DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
311

KAIE LOKK

Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
311

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
311

KAIE LOKK

Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue

UNIVERSITY OF TARTU
Press

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

This dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Molecular Diagnostics on 18.11.2016 by the Council of the Institute of Molecular and Cell Biology at the University of Tartu.

Supervisors: Neeme Tõnisson, MD, PhD
Estonian Genome Center
University of Tartu
Estonia

Lili Azin Milani, PhD
Estonian Genome Center
University of Tartu
Estonia

Reviewer: Viljar Jaks, MD, PhD
Department of Cell Biology
Institute of Molecular and Cell Biology
University of Tartu
Estonia

Opponent: Prof. Stephan Beck, PhD
UCL Cancer Institute
University College London
United Kingdom

Commencement: Room No. 105, 23B Riia St., Tartu, on January 20, 2017, at 10.15 am.

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

ISSN 1024-6479
ISBN 978-9949-77-312-1 (print)
ISBN 978-9949-77-313-8 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS 7

ABBREVIATIONS 8

INTRODUCTION 9

REVIEW OF THE LITERATURE 11
1.1 DNA methylation and demethylation 11
1.2 DNMTs 13
1.3 DNA methylation in healthy tissues 14
1.3.1 DNA methylation during development 14
1.3.2 Tissue-specific DNA methylation 16
1.3.3 Epigenetic regulation of hepatic gene expression 17
1.4 DNA methylation and tumorigenesis 19
1.4.1 Driver mutations and epimutations in cancer 19
1.4.2 Cancer stem cells (CSCs) 20
1.4.3 Epigenetics of lung cancer 22
1.5 DNA hydroxymethylation 24
1.6 Current methods of epigenome analysis 25

AIMS OF THE STUDY 28

RESULTS AND DISCUSSION 29
3.1 Methylation markers of early-stage NSCLC (Ref I) 29
3.1.1 Cohort description 29
3.1.2 Description of DNA methylation in NSCLC 30
3.1.3 Correlation between DNA methylation and gene expression . 30
3.1.4 Ingenuity pathway analysis 31
3.1.5 Survival and smoking analysis 31
3.2 DNA methylome profiling of human tissues to identify global and tissue-specific methylation patterns (Ref II) 34
3.2.1 Samples analyzed 34
3.2.2 Genome-wide DNA methylation patterns 34
3.2.3 Comparison of DNA methylation in CGI and non-CGI regions 36
3.2.4 Tissue-specific differentially methylated regions 37
3.2.5 Inter-individual methylation variations 40
3.2.6 Relationship between gene expression and global DNA methylation 40
3.2.7 Gene expression and methylation in tDMRs 41
3.3 Genetic and epigenetic regulation of gene expression in fetal and adult human livers (Ref III) 42
3.3.1 Samples analyzed 42
3.3.2 Developmental regulation of hepatic gene expression 43
3.3.2.1 The epigenome of the developing human liver 43

3.3.2.2 Transcriptome of the developing liver 43
3.3.2.3 Orchestration of epigenetics and transcriptomics in regulating liver development 44
3.3.3 Genetic and epigenetic effects on inter-individual variability in gene expression 44
3.3.3.1 Correlation in DNA methylation and gene expression . 44
3.3.3.2 Regulation of gene expression by genetic polymorphisms 45
3.3.3.3 Contribution of genetic variants and DNA methylation to variation in hepatic gene expression 46
3.3.4 Tissue-specificity of eQTLs and eQTMs 47
CONCLUSIONS 48
SUMMARY IN ESTONIAN 49
REFERENCES 51
ACKNOWLEDGEMENTS 63
PUBLICATIONS 65
CURRICULUM VITAE 110
ELULOOKIRJELDUS 111

LIST OF ORIGINAL PUBLICATIONS

Ref I **Lokk K**, Vooder T, Kolde R, Välk K, Võsa U, Roosipuu R, Milani L, Fischer K, Koltsina M, Urgard E, Annilo T, Metspalu A, Tõnisson N. (2012) Methylation Markers of Early-Stage Non-Small Cell Lung Cancer. PLoS ONE 7(6): e39813. doi:10.1371/journal.pone.0039813

Ref II **Lokk K**, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R, Koltšina M, Nilsson TK, Vilo J, Salumets A, Tõnisson N. (2014). DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biology, 15(4):r54. doi: 10.1186/gb-2014-15-4-r54

Ref III 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. doi: 10.1186/1471-2164-15-860

The articles listed above have been reprinted with permission of the copyright owners.

My contributions to the above publications are as follows:

Ref I The author conceived and designed the study, performed most of the experiments, participated in the data analysis, wrote the manuscript.

Ref II The author carried out the studies, participated in the data analysis and wrote the manuscript.

Ref III The author participated in the data analysis and writing of the manuscript.

This research was supported by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) GENTRANSMED, by IUT20-60 and PUT735 from the Estonian Research Council, EU 2020 (Grant no. 692145), and the EU Social Fund support for core facilities through the Estonian Research Council IUT24-6 Estonian Centre for Genomics.

ABBREVIATIONS

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
ADME	Absorption, distribution, metabolism and excretion
ANOVA	Analysis of variance
AQP1	Aquaporin 1
CGI	CpG island
CHARM	Comprehensive high-throughput analysis for relative methylation
CSC	Cancer stem cell
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNMT	DNA methyltransferase
eCGI	Experimentally defined CpG island
eQTL	Expression quantitative trait locus
eQTM	Expression quantitative trait DNA methylation
FDR	False discovery rate
GO	Gene ontology
HCP	High CpG density promoter
HELP	HpaII tiny fragment enrichment by ligation-mediated PCR
ICP	Intermediate CpG density promoter
LCP	Low CpG density promoter
MBD-seq	Methyl-binding domain sequencing
meDIP-seq	Methylated DNA immunoprecipitation sequencing
NSCLC	Non-small cell lung cancer
PCC	Pearson correlation coefficient
PGC	Primordial germ cell
RRBS	Reduced representation bisulfite sequencing
SAT	subcutaneous adipose tissue
tDMR	Tissue-specific differentially methylated region
TET	Ten-eleven translocation
TSS	Transcription start site
VAT	Visceral adipose tissue

INTRODUCTION

The human genome contains about 29 million CpG dinucleotides where a cytosine base is followed by a guanine base. Cytosine in the CpG dinucleotide can exist in either a methylated or unmethylated state. DNA methyltransferase (DNMT) enzymes can add a methyl group to cytosine to form 5-methylcytosine (5mC), and this methylated form of cytosine is considered the "fifth" DNA base. Methylated cytosines have a crucial role in mammalian development, and the proportion of methylated CpG sites can vary greatly over a genome. Repeated sequences tend to be the most heavily methylated, while CpG-rich regulatory regions are almost exclusively unmethylated in all human tissues. As DNA methylation is vital for the normal functioning of organism, changes in the epigenome can account for individual differences in drug responses or the incidence of severe diseases, especially cancer. DNA methylation patterns have been investigated over several decades and various methods have been applied. However, the mechanistic details regarding the relationship between epigenetics and transcription remain unclear. Several studies have linked methylation events in promoter-CpG islands with gene expression regulation. Despite these correlations, tissue-specific differentially methylated regions have been increasingly found among intragenic regions, such as in gene bodies or even in intergenic regions. But the relationship between DNA methylation in non-promoter regions and gene expression remains unclear.

Cancer, especially lung cancer, is currently a major health problem worldwide. Over the past few years, several biomarkers have been discovered, both genetic and epigenetic, that may potentially help clinicians diagnose cancer in its early stage, or help provide a prognosis regarding treatment outcome or overall survival. Despite these advances, however, only a handful of these biomarkers have been successfully validated and used in the clinic.

In the first part of the current thesis, an overview of DNA methylation is provided. Specifically, the dynamics of DNA methylation patterns during mammalian development are described and the enzymes involved. Insight regarding specific methylation differences and their influence on gene expression in several adult tissues are also presented. Next, an overview of aberrant genetic and epigenetic events in various cancerous tissues is provided, particularly as these events relate to lung cancer. Then, finally, the current methods used in epigenome analyses are presented.

The experimental part of this thesis consists of three sections that address the global methylation patterns observed in healthy human somatic tissues, including fetal and adult liver tissue, and non-small cell lung cancer (NSCLC) tissue. In the first study, a set of genes that exhibited alterations in their methylation profile in NSCLC tissues were identified and correlated with transcription levels. A number of genes that exhibited a significant association between DNA methylation and survival rates were also identified. In the second study, tissue-specific differences in methylated regions in 17 human tissues

were described, and many of the methylated regions were found to contribute to tissue-specific functions. In the third study, a comprehensive genome-wide analysis was performed to investigate genomic and epigenomic variations and their associations with gene expression in fetal and adult human liver tissues. These results demonstrate that both genetic and epigenetic factors have a significant influence on transcription, and this influence affects tissue-specific gene expression as well.

REVIEW OF THE LITERATURE

1.1 DNA methylation and demethylation

One of the earliest studied epigenetic marks in eukaryotic DNA was DNA methylation (Lujambio et al., 2008). It has important roles in embryogenesis, differentiation, imprinting, X-inactivation, and genomic stability (Yanagawa et al., 2003). In mammals, methylation almost exclusively affects cytosines in CpG dinucleotides, of which approximately 70–80% are methylated throughout the genome (Busche et al., 2015). The remaining unmethylated CpG dinucleotides are mostly grouped in dense clusters called CpG islands (CGIs), and these are often found near gene promoters and other regulatory regions. Housekeeping genes have CGIs proximal to their promoters, and these CGIs are typically unmethylated even when the housekeeping gene is transcriptionally inactive (Weber et al., 2007; Yan et al.).

In general, DNA methylation often acts as a repression mark that causes gene expression inhibition, either by directly preventing the binding of transcription factors (Comb and Goodman, 1990) or by recruiting methyl-binding proteins which alter the structure of chromatin (Harikrishnan et al., 2005). DNA methylation profiles have been shown to vary with age (Madrigano et al., 2012), and in response to environmental factors (Bind et al., 2012), nutritional factors, and pathogenic factors (Leonard et al., 2012; Liu et al., 2003). Disease-specific methylation patterns have also been observed (Montano et al., 2016; Sun et al., 2016), especially for cancer (Jones and Baylin, 2007). To date, almost all types of cancer have exhibited both aberrant hypermethylation of tumor suppressor genes and global hypomethylation of their genomic DNA (Esteller, 2007).

Based on the CpG content and length of CGI-associated promoters, they have been classified into three categories. Promoters with a high density of CpG sites, referred to as HCPs, are rarely methylated, even when the genes are not transcriptionally active (Weber et al., 2007; Ziller et al., 2013) In the cases where HCPs are methylated, the genes are often efficiently silenced (Meissner et al., 2008; Weber et al., 2007). Promoters with an intermediate density of CpG sites, referred to as ICPs, also repress gene expression in response to methylation (Weber et al., 2007). ICPs are common in differentiation-dependent hypermethylation events where DNA methylation is needed to ensure stable silencing during differentiation (Borgel et al., 2010; Weber et al., 2007). In contrast, low CpG density promoters (LCPs) undergo varying levels of methylation, and in general, they remain hypermethylated (Ziller et al., 2013). However, the transcriptional status of LCP genes lack a clear correlation with their methylation state, thereby suggesting that a low concentration of methylated CpG sites in a promoter region may not be sufficient to influence gene activity (Weber et al., 2007).

DNA methylation has traditionally been considered a relatively stable epigenetic mark. However, numerous studies have demonstrated that DNA methylation exhibits dynamic characteristics. For example, DNA can be *de novo* methylated

as described in Chapter 1.2 (Figure 1), yet DNA demethylation can also occur via passive and active mechanisms (Figure 1). Passive DNA demethylation involves the loss of a methyl group over rounds of DNA replication. It is hypothesized that this mechanism occurs during early embryonic development in mammals, and it involves the gradual demethylation of maternal DNA in a pre-implanted blastocyst over several cell divisions in a replication-dependent manner (Howlett and Reik, 1991). In contrast, active DNA demethylation involves an enzymatic process by which the methyl group is removed from 5mC via

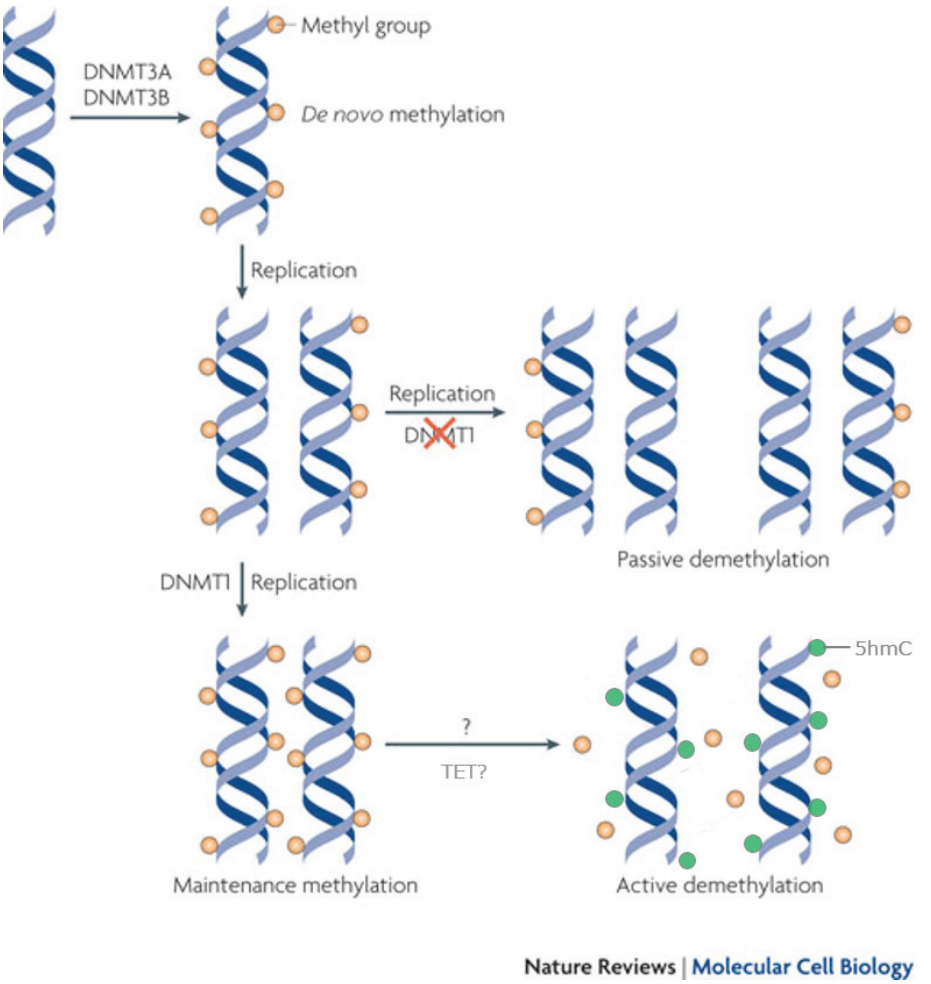

Figure 1. Passive and active DNA demethylation. During development, DNA methylation patterns are established by DNMTs. Passive demethylation occurs during rounds of replication when DNMT1 is not present. In contrast, active DNA demethylation is an enzymatic process and it is hypothesized that members of the TET family mediate the oxidation of 5mC to 5hmC. Figure adapted from Wu and Zhang (2010) and reprinted with permission of the Nature publishing group.

breakage of the carbon-carbon bond. Several mechanisms are proposed to explain active DNA demethylation, and accumulating evidence indicates that oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) is mediated by members of the Ten-eleven translocation (TET) family (Tahiliani et al., 2009).

1.2 DNMTs

DNMTs are responsible for catalyzing cytosine methylation. The mammalian family of DNMTs includes three members which have active catalytic domains – DNMT1, DNMT3A, DNMT3B and in addition DNMTL, which can interact with DNMT3A/DNMT3B (Figure 2).

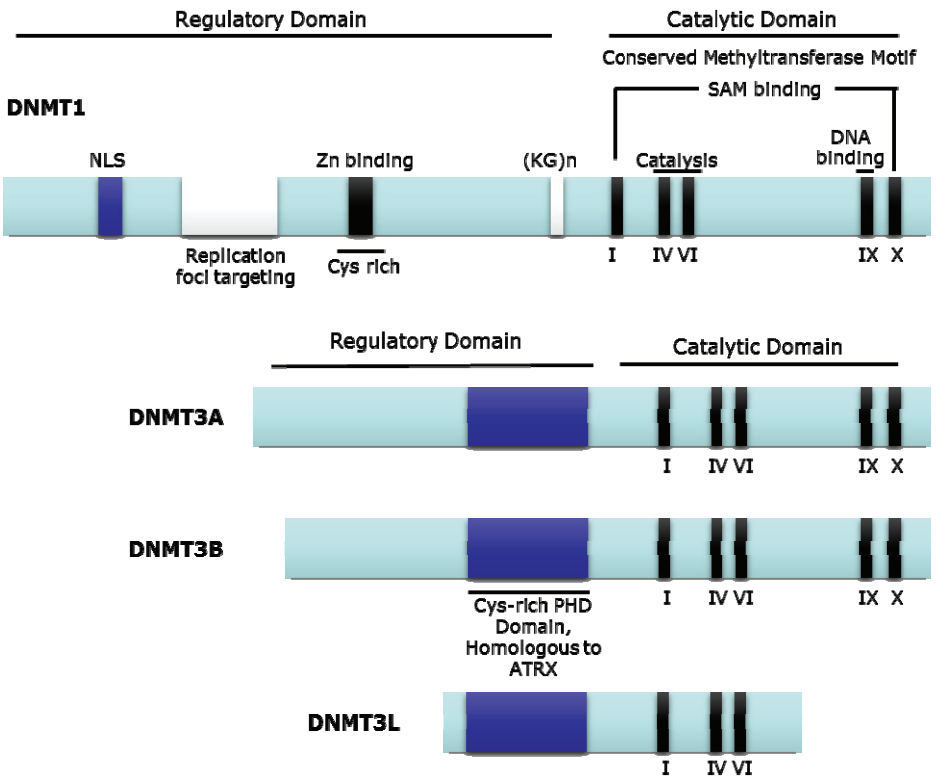

Figure 2. Schematic structure of the DNMTs, DNMT1, DNMT3A, DNMT3B, and DNMT3L. The regulatory domains contain motifs for mediating interactions with proteins or DNA. The catalytic domains contain conserved methyltransferase motifs. SAM: S-adenosyl methionine; NLS: nuclear localization signal; (KG)n: lysine-glycine repeats; ATRX: chromatin remodeler gene; PHD: plant homology domain.

It has been observed most often that DNMT1 is responsible for maintaining DNA methylation patterns. It has exhibited up to a 40-fold preference for hemimethylated DNA over unmethylated DNA and it localizes to replication forks during DNA replication where it methylates newly biosynthesized DNA (Fatemi et al., 2001; Goyal et al., 2006). Furthermore, mouse models with reduced levels of DNMT1 at an embryonic age have been shown to be nonviable (Li et al., 1992).

In contrast, DNMT3a and DNMT3b are not able to differentiate between hemimethylated and unmethylated DNA and are considered to be responsible for establishing *de novo* methylation patterns during early embryonic development and in germ cells (Gowher and Jeltsch, 2001; Okano et al., 1999; Okano et al., 1998). DNMT3a and DNMT3b are both highly expressed in embryonic tissues. DNMT3a is primarily responsible for establishing imprinting patterns in maturing gametes, and thus, is expressed in the later embryonic stages (Kaneda et al., 2004; Watanabe et al., 2002). DNMT3a is also involved in methylating postnatal stem cells (Challen et al., 2012; Wu et al., 2010). Conversely, DNMT3b is more predominant in the early embryonic stages, and has been being detected in preimplantation embryos and epiblasts (Borgel et al., 2010; Watanabe et al., 2002). DNMT3b is also essential for maintaining methylation events in pericentromeric repeats which are necessary for proper cell division (Gopalakrishnan et al., 2009). While DNMT3L has no detectable methyltransferase activity, it has been found to interact with DNMT3A/DNMT3B (Suetake et al., 2004). DNMT3L is also essential for establishing maternal genomic imprinting in oocytes and for spermatogenesis (Hata et al., 2002).

1.3 DNA methylation in healthy tissues

1.3.1 DNA methylation during development

DNA methylation undergoes major reprogramming during embryonic development in mammals. First, methylation levels globally decrease, then new methylation patterns are established upon implantation. The latter step is important for establishing pluripotency.

Demethylation occurs asymmetrically in paternal and maternal pronuclei in formed zygotes (Guo et al., 2014b). Paternal DNA rapidly loses almost all of its methylation primarily via the oxidation of 5mC to 5hmC by TET proteins (Guo et al., 2014a; Mayer et al., 2000). At the same time, it is hypothesized that maternal DNA loses its methylation over several replication events during cell divisions (Mayer et al., 2000). The only exceptions are imprinted loci, certain transposable elements, and centromeric heterochromatin which maintain their methylation patterns during this demethylation wave (Guo et al., 2014b; Lane et al., 2003; Nakamura et al., 2007).

After a global erasing of DNA methylation has occurred, new patterns are established at the time of embryo implantation (Figure 3). DNMT3A and

DNMT3B, together with DNMT1 are responsible for *de novo* methylation (Okano et al., 1999). During this process, some CGIs remain unmethylated while other CGIs are newly methylated (Brandeis et al., 1994). Sequence-specific methylation events take place after implantation, genes responsible for pluripotency (e.g., *Oct3/4*, *Nanog*) undergo methylation and are silenced (Feldman et al., 2006). *De novo* methylation also occurs on the X-chromosome, where promoters containing CGIs become methylated as a last event (Gendrel et al., 2012). It has been demonstrated that DNA methylation itself is not necessary for initiating silencing, although it is important for maintaining gene repression over several cell cycles throughout an organism's lifespan (Epsztejn-Litman et al., 2008). In addition, it has been observed that many tissue-specific genes become methylated during development, and then are demethylated in the later stages of development when their expression becomes tissue-specific (Song et al., 2009).

Another major DNA methylation reprogramming takes place in primordial germ cells. Demethylation is carried out almost entirely genome-wide and it is completed by approximately E12.5 (Figure 3). The last demethylation events involve parent-of-origin dependent imprinting loci, CGIs of inactive X-chromosome genes, and promoters of germ cell-specific genes (Seisenberger et al., 2012). Only a subset of repetitive sequences are protected from demethylation, and it is hypothesized that these sequences are essential for overall chromosome stability (Hajkova et al., 2002). In PGCs, the loss of 5mC is likely to involve both active and passive mechanisms (Hackett et al., 2013). Global demethylation is followed by 5mC remethylation to establish a unique epigenome in PGCs to form mature gametes (Davis et al., 2000). Sperm- and oocyte-specific patterns are then created at different times and in different cellular environments (Sasaki and Matsui, 2008), with male germ cells progressively establishing imprints up to the newborn stage (Kato et al., 2007) and female PGCs undergoing remethylation after birth during oocyte growth (Hiura et al., 2006).

DNA demethylation occurs during embryonic development, but it has also been observed in hematopoietic cells. In the latter DNA demethylation has a role in regulating the activation of hematopoietic-specific genes (Calvanese et al., 2012). Similarly, DNA demethylation events have been associated with myogenic genes during muscle differentiation (Tsumagari et al., 2013). Studies are also ongoing of DNA demethylation events that have been detected in neurons (Guo et al., 2011) and in liver (Waterland et al., 2009).

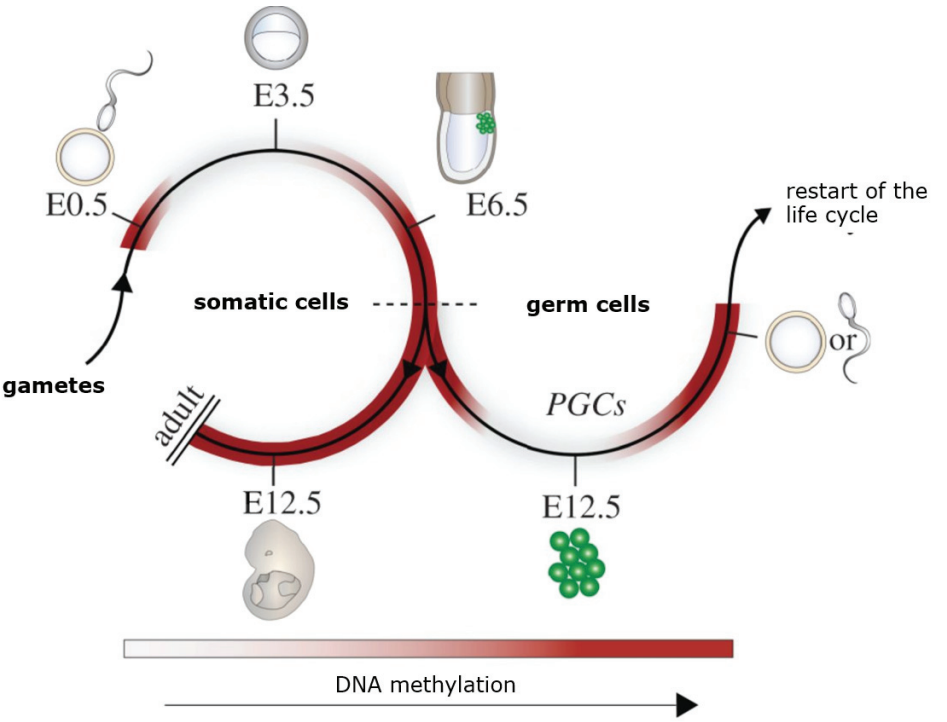

Figure 3. Global DNA methylation dynamics during mammalian embryonic development. Demethylation occurs in the formed zygote differentially in paternal and maternal pronuclei. At approximately embryonic day 3.5, remethylation begins at the time of implantation. Both DNMT3A and DNMT3B, in combination with DNMT1, are responsible for *de novo* methylation events. At E6.5, cells commit to a development program to become somatic cells or primordial germ cells (PGCs). PGCs undergo another major DNA methylation reprogramming event that is completed by approximately E12.5. Global demethylation is then followed by 5mC remethylation to establish a unique epigenome in PGCs that lead to the formation of mature gametes. Figure adapted from Hackett and Surani (2013).

1.3.2 Tissue-specific DNA methylation

To obtain more detailed information regarding the spatial distribution of DNA methylation and its functional effects, numerous studies have investigated the methylation profiles of several diverse tissue types and have confirmed that tissue-specific methylation patterns exist (Lowe et al., 2015; Varley et al., 2013; Yang et al., 2015). While tissue-specific differentially methylated regions have been well characterized in promoter regions of the genome (Nagae et al., 2011), recent studies have elucidated that these methylated regions are more often represented in gene bodies (Deaton et al., 2011), outside of CGIs, and in non-coding regions (Yang et al., 2015). Moreover, in contrast with tissue-specific hypermethylation, genes that exhibit tissue-specific hypomethylation have a

closer relation to tissue functions according to gene ontology analyses (Nagae et al., 2011; Yang et al., 2015; Zilbauer et al., 2013).

The classic model of the relationship between DNA methylation and gene expression has included methylation of gene promoters as a mechanism by which transcription is suppressed, while the methylation of gene bodies mediates active expression (Rakyan et al., 2008; Song et al., 2005). However, recent studies have revealed that the relationship between DNA methylation and gene expression is far more complicated than previously thought. For example, it has been shown that CGIs in gene bodies can have a bimodal distribution of correlation between methylation and gene expression (Varley et al., 2013), thereby suggesting that alternative promoters may be methylated in a promoter-like manner (Deaton et al., 2011; Varley et al., 2013). Conversely, Varley *et al*. demonstrated that binding sites for the transcription coactivator, EP300, in gene body CGIs may account for the inverse correlation between DNA methylation and gene expression that has been observed in certain cases (Varley et al., 2013).

Recently, whole-genome bisulfite sequencing was used to establish a novel and experimentally defined CGI (eCGI) catalog (Mendizabal and Yi, 2016). The authors claim that this approach enabled the identification of many CGIs that were otherwise undetectable by traditional computational methods, such as CGIs in repetitive elements (e.g., SINE, LINE, STR). Many of the novel eCGIs were predicted to be tissue-specific, and almost all of the identified eCGIs (98.1%) had at least one transcription start site (TSS), while many (81%) overlapped with transcription factor binding sites. A number of eCGIs were also located in gene body regions.

1.3.3 Epigenetic regulation of hepatic gene expression

The liver metabolizes drugs, chemicals, and other xenobiotics, while also regulating the balance of hormones and vitamins in the body. Genes that are engaged in the absorption, distribution, metabolism, and excretion (ADME) of drugs display very high interindividual heterogeneity in their gene expression profiles, which is responsible for the differences in drug response and adverse drug reactions that are observed in patients (Pirmohamed, 2014). These differences in gene expression can be attributed, in part, to variations in genetic factors so that a significant amount may be due to epigenetic factors (Ivanov et al., 2016; Sim et al., 2013). Several studies have investigated the possible role of epigenetic regulation in the transcription for ADME genes (Kacevska et al., 2012a).

Much of the available data regarding the relationship between DNA methylation and gene expression regulation have derived from cancer tissues (Muggerud et al., 2010; Silviera et al., 2012) or other medical conditions (Anttila et al., 2003; Kaut et al., 2012; Tobi et al., 2009). However, there are a few studies which have investigated methylation levels of genes in liver or

various cell lines. In liver tissue, a significant negative correlation was observed between DNA methylation in the CGI located in the promoter region of *CYP1A2* and total *CYP1A2* mRNA expression (Ghotbi et al., 2009). Similarly, *CYP3A4* displays highly variable CpG methylation in the proximal promoter, corresponding to the binding sites for several transcription factors (Kacevska et al., 2012b). Some single CpGs in the region were found inversely correlate with gene expression, thereby suggesting that they may have a role in transcription regulation. It has also been demonstrated that the liver has abundant levels of 5hmC, and bisulfite treatment is unable to distinguish between 5hmC and 5mC. Therefore, the authors clearly demonstrated the importance of studying 5hmC and 5mC separately in liver tissue.

When Dannenberg treated the HepG2 cell line with DNMT inhibitor, 5-aza-2'deoxycytidine, the expression of several liver-specific genes involved in xenobiotic metabolism, steroid biosynthesis, and CCAAT element binding were restored suggesting that the expression of these genes might be regulated by DNA methylation (Dannenberg and Edenberg, 2006). Namely, the cytochrome P450 family of proteins, including the *CYP3A* gene involved in xenobiotic metabolism, and *CYP19A1* and *CYP17A1* with roles in steroid biosynthesis, are possibly regulated by DNA methylation.

In a recent study, differences in DNA methylation and gene expression between human fetal and adult liver tissues were investigated (Huse et al., 2015). Most of the genes that were found to be differentially methylated were previously shown to have roles in the metabolism of drugs and xenobiotics. In addition, the genes that were overexpressed in the adult tissues exhibited an inverse correlation with the DNA methylation events detected, especially those located in the CpGs at the TSS and in the first exon. These changes potentially indicate that over the course of the differentiation of fetal tissue into adult tissue, DNA methylation plays a major role in transcription induction. However, a high proportion of the DNA methylation changes showed no significant change in the respective gene expression patterns, thereby suggesting that the relationship between DNA methylation and gene expression remains unpredictable.

When separate analyses of 5mC and 5hmC in relation to ADME gene expression in liver tissue were conducted, 5hmC was found to be more abundant and it exhibited greater variability than 5mC (Ivanov et al., 2016). In contrast, data from a conventional bisulfite-based methylation analysis did not correlate with the gene expression data. Therefore, based on the observation that a negative correlation was associated with 5mC and a positive correlation was associated with 5hmC, the importance of analyzing these DNA modifications separately in 5hmC-rich tissues was demonstrated.

1.4 DNA methylation and tumorigenesis

1.4.1 Driver mutations and epimutations in cancer

Most cancers arise sporadically, with many alleles only weakly contributing to cancer risk. In familial types of cancer, alleles can have a larger effect on tumorigenesis, whereas monogenic cancers have highly pathogenic risk alleles. There are many genes that contain mutations which contribute to hereditary cancer syndromes, including *MLH1* and *MSH2* for Lynch syndrome (Liu et al., 1996) and *BRCA1* (Miki et al., 1994) and *BRCA2* (Wooster et al., 1995) for hereditary breast and ovarian cancer syndrome. Cancer genomes are characterized by very high mutation rates and genomic instability, and recent sequencing studies have shown that these genomes can contain thousands to tens of thousands of somatic base substitutions (Nik-Zainal et al., 2016; Pleasance et al., 2010). However, only a small subset of these mutations represent "driver mutations" which are considered to be mutations that initiate and maintain cell malignancy by providing a proliferation advantage and by being positively selected for in the cancer evolution (Maley et al., 2004; Stratton et al., 2009). Mathematical modeling has estimated that of all the mutations found in cancers, the average number of driver mutations is less than ten per genome, while the others are passenger mutations (Kandoth et al., 2013).

Cancer genomes also show changes in their epigenetic patterns compared to normal cells, with both hypermethylated CGIs and global hypomethylation promoting chromosomal instability (Jones and Baylin, 2002) (Figure 4). In the context of epigenetics, "driver methylation" events represent a promoting factor of tumorigenesis (Carter et al., 2009). This includes events where hypermethylation inactivates tumor suppressor genes, as well as events involving the activation of oncogenes by global hypomethylation. For hereditary and sporadic forms of cancer, Knudson's "two hit" theory has postulated that both alleles of a gene need to be affected to promote tumorigenesis (Knudson, 1971). In the case of hereditary cancer, the first hit can be an inherited mutation in a tumor suppressor gene, and the second mutation can occur in a somatic cancer progenitor cell in a target tissue. In sporadic cancers, both hits take place in a target tissue prior to tumor progression, and it has been suggested that one hit or both can either be genetic or epigenetic (Peltomäki, 2012).

Epimutations are described as changes that do not affect DNA sequences, yet they affect gene expression and can be heritable to daughter cells, sometimes even in a transgenerational manner (Gazzoli et al., 2002; Holliday, 1987; Peltomäki, 2012). Epimutations can be primary or secondary, and several have been described to date: *MLH1*, *MSH2*, *CDH1*, *PTPRJ*, and *KLLN*, explaining a different proportion of "missing heritability" (Peltomäki, 2012). Heritability differs between primary and secondary epimutations. For example, a primary epimutation of *MLH1* exhibits non-Mendelian heritability whereby methylation can be erased during epigenomic reprogramming in embryogenesis (Hitchins et al., 2007). In contrast, a secondary epimutation of *MSH2*, which follows the

primary genetic mutation, exhibits a classical Mendelian inheritance (Ligtenberg et al., 2009). Another theory suggests that secondary epimutations are not inherited, but rather they are reinstated in progeny as a result of *cis*- or *trans*-acting genetic factors that increase susceptibility to epimutations (Hitchins et al., 2011).

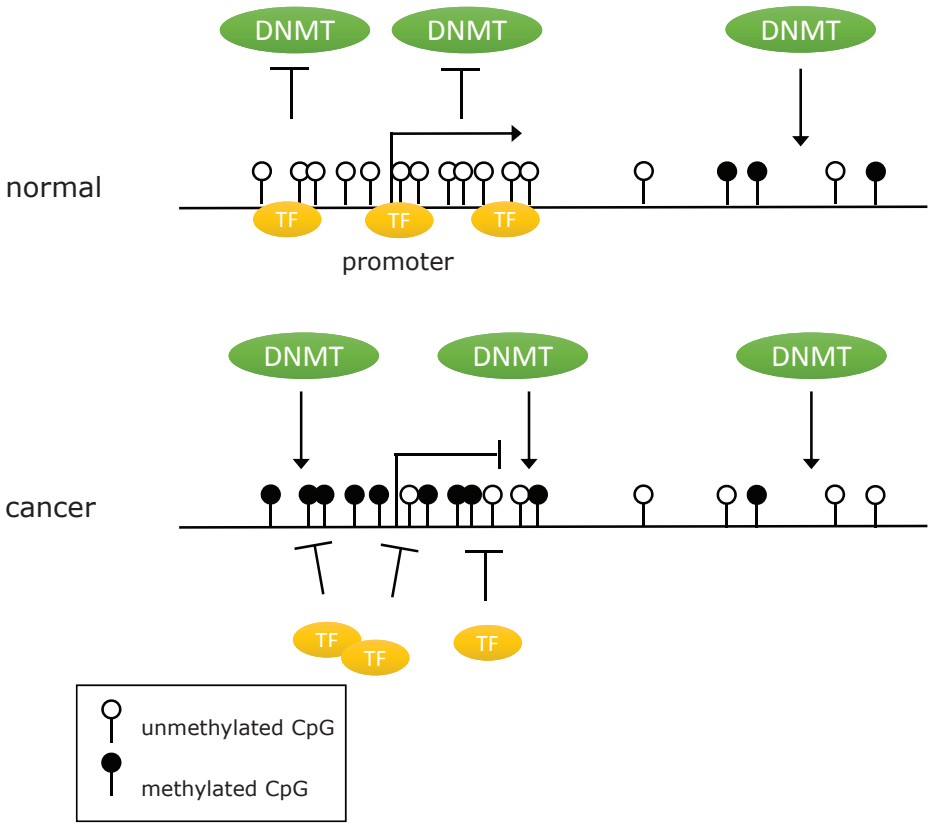

Figure 4. DNA methylation of a typical tumor suppressor gene in normal and cancer cells. In a transcriptionally active normal cell, CGIs in the promoter regions are kept unmethylated and accessible to transcription factors, while CpG sites in intragenic regions undergo methylation by DNMTs. The unmethylated state of the promoter enables active gene expression. In contrast, CGIs in the promoter regions in cancer cells are hypermethylated and this represses gene transcription. TF: transcription factor.

1.4.2 Cancer stem cells (CSCs)

Historically, cancer heterogeneity has been explained by a clonal evolution model. This model postulates that cancer cells arise due to an accumulation of genetic and epigenetic mutations that provide a growth advantage at the cellular level and that result in the generation of new tumors (Nowell, 1976). In the late

20th century, the concept of CSCs emerged and this model is thought to exist together with the clonal model (Iliopoulos et al., 2011; Notta et al., 2011). The CSC model proposes that a subset of cancer cells have stem cell-like properties, and these cells are able to renew and differentiate as part of the initiation, progression, and metastasis of cancer (Bonnet and Dick, 1997; Hamburger and Salmon, 1977). In addition to the ability of CSCs to initiate cancer, they also play a major role in resistance to chemo- and radiotherapy. Moreover, the elimination of rapidly proliferating cancer cells by chemotherapy provides a greater volume for the expansion of CSCs, and this may cause a cancer to become more malignant as its self-renewal capacity increases (Hermann et al., 2007; Li et al., 2008).

There are several factors that have been proposed to mediate treatment resistance. These include increased expression of drug transporters, intracellular detoxification enzymes, up-regulation of anti-apoptotic proteins, and tumor microenvironmental factors (Alison et al., 2012). Activation of various signaling pathways (e.g., Notch, Hedgehog, Jak/Stat) has also been associated with the resistance of CSCs to therapy (Marotta et al., 2011; McAuliffe et al., 2012; Zhao et al., 2009). Correspondingly, when chemotherapy has been administered in combination with drugs that inactivate different components of these signaling pathways, the resistance of CSCs to conventional chemotherapy has been reduced (Liu et al., 2013d; Ulasov et al., 2011). However, these targeted signaling pathways are also vital for the homeostasis and survival of normal stem cells, and thus, it is necessary to compromise these targets only in CSCs.

The precise origin of CSCs is a topic of ongoing research. CSCs were first discovered in leukemia (Bonnet and Dick, 1997), and have subsequently been found in breast, brain, lung, colon, melanoma, prostate, and pancreatic cancers (Al-Hajj et al., 2003; Collins et al., 2005; Eramo et al., 2008; Fang et al., 2005; Li et al., 2007; O'Brien et al., 2007; Singh et al., 2003). There is some evidence to indicate that CSCs may originate from normal stem cells via an accumulation of malignant genetic and epigenetic mutations (Smalley and Ashworth, 2003). Another theory suggests that CSCs are mature cancer cells that were dedifferentiated via the epithelial-mesenchymal transition (Mani et al., 2008; Shuang et al., 2014).

The actual frequency of CSCs has also been found to vary, with the incidence of CSCs reported to range from 0.1–3% (Collins et al., 2005; Singh et al., 2003; Todaro et al., 2007). However, some groups have shown that CSCs can be found in fairly large numbers (Boiko et al., 2010; Kelly et al., 2007). Their hypothetical range is vast and debatable, probably varying according to the type and stage of cancer. Also, different experimental conditions may cause biased CSCs estimates (Islam et al., 2015b).

CSCs express a variety of cell surface markers, and this enables CSCs to be isolated from tumor tissues (Ginestier et al., 2007). Some of the more commonly used markers include ALDH, CD24, CD44, CD29, CD133, and EpCAM (Islam et al., 2015a). These markers exhibit tissue-specific expression profiles, and in some cases, are cancer subtype- or cancer stage-specific. However, the

development of CSC-targeted therapies based on cell surface markers is potentially more complicated if cell surface markers for every tumor type need to be individually specified (Visvader and Lindeman, 2012).

The currently accepted "gold standard" for studying CSCs involves the inoculation of cancer cells into immunodeficient mice to establish a xenotransplantation model. It continues to be discussed whether the ability of human cells to grow in a mouse model reflects the ability of CSCs to proliferate in a tumor – mainly because the cells often undergo various *in vitro* manipulations prior to transplantation. In addition, the microenvironment is not the same as in the original location (De Palma et al., 2005; Magee et al., 2012; Puglisi et al., 2013).

From a clinical point of view, the discovery of CSCs offers new possibilities for improving cancer patient prognosis either by prolonging the time to progression or improving overall survival. The combination of existing chemotherapy drugs with drugs targeting CSCs may be the best treatment approach. However, due to the similarities between CSCs and normal stem cells, this approach affects both populations and would be improved if the therapeutic specificity would preferentially target CSCs (O'Connor et al., 2014). Alternatively, nucleic acid aptamers and chemical antibodies have exhibited high affinity and specificity for CSCs and they represent promising agents for the selective delivery of drugs to CSCs (Shigdar et al., 2013). Already, synthesized RNA aptamers have been shown to target the CSC markers, EpCAM and CD133 (Shigdar et al., 2011; Shigdar et al., 2013).

1.4.3 Epigenetics of lung cancer

Currently, lung cancer is the most common cancer worldwide. In 2012, it represented 13% of the total number of cancer cases diagnosed (Torre et al., 2015). In Estonia, lung cancer is the second most common carcinoma among men (14.8% of total) and the fourth most common among women (5.8% of total) (Estonian National Institute for Health Development, *http://www.tai.ee/*). Similarly, lung cancer is the leading cause of death among males worldwide (Torre et al., 2015). Lung cancer recurrence is also a major problem despite surgical resection being performed for early-stage patients. For example, approximately 30–55% of patients experience a relapse that is fatal (Goldstraw et al., 2007).

Lung cancer can be classified as small cell lung cancer or NSCLC according to histology and clinical course. NSCLC comprises approximately 85% of all lung cancer cases and can be divided into major subgroups that include: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Balgkouranidou et al., 2013). Genomic aberrations in adenocarcinomas have been well characterized and include somatic copy number aberrations, indels, and point mutations. In addition, common driver mutations in *EGFR*, *KRAS* and *ALK* genes are well characterized for adenocarcinomas (Devarakonda et al., 2015). In contrast, squamous cell carcinoma of the lung is more related to smoking habits, and its

genomic landscape is much more variable (Khuder and Mutgi, 2001; Pao and Girard, 2011).

In addition to global genomic rearrangements and mutations, cancer genomes undergo dramatic changes in their DNA methylation patterns. These changes can subsequently lead to changes in gene expression and compromised genome stability. Namely, repetitive sequences and oncogenes are commonly hypomethylated, while CGIs of tumor suppressor genes are highly methylated (Rauch et al., 2008; Rodriguez et al., 2006).

DNA methylation has been described as an early event in cancer progression, thereby making it an attractive biomarker target (Belinsky et al., 1998; Diaz-Lagares et al., 2016). Moreover, since epigenetic modifications are relatively stable and can be also detected in different biological fluids (e.g., sputum, bronchoalveolar lavage, saliva, and blood), they also hold great promise for application in non-invasive cancer detection. For example, methylation changes in *p16* and *MGMT* genes were found to be detectable in sputum up to three years before cancer was diagnosed (Palmisano et al., 2000). Recently, a diagnostic signature of DNA methylation biomarkers was presented and independently validated (Diaz-Lagares et al., 2016). In particular, four genes were selected as an early lung cancer methylation signature (*CDO1*, *BCAT1*, *TRIM58*, *ZNF177*) due to their high specificity and sensitivity, even in respiratory samples such as bronchial aspirates, bronchoalveolar lavages, and sputum. Another group reported a panel of genes (*CDO1*, *HOXA9*, *TAC1*) that were found to be highly specific and sensitive in NSCLC testing (Wrangle et al., 2014). These advances that facilitate the early detection of lung cancer have the potential to increase patient survival.

In addition to early cancer detection, aberrant DNA methylation can also be used to obtain a prognosis regarding cancer outcome or to support conventional therapy methods. In early-stage NSCLC patients, aberrant methylation of the promoter regions of four genes, *p16*, *CDH13*, *APC*, and *RASSF1A*, was found to be associated with tumor recurrence (Brock et al., 2008). The authors hypothesized that the recurrence was due to microscopically undetectable metastases (Brock et al., 2008). More recently, a panel of five genes (*HIST1H4F*, *PCDHGB6*, *NPBWR1*, *ALX1*, and *HOXA9*) were found to be significantly associated with the recurrence of stage I NSCLC. Furthermore, hypermethylation of two or more of these genes has been correlated with a shorter relapse-free survival period (Sandoval et al., 2013).

Compared to genetic mutations, epigenetic changes are reversible, thereby making them a promising target for cancer treatment. Several studies have evaluated the capacity for epigenetic drugs to reverse aberrant methylation patterns in lung cancer patients, particularly the DNMTs inhibitors, 5-aza-2′-deoxycytidine (decitabine) and 5-azacytidine (azacitidine) (Chu et al., 2013; Wrangle et al., 2013). While these agents were shown to be effective in demethylating cancer-specific genes in a few lung cancer patients for a limited time (Juergens et al., 2011), they also reduced genome-wide methylation and caused many toxic side effects, or did not exhibit antitumor activity at all (Nervi

et al., 2015). Many ongoing clinical trials involve the application of hypomethylating agents and histone deacetylase inhibitors for the treatment of lung cancer (Ansari et al., 2016), yet no remarkable improvements have been observed.

Most lung cancer patients have a history of smoking, although not all smokers develop tumors. Thus, smoking is considered one of the main reasons for the high prevalence of lung cancer. DNA methylation of certain genes (*p16*, *APC, MGMT*) has been found to be associated with smoking status and longer exposures to cigarette smoke (Liu et al., 2006; Toyooka et al., 2003). The carcinogens in tobacco smoke have also been found to increase the expression of DNMT1 and to correlate with poor prognosis (Lin et al., 2010).

1.5 DNA hydroxymethylation

Oxidation of 5mC to form 5hmC occurs in several tissues and cell types. The highest levels of 5hmC have been detected in neuronal cells (Kriaucionis and Heintz, 2009; Li and Liu, 2011), embryonic stem cells (Ficz et al., 2011; Ito et al., 2010), and liver (Ivanov et al., 2013). In contrast, levels of 5hmC are remarkably reduced in cancer cells (Li and Liu, 2011).

It has been hypothesized that the conversion of 5mC to 5hmC represents the first step in a DNA demethylation pathway. In mammals, enzymes of the TET family of proteins (TET1, TET2, TET3) oxidize 5mC to 5hmC (Wu and Zhang, 2011), and then mediate the conversion of 5hmC further to 5-formylcytosine (5fC) and then to 5-carboxylcytosine (5caC) (Ito et al., 2011). It has been proposed that 5caC is subsequently recognized and excised by a thymine-DNA glycosylase-initiated base excision repair mechanism (He et al., 2011). A passive demethylation pathway also exists, and in this pathway, DNMT1 does not recognize 5hmC, so hydroxymethylated cytosine cannot be copied during replication and this results in loss of methylation at the site (Valinluck and Sowers, 2007).

To date, the functional role of 5hmC remains unclear. It has been shown that fertilization of an oocyte by a sperm leads to a global reduction in DNA methylation in the paternal pronucleus and 5mC is oxidized to form 5hmC (Guo et al., 2014a). At the same time, the maternal genome retains its methylation status and is protected from Tet-catalyzed oxidation (Iqbal et al., 2011; Nakamura et al., 2012).

It has been observed that 5hmC is most abundant in brain tissue, thereby suggesting that it is necessary for brain development and function. In addition to being an intermediate in the process of demethylation, 5hmC is also stably maintained during neurodevelopment and throughout the life span of the brain (Chen et al., 2014). Further studies have found that 5hmC is highly associated with euchromatin and histone codes, H3K4me2, H3K4me1, and H3K27ac, while it also co-localizes with *Pol*II during neuronal differentiation and is enriched in actively transcribed genes (Chen et al., 2014; Stroud et al., 2011).

Human malignant tumors have remarkably lower 5hmC levels than corresponding normal tissues, which suggests that cell proliferation in cancers may lead to loss of 5hmC (Jin et al., 2011). Several mechanisms have been proposed to explain the altered 5hmC levels that are observed in malignancies. In particular, mutations in *TET2*, *IDH1*, and *IDH2* have been found to play important roles (Ko et al., 2010; Kroeze et al., 2014). However, other studies did not find that mutations in these genes led to reduced levels of 5hmC (Jin et al., 2011; Lian et al., 2012). Furthermore, in some solid cancers (e.g., various brain tumors, hepatocellular carcinoma, breast cancer), 5hmC levels were found to correlate with the stages of disease and patient survival (Kraus et al., 2012; Liu et al., 2013a; Tsai et al., 2015).

1.6 Current methods of epigenome analysis

Genome-wide DNA methylation analysis methods that are commonly used can be divided into three main categories – chemical bisulfite modification, affinity-based enrichment, and differential enzymatic digestion (Table 1). These approaches can also be combined with sequencing or microarray methods to cover a greater number of CpG sites.

Bisulfite treatment converts unmethylated cytosines into thymines, while methylated cytosines are resistant. In combination with direct sequencing (Frommer et al., 1992), bisulfite treatment can be a sensitive method for quantitating methylated cytosines across an entire genome (Busche et al., 2015; Mendizabal and Yi, 2016). Despite the uniqueness of this method, however, there are some drawbacks. First, bisulfite conversion is not capable of distinguishing between 5mC and 5hmC, or between unmethylated cytosine and 5caC (He et al., 2011; Huang et al., 2010). Second, bisulfite conversion reduces a genome's complexity mostly to three nucleotides, and this makes the alignment of DNA sequences quite challenging. Third, bisulfite treatment leads to the fragmentation of DNA into relatively short pieces, and this can lead to bias in further methylation analyses (Kurdyukov and Bullock, 2016; Sun et al., 2015). Fourth, genome-wide bisulfite sequencing is quite costly. An alternative approach can be RRBS, which analyses only the part of the epigenome that is enriched in CGIs, promoters, and genic regions (Meissner et al., 2005; Wang et al., 2012). Equally important are assays that focus on specific regions of the genome, such as the Illumina Infinium assays. In these assays, thousands of CpGs are analyzed, thereby providing a cost-effective high-throughput technology (Cicek et al., 2013; Kozlenkov et al., 2014; Moran et al., 2016).

Table 1. Comparison of current methods of epigenome analysis

Feature	Bisulfite-based methods	Affinity-based methods	Enzymatic digestion-based methods
Description	Uses bisulfite treatment, where 5mC→ C and C→T	Antibodies or proteins are used to enrich for methylated DNA	DNA is digested with methylation sensitive restriction enzymes
Resolution	Single CpG up to whole genome	No information on single CpGs	Limited to available restriction enzyme sites
Advantages	High resolution, cost-effective in combination with microarrays	Distinguishes 5hmC, high genome coverage	High throughput, high accuracy, cost-effective
Disadvantages	Cannot distinguish 5hmC, expensive for whole genome sequencing	Not quantitative, CpG-dense regions are not covered in the same way	Cannot distinguish 5hmC, low resolution, limited to restriction sites
Examples	WGSGS RRBS Infinium MethyLight	MeDIP-seq MBD-seq MIRA	CHARM HELP RLGS

WGSGS: whole genome shotgun bisulfite sequencing; RRBS: reduced representation bisulfite sequencing; MeDIP-seq: methylated DNA immunoprecipitation sequencing; MBD-seq: methyl-binding domain sequencing; MIRA: methylated CpG island recovery assay; CHARM: comprehensive high-throughput arrays for relative methylation; HELP: HpaII tiny fragment enrichment by ligation-mediated PCR; RLGS: restriction landmark genomic scanning.

Affinity-based methods use antibodies or proteins to enrich for methylated DNA fragments, and then these specific regions are further characterized with microarray or direct sequencing methods. The data obtained are then mapped back to the genome. Genome-wide methylated DNA immunoprecipitation (meDIP-seq) can distinguish between 5mC and 5hmC, and this distinction is becoming more important in epigenetic studies (Ivanov et al., 2016; Mellén et al., 2012). Another enrichment-based method is MBD-seq, and this method utilizes the methyl binding proteins, MBD2 and MBD3L1, to capture methylated DNA (Rauch and Pfeifer, 2010; Serre et al., 2010). Affinity-based approaches survey CpG-dense regions in different ways: meDIP-seq preferentially enriches areas with lower CpG densities and MBD-seq enriches regions with medium and high CpG densities. For both methods, DNA methylation is not presented in a quantitative manner nor is single CpG resolution achieved (Harris et al., 2010).

Methods that employ enzymatic digestion steps are limited to regions of the genome that have matching restriction sites, and this compromises the resolution of the data obtained. However, there are several approaches that do use enzymatic digestion and are able to cover the entire genome, or that are suitable for determining the methylation levels of individual genes. The CHARM method offers rather high precision and accuracy that facilitate a genome-wide methylation analysis (Irizarry et al., 2008). The HELP method also uses methylation-sensitive restriction enzymes to cut genomic DNA and can be combined with different microarrays or direct sequencing. when the methylation-insensitive isoschizomer, MspI, is added as a control, a certain level of methylation quantification can be achieved (Khulan et al., 2006; Oda et al., 2009; Suzuki and Greally, 2010).

Unfortunately, there is no single method currently available that combines good coverage and resolution with cost-effective high-throughput sample processing. Several studies have compared different methylation profiling methods, and they have demonstrated that while all of the methods examined had high concordance, a subset had method-specific drawbacks (Clark et al., 2012; Stevens et al., 2013; Walker et al., 2015). In a detailed and comprehensive analysis of four sequencing-based methods and the Infinium HumanMethylation450 BeadChip (Illumina, Inc., CA, USA), a high correlation and high concordance between the different approaches were observed (Walker et al., 2015). Sequencing-based methods do have the advantage that they cover the majority of CpG loci across a genome, while the Illumina 450K chip is restricted according to the design of the array. Coverage of the RefSeq genes and CGIs by the Illumina 450K chip is well comparable with other methods. Comparison of single CpG methylation levels has revealed that the concordance between sequencing methods and the 450K array was approximately 80%. Therefore, the choice of method for methylation profiling depends largely on the regions of interest and the number of samples, as well as the cost of analysis.

AIMS OF THE STUDY

The overall aim of the current thesis was to describe the methylation patterns in different types of tissues, both healthy and diseased, and their effect on gene expression. The specific aims were as follows:

1 To describe differentially methylated CpG sites in early-stage NSCLC samples and to identify potential diagnostic or prognostic biomarkers for NSCLC (Ref I).
2 To describe the general patterns of globally conserved and tissue-specific DNA methylation patterns in various human tissues and to identify their functional consequences in gene expression regulation (Ref II).
3 To compare epigenomes and transcriptomes of human fetal and adult liver tissues in order to investigate the relationship between DNA methylation and gene expression regulation, and to describe genetic and epigenetic effects on inter-individual variability in gene expression in adult liver tissues (Ref III).

RESULTS AND DISCUSSION

3.1 Methylation markers of early-stage NSCLC (Ref I)

3.1.1 Cohort description

To find biomarkers that describe the methylation changes in early lung cancer, we performed a genome-wide DNA methylation study of stage I NSCLC using the HumanMethylation27 BeadChips (Illumina, Inc). We analyzed 48 cancer samples and 18 macroscopically cancer-free "normal" lung control samples. All the specimens were isolated during lung surgery at Tartu University Hospital, Estonia. Patients did not receive any preoperative chemo- or radiotherapy. The patients with adenocarcinoma (n=6, 12.5%) and its subtype bronchioalveolar carcinoma (n=10, 20.8%) were analyzed as one group (n=16, 33.3%). The remaining 32 (66.7%) of the analyzed patients had squamous cell carcinoma. A detailed overview of the patients involved in our study is summarized in Table 2.

Table 2. Clinical description of the patient cohort.

Characteristics	Number of patients	%
Patients involved	48	
Age range (y)	41–80	
Male	40	83.3%
Range (y)	44–80	
Median (y)	67.5	
Female	8	16.7%
Range (y)	41–79	
Median (y)	67	
Histology		
Adenocarcinoma	6	12.5%
Bronchioalveolar carcinoma	10	20,8%
Squamous cell carcinoma	32	66.7%
History of smoking		
Yes	46	95.8%
No	2	4.2%
NSCLC stage		
Ia	13	27.1%
Ib	35	72.9%

3.1.2 Description of DNA methylation in NSCLC

A methylation analysis was performed by using Infinium® HumanMethylation27 RevB BeadChips (Illumina). At the time of the study, this chip was the most comprehensive and cost-effective microarray-based method for analyzing genome-wide DNA methylation. Details regarding the preparation of samples and data preprocessing are described in Reference I (Ref I).

A total of 496 hypermethylated CpGs associated with 379 genes, and 373 hypomethylated CpGs associated with 336 genes, were detected in samples from early-stage NSCLC tissue samples. A complete list of these sites is provided in Ref I Supplementary Table 1. The methylation profile of CpG sites between the normal and cancer tissue samples did not markedly differ, with only 17% (149 CpG sites) of the statistically significant sites showing a difference in their mean methylation levels that was greater than 0.2. The maximum value was 0.39. Hypermethylated sites were more often located in the CGIs near TSSs, while hypomethylated sites were found 5' upstream of the TSSs and in the CGI shores (the 2 kb regions upstream and downstream of a CGI) ($p < 0.0001$, Welch Two Sample t-test). Six cancer samples were clustered with tumor-free lung tissue samples, and these exhibited a methylation pattern that differed from the other tumor samples. Pathological examinations of these samples further revealed either low tumor content (10–30%) or a very heterogeneous tumor cell composition. Therefore, these samples were excluded from further analysis.

Among the aberrantly methylated genes, several well-known methylation markers for NSCLC were detected: *CDKN2A*, *MGMT*, *HOXA9*, *TRIM58*, and *CDO1* (Diaz-Lagares et al., 2016; Lin et al., 2014; Sandoval et al., 2013). Some of these markers were previously associated with early recurrence of stage I lung cancer, they displayed high specificity in initial diagnostics, and were recently shown to be detectable in various biological samples (Diaz-Lagares et al., 2016; Sandoval et al., 2013).

3.1.3 Correlation between DNA methylation and gene expression

The second specific aim of our study was to correlate changes in methylation levels with gene expression data. Pearson's analysis confirmed an inverse correlation between methylation levels and gene expression values for 378/869 of the CpG sites (43.5%). The highest inverse correlation values were observed between the hypermethylated CpGs and the expression values detected for *AGER* (–0.78), *EPOR* (–0.65), and *AQP1* (–0.63). Among the hypomethylated genes, the highest inverse correlation was observed for *MB* (–0.63), *ADA* (–0.60), and *MAGEA6* (–0.60). After multiple testing using permutation analysis, *AGER* (–0.81, $p = 0.0143$) and *MB* (–0.78, $p = 0.039$) showed statistically significant negative correlations between DNA methylation and gene expression

values. These results suggest that DNA methylation in the promoter region of these genes may have a functional role in regulating gene expression. In NSCLC cases, aquaporin 1 (AQP1) has been found to be hypermethylated and downregulated (Ehrich et al., 2006). Similarly, the cell surface receptor for advanced glycation end products (AGER) has been found to be downregulated in NSCLC cases, although the treatment with 5-aza-2′-deoxycytidine did not reinduce the expression of *AGER* mRNA (Bartling et al., 2005). Despite the observation that the inverse correlation was found between DNA methylation and gene expression for approximately half of the CpGs examined, it is important to note that the HumanMethylation27 chip analyzes the methylation status of only a few CpGs per gene, and this might not reflect the overall methylation level of any given promoter. Thus, there is potential for bias to be present in the analyses of the chip data.

3.1.4 Ingenuity pathway analysis

In silico functional and interaction analyses of the differentially methylated genes identified were performed using Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, CA, USA). This analysis revealed that the dominant functions of the differentially methylated genes involved cell-to-cell signaling and interactions, DNA replication and repair, cellular growth and proliferation, cell death, cancer, and the inflammatory response. These functions are consistent with the large body of data that indicate that genes that mediate cell growth, proliferation, and cell death are directly involved in cancer progression (Ansari et al., 2016). Tumor suppressor genes are also responsible for many of the gene functions mentioned above, aberrant methylation of CGIs located in tumor suppressor gene promoters being well characterized (Jones and Baylin, 2002). Many of the genes involved in DNA repair, apoptosis, cellular movement, and invasion have also been found to be dysregulated by aberrant DNA methylation in different tumors (Ansari et al., 2016).

3.1.5 Survival and smoking analysis

Two types of survival analysis were performed to identify potential prognostic methylation markers. First, a Kaplan-Meier survival test was performed for each of the CpG sites. The corresponding methylation values were categorized as low (0–0.25), medium (0.25–0.75), and high (0.75–1). Among these, the methylation levels associated with 10 CpGs in 10 genes were associated with differences in patient survival (Figure 5).

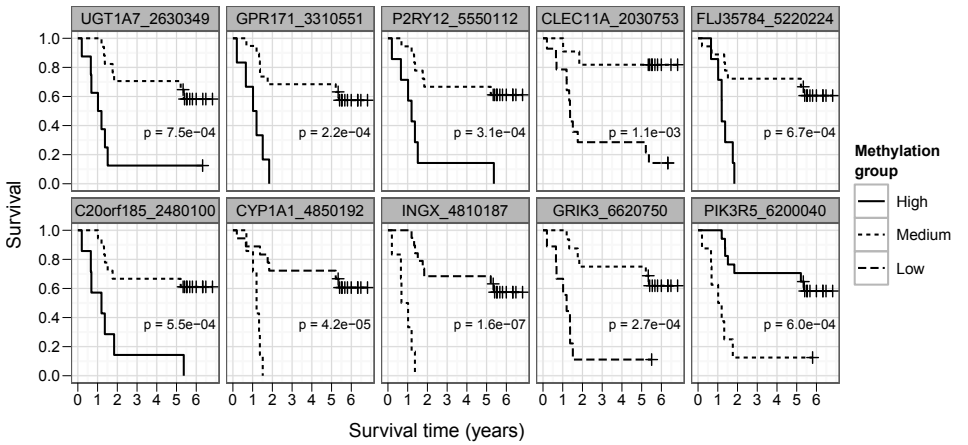

Figure 5. Survival curves of ten differentially methylated CpG sites. Survival test was performed for each of the indicated CpG sites. The methylation values obtained were categorized as low (0–0.25), medium (0.25–0.75), or high (0.75–1). The methylation levels for the ten CpG sites shown differed according to patient survival. The x-axes represent the years of survival and the y-axes show the cumulative probability of survival.

Patients with medium levels of methylation at the CpG sites in *GPR171*, *UGT1A7*, *P2RY12*, *FLJ35784*, and *C20orf185* exhibited better survival than the patients with high levels of methylation at the corresponding sites. Similarly, patients with medium methylation levels at the CpG sites in *CLEC11A* and *GRIK3* exhibited better survival than the patients with low levels of methylation at the same sites. In contrast, patients with medium levels of methylation at the CpG sites in *CYP1A1* and *INGX* had worse survival that those with low methylation level. Patients with high levels of methylation at the CpG sites in *PIK3R5* were associated with better survival than the patients with medium levels of methylation at the same sites. A recent study shows that overexpression of *GPR171* is responsible for inducing NSCLC proliferation, invasion, and migration of lung cancer cells (Dho et al., 2016). In contrast, the patients in our study that had higher DNA methylation levels exhibited worse survival. Nevertheless, an online survival analysis software (Győrffy et al., 2013) showed similarities to our study using gene expression data. For example, the patients with low levels of *GPR171* expression had a worse survival ($p = 1.2 \times 10^{-9}$). *UGT1A7* is an enzyme involved in the metabolism of (pre)carcinogens present in tobacco smoke. Precarcinogens and their metabolites are considered to play an important role in the carcinogenesis of tobacco smoke-related cancers (Strassburg et al., 1999). Correspondingly, polymorphisms in the *UGT1A7* gene have been associated with lung cancer, and it is hypothesized that these polymorphisms reduce the enzymatic activity of *UGT1A7* (Araki et al., 2005). A high methylation level could also potentially affect *UGT1A7* activity and lead

to a poor prognosis for NSCLC patients. *CYP1A1*, which belongs to the cytochrome P450 superfamily, catalyzes many of the reactions that are involved in drug metabolism, the conversion of polycyclic aromatic hydrocarbons into reactive metabolites, and the detoxification of environmental carcinogens. Previously, *CYP1A1* was found to be hypermethylated in lung cancer samples, and this corresponded with reduced mRNA levels of *CYP1A1* in the same samples (Tekpli et al., 2012). In our study, higher methylation levels in the lung cancer samples were associated with poor survival, and these results support the hypothesis that *CYP1A1* has a protective role in cancer progression.

A second type of survival analysis was performed and it combined a Cox proportional hazard analysis and the Wilcoxon rank-sum test to analyze differential methylation. Twelve patients with a post-surgery survival period of less than 24 months, and 15 patients who survived 60 months or more after surgery, were analyzed. A total of 15 differentially methylated genes in relation to patient survival were identified, and these included: *SOCS2*, *ACTA1*, *ALDH1A3*, *RTEL1*, *MTM1*, *DXS9879E, SCUBE3, SYT2, GRIK3, CRB1, ZNF660, MDFI, SRD5A2, KCNC4*, and *KCNC3* (Ref I, Figure S6). Among these, *RTEL1* is a regulator of telomere elongation helicase 1 and is required for the suppression of inappropriate homologous recombination events. Accordingly, with central roles in DNA repair and the maintenance of genomic stability, *RTEL1* mediates tumor suppressive functions (Uringa et al., 2011). In the patients of our cohort that exhibited poor survival, *RTEL1* was found to be hypermethylated ($p = 0.01$). *MTM1*, the gene encoding myotubularin 1, is required for muscle cell differentiation. In our cohort, *MTM1* was hypermethylated in the patients with poor survival ($p = 0.007$). The online survival analysis software we used also showed that lower levels of *MTM1* expression were associated with poor survival, while higher levels were associated with better survival ($p = 1.8 \times 10^{-6}$) (Győrffy et al., 2013). In several types of cancer, potassium channels have been found to be overexpressed, and this enhances tumorigenic processes such as proliferation and metastasis (Huang and Jan, 2014). In our dataset, hypomethylation of *KCNC3* ($p = 0.0006$) and *KCNC4* ($p = 0.004$) were associated with a shorter survival period. In patients with NSCLC, overexpression of *SCUBE3* has correlated with significantly shorter survival times compared to patients with low levels of *SCUBE3* (Zhao et al., 2013). In our dataset, hypomethylation of *SCUBE3* was associated with a short survival period after surgical resection ($p = 0.03$). However, in a previous study of renal cell carcinoma, hypermethylation of *SCUBE3* was associated with a significantly worse survival (Morris et al., 2011). Taken together, these results suggest that aberrant methylation of *SCUBE3* may be a possible prognostic factor for NSCLC.

To investigate whether smoking affects the DNA methylation patterns in a tumor, linear regression was applied based on pack-years data. Our analysis of tumor samples did not detect any differentially methylated genes that were related to the extent of tobacco smoking. However, when we compared data from the limited number of non-smokers in our cohort ($n = 3$, 6.4%) with the

data from the smokers (n = 44, 93.6%), four differentially methylated CpG sites were identified in three genes ($p < 0.05$, False discovery rate (FDR) adjusted): *CXorf38*, *MTHFD2*, and *TLL2*. Furthermore, all three genes were hypo-methylated in the smokers group.

3.2 DNA methylome profiling of human tissues to identify global and tissue-specific methylation patterns (Ref II)

3.2.1 Samples analyzed

To study tissue-specific DNA methylation patterns, 17 postmortem human somatic tissues were collected from four individuals at the time of autopsy. The somatic tissues included: abdominal and subcutaneous adipose tissue, bone, joint cartilage, yellow and red bone marrow, coronary and splenic artery, abdominal and thoracic aorta, gastric mucosa, lymph node, tonsils, bladder, gall bladder, medulla oblongata, and ischiatic nerve. The causes of death for the individuals included: intracerebral hemorrhage (BM419/4; female, 60 years old), heart attack with acute cardiac insufficiency (KA522; male, 53 years old), heart attack (KT538; male, 40 years old), and intracerebral hemorrhage (SJ600-5; male, 54 years old).

3.2.2 Genome-wide DNA methylation patterns

The expanded Illumina Infinium HumanMethylation450 BeadChip was used to survey 486,428 CpG sites in the human genome. This advanced platform offers unbiased coverage of gene and CpG island regions reaching up to 99% and 96%, respectively. In addition, the CGI shores (2 kb regions upstream and downstream of the CpG islands) and CGI shelves (2 kb regions upstream and downstream of the CGI shores) to reveal a genome-wide methylation profile (Bibikova et al., 2011). Detailed information regarding the preparation and processing of DNA for this chip analysis are provided in Reference II (Ref II).

The methylation profiles obtained were well conserved between the 17 tissues studied, with the pairwise correlation ranging from 0.93 to 0.99. Thus, only subtle dissimilarities existed between this diverse set of tissues (Ref II, Figure 1). The highest correlations were found among functionally similar tissues, such as the different arteries and aortas, red and yellow bone marrow, and bone and joint cartilage (Pearson correlation coefficient (PCC): ≥ 0.99). Although the number of analyzed individuals was limited, hierarchical clustering of the methylation profiles showed that most of the similar tissues (e.g., the aortas and arteries) were co-clustered (Figure 6). Based on the strong correlations that were observed between similar tissues, it appears that tissue-specific methylation profiles were obtained.

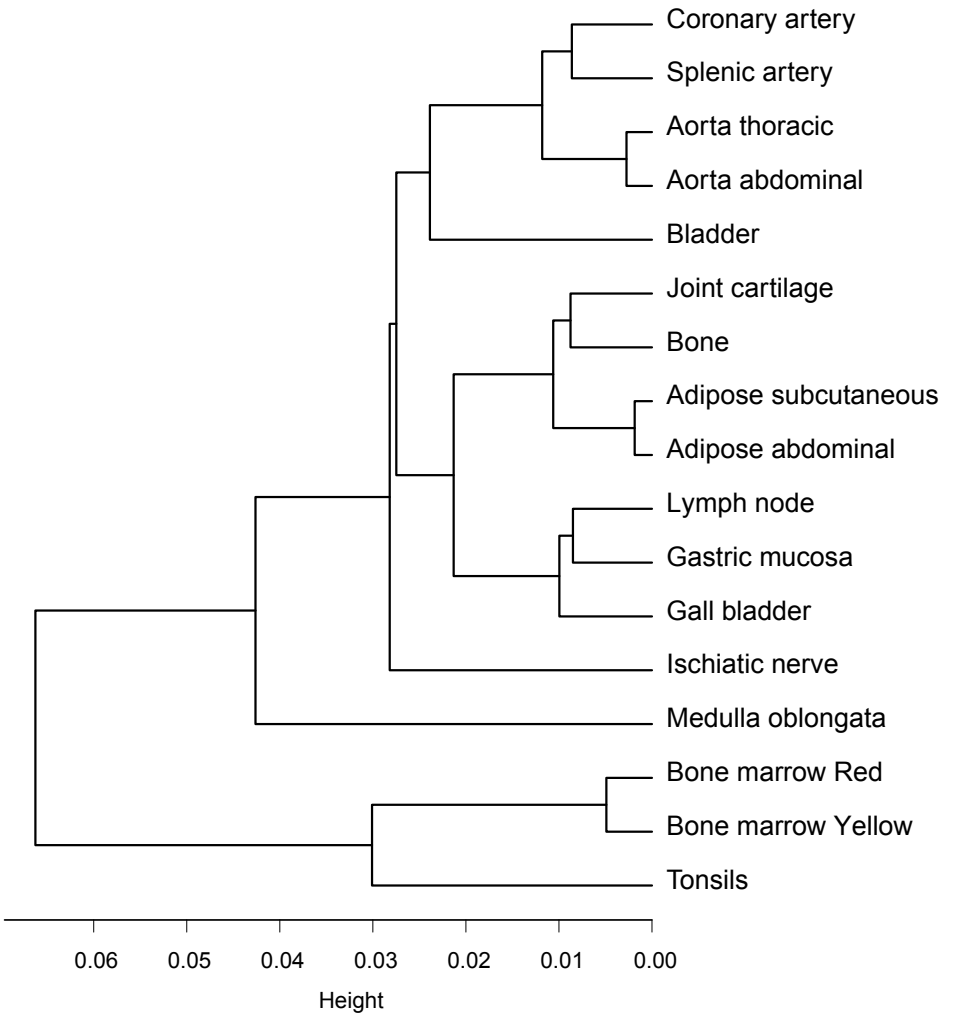

Figure 6. Hierarchical clustering of the 17 tissues studied. A hierarchical clustering analysis was performed using the hclust command in R. All of the samples were merged according to their corresponding tissues. As a result, a matrix of the mean beta values for all of the CpG sites detected in the 17 total tissues was generated. A clustering tree was generated using the complete method and a strong correlation between similar tissue types was observed.

Global distribution of the methylated CpGs that were detected in the somatic tissues analyzed indicates that a significant portion of the detected CpGs were either unmethylated (0%) or fully methylated (100%) (Ref II, Additional File 2). The collected data for all 17 tissues reveal that only 2.2% of all of the CpGs were hypermethylated in all of the samples (10,707 CpGs representing 4,416 genes; beta values > 0.9). These invariably methylated CpGs were mostly located in gene bodies, in the 3'-untranslated regions (UTRs) (66.8%, 7,150

CpGs; Ref II, Figure 3), or in the non-CGI regions (77.4%, 8,287 CpGs; Ref II, Figure 4A) (Fisher's exact test, $p < 2.2 \times 10^{-16}$). Thus, DNA methylation appears to be more prominent in the areas where CpG density is low and transcription is not often initiated.

On the other hand, 14.9% of CpGs (72,444 CpGs representing 12,604 genes) were hypomethylated in the samples examined (beta values < 0.1). These invariably hypomethylated CpGs were mostly located in gene promoter areas (73.2%, 53,057 CpGs), including the sequence region –200 to –1500 nt upstream of the TSS (TSS1500), the region –200 nts upstream of the TSS (TSS200), and the region extending from the 5'-UTR through the first exon (Ref II, Figure 3). In addition, hypomethylated CpGs were mostly found in CGI regions (73%, 52,862 CpGs; Ref II, Figure 4A) (Fisher's exact test, $p < 2.2 \times 10^{-16}$). These findings are consistent with the general consensus that gene promoter areas and CGI regions of actively transcribed genes are largely unmethylated to maintain their accessibility to transcription factors.

A gene ontology (GO) analysis was performed by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al., 2003). This analysis revealed that many of the genes with hypermethylated CGI-promoters had functions related to the reproductive system. In contrast, many of the genes with hypomethylated CGI-promoters had functions associated with housekeeping processes, including RNA processing and the cell cycle. A number of housekeeping genes were identified by expression profiling by Chang et al. (Chang et al., 2011), we found 93% overlap with our housekeeping genes that had hypomethylation at CGI-promoters.

3.2.3 Comparison of DNA methylation in CGI and non-CGI regions

It is well recognized that DNA methylation patterns can differ significantly across the different regions of CpG islands, with methylation levels increasing at the boundaries. Similarly, we found that the highest methylation levels were observed in the CGI shores and shelves that were examined (Ref II, Figure 4A). These results are in agreement with those of previous studies (Davies et al., 2012; Eckhardt et al., 2006), where the majority of CGIs were found to be unmethylated. CGI methylation patterns were also similar when the intergenic and intragenic regions were analyzed separately (Ref II, Figure 4, B and C). It is possible that maintenance of an unmethylated state in CGIs may provide protection from mutations that can occur via spontaneous deamination of methylated cytosines in functionally important regions (Duncan and Miller, 1980; Pfeifer, 2000).

Overall, CGI methylation has been found to vary across different gene regions, and this observation was also made in our analysis. We found that the promoter regions (TSS1500, TSS200, and 5'-UTR) and the first exon regions were almost exclusively unmethylated, while variable CGI methylation levels

were detected in the gene bodies and 3'UTRs (Ref II, Figure 5A). Moreover, this pattern has been observed in other studies as well (Eckhardt et al., 2006; Rakyan et al., 2008). In contrast, the CpGs that were located in non-CGIs were mostly methylated, and they showed little variation between the different genomic locations (Ref II, Figure 5B).

3.2.4 Tissue-specific differentially methylated regions

Next, regions with distinct methylation patterns in certain tissues were analyzed in detail. An algorithm was applied to identify statistically significant differential methylation events that existed between two sets of samples in three or more consecutive CpG probes (Kolde et al., 2016). This new method is based on fitting analysis of variance (ANOVA) models in moving windows of different lengths, encompassing up to 50 probes. The optimal region boundaries were selected according to the minimum description length (MDL) principle. As a result, every region consists of probes with similar methylation patterns. Since the HumanMethylation450 BeadChip is focused more on genes and promoter areas, this robust approach finds more likely areas with higher CpG probe densities.

When this method was used to compare tDMRs present in one tissue of interest with all of the other tissues that were examined, we found that the data from certain tissues could be combined due to high level of methylation similarities that existed between them. Specifically, the abdominal and subcutaneous adipose tissues were processed together, as were the thoracic and abdominal aorta, coronary and splenic artery, joint cartilage and bone, and red and yellow bone marrow tissues.

The numbers of different tDMR CpG blocks were found to greatly differ between the studied tissues (Table 3). The highest number of hypermethylated tDMRs were found in tonsil tissues, followed by medulla oblongata and the aorta tissues (abdominal and thoracic). The lowest number of hypermethylated tDMRs were observed in the lymph nodes. It is possible that the large number of hypermethylated blocks found in the medulla oblongata is biased by the 5hmC content of that tissue (Chen et al., 2014), with bisulfite treatment analysis unable to distinguish between 5mC and 5hmC. The largest number of hypomethylated tDMRs were found in both red and yellow bone marrow, as well as in both aorta tissues (abdominal and thoracic) and the ischiatic nerve. In contrast, lower numbers of hypomethylated tDMRs were found in the lymph nodes.

Table 3. Summary of tDMR data.

Tissue	No. of hyper-methylated blocks with gene annotation	No. of hypo-methylated blocks with gene annotation
Adipose (subcutaneous, abdominal)	65	301
Artery (coronary, splenic)	280	219
Bone, joint cartilage	73	104
Bone marrow (red, yellow)	150	1,028
Gastric mucosa	54	22
Lymph node	42	3
Tonsils	3,893	924
Bladder	274	566
Gall bladder	47	66
Aorta (thoracic, abdominal)	453	888
Medulla oblongata	495	278
Ischiatic nerve	156	861

Of the 14,441 tDMRs that were identified (Ref II, Additional File 3), 11,242 (77.8%) mapped to genes. Among the latter, 41.7% were in gene promoter areas, with 36.5% of these present in CGIs (Fisher's exact test, $p < 2.2 \times 10^{-16}$); while 58.3% were in gene body regions, with 44.1% of these present in CGIs (Fisher's exact test, $p < 2.2 \times 10^{-16}$). Based on the observation that > 50% of the tDMRs were located in gene bodies and not within promoter areas might indicate the presence of alternative promoters (Maunakea et al., 2010). However, it was previously obsesrved that tDMRs in gene bodies, or even in intergenic regions, more frequently exhibit tissue-specific methylation patterns (Deaton et al., 2011; Yang et al., 2015). In our study, 45.8% of the intergenic tDMRs co-localized with CGIs (Fisher's exact test, $p = 0.0003$). We hypothesize that these intergenic regions may act as repressors or enhancers of gene expression to maintain tissue-specific gene expression.

To study the most variable regions between tissues, the proportion of variance explained by the tissues between different gene regions and CGIs, shores, and shelves were compared (Ref II, Additional File 4). Our results show that large differences can be found in gene body, 3'UTR, intergenic, and non-island regions, as well as in CGI shores and shelves.

To characterize the functions of the genes related to the detected tDMRs, a GO analysis with the DAVID database was performed. A custom background in the GO enrichment analysis was used which contained all of the genes identified as tDMRs. Thus, the distribution of CpG probes in the microarrays would be accounted for. As shown in Table 4, the hypomethylated genes in certain tissues were frequently associated with tissue-specific functions.

Table 4. GO analysis of hypomethylated tDMRs

Tissue	GO term	No. of genes	P-value
Adipose tissue	Lipid homeostasis	5	0.0096
(abdominal,	White fat cell differentiation	3	0.0172
subcutaneous)	Fat cell differentiation	4	0.0532
Artery	Blood vessel morphogenesis	12	3.24E-04
(coronary, splenic)	Angiogenesis	10	4.25E-04
	Blood vessel development	13	4.60E-04
Aorta	Cardiac muscle tissue development	11	5.91E-04
(thoracic, abdominal)	Muscle organ development	24	8.21E-04
	Striated muscle tissue development	16	9.72E-04
Bone, joint cartilage	Chondrocyte differentiation	3	0.0067
	Cartilage development	4	0.0253
	Skeletal system development	7	0.0553
Bone marrow	Cell activation	41	5.07E-07
(red, yellow)	Leukocyte activation	33	7.98E-06
	Immune response	62	1.97E-05
Lymph node	–		
Tonsils	Immune response	59	7.00E-06
	Regulation of T cell activation	17	4.38E-05
	Defense response	50	7.22E-05
Gastric mucosa	Regulation of pH	2	0.0530
	Monovalent inorganic cation homeostasis	2	0.0677
Bladder	Muscle contraction	14	0.0034
	Excretion	7	0.0266
	Secretion	17	0.0379
Gall bladder	Negative regulation of granulocyte differentiation	2	0.0417
	Negative regulation of immune system process	3	0.0496
	Regulation of granulocyte differentiation	2	0.0519
Medulla oblongata	Homophilic cell adhesion	15	8.78E-06
	Cell-cell adhesion	18	5.00E-04
	Cell adhesion	25	0.0151
Ischiatic nerve	Filopodium assembly	5	0.0023
	Regulation of action potential in neuron	10	0.0036
	Negative regulation of neurogenesis	7	0.0074

For example, the hypomethylated genes detected in arteries (including *COL18A1*, *EPAS1*, *ENPEP*, *ANGPT2*, and *APOLD1*) have been characterized as mediators of blood vessel development and morphogenesis, while those detected in tonsils (including *LAX1*, *TNFSF14*, *LCK*, and *RHOH*) have been shown to have roles in the immune response and leukocyte activation. In contrast, and in agreement with previous results, none of the hypermethylated genes were associated with tissue-specific functions (Ref II, Additional File 5) (Nagae et al., 2011; Yang et al., 2015; Zilbauer et al., 2013). Taken together, these results strongly support the hypothesis that hypomethylation, rather than hypermethylation, in some regions of the genome is more likely to be associated with tissue-specific functions.

3.2.5 Inter-individual methylation variations

We analyzed the rate of inter-individual variation to understand whether differences between individuals or tissues could account for the observed variability between samples. For this, the proportion of variance of the beta values explained by the individuals and the proportion of variance explained by the tissues were compared. While differences between individuals only accounted for 6.4% of the variance observed, tissue-based differences accounted for 51.2% of the variance observed (Ref II, Figure 6). Thus, variance between individuals for the data examined was insignificant.

Hierarchical clustering of all the samples studied also showed that the similarity between tissues was higher than between individuals, as the tissue data mostly clustered together (Ref II, Additional File 6).

3.2.6 Relationship between gene expression and global DNA methylation

To further investigate the role of DNA methylation in regulating gene expression, the detected methylation patterns were compared with publicly available gene expression data maintained in the Gene Expression Omnibus and ArrayExpress databases. Only tissues with gene expression data obtained using a single platform (Human Genome U133A arrays; Affymetrix, CA, USA) were selected to reduce the impact of potential confounding factors. As a result, correlations between gene expression levels were performed for 8/17 of the tissues used in the original analysis: aorta, bladder, bone, bone marrow, coronary artery, lymph node, medulla oblongata, and tonsil.

The method by which the global methylation data were correlated with the gene expression data relied on averaging beta values across the comprehensive gene panel. PCCs were calculated for 10,120 genes across the eight tissues (Ref II, Table 3) and showed a slight bias towards negatively correlated genes' expression (5,710 *vs*. 4,410 positively-correlated genes). Also, nearly twice as many

genes exhibited a strong inverse correlation (1,713 genes, PCCs: < -0.5) compared with the genes that exhibited a strong positive correlation (1,090 genes, PCCs: > 0.5) (Fisher's exact test, $p < 2.2 \times 10^{-16}$).

When analyzing the global methylation data within different gene regions, the number of negatively correlated genes in the CGI-promoter regions (56.7%) was found to be almost the same as that in the gene bodies (52.7%). A similar observation was made for the strong inverse correlation, with both the CGI-promoter and body region showed a bit more negatively correlated genes than positively correlated genes (11.5% and 10.6% in promoter-CGI regions and 16.1% and 12.1% in the gene body regions, respectively in each case) (Fisher's exact test, $p = 0.005$).

The methylation and gene expression data were plotted onto a single figure to further assess the integrity of the correlation between the CGI-promoter and gene body regions (Ref II, Figure 7). A large proportion of the CpG sites in the CGI-promoter regions appeared to be unmethylated, an L-shaped correlation can be seen, where high levels of methylation were found to be associated with lower levels of gene expression and low levels of methylation associated with various levels of gene expression. However, no relationship was observed for the data related to gene bodies (Ref II, Figure 7B).

Our analysis of highly methylated promoters suggested a possible link between promoter methylation and suppressed gene expression. These results are consistent with those of previous studies where genes with unmethylated promoters were found to exhibit variable levels of transcription activity (Eckhardt et al., 2006; Mendizabal and Yi, 2016; Rakyan et al., 2008). In contrast, methylation analysis within the gene body regions did not reveal any clear relationship with mRNA expression levels, despite previous studies reporting a positive correlation (Rakyan et al., 2008; Yang et al., 2015) or a bell-shaped correlation pattern for their data (Jjingo et al., 2012). Many genes harbor several alternative TSSs, and these are located throughout the gene body and yield different splice isoforms. Methylation of these yet unrecognized sites could confound a correlation analysis of gene body methylation and gene expression, and further studies are needed to confirm this possibility.

3.2.7 Gene expression and methylation in tDMRs

We also performed a correlation analysis of gene expression and DNA methylation levels separately on tDMRs. Collectively, there were more negative than positive correlation coefficients (63.2%, 2,288 *vs*. 36.8%, 1,322; Figure 7; Ref II, Table 4), as expected. In addition, strongly negative PCCs prevailed over the strongly positive PCCs (20.7%, 749 vs. 10.3%, 372, respectively) (Fisher's exact test, $p < 2.2 \times 10^{-16}$).

Surprisingly, however, there were relatively more negative correlations also in the gene body regions (60.9%, 1,148), and in general, gene body methylation was not related to low gene expression levels. The finding that a high number of

inversely correlated CpG sites in the CGI-promoter regions (78.5%, 489) and the genes with highly methylated promoter regions were both associated with suppressed gene expression suggest that methylation in promoter regions correspond to changes in gene expression. However, it remains unclear how gene body tDMRs may function as regulators of gene expression. Based on the observation that DNA methylation is enriched in alternatively spliced exons, it has been proposed that methylation in intragenic regions could modulate alternative splicing (Maunakea et al., 2013).

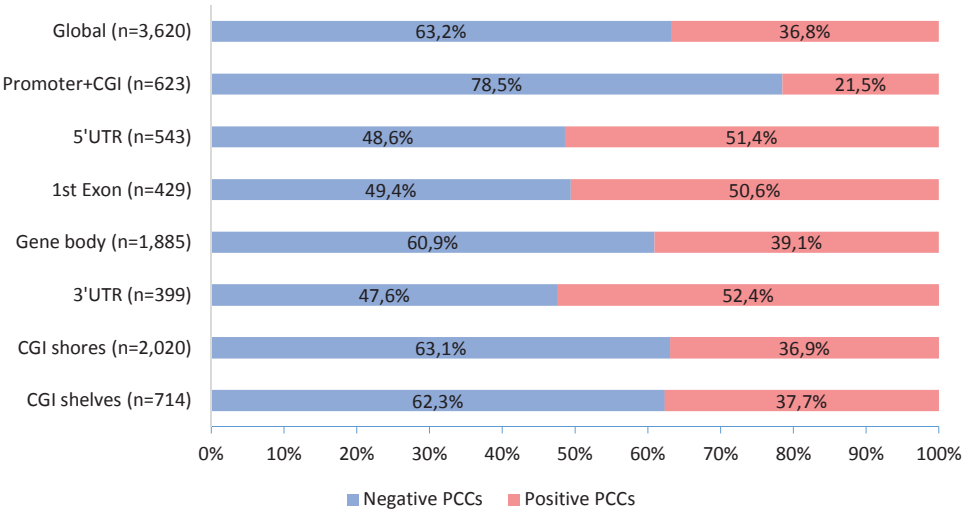

Figure 7. Correlation between gene expression and methylation in tDMRs. The percentages of negative (shown in blue) and positive (shown in red) PCCs are shown for the various genomic regions within the analyzed tDMRs. The total number of tDMRs in each category is indicated in parentheses at the far left.

3.3 Genetic and epigenetic regulation of gene expression in fetal and adult human livers (Ref III)

3.3.1 Samples analyzed

A total of 96 adult livers and 14 fetal livers obtained from 8- to 21-week-old human fetuses (Karolinska Liver Bank) were used to investigate gene expression regulation and DNA methylation levels. Data from the adult livers were combined with data from 85 adult liver samples obtained from a Dutch population (Fu et al., 2012; Wolfs et al., 2010) to investigate the regulation of gene expression by genetic and epigenetic factors. In addition, tissue specificity was explored by using single nucleotide polymorphism (SNP), methylation, and gene expression data from muscle, subcutaneous- (SAT), and visceral adipose tissue (VAT) that were collected from the same individuals in the Dutch cohort. A detailed description of the samples used in this study is provided in Ref III Additional File 1.

3.3.2 Developmental regulation of hepatic gene expression

Previous studies on the regulation of gene expression in human liver have mainly concentrated on the effect of genetic variations in adult samples (Fu et al., 2012; Greenawalt et al., 2011; Innocenti et al., 2011; Schadt et al., 2008; Schröder et al., 2013). In this study, we investigated the developmental regulation of gene expression in human livers by comparing the expression levels and methylation levels of genes in adult and fetal livers. In addition, both genetic variants and differences in DNA methylation were examined to explain the observed variability in transcript levels in adult livers.

3.3.2.1 The epigenome of the developing human liver

DNA methylation in fetal and adult liver tissues were analyzed with Infinium HumanMethylation450 BeadChips (Illumina). A total of 28,917 CpG sites in 12,619 unique genes showed significant differences (delta beta > 0.2, FDR < 0.05) in fetal liver tissue versus adult liver tissue. Moreover, the number of hypomethylated CpG sites in fetal liver (53.4%) was similar to the number of hypermethylated sites (46.6%).

To explore the functions of these differentially methylated genes, the GREAT pathway tool (McLean et al., 2010) was used. The CpG sites that were found to be hypomethylated in the adult liver were largely associated with metabolic pathways (Ref III, Table 1A). On the other hand, the genes that were hypomethylated in the fetal tissues were strongly enriched for the pathways of insulin receptor signaling, regulation of glycogen synthase activity, differentiation processes, and developmental functions (Ref III, Table 1B). Similar results were previously described for the transcriptome of mouse fetal livers at different stages of development (Jochheim-Richter et al., 2006; Lee et al., 2012; Li et al., 2009).

3.3.2.2 Transcriptome of the developing liver

Comparison of gene expression levels between fetal and adult liver samples yielded 3,284 differentially expressed probes (absolute log2-fold change > 1.0, FDR < 0.05; Ref III, Additional File 4). A GO analysis using Gene Network (Cvejic et al., 2013) further revealed that genes with higher levels of expression in adult liver were strongly enriched for metabolic processes involving monocarboxylic acid, steroids, and bile acid (Ref II, Table 2A). In contrast, the genes that were highly expressed in the fetal liver tissues were associated with regulation of organelle organization, chromosome organization, and tetrapyrrole (e.g., hemoglobin) biosynthetic processes (Ref II, Table 2B). These observations are consistent with the process of fetal development which is characterized by tissue differentiation and growth. Furthermore, the liver is predominantly a hematopoietic organ during this period (Moscovitz and Aleksunes, 2013).

3.3.2.3 Orchestration of epigenetics and transcriptomics in regulating liver development

There were 1,655 genes that exhibited both differential expression and differential methylation in adult versus fetal livers (Ref III, Additional File 5). The location of the differentially methylated CpG sites differed significantly in relation to the CGIs, depending on the expression and methylation differences between the fetal and adult tissues (Chi-squared test, $p < 2.2 \times 10^{-16}$; Ref III, Figure 1C). For example, the CpG sites that were hypomethylated in the fetal liver tissues were more often located in the CGIs, and were associated with both increased and decreased gene expression. In contrast, the hypermethylated CpG sites in the same tissues were more distant from the CGI regions, yet were still associated with both increased and decreased gene expression.

The twenty genes that exhibited the largest differences in expression and methylation are listed in Ref III Table 3. These results clearly illustrate that genes involved in differentiation and hematopoiesis (e.g., *DLK1*, *HBZ*, *HBM*, *AHSP*, *EPB42,* and *NFE2*) exhibited fetal liver-specific expression profiles, while the adult liver-specific expression profiles involved drug metabolism, catabolism, and other biosynthesis processes. Two cytochrome P450 genes exhibited the most significant differences in expression levels between the fetal and adult liver tissues, with the expression levels of *CYP2E1* and *CYP2C8* being approximately 7-fold higher in the adult liver.

3.3.3 Genetic and epigenetic effects on inter-individual variability in gene expression

3.3.3.1 Correlation in DNA methylation and gene expression

We next assessed whether DNA methylation correlated with gene expression levels in the adult liver samples. Fetal samples were excluded from this analysis due to the significant developmental differences reported above. Furthermore, we estimated that the fetal samples would not add any considerable statistical power for the analyses. Data from the Karolinska Liver Bank and Dutch liver samples were combined to provide 158 samples with available expression and methylation data. Expression probes with CpG sites that mapped within 250 kb of these probes were then compared. A total of 3,238 significant methylation-expression associations (eQTM; Ref III, Additional File 7) were identified, and these comprised 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. As expected, a greater number of eQTMs showed a negative correlation between gene expression and CpG methylation levels (58.4%), irrespective of the location of the CpG site in relation to the CGIs. Both negative and positive correlations have been described by other groups (Bell and Spector, 2012; Gutierrez-Arcelus et al., 2013; Huse et al., 2015), with a modest, yet significant, excess of negative correlations between DNA methylation and variation in gene expression levels

observed across individuals. Furthermore, for CpG sites with a strong correlation between expression and methylation levels, and/or a location within 50 kb of the expression probes, we observed an overrepresentation of negative correlations (Chi-squared test, $p < 2.2 \times 10^{-16}$; Figure 8).

It was previously reported that the role of DNA methylation appears to depend on genomic context (Jones, 2012). For example, CpG sites located near genes and/or CpG sites exhibiting a stronger correlation between methylation and expression are considered more likely to display a negative correlation. Interestingly, CpG sites downstream of the expression probes in the present study displayed fewer negative correlations than those that were upstream of the probes, thereby indicating an association between methylation in gene bodies and active gene expression. This phenomenon was also noted in the early days of DNA methylation research (Jones, 1999; Wolf et al., 1984). Both of these sets of observations are explained by the fact that, in mammals, DNA methylation silences transcription initiation, but not transcription elongation (Jones, 2012).

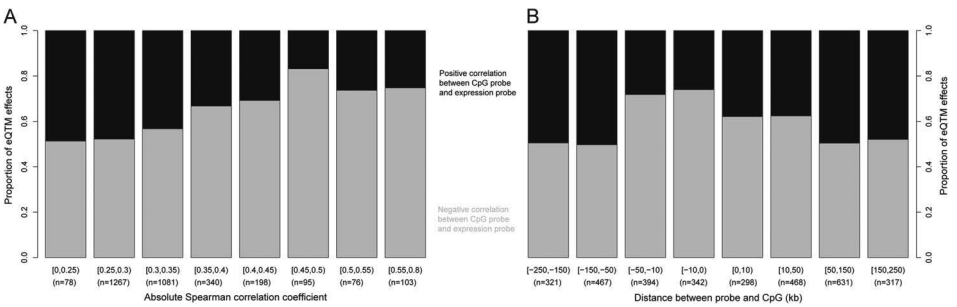

Figure 8. Distribution of the direction of expression and methylation correlation coefficients. The proportion of eQTM effects (y-axis) are grouped according to: (**A**) the absolute Spearman correlation coefficient values and (**B**) the distance between the expression probe and the CpG site in kilobases (kb). The grey and black shaded bars represent the negative and positive correlations between the expression probes and methylation CpG sites, respectively.

3.3.3.2 Regulation of gene expression by genetic polymorphisms

We next explored the effects of genetic variations on gene expression levels. To conduct a meta-analysis of the Karolinska Liver Bank and Dutch liver samples, eQTL mapping of the adult liver samples was performed. As a result, 171 samples with expression and genotype data available were identified. A total of 47,168 significant SNP-probe pair correlations (FDR < 0.05) were identified, and these represented 751 unique genes (Ref III, Additional File 8). The eQTL probes were significantly enriched for liver-specific genes ($p < 4 \times 10^{-57}$, as reported by Gene Network) and for genes encoding drug-metabolizing enzymes ($p < 2.0 \times 10^{-19}$). In addition to the 667 eQTL genes that were previously reported

(Greenawalt et al., 2011; Innocenti et al., 2011; Schadt et al., 2008; Schröder et al., 2013; Yang et al., 2010), 84 new associations were identified in the present analysis (Ref III, Additional File 8). These new associations are potentially due to the inclusion of a larger number of samples and imputation of SNPs that were not previously used in genotyping arrays and that derived from the 1000 Genomes project (Auton et al., 2015).

3.3.3.3 Contribution of genetic variants and DNA methylation to variation in hepatic gene expression

After the eQTL and eQTMs were identified, it was investigated to what extent both SNPs and DNA methylation events explain variations in liver gene expression levels. For this, 293 expression probes (reflecting 274 unique genes) that had both a significant *cis*-eQTL and a significant eQTM effect were selected. Four different linear models (Ref III, Additional File 1) were then tested to assess the proportion of variations in gene expression that could be explained. For 83% of these 293 expression probes, most of the variations were explained by a SNP (Ref III, Additional File 11), whereas the expression variance for the remaining 17% was most strongly explained by a specific CpG site. As expected, when we combined the SNP genotype and CpG methylation effects, we could explain more of the expression variations than by using the SNP or methylation levels alone. Furthermore, the SNP sites and CpG sites with particularly high correlations with the expression levels were found to be closer to the TSS of the corresponding genes (Ref III, Additional File 13).

Inter-individual variability in ADME gene expression has been shown to affect drug efficacy, toxicity, and susceptibility to environmental toxins (Ingelman-Sundberg et al., 2007). In the present study, 16 unique ADME genes had both a significant eQTL and eQTMs. These genes included members of the glutathione S-transferase (GST) family of phase II ADME isozymes (*GST4A*, *GSTM1*, *GSTM3*, *GSTO2*, and *GSTT1*), solute carrier transporters (*SLC19A1* and *SLC22A18*) responsible for the transmembrane transfer of multiple drugs and endogenous compounds, as well as *FMO4*, *GPX7*, *PON1*, and *UGT1A1* (Ref III, Additional File 15, Table 4). *GSTM1* functions in the detoxification of exogenous/endogenous toxins, while *GSTT1* is involved in the conjugation of a variety of compounds (Marinković et al., 2013; Ramos et al., 2011; Tulsyan et al., 2013; Zhong et al., 2006). The effects of epigenetic modifications on the expression of these genes have been reported in both blood and brain tissues (Liu et al., 2013c; Sintupisut et al., 2013). In our analysis, DNA methylation explained a larger proportion of the variation in the gene expression levels of *GSTM1*, *GSTO2*, *GSTT1*, *PON1*, and *UGT1A1*.

Overall, we found that both SNPs and DNA methylation contribute to the variable expression profiles observed for ADME genes. For example, SNP rs2739330, which is located downstream of the *GSTT1* gene and upstream of the *DDT* gene, has been reported to be associated with gamma-glutamyl transferase levels in plasma (Chambers et al., 2011). This SNP, together with

methylation levels of a nearby CpG cite, cg05380919, explain 78% of the observed variability in the expression of *GSTT1*. It is also possible that methylation levels of the CpG site may provide a stronger contribution. Similarly, for *GSTM1*, the strongest SNP only explains 11% of the variation in its expression, while methylation levels of the CpG site, cg18938907, have a much stronger association with expression of this gene and may be responsible for up to 55% of the observed variation. The latter CpG site also falls within a CGI that spans the promoter and a portion of the gene's first intron.

A substantial portion of the overall phenotypic variance in the activity of the hepatic enzyme, PON1, between individuals currently remains unexplained. In addition to contributions made by a variety of non-genetic factors, numerous transcription factors (Fuhrman, 2012), and miRNA regulation (Liu et al., 2013b), various functional PON1 polymorphisms have been shown to influence PON1 activity and its serum levels (Deakin et al., 2003; Fuhrman, 2012). For example, the SNP, rs705379, has been shown to be associated with approximately a 50% mean reduction in serum PON1 protein levels, as well as a reduction in PON1 transcript levels (Brophy et al., 2001; Deakin et al., 2003). In our study, it was interesting to observe that this SNP was associated with increased methylation of nine proximal CpG sites and lower expression of PON1.

3.3.4 Tissue-specificity of eQTLs and eQTMs

Since methylation and expression data for the present cohort were available for three tissues (muscle, SAT, and VAT) that were also examined in the individuals from the Dutch sample cohort, we could assess the tissue-specificity of both the detected eQTL and eQTM effects. For liver eQTL, approximately 40–50% of the effects found in one tissue could also be significantly detected in another tissue (Ref III, Figure 4, Additional File 16A). Similar results have been reported in previous studies (Fu et al., 2012; Innocenti et al., 2011; Schadt et al., 2008). However, the eQTL effects (n = 32,863) that were only present in liver tissue were found to be related to genes strongly specific to liver function ($p = 5 \times 10^{-53}$), as well as metabolic and catabolic processes ($p < 5 \times 10^{-20}$).

As expected, we observed very strong tissue-specificity for the identified eQTMs. For example, only up to 4% of the eQTMs found in one tissue were also detectable with the same effect in another tissue (Ref III, Figure 4C, Additional File 16E). These results are in accordance with the fact that DNA methylation plays an important role in regulating tissue-specific gene expression. Thus, conclusions drawn from eQTL or eQTM data in one tissue cannot be extrapolated to other tissues, although the effects of SNPs on gene expression are more likely to be detectable in an alternative tissue. Ideally, detection of these effects in blood would represent a readily accessible source for testing.

CONCLUSIONS

The aim of the experimental work presented in this thesis was to describe methylation patterns and their effect on gene expression in different tissue types, both healthy and cancerous. To investigate DNA methylation patterns in early-stage lung cancer patients, a genome-wide DNA methylation study was performed with stage I NSCLC samples. Most of the identified genes represent novel markers for NSCLC. Furthermore, a gene expression and DNA methylation correlation analysis showed that approximately half of the CpG sites examined exhibited an expected negative correlation, which can be altered by the reduced resolution provided by the HumanMethylation27 microarray. GO was examined with IPA software and this revealed that differentially methylated genes were closely related to cancer progression. Furthermore, a survival analysis identified a number of CpG sites whose methylation levels differed according to patient survival. Accordingly, the latter sites represent CpGs that could potentially serve as prognostic markers.

Genome-wide DNA methylation patterns were also described for 17 human somatic tissues, and these patterns were largely in agreement with patterns previously observed in similar studies. In addition, our analysis confirmed a clear correlation between DNA methylation in promoter regions and gene expression. The DNA methylation patterns also clearly reflected tissue-specific functions as demonstrated with hierarchical clustering and GO analyses of hypomethylated tDMRs. Furthermore, the tDMR analysis revealed that a large number of methylated regions were within gene body regions, yet we were not able to show how these tDMRs mechanistically contribute to tissue-specific functions.

When genome-wide analyses of genomic and epigenetic variations and their association with gene expression were examined for fetal and adult liver tissues, a strong negative correlation between DNA methylation and gene expression was found in the promoter regions, while methylation in the gene bodies were associated with active gene expression. eQTL mapping was also performed and a large number of liver-specific genes were discovered, many of which had not previously been reported. The effect of eQTL and eQTMs on tissue-specificity were also assessed on another set of human somatic tissues. The results from our two published studies confirm that DNA methylation has a strong tissue-specific effect on gene expression. Meanwhile, SNPs had a much weaker influence on tissue-specific gene expression, and approximately half of the SNP effects found in one tissue could also be found in another tissue.

Based on the observation that inter-individual variability in ADME gene expression affects drug efficacy, toxicity, and susceptibility to environmental toxins, we determined to what extent SNPs and DNA methylation can jointly explain variations in ADME gene expression in liver. As expected, the combination of SNP genotype and CpG site methylation levels data explained more of the observed expression variations than the use of SNP or methylation levels alone.

SUMMARY IN ESTONIAN

Inimese tervete kudede ja mitteväikerakulise kopsuvähi võrdlevad ülegenoomsed DNA metülatsiooni uuringud

Metüleeritud tsütosiini peetakse „DNA viiendaks aluspaariks“ ning sellel on märkimisväärselt suur roll imetajate organismi arengus. Inimese genoomis on umbes 29 miljonit CpG dinukleotiidi, mis võivad olla kas metüleeritud või metüleerimata olekus. CpG dinukleotiidide metülatsioonitase on genoomi lõikes väga erinev. Kordusjärjestused on enamasti tugevalt metüleeritud, samas kui CpG rikkad promootorjärjestused hoitakse peamiselt mittemetüleerituna kõikides inimese kudedes. Kuna DNA metülatsioon on eluliselt vajalik organismi funktsioneerimiseks, tekitavad muutused epigenoomis erinevaid haiguslikke seisundeid, näiteks vähki. Samuti põhjustab erinev metüleerimise tase inimestevahelist varieeruvust ravimite metaboliseerimises, mis toob omakorda kaasa ebasoovitavaid ravimite kõrvaltoimeid. DNA metülatsioonimustreid on uuritud erinevate meetoditega juba aastakümneid, samas ei ole suudetud täielikult seletada kuidas epigenoom ning transkriptoom omavahel suhtlevad. Mitmed uuringud seovad promootorpiirkonnas asuvate CpG saarekeste metülatsiooni geeniekspressiooni vaigistamisega. Kuid samal ajal leitakse koespetsiifilisi erinevalt metüleeritud regioone aina enam just geenisisestelt aladelt ning isegi geenide vahelistest piirkondadest. Nende piirkondade ja geeniekspressiooni regulatsiooni vahel ei ole veel suudetud leida lõplikke seoseid.

Vähk, eriti kopsuvähk on tänapäeva maailmas suureks probleemiks. Paljude aastate jooksul on erinevad tööd toonud välja sadu biomarkereid, nii geneetilisi kui epigeneetilisi, mis võiksid aidata diagnoosida vähki varajases staadiumis või prognoosida patsiendi ravi tulemuslikkust ning üleüldist elulemust. Hoolimata rohketest uuringutest, on ainult väike hulk neist biomarkeritest edukalt valideeritud ning kasutusel igapäevases praktikas.

Käesolevas doktoritöös antakse kirjanduse põhjal ülevaade erinevatest DNA metülatsiooni tahkudest. Kirjeldatakse kuidas DNA metülatsiooni tase on dünaamilises muutumises organismi arengu jooksul, millised ensüümid vastutavad DNA metüleerimise eest ning kuidas geeniekspressioon ning DNA metülatsioon on omavahel seotud mitmesugustes inimkudedes. Antakse ülevaade muutunud geneetilistest ja epigeneetilistest teguritest erinevates vähivormides ning eraldi ka kopsuvähis. Välja on toodud ka mitmed tänapäevased meetodid DNA metülatsiooni uurimiseks.

Töö eksperimentaalse osa eesmärgiks oli kirjeldada DNA metülatsioonimustreid ning nende mõju geeniekspressioonile erinevates inimkudedes, nii tervetes kui ka tuumoris. Ülegenoomse DNA metülatsiooni mustreid kirjeldati 17-nes inimese somaatilises koes. Uuringute põhjal järeldasime, et promootorpiirkonna metülatsiooni ning geeniekspressiooni vahel on selge funktsionaalne seos. Samas ei suutnud me leida seoseid geenisisese metülatsiooni ning mRNA ekspressiooni tasemete vahel. DNA metülatsiooni mustrid peegeldasid efektiivselt kudede funktsioone, nii geeni ontoloogia analüüsi kui ka hierarhilise

klasterdamise põhjal. Koe-spetsiifiliste erinevalt metüleeritud regioonide (tDMRs) analüüs leidis suurel hulgal piirkondi, mis paiknesid geenisiseselt, kuid me ei suutnud leida seoseid, kuidas need tDMR-d põhjustavad koe-spetsiifilist transkriptsiooni.

Praktilise poole teises osas kirjeldatakse inimese loote ning täiskasvanu maksakoe ülegenoomseid geneetilisi ning epigeneetilisi muutusi ning nende seost geenisekspressiooniga. DNA metülatsiooni ning geenisekspressiooni vahelise korrelatsiooni analüüsi tulemusel leiti, et tugevaim negatiivne seos oli geeni promootorpiirkonnas, samal ajal seostati geenisisest metülatsiooni aktiivse transkriptsiooniga. Geenide ekspressiooni mõjutavate geneetiliste variantide (eQTL) kaardistamise tulemusel leidsime suurel hulgal maksa-spetsiifilisi geene, millest mitmeid pole varasemalt kirjeldatud. Geenide, mis on seotud absorptsiooni, jaotuse, metabolismi ning väljutusega (ADME) ekspressioonis on leitud suur indiviidide vaheline varieeruvus, mis omakorda mõjutab ravimvastust, toksilisust ning vastuvõtlikkust keskkonna toksiinidele. Oma uuringus püüdsime määrata, kui suurel hulgal SNP-d ning DNA metülatsioon on võimelised üheskoos kirjeldama ADME geenide ekspressiooni varieeruvust täiskasvanute maksakoes. Ootuspäraselt leidsime, et SNP genotüübi ning CpG metülatsioonitaseme kombineerimine võimaldab kirjeldada ekspressiooni variatsiooni paremini, kui kasutades ainult SNP-d või CpG metülatsiooni taset eraldiseisvalt. Me hindasime ka eQTL-ide ning eQTM-ide (geenide ekspressiooni mõjutavad metülatsiooni tasemed) mõju koespetsiifilisusele, kasutades lisaks maksale veel kolme koetüüpi (lihas, nahaalune- ja vistseraalne rasvkude). Need tulemused kinnitasid samuti metülatsiooni tugevat koespetsiifilist mõju geenisekspressioonile. Samas SNP-del oli palju väiksem mõju koespetsiifilisele geenisekspressioonile, umbes pooled efektid olid leitavad ka kõikides teistes kudedes.

Töö kolmas osa keskendus DNA metülatsiooni kirjeldamisele varajase staadiumi mitteväikerakulise kopsuvähi (NSCLC) patsientidel. Uuringu tulemusel leidsime suurel hulgal erinevalt metüleeritud CpG saite. Mitmed neist on tuntud vähiseoselised biomarkerid, kuid enamik kirjeldatud geenidest on uued kopsuvähi markerid. Geenisekspressiooni ja DNA metülatsiooni korrelatsioonianalüüsi tulemusel leidsime oodatud negatiivseid seoseid pooltel CpG saitidel. Geenide ontoloogiaanalüüs, erinevalt metüleeritud geenidega näitas, et nende geenide funktsioonid olid tugevalt seotud vähi progresseerumisega. Elulemusanalüüsi tulemusel leidsime hulga CpG saite, mille metülatsiooni tasemed erinesid erinevates elulemusgruppides. Mõned neist geenidest võivad olla head prognostiliste markerite kandidaadid.

REFERENCES

Al-Hajj, M., et al., 2003. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 100, 3983–8.

Alison, M. R., et al., 2012. Cancer stem cells: in the line of fire. Cancer Treat Rev. 38, 589–98.

Ansari, J., et al., 2016. Epigenetics in non-small cell lung cancer: from basics to therapeutics. Transl Lung Cancer Res. 5, 155–71.

Anttila, S., et al., 2003. Methylation of cytochrome P4501A1 promoter in the lung is associated with tobacco smoking. Cancer Res. 63, 8623–8.

Araki, J., et al., 2005. Polymorphism of UDP-glucuronosyltransferase 1A7 gene: a possible new risk factor for lung cancer. Eur J Cancer. 41, 2360–5.

Auton, A., et al., 2015. A global reference for human genetic variation. Nature. 526, 68–74.

Balgkouranidou, I., et al., 2013. Lung cancer epigenetics: emerging biomarkers. Biomark Med. 7, 49–58.

Bartling, B., et al., 2005. Down-regulation of the receptor for advanced glycation end-products (RAGE) supports non-small cell lung carcinoma. Carcinogenesis. 26, 293–301.

Belinsky, S. A., et al., 1998. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A. 95, 11891–6.

Bell, J. T., Spector, T. D., 2012. DNA methylation studies using twins: what are they telling us? Genome Biol. 13, 172.

Bibikova, M., et al., 2011. High density DNA methylation array with single CpG site resolution. Genomics. 98, 288–95.

Bind, M. A., et al., 2012. Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigene-environment interactions in an elderly cohort. Epidemiology. 23, 332–40.

Boiko, A. D., et al., 2010. Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. Nature. 466, 133–7.

Bonnet, D., Dick, J. E., 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 3, 730–7.

Borgel, J., et al., 2010. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 42, 1093–100.

Brandeis, M., et al., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature. 371, 435–8.

Brock, M. V., et al., 2008. DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med. 358, 1118–28.

Brophy, V. H., et al., 2001. Polymorphisms in the human paraoxonase (PON1) promoter. Pharmacogenetics. 11, 77–84.

Busche, S., et al., 2015. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. Genome Biol. 16, 290.

Calvanese, V., et al., 2012. A promoter DNA demethylation landscape of human hematopoietic differentiation. Nucleic Acids Res. 40, 116–31.

Carter, H., et al., 2009. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res. 69, 6660–7.

Challen, G. A., et al., 2012. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet. 44, 23–31.
Chambers, J. C., et al., 2011. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet. 43, 1131–8.
Chang, C. W., et al., 2011. Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. PLoS One. 6, e22859.
Chen, Y., et al., 2014. Diversity of two forms of DNA methylation in the brain. Front Genet. 5, 46.
Chu, B. F., et al., 2013. Phase I study of 5-aza-2'-deoxycytidine in combination with valproic acid in non-small-cell lung cancer. Cancer Chemother Pharmacol. 71, 115–21.
Cicek, M. S., et al., 2013. Epigenome-wide ovarian cancer analysis identifies a methylation profile differentiating clear-cell histology with epigenetic silencing of the HERG K+ channel. Hum Mol Genet. 22, 3038–47.
Clark, C., et al., 2012. A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling. PLoS One. 7, e50233.
Collins, A. T., et al., 2005. Prospective identification of tumorigenic prostate cancer stem cells. Cancer Res. 65, 10946–51.
Comb, M., Goodman, H. M., 1990. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. Nucleic Acids Res. 18, 3975–82.
Cvejic, A., et al., 2013. SMIM1 underlies the Vel blood group and influences red blood cell traits. Nat Genet. 45, 542–5.
Dannenberg, L. O., Edenberg, H. J., 2006. Epigenetics of gene expression in human hepatoma cells: expression profiling the response to inhibition of DNA methylation and histone deacetylation. BMC Genomics. 7, 181.
Davies, M. N., et al., 2012. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. 13, R43.
Davis, T. L., et al., 2000. The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. Hum Mol Genet. 9, 2885–94.
De Palma, M., et al., 2005. Tie2 identifies a hematopoietic lineage of proangiogenic monocytes required for tumor vessel formation and a mesenchymal population of pericyte progenitors. Cancer Cell. 8, 211–26.
Deakin, S., et al., 2003. Paraoxonase-1 promoter haplotypes and serum paraoxonase: a predominant role for polymorphic position - 107, implicating the Sp1 transcription factor. Biochem J. 372, 643–9.
Deaton, A. M., et al., 2011. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res. 21, 1074–86.
Dennis, G., et al., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, P3.
Devarakonda, S., et al., 2015. Genomic alterations in lung adenocarcinoma. Lancet Oncol. 16, e342–51.
Dho, S. H., et al., 2016. GPR171 expression enhances proliferation and metastasis of lung cancer cells. Oncotarget. 7, 7856–65.
Diaz-Lagares, A., et al., 2016. A Novel Epigenetic Signature for Early Diagnosis in Lung Cancer. Clin Cancer Res.

Duncan, B. K., Miller, J. H., 1980. Mutagenic deamination of cytosine residues in DNA. Nature. 287, 560–1.
Eckhardt, F., et al., 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 38, 1378–85.
Ehrich, M., et al., 2006. Cytosine methylation profiles as a molecular marker in non-small cell lung cancer. Cancer research. 66, 1–1.
Epsztejn-Litman, S., et al., 2008. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat Struct Mol Biol. 15, 1176–83.
Eramo, A., et al., 2008. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death Differ. 15, 504–14.
Esteller, M., 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 8, 286–98.
Fang, D., et al., 2005. A tumorigenic subpopulation with stem cell properties in melanomas. Cancer Res. 65, 9328–37.
Fatemi, M., et al., 2001. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol. 309, 1189–99.
Feldman, N., et al., 2006. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat Cell Biol. 8, 188–94.
Ficz, G., et al., 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 473, 398–402.
Frommer, M., et al., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 89, 1827–31.
Fu, J., et al., 2012. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. 8, e1002431.
Fuhrman, B., 2012. Regulation of hepatic paraoxonase-1 expression. J Lipids. 2012, 684010.
Gazzoli, I., et al., 2002. A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. Cancer Res. 62, 3925–8.
Gendrel, A. V., et al., 2012. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev Cell. 23, 265–79.
Ghotbi, R., et al., 2009. Allele-specific expression and gene methylation in the control of CYP1A2 mRNA level in human livers. Pharmacogenomics J. 9, 208–17.
Ginestier, C., et al., 2007. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell. 1, 555–67.
Goldstraw, P., et al., 2007. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Oncol. 2, 706–14.
Gopalakrishnan, S., et al., 2009. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum Mol Genet. 18, 3178–93.
Gowher, H., Jeltsch, A., 2001. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive

manner and also methylates non-CpG [correction of non-CpA] sites. J Mol Biol. 309, 1201–8.
Goyal, R., et al., 2006. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Res. 34, 1182–8.
Greenawalt, D. M., et al., 2011. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. 21, 1008–16.
Guo, F., et al., 2014a. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell. 15, 447–58.
Guo, H., et al., 2014b. The DNA methylation landscape of human early embryos. Nature. 511, 606–10.
Guo, J. U., et al., 2011. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell. 145, 423–34.
Gutierrez-Arcelus, M., et al., 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife. 2, e00523.
Győrffy, B., et al., 2013. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One. 8, e82241.
Hackett, J. A., et al., 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science. 339, 448–52.
Hackett, J. A., Surani, M. A., 2013. DNA methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond B Biol Sci. 368, 20110328.
Hajkova, P., et al., 2002. Epigenetic reprogramming in mouse primordial germ cells. Mech Dev. 117, 15–23.
Hamburger, A. W., Salmon, S. E., 1977. Primary bioassay of human tumor stem cells. Science. 197, 461–3.
Harikrishnan, K. N., et al., 2005. Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet. 37, 254–64.
Harris, R. A., et al., 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 28, 1097–105.
Hata, K., et al., 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 129, 1983–93.
He, Y. F., et al., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 333, 1303–7.
Hermann, P. C., et al., 2007. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. Cell Stem Cell. 1, 313–23.
Hitchins, M. P., et al., 2011. Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. Cancer Cell. 20, 200–13.
Hitchins, M. P., et al., 2007. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med. 356, 697–705.
Hiura, H., et al., 2006. Oocyte growth-dependent progression of maternal imprinting in mice. Genes Cells. 11, 353–61.
Holliday, R., 1987. The inheritance of epigenetic defects. Science. 238, 163–70.
Howlett, S. K., Reik, W., 1991. Methylation levels of maternal and paternal genomes during preimplantation development. Development. 113, 119–27.

Huang, X., Jan, L. Y., 2014. Targeting potassium channels in cancer. J Cell Biol. 206, 151–62.
Huang, Y., et al., 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One. 5, e8888.
Huse, S. M., et al., 2015. Patterns of gene expression and DNA methylation in human fetal and adult liver. BMC Genomics. 16, 981.
Iliopoulos, D., et al., 2011. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. Proc Natl Acad Sci U S A. 108, 1397–402.
Ingelman-Sundberg, M., et al., 2007. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. Pharmacol Ther. 116, 496–526.
Innocenti, F., et al., 2011. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 7, e1002078.
Iqbal, K., et al., 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A. 108, 3642–7.
Irizarry, R. A., et al., 2008. Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res. 18, 780–90.
Islam, F., et al., 2015a. Translational potential of cancer stem cells: A review of the detection of cancer stem cells and their roles in cancer recurrence and cancer treatment. Exp Cell Res. 335, 135–47.
Islam, F., et al., 2015b. Cancer stem cell: fundamental experimental pathological concepts and updates. Exp Mol Pathol. 98, 184–91.
Ito, S., et al., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 466, 1129–33.
Ito, S., et al., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 333, 1300–3.
Ivanov, M., et al., 2013. Ontogeny, distribution and potential roles of 5-hydroxymethylcytosine in human liver function. Genome Biol. 14, R83.
Ivanov, M., et al., 2016. Single base resolution analysis of 5-hydroxymethylcytosine in 188 human genes: implications for hepatic gene expression. Nucleic Acids Res.
Jin, S. G., et al., 2011. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res. 71, 7360–5.
Jjingo, D., et al., 2012. On the presence and role of human gene-body DNA methylation. Oncotarget. 3, 462–74.
Jochheim-Richter, A., et al., 2006. Gene expression analysis identifies novel genes participating in early murine liver development and adult liver regeneration. Differentiation. 74, 167–73.
Jones, P., Baylin, S., 2002. The fundamental role of epigenetic events in cancer. Nat Rev Genet. 3, 415–28.
Jones, P. A., 1999. The DNA methylation paradox. Trends Genet. 15, 34–7.
Jones, P. A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 13, 484–92.
Jones, P. A., Baylin, S. B., 2007. The epigenomics of cancer. Cell. 128, 683–92.
Juergens, R. A., et al., 2011. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov. 1, 598–607.

Kacevska, M., et al., 2012a. Epigenetic-dependent regulation of drug transport and metabolism: an update. Pharmacogenomics. 13, 1373–85.
Kacevska, M., et al., 2012b. DNA methylation dynamics in the hepatic CYP3A4 gene promoter. Biochimie. 94, 2338–44.
Kandoth, C., et al., 2013. Mutational landscape and significance across 12 major cancer types. Nature. 502, 333–9.
Kaneda, M., et al., 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 429, 900–3.
Kato, Y., et al., 2007. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet. 16, 2272–80.
Kaut, O., et al., 2012. Genome-scale methylation analysis of Parkinson's disease patients' brains reveals DNA hypomethylation and increased mRNA expression of cytochrome P450 2E1. Neurogenetics. 13, 87–91.
Kelly, P. N., et al., 2007. Tumor growth need not be driven by rare cancer stem cells. Science. 317, 337.
Khuder, S. A., Mutgi, A. B., 2001. Effect of smoking cessation on major histologic types of lung cancer. Chest. 120, 1577–83.
Khulan, B., et al., 2006. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res. 16, 1046–55.
Knudson, A. G., 1971. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 68, 820–3.
Ko, M., et al., 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 468, 839–43.
Kolde, R., et al., 2016. seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics.
Kozlenkov, A., et al., 2014. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. Nucleic Acids Res. 42, 109–27.
Kraus, T. F., et al., 2012. Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors. Int J Cancer. 131, 1577–90.
Kriaucionis, S., Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 324, 929–30.
Kroeze, L. I., et al., 2014. Characterization of acute myeloid leukemia based on levels of global hydroxymethylation. Blood. 124, 1110–8.
Kurdyukov, S., Bullock, M., 2016. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel). 5.
Lane, N., et al., 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis. 35, 88–93.
Lee, J. S., et al., 2012. Transcriptional ontogeny of the developing liver. BMC Genomics. 13, 33.
Leonard, S. M., et al., 2012. Oncogenic human papillomavirus imposes an instructive pattern of DNA methylation changes which parallel the natural history of cervical HPV infection in young women. Carcinogenesis. 33, 1286–93.
Li, C., et al., 2007. Identification of pancreatic cancer stem cells. Cancer Res. 67, 1030–7.
Li, E., et al., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 69, 915–26.

Li, T., et al., 2009. Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. Genomics. 93, 235–42.
Li, W., Liu, M., 2011. Distribution of 5-hydroxymethylcytosine in different human tissues. J Nucleic Acids. 2011, 870726.
Li, X., et al., 2008. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst. 100, 672–9.
Lian, C. G., et al., 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell. 150, 1135–46.
Ligtenberg, M. J., et al., 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet. 41, 112–7.
Lin, R. K., et al., 2010. The tobacco-specific carcinogen NNK induces DNA methyltransferase 1 accumulation and tumor suppressor gene hypermethylation in mice and lung cancer patients. J Clin Invest. 120, 521–32.
Lin, S. H., et al., 2014. Genes suppressed by DNA methylation in non-small cell lung cancer reveal the epigenetics of epithelial-mesenchymal transition. BMC Genomics. 15, 1079.
Liu, B., et al., 1996. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. Nat Med. 2, 169–74.
Liu, C., et al., 2013a. Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. PLoS One. 8, e62828.
Liu, L., et al., 2003. Aging, cancer and nutrition: the DNA methylation connection. Mech Ageing Dev. 124, 989–98.
Liu, M. E., et al., 2013b. A functional polymorphism of PON1 interferes with microRNA binding to increase the risk of ischemic stroke and carotid atherosclerosis. Atherosclerosis. 228, 161–7.
Liu, Y., et al., 2013c. Methylomics of gene expression in human monocytes. Hum Mol Genet. 22, 5065–74.
Liu, Y., et al., 2006. Aberrant promoter methylation of p16 and MGMT genes in lung tumors from smoking and never-smoking lung cancer patients. Neoplasia. 8, 46–51.
Liu, Y. P., et al., 2013d. Cisplatin selects for multidrug-resistant CD133+ cells in lung adenocarcinoma by activating Notch signaling. Cancer Res. 73, 406–16.
Lowe, R., et al., 2015. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. Epigenetics. 10, 274–81.
Lujambio, A., et al., 2008. A microRNA DNA methylation signature for human cancer metastasis. Proc Natl Acad Sci U S A. 105, 13556–61.
Madrigano, J., et al., 2012. Aging and epigenetics: Longitudinal changes in gene-specific DNA methylation. Epigenetics. 7.
Magee, J. A., et al., 2012. Cancer stem cells: impact, heterogeneity, and uncertainty. Cancer Cell. 21, 283–96.
Maley, C. C., et al., 2004. Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. Cancer Res. 64, 3414–27.
Mani, S. A., et al., 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 133, 704–15.
Marinković, N., et al., 2013. Polymorphisms of genes involved in polycyclic aromatic hydrocarbons' biotransformation and atherosclerosis. Biochem Med (Zagreb). 23, 255–65.

Marotta, L. L., et al., 2011. The JAK2/STAT3 signaling pathway is required for growth of $CD44^{+}CD24^{-}$ stem cell-like breast cancer cells in human tumors. J Clin Invest. 121, 2723–35.
Maunakea, A. K., et al., 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. 23, 1256–69.
Maunakea, A. K., et al., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 466, 253–7.
Mayer, W., et al., 2000. Demethylation of the zygotic paternal genome. Nature. 403, 501–2.
McAuliffe, S. M., et al., 2012. Targeting Notch, a key pathway for ovarian cancer stem cells, sensitizes tumors to platinum therapy. Proc Natl Acad Sci U S A. 109, E2939–48.
McLean, C. Y., et al., 2010. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 28, 495–501.
Meissner, A., et al., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33, 5868–77.
Meissner, A., et al., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 454, 766–70.
Mellén, M., et al., 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell. 151, 1417–30.
Mendizabal, I., Yi, S. V., 2016. Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. Hum Mol Genet. 25, 69–82.
Miki, Y., et al., 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 266, 66–71.
Montano, C., et al., 2016. Association of DNA Methylation Differences With Schizophrenia in an Epigenome-Wide Association Study. JAMA Psychiatry.
Moran, S., et al., 2016. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 8, 389–99.
Morris, M. R., et al., 2011. Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene. 30, 1390–401.
Moscovitz, J. E., Aleksunes, L. M., 2013. Establishment of metabolism and transport pathways in the rodent and human fetal liver. Int J Mol Sci. 14, 23801–27.
Muggerud, A. A., et al., 2010. Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. Breast Cancer Res. 12, R3.
Nagae, G., et al., 2011. Tissue-specific demethylation in CpG-poor promoters during cellular differentiation. Hum Mol Genet. 20, 2710–21.
Nakamura, T., et al., 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol. 9, 64–71.
Nakamura, T., et al., 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature. 486, 415-9.
Nervi, C., et al., 2015. Epigenetic treatment of solid tumours: a review of clinical trials. Clin Epigenetics. 7, 127.
Nik-Zainal, S., et al., 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 534, 47–54.

Notta, F., et al., 2011. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. Nature. 469, 362–7.
Nowell, P. C., 1976. The clonal evolution of tumor cell populations. Science. 194, 23–8.
O'Brien, C. A., et al., 2007. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature. 445, 106–10.
O'Connor, M. L., et al., 2014. Cancer stem cells: A contentious hypothesis now moving forward. Cancer Lett. 344, 180–7.
Oda, M., et al., 2009. High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. Nucleic Acids Res. 37, 3829–39.
Okano, M., et al., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 99, 247–57.
Okano, M., et al., 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 19, 219–20.
Palmisano, W. A., et al., 2000. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res. 60, 5954–8.
Pao, W., Girard, N., 2011. New driver mutations in non-small-cell lung cancer. Lancet Oncol. 12, 175–80.
Peltomäki, P., 2012. Mutations and epimutations in the origin of cancer. Exp Cell Res. 318, 299–310.
Pfeifer, G. P., 2000. p53 mutational spectra and the role of methylated CpG sequences. Mutat Res. 450, 155–66.
Pirmohamed, M., 2014. Personalized pharmacogenomics: predicting efficacy and adverse drug reactions. Annu Rev Genomics Hum Genet. 15, 349–70.
Pleasance, E. D., et al., 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 463, 191–6.
Puglisi, M. A., et al., 2013. Colon cancer stem cells: controversies and perspectives. World J Gastroenterol. 19, 2997–3006.
Rakyan, V. K., et al., 2008. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res. 18, 1518–29.
Ramos, D. L., et al., 2011. Genotoxic effects of doxorubicin in cultured human lymphocytes with different glutathione S-transferase genotypes. Mutat Res. 724, 28–34.
Rauch, T., et al., 2008. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proc Natl Acad Sci U S A. 105, 252–7.
Rauch, T. A., Pfeifer, G. P., 2010. DNA methylation profiling using the methylated-CpG island recovery assay (MIRA). Methods. 52, 213–7.
Rodriguez, J., et al., 2006. Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. Cancer Res. 66, 8462–9468.
Sandoval, J., et al., 2013. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J Clin Oncol. 31, 4140–7.
Sasaki, H., Matsui, Y., 2008. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet. 9, 129–40.
Schadt, E. E., et al., 2008. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 6, e107.
Schröder, A., et al., 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.

Seisenberger, S., et al., 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 48, 849–62.
Serre, D., et al., 2010. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. 38, 391–9.
Shigdar, S., et al., 2011. RNA aptamer against a cancer stem cell marker epithelial cell adhesion molecule. Cancer Sci. 102, 991–8.
Shigdar, S., et al., 2013. RNA aptamers targeting cancer stem cell marker CD133. Cancer Lett. 330, 84–95.
Shuang, Z. Y., et al., 2014. Transforming growth factor-β1-induced epithelial-mesenchymal transition generates ALDH-positive cells with stem cell properties in cholangiocarcinoma. Cancer Lett. 354, 320–8.
Silviera, M. L., et al., 2012. Epigenetic differences in normal colon mucosa of cancer patients suggest altered dietary metabolic pathways. Cancer Prev Res (Phila). 5, 374–84.
Sim, S. C., et al., 2013. Pharmacogenomics of drug-metabolizing enzymes: a recent up-date on clinical implications and endogenous effects. Pharmacogenomics J. 13, 1–11.
Singh, S. K., et al., 2003. Identification of a cancer stem cell in human brain tumors. Cancer Res. 63, 5821–8.
Sintupisut, N., et al., 2013. An integrative characterization of recurrent molecular aberrations in glioblastoma genomes. Nucleic Acids Res. 41, 8803–21.
Smalley, M., Ashworth, A., 2003. Stem cells and breast cancer: A field in transit. Nat Rev Cancer. 3, 832–44.
Song, F., et al., 2009. Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development. Genomics. 93, 130–9.
Song, F., et al., 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A. 102, 3336–41.
Stevens, M., et al., 2013. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. Genome Res. 23, 1541–53.
Strassburg, C. P., et al., 1999. Regulation and function of family 1 and family 2 UDP-glucuronosyltransferase genes (UGT1A, UGT2B) in human oesophagus. Biochem J. 338 (Pt 2), 489–98.
Stratton, M. R., et al., 2009. The cancer genome. Nature. 458, 719–24.
Stroud, H., et al., 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 12, R54.
Suetake, I., et al., 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem. 279, 27816–23.
Sun, B., et al., 2016. DNA methylation perspectives in the pathogenesis of autoimmune diseases. Clin Immunol. 164, 21–7.
Sun, Z., et al., 2015. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. Epigenomics. 7, 813–28.
Suzuki, M., Greally, J. M., 2010. DNA methylation profiling using HpaII tiny fragment enrichment by ligation-mediated PCR (HELP). Methods. 52, 218–22.
Tahiliani, M., et al., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 324, 930–5.
Tekpli, X., et al., 2012. DNA methylation of the CYP1A1 enhancer is associated with smoking-induced genetic alterations in human lung. Int J Cancer. 131, 1509–16.

Tobi, E. W., et al., 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet. 18, 4046–53.
Todaro, M., et al., 2007. Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. Cell Stem Cell. 1, 389–402.
Torre, L. A., et al., 2015. Global cancer statistics, 2012. CA Cancer J Clin. 65, 87–108.
Toyooka, S., et al., 2003. Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. Int J Cancer. 103, 153–60.
Tsai, K. W., et al., 2015. Reduction of global 5-hydroxymethylcytosine is a poor prognostic factor in breast cancer patients, especially for an ER/PR-negative subtype. Breast Cancer Res Treat. 153, 219–34.
Tsumagari, K., et al., 2013. Early de novo DNA methylation and prolonged de-methylation in the muscle lineage. Epigenetics. 8, 317–32.
Tulsyan, S., et al., 2013. Pharmacogenetic influence of GST polymorphisms on anthracycline-based chemotherapy responses and toxicity in breast cancer patients: a multi-analytical approach. Mol Diagn Ther. 17, 371–9.
Ulasov, I. V., et al., 2011. Inhibition of Sonic hedgehog and Notch pathways enhances sensitivity of CD133(+) glioma stem cells to temozolomide therapy. Mol Med. 17, 103–12.
Uringa, E. J., et al., 2011. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. Nucleic Acids Res. 39, 1647–55.
Valinluck, V., Sowers, L. C., 2007. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 67, 946–50.
Varley, K. E., et al., 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23, 555–67.
Visvader, J. E., Lindeman, G. J., 2012. Cancer stem cells: current status and evolving complexities. Cell Stem Cell. 10, 717–28.
Walker, D. L., et al., 2015. DNA methylation profiling: comparison of genome-wide sequencing methods and the Infinium Human Methylation 450 Bead Chip. Epigenomics. 7, 1287–302.
Wang, L., et al., 2012. Systematic assessment of reduced representation bisulfite sequencing to human blood samples: A promising method for large-sample-scale epigenomic studies. J Biotechnol. 157, 1–6.
Watanabe, D., et al., 2002. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech Dev. 118, 187–90.
Waterland, R. A., et al., 2009. Epigenomic profiling indicates a role for DNA methylation in early postnatal liver development. Hum Mol Genet. 18, 3026–38.
Weber, M., et al., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 39, 457–66.
Wolf, S. F., et al., 1984. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci U S A. 81, 2806–10.
Wolfs, M. G., 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.
Wooster, R., et al., 1995. Identification of the breast cancer susceptibility gene BRCA2. Nature. 378, 789–92.

Wrangle, J., et al., 2014. Functional identification of cancer-specific methylation of CDO1, HOXA9, and TAC1 for the diagnosis of lung cancer. Clin Cancer Res. 20, 1856–64.
Wrangle, J., et al., 2013. Alterations of immune response of Non-Small Cell Lung Cancer with Azacytidine. Oncotarget. 4, 2067–79.
Wu, H., et al., 2010. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science. 329, 444–8.
Wu, H., Zhang, Y., 2011. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. 25, 2436–52.
Wu, S. C., Zhang, Y., 2010. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol. 11, 607–20.
Yan, L., et al., Specific inhibition of DNMT1 by antisense oligonucleotides induces re-expression of estrogen receptor-alpha (ER) in ER-negative human breast cancer cell lines. Cancer Biol Ther. 2, 552–6.
Yanagawa, N., et al., 2003. Promoter hypermethylation of tumor suppressor and tumor-related genes in non-small cell lung cancers. Cancer Sci. 94, 589–92.
Yang, X., et al., 2015. Systematic DNA methylation analysis of multiple cell lines reveals common and specific patterns within and across tissues of origin. Hum Mol Genet. 24, 4374–84.
Yang, X., et al., 2010. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. Genome Res. 20, 1020–36.
Zhao, C., et al., 2009. Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. Nature. 458, 776–9.
Zhao, C., et al., 2013. SCUBE3 overexpression predicts poor prognosis in non-small cell lung cancer. Biosci Trends. 7, 264–9.
Zhong, S., et al., 2006. Relationship of glutathione S-transferase genotypes with side-effects of pulsed cyclophosphamide therapy in patients with systemic lupus erythematosus. Br J Clin Pharmacol. 62, 457–72.
Zilbauer, M., et al., 2013. Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypo-methylated regions. Blood. 122, e52–60.
Ziller, M. J., et al., 2013. Charting a dynamic DNA methylation landscape of the human genome. Nature. 500, 477–81.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Neeme Tõnisson, for his continuous support of my studies over these years and for his patience, motivation, and knowledge. And not less, I would like to thank my co-supervisor, Lili Milani, for the inspiration and motivation she provided me during the final years of my Ph.D. studies. She also showed me that family life with small kids and science can be successfully combined. I also would like to thank Prof. Andres Metspalu for providing me with a good scientific atmosphere during all of my studies.

My sincere thanks also goes to all my colleagues in the Biotechnology Department for keeping up a good scientific and social vibe. Especially I would like to thank Viljo Soo for excellent work in the lab providing me with methylation and gene expression data. Also Heidi Saulep for patiently answering my every (stupid) question that I had, particularly when I was beginning my wet lab work. Last, but not least, I would like to thank our secretary, Krista Liiv. Krista, without you, we all would be lost in the administrative matters.

I thank Tõnu Vooder and Andres Salumets for making it possible for me to study the methylation of such interesting samples. I thank Raivo Kolde and Kaspar Märtens for all their help with the data analysis, as well as Vijayachitra Modhukur and Balaji Rajashekar. I thank all the other co-authors of the papers we published for their contributions and help. I thank my fellow labmates for all the scientific discussions and fun we had after-hours and at conferences. Especially Kats, Anu, Sven, and Ott – without you, it would not have been the same!

Last, but not least, I would like to thank my family: my parents, my husband, and our little girl for supporting and motivating me spiritually through all these years. I could not have done it without you!

PUBLICATIONS

CURRICULUM VITAE

Name: Kaie Lokk
Date of birth: October 21, 1984, Tartu, Estonia
Contact: University of Tartu, The Institute of Molecular and Cell Biology, Department of Biotechnology, Riia 23, 51010, Tartu
E-mail: kaie.lokk@ut.ee

Education:

2008–present	University of Tartu, PhD student (Gene technology)
2006–2008	University of Tartu, Master's degree (Gene technology)
2003–2006	University of Tartu, Bachelor's degree (Gene technology)
2000–2003	Tartu Kommertsgümnaasium

Work Experience:

2016–present	University of Tartu, Estonian Genome Center, Sequencing manager
2010–2014	Tartu University Hospital, United Laboratories of Tartu University Hospital, Department of Genetics, Specialist

Publications:

Kolde, R; Märtens, K; **Lokk, K**; Laur, S; Vilo, J (2016). seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics, Sep 1;32(17):2604–10. doi: 10.1093/bioinformatics/btw304.

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, Oct 4;15:860. doi: 10.1186/1471-2164-15-860.

Lokk, K; Modhukur, V; Rajashekar, B; Märtens, K; Mägi, R; Kolde, R; Koltšina, M; Nilsson, TK; Vilo, J; Salumets, A; Tõnisson, N (2014). DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biology, Apr 1;15(4):r54. doi: 10.1186/gb-2014-15-4-r54.

Lokk, K; Vooder, T; Kolde, R; Välk, K; Võsa, U; Roosipuu, R; Milani, L; Fischer, K; Koltšina, M; Urgard, E; Annilo, T; Metspalu, A; Tõnisson, N (2012). Methylation markers of early-stage non-small cell lung cancer. PLoS One, 7(6):e39813. doi: 10.1371/journal.pone.0039813.

Supervised dissertations:

Co-supervision of the master's thesis of Marina Koltšina "Genomic copy number alterations in non-small cell lung cancer" ("Genoomse koopiaarvu muutused mitteväikerakulise kopsuvähi korral") (2010).

ELULOOKIRJELDUS

Nimi: Kaie Lokk
Sünniaeg: 21.oktoober, 1984, Tartu
Aadress: Tartu Ülikool, Molekulaar- ja Rakubioloogia Instituut, Biotehnoloogia õppetool, Riia 23, 51010, Tartu
E-post: kaie.lokk@ut.ee

Haridus:

2008– …	Tartu Ülikool, Doktoriõpe (Geenitehnoloogia)
2006–2008	Tartu Ülikool, Magistriõpe (Geenitehnoloogia)
2003–2006	Tartu Ülikool, Bakalaureuseõpe (Geenitehnoloogia)
2000–2003	Tartu Kommertsgümnaasium

Töökogemus:

2016–…	Tartu Ülikool, Tartu Ülikooli Geenivaramu, Sekveneerimislabori projektijuht
2010–2014	Tartu Ülikooli Kliinikum, Ühendlabor, Geneetikakeskus, spetsialist

Teaduspublikatsioonid:

Kolde, R; Märtens, K; **Lokk, K**; Laur, S; Vilo, J (2016). seqlm: an MDL based method for identifying differentially methylated regions in high density methylation array data. Bioinformatics, Sep 1;32(17):2604–10. doi: 10.1093/bioinformatics/btw304.

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, Oct 4;15:860. doi: 10.1186/1471-2164-15-860.

Lokk, K; Modhukur, V; Rajashekar, B; Märtens, K; Mägi, R; Kolde, R; Koltšina, M; Nilsson, TK; Vilo, J; Salumets, A; Tõnisson, N (2014). DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biology, Apr 1;15(4):r54. doi: 10.1186/gb-2014-15-4-r54.

Lokk, K; Vooder, T; Kolde, R; Välk, K; Võsa, U; Roosipuu, R; Milani, L; Fischer, K; Koltšina, M; Urgard, E; Annilo, T; Metspalu, A; Tõnisson, N (2012). Methylation markers of early-stage non-small cell lung cancer. PLoS One, 7(6):e39813. doi: 10.1371/journal.pone.0039813.

Juhendatud väitekirjad:

Kaasjuhendamine Marina Koltšina magistritööle "Genoomse koopiaarvu muutused mitteväikerakulise kopsuvähi korral" ("Genomic copy number alterations in non-small cell lung cancer") (2010).

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.