TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Experimentelle Genetik

Helmholtz Zentrum München, Neuherberg

Genome-wide DNA methylation analyses of human blood lipid levels and functional analyses of lipid-associated CpG sites

Liliane Chantal Isabelle Pfeiffer

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr.rer.nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. J.J. Hauner
Prüfer der Dissertation: 1. apl. Prof. Dr. J. Adamski
                           2. Univ.-Prof. Dr. H. Witt

Die Dissertation wurde am 24.08.2016 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 19.12.2016 angenommen.

TABLE OF CONTENTS

SUMMARY

Blood lipid levels, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides and total cholesterol levels, are considered heritable, modifiable risk factors for cardiometabolic diseases. They can be influenced by drug intake or lifestyle factors such as physical activity and diet. Several studies have also revealed a genetic background of disturbed blood lipid levels. In addition, studies in non-population-based cohorts have shown that epigenetic mechanisms may also be involved in inter-individual lipid level variability. However, no study yet has examined the association between lipid levels and genome-wide DNA methylation in a population-based cohort.

The objectives of this thesis were therefore: 1) to perform an epigenome-wide association study (EWAS) to systematically investigate the association between genome-wide DNA methylation in whole blood and main blood lipid levels (HDL-C, LDL-C, triglycerides and total cholesterol) of the population-based cohort KORA F4; 2) to conduct an EWAS in the same cohort to examine the association between genome-wide DNA methylation and serum levels of lipoprotein subfractions as well as other metabolites measured by ^{1}H NMR spectroscopy; 3) to investigate the effect of the methylation status of selected lipid-associated CpG sites on gene regulation by functional analyses.

The EWAS of main blood lipids discovered 12 associations between the methylation status of a given cytosine-phospho-guanosine dinucleotide (CpG) site and the level of a given lipid, with p-values ranging from 1.21E-27 to 9.66E-08. A meta-analysis of the replication results of two independent cohorts confirmed nine of the 12 associations. The EWAS of NMR-measured metabolites revealed 237 significant associations between the methylation status of a given CpG site and the level of a given metabolite, with p-values ranging from 5.76E-51 to 4.71E-10. 118 of the 237 associations were validated by a meta-analysis of the replication results of three cohorts. Some of the metabolite-associated CpG sites overlapped with those identified in the EWAS of main blood lipid levels.

Summarizing the results of both EWAS, a total of 17 different lipid- and fatty acid-related CpG sites were identified. These CpG sites were located in genes which are involved in lipid (*ABCG1*, *TXNIP*, *TNIP1*), fatty acid (*FADS2*, *SCD*, *CPT1A*, *SREBF1*, *MIR33B/SREBF1*) and amino acid (*PHGDH*, *SLC7A11*) metabolism. Replication in human adipose tissue and skin samples

indicated tissue-specificity for certain associations. Further analyses in KORA F4 showed that methylation of CpG sites located in *ABCG1*, *PHGDH* and *SCD* was additionally associated with the expression of their genes. Functional assays of lipid-associated CpG sites located in the genes *ABCG1* and *MIR33B/SREBF1* detected methylation-dependent protein-DNA binding affinity and revealed methylation-dependent reporter gene expression. Investigation of genetic influence on the associations identified methylation quantitative trait loci for the *FADS2* gene. CpG sites located in *ABCG1*, *CPT1A* and *TXNIP* also showed associations with type 2 diabetes or myocardial infarction in KORA F4. The results of this thesis may therefore contribute to a better understanding of the pathogenesis of complex lipid-related diseases.

ZUSAMMENFASSUNG

Blutspiegel von Lipiden wie High-Density-Lipoprotein-Cholesterin (HDL-C), Low-Density-Lipoprotein-Cholesterin (LDL-C), Triglyzeriden und Gesamtcholesterin gelten als vererbbare und auch beeinflussbare Risikofaktoren für kardiometabolische Erkrankungen. Medikamenteneinnahme und Lebensstilfaktoren wie Ernährung und sportliche Betätigung können die Lipidspiegel verändern. Einige Studien haben zudem einen genetischen Hintergrund von veränderten Lipidspiegeln festgestellt. Des Weiteren haben Studien in nicht-populationsbasierten Kohorten gezeigt, dass auch epigenetische Mechanismen zu den interindividuellen Schwankungen der Lipidlevel beitragen können. Allerdings gibt es noch keine Studie, die den Zusammenhang zwischen Lipidlevel und genomweiter DNA Methylierung in einer populationsbasierten Kohorte untersucht.

Die Ziele dieser Arbeit waren daher: 1) Durchführung einer epigenomweiten Assoziationsstudie (EWAS), um systematisch den Zusammenhang zwischen genomweiten DNA-Methylierungsmustern in Vollblut und Blutlipid-Spiegeln (HDL-C, LDL-C, Triglyzeride, Gesamtcholesterin) in der populationsbasierten KORA F4 Kohorte zu untersuchen; 2) Durchführung einer weiteren EWAS in der gleichen Kohorte, um den Zusammenhang zwischen genomweiten DNA-Methylierungsmustern und Serumspiegeln von Lipoprotein Subfraktionen und weiteren Metaboliten zu untersuchen, die mittels ^{1}H NMR-Spektroskopie gemessen wurden; 3) Durchführung von funktionellen Analysen um den Effekt des Methylierungsstatus von ausgewählten Lipid-assoziierten CpG Sites auf die Genregulation zu untersuchen.

In der EWAS der Blutlipid-Spiegel wurden 12 Assoziationen zwischen Methylierungszustand bestimmter Cytosin-Phosphat-Guanin Dinukleotid (CpG) Sites und bestimmter Lipidspiegel gefunden, wobei die p-Werte zwischen 1.21E-27 und 9.66E-08 lagen. Eine Metaanalyse der Replikationsergebnisse zweier unabhängigen Kohorten bestätigte neun der 12 Assoziationen. Die EWAS der NMR-gemessenen Metaboliten ergab 237 signifikante Assoziationen zwischen dem Methylierungszustand bestimmter CpG Sites und bestimmten Metaboliten, mit p-Werten zwischen 5.76E-51 und 4.71E-10. 118 der 237 Assoziationen wurden durch eine Metaanalyse der Replikationsergebnisse dreier unabhängiger Kohorten

validiert. Einige der CpG Sites, die Assoziationen zu Metaboliten zeigten, überlappten mit denen, die in der EWAS der Blutlipid-Spiegel identifiziert wurden.

Insgesamt konnten durch die zwei EWAS 17 CpG Sites identifiziert werden, die mit Lipid- und Fettsäure-Spiegeln assoziiert waren. Diese CpG Sites sind in Genen lokalisiert, die eine Rolle im Lipid- (*ABCG1*, *TXNIP*, *TNIP1*), Fettsäuren- (*FADS2*, *SCD*, *CPT1A*, *SREBF1*, *MIR33B/SREBF1*) und Aminosäuren-Metabolismus (*PHGDH*, *SLC7A11*) spielen. Die Replikation in humanem Fettgewebe und in Hautproben wies in einigen Fällen auf gewebsspezifische Zusammenhänge hin. Weitere Analysen in der KORA F4 Kohorte zeigten, dass die Methylierung von CpG Sites in *ABCG1*, *PHGDH* und *SCD* auch mit der Expression ihrer Gene zusammenhing. Funktionelle Analysen von Lipid-assoziierten CpG Sites, die in den Genen *ABCG1* und *MIR33B/SREBF1* liegen, zeigten zudem eine methylierungsabhängige Protein-DNA Bindungsaffinität und eine methylierungsabhängige Expression eines Reportergens. Die Untersuchung des genetischen Hintergrundes der Assoziationen führte zur Identifizierung von quantitativen Methylierungs Trait Loci im *FADS2*-Gen. CpG Sites, die in *ABCG1*, *CPT1A* und *TXNIP* lokalisiert sind, zeigten zudem einen Zusammenhang mit Typ 2 Diabetes und Myokardinfarkt in KORA F4. Die Ergebnisse der vorliegenden Arbeit können daher zu einem besseren Verständnis der Pathogenese von komplexen, Lipid-assoziierten Erkrankungen beitragen.

LIST OF ABBREVIATIONS

-C	total cholesterol	***et al.***	*et alteri*
-CE	cholesterol ester	**EWAS**	epigenome-wide association study
-FC	free cholesterol		
-L	total lipids	**(F)FA**	(free) fatty acids
-PL	phospholipids	**FH**	familial hypercholesterolemia
5-hmC	5-hydroxymethylcytosine	**GWAS**	genome-wide association study
5-mC	5-methylcytosine		
ALA	α-linolenic acid	**H**	histone
BMI	body mass index	**H3K9**	lysine 9 of histone H3
BMIQ	beta-mixture quantile normalization	**HbA1c**	hemoglobin A1c
		HDL-C	high-density lipoprotein cholesterol
bp	base pairs		
cDNA	complementary DNA	**HL**	hepatic lipase
CE	cholesterol ester	**IDL**	intermediate-density lipoprotein
CETP	cholesterol ester transfer protein		
CHD	coronary heart disease	**lnCHIANTI**	Invecchiare in Chianti, Aging in the Chianti Area
Chr	chromosome		
CpG	cytosine-phospho-guanosine dinucleotide	**kb**	kilobase pairs
		KORA	The Cooperative Health Research in the Region of Augsburg
CVD	cardiovascular diseases		
Cy5	cyanine 5		
DHA	docosahexaenoic acid	**LA**	linoleic acid
DNA	deoxyribonucleic acid	**LCAT**	lecithin-cholesterol acyl transferase
DNase1	deoxyribonuclease I		
DNMT	DNA methyltransferase	**LD**	Linkage Disequilibrium
E.coli	*Escherichia coli*	**LDL-C**	low-density lipoprotein cholesterol
e.g.	*exempli gratia*		
EMSA	electrophoretic mobility shift assay	**LDL-R**	low-density lipoprotein receptor
ENCODE	Encyclopedia of DNA Elements		

LOLIPOP	London Life Sciences Prospective Population Study
LPL	lipoprotein lipase
Mb	megabase pair
methQTL	methylation quantitative trait loci
mg	milligrams
min	minute
miRNA	microRNA
ml	milliliter
mRNA	messenger RNA
MUFA	monounsaturated fatty acid
MuTHER	Multiple Tissue Human Expression Resource
N	number of subjects
NFBC1966	Northern Finnish Birth Cohort 1966
ng	nanogram
nl	nanoliter
nm	nanometer
NMR	nuclear magnetic resonance
nt	nucleotides
PCR	polymerase chain reaction
PL	phospholipids
PLTP	phospholipid transfer protein
PPAR	peroxisome proliferator-activated receptors
PUFA	polyunsaturated fatty acid
RAR	retinoic acid receptors
RISC	RNA-inducing silencing complex
RNA	ribonucleic acid

RNAi	RNA interference
rpm	revolutions per minute
RT	room temperature
SAM	S-adenosylmethionine
SE	standard error
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SR-BI	scavenger receptor class B member 1
T2D	type 2 diabetes
TET	ten eleven translocation proteins
TF	transcription factor
TFBS	transcription factor binding site
TG	triglycerides
U	unit
UCSC	University of California Santa Cruz
UTR	untranslated region
UV	ultraviolet
VLDL	very-low-density lipoprotein
wbc	white blood cell count
x g	centrifuge rotor speed in g
YFS	Cardiovascular Risk in Young Finns Study
λ	wavelength
µl	microliters

1 INTRODUCTION

1.1 Epigenetics

1.1.1 Concept of epigenetics

The term *epigenetics* was defined by Conrad Hal Waddington in 1942 as "the branch of biology that studies the causal interactions between genes and their products, which bring the phenotype into being" (Goldberg *et al.* 2007). Today the term *epigenetics*, which is a composition of the ancient Greek prefix *epi-* (above or in addition to), and the word *genetics* indicates processes and mechanisms of chemical modifications of deoxyribonucleic acid (DNA) or of the structural and regulatory proteins bound to it, without changing the nucleotide sequence itself. Epigenetic mechanisms include DNA methylation, histone modifications and ribonucleic acid (RNA) interference (explained in detail in **chapter 1.1.2**). In contrast to the genome, which is largely static, the epigenome is much more dynamic and flexible and can change in response to different environmental conditions. With the discovery of epigenetic mechanisms that regulate the development and function of cells and tissues, a novel area of disease-relevant research was entered.

Data from human studies as well as animal models are providing evidence that the environment can affect epigenetic marks and thus may contribute to the development and progression of diseases. In studies with monozygotic twins the global methylation pattern of the individuals changed with increasing age (Fraga *et al.* 2005) and additional acquired differences in epigenetic state during life (Wong *et al.* 2010) may be a result of different environmental exposures. For example, several studies support the hypothesis that nutrients may modify the pattern of DNA methylation, either at the global scale or at locus-specific sites (Jimenez-Chillaron *et al.* 2012). Dietary components such as folate, methionine, vitamin B6 or betaine act as cofactors for reactions in the one-carbon metabolism which provide the bioavailability of S-adenosylmethionine (SAM), a methyl-donor for DNA methylation (Van den Veyver 2002). Limiting folate supply in humans results in increasing levels of homocysteine and reduced DNA methylation (Jacob *et al.* 1998; Rampersaud *et al.* 2000), whereas folate supplementation increases genomic DNA methylation in subjects with resected colorectal adenoma or cancer (Cravo *et al.* 1994; Pufulete *et al.* 2005). Dietary

methyl-donors may also contribute to change patterns of histone methylation through the provision of SAM and may also alter expression of specific RNAs (Jimenez-Chillaron *et al.* 2012).

Beside nutritional factors, other environmental toxins such as air pollution, dioxin, metals, bisphenol A or benzene also have an influence on the human epigenetic state (Christensen and Marsit 2011; Hou *et al.* 2012). The same applies for other environmental exposures such as alcohol consumption (Philibert *et al.* 2012; Krishnan *et al.* 2014), tobacco smoking (Breitling *et al.* 2012; Zeilinger *et al.* 2013; Gao *et al.* 2015) or psychological stress (Alegria-Torres *et al.* 2011; Klengel and Binder 2015).

Although there is evidence of environmental influence on epigenetic alterations, the mechanisms by which toxicants and the environment modulate the epigenetic landscape of individual cells need to be elucidated (Christensen and Marsit 2011).

1.1.2 Epigenetic mechanisms

Key players of epigenetic mechanisms are histone modifications (acetylation, methylation, ubiquitylation etc.), RNA interference (RNAi) and DNA methylation (Vaissiere *et al.* 2008; Delcuve *et al.* 2009). All epigenetic mechanisms interact with and influence each other, forming a complex regulatory machinery (Liep *et al.* 2012) (**Figure 1**).

Histone modification

The main component of the chromatin structure, which carries the genetic information, is the nucleosome, a structure consisting of an octamer of four core histone proteins around which a DNA strand of 147 base pairs (bp) length is wrapped (Kornberg and Lorch 1999). The octamer is composed of a H3-H4 histone tetramer interacting with two H2A-H2B dimers. The four histone proteins (H2A, H2B, H3 and H4) are each composed of a globular domain and an unstructured tail domain, on which a large number of covalent modifications occur including methylation, acetylation, phosphorylation and ubiquitylation (Strahl and Allis 2000; Khorasanizadeh 2004; Kouzarides 2007). The histone code is very complex, for example methylation of lysine 9 of histone H3 (H3K9) is linked to heterochromatin formation and gene silencing, whereas methylation of lysine 4 of H3 is linked to transcriptional activation (Martin and Zhang 2005). Also DNA methylation, another

component of the epigenome, interacts with the chromatin to control gene expression. Transcriptionally active regions of the chromatin are associated with hypomethylated DNA, whereas inactive chromatin is enriched with hypermethylated DNA (Razin and Cedar 1977). DNA methylation can trigger the methylation of H3K9 and interacts also with histone deacetylases, histone methyltransferases and methylcytosine-binding proteins in a complex network (Nan *et al.* 1998; Fuks *et al.* 2000; Fuks *et al.* 2003). The modifications regulate the open (eu-) and closed (hetero-) chromatin states, influencing the gene accessibility of transcription factors (Kouzarides 2007). Heterochromatin is highly condensed, gene-poor, transcriptionally silenced and typically enriched in hypoacetylated histones H3 and H4 and methylated H3K9. On the contrary, euchromatin is less condensed, gene-rich, transcriptionally active and characterized by highly acetylated histone tails and by H3K4 methylation (Rea *et al.* 2000; Jenuwein and Allis 2001; Nakayama *et al.* 2001; Noma *et al.* 2001; Huisinga *et al.* 2006).
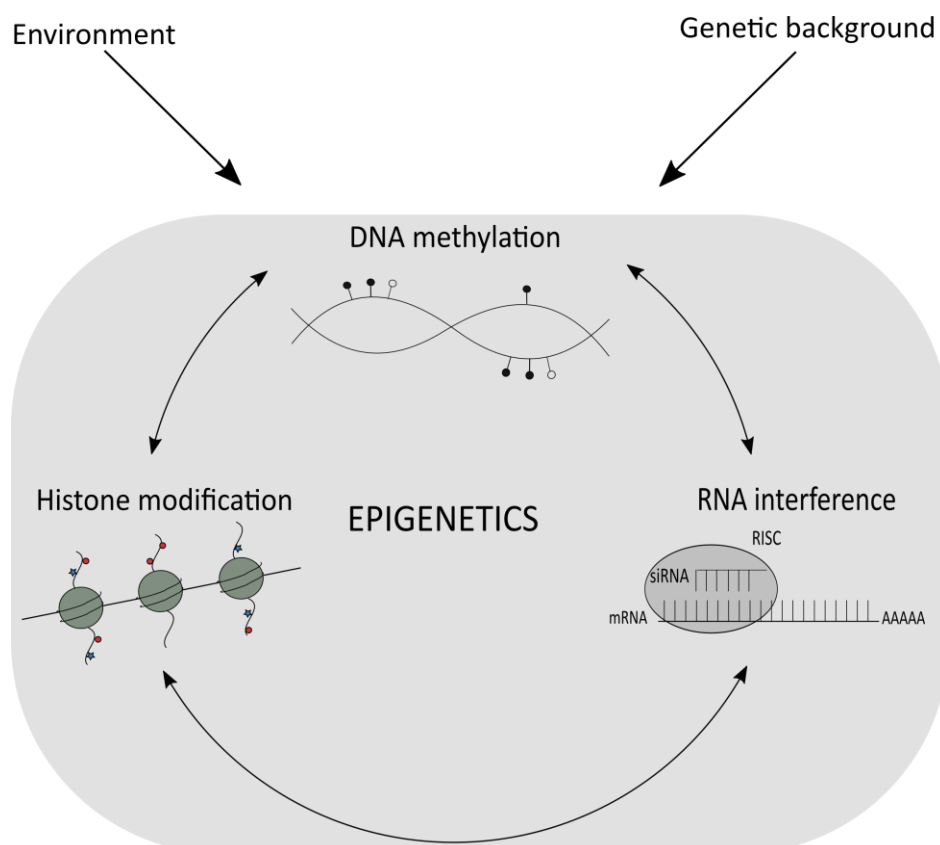
Figure 1: Interaction of epigenetic mechanisms. Enzymes of DNA methylation and histone modification define together with RNAi the chromatin state that is directly associated with the transcriptional state of a gene. Genetic background and environmental factors (e.g. tobacco smoking) additionally mediate epigenetic modifications of the genome, which might facilitate the establishment of certain diseases such as cancer (Christensen and Marsit 2011).

RNA interference (RNAi)

RNAi is a non-coding RNA dependent, sequence-specific regulation of gene expression. Epigenetic related non-coding RNAs include among others small interfering RNAs (siRNAs) and microRNAs (miRNAs). Their contribution in epigenetic gene regulation and transcriptional processes as well as their roles in several diseases give an indication of the existence of a regulatory network controlled by RNA (Mattick and Makunin 2005).

MiRNAs, which are small non-coding RNAs (~20-30 nucleotides) encoded in the genome, are considered as key post-transcriptional regulators of the expression of several protein-coding genes (Bartel 2009), including genes involved in the high-density lipoprotein metabolism (Davalos and Fernandez-Hernando 2013). Currently, about 1,881 human miRNAs are known according to the latest version (Release 21) of the microRNA database miRBase (mirbase.org) (Kozomara and Griffiths-Jones 2014). MiRNAs thus compose one of the largest families of gene regulatory molecules in the human genome.

MiRNAs are usually transcribed by RNA polymerase II in the nucleus to generate primary-miRNAs which are subsequently processed by Drosha (a RNase III) to form pre-mature miRNAs (Rottiers and Naar 2012). After translocation to the cytoplasm the pre-miRNAs are further processed by a miRNA-specific nuclease (Dicer) into a mature miRNA duplex. One strand of the duplex is incorporated into a RNA-inducing silencing complex (RISC) by associating with Argonaute proteins. The complex harboring the miRNA binds to the 3´-UTR of the target messenger RNA (mRNA) and inhibits gene expression via induction of mRNA degradation or translational repression (Rottiers and Naar 2012). The other strand of the duplex may be degraded or released from the cell through different export mechanisms (Creemers *et al.* 2012). Similarly to miRNA, siRNAs also inhibit post-transcriptional gene expression. The major difference between these two classes of non-coding RNAs is that miRNAs are endogenous products of an organism´s genome, whereas siRNAs originate from an exogenous source (derived e.g. from a virus) (Carthew and Sontheimer 2009). Also, each siRNA inhibits the expression of one specific target mRNA, whereas a miRNA can regulate the expression of multiple mRNAs (Lam *et al.* 2015).

The RNA machinery also interacts with other epigenetic mechanisms. One study has shown that the expression of miRNAs can be regulated by DNA methylation of their promoters

(Saito and Jones 2006). MiRNAs can also directly target DNA methyltransferases (DNMTs) (Duursma *et al.* 2008) and chromatin remodeling enzymes (Chen *et al.* 2006).

Studies indicate that miRNAs play a role in the development of several diseases such as carcinogenesis or cardiovascular diseases (Macfarlane and Murphy 2010; Gurha 2016). Also, multiple miRNAs appear to be attractive therapeutic targets (van Rooij *et al.* 2012). In 2013, the first miRNA-based anti-cancer drug, MXR-34, a liposome-based miR-34 mimic, reached phase 1 clinical trials for the treatment of hepatocellular carcinoma (Agostini and Knight 2014). Moreover, plasma or serum miRNA expression profiles can be used as biomarkers to determine the outcome of certain diseases (Moldovan *et al.* 2014). In addition, lipoprotein-bound miRNAs have a great potential to become a novel class of disease biomarkers. Extracellular miRNAs in plasma are protected from circulating nucleases by being packaged in exosomes, microvesicles, or binding to ribonucleoproteins such as Argonaute 2. But also high-density lipoproteins (HDL) and low-density lipoproteins (LDL) are miRNA transporters (Michell and Vickers 2016). This lipoprotein transport is a new research field and a lot of questions are yet unanswered. However, one study has shown that levels of miRNAs associated with HDL and LDL are altered in different cardiometabolic diseases in humans (Niculescu *et al.* 2015), giving hope for new potential drug targets and/or a novel class of disease biomarkers.

1.1.3   DNA methylation

1.1.3.1   Basic principles

DNA methylation was the first mechanism to be recognized as an epigenetic silencing mechanism (Holliday and Pugh 1975; Riggs 1975). Following research has demonstrated that DNA methylation has a critical role in normal development and cell functions, including imprinting, X-chromosome inactivation and tissue-specific gene expression (Csankovszki *et al.* 2001; Jones and Takai 2001; Kaneda *et al.* 2004).

In mammals, DNA methylation mainly occurs in the context of cytosine-phospho-guanosine dinucleotides (so called CpG sites) and the majority (~74%) of CpG sites throughout the genome is methylated (Tost 2010). CpG sites are generally underrepresented in the genome, but enriched in CpG islands which are CpG-rich clusters with a ~4 kilo base pairs

(kb) length (Illingworth and Bird 2009). DNA methylation is introduced by DNA methyltransferases (DNMTs) which catalyze the transfer of a methyl group from the methyl donor, S-adenosylmethionine, to the 5-position of cytosine residues in DNA. Methylation patterns are established by the de novo methyltransferases Dnmt3a and Dnmt3b and then conserved by the maintenance methyltransferase Dnmt1 (Jeltsch 2006; Jurkowska *et al.* 2011) (**Figure 2**).
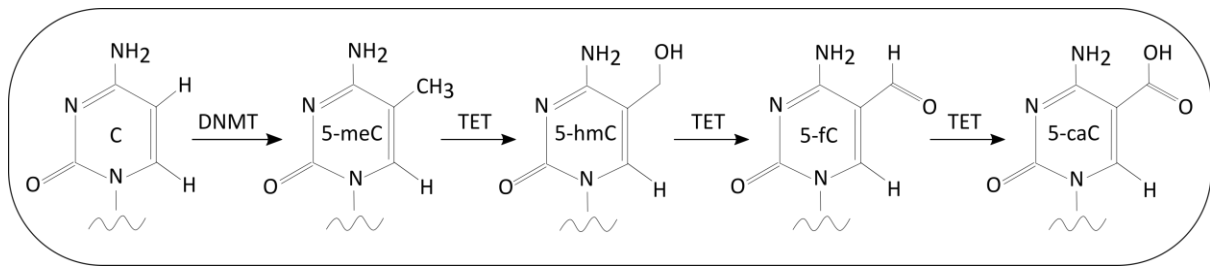
Figure 2: DNA cytosine methylation and TET-initiated DNA demethylation pathways. DNA methylation (5-mC) is established and maintained by DNMT. TET proteins can oxidize 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC).

Beside this, recent discoveries have shown that ten eleven translocation (TET) proteins can oxidize 5-methylcytosine (5-mC) resulting in the formation of 5-hydroxymethylcytosine (5-hmC) (Kriaucionis and Heintz 2009; Tahiliani *et al.* 2009), providing a new type of DNA modification. TET-mediated 5-mC oxidation is able to proceed further to generate 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (Ito *et al.* 2011) (**Figure 2**). Studies have indicated that 5-hmC and 5-fC are broadly present in the genomic DNA of various tissues and cell types whereas the presence of 5-caC appears to be more limited (Ito *et al.* 2011). 5-hmC enriched regions were found in promoter regions, transcription start sites, exons and in gene bodies of actively transcribed genes, suggesting a role of 5-hmC in regulating gene transcription (Ficz *et al.* 2011; Pastor *et al.* 2011; Wu *et al.* 2011). However, further research is needed to address the precise function of 5-hmC and the other oxidized 5-mC derivatives.

1.1.3.2   DNA methylation and gene expression

The function of DNA methylation on gene expression is very complex. The focus of previous studies about DNA methylation and its impact on gene expression was on CpG islands at transcription start sites leading to the suggestion that hypermethylation of CpG islands is

associated with epigenetic silencing and hypomethylation with transcriptional activity (Jones and Takai 2001; Weber *et al.* 2007; Illingworth and Bird 2009). However, it is not necessarily the case that DNA methylation of CpG islands and gene expression are inversely correlated. Recent studies also showed positive associations between expression and methylation levels (Gibbs *et al.* 2010; Bell *et al.* 2011; van Eijk *et al.* 2012). It was found that *trans*-associations (outside of a 500 kb interval between expression and methylation locus) between methylation and expression levels were enriched for positive correlations, whereas *cis*-associations (within the 500 kb interval) were both positive and negative (van Eijk *et al.* 2012). Negative associations were found to be enriched especially in promoter regions (Bell *et al.* 2011) and CpG islands (Gibbs *et al.* 2010).

On gene level, hypermethylation of a specific core region of the gene promoter (which often includes the transcription start site) does not automatically silence the gene, but does have a modifying effect on the expression of the gene (Ushijima 2005). Additionally, the extent of this modification depends on the position of the methylation in the transcriptional unit. For example, methylation in very close distance to a transcription start site may have a silencing effect, whereas methylation of the gene body might even activate expression (Jones 2012). Interestingly, gene body methylation was also recently linked with splicing (Shukla *et al.* 2011).

Numerous processes how DNA methylation can influence gene transcription have been proposed. DNA methylation can directly prevent transcription factor (TF) binding (Comb and Goodman 1990) as certain TFs are unable to bind to their recognition sites within DNA when 5-methylcytosine occurs within a critical base. Additionally, DNA methylation can recruit methylcytosine-binding proteins preventing thereby the binding of potentially activating transcription factors (Wade 2001; Tost 2010). DNA methylation can also lead to changes in the chromatin structure, which is the mechanism to heritably lock the gene in its inactive state (Bird 2002).

In summary, DNA methylation represents a complex regulatory mechanism of gene expression. Further studies, particularly at the genome-wide level, would help to understand this intricate machinery.

1.1.3.3   Genetic influence on DNA methylation levels by methylation quantitative trait loci

Several studies have shown that genetic variation is often associated with quantitative changes in methylation levels (Gibbs *et al.* 2010; Zhang *et al.* 2010; Bell *et al.* 2011; Gutierrez-Arcelus *et al.* 2013). More *cis-* than *trans-*associations between genetic variations and DNA methylation have been identified with peak enrichment for methylation quantitative trait loci (methQTLs) located in regions with a distance of less than 1500 bp to a transcription start site (Gibbs *et al.* 2010; Zhang *et al.* 2010; Bell *et al.* 2011).

Recently it was reported that the correlation of DNA methylation within a cluster of CpG sites was similar to the LD (linkage disequilibrium) correlation in genetic single nucleotide polymorphism (SNP) variation, but for much shorter distances when contiguous CpG sites were considered (Liu *et al.* 2014). Some of these CpG clusters appeared to be genetically driven as SNPs in LD blocks showed associations with correlated methylated CpG sites (Liu *et al.* 2014). Banovich et al. reported a frequent association between methQTLs and changes in methylation at multiple CpG sites across regions of up to 3 kb (Banovich *et al.* 2014).

MethQTLs are also associated with variation in a variety of other types of chromatin changes including histone modifications, DNase I and chromatin accessibility as well as with expression changes of nearby genes (Banovich *et al.* 2014). Interestingly, SNPs that show an allele-specific change in predicted TF binding affinities are more likely to be associated with DNA methylation than SNPs outside of TF binding regions. Thus, variations in TF binding sites may result in changes in methylation status of nearby CpG sites ending in altered regulatory processes (Banovich *et al.* 2014).

All previous findings suggest a strong genetic component to inter-individual variation in epigenetic signature, but the mechanistic link between methQTLs and DNA methylation still needs to be investigated in detail.

1.2 Introducing the world of blood lipids

1.2.1 Role of blood lipid levels in health and disease

Altered lipid levels are associated with metabolic and vascular diseases such as type 2 diabetes and atherosclerosis. Atherosclerosis is the underlying cause of cardiovascular diseases (CVD) which is the major cause of deaths world-wide (Elder *et al.* 2009).

The recommended values for normal levels of the human blood lipids total cholesterol, triglycerides (TG), HDL-C and LDL-C are shown in **Table 1**. The characteristics of dyslipidemia are elevated TG and total or LDL cholesterol levels, and reduced HDL-C levels. The combined altered levels are commonly referred to as the atherogenic lipoprotein phenotype (Mooradian 2009). This phenotypic pattern is characteristic for patients with diabetes or metabolic syndrome (National Cholesterol Education Program Expert Panel on Detection and Treatment of High Blood Cholesterol in 2002) and this combination increases significantly the risk for CVD (Isomaa *et al.* 2001; Lakka *et al.* 2002).

Table 1: Recommended human blood lipid levels according to the guidelines of the National Cholesterol Education Program (NCEP) (Expert Panel on Detection and Treatment of High Blood Cholesterol in 2001)

Parameter	Concentration mg/dl (mmol/l)
Total cholesterol	<200 (<5.16)
LDL-C: dependent on the global risk factor*	
0-1 risk factors	<160 (<4.13)
2 or more risk factors	<130 (<3.35)
Coronary heart disease/Diabetes Mellitus	<100 (<2.58)
HDL-C	≥40 (≥1.03)
Triglycerides	<150 (<1.69)

*risk factors: age (men >45 years, women >55 years), HDL-C <40 mg/dl (<1.03 mmol/l), smoker, hypertension, positive family history on coronary heart disease)

Furthermore, each particular lipid level (HDL-C, LDL-C or total cholesterol) is also a well-established independent risk factor for developing of atherosclerosis and CVD (National Cholesterol Education Program Expert Panel on Detection and Treatment of High Blood Cholesterol in 2002). It has been suggested that small, dense LDL particles have a greater oxidative susceptibility and thus an enhanced atherogenic potential. However, it is not clear

to what extent increased levels of these particles contribute to the disease development (Berneis and Krauss 2002; Krauss 2010). In contrast, there is an inverse relationship between plasma HDL-C levels and the risk for cardiovascular diseases (Gordon *et al.* 1989). The anti-atherosclerotic effect of HDL-C may be due to its ability to mobilize excess cholesterol from artery wall macrophages in context of the reverse cholesterol transport (Mazzone *et al.* 2008).

How TG are associated with CVD is an ongoing debate. The controversy is due to the inverse correlation of TG levels with HDL-C levels; adjustment for HDL-C attenuates the relationship between TG and CVD (Bansal *et al.* 2007). Another important factor is the time point of TG measurement: it is recommended that for blood lipid profiles the blood should be drawn after an 8- to 12-hour fast (National Cholesterol Education Program Expert Panel on Detection and Treatment of High Blood Cholesterol in 2002) as TG levels can increase substantially in the postprandial state and fasting measurements avoid the meal-associated variability. However, postprandial lipids may play a crucial role in the pathogenesis of CVD as the hydrolysis of TG-rich lipoproteins were shown to result in atherogenic cholesterol-enriched remnant lipoproteins which support the building of atherosclerotic plaques (Proctor and Mamo 1998). Accordingly, one study showed a strong correlation between such atherogenic cholesterol-enriched remnant lipoproteins and non-fasting TG (Nordestgaard *et al.* 2007). Additionally, two other studies identified non-fasting TG as a superior predictor of incident CVD compared with fasting TG levels (Bansal *et al.* 2007; Nordestgaard *et al.* 2007).
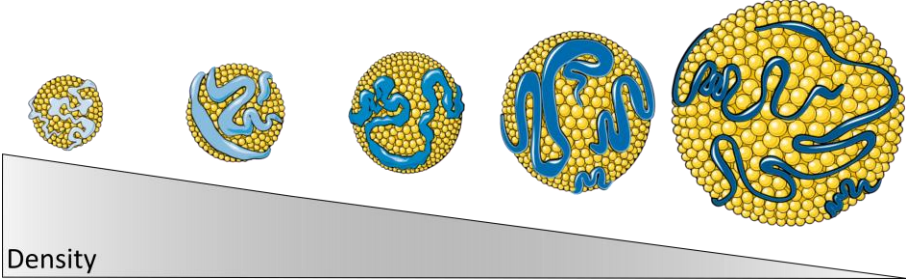
1.2.2   Lipoproteins and their subclasses

Lipids, including triglycerides (TG) and phospholipids (PL), are involved in many main biological functions such as molecular reactions, energy storage and building of structural components of cell membranes (Subramaniam *et al.* 2011). Due to their poor solubility in blood, lipids must be transported in lipoprotein particles which are complex particles consisting of a hydrophobic core of TG and cholesteryl esters surrounded by a monolayer of cholesterol, phospholipids and apolipoproteins (Feingold and Grunfeld 2000). The apolipoproteins such as ApoE and ApoB have several functions including receptor binding properties and activation of enzymes (Ramasamy 2014). Plasma lipoproteins can be

categorized according to their density, lipid and lipoprotein compositions or particle size. The latter one is often used to classify lipoproteins into five groups: chylomicrons, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL) and high-density lipoproteins (HDL). Using this categorization the principle applied that the less dense the particles are, the larger they are in size and compose more lipids than proteins (Feingold and Grunfeld 2000) (**Table 2**). Two HDL-subclasses can be identified using analytic ultracentrifugation: HDL2 and HDL3. HDL2 corresponds to the less dense (1.063 – 1.125 g/ml) and relatively lipid-rich form of HDL particles, whereas HDL3 consists of the more dense (1.125 – 1.21 g/ml) and relatively protein rich form of HDL particles (De Lalla and Gofman 1954). These two particles form the major subpopulations of circulating HDL in human plasma.

Table 2: Characteristics of lipoproteins. Density, diameter, constituent/weight percentages for the lipid components, and main apolipoprotein for the main lipoprotein particles

	HDL	LDL	IDL	VLDL	Chylomicrons
Density (g/ml)	1.06–1.21	1.02–1.06	1.01–1.02	0.95–1.01	<0.95
Diameter (nm)	5–15	18–28	25–50	30-80	100–500
Cholesterol (%)	30	50	29	22	8
Phospholipid (%)	29	21	22	18	7
Triglyceride (%)	8	4	31	50	84
Protein (%)	33	25	18	10	1-2
Main apolipoprotein	ApoA	ApoB	ApoB	ApoB, ApoC	ApoB

Modified from R. Garrett & C. Grisham, Biochemistry 4th Edition, 2010, USA; illustration was prepared using a template on the Servier medical art website (http://www.servier.com/Powerpoint-image-bank)

The lipoproteins can be further subdivided into more detailed lipoprotein subclasses that further characterize the different sizes (e.g. small, medium or large) and compositions of the particles (e.g. phospholipids or cholesteryl-ester in the particle). These subclasses have

attracted a lot of attention as they may differ in metabolic and functional relevance (Asztalos *et al.* 2011; Basu *et al.* 2016). Especially the association between lipoprotein subfractions and their role in cardiovascular events has been discussed in detail (Superko *et al.* 2012; Hoogeveen *et al.* 2014; Tian *et al.* 2014).

1.2.3 Lipoprotein metabolism

Three pathways are responsible for the generation and transport of lipids within the body, the exogenous, the endogenous and the reverse cholesterol transport pathway (Feingold and Grunfeld 2000).

The exogenous pathway: transport of dietary lipids

The exogenous pathway of lipoprotein metabolism permits efficient transport of dietary lipids. After digestion and absorption of dietary fat, TG and cholesterol are accumulated to form chylomicrons in the epithelial cells of the intestine, which are then delivered directly to the systemic circulation (Feingold and Grunfeld 2000). Inside the blood vessels, the enzyme lipoprotein lipase (LPL), which is anchored to the luminal surface of capillary endothelial cells, hydrolyzes the TG of the chylomicrons and free fatty acids are released. Some free fatty acids bind to albumin and are transported to other tissues, especially the liver (Tulenko and Sumner 2002). Due to the release of TG, the chylomicron particles progressively shrink in size resulting in smaller particles, the chylomicron remnants. The remnant particles are rapidly taken up from the circulation into the liver mainly through LDL receptors (LDL-R) or LDL-R related proteins (Feingold and Grunfeld 2000; Kwiterovich 2000; Tulenko and Sumner 2002; Ramasamy 2014). Chylomicrons are mainly present in human blood in the postprandial state (Feingold and Grunfeld 2000; Ramasamy 2014), thus lipoprotein measurements in fasting individuals do normally not include chylomicrons.

The endogenous pathway: transport of hepatic lipids

The endogenous pathway begins with the synthesis of TG-rich VLDL particles in the liver (Feingold and Grunfeld 2000; Ramasamy 2014) (**Figure 3**). The TG of VLDL particles are derived predominantly from the esterification of long-chain fatty acids. Once secreted to the plasma, TG of VLDL particles are hydrolyzed by LPL and fatty acids are released which are taken up by muscle cells for energy and by adipose cells for storage. The resulting IDL

particles, which are relatively enriched in cholesterol esters and acquire ApoE from HDL particles, can be removed from the circulation by the LDL-R or LDL-R related protein receptors (Feingold and Grunfeld 2000). Most of the remaining TG in the IDL particles are hydrolyzed by the hepatic lipase (HL) and all apolipoproteins except ApoB-100 are transferred to other lipoproteins, leading to formation of LDL particles (Feingold and Grunfeld 2000; Kwiterovich 2000). Approximately 70% of the circulating LDL particles are cleared by LDL-R-mediated endocytosis into the liver; the remainder is taken up by peripheral tissues (Feingold and Grunfeld 2000).

Figure 3: Schematic illustration of the endogenous lipoprotein metabolism and reverse cholesterol transport. In blood plasma, TG of hepatic VLDL particles are hydrolyzed by LPL to IDL releasing FFA **(a)**. IDL is taken up by the liver or further hydrolyzed by HL to LDL **(b)**. LDL particles are taken up by peripheral tissues or are removed by the liver **(c)**. Hepatic HDL is first converted via LCAT to HDL3 and then to HDL2 by the efflux of cellular free cholesterol (e.g. from macrophages by the ABCA1 and ABCG1 transporters), which is taken up from the liver via SR-BI **(d)**. Alternatively, cholesterol esters of HDL2 are transferred in exchange for TG to VLDL and LDL by CETP **(e)**. PLTP transfers phospholipid between HDL2 and VLDL **(f)**. The dashed arrows indicate steps of the reverse cholesterol transport, by which peripheral cholesterol is transported back to the liver.
LPL: lipoprotein lipase; FFA: free fatty acids; HL: hepatic lipase; LCAT: lecithin-cholesterol acyl transferase; SR-BI: scavenger receptor class B member 1; CETP: cholesterol ester transfer protein; PLTP: phospholipid transfer protein; Figure was prepared using a template on the Servier medical art website (http://www.servier.com/Powerpoint-image-bank).

Reverse cholesterol transport

Peripheral cells accumulate cholesterol through the uptake of circulating lipoproteins and de novo cholesterol synthesis. For an efficient metabolism and excretion of cholesterol a reverse cholesterol transport is required where cholesterol in peripheral cells is transported from the plasma membranes of peripheral cells to the liver by an HDL-mediated process (Ramasamy 2014).

On of the first steps of the HDL biogenesis is the secretion of ApoA1 by the intestine and the liver, which then conglomerates with phospholipids to form discoidal nascent HDL that can bind cholesterol (Joy and Hegele 2008). The nascent HDL particles functionally interact with the ATP binding cassette subfamily A member 1 (ABCA1) transporter. This leads to the transfer of cellular phospholipids and free cholesterol, for example from macrophages in the arterial wall, to the nascent HDL (**Figure 3**). The cholesterol is then esterified by lecithin-cholesterol acyl transferase (LCAT) resulting in small and dense HDL3 particles (Joy and Hegele 2008; Ramasamy 2014). HDL is then transformed into spherical HDL2 after acquiring additional cholesterol by the ATP binding cassette subfamily G member 1 (ABCG1) transporter and esterification of the cholesterol by LCAT (Joy and Hegele 2008). This spherical, mature HDL is the main form of HDL responsible for cholesterol transport to the liver (Ramasamy 2014). *In vitro* studies have shown that ABCA1 and ABCG1 synergistically mediate cholesterol efflux to HDL (Gelissen *et al.* 2006). The cholesterol esters originating mainly from the mature large HDL particles are delivered directly to the liver through the scavenger receptor class B member 1 (SR-BI) protein, the surface components of HDL are recycled and the lipid-depleted, nascent HDL particles again enter the HDL maturation process (Feingold and Grunfeld 2000) (**Figure 3**). The cholesterol can be recycled or be catabolized to bile acids (Ramasamy 2014). Alternatively, cholesterol esters of HDL2 are transferred in exchange for TG to VLDL and LDL particles by the cholesterol ester transfer protein (CETP) with subsequent uptake of LDL by hepatic LDL receptors. This process is termed as the indirect reverse cholesterol transport pathway (Ramasamy 2014). Another enzyme, the phospholipid transfer protein (PLTP), transfers phospholipids between HDL and VLDL particles and also between different HDL particles (Ramasamy 2014). The role of PLTP in the reverse cholesterol transport is however not yet clarified (Yazdanyar *et al.* 2011).

1.2.4   Genetic background of human blood lipid levels

According to the NHGRI GWAS Catalog (Welter *et al.* 2014), over the past decade genome-wide association studies (GWAS) identified about 300 genetic variants being associated with various lipids or lipoprotein measurements. The most recent update on the GWAS-generated list of loci was reported by Willer *et al.*, which added 62 loci to the original loci reported in 2010 by Teslovich *et al.*, ending with 157 lipid associated loci in total (Teslovich *et al.* 2010; Global Lipids Genetics *et al.* 2013). 32 of the 62 new loci were already previously connected to lipoprotein metabolism, whereas the remaining 30 have not yet been reported to be associated with lipids. 92 loci showed associations with only one of the four main lipid traits (total cholesterol, TG, LDL-C, HDL-C), whereas 65 were associated with two or more of the lipid traits including four loci which even associated with all four lipid traits.

In addition to the mentioned studies, Chasman *et al.* analyzed the association between 17 nuclear magnetic resonance (NMR)-based measures (concentration of eight lipoprotein subclasses, the total number of particles for four lipoprotein classes, mean particle sizes for three lipoprotein classes, and the estimates for HDL-C and TG) and five lipoprotein and apolipoprotein measures determined using conventional assays (Chasman *et al.* 2009). 31 loci were identified to be associated with one or more of the 22 lipoprotein measures. Also Kettunen *et al.* performed a genome-wide association analysis of serum metabolic phenotypes assessed by NMR of serum samples including lipid and lipoprotein measures (Kettunen *et al.* 2012). They identified 18 associations of which four have not previously been reported (Kettunen *et al.* 2012). In addition, Petersen *et al.* conducted an association analysis between the 95 lipid loci identified by Teslovich *et al.* and 15 NMR-measured lipoprotein subclasses and concludes that NMR-based fine mapping of lipoprotein subfractions strengthens the association with lipid loci (Petersen *et al.* 2012).

Despite all efforts, the variance of lipid levels explained by the currently known genetic variants is modest. All lipid-associated SNPs together explain less than 15% of the variation in the plasma lipid traits (Global Lipids Genetics *et al.* 2013), although the estimated heritable variance of lipids is estimated to be at least 50% (Goode *et al.* 2007). This is called the missing heritability, which perhaps may be partly explained by epigenetic processes such as DNA methylation (Johannes *et al.* 2008).

1.3   Blood lipid levels and DNA methylation – what is known so far?

Some studies have already investigated the association between DNA methylation and lipid levels. Guay *et al*. published three studies examining the association between DNA methylation of several specific gene loci and lipid levels as well as CVD in patients with familial hypercholesterolemia (Guay *et al.* 2012a; Guay *et al.* 2013; Guay *et al.* 2014). Here, CpG sites located in *ABCA1* and *LPL* showed an association with HDL-C and HDL particle size, which was also the case for CpG sites in the *CETP* gene in men (Guay *et al.* 2012a; Guay *et al.* 2013). Furthermore, CpG sites in the genes *ABCG1*, *LIPC* (hepatic lipase C) and *PLTP* were associated with HDL-C, LDL-C and TG levels in a sex-specific manner (Guay *et al.* 2014). Additionally, a relationship was revealed between the previous history of CVD and methylation of the genes *ABCA1* and *LIPC* (Guay *et al.* 2012a; Guay *et al.* 2014). Another study discovered a prenatal malnutrition-associated differentially methylated region in the *CPT1A* gene which was also associated with LDL-C levels in the later life (Tobi *et al.* 2014). In very young children (17 months old), whole blood total cholesterol and HDL-C levels showed associations with methylation of *TNFα* (tumor necrosis factor-alpha) and *LEP* (leptin) (Wijnands *et al.* 2015).

There have also been some studies exploring associations between lipids and DNA methylation using a genome-wide approach. An EWAS in patients of familial hypercholesterolemia identified *TNNT1* (troponin T1) methylation to be associated with HDL particle size, HDL-phospholipids, HDL-ApoA1 and HDL-C (Guay *et al.* 2012b). Another EWAS discovered that differentially methylated regions in *STAT5A* (signal transducer and activator of transcription 5A), which were associated with age-related phenotypes, also showed associations with LDL levels. Additionally, age-related DNA methylation differences in the *CSMD1* (CUB and Sushi multiple domains 1) gene in monozygotic twins were associated with total cholesterol and LDL levels (Bell *et al.* 2012). However, the results of both studies have to be considered carefully as they were both measured with the Infinium HumanMethylation27 BeadChip Array (Illumina) which has a limited density of probe coverage and focusses mostly on the cancer pathways (Bibikova *et al.* 2009). Additionally, both analyses were done using DNA methylation information of whole blood DNA without correction for cellular heterogeneity (Guay *et al.* 2012b) or with only indirect testing for blood cell subtypes variability (Bell *et al.* 2011). Two other EWAS circumvent the issue of cell

heterogeneity using DNA methylation data obtained from one blood cell type, namely CD4[+] T cells (Frazier-Wood *et al.* 2014; Irvin *et al.* 2014). Frazier-Wood *et al*. used DNA methylation data analyzed with the Infinium HumanMethylation450 BeadChip Array (Illumina) to investigate the associations between 14 lipoprotein subfraction measures and genome-wide DNA methylation in N=663 participants of the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN). They discovered two CpG sites (namely cg00574958 and cg17058475) in the *CPT1A* gene to be associated with total LDL, small LDL, the average diameter of LDL particles (only cg00574958), as well as total, medium and large VLDL particles (both CpG sites) (Frazier-Wood *et al.* 2014). The results were confirmed by another study which additionally found associations between the CpG sites and TG levels (Gagnon *et al.* 2014). A separate study in GOLDN (N=991) described an inverse relationship of the same *CPT1A* loci and two nearby CpG sites with fasting TG as well as VLDL-C, a finding that was replicated in the Framingham Heart Study (Irvin *et al.* 2014). A methQTL analysis showed that an influence of nearby variants from GWAS (located within 1 megabase pair (Mb) up- or downstream of the locus) can be excluded (Irvin *et al.* 2014). An expression analysis revealed a decreased *CPT1A* expression in case of increased cg00574958 methylation (Irvin *et al.* 2014).

1.4 Aim of this thesis

The first aim of this thesis was to perform an epigenome-wide association study (EWAS) to systematically examine the association between genome-wide DNA methylation and main blood lipid levels of HDL-C, LDL-C, triglycerides and total cholesterol in whole blood of a large population-based cohort. Results were to be validated through replication in whole blood, human adipose tissue and skin samples of independent cohorts.

The second aim was to conduct an EWAS of lipoprotein subfractions as well as other metabolites measured by ^{1}H NMR spectroscopy in the same cohort to get an in-depth insight into the relation between epigenetic processes and lipid metabolism.

The third aim of the thesis was to assess the biological relevance of selected lipid-associated CpG loci through functional analyses. For this purpose, electrophoretic mobility shift assays as well as Luciferase Reporter Assays were to be conducted to investigate both methylation-dependent protein-DNA binding affinity and methylation-dependent gene expression.

This work was dedicated to exploring the interrelation between DNA methylation and the regulation of inter-individual lipid level variability in humans, and to achieving new insights into the complex picture of lipid-related complex diseases.

2 METHODS

2.1 Study populations

2.1.1 Discovery cohort

The Cooperative Health Research in the Region of Augsburg (KORA) study is a series of independent population-based epidemiological surveys and follow-up studies of participants living in the region of Augsburg, Southern Germany. The studies have been conducted according to the principles expressed in the Declaration of Helsinki and were approved by the local ethics committee (Bayerische Landesärztekammer). Written informed consent has been given by each participant. The **KORA F4** study, a seven-year follow-up study of the KORA S4 survey (examined between 1999 and 2001), was conducted between 2006 and 2008. The standardized examinations applied in the survey have been described in detail elsewhere (Holle *et al.* 2005; Wichmann *et al.* 2005). A total of 3,080 subjects with ages ranging from 32 to 81 years participated in the examination. At the time of the KORA F4 visit, blood samples for metabolic analyses, DNA and RNA extraction were collected. In a random subgroup of 1,802 subjects the genome-wide DNA methylation patterns were analyzed using the Infinium HumanMethylation450 BeadChip (Illumina).

2.1.2 Replication cohorts for the EWAS of main blood lipids

The **KORA F3 cohort** is a ten years follow-up survey of the KORA S3 survey from 1994-1995, as described previously (Lowel *et al.* 2005; Wichmann *et al.* 2005). The KORA F3 and F4 survey are completely independent with no overlap of individuals. In KORA F3 Infinium HumanMethylation 450 BeadChip (Illumina) data for 499 subjects were available (comprised of 250 smokers and 250 non-smokers; one sample was excluded after quality control (Zeilinger *et al.* 2013)).

The **Invecchiare in Chianti, Aging in the Chianti Area (InCHIANTI)** study is a population-based study initiated in 1998 in the Chianti region of Tuscany, Italy. Data for the replication were taken from the participants´ third follow-up visit in 2007/2008. DNA methylation data were available for 472 subjects, analyzed using the Infinium HumanMethylation450 BeadChip (Illumina).

The **Multiple Tissue Human Expression Resource (MuTHER)** study includes 856 female twins recruited from the TwinsUK Adult twin registry (154 monozygotic twin pairs, 232 dizygotic twin pairs and 84 singletons) (Grundberg *et al.* 2012). Infinium HumanMethylation 450 BeadChip (Illumina) data were available for N=648 individuals on adipose tissue biopsies (Grundberg *et al.* 2013) and for N=469 on skin biopsies. Due to missing phenotype information the final sample size for the analysis was 634 for adipose samples and 395 for skin samples. Blood lipids and other phenotypes were measured at the same time point as the biopsies as described elsewhere (Aulchenko *et al.* 2009).

2.1.3 Replication cohorts for the EWAS of NMR-measured metabolites

The **London Life Sciences Prospective Population Study (LOLIPOP)** is a prospective cohort study of ~28K Indian Asian and European men and women, recruited from the lists of 58 General Practitioners in West London, United Kingdom between 2003 and 2008. The LOLIPOP study is approved by the National Research Ethics Service (07/H0712/150) and all participants gave written informed consent. In 4,060 samples of Indian-Asian subjects DNA methylation was quantified in bisulfite converted genomic DNA from whole blood, using the Illumina Infinium HumanMethylation450 BeadChip (Illumina).

The **Northern Finnish Birth Cohort 1966 (NFBC1966)** is a prospective population-based birth cohort, including all the mothers (N=12,055) with children whose expected date of birth was in the year 1966 in the two northernmost provinces of Finland. In 1997-1998, a postal questionnaire on health, social status and lifestyle was sent to the living cohort members, and those living in the original target area or in the capital area were invited for a clinical examination, including blood sample collection. Ethical approval was obtained from the ethical committee of the Northern Ostrobothnia Hospital District and all participants gave written informed consent. DNA methylation patterns were analyzed for 809 subjects using the Infinium HumanMethylation450 BeadChip (Illumina).

The **Cardiovascular Risk in Young Finns Study (YFS)** is an ongoing multicentre Finnish longitudinal population study sample on the evolution of cardiovascular risk factors from childhood to adulthood. The study began in 1980, when 3,596 participants between the ages of 3 and 18 were randomly selected from the national population registers. YFS was approved by the Ethics Committee of the Hospital District of Southwest Finland. All

participants gave their written informed consent. In a subsample of 184 individuals randomly assigned from a follow-up in 2011, the genome-wide DNA methylation patterns were determined using the Illumina Infinium HumanMethylation450 BeadChip (Illumina).

2.2 Study designs
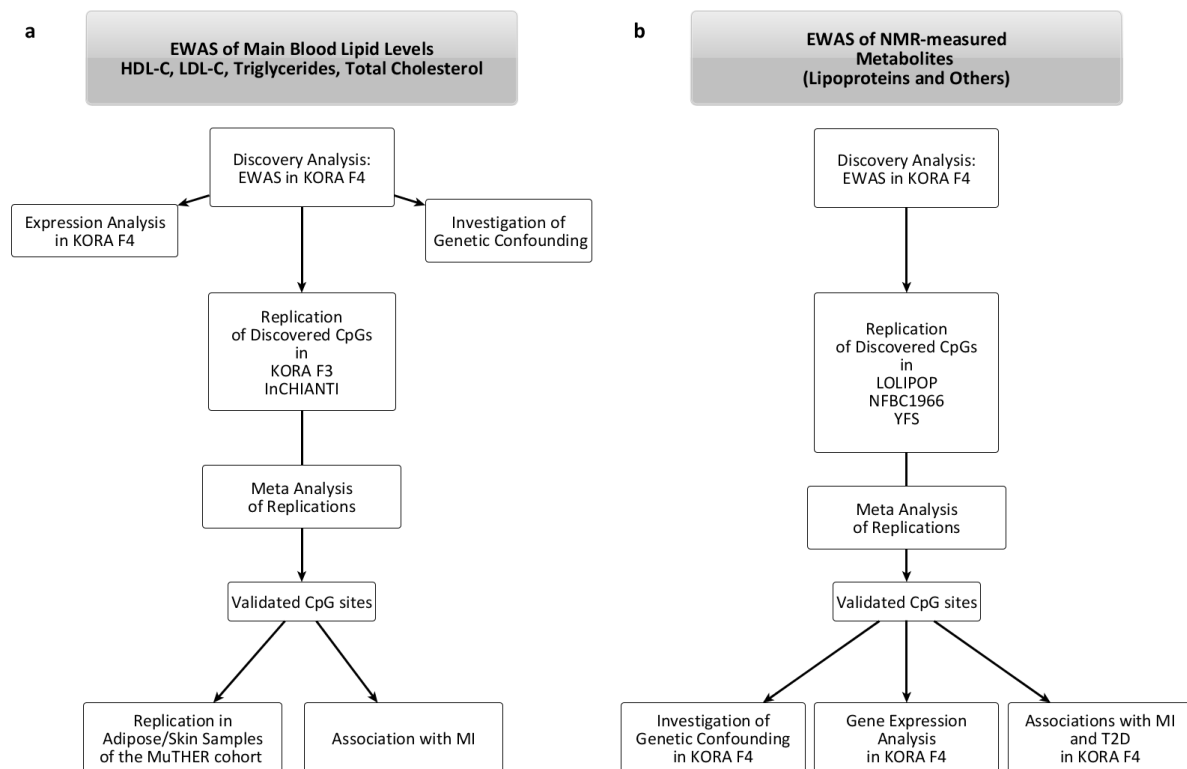
Figure 4: Study designs of the EWAS of main blood lipids (a) and lipoproteins (b). EWAS epigenome-wide association analysis; MI myocardial infarction; T2D type 2 diabetes.

For the epigenome-wide association study (EWAS) of main blood lipids, the discovery analysis was performed in the KORA F4 cohort. After meta-analysis of replication results of KORA F3 and InCHIANTI, confirmed associations were also tested in methylation datasets of human adipose and skin tissue of the MuTHER cohort. In addition, the associations of lipid-associated CpG sites with gene expression and genetic confounding as well as the relationship to myocardial infarction in KORA F4 were analyzed (**Figure 4**).

To get a detailed insight into the associations between lipoprotein subfractions and DNA methylation, an EWAS of NMR-measured metabolites was performed. The results of the discovery cohort KORA F4 were replicated in the LOLIPOP, NFBC1966 and YFS cohorts. After

meta-analysis of the replication results, confirmed CpG sites were further analyzed to explore genetic confounding of the associations and the relationship to gene expression as well as to lipid-related diseases such as myocardial infarction and type 2 diabetes. **Figure 4** gives an overview of the study designs of both EWAS.

For the studies, the datasets of the main blood lipids, the ^{1}H NMR-measured metabolites (KORA F4), the genome-wide DNA methylation data (KORA F3/F4), the gene expression data, the genotype data as well as data of different phenotypes were provided by the KORA meta-database for phenotypes/variables and genetics/omics (KORAgen). Detailed descriptions of the applied methods for measurements are provided in the following chapters.

2.3 Assessment of lipid levels in KORA F4

2.3.1 Determination of main blood lipid levels using enzymatic methods

In KORA F4 lipid levels were determined in fasting fresh blood samples, in KORA F3 also non-fasting samples were included. Measurements were performed in the Medical Center Augsburg at the latest six hours after blood collection. HDL-C and LDL-C were determined enzymatically with the CHOP-PAP (cholesterol oxidase-peroxidase-4-aminophena-zone) assay after selective release of HDL-C or dissolution of all non-LDL cholesterol, respectively (AHDL Flex/ALDL Flex, Dade-Behring, Germany). Total triglyceride levels were measured with the GPO-PAP (glycerophosphate oxidase-peroxidase-4-aminophena-zone) assay (TGL Flex, Dade-Behring, Germany). Total cholesterol measurement was performed with an enzymatic method which uses cholesterol esterase, cholesterol oxidase and peroxidase (CHOL Flex, Dade-Behring, Germany).

2.3.2 Metabolomics measurement using ^{1}H NMR spectroscopy

2.3.2.1 Laboratory process

For determination of advanced lipoprotein profile in KORA F4 a high-throughput proton (^{1}H) nuclear magnetic resonance (NMR) metabolomics platform was applied (Brainshake´s Biomarker Analysis Platform; http://www.brainshake.fi). This methodology can provide

information about multiple serum measures such as lipoprotein subclasses, lipoprotein particle concentration, low molecular weight metabolites and serum lipid extracts. The methodology is based on three molecular windows to improve the detection of metabolites from different molecular weight classes. Two of the windows, the lipoprotein lipids (LIPO) and the low-molecular-weight molecules (LMWM) window, were applied on serum providing quantification of lipoprotein subclass distribution and composition as well as of low-molecular-weight metabolites including amino acids, ketone bodies and carbohydrate metabolites. The third window, the LIPID window, was applied on serum lipid extracts and gave information about serum lipid constituents including fatty acids, cholesterol and sphingomyelins (Tukiainen *et al.* 2008). The precise experimental methodology as well as the computational strategies of metabolite identification and quantification from the NMR spectra has been described elsewhere (Tukiainen *et al.* 2008; Soininen *et al.* 2009; Inouye *et al.* 2010; Soininen *et al.* 2015).

In KORA F4 a total of 228 metabolite concentrations and derived measures were obtained from the NMR spectroscopy platform in N=1,790 participants. Measurements were conducted by the cooperation partners of the Faculty of Medicine, Computational Medicine, University of Oulu, Oulu, Finland. A detailed list of measured traits in KORA F4 is provided in the **Appendix Table I.** The same metabolomics platform was applied to measure the metabolites in the replication cohorts LOLIPOP, NFBC1966 and YFS.

2.3.2.2 Data preprocessing

Data preprocessing of the KORA F4 data was done by a statistician of the doctoral candidate´s research group, data of the replication cohorts LOLIPOP, NFBC1966 and YFS were preprocessed by the respective cooperation partners.

Sample size

For 1,788 KORA F4 subjects metabolite data were obtained. In 36 subjects the metabolites showed detection rates of less than 95% over all metabolites and were eliminated from further analysis. Eight individuals were eliminated due to non-fasting status at the time of blood sampling and 82 due to lack of valid methylation data, leaving a final sample size of 1,662 subjects.

Metabolite Data Preprocessing

Using the ^{1}H NMR metabolomics platform, 228 serum metabolites were measured. Of these, two metabolites (glycine and glycerol) had missing values or zero values in greater than 10% of the samples; these metabolites were considered untrustworthy and removed from further analysis. The natural logarithm was taken of the metabolite concentrations to achieve more normal distributions. Prior to taking the logarithm the zero values were set to the lowest non-zero value seen for that metabolite divided by two. Metabolite outliers were defined as those concentrations further than 5 standard deviations from the mean for that metabolite. These values were set to NA (not available).

For the initial epigenome-wide association discovery in KORA F4, imputation was performed for all missing values for the methylation, metabolite and covariate data. For the methylation data mean imputation was performed automatically by the omicABEL software. For the missing values for the metabolite and covariate data, multiple imputation by chained equations was performed to create 10 complete covariate and metabolite datasets. This was done using the R-package mice (Buuren and Groothuis-Oudshoorn 2011).

2.4 Determination of genome-wide DNA methylation pattern in KORA F4

Genome-wide DNA methylation measurement was performed using the Infinium HumanMethylation450 BeadChip® (Illumina, USA). This array covers 99% of RefSeq genes and 96% of CpG islands with multiple sites in CpG islands, shores (2 kb flanking the CpG island) and shelves (2 kb flanking the shores) as well as island-independent CpG sites (Bibikova *et al.* 2011).

In KORA F4, data about genome-wide DNA methylation were available for N=1,802 participants, in KORA F3 for N=500 participants (comprising smokers and never smokers (Zeilinger *et al.* 2013)).

2.4.1 Laboratory process of the Infinium HumanMethylation450 BeadChip

For sample preparation whole blood genomic DNA was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research, USA) according to the manufacturer´s procedure with the alternative incubation conditions recommended when using the Illumina Infinium

Methylation Assay. Bisulfite converted samples were then subjected to whole genome amplification, followed by enzymatic fragmentation and hybridization to the BeadChip. After a single-base extension and staining process, the BeadChip was scanned with the Illumina HiScan SQ scanner. The laboratory process for the array has been described in detail elsewhere (Zeilinger *et al.* 2013; Petersen *et al.* 2014).

2.4.2 Preprocessing of the genome-wide DNA methylation data

The raw methylation data were extracted using the methylation module (version 1.9.0) of the Illumina Genome Studio (version 2011.1). A methylated and an unmethylated signal count per CpG sites are obtained. Counts are commonly combined to β-values, defined as the ratio of the methylated signal intensity to the overall signal intensity (Du *et al.* 2010; Bibikova *et al.* 2011). Preprocessing of the data was done by a statistician of the doctoral candidate´s research group who used the R software (version 3.0.1).

For the EWAS of main blood lipid levels, probes with signals from less than three functional beads, and probes with a detection p-value>0.01 were defined as low-confidence probes. Data from probes that covered SNPs with a minor allele frequency in Europeans >5% were excluded from the data set. A color bias adjustment was performed with the R package lumi (version 2.12.0) by smooth quantile normalization and background correction based on negative control probes present on the Infinium HumanMethylation BeadChip. This was performed separately for the two color channels and the chips. β-values corresponding to low-confidence probes were set to missing. A 95% call rate threshold was applied on samples and CpG sites. The data were normalized by beta-mixture quantile normalization (BMIQ) using the R package wateRmelon, version 1.0.3. Plate and batch effects were investigated by principle component analysis and eigenR2 analysis, since KORA F4 samples were processed on 20 96-well plates and measured in 9 batches.

For the EWAS of NMR-measured metabolites, normalization of the methylation data was conducted according to the pipeline of Lehne *et al*. (Lehne *et al.* 2015), beginning with exclusion of 65 SNP markers and background correction using the minfi R package. Probes were set to NA if the detection p-value was ≥0.01 or the number of beads was ≤3. Samples were excluded if the detection rate was ≤0.95. Quantile normalization was performed on intensity values separated by color channel, probe type and M/U subtypes. The resulting

methylated and unmethylated signals were then used to calculate the CpG site β-values, a measure of percent methylation between 0 and 1. To eliminate technical effects, the principal components of the methylation control probes were calculated on all green and red intensities of all (non-negative, autosomal) control probes together. For each CpG site, the residuals of a linear regression model with β-value as outcome and the first 20 control probe PCs as covariates were used as the "technically adjusted β-values" in the analysis. Technically adjusted β-values that showed more than 5 standard deviations from the mean for each CpG site were defined as "methylation outliers". Outliers were replaced with NA values. Finally, those CpG sites with NA for more than 20% of the samples were eliminated from all further analysis, as were those sites annotated to the X or Y chromosomes, leaving a total of 468,151 CpG sites, with 0.7% missing values total.

2.5   Gene expression profiling in KORA F4

In KORA F4 gene expression measurement was performed using the Illumina Human HT-12 v3 Expression BeadChip as described elsewhere (Schurmann *et al.* 2012). Briefly, total RNA was extracted from whole blood fasting samples using the PAXgene™ Blood miRNA Kit (Quiagen, Germany). Purity and integrity of the RNA was assessed on the Agilent Bioanalyzer with the 6000 Nano LabChip reagent set (Agilent Technologies, Germany). Using the Illumina TotalPrep-96 RNA Amp Kit (Ambion, Darmstadt, Germany), 500 ng of RNA were reverse transcribed into cRNA and thereby labelled with biotin-UTP. A total of 3 µg of cRNA was hybridized to the Illumina Human HT-12 v3 Expression BeadChip, followed by washing steps as described in the Illumina protocol. BeadChips were scanned using the Illumina Bead Array Reader.

Sample quality control and imputation of missing values was performed with the Illumina Software Genome Studio, version 2010.1. Data were quantile normalized (Bolstad *et al.* 2003) using the R package lumi, version 2.8.0 from the Bioconductor open source software (http://www.bioconductor.org) (Du *et al.* 2008). For the KORA data, sample storage time between blood sampling and RNA isolation, the RNA integrity number and the assignment to one of 30 amplification plates were important technical aspects which explain a large proportion of variability in the gene expression data (Schurmann *et al.* 2012).

2.6   Genotyping of single nucleotide polymorphisms in KORA F4

KORA F4 samples were genotyped using the Affymetrix Axiom chip array. Genotypes were called with the Affymetrix software and were annotated to NCBI build 37.

To ensure data quality samples with a missmatch of phenotypic and genetic gender were removed. In addition samples which represented population outliers and/or heterozygosity outliers were excluded from the data set. Furthermore, individuals were excluded with a call rate <97% to ensure data reliability. Similarly a callrate threshold of 98% was applied for SNPs. Another important quality measure is the Hardy-Weinberg equilibrium (HWE) which states that allele and genotype frequencies remain constant across the generations of a population under certain conditions and that therefore a fix relationship between allele and genotype frequencies should exist (Ziegler and König 2010). SNPs with HWE violation ($p<5E-06$) were also excluded from the dataset as well as SNPs with a minor allele frequency below 1%. Genotype imputation was performed based on the 1000G phase 1 reference panel using IMPUTE v2.3.0, with SHAPEIT v2 as a pre-phasing tool.

2.7   Statistical methods

2.7.1   Epigenome-wide association study of main lipid levels

All statistical analyses including the EWAS in KORA F4, the replication in KORA F3, the meta-analysis of replication results, the expression analyses and the association analysis with myocardial infarction as well as the SNP analysis were done by the doctoral candidate. The statistical methods are also described in Pfeiffer *et al.* (Pfeiffer *et al.* 2015).

2.7.1.1   Analysis in the discovery cohort KORA F4

Associations between DNA methylation β-values and lipid levels (HDL-C, LDL-C, TG and total cholesterol) were analyzed using linear mixed effects models (nlme package in R) with lipid levels as response variable. In KORA F4, 26 of 1,802 individuals were excluded from the analysis due to missing information in covariates or due to non-fasting status at the time point of blood collection, resulting in a final sample size of 1,776 KORA F4 subjects. To normalize lipid levels, logarithmic (HDL-C and TG) and square root (total cholesterol and

LDL-C) transformations were applied, followed by standardization to a mean of zero and a standard deviation of 1. The covariates age, sex, body mass index (BMI), smoking status, intake of lipid lowering drugs, physical activity, alcohol consumption, current hypertension, history of myocardial infarction, hemoglobin A1c levels (HbA1c), C-reactive protein levels, and white blood cell count were included in the statistical models as potential confounders. The experimental plate number was considered as a random effect. The Bonferroni procedure was applied to correct for multiple comparisons, resulting in a genome-wide significance level of 0.05/441552=1.1E-07. Because whole blood DNA samples consist of a mixture of cells with different DNA methylation patterns, cell heterogeneity had also to be considered as a confounder. As no measured cell count information was available in KORA F4, a statistical method described by Houseman *et al.* was applied to obtain sample-specific estimates of the proportion of the major white blood cell types (Houseman *et al.* 2012). Significant associations of the first statistical model were recalculated, additionally adjusting for the estimated white blood cell proportions (CD4$^+$ T cells, CD8$^+$ T cells, B-lymphocytes, monocytes, granulocytes, and natural killer cells). R2 statistics were calculated (Edwards *et al.* 2008) to get a measure of the variance in the lipid levels explained by methylation levels. This was done using the R package pbkrtest, version 0.3–7.

2.7.1.2 Replication of identified associations

CpG sites which were significant associated with blood lipids in KORA F4 were replicated in KORA F3 (N=499) and InCHIANTI (N=472) using the same statistical model. InCHIANTI data were analyzed by the cooperation partners.

In KORA F3, no data about the C-reactive protein were available and therefore an adjustment for C-reactive protein was not possible. A fixed-effects meta-analysis of KORA F3 and InCHIANTI results was done with the R package metafor, version 1. 9–2. All results were corrected according to the Bonferroni procedure (level of significance 0.05/11=4.5E-03). Results were also replicated by the cooperation partners of the MuTHER (The Multiple Tissue Human Expression Resource) cohort. The Infinium HumanMethylation450 BeadChip Array signal intensities were quantile normalized and methylation β-values were calculated using the R software version 2.12 as previously described (Grundberg *et al.* 2013). For the final analyses data of N=634 human adipose and N=395 human skin samples were available.

There were no DNA methylation data available for the CpG site cg06500161 in the final data set, as the data did not pass the quality control filters of the cohort. A linear mixed effects model was conducted using the lme4 package in R. The model was adjusted for age, body mass index, smoking status, statin intake, technical covariates (fixed effects), and family relationship and zygosity (random effects). To assess significance, a likelihood ratio test was used, and the p-values were calculated from the χ2 distribution with one degree of freedom using −2 log (likelihood ratio) as the test statistic. Results were corrected according to the Bonferroni procedure (level of significance: 0.05/7=7.14E-03).

2.7.1.3   Association of DNA methylation with gene expression

To test whether DNA methylation of identified lipid-associated CpG sites were also associated with gene expression, a gene expression analysis was conducted. For this analysis only N=724 of the original N=1,776 KORA F4 subjects were included, as only for this number of subjects both DNA methylation and gene expression data were available. Using the data of these N=724 KORA F4 subjects, the association between lipid level and DNA methylation was recalculated for each significantly associated lipid-methylation pair. Next, the association between DNA methylation of the CpG site and transcripts of the gene (which is annotated to the CpG site) as well as between transcripts and the respective lipid level were determined (except for the CpG site cg07504977 which has no annotation to a gene according to the University of California Santa Cruz (UCSC) Genome Browser; **Appendix Table II**). To investigate the role of transcripts, the associations between DNA methylation and lipids was recalculated additionally adjusting for the transcripts of the annotated gene. A likelihood ratio test was done to determine the p-value for the association. The association between DNA methylation and transcript levels, as well as between lipid levels and transcript levels were recalculated. All models were also adjusted for age, sex, body mass index, alcohol consumption, intake of lipid lowering drugs, physical activity, smoking, current hypertension, history of myocardial infarction, hemoglobin A1c levels, and C-reactive protein levels as well as for white blood cell count and estimated white blood cell proportions. Models including expression data were additionally adjusted for technical parameters such as the sample storage time, RNA integrity number, and RNA amplification batch (Schurmann *et al.* 2012). The level of significance was set to 8.3E-04.

2.7.1.4   Investigation of the genetic background of the observed associations

It was investigated whether the observed associations between lipid and DNA methylation in KORA F4 were caused by lipid-associated SNPs. For this 157 lipid-associated SNPs identified by the Global Lipids Genetics Consortium were included in the analysis (Global Lipids Genetics *et al.* 2013). In KORA F4, no genotype data were available for the SNP rs9411489. Data of the 156 lipid-associated SNPs were available for N=1,710 KORA F4 participants. In a first step, a preselection was made to reveal the lipid-associated SNPs which were at the same time nominally associated (p-value<0.05) with differentially methylated lipid-associated CpG sites. In a second step, models for each significant CpG–lipid pair were recalculated with additional adjustment for the respective preselected SNPs to see if the association stays significant. An increase of the p-value indicates SNP confounding. Level of significance was the same as used for the EWAS of main blood lipid levels in KORA F4 (1.1E-07).

2.7.1.5   Association of DNA methylation with prevalent myocardial infarction

To assess the association of the observed lipid-associated CpG sites with previous hospitalized myocardial infarction in KORA F4, generalized linear mixed effects models were implemented with adaptive Gauss-Hermite quadrature (R package lme4, version 1.0–4). The following potential confounders were included as covariates: age, sex, body mass index, physical activity, smoking status, alcohol consumption, current hypertension, hemoglobin A1c levels, C-reactive protein levels, lipid variables (HDL-C, LDL-C, triglycerides, and total cholesterol), white blood cell count and estimated white blood cell proportions. Results were corrected according to the Bonferroni procedure (level of significance: 0.05/8=6.3E-03). The same analysis was done in the KORA F3 and InCHIANTI cohort.

2.7.2   Epigenome-wide association study of NMR-measured metabolites

The statistical analyses were performed in collaboration with a statistician of the doctoral candidate´s research group.

2.7.2.1 Analysis in the discovery cohort KORA F4

In KORA F4 (N=1,662), associations between DNA methylation (β-values) and NMR-measured metabolites were analyzed in 226 epigenome-wide association studies, one per metabolite. Specifically, for each metabolite, 468,151 linear regression models were examined, one per CpG site. Each model used the (natural) logarithm of the metabolite as the dependent variable, and the technically adjusted β-values and covariates as explanatory variables. The following covariates were used: age, sex, BMI, C-reactive protein levels, hemoglobin A1c levels, smoking status (current smoker, ex-smoker or never smoker), intake of lipid-lowering drugs, current hypertension, history of myocardial infarction, level of physical activity (high/low), white blood cell count and estimated proportions of white blood cell type (Houseman *et al.* 2012). The EWAS for the discovery stage were carried out using omicABEL (Fabregat-Traver *et al.* 2014), a software package allowing rapid (epi-) genome-wide association analyses. The beta version was used, in combination with DatABEL v.0.9-6. As there were 10 MICE-imputed datasets, each model was run 10 times and the results were combined using the method of (Li *et al.* 1991). The combining of the coefficients and test statistics of the resulting models was performed in the software R using the package miceadds version 1.7-8. A CpG site–metabolite pair was carried forward to further analysis if the p-value of the coefficient of the methylation variable was lower than a Bonferroni-corrected threshold of 4.73E-10, for a family-wise error rate of 0.05.

To investigate the validity of the regression models used in the discovery analysis the models of the significant CpG-trait pairs were re-run, but firstly replacing the 0 values for the metabolites with NAs and using complete case analysis on all data, rather than any imputation methods. A model was considered unstable if the coefficient for the methylation changed by 10% from its original (discovery analysis) value or if the p-value increased by a factor of 100, resulting in the methylation coefficient no longer being statistically significantly different to 0, based on the original definition of statistical significance. The models failing these criteria were eliminated from future analysis. Finally, the qq-plots of the residuals of the sensitivity models were examined to verify that model assumptions were approximately achieved. Those CpG site-metabolite pairs whose model residuals had a correlation of less than 0.98 with the expected distribution were excluded from further analysis.

Of the 282 CpG-metabolite pairs that were significantly associated in the discovery analysis, 16 models were considered unstable by the definition above, and 29 failed the residuals criterion, leaving a total of 237 models to be carried forward to the meta-analysis of the results of the replication partners.

2.7.2.2 Replication of identified associations

Significant CpG-trait pairs were replicated in three independent cohorts, LOLIPOP (N=2,805), NFBC1966 (N=771) and YFS (N=176), using the same statistical methods as in the discovery cohort. Analyses were performed by the cooperation partners. In LOLIPOP, an adjustment for intake of lipid-lowering drugs was not possible as this variable was not available. The same applied to the hemoglobin A1c levels variable in NFBC1966.

Afterwards, a random effects meta-analysis was performed on the replication results using the DerSimonian-Laird estimator and the rma command of the R package metaphor, version 1.9-8. Statistical significance for overall effect being different to 0 was achieved using a Bonferroni-corrected threshold of 0.05/237=2.1E-04, for a family-wise error rate of 0.05.

2.7.2.3 Association of DNA methylation with gene expression

To study the association between the identified CpG sites and gene expression, those gene transcripts lying within 1 Mb of the significant and successfully replicated CpG sites were examined. Analysis was done using N=695 KORA F4 participants whose gene expression data were also available. For each CpG site and each expression probe within the 1 Mb-region, a mixed-effects linear model was used with log2 of the expression value as outcome, technically adjusted β-value, RNA integrity number, sample storage time and all covariates from the discovery analysis as fixed effects, and RNA amplification batch as random effect. The analysis of the linear mixed models was done using the R package lme4 version 1.1-10. A total of 790 CpG site-expression probe pairs were examined (covering all 15 CpG sites and 346 expression probes), leading to a Bonferroni-corrected significance threshold of 6.33E-05, for a family-wise error rate of 0.05.

2.7.2.4 Investigation of the genetic background of the observed associations

The aim of this analysis was to investigate possible genetic effects on the associations between methylation and metabolite measurements established in the discovery analysis and replicated in the meta-analysis. N=1,598 of the 1,662 KORA F4 participants also had genotype data available.
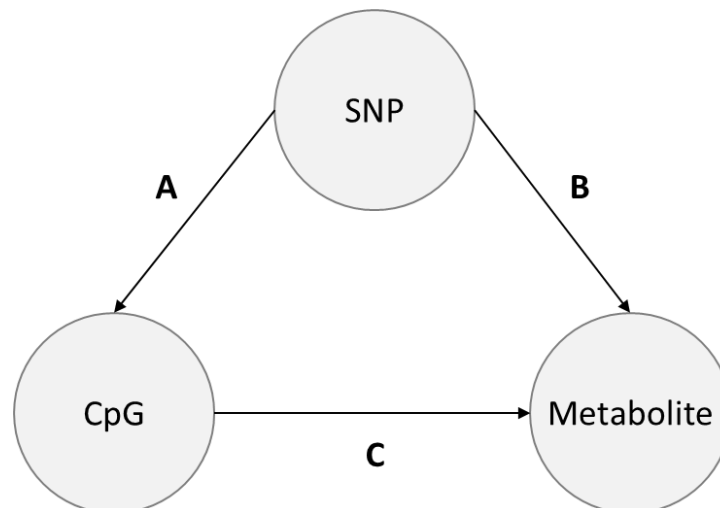
Figure 5: Investigation of possible genetic effects on the associations between methylation and metabolite measurements using conditional analysis. The effects of SNPs located up to 1 Mb away from the CpG sites were tested. A) First it was determined which SNPs were associated with each of the CpG sites. B) In a second step it was determined if these SNPs were in turn associated with the respective CpG-associated metabolites. C) If a SNP was found to be associated with both a CpG site and one of this site´s associated metabolites, it was explored whether this CpG-metabolite association was confounded by the SNP. For this the association between CpG and metabolite was re-calculated, including the SNP as covariate.

As shown in **Figure 5** the investigation was performed using conditional analysis: A) It was first determined which SNPs were associated with each of the CpG sites. Here, for each of the 15 significant CpG sites those SNPs within 1 Mb were examined. To determine if these SNPs and CpG sites were associated, for each pair a multiple regression model under the additive genetic assumption was run with methylation (technically adjusted β-values) as outcome and SNP data and all covariates as in the discovery analysis as the independent variables. Those pairs whose SNP coefficient achieved a p-value of less than 0.01 were carried forward to the next step of the analysis. B) In a second step it was determined if the SNPs of A) were in turn associated with the respective CpG-associated metabolites. Again,

for each pair a multiple regression additive model was run with metabolite value as outcome, and SNP data and all covariates as in the discovery analysis as the independent variables. Those pairs whose SNP coefficient achieved a p-value of less than 0.01 were carried forward to the next step of the analysis. C) If a SNP was found to be associated with both a CpG site and one of this site´s associated metabolites, it was explored whether this CpG-metabolite association was confounded by the SNP. For this the discovery multiple analysis regression model was re-run, but including the SNP as a potential confounder. It was then determined if the association lost statistical significance, according to the discovery analysis threshold (4.73E-10), with the addition of any single SNP to the model.

2.7.2.5 Association of DNA methylation with lipid-related diseases

To assess the association of the observed lipid-associated CpG sites with previous hospitalized myocardial infarction and prevalent type 2 diabetes in KORA F4, logistic regression models were performed with methylation as the independent variable and presence of the outcome as dependent. Two models were used for each outcome. The first model included the same potential confounders as in the discovery analysis (see **chapter 2.7.2.1**), with the exception that it was not adjusted for intake of lipid-lowering drugs. Clearly, the myocardial infarction models were not further adjusted for myocardial infarction, and the type 2 diabetes models were not adjusted for hemoglobin A1c levels, as this was considered a surrogate for diabetes. The second model included the additional adjustment for the intake of lipid-lowering drugs. The Bonferroni correction was used with a significance level of 0.05/(2x15)=1.7E-03.

2.7.2.6 CpG-by-sex interaction analysis

To assess whether the identified and validated associations (N=118) between methylation of CpG sites and metabolites were sex-specific, a CpG-by-sex interaction analysis was performed. For all CpG-metabolite pairs a linear regression was performed with methylation as the independent variable and logarithm of the metabolite as the dependent variable, and sex*CpG as an additional independent variable. It was adjusted for the same covariates as used in the discovery analysis (see **chapter 2.7.2.1**). P-values of the interaction term were corrected according to the Bonferroni procedure (0.05/118=4.2E-04). A significant p-value

would imply that the association between methylation and metabolite of that specific pair is different for males as it is for females.

2.8 Functional analysis of methylation-dependent protein-DNA interaction

To assess whether methylation of the CpG sites found to be associated with lipid levels has an influence on protein-DNA binding affinity, electrophoretic mobility shift assays were carried out for the CpG sites cg06500161 (*ABCG1*) and cg20544516 (*MIR33B/SREBF1*). The CpG site cg06500161 was chosen because it showed the strongest association with both TG and HDL-C levels in the EWAS of main blood lipid levels. The CpG site cg20544516 was selected because of its functionally interesting location in the *SREBF1* gene in a region coding for a microRNA (miR-33b).

2.8.1 Preparation of nuclear protein extracts

Nuclear extracts of THP-1 cells were prepared using the Nuclear Extraction Kit (Active Motif) according to the manufacturer´s protocol. Protein concentration was measured with the BCA Protein Assay Reagent (see **2.8.2**). Nuclear extracts were stored in 100 µl aliquots at -80°C.

2.8.2 Determination of total protein concentration

Total protein concentration of nuclear protein extracts were determined using the Pierce™ BCA Protein Assay Kit (Thermo Scientific) and the UV-VIS Spectrophotometer DU530 (Beckman). The assay was performed according to the manufacturer´s standard protocol. First, diluted bovine serum albumin standards were prepared. For all samples 50 µl of 1:10 and 1:20 dilutions were prepared and 50 µl of the samples and of each of the standards of the dilution series were mixed with 1 ml freshly prepared Working Reagent. After incubation for 30 min at 37°C, the absorbance of all standards and samples was measured at 562 nm within 10 min. The concentration of the samples was calculated using the linear calibration line.

2.8.3 Preparation of probes for electrophoretic mobility shift assays

Oligonucleotides with the consensus sequence for Oct1 and Cy5-labeled and unlabeled oligonucleotides containing the methylated and unmethylated cytosine on the loci cg06500161/cg20544516 were obtained from Metabion (Germany). A detailed list of the oligonucleotides can be found in the **Appendix Table III**. To obtain double-stranded DNA probes the oligonucleotides were annealed mixing 7 µl of both, forward and reverse oligonucleotides (100 pmol/µl) and 6 µl 1X TE. Samples were incubated at 95°C for 2 min, followed by 10 min at 76°C. Annealed probes were left at RT overnight to cool down. On the next day probes were purified by 12% polyacrylamide gel electrophoresis (240 Volt, 4 hours). Probes were cut out of the gel using a UV lamp, gel pieces were shredded with a pipette tip in a tube and incubated with 200 µl 1X TE overnight at 37°C in a shaker (1000 rpm). Afterwards, probes were transferred to a filter of a cut filter pipette tip and centrifuged (16 000 x g, 5 min). Concentration of the annealed probes was determined using the Nanodrop (see **2.9.4.1**). Labeled probes were diluted to 1 ng/µl, unlabeled competitor probes to 40 ng/µl.

2.8.4 Electrophoretic mobility shift assay

To investigate methylation-specific protein-binding, electrophoretic mobility shift assays (EMSAs) were performed. The CpG site cg06500161 was analyzed using self-prepared THP-1 nuclear extracts and for EMSAs of cg20544516 commercial THP-1 nuclear extract (Active Motif) was used. For the binding reaction, 5 µg of nuclear extract, 1 ng of labeled probe and either no or rising concentrations (10 and 100 ng) of unlabeled competitor probes were incubated in binding buffer (4% v/v glycerol, 1 mM MgCl2, 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl, 10 mM TrisHCl pH 7.5) with 0.05 mg poly(dI-dC) (Roche Diagnostics) in a total volume of 10 µl for 20 min at 4°C. DNA-protein complexes were mixed with 10X gel loading dye (Orange G) and separated by electrophoresis using a 5.3% polyacrylamide gel and a 0.5X TBE running buffer at 200 V for 4 hours at 4°C. The gels were scanned using the Typhoon Trio+ Imager (GE Healthcare) with the setting of Acquisition mode: Fluorescence; Emission Filter: 670 BP 30 CY5; PMT: 650; Laser: Red (633 nm); Pixel size: 100 microns. Experiments were performed at least twice.

2.9 Functional analysis of methylation-dependent gene expression

To investigate whether the methylation status of the CpG sites cg06500161 (*ABCG1*) and cg20544516 (*MIR33B/SREBF1*) as well as the methylation of the surrounding CpG sites causes alterations in the transcriptional activity, Luciferase Reporter Assays were performed.

2.9.1 Experimental design

For Luciferase Reporter Assays, DNA fragments of different sizes containing the CpG site of interest (cg06500161 or cg20544516) as well as different numbers of additional adjacent CpG sites were cloned as inserts into the pCpGfree-promoter-Lucia plasmid (Invivogen) (see **2.9.4.5**). The sequences of the different DNA fragments are provided in the **Appendix Table IV**. Also the respective reverse complementary (experiments for cg06500161) or the reverse (experiments for cg20544516) DNA fragments were cloned into the pCpGfree-promoter-Lucia plasmid to act as controls for the experiments. **Table 3** provides an overview of the created constructs and their names which are used in the following chapters.

Table 3: pCpGfree-promoter-Lucia plasmid constructs for the Luciferase Reporter Assays

Name of the construct	CpG site of interest	Length of inserts	Number of CpG sites of the construct
Empty vector	-	-	0
*ABCG1*_104 bp	cg06500161	104 bp	2
*ABCG1*_104 bp_rc	cg06500161	104 bp	2
*ABCG1*_518 bp	cg06500161	518 bp	7
*ABCG1*_518 bp_rc	cg06500161	518 bp	7
*ABCG1*_1607 bp	cg06500161	1607 bp	29
*ABCG1*_1607 bp_rc	cg06500161	1607 bp	29
*MIR33B/SREBF1*_24 bp	cg20544516	24 bp	1
*MIR33B/SREBF1*_24 bp_r	cg20544516	24 bp	5
*MIR33B/SREBF1*_32 bp	cg20544516	32 bp	3
*MIR33B/SREBF1*_32 bp_r	cg20544516	32 bp	6

_rc = reverse complementary sequence of insert; _r = reverse sequence of insert

The pCpGfree-promoter-Lucia plasmid constructs containing the different inserts were then methylated using the methyltransferase M.SssI (see **2.9.4.6**). For the experiments regarding cg06500161, the constructs containing the two shorter inserts (*ABCG1*_104 bp and *ABCG1*_518 bp) were also treated with methyltransferase HpaII, which should specifically methylate the CpG site of interest. HpaII was not suitable for the third construct (*ABCG1*_1607 bp) as this construct carries the enzyme-specific sequence (5`…CCGG…3´) more than once. After methylation, the constructs were transfected into three different cell lines (THP-1, HepG2 and HEK293 as control cell line; see **2.9.3.3**) and luciferase activity was determined measuring samples in triplicates (=technical replicates). Experiments were replicated at least three times (=biological replicates; see **2.9.3.4**).

A detailed scheme of the experimental design of the Luciferase Reporter Assays is shown in **Figure 6**. The reporter assay on cg20544516 was performed by a student as part of a Master´s Thesis which was supervised by the doctoral candidate.
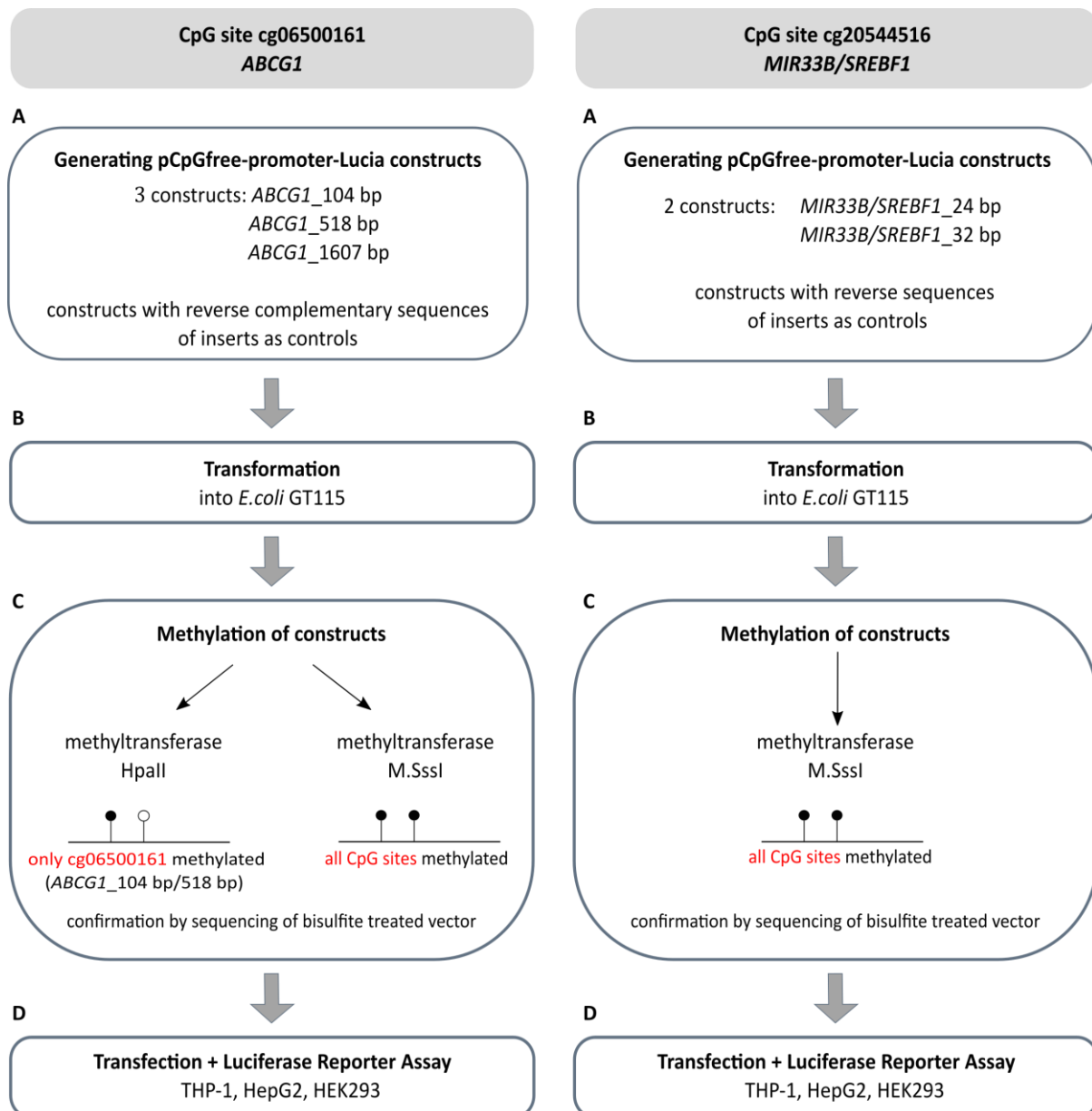
Figure 6: Experimental design of the Luciferase Reporter Assays for the CpG sites cg06500161 and cg20544516. Constructs were created by cloning DNA fragments of different lengths into the pCpGfree-promoter-Lucia plasmid. Constructs carrying either the respective reverse complementary inserts or the reverse inserts were used as controls (**A**). After transformation into *E.coli* (**B**), constructs were methylated (**C**). DNA was then transfected into three different cell lines and luciferase signals were measured 30 hours post-transfection (**D**). ⏄methylated CpG site ⏁unmethylated CpG site

2.9.2 Bacteria protocols

2.9.2.1 Liquid culture

Liquid culture of bacteria was performed in LB_0 medium containing the appropriate selection antibiotic depending on the resistance gene carried by the transformed plasmid.

The culture was incubated overnight at 37°C in a shaking incubator (200 rpm). For a PCR screen (see **2.9.5.2**), colonies picked from an agar plate were cultivated in LB_0 medium with the appropriate antibiotic for two hours at 37°C and 1000 rpm (thermomixer comfort, Eppendorf).

2.9.2.2 Long term storage

For long term storage glycerol stocks of the clones were prepared by mixing 900 µl of a liquid bacteria culture with 900 µl of 80% sterile glycerol in a cryogenic tube. A glycerol stock can be stored at -80°C for several years.

2.9.2.3 Production of chemically competent *E.coli*

To facilitate the entry of plasmid DNA into the bacterial cell chemically competent cells were produced. Bacteria were plated onto an agar plate and incubated overnight at 37°C. Colonies picked from the agar plate were incubated in 150 µl LB_0 medium for 3 hours at 37°C and 1000 rpm (thermomixer comfort). 20 µl of this culture were then added to 6 ml LB_0 medium and incubated overnight at 39°C and 175 rpm (thermomixer comfort). 500 µl of the overnight culture were cultivated under vigorous shaking in 100 ml LB_0 medium at 37°C to an OD_{600} of 0.6-0.8. Bacteria were put on ice for 15 min and were then centrifuged at 4000 x g for 10 min at 4 °C. The pellet was resuspended in 15 ml ice-cold TFB1 buffer pH 5.8 (30 mM K-Acetate, 100 mM RbCl, 10 mM CaCl2, 50 mM MnCl2, 15% v/v Glycerol). Afterwards the cells were pelleted by centrifugation again. After addition of 4 ml TFB2 buffer pH 6.5 (10 mM MOPS, 75 mM CaCl2, 10 mM RbCl, 15% v/v Glycerol) and incubation for 15 min on ice, 100 µl aliquots were prepared and stored at -80°C.

2.9.2.4 Transformation of competent *E.coli*

For transformation of ligation products into *E.coli*, 8 µl of the ligation reaction (see **2.9.4.5**) were added to 100 µl competent *E.coli* DH5α (transformation of pGL4.53 plasmid) or *E.coli* GT115 cells (transformation of pCpGfree-promoter-Lucia plasmid), respectively, and incubated on ice for 20 min. Afterwards the cells were heat-shocked for 45 seconds in a 42°C water bath and put back on ice for two min. 150 µl LB_0 medium were added and the cells were incubated at 37°C for one hour in a shaker (1000 rpm). After incubation two

aliquots (50 µl and 100 µl) of the cells were plated onto agar plates containing the appropriate selective antibiotic. The plates were incubated overnight at 37°C.

2.9.3 Cell culture protocols

2.9.3.1 Cultivation

Cells were cultivated under sterile conditions at 37°C and 5% CO_2 in a humidified atmosphere. THP-1 cells were cultivated in suspension in RPMI-1640 medium supplemented with 10% FBS and 2 mM L-glutamine and adhesive HepG2 and HEK293 cells in D-MEM High Glucose (4.5 g/L) medium also containing 10% FBS and 2 mM L-glutamine.

2.9.3.2 Thawing, splitting and freezing of cells

Cryopreserved cells were thawed quickly in a 37°C water bath and diluted in 10 ml fresh medium. After centrifugation (5 min, 1000 x g), the cell pellet was resuspended in 6 ml fresh medium and incubated in a T25 cell culture flask overnight. At a confluence of 80-100% cells were transferred into a 75 cm^2 cell culture flask (T75) for further cultivation.

Adhesive cells were split at a confluence of 80–100%. Before splitting, the adherent HepG2 and HEK293 cells were washed with 1X PBS and trypsinized with 2 ml trypsin-EDTA. After a short incubation at 37°C, the peptidase activity of the trypsin was stopped by adding a desired amount of fresh medium. After resuspension of the cells, an adequate amount of the cell suspension - depending on the splitting ratio - was added into a new T75 flask. Splitting ratio was typically 1:3–1:4. The suspension cell line THP-1 was split every second day. Cells were resuspended and split in a ratio of 1:2.

To freeze the cells for storage, cells were counted using the automated cell counter (Cellometer Auto T4, Nexcelom). 1x10^6 cells were suspended in appropriate freeze medium which contained 20% FBS and 10% DMSO, transferred to cryogenic tubes and cooled down to -80°C in CoolCell® Cell Freezing Container (Biocision). For long term storage cells were transferred to a liquid nitrogen tank.

2.9.3.3 Transfection of eukaryotic cell lines

To transfect the plasmid into cells, the lipid-based ViaFect™ Transfection Reagent (Promega) was used. To achieve optimal transfection efficiency, cells should be approximately 75% confluent on the day of transfection. Therefore, THP-1 cells were seeded at a density of $1x10^5$/well in a 12-well plate with 1 ml RPMI-1640 medium supplemented with L-glutamine and 10% FBS. HepG2 and HEK293 cells were seeded at a density of $2x10^5$ cells/well in 12-well plates in 1 ml DMEM medium containing stable L-glutamine and 10% FBS. For cell counting, the automated cell counter (Cellometer Auto T4, Nexcelom) was used. After 18 hours in culture, the medium was replaced by 1 ml of fresh medium.

For transfection, the optimal ViaFect™ Transfection Reagent:DNA (µl reagent: µg DNA) ratio was found to be 3:1. In addition to the pCpGfree-promoter-Lucia plasmid constructs (test plasmids), which express Lucia luciferase, the pGL4.53 plasmid (Promega) which constitutively expresses firefly luciferase was co-transfected for normalization. 750 ng of test plasmid and 50 ng of control plasmid (pGL4.53) were mixed with medium before adding 2.25 µl ViaFect. The total volume of transfection complex was 100 µl per well. The ViaFect™ Transfection Reagent:DNA mixture was incubated for 20 min at RT and then added dropwise to the cells. Transfected cell were incubated for 31 hours (at 37°C and 5% CO_2 in a humidified atmosphere).

2.9.3.4 Dual Luciferase Reporter Assay

Dual Luciferase Reporter Assays were conducted in three different cell lines, THP-1, HepG2 and HEK293 cells, which had been transfected as described in **chapter 2.9.3.3**. Luciferase activity was measured 30 hours post transfection. For this purpose, the transfected cells were washed in PBS buffer once before lysis in 200 µl of 1X passive lysis buffer provided by the Dual Luciferase Reporter Assay System (Promega). After shaking for 30 min at room temperature, the lysates were frozen at -80°C overnight.

20 µl of the samples were transferred to a 96-well plate and luciferase activity was measured with the GloMax®-Multi Detection System (Promega) using the Dual Luciferase Reporter Assay System according to the manufacturer`s instructions. Briefly, 50 µl Luciferase Assay Reagent II were added to the samples initiating the firefly luciferase reporter assay. Adding another 50 µl of Stop & Glo® Reagent the firefly luciferase luminescence is quenched

and the Lucia luciferase is activated. For technical replicates, samples were measured in triplicates. Experiments were replicated at least three times (=biological replicates). Lucia luciferase activity was normalized against the firefly activity to get the relative promoter activity of the constructs. To display the overall effect of the tested sequence on the promoter, activity of the different pCpGfree-Lucia-plasmid constructs was set out against the pCpGfree-Lucia-plasmid without any inserts (empty vector). For data evaluation, data were first tested for normal distribution by Kolmogorov-Smirnov-tests; no significant deviation from normal distribution was shown. To identify statistically significant differences in promoter activities between the constructs, linear mixed effects models (LME) with random intercept were used. LME models are the appropriate statistical method as they consider the correlation among the technical replicates, which is not assured when using the Student´s test or comparable statistical models. Results were analyzed using R, version 3.0.2.

2.9.4 DNA-based methods

2.9.4.1 Quantification of nucleic acids by measuring optical density

Nucleic acids were quantified by measuring the optical density at $\lambda=260$ nm using a NanoDrop 8000. To verify the sample quality, the ratios OD_{260nm}/OD_{280nm} and OD_{260nm}/OD_{230nm} were determined. The first ratio (OD_{260nm}/OD_{280nm}) represents the contamination of nucleic acids with proteins absorbing light at 280 nm. A ratio of 1.8 for DNA and 2.0 for RNA is generally accepted as pure sample. The OD_{260nm}/OD_{230nm} ratio, which should be between 1.8 and 2.2, indicates the presence of residual phenol, carbohydrates, or guanidine etc.

2.9.4.2 Agarose gel electrophoresis

For detection or qualitative analysis of DNA, PCR products or other DNA samples were analyzed by gel electrophoresis. Samples were mixed with 6X Blue Juice loading dye and analyzed on 1–1.5% agarose gels (1-1.5% w/v agarose, 89 mM Tris, 89 mM Boric Acid, 2 mM EDTA (pH 8.0), 0.025% v/v Midori Green), applying an electric field of 100–120 V for 30-120 min. As standards 7 µl pUC8 Mix (Fermentas) or GeneRuler 100 bp Plus DNA Ladder

(Thermo Scientific) were used. DNA bands were then visualized using the gel documentation System Felix® 2000 (Biostep).

2.9.4.3   Purification of DNA

To purify DNA from solutions containing enzymes, dNTPs or salts from buffers of e.g. PCR reactions, the Wizard® SV Gel and PCR Clean-Up Kit (Promega) was used according to the manufacturer´s protocol. The DNA was eluted from the columns two times with 20 μl LiChrosolv® water (MERCK). The DNA concentration was determined (see **2.9.4.1**) and the DNA was stored at -20°C until use.

2.9.4.4   Isolation of plasmid DNA

For the isolation of plasmids from bacteria, two different scales were used depending on the volume of the liquid culture. The NucleoSpin Plasmid Mini Kit (Macherey&Nagel) was used for small scale preparations (6 ml overnight culture), the NucleoSpin Plasmid Midi Kit (Macherey&Nagel) was used for large scale preparations (50 ml overnight cultures).

The overnight culture was centrifuged for 15 min at 4°C. The supernatant was discarded and the pellet was resuspended in resuspension buffer provided with the kit. The isolation procedure was done according to the manufacturer´s protocol with slight modifications. DNA from mini preparations was eluted from the column two times with 25 μl LiChrosolv® water. DNA from midi preparations was eluted from the column using 5 ml elution buffer provided with the kit and the eluted DNA was evenly divided into four 2 ml tubes. After precipitation with 700 μl isopropanol per tube the DNA was centrifuged for 15 min at 15 000 x g and 4°C and the DNA pellet was washed with 700 μl of 70% ethanol per tube by centrifugation for 5 min at 18 000 x g and RT. The ethanol was completely removed and the pellet was dried at RT. The DNA was reconstituted with 25 μl LiChrosolv® water per tube and mixed for one hour on a shaker (1000 rpm) at RT. At the end the four tubes were combined and the DNA concentration was measured (**2.9.4.1**). The DNA was stored at -20°C.

2.9.4.5   Cloning via two restriction sites

Different sizes of *ABCG1* DNA fragments containing the CpG site cg06500161 and *MIR33B/SREBF1* fragments with cg20544516 were cloned into the pCpGfree-promoter-Lucia

plasmid (Invivogen). The specificity of this plasmid is the completely CpG site free plasmid backbone thus making it a useful tool to study the effect of DNA methylation in regulatory elements. It also contains the human EF-1α promoter and the Lucia luciferase reporter gene that codes for a secreted coelenterazine-utilizing luciferase, namely Lucia.

ABCG1 DNA fragments were obtained by PCR amplification (see **2.9.5.1**) from human genomic DNA (Promega). For *MIR33B/SREBF1* oligonucleotides were purchased (Metabion), annealed to double-stranded DNA (see **2.8.3**), and used as inserts for cloning into the pCpGfree-promoter-Lucia plasmid.

Restriction digestion

For the restriction digestion 1 µg of DNA (pCpGfree-promoter-Lucia vector or insert) was digested at 37°C for 2 hours in a 20 µl reaction containing 10X buffer (NEB) and the appropriate amount of the desired restriction enzymes (BamHI and NsiI, both NEB) as well as 10X BSA. The appropriate amount of the restriction enzymes was calculated based on the assumption that 1 U of enzyme digests 1 µg DNA in 1 hour under ideal conditions. After the digestion the restriction enzymes were heat inactivated at 65°C for 20 min. In some cases, a gel electrophoresis (see **2.9.4.2**) was performed to check efficiency of the restriction digestion. The digested sample was stored at -20°C.

Ligation

For ligation of vector and insert a T4 DNA ligase was used. 150 ng of vector DNA and an amount of insert DNA corresponding to a 1:5 molar ratio of insert to vector was added to a 20 µl reaction also containing 10X T4 ligation buffer and an appropriate amount of T4 DNA ligase. Ligation was carried out at room temperature for 2.5 hours. After ligation it was immediately proceeded with the transformation (see **2.9.2.4**).

2.9.4.6   Methylation of pCpGfree-promoter-Lucia plasmid constructs

For methylation, 20 µg DNA were incubated with M.SssI (2.5 U/µg DNA; NEB) in the presence of 160 µM S-adenosylmethionine (SAM) and NEBuffer 2 (10X) for 4 hours at 37°C, with another 160 µM SAM being added after the first two hours of incubation. For HpaII methylation the incubation time was 2 hours, using HpaII (2.5 U/µg DNA; NEB), 80 µM SAM and HpaII Methyltransferase Reaction Buffer (10X) (NEB), adding another 80 µM SAM after

one hour incubation. To stop the reaction the samples were incubated at 65°C for 20 min. Additional samples were treated as above but without methyltransferases and SAM, which were then considered as unmethylated constructs. Samples were purified and afterwards eluted in 40 µl LiChrosolv® water (see **2.9.4.3**). DNA concentrations were determined using the NanoDrop (see **2.9.4.1**).

2.9.4.7   Bisulfite treatment of plasmid DNA

The treatment of single-stranded DNA with sodium bisulfite converts unmethylated cytosines to uracil, whereas methylated cytosines are protected and remain unchanged. The uracils are amplified in subsequent PCR or sequencing reactions as thymines, ending in eventually methylation-dependent sequence changes of cytosine (C) to thymine (T) (**Figure 7**).
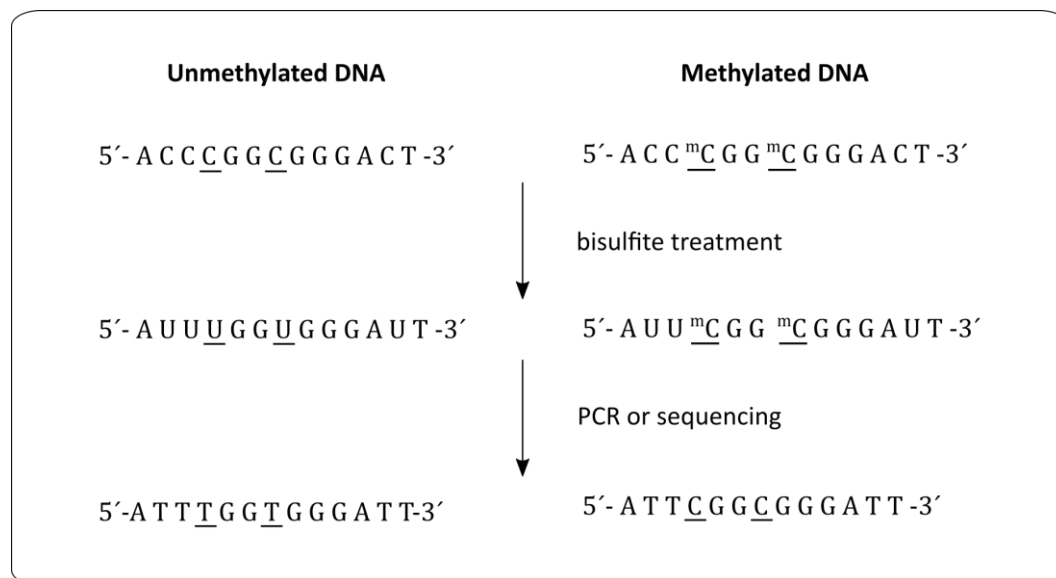
Figure 7: Sequence-changes caused by bisulfite treatment of DNA. Bisulfite conversion can be used to determine the methylation status of CpG sites as methylation-dependent changes in the sequence occur: unmethylated cytosines are converted into uracil and show peaks for thymines after sequencing, whereas methylated cytosines (^{m}C) are protected against the conversion.

To verify DNA methylation of the constructs, the samples were bisulfite converted using the EpiTect® Fast DNA Bisulfite Kit (Qiagen) according to the manufacturer´s protocol. In short, bisulfite reactions containing 500 ng DNA, 85 µl Bisulfite Solution and 35 µl DNA Protect Buffer in a total volume of 140 µl were incubated using the following PCR-cycler conditions: 5 min at 95°C for denaturation, 20 min at 60°C for incubation, with one repetition of these

steps. After conversion, DNA samples were purified according to the manufacturer´s protocol and eluted in 15 µl LiChrosolv® water.

Bisulfite converted samples were then immediately used for Sanger sequencing to verify the bisulfite treatment (see **2.9.5.3**). Here, it should be noted that after bisulfite treatment two non-complementary DNA single-strands remain. Therefore only reverse primers (for both different strands) can be used for the sequencing reactions. The respective primers are listed in **Appendix Table III**.

2.9.5 PCR-based methods

2.9.5.1 Polymerase chain reaction (PCR)

PCRs were performed in 25 µl and 50 µl reactions containing 0.2 mM dNTP mix, 1 µM forward and reverse primer each and 0.5-2.5 U polymerase in 1X PCR buffer.

Depending on the applications different polymerases were used: for screening-PCRs, a lab-made Taq Polymerase (GAC-Taq Polymerase) was used, for applications with higher accuracy like PCRs for cloning and sequencing, Pfu turbo Polymerase (Stratagene) or HotStart Taq polymerase (Qiagen) were used.

A standard PCR program was used consisting of an initial denaturation step for 10 min at 95°C followed by 35 cycles with 95°C for 35 seconds (melting), a specific primer dependent annealing temperature for 30 seconds (annealing) and 72°C for one min per kb (elongation). The final step was performed at 72°C for 10 min.

The annealing temperature was generally 3°C below the melting temperature (T_m) of the applied primer with the lowest melting temperature.

2.9.5.2 Colony screen with bacterial lysates

After cloning and transformation of vectors containing the desired insert into *E.coli* bacteria, single colonies were picked from the agar plates. The colonies were incubated at 150 µl LB medium with the appropriate selection antibiotic for 2 h using a thermo shaker (37°C, 1000 rpm). 3 µl of these pre-cultures were used as templates for a 20 µl PCR reaction using GAC-Taq polymerase and vector- and insert-specific primers (**Appendix Table III**). After PCR

size and integrity of the PCR-products were analyzed by agarose gel electrophoresis (see **2.9.4.2**).

2.9.5.3   Sequencing of DNA

Before the sequencing reaction a 5 µl PCR reaction had to be performed containing 0.5-2 µl template DNA (100 ng per 1 kb template DNA sequence), 10 µM of the sequencing primers and 0.8 µl BigDye® Terminator V3.1 (Applied Biosystems). The reaction was filled up to 5 µl with Ultrapure LiChrosolv® water. The PCR program consisted of following steps: initial denaturation for 5 min at 95°C, 35 cycles with 95°C for 30 seconds, T_m - 3°C for 45 seconds, 60°C for 4 min.

After the PCR reaction and before sequencing the PCR products were purified with the 96-well format Montage^{TM} SEQ$_{96}$ Sequencing Reaction Cleanup Kit (Millipore). Purification was carried out with slight changes to the manufacturer´s protocol: PCR products were eluted from the filter membrane adding 25 µl of injection solution.

PCR products or plasmids were sequenced on the ABI PRISM® 3730 DNA Analyzer (Applied Biosystems) by the Sanger didesoxy method.

2.10   Bioinformatics

2.10.1   DNA sequence analysis (Vector NTI, UCSC)

The Vector NTI component AlignX was used to analyze similarities between nucleotide sequences. A template sequence containing the desired nucleotide arrangement was compared to a sequenced product e.g. to verify cloning of a nucleotide sequence of interest.

Location of CpG sites were analyzed using the UCSC Genome Browser assembly NCBI36/hg18 and GRCh37/hg19.

2.10.2   Transcription factor binding site analysis

The prediction of transcription factor binding sites at cg06500161 and cg20544516 was performed using the Genomatix MatInspector software (Cartharius *et al.* 2005).

3 MATERIALS

3.1 Cell lines and organisms

3.1.1 Human cell lines

Cell line	Cell type	Distributor	Accession number
THP-1	human acute monocytic leukemia	DSMZ	ACC-16
HepG2	human hepatocellular carcinoma	DSMZ	ACC-180
HEK-293	human embryonal kidney	ATCC®	CRL-1573™

THP-1 nuclear extract   Active motif

3.1.2 Bacteria strains

E.coli DH5α (Stratagene)

F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–

E.coli GT115 (Invivogen)

F- mcrA Δ(mrr-hsdRMS-mcrBC) f80lacZDM15 ΔlacX74 recA1 rpsL (StrA) endA1 Ddcm uidA(DMluI)::pir-116 ΔsbcC-sbcD

3.2 Media

3.2.1 Media for human cell lines

Medium for THP-1 cells   RPMI Medium 1640 (1X) (Gibco)

2 mM L-glutamine (Gibco)

10% FBS (Merck Millipore)

Medium for HepG2 and HEK293 cells   DMEM (1X)

2 mM L-glutamine (Gibco)

10% FBS (Merck Millipore)

3.2.2 Media for bacterial strains

LB_0 medium   10 g BACTO Peptone

5 g Yeast extract

add 1000 ml H_2O (pH 7.4)

LB-agar LB_0-medium

10 g/l BACTO-agar

LB_{Zeocin} LB_0-medium/LB-agar

25 µg/ml

LB_{Amp} LB_0-medium/LB-agar

50 mg/l ampicillin

3.2.3 Antibiotics and supplements

Ampicillin (50 mg/ml)   Sigma

Zeocin (100 mg/ml)   Invitrogen

Fetal Bovine Serum Superior   Merck Millipore

L-Glutamine 200 mM (100X)   Gibco

3.3 Vectors

For detailed vector maps information see **Appendix Figure I and II**.

pCpGfree-promoter-Lucia   Invivogen

pGL4.53   Promega

3.4   Enzymes

3.4.1   Polymerases

Hot start Taq polymerase (5 U/µl)	Qiagen
GAC Taq DNA polymerase (lab-made on 24.5.2004)	GAC
Pfu turbo (2.5 U/µl) polymerase	Agilent

3.4.2   Restriction endonucleases

BamHI	NEB
NsiI	NEB

3.4.3   Methyltransferases

HpaII 4,000 units/ml	NEB
M.SssI 20,000 units/ml	NEB

3.4.4   Further enzymes

T4 DNA Ligase (400 000 u/ml)	NEB
0.05% Trypsin-EDTA (1X)	Gibco

3.5   DNA Marker

pUC Mix Marker 8	Fermentas
GeneRuler 100 bp DNA Ladder	Fermentas

3.6   Chemicals, Supplements, Media

10X cloned Pfu Reaction Buffer	Agilent
2-Propanol	Merck
3-(N-morpholino)propanesulfonic acid (MOPS)	Sigma
Agarose	Biozym
Ammonium peroxodisulfate (APS)	Biozym
Boric acid	Roth

Dimethylsulfoxide (DMSO)	Sigma
Dithiothreitol (DTT)	Fermentas
DMEM	Gibco
dNTPs	Fermentas
EDTA	Merck
Ethanol	Merck
Fetal bovine serum (FBS) Superior	Gibco
Ficoll 400	Amersham
Glycerol	neoLab
Glycine	Biomol
Hydrogen chloride	Sigma
L-Glutamine 200 mM (100X)	Gibco
LiChrosolv® water	Millipore
Magnesium chloride (25 mM)	Qiagen
Methanol	Merck
Midori Green	Biozym
Orange G	Sigma
PBS pH 7.4 (1X)	Gibco
PCR buffer 10X	Qiagen
Poly(d(I-C))	Roche Diagnostics
Potassium acetate	Merck
Potassium chloride	Merck
Rotiphorese® Gel 40 (37,5:1)/30 (37,5:1)	Roth
RPMI 1640	Gibco
Sodium chloride	Merck
Sodium phosphate	Merck
Tetramethylethylenediamine (TEMED)	Sigma
Tris	Merck

3.7   Consumables

96F Nunclon® MicroWell plates white	Nunc
Cellstar plastic pipettes 5 ml/ 10 ml/ 25 ml	Greiner bio-one
Cryo Tube^{TM} Vials	Nunc
Falcon tubes 15 ml/ 50 ml	BD
Multiwell^{TM} 12 well plates	Falcon
Parafilm M® Laboratory Film	Pechiney Plastic Packaging
Safe-Lock Tubes 1.5 ml/ 2.0 ml	Eppendorf
TC Flask (T25/T75)	Greiner bio-one

3.8   Laboratory Equipment

ABI PRISM® 3730 DNA analyzer	Applied Biosystems
Cellometer Auto T4	Nexcelom
Centrifuge Typ 3500	Hettich Centrifuges
Centrifuge Universal 32R	Hettich Centrifuges
CO_2 Incubator	Sanyo
CoolCell® Cell Freezing Container	Biocision
Gel electrophoresis chamber	Biozym
Gel documentation System Felix® 2000	Biostep
GloMax® 96 Microplate Luminometer	Promega
Haake K20 + Haake DC10 Kühlbad	Thermo Haake
Haake Water bath DC10	Thermo Haake
Heraeus Fresco 21 Centrifuge	Thermo Scientific
Incubator Function line B6	Heraeus instruments
Maxigel System	Biometra
Microflow	Nunc
Micropipettes	Rainin
Microscope Primo Vert	Zeiss
Mikro200	Hettich Centrifuges
Mikrowelle 900 & Grill	Severin

Mixer Uzusio VTX-3000 L	LMS
Multi Tube Vortexer DVX-2500	VWR
NanoDrop ND 8000 Spectrophotometer	Peqlab
P25T Standard Power Pack	Biometra
Peltier Thermal Cycler	MJ Research
pH Meter 766 calimatric	Knick
Power PAC 200 and 300	BioRad
Refrigerated incubator shaker Innova 4230	New Brunswick Scientific
RoboCycler 96 Temperature Cycler Gradient	Stratagene
RoboCycler 96 Temperature Cycler	Stratagene
Rottana 46S	Hettich Centrifuges
Scale 57245	Kern
Sequencer ABI 3100 Genetic Analyzer	Applied Biosystems
Shaker Innova 4230	New Brunswick Scientific
SubCell® GT electrophoresis chamber	BioRad
Thermo Block HBT 130	HLC
Thermomixer comfort	Eppendorf
Thermomixer compact	Eppendorf
Typhoon Trio +	GE Healthcare
Unitwist 3-D	UniEquip
UV-VIS Spectrophotometer DU530	Beckman
VWB12 water bath	VWR

3.9 Kits

BigDye terminator v3.1 cycle Sequencing Kit	Applied Biosystems
Dual-Luciferase® Assay Kit	Promega
EpiTect® Fast DNA Bisulfite Kit	Qiagen
MontageTM SEQ96-Sequencing Reaction Cleanup Kit	Millipore
Nuclear Extraction Kit	Active Motif
NucleoBond® PC 100 (Midi) Plasmid DNA purification Kit	Machery-Nagel
NucleoSpin® Plasmid Kit	Machery-Nagel

Pierce® BCA Protein Assay Kit   Thermo

Wizard® SV PCR and Gel Clean-Up System   Promega

3.10 Gel compositions and buffers

1X binding buffer for EMSA	4% v/v Glycerol
	1mM MgCl$_2$
	0.5 mM EDTA
	0.5 mM DTT
	50 mM NaCl
	10 mM TrisHCl pH7.5
TBE 5X	54 g Tris Base
	27.5 g boric acid
	20 ml 0.5 M EDTA pH 8.0
	add 1 L H$_2$O, final pH ~8.3
12% polyacrylamide gel	25 ml TBE 1X
	15 ml 37, 5:1 acrylamide/bisacrylamide (40% w/v)
	1562.5 µl 80% v/v Glycerol
	7.0 ml H$_2$O
	375 µl 10% APS
	25 µl TEMED
5.3% polyacrylamide gel	25 ml TBE 1X
	6.5 ml 37, 5:1 acrylamide/bisacrylamide (40% w/v)
	1562.5 µl 80% v/v Glycerol
	15.5 ml H$_2$O
	375 µl 10% APS
	25 µl TEMED
Composition of 10X PCR buffer	100 mM Tris-HCl pH 9.0
for lab-made GAC-Taq	500 mM KCl
	15 mM MgCl$_2$

6X Blue Juice gel loading dye	30% Ficoll
for DNA gel electrophoresis	0.5 M EDTA
	0.025% bromophenol blue

10X gel loading dye (Orange G)	
for EMSA	250 mM TrisHCl pH 7.5
	0.2% Orange G
	40% v/v Glycerol

3.11 Computer Software and Programs

3.11.1 Online tools, databases and freeware

BioEdit (free alignment program): http://www.mbio.ncsu.edu/BioEdit/bioedit.html

BLAST (Basic Local Alignment Search Tool): http://www.ncbi.nlm.nih.gov/BLAST

Metabion Oligonucleotide Calculator: http://www.metabion.com/biocalc/index.html

NCBI (National center of Biotechnology Information): http://www.ncbi.nlm.nih.gov/

UCSC Genome Bioinformatics Browser: http://genome.ucsc.edu/

Reverse complement–Sequence converter:

http://www.bioinformatics.org/sms/rev_comp.html

Bisulfite Primer Seeker: http://www.zymoresearch.com/tools/bisulfite-primer-seeker

3.11.2 Computer software

BioEdit Sequence Alignment Editor (Ibis Biosciences)

GloMax® 96 Microplate Luminometer Software version 1.9.2 (Promega)

Bio-Imaging software ArgusX1 (Biostep)

Microsoft Office 2010

NanoDrop version 3.1.0 NanoDrop Technologies

R software (R Foundation)

Vector NTI (Invitrogen)

Cellometer Auto T4_V335 (Nexcelom)

Inkscape Version 0.91

yEd – Graph editor Version 3.15.0.2

4 RESULTS

4.1 Epigenome-wide association study of main blood lipid levels

Results of the EWAS of main blood lipid levels are published in Pfeiffer *et al.* (Pfeiffer *et al.* 2015).

4.1.1 Characteristics of the cohorts

To assess the relationship between genome-wide DNA methylation and human main blood lipid levels, an epigenome-wide association study (EWAS) was performed using data of the KORA F4 cohort (N=1,776). Significant associations were replicated in the KORA F3 (N=499) and InCHIANTI (N=472) cohort. After meta-analysis of the replication results of KORA F3 and InCHIANTI, successfully replicated associations were also tested in methylation datasets of human adipose and skin tissue of the MuTHER cohort (N=856). Characteristics of the subjects of the discovery cohort as well as of the three replication cohorts are shown in **Table 4**.

The mean age of all cohorts was between 52.9 (KORA F3) and 71.2 (InCHIANTI) years. In KORA F4, F3 and InCHIANTI the sex distribution was almost equal, the MuTHER cohort consisted only of females. In KORA F4 and InCHIANTI, 16.3% and 13.0% of the individuals were taking lipid-lowering drugs, respectively, whereas in the KORA F3 and MuTHER cohort the intake of lipid-lowering drugs was lower (6.2% of KORA F3 subjects, 8.1% of MuTHER subjects). The lipid levels were within the normal ranges except the slightly elevated triglyceride levels in KORA F3 and total cholesterol levels in all cohorts. In KORA F4 N=60 cases of hospitalized myocardial infarcts were available, KORA F3 comprised N=8 patients and InCHIANTI N=36. There was no information about myocardial infarction available in the MuTHER cohort.

Table 4: Characteristics of the study subjects

Characteristic	KORA F4 (N=1,776)	KORA F3 (N=499)	InCHIANTI (N=472)	MuTHER cohort (N=856)
Age (years)	60.8 (8.9)	52.9 (9.6)	71.2 (16.0)	59.4 (9.0)
Sex=male	867 (48.8%)	259 (51.9%)	215 (45.6%)	0 (0.0%)
BMI (kg/m^2)*	28.2 (4.8)	27.2 (4.5)	27.0 (4.3)	26.6 (4.9)
Current smoker	258 (14.5%)	249 (49.9%)	206 (43.7%)†	84 (9.8%)
Physically active	1021 (57.5%)	249 (49.9%)	238 (50.5%)	NA‡
Alcohol consumption (g/day)	15.5 (20.4)	16.1 (19.6)	12.6 (15.1)	NA‡
HDL-C (mg/dl)	56.5 (14.6)	58.2 (17.8)	56.8 (14.8)	71.5 (18.2)
LDL-C (mg/dl)	140.0 (35.1)	131.0 (33.2)	124.5 (32.4)	124.5 (37.9)
Triglyceride (mg/dl)	133.1 (94.7)	164.6 (121.9)	119.7 (57.9)	99.2 (49.6)
Cholesterol (mg/dl)	221.9 (39.3)	220.5 (38.2)	205.2 (37.9)	218.9 (38.7)
C-reactive protein (mg/l)	2.5 (5.1)	NA‡	4.0 (8.0)	NA‡
WBC count (WBC/nl)§	5.9 (1.6)	7.3 (2.1)	6.3 (1.6)	6.5 (1.8)
HbA1c (%)‖	5.6 (0.6)	5.3 (0.5)	4.9 (0.8)#	NA‡
Self-reported history:				
Hypertension**	811 (45.7%)	211 (42.3%)	124 (26.3%)	172 (20.1%)
Hospitalized myocardial infarction	60 (3.4%)	8 (1.6%)	36 (7.6%)	NA‡
Intake of lipid-lowering drugs (excl. herbal substances)	290 (16.3%)	31 (6.2%)	61 (13.0%)	69 (8.1%)
Fasting at time of blood collection††	1776 (100.0%)	47 (9.4%)	472 (100.0%)	844 (98.5%)

Continuous and categorical characteristics are given as mean (sd) or absolute numbers and relative proportions, respectively. *BMI: body mass index; †current smokers variable comprises in InCHIANTI current and former smokers; ‡NA: variable not available; §WBC: white blood cell count; ‖HbA1c: hemoglobin A1c; #in InCHIANTI, HbA1c levels were calculated using the formula (46.7+glucose level)/28.7, in KORA F3/F4 they were analyzed using the HPLC method; **>140/90 mmHg or medically controlled; ††overnight fast of at least 8 hours

4.1.2 Blood lipid levels were associated with DNA methylation

To investigate the association between genome-wide DNA methylation and human main blood lipid levels (HDL-C, LDL-C, TG and total cholesterol), an epigenome-wide association study (EWAS) was performed in KORA F4. The EWAS discovered 12 associations between the methylation status of 11 CpG sites and lipid levels. Methylation of one CpG site was significantly associated (p-value<1.1E-07) with HDL-C and LDL-C, respectively. Ten CpG sites showed an association with TG; no CpG site was associated with total cholesterol. P-values ranged from 1.21E-27 to 9.66E-08, with percentage of explained lipid level variance ranging from 1.6% to 6.5% (**Table 5**).

The CpG site cg06500161, located in the *ABCG1* gene, was associated in opposite directions with HDL-C (β=-0.049, p-value=8.26E-17) and TG levels (β=0.070, p-value=1.21E-27). TG levels were associated with nine additional CpG sites located in genes including *ABCG1*, *MIR33B*, *SREBF1*, *CPT1A*, *TXNIP* and *APOA5*. Two further CpG sites, cg07504977 and cg07815238, are located in intergenic regions. The CpG site cg07504977 is located between *SCD* and a gene coding for a long non-coding RNA (*LINC00263*) on chromosome 10 (chr10:102,131,013). The CpG site cg07815238 is located between the *DAPK2* and *FAM96A* gene on chromosome 15 (chr15:6,434,100). LDL-C showed a positive association with the methylation status of a CpG site located in *TNIP1*.

The lipid-associated CpG sites were replicated in two independent cohorts, KORA F3 and InCHIANTI. As a meta-analysis is a powerful tool to summarize data and to effectively increase sample size to provide a more valid pooled estimate (Greco *et al.* 2013), a meta-analysis of the replications was performed. This analysis confirmed nine of the 12 associations, with p-values ranging from 9.00E-11 to 3.78E-03 (**Table 5**). Associations between TG levels and cg07815238 (chr15:6,434,100) and CpG sites located in *TXNIP* and *APOA5* were not replicated.

Table 5: Associations between genome-wide DNA methylation and human blood lipid levels

Lipid	CpG	Chr[†]	Gene	β coef[‡]	SE[§]	p-value	exp var (%)[∥]	β coef	SE	p-value	β coef	SE	p-value	p-value
				KORA F4				KORA F3			InCHIANTI			Meta-analysis[*]
HDL-C	cg06500161	21	ABCG1	-0.049	0.006	8.26E-17	3.9	-0.065	0.014	2.97E-06	-0.071	0.016	1.13E-05	9.00E-11[#]
TG	cg06500161	21	ABCG1	0.070	0.006	1.21E-27	6.5	0.072	0.015	1.89E-06	0.063	0.016	1.03E-04	5.56E-10[#]
	cg19693031	1	TXNIP	-0.030	0.003	1.89E-17	4.1	-0.014	0.007	5.65E-02	-0.023	0.011	3.54E-02	5.67E-03
	cg11024682	17	SREBF1	0.059	0.008	5.54E-14	3.2	0.031	0.014	2.89E-02	0.030	0.013	2.36E-02	1.60E-03[#]
	cg00574958	11	CPT1A	-0.118	0.016	3.15E-13	3.2	-0.103	0.028	2.42E-04	-0.058	0.020	4.86E-03	7.88E-06[#]
	cg27243685	21	ABCG1	0.064	0.009	3.24E-13	3.0	0.050	0.014	5.87E-04	0.054	0.022	1.63E-02	2.49E-05[#]
	cg07504977	10	NA**	0.026	0.004	3.93E-12	2.7	0.026	0.008	1.87E-03	0.027	0.009	3.74E-03	1.91E-05[#]
	cg20544516	17	MIR33B/ SREBF1	0.043	0.007	2.84E-09	2.7	0.032	0.013	1.39E-02	0.032	0.018	7.16E-02	2.22E-03[#]
	cg12556569	11	APOA5	0.005	0.001	6.43E-09	1.9	0.002	0.002	2.56E-01	0.004	0.002	1.25E-02	1.20E-02
	cg07397296	21	ABCG1	0.027	0.005	9.48E-08	2.1	0.034	0.010	1.03E-03	0.008	0.011	4.63E-01	3.78E-03[#]
	cg07815238	15	NA**	0.048	0.009	9.66E-08	1.6	0.015	0.017	3.69E-01	0.003	0.014	8.41E-01	4.61E-01
LDL-C	cg22178392	5	TNIP1	0.040	0.007	4.27E-09	2.1	0.049	0.015	1.11E-03	0.020	0.014	1.45E-01	1.04E-03[#]

*meta-analysis of the results of the replication in KORA F3 and InCHIANTI; [†]Chr: chromosome; [‡]β coef: β coefficient; [§]SE: standard error; [∥]exp var: explained variance; [#]lipid-associated CpG site confirmed by replication meta-analysis; level of significance: 1.1E-07 (discovery cohort), 4.5E-03 (replication meta-analysis); **no gene annotation for this CpG site according to the UCSC Genome Browser

To investigate tissue specificity, associations were also replicated in adipose tissue as well as in skin tissue (negative control) in cooperation with the MuTHER cohort. Five of the replicated associations were also significant in adipose tissue (**Table 6**). Here, the CpG site cg20544516 (*MIR33B/SREBF1*) showed the strongest association with TG levels (β=0.012, p-value=1.20E-10), followed by CpG sites located in *ABCG1* (cg27243685, cg07397296; β=0.013, p-value=5.86E-08 and β=0.008, p-value=6.59E-07, respectively) and *SREBF1* (cg11024682, β=0.007, p-value=6.72E-04). The association between LDL-C and cg22178392 (*TNIP1*) was also found to be significant in adipose tissue (β=0.002, p-value=6.02E-03). In skin tissue no associations could be determined except for TG levels and cg11024682 (*SREBF1*) and cg00574958 (*CPT1A*) (β=0.006, p-value=4.07E-04 and β=-0.005, p-value=2.81E-03, respectively).

Associations of the CpG site cg06500161 (*ABCG1*) could not be replicated in the MuTHER cohort as the methylation data of this CpG site did not pass the quality control filters of the cohort and were thus not available in the dataset.

Table 6: **Associations between blood lipid levels and lipid-associated CpG sites in adipose and skin tissue of the MuTHER cohort**

Lipid	CpG	Gene	Adipose tissue (N=634)			Skin (N=395)		
			β coef*	SE[†]	p-value	β coef*	SE[†]	p-value
HDL-C	cg06500161	*ABCG1*	na[‡]	na[‡]	na[‡]	na[‡]	na[‡]	na[‡]
TG	cg06500161	*ABCG1*	na[‡]	na[‡]	na[‡]	na[‡]	na[‡]	na[‡]
	cg11024682	*SREBF1*	0.007	0.002	6.72E-04[||]	0.006	0.002	4.07E-04[||]
	cg00574958	*CPT1A*	0.0001	0.001	8.16E-01	-0.005	0.002	2.81E-03[||]
	cg27243685	*ABCG1*	0.013	0.002	5.86E-08[||]	0.003	0.003	3.91E-01
	cg07504977	NA[§]	0.006	0.003	4.60E-02	0.004	0.002	1.56E-01
	cg20544516	*MIR33B/ SREBF1*	0.012	0.002	1.20E-10[||]	-0.001	0.002	5.04E-01
	cg07397296	*ABCG1*	0.008	0.002	6.59E-07[||]	0.001	0.002	6.64E-01
LDL-C	cg22178392	*TNIP1*	0.002	0.001	6.02E-03[||]	0.0003	0.001	7.49E-01

*β coef: β coefficient; [†]SE: standard error; [‡]na: no DNA methylation data available for this CpG site; [§]NA: no gene annotation for this CpG according to the UCSC Genome Browser; [||]significant associations; level of significance: 7.14E-03

4.1.3 DNA methylation of *ABCG1* was associated with gene expression

Aim of the expression analysis was to disentangle the relationships between methylation at the CpG sites, expression of the corresponding annotated gene, and lipid levels. For this purpose, an *ad hoc* approach was applied based on a sequence of regression models with and without adjusting for the third of the three components.

The gene expression analysis discovered a negative association between methylation of a CpG site located in *ABCG1* (cg06500161) and mRNA levels of the six in KORA F4 available *ABCG1* transcripts (p-value=5.42E-14; **Figure 8a** ①). Here, the transcript ILMN_2329927 of the six *ABCG1* transcripts showed the strongest association (β=-0.151, p-value=5.22E-15) (**Appendix Table V**). When additionally adjusted for the respective lipid (HDL-C), the association became less significant (p-value=5.15E-11; **Figure 8a** ①).
ABCG1 transcript levels showed also a strong positive association with HDL-C levels (p-value=7.76E-13). The significance of the association was reduced when adjusting for cg06500161 methylation (p-value=1.79E-09, **Figure 8a** ②). The association between cg06500161 and HDL-C weakened (from a p-value of 4.10E-07 to a p-value of 1.30E-02) after adjusting for *ABCG1* transcripts (**Figure 8a** ③).

Similar results were found for the association between methylation of cg06500161 and TG levels. The negative association between cg06500161 and *ABCG1* transcripts became less significant when additionally adjusted for TG levels (marginal p-value=5.42E-14, p-value after adjustment for TG levels = 1.23E-06; **Figure 8b** ①). In contrast to the association between the *ABCG1* transcripts and HDL-C levels, transcripts were negative correlated with TG levels (p-value=1.25E-33). Significance of the association was reduced when adjusting for cg06500161 methylation (p-value=1.20E-25); **Figure 8b** ②). The association between cg06500161 and TG levels weakened after adjusting for *ABCG1* transcripts (p-value without adjustment for *ABCG1* transcripts =5.56E-13; p-value after adjustment for *ABCG1* transcripts =1.36E-05; **Figure 8b** ③).
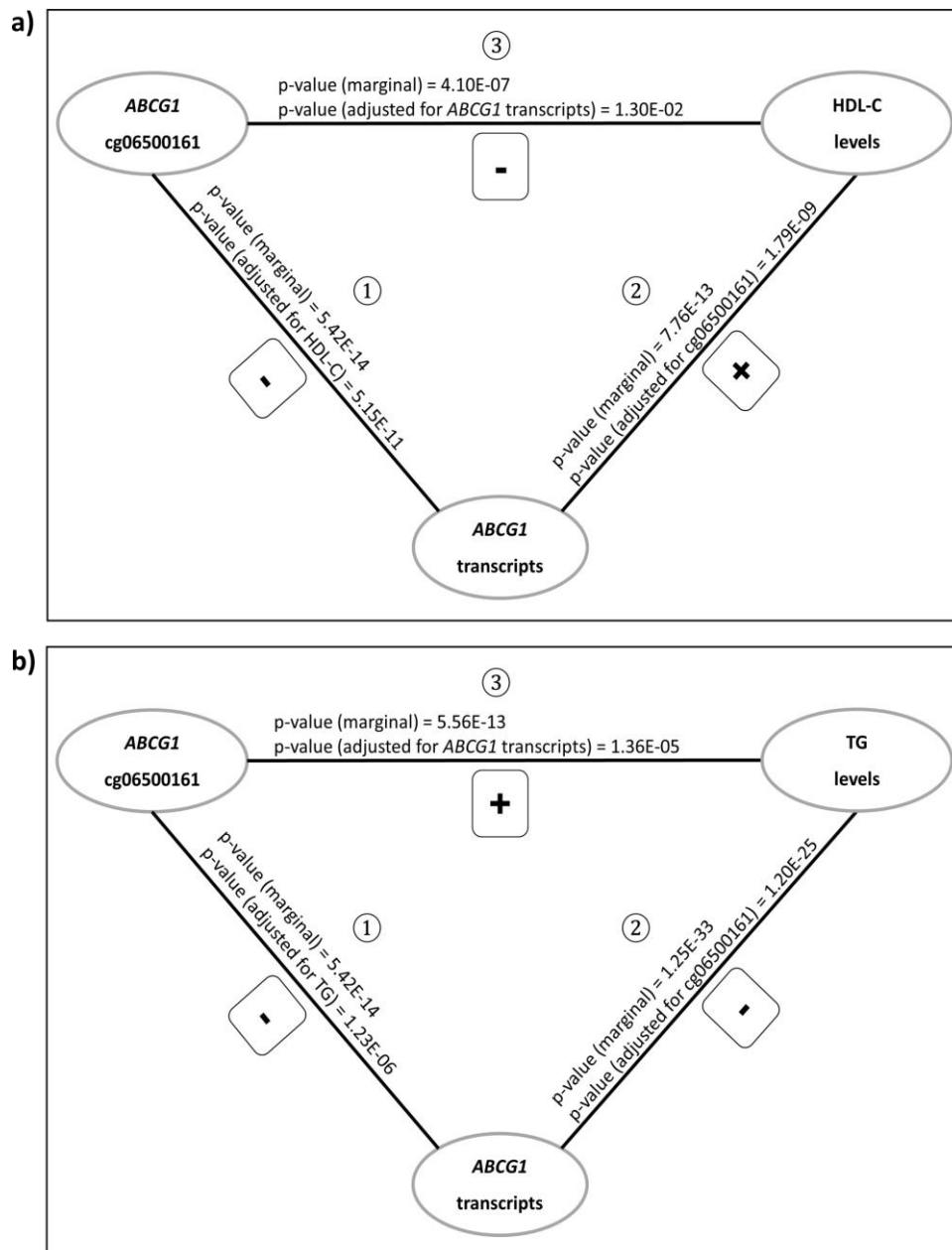
Figure 8: Triangular relationship between *ABCG1* methylation, *ABCG1* gene expression and lipid levels. ① Methylation of cg06500161 was negatively associated with *ABCG1* mRNA levels. ② *ABCG1* mRNA levels were positively associated with HDL-C (a) and negatively associated with TG levels (b). ③ The association between *ABCG1* methylation and lipid levels became less significant after adjusting for *ABCG1* transcripts. Negative and positive associations are indicated by minus or plus, respectively. Analysis was done in KORA F4 (N=724); Level of significance=8.3E-04.

The same analyses were done for the association between the methylation of cg27243685 (*ABCG1*) and TG levels. The CpG site cg27243685 was negatively associated with *ABCG1* transcripts (p-value=1.86E-07). Additional adjustment for TG levels weakened the association (p-value=8.11E-04; **Appendix Figure III** ①). Significance of the association

between *ABCG1* transcripts and TG levels were lowered after adjusting for cg27243685 (p-value without adjustment for cg27243685 =1.25E-33; p-value after adjustment =7.08E-30; **Appendix Figure III** ②). The association between methylation of cg27243685 and TG levels (p-value=2.36E-07) became insignificant (p-value=3.24E-03) when it was additionally adjusted for *ABCG1* transcripts; **Appendix Figure III** ③).

4.1.4 The association between triglyceride levels and cg12556569 (*APOA5*) was confounded by methQTLs

Investigation of genetic confounding was carried out to identify whether the observed associations between lipids and methylation levels in KORA F4 were due to SNPs which are associated with both lipid levels and DNA methylation. First, those known lipid-associated SNPs were preselected which were nominally associated (p-value<0.05) with DNA methylation at the identified lipid-associated CpG sites in KORA F4 and which thus act as potential confounders. P-values of associations ranged between 4.99E-02 and 3.42E-05, except for one SNP (rs964184), which was highly significantly associated with DNA methylation of cg12556569 (*APOA5*, p-value=3.75E-289). Models for each CpG-lipid pair were recalculated with additional adjustment for the respective pre-selected SNP. These analyses showed that only the association between methylation of *APOA5* (cg12556569) and TG was considerably genetically confounded. This was indicated by an increase of the p-value compared to the discovery p-value (6.43E-09 vs. 2.05E-01 after adjusting for the SNPs; **Appendix Table VI**). Further analyses identified rs964184 of the preselected SNPs as cause of the genetic confounding.

4.1.5 Methylation of cg06500161 (*ABCG1*) was associated with prevalent myocardial infarction

The lipid-associated CpG sites of the discovery cohort were tested for an association with previous hospitalized myocardial infarction in KORA F4 (N=1,776 with N=60 myocardial infarction cases). The methylation of the CpG site cg06500161 (*ABCG1*) showed a positive association with myocardial infarction, independent of lipid levels (odds ratio 1.15; 95% confidence interval=1.06-1.25; **Appendix Table VII**). The results could not be replicated in KORA F3 (N=8 cases) and InCHIANTI (N=36 cases).

4.2 Epigenome-wide association study of NMR-measured metabolites

4.2.1 Characteristics of the cohorts

The discovery EWAS was performed in KORA F4 (N=1,662), significant results were subsequently replicated by three independent cohorts: LOLIPOP (N=2,805), NFBC1966 (N=771) and YFS (N=176). Characteristics of the subjects of the discovery cohort as well as the three replication cohorts are shown in **Table 7**.

The mean age of all cohorts ranged between 31.01 (NFBC1966) and 60.99 (KORA F4) years. In KORA F4 the sex distribution was almost equal, in LOLIPOP there were more men than women (68.0% men), whereas in NFBC1966 and YFS the proportion of men was lower than of women.

Participants of all cohorts were in a state of fasting when blood samples were collected. The main lipid levels (HDL-C, LDL-C, TG, total cholesterol) were within normal ranges except the slightly elevated total cholesterol levels in KORA F4, NFBC1966 and YFS. In KORA F4 16.2% of the participants were taking lipid-lowering drugs. In the YFS cohort the intake was quite low (4.0%) and in the NFBC1966 cohort no one was taking lipid-lowering drugs. In the LOLIPOP cohort there was no information available about the intake of lipid-lowering drugs.

Table 7: Characteristics of the subjects of the discovery cohort and the replication cohorts for the EWAS of NMR-measured metabolites

	KORA F4 (N=1,662)	LOLIPOP (N=2,805)	NFBC1966 (N=771)	YFS (N=176)
Age (years)	60.99 (8.92)	51.39 (10.14)	31.01 (0.33)	44.2 (3.34)
Sex=male	815 (49%)	1908 (68.0%)	337 (43.7%)	66 (37.5%)
BMI (kg/m^2)*	28.1 (4.75)	27.68 (4.37)	24.39 (3.73)	25.9 (4.46)
Current smokers	239 (14.4%)	249 (8.9%)	202 (26.2%)	29 (16.5%)
Ex-smokers	689 (41.5%)	244 (8.7%)	148 (19.2%)	45 (25.6%)
Never smokers	732 (44.1%)	2312 (82.4%)	411 (53.3%)	102 (58.0%)
Physically active†	955 (57.5%)	785 (28.0%)	459 (59.5%)	-
Physical activity index (PAI)†	-	-	-	9.16 (2.00)
Alcohol consumption (g/day)	15.49 (20.26)	6.38 (14.19)	8.66 (14.28)	0.716 (0.95)‡
HDL-C (mmol/l)	1.67 (0.43)	1.26 (0.31)	1.74 (0.51)	1.71 (0.42)
LDL-C (mmol/l)	2.09 (0.59)	1.91 (0.57)	1.99 (0.74)	2.06 (0.58)
VLDL-C (mmol/l)	0.99 (0.35)	0.89 (0.31)	0.86 (0.34)	0.713 (0.26)
Total cholesterol (mmol/l)	5.66 (1.04)	4.86 (1.08)	5.50 (1.45)	5.34 (0.99)
Total TG§ (mmol/l)	1.60 (0.85)	1.46 (0.68)	1.17 (0.66)	1.16 (0.65)
HDL-TG (mmol/l)	0.16 (0.05)	0.15 (0.04)	0.16 (0.07)	0.138 (0.04)
LDL-TG (mmol/l)	0.2 (0.06)	0.22 (0.07)	0.18 (0.09)	0.185 (0.05)
VLDL-TG (mmol/l)	1.11 (0.76)	0.95 (0.56)	0.71 (0.53)	0.716 (0.59)
Total FA (mmol/l)	14.52 (2.39)	12.55 (3.02)	13.39 (3.46)	12.9 (2.31)
C-reactive protein (mg/l)	2.5 (5.16)	4.24 (7.18)	1.85 (3.35)	1.44 (2.41)
WBC count (WBC/nl)‖	5.9 (1.63)	4.96 (0.51)	5.94 (1.67)	5.62 (1.60)
HbA1c# (%)	5.63 (0.63)	5.66 (0.76)	NA**	5.49 (0.34)
Self-reported history:				
Hypertension	757 (45.5%)	1066 (38.0%)	129 (16.7%)	34 (19.3%)
Hospitalized myocardial infarction	59 (3.6%)	108 (3.9%)	0 (0.0%)	0 (0.0%)
Intake of lipid-lowering drugs	269 (16.2%)	NA**	0 (0.0%)	7 (4.0%)

Continuous and categorical characteristics are given as mean (sd) or absolute numbers and relative proportions, respectively. *BMI: body mass index; †in KORA individuals who participated in leisure time physical activity during summer and winter and were active for at least one hour per week in either season were classified as being physically active; in YFS physical activity was specified using the Physical Activity Index (PAI) which was calculated as described in (Telama *et al.* 2005); ‡Drinks per day (directly proportional to g/day); §TG: triglycerides; ‖WBC: white blood cell count; #HbA1c: hemoglobin A1c; **NA: variable not available

4.2.2 NMR-measured metabolites were associated with DNA methylation

In the discovery cohort KORA F4 (N=1,662), the EWAS of NMR-measured metabolites revealed 282 significant associations between DNA methylation and metabolites, of which 45 did not pass the quality controls. The p-values of the remaining 237 associations ranged from 5.76E-51 to 4.71E-10. Methylation levels of 24 unique CpG sites located in 13 different genes were associated with 60 unique metabolites or their derivatives (**Appendix Figure IV**).

Results were validated by replication in three independent cohorts. In LOLIPOP (N=2,805), 227 associations were replicated (p-value<2.1E-04). In the Finnish replication cohorts, 17 (NFBC1966, N=771) and 19 (YFS, N=176) associations were replicated. A meta-analysis of the replication results confirmed 118 of the 237 associations (49.8%), where 15 different CpG sites showed significant associations with 48 unique metabolites or their derivatives (**Figure 9**). The percentage of explained metabolite level variance ranged from 1.2% to 8.0%. A CpG-by-sex interaction analysis revealed no evidence of a significant difference between men and women for the associations. CpG sites which showed associations with mainly lipid-related metabolites were located in genes such as *ABCG1* (cg06500161, cg27243685), *TXNIP* (cg19693031), *CPT1A* (cg00574958), *MIR33B/SREBF1* (cg20544516) and *SREBF1* (cg11024682). In contrast, CpG sites located in *FADS2* (three CpG sites in total) or nearby (cg27386326), *PHGDH* (cg16246545), *SCD* (two CpG sites), *SLC7A11* (cg06690548) and in an intergenic region between *SCD* and a gene coding for a long non-coding RNA (*LINC00263*) on chromosome 10 (cg07504977), were associated with fatty acid levels. A list with detailed results is provided in the **Appendix Table VIII**.

Epigenetic associated lipoprotein subfractions were mainly VLDL particles of different sizes (extra small - XS, small - S, medium - M, and large - L) and their content of TG, phospholipids (PL), total lipids (L), free cholesterol (FC), cholesterol esters (CE) and total cholesterol (C). All of those traits were highly correlated with each other in KORA F4 (**Appendix Figure V**). Additionally, DNA methylation of certain genes was associated with the content of TG in IDL, LDL and HDL particles as well as with total serum TG levels. Discovered epigenetic associated fatty acid composition traits were serum levels of monounsaturated fatty acids (MUFA) and omega-3 fatty acids (total levels and in particular level of docosahexaenoic acid DHA). Also ratios such as DHA, MUFA, PUFA or omega-3 and omega-6 fatty acid to total fatty acids were related to the methylation state of certain CpG sites (**Figure 9**).
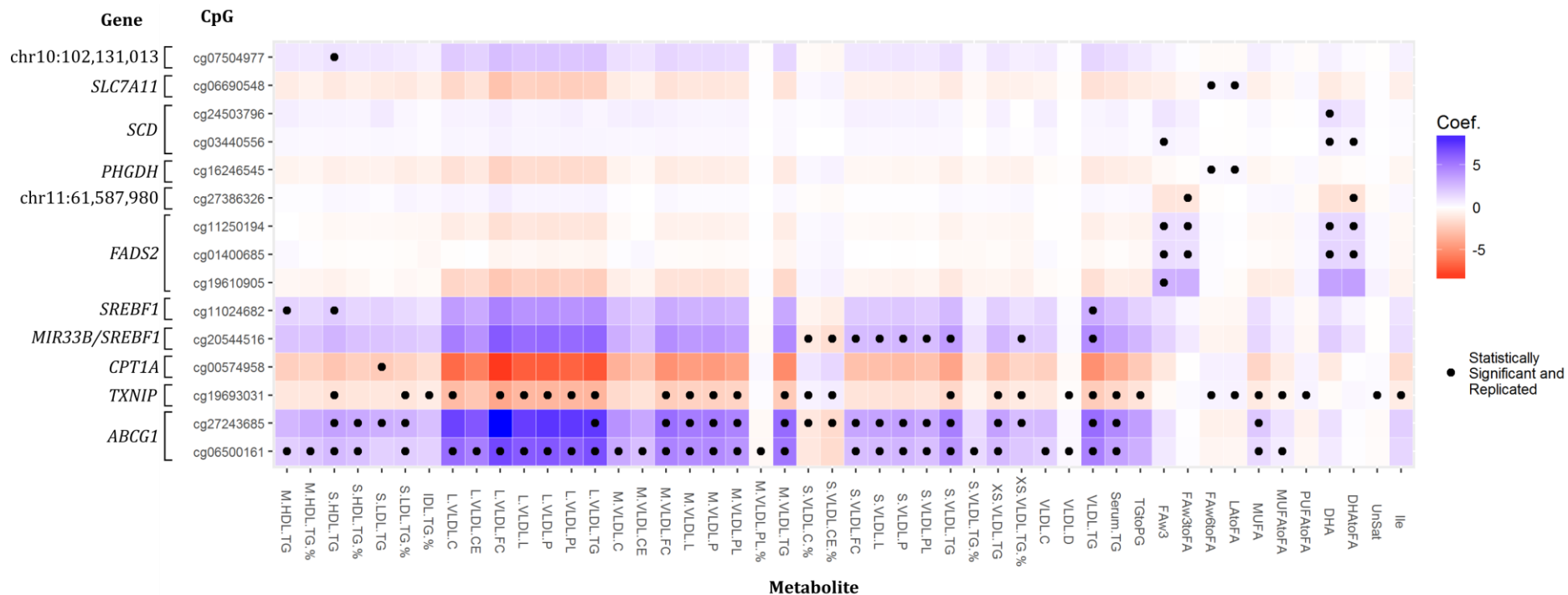
Figure 9: EWAS revealed associations between DNA methylation of certain CpG sites and NMR-measured metabolites. Results were confirmed by meta-analysis of replication results of LOLIPOP, NFBC1966 and YFS. Black points indicate statistical significance and confirmed associations between CpG sites and metabolites (N=118). Boxes are colored according to the signs and values of the coefficients. Level of significance: 2.1E-04

4.2.3 Gene expression analysis revealed associations between DNA methylation and gene expression

As DNA methylation may also effect gene expression, it was tested whether the identified metabolite-related CpG sites are associated with gene transcripts located within 1 Mb of the CpG site´s location.

The gene expression analysis revealed eight CpG sites to be significantly associated (p-value<6.3E-05) with gene expression in KORA F4 (N=692). Here, not only intragenic associations were discovered, also associations with transcripts of genes located nearby were found. CpG sites located in *FADS2* showed negative associations with transcripts of the genes *FADS2*, *FADS1* or *C11ORF10* (**Table 8**). In contrast, cg27386326, located in an intergenic region between *FADS1* and *FADS2*, showed a positive relationship with expression of the *FADS1* gene.

In addition, expression analysis discovered a negative association between two CpG sites of the *SCD* gene and *SCD* expression. Similar results were observed for *PHGDH* and *ABCG1* where a CpG site (cg16246545 and cg06500161, respectively) was negatively associated with expression of the respective gene (**Table 8**).

Table 8: Results of gene expression analyses of CpG sites which were associated with NMR-measured metabolites in KORA F4

CpG	Gene	Gene of transcript	Distance (bp)*	β coef†	p-value
cg16246545	*PHGDH*	*PHGDH*	0	-0.522	2.24E-24
cg11250194	*FADS2*	*FADS1*	17408	-0.331	9.68E-15
cg11250194	*FADS2*	*FADS2*	0	-0.145	2.77E-07
cg19610905	*FADS2*	*FADS1*	11804	-0.439	2.76E-08
cg19610905	*FADS2*	*C11ORF10*	36248	-0.269	7.11E-06
cg19610905	*FADS2*	*FADS2*	0	-0.210	4.82E-05
cg01400685	*FADS2*	*FADS1*	13496	-0.356	3.85E-11
cg27386326	NA‡	*FADS1*	3451	0.315	5.54E-07
cg24503796	*SCD*	*SCD*	0	-0.306	1.92E-08
cg03440556	*SCD*	*SCD*	0	-0.121	5.32E-07
cg06500161	*ABCG1*	*ABCG1*	0	-0.515	2.61E-07

*Distance (bp): distance between CpG site and beginning position of the transcript measured in base pairs †β coef: β coefficient ‡no gene annotation for this CpG according to the UCSC Genome Browser; CpG site is located in an intergenic region between the *FADS1* and *FADS2* gene; Level of significance: 6.3E-05

4.2.4 Associations between lipids and CpG sites located in the *FADS2* gene were confounded by methQTLs

To explore whether the identified associations between DNA methylation and metabolites (N=118) have a genetic background, associations in KORA F4 were tested on confounding by SNPs.

It was first determined whether SNPs, that are located within 1 Mb of each of the 15 CpG sites, were associated with the methylation of the respective CpG site. This analysis revealed that in total 6,607 SNPs were associated with methylation of at least one of the 15 CpG sites (level of significance: 0.01). 741 of those SNPs were also significant associated with at least one metabolite which was in turn also associated with methylation of the respective CpG site. Next, the original regression models were re-run where it was additionally adjusted for each SNP which showed associations with methylation of at least one of the 15 CpG sites and metabolites. 23 of the original 118 CpG-metabolite associations became insignificant (p-value>4.73E-10) with addition of one of these SNPs to the regression model indicating genetic confounding (**Table 9**). Eleven of the associations showed a drastic increase in p-values (e.g. from p-value=3.3E-28 (EWAS) to p-value=1.00 after adjustment for SNPs), where all affected associations included CpG sites located in the *FADS2* gene or nearby. Here, up to 366 SNPs caused the loss of significance.

The p-values of the other CpG-metabolite associations of the EWAS were just under the level of significance and increased slightly after adjusting for the SNPs, dropping just over the level of significance (**Table 9**).

Table 9: Genetically confounded CpG-metabolite associations

Gene	CpG	Metabolite	EWAS β coef*	EWAS p-value*	SNP β coef†	SNP p-value†	N SNPs‡	SNP causing greatest effect
ABCG1	cg06500161	M-VLDL-CE	2.444	5.36E-11	2.327	4.90E-10	2	rs234720
SCD	cg24503796	DHA	1.248	1.55E-10	1.221	5.88E-10	2	rs148785024
MIR33B	cg20544516	S-VLDL-C_%	-0.975	2.42E-11	-0.912	6.91E-10	6	rs1110353
ABCG1	cg27243685	L-VLDL-TG	7.396	4.14E-10	7.445	7.69E-10	1	rs200703373
ABCG1	cg27243685	M-VLDL-PL	4.632	4.91E-11	4.384	9.80E-10	1	rs180959858
ABCG1	cg27243685	M-VLDL-L	4.916	6.58E-11	4.644	1.36E-09	1	rs180959858
ABCG1	cg27243685	M-VLDL-P	4.990	6.73E-11	4.712	1.39E-09	1	rs180959858
ABCG1	cg27243685	S-LDL-TG	3.004	1.49E-10	2.899	1.40E-09	3	rs183163607
PHGDH	cg16246545	FAw6toFA	0.296	3.62E-10	0.296	1.83E-09	1	rs79900295
ABCG1	cg27243685	M-VLDL-TG	5.579	1.30E-10	5.255	2.73E-09	1	rs180959858
ABCG1	cg27243685	MUFA	2.045	2.61E-10	1.943	4.36E-09	1	rs183163607
SCD	cg03440556	DHAtoFA	0.461	4.16E-10	0.433	6.11E-09	6	rs35329375
FADS2	cg19610905	FAw3	2.950	2.46E-32	0.779	0.03#	65	rs174562
FADS2	cg01400685	DHA	1.588	1.93E-13	0.320	0.18#	159	rs174538
FADS2	cg01400685	FAw3	1.325	4.52E-14	0.225	0.25#	149	rs174555
FADS2	cg01400685	DHAtoFA	1.481	2.07E-16	0.160	0.41#	139	rs174538
FADS2	cg11250194	DHA	1.483	9.13E-18	0.086	0.69#	135	rs174550
FADS2	cg11250194	FAw3	1.264	1.63E-19	0.062	0.72#	100	rs174546
FADS2	cg01400685	FAw3toFA	1.219	1.80E-17	0.035	0.82#	137	rs174549
FADS2	cg11250194	DHAtoFA	1.479	1.36E-24	0.020	0.91#	84	rs174550
NA§	cg27386326	DHAtoFA	-1.307	2.22E-10	0.010	0.97#	366	rs174559
NA§	cg27386326	FAw3toFA	-1.148	2.18E-12	-0.005	0.98#	203	rs174559
FADS2	cg11250194	FAw3toFA	1.260	3.30E-28	0.000	1.00#	78	rs174561

*EWAS β coef/p-value: β coefficient and p-value of the EWAS in KORA F4; †SNP β coef/p-value: β coefficient and p-value after adjusting for the SNP which caused the greatest effect on the CpG-metabolite associations; level of significance: 4.73E-10; ‡N SNPs: number of SNPs which, when included as covariates in the statistical model, causes the association to lose significance; §no gene annotation for this CpG according to the UCSC Genome Browser; #drastic increase of p-values after addition of the SNP to the regression model

4.2.5 Methylation of CpG sites located in *ABCG1*, *CPT1A* and *TXNIP* were associated with lipid-related diseases

To investigate the relevance of the results for lipid-related diseases, it was tested whether the CpG sites, which showed associations with metabolites, were also associated with prevalent myocardial infarction and prevalent type 2 diabetes. Analyses were done in KORA

F4, where N=54 myocardial infarction cases and N=78 subjects with type 2 diabetes were available.

The methylation of the CpG site cg06500161 in the *ABCG1* gene was positively associated with myocardial infarction (odds ratio 1.86; 95% confidence interval=1.36-2.53; **Appendix Table IX**). The association lost its significance when additionally adjusting for intake of lipid lowering drugs in the second statistical model (odds ratio 1.37; confidence interval=0.97-1.92; **Appendix Table IX**).

CpG sites in *TXNIP* (cg19693031) and *CPT1A* (cg00574958) showed a significant negative association with prevalent type 2 diabetes (*TXNIP*: odds ratio 0.59; 95% confidence interval=0.49-0.71; *CPT1A*: odds ratio 0.73; 95% confidence interval=0.6-0.89), which was independent of the intake of lipid lowering drugs (**Appendix Table X**).

4.3 Functional analysis of lipid-associated CpG sites

As DNA methylation could have an effect on biological processes, two selected lipid-associated CpG sites were functionally analyzed. One of the CpG site was cg06500161 located in the *ABCG1* gene. It was chosen because it showed the strongest association with both HDL-C and TG levels in the EWAS of main blood lipid levels. The second CpG site, cg20544516 (*MIR33B/SREBF1*), was selected because of its functionally interesting location in the *SREBF1* gene in a region coding for a microRNA (miR-33b).

To assess whether the methylation on this two CpG sites have an influence on protein-DNA binding affinity, electrophoretic mobility shift assays were carried out. Moreover, it was also of interest to investigate whether the methylation status of the two CpG sites as well as the methylation of the surrounding CpG sites causes alterations in transcriptional activity. For this purpose Luciferase Reporter Assays were performed.

4.3.1 Functional analysis of the CpG site cg06500161 (*ABCG1*)

4.3.1.1 Methylation of cg06500161 impairs protein-DNA binding affinity

To investigate cg06500161-methylation-specific protein-DNA binding, electrophoretic mobility shift assays (EMSAs) were carried out. For this purpose THP-1, a monocytic cell line derived from peripheral blood, was considered to be the appropriate cell line as the EWAS

were performed using DNA methylation data obtained from whole blood samples. THP-1 nuclear extract, Cy5-labeled probes and either no or rising concentrations of unlabeled competitor probes were incubated and the protein-DNA complexes were then separated on a polyacrylamide gel by electrophoresis (see also **chapter 2.8**).

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Labeled Probe	um	m				um							m			
Competitor	-	-	-	10x m	100x m	10x um	100x um	10x Oct1	100x Oct1	-	10x um	100x um	10x m	100x m	10x Oct1	100x Oct1
THP-1 NE	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
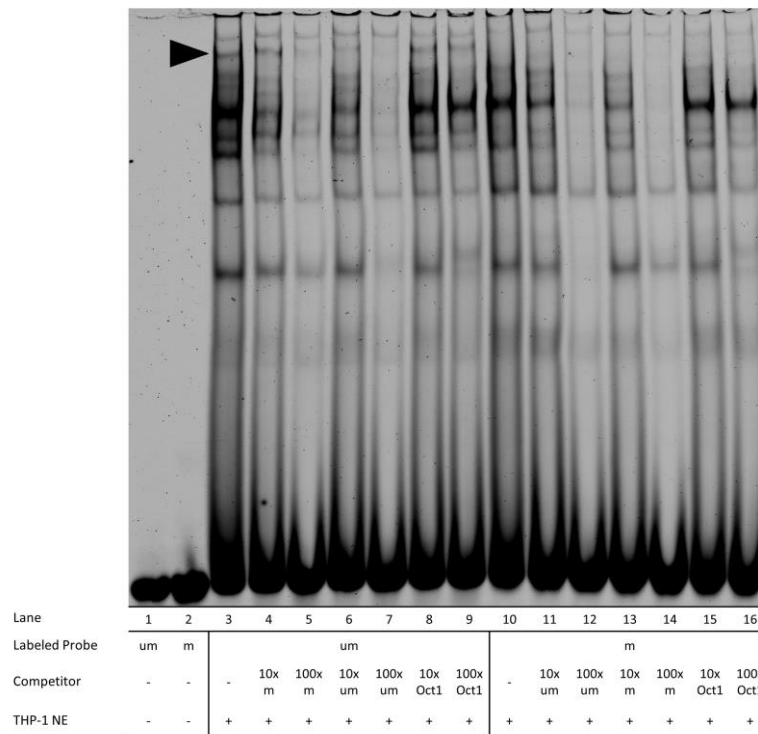
Figure 10: Methylation of cg06500161 (*ABCG1*) lowered the DNA binding affinity of an unknown protein or protein complex in THP-1 nuclear extracts. Labeled probes carrying a methylated or unmethylated cytosine on the position of cg06500161 revealed differential protein binding (lane 3 vs. lane 10, indicated by arrowhead). Competition was performed with unlabeled methylated probes (lane 4, 5, 13, 14) and with unlabeled unmethylated probes (lane 6, 7, 11, 12). For competition different concentrations were used (10 and 100 ng/µl). For verification of specificity, competitions with an unlabeled Oct1 consensus probe (lane 8, 9, 15, 16) were done. Probes without nuclear extract did not reveal any protein-DNA binding (lanes 1-2). NE nuclear extract; um unmethylated; m methylated; Oct1 = octamer binding transcription factor 1.

A higher binding affinity for a yet unidentified protein complex in the presence of the Cy5-labeled unmethylated cg06500161 probe (cg06500161_um) was identified, which was not visible when cg06500161 was methylated (cg06500161_m; **Figure 10** lane 3 vs. lane 10, indicated by arrowhead). With increasing amounts of unlabeled cg06500161_um competitor DNA an almost complete displacement of the protein-DNA complex was achieved, the band almost vanished **(Figure 10** lane 6 and 7). On the other hand, a high concentration (100 ng/µl) of the competitor cg06500161_m could only slightly decrease the

intensity of the protein-DNA complex (**Figure 10** lane 5). Specificity of binding was verified with an unlabeled Oct1 consensus probe which showed no visible influence on the binding affinity of the protein complex (**Figure 10** lane 8, 9).

4.3.1.2 Motif search for Luciferase Reporter Assays

As the EMSA indicated a methylation-dependent formation of a protein-DNA complex for cg06500161, potential alterations in the transcriptional activity caused by the methylation status of the CpG site and the CpG site-surrounding sequence were analyzed.
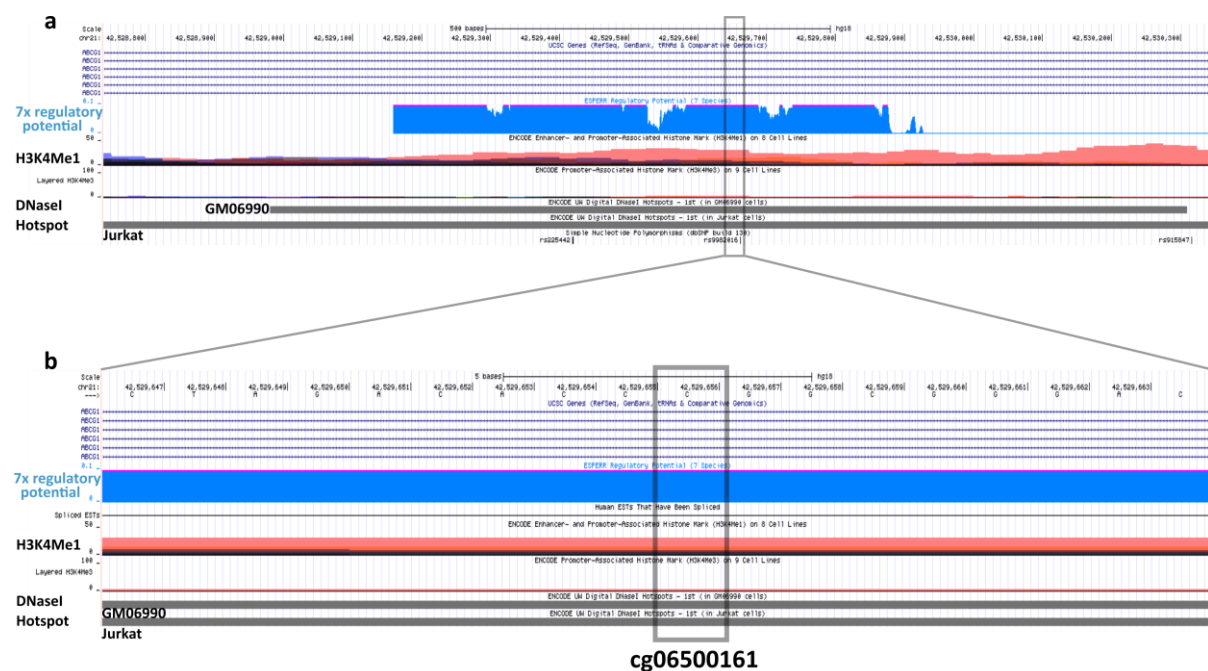
Figure 11: Region of DNaseI Hotspot of Jurkat cells in the *ABCG1* gene containing the CpG site cg06500161 on human chromosome 21. a) A high 7x Regulatory Potential is predicted (blue colored area) for the shown region (NCBI36/hg18, chr21:42,528,740–42,530,346). It also overlaps a potential enhancer region (pink curves). **b)** The CpG site cg06500161 is located in a region which was described in two leucocyte cell lines (Jurkat and GM06990) to be covered by DNaseI Hotspots (indicated by grey bars).

The CpG site cg06500161 is located on human chromosome 21 (**Figure 11**). The locus possesses a high 7x Regulatory Potential (regulatory potential scores computed from seven mammals; UCSC NCBI36/hg18) and data from the ENCODE project provided evidence for DNaseI hypersensitive sites in GM06990 (human lymphoblastoid cells) and Jurkat cells (human T lymphocyte cells), indicating regulatory regions. The locus overlaps also a potential enhancer region indicated by the ENCODE enhancer- and promoter-associated

histone mark H3K4Me1 which is associated with enhancers and with DNA regions downstream of transcription starts (**Figure 11**).

To investigate the influence of the methylation status of the CpG site cg06500161 (located in *ABCG1*) on promoter activity, DNA fragments of three different sizes were cloned in front of the human minimal promoter EF-1α which controls the luciferase gene of the CpG-free luciferase plasmid (pCpGfree-promoter-Lucia, Invivogen). The longest DNA fragment covered the sequence of the DNaseI Hotspot of Jurkat cells, which also overlaps the DNaseI Hotspot of GM06990 cells. GM06990 and Jurkat cells are leucocytes, just as THP-1 cells which were used for the EMSA experiments. The DNA fragment had a length of 1607 bp and contained 29 CpG sites. The construct carrying this insert is further designated as *ABCG1*_1607 bp. The two other DNA fragments were smaller parts of the long insert (518 bp and 104 bp long containing 7 and 2 CpG sites, respectively), in which the CpG site cg06500161 was positioned nearly in the center (further designated as *ABCG1*_518 bp and *ABCG1*_104 bp). **Table 3** in **chapter 2.9.1** provides an overview of the used pCpGfree-promoter-Lucia plasmid constructs.

To obtain methylated constructs, plasmids were treated with two different methyltransferases, HpaII and M.SssI. HpaII recognizes a specific sequence and modifies only the internal cytosine residue of this sequence (5`…CCGG…3′). As the CpG site cg06500161 and its directly adjacent bases correspond to the enzyme-specific sequence, cg06500161 could be selectively methylated. HpaII was used for the constructs containing the two shorter inserts (*ABCG1*_104 bp and *ABCG1*_518 bp). It was not suitable for the third construct (*ABCG1*_1607 bp) as this construct carries the enzyme-specific sequence more than once. In contrast, the methyltransferase M.SssI, which methylates all available CpG sites, was used for all three constructs. Additional samples of the same constructs were processed the same way but without enzymes and substrates, and were considered as unmethylated constructs. To examine whether the methylation status alone or the orientation of the inserts are essential for changes of promoter activity as well, constructs with the corresponding reverse complementary (_rc) inserts were included in the assay. Luciferase Reporter Assays were performed in three different cell lines, THP-1, HepG2 and HEK293. THP-1 was chosen as the methylation-specific protein binding was shown in EMSA experiments using THP-1 nuclear extracts. HepG2 cells, a liver cell line, were selected due to

the relation of the CpG site cg06500161 located in the *ABCG1* gene to the lipid metabolism. The human kidney cell line HEK293 was selected as a control cell line. Measurements were done in triplicates (=technical replicates) and experiments were replicated at least three times (=biological replicates). A detailed description of the experimental design is provided in **chapter 2.9.1**.

4.3.1.3   Influence of the methylation status of cg06500161 on reporter gene expression

Aim of the analyses was to examine the effect of the methylation of the CpG site cg06500161 on the transcriptional activity of the minimal promoter EF-1α and thus on the expression of the luciferase reporter gene. Therefore, Luciferase Reporter Assays were conducted using the *ABCG1*_104 bp and *ABCG1*_518 bp constructs, which were methylated by the HpaII methyltransferase that selectively methylated the CpG site of interest.

Lucia luciferase activity of the pCpGfree-promoter-Lucia vector was normalized against the firefly luciferase activity of the control vector pGL4.53 to get the relative promoter activity of the constructs. To make the overall effect of the different inserts on transcriptional activity visible, the normalized fold change was calculated relative to the luciferase activity of the empty pCpGfree-promoter-Lucia vector. The relative light unit of the empty vector was therefore set to 100% (see **Figure 12a**).

In THP-1 and HepG2 cells the inserts of the constructs *ABCG1*_104 bp and *ABCG1*_518 bp had a repressor function on the tested minimal promoter EF-1α (up to 71% decrease in luciferase activity) compared to the empty vector, however, independent of the methylation status of cg06500161 (**Figure 12b and c**, shaded bars). In contrast, in HEK293 cells the inserts had no significant effect on the minimal promoter. Similar results were observed for the constructs carrying the corresponding reverse complementary inserts (_rc-constructs) except in HepG2 where *ABCG1*_104 bp induced almost no change in luciferase activity and *ABCG1*_518 bp induced a methylation-dependent decrease of luciferase activity compared to the empty vector (**Figure 12b and c**, unfilled bars).

To determine the differences between the effects of methylated and unmethylated constructs on transcriptional activity, the activity of the methylated constructs was calculated relative to the activity of the unmethylated constructs. In THP-1 cells, *ABCG1*_104 bp- and *ABCG1*_518 bp-constructs methylated at cg06500161 induced a

significant stronger decrease of luciferase activity than unmethylated constructs (66% versus 39% and 71% versus 60% decrease compared to the empty vector, respectively; **Figure 12b and c**, shaded bars; significance of p-values is indicated by curly brackets). In HepG2 cells only the *ABCG1*_518 bp-constructs induced a cg06500161 methylation-dependent decrease of luciferase activity (**Figure 12c**, shaded bars, curly brackets). Methylation of cg06500161 in each construct had no influence on luciferase activity in HEK293 indicating a cell-type dependent effect (**Figure 12b and c**, shaded bars, curly brackets).

To examine whether the methylation status alone or the orientation of the inserts are essential for changed promoter activity in THP-1 and HepG2 cells, the effects of the _rc-constructs and the forward constructs were compared with each other. The effect sizes of the _rc-constructs were significantly different to the forward constructs suggesting an orientation-specific effect (**Figure 12b and c**, unfilled bars, significance of p-values is indicated by square brackets). Only the effect of the unmethylated *ABCG1*_104 bp-construct in THP-1 was not significantly different to the effect of the respective unmethylated _rc-construct (**Figure 12b**, square bracket).

In summary, a cg06500161-methylation-specific decreasing effect on promoter activity was shown, which was specific for the orientation of the insert as well as cell type-specific for THP-1 and HepG2 cells.
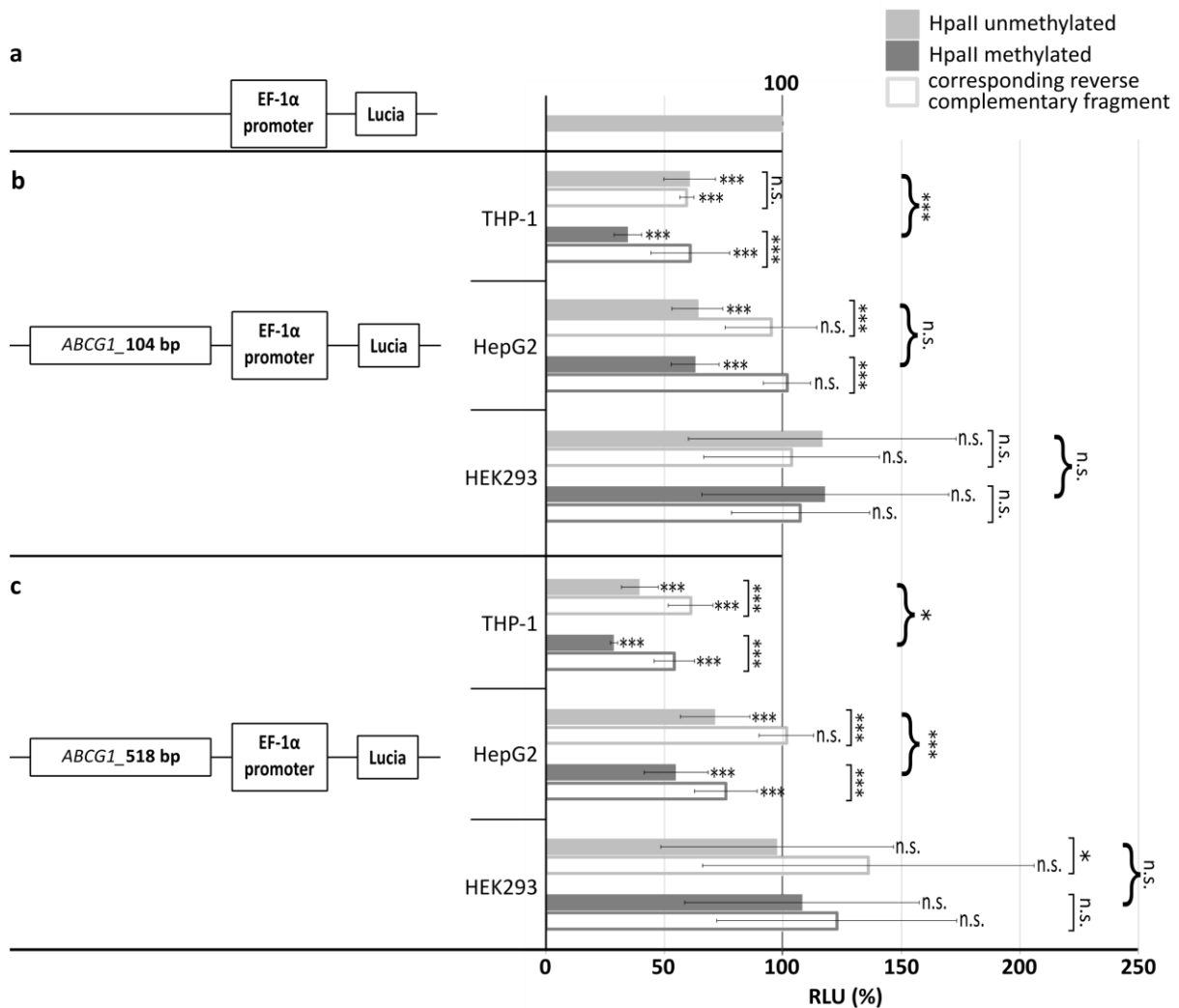
Figure 12: Methylation of cg06500161 decreased promoter activity in a cell-specific manner. Assays were performed in THP-1, HepG2 and HEK293 cells with minimal promoter plasmids **a)** without any insert, **b)** plasmids with *ABCG1*_104 bp fragment (replication of experiments: THP-1 N=5, HepG2 N=6, HEK293 N=3) and **c)** *ABCG1*_518 bp fragment (THP-1, HepG2 N=7, HEK293 N=3) containing cg06500161 at mid-position. Significance of p-values are given for each construct in comparison with the empty luciferase vector, unless otherwise noted by brackets:] / } indicate significance of p-values of differences between forward and _rc-constructs / between unmethylated and methylated forward constructs. P-values were derived from linear mixed effects models and corrected for multiple testing using Bonferroni method. * p<0.05, ** p<0.01, *** p<0.001, n.s.=p-value not significant; RLU=relative light unit.

4.3.1.4 Influence of the methylation status of cg06500161 and its adjacent CpG sites on reporter gene expression

Aim of the analyses was to examine the effect of the methylation of cg06500161 and its adjacent CpG sites on transcriptional activity of the minimal promoter EF-1α and thus on the expression of the luciferase reporter gene. Therefore, Luciferase Reporter Assays were

conducted using the *ABCG1*_104 bp-, the *ABCG1*_518 bp- and the *ABCG1*_1607 bp-constructs, which carried 2, 7 and 29 CpG sites, respectively. All CpG sites in the constructs were methylated by the M.SssI methyltransferase.

Lucia luciferase activity was normalized against the firefly luciferase activity to get the relative promoter activity of the constructs. To make the overall effect of the different inserts on transcriptional activity visible, the normalized fold change was calculated relative to the luciferase activity of the empty vector. The relative light unit of the empty vector was therefore set to 100% (see **Figure 13a**).

In THP-1 and HepG2 cells the inserts of the *ABCG1*_104 bp- and *ABCG1*_518 bp-constructs had a repressive effect on the tested minimal promoter (up to 62% decrease in luciferase activity; shaded bars in **Figure 13b and c**). In contrast, the insert of the *ABCG1*_1607 bp-construct showed an enhancing effect which was significantly different between the unmethylated constructs and the empty vector (**Figure 13d**, shaded bars). In HepG2 cells, the enhancing function of the unmethylated *ABCG1*_1607 bp-construct was very strong and highly significant (1599% increase of promoter activity; **Figure 13d**, shaded bars).

In HEK293, which was used as control cell line, the overall effect of the different constructs was the other way round: the methylated *ABCG1*_104 bp-construct acted as enhancer (**Figure 13b**, shaded bars) whereas the *ABCG1*_1607 bp-construct showed a methylation-independent repressive function on promoter activity (**Figure 13d**, shaded bars). The *ABCG1*_518 bp construct did not induce a significant change in luciferase activity (**Figure 13c**, shaded bars).

To determine the differences between the effects of methylated and unmethylated constructs on transcriptional activity, the activity of the methylated constructs was calculated relative to the activity of the unmethylated constructs.

In THP-1 cells, a significant methylation-dependent enhancer effect was visible for the insert of the *ABCG1*_1607 bp-construct which was lower for the methylated insert than for the unmethylated one (**Figure 13d**, shaded bars; significance of p-value is indicated by a curly bracket).

In HepG2 cells, methylation strengthened significantly the repressive effect of the *ABCG1*_518 bp-construct (**Figure 13c**, shaded bars; curly brackets). A very strong methylation-dependent effect was observed for the *ABCG1*_1607 bp-construct. Methylation

of the CpG sites sharply reduced the enhancing effect (up to 1448% decrease of luciferase activity) (**Figure 13d**, shaded bars; curly brackets). In both cell lines, THP-1 and HepG2, the effect of the *ABCG1*_104 bp-construct was the other way round: unmethylated constructs showed a (significant) stronger repressive effect than methylated ones (**Figure 13b**, shaded bars; curly bracket).

In HEK293 cells a methylation-dependent effect was detected only for the *ABCG1*_104 bp-construct. Here, the methylated construct showed a stronger enhancer role than the unmethylated one (**Figure 13b**, shaded bars; curly bracket).

To examine whether the methylation status alone or the orientation of the inserts are essential for changed promoter activity, also the effects of the _rc-constructs and the forward constructs were compared with each other. This was especially interesting for the constructs which have shown a significant methylation-dependent transcriptional effect (indicated by curly brackets and stars in **Figure 13**). For these constructs the respective _rc-constructs induced also effects on transcriptional activity (_rc-constructs are represented by unfilled bars in **Figure 13**). However, the effect sizes between each construct and its corresponding _rc-construct were significantly different, except for the effect of the *ABCG1*_104 bp-construct in HEK293 (**Figure 13b,** significance of p-values is indicated by square brackets) and for the effect of the *ABCG1*_1607 bp-construct in HepG2 cells (**Figure 13d**).
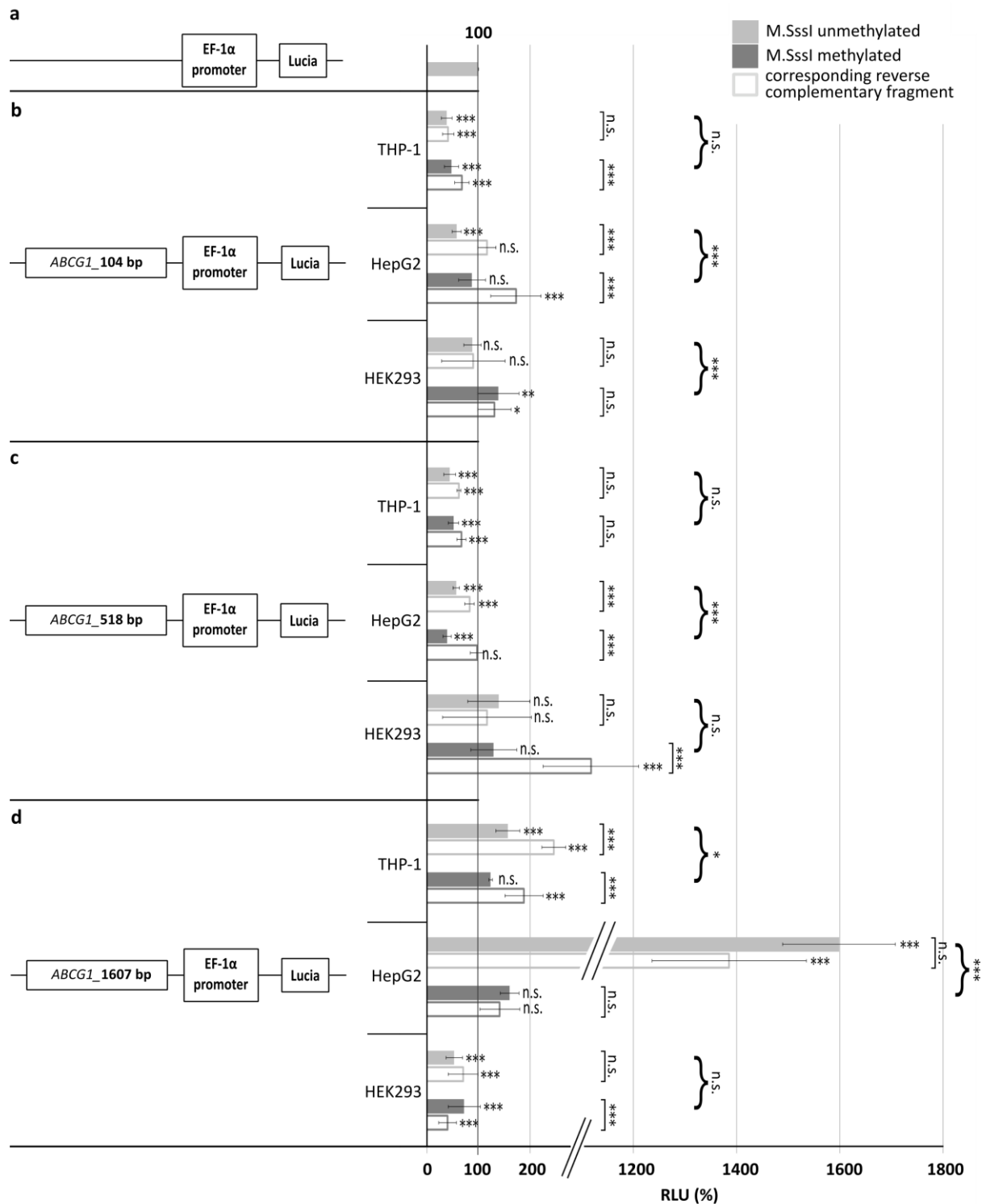
Figure 13: Methylation of cg06500161 and its adjacent CpG sites induced a cell type- and construct-specific effect on promoter activity. Assays were performed in THP-1, HepG2 and HEK293 cell lines with minimal promoter vectors (a) without any inserts, (b) vectors with *ABCG1*_104 bp fragments (replication of experiments: THP-1, HepG2 N=6, HEK293 N=3), (c) *ABCG1*_518 bp (THP-1 N=6, HepG2 N=7, HEK293 N=3) fragments and (d) *ABCG1*_1607 bp (THP-1, HepG2, HEK293: N=3) fragments, containing cg06500161 in mid-position. P-values are given for each construct in comparison with the empty luciferase vector, unless otherwise noted by brackets:] / } indicate significance of p-values of differences between forward and _rc-constructs/ between unmethylated and methylated forward constructs. P-values were derived from linear mixed effects models and corrected for multiple testing using Bonferroni method. * p<0.05, ** p<0.01, *** p<0.001, n.s.=p-value not significant; RLU=relative light unit.

Luciferase Reporter Assays revealed that the methylation of cg06500161 and its adjacent CpG sites had an effect on promoter activity. This effect was cell type-specific; in HepG2 cells the *ABCG1*_1607 bp-construct, which comprises 29 CpG sites, showed a strong enhancing effect which was sharply reduced when the construct was methylated. Similar results were observed in THP-1 cells, but the overall enhancing effect of the *ABCG1*_1607 bp-construct as well as the repressive effect of the methylation was much weaker than in HepG2 cells. The effect was orientation-dependent in the THP-1 cells, but not in the HepG2 cells. The methylation-dependent effect on promoter activity was also dependent on the length of the inserts and thus the number of CpG sites. In HepG2 cells, all three construct had a methylation-dependent effect which differed in effect size. Methylation of the construct carrying the insert with 29 CpG sites induced a much stronger repressor effect than inserts with less (2 or 7) CpG sites. In THP-1, a methylation-dependent effect on promoter activity was only observed with the *ABCG1*_1607 bp-construct. In HEK293, the control cell line, methylation of *ABCG1*_104 bp-constructs had an orientation-independent enhancer role, which was not seen for the other two constructs.

To summarize, methylation of cg06500161 and its adjacent CpG sites had an effect on promoter activity. The effect was cell type-specific and the direction of the effect and the effect size differed dependent on the length and thus the number of CpG sites of the constructs, and the orientation of the inserts.

4.3.1.5 Computational analysis of the CpG site cg06500161

To identify transcription factors whose binding sites overlap the CpG site cg06500161 and which therefore may bind methylation-dependent, a computational analysis was performed using the Genomatix MatInspector software with standard settings (Cartharius *et al.* 2005). To predict transcription factor binding sites (TFBS), this program utilizes a library of position weight matrices derived form a large set of known binding sites from several publications.

Analysis of the sequence of the oligonucleotide used for the EMSA (5´ CCTTCTCTAGACACC**CG**GCGGGACTAGTTCCT 3´) revealed that the CpG site cg06500161 was directly located in the binding sites of the following proteins: Sp2 and the Zinc finger protein of the cerebellum (Zic3) on the forward strand as well as the erythroid krueppel like factor (EKLF) and tumor suppressor p53 on the reverse strand. The binding site of another

transcription factor, the E2F transcription factor 1 (E2F1), was located right next to cg06500161 and had the highest matrix similarity (0.98) (**Table 10**). The matrix similarity is an index of the overlap of the candidate sequence and the most conserved nucleotides of the transcription factor binding site.

Table 10: **Potential transcription factors whose binding sites involve the CpG site cg06500161 or cg20544516**

CpG site	Transcription factor	Tissue	Strand*	Matrix similarity†	Consensus sequence of the TFBS in the oligonucleotide sequence‡
cg06500161	E2F transcription factor 1	ubiquitous	+	0.98	acc**c**GGCGggactagtt
	Sp2	ubiquitous	+	0.927	acacc**c**ggcgGGACtag
	Zinc finger protein of the cerebellum (Zic3)	Brain, Central Nervous System, Embryonic Structures, Eye, Neurons	+	0.913	acaccCGGCgggact
	Erythroid krueppel like factor (EKLF)	Blood Cells, Bone, Marrow Cells, Embryonic Structures, Erythrocytes, Hematopoietic System	-	0.903	ctagtcccgc**c**GGGTgtct
	Tumor suppressor p53	ubiquitous	-	0.772	aggaaCTAGtcccgc**c**gggtgtcta
cg20544516	PAX-5	Antibody-Producing Cells, Endocrine System, Immune System, Thyroid Gland	+	0.806	ccgggctgCActgc**c**gaggcactgcaccc

*forward (+) or reverse (-) DNA strand; †Matrix similarity = 1 when the candidate sequence corresponds to the most conserved nucleotide at each position of the matrix; ‡underlined letters = consensus sequence of transcription factor binding sites; CAPITAL LETTERS = core binding motif; small underlined letters = conserved binding motif; red letters = position of the CpG site

4.3.2   Functional analysis of the CpG site cg20544516 (*MIR33B/SREBF1*)

4.3.2.1   Methylation-specific protein-DNA binding for cg20544516

To investigate methylation-dependent protein-DNA interaction, EMSA experiments with cg20544516-methylated and -unmethylated Cy5-labeled probes were performed using nuclear protein extracts of THP-1 cells. Similar to the EMSA experiments for cg06500161, THP-1 nuclear extract, Cy5-labeled probes and either no or rising concentrations of unlabeled competitor probes were incubated and the protein-DNA complexes were then separated on a polyacrylamide gel by electrophoresis (see **chapter 2.8**).

A binding affinity for a yet unidentified protein complex was visible for Cy5-labeled probes with methylated cg20544516 (cg20544516_m), which was not visible for unmethylated cg20544516 probes (cg20544516_um, **Figure 14** lane 10 vs. lane 3, indicated by arrowhead). Competition was performed with unlabeled methylated probes, showing a decrease of the intensity of the protein-DNA complex when using a high amount of competitor (100 ng/µl) (**Figure 14** lane 14). There was no visible decrease of the intensity for unlabeled unmethylated competitors (**Figure 14** lane 11, 12). Specificity of binding was verified with an unlabeled Oct1 consensus probe which showed no visible influence on the binding affinity of the protein complex (**Figure 14** lane 15, 16).

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Labeled Probe	um	m				um							m			
Competitor	-	-	-	10x m	100x m	10x um	100x um	10x Oct1	100x Oct1	-	10x um	100x um	10x m	100x m	10x Oct1	100x Oct1
THP-1 NE	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
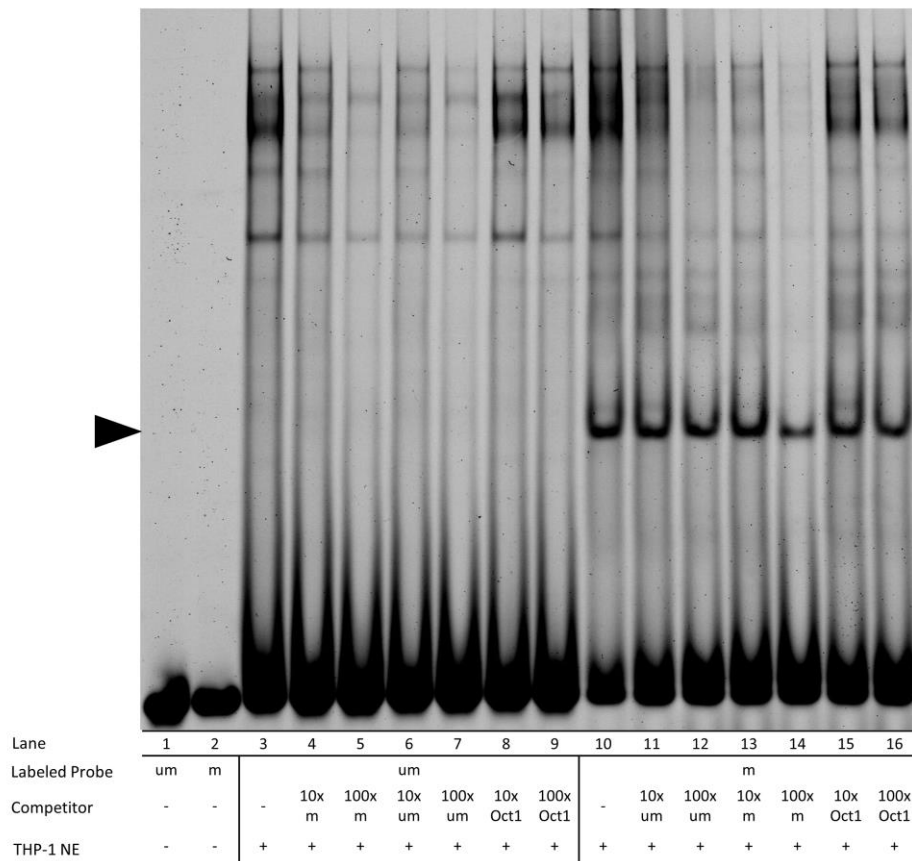
Figure 14: Methylation of cg20544516 (*MIR33B/SREBF1*) enhanced the binding affinity of an unknown protein or protein complex in THP-1 nuclear extracts. Labeled probes carrying a methylated or unmethylated cytosine on the position of cg20544516 revealed differential protein binding (lane 3 vs. lane 10, indicated by arrowhead). Competition was performed with unlabeled methylated probes (lane 4, 5, 13, 14) and with unlabeled unmethylated probes (lane 6, 7, 11, 12). For competition different concentrations were used (10 and 100 ng/µl). For verification of specificity, competitions with an unlabeled Oct1 consensus probe (lane 8, 9, 15, 16) were done. Probes without nuclear extract did not reveal any protein-DNA binding (lanes 1-2). NE nuclear extract; um unmethylated; m methylated; Oct1 = octamer binding transcription factor 1.

4.3.2.2 Motif search for Luciferase Reporter Assays

Results of the EMSAs showed a cg20544516 methylation-specific protein-DNA binding. It was therefore further investigated whether methylation of this CpG site and its adjacent CpG sites also have an effect on transcriptional activity.

The CpG site cg20544516 is located on human chromosome 17 in the gene *SREBF1* at a locus which also encodes the microRNA 33b. The region also possesses a high 7x Regulatory Potential according to UCSC NCBI36/hg18. For the Luciferase Reporter Assays of cg20544516, the sequence of the DNA fragment, which was cloned into the pCpGfree-promoter-Lucia plasmid, was the same as used for the EMSA. The DNA fragment had a

length of 32 bp and contained in sum three CpG sites. The construct carrying this fragment is further designated as *MIR33B/SREBF1*_32 bp. The second construct carried a DNA fragment which was a smaller part of the first one. It had a length of 24 bp and contained only the CpG of interest (further designated as *MIR33B/SREBF1*_24 bp). **Table 3** in **chapter 2.9.1** provides an overview of the used pCpGfree-promoter-Lucia constructs.

To obtain methylated constructs, plasmids were treated with the methyltransferase M.SssI. Additional samples of the same constructs were processed the same way but without enzymes and substrates, and were considered as unmethylated constructs. To examine whether the methylation status alone or the orientation of the fragments are essential for changes of promoter activity as well, constructs with the corresponding reverse (_r) DNA fragment sequence were included in the assay. Luciferase Reporter Assays were performed in three different cell lines, THP-1, HepG2 and HEK293. THP-1 was chosen because the methylation-specific protein binding was shown in EMSA experiments using THP-1 nuclear extracts. HepG2 cells, a liver cell line, were selected due to the relation of the CpG site cg20544516 located in the *SREBF1* gene to the lipid metabolism. The human kidney cell line HEK293 was selected as a control cell line. Measurements were done in triplicates (=technical replicates) and experiments were replicated at least three times (=biological replicates). A detailed description of the experimental design is provided in **chapter 2.9.1**.

4.3.2.3 Influence of the methylation status of cg20544516 and its adjacent CpG sites on reporter gene expression

Lucia luciferase activity was normalized against the firefly luciferase activity to get the relative promoter activity of the constructs. To make the overall effect of the different inserts on transcriptional activity visible, the normalized fold change was calculated relative to the luciferase activity of the empty pCpGfree-promoter-Lucia vector. The relative light unit of the empty vector was therefore set to 100% (see **Figure 15a**).

In all cell lines the methylated constructs showed significant enhancing effects on the promoter activity compared to the empty vector (**Figure 15b and c**, dark grey bars), which was not the case for the unmethylated constructs (light grey bars). A particularly strong enhancer function was visible for the *MIR33B/SREBF1*_32 bp-construct in HEK293 cells (up to 980% increase of luciferase activity, **Figure 15c**).

To determine the differences between the effects of methylated and unmethylated constructs on transcriptional activity, the activity of the methylated constructs was calculated relative to the activity of the unmethylated constructs. Both constructs (*MIR33B/SREBF1*_32 bp and *MIR33B/SREBF1*_24 bp) induced in all cell lines a significant methylation-dependent effect on promoter activity (**Figure 15b and c**, shaded bars, significance of p-values are indicated by curly brackets).

To examine whether the methylation status alone or the orientation of the inserts are essential for changed promoter activity, the effects of the _r-constructs and the forward constructs were compared with each other. Similar to the forward constructs, a methylation-dependent effect was visible for all _r-constructs (**Figure 15b and c**, unfilled bars). However, for *MIR33B/SREBF1*_24 bp the enhancing effect of the methylated constructs were significantly different to the corresponding _r-constructs in THP-1 and HepG2 cells (**Figure 15b**, significance of p-values is indicated by square brackets). This was similar for the methylated *MIR33B/SREBF1*_32 bp–constructs, which showed a significant different effect than the methylated *MIR33B/SREBF1*_32 bp_r-constructs in all three cell lines (**Figure 15c**, square brackets).

To sum up, methylation of cg20544516 (*MIR33B/SREBF1*) and its adjacent CpG sites showed an insert-orientation-specific enhancing effect on promoter activity. The enhancer function was particularly strong for the *MIR33B/SREBF1*_32 bp-construct in HEK293 cells.
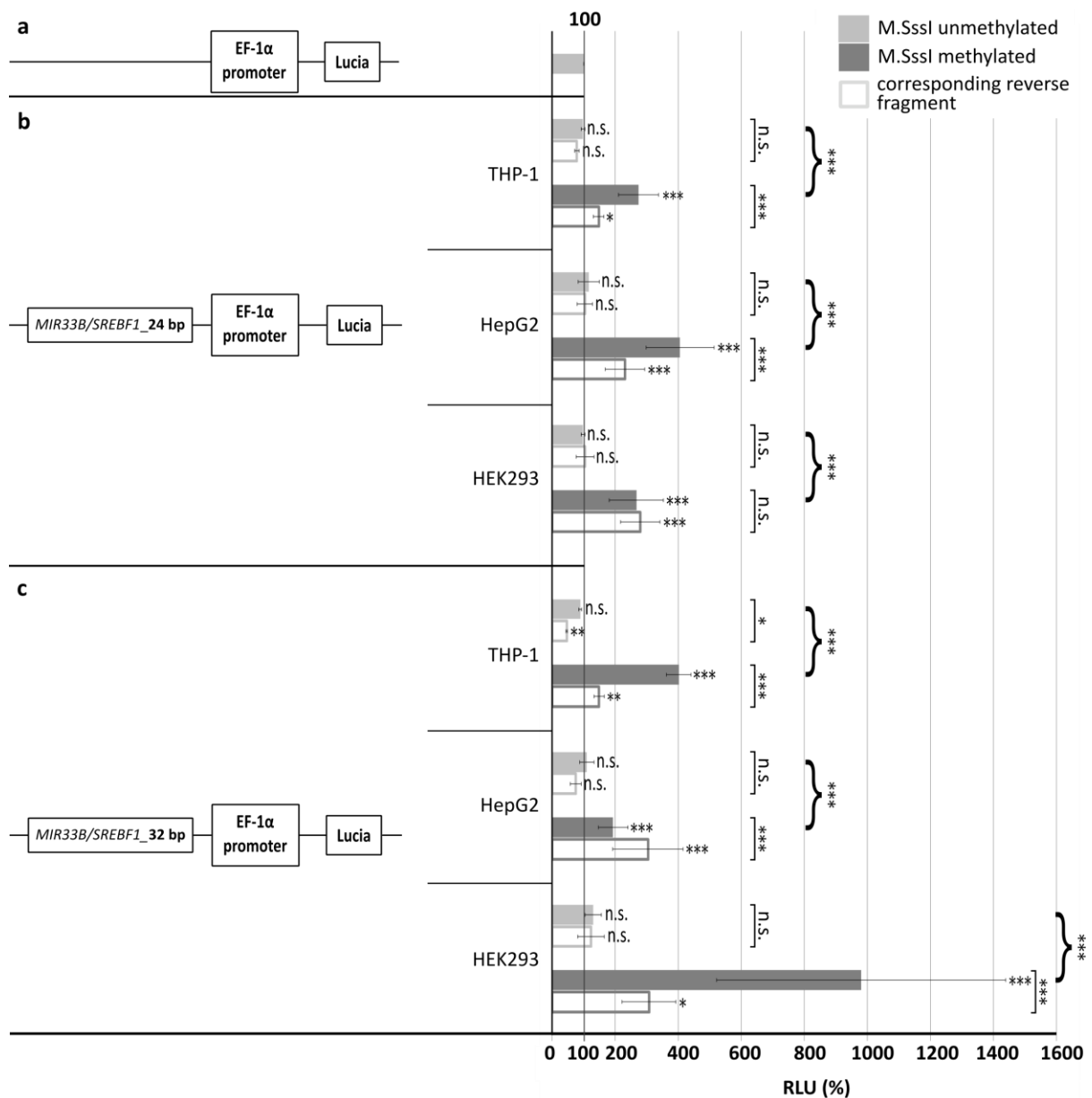
Figure 15: Methylation of cg20544516 and its adjacent CpG sites (*MIR33B/SREBF1*) induced an insert-orientation-specific enhancing effect on promoter activity. Assays were performed in THP-1, HepG2 and HEK293 cell lines with minimal promoter vectors (a) without any inserts, (b) vectors with *MIR33B/SREBF1_*24 bp fragments and (c) *MIR33B/SREBF1_*32 bp fragments. Replication of experiments was N=3 (THP-1 and HEK293) and N=4 (HepG2). P-values are given for each construct in comparison with the empty luciferase vector, unless otherwise noted in brackets:] / } indicate significance of p-values of differences between forward and _r-constructs/ between unmethylated and methylated forward constructs. P-values were derived from linear mixed effects models and corrected for multiple testing using Bonferroni method. * p<0.05, ** p<0.01, *** p<0.001, n.s.=p-value not significant; RLU=relative light unit.

4.3.2.4   Computational analysis of the CpG site cg20544516

To predict possible transcription factors which may bind to the DNA dependent on the methylation status of cg20544516, computational analyses were performed using the Genomatix MatInspector software with standard settings as described in **chapter 4.3.1.5**. Analysis of the sequence of the oligonucleotide used for the EMSA (5´ TCCGGGCTGCACTGC**CG**AGGCACTGCACCCGC 3´) revealed PAX-5 as potential transcription factor binding on the forward strand. The binding site did not directly overlap with cg20544516, the matrix similarity was 0.806 (see **Table 10** in **chapter 4.3.1.5**).

5   DISCUSSION

5.1   Epigenome-wide association studies discovered lipid- and fatty acid-related CpG sites

Previous epigenome-wide association studies (EWAS) have already provided first indications of a relationship between epigenetic changes and blood lipid levels in patients with hypercholesterolemia or in non-population-based cohorts (Guay *et al.* 2012b; Irvin *et al.* 2014). This thesis comprises a systematical approach to investigate associations between genome-wide DNA methylation and blood lipid levels as well as lipoprotein subfractions using data of the population-based cohort KORA F4, with subsequent validation of the results by replication in independent cohorts. The two EWAS revealed in total 17 lipid- and fatty acid-related CpG sites located in a total of ten different genes and two gene locations (**Figure 16**).

In the EWAS of main blood lipid levels seven new lipid-associated CpG sites located in *ABCG1* (HDL-C, TG), *MIR33B/SREBF1*, in an intergenic region between *SCD* and *LINC00263* (TG), and in *TNIP1* (LDL-C) were identified, which have been validated by replication in independent cohorts. In addition, one CpG site was confirmed (cg05574958 in *CPT1A*) which was already found to be associated with TG levels in CD4$^+$ T cells in the GOLDN study (N=991) (Irvin *et al.* 2014). Five of these associations were also found in human adipose tissue of the MuTHER cohort. Here, the strongest association was observed between TG levels and *MIR33B/SREBF1* as well as *ABCG1* DNA methylation. These associations were not significant in skin samples indicating tissue specificity. In contrast, the associations between TG levels and methylation of the CpG site in *SREBF1* and *CPT1A* did at least not differ between skin and adipose tissue.

The EWAS of NMR-measured metabolites revealed after validation 118 associations between 15 different CpG sites and 48 unique metabolites. With the exception of the association between cg07397296 (*ABCG1*) and TG levels and the association between *TNIP1* (cg22178392) and LDL-C, the discovery EWAS in KORA F4 replicated all results of the EWAS of main blood lipids. However, it was not possible to validate them by the meta-analysis of the replication results of the LOLIPOP, NFBC1966 and YFS cohort. Furthermore, the CpG site

located in the intergenic region between *SCD* and *LINC00263* (cg07504977) was associated with TG content of S-HDL particles but not with serum TG levels as observed in the EWAS of main blood lipids. In the EWAS of NMR-measured metabolites, the validated lipid-associated CpG sites of the EWAS of main blood lipids showed additional associations with different lipoprotein subfractions, mainly with several fractions of different-sized VLDL particles and with the TG content in IDL, LDL and HDL particles. Moreover, methylation of the CpG sites located in *ABCG1* and *TXNIP* as well as CpG sites in *FADS2*, *SCD*, *PHGDH* and *SLC7A11* were associated with fatty acids (**Figure 16**), which are not only known as an important source of energy and mediators of inflammation but also as regulators of gene expression (Sampath and Ntambi 2005) (see also **chapter 5.3**). The results of this EWAS could not be replicated in the tissue data of the MuTHER cohort as no NMR data were available.
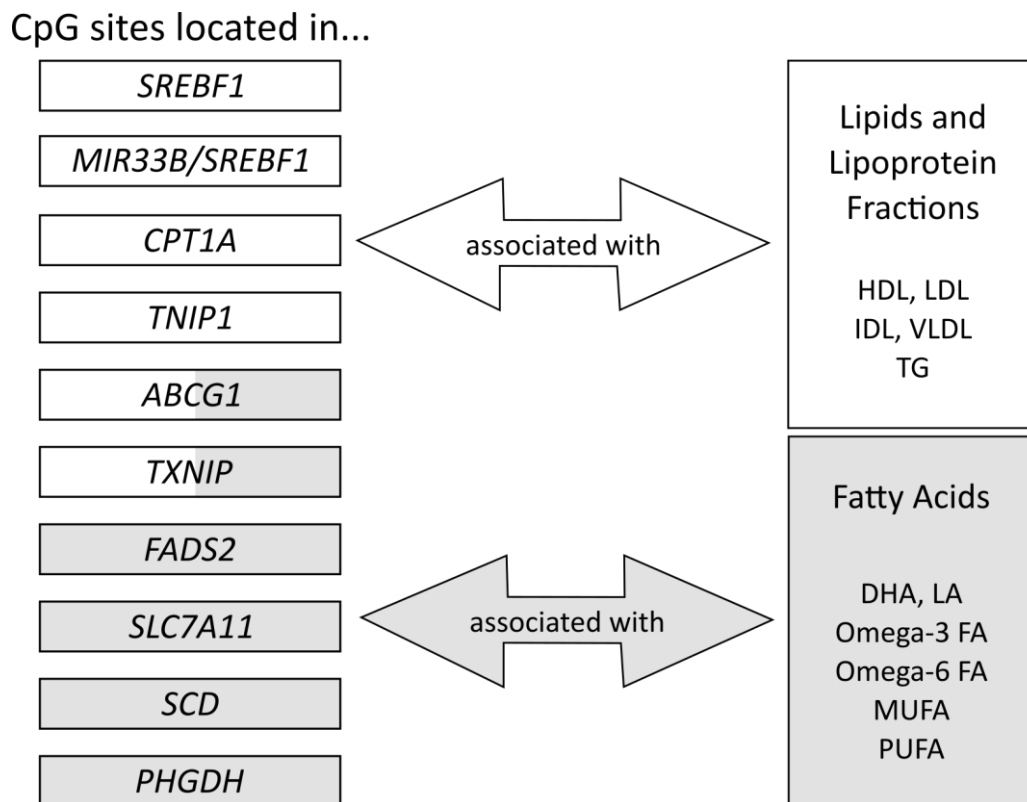
Figure 16: CpG sites located in ten different genes were associated with lipid and fatty acid levels. CpG sites of two genes (*TXNIP*, *ABCG1*) were associated with both lipids (white background) and fatty acids (grey background). TG triglycerides; DHA docosahexaenoic acid; LA linoleic acid; FA fatty acids; MUFA/PUFA mono/poly-unsaturated fatty acids.

All methylation-related metabolites showed a high correlation in KORA F4, most likely due to their metabolic interrelationship. HDL3 particles acquire cholesterol through the

cholesterol transporter ABCG1 resulting in spherical HDL2 particles. Cholesterol esters (CE) of HDL2 are transferred in exchange for TG to VLDL particles via the cholesterol ester transfer protein (CETP). Additionally phospholipids are transferred between HDL and VLDL particles (see also **chapter 1.2.3**). Also the metabolism of fatty acids (FA) and lipoproteins are interlinked. Increased free fatty acid concentrations leads to excessive accumulation of skeletal muscle fat and increased concentrations of TG and CE in the liver. High hepatic liver TG leads to high levels of blood VLDL particles which can induce CETP activity resulting in an increased transfer of TG from VLDL to HDL particles and a subsequent increase in HDL clearance and decreased HDL concentrations (Zivkovic *et al.* 2007).

5.2 Metabolite-related CpG sites are located in genes which are involved in lipid, fatty acid and amino acid metabolism

As mentioned above, the lipid- and fatty acid-related CpG sites are located in a total of ten different genes. All of these genes are directly or indirectly involved in lipid, fatty acid or amino acid metabolism. The following subchapters will discuss the role of the genes and their products in the corresponding metabolic pathways.

5.2.1 CpG sites within genes involved in lipid metabolism

In the EWAS of main blood lipids, results indicated that DNA methylation of cg06500161 located in the ***ABCG1*** gene was associated in opposite directions with HDL-C and TG levels. These results were confirmed by the EWAS using the NMR-data in KORA F4, but it was not validated by replication in the LOLIPOP, NFBC1966 and YFS cohort. *ABCG1* methylation was also associated with lipoprotein subfractions. The gene encodes the ATP binding cassette subfamily G member 1, a cholesterol transporter which plays a role in cellular lipid homeostasis. It has been shown that ABCG1 functions cooperatively with ABCA1, another cholesterol transporter (Gelissen *et al.* 2006). ABCG1 has mature HDL particles as its acceptor, whereas ABCA1 transports phospholipids and cholesterol to lipid-poor HDL subclasses such as ApoA1 (Wang *et al.* 2004; Kennedy *et al.* 2005).

Associations between *ABCG1* methylation and lipid levels were already reported in patients with familial hypercholesterolemia (FH), a monogenic disorder mostly caused by mutations in the *LDLR* gene (Guay *et al.* 2014). Likewise, in FH patients also methylation of *ABCA1* was

related to HDL particles (Guay *et al.* 2012a). One reason why this was not observed in the KORA F4 cohort could be the strong genetic background of the FH disease and the severe elevations in total cholesterol and LDL-C which are not comparable with the range of lipid levels in a population-based cohort.

Methylation of a CpG site located in the **TNIP1** gene was positively associated with LDL-C levels. This gene encodes the tumor necrosis factor-α-induced protein 3 (TNFAIP3)-interacting protein 1 which controls signaling in diverse receptor pathways (Ramirez *et al.* 2012). As an example, studies have shown that this protein acts as co-repressor of agonist-bound peroxisome proliferator-activated receptors (PPAR) and retinoic acid receptors (RAR) (Gurevich and Aneskievich 2009; Flores *et al.* 2011; Ramirez *et al.* 2012). Both receptors play a role in controlling lipid metabolism. PPARs regulate e.g. fatty acid oxidation and lipoprotein metabolism (Li and Glass 2004). RAR and PPARδ together mediate lipid homeostasis (Berry and Noy 2009). Interestingly, the methylation between *TNIP1* and LDL-C was also detected in human adipose tissue but not in skin tissue of the MuTHER cohort, indicating a tissue-specific relationship.

In the EWAS of NMR-measured metabolites, methylation of a CpG site located in **TXNIP** was associated with different lipoprotein subfractions and fatty acids. A published study that also used KORA F4 data had already revealed associations between *TXNIP* methylation and chylomicrons which were measured using a different analysis platform (Petersen *et al.* 2014). *TXNIP* (thioredoxin interacting protein) encodes a pro-apoptotic and pro-oxidant protein that acts as an endogenous inhibitor of thioredoxin which plays a crucial role in modulating intra- and extracellular signaling pathways (Junn *et al.* 2000). TXNIP is an important element in the pathogenesis of many cancers and metabolic diseases (Watanabe *et al.* 2010). It was shown that glucose stimulates *TXNIP* transcription through a carbohydrate response element in the *TXNIP* promoter, resulting in elevated *TXNIP* mRNA expression (Chen *et al.* 2008). A nonsense mutation in the *TNXIP* gene was identified as the cause for the phenotype of a mutant mouse which resembles familial combined hyperlipidemia (Bodnar *et al.* 2002). Disruption of *TXNIP* in mice resulted in elevated plasma levels of free FA, TG, total cholesterol, phospholipids and ketone bodies during fasting. Comparison of the lipoprotein profiles between the *TXNIP* KO and wildtype mice lead to the suggestion that VLDL- and chylomicron-triglyceride degradation is suppressed in *TXNIP* KO

mice, explained by impaired expression of *LPL* (Oka *et al.* 2009). These results strengthen the role of *TXNIP* in lipid metabolism which needs to be further investigated to get insights into the mechanistic background.

5.2.2 CpG sites within genes involved in fatty acid metabolism

CpG sites located in different genes (*ABCG1*, *TXNIP*, *FADS2*, *SLC7A11*, *SCD*, and *PHGDH*) showed associations with fatty acids. Linoleic acid (LA), an omega-6 fatty acid, and α-linolenic acid (ALA), an omega-3 fatty acid, are considered essential fatty acids because they cannot be synthesized by humans. They are endogenously metabolized into very long chain polyunsaturated fatty acids (PUFA) through a series of desaturases and elongases. The key enzymes are here the delta-5 and delta-6 desaturases, which are encoded by the *FADS1* and *FADS2* gene, respectively (Cho *et al.* 1999a; Cho *et al.* 1999b). They are the rate-limiting enzymes in the synthesis of long-chain PUFAs (LC-PUFAs) such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (both omega-3 fatty acids) and arachidonic acid (omega-6 fatty acid). Genetic variants in these genes are important regulators of LC-PUFA synthesis (Schaeffer *et al.* 2006; Lattka *et al.* 2009; Lattka *et al.* 2010).

In this study three CpG sites located in the ***FADS2*** (fatty acid desaturase 2 or delta-6 desaturase) showed associations with omega-3 fatty acids, in particular with DHA as well as the ratios of omega-3 fatty acids and DHA to total FA. One CpG site (cg27386326) located in an intergenic region between *FADS1* and *FADS2* was also related to the ratio of omega-3 fatty acids and DHA to total FA. A study in mice already brought ALA, a precursor of DHA, and DNA methylation of *FADS2* in one context: in mice receiving ALA supplementation during lactation, *FADS2* promoter methylation was changed which was also negatively correlated to *FADS2* expression (Niculescu *et al.* 2013). Also in KORA F4 an association between methylation of *FADS2* CpG sites and gene expression was identified leading to the assumption of the existence of a feedback mechanism which is controlled by epigenetic mechanisms that regulate the expression of the enzymes and thus the FA levels. However, testing on genetic background revealed that associations between methylation of *FADS2* and metabolites were confounded by mainly *FADS1/2* methQTLs (see also **chapter 5.4**). This means that not only genetic variants alone, but also their interplay with DNA methylation of CpG sites could be important regulators of LC-PUFA synthesis. The results indicate a

complex regulation of gene expression and metabolite level by a genetic and epigenetic interplay.

Two CpG sites of the **SCD** gene were related to omega-3 fatty acids and more precisely to DHA, a polyunsaturated fatty acid. *SCD* encodes the stearoyl-CoA desaturase-1, the rate-limiting enzyme in the cellular synthesis of monounsaturated fatty acids (MUFA) from saturated fatty acids (Paton and Ntambi 2009). Interestingly, animal experiments and human studies have shown that PUFAs can influence gene expression (Simopoulos 2010). One study examined the effects of supplementation with EPA and DHA on peripheral blood mononuclear cell gene expression profiles in humans. One gene that was significantly down-regulated in subjects with high EPA and DHA intake was *SCD* (Bouwens *et al.* 2009). In regard to the EWAS results which showed an association between *SCD* methylation and DHA, one hypothesis could be that the effect of the PUFAs on gene expression is driven by epigenetic processes initialized by the PUFAs.

The gene **CPT1A** encodes the carnitine palmitoyltransferase 1A which is located in the outer mitochondrial membrane and facilitates the transport of long chain fatty acids into the mitochondria for their oxidation (McGarry and Brown 1997). In case of inhibition of *CPT1A*, long chain fatty acids are directed toward esterification to TG, the main substrate for hepatic VLDL-TG production (McGarry and Brown 1997; Nielsen and Karpe 2012). In KORA F4 the methylation status of the CpG site cg00574958 located in *CPT1A* showed inverse associations with TG content of HDL and LDL particles, with several constituents of VLDL particles and with serum TG levels as well as with MUFAs and total FA levels. This was also replicated in LOLIPOP but not in the Finnish cohorts and therefore meta-analysis of the replication results confirmed only the association with TG content of small LDL particles. However, the results in KORA F4 are in line with previously published EWAS which reported associations between methylation of cg00574958 and total and small LDL particles as well as average diameter of LDL particles and with total, medium and large VLDL and fasting TG levels (Frazier-Wood *et al.* 2014; Gagnon *et al.* 2014; Irvin *et al.* 2014). In contrast, other reported lipid-associated CpG sites than cg00574958 in *CPT1A* (Frazier-Wood *et al.* 2014; Gagnon *et al.* 2014; Irvin *et al.* 2014) were not discovered in the KORA EWAS.

Methylation of CpG sites located in **SREBF1** showed associations with TG in VLDL and HDL particles. SREBP-1, encoded by *SREBF1* (sterol regulatory element binding transcription

factor 1), is a member of the SREBP family of proteins which are involved in activation of different genes dedicated to the synthesis of cholesterol, FA, TG and phospholipids. The isoform SREBP-1c favors the FA biosynthetic pathway and SREBP-2 favors cholesterol synthesis (Horton *et al.* 2002). The genetic loci of the SREBP, *SREBF1* and *SREBF2*, contain highly conserved miRNAs (miR-33b and miR-33a, respectively) that regulate cholesterol export and fatty acid oxidation. (Najafi-Shoushtari *et al.* 2010; Rayner *et al.* 2010). Interestingly, one CpG site located in *SREBF1* in a region where also miR-33b is encoded (**MIR33B/SREBF1**) showed associations with serum TG, VLDL-TG and S-VLDL particle constituents like phospholipids, free cholesterol, and total lipids.

5.2.3 CpG sites within genes involved in amino acid metabolism

Two CpG sites which were associated with omega-6 fatty acids were located in genes which are linked to the metabolism of amino acids: *PHGDH* (phosphoglycerate dehydrogenase) gene and *SLC7A11* (solute carrier family 7 member 11).

PHGDH is the key metabolic enzyme that drives the rate-limiting step of the serine biosynthesis pathway. It catalyzes the conversion of the glycolytic intermediate 3-phosphoglycerate to the serine precursor 3-phosphohydroxypyruvate which is then further processed to serine (Mullarky *et al.* 2011) (**Figure 17**). Serine is converted to glycine by donation of a methyl group to the folate cycle of the cellular one-carbon metabolism (Mullen and DeBerardinis 2012). Beside the folate cycle the one-carbon metabolism also comprises the methionine cycle as well as the transsulfuration pathway (**Figure 17**). The folate metabolism delivers products which are necessary for the nucleotide metabolism and for the methionine cycle. During the methionine cycle, S-adenosylmethionine (SAM) is synthesized which acts as major methyl donor in cells and is involved in histone, DNA and RNA methylation as well as general protein lysine and arginine methylation. The transsulfuration pathway is coupled to the methionine cycle and results in generation of glutathione, one of the major redox-regulating metabolic systems in cells. Serine acts not only as methyl donor for the folate cycle, it also can be directly metabolized through transsulfuration and thus contributes to the maintenance of cellular redox balance (Locasale 2013) (**Figure 17**).

PHGDH is often cited in connection with cancer pathogenesis. It is overexpressed in approximately 70% of estrogen receptor-negative breast cancers as it drives metabolic

changes leading to rapid cellular growth (Possemato *et al.* 2011). While the role of PHGDH in oncogenesis is well characterized, the effect in metabolic phenotypes needs to be further clarified.
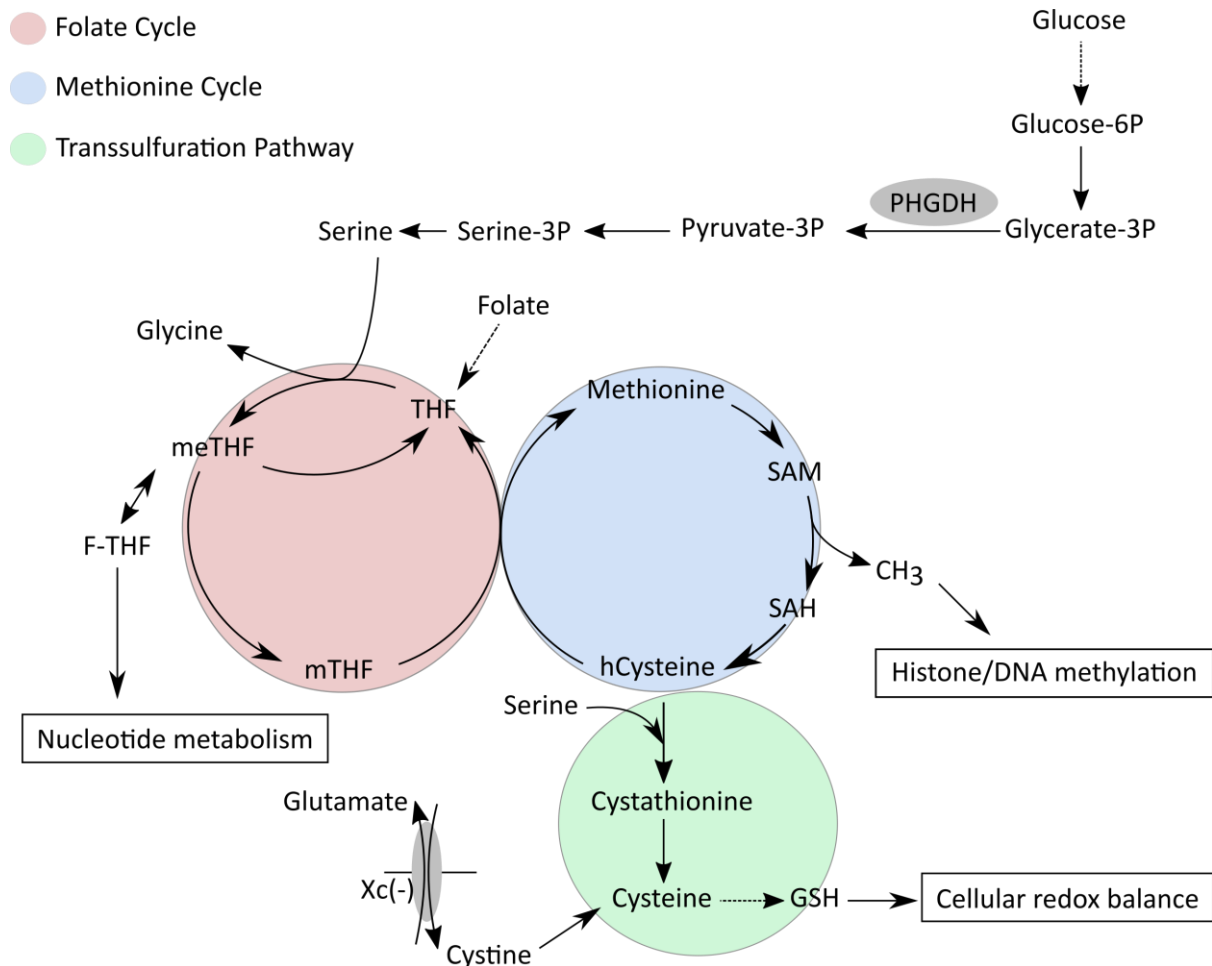
Figure 17: PHGDH is a key enzyme in serine biosynthesis pathway which donates a methyl-group to the folate cycle of the one-carbon metabolism. THF tetrahydrofolate; meTHF 5,10-methylene-THF; mTHF 5-methyltetrahydrofolate; F-THF 10-formyltetrahydrofolate; SAM S-adenosylmethionine; SAH S-adenosylhomocysteine; hCysteine homocysteine; GSH Glutathione; dashed arrows indicate multiple biochemical steps. Modified from (Locasale 2013).

SLC7A11 encodes a member (xCT) of a heteromeric anionic amino acid transport system that is highly specific for cystine and glutamate. This system (designated as Xc(-)) transports the anionic form of cystine in exchange of glutamate (Lim and Donaldson 2011). One of its two major functions is the cellular uptake of cystine for the maintenance of intracellular levels of glutathione which is essential for protection of cells from oxidative stress (**Figure 17**). Secondly, it is involved in maintaining the redox balance between extracellular cystine

and cysteine (Lo *et al.* 2008). Radical-initiated degradation of omega-6 PUFAs such as linoleic and arachidonic acids forms trans-4-hydroxy-2-nonenal (HNE) which can be considered as a major product of lipid peroxidation caused by oxidative stress (Esterbauer *et al.* 1991; Spickett 2013). It was shown that HNE stimulates the Nrf2 (NFE2-related factor) mediated upregulation of antioxidant enzymes such as xCT (Ishii *et al.* 2000; Habib *et al.* 2015). Other studies have shown that another transcription factor, ATF4, is also involved in upregulation of xCT either as direct transcriptional target or as heterodimerization partner of Nfr2 (Lewerenz *et al.* 2012; Ye *et al.* 2014). ATF4 was found to bind methylation dependent on DNA (Mann *et al.* 2013). In addition, HNE itself is able to interact with DNA directly, and one study showed that methylation at CpG sites in a HNE binding spectrum enhances HNE-DNA binding in the human tumor suppressor gene p53 (Hu *et al.* 2002). All studies indicate an epigenetic regulated response to cellular stress which is also affirmed by the results of the EWAS that revealed an association between methylation of *SLC7A11* and omega-6 fatty acid traits.

5.3 Methylation of lipid- and fatty acid-related CpG sites has an impact on gene expression

DNA methylation can directly prevent transcription factor binding by physical obstruction of the transcription factor (Tate and Bird 1993; Hu *et al.* 2013) or by recruitment of methylcytosine-specific repressive factors (Tate and Bird 1993; Baubec *et al.* 2013). It is therefore not surprising that DNA methylation has also been associated with transcriptional regulation (Jones 2012). Studies provide evidence that DNA methylation can cause a loss of gene expression as well as that expression changes can alter DNA methylation patterns (Bestor *et al.* 2015). It was therefore not only of interest to investigate associations between DNA methylation and gene expression in KORA F4, but also to perform functional analyses for selected CpG sites. These analyses should provide information about methylation-dependent protein-DNA binding and whether the methylation has an enhancing or repressive effect on transcriptional activity.

5.3.1   *ABCG1* **methylation has an repressive effect on gene expression**

In KORA F4, the methylation of the lipid-associated CpG sites located in *ABCG1* showed an inverse relationship with *ABCG1* mRNA. Similar results were observed in a study with familial hypercholesterolemia patients where higher *ABCG1* DNA methylation was associated with repression of gene expression (Guay *et al.* 2014). In KORA F4, the identified negative association between *ABCG1* methylation (cg06500161) and *ABCG1* mRNA levels is possibly mediated by methylation-dependent transcription factor binding as observed in the electrophoretic mobility shift assay experiments for cg06500161. Here, methylation of this CpG site inhibits DNA binding of a yet unidentified protein or protein complex. Computational analysis indicated that cg06500161 was located in the binding site of transcription factors such as E2F1. E2F1 is a member of the E2F family of transcription factors (TF) which plays a key role in the regulation of cell growth and apoptosis. E2F binding sites contain at least one CpG motif and are therefore potential TF to be regulated by DNA methylation. Indeed, one study has shown that methylation of the E2F binding sites inhibits the binding of E2F1 (Campanero *et al.* 2000). However, computational analyses discovered also other potential TFs than E2F1 and further studies are necessary for a final identification of the methylation-dependent protein or protein-complex.

In vitro results revealed a repressive impact of cg06500161 methylation on transcriptional activity of a reporter gene. However this effect was only detected in human leucocytes and liver cells but not in human kidney cells indicating tissue specificity. Tissue-specific associations were also discovered in the MuTHER cohort, where the associations between *ABCG1* methylation and blood lipid levels (TG/HDL-C) were found in adipose tissue but not in skin tissue.

Further *in vitro* experiments, which investigated the effect of the additional methylation of up to 28 adjacent CpG sites of cg06500161 on reporter gene expression, underlined the complexity of gene regulation by DNA methylation. In a human monocyte cell line the repressive effect of cg06500161 methylation was abrogated when one or six adjacent CpG sites were simultaneously methylated. However, in case of methylation of all 29 CpG sites within the cg06500161 containing DNaseI hotspot, the repressive effect was again detectable. In liver cells the effect of DNA methylation was independent of the numbers of methylated CpG sites. Targeted methylation of only cg06500161 in a DNA segment with six

additional CpG sites induced a repressive effect on transcriptional activity, emphasizing the functional role of this CpG site. However, only the methylation of all CpG sites within the DNaseI hotspot induced a particularly strong effect, sharply reducing the enhancing effect of the DNaseI hotspot. The results obtained in both cell lines let conclude that the methylation of only one CpG site can already have an effect on gene expression, but also the methylation pattern of the adjacent CpG sites contributes to the transcriptional activity of a gene.

Further experiments revealed that the DNA methylation-dependent enhancer activity of the DNaseI hotspot region around cg06500161 was orientation-independent in hepatic cells, but not in monocytes. These observations as well as the observation that DNA methylation of this region had no influence on reporter gene expression in kidney cells demonstrate a tissue-specific epigenetic gene regulation.

5.3.1.1 Triangular relationship between *ABCG1* methylation, *ABCG1* transcripts and lipid levels

In KORA F4 a triangular relationship between methylation of *ABCG1*, gene transcripts and lipid levels (TG, HDL-C) was observed as *ABCG1* mRNA levels were not only associated with *ABCG1* methylation but also with HDL-C and TG levels (**Figure 8** in **chapter 4.1.3**). Further statistical analyses revealed that the negative association between *ABCG1* methylation (cg06500161) and HDL-C might be partly mediated by the expression of *ABCG1* as the significance of the association was greatly weakened after additional adjustment for *ABCG1* transcripts. Similar results were observed for the positive association between *ABCG1* methylation (cg06500161, cg27243685) and TG levels. However, also the relationship between methylation and transcripts of *ABCG1* seemed to be mediated by TG levels leading to the hypothesis that TG levels may have an effect on DNA methylation. In this context, a very recent study performed a stepwise Mendelian Randomization approach to investigate the direction of causality of the associations between blood lipids and DNA methylation (Dekkers *et al.* 2016). Their analyses provide evidence that altered lipid levels (lowered TG or elevated HDL-C levels) induce lower methylation of the CpG sites cg06500161 and cg27243685 which lead to altered gene expression (Dekkers *et al.* 2016). Nevertheless, further experiments are necessary to finally clarify the direction of the causality and also to investigate a possible epigenetic-driven feedback regulation of lipid levels. For this purpose, one investigative approach could be the examination whether DNA methylation of *ABCG1*

not only affect the transcriptional level but also at the translational level of *ABCG1* resulting in altered ABCG1 protein levels. Also intervention studies in e.g. mice may help to get information about a possible feedback regulation of lipid levels.

5.3.1.2 Fatty acids may repress *ABCG1* expression via epigenetic mechanisms

ABCG1 expression is modulated by different regulatory regions, transcription factors and signaling molecules. Interestingly, studies provide hints that also fatty acids have an influence on *ABCG1* expression. Studies have shown that unsaturated fatty acids repress *ABCG1* expression (Uehara *et al.* 2007; Mauerer *et al.* 2009; Ku *et al.* 2012). This is of interest in regard to the results of the EWAS, where *ABCG1* methylation was positively associated with MUFA levels and with the ratio of MUFAs to total fatty acids. The fatty acid induced altered gene expression is driven by a LXR-dependent mechanism (Uehara *et al.* 2007; Mauerer *et al.* 2009; Ku *et al.* 2012). It is known that LXR plays an important role in *ABCG1* promoter activation, various LXR ligands upregulate *ABCG1* expression (Peng *et al.* 2011; Jun *et al.* 2013). Transcriptional activity of LXR is regulated by binding its natural ligands and by interaction with additional co-regulatory proteins. Interestingly, these coactivator complexes contain enzymatic functions that include histone acetyltransferase and histone methyltransferase activity, therefore they may directly regulate epigenetic changes of a gene and its region (Li and Glass 2004; Rosenfeld *et al.* 2006). One recent study detected that an LXR agonist induced a decrease of DNA methylation of synapse-related and neurogenesis-associated genes in brains of mice (Sandoval-Hernandez *et al.* 2016). The same agonist was also shown to induce *ABCG1* expression in mice macrophages (Tarling and Edwards 2011). Considering all these findings, the results of the EWAS may provide hints for a fatty acid induced LXR-dependent epigenetic modification resulting in controlled expression of *ABCG1*. However, further studies are mandatory to investigate in detail if and how fatty acids are involved in epigenetic-driven transcriptional regulation of *ABCG1*.

5.3.2 Inverse relationship between *SCD* methylation and gene transcripts

In the EWAS, *SCD* methylation was associated with omega-3 fatty acids and precisely with DHA. Moreover, an inverse association between *SCD* methylation and gene expression was detected in KORA F4, which is in accordance with recent mice studies that investigated diet-

dependent alterations of *SCD* expression induced by epigenetic changes (Schwenk *et al.* 2013; Chaplin *et al.* 2015). Schwenk *et al.* showed an inverse correlation of *SCD* methylation and SCD levels in livers of mice fed with a high-fat diet (Schwenk *et al.* 2013). Chaplin *et al.* recognized a positive effect of conjugated linoleic acid and calcium supplementation on the methylation profile of *SCD* in epididymal adipose tissue in mice. In addition, *SCD* methylation was negatively correlated with *SCD* expression (Chaplin *et al.* 2015). In humans EPA and DHA supplementation induced a downregulation of *SCD* expression (Bouwens *et al.* 2009) and considering the animal studies and the EWAS results it appears likely that also this effect might be driven by epigenetic processes.

SCD is not only involved in the biosynthesis of MUFAs, it is also a key adipogenic enzyme in rodents and humans (Miyazaki and Ntambi 2003) and plays a significant role in the regulation of adipocyte inflammation via a mechanism that involves the regulation of inflammatory gene expression (Liu *et al.* 2010). Interestingly, a recent study has shown that SCD might control inflammatory gene expression by regulating the DNA methylation of those genes (Malodobra-Mazur *et al.* 2014) . Further studies are needed to enlighten the interrelation between epigenetic induced *SCD* expression changes and the consequences of this mechanism on (epigenetic) regulation of further genes.

5.3.3   *PHGDH* methylation is strongly inversely associated with gene transcripts

The association with the highest significance was between *PHGDH* methylation (cg16246545) and *PHGDH* gene expression. As serine is involved in controlling intracellular SAM levels, PHGDH amplification might affect biosynthesis of SAM and thus the methylation status of DNA and histones, which might have an impact in regulation of gene expression (DeBerardinis 2011; Mullarky *et al.* 2011) (see also **Figure 17**). Interestingly, the response to serine deprivation is regulated by epigenetic mechanisms leading to transcriptional activation of genes involved in the serine biosynthetic pathway such as *PHGDH* (Ye *et al.* 2012; Ding *et al.* 2013). This is modulated by the histone H3 lysine K9 (H3K9) methyltransferase G9A by specifically marking the pathway enzyme genes with monomethylation of H3K9 (H3K9me1) (Ding *et al.* 2013), a marker for transcriptional activation (Mosammaparast and Shi 2010). In reference to this, bioinformatic analysis using UCSC Genome Browser GRCh37/hg19 revealed that in some cell lines (Huvec, NHEK and

K562 cells) the CpG site cg16246545 is located in a region marked by the histone modification H3K9me1. An adjacent CpG site (cg14476101), also located in this region, was also found to be associated with *PHGDH* expression in KORA F4 strengthening the possible regulatory role of this locus.

PHGDH methylation was associated with omega-6 fatty acid traits. Interestingly, *PHGDH* expression is – just as *SLC7A11* expression - controlled by the transcription factor Nrf2 via ATF4 (DeNicola *et al.* 2015) which is stimulated by HNE (Chen *et al.* 2005; Cheng *et al.* 2011), a product of omega-6 PUFAs oxidation. DNA binding of transcription factor ATF4 was shown to be methylation dependent (Mann *et al.* 2013). ATF4 is also required for transcriptional activation by G9A in the state of serine deprivation (Ding *et al.* 2013).

All results give hints for a complex machinery in which epigenetic processes may act as important regulatory mechanisms for *PHGDH* expression. Omega-6 fatty acids might be involved in maintenance of redox balance by epigenetic regulation of respective genes. Omega-6 fatty acid induced *PHGDH* regulation may also affect cellular SAM levels and thus methylation reactions. Further studies are needed to unravel the detailed epigenetic regulatory mechanisms of *PHGDH* and thus of the serine-glycine biosynthetic pathway.

5.3.4 Methylation of cg20544516 (*MIR33B/SREBF1*) enhances transcriptional activity

Although *SREBF1* methylation was not associated with its gene expression in KORA F4, the CpG site cg20544516 (*MIR33B/SREBF1*) was further investigated as this locus attracted the attention due to its location in a microRNA encoding sequence. As mentioned in **chapter 5.2.2 and 5.5**, miR-33b plays a pivotal role in a variety of biological processes including cholesterol homeostasis maintenance, HDL metabolism and fatty acid oxidation.

Functional analysis of cg20544516 uncovered a methylation-specific DNA binding of a yet unknown protein or protein complex. Computational analysis identified PAX-5 as a potential transcription factor, however, the binding site of PAX-5 overlaps not directly with the CpG site. Further analyses are necessary to identify the methylation-specific DNA-binding proteins as the detected protein complex could also be methylcytosine-binding proteins which function in transcriptional repression (Bogdanovic and Veenstra 2009). A study of Grundberg *et al.* reported a negative association between methylation of cg20544516 and *SREBF1* expression in adipose tissue from twins (Grundberg *et al.* 2013), supporting the

hypothesis that methylation causes binding of methylcytosine-binding proteins which leads to transcriptional repression. In contrast, this association was not observed in the KORA F4 cohort. In addition, results of Luciferase Reporter Assays support the assumption of a methylation-specific binding of a transcription factor as the analyses revealed a methylation-specific enhancing effect on reporter gene expression. The conducted functional analyses however do not provide information whether methylation of cg20544516 also affects *MIR33B/SREBF1* expression and thus causes altered miR-33b levels in the human body, which could lead to regulatory changes of target genes. In general, one study has already reported an epigenetic regulation of miR-33b levels; Yin *et al.* observed an association between enhanced methylation levels of a CpG island in *SREBF1* upstream of *MIR33B/SREBF1* and down-regulated miR-33b levels (Yin *et al.* 2016).

In summary, this is the first time that methylation of *MIR33B/SREBF1* and blood lipid levels are placed in context, and the additional discovery of the impact of the epigenetic alteration on gene expression might open new doors for further investigative approaches. One of these approaches could be the examination of a potential association between methylation of *MIR33B/SREBF1* and miR-33b plasma levels in humans.

5.4   The role of methylation quantitative trait loci

It is known that genetic variation is often associated with quantitative changes in methylation levels (Gibbs *et al.* 2010; Zhang *et al.* 2010; Bell *et al.* 2011; Gutierrez-Arcelus *et al.* 2013). It was therefore of interest to identify methylation quantitative trait loci (methQTLs) and thus to investigate the genetic background of the discovered associations between lipids and DNA methylation.

In the EWAS of main blood lipids, it was examined whether the observed associations between main lipids and methylation levels in KORA F4 were based on genetic confounding by known lipid-associated SNPs which have been identified by GWAS. Firstly, those lipid-associated SNPs were pre-selected which were nominally associated with DNA methylation at the respective CpG site. Secondly, associations between lipids and DNA methylation were re-calculated additional adjusting for the pre-selected SNPs. Most associations remained significant after adjustment for the respective SNPs, indicating that the observed nominal associations between the pre-selected SNPs and the CpG sites were dependent on lipids as

both the SNPs and the CpG sites are associated with lipid levels. Only one association lost its significance after adjustment for the respective SNPs and further analyses showed that the association between DNA methylation at the CpG site cg12556569 (located in the promoter region of *APOA5*) and TG levels was confounded by the SNP rs964184. This SNP is known to primarily affect triglyceride levels (Global Lipids Genetics *et al.* 2013) and one study had already previously identified this SNP as a methQTL (Moen *et al.* 2013). The association between cg12556569 (*APOA5*) and TG however was anyway not successfully replicated in KORA F3 and InCHIANTI. The other identified and also validated lipid-DNA methylation associations were not due to genetic confounding.

To investigate the genetic background of the results of the EWAS of NMR-measured metabolites, another approach was applied. As there are not many GWAS for NMR-measured lipoprotein subclasses and thus only a low number of identified genetic loci which are associated with lipoprotein measures are known, the influence of SNPs located in a 2 Mb window of the respective CpG was investigated. It was tested whether the CpG-metabolite associations remained significant when adjusting for SNPs which were nominally associated with both methylation of the CpG site and with the respective CpG-associated metabolites. 23 CpG-metabolite pairs lost significance after additional adjustment for the pre-selected SNPs. However, p-values of 12 associations showed only slight changes and were just over the level of significance. The other associations, which clearly lost significance, were those between *FADS2* methylation and FA. Genetic variants in *FADS* genes (*FADS1* and *2*) are well-known to be related to FA levels in humans (Lattka *et al.* 2010). Additionally, genetic variants are also associated with gene expression (Reardon *et al.* 2012; Wang *et al.* 2015). It is therefore plausible that the EWAS results regarding *FADS2* methylation and omega-3 fatty acids were found to be genetically confounded by methQTLs in the *FADS* region. In addition, Liu *et al.* already identified the *FADS1/2* region as a region where SNPs in LD blocks are associated with correlated methylated CpG sites, which means that some clustering of methylated CpG sites in the *FADS1/2* genes appear to be genetically driven (Liu *et al.* 2014). Interestingly, in KORA F4 the methylation of fatty acid-related CpG sites located in *FADS2* showed also associations with *FADS* gene expression. It is therefore possible that SNPs, which were identified to be related to fatty acid levels, do not have a direct influence on fatty acid levels (e.g. by impairing transcription factor binding leading to

altered gene expression), but do have an indirect influence by determining the methylation patterns in the *FADS* region and thus the expression levels of the *FADS* genes. Further studies are necessary to clarify the complex interplay between fatty acids, genetic variants, DNA methylation and gene expression in the region of the *FADS* genes.

5.5 Interaction of genes of lipid- and fatty acid-associated CpG sites

Some of the genes in which the identified lipid- and fatty acid-related CpG sites are located and their products interact with one another (**Figure 18**).

SREBPs (sterol regulatory element-binding proteins) are key transcriptional factors that control lipogenesis and lipid uptake. The two miRNAs conserved in *SREBF1* and *SREBF2* (miR-33b and miR-33a, respectively) are identical in their seed sequences, and thus have been predicted to repress the same set of genes with similar specificities (Najafi-Shoushtari *et al.* 2010; Rayner *et al.* 2010). They play an important role for the regulation of genes involved in fatty acid oxidation, glucose metabolism, and in the reverse cholesterol transport (Horie *et al.* 2010; Davalos *et al.* 2011; Ramirez *et al.* 2013). Being more precisely, studies have shown that both miR-33b and miR-33a target *CPT1A* which is involved in regulation of fatty acid oxidation (Gerin *et al.* 2010; Davalos *et al.* 2011) (**Figure 18**). MiR-33a also targets *ABCA1* in humans and mice (Gerin *et al.* 2010; Najafi-Shoushtari *et al.* 2010). The cholesterol transporters ABCA1 and ABCG1 mediate synergistically cholesterol efflux to HDL (Gelissen *et al.* 2006), however, the role of miR-33 in *ABCG1* expression in humans is not yet clarified. MiR-33a also targets *ABCG1* in mice (Rayner *et al.* 2010). In contrast miR-33a repression of *ABCG1* is not conserved in human cells due to the loss of miR-33a binding sites in 3' UTR of *ABCG1* (Rayner *et al.* 2010). MiR-33b is present in primates, but lacking in rodents and lower organisms (Brown *et al.* 2010), which is one reason why the determination of the precise mechanisms of miR-33b is not possible in, for example, mouse models. Horie *et al*. established a miR-33b knock-in (KI) mouse and noticed that in these mice expression of *ABCA1*, *ABCG1* and *SREBF1* was diminished. Also HDL-C levels were reduced by almost 35% in miR-33b KI homozygous mice compared with the control mice. In contrast, TG levels did not show any changes (Horie *et al.* 2014). Beside this, Rayner *et al*. showed that inhibiting of miR-33a and miR-33b in healthy male non-human primates resulted in increased circulating HDL-C levels and in a coincident decrease in VLDL

triglycerides (Rayner *et al*. 2011). Additionally, Rottiers *et al*. saw the same effect on HDL-C levels when inhibiting both isoforms of miR-33 by an 8-mer LNA-modified anti-miR in female obese non-human primates (Rottiers *et al.* 2013).

Expression of *ABCG1* is also regulated by another product of *SREBF1*, the SREBP-1c. As an example, Ecker *et al*. showed trans-9,trans-11-CLA (conjugated linoleic acids) induced activation of *ABCG1* via SREBP-1c in MCSF (macrophage colony stimulating factor)-differentiated monocytes (Ecker *et al.* 2007). SREBP-1c plays also an important role in the transcriptional regulation of *SCD* (Ntambi 1999). Mouse models with SREBP-1 deficiency have a decreased hepatic SCD1 mRNA expression whereas mice expressing a constitutively active form have increased expression (Shimomura *et al.* 1998; Liang *et al.* 2002). *SCD* expression is also regulated by PUFAs, probably also via mechanism where SREBP-1c is involved (Yoshikawa *et al.* 2002). SREBP1 is also involved in the effect of dietary fatty acids on desaturase transcription (Matsuzaka *et al.* 2002). Dietary fatty acids have an influence on SREBP processing and stability, which in turn has an effect on *FADS2* gene expression (Xu *et al.* 1999; Yoshikawa *et al.* 2002).
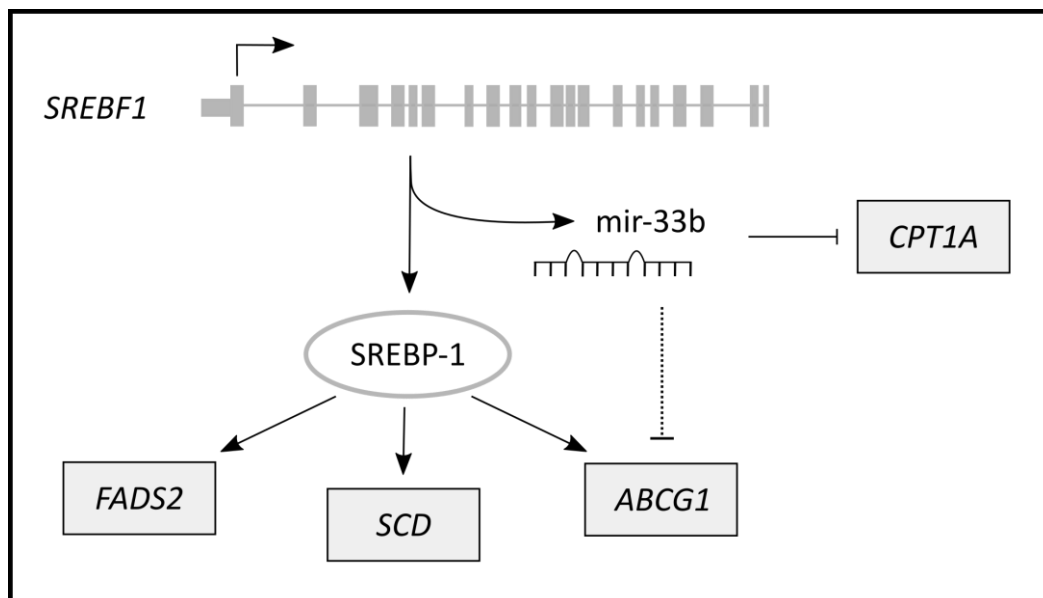
Figure 18: Genes of lipid- and fatty acid-related CpG sites interact with one another. *SREBF1* encodes the transcription factor SREBP-1 that targets genes such as *FADS2*, *SCD* and *ABCG1* which are involved in lipid and fatty acid metabolism. *SREBF1* also encodes microRNA 33b, which has an repressive effect on *CPT1A* and *ABCG1*. The dashed line indicates lack of evidence in humans.

There also seems to be an interrelation between *TXNIP*, *SREBF1* and *SCD* in connection with the mechanism of fructose-induced hepatic disturbance. In rats, fructose induced increase of *TXNIP* expression triggered inflammatory signaling that caused transcriptomic dysregulation of PPAR-α, SREBP1, and SCD1 (Zhang *et al.* 2015).

TNIP1 encodes a protein which acts as co-repressor of PPARs and RARs (Gurevich and Aneskievich 2009; Flores *et al.* 2011; Ramirez *et al.* 2012). Interestingly, ligand-activated RAR influences *ABCA1* and *ABCG1* expression in human macrophages. It increases the expression by modulation of the *ABCG1* promoter activity via LXR responsive elements-dependent mechanisms (Ayaori *et al.* 2012). Additionally, studies have shown that PPARα/γ-activators induce *ABCA1* expression in macrophages (Chinetti *et al.* 2001) and PPARγ induce *ABCG1* expression (Li *et al.* 2004). Therefore, *TNIP1* may have an indirect impact on the expression of *ABCA1* and *ABCG1*.

There might also be an interrelation between *PHGDH* and *SLC7A11* pathways. As mentioned above, *SLC7A11* encodes a member of an amino acid transport system that is highly specific for cystine and glutamate (Lim and Donaldson 2011). The *PHGDH* product is also used to convert glutamate to α-ketoglutarate which can enter the citric acid cycle (Mullen and DeBerardinis 2012). Also, both are involved in maintenance of intracellular levels of glutathione, *SLC7A11* by mediating cellular uptake of cystine (Lo *et al.* 2008) and *PHGDH* by being involved in biosynthesis of serine, a carbon donator for the one-carbon metabolism (DeBerardinis 2011) (see also **Figure 17** in **chapter 5.2.3**). In addition, *PHGDH* and *SLC7A11* are both controlled by processes where the transcription factor Nrf2 is involved, which is stimulated by HNE, the end product formed from the oxidation of omega-6 PUFAs (Sasaki *et al.* 2002; DeNicola *et al.* 2015).

In summary, lipid levels were associated with the methylation level of genes which are involved in lipid metabolism and which interrelate with each other. This indicates a complex regulation of the human metabolism with the involvement of epigenetic processes. Further studies are needed to clarify detailed mechanisms of this complex machinery.

5.6 Link between DNA methylation and lipid-related diseases

As blood lipid levels are risk factors for cardiometabolic diseases, it was therefore of interest to examine whether methylation states of lipid-associated CpG sites are associated with lipid-related diseases such as myocardial infarction. The identified positive association between *ABCG1* methylation and prevalent myocardial infarction in KORA F4 are in line with results of a study in Chinese coronary heart disease (CHD) patients which showed that promoter *ABCG1* DNA methylation was associated with risk of CHD (Peng *et al.* 2014). However, the in KORA F4 identified association with myocardial infarction was not validated by replication in KORA F3 and InCHIANTI, possibly due to the low number of cases of myocardial infarction in these cohorts. Furthermore, the association was not significant anymore when additionally adjusting for intake of lipid-lowering drugs. One explanation for this observation might be the high correlation between patients with myocardial infarction and intake of lipid lowering medication. Studies also have found effects of lipid-lowering drugs on *ABCG1* gene expression probably through a LXR-dependent pathway (Genvigir *et al.* 2010; Wang *et al.* 2013), where also epigenetic modifications could be involved (see also **chapter 5.3.1.2**). However, further studies with larger numbers of cases with myocardial infarction or other CHD cases are needed to validate these results.

Beside prevalent myocardial infarction it was also investigated whether lipid-associated CpG sites were associated with prevalent type 2 diabetes (T2D). Although in this study *ABCG1* methylation was not associated with prevalent T2D, other epigenome-wide association studies have reported associations between methylation of cg06500161 and future T2D incidence (Chambers *et al.* 2015; Dayeh *et al.* 2016). There were also links between *ABCG1* methylation and diabetes-related risk factors such as fasting insulin (Hidalgo *et al.* 2014; Kriebel *et al.* 2016), HOMA-IR (Hidalgo *et al.* 2014; Kriebel *et al.* 2016), fasting glucose (Kriebel *et al.* 2016), hypertriglyceridemic waist (Mamtani *et al.* 2016) as well as BMI (Demerath *et al.* 2015; Ding *et al.* 2015) and waist circumference (Demerath *et al.* 2015).

Similar results were found for methylation of *CPT1A*. Methylation of cg00574958 was not only negatively associated with prevalent T2D in KORA F4, but also in an Arab population (Al Muftah *et al.* 2016). EWAS have also reported associations with metabolic syndrome (Das *et al.* 2016), obesity traits such as BMI, BMI change, waist circumference (Aslibekyan *et al.*

2015; Demerath *et al.* 2015; Al Muftah *et al.* 2016) and hypertriglyceridemic waist (Mamtani *et al.* 2016) as well as with fasting glucose levels (Kriebel *et al.* 2016). Similarly, *TXNIP* showed a strong negative association with prevalent T2D in KORA F4 which was also reported in other cohorts (Chambers *et al.* 2015; Al Muftah *et al.* 2016).

Furthermore, studies provide results showing that also other lipid-associated CpG sites are related to metabolic-related traits. Methylation of the CpG site cg07504977 mapped to the promoter of the long intergenic non-protein coding RNA 263 was linked to obesity traits such as BMI and waist circumference (Aslibekyan *et al.* 2015) as well as with glucose measures (Kriebel *et al.* 2016). Studies have already shown that long non-coding RNAs (lncRNAs) are important regulators of lipid metabolism and adipogenesis, the understanding of the role of lncRNAs is however yet unexplored (Chen 2016). In addition *PHGDH* and *SREBF1* methylation was found to be related to BMI and waist circumference (Aslibekyan *et al.* 2015; Demerath *et al.* 2015; Al Muftah *et al.* 2016). Methylation of *SREBF1* was also associated with T2D (Chambers *et al.* 2015) and fasting glucose (Kriebel *et al.* 2016). A prospective cohort intervention study detected relations between weight changes and *SCD1* gene promoter methylation (Martin-Nunez *et al.* 2014).

In summary, this thesis as well as several other studies identified a relationship between lipid-related diseases and methylation of CpG sites which were also associated with lipid levels. Together with the findings of Dekkers *et al.*, who reports that lipid levels influence DNA methylation and not the other way around, one hypothesis is that altered lipid levels induce DNA methylation changes which contribute to the development of diseases for which lipid levels are known risk factors (Dekkers *et al.* 2016).

Another interesting aspect is the clinical relevance of methylation-regulated expression of microRNAs in diseases. The EWAS discovered lipid-associated CpG sites located in the *SREBF1* gene as well as in the *MIR33B/SREBF1* gene encoding the miR-33b. Plasma levels of this microRNA were found to be upregulated in patients with familial hypercholesterolemia in paediatric age (Martino *et al.* 2015). Another study discovered differential methylation of the same CpG site in adipose tissue from diabetic versus nondiabetic twins (Nilsson *et al.* 2014). These studies as well as the results of the functional analyses support the hypothesis of an association between epigenetic influence on *MIR33B* expression, miR-33b levels and metabolic diseases. In fact, miR-33 is known to be involved in regulation of fatty acid, lipid

and glucose metabolism as well as in cell cycle progression/proliferation and inflammatory response and thus may be involved in the onset of atherosclerosis. It is therefore also a potential candidate for a novel microRNA based therapeutic strategy against lipid disorders and/or atherosclerosis (Chen *et al.* 2013; Ono *et al.* 2015). However, before that, further investigations are required to understand the complexity of the miR-33 biology. The results of this study might contribute to this understanding.

5.7 Strengths and limitations of the studies

Major strengths of the thesis are the relatively large samples size of the population-based discovery cohort KORA F4 as well as the genome-wide approach to investigate associations between DNA methylation patterns and metabolites. The network of cooperation partners and thus the possibility to involve other large-scale studies allowed a validation of the results. The large pool of different large-scale datasets in KORA F4 (DNA methylation, gene expression and genotypes) provided by the KORA meta-database for phenotypes/variables and genetics/omics (KORAgen) enabled a detailed analysis of the associations. Quality assurance of the Infinium HumanMethylation450 BeadChip (Illumina) data was performed applying several quality control steps during the laboratory processes as well as data processing and assessment. In addition, statistical models were adjusted for a large number of potential confounders. *In vitro* experiments were applied to confirm and sustain some of the results and to further investigate the consequences of DNA methylation changes on gene regulation.

The thesis has also certain limitations. Cross-sectional studies do not provide any information about the causation of DNA methylation. However, a very recent study provides results that imply an effect of lipid levels on DNA methylation – and not the other way around (Dekkers *et al.* 2016). Also fatty acids regulate gene expression via epigenetic mechanisms as discussed in **chapter 5.2.2**. Further *in vitro* and/or *in vivo* studies could finally clarify the causes as well as the functional consequences of lipid-related DNA methylation alterations.

Another limitation is the use of DNA samples isolated from whole blood which consists of a mixture of cell types with different methylation patterns. Normal tissue development, individual cellular differentiation and cellular lineage determination are regulated by

epigenetic mechanisms (Khavari *et al.* 2010) which result in tissue- and cell-specific DNA methylation patterns (Christensen *et al.* 2009; Reinius *et al.* 2012). Therefore results of whole blood EWAS need to be considered critically as they might be driven by cell heterogeneity. The ideal method to overcome this problem would be the use of DNA methylation data measured in a single cell type. If only whole blood DNA is available the best approach would be a direct measurement of various leukocyte counts and to adjust for those cell type proportions. However, it is expensive and logistically difficult to measure these counts in a large study population. In this thesis the issue of cellular heterogeneity was addressed by a statistical approach which estimates cell proportions based on published cell-specific methylation signatures (Houseman *et al.* 2014). Although all statistical models were adjusted for those estimated cell proportions, the cell proportion data are limited by the quality of external data and would be more accurate by a direct analysis of the blood.

Furthermore, results detected with the HumanMethylation450 BeadChip were not further evaluated using different laboratory techniques such as EpiTYPER (Agena Bioscience; targeted approach) or bisulfite sequencing. It is also important to note, that the HumanMethylation450 BeadChip array does not distinguish between 5mC and 5hmC, which was also recently found to be associated with human diseases (Wang *et al.* 2014). As there are already hints for associations between 5hmC and LDL-C in visceral adipose tissue (Rohde *et al.* 2015), further studies using methods which allow the determination of 5hmC levels are necessary to also investigate the relationship between 5hmC and lipid measurements.

Also the *in vitro* experiments have some limitations. Transfection was done using a lipid-based transfection reagent. The lipid-DNA complexes are often unstable in solution resulting in the formation of aggregates which contribute to lower transfection efficiencies (Liu *et al.* 1997). The efficiency also varies depending on cell division and cell type. For comparable results a dual luciferase reporter system was applied and luciferase activity of the test plasmid was normalized against the luciferase activity of a control vector. Other transfection methods, e.g. the nucleofection technique where the plasmid DNA is delivered directly into the cell nucleus (Gresch *et al.* 2004), would ensure more reproducible transfection efficiencies. Especially results of experiments conducted in HepG2 cells, a known difficult-to-transfect cell line, would be more reliable using nucleofection. Another limitation is the limited ability of CpG site-specific DNA methylation. In these experiments, a

methyltransferase was used which specifically methylated the CpG site of interest. However, in constructs carrying a longer DNA fragment, a CpG site-specific methylation was not possible anymore as the fragment carried the enzyme-specific sequence more than once. Furthermore, when interpreting the results of experiments conducted in *in vitro* systems, it should be considered that also further regulatory mechanisms such as long non-coding RNAs or miRNAs exist which might have an additional impact on gene regulation.

5.8 Conclusion

Epigenome-wide association studies in whole blood and subsequent replication in independent cohorts revealed 17 lipid-associated CpG sites located in a total of ten different genes, all of which are involved in lipid metabolism (*ABCG1*, *TXNIP*, *TNIP1*), fatty acid metabolism (*FADS2*, *SCD*, *CPT1A*, *SREBF1*, *MIR33B/SREBF1*) and amino acid metabolism (*PHGDH*, *SLC7A11*). Replication in human tissue indicated tissue-specificity for certain associations. Methylation of CpG sites within *ABCG1*, *PHGDH* and *SCD* was additionally associated with expression of their genes in blood cells. *In vitro* studies of CpG sites located in *ABCG1* and *MIR33B/SREBF1* detected a methylation-dependent protein-DNA binding affinity and revealed methylation-dependent reporter gene expression. Investigation of the genetic background of the identified CpG-metabolite associations identified methylation quantitative trait loci in the *FADS2* gene. Furthermore, CpG sites located in *ABCG1*, *CPT1A* and *TXNIP* were connected to lipid-related diseases such as type 2 diabetes or myocardial infarction. In revealing that epigenetic changes are associated with both dyslipidemia and metabolic diseases, the results of this thesis may help to understand their respective pathogenesis and may contribute to the development of new classes of pharmacological agents for the treatment of lipid-related disorders.

5.9 Outlook

This thesis comprises the results of EWAS of lipids and lipoprotein subfractions and provides further insights into the complex regulatory relationship between methylation of lipid-associated CpG sites, gene expression and genetic variation.

Functional experiments underlined the biological relevance of the methylation state of lipid-associated CpG sites. Carrying out electrophoretic mobility shift assays for further lipid-associated CpG sites, as well as identifying the corresponding methylation-specific transcription factors (e.g. by mass spectroscopy), could be the subjects of further studies. Furthermore, other *in vivo* and *in vitro* studies are necessary to analyse the mechanisms that contribute to the formation of locus-specific DNA methylation. *In vivo* diet intervention studies, for instance, could help to clarify causation of DNA methylation changes. One *in vitro* approach could be epigenome editing using a CRISPR-Cas9-based tool for specific DNA methylation (McDonald *et al.* 2016; Vojta *et al.* 2016). This method enables the investigation of site-specific DNA methylation effects on gene expression and other functional consequences in a genome- and transcriptome-wide manner.

The Infinium HumanMethylation450 BeadChip (Illumina) only queries a subset of all methylated sites in the human genome. Whole-genome bisulfite sequencing would provide more information, however the associated currently high costs may limit the possibility of application. Alternatively, analyses could be done using the newly released Infinium MethylationEPIC BeadChip (Illumina) which features almost twice as many CpG sites as the HumanMethylation450 BeadChip. These analyses would be interesting not only using whole blood DNA but also using DNA of different tissue samples (e.g., blood vessel, liver). This would enable the examination of tissue specificity of detected lipid-related associations as well as new tissue-specific associations which do not occur in whole blood samples.

Additionally, performing a meta-analysis of multiple cohorts would not only strengthen the evidence for the identified metabolite-CpG associations but could also potentially detect further associated CpG sites due to the increased sample size and thereby statistical power. Finally, follow-up analyses on patients would be interesting to investigate the epigenetic predisposition for development of lipid-related complex diseases.

6 REFERENCES

Agostini M, Knight RA. 2014. miR-34: from bench to bedside. *Oncotarget* **5**: 872-881.

Al Muftah WA, Al-Shafai M, Zaghlool SB, Visconti A, Tsai PC, Kumar P, Spector T, Bell J, Falchi M, Suhre K. 2016. Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clinical epigenetics* **8**: 13.

Alegria-Torres JA, Baccarelli A, Bollati V. 2011. Epigenetics and lifestyle. *Epigenomics* **3**: 267-277.

Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, Sha J, Pankow JS, Liu C, Irvin MR *et al.* 2015. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity* **23**: 1493-1501.

Asztalos BF, Tani M, Schaefer EJ. 2011. Metabolic and functional relevance of HDL subspecies. *Current opinion in lipidology* **22**: 176-185.

Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T *et al.* 2009. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature genetics* **41**: 47-55.

Ayaori M, Yakushiji E, Ogura M, Nakaya K, Hisada T, Uto-Kondo H, Takiguchi S, Terao Y, Sasaki M, Komatsu T *et al.* 2012. Retinoic acid receptor agonists regulate expression of ATP-binding cassette transporter G1 in macrophages. *Biochimica et biophysica acta* **1821**: 561-572.

Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, Roux J, Pritchard JK, Gilad Y. 2014. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS genetics* **10**: e1004663.

Bansal S, Buring JE, Rifai N, Mora S, Sacks FM, Ridker PM. 2007. Fasting compared with nonfasting triglycerides and risk of cardiovascular events in women. *JAMA : the journal of the American Medical Association* **298**: 309-316.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215-233.

Basu A, Jenkins AJ, Zhang Y, Stoner JA, Klein RL, Lopes-Virella MF, Garvey WT, Lyons TJ, Group DER. 2016. Nuclear magnetic resonance-determined lipoprotein subclasses and carotid intima-media thickness in type 1 diabetes. *Atherosclerosis* **244**: 93-100.

Baubec T, Ivanek R, Lienert F, Schubeler D. 2013. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**: 480-492.

Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome biology* **12**: R10.

Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A *et al.* 2012. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics* **8**: e1002629.

Berneis KK, Krauss RM. 2002. Metabolic origins and clinical significance of LDL heterogeneity. *Journal of lipid research* **43**: 1363-1379.

Berry DC, Noy N. 2009. All-trans-retinoic acid represses obesity and insulin resistance by activating both peroxisome proliferation-activated receptor beta/delta and retinoic acid receptor. *Molecular and cellular biology* **29**: 3286-3296.

Bestor TH, Edwards JR, Boulard M. 2015. Notes on the role of dynamic DNA methylation in mammalian development. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 6796-6799.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL *et al.* 2011. High density DNA methylation array with single CpG site resolution. *Genomics* **98**: 288-295.

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. 2009. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* **1**: 177-200.

Bird A. 2002. DNA methylation patterns and epigenetic memory. *Genes & development* **16**: 6-21.

Bodnar JS, Chatterjee A, Castellani LW, Ross DA, Ohmen J, Cavalcoli J, Wu C, Dains KM, Catanese J, Chu M *et al.* 2002. Positional cloning of the combined hyperlipidemia gene Hyplip1. *Nature genetics* **30**: 110-116.

Bogdanovic O, Veenstra GJ. 2009. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* **118**: 549-565.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185-193.

Bouwens M, van de Rest O, Dellschaft N, Bromhaar MG, de Groot LC, Geleijnse JM, Muller M, Afman LA. 2009. Fish-oil supplementation induces antiinflammatory gene expression profiles in human blood mononuclear cells. *The American journal of clinical nutrition* **90**: 415-424.

Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H. 2012. Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. *European heart journal* **33**: 2841-2848.

Brown MS, Ye J, Goldstein JL. 2010. Medicine. HDL miR-ed down by SREBP introns. *Science* **328**: 1495-1496.

Buuren Sv, Groothuis-Oudshoorn K. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45**.

Campanero MR, Armstrong MI, Flemington EK. 2000. CpG methylation as a mechanism for the regulation of E2F activity. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 6481-6486.

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**: 2933-2942.

Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642-655.

Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, Wahl S, Elliott HR, Rota F, Scott WR *et al.* 2015. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol* **3**: 526-534.

Chaplin A, Palou A, Serra F. 2015. Methylation analysis in fatty-acid-related genes reveals their plasticity associated with conjugated linoleic acid and calcium supplementation in adult mice. *Eur J Nutr* doi:10.1007/s00394-015-1135-3.

Chasman DI, Pare G, Mora S, Hopewell JC, Peloso G, Clarke R, Cupples LA, Hamsten A, Kathiresan S, Malarstig A *et al.* 2009. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS genetics* **5**: e1000730.

Chen J, Saxena G, Mungrue IN, Lusis AJ, Shalev A. 2008. Thioredoxin-interacting protein: a critical link between glucose toxicity and beta-cell apoptosis. *Diabetes* **57**: 938-944.

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang DZ. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature genetics* **38**: 228-233.

Chen WJ, Zhang M, Zhao GJ, Fu Y, Zhang DW, Zhu HB, Tang CK. 2013. MicroRNA-33 in atherosclerosis etiology and pathophysiology. *Atherosclerosis* **227**: 201-208.

Chen Z. 2016. Progress and prospects of long noncoding RNAs in lipid homeostasis. *Mol Metab* **5**: 164-170.

Chen ZH, Saito Y, Yoshida Y, Sekine A, Noguchi N, Niki E. 2005. 4-Hydroxynonenal induces adaptive response and enhances PC12 cell tolerance primarily through induction of thioredoxin reductase 1 via activation of Nrf2. *The Journal of biological chemistry* **280**: 41921-41927.

Cheng X, Siow RC, Mann GE. 2011. Impaired redox signaling and antioxidant gene expression in endothelial cells in diabetes: a role for mitochondria and the nuclear factor-E2-related factor 2-Kelch-like ECH-associated protein 1 defense pathway. *Antioxidants & redox signaling* **14**: 469-487.

Chinetti G, Lestavel S, Bocher V, Remaley AT, Neve B, Torra IP, Teissier E, Minnich A, Jaye M, Duverger N *et al.* 2001. PPAR-alpha and PPAR-gamma activators induce cholesterol removal from human macrophage foam cells through stimulation of the ABCA1 pathway. *Nature medicine* **7**: 53-58.

Cho HP, Nakamura M, Clarke SD. 1999a. Cloning, expression, and fatty acid regulation of the human delta-5 desaturase. *The Journal of biological chemistry* **274**: 37335-37339.

Cho HP, Nakamura MT, Clarke SD. 1999b. Cloning, expression, and nutritional regulation of the mammalian Delta-6 desaturase. *The Journal of biological chemistry* **274**: 471-477.

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R *et al.* 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS genetics* **5**: e1000602.

Christensen BC, Marsit CJ. 2011. Epigenomics in environmental health. *Frontiers in genetics* **2**: 84.

Comb M, Goodman HM. 1990. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. *Nucleic acids research* **18**: 3975-3982.

Cravo M, Fidalgo P, Pereira AD, Gouveia-Oliveira A, Chaves P, Selhub J, Mason JB, Mira FC, Leitao CN. 1994. DNA methylation as an intermediate biomarker in colorectal cancer: modulation by folic acid supplementation. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation* **3**: 473-479.

Creemers EE, Tijsen AJ, Pinto YM. 2012. Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease? *Circulation research* **110**: 483-495.

Csankovszki G, Nagy A, Jaenisch R. 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *The Journal of cell biology* **153**: 773-784.

Das M, Sha J, Hidalgo B, Aslibekyan S, Do AN, Zhi D, Sun D, Zhang T, Li S, Chen W *et al.* 2016. Association of DNA Methylation at CPT1A Locus with Metabolic Syndrome in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study. *PloS one* **11**: e0145789.

Davalos A, Fernandez-Hernando C. 2013. From evolution to revolution: miRNAs as pharmacological targets for modulating cholesterol efflux and reverse cholesterol transport. *Pharmacological research* **75**: 60-72.

Davalos A, Goedeke L, Smibert P, Ramirez CM, Warrier NP, Andreo U, Cirera-Salinas D, Rayner K, Suresh U, Pastor-Pareja JC *et al.* 2011. miR-33a/b contribute to the regulation of fatty acid metabolism and insulin signaling. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 9232-9237.

Dayeh T, Tuomi T, Almgren P, Perfilyev A, Jansson PA, de Mello VD, Pihlajamaki J, Vaag A, Groop L, Nilsson E *et al.* 2016. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. *Epigenetics : official journal of the DNA Methylation Society* doi:10.1080/15592294.2016.1178418: 1-7.

De Lalla OF, Gofman JW. 1954. Ultracentrifugal analysis of serum lipoproteins. *Methods of biochemical analysis* **1**: 459-478.

DeBerardinis RJ. 2011. Serine metabolism: some tumors take the road less traveled. *Cell metabolism* **14**: 285-286.

Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, Mei H, Zhernakova DV, van den Berg LH, Deelen J *et al.* 2016. Blood lipids influence DNA methylation in circulating cells. *Genome biology* **17**: 138.

Delcuve GP, Rastegar M, Davie JR. 2009. Epigenetic control. *Journal of cellular physiology* **219**: 243-250.

Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, Hedman AK, Sandling JK, Li LA, Irvin MR *et al.* 2015. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Human molecular genetics* **24**: 4464-4479.

REFERENCES

DeNicola GM, Chen PH, Mullarky E, Sudderth JA, Hu Z, Wu D, Tang H, Xie Y, Asara JM, Huffman KE *et al.* 2015. NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nature genetics* **47**: 1475-1481.

Ding J, Li T, Wang X, Zhao E, Choi JH, Yang L, Zha Y, Dong Z, Huang S, Asara JM *et al.* 2013. The histone H3 methyltransferase G9A epigenetically activates the serine-glycine synthesis pathway to sustain cancer cell survival and proliferation. *Cell metabolism* **18**: 896-907.

Ding J, Reynolds LM, Zeller T, Muller C, Lohman K, Nicklas BJ, Kritchevsky SB, Huang Z, de la Fuente A, Soranzo N *et al.* 2015. Alterations of a Cellular Cholesterol Metabolism Network Are a Molecular Feature of Obesity-Related Type 2 Diabetes and Cardiovascular Disease. *Diabetes* **64**: 3464-3474.

Du P, Kibbe WA, Lin SM. 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**: 1547-1548.

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics* **11**: 587.

Duursma AM, Kedde M, Schrier M, le Sage C, Agami R. 2008. miR-148 targets human DNMT3b protein coding region. *Rna* **14**: 872-877.

Ecker J, Langmann T, Moehle C, Schmitz G. 2007. Isomer specific effects of Conjugated Linoleic Acid on macrophage ABCG1 transcription by a SREBP-1c dependent mechanism. *Biochemical and biophysical research communications* **352**: 805-811.

Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O. 2008. An R2 statistic for fixed effects in the linear mixed model. *Statistics in medicine* **27**: 6137-6157.

Elder SJ, Lichtenstein AH, Pittas AG, Roberts SB, Fuss PJ, Greenberg AS, McCrory MA, Bouchard TJ, Jr., Saltzman E, Neale MC. 2009. Genetic and environmental influences on factors associated with cardiovascular disease and the metabolic syndrome. *Journal of lipid research* **50**: 1917-1926.

Esterbauer H, Schaur RJ, Zollner H. 1991. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free radical biology & medicine* **11**: 81-128.

Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. 2001. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA : the journal of the American Medical Association* **285**: 2486-2497.

Fabregat-Traver D, Sharapov SZ, Hayward C, Rudan I, Campbell H, Aulchenko Y, Bientinesi P. 2014. High-Performance Mixed Models Based Genome-Wide Association Analysis with omicABEL software [version 1;referees: 2 approved, 1 approved with reservations]. *F1000Research 2014* **3:200**.

Feingold KR, Grunfeld C. 2000. Introduction to Lipids and Lipoproteins. In *Endotext*, (ed. LJ De Groot, *et al.*), South Dartmouth (MA).

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**: 398-402.

Flores AM, Gurevich I, Zhang C, Ramirez VP, Devens TR, Aneskievich BJ. 2011. TNIP1 is a corepressor of agonist-bound PPARs. *Archives of biochemistry and biophysics* **516**: 58-66.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J *et al.* 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 10604-10609.

Frazier-Wood AC, Aslibekyan S, Absher DM, Hopkins PH, Sha J, Tsai MY, Tiwari HK, Waite LL, Zhi D, Arnett DK. 2014. Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. *Journal of lipid research* doi:10.1194/jlr.M048504.

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. 2000. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature genetics* **24**: 88-91.

118

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. 2003. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *The Journal of biological chemistry* **278**: 4035-4040.

Gagnon F, Aissi D, Carrie A, Morange PE, Tregouet DA. 2014. Robust validation of methylation levels association at CPT1A locus with lipid plasma levels. *Journal of lipid research* doi:10.1194/jlr.E051276.

Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. 2015. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clinical epigenetics* **7**: 113.

Gelissen IC, Harris M, Rye KA, Quinn C, Brown AJ, Kockx M, Cartland S, Packianathan M, Kritharides L, Jessup W. 2006. ABCA1 and ABCG1 synergize to mediate cholesterol export to apoA-I. *Arteriosclerosis, thrombosis, and vascular biology* **26**: 534-540.

Genvigir FD, Rodrigues AC, Cerda A, Arazi SS, Willrich MA, Oliveira R, Hirata MH, Dorea EL, Bernik MM, Curi R *et al.* 2010. Effects of lipid-lowering drugs on reverse cholesterol transport gene expressions in peripheral blood mononuclear and HepG2 cells. *Pharmacogenomics* **11**: 1235-1246.

Gerin I, Clerbaux LA, Haumont O, Lanthier N, Das AK, Burant CF, Leclercq IA, MacDougald OA, Bommer GT. 2010. Expression of miR-33 from an SREBP2 intron inhibits cholesterol export and fatty acid oxidation. *The Journal of biological chemistry* **285**: 33652-33661.

Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J *et al.* 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS genetics* **6**: e1000952.

Global Lipids Genetics C Willer CJ Schmidt EM Sengupta S Peloso GM Gustafsson S Kanoni S Ganna A Chen J Buchkovich ML *et al.* 2013. Discovery and refinement of loci associated with lipid levels. *Nature genetics* **45**: 1274-1283.

Goldberg AD, Allis CD, Bernstein E. 2007. Epigenetics: a landscape takes shape. *Cell* **128**: 635-638.

Goode EL, Cherny SS, Christian JC, Jarvik GP, de Andrade M. 2007. Heritability of longitudinal measures of body mass index and lipid and lipoprotein levels in aging twins. *Twin research and human genetics : the official journal of the International Society for Twin Studies* **10**: 703-711.

Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, Jacobs DR, Jr., Bangdiwala S, Tyroler HA. 1989. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation* **79**: 8-15.

Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. 2013. Meta-analysis: pitfalls and hints. *Heart, lung and vessels* **5**: 219-225.

Gresch O, Engel FB, Nesic D, Tran TT, England HM, Hickman ES, Korner I, Gan L, Chen S, Castro-Obregon S *et al.* 2004. New non-viral method for gene transfer into primary cells. *Methods* **33**: 151-163.

Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, Busche S, Yuan W, Nisbet J, Sekowska M *et al.* 2013. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *American journal of human genetics* **93**: 876-890.

Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A *et al.* 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature genetics* **44**: 1084-1089.

Guay SP, Brisson D, Lamarche B, Gaudet D, Bouchard L. 2014. Epipolymorphisms within lipoprotein genes contribute independently to plasma lipid levels in familial hypercholesterolemia. *Epigenetics : official journal of the DNA Methylation Society* **9**.

Guay SP, Brisson D, Lamarche B, Marceau P, Vohl MC, Gaudet D, Bouchard L. 2013. DNA methylation variations at CETP and LPL gene promoter loci: New molecular biomarkers associated with blood lipid profile variability. In *Atherosclerosis*, Vol 228, pp. 413-420.

Guay SP, Brisson D, Munger J, Lamarche B, Gaudet D, Bouchard L. 2012a. ABCA1 gene promoter DNA methylation is associated with HDL particle profile and coronary artery disease in familial hypercholesterolemia. *Epigenetics : official journal of the DNA Methylation Society* **7**: 464-472.

Guay SP, Voisin G, Brisson D, Munger J, Lamarche B, Gaudet D, Bouchard L. 2012b. Epigenome-wide analysis in familial hypercholesterolemia identified new loci associated with high-density lipoprotein cholesterol concentration. *Epigenomics* **4**: 623-639.

Gurevich I, Aneskievich BJ. 2009. Liganded RARalpha and RARgamma interact with but are repressed by TNIP1. *Biochemical and biophysical research communications* **389**: 409-414.

Gurha P. 2016. MicroRNAs in cardiovascular disease. *Current opinion in cardiology* doi:10.1097/HCO.0000000000000280.

Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A *et al.* 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**: e00523.

Habib E, Linher-Melville K, Lin HX, Singh G. 2015. Expression of xCT and activity of system xc(-) are regulated by NRF2 in human breast cancer cells in response to oxidative stress. *Redox biology* **5**: 33-42.

Hidalgo B, Irvin MR, Sha J, Zhi D, Aslibekyan S, Absher D, Tiwari HK, Kabagambe EK, Ordovas JM, Arnett DK. 2014. Epigenome-Wide Association Study of Fasting Measures of Glucose, Insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network Study. *Diabetes* **63**: 801-807.

Holle R, Happich M, Lowel H, Wichmann HE, Group MKS. 2005. KORA--a research platform for population based health research. *Gesundheitswesen* **67 Suppl 1**: S19-25.

Holliday R, Pugh JE. 1975. DNA modification mechanisms and gene activity during development. *Science* **187**: 226-232.

Hoogeveen RC, Gaubatz JW, Sun W, Dodge RC, Crosby JR, Jiang J, Couper D, Virani SS, Kathiresan S, Boerwinkle E *et al.* 2014. Small dense low-density lipoprotein-cholesterol concentrations predict risk for coronary heart disease: the Atherosclerosis Risk In Communities (ARIC) study. *Arteriosclerosis, thrombosis, and vascular biology* **34**: 1069-1077.

Horie T, Nishino T, Baba O, Kuwabara Y, Nakao T, Nishiga M, Usami S, Izuhara M, Nakazeki F, Ide Y *et al.* 2014. MicroRNA-33b knock-in mice for an intron of sterol regulatory element-binding factor 1 (Srebf1) exhibit reduced HDL-C in vivo. *Scientific reports* **4**: 5312.

Horie T, Ono K, Horiguchi M, Nishi H, Nakamura T, Nagao K, Kinoshita M, Kuwabara Y, Marusawa H, Iwanaga Y *et al.* 2010. MicroRNA-33 encoded by an intron of sterol regulatory element-binding protein 2 (Srebp2) regulates HDL in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17321-17326.

Horton JD, Goldstein JL, Brown MS. 2002. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *The Journal of clinical investigation* **109**: 1125-1131.

Hou L, Zhang X, Wang D, Baccarelli A. 2012. Environmental chemical exposures and human epigenetics. *International journal of epidemiology* **41**: 79-105.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* **13**: 86.

Houseman EA, Molitor J, Marsit CJ. 2014. Reference-Free Cell Mixture Adjustments in Analysis of DNA Methylation Data. *Bioinformatics* doi:10.1093/bioinformatics/btu029.

Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C *et al.* 2013. DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**: e00726.

Hu W, Feng Z, Eveleigh J, Iyer G, Pan J, Amin S, Chung FL, Tang MS. 2002. The major lipid peroxidation product, trans-4-hydroxy-2-nonenal, preferentially forms DNA adducts at codon 249 of human p53 gene, a unique mutational hotspot in hepatocellular carcinoma. *Carcinogenesis* **23**: 1781-1789.

Huisinga KL, Brower-Toland B, Elgin SC. 2006. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**: 110-122.

Illingworth RS, Bird AP. 2009. CpG islands--'a rough guide'. *FEBS letters* **583**: 1713-1720.

Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hamalainen E, Jousilahti P, Kangas AJ, Mannisto S *et al.* 2010. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Molecular systems biology* **6**: 441.

Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, Thibeault KS, Patel N, Day K, Jones LW *et al.* 2014. Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study. *Circulation* **130**: 565-572.

Ishii T, Itoh K, Takahashi S, Sato H, Yanagawa T, Katoh Y, Bannai S, Yamamoto M. 2000. Transcription factor Nrf2 coordinately regulates a group of oxidative stress-inducible genes in macrophages. *The Journal of biological chemistry* **275**: 16023-16029.

Isomaa B, Almgren P, Tuomi T, Forsen B, Lahti K, Nissen M, Taskinen MR, Groop L. 2001. Cardiovascular morbidity and mortality associated with the metabolic syndrome. *Diabetes care* **24**: 683-689.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**: 1300-1303.

Jacob RA, Gretz DM, Taylor PC, James SJ, Pogribny IP, Miller BJ, Henning SM, Swendseid ME. 1998. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *The Journal of nutrition* **128**: 1204-1212.

Jeltsch A. 2006. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics : official journal of the DNA Methylation Society* **1**: 63-66.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074-1080.

Jimenez-Chillaron JC, Diaz R, Martinez D, Pentinat T, Ramon-Krauel M, Ribo S, Plosch T. 2012. The role of nutrition on epigenetic modifications and their implications on health. *Biochimie* **94**: 2242-2263.

Johannes F, Colot V, Jansen RC. 2008. Epigenome dynamics: a quantitative genetics perspective. *Nature reviews Genetics* **9**: 883-890.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews Genetics* **13**: 484-492.

Jones PA, Takai D. 2001. The role of DNA methylation in mammalian epigenetics. *Science* **293**: 1068-1070.

Joy T, Hegele RA. 2008. Is raising HDL a futile strategy for atheroprotection? *Nature reviews Drug discovery* **7**: 143-155.

Jun HJ, Hoang MH, Yeo SK, Jia Y, Lee SJ. 2013. Induction of ABCA1 and ABCG1 expression by the liver X receptor modulator cineole in macrophages. *Bioorganic & medicinal chemistry letters* **23**: 579-583.

Junn E, Han SH, Im JY, Yang Y, Cho EW, Um HD, Kim DK, Lee KW, Han PL, Rhee SG *et al.* 2000. Vitamin D3 up-regulated protein 1 mediates oxidative stress via suppressing the thioredoxin function. *Journal of immunology* **164**: 6287-6295.

Jurkowska RZ, Jurkowski TP, Jeltsch A. 2011. Structure and function of mammalian DNA methyltransferases. *Chembiochem : a European journal of chemical biology* **12**: 206-222.

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**: 900-903.

Kennedy MA, Barrera GC, Nakamura K, Baldan A, Tarr P, Fishbein MC, Frank J, Francone OL, Edwards PA. 2005. ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. *Cell metabolism* **1**: 121-131.

Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikainen LP, Kangas AJ, Soininen P, Wurtz P, Silander K *et al.* 2012. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature genetics* **44**: 269-276.

REFERENCES

Khavari DA, Sen GL, Rinn JL. 2010. DNA methylation and epigenetic control of cellular differentiation. *Cell cycle* **9**: 3880-3883.

Khorasanizadeh S. 2004. The nucleosome: from genomic organization to genomic regulation. *Cell* **116**: 259-272.

Klengel T, Binder EB. 2015. Epigenetics of Stress-Related Psychiatric Disorders and Gene x Environment Interactions. *Neuron* **86**: 1343-1357.

Kornberg RD, Lorch Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**: 285-294.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* **42**: D68-73.

Krauss RM. 2010. Lipoprotein subfractions and cardiovascular disease risk. *Current opinion in lipidology* **21**: 305-311.

Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**: 929-930.

Kriebel J, Herder C, Rathmann W, Wahl S, Kunze S, Molnos S, Volkova N, Schramm K, Carstensen-Kirberg M, Waldenberger M *et al.* 2016. Association between DNA Methylation in Whole Blood and Measures of Glucose Metabolism: KORA F4 Study. *PloS one* **11**: e0152314.

Krishnan HR, Sakharkar AJ, Teppen TL, Berkel TD, Pandey SC. 2014. The epigenetic landscape of alcoholism. *International review of neurobiology* **115**: 75-116.

Ku CS, Park Y, Coleman SL, Lee J. 2012. Unsaturated fatty acids repress expression of ATP binding cassette transporter A1 and G1 in RAW 264.7 macrophages. *The Journal of nutritional biochemistry* **23**: 1271-1276.

Kwiterovich PO, Jr. 2000. The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides: a current review. *The American journal of cardiology* **86**: 5L-10L.

Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, Salonen JT. 2002. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA : the journal of the American Medical Association* **288**: 2709-2716.

Lam JK, Chow MY, Zhang Y, Leung SW. 2015. siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular therapy Nucleic acids* **4**: e252.

Lattka E, Illig T, Heinrich J, Koletzko B. 2009. FADS gene cluster polymorphisms: important modulators of fatty acid levels and their impact on atopic diseases. *Journal of nutrigenetics and nutrigenomics* **2**: 119-128.

Lattka E, Illig T, Koletzko B, Heinrich J. 2010. Genetic variants of the FADS1 FADS2 gene cluster as related to essential fatty acid metabolism. *Current opinion in lipidology* **21**: 64-69.

Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P *et al.* 2015. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome biology* **16**: 37.

Lewerenz J, Sato H, Albrecht P, Henke N, Noack R, Methner A, Maher P. 2012. Mutation of ATF4 mediates resistance of neuronal cell lines against oxidative stress by inducing xCT expression. *Cell death and differentiation* **19**: 847-858.

Li AC, Binder CJ, Gutierrez A, Brown KK, Plotkin CR, Pattison JW, Valledor AF, Davis RA, Willson TM, Witztum JL *et al.* 2004. Differential inhibition of macrophage foam-cell formation and atherosclerosis in mice by PPARalpha, beta/delta, and gamma. *The Journal of clinical investigation* **114**: 1564-1576.

Li AC, Glass CK. 2004. PPAR- and LXR-dependent pathways controlling lipid metabolism and the development of atherosclerosis. *Journal of lipid research* **45**: 2161-2173.

Li K-H, Meng X-L, Raghunathan TE, Rubin DB. 1991. Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica* **1**: 65-92.

Liang G, Yang J, Horton JD, Hammer RE, Goldstein JL, Brown MS. 2002. Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *The Journal of biological chemistry* **277**: 9520-9528.

Liep J, Rabien A, Jung K. 2012. Feedback networks between microRNAs and epigenetic modifications in urological tumors. *Epigenetics : official journal of the DNA Methylation Society* **7**: 315-325.

Lim JC, Donaldson PJ. 2011. Focus on molecules: the cystine/glutamate exchanger (System x(c)(-)). *Experimental eye research* **92**: 162-163.

Liu X, Miyazaki M, Flowers MT, Sampath H, Zhao M, Chu K, Paton CM, Joo DS, Ntambi JM. 2010. Loss of Stearoyl-CoA desaturase-1 attenuates adipocyte inflammation: effects of adipocyte-derived oleate. *Arteriosclerosis, thrombosis, and vascular biology* **30**: 31-38.

Liu Y, Li X, Aryee MJ, Ekstrom TJ, Padyukov L, Klareskog L, Vandiver A, Moore AZ, Tanaka T, Ferrucci L *et al.* 2014. GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *American journal of human genetics* **94**: 485-495.

Liu Y, Mounkes LC, Liggitt HD, Brown CS, Solodin I, Heath TD, Debs RJ. 1997. Factors influencing the efficiency of cationic liposome-mediated intravenous gene delivery. *Nature biotechnology* **15**: 167-173.

Lo M, Wang YZ, Gout PW. 2008. The x(c)- cystine/glutamate antiporter: a potential target for therapy of cancer and other diseases. *Journal of cellular physiology* **215**: 593-602.

Locasale JW. 2013. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nature reviews Cancer* **13**: 572-583.

Lowel H, Meisinger C, Heier M, Hormann A. 2005. The population-based acute myocardial infarction (AMI) registry of the MONICA/KORA study region of Augsburg. *Gesundheitswesen* **67 Suppl 1**: S31-37.

Macfarlane LA, Murphy PR. 2010. MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics* **11**: 537-561.

Malodobra-Mazur M, Dziewulska A, Kozinski K, Dobrzyn P, Kolczynska K, Janikiewicz J, Dobrzyn A. 2014. Stearoyl-CoA desaturase regulates inflammatory gene expression by changing DNA methylation level in 3T3 adipocytes. *Int J Biochem Cell Biol* **55**: 40-50.

Mamtani M, Kulkarni H, Dyer TD, Goring HH, Neary JL, Cole SA, Kent JW, Kumar S, Glahn DC, Mahaney MC *et al.* 2016. Genome- and epigenome-wide association study of hypertriglyceridemic waist in Mexican American families. *Clinical epigenetics* **8**: 6.

Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR, Vinson C. 2013. CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome research* **23**: 988-997.

Martin-Nunez GM, Cabrera-Mulero R, Rubio-Martin E, Rojo-Martinez G, Olveira G, Valdes S, Soriguer F, Castano L, Morcillo S. 2014. Methylation levels of the SCD1 gene promoter and LINE-1 repeat region are associated with weight change: an intervention study. *Mol Nutr Food Res* **58**: 1528-1536.

Martin C, Zhang Y. 2005. The diverse functions of histone lysine methylation. *Nature reviews Molecular cell biology* **6**: 838-849.

Martino F, Carlomosti F, Avitabile D, Persico L, Picozza M, Barilla F, Arca M, Montali A, Martino E, Zanoni C *et al.* 2015. Circulating miR-33a and miR-33b are up-regulated in familial hypercholesterolaemia in paediatric age. *Clin Sci (Lond)* **129**: 963-972.

Matsuzaka T, Shimano H, Yahagi N, Amemiya-Kudo M, Yoshikawa T, Hasty AH, Tamura Y, Osuga J, Okazaki H, Iizuka Y *et al.* 2002. Dual regulation of mouse Delta(5)- and Delta(6)-desaturase gene expression by SREBP-1 and PPARalpha. *Journal of lipid research* **43**: 107-114.

Mattick JS, Makunin IV. 2005. Small regulatory RNAs in mammals. *Human molecular genetics* **14 Spec No 1**: R121-132.

Mauerer R, Ebert S, Langmann T. 2009. High glucose, unsaturated and saturated fatty acids differentially regulate expression of ATP-binding cassette transporters ABCA1 and ABCG1 in human macrophages. *Experimental & molecular medicine* **41**: 126-132.

REFERENCES

Mazzone T, Chait A, Plutzky J. 2008. Cardiovascular disease risk in type 2 diabetes mellitus: insights from mechanistic studies. *Lancet* **371**: 1800-1809.

McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwards JR, Challen GA. 2016. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biology open* doi:10.1242/bio.019067.

McGarry JD, Brown NF. 1997. The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. *European journal of biochemistry / FEBS* **244**: 1-14.

Michell DL, Vickers KC. 2016. Lipoprotein carriers of microRNAs. *Biochimica et biophysica acta* doi:10.1016/j.bbalip.2016.01.011.

Miyazaki M, Ntambi JM. 2003. Role of stearoyl-coenzyme A desaturase in lipid metabolism. *Prostaglandins Leukot Essent Fatty Acids* **68**: 113-121.

Moen EL, Zhang X, Mu W, Delaney SM, Wing C, McQuade J, Myers J, Godley LA, Dolan ME, Zhang W. 2013. Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. *Genetics* **194**: 987-996.

Moldovan L, Batte KE, Trgovcich J, Wisler J, Marsh CB, Piper M. 2014. Methodological challenges in utilizing miRNAs as circulating biomarkers. *Journal of cellular and molecular medicine* **18**: 371-390.

Mooradian AD. 2009. Dyslipidemia in type 2 diabetes mellitus. *Nature clinical practice Endocrinology & metabolism* **5**: 150-159.

Mosammaparast N, Shi Y. 2010. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annual review of biochemistry* **79**: 155-179.

Mullarky E, Mattaini KR, Vander Heiden MG, Cantley LC, Locasale JW. 2011. PHGDH amplification and altered glucose metabolism in human melanoma. *Pigment cell & melanoma research* **24**: 1112-1115.

Mullen AR, DeBerardinis RJ. 2012. Genetically-defined metabolic reprogramming in cancer. *Trends in endocrinology and metabolism: TEM* **23**: 552-559.

Najafi-Shoushtari SH, Kristo F, Li Y, Shioda T, Cohen DE, Gerszten RE, Naar AM. 2010. MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis. *Science* **328**: 1566-1569.

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**: 110-113.

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**: 386-389.

National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. 2002. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* **106**: 3143-3421.

Niculescu LS, Simionescu N, Sanda GM, Carnuta MG, Stancu CS, Popescu AC, Popescu MR, Vlad A, Dimulescu DR, Simionescu M *et al.* 2015. MiR-486 and miR-92a Identified in Circulating HDL Discriminate between Stable and Vulnerable Coronary Artery Disease Patients. *PloS one* **10**: e0140958.

Niculescu MD, Lupu DS, Craciunescu CN. 2013. Perinatal manipulation of alpha-linolenic acid intake induces epigenetic changes in maternal and offspring livers. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **27**: 350-358.

Nielsen S, Karpe F. 2012. Determinants of VLDL-triglycerides production. *Current opinion in lipidology* **23**: 321-326.

Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, Poulsen P, Ribel-Madsen R, Pedersen NL, Almgren P *et al.* 2014. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* **63**: 2962-2976.

Noma K, Allis CD, Grewal SI. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**: 1150-1155.

Nordestgaard BG, Benn M, Schnohr P, Tybjaerg-Hansen A. 2007. Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. *JAMA : the journal of the American Medical Association* **298**: 299-308.

Ntambi JM. 1999. Regulation of stearoyl-CoA desaturase by polyunsaturated fatty acids and cholesterol. *Journal of lipid research* **40**: 1549-1558.

O'Neill EA, Fletcher C, Burrow CR, Heintz N, Roeder RG, Kelly TJ. 1988. Transcription factor OTF-1 is functionally identical to the DNA replication factor NF-III. *Science* **241**: 1210-1213.

Oka S, Yoshihara E, Bizen-Abe A, Liu W, Watanabe M, Yodoi J, Masutani H. 2009. Thioredoxin binding protein-2/thioredoxin-interacting protein is a critical regulator of insulin secretion and peroxisome proliferator-activated receptor function. *Endocrinology* **150**: 1225-1234.

Ono K, Horie T, Nishino T, Baba O, Kuwabara Y, Yokode M, Kita T, Kimura T. 2015. MicroRNA-33a/b in Lipid Metabolism. *Circulation journal : official journal of the Japanese Circulation Society* **79**: 278-284.

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P *et al.* 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**: 394-397.

Paton CM, Ntambi JM. 2009. Biochemical and physiological function of stearoyl-CoA desaturase. *American journal of physiology Endocrinology and metabolism* **297**: E28-37.

Peng D, Hiipakka RA, Xie JT, Dai Q, Kokontis JM, Reardon CA, Getz GS, Liao S. 2011. A novel potent synthetic steroidal liver X receptor agonist lowers plasma cholesterol and triglycerides and reduces atherosclerosis in LDLR(-/-) mice. *British journal of pharmacology* **162**: 1792-1804.

Peng P, Wang L, Yang X, Huang X, Ba Y, Chen X, Guo J, Lian J, Zhou J. 2014. A preliminary study of the relationship between promoter methylation of the ABCG1, GALNT2 and HMGCR genes and coronary heart disease. *PloS one* **9**: e102265.

Petersen AK, Stark K, Musameh MD, Nelson CP, Romisch-Margl W, Kremer W, Raffler J, Krug S, Skurk T, Rist MJ *et al.* 2012. Genetic associations with lipoprotein subfractions provide information on their biological nature. *Human molecular genetics* **21**: 1433-1443.

Petersen AK, Zeilinger S, Kastenmuller G, Romisch-Margl W, Brugger M, Peters A, Meisinger C, Strauch K, Hengstenberg C, Pagel P *et al.* 2014. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Human molecular genetics* **23**: 534-545.

Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, Holdt LM, Kretschmer A, Schramm K, Adamski J *et al.* 2015. DNA methylation of lipid-related genes affects blood lipid levels. *Circulation Cardiovascular genetics* **8**: 334-342.

Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SR. 2012. The impact of recent alcohol use on genome wide DNA methylation signatures. *Frontiers in genetics* **3**: 54.

Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, Jha AK *et al.* 2011. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**: 346-350.

Proctor SD, Mamo JC. 1998. Retention of fluorescent-labelled chylomicron remnants within the intima of the arterial wall--evidence that plaque cholesterol may be derived from post-prandial lipoproteins. *European journal of clinical investigation* **28**: 497-503.

Pufulete M, Al-Ghnaniem R, Khushal A, Appleby P, Harris N, Gout S, Emery PW, Sanders TA. 2005. Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. *Gut* **54**: 648-653.

Ramasamy I. 2014. Recent advances in physiological lipoprotein metabolism. *Clinical chemistry and laboratory medicine : CCLM / FESCC* **52**: 1695-1727.

Ramirez CM, Goedeke L, Rotllan N, Yoon JH, Cirera-Salinas D, Mattison JA, Suarez Y, de Cabo R, Gorospe M, Fernandez-Hernando C. 2013. MiR-33 regulates glucose metabolism. *Molecular and cellular biology* doi:10.1128/MCB.00016-13.

REFERENCES

Ramirez VP, Gurevich I, Aneskievich BJ. 2012. Emerging roles for TNIP1 in regulating post-receptor signaling. *Cytokine & growth factor reviews* **23**: 109-118.

Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. 2000. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. *The American journal of clinical nutrition* **72**: 998-1003.

Rayner KJ, Esau CC, Hussain FN, McDaniel AL, Marshall SM, van Gils JM, Ray TD, Sheedy FJ, Goedeke L, Liu X *et al.* 2011. Inhibition of miR-33a/b in non-human primates raises plasma HDL and lowers VLDL triglycerides. *Nature* **478**: 404-407.

Rayner KJ, Suarez Y, Davalos A, Parathath S, Fitzgerald ML, Tamehiro N, Fisher EA, Moore KJ, Fernandez-Hernando C. 2010. MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* **328**: 1570-1573.

Razin A, Cedar H. 1977. Distribution of 5-methylcytosine in chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 2725-2728.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD *et al.* 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593-599.

Reardon HT, Zhang J, Kothapalli KS, Kim AJ, Park WJ, Brenna JT. 2012. Insertion-deletions in a FADS2 intron 1 conserved regulatory locus control expression of fatty acid desaturases 1 and 2 and modulate response to simvastatin. *Prostaglandins Leukot Essent Fatty Acids* **87**: 25-33.

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J. 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PloS one* **7**: e41361.

Riggs AD. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenetics and cell genetics* **14**: 9-25.

Rohde K, Keller M, Stumvoll M, Dietrich A, Bluher M, Bottcher Y. 2015. DNA 5-hydroxymethylation in human adipose tissue differs between subcutaneous and visceral adipose tissue depots. *Epigenomics* **7**: 911-920.

Rosenfeld MG, Lunyak VV, Glass CK. 2006. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes & development* **20**: 1405-1428.

Rottiers V, Naar AM. 2012. MicroRNAs in metabolism and metabolic disorders. *Nature reviews Molecular cell biology* **13**: 239-250.

Rottiers V, Obad S, Petri A, McGarrah R, Lindholm MW, Black JC, Sinha S, Goody RJ, Lawrence MS, deLemos AS *et al.* 2013. Pharmacological inhibition of a microRNA family in nonhuman primates by a seed-targeting 8-mer antimiR. *Science translational medicine* **5**: 212ra162.

Saito Y, Jones PA. 2006. Epigenetic activation of tumor suppressor microRNAs in human cancer cells. *Cell cycle* **5**: 2220-2222.

Sampath H, Ntambi JM. 2005. Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annual review of nutrition* **25**: 317-340.

Sandoval-Hernandez AG, Hernandez HG, Restrepo A, Munoz JI, Bayon GF, Fernandez AF, Fraga MF, Cardona-Gomez GP, Arboleda H, Arboleda GH. 2016. Liver X Receptor Agonist Modifies the DNA Methylation Profile of Synapse and Neurogenesis-Related Genes in the Triple Transgenic Mouse Model of Alzheimer's Disease. *Journal of molecular neuroscience : MN* **58**: 243-253.

Sasaki H, Sato H, Kuriyama-Matsumura K, Sato K, Maebara K, Wang H, Tamba M, Itoh K, Yamamoto M, Bannai S. 2002. Electrophile response element-mediated induction of the cystine/glutamate exchange transporter gene expression. *The Journal of biological chemistry* **277**: 44765-44771.

Schaeffer L, Gohlke H, Muller M, Heid IM, Palmer LJ, Kompauer I, Demmelmair H, Illig T, Koletzko B, Heinrich J. 2006. Common genetic variants of the FADS1 FADS2 gene cluster and their reconstructed haplotypes are associated with the fatty acid composition in phospholipids. *Human molecular genetics* **15**: 1745-1756.

Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, Dorr M, Endlich K, Felix SB, Gieger C, Grallert H *et al.* 2012. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PloS one* **7**: e50938.

Schwenk RW, Jonas W, Ernst SB, Kammel A, Jahnert M, Schurmann A. 2013. Diet-dependent alterations of hepatic Scd1 expression are accompanied by differences in promoter methylation. *Horm Metab Res* **45**: 786-794.

Shimomura I, Shimano H, Korn BS, Bashmakov Y, Horton JD. 1998. Nuclear sterol regulatory element-binding proteins activate genes responsible for the entire program of unsaturated fatty acid biosynthesis in transgenic mouse liver. *The Journal of biological chemistry* **273**: 35299-35306.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74-79.

Simopoulos AP. 2010. Genetic variants in the metabolism of omega-6 and omega-3 fatty acids: their role in the determination of nutritional requirements and chronic disease risk. *Experimental biology and medicine* **235**: 785-795.

Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. 2015. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation Cardiovascular genetics* **8**: 192-206.

Soininen P, Kangas AJ, Wurtz P, Tukiainen T, Tynkkynen T, Laatikainen R, Jarvelin MR, Kahonen M, Lehtimaki T, Viikari J *et al.* 2009. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *The Analyst* **134**: 1781-1785.

Spickett CM. 2013. The lipid peroxidation product 4-hydroxy-2-nonenal: Advances in chemistry and analysis. *Redox biology* **1**: 145-152.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41-45.

Subramaniam S, Fahy E, Gupta S, Sud M, Byrnes RW, Cotter D, Dinasarapu AR, Maurya MR. 2011. Bioinformatics and systems biology of the lipidome. *Chemical reviews* **111**: 6452-6490.

Superko HR, Pendyala L, Williams PT, Momary KM, King SB, 3rd, Garrett BC. 2012. High-density lipoprotein subclasses and their relationship to cardiovascular disease. *Journal of clinical lipidology* **6**: 496-523.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L *et al.* 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.

Tarling EJ, Edwards PA. 2011. ATP binding cassette transporter G1 (ABCG1) is an intracellular sterol transporter. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 19719-19724.

Tate PH, Bird AP. 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. *Current opinion in genetics & development* **3**: 226-231.

Telama R, Yang X, Viikari J, Valimaki I, Wanne O, Raitakari O. 2005. Physical activity from childhood to adulthood: a 21-year tracking study. *American journal of preventive medicine* **28**: 267-273.

Teslovich TM Musunuru K Smith AV Edmondson AC Stylianou IM Koseki M Pirruccello JP Ripatti S Chasman DI Willer CJ *et al.* 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**: 707-713.

Tian L, Li C, Liu Y, Chen Y, Fu M. 2014. The value and distribution of high-density lipoprotein subclass in patients with acute coronary syndrome. *PloS one* **9**: e85114.

Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, Slieker RC, Stok AP, Thijssen PE, Muller F *et al.* 2014. DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nature communications* **5**: 5592.

Tost J. 2010. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Molecular biotechnology* **44**: 71-81.

REFERENCES

Tukiainen T, Tynkkynen T, Makinen VP, Jylanki P, Kangas A, Hokkanen J, Vehtari A, Grohn O, Hallikainen M, Soininen H *et al.* 2008. A multi-metabolite analysis of serum by 1H NMR spectroscopy: early systemic signs of Alzheimer's disease. *Biochemical and biophysical research communications* **375**: 356-361.

Tulenko TN, Sumner AE. 2002. The physiology of lipoproteins. *Journal of nuclear cardiology : official publication of the American Society of Nuclear Cardiology* **9**: 638-649.

Uehara Y, Miura S, von Eckardstein A, Abe S, Fujii A, Matsuo Y, Rust S, Lorkowski S, Assmann G, Yamada T *et al.* 2007. Unsaturated fatty acids suppress the expression of the ATP-binding cassette transporter G1 (ABCG1) and ABCA1 genes via an LXR/RXR responsive element. *Atherosclerosis* **191**: 11-21.

Ushijima T. 2005. Detection and interpretation of altered methylation patterns in cancer cells. *Nature reviews Cancer* **5**: 223-231.

Vaissiere T, Sawan C, Herceg Z. 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutation research* **659**: 40-48.

Van den Veyver IB. 2002. Genetic effects of methylation diets. *Annual review of nutrition* **22**: 255-282.

van Eijk KR, de Jong S, Boks MP, Langeveld T, Colas F, Veldink JH, de Kovel CG, Janson E, Strengman E, Langfelder P *et al.* 2012. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC genomics* **13**: 636.

van Rooij E, Purcell AL, Levin AA. 2012. Developing microRNA therapeutics. *Circulation research* **110**: 496-507.

Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, Klasic M, Zoldos V. 2016. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic acids research* doi:10.1093/nar/gkw159.

Wade PA. 2001. Methyl CpG-binding proteins and transcriptional repression. *BioEssays : news and reviews in molecular, cellular and developmental biology* **23**: 1131-1137.

Wang J, Tang J, Lai M, Zhang H. 2014. 5-Hydroxymethylcytosine and disease. *Mutation research Reviews in mutation research* **762**: 167-175.

Wang L, Athinarayanan S, Jiang G, Chalasani N, Zhang M, Liu W. 2015. Fatty acid desaturase 1 gene polymorphisms control human hepatic lipid composition. *Hepatology* **61**: 119-128.

Wang N, Lan D, Chen W, Matsuura F, Tall AR. 2004. ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 9774-9779.

Wang W, Song W, Wang Y, Chen L, Yan X. 2013. HMG-CoA reductase inhibitors, simvastatin and atorvastatin, downregulate ABCG1-mediated cholesterol efflux in human macrophages. *J Cardiovasc Pharmacol* **62**: 90-98.

Watanabe R, Nakamura H, Masutani H, Yodoi J. 2010. Anti-oxidative, anti-cancer and anti-inflammatory actions by thioredoxin 1 and thioredoxin-binding protein-2. *Pharmacology & therapeutics* **127**: 261-270.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* **39**: 457-466.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L *et al.* 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* **42**: D1001-1006.

Wichmann HE, Gieger C, Illig T, Group MKS. 2005. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**: S26-30.

Wijnands KP, Obermann-Borst SA, Steegers-Theunissen RP. 2015. Early life lipid profile and metabolic programming in very young children. *Nutrition, metabolism, and cardiovascular diseases : NMCD* doi:10.1016/j.numecd.2015.02.010.

Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J. 2010. A longitudinal study of epigenetic variation in twins. *Epigenetics : official journal of the DNA Methylation Society* **5**: 516-526.

Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & development* **25**: 679-684.

Xu J, Nakamura MT, Cho HP, Clarke SD. 1999. Sterol regulatory element binding protein-1 expression is suppressed by dietary polyunsaturated fatty acids. A mechanism for the coordinate suppression of lipogenic genes by polyunsaturated fats. *The Journal of biological chemistry* **274**: 23577-23583.

Yazdanyar A, Yeang C, Jiang XC. 2011. Role of phospholipid transfer protein in high-density lipoprotein- mediated reverse cholesterol transport. *Current atherosclerosis reports* **13**: 242-248.

Ye J, Mancuso A, Tong X, Ward PS, Fan J, Rabinowitz JD, Thompson CB. 2012. Pyruvate kinase M2 promotes de novo serine synthesis to sustain mTORC1 activity and cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 6904-6909.

Ye P, Mimura J, Okada T, Sato H, Liu T, Maruyama A, Ohyama C, Itoh K. 2014. Nrf2- and ATF4-dependent upregulation of xCT modulates the sensitivity of T24 bladder carcinoma cells to proteasome inhibition. *Molecular and cellular biology* **34**: 3421-3434.

Yin H, Song P, Su R, Yang G, Dong L, Luo M, Wang B, Gong B, Liu C, Song W *et al.* 2016. DNA Methylation mediated down-regulating of MicroRNA-33b and its role in gastric cancer. *Scientific reports* **6**: 18824.

Yoshikawa T, Shimano H, Yahagi N, Ide T, Amemiya-Kudo M, Matsuzaka T, Nakakuki M, Tomita S, Okazaki H, Tamura Y *et al.* 2002. Polyunsaturated fatty acids suppress sterol regulatory element-binding protein 1c promoter activity by inhibition of liver X receptor (LXR) binding to LXR response elements. *The Journal of biological chemistry* **277**: 1705-1711.

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A *et al.* 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS one* **8**: e63812.

Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, Craig DW, Redman M, Gershon ES, Liu C. 2010. Genetic control of individual differences in gene-specific methylation in human brain. *American journal of human genetics* **86**: 411-419.

Zhang X, Zhang JH, Chen XY, Hu QH, Wang MX, Jin R, Zhang QY, Wang W, Wang R, Kang LL *et al.* 2015. Reactive oxygen species-induced TXNIP drives fructose-mediated hepatic inflammation and lipid accumulation through NLRP3 inflammasome activation. *Antioxidants & redox signaling* **22**: 848-870.

Ziegler A, König I. 2010. A statistical approach to genetic epidemiology: concepts and applications. *Wiley-VCH, Weilheim* **2. Edition**.

Zivkovic AM, German JB, Sanyal AJ. 2007. Comparative review of diets for the metabolic syndrome: implications for nonalcoholic fatty liver disease. *The American journal of clinical nutrition* **86**: 285-300.

7 APPENDIX

7.1 Tables

Table I: List of metabolites quantified via the ^{1}H NMR platform in KORA F4. N metabolites=228

Name	Abbr.	Unit
Concentration of chylomicrons and extremely large VLDL particles	XXL-VLDL-P	mol/l
Total lipids in chylomicrons and extremely large VLDL	XXL-VLDL-L	mmol/l
Phospholipids in chylomicrons and extremely large VLDL	XXL-VLDL-PL	mmol/l
Total cholesterol in chylomicrons and extremely large VLDL	XXL-VLDL-C	mmol/l
Cholesterol esters in chylomicrons and extremely large VLDL	XXL-VLDL-CE	mmol/l
Free cholesterol in chylomicrons and extremely large VLDL	XXL-VLDL-FC	mmol/l
Triglycerides in chylomicrons and extremely large VLDL	XXL-VLDL-TG	mmol/l
Concentration of very large VLDL particles	XL-VLDL-P	mol/l
Total lipids in very large VLDL	XL-VLDL-L	mmol/l
Phospholipids in very large VLDL	XL-VLDL-PL	mmol/l
Total cholesterol in very large VLDL	XL-VLDL-C	mmol/l
Cholesterol esters in very large VLDL	XL-VLDL-CE	mmol/l
Free cholesterol in very large VLDL	XL-VLDL-FC	mmol/l
Triglycerides in very large VLDL	XL-VLDL-TG	mmol/l
Concentration of large VLDL particles	L-VLDL-P	mol/l
Total lipids in large VLDL	L-VLDL-L	mmol/l
Phospholipids in large VLDL	L-VLDL-PL	mmol/l
Total cholesterol in large VLDL	L-VLDL-C	mmol/l
Cholesterol esters in large VLDL	L-VLDL-CE	mmol/l
Free cholesterol in large VLDL	L-VLDL-FC	mmol/l
Triglycerides in large VLDL	L-VLDL-TG	mmol/l
Concentration of medium VLDL particles	M-VLDL-P	mol/l
Total lipids in medium VLDL	M-VLDL-L	mmol/l
Phospholipids in medium VLDL	M-VLDL-PL	mmol/l
Total cholesterol in medium VLDL	M-VLDL-C	mmol/l
Cholesterol esters in medium VLDL	M-VLDL-CE	mmol/l
Free cholesterol in medium VLDL	M-VLDL-FC	mmol/l
Triglycerides in medium VLDL	M-VLDL-TG	mmol/l
Concentration of small VLDL particles	S-VLDL-P	mol/l
Total lipids in small VLDL	S-VLDL-L	mmol/l
Phospholipids in small VLDL	S-VLDL-PL	mmol/l
Total cholesterol in small VLDL	S-VLDL-C	mmol/l
Cholesterol esters in small VLDL	S-VLDL-CE	mmol/l
Free cholesterol in small VLDL	S-VLDL-FC	mmol/l
Triglycerides in small VLDL	S-VLDL-TG	mmol/l
Concentration of very small VLDL particles	XS-VLDL-P	mol/l
Total lipids in very small VLDL	XS-VLDL-L	mmol/l
Phospholipids in very small VLDL	XS-VLDL-PL	mmol/l

Name	Abbr.	Unit
Total cholesterol in very small VLDL	XS-VLDL-C	mmol/l
Cholesterol esters in very small VLDL	XS-VLDL-CE	mmol/l
Free cholesterol in very small VLDL	XS-VLDL-FC	mmol/l
Triglycerides in very small VLDL	XS-VLDL-TG	mmol/l
Concentration of IDL particles	IDL-P	mol/l
Total lipids in IDL	IDL-L	mmol/l
Phospholipids in IDL	IDL-PL	mmol/l
Total cholesterol in IDL	IDL-C	mmol/l
Cholesterol esters in IDL	IDL-CE	mmol/l
Free cholesterol in IDL	IDL-FC	mmol/l
Triglycerides in IDL	IDL-TG	mmol/l
Concentration of large LDL particles	L-LDL-P	mol/l
Total lipids in large LDL	L-LDL-L	mmol/l
Phospholipids in large LDL	L-LDL-PL	mmol/l
Total cholesterol in large LDL	L-LDL-C	mmol/l
Cholesterol esters in large LDL	L-LDL-CE	mmol/l
Free cholesterol in large LDL	L-LDL-FC	mmol/l
Triglycerides in large LDL	L-LDL-TG	mmol/l
Concentration of medium LDL particles	M-LDL-P	mol/l
Total lipids in medium LDL	M-LDL-L	mmol/l
Phospholipids in medium LDL	M-LDL-PL	mmol/l
Total cholesterol in medium LDL	M-LDL-C	mmol/l
Cholesterol esters in medium LDL	M-LDL-CE	mmol/l
Free cholesterol in medium LDL	M-LDL-FC	mmol/l
Triglycerides in medium LDL	M-LDL-TG	mmol/l
Concentration of small LDL particles	S-LDL-P	mol/l
Total lipids in small LDL	S-LDL-L	mmol/l
Phospholipids in small LDL	S-LDL-PL	mmol/l
Total cholesterol in small LDL	S-LDL-C	mmol/l
Cholesterol esters in small LDL	S-LDL-CE	mmol/l
Free cholesterol in small LDL	S-LDL-FC	mmol/l
Triglycerides in small LDL	S-LDL-TG	mmol/l
Concentration of very large HDL particles	XL-HDL-P	mol/l
Total lipids in very large HDL	XL-HDL-L	mmol/l
Phospholipids in very large HDL	XL-HDL-PL	mmol/l
Total cholesterol in very large HDL	XL-HDL-C	mmol/l
Cholesterol esters in very large HDL	XL-HDL-CE	mmol/l
Free cholesterol in very large HDL	XL-HDL-FC	mmol/l
Triglycerides in very large HDL	XL-HDL-TG	mmol/l
Concentration of large HDL particles	L-HDL-P	mol/l
Total lipids in large HDL	L-HDL-L	mmol/l
Phospholipids in large HDL	L-HDL-PL	mmol/l
Total cholesterol in large HDL	L-HDL-C	mmol/l
Cholesterol esters in large HDL	L-HDL-CE	mmol/l
Free cholesterol in large HDL	L-HDL-FC	mmol/l

Name	Abbr.	Unit
Triglycerides in large HDL	L-HDL-TG	mmol/l
Concentration of medium HDL particles	M-HDL-P	mol/l
Total lipids in medium HDL	M-HDL-L	mmol/l
Phospholipids in medium HDL	M-HDL-PL	mmol/l
Total cholesterol in medium HDL	M-HDL-C	mmol/l
Cholesterol esters in medium HDL	M-HDL-CE	mmol/l
Free cholesterol in medium HDL	M-HDL-FC	mmol/l
Triglycerides in medium HDL	M-HDL-TG	mmol/l
Concentration of small HDL particles	S-HDL-P	mol/l
Total lipids in small HDL	S-HDL-L	mmol/l
Phospholipids in small HDL	S-HDL-PL	mmol/l
Total cholesterol in small HDL	S-HDL-C	mmol/l
Cholesterol esters in small HDL	S-HDL-CE	mmol/l
Free cholesterol in small HDL	S-HDL-FC	mmol/l
Triglycerides in small HDL	S-HDL-TG	mmol/l
Phospholipids to total lipds ratio in chylomicrons and extremely large VLDL	XXL-VLDL-PL_%	%
Total cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-C_%	%
Cholesterol esters to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-CE_%	%
Free cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-FC_%	%
Triglycerides to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-TG_%	%
Phospholipids to total lipds ratio in very large VLDL	XL-VLDL-PL_%	%
Total cholesterol to total lipids ratio in very large VLDL	XL-VLDL-C_%	%
Cholesterol esters to total lipids ratio in very large VLDL	XL-VLDL-CE_%	%
Free cholesterol to total lipids ratio in very large VLDL	XL-VLDL-FC_%	%
Triglycerides to total lipids ratio in very large VLDL	XL-VLDL-TG_%	%
Phospholipids to total lipds ratio in large VLDL	L-VLDL-PL_%	%
Total cholesterol to total lipids ratio in large VLDL	L-VLDL-C_%	%
Cholesterol esters to total lipids ratio in large VLDL	L-VLDL-CE_%	%
Free cholesterol to total lipids ratio in large VLDL	L-VLDL-FC_%	%
Triglycerides to total lipids ratio in large VLDL	L-VLDL-TG_%	%
Phospholipids to total lipds ratio in medium VLDL	M-VLDL-PL_%	%
Total cholesterol to total lipids ratio in medium VLDL	M-VLDL-C_%	%
Cholesterol esters to total lipids ratio in medium VLDL	M-VLDL-CE_%	%
Free cholesterol to total lipids ratio in medium VLDL	M-VLDL-FC_%	%
Triglycerides to total lipids ratio in medium VLDL	M-VLDL-TG_%	%
Phospholipids to total lipds ratio in small VLDL	S-VLDL-PL_%	%
Total cholesterol to total lipids ratio in small VLDL	S-VLDL-C_%	%
Cholesterol esters to total lipids ratio in small VLDL	S-VLDL-CE_%	%

Name	Abbr.	Unit
Free cholesterol to total lipids ratio in small VLDL	S-VLDL-FC_%	%
Triglycerides to total lipids ratio in small VLDL	S-VLDL-TG_%	%
Phospholipids to total lipds ratio in very small VLDL	XS-VLDL-PL_%	%
Total cholesterol to total lipids ratio in very small VLDL	XS-VLDL-C_%	%
Cholesterol esters to total lipids ratio in very small VLDL	XS-VLDL-CE_%	%
Free cholesterol to total lipids ratio in very small VLDL	XS-VLDL-FC_%	%
Triglycerides to total lipids ratio in very small VLDL	XS-VLDL-TG_%	%
Phospholipids to total lipds ratio in IDL	IDL-PL_%	%
Total cholesterol to total lipids ratio in IDL	IDL-C_%	%
Cholesterol esters to total lipids ratio in IDL	IDL-CE_%	%
Free cholesterol to total lipids ratio in IDL	IDL-FC_%	%
Triglycerides to total lipids ratio in IDL	IDL-TG_%	%
Phospholipids to total lipds ratio in large LDL	L-LDL-PL_%	%
Total cholesterol to total lipids ratio in large LDL	L-LDL-C_%	%
Cholesterol esters to total lipids ratio in large LDL	L-LDL-CE_%	%
Free cholesterol to total lipids ratio in large LDL	L-LDL-FC_%	%
Triglycerides to total lipids ratio in large LDL	L-LDL-TG_%	%
Phospholipids to total lipds ratio in medium LDL	M-LDL-PL_%	%
Total cholesterol to total lipids ratio in medium LDL	M-LDL-C_%	%
Cholesterol esters to total lipids ratio in medium LDL	M-LDL-CE_%	%
Free cholesterol to total lipids ratio in medium LDL	M-LDL-FC_%	%
Triglycerides to total lipids ratio in medium LDL	M-LDL-TG_%	%
Phospholipids to total lipds ratio in small LDL	S-LDL-PL_%	%
Total cholesterol to total lipids ratio in small LDL	S-LDL-C_%	%
Cholesterol esters to total lipids ratio in small LDL	S-LDL-CE_%	%
Free cholesterol to total lipids ratio in small LDL	S-LDL-FC_%	%
Triglycerides to total lipids ratio in small LDL	S-LDL-TG_%	%
Phospholipids to total lipds ratio in very large HDL	XL-HDL-PL_%	%
Total cholesterol to total lipids ratio in very large HDL	XL-HDL-C_%	%
Cholesterol esters to total lipids ratio in very large HDL	XL-HDL-CE_%	%
Free cholesterol to total lipids ratio in very large HDL	XL-HDL-FC_%	%
Triglycerides to total lipids ratio in very large HDL	XL-HDL-TG_%	%
Phospholipids to total lipds ratio in large HDL	L-HDL-PL_%	%
Total cholesterol to total lipids ratio in large HDL	L-HDL-C_%	%
Cholesterol esters to total lipids ratio in large HDL	L-HDL-CE_%	%
Free cholesterol to total lipids ratio in large HDL	L-HDL-FC_%	%
Triglycerides to total lipids ratio in large HDL	L-HDL-TG_%	%
Phospholipids to total lipds ratio in medium HDL	M-HDL-PL_%	%
Total cholesterol to total lipids ratio in medium HDL	M-HDL-C_%	%
Cholesterol esters to total lipids ratio in medium HDL	M-HDL-CE_%	%
Free cholesterol to total lipids ratio in medium HDL	M-HDL-FC_%	%
Triglycerides to total lipids ratio in medium HDL	M-HDL-TG_%	%

Name	Abbr.	Unit
Phospholipids to total lipds ratio in small HDL	S-HDL-PL_%	%
Total cholesterol to total lipids ratio in small HDL	S-HDL-C_%	%
Cholesterol esters to total lipids ratio in small HDL	S-HDL-CE_%	%
Free cholesterol to total lipids ratio in small HDL	S-HDL-FC_%	%
Triglycerides to total lipids ratio in small HDL	S-HDL-TG_%	%
Mean diameter for VLDL particles	VLDL-D	nm
Mean diameter for LDL particles	LDL-D	nm
Mean diameter for HDL particles	HDL-D	nm
Serum total cholesterol	Serum-C	mmol/l
Total cholesterol in VLDL	VLDL-C	mmol/l
Remnant cholesterol (non-HDL, non-LDL -cholesterol)	Remnant-C	mmol/l
Total cholesterol in LDL	LDL-C	mmol/l
Total cholesterol in HDL	HDL-C	mmol/l
Total cholesterol in HDL2	HDL2-C	mmol/l
Total cholesterol in HDL3	HDL3-C	mmol/l
Esterified cholesterol	EstC	mmol/l
Free cholesterol	FreeC	mmol/l
Serum total triglycerides	Serum-TG	mmol/l
Triglycerides in VLDL	VLDL-TG	mmol/l
Triglycerides in LDL	LDL-TG	mmol/l
Triglycerides in HDL	HDL-TG	mmol/l
Total phosphoglycerides	TotPG	mmol/l
Ratio of triglycerides to phosphoglycerides	TG/PG	
Phosphatidylcholine and other cholines	PC	mmol/l
Sphingomyelins	SM	mmol/l
Total cholines	TotCho	mmol/l
Apolipoprotein A-I	ApoA1	g/l
Apolipoprotein B	ApoB	g/l
Ratio of apolipoprotein B to apolipoprotein A-I	ApoB/ApoA1	
Total fatty acids	TotFA	mmol/l
Estimated degree of unsaturation	UnSat	
22:6, docosahexaenoic acid	DHA	mmol/l
18:2, linoleic acid	LA	mmol/l
Omega-3 fatty acids	FAw3	mmol/l
Omega-6 fatty acids	FAw6	mmol/l
Polyunsaturated fatty acids	PUFA	mmol/l
Monounsaturated fatty acids 16:1, 18:1	MUFA	mmol/l
Saturated fatty acids	SFA	mmol/l
Ratio of 22:6 docosahexaenoic acid to total fatty acids	DHA/FA	%
Ratio of 18:2 linoleic acid to total fatty acids	LA/FA	%
Ratio of omega-3 fatty acids to total fatty acids	FAw3/FA	%
Ratio of omega-6 fatty acids to total fatty acids	FAw6/FA	%
Ratio of polyunsaturated fatty acids to total fatty acids	PUFA/FA	%
Ratio of monounsaturated fatty acids to total fatty acids	MUFA/FA	%
Ratio of saturated fatty acids to total fatty acids	SFA/FA	%

Name	Abbr.	Unit
Glucose	Glc	mmol/l
Lactate	Lac	mmol/l
Pyruvate	Pyr	mmol/l
Citrate	Cit	mmol/l
Glycerol	Glol	mmol/l
Alanine	Ala	mmol/l
Glutamine	Gln	mmol/l
Glycine	Gly	mmol/l
Histidine	His	mmol/l
Isoleucine	Ile	mmol/l
Leucine	Leu	mmol/l
Valine	Val	mmol/l
Phenylalanine	Phe	mmol/l
Tyrosine	Tyr	mmol/l
Acetate	Ace	mmol/l
Acetoacetate	AcAce	mmol/l
3-hydroxybutyrate	bOHBut	mmol/l
Creatinine	Crea	mmol/l
Albumin	Alb	signal area
Glycoprotein acetyls, mainly a1-acid glycoprotein	Gp	mmol/l

Table II: Probes of the Illumina Human HT-12 v3 Expression BeadChip used for expression analysis of lipid-associated CpG sites obtained by the EWAS of main blood lipids in KORA F4

Probe ID	Array Address ID	Annotated gene	Chr*	Probe Chr Orientation
ILMN_1658176	460593	*ABCG1*	21	+
ILMN_2262362	3440292	*ABCG1*	21	+
ILMN_1743638	4290020	*ABCG1*	21	+
ILMN_2329927	5860377	*ABCG1*	21	+
ILMN_1794782	6060377	*ABCG1*	21	+
ILMN_1695968	6450059	*ABCG1*	21	+
ILMN_1687589	2680458	*CPT1A*	11	-
ILMN_1710052	4150091	*CPT1A*	11	-
ILMN_1696316	5290358	*CPT1A*	11	-
ILMN_1663035	3390343	*SREBF1*	17	-
ILMN_1695378	4230521	*SREBF1*	17	-
ILMN_2328986	6840044	*SREBF1*	17	-
ILMN_1697448	1240440	*TXNIP*	1	+
ILMN_1703650	2680100	*TNIP1*	5	-

*Chr = chromosome

Table III: Sequences of the EMSA oligonucleotides and of the primers used for the functional experiments

Primer code	5´→ 3´ sequence	Product length (bp)	purpose
For_cg06500161	Cy5 - ccttctctagacacccggcgggactagttcct	32	EMSA
Rev_cg06500161	aggaactagtcccgccgggtgtctagagaagg	32	EMSA
For_cg06500161	ccttctctagacacccggcgggactagttcct	32	EMSA
For_(me)cg06500161	Cy5 - ccttctctagacacc-(me)c-ggcgggactagttcct	32	EMSA
Rev_(me)cg06500161	aggaactagtcccgc-(me)c-gggtgtctagagaagg	32	EMSA
For_(me)cg06500161	ccttctctagacacc-(me)c-ggcgggactagttcct	32	EMSA
For_cg20544516	Cy5 - tccgggctgcactgccgaggcactgcacccgc	32	EMSA
Rev_cg20544516	gcgggtgcagtgcctcggcagtgcagcccgga	32	EMSA
For_cg20544516	tccgggctgcactgccgaggcactgcacccgc	32	EMSA
For_(me)cg20544516	Cy5 - tccgggctgcactgc-(me)c-gaggcactgcacccgc	32	EMSA
Rev_(me)cg20544516	gcgggtgcagtgcct-(me)c-ggcagtgcagcccgga	32	EMSA
For_(me)cg20544516	tccgggctgcactgc-(me)c-gaggcactgcacccgc	32	EMSA
Oct1(O'Neill *et al.* 1988)	tgtcgaatgcaaatcactagaa	22	EMSA
RVprimer3_for (Promega)	tagcaaaataggctgtccc		Sequencing pGL4.53
*ABCG1*_cg06500161_for	cctcaggagaccaggagaaca	1803	Insert Luciferase Assay
*ABCG1*_cg06500161_rev	tgctggccaggctaagctctt	1803	Insert Luciferase Assay
*ABCG1*_104 bp_for (NsiI)	ttttatgcattttgctgtctctggagacca	124	Insert Luciferase Assay
*ABCG1*_104 bp _rev (BamHI)	ttttggatcctctgaaagacaatgtttaacatttc	124	Insert Luciferase Assay
*ABCG1*_104 bp _invers_for (NsiI)	ttttatgcattctgaaagacaatgtttaacatttcaca	124	Insert Luciferase Assay
*ABCG1*_104 bp _invers_rev (BamHI)	ttttggatcctttgctgtctctggagacc	124	Insert Luciferase Assay
*ABCG1*_518 bp _for (NsiI)	ttttatgcatagtcacttctaccctaggca	538	Insert Luciferase Assay
*ABCG1*_518 bp _rev (BamHI)	ttttggatcctttcatccaccaccatctct	538	Insert Luciferase Assay

Primer code	5´→ 3´ sequence	Product length (bp)	purpose
*ABCG1*_518 bp _invers_for (NsiI)	ttttatgcattttcatccaccaccatctctgg	538	Insert Luciferase Assay
*ABCG1*_518 bp _invers_rev (BamHI)	ttttggatccagtcacttctaccctaggc	538	Insert Luciferase Assay
*ABCG1*_1607 bp _for (NsiI)	ttttatgcatcattggttccgtctagaaaggtg	1627	Insert Luciferase Assay
*ABCG1*_1607 bp _rev (BamHI)	ttttggatccaagagagaaggtgacaaaatgatg	1627	Insert Luciferase Assay
*ABCG1*_1607 bp _invers_for (NsiI)	ttttatgcataagagagaaggtgacaaaatgatgc	1627	Insert Luciferase Assay
*ABCG1*_1607 bp _invers_rev (BamHI)	ttttggatcccattggttccgtctagaaag	1627	Insert Luciferase Assay
pCpGfree_*ABCG1*_all inserts_for	gttcagctctctgtttctat		Sequencing Insert Luciferase Assay
pCpGfree_*ABCG1*_all inserts _rev	gcacttatatatggttctcc		Sequencing Insert Luciferase Assay
pCpGfree_*ABCG1*_ ABCG1_1607 bp _for	ttagaggcaaggacagatga		Sequencing Insert Luciferase Assay
pCpGfree_ *ABCG1*_1607 bp _rev	tatgcccaaaggaactagtc		Sequencing Insert Luciferase Assay
pCpGfree_*ABCG1*_ ABCG1_1607 bp_2_for	ttggaattgcccttgggaat		Sequencing Insert Luciferase Assay
pCpGfree_ *ABCG1*_1607 bp _2_rev	atgtacccttccttcttgtt		Sequencing Insert Luciferase Assay
pCpGfree_ABCG1_all inserts_BC_1	accttctctaaacaccaattcaatta		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_ABCG1_all inserts_BC_2	aactactacacttatatataattct		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_ABCG1_ 518 bp_BC	cacaatcacaaaatcaaataaacaaaca		
pCpGfree_ABCG1_ 1607 bp_BC_1	ataaaacttaataaaaaaactcctccc		Sequencing Insert Luciferase Assay after bisulfite conversion

Primer code	5´→ 3´ sequence	Product length (bp)	purpose
pCpGfree_*ABCG1*_ 1607 bp_BC_2	tccactaaaataaccccaaacctt		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_*ABCG1*_ 1607 bp_BC_3	ccttaaccaaaacactctaaaatttc		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_*ABCG1*_ 518 bp_invers_BC	acaaacatcatttccatcaactatac		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_*ABCG1*_ 1607 bp_invers_BC_1	tccatccccataattcaaaaaacc		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_*ABCG1*_ 1607 bp_ invers_BC_2	acaaacatcatttccatcaactatac		Sequencing Insert Luciferase Assay after bisulfite conversion
pCpGfree_*ABCG1*_ 1607 bp_ invers_BC_3	tcactaacataaaaatatactctacaca		Sequencing Insert Luciferase Assay after bisulfite conversion

Table IV: Sequences of the DNA fragments which were cloned into the pCpGfree-promoter-Lucia plasmid for Luciferase Reporter Assays. CpG sites of interest (cg06500161 in *ABCG1* and cg20544516 in *MIR33B/SREBF1)* are highlighted

Insert name	Sequence
*ABCG1*_104 bp	ttt gct gtc tct gga gac cag ggc ctt ctc tag aca cc**C** ggc ggg act agt tcc ttt ggg cat aat tta ggt gtt ttg tga aat gtt aaa cat tgt ctt tca ga
*ABCG1*_104 bp_rc	tct gaa aga caa tgt tta aca ttt cac aaa aca cct aaa tta tgc cca aag gaa cta gtc ccg c**C**g ggt gtc tag aga agg ccc tgg tct cca gag aca gca aa
*ABCG1*_518 bp	agtcacttc tac cct agg cag gtt aga ggc aag gac aga tga cag ctg aac ttt acg gag tgt ttt cca ggg cat cta tta cag ttt tgt gtg tgt gaa ccc aaa agt gtg ata aca aac atc att tcc atc agc tgt act gtg ttt gag tat ccg caa ggc ctg ggg cca cct cag tgg agt cgg agg tgc cac gct ttg ctg tct ctg gag acc agg gcc ttc tct aga cac c**C**g gcg gga cta gtt cct ttg ggc ata att tag gtg ttt tgt gaa atg tta aac att gtc ttt cag acg ctt tgc tca acc ttt tgg gct cct gac caa atg caa aga gac cca agg gct tgc ttt tcc ttg cac tct tgg gtt ttc att gcc tgc ttg ctt att tga ttt tgt gac tgt gga gct ggc att tct gta agc cca tat aga aat tac agt ttc caa gcc taa acc cac taa gcc aac agg ggc aca cag gaa acc aga gat ggt ggt gga tga aa
*ABCG1*_518 bp_rc	ttt cat cca cca cca tct ctg gtt tcc tgt gtg ccc ctg ttg gct tag tgg gtt tag gct tgg aaa ctg taa ttt cta tat ggg ctt aca gaa atg cca gct cca cag tca caa aat caa ata agc aag cag gca atg aaa acc caa gag tgc aag gaa aag caa gcc ctt ggg tct ctt tgc att tgg tca gga gcc caa aag gtt gag caa agc gtc tga aag aca atg ttt aac att tca caa aac acc taa att atg ccc aaa gga act agt ccc gc**C** ggg tgt cta gag aag gcc ctg tct cc aga gac agc aaa gcg tgg cac ctc cga ctc cac tga ggt ggc ccc agg cct tgc gga tac tca aac aca gta cag ctg atg gaa atg atg ttt gtt atc aca ctt ttg ggt tca cac aca caa aac tgt aat aga tgc cct gga aaa cac tcc gta aag ttc agc tgt cat ctg tcc ttg cct cta acc tgc cta ggg tag aag tga ct
*ABCG1*_1607 bp	cat tgg ttc cgt cta gaa agg tgg gac agc cca aag ggg gac ttc cag gtt ata ggt aaa ttt aaa ttt ttt tct aat tgg tta agt tat tat cga tag aaa ggg atg tct ggg tta cga tac ggg gct gtg ggg acc tag gtt gta tca tgc gga tga ggc ctc cag gca gca ggc ttg aga gag aac aga ctg taa atg ctt cgt atc aga ctt aag gtc tgt gtt gat gtt aat gtt aca ggc tac gat aag gca tgt cag acg gac cgc cac ttc ccg tca tgg cct gaa cca gtc ttt cag ggg aaa ttt tag agt gcc ctg gcc aag gag gga gtc cat ccc cat ggt tca ggg ggc ctt tga att tta ttt ttg att tac agg tct tta ctt cag gtg gca caa ttt tcc tag gag aca gtt gat ttt tga ttc ctt ata aaa cgt gct ggt ttt gtt ttt agc tga cag tgt ggg gtg aag ctg taa ggg agc cag ggc ctc ctg ggg ttg atc caa gca tgt ctg agc ctt ctc tgg ggt cag cac tga aaa gtg gcc agg gtg ggg ttt gga acc ttg gaa ttg ccc ttg gga atc tgc agc gtt ggc tcc gct aag ccg gct gga atc cgg aga tca tgc ggc aag gtc acg ctt cc**C** ggc tgc ctg ccg agt cac ttc tac cct agg cag gtt aga ggc aag gac aga tga cag ctg aac ttt acg gag tgt ttt cca ggg cat cta tta cag ttt tgt gtg tgt gaa ccc aaa agt gtg ata aca aac atc att tcc atc agc tgt act gtg ttt gag tat ccg caa ggc ctg ggg cca cct cag tgg agt cgg agg tgc cac gct ttg ctg tct ctg gag acc agg gcc ttc tct aga cac ccg gcg gga cta gtt cct ttg ggc ata att tag gtg ttt tgt gaa atg tta aac att gtc ttt cag

acg ctt tgc tca acc ttt tgg gct cct gac caa atg caa aga gac cca agg gct tgc ttt tcc ttg cac tct tgg gtt ttc att gcc tgc ttg ctt att tga ttt tgt gac tgt gga gct ggc att tct gta agc cca tat aga aat tac agt ttc caa gcc taa acc cac taa gcc aac agg ggc aca cag gaa acc aga gat ggt ggt gga tga aaa cct ttg caa caa gaa gga agg gta cat cta tga taa cta tgg tgt tac att act aaa aac agt aag gat gag tgt cac taa cat gag aat gta ctc tac aca aaa gtc tgg tga aaa agg gag gag ccc ttt cac taa gtt tca tat tta tca gca ctc acg tta aga att aca gca gaa gct gca ggt ttc ctg ttt ctg aag acc tgc gaa tga gtt tgt cca gaa aat ggc cgt gtc ctg aat ggt gtg agc agg agg cca tcc agc ctt gtg cct cta ggc ctc ctc ccc ttc agt tgt cat aaa tgt gcc atg tgc ccc tgc cca aac cca gca gcc ttt ctc tga att tca tct ggt tta gag agg aaa atg ggc agc atc att ttg tca cct tct ctc tt

*ABCG1*_1607 bp_rc

aag aga gaa ggt gac aaa atg atg ctg ccc att ttc ctc tct aaa cca gat gaa att cag aga aag gct gct ggg ttt ggg cag ggg cac atg gca cat tta tga caa ctg aag ggg agg agg cct aga ggc aca agg ctg gat ggc ctc ctg ctc aca cca ttc agg aca cgg cca ttt tct gga caa act cat tcg cag gtc ttc aga aac agg aaa cct gca gct tct gct gta att ctt aac gtg agt gct gat aaa tat gaa act tag tga aag ggc tcc tcc ctt ttt cac cag act ttt gtg tag agt aca ttc tca tgt tag tga cac tca tcc tta ctg ttt tta gta atg taa cac cat agt tat cat aga tgt acc ctt cct tct tgt tgc aaa ggt ttt cat cca cca cca tct ctg gtt tcc tgt gtg ccc ctg ttg gct tag tgg gtt tag gct tgg aaa ctg taa ttt cta tat ggg ctt aca gaa atg cca gct cca cag tca caa aat caa ata agc aag cag gca atg aaa acc caa gag tgc aag gaa aag caa gcc ctt ggg tct ctt tgc att tgg tca gga gcc caa aag gtt gag caa agc gtc tga aag

aca atg ttt aac att tca caa aac acc taa att atg ccc aaa gga act agt ccc gc**C** ggg tgt cta gag aag gcc ctg gtc tcc aga gac agc aaa gcg tgg cac ctc cga ctc cac tga ggt ggc ccc agg cct tgc gga tac tca aac aca gta cag ctg atg gaa atg atg ttt gtt atc aca ctt ttg ggt tca cac aca caa aac tgt aat aga tgc cct gga aaa cac tcc gta aag ttc agc tgt cat ctg tcc ttg cct cta acc tgc cta ggg tag aag tga ctc ggc agg cag ccg gga agc gtg acc ttg ccg cat gat ctc cgg att cca gcc ggc tta gcg gag cca acg ctg cag att ccc aag ggc aat tcc aag gtt cca aac ccc acc ctg gcc act ttt cag tgc tga ccc cag aga agg ctc aga cat gct tgg atc aac ccc agg agg ccc tgg ctc cct tac agc ttc acc cca cac tgt cag cta aaa aca aaa cca gca cgt ttt ata agg aat caa aaa tca act gtc tcc tag gaa aat tgt gcc acc tga agt aaa gac ctg taa atc aaa aat aaa att caa agg ccc cct gaa cca tgg gga tgg act ccc tcc ttg gcc agg gca ctc taa aat ttc ccc tga aag act ggt tca ggc cat gac ggg aag tgg cgg tcc gtc tga cat gcc tta tcg tag cct gta aca tta aca tca aca cag acc tta agt ctg ata cga agc att tac agt ctg ttc tct ctc aag cct gct gcc tgg agg cct cat ccg cat gat aca acc tag gtc ccc aca gcc ccg tat cgt aac cca gac atc cct ttc tat cga taa taa ctt aac caa tta gaa aaa aat tta aat tta cct ata acc tgg aag tcc ccc ttt ggg ctg tcc cac ctt tct aga cgg aac caa tg

*MIR33B/SREBF1*_24 bp

ggg ctg cac tgc **C**ga ggc act gca

*MIR33B/SREBF1*_24 bp_r

acg tca cgg agc cgt cac gtc ggg

*MIR33B/SREBF1*_32 bp

tcc ggg ctg cac tgc **C**ga ggc act gca ccc gc

*MIR33B/SREBF1*_32 bp_r

cgc cca cgt cac gga gcc gtc acg tcg ggc ct

Table V: EWAS of main blood lipids: Results of the gene expression analyses of lipid-associated CpG sites. Results of the expression analyses **a)** The association between lipid level and *ABCG1* DNA methylation was recalculated in N=724 KORA F4 participants. **b)** Investigation of the association between *ABCG1* DNA methylation and *ABCG1* transcripts. **c)** Analysis of the association between *ABCG1* transcripts and lipids. Results of models with adjustment are shown on the right side of the table. Level of significance: 8.3E-04

a) Associations between lipid levels and *ABCG1* DNA methylation with and without adjusting for *ABCG1* expression (N=724)

Without adjustment					*adjusting for ABCG1 transcripts*			
Lipid	CpG	β coef*	SE†	p-value	*ABCG1* transcript	β coef*	SE†	p-value
HDL	cg06500161	-0.050	0.01	4.10E-07	all 6 transcripts	-0.025	0.01	1.30E-02
					ILMN_2329927	-0.029	0.01	4.10E-03
					ILMN_1794782	-0.043	0.01	2.31E-05
					ILMN_2262362	-0.047	0.01	2.94E-06
					ILMN_1658176	-0.047	0.01	2.83E-06
					ILMN_1695968	-0.046	0.01	3.45E-06
					ILMN_1743638	-0.047	0.01	2.62E-06
TG	cg06500161	0.075	0.01	5.56E-13	all 6 transcripts	0.044	0.01	1.36E-05
					ILMN_2329927	0.049	0.01	1.34E-06
					ILMN_1794782	0.073	0.01	4.69E-12
					ILMN_2262362	0.078	0.01	1.81E-13
					ILMN_1658176	0.078	0.01	1.56E-13
					ILMN_1695968	0.077	0.01	2.63E-13
					ILMN_1743638	0.078	0.01	1.82E-13
TG	cg27243685	0.075	0.01	2.36E-07	all 6 transcripts	0.040	0.01	3.24E-03
					ILMN_2329927	0.049	0.01	3.30E-04
					ILMN_1794782	0.067	0.01	7.23E-06
					ILMN_2262362	0.074	0.01	6.21E-07
					ILMN_1658176	0.074	0.01	5.73E-07
					ILMN_1695968	0.073	0.01	8.04E-07
					ILMN_1743638	0.074	0.01	5.46E-07

b) Associations between *ABCG1* DNA methylation and *ABCG1* transcripts with and without adjusting for blood lipids

Without adjustment

adjusting for blood lipids

CpG	*ABCG1* transcript	β coef*	SE†	p-value	CpG	Lipid	*ABCG1* transcript	β coef*	SE†	p-value
cg06500161	all transcripts			5.42E-14	cg06500161	HDL	all transcripts			5.15E-11
	ILMN_2329927	-0.151	0.02	5.22E-15			ILMN_2329927	-0.14	0.02	3.42E-12
	ILMN_1794782	-0.135	0.04	9.98E-05			ILMN_1794782	-0.12	0.04	6.72E-04
	ILMN_2262362	0.016	0.05	7.41E-01			ILMN_2262362	0.02	0.05	7.37E-01
	ILMN_1658176	0.011	0.04	7.57E-01			ILMN_1658176	0.02	0.04	6.26E-01
	ILMN_1695968	0.037	0.04	3.25E-01			ILMN_1695968	0.03	0.04	3.82E-01
	ILMN_1743638	-0.013	0.05	8.06E-01			ILMN_1743638	-0.02	0.05	6.48E-01
						TG	all transcripts			1.23E-06
							ILMN_2329927	-0.11	0.02	7.53E-08
							ILMN_1794782	-0.10	0.04	3.48E-03
							ILMN_2262362	0.00	0.05	9.94E-01
							ILMN_1658176	0.02	0.04	6.02E-01
							ILMN_1695968	0.02	0.04	5.81E-01
							ILMN_1743638	-0.02	0.05	7.42E-01
cg27243685	all transcripts			1.86E-07	cg27243685	TG	all transcripts			8.11E-04
	ILMN_2329927	-0.185	0.04	4.34E-06			ILMN_2329927	-0.11	0.05	8.98E-03
	ILMN_1794782	-0.286	0.07	5.72E-05			ILMN_1794782	-0.23	0.07	1.00E-03
	ILMN_2262362	0.124	0.10	1.99E-01			ILMN_2262362	0.10	0.10	2.93E-01
	ILMN_1658176	-0.064	0.08	3.88E-01			ILMN_1658176	-0.05	0.08	4.64E-01
	ILMN_1695968	0.139	0.08	7.21E-02			ILMN_1695968	0.11	0.08	1.35E-01
	ILMN_1743638	0.118	0.11	2.80E-01			ILMN_1743638	0.11	0.11	2.94E-01

c) Associations between *ABCG1* transcripts and lipids with and without adjusting for *ABCG1* DNA methylation

Without adjustment   *adjusting for ABCG1 DNA methylation*

Lipid	*ABCG1* transcript	β coef*	SE†	p-value	Lipid	CpG	*ABCG1* transcript	β coef*	SE†	p-value
HDL-C	all transcripts			7.76E-13	HDL-C	cg06500161	all transcripts			1.79E-09
	ILMN_2329927	0.15	0.02	2.45E-13			ILMN_2329927	0.13	0.02	3.89E-10
	ILMN_1794782	0.13	0.04	2.22E-04			ILMN_1794782	0.11	0.04	2.10E-03
	ILMN_2262362	0.01	0.05	8.65E-01			ILMN_2262362	0.01	0.05	8.65E-01
	ILMN_1658176	0.05	0.04	2.07E-01			ILMN_1658176	0.05	0.04	1.99E-01
	ILMN_1695968	-0.04	0.04	2.89E-01			ILMN_1695968	-0.03	0.04	3.73E-01
	ILMN_1743638	-0.05	0.06	3.44E-01			ILMN_1743638	-0.06	0.06	2.85E-01
TG	all transcripts			1.25E-33	TG	cg06500161	all transcripts			1.20E-25
	ILMN_2329927	-0.49	0.04	1.55E-27			ILMN_2329927	-0.43	0.04	1.42E-20
	ILMN_1794782	-0.32	0.08	2.51E-05			ILMN_1794782	-0.25	0.08	9.82E-04
	ILMN_2262362	0.16	0.11	1.33E-01			ILMN_2262362	0.15	0.10	1.45E-01
	ILMN_1658176	-0.09	0.08	2.66E-01			ILMN_1658176	-0.10	0.08	2.01E-01
	ILMN_1695968	0.15	0.09	6.38E-02			ILMN_1695968	0.13	0.08	9.86E-02
	ILMN_1743638	0.04	0.12	7.09E-01			ILMN_1743638	0.05	0.12	6.50E-01
					TG	cg27243685	all transcripts			7.08E-30
							ILMN_2329927	-0.47	0.04	9.56E-25
							ILMN_1794782	-0.27	0.08	4.55E-04
							ILMN_2262362	0.14	0.11	1.72E-01
							ILMN_1658176	-0.08	0.08	3.14E-01
							ILMN_1695968	0.13	0.09	1.13E-01
							ILMN_1743638	0.02	0.12	8.92E-01

*β coef: β coefficient; †SE: standard error

Table VI: EWAS of main blood lipids: Investigation of genetic confounding of the associations between DNA methylation and main blood lipid levels in KORA F4. Associations between CpG sites and lipid levels were recalculated adjusting for known lipid-associated SNPs which were nominally associated with DNA methylation of the respective CpG site. An increase of the p-value indicates SNP confounding of the association.

Lipid	CpG	Gene	β coef*	p-value SNP analysis	Discovery p-value†
HDL-C	cg06500161	*ABCG1*	-0.048	5.61E-15	8.26E-17
TG	cg06500161	*ABCG1*	0.065	7.41E-24	1.21E-27
	cg19693031	*TXNIP*	-0.029	1.86E-16	1.89E-17
	cg11024682	*SREBF1*	0.054	1.83E-11	5.54E-14
	cg00574958	*CPT1A*	-0.112	6.35E-12	3.15E-13
	cg27243685	*ABCG1*	0.062	3.10E-12	3.24E-13
	cg07504977	NA‡	0.023	1.08E-09	3.93E-12
	cg20544516	*MIR33B/SREBF1*	0.039	1.23E-07	2.84E-09
	cg12556569	*APOA5*	-0.001	2.05E-01	6.43E-09
	cg07397296	*ABCG1*	0.024	3.74E-06	9.48E-08
	cg07815238	NA‡	0.049	1.95E-07	9.66E-08
LDL-C	cg22178392	*TNIP1*	0.042	1.42E-09	4.27E-09

Genetic confounding was examined for each significant CpG–lipid pair. *β coef: β coefficient; †p-value of CpG-lipid associations in KORA F4; ‡no gene annotation for this CpG according to the UCSC Genome Browser.

Table VII: EWAS of main blood lipids: Association of significant lipid-associated CpG sites with prevalent myocardial infarction in KORA F4 (N=1,776 with N=60 cases with myocardial infarction)

CpG	Chr*	gene	β coef†	SE‡	p-value	OR (95% CI)§
cg00574958	11	*CPT1A*	0.063	0.11	5.59E-01	1.06 (0.86/1.31)
cg06500161	21	*ABCG1*	0.141	0.04	1.30E-03[‖]	1.15 (1.06/1.25)
cg07397296	21	*ABCG1*	0.005	0.03	8.82E-01	1.00 (0.94/1.07)
cg07504977	10	NA[#]	0.015	0.02	5.30E-01	1.02 (0.97/1.06)
cg11024682	17	*SREBF1*	0.089	0.05	7.14E-02	1.09 (0.99/1.20)
cg20544516	17	*MIR33B/ SREBF1*	0.008	0.05	8.63E-01	1.01 (0.92/1.10)
cg22178392	5	*TNIP1*	0.022	0.04	5.91E-01	1.02 (0.94/1.11)
cg27243685	21	*ABCG1*	0.133	0.07	4.83E-02	1.14 (1.00/1.30)

Lipid-associated CpG sites were tested for an association with previous hospitalized myocardial infarction in KORA F4. *Chr: chromosome; †β coef: β coefficient; ‡SE: standard error; §OR: Odds Ratio; CI: confidence interval; ‖significant, level of significance: 6.3E-03; #no gene annotation for this CpG according to the UCSC Genome Browser

Table VIII: EWAS of NMR-measured metabolites: CpG-metabolite pairs which were confirmed by the meta-analysis of the replication results of the LOLIPOP, NFBC1966 and YFS cohort.
Table provides β coefficients and p-values of the EWAS in KORA F4 as well of the meta-analysis. Level of significance of the EWAS in KORA F4: 4.73E-10; level of significance for meta-analysis: 2.1E-04

CpG	Gene	Trait	EWAS β coef*	EWAS p-value	Explained variance (%)	meta-analysis β coef*	meta-analysis p-value
cg06500161	*ABCG1*	L-VLDL-L	6.279	1.10E-18	3.6	7.721	3.67E-20
cg06500161	*ABCG1*	L-VLDL-P	6.401	6.80E-18	3.6	7.774	3.92E-20
cg11024682	*SREBF1*	S-HDL-TG	2.004	6.60E-12	2.2	1.723	2.75E-19
cg06500161	*ABCG1*	S-VLDL-P	2.820	4.45E-19	3.8	2.162	3.41E-19
cg06500161	*ABCG1*	M-HDL-TG_%	2.163	2.73E-13	2.7	2.149	8.46E-19
cg01400685	*FADS2*	FAw3	1.325	4.52E-14	3.0	1.430	9.31E-19
cg07504977	NA†	S-HDL-TG	1.076	1.16E-10	1.8	1.127	1.65E-18
cg06500161	*ABCG1*	L-VLDL-TG	6.717	4.82E-20	4.2	8.721	3.03E-18
cg06500161	*ABCG1*	M-VLDL-FC	4.722	1.44E-20	4.2	3.998	3.26E-18
cg06500161	*ABCG1*	S-VLDL-PL	2.399	9.13E-18	3.6	1.979	3.39E-18
cg27243685	*ABCG1*	M-VLDL-TG	5.579	1.30E-10	2.1	5.020	5.10E-18
cg27243685	*ABCG1*	VLDL-TG	5.494	2.20E-12	2.4	4.299	1.18E-17
cg06500161	*ABCG1*	S-VLDL-L	2.646	5.32E-18	3.6	1.993	3.15E-17
cg06690548	*SLC7A11*	LAtoFA	0.508	4.29E-13	2.5	0.529	7.61E-17
cg27243685	*ABCG1*	S-VLDL-TG	4.812	1.52E-13	2.7	3.576	8.79E-17
cg06500161	*ABCG1*	L-VLDL-PL	6.083	1.91E-16	3.3	7.906	9.16E-17
cg19693031	*TXNIP*	VLDL-D	-0.144	2.23E-12	2.3	-0.108	1.10E-16
cg06500161	*ABCG1*	L-VLDL-FC	6.984	1.01E-13	2.8	9.351	1.15E-16
cg19693031	*TXNIP*	XS-VLDL-TG_%	-0.991	1.92E-13	2.6	-0.773	2.33E-16
cg19693031	*TXNIP*	S-VLDL-CE_%	0.737	1.14E-11	2.3	0.647	2.72E-16
cg27386326	NA†	FAw3toFA	-1.148	2.18E-12	2.9	-1.866	3.57E-16
cg11250194	*FADS2*	FAw3toFA	1.260	3.30E-28	6.8	1.566	6.02E-16
cg27243685	*ABCG1*	M-VLDL-P	4.990	6.73E-11	2.1	4.055	6.85E-16
cg06500161	*ABCG1*	M-HDL-TG	1.848	1.90E-12	2.8	1.796	7.70E-16
cg06690548	*SLC7A11*	FAw6toFA	0.398	8.90E-11	1.9	0.432	1.00E-15
cg06500161	*ABCG1*	L-VLDL-C	5.654	1.46E-14	3.2	8.307	1.09E-15
cg11024682	*SREBF1*	M-HDL-TG	1.656	3.16E-10	2.0	1.747	1.30E-15
cg01400685	*FADS2*	FAw3toFA	1.219	1.80E-17	3.9	1.440	1.54E-15
cg19693031	*TXNIP*	S-VLDL-C_%	0.482	1.81E-11	2.2	0.441	1.69E-15
cg27243685	*ABCG1*	M-VLDL-L	4.916	6.58E-11	2.1	3.934	2.51E-15
cg06500161	*ABCG1*	M-VLDL-C	3.298	1.20E-15	3.2	2.790	3.22E-15
cg19693031	*TXNIP*	L-VLDL-L	-3.306	7.67E-12	2.1	-3.840	6.43E-15
cg19693031	*TXNIP*	L-VLDL-P	-3.356	2.87E-11	2.1	-3.862	7.43E-15
cg06500161	*ABCG1*	M-VLDL-PL_%	-0.314	3.26E-16	3.7	-0.255	1.77E-14
cg27243685	*ABCG1*	M-VLDL-FC	5.493	2.14E-11	2.3	4.252	3.02E-14
cg06500161	*ABCG1*	S-VLDL-TG	4.064	7.34E-24	4.7	3.324	3.82E-14
cg19693031	*TXNIP*	MUFAtoFA	-0.502	8.65E-15	3.1	-0.320	4.35E-14
cg19693031	*TXNIP*	L-VLDL-FC	-4.011	3.16E-10	2.0	-4.965	6.25E-14
cg06500161	*ABCG1*	S-VLDL-FC	2.479	1.03E-16	3.4	1.824	6.77E-14

cg06500161	*ABCG1*	L-VLDL-CE	4.884	8.10E-13	2.9	7.664	7.43E-14
cg27243685	*ABCG1*	L-VLDL-TG	7.396	4.14E-10	1.8	10.179	1.09E-13
cg19693031	*TXNIP*	L-VLDL-TG	-3.407	7.72E-12	2.3	-4.314	2.91E-13
cg19693031	*TXNIP*	L-VLDL-PL	-3.298	5.15E-11	2.1	-4.038	3.23E-13
cg27243685	*ABCG1*	S-HDL-TG	3.488	1.15E-13	2.7	1.952	3.83E-13
cg19693031	*TXNIP*	S-LDL-TG_%	-1.221	8.93E-14	2.8	-0.768	8.15E-13
cg27243685	*ABCG1*	XS-VLDL-TG	3.529	5.88E-14	2.9	2.259	9.76E-13
cg19693031	*TXNIP*	IDL-TG_%	-0.887	6.40E-11	2.2	-0.724	1.45E-12
cg27243685	*ABCG1*	S-VLDL-P	3.614	1.38E-12	2.5	2.305	1.54E-12
cg24503796	*SCD*	DHA	1.248	1.55E-10	2.3	1.704	6.09E-12
cg19693031	*TXNIP*	L-VLDL-C	-3.115	4.19E-10	2.1	-4.134	1.08E-11
cg19693031	*TXNIP*	Serum TG	-1.849	1.16E-13	2.7	-1.465	1.31E-11
cg27243685	*ABCG1*	S-HDL-TG_%	3.246	2.18E-12	2.4	1.839	1.33E-11
cg27243685	*ABCG1*	S-VLDL-L	3.448	3.00E-12	2.4	2.136	2.04E-11
cg27243685	*ABCG1*	S-VLDL-PL	3.191	1.51E-12	2.5	2.030	4.17E-11
cg06500161	*ABCG1*	VLDL-TG	4.835	1.87E-23	4.6	3.795	1.23E-10
cg11250194	*FADS2*	FAw3	1.264	1.63E-19	4.6	1.527	2.83E-10
cg06500161	*ABCG1*	M-VLDL-TG	5.233	2.12E-22	4.4	4.337	6.52E-10
cg27243685	*ABCG1*	M-VLDL-PL	4.632	4.91E-11	2.2	3.508	6.92E-10
cg19693031	*TXNIP*	M-VLDL-FC	-2.333	1.36E-11	2.2	-1.926	7.83E-10
cg06500161	*ABCG1*	M-VLDL-CE	2.444	5.36E-11	2.2	2.258	8.39E-10
cg06500161	*ABCG1*	Serum TG	3.577	1.84E-22	4.5	2.736	1.19E-09
cg27243685	*ABCG1*	S-LDL-TG_%	2.912	7.18E-14	2.9	1.517	1.20E-09
cg27386326	NA†	DHAtoFA	-1.307	2.22E-10	2.4	-2.009	1.45E-09
cg19693031	*TXNIP*	PUFAtoFA	0.433	3.17E-17	3.4	0.292	1.56E-09
cg19693031	*TXNIP*	UnSat	0.178	1.66E-12	2.6	0.125	1.57E-09
cg19693031	*TXNIP*	VLDL-TG	-2.370	5.99E-13	2.2	-1.936	1.58E-09
cg19693031	*TXNIP*	XS-VLDL-TG	-1.297	5.31E-11	2.4	-1.062	1.58E-09
cg06500161	*ABCG1*	XS-VLDL-TG	2.598	4.44E-19	3.8	2.176	2.98E-09
cg06500161	*ABCG1*	MUFA	1.424	1.11E-12	2.7	1.109	4.21E-09
cg27243685	*ABCG1*	S-VLDL-FC	3.335	4.61E-12	2.4	1.893	8.78E-09
cg06500161	*ABCG1*	M-VLDL-L	4.346	1.14E-20	4.2	3.397	1.25E-08
cg19693031	*TXNIP*	Ile	-0.891	3.56E-10	1.2	-0.531	2.08E-08
cg19693031	*TXNIP*	M-VLDL-TG	-2.443	2.21E-11	2.1	-2.164	2.22E-08
cg27243685	*ABCG1*	Serum TG	4.404	1.06E-13	2.8	2.830	3.50E-08
cg19693031	*TXNIP*	S-HDL-TG	-1.375	3.47E-12	2.2	-0.792	4.59E-08
cg06500161	*ABCG1*	M-VLDL-P	4.456	4.88E-21	4.2	3.452	6.35E-08
cg19693031	*TXNIP*	TGtoPG	-1.326	5.27E-12	2.0	-0.637	9.61E-08
cg01400685	*FADS2*	DHAtoFA	1.481	2.07E-16	3.8	1.503	1.27E-07
cg06500161	*ABCG1*	VLDL-C	1.810	4.21E-11	2.3	1.302	1.53E-07
cg01400685	*FADS2*	DHA	1.588	1.93E-13	3.0	1.476	2.13E-07
cg27243685	*ABCG1*	S-LDL-TG	3.004	1.49E-10	2.3	1.776	2.72E-07
cg27243685	*ABCG1*	XS-VLDL-TG_%	2.494	6.76E-15	2.9	1.668	3.03E-07
cg20544516	*MIR33B*	VLDL-TG	4.354	7.72E-11	1.9	2.355	3.05E-07
cg27243685	*ABCG1*	S-VLDL-C_%	-1.091	1.65E-10	2.1	-0.935	3.36E-07
cg06500161	*ABCG1*	M-VLDL-PL	4.032	2.48E-20	4.1	3.091	3.80E-07

cg20544516	*MIR33B*	S-VLDL-TG	3.749	1.72E-11	2.0	1.955	5.57E-07
cg06500161	*ABCG1*	S-LDL-TG_%	1.887	4.87E-15	3.1	1.748	1.75E-06
cg27243685	*ABCG1*	S-VLDL-CE_%	-1.654	1.51E-10	2.1	-1.312	2.14E-06
cg11250194	*FADS2*	DHAtoFA	1.479	1.36E-24	6.0	1.675	2.75E-06
cg19693031	*TXNIP*	MUFA	-0.975	7.68E-13	2.7	-0.911	3.13E-06
cg20544516	*MIR33B*	S-VLDL-P	2.770	2.13E-10	1.8	1.376	3.72E-06
cg19693031	*TXNIP*	FAw6toFA	0.460	2.20E-17	3.4	0.284	4.04E-06
cg19693031	*TXNIP*	M-VLDL-L	-2.114	2.46E-11	2.1	-1.710	4.12E-06
cg19693031	*TXNIP*	M-VLDL-P	-2.153	2.16E-11	2.1	-1.757	4.25E-06
cg06500161	*ABCG1*	S-HDL-TG	2.634	1.43E-19	3.7	1.933	6.34E-06
cg20544516	*MIR33B*	S-VLDL-L	2.632	4.71E-10	1.8	1.309	6.63E-06
cg20544516	*MIR33B*	S-VLDL-PL	2.501	9.08E-11	1.9	1.261	6.77E-06
cg03440556	*SCD*	DHAtoFA	0.461	4.16E-10	2.7	0.343	8.43E-06
cg16246545	*PHGDH*	LAtoFA	0.351	5.83E-11	2.1	0.229	1.08E-05
cg03440556	*SCD*	DHA	0.611	3.33E-12	3.1	0.650	1.43E-05
cg16246545	*PHGDH*	FAw6toFA	0.296	3.62E-10	1.9	0.190	1.44E-05
cg19693031	*TXNIP*	M-VLDL-PL	-1.982	2.25E-11	2.1	-1.598	1.85E-05
cg27243685	*ABCG1*	MUFA	2.045	2.61E-10	2.3	1.082	1.92E-05
cg20544516	*MIR33B*	S-VLDL-FC	2.580	3.89E-10	1.8	1.276	1.99E-05
cg06500161	*ABCG1*	MUFAtoFA	0.724	2.98E-14	2.9	0.403	2.10E-05
cg06500161	*ABCG1*	S-VLDL-TG_%	1.418	2.87E-25	4.9	1.228	2.51E-05
cg06500161	*ABCG1*	S-HDL-TG_%	2.343	2.76E-16	3.1	1.743	3.65E-05
cg00574958	*CPT1A*	S-LDL-TG	-2.431	6.21E-12	2.4	-3.588	5.15E-05
cg11024682	*SREBF1*	VLDL-TG	3.156	8.42E-11	1.9	2.851	6.05E-05
cg03440556	*SCD*	FAw3	0.509	7.29E-13	3.1	0.606	6.26E-05
cg11250194	*FADS2*	DHA	1.483	9.13E-18	4.3	1.632	6.32E-05
cg19693031	*TXNIP*	LAtoFA	0.477	1.10E-14	3.1	0.309	6.38E-05
cg20544516	*MIR33B*	S-VLDL-CE_%	-1.517	6.50E-12	2.1	-0.638	7.47E-05
cg19693031	*TXNIP*	S-VLDL-TG	-1.860	1.15E-11	2.2	-1.599	1.51E-04
cg20544516	*MIR33B*	XS-VLDL-TG_%	1.948	1.15E-12	2.3	0.735	1.55E-04
cg06500161	*ABCG1*	VLDL-D	0.289	1.32E-21	4.4	0.194	1.61E-04
cg20544516	*MIR33B*	S-VLDL-C_%	-0.975	2.42E-11	2.0	-0.424	1.75E-04
cg19610905	*FADS2*	FAw3	2.950	2.46E-32	8.0	2.622	1.93E-04

*β coef**:** β coefficient; †NA: no gene annotation for this CpG according to the UCSC Genome Browser

Table IX: **EWAS of NMR-measured metabolites: Results of the investigation of associations between metabolite-related CpG sites and prevalent myocardial infarction in KORA F4 (N=1,662, with N=54 cases of myocardial infarction)**

CpG	gene	1st model*				2nd model*			
		β coef†	SE‡	p-value	OR (95% CI)§	β coef†	SE‡	p-value	OR (95% CI)§
cg06500161	*ABCG1*	0.619	0.16	8.66E-05**	1.86 (1.36-2.53)	0.312	0.17	7.03E-02	1.37 (0.97-1.92)
cg11024682	*SREBF1*	0.370	0.15	1.33E-02	1.45 (1.08-1.94)	0.385	0.17	2.67E-02	1.47 (1.05-2.06)
cg27243685	*ABCG1*	0.361	0.15	1.46E-02	1.43 (1.07-1.92)	0.106	0.16	5.16E-01	1.11 (0.81-1.53)
cg16246545	*PHGDH*	-0.163	0.12	1.90E-01	0.85 (0.67-1.08)	-0.160	0.14	2.55E-01	0.85 (0.65-1.12)
cg20544516	*MIR33B/ SREBF1*	0.180	0.15	2.22E-01	1.2 (0.9-1.6)	0.167	0.16	3.05E-01	1.18 (0.86-1.62)
cg27386326	NA#	0.186	0.16	2.33E-01	1.2 (0.89-1.64)	0.202	0.17	2.44E-01	1.22 (0.87-1.72)
cg19693031	*TXNIP*	0.172	0.15	2.37E-01	1.19 (0.89-1.58)	0.085	0.16	5.85E-01	1.09 (0.8-1.47)
cg00574958	*CPT1A*	0.131	0.14	3.43E-01	1.14 (0.87-1.5)	0.096	0.16	5.43E-01	1.1 (0.81-1.5)
cg01400685	*FADS2*	0.093	0.14	4.98E-01	1.1 (0.84-1.43)	0.136	0.15	3.76E-01	1.15 (0.85-1.55)
cg06690548	*SLC7A11*	-0.069	0.12	5.78E-01	0.93 (0.73-1.19)	-0.009	0.13	9.49E-01	0.99 (0.77-1.28)
cg19610905	*FADS2*	0.055	0.13	6.77E-01	1.06 (0.81-1.37)	-0.004	0.15	9.80E-01	1 (0.74-1.33)
cg07504977	NA#	0.057	0.14	6.83E-01	1.06 (0.8-1.39)	0.017	0.16	9.15E-01	1.02 (0.74-1.39)
cg11250194	*FADS2*	-0.032	0.15	8.37E-01	0.97 (0.72-1.31)	0.047	0.17	7.85E-01	1.05 (0.75-1.46)
cg03440556	*SCD*	0.016	0.14	9.11E-01	1.02 (0.77-1.34)	0.122	0.16	4.40E-01	1.13 (0.83-1.54)
cg24503796	*SCD*	-0.015	0.15	9.20E-01	0.99 (0.73-1.32)	-0.046	0.16	7.77E-01	0.95 (0.69-1.31)

*1st model: logistic regression model that included the following covariates: age, sex, BMI, C-reactive protein levels, hemoglobin A1c levels, smoking status, current hypertension, physical activity, white blood cell count and estimated proportions of white blood cell type; 2nd model: model 1 with intake of lipid lowering drug as additional covariate; †β coef: β coefficient; ‡SE: standard error; §OR: Odds Ratio CI: confidence interval; #NA: no gene annotation for this CpG according to the UCSC Genome Browser; **significant; level of significance: 1.7E-03

Table X: **EWAS of NMR-measured metabolites: Results of the investigation of associations between metabolite-related CpG sites and prevalent type diabetes in KORA F4 (N=1,662, with N=78 cases of prevalent type 2 diabetes).**

CpG	gene	1st model*				2nd model*			
		β coef†	SE‡	p-value	OR (95% CI)§	β coef†	SE‡	p-value	OR (95% CI)§
cg19693031	*TXNIP*	-0.506	0.09	3.59E-08**	0.6 (0.5-0.72)	-0.531	0.09	1.28E-08**	0.59 (0.49-0.71)
cg00574958	*CPT1A*	-0.316	0.10	1.58E-03**	0.73 (0.6-0.89)	-0.329	0.10	1.30E-03**	0.72 (0.59-0.88)
cg11024682	*SREBF1*	0.293	0.10	4.41E-03	1.34 (1.1-1.64)	0.288	0.11	6.04E-03	1.33 (1.09-1.64)
cg27243685	*ABCG1*	0.251	0.10	1.13E-02	1.29 (1.06-1.56)	0.189	0.10	6.12E-02	1.21 (0.99-1.47)
cg06500161	*ABCG1*	0.228	0.10	2.36E-02	1.26 (1.03-1.53)	0.132	0.10	2.07E-01	1.14 (0.93-1.4)
cg20544516	*MIR33B/ SREBF1*	0.161	0.09	9.03E-02	1.17 (0.98-1.41)	0.155	0.10	1.09E-01	1.17 (0.97-1.41)
cg07504977	NA#	0.141	0.09	1.33E-01	1.15 (0.96-1.38)	0.135	0.10	1.57E-01	1.14 (0.95-1.38)
cg16246545	*PHGDH*	-0.122	0.08	1.50E-01	0.88 (0.75-1.05)	-0.137	0.09	1.15E-01	0.87 (0.74-1.03)
cg06690548	*SLC7A11*	-0.100	0.08	2.22E-01	0.9 (0.77-1.06)	-0.094	0.08	2.54E-01	0.91 (0.77-1.07)
cg01400685	*FADS2*	-0.049	0.09	5.72E-01	0.95 (0.8-1.13)	-0.039	0.09	6.65E-01	0.96 (0.81-1.15)
cg11250194	*FADS2*	0.058	0.10	5.80E-01	1.06 (0.86-1.3)	0.076	0.11	4.76E-01	1.08 (0.88-1.33)
cg24503796	*SCD*	0.013	0.10	8.99E-01	1.01 (0.83-1.23)	-0.010	0.10	9.18E-01	0.99 (0.81-1.21)
cg27386326	NA#	0.010	0.10	9.22E-01	1.01 (0.83-1.24)	0.020	0.10	8.49E-01	1.02 (0.83-1.25)
cg19610905	*FADS2*	0.007	0.09	9.35E-01	1.01 (0.85-1.2)	-0.015	0.09	8.71E-01	0.99 (0.82-1.18)
cg03440556	*SCD*	-0.003	0.10	9.74E-01	1 (0.83-1.2)	0.020		8.35E-01	1.02 (0.84-1.24)

*1st model: logistic regression model that included the following covariates: age, sex, BMI, C-reactive protein levels, smoking status, current hypertension, history of myocardial infarction, physical activity, white blood cell count and estimated proportions of white blood cell type; 2nd model: model 1 with intake of lipid lowering drug as additional covariate; †β coef: β coefficient; ‡SE: standard error; §OR: Odds Ratio; CI: confidence interval; #NA: no gene annotation for this CpG according to the UCSC Genome Browser; **significant; level of significance: 1.7E-03
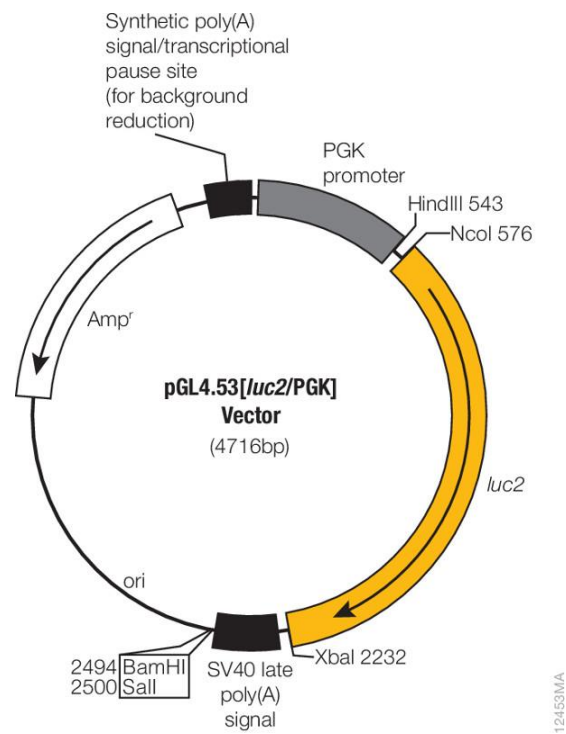
7.2 Figures

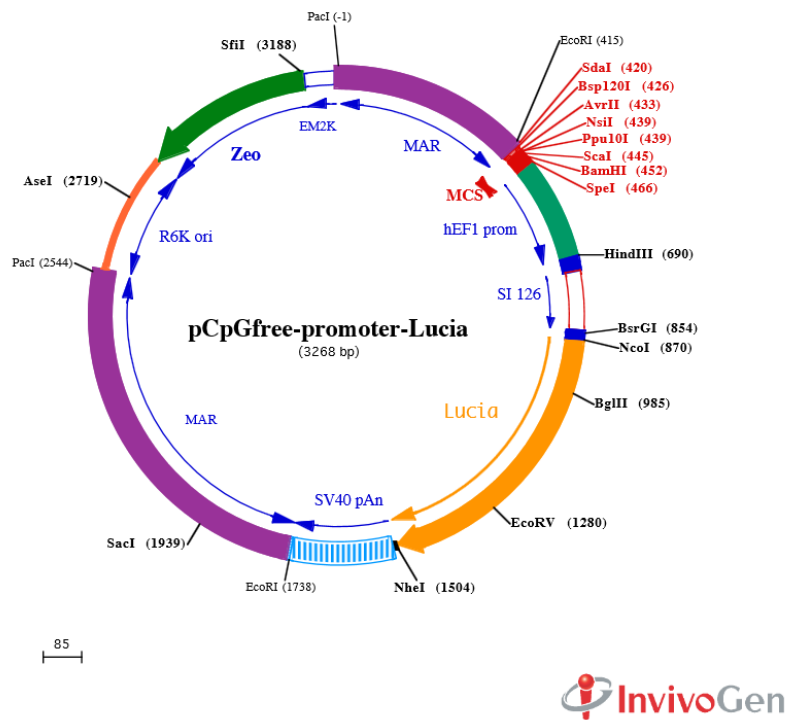**Figure I: Vector map of pGL4.53(luc2/PGK) vector (Promega)**
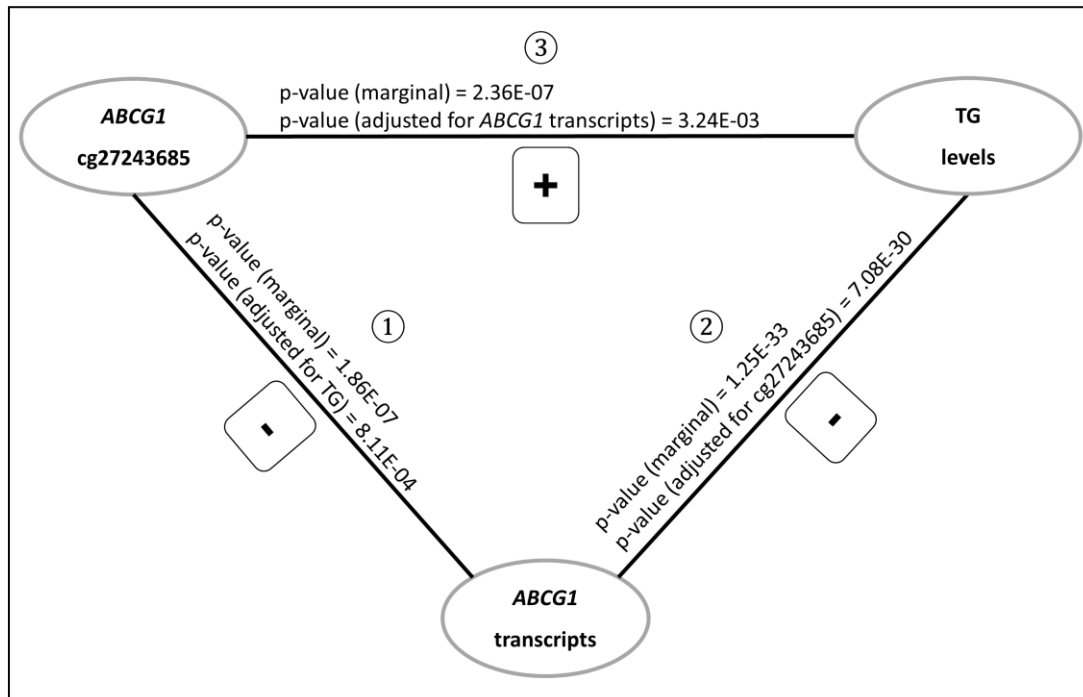

**Figure II: Vector map of pCpGfree-promoter-Lucia plasmid (Invivogen)**

Figure III: EWAS of main blood lipids: Triangluar relationship between methylation of cg27243685 (*ABCG1*), *ABCG1* gene expression and triglyceride levels. ① Methylation of cg27243685 was negatively associated with *ABCG1* mRNA levels. ② *ABCG1* mRNA levels were negatively associated with TG levels. ③ The association between *ABCG1* methylation and TG levels became less significant after adjusting for *ABCG1* transcripts. Negative and positive associations are indicated by minus or plus, respectively. N=724; Level of significance=8.3E-04.
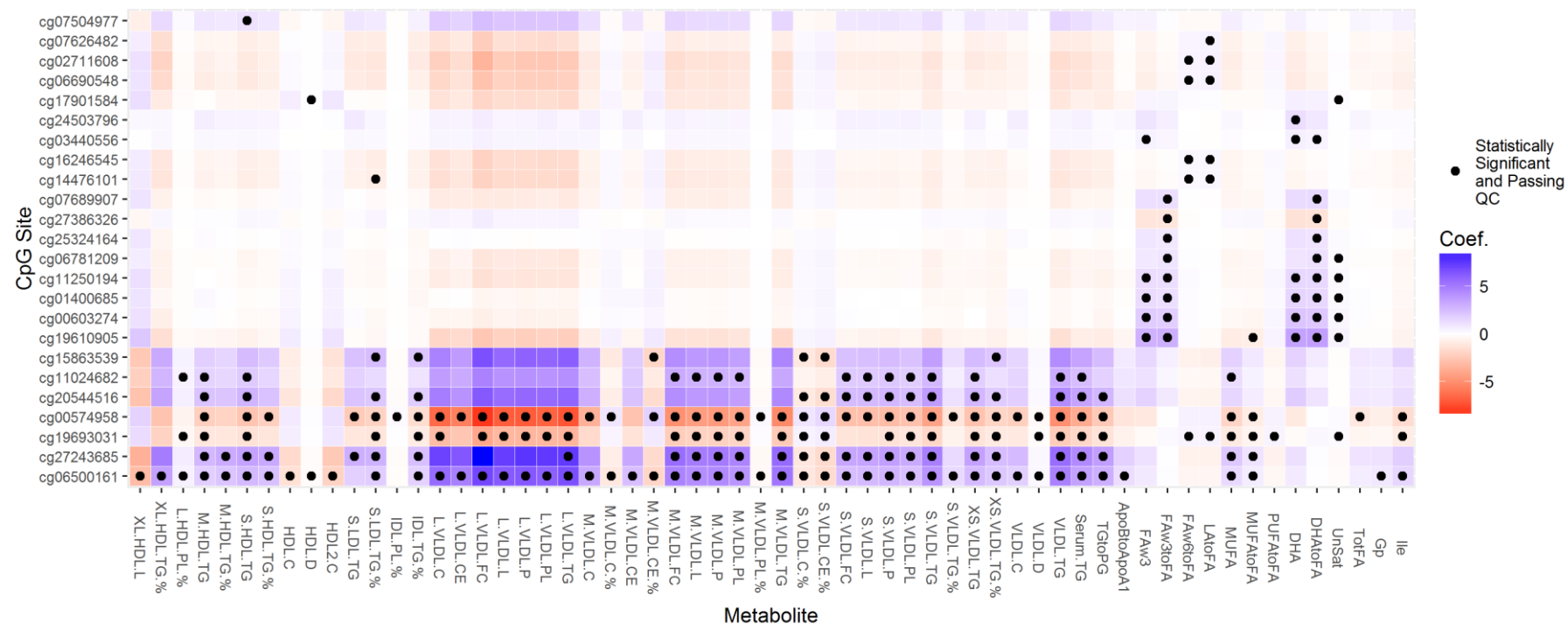
Figure IV: EWAS revealed associations between methylation of certain CpG sites and NMR-measured metabolites in KORA F4 (N=1,662). Black points indicate the N=237 statistically significant associations between CpG sites (listed left) and metabolites, which also passed the quality control (see **chapter 2.7.2.1**.) Boxes are colored according to the signs and values of the coefficients. Level of significance: 4.73E-10.
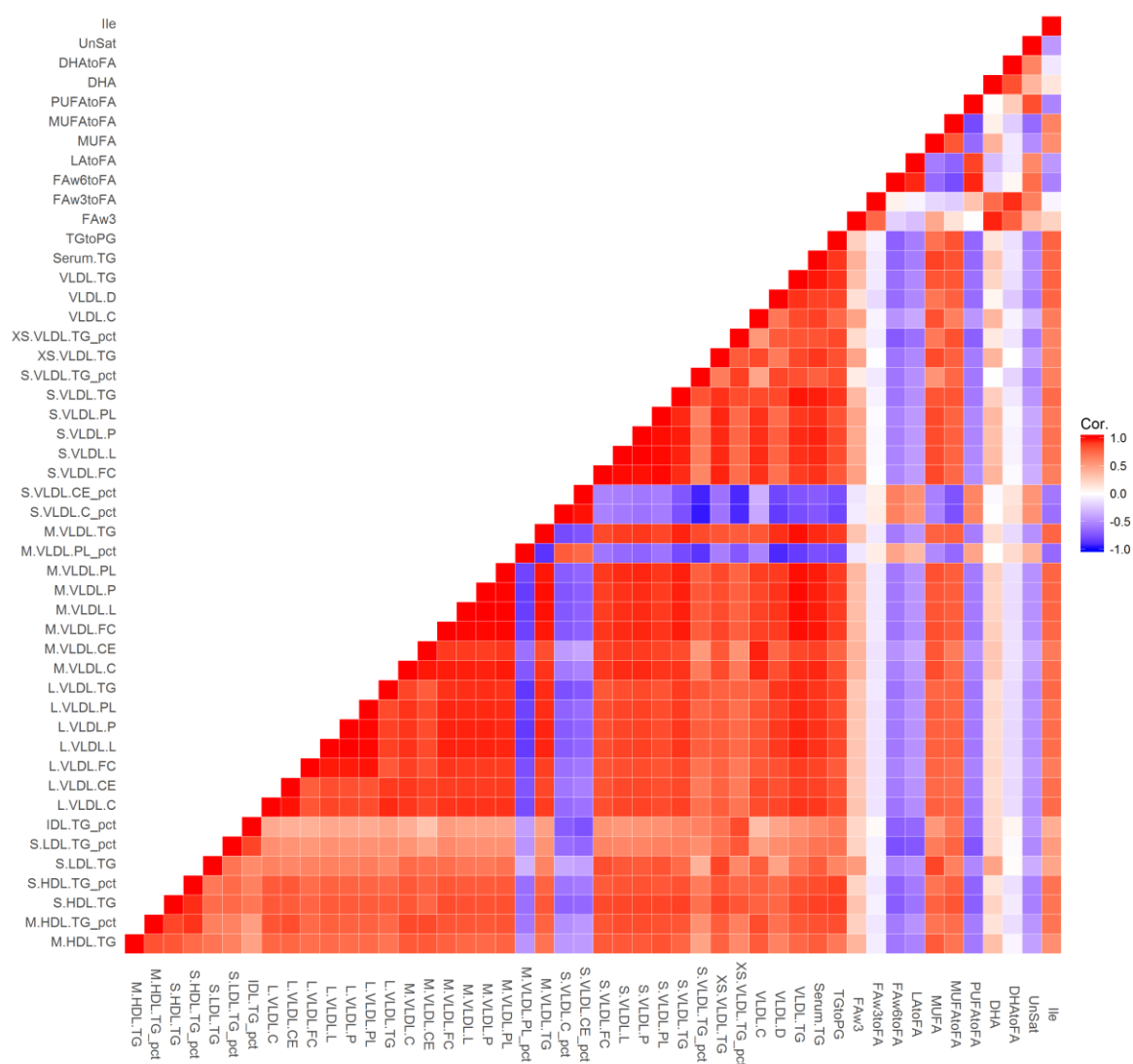
Figure V: NMR-measured metabolites, which were associated with DNA methylation, are highly correlated with each other in KORA F4. Boxes are colored according to the the sign and strength of the correlation.

LIST OF PUBLICATIONS AND PRESENTATIONS

Publications included in this thesis

Pfeiffer L., Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, Holdt LM, Kretschmer A, Schramm K, Adamski J, Klopp N, Illig T, Hedman ÅK, Roden M, Hernandez DG, Singleton AB, Thasler WE, Grallert H, Gieger C, Herder C, Teupser D, Meisinger C, Spector TD, Kronenberg F, Prokisch H, Melzer D, Peters A, Deloukas P, Ferrucci L, Waldenberger M.: DNA Methylation of lipid-related genes affects blood lipid levels. Circ Cardiovasc Genet. 2015 Apr;8(2):334-42. doi: 10.1161/CIRCGENETICS.114.000804

Publications not included in this thesis

Sigurdson AJ, Brenner AV, Roach JA, Goudeva L, Müller JA, Nerlich K, Reiners C, Schwab R, **Pfeiffer L**, Waldenberger M, Braganza M, Xu L, Sturgis EM, Yeager M, Chanock SJ, Pfeiffer RM, Abend M, Port M: Selected single-nucleotide polymorphisms in *FOXE1*, *SERPINA5*, *FTO*, *EVPL*, *TICAM1* and *SCARB1* are associated with papillary and follicular thyroid cancer risk: replication study in a German population. Carcinogenesis 2016 doi: 10.1093/carcin/bgw047

Chunyu Liu*, Riccardo E Marioni*, Åsa Hedman*, **Liliane Pfeiffer***, Pei-Chien Tsai *, Lindsay Reynolds*, Allan C Just*, Qing Duan*, Cindy G Boer*, Toshiko Tanaka*, Cathy E Elks, Stella Aslibekyan, Jennifer A Brody, Brigitte Kühnel, Christian Herder, Degui Zhi, Tianxiao Huan, Chen Yao, Michael Mendelson, Roby Joehanes, Liming Liang, Shelly-Ann Love, Weihua Guan, Sonia Shah, Allan F McRae, Anja Kretschmer, Holger Prokisch, Konstantin Strauch, Annette Peters, Peter M Visscher, Naomi R Wray, Xiuqing Guo, Kerri L Wiggins, Lynn M Almli, Alicia K Smith, Elisabeth B Binder, Kerry J Ressler, Ryan Irvin, Devin M Absher, Dena Hernandez, Luigi Ferrucci, Stefania Bandinelli, Kurt Lohman, Jingzhong Ding, Letizia Trevisi, Stefan Gustafsson, Johanna Sandling, Lisette Stolk, André G Uitterlinden, Idil Yet, Juan Castillo-Fernandez, Timothy D Spector, Joel D Schwartz, Pantel Vokonas, Lars Lind, Yun Li, Myriam Fornage, Donna K Arnett, Nona Sotoodehnia, Ken K Ong, Joyce BJ van Meurs, Karen N Conneely†, Andrea A Baccarelli†, Ian J Deary†, Jordana T Bell†, Kari E North†, Yongmei Liu†, Melanie Waldenberger†, Stephanie London†, Erik Ingelsson†, Daniel Levy†: A DNA Methylation Biomarker of Alcohol Consumption. Submitted. *joint first authors

Rory Wilson, Simone Wahl, **Liliane Pfeiffer**, Cavin K. Ward-Caviness, Sonja Kunze, Anja Kretschmer, Eva Reischl, Annette Peters, Christian Gieger, Melanie Waldenberger: DNA Methylation: Smoking and Quitting, A Longitudinal Analysis. Submitted

Brian H. Chen*, Claudia Schurmann*, Marjolein J. Peters*, Katharina Schramm*, Luke C. Pilling*, **Liliane Pfeiffer***, Roby Joehanes*, Stefania Bandinelli, Maren Carstensen-Kirberg, Paul Courchesne, L. Adrienne Cupples, Valur Emilsson, Tonu Esko, Stephan B. Felix, Lita Freeman, Harald Grallert, Dena G. Hernandez, Albert Hofman, Georg Homuth, Tianxiao Huan, Till Ittermann, Andrew D. Johnson, Sekar Kathiresan, Thomas Meitinger, Peter J. Munson, Matthias Nauck, Gina Pelos, Annette Peters, Eva Reischl, Alan T. Remaley, Michael Roden, Andrew B. Singleton, Seth G. Thacker, André G. Uitterlinden, Cornelia van Dujin, Joyce B.J. van Meurs, Melanie Waldenberger, Hanieh Yaghootkar, Saixia Ying, David Melzer,Luigi Ferrucci Holger Prokisch, Christian Herder, Aaron Isaacs, Daniel Levy, Alexander Teumer on behalf of the CHARGE Consortium Gene Expression Working Group. Meta-Analysis of Whole-Blood Gene Expression Associations with Circulating Lipid Levels. Submitted. Submitted. *joint first authors

So-Youn Shin, Andrew Simpkin, **Liliane Pfeiffer**, Hannah R Elliott, Stephen Burgess, Åsa K Hedman, Alexessander Couto Alves, Geoff Woodward, Tom R Gaunt, Sue Ring, Wendy McArdle, the MuTHER Consortium, Marjo-Riitta Jarvelin, Annette Peters, Mauro Santibanez-Koref, Debbie A Lawlor, Scott M Nelson, Melanie Waldenberger, George Davey Smith, Caroline L Relton: Analysis of causal pathways linking DNA methylation and lipid levels in the Avon Longitudinal Study of Parents and Children. Submitted

Rebecca Thwing Emeny, PhD, MPH; Jens Baumert; Anthony Zannas; Elisabeth B Binder; Sonja Kunze; Simone Wahl; Stella Iurato; Angelika Erhardt; Peter Weber; Anja Kretschmer; **Liliane Pfeiffer**; Johannes Kruse; Konstantin Strauch; Christian Gieger; Melanie Waldenberger; Annette Peters; Karl-Heinz Ladwig: Anxiety Associated Increased CpG Methylation in the Promoter of ASB1: A translational approach, evidenced by cohort and clinical studies, in vivo experiments, and a murine model. Submitted

Fu, Z.; Gilbert, E.; **Pfeiffer, L.**; Zhang, Y.; Fu, Y.; Liu, D.: Genistein ameliorates hyperglycemia in a mouse model of non-genetic type 2 diabetes. Appl Physiol Nutr Metab. 2012 Jun;37(3):480- 8. doi: 10.1139/h2012-005

Poster presentations

Pfeiffer *et al*.: Associations between genome-wide DNA methylation and lipid levels; 4th Clinical Epigenetics International Meeting CLEPSO 2014, Düsseldorf, Germany

Pfeiffer *et al*.: DNA methylation of lipid-related genes affects blood lipid levels: a genomewide screen; ASHG 2014 Annual meeting, San Diego, USA

Pfeiffer *et al*.: Associations between genome-wide DNA methylation and lipid levels; Epigenomics of common diseases 2013, Cambridge, UK

Oral presentations

84th EAS Congress Innsbruck 2016: "Associations between lipoprotein subfractions and genome-wide DNA methylation in KORA F4"

On the Tip of Omics: The 6th Grainau Workshop of Genetic Epidemiology, Grainau, 2016: "Genome-wide DNA methylation analyses of human blood lipid levels"

XIX Lipid Meeting, Leipzig, 2015: "Associations between lipoprotein subfractions and genome-wide DNA methylation in KORA F4"

5th Workshop of Genetic Epidemiology, Bad Aibling, 2015: "DNA Methylation of lipid related gene affects blood lipid levels"

Genetic Variants, Omics, and Individualized Prevention, Grainau, 2013: "Associations between epigenome-wide DNA methylation and lipid levels"

XVIII Lipid Meeting, Leipzig, 2013: "Associations between genome-wide DNA methylation and lipid levels"

DANKSAGUNG (ACKNOWLEDGEMENTS)

CURRICULUM VITAE

Persönliche Daten:

Name: Liliane Pfeiffer

Werdegang:

Seit 2012   Doktorandin und wissenschaftliche Mitarbeiterin
                                                Helmholtz Zentrum München
                                                Institut für Epidemiologie II
                                                Abteilung für molekulare Epidemiologie
                                                Arbeitsgruppe Complex Diseases

2009 - 2012   Masterstudium der Ernährungswissenschaft
                                                Technische Universität München
                                                Abschluss: Master of Science (M.Sc.)
                                                Masterarbeit: "Identification of functional variants in
                                                the *ACADS* gene for inter-individual differences in
                                                acyl-carnitine levels"

2008 – 2009   Gastwissenschaftlerin
                                                Department of Human Nutrition, Food and Exercises
                                                Virginia Polytechnic Institute and University
                                                Blacksburg, Virginia, USA

2005 – 2008   Bachelorstudium der Ernährungswissenschaft
                                                Technische Universität München
                                                Abschluss: Bachelor of Science (B.Sc.)
                                                Bachelorarbeit: „Vergleich PCR-basierter Methoden
                                                zum Nachweis von Lupinen-DNA"

2005   Abschluss der Allgemeinen Hochschulreife (Abitur)
                                                Wittelsbacher Gymnasium München
                                                Leistungskurse: Biologie, Physik