTL mapping analysis cookbook by Prof. Lude Franke's group provides tools for normalization, correction for population stratification, identifying sample mix-ups, correction for confounders, and *cis*- and *trans*-eQTL mapping (Westra et al., 2013).

To minimize the number of artificial associations, data pre-processing is essential as a first step in every analysis pipeline involving microarray data. Raw data sets usually contain systematic variation to some extent (Leek et al., 2010). Both technical (for example, the day the experiments were run, several lots of reagents or chips used) and biological (for example, different age and sex of the individuals) factors contribute to the overall variation. To ensure the comparability of measurements for individual samples, firstly the raw gene expression data is quantile normalized to match the percentiles of each array (Bolstad et al., 2003), $\log_2$-transformed, and expression values of probes are

centered and expression values of samples are standardized (to have mean 0 and variance 1). Next, normalized expression data is corrected for possible population stratification by regressing out the four multidimensional scaling vectors obtained from the pruned SNP genotype data, and the resulting residuals are used for subsequent steps.

As normalization does not remove batch effects, additional effects of some covariates need to be removed. Firstly, the components capturing majority of the variation between the samples are established by principal component (PC) analysis on the sample correlation matrix calculated on the normalized gene expression data adjusted for population stratification. The first PCs primarily capture sample differences in expression that are reflected by the major sources of variation due to technical, environmental, demographic, or genetic factors (Leek and Storey, 2007). Thus, by regressing out the first PCs, the resulting residual expression data would be more strongly determined by the genetic variants and the power to detect eQTLs would increase (Fehrmann et al., 2011; Westra et al., 2013). To ensure that no genetic effects are removed, only the PCs showing no association with SNPs are accounted for. The optimal number of non-genetic-PCs to remove is determined by the highest number of eQTLs detected in iteratively performing eQTL mapping on the residual expression data adjusted for an increasing number of PCs each time. To quicken the process, only a subset of SNPs is used.

Standard settings from GWAS are applied for SNPs used for eQTL mapping, including minor allele frequency > 0.01 or 0.05, call rate > 0.95, and Hardy-Weinberg *P*-value > 0.0001. In addition, polymorphisms in the probe sequence may cause hybridization differences captured by the microarray probe (Alberts et al., 2007). To minimize the number of false-positive *cis*-eQTLs driven by such SNPs that are in LD ($r^2 > 0.2$) with the target SNP of interest, any target SNP-probe pair fulfilling the criteria is excluded from the analysis. However, false-positive *trans*-eQTLs may arise by cross-hybridizing probes. Therefore, the list of probes with *trans*-eQTLs are mapped again to a region of 5 Mb of the *trans*-eQTL with relaxed settings. Any SNP-probe pair, where the probe maps to the tested region with at least 15 identical bases, is considered to be false positive and excluded from the downstream analyses.

Both *cis*- and *trans*-associations are estimated using the Spearman's rank correlation coefficient (ρ) between the imputed allele dosages ranging from 0 to 2 and the residual gene expression levels obtained after normalization, correction for population stratification, and correction for technical and biological factors. Namely, the Pearson correlation coefficient (r) is computed on the ranks of the two aforementioned variables:

$$\rho = r_{rg_X, rg_Y} = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}},$$

where rg_X and rg_Y denotes the ranks of the respective variable (allele dosages or residual gene expression), $cov(rg_X, rg_Y)$ is the covariance of the rank variables, σ_{rg_X} and σ_{rg_Y} are the standard deviations of the respective rank variable.

To determine the significance of the correlation coefficient, firstly the Spearman's correlation coefficients are converted to *t*-values, which follow *t*-distribution with $n - 2$ degrees of freedom:

$$t = \frac{\rho}{\sqrt{\frac{1-\rho^2}{n-2}}},$$

where ρ is the Spearman's correlation coefficient and n is the sample size. Then *t*-values are converted to *Z*-scores by matching the quantiles of *t*-values to the corresponding quantiles in the standard normal distribution. Finally, two-sided *P*-values are calculated using the standard normal distribution.

In a case of meta-analysis, the sum of *Z*-scores weighted by the square root of the sample size in a given cohort and divided by the square root of the number of total individuals with both genotype and expression data available is found over the datasets:

$$Z_{meta_{SNP_j Probe_k}} = \frac{\sum \sqrt{n_{Dataset_i SNP_j}} \times Z_{Dataset_i SNP_j Probe_k}}{\sqrt{N}},$$

where $Z_{Dataset_i SNP_j Probe_k}$ denotes the *Z*-score between the probe *k* and the SNP *j* in the cohort *i*, $n_{Dataset_i SNP_j}$ is the number of individuals with genotype data available for the SNP *j* in the cohort *i*, and *N* is the total number of individuals with both genotype and expression data available over the cohorts.

To control for multiple testing, a false discovery rate (FDR) procedure using permutations is applied on probe-level to account for the number of SNPs tested per probe. In short, firstly, a null distribution is generated by performing eQTL mapping on the permuted residual gene expression data with shuffled sample labels and repeating the procedure ten times. Shuffling the sample labels in the gene expression data preserves the LD-structure in the genotype data and the co-expression structure in the gene expression data, but breaks the correct link between the genotypes and the expression levels. Then, for both the real and the permuted *P*-value distributions, only the most significant SNP per probe is used to determine the FDR, the proportion of the incorrect rejections (false discoveries or false positive results) among all rejected null hypothesis. Namely, for any given $q > 0$, if we claim any SNP-probe pair satisfying $p < q$ to be significant, we can estimate FDR as the number of false positives divided by the number of true positives:

$$\widehat{FDR}(q) = \frac{\widehat{FP}(q)}{\widehat{TP}(q)} = \frac{\sum_{b=1}^{B} \#\{i: p_{perm_i}{}^{(b)} < q\}/B}{\#\{i: p_{real_i} < q\}},$$

where $b = 1, \ldots, B$ is the number of the permutation, i denotes the set of SNP-probe pairs, where only the most significant SNP per probe is included to the estimation of FDR, *FP* is the number of false positives, and *TP* is the number of true positives.

Restricting the calculations only for the most significant SNP per probe helps to overcome the potential issue with the loci of extensive LD between the SNPs correlating with one probe. The distribution of all *P*-values for calculation of FDR at 5% accurately gives control over the number of false-positive SNP-probe pairs among the significant association pairs. However, the number of unique false-positive probes with eQTLs will likely be inflated due to the disproportionate number of real eQTLs from the regions of extensive LD. The probe-level FDR is somewhat conservative approach that does not consider the presence of potential independent secondary effects. Of note, ten permutations is considered to be enough for a stable estimate of the false discovery threshold after ascertaining the fluctuation of the significance threshold using the increasing number of permutations from one to twenty. The significance thresholds for the *cis*- and *trans*-eQTL are assessed separately.

1.3.1.2. Multi-tissue joint analysis

The multi-tissue eQTL analysis methods by Prof. Matthew Stephens' lab implemented in the eQtlBma program allow for *cis*-eQTL mapping jointly in all tissues by a hierarchical model based on the multivariate regression in a Bayesian framework fitted via an expectation-maximization (EM) algorithm (Flutre et al., 2013). In general, multiple tissues can be extended to multiple subgroups, such as different populations, platforms, treatments, etc. As an advantage, by modelling the inactivity of eQTLs in some tissues, the proportion of eQTLs shared across different tissues is formally estimated as a model parameter. Additionally, the benefits of acknowledging sharing information are greater for the association pairs, where SNPs have modest effects on the gene expression levels across tissues, and for tissues with small sample size.

Briefly, in each tissue, $s = 1, \ldots, S$, a linear regression model is used to evaluate the association between a target gene and a target SNP:

$$y_{si} = \mu_s + \beta_s g_i + \varepsilon_{si},$$

where y_{si} marks the suitably normalized and transformed expression levels of the target gene in the tissue *s* for the i^{th} individual, μ_s is the mean expression level of this gene in the tissue *s*, β_s is the effect of the SNP on the expression levels in the tissue *s*, g_i is the allele dosages of the i^{th} individual for the SNP (ranging from 0 to 2, indicating the number of copies of a reference allele), and

ε_{si} is the residual error for the tissue s and the individual i, $\varepsilon_{si} \sim N(0, \sigma_s^2)$. Residual variances are allowed to be different in each tissue and $\varepsilon_{1i}, \ldots, \varepsilon_{Si}$ may be correlated when the same set of individuals have the gene expression measurements in different tissues (correlation matrix is estimated from the data).

Whether the SNP is an eQTL in any tissue is evaluated using configurations. A configuration, $\gamma = (\gamma_1, \ldots, \gamma_S)$, is a binary vector, where $\gamma_s \in \{0,1\}$ indicates whether the SNP is an eQTL in the tissue s. The global null hypothesis H_0 is that the SNP is not an eQTL in any tissue, $\gamma = (0, \ldots, 0)$. To perform inference on γ, a flexible family of distributions, $P(\beta|\gamma, \theta)$, is used, where θ denotes the various combinations of hyper-parameters for the typical effect size and the heterogeneity of effects across tissues, i.e. grid points, thus allowing for differences among eQTLs present in each tissue.

The steps needed for the inference on γ can be divided into three main parts: 1) computation of the Bayes factors, i.e. the likelihood ratios of the support in the data for each configuration relative to the null model, 2) combining information across all genes by the hierarchical model (Gelman and Hill, 2006) to estimate the hyper-parameters from the data, 3) Bayesian model averaging across the configurations to detect genes with eQTLs and calculation of posterior probabilities for each configuration to identify tissues where the eQTLs are active.

Firstly, to assess the support in the data for each SNP-gene pair being an eQTL in the tissue s relative to the H_0, the Bayes factors (BF) are computed for each configuration γ and each grid point:

$$BF_\gamma = \frac{P(\text{data}|\text{true configuration is } \gamma)}{P(\text{data}|H_0 \text{ true})}.$$

So, the BF_γ quantifies the support in the data for the given configuration γ. The BFs are used to build test statistics to detect genes that have an eQTL in at least one tissue and to identify the tissue(s) in which these eQTLs are active.

Secondly, information across all genes is combined by the hierarchical model with an EM algorithm to get the maximum-likelihood estimates of the hyper-parameters (configuration probabilities and grid points) from the data. By learning the extent of sharing between tissues from the data, the hierarchical model is expected to exploit, for example, that *Tissue1* shares more eQTLs with *Tissue2* than with the other tissues.

Thirdly, the strength of the evidence over the different possible alternative configurations against H_0 at the candidate SNP is achieved by the Bayesian model averaging (BMA) across the configurations using the Bayes factors weighted by the estimated hyper-parameters obtained with the hierarchical model (HM):

$$BF_{BMA}^{HM} = \frac{P(\text{data}|H_0 \text{ false})}{P(\text{data}|H_0 \text{ true})} = \sum_{\gamma \neq (0,\ldots,0)} \eta_\gamma BF_\gamma \,,$$

where $\eta_\gamma = P(\gamma|H_0\ false)$. To detect genes that have an eQTL in at least one tissue under an assumption of at most one eQTL per gene and the equality of likeliness of each SNP to be the eQTL, BF_{BMA}^{HM} is averaged across all candidate SNPs in the *cis*-region to give the overall evidence against H_0 for the given gene.

To obtain the posterior probabilities, an estimate of the probability for a gene to have no eQTL in any tissue, π_0, is needed. One approach is to use the gene-level *P*-values obtained by permutations and estimate the π_0 using the *q*-value approach (Storey and Tibshirani, 2003). Specifically, the list of *P*-values is used to estimate the overall proportion of the true null hypotheses, π_0, i.e. those following the $U(0,1)$ distribution. Of note, an estimate of the true alternative tests is $\pi_1 = 1 - \pi_0$, i.e. π_1 is the proportion of true positives and is used to estimate the proportion of shared eQTLs between tissues in a tissue-by-tissue analysis context. The second approach is to use the computationally less intensive EBF or QBF procedure (Wen, 2016). Particularly, for EBF only the gene-level Bayes factors averaged over the grid and the configuration weights (estimated via the EM algorithm) and for QBF the median Bayes factors under the null are needed to estimate π_0.

Finally, to identify the tissue(s) in which the eQTLs have an effect, the posterior probabilities for the eQTL to be active in a given configuration are obtained:

$$P(\text{true configuration is } \gamma|\text{data}, H_0 \text{ false}) = \frac{\eta_\gamma BF_\gamma}{\sum_{\gamma \neq (0,\ldots,0)} \eta_\gamma BF_\gamma},$$

and the posterior probability that the SNP is an eQTL in the tissue *s* is the sum of the probabilities over the configurations in which $\gamma_s = 1$:

$$P(\text{eQTL in tissue } s|\text{data}, H_0 \text{ false}) = \sum_{\gamma:\, \gamma_s = 1} P(\text{true configuration is } \gamma|\text{data}, H_0 \text{ false}).$$

As a final result, the best SNP per gene can be picked based on the posterior probability for a SNP to be "the" eQTL for the given gene. The genes having an eQTL at a given FDR threshold are identified using the *q*-value approach (Storey and Tibshirani, 2003) on the gene-level *P*-values obtained by per-mutations. Additionally for each gene, the significant gene-SNP pairs can be determined in a fine-mapping context (Wen, 2014; Wen et al., 2015, 2016).

Altogether, the tissue-by-tissue approaches are powerful for detecting eQTLs present in a single tissue, but are outperformed by the multi-tissue approaches if

the number of tissues sharing the eQTL effect increases. The multi-tissue joint analysis allows heterogeneity in eQTL effects among tissues, but unlike ANOVA, also allows for varying variances of expression levels in different tissues, thus in practice, outperforming also ANOVA or simple linear regression with the interaction term.

1.4. Future perspectives of eQTL studies towards precision medicine

During the past ten years of the successful GWAS era, the main purpose was to identify genetic loci contributing to human traits. The major challenge for the following ten years is to ascertain the functional mechanisms of these discovered loci to ultimately improve human health by enhanced prognostic, preventive, and therapeutic measures (Edwards et al., 2013; Huang, 2015).

To this end, mapping of eQTLs has proven highly informative in suggesting mechanisms and target genes affected by GWAS loci. However, the effect of the functional variant is dependent on the proximity to the gene and on the accompanied epigenomic landscape (Knight, 2005). To prioritize candidate variants in transcriptionally active regions for functional validation, it is necessary to integrate regulatory elements, for example by mapping of DNase I hypersensitivity sites to indicate chromatin accessibility and ChIP-seq peaks to indicate DNA binding sites for proteins like transcription factors. Furthermore, gene-by-gene and gene-by-environment interactions are substantial components of complex traits, suggesting the importance of incorporation of different layers of biological data, such as proteins, metabolites, and also gut microbiota, to fully understand the mechanisms for complex traits before clinical applications (Civelek and Lusis, 2013).

The potential of genetic risk scores (GRS), the sum of weighted risk alleles, has been shown for many complex diseases, such as schizophrenia and bipolar disorder (Reginsson et al., 2017), coronary heart disease (Abraham et al., 2016), and type 2 diabetes (Läll et al., 2017). Following this idea, another use of eQTLs in clinical practice would be in personalized risk assessment as an additional instrument in the form of transcriptional risk score (TRS, Gibson et al., 2015). Namely, TRS is a sum of standardized gene expression measures of genes that are affected by genetic variants, jointly linked with a disease of interest and gene expression levels, which are preferably measured in the relevant tissue or cell type. TRS can be further weighted by the eQTL effect size and polarized to be positively correlated with the GRS. It is argued that integrating genotypic information with gene expression levels would result in a better predictor compared to GRS alone as the effect of genotypes on disease susceptibility is obtained through altered gene expression profiles (Gibson et al., 2015).

Fundamentally, experimental validation of the statistical association in the native genomic context is the ultimate proof of causality. Rapid advances in genome editing open up new possibilities in addition to the *in vitro* studies in

cultured cell lines or primary tissues or *in vivo* studies with model organisms. Namely, the clustered, regularly interspaced, short palindromic repeat-associated protein 9 (CRISPR-Cas9) technology is a new versatile approach to analyze the effect of putative causal variants (Cong et al., 2013; Sander and Joung, 2014). As a result, the short RNA-guided nuclease, Cas9, induces precise cleavage at genomic loci within a specific targeting range in the mammalian genome (Hsu et al., 2014; Mali et al., 2013; Shen et al., 2014). Upon cleavage, the double-strand breaks at the target locus undergo DNA damage repair, which can result in desired insertions (also used to introduce gene knockouts), deletions, or substitutions at the target site (using Cas9 together with repair templates bearing the mutation of interest), but with potential off-target modifications (Ran et al., 2013). By the using CRISPR-Cas9 approach, the modified cell lines provide validation for both *cis*- and *trans*-eQTLs in their native genomic context (Claussnitzer et al., 2014; Lee et al., 2014; Raghavan et al., 2016; Soldner et al., 2016). For example, Soldner and colleagues combined epigenetic information with CRISPR-Cas9 genome editing in human pluripotent stem cells to link a common Parkinson's disease-associated risk variant in a distal enhancer with the expression of a Parkinson's disease development gene, *SNCA*. They found that *SNCA* expression is modulated by sequence-specific binding of the brain-specific transcription factors EMX2 and NKX6-1 at the distal enhancer. Additionally, CRISPR-Cas9-based epigenome editing allows for loss- and gain-of-function screens to identify regulatory elements in the human genome (Thakore et al., 2015; Fulco et al., 2016; Klann et al., 2017), widening the opportunities even more.

2. AIMS OF THE STUDY

The aim of this thesis was to characterize the genetic regulation of gene expression in general and to improve the eQTL maps for tissues and purified cell types.

The specific objectives of the thesis were as follows:

1. To investigate the regulation of hepatic gene expression and estimate the proportion of variation in gene expression that could be attributed to genetic and epigenetic variation.

2. To describe the extent of genetic control of gene expression in $CD4^+$ and $CD8^+$ T cells, assess the specificity and proportion of shared eQTLs between the two cell types, and find functional consequences of disease-associated variants in T cells.

3. To explore eQTL SNPs that either disrupt an existing miRNA binding site or create a new one and their effect on target gene expression levels.

3. RESULTS AND DISCUSSION

3.1. Genetic and epigenetic regulation of hepatic gene expression (Ref. I)

The liver plays a central role in the maintenance of homeostasis and health in general. Given the substantial inter-individual variation seen in metabolism, regulation of nutrients, protein synthesis, and detoxification of xenobiotics, it is essential to have a better understanding of the sources of inter-individual variation in hepatic gene expression, such as genetic variants, DNA methylation, or different developmental stages.

Several SNPs associated with liver function and related diseases are gleaned from GWAS (Adams et al., 2013; Ellinghaus et al., 2013; Suhre et al., 2011) and further explored to enlighten their effect on liver gene expression levels by eQTL mapping (Innocenti et al., 2011; Schadt et al., 2008; Schröder et al., 2013). Apart from genetic variation, epigenetic mechanisms, such as DNA methylation and histone modifications, play an important role in regulating tissue- and condition-specific gene expression (Ghosh et al., 2010; Lister et al., 2009; Varley et al., 2013). Hence, it is advantageous to incorporate epigenetic variation when studying the regulation of hepatic gene expression, to enhance our knowledge base.

By analyzing the methylomes and transcriptomes of 14 fetal and 181 adult livers, the first part of the thesis generates a comprehensive resource of factors involved in the regulation of hepatic gene expression, and allows us to estimate the proportion of variation in gene expression that could be attributed to genetic and epigenetic variation, both crucial in understanding differences in drug response and the etiology of diseases involving the liver.

3.1.1. Description of cohort and materials

The study was performed on two different cohorts, the Karolinska Liver Bank cohort and the Dutch tissue cohort MORE (BBMRI obesity cohort).

The Karolinska Liver Bank sample set included 96 adult and 14 fetal tissue samples. Adult tissues were collected from organ donors who had died in accidents (52 samples) and from patients undergoing liver resection due to malignant tumors (44 samples), most commonly from patients with metastatic colon cancers. Liver biopsies from these patients were collected from 'healthy' tissue that showed no visible pathological changes compared to the adjacent tumor. Liver samples from the 14 fetuses were obtained at gestational week 8 to 21. The datasets have been described in previous publications in detail (Ivanov et al., 2013; Kacevska et al., 2012). The Ethics Committees at Karolinska University Hospital approved the use of all samples for the purposes of this study.

A second set of 85 samples was collected from morbidly obese Dutch individuals with a body-mass index between 35 and 70. Tissue samples were

collected from their liver, muscle, subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT), as described before (Fu et al., 2012; Wolfs et al., 2010).

All of the following steps were done according to standard protocols and manufacturer's instructions. DNA from the samples was genotyped using Illumina HumanOmni BeadChips. We imputed both datasets using the GIANT release from the 1000 Genomes project, resulting in 5,763,069 unique SNPs, which were used in all downstream analyses. Gene expression data was generated using Illumina HumanHT-12 BeadChips and bisulfite-converted DNA samples were hybridized to Illumina Infinium HumanMethylation450 BeadChips to get methylation signals.

3.1.2. Developmental regulation of hepatic gene expression

To disentangle the regulation of the developing liver, we compared the transcriptomes and epigenomes of the fetal and adult liver samples. Fetal development is characterized by tissue differentiation and growth. During this period, the liver is predominantly a hematopoietic organ (Moscovitz and Aleksunes, 2013). Our observed differences in gene expression and DNA methylation are well in line with the known biological functions.

Namely, the 1,396 genes expressed at higher levels in the adult livers compared to the fetal livers were strongly enriched for metabolic functions like monocarboxylic acid, steroid and bile acid metabolic processes, as well as the response to xenobiotic processes (Table 2A in Ref. I). Likewise, genes that were associated with hypomethylation in adult livers compared to the fetal livers were highly enriched for metabolic pathways, such as steroid metabolism, the regulation of lipid metabolism, the regulation of the generation of precursor metabolites and energy, and regulation of glycolysis (Table 1A in Ref. I)

In contrast, 1,277 genes expressed at higher levels in the fetal livers compared to the adult livers were associated with regulating organelle organization, chromosome organization, and hemoglobin biosynthetic processes (Table 2B in Ref. I). Similarly, the genes that were associated with hypomethylated CpG sites in the fetal livers compared to the adult livers were strongly enriched for pathways of insulin receptor signaling, regulation of glycogen synthase activity, differentiation processes, and developmental functions (Table 1B in Ref. I).

In total, we found 1,655 genes that showed both differential expression and differential methylation in comparison of the two groups. More specifically, 657 genes were linked to probes with higher expression levels in adults and 1,000 genes were linked to probes that were more highly expressed in fetal livers (with an overlap of two genes). As expected, these genes were even more significantly enriched for developmental stage-specific functions, such as drug response for the adult cohort and liver development for the fetal cohort. Also, the regions within 2 kb of the TSS of the aforementioned genes were enriched for binding sequences of transcription factors essential for the development or

function of the liver, specifically HNF4A and HNF1A; hematopoietic transcription factors GATA1, STAT5A, and STAT5B; and YY1, which plays a fundamental role in embryogenesis and differentiation. Accordingly, the genes that encode the transcription factors showed clear and significant differences in expression levels between fetal and adult liver samples.

We highlighted a couple of genes with the greatest changes during liver development (Fig. 2 in Ref. I), including the fetal-specific expression of genes involved in differentiation and hematopoiesis (e.g. *DLK1*, *HBZ*, *HBM*, *AHSP*, *EPB42* and *NFE2*) and the adult-specific expression of genes involved in drug metabolism, catabolism and other biosynthesis processes (e.g. *CYP2E1* and *CYP2C8* that are the cytochrome P450 (*CYP*) genes), that showed up to 9-fold difference in expression and up to 50% change in methylation, illustrating the magnitude of developmental regulation of hepatic gene expression.

3.1.3. Contribution of genetic and epigenetic variants to variation in hepatic gene expression

Next, we aimed to ascertain to what extent SNPs and DNA methylation could jointly explain the variation in liver gene expression levels. Firstly, we mapped local genetic variants (*cis*-eQTLs, within 250 kb of the expression probe) and methylation CpG sites (eQTM – expression quantitative trait methylation, within 250 kb of the expression probe) that correlate with gene expression in the adult livers from the two cohorts. Also, we studied the effects of SNPs on CpG methylation (meQTL – methylation quantitative trait loci, within 250 kb of the CpG site).

In total, we found 47,168 significant *cis*-eQTLs (FDR < 0.05), representing 751 unique genes (including 84 novel genes) which were strongly liver-specific and enriched for drug metabolizing functions. In contrast, the 12,054 unique genes with significant *cis*-meQTLs (FDR < 0.05) did not show any enrichment of liver functions. We discovered a total of 3,238 significant eQTMs, comprising 1,988 unique expression probes (in 1,798 genes) and 2,980 CpG sites (reflecting 2,057 unique genes), with a permutation *P*-value < 0.05. For example, in case of CpG sites with strong correlation between expression and methylation levels, and/or within 50 kb of the expression probes, we observed an overrepresentation of negative correlations (Fig. 3 in Ref. I). Interestingly, CpG sites downstream of the expression probes displayed less negative correlations than those upstream of the probes, indicating that the methylation in gene bodies is associated with active gene expression, as known from the early days of DNA methylation research (Jones, 1999).

To determine the proportion of explained variation in gene expression, we selected 293 expression probes (reflecting 274 unique genes) that had both a significant eQTL and significant eQTM effect. For 83% of probes, most of the variability in expression was explained by a SNP, whereas for the remaining 17%, the variability in expression was mostly explained by a specific CpG site.

For the latter cases, we observed that these expression-associated CpG sites were also likely affected by a SNP (meQTL). As expected, using SNPs jointly with CpG sites yielded even higher estimates of variation explained. Overall, the unique contribution of genotypes in describing the variation in gene expression was greater compared to methylation levels (median 0.1 and standard deviation 0.122 *versus* median 0.029 and standard deviation 0.049, respectively).

The extent of variation explained in genes involved in the absorption, distribution, metabolism and excretion of drugs (ADME genes) ranged from 13% to 84% with no clear patterns of favoring the importance of SNPs over CpG sites (Fig. 3). This clearly illustrates the necessity to extensively study the inter-individual variability in ADME genes driven by genetic polymorphisms, DNA methylation marks, and their interaction to be able to dissect their effect in drug efficacy, toxicity, and susceptibility to environmental toxins.

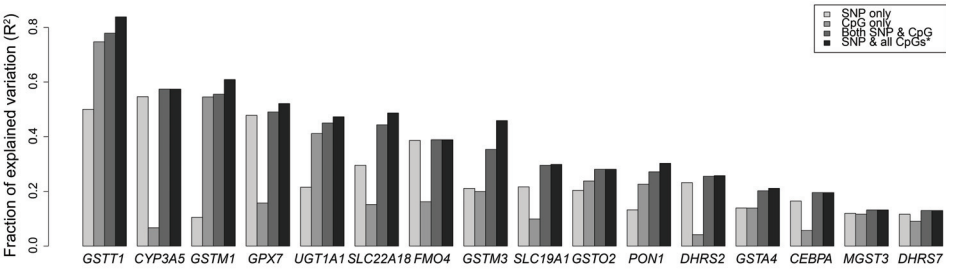

Figure 3. The contributions of SNPs and DNA methylation levels to the fraction of variation explained in gene expression levels of 16 ADME genes. Shown are fractions of explained variation in gene expression of 16 ADME genes by a SNP (eQTL), a CpG (eQTM), both a SNP and a CpG site (eQTL+eQTM) or a SNP and all CpG sites (eQTL+eQTMs).

3.1.4. Tissue-specificity of eQTLs, meQTLs and eQTMs

Three other tissues (muscle, SAT, VAT) were used to assess the tissue-specificity of eQTL, meQTL and eQTM effects. With the given sample sizes and comparing only the overlap of significant effects between tissues, we observed a stronger tissue-specificity for eQTMs and eQTLs compared to meQTLs. We note that the replication rates are probably higher when using methods that take into account the problems with incomplete power.

For liver eQTLs, approximately 40–50% of the effects found in one tissue could also be significantly detected in another tissue. We identified only a few opposite allelic effects (<1%) between the tissues, suggesting that if a SNP affects expression in multiple tissues, the allelic direction is mostly identical. The eQTL effects that were only present in liver and not in the other three tissues were related to genes strongly specific to liver function and metabolic and catabolic processes. In contrast, SNP-methylation correlations were less

tissue-specific than SNP-expression correlation. On average, 70% of the meQTLs are shared between at least two tissues, with over 98% of their effects having the same allelic direction. On the other hand, we observed that DNA methylation associated with the expression levels (i.e. eQTMs) is highly tissue-specific, in accordance with the fact that DNA methylation plays an important role in regulating tissue-specific gene expression. Thus, conclusions drawn from eQTM data in one tissue cannot be extrapolated to other tissues, whereas the effect of SNPs on methylation is more likely to be detectable in another tissue.

3.2. Genetic regulation of gene expression in $CD4^+$ and $CD8^+$ T cells (Ref. II)

While whole blood is one of the most widely used biological materials, and can serve as a bulk tissue for finding eQTLs, the challenge to distinguish the cells responsible for the associations remains. The power of eQTL mapping in purified cell types has now been illustrated for B cells, monocytes, neutrophils, T regulatory cells, and $CD4^+$ T cells (Fairfax et al., 2012; Ferraro et al., 2014; Raj et al., 2014; Andiappan et al., 2015; Naranbhai et al., 2015), which allowed the identification of functional roles for several polymorphisms at autoimmune and even neurodegenerative disease loci.

An expanded survey of cells is required for studying the consequences of the risk alleles in biologically relevant contexts for different diseases. For example, $CD4^+$ and $CD8^+$ T cells are essential elements of the adaptive immune response (Swain et al., 2012). $CD4^+$ T cells, together with an appropriate cytokine environment, are required for the activation and differentiation of $CD8^+$ T cells that mediate defense and pathogen clearance during various infections (Tscharke et al., 2015). The involvement of $CD4^+$ T cells is also necessary for B cells and macrophages to execute their protective functions. Faulty activation or inadequate regulation of $CD4^+$ and $CD8^+$ T cells may contribute to the initiation and progression of multiple autoimmune diseases, including type 1 diabetes (T1D), rheumatoid arthritis (RA), autoimmune thyroiditis, systemic lupus erythematosus, multiple sclerosis, psoriasis, inflammatory bowel disease, as well as allergy and asthma (Liblau et al., 2002; Walter and Santamaria, 2005).

To this end, we purified $CD4^+$ and $CD8^+$ T cells from the peripheral blood of 313 healthy individuals for genome-wide mapping of genetic variation affecting the expression of genes involved in immune response. This part of the thesis provides insights into the extent of genetic regulation of gene expression in T cells and advances our understanding of the pathways in the adaptive immune system involved with disease susceptibility.

3.2.1. Description of cohort and materials

The participants of this study were healthy individuals from the Estonian Biobank of the Estonian Genome Center of the University of Tartu (Leitsalu et al., 2015). In total, 313 subjects were selected for the study, with median age 54 (standard deviation 17.8), 154 females and 159 males. The study was approved by the Ethics Review Committee of Human Research of the University of Tartu, Estonia (permission no 206/T-4, date of issue 25th August 2011) and it was carried out in compliance with the Helsinki Declaration. A written informed consent to participate in the study was obtained from each individual prior to recruitment.

All of the following steps were done according to standard protocols and manufacturer's instructions. We purified $CD4^+$ and $CD8^+$ T cells from peripheral blood mononuclear cells (PBMCs) by consecutive positive separation using microbeads and AutoMACS technology. Purified cells were then subjected to genome-wide gene expression analysis (Illumina HumanHT-12v4 BeadChips), genotyping (Illumina HumanOmniExpress BeadChips), and imputation using the 1000 Genomes reference panel. After stringent quality control and filtering, close to 6 million SNPs, and gene expression data from 38,839 probes covering 23,704 genes from 303 unique individuals (293 for $CD4^+$ and 283 for $CD8^+$ T cells) were included in the analysis.

3.2.2. Landscape of *cis*-eQTLs and their cellular specificity in T cells

To characterize the extent of genetic control of gene expression in T cells and its cellular specificity, we first searched for the association between SNPs and gene expression within 1 Mb intervals, referred to as *cis*-eQTLs. In total, we identified *cis* regulatory SNPs for approximately 10% of the tested genes at probe-level FDR < 0.05 (2,605 genes in $CD4^+$ T cells and 2,056 genes in $CD8^+$ T cells with an overlap of 1,637 genes; Fig. 1 and Table S1 in Ref. II) by using a non-parametric Spearman correlation in a tissue-by-tissue analysis framework (see section 1.3.1.1. "Tissue-by-tissue analysis").

On one hand, the similarity of the cell types is reflected in the high proportion of shared effects, $\pi_1 = 0.99$, and is confirmed by the multi-tissue analysis using Bayesian methods (see section 1.3.1.2 "Multi-tissue joint analysis"). Specifically, of 3,871 genes associated with eQTLs at FDR < 0.05 in the multi-tissue analysis, all showed very strong posterior probability to be shared between $CD4^+$ and $CD8^+$ T cells (S2 Table in Ref. II). The seemingly cell type-specific effects in the tissue-by-tissue analysis mostly reflected genes with eQTLs that have modest effects, highlighting the limitations due to incomplete power and arbitrary cut-off levels for significance.

On the other hand, only about a half of the significant *cis*-eQTLs detected in a large meta-analysis of peripheral blood from 5,000 individuals (Westra et al., 2013) could be replicated in the $CD4^+$ and $CD8^+$ T cells ($\pi_1 = 0.51$ and 0.45, respectively), indicating a high level of specificity in *cis*-eQTLs originating

from different blood cells, as well as a remarkable sensitivity despite our several fold smaller sample size.

3.2.3. Landscape of *trans*-eQTLs in T cells

For the analysis of SNPs affecting the expression of distal genes (>5 Mb apart), referred to as *trans*-eQTLs, we selected all 4,638 genome-wide significant (P-value < $5x10^{-8}$) SNPs from the GWAS catalog (Welter et al., 2014; accessed 24/03/2015). After correcting gene expression levels for *cis*-eQTL effects to increase power, we identified 36 and 40 GWAS SNPs associated with the expression levels of 209 and 378 distal genes in $CD4^+$ and $CD8^+$ T cells, respectively (overlap of 21 SNPs and 133 genes; Fig. 3A and Table S4 in Ref. II). The functions of the genes associated with the *trans*-acting GWAS SNPs highlighted their role in T1D (Ingenuity pathway analysis, $P = 4.39x10^{-5}$) and mTOR signaling ($P = 3.84x10^{-3}$) in $CD4^+$ and $CD8^+$ T cells, respectively.

Interestingly, we observed a *trans*-acting regulatory locus at 12q13.2 in $CD4^+$ and $CD8^+$ T cells with a broad-range associations with hundreds of genes. Of note, five SNPs in that region had previously also been implicated in B cells, but not in monocytes (Fairfax et al., 2012), suggesting lymphoid-specific regulation. Specifically, known susceptibility SNPs for autoimmune and inflammatory diseases, such as T1D (Hakonarson et al., 2008; Todd et al., 2007; Wellcome Trust Case Control Consortium, 2007), vitiligo (Jin et al., 2012; Tang et al., 2013), RA (Okada et al., 2014), asthma (Hirota et al., 2011), alopecia areata (Petukhova et al., 2010), and polycystic ovary syndrome (Shi et al., 2012), exert an effect on the expression levels of 187 and 351 genes in $CD4^+$ and $CD8^+$ T cells, respectively, with an overlap of 124 genes (Fig. 3B in Ref. II). Many of these genes such as *CTLA4*, the inhibitory immune checkpoint gene (Scalapino and Daikh, 2008), the human granzyme genes *GZMA, GZMB,* and *GZMH* that induce cell death to protect the host against viruses and tumors (Bots and Medema, 2006)*, GNLY* that encodes granulysin, a molecule of anti-microbial activity against microbial pathogens (Stenger et al., 1998), are highly expressed and have important roles in T cells.

After *trans*-eQTL mapping with all the variants in this region, we identified rs1131017 as the lead SNP with the lowest association P-value mapping to the oligopyrimidine tract of the 5'UTR of the ribosomal small subunit protein 26 (*RPS26*) gene. Its C allele is in strong LD with the T1D risk allele and has been reported to correlate positively with *RPS26* expression levels (Dixon et al., 2007; Schadt et al., 2008). We hypothesized that the *RPS26* gene may constitute a mechanism for the detected *trans*-eQTL effects. Namely, the oligopyrimidine tract controls the translation process by repression of many mammalian ribosomal protein genes (Levy et al., 1991). The effect of the SNP on RPS26 ribosomal distribution has been reported before, with the G allele showing higher translational efficiency by producing more protein (Li et al., 2013). RPS26 is a main component of the ribosomal region involved in the recruitment

of cellular mRNA during translational initiation and in maintenance of the path of mRNA molecules to the ribosomal exit site (Sharifulin et al., 2012). Hence, it is conceivable that the altered RPS26 protein levels may affect the stability or translational efficiency of a large number of cytosolic mRNAs. Nevertheless, the exact functional role of rs1131017 in T1D remains unknown, as well as the effect of other candidate genes in this region, such as *CDK2*, *RAB5B*, *SUOX, IKZF4* and *ERBB3*.

3.2.4. Missense variant in *IL27* as a candidate disease variant for T1D

By comparing the overlap of *trans*-eQTLs, we identified a GWAS SNP rs4788084 at 16p11.2 close to the cytokine *IL27* gene. In $CD4^+$ T cells, the T allele of the SNP is associated with lower expression of the *IRF1* ($P = 1.84 \times 10^{-9}$) and *STAT1* ($P = 2.91 \times 10^{-8}$) genes, both involved in interferon-γ signaling. The rs4788084[T] was also associated with the lower expression of *STAT1* in peripheral blood (Westra et al., 2013) and decreased risk for T1D (Barrett et al., 2009; Plagnol et al., 2011). By targeted *trans*-eQTL mapping with 829 SNPs in the *IL27* region, we found an even stronger signal for a common missense SNP rs181206 within the *IL27* gene in $CD4^+$ T cells (Fig. 4B in Ref. II) and reduced expression levels of *IRF1*, *STAT1*, and *REC8* (Fig 4C in Ref. II) for its G allele. Importantly, we also confirmed the absence of the signal in B cells and monocytes by re-analyzing the data from Fairfax et al. (2012).

Notably, the *trans*-eQTL locus had no effects on the expression level of *IL27* in *cis*. In agreement with the cell type-specific effect, we found higher expression levels of the *IL27RA* and *IL6ST* (gp130) genes, which act together as a receptor for the IL-27 cytokine, in $CD4^+$ T cells compared to $CD8^+$ T cells (Fig. 4D in Ref. II), suggesting a suitable cellular environment for downstream effects of the missense SNP particularly in $CD4^+$ T cells.

To get more insights into the region, we assessed whether there is a single variant in the region driving both the GWAS and eQTL signals by a Bayesian test for colocalization (Giambartolomei et al., 2014). Indeed, T1D susceptibility and changes in the expression of *IRF1* and *STAT1* were colocalized, and moreover, with the most posterior support for rs181206 to be the shared causal variant. Next, given the strong positive correlation between the expression levels of *IRF1* and *STAT1*, we used structural equation modelling to determine whether the *IL27* SNP rs181206 affects both genes independently or via each other. Overall, the best-fitting scenario suggested that *IRF1* mediates the SNP and *STAT1* relationship. The finding was supported by a simulation experiment and suggested a mechanism for the effect of IL-27 on *IRF1* and *STAT1* expression (Fig. 4).

Specifically, IL-27 is produced by innate immune cells. After forming a heterodimer with EBI3, it interacts with its receptor IL27RA, and activates the STAT1/STAT3 pathway in T cells (Yoshida and Hunter, 2015). After binding to interferon-stimulated response elements, the STAT1/STAT3 pathway induces

transcription of several interferon-induced genes, including *IRF1* and *STAT1* itself. Moreover, IRF1 is a transcription factor that enhances the expression of the *STAT1* gene. Additionally, the effect of IL-27 in T cells is regarded as anti-inflammatory but it has also been shown as a growth and survival factor for T cells (Yoshida and Hunter, 2015).

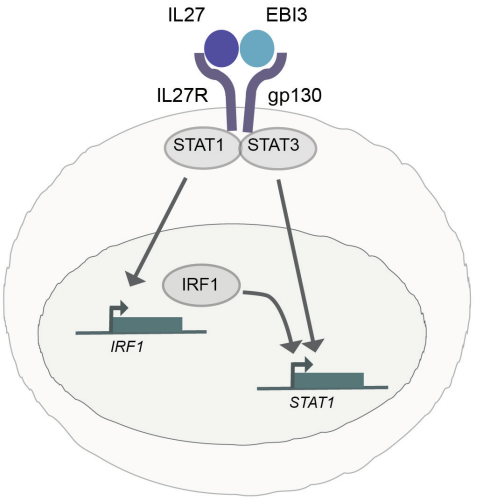

Figure 4. A simplified scheme of the role of IL-27 in the activation of the transcription of *IRF1* and *STAT1*.

Based on the statistical evidence, we hypothesized that the mutated form of IL-27 has a decreased capacity to activate the STAT1/STAT3 pathway, resulting in lower expression of *IRF1* and *STAT1*. The aforementioned G allele, in strong LD with the protective allele for T1D, causes an amino acid change (Leu119Pro) in the alpha-helical domain of IL-27. Therefore, we cloned cDNA variants of the IL-27 wild-type (Leu119) and missense (Pro119), as well as EBI3. After transfection into HEK293 cells, we combined the cell supernatants containing either the IL-27 Leu119 or the Pro119 protein with an equal amount of EBI3 protein and studied their effect on IRF1 and STAT1 expression using real-time PCR in human PBMCs from four healthy individuals. Indeed, the missense SNP resulting in Pro119 in IL-27 induced significantly lower STAT1 and IRF1 transcript levels compared to the IL-27 Leu119 ($P = 5.52\text{x}10^{-13}$ for IRF1 and $P = 4.86\text{x}10^{-10}$ for STAT1, Fig. 4E in Ref. II). Our results also suggest that IL-27 may promote autoimmunity toward pancreatic islets via the upregulation of the STAT1/STAT3 pathway.

Furthermore, our findings are supported by studies of a T1D mouse model with high levels of IL-27 and delayed T1D onset after treatment with an IL-27 blocking antibody (Wang et al., 2008). Altogether, these results suggest that the G allele of the rs181206 missense variant in the *IL27* gene confers protection against T1D through the inhibited expression of *IRF1* and *STAT1* in $CD4^+$ T cells.

3.3. Regulation of gene expression by miRNA binding site polymorphisms (Ref. III)

A majority of eQTL studies have performed inference on steady-state gene expression levels. However, the balance between mRNA transcription (including transcript initiation, elongation, and processing) and mRNA decay (including spontaneous and targeted degradation of transcripts) determine the steady-state gene expression levels (Dori-Bachash et al., 2011). While the emerging body of work have focused on the role of transcription factor binding, DNA methylation, histone modification, chromatin accessibility, and splicing in variation in gene expression, the extent of the contribution of altered miRNA binding sites and AU-rich elements affecting mRNA decay rate among genes is unclear (Pai et al., 2012).

MiRNAs, small endogenous non-coding RNAs approximately 22-nucleotides long, are well-established as negative regulators of gene expression. Changes in mRNA levels closely reflect the impact of miRNAs on gene expression (Guo et al., 2010). MiRNAs bind to specific sequence motifs called miRNA response elements (MREs) in the 3'UTR of mRNAs, repressing the activity of their targets by affecting mRNA stability and/or protein translation. Also, SNPs located in MREs may influence mRNA expression post-transcriptionally by disrupting or creating functional miRNA binding sites, by changing the effectiveness of MREs, or by replacing the binding site of one miRNA with that of another.

Following the importance of miRNA-mediated gene expression variation, the last part of the thesis dissects the role of polymorphisms in miRNA response elements (MRE-SNPs) that either disrupt miRNA binding sites or create new ones and their effect on the allele-specific expression of target genes. By integrating public eQTL data, miRNA binding site predictions, and small RNA sequencing, we identify genetic variants that can affect gene expression by modulating miRNA binding efficiency and provide possible causative mechanisms for associations with complex traits.

3.3.1. Identification of MRE-SNPs affecting gene expression

To create a list of MRE-SNPs affecting gene expression, *cis*-eQTLs in peripheral blood from the most comprehensive report at the time of the study (Westra et al., 2013; meta-analysis of seven cohorts yielding a sample size of 5,311 individuals and detecting 664,097 unique SNP-probe pairs at probe-level FDR < 0.05) and their perfect proxies ($R^2 = 1$, 1000G CEU population) were mapped to 3'UTR of the *cis*-regulated genes. In total, we identified 9,334 *cis*-acting SNPs and their perfect proxies residing in the 3'UTR of the corresponding *cis*-affected transcript, referred to as UTR-SNP-probe associations (Fig.1 in Ref. III and a detailed analysis workflow in Fig. 5).

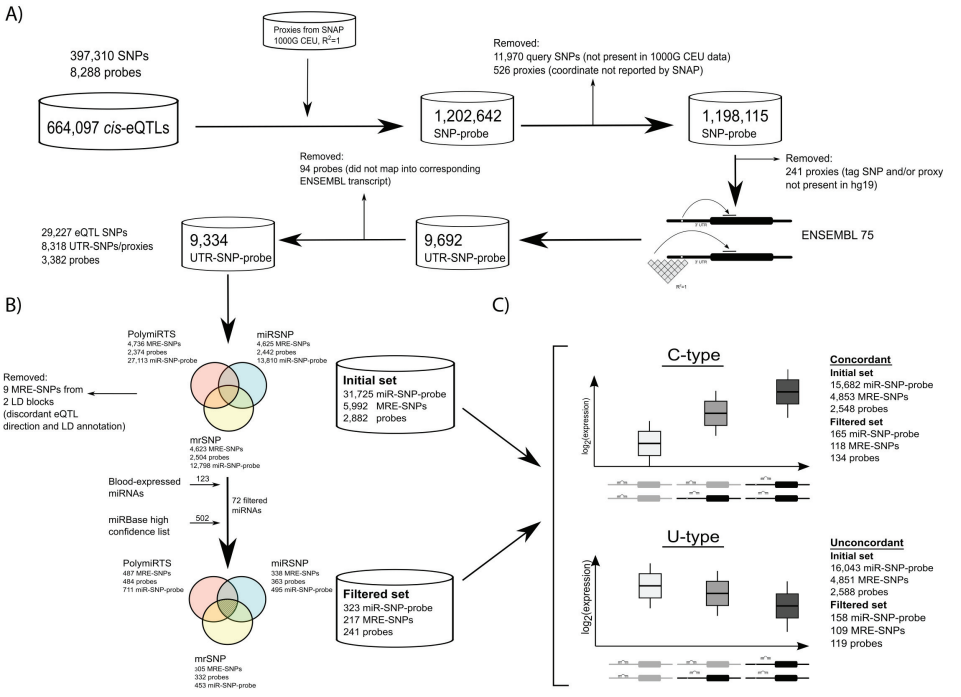

Figure 5. Detailed analysis workflow. (A) *Cis*-eQTLs (FDR < 0.05) were acquired from Westra et al. (2013). Perfect proxies (1000G CEU, $R^2 = 1$) were added using the SNAP v2.2 webtool, and the resulting associations are referred to as SNP-probe associations. *Cis*-acting SNPs and their perfect proxies mapping into the 3'UTR of corresponding *cis*-affected transcripts were used in subsequent analyses and are indicated as UTR-SNP-probe associations. (B) UTR-SNP-probe associations were intersected against resources containing information about *in silico* predicted MRE-SNPs. The resulting miR-SNP-probe associations (unfiltered set) were then filtered based on: i) overlap between all three of the target prediction methods, ii) their inclusion in a "blood-expressed" miRNA consensus list, and iii) their presence in the miRBase "high confidence" list. (C) Unfiltered and filtered sets of miR-SNP-probe associations were queried for concordance using the logic of miRNA-mediated regulation and classified as either concordant (C-type) or unconcordant (U-type).

Next, to identify the SNPs that have the potential to disrupt an existing MRE or create a new MRE, referred to as MRE-SNPs, the UTR-SNP-probe associations were intersected with public databases containing information about *in silico* predicted MRE-SNPs. As a result, in union over the databases, we identified 5,992 *cis*-eQTL SNPs and proxies in the putative MREs of 2,545 genes, which corresponds to 72% of all 8,318 SNPs and proxies mapped to 3'UTRs of the affected genes, and accounted for approximately 4% of all SNP-probe pairs identified by Westra et al. (2013). The expression of those genes was detected by 2,882 unique Illumina probes, which made up ~35% of 8,201 probes with *cis*-eQTLs identified by Westra et al. (2013). Also, the binding of almost all

miRNAs present in the miRBase (v20) was predicted to be influenced by at least one of the SNPs (2,573 out of 2,578 miRNAs).

Among all MRE-SNPs, 1,137 were predicted to be exclusively MRE-disrupting, 1,191 exclusively MRE-creating, and 3,664 both MRE-disrupting and MRE-creating. The average number of MREs affected per SNP was 4.72, verifying that most of the 3'UTRs contain binding sites for several different miRNAs, and quite often, there is more than one site for a specific miRNA. Altogether, 31,725 putative miRNA-SNP-probe associations were identified and the average number of MRE-SNPs per miRNA was 11.15.

However, as each of the prediction algorithms has its limitations, we narrowed the 5,992 putative MRE-SNPs (unfiltered) down to a prioritized set using more stringent filtering criteria: i) intersection of the three target prediction methods; ii) their inclusion in a "blood-expressed" miRNA consensus list, based on publicly available small RNA sequencing datasets, as the original eQTL mapping was performed using peripheral blood samples; iii) their presence in the miRBase "high confidence" list showing the high quality of miRNA annotations. Applying these criteria, we identified a filtered set of 323 miRNA-SNP-probe associations consisting of 217 MRE-SNPs, 57 miRNAs, and 241 probes corresponding to 206 genes. Of them, the minor allele was disrupting 163 MREs and creating 160 MREs (Fig. 2 in Ref. III for an example subset of filtered associations).

3.3.2. Concordance with miRNA-mediated regulation

A functional miRNA and mRNA interaction is assumed to result in the down-regulation of the target transcript. So, an MRE-SNP disrupting the miRNA binding site should result in the upregulation of the target mRNA and therefore, show a positive correlation with the target mRNA. Following this rationale, we assessed the possibility of a miRNA-mediated regulation of *cis*-eQTLs. We refer to combinations where the correlation between expression levels and MRE-SNP genotypes follows the logic of miRNA-mediated mechanism as "concordant" and to combinations with the opposite effect as "unconcordant".

Among both unfiltered and filtered miRNA-SNP-probe pairs, roughly 50% were concordant (49.4% for unfiltered and 51% filtered sets), indicating no significant overrepresentation of MRE-SNPs for which the *cis*-eQTL direction and effect on miRNA binding would be in line. The effect sizes did not differ between exclusively concordant and exclusively unconcordant *cis*-eQTLs. Likewise, we did not observe any significant association between the average effect of exclusively MRE-breaking or MRE-creating SNPs. Also, there was no significant correlation between the miRNA regulation concordance and miRNA binding efficiency.

The ambiguity in our results can be explained by several factors. First, most of the 3'UTRs contain binding sites for several different miRNAs, and quite often, there is more than one site for a specific miRNA. This means that the

effect of disrupting or creating a single binding site may be reduced by the action of other sites. Second, the effect of a 3'UTR SNP can be manifested through different mechanisms, since both the miRNA binding and mRNA stability in general are affected by several different factors. These mechanisms may include alternative polyadenylation or splicing, mRNA decay, mRNA structural alterations, or the accessibility to an RNA-induced silencing complex. Some of those mechanisms are already addressed, suggesting that the majority of 3'UTR SNPs influence MREs rather than splicing sites or 3'UTR folding (Arnold et al., 2012). Compared to all 3'UTR SNPs, the SNPs within miRNA binding sites are enriched for associations with variation in decay rates (Pai et al., 2012).

3.3.3. Impact of complex traits-associated SNPs within MREs

Regardless of the unclear role of SNPs in miRNA response elements, we identified four trait-associated concordant-type MRE-SNPs as a proof of concept, for which three variants were related to cancer (Fig. 4 in Ref. III). Altogether, comparing 5,992 putative MRE-SNPs and their proxies against the catalog of published GWAS associations revealed an overlap between 208 (3.5%) MRE-SNPs and 154 GWAS SNPs or their proxies. The filtered set of MRE-SNPs contains 10 (4.6%) variants, associated with 12 traits that overlap with GWAS hits.

One of the most interesting findings of the four is the rs4245739 SNP. The minor C allele creates a functionally verified MRE for miR-191-5p in the *MDM4* gene (Fig. 3D and 3F, Fig. 4A in Ref. III). The *MDM4*-encoded protein inhibits the tumor suppressor p53 post-translationally and is upregulated in tumors (Bartel et al., 2005; Han et al., 2007). The minor allele of rs4245739, carried by approximately 20% of the European population, is associated with a protective effect for several cancers (Eeles et al., 2013; Garcia-Closas et al., 2013; Purrington et al., 2014; Wynendaele et al., 2010), and may serve as a potential biomarker. Most importantly, the effect of rs4245739 on miR-191-5p binding and subsequent down-regulation of *MDM4* mRNA and protein expression has been experimentally verified in ovarian cancer cell lines (Wynendaele et al., 2010), serving as an example of a functional MRE-SNP identified independently of our systematic genome-wide approach.

The second gene containing both a GWAS hit and an MRE-SNP is *N4BP1*. The SNP rs6500395, located in the first intron of *N4BP1*, has been associated with the response of rheumatoid arthritis patients to tocilizumab treatment (Wang et al., 2013). This gene also contains an AGO-CLIP-supported C-type MRE-SNP proxy (rs1224) for miR-330-3p in its 3'UTR (Fig. 4C in Ref. III).

In the last two cases, the absolute proxies of *cis*-eQTLs were located in the 3'UTR MRE of a nearby gene under genetic regulation (Fig. 4B in Ref. III). The intronic region of the *FARP2* gene consists of the SNP rs3771570, associated with aggressive prostate cancer (Eeles et al., 2013). Its perfect proxy, rs1056801, within the 3'UTR of a gene next to it, *SEPT2*, disrupts the binding

of cancer-associated miR-17-92 family members. Aberrant expression of *SEPT2* has been reported in different tumor types (Liu et al., 2010). As *SEPT2* is also the only gene influenced by a significant *cis*-eQTL in a corresponding LD block, we propose that MRE-SNP-mediated alterations in the binding of miR-17-92 family of miRNAs may be related to abnormal expression of *SEPT2*.

In the esophageal squamous cell carcinoma susceptibility region tagged by rs2239815 (Wu et al., 2012), we identified an MRE-SNP within the 3'UTR of *CCDC117* (Fig. 4D in Ref. III). This LD block contains two apparent candidate genes for cancer susceptibility (*XBP1* and *CHEK2*). Although all three of these genes are affected by *cis*-eQTLs, the largest effect of this LD block is associated with the *XBP1* gene, casting doubt on the miRNA-mediated *cis*-eQTL mechanism.

CONCLUSIONS

Knowing the genetic regulators that modulate gene expression contributes to our understanding about the mechanistic basis of variation in human phenotypes. Comprehensive eQTL maps shed light on the relevant tissues or cell types, molecular mechanisms and pathways through which susceptibility variants exert their effects, and provide bases for further explorative studies of potential drug targets and other disease interventions for the benefit of human health.

The main conclusions drawn from this thesis are as follows:

- The transcriptome and epigenome of the fetal liver depicts its role in developmental and hematopoietic processes, whereas the genes active in the adult liver are mainly responsible for catabolic and metabolic processes, well in line with the known developmental function of these tissues. Genes with eQTL effects in adults show strong liver-specificity and are enriched for genes encoding drug metabolizing enzymes. Inter-individual variability in hepatic gene expression is affected by both genetic and epigenetic factors, whereas, on average, genetic variants uniquely explain a greater proportion of the variation compared to DNA methylation levels.
- The expression of genes in $CD4^+$ and $CD8^+$ T cells are under a substantial genetic control. Thereby, patterns of local genetic control of gene expression in $CD4^+$ and $CD8^+$ T cells are highly similar, but show differences from profiles in whole blood. The advantage of studying purified cell types instead of a bulk tissue, like whole blood, is illustrated by the discovery of a $CD4^+$-specific *trans*-eQTL. Namely, the missense SNP in the *IL27* gene has a decreased capacity to activate the STAT1/STAT3 pathway, resulting in lower expression of the *IRF1* and *STAT1* genes, which plausibly confers protection against T1D.
- *Cis*-eQTLs disrupting or creating miRNA binding sites that are concordant with the logic of miRNA-mediated regulation of gene expression were not overrepresented among all of the *cis*-eQTLs or their perfect proxies harboring miRNA binding sites. This shows the complexity of post-transcriptional regulation of gene expression. However, the prioritized list of potentially miRNA-driven *cis*-eQTL effects show the importance of miRNAs in the regulation of gene expression in cancer.

SUMMARY IN ESTONIAN

Geeniekspressiooni geneetiline regulatsioon: koe- ja rakutüübi-spetsiifiliste efektide leidmine

Tänu sekveneerimistehnoloogia kiirele arengule sai võimalikuks inimese genoomi DNA järjestuse määramine 2000. aastate alguses. Sealt edasi on toimunud tohutu hüpe erinevate tunnuste ja komplekshaiguste kujunemist mõjutavate geneetiliste markerite ülegenoomsel kaardistamisel. Enamik leitud variantidest asuvad genoomi mittekodeerivates alades, mõjutades pigem geenide avaldumist kui geenikspressiooni lõpp-produkti valku.

Geenikspressiooni tase on vahepealne tunnus DNAs peituva info ja keskkonna toimel avalduva fenotüübi vahel. Erinevates rakutüüpides on geenide avaldumine erinevalt reguleeritud ja see sõltub nii keskkonnamõjudest kui ka rakus toimuvatest protsessidest. Geenide avaldumise regulatsioonil on tähtis osa geneetilistel markeritel. Nad mõjutavad nii transkriptsiooni, muutes transkriptsioonifaktorite seondumise efektiivsust promootor- või enhanseraladele, RNA töötlemist splaissingu näol kui ka transkriptsioonijärgselt mikroRNAsid aktiveerides, millede seondumisel mRNAga toimub geenide vaigistamine.

Viimastel aastatel on rohkesti kogutud ja täiendatud infot inimese geenikspressiooni kvantitatiivsete tunnusete lookuste ehk eQTLide (*expression quantitative trait loci*) kohta. eQTLi näol on tegemist ühe nukleotiidi polümorfismiga (*single nucleotide polymorphism*, SNP), mille genotüüp korreleerub geeniekspressiooni tasemega. Ühest küljest aitavad eQTLid selgitada ülegenoomse olulisusega mittekodeerivate SNPide funktsionaalset rolli, sidudes need muutustega geenikspressioonis. Teisalt, vaadates eQTLide asukohti genoomis ja epigenoomis, aitab see kirjeldada bioloogilist mehhanismi, mille kaudu SNP mõjutab geeni ja seeläbi huvipakkuva tunnuse või haiguse kujunemist. Seega, eQTL-uuringud on hea lisavahend, mis aitab õppida geneetilist koodi funktsionaalselt lugema ja seda õigesti tõlgendama inimeste tervise hüvanguks.

Käesolev doktoritöö uuris koe- ja rakutüübi-spetsiifilise geenikspressiooni geneetilist regulatsiooni. Teaduskirjanduse põhjal anti ülevaade eQTLidest, keskendudes nende bioloogilisele olemusele, omadustele ning tähtsusele inimese genoomikas. Lisaks kirjeldati eQTLide analüüsiraamistikku ning statistilisi analüüsimeetodeid, mida kasutatakse eQTLide leidmiseks nii üksikus koes või rakus kui ka analüüsides neid koos. Samuti toodi välja tulevikuperspektiivid, mis aitaksid eQTL-uuringute tulemusi kasutada personaalses meditsiinis. Töö eksperimentaalosa eesmärk oli kirjeldada geenikspressiooni regulatsiooni maksakoes ja T-rakkudes ning hinnata mikroRNAde seondumisalades asuvate SNPide rolli geenide avaldumises.

Maksa tähtsate funktsioonide hulka kuuluvad nii imendunud toitainete ja vitamiinide töötlemine kui ka organismile kahjulike produktide eemaldamine. Võrreldes omavahel loote ja täiskasvanu maksakoe geenikspressiooni ja DNA metülatsiooni mustreid, leidsime mitmeid arengufaasile iseloomulikke muutusi. Täiskasvanu maksas ülesreguleeritud või hüpometüleeritud geenid on seotud

metaboolsete funktsioonide täitmisega, loote maksas ülesreguleeritud või hüpometüleeritud geenid seotud aga diferentseerumise ja kasvuga ning vereloome protsessidega, mis on kooskõlas sellega, et loote arengu ajal täidab maks vereloomeorgani ülesandeid. Lisaks leidsime, et geenid, mille ekspressioonitase korreleerub geneetiliste markeritega, on tugevalt maksa-spetsiifilised ja nende hulgas on üleesindatud geenid, mis kodeerivad ravimeid metaboliseerivaid ensüüme. Oma uuringu tulemusena kirjeldasime nii geneetilisi kui ka epigeneetilisi faktoreid, mis mõjutavad geeniekspressiooni regulatsiooni. Saadud teadmised aitavad paremini mõista erinevusi ravimi vastustes ja maksahaiguste etioloogias.

Kui maksal on täita tähtis roll inimese tervise seisukohalt üldiselt, siis $CD4^+$ ja $CD8^+$ T-rakud on olulised elemendid omandatud immuunsüsteemis. $CD4^+$ T-rakud juhivad üldist immuunvastuse kujunemist ja $CD8^+$ T-rakud osalevad patogeenide tõrjes infektsioonide ajal. $CD4^+$ ja $CD8^+$ T-rakkude sobimatu või puudulik regulatsioon võib soodustada mitmete autoimmuun- ja põletikuliste haiguste teket ja edasist kulgu. Me uurisime perifeersest verest eraldatud $CD4^+$ ja $CD8^+$ T-rakkude geeniekspressiooni ja geneetiliste markerite vahelist seost. Leidsime, et ligi 10% uuritud geenide avaldumine on mõjutatud lähedalasuvate SNPide poolt. Seejuures on $CD4^+$ ja $CD8^+$ T-rakud väga sarnaselt reguleeritud, kuid kõigest pool täisveres leitud SNP-geen seostest on olemas ka uuritud T-rakkudes. See näitab, et rakkude puhastamine täisverest on küll keerukas ja kulukas, kuid vajalik protseduur, et leida rakutüübi-spetsiifilisi efekte. Leidsime ka selliseid SNPe, mis mõjutavad temast kaugemal või hoopis teisel kromosoomil asuvate geenide avaldumist. Näiteks, 16. kromosoomil *IL27* geenis asub esimest tüüpi diabeeti haigestumisriskiga seotud mittesünonüümne SNP, mis ainult $CD4^+$ T-rakkudes mõjutab immuunsüsteemi seisukohalt kahe olulise geeni avaldumist: *STAT1* (kromosoom 2) ja *IRF1* (kromosoom 6). Funktsionaalse katsega näitasime, et leitud SNP-geenide seos ei ole üksnes statistiliselt oluline korrelatsioon, vaid need seosed on olemas ka inimese perifeerse vere mononukleaarsetes rakkudes. Meie hüpoteesi kohaselt mittesünonüümne SNP pärsib IL-27 seondumist oma retseptoriga, mistõttu on häiritud STAT1/STAT3 geeniraja aktivatsioon. Selle tulemusel väheneb geenide *STAT1* ja *IRF1* ekspressioon $CD4^+$ T-rakkudes, mis annab võimaliku kaitse esimest tüübi diabeedi suhtes. Antud hüpoteesi kinnitab ka esimest tüüpi diabeedi hiiremudel, kus näidati, et hiirtel, keda raviti IL-27 valku blokeeriva antikehaga, oli kõrgem IL-27 tase ja esimest tüübi diabeedi sümptomid avaldusid hiljem võrreldes nende hiirtega, kes ravi ei saanud. Kokkuvõtlikult leidsime uuringu käigus võimaliku põhjusliku SNPi ning kirjeldasime bioloogilist mehhanismi, mis on loonud eeldused lisauuringuteks, et testida leiu potentsiaali rakendamisel kliinilises praktikas.

Selleks, et eQTL-uuringute tulemused oleks rakendatavad näiteks ravimiarenduses, on vaja teade täpset sihtmärkgeeni ja bioloogilist mehhanismi. Transkriptsiooni-järgselt pärsivad geeniekspressiooni näiteks mikroRNAd. MikroRNAd on väikesed, umbes 22 nukleotiidi pikkused RNA molekulid, mis seonduvad mRNA transkriptide komplementaarsetele järjestustele. Me uurisime

süstemaatiliselt, kas mikroRNAde seondumisalades olevate ja geenieksprессiooni mõjutavate SNPide roll geenide üles- ja allareguleerimisel läheb kokku funktsionaalse mikroRNA rolliga sihtmärkgeeni ekspressioonile. Kogutud andmed ei tõendanud, et SNPid mikroRNAde seondumisalades oleks kooskõlas mikroRNAde toimemehhanismi loogikaga. See annab alust arvata, et geeniekspressiooni transkriptsiooni-järgne regulatsioon on väga kompleksne ning mikroRNAde seondumisalades asuvad SNPid võivad omakorda olla mõjutatud teiste mehhanismide poolt. Siiski, meil õnnestus leida mitu juhtu, kus haigusseoselise SNPi mõju geenieksprессioonile toetab mikroRNAde rolli geenide avaldumisele.

REFERENCES

Abraham, G., Havulinna, A.S., Bhalala, O.G., Byars, S.G., De Livera, A.M., Yetukuri, L., Tikkanen, E., Perola, M., Schunkert, H., Sijbrands, E.J., et al. (2016). Genomic prediction of coronary heart disease. Eur. Heart J. *37*, 3267–3278.

Adams, L.A., White, S.W., Marsh, J.A., Lye, S.J., Connor, K.L., Maganga, R., Ayonrinde, O.T., Olynyk, J.K., Mori, T.A., Beilin, L.J., et al. (2013). Association between liver-specific gene polymorphisms and their expression levels with nonalcoholic fatty liver disease. Hepatology *57*, 590–600.

Aguet, F., Brown, A.A., Castel, S., Davis, J.R., Mohammadi, P., Segre, A. V, Zappala, Z., Abell, N.S., Fresard, L., Gamazon, E.R., et al. (2016). Local genetic effects on gene expression across 44 human tissues. bioRxiv 74450.

Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Consortium, H., Hale, C., Dougan, G., and Gaffney, D.J. (2017). Genetic effects on chromatin accessibility foreshadow gene expression changes in macrophage immune response. bioRxiv 102392.

Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. *16*, 197–212.

Alberts, R., Terpstra, P., Li, Y., Breitling, R., Nap, J.-P., and Jansen, R.C. (2007). Sequence polymorphisms cause many false cis eQTLs. PLoS One *2*, e622.

Andiappan, A.K., Melchiotti, R., Poh, T.Y., Nah, M., Puan, K.J., Vigano, E., Haase, D., Yusof, N., San Luis, B., Lum, J., et al. (2015). Genome-wide analysis of the genetic regulation of gene expression in human neutrophils. Nat. Commun. *6*, 7971.

Arnold, M., Ellwanger, D.C., Hartsperger, M.L., Pfeufer, A., and Stümpflen, V. (2012). Cis-Acting Polymorphisms Affect Complex Traits through Modifications of MicroRNA Regulation Pathways. PLoS One *7*, e36694.

Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C., et al. (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat. Genet. *41*, 703–707.

Bartel, F., Schulz, J., Böhnke, A., Blümke, K., Kappler, M., Bache, M., Schmidt, H., Würl, P., Taubert, H., and Hauptmann, S. (2005). Significance of HDMX-S (or MDM4) mRNA splice variant overexpression and HDMX gene amplification on primary soft tissue sarcoma prognosis. Int. J. Cancer *117*, 469–475.

Battle, A., and Montgomery, S.B. (2014). Determining causality and consequence of expression quantitative trait loci. Hum. Genet. *133*, 727–735.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Stamatoyannopoulos, J.A., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Bots, M., and Medema, J.P. (2006). Granzymes at a glance. J. Cell Sci. *119*.

Breitling, R., Li, Y., Tesson, B.M., Fu, J., Wu, C., Wiltshire, T., Gerrits, A., Bystrykh, L. V., de Haan, G., Su, A.I., et al. (2008). Genetical Genomics: Spotlight on QTL Hotspots. PLoS Genet. *4*, e1000232.

Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science (80-.). *296*, 752–755.

Brown, M.S., and Goldstein, J.L. (1996). Heart attacks: gone with the century? Science *272*, 629.

Brown, C.D., Mangravite, L.M., Engelhardt, B.E., Blanche, P., and Waters, D. (2013). Integrative Modeling of eQTLs and Cis-Regulatory Elements Suggests Mechanisms Underlying Cell Type Specificity of eQTLs. PLoS Genet. *9*, e1003649.

Brynedal, B., Choi, J., Raj, T., Bjornson, R., Stranger, B.E., Neale, B.M., Voight, B.F., and Cotsapas, C. (2017). Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation. Am. J. Hum. Genet. *100*, 581–591.

Bryois, J., Buil, A., Evans, D.M., Kemp, J.P., Montgomery, S.B., Conrad, D.F., Ho, K.M., Ring, S., Hurles, M., Deloukas, P., et al. (2014). Cis and Trans Effects of Human Genomic Variants on Gene Expression. PLoS Genet. *10*, e1004461.

Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-Wide Association Studies. PLoS Comput. Biol. *8*, e1002822.

Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell *167*, 1398–1414.e24.

Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Brüning, J.C., Nolan, P.M., Ashcroft, F.M., et al. (2010). Overexpression of Fto leads to increased food intake and results in obesity. Nat. Genet. *42*, 1086–1092.

Civelek, M., and Lusis, A.J. (2013). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

Claussnitzer, M., Dankel, S.N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., et al. (2014). Leveraging Cross-Species Transcription Factor Binding Site Patterns: From Diabetes Risk Loci to Disease Mechanisms. Cell *156*, 343–358.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/ Cas systems. Science *339*, 819–823.

Cooper, G.M. (2000). The Cell: A Molecular Approach. 2nd edition (Sunderland (MA): Sinauer Associates).

Dermitzakis, E.T. (2008). From gene expression to disease risk. Nat. Genet. *40*, 492–493.

Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L.M.S., Kiess, W., Vatin, V., Lecoeur, C., et al. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. Nat. Genet. *39*, 724–726.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., et al. (2007). A genome-wide association study of global gene expression. Nat. Genet. *39*, 1202–1207.

Dori-Bachash, M., Shema, E., and Tirosh, I. (2011). Coupled Evolution of Transcription and mRNA Degradation. PLoS Biol. *9*, e1001106.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: Illuminating the Dark Road from Association to Function. Am. J. Hum. Genet. *93*, 779–797.

Eeles, R.A., Olama, A.A. Al, Benlloch, S., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghoussaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S., et al. (2013). Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat. Genet. *45*, 385–391, 391-2.

Ellinghaus, D., Folseraas, T., Holm, K., Ellinghaus, E., Melum, E., Balschun, T., Laerdahl, J.K., Shiryaev, A., Gotthardt, D.N., Weismüller, T.J., et al. (2013). Genome-wide association analysis in Primary sclerosing cholangitis and ulcerative colitis identifies risk loci at *GPR35* and *TCF4*. Hepatology *58*, 1074–1083.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science (80-.). *306*, 636–640.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Fairfax, B.P., and Knight, J.C. (2014). Genetics of gene expression in immunity to infection. Curr. Opin. Immunol. *30*, 63–71.

Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O., and Knight, J.C. (2012). Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. Nat. Genet. *44*, 502–510.

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., et al. (2014). Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. Science (80-.). *343*, 1246949–1246949.

Fehrmann, R.S.N., Jansen, R.C., Veldink, J.H., Westra, H.-J., Arends, D., Bonder, M.J., Fu, J., Deelen, P., Groen, H.J.M., Smolonska, A., et al. (2011). Trans-eQTLs Reveal That Independent Genetic Variants Associated with a Complex Phenotype Converge on Intermediate Genes, with a Major Role for the HLA. PLoS Genet. *7*, e1002197.

Ferraro, A., D'Alise, A.M., Raj, T., Asinovski, N., Phillips, R., Ergun, A., Replogle, J.M., Bernier, A., Laffel, L., Stranger, B.E., et al. (2014). Interindividual variation in human T regulatory cells. Proc. Natl. Acad. Sci. U. S. A. *111*, E1111-20.

Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Brüning, J.C., and Rüther, U. (2009). Inactivation of the Fto gene protects from obesity. Nature *458*, 894–898.

Flutre, T., Wen, X., Pritchard, J., Stephens, M., Frazer, K., Murray, S., Schork, N., Topol, E., Montgomery, S., Dermitzakis, E., et al. (2013). A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. PLoS Genet. *9*, e1003486.

Fraser, H.B., and Xie, X. (2009). Common polymorphic transcript variation in human disease. Genome Res. *19*, 567–575.

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007). A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. Science (80-.). *316*.

Fu, J., Wolfs, M.G.M., Deelen, P., Westra, H.-J., Fehrmann, R.S.N., Te Meerman, G.J., Buurman, W.A., Rensen, S.S.M., Groen, H.J.M., Weersma, R.K., et al. (2012). Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. *8*, e1002431.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer–promoter connections with CRISPR interference. Science (80-.). *354*, 769–773.

Gaffney, D.J. (2013). Global Properties and Functional Complexity of Human Gene Regulatory Variation. PLoS Genet. *9*, e1003501.

Garcia-Closas, M., Couch, F.J., Lindstrom, S., Michailidou, K., Schmidt, M.K., Brook, M.N., Orr, N., Rhie, S.K., Riboli, E., Feigelson, H.S., et al. (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat. Genet. *45*, 392–398, 398-2.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat. Rev. Mol. Cell Biol. *8*, 113–126.

Gelman, A., and Hill, J. (2006). Applied regression and multilevel/hierarchical models (Cambridge; New York: Cambridge University Press, 1 edition).

Ghosh, S., Yates, A.J., Frühwald, M.C., Miecznikowski, J.C., Plass, C., and Smiraglia, D. (2010). Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics *5*, 527–538.

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. *10*, e1004383.

Gibson, G., Powell, J.E., and Marigorta, U.M. (2015). Expression quantitative trait locus analysis for translational medicine. Genome Med. *7*, 60.

Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabási, A.-L. (2007). The human disease network. Proc. Natl. Acad. Sci. U. S. A. *104*, 8685–8690.

Grundberg, E., Small, K.S., Hedman, Å.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.-P., Meduri, E., Barrett, A., et al. (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. *44*, 1084–1089.

GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science (80-.). *348*, 648–660.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Hakonarson, H., Qu, H.-Q., Bradfield, J.P., Marchand, L., Kim, C.E., Glessner, J.T., Grabs, R., Casalunovo, T., Taback, S.P., Frackelton, E.C., et al. (2008). A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes *57*, 1143–1146.

Han, X., Garcia-Manero, G., McDonnell, T.J., Lozano, G., Medeiros, L.J., Xiao, L., Rosner, G., Nguyen, M., Fernandez, M., Valentin-Vega, Y.A., et al. (2007). HDM4 (HDMX) is widely expressed in adult pre-B acute lymphoblastic leukemia and is a potential therapeutic target. Mod. Pathol. *20*, 54–62.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. *106*, 9362–9367.

Hirota, T., Takahashi, A., Kubo, M., Tsunoda, T., Tomita, K., Doi, S., Fujita, K., Miyatake, A., Enomoto, T., Miyagawa, T., et al. (2011). Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat. Genet. *43*, 893–896.

Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet. *99*, 1245–1260.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Huang, Q. (2015). Genetic Study of Complex Diseases in the Post-GWAS Era. J. Genet. Genomics *42*, 87–98.

Innocenti, F., Cooper, G.M., Stanaway, I.B., Gamazon, E.R., Smith, J.D., Mirkov, S., Ramirez, J., Liu, W., Lin, Y.S., Moloney, C., et al. (2011). Identification, Replication, and Functional Fine-Mapping of Expression Quantitative Trait Loci in Primary Human Liver Tissue. PLoS Genet. *7*, e1002078.

Ivanov, M., Kals, M., Kacevska, M., Barragan, I., Kasuga, K., Rane, A., Metspalu, A., Milani, L., and Ingelman-Sundberg, M. (2013). Ontogeny, distribution and potential roles of 5-hydroxymethylcytosine in human liver function. Genome Biol. *14*, R83.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33*, 245–254.

Jansen, R.C., and Nap, J.P. (2001). Genetical genomics: the added value from segregation. Trends Genet. *17*, 388–391.

Jin, Y., Birlea, S.A., Fain, P.R., Ferrara, T.M., Ben, S., Riccardi, S.L., Cole, J.B., Gowan, K., Holland, P.J., Bennett, D.C., et al. (2012). Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. Nat. Genet. *44*, 676–680.

Jo, B., He, Y., Strober, B.J., Parsana, P., Aguet, F., Brown, A.A., Castel, S.E., Gamazon, E.R., Gewirtz, A., Gliner, G., et al. (2016). Distant regulatory effects of genetic variation in multiple human tissues. bioRxiv 74419.

Joehanes, R., Zhang, X., Huan, T., Yao, C., Ying, S., Nguyen, Q.T., Demirkale, C.Y., Feolo, M.L., Sharopova, N.R., Sturcke, A., et al. (2017). Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. Genome Biol. *18*, 16.

Jones, P.A. (1999). The DNA methylation paradox. Trends Genet. *15*, 34–37.

Kacevska, M., Ivanov, M., Wyss, A., Kasela, S., Milani, L., Rane, A., and Ingelman-Sundberg, M. (2012). DNA methylation dynamics in the hepatic CYP3A4 gene promoter. Biochimie *94*, 2338–2344.

Kathiresan, S., Voight, B.F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P.M., Anand, S., Engert, J.C., Samani, N.J., Schunkert, H., et al. (2009). Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat. Genet. *41*, 334–341.

Kilpeläinen, T.O., Qi, L., Brage, S., Sharp, S.J., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C.H., et al. (2011). Physical Activity Attenuates the Influence of FTO Variants on Obesity Risk: A Meta-Analysis of 218,166 Adults and 19,268 Children. PLoS Med. *8*, e1001116.

Kim-Hellmuth, S., Bechheim, M., Puetz, B., Mohammadi, P., Nedelec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E., et al. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. bioRxiv 116376.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol.

Knight, J.C. (2005). Regulatory polymorphisms underlying complex disease traits. J. Mol. Med. *83*, 97–109.

Knight, J.C. (2014). Approaches for establishing the function of regulatory genetic variants involved in disease. Genome Med. *6*, 92.

Lage, K., Hansen, N.T., Karlberg, E.O., Eklund, A.C., Roque, F.S., Donahoe, P.K., Szallasi, Z., Jensen, T.S., and Brunak, S. (2008). A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proc. Natl. Acad. Sci. U. S. A. *105*, 20870–20875.

Läll, K., Mägi, R., Morris, A., Metspalu, A., and Fischer, K. (2017). Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. Genet. Med. *19*, 322–329.

Lappalainen, T. (2015). Functional genomics bridges the gap between quantitative genetics and molecular biology. Genome Res. *25*, 1427–1431.

Lappalainen, T., Sammeth, M., Friedländer, M.R., ‘t Hoen, P.A.C., Monlong, J., Rivas, M.A., Gonzàlez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature *501*, 506–511.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.

Lee, M.N., Ye, C., Villani, A.-C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K., et al. (2014). Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. Science (80-.). *343*, 1246980–1246980.

Leek, J.T., and Storey, J.D. (2007). Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. PLoS Genet. *3*, e161.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nat. Rev. Genet. *11*, 733–739.

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., Perola, M., Ng, P.C., Mägi, R., Milani, L., et al. (2015). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int. J. Epidemiol. *44*, 1137–1147.

Lek, M., Karczewski, K.J., Minikel, E. V., Samocha, K.E., Banks, E., Fennell, T., O’Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. Annu. Rev. Genet. *46*, 43–68.

Levy, S., Avni, D., Hariharan, N., Perry, R.P., and Meyuhas, O. (1991). Oligo-pyrimidine tract at the 5’ end of mammalian ribosomal protein mRNAs is required for their translational control. Proc. Natl. Acad. Sci. U. S. A. *88*, 3319–3323.

Li, Q., Makri, A., Lu, Y., Marchand, L., Grabs, R., Rousseau, M., Ounissi-Benkalha, H., Pelletier, J., Robert, F., Harmsen, E., et al. (2013). Genome-wide search for exonic variants affecting translational efficiency. Nat. Commun. *4*, 2260.

Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. Science (80-.). *352*.

Liblau, R.S., Wong, F.S., Mars, L.T., and Santamaria, P. (2002). Autoreactive CD8 T cells in organ-specific autoimmunity: emerging targets for therapeutic intervention. Immunity *17*, 1–6.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, M., Shen, S., Chen, F., Yu, W., and Yu, L. (2010). Linking the septin expression with carcinogenesis. Mol. Biol. Rep. *37*, 3601–3608.

Liu, X., Finucane, H.K., Gusev, A., Bhatia, G., Gazal, S., O'Connor, L., Bulik-Sullivan, B., Wright, F.A., Sullivan, P.F., Neale, B.M., et al. (2017). Functional Architectures of Local and Distal Regulation of Gene Expression in Multiple Human Tissues. Am. J. Hum. Genet. *11*, e1004958.

Lloyd-Jones, L.R., Holloway, A., McRae, A., Yang, J., Small, K., Zhao, J., Zeng, B., Bakshi, A., Metspalu, A., Dermitzakis, M., et al. (2017). The Genetic Architecture of Gene Expression in Peripheral Blood. Am. J. Hum. Genet. *100*, 228–237.

Lu, J., and Clark, A.G. (2012). Impact of microRNA regulation on variation in human gene expression. Genome Res. *22*, 1243–1254.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. *9*, 356–369.

Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E.T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. Nature *464*, 773–777.

Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. (2004). Genetic analysis of genome-wide variation in human gene expression. Nature *430*, 743–747.

Moscovitz, J., and Aleksunes, L. (2013). Establishment of Metabolism and Transport Pathways in the Rodent and Human Fetal Liver. Int. J. Mol. Sci. *14*, 23801–23827.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K. V, Li, X., Li, H., Kuperwasser, N., Ruda, V.M., et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466*, 714–719.

Naranbhai, V., Fairfax, B.P., Makino, S., Humburg, P., Wong, D., Ng, E., Hill, A.V.S., and Knight, J.C. (2015). Genomic modulators of gene expression in human neutrophils. Nat. Commun. *6*, 7545.

Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. PLoS Genet. *6*, e1000895.

Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLoS Genet. *6*, e1000888.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.

Pai, A.A., Cain, C.E., Mizrahi-Man, O., De Leon, S., Lewellen, N., Veyrieras, J.-B., Degner, J.F., Gaffney, D.J., Pickrell, J.K., Stephens, M., et al. (2012). The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. PLoS Genet. *8*, e1003000.

Pai, A.A., Pritchard, J.K., Gilad, Y., Gerber, G., and Gifford, D. (2015). The Genetic and Mechanistic Basis for Variation in Gene Regulation. PLoS Genet. *11*, e1004857.

Petukhova, L., Duvic, M., Hordinsky, M., Norris, D., Price, V., Shimomura, Y., Kim, H., Singh, P., Lee, A., Chen, W. V, et al. (2010). Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature *466*, 113–117.

Phillips, T. (2008). Regulation of Transcription and gene expression in eukaryotes. Nat. Educ. *1*, 199.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature *464*, 768–772.

Pierce, B.L., Tong, L., Chen, L.S., Rahaman, R., Argos, M., Jasmine, F., Roy, S., Paul-Brutus, R., Westra, H.-J., Franke, L., et al. (2014). Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. PLoS Genet. *10*, e1004818.

Plagnol, V., Howson, J.M.M., Smyth, D.J., Walker, N., Hafler, J.P., Wallace, C., Stevens, H., Jackson, L., Simmonds, M.J., Bingley, P.J., et al. (2011). Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. PLoS Genet. *7*, e1002216.

Price, A.L., Helgason, A., Thorleifsson, G., McCarroll, S.A., Kong, A., Stefansson, K., and Spielman, R. (2011). Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals. PLoS Genet. *7*, e1001317.

Purrington, K.S., Slager, S., Eccles, D., Yannoukakos, D., Fasching, P.A., Miron, P., Carpenter, J., Chang-Claude, J., Martin, N.G., Montgomery, G.W., et al. (2014). Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis *35*, 1012–1019.

Rader, D.J., Cohen, J., and Hobbs, H.H. (2003). Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. J. Clin. Invest. *111*, 1795–1803.

Raghavan, A., Wang, X., Rogov, P., Wang, L., Zhang, X., Mikkelsen, T.S., and Musunuru, K. (2016). High-throughput Screening and CRISPR-Cas9 Modeling of Causal Lipid-associated Expression Quantitative Trait Locus Variants. bioRxiv.

Raj, T., Kuchroo, M., Replogle, J.M., Raychaudhuri, S., Stranger, B.E., De Jager, P.L., Taylor, J., Burnett, E., Gut, I., Farrall, M., et al. (2013). Common risk alleles for inflammatory diseases are targets of recent positive selection. Am. J. Hum. Genet. *92*, 517–529.

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I., et al. (2014). Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. Science (80-.). *344*, 519–523.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Reginsson, G.W., Ingason, A., Euesden, J., Bjornsdottir, G., Olafsson, S., Sigurdsson, E., Oskarsson, H., Tyrfingsson, T., Runarsdottir, V., Hansdottir, I., et al. (2017). Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. Addict. Biol.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Rockman, M. V., and Kruglyak, L. (2006). Genetics of global gene expression. Nat. Rev. Genet. *7*, 862–872.

Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.-E., et al. (2007). Genomewide Association Analysis of Coronary Artery Disease. N. Engl. J. Med. *357*, 443–453.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355.

Scalapino, K.J., and Daikh, D.I. (2008). CTLA-4: a key regulatory point in the control of autoimmune disease. Immunol. Rev. *223*, 143–155.

Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the Genetic Architecture of Gene Expression in Human Liver. PLoS Biol. *6*, e107.

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res. *22*, 1748–1759.

Schröder, A., Klein, K., Winter, S., Schwab, M., Bonin, M., Zell, A., and Zanger, U.M. (2013). Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. Pharmacogenomics J. *13*, 12–20.

Scuteri, A., Sanna, S., Chen, W.-M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orrú, M., Usala, G., et al. (2007). Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. PLoS Genet. *3*, e115.

Sharifulin, D., Khairulina, Y., Ivanov, A., Meschaninova, M., Ven'yaminova, A., Graifer, D., and Karpova, G. (2012). A central fragment of ribosomal protein S26 containing the eukaryote-specific motif YxxPKxYxK is a key component of the ribosomal binding site of mRNA region 5' of the E site codon. Nucleic Acids Res. *40*, 3056–3065.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., et al. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat. Methods *11*, 399–402.

Shi, Y., Zhao, H., Shi, Y., Cao, Y., Yang, D., Li, Z., Zhang, B., Liang, X., Li, T., Chen, J., et al. (2012). Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat. Genet. *44*, 1020–1025.

Smemo, S., Tena, J.J., Kim, K.-H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375.

Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A., and Jaenisch, R. (2016). Parkinson-

associated risk variant in distal enhancer of α-synuclein modulates target gene expression. Nature *533*, 95–99.

Stenger, S., Hanson, D.A., Teitelbaum, R., Dewan, P., Niazi, K.R., Froelich, C.J., Ganz, T., Thoma-Uszynski, S., Melián, A., Bogdan, C., et al. (1998). An Antimicrobial Activity of Cytolytic T Cells Mediated by Granulysin. Science (80-.). *282*.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. *100*, 9440–9445.

Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M., et al. (2012). Patterns of Cis Regulatory Variation in Diverse Human Populations. PLoS Genet. *8*, e1002639.

Suhre, K., Shin, S.-Y., Petersen, A.-K., Mohney, R.P., Meredith, D., Wägele, B., Altmaier, E., CARDIoGRAM, Deloukas, P., Erdmann, J., et al. (2011). Human metabolic individuality in biomedical and pharmaceutical research. Nature *477*, 54–60.

Swain, S.L., McKinstry, K.K., and Strutt, T.M. (2012). Expanding roles for $CD4^+$ T cells in immunity to viruses. Nat. Rev. Immunol. *12*, 136–148.

Tang, X.-F., Zhang, Z., Hu, D.-Y., Xu, A.-E., Zhou, H.-S., Sun, L.-D., Gao, M., Gao, T.-W., Gao, X.-H., Chen, H.-D., et al. (2013). Association analyses identify three susceptibility Loci for vitiligo in the Chinese Han population. J. Invest. Dermatol. *133*, 403–410.

Teslovich, T.M., Musunuru, K., Smith, A. V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F., et al. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat. Genet. *39*, 857–864.

Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. *45*, 124–130.

Tscharke, D.C., Croft, N.P., Doherty, P.C., and La Gruta, N.L. (2015). Sizing up the key determinants of the CD8+ T cell response. Nat. Rev. Immunol. *15*, 705–716.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. *23*, 555–567.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five Years of GWAS Discovery. Am. J. Hum. Genet. *90*, 7–24.

Wallace, C., Rotival, M., Cooper, J.D., Rice, C.M., Yang, J.H.M., McNeill, M., Smyth, D.J., Niblett, D., Cambien, F., Cardiogenics Consortium, et al. (2012). Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. Hum. Mol. Genet. *21*, 2815–2824.

Walter, U., and Santamaria, P. (2005). CD8+ T cells in autoimmunity. Curr. Opin. Immunol. *17*, 624–631.

Wang, J., Bansal, A.T., Martin, M., Germer, S., Benayed, R., Essioux, L., Lee, J.S., Begovich, A., Hemmings, A., Kenwright, A., et al. (2013). Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. Pharmacogenomics J. *13*, 235–241.

Wang, R., Han, G., Wang, J., Chen, G., Xu, R., Wang, L., Li, X., Shen, B., and Li, Y. (2008). The pathogenic role of interleukin-27 in autoimmune diabetes. Cell. Mol. Life Sci. *65*, 3851–3860.

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*, 661–678.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001-6.

Wen, X. (2014). Bayesian model selection in complex linear systems, as illustrated in genetic association studies. Biometrics *70*, 73–83.

Wen, X. (2016). Robust Bayesian FDR Control Using Bayes Factors, with Applications to Multi-tissue eQTL Discovery. Stat. Biosci. 1–22.

Wen, X., Luca, F., Pique-Regi, R., Naranbhai, V., and Wong, D. (2015). Cross-Population Joint Analysis of eQTLs: Fine Mapping and Functional Annotation. PLOS Genet. *11*, e1005176.

Wen, X., Lee, Y., Luca, F., Pique-Regi, R., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., et al. (2016). Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors. Am. J. Hum. Genet. *98*, 1114–1129.

Westra, H.-J., and Franke, L. (2014). From genome to function by studying eQTLs. Biochim. Biophys. Acta - Mol. Basis Dis. *1842*, 1896–1902.

Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.

Westra, H.-J., Arends, D., Esko, T., Peters, M.J., Schurmann, C., Schramm, K., Kettunen, J., Yaghootkar, H., Fairfax, B.P., Andiappan, A.K., et al. (2015). Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet. *11*, e1005223.

White, R.J., and Sharrocks, A.D. (2010). Coordinated control of the gene expression machinery. Trends Genet. *26*, 214–220.

Wolfs, M.G.M., Rensen, S.S., Bruin-Van Dijk, E.J., Verdam, F.J., Greve, J.-W., Sanjabi, B., Bruinenberg, M., Wijmenga, C., van Haeften, T.W., Buurman, W.A., et al. (2010). Co-expressed immune and metabolic genes in visceral and subcutaneous adipose tissue from severely obese individuals are associated with plasma HDL and glucose levels: a microarray study. BMC Med. Genomics *3*, 34.

Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.-H., et al. (2014). Heritability and genomics of gene expression in peripheral blood. Nat. Genet. *46*, 430–437.

Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., Hu, Z., He, Z., Jia, W., Abnet, C.C., et al. (2012). Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. Nat. Genet. *44*, 1090–1097.

Wynendaele, J., Böhnke, A., Leucci, E., Nielsen, S.J., Lambertz, I., Hammer, S., Sbrzesny, N., Kubitza, D., Wolf, A., Gradhand, E., et al. (2010). An illegitimate microRNA target site within the 3' UTR of MDM4 affects ovarian cancer progression and chemosensitivity. Cancer Res. *70*, 9641–9649.

Yao, C., Joehanes, R., Johnson, A.D., Huan, T., Liu, C., Freedman, J.E., Munson, P.J., Hill, D.E., Vidal, M., and Levy, D. (2017). Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. Am. J. Hum. Genet. *100*, 571–580.

Yoon, O.K., Hsu, T.Y., Im, J.H., and Brem, R.B. (2012). Genetics and regulatory impact of alternative polyadenylation in human B-lymphoblastoid cells. PLoS Genet. *8*, e1002882.

Yoshida, H., and Hunter, C.A. (2015). The Immunobiology of Interleukin-27. Annu. Rev. Immunol. *33*, 417–443.

Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., Maouche, S., Germain, M., Lackner, K., Rossmann, H., et al. (2010). Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One *5*, e10693.

Zhernakova, D. V, Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.-J., et al. (2016). Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. *49*, 139–145.

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my supervisors Krista Fischer, Lili Milani, and Prof. Andres Metspalu for the guidance and support over the years. Krista, thank you for introducing me the world of genetics from a perspective of a biostatistician. I admire your ability to speak "statistics" understandably for those without a formal background. Lili, you are my role model as a scientist and as a person. I am doing my best to learn how to dream big! Andres, I am very grateful for the warm welcome to the family of the Estonian Genome Center. Thank you for taking good care of us all and providing excellent conditions to do high-impact science.

I thank Prof. Maido Remm for reviewing my thesis and suggesting helpful corrections. I also thank all my co-authors, I truly value your expertise and knowledge. I wish to thank Prof. Pärt Peterson and Kai Kisand for introducing me to T cells and immunology, and Prof. Lude Franke and Harm-Jan Westra for the guidance on the eQTLMapper.

I am very happy to be a part of the awesome team at the Estonian Genome Center. Thank you my dear colleagues and fellow PhD students for your support and friendly atmosphere at work.

I would like to thank my dear course mates forming the group of "Active Triin and the other passive statisticians/mathematicians". It has been a pleasure to study together with you for the exams, have our own summer days in Ruhnu and Kehra, and celebrate together the 20th anniversary of the start of the school year on September 1st. May all your confidence intervals be small! A very special thanks to Kristi, my support person and work wife, I am so grateful for your friendship.

I am deeply grateful to my family for their unconditional support and everlasting belief in me. My dear Märt, thank you for having this adventure of a lifetime with me!

PUBLICATIONS

CURRICULUM VITAE

Name: Silva Kasela
Date of birth: February 21, 1989
Contact: Estonian Genome Center, University of Tartu
Riia 23B, 51010, Tartu, Estonia
E-mail: silva.kasela@ut.ee

Education:

2013–…	PhD, Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu, Estonia
09/2012–02/2013	Exchange student, Master of Statistics, Specialization in Biostatistics, Universiteit Hasselt, Belgium
2011–2013	Master's studies (*cum laude*), Mathematical Statistics, Faculty of Mathematics and Computer Science, University of Tartu, Estonia
2008–2011	Bachelor's studies (*cum laude*), Mathematical Statistics, Faculty of Mathematics and Computer Science, University of Tartu, Estonia
1996–2008	Kilingi-Nõmme Gymnasium (graduated with honours)

Professional employment:

O7/2013–…	Estonian Genome Center, University of Tartu; specialist
02/2013-06/2013	Institute of Mathematical Statistics, Faculty of Mathematics and Computer Science, University of Tartu; course instructor of "MTMS.02.027 Probability and Mathematical Statistics"

Administrative work:

2014−…	Member of the Nordic-Baltic Region of the International Biometric Society
2014−...	Organizer of the international course "Statistical Practice in Epidemiology using R" (Tartu University)
2014−2016	Secretary of the Nordic-Baltic Region of the International Biometric Society
2014−2015	Participating in the Researches' Night program "Scientists in Schools"
2013−...	Member of the Estonian Statistics Society

Publications:

Org, E., Blum, Y., **Kasela, S.**, Mehrabian, M., Kuusisto, J., Kangas, A.J., Soininen, P., Wang, Z., Ala-Korpela, M., Hazen, S.L., et al. (2017). Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. Genome Biol. 18, 70.

Kasela, S., Kisand, K., Tserel, L., Kaleviste, E., Remm, A., Fischer, K., Esko, T., Westra, H.-J., Fairfax, B.P., Makino, S., et al. (2017). Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. PLOS Genet. 13, e1006643.

Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.-C., Ried, J.S., Zhang, W., Yang, Y., …, **Kasela S.**, et al. (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541, 81–86.

Pervjakova, N., **Kasela, S.**, Morris, A.P., Kals, M., Metspalu, A., Lindgren, C.M., Salumets, A., and Mägi, R. (2016). Imprinted genes and imprinting control regions show predominant intermediate methylation in adult somatic tissues. Epigenomics 8, 789–799.

Milani, F., Dumas, M., Matulevičius, R., Ahmed, N., and **Kasela, S.** (2016). Criteria and Heuristics for Business Process Model Decomposition. Bus. Inf. Syst. Eng. 58, 7–17.

Võsa, U., Esko, T., **Kasela, S.**, and Annilo, T. (2015). Altered Gene Expression Associated with microRNA Binding Site Polymorphisms. PLoS One 10, e0141351.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., …, **Kasela S.**, et al. (2015). The transcriptional landscape of age in human peripheral blood. Nat. Commun. 6, 8570.

Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., Kelly, T.N., Saleheen, D., Lehne, B., Leach, I.M., …, **Kasela S.** et al. (2015). Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat. Genet. 47, 1282–1293.

Tserel, L., Kolde, R., Limbach, M., Tretyakov, K**., Kasela, S.**, Kisand, K., Saare, M., Vilo, J., Metspalu, A., Milani, L., et al. (2015). Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. Sci. Rep. 5, 13107.

Westra, H.-J., Arends, D., Esko, T., Peters, M.J., Schurmann, C., Schramm, K., Kettunen, J., Yaghootkar, H., Fairfax, B.P., Andiappan, A.K., …, **Kasela S.**, et al. (2015). Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet. 11, e1005223.

Putku, M., Kals, M., Inno, R., **Kasela, S.**, Org, E., Kožich, V., Milani, L., and Laan, M. (2015). CDH13 promoter SNPs with pleiotropic effect on cardio-metabolic parameters represent methylation QTLs. Hum. Genet. 134, 291–303.

Bonder, M.J.*, **Kasela, S.***, Kals, M., Tamm, R., Lokk, K., Barragan, I., Buurman, W.A., Deelen, P., Greve, J.-W., Ivanov, M., et al. (2014). Genetic and epigenetic regulation of gene expression in fetal and adult human livers. BMC Genomics 15, 860.

Kacevska, M., Ivanov, M., Wyss, A., **Kasela, S.**, Milani, L., Rane, A., and Ingelman-Sundberg, M. (2012). DNA methylation dynamics in the hepatic CYP3A4 gene promoter. Biochimie 94, 2338–2344.

Supervised dissertations:

2016 Co-supervision of the master's thesis of Marili Palover "Associations between telomere length and DNA methylation levels" (Gene Technology)

2015 Co-supervision of the bachelor's thesis of Kaupo Koppel "PheWAS in theory and in practice, based on data from Estonian Genome Center, University of Tartu" (Mathematical Statistics)

2014 Co-supervision of the bachelor's thesis of Carmen-Kristiina Parik "Asthma associated DNA methylation changes in peripheral blood" (Gene Technology)

Awards and stipends:

2016 Kristjan Jaak Scholarship for foreign visits: poster presentation in the 28th International Biometric Conference, Victoria, Canada

2014 Kristjan Jaak Scholarship for foreign visits: platform presentation in the 23rd Annual Conference of the International Genetic Epidemiology Society, Vienna, Austria

2014 Scholarship of the Graduate School in Biomedicine and Biotechnology: platform presentation in the 23rd Annual Conference of the International Genetic Epidemiology Society, Vienna, Austria

2012 Kristjan Jaak Scholarship for part-time studies in the Universiteit Hasselt, Hasselt, Belgium

ELULOOKIRJELDUS

Nimi: Silva Kasela
Sünniaeg: 21. veebruar 1989
Aadress: Tartu Ülikooli Eesti Geenivaramu
Riia 23B, 51010, Tartu, Eesti
E-post: silva.kasela@ut.ee

Haridus:

2013–…	Doktoriõpe, molekulaar- ja rakubioloogia, loodus- ja täppisteaduste valdkond, Tartu Ülikool, Tartu, Eesti
09/2012–02/2013	Vahetustudeng, biostatistika magistriõpe, Universiteit Hasselt, Belgia
2011–2013	Magistriõpe (*cum laude*), matemaatiline statistika, matemaatika-informaatikateaduskond, Tartu Ülikool, Tartu, Eesti
2008–2011	Bakalaureuseõpe (*cum laude*), matemaatiline statistika, matemaatika-informaatikateaduskond, Tartu Ülikool, Tartu, Eesti
1996–2008	Kilingi-Nõmme Gümnaasium (kuldmedal)

Teenistuskäik:

07/2013–…	Tartu Ülikooli Eesti Geenivaramu; spetsialist
02/2013-06/2013	Matemaatilise statistika instituut, matemaatika- ja informaatikateaduskond, Tartu Ülikool; "MTMS.02.027 Tõenäosusteooria ja matemaatilise statistika" praktikumijuhendaja

Teadusorganisatsiooniline ja- administratiivne tegevus:

2014−…	Rahvusvahelise biomeetriaühingu "*International Biometric Society*" Põhja-Balti regiooni liige
2014−...	Rahvusvahelise kursuse "*Statistical Practice in Epidemiology using R*" (Tartu Ülikool) korraldaja
2014−2016	Rahvusvahelise biomeetriaühingu "*International Biometric Society*" Põhja-Balti regiooni sekretär
2014−2015	Teadlaste Öö festivali teadussaadik
2013−...	Eesti Statistikaseltsi liige

Teaduspublikatsioonid:

Org, E., Blum, Y., **Kasela, S.**, Mehrabian, M., Kuusisto, J., Kangas, A.J., Soininen, P., Wang, Z., Ala-Korpela, M., Hazen, S.L., et al. (2017). Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. Genome Biol. 18, 70.

Kasela, S., Kisand, K., Tserel, L., Kaleviste, E., Remm, A., Fischer, K., Esko, T., Westra, H.-J., Fairfax, B.P., Makino, S., et al. (2017). Pathogenic

implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. PLOS Genet. 13, e1006643.

Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.-C., Ried, J.S., Zhang, W., Yang, Y., …, **Kasela S.**, et al. (2017). Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541, 81–86.

Pervjakova, N., **Kasela, S.**, Morris, A.P., Kals, M., Metspalu, A., Lindgren, C.M., Salumets, A., and Mägi, R. (2016). Imprinted genes and imprinting control regions show predominant intermediate methylation in adult somatic tissues. Epigenomics 8, 789–799.

Milani, F., Dumas, M., Matulevičius, R., Ahmed, N., and **Kasela, S.** (2016). Criteria and Heuristics for Business Process Model Decomposition. Bus. Inf. Syst. Eng. 58, 7–17.

Võsa, U., Esko, T., **Kasela, S.**, and Annilo, T. (2015). Altered Gene Expression Associated with microRNA Binding Site Polymorphisms. PLoS One 10, e0141351.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., …, **Kasela S.**, et al. (2015). The transcriptional landscape of age in human peripheral blood. Nat. Commun. 6, 8570.

Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., Kelly, T.N., Saleheen, D., Lehne, B., Leach, I.M., …, **Kasela S.** et al. (2015). Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat. Genet. 47, 1282–1293.

Tserel, L., Kolde, R., Limbach, M., Tretyakov, K**., Kasela, S.**, Kisand, K., Saare, M., Vilo, J., Metspalu, A., Milani, L., et al. (2015). Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. Sci. Rep. 5, 13107.

Westra, H.-J., Arends, D., Esko, T., Peters, M.J., Schurmann, C., Schramm, K., Kettunen, J., Yaghootkar, H., Fairfax, B.P., Andiappan, A.K., …, **Kasela S.**, et al. (2015). Cell Specific eQTL Analysis without Sorting Cells. PLoS Genet. 11, e1005223.

Putku, M., Kals, M., Inno, R., **Kasela, S.**, Org, E., Kožich, V., Milani, L., and Laan, M. (2015). CDH13 promoter SNPs with pleiotropic effect on cardio-metabolic parameters represent methylation QTLs. Hum. Genet. 134, 291–303.

Bonder, M.J.*, **Kasela, S.***, Kals, M., Tamm, R., Lokk, K., Barragan, I., Buurman, W.A., Deelen, P., Greve, J.-W., Ivanov, M., et al. (2014). Genetic and epigenetic regulation of gene expression in fetal and adult human livers. BMC Genomics 15, 860.

Kacevska, M., Ivanov, M., Wyss, A., **Kasela, S.**, Milani, L., Rane, A., and Ingelman-Sundberg, M. (2012). DNA methylation dynamics in the hepatic CYP3A4 gene promoter. Biochimie 94, 2338–2344.

Juhendatud väitekirjad:

2016 Marili Paloveri magistritöö „Telomeeride pikkuste seosed DNA metülatsiooniga“ kaasjuhendamine (geenitehnoloogia õppekava)

2015 Kaupo Koppeli bakalaureusetöö „PheWAS ja selle praktiline läbiviimine Tartu Ülikooli Eesti Geenivaramu andmete põhjal“ kaasjuhendamine (matemaatilise statistika õppekava)

2014 Carmen-Kristiina Pariku bakalaureusetöö „Astmaga seotud DNA metülatsiooni muutused veres“ kaasjuhendamine (geenitehnoloogia õppekava)

Stipendiumid:

2016 Kristjan Jaagu välissõidu stipendium: stendiettekanne 28. Rahvusvahelise Biomeetriaühingu konverentsil (28th *International Biometric Conference*), Victoria. Kanada

2014 Kristjan Jaagu välissõidu stipendium: suuline ettekanne 23. Rahvusvahelise Geneetilise Epidemioloogia Ühingu (23rd *International Genetic Epidemiology Society*) aastakonverentsil, Viin, Austria

2014 Biomeditsiini ja Biotehnoloogia doktorikooli stipendium: suuline ettekanne Rahvusvahelise Geneetilise Epidemioloogia Ühingu (23rd *International Genetic Epidemiology Society*) aastakonverentsil, Viin, Austria

2012 Kristjan Jaagu osalise õppe stipendium: vahetussemester Hasselti Ülikoolis, Belgias

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 Uusküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.

245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.
246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreo-nemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and 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.