TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Experimentelle Genetik

Association analysis between genome-wide DNA methylation patterns and tobacco smoking

Sonja Zeilinger

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.

Vorsitzende: Univ.-Prof. A. Schnieke, Ph.D.

Prüfer der Dissertation:

1. apl. Prof. Dr. J. Adamski
2. Univ.-Prof. Dr. H.-R. Fries
3. apl. Prof. Dr. Th. Illig,
   Medizinische Hochschule Hannover

Die Dissertation wurde am 30.04.2013 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 30.10.2013 angenommen.

Smoking kills. If you're killed, you've lost a very important part of your life.
(Brooke Shields, American actress)

Table of contents

Summary

Environmental factors have been shown to modulate the establishment and maintenance of epigenetic modifications, such as DNA methylation, thereby inducing changes in gene expression and having a profound role on development and health. Studies which have investigated the effects of tobacco smoking on DNA methylation have been largely restricted to promoter regions of malignancy-related genes, whereas only a few have assessed global DNA methylation with a small number of subjects. Nevertheless, these studies provided first evidence that tobacco smoking leads to aberrant DNA methylation, which might have a tremendous effect on the body. In addition, it has been shown that smoking cessation can significantly reduce the risk of developing smoking-related diseases. Reestablishment of the original DNA methylation patterns due to smoking cessation may be a good explanation for this observation, but the effect of cessation time on DNA methylation, and the time period over which DNA methylation is subject to change, is not known. Only recently, improvements in array-based technologies have enabled the assessment of the molecular effects of tobacco smoking in larger populations.

The objectives of this thesis were: 1) the establishment of a platform (Infinium HumanMethylation 450K BeadChip array) that allows to conduct a genome-wide hypothesis free screen for aberrant DNA methylation patterns in whole blood genomic DNA from the KORA (Cooperative Health Research in the Region of Augsburg) cohort, and the investigation of the effect of current tobacco smoking on the epigenome in a relatively large discovery (N=1814) and replication panel (N=479), 2) the investigation of the effect of former tobacco smoking on the epigenome with the 450K BeadChip array in general and in respect to cessation time and cumulative smoke exposure, and 3) the establishment of a platform (Sequenom EpiTYPER) that enables DNA methylation analysis of selected regions in high resolution, in order to validate and fine map five selected loci detected with the 450K BeadChip in 42 current and never smokers of the KORA F4 survey.

Evidence of wide-spread differences in the degree of site-specific methylation as a function of tobacco smoking could be observed in each of the 22 autosomes, identifying 187 genome-wide, significant, differentially methylated CpG sites that correspond to 95 genes, with the percent of variance explained by smoking ranging from 1.31 to 41.02 and p-values ranging from 9.31E-08 to 2.54E-182. Current tobacco smoking was associated with genome-wide hypomethylation, and a link between global hypomethylation and, for example, cardiovascular diseases, Alzheimer's disease, mental disorders, and different forms of cancer has been demonstrated in various studies. Several novel loci that had not previously been implicated in responses to tobacco smoke could be detected, and numerous loci that

had shown smoking-associated changes in DNA methylation in other studies could be replicated. Furthermore, the smoking-associated differential DNA methylation for the five most significant loci could be validated by the EpiTYPER method, backing up the reliability of the 450K BeadChip array. 13 CpG sites also remained significant in former smokers, but overall, depending on cessation time and cumulative smoke exposure, the methylation levels in former smokers were found to be close to levels seen in never smokers. These results may fit to the common observation that former smokers have a significantly reduced risk of developing smoking-related disease, but are, compared to never smokers, still at a higher risk for developing, for example, COPD and various cancers, which might possibly be mediated by smoking-specific methylation patterns that do not recover even after years of cessation. The most significant CpG site in terms of median methylation difference, p-value, and explained variance in current compared to never smokers, as well as in former compared to never smokers, was cg05575921 in the *AHRR* gene. EMSA experiments performed for this site additionally showed methylation-specific binding of protein complexes and therefore suggest a regulatory role on gene expression for this site.

The significant genes detected make sense in a biological context when it comes to the impact of smoking, as they play roles in the development and function of the cellular, hematological, immune, cardiovascular, tumorigenic, and reproductive systems, and therefore may be involved in mediating the effects of smoking in development and progression of disease. Overall, the results confirm the broad effects of tobacco smoking on the human organism and indicate that quitting tobacco smoking allows regaining the DNA methylation state of never smokers for most of the investigated sites. Furthermore, this thesis opens several research topics for possible follow-up projects, for example, to reveal the underlying molecular mechanisms that alter the epigenome due to environmental triggers.

Zusammenfassung

Umweltfaktoren können epigenetische Modifizierungen, wie die der Methylierung von DNA, modulieren, dadurch die Expression von Genen verändern und somit eine wichtige Rolle für die Entwicklung und die Gesundheit spielen. Studien die bisher den Einfluss des Zigarettenrauchens auf die DNA-Methylierung untersuchten, waren zum großen Teil auf den Promoterbereich Tumor-assoziierter Gene beschränkt und nur in wenigen Studien mit einer kleinen Anzahl an Probanden wurde DNA-Methylierung genomweit bestimmt. Dennoch lieferten diese Studien erste Hinweise, dass das Rauchen von Zigaretten zu veränderter DNA-Methylierung führt, welche enorme Auswirkungen auf den Körper haben könnte. Es konnte zudem gezeigt werden, dass das Risiko, eine Folgeerkrankung des Rauchens zu entwickeln, durch Aufgeben des Rauchens signifikant reduziert werden kann. Die Wiederherstellung der ursprünglichen DNA-Methylierungsmuster durch das Aufgeben des Rauchens könnte eine gute Erklärung für diese Beobachtung sein, aber die Auswirkung der rauchfreien Zeit auf die DNA-Methylierung, sowie der zeitliche Verlauf in dem Veränderungen in der DNA-Methylierung stattfinden können, sind nicht bekannt. Erst seit kurzem ermöglichen es „array-basierte“ Technologien, die molekularen Auswirkungen des Rauchens in größeren Populationen zu bestimmen.

Die Ziele dieser Arbeit waren: 1.) Die Etablierung einer Plattform (Infinium HumanMethylation 450K BeadChip array), welche die Durchführung einer genomweiten hypothesenfreien Suche nach veränderten DNA-Methylierungsmustern in aus Vollblut gewonnener genomischer DNA der KORA (Kooperative Gesundheitsforschung im Raum Augsburg) Kohorte ermöglicht, und Untersuchung der Auswirkungen des Zigarettenrauchens auf das Epigenom einer relativ großen „Discovery-“ (N=1814) sowie einer Replikations- (N=479) Studie. 2.) Die Ermittlung der Auswirkungen ehemaligen Zigarettenrauchens auf das Epigenom mit dem 450K „BeadChip Array“ im Allgemeinen und im Hinblick auf die Zeit seit der Aufgabe des Rauchens und der kumulativen Rauchexposition. 3.) Die Etablierung einer Plattform (Sequenom EpiTYPER), welche eine höhere Auflösung der DNA-Methylierungsanalyse in ausgewählten Regionen ermöglicht, um eine Auswahl von fünf der mit dem 450K „BeadChip Array“ detektierten Loci in 42 Rauchern und Nie-Rauchern der KORA F4 Studie zu validieren und fein zu kartieren.

In allen 22 Autosomen konnten Unterschiede des „CpG site“ -spezifischen Methylierungsgrades in Abhängigkeit des Rauchens von Zigaretten festgestellt werden. 187 der unterschiedlich methylierten „CpG-sites“, welche sich 95 Genen zuordnen lassen, konnten mit einer durch das Rauchen erklärbaren prozentualen Varianz von 1.31 bis 41.02 und p-Werten von 9.31E-08 bis 2.54E-182 als genomweit signifikant identifiziert werden.

Das Rauchen von Zigaretten ging einher mit genomweiter Hypomethylierung, und ein Zusammenhang zwischen globaler Hypomethylierung und zum Beispiel Herz-Kreislauf-Erkrankungen, Alzheimer, psychischen Erkrankungen und unterschiedlichen Formen von Krebs, konnte bereits in verschiedenen Studien demonstriert werden. Einige neue „Loci“ die zuvor nicht mit Reaktionen auf das Rauchen von Zigaretten in Verbindung gebracht wurden, konnten detektiert werden und zahlreiche „Loci“, die in anderen Studien bereits eine mit dem Rauchen assoziierte veränderte DNA-Methylierung zeigten, konnten bestätigt werden. Zudem konnte die mit dem Rauchen assoziierte unterschiedliche DNA-Methylierung für die fünf „Loci“ mit der höchsten Signifikanz durch die EpiTYPER Methode validiert werden, was die Zuverlässigkeit des 450K „BeadChip Arrays“ bestätigt. 13 „CpG-sites“ blieben auch bei ehemaligen Rauchern noch signifikant, im Großen und Ganzen jedoch war der Methylierungsgrad, abhängig von der Zeit seit Aufgabe des Rauchens und der kumulativen Rauchexposition, in ehemaligen Rauchern dem von Nie-Rauchern sehr ähnlich. Diese Ergebnisse könnten zur allgemeinen Beobachtung passen, dass ehemalige Raucher zwar ein signifikant geringeres Risiko auf das Rauchen zurückführbare Erkrankungen zu entwickeln haben, aber verglichen mit Nie-Rauchern dennoch ein höheres Risiko für die Entwicklung von zum Beispiel COPD und verschiedenen Krebserkrankungen aufweisen. Möglicherweise wird dies durch das Rauchen verursachte spezifische Methylierungsmuster, welche sich auch nach mehreren rauchfreien Jahren nicht regenerieren, vermittelt. Die bedeutendste „CpG-site“ im Hinblick auf mittlere Methylierungsdifferenz, p-Wert und erklärte Varianz bei Rauchern im Vergleich zu Nie-Rauchern, als auch bei ehemaligen Rauchern im Vergleich zu Nie-Rauchern, war cg05575921 im *AHRR* Gen. EMSA Experimente, die für diese CpG-site durchgeführt wurden, zeigten zudem methylierungsspezifische Bindungen von Proteinkomplexen und weisen somit auf eine regulatorische Rolle dieser „CpG-site“ in der Genexpression hin.

Die detektierten signifikanten Gene sind bezogen auf die Auswirkungen des Rauchens im biologischen Kontext plausibel, da sie in der Entwicklung und Funktion des zellulären, hämatologischen, immunologischen, kardiovaskulären, tumorspezifischen und reproduktiven Systems eine Rolle spielen. Somit könnten sie bei der Vermittlung der Auswirkungen des Rauchens auf Entwicklung und Progression von Krankheiten involviert sein. Insgesamt bestätigen die hier gewonnenen Resultate weitreichende Auswirkungen des Rauchens auf den menschlichen Organismus und indizieren, dass das Aufgeben des Rauchens vermutlich das Wiedererlangen des Methylierungsstatus von Nie-Rauchern für die meisten untersuchten „CpG-sites“ ermöglicht. Diese Dissertation eröffnet zudem etliche Forschungsthemen für mögliche Folgeprojekte, um zum Beispiel die zugrundeliegenden molekularen Mechanismen, welche aufgrund von Auslösern aus der Umwelt das Epigenom verändern, aufzudecken.

LIST OF ABBREVIATIONS

BMI	body mass index
bp	base pairs
CGI	CpG island
COPD	chronic obstructive pulmonary disease
CpG	cytosin-phosphate-guanin site
ddNTP	dideoxynucleotides
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxynucleotides
EDTA	ethylenediamine tetraacetic acid
EMSA	electrophoretic mobility shift assay
h	hours
kb	kilobase
KORA	Cooperative Health Research in the region of Augsburg
m/z	mass-to-charge ratio
MALDI	matrix-assisted laser desorption/ionisation
mg	milligrams
min	minute
ml	milliliter
mM	millimolar
mRNA	messenger ribonucleic acid
MS	mass spectrometry
N	number of subjects
ng	nanogram
nl	nanoliter
nm	nanometer
PCR	polymerase chain reaction
RCF	relative centrifugal force
RNA	ribonucleic acid
RT	room temperature
SAP	shrimp alkaline phosphatase
sec	seconds
SNP	single nucleotide polymorphism
TBE	tris-borate-EDTA
TE	Tris (tris(hydroxymethyl)aminomethane) + EDTA
TOF	time-of-flight
TSS	transcriptional start site
UTR	untranslated region
UV	ultraviolet
WBC	white blood cell
x g	centrifuge rotor speed in g
µl	microliters

1. Introduction

1.1. Epigenetics

1.1.1. Concept of Epigenetics

In principal, all cells of a multicellular organism share an identical genotype, but still these cells obviously display a broad morphological and functional diversity within the organism which is accomplished by differential gene expression. The identity and developmental potential of an individual cell is therefore not only defined by its genetic component, but also by changes that occur without a change in the genome sequence. This branch of research is known as “epigenetics” with regard to chemical interactions that modify either DNA itself or its packaging proteins, the histones, or change micro RNA expression [1]. The word “epi” comes from the Greek and means the same as “over”, “above” or “in addition to”, so the epigenetic code can be seen as a second layer of information on top of the genetic code. In 1942 Conrad Hal Waddington first introduced the term epigenetics as “the study of the interactions between genes and their products which bring phenotype into being“ [2]. Epigenetics now is understood as the study of mitotically and/or meiotically heritable changes in gene expression that are not caused by changes in the DNA sequence, but by biochemical modifications [3].

Epigenetics, in a broad sense, is a connection between genotype and phenotype that alters the final outcome of genomes without changing the underlying DNA sequences. This includes covalent and non-covalent modifications of DNA and histone proteins by which the overall chromatin structure is influenced. Epigenetic modifications can be passed from one cell generation to the next (mitotic inheritance) and in some cases between generations of a species (meiotic inheritance). That epigenetic traits can be inherited from one generation to the next is well established in plants [4]. Not much information is available about the inheritance of epigenetic modifications between generations in mammals, however a few studies could identify environmentally induced epigenetic transgenerational inheritance of adult onset disease phenotypes in F3 generations [5]. For example, the environmental toxicant 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) has been shown to negatively affect fertility, dioxin has been shown to promote ovarian diseases, and nutritional alterations have been shown to promote obesity, all in the F3 generation via epigenetic transgenerational inheritance [6-8].

1.1.2. Basics of Epigenetics

1.1.2.1. Epigenetic mechanisms

Mechanisms of epigenetic regulation can alter the accessibility of gene promoters and regulatory regions and include DNA methylation, histone modification, chromatin remodeling complexes, and RNA interference [9,10]. DNA methyltransferases and histone-modifying as well as chromatin remodeling enzymes work in concert with miRNA mechanisms to establish a specific chromatin structure that defines the transcriptional state of genes (Figure 1).

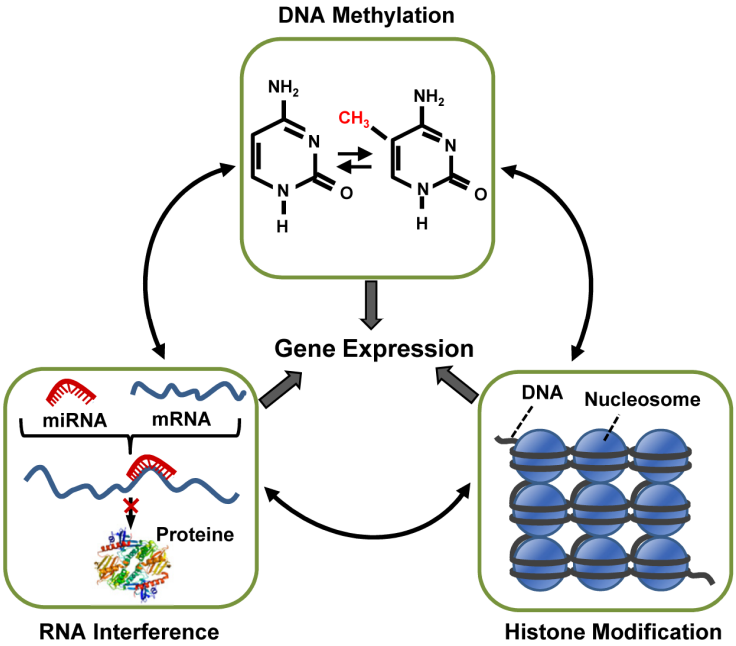


Figure 1. Mechanisms of epigenetic regulation. DNA methylation, histone modifications, and RNA-mediated gene silencing constitute three distinct mechanisms of epigenetic regulation. DNA methylation refers to the covalent attachment of a methyl group to a cytosine (C) that is located 5' to a guanine (G) in a CpG dinucleotide, to form 5-methylcytosines. Histone modifications are covalent post-translational modifications of N-terminal tails of four core histones (H3, H4, H2A, and H2B). The most recent mechanism of epigenetic inheritance involves small RNAs, which interfere with mRNA. Each of these mechanisms controls and is controlled by the others creating a complex regulatory machinery. Modified from [11-14].

Epigenetic mechanisms are crucial to normal development and differentiation of distinct cell lineages in the adult organism and are critical for maintaining cell fate during development [15]. Although almost every cell in the human body contains the same DNA, epigenetic marks act to program the cell to express genes that are relevant for a particular tissue type. A neuronal cell, for example, expresses genes that help it to develop dendrites and axons. In a liver cell those same genes are marked with epigenetic tags that cause tighter binding of the DNA, making it inaccessible to the transcription machinery. Epigenetic mechanisms allow

the stable propagation of gene expression states from one generation of cells to the next, are not definitive but flexible with changes occurring with age in a stochastic manner as well as in response to environmental stimuli [1,16,17]. They therefore mediate genomic adaption to an environment, ultimately contributing toward a specific phenotype.

DNA methylation is the best understood, frequent and stable epigenetic mechanism and refers to the covalent attachment of a methyl group to the C5 position of cytosine residues to form 5-methylcytosines [18]. In mammals this predominantly occurs at 5`-CpG-3`dinucleotides which are cytosines followed by guanines on one strand [3]. Since this epigenetic modification is the relevant one for this doctoral thesis, it will be separately described in detail in chapter 1.2., the others will be introduced within the following two subchapters.

1.1.2.1.1. Histone modifications

The genetic information of eukaryotic organisms is packaged into a compacted chromatin structure. The main component of this structure is the nucleosome, which individually assembles 147 DNA base pairs around a core histone octamer. Each octamer contains two H3-H4 histone dimers bridged together as a stable tetramer that is flanked by two separate H2A-H2B dimers (Figure 2a). Through a histone fold domain, also called the globular domain, nucleosomal histones can interact with each other. Open (eu-) and closed (hetero-) chromatin states, which define the gene accessibility to the transcriptional machinery, are regulated by modifications to both DNA and histone tails, including acetylation, methylation, phosphorylation, and ubiquitylation [19]. The open, less densely packed chromatin state allows binding of transcription factors (TF) and the general transcriptional machinery to the gene regulatory region, and is therefore transcriptionally active. Open chromatin is typically associated with demethylated DNA and highly acetylated histone tails (Ac). The closed chromatin structure is more densely packed, does not allow the binding of transcription factors and is therefore transcriptionally silent. It is associated with repressive chromatin marks, such as DNA methylation (Me), and deacetylated and methylated histone lysine (H3K9) residues (Figure 2b).

The histone code hypothesis implies that distinct histone modifications, on one or more tails, act sequentially or in combination to form a `histone code' that is read by other proteins and leads to distinct downstream events [20]. But the function of these chromatin marks is even more complex and in many cases there is not a strict division between active and repressive modification states. Instead, many of these histone modification marks have several, seemingly conflicting roles. A prominent feature which shows the dynamic behavior of histone modification, for example, is a mechanism in which modifications establish gene

activation and then reinstate repression, therefore the presence of certain modifications may not indicate a unique regulatory “on” or “off” status [21].

Also, chromatin remodeling complexes alter nucleosome architecture to expose or hide regions of DNA for transcriptional regulation. There are currently four different families of chromatin remodeling complexes known: SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation switch), CHD (chromodomain, helicase, DNA binding) and INO80 (inositol requiring 80). All of them utilize ATP hydrolysis to alter histone-DNA and work in concert with chromatin modifiers to direct nucleosome dynamics [22] (Figure 2b).

DNA methylation is highly related to certain chromatin modifications and it has been shown that DNA and histone modifying enzymes directly interact and constitute links between local DNA methylation and regional chromatin structure [23]. In fact, DNA methyltransferases (DNMTs) and histone deacetylases (HDACs) are found in the same multi-protein complexes and methyl-binding proteins interact with HDACs and histone methyltransferases, as well as with the chromatin remodeling complexes. Changes in the chromatin structure are followed by DNA methylation in many cases, heritably locking the gene in its inactive state [3].

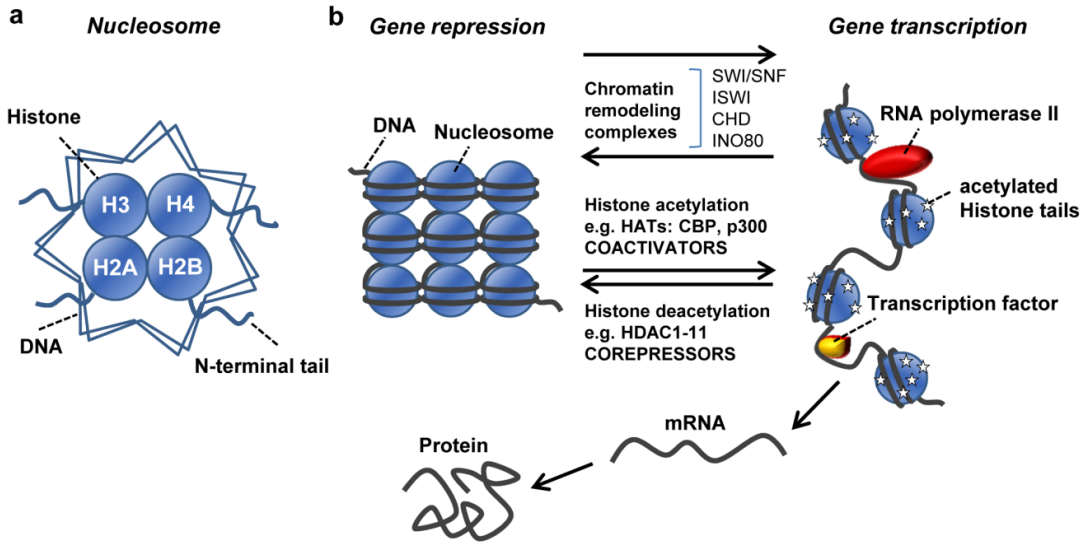


Figure 2. The structure of chromatin and schematic representation of eu- and heterochromatin. **a)** The structure of chromatin. DNA is wound around nucleosomes which are composed of eight histone molecules with two copies of histones H2A, H2B, H3, and H4. Each histone molecule has a long tail rich in lysine residues (K), which are the sites of enzymatic modification, such as acetylation, thus changing the charge of the molecule and leading to DNA unwinding. **b)** Schematic representation of euchromatin and heterochromatin. Activation and repression of genes is regulated by acetylation of core histones, mediated by coactivators that have intrinsic histone acetyltransferase (HAT) activity. The open chromatin structure allows binding of RNA polymerase II and transcription factors that were unable to bind DNA in the closed chromatin configuration. Corepressors, which include histone deacetylases (HDACs), can reverse this acetylation and cause gene silencing. CBP: CREB-binding protein; PCAF: p300/CBP-associated factor; NCoR: nuclear receptor corepressor. Modified from [11])

1.1.2.1.2. RNA interference (RNAi)

RNAi is a biological mechanism in eukaryotic cells that can act through either microRNA (miRNA) or small interfering RNA (siRNA), and takes part in controlling which genes are active and how active they are. miRNAs differ from siRNAs in their biogenesis, not in their functions. First, miRNAs are endogenous expressed products of an organism's own genome, whereas siRNAs are primarily exogenous in origin, derived directly from the virus or transposon. Second, miRNAs are processed from stem-loop precursors with incomplete double-stranded character, whereas siRNAs are excised from long, fully complementary double-stranded RNAs (dsRNAs) [24].

The RNAi pathway is initiated by the enzyme dicer which cleaves long dsRNA molecules into short fragments of ~20-25 nucleotides. The siRNA will be unwound into a passenger and a guide strand, followed by degradation of the passenger strand and incorporation of the guide strand into the RNA-induced silencing complex (RISC). The miRNA and siRNA strands subsequently guide the RISCs to complementary RNA molecules, resulting in translational repression or target degradation (effector step) [24,25] (Figure 3).

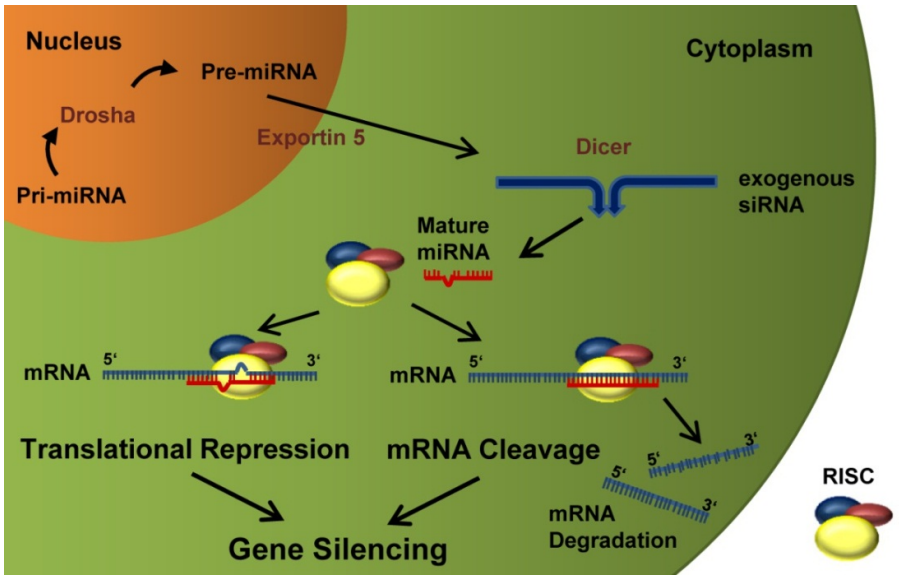


Figure 3. The principal mechanism of RNA interference. Inside the nucleus, pri-miRNAs are cleaved by Drosha to pre-miRNAs. With the aid of Exportin 5, the pre-miRNAs are then exported into the cytoplasm. The subsequent cascade is shared with exogenous siRNA and mainly involves Dicer and RISC. The Dicer processes the pre-miRNAs and exogenous siRNAs to mature miRNAs or functional siRNAs, which bind to RISC and to complementary RNA sequences. mRNAs with perfect complementary sequences to the siRNAs are targeted to the degradation while transcripts captured by miRNAs with incomplete complementary undergo translation repression. Modified from [12,13].

1.2. DNA methylation

1.2.1. Basic principles of DNA methylation

DNA methylation refers to the covalent attachment of a methyl group ($-CH_3$) to the carbon-5 position of the cytosine pyrimidine-residue to form 5-methylcytosine [18], as illustrated in Figure 4. The reaction is catalyzed by enzymes called DNA methyltranferases (DNMTs), with methyl groups donated from *S*-adenosyl methionine (SAM) [14].

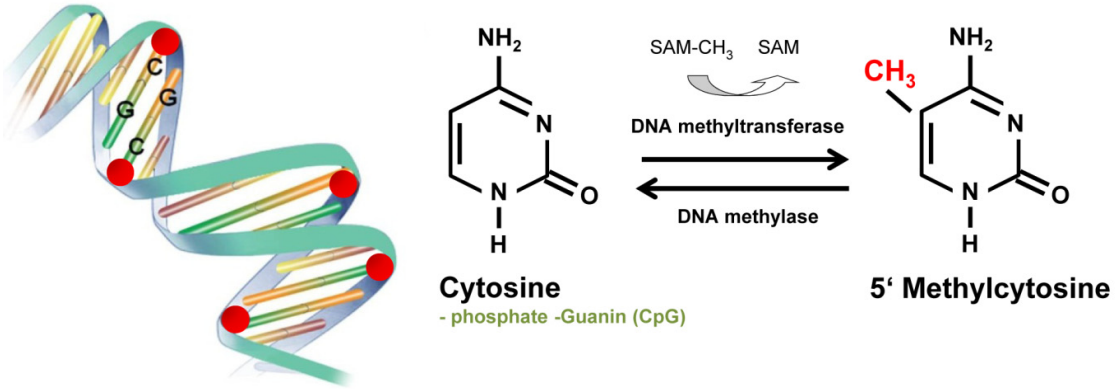


Figure 4. Basic principal of DNA methylation. In mammals, a cytosine base located in the context of CpG can be methylated by the transfer of a methyl group from the methyl donor SAM to the 5′ position of the cytosine ring by DNMT. This reaction is potentially reversible. Modified from [14,26].

In mammals this modification is predominantly found on the 5’ position of the pyrimidine ring of cytosines in the context of the dinucleotide sequence CpG [3], therefore resulting in two methylated cytosine residues sitting diagonally to each other on opposing DNA strands [27]. However, it also has been shown that hemimethylated CpG dyads may be intermediates in active demethylation during carcinogenesis [28]. 5-methylcytosine accounts for ~1% of all bases in mammalian genomes, varying slightly in different tissue types, and the majority (75%) of CpG dinucleotides are methylated [29]. Methylation of cytosines in other contexts such as CpNpG or CpA sequences is widespread in plants and some fungi and have furthermore been recently detected in mammal embryonic stem cells [30-32]. The function of non-CpG methylation in mammals is currently unknown.

Short CpG-rich regions, known as CpG islands (CGIs), can be found in more than half of the genes in the genome of vertebrates [33]. Depending on the employed set of parameters, according to Gardiner-Garden and Frommer (1987) or Takai and Jones (2002), a CGI is defined as having a G+C content of more than 50% (55%), an observed versus expected ratio for the occurrence of CPG sites (CpGs) of more than 0.6 (0.65), and a minimum size of

200 (500) bp [34,35]. In general, CpGs are underrepresented in the genome, which might be due to the conversion of methylated CpGs to TpGs by spontaneous or enzymatic deamination [36]. In comparison to other transitional mutations, the mutation rates at CpGs have been estimated to be about 10-50 times higher, which has probably led to the depletion of the dinucleotide during evolution [29].

A family of DNA methyltransferase enzymes is involved in *de novo* DNA methylation and its maintenance. The primary establishment of DNA methylation patterns during early embryogenesis is carried out by DNMT3A and DNMT3B, therefore they are fundamental to embryonic development in mammals. The DNMT3 family contains a third member, DNMT3L, which has no DNMT activity, acts as a general stimulatory factor for DNMT3a and DNMT3b and is required for establishing maternal genomic imprinting. It has long been assumed that the maintenance of DNA methylation is accomplished exclusively by DNMT1, which restores hemimethylated DNA generated during DNA replication to the fully methylated state by adding methyl groups to the non-methylated daughter strand. In this manner, DNA methylation and its corresponding effects on gene expression can be passed from cell to cell. The correct maintenance of DNA methylation is essential for embryonic development, imprinting, and X-inactivation. However, several recent studies show that all three DNMTs (DNMT1, DNMT3A and DNMT3B) cooperate and participate in *de novo* methylation and methylation maintenance [36-40].

1.2.2. Key processes regulated by DNA methylation

DNA methylation is dynamically regulated and contributes to the epigenetic regulation of many key developmental processes in mammals:

***Embryogenesis*:** Two phases of epigenetic reprogramming occur in early embryogenesis - a widespread demethylation of the genome followed by de novo methylation and therefore reestablishment of the DNA methylation, which allows to stably transmitting methyl groups to daughter cells after cell division. Perturbations of this process might lead to altered methylation in the cell lineages of the offspring [9,41].

Imprinting: Certain genes are either expressed only from the allele inherited from the mother or from the father (parent-of-origin-specific manner) [42]. Although up to 200 imprinted genes have been computationally predicted in human, about 50 imprinted genes are known [43]. The alleles of these imprinted genes are generally located in clusters and

next to other epigenetic modifications, such as histone acetylation/deacetylation, differentially marked by DNA methylation [42,44].

Inactivation of X-chromosome: X-chromosome inactivation is a process by which one of the copies of the X-chromosome present in female mammals is inactivated. The choice is random and remains inactive throughout the lifetime of the cell. First, the Xist (X-inactive specific transcript) RNA is expressed, which than coats the future inactive chromosome X. After various histone modifications and gene silencing, DNA methylation patterns are established on the inactive X-chromosome, which are important in maintaining the repressed state of inactive X-linked genes [45,46].

Genome stability: A stable genome impedes rearrangements of the genome, whereas an instable genome is susceptible for these mechanisms. DNA methylation and the addition of methyl groups, termed hypermethylation, together with specific histone modifications, stabilizes the genome and prevents recombination events. On the other hand, a loss of methyl groups, termed hypomethylation, allows the occurrence of rearrangement events, such as homologous recombination, which contribute to the etiology of many diseases, especially different forms of cancer [47,48].

Regulation of gene expression: Gene expression refers to the processes needed to turn the information within genes into gene products (by transcription and translation). Most past work on DNA methylation and its role in gene expression focused on CGIs at transcriptional start sites (TSSs), which led to the general assumption that hypomethylation of CGIs corresponds to the maintenance of an open chromatin structure and a potentially active state of transcription, whereas hypermethylation of CGIs is associated with epigenetic silencing [33,49,50]. Suppression of transcription by CpG methylation can be accomplished by several mechanisms. DNA recognition and binding of transcription factors may be directly blocked by the presence of methyl groups at specific CpG sites. It has been shown that transcriptional activation at GC-boxes is inhibited by methylation which prevents for example, binding of Sp1 and Sp3 transcription factors [51,52]. In other cases transcription factor access to regulatory elements might be blocked [37]. For example, the binding of MeCP2 and other family members to methyl CpG contributes to transcriptional repression by the recruitment of histone-modifying proteins, such as histone deacetylases (HDAC) [53], and histone deacetylation promotes chromatin condensation, further repressing transcription [54]. This shows how DNA methylation and certain histone modifications function together to establish a transcriptional on or off state of genes subject to epigenetic modification.

Though, the function of DNA methylation is even more complex. An unmethylated CGI does not necessarily correlate with gene transcription, but rather that the gene can be potentially activated. Also, a methylated CGI does not necessarily induce gene silencing. The expression of the associated gene is only modified when a specific core region of the promoter - usually located at the transcription start site - becomes hypermethylated [55]. This has especially been emphasized with recent approaches that enable epigenome-wide studies, as the impact of methylation on gene expression is influenced by its position in the transcriptional unit [36]. So DNA methylation blocks initiation at the TSS, but does not block transcription in the gene body and even might stimulate it. Recently, a potential mechanism has been proposed linking gene body methylation with splicing [56]. Also, DNA methylation may be activating if it prevents binding or limits expression of transcriptional repressors [37].
All together DNA methylation represents an essential mechanism in the regulation of gene expression.

1.2.3. DNA methylation as a link between environment and disease

Epigenetic modifications become increasingly divergent with age as has been demonstrated by monozygotic twins that experience an epigenetic drift in relation to one another with advancing age, as well as locus-specific inter-individual DNA methylation differences [17,57]. They begin and end life with the same genetic make-up, but as they grow they will experience differences in their environment, some of which might alter their appearance and behavior. Gene expression profiles, which are almost identical between 3 year-old twin pairs, can differentiate by epigenetic marks during lifetime and lead to very different phenotypes by age 50 [17].
Various environmental and lifestyle factors have been shown to modulate the establishment and maintenance of epigenetic modifications, such as DNA methylation, thereby inducing changes in gene expression and having a profound role on development and health as well as development and progression of disease. However, the exact mechanisms are still not understood.

Ageing: Age- and environment-related changes in DNA methylation play a role in phenotypic differences during aging and are relevant to disease development [58]. Hypermethylation of CGIs, for example, can lead to silencing of tumor-suppressor genes, promoting the development of cancer [59,60]; conversely, hypomethylation can lead to gene overexpression, which may also promote cancer [61]. Alterations in DNA methylation have

been furthermore associated with the aging-related diseases atherosclerosis, Alzheimer's and other neurodegenerative and autoimmune diseases [58].

Nutrition: Many dietary components, such as folate, methionine, betaine, or vitamin B6 and B12, are associated with DNA methylation changes in mammals [62-64]. They influence the one-carbon metabolism and therefore the amount of available S-adenosylmethionine (SAM), which is the methyl donor for DNA methylation [62,63]. Folate and other dietary supplements, such as vitamins, affect the activity of enzymes supplying methyl groups and influence the rate of disease manifestation. Reduced amounts of folate, for example, have been associated with genomic instability, neuronal tube defects and genomic hypomethylation. DNA methylation is also influenced by other dietary factors, including polyphenols and selenium, which mediate effects independently of folate metabolism [62,64]. The gestational period is particularly susceptible to epigenetic perturbation and nutrition can have different effects on the placenta and the embryo, which has been shown by various studies linking prenatal nutrition and DNA methylation [65-67].

Environmental toxicants: Various studies have shown an association between DNA methylation and environmental toxicants, such as metals, air pollution, dioxin, benzene or bisphenol A (BPA). Environmental metals, including nickel, cadmium, lead, and particularly arsenic, have been linked to several diseases, such as cancer, neurological disorders, autoimmune and cardiovascular diseases. Air pollution, especially particulate matter (PM), has been associated with cardiovascular and respiratory disease, as well as with lung cancer risk. Dioxin has been associated with carcinogenesis and benzene with increased risk of hematological malignancies. The endocrine disruptor BPA has potential reproductive effects, and is associated with increased cancer risk in adult life through fetal exposures [1,64,68,69]. Further, candidate gene studies have demonstrated that tobacco smoke has the potential to change DNA methylation patterns, which furthermore may be causally related to a variety of disease conditions. As this environmental toxicant will be the main issue of this thesis, it will be further introduced in chapter 1.3.

Others: In addition, DNA methylation changes have been associated with chronic exposure to sunlight [70], alcohol consumption [69,71], UV radiation [72,73], and psychological stress [64,74].

1.2.4. Genetic effects on DNA methylation

Epigenetic regulation is influenced by genetic differences between individuals, and dizygotic twins, for example, show more epigenetic differences than monozygotic twins [75]. Various studies have shown that the level of methylation at CpG sites is affected by single nucleotide polymorphisms (SNPs) nearby [76,77]. However, it is still unclear to which extent the DNA sequence determines the levels of epigenetic modification at specific loci.

1.3. Tobacco smoking

1.3.1. The burden of tobacco smoking

Tobacco smoking, a theoretically easy preventable environmental factor, is a leading cause of disease and premature death worldwide [78-80]. Through its complex, dynamic and reactive mixture of an estimated 7,000 chemicals, tobacco use and its related complications killed almost 6 million people worldwide in 2011, including approximately 600,000 never smokers that died from exposure to second-hand smoke. If this trend continues, the number will increase to 8 million deaths annually by 2030. Worldwide tobacco use already killed 100 million people in the 20th century and will cause one billion deaths in the 21st century, unless urgent action is taken. As tobacco therefore kills more than tuberculosis, human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) and malaria combined, the threat of tobacco smoking to the public health of the world can be called epidemic [81,82].

Within Europe, different stages of the tobacco epidemic can be observed, with Northern countries having a lower percentage of smokers than Mediterranean countries. Furthermore, within the past 20 years the proportion of male smokers has decreased in most countries, whereas the number of female smokers increased. Tobacco smoking leads to the death of more than 650,000 Europeans each year, which equates to one in seven deaths across the EU, and an additional 13 million people suffer from a serious chronic disease as a result of smoking. In the year 2000, for every 1,000 EU citizens who smoke, one was murdered, seven were killed in road crashes, but 500 died from smoking [83].

Cigarette smoking causes a wide spectrum of diseases and affects every organ system in the body [84], as a consequence male and female smokers lose an average of 13.2 and 14.5 years of life, respectively [71]. The most common tobacco smoke related deaths, for active

as well as passive smokers, are due to cardiovascular and respiratory diseases and various types of cancer, in particular lung cancer, through mechanisms that include DNA damage, inflammation, and oxidative stress [71,84,85]. In addition, environmental tobacco smoke also significantly increases the risk to develop these and other diseases [85].

As tobacco smoking is a major cause of various diseases, it is responsible for hundreds of billions of dollars of economic damage worldwide each year [81,82]. The money raised from duties and taxes of tobacco products does not cover all the health costs resulting directly or indirectly from tobacco-related diseases. For example, in the year 2003 in Germany the total healthcare costs directly linked to tobacco use were calculated to be as high as €21 billion [86].

While in high-income countries smoking prevalence is on the decline, in many middle- and low-income countries smoking prevalence is rising, and so most of tobacco-related deaths (nearly 80%) move from developed to developing parts of the world [81,82]. According to the WHO (world health organization), stronger global anti-tobacco measures would be necessary, including tobacco taxes, smoke-free public places, mass media campaigns and effective health warnings. The problem is, that the countries with the most need for anti-tobacco measures have the fewest resources to spend on implementing them [81].

1.3.2. Chemistry of tobacco smoke and biomarkers of harm

Tobacco is the only legal consumer product that kills many of its users when used exactly as intended by manufacturers. The first time a person smokes, the body makes it very clear that it is being poisoned, with lungs that may feel like they are burning and violent coughing. But most people ignore the signs and do not pay attention to their body, addicting it to the toxic components. The chemicals in tobacco smoke reach the lungs quickly when they are inhaled, go rapidly from the lungs into the blood and are then carried to every organ in the body, where they inflame and damage the tissues. It is not only that the poisons in smoke cause tremendous problems to the health and various diseases, beyond that, they disrupt the way the body heals itself [84].

By the early 1950s, mounting scientific evidence began to implicate a role of tobacco smoking in the development of various respiratory and cardiovascular disease as well as cancer. This evidence led to major cigarette design changes focusing on efforts to reduce the overall tar and nicotine yields of cigarettes, starting with using filters. But the changes in cigarette design and manufacturing over the last sixty years did not have a beneficial impact on public health [87].

The cigarette smoke of modern cigarettes contains at least 69 known carcinogens and many other toxicants implicated in major diseases [88]. 599 different cigarette additives, submitted by the five major American cigarette companies, have been approved by the US Government (U.S. Food and Drug Administration). These additives are not only used to adjust the pH, control the burn rate or give a special flavor to the smoke, but are also used by the tobacco industry to cover the harshness of lower-quality tobacco, increase free base nicotine and addiction potential and to mask and treat symptoms. One significant issue is that while these ingredients are approved as additives for foods, they were not tested by burning [84]. But burning actually changes the properties of these substances, creating over 7,000 chemical compounds, of which many are highly toxic and/or carcinogenic. Nicotine, benzene, phenols, polyaromatic hydrocarbons (PAHs), inorganic compounds, and tobacco-specific nitrosamines (TSNAs) are all present in cigarette smoke, just to name a few [88,89]. Other consumer products, which contain these chemicals, have warning labels that make aware of the danger of the poisons in these products, but no such warning exists for theses toxins in tobacco smoke. Figure 5 demonstrates some of the chemicals in tobacco smoke and in which other consumer products they can be found.

Nicotine is one of the most harmful substances in tobacco products, as it is a psychoactive drug that is strongly addictive and many smokers who would like to quit smoking are unable to do so as the result of nicotine dependence [90-92]. Research suggests that nicotine may be as addictive as heroin, cocaine, or alcohol [84,91]. Following abstinence, the nicotine effects disappear quick, leading to nicotine withdrawal syndrome and continuation of smoking [92,93]. The symptoms of nicotine withdrawal include, for example, intense cigarette craving, impaired concentration, nervousness, anxiety, depression, disturbed sleep, tension, headaches, restlessness, and an increased appetite leading to weight gain [92,94]. It has been shown that the severity of these symptoms depends as well on the number of cigarettes smoked daily as on the duration of usage [95].

Tobacco smoking can cause both physical and mental addiction. Nicotine changes the way the brain works, like heroin or cocaine, and causes smokers to crave more and more nicotine leading to physical addiction. These cravings can be very powerful, making it hard to think about anything else and therefore leading to mental addiction. Since decades, tobacco companies know very well that nicotine addiction helps sell their products, and as mentioned before, put additives and chemicals in cigarettes that make them even more addictive [96]. Nowadays, more nicotine is delivered within one cigarette and quicker as ever before, racing from the lungs to the heart and brain. Labels as “filtered,” “low-tar,” or “light” cigarettes are misleading, because they are not less dangerous than others [84].

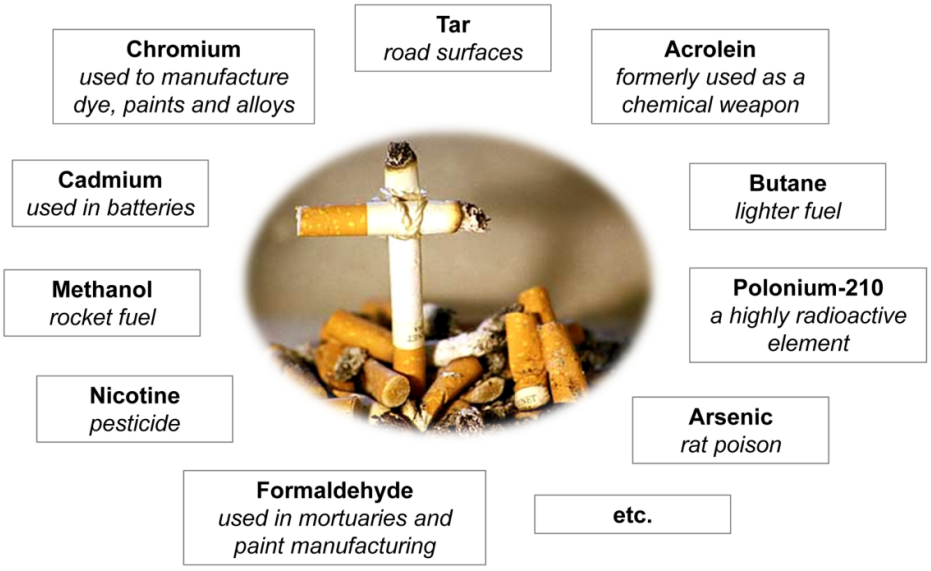


Figure 5. Some of the harmful chemicals found in cigarette smoke. Cigarette smoke contains over 7,000 chemical compounds, of which many are highly toxic and/or carcinogenic. Only a few of them are demonstrated in this figure (information from http://www.cancerresearchuk.org/; picture from http://www.gesundheitlicheaufklaerung.de/).

1.3.3. Former tobacco smoking – the benefits of smoking cessation

As described above, tobacco smoking usually leads to nicotine dependence, which is a chronic condition that often requires repeated interventions and effective treatments in order to overcome the urge to smoke the next cigarette. It is needless to say that quitting smoking is extremely difficult. But in fact, today there are more former smokers than current smokers, so smokers can and do quit smoking [97].

Smoking cessation can significantly reduce the risk of developing smoking-related diseases, and gives the body a chance to heal the damage, even though the impact of smoking on malignant, cardiovascular and respiratory diseases can persist decades after smoking cessation [98-103]. Nevertheless, when smokers quit, heart attack risk drops sharply after just 1 year; the risk to get a stroke is about the same as the one of a never smoker after 2-5 years; after 5 years the risks for cancer of the mouth, throat, esophagus and bladder are cut in half and after 10 years the risk for dying of lung cancer drops by half [84].

It is never too late to stop smoking, which was demonstrated once again by a recent review that found that even the oldest smokers (> 60 years of age) can reap significant benefits from quitting smoking [104], in order to prevent or at least attenuate the development and progression of tobacco-related diseases as will be described within the following chapter.

1.3.4. DNA methylation and tobacco smoking-associated diseases

Cigarette smoke causes a wide spectrum of diseases and affects every organ system in the body through its complex, dynamic and reactive mixture of highly toxic and/or carcinogenic chemical compounds. In general, heavier smoking for a longer duration results in a higher risk of disease, as a strong relationship can be found between the amount of cigarettes smoked by an individual and disease [83,84]. Until the end of 2009, and therefore the beginning of this doctoral thesis, studies which investigated the effect of tobacco smoking on aberrant DNA methylation have been largely restricted to promoter regions of malignancy-related genes and have enrolled small numbers of subjects [105-110]. Only a few studies assessed global DNA methylation levels [111,112] or the effect of smoking on diseases, such as cardiovascular diseases [113].

1.3.4.1. Cancer

Cigarettes contain at least 69 known carcinogens that can lead to cancer by changing the cells instructions or even destroy it [84]. Lung cancer is the leading cause of cancer mortality worldwide and over 85 percent of lung cancers are attributable to smoking [114], making cigarette smoking the major cause of lung cancer. Even though smoking cessation reduces the relative risk of lung and other cancers, former smokers still have an increased risk compared to never smokers, even decades after quitting smoking [84,115].

An important epigenetic pathway, for example, is the enzymatic promoter hypermethylation of genes which can result in gene silencing. The result of this process can be loss of gene transcription and therefore the silencing of gene function. DNA hypermethylation of CGIs in the promoter region of known tumor suppressor genes is a common hallmark of human tumors in general [116] and of lung cancer in particular [117,118]. The most prominent genes for which promoter hypermethylation has been detected are listed in Table 1.

Table 1. Impact of tobacco smoking on promoter DNA methylation in various cancers.

Gene	Function
CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1) [119]	involved in the activation and detoxification of carcinogens 1
P16 (CDKN2A; cyclin-dependent kinase inhibitor 2A) [117,118,120-130]	plays an important role in cell cycle
RASSF1A (Ras association (RalGDS/AF-6) domain family member 1) [125,131-135]	plays an important role in signal transduction pathways that control cell proliferation, differentiation and death
GSTP1 (glutathione S-transferase pi 1) [136,137]	catalyzes glutathione conjugation of tobacco smoke carcinogens
DAPK (death-associated protein kinase) [116,117,125,128,129,133,138-140]	important in ligand-induced programmed cell death
E-CADHERIN (epithelial; also known as CDH1- cadherin 1, type 1) [133,141]	involved in mechanisms regulating cell-cell adhesions, mobility and proliferation of epithelial cells
RAR-β (retinoic acid receptor β) [142,143]	mediates cell differentiation and proliferation
p15 (CDKN2B; cyclin-dependent kinase inhibitor 2B) [144,145]	functions as a cell growth regulator that controls cell cycle G1 progression
MGMT (DNA repair gene O6-methylguanine DNA methyltransferase) [124,125,128,129]	removes adducts from the O6 position of guanine and protects cells from the carcinogenic effects of alkylating agents

Listed are the most studied candidate genes concerning DNA methylation associated with tobacco smoking and various cancers, until the end of 2009.

1.3.4.2. Cardiovascular disease

Cigarette smoking and second-hand exposure to cigarette smoke are major causes of coronary heart disease (CHD), stroke, aortic aneurysm, and peripheral arterial disease (PAD). The risk to develop these cardiovascular diseases (CVD) increases with the number of cigarettes smoked and with the duration of smoking. However, it is a nonlinear dose response, as the risk is substantially increased even by exposure to low levels of cigarette smoke, such as occasional smoking, smoking a few cigarettes per day or exposure to second-hand smoke. The constituents of cigarette smoke believed to be responsible for CVD include nicotine, carbon monoxide (CO), oxidant gases, polycyclic aromatic hydrocarbons (PAHs), and particulate matter. Cigarette smoking leads to endothelial injury and dysfunction in both coronary and peripheral arteries, leads to an increased risk of thrombosis, produces a chronic inflammatory state that contributes to the atherogenic disease processes and elevates levels of biomarkers of inflammation, known to be powerful predictors of cardiovascular events. Smoking cessation reduces the risk of cardiovascular morbidity and mortality [146-150].

Not much is known about aberrant DNA methylation in cardiovascular diseases, especially in the context of tobacco smoking. Patients with coronary artery disease (CAD) showed genomic hypermethylation compared to controls [151]. Global or LINE-1 DNA hypomethylation has been shown in human individuals with atherosclerotic disease both in

atherosclerotic lesions and blood [152,153]. As atherosclerosis is largely an inflammatory disease, circulating factors related to inflammation and inflammation-related endothelial dysfunction are predictors of cardiovascular disease [154]. Vascular Cell Adhesion Molecule-1 (VCAM-1) has been identified as such an inflammatory marker and therefore predictor of cardiovascular disease in humans. In a recent study, LINE-1 element hypomethylation has been associated with higher serum VCAM-1, which might be a mechanisms linking DNA hypomethylation to increased cardiovascular risk [155]. Also, specific methylation changes, usually hypermethylation, have been found in the promoter region of genes associated with atherosclerosis, such as extracellular superoxide dismutase (*SOD*), estrogen receptor α (*ESR1*), endothelial nitric oxide synthase (*NOS*), and 15-lipoxygenase (*ALOX15*) [156-158], but they all have not been studied concerning the effect of tobacco smoking on these changes. One study that analyzed the methylation frequency of the monoamine oxidase-B (*MAOB*) gene promoter in platelets could show a markedly lower methylation in current vs. never smokers due to cigarette smoke-induced increase of nucleic acid demethylases. MAOB plays a role in cardiovascular damage [113]. Strikingly, in an animal model for CVD (apolipoproteinE (apoE) mutant mouse) altered methylation patterns in peripheral blood mononuclear cells (PBMCs) DNA, that were found within coding sequences as well as repeated interspersed sequences, could be detected before any noticeable atherosclerotic lesions were present, suggesting that DNA methylation plays a causative role during the course of CVD development [159].

1.3.4.3. Respiratory and other diseases

The chemicals in tobacco smoke inflame the delicate lining of the lungs and can cause permanent damage that reduces the ability of the lungs to exchange air efficiently and leads to chronic obstructive pulmonary disease (COPD), which includes emphysema and chronic bronchitis [84]. In fact, cigarette smoking is the most important risk factor for obstruction of airflow in COPD. Active and passive smoking increases the risk to develop asthma and worsens the progression [160]. Numerous studies associated *in utero* tobacco smoke exposure as well as smoking of the mother previous to pregnancy to the development of asthma in the child [161,162]. Moreover, it was demonstrated that smoking by the grandmother during the mother's fetal period was associated with increased asthma risk in her grandchildren and the authors speculate that this could be explained by epigenetic transgenerational inheritance [163].

Tobacco smoking reduces a womans chance of getting pregnant, increases the risk of pregnancy complications, premature delivery, low birth weight infants, stillbirth, and sudden

infant death syndrome (SIDS). Smoking might also damage the DNA in men's sperm, leading to decreased fertility, birth defects or miscarriage [84].

Adverse effects of tobacco smoking on type 2 diabetes are generally under-recognized even though various studies have shown that smoking is a risk factor for this disease [164,165]. In a meta-analysis that included 1.2 million participants, active and past smoking was shown to be significantly associated with an increased risk of type 2 diabetes, with active smokers having a 44 percent increased risk compared to never smokers [166]. Furthermore, smoking-related risk of diabetes probably increases with the number of cigarettes smoked, suggesting a dose-response relationship between smoking and diabetes [166,167].

No studies were found that assessed the effect of aberrant DNA methylation on these diseases due to tobacco smoking.

1.4. Aim of the study

1) It has been shown that environmental factors have the potential to change DNA methylation patterns, which may lead to changes in gene expression and the development or progression of disease. There is evidence that tobacco smoking also has an effect on aberrant DNA methylation, but studies have been largely restricted to promoter regions of malignancy-related genes, whereas only a few have assessed global DNA methylation with a small number of subjects. The first aim of the present thesis was to conduct a genome-wide hypothesis free screen for aberrant DNA methylation patterns using the Infinium HumanMethylation 450K BeadChip array and to investigate the effect of current tobacco smoking on the DNA methylome in a relatively large discovery and replication panel.

2) Smoking cessation can significantly reduce the risk of developing smoking-related diseases. Reestablishment of the original DNA methylation patterns due to cessation may be a good explanation for this observation, but the effect of cessation time on DNA methylation, and the time period over which DNA methylation is subject to change, is not known. The second aim of the present thesis was to investigate the effect of former tobacco smoking on the epigenome with the 450K BeadChip array in general and in respect to cessation time and pack-years.

3) The third aim of the present thesis was to establish a platform that enables DNA methylation analysis of selected regions in high resolution (Sequenom EpiTYPER), in order to validate and fine map selected loci detected with the 450K BeadChip array.

A further objective of this doctoral thesis was to participate in various other projects with responsibility for the genome-wide (Illumina 450K) or region-specific (Sequenom EpiTYPER) DNA methylation analysis, or with responsibilty for the genotyping of SNPs (SNP selection, assay design, genotyping by MALDI-TOF MS and genotype calling). This led to the successful publication or in some cases to the submission, of various manuscripts listed in the appendix section on page 182.

2. Methods and Material

2.1. Study population

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. Written informed consent has been given by each participant. The study, including the protocols for subject recruitment and assessment and the informed consent for participants, was reviewed and approved by the local ethical committee (Bayerische Landesärztekammer). The KORA S4 survey was conducted in 1999/2001 and the standardized examinations applied in the survey (4261 participants) have been described in detail elsewhere [168-170]. A total of 3080 subjects participated in a follow-up examination of S4 in 2006-08 (KORA F4), comprising individuals who, at that time, were aged 32–81 years. The KORA F3 cohort is a ten years follow-up survey of the KORA S3 survey examined in 1994–1995 as described previously [170,171]. No evidence of population stratification was found in multiple published analyses using the KORA cohort [172]. The KORA F3 and F4 surveys are completely independent with no overlap of individuals.

Samples choice for the Infinium HumanMethylation450K BeadChip

Within the KORA F4 study, 1814 randomly selected participants were genome-wide genotyped using the Affymetrix GeneChip array 6.0 some years ago. To allow comparison of data gained by different omics-technologies, the same 1814 participants were included in the KORA F4 discovery panel for the 450K DNA methylation analysis. The KORA F3 replication panel included 479 samples that were selected on a case-control basis, including about 50% current smokers as cases (N=240) and about 50% never smokers (N=239) as controls.

Samples choice for the Sequenom EpiTYPER assay

21 current and 21 never smokers, selected out of the 1814 KORA F4 samples used for 450K analysis, were used for the region-specific DNA methylation analysis with the EpiTYPER.

2.1.1. Assessment of smoking status

The category of current smokers comprised regular smokers (smoking daily) and occasional smokers (not smoking daily). The baseline questionnaire included the smoking status (regular/occasional/former/never smoker), the number of cigarettes smoked daily (for regular smokers only), the largest number of cigarettes ever smoked daily for a whole year (for current and past smokers), and the year of beginning and (in case of past smokers) of stopping smoking. Assuming 20 cigarettes per pack, pack-years were calculated using the formula ‘‘(cigarettes per day/20) * number of years smoked’’.

2.2. DNA concentration measurement

Spectrophotometry is the quantitative measurement of the reflection or transmission properties of a material as a function of wavelength. DNA absorbs ultraviolet (UV) light very efficiently. The nitrogenous bases in nucleotides have an absorption maximum at about 260 nm. In contrast to nucleic acids, proteins have a UV absorption maximum of 280 nm, due mostly to the aromatic residues. The ratio of the absorbance at 260 nm and the absorbance at 280 nm gives an estimate of the protein contamination of the DNA sample. Concentration and quality of the DNA samples were measured with a 8-Sample Spectrophotometer (Peqlab, Germany) following the manufacturer`s protocol and using standard settings. Only good-quality DNA with an A260/A280 ratio of 1.8-2.0 was used for further processing.

2.3. Polymerase Chain Reaction (PCR)

With the introduction of the PCR [173], it became possible to exponentially amplify specific DNA regions and thus generate billion copies of the fragment to be analyzed. A PCR reaction starts with a denaturing step where samples are heated to 94°C. Prior to the first cycle, an additional denaturation step of several minutes is used to disentangle the complex structure of the template DNA. For annealing, the temperature is lowered. This allows the excess of primers to anneal to their complementary sequences. The primers are usually only 18 to 25 base pairs long and designed to bracket the DNA region to be amplified. The annealing temperature of this stage depends on the primers and is usually 5°C below their melting temperature. During the following elongation step, the temperature is raised to 72°C,

the optimum temperature of the polymerase. The polymerase attaches at each priming site and uses the 3'OH ends of the primers and the provided deoxynucleotides (dNTPs) to catalyze the synthesis of the new DNA strands. A final elongation step is frequently used after the last cycle to ensure that any remaining single stranded DNA is completely copied. The amplification consists of 30-45 cycles of denaturation, annealing and elongation.

2.4. Agarose gel electrophoresis

Agarose gel electrophoresis is used to separate DNA fragments by size. The size of the separated fragments can be estimated by comparison with known fragments (DNA standards). Nucleic acids migrate to the anode in electric fields due to the natural negative charge carried on their phosphodiester backbone. Shorter fragments move faster than longer ones. 3% agarose gels were made by dissolving agarose in 1x tris-borate-EDTA (TBE) buffer in the microwave and then adding ethidium bromide (0.5 µg/ml), that intercalates with DNA and fluoresces under UV light. 6x Loading Dye (purchased as ready-to-use solution from Fermentas) was added to the DNA solution in order to sediment it in the gel well and the solution then applied to the gel. As a standard, 5 µl of pUC8 100 bp DNA ladder (Fermentas) was used. The DNA fragments were separated in a constant electric field of 120 V and 400 mA for about 40 minutes, on the basis of their molecular weight. After the separation was completed, the DNA bands on the gel were visualized using a UV-transilluminator with a wavelength of λ=254 nm and photographed using a Bio Vision gel documentation system (PeqLab, Germany).

10x TBE buffer: 108 g Tris
55 g Boric acid
9.3 g EDTA pH 8.0
H_2O_{bidest} up to 1 L

2.5. Bisulfite treatment of genomic DNA

2.5.1. Principles

Treatment of denatured single-stranded genomic DNA with sodium bisulfite converts unmethylated cytosines to uracil, whereas methylated cytosine bases are protected from conversion and remain unchanged. The uracils are amplified in subsequent PCR or WGA

(whole genome amplification) reactions as thymines, producing methylation-dependent sequence variations of C to T (Figure 6).

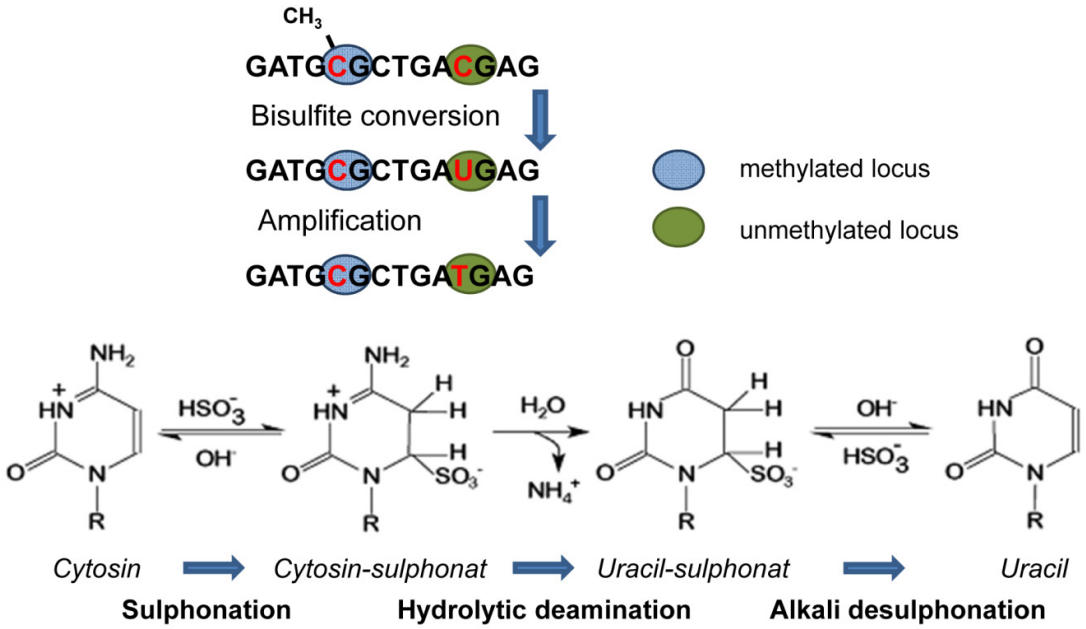


Figure 6. Principle of the bisulfite conversion and chemical background. The chemistry of cytosine deamination by sodium bisulfite involves three steps: Sulphonation (the addition of bisulfite to the 5-6 double bond of cytosine), Hydrolic deamination of the resulting cytosine-bisulfite derivative to give a uracil-bisulfite derivative and Alkli desulphonation (the removal of the sulphonate group by an alkali treatment) to give uracil.

2.5.2. Lab procedure of bisulfite conversion

Protocols were performed according to the manufacturer's instructions (Zymo Research EZ DNA methylation kit) with slight modifications as recommended by Illumina and Sequenom, respectively. At first, a 45 µl DNA mix was prepared in a Conversion Plate (96-well format) for each of the DNA samples. Therefore, a certain amount of genomic DNA - 1000 ng for the Illumina process and 800 ng for the Sequenom process - was added to the respective volume of nanopure water, depending on the DNA concentration. This preparation of the Conversion Plate was done either the day before and then stored at -20°C or right before the next step, in which 5 µl M-Dilution Buffer was added to each DNA mix and mixed by pipetting up and down. The samples were then incubated at 37°C for 15 min. During the incubation time of the samples the CT Conversion Reagent was prepared by adding 750 µl of nanopure water and 210 µl of M-Dilution Buffer to one bottle of CT Conversion Reagent, then the tube was fixed to the vortexer by adhesive tape and mixed with frequent vortexing for at least 10 min at room temperature (RT). As the CT Conversion reagent is light sensitive, the exposure

to light was minimized by darkening the lab room. For best results the prepared CT Conversion Reagent was used immediately after preparation for all experiments described in this thesis (theoretically, if not used immediately, the CT Conversion Reagent solution can be stored overnight at room temperature, one week at 4°C or up to one month at -20°C. Stored CT Conversion Reagent solution must be warmed to 37°C, then vortexed well prior to use). 100 µl of the reagent was added to each sample after incubation, mixed by pipetting up and down and incubated on a thermal cycler with the following alternative cycling protocols.

Alternative cycling protocol for Illumina process

Temperature	Time	
95ºC	30 sec	**16 cycles**
50ºC	60 min	
4ºC	∞	

Alternative cycling protocol for Sequenom process

Temperature	Time	
95ºC	30 sec	**20 cycles**
50ºC	15 min	
4ºC	∞	

A complete bisulfite conversion is essential for the DNA methylation analysis, as incomplete conversion of cytosine to uracil can result in false-positive methylation signals. Incomplete bisulfite conversion can arise from incomplete denaturation before bisulfite treatment or reannealing during the bisulfite conversion, as the conversion only takes place within single stranded DNA. Thus, repeated denaturation cycles were part of the bisulfite conversion process. On the other hand overtreatment with bisulfite degrades DNA and may lead to an increased incidence of methylated cytosines converting to thymine residues, which results in under-reporting of DNA methylation. The reason to choose a shorter cycling protocol for the Sequenom process (about 5.5 hours compared to 16.5 hours for the Illumina process) is that larger fragments are needed for this process, as regions of up to 500 bp have to be amplified by PCR. Bisulfite treatment leads to fragmentation of the DNA, and the longer the incubation time the more fragmentation is likely. The complete bisulfite conversion of the regions analyzed can be checked at a later time. On the contrary, for the Illumina process the longer incubation time should assure that the entire genome has been completely converted, DNA

fragmentation is not such a big issue in the whole genome amplification (WGA) step used for this process.

After the incubation step, the samples were kept on ice for 10 min and the M-Wash Buffer was prepared by adding 144 ml of 100% ethanol to the 36 ml M-Wash Buffer concentrate. The centrifuge was set to a temperature of 25°C.

The following applied for all coming centrifuge steps within the clean-up steps: It was important to make sure that all wells were completely dry after each centrifugation step, but usually a few wells were not dry even after an extended centrifugation time which could be seen by swaying the plate under the bench light. This problem was solved by rotating the plate 180° in between the centrifugation steps, probably due to a slight change of the centrifugal force. Also, the flow-through was discarded after each step to prevent contamination of the column contents.

400 µl of M-Binding Buffer was added to each well of a Silicon-ATM Binding Plate, which was stacked on top of a Collection Plate. The samples, which have been on ice for 10 min, were loaded into the wells of the Silicon-ATM Binding Plate, containing the M-Binding Buffer, and mixed by pipetting up and down as long as smear was observable. The plates were centrifuged for 3 + 3 min at ≥ 3000 x g (RCF 4500). 500 µl of M-Wash-Buffer was added to each well and centrifuged for 3 + 3 min at ≥ 3000 x g (RCF 4500). Then 200 µl of M-Desulphonation Buffer was added to each well, incubated at RT (20°C - 30°C) for 15 - 20 min, and centrifuged for 3 + 3 min at ≥ 3000 x g (RCF 4500). Again 500 µl of M-Wash Buffer was added to each well and centrifuged. This was followed by the addition of another 500 µl of M-Wash Buffer and centrifugation for 6 + 6 min at ≥ 3000 x g (RCF 4500). The Silicon-ATM Binding Plate was then placed onto an Elution Plate and, depending on the subsequent process the DNA was needed for, eluted with 15 µl of M-Elution Buffer (Illumina) or 60 µl of nanopure water (Sequenom). Therefore, both elution fluids were preheated to a temperature of about 50°C by placing them into the hybridization oven about 15 min prior to use, then directly pipetted to the binding matrix in each well, incubated for about 2 min and centrifuged for 3 + 3 min at ≥ 3000 x g (RCF 4500). The bisulfite-converted DNA was used immediately within the Illumina process, and was stored at 4°C overnight or below -20°C for later use within the Sequenom process.

Since the DNA was single stranded due to the conversion process, with limited non-specific base-pairing at room temperature, an absorption coefficient of 260nm, which resembles that of RNA, had to be used to determine the concentration of the recovered bisulfite-treated DNA (value of 40 µg/ml for Abs260=1.0).

2.5.3. Test for complete bisulfite conversion by PCR

In order to assure a complete bisulfite conversion which is essential for the following DNA methylation analysis, as incomplete conversion of cytosine to uracil can result in false-positive methylation signals, a test PCR was carried out with primer pairs specific for bisulfite-converted DNA. These primers cover short sequences that contain many non-CpG cytosines (Cs) which are converted to thymines (Ts) during the bisulfite treatment. The primers are designed so that all Cs within the 1st Primer are replaced by Ts and all guanines (Gs) within the 2nd Primer are replaced by adenines (As), as a consequence, these primers will only bind to fully converted templates. If the PCR products then show a defined, single band on the agarose gel the bisulfite conversion is completed.

1st Primer: 5'-GTTTGGGTTTGGTTTTTGTT-3'
2nd Primer: 5'-ACCCTTTACCAATCTCTTAACTTTC-3'

First, 1 µM PCR primer mix was prepared in a 1.5 ml tube by adding 4 µl of both forward and reverse primers for each assay to 392 µl of nanopure water. The master mix used for the test PCR is shown in Table 2. Thermal cycling was performed on Thermal Cycler C1000™ (BIO-RAD Laboratories, Munich, Germany).

Table 2. PCR cocktail for bisulfite conversion test

Reagent	Volume for single reaction (in µl)	Volume for 96-well plate* (in µl)
ddH_20	1.42	149.96
10x PCR-Buffer (containing 20mM $MgCl_2$)	0.50	52.80
dNTP (25 mM)	0.04	4.22
Primermix (1 µM)	2.0	211.20
PCR enzyme (Sequenom, 5 U/µl)	0.04	4.22
Mastermix volume:	4.00	2496
BT DNA	1.00	422.4
Total volume:	5.00	

* Volumes for a 96-well plate include 10% overhang to account for possible pipetting loss and the negative control; BT DNA: bisulfite treated DNA

PCR cycling protocol: Test for complete bisulfite conversion

Temperature	Time	
95ºC	15 min	
94ºC	30 sec	**45 cycles**
59ºC	30 sec	
72ºC	1 min	
72ºC	5 min	
4ºC	∞	

2.6. Infinium HumanMethylation450K BeadChip

2.6.1. Principles

The Infinium HumanMethylation450K BeadChip® (Illumina, Inc. CA, USA) detects cytosine methylation at CpG sites based on highly multiplexed genotyping of bisulfite-converted genomic DNA. The 450K method makes it possible to assess the methylation status of over 485,000 cytosines distributed over the entire genome. A single BeadChip accommodates 12 samples and 1 µg genomic DNA of each sample is required for the initial bisulfite conversion step prior to performing the automated Infinium Assay. Unmethylated cytosines are chemically deaminated to uracil in the presence of bisulfite, while methylated cytosines are refractory to the effects of bisulfite and remain cytosine.

After bisulfite conversion each sample will be whole-genome amplified (WGA), enzymatically fragmented and applied to the BeadChips. During hybridization, the WGA-DNA molecules anneal to locus-specific DNA oligomers linked to individual bead types. The 450K Methylation BeadChip applies both Infinium I & II assay chemistry technologies (Figure 7). Infinium I assay design employs two bead types per CpG locus, one each for the methylated (C) and unmethylated (T) states. Allele-specific primer annealing is followed by single-base extension using DNP- and Biotin-labeled ddNTPs. Both bead types for the same CpG locus will incorporate the same type of labeled nucleotide, determined by the base preceding the interrogated “C” in the CpG locus, and therefore will be detected in the same colour channel.

Infinium II design uses only one bead type with a unique type of probe allowing detection of both alleles. The methylated and unmethylated signals are generated respectively in the green and the red channels. In both cases the percentage of methylation of a given cytosine

is reported as a ß-value, which is a continuous variable between 0 and 1, corresponding to the ratio of the methylated signal over the sum of the methylated and unmethylated signals.

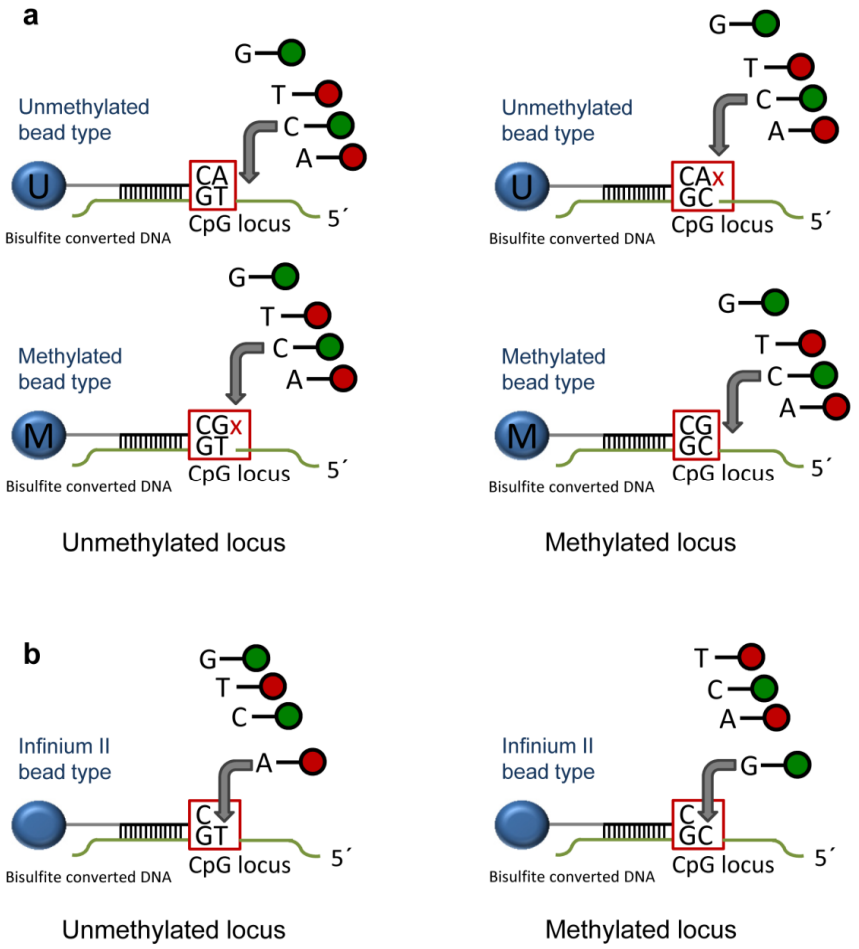


Figure 7. Chemistry technology of the Infinium HumanMethylation450K BeadChip. a) Infinium I assay design employs two bead types per CpG locus, one each for the methylated (C) and unmethylated (T) states, and uses allele-specific primer extension (ASPE). **b)** Infinium II design employs only one bead type with a unique type of probe allowing detection of both alleles and uses single base extension (SBE). Adapted from [174].

2.6.2. Region definition

At the gene level the Infinium HumanMethylation450K BeadChip covers 99% of RefSeq genes with multiple sites in the annotated promoter (TSS200 and TSS1500; 200 bp or 1500 bp upstream of transcription start site), 5' UTR, 1st Exon, gene body and 3' UTR. From the CpG context, the 450K BeadChip covers 96% of CGIs with multiple sites in the annotated CGI, shores (2 kb flanking the CGI) and shelves (2 kb flanking the shores) as well as island-

independent CpGs, which are not necessarily gene-based, depending on their distance to the nearest genes [174]. The usual definition of a CGI within the 450K BeadChip is based on information given by the UCSC Genome Browser, where algorithms were used that search for regions satisfying a definition of CGI proposed by Gardiner-Garden and Frommer in 1987 [34].

2.6.3. Lab procedure of the HumanMethylation450K BeadChip

The following protocol refers to the processing of 96 samples (one 96-well plate; eight 450K BeadChips). Up to three 96-well plates can be processed in parallel. An overview of the lab procedure is given in Figure 8.

Day 1

Making bisulfite-converted DNA (BCD)

Principle: DNA is treated with sodium bisulfite, converting unmethylated cytosines to uracil whereas methylated cytosines remain unchanged, as described above (chapter 2.5.).

Day 2

Denaturing and Neutralizing BCD

Principle: The bisulfite-converted DNA (BCD) is denatured and neutralized, and prepared for amplification.

MA1 (Multi-Sample Amplification 1 Mix), RPM (Random Primer Mix) and MSM (Multi-Sample Amplification Master Mix) were thawed completely at RT. In preparation for the incubation step the hybridization oven was preheated to 37°C. 4 µl of the bisulfite-converted DNA (BCD) were moved into a new 0.8 ml microtiter storage plate (MIDI; referred to as 'sample plate' within this chapter). Then 20 µl MA1 followed by 4 µl 0.1 N N_aOH were added and incubated at RT for 10 min to denature the BCD samples. During the incubation time the sealed sample plate was vortexed at 1600 rpm for 1 min and then centrifuged to 280 x g for 1 min at RT. After the incubation step 68 µl RPM and 75 µl MSM were added and mixed gently by the use of the pipette, then the sample plate was sealed with the 96-well cap mat and inverted at least 10 times to mix the contents. The plate was centrifuged to 280 x g for 1 min and the next step proceeded immediately.

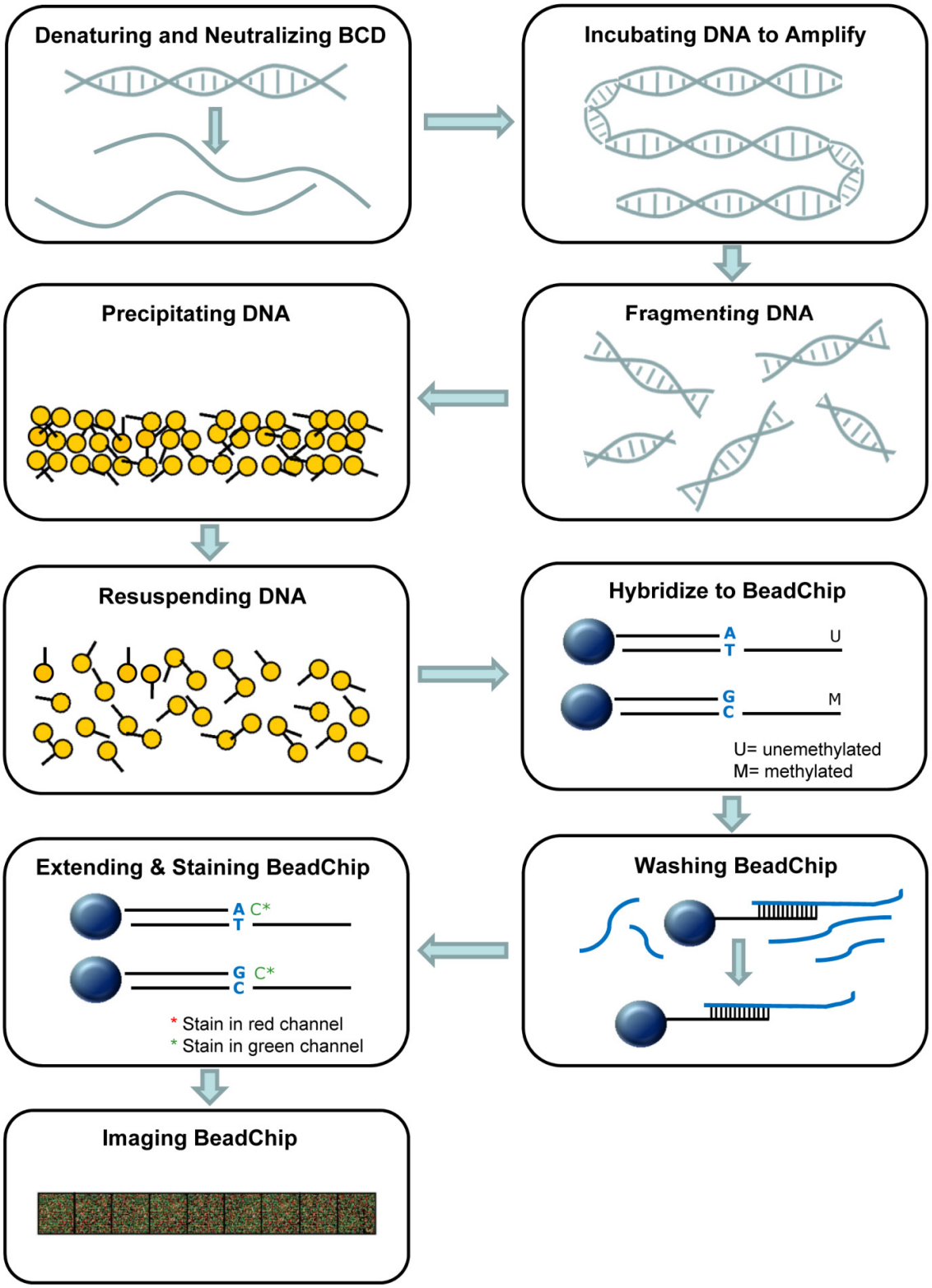


Figure 8. Workflow for processing 450K BeadChips. Adapted from Infinium HD Methylation Assay Guide (part # 15019519).

Incubating DNA to Amplify

Principle: The sample plate is incubated to isothermally amplify the denatured DNA (WGA-whole genome amplification) and generate sufficient quantity of each individual DNA sample in an overnight step. As this amplification process increases the quantity of input DNA by several thousand folds, it is essential that the laboratory spaces for pre-amplification processes (quantification and bisulfite conversion) and post-amplification processes (WGA and following steps) are physically separated.

The sample plate was incubated in the preheated hybridization oven for at least 20 but no more than 24 hours at 37°C. It is possible to freeze the sample plate after the amplification step at -20°C, if necessary.

Day 3

Fragmenting DNA

Principle: The amplified DNA samples are fragmented by a controlled enzymatic process that avoids over-fragmenting the samples and does not require gel electrophoresis.

The heat block with the MIDI plate insert was preheated to 37°C; the FMS (Fragmentation Solution) tubes thawed at RT and gently inverted at least 10 times to mix contents. The sample plate was removed from the hybridization oven and centrifuged at 50 x g for 1 min. 50 µl FMS was added to each sample, the plate was sealed with the cap mat and vortexed at 1500 rpm for 1 min, followed by a centrifugation step at 50 x g for 1 min. The sealed plate was placed on the preheated 37°C heat block and incubated for 1 hour. It is possible to freeze the sample plate after the fragmentation step at -20°C, if necessary.

Precipitating DNA

Principle: The fragmented DNA samples are precipitated by PM1 and 2-propanol and then collected by centrifugation at 4°C.

PM1 (Precipitation Solution) was brought to RT and gently inverted at least 10 times to mix contents. First, the sample plate was removed from the heat block and centrifuged at 50 g for 1 min. Then 100 µl PM1 was added to each sample and the plate was vortexed at 1500 rpm for 1 min followed by a centrifugation step at 50 x g for 1 min. The plate was again incubated at 37°C for 5 minutes and centrifuged to 50 x g. 300 µl 100% 2-propanol was added and the sample plate was carefully sealed with a new, dry cap mat. The plate was inverted at least 10 times to mix contents thoroughly, followed by an incubation step at 4°C (fridge) for 30 min. Meanwhile, the centrifuge was set to 4°C (3,000 x g) in preparation. Following incubation, the plate was centrifuged at 3,000 x g at 4°C for 30 min. The next step had to be performed immediately to avoid dislodging the blue pellet, therefore, the sample plate was instantly

removed from the centrifuge and the cap mat was discarded. The supernatant was decanted over a sink by quickly inverting the sample plate, and the liquid drained onto an absorbent pad. Then the plate was smacked down, avoiding the liquid that was just drained onto the pad, and tapped firmly several times for 1 minute or until all wells were devoid of liquid. During these steps the plate had to be kept inverted at any time and it was important to not allow supernatant in wells to pour into other wells to ensure optimal performance. The uncovered, inverted plate was left on the tube rack for at least 1 hour (2 hours maximum) at RT to air dry the blue pellets, which should be present at the bottoms of the wells. It was important not to over-dry the pellets since otherwise they were difficult to resuspend, and poorly resuspended samples would lead to poor data quality. It is possible to freeze the sample plate at this point at -20°C, if necessary.

Resuspending DNA

Principle: The precipitated DNA samples are resuspended in hybridization buffer (RA1).

RA1 (Resuspension, Hybridization, and Wash Solution) was thawed at room temperature and gently mixed to dissolve any crystals. Also, as it was important to use fresh RA1 for the corresponding protocol steps, only the amount needed for the step in progress was poured out. In preparation the heat sealer was preheated to 165°C and the hybridization oven was preheated to 48°C. 46 µl RA1 was added to each sample and then a foil seal was applied to the plate (mat side down) by firmly holding the heat sealer block down for 4 sec. The sample plate was incubated in the preheated hybridization oven for 1 hour at 48°C, then vortexed at 1700 rpm for 1 min and centrifuged to 280 x g. The blue pellets needed to be completely resuspended, otherwise the plate had to be re-vortexed and centrifuged. The sealed plate can be stored at -20° for no more than 24 hours, or at -80°C for more than 24 hours.

Hybridize to BeadChip

Principle: Twelve fragmented and resuspended WGA-DNA samples are dispensed onto each Infinium HumanMethylation450K BeadChip, counting up to eight BeadChips per 96-well plate, which are then incubated overnight in the hybridization oven to allow the samples to hybridize onto the BeadChips. During hybridization the DNA samples anneal to locus-specific 50mers, which are covalently linked to one of the over 485,000 bead types.

In preparation the heat block was preheated to 95°C and the hybridization oven to 48°C. Eight BeadChips, stored at 4°C, were brought to room temperature. The sample plate was removed from the hybridization oven, vortexed at 1700 rpm for 1 min, and centrifuged at 280x g for 1 min. Then the plate was placed on the heat block to denature the samples at 95°C for 20 min. During the incubation step the Hyb (hybridization) Chambers were prepared

by placing the BeadChip Hyb Chamber gaskets into the BeadChip Hyb Chambers and dispensing 400 µl PB2 (humidifying buffer) into the reservoirs in the Hyb Chambers (PB2 prevents the sample from evaporating during hybridization). The lid was loosely placed on the Hyb Chamber right away to prevent evaporation and the BeadChips were loaded within one hour.

After the 20 min incubation step the sample plate was removed from the heat block, allowed to cool down to RT for 30 min and centrifuged to 280 x g for 1 min. At the robot PC the 'Infinium Robot Control' was opened and a system wash was carried out prior to robot use. All BeadChips were removed from their ziplock bags and mylar packages and placed into the Robot BeadChip Alignment Fixtures (in the Tecan robot) with the barcode end aligned to the ridges on the fixture. The BeadChip were not held by the sides near the sample inlets to avoid contact. The sample plate was placed onto the corresponding robot bed and the robot dispensed the samples to the BeadChips. Afterwards, each BeadChip was carefully removed from the Robot BeadChip Alignment Fixtures and placed in a Hyb Chamber insert (the Hyb Chamber inserts containing the BeadChips were kept steady and level when lifting or moving). The clamps were closed on both sides of the Hyb Chamber in a kitty-corner fashion (closing first the top left clamp, then the bottom right, then the top right followed by the bottom left), so that the lid was secure. The Hyb Chamber was placed into the preheated hybridization oven and incubated at 48°C for at least 16 hours but no more than 24 hours. During this time each sample was hybridized to an individual section of the BeadChip.

In preparation for the XStain protocol 330 ml 100% EtOH was added to XC4 (XStain BeadChip solution 4), and the bottle was vigorously shaken to ensure complete resuspension. XC4 was used at room temperature, but can be stored at 2 to 8°C for two weeks, as the solution allows to process up to 24 BeadChips.

After every run, bleaching was carried out with 0.4% hypochlorite (program 'QC Tasks: Bleach Wash') followed by a system wash for three times.

Day 4

Washing BeadChip

Principle: The BeadChip is prepared for the staining and extension process by washing away unhybridized and non-specifically hybridized DNA.

Before the chips were washed the Tecan robot was prepared for the staining and extension process (see next paragraph).

Each Hyb Chamber was removed from the hybridization oven and allowed to cool down at RT for 25 min prior to opening. The wash rack was submerged in the wash dish containing 200 ml PB1. After removing the cover seal each BeadChip was immediately and carefully

slide into the wash rack, making sure that the BeadChip was completely submerged in the PB1. Once all BeadChips were in the wash rack, the wash rack was moved up and down for 1 minute, breaking the surface of the PB1 with gentle, slow agitation. This step was repeated with a second wash dish containing fresh 200 ml PB1. Next, the Flow-Through Chambers were assembled. The Multi-Sample BeadChip Alignment Fixture was filled with 150 ml PB1 (fresh PB1 had to be used for every additional set of eight BeadChips) and for each BeadChip a black frame was placed into it, followed by the BeadChips. Then a clear spacer was placed onto the top of each BeadChip and the Alignment Bar was placed onto the Alignment Fixture. A clean glass back plate was placed on top of the clear spacer covering each BeadChip and the metal clamps were attached to the Flow-Through Chambers. Finally, the ends of the clear plastic spacers were trimmed from the Flow-Through Chamber assembly using scissors. At the end the Hyb Chamber reservoirs were immediately washed with DiH_2O, ensuring that no PB1 remained.

Extending & Staining BeadChip

Principle: Following hybridization the DNA samples were annealed to locus-specific 50mers and can now be used as templates for the single-base extension of the oligos on the BeadChip, which allows the incorporation of detectable labels on the BeadChip and determination of the methylation level of the query CpG sites.

The extension and staining process was carried out in capillary flow-through chambers and the preparation steps (until the chips were loaded into the Tecan robot) were carried out prior to washing the BeadChips (see previous paragraph).

RA1 was thawed at RT and gently mixed to dissolve any crystals. Also, XC1 (XStain BeadChip solution 1), XC2 (XStain BeadChip solution 2), STM (Superior Two-Color Master Mix), TEM (Two-Color Extension Master Mix) and ATM (Anti-Stain Two-Color Master Mix) were thawed at RT and gently mixed. Next, the Chamber Rack was set up to a temperature of 44°C and bubbles trapped in the Chamber Rack had to be removed each time this process was run. 95% formamide/1 mM EDTA was prepared by adding 23.75 ml formamide and 50 µl EDTA to 1200 µl H_2O in a reservoir. This and two other reservoirs, one containing 30 ml RA1 and the other containing 145 ml XC3, were placed in the reservoir frame, according to the robot bed map. Each reagent tube (XC1, XC2, TEM, STM, and ATM) was placed in the robot tube rack, according to the bed map, and the caps were removed. The program was started and run until the stain temperature had to be entered (32°C) to proceed. When this prompt appeared it was important to wait for the Chamber Rack to reach 44°C and then click OK. The BeadChips were loaded when prompted and the program run for 2 – 3 hours. The RA1 reagent was used to wash away unhybridized and non-specifically

hybridized DNA samples. XC1 and XC2 were added to condition the BeadChip surface for the extension reaction. TEM reagents performed single-base extension of primers hybridized to DNA on the BeadChip, incorporating labeled nucleotides into the extended primers. This reaction was neutralized by the XC3 reagent. 95% formamide/1 mM EDTA was added to remove the hybridized DNA. When the program was finished the BeadChips were removed from the Chamber Rack immediately and horizontally placed on the lab bench waiting to be washed and coated.

For preparation, a top-loading wash dish was filled with 310 ml PB1, another one with XC4. Each Flow-Through Chamber was disassembled one at a time: the two metal clamps were removed with the dismantling tool, then the glass back plate, the spacer and finally the BeadChip were removed. The BeadChips were placed in the staining rack containing the PB1 and had to be completely submerged. The staining rack was slowly moved up and down 10 times, breaking the surface of the reagent, and then the BeadChips were allowed to soak for an additional 5 min (30 min maximum) to be thoroughly washed. Metal clamps, glass back plates and spacer were washed with H_2O and air-dried. Next, 310 ml XC4 were poured into a second wash dish after the XC4 has been completely resuspended by vigorously shaking the bottle, and the staining rack was removed from the PB1 dish into the XC4 dish. The staining rack was again slowly moved up and down 10 times, breaking the surface of the reagent, and then the BeadChips were allowed to soak for an additional 5 min to be thoroughly stained. Afterwards all BeadChips were removed on a prepared tube rack, the tube rack was placed in the vacuum desiccator and the BeadChips dried under vacuum for 45 - 60 min. The underside of the BeadChip was manually cleaned to remove any excess XC4. The BeadChips were then ready to be imaged.

Imaging BeadChip

Principle: A laser is used to excite the fluorophore of the single-base extension product on the beads and the scanner records the high-resolution images of the light emitted from the fluorophores.

The BeadChips were scanned by the Illumina iScan using the ‘Methylation NXT scan’ setting. The Illumina iScan is a two-channel high-resolution laser imager which scans BeadChips at two wavelengths simultaneously and creates an image file for each channel (i.e., two per array). The GenomeScan software determines intensity values for each bead type and creates data files for each channel. GenomeStudio uses these data files in conjunction with the Infinium HumanMethylation450K manifest file to analyze the data from the assay.

2.6.4. The GenomeStudio Module

The GenomeStudio Methylation Module (version 1.8.5) is included in the Illumina Infinium Methylation Assay system and was used for extracting the genome-wide DNA methylation data from the data files gained by the BeadChip scan.
Also, initial quality assessment of assay performance was conducted for all samples processed with the 450K BeadChip using the "Control Dashboard" in the software package. The approach is based on built-in controls that accompany each experiment and includes DNP and Biotin staining, extension, hybridization, target removal, bisulfite conversion I and II, specifity, negative, and non-polymorphic controls. For samples that did not pass this quality control step the complete Infinium HD Methylation protocol was repeated, the ones that then still showed deviations from optimal performance were excluded from further analysis.

2.6.5. Statistical methods for genome-wide analysis

2.6.5.1. Data pre-processing

For data pre-processing of the Infinium HumanMethylation 450K BeadChip the pipeline described in Touleimat & Tost 2012 was used with default parameter settings to avoid bias in the analysis, since the assay combines two different chemistries [175]. Touleimat & Tost implemented the pre-processing, correction and normalization steps in R with Bioconductor packages and adapted some of the steps from earlier R packages, such as lumi [176]. The pre-processing steps that were carried out for the 450K BeadChip raw data of the KORA cohorts by this pipeline are displayed in Figure 9.

The *Quality Control* step estimated the quality of a data set and selected reliable probes and samples. First, probes with signals obtained from less than three functional beads were marked by setting the detection p-value to 1. According to Illumina, the detection p-value is the probability of the probe being not distinguishable above the background (negative controls). Low-confidence probes with a detection p-value > 0.01 were removed (Bead Number filtering). Then the percentage of high-confidence probes was calculated for each sample (Sample QC).

$$\text{Sample-wise callrate} = \frac{\#\ \text{high-confidence probes for sample i}}{\#\ \text{probes in total for sample i}}$$

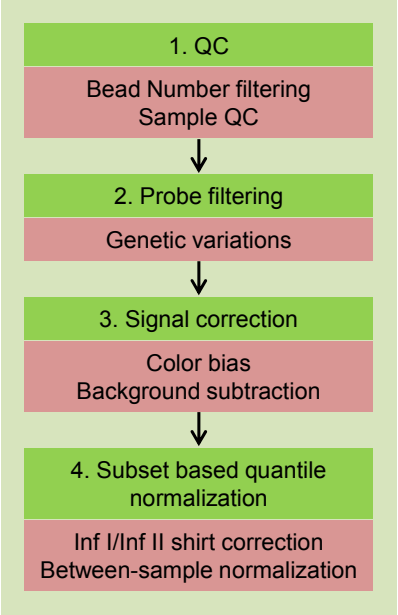


Figure 9. Pre-processing pipeline by Touleimat & Tost (2012).

Within the *Probe Filtering* step CpG sites located in close proximity (50bp) to common single nucleotide polymorphisms (SNPs) with a minor allele frequency of at least 5% were removed, to avoid that the observed variation of the methylation level might be due to genetic variations.

The *Signal Correction* step adjusted potential color-bias which might be present due to imbalances of the two color channels, using smooth quantile normalization (adapted by the R package lumi). For correction of the background level a background estimation approach based on the negative control probes provided by Illumina was carried out and the background noise was subtracted.

The S*ubset based Quantile Normalization* step corrected the InfI/InfII shift which may have induced a bias in the analysis, as the methylation signals corresponding to the two types of assays were analyzed together. Within this step, also a between sample normalization was performed to eliminate non-biological variation between the samples.

CpG-categories were built using the 'relation to CpG island' information (South shore, South shelf, North shore, North shelf, and Open sea) from the Illumina file. Computational challenges, such as memory capacity and computation time, which occurred due to the large amount of data, were solved as followed: During chip-by-chip normalization the RAM was cleaned after each step and only relevant data was saved in portions on the physical drive as

".RData" files, which are faster to read in. For correction of the InfI/InfII shift all .RData files were read in after memory clearance and quantile normalization across all individuals were performed. For data analysis the normalized data was split into chunks and analyzed in parallel.

2.6.5.2. Beta- versus M-value

Beta-value (ß-value) and M-value can be used for DNA methylation analyses, and both have certain strengths and limitations [177]. ß-value is the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities) and allows biological interpretation as approximate percentage of methylation of a CpG site in a sample. However, a study of Du and colleagues could show that ß-values show significant heteroscedasticity (= heterogeneity of variance) in the low and high methylation range, which might be a problem as classical statistical tests assume a normal distribution of outcome. They compared the ß-value method with the M-value method. M-value is calculated as the $\log_2$ ratio of the intensities of methylated probe versus unmethylated probe and is more statistically valid as it is approximately homoscedastic. But, M-value statistic does not have an intuitive biological meaning. According to Du et al., the problem can be solved by using the M-value for conducting differential methylation analysis and including the ß-value statistics when reporting the results [177]. Therefore, the ß-values generated by the Tost pipeline were transformed to M-values.

$$\text{M-value} = \log_2 [\beta\text{-value}/(1- \beta\text{-value})]$$

The final statistical analyses were carried out with M-values, and the descriptive ß-values were included to display the approximate percentage of methylation of the corresponding CpG site.

2.6.5.3. Choosing appropriate covariates

Significant associations of DNA methylation with age and age-related disease have been demonstrated in various studies [178-180] and in general DNA methylation declines with age [181]; therefore, all statistical models were corrected for age. Studies that addressed the question of gender-specific DNA methylation differences showed that men and women have quite similar patterns within the autosomes, nevertheless differences could be detected [182-184], thus all statistical models were corrected for sex. Also, a correlation of DNA methylation with body mass index (BMI) [185,186] and alcohol drinking [187,188] was suggested and both included as covariates. Numerous studies have found increased leukocyte blood counts in current compared to never smokers [189-194], and of up to nearly 30% higher in heavy

smokers who inhale [192]. Some studies have reported that leukocyte count decreases after cessation of smoking [189,191,195,196]. This observation could be confirmed within this study by comparing the numbers of white blood cells (WBCs) between current, former and never smokers (please see chapter 3.1.2.).

As 161 of the 1802 probands in KORA F4 had a history of cancer it was tested if they could influence the results concerning smoking, since it is well known that DNA methylation levels change in a variety of genes in people with cancer. No significant impact on the methylation differences between current and never smokers could be determined by including or excluding probands with a history of cancer, therefore they remained in the data set.

Summarized, at the end the statistical models were not only adjusted for the classic potential confounders of DNA methylation, such as sex and age, but also for alcohol consumption, BMI and WBC.

2.6.5.4. Data analysis

Associations between smoking and methylation M-values were analyzed using multivariable linear regression. A particular methylation M-value was the response variable, with smoking status being the explanatory variable and sex, age, BMI, alcohol consumption as well as white blood cell count as covariates. Analyses of current vs. never smokers as well as of former vs. never smokers were performed by means of smoking status coded as a factor variable with three levels. Also, an interaction model with sex was calculated, where the interaction of the smoking factor variable with sex was included in the latter model. Besides this, the stratified analyses were calculated for males and females separately. In addition to the earlier described covariates (except for sex), this model was also adjusted for pack-years, due to the significant difference of this variable in males and females.

In addition, linear models that included former smokers only were calculated with the metric explanatory variables pack-years and/or time since quitting instead of smoking status. The methylation level in former smokers at the majority of CpG sites approached the corresponding level of never smokers within increasing time since quitting, starting approximately from the level of current smokers for those who only recently quit smoking. Therefore, a smooth loess curve (smooth factor = 0.5) was plotted in a scatterplot of methylation (ß-value, only former smokers) and time since quitting, in order to visualize which impact years or decades of cessation might have on DNA methylation. The descriptive median methylation ß-values of current and never smokers were also displayed as a brown respective green line. These plots were used to get an idea of the time since quitting at which the methylation state of former smokers is closer to or equals the one of the original median difference between current and never smokers. The same procedure was carried out with

pack-years to get an idea of the influence of cumulative smoke exposure on the methylation state of former smokers.

The explained variance was calculated in the linear model from the ANOVA table, taking the deviance of the variable (e.g. smoking, pack-years) divided by the null-deviance (i.e. residual deviance in the model without covariates). In calculating the explained variance of smoking, a two-stage-variable was used (never and current respective never and former).

Methylation β-values were used for the presentation of the scatterplots, since they allow for a straightforward interpretation of the results. In the linear models with covariates the M-values were used, since they showed better statistical ability (2.6.5.2.). The assumption of a normal distribution was verified for all CpG sites that showed a significant result using density plots of the residuals obtained from the multivariable linear regression as well as corresponding QQ plots. All significantly associated sites showed approximately normal distributed residuals except for cg23576855 which was therefore excluded from further analysis.

Regarding the discovery sample (KORA F4), the global significance level of 5% was corrected for multiple comparisons of CpG sites with smoking status, following the Bonferroni procedure (0.05/468316 = approx. p= 1E-07). In the replication sample (KORA F3) the correction was made for the number of significant CpG sites in the discovery sample (0.05/972 = approx. p= 5E-05).

All analyses were performed using the statistical package R Version 2.14 (http://www.r-project.org/), including the packages: base, datasets, graphics, grDevices, methods, stats and utils. The meta-analysis for KORA F4 and F3 was performed with the software METAL (http://www.sph.umich.edu/csg/abecasis/Metal/; release 2011-03-25) with cohorts weighted by their sample size.

2.7. EpiTYPER™ methylation assay (Sequenom)

2.7.1. Principles

To be able to validate and analyze the top hits found on the 450K BeadChip with a different method, a new approach - the EpiTYPER method (Sequenom) - for region-specific high-throughput DNA methylation analysis was established in our lab. The EpiTYPER method is a tool for the detection and quantitative analysis of DNA methylation using base-specific cleavage of bisulfite-converted genomic DNA and Matrix-Assisted Laser Desorption/

Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) [197]. An overview of the concept of the EpiTYPER method is given in Figure 10.

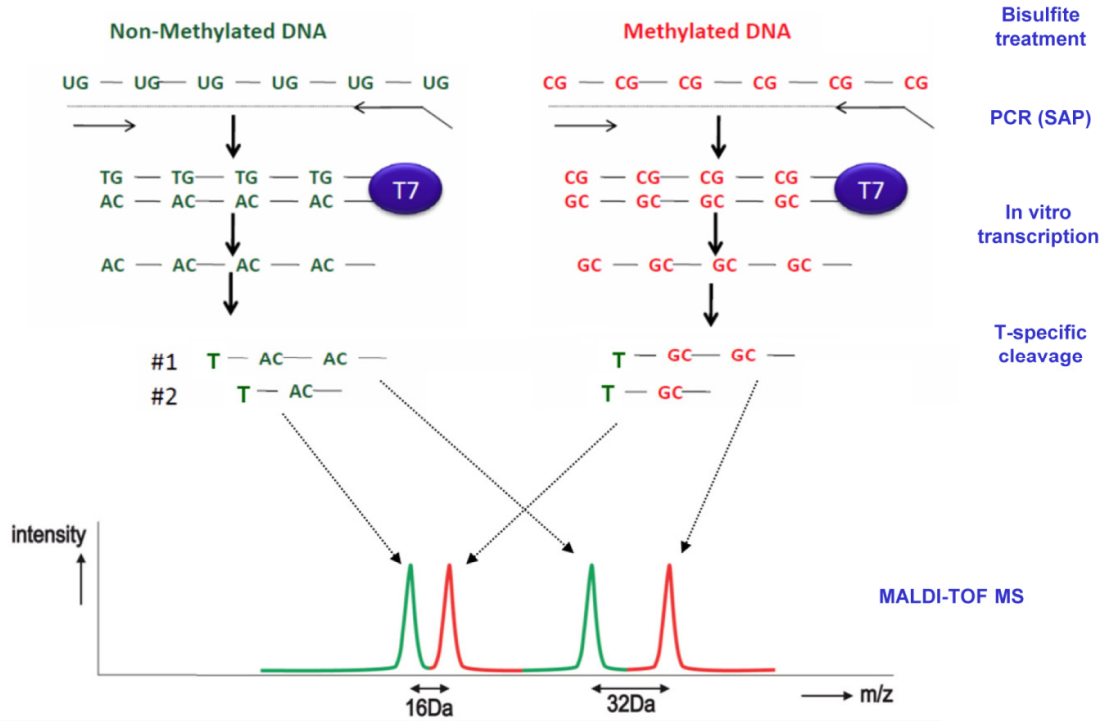


Figure 10. Schematic of the EpiTYPER technique. Bisulfite treated DNA is amplified by PCR using primers located around the CpG site of interest, one of them being tagged with a T7-promoter sequence. The PCR product is transcribed into a RNA transcript and cleaved in a base-specific manner after each T (thymin). The cleavage products are then analyzed by matrix-assisted laser desorption/ionization – time of flight (MALDI-TOF) mass spectrometry (MS) and a characteristic mass signal pattern can be obtained. The green signal represents non-methylated DNA and the red one represents methylated DNA. This figure was adapted from [197].

The method starts with bisulfite treatment of genomic DNA followed by PCR amplification in which a T7- promoter tag is introduced by a T7-promoter tagged reverse primer. Also, a 10mer tag sequence is added to the forward primer to balance the PCR primer length. The PCR primers should be designed to yield a product within a 200-550 bp range. The PCR primers are independent of the methylation state of the genomic DNA, meaning they bind to both methylated and non-methylated template. After the SAP reaction, that removes unincorporated dNTPs, in vitro RNA transcription is performed on the reverse strand followed by base-specific T-cleavage, in which the reverse strand is cleaved by RNase A. MALDI-TOF MS analyzes the mixture of cleavage products differing in length and mass. The methylation dependent C/T sequence changes introduced by bisulfite treatment are presented as G/A changes on the reverse strand and therefore result in a mass difference of 16 Da (Dalton) for each CpG site enclosed in the cleavage products generated from the RNA transcript. A

distinct signal pair pattern results from the mass signals representing methylated and non-methylated template DNA, which is representative for the CpG sites within the analyzed sequence substring. The relative amount of methylated DNA can be calculated by the ratio of these signal intensities and quantitative results can be obtained for each sequence-defined CpG unit, which contains either one individual CpG site or an aggregate of subsequent CpG sites. Depended on the sequence content and the distribution of CpGs, up to 85% of the CpGs within a target region are analyzed.

2.7.2. Lab procedure of EpiTYPER^{TM} methylation assay

2.7.2.1. Bisulfite treatment

DNA was treated with sodium bisulfite, converting unmethylated cytosines to uracil whereas methylated cytosines remained unchanged, as described above (chapter 2.5.).

2.7.2.2. Primer design

Primers for the five target regions (cg05575921-*AHRR*, cg21566642-*ALPP/ALPP2*, cg03636183-*F2RL3*, cg06126421-not annotated, cg19572487-*RARA*) were designed using the Sequenom Primer Design software (www.epidesigner.com) with the following settings:

Setting	Minimum	Optimum	Maximum
Primer T_m (^oC)	52	56	60
Primer size (bp)	20	25	30
Product size (bp)	100	300	500
Primer non-CpG 'C's	4		

Primers were designed to contain at least 4 non-CpG ‘C’s, in order to allow selective amplification of fully converted templates. The T7-promoter tag and 10mer tag were added automatically by the software upon export of the designed PCR primers. The target regions were amplified using the primer pairs and annealing temperatures (T_a) described in Table 3.
An overview of the target sequence and the covered CpG sites is exemplarily shown for the amplicon of the *AHRR* gene, including the 450K site cg05575921, in Figure 11.

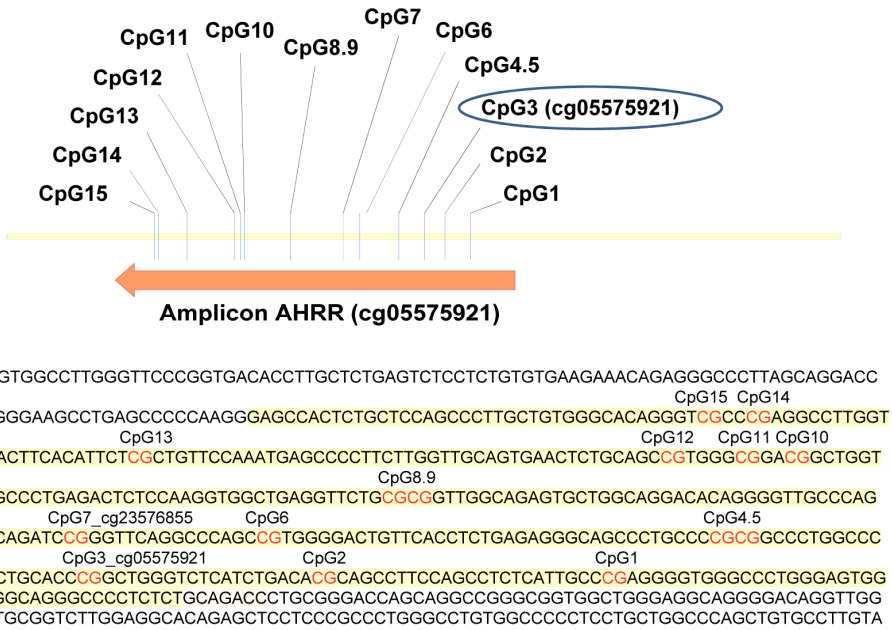


Figure 11. Target sequence for the AHRR amplicon (cg05575921). The target sequence (marked in yellow) and its covered CpG sites (written in red) of the AHRR amplicon (cg05575921) is displayed.

2.7.2.3. Polymerase Chain Reaction (PCR)

First,1 µM PCR primer mixes were prepared in 1.5 ml tubes by adding 4 µl of both forward and reverse primers to 392 µl of nanopure water for each assay. To obtain a defined, single band on the agarose gel (without any additional non-specific bands) and thus the ideal annealing temperature of the primers, PCR incubation conditions were optimized by test PCR. Therefore, PCR cocktails were prepared in 1.5 ml tubes according to Table 4 and a Gradient-PCR (52 - 62°C) was run to determine the optimum annealing temperature. If primers did not show good results within the test PCR, they were re-designed and tested again. DNA used for this test process was bisulfite treated prior to the DNA samples that were used for the actual later experiment. This was done to assure that the actual experiment was carried out on fresh bisulfite treated DNA. The final annealing temperatures (T_a) used are also displayed in Table 3.

The bisulfite-converted DNA of 21 current and 21 never smokers of KORA F4 (see chapter 2.1.) was used for analyzing the DNA methylation status of the target regions. In addition, 0%, 25%, 75%, and 100% methylation standards (EpigeneDX) were included within each run as an indicator of the intra-run variation in the degree of bisulfite-related bias. To not waste material and chips on a half-full plate, the 384-well plate was shared with another project (not shown).

Table 3. Primers used for EpiTYPER methylation analysis.

CpG	UCSC RefGene	Primer	Sequence	Product size (bp)	T_a (°C)	Total No. of CpG sites in amplicon	No. of analyzed CpG sites per amplicon*	
							Single	Composite
cg05575921	AHRR	forward	aggaagagagAGAGAGGGGTTTTGTTAGGATTATT	389	56	15	6	8
		revers	cagtaatacgactcactatagggagaaggctAAACCACTCTACTCCAACCCTTACT					
cg21566642	ALPPL2[b]	forward	aggaagagagGAGGATAGGGGAGAAGATTTTATTG	276	56	18	4	6
		revers	cagtaatacgactcactatagggagaaggctTACAACCCTAAACCACCTAAACTAT					
cg03636183	F2RL3	forward	aggaagagagTTGTGTTAATGATAGTGATATTTTGGAG	452	56	30	7	10
		revers	cagtaatacgactcactatagggagaaggctCCATAAACCAAACAACCATACAAA					
cg06126421	x[a]	forward	aggaagagagAGGGTTAGGGAGTTGTAGATAGGTT	379	56	5	3	2
		revers	cagtaatacgactcactatagggagaaggctCAACTCCAATATCAACCATTAAAAAA					
cg19572487	RARA	forward	aggaagagagATAGGGTTTTATGAAGAGGTTGTGT	259	56	4	2	2
		revers	cagtaatacgactcactatagggagaaggctCAAACAATACTAATATCCCAACTCC					

Listed are the sequences of PCR tagged primers used for EpiTYPER methylation analysis; product size, annealing temperature (T_a) and informative CpG sites of each amplicon. *The total No. of informative CpG sites in the amplicon is divided into single CpG sites and composite sites (two or more adjacent CpG sites fall within one fragment, or when fragment masses are overlapping); [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG island) have a predicted function on the ALPPL2 gene, which is located several kb apart from this CpG island.

PCR cocktails were prepared as displayed in Table 4 without adding the DNA.

Table 4. PCR cocktail for EpiTYPER methylation analysis.

Reagent	Volume for single reaction (in µl)	Volume for 384-well plate* (in µl)
ddH_20	1.42	654.34
10x PCR-Buffer (containing 20mM $MgCl_2$)	0.50	230.40
dNTP (25 mM)	0.04	18.43
PCR enzyme (Sequenom, 5 U/µl)	0.04	18.43
Mastermix volume:	4.00	1843.20
Primermix (1 µM) (T7Reverse & 10mer Forward Tags)	2.00	
BT DNA	1.00	
Total volume:	5.00	

* Volumes for a 384-well plate include 20% overhang to account for possible pipetting loss and the negative control;
BT DNA: bisulfite treated DNA

The preparation of the 384-well plate for PCR was done on ice, as depending on the number of amplicons and samples within one plate, the pipetting process was quite time-consuming. 2 µl of the PCR cocktail were dispensed into the wells of a 384-well plate, sealed and centrifuged. Then 2 µl of the PCR primer mixes were dispensed to each well as per the layout in Figure 12a, sealed and centrifuged. Finally, 1 µl of the bisulfite-converted DNA (10 ng/µl) was dispensed to each well as per layout in Figure 12b, sealed with a plate-sealing film and centrifuged at 560 x g for 1 min.

a **Amplicons**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
B	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
C	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
D	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
E	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
F	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
G	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
H	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
I	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
J	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
K	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
L	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
M	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
N	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
O	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5
P	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5

Amplicon 1: cg05575921 Amplicon 2: cg21566642 Amplicon 3: cg03636183 Amplicon 4: cg06126421 Amplicon 5: cg19572487

b **Samples**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	1	17	33	1	17	33	1	17	33	1	17	33	1	17	33
B	2	18	34	2	18	34	2	18	34	2	18	34	2	18	34
C	3	19	35	3	19	35	3	19	35	3	19	35	3	19	35
D	4	20	36	4	20	36	4	20	36	4	20	36	4	20	36
E	5	21	37	5	21	37	5	21	37	5	21	37	5	21	37
F	6	22	38	6	22	38	6	22	38	6	22	38	6	22	38
G	7	23	39	7	23	39	7	23	39	7	23	39	7	23	39
H	8	24	40	8	24	40	8	24	40	8	24	40	8	24	40
I	9	25	41	9	25	41	9	25	41	9	25	41	9	25	41
J	10	26	42	10	26	42	10	26	42	10	26	42	10	26	42
K	11	27	0%	11	27	0%	11	27	0%	11	27	0%	11	27	0%
L	12	28	25%	12	28	25%	12	28	25%	12	28	25%	12	28	25%
M	13	29	50%	13	29	50%	13	29	50%	13	29	50%	13	29	50%
N	14	30	75%	14	30	75%	14	30	75%	14	30	75%	14	30	75%
O	15	31	100%	15	31	100%	15	31	100%	15	31	100%	15	31	100%
P	16	32	NTC	16	32	NTC	16	32	NTC	16	32	NTC	16	32	NTC

0%: unmethylated DNA; 25 - 100%: methylated DNA; NTC: Negative control

Figure 12. Amplicon a) and sample b) layout for the preparation of the PCR plate.

The 384-well plate was incubated on a thermal cycler using the following program:

PCR cycling protocol

Temperature	Time	
94ºC	4 min	
94ºC	20 sec	**45 cycles**
*56ºC	30 sec	
72ºC	1 min	
72ºC	3 min	
4ºC	∞	

* Annealing temperature adjusted according to Table 3.

After the incubation step, a few samples of each amplicon were randomly selected and run on a 3% agarose gel; please see chapter 2.4.

2.7.2.4. SAP reaction

To reduce the presence of by-products, the amplification products were treated with shrimp alkaline phosphatase (SAP) which cleaves phosphates of unincorporated dNTPs, rendering

them unavailable for future polymerase reactions. The SAP enzyme solution was prepared in a 1.5 ml tube according to Table 5.

Table 5: SAP enzyme solution.

Reagent	Volume for single reaction (in µl)	Volume for 384-well plate* (in µl)
RNase-free ddH_2O	1.70	783.36
Shrimp alkaline phosphatase (SAP)	0.30	138.24
Total volume	2.00	921.60

* Volumes for a 384-well microtiter plate include 20% overhang to account for possible pipetting loss

The SAP enzyme solution was vortexed to thoroughly mix the solution and shortly centrifuged. 85 µl of the solution was pipetted into each well of row H of a new 96-well plate, then 10 µl were drawn from the wells in row H and distributed to each well in rows A-G using a twelve-channel pipette. As the SAP enzyme solution was moderately viscous, this had to be done with care to minimize loss of solution due to adhesion to the pipette tips. The 384-well sample plate and the 96-well SAP solution plate were centrifuged at 560 x g for 1 min, then 2 µl SAP mix was dispensed to each well of the sample plate. After the plate was sealed with plate sealing film, vortexed and centrifuged at 3,000 x g for 1 min, it was incubated on a thermal cycler using the following program:

SAP cycling protocol

Temperature	Time
37ºC	20 min
85ºC	5 min
4ºC	∞

2.7.2.5. MassCLEAVE reaction

Next, *in vitro* RNA transcription was performed on the reverse strand, followed by base specific cleavage using RNase A which yielded fragmented RNA molecules. Within this step, cleavage products were generated for the reverse transcription reaction for U (thymidine (T)). Since T is found in DNA it is used from this point forward, however, cleavage actually happens at U in an RNA molecule. Both methylated and non-methylated regions were cleaved at every T to produce fragments and the cleavage products resulting from methylated and non-methylated DNA respectively had the same length and differed only in their nucleotide composition. Later on, MALDI-TOF MS will analyze the cleavage products,

and a distinct signal pair pattern will result from the methylated and non-methylated template DNA. The T-cleavage transcription/RNase A cocktail was prepared according to Table 6.

Table 6: T-cleavage transcription/RNase cocktail.

Reagent	Volume for single reaction (in µl)	Volume for 384-well plate* (in µl)
RNase free- ddH_20	3.21	1479.16
5x T7-polymerase buffer	0.89	410.11
T-cleavage mix	0.22	101.38
DTT, 100mM	0.22	101.38
T7-RNA & DNA polymerase	0.40	184.32
RNase A (10 mg/ml)	0.06	27.65
Total volume:	5.00	2304.00

* Volumes for a 384-well microtiter plate include 20% overhang to account for possible pipetting loss

2 µl of each PCR product of the 384-well sample plate were transferred 'one to one' into a new 384-well plate, followed by 5 µl of the MassCLEAVE mix per well. The sample plate was then sealed, vortexed, centrifuged and incubated on a thermal cycler using the following program:

T-cleavage cycling protocol

Temperature	Time
37ºC	3 hours
4ºC	∞

Upon completion, the sample plate was either conditioned immediately (next step) or stored at -20°C.

2.7.2.6. Conditioning

After the MassCLEAVE reaction a cation exchange resin was added to the products (Clean Resin; SEQUENOM, San Diego, CA) to remove residual salt that could interfere with mass spectrometry analyses. For this process 6 mg/well Clean Resin was spread out on a 384-well Clean Resin plate (also called dimple plate) with the help of a Clean Resin spoon and scraper, and left aside until dried (about 15 to 25 min). Meanwhile, the 384-well sample plate was centrifuged at 540 x g for 1 min, and the samples were diluted with 20 µl of nanopure water. Then the plate was sealed and again centrifuged at 540 x g for 1 min. The Clean Resin was added to the products by gently placing the 384-well sample plate, upside-down,

onto the Clean Resin plate. The two plates were flipped over and the Clean Resin plate was tapped to help the Clean Resin fall into the sample plate. The plate was rotated for about 1 hour at RT using a rotator, followed by a centrifugation step at 3,200 x g for 5 min. The reaction products were then ready for transfer to a SpectroCHIP®, which was done right away (alternatively, the sample plate can be tightly sealed by an adhesive sealing foil and stored at -20ºC until ready to do so).

2.7.2.7. Nanodispensing

The reaction products were transferred to a SpectroCHIP® using the MassARRAY® Nanodispenser after a volume check was performed to figure out which dispense speed was required to optimize the amount of product spotted. This step was necessary, since the volume deposited on the chip is affected by the low viscosity of the reactions. The volume check led to a dispense speed of 90 mm/sec for the deposition of an optimal volume of 20 to 25 nl. In addition to the 384 samples a 4-point calibrant, a mix of four oligonucleotides with known masses, was dispensed.

2.7.2.8. MALDI-TOF-MS measurements

The Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS) was originally established for the genotyping of SNPs (Single Nucleotide Polymorphism) in our lab, but also allows measurement of the cleavage products of the EpiTYPER assay, giving a distinct signal pair pattern that results from the methylated and non-methylated template DNA [197]. During MALDI-TOF mass spectrometry, the sample is staggered with a 100 to 1,000 fold excess of matrix, co-crystallized on a sample plate and irradiated with an intensive laser pulse for a few nanoseconds in the high vacuum chamber of the mass spectrometer [198,199]. The SpectroCHIP® matrix consists of 3-hydroxypicolinic acid, which absorbs the applied laser energy and induces the ionization of the analyte molecule. Additionally, the matrix prevents a photolytic damage of the analyte and avoids interaction of analyte molecules with each other or with the sample carrier [200].

The SpectroCHIP® was loaded into a MassARRAY mass spectrometer (Bruker–Sequenom) and ran on a MassARRAY Workstation Compact, with settings for MassCLEAVE (MassCleave.par) for both FlexControl and SpectroAcquire. After loading, the chip was transferred to a metallic sample carrier and put in the vacuum lock of the MassARRAY™ MS. Before starting the run, the 4-point calibrant was manually controlled on stage position “F0”. The solution has four standardized spectra mass signals for calibrating the analysis system (in Da) at 1479.0, 3004.0, 5044.4, and 8486.6, which should not be shifted for more than 2 to 3 Daltons. The spectra mass signals were in place for all the experiments carried out.

The transfer of laser energy to sample molecules in the matrix generates mainly single charged molecule ions that trespass into the gas phase. Under high vacuum conditions, the matrix crystals are irradiated with nanosecond duration laser pulses, leading to formation of a plume of volatilized matrix and analyte as well as charge transfer from matrix ions to analyte molecules. After electric field-induced acceleration in the mass spectrometer source region, the gas phase ions travel through a field-free region at a velocity inversely proportional to their mass-to-charge ratios (m/z), until they hit the detector. Ions with low m/z values are faster than ions with higher m/z values and reach the detector earlier. The TOF-analyzer measures exactly the time until the ions hit the detector [200,201].
The resulting time-resolved spectrum is translated into a mass spectrum upon calibration. These mass spectra were further processed and analyzed by the software EpiTYPER® that performs baseline correction, peak identification and quality assessments.

2.7.3. Statistical methods for EpiTYPER™ methylation analysis

2.7.3.1. Test for complete bisulfite conversion by PCR

It is important to test for complete bisulfite conversion, as incomplete conversion of cytosine to uracil can result in false-positive methylation signals. This can theoretically be done by a test PCR as described in chapter 2.5.3. Unfortunately, within the 450K BeadChip analysis it became apparent that a test PCR covering only a small region of the genome is not really trustworthy, as even though the test PCR shows good results other parts of the genome than the one covered by the test primers might still lack complete conversion (see chapter 3.1.1.1. in the result section). So, for the EpiTYPER methylation analysis the pipeline of Thompson et al. was used [202], which offers a bisulfite conversion efficiency calculation by measuring levels of unconverted non-CpG cytosines in a given sample using R [202]. As cytosine methylation in mammals usually occurs in the context of CpG, unmethylated cytosines outside this context can be used to measure bisulfite treatment efficiency, acting as ‘conversion controls’. It is essential to measure the extent of bisulfite conversion, because even though the primers are designed to enrich for completely converted sequences (they contain at least 4 non-CpG ‘C’s), some proportion of amplicons may contain remnant unconverted cytosines which may cause a molecular weight shift in the subsequent fragmentation profile.

2.7.3.2. Data analysis

DNA methylation values were generated as ß-values, determined by comparing the signal intensities between the mass signals of methylated and non-methylated templates. The ß-values were then transformed into M-values for statistical analysis:

$$M\text{-value} = \log_2 [ß\text{-value} / (1 - ß\text{-value})]$$

Association with smoking status was assessed by linear regression using M-values as the response variable, smoking status as the explanatory variable and sex, age, BMI, alcohol consumption as well as WBC count as covariates. Statistical analysis was carried out by R 2.14 (http://www.r-project.org/).

2.7.3.3. SNP prediction analysis

SNPs can hamper correct quantification of methylation status at one or more CpG sites and may therefore complicate interpretation of results. Each novel peak among the MassArray spectrum can be explained by any number of potential SNPs. By using an exhaustive string substitution approach as implemented in the R package "MassArray", putative SNPs can be identified by comparing expected and observed data. The putative nucleotide sequences underlying novel peaks are identified by the EpiTYPER software and can be used for comparison with the original input sequence. The applied algorithm substitutes each base pair at a time in the original input sequence with the other three remaining bases or a gap, i.e., a deletion, and then assesses the ability of the altered input sequence to explain the observed fragmentation pattern due to new peaks. This is done by fragmentation of the altered input sequence and finding base compositional matches to the putative base pair composition of the new peak. Once these new peaks are mapped to the appropriate fragments, the expected peaks corresponding to these fragments are analyzed in order to determine if they are missing or if there is a diminished signal-to-noise ratio (SNR), which is done by comparison of the expected peak SNR to the average SNR of the sample. The SNR of the novel peak is also compared to the average SNR of the sample and granted more reliability if it exceeds this average. Finally, the SNPs quality is then calculated as function of new peak SNR and expected peak SNR.

2.8. Functional analysis by EMSA

Electrophoretic mobility shift assays (EMSA) have been used to assess the potential biological relevance of DNA methylation differences caused by tobacco smoking for the most significant CpG site (cg05575921 – AHRR). THP1 (human monocytic cell line) and Raji (human B-lymphoblastoid cell line) nuclear extracts were purchased from Active Motif. Cy5-labelled and unlabeled oligonucleotides containing the methylated or unmethylated CpG site cg05575921 were purchased from Metabion. Oligonucleotides were annealed and purified in a 12% polyacrylamide gel. Binding reaction was carried out with or without different concentrations of unlabeled competitor oligonucleotides using 5 µg of nuclear extract in 5x binding buffer (20% v/v Glycerol, 5 mM $MgCl_2$, 2.5 mM EDTA, 2.5 mM DTT, 250 mM NaCl, 50 mM TrisHCl pH7.5) with 0.5 µg poly dI-dC (Roche Diagnostics) and 1 ng of labeled probe in a total volume of 10 µl for 20 min at 4°C. Protein-DNA complexes were separated on a 5.3% polyacrylamide gel by electrophoresis in 0.5 x tris-borate-EDTA (TBE) buffer. The gels were visualized by scanning with the Thyphoon Trio + (GE Healthcare).

2.9. Material

2.9.1. Equipment and labware

Centrifuge:	Sigma 4K15C (Sigma Laborzentrifugen, Osterode, Germany) Rotanta 46 RS (Hettich, Tuttlingen, Germany) Mikrozentrifuge (NeoLab, Heidelberg, Germany) Centrifuge 5417R (Eppendorf AG, Hamburg, Germany)
Gel electrophoresis documentation system:	UVT-40 M Transilluminator (Herolab, Wiesloch, Germany) E.A.S.Y. 429 K Camera (Herolab, Wiesloch, Germany)
Gel electrophoresis device:	Bio-Rad Power Pac 300/3000 (BIO-RAD Laboratories, Munich, Germany)
Mass spectrometer Bruker™ Autoflex	Sequenom, Inc, Hamburg, Germany
Mass Array™ Nanodispenser	Sequenom, Inc, Hamburg, Germany
PCR cycler:	Thermal Cycler C1000™ (BIO-RAD Laboratories, Munich, Germany) DNA Engine Tetrad (MJ Research, South San Francisco, USA)
Pipetting robots:	Temo (Tecan AG, Crailsheim, Germany) Genesis RSP 150 Work Station (Tecan AG, Crailsheim, Germany)

8-Sample Spectrophotometer	Peqlab, Erlangen, Germany
2 Hybridization Ovens	Illumina, Inc., Orange, CA, USA
Vacuum Desiccator (Nalgene)	Thermo Fisher Scientific, Waltham, MA, USA
Kimwipes	Kimberly-Clark, Koblenz-Reinh., Germany
Water circulator	VWR, Darmstadt, Germany
Heat sealer	Thermo Fisher Scientific, Waltham, MA, USA
Illumina iScan	Illumina, Inc., Orange, CA, USA
Hybex Microsample Incubator	SciGene, Sunnyvale, CA, USA
MIDI plate heat block insert	SciGene, Sunnyvale, CA, USA
450K BeadChips and accompanying CD	Illumina, Inc., Orange, CA, USA
BeadChip Hyb Chambers	Illumina, Inc., Orange, CA, USA
Hyb Chamber gaskets	Illumina, Inc., Orange, CA, USA
BeadChip Hyb Chamber inserts	Illumina, Inc., Orange, CA, USA
Flow-Through Chambers	Tecan AG, Crailsheim, Germany)
Flow Cell storage box	Illumina, Inc., Orange, CA, USA
48-position Chamber Rack	Tecan AG, Crailsheim, Germany)
Spacers	Tecan AG, Crailsheim, Germany)
Temperature probe	Illumina, Inc., Orange, CA, USA
Temperature probe block	Illumina, Inc., Orange, CA, USA
Staining rack (1 staining rack per 8 BeadChips processed simultaneously)	Illumina, Inc., Orange, CA, USA
Dismantling tool	Illumina, Inc., Orange, CA, USA
Glass wash dishes	Illumina, Inc., Orange, CA, USA
BeadChip wash racks and handles	Illumina, Inc., Orange, CA, USA
Slide Storage Box	Illumina, Inc., Orange, CA, USA
Braided silicone tubing 0.375” ID x 0.625” OD, 0.125” inch wall thickness, 25' long	Illumina, Inc., Orange, CA, USA
Stainless steel hose clamps 7/32" x 5/8"	Illumina, Inc., Orange, CA, USA
15 ml conical tubes	BD Biosciences, Heidelberg, Germany
0.8 ml storage plate (MIDI), conical well bottom	Thermo Fisher Scientific, Waltham, MA, USA
Heat Sealing foil sheets, Thermo-Seal	Thermo Fisher Scientific, Waltham, MA, USA
96-well cap mats (piercable, non-autoclavable)	Thermo Fisher Scientific, Waltham, MA, USA
Aerosol filter pipet tips (20 µl, 200 µl, 1000 µl)	Mettler Toledo, Gießen, Germany
2 disposable multi-channel reagent reservoirs	Roth, Karlsruhe, Germany
Pipette tips (10 µl, 20 µl, 200 µl, 1000 µl)	Mettler Toledo, Gießen, Germany
Two quarter-module sterile reservoirs	Beckman Coulter, Fullerton, USA
Forceps	VWR, Darmstadt, Germany
Multi-Sample BeadChip Alignment Fixture	Illumina, Inc., Orange, CA, USA
Robot BeadChip Alignment Fixture (6)	Illumina, Inc., Orange, CA, USA
Robot Tip Alignment Guide-B inserts	Illumina, Inc., Orange, CA, USA
Robot Tip Alignment Guide-B	Illumina, Inc., Orange, CA, USA
Reservoir, full, 150 ml	Beckman Coulter, Fullerton, USA
Reservoir, half, 75 ml	Beckman Coulter, Fullerton, USA
Reservoir, quarter, 40 ml	Beckman Coulter, Fullerton, USA
Reservoir frames, 2 (per TECAN)	Beckman Coulter, Fullerton, USA
Tube racks for vacuum desiccator	VWR, Darmstadt, Germany
Vacuum source (> 508 mm Hg (0.68 bar)	VWR, Darmstadt, Germany

Velcro for high-speed shaker platform	Illumina, Inc., Orange, CA, USA
384-well microplates (ABGene Thermo-Fast)	Thermo Fisher Scientific, Waltham, MA, USA
96-well PCR microplates	Thermo Fisher Scientific, Waltham, MA, USA
MicroAmp clear adhesive films	Thermo Fisher Scientific, Waltham, MA, USA
Single pipettors (Rainin) 1–10 µl, 5–50 µl, 50–300 µl, 100 – 1000 µl	Mettler Toledo, Gießen, Germany
Multi-channel pipettes (two separate sets) (Rainin) 1–10 µl, 5–50 µl, 50–300 µl, 100 – 1200 µl	Mettler Toledo, Gießen, Germany
12-channel precision pipette (0.5 – 12 µl)	Thermo Fisher Scientific, Waltham, MA, USA
Repeater pipettors with Combitips Plus 0.1 ml	Eppendorf AG, Hamburg, Germany
Microtubes 1.5 ml and 2 ml	Eppendorf AG, Hamburg, Germany
Safe-Lock Tubes 1.5 ml, black	A.Hartenstein, Würzburg, Germany
StableStak Racks (20 µl, 200 µl, 1000 µl)	Mettler Toledo, Gießen, Germany
Maxigel System	Biometra, Goettingen, Germany
P25T Standard Power Pack	Biometra, Goettingen, Germany
Typhoon Trio +	GE Healthcare, Munich, Germany
Aluminum foil	

2.9.2. Buffer, solutions, reagents, and enzymes

Nanopure Water	Roth, Karlsruhe, Germany
EZ DNA M-Dilution Buffer	Zymo Research, Orange, CA, USA
EZ DNA CT Conversion Reagent	Zymo Research, Orange, CA, USA
EZ DNA M-Binding Buffer	Zymo Research, Orange, CA, USA
EZ DNA Spin I Column	Zymo Research, Orange, CA, USA
EZ DNA M-Wash Buffer	Zymo Research, Orange, CA, USA
EZ DNA M-Desulphonation Buffer	Zymo Research, Orange, CA, USA
ATM - Anti-Stain Two-Color Master Mix	Illumina, Inc., Orange, CA, USA
FMS - Fragmentation solution	Illumina, Inc., Orange, CA, USA
MA1 - Multi-Sample Amplification 1 Mix	Illumina, Inc., Orange, CA, USA
RPM - Random Primer Mix	Illumina, Inc., Orange, CA, USA
MSM - Multi-Sample Amplification Master Mix	Illumina, Inc., Orange, CA, USA
PB1 - Wash Solution	Illumina, Inc., Orange, CA, USA
PB2 - Humidifying buffer used during hybridization	Illumina, Inc., Orange, CA, USA
PM1 - Precipitation solution	Illumina, Inc., Orange, CA, USA
RA1 - Resuspension, hybridization, and wash solution	Illumina, Inc., Orange, CA, USA
STM - Superior Two-Color Master Mix	Illumina, Inc., Orange, CA, USA
TEM - Two-Color Extension Master Mix	Illumina, Inc., Orange, CA, USA
XC1 - XStain BeadChip solution 1	Illumina, Inc., Orange, CA, USA
XC2 - XStain BeadChip solution 2	Illumina, Inc., Orange, CA, USA
XC3 - XStain BeadChip solution 3	Illumina, Inc., Orange, CA, USA
XC4 - XStain BeadChip solution 4	Illumina, Inc., Orange, CA, USA
0.1N NaOH (Sodium hydroxide)	Sigma-Aldrich, Osterode, Germany
100% Ethanol	Merck, Darmstadt, Germany
100% 2-propanol	Merck, Darmstadt, Germany
Deionized water	Merck, Darmstadt, Germany

0.5 M EDTA (100 ml)	Merck, Darmstadt, Germany
1x TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA)	Merck, Darmstadt, Germany
95% formamide/1 mM EDTA	Roth, Karlsruhe, Germany
100% EtOH	Merck, Darmstadt, Germany
5% sodium hypochlorite (10% bleach)	Sigma-Aldrich, Osterode, Germany
PCR Buffer, 10x (contains 20 mM MgCl2)	Sequenom, Inc, Hamburg, Germany
MgCl2 , 25mM	Sequenom, Inc, Hamburg, Germany
PCR Polymerase, 5 U/µl	Sequenom, Inc, Hamburg, Germany
dNTP Mix, 25mM each	Sequenom, Inc, Hamburg, Germany
SAP (Shrimp Alkaline Phosphatase) 1.7 U/µl	Sequenom, Inc, Hamburg, Germany
5x T& Polymerase Buffer	Sequenom, Inc, Hamburg, Germany
T-cleavage Mix	Sequenom, Inc, Hamburg, Germany
DTT, 100mM	Sequenom, Inc, Hamburg, Germany
T7 RNA & DNA Polymerase (50U/uL)	Sequenom, Inc, Hamburg, Germany
RNase A (10ug/uL)	Sequenom, Inc, Hamburg, Germany
Clean Resin	Sequenom, Inc, Hamburg, Germany
4-Pt Calibrant	Sequenom, Inc, Hamburg, Germany
SpectroCHIPTM G384+10	Sequenom, Inc, Hamburg, Germany
Boric acid	Alfa Aesar, Karlsruhe, Germany
EDTA	Merck, Darmstadt, Germany
Glycerol	NeoLab, Heidelberg, Germany
Magnesium chloride	Roth, Karlsruhe, Germany
Orange G	Sigma-Aldrich, Osterode, Germany
Poly[d(I-C)]	Roche Diagnostics, Penzberg, Germany
Tris	Merck, Darmstadt, Germany
Tetramethylethylenediamine (TEMED)	Sigma-Aldrich, Osterode, Germany
Sodium chloride	Merck, Darmstadt, Germany
Rotriphorese® Gel 40 (37,5:1)/30 (37,5:1)	Roth, Karlsruhe, Germany
Ammonium peroxodisulfate (APS)	Biozym, Hessisch Oldendorf, Germany
Dimethylsulfoxide (DMSO)	Sigma-Aldrich, Osterode, Germany

2.9.3. Software

2.9.3.1. Software for 450K BeadChip and EpiTYPER assay

iScan Control Software	Illumina, Inc., Orange, CA, USA
Illumina GenomeStudio (Methylation Module)	Illumina, Inc., Orange, CA, USA
RT Workstation 3.4: Chip Linker Caller Acquire	Sequenom, Inc, Hamburg, Germany
MassARRAY EpiTyper 1.2 EpiDesigner.com Plate Editor Analyzer	Sequenom, Inc, Hamburg, Germany
ImageQuant TL	Amersham Biosciences, Freiburg, Germany
Vector NTI Suite 9	www.informaxinc.com

2.9.3.2. Statistical software

R 2.14	www.r-project.org
MassArray package	Thompson and Greally (2009). MassArray: Analytical Tools for MassArray Data. R package version 1.10.0.
METAL	http://www.sph.umich.edu/csg/abecasis/Metal/; release 2011-03-25

2.9.4. Online databases and programs

Ensembl	www.ensembl.org
National Center for Biotechnology Information	www.ncbi.nlm.nih.gov
UCSC Genome Browser	http://genome.ucsc.edu/cgi-bin/hgGateway
Sequenom's Primer Design	www.epidesigner.com

3. Results

The first part of this section presents the establishment of the Illumina 450K BeadChip in our laboratory and the results of the effect of tobacco smoking on genome-wide DNA methylation in whole blood, measured in the discovery (KORA F4) and replication (KORA F3) panel. This includes the comparison of current smokers and former smokers to never smokers in general, with detailed description of the genes that correspond to the most significant CpG sites detected, addresses the question of the effect of cessation time and cumulative smoke exposure on the DNA methylation in former smokers and gives insights into gender-specific effects. Furthermore, the results of methylation-specific DNA-protein binding analysis by electrophoretic mobility shift assay (EMSA) carried out exemplarily for the most striking CpG site, in order to assess the potential biological relevance of DNA methylation differences caused by tobacco smoking, are shown.
The second part demonstrates the establishment of Sequenom's EpiTYPER platform, which allows region-specific DNA methylation analysis. With this approach, the earlier detected 450K BeadChip DNA methylation differences for the five most significant loci of the current vs. never smoker analysis were validated.

3.1. Genome-wide effect of tobacco smoking on the DNA methylation status

3.1.1. Establishment of the Illumina 450K BeadChip array

In order to investigate the effect of tobacco smoking on genome-wide DNA methylation, DNA methylation analyses had to be performed with the Illumina 450K BeadChip using DNA obtained from whole blood. The Illumina microarray platform was already in use in our laboratory for SNP detection analysis, but the newly developed Infinium HD assay for methylation first had to be established in order to run the 1814 DNA samples of KORA F4 and 479 DNA samples of KORA F3. The pilot experiment was started with 48 DNA samples which corresponds to a half 96-well plate or four 450K BeadChips.

3.1.1.1. Bisulfite conversion

As the DNA methylation information is erased by the whole genome amplification (WGA) step during the Illumina process, a methylation-dependent pretreatment of the DNA had to be carried out to reveal the presence or absence of the methyl group at cytosine residues. Bisulfite treatment was used for this purpose, as recommended by Illumina, followed by a test PCR with primer pairs specific for bisulfite-converted DNA. It is essential to test if the bisulfite treatment was carried out successfully as incomplete conversion of cytosine to uracil can result in false-positive methylation signals The resulting DNA bands were as expected (Figure 13), therefore it was assumed that bisulfite conversion was completed and the 48 DNA samples were further processed according to the Illumina protocol.

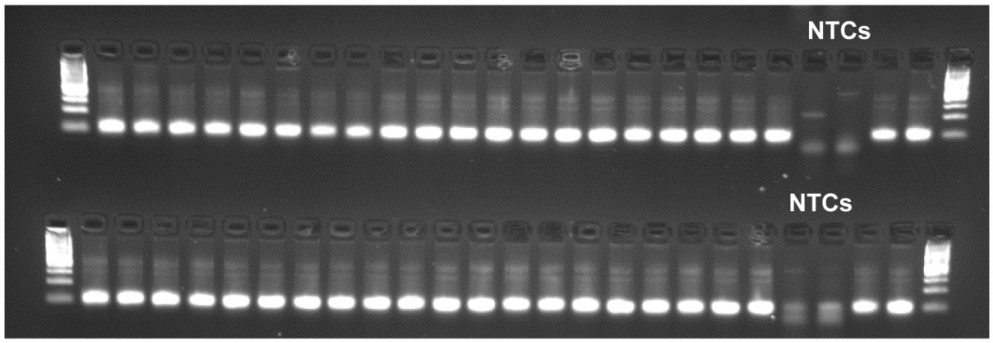


Figure 13. Agarose gel of bisulfite conversion PCR test products. NTCs = Negative controls (H_2O)

After the four BeadChips were scanned with the Illumina iScan Reader, quality assessment of assay performance, based on built-in controls that accompany each experiment, was conducted using the "Control Dashboard" of the Illumina GenomeStudio Methylation Module. The control diagrams for DNP and Biotin staining, extension, hybridization, target removal, specificity, negative, and non-polymorphic controls looked as expected, but the ones for bisulfite conversion I and II, which assess the efficiency of bisulfite conversion of the genomic DNA, differed as displayed by Figure 14.

Even though the test PCR seemed to assure a complete conversion of the DNA samples, this obviously was not the case. Therefore, the built-in controls for bisulfite conversion were used instead of the test PCR to control for completeness of conversion, and after optimization of the bisulfite conversion the assay was run on all of the 1814 samples of KORA F4 and subsequently on all 479 samples of KORA F3.

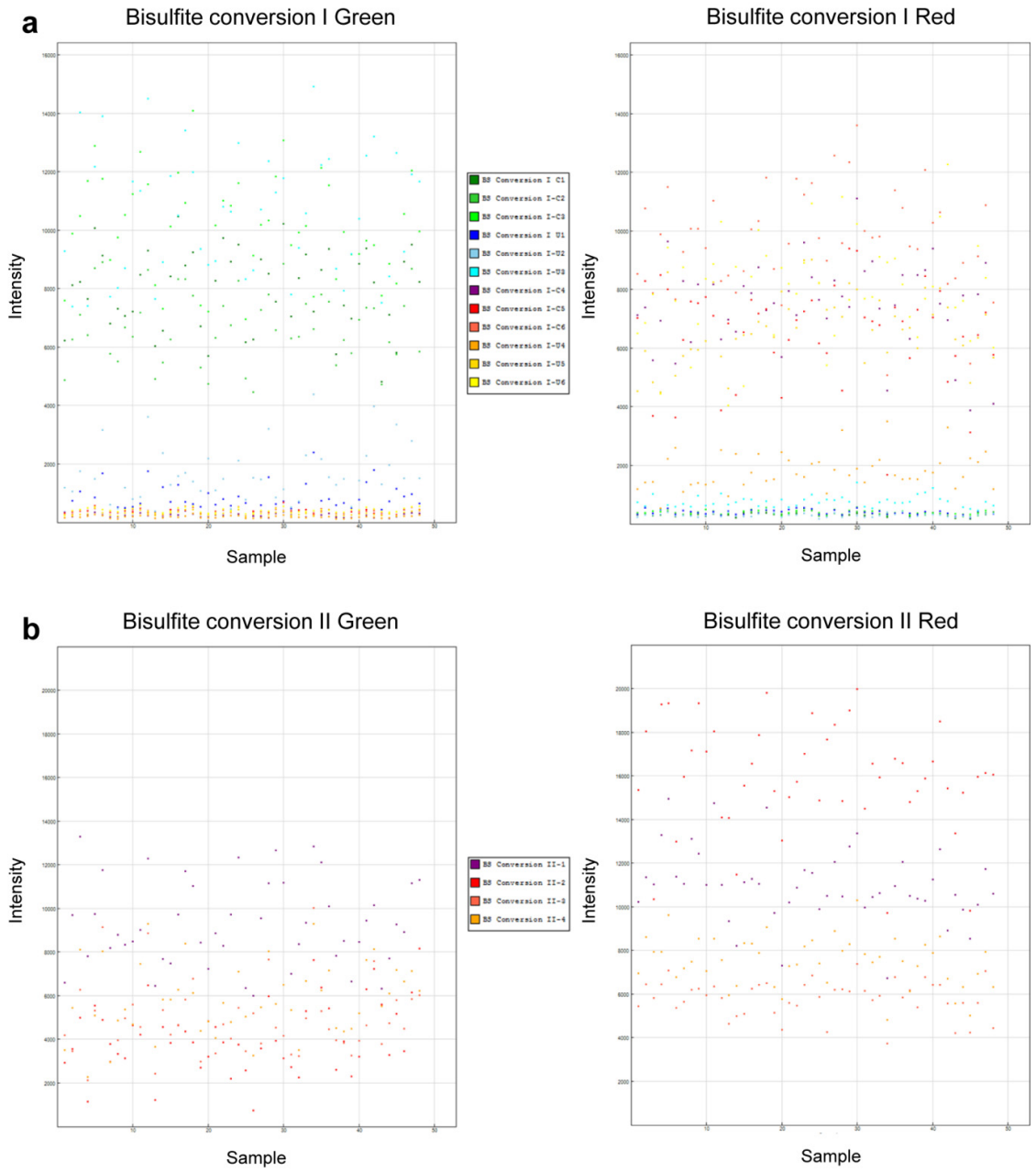


Figure 14. Control diagrams for bisulfite conversion I and II of four Testchips. Visualized are **a)** the controls of bisulfite conversion I that use the Infinium I probe design and allele-specific single base extension, and **b)** the controls of bisulfite conversion II that use the Infinium II probe design and single base extension to display efficiency of the bisulfite conversion. Ad **a)** If the bisulfite conversion reaction was successful, the converted probes ("C") will match the converted sequence and be extended. If the sample has unconverted DNA, the unconverted probes ("U") will be extended. Performance of bisulfite conversion controls C1, C2, and C3 should be monitored in the Green channel, and controls C4, C5, and C6 should be monitored in Red channel. Ad **b)** If the bisulfite conversion reaction was successful, the "A" base will get incorporated and the probe will have intensity in the Red channel. If the sample has unconverted DNA, the "G" base will get incorporated across the unconverted cytosine, and the probe will have elevated signal in the Green channel (slides and description are adapted from the “Control Dashboard” of the Illumina GenomeStudio Methylation Module).

3.1.1.2. Quality of data and data pre-processing

The diagrams for DNP and Biotin staining, extension, hybridization, target removal, bisulfite conversion I and II, specificity, negative, and non-polymorphic controls were monitored for all 1814 DNA samples of KORA F4 and 479 DNA samples of KORA F3 to assure good quality of the data and are displayed in Figure A.1 and Figure A.2, respectively. A final technical call rate of how many percent of processed samples passed the QC through the "Control Dashboard" of the Illumina GenomeStudio Methylation Module was assessed. In KORA F4, 1814 samples were processed, 54 samples (~ 3%) were repeated because they did not pass the technical QC and 9 samples (~ 0.5%) dropped out after rerun, leading to a final technical call rate of 99.5%. In KORA F3, 479 samples were processed, 24 samples (~ 4.8%) were repeated because of chip-based problems, 9 samples (~ 1.8%) were repeated because they did not pass the technical QC and no sample dropped out after rerun, leading to a final technical call rate of 100%. Hence, the quality of the data was very good and only 9 samples of KORA F4 had to be excluded because of deviations from optimal performance that also remained when the complete Illumina Infinium HD Methylation protocol was repeated, suggesting insufficient DNA quality.

Pre-processing and normalization of the 450K data of KORA F4 and F3 was carried out with the pipeline of Touleimat & Tost [175]. Table 7 summarizes the number of samples and probes removed prior to 450K data analysis. The *Quality Control* step revealed that three samples of KORA F4 had less than 80% of high-confidence probes, which were therefore removed from further analysis; all samples of KORA F3 remained. The *Probe Filtering* step led to the exclusion of 17,196 CpG sites located in 50bp proximity to a single nucleotide polymorphisms (SNP) with a minor allele frequency of at least 5%.

Table 7. Number of samples and probes removed prior to 450K data analysis.

	KORA F4	***KORA F3***
Probes		
Total number of probes measured	485577	485577
Removal of SNPs on 450K BeadChip	65	65
Loaded into pipeline (Touleimat & Tost 2012)	485512	485512
Removal of probes containing SNPs (MAF < 0.05)	17196	17196
Final N of probes included in analysis	**468316**	**468316**
Samples		
Total number of samples measured	1814	479
Removal of samples:		
Did not pass the initial quality assessment of assay performance	9	0
< 80% of high-confidence probes* (accomplished in pipeline Touleimat & Tost 2012)	3	0
Missing phenotype data	9	0
Final N of samples included in analysis	**1793**	**479**

* After removal of low-confidence probes with a detection p-value > 0.01; calculation for percentage of high-confidence probes: # high−confidence probes for sample i / # probes in total for sample i

Furthermore, the 450K BeadArray is a two-color assay and a comparison of the signal intensities of the green and red channel within the measurement of KORA F4 and F3 revealed an imbalance in the intensities, visualized by a line plot in Figure 15. These color-biases were adjusted by the *Signal Correction* step within the pipeline.

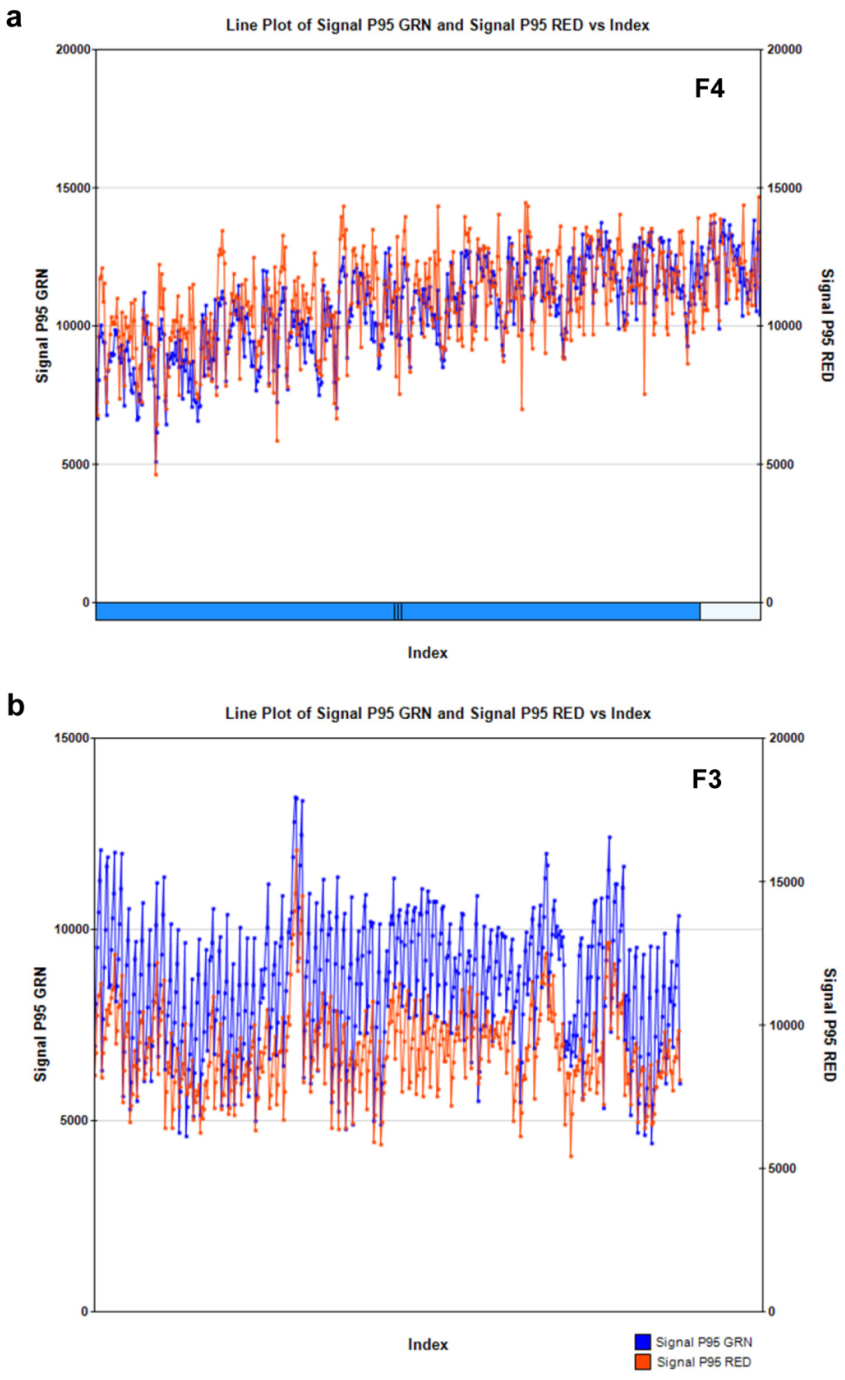


Figure 15. Comparison of the signal intensities of the green and red channel. The 95th percentile of intensity in the green channel and the 95th percentile of intensity in the red channel are plotted versus the Index, which represents **a)** 1814 samples of KORA F4 and **b)** 479 samples of KORA F3.

3.1.1.3. Reproducibility of the 450K BeadArray methylation data

To evaluate reproducibility of the methylation data gained from the 450K BeadChip, a positive control was included per Illumina run, summing up to a total of six replicates. The data showed very good reproducibility, with Pearson correlation coefficients ranging from 0.990 to 1.000 as displayed in Table 8.

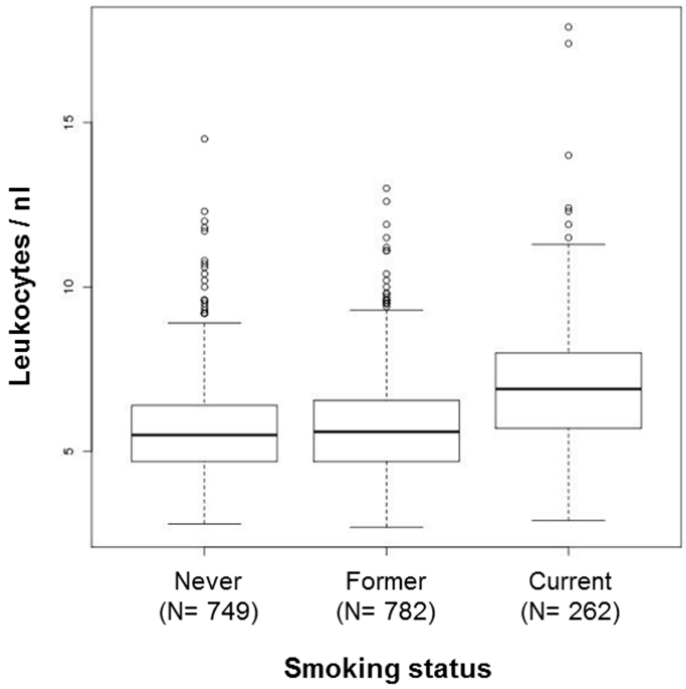
Table 8. Pearson correlation coefficient for technical replicates.

	PC1	PC2	PC3	PC4	PC5	PC6
PC1	1.000	0.997	0.995	0.996	0.997	0.992
PC2	0.997	1.000	0.995	0.997	0.997	0.991
PC3	0.995	0.995	1.000	0.995	0.993	0.991
PC4	0.996	0.997	0.995	1.000	0.995	0.991
PC5	0.997	0.997	0.993	0.995	1.000	0.990
PC6	0.992	0.991	0.991	0.991	0.990	1.000

PC: Positive Control

3.1.2. Including white blood cell (WBC) count as covariate

As numerous studies have found increased leukocyte blood counts in current compared to never smokers and reported that leukocyte count decreases after cessation of smoking (please see chapter 2.6.5.3.), it was of interest if increased leukocyte blood counts could also be observed in current compared to never smokers in the KORA F4 cohort and what the numbers would be in former smokers. Figure 16 shows that current smokers, but not former smokers, did have a significant higher number of leukocytes/nl in their blood compared to never smokers. Cigarette smoke is a powerful inducer of inflammatory mediators and results in a chronic inflammatory state; the increased circulating leukocytes are the principal cellular mediators of inflammation. To avoid measurement of inflammation-related instead of tobacco smoking-associated DNA methylation differences, all statistical models were adjusted for the number of WBCs, in addition to the classic covariates described in 2.6.5.3.



	Never vs. former smokers	Never vs. current smokers
Beta	0.058	1.437
SD	0.082	0.114
p-value	0.478	< 2e-16

SD: Standard Deviation

Figure 16. Comparison of WBC (leukocytes/nl) between smoking groups. The box part within the box plot represents the central 50% of the data or the Interquartile Range (IQR). The lower edge of the box plot is the first quartile or 25th percentile. The upper edge of the box plot is the third quartile or 75th percentile. The whiskers in the box plot represent the tails of the distribution, with the upper whisker extending upward to the third quartile (Q3 +1.5(IQR)) and the lower whisker extending downward to the first quartile (Q1-1.5(IQR)). The points outside the range represent outliers.

3.1.3. Effect of tobacco smoking on genome-wide DNA methylation

During the time of chip processing the KORA cohorts were revised concerning their information about smoking status when inconsistencies were found between what was stated in the baseline survey (KORA S4 and S3) and what was stated several years later in the follow-up study (KORA F4 and F3). For example, if one and the same person stated to be current smoker in KORA S4 and stated to be never smoker in KORA F4, this person was revised as former smoker in KORA F4. Of the 479 KORA F3 samples chosen for replication, 11 of the original never and current smokers had to be revised to former smokers and were

therefore excluded for further analysis. The characteristics of the final cross-sectional KORA F4 discovery and case-control KORA F3 replication panel are summarized in Table 9. The results of the genome-wide distribution of the significant, differentially methylated CpG sites of current vs. never smokers in the discovery (KORA F4; current N=262, never N=749) and replication (KORA F3; current N=236, never N=232) panel are represented as Manhattan Plots in Figure 17a and 17b, respectively.

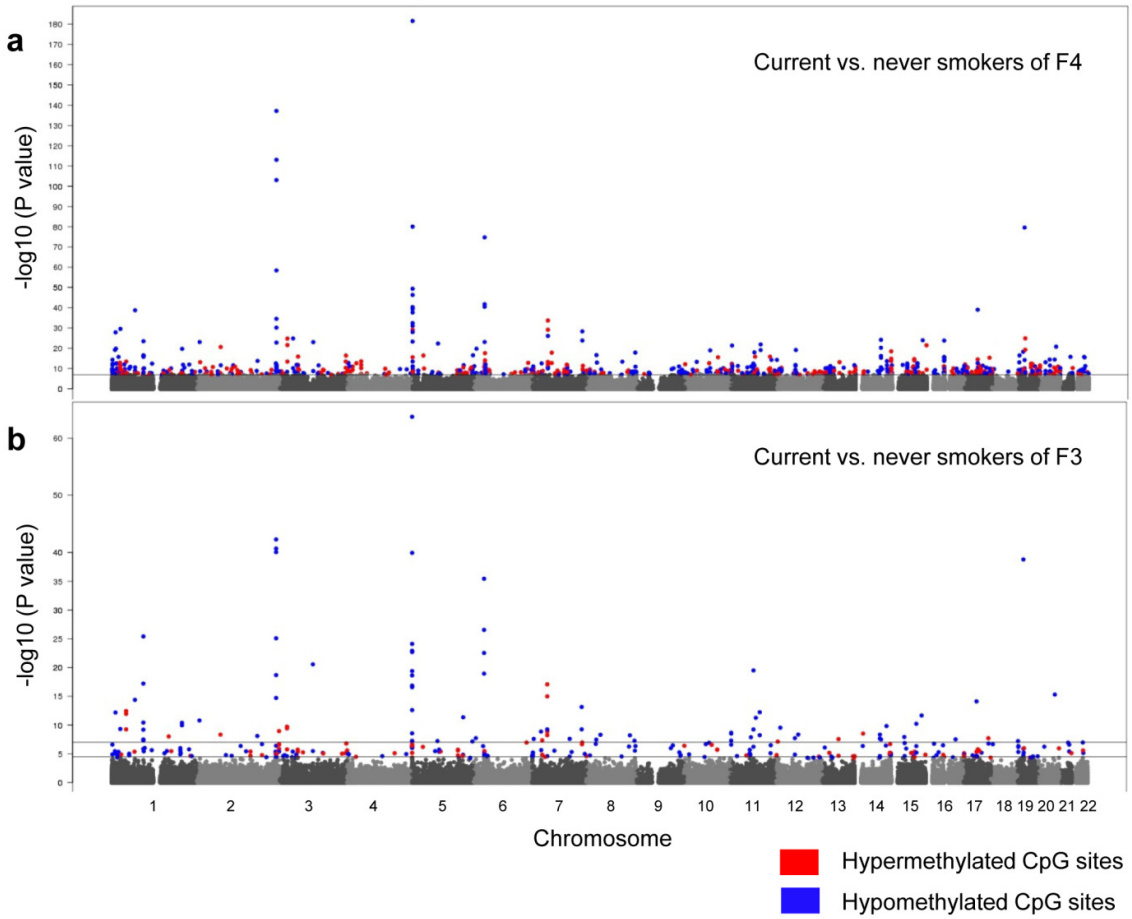


Figure 17. Genome-wide effect of current tobacco smoking on DNA methylation. The continuous lines in the Manhattan Plots mark the 1E-07 significance thresholds; the lower line in Figure 17b marks the 5E-05. The significant CpG sites of **a)** current vs. never smokers of the KORA F4 discovery panel and **b)** current vs. never smokers of the KORA F3 replication panel, are color coded with the direction of the aberration in current smokers, using blue for hypomethylated and red for hypermethylated CpG sites.

Table 9. Characteristics of the study population.

	a) F4 Discovery panel					**b) F3 Replication panel**		
	Never smokers	Former smokers	p-value test never-former*	Current smokers	p-value test never-current*	Never smokers	Current smokers	p-value test never-current*
N	749 (41.77%)	782 (43.61%)		262 (14.61%)		232 (49.57%)	236 (50.43%)	
Age (years)	62.09 (31-81)	61.02 (35-81)	0.01 [a]	56.96 (46-76)	<0.001 [a]	53.16 (34-83)	52.78 (35-80)	ns [a]
Gender			<0.001 [b]		<0.001 [b]			ns [b]
Female	490 (65.42%)	310 (39.64%)		120 (45.80%)		115 (49.57%)	116 (49.15%)	
Male	259 (34.58%)	472 (60.36%)		142 (54.20%)		117 (50.43%)	120 (50.85%)	
Smoking								
Pack-years	0	23.00 (0.60-195)		33.64 (0.54-127)				
Time since quitting (years)		21.83 (0.08-62.5)						
BMI (kg/m^2)	28.05 (18.99-45.17)	28.61 (17.50-55.99)	ns [a]	27.11 (18.68-47.31)	<0.001 [a]	27.31 (15.07-45.31)	26.74 (17.50-45.28)	ns [a]
WBC count (WBC/nl)	5.64 (2.80-14.50)	5.75 (2.70-13.00)	ns [a]	7.08 (2.90-17.90)	<0.001 [a]	6.61 (3.40-29.10)	8.06 (3.90-15.10)	0.02 [a]
Alcohol (g/kg/day)	0.15 (0.00-1.38)	0.22 (0.00-1.42)	<0.001 [a]	0.23 (0.00-1.55)	0.005 [a]	0.17 (0.00-0.95)	0.23 (0.00-1.21)	<0.001 [a]

BMI: Body Mass Index, WBC: White Blood Cell count; each individual characteristic is given in mean with range in parentheses. * Group comparison - ns: not significant according to [a] paired Wilcoxon rank sum test with never smokers as reference and [b] Fisher's exact test for gender.

Depending upon the smoking status, 972 CpG sites could be identified with differential methylation levels after conservative correction for multiple testing ($p \leq 1E\text{-}07$) throughout the genome in KORA F4 (Table A.1), of which 187 CpGs could be replicated in KORA F3 ($p \leq 5E\text{-}5$; Table A.2). Overall, genome-wide, significant, differentially methylated CpGs could be detected in each of the 22 autosomes with p-values ranging from 9.31E-08 to 2.54E-182 and with a percent of variance explained by smoking of 1.31 to 41.02 (Table A.2). The differences in the methylation levels between current and never smokers for the shared significant CpGs of KORA F4 and F3 were similar.

Table 10 displays all replicated CpG sites of current vs. never smokers with a methylation difference of more than 5% in both panels. Some of the detected CpG sites displayed in Table 10 had very low p-values which indicate that the smoking-associated differences in methylation were not measured by chance, especially for these sites. Also, the percent of explained variance was surprisingly high, for example, smoking status (current or never) explained 41.02% of the variance in the degree of methylation at cg05575921. Since the replicated 187 CpGs correspond to 95 loci, this thesis will primary focus on the CpGs described in this table and their corresponding loci, paying particular attention to two remarkably clusters of CpGs with smoking dependent changes in methylation patterns identified on chromosome 2q37.1 and 5p15.3 (Figure 17, Table 10, Table A.1 and A.2).

Table 10. Genome-wide differentially methylated CpGs of current compared to never smokers.

			a) F4 Discovery panel			**b)** F3 Replication panel			**c**) Meta-analysis	Methylation ß-value as median (first quartile - third quartile)			
										a) F4 Discovery panel		**b)** F3 Replication panel	
CpG	Chr	Gene	Median ß-value methylation difference in %	p-value	Explained variance in %	Median ß-value methylation difference in %	p-value	Explained variance in %	p-value	Never smokers	Current smokers	Never smokers	Current smokers
cg04885881	1	*x*[a]	-7.41	1.35E-28	9.57	-6.71	6.62E-13	9.83	7.19E-40	0.3115 (0.25-0.39)	0.2374 (0.18-0.31)	0.2899 (0.24-0.36)	0.2228 (0.17-0.29)
cg15542713	1	*HIVEP3*	9.52	2.54E-08	2.03	9.95	3.65E-13	9.51	3.56E-18	0.4002 (0.31-0.50)	0.4954 (0.38-0.60)	0.3784 (0.30-0.46)	0.4779 (0.40-0.57)
cg25189904	1	*GNG12*	-8.19	1.71E-39	14.27	-7.40	4.13E-15	11.42	9.11E-53	0.2740 (0.23-0.33)	0.1921 (0.16-0.24)	0.2856 (0.24-0.33)	0.2116 (0.17-0.26)
cg12876356	1	*GFI1*	-8.48	2.92E-17	5.46	-22.63	6.18E-18	13.59	2.47E-32	0.5475 (0.49-0.62)	0.4627 (0.37-0.55)	0.7010 (0.56-0.80)	0.4746 (0.33-0.60)
cg09935388	1	*GFI1*	-15.31	3.27E-24	7.39	-16.47	3.98E-26	18.93	1.26E-46	0.5765 (0.49-0.70)	0.4234 (0.32-0.54)	0.4757 (0.40-0.54)	0.3110 (0.24-0.39)
cg27241845	2	*x*[a]	-8.74	2.79E-35	10.96	-8.36	2.07E-19	13.78	6.03E-53	0.4617 (0.41-0.52)	0.3743 (0.31-0.42)	0.4457 (0.41-0.49)	0.3621 (0.31-0.43)
cg21566642	2	*ALPPL2*[b]	-16.70	6.90E-138	36.24	-15.58	8.82E-41	27.13	7.86E-175	0.5658 (0.51-0.61)	0.3988 (0.35-0.45)	0.5511 (0.50-0.59)	0.3953 (0.31-0.44)
cg01940273	2	*ALPPL2*[b]	-7.89	9.28E-114	31.50	-7.53	5.46E-43	28.33	1.92E-154	0.3437 (0.32-0.37)	0.2648 (0.24-0.29)	0.2941 (0.27-0.31)	0.2189 (0.20-0.25)
cg15693572	3	*x*[a]	10.87	3.84E-14	4.50	6.90	1.74E-06	4.60	3.69E-19	0.2942 (0.22-0.41)	0.4029 (0.29-0.53)	0.3159 (0.23-0.40)	0.3850 (0.29-0.50)
cg23480021	3	*x*[a]	18.63	2.67E-22	8.74	15.77	1.95E-10	8.09	3.63E-31	0.4259 (0.30-0.60)	0.6122 (0.44-0.78)	0.4266 (0.29-0.58)	0.5844 (0.43-0.78)
cg03274391	3	*x*[a]	18.06	1.67E-25	8.79	11.93	3.74E-10	7.78	5.47E-34	0.3190 (0.23-0.47)	0.4997 (0.34-0.70)	0.2872 (0.21-0.41)	0.4065 (0.29-0.55)
cg15417641	3	*CACNA1D*	7.92	1.25E-16	4.90	5.28	1.90E-06	4.61	1.65E-21	0.6818 (0.58-0.76)	0.7610 (0.68-0.83)	0.6019 (0.52-0.68)	0.6546 (0.60-0.73)
cg21188533	3	*CACNA1D*	10.18	1.12E-10	3.10	5.56	4.38E-05	3.42	2.33E-14	0.5724 (0.45-0.69)	0.6742 (0.55-0.76)	0.5309 (0.41-0.63)	0.5865 (0.50-0.70)
cg03991871	5	*AHRR*	-5.73	4.84E-41	12.72	-7.88	1.23E-23	17.61	8.03E-63	0.9469 (0.92-0.96)	0.8896 (0.80-0.93)	0.9175 (0.89-0.94)	0.8387 (0.69-0.89)
cg05575921	5	*AHRR*	-24.40	2.54E-182	41.02	-23.29	1.81E-64	39.69	8.70E-244	0.8841 (0.86-0.90)	0.6401 (0.60-0.73)	0.9028 (0.88-0.92)	0.6699 (0.58-0.77)
cg26703534	5	*AHRR*	-6.17	2.14E-38	10.90	-5.77	7.83E-25	17.21	3.66E-61	0.6269 (0.60-0.65)	0.5652 (0.52-0.60)	0.6180 (0.60-0.63)	0.5603 (0.53-0.60)
cg25648203	5	*AHRR*	-7.96	4.73E-47	12.98	-7.55	2.03E-23	16.96	1.01E-68	0.8642 (0.83-0.89)	0.7846 (0.72-0.84)	0.8473 (0.81-0.87)	0.7719 (0.71-0.82)
cg21161138	5	*AHRR*	-10.45	8.58E-81	21.63	-10.60	1.19E-40	28.49	1.49E-119	0.6998 (0.67-0.73)	0.5954 (0.53-0.65)	0.7304 (0.70-0.75)	0.6244 (0.57-0.68)
cg06126421	6	*x*[a]	-17.05	1.72E-75	23.60	-17.89	3.73E-36	24.58	8.42E-110	0.7724 (0.70-0.82)	0.6019 (0.48-0.69)	0.6919 (0.63-0.74)	0.5131 (0.42-0.61)
cg14753356	6	*x*[a]	-5.37	8.14E-24	8.58	-6.41	2.98E-23	17.61	5.90E-44	0.2383 (0.20-0.28)	0.1846 (0.15-0.22)	0.2451 (0.20-0.28)	0.1810 (0.15-0.22)
cg00931843	6	*TIAM2*	6.37	1.69E-10	2.98	5.66	1.09E-07	5.73	1.37E-16	0.2963 (0.23-0.39)	0.3600 (0.27-0.47)	0.3293 (0.25-0.42)	0.3859 (0.29-0.46)
cg22132788	7	*MYO1G*	6.68	1.99E-34	11.04	5.02	8.44E-18	13.68	1.54E-50	0.8599 (0.81-0.90)	0.9267 (0.88-0.95)	0.8727 (0.83-0.90)	0.9229 (0.88-0.95)
cg12803068	7	*MYO1G*	14.96	7.08E-30	10.34	15.12	1.04E-15	12.29	6.33E-44	0.7382 (0.60-0.85)	0.8878 (0.77-0.94)	0.7014 (0.60-0.81)	0.8527 (0.71-0.92)
cg11207515	7	*CNTNAP2*	8.28	2.53E-10	3.44	5.29	2.16E-07	5.30	3.77E-16	0.2725 (0.21-0.35)	0.3552 (0.27-0.44)	0.2166 (0.17-0.27)	0.2695 (0.21-0.33)
cg25305703	8	*x*[a]	-6.05	2.09E-08	2.97	-6.17	6.16E-09	6.16	2.75E-15	0.7103 (0.65-0.76)	0.6498 (0.58-0.73)	0.6679 (0.62-0.71)	0.6062 (0.54-0.67)
cg26361535	8	*ZC3H3*	-5.54	4.14E-09	2.83	-6.35	4.39E-07	4.90	1.35E-14	0.7593 (0.70-0.81)	0.7040 (0.63-0.77)	0.7442 (0.69-0.79)	0.6807 (0.64-0.75)
cg21611682	11	*LRP5*	-5.24	1.09E-18	6.50	-5.38	3.16E-20	14.13	9.90E-36	0.5276 (0.50-0.56)	0.4752 (0.43-0.52)	0.4844 (0.46-0.51)	0.4306 (0.40-0.47)
cg23126342	13	*PCDH9*	9.11	7.16E-14	4.38	7.56	2.73E-08	6.20	1.23E-20	0.5136 (0.43-0.60)	0.6047 (0.52-0.67)	0.4865 (0.39-0.57)	0.5621 (0.46-0.64)
cg01208318	14	*x*[a]	-7.74	2.54E-12	4.15	-6.20	1.20E-05	3.59	1.57E-16	0.2836 (0.20-0.37)	0.2063 (0.16-0.28)	0.2246 (0.15-0.30)	0.1626 (0.12-0.23)
cg19572487	17	*RARA*	-10.02	9.37E-40	14.02	-7.81	7.56E-15	11.92	9.85E-53	0.4477 (0.40-0.51)	0.3475 (0.30-0.43)	0.4502 (0.40-0.51)	0.3722 (0.31-0.44)
cg00835193	19	*LINGO-3*	-8.23	2.44E-08	2.26	-11.66	4.40E-07	5.01	9.20E-14	0.8822 (0.75-0.93)	0.7999 (0.61-0.92)	0.8938 (0.77-0.94)	0.7772 (0.59-0.91)
cg03636183	19	*F2RL3*	-14.74	2.42E-80	22.45	-17.63	1.65E-39	26.94	5.58E-118	0.4930 (0.45-0.54)	0.3456 (0.28-0.43)	0.5152 (0.47-0.56)	0.3390 (0.28-0.43)

Displayed are **a)** the results of the linear model calculated with M-value adjusted for age, sex, BMI, alcohol, and white blood cell count (p-value and explained variance), as well as the median ß-value methylation difference between current and never smokers for the F4 discovery panel with genome-wide significance (p-value ≤ 1E-07) for all CpG sites with a DNA methylation difference of > 5% in F4 and F3, **b)** the corresponding results of the same CpG sites for the F3 replication panel for comparison (significance level ≤ 5E-05), and **c)** the corresponding p-value gained by meta-analysis of F4 and F3, sorted by chromosome and mapinfo (Genome build 37). [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

3.1.3.1. Effect of tobacco smoking on the methylation status of *AHRR*

The most striking and significant CpG site, cg05575921 (current smokers; KORA F4: -24.40%, p = 2.54E-182, explained variance = 41.02%; KORA F3: -23.29%, p = 1.81E-64, explained variance = 39.69%), is located in the region of chromosome 5p15.3 within the *AHRR* (aryl hydrocarbon receptor (*AHR*) repressor) gene (Table A.1, Table A.2 and Figure 17). *AHRR* consists of 12 exons, with the first one being non-coding. 11 alternatively spliced transcript variants encoding different isoforms have been described for this gene, 9 of them being protein coding (http://www.ensembl.org). Within the *AHRR* gene 20 smoking-dependent, differentially methylated CpGs could be detected in KORA F4, of which 16 could be replicated in KORA F3 (Table A.1, Table A.2). 13 of the replicated CpGs were hypomethylated in smokers compared to never smokers, 3 CpGs were hypermethylated, all located in introns of the gene body. An overview of the *AHRR* gene structure and the differentially methylated CpGs is given in Figure 18.

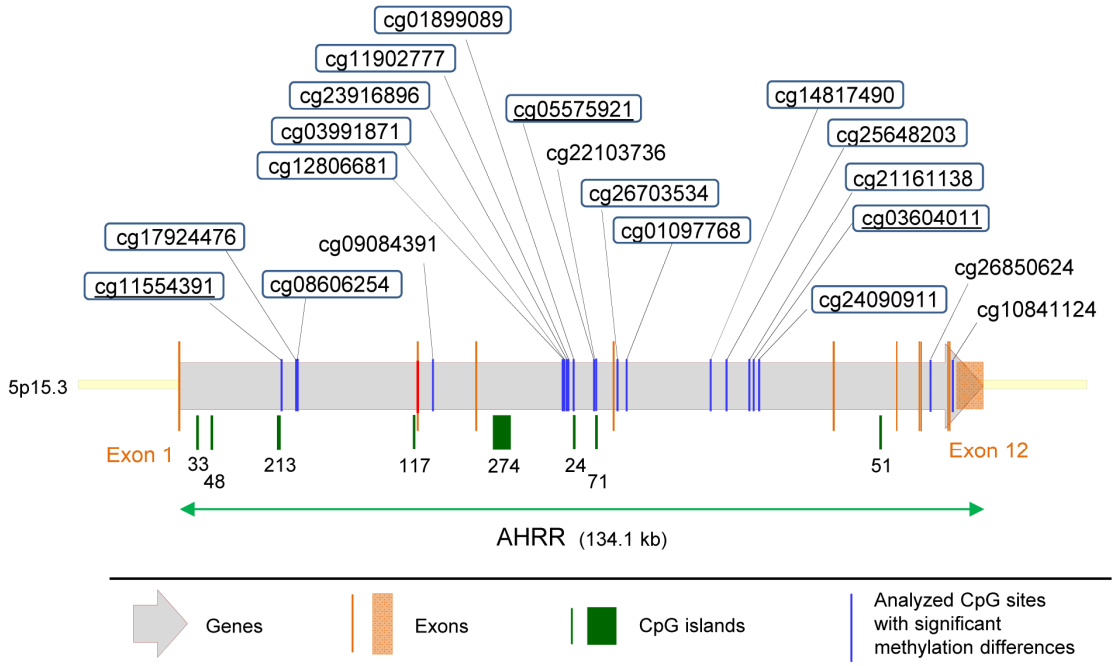


Figure 18. Overview of the results for the *AHRR gene*. The gene structure and the significant differentially methylated CpG sites of *AHRR* (aryl hydrocarbon receptor (*AHR*) repressor) are displayed in current compared to never smokers of the KORA F4 discovery panel. CpG sites which remain significant in the KORA F3 replication panel are framed; CpG sites that were found to be still significant in former smokers are underlined.

3.1.3.2. Effect of tobacco smoking on the methylation status of *ALPP/ALPPL2*

The second most striking region on chromosome 2q37.1 comprises 13 smoking-dependent, differentially methylated CpG sites that could be detected in KORA F4, of which 10 could be replicated in KORA F3 (Table A.1, Table A.2, Figure 17). Three closely related alkaline phosphatase genes, placental (*ALPP*), placental-like (*ALPPL2*) and intestinal (*ALPI*) are located within this region (Figure 19). All CpGs were hypomethylated in current compared to never smokers. Five of the replicated CpGs (cg06644428, cg05951221, cg21566642, cg01940273, and cg13193840), including the second most outstanding CpG respective to significance and level of detectable changes in DNA methylation patterns associated with tobacco smoking (cg21566642; KORA F4: -16.70%, p = 6.90E-138, explained variance = 36.24%; KORA F3: -15.58%; p = 8.82E-41, explained variance = 27.13%), were located in a CGI 9kb apart from the 3'UTR of the *ALPPL2* gene, one (cg03329539) within its shore (which means 0 - 2 kb apart from the corresponding CpG island). This CGI is located in the promoter region of the non-coding RNA *ECEL1P1* (endothelin converting enzyme-like 1, pseudogene 1). Within the *ALPP* gene two CpGs (cg23667432, cg03188382) were located in introns (intron 4 and 9), one (cg19713851) in the 3' UTR, all of them either in a CGI itself (cg23667432) or within the shore (cg03188382 and cg19713851). One CpG (cg27241845) was located within the non-coding RNA *ECEL1P2* (endothelin converting enzyme-like 1, pseudogene 2), a few base pairs after the 3'UTR.

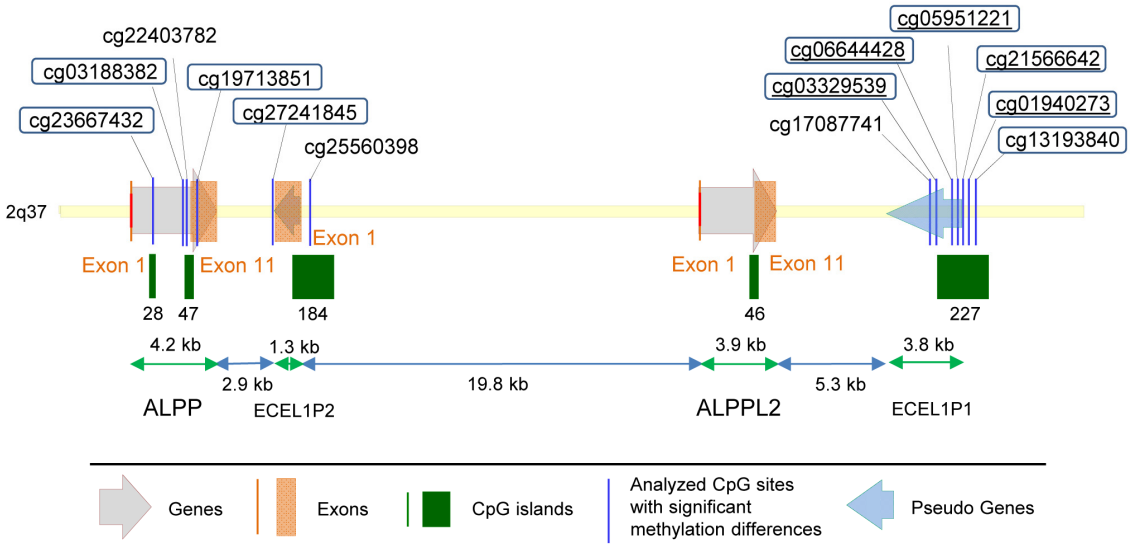


Figure 19. Overview of the results for *ALPP/ALPPL2*. The gene structure and the significantly differentially methylated CpG sites of *ALPP/ALPPL2* (alkaline phosphatase, placental/placental-like) are displayed in current compared to never smokers of the KORA F4 discovery panel. CpG sites which remain significant in the KORA F3 replication panel are framed; CpG sites which were found to be still significant in former smokers are underlined.

3.1.3.3. Effect of tobacco smoking on the methylation status of further loci

An additional 25 CpG sites showed DNA methylation differences of more than 5% (Table 10), located in the genes *HIVEP3, GNG12, GFI1, CACNA1D, TIAM2, MYO1G*, *CNTNAP2, ZC3H3, LRP5, PCDH9, RARA, LINGO-3,* and *F2RL3.* CpG sites at *HIVEP3, CACNA1D, TIAM2*, and PCDH9 were all hypermethylated, sites at *GNG12, GFI1, ZC3H3, LRP5, LINGO-3,* and *F2RL3* were all hypomethylated, whereas loci *MYO1G*, *CNTNAP2, and RARA* showed both, hyper- and hypomethylated sites, in current compared to never smokers. For another 9 CpG sites that were detected, no annotated transcripts are known according to the given annotation file of Illumina. All of them were either located in the shore (cg27241845, cg15693572, cg23480021, cg03274391) or shelf (cg04885881, cg01208318) of a CGI and/or UCSC enhancer binding sites (cg03274391, cg06126421, cg14753356, cg25305703).

3.1.3.4. Meta-analysis of KORA F4 and F3

A meta-analysis of the KORA F4 and F3 data sets was performed to find the general trend for DNA methylation changes due to tobacco smoking across these two panels, and to identify possible disagreements among the results concerning hyper- and hypomethylation of the specific sites. The results are displayed by the corresponding p-value in Table 10 and Table A.2 and furthermore visually presented as a Manhattan Plot in Figure 20. 1498 CpG sites were found to be genome-wide significant ($p \leq 1E\text{-}07$) in the combined analysis of KORA F4 and F3, all 972 significant CpGs of KORA F4 remained genome-wide significant (data not shown). This confirms that the direction of methylation change for the CpG sites in both panels is accordant and underlines the reliability of the gained results.

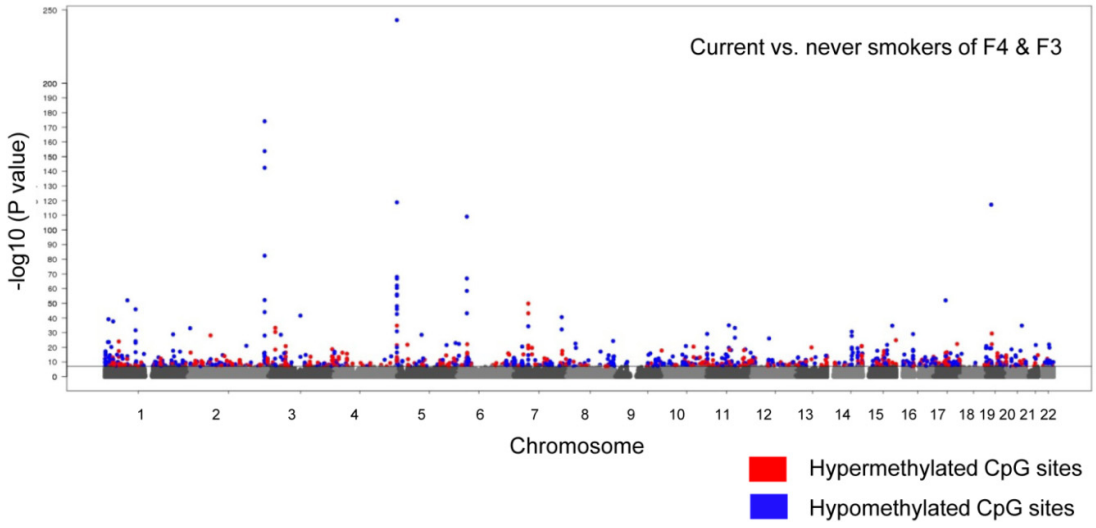


Figure 20. Effect of current tobacco smoking on DNA methylation in the meta-analysis. The meta-analysis included both KORA F4 and F3. The continuous line marks the 1E-07 significance threshold. The significant CpG sites are color coded with the direction of the aberration in current smokers, using blue for hypomethylated and red for hypermethylated CpG sites.

3.1.3.5. Gender-specific effects of tobacco smoking

Women may be more sensitive to some of the harmful effects of smoking, as for example, various studies have shown that tobacco smoking is a higher risk factor for myocardial infarction in women than in men [203,204]. In order to test for gender-specific effects of tobacco smoking on differential DNA methylation, an interaction model was analyzed with the use of the KORA F4 discovery panel where the smoking status * sex interaction was included in the main model. No significant CpG sites were detected for the interaction term, suggesting no difference between males and females in methylation change due to smoking. Nonetheless, as a few studies have detected small gender differences by separately analyzing female and male subjects (please see chapter 4.2.1.4. in the discussion part) this approach was also accomplished in the KORA F4 panel. As men and women showed a considerable difference in the number of pack-years ($p < 0.001$) this model was, in addition to the previously mentioned covariates (except for sex), also adjusted for pack-years. In males, 42 CpGs were found to be differentially methylated in current compared to never smokers (Figure 21a, Table 11). Compared with the general analysis that included both sexes, 35 of these sites have been replicated in KORA F3, 5 were only significant in KORA F4 but not F3, and two sites were found to be only significant in the separate male analysis (cg05498905 and cg00395697). In females, only 10 CpGs were found to be differentially methylated in current compared to never smokers (Figure 21b, Table 11). All but one (cg12806681; $p = 2.00E-05$ in males) were also present within the significant sites of the male analysis and all were replicated in KORA F3.

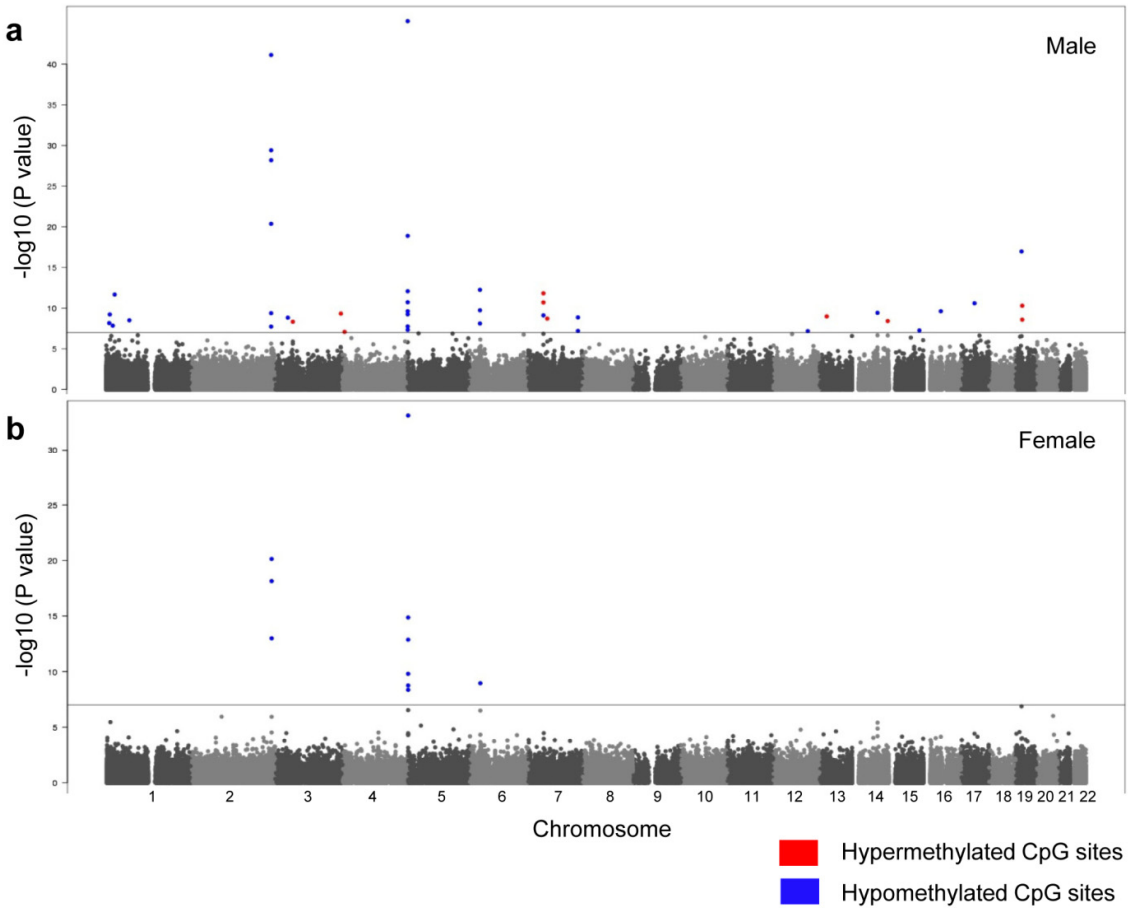


Figure 21. Gender-specific effect of current tobacco smoking on DNA methylation. The continuous lines mark the 1E-07 significance thresholds. The significant CpG sites of **a)** male current vs. never smokers (KORA F4) and **b)** female current vs. never smokers (KORA F4), are color coded with the direction of the aberration in current smokers, using blue for hypomethylated and red for hypermethylated CpG sites.

Table 11. Differentially methylated CpGs of current compared to never smokers in males and females.

			a) Male			**b)** Female			Methylation ß-value as median (first quartile - third quartile)			
									a) Male		**b)** Female	
CpG	Chr	Gene	Median ß-value methylation difference in %	p-value	Explained variance in %	Median ß-value methylation difference in %	p-value	Explained variance in %	Never smokers	Current smokers	Never smokers	Current smokers
cg12547807	1	*x*[a]	-2.97	6.94E-09	3.58	-1.86	7.15E-04	0.35	0.2001 (0.17-0.23)	0.1704 (0.15-0.20)	0.1819 (0.16-0.22)	0.1633 (0.14-0.19)
cg04885881	1	*x*[a]	-8.03	6.03E-10	1.33	-5.90	3.48E-06	1.50	0.3093 (0.24-0.39)	0.2290 (0.18-0.31)	0.3127 (0.25-0.39)	0.2537 (0.18-0.32)
cg19713429	1	*CAPZB*	-1.01	1.43E-08	2.57	-0.18	1.31E-01	0.00	0.1052 (0.10-0.12)	0.0951 (0.09-0.10)	0.0957 (0.09-0.10)	0.0939 (0.08-0.10)
cg27537125	1	*x*[a]	-1.47	2.14E-12	4.84	-1.09	3.16E-04	0.36	0.0897 (0.08-0.10)	0.0750 (0.07-0.08)	0.0869 (0.08-0.10)	0.0760 (0.07-0.09)
cg25189904	1	*GNG12*	-9.16	3.16E-09	2.16	-7.62	6.29E-04	0.97	0.2815 (0.23-0.33)	0.1899 (0.16-0.23)	0.2717 (0.23-0.33)	0.1955 (0.16-0.25)
cg27241845	2	*x*[a]	-8.58	4.22E-10	2.07	-8.51	1.22E-03	1.18	0.4225 (0.38-0.47)	0.3367 (0.29-0.39)	0.4825 (0.43-0.53)	0.3974 (0.36-0.46)
cg03329539	2	*ALPPL2*[b]	-2.79	4.33E-21	6.71	-2.11	1.15E-06	1.12	0.1635 (0.15-0.18)	0.1356 (0.13-0.15)	0.1577 (0.14-0.17)	0.1366 (0.12-0.15)
cg06644428	2	*ALPPL2*[b]	-10.13	1.88E-08	4.88	-5.93	3.00E-05	2.51	0.2542 (0.19-0.32)	0.1529 (0.11-0.20)	0.1989 (0.16-0.24)	0.1396 (0.11-0.18)
cg05951221	2	*ALPPL2*[b]	-5.70	6.61E-29	9.64	-5.38	1.00E-13	4.93	0.2313 (0.21-0.25)	0.1743 (0.16-0.19)	0.2126 (0.19-0.23)	0.1588 (0.15-0.17)
cg21566642	2	*ALPPL2*[b]	-16.35	7.61E-42	9.10	-19.64	7.23E-21	5.01	0.5821 (0.55-0.62)	0.4186 (0.38-0.46)	0.5515 (0.49-0.61)	0.3551 (0.32-0.41)
cg01940273	2	*ALPPL2*[b]	-7.37	3.92E-30	9.33	-7.76	7.09E-19	5.01	0.3360 (0.32-0.36)	0.2623 (0.24-0.29)	0.3470 (0.32-0.37)	0.2694 (0.24-0.29)
cg00501876	3	*CSRNP1*	-5.07	1.50E-09	2.65	-4.02	1.37E-03	0.45	0.3737 (0.34-0.41)	0.3230 (0.29-0.36)	0.3663 (0.33-0.40)	0.3261 (0.30-0.37)
cg15417641	3	*CACNA1D*	10.10	4.77E-09	2.89	7.69	8.30E-03	0.24	0.6583 (0.56-0.75)	0.7593 (0.67-0.83)	0.6884 (0.60-0.77)	0.7653 (0.68-0.83)
cg13185177	3	*GP5*	8.08	4.74E-10	2.41	7.98	6.39E-04	0.46	0.4083 (0.36-0.48)	0.4891 (0.43-0.57)	0.3565 (0.31-0.43)	0.4363 (0.35-0.50)
cg05498905	4	*TBC1D14*	1.49	8.43E-08	3.11	0.16	2.65E-01	0.03	0.8357 (0.81-0.86)	0.8506 (0.83-0.87)	0.8409 (0.82-0.86)	0.8425 (0.83-0.87)
cg12806681	5	*AHRR*	-2.47	2.36E-05	0.71	-2.42	4.34E-09	0.76	0.9540 (0.94-0.96)	0.9293 (0.90-0.94)	0.9535 (0.94-0.96)	0.9293 (0.90-0.95)
cg23916896	5	*AHRR*	-2.72	4.70E-08	3.27	-1.79	5.09E-05	0.85	0.1366 (0.12-0.16)	0.1094 (0.10-0.12)	0.1234 (0.11-0.14)	0.1055 (0.09-0.12)
cg11902777	5	*AHRR*	-1.26	2.41E-10	2.33	-1.05	4.45E-05	0.55	0.0401 (0.03-0.05)	0.0275 (0.02-0.03)	0.0364 (0.03-0.04)	0.0259 (0.02-0.03)
cg01899089	5	*AHRR*	-3.09	1.78E-08	3.79	-2.44	3.45E-05	0.79	0.2549 (0.24-0.28)	0.2240 (0.20-0.25)	0.2413 (0.22-0.26)	0.2169 (0.20-0.24)
cg05575921	5	*AHRR*	-24.63	5.27E-46	9.22	-21.29	7.15E-34	5.57	0.8784 (0.85-0.90)	0.6321 (0.60-0.70)	0.8862 (0.86-0.91)	0.6733 (0.59-0.79)
cg26703534	5	*AHRR*	-6.13	1.90E-11	2.45	-5.95	1.34E-13	3.21	0.6181 (0.60-0.64)	0.5568 (0.51-0.59)	0.6316 (0.61-0.66)	0.5721 (0.53-0.60)
cg14817490	5	*AHRR*	-4.69	8.51E-13	3.22	-3.64	1.76E-09	1.24	0.1447 (0.12-0.17)	0.0978 (0.08-0.12)	0.1323 (0.11-0.15)	0.0959 (0.07-0.11)
cg25648203	5	*AHRR*	-7.75	5.86E-10	0.96	-7.98	1.57E-10	1.75	0.8621 (0.83-0.89)	0.7846 (0.72-0.84)	0.8648 (0.83-0.89)	0.7850 (0.73-0.84)
cg21161138	5	*AHRR*	-11.10	1.32E-19	3.50	-9.19	1.33E-15	2.21	0.6956 (0.66-0.73)	0.5846 (0.53-0.64)	0.7009 (0.67-0.73)	0.6090 (0.55-0.66)
cg06126421	6	*x*[a]	-19.49	5.67E-13	3.23	-14.76	1.10E-09	1.96	0.7555 (0.69-0.81)	0.5606 (0.46-0.67)	0.7824 (0.72-0.83)	0.6348 (0.51-0.71)
cg24859433	6	*x*[a]	-5.07	1.84E-10	1.42	-3.06	4.73E-05	0.87	0.9323 (0.91-0.94)	0.8816 (0.84-0.92)	0.9321 (0.91-0.95)	0.9015 (0.87-0.93)
cg15342087	6	*x*[a]	-3.78	7.80E-09	0.72	-3.09	3.20E-07	1.04	0.9343 (0.92-0.95)	0.8965 (0.86-0.93)	0.9387 (0.92-0.95)	0.9078 (0.87-0.93)
cg22132788	7	*MYO1G*	7.71	1.52E-12	4.70	5.29	1.10E-04	0.33	0.8533 (0.81-0.90)	0.9304 (0.89-0.96)	0.8615 (0.81-0.91)	0.9144 (0.87-0.95)
cg12803068	7	*MYO1G*	17.43	2.02E-11	5.38	11.25	5.91E-03	0.08	0.7278 (0.59-0.83)	0.9021 (0.81-0.94)	0.7468 (0.61-0.86)	0.8593 (0.73-0.94)
cg07826859	7	*MYO1G*	-5.75	7.79E-10	2.72	-4.55	3.40E-05	0.83	0.5746 (0.52-0.62)	0.5171 (0.46-0.56)	0.5911 (0.55-0.63)	0.5456 (0.50-0.59)
cg19956914	7	*SUMF2*	8.34	1.96E-09	3.06	6.11	2.16E-03	0.35	0.5277 (0.48-0.58)	0.6111 (0.55-0.65)	0.5570 (0.50-0.61)	0.6181 (0.55-0.66)
cg21322436	7	*CNTNAP2*	-6.07	6.37E-08	3.24	-4.20	1.27E-02	0.00	0.3099 (0.28-0.34)	0.2492 (0.22-0.28)	0.2732 (0.24-0.30)	0.2312 (0.20-0.26)
cg25949550	7	*CNTNAP2*	-1.70	1.39E-09	2.45	-1.25	6.35E-03	0.43	0.0894 (0.08-0.10)	0.0724 (0.06-0.08)	0.0834 (0.08-0.09)	0.0709 (0.06-0.08)
cg00395697	12	*UHRF1BP1L*	-0.32	6.63E-08	2.92	0.02	7.72E-01	0.00	0.0256 (0.02-0.03)	0.0224 (0.02-0.03)	0.0242 (0.02-0.03)	0.0244 (0.02-0.03)
cg06648759	13	*x*[a]	4.19	1.05E-09	3.07	0.64	1.93E-01	0.21	0.4991 (0.45-0.55)	0.5410 (0.50-0.59)	0.5467 (0.51-0.59)	0.5531 (0.51-0.60)
cg24996979	14	*C14orf43*	-0.75	3.81E-10	4.35	-0.52	6.46E-05	0.94	0.0829 (0.08-0.09)	0.0754 (0.07-0.08)	0.0787 (0.07-0.09)	0.0735 (0.07-0.08)
cg14977938	14	*ZFYVE21*	4.97	3.88E-09	4.38	4.09	1.69E-02	0.05	0.6635 (0.62-0.70)	0.7132 (0.68-0.74)	0.6633 (0.62-0.70)	0.7042 (0.68-0.74)
cg23161492	15	*ANPEP*	-1.96	5.51E-08	2.78	-1.43	1.31E-01	0.00	0.1253 (0.11-0.14)	0.1057 (0.09-0.12)	0.1189 (0.11-0.13)	0.1046 (0.10-0.12)
cg09099830	16	*ITGAL*	-1.76	2.41E-10	3.06	-3.31	7.28E-05	1.33	0.6102 (0.59-0.63)	0.5926 (0.57-0.61)	0.5806 (0.55-0.61)	0.5475 (0.53-0.58)
cg19572487	17	*RARA*	-10.03	2.53E-11	2.29	-9.53	3.81E-05	1.12	0.4457 (0.40-0.52)	0.3454 (0.29-0.43)	0.4488 (0.39-0.51)	0.3535 (0.30-0.44)
cg03636183	19	*F2RL3*	-16.15	1.09E-17	2.80	-13.26	1.34E-07	1.27	0.4925 (0.45-0.54)	0.3310 (0.28-0.41)	0.4932 (0.45-0.54)	0.3606 (0.29-0.44)
cg23973524	19	*CRTC1*	3.74	5.05E-11	2.73	4.14	9.89E-05	0.38	0.3198 (0.30-0.35)	0.3572 (0.33-0.39)	0.2965 (0.27-0.33)	0.3379 (0.31-0.37)
cg21473814	19	*CRTC1*	5.37	2.64E-09	1.53	4.74	7.33E-03	0.15	0.3300 (0.30-0.37)	0.3837 (0.34-0.42)	0.3180 (0.28-0.36)	0.3654 (0.32-0.41)

The results of the linear model calculated with M-value adjusted for age, BMI, alcohol, white blood cell count, and pack-years (p-value and explained variance), as well as the median ß-value methylation difference between current and never smokers are displayed for the **a)** male and **b)** female subpopulation of F4 with genome-wide significance ($p \leq 1E-07$); sorted by chromosome and mapinfo (Genome build 37). [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

3.1.4. Effect of former tobacco smoking on genome-wide DNA methylation

To investigate if these changes in DNA methylation remain after quitting tobacco smoking, the DNA methylation level of former smokers compared to never smokers was analyzed in the KORA F4 panel (former N=782, never N=749; see Table 9 for characteristics of the study population). The results are shown in Figure 22.

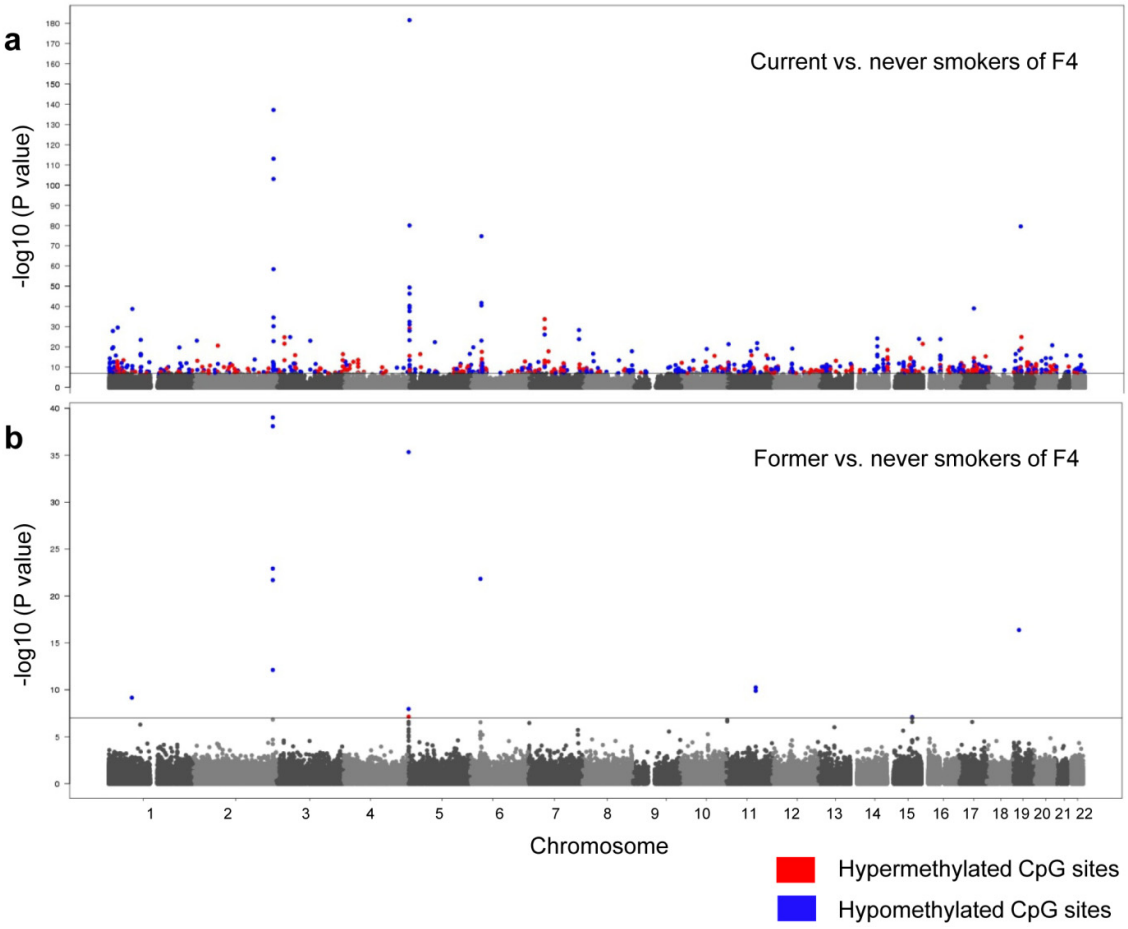


Figure 22. Genome-wide effect of former tobacco smoking on DNA methylation. The continuous lines in the Manhattan Plots mark the 1E-07 significance thresholds. The significant CpG sites of **a)** current vs. never smokers (KORA F4) and **b)** former vs. never smokers (KORA F4), are color coded with the direction of the aberration in former and current smokers, respectively, using blue for hypomethylated and red for hypermethylated CpG sites.

In former smokers, the methylation levels of 174 of the 187 replicated current smoking-associated CpG sites were almost comparable to the state found in never smokers. However, 13 of the 187 CpGs still showed differential methylation levels in former smokers compared to never smokers, although differences were less pronounced (Table 12). Except for cg03604011, all of the significant CpGs in former smokers were hypomethylated compared to never smokers (Figure 22 and Table 12).

Table 12. Differentially methylated CpG sites of former compared to never smokers (KORA F4).

			a) Former smokers			**b)** Current smokers			Methylation ß-value as median (first quartile - third quartile)		
CpG	Chr.	Gene	Median ß-value methylation difference in %	p-value	Explained variance in %	Median ß-value methylation difference in %	p-value	Explained variance in %	Never smokers	Former smokers	Current smokers
cg25189904	1	*GNG12*	-2.36%	6.85E-10	2.54	-8.19%	1.71E-39	14.27	0.2740 (0.23-0.33)	0.2504 (0.21-0.31)	0.1921 (0.16-0.24)
cg03329539	2	*ALPPL2*[b]	-0.61%	7.58E-13	3.26	-2.39%	3.66E-59	19.04	0.1598 (0.15-0.17)	0.1537 (0.14-0.17)	0.1359 (0.13-0.15)
cg06644428	2	*ALPPL2*[b]	-3.09%	1.20E-23	6.20	-7.05%	6.37E-31	12.14	0.2142 (0.17-0.27)	0.1833 (0.14-0.23)	0.1437 (0.11-0.19)
cg05951221	2	*ALPPL2*[b]	-1.87%	8.18E-39	9.77	-5.21%	8.92E-104	31.07	0.2199 (0.20-0.24)	0.2012 (0.18-0.22)	0.1678 (0.15-0.19)
cg21566642	2	*ALPPL2*[b]	-4.31%	9.49E-40	10.22	-16.70%	6.90E-138	36.24	0.5658 (0.51-0.61)	0.5227 (0.46-0.57)	0.3988 (0.35-0.45)
cg01940273	2	*ALPPL2*[b]	-2.26%	2.01E-22	5.56	-7.89%	9.28E-114	31.50	0.3437 (0.32-0.37)	0.3211 (0.30-0.35)	0.2648 (0.24-0.29)
cg11554391	5	*AHRR*	-0.15%	1.10E-08	2.28	-1.03%	5.07E-24	7.71	0.0911 (0.08-0.10)	0.0895 (0.08-0.10)	0.0808 (0.07-0.09)
cg05575921	5	*AHRR*	-3.31%	4.62E-36	9.25	-24.40%	2.54E-182	41.02	0.8841 (0.86-0.90)	0.8510 (0.79-0.89)	0.6401 (0.60-0.73)
cg03604011	5	*AHRR*	1.58%	7.21E-08	1.84	3.40%	7.87E-30	9.49	0.0759 (0.06-0.10)	0.0917 (0.07-0.12)	0.1099 (0.08-0.15)
cg06126421	6	*x*[a]	-5.65%	1.50E-22	5.34	-17.05%	1.72E-75	23.60	0.7724 (0.70-0.82)	0.7159 (0.63-0.79)	0.6019 (0.48-0.69)
cg11660018	11	*PRSS23*	-1.67%	1.28E-10	2.46	-3.87%	1.29E-22	6.88	0.2926 (0.27-0.32)	0.2759 (0.25-0.30)	0.2538 (0.23-0.28)
cg23771366	11	*PRSS23*	-0.98%	5.55E-11	2.93	-2.26%	7.62E-20	7.05	0.1993 (0.18-0.22)	0.1895 (0.17-0.21)	0.1767 (0.16-0.19)
cg03636183	19	*F2RL3*	-3.48%	4.21E-17	4.75	-14.74%	2.42E-80	22.45	0.4930 (0.45-0.54)	0.4582 (0.40-0.51)	0.3456 (0.28-0.43)

Displayed are **a)** the results of the linear model calculated with M-value adjusted for age, sex, BMI, alcohol, and white blood cell count (p-value and explained variance), as well as the median ß-value methylation difference between former and never smokers of KORA F4 with genome-wide significance ($p \leq 1E-07$) and **b)** the corresponding results of the same CpG sites for current smokers; sorted by chromosome and mapinfo (Genome build 37). [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

3.1.4.1. The effect of cessation time on DNA methylation

As the DNA methylation levels of former tobacco smokers were found to be close to the ones seen in never smokers, it was of interest which role cessation time and pack-years may play in that matter. The time course over which DNA methylation is subject to change is not known, but it is assumed that it occurs in a CpG site-specific manner. Therefore, the effect of time after quitting smoking on the degree of DNA methylation in former smokers of the KORA F4 panel was assessed. This was found to be significant in 36 of the 187 CpGs (p = 8.44E-08 - 7.73E-44, explained variance = 3.15 - 21.48%; Table 13).

Table 13. Differentially methylated CpGs in relation to cessation time.

CpG	Chr.	Gene	Median methylation ß-value in % Current smokers	Median methylation ß-value in % Never smokers	Median methylation ß-value in % Former smokers	p-value calculated with M-value (linear model)	Explained variance calculated with M-value in %
cg25189904	1	*GNG12*	19.21	27.40	25.04	5.19E-13	6.80
cg23079012	2	*x*[a]	94.84	95.94	95.80	2.48E-14	7.49
cg03329539	2	*ALPPL2*[b]	13.59	15.98	15.37	1.92E-14	7.22
cg06644428	2	*ALPPL2*[b]	14.37	21.42	18.33	5.71E-12	5.80
cg05951221	2	*ALPPL2*[b]	16.78	21.99	20.12	1.06E-29	14.92
cg21566642	2	*ALPPL2*[b]	39.88	56.58	52.27	7.26E-34	16.75
cg01940273	2	*ALPPL2*[b]	26.48	34.37	32.11	5.94E-22	11.29
cg15693572	3	*x*[a]	40.29	29.42	32.53	4.84E-09	4.45
cg03274391	3	*x*[a]	49.97	31.90	37.32	2.15E-08	4.10
cg11554391	5	*AHRR*	8.08	9.11	8.95	1.91E-09	4.51
cg17924476	5	*AHRR*	32.08	26.28	27.95	6.03E-08	3.83
cg03991871	5	*AHRR*	88.96	94.69	93.59	4.14E-12	6.16
cg11902777	5	*AHRR*	2.67	3.79	3.37	3.97E-11	5.71
cg05575921	5	*AHRR*	64.01	88.41	85.10	7.73E-44	21.48
cg14817490	5	*AHRR*	9.68	13.50	13.33	2.79E-08	3.96
cg21161138	5	*AHRR*	59.54	69.98	68.05	9.81E-15	7.57
cg03604011	5	*AHRR*	10.99	7.59	9.17	2.29E-09	4.58
cg06126421	6	*x*[a]	60.19	77.24	71.59	3.84E-24	11.03
cg14753356	6	*x*[a]	18.46	23.83	21.57	4.60E-08	3.28
cg24859433	6	*x*[a]	89.19	93.22	92.46	8.17E-10	4.55
cg15342087	6	*x*[a]	90.36	93.71	93.16	7.48E-08	3.45
cg23565821	6	*CUTA*	14.55	14.72	15.09	6.29E-08	3.15
cg12803068	7	*MYO1G*	88.78	73.82	76.30	7.47E-11	5.51
cg21322436	7	*CNTNAP2*	23.81	28.55	27.95	1.35E-08	3.94
cg25949550	7	*CNTNAP2*	7.14	8.52	8.31	2.31E-09	4.57
cg24540678	8	*x*[a]	7.87	8.55	8.36	3.79E-10	4.86
cg11660018	11	*PRSS23*	25.38	29.26	27.59	7.45E-10	4.54
cg23771366	11	*PRSS23*	17.67	19.93	18.95	3.56E-13	6.91
cg09084200	11	*VPS26B/NCAPD3*	14.32	14.85	14.93	8.44E-08	3.22
cg02583484	12	*HNRNPA1*	12.67	14.76	14.22	3.56E-08	3.81
cg23161492	15	*ANPEP*	10.53	12.12	11.75	4.82E-08	3.84
cg19572487	17	*RARA*	34.75	44.77	42.73	1.71E-08	4.20
cg03636183	19	*F2RL3*	34.56	49.30	45.82	1.73E-32	16.88
cg12303084	20	*ZMYND8*	9.28	9.91	10.00	2.33E-08	3.93
cg00871610	21	*MIR802*	48.34	52.67	50.85	2.18E-12	6.01
cg01127300	22	*x*[a]	41.93	45.76	44.69	1.15E-08	4.02

Displayed are the results of the linear model for time since quitting with genome-wide significance level p ≤ 1E-07, calculated with M-value and adjusted for age, sex, BMI, alcohol, and white blood cell count (p-value and explained variance), including former smokers of KORA F4 only, as well as the median ß-value methylation levels for current, never and former smokers; CpG sites are sorted by chromosome and mapinfo (Genome build 37); [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

To get an impression of the time period that may be needed for former smokers to achieve the median ß-value methylation state of never smokers, a smooth curve was plotted in the scatter plot. Years required for former smokers to obtain a median ß-value methylation state that is closer to or equals the one of never smokers are visualized for the most significant CpG site, cg05575921 of *AHRR,* by scatter plot in Figure 23, for the other 35 significant CpGs by scatter plots in Figure A.3. While in the majority of cases a relatively fast approach to the level of never smokers could be detected in former smokers who have quit recently, this seemed to slow down substantially depending on how many years or decades ago a person quit smoking. The period of time required for the 36 CpGs varied from about 1 to 20 years to gain a methylation state that is 50% closer to the DNA methylation state of never smokers and about 5 to 60 years to completely achieve the DNA methylation state of never smokers, but with 12 of the 36 CpGs never reaching the 95% line.

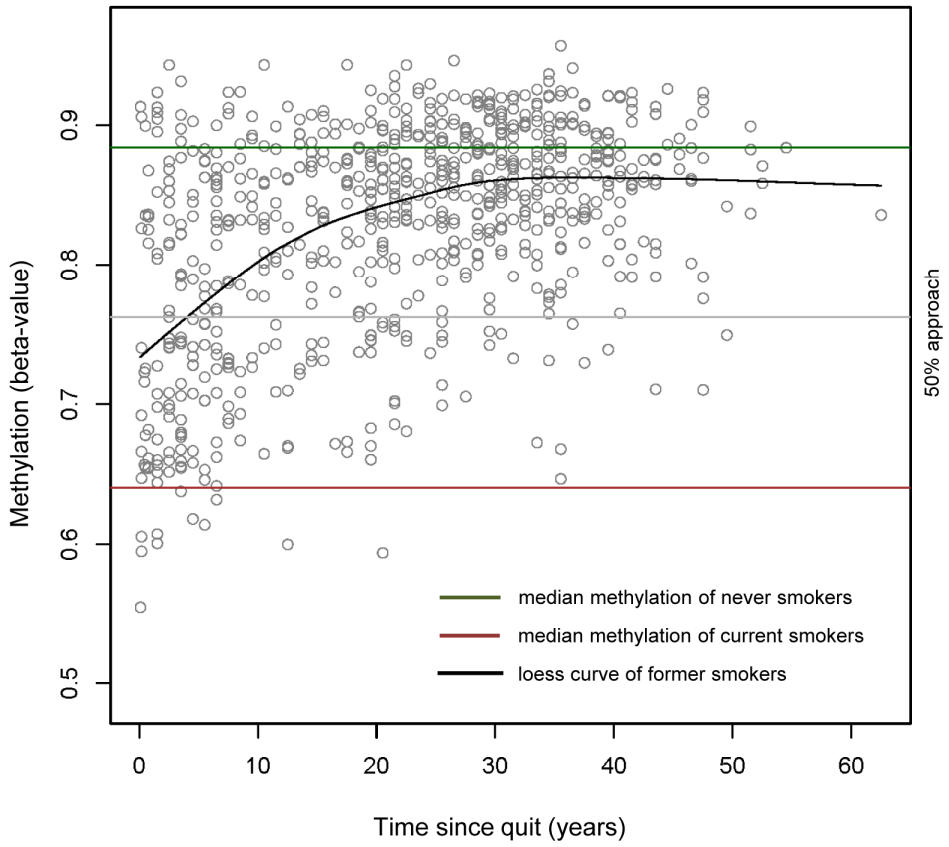


Figure 23. Effect of cessation time on the DNA methylation state of cg05575921. The years required for former smokers to obtain a median ß-value methylation state at CpG site cg05575921 that is closer to or equals the one of never smokers is illustrated by a loess curve in the scatterplot; the x-axis displays the cessation time in years, the y-axis displays the methylation level with the use of numbers between 0 (for 0% methylation) and 1 (for 100% methylation); horizontal brown line: median methylation level of current smokers; horizontal green line: median methylation level of never smokers; horizontal grey line: center line of current and never smokers median ß-value methylation; please see Table 13 for detailed data.

3.1.4.2. The effect of pack-years on DNA methylation

It has been shown that dynamic changes in DNA methylation not only occur in response to cessation time, but also in response to cumulative smoke exposure (pack-years). Therefore, the linear effect of pack-years on the degree of DNA methylation in the former smokers of the KORA F4 panel was assessed and found to be significant in 14 of the 187 CpGs. All 14 CpGs were also significant in time since quitting and replicated in KORA F3 in the smoking model (Table 14). The number of pack-years needed for former smokers to reach a median ß-value methylation state that is closer to or equals the one of current smokers is visualized by scatterplots for the most significant CpG site cg05575921 of *AHRR* in Figure 24, and for the other 13 ones in Figure A.4.

Table 14. Differentially methylated CpGs in relation to pack-years.

CpG	Chr.	Gene	Median methylation ß-value in % Current smokers	Median methylation ß-value in % Never smokers	Median methylation ß-value in % Former smokers	p-value calculated with M-value (linear model)	Explained variance calculated with M-value in %
cg25189904	1	*GNG12*	19.21	27.40	25.04	8.44E-10	5.28
cg23079012	2	*x*[a]	94.84	95.94	95.80	4.27E-08	4.17
cg06644428	2	*ALPPL2*[b]	14.37	21.42	18.33	5.71E-08	3.85
cg05951221	2	*ALPPL2*[b]	16.78	21.99	20.12	3.71E-21	11.33
cg21566642	2	*ALPPL2*[b]	39.88	56.58	52.27	8.11E-21	11.05
cg01940273	2	*ALPPL2*[b]	26.48	34.37	32.11	1.06E-14	7.91
cg11554391	5	*AHRR*	8.08	9.11	8.95	5.91E-08	3.91
cg05575921	5	*AHRR*	64.01	88.41	85.10	2.44E-27	14.68
cg21161138	5	*AHRR*	59.54	69.98	68.05	9.29E-08	3.96
cg06126421	6	*x*[a]	60.19	77.24	71.59	1.29E-17	8.49
cg14753356	6	*x*[a]	18.46	23.83	21.57	5.12E-08	3.44
cg11660018	11	*PRSS23*	25.38	29.26	27.59	1.14E-10	5.25
cg23771366	11	*PRSS23*	17.67	19.93	18.95	4.41E-13	7.26
cg03636183	19	*F2RL3*	34.56	49.30	45.82	2.15E-17	9.60

Displayed are the results of the linear model for pack-years with genome-wide significance level $p \leq 1E-07$, calculated with M-value and adjusted for age, sex, BMI, alcohol, and white blood cell count (p-value and explained variance), including former smokers of KORA F4 only, as well as the median ß-value methylation levels for current, never and former smokers; sorted by chromosome and mapinfo (Genome build 37). [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

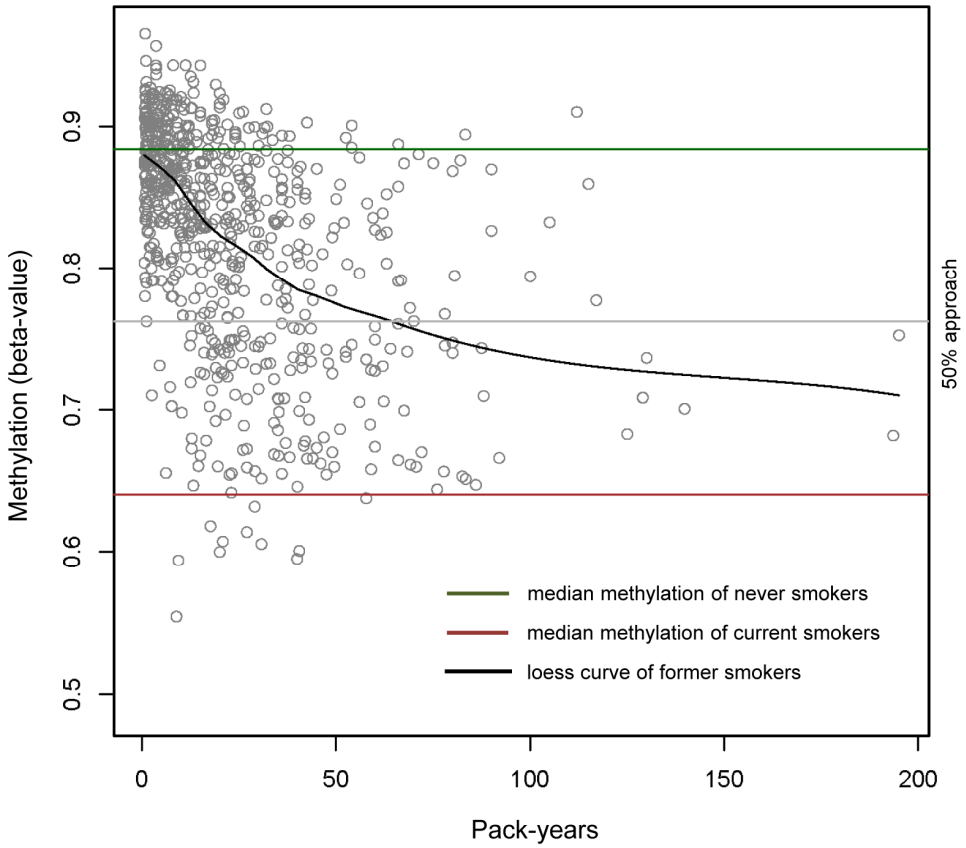


Figure 24. Effect of pack-years on the DNA methylation state of cg05575921. The pack-years required for former smokers to obtain a median ß-value methylation state at CpG site cg05575921 that is closer to or equals the one of current smokers is displayed by a loess curve in the scatterplot; the x-axis displays the number of pack- years, the y-axis displays the methylation level with the use of numbers between 0 (for 0% methylation) and 1 (for 100% methylation); horizontal brown line: median methylation of current smokers in %; horizontal green line: median methylation of never smokers; horizontal grey line: center line of current and never smokers median ß-value methylation; please see Table 14 for detailed data.

3.1.4.3. Combined effect of cessation time and pack-years on DNA methylation

Since both cessation time and cumulative smoke exposure showed an effect on the DNA methylation state, it was of interest if a correlation between these two variables and DNA methylation exists. Therefore, another model was calculated that included both 'time since quitting' and 'pack-years' to analyze the combined effect of these variables. This approach showed that the combination of these two variables had an influence on the DNA methylation state of former smokers (Table 15). The longer the cessation time the smaller the effect of smoking on DNA methylation, therefore approaching the DNA methylation level of never smokers. On the contrary, the more cumulative smoke exposure the larger the effect of smoking on DNA methylation, therefore departing from the DNA methylation level of never smokers. An overall trend could be observed with the methylation levels of subjects with the longest cessation time and the lowest cumulative smoke exposure being closest to the levels detected in never smokers, as can be seen by the opposed effect estimate in Table 15. Three CpGs were found to have genome-wide significance in both variables and interestingly represent top hits of both the *AHRR* (cg05575921) and *ALPP/ALPPL2* (cg05951221, cg21566642) regions. The interplay between cessation time and cumulative smoke exposure for these three sites is illustrated in Figure 25.

Additionally, an interaction model defined two CpGs that showed genome-wide significance between time since quitting and pack-years (cg24128853, p = 2.80E-08, effect of interaction = 0.00029; cg24504601, p = 7.44E-08, effect of interaction = 0.00040). However, these two CpGs were neither among the 36 that were found to have a significant effect of time after smoking cessation on the degree of DNA methylation in the former smokers nor in the combined model. Furthermore, these two sites were not found in the general smoking model and were therefore not further observed.

Table 15. The combined effect of cessation time and pack-years on DNA methylation.

CpG	Chr.	Mapinfo	Gene	**a)** Effect estimate time since quitting	p-value	**b)** Effect estimate pack-years	p-value	**c)** Effect estimate time since quitting adjusted	p-value	**d)** Effect estimate pack-years adjusted	p-value
cg25189904	1	68299493	*GNG12*	0.012774476	5.19E-13	-0.006086506	8.44E-10	0.009230685	1.06E-05	-0.004153401	1.08E-04
cg23079012	2	8343710	*x*[a]	0.010826481	2.48E-14	-0.004375023	4.27E-08	0.010540617	4.15E-10	-0.002100705	1.36E-02
cg03329539	2	233283329	*ALPPL2*[b]	0.004853206	1.92E-14	-0.001599676	1.02E-05	0.004312098	1.46E-08	-0.000589666	1.27E-01
cg06644428	2	233284112	*ALPPL2*[b]	0.0139032	5.71E-12	-0.006133357	5.71E-08	0.008303562	4.72E-04	-0.004345463	3.64E-04
cg05951221	2	233284402	*ALPPL2*[b]	0.008476554	1.06E-29	-0.00394502	3.71E-21	0.006052889	1.03E-12	-0.002577901	2.69E-09
cg21566642	2	233284661	*ALPPL2*[b]	0.015551704	7.26E-34	-0.006819629	8.11E-21	0.012005841	2.69E-16	-0.004161692	2.73E-08
cg01940273	2	233284934	*ALPPL2*[b]	0.007372237	5.94E-22	-0.003405517	1.06E-14	0.006098512	8.53E-12	-0.002035633	7.23E-06
cg15693572	3	22412385	*x*[a]	-0.016116999	4.84E-09	0.006143252	8.71E-05	-0.015523785	3.37E-06	0.002974916	8.06E-02
cg03274391	3	22413232	*x*[a]	-0.02019266	2.15E-08	0.009350908	4.30E-06	-0.018369508	2.21E-05	0.005562546	1.17E-02
cg11554391	5	321320	*AHRR*	0.004561621	1.91E-09	-0.002284664	5.91E-08	0.00338126	1.55E-04	-0.001592762	5.10E-04
cg17924476	5	323794	*AHRR*	-0.007422749	6.03E-08	0.003957882	1.12E-07	-0.00535159	7.00E-04	0.002844057	4.49E-04
cg03991871	5	368447	*AHRR*	0.019979938	4.14E-12	-0.008462903	2.11E-07	0.017795122	2.69E-07	-0.004555672	9.73E-03
cg11902777	5	368843	*AHRR*	0.00924539	3.97E-11	-0.003599043	3.68E-06	0.007017155	2.31E-05	-0.002089958	1.36E-02
cg05575921	5	373378	*AHRR*	0.028027825	7.73E-44	-0.012681468	2.44E-27	0.023024556	3.75E-23	-0.007578304	8.40E-11
cg14817490	5	392920	*AHRR*	0.007449846	2.79E-08	-0.003241823	1.34E-05	0.006754998	1.90E-05	-0.001889137	1.92E-02
cg21161138	5	399360	*AHRR*	0.009076502	9.81E-15	-0.003484034	9.29E-08	0.006927905	4.48E-07	-0.002083254	2.92E-03
cg03604011	5	400201	*AHRR*	-0.012108191	2.29E-09	0.004895609	1.41E-05	-0.009055977	1.68E-04	0.002904678	1.83E-02
cg06126421	6	30720080	*x*[a]	0.023261814	3.84E-24	-0.011055098	1.29E-17	0.018513487	2.61E-12	-0.007045216	1.69E-07
cg14753356	6	30720108	*x*[a]	0.007424528	4.60E-08	-0.004017977	5.12E-08	0.005502075	3.93E-04	-0.002892271	2.83E-04
cg24859433	6	30720203	*x*[a]	0.010720984	8.17E-10	-0.004113007	3.06E-05	0.009509394	5.56E-06	-0.002215507	3.80E-02
cg15342087	6	30720209	*x*[a]	0.009101885	7.48E-08	-0.003380871	4.79E-04	0.008747995	1.59E-05	-0.001513031	1.43E-01
cg23565821	6	33385056	*CUTA*	0.003758547	6.29E-08	-0.001129069	3.74E-03	0.003743634	6.29E-06	-0.000308437	4.65E-01
cg12803068	7	45002919	*MYO1G*	-0.025219825	7.47E-11	0.008051484	1.77E-04	-0.020526508	6.37E-06	0.003999853	8.45E-02
cg21322436	7	145812842	*CNTNAP2*	0.006851939	1.35E-08	-0.001879379	4.73E-03	0.006431848	5.89E-06	-0.000634697	3.80E-01
cg25949550	7	145814306	*CNTNAP2*	0.005014016	2.31E-09	-0.001751718	1.66E-04	0.004164323	2.07E-05	-0.001020342	4.10E-02
cg24540678	8	28258603	*x*[a]	0.004009471	3.79E-10	-0.001724378	1.03E-06	0.003674462	9.69E-07	-0.000958047	1.22E-02
cg11660018	11	86510915	*PRSS23*	0.005136447	7.45E-10	-0.003069423	1.14E-10	0.004017206	3.65E-05	-0.002267364	5.73E-06
cg23771366	11	86510998	*PRSS23*	0.005547923	3.56E-13	-0.003086321	4.41E-13	0.004320094	1.02E-06	-0.002169611	1.71E-06
cg09084200	11	134095863	*VPS26B/NCAPD3*	0.00316939	8.44E-08	-0.001243043	1.35E-04	0.00277668	6.13E-05	-0.00062632	7.67E-02
cg02583484	12	54677008	*HNRNPA1*	0.006888194	3.56E-08	-0.002766366	6.48E-05	0.005446731	1.97E-04	-0.001708307	2.25E-02
cg23161492	15	90357202	*ANPEP*	0.004303915	4.82E-08	-0.001927201	8.74E-06	0.003704331	6.15E-05	-0.001213823	1.02E-02
cg19572487	17	38476024	*RARA*	0.0104466	1.71E-08	-0.003762921	3.61E-04	0.008555266	1.19E-04	-0.001912849	9.59E-02
cg03636183	19	17000585	*F2RL3*	0.016911144	1.73E-32	-0.006797177	2.15E-17	0.01517163	1.26E-21	-0.003498853	9.94E-06
cg12303084	20	45985741	*ZMYND8*	0.005218624	2.33E-08	-0.001787353	5.80E-04	0.004480121	5.36E-05	-0.000913024	1.07E-01
cg00871610	21	37093012	*MIR802*	0.010771423	2.18E-12	-0.003638815	2.88E-05	0.009179189	3.06E-07	-0.001530863	9.33E-02
cg01127300	22	38614796	*x*[a]	0.009074592	1.15E-08	-0.003493628	6.08E-05	0.007891918	1.77E-05	-0.001889634	4.40E-02

The results of the linear model are displayed, calculated with M-value adjusted for age, sex, BMI, alcohol, and white blood cell count for former smokers of KORA F4 for **a)** time since quit, **b)** pack-years, **c)** time since quit after adjustment for pack-years, and **d)** pack-years after adjustment for time since quit with a genome-wide significance level p ≤ 1E-07; sorted by chromosome and mapinfo (Genome build 37). [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

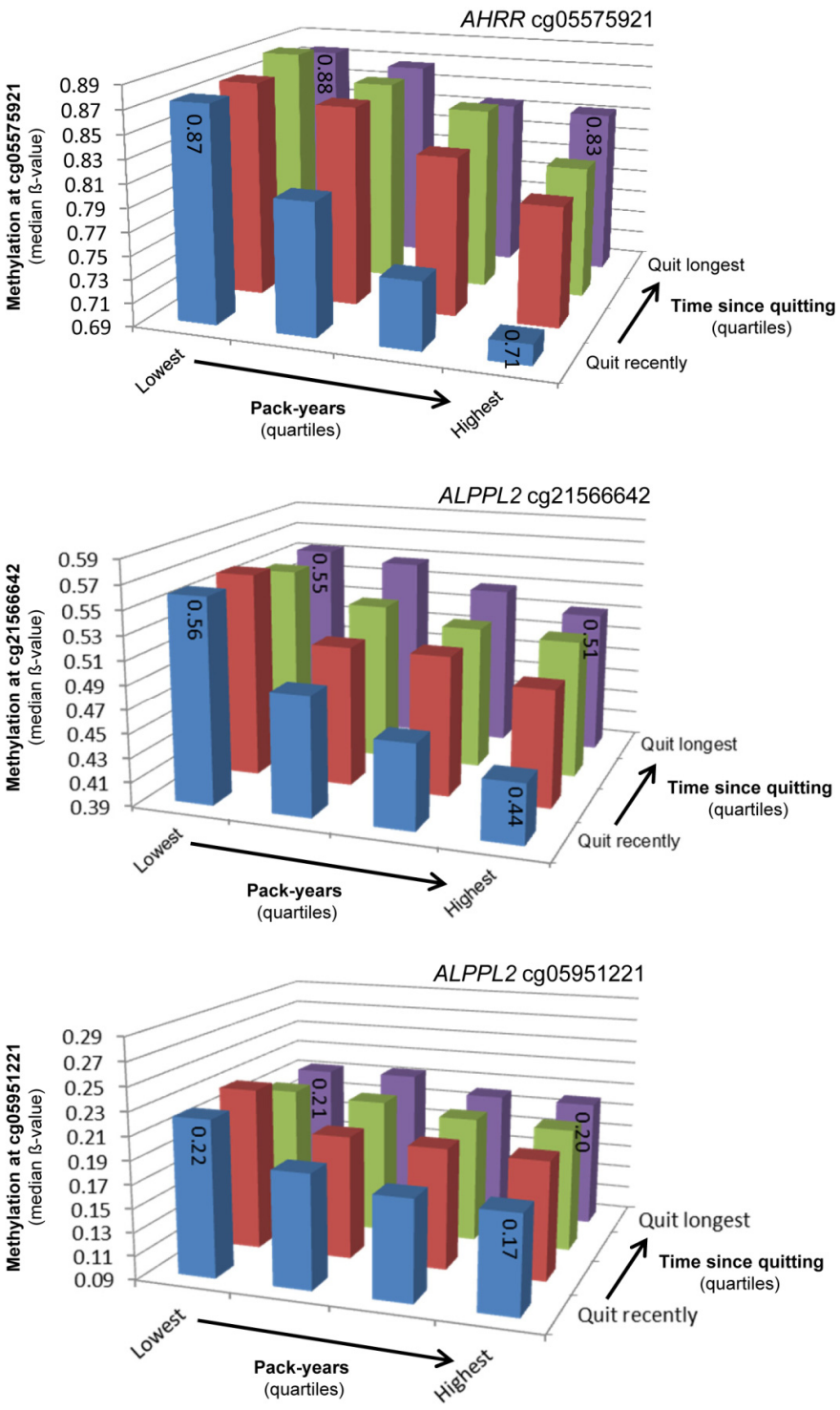


Figure 25. Combined effect of cessation time and pack-years on the DNA methylation state. Displayed are the CpG sites cg05575921, cg21566642, and cg05951221. Quartiles of time since quitting (in years) are plotted on the x-axis, quartiles of pack-years (cumulative smoke exposure) are plotted on the z-axis, while the median methylation of each quartile (in the range from 0, unmethylated, to 1, 100% methylated) is plotted on the y-axis.

3.1.5. Genomic distribution of smoking-associated significant CpGs

The following chapter presents noticeable observations within smoking-associated CpG sites concerning their direction of methylation change (hypo- or hypermethylation) and their location in relation to RefGeneGroups and CpG islands.

3.1.5.1. Distribution of hypo- and hypermethylated CpGs

The most prominent alterations in DNA methylation that lead to disease include generalized genome-wide hypomethylation and locus-specific hypermethylation, especially in the case of cancer. Therefore, the distribution of significant smoking-associated CpGs concerning hyper- and hypomethylation was of interest. By taking the 187 replicated CpG sites into account, about three quarters (72.7%) were hypomethylated in current compared to never smokers. Also, by taking a cutoff of 2.5% and 5% differential DNA methylation, hypomethylation in current smokers remained within about two thirds (65.7% and 61.1%, respectively) of the significant CpG sites. As for the former smokers, 12 of the significant CpG sites were hypomethylated compared to never smokers, only cg03604011 was hypermethylated (Figure 26).

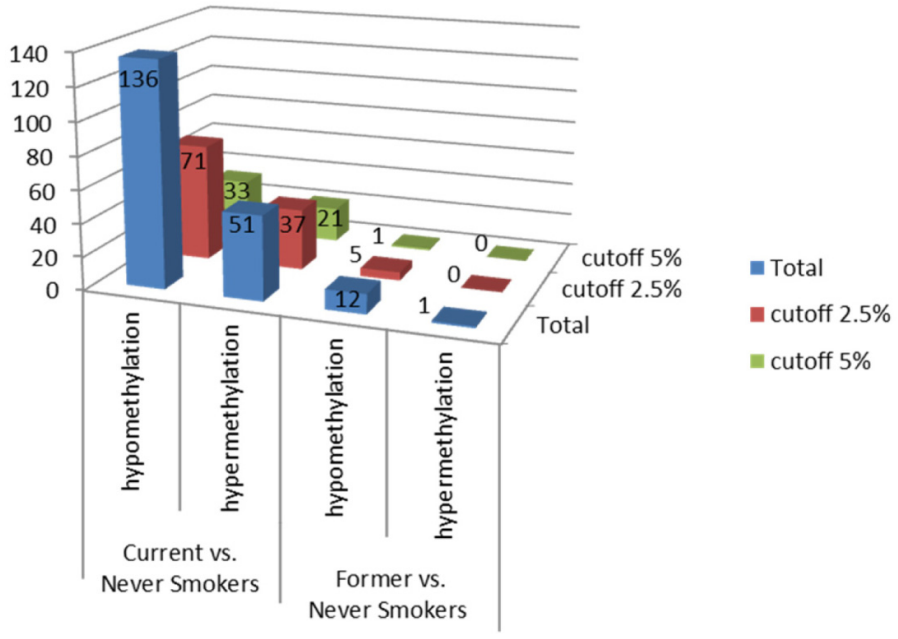


Figure 26. Distribution of hypo- and hypermethylated CpG sites in current and former smokers. Displayed are the absolute numbers of CpG sites found to be significant in current and former smokers compared to never smokers in relation to their direction of DNA methylation change (hypo- or hypermethylated). The various groups of hypo- and hypermethylation are plotted on the x-axis, the various groups depending on the percent of methylation change found within the sites are plotted on the z-axis, while the numbers of corresponding CpG sites within the single groups are plotted on the y-axis.

3.1.5.2. Relation to RefGeneGroups

As DNA methylation can occur at all regions of a gene, it was investigated whether the significant smoking-associated CpGs were targeted to certain gene sequences. At the gene level, the 450K BeadChip covers 99% of RefSeq genes with multiple sites in the annotated promoter (TSS200 and TSS1500; 200 bp or 1500 bp upstream of transcription start site), 5' UTR, 1st Exon, gene body and 3' UTR.

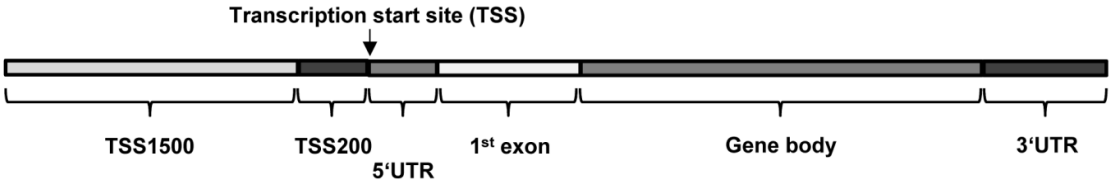
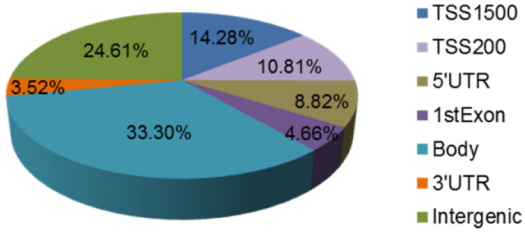


Figure 27 displays the functional genomic distribution of all 468316 CpGs analyzed, of the significant and replicated 187 CpGs associated with current smoking, and of the significant 13 CpGs associated with former smoking.

Comparing the distribution within RefGeneGroups of smoking-associated sites (N=187) with the overall analyzed sites of the 450K BeadChip (N= 468316), more smoking-associated sites were located in gene bodies (49.73% vs. 33.30%). On the other hand, less smoking-associated sites were located in TSS200 (1.07% vs. 10.81%), TSS1500 (10.70% vs. 14.28%), and 1st Exon (1.07% vs. 4.66%) than would be expected by percentage (Figure 27 a & b). Within the CpGs that remained significant in former smokers, an overrepresentation of sites located in the intergenic region was detected with about twice as many sites as would have been expected (46.15% vs. 24.61%).

a

FGD	CpGs	Subgroup	CpGs
Promoter	180641 (38.57%)	TSS1500	66870 (14.28%)
		TSS200	50617 (10.81%)
		5'UTR	41310 (8.82%)
		1stExon	21844 (4.66%)
Body	21844 (33.30%)		
3'UTR	16467 (3.52%)		
Intergenic	115269 (24.61%)		



b

FGD	CpGs	Subgroup	CpGs
Promoter	45 (24.07%)	TSS1500	20 (10.70%)
		TSS200	2 (1.07%)
		5'UTR	21 (11.23%)
		1stExon	2 (1.07%)
Body	93 (49.73%)		
3'UTR	7 (3.74%)		
Intergenic	42 (22.46%)		

c

FGD	CpGs	Subgroup	CpGs
Promoter	3 (23.08%)	TSS1500	3 (23.08%)
		TSS200	0 (0.00%)
		5'UTR	0 (0.00%)
		1stExon	0 (0.00%)
Body	4 (30.77%)		
3'UTR	0 (0.00%)		
Intergenic	6 (46.15%)		

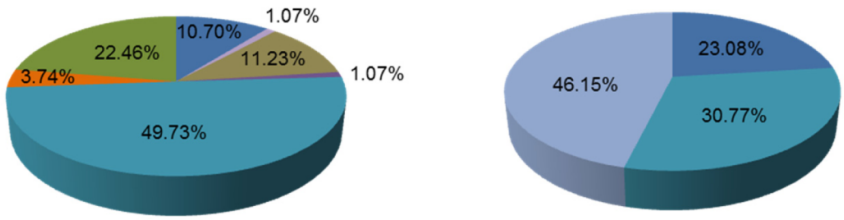


Figure 27. Significant CpGs and their relation to RefGeneGroups. Displayed is the distribution of CpG sites classified in four different major groups (promoter, body, 3'UTR, intergenic) and four promoter subgroups (TSS200, TSS1500, 5'UTR, 1st Exon) according to their position within RefGenes for **a)** all analyzed CpG sites of the 450K (N= 468316), **b)** the current smoking-associated differentially methylated CpG sites of KORA F4 that have been replicated in KORA F3 (N=187), and **c)** the former smoking-associated differentially methylated CpG sites of KORA F4 (N=13).

The genomic position of the hypo- and hypermethylated CpGs can influence the expression of genes. Aberrant hypermethylation of promoter regions, which are usually unmethylated to allow transcription, can lead to transcriptional inactivation whereas aberrant hypomethylation of gene bodies and intergenic regions tends to allow unwanted transcription. By comparing the 187 significant and replicated CpGs of current vs. never smokers in KORA F4 with regard to the distribution of hypo- and hypermethylated CpGs within the RefGene Groups, the outcome was similar for the gene body (47.79% vs. 54.90%) and for the 5'UTR (11.03% to 11.76%). Still, the intergenic region was more predominant in hypomethylated CpG sites (26.47% vs. 11.76%). On the other hand, more hypermethylated CpG sites were detected within the TSS200 (3.92% to 0.00%), 1st Exon (1.96% vs. 0.74%) and 3' UTR (7.84% vs. 2.21%) (Figure 28 a & b).

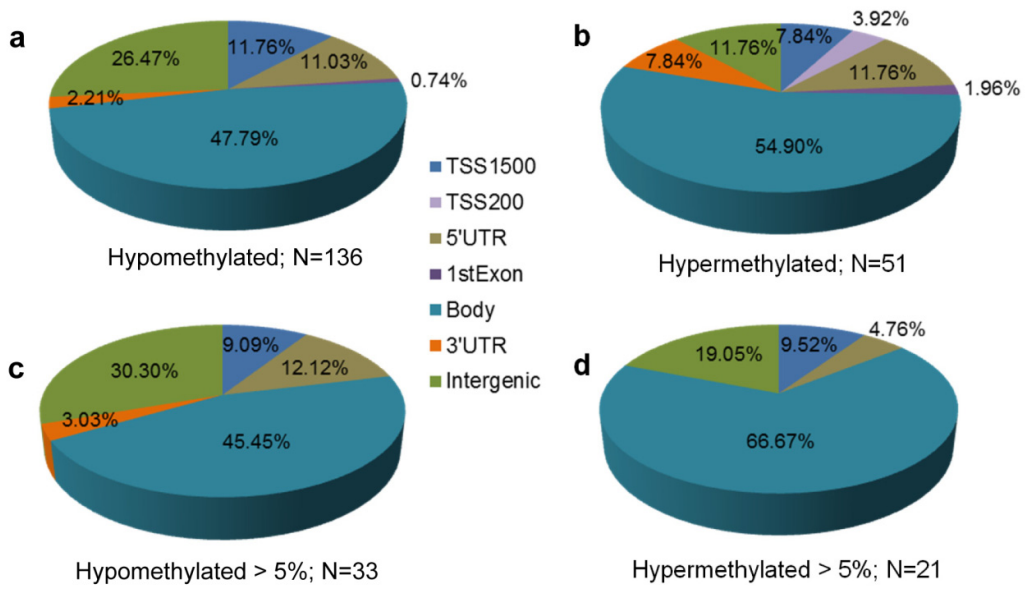


Figure 28. Significant hypo- and hypermethylated CpGs and their relation to RefGeneGroups. Displayed are the significant **a)** hypomethylated and **b)** hypermethylated CpG sites of current smokers in KORA F4 that have been replicated in KORA F3, and their relation to RefGeneGroups. And the corresponding results including only **c)** hypomethylated and **d)** hypermethylated CpG sites with a smoking-associated methylation difference of over 5% in KORA F4.

Within the sites that showed a smoking-associated methylation difference of over 5% in KORA F4, the intergenic region was still more predominant in hypomethylated CpGs (30.30% vs. 19.05%), as well as the 3' UTR (3.03% vs. 0.00%), and the 5' UTR (12.12% vs. 4.76%). On the other hand, more hypermethylated sites were found to be located in gene bodies (66.67% vs. 45.45%) (Figure 28 c & d).

3.1.5.3. Relation to CpG islands

In general, CpGs are underrepresented in the genome but certain regions, the CpG islands, have a high concentration of CpG sites. As recent work showed that not necessarily CpG islands but their flanking regions may play a significant role in gene expression [205], the location of the significant smoking-associated CpGs in relation to CGIs was of interest. From the CpG context, the 450K BeadChip covers 96% of CGIs with multiple sites in the annotated CGI, shores (2 kb flanking the CGI) and shelves (2 kb flanking the shores) as well as island-independent CpGs, which are not necessarily gene-based, depending on their distance to the nearest genes.

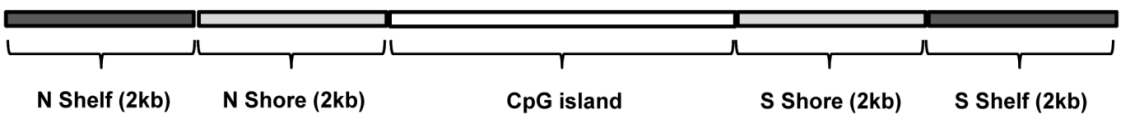
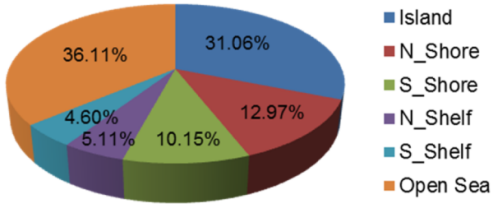


Figure 29 displays the distribution of all 468316 CpGs analyzed, of the significant and replicated 187 CpGs associated with current smoking, and of the significant 13 CpGs associated with former smoking with regard to their relation to CGI.

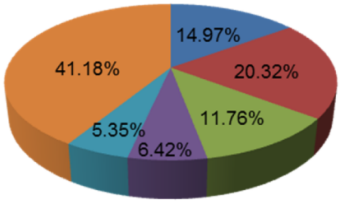
a

CpG content	CpGs	Subgroup	CpGs
Island	145467 (31.06%)		
Shore	108268 (23.12%)	N-Shore	60720 (12.97%)
		S-Shore	47548 (10.15%)
Shelf	45476 (9.71%)	N-Shelf	23938 (5.11%)
		S-Shelf	21538 (4.60%)
Open Sea	169105 (36.11%)		



b

CpG content	CpGs	Subgroup	CpGs
Island	28 (14.97%)		
Shore	60 (32.08%)	N-Shore	38 (20.32%)
		S-Shore	22 (11.76%)
Shelf	22 (11.77%)	N-Shelf	12 (6.42%)
		S-Shelf	10 (5.35%)
Open Sea	77 (41.18%)		



c

CpG content	CpGs	Subgroup	CpGs
Island	5 (38.46%)		
Shore	6 (46.15%)	N-Shore	5 (38.46%)
		S-Shore	1 (7.69%)
Shelf	0 (0.00%)	N-Shelf	0 (0.00%)
		S-Shelf	0 (0.00%)
Open Sea	2 (15.38%)		

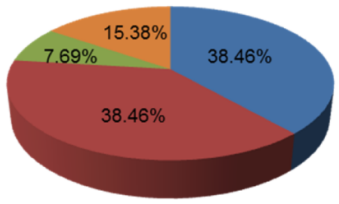


Figure 29. Significant CpGs and their relation to CpG islands. Displayed is the distribution of CpGs according to their neighborhood context classified as Island, Shore, Shelf and Open Sea, with additionally sub-dividing Shore and Shelf (N-Shore, S-Shore, N-Shelf, S-Shelf), for **a)** all analyzed CpGs of the 450K (N= 468316), **b)** the current smoking-associated differentially methylated CpGs of KORA F4 that have been replicated in F3 (N=187), and **c)** the former smoking-associated differentially methylated CpGs of KORA F4 (N=13).

Comparing the relation to CGI of smoking-associated sites (N=187) with the overall analyzed sites of the 450K BeadChip (N= 468316), twice as many of the smoking-associated sites would have been expected to be located in a CGI (14.97% vs. 31.06%), whereas an overrepresentation of sites located in the N-Shore (20.32% vs. 12.97%) and also a slight one of sites located in the Open Sea (41.18% vs. 36.11%) could be detected (Figure 29 a & b).
Again, former smokers differed from current smokers as differentially methylated sites detected in former vs. never smokers were not as prevalently located in the Open Sea as sites detected in current vs. never smokers (with only 15.38% vs. 41.18%). Conversely, sites located in the N-Shores of CGIs and in Islands were much more predominant in former

compared to current smokers (Island: 38.46% vs. 14.97%; N-Shores: 38.46% vs. 20.32%) (Figure 29 b & c). By comparing the 187 significant and replicated CpG sites of current vs. never smokers in KORA F4 with regard to the distribution of hypo- and hypermethylated CpG sites within the neighborhood of CGIs, the results were quite similar, with most of them being located in the Open Sea (both 41.18%) and 15.44% (hypo), respectively 13.73% (hyper), being located in the Island. Nevertheless, the distribution in the shores and shelves of CGIs varied with N-Shores and S-Shelves being more predominant in hypomethylated sites and S-Shores and N-Shelves being more predominant in hypermethylated sites (Figure 30 a & b).

By including only the significant CpG sites of current vs. never smokers with a smoking-associated methylation difference of more than 5% in KORA F4, the outcome was different. Even though most of the smoking-associated CpG sites were again located in the Open Sea in both groups, the percentage of hypermethylated sites included (57.14%) was even higher than the one of hypomethylated sites (42.42%). Sites located in S-Shores were also found to be more predominant in the hypermethylated group (14.29% vs. 9.09%). On the other hand, sites located in the Island itself were more predominant in the hypomethylated group (21.21% vs. 9.52%), as well as S-Shelves that were not present in the hypermethylated group at all (6.06% vs. 0.00%) (Figure 30 c & d).

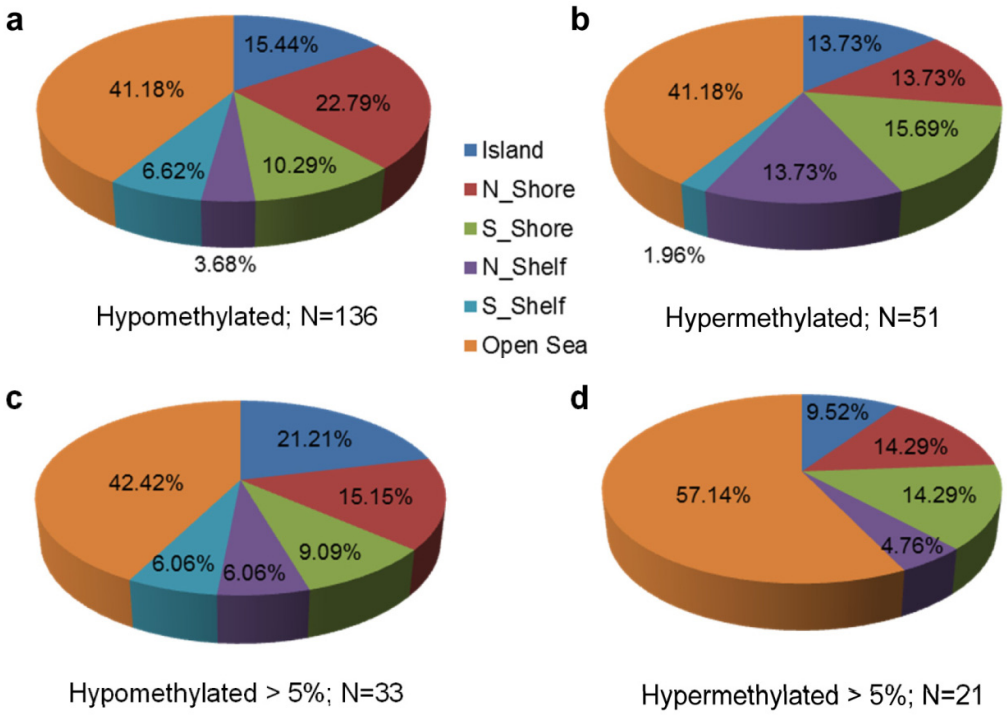


Figure 30. Significant hypo- and hypermethylated CpG sites and their relation to CGIs. Displayed are the significant **a)** hypomethylated and **b)** hypermethylated CpG sites of current smokers in KORA F4 that have been replicated in KORA F3, and their relation to CpG islands. And the corresponding results including only **c)** hypomethylated, and **d)** hypermethylated CpG sites with a smoking-associated methylation difference of over 5% in KORA F4.

3.1.6. Methylation-specific DNA-protein binding analysis for *AHRR*

To assess the potential biological relevance of DNA methylation differences caused by tobacco smoking, methylation-specific DNA-protein binding analysis by electrophoretic mobility shift assay (EMSA) were carried out exemplarily for CpG cg05575921 (*AHRR*). This site was chosen because it was the most outstanding site with respect to significance and level of detectable changes in DNA methylation associated with tobacco smoking. Figure 31 depicts the fluorescence scan of two independent EMSA experiments and shows that methylation-specific DNA-protein binding patterns could be detected for this site in the nuclear extract of two cell lines, Raji (human B-lymphoblastoid cell line; Figure 31 a) and THP1 (human monocytic cell line; Figure 31 b). DNA-protein complex C1 showed higher binding affinity to the methylated site, whereas complexes C2 and C3 preferably bound to the unmethylated state of cg05575921. Binding specificity was validated by using a competitive approach.

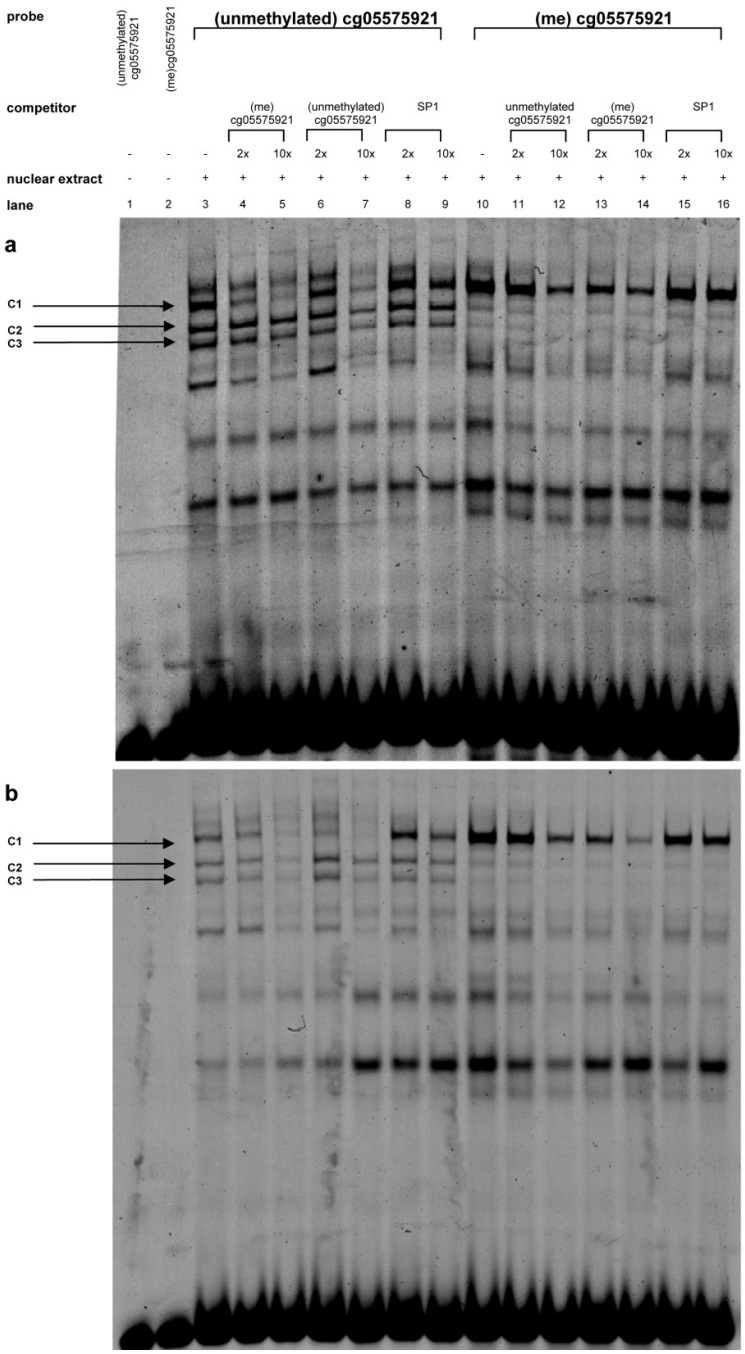


Figure 31. Methylation-specific protein binding patterns of cg05575921 (*AHRR*). Methylated and unmethylated Cy5-labelled probes carrying the cg05575921 site were used in competition EMSAs using **a)** Raji and **b)** THP1 nuclear extracts. Arrows indicate shifted protein-DNA complexes showing methylation-specific binding patterns (C1-C3). cg: unmethylated cg05575921, (me)cg: methylated cg05575921, SP1 = Specificity protein 1 (used as unspecific competitor). In lane 1 and 2, free oligonucleotides without incubation with nuclear extracts are shown. Lane 3 and 10 show the results for EMSAs for the unmethylated and methylated variant without competition. In lane 4, 5, 11, and 12 competitions with the unlabeled adverse oligonucleotides were performed, whereas competitions with the same unlabeled oligonucleotides were performed in lane 6, 7, 13, and 14. To ensure specificity, competitions with unlabeled SP1-consensus oligonucleotides were performed in lane 8, 9, 15, and 16. (me)cg: methylated cg05575921; SP1 = Specificity protein 1.

3.2. Validation of the 450K DNA methylation analysis

As the 450K BeadArray was a new approach, not knowing how reliable gained data would really be, validation of the results with another method was needed. Also, the target areas identified by whole genome scans only provide methylation data for a few individual CpGs within a region. Thus, the EpiTYPER method of Sequenom was established in order to perform a high-resolution scan of selected regions.

3.2.1. Establishment of the Sequenom EpiTYPER array

This approach was carried out on 21 randomly selected current and never smokers of the KORA F4 panel. The five most significant current smoking-associated loci (cg05575921-*AHRR*, cg21566642-*ALPP/ALPP2*, cg03636183-*F2RL3*, cg06126421-not annotated, cg19572487-*RARA*) were selected for validation.

3.2.1.1. Quality of assay performance

The EpiTYPER Analyzer program allows comparison of the mass spectra, the nucleotide sequences that produced the spectra, and the amplicon-specific methylation data. Figure 32 exemplarily displays the methylation data of 42 selected KORA F4 DNA samples for the amplicon of the *AHRR* gene, including the 450K site cg05575921. Sample 15 did not show any or in some cases not trustworthy results for any of the amplicons and was therefore discarded in the subsequent statistical analysis.

For quality control issues, DNA methylation standards (0, 25, 50, 75, and 100%) and a negative control (NTC = H_2O) were additionally analyzed for each amplicon to control for proper dynamic ranges of measurements, potential PCR amplification biases, and contaminations (Figure 32). All negative controls did not show any result as expected, therefore ruling out possible contamination during the EpiTYPER process. The results of the DNA methylation standards showed that within some of the CpG sites the methylation level observed did not accurately match the expected methylation level given by the DNA methylation standards, even though the trend was given for each CpG site.

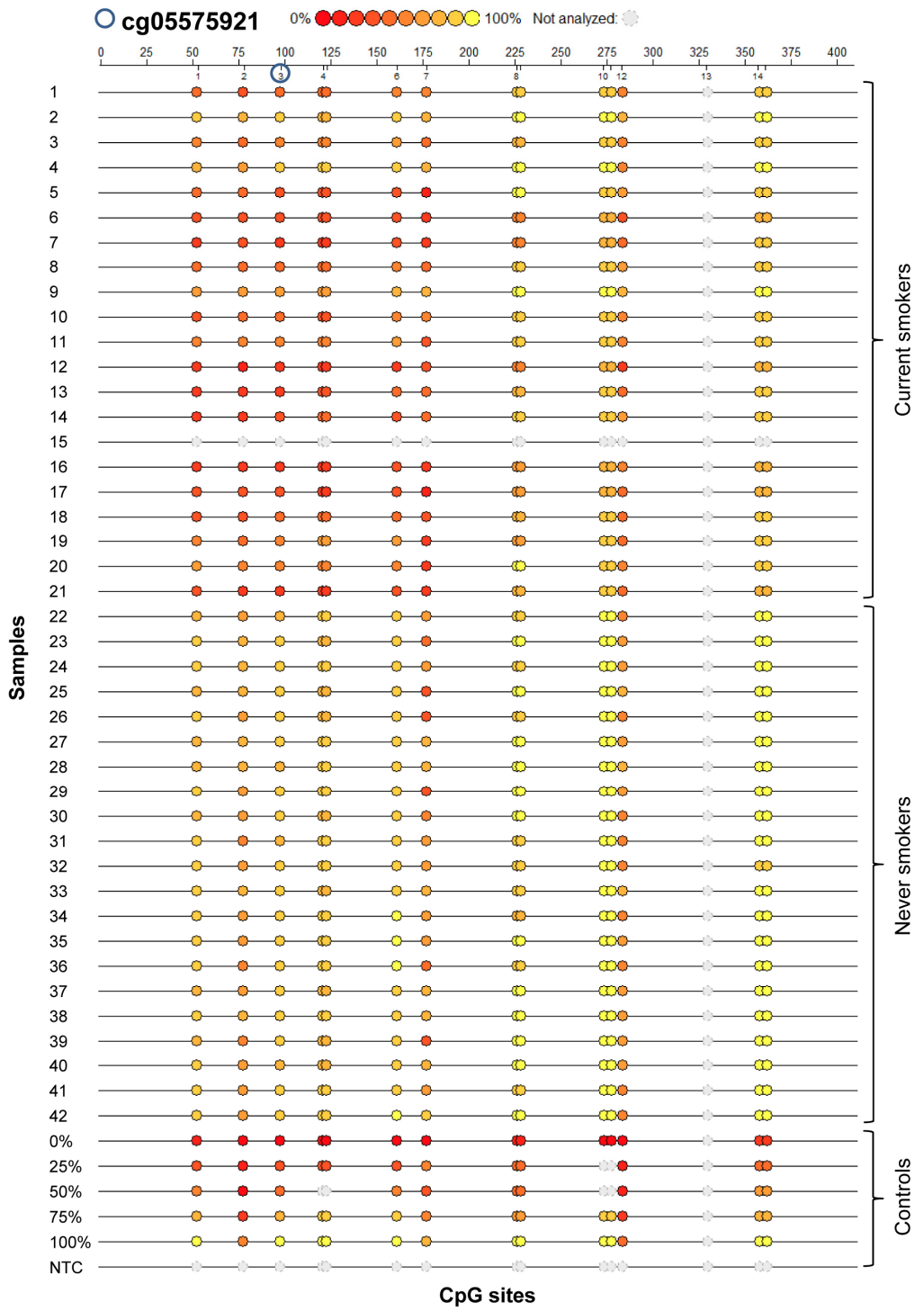


Figure 32. EpiTYPER methylation data for the AHRR amplicon (cg05575921). The analyzed region of the *AHRR* gene is displayed as it can be viewed in the EpiGram tab of the EpiTYPER Analyzer program. The EpiGram tab is a graphical representation of methylation ratio, each circle displays a CpG site, the colour code within the circles denotes the level of methylation found at this particular site in the selected sample. The default colour spectrum ranges from red (0% methylated) to yellow (100% methylated). Above the set of samples the nucleotide position and the CpG site number are given. NTC = negative control.

As the subsequent statistical analyses were planned to focus on a case-control study design, reporting only the difference in methylation levels between the two groups and not taking the absolute number into account, this was not a big issue. Moreover, for all particular CpGs that covered the corresponding 450K site, the graphs originating from samples with 0 to 100% methylation showed proper dynamic ranges of measurements with clearly distinguishable peaks, as exemplarily displayed for CpG3 of *AHRR* (cg05575921) in Figure 33.

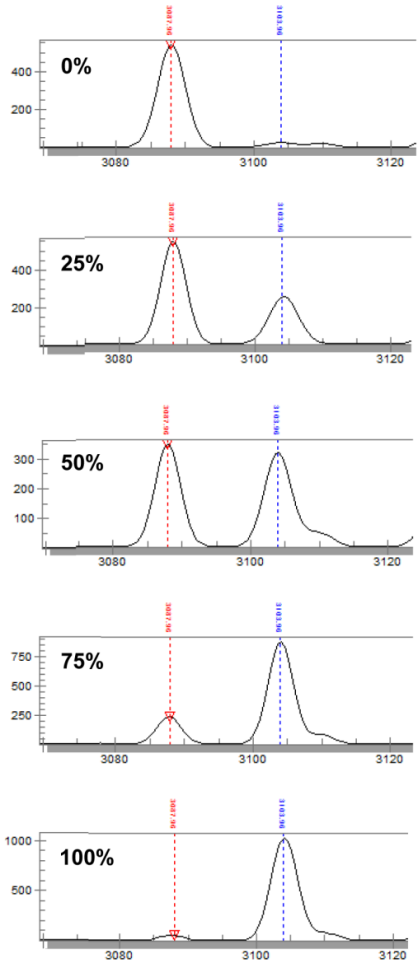


Figure 33. Results of DNA methylation standards for CpG3 of the AHRR amplicon. The detected mass signals of the DNA methylation standards are displayed for CpG3 of the AHRR amplicon, which corresponds to cg05575921. X-axis display the fragments of different mass for originally methylated and non-methylated CpG sites, y-axis display the detection intensity of amplicon fragments. The shift in mass, analyzed by the MALDI-TOF mass spectrometer, for non-methylated to methylated fragments for a single CpG site is 16 Da, due to the presence of an “A” residue in the place of a “G” in the sequence. If samples contain both methylated and non-methylated target sequences, two mass signals will be observed. The peak area ratio of these two mass signals representing methylated and non-methylated DNA are used to estimate the degree of methylation of the targeted CpG site.

3.2.1.2. Bisulfite conversion

As earlier described for the 450K BeadChip in chapter 3.1.1.1., the DNA methylation information is erased by amplification, which is accomplished by PCR in the case of the EpiTYPER technology. So bisulfite treatment was used as well for the region-specific approach of Sequenom. As could be shown within that same chapter, a test PCR covering only a small region of the genome is not really trustworthy, as other parts of the genome than the one covered by the test primers might lack complete conversion. Thus, another way to check for complete bisulfite conversion was needed and so the pipeline of Thompson et al., which offers a bisulfite conversion efficiency calculation by measuring levels of unconverted non-CG cytosines in a given sample, was used [202]. The bisulfite conversion efficiency was calculated for all amplicons and samples in the validation process. The results for three amplicons are displayed in the form of color-filled bars in Figure 34 and are representative for all of the five amplicons, as mentioned in the figure text. Within two amplicons (corresponding to CpG sites cg03636183 and cg19572487) no non-CG cytosines were present, consequently they could not be measured for complete bisulfite conversion and all bars are colored in blue (Figure 34 c). Therefore, complete bisulfite conversion could be detected for all samples, excluding Sample 15 and the negative control, and for three out of the five amplicons (Figure 34 a & b). To allow assignment, for this figure the amplicons are named according to their corresponding 450K CpG site even though a number of additional CpG sites within the amplicon are measured by the EpiTYPER, as not all of the five amplicons correspond to an annotated gene.

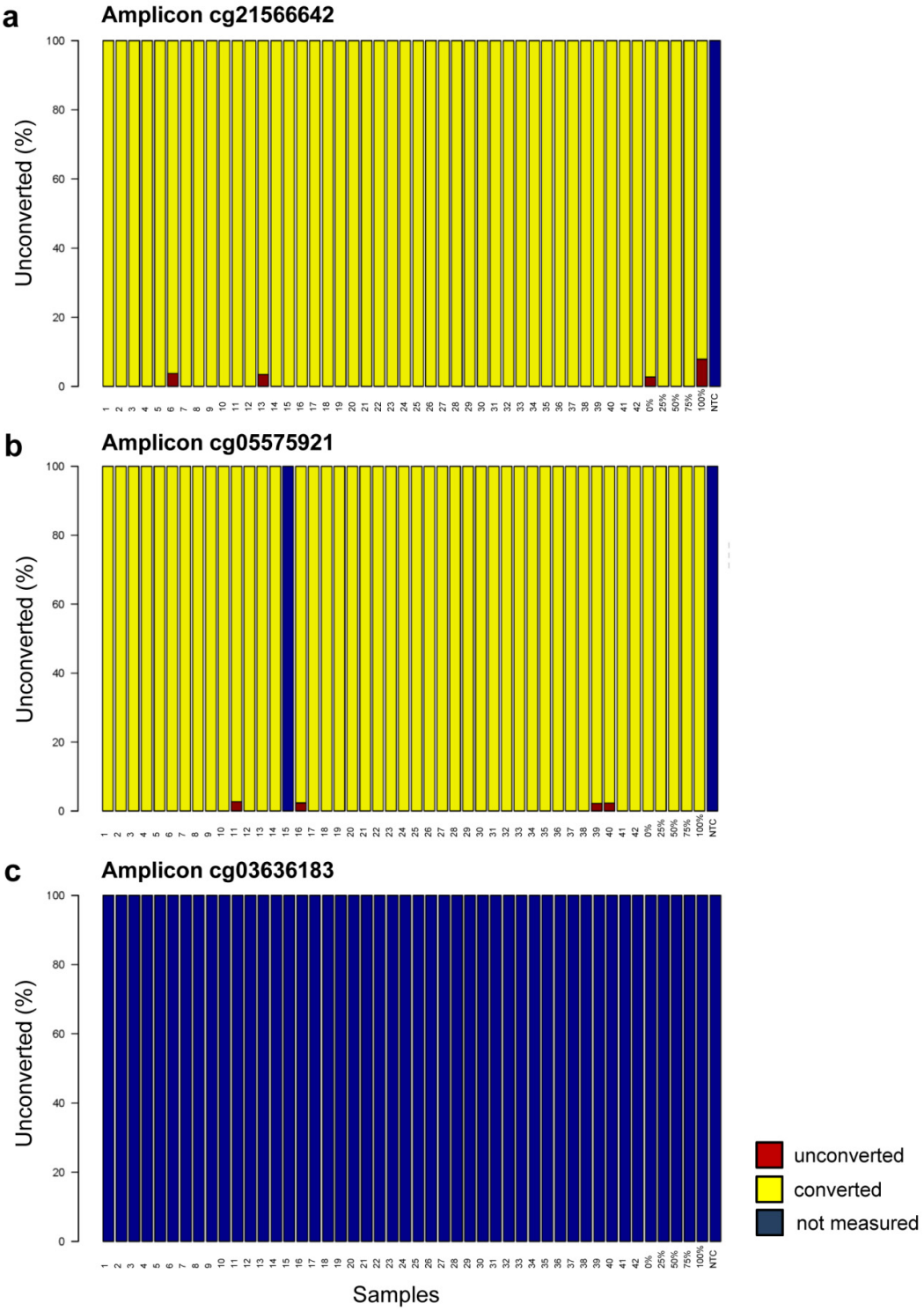


Figure 34. Measurement of bisulfite conversion controls. The results of the three displayed amplicons are representative for all of the five amplicons, as two render the results of another one, as followed: cg21566642 for cg06126421; cg03636183 for cg19572487; only cg05575921 for itself. Bar height in red color depicts the percentage of unconverted cytosines detected, with 0% indicating complete conversion of measured cytosines and 100% indicating a complete failure of conversion. Bars in blue indicate that no non-CG cytosines were present in the amplicon and bisulfite conversion could not be measured.

3.2.2. Region-specific validation of the 450K results

To validate the smoking-associated differential DNA methylation detected with the 450K BeadChip, the Sequenom's EpiTYPER approach was carried out on 21 randomly selected current and never smokers of the KORA F4 panel. However, one of the case samples (Sample 15) was discarded after bisulfite conversion control (see subchapter 3.2.1.2.) and so one of the control samples (Sample 35) that matched Sample 15 was discarded as well to allow an equal sample size. The characteristics of the final study population used for statistical analysis are summarized in Table 16.

Table 16. Characteristics of the study population for the EpiTYPER assay.

	Never smokers	Current smokers
N	20 (50%)	20 (50%)
Age (years)	63.00 (44-77)	55.70 (48-64)
Gender		
Female	13 (65%)	9 (45%)
Male	7	11
Smoking		
Pack-years	0	38.90 (10.28-74)
BMI (kg/m^2)	27.56 (20.36-32.97)	27.35 (20.04-38.20)
WBC count (WBC/nl)	5.17 (3.30-7.30)	7.62 (4.80-11.00)
Alcohol (g/kg/day)	0.14 (0.00-0.74)	0.31 (0.00-1.55)

BMI: Body Mass Index, WBC: White Blood Cell count; each individual characteristic is given in mean with range in parentheses

The association with smoking status for the five loci was assessed by linear regression using M-values as response, smoking status as explanatory variable and sex, age, BMI, alcohol consumption as well as WBC count as covariates. The results, covering several CpGs within these regions, are displayed in Table 17. Three CpGs were validated directly and two CpGs were not directly covered by the EpiTYPER assay, owing to low mass (cg03636183, < 1.500 Da) and high mass (cg21566642, > 7.000 Da) of the cleavage product, thus lying outside the analytical window of the mass spectrometry. Within the five regions assayed in the EpiTYPER analysis only one additional CpG, CpG7 of the *AHRR* locus, corresponded to another 450K CpG, cg23576855. This CpG was actually also significantly associated with current smoking in the 450K analysis, but had to be excluded as it did not show normally distributed residuals (please see method section for more information).

Table 17. Results of the validation by EpiTYPER assay.

Gene CpG	450K: Median ß-value methylation difference	Median ß-value methylation difference in %	p-value	Pearson r^2	ß-value as median (first quartile - third quartile)	
					Never smokers	Current smokers
AHRR						
CpG1*		-42.50	8.52E-07		0.820 (0.798 - 0.823)	0.395 (0.295 - 0.565)
CpG2*		-24.50	3.53E-04		0.640 (0.610 - 0.710)	0.395 (0.308 - 0.488)
CpG3*		-38.00	3.46E-07	0.976456	0.820 (0.800 - 0.833)	0.440 (0.305 - 0.560)
CpG4.5*	CpG3	-40.00	6.11E-06		0.835 (0.783 - 0.850)	0.435 (0.305 - 0.555)
CpG6*	cg05575921:	-40.00	2.41E-08		0.885 (0.878 - 0.893)	0.485 (0.388 - 0.613)
CpG7	-24.40%	-26.50	4.11E-02		0.650 (0.420 - 0.713)	0.385 (0.260 - 0.485)
CpG8.9		-12.50	1.63E-02		0.935 (0.855 - 0.960)	0.810 (0.675 - 0.900)
CpG10.11*		-8.00	2.05E-04		0.930 (0.920 - 0.940)	0.850 (0.808 - 0.870)
CpG12		-5.50	1.24E-01		0.615 (0.565 - 0.670)	0.560 (0.443 - 0.608)
CpG14.15		-7.00	6.71E-03		0.930 (0.920 - 0.940)	0.860 (0.815 - 0.890)
ALPPL2[b]						
CpG1.2*		-23.00	8.06E-08		0.395 (0.340 - 0.460)	0.165 (0.130 - 0.233)
CpG3*		-19.00	2.40E-08		0.310 (0.265 - 0.365)	0.120 (0.100 - 0.150)
CpG6	CpG12	-9.50	1.45E-01		0.180 (0.145 - 0.220)	0.085 (0.078 - 0.113)
CpG12	cg21566642:	not covered				
CpG13*	-16.70%	-11.50	2.40E-06	0.797046	0.195 (0.150 - 0.223)	0.080 (0.070 - 0.133)
CpG14*		-5.00	8.20E-07		0.130 (0.120 - 0.150)	0.080 (0.070 - 0.100)
CpG15.16.17.18*		-15.50	1.66E-05		0.355 (0.295 - 0.403)	0.200 (0.160 - 0.215)
F2RL3						
CpG1*		-26.00	1.43E-03		0.750 (0.708 - 0.810)	0.490 (0.445 - 0.680)
CpG2		-25.50	4.93E-03		0.660 (0.620 - 0.720)	0.405 (0.360 - 0.488)
CpG3*		-18.00	2.43E-04		0.870 (0.830 - 0.890)	0.690 (0.638 - 0.805)
CpG6		not covered				
CpG7*		-20.50	7.19E-06	0.8893826	0.850 (0.828 - 0.870)	0.645 (0.560 - 0.755)
CpG8.9	CpG6	-12.00	2.43E-01		0.765 (0.500 - 0.823)	0.645 (0.478 - 0.753)
CpG13	cg03636183:	-1.50	1.05E-01		0.905 (0.890 - 0.933)	0.890 (0.870 - 0.905)
CpG14	-14.74%	-2.50	3.58E-01		0.915 (0.890 - 0.923)	0.890 (0.880 - 0.903)
CpG15.16		0.00	8.28E-01		0.980 (0.980 - 0.983)	0.980 (0.970 - 0.983)
CpG17.18		-1.00	7.79E-03		0.980 (0.970 - 0.990)	0.970 (0.970 - 0.980)
CpG19.20		0.00	2.57E-01		0.860 (0.850 - 0.880)	0.860 (0.850 - 0.870)
CpG21.22		-0.50	6.88E-01		0.930 (0.930 - 0.940)	0.925 (0.920 - 0.930)
CpG23		1.00	6.89E-01		0.870 (0.855 - 0.890)	0.880 (0.860 - 0.890)
x[a]						
CpG1.2		-5.00	5.03E-02		0.940 (0.930 - 0.940)	0.890 (0.870 - 0.923)
CpG3	CpG4	-4.50	9.31E-01		0.420 (0.388 - 0.433)	0.375 (0.340 - 0.463)
CpG4*	cg06126421:	-16.00	5.19E-04	0.9657657	0.680 (0.638 - 0.730)	0.520 (0.400 - 0.600)
CpG5	-17.05%	-11.50	1.04E-01		0.730 (0.678 - 0.753)	0.615 (0.548 - 0.693)
RARA						
CpG1*	CpG2	-7.50	1.57E-03		0.725 (0.690 - 0.740)	0.650 (0.610 - 0.695)
CpG2*	cg19572487:	-13.00	1.48E-05	0.886522	0.470 (0.438 - 0.500)	0.340 (0.310 - 0.405)
CpG3.4	- 10.02%	-11.00	6.02E-02		0.370 (0.305 - 0.463)	0.260 (0.205 - 0.380)

Displayed are the results of current vs. never smokers of the linear model adjusted for age, sex, BMI, alcohol, and white blood cell count (p-value) and the corresponding results of the 450K analysis for KORA F4. For the EpiTYPER validation, the correlation coefficient (Pearson r^2) between the Illumina array derived b-values and the EpiTYPER methylation is reported for the associated CpG site (except for cg21566642 and cg03636183, where a related CpG site was validated as indicated in the text and table). * Significant after Bonferroni $p \leq 0.05/28 = 0.0018$. [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island.

To illustrate technical validation, the methylation intensities for each individual from the 450K BeadChip and from the EpiTYPER were plotted against each other (Figure 35); the corresponding Pearson correlation coefficient is also given in Table 17.

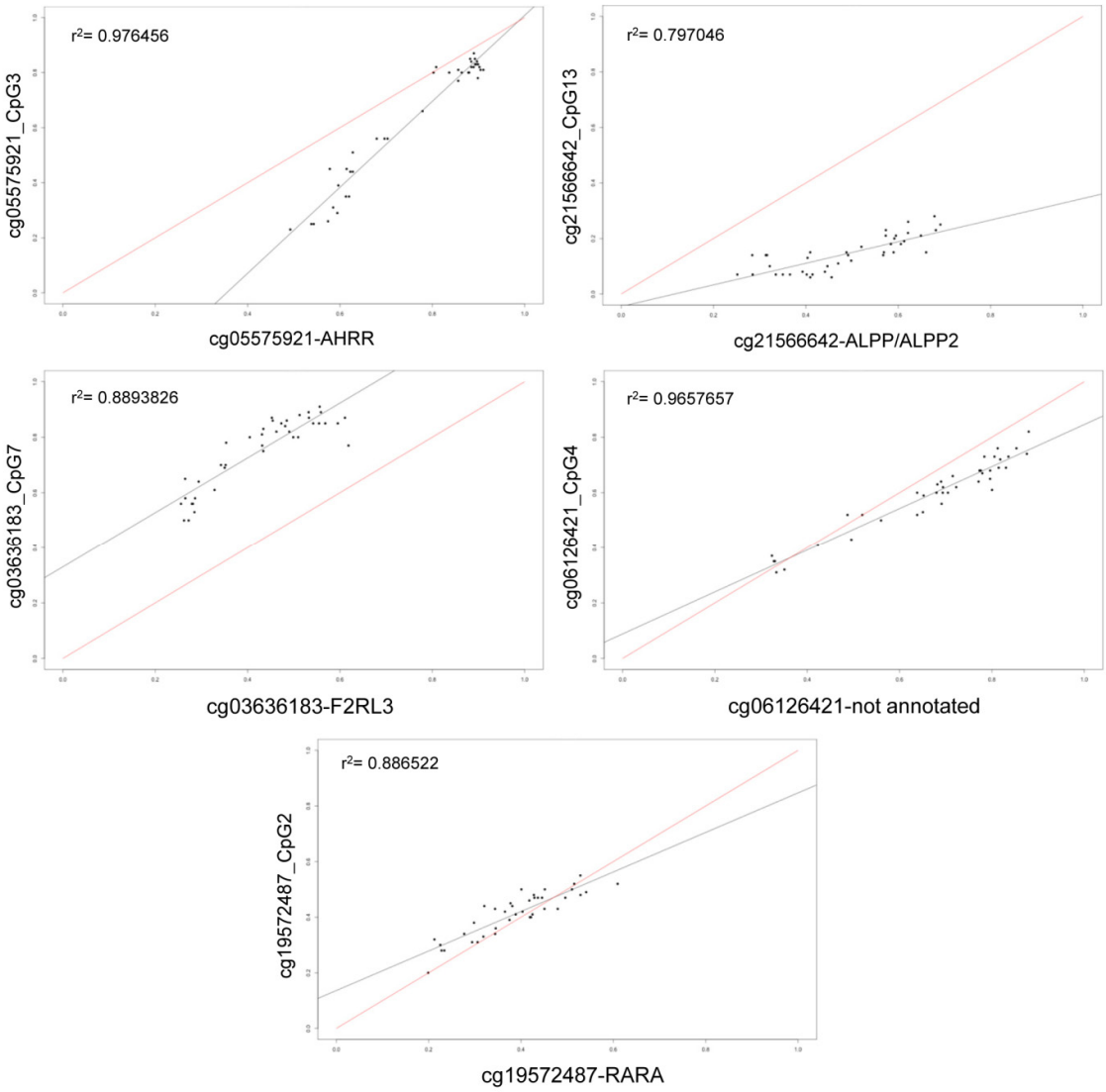


Figure 35. Technical validation of the 450K BeadChip methylation intensities. The methylation intensities of the 450K assay and the EpiTYPER method are compared and plotted against each other. Methylation intensities for each individual are plotted on the x-axis for the 450K assay and on the y-axis for the EpiTYPER method, with the blue line displaying the correlation of methylation data. The red line displays the angle bisector. Data points above the angle bisector indicate higher, items beneath this line lower EpiTYPER than Illumina 450K methylation levels. Pearson correlation is given in r^2.

As stated above, two sites were not directly covered by the EpiTYPER assay, so alternatively CpG13 was used for cg21566642 (instead of CpG12) and CpG7 was used for cg03636183 (instead of CpG6), to allow at least a rough estimate of correlation. Nonetheless, the correlation was very high, even though the correlation for cg03636183 showed a parallel shift to the ideal angle bisector and the one of cg21566642 was below 0.8. For cg05575921, EpiTYPER methylation intensities were lower compared to the chip. Altogether, the correlation with the proper CpG sites from the chip were all above 0.88, showing a good

reproducibility of the intensities, and the DNA methylation differences in current compared to never smokers could be replicated.

3.2.2.1. Determination of possible genetic effects on DNA methylation analysis

DNA methylation analysis within the EpiTYPER approach can be effected by nearby SNPs, which may lead to misinterpretation of the results. However, these SNPs cause an altered peak pattern and can also be used for a high-throughput analysis of the allelic distribution of SNPs in the sample cohort. Expected and observed data was therefore compared and an exhaustive string substitution approach used [202] to identify putative SNPs. The results from SNP prediction analysis are displayed in Figure 36.

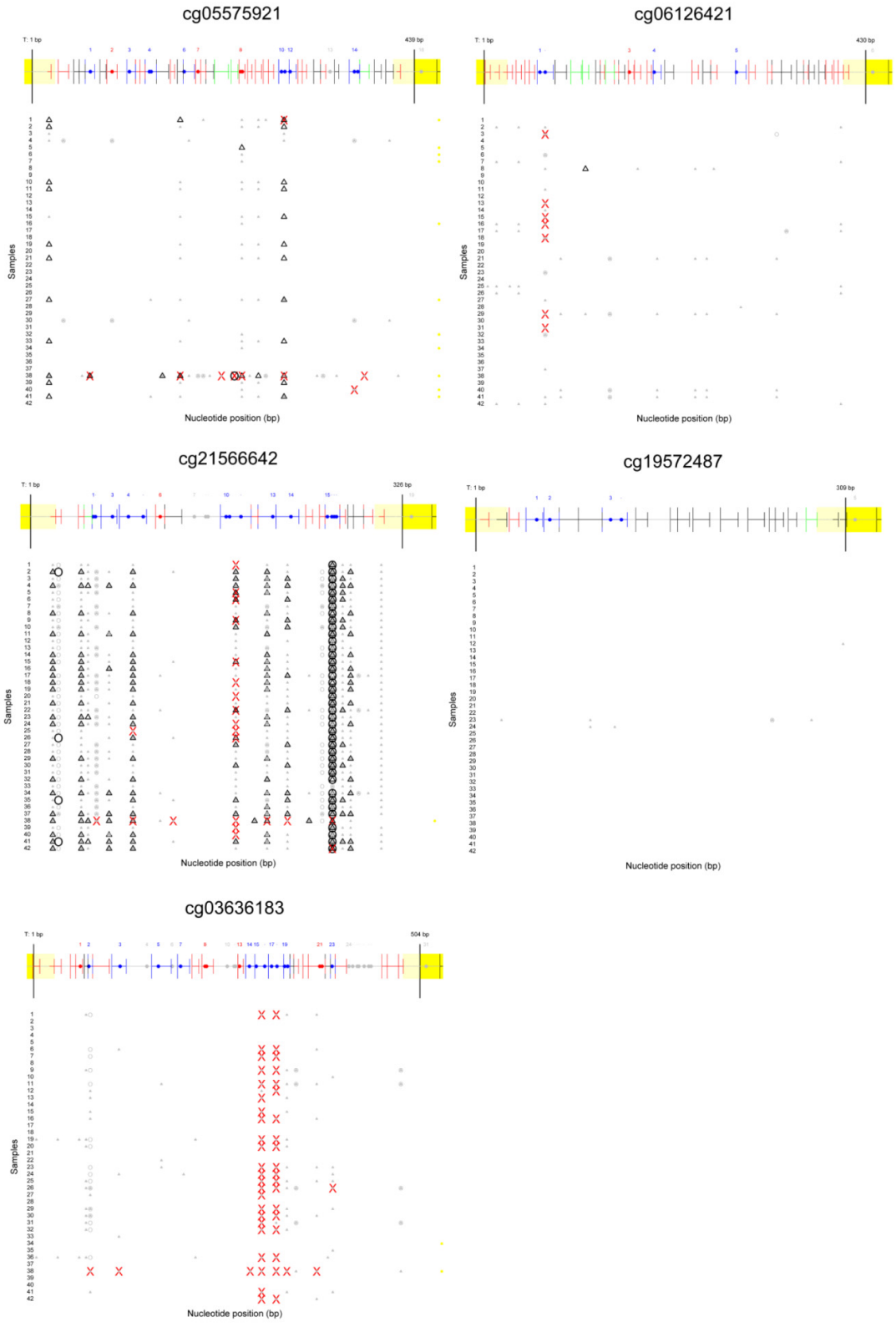


Figure 36. Graphical output from SNP detection analysis for a single amplicon. The caption for this figure is located on the next page.

Figure 36. Graphical output from SNP detection analysis for a single amplicon. The expected T-cleavage fragmentation pattern is shown in the top panel. The beginning of a fragment is illustrated by long vertical bars, the end by short vertical bars. CpG dinucleotides are numbered and colored according to their ability to be assayed: uniquely assayable site (blue), fragment molecular weight overlaps with another fragment (red), fragment contains potential conversion controls, i.e., non-CpG cytosines (green), fragment is uniquely assayable but contains no CpGs (black), or fragment molecular weight lies outside the testable mass window (gray). If multiple CpGs lie within the same fragment, only the first one is numbered, and all others are marked with a "-". In the lower panel, each row corresponds to a sample with its identified potential SNPs depicted as circles, triangles, or X's. Triangles indicate base pair substitutions, while circles correspond to base pair deletions. An 'X' marks SNPs that explain expected peaks that are entirely missing in the data. SNPs attributed a higher confidence are depicted in large black symbols, whereas low-confidence SNPs are depicted in small grey symbols.

4. Discussion

4.1. Study design

4.1.1 General remarks

Environmental factors have been shown to modulate the establishment and maintenance of epigenetic modifications, such as DNA methylation, thereby inducing changes in gene expression and having a profound role on development and health. At the beginning of this doctoral thesis, studies which have investigated the effects of tobacco smoking on aberrant DNA methylation have been largely restricted to promoter regions of malignancy-related genes and have enrolled small numbers of subjects. Nevertheless, these studies found great evidence that tobacco smoking leads to aberrant DNA methylation which might have a tremendous effect on the body. Only recently, improvements in array-based technologies have enabled the assessment of the molecular effects of tobacco smoking in larger populations. With the introduction of the Illumina 450K BeadArray, analyses of the DNA methylation pattern across the genome with a high coverage of promoters, 5'UTRs, 1st Exons, gene bodies and 3'UTRs were made possible.

Accordingly, an increasing number of studies were published on tobacco smoking and genome-wide DNA methylation very recently, which will be referred to in the following chapters of the discussion part. These studies underline the great interest in this subject in general and support the theory and results of this doctoral thesis.

The aim of this study was to assess the impact of current and former tobacco smoking on DNA methylation in an epigenome-wide approach using the 450K BeadArray and whole blood samples from healthy individuals. It was built on the hypothesis that smoking induces gene-specific methylation and that these exposure induced changes may help to unravel the pathogenesis of complex diseases.

4.1.2. Challenges in data pre-processing and normalization

Over the past decades various methods have been developed to carry out DNA methylation analysis, but improved technologies for epigenome-wide association studies (EWAs) using microarray or second-generation sequencing-based approaches have been devised only

recently. Consequently, when the raw data of the 450K BeadChip was processed, pre-processing and statistical approaches were at an early stage, which posed great challenges. Questions concerning normalization strategies and statistical models that can be used for the analyses are discussed quite contrary in the community and method papers that give a clue of how to precede with the 450K data were published only very recently. The first statistical analyses were carried out without any normalization or pre-processing steps. Then the same analyses were carried out with GenomeStudio normalized beta-values, even though it was not known how this was exactly accomplished by the software and therefore not accepted within the community. Also, as described in the methods section, the use of Beta-values for the statistical analyses was not state of the art as M-values are more statistical reliable. Publishing the results was therefore not an option. Finally, various packages (IMA [206], MINFI [207], lumi [176]) for methylation files outputted by Illumina GenomeStudio software were tested, but these packages all did not provide a satisfying solution. Fortunately a complete pre-processing pipeline for 450K BeadChip data was published by Touleimat & Tost in June 2012 [175] that provided solutions for the main critical points that came up with the design of the 450K BeadChip and was used for the 450K data of KORA F4 and F3. First, the Infinium 450K BeadChip consists of two different chemical assays, Infinium I and II. It has been demonstrated that these two show different dynamic behavior in the assay that leads to a shift in the density curves between them [208]. The two datasets therefore had to be viewed as two distinct datasets and some processing of the data was required to make them comparable. The *Subset based Quantile Normalization* step corrected the InfI/InfII shift and also eliminated non-biological variation between samples via 'between sample normalization'. Second, within the measurement of KORA F4 and F3 an imbalance in the signal intensities of the green and red channel was observed. It has been shown that the labeling efficiency and scanning properties of the two color channels can differ from each other and that the intensities measured in the two color channels might be imbalanced [175]. A color balance adjustment therefore had to be accomplished, which was part of the *Signal Correction* step. Third, some of the probes present on the 450K BeadChip potentially contain or extend on SNPs. But microarray-based methylation chips are susceptible to interference with SNPs, as genetic variance near a CpG site may interfere with the oligo-based determination of CpG methylation and therefore affecting probe binding or read-out [209]. The *Probe Filtering* step accounted for this problem and removed the corresponding CpG sites.

In general, analyzing these types of arrays is ongoing research and new pipelines and statistical approaches have been published. An earlier publication, for example, proposed a peak-based correction (PBC) method to correct for the bias of Infinium I and II probes, which

normalizes type2 design probes so as to render them comparable to type1 probes [208]. This PBC method was implemented in IMA package, for example [206]. But two studies have uncovered potential problems with PBC and proposed subset quantile normalization methods (SQN & SWAN) to correct for the type2 bias, one of them being the pipeline used for this thesis (Touleimat and Tost, 2012) [175,210]. A very recently published study now shows that PBC often leads to discontinuities in the type2 density distribution and propose a mixture model-based normalization algorithm, called Beta MIxture Quantile dilation (BMIQ) to address this problem [211]. Many more advances in this area are anticipated in the near future.

4.2. Tobacco smoking leads to extensive changes in the human methylome

Overall, after the Illumina 450K BeadChip was successfully established in our laboratory, the actual processing of the 450K BeadChips for the 1814 KORA F4 and 479 KORA F3 samples was easy to handle. Also, after solving the problems concerning normalization and pre-processing steps, the data was reliable and of good quality. The following subchapters will discuss the results of the 450K analysis in current vs. never smokers, with which the first aim of this thesis was achieved.

4.2.1. A variety of genes are affected by tobacco smoking

Wide-spread differences in the degree of site-specific methylation as a function of tobacco smoking could be observed in each of the 22 autosomes, identifying 187 genome-wide, significant, differentially methylated CpG sites with the percent of variance explained by smoking ranging from 1.31 to 41.02 and p-values ranging from 9.31E-08 to 2.54E-182.
As indicated by the Manhatten Plot in Figure 22 and shown within Figure 26, current and former tobacco smoking was associated with genome-wide hypomethylation. Furthermore, hypomethylation seemed to be more subject to remain significant after smoking cessation with a total of 73% of sites being hypomethylated in current smokers, but a total of 92% being hypomethylated in former smokers. In current smokers, hypermethylation (defined as an increase in DNA methylation) was found to be more relevant if only sites with a percent of methylation difference that exceeded the 2.5% threshold (34% compared to 27%), and even

more if they exceeded the 5% threshold (39% compared to 27%), were taken into account. The same could not be determined in former smokers. Though, it is important to note that the number of significant sites in current (N=187) and former (N=13) smokers greatly differed and might not be directly comparable.

The observation of global hypomethylation in current and former smokers is in the line with what is known about cancer, one of the most common tobacco smoke-related diseases. Though 'targeted' hypermethylation is more recognized in cancer development, global hypomethylation is almost always linked to malignancy and disease tumor progression [139,212]. In normal human cells, more than twice as many cytosine residues are methylated compared to tumor cell DNA [213]. It also has been shown that global DNA hypomethylation is highly associated with cardiovascular diseases, Alzheimer's disease, and mental disorders [152,212]. Hypermethylation together with specific histone modifications stabilizes the genome, whereas hypomethylation allows rearrangement events, such as homologous recombination, which contribute to the etiology of many diseases, especially cancer [48].

The mechanism that leads to global hypomethylation in smokers is not known, but various hypotheses exist. It has been shown that organic nitrites, nitrous oxide, cyanates, and isocyanates present in tobacco smoke interact with folic acid and Vitamin B12 coenzymes and transform them into biologically inactive compounds [214,215]. This could result in low folic acid concentration in smokers and thus may alter the availability and enzymatic pathways responsible for methylating DNA appropriately [216]. Another study supporting this hypothesis demonstrated that oral mucosal folate levels were much lower in the buccal mucosal cells of current than in never smokers [217]. A further hypothetical unifying mechanism linking DNA hypomethylation to chemicals in tobacco smoke may be the glutathione (GSH) synthesis pathway [216], as exposure to chemicals markedly increases the need for the GSH to maintain its detoxification pathways [218]. Methyl groups from S-adenosylmethionine (SAM) are needed for DNA methylation [219]. The presence of chemicals that deplete GSH can impair synthesis of SAM and perturb DNA methylation because the methylation cycle and the GSH synthesis pathways are biochemically linked [216]. Interestingly, one gene that encodes for a member of the GST family, *GST-O1 (Omega),* was found to be significant in KORA F4, but unfortunately could not be replicated in KORA F3. *GST-O1* encodes for glutathione S-transferase-Omega, also known as CHST3 (carbohydrate (chondroitin 6) sulfotransferase 3). The other members of the GST family (*GST-Alpha, -Mu, -Pi, -Theta*) are one of the most frequently studied genes concerning genetic effects of tobacco smoking and have been shown to be involved in the metabolism of cigarette smoke leading to individual susceptibility against various pathologies including cancer, cardiovascular and respiratory diseases [220,221]. But none of these members could

be detected within this methylation study. Aberrant DNA methylation at *GST-O1* could also be detected in placental samples, demonstrating an association with maternal smoking during pregnancy [222]. Furthermore, GST-O1 levels have been shown to be decreased in COPD patients and current smokers compared to never smokers [223] and probably play a role in Alzheimer's and Parkinson disease [224].

Within candidate gene studies high levels of DNA methyltransferase 1 (DNMT1) have been linked to tobacco-caused hypermethylation of tumor suppressor genes [225,226] and have been reported in neurons and human bronchial epithelial cell cultures exposed to nicotine or cigarette smoke extract [227,228]. DNMT1 is the most abundant form of DNMTs in mammalian cells and functions primarily in maintenance methylation during DNA replication using hemimethylated DNA as substrates [229]. It has been shown that the tobacco-specific carcinogen nicotine-derived nitrosamine ketone (NNK) induces DNMT1 accumulation by inhibiting DNMT1 protein degradation through Akt signaling [230-232]. Furthermore, nicotine itself has been shown to influence not only the expression of *DNMT1*, but also promoter methylation levels in GABAergic neurons, providing a plausible link to nicotine addiction [227].

Since the 187 smoking-associated CpG sites found in this thesis correspond to 95 unique genes, not all can be discussed in detail. Therefore, the following paragraphs mainly discuss the loci of replicated CpG sites of current vs. never smokers with a methylation difference of more than 5% in both panels, described in Table 10, with special focus on the *AHRR* and *ALPP/ALPPL2* loci. Additionally, some of the genes listed in Table A.1 and A.2 that have also shown smoking-associated aberrant DNA methylation in other studies or are particulary interesting considering the known effects of tobacco smoking in general, are included in the discussion. The human nucleotide sequences of the candidate genes are annotated and can be found in several publicly available databases.

4.2.1.1. *AHRR* (aryl hydrocarbon receptor (AHR) repressor)

The human *AHRR* (aryl hydrocarbon receptor (AHR) repressor), which includes the most significant and remarkable CpG site detected (cg05575921), codes for an evolutionary conserved bHLH-PAS (basic helix-loop-helix/Per-AHR nuclear translocator (ARNT)-Sim) protein. This protein is part of the aryl hydrocarbon receptor (AHR) signaling cascade, which mediates dioxin toxicity, is involved in regulation of cell growth and differentiation [233,234] and the modulation of the immune system (by effecting for example, the development, activation, proliferation, and/or differentiation of DCs, T(reg)s, T(H)17 cells and B cells) [235]. Furthermore, evidence exists for AHR crosstalk with estrogen receptor (ER) signaling, thereby impacting cell proliferation and metabolism by P450 enzymes [236]. Tobacco smoke

is a remarkable source of polycyclic aromatic hydrocarbons (PAHs) that trigger the AHR signaling pathway [237-239], leading to several pathological effects in humans, such as cancer, asthma, COPD, and inflammatory bowel disease through AHR-dependent changes in gene expression [240-243]. Activation of AHR also promotes allograft-specific tolerance and it has been shown that tobacco smoke exposure in either donors or recipients leads to accelerated allograft rejection, vascular inflammation, and graft loss [244,245]. Furthermore, activation of AHR causes various tobacco smoking-associated cardiovascular diseases [246,247]. Although it is well established that AHRR inhibits AHR signaling, the mechanism by which this is achieved remains unclear. It was initially hypothesized that the mechanism of AHRR-mediated AHR repression involves competition between AHR and AHRR for binding to ARNT and competition of AHR-ARNT and AHRR-ARNT dimers for binding to XREs (xenobiotic responsive enhancer elements), also called AHR-responsive enhancer elements (AHREs) [233]. But this mechanism does not seem to be the whole story. A study of Evans et al. indicates that the inhibition of AHR activity through AHRR is more complicated as it demonstrates that AHRR can repress AHR function by a mechanism that is independent of both competition for ARNT and XRE binding, since AHRR is still able to act as a repressor when both mechanisms are eliminated. The authors conclude that there must be additional mechanism(s) by which the AHRR interacts with the AHR complex in a DNA binding-independent manner that requires AHRR-dependent protein-protein interactions [248]. However, the specific proteins involved are still unknown. Furthermore, as the expression of *AHRR* is induced by a variety of AHR agonists, it probably forms a negative feedback loop with the AHR [233,249] (Figure 37).

AHRR is a known tumor suppressor, mediating detoxification of PAHs, which are the principle carcinogenic agents causing tobacco-related lung and other cancers [250]. If the detected changes in DNA methylation for *AHRR* inhibit the translation of *AHRR* mRNA, this might result in superinduction of mRNAs, such as *CYP1A1*, by dioxins of tobacco smoke. *CYP1A1* encodes for Cytochrome P4501A1 and is involved in the metabolic activation of polycyclic aromatic hydrocarbon pro-carcinogens, produces reactive oxygen species (ROS) and oxidative DNA damage, and probably plays an important role in various cancers [251,252]. Even though, only one CpG site, located at the promoter region, was found to be significant in KORA F4 and was not replicated in KORA F3, it has to be mentioned that CYP1A1 is induced in the human lung up to 100-fold because of tobacco smoking, and promoter demethylation has been shown to play a role in the smoking-associated expression of this enzyme [119]. Furthermore, a recent study of Joubert et al. carried out in cord blood of newborns (450K BeadChip), as well as a study of Suter et al. carried out in placental

samples (27K BeadChip), have shown an association of differential DNA methylation in *CYP1A1* with maternal smoking during pregnancy [222,253].

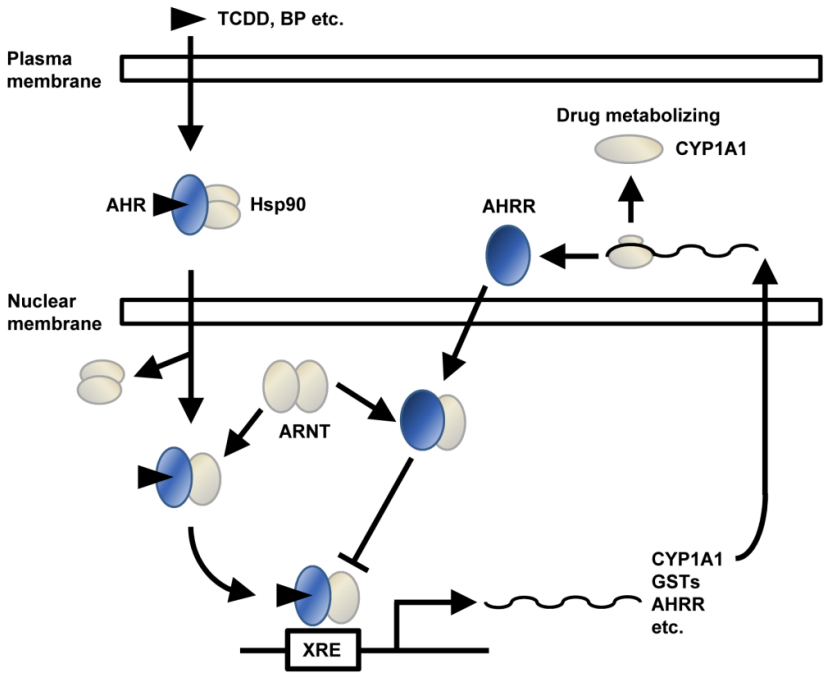


Figure 37. Mechanism of negative feedback regulation of AHR function by AHRR. The transcription factor AHR is located in the cytosol and, bound to several co-chaperones, normally inactive. Upon ligand binding of chemicals to AHR that are present in tobacco smoke, such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) or benzo(a)pyrene (BP), the chaperones dissociate, causing the translocation of AHR into the nucleus and the formation of a heterodimer with ARNT (AHR nuclear translocator). The AHR/ARNT heterodimer binds DNA regulatory sequences, known as xenobiotic response elements (XREs), which are present in the promoter region of a variety of genes, such as *CYP1A1* and other genes involved in detoxification of these chemicals, leading to changes in the transcription of these genes. Also, the expression of *AHRR* is induced by the ligand-activated AHR/ARNT heterodimer and as AHRR inhibits AHR signal transduction by competing with AHR for ARNT a negative feedback loop for the regulation of AHR is formed. Modified from [233].

Recently, a differential methylation of CpG sites in smokers within the *AHRR* gene has also been demonstrated in lymphoblasts and pulmonary macrophages by Monick et al. [254]. Moreover, the study of Joubert et al. also found cg05575921 in *AHRR* to be the most statistically significant CpG site and showed that lower methylation was associated with higher levels of maternal smoking. Furthermore, CpG sites in *CYP1A1* showed contrasting effects by higher methylation due to maternal smoking, which is interesting because of the divergent function of these genes in the AHR pathway [253]. *AHRR* was also found to be differentially methylated in a recent study of Shenker et al. carried out in whole blood. The expression of *AHR* and *CYP1A1* increased in response to cigarette smoke, and the expression of *AHRR* initially decreased after 3 days, before increasing after 28 days [255]. Another study concerning gene expression found *AHRR* to be the most significant probe set

between never and current smokers with a fold change of 6.1 [256]. Very recently it has been shown that *AHRR* demethylation at cg05575921 in peripheral cells may serve as an early, sensitive biomarker for even low levels of exposure to tobacco smoke, providing a non-self-report alternative for nascent exposure to tobacco smoke [257].

4.2.1.2. *ALPP* (placental) and *ALPPL2* (placental-like alkaline phosphatases)

The second most outstanding CpG site respective to significance and level of detectable changes in DNA methylation patterns associated with tobacco smoking (cg21566642) was located in a relatively large CGI, not annotated to any gene. Another five CpG sites (cg06644428, cg05951221, cg01940273, cg13193840, and cg03329539) were also located within or in the shore of this CGI, which is located 9 kb apart from the 3'UTR of the *ALPPL2* gene and lies within the promoter region of the non-coding RNA *ECEL1P1*. Nonetheless, SNPs within this CGI are predicted to have a functional impact on the *ALPPL2* gene (http://genome.ucsc.edu/; UCSC's predicted function relative to selected gene tracks) and therefore aberrant methylation might also have an effect on this gene. Moreover, another three of the detected and replicated CpG sites (cg23667432, cg03188382, and cg19713851) were located within or in the shore of CGIs of the related *ALPP* gene.

At least four distinct but related alkaline phosphatases (ALPs) can be found in humans: The placental (ALPP), placental-like (ALPPL2; also known as Nagao isozyme or germ cell alkaline phosphatase), and intestinal (ALPI) forms are located together on chromosome 2, while the liver/bone/kidney (ALPL; tissue non-specific) one is located on chromosome 1. These enzymes are responsible for the dephosphorylation of various molecules, such as proteins, nucleotides, or alkaloids and are present in all tissues throughout the body with *ALPP* being expressed primarily in the placenta, whereas ALPPL2 is localized in testis, thymus, and certain germ cell tumors. *ALPP* and *ALPPL2* show 98% amino acid sequence homology [258-260]. Quantitative variations of circulating ALP concentrations are associated with premature birth [261,262], low birth weight [263,264], and pre-eclampsia [265]. Serum ALPP and ALPPL2 enzyme levels are increased up to 10-fold in 80% of cigarette smokers [266-268], probably originating from the lung [269], and were shown to be elevated in patients with a number of cancers, especially seminoma [270,271].

CpGs in this region were also found to be differentially methylated in pulmonary macrophages within the study of Monick et al. [254] and in whole blood within the study of Shenker [255]. The same group could further show an association of cg01940273 (KORA F4: -7.89%, p = 9.28E-114, explained variance = 31.50%; KORA F3: -7.53%; p = 5.46E-43, explained variance = 28.33%) with developing breast cancer. It is important to mention that this group hypothesizes that the mechanism may involve *in cis* regulation of the

developmentally regulated *ECEL1*, as the region including cg01940273 is a potential pseudogene of *ECEL1*, possessing a high level of sequence homology (>95%) [255].

4.2.1.3. Further smoking-associated genes

The smoking-associated genes detected are involved in a variety of biological processes.

Nicotine dependence

CHRNA5 encodes for cholinergic receptor, nicotinic, alpha 5 (neuronal) and was found to be significant in KORA F4. Genome-wide association studies identified *CHRNA5* and also *CHRNA3* (cholinergic receptor, nicotinic, alpha 3 (neuronal)) as susceptibility loci for lung cancer [272,273]. Furthermore, smokers which carried the *CHRNA3* and *CHRNA5* variants were shown to be exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine [274]. Nicotine is one of the most harmful substances in tobacco products that is strongly addictive [92]. Various studies of *CHRNA5* and CHRNA3 were published concerning genetic effects and their role in nicotine dependence and smoking intensity as well as various diseases, such as COPD and lung cancer [275,276], but no study until now has shown epigenetic effects in association with smoking.

DNA repair

RAD52 encodes for RAD52 homolog (S. cerevisiae) and was found to be significant in KORA F4 and F3. RAD52 has been shown to interact with RAD51, which suggests a role of RAD52 in RAD51 related DNA recombination and repair [277]. RAD51 participates in homologous recombination (HR) to maintain chromosome stability and repairs chromosomal damage, such as breaks, translocations, and deletions [278]. A study could show that knock down of *RAD52* is synthetically lethal (approached by selective cell killing based on genotype) in *BRCA2* (breast cancer 2, early onset)-depleted cells, resulting from decreased RAD51 foci formation and double-strand break-induced HR [279]. In addition to RAD51 several RAD51 paralogs exist (RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3) that play roles in DNA damage signalling, recombinational repair, and tumorigenesis [280]. Two of them, *XRCC3* (X-ray repair complementing defective repair in Chinese hamster cells 3) and *RAD51B* (RAD51 homolog B (S. cerevisiae); named *RAD51L1* in the Illumina annotation), were found to be significant in the discovery panel KORA F4. RAD51B has been shown to form a stable heterodimer with RAD51C, which further interacts with the other RAD51 paralogs [281]. Moreover, recently two distinct complexes, RAD51B-RAD51C-RAD51D-XRCC2 (BCDX2) and RAD51C-XRCC3 (CX3), were found to act at different stages of the HR pathway, both being epistatic with BRCA2 and synthetically lethal with RAD52 [282]. Therefore, RAD52

may play a key survival role in cells lacking the function of the BRCA1-BRCA2 pathway of HR and may provide back-up RAD51 function for all members of the BRCA1-BRCA2 pathway [283]. Importantly, *BRCA2* also was found to be significant in KORA F4. *BRCA2* is a tumor suppressor gene that is involved in maintenance of genome stability. Human BRCA2 protein might play a role in delivering and loading RAD51 onto single-stranded DNA generated after resection of DNA double-strand breaks [284]. Furthermore, BRCA2 has very recently been shown to be epistatic to the RAD51 paralogs in response to DNA damage [285].

Tobacco smoke contains many carcinogens and reactive oxygen species that produce DNA adducts, cross-links, and DNA strand breaks and various repair pathways exist, which normally protect a human cell against these DNA damages including homologous recombination [286,287]. Reduced capacity for repair of DNA damage, or only at a slower rate, is associated with cancer, infertility, cardiovascular disease, neurodegeneration, and aging [286-288] (Figure 38). Epigenetic inactivation of genome stability control genes by promoter hypermethylation has already been detected for *BRCA1, BRCA2, XRCC3,* and *XRCC5* [288,289]. Also, variants in the *RAD51B* gene were associated with nasopharyngeal carcinoma (NPC) risk in a Chinese population and the authors state that cigarette smoking showed potential interactions with these gene variations [290]. The novel finding of this thesis is that tobacco smoking is associated with DNA methylation changes in *BRCA2, XRCC3, RAD52,* and *RAD51B*, which play an important role in homologous recombination and maintenance of genome stability. Possible aberrant expression of these genes due to DNA methylation changes caused by tobacco smoking may lead to genome instability and various types of diseases (Figure 38).

Cell cycle control

CDKs encode for cyclin-dependent kinases and cyclin-dependent kinase inhibitors which function as checkpoint regulators for the G1/S and G2/M phases. Aberrant DNA methylation in response to tobacco smoking has been demonstrated for *CDKN2A-p16* and *CDKN2B-p15* [123,128,144,145], but not for *CDKN1A-p21*. Within the study of this doctoral thesis, *p21* was found to be significant in KORA F4 and F3.

It has been demonstrated that cigarette smoke induces cellular senescence, the phenomenon by which normal diploid cells lose the ability to divide, which may lead to smoking-related chronic lung diseases, such as COPD [291,292]. Cellular senescence may be induced through either the ARF-p53 pathway or the p16-retinoblastoma protein (Rb) pathway. Three genes that showed aberrant smoking-associated DNA methylation play a role in these pathways, including *CDKN1A* and further *Rb* and *CDK6,* which were found to be

significant in KORA F4, but could not be replicated in KORA F3. ARF docks the mouse double minute (MDM2) protein and extenuates MDM2-dependent Tumor protein p53 (TP53) degradation, leading to elevated p53 levels. P53 drives the transcription of genes associated with cell cycle arrest, DNA repair, and apoptosis, like *p21*, which functions as a checkpoint regulator for the G1/S and G2/M phases and causes cell cycle arrest. p16, an activator of Rb, docks CDKs as CDK4 and CDK6 and suppresses CDK activation through inhibition of cycling binding, resulting in Rb activation. Activated Rb induces chromatin modifications that suppress growth-promoting transcription factors, such as E2F transcription factor 1 (E2F) [292]. All of these genes have been subject of tobacco smoking association studies prior to this thesis, but mainly in relation to genetic and not epigenetic effects. Another gene that was found to be significant in KORA F4 and F3, *SKI*, might also play a role in these pathways. *SKI* encodes for v-ski sarcoma viral oncogene homolog (avian) which functions as a repressor of TGF-beta signaling and has been shown to activate Wnt/beta-catenin signaling in human melanoma [293]. It recently has been demonstrated that overexpression of *SKI* leads to p53 degradation through up-regulation of MDM2 activity by SKI [294]. SKI has not previously been implicated in responses to tobacco smoke.

Furthermore, it has been shown that DNA damage contributes to cellular senescence [295]. To afford adequate time for DNA repair to occur, cell cycle checkpoints delay the cell cycle progression [296]. If the damage cannot be repaired, checkpoint signaling also activates pathways leading to apoptosis [297]. Defects in cell cycle checkpoints caused by tobacco smoking that may result because of smoking-associated aberrant DNA methylation in various detected genes may therefore contribute to the ageing process and to the pathogenesis of various smoking-associated diseases, such as coronary heart disease (CAD), osteoporosis, atherosclerosis, COPD, and cancer [292,298] (Figure 38).

Regulation of transcription

HIVEP3 encodes for the transcription factor human immunodeficiency virus type I enhancer binding protein 3 and is also known as *Schnurri-3*. It plays a critical role in inflammatory and apoptotic responses as well as cell growth [299,300]. HIVEP3 is induced by T-cell receptor signaling and positively regulates the expression of *IL-2* in T-cells [301]. Furthermore, it is an essential regulator of adult bone formation [302]. Aberrant DNA methylation at *HIVEP3* could also be detected in placental samples, demonstrating an association with maternal smoking during pregnancy [222].

GFI1 encodes for the growth factor independent 1 transcription repressor and functions as a transcriptional repressor by controlling histone modifications that lead to silencing of the target gene promoters [303]. It plays a role in various developmental contexts [304] including

lymphocyte development and activation [305], plays a role in proliferation, apoptosis, differentiation, cell fate specification and oncogenesis [303], and is aberrantly expressed in lung tumors [306]. GFI1 plays a critical role in enhancing Th2 cell expansion and in repressing induction of Th17 and CD103(+) iTreg cells [307]. GFI1 coordinates epigenetic repression of p21Cip/WAF1, which is an important inhibitory regulator of cycle progression after oxidative stress [303]. *P21* was also found to be significant in this thesis and was already discussed in the paragraph concerning cell cycle control genes (page 111). Tobacco smoking leads to oxidative stress, and oxidative stress might contribute to chronic inflammation in the airways through increased p21 levels and therefore reduced apoptosis of alveolar macrophages and bronchial epithelial cells [308]. Also, the transcription of this gene has been shown to be altered by tobacco smoking in lymphocytes [309].

RARA (retinoic acid receptor, alpha) regulates the expression of target genes in a ligand-dependent manner and plays a role in acute promyelocytic leukaemia [310], germ cell development during spermatogenesis [311], and CD4+ T-cell immunity and homeostasis [312]. It plays an important role in cellular memory and imprinting by regulating the CpG methylation status of specific promoter regions [313]. Associations between altered expression of nuclear retinoic acid (RA) receptors and the malignant transformation of human cells have been demonstrated in several studies, and in some patients smoking and alcohol consumption may contribute to the alteration of RA receptor levels in esophagus and result in the development of squamous-cell carcinoma [314]. Not *RAR-alpha* but *RAR-beta,* which belongs to the same family but is located on a different chromosome, has been shown to be frequently silenced in human lung cancers by promoter hypermethylation [142,143].

ZC3H3 encodes for zinc finger CCCH-type containing 3. Very little is known about this gene but one study could show that it regulates mRNA nuclear adenylation and export [315].

Signal transduction

GNG12 encodes for the guanine nucleotide binding protein (G protein), gamma 12. G proteins are involved as modulators or transducers in various transmembrane signaling systems and are composed of 3 units, alpha, beta, and gamma [316]. The beta and gamma chains are required for the GTPase activity, for replacement of GDP by GTP, and for G protein-effector interaction. GNG12 may be a negative regulator of lipopolysaccharide-induced inflammation [317]. A genome-wide association for smoking cessation success detected *GNG12* as one of the genomic regions that contain clustered, nominally positive SNPs for success in smoking cessation [318].

CACNA1D encodes for calcium channel, voltage-dependent, L type, alpha 1D subunit, also known as *Cav1.3*. Voltage-gated Ca^{2+} channels divert Ca^{2+} signals to different cellular

processes within different cell types, such as muscle contraction, neurotransmitter release, hormone secretion, gene expression, cell motility, cell division, and cell death [319]. Cav1.3 can signal to transcriptional events and induce long lasting alterations of neuronal responsiveness [91,320], and mutations in the gene contribute to hearing impairment [321]. It recently has been shown that CaV1.3 may play a crucial role in osmotic stress-induced Ca^{2+} influx and tight junction disruption in the intestinal epithelium [92]. A very recent study showed that CACNA1D is important in human glucose-induced insulin secretion and that common variants in the gene might contribute to type 2 diabetes [322]. Various studies have associated tobacco smoking with an increased risk of type 2 diabetes [164,166]. Moreover, a study that was assessed in the same study cohort used for the 450K BeadChip analysis, KORA F4, could show an association of passive and active smoking with type 2 diabetes [165]. A causality between smoking and diabetes might be that smoking is a possible risk factor for insulin resistance, a precursor for diabetes, and that it might deteriorate glucose metabolism, which may lead to the onset of type 2 diabetes [323-325]. Interestingly, various studies could show that smoking may have a significant effect on calcium metabolism and impaired calcium absorption [326-328].

F2RL3 encodes for coagulation factor II (thrombin) receptor-like 3 and is also known as *PAR4* (protease-activated receptor 4). The F2RL3 protein is relevant for cardiovascular physiology and plays a role in platelet activation [329] and cell signaling [330]. High glucose enhances smooth muscle cell responsiveness to thrombin through transcriptional upregulation of *F2RL3*, which may play an important role in the vascular complications of diabetes [331]. Furthermore, *F2RL3* expression is down-regulated in gastric cancers [332]. Very recently a study reported an association of *F2RL3* methylation with mortality among patients with stable coronary heart disease [333].

TIAM2 encodes for T-cell lymphoma invasion and metastasis 2, which is a guanine nucleotide exchange factor that stimulates the GDP-GTP exchange activity of RHO-like GTPases and activates them. It connects extracellular signals to cytoskeletal activities. The encoded protein may play a role in neural cell development [334,335] and regulates cell migration by microtubule-mediated focal adhesion disassembly [336]. Recently, it has been shown that the expression of *TIAM2* promotes proliferation and invasion of liver cancer [337].

GPR15 encodes for G protein-coupled receptor 15 and acts as a co-receptor for the human immunodeficiency virus [338]. *GPR15* shares homology with receptors for Angiotensin II, angiotensin AT1 and AT2, which is a potent vasopressor hormone that controls blood pressure and volume in the cardiovascular system [339]. It has been shown that AT1-receptor blockade prevents smoking-induced impairment of endothelium-dependent vasodilation in rats [340].

Cell adhesion

CNTNAP2 encodes for contactin associated protein-like 2, a member of the neurexin family which functions in the vertebrate nervous system as cell adhesion molecule and receptor. *CNTNAP2* has been associated with a wide spectrum of neuropsychiatric disorders, such as developmental language and autism spectrum disorders, epilepsy, and schizophrenia [341]. Furthermore, it undergoes aberrant methylation in pancreatic cancer [342,343] and acts as a tumor suppressor gene in glioma [344]. The transcription of this gene has been shown to be altered by cigarette smoke condensate in human bronchial epithelial cells [345].

PCDH9 encodes for protocadherin 9. Protocadherins are a subfamily of cadherins, a large group of related glycoproteins that mediate calcium-dependent cell-to-cell adhesion via a homophilic mechanism. PCDH9 is localized to the cell membrane and expressed primarily in the brain. It is found in synaptic junctions where it functions as a neuronal receptor and is involved in signal transduction and maintaining specific neuronal connections [346]. Expression of *PCDH9* is found in hairy cell leukemia [347] and PCDH9 might function as a tumor suppressor during cancer development and progression [348]. *PCDH9* might furthermore be a susceptibility locus for rheumatoid arthritis [349]. Recently, a study characterized intra- and inter-individual methylomic variation across whole blood and multiple regions of the brain from multiple donors and found tissue-specific differentially methylated regions to be significantly enriched near genes involved in functional pathways related to neurodevelopment and neuronal differentiation, including *PCDH9* [350]. Rare variants in *PCDH9* have been described as genetic risk factors for Autism Spectrum Disorders (ASD), accompanied by mutations in two other genes detected in this thesis, *CNTNAP2* (common and rare variants) and *CACNA1D* (functional variants) [351].

Membrane trafficking

MYO1G encodes for myosin 1G, which is a plasma membrane-associated class I myosin. It is present in T and B lymphocytes and mast cells [352,353] and regulates cell elasticity [354]. Depletion of MYO1G induces non-apoptotic cancer cell death [355].

LRP5 encodes for low density lipoprotein receptor-related protein 5, which binds and internalizes ligands in the process of receptor-mediated endocytosis. LRP5 is involved in canonical Wnt signaling that regulates bone metabolism, plays a role in regulating bone mass [356], and may play a role in smoke-induced bone loss [357]. It is involved in the development of lung microvessels and alveoli through the angiopoietin-Tie2 pathway [358]. LRP5 may further contribute to the glucose-induced insulin secretion in the islets [359] and play a role in the pathophysiology of diabetes and metabolic syndrome [360]. Increased DNA methylation at a promoter CpG island of *LRP5* could be detected in ovarian tumors [361].

LINGO-3 encodes for leucine rich repeat and Ig domain containing 3 and belongs to the leucine-rich repeat transmembrane neuronal family (LRRTM) together with three other human paralogs (*LINGO-1, -2* and *-4*). *LINGO-3* shares 53% amino acid identity with *LINGO-1*. LINGO-1 shows similarity to proteins involved in axonal guidance and migration, nervous system development, and regeneration processes and may play a role in the molecular and cellular neurobiology of vertebrates [362].

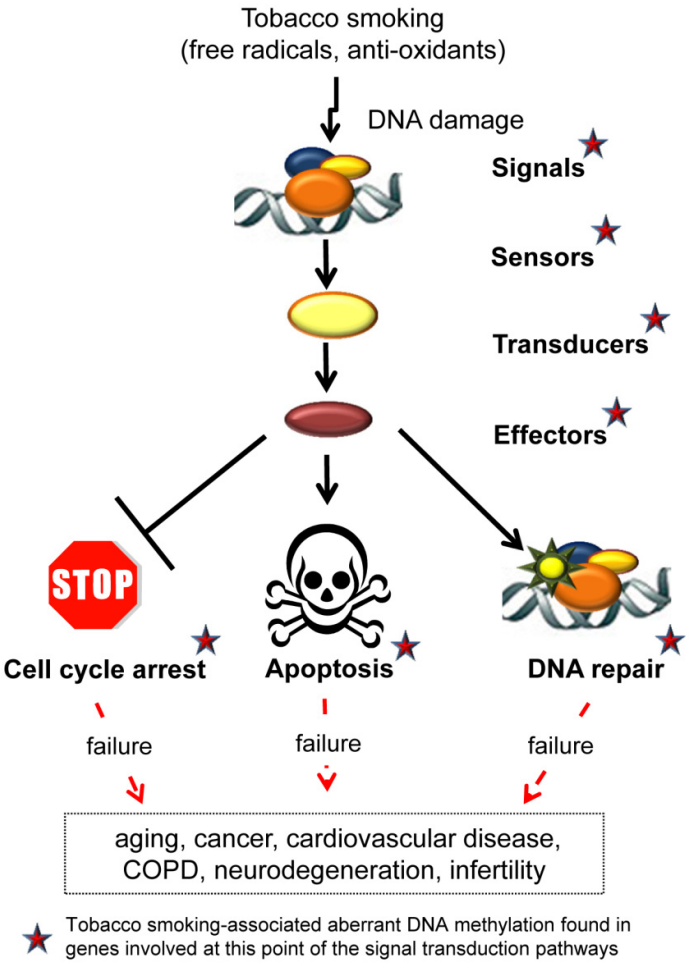


Figure 38. Possible effect of smoking-related methylation changes on DNA damage response. DNA damage that may occur from various chemicals present in tobacco smoke is recognized by sensor proteins that then initiate a network of signal transduction pathways. This ultimately results in the activation of effector proteins that execute the functions of the DNA damage response, including recruitment of DNA repair proteins, cell cycle arrest, or the induction of apoptosis. Tobacco smoke may not only lead to DNA damage but may also influence the network of signal transduction pathways, that normally help the body to stay protected from further damage, through changing the DNA methylation patterns of the involved genes. Therefore, tobacco smoking may cause DNA damage and may further hamper the body from fighting it by blocking the network of signal transduction pathways at various points, leading to the development and progression of various pathological outcomes. DNA damage part (not the connection with tobacco smoking) adapted from [363].

Inflammation

MMP9 encodes for matrix metallopeptidase 9 and was found to be significant in KORA F4. It plays a role in the pathogenesis of the cigarette smoke-induced COPD [364], with MMP9 levels being elevated in the lungs of cigarette smoke-exposed mice and reduction to normal levels after smoking cessation [365]. A study found significantly higher *MMP9* expression levels in placental sections from preeclampsia tissue and the increased expression was well correlated with promoter demethylation [366]. Paradoxically, preeclampsia risk is reduced in smokers by as much as 50 percent [367]. Another study showed that eosinophil stimulation with the combination of IL-3 (interleukin 3) plus TNF (tumor necrosis factor) led to increased steady-state levels of *MMP-9* mRNA, prolonged mRNA stabilization, and enhanced activation of ERK1/2 phosphorylation [368]. Interestingly, also *TNF* and *IL-3* were found to be significant in KORA F4, even though it has to be mentioned that for all three genes a smoking-dependent methylation difference of only below 1% was detected. The ERK cascade plays a role in cell adhesion, cell cycle progression, cell migration, cell survival, differentiation, metabolism, proliferation, and transcription [369].

Tobacco smoke has been shown to alter a range of immunological functions including innate and adaptive immune responses [370], and inflammations as a result of tobacco smoke have been thought to play a role in carcinogenesis of a variety of cancers [371].

Unknown function

The detected sites within the intergenic region at 6p21.33 are located in close proximity to each other in a gene desert which maps a DNase I hypersensitivity site. Various transcription factor binding sites are located in this region and are associated with an H3K27 acetylated chromatin site (http://genome.ucsc.edu/), which is often associated with active regulatory elements. Nothing is known about the possible function of this region as it is not associated with any gene or SNP and therefore might be an interesting research question for future studies.

Many loci detected have not previously been implicated in responses to tobacco smoke, such as *AHRR, HIVEP3, LINGO-3, GPR15, F2RL3, PCDH9, CNTNAP2, ZC3H3, LRP5, GFI1, GNG12, RARalpha* (only beta was known), or the intergenic region at 6p21.33. During the time of this doctoral thesis, other studies have also reported a significant association of tobacco smoking with aberrant DNA methylation for some of the above-mentioned genes: CpG site cg03636183, located within the *F2RL3* gene (current smokers; KORA F4: -14.74%, p = 2.42E-80; KORA F3: -17.63%, p = 1.65E-39) could be detected in most studies carried out with the 27K or 450K BeadChip [254,345,372,373]. Replication of sites found within

Shenker and co-workers, accompanied by additional findings for the corresponding regions (please see Table A.2), could be achieved for the genes *GNG12* (cg25189904), *GFI1* (cg09935388), *CNTNAP2* (cg25949550), and *LRP5* (cg21611682). Furthermore, this group found one of the four CpG sites of the intergenic region at 6p21.33 detected in this thesis (cg06126421) [255]. *CNTNAP2* was also significant in the study of Wan et al., but with another CpG site (cg16254309) [373]. Also, *GPR15* was found within the study of Wan et al. and in the study of Breitling et al. [372,373]. *GNG12* (cg25189904, cg26764244), *GFI1* (cg09662411 and cg09935388), *MYOIG* (cg22132788 and cg04180046), and *CNTNAP2* (cg25949550) could be identified at genome –wide significance in relation to maternal smoking as well, by the study of Joubert et al. [253].

Summarized, all genes that showed significant changes in methylation patterns between current and never smokers make sense in a biological context when it comes to the impact of smoking. They play roles in the development and function of the cellular, hematological, immune, cardiovascular, tumorigenic, and reproduction systems and have been linked to all kind of tobacco smoking-associated diseases. The results of this thesis suggest that the influence of the identified genes on smoking-associated diseases might be mediated by a change in the methylation status of these genes. However, clarification of the exact mechanisms and causality are subject to further studies.

4.2.1.4. Gender-specific effects of tobacco smoking are marginal

Two sites were found to be only significant in the separate male analysis (Table 11). Cg00395697, annotated to the *UHRF1BP1L* (UHRF1 binding protein 1-like; also known as SHIP164) gene, which regulates syntaxin 6-dependent sorting from early endosomes [374] and cg05498905, annotated to the *TBC1D14* (TBC1 domain family, member 14) gene. TBC1D14 has recently been shown to regulate autophagosome formation via Rab11- and ULK1-positive recycling endosomes [375] and autophagy is an evolutionarily conserved degradative process, which dysfunction is associated with cancer, neurodegeneration, microbial infection, and ageing [376]. Even though both sites showed only a slight difference in methylation between current and never smokers (cg05498905: 1.49%; cg00395697: -0.32%), it is of notice that both play a role in the regulation of endosomes. Overall, the difference in DNA methylation between current and never smokers was found to be only slightly more pronounced in males than in females. Most CpG sites detected in the model for men, in addition to the nine overlapping CpG sites, were close to the genome-wide significance level observed in the female model, which explains why no significant CpG sites were detected for the interaction term (Table 11).

A genome-wide study that investigated the influence of sex on methylation in saliva and blood cells, found 81% of genes on the X chromosome and 8% of genes on autosomes to be affected by sex [182]. Another study that analyzed the methylation difference of a large number of CpGs on three human chromosomes also identified a relatively small mean methylation difference (0.1%) between males and females [183]. Within a candidate gene study, infants born to smokers had higher methylation at the *IGF2* loci and the smoking-related increase in methylation was most pronounced in male offspring [377]. Another study that investigated DNA methylation states in cancer-associated genes showed that methylation levels in *RASSF1A* were influenced by sex, with males showing higher levels of methylation, but this gene was not associated with smoking [378]. Even though a few studies exist showing gender differences in DNA methylation patterns, these differences seem to be small, which is in line with the results of this doctoral thesis. Still, small differences in methylation patterns, if present at critical regulatory genes, may be a basis for differential susceptibility to complex diseases (most notably cancer) between men and women [379], therefore being an important aspect for individualized medicine needs.

4.2.2. The positive effect of quitting tobacco smoking

The next question was if former tobacco smoking would have an effect on DNA methylation in general and which role cessation time and pack-years may play in that matter. With the investigation of this subject the second aim of the present thesis could be achieved. Compared to the 187 replicated sites in current smokers, the 13 CpGs that were still significant in former smokers (Table 12) were among the top 55 CpGs in respect to significance, and among the top 156 CpGs in respect to change of methylation level. Thus, the significance and level of DNA methylation change of certain CpGs in current vs. never smokers did not necessarily allow an assumption of which CpGs would remain significant in former smokers. Nevertheless, the top four CpGs in respect to significance and the top three CpGs in respect to change of methylation level remained, and the regions of *AHRR* and *ALPPL2* were still the most predominant ones. The corresponding genes were already discussed in chapter 4.2.1., only the gene *PRSS23* (protease, serine 23) that corresponds to two of the CpGs (cg11660018, cg23771366) has not yet been discussed, as it showed a methylation difference of under 5% in the current smoker analysis. A recent study showed that *PRSS23* is up-regulated by estrogen receptor α and associated with proliferation of breast cancer cells [380]. Moreover, the transcription of this gene has been shown to be

altered by tobacco smoking in lymphocytes [309] and by cigarette smoke condensate in human bronchial epithelial cells [345].

Summarized, even though a few CpG sites were still significant the genome-wide DNA methylation levels in former smokers were found to be close to the ones seen in never smokers, suggesting that quitting tobacco smoking potentially allows regaining the DNA methylation state of never smokers. One of the first studies concerning the reversible smoke effects on DNA methylation has shown that knockdown of DNA methyltransferase 1 (DNMT1) leads to promoter demethylation and gene re-expression in human lung and breast cancer cells [381]. As mentioned earlier, the tobacco-specific carcinogen NNK induces DNMT1 accumulation by inhibiting DNMT1 protein degradation through Akt signaling [230,231,369]. Further, reversible smoke effects on DNA methylation have been documented in cultured lung cancer cells where the oncogene synuclein-γ (*SNCG*) is silenced by CGI methylation. Cigarette smoke extract induced *SNCG* demethylation, associated with a twofold decrease of *DNMT3B* mRNA, was accompanied by an overexpression of the gene in just three days of treatment. By ending the treatment the *DNMT3B* expression was normalized, accompanied by the reestablishment of the *SNCG* CpG methylation [382]. As described in the introduction part, the reactions of de novo methylation using unmethylated DNA are thought to be primarily catalyzed by DNMT3A and DNMT3B [39,40]. However, no compelling evidence has yet been found for differential expression in DNA methyltransferases by current smoking status in whole blood, histologically normal lung tissue, or bronchial epithelium. Therefore, alternative mechanisms may contribute to the establishment of locus-specific differential DNA methylation detected in this study [373].

4.2.2.1. The effect of cessation time on DNA methylation

Cessation time was found to have an effect on the degree of DNA methylation in 36 CpGs (Table 13), including all 13 CpGs that remained significant in former smokers. All of these 36 sites were also current smoking-associated sites, with 33 of them being significant in KORA F4 and F3, and three of them being significant only in F4 (cg23565821: *CUTA* (cutA divalent cation tolerance homolog (E. coli)); cg12303084: *ZMYND8* (zinc finger, MYND-type containing 8); cg09084200: *VPS26B* (vacuolar protein sorting 26 homolog B (S. pombe))). Interestingly, these three sites show opposed methylation differences with hypomethylation in current and hypermethylation in former compared to never smokers, but all with a methylation difference of less than 1%. The three additional genes in Table 13 are *HNRNPA1* (heterogeneous nuclear ribonucleoprotein A1), *ANPEP* (alanyl (membrane) aminopeptidase) and *MIR802* (microRNA 802). The recent study of Wan et al., carried out with the 27K BeadChip, was able to detect three sites that were differentially methylated

according to time since quitting. Within the study of this doctoral thesis, only the site within the *F2RL3* gene (cg03636183) could be replicated, the others (cg19859270 - *GPR15* and cg09837977 - *LRRN3)* could not be confirmed [373].

Overall, the degree of methylation difference between current vs. never smokers did not seem to have an impact on how close former smokers could come to the state of never smokers. For example, cg05575921 within *AHRR* which exhibited the highest difference in median ß-value methylation (current smokers; -21.09%), showed a relatively fast approach to the methylation level of never smokers within the first years of quitting. Although, this approach seemed to stagnate after a few decades as the median ß-value methylation level of former smokers never completely approached the level of never smokers (former smokers; -3.31%). Another CpG site, cg06644428 of *ALPPL2*, showed a much lower methylation difference in current compared to never smokers (current smokers; -7.05%) but remained significant in former smokers with a methylation difference of -3.09%, which is comparable to the one of cg05575921. This might explain why the 13 CpGs that remained significant in former smokers were not necessarily the ones showing the highest percent of methylation difference in current smokers. A recent large-scale whole-genome gene expression study by Bosse et al., carried out on non-tumor lung tissue from patients with lung cancer, showed that the expression of most genes with altered smoking-dependent expression reverted to the levels of never smokers, but some genes also showed very slow or no reversibility in expression [256]. Moreover, within this study *AHRR* was found to be the most significant probe set between never and current smokers with a fold change of 6.1. Upon smoking cessation the expression of this gene fell extensively, but changes slowed down substantially in later years never reaching the level of never smokers, which corresponds to the progress of DNA methylation changes detected within this gene (Figure 23) [256].

Summarized, these results indicate a long-term gene expression consequence of prior smoking history, at least for some genes, including *AHRR*, which may be locked in by DNA methylation changes.

That aberrant DNA methylation changes can still be detected after quitting tobacco smoking fits to the common observation that former smokers are still at a higher risk for various diseases, such as COPD and various cancers, even after years of cessation. Moreover, the genes that have been found to be still differentially methylated in former smokers are highly plausible. For example, the risks of former smokers for cancer of the mouth, throat, esophagus, and bladder drop by half after 5 years and for lung cancer after 10 years, but are still higher even decades after cessation when compared to never smokers [84]. As mentioned above, the feedback modulation of AHRR plays a pivotal role in AHR regulation

and may be critical in moderating AHR role in oncogenesis and altered immune function, leading to various types of cancers, COPD, and asthma [243,250]. Interestingly, within the former smokers of the study of this thesis, the period of time required for the CpG sites of *AHRR* to gain a methylation state that is 50% closer to the DNA methylation state of never smokers was roughly 5 to 10 years (Figure 23 and Figure A.3).

Further, a causative relationship between smoking and coronary heart disease (CHD) is well established and also that smoking cessation reduces the risk of cardiovascular morbidity and mortality for smokers with or without coronary heart disease. How many years former smokers need to recover is debated, as some studies show large reductions in risk after only 2 to 3 years, whereas other studies show higher risks even 20 years after quitting or lifelong [147,150]. Interestingly, for *F2RL3* - one of the genes still significant in former smokers in the study of this thesis (Table 12) - a study recently reported an association of methylation at the same CpG site with mortality among patients with stable coronary heart disease [333].

4.2.2.2. Cessation time and pack-years have an effect on DNA methylation

Pack-years were found to have an effect on the degree of DNA methylation in former smokers in 14 of the 187 CpGs (Table 14), including all of the 13 CpGs that remained significant in former smokers. Also, with this approach the regions *AHRR* and *ALPP/ALPPL2* remained the most significant ones. The study of Wan et al. also assessed the effect of pack-years on aberrant DNA methylation and detected two sites that were significant (cg03636183 – *F2RL3* and cg19859270 - *GPR15*) [373]. Again, within the study of this doctoral thesis, only the site within the *F2RL3* gene could be replicated.

The DNA methylation state of three CpG sites was found to be influenced by both, cessation time and cumulative smoke exposure, and interestingly again these sites represented top hits of both the *AHRR* (cg05575921) and *ALPP/ALPPL2* (cg05951221, cg21566642) regions. For these three CpG sites it could be shown that the greater the number of pack-years, the greater the influence of cessation time on the level of DNA methylation (Figure 25). The largest percentage change in DNA methylation to the levels of never smokers was detected in former smokers who had the highest number of pack-years and quit recently, which underlines the value of quitting even in the case of a long and intensive history of smoking. However, even though showing great convergence to the level of never smokers, probands with the highest pack-years never reached the state of never smokers even after a long time of cessation. For former smokers who had only low pack-years the difference between having quit recently or long ago was not as pronounced, as they showed DNA methylation levels comparable to the ones of never smokers. Therefore, pack-years might have a greater influence on whether aberrant DNA methylation regenerates to the methylation state of never

smokers than cessation time, suggesting a dose-response relationship. A recent study also showed that methylation levels of subjects with the longest cessation time and the lowest cumulative smoke exposure were closest to the levels observed in never smokers [373].
Interestingly, these results perfectly fit with previous knowledge about the dose-response relationship of smoking and disease. Heavier smoking for a longer duration results in a higher risk of disease, as a strong relationship can be found between the overall amount of cigarettes smoked by an individual and disease [84].

4.2.3. Genomic location of significant CpG sites may play a role

Until recently the majority of studies have focused on promoters and CpG islands (CGIs) where DNA methylation was thought to have the highest functional relevance to gene expression control, which led to the identification of many candidate genes associated with various disease phenotypes. Nevertheless, as only 7% of CpG sites are located within CGIs, it is highly probable that many potentially informative sites have not yet been found [383]. Recent work has identified regions up to 2 kb from the islands, termed CGI shores, that play a more significant role in the regulation of gene expression than do CGIs themselves and show variable tissue- and disease-specific methylation [205,384].
Within the study of this thesis this observation could be confirmed, as an overrepresentation of current smoking-associated sites located in the N-Shore and an underrepresentation of sites located in a CGI were detected, emanating from the percentage of covered sites by the 450K BeadChip (Figure 29). For example, according to UCSC Genome Browser (http://genome.ucsc.edu/), the *AHRR* gene contains eight CGIs and 10 out of the 16 significant CpG sites were located in relation to these CpG rich regions. Interestingly, only one site (cg11554391) was actually located in a CGI itself, the other 9 sites were found in the shores and shelves of the CGIs. Within the *ALPP/ALPPL2* region, 6 of the 10 significant CpGs were located in a CGI itself and 4 were found in shores.
In former smokers, significant sites located in the N-Shores of CGIs and in Islands were much more predominant than sites located in the Open Sea, N-Shelf, and S-Shelf (Figure 29). Though it is important to note that the number of significant sites within this group is very small (N=13). One explanation for this observation could be that although methylation patterns may not be maintained at the level of a single CpG nucleotide, CGIs keep their overall unmethylated or methylated state extremely stable through multiple cell generations [3]. Above a critical CpG density, neighboring CpGs influence each other's DNA methylation states so strongly that individual CpGs cannot stably maintain DNA methylation states

deviating from those of their neighbors. In contrast, individual CpGs in CpG-poor regions lack this pressure from neighboring CpGs [385]. Summarized, sites located in the N-Shores of CGIs and in Islands may more likely be locked in their DNA methylation state, whereas other sites may more likely reestablish their DNA methylation states upon smoking cessation.

DNA methylation can occur in different regions of the genome and the alteration of normal patterns can lead to aberrant gene expression in the cells and therefore disease. CpG islands and shores at the promoter regions of genes are normally unmethylated, allowing transcription, whereas aberrant hypermethylation leads to transcriptional inactivation. The majority of methylated CGIs can be detected in intragenic (gene body) and intergenic regions, where methylation facilitates transcription preventing spurious transcription initiations. In contrast to the promoter regions the gene body tends to demethylate in disease, allowing transcription to be initiated at several incorrect sites [38,386].
In current smokers, sites with aberrant methylation were overpresented in the gene bodies and underrepresented in the promoter regions; sites in the intergenic regions were present as expected from the content of the 450K chip. In former smokers, sites with aberrant methylation were overpresented in the intergenic regions, with about twice as many sites as would be expected (46.15% vs. 24.61%), and underrepresented in the promoter regions; sites at the gene bodies were present as expected from the content of the 450K chip (Figure 27).
In summary, it may be hypothesized that the overrepresentation of significant sites in gene bodies and/or intergenic regions, which are normally methylated, and simultaneously underrepresentation of significant sites in promoter regions, which are normally unmethylated, may lead to the general hypomethylation in current and former smokers. On the other hand, it also may be hypothesized that hypomethylation in general is more sensitive to measurements and therefore easier detectable than hypermethylation, but no studies could be found which indicate this.

4.2.4. Possible regulatory role of cg05575921 (*AHRR*) for gene expression

Methylation-specific DNA-protein binding patterns for the most significant site cg05575921 (*AHRR*) were detected using both Raji (human B-lymphoblastoid cell line; Figure 31 a) and THP1 (human monozytic cell line; Figure 31 b) nuclear extracts in two independent EMSA experiments for each cell line.

Corroborating the observations made by the EMSA experiments, Monick et al. recently showed that an increase in methylation at cg05575921 was associated with a decrease in lymphoblast *AHRR* gene expression ($p < 0.03$, $N = 108$) [254]. And, as mentioned earlier, the study by Bosse et al. found *AHRR* to be the most significant probe set between never and current smokers with a fold change of 6.1 [256]. Furthermore, a recent study of Shenker et al. demonstrated *AHRR* expression to be 5.7 fold increased in human lung samples from current vs. never smokers, which inversely correlated with methylation levels [255]. This underscores the EMSA findings and suggests that this CpG site may have a regulatory role on gene expression, possibly mediated by differential binding of methylation-specific transcription factors.

Excitingly, an examination of this CpG site with the Transcription Factor ChIP-seq track from ENCODE (UCSC Genome Browser) revealed that this region is a possible binding site for three transcription factors, proteins responsible for modulating gene transcription. One of them is BAF155 which targets *SMARCC1* (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1), a ubiquitous component of the SWI/SNF chromatin-remodeling complex, in a cervical carcinoma cell line (HeLa-S3). The other two are Pol2 and Pol2-4H8, which both target POLR2A (the largest subunit of RNA polymerase II, the polymerase which is responsible for synthesizing messenger RNA in eukaryotes), with Pol2 in an acute promyelocytic leukemia cell line (NB4) and in peripheral blood-derived erythroblasts (PBDE), and Pol2-4H8 in GM12891 cells (B-lymphocytes treated with transformed Epstein-Barr Virus).

4.3. Smoking-associated CpG sites can be validated

With the establishment of the EpiTYPER assay that allowed a rapid and quantitative measurement of allele-specific DNA methylation, the third aim of this thesis could be achieved. The smoking-associated differential DNA methylation for the five selected loci could be validated by this technique, showing a high Pearson correlation coefficient, therefore backing up the reliability of the 450K BeadChip array (Table 17, Figure 35). The loci chosen for technical validation (cg05575921-*AHRR*, cg21566642-*ALPP/ALPP2*, cg03636183-*F2RL3*, cg06126421-not annotated, cg19572487-*RARA*) have already been discussed in chapter 4.2.1.

The bisulfite conversion efficiency calculation of Thompson et al. assured that the bisulfite conversion for all samples was completed, so false-positive methylation signals could be ruled out (Figure 34). The disadvantage of this approach is that it can only be implemented

after the methylation analysis have been carried out, as an output file of the methylation data for each amplicon has to be generated in order to run the pipeline.

Also, for the EpiTYPER method a major drawback is evident: during the primer design with the EpiDesigner all CpG sites are weighted equally, whether or not there are specific CpGs that the user is particularly interested in interrogating. This complicated the validation of the results of the specific CpGs of the 450K BeadChip. Two of the five CpGs were not directly covered by the EpiTYPER assay, however, both target and their flanking CpGs are located in or in the shore of a CGI and distribution of DNA methylation within a definite genomic element as a CGI is known to be relatively homogeneous. This uniformity leads to similar levels of DNA methylation and therefore allows the representative analysis of CpGs neighboring the actual target CpG [387].

Furthermore, the EpiTYPER assay fundamentally assumes an accurate amplicon sequence. Nevertheless, SNPs within the target sequence of an amplicon could have an effect on the DNA methylation analysis, whether in the form of a fragment mass shift or a different fragmentation pattern, which can lead to misinterpretation of the results [388]. Even though SNP detection analyses were carried out by the pipeline of Thompson et al. (Figure 36), it is of note that these analyses cannot directly confirm or deny the presence or absence of a SNP. However, if samples show a putative high-confidence SNP that maps to a fragment containing one or more CpGs, methylation data from that site should be interpreted with caution [202]. Only the amplicon within the *ALPP/ALPPL2* region showed several putative SNPs. This amplicon included cg21566642, which could not directly be covered by the EpiTYPER assay.

Another possible method to validate the results of the 450K BeadArray would have been bisulfite sequencing. This method permits interrogation of each individual CpG site in the regions interrogated, whereas the EpiTYPER assay only examines the average methylation data per cleavage fragment. In best case, a fragment contains an isolated site and has a unique molecular weight. However, a fragment may also contain multiple CpGs or may have a molecular weight that overlaps with another CpG-containing fragment and then the constituent CpGs have to be measured in aggregates. On the other hand, bisulfite sequencing is time consuming and costly for large sets of samples, the EpiTYPER assay is therefore favorable in a high-throughput experiment. Furthermore, the EpiTYPER assay is highly quantitative and bisulfite sequencing analyses may be easily biased by the limited number of samples analyzed [388].

Another frequently used high-throughput method for quantitative DNA methylation analyses is pyrosequencing, which is a real-time DNA sequencing by synthesis method that can be used for direct quantitative methylation analyses after bisulfite treatment. Compared with the

EpiTYPER assay, both have the advantages of accuracy and high-throughput nature and the sensitivity is similar (5%). The main limitation of pyrosequencing is that only 25-30 base pairs can be sequenced in each reaction, compared to a maximum of 800 bp via the EpiTYPER method, limiting the number of CpGs that can be analyzed in one sequencing reaction [388,389].

4.4. Strengths and Limitations

The major strengths of this study are the relatively large sample size of the population-based discovery and the selected replication panel, as well as the information about former smoking. For quality assurance of the 450K and EpiTYPER methylation data, various quality control steps were applied and all statistical models were adjusted for a large number of potential confounders.

There are also limitations to this study: despite thorough assessment of the smoking status by several questions, no cotinine measurements are available in the KORA study to directly assess smoking burden. Passive smoking, which might also have an effect on DNA methylation, was not taken into account. The design of the present study is cross-sectional in nature; therefore only a suggestion that quitting tobacco smoking presumably allows reformation of the DNA methylation state in former smokers can be made. Longitudinal studies are needed to confirm these results. Furthermore, the present study explores whole blood, which consists of a complex composition of cells that show individual methylation patterns [183]. However, Shenker et al. analyzed the relationship between different blood cell fractions and whole blood DNA from the same individual by the 450K BeadChip. These analyses could show no evidence that any of the blood cell types have significantly different methylation levels that would confound an association with smoking. In addition, the methylation levels of sites in the *AHRR* gene between lung tissues and PBMCs were compared and found to be identical [255]. Besides, a similar correlation has also been reported in lymphoblasts and pulmonary macrophages by Monick et al. [254]. Additionally, several of the smoking-associated genes detected within the study of this thesis (*AHRR*, *GFI1*, *MYO1G*, *CNTNAP2*) were also reported to be differentially methylated in cord blood samples due to maternal smoking. This study by Joubert et al. directly addressed the potential impact of differential cell counts by additionally measuring polymorphonuclear and mononuclear cells with the 450K BeadChip. The differences in methylation by cell type were much smaller than the differences in methylation by smoking observed in whole blood, indicating that their findings are not explained by cell type confounding [253]. These studies

show that even though DNA methylation is tissue specific and the sensitivity depends on the tissue type, changes in DNA methylation may at least in some cases be reflected in whole blood, and thus strengthen the findings of this thesis. This certainly has high clinical relevance, as blood is an easily accessible biomaterial and therefore an attractive tissue for the identification and subsequent use of biomarkers. Nevertheless, a recent published study showed variation in the methylation profiles of whole blood, mononuclear cells, granulocytes, and cells from seven selected purified lineages [390], thus a way has to be found to solve this problem as no information is available on the number of the single cell types in the KORA samples. One way may be a cell mixture reconstruction by a method published by Houseman et.al [391].

The pipeline of Touleimat & Tost does not provide a correction approach for possible batch effects. However, that the results of this study are due to batch effects is very unlikely because of various reasons. First, current and never smokers were equally distributed on the plates and chips. Second, a recent study that analyzed methylation levels of current and never smokers with the 27K BeadChip mentioned that even significant batch effects had very limited confounding potential [372]. Third, as shown in chapter 3.1.1.3., the six positive control samples that were included in each Illumina run, placed on different chips and plates, showed highest reproducibility. Last but not least, many results that have been found in other smoking-related studies could be replicated.

4.5. Conclusions

In summary, evidence of significant differences in the degree of site-specific methylation in each of the 22 autosomes as a function of tobacco smoking was observed, identifying 187 differentially methylated CpG sites by array-based DNA methylation analysis. Several novel loci which play roles mostly in the development and function of the cellular, hematological, immune, cardiovascular, tumorigenic, and reproduction systems, and may be involved in mediating the effects of smoking in the development of disease, have been revealed. Depending on cessation time and pack-years, methylation levels in former smokers were found to be close to levels seen in never smokers. Methylation-specific protein binding patterns observed in EMSA experiments suggest a regulatory role of CpG site cg05575921 for gene expression. The results of this study confirm the broad effects of tobacco smoking on the human organism. Revealing the underlying molecular mechanisms that alter the epigenome due to environmental triggers will be an important aspect of future studies.

4.6. Outlook

The work of this thesis has opened several research topics for possible follow-up projects and therefore provided the basis for interesting new studies: A few examples of potential future projects are given within the next paragraphs.

Within this thesis, methylation-specific DNA-protein binding patterns have been exemplarily demonstrated by EMSA for the most significant CpG site of the *AHRR* gene (cg05575921), suggesting a regulatory role of this site on gene expression. This experiment underlines the potential biological relevance of DNA methylation differences caused by tobacco smoking. Carrying out EMSAs for further smoking-associated sites, as well as identifying the corresponding methylation-specific transcription factors, starting with the three transcription factors (BAF155, Pol2 and Pol2-4H8) that possibly bind to the region around cg05575921 according to Transcription Factor ChIP-seq track from ENCODE (UCSC Genome Browser), will be the subject of future studies.

Further, according to the UCSC genome browser database, three large CGIs are interspersed throughout the *AHRR* gene and at least 11 *AHRR* transcripts exist that may have unique competitive properties with respect to AHR. More research is needed to better understand the relationship between methylation and gene expression of the specific *AHRR* transcripts and their effect on AHR function.

The genome-wide DNA methylation levels in former smokers were shown to be close to the ones seen in never smokers, even though a few CpG sites were still significant. Together with other published work concerning gene expression in former smokers, these results indicate a long-term gene expression consequence of prior smoking history, at least for some genes, including *AHRR*, which may be locked in by DNA methylation changes (see chapter 4.2.2.1.). It has been shown that the effect of smoking on malignant, cardiovascular, and respiratory diseases can persist for extended periods after smoking cessation and may involve epigenetic reprogramming [373]. Identifying genes that do not normalize following smoking cessation in subjects who develop a certain disease compared with those who do not develop this disease might be an important step to uncover disease candidate genes. Therefore, another extremely interesting research project would be to identify molecular pathways that contribute to the delay between exposure and disease, especially as it may offer opportunities for targeted diagnostic and therapeutic interventions.

How hypo- and hypermethylation in current smokers and later reestablishment of the methylation patterns in former smokers are accomplished also remains to be elucidated. Even though a few studies showed that one possible mechanism of aberrant methylation and demethylation involves DNMTs, differential expression in DNA methyltransferases by current

smoking status is missing in many studies (chapter 4.2.2.). Therefore, in vitro and in vivo studies are needed to explore alternative mechanisms that may contribute to the establishment of locus-specific differential DNA methylation.

As mentioned in the limitations section, only a suggestion that quitting tobacco smoking presumably allows reformation of the DNA methylation state in former smokers can be made, as the study of this thesis is cross-sectional in nature. Longitudinal studies would allow assessing the real effect of quitting tobacco smoking. The great advantage of the KORA cohort is that a baseline survey (KORA S4) as well as a follow-up study (KORA F4) does exist. Presently, the samples of S4 are prepared to be analyzed with the 450K BeadChip to enable exciting longitudinal studies in the future that may also allow to investigate changes in DNA methylation patterns in relation to the development of disease.

A family of enzymes, called ten eleven translocation (TET) proteins, has been shown to be able to oxidize 5-methylcytosines (5mC) to 5-hydroxymethylcytosine (5hmC). The function of 5hmC and its effects on gene expression are not yet exactly known, but 5hmC has been found in multiple mammalian tissues and it is speculated that it might play a role in epigenetic regulation of mammalian development and differentiation. Furthermore, it has been shown that oxidation of 5mC may contribute to active DNA demethylation. TET enzymes are mutated in several types of cancer and might therefore play a role in the global hypomethylation during cancer development and progression [392]. As in all bisulfite-based analyses techniques, it is not possible to differentiate between cytosine methylation and hydroxymethylation within the 450K array and the EpiTYPER approach. Investigating the role of TET proteins and 5hmC in smoking-associated gene regulation and disease might therefore be another interesting project for the future.

Epigenetic regulation is influenced by genetic differences between individuals, so dizygotic twins, for example, show more epigenetic differences than monozygotic twins [75]. And various studies have shown that the level of methylation of CpG sites is affected by SNPs nearby [76,77]. However, it is still unclear to which extent the DNA sequence determines the levels of epigenetic modification at specific loci. More research is needed to understand and account for the genetic predisposition to the effect of smoking.

Last but not least, as DNA methylation closely interacts with histone modifications, miRNA mechanism, and chromatin-remodeling complexes to form the genomic chromatin landscape, investigating the role of the smoking-associated DNA methylation differences within all of these epigenetic regulatory mechanisms would be an interesting subject for future research. In this respect, the question of whether aberrant DNA methylation is cause or consequence of deviant regulation within this complex epigenetic regulatory machinery and/or disease remains to be elucidated.

5. Literature

1. Baccarelli A, Bollati V (2009) Epigenetics and environmental chemicals. Curr Opin Pediatr 21: 243-251.
2. Waddington CH (2012) The epigenotype. 1942. Int J Epidemiol 41: 10-13.
3. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6-21.
4. Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157-161.
5. Guerrero-Bosagna C, Skinner MK (2012) Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. Mol Cell Endocrinol 354: 3-8.
6. Nilsson E, Larsen G, Manikkam M, Guerrero-Bosagna C, Savenkova MI, et al. (2012) Environmentally induced epigenetic transgenerational inheritance of ovarian disease. PLoS One 7: e36129.
7. Waterland RA, Travisano M, Tahiliani KG, Rached MT, Mirza S (2008) Methyl donor supplementation prevents transgenerational amplification of obesity. Int J Obes (Lond) 32: 1373-1379.
8. Bruner-Tran KL, Osteen KG (2011) Developmental exposure to TCDD reduces fertility and negatively affects pregnancy outcomes across multiple generations. Reprod Toxicol 31: 344-350.
9. Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293: 1089-1093.
10. Vaissiere T, Sawan C, Herceg Z (2008) Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutat Res 659: 40-48.
11. Barnes PJ, Adcock IM, Ito K (2005) Histone acetylation and deacetylation: importance in inflammatory lung diseases. Eur Respir J 25: 552-563.
12. Yeung ML, Bennasser Y, Le SY, Jeang KT (2005) siRNA, miRNA and HIV: promises and challenges. Cell Res 15: 935-946.
13. Morales Prieto DM, Markert UR (2011) MicroRNAs in pregnancy. J Reprod Immunol 88: 106-111.
14. Ulrey CL, Liu L, Andrews LG, Tollefsbol TO (2005) The impact of metabolism on DNA methylation. Hum Mol Genet 14 Spec No 1: R139-147.
15. Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447: 425-432.
16. Probst AV, Dunleavy E, Almouzni G (2009) Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol 10: 192-206.
17. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102: 10604-10609.
18. Miller OJ, Schnedl W, Allen J, Erlanger BF (1974) 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature 251: 636-637.
19. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693-705.
20. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41-45.
21. Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447: 407-412.
22. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annu Rev Biochem 78: 273-304.
23. Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10: 295-304.
24. Tomari Y, Zamore PD (2005) Perspective: machines for RNAi. Genes Dev 19: 517-529.
25. Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. Nature 431: 343-349.
26. Barros SP, Offenbacher S (2009) Epigenetics: connecting environment and genotype to phenotype and disease. J Dent Res 88: 400-408.

27. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465-476.
28. Shao C, Lacey M, Dubeau L, Ehrlich M (2009) Hemimethylation footprints of DNA demethylation in cancer. Epigenetics 4: 165-175.
29. Tost J (2010) DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. Mol Biotechnol 44: 71-81.
30. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315-322.
31. Gupta R, Nagarajan A, Wajapeyee N (2010) Advances in genome-wide DNA methylation analysis. Biotechniques 49: iii-xi.
32. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, et al. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A 97: 5237-5242.
33. Illingworth RS, Bird AP (2009) CpG islands--'a rough guide'. FEBS Lett 583: 1713-1720.
34. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196: 261-282.
35. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99: 3740-3745.
36. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13: 484-492.
37. Handy DE, Castro R, Loscalzo J (2011) Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation 123: 2145-2156.
38. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28: 1057-1068.
39. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99: 247-257.
40. Hermann A, Gowher H, Jeltsch A (2004) Biochemistry and biology of mammalian DNA methyltransferases. Cell Mol Life Sci 61: 2571-2587.
41. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204-220.
42. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2: 21-32.
43. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, et al. (2007) Computational and experimental identification of novel human imprinted genes. Genome Res 17: 1723-1730.
44. Holmes R, Soloway PD (2006) Regulation of imprinted DNA methylation. Cytogenet Genome Res 113: 122-129.
45. Chow J, Heard E (2009) X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol 21: 359-366.
46. Wolf SF, Migeon BR (1982) Studies of X chromosome DNA methylation in normal human cells. Nature 295: 667-671.
47. Boyko A, Kathiria P, Zemp FJ, Yao Y, Pogribny I, et al. (2007) Transgenerational changes in the genome stability and methylation in pathogen-infected plants: (virus-induced plant genome instability). Nucleic Acids Res 35: 1714-1725.
48. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R (1998) DNA hypomethylation leads to elevated mutation rates. Nature 395: 89-93.
49. Jones PA, Takai D (2001) The role of DNA methylation in mammalian epigenetics. Science 293: 1068-1070.
50. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, et al. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39: 457-466.

51. Nomura J, Hisatsune A, Miyata T, Isohama Y (2007) The role of CpG methylation in cell type-specific expression of the aquaporin-5 gene. Biochem Biophys Res Commun 353: 1017-1022.
52. Aoyama T, Okamoto T, Nagayama S, Nishijo K, Ishibe T, et al. (2004) Methylation in the core-promoter region of the chondromodulin-I gene determines the cell-specific expression by regulating the binding of transcriptional activator Sp3. J Biol Chem 279: 28789-28797.
53. Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18: 6538-6547.
54. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386-389.
55. Ushijima T (2005) Detection and interpretation of altered methylation patterns in cancer cells. Nat Rev Cancer 5: 223-231.
56. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, et al. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479: 74-79.
57. Wong CC, Caspi A, Williams B, Craig IW, Houts R, et al. (2010) A longitudinal study of epigenetic variation in twins. Epigenetics 5: 516-526.
58. Rodriguez-Rodero S, Fernandez-Morera JL, Fernandez AF, Menendez-Torre E, Fraga MF (2010) Epigenetic regulation of aging. Discov Med 10: 225-233.
59. Esteller M, Fraga MF, Paz MF, Campo E, Colomer D, et al. (2002) Cancer epigenetics and methylation. Science 297: 1807-1808; discussion 1807-1808.
60. Baylin SB, Herman JG (2000) DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet 16: 168-174.
61. Ehrlich M (2009) DNA hypomethylation in cancer cells. Epigenomics 1: 239-259.
62. Feil R, Fraga MF (2011) Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet 13: 97-109.
63. Van den Veyver IB (2002) Genetic effects of methylation diets. Annu Rev Nutr 22: 255-282.
64. Alegria-Torres JA, Baccarelli A, Bollati V (2011) Epigenetics and lifestyle. Epigenomics 3: 267-277.
65. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, et al. (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci U S A 105: 17046-17049.
66. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, et al. (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18: 4046-4053.
67. Waterland RA, Kellermayer R, Laritsky E, Rayco-Solon P, Harris RA, et al. (2010) Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. PLoS Genet 6: e1001252.
68. Hou L, Zhang X, Wang D, Baccarelli A (2012) Environmental chemical exposures and human epigenetics. Int J Epidemiol 41: 79-105.
69. Christensen BC, Marsit CJ (2011) Epigenomics in environmental health. Front Genet 2: 84.
70. Gronniger E, Weber B, Heil O, Peters N, Stab F, et al. (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet 6: e1000971.
71. Preventio CfDCa (2008) Annual Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses—United States, 2000–2004. Morbidity and Mortality Weekly Report 57(45): 1226–1228.
72. Vogel U, Christensen J, Wallin H, Friis S, Nexo BA, et al. (2008) Polymorphisms in genes involved in the inflammatory response and interaction with NSAID use or smoking in relation to lung cancer risk in a prospective study. Mutat Res 639: 89-100.
73. Stein RA (2012) Epigenetics and environmental exposures. J Epidemiol Community Health 66: 8-13.
74. Groom A, Elliott HR, Embleton ND, Relton CL (2011) Epigenetics and child health: basic principles. Arch Dis Child 96: 863-869.

75. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, et al. (2009) DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet 41: 240-245.
76. Gertz J, Varley KE, Reddy TE, Bowling KM, Pauli F, et al. (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. PLoS Genet 7: e1002228.
77. Hellman A, Chess A (2010) Extensive sequence-influenced DNA methylation polymorphism in the human genome. Epigenetics Chromatin 3: 11.
78. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3: e442.
79. Thun MJ, DeLancey JO, Center MM, Jemal A, Ward EM (2010) The global burden of cancer: priorities for prevention. Carcinogenesis 31: 100-110.
80. Ezzati M, Lopez AD (2003) Estimates of global mortality attributable to smoking in 2000. Lancet 362: 847-852.
81. Mackay J EM, Ross H. (2012) The Tobacco Atlas, 4th edn. American Cancer Society, Atlanta, GA, USA.
82. Oberg M, Jaakkola MS, Woodward A, Peruga A, Pruss-Ustun A (2011) Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. Lancet 377: 139-146.
83. Commission E (2004) Tobacco or health in the European Union - Past, present and future Luxembourg: European Communities.
84. Services USDoHaH (2010) How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease. A Report of the Surgeon General.
85. Services USDoHaH (2006) The Health Consequences of Involuntary Exposure to Tobacco Smoke. A Report of the Surgeon General.
86. Neubauer S, Welte R, Beiche A, Koenig HH, Buesch K, et al. (2006) Mortality, morbidity and costs attributable to smoking in Germany: update and a 10-year comparison. Tob Control 15: 464-471.
87. Hoffmann D, Hoffmann I (1997) The changing cigarette, 1950-1995. J Toxicol Environ Health 50: 307-364.
88. Rodgman A PT (2009) The Chemical Components of Tobacco and Tobacco Smoke. Boca Raton (FL): CRC Press, Taylor & Francis Group.
89. Borgerding M, Klus H (2005) Analysis of complex mixtures--cigarette smoke. Exp Toxicol Pathol 57 Suppl 1: 43-73.
90. Benowitz NL (1996) Pharmacology of nicotine: addiction and therapeutics. Annu Rev Pharmacol Toxicol 36: 597-613.
91. Abuse NIoD (2009) Research Report Series: Tobacco Addiction. National Institutes of Health.
92. Fiore MC JC, Baker TB, et al. (2008) Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: U.S. Department of Health and Human Services: Public Health Service.
93. Hendricks PS, Ditre JW, Drobes DJ, Brandon TH (2006) The early time course of smoking withdrawal effects. Psychopharmacology (Berl) 187: 385-396.
94. Cappelleri JC, Bushmakin AG, Baker CL, Merikle E, Olufade AO, et al. (2005) Revealing the multidimensional framework of the Minnesota nicotine withdrawal scale. Curr Med Res Opin 21: 749-760.
95. Anderson JE, Jorenby DE, Scott WJ, Fiore MC (2002) Treating tobacco use and dependence: an evidence-based clinical practice guideline for tobacco cessation. Chest 121: 932-941.
96. Henningfield JE, Rose CA, Zeller M (2006) Tobacco industry litigation position on addiction: continued dependence on past views. Tob Control 15 Suppl 4: iv27-36.
97. Prevention CfDCa (2011) Quitting Smoking Among Adults—United States, 2001–2010. Morbidity and Mortality Weekly Report
98. Vineis P, Alavanja M, Buffler P, Fontham E, Franceschi S, et al. (2004) Tobacco and cancer: recent epidemiological evidence. J Natl Cancer Inst 96: 99-106.

99. Kawachi I, Colditz GA, Stampfer MJ, Willett WC, Manson JE, et al. (1993) Smoking cessation and decreased risk of stroke in women. JAMA 269: 232-236.
100. Conen D, Everett BM, Kurth T, Creager MA, Buring JE, et al. (2011) Smoking, smoking cessation, [corrected] and risk for symptomatic peripheral artery disease in women: a cohort study. Ann Intern Med 154: 719-726.
101. Willemse BW, ten Hacken NH, Rutgers B, Lesman-Leegte IG, Postma DS, et al. (2005) Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. Eur Respir J 26: 835-845.
102. Bouloukaki I, Tsiligianni IG, Tsoumakidou M, Mitrouska I, Prokopakis EP, et al. (2011) Sputum and nasal lavage lung-specific biomarkers before and after smoking cessation. BMC Pulm Med 11: 35.
103. Braber S, Henricks PA, Nijkamp FP, Kraneveld AD, Folkerts G (2010) Inflammatory changes in the airways of mice caused by cigarette smoke exposure are only partially reversed after smoking cessation. Respir Res 11: 99.
104. Gellert C, Schottker B, Brenner H (2012) Smoking and all-cause mortality in older people: systematic review and meta-analysis. Arch Intern Med 172: 837-844.
105. Tessema M, Yu YY, Stidley CA, Machida EO, Schuebel KE, et al. (2009) Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers. Carcinogenesis 30: 1132-1138.
106. Risch A, Plass C (2008) Lung cancer epigenetics and genetics. Int J Cancer 123: 1-7.
107. Kanai Y (2008) Alterations of DNA methylation and clinicopathological diversity of human cancers. Pathol Int 58: 544-558.
108. Russo AL, Thiagalingam A, Pan H, Califano J, Cheng KH, et al. (2005) Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. Clin Cancer Res 11: 2466-2470.
109. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, et al. (2002) Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res 62: 2370-2377.
110. Enokida H, Shiina H, Urakami S, Terashima M, Ogishima T, et al. (2006) Smoking influences aberrant CpG hypermethylation of multiple genes in human prostate carcinoma. Cancer 106: 79-86.
111. Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, et al. (2007) Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev 16: 108-114.
112. Breton CV, Byun HM, Wenten M, Pan F, Yang A, et al. (2009) Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. Am J Respir Crit Care Med 180: 462-467.
113. Launay JM, Del Pino M, Chironi G, Callebert J, Peoc'h K, et al. (2009) Smoking induces long-lasting effects through a monoamine-oxidase epigenetic regulation. PLoS One 4: e7959.
114. Ries LAG EM, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Feuer EJ, Edwards BK, editors (2004) SEER Cancer Statistics Review, 1975–2001. Bethesda (MD): National Cancer Institute.
115. Halpern MT, Gillespie BW, Warner KE (1993) Patterns of absolute risk of lung cancer mortality in former smokers. J Natl Cancer Inst 85: 457-464.
116. Esteller M (2007) Epigenetic gene silencing in cancer: the DNA hypermethylome. Hum Mol Genet 16 Spec No 1: R50-59.
117. Zochbauer-Muller S, Minna JD, Gazdar AF (2002) Aberrant DNA methylation in lung cancer: biological and clinical implications. Oncologist 7: 451-457.
118. Belinsky SA (2005) Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. Carcinogenesis 26: 1481-1487.
119. Anttila S, Hakkola J, Tuominen P, Elovaara E, Husgafvel-Pursiainen K, et al. (2003) Methylation of cytochrome P4501A1 promoter in the lung is associated with tobacco smoking. Cancer Res 63: 8623-8628.

120. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, et al. (1998) Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A 95: 11891-11896.
121. Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, et al. (1995) Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res 55: 4525-4530.
122. Jarmalaite S, Kannio A, Anttila S, Lazutka JR, Husgafvel-Pursiainen K (2003) Aberrant p16 promoter methylation in smokers and former smokers with nonsmall cell lung cancer. Int J Cancer 106: 913-918.
123. Lea JS, Coleman R, Kurien A, Schorge JO, Miller DS, et al. (2004) Aberrant p16 methylation is a biomarker for tobacco exposure in cervical squamous cell carcinogenesis. Am J Obstet Gynecol 190: 674-679.
124. Pulling LC, Divine KK, Klinge DM, Gilliland FD, Kang T, et al. (2003) Promoter hypermethylation of the O6-methylguanine-DNA methyltransferase gene: more common in lung adenocarcinomas from never-smokers than smokers and associated with tumor progression. Cancer Res 63: 4842-4848.
125. Schwartz AG, Prysak GM, Bock CH, Cote ML (2007) The molecular epidemiology of lung cancer. Carcinogenesis 28: 507-518.
126. Ai L, Stephenson KK, Ling W, Zuo C, Mukunyadzi P, et al. (2003) The p16 (CDKN2a/INK4a) tumor-suppressor gene in head and neck squamous cell carcinoma: a promoter methylation and protein expression study in 100 cases. Mod Pathol 16: 944-950.
127. Le Calvez F, Mukeria A, Hunt JD, Kelm O, Hung RJ, et al. (2005) TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: distinct patterns in never, former, and current smokers. Cancer Res 65: 5076-5083.
128. Rosas SL, Koch W, da Costa Carvalho MG, Wu L, Califano J, et al. (2001) Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients. Cancer Res 61: 939-942.
129. Fujiwara K, Fujimoto N, Tabata M, Nishii K, Matsuo K, et al. (2005) Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. Clin Cancer Res 11: 1219-1225.
130. Marsit CJ, Karagas MR, Danaee H, Liu M, Andrew A, et al. (2006) Carcinogen exposure and gene promoter hypermethylation in bladder cancer. Carcinogenesis 27: 112-116.
131. Dammann R, Schagdarsurengin U, Strunnikova M, Rastetter M, Seidel C, et al. (2003) Epigenetic inactivation of the Ras-association domain family 1 (RASSF1A) gene and its function in human carcinogenesis. Histol Histopathol 18: 665-677.
132. Burbee DG, Forgacs E, Zochbauer-Muller S, Shivakumar L, Fong K, et al. (2001) Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. J Natl Cancer Inst 93: 691-699.
133. Hasegawa M, Nelson HH, Peters E, Ringstrom E, Posner M, et al. (2002) Patterns of gene promoter methylation in squamous cell cancer of the head and neck. Oncogene 21: 4231-4236.
134. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, et al. (2003) Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. Int J Cancer 107: 612-616.
135. Kim DH, Kim JS, Ji YI, Shim YM, Kim H, et al. (2003) Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. Cancer Res 63: 3743-3746.
136. Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, et al. (1998) Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res 58: 4515-4518.

137. Tchou JC, Lin X, Freije D, Isaacs WB, Brooks JD, et al. (2000) GSTP1 CpG island DNA hypermethylation in hepatocellular carcinomas. Int J Oncol 16: 663-676.
138. Tang X, Khuri FR, Lee JJ, Kemp BL, Liu D, et al. (2000) Hypermethylation of the death-associated protein (DAP) kinase promoter and aggressiveness in stage I non-small-cell lung cancer. J Natl Cancer Inst 92: 1511-1516.
139. Anisowicz A, Huang H, Braunschweiger KI, Liu Z, Giese H, et al. (2008) A high-throughput and sensitive method to measure global DNA methylation: application in lung cancer. BMC Cancer 8: 222.
140. Pulling LC, Vuillemenot BR, Hutt JA, Devereux TR, Belinsky SA (2004) Aberrant promoter hypermethylation of the death-associated protein kinase gene is early and frequent in murine lung tumors induced by cigarette smoke and tobacco carcinogens. Cancer Res 64: 3844-3848.
141. Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, et al. (1995) E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. Cancer Res 55: 5195-5199.
142. Vuillemenot BR, Pulling LC, Palmisano WA, Hutt JA, Belinsky SA (2004) Carcinogen exposure differentially modulates RAR-beta promoter hypermethylation, an early and frequent event in mouse lung carcinogenesis. Carcinogenesis 25: 623-629.
143. Virmani AK, Rathi A, Zochbauer-Muller S, Sacchi N, Fukuyama Y, et al. (2000) Promoter methylation and silencing of the retinoic acid receptor-beta gene in lung carcinomas. J Natl Cancer Inst 92: 1303-1307.
144. Wong TS, Man MW, Lam AK, Wei WI, Kwong YL, et al. (2003) The study of p16 and p15 gene methylation in head and neck squamous cell carcinoma and their quantitative evaluation in plasma by real-time PCR. Eur J Cancer 39: 1881-1887.
145. Chang HW, Ling GS, Wei WI, Yuen AP (2004) Smoking and drinking can induce p15 methylation in the upper aerodigestive tract of healthy individuals and patients with head and neck squamous cell carcinoma. Cancer 101: 125-132.
146. Ambrose JA, Barua RS (2004) The pathophysiology of cigarette smoking and cardiovascular disease: an update. J Am Coll Cardiol 43: 1731-1737.
147. Bakhru A, Erlinger TP (2005) Smoking cessation and cardiovascular disease risk factors: results from the Third National Health and Nutrition Examination Survey. PLoS Med 2: e160.
148. Barnoya J, Glantz SA (2005) Cardiovascular effects of secondhand smoke: nearly as large as smoking. Circulation 111: 2684-2698.
149. Burns DM (2003) Epidemiology of smoking-induced cardiovascular disease. Prog Cardiovasc Dis 46: 11-29.
150. Critchley JA, Capewell S (2003) Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. JAMA 290: 86-97.
151. Sharma P, Kumar J, Garg G, Kumar A, Patowary A, et al. (2008) Detection of altered global DNA methylation in coronary artery disease patients. DNA Cell Biol 27: 357-365.
152. Castro R, Rivera I, Struys EA, Jansen EE, Ravasco P, et al. (2003) Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. Clin Chem 49: 1292-1296.
153. Hiltunen MO, Turunen MP, Hakkinen TP, Rutanen J, Hedman M, et al. (2002) DNA hypomethylation and methyltransferase expression in atherosclerotic lesions. Vasc Med 7: 5-11.
154. Lind L (2003) Circulating markers of inflammation and atherosclerosis. Atherosclerosis 169: 203-214.
155. Baccarelli A, Tarantini L, Wright RO, Bollati V, Litonjua AA, et al. (2010) Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. Epigenetics 5.

156. Post WS, Goldschmidt-Clermont PJ, Wilhide CC, Heldman AW, Sussman MS, et al. (1999) Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. Cardiovasc Res 43: 985-991.
157. Turunen MP, Aavik E, Yla-Herttuala S (2009) Epigenetics and atherosclerosis. Biochim Biophys Acta 1790: 886-891.
158. Dong C, Yoon W, Goldschmidt-Clermont PJ (2002) DNA methylation and atherosclerosis. J Nutr 132: 2406S-2409S.
159. Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, et al. (2004) DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. J Biol Chem 279: 29147-29154.
160. Pietinalho A, Pelkonen A, Rytila P (2009) Linkage between smoking and asthma. Allergy 64: 1722-1727.
161. Prescott SL, Clifton V (2009) Asthma and pregnancy: emerging evidence of epigenetic interactions in utero. Curr Opin Allergy Clin Immunol 9: 417-426.
162. Xepapadaki P, Manios Y, Liarigkovinos T, Grammatikaki E, Douladiris N, et al. (2009) Association of passive exposure of pregnant women to environmental tobacco smoke with asthma symptoms in children. Pediatr Allergy Immunol 20: 423-429.
163. Li YF, Langholz B, Salam MT, Gilliland FD (2005) Maternal and grandmaternal smoking patterns are associated with early childhood asthma. Chest 127: 1232-1241.
164. Radzeviciene L, Ostrauskas R (2009) Smoking habits and the risk of type 2 diabetes: a case-control study. Diabetes Metab 35: 192-197.
165. Kowall B, Rathmann W, Strassburger K, Heier M, Holle R, et al. (2010) Association of passive and active smoking with incident type 2 diabetes mellitus in the elderly population: the KORA S4/F4 cohort study. Eur J Epidemiol 25: 393-402.
166. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J (2007) Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. JAMA 298: 2654-2664.
167. Will JC, Galuska DA, Ford ES, Mokdad A, Calle EE (2001) Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. Int J Epidemiol 30: 540-546.
168. Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, et al. (2010) A genome-wide perspective of genetic variation in human metabolism. Nat Genet 42: 137-141.
169. Holle R, Happich M, Lowel H, Wichmann HE (2005) KORA--a research platform for population based health research. Gesundheitswesen 67 Suppl 1: S19-25.
170. Wichmann HE, Gieger C, Illig T (2005) KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67 Suppl 1: S26-30.
171. 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.
172. Steffens M, Lamina C, Illig T, Bettecken T, Vogler R, et al. (2006) SNP-based analysis of genetic substructure in the German population. Hum Hered 62: 20-29.
173. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-491.
174. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High density DNA methylation array with single CpG site resolution. Genomics 98: 288-295.
175. Touleimat N, Tost J (2012) Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics 4: 325-341.
176. Du P, Kibbe WA, Lin SM (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics 24: 1547-1548.

177. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, et al. (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11: 587.
178. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, et al. (2011) Epigenetic predictor of age. PLoS One 6: e14821.
179. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, et al. (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet 8: e1002629.
180. Burgess DJ (2013) Human epigenetics: Showing your age. Nat Rev Genet 14: 6.
181. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, et al. (2012) Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A 109: 10522-10527.
182. Liu J, Morgan M, Hutchison K, Calhoun VD (2010) A study of the influence of sex on genome wide methylation. PLoS One 5: e10028.
183. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38: 1378-1385.
184. Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, et al. (2011) Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics 6: 623-629.
185. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, et al. (2010) Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med 2: 49ra67.
186. Tao MH, Marian C, Nie J, Ambrosone C, Krishnan SS, et al. (2011) Body mass and DNA promoter methylation in breast tumors in the Western New York Exposures and Breast Cancer Study. Am J Clin Nutr 94: 831-838.
187. Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, et al. (2012) Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. Int J Epidemiol 41: 126-139.
188. Manzardo AM, Henkhaus RS, Butler MG (2012) Global DNA promoter methylation in frontal cortex of alcoholics and controls. Gene 498: 5-12.
189. Petitti DB, Kipp H (1986) The leukocyte count: associations with intensity of smoking and persistence of effect after quitting. Am J Epidemiol 123: 89-95.
190. Schwartz J, Weiss ST (1994) Cigarette smoking and peripheral blood leukocyte differentials. Ann Epidemiol 4: 236-242.
191. Parry H, Cohen S, Schlarb JE, Tyrrell DA, Fisher A, et al. (1997) Smoking, alcohol consumption, and leukocyte counts. Am J Clin Pathol 107: 64-67.
192. Corre F, Lellouch J, Schwartz D (1971) Smoking and leucocyte-counts. Results of an epidemiological survey. Lancet 2: 632-634.
193. Schwartz J, Weiss ST (1991) Host and environmental factors influencing the peripheral blood leukocyte count. Am J Epidemiol 134: 1402-1409.
194. Hughes DA, Haslam PL, Townsend PJ, Turner-Warwick M (1985) Numerical and functional alterations in circulatory lymphocytes in cigarette smokers. Clin Exp Immunol 61: 459-466.
195. Abel GA, Hays JT, Decker PA, Croghan GA, Kuter DJ, et al. (2005) Effects of biochemically confirmed smoking cessation on white blood cell count. Mayo Clin Proc 80: 1022-1028.
196. Van Tiel E, Peeters PH, Smit HA, Nagelkerke NJ, Van Loon AJ, et al. (2002) Quitting smoking may restore hematological characteristics within five years. Ann Epidemiol 12: 378-388.
197. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, et al. (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 102: 15785-15790.
198. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60: 2299-2301.
199. Kirpekar F, Nordhoff E, Larsen LK, Kristiansen K, Roepstorff P, et al. (1998) DNA sequence analysis by MALDI mass spectrometry. Nucleic Acids Res 26: 2554-2559.

200. Hillenkamp F, Karas M, Beavis RC, Chait BT (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem 63: 1193A-1203A.
201. Griffin TJ, Smith LM (2000) Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. Trends Biotechnol 18: 77-84.
202. Thompson RF, Suzuki M, Lau KW, Greally JM (2009) A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry. Bioinformatics 25: 2164-2170.
203. Njolstad I, Arnesen E, Lund-Larsen PG (1996) Smoking, serum lipids, blood pressure, and sex differences in myocardial infarction. A 12-year follow-up of the Finnmark Study. Circulation 93: 450-456.
204. Prescott E, Hippe M, Schnohr P, Hein HO, Vestbo J (1998) Smoking and risk of myocardial infarction in women and men: longitudinal population study. BMJ 316: 1043-1047.
205. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41: 178-186.
206. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, et al. (2012) IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics 28: 729-730.
207. Hansen KD AM (2012) Minfi: analyze Illumina's 450 k methylation arrays. R version 2.15.1 ed.
208. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, et al. (2011) Evaluation of the Infinium Methylation 450K technology. Epigenomics 3: 771-784.
209. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, et al. (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum Mol Genet 18: 4808-4817.
210. Maksimovic J, Gordon L, Oshlack A (2012) SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 13: R44.
211. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, et al. (2013) A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29: 189-196.
212. Pogribny IP, Beland FA (2009) DNA hypomethylation in the origin and pathogenesis of human diseases. Cell Mol Life Sci 66: 2249-2261.
213. Dunn BK (2003) Hypomethylation: one side of a larger picture. Ann N Y Acad Sci 983: 28-42.
214. Francis KT, Thompson RW, Krumdieck CL (1977) Reaction of tetrahydrofolic acid with cyanate from urea solutions: formation of an inactive folate derivative. Am J Clin Nutr 30: 2028-2032.
215. Erdemir EO, Bergstrom J (2007) Effect of smoking on folic acid and vitamin B12 after nonsurgical periodontal intervention. J Clin Periodontol 34: 1074-1081.
216. Lee DH, Jacobs DR, Jr., Porta M (2009) Hypothesis: a unifying mechanism for nutrition and chemicals as lifelong modulators of DNA hypomethylation. Environ Health Perspect 117: 1799-1802.
217. Piyathilake CJ, Hine RJ, Dasanayake AP, Richards EW, Freeberg LE, et al. (1992) Effect of smoking on folate levels in buccal mucosal cells. Int J Cancer 52: 566-569.
218. Jones DP, Brown LA, Sternberg P (1995) Variability in glutathione-dependent detoxication in vivo and its relevance to detoxication of chemical mixtures. Toxicology 105: 267-274.
219. Loenen WA (2006) S-adenosylmethionine: jack of all trades and master of everything? Biochem Soc Trans 34: 330-333.
220. Kato S, Bowman ED, Harrington AM, Blomeke B, Shields PG (1995) Human lung carcinogen-DNA adduct levels mediated by genetic polymorphisms in vivo. J Natl Cancer Inst 87: 902-907.
221. Hayes JD, Flanagan JU, Jowsey IR (2005) Glutathione transferases. Annu Rev Pharmacol Toxicol 45: 51-88.
222. Suter M, Ma J, Harris A, Patterson L, Brown KA, et al. (2011) Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. Epigenetics 6: 1284-1294.

223. Harju TH, Peltoniemi MJ, Rytila PH, Soini Y, Salmenkivi KM, et al. (2007) Glutathione S-transferase omega in the lung and sputum supernatants of COPD patients. Respir Res 8: 48.
224. Allen M, Zou F, Chai HS, Younkin CS, Miles R, et al. (2012) Glutathione S-transferase omega genes in Alzheimer and Parkinson disease risk, age-at-diagnosis and brain gene expression: an association study with mechanistic implications. Mol Neurodegener 7: 13.
225. Damiani LA, Yingling CM, Leng S, Romo PE, Nakamura J, et al. (2008) Carcinogen-induced gene promoter hypermethylation is mediated by DNMT1 and causal for transformation of immortalized bronchial epithelial cells. Cancer Res 68: 9005-9014.
226. Robert MF, Morin S, Beaulieu N, Gauthier F, Chute IC, et al. (2003) DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. Nat Genet 33: 61-65.
227. Satta R, Maloku E, Zhubi A, Pibiri F, Hajos M, et al. (2008) Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. Proc Natl Acad Sci U S A 105: 16356-16361.
228. Liu F, Killian JK, Yang M, Walker RL, Hong JA, et al. (2010) Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. Oncogene 29: 3650-3664.
229. Robertson KD, Keyomarsi K, Gonzales FA, Velicescu M, Jones PA (2000) Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells. Nucleic Acids Res 28: 2108-2113.
230. Weber SM, Bornstein S, Li Y, Malkoski SP, Wang D, et al. (2011) Tobacco-specific carcinogen nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone induces AKT activation in head and neck epithelia. Int J Oncol 39: 1193-1198.
231. Lin RK, Hsieh YS, Lin P, Hsu HS, Chen CY, et al. (2010) The tobacco-specific carcinogen NNK induces DNA methyltransferase 1 accumulation and tumor suppressor gene hypermethylation in mice and lung cancer patients. J Clin Invest 120: 521-532.
232. Wang J, Xu Y, Li J, Sun X, Wang LP, et al. (2012) The tobacco-specific carcinogen NNK induces DNA methyltransferase 1 accumulation in laryngeal carcinoma. Oral Oncol 48: 541-546.
233. Mimura J, Ema M, Sogawa K, Fujii-Kuriyama Y (1999) Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. Genes Dev 13: 20-25.
234. Haarmann-Stemmann T, Bothe H, Kohli A, Sydlik U, Abel J, et al. (2007) Analysis of the transcriptional regulation and molecular function of the aryl hydrocarbon receptor repressor in human cell lines. Drug Metab Dispos 35: 2262-2269.
235. Lawrence BP, Sherr DH (2012) You AhR what you eat? Nat Immunol 13: 117-119.
236. Matthews J, Gustafsson JA (2006) Estrogen receptor and aryl hydrocarbon receptor signaling pathways. Nucl Recept Signal 4: e016.
237. Fernandez-Salguero PM, Hilbert DM, Rudikoff S, Ward JM, Gonzalez FJ (1996) Aryl-hydrocarbon receptor-deficient mice are resistant to 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced toxicity. Toxicol Appl Pharmacol 140: 173-179.
238. Mimura J, Yamashita K, Nakamura K, Morita M, Takagi TN, et al. (1997) Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. Genes Cells 2: 645-654.
239. Kasai A, Hiramatsu N, Hayakawa K, Yao J, Maeda S, et al. (2006) High levels of dioxin-like potential in cigarette smoke evidenced by in vitro and in vivo biosensing. Cancer Res 66: 7143-7150.
240. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ (2004) Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem 279: 23847-23850.
241. Mimura J, Fujii-Kuriyama Y (2003) Functional role of AhR in the expression of toxic effects by TCDD. Biochim Biophys Acta 1619: 263-268.

242. Arsenescu R, Arsenescu V, Zhong J, Nasser M, Melinte R, et al. (2011) Role of the xenobiotic receptor in inflammatory bowel disease. Inflamm Bowel Dis 17: 1149-1162.
243. Chiba T, Chihara J, Furue M (2012) Role of the Arylhydrocarbon Receptor (AhR) in the Pathology of Asthma and COPD. J Allergy (Cairo) 2012: 372384.
244. Hauben E, Gregori S, Draghici E, Migliavacca B, Olivieri S, et al. (2008) Activation of the aryl hydrocarbon receptor promotes allograft-specific tolerance through direct and dendritic cell-mediated effects on regulatory T cells. Blood 112: 1214-1222.
245. Khanna AK, Xu J, Uber PA, Burke AP, Baquet C, et al. (2009) Tobacco smoke exposure in either the donor or recipient before transplantation accelerates cardiac allograft rejection, vascular inflammation, and graft loss. Circulation 120: 1814-1821.
246. Savouret JF, Berdeaux A, Casper RF (2003) The aryl hydrocarbon receptor and its xenobiotic ligands: a fundamental trigger for cardiovascular diseases. Nutr Metab Cardiovasc Dis 13: 104-113.
247. Kerley-Hamilton JS, Trask HW, Ridley CJ, Dufour E, Lesseur C, et al. (2012) Inherent and benzo[a]pyrene-induced differential aryl hydrocarbon receptor signaling greatly affects life span, atherosclerosis, cardiac gene expression, and body and heart growth in mice. Toxicol Sci 126: 391-404.
248. Evans BR, Karchner SI, Allan LL, Pollenz RS, Tanguay RL, et al. (2008) Repression of aryl hydrocarbon receptor (AHR) signaling by AHR repressor: role of DNA binding and competition for AHR nuclear translocator. Mol Pharmacol 73: 387-398.
249. Karchner SI, Franks DG, Powell WH, Hahn ME (2002) Regulatory interactions among three members of the vertebrate aryl hydrocarbon receptor family: AHR repressor, AHR1, and AHR2. J Biol Chem 277: 6949-6959.
250. Zudaire E, Cuesta N, Murty V, Woodson K, Adams L, et al. (2008) The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. J Clin Invest 118: 640-650.
251. Androutsopoulos VP, Tsatsakis AM, Spandidos DA (2009) Cytochrome P450 CYP1A1: wider roles in cancer progression and prevention. BMC Cancer 9: 187.
252. Zhan P, Wang Q, Qian Q, Wei SZ, Yu LK (2011) CYP1A1 MspI and exon7 gene polymorphisms and lung cancer risk: an updated meta-analysis and review. J Exp Clin Cancer Res 30: 99.
253. Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, et al. (2012) 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect 120: 1425-1431.
254. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, et al. (2012) Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. Am J Med Genet B Neuropsychiatr Genet 159B: 141-151.
255. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, et al. (2012) Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet.
256. Bosse Y, Postma DS, Sin DD, Lamontagne M, Couture C, et al. (2012) Molecular signature of smoking in human lung tissues. Cancer Res 72: 3753-3763.
257. Philibert RA, Beach SR, Brody GH (2012) Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. Epigenetics 7: 1331-1338.
258. Millan JL, Eriksson A, Stigbrand T (1982) A possible new locus of alkaline phosphatase expressed in human testis. Hum Genet 62: 293-295.
259. Knoll BJ, Rothblum KN, Longley M (1988) Nucleotide sequence of the human placental alkaline phosphatase gene. Evolution of the 5' flanking region by deletion/substitution. J Biol Chem 263: 12020-12027.
260. Millan JL, Manes T (1988) Seminoma-derived Nagao isozyme is encoded by a germ-cell alkaline phosphatase gene. Proc Natl Acad Sci U S A 85: 3024-3028.

261. Meyer RE, Thompson SJ, Addy CL, Garrison CZ, Best RG (1995) Maternal serum placental alkaline phosphatase level and risk for preterm delivery. Am J Obstet Gynecol 173: 181-186.
262. Moawad AH, Goldenberg RL, Mercer B, Meis PJ, Iams JD, et al. (2002) The Preterm Prediction Study: the value of serum alkaline phosphatase, alpha-fetoprotein, plasma corticotropin-releasing hormone, and other serum markers for the prediction of spontaneous preterm birth. Am J Obstet Gynecol 186: 990-996.
263. Brock DJ, Barron L (1988) Measurement of placental alkaline phosphatase in maternal plasma as an indicator of subsequent low birthweight outcome. Br J Obstet Gynaecol 95: 79-83.
264. Mosbah AA, Abd-Ellatif NA, Sorour EI, El-Halaby AF (2011) Placental alkaline phosphatase activity and its relation to foetal growth and nutrition in appropriate and small for gestational age newborns at term. J Egypt Soc Parasitol 41: 745-752.
265. Fox H, Agrafojo-Blanco A (1974) Scanning electron microscopy of the human placenta in normal and abnormal pregnancies. Eur J Obstet Gynecol Reprod Biol 4: 45-50.
266. Nielsen OS, Munro AJ, Duncan W, Sturgeon J, Gospodarowicz MK, et al. (1990) Is placental alkaline phosphatase (PLAP) a useful marker for seminoma? Eur J Cancer 26: 1049-1054.
267. Koshida K, Stigbrand T, Munck-Wikland E, Hisazumi H, Wahren B (1990) Analysis of serum placental alkaline phosphatase activity in testicular cancer and cigarette smokers. Urol Res 18: 169-173.
268. Muensch H, Maslow W, Azama F (1984) Serum heat stable alkaline phosphatase activity in smokers and non-smokers. Prog Clin Biol Res 166: 317-325.
269. Kallioniemi OP, Nieminen MM, Lehtinen J, Veneskoski T, Koivula T (1987) Increased serum placental-like alkaline phosphatase activity in smokers originates from the lungs. Eur J Respir Dis 71: 170-176.
270. Koshida K, Uchibayashi T, Yamamoto H, Hirano K (1996) Significance of placental alkaline phosphatase (PLAP) in the monitoring of patients with seminoma. Br J Urol 77: 138-142.
271. Tucker DF, Oliver RT, Travers P, Bodmer WF (1985) Serum marker potential of placental alkaline phosphatase-like activity in testicular germ cell tumours evaluated by H17E2 monoclonal antibody assay. Br J Cancer 51: 631-639.
272. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, et al. (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452: 633-637.
273. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, et al. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 40: 616-622.
274. Le Marchand L, Derby KS, Murphy SE, Hecht SS, Hatsukami D, et al. (2008) Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. Cancer Res 68: 9137-9140.
275. Picciotto MR, Kenny PJ (2012) Molecular Mechanisms Underlying Behaviors Related to Nicotine Addiction. Cold Spring Harb Perspect Med.
276. Bierut LJ (2010) Convergence of genetic findings for nicotine dependence and smoking related diseases with chromosome 15q24-25. Trends Pharmacol Sci 31: 46-51.
277. Sugiyama T, Kantake N (2009) Dynamic regulatory interactions of rad51, rad52, and replication protein-a in recombination intermediates. J Mol Biol 390: 45-55.
278. Richardson C (2005) RAD51, genomic stability, and tumorigenesis. Cancer Lett 218: 127-139.
279. Lok BH, Carley AC, Tchang B, Powell SN (2012) RAD52 inactivation is synthetically lethal with deficiencies in BRCA1 and PALB2 in addition to BRCA2 through RAD51-mediated homologous recombination. Oncogene.
280. Suwaki N, Klare K, Tarsounas M (2011) RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. Semin Cell Dev Biol 22: 898-905.
281. Lio YC, Mazin AV, Kowalczykowski SC, Chen DJ (2003) Complex formation by the human Rad51B and Rad51C DNA repair proteins and their activities in vitro. J Biol Chem 278: 2469-2478.

282. Chun J, Buechelmaier ES, Powell SN (2013) Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. Mol Cell Biol 33: 387-395.
283. Lok BH, Powell SN (2012) Molecular pathways: understanding the role of Rad52 in homologous recombination for therapeutic advancement. Clin Cancer Res 18: 6400-6406.
284. Goodsell DS (2005) The molecular perspective: RAD51 and BRCA2. Oncologist 10: 555-556.
285. Jensen RB OA, Kim T, Estep A, Kowalczykowski SC. (2013) BRCA2 is epistatic to the RAD51 paralogs in response to DNA damage. DNA Repair (Amst) Epub ahead of print.
286. Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. Nature 461: 1071-1078.
287. Ciccia A, Elledge SJ (2010) The DNA damage response: making it safe to play with knives. Mol Cell 40: 179-204.
288. Leng S, Stidley CA, Willink R, Bernauer A, Do K, et al. (2008) Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. Cancer Res 68: 3049-3056.
289. Lee MN, Tseng RC, Hsu HS, Chen JY, Tzao C, et al. (2007) Epigenetic inactivation of the chromosomal stability control genes BRCA1, BRCA2, and XRCC5 in non-small cell lung cancer. Clin Cancer Res 13: 832-838.
290. Qin HD, Shugart YY, Bei JX, Pan QH, Chen L, et al. (2011) Comprehensive pathway-based association study of DNA repair gene variants and the risk of nasopharyngeal carcinoma. Cancer Res 71: 3000-3008.
291. Tsuji T, Aoshiba K, Nagai A (2004) Cigarette smoke induces senescence in alveolar epithelial cells. Am J Respir Cell Mol Biol 31: 643-649.
292. Nyunoya T, Monick MM, Klingelhutz A, Yarovinsky TO, Cagley JR, et al. (2006) Cigarette smoke induces cellular senescence. Am J Respir Cell Mol Biol 35: 681-688.
293. Chen D, Xu W, Bales E, Colmenares C, Conacci-Sorrell M, et al. (2003) SKI activates Wnt/beta-catenin signaling in human melanoma. Cancer Res 63: 6626-6634.
294. Ding B, Sun Y, Huang J (2012) Overexpression of SKI oncoprotein leads to p53 degradation through regulation of MDM2 protein sumoylation. J Biol Chem 287: 14621-14630.
295. Lou Z, Chen J (2006) Cellular senescence and DNA repair. Exp Cell Res 312: 2641-2646.
296. Kaufmann WK, Paules RS (1996) DNA damage and cell cycle checkpoints. FASEB J 10: 238-247.
297. Norbury CJ, Zhivotovsky B (2004) DNA damage-induced apoptosis. Oncogene 23: 2797-2808.
298. Burton DG (2009) Cellular senescence, ageing and disease. Age (Dordr) 31: 1-9.
299. Oukka M, Kim ST, Lugo G, Sun J, Wu LC, et al. (2002) A mammalian homolog of Drosophila schnurri, KRC, regulates TNF receptor-driven responses and interacts with TRAF2. Mol Cell 9: 121-131.
300. Wu LC (2002) ZAS: C2H2 zinc finger proteins involved in growth and development. Gene Expr 10: 137-152.
301. Oukka M, Wein MN, Glimcher LH (2004) Schnurri-3 (KRC) interacts with c-Jun to regulate the IL-2 gene in T cells. J Exp Med 199: 15-24.
302. Jones DC, Wein MN, Glimcher LH (2007) Schnurri-3 is an essential regulator of osteoblast function and adult bone mass. Ann Rheum Dis 66 Suppl 3: iii49-51.
303. Duan Z, Zarebski A, Montoya-Durango D, Grimes HL, Horwitz M (2005) Gfi1 coordinates epigenetic repression of p21Cip/WAF1 by recruitment of histone lysine methyltransferase G9a and histone deacetylase 1. Mol Cell Biol 25: 10338-10351.
304. Moroy T (2005) The zinc finger transcription factor Growth factor independence 1 (Gfi1). Int J Biochem Cell Biol 37: 541-546.
305. Moroy T, Khandanpour C (2011) Growth factor independence 1 (Gfi1) as a regulator of lymphocyte development and activation. Semin Immunol 23: 368-378.

306. Kazanjian A, Wallis D, Au N, Nigam R, Venken KJ, et al. (2004) Growth factor independence-1 is expressed in primary human neuroendocrine lung carcinomas and mediates the differentiation of murine pulmonary neuroendocrine cells. Cancer Res 64: 6874-6882.
307. Zhu J, Davidson TS, Wei G, Jankovic D, Cui K, et al. (2009) Down-regulation of Gfi-1 expression by TGF-beta is important for differentiation of Th17 and CD103+ inducible regulatory T cells. J Exp Med 206: 329-341.
308. Tomita K, Caramori G, Lim S, Ito K, Hanazawa T, et al. (2002) Increased p21(CIP1/WAF1) and B cell lymphoma leukemia-x(L) expression and reduced apoptosis in alveolar macrophages from smokers. Am J Respir Crit Care Med 166: 724-731.
309. Charlesworth JC, Curran JE, Johnson MP, Goring HH, Dyer TD, et al. (2010) Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. BMC Med Genomics 3: 29.
310. Wells RA, Catzavelos C, Kamel-Reid S (1997) Fusion of retinoic acid receptor alpha to NuMA, the nuclear mitotic apparatus protein, by a variant translocation in acute promyelocytic leukaemia. Nat Genet 17: 109-113.
311. Doyle TJ, Braun KW, McLean DJ, Wright RW, Griswold MD, et al. (2007) Potential functions of retinoic acid receptor A in Sertoli cells and germ cells during spermatogenesis. Ann N Y Acad Sci 1120: 114-130.
312. Hall JA, Cannons JL, Grainger JR, Dos Santos LM, Hand TW, et al. (2011) Essential role for retinoic acid in the promotion of CD4(+) T cell effector responses via retinoic acid receptor alpha. Immunity 34: 435-447.
313. Laursen KB, Wong PM, Gudas LJ (2012) Epigenetic regulation by RARalpha maintains ligand-independent transcriptional activity. Nucleic Acids Res 40: 102-115.
314. Bergheim I, Wolfgarten E, Bollschweiler E, Holscher AH, Bode CH, et al. (2005) Role of retinoic acid receptors in squamous-cell carcinoma in human esophagus. J Carcinog 4: 20.
315. Hurt JA, Obar RA, Zhai B, Farny NG, Gygi SP, et al. (2009) A conserved CCCH-type zinc finger protein regulates mRNA nuclear adenylation and export. J Cell Biol 185: 265-277.
316. Hurowitz EH, Melnyk JM, Chen YJ, Kouros-Mehr H, Simon MI, et al. (2000) Genomic characterization of the human heterotrimeric G protein alpha, beta, and gamma subunit genes. DNA Res 7: 111-120.
317. Larson KC, Lipko M, Dabrowski M, Draper MP (2010) Gng12 is a novel negative regulator of LPS-induced inflammation in the microglial cell line BV-2. Inflamm Res 59: 15-22.
318. Uhl GR, Drgon T, Johnson C, Ramoni MF, Behm FM, et al. (2010) Genome-wide association for smoking cessation success in a trial of precessation nicotine replacement. Mol Med 16: 513-526.
319. Pot C (2012) Aryl hydrocarbon receptor controls regulatory CD4+ T cell function. Swiss Med Wkly 142: w13592.
320. Gomolka E, Radomska M, Sulka A (2009) [Tobacco smoking by patients treat for alcohol addiction--interaction of psychoactive substances]. Przegl Lek 66: 624-627.
321. Satheesh SV, Kunert K, Ruttiger L, Zuccotti A, Schonig K, et al. (2012) Retrocochlear function of the peripheral deafness gene Cacna1d. Hum Mol Genet 21: 3896-3909.
322. Reinbothe TM, Alkayyali S, Ahlqvist E, Tuomi T, Isomaa B, et al. (2013) The human L-type calcium channel Ca(v)1.3 regulates insulin release and polymorphisms in CACNA1D associate with type 2 diabetes. Diabetologia 56: 340-349.
323. Attvall S, Fowelin J, Lager I, Von Schenck H, Smith U (1993) Smoking induces insulin resistance--a potential link with the insulin resistance syndrome. J Intern Med 233: 327-332.
324. Houston TK, Person SD, Pletcher MJ, Liu K, Iribarren C, et al. (2006) Active and passive smoking and development of glucose intolerance among young adults in a prospective cohort: CARDIA study. BMJ 332: 1064-1069.
325. Facchini FS, Hollenbeck CB, Jeppesen J, Chen YD, Reaven GM (1992) Insulin resistance and cigarette smoking. Lancet 339: 1128-1130.

326. Krall EA, Dawson-Hughes B (1999) Smoking increases bone loss and decreases intestinal calcium absorption. J Bone Miner Res 14: 215-220.
327. Krall EA, Dawson-Hughes B (1991) Smoking and bone loss among postmenopausal women. J Bone Miner Res 6: 331-338.
328. Brot C, Jorgensen NR, Sorensen OH (1999) The influence of smoking on vitamin D status and calcium metabolism. Eur J Clin Nutr 53: 920-926.
329. Kahn ML, Nakanishi-Matsui M, Shapiro MJ, Ishihara H, Coughlin SR (1999) Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. J Clin Invest 103: 879-887.
330. Chen HT, Tsou HK, Tsai CH, Kuo CC, Chiang YK, et al. (2010) Thrombin enhanced migration and MMPs expression of human chondrosarcoma cells involves PAR receptor signaling pathway. J Cell Physiol 223: 737-745.
331. Dangwal S, Rauch BH, Gensch T, Dai L, Bretschneider E, et al. (2011) High glucose enhances thrombin responses via protease-activated receptor-4 in human vascular smooth muscle cells. Arterioscler Thromb Vasc Biol 31: 624-633.
332. Zhang Y, Yu G, Jiang P, Xiang Y, Li W, et al. (2011) Decreased expression of protease-activated receptor 4 in human gastric cancer. Int J Biochem Cell Biol 43: 1277-1283.
333. Breitling LP, Salzmann K, Rothenbacher D, Burwinkel B, Brenner H (2012) Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease. Eur Heart J.
334. Goto A, Hoshino M, Matsuda M, Nakamura T (2011) Phosphorylation of STEF/Tiam2 by protein kinase A is critical for Rac1 activation and neurite outgrowth in dibutyryl cAMP-treated PC12D cells. Mol Biol Cell 22: 1780-1790.
335. Yoshizawa M, Hoshino M, Sone M, Nabeshima Y (2002) Expression of stef, an activator of Rac1, correlates with the stages of neuronal morphological development in the mouse brain. Mech Dev 113: 65-68.
336. Rooney C, White G, Nazgiewicz A, Woodcock SA, Anderson KI, et al. (2010) The Rac activator STEF (Tiam2) regulates cell migration by microtubule-mediated focal adhesion disassembly. EMBO Rep 11: 292-298.
337. Chen JS, Su IJ, Leu YW, Young KC, Sun HS (2012) Expression of T-cell lymphoma invasion and metastasis 2 (TIAM2) promotes proliferation and invasion of liver cancer. Int J Cancer 130: 1302-1313.
338. Blaak H, Boers PH, Gruters RA, Schuitemaker H, van der Ende ME, et al. (2005) CCR5, GPR15, and CXCR6 are major coreceptors of human immunodeficiency virus type 2 variants isolated from individuals with and without plasma viremia. J Virol 79: 1686-1700.
339. Heiber M, Marchese A, Nguyen T, Heng HH, George SR, et al. (1996) A novel human gene encoding a G-protein-coupled receptor (GPR15) is located on chromosome 3. Genomics 32: 462-465.
340. Iida H, Iida M, Takenaka M, Fujiwara H, Dohi S (2006) Angiotensin II type 1 (AT1)-receptor blocker prevents impairment of endothelium-dependent cerebral vasodilation by acute cigarette smoking in rats. Life Sci 78: 1310-1316.
341. Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, et al. (2011) Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC Med Genet 12: 106.
342. Kisiel JB, Yab TC, Taylor WR, Chari ST, Petersen GM, et al. (2012) Stool DNA testing for the detection of pancreatic cancer: assessment of methylation marker candidates. Cancer 118: 2623-2631.
343. Omura N, Li CP, Li A, Hong SM, Walter K, et al. (2008) Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma. Cancer Biol Ther 7: 1146-1156.
344. Bralten LB, Gravendeel AM, Kloosterhof NK, Sacchetti A, Vrijenhoek T, et al. (2010) The CASPR2 cell adhesion molecule functions as a tumor suppressor gene in glioma. Oncogene 29: 6138-6148.

345. Hu YC, Yang ZH, Zhong KJ, Niu LJ, Pan XJ, et al. (2009) Alteration of transcriptional profile in human bronchial epithelial cells induced by cigarette smoke condensate. Toxicol Lett 190: 23-31.
346. Kim SY, Mo JW, Han S, Choi SY, Han SB, et al. (2010) The expression of non-clustered protocadherins in adult rat hippocampal formation and the connecting brain regions. Neuroscience 170: 189-199.
347. Basso K, Liso A, Tiacci E, Benedetti R, Pulsoni A, et al. (2004) Gene expression profiling of hairy cell leukemia reveals a phenotype related to memory B cells with altered expression of chemokine and adhesion receptors. J Exp Med 199: 59-68.
348. Wang C, Yu G, Liu J, Wang J, Zhang Y, et al. (2012) Downregulation of PCDH9 predicts prognosis for patients with glioma. J Clin Neurosci 19: 541-545.
349. Wei Z, Li M (2007) Genome-wide linkage and association analysis of rheumatoid arthritis in a Canadian population. BMC Proc 1 Suppl 1: S19.
350. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, et al. (2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol 13: R43.
351. Betancur C (2011) Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res 1380: 42-77.
352. Patino-Lopez G, Aravind L, Dong X, Kruhlak MJ, Ostap EM, et al. (2010) Myosin 1G is an abundant class I myosin in lymphocytes whose localization at the plasma membrane depends on its ancient divergent pleckstrin homology (PH) domain (Myo1PH). J Biol Chem 285: 8675-8686.
353. Pierce RA, Field ED, Mutis T, Golovina TN, Von Kap-Herr C, et al. (2001) The HA-2 minor histocompatibility antigen is derived from a diallelic gene encoding a novel human class I myosin protein. J Immunol 167: 3223-3230.
354. Olety B, Walte M, Honnert U, Schillers H, Bahler M (2010) Myosin 1G (Myo1G) is a haematopoietic specific myosin that localises to the plasma membrane and regulates cell elasticity. FEBS Lett 584: 493-499.
355. Groth-Pedersen L, Aits S, Corcelle-Termeau E, Petersen NH, Nylandsted J, et al. (2012) Identification of cytoskeleton-associated proteins essential for lysosomal stability and survival of human cancer cells. PLoS One 7: e45381.
356. Kubota T, Michigami T, Ozono K (2009) Wnt signaling in bone metabolism. J Bone Miner Metab 27: 265-271.
357. Fusby JS, Kassmeier MD, Palmer VL, Perry GA, Anderson DK, et al. (2010) Cigarette smoke-induced effects on bone marrow B-cell subsets and CD4+:CD8+ T-cell ratios are reversed by smoking cessation: influence of bone mass on immune cell response to and recovery from smoke exposure. Inhal Toxicol 22: 785-796.
358. Mammoto T, Chen J, Jiang E, Jiang A, Smith LE, et al. (2012) LRP5 regulates development of lung microvessels and alveoli through the angiopoietin-Tie2 pathway. PLoS One 7: e41596.
359. Fujino T, Asaba H, Kang MJ, Ikeda Y, Sone H, et al. (2003) Low-density lipoprotein receptor-related protein 5 (LRP5) is essential for normal cholesterol metabolism and glucose-induced insulin secretion. Proc Natl Acad Sci U S A 100: 229-234.
360. Palsgaard J, Emanuelli B, Winnay JN, Sumara G, Karsenty G, et al. (2012) Cross-talk between insulin and Wnt signaling in preadipocytes: role of Wnt co-receptor low density lipoprotein receptor-related protein-5 (LRP5). J Biol Chem 287: 12016-12026.
361. Dai W, Teodoridis JM, Zeller C, Graham J, Hersey J, et al. (2011) Systematic CpG islands methylation profiling of genes in the wnt pathway in epithelial ovarian cancer identifies biomarkers of progression-free survival. Clin Cancer Res 17: 4052-4062.
362. Carim-Todd L, Escarceller M, Estivill X, Sumoy L (2003) LRRN6A/LERN1 (leucine-rich repeat neuronal protein 1), a novel gene with enriched expression in limbic system and neocortex. Eur J Neurosci 18: 3167-3182.

363. Zhou BB, Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. Nature 408: 433-439.
364. Mocchegiani E, Giacconi R, Costarelli L (2011) Metalloproteases/anti-metalloproteases imbalance in chronic obstructive pulmonary disease: genetic factors and treatment implications. Curr Opin Pulm Med 17 Suppl 1: S11-19.
365. Braber S, Koelink PJ, Henricks PA, Jackson PL, Nijkamp FP, et al. (2011) Cigarette smoke-induced lung emphysema in mice is associated with prolyl endopeptidase, an enzyme involved in collagen breakdown. Am J Physiol Lung Cell Mol Physiol 300: L255-265.
366. Wang Z, Lu S, Liu C, Zhao B, Pei K, et al. (2010) Expressional and epigenetic alterations of placental matrix metalloproteinase 9 in preeclampsia. Gynecol Endocrinol 26: 96-102.
367. England L, Zhang J (2007) Smoking and risk of preeclampsia: a systematic review. Front Biosci 12: 2471-2483.
368. Kelly EA, Liu LY, Esnault S, Quinchia Johnson BH, Jarjour NN (2012) Potent synergistic effect of IL-3 and TNF on matrix metalloproteinase 9 generation by human eosinophils. Cytokine 58: 199-206.
369. Roskoski R, Jr. (2012) ERK1/2 MAP kinases: structure, function, and regulation. Pharmacol Res 66: 105-143.
370. Sopori M (2002) Effects of cigarette smoke on the immune system. Nat Rev Immunol 2: 372-377.
371. Coussens LM, Werb Z (2002) Inflammation and cancer. Nature 420: 860-867.
372. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 88: 450-457.
373. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, et al. (2012) Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. Hum Mol Genet.
374. Otto GP, Razi M, Morvan J, Stenner F, Tooze SA (2010) A novel syntaxin 6-interacting protein, SHIP164, regulates syntaxin 6-dependent sorting from early endosomes. Traffic 11: 688-705.
375. Longatti A, Lamb CA, Razi M, Yoshimura S, Barr FA, et al. (2012) TBC1D14 regulates autophagosome formation via Rab11- and ULK1-positive recycling endosomes. J Cell Biol 197: 659-675.
376. Mizushima N, Levine B, Cuervo AM, Klionsky DJ (2008) Autophagy fights disease through cellular self-digestion. Nature 451: 1069-1075.
377. Murphy SK, Adigun A, Huang Z, Overcash F, Wang F, et al. (2012) Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke. Gene 494: 36-43.
378. Vaissiere T, Hung RJ, Zaridze D, Moukeria A, Cuenin C, et al. (2009) Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. Cancer Res 69: 243-252.
379. Kaminsky Z, Wang SC, Petronis A (2006) Complex disease, gender and epigenetics. Ann Med 38: 530-544.
380. Chan HS, Chang SJ, Wang TY, Ko HJ, Lin YC, et al. (2012) Serine protease PRSS23 is upregulated by estrogen receptor alpha and associated with proliferation of breast cancer cells. PLoS One 7: e30397.
381. Suzuki M, Sunaga N, Shames DS, Toyooka S, Gazdar AF, et al. (2004) RNA interference-mediated knockdown of DNA methyltransferase 1 leads to promoter demethylation and gene re-expression in human lung and breast cancer cells. Cancer Res 64: 3137-3143.
382. Liu H, Zhou Y, Boggs SE, Belinsky SA, Liu J (2007) Cigarette smoke induces demethylation of prometastatic oncogene synuclein-gamma in lung cancer cells by downregulation of DNMT3B. Oncogene 26: 5900-5910.
383. Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, et al. (2011) Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. Genome Res 21: 515-524.

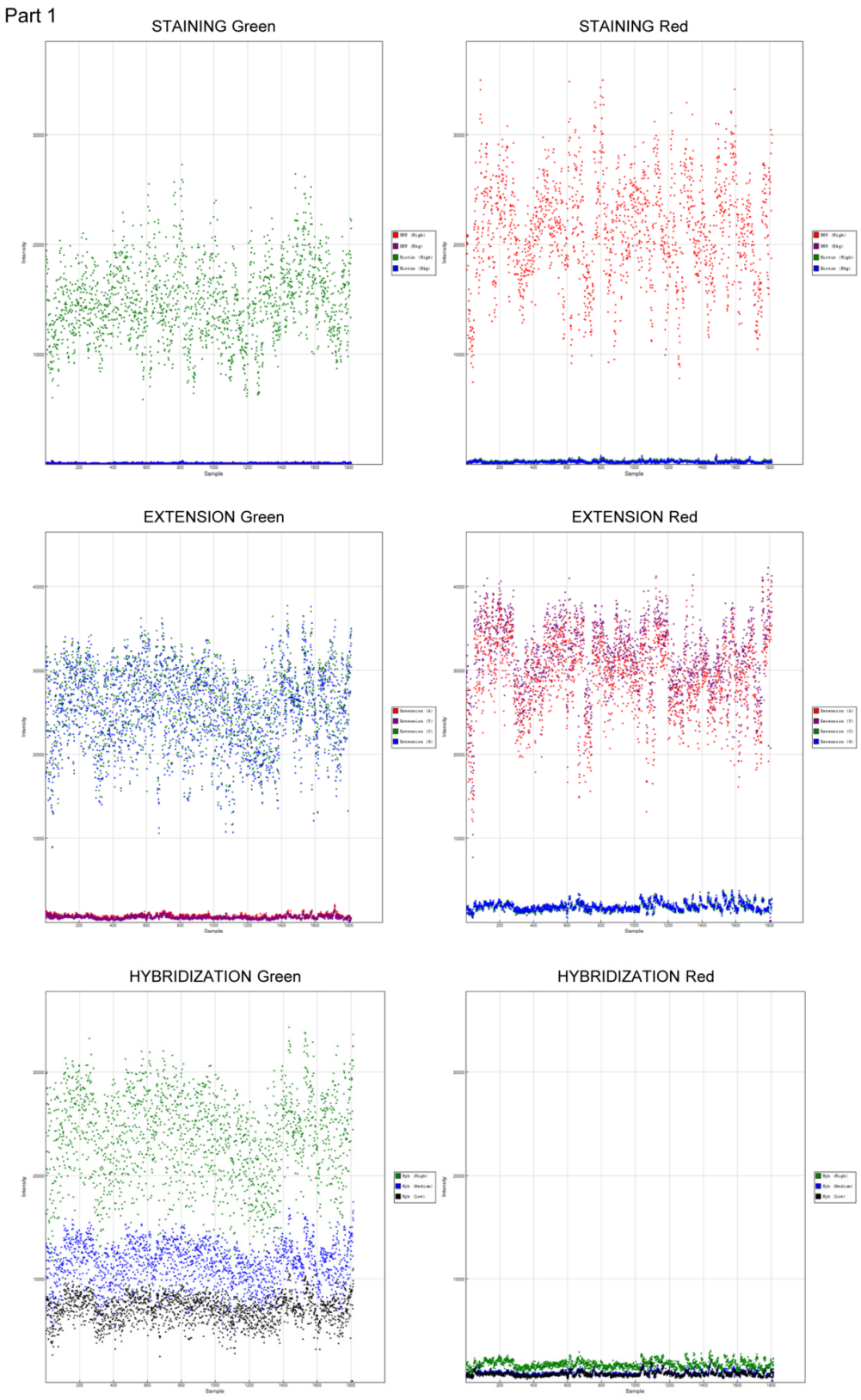
384. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, et al. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 41: 1350-1353.
385. Bock C, Walter J, Paulsen M, Lengauer T (2008) Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. Nucleic Acids Res 36: e55.
386. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466: 253-257.
387. Barrera V, Peinado MA (2012) Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale. Nucleic Acids Res 40: 11490-11498.
388. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ (2007) Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 35: e119.
389. Shen L, Waterland RA (2007) Methods of DNA methylation analysis. Curr Opin Clin Nutr Metab Care 10: 576-581.
390. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, et al. (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 7: e41361.
391. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, et al. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13: 86.
392. Kinney SR, Pradhan S (2013) Ten eleven translocation enzymes and 5-hydroxymethylation in mammalian development and cancer. Adv Exp Med Biol 754: 57-79.

Appendix

Figure A.1. Quality control of assay performance by GenomeStudio for KORA F4 &
Figure A.2. Quality control of assay performance by GenomeStudio for KORA F3

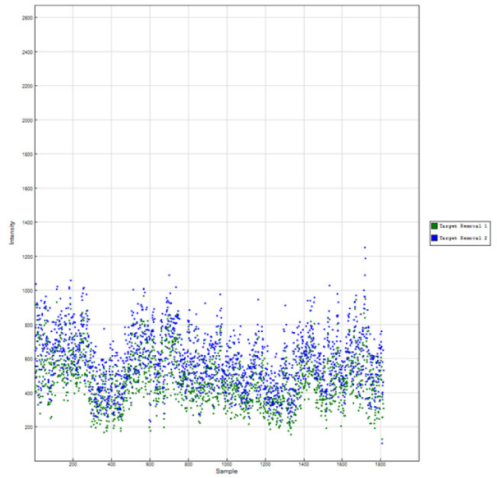
Staining controls have dinitrophenyl (DNP) or biotin, attached to the beads and monitor the performance of the staining step in both the red and green channels.
Extension controls examine the extension efficiency of A, T, C, and G nucleotides from a hairpin probe, and are monitored in both red (A, T) and green (C, G) channels.
Hybridization controls test the overall performance of the entire Methylation Assay using synthetic targets, which function as templates for the probes, instead of amplified DNA. These synthetic targets are present in the hybridization buffer (RA1) in three concentrations (high-concentration (5 pM), medium-concentration (1 pM), and low-concentration (0.2 pM)), and signals with various intensities corresponding to these concentrations should only be monitored in the green channel.
Target removal controls are present in the hybridization buffer RA1 and test the efficiency of the stripping step after the extension reaction. Their performance should result in low signal only in the green channel.
Bisulfite Conversion I and II controls use Infinium I respectively II probe design and allele-specific single base extension to monitor efficiency of bisulfite conversion. Bisulfite Conversion I: If the bisulfite conversion reaction was successful, the converted probes ("C") will match the converted sequence and be extended. If the sample has unconverted DNA, the unconverted probes ("U") will be extended. Performance of bisulfite conversion controls C1, C2, and C3 should be monitored in the Green channel, and controls C4, C5, and C6 should be monitored in Red channel. Bisulfite Conversion II: If the bisulfite conversion reaction was successful, the "A" base will get incorporated and the probe will have intensity in the Red channel. If the sample has unconverted DNA, the "G" base will get incorporated across the unconverted cytosine, and the probe will have elevated signal in the Green channel.
Specificity I and II controls monitor potential non-specific primer extension for Infinium I and Infinium II assay probes. Specificity I: Query probes for unmethylated and methylated state of each CpG locus are used. The A/T match corresponds to the unmethylated status of the interrogated C, and the G/C match corresponds to the methylated status of C. G/T mismatch controls check for non-specific detection of methylation signal over unmethylated background. PM controls correspond to A/T perfect match and should give high signal. MM controls correspond to G/T mismatch and should give low signal. Performance of GT Mismatch controls should be monitored in both green and red channels. Specificity II: These probes check for potential non-specific detection of methylation signal over unmethylated background, should incorporate the "A" base across the non-polymorphic T and have intensity in the Red channel. In case of nonspecific incorporation of the "G" base, the probe will have elevated signal in the Green channel.
Negative controls target bisulfite-converted sequences that do not contain CpGs and the mean signal of these probes defines the system background. Their performance should be monitored in both the green channel and the red channel.
Non-polymorphic controls test the overall performance of the assay, from amplification to detection, and allow comparison of assay performance across different samples. One non-polymorphic control has been designed for each of the four nucleotides (A, T, C, and G).
(Slides and description are adapted from the “Control Dashboard” of the Illumina GenomeStudio Methylation Module.)

Figure A.1. Quality control of assay performance by GenomeStudio for KORA F4.

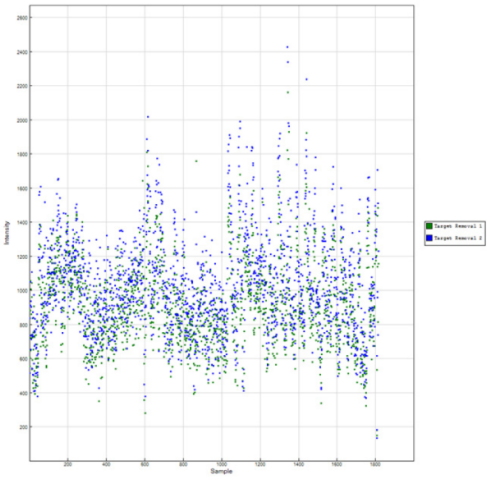


Part 2

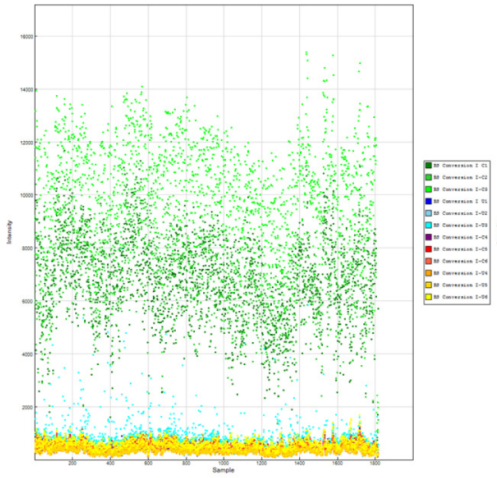
TARGET REMOVAL Green



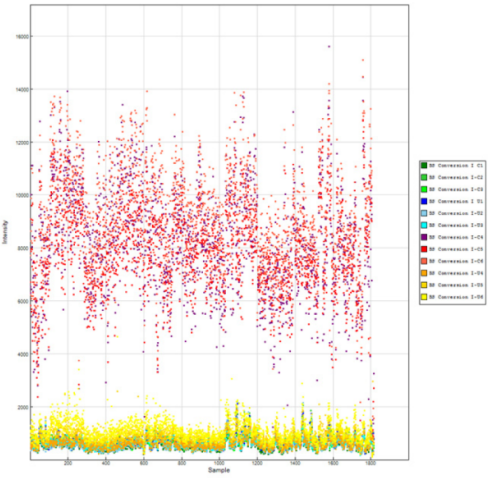
TARGET REMOVAL Red



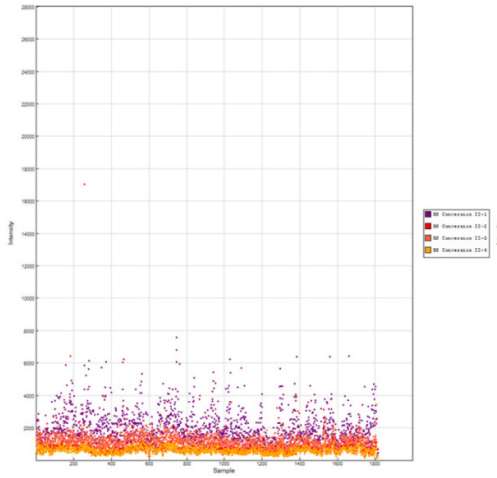
BISULFITE CONVERSION I Green



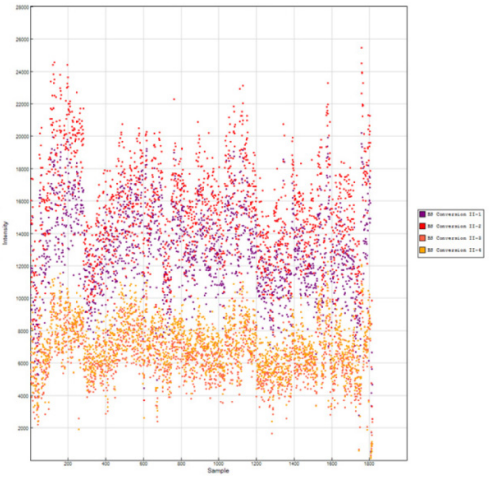
BISULFITE CONVERSION I Red



BISULFITE CONVERSION II Green

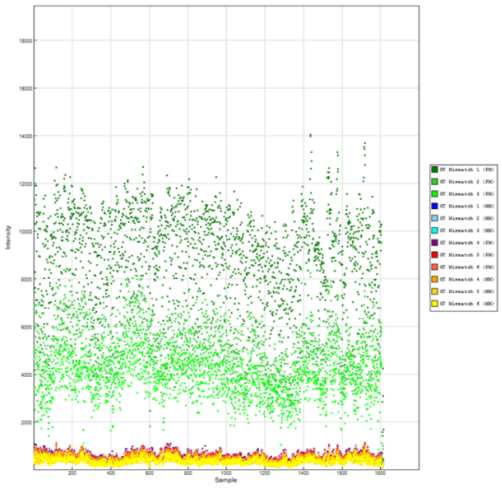


BISULFITE CONVERSION II Red

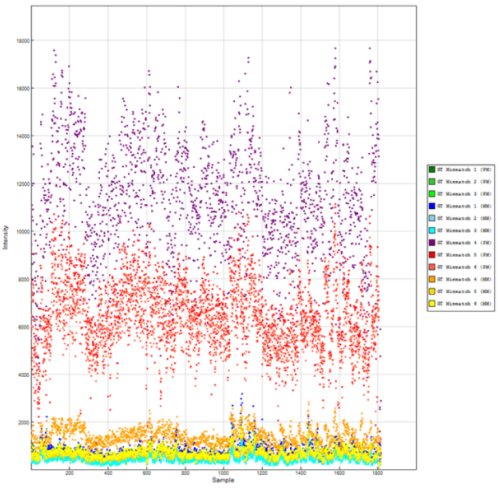


Part 3

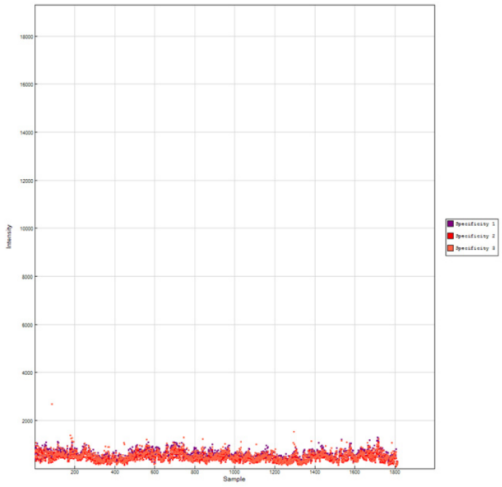
SPECIFICITY I Green



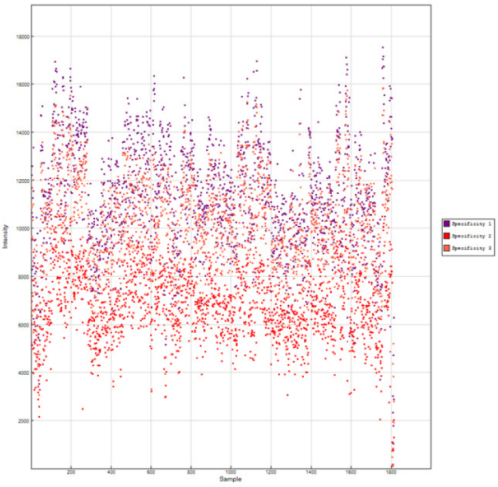
SPECIFICITY II Red



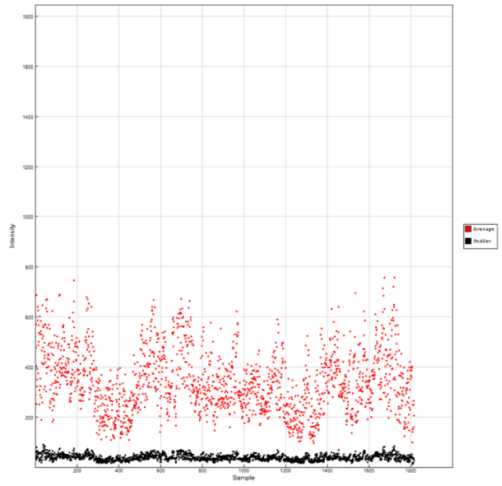
SPECIFICITY II Green



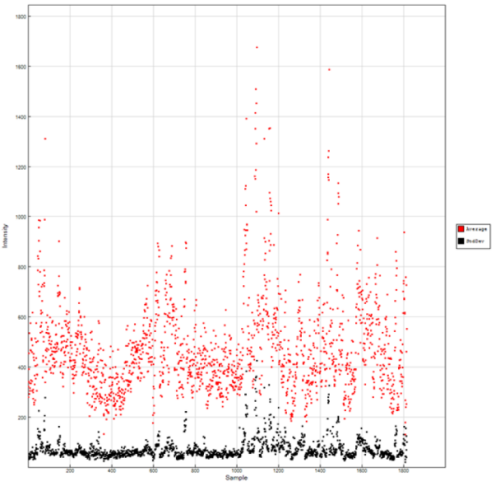
SPECIFICITY II Red



NEGATIVE Green



NEGATIVE Red



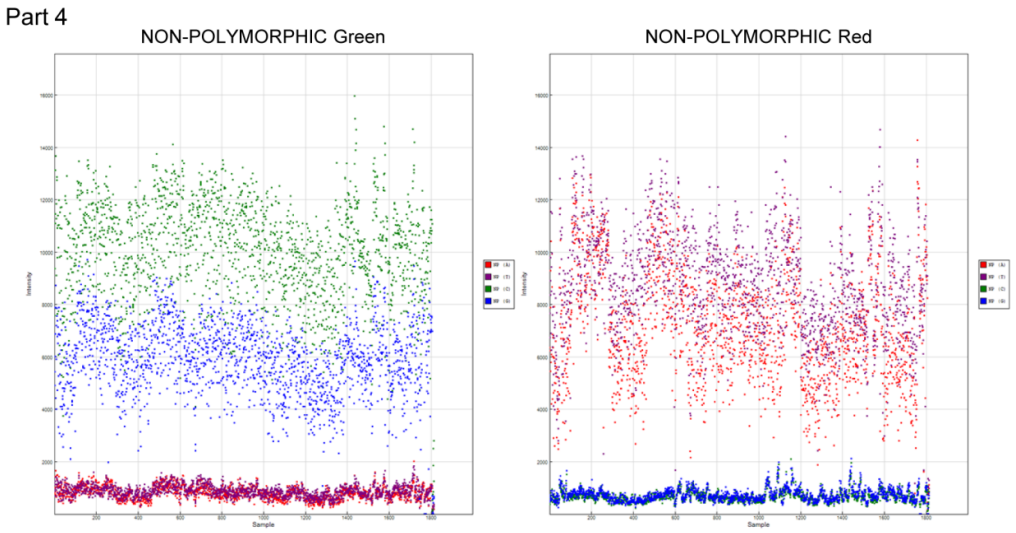
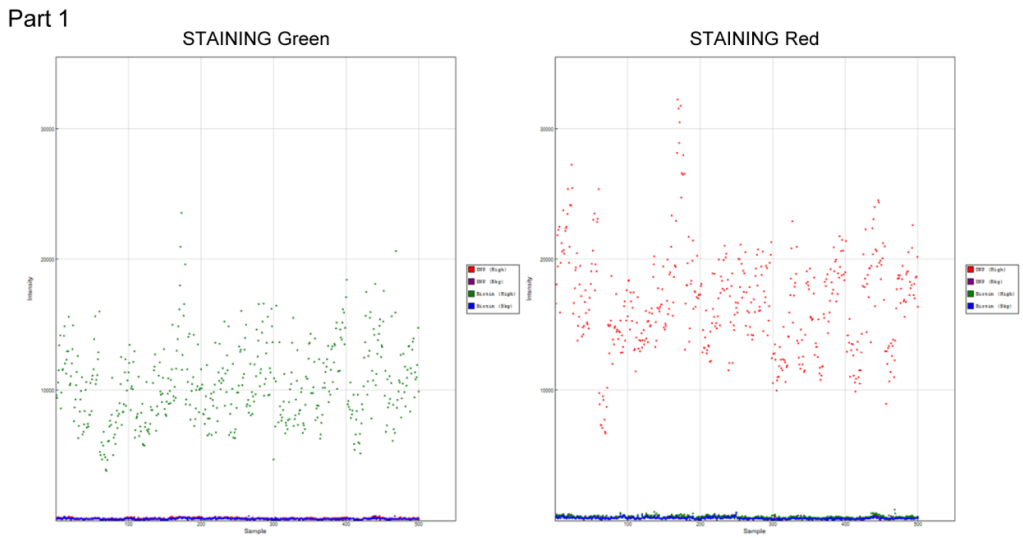
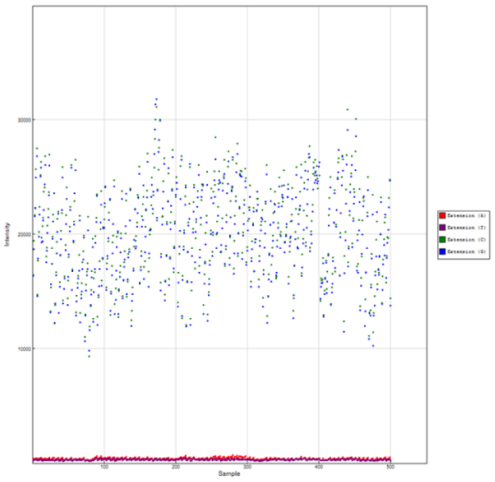


Figure A.2. Quality control of assay performance by GenomeStudio for KORA F3.

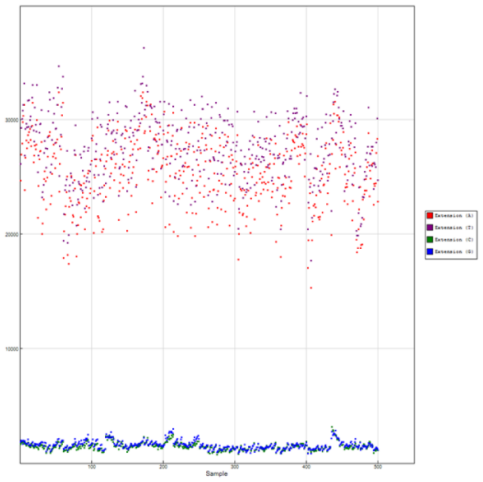


Part 2

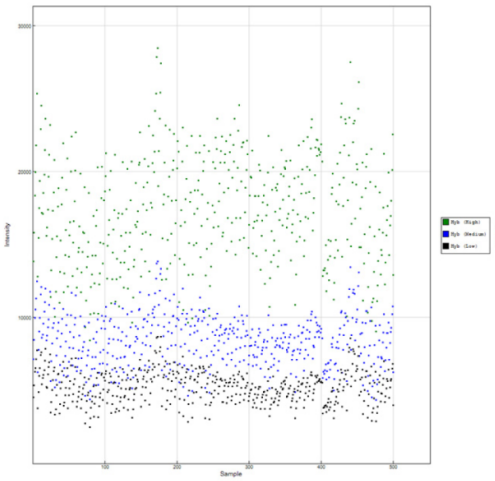
EXTENSION Green



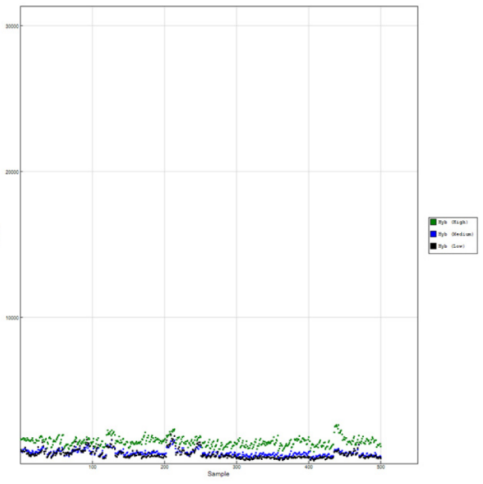
EXTENSION Red



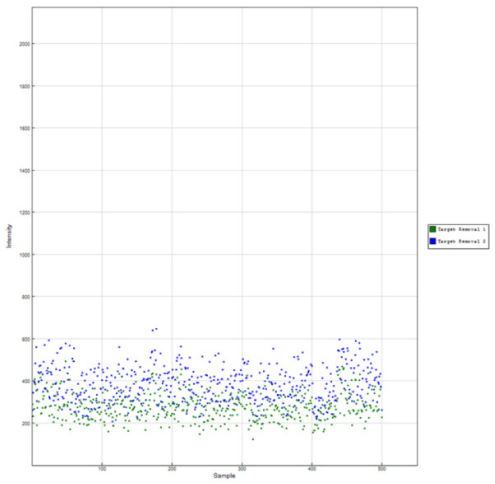
HYBRIDIZATION Green



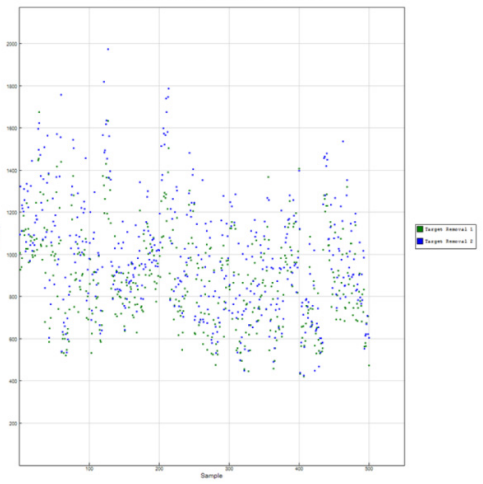
HYBRIDIZATION Red



TARGET REMOVAL Green

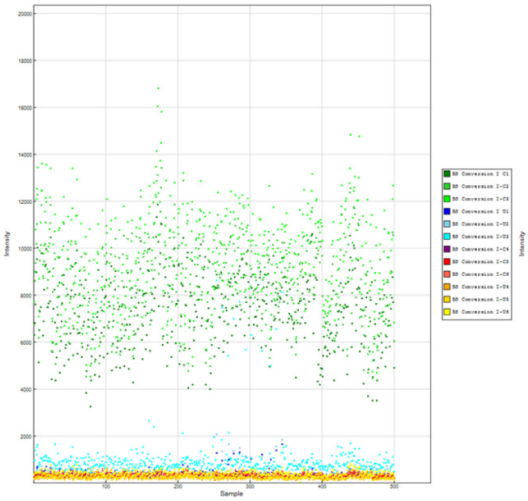


TARGET REMOVAL Red

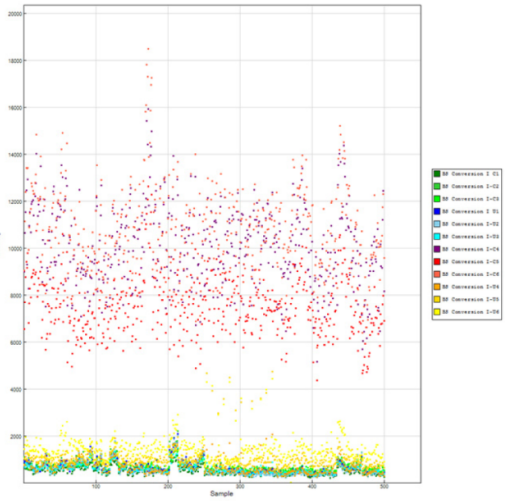


Part 3

BISULFITE CONVERSION I Green

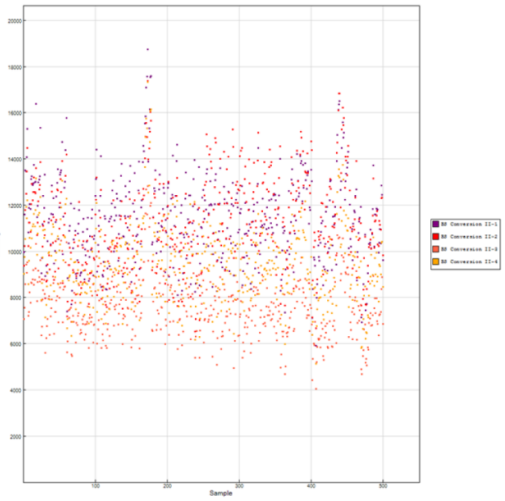


BISULFITE CONVERSION I Red

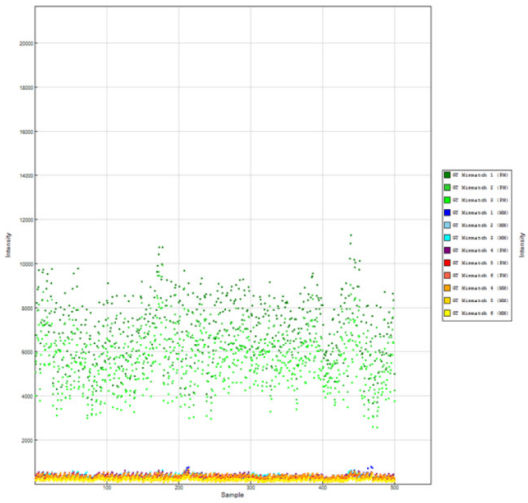


BISULFITE CONVERSION II Green

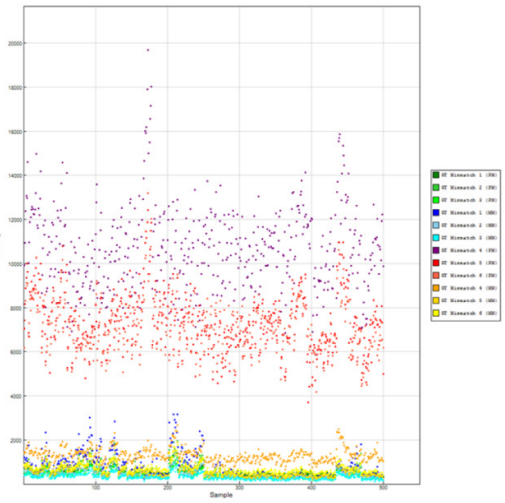
BISULFITE CONVERSION II Red



SPECIFICITY I Green

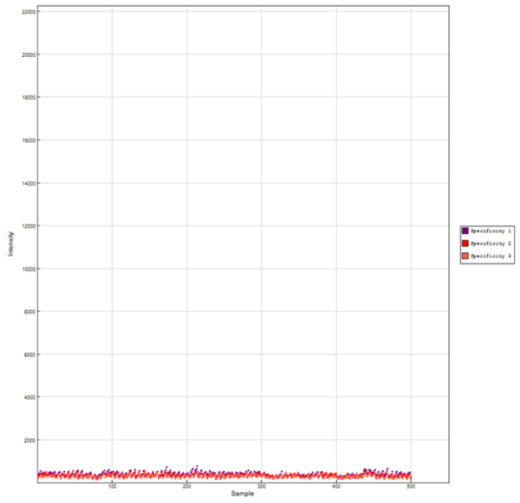


SPECIFICITY II Red

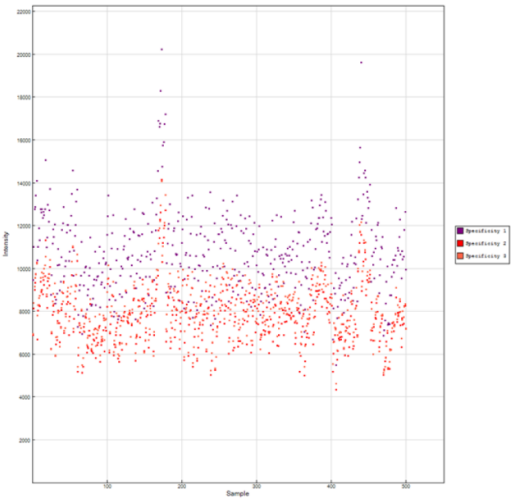


Part 4

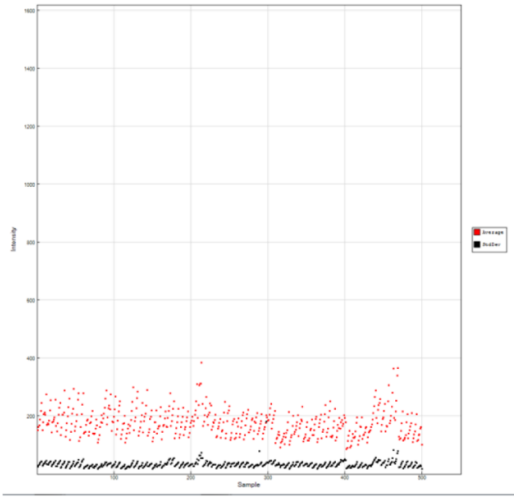
SPECIFICITY II Green



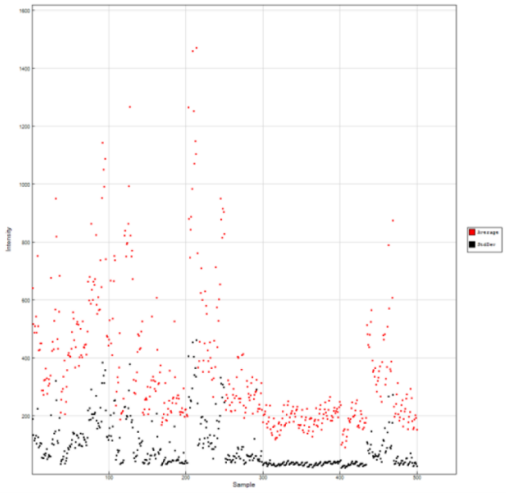
SPECIFICITY II Red



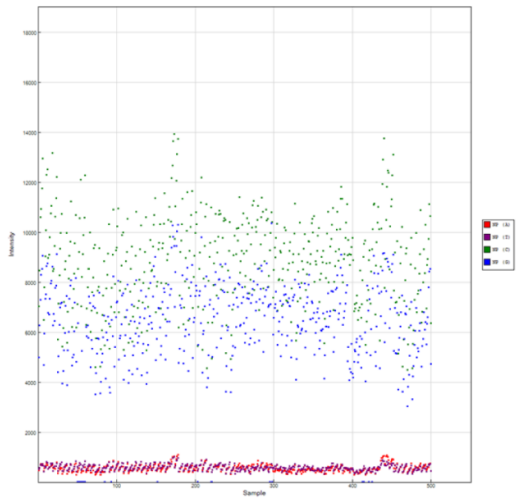
NEGATIVE Green



NEGATIVE Red



NON-POLYMORPHIC Green



NON-POLYMORPHIC Red

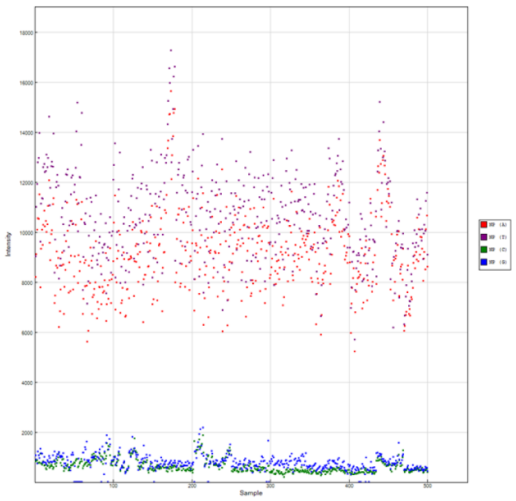
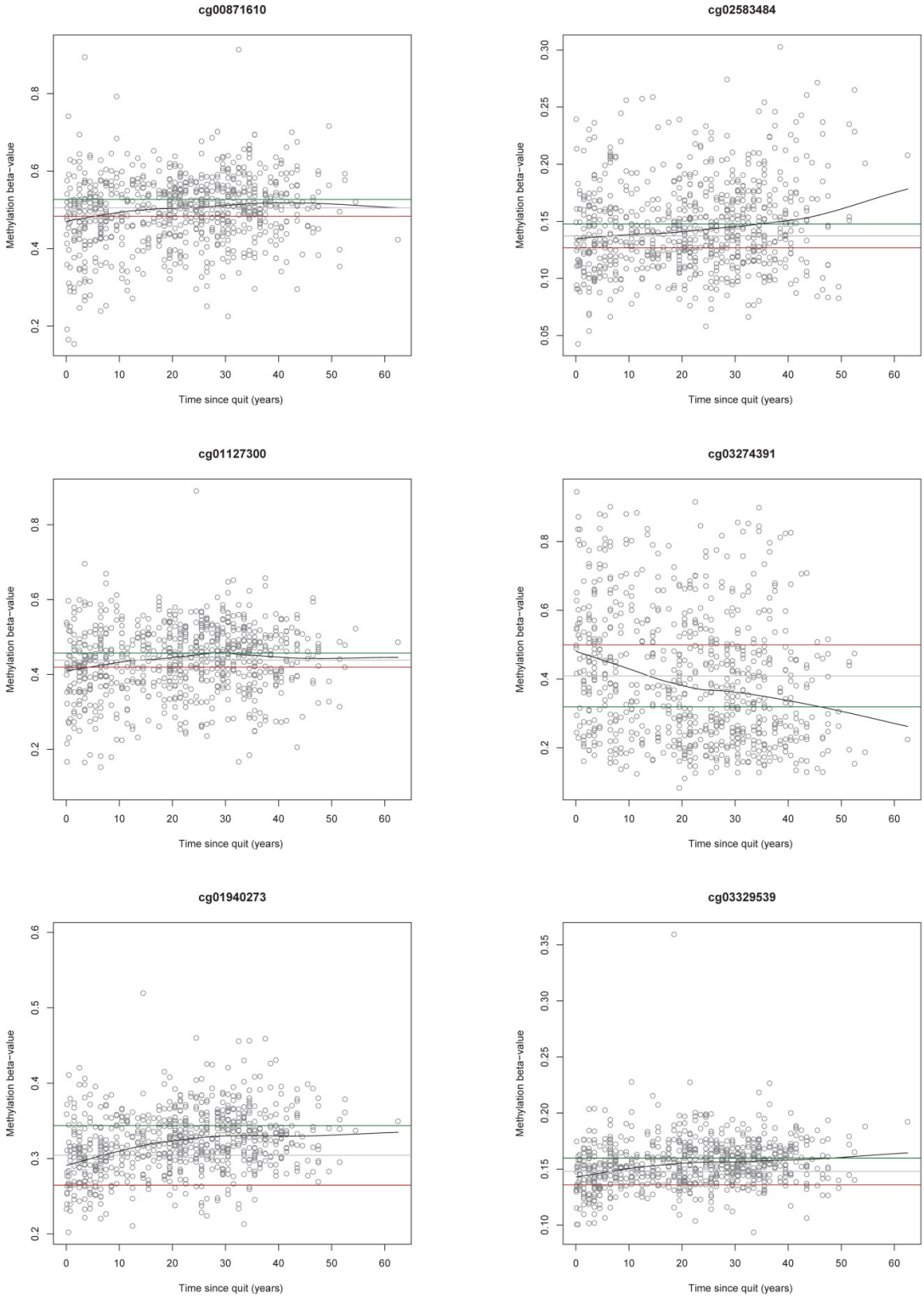
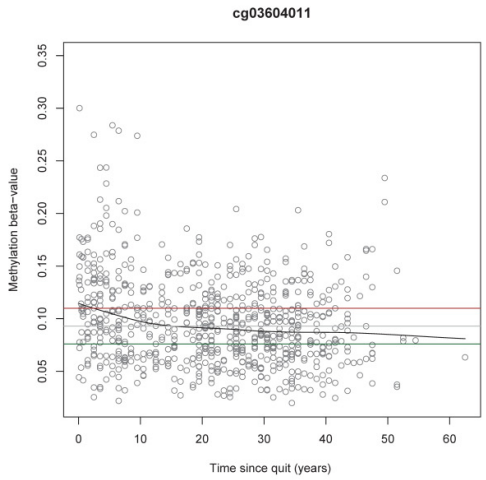
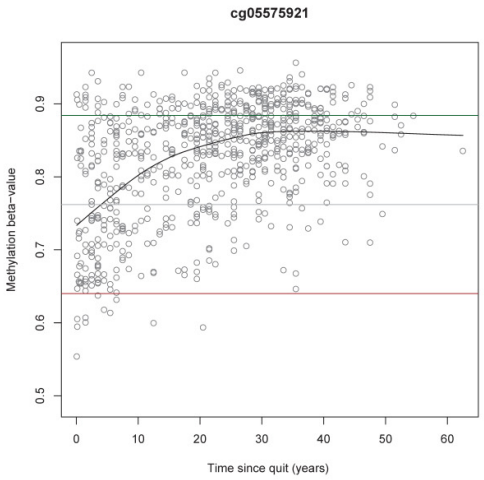


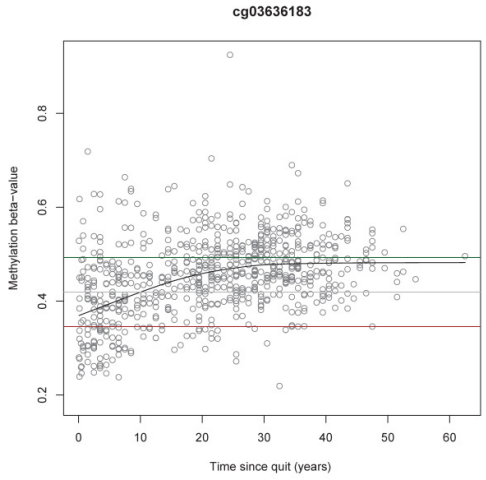
Figure A.3. Influence of time since smoking cessation on the DNA methylation state.
The years required for former smokers to obtain a median ß-value methylation state at single CpG sites that are closer to or equal the one of never smokers are illustrated by loess curves in the scatterplots; the x-axis display the cessation time in years, the y-axis display the methylation level with the use of numbers between 0 (for 0% methylation) and 1 (for 100% methylation); horizontal brown lines: median methylation level of current smokers; horizontal green lines: median methylation level of never smokers; horizontal grey lines: center line of current and never smokers median ß-value methylation; please see Table 13 for detailed data.

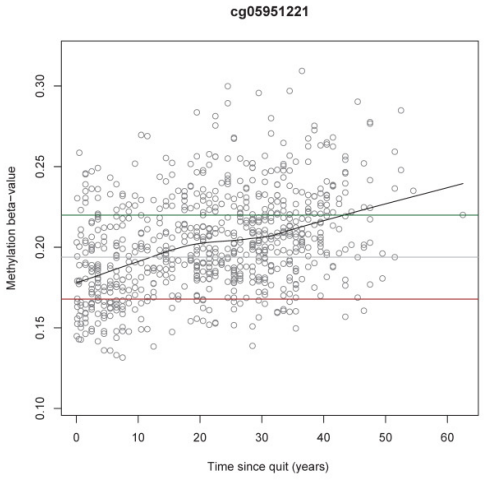
Figure A.3. Influence of time since smoking cessation on the DNA methylation state.

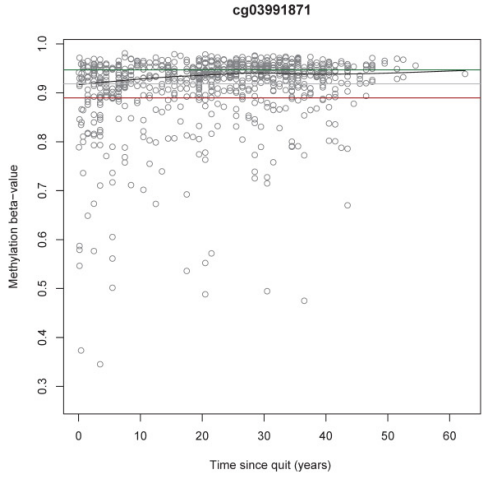


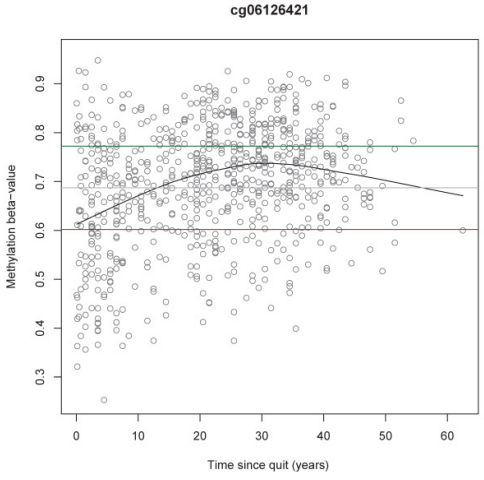

cg03604011
Methylation beta−value
Time since quit (years)

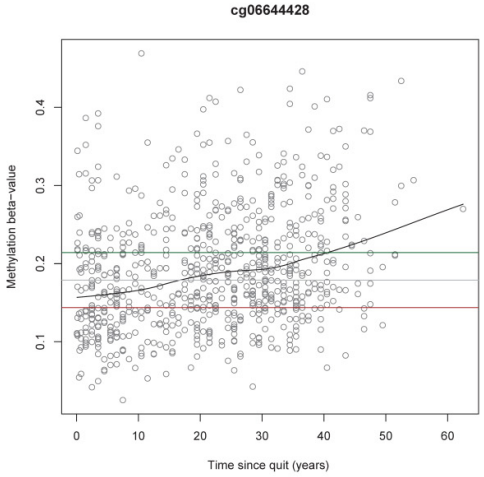

cg05575921
Methylation beta−value
Time since quit (years)

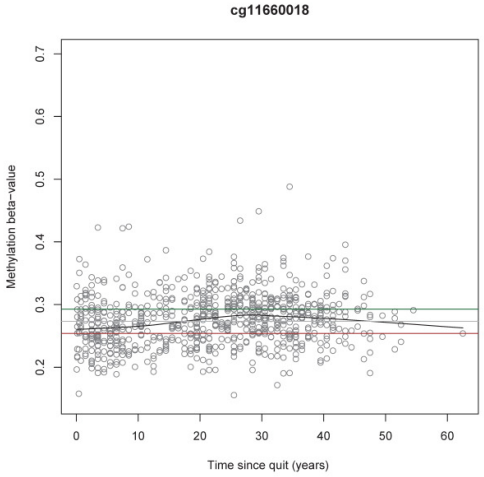

cg03636183
Methylation beta−value
Time since quit (years)

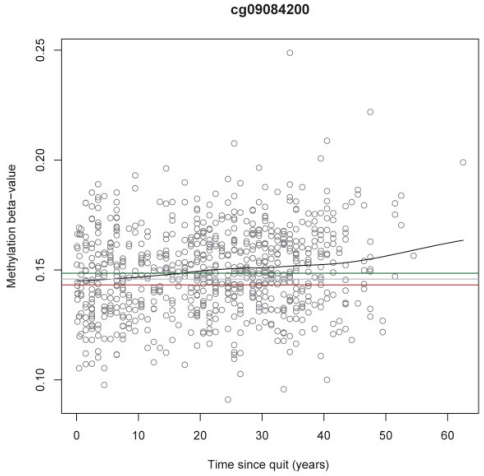

cg05951221
Methylation beta−value
Time since quit (years)

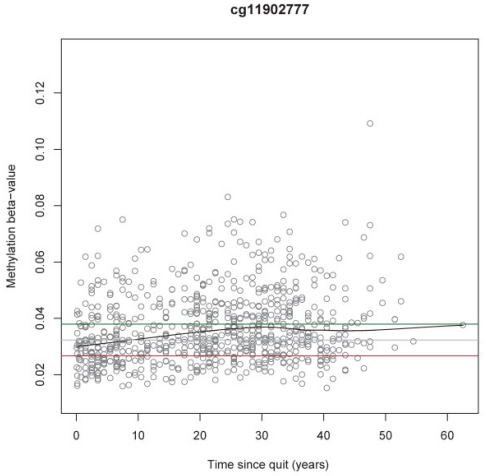

cg03991871
Methylation beta−value
Time since quit (years)

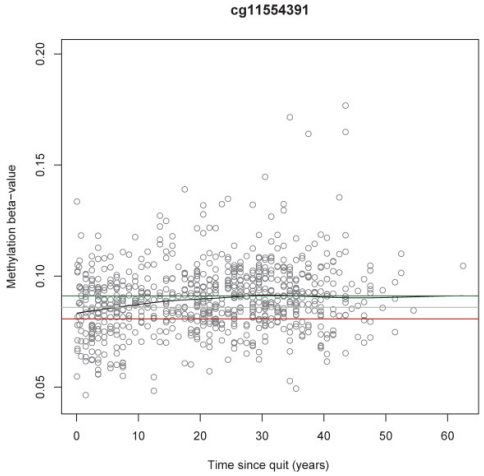

cg06126421
Methylation beta−value
Time since quit (years)

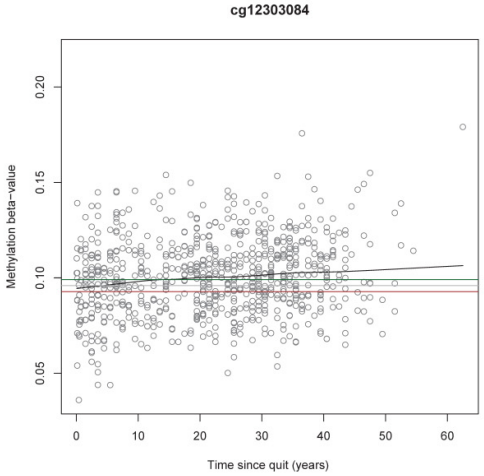

cg06644428
Methylation beta−value
Time since quit (years)

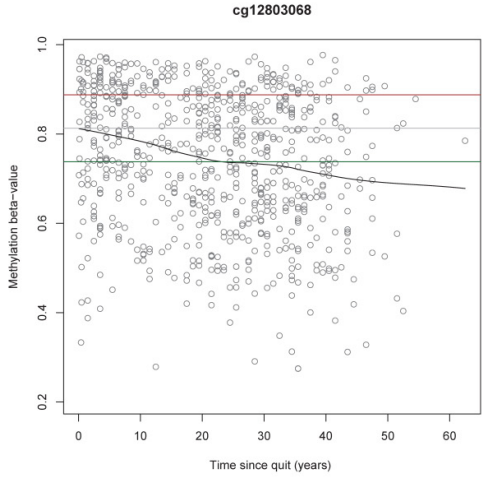

cg11660018
Methylation beta−value
Time since quit (years)

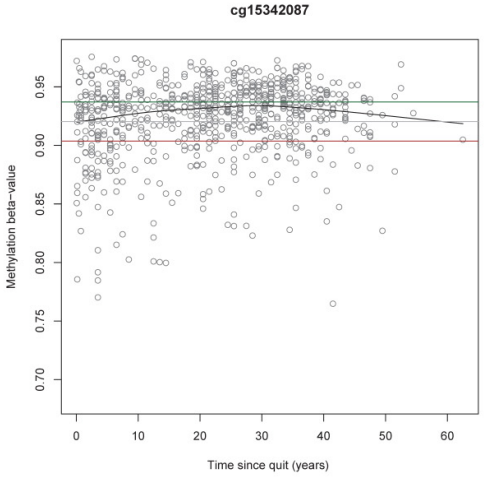

cg09084200
Methylation beta−value
Time since quit (years)

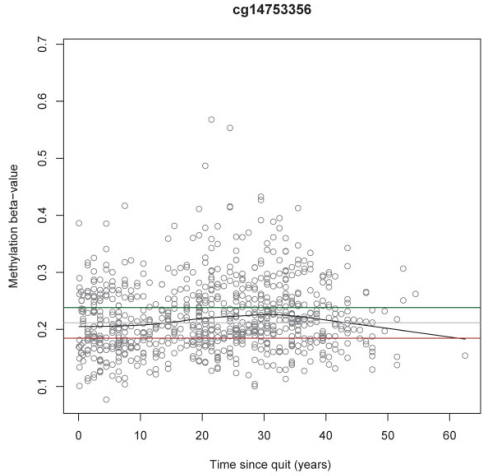

cg11902777
Methylation beta−value
Time since quit (years)

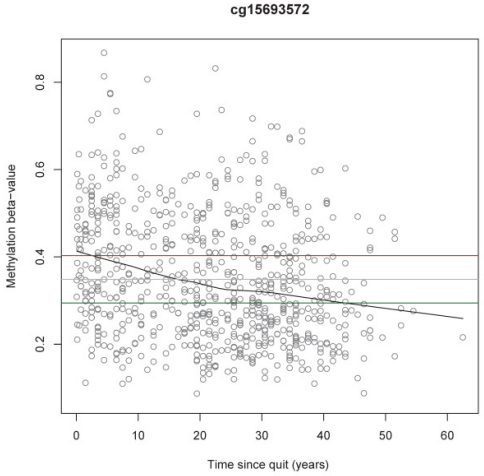

cg11554391
Methylation beta−value
Time since quit (years)

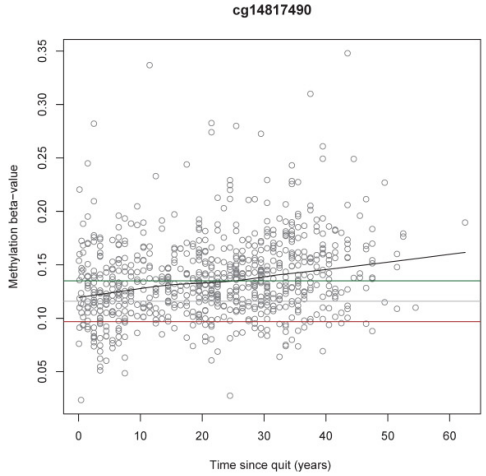

cg12303084
Methylation beta−value
Time since quit (years)

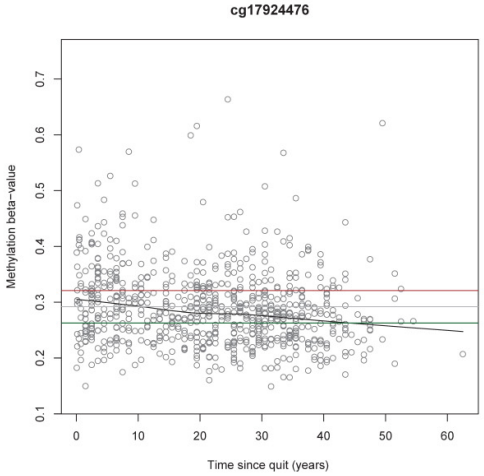

cg12803068
Methylation beta−value
Time since quit (years)

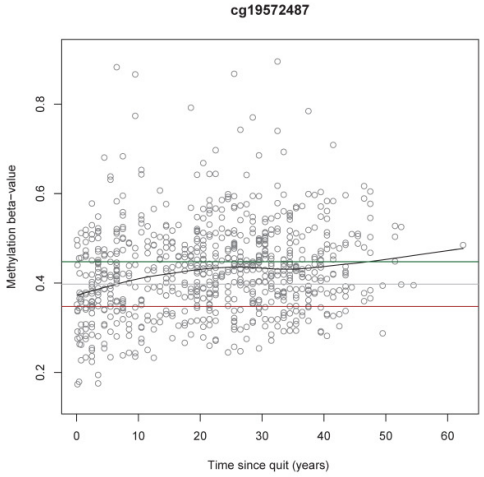

cg15342087
Methylation beta−value
Time since quit (years)

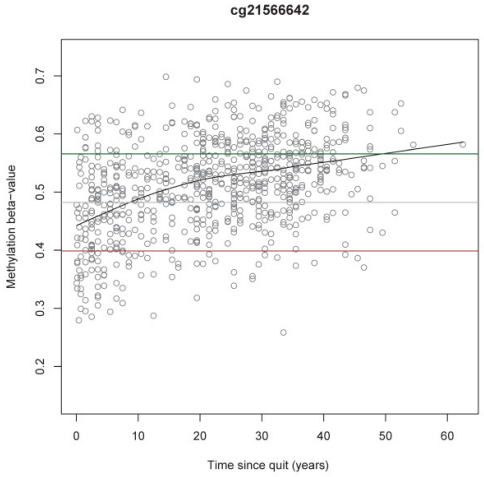

cg14753356
Methylation beta−value
Time since quit (years)

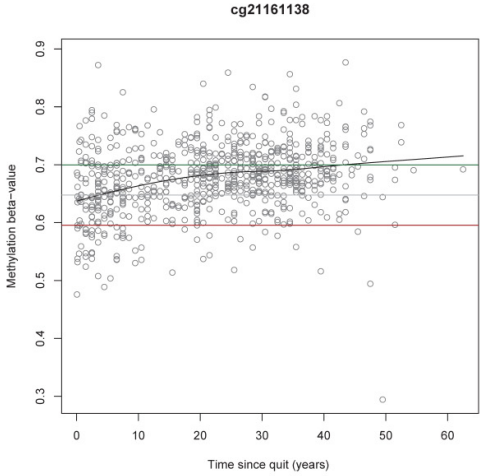

cg15693572
Methylation beta−value
Time since quit (years)

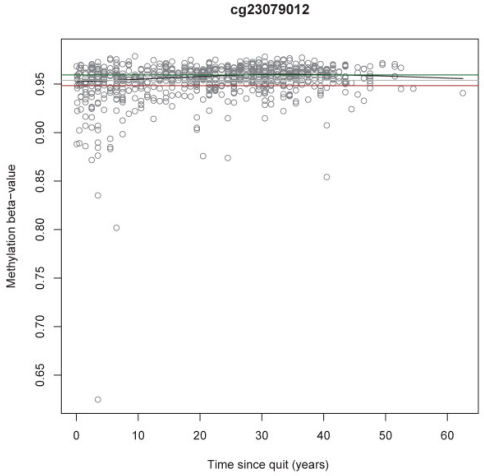

cg14817490
Methylation beta−value
Time since quit (years)

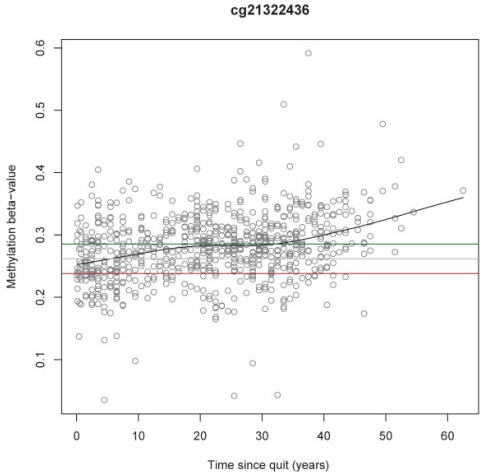

cg17924476
Methylation beta−value
Time since quit (years)

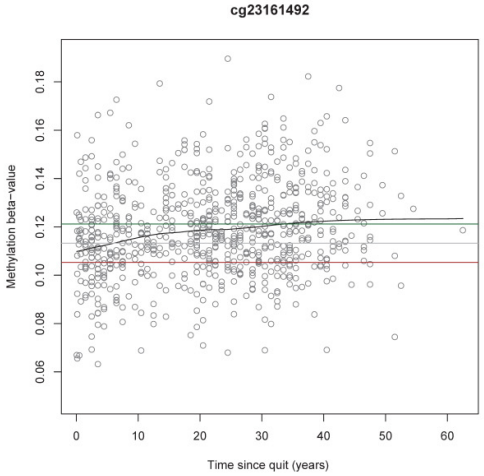

cg19572487
Methylation beta−value
Time since quit (years)

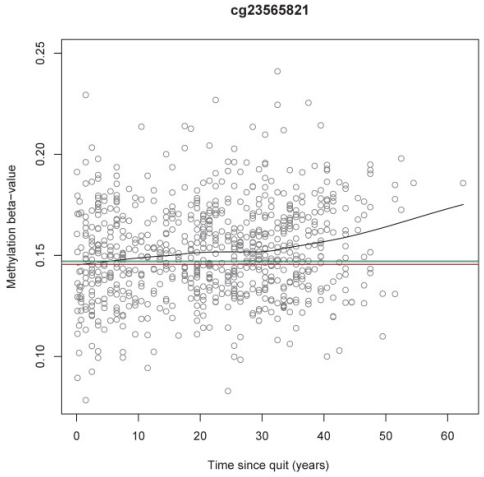

cg21566642
Methylation beta−value
Time since quit (years)

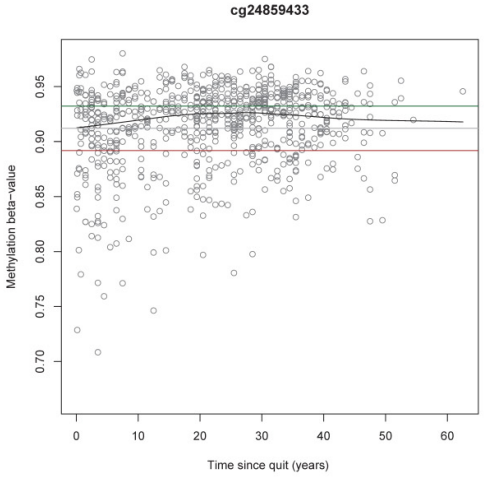

cg21161138
Methylation beta−value
Time since quit (years)

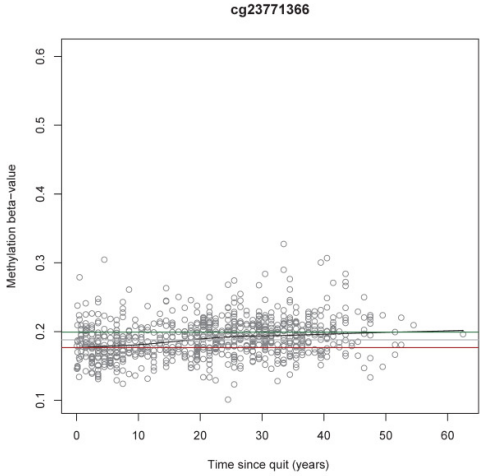

cg23079012
Methylation beta−value
Time since quit (years)

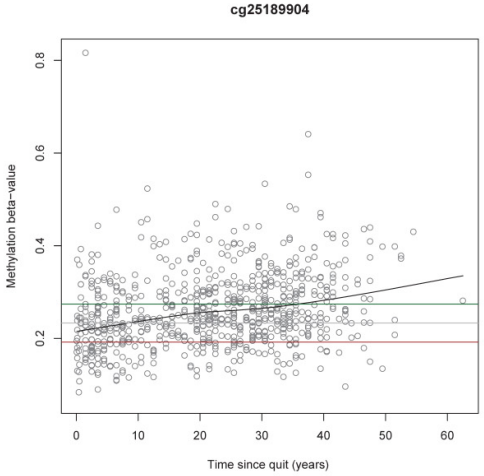

cg21322436
Methylation beta−value
Time since quit (years)

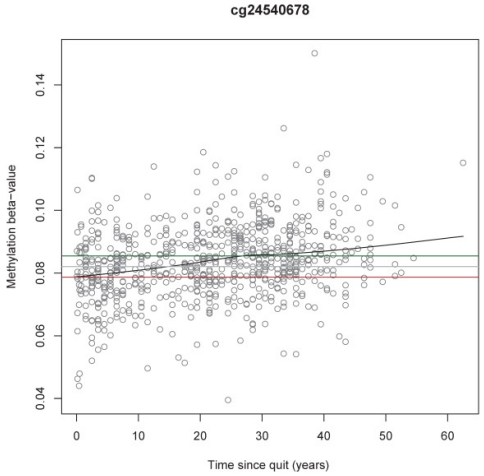

cg23161492
Methylation beta−value
Time since quit (years)


cg23565821
Methylation beta−value
Time since quit (years)


cg24859433
Methylation beta−value
Time since quit (years)


cg23771366
Methylation beta−value
Time since quit (years)


cg25189904
Methylation beta−value
Time since quit (years)


cg24540678
Methylation beta−value
Time since quit (years)

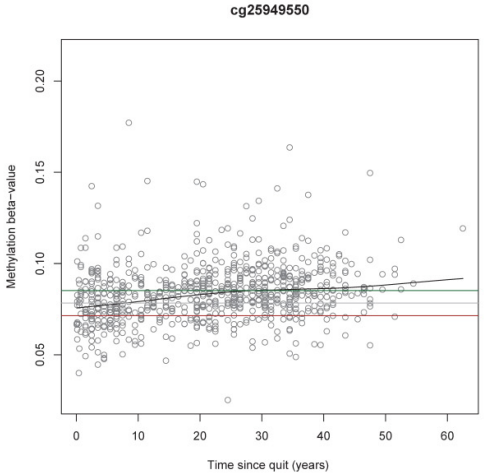
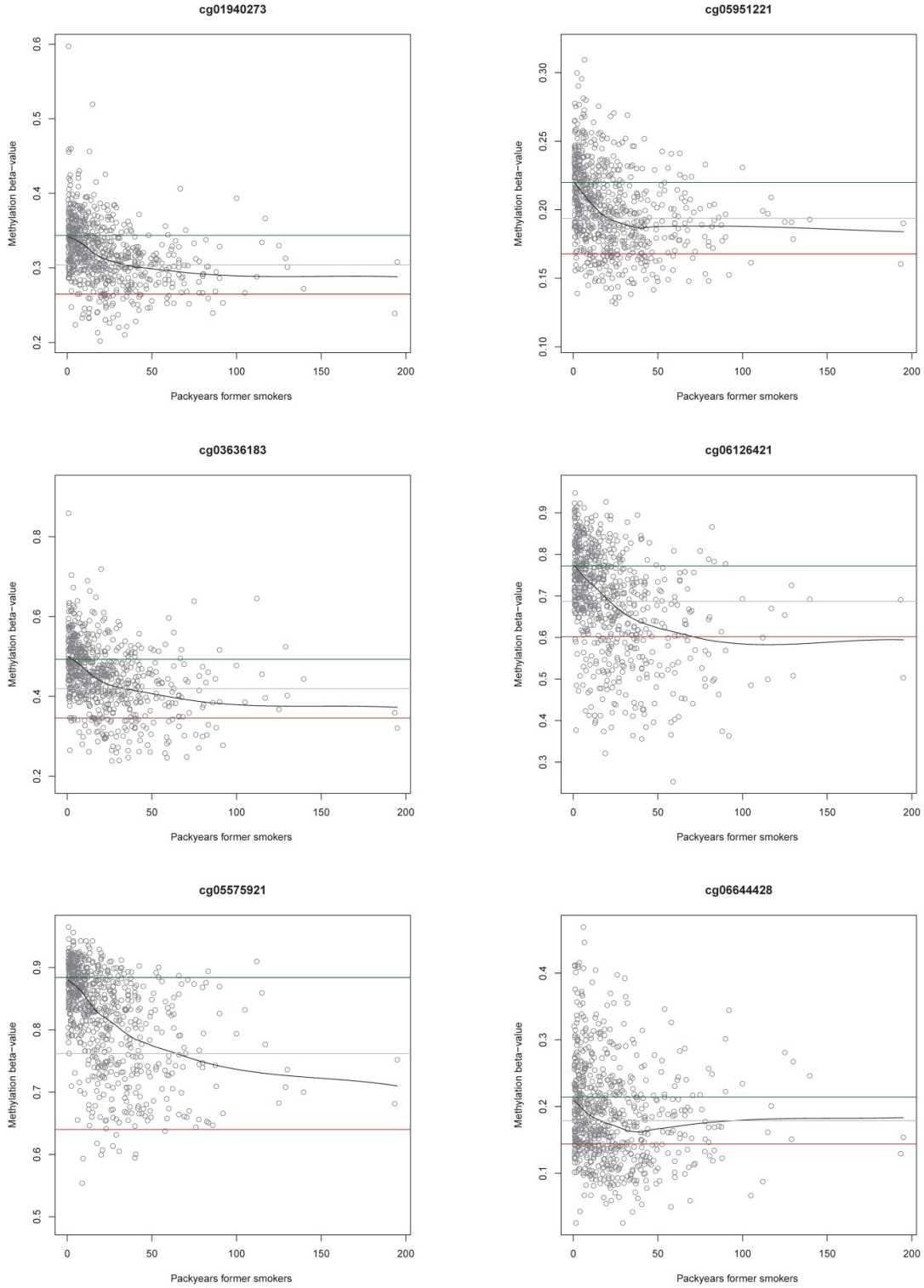
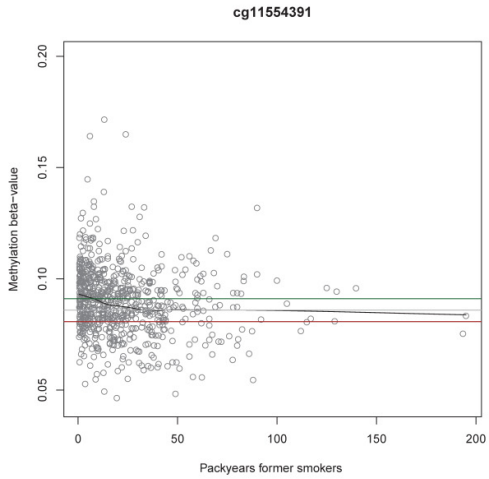
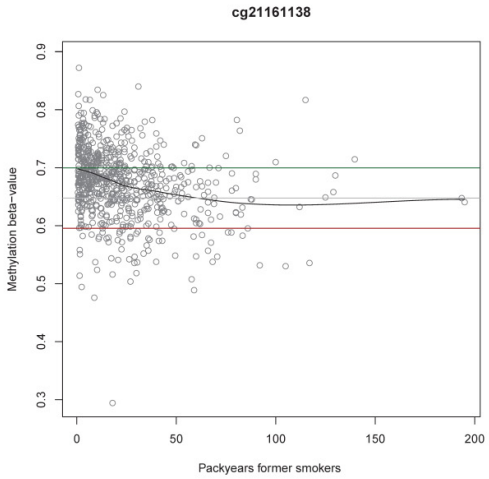

cg25949550
Methylation beta−value
Time since quit (years)

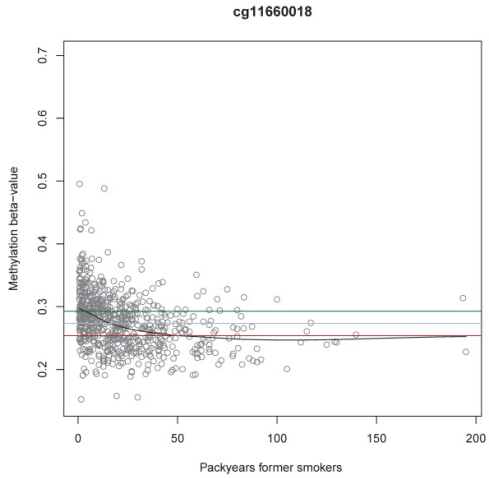
Figure A.4. Influence of pack-years on the DNA methylation state of former smokers.
The pack-years required for former smokers to obtain a median ß-value methylation state at single CpG sites that are closer to or equal the one of current smokers are displayed by loess curves in the scatterplots; the x-axis display the numbers of pack- years, the y-axis display the methylation levels with the use of numbers between 0 (for 0% methylation) and 1 (for 100% methylation); horizontal brown lines: median methylation of current smokers in %; horizontal green lines: median methylation of never smokers; horizontal grey lines: center line of current and never smokers median ß-value methylation; please see Table 14 for detailed data.

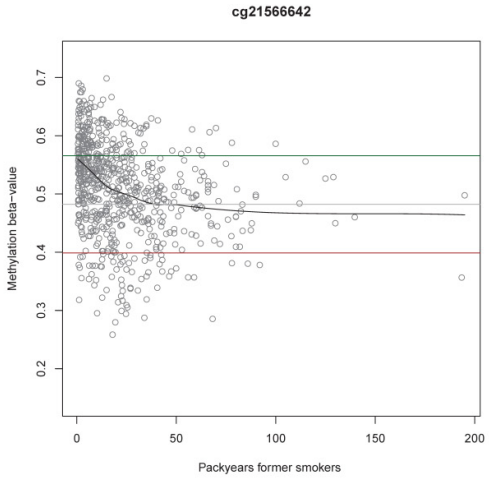
Figure A.4. Influence of pack-years on the DNA methylation state of former smokers.

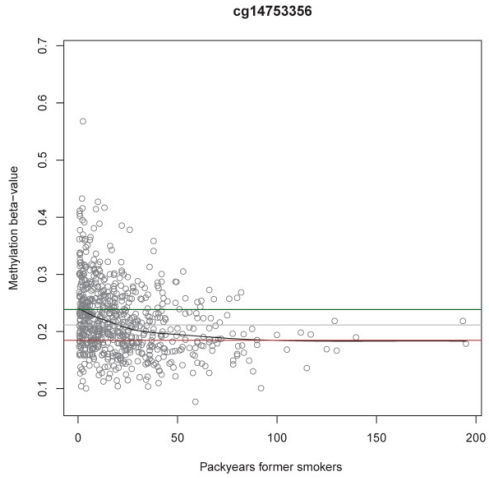


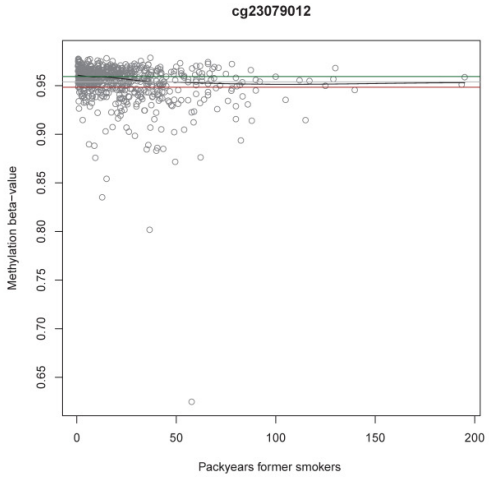

cg11554391
Methylation beta−value
Packyears former smokers

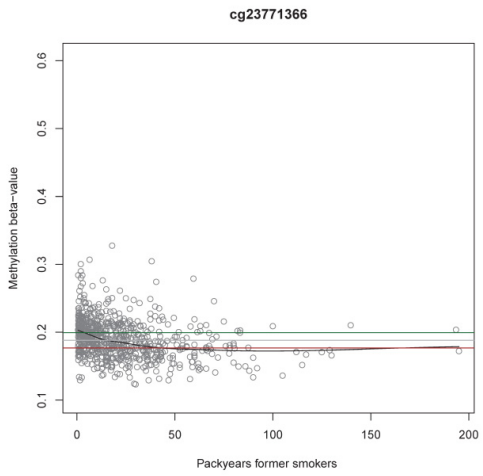

cg21161138
Methylation beta−value
Packyears former smokers


cg11660018
Methylation beta−value
Packyears former smokers


cg21566642
Methylation beta−value
Packyears former smokers


cg14753356
Methylation beta−value
Packyears former smokers


cg23079012
Methylation beta−value
Packyears former smokers


cg23771366
Methylation beta−value
Packyears former smokers
0.1
0.2
0.3
0.4
0.5
0.6
0
50
100
150
200

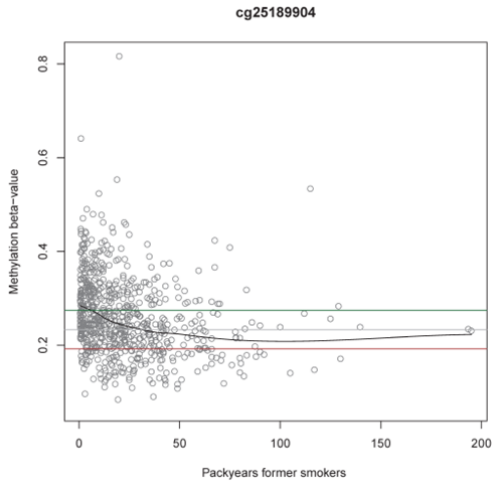

cg25189904
Methylation beta−value
Packyears former smokers
0.2
0.4
0.6
0.8
0
50
100
150
200

Table A.1. Differentially methylated current smoking-associated CpG sites of KORA F4.

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg11405655	1	x[a]	-0,50%	1,60E-10	0,16%	1,74E-01	0.1047 (0.09-0.12)	0.1063 (0.09-0.12)	0.0997 (0.09-0.11)
cg14531564	1	SDF4	-0,44%	2,83E-10	-0,02%	7,78E-01	0.7167 (0.69-0.74)	0.7165 (0.69-0.74)	0.7123 (0.68-0.74)
cg09276842	1	SDF4	-0,54%	5,18E-09	-0,05%	4,20E-01	0.9447 (0.93-0.96)	0.9441 (0.93-0.95)	0.9393 (0.92-0.96)
cg05818515	1	SDF4	3,01%	8,95E-10	-0,93%	2,02E-01	0.8373 (0.79-0.87)	0.8280 (0.78-0.87)	0.8674 (0.83-0.89)
cg27528222	1	C1orf86	-0,57%	1,41E-09	-0,07%	8,26E-01	0.9506 (0.94-0.96)	0.9499 (0.94-0.96)	0.9449 (0.93-0.96)
cg01979157	1	SKI	-1,79%	2,23E-12	-0,19%	6,42E-01	0.1476 (0.12-0.18)	0.1457 (0.12-0.18)	0.1297 (0.11-0.15)
cg05603985	1	SKI	-0,25%	4,80E-15	0,45%	3,64E-01	0.1457 (0.14-0.16)	0.1501 (0.14-0.16)	0.1432 (0.14-0.15)
cg09469355	1	SKI	-1,76%	3,34E-10	-0,61%	1,98E-01	0.3299 (0.30-0.36)	0.3238 (0.30-0.35)	0.3123 (0.28-0.34)
cg08884752	1	SKI	-4,01%	4,43E-13	-1,41%	2,22E-01	0.6803 (0.63-0.73)	0.6662 (0.62-0.72)	0.6401 (0.59-0.69)
cg15564619	1	SKI	-0,88%	7,03E-08	-0,36%	1,31E-02	0.5939 (0.56-0.63)	0.5902 (0.55-0.63)	0.5851 (0.54-0.62)
cg26687579	1	x[a]	-2,33%	1,01E-08	-0,57%	7,04E-01	0.2488 (0.21-0.28)	0.2431 (0.21-0.28)	0.2255 (0.19-0.27)
cg19825898	1	x[a]	-1,01%	4,81E-09	-0,22%	2,98E-01	0.8472 (0.82-0.88)	0.8451 (0.81-0.88)	0.8371 (0.81-0.87)
cg24150662	1	x[a]	-0,88%	5,18E-08	-0,37%	3,24E-01	0.8127 (0.78-0.85)	0.8091 (0.77-0.85)	0.8039 (0.76-0.84)
cg04134748	1	PRDM16	0,88%	4,75E-08	-0,71%	1,12E-01	0.7520 (0.70-0.80)	0.7449 (0.69-0.79)	0.7608 (0.71-0.82)
cg01309569	1	NPHP4	0,92%	7,65E-08	0,61%	1,03E-01	0.3718 (0.32-0.44)	0.3778 (0.33-0.44)	0.3810 (0.34-0.45)
cg01715088	1	ESPN	-0,31%	1,83E-09	0,15%	9,52E-01	0.0851 (0.08-0.09)	0.0866 (0.08-0.09)	0.0820 (0.07-0.09)
cg21012139	1	PLEKHG5	-1,57%	1,01E-08	0,05%	2,80E-01	0.6782 (0.64-0.72)	0.6787 (0.63-0.72)	0.6625 (0.63-0.70)
cg12547807	1	x[a]	-2,14%	6,52E-20	-0,33%	6,18E-02	0.1885 (0.16-0.22)	0.1852 (0.16-0.22)	0.1672 (0.15-0.19)
cg07793148	1	C1orf200/PIK3CD	-0,19%	9,61E-08	0,16%	7,40E-01	0.0783 (0.07-0.09)	0.0799 (0.07-0.09)	0.0763 (0.06-0.09)
cg23098018	1	PIK3CD	-0,61%	2,93E-08	0,67%	4,85E-01	0.7569 (0.73-0.79)	0.7637 (0.73-0.80)	0.7509 (0.72-0.78)
cg01320698	1	PIK3CD	-2,28%	1,06E-09	0,26%	3,28E-01	0.6230 (0.55-0.70)	0.6256 (0.55-0.70)	0.6002 (0.54-0.68)
cg07035320	1	CLSTN1	0,59%	4,22E-08	0,36%	5,42E-02	0.1357 (0.12-0.15)	0.1394 (0.13-0.15)	0.1416 (0.13-0.16)
cg04885881	1	x[a]	-7,41%	1,35E-28	-1,06%	1,42E-03	0.3115 (0.25-0.39)	0.3009 (0.23-0.37)	0.2374 (0.18-0.31)
cg21826784	1	AGTRAP	1,53%	1,67E-08	1,04%	5,35E-02	0.4048 (0.37-0.43)	0.4153 (0.39-0.44)	0.4201 (0.39-0.45)
cg19940644	1	x[a]	-2,43%	3,38E-13	-0,41%	5,86E-02	0.1633 (0.13-0.21)	0.1592 (0.13-0.20)	0.1390 (0.12-0.18)
cg21393163	1	x[a]	-0,69%	1,54E-20	-0,24%	1,71E-02	0.0445 (0.04-0.05)	0.0422 (0.04-0.05)	0.0376 (0.03-0.05)
cg09069072	1	TMEM51	-3,59%	2,95E-13	-0,66%	7,06E-02	0.8619 (0.83-0.89)	0.8553 (0.83-0.88)	0.8260 (0.79-0.87)
cg21913886	1	TMEM51	-3,38%	7,26E-11	-0,91%	1,82E-03	0.8917 (0.86-0.92)	0.8827 (0.84-0.91)	0.8579 (0.81-0.90)
cg04211179	1	ZBTB17	-0,89%	1,82E-09	0,18%	6,13E-01	0.2671 (0.24-0.30)	0.2688 (0.24-0.31)	0.2582 (0.23-0.29)
cg09012001	1	IFFO2	0,45%	7,50E-09	0,12%	8,72E-01	0.1252 (0.11-0.14)	0.1264 (0.11-0.14)	0.1297 (0.11-0.15)
cg17163404	1	UBR4	-1,35%	2,79E-08	-0,23%	8,42E-01	0.7926 (0.76-0.83)	0.7903 (0.76-0.83)	0.7791 (0.74-0.81)
cg19713429	1	CAPZB	-0,45%	1,90E-16	0,11%	6,94E-03	0.0987 (0.09-0.11)	0.0998 (0.09-0.11)	0.0942 (0.09-0.10)
cg01693305	1	CAPZB	-0,51%	2,47E-09	-0,13%	4,02E-01	0.9354 (0.92-0.95)	0.9341 (0.92-0.95)	0.9303 (0.91-0.94)
cg19276111	1	CNR2	3,95%	6,31E-09	2,86%	1,47E-01	0.4172 (0.37-0.47)	0.4458 (0.40-0.49)	0.4567 (0.40-0.50)
cg24452347	1	CNR2	4,12%	5,08E-11	2,60%	1,18E-02	0.2920 (0.25-0.34)	0.3180 (0.27-0.37)	0.3331 (0.28-0.38)
cg26404511	1	CNR2	2,25%	1,33E-13	1,06%	6,88E-03	0.4752 (0.44-0.50)	0.4858 (0.45-0.51)	0.4978 (0.47-0.52)
cg03611151	1	CNR2	1,59%	1,33E-09	0,84%	4,67E-02	0.4481 (0.42-0.47)	0.4566 (0.43-0.48)	0.4640 (0.44-0.49)
cg07967717	1	CNR2	1,43%	4,81E-10	0,59%	7,52E-04	0.2276 (0.20-0.25)	0.2335 (0.21-0.26)	0.2419 (0.22-0.27)
cg26856289	1	SFRS13A	-1,06%	1,51E-09	0,00%	8,16E-02	0.1674 (0.15-0.18)	0.1675 (0.15-0.18)	0.1569 (0.14-0.17)
cg04991639	1	x[a]	0,55%	4,27E-12	0,13%	3,40E-01	0.0598 (0.05-0.07)	0.0611 (0.05-0.07)	0.0654 (0.06-0.07)
cg15498134	1	RUNX3	-0,97%	4,31E-09	-0,11%	3,88E-01	0.7682 (0.73-0.80)	0.7671 (0.73-0.80)	0.7586 (0.71-0.79)
cg10951873	1	RUNX3	-0,24%	2,69E-09	-0,02%	7,99E-02	0.0452 (0.04-0.05)	0.0450 (0.04-0.05)	0.0428 (0.04-0.05)
cg27537125	1	x[a]	-1,21%	2,70E-30	-0,21%	1,69E-02	0.0877 (0.08-0.10)	0.0856 (0.08-0.10)	0.0756 (0.07-0.09)
cg22442430	1	LDLRAP1	-0,42%	9,58E-09	0,80%	3,86E-01	0.3961 (0.37-0.42)	0.4041 (0.38-0.43)	0.3920 (0.37-0.42)
cg03519967	1	MAN1C1	2,84%	1,84E-08	0,10%	5,51E-02	0.6937 (0.66-0.73)	0.6946 (0.66-0.74)	0.7221 (0.69-0.75)
cg16417374	1	MAN1C1	0,99%	4,30E-08	0,47%	3,44E-02	0.5584 (0.52-0.60)	0.5631 (0.52-0.60)	0.5683 (0.54-0.62)
cg26937868	1	x[a]	0,62%	4,73E-10	0,12%	9,42E-01	0.1122 (0.10-0.12)	0.1134 (0.10-0.12)	0.1184 (0.11-0.13)
cg24313303	1	PIGV	0,19%	6,58E-08	0,00%	6,41E-01	0.0401 (0.03-0.05)	0.0402 (0.03-0.05)	0.0421 (0.04-0.05)
cg22429121	1	AHDC1	1,01%	5,12E-08	1,12%	4,95E-03	0.2168 (0.20-0.23)	0.2279 (0.21-0.24)	0.2269 (0.21-0.24)
cg14510299	1	AHDC1	1,43%	4,15E-08	0,65%	8,53E-01	0.4403 (0.40-0.47)	0.4468 (0.41-0.48)	0.4546 (0.42-0.49)
cg12465678	1	FGR	1,04%	2,79E-08	0,11%	1,24E-01	0.1303 (0.11-0.15)	0.1314 (0.11-0.16)	0.1407 (0.12-0.17)
cg21115433	1	FGR	1,77%	7,73E-08	0,60%	5,24E-01	0.2723 (0.23-0.32)	0.2783 (0.23-0.33)	0.2900 (0.24-0.34)
cg23840027	1	EPB41	1,18%	5,57E-09	0,38%	8,49E-02	0.2783 (0.23-0.33)	0.2821 (0.24-0.33)	0.2901 (0.24-0.36)
cg15059804	1	ZNF362	4,63%	1,19E-09	0,19%	4,25E-02	0.2572 (0.21-0.32)	0.2592 (0.21-0.34)	0.3036 (0.23-0.38)
cg25722983	1	STK40	-1,84%	9,04E-13	-0,33%	5,26E-01	0.4696 (0.43-0.51)	0.4663 (0.43-0.50)	0.4512 (0.41-0.49)
cg01937809	1	ZC3H12A	-2,38%	1,76E-10	-0,03%	2,81E-01	0.3831 (0.34-0.42)	0.3828 (0.34-0.43)	0.3593 (0.32-0.41)
cg23093580	1	NDUFS5	2,50%	5,54E-09	-0,09%	3,53E-01	0.7273 (0.69-0.77)	0.7264 (0.69-0.77)	0.7524 (0.72-0.78)
cg24142464	1	PABPC4	-0,37%	7,24E-09	0,09%	2,09E-01	0.0729 (0.07-0.08)	0.0738 (0.07-0.08)	0.0691 (0.06-0.08)
cg16736826	1	EDN2	-0,86%	3,22E-08	0,12%	8,84E-01	0.3123 (0.28-0.34)	0.3135 (0.27-0.34)	0.3037 (0.27-0.33)
cg23762517	1	HIVEP3	6,97%	9,41E-08	2,30%	2,50E-01	0.2778 (0.21-0.37)	0.3009 (0.23-0.40)	0.3476 (0.28-0.47)
cg26038582	1	HIVEP3	9,75%	1,44E-08	3,41%	8,03E-02	0.4536 (0.34-0.56)	0.4877 (0.37-0.59)	0.5511 (0.44-0.68)
cg15937073	1	HIVEP3	3,16%	2,18E-08	1,07%	4,51E-02	0.1736 (0.14-0.21)	0.1843 (0.15-0.22)	0.2052 (0.17-0.25)
cg15542713	1	HIVEP3	9,52%	2,54E-08	1,87%	7,61E-01	0.4002 (0.31-0.50)	0.4189 (0.32-0.52)	0.4954 (0.38-0.60)
cg24049493	1	HIVEP3	4,05%	3,73E-14	1,38%	1,37E-01	0.1219 (0.10-0.15)	0.1357 (0.11-0.17)	0.1624 (0.12-0.21)
cg13985198	1	SNORD46/RPS8	-0,41%	6,47E-08	0,38%	5,39E-02	0.1276 (0.10-0.15)	0.1314 (0.11-0.16)	0.1235 (0.10-0.15)
cg05371506	1	TSPAN1	0,35%	3,55E-08	0,01%	9,79E-01	0.0932 (0.08-0.10)	0.0933 (0.08-0.11)	0.0967 (0.08-0.11)
cg23090529	1	x[a]	-2,28%	9,90E-11	-0,60%	6,11E-04	0.2061 (0.17-0.25)	0.2001 (0.17-0.23)	0.1834 (0.16-0.22)
cg21140898	1	x[a]	-3,28%	1,08E-10	-0,59%	8,18E-03	0.3534 (0.31-0.40)	0.3475 (0.31-0.40)	0.3206 (0.28-0.37)
cg24741609	1	GLIS1	-1,92%	1,39E-11	0,23%	3,86E-01	0.1978 (0.17-0.24)	0.2001 (0.17-0.24)	0.1786 (0.15-0.21)
cg19406367	1	SGIP1	3,11%	2,79E-08	1,00%	1,22E-01	0.8147 (0.76-0.86)	0.8247 (0.76-0.87)	0.8458 (0.80-0.89)
cg13184736	1	GNG12	-6,07%	1,32E-09	-1,93%	1,39E-03	0.4204 (0.36-0.48)	0.4011 (0.34-0.46)	0.3597 (0.29-0.43)
cg13399816	1	GNG12	-1,66%	2,73E-11	-0,25%	3,30E-03	0.1758 (0.16-0.20)	0.1733 (0.16-0.19)	0.1592 (0.14-0.18)
cg25189904*	1	GNG12	-8,19%	1,71E-39	-2,36%	6,85E-10	0.2740 (0.23-0.33)	0.2504 (0.21-0.31)	0.1921 (0.16-0.24)
cg15553397	1	ST6GALNAC3	2,15%	1,93E-08	0,70%	1,06E-01	0.6959 (0.67-0.73)	0.7028 (0.67-0.74)	0.7173 (0.68-0.76)
cg08848903	1	TGFBR3	1,46%	2,21E-08	1,02%	8,25E-02	0.2372 (0.21-0.27)	0.2474 (0.22-0.28)	0.2518 (0.23-0.28)
cg00563926	1	TGFBR3	0,75%	1,42E-10	0,31%	7,83E-02	0.1280 (0.11-0.14)	0.1310 (0.12-0.14)	0.1355 (0.12-0.15)
cg09662411	1	GFI1	-3,73%	8,01E-09	-0,99%	8,98E-03	0.4805 (0.44-0.52)	0.4706 (0.43-0.51)	0.4432 (0.39-0.50)
cg18146737	1	GFI1	-1,55%	4,79E-17	-0,29%	8,21E-03	0.9565 (0.95-0.96)	0.9536 (0.94-0.96)	0.9410 (0.88-0.96)
cg12876356	1	GFI1	-8,48%	2,92E-17	-1,62%	3,73E-03	0.5475 (0.49-0.62)	0.5313 (0.47-0.58)	0.4627 (0.37-0.55)
cg18316974	1	GFI1	-1,86%	1,60E-16	-0,29%	4,05E-03	0.9544 (0.94-0.96)	0.9514 (0.94-0.96)	0.9357 (0.90-0.96)
cg09935388	1	GFI1	-15,31%	3,27E-24	-3,79%	5,06E-07	0.5765 (0.49-0.70)	0.5385 (0.45-0.64)	0.4234 (0.32-0.54)
cg24049050	1	ARHGAP29	-0,38%	7,53E-11	0,08%	8,27E-02	0.1052 (0.10-0.11)	0.1060 (0.10-0.11)	0.1014 (0.10-0.11)
cg23616212	1	SORT1	1,06%	6,74E-08	0,40%	6,71E-01	0.1791 (0.16-0.20)	0.1830 (0.17-0.20)	0.1897 (0.17-0.21)
cg15995265	1	x[a]	-3,04%	1,38E-08	-0,08%	5,26E-01	0.2418 (0.20-0.30)	0.2411 (0.19-0.29)	0.2114 (0.17-0.27)
cg08737116	1	SLC22A15	-0,08%	1,79E-08	-0,16%	7,68E-03	0.0960 (0.08-0.11)	0.0945 (0.08-0.11)	0.0952 (0.08-0.11)
cg27449150	1	x[a]	-0,55%	3,76E-13	-0,06%	1,55E-01	0.1094 (0.10-0.13)	0.1088 (0.10-0.12)	0.1039 (0.09-0.12)
cg20607293	1	x[a]	-0,27%	9,71E-08	0,02%	2,77E-01	0.1136 (0.10-0.13)	0.1138 (0.10-0.13)	0.1109 (0.10-0.12)
cg03399898	1	x[a]	-0,68%	9,47E-08	-0,16%	3,92E-02	0.0816 (0.07-0.09)	0.0800 (0.07-0.09)	0.0748 (0.07-0.08)
cg02976588	1	x[a]	0,73%	6,22E-08	0,34%	5,97E-01	0.1551 (0.13-0.19)	0.1585 (0.13-0.19)	0.1624 (0.14-0.20)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg21937128	1	FAM63A	0,74%	2,66E-08	0,33%	8,05E-01	0.1346 (0.12-0.16)	0.1378 (0.12-0.16)	0.1420 (0.12-0.17)
cg09820519	1	C1orf230	0,34%	2,83E-08	0,08%	2,35E-01	0.0563 (0.05-0.06)	0.0570 (0.05-0.06)	0.0596 (0.05-0.06)
cg24158160	1	C1orf230	0,95%	3,86E-08	0,37%	4,12E-01	0.1240 (0.12-0.14)	0.1277 (0.12-0.14)	0.1335 (0.12-0.14)
cg13869942	1	THEM4	-0,72%	1,05E-09	0,05%	6,13E-01	0.1284 (0.11-0.16)	0.1289 (0.11-0.16)	0.1212 (0.10-0.14)
cg16139316	1	S100A9	1,20%	2,29E-08	0,30%	8,58E-01	0.1576 (0.13-0.19)	0.1606 (0.13-0.20)	0.1696 (0.14-0.21)
cg08056629	1	x[a]	-0,95%	4,72E-09	0,04%	2,48E-01	0.8501 (0.81-0.88)	0.8505 (0.81-0.89)	0.8406 (0.80-0.88)
cg00867472	1	HDGF	1,17%	5,74E-08	-0,29%	2,11E-01	0.6068 (0.57-0.64)	0.6039 (0.57-0.64)	0.6186 (0.58-0.65)
cg11231349	1	NOS1AP	-1,92%	3,94E-09	-1,13%	7,23E-05	0.8372 (0.79-0.87)	0.8259 (0.77-0.87)	0.8180 (0.77-0.85)
cg07807219	1	x[a]	2,40%	4,24E-08	1,00%	2,91E-01	0.8121 (0.78-0.85)	0.8221 (0.79-0.85)	0.8361 (0.80-0.87)
cg24304617	1	ATP1B1	0,35%	4,40E-10	0,01%	2,76E-01	0.0591 (0.05-0.07)	0.0592 (0.05-0.07)	0.0626 (0.05-0.07)
cg16290996	1	GAS5/SNORD78	-0,58%	8,62E-11	0,09%	5,43E-02	0.1202 (0.11-0.13)	0.1210 (0.11-0.13)	0.1144 (0.10-0.13)
cg17025683	1	SNORD78/ZBTB37	-0,27%	7,30E-08	0,20%	1,41E-01	0.0793 (0.07-0.09)	0.0813 (0.07-0.09)	0.0765 (0.07-0.09)
cg20240091	1	TNN	1,68%	4,54E-08	0,11%	1,27E-01	0.6831 (0.64-0.72)	0.6843 (0.64-0.72)	0.6999 (0.65-0.74)
cg23036452	1	ATP2B4	1,15%	1,35E-08	0,01%	1,61E-01	0.6948 (0.66-0.72)	0.6949 (0.66-0.73)	0.7062 (0.67-0.74)
cg08709672	1	AVPR1B	-2,58%	1,93E-20	-0,11%	5,56E-01	0.6776 (0.66-0.70)	0.6765 (0.66-0.70)	0.6518 (0.64-0.67)
cg20295214	1	AVPR1B	-2,98%	1,13E-09	-0,25%	9,24E-01	0.8467 (0.82-0.87)	0.8442 (0.81-0.87)	0.8169 (0.78-0.85)
cg06872036	1	RASSF5	-1,18%	8,12E-09	-0,19%	1,56E-01	0.9032 (0.88-0.92)	0.9013 (0.88-0.92)	0.8915 (0.87-0.91)
cg17220933	1	RASSF5	-0,59%	6,73E-09	0,05%	2,63E-01	0.7661 (0.73-0.80)	0.7666 (0.73-0.80)	0.7602 (0.72-0.79)
cg22049321	1	CR1	0,19%	4,19E-08	0,21%	3,22E-01	0.0991 (0.09-0.11)	0.1012 (0.09-0.11)	0.1009 (0.09-0.11)
cg00541303	1	x[a]	-2,84%	1,33E-12	0,79%	3,40E-01	0.2530 (0.21-0.31)	0.2609 (0.21-0.31)	0.2246 (0.19-0.27)
cg05376738	1	NSL1	-1,28%	3,66E-11	0,34%	9,48E-01	0.1803 (0.16-0.21)	0.1837 (0.16-0.22)	0.1674 (0.14-0.19)
cg03547355	1	x[a]	-2,76%	1,06E-12	-1,16%	9,06E-04	0.6024 (0.57-0.63)	0.5909 (0.55-0.63)	0.5748 (0.53-0.61)
cg26314722	1	x[a]	-0,92%	1,42E-09	-0,25%	9,40E-02	0.1358 (0.12-0.16)	0.1333 (0.12-0.16)	0.1266 (0.11-0.15)
cg17819085	1	x[a]	-2,30%	6,87E-08	-0,53%	7,44E-01	0.7041 (0.67-0.73)	0.6988 (0.67-0.73)	0.6811 (0.65-0.71)
cg04311403	1	EDARADD	-1,93%	2,12E-08	-0,75%	2,22E-01	0.5380 (0.48-0.60)	0.5305 (0.47-0.60)	0.5187 (0.45-0.59)
cg23166773	1	SDCCAG8	2,11%	4,44E-09	0,20%	1,99E-01	0.3061 (0.26-0.36)	0.3081 (0.26-0.36)	0.3273 (0.28-0.39)
cg23079012	2	x[a]	-1,10%	8,29E-24	-0,14%	2,81E-03	0.9594 (0.95-0.96)	0.9580 (0.95-0.96)	0.9484 (0.93-0.96)
cg08035323	2	x[a]	4,09%	7,35E-14	-0,15%	9,10E-05	0.1565 (0.12-0.20)	0.1551 (0.12-0.21)	0.1974 (0.14-0.25)
cg03959186	2	x[a]	-1,13%	2,19E-09	-0,36%	1,94E-01	0.1029 (0.08-0.13)	0.0993 (0.08-0.12)	0.0916 (0.07-0.12)
cg21356710	2	MFSD2B	2,30%	7,24E-11	0,37%	8,37E-02	0.5941 (0.56-0.63)	0.5978 (0.56-0.63)	0.6171 (0.58-0.65)
cg00483217	2	EFR3B	1,06%	4,26E-08	0,87%	1,43E-02	0.2948 (0.25-0.35)	0.3035 (0.26-0.35)	0.3054 (0.27-0.37)
cg23401251	2	IFT172	1,82%	2,29E-08	-0,12%	2,70E-01	0.5469 (0.49-0.60)	0.5457 (0.49-0.60)	0.5651 (0.52-0.62)
cg06142740	2	PPP1CB	-0,05%	3,00E-09	0,18%	1,53E-01	0.1124 (0.10-0.12)	0.1142 (0.11-0.12)	0.1119 (0.10-0.12)
cg04268813	2	x[a]	1,72%	6,09E-08	0,12%	2,35E-01	0.7208 (0.68-0.76)	0.7219 (0.68-0.76)	0.7379 (0.70-0.78)
cg19427338	2	x[a]	1,92%	4,37E-09	0,01%	7,52E-01	0.7504 (0.72-0.78)	0.7505 (0.72-0.78)	0.7696 (0.74-0.80)
cg23008718	2	x[a]	0,67%	7,79E-08	-0,04%	6,71E-01	0.1522 (0.13-0.17)	0.1518 (0.13-0.17)	0.1589 (0.14-0.18)
cg11253886	2	x[a]	3,53%	1,70E-11	0,43%	2,26E-03	0.6195 (0.58-0.67)	0.6238 (0.58-0.67)	0.6548 (0.61-0.70)
cg06949933	2	AHSA2	-1,45%	2,30E-09	0,43%	1,89E-01	0.8465 (0.80-0.88)	0.8508 (0.81-0.89)	0.8320 (0.78-0.88)
cg22001211	2	XPO1	-0,24%	6,13E-08	0,13%	6,43E-01	0.0868 (0.08-0.09)	0.0881 (0.08-0.10)	0.0844 (0.08-0.09)
cg14209730	2	x[a]	1,79%	6,20E-08	-0,72%	2,06E-01	0.4711 (0.39-0.55)	0.4639 (0.39-0.55)	0.4890 (0.42-0.57)
cg03233656	2	SLC1A4	-0,80%	1,26E-08	0,58%	9,02E-01	0.1470 (0.13-0.17)	0.1528 (0.13-0.17)	0.1389 (0.12-0.16)
cg13060970	2	PLEK	1,47%	2,35E-08	1,12%	7,59E-02	0.2932 (0.25-0.33)	0.3044 (0.26-0.34)	0.3079 (0.28-0.34)
cg04872689	2	PLEK	0,43%	9,52E-08	0,02%	7,18E-01	0.0674 (0.06-0.08)	0.0676 (0.06-0.08)	0.0717 (0.06-0.08)
cg10812236	2	PLEK	0,59%	1,59E-08	0,25%	1,91E-01	0.0893 (0.08-0.10)	0.0918 (0.08-0.11)	0.0952 (0.08-0.11)
cg06635952	2	ANXA4	2,00%	2,37E-21	0,42%	1,07E-03	0.1388 (0.12-0.16)	0.1430 (0.13-0.16)	0.1588 (0.14-0.18)
cg02039404	2	x[a]	-3,53%	6,74E-08	-0,69%	2,89E-01	0.7783 (0.71-0.83)	0.7714 (0.70-0.82)	0.7430 (0.68-0.80)
cg15408734	2	x[a]	-0,27%	2,49E-12	0,03%	1,43E-01	0.0583 (0.05-0.07)	0.0586 (0.05-0.07)	0.0556 (0.05-0.06)
cg23230929	2	x[a]	3,52%	9,48E-11	0,81%	4,63E-02	0.3567 (0.32-0.40)	0.3648 (0.32-0.41)	0.3919 (0.36-0.43)
cg03147185	2	NCAPH	-1,16%	3,44E-08	-0,33%	8,78E-01	0.4391 (0.39-0.48)	0.4358 (0.39-0.48)	0.4275 (0.38-0.47)
cg02592133	2	C2orf55	0,31%	4,60E-09	-0,01%	9,04E-01	0.1035 (0.09-0.12)	0.1034 (0.09-0.12)	0.1066 (0.09-0.12)
cg13522882	2	MAP4K4	-1,17%	8,31E-08	-0,05%	1,72E-01	0.7839 (0.74-0.83)	0.7834 (0.73-0.83)	0.7722 (0.72-0.82)
cg13399261	2	x[a]	0,66%	4,91E-08	0,09%	9,57E-01	0.1229 (0.11-0.14)	0.1238 (0.11-0.14)	0.1295 (0.11-0.15)
cg06880612	2	x[a]	0,64%	1,99E-08	0,27%	6,31E-01	0.1126 (0.10-0.12)	0.1153 (0.10-0.13)	0.1190 (0.11-0.13)
cg27598107	2	x[a]	2,38%	1,23E-10	0,72%	3,31E-02	0.2133 (0.18-0.25)	0.2205 (0.19-0.26)	0.2371 (0.20-0.28)
cg00295485	2	UXS1	-4,47%	2,65E-12	-0,83%	2,80E-04	0.4991 (0.43-0.56)	0.4908 (0.41-0.55)	0.4544 (0.39-0.51)
cg09570614	2	x[a]	3,43%	1,16E-11	0,83%	1,81E-02	0.3590 (0.32-0.40)	0.3674 (0.33-0.41)	0.3934 (0.36-0.44)
cg05616969	2	x[a]	0,44%	6,62E-08	0,18%	3,81E-01	0.1261 (0.11-0.14)	0.1279 (0.11-0.15)	0.1306 (0.11-0.15)
cg18859174	2	x[a]	0,51%	6,87E-10	0,21%	2,98E-01	0.1315 (0.12-0.14)	0.1336 (0.12-0.15)	0.1366 (0.12-0.15)
cg08129583	2	GYPC	0,72%	6,50E-09	0,20%	6,75E-01	0.1512 (0.13-0.17)	0.1532 (0.14-0.17)	0.1584 (0.14-0.18)
cg04019522	2	BIN1	3,73%	7,66E-09	-0,48%	6,76E-01	0.4664 (0.40-0.52)	0.4616 (0.40-0.52)	0.5037 (0.45-0.55)
cg19375644	2	x[a]	0,60%	2,00E-08	0,12%	1,33E-01	0.1005 (0.09-0.11)	0.1017 (0.09-0.11)	0.1065 (0.10-0.11)
cg11014740	2	x[a]	-0,28%	5,25E-09	0,02%	5,56E-01	0.1750 (0.16-0.19)	0.1751 (0.16-0.19)	0.1722 (0.16-0.18)
cg06240690	2	GALNT5	0,37%	5,80E-10	0,20%	6,00E-01	0.0698 (0.06-0.08)	0.0718 (0.06-0.08)	0.0735 (0.06-0.08)
cg15352315	2	CD302	0,67%	1,94E-08	0,30%	2,96E-01	0.1216 (0.11-0.14)	0.1246 (0.11-0.14)	0.1283 (0.11-0.14)
cg04737150	2	x[a]	-1,44%	3,41E-08	-0,01%	9,15E-01	0.8125 (0.77-0.85)	0.8124 (0.77-0.85)	0.7981 (0.76-0.84)
cg17825400	2	WIPF1	-1,08%	1,40E-09	-0,12%	1,65E-01	0.4399 (0.39-0.49)	0.4387 (0.39-0.49)	0.4291 (0.37-0.47)
cg26271591	2	NFE2L2	-5,57%	1,75E-14	-0,90%	7,07E-02	0.2569 (0.19-0.32)	0.2479 (0.19-0.32)	0.2012 (0.15-0.26)
cg04361126	2	SLC40A1	0,50%	6,67E-08	0,29%	2,87E-01	0.1287 (0.12-0.14)	0.1316 (0.12-0.14)	0.1337 (0.12-0.15)
cg13651908	2	CD28	-1,07%	2,62E-08	0,25%	2,37E-01	0.8018 (0.76-0.84)	0.8043 (0.76-0.84)	0.7910 (0.76-0.83)
cg24715680	2	BARD1	2,28%	5,35E-08	0,12%	7,09E-01	0.2613 (0.22-0.30)	0.2625 (0.22-0.30)	0.2841 (0.24-0.33)
cg25941354	2	CXCR2	1,13%	4,37E-10	-0,06%	4,00E-01	0.1456 (0.13-0.17)	0.1450 (0.12-0.17)	0.1568 (0.14-0.18)
cg20456243	2	SPEG	-1,53%	7,86E-08	-0,94%	9,29E-01	0.5379 (0.49-0.57)	0.5285 (0.48-0.57)	0.5226 (0.47-0.56)
cg05940122	2	x[a]	-1,27%	1,71E-09	-0,05%	2,84E-02	0.7837 (0.73-0.83)	0.7832 (0.73-0.83)	0.7710 (0.71-0.81)
cg19827923	2	GPR55	-2,03%	2,69E-13	-0,31%	7,08E-01	0.8675 (0.84-0.89)	0.8644 (0.84-0.89)	0.8472 (0.82-0.87)
cg03164561	2	NMUR1	2,64%	8,26E-08	0,17%	1,75E-01	0.6905 (0.65-0.74)	0.6922 (0.65-0.73)	0.7169 (0.68-0.76)
cg23667432	2	ALPP	-1,57%	1,55E-11	0,12%	4,17E-01	0.3401 (0.32-0.36)	0.3413 (0.32-0.36)	0.3244 (0.31-0.35)
cg03188382	2	ALPP	-1,19%	6,73E-11	-0,38%	8,78E-02	0.2432 (0.23-0.26)	0.2394 (0.22-0.26)	0.2313 (0.21-0.25)
cg22403782	2	ALPP	-0,97%	4,25E-10	-0,22%	7,90E-04	0.1665 (0.15-0.19)	0.1643 (0.15-0.18)	0.1569 (0.14-0.17)
cg19713851	2	ALPP	-6,47%	1,28E-12	-2,89%	5,88E-05	0.4950 (0.43-0.56)	0.4661 (0.40-0.54)	0.4302 (0.37-0.50)
cg27241845	2	x[a]	-8,74%	2,79E-35	-3,43%	2,04E-05	0.4617 (0.41-0.52)	0.4274 (0.37-0.49)	0.3743 (0.31-0.42)
cg25560398	2	ECEL1P2	-1,44%	5,75E-12	-0,31%	4,09E-03	0.7081 (0.69-0.73)	0.7051 (0.69-0.72)	0.6938 (0.68-0.71)
cg17087741	2	ALPPL2[b]	-10,53%	3,06E-09	-2,88%	9,67E-02	0.6607 (0.55-0.78)	0.6320 (0.53-0.75)	0.5555 (0.45-0.68)
cg03329539*	2	ALPPL2[b]	-2,39%	3,66E-59	-0,61%	7,58E-13	0.1598 (0.15-0.17)	0.1537 (0.14-0.17)	0.1359 (0.13-0.15)
cg06644428*	2	ALPPL2[b]	-7,05%	6,37E-31	-3,09%	1,20E-23	0.2142 (0.17-0.27)	0.1833 (0.14-0.23)	0.1437 (0.11-0.19)
cg05951221*	2	ALPPL2[b]	-5,21%	8,92E-104	-1,87%	8,18E-39	0.2199 (0.20-0.24)	0.2012 (0.18-0.22)	0.1678 (0.15-0.19)
cg21566642*	2	ALPPL2[b]	-16,70%	6,90E-138	-4,31%	9,49E-40	0.5658 (0.51-0.61)	0.5227 (0.46-0.57)	0.3988 (0.35-0.45)
cg01940273*	2	ALPPL2[b]	-7,89%	9,28E-114	-2,26%	2,01E-22	0.3437 (0.32-0.37)	0.3211 (0.30-0.35)	0.2648 (0.24-0.29)
cg13193840	2	ALPPL2[b]	-2,41%	1,53E-23	-0,37%	1,48E-07	0.1525 (0.13-0.17)	0.1488 (0.13-0.17)	0.1285 (0.11-0.15)
cg05208178	2	INPP5D	-0,32%	5,81E-08	0,01%	3,50E-01	0.1016 (0.09-0.12)	0.1017 (0.09-0.12)	0.0984 (0.09-0.11)
cg25488284	2	PASK	-1,26%	2,51E-09	-0,41%	7,57E-02	0.9108 (0.88-0.93)	0.9067 (0.88-0.93)	0.8982 (0.87-0.92)
cg19450479	3	x[a]	0,27%	1,46E-08	0,03%	7,63E-01	0.0397 (0.03-0.05)	0.0400 (0.03-0.05)	0.0424 (0.04-0.05)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg14766620	3	x[a]	0,56%	2,08E-09	0,16%	9,87E-02	0.0797 (0.07-0.09)	0.0813 (0.07-0.09)	0.0853 (0.08-0.09)
cg00085013	3	FGD5	3,20%	3,64E-08	1,17%	2,68E-02	0.2443 (0.22-0.28)	0.2560 (0.23-0.29)	0.2763 (0.24-0.31)
cg02930866	3	EAF1	0,51%	3,67E-08	-0,10%	7,86E-01	0.0939 (0.08-0.11)	0.0929 (0.08-0.11)	0.0990 (0.09-0.11)
cg17024919	3	ZNF385D	-3,09%	4,06E-09	-0,53%	6,63E-02	0.1708 (0.13-0.25)	0.1654 (0.12-0.25)	0.1398 (0.10-0.21)
cg15693572	3	x[a]	10,87%	3,84E-14	3,10%	1,03E-03	0.2942 (0.22-0.41)	0.3253 (0.24-0.45)	0.4029 (0.29-0.53)
cg23480021	3	x[a]	18,63%	2,67E-22	4,33%	4,33E-05	0.4259 (0.30-0.60)	0.4692 (0.32-0.63)	0.6122 (0.44-0.78)
cg03274391	3	x[a]	18,06%	1,67E-25	5,42%	2,56E-05	0.3190 (0.23-0.47)	0.3732 (0.25-0.52)	0.4997 (0.34-0.70)
cg03540589	3	CSRNP1	-3,15%	6,74E-11	0,24%	2,78E-01	0.7917 (0.75-0.83)	0.7941 (0.75-0.83)	0.7602 (0.71-0.81)
cg00501876	3	CSRNP1	-4,43%	1,39E-25	-0,56%	4,62E-01	0.3692 (0.33-0.40)	0.3636 (0.33-0.40)	0.3249 (0.30-0.36)
cg10911980	3	C3orf39	-0,78%	3,27E-09	0,42%	1,21E-02	0.5283 (0.50-0.55)	0.5324 (0.50-0.56)	0.5204 (0.50-0.55)
cg03194226	3	CLEC3B	-0,80%	1,13E-08	-0,36%	1,34E-02	0.9379 (0.92-0.95)	0.9343 (0.92-0.95)	0.9299 (0.91-0.94)
cg18642234	3	GPX1	-1,62%	1,44E-12	-0,15%	1,29E-02	0.2316 (0.21-0.25)	0.2300 (0.21-0.25)	0.2154 (0.20-0.23)
cg25100871	3	x[a]	-0,95%	8,96E-08	-0,37%	2,45E-02	0.2117 (0.18-0.24)	0.2080 (0.18-0.24)	0.2022 (0.18-0.23)
cg03929796	3	ALAS1	-0,19%	3,81E-12	0,07%	3,16E-03	0.0884 (0.08-0.10)	0.0891 (0.08-0.10)	0.0865 (0.08-0.10)
cg07757252	3	TWF2	-1,19%	6,12E-09	-0,07%	2,73E-01	0.8489 (0.82-0.87)	0.8482 (0.82-0.87)	0.8370 (0.80-0.87)
cg15417641	3	CACNA1D	7,92%	1,25E-16	0,37%	8,95E-02	0.6818 (0.58-0.76)	0.6855 (0.59-0.77)	0.7610 (0.68-0.83)
cg00336149	3	CACNA1D	6,46%	1,57E-12	0,88%	5,85E-02	0.2530 (0.20-0.32)	0.2617 (0.21-0.32)	0.3175 (0.25-0.39)
cg21188533	3	CACNA1D	10,18%	1,12E-10	0,92%	3,21E-01	0.5724 (0.45-0.69)	0.5816 (0.46-0.68)	0.6742 (0.55-0.76)
cg08052292	3	ARHGEF3	0,43%	3,88E-10	0,11%	3,43E-01	0.0519 (0.04-0.06)	0.0530 (0.05-0.06)	0.0562 (0.05-0.07)
cg00505318	3	UBA3	-1,43%	1,56E-08	0,59%	7,24E-01	0.5867 (0.53-0.65)	0.5927 (0.53-0.65)	0.5725 (0.52-0.63)
cg02336104	3	FOXP1	-1,22%	2,55E-08	-0,40%	5,62E-02	0.6758 (0.65-0.70)	0.6718 (0.65-0.69)	0.6637 (0.64-0.68)
cg19859270	3	GPR15	-1,31%	9,00E-24	-0,33%	2,82E-05	0.9560 (0.95-0.96)	0.9527 (0.94-0.96)	0.9430 (0.93-0.95)
cg02657160	3	CPOX	-1,48%	1,67E-09	-0,45%	2,15E-02	0.9255 (0.91-0.94)	0.9210 (0.90-0.94)	0.9107 (0.88-0.93)
cg24323726	3	ZBED2/CD96	-1,14%	2,43E-08	0,60%	4,10E-01	0.7791 (0.74-0.82)	0.7851 (0.75-0.82)	0.7677 (0.73-0.81)
cg18548864	3	BOC	0,89%	1,86E-08	0,40%	8,91E-01	0.4105 (0.38-0.44)	0.4145 (0.39-0.44)	0.4194 (0.39-0.45)
cg22870429	3	TIGIT	-2,24%	2,78E-12	-0,66%	1,71E-01	0.6114 (0.57-0.66)	0.6048 (0.56-0.65)	0.5890 (0.55-0.63)
cg03631078	3	ZBTB20	2,84%	9,30E-09	0,20%	1,72E-01	0.7274 (0.68-0.77)	0.7294 (0.68-0.78)	0.7558 (0.71-0.79)
cg16904330	3	ADPRH	0,56%	1,38E-08	0,32%	1,89E-01	0.1049 (0.09-0.12)	0.1081 (0.10-0.12)	0.1105 (0.10-0.12)
cg11854227	3	ILDR1	0,50%	6,14E-08	0,34%	1,56E-01	0.1196 (0.11-0.13)	0.1230 (0.11-0.14)	0.1246 (0.11-0.14)
cg00638210	3	PLXNA1	2,50%	3,80E-10	0,81%	1,84E-01	0.3029 (0.27-0.34)	0.3110 (0.28-0.35)	0.3279 (0.29-0.36)
cg19982684	3	x[a]	1,74%	8,51E-10	0,05%	2,54E-01	0.1587 (0.13-0.20)	0.1592 (0.13-0.20)	0.1761 (0.15-0.22)
cg25197194	3	CCDC48	4,62%	5,71E-10	1,27%	2,57E-01	0.4321 (0.37-0.50)	0.4448 (0.38-0.52)	0.4783 (0.42-0.55)
cg05279738	3	x[a]	-2,71%	1,47E-08	-0,76%	1,56E-02	0.3555 (0.30-0.41)	0.3479 (0.29-0.40)	0.3284 (0.28-0.40)
cg22571654	3	MED12L	2,71%	9,53E-08	1,17%	1,93E-01	0.2784 (0.24-0.32)	0.2901 (0.25-0.34)	0.3055 (0.27-0.34)
cg02499139	3	TNIK	0,67%	7,41E-08	0,14%	5,24E-01	0.1268 (0.11-0.14)	0.1282 (0.11-0.15)	0.1335 (0.11-0.15)
cg17566735	3	PLD1	0,07%	4,29E-08	0,03%	7,29E-01	0.0687 (0.06-0.07)	0.0690 (0.06-0.08)	0.0694 (0.06-0.08)
cg06294803	3	EIF4G1	-1,00%	3,57E-08	0,17%	2,59E-03	0.2654 (0.23-0.30)	0.2671 (0.23-0.30)	0.2554 (0.22-0.30)
cg16588163	3	IL1RAP	1,60%	1,93E-08	0,74%	3,28E-01	0.3939 (0.37-0.42)	0.4012 (0.37-0.43)	0.4099 (0.38-0.43)
cg21664281	3	x[a]	8,52%	4,57E-11	1,94%	1,57E-02	0.4526 (0.39-0.57)	0.4719 (0.40-0.58)	0.5377 (0.43-0.63)
cg22891595	3	x[a]	3,46%	3,61E-14	-0,10%	4,19E-02	0.6391 (0.60-0.68)	0.6382 (0.60-0.68)	0.6738 (0.63-0.70)
cg06012804	3	x[a]	-3,16%	1,89E-08	0,09%	1,94E-02	0.5423 (0.47-0.61)	0.5433 (0.47-0.61)	0.5107 (0.44-0.59)
cg07141504	3	LOC100131551	0,51%	1,62E-08	0,35%	1,25E-01	0.1042 (0.09-0.12)	0.1077 (0.09-0.12)	0.1093 (0.10-0.12)
cg13185177	3	GP5	8,89%	3,92E-17	2,92%	1,82E-02	0.3779 (0.32-0.44)	0.4071 (0.35-0.47)	0.4668 (0.40-0.55)
cg14880079	3	GP5	0,73%	5,44E-08	0,50%	1,13E-01	0.1274 (0.11-0.14)	0.1324 (0.12-0.15)	0.1347 (0.12-0.15)
cg15251319	3	C3orf21	0,91%	4,19E-08	0,30%	1,98E-01	0.7639 (0.74-0.79)	0.7669 (0.74-0.79)	0.7730 (0.75-0.80)
cg08202836	3	LRRC33	1,51%	3,85E-08	0,08%	8,24E-01	0.8802 (0.85-0.90)	0.8809 (0.86-0.90)	0.8952 (0.88-0.91)
cg10631854	3	x[a]	-1,87%	2,59E-08	0,90%	8,29E-01	0.5595 (0.49-0.64)	0.5685 (0.50-0.64)	0.5408 (0.48-0.62)
cg10843276	4	PCGF3	3,76%	4,38E-11	2,00%	1,53E-02	0.5301 (0.48-0.59)	0.5501 (0.50-0.60)	0.5677 (0.52-0.63)
cg26923863	4	CTBP1	3,57%	5,17E-10	2,83%	1,41E-01	0.3029 (0.26-0.35)	0.3312 (0.29-0.39)	0.3386 (0.29-0.39)
cg07094298	4	TNIP2	-1,41%	1,90E-13	0,29%	6,18E-03	0.3099 (0.26-0.36)	0.3128 (0.26-0.37)	0.2958 (0.24-0.35)
cg00741986	4	TNIP2	-1,16%	6,36E-08	-0,15%	5,03E-01	0.6325 (0.59-0.67)	0.6310 (0.59-0.67)	0.6208 (0.57-0.65)
cg08763102	4	HTT	-2,23%	5,65E-08	0,26%	6,03E-01	0.6034 (0.57-0.64)	0.6060 (0.57-0.64)	0.5811 (0.54-0.62)
cg21121843	4	HTT	-2,07%	1,79E-13	-1,27%	9,34E-04	0.1334 (0.11-0.16)	0.1207 (0.10-0.15)	0.1127 (0.10-0.13)
cg16324409	4	RGS12	0,64%	1,23E-08	0,40%	6,76E-02	0.1317 (0.12-0.15)	0.1357 (0.12-0.15)	0.1381 (0.12-0.15)
cg19906672	4	TBC1D14	-2,80%	8,79E-12	-0,57%	3,90E-01	0.7040 (0.67-0.74)	0.6983 (0.66-0.74)	0.6760 (0.64-0.72)
cg06496803	4	TBC1D14	1,09%	3,02E-09	0,35%	2,63E-02	0.2634 (0.23-0.30)	0.2669 (0.23-0.31)	0.2743 (0.24-0.32)
cg03254067	4	CLNK	0,40%	1,71E-09	0,07%	3,59E-01	0.0473 (0.04-0.05)	0.0480 (0.04-0.05)	0.0513 (0.05-0.06)
cg04332373	4	CD38	0,58%	9,85E-12	0,17%	2,31E-02	0.0859 (0.08-0.09)	0.0876 (0.08-0.09)	0.0917 (0.08-0.10)
cg04811945	4	CD38	0,54%	4,93E-10	0,22%	8,97E-03	0.0775 (0.07-0.08)	0.0798 (0.07-0.09)	0.0829 (0.08-0.09)
cg19719391	4	x[a]	4,18%	2,32E-13	0,53%	4,63E-03	0.5305 (0.48-0.57)	0.5358 (0.49-0.58)	0.5723 (0.52-0.61)
cg22601421	4	RBM47	0,23%	6,99E-11	0,12%	5,28E-02	0.0322 (0.03-0.04)	0.0334 (0.03-0.04)	0.0345 (0.03-0.04)
cg20435896	4	RBM47	2,00%	2,77E-14	1,61%	1,78E-03	0.4236 (0.40-0.45)	0.4398 (0.42-0.46)	0.4436 (0.42-0.46)
cg27076223	4	RBM47	1,13%	1,59E-12	0,47%	6,23E-02	0.2152 (0.19-0.24)	0.2199 (0.19-0.25)	0.2265 (0.20-0.26)
cg08949974	4	RBM47	1,13%	4,17E-08	0,48%	1,59E-01	0.1877 (0.17-0.21)	0.1925 (0.17-0.22)	0.1990 (0.17-0.22)
cg04307083	4	HOPX	0,60%	2,18E-08	0,30%	1,41E-02	0.1088 (0.10-0.12)	0.1118 (0.10-0.12)	0.1148 (0.11-0.12)
cg12623364	4	LEF1	-0,30%	4,40E-08	-0,05%	4,50E-02	0.0979 (0.09-0.11)	0.0974 (0.08-0.11)	0.0949 (0.08-0.11)
cg03835987	4	ELOVL6	0,84%	1,30E-10	0,55%	2,59E-01	0.1861 (0.16-0.21)	0.1916 (0.17-0.22)	0.1945 (0.17-0.22)
cg25198847	4	x[a]	1,75%	9,23E-08	0,85%	3,42E-01	0.4115 (0.33-0.50)	0.4199 (0.34-0.51)	0.4290 (0.35-0.53)
cg13643914	4	QRFPR	0,67%	8,66E-09	0,06%	2,14E-01	0.0448 (0.03-0.06)	0.0454 (0.04-0.06)	0.0515 (0.04-0.06)
cg03814093	4	KIAA0922	-0,79%	1,91E-10	-0,16%	7,99E-03	0.0763 (0.06-0.09)	0.0747 (0.06-0.09)	0.0684 (0.06-0.08)
cg24556382	4	GALNT7	-4,25%	2,64E-10	-0,84%	2,00E-02	0.7934 (0.74-0.85)	0.7851 (0.72-0.84)	0.7510 (0.68-0.81)
cg21960364	4	x[a]	-2,98%	2,59E-08	-0,70%	8,84E-02	0.7444 (0.70-0.78)	0.7374 (0.69-0.78)	0.7147 (0.67-0.76)
cg11554391*	5	AHRR	-1,03%	5,07E-24	-0,15%	1,10E-08	0.0911 (0.08-0.10)	0.0895 (0.08-0.10)	0.0808 (0.07-0.09)
cg17924476	5	AHRR	5,80%	2,67E-16	1,67%	2,38E-04	0.2628 (0.23-0.30)	0.2795 (0.24-0.32)	0.3208 (0.26-0.36)
cg08606254	5	AHRR	0,70%	9,25E-09	0,09%	1,74E-01	0.9490 (0.93-0.96)	0.9499 (0.93-0.96)	0.9560 (0.94-0.97)
cg09084391	5	AHRR	-1,09%	7,59E-08	-0,41%	3,55E-03	0.8893 (0.87-0.91)	0.8852 (0.87-0.90)	0.8785 (0.86-0.90)
cg12806681	5	AHRR	-2,44%	3,63E-33	-0,33%	4,55E-03	0.9537 (0.94-0.96)	0.9504 (0.93-0.96)	0.9293 (0.90-0.95)
cg03991871	5	AHRR	-5,73%	4,84E-41	-1,10%	4,45E-07	0.9469 (0.92-0.96)	0.9359 (0.90-0.95)	0.8896 (0.80-0.93)
cg23916896	5	AHRR	-2,11%	9,79E-29	-0,37%	1,04E-05	0.1286 (0.11-0.15)	0.1249 (0.11-0.14)	0.1075 (0.10-0.12)
cg11902777	5	AHRR	-1,12%	4,04E-40	-0,42%	2,58E-07	0.0379 (0.03-0.05)	0.0337 (0.03-0.04)	0.0267 (0.02-0.03)
cg01899089	5	AHRR	-2,86%	8,95E-32	-0,37%	2,64E-05	0.5969 (0.44-0.66)	0.5139 (0.31-0.61)	0.2514 (0.20-0.34)
cg05575921*	5	AHRR	-24,40%	2,54E-182	-3,31%	4,62E-36	0.8841 (0.86-0.90)	0.8510 (0.79-0.89)	0.6401 (0.60-0.73)
cg22103736	5	AHRR	-0,58%	2,94E-08	0,18%	7,14E-02	0.1547 (0.14-0.17)	0.1565 (0.14-0.18)	0.1489 (0.13-0.17)
cg26703534	5	AHRR	-6,17%	2,14E-38	-0,28%	5,09E-01	0.6269 (0.60-0.65)	0.6241 (0.60-0.65)	0.5652 (0.52-0.60)
cg01097768	5	AHRR	-2,40%	1,22E-11	-0,13%	8,45E-01	0.7457 (0.70-0.79)	0.7444 (0.70-0.79)	0.7217 (0.67-0.76)
cg14817490	5	AHRR	-3,81%	4,08E-50	-0,16%	4,17E-02	0.1350 (0.11-0.16)	0.1333 (0.11-0.16)	0.0968 (0.08-0.12)
cg25648203	5	AHRR	-7,96%	4,73E-47	-0,76%	2,99E-02	0.8642 (0.83-0.89)	0.8565 (0.82-0.89)	0.7846 (0.72-0.84)
cg21161138	5	AHRR	-10,45%	8,58E-81	-1,93%	2,92E-06	0.6998 (0.67-0.73)	0.6805 (0.64-0.71)	0.5954 (0.53-0.65)
cg03604011*	5	AHRR	3,40%	7,87E-30	1,58%	7,21E-08	0.0759 (0.06-0.10)	0.0917 (0.07-0.12)	0.1099 (0.08-0.15)
cg24090911	5	AHRR	-4,08%	3,41E-14	-1,18%	2,62E-03	0.8246 (0.79-0.86)	0.8128 (0.78-0.85)	0.7838 (0.74-0.83)
cg26850624	5	AHRR	3,31%	3,79E-09	1,30%	1,15E-04	0.6610 (0.62-0.70)	0.6741 (0.63-0.72)	0.6942 (0.66-0.72)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg10841124	5	AHRR	2,23%	6,61E-09	0,93%	1,51E-06	0.8643 (0.84-0.89)	0.8736 (0.85-0.89)	0.8866 (0.86-0.91)
cg03999941	5	x[a]	-1,57%	3,25E-10	0,01%	9,20E-01	0.8145 (0.79-0.84)	0.8146 (0.79-0.84)	0.7988 (0.77-0.83)
cg21747070	5	x[a]	-1,16%	8,06E-12	0,43%	3,78E-01	0.5407 (0.52-0.57)	0.5449 (0.52-0.57)	0.5291 (0.50-0.55)
cg24908166	5	TERT	-0,63%	3,33E-09	-0,22%	1,43E-01	0.9537 (0.94-0.96)	0.9515 (0.94-0.96)	0.9475 (0.94-0.96)
cg05255351	5	LPCAT1	0,74%	2,40E-09	0,03%	2,32E-01	0.1133 (0.10-0.13)	0.1136 (0.10-0.13)	0.1207 (0.11-0.14)
cg08445550	5	MYO10	-1,47%	9,81E-10	-0,09%	4,97E-01	0.8613 (0.82-0.89)	0.8604 (0.82-0.89)	0.8466 (0.80-0.88)
cg25537119	5	MYO10	-1,01%	4,12E-08	-0,07%	7,05E-01	0.9291 (0.91-0.94)	0.9284 (0.91-0.94)	0.9190 (0.90-0.94)
cg13039251	5	PDZD2	7,15%	3,81E-17	2,86%	3,82E-05	0.7536 (0.69-0.82)	0.7822 (0.71-0.84)	0.8251 (0.77-0.87)
cg18752880	5	C1QTNF3	0,43%	1,01E-10	0,11%	3,82E-01	0.0855 (0.08-0.10)	0.0866 (0.08-0.10)	0.0897 (0.08-0.10)
cg23977888	5	x[a]	-1,25%	5,48E-08	-0,12%	2,62E-02	0.6926 (0.65-0.73)	0.6914 (0.65-0.73)	0.6801 (0.64-0.71)
cg16578883	5	PDE4D	-2,21%	4,35E-09	-0,65%	3,15E-02	0.8472 (0.80-0.88)	0.8407 (0.79-0.88)	0.8251 (0.78-0.87)
cg03227963	5	x[a]	-1,11%	7,20E-08	-0,25%	3,78E-03	0.8807 (0.85-0.91)	0.8782 (0.85-0.90)	0.8695 (0.83-0.90)
cg05673882	5	POLK	-2,65%	4,49E-23	-1,15%	2,05E-03	0.1515 (0.12-0.19)	0.1400 (0.11-0.18)	0.1250 (0.10-0.16)
cg00499700	5	F2RL1	1,07%	8,95E-09	0,91%	1,54E-01	0.1787 (0.17-0.20)	0.1878 (0.17-0.20)	0.1894 (0.17-0.21)
cg06677021	5	SERINC5	-8,36%	2,58E-08	0,14%	1,60E-01	0.4659 (0.38-0.55)	0.4673 (0.37-0.55)	0.3822 (0.28-0.48)
cg26908328	5	SERINC5	-0,60%	2,73E-08	0,11%	6,41E-01	0.0824 (0.07-0.09)	0.0835 (0.07-0.10)	0.0765 (0.07-0.09)
cg07870237	5	ACSL6	0,78%	7,01E-10	0,41%	5,03E-01	0.1628 (0.15-0.18)	0.1669 (0.15-0.18)	0.1706 (0.15-0.19)
cg16545743	5	IL3	1,89%	7,96E-08	0,03%	6,99E-01	0.6703 (0.63-0.72)	0.6706 (0.62-0.71)	0.6892 (0.64-0.74)
cg17983064	5	IL3	3,73%	3,92E-08	0,06%	4,79E-01	0.4791 (0.42-0.54)	0.4796 (0.41-0.54)	0.5164 (0.46-0.58)
cg16151960	5	PHF15	-2,08%	1,34E-08	-0,42%	6,93E-01	0.7222 (0.69-0.75)	0.7180 (0.69-0.75)	0.7015 (0.68-0.73)
cg13009654	5	EGR1	0,34%	1,76E-09	1,17%	4,60E-01	0.1704 (0.16-0.19)	0.1821 (0.16-0.20)	0.1738 (0.16-0.19)
cg23205886	5	SNHG4/MATR3	-1,60%	7,38E-10	0,15%	9,57E-01	0.4554 (0.40-0.50)	0.4569 (0.41-0.50)	0.4393 (0.38-0.49)
cg14117392	5	x[a]	1,29%	1,01E-09	0,59%	2,64E-01	0.2327 (0.21-0.26)	0.2387 (0.21-0.27)	0.2456 (0.22-0.28)
cg18814344	5	CXXC5	0,61%	3,30E-09	0,40%	1,47E-01	0.1600 (0.15-0.17)	0.1640 (0.15-0.18)	0.1661 (0.15-0.18)
cg15824323	5	x[a]	-0,63%	1,25E-08	-0,07%	6,92E-01	0.8898 (0.86-0.91)	0.8891 (0.86-0.91)	0.8835 (0.85-0.91)
cg01553231	5	x[a]	-0,62%	2,87E-09	0,21%	6,15E-01	0.7884 (0.76-0.81)	0.7905 (0.76-0.82)	0.7823 (0.75-0.81)
cg11963365	5	KIAA0141	1,95%	9,66E-08	-0,38%	2,05E-01	0.5356 (0.50-0.58)	0.5318 (0.48-0.57)	0.5551 (0.52-0.60)
cg19514721	5	x[a]	3,54%	1,72E-08	-0,16%	8,18E-01	0.6216 (0.57-0.67)	0.6200 (0.57-0.67)	0.6570 (0.60-0.70)
cg16786458	5	PPARGC1B	2,16%	2,79E-11	0,39%	5,41E-01	0.1744 (0.16-0.20)	0.1783 (0.16-0.20)	0.1960 (0.17-0.22)
cg14580211	5	C5orf62	-3,91%	2,19E-12	-0,94%	3,51E-03	0.6909 (0.65-0.73)	0.6816 (0.64-0.72)	0.6519 (0.61-0.69)
cg18188717	5	C5orf4	0,61%	9,85E-09	0,10%	8,00E-02	0.1015 (0.09-0.12)	0.1026 (0.09-0.12)	0.1076 (0.09-0.13)
cg17668731	5	CCNJL	0,44%	1,74E-08	0,15%	1,14E-01	0.0952 (0.08-0.11)	0.0967 (0.08-0.11)	0.0996 (0.09-0.11)
cg00357551	5	FAM196B/DOCK2	3,21%	2,73E-10	0,44%	7,44E-01	0.2961 (0.24-0.36)	0.3006 (0.25-0.37)	0.3282 (0.26-0.39)
cg17038235	5	ERGIC1	2,35%	3,15E-11	0,67%	3,99E-02	0.4867 (0.44-0.53)	0.4934 (0.45-0.55)	0.5102 (0.47-0.57)
cg17747551	5	SFXN1	-0,77%	1,29E-08	0,14%	5,91E-01	0.8723 (0.84-0.90)	0.8737 (0.84-0.90)	0.8646 (0.82-0.89)
cg08866634	5	RNF44	-0,30%	2,91E-10	-0,02%	3,16E-01	0.0823 (0.08-0.09)	0.0821 (0.08-0.09)	0.0793 (0.07-0.09)
cg19589060	5	HK3	0,36%	2,75E-08	-0,08%	3,75E-01	0.0636 (0.06-0.07)	0.0628 (0.06-0.07)	0.0672 (0.06-0.08)
cg24711224	5	NSD1	-0,69%	5,07E-08	0,11%	2,71E-01	0.1542 (0.14-0.17)	0.1553 (0.13-0.17)	0.1474 (0.13-0.17)
cg12513616	5	x[a]	-0,87%	2,88E-17	0,02%	1,97E-03	0.2101 (0.20-0.23)	0.2103 (0.20-0.23)	0.2014 (0.19-0.21)
cg23001918	6	x[a]	-0,92%	1,80E-08	0,15%	8,98E-01	0.7961 (0.76-0.83)	0.7976 (0.76-0.83)	0.7868 (0.75-0.82)
cg09757644	6	x[a]	-1,02%	7,11E-08	-0,05%	9,55E-01	0.8283 (0.79-0.86)	0.8277 (0.79-0.87)	0.8181 (0.78-0.86)
cg13299325	6	x[a]	-1,00%	6,67E-12	0,34%	8,02E-01	0.7343 (0.71-0.77)	0.7377 (0.70-0.77)	0.7243 (0.69-0.76)
cg01882991	6	x[a]	-1,40%	1,74E-08	-0,51%	4,93E-03	0.6357 (0.62-0.66)	0.6306 (0.61-0.65)	0.6217 (0.60-0.64)
cg07090714	6	x[a]	-2,10%	1,66E-20	0,10%	5,82E-03	0.4239 (0.40-0.45)	0.4249 (0.39-0.45)	0.4029 (0.37-0.43)
cg06193328	6	x[a]	-1,72%	8,25E-08	0,30%	3,54E-01	0.8070 (0.75-0.86)	0.8101 (0.76-0.85)	0.7898 (0.75-0.83)
cg03446062	6	x[a]	1,56%	6,39E-08	0,30%	1,51E-01	0.2344 (0.19-0.28)	0.2374 (0.20-0.29)	0.2499 (0.21-0.30)
cg12539124	6	FLJ22536	-0,94%	1,06E-08	0,30%	4,06E-01	0.1090 (0.10-0.13)	0.1120 (0.09-0.13)	0.0996 (0.09-0.12)
cg11435441	6	x[a]	-0,76%	2,30E-08	0,31%	9,66E-01	0.7486 (0.71-0.79)	0.7517 (0.72-0.79)	0.7410 (0.71-0.77)
cg12423733	6	MAS1L	1,52%	1,62E-11	0,77%	6,76E-06	0.0890 (0.08-0.11)	0.0967 (0.08-0.12)	0.1042 (0.09-0.13)
cg25730428	6	MAS1L	1,14%	5,19E-09	0,39%	6,39E-02	0.1375 (0.12-0.16)	0.1414 (0.12-0.16)	0.1489 (0.13-0.18)
cg12600843	6	TRIM26	-3,61%	3,72E-08	-1,18%	3,54E-01	0.7510 (0.66-0.82)	0.7392 (0.66-0.81)	0.7149 (0.62-0.78)
cg10145196	6	KIAA1949	-1,06%	2,36E-08	-0,54%	2,20E-01	0.7253 (0.67-0.77)	0.7199 (0.67-0.77)	0.7147 (0.67-0.75)
cg06126421*	6	x[a]	-17,05%	1,72E-75	-5,65%	1,50E-22	0.7724 (0.70-0.82)	0.7159 (0.63-0.79)	0.6019 (0.48-0.69)
cg14753356	6	x[a]	-5,37%	8,14E-24	-2,26%	2,89E-07	0.2383 (0.20-0.28)	0.2157 (0.18-0.26)	0.1846 (0.15-0.22)
cg24859433	6	x[a]	-4,03%	3,06E-41	-0,76%	3,15E-06	0.9322 (0.91-0.95)	0.9246 (0.90-0.94)	0.8919 (0.85-0.92)
cg15342087	6	x[a]	-3,36%	1,71E-42	-0,56%	1,57E-05	0.9371 (0.92-0.95)	0.9316 (0.91-0.94)	0.9036 (0.87-0.93)
cg08951271	6	DDR1	-2,54%	6,05E-08	-0,26%	5,02E-01	0.8321 (0.78-0.87)	0.8295 (0.78-0.87)	0.8067 (0.75-0.86)
cg08827454	6	x[a]	0,42%	2,71E-08	0,24%	5,59E-01	0.1228 (0.11-0.13)	0.1251 (0.11-0.14)	0.1269 (0.12-0.14)
cg09637172	6	TNF	-0,74%	4,64E-08	0,19%	8,54E-01	0.8560 (0.83-0.88)	0.8579 (0.82-0.89)	0.8486 (0.82-0.88)
cg14868222	6	LTB	-0,27%	6,38E-09	0,53%	8,67E-01	0.5269 (0.51-0.55)	0.5321 (0.51-0.55)	0.5242 (0.50-0.54)
cg02249390	6	LTB	-0,86%	2,76E-08	0,45%	6,34E-01	0.6168 (0.57-0.67)	0.6212 (0.57-0.68)	0.6082 (0.56-0.65)
cg23781467	6	VARS	1,68%	2,17E-08	0,67%	1,39E-03	0.6611 (0.64-0.69)	0.6678 (0.64-0.70)	0.6779 (0.65-0.71)
cg05302489	6	VARS	4,56%	2,52E-10	1,41%	7,73E-03	0.7120 (0.66-0.76)	0.7260 (0.68-0.78)	0.7575 (0.71-0.80)
cg04368724	6	VARS	2,24%	1,78E-11	0,13%	2,65E-02	0.8798 (0.84-0.91)	0.8812 (0.85-0.92)	0.9022 (0.86-0.93)
cg24771152	6	VARS	0,80%	4,03E-08	0,44%	2,91E-03	0.9486 (0.93-0.96)	0.9530 (0.94-0.96)	0.9566 (0.94-0.97)
cg04018738	6	VARS	2,91%	2,40E-13	0,93%	2,51E-03	0.8450 (0.81-0.88)	0.8543 (0.82-0.89)	0.8740 (0.83-0.90)
cg17619755	6	VARS	4,05%	2,64E-18	0,75%	9,20E-04	0.6519 (0.61-0.69)	0.6594 (0.62-0.70)	0.6924 (0.64-0.73)
cg08899667	6	VARS	2,98%	9,31E-15	0,38%	8,95E-03	0.4997 (0.46-0.54)	0.5036 (0.46-0.55)	0.5295 (0.48-0.57)
cg10807309	6	VARS	1,83%	1,25E-12	0,52%	8,00E-03	0.1534 (0.13-0.18)	0.1586 (0.14-0.18)	0.1717 (0.15-0.20)
cg06857094	6	VARS	-0,31%	6,30E-08	0,21%	1,05E-01	0.1573 (0.14-0.17)	0.1594 (0.14-0.18)	0.1542 (0.14-0.17)
cg13541527	6	C6orf48/SNORD52	-1,53%	9,91E-10	0,04%	1,79E-02	0.2186 (0.18-0.26)	0.2190 (0.18-0.26)	0.2033 (0.17-0.24)
cg05789250	6	C6orf48/SNORD52	-0,47%	1,22E-08	-0,13%	5,84E-02	0.1233 (0.11-0.14)	0.1220 (0.11-0.14)	0.1186 (0.10-0.13)
cg03280235	6	PBX2/GPSM3	-1,16%	2,09E-08	-0,49%	5,05E-01	0.6234 (0.58-0.66)	0.6185 (0.57-0.66)	0.6118 (0.57-0.65)
cg00442389	6	GPSM3	-0,17%	5,00E-08	-0,09%	5,24E-02	0.0420 (0.04-0.05)	0.0411 (0.04-0.05)	0.0403 (0.03-0.05)
cg08331398	6	PSMB8	-1,81%	4,38E-10	-0,31%	5,84E-01	0.8846 (0.85-0.91)	0.8815 (0.84-0.91)	0.8664 (0.83-0.90)
cg06473288	6	TAP1	-0,18%	1,19E-08	0,81%	7,48E-01	0.6938 (0.60-0.79)	0.7019 (0.61-0.78)	0.6920 (0.59-0.77)
cg04772644	6	RXRB	-0,29%	1,13E-09	-0,01%	5,91E-01	0.0624 (0.06-0.07)	0.0623 (0.06-0.07)	0.0596 (0.05-0.07)
cg17567838	6	SLC39A7/RXRB	0,16%	2,45E-08	0,46%	9,04E-01	0.1930 (0.18-0.21)	0.1976 (0.18-0.21)	0.1946 (0.18-0.21)
cg11222065	6	VPS52/RPS18	-0,55%	1,84E-08	0,31%	9,20E-02	0.2220 (0.20-0.24)	0.2251 (0.20-0.25)	0.2165 (0.20-0.25)
cg26283141	6	VPS52/RPS18	-0,72%	1,71E-09	1,25%	1,97E-01	0.4302 (0.38-0.49)	0.4427 (0.39-0.49)	0.4230 (0.37-0.46)
cg27182159	6	RPS18	-1,93%	2,45E-08	0,45%	6,37E-01	0.5319 (0.48-0.58)	0.5364 (0.48-0.58)	0.5126 (0.46-0.57)
cg03702686	6	B3GALT4/RPS18	-3,08%	9,91E-08	-0,84%	4,56E-01	0.6630 (0.61-0.72)	0.6546 (0.60-0.71)	0.6322 (0.57-0.70)
cg15454110	6	CUTA	-2,47%	6,37E-09	0,05%	7,00E-01	0.8045 (0.73-0.85)	0.8050 (0.73-0.86)	0.7798 (0.71-0.83)
cg09287933	6	CUTA	-0,95%	3,48E-08	0,27%	4,34E-01	0.8549 (0.81-0.89)	0.8576 (0.81-0.90)	0.8454 (0.79-0.88)
cg01300096	6	CUTA	-1,54%	7,16E-11	-0,05%	2,76E-01	0.4386 (0.39-0.49)	0.4381 (0.39-0.49)	0.4232 (0.37-0.46)
cg14021478	6	CUTA	-0,82%	4,36E-11	0,17%	2,96E-01	0.6334 (0.60-0.66)	0.6351 (0.61-0.66)	0.6252 (0.60-0.65)
cg23565821	6	CUTA	-0,17%	1,24E-12	0,37%	4,32E-02	0.1472 (0.13-0.16)	0.1509 (0.14-0.17)	0.1455 (0.13-0.16)
cg02654940	6	HMGA1	-0,72%	3,69E-09	0,50%	6,54E-01	0.3258 (0.29-0.35)	0.3308 (0.30-0.36)	0.3186 (0.28-0.35)
cg22797644	6	HMGA1	-0,10%	3,27E-09	0,62%	3,02E-01	0.1893 (0.17-0.21)	0.1954 (0.18-0.21)	0.1883 (0.17-0.20)
cg08335767	6	HMGA1	-0,13%	9,00E-08	0,25%	9,95E-02	0.1044 (0.09-0.11)	0.1069 (0.10-0.12)	0.1031 (0.09-0.11)
cg09447622	6	TCP11	0,47%	4,65E-08	0,06%	4,72E-01	0.0850 (0.08-0.09)	0.0855 (0.08-0.09)	0.0897 (0.08-0.10)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg15474579	6	CDKN1A	-3,23%	9,30E-09	-1,12%	6,24E-06	0.3563 (0.31-0.41)	0.3452 (0.30-0.39)	0.3240 (0.29-0.37)
cg23575275	6	CDKN1A	0,59%	7,23E-08	0,25%	9,69E-01	0.1711 (0.15-0.19)	0.1736 (0.16-0.19)	0.1770 (0.16-0.20)
cg03957124	6	x[a]	-1,50%	2,44E-08	0,17%	2,64E-01	0.5163 (0.45-0.58)	0.5180 (0.45-0.57)	0.5013 (0.44-0.56)
cg02524834	6	CCND3	2,71%	8,73E-09	0,55%	8,59E-02	0.8749 (0.83-0.91)	0.8804 (0.84-0.91)	0.9020 (0.86-0.92)
cg11854981	6	TRERF1	1,48%	1,19E-08	-0,22%	1,05E-01	0.2013 (0.17-0.24)	0.1991 (0.17-0.24)	0.2161 (0.18-0.26)
cg20819617	6	PTK7	0,46%	4,21E-10	0,44%	8,44E-02	0.1042 (0.10-0.11)	0.1086 (0.10-0.12)	0.1088 (0.10-0.12)
cg20003368	6	x[a]	0,29%	4,26E-09	0,07%	8,71E-02	0.0344 (0.03-0.04)	0.0351 (0.03-0.04)	0.0372 (0.03-0.04)
cg14431165	6	x[a]	0,81%	1,35E-09	0,05%	1,62E-01	0.1025 (0.09-0.11)	0.1030 (0.09-0.12)	0.1106 (0.09-0.12)
cg06560379	6	NFKBIE	-0,27%	1,05E-09	-0,01%	1,39E-01	0.0510 (0.05-0.06)	0.0508 (0.05-0.06)	0.0483 (0.04-0.05)
cg04326499	6	x[a]	-2,12%	2,86E-08	-0,33%	1,45E-01	0.8427 (0.80-0.88)	0.8394 (0.80-0.87)	0.8215 (0.79-0.86)
cg00937359	6	SYNCRIP	-0,69%	6,43E-08	0,49%	8,23E-01	0.2804 (0.25-0.32)	0.2853 (0.25-0.33)	0.2735 (0.24-0.31)
cg23279756	6	x[a]	2,60%	4,15E-08	0,17%	7,17E-01	0.2860 (0.24-0.33)	0.2877 (0.25-0.33)	0.3120 (0.27-0.35)
cg04642300	6	ARMC2	0,89%	7,69E-08	0,11%	1,87E-01	0.1098 (0.09-0.13)	0.1109 (0.09-0.13)	0.1186 (0.10-0.14)
cg02578070	6	MAN1A1	0,51%	7,18E-09	0,27%	7,83E-03	0.0567 (0.05-0.06)	0.0594 (0.05-0.07)	0.0618 (0.05-0.07)
cg02673417	6	MAP3K5	0,56%	6,84E-09	0,16%	9,78E-02	0.1157 (0.10-0.13)	0.1173 (0.10-0.14)	0.1213 (0.11-0.14)
cg17478979	6	ZC3H12D	2,11%	4,29E-08	2,22%	6,42E-02	0.5344 (0.50-0.57)	0.5566 (0.52-0.58)	0.5555 (0.52-0.58)
cg04494800	6	ZC3H12D	0,61%	5,49E-09	0,26%	1,86E-02	0.1663 (0.15-0.18)	0.1689 (0.16-0.18)	0.1724 (0.16-0.19)
cg15132169	6	ZC3H12D	-1,90%	6,78E-09	-1,08%	1,24E-01	0.1814 (0.14-0.23)	0.1706 (0.14-0.22)	0.1624 (0.13-0.20)
cg17501395	6	ZC3H12D	-1,24%	9,25E-10	-0,59%	3,03E-02	0.0927 (0.08-0.12)	0.0868 (0.07-0.11)	0.0803 (0.06-0.10)
cg18082788	6	ZC3H12D	-3,37%	1,91E-09	-1,38%	1,29E-02	0.1652 (0.12-0.22)	0.1515 (0.10-0.20)	0.1315 (0.09-0.18)
cg06762457	6	ZC3H12D	-4,95%	2,84E-10	-2,59%	3,07E-02	0.4930 (0.42-0.55)	0.4671 (0.38-0.53)	0.4435 (0.37-0.52)
cg14030904	6	ZC3H12D	-3,15%	2,43E-08	-1,68%	1,06E-01	0.3087 (0.25-0.36)	0.2919 (0.23-0.35)	0.2772 (0.23-0.34)
cg20335735	6	NUP43	-1,34%	1,07E-09	-0,39%	2,00E-01	0.7635 (0.71-0.81)	0.7597 (0.71-0.80)	0.7502 (0.70-0.79)
cg00931843	6	TIAM2	6,37%	1,69E-10	1,90%	5,54E-01	0.2963 (0.23-0.39)	0.3153 (0.24-0.41)	0.3600 (0.27-0.47)
cg26701785	6	SYNJ2	3,01%	1,62E-13	0,09%	2,12E-02	0.8498 (0.81-0.88)	0.8507 (0.81-0.89)	0.8800 (0.84-0.91)
cg08557970	6	RPS6KA2	0,21%	4,94E-08	0,22%	7,38E-02	0.2382 (0.20-0.29)	0.2404 (0.20-0.29)	0.2403 (0.20-0.30)
cg05877104	6	x[a]	-0,07%	3,79E-10	1,64%	5,39E-01	0.5410 (0.51-0.57)	0.5574 (0.52-0.58)	0.5404 (0.51-0.57)
cg03120555	7	PRKAR1B	2,30%	7,02E-08	1,03%	4,81E-01	0.4202 (0.37-0.47)	0.4306 (0.39-0.47)	0.4433 (0.40-0.49)
cg10251229	7	PRKAR1B	1,13%	1,14E-08	0,72%	8,58E-02	0.2193 (0.20-0.24)	0.2265 (0.21-0.25)	0.2306 (0.21-0.26)
cg00541718	7	PRKAR1B	3,12%	8,72E-09	0,74%	1,48E-01	0.3129 (0.28-0.36)	0.3203 (0.28-0.36)	0.3441 (0.30-0.38)
cg14835981	7	PRKAR1B	1,55%	8,98E-09	0,63%	9,67E-01	0.2337 (0.21-0.27)	0.2400 (0.21-0.27)	0.2492 (0.22-0.28)
cg17593625	7	PRKAR1B	1,79%	4,15E-08	1,20%	4,30E-01	0.2436 (0.20-0.29)	0.2556 (0.21-0.30)	0.2615 (0.22-0.31)
cg00921574	7	INTS1	-0,45%	8,71E-10	0,39%	9,56E-01	0.1005 (0.09-0.11)	0.1044 (0.09-0.12)	0.0960 (0.08-0.11)
cg20697025	7	x[a]	1,79%	2,87E-08	1,06%	8,15E-01	0.2748 (0.24-0.32)	0.2855 (0.24-0.32)	0.2927 (0.25-0.33)
cg17551891	7	MAD1L1	-2,59%	8,75E-12	-1,02%	2,64E-04	0.7785 (0.74-0.82)	0.7683 (0.73-0.81)	0.7525 (0.71-0.79)
cg15309361	7	MAD1L1	0,84%	2,23E-08	-0,02%	4,41E-01	0.1443 (0.12-0.17)	0.1442 (0.13-0.17)	0.1527 (0.13-0.18)
cg20161791	7	LFNG	0,14%	3,92E-09	1,29%	3,42E-01	0.2094 (0.18-0.24)	0.2223 (0.19-0.26)	0.2108 (0.18-0.25)
cg11884933	7	GNA12	0,62%	3,13E-08	0,17%	8,44E-01	0.1546 (0.14-0.17)	0.1563 (0.14-0.17)	0.1607 (0.14-0.18)
cg01492538	7	GNA12	1,06%	1,03E-10	0,75%	4,54E-01	0.1670 (0.16-0.18)	0.1746 (0.16-0.19)	0.1776 (0.16-0.19)
cg19717773	7	GNA12	-4,05%	1,15E-08	-1,37%	3,42E-07	0.7015 (0.65-0.76)	0.6879 (0.63-0.74)	0.6610 (0.60-0.72)
cg01741041	7	ACTB	-0,18%	8,22E-08	0,19%	2,43E-01	0.1478 (0.14-0.16)	0.1496 (0.14-0.16)	0.1459 (0.14-0.16)
cg05971148	7	ACTB	-0,20%	9,02E-08	-0,03%	2,69E-02	0.0570 (0.05-0.06)	0.0567 (0.05-0.06)	0.0550 (0.05-0.06)
cg12192749	7	x[a]	1,46%	7,58E-09	0,43%	1,78E-01	0.2295 (0.19-0.28)	0.2338 (0.19-0.29)	0.2442 (0.21-0.30)
cg04907244	7	SNORD93	-0,38%	2,37E-08	0,08%	3,90E-02	0.0805 (0.07-0.09)	0.0812 (0.07-0.09)	0.0767 (0.06-0.09)
cg08822075	7	NFE2L3	-1,49%	3,86E-09	-0,17%	4,13E-02	0.2027 (0.17-0.25)	0.2010 (0.17-0.24)	0.1878 (0.16-0.22)
cg02315870	7	HNRNPA2B1	-1,33%	1,17E-08	-0,35%	1,33E-01	0.7485 (0.71-0.79)	0.7450 (0.71-0.78)	0.7352 (0.69-0.77)
cg02451831	7	KIAA0087	-2,30%	2,79E-13	0,09%	1,08E-01	0.8243 (0.80-0.86)	0.8252 (0.79-0.86)	0.8013 (0.76-0.83)
cg16880946	7	HOXA6	-0,34%	8,28E-08	0,48%	2,03E-02	0.1682 (0.15-0.19)	0.1730 (0.16-0.19)	0.1648 (0.15-0.18)
cg08972170	7	C7orf41	5,33%	1,53E-09	1,91%	7,73E-02	0.5272 (0.47-0.59)	0.5463 (0.48-0.61)	0.5805 (0.51-0.64)
cg19089201	7	MYO1G	1,86%	9,26E-13	0,17%	5,81E-01	0.9094 (0.88-0.93)	0.9111 (0.88-0.93)	0.9280 (0.90-0.95)
cg22132788	7	MYO1G	6,68%	1,99E-34	0,61%	2,64E-01	0.8599 (0.81-0.90)	0.8660 (0.82-0.91)	0.9267 (0.88-0.95)
cg04180046	7	MYO1G	2,17%	5,92E-14	0,62%	1,31E-02	0.2316 (0.21-0.25)	0.2378 (0.22-0.26)	0.2533 (0.23-0.28)
cg12803068	7	MYO1G	14,96%	7,08E-30	2,49%	1,08E-02	0.7382 (0.60-0.85)	0.7630 (0.63-0.88)	0.8878 (0.77-0.94)
cg07826859	7	MYO1G	-5,12%	8,03E-27	-0,71%	6,15E-01	0.5836 (0.54-0.62)	0.5765 (0.53-0.61)	0.5324 (0.48-0.57)
cg03440944	7	C7orf40	-2,42%	1,68E-11	-0,28%	7,61E-01	0.7533 (0.72-0.78)	0.7505 (0.72-0.78)	0.7290 (0.70-0.76)
cg16459265	7	C7orf40/SNORA9	-5,34%	9,57E-11	0,82%	7,20E-01	0.6252 (0.55-0.71)	0.6334 (0.56-0.71)	0.5718 (0.51-0.64)
cg19956914	7	SUMF2	6,51%	1,51E-18	0,43%	7,30E-03	0.5468 (0.49-0.60)	0.5511 (0.50-0.61)	0.6119 (0.54-0.66)
cg26790897	7	SUMF2	3,42%	1,91E-13	1,42%	2,79E-04	0.3513 (0.30-0.41)	0.3655 (0.31-0.43)	0.3856 (0.33-0.46)
cg19072291	7	CRCP	1,70%	2,43E-08	0,02%	6,67E-01	0.8497 (0.82-0.87)	0.8500 (0.82-0.88)	0.8667 (0.84-0.89)
cg01712428	7	STX1A	0,26%	1,05E-09	0,09%	1,20E-01	0.0652 (0.06-0.07)	0.0660 (0.06-0.07)	0.0677 (0.06-0.08)
cg18478531	7	x[a]	2,06%	3,03E-08	1,08%	1,14E-01	0.2701 (0.23-0.33)	0.2809 (0.24-0.34)	0.2908 (0.24-0.36)
cg21008363	7	POM121C	2,07%	9,42E-08	-0,06%	9,36E-01	0.8408 (0.81-0.87)	0.8402 (0.81-0.87)	0.8615 (0.83-0.88)
cg04776231	7	PTPN12	1,54%	1,09E-08	0,80%	1,41E-01	0.2286 (0.19-0.27)	0.2366 (0.20-0.28)	0.2440 (0.21-0.30)
cg27457191	7	PHTF2	-1,09%	8,11E-08	0,48%	7,52E-01	0.7253 (0.66-0.78)	0.7301 (0.67-0.79)	0.7144 (0.64-0.78)
cg07413467	7	CDK6	1,87%	1,72E-08	0,37%	4,82E-02	0.6740 (0.63-0.73)	0.6777 (0.62-0.75)	0.6927 (0.63-0.76)
cg06688763	7	CDK6	2,19%	1,09E-08	0,11%	1,32E-01	0.7657 (0.72-0.81)	0.7668 (0.72-0.81)	0.7876 (0.73-0.83)
cg15261712	7	CDK6	2,03%	8,06E-09	-0,52%	1,79E-01	0.7813 (0.72-0.83)	0.7761 (0.72-0.84)	0.8016 (0.75-0.86)
cg08038054	7	GNG11	3,81%	2,92E-10	0,94%	1,52E-01	0.2023 (0.17-0.25)	0.2118 (0.17-0.26)	0.2404 (0.20-0.29)
cg08486903	7	SMURF1	-1,70%	8,29E-08	0,72%	8,28E-01	0.7599 (0.69-0.82)	0.7671 (0.69-0.83)	0.7429 (0.67-0.80)
cg21873567	7	GATS/PVRIG	-0,56%	6,82E-08	-0,22%	1,51E-01	0.9390 (0.92-0.95)	0.9368 (0.92-0.95)	0.9335 (0.92-0.95)
cg22728904	7	x[a]	2,41%	1,28E-12	1,14%	4,93E-02	0.3435 (0.31-0.37)	0.3548 (0.32-0.39)	0.3676 (0.33-0.40)
cg09762515	7	CUX1	4,59%	2,66E-12	0,20%	6,19E-01	0.5606 (0.51-0.60)	0.5626 (0.52-0.61)	0.6065 (0.56-0.65)
cg14012925	7	CUX1	2,21%	2,55E-09	0,73%	3,56E-01	0.4598 (0.44-0.49)	0.4671 (0.44-0.50)	0.4819 (0.46-0.51)
cg27405731	7	CUX1	2,59%	5,29E-09	0,90%	9,75E-01	0.4717 (0.43-0.51)	0.4807 (0.44-0.52)	0.4976 (0.46-0.53)
cg02856420	7	CUX1	1,63%	1,16E-08	1,01%	4,36E-01	0.2476 (0.21-0.29)	0.2577 (0.22-0.30)	0.2639 (0.23-0.31)
cg22103219	7	SH2B2	-0,41%	4,93E-11	0,43%	5,68E-01	0.1604 (0.15-0.18)	0.1647 (0.15-0.18)	0.1563 (0.14-0.17)
cg15873112	7	ATXN7L1	-1,93%	5,48E-09	0,33%	8,35E-01	0.8642 (0.83-0.89)	0.8675 (0.84-0.89)	0.8449 (0.81-0.88)
cg09837977	7	LRRN3/IMMP2L	-1,34%	3,27E-09	-0,37%	8,72E-05	0.9391 (0.92-0.95)	0.9354 (0.92-0.95)	0.9257 (0.90-0.94)
cg21322436	7	CNTNAP2	-4,74%	1,43E-24	-0,60%	6,06E-06	0.2855 (0.25-0.32)	0.2795 (0.25-0.31)	0.2381 (0.21-0.27)
cg25949550	7	CNTNAP2	-1,38%	4,53E-29	-0,21%	1,88E-06	0.0852 (0.08-0.09)	0.0831 (0.07-0.09)	0.0714 (0.06-0.08)
cg11207515	7	CNTNAP2	8,28%	2,53E-10	3,05%	4,92E-05	0.2725 (0.21-0.35)	0.3030 (0.24-0.38)	0.3552 (0.27-0.44)
cg17372101	7	CNTNAP2	7,23%	4,17E-12	2,45%	3,66E-02	0.3192 (0.27-0.38)	0.3437 (0.29-0.40)	0.3915 (0.33-0.47)
cg07632771	7	x[a]	-1,42%	2,10E-09	0,20%	2,45E-01	0.7922 (0.75-0.83)	0.7942 (0.76-0.83)	0.7780 (0.73-0.81)
cg02892367	7	PDIA4	-1,49%	7,98E-08	0,66%	6,79E-01	0.8713 (0.83-0.90)	0.8780 (0.83-0.91)	0.8564 (0.81-0.90)
cg12644845	7	GIMAP8	-0,80%	6,53E-08	0,09%	6,99E-02	0.8734 (0.85-0.89)	0.8744 (0.85-0.89)	0.8654 (0.84-0.89)
cg19421584	7	WDR60	1,65%	1,07E-09	0,55%	1,03E-01	0.5782 (0.56-0.60)	0.5836 (0.56-0.61)	0.5947 (0.57-0.62)
cg13582028	8	ERICH1	1,37%	3,91E-08	0,12%	1,06E-01	0.2633 (0.23-0.30)	0.2646 (0.23-0.30)	0.2770 (0.25-0.31)
cg12077460	8	MFHAS1	-1,46%	1,17E-09	-0,15%	1,21E-01	0.8201 (0.79-0.85)	0.8186 (0.79-0.85)	0.8055 (0.78-0.83)
cg27280535	8	MFHAS1	-3,38%	9,85E-09	-0,23%	1,41E-01	0.7501 (0.65-0.83)	0.7479 (0.66-0.82)	0.7163 (0.63-0.80)
cg26711333	8	x[a]	0,77%	4,17E-09	0,70%	4,24E-01	0.1506 (0.14-0.16)	0.1576 (0.15-0.17)	0.1583 (0.14-0.17)
cg11896066	8	x[a]	0,77%	3,40E-08	0,61%	2,30E-01	0.1331 (0.12-0.14)	0.1392 (0.13-0.15)	0.1408 (0.13-0.15)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg03899229	8	x[a]	0,60%	2,65E-10	0,29%	3,42E-01	0.1267 (0.11-0.14)	0.1296 (0.11-0.15)	0.1327 (0.12-0.15)
cg12276019	8	XKR6	-0,77%	8,75E-08	0,36%	5,45E-02	0.1235 (0.11-0.14)	0.1271 (0.11-0.15)	0.1158 (0.10-0.14)
cg11769400	8	PEBP4	0,19%	4,18E-09	0,07%	5,44E-02	0.0320 (0.03-0.04)	0.0327 (0.03-0.04)	0.0338 (0.03-0.04)
cg26837477	8	ENTPD4	-2,30%	2,13E-08	-0,59%	6,07E-01	0.7724 (0.72-0.82)	0.7666 (0.72-0.82)	0.7495 (0.70-0.80)
cg24540678	8	x[a]	-0,68%	2,30E-17	-0,19%	1,91E-05	0.0855 (0.08-0.09)	0.0836 (0.08-0.09)	0.0787 (0.07-0.09)
cg13518625	8	x[a]	-0,49%	1,15E-13	-0,05%	4,40E-01	0.0565 (0.05-0.07)	0.0560 (0.05-0.07)	0.0515 (0.04-0.06)
cg01419713	8	PLAT	-1,17%	2,76E-10	-0,48%	1,86E-02	0.8975 (0.87-0.92)	0.8927 (0.86-0.92)	0.8859 (0.86-0.91)
cg22855020	8	SLC20A2	1,68%	3,02E-08	0,65%	3,62E-01	0.2502 (0.20-0.31)	0.2568 (0.21-0.31)	0.2670 (0.22-0.34)
cg25801976	8	KIAA0146	0,46%	1,27E-08	0,12%	3,29E-01	0.1278 (0.11-0.14)	0.1290 (0.11-0.15)	0.1324 (0.12-0.15)
cg00007076	8	RRS1	-0,34%	1,86E-08	0,66%	8,71E-01	0.2059 (0.18-0.24)	0.2124 (0.19-0.24)	0.2025 (0.18-0.22)
cg14439572	8	x[a]	-1,23%	9,07E-08	-0,08%	1,72E-01	0.1214 (0.10-0.15)	0.1206 (0.09-0.15)	0.1091 (0.08-0.13)
cg19589396	8	x[a]	-3,86%	4,23E-14	-0,40%	4,70E-01	0.7511 (0.71-0.80)	0.7470 (0.71-0.80)	0.7125 (0.67-0.77)
cg26120971	8	x[a]	2,96%	4,48E-09	0,51%	9,67E-01	0.5900 (0.54-0.65)	0.5952 (0.53-0.66)	0.6196 (0.55-0.68)
cg13074203	8	SNTB1	1,88%	7,88E-10	0,54%	2,91E-02	0.1471 (0.13-0.18)	0.1525 (0.13-0.18)	0.1659 (0.14-0.20)
cg25305703	8	x[a]	-6,05%	2,09E-08	-0,89%	2,67E-02	0.7103 (0.65-0.76)	0.7013 (0.64-0.76)	0.6498 (0.58-0.73)
cg25247520	8	MIR1204/PVT1	-0,67%	3,90E-09	0,08%	2,09E-01	0.5453 (0.49-0.61)	0.5461 (0.49-0.60)	0.5386 (0.47-0.60)
cg11201447	8	MIR1204/PVT1	-0,14%	6,75E-09	0,24%	1,04E-01	0.4216 (0.36-0.49)	0.4239 (0.35-0.49)	0.4201 (0.34-0.48)
cg25020550	8	TG	-1,22%	3,72E-08	-0,18%	8,67E-01	0.1442 (0.12-0.17)	0.1425 (0.12-0.17)	0.1320 (0.11-0.15)
cg22845912	8	SLA/TG	-1,32%	2,58E-08	-0,10%	3,62E-01	0.8402 (0.80-0.87)	0.8392 (0.80-0.88)	0.8270 (0.79-0.86)
cg12140977	8	ST3GAL1	-1,90%	2,85E-09	-0,93%	1,27E-01	0.4570 (0.41-0.50)	0.4477 (0.41-0.49)	0.4380 (0.38-0.48)
cg11130692	8	EIF2C2	-0,76%	2,85E-09	0,18%	1,26E-01	0.1995 (0.18-0.21)	0.2013 (0.19-0.22)	0.1919 (0.18-0.21)
cg12075928	8	PTK2	-7,80%	1,40E-18	-2,62%	5,10E-03	0.4672 (0.39-0.53)	0.4411 (0.37-0.52)	0.3892 (0.32-0.48)
cg12873476	8	x[a]	-1,71%	1,28E-11	-1,34%	3,98E-03	0.6959 (0.67-0.72)	0.6825 (0.65-0.71)	0.6787 (0.65-0.71)
cg11827514	8	x[a]	-4,40%	6,94E-09	-1,47%	5,85E-03	0.8756 (0.82-0.91)	0.8610 (0.79-0.91)	0.8316 (0.74-0.89)
cg11848483	8	ZC3H3	-1,46%	4,93E-09	0,07%	1,54E-01	0.8802 (0.84-0.92)	0.8809 (0.84-0.92)	0.8656 (0.83-0.90)
cg20156237	8	ZC3H3	-1,30%	4,01E-08	0,42%	4,98E-01	0.8155 (0.77-0.87)	0.8197 (0.77-0.87)	0.8025 (0.75-0.85)
cg26361535	8	ZC3H3	-5,54%	4,14E-09	-1,77%	7,48E-04	0.7593 (0.70-0.81)	0.7417 (0.69-0.80)	0.7040 (0.63-0.77)
cg14069088	9	CDKN2BAS	1,54%	5,10E-08	1,06%	5,87E-02	0.2475 (0.21-0.28)	0.2581 (0.22-0.30)	0.2630 (0.22-0.30)
cg12971694	9	CD72	0,56%	1,79E-08	0,57%	1,82E-01	0.1741 (0.16-0.19)	0.1798 (0.16-0.20)	0.1797 (0.16-0.20)
cg13832290	9	SIT1	-0,72%	1,01E-08	0,08%	7,00E-01	0.7764 (0.75-0.81)	0.7772 (0.75-0.81)	0.7692 (0.74-0.80)
cg14178364	9	FBXO10	-1,02%	4,55E-08	0,43%	5,49E-01	0.8686 (0.84-0.90)	0.8728 (0.84-0.90)	0.8583 (0.82-0.89)
cg14235811	9	MCART1	-0,17%	2,79E-08	0,27%	1,25E-01	0.0917 (0.08-0.10)	0.0944 (0.08-0.10)	0.0900 (0.08-0.10)
cg14270346	9	SHB	-1,28%	4,26E-08	-0,49%	1,38E-01	0.9226 (0.91-0.94)	0.9177 (0.90-0.93)	0.9099 (0.89-0.93)
cg13787850	9	x[a]	-3,10%	2,14E-10	-0,50%	6,41E-01	0.2680 (0.21-0.33)	0.2630 (0.21-0.32)	0.2370 (0.18-0.29)
cg01692968	9	x[a]	-1,42%	7,30E-11	-1,01%	2,80E-06	0.1619 (0.15-0.18)	0.1518 (0.14-0.17)	0.1477 (0.13-0.16)
cg02388253	9	PHF19	-0,50%	2,88E-09	-0,12%	4,39E-03	0.1314 (0.12-0.15)	0.1302 (0.12-0.15)	0.1264 (0.11-0.14)
cg13501527	9	PHF19	-0,61%	9,68E-09	0,25%	5,40E-02	0.3840 (0.35-0.42)	0.3864 (0.35-0.42)	0.3779 (0.34-0.41)
cg04236137	9	x[a]	-2,03%	9,54E-12	1,14%	8,92E-01	0.5747 (0.51-0.65)	0.5861 (0.52-0.66)	0.5545 (0.48-0.61)
cg13418576	9	x[a]	-3,59%	4,65E-09	-0,37%	7,88E-01	0.2705 (0.22-0.33)	0.2669 (0.22-0.33)	0.2347 (0.18-0.29)
cg13618969	9	FAM125B	0,67%	8,49E-08	1,13%	1,00E-01	0.4516 (0.37-0.54)	0.4629 (0.39-0.54)	0.4583 (0.39-0.55)
cg13992008	9	FAM102A	-1,96%	6,61E-10	0,02%	2,87E-01	0.7934 (0.74-0.84)	0.7936 (0.75-0.85)	0.7738 (0.73-0.82)
cg13910681	9	FAM102A	-0,41%	5,92E-08	0,19%	5,01E-01	0.0782 (0.07-0.08)	0.0801 (0.07-0.09)	0.0741 (0.07-0.08)
cg06901890	9	FNBP1	-0,36%	3,89E-10	-0,04%	2,01E-01	0.0869 (0.08-0.09)	0.0865 (0.08-0.10)	0.0834 (0.07-0.09)
cg02257517	9	SLC2A6	1,22%	5,57E-09	0,44%	4,90E-04	0.2841 (0.27-0.30)	0.2885 (0.27-0.31)	0.2963 (0.28-0.31)
cg14108380	9	SDCCAG3	-0,91%	3,88E-08	0,27%	7,19E-01	0.8861 (0.86-0.91)	0.8888 (0.86-0.91)	0.8770 (0.86-0.90)
cg14435720	9	EGFL7/MIR126	-0,52%	4,12E-08	0,19%	2,90E-01	0.9545 (0.94-0.97)	0.9564 (0.95-0.97)	0.9493 (0.93-0.96)
cg22539182	10	x[a]	1,41%	7,40E-13	0,43%	7,99E-03	0.1902 (0.18-0.20)	0.1945 (0.18-0.21)	0.2044 (0.19-0.22)
cg05931497	10	LARP4B	0,72%	6,69E-08	0,23%	3,99E-01	0.2487 (0.20-0.29)	0.2510 (0.21-0.29)	0.2559 (0.22-0.30)
cg03444838	10	WDR37	-1,23%	7,69E-09	0,07%	7,96E-01	0.8434 (0.80-0.89)	0.8441 (0.80-0.89)	0.8311 (0.79-0.87)
cg25529535	10	x[a]	-1,35%	3,62E-08	-0,32%	1,32E-01	0.6749 (0.63-0.71)	0.6716 (0.63-0.72)	0.6613 (0.62-0.70)
cg19726711	10	x[a]	-0,56%	6,36E-08	0,17%	9,54E-01	0.7535 (0.70-0.81)	0.7551 (0.71-0.81)	0.7479 (0.70-0.80)
cg11430077	10	GATA3	-0,47%	2,58E-08	-0,23%	4,01E-02	0.0612 (0.05-0.07)	0.0589 (0.05-0.07)	0.0565 (0.05-0.07)
cg00463367	10	GATA3	-0,42%	2,74E-08	-0,14%	2,99E-01	0.0662 (0.06-0.08)	0.0648 (0.06-0.07)	0.0620 (0.06-0.07)
cg17256364	10	ST8SIA6	0,34%	9,19E-08	0,23%	7,31E-01	0.1102 (0.10-0.12)	0.1125 (0.10-0.12)	0.1136 (0.10-0.12)
cg04813697	10	PIP4K2A	-1,91%	3,69E-10	-0,98%	2,08E-02	0.5244 (0.48-0.56)	0.5146 (0.47-0.55)	0.5053 (0.46-0.54)
cg10317175	10	x[a]	-0,73%	1,36E-08	0,02%	8,79E-03	0.1697 (0.15-0.20)	0.1699 (0.15-0.19)	0.1624 (0.14-0.19)
cg25212025	10	PARD3	-3,73%	5,07E-14	-0,10%	3,97E-03	0.2477 (0.20-0.32)	0.2467 (0.20-0.29)	0.2103 (0.17-0.27)
cg06112654	10	x[a]	0,51%	9,59E-10	0,16%	1,32E-01	0.0926 (0.08-0.10)	0.0941 (0.08-0.10)	0.0976 (0.09-0.11)
cg04641860	10	HNRNPF	-1,66%	6,60E-08	0,96%	8,53E-01	0.5005 (0.43-0.58)	0.5101 (0.43-0.60)	0.4839 (0.42-0.57)
cg25953130	10	ARID5B	-4,69%	6,34E-10	-0,87%	2,95E-03	0.4928 (0.41-0.56)	0.4841 (0.40-0.56)	0.4459 (0.36-0.52)
cg01005506	10	ADO	0,33%	4,41E-09	1,07%	1,78E-02	0.1505 (0.14-0.17)	0.1611 (0.14-0.18)	0.1538 (0.14-0.17)
cg12474798	10	ADO	0,74%	1,09E-08	3,09%	6,84E-02	0.5070 (0.45-0.56)	0.5379 (0.48-0.58)	0.5144 (0.46-0.57)
cg00210249	10	HK1	5,59%	2,26E-13	0,70%	6,73E-02	0.7513 (0.69-0.80)	0.7583 (0.70-0.82)	0.8072 (0.74-0.85)
cg15013801	10	ASCC1/C10orf104	-0,74%	7,88E-09	0,04%	5,30E-01	0.9012 (0.89-0.92)	0.9017 (0.89-0.91)	0.8938 (0.88-0.91)
cg12147622	10	x[a]	-4,27%	1,08E-19	-1,14%	5,68E-03	0.3941 (0.34-0.45)	0.3826 (0.33-0.44)	0.3514 (0.30-0.40)
cg01839993	10	DDIT4	0,05%	5,65E-11	0,73%	8,75E-02	0.1753 (0.16-0.19)	0.1826 (0.17-0.20)	0.1758 (0.16-0.19)
cg10592592	10	DDIT4	-0,08%	2,24E-08	0,64%	1,41E-01	0.1514 (0.14-0.17)	0.1578 (0.14-0.17)	0.1506 (0.13-0.17)
cg02370832	10	IFIT3	-1,51%	2,48E-08	-0,40%	4,16E-01	0.6762 (0.63-0.73)	0.6723 (0.63-0.72)	0.6611 (0.61-0.71)
cg20180364	10	HHEX	0,41%	1,86E-09	0,21%	2,46E-01	0.1237 (0.11-0.14)	0.1259 (0.11-0.14)	0.1278 (0.12-0.14)
cg27312979	10	SORBS1	5,02%	2,74E-16	1,57%	6,47E-03	0.5558 (0.51-0.60)	0.5715 (0.52-0.61)	0.6061 (0.57-0.65)
cg25421530	10	SORBS1	1,60%	4,22E-09	0,23%	4,30E-01	0.8762 (0.84-0.90)	0.8786 (0.85-0.91)	0.8922 (0.86-0.92)
cg13383814	10	C10orf12	0,34%	8,67E-09	0,07%	7,45E-01	0.0945 (0.08-0.11)	0.0952 (0.08-0.11)	0.0979 (0.09-0.11)
cg17296078	10	MMS19/UBTD1	0,76%	3,36E-08	0,41%	2,15E-01	0.1801 (0.16-0.20)	0.1842 (0.16-0.21)	0.1877 (0.17-0.21)
cg17333042	10	KAZALD1	0,37%	2,14E-08	0,67%	4,66E-01	0.1818 (0.17-0.20)	0.1885 (0.18-0.20)	0.1855 (0.17-0.20)
cg09701700	10	MIR146B	-0,69%	3,40E-10	-0,29%	4,78E-02	0.1325 (0.11-0.16)	0.1296 (0.11-0.16)	0.1256 (0.11-0.15)
cg07631309	10	GSTO1	-3,23%	2,45E-08	-0,88%	1,39E-02	0.4471 (0.37-0.50)	0.4383 (0.37-0.50)	0.4148 (0.34-0.49)
cg05329352	10	ADRA2A	-2,54%	6,72E-11	-1,05%	5,86E-03	0.3008 (0.28-0.33)	0.2903 (0.27-0.32)	0.2755 (0.25-0.29)
cg07743805	10	x[a]	1,05%	2,25E-10	0,50%	9,88E-02	0.1625 (0.15-0.17)	0.1674 (0.16-0.18)	0.1729 (0.16-0.18)
cg15806304	10	SLC18A2	3,07%	5,92E-08	0,36%	3,83E-02	0.4987 (0.45-0.54)	0.5023 (0.46-0.55)	0.5294 (0.47-0.58)
cg27009448	10	C10orf46	-1,69%	2,01E-08	-0,22%	1,11E-01	0.9031 (0.87-0.93)	0.9009 (0.86-0.93)	0.8862 (0.85-0.92)
cg00291478	10	RGS10	-3,23%	1,28E-09	-0,84%	2,74E-01	0.4976 (0.42-0.58)	0.4892 (0.42-0.57)	0.4653 (0.37-0.54)
cg07858728	10	DMBT1	0,47%	1,92E-09	0,22%	2,10E-01	0.0654 (0.06-0.07)	0.0676 (0.06-0.08)	0.0701 (0.06-0.08)
cg04726013	10	LHPP	0,60%	7,67E-08	0,04%	6,90E-01	0.1437 (0.12-0.17)	0.1441 (0.12-0.17)	0.1497 (0.13-0.18)
cg00424785	10	CTBP2	-0,80%	2,38E-08	0,10%	8,75E-01	0.9291 (0.91-0.95)	0.9301 (0.91-0.94)	0.9210 (0.90-0.94)
cg08881175	10	x[a]	0,70%	1,03E-09	0,34%	1,70E-01	0.0835 (0.07-0.10)	0.0869 (0.07-0.10)	0.0906 (0.08-0.10)
cg15705175	10	x[a]	0,58%	4,41E-08	0,44%	1,38E-02	0.1725 (0.16-0.18)	0.1768 (0.17-0.19)	0.1783 (0.17-0.19)
cg01971407	11	IFITM1	-0,85%	1,03E-09	0,47%	7,77E-01	0.3805 (0.32-0.42)	0.3852 (0.33-0.43)	0.3720 (0.30-0.42)
cg20045320	11	x[a]	-0,48%	1,55E-08	0,73%	1,77E-01	0.6058 (0.57-0.64)	0.6131 (0.58-0.65)	0.6011 (0.56-0.64)
cg09162537	11	AP2A2	1,37%	5,33E-08	0,81%	7,52E-02	0.1465 (0.14-0.16)	0.1546 (0.14-0.17)	0.1602 (0.15-0.17)
cg14486857	11	AP2A2	9,30%	2,02E-10	1,12%	1,15E-01	0.7304 (0.62-0.84)	0.7416 (0.64-0.85)	0.8234 (0.72-0.89)
cg11095027	11	TOLLIP	2,54%	2,22E-09	0,02%	3,71E-01	0.3772 (0.34-0.41)	0.3774 (0.35-0.41)	0.4026 (0.37-0.44)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg04258133	11	TNNT3	1,69%	4,54E-13	-0,05%	1,66E-01	0.8174 (0.79-0.84)	0.8169 (0.79-0.84)	0.8343 (0.81-0.86)
cg19670883	11	KCNQ1	0,59%	8,18E-09	0,25%	5,97E-01	0.0961 (0.09-0.11)	0.0986 (0.09-0.11)	0.1020 (0.09-0.12)
cg07824422	11	KCNQ1	1,05%	8,29E-08	0,74%	1,13E-01	0.1914 (0.18-0.20)	0.1988 (0.19-0.21)	0.2019 (0.19-0.21)
cg07556018	11	KCNQ1	2,69%	1,76E-12	0,47%	1,06E-01	0.2299 (0.19-0.28)	0.2345 (0.20-0.28)	0.2568 (0.22-0.31)
cg13428066	11	KCNQ1	1,80%	8,59E-10	0,43%	3,25E-01	0.1819 (0.16-0.21)	0.1862 (0.16-0.22)	0.1999 (0.17-0.24)
cg01744331	11	KCNQ1OT1	-1,21%	3,51E-10	-0,38%	2,24E-03	0.9380 (0.93-0.95)	0.9342 (0.92-0.95)	0.9259 (0.91-0.94)
cg07123182	11	KCNQ1OT1	-0,91%	8,64E-13	-0,22%	2,32E-03	0.9527 (0.94-0.96)	0.9505 (0.94-0.96)	0.9436 (0.93-0.95)
cg16556677	11	KCNQ1OT1	-3,84%	6,58E-10	-2,06%	2,29E-07	0.8507 (0.81-0.88)	0.8301 (0.79-0.86)	0.8123 (0.77-0.85)
cg26963277	11	KCNQ1OT1	-2,86%	4,52E-22	-1,01%	1,59E-07	0.9013 (0.88-0.92)	0.8912 (0.87-0.91)	0.8726 (0.84-0.90)
cg24497361	11	RHOG	-1,32%	8,08E-12	-0,23%	2,85E-01	0.1546 (0.13-0.18)	0.1523 (0.13-0.18)	0.1414 (0.12-0.17)
cg04039799	11	NAV2	-1,62%	5,24E-12	-0,46%	1,35E-01	0.1954 (0.17-0.23)	0.1908 (0.16-0.22)	0.1792 (0.16-0.20)
cg05533340	11	PRMT3	1,02%	1,20E-09	-0,14%	2,70E-01	0.9016 (0.87-0.92)	0.9002 (0.87-0.92)	0.9118 (0.88-0.93)
cg25557175	11	x[a]	-0,62%	1,80E-08	-0,04%	6,25E-01	0.1104 (0.10-0.12)	0.1100 (0.10-0.12)	0.1043 (0.09-0.12)
cg24453664	11	CD59	1,08%	4,63E-08	0,51%	4,90E-01	0.1793 (0.16-0.20)	0.1844 (0.16-0.21)	0.1902 (0.16-0.22)
cg16568681	11	CD59	0,86%	4,12E-08	0,18%	8,57E-01	0.1545 (0.14-0.18)	0.1563 (0.14-0.18)	0.1631 (0.14-0.18)
cg02211741	11	DGKZ	0,62%	6,94E-09	-0,12%	9,73E-01	0.1311 (0.12-0.15)	0.1300 (0.12-0.15)	0.1373 (0.12-0.16)
cg14099685	11	CUGBP1	-0,67%	2,65E-11	-0,07%	4,30E-01	0.1320 (0.12-0.15)	0.1313 (0.12-0.15)	0.1253 (0.11-0.14)
cg26468891	11	PTPRJ	-1,38%	1,20E-08	-1,27%	5,35E-04	0.2575 (0.21-0.31)	0.2449 (0.20-0.29)	0.2437 (0.19-0.29)
cg09197783	11	SLC43A3	-3,04%	5,23E-08	-0,68%	4,05E-02	0.7554 (0.72-0.80)	0.7486 (0.71-0.79)	0.7250 (0.69-0.76)
cg16611234	11	x[a]	-1,25%	8,31E-12	0,10%	3,50E-03	0.1449 (0.13-0.16)	0.1459 (0.13-0.16)	0.1323 (0.12-0.15)
cg00706994	11	OR5A2	-2,09%	1,01E-08	-0,08%	2,12E-01	0.6654 (0.62-0.71)	0.6645 (0.61-0.71)	0.6445 (0.60-0.69)
cg19254163	11	GPR44	-3,88%	4,82E-09	-1,79%	1,59E-03	0.5778 (0.54-0.61)	0.5599 (0.52-0.60)	0.5390 (0.50-0.58)
cg11465943	11	x[a]	-1,03%	9,51E-12	0,01%	1,18E-01	0.7588 (0.73-0.79)	0.7589 (0.73-0.79)	0.7485 (0.72-0.77)
cg24068972	11	x[a]	-0,66%	1,60E-11	0,27%	1,13E-01	0.7355 (0.71-0.77)	0.7382 (0.71-0.76)	0.7289 (0.70-0.75)
cg13692972	11	OTUB1	1,09%	3,96E-08	0,26%	7,80E-02	0.1840 (0.16-0.21)	0.1866 (0.16-0.22)	0.1950 (0.17-0.23)
cg07019638	11	EHD1	-0,47%	2,39E-09	0,36%	6,06E-01	0.8255 (0.79-0.86)	0.8292 (0.80-0.86)	0.8208 (0.79-0.85)
cg20889322	11	x[a]	-2,02%	3,40E-13	0,39%	4,17E-01	0.2809 (0.25-0.31)	0.2849 (0.26-0.32)	0.2608 (0.24-0.30)
cg05363534	11	MIR612	1,01%	4,17E-08	-0,03%	1,69E-01	0.2044 (0.17-0.24)	0.2041 (0.17-0.25)	0.2145 (0.18-0.26)
cg07029024	11	x[a]	3,57%	7,40E-09	0,07%	8,80E-01	0.4619 (0.39-0.54)	0.4626 (0.39-0.53)	0.4976 (0.43-0.56)
cg02740606	11	CORO1B	-0,86%	8,65E-09	0,05%	5,38E-01	0.7296 (0.70-0.76)	0.7301 (0.70-0.76)	0.7210 (0.69-0.75)
cg19528654	11	AIP	-1,57%	4,43E-09	-0,02%	2,06E-01	0.7418 (0.68-0.80)	0.7415 (0.67-0.81)	0.7260 (0.67-0.78)
cg01972009	11	PITPNM1	-0,12%	6,93E-09	0,05%	4,02E-02	0.1095 (0.10-0.12)	0.1099 (0.10-0.12)	0.1083 (0.10-0.12)
cg04383058	11	PITPNM1	-0,97%	2,41E-10	0,39%	5,13E-01	0.3660 (0.33-0.39)	0.3699 (0.34-0.40)	0.3563 (0.32-0.39)
cg24051242	11	LRP5	0,69%	1,09E-08	0,10%	3,06E-01	0.1182 (0.10-0.13)	0.1192 (0.10-0.14)	0.1251 (0.11-0.14)
cg21611682	11	LRP5	-5,24%	1,09E-18	-1,05%	8,14E-02	0.5276 (0.50-0.56)	0.5171 (0.48-0.55)	0.4752 (0.43-0.52)
cg14624207	11	LRP5	-5,15%	2,94E-12	-1,70%	3,59E-02	0.4747 (0.44-0.51)	0.4577 (0.42-0.50)	0.4233 (0.37-0.47)
cg11152384	11	x[a]	-0,40%	8,66E-08	-0,22%	5,94E-02	0.0546 (0.05-0.06)	0.0524 (0.04-0.06)	0.0506 (0.04-0.06)
cg08885142	11	x[a]	1,39%	2,87E-09	0,72%	1,87E-01	0.3325 (0.30-0.36)	0.3396 (0.31-0.37)	0.3463 (0.31-0.37)
cg16061668	11	CTTN	0,57%	1,38E-16	0,12%	1,27E-01	0.1000 (0.09-0.11)	0.1012 (0.09-0.11)	0.1057 (0.10-0.12)
cg01901332	11	ARRB1	-5,07%	1,70E-09	-1,38%	6,67E-02	0.6741 (0.63-0.72)	0.6603 (0.61-0.72)	0.6234 (0.57-0.68)
cg17472111	11	x[a]	0,65%	2,22E-10	0,40%	4,37E-02	0.1074 (0.09-0.12)	0.1113 (0.10-0.12)	0.1139 (0.10-0.13)
cg10788371	11	LRRC32	-1,65%	7,16E-09	-0,33%	3,40E-02	0.2693 (0.24-0.29)	0.2660 (0.24-0.29)	0.2529 (0.23-0.28)
cg13985437	11	LRRC32	-0,32%	1,19E-09	0,15%	1,90E-02	0.1070 (0.10-0.12)	0.1085 (0.10-0.12)	0.1038 (0.10-0.11)
cg07202214	11	LRRC32	-0,49%	1,88E-09	0,03%	3,26E-02	0.1138 (0.10-0.12)	0.1141 (0.10-0.12)	0.1089 (0.10-0.12)
cg20886049	11	TSKU	-1,81%	3,53E-08	-1,35%	5,79E-04	0.5383 (0.48-0.59)	0.5249 (0.47-0.58)	0.5203 (0.46-0.57)
cg06562283	11	ODZ4	-1,26%	2,19E-08	-0,08%	5,48E-01	0.4806 (0.43-0.53)	0.4798 (0.44-0.52)	0.4680 (0.42-0.51)
cg11660018*	11	PRSS23	-3,87%	1,29E-22	-1,67%	1,28E-10	0.2926 (0.27-0.32)	0.2759 (0.25-0.30)	0.2538 (0.23-0.28)
cg23771366*	11	PRSS23	-2,26%	7,62E-20	-0,98%	5,55E-11	0.1993 (0.18-0.22)	0.1895 (0.17-0.21)	0.1767 (0.16-0.19)
cg24088496	11	MAML2	-3,17%	3,15E-08	1,25%	9,28E-01	0.2981 (0.24-0.36)	0.3106 (0.25-0.38)	0.2664 (0.22-0.33)
cg06106428	11	ARHGAP20	5,54%	2,49E-10	1,02%	6,12E-01	0.3645 (0.31-0.43)	0.3747 (0.33-0.43)	0.4199 (0.37-0.48)
cg03597491	11	ZBTB16	1,34%	6,20E-09	0,14%	6,78E-01	0.2355 (0.21-0.27)	0.2369 (0.21-0.27)	0.2489 (0.22-0.29)
cg04226002	11	ZBTB16	0,64%	9,54E-08	0,19%	2,60E-01	0.1466 (0.13-0.16)	0.1484 (0.13-0.17)	0.1530 (0.13-0.17)
cg10827488	11	ZBTB16	0,78%	1,43E-16	0,35%	7,48E-02	0.0892 (0.08-0.10)	0.0927 (0.08-0.10)	0.0970 (0.08-0.11)
cg17518710	11	ZBTB16	0,85%	7,73E-11	0,38%	2,63E-01	0.1126 (0.10-0.13)	0.1164 (0.10-0.13)	0.1211 (0.11-0.14)
cg03725573	11	ZBTB16	1,73%	1,18E-08	0,24%	9,93E-02	0.7858 (0.76-0.82)	0.7882 (0.76-0.82)	0.8032 (0.77-0.83)
cg03044513	11	x[a]	0,49%	4,67E-08	0,04%	8,45E-01	0.1565 (0.14-0.17)	0.1570 (0.15-0.17)	0.1614 (0.15-0.17)
cg03234777	11	AMICA1	-0,69%	5,16E-15	0,04%	9,72E-02	0.0561 (0.05-0.07)	0.0565 (0.05-0.07)	0.0492 (0.04-0.06)
cg20483374	11	C1QTNF5/MFRP	0,78%	6,19E-08	0,63%	7,85E-01	0.1616 (0.15-0.17)	0.1679 (0.16-0.18)	0.1694 (0.16-0.18)
cg13956924	11	x[a]	0,34%	8,93E-10	0,02%	7,87E-02	0.0882 (0.08-0.10)	0.0884 (0.08-0.10)	0.0916 (0.08-0.10)
cg12689529	11	KIRREL3	-0,36%	6,28E-08	-0,07%	2,57E-01	0.0535 (0.05-0.06)	0.0528 (0.05-0.06)	0.0499 (0.04-0.06)
cg25068347	11	ETS1	-1,52%	1,32E-09	0,60%	5,10E-01	0.7590 (0.71-0.81)	0.7650 (0.71-0.82)	0.7438 (0.69-0.79)
cg03295554	11	ETS1	-5,55%	3,20E-08	-1,63%	6,78E-02	0.5605 (0.48-0.63)	0.5441 (0.44-0.63)	0.5049 (0.41-0.58)
cg16983588	11	PRDM10	6,45%	3,72E-08	0,39%	1,47E-01	0.3331 (0.27-0.41)	0.3370 (0.27-0.41)	0.3975 (0.31-0.49)
cg09084200	11	VPS26B/NCAPD3	-0,53%	5,29E-15	0,08%	1,15E-03	0.1485 (0.14-0.16)	0.1493 (0.14-0.16)	0.1432 (0.13-0.15)
cg26282236	12	RAD52	2,46%	8,24E-08	-0,16%	2,53E-01	0.5699 (0.52-0.63)	0.5683 (0.52-0.62)	0.5945 (0.55-0.65)
cg17150898	12	FBXL14	-0,58%	1,98E-08	0,59%	8,60E-01	0.6895 (0.66-0.72)	0.6954 (0.66-0.73)	0.6838 (0.65-0.71)
cg19369955	12	x[a]	1,20%	6,30E-08	0,35%	4,53E-01	0.2430 (0.20-0.28)	0.2465 (0.21-0.29)	0.2550 (0.22-0.30)
cg03131366	12	x[a]	2,03%	1,61E-09	0,65%	1,31E-01	0.1441 (0.13-0.17)	0.1506 (0.13-0.18)	0.1644 (0.14-0.19)
cg07986199	12	CACNA1C	-0,48%	1,17E-08	0,37%	4,65E-02	0.5055 (0.48-0.53)	0.5092 (0.48-0.53)	0.5007 (0.47-0.52)
cg23219570	12	FGF23	0,46%	3,96E-08	0,21%	1,23E-02	0.0896 (0.08-0.10)	0.0917 (0.08-0.10)	0.0941 (0.08-0.11)
cg26276120	12	TPI1	-0,16%	1,68E-10	0,08%	9,32E-03	0.0924 (0.08-0.10)	0.0932 (0.08-0.10)	0.0908 (0.08-0.10)
cg22335340	12	PTPN6/C12orf57	-2,03%	1,84E-10	-0,39%	1,84E-01	0.7771 (0.74-0.81)	0.7732 (0.74-0.81)	0.7568 (0.71-0.79)
cg21990700	12	LOC283314/C1RL	-0,41%	9,61E-09	-0,26%	3,70E-02	0.0963 (0.08-0.11)	0.0937 (0.08-0.11)	0.0922 (0.08-0.10)
cg06516865	12	ETV6	-0,55%	2,87E-09	-0,13%	3,43E-01	0.1171 (0.10-0.14)	0.1158 (0.10-0.14)	0.1116 (0.10-0.13)
cg00046744	12	LRP6	0,47%	4,38E-11	0,25%	6,02E-02	0.0831 (0.07-0.09)	0.0856 (0.08-0.09)	0.0878 (0.08-0.10)
cg06538684	12	LOH12CR2	-1,30%	1,22E-09	0,87%	6,87E-01	0.3695 (0.32-0.42)	0.3782 (0.33-0.43)	0.3566 (0.31-0.41)
cg06826457	12	x[a]	-0,52%	3,26E-10	-0,30%	2,69E-02	0.0998 (0.09-0.12)	0.0968 (0.08-0.11)	0.0946 (0.08-0.11)
cg12496975	12	RAPGEF3	-1,06%	2,56E-08	-0,05%	9,56E-01	0.8652 (0.83-0.89)	0.8646 (0.84-0.89)	0.8546 (0.82-0.88)
cg08140412	12	SCN8A	-1,82%	8,90E-09	-0,72%	9,29E-02	0.4331 (0.37-0.48)	0.4258 (0.37-0.48)	0.4148 (0.36-0.47)
cg22566906	12	GRASP	0,60%	1,10E-09	0,23%	2,21E-01	0.1162 (0.10-0.13)	0.1185 (0.11-0.13)	0.1222 (0.11-0.13)
cg15074033	12	x[a]	1,18%	2,35E-08	0,47%	8,04E-01	0.0963 (0.08-0.12)	0.1010 (0.08-0.12)	0.1081 (0.09-0.13)
cg24504361	12	KRT8	-1,60%	4,03E-10	-0,56%	2,90E-03	0.2955 (0.26-0.32)	0.2900 (0.26-0.32)	0.2795 (0.25-0.31)
cg10843343	12	KRT8	-1,19%	1,57E-08	-0,34%	1,14E-03	0.6159 (0.59-0.65)	0.6125 (0.58-0.63)	0.6040 (0.56-0.63)
cg13937905	12	RARG	-2,09%	1,76E-12	-0,33%	1,24E-03	0.8769 (0.84-0.91)	0.8736 (0.83-0.90)	0.8560 (0.82-0.88)
cg10592478	12	RARG	-0,93%	4,46E-12	-0,32%	6,99E-05	0.9229 (0.91-0.94)	0.9197 (0.90-0.93)	0.9135 (0.90-0.93)
cg02583484	12	HNRNPA1	-2,09%	7,10E-20	-0,54%	9,24E-04	0.1476 (0.12-0.18)	0.1422 (0.12-0.17)	0.1267 (0.11-0.15)
cg04158018	12	NFE2	-1,25%	8,40E-08	0,17%	6,20E-01	0.1587 (0.13-0.19)	0.1603 (0.14-0.19)	0.1462 (0.13-0.17)
cg26477856	12	DGKA	-0,76%	3,25E-08	0,12%	6,77E-01	0.2732 (0.25-0.30)	0.2744 (0.25-0.30)	0.2656 (0.24-0.29)
cg22876908	12	DGKA	-0,05%	1,30E-08	0,02%	3,76E-02	0.1784 (0.16-0.19)	0.1786 (0.16-0.19)	0.1779 (0.16-0.19)
cg06739462	12	DGKA	-0,66%	1,40E-08	0,09%	2,56E-02	0.8903 (0.85-0.92)	0.8911 (0.85-0.92)	0.8836 (0.84-0.91)
cg23378033	12	MSRB3	0,19%	8,98E-08	0,19%	6,92E-01	0.1416 (0.13-0.16)	0.1435 (0.13-0.16)	0.1435 (0.13-0.16)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg11649376	12	ACSS3	-0,53%	5,24E-10	-0,02%	2,60E-01	0.4789 (0.44-0.52)	0.4787 (0.44-0.52)	0.4736 (0.42-0.51)
cg17740822	12	DUSP6	0,32%	2,05E-08	0,15%	5,97E-01	0.0896 (0.08-0.10)	0.0911 (0.08-0.10)	0.0927 (0.08-0.10)
cg15408080	12	x[a]	-0,63%	5,32E-08	0,31%	2,19E-01	0.1628 (0.13-0.21)	0.1659 (0.14-0.21)	0.1565 (0.13-0.20)
cg07905054	12	DRAM1	0,22%	6,00E-08	0,27%	2,75E-01	0.1177 (0.11-0.13)	0.1204 (0.11-0.13)	0.1199 (0.11-0.14)
cg24124443	12	BTBD11	1,89%	9,86E-08	0,88%	3,91E-01	0.1615 (0.13-0.20)	0.1703 (0.14-0.21)	0.1804 (0.14-0.23)
cg14642045	12	UNG	0,32%	2,00E-09	-0,32%	5,04E-01	0.3377 (0.28-0.39)	0.3345 (0.29-0.38)	0.3409 (0.31-0.39)
cg11930955	12	TPCN1/IQCD	-0,01%	9,40E-08	0,48%	5,61E-01	0.1603 (0.15-0.17)	0.1651 (0.15-0.18)	0.1602 (0.15-0.17)
cg13645530	12	x[a]	1,73%	4,67E-08	0,46%	3,54E-01	0.7907 (0.76-0.82)	0.7954 (0.77-0.83)	0.8081 (0.78-0.83)
cg04682193	12	KDM2B	0,84%	7,98E-09	-0,25%	3,25E-01	0.1418 (0.12-0.16)	0.1394 (0.12-0.17)	0.1503 (0.13-0.18)
cg12800105	12	KDM2B	-1,26%	3,49E-08	0,07%	6,69E-01	0.8548 (0.82-0.89)	0.8555 (0.82-0.89)	0.8422 (0.80-0.88)
cg08120831	12	LRRC43	0,92%	9,82E-09	0,35%	7,92E-01	0.1546 (0.14-0.17)	0.1581 (0.14-0.18)	0.1638 (0.14-0.18)
cg05298628	12	VPS37B	0,48%	3,30E-08	0,38%	2,56E-01	0.1207 (0.11-0.13)	0.1245 (0.11-0.13)	0.1255 (0.12-0.14)
cg20181887	12	CDK2AP1	1,28%	5,29E-08	0,65%	5,09E-01	0.5274 (0.48-0.57)	0.5339 (0.50-0.57)	0.5402 (0.50-0.58)
cg04347477	12	NCOR2	4,08%	2,78E-09	4,25%	4,35E-01	0.3905 (0.35-0.45)	0.4330 (0.37-0.48)	0.4314 (0.37-0.48)
cg04929759	12	x[a]	0,88%	7,70E-09	0,49%	4,31E-01	0.1901 (0.15-0.22)	0.1950 (0.16-0.23)	0.1989 (0.16-0.24)
cg00893603	13	ATP8A2	-4,63%	7,11E-11	-1,91%	1,36E-02	0.7660 (0.69-0.83)	0.7469 (0.66-0.82)	0.7197 (0.61-0.79)
cg23681440	13	x[a]	-1,31%	1,12E-08	-0,38%	1,84E-02	0.2499 (0.21-0.30)	0.2461 (0.21-0.29)	0.2368 (0.20-0.28)
cg26885400	13	x[a]	1,05%	6,33E-09	0,27%	3,91E-01	0.1720 (0.15-0.20)	0.1746 (0.15-0.21)	0.1825 (0.16-0.22)
cg07756788	13	x[a]	-1,47%	1,02E-08	0,10%	1,51E-01	0.8790 (0.85-0.91)	0.8800 (0.84-0.91)	0.8644 (0.83-0.89)
cg24366665	13	x[a]	-1,15%	5,88E-11	-0,35%	1,76E-01	0.8862 (0.85-0.91)	0.8827 (0.85-0.91)	0.8747 (0.84-0.90)
cg12836863	13	BRCA2	1,23%	5,06E-11	-0,08%	1,19E-01	0.1814 (0.16-0.20)	0.1806 (0.16-0.21)	0.1936 (0.17-0.22)
cg20749792	13	x[a]	-1,60%	5,63E-10	-0,35%	1,13E-01	0.6971 (0.64-0.75)	0.6936 (0.64-0.75)	0.6811 (0.63-0.73)
cg06133910	13	x[a]	0,04%	9,47E-09	0,02%	5,81E-01	0.7536 (0.72-0.79)	0.7538 (0.72-0.79)	0.7540 (0.72-0.78)
cg21565496	13	x[a]	-0,68%	2,28E-10	0,11%	1,43E-01	0.8597 (0.83-0.89)	0.8608 (0.83-0.89)	0.8528 (0.82-0.87)
cg06329392	13	x[a]	-1,27%	5,58E-12	-0,08%	6,70E-02	0.8356 (0.80-0.87)	0.8348 (0.80-0.87)	0.8229 (0.79-0.85)
cg22488891	13	x[a]	-1,03%	8,13E-09	-0,07%	1,01E-01	0.7825 (0.75-0.82)	0.7818 (0.75-0.82)	0.7723 (0.74-0.81)
cg06648759	13	x[a]	1,48%	7,96E-09	-0,72%	1,46E-01	0.5338 (0.48-0.58)	0.5266 (0.48-0.57)	0.5486 (0.51-0.59)
cg11978634	13	ENOX1	2,79%	4,15E-08	0,52%	8,62E-02	0.8034 (0.75-0.85)	0.8087 (0.76-0.85)	0.8313 (0.78-0.87)
cg03646329	13	LPAR6/RB1	-3,94%	1,96E-08	-0,71%	3,09E-03	0.7780 (0.72-0.84)	0.7709 (0.70-0.84)	0.7386 (0.67-0.80)
cg02003272	13	x[a]	-0,80%	6,41E-10	0,07%	3,93E-01	0.9056 (0.89-0.92)	0.9063 (0.89-0.92)	0.8976 (0.89-0.91)
cg00843117	13	x[a]	-0,32%	6,86E-11	-0,02%	1,06E-01	0.9658 (0.96-0.97)	0.9656 (0.96-0.97)	0.9627 (0.95-0.97)
cg27467282	13	x[a]	1,76%	9,49E-10	0,34%	2,08E-02	0.8900 (0.86-0.91)	0.8933 (0.86-0.92)	0.9076 (0.89-0.92)
cg23126342	13	PCDH9	9,11%	7,16E-14	5,54%	9,74E-07	0.5136 (0.43-0.60)	0.5690 (0.48-0.65)	0.6047 (0.52-0.67)
cg25491122	13	PCDH9	2,36%	1,49E-08	1,10%	1,48E-04	0.9053 (0.86-0.93)	0.9163 (0.88-0.94)	0.9289 (0.90-0.95)
cg16969872	13	RBM26	-2,57%	3,92E-11	-0,37%	2,63E-02	0.7776 (0.74-0.82)	0.7739 (0.72-0.82)	0.7520 (0.68-0.79)
cg24714606	13	MBNL2	2,82%	7,82E-09	0,22%	8,53E-01	0.3563 (0.32-0.41)	0.3585 (0.31-0.41)	0.3845 (0.34-0.44)
cg02030958	13	x[a]	2,81%	6,36E-09	-0,21%	8,33E-01	0.7105 (0.68-0.75)	0.7085 (0.68-0.74)	0.7386 (0.70-0.78)
cg20124610	13	CARS2	-1,16%	2,18E-12	0,22%	4,72E-01	0.2431 (0.22-0.27)	0.2453 (0.22-0.27)	0.2315 (0.21-0.26)
cg01229865	13	ATP11A	0,72%	5,21E-08	0,19%	1,75E-01	0.1554 (0.13-0.18)	0.1573 (0.13-0.18)	0.1626 (0.14-0.19)
cg11668844	13	MCF2L	0,56%	6,03E-09	0,40%	5,47E-01	0.1090 (0.10-0.12)	0.1130 (0.10-0.12)	0.1145 (0.10-0.13)
cg06885459	13	MCF2L	4,88%	2,09E-11	1,25%	2,85E-01	0.3917 (0.34-0.44)	0.4042 (0.36-0.45)	0.4406 (0.39-0.50)
cg01435643	13	MCF2L	6,23%	9,52E-12	1,66%	1,12E-01	0.4687 (0.40-0.52)	0.4853 (0.43-0.54)	0.5311 (0.47-0.58)
cg03641640	13	RASA3	-0,84%	3,88E-08	-0,14%	2,80E-01	0.8788 (0.86-0.90)	0.8774 (0.86-0.90)	0.8704 (0.85-0.89)
cg17487894	13	RASA3	2,80%	4,72E-10	0,48%	2,87E-01	0.4107 (0.36-0.45)	0.4155 (0.37-0.46)	0.4387 (0.40-0.48)
cg00609473	14	REM2	0,21%	3,35E-09	0,05%	8,45E-02	0.0798 (0.07-0.09)	0.0804 (0.07-0.09)	0.0820 (0.08-0.09)
cg24681208	14	REM2	0,23%	2,87E-08	0,04%	7,66E-02	0.0769 (0.07-0.08)	0.0773 (0.07-0.08)	0.0792 (0.07-0.09)
cg16473141	14	REC8	0,64%	2,29E-09	0,67%	1,26E-01	0.1399 (0.13-0.15)	0.1466 (0.13-0.16)	0.1463 (0.13-0.16)
cg04545963	14	NFKBIA	-0,38%	3,05E-09	0,87%	5,64E-01	0.4083 (0.37-0.45)	0.4170 (0.38-0.46)	0.4046 (0.36-0.44)
cg01860774	14	ZBTB25	-1,50%	1,56E-09	0,97%	7,04E-01	0.6009 (0.52-0.69)	0.6106 (0.53-0.69)	0.5859 (0.50-0.67)
cg22930549	14	RAD51L1	-1,34%	4,57E-08	0,22%	6,06E-01	0.8705 (0.84-0.90)	0.8727 (0.84-0.90)	0.8571 (0.82-0.89)
cg08280368	14	TTC9	-3,34%	1,36E-09	-0,72%	2,23E-01	0.8259 (0.75-0.87)	0.8187 (0.74-0.87)	0.7926 (0.70-0.85)
cg01731783	14	C14orf43	-2,05%	3,48E-11	-0,30%	5,50E-01	0.6572 (0.63-0.69)	0.6541 (0.62-0.69)	0.6366 (0.60-0.67)
cg22851561	14	C14orf43	-6,06%	5,47E-21	-0,49%	3,92E-01	0.4718 (0.41-0.52)	0.4670 (0.41-0.52)	0.4112 (0.36-0.48)
cg24996979	14	C14orf43	-0,56%	6,84E-25	-0,05%	4,34E-03	0.0802 (0.07-0.09)	0.0797 (0.07-0.09)	0.0746 (0.07-0.08)
cg10919522	14	C14orf43	-0,89%	6,20E-17	0,18%	4,65E-01	0.1323 (0.12-0.15)	0.1341 (0.12-0.15)	0.1234 (0.11-0.14)
cg03519879	14	C14orf43	-1,84%	2,61E-11	-0,36%	8,15E-01	0.5000 (0.46-0.53)	0.4964 (0.47-0.53)	0.4816 (0.45-0.51)
cg13976502	14	C14orf43	-1,01%	4,67E-16	0,31%	1,33E-01	0.2701 (0.25-0.29)	0.2733 (0.25-0.30)	0.2600 (0.24-0.28)
cg26217402	14	C14orf43	-0,76%	4,36E-10	-0,05%	1,98E-01	0.6547 (0.62-0.69)	0.6543 (0.62-0.69)	0.6471 (0.61-0.68)
cg13038618	14	x[a]	-4,49%	1,86E-10	-0,36%	1,43E-02	0.4082 (0.35-0.46)	0.4046 (0.34-0.46)	0.3633 (0.29-0.42)
cg05875421	14	GPR68	-0,52%	2,15E-14	0,13%	6,15E-01	0.0755 (0.07-0.08)	0.0768 (0.07-0.08)	0.0703 (0.06-0.08)
cg10814005	14	GPR68	-0,49%	2,21E-12	-0,22%	1,25E-04	0.0556 (0.05-0.06)	0.0534 (0.05-0.06)	0.0507 (0.05-0.06)
cg20303561	14	CCDC88C	-5,81%	1,41E-10	-1,92%	8,26E-02	0.5709 (0.51-0.64)	0.5517 (0.49-0.62)	0.5128 (0.46-0.58)
cg02268192	14	RIN3	0,67%	9,33E-08	0,36%	1,57E-01	0.1496 (0.13-0.17)	0.1532 (0.13-0.18)	0.1563 (0.14-0.18)
cg01288337	14	RIN3	0,88%	2,28E-09	0,03%	6,47E-01	0.1657 (0.14-0.19)	0.1660 (0.14-0.19)	0.1745 (0.15-0.20)
cg05284742	14	ITPK1	-3,76%	3,57E-09	-0,81%	4,27E-02	0.7424 (0.71-0.78)	0.7343 (0.69-0.77)	0.7048 (0.66-0.74)
cg01956154	14	ASB2	-0,11%	4,88E-09	0,13%	9,15E-01	0.1100 (0.10-0.13)	0.1113 (0.10-0.12)	0.1089 (0.10-0.12)
cg07015803	14	BCL11B	-0,29%	1,20E-09	0,32%	3,75E-01	0.8032 (0.77-0.84)	0.8064 (0.77-0.85)	0.8004 (0.76-0.83)
cg19682870	14	BCL11B	-1,36%	4,27E-08	0,19%	6,68E-01	0.8036 (0.75-0.85)	0.8055 (0.75-0.86)	0.7900 (0.75-0.84)
cg06819357	14	TECPR2	5,70%	4,81E-11	0,55%	6,88E-01	0.6637 (0.62-0.72)	0.6693 (0.63-0.72)	0.7207 (0.67-0.77)
cg27170268	14	XRCC3	0,78%	1,72E-08	0,10%	4,35E-01	0.4133 (0.36-0.46)	0.4143 (0.36-0.47)	0.4211 (0.37-0.48)
cg26242531	14	ZFYVE21	4,02%	1,37E-15	0,72%	9,70E-02	0.2288 (0.19-0.27)	0.2359 (0.20-0.28)	0.2690 (0.24-0.31)
cg03509898	14	ZFYVE21	1,46%	1,83E-08	0,16%	3,08E-01	0.9126 (0.89-0.93)	0.9142 (0.89-0.94)	0.9272 (0.91-0.95)
cg14977938	14	ZFYVE21	4,60%	3,21E-19	1,30%	3,21E-03	0.6633 (0.62-0.70)	0.6763 (0.64-0.71)	0.7093 (0.68-0.74)
cg15963913	14	ZFYVE21	2,72%	1,77E-08	0,30%	3,65E-01	0.8213 (0.77-0.86)	0.8243 (0.77-0.87)	0.8484 (0.81-0.88)
cg19838043	14	ZFYVE21	4,85%	1,82E-14	0,44%	3,03E-02	0.6186 (0.57-0.67)	0.6230 (0.57-0.68)	0.6671 (0.61-0.72)
cg08200043	14	x[a]	-1,03%	3,29E-08	0,06%	7,64E-01	0.9253 (0.90-0.94)	0.9259 (0.90-0.95)	0.9150 (0.89-0.94)
cg13578465	14	INF2	1,99%	7,84E-09	0,46%	1,34E-02	0.7629 (0.73-0.80)	0.7675 (0.73-0.81)	0.7828 (0.75-0.82)
cg11730703	14	INF2	2,06%	2,04E-10	0,30%	4,33E-02	0.7804 (0.75-0.81)	0.7834 (0.75-0.81)	0.8010 (0.77-0.83)
cg14359680	14	GPR132	-0,21%	4,98E-08	0,11%	8,03E-01	0.1021 (0.09-0.11)	0.1031 (0.09-0.12)	0.1000 (0.09-0.11)
cg00808648	14	PACS2	0,30%	8,74E-09	0,21%	3,18E-01	0.1267 (0.11-0.14)	0.1288 (0.12-0.14)	0.1297 (0.12-0.15)
cg01208318	14	x[a]	-7,74%	2,54E-12	-0,28%	7,68E-02	0.2836 (0.20-0.37)	0.2808 (0.20-0.39)	0.2063 (0.16-0.28)
cg16558846	14	x[a]	-4,95%	1,63E-08	0,36%	3,09E-01	0.5912 (0.52-0.67)	0.5948 (0.51-0.67)	0.5417 (0.45-0.63)
cg04189320	15	x[a]	-1,34%	1,68E-12	-0,33%	1,40E-01	0.8402 (0.80-0.87)	0.8370 (0.80-0.87)	0.8268 (0.78-0.86)
cg09735851	15	x[a]	-0,06%	6,61E-08	0,62%	3,98E-01	0.1971 (0.19-0.21)	0.2032 (0.19-0.21)	0.1964 (0.18-0.21)
cg17750114	15	EHD4	0,73%	3,40E-09	0,51%	2,45E-02	0.1274 (0.12-0.14)	0.1325 (0.12-0.15)	0.1347 (0.12-0.15)
cg22107533	15	TRIM69	-2,53%	2,95E-12	-0,33%	3,03E-01	0.4010 (0.35-0.45)	0.3977 (0.35-0.44)	0.3758 (0.32-0.43)
cg05439368	15	TRIM69	-1,93%	7,39E-11	-0,29%	5,66E-01	0.5289 (0.48-0.57)	0.5260 (0.48-0.57)	0.5096 (0.45-0.56)
cg15022400	15	TRIM69	-0,73%	3,36E-13	-0,09%	1,24E-01	0.0804 (0.07-0.09)	0.0795 (0.07-0.09)	0.0731 (0.06-0.08)
cg10439456	15	TRIM69	-3,04%	6,86E-09	-0,48%	1,28E-01	0.3368 (0.26-0.42)	0.3320 (0.26-0.41)	0.3064 (0.23-0.37)
cg01138448	15	TRIM69	-3,95%	2,74E-11	-0,33%	5,93E-01	0.6841 (0.63-0.73)	0.6808 (0.62-0.73)	0.6446 (0.58-0.71)
cg24507955	15	SECISBP2L	-1,54%	4,27E-08	0,01%	1,52E-01	0.2599 (0.23-0.29)	0.2600 (0.23-0.29)	0.2445 (0.22-0.28)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg10858677	15	x[a]	-3,77%	4,30E-09	-1,42%	2,25E-02	0.4714 (0.41-0.52)	0.4572 (0.39-0.52)	0.4337 (0.37-0.49)
cg27332104	15	x[a]	3,12%	6,05E-09	-0,05%	4,97E-03	0.5481 (0.49-0.60)	0.5476 (0.49-0.60)	0.5793 (0.53-0.63)
cg24687805	15	RAB27A	0,40%	6,06E-09	0,21%	6,41E-02	0.0484 (0.04-0.06)	0.0505 (0.04-0.06)	0.0524 (0.04-0.06)
cg21756476	15	RORA	5,94%	2,74E-11	1,39%	2,29E-02	0.5055 (0.45-0.56)	0.5194 (0.46-0.57)	0.5649 (0.50-0.61)
cg10403394	15	TPM1	5,07%	7,00E-12	1,63%	2,62E-02	0.1560 (0.12-0.23)	0.1722 (0.13-0.25)	0.2067 (0.15-0.33)
cg08783514	15	x[a]	-2,49%	5,88E-08	0,01%	2,87E-02	0.7054 (0.66-0.75)	0.7055 (0.65-0.75)	0.6805 (0.63-0.73)
cg14004161	15	SNX22	1,45%	7,92E-08	1,15%	4,75E-01	0.3639 (0.34-0.40)	0.3755 (0.34-0.40)	0.3785 (0.35-0.41)
cg07196571	15	SNX22	1,01%	8,50E-09	0,92%	2,56E-01	0.4533 (0.43-0.47)	0.4625 (0.44-0.48)	0.4634 (0.44-0.49)
cg18926409	15	x[a]	0,55%	2,10E-09	0,14%	1,29E-01	0.0764 (0.07-0.09)	0.0778 (0.07-0.09)	0.0819 (0.07-0.09)
cg19459791	15	x[a]	2,67%	3,67E-11	0,03%	1,74E-02	0.7968 (0.75-0.84)	0.7971 (0.75-0.84)	0.8236 (0.78-0.86)
cg03489965	15	LOC390594	2,66%	9,52E-15	0,89%	2,54E-04	0.2217 (0.19-0.25)	0.2306 (0.20-0.27)	0.2483 (0.22-0.30)
cg10619342	15	LOC390594	4,11%	2,04E-15	0,91%	1,39E-03	0.2781 (0.24-0.33)	0.2872 (0.24-0.34)	0.3191 (0.27-0.39)
cg27514333	15	SMAD6	-0,41%	1,65E-10	-0,04%	1,42E-02	0.1252 (0.11-0.14)	0.1249 (0.11-0.14)	0.1212 (0.11-0.13)
cg21963854	15	MIR548H4	0,63%	2,17E-08	0,41%	5,01E-01	0.1084 (0.10-0.12)	0.1124 (0.10-0.12)	0.1147 (0.10-0.13)
cg26971042	15	TLE3	-0,53%	5,34E-09	0,10%	2,81E-01	0.0855 (0.08-0.09)	0.0865 (0.08-0.09)	0.0803 (0.07-0.09)
cg15903956	15	x[a]	-1,66%	9,49E-08	0,31%	7,17E-01	0.6405 (0.59-0.69)	0.6437 (0.59-0.69)	0.6239 (0.58-0.67)
cg18335991	15	SEMA7A	-3,94%	4,90E-13	-1,33%	3,49E-01	0.6581 (0.60-0.71)	0.6448 (0.60-0.70)	0.6187 (0.55-0.67)
cg00310412	15	SEMA7A	-1,97%	2,34E-13	-0,60%	3,51E-01	0.2994 (0.27-0.32)	0.2934 (0.27-0.32)	0.2797 (0.25-0.31)
cg11152412	15	EDC3	-0,30%	1,37E-11	-0,10%	1,57E-02	0.0522 (0.05-0.06)	0.0513 (0.05-0.06)	0.0492 (0.04-0.05)
cg12101586	15	CYP1A1	-3,31%	5,46E-08	-2,31%	7,90E-08	0.4771 (0.42-0.53)	0.4540 (0.40-0.51)	0.4439 (0.40-0.50)
cg19696491	15	CHRNA5	2,45%	5,60E-09	0,43%	7,87E-01	0.2234 (0.19-0.26)	0.2277 (0.20-0.26)	0.2479 (0.21-0.29)
cg23161492	15	ANPEP	-1,60%	1,12E-24	-0,37%	1,58E-05	0.1212 (0.11-0.13)	0.1175 (0.11-0.13)	0.1053 (0.10-0.12)
cg15188623	15	ZNF710	0,70%	1,92E-08	0,32%	4,03E-01	0.1408 (0.13-0.16)	0.1440 (0.13-0.16)	0.1479 (0.13-0.16)
cg05194346	15	x[a]	9,98%	3,36E-22	1,65%	6,73E-02	0.3928 (0.32-0.48)	0.4093 (0.34-0.50)	0.4926 (0.41-0.57)
cg20357538	15	CHSY1	2,44%	3,71E-10	0,71%	2,58E-01	0.2775 (0.23-0.32)	0.2846 (0.24-0.33)	0.3019 (0.25-0.35)
cg09465703	16	JMJD8	-1,06%	4,51E-09	1,78%	3,37E-02	0.2240 (0.18-0.28)	0.2417 (0.19-0.31)	0.2134 (0.17-0.27)
cg10244976	16	LMF1	-0,86%	5,87E-09	0,36%	6,45E-01	0.8544 (0.82-0.88)	0.8580 (0.83-0.88)	0.8458 (0.81-0.87)
cg08438529	16	x[a]	-0,29%	1,47E-08	0,18%	9,29E-01	0.6303 (0.60-0.66)	0.6321 (0.61-0.66)	0.6274 (0.59-0.65)
cg19030554	16	NME3/MRPS34	-0,45%	8,13E-09	0,33%	1,80E-01	0.1098 (0.10-0.12)	0.1131 (0.10-0.13)	0.1053 (0.09-0.12)
cg01207684	16	ADCY9	-1,83%	6,29E-10	-0,81%	1,59E-05	0.7670 (0.73-0.80)	0.7590 (0.73-0.79)	0.7487 (0.70-0.78)
cg12486287	16	VASN/CORO7	1,13%	6,38E-08	1,18%	3,96E-02	0.1496 (0.14-0.16)	0.1615 (0.15-0.17)	0.1609 (0.15-0.17)
cg26504467	16	x[a]	-1,13%	9,98E-08	0,03%	7,25E-01	0.8152 (0.78-0.85)	0.8155 (0.78-0.85)	0.8039 (0.76-0.84)
cg26988253	16	LAT	-0,85%	7,75E-09	-0,19%	6,76E-01	0.8555 (0.82-0.89)	0.8536 (0.82-0.89)	0.8470 (0.81-0.88)
cg06177555	16	SPN	3,46%	8,43E-12	0,27%	1,27E-01	0.2508 (0.21-0.31)	0.2535 (0.20-0.31)	0.2854 (0.23-0.35)
cg27583010	16	CORO1A	-0,27%	3,89E-08	-0,09%	6,07E-01	0.5939 (0.56-0.62)	0.5930 (0.56-0.63)	0.5911 (0.55-0.62)
cg07635132	16	DCTPP1	-1,51%	1,42E-08	-0,12%	2,86E-01	0.6054 (0.58-0.64)	0.6041 (0.57-0.64)	0.5903 (0.57-0.62)
cg09099830	16	ITGAL	-1,72%	1,55E-24	0,56%	4,39E-02	0.5947 (0.57-0.62)	0.6003 (0.57-0.63)	0.5775 (0.55-0.60)
cg16391678	16	ITGAL	-0,64%	1,04E-14	-0,01%	1,85E-01	0.6997 (0.67-0.73)	0.6996 (0.67-0.73)	0.6933 (0.66-0.72)
cg04716530	16	ITGAL	-0,99%	1,69E-14	-0,25%	3,29E-01	0.8148 (0.78-0.85)	0.8123 (0.78-0.84)	0.8050 (0.76-0.84)
cg16519923	16	ITGAL	-3,09%	2,46E-16	-0,58%	3,59E-01	0.8142 (0.76-0.86)	0.8084 (0.76-0.85)	0.7833 (0.73-0.83)
cg06235438	16	ITGAL	-1,99%	2,00E-16	-0,29%	2,12E-02	0.8895 (0.85-0.92)	0.8866 (0.84-0.91)	0.8696 (0.83-0.91)
cg06972908	16	ITGAL	-2,06%	1,32E-10	-1,16%	2,51E-03	0.7788 (0.74-0.82)	0.7672 (0.73-0.81)	0.7582 (0.72-0.79)
cg04232972	16	x[a]	-0,41%	9,71E-12	0,00%	8,05E-03	0.0970 (0.09-0.11)	0.0970 (0.09-0.11)	0.0928 (0.08-0.10)
cg07069636	16	x[a]	-1,33%	1,36E-15	0,42%	1,23E-01	0.3380 (0.31-0.37)	0.3422 (0.31-0.37)	0.3247 (0.29-0.36)
cg01305745	16	VKORC1	1,17%	1,20E-08	0,03%	4,55E-02	0.7496 (0.73-0.77)	0.7499 (0.73-0.77)	0.7612 (0.74-0.78)
cg15337006	16	ITGAM	0,96%	1,81E-09	0,19%	4,33E-01	0.1441 (0.13-0.16)	0.1460 (0.13-0.17)	0.1538 (0.13-0.17)
cg00833777	16	ITGAM	0,30%	1,69E-08	0,14%	3,40E-01	0.0695 (0.06-0.08)	0.0709 (0.06-0.08)	0.0725 (0.07-0.08)
cg08944236	16	CHD9	0,65%	2,46E-09	0,20%	8,33E-01	0.1234 (0.11-0.14)	0.1254 (0.11-0.14)	0.1299 (0.11-0.15)
cg16547186	16	DOK4	0,42%	6,17E-08	0,18%	1,22E-01	0.1099 (0.09-0.12)	0.1117 (0.10-0.13)	0.1141 (0.10-0.13)
cg00407537	16	GPR114	0,43%	8,86E-11	0,14%	1,94E-01	0.0740 (0.07-0.08)	0.0754 (0.07-0.09)	0.0783 (0.07-0.09)
cg07877900	16	GPR114	1,04%	5,09E-08	0,80%	5,98E-01	0.1984 (0.17-0.23)	0.2064 (0.18-0.24)	0.2087 (0.18-0.24)
cg03155159	16	x[a]	-1,19%	9,31E-08	0,30%	2,86E-01	0.3190 (0.24-0.39)	0.3220 (0.25-0.39)	0.3071 (0.23-0.36)
cg13500388	16	CBFB	-2,43%	1,14E-10	-0,02%	3,60E-02	0.2865 (0.26-0.33)	0.2863 (0.26-0.32)	0.2622 (0.23-0.29)
cg00496272	16	CENPT/THAP11	-0,27%	2,10E-09	0,14%	4,63E-02	0.0923 (0.08-0.10)	0.0937 (0.08-0.10)	0.0896 (0.08-0.10)
cg05969150	16	ZFHX3	2,60%	6,33E-09	0,35%	3,07E-01	0.2702 (0.24-0.30)	0.2738 (0.25-0.31)	0.2962 (0.26-0.32)
cg27262054	16	ZFHX3	-1,77%	6,07E-09	-0,16%	1,62E-01	0.7544 (0.73-0.78)	0.7529 (0.73-0.78)	0.7367 (0.71-0.76)
cg09985344	16	COTL1	-1,30%	3,05E-10	-0,19%	1,02E-01	0.6888 (0.65-0.73)	0.6869 (0.64-0.73)	0.6758 (0.62-0.72)
cg09938443	16	ZDHHC7	-1,42%	1,89E-08	-0,34%	3,28E-01	0.8895 (0.86-0.92)	0.8860 (0.86-0.91)	0.8753 (0.84-0.90)
cg02328930	16	x[a]	0,41%	8,24E-09	0,15%	6,10E-01	0.0818 (0.07-0.09)	0.0833 (0.08-0.09)	0.0859 (0.08-0.09)
cg03848483	16	x[a]	0,90%	5,75E-09	0,47%	4,83E-01	0.2102 (0.18-0.23)	0.2150 (0.18-0.24)	0.2192 (0.19-0.25)
cg16497398	16	ZCCHC14	-1,45%	3,76E-10	-0,44%	1,17E-01	0.7286 (0.70-0.75)	0.7242 (0.69-0.75)	0.7141 (0.68-0.74)
cg09636313	16	x[a]	2,86%	5,90E-08	1,20%	3,31E-02	0.3091 (0.27-0.36)	0.3211 (0.28-0.37)	0.3377 (0.30-0.40)
cg27150870	16	ANKRD11	-1,16%	5,05E-12	-0,05%	6,05E-02	0.8970 (0.87-0.92)	0.8965 (0.87-0.92)	0.8854 (0.86-0.91)
cg00759807	16	ANKRD11	-0,31%	8,29E-09	0,30%	6,95E-01	0.7729 (0.74-0.80)	0.7759 (0.74-0.80)	0.7698 (0.74-0.80)
cg05978306	17	MYO1C	1,17%	5,93E-08	1,08%	1,19E-01	0.4806 (0.46-0.50)	0.4914 (0.47-0.51)	0.4923 (0.47-0.52)
cg20426671	17	PITPNA	2,37%	1,28E-11	0,16%	6,63E-01	0.6473 (0.62-0.68)	0.6489 (0.62-0.68)	0.6710 (0.64-0.70)
cg12814335	17	PITPNA	2,96%	9,36E-13	0,55%	1,40E-01	0.6350 (0.59-0.68)	0.6404 (0.60-0.68)	0.6645 (0.62-0.72)
cg11719157	17	PITPNA	3,21%	9,42E-12	-0,06%	6,70E-01	0.6272 (0.58-0.67)	0.6266 (0.58-0.67)	0.6593 (0.62-0.71)
cg25212453	17	SLC43A2	-0,81%	2,66E-10	0,15%	9,04E-02	0.9339 (0.92-0.94)	0.9354 (0.93-0.95)	0.9258 (0.91-0.94)
cg21949830	17	SLC43A2	-0,67%	1,13E-08	0,40%	6,73E-02	0.8819 (0.86-0.90)	0.8859 (0.87-0.90)	0.8753 (0.85-0.89)
cg00911794	17	HIC1	4,60%	6,84E-10	3,90%	4,22E-01	0.2870 (0.25-0.35)	0.3260 (0.26-0.39)	0.3330 (0.27-0.40)
cg23621097	17	HIC1	3,47%	4,23E-08	2,86%	3,40E-01	0.2768 (0.23-0.33)	0.3053 (0.25-0.37)	0.3115 (0.25-0.38)
cg12077963	17	ANKFY1	-1,07%	3,43E-09	-0,47%	6,52E-02	0.7765 (0.74-0.81)	0.7718 (0.73-0.81)	0.7658 (0.73-0.80)
cg07697428	17	ANKFY1	-0,43%	1,36E-11	0,01%	3,73E-01	0.7243 (0.70-0.76)	0.7244 (0.70-0.75)	0.7200 (0.69-0.75)
cg04970434	17	GP1BA	3,10%	1,28E-09	0,33%	6,57E-01	0.6665 (0.62-0.72)	0.6698 (0.62-0.72)	0.6976 (0.66-0.75)
cg14897188	17	RNASEK/C17orf49	-2,08%	2,19E-08	0,53%	4,20E-01	0.5010 (0.44-0.56)	0.5063 (0.45-0.56)	0.4803 (0.43-0.53)
cg26946288	17	PLSCR3	-3,96%	3,48E-09	-1,25%	3,05E-01	0.7887 (0.67-0.88)	0.7762 (0.66-0.88)	0.7491 (0.63-0.85)
cg15227911	17	CHD3	-0,22%	1,51E-08	0,33%	3,22E-01	0.3530 (0.31-0.40)	0.3563 (0.30-0.41)	0.3508 (0.30-0.38)
cg04674060	17	CHD3	-1,80%	2,44E-08	-0,05%	2,62E-01	0.5183 (0.46-0.57)	0.5178 (0.45-0.57)	0.5003 (0.43-0.55)
cg25921609	17	MYH10	1,97%	6,02E-11	0,80%	1,42E-01	0.6682 (0.63-0.70)	0.6762 (0.64-0.71)	0.6879 (0.65-0.72)
cg05460226	17	PIK3R5	-3,94%	2,41E-11	-0,81%	3,48E-03	0.2394 (0.20-0.29)	0.2313 (0.20-0.28)	0.2000 (0.17-0.25)
cg11122944	17	GAS7	0,49%	4,62E-08	-0,01%	2,91E-01	0.1275 (0.11-0.14)	0.1274 (0.11-0.14)	0.1324 (0.11-0.15)
cg07435331	17	x[a]	1,13%	4,97E-09	0,25%	2,64E-01	0.1847 (0.17-0.20)	0.1873 (0.17-0.21)	0.1961 (0.18-0.22)
cg04862556	17	SLFN13	0,70%	1,49E-09	0,45%	3,30E-01	0.1941 (0.18-0.21)	0.1986 (0.18-0.22)	0.2011 (0.18-0.22)
cg13443575	17	SLFN13	1,43%	5,16E-08	1,09%	7,06E-01	0.4563 (0.43-0.48)	0.4672 (0.44-0.49)	0.4706 (0.44-0.50)
cg10295552	17	ACACA	2,00%	1,04E-08	0,04%	3,17E-02	0.6115 (0.56-0.67)	0.6119 (0.56-0.67)	0.6315 (0.58-0.69)
cg00808555	17	ARL5C	0,77%	1,01E-08	0,65%	7,59E-02	0.1452 (0.13-0.16)	0.1517 (0.14-0.16)	0.1529 (0.14-0.17)
cg09858022	17	RARA	1,20%	5,46E-08	0,36%	9,17E-01	0.3923 (0.36-0.42)	0.3959 (0.37-0.42)	0.4043 (0.38-0.43)
cg26654286	17	RARA	0,22%	6,97E-09	-0,06%	9,49E-01	0.0495 (0.04-0.06)	0.0489 (0.04-0.06)	0.0517 (0.04-0.06)
cg19572487	17	RARA	-10,02%	9,37E-40	-2,04%	2,64E-07	0.4477 (0.40-0.51)	0.4273 (0.36-0.49)	0.3475 (0.30-0.43)
cg26878655	17	RARA	-1,82%	2,01E-13	0,15%	1,87E-01	0.7795 (0.75-0.81)	0.7810 (0.74-0.81)	0.7613 (0.72-0.80)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg27453606	17	RARA	0,12%	1,50E-09	0,02%	2,53E-01	0.0233 (0.02-0.03)	0.0236 (0.02-0.03)	0.0245 (0.02-0.03)
cg16276850	17	RARA	0,98%	6,50E-09	0,81%	6,96E-01	0.1854 (0.17-0.20)	0.1934 (0.18-0.21)	0.1952 (0.18-0.21)
cg05824218	17	RARA	2,83%	2,99E-15	0,31%	8,84E-02	0.4097 (0.37-0.44)	0.4128 (0.38-0.45)	0.4380 (0.40-0.47)
cg04956244	17	RARA	2,50%	3,44E-10	0,50%	1,58E-02	0.5979 (0.57-0.63)	0.6030 (0.57-0.63)	0.6229 (0.59-0.65)
cg16255816	17	HAP1	-0,57%	9,19E-11	-0,07%	1,42E-02	0.1406 (0.13-0.15)	0.1399 (0.13-0.15)	0.1348 (0.12-0.14)
cg03373393	17	HAP1	-0,34%	1,38E-11	-0,02%	5,70E-03	0.0849 (0.08-0.09)	0.0846 (0.08-0.09)	0.0814 (0.07-0.09)
cg08783253	17	AOC2	1,92%	9,79E-11	0,32%	3,71E-01	0.3062 (0.26-0.37)	0.3094 (0.26-0.37)	0.3254 (0.27-0.39)
cg04951104	17	MEOX1	-1,45%	1,03E-08	0,36%	7,38E-01	0.8209 (0.78-0.86)	0.8245 (0.79-0.86)	0.8064 (0.76-0.84)
cg23766254	17	FAM171A2	-1,08%	2,51E-08	1,42%	1,92E-01	0.5015 (0.46-0.54)	0.5157 (0.47-0.55)	0.4907 (0.43-0.53)
cg22078805	17	FAM171A2	-1,94%	5,73E-08	0,52%	1,13E-01	0.5353 (0.49-0.57)	0.5405 (0.50-0.58)	0.5159 (0.46-0.55)
cg25809905	17	ITGA2B	6,53%	2,74E-12	1,01%	5,35E-03	0.5129 (0.44-0.58)	0.5230 (0.46-0.59)	0.5782 (0.52-0.64)
cg06649191	17	GJC1	0,17%	6,76E-08	0,00%	3,08E-01	0.0658 (0.06-0.07)	0.0657 (0.06-0.07)	0.0675 (0.06-0.08)
cg16792613	17	FMNL1	0,03%	4,18E-09	0,14%	5,45E-02	0.1362 (0.13-0.15)	0.1376 (0.13-0.15)	0.1365 (0.13-0.14)
cg20864568	17	MAP3K14	-0,19%	3,78E-08	-0,01%	6,14E-01	0.1056 (0.09-0.12)	0.1055 (0.10-0.12)	0.1037 (0.09-0.11)
cg19325791	17	x[a]	4,16%	5,15E-08	0,78%	3,93E-01	0.5504 (0.49-0.61)	0.5582 (0.51-0.62)	0.5921 (0.53-0.65)
cg21280392	17	PHOSPHO1	2,37%	1,56E-08	0,67%	5,66E-02	0.2667 (0.24-0.29)	0.2733 (0.25-0.30)	0.2903 (0.27-0.32)
cg26664457	17	ZNF652	0,60%	5,88E-09	0,06%	5,10E-01	0.1628 (0.14-0.19)	0.1635 (0.13-0.20)	0.1689 (0.14-0.20)
cg21483922	17	SAMD14	0,60%	1,67E-10	0,16%	6,09E-01	0.0822 (0.07-0.09)	0.0838 (0.08-0.09)	0.0882 (0.08-0.10)
cg07465627	17	STXBP4	-1,38%	3,67E-12	0,11%	6,90E-01	0.1802 (0.16-0.21)	0.1814 (0.16-0.21)	0.1665 (0.15-0.19)
cg25537245	17	x[a]	-0,26%	1,23E-08	-0,14%	1,13E-01	0.0576 (0.05-0.06)	0.0562 (0.05-0.06)	0.0551 (0.05-0.06)
cg10588135	17	BCAS3	0,45%	2,43E-09	0,00%	9,59E-01	0.1165 (0.10-0.13)	0.1165 (0.10-0.13)	0.1210 (0.11-0.14)
cg04543901	17	x[a]	-3,03%	3,18E-11	-1,31%	2,83E-02	0.4495 (0.35-0.52)	0.4364 (0.35-0.52)	0.4191 (0.32-0.50)
cg25909396	17	PRKCA	-1,24%	2,25E-08	-0,23%	1,39E-01	0.8959 (0.87-0.92)	0.8936 (0.86-0.92)	0.8835 (0.85-0.91)
cg02186444	17	ARMC7	3,82%	4,67E-16	0,39%	1,33E-02	0.5410 (0.50-0.58)	0.5449 (0.51-0.58)	0.5792 (0.53-0.61)
cg20338276	17	GGA3	-1,22%	1,15E-10	-0,05%	7,29E-01	0.8192 (0.79-0.84)	0.8187 (0.79-0.85)	0.8070 (0.77-0.84)
cg23599820	17	KIAA0195	0,74%	6,82E-08	-0,23%	8,22E-01	0.1727 (0.15-0.20)	0.1705 (0.15-0.20)	0.1802 (0.15-0.21)
cg07251887	17	RECQL5	-5,13%	5,61E-09	-2,20%	3,98E-04	0.3329 (0.29-0.38)	0.3110 (0.26-0.36)	0.2816 (0.24-0.32)
cg19468528	17	LOC100130933	0,59%	3,67E-09	0,23%	4,87E-01	0.1123 (0.10-0.13)	0.1146 (0.10-0.13)	0.1181 (0.11-0.13)
cg00871371	17	SEPT9	-0,63%	3,36E-08	0,58%	9,85E-01	0.2695 (0.24-0.30)	0.2754 (0.24-0.31)	0.2633 (0.23-0.29)
cg07324245	17	SEPT9	-1,96%	5,19E-08	0,06%	7,26E-02	0.5573 (0.50-0.62)	0.5579 (0.49-0.62)	0.5377 (0.48-0.59)
cg00324097	17	SEPT9	-0,71%	3,05E-10	-0,09%	1,72E-01	0.2406 (0.22-0.26)	0.2397 (0.22-0.26)	0.2335 (0.21-0.26)
cg19654743	17	SEPT9	-0,97%	6,86E-08	0,30%	5,40E-01	0.4705 (0.41-0.54)	0.4734 (0.42-0.53)	0.4607 (0.40-0.51)
cg06513247	17	SEPT9	-0,35%	4,63E-08	0,21%	5,16E-01	0.2680 (0.25-0.30)	0.2701 (0.24-0.30)	0.2644 (0.24-0.29)
cg27627381	17	SEPT9	-0,56%	1,83E-09	-0,19%	1,19E-01	0.7875 (0.77-0.81)	0.7856 (0.76-0.81)	0.7819 (0.76-0.80)
cg18758622	17	BAIAP2	1,05%	1,38E-10	0,42%	5,88E-02	0.2013 (0.19-0.22)	0.2055 (0.19-0.22)	0.2118 (0.19-0.23)
cg06710464	17	BAIAP2	1,48%	5,56E-08	1,15%	1,10E-01	0.7811 (0.73-0.83)	0.7927 (0.73-0.84)	0.7960 (0.75-0.85)
cg10909080	17	MAFG	0,97%	2,59E-08	0,82%	3,70E-01	0.2709 (0.25-0.30)	0.2791 (0.25-0.31)	0.2805 (0.25-0.32)
cg21773646	17	CCDC57	0,11%	4,32E-09	0,39%	3,98E-01	0.6772 (0.64-0.71)	0.6811 (0.65-0.72)	0.6784 (0.64-0.71)
cg08999812	17	TBCD	2,73%	3,92E-08	1,15%	8,93E-02	0.2810 (0.24-0.32)	0.2925 (0.25-0.34)	0.3083 (0.26-0.35)
cg23989207	17	TBCD	2,49%	9,43E-10	0,20%	1,46E-01	0.5550 (0.50-0.60)	0.5570 (0.51-0.60)	0.5798 (0.53-0.62)
cg11526020	17	TBCD	0,64%	3,19E-09	0,39%	8,64E-02	0.1930 (0.18-0.21)	0.1969 (0.18-0.22)	0.1994 (0.18-0.22)
cg15727507	17	B3GNTL1	1,44%	1,39E-09	0,10%	6,89E-01	0.2459 (0.23-0.26)	0.2470 (0.23-0.27)	0.2603 (0.24-0.28)
cg21088259	17	METRNL	1,29%	1,34E-08	0,88%	5,78E-02	0.3066 (0.26-0.36)	0.3154 (0.27-0.37)	0.3194 (0.28-0.38)
cg06459104	18	EPB41L3	-5,45%	1,78E-10	-3,04%	8,04E-05	0.2918 (0.21-0.38)	0.2615 (0.19-0.36)	0.2373 (0.17-0.32)
cg01017464	18	SNORD58A/RPL17	-0,38%	3,03E-09	0,17%	8,65E-01	0.0980 (0.09-0.11)	0.0997 (0.09-0.11)	0.0942 (0.08-0.10)
cg27262850	18	MBP	0,75%	2,56E-08	-0,39%	7,00E-01	0.1451 (0.12-0.17)	0.1411 (0.12-0.17)	0.1526 (0.13-0.19)
cg16604566	18	MBP	0,70%	3,26E-08	0,52%	4,53E-01	0.2031 (0.18-0.23)	0.2083 (0.18-0.23)	0.2101 (0.18-0.24)
cg22324981	18	NFATC1	-1,58%	5,23E-10	0,29%	3,40E-02	0.7688 (0.73-0.80)	0.7717 (0.73-0.80)	0.7530 (0.72-0.79)
cg06784563	18	NFATC1	0,12%	6,38E-08	0,65%	9,92E-01	0.5981 (0.56-0.63)	0.6046 (0.57-0.64)	0.5993 (0.56-0.63)
cg00073090	19	x[a]	-0,26%	9,10E-14	0,31%	8,03E-03	0.1727 (0.16-0.18)	0.1759 (0.17-0.19)	0.1701 (0.16-0.18)
cg15187398	19	MOBKL2A	-2,75%	8,51E-09	-0,43%	4,01E-02	0.2507 (0.22-0.28)	0.2465 (0.22-0.27)	0.2232 (0.20-0.25)
cg07381806	19	MOBKL2A	-5,12%	3,46E-17	-0,46%	4,40E-05	0.4132 (0.36-0.47)	0.4086 (0.35-0.46)	0.3621 (0.30-0.42)
cg21869609	19	LINGO3	-2,83%	2,01E-08	0,07%	5,72E-02	0.7228 (0.68-0.77)	0.7235 (0.67-0.78)	0.6945 (0.65-0.74)
cg00835193	19	LINGO3	-8,23%	2,44E-08	-0,43%	4,23E-03	0.8822 (0.75-0.93)	0.8780 (0.73-0.93)	0.7999 (0.61-0.92)
cg13561409	19	PIP5K1C	2,85%	2,77E-08	1,35%	1,92E-04	0.3813 (0.33-0.43)	0.3947 (0.35-0.45)	0.4097 (0.36-0.46)
cg03715305	19	NDUFA11	1,70%	6,89E-09	0,21%	2,36E-02	0.5767 (0.54-0.61)	0.5788 (0.54-0.61)	0.5937 (0.56-0.62)
cg08243524	19	DENND1C	-0,75%	3,42E-10	-0,28%	7,54E-03	0.0774 (0.07-0.09)	0.0746 (0.06-0.09)	0.0699 (0.06-0.08)
cg15055101	19	SH2D3A	-1,95%	1,91E-08	-0,30%	6,66E-01	0.4265 (0.38-0.48)	0.4234 (0.38-0.48)	0.4070 (0.36-0.45)
cg04770751	19	STXBP2	-0,42%	1,70E-08	0,24%	2,98E-03	0.1663 (0.16-0.18)	0.1687 (0.15-0.18)	0.1621 (0.15-0.18)
cg14074174	19	SNAPC2	-0,67%	7,74E-09	0,40%	2,84E-01	0.3031 (0.28-0.33)	0.3071 (0.28-0.33)	0.2964 (0.27-0.32)
cg18745507	19	ZGLP1	2,48%	2,47E-08	0,36%	7,75E-02	0.5514 (0.52-0.59)	0.5550 (0.52-0.60)	0.5762 (0.54-0.62)
cg05323251	19	ICAM3	-0,47%	7,80E-09	0,07%	6,51E-01	0.0699 (0.06-0.08)	0.0707 (0.06-0.08)	0.0653 (0.06-0.07)
cg07512814	19	LDLR	-3,72%	2,56E-10	-0,89%	7,51E-02	0.4219 (0.36-0.48)	0.4130 (0.36-0.47)	0.3847 (0.33-0.44)
cg03045323	19	C19orf39	0,26%	2,76E-08	0,10%	1,40E-01	0.0591 (0.05-0.06)	0.0601 (0.05-0.07)	0.0617 (0.06-0.07)
cg11621113	19	MORG1/MAN2B1	-0,57%	5,51E-19	0,10%	8,59E-02	0.1214 (0.11-0.13)	0.1225 (0.11-0.13)	0.1157 (0.11-0.13)
cg22027399	19	SYDE1	0,32%	1,98E-09	0,22%	4,11E-01	0.1110 (0.10-0.12)	0.1132 (0.10-0.12)	0.1142 (0.11-0.12)
cg23661483	19	ILVBL	1,64%	1,96E-10	0,44%	1,63E-02	0.6865 (0.66-0.72)	0.6909 (0.66-0.73)	0.7029 (0.67-0.74)
cg07408456	19	PGLYRP2	1,65%	4,85E-08	0,50%	1,71E-01	0.1877 (0.17-0.21)	0.1928 (0.17-0.22)	0.2043 (0.18-0.23)
cg03384915	19	SIN3B	-9,79%	7,60E-11	-0,07%	3,54E-01	0.7198 (0.61-0.85)	0.7191 (0.60-0.85)	0.6219 (0.53-0.74)
cg14712058	19	SIN3B	-1,48%	2,66E-08	-0,29%	3,36E-01	0.8068 (0.79-0.83)	0.8039 (0.78-0.83)	0.7920 (0.77-0.82)
cg03636183*	19	F2RL3	-14,74%	2,42E-80	-3,48%	4,21E-17	0.4930 (0.45-0.54)	0.4582 (0.40-0.51)	0.3456 (0.28-0.43)
cg15159987	19	CPAMD8	-3,53%	6,39E-15	-0,52%	4,60E-01	0.6339 (0.59-0.67)	0.6287 (0.59-0.66)	0.5986 (0.56-0.64)
cg02639359	19	FCHO1	-0,45%	2,55E-09	0,13%	1,43E-01	0.1180 (0.10-0.15)	0.1193 (0.10-0.15)	0.1135 (0.10-0.14)
cg23931381	19	ARRDC2	-1,20%	6,51E-09	-0,35%	1,40E-03	0.9214 (0.90-0.94)	0.9179 (0.90-0.93)	0.9094 (0.88-0.93)
cg05206276	19	x[a]	-1,66%	8,28E-08	-0,24%	1,31E-01	0.7655 (0.71-0.81)	0.7632 (0.71-0.82)	0.7489 (0.69-0.80)
cg23973524	19	CRTC1	4,21%	1,23E-25	1,15%	1,08E-02	0.3062 (0.28-0.34)	0.3177 (0.29-0.35)	0.3483 (0.32-0.38)
cg21473814	19	CRTC1	5,14%	5,50E-20	0,96%	3,70E-02	0.3233 (0.29-0.36)	0.3329 (0.30-0.38)	0.3746 (0.34-0.42)
cg00591868	19	PBX4	-0,16%	9,46E-09	-0,21%	7,88E-01	0.6814 (0.66-0.71)	0.6793 (0.66-0.70)	0.6798 (0.65-0.70)
cg10521852	19	LPAR2	0,56%	1,72E-08	0,40%	9,48E-02	0.1122 (0.10-0.12)	0.1162 (0.10-0.13)	0.1179 (0.11-0.13)
cg24715735	19	HPN	0,22%	6,04E-08	0,07%	6,37E-01	0.0834 (0.08-0.09)	0.0841 (0.08-0.09)	0.0857 (0.08-0.09)
cg06151964	19	HPN	0,48%	2,50E-11	0,22%	4,33E-02	0.1101 (0.10-0.12)	0.1122 (0.10-0.12)	0.1149 (0.11-0.12)
cg11902728	19	MAG	5,76%	4,05E-13	5,73%	5,36E-02	0.3524 (0.32-0.42)	0.4097 (0.35-0.45)	0.4100 (0.35-0.46)
cg24091474	19	TYROBP	1,28%	4,83E-12	0,46%	1,17E-01	0.1650 (0.14-0.19)	0.1696 (0.15-0.19)	0.1779 (0.15-0.20)
cg07748854	19	x[a]	0,15%	7,43E-08	0,70%	6,19E-01	0.1385 (0.13-0.15)	0.1455 (0.13-0.16)	0.1400 (0.13-0.16)
cg17651625	19	ACTN4	0,97%	4,65E-13	0,73%	9,14E-03	0.2134 (0.20-0.23)	0.2207 (0.20-0.24)	0.2231 (0.21-0.25)
cg25420507	19	LGALS7	-0,70%	4,42E-09	0,16%	2,19E-01	0.1145 (0.10-0.13)	0.1160 (0.10-0.13)	0.1075 (0.10-0.12)
cg22649124	19	LGALS7B	-0,71%	1,15E-09	0,10%	2,97E-01	0.1161 (0.10-0.13)	0.1171 (0.11-0.13)	0.1090 (0.10-0.12)
cg12406027	19	PAK4	0,25%	2,09E-08	0,75%	3,93E-01	0.1410 (0.13-0.15)	0.1484 (0.14-0.16)	0.1434 (0.13-0.15)
cg15393221	19	PRX	2,51%	5,29E-09	-0,15%	6,70E-01	0.1982 (0.17-0.23)	0.1968 (0.17-0.22)	0.2233 (0.19-0.25)
cg13668129	19	HNRNPUL1	-0,16%	2,93E-08	0,10%	6,65E-01	0.1019 (0.10-0.11)	0.1029 (0.10-0.11)	0.1003 (0.09-0.11)
cg11666770	19	x[a]	-2,81%	5,97E-09	-1,31%	4,30E-03	0.8740 (0.83-0.92)	0.8609 (0.81-0.90)	0.8459 (0.80-0.88)

CpG	Chr.	Gene	**a)** ß-value current vs. never smokers	p-value	**b)** ß-value former vs. never smokers	p-value	Methylation ß-value as median (first quartile-third quartile)		
							Never smokers	Former smokers	Current smokers
cg03214420	19	C19orf61	-1,72%	2,96E-08	0,13%	4,93E-01	0.8212 (0.76-0.87)	0.8226 (0.77-0.87)	0.8040 (0.74-0.86)
cg21629821	19	DKFZp434J0226	0,89%	3,40E-08	0,35%	1,47E-02	0.0836 (0.07-0.10)	0.0871 (0.07-0.10)	0.0925 (0.08-0.11)
cg07626482	19	SLC1A5	-0,30%	7,56E-12	0,02%	5,41E-02	0.1397 (0.13-0.15)	0.1399 (0.13-0.15)	0.1367 (0.13-0.15)
cg24143611	19	SNORD23	1,02%	9,61E-08	0,13%	7,89E-01	0.8873 (0.87-0.90)	0.8886 (0.87-0.91)	0.8975 (0.88-0.91)
cg10126923	19	NKG7	1,18%	1,87E-10	-0,05%	1,63E-01	0.1571 (0.13-0.18)	0.1566 (0.13-0.18)	0.1689 (0.14-0.20)
cg21166330	19	FAM71E2	-2,16%	4,60E-09	-0,20%	8,40E-01	0.6606 (0.62-0.70)	0.6586 (0.61-0.70)	0.6390 (0.60-0.69)
cg11701312	19	RPS5	-1,41%	3,40E-10	-0,12%	5,71E-01	0.3002 (0.28-0.32)	0.2990 (0.28-0.32)	0.2861 (0.27-0.31)
cg21524538	20	SDCBP2	-1,48%	5,21E-10	0,06%	5,69E-01	0.8517 (0.83-0.88)	0.8523 (0.83-0.87)	0.8369 (0.81-0.86)
cg03980408	20	FKBP1A	-1,15%	3,82E-08	-0,16%	1,32E-01	0.9105 (0.88-0.93)	0.9089 (0.87-0.93)	0.8990 (0.86-0.93)
cg25551168	20	AVP	1,18%	2,20E-08	0,79%	5,88E-03	0.2428 (0.23-0.26)	0.2507 (0.23-0.27)	0.2546 (0.24-0.27)
cg16536918	20	AVP	1,10%	9,36E-09	0,57%	2,35E-03	0.4071 (0.38-0.44)	0.4128 (0.38-0.44)	0.4181 (0.39-0.46)
cg24257309	20	AVP	1,37%	2,80E-12	0,90%	2,93E-04	0.2470 (0.23-0.27)	0.2560 (0.24-0.27)	0.2607 (0.24-0.28)
cg04632887	20	AVP	1,48%	1,86E-08	0,90%	8,20E-03	0.5489 (0.53-0.57)	0.5579 (0.54-0.57)	0.5637 (0.54-0.59)
cg23169111	20	AVP	1,79%	2,99E-11	0,93%	1,02E-03	0.3708 (0.34-0.40)	0.3801 (0.35-0.41)	0.3887 (0.36-0.42)
cg16837898	20	x[a]	-0,36%	4,44E-09	0,02%	3,45E-01	0.1153 (0.10-0.13)	0.1155 (0.10-0.13)	0.1117 (0.10-0.13)
cg11436113	20	x[a]	-0,63%	4,91E-08	-0,02%	7,59E-01	0.3030 (0.27-0.33)	0.3028 (0.27-0.33)	0.2967 (0.26-0.33)
cg16201146	20	x[a]	-3,57%	7,91E-14	-0,38%	5,05E-02	0.6390 (0.57-0.72)	0.6352 (0.56-0.71)	0.6033 (0.55-0.68)
cg20244340	20	SLC24A3	-1,06%	1,50E-09	0,03%	1,56E-03	0.1822 (0.16-0.21)	0.1824 (0.16-0.20)	0.1716 (0.15-0.20)
cg24484138	20	C20orf112	0,96%	1,30E-08	0,56%	1,80E-01	0.1415 (0.13-0.16)	0.1471 (0.13-0.16)	0.1511 (0.13-0.16)
cg13127741	20	COMMD7	-1,29%	3,81E-15	0,24%	1,30E-02	0.3025 (0.27-0.35)	0.3048 (0.27-0.34)	0.2896 (0.25-0.33)
cg18884368	20	COMMD7	-0,30%	3,77E-09	0,02%	2,14E-01	0.0428 (0.04-0.05)	0.0430 (0.04-0.05)	0.0398 (0.03-0.05)
cg19593285	20	E2F1	-1,81%	3,09E-12	0,27%	7,75E-01	0.8546 (0.82-0.88)	0.8573 (0.83-0.89)	0.8365 (0.81-0.87)
cg26347170	20	CHMP4B	-1,19%	1,78E-09	0,04%	2,10E-02	0.8881 (0.86-0.91)	0.8885 (0.86-0.91)	0.8762 (0.85-0.90)
cg04759756	20	SLA2	-1,36%	5,85E-09	-0,09%	3,42E-01	0.8169 (0.77-0.86)	0.8160 (0.77-0.86)	0.8033 (0.76-0.85)
cg02121447	20	SNX21	2,23%	2,51E-08	-0,10%	4,41E-01	0.3844 (0.35-0.43)	0.3834 (0.35-0.43)	0.4067 (0.37-0.46)
cg13334949	20	SNX21/ACOT8	0,78%	3,43E-11	0,50%	1,88E-01	0.1621 (0.15-0.18)	0.1671 (0.15-0.18)	0.1699 (0.15-0.19)
cg18059369	20	x[a]	1,17%	4,29E-08	0,19%	3,24E-01	0.2693 (0.23-0.31)	0.2712 (0.23-0.31)	0.2810 (0.24-0.32)
cg02310296	20	MMP9	0,23%	5,64E-08	0,16%	1,54E-01	0.0714 (0.06-0.08)	0.0730 (0.07-0.08)	0.0737 (0.07-0.08)
cg10505873	20	MMP9	0,21%	8,48E-08	0,10%	3,50E-02	0.0745 (0.07-0.08)	0.0755 (0.07-0.08)	0.0766 (0.07-0.08)
cg12303084	20	ZMYND8	-0,63%	3,76E-15	0,08%	3,44E-02	0.0991 (0.09-0.11)	0.1000 (0.09-0.11)	0.0928 (0.08-0.11)
cg22214889	20	BCAS4	0,36%	6,65E-10	-0,04%	6,26E-01	0.0955 (0.09-0.10)	0.0950 (0.09-0.10)	0.0990 (0.09-0.11)
cg07339236	20	ATP9A	-1,10%	1,57E-21	-0,21%	1,43E-05	0.0828 (0.07-0.09)	0.0807 (0.07-0.09)	0.0717 (0.06-0.08)
cg04628369	20	PMEPA1	0,84%	2,59E-09	0,65%	6,59E-01	0.0839 (0.07-0.10)	0.0904 (0.08-0.10)	0.0922 (0.08-0.11)
cg16178855	20	PMEPA1	1,78%	1,00E-11	-0,20%	9,07E-02	0.7970 (0.77-0.82)	0.7950 (0.76-0.83)	0.8147 (0.79-0.84)
cg00071265	20	PSMA7	-0,25%	9,27E-09	0,11%	1,23E-01	0.1389 (0.13-0.15)	0.1400 (0.13-0.15)	0.1363 (0.13-0.15)
cg25143652	20	PTK6	-0,28%	7,61E-11	0,03%	3,61E-01	0.2903 (0.27-0.31)	0.2906 (0.28-0.31)	0.2875 (0.27-0.30)
cg17474524	20	PTK6	-0,50%	5,42E-11	-0,13%	4,82E-02	0.0908 (0.08-0.10)	0.0895 (0.08-0.10)	0.0858 (0.08-0.10)
cg14010720	20	PTK6	-0,36%	8,17E-08	0,11%	1,15E-01	0.2696 (0.24-0.30)	0.2706 (0.24-0.30)	0.2659 (0.23-0.30)
cg17132148	20	TPD52L2	2,83%	1,38E-09	0,47%	8,40E-01	0.7208 (0.68-0.76)	0.7255 (0.69-0.76)	0.7491 (0.71-0.78)
cg00871610	21	MIR802	-4,33%	5,46E-08	-1,81%	9,52E-05	0.5267 (0.47-0.57)	0.5085 (0.45-0.56)	0.4834 (0.42-0.54)
cg06595162	21	NCRNA00114	-2,71%	2,65E-08	-1,11%	4,58E-02	0.8358 (0.80-0.86)	0.8248 (0.79-0.85)	0.8088 (0.77-0.84)
cg23110422	21	ETS2	-3,68%	1,72E-16	-1,00%	2,92E-03	0.9005 (0.87-0.92)	0.8904 (0.86-0.92)	0.8637 (0.82-0.90)
cg08039560	21	x[a]	-1,23%	1,98E-08	0,31%	5,09E-01	0.7564 (0.72-0.79)	0.7596 (0.72-0.80)	0.7441 (0.71-0.79)
cg26472802	21	AIRE	-0,42%	1,40E-08	0,19%	6,60E-01	0.6904 (0.66-0.73)	0.6923 (0.66-0.74)	0.6862 (0.66-0.72)
cg22635096	21	ADARB1	3,36%	5,81E-11	0,26%	9,39E-01	0.2511 (0.20-0.30)	0.2537 (0.20-0.31)	0.2847 (0.23-0.36)
cg18704527	22	ARVCF	1,53%	2,44E-08	0,74%	1,80E-02	0.1678 (0.15-0.19)	0.1752 (0.16-0.20)	0.1831 (0.16-0.20)
cg06861736	22	MED15	-0,58%	4,36E-09	0,15%	3,55E-01	0.1817 (0.15-0.21)	0.1832 (0.16-0.21)	0.1759 (0.15-0.20)
cg26486070	22	GGT1	-2,57%	9,35E-10	-0,50%	3,84E-02	0.6481 (0.60-0.70)	0.6431 (0.58-0.69)	0.6224 (0.56-0.68)
cg26564117	22	PITPNB	1,56%	9,00E-08	0,37%	3,21E-02	0.1445 (0.12-0.18)	0.1482 (0.12-0.18)	0.1601 (0.13-0.19)
cg07127410	22	ZNRF3	4,82%	4,87E-09	-0,43%	9,46E-01	0.4102 (0.36-0.46)	0.4059 (0.36-0.46)	0.4584 (0.40-0.51)
cg07713946	22	LIMK2	-2,29%	3,38E-10	-0,32%	6,25E-02	0.8369 (0.79-0.88)	0.8337 (0.79-0.88)	0.8141 (0.77-0.86)
cg08548559	22	PIK3IP1	-0,56%	3,32E-10	0,16%	2,98E-01	0.0972 (0.09-0.11)	0.0988 (0.09-0.11)	0.0916 (0.08-0.10)
cg13003054	22	PATZ1	-0,21%	1,56E-09	0,14%	7,21E-02	0.0722 (0.07-0.08)	0.0736 (0.07-0.08)	0.0701 (0.06-0.08)
cg08305533	22	SFI1	4,96%	7,54E-10	1,53%	6,57E-02	0.3930 (0.34-0.44)	0.4083 (0.35-0.46)	0.4426 (0.38-0.49)
cg02532700	22	NCF4	-1,15%	1,88E-16	-0,29%	1,01E-02	0.0924 (0.08-0.11)	0.0895 (0.08-0.11)	0.0809 (0.07-0.10)
cg12741639	22	CYTH4	-1,57%	1,36E-10	0,50%	8,04E-01	0.1833 (0.15-0.23)	0.1883 (0.15-0.23)	0.1676 (0.14-0.21)
cg19035792	22	x[a]	0,72%	9,54E-08	0,25%	8,70E-01	0.9318 (0.92-0.95)	0.9343 (0.92-0.95)	0.9390 (0.93-0.95)
cg01127300	22	x[a]	-3,83%	3,08E-16	-1,08%	1,07E-02	0.4576 (0.40-0.51)	0.4469 (0.38-0.50)	0.4193 (0.34-0.48)
cg21649604	22	RPL3	-0,06%	3,54E-08	0,21%	2,65E-01	0.1023 (0.09-0.11)	0.1044 (0.10-0.11)	0.1017 (0.09-0.11)
cg18405341	22	ATF4	-1,88%	1,58E-08	0,25%	3,75E-01	0.4531 (0.42-0.49)	0.4556 (0.42-0.49)	0.4343 (0.40-0.47)
cg25636159	22	ADSL	-0,91%	5,84E-12	-0,07%	8,46E-03	0.9046 (0.88-0.93)	0.9039 (0.88-0.93)	0.8955 (0.86-0.92)
cg13370427	22	ADSL	0,33%	9,86E-08	0,26%	9,54E-02	0.2891 (0.26-0.32)	0.2917 (0.26-0.32)	0.2924 (0.26-0.31)
cg04873169	22	GRAMD4	1,63%	5,83E-09	-0,16%	8,97E-01	0.8241 (0.80-0.85)	0.8225 (0.80-0.84)	0.8404 (0.82-0.86)
cg20698113	22	PIM3	-5,02%	2,59E-08	-0,96%	8,17E-01	0.8182 (0.68-0.90)	0.8086 (0.69-0.89)	0.7680 (0.61-0.87)

Displayed are **a)** the results of the linear model calculated with M-value adjusted for age, sex, BMI, alcohol and white blood cell count (p-value), as well as the median ß-value methylation difference between current and never smokers for the F4 discovery panel with genome-wide significance ($p \leq 1E-07$) and **b)** the corresponding results of the same CpG sites for former smokers; sorted by chromosome and mapinfo (Genome build 37); [a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites; [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island; * CpG sites that remain significant in former smokers.

Table A.2. Replicated differentially methylated current smoking-associated CpG sites.

				a) F4 Discovery Panel			**b)** F3 Replication Panel				Methylation ß-value as median (first quartile-third quartile)			
CpG	Chr.	Mapinfo	Gene	Median ß-value in %	p-value	Explained variance in %	Median ß-value in %	p-value	Explained variance in %	**c)** Meta p-value	**a)** F4 Discovery Panel		**b)** F3 Replication Panel	
											Never smokers	Current smokers	Never smokers	Current smokers
cg09469355	1	2161886	SKI	-1,76	3,34E-10	3,26	-1,61	2,48E-07	5,44	5,71E-16	0.3299 (0.30-0.36)	0.3123 (0.28-0.34)	0.3021 (0.28-0.33)	0.2860 (0.26-0.31)
cg08884752	1	2162001	SKI	-4,01	4,43E-13	4,30	-2,82	1,25E-05	3,98	3,04E-17	0.6803 (0.63-0.73)	0.6401 (0.59-0.69)	0.5889 (0.54-0.63)	0.5607 (0.51-0.61)
cg12547807	1	9473751	x[a]	-2,14	6,52E-20	6,41	-1,46	3,43E-06	4,43	2,82E-24	0.1885 (0.16-0.22)	0.1672 (0.15-0.19)	0.1623 (0.14-0.19)	0.1476 (0.13-0.17)
cg04885881	1	11123118	x[a]	-7,41	1,35E-28	9,57	-6,71	6,62E-13	9,83	7,19E-40	0.3115 (0.25-0.39)	0.2374 (0.18-0.31)	0.2899 (0.24-0.36)	0.2228 (0.17-0.29)
cg21393163	1	12217629	x[a]	-0,69	1,54E-20	6,86	-0,68	9,33E-06	3,98	2,55E-24	0.0445 (0.04-0.05)	0.0376 (0.03-0.05)	0.0572 (0.05-0.07)	0.0503 (0.04-0.06)
cg21913886	1	15485346	TMEM51	-3,38	7,26E-11	3,45	-3,51	7,03E-06	4,05	2,49E-15	0.8917 (0.86-0.92)	0.8579 (0.81-0.90)	0.8314 (0.79-0.86)	0.7963 (0.74-0.84)
cg19713429	1	19810690	CAPZB	-0,45	1,90E-16	5,50	-0,49	4,52E-06	3,92	6,42E-21	0.0987 (0.09-0.11)	0.0942 (0.09-0.10)	0.0961 (0.09-0.11)	0.0912 (0.08-0.10)
cg27537125	1	25349681	x[a]	-1,21	2,70E-30	10,45	-0,73	4,79E-10	7,96	2,07E-38	0.0877 (0.08-0.10)	0.0756 (0.07-0.09)	0.0840 (0.08-0.09)	0.0767 (0.07-0.08)
cg15542713	1	42385581	HIVEP3	9,52	2,54E-08	2,03	9,95	3,65E-13	9,51	3,56E-18	0.4002 (0.31-0.50)	0.4954 (0.38-0.60)	0.3784 (0.30-0.46)	0.4779 (0.40-0.57)
cg24049493	1	42385941	HIVEP3	4,05	3,73E-14	3,77	3,83	1,13E-12	9,16	1,07E-24	0.1219 (0.10-0.15)	0.1624 (0.12-0.21)	0.1112 (0.10-0.14)	0.1496 (0.12-0.18)
cg23090529	1	51442133	x[a]	-2,28	9,90E-11	3,66	-1,64	9,52E-06	4,06	4,55E-15	0.2061 (0.17-0.25)	0.1834 (0.16-0.22)	0.1670 (0.15-0.19)	0.1506 (0.13-0.17)
cg21140898	1	51442318	x[a]	-3,28	1,08E-10	3,55	-2,62	1,83E-05	3,62	9,33E-15	0.3534 (0.31-0.40)	0.3206 (0.28-0.37)	0.3454 (0.31-0.38)	0.3192 (0.28-0.35)
cg19406367	1	66999929	SGIP1	3,11	2,79E-08	2,63	4,54	4,25E-06	4,41	7,04E-13	0.8147 (0.76-0.86)	0.8458 (0.80-0.89)	0.6663 (0.62-0.72)	0.7117 (0.65-0.77)
cg25189904	1	68299493	GNG12	-8,19	1,71E-39	14,27	-7,40	4,13E-15	11,42	9,11E-53	0.2740 (0.23-0.33)	0.1921 (0.16-0.24)	0.2856 (0.24-0.33)	0.2116 (0.17-0.26)
cg09662411	1	92946132	GFI1	-3,73	8,01E-09	2,56	-9,49	3,76E-11	8,29	2,11E-17	0.4805 (0.44-0.52)	0.4432 (0.39-0.50)	0.5557 (0.49-0.63)	0.4608 (0.37-0.53)
cg18146737	1	92946700	GFI1	-1,55	4,79E-17	4,80	-2,87	3,22E-08	5,86	9,33E-24	0.9565 (0.95-0.96)	0.9410 (0.88-0.96)	0.9374 (0.92-0.95)	0.9086 (0.83-0.94)
cg12876356	1	92946825	GFI1	-8,48	2,92E-17	5,46	-22,63	6,18E-18	13,59	2,47E-32	0.5475 (0.49-0.62)	0.4627 (0.37-0.55)	0.7010 (0.56-0.80)	0.4746 (0.33-0.60)
cg18316974	1	92947035	GFI1	-1,86	1,60E-16	4,43	-1,73	6,34E-10	7,29	7,28E-25	0.9544 (0.94-0.96)	0.9357 (0.90-0.96)	0.9496 (0.94-0.96)	0.9323 (0.89-0.95)
cg09935388	1	92947588	GFI1	-15,31	3,27E-24	7,39	-16,47	3,98E-26	18,93	1,26E-46	0.5765 (0.49-0.70)	0.4234 (0.32-0.54)	0.4757 (0.40-0.54)	0.3110 (0.24-0.39)
cg11231349	1	162050656	NOS1AP	-1,92	3,94E-09	2,51	-3,34	3,21E-06	4,23	7,09E-14	0.8372 (0.79-0.87)	0.8180 (0.77-0.85)	0.8281 (0.79-0.85)	0.7947 (0.74-0.85)
cg08709672	1	206224334	AVPR1B	-2,58	1,93E-20	7,09	-2,44	1,10E-10	7,83	1,46E-29	0.6776 (0.66-0.70)	0.6518 (0.64-0.67)	0.6442 (0.62-0.66)	0.6198 (0.60-0.64)
cg20295214	1	206226794	AVPR1B	-2,98	1,13E-09	2,99	-4,36	4,14E-11	7,93	2,22E-18	0.8467 (0.82-0.87)	0.8169 (0.78-0.85)	0.7816 (0.75-0.81)	0.7380 (0.71-0.78)
cg03547355	1	227003060	x[a]	-2,76	1,06E-12	4,75	-2,21	1,65E-06	4,49	9,10E-18	0.6024 (0.57-0.63)	0.5748 (0.53-0.61)	0.5325 (0.50-0.55)	0.5104 (0.47-0.54)
cg17819085	1	235122803	x[a]	-2,30	6,87E-08	2,03	-1,98	2,53E-05	3,79	8,54E-12	0.7041 (0.67-0.73)	0.6811 (0.65-0.71)	0.6214 (0.59-0.64)	0.6016 (0.58-0.63)
cg23079012	2	8343710	x[a]	-1,10	8,29E-24	7,15	-0,96	1,59E-11	8,43	9,51E-34	0.9594 (0.95-0.96)	0.9484 (0.93-0.96)	0.9578 (0.95-0.96)	0.9482 (0.93-0.96)
cg06635952	2	70025869	ANXA4	2,00	2,37E-21	6,76	1,38	4,59E-09	7,10	7,88E-29	0.1388 (0.12-0.16)	0.1588 (0.14-0.18)	0.1239 (0.11-0.14)	0.1378 (0.12-0.15)
cg26271591	2	178125956	NFE2L2	-5,57	1,75E-14	4,91	-4,80	8,08E-09	6,73	9,43E-22	0.2569 (0.19-0.32)	0.2012 (0.15-0.26)	0.2535 (0.20-0.30)	0.2055 (0.16-0.25)
cg23667432	2	233244439	ALPP	-1,57	1,55E-11	3,27	-1,46	4,59E-05	3,46	3,60E-15	0.3401 (0.32-0.36)	0.3244 (0.31-0.35)	0.3272 (0.30-0.35)	0.3127 (0.29-0.33)
cg03188382	2	233245886	ALPP	-1,19	6,73E-11	3,66	-1,15	2,70E-05	3,59	8,69E-15	0.2432 (0.23-0.26)	0.2313 (0.21-0.25)	0.2194 (0.20-0.24)	0.2079 (0.19-0.23)
cg19713851	2	233246594	ALPP	-6,47	1,28E-12	4,79	-4,45	6,72E-06	4,12	4,47E-17	0.4950 (0.43-0.56)	0.4302 (0.37-0.50)	0.4462 (0.39-0.51)	0.4017 (0.33-0.47)
cg27241845	2	233250370	x[a]	-8,74	2,79E-35	10,96	-8,36	2,07E-19	13,78	6,03E-53	0.4617 (0.41-0.52)	0.3743 (0.31-0.42)	0.4457 (0.41-0.49)	0.3621 (0.31-0.43)
cg03329539	2	233283329	ALPPL2[b]	-2,39	3,66E-59	19,04	-2,57	8,34E-26	19,36	3,73E-83	0.1598 (0.15-0.17)	0.1359 (0.13-0.15)	0.1558 (0.14-0.17)	0.1301 (0.12-0.14)
cg06644428	2	233284112	ALPPL2[b]	-7,05	6,37E-31	12,14	-3,75	1,94E-15	11,75	1,04E-44	0.2142 (0.17-0.27)	0.1437 (0.11-0.19)	0.1147 (0.08-0.15)	0.0772 (0.06-0.10)
cg05951221	2	233284402	ALPPL2[b]	-5,21	8,92E-104	31,07	-4,53	2,06E-41	28,00	4,39E-143	0.2199 (0.20-0.24)	0.1678 (0.15-0.19)	0.1757 (0.16-0.19)	0.1304 (0.12-0.14)
cg21566642	2	233284661	ALPPL2[b]	-16,70	6,90E-138	36,24	-15,58	8,82E-41	27,13	7,86E-175	0.5658 (0.51-0.61)	0.3988 (0.35-0.45)	0.5511 (0.50-0.59)	0.3953 (0.31-0.44)
cg01940273	2	233284934	ALPPL2[b]	-7,89	9,28E-114	31,50	-7,53	5,46E-43	28,33	1,92E-154	0.3437 (0.32-0.37)	0.2648 (0.24-0.29)	0.2941 (0.27-0.31)	0.2189 (0.20-0.25)
cg13193840	2	233285289	ALPPL2[b]	-2,41	1,53E-23	8,21	-1,91	3,95E-07	4,85	9,89E-29	0.1525 (0.13-0.17)	0.1285 (0.11-0.15)	0.1388 (0.12-0.16)	0.1197 (0.10-0.14)
cg17024919	3	21792248	ZNF385D	-3,09	4,06E-09	3,38	-2,47	2,57E-05	3,57	4,81E-13	0.1708 (0.13-0.25)	0.1398 (0.10-0.21)	0.1593 (0.12-0.24)	0.1346 (0.10-0.19)
cg15693572	3	22412385	x[a]	10,87	3,84E-14	4,50	6,90	1,74E-06	4,60	3,69E-19	0.2942 (0.22-0.41)	0.4029 (0.29-0.53)	0.3159 (0.23-0.40)	0.3850 (0.29-0.50)
cg23480021	3	22412746	x[a]	18,63	2,67E-22	8,74	15,77	1,95E-10	8,09	3,63E-31	0.4259 (0.30-0.60)	0.6122 (0.44-0.78)	0.4266 (0.29-0.58)	0.5844 (0.43-0.78)
cg03274391	3	22413232	x[a]	18,06	1,67E-25	8,79	11,93	3,74E-10	7,78	5,47E-34	0.3190 (0.23-0.47)	0.4997 (0.34-0.70)	0.2872 (0.21-0.41)	0.4065 (0.29-0.55)
cg00501876	3	39193251	CSRNP1	-4,43	1,39E-25	8,79	-3,12	4,12E-06	4,59	2,71E-29	0.3692 (0.33-0.40)	0.3249 (0.30-0.36)	0.4298 (0.38-0.47)	0.3986 (0.35-0.43)
cg18642234	3	49394622	GPX1	-1,62	1,44E-12	4,19	-1,11	1,41E-05	3,61	1,08E-16	0.2316 (0.21-0.25)	0.2154 (0.20-0.23)	0.2206 (0.21-0.24)	0.2095 (0.19-0.22)
cg15417641	3	53700141	CACNA1D	7,92	1,25E-16	4,90	5,28	1,90E-06	4,61	1,65E-21	0.6818 (0.58-0.76)	0.7610 (0.68-0.83)	0.6019 (0.52-0.68)	0.6546 (0.60-0.73)
cg00336149	3	53700195	CACNA1D	6,46	1,57E-12	3,53	3,71	4,40E-06	4,27	3,54E-17	0.2530 (0.20-0.32)	0.3175 (0.25-0.39)	0.2315 (0.19-0.29)	0.2686 (0.22-0.33)
cg21188533	3	53700263	CACNA1D	10,18	1,12E-10	3,10	5,56	4,38E-05	3,42	2,33E-14	0.5724 (0.45-0.69)	0.6742 (0.55-0.76)	0.5309 (0.41-0.63)	0.5865 (0.50-0.70)
cg19859270	3	98251294	GPR15	-1,31	9,00E-24	8,42	-1,94	2,79E-21	16,20	2,42E-42	0.9560 (0.95-0.96)	0.9430 (0.93-0.95)	0.9502 (0.94-0.96)	0.9307 (0.92-0.95)

CpG	Chr.	Mapinfo	Gene	**a)** F4 Discovery Panel Median ß-value in %	p-value	Explained variance in %	**b)** F3 Replication Panel Median ß-value in %	p-value	Explained variance in %	**c)** Meta p-value	Methylation ß-value as median (first quartile-third quartile) **a)** F4 Discovery Panel Never smokers	Current smokers	**b)** F3 Replication Panel Never smokers	Current smokers
cg02657160	3	98311063	CPOX	-1,48	1,67E-09	3,36	-1,37	3,34E-06	4,31	3,01E-14	0.9255 (0.91-0.94)	0.9107 (0.88-0.93)	0.9111 (0.90-0.93)	0.8974 (0.87-0.92)
cg25197194	3	128758787	CCDC48	4,62	5,71E-10	3,07	4,88	8,69E-06	4,21	2,42E-14	0.4321 (0.37-0.50)	0.4783 (0.42-0.55)	0.4084 (0.36-0.48)	0.4571 (0.40-0.51)
cg08202836	3	196373751	LRRC33	1,51	3,85E-08	2,00	1,85	1,60E-07	5,70	6,75E-14	0.8802 (0.85-0.90)	0.8952 (0.88-0.91)	0.8700 (0.84-0.89)	0.8885 (0.87-0.91)
cg21121843	4	3203982	HTT	-2,07	1,79E-13	4,66	-1,32	9,78E-06	3,99	9,83E-18	0.1334 (0.11-0.16)	0.1127 (0.10-0.13)	0.1209 (0.11-0.14)	0.1078 (0.09-0.13)
cg19719391	4	26789915	x[a]	4,18	2,32E-13	4,36	3,14	3,22E-05	3,54	4,53E-17	0.5305 (0.48-0.57)	0.5723 (0.52-0.61)	0.4848 (0.45-0.52)	0.5163 (0.47-0.55)
cg24556382	4	174173455	GALNT7	-4,25	2,64E-10	3,49	-4,87	1,10E-05	3,88	1,39E-14	0.7934 (0.74-0.85)	0.7510 (0.68-0.81)	0.7247 (0.67-0.77)	0.6760 (0.62-0.73)
cg11554391	5	321320	AHRR	-1,03	5,07E-24	7,71	-0,69	2,72E-09	6,27	1,22E-31	0.0911 (0.08-0.10)	0.0808 (0.07-0.09)	0.0765 (0.07-0.08)	0.0696 (0.06-0.08)
cg17924476	5	323794	AHRR	5,80	2,67E-16	5,79	3,25	2,13E-07	5,39	3,41E-22	0.2628 (0.23-0.30)	0.3208 (0.26-0.36)	0.2556 (0.23-0.29)	0.2881 (0.25-0.33)
cg08606254	5	323969	AHRR	0,70	9,25E-09	3,43	0,69	6,18E-07	5,19	4,26E-14	0.9490 (0.93-0.96)	0.9560 (0.94-0.97)	0.9440 (0.93-0.95)	0.9509 (0.94-0.96)
cg12806681	5	368394	AHRR	-2,44	3,63E-33	10,38	-2,42	2,60E-17	13,06	8,60E-49	0.9537 (0.94-0.96)	0.9293 (0.90-0.95)	0.9459 (0.93-0.96)	0.9217 (0.89-0.94)
cg03991871	5	368447	AHRR	-5,73	4,84E-41	12,72	-7,88	1,23E-23	17,61	8,03E-63	0.9469 (0.92-0.96)	0.8896 (0.80-0.93)	0.9175 (0.89-0.94)	0.8387 (0.69-0.89)
cg23916896	5	368804	AHRR	-2,11	9,79E-29	10,21	-1,94	4,14E-20	15,62	9,07E-47	0.1286 (0.11-0.15)	0.1075 (0.10-0.12)	0.1158 (0.10-0.13)	0.0964 (0.09-0.11)
cg11902777	5	368843	AHRR	-1,12	4,04E-40	14,09	-0,95	1,38E-17	13,62	5,48E-56	0.0379 (0.03-0.05)	0.0267 (0.02-0.03)	0.0410 (0.03-0.05)	0.0315 (0.03-0.04)
cg01899089	5	369969	AHRR	-2,86	8,95E-32	11,21	-2,26	2,49E-13	10,12	2,05E-43	0.2486 (0.23-0.27)	0.2200 (0.20-0.24)	0.2381 (0.22-0.26)	0.2155 (0.20-0.24)
cg05575921	5	373378	AHRR	-24,40	2,54E-182	41,02	-23,29	1,81E-64	39,69	8,70E-244	0.8841 (0.86-0.90)	0.6401 (0.60-0.73)	0.9028 (0.88-0.92)	0.6699 (0.58-0.77)
cg26703534	5	377358	AHRR	-6,17	2,14E-38	10,90	-5,77	7,83E-25	17,21	3,66E-61	0.6269 (0.60-0.65)	0.5652 (0.52-0.60)	0.6180 (0.60-0.63)	0.5603 (0.53-0.60)
cg01097768	5	378854	AHRR	-2,40	1,22E-11	3,74	-1,89	5,02E-07	5,45	3,45E-17	0.7457 (0.70-0.79)	0.7217 (0.67-0.76)	0.6593 (0.63-0.69)	0.6404 (0.59-0.68)
cg14817490	5	392920	AHRR	-3,81	4,08E-50	14,40	-3,17	2,37E-19	13,72	1,45E-67	0.1350 (0.11-0.16)	0.0968 (0.08-0.12)	0.1327 (0.12-0.16)	0.1010 (0.08-0.13)
cg25648203	5	395444	AHRR	-7,96	4,73E-47	12,98	-7,55	2,03E-23	16,96	1,01E-68	0.8642 (0.83-0.89)	0.7846 (0.72-0.84)	0.8473 (0.81-0.87)	0.7719 (0.71-0.82)
cg21161138	5	399360	AHRR	-10,45	8,58E-81	21,63	-10,60	1,19E-40	28,49	1,49E-119	0.6998 (0.67-0.73)	0.5954 (0.53-0.65)	0.7304 (0.70-0.75)	0.6244 (0.57-0.68)
cg03604011	5	400201	AHRR	3,40	7,87E-30	9,49	1,93	5,66E-08	6,18	1,69E-35	0.0759 (0.06-0.10)	0.1099 (0.08-0.15)	0.0681 (0.05-0.08)	0.0874 (0.06-0.11)
cg24090911	5	400732	AHRR	-4,08	3,41E-14	4,20	-4,21	5,68E-08	5,97	1,14E-20	0.8246 (0.79-0.86)	0.7838 (0.74-0.83)	0.7998 (0.76-0.84)	0.7577 (0.71-0.80)
cg13039251	5	32018601	PDZD2	7,15	3,81E-17	5,53	4,67	6,24E-07	4,87	1,60E-22	0.7536 (0.69-0.82)	0.8251 (0.77-0.87)	0.7054 (0.64-0.77)	0.7521 (0.69-0.82)
cg05673882	5	74862702	POLK	-2,65	4,49E-23	8,96	-1,99	6,02E-08	5,97	2,99E-29	0.1515 (0.12-0.19)	0.1250 (0.10-0.16)	0.1427 (0.12-0.17)	0.1228 (0.10-0.15)
cg26908328	5	79552537	SERINC5	-0,60	2,73E-08	2,76	-0,65	1,83E-06	4,46	3,36E-13	0.0824 (0.07-0.09)	0.0765 (0.07-0.09)	0.0783 (0.07-0.09)	0.0718 (0.07-0.08)
cg16786458	5	149108820	PPARGC1B	2,16	2,79E-11	3,24	1,57	2,34E-05	3,61	3,18E-15	0.1744 (0.16-0.20)	0.1960 (0.17-0.22)	0.1752 (0.16-0.20)	0.1909 (0.17-0.22)
cg14580211	5	150161299	C5orf62	-3,91	2,19E-12	3,94	-4,61	4,44E-12	8,85	3,07E-22	0.6909 (0.65-0.73)	0.6519 (0.61-0.69)	0.6447 (0.61-0.68)	0.5986 (0.55-0.64)
cg12513616	5	177370977	x[a]	-0,87	2,88E-17	5,54	-1,20	6,99E-08	5,59	1,24E-23	0.2101 (0.20-0.23)	0.2014 (0.19-0.21)	0.1823 (0.17-0.19)	0.1703 (0.16-0.18)
cg01882991	6	6677756	x[a]	-1,40	1,74E-08	3,36	-1,49	1,84E-08	5,69	5,15E-15	0.6357 (0.62-0.66)	0.6217 (0.60-0.64)	0.6087 (0.59-0.62)	0.5937 (0.58-0.61)
cg06126421	6	30720080	x[a]	-17,05	1,72E-75	23,60	-17,89	3,73E-36	24,58	8,42E-110	0.7724 (0.70-0.82)	0.6019 (0.48-0.69)	0.6919 (0.63-0.74)	0.5131 (0.42-0.61)
cg14753356	6	30720108	x[a]	-5,37	8,14E-24	8,58	-6,41	2,98E-23	17,61	5,90E-44	0.2383 (0.20-0.28)	0.1846 (0.15-0.22)	0.2451 (0.20-0.28)	0.1810 (0.15-0.22)
cg24859433	6	30720203	x[a]	-4,03	3,06E-41	13,04	-4,92	1,14E-19	14,65	3,34E-59	0.9322 (0.91-0.95)	0.8919 (0.85-0.92)	0.8774 (0.85-0.90)	0.8282 (0.77-0.87)
cg15342087	6	30720209	x[a]	-3,36	1,71E-42	13,50	-5,42	2,89E-27	19,61	1,19E-67	0.9371 (0.92-0.95)	0.9036 (0.87-0.93)	0.9067 (0.89-0.92)	0.8525 (0.81-0.89)
cg17619755	6	31760629	VARS	4,05	2,64E-18	5,64	2,62	3,32E-06	4,47	8,30E-23	0.6519 (0.61-0.69)	0.6924 (0.64-0.73)	0.6169 (0.59-0.65)	0.6431 (0.60-0.69)
cg10807309	6	31761062	VARS	1,83	1,25E-12	3,46	2,01	8,47E-06	4,05	5,50E-17	0.1534 (0.13-0.18)	0.1717 (0.15-0.20)	0.1440 (0.13-0.17)	0.1641 (0.13-0.20)
cg15474579	6	36645812	CDKN1A	-3,23	9,30E-09	2,77	-2,14	2,94E-05	3,26	1,26E-12	0.3563 (0.31-0.41)	0.3240 (0.29-0.37)	0.2999 (0.27-0.33)	0.2784 (0.24-0.31)
cg00931843	6	155442993	TIAM2	6,37	1,69E-10	2,98	5,66	1,09E-07	5,73	1,37E-16	0.2963 (0.23-0.39)	0.3600 (0.27-0.47)	0.3293 (0.25-0.42)	0.3859 (0.29-0.46)
cg00921574	7	1513873	INTS1	-0,45	8,71E-10	2,82	-0,52	4,90E-05	3,04	1,93E-13	0.1005 (0.09-0.11)	0.0960 (0.08-0.11)	0.0767 (0.07-0.08)	0.0715 (0.07-0.08)
cg19717773	7	2847554	GNA12	-4,05	1,15E-08	3,14	-4,12	1,95E-05	3,70	1,07E-12	0.7015 (0.65-0.76)	0.6610 (0.60-0.72)	0.6278 (0.58-0.68)	0.5866 (0.54-0.64)
cg02451831	7	26578098	KIAA0087	-2,30	2,79E-13	4,22	-3,10	1,35E-09	7,17	3,46E-21	0.8243 (0.80-0.86)	0.8013 (0.76-0.83)	0.7564 (0.73-0.79)	0.7254 (0.69-0.77)
cg08972170	7	30185776	C7orf41	5,33	1,53E-09	2,69	4,95	4,51E-08	6,05	7,01E-16	0.5272 (0.47-0.59)	0.5805 (0.51-0.64)	0.4908 (0.44-0.54)	0.5403 (0.48-0.60)
cg19089201	7	45002287	MYO1G	1,86	9,26E-13	4,20	2,87	5,82E-09	6,63	4,37E-20	0.9094 (0.88-0.93)	0.9280 (0.90-0.95)	0.8742 (0.84-0.90)	0.9030 (0.87-0.93)
cg22132788	7	45002486	MYO1G	6,68	1,99E-34	11,04	5,02	8,44E-18	13,68	1,54E-50	0.8599 (0.81-0.90)	0.9267 (0.88-0.95)	0.8727 (0.83-0.90)	0.9229 (0.88-0.95)
cg04180046	7	45002736	MYO1G	2,17	5,92E-14	4,32	2,10	1,33E-09	7,32	6,63E-22	0.2316 (0.21-0.25)	0.2533 (0.23-0.28)	0.1987 (0.18-0.22)	0.2196 (0.19-0.25)
cg12803068	7	45002919	MYO1G	14,96	7,08E-30	10,34	15,12	1,04E-15	12,29	6,33E-44	0.7382 (0.60-0.85)	0.8878 (0.77-0.94)	0.7014 (0.60-0.81)	0.8527 (0.71-0.92)
cg07826859	7	45020086	MYO1G	-5,12	8,03E-27	9,12	-3,78	5,60E-10	8,01	4,66E-35	0.5836 (0.54-0.62)	0.5324 (0.48-0.57)	0.5631 (0.53-0.59)	0.5254 (0.48-0.56)
cg03440944	7	45023329	C7orf40	-2,42	1,68E-11	3,45	-1,42	2,54E-05	3,67	2,11E-15	0.7533 (0.72-0.78)	0.7290 (0.70-0.76)	0.6934 (0.66-0.73)	0.6792 (0.64-0.71)
cg21322436	7	145812842	CNTNAP2	-4,74	1,43E-24	8,80	-4,46	5,47E-10	6,78	6,49E-33	0.2855 (0.25-0.32)	0.2381 (0.21-0.27)	0.2849 (0.26-0.31)	0.2403 (0.22-0.27)
cg25949550	7	145814306	CNTNAP2	-1,38	4,53E-29	9,12	-0,95	7,21E-14	10,50	2,61E-41	0.0852 (0.08-0.09)	0.0714 (0.06-0.08)	0.0840 (0.08-0.09)	0.0745 (0.07-0.08)

CpG	Chr.	Mapinfo	Gene	**a)** F4 Discovery Panel Median ß-value in %	p-value	Explained variance in %	**b)** F3 Replication Panel Median ß-value in %	p-value	Explained variance in %	**c)** Meta p-value	Methylation ß-value as median (first quartile-third quartile) **a)** F4 Discovery Panel Never smokers	Current smokers	**b)** F3 Replication Panel Never smokers	Current smokers
cg11207515	7	146904205	CNTNAP2	8,28	2,53E-10	3,44	5,29	2,16E-07	5,30	3,77E-16	0.2725 (0.21-0.35)	0.3552 (0.27-0.44)	0.2166 (0.17-0.27)	0.2695 (0.21-0.33)
cg17372101	7	147500722	CNTNAP2	7,23	4,17E-12	2,97	4,09	1,05E-07	5,06	2,74E-18	0.3192 (0.27-0.38)	0.3915 (0.33-0.47)	0.3247 (0.28-0.38)	0.3656 (0.32-0.42)
cg12276019	8	11057947	XKR6	-0,77	8,75E-08	1,92	-0,83	1,89E-05	3,31	8,52E-12	0.1235 (0.11-0.14)	0.1158 (0.10-0.14)	0.0974 (0.09-0.11)	0.0891 (0.08-0.10)
cg24540678	8	28258603	x[a]	-0,68	2,30E-17	6,23	-0,51	1,91E-07	5,40	2,82E-23	0.0855 (0.08-0.09)	0.0787 (0.07-0.09)	0.0856 (0.08-0.09)	0.0805 (0.07-0.09)
cg13518625	8	29522838	x[a]	-0,49	1,15E-13	4,48	-0,70	3,45E-08	6,21	2,46E-20	0.0565 (0.05-0.07)	0.0515 (0.04-0.06)	0.0567 (0.05-0.06)	0.0497 (0.04-0.06)
cg19589396	8	103937374	x[a]	-3,86	4,23E-14	4,85	-1,83	2,89E-05	3,65	8,09E-18	0.7511 (0.71-0.80)	0.7125 (0.67-0.77)	0.6884 (0.65-0.74)	0.6701 (0.62-0.71)
cg25305703	8	128378218	x[a]	-6,05	2,09E-08	2,97	-6,17	6,16E-09	6,16	2,75E-15	0.7103 (0.65-0.76)	0.6498 (0.58-0.73)	0.6679 (0.62-0.71)	0.6062 (0.54-0.67)
cg12075928	8	141801307	PTK2	-7,80	1,40E-18	6,80	-4,97	5,27E-08	6,03	4,90E-25	0.4672 (0.39-0.53)	0.3892 (0.32-0.48)	0.4791 (0.41-0.54)	0.4294 (0.36-0.48)
cg26361535	8	144576604	ZC3H3	-5,54	4,14E-09	2,83	-6,35	4,39E-07	4,90	1,35E-14	0.7593 (0.70-0.81)	0.7040 (0.63-0.77)	0.7442 (0.69-0.79)	0.6807 (0.64-0.75)
cg13787850	9	102195951	x[a]	-3,10	2,14E-10	2,87	-3,83	1,04E-06	5,05	1,26E-15	0.2680 (0.21-0.33)	0.2370 (0.18-0.29)	0.2767 (0.24-0.32)	0.2383 (0.19-0.29)
cg01692968	9	108005349	x[a]	-1,42	7,30E-11	4,23	-1,21	3,33E-07	4,93	1,50E-16	0.1619 (0.15-0.18)	0.1477 (0.13-0.16)	0.1370 (0.12-0.15)	0.1249 (0.11-0.14)
cg13910681	9	130740599	FAM102A	-0,41	5,92E-08	2,28	-0,40	1,65E-05	3,35	5,02E-12	0.0782 (0.07-0.08)	0.0741 (0.07-0.08)	0.0737 (0.07-0.08)	0.0697 (0.07-0.08)
cg22539182	10	850393	x[a]	1,41	7,40E-13	3,55	0,93	3,99E-07	5,27	1,62E-18	0.1902 (0.18-0.20)	0.2044 (0.19-0.22)	0.1757 (0.17-0.19)	0.1850 (0.17-0.20)
cg25953130	10	63753550	ARID5B	-4,69	6,34E-10	3,41	-5,41	2,09E-07	5,07	9,72E-16	0.4928 (0.41-0.56)	0.4459 (0.36-0.52)	0.4823 (0.41-0.54)	0.4282 (0.33-0.51)
cg27312979	10	97094453	SORBS1	5,02	2,74E-16	5,30	3,86	1,98E-06	4,26	3,65E-21	0.5558 (0.51-0.60)	0.6061 (0.57-0.65)	0.5413 (0.51-0.59)	0.5799 (0.54-0.62)
cg25421530	10	97200987	SORBS1	1,60	4,22E-09	3,02	2,78	1,88E-06	4,53	4,78E-14	0.8762 (0.84-0.90)	0.8922 (0.86-0.92)	0.8117 (0.77-0.84)	0.8395 (0.81-0.87)
cg01744331	11	2722358	KCNQ1OT1	-1,21	3,51E-10	3,11	-1,35	4,98E-08	5,73	1,53E-16	0.9380 (0.93-0.95)	0.9259 (0.91-0.94)	0.9294 (0.91-0.94)	0.9160 (0.90-0.93)
cg07123182	11	2722391	KCNQ1OT1	-0,91	8,64E-13	4,57	-0,81	3,11E-09	6,57	2,36E-20	0.9527 (0.94-0.96)	0.9436 (0.93-0.95)	0.9479 (0.94-0.95)	0.9397 (0.92-0.95)
cg16556677	11	2722401	KCNQ1OT1	-3,84	6,58E-10	3,70	-3,29	2,57E-07	4,91	1,21E-15	0.8507 (0.81-0.88)	0.8123 (0.77-0.85)	0.8239 (0.79-0.85)	0.7910 (0.74-0.83)
cg26963277	11	2722407	KCNQ1OT1	-2,86	4,52E-22	8,68	-1,93	2,16E-09	6,78	7,20E-30	0.9013 (0.88-0.92)	0.8726 (0.84-0.90)	0.9110 (0.89-0.93)	0.8916 (0.86-0.91)
cg04039799	11	19745484	NAV2	-1,62	5,24E-12	3,96	-0,74	2,46E-05	3,90	6,64E-16	0.1954 (0.17-0.23)	0.1792 (0.16-0.20)	0.1845 (0.16-0.21)	0.1770 (0.15-0.20)
cg09197783	11	57176720	SLC43A3	-3,04	5,23E-08	2,83	-2,57	1,99E-06	4,54	7,27E-13	0.7554 (0.72-0.80)	0.7250 (0.69-0.76)	0.6918 (0.66-0.72)	0.6661 (0.64-0.70)
cg16611234	11	58870075	x[a]	-1,25	8,31E-12	3,18	-0,98	1,53E-05	3,51	6,36E-16	0.1449 (0.13-0.16)	0.1323 (0.12-0.15)	0.1228 (0.11-0.13)	0.1130 (0.10-0.13)
cg19254163	11	60623782	GPR44	-3,88	4,82E-09	3,30	-2,92	1,33E-08	6,04	9,35E-16	0.5778 (0.54-0.61)	0.5390 (0.50-0.58)	0.5264 (0.50-0.55)	0.4972 (0.46-0.53)
cg21611682	11	68138269	LRP5	-5,24	1,09E-18	6,50	-5,38	3,16E-20	14,13	9,90E-36	0.5276 (0.50-0.56)	0.4752 (0.43-0.52)	0.4844 (0.46-0.51)	0.4306 (0.40-0.47)
cg14624207	11	68142198	LRP5	-5,15	2,94E-12	3,77	-3,59	5,74E-10	7,04	2,11E-20	0.4747 (0.44-0.51)	0.4233 (0.37-0.47)	0.4327 (0.40-0.47)	0.3968 (0.36-0.43)
cg01901332	11	75031054	ARRB1	-5,07	1,70E-09	3,57	-4,34	5,57E-12	8,98	8,35E-19	0.6741 (0.63-0.72)	0.6234 (0.57-0.68)	0.6406 (0.61-0.68)	0.5972 (0.55-0.64)
cg11660018	11	86510915	PRSS23	-3,87	1,29E-22	6,88	-2,96	5,88E-13	9,22	6,25E-34	0.2926 (0.27-0.32)	0.2538 (0.23-0.28)	0.2608 (0.24-0.28)	0.2312 (0.21-0.25)
cg23771366	11	86510998	PRSS23	-2,26	7,62E-20	7,05	-1,79	5,86E-09	6,37	2,97E-27	0.1993 (0.18-0.22)	0.1767 (0.16-0.19)	0.1859 (0.17-0.20)	0.1680 (0.15-0.18)
cg03234777	11	118095544	AMICA1	-0,69	5,16E-15	5,03	-0,56	7,93E-06	4,08	2,68E-19	0.0561 (0.05-0.07)	0.0492 (0.04-0.06)	0.0626 (0.06-0.07)	0.0570 (0.05-0.06)
cg26282236	12	1025755	RAD52	2,46	8,24E-08	2,58	3,65	1,66E-05	3,88	7,11E-12	0.5699 (0.52-0.63)	0.5945 (0.55-0.65)	0.4957 (0.44-0.55)	0.5322 (0.49-0.59)
cg02583484	12	54677008	HNRNPA1	-2,09	7,10E-20	6,72	-1,77	1,90E-08	6,35	9,38E-27	0.1476 (0.12-0.18)	0.1267 (0.11-0.15)	0.1380 (0.12-0.16)	0.1203 (0.10-0.14)
cg04158018	12	54696210	NFE2	-1,25	8,40E-08	2,41	-1,21	1,38E-05	3,89	6,21E-12	0.1587 (0.13-0.19)	0.1462 (0.13-0.17)	0.1395 (0.12-0.16)	0.1275 (0.11-0.14)
cg23681440	13	27498239	x[a]	-1,31	1,12E-08	2,92	-2,68	7,98E-07	5,04	6,47E-14	0.2499 (0.21-0.30)	0.2368 (0.20-0.28)	0.2537 (0.22-0.29)	0.2270 (0.19-0.27)
cg23126342	13	67801125	PCDH9	9,11	7,16E-14	4,38	7,56	2,73E-08	6,20	1,23E-20	0.5136 (0.43-0.60)	0.6047 (0.52-0.67)	0.4865 (0.39-0.57)	0.5621 (0.46-0.64)
cg25491122	13	67801551	PCDH9	2,36	1,49E-08	2,58	2,46	2,17E-05	3,58	1,54E-12	0.9053 (0.86-0.93)	0.9289 (0.90-0.95)	0.8605 (0.80-0.90)	0.8851 (0.84-0.91)
cg06885459	13	113689728	MCF2L	4,88	2,09E-11	2,70	3,01	4,91E-05	3,27	5,15E-15	0.3917 (0.34-0.44)	0.4406 (0.39-0.50)	0.3839 (0.34-0.43)	0.4140 (0.37-0.46)
cg17487894	13	114817820	RASA3	2,80	4,72E-10	3,65	2,63	2,49E-05	3,66	5,43E-14	0.4107 (0.36-0.45)	0.4387 (0.40-0.48)	0.3790 (0.33-0.42)	0.4053 (0.36-0.43)
cg01731783	14	74211788	C14orf43	-2,05	3,48E-11	3,85	-1,64	3,76E-05	3,75	6,39E-15	0.6572 (0.63-0.69)	0.6366 (0.60-0.67)	0.6395 (0.62-0.66)	0.6231 (0.60-0.65)
cg22851561	14	74214183	C14orf43	-6,06	5,47E-21	7,58	-3,60	4,60E-09	7,03	1,78E-28	0.4718 (0.41-0.52)	0.4112 (0.36-0.48)	0.4375 (0.40-0.49)	0.4015 (0.35-0.46)
cg24996979	14	74223355	C14orf43	-0,56	6,84E-25	9,76	-0,48	2,27E-08	6,37	1,96E-31	0.0802 (0.07-0.09)	0.0746 (0.07-0.08)	0.0813 (0.08-0.09)	0.0765 (0.07-0.08)
cg10919522	14	74227441	C14orf43	-0,89	6,20E-17	5,30	-0,61	2,46E-05	3,46	1,60E-20	0.1323 (0.12-0.15)	0.1234 (0.11-0.14)	0.1256 (0.11-0.14)	0.1195 (0.11-0.13)
cg13976502	14	74227875	C14orf43	-1,01	4,67E-16	5,64	-0,96	3,86E-05	3,40	1,73E-19	0.2701 (0.25-0.29)	0.2600 (0.24-0.28)	0.3003 (0.28-0.33)	0.2906 (0.27-0.31)
cg13038618	14	77467391	x[a]	-4,49	1,86E-10	3,59	-5,34	3,25E-08	5,87	5,41E-17	0.4082 (0.35-0.46)	0.3633 (0.29-0.42)	0.4037 (0.36-0.45)	0.3503 (0.29-0.41)
cg05875421	14	91709951	GPR68	-0,52	2,15E-14	4,60	-0,44	3,63E-07	5,32	4,32E-20	0.0755 (0.07-0.08)	0.0703 (0.06-0.08)	0.0665 (0.06-0.07)	0.0621 (0.06-0.07)
cg05284742	14	93552128	ITPK1	-3,76	3,57E-09	3,02	-4,32	1,48E-10	7,77	2,18E-17	0.7424 (0.71-0.78)	0.7048 (0.66-0.74)	0.6985 (0.67-0.73)	0.6553 (0.63-0.69)
cg06819357	14	102928437	TECPR2	5,70	4,81E-11	2,93	2,28	1,32E-05	3,77	3,08E-15	0.6637 (0.62-0.72)	0.7207 (0.67-0.77)	0.5934 (0.55-0.62)	0.6162 (0.58-0.65)
cg26242531	14	104190678	ZFYVE21	4,02	1,37E-15	5,04	3,28	2,01E-07	5,53	1,59E-21	0.2288 (0.19-0.27)	0.2690 (0.24-0.31)	0.2211 (0.19-0.26)	0.2539 (0.22-0.29)
cg11730703	14	105167607	INF2	2,06	2,04E-10	2,76	1,23	8,21E-06	4,06	8,12E-15	0.7804 (0.75-0.81)	0.8010 (0.77-0.83)	0.8144 (0.78-0.84)	0.8266 (0.80-0.86)

CpG	Chr.	Mapinfo	Gene	**a)** F4 Discovery Panel Median ß-value in %	p-value	Explained variance in %	**b)** F3 Replication Panel Median ß-value in %	p-value	Explained variance in %	**c)** Meta p-value	Methylation ß-value as median (first quartile-third quartile) **a)** F4 Discovery Panel Never smokers	Current smokers	**b)** F3 Replication Panel Never smokers	Current smokers
cg01208318	14	106329652	x[a]	-7,74	2,54E-12	4,15	-6,20	1,20E-05	3,59	1,57E-16	0.2836 (0.20-0.37)	0.2063 (0.16-0.28)	0.2246 (0.15-0.30)	0.1626 (0.12-0.23)
cg15022400	15	45028161	TRIM69	-0,73	3,36E-13	3,94	-0,70	1,22E-06	4,52	2,17E-18	0.0804 (0.07-0.09)	0.0731 (0.06-0.08)	0.0842 (0.08-0.09)	0.0773 (0.07-0.09)
cg03489965	15	65368982	LOC390594	2,66	9,52E-15	4,81	2,46	4,23E-05	3,43	3,10E-18	0.2217 (0.19-0.25)	0.2483 (0.22-0.30)	0.2349 (0.21-0.28)	0.2595 (0.22-0.33)
cg18335991	15	74724562	SEMA7A	-3,94	4,90E-13	4,40	-3,00	4,53E-07	5,19	1,21E-18	0.6581 (0.60-0.71)	0.6187 (0.55-0.67)	0.5936 (0.55-0.64)	0.5636 (0.52-0.61)
cg00310412	15	74724918	SEMA7A	-1,97	2,34E-13	5,17	-2,62	6,10E-11	8,26	2,09E-22	0.2994 (0.27-0.32)	0.2797 (0.25-0.31)	0.2711 (0.25-0.30)	0.2449 (0.23-0.27)
cg11152412	15	74927688	EDC3	-0,30	1,37E-11	3,59	-0,34	6,16E-06	4,14	4,19E-16	0.0522 (0.05-0.06)	0.0492 (0.04-0.05)	0.0633 (0.06-0.07)	0.0599 (0.05-0.07)
cg23161492	15	90357202	ANPEP	-1,60	1,12E-24	8,24	-1,52	2,15E-12	9,18	1,80E-35	0.1212 (0.11-0.13)	0.1053 (0.10-0.12)	0.1142 (0.10-0.13)	0.0989 (0.09-0.11)
cg05194346	15	101305443	x[a]	9,98	3,36E-22	7,02	4,89	1,43E-05	3,77	1,42E-25	0.3928 (0.32-0.48)	0.4926 (0.41-0.57)	0.4374 (0.36-0.50)	0.4863 (0.43-0.54)
cg01207684	16	4103167	ADCY9	-1,83	6,29E-10	2,77	-3,42	1,76E-07	5,08	8,30E-16	0.7670 (0.73-0.80)	0.7487 (0.70-0.78)	0.7476 (0.72-0.78)	0.7134 (0.68-0.75)
cg09099830	16	30485485	ITGAL	-1,72	1,55E-24	7,63	-2,57	3,23E-07	5,15	9,50E-30	0.5947 (0.57-0.62)	0.5775 (0.55-0.60)	0.6128 (0.59-0.64)	0.5871 (0.56-0.62)
cg03155159	16	59627519	x[a]	-1,19	9,31E-08	2,65	-2,72	3,91E-05	3,28	1,72E-11	0.3190 (0.24-0.39)	0.3071 (0.23-0.36)	0.3001 (0.26-0.35)	0.2729 (0.22-0.32)
cg00911794	17	1962132	HIC1	4,60	6,84E-10	2,05	1,99	7,89E-06	3,39	2,65E-14	0.2870 (0.25-0.35)	0.3330 (0.27-0.40)	0.2034 (0.17-0.25)	0.2233 (0.18-0.28)
cg23621097	17	1962236	HIC1	3,47	4,23E-08	1,90	1,83	2,40E-05	3,12	4,94E-12	0.2768 (0.23-0.33)	0.3115 (0.25-0.38)	0.1906 (0.16-0.24)	0.2090 (0.17-0.27)
cg09858022	17	38465333	RARA	1,20	5,46E-08	2,69	1,04	2,06E-05	3,80	5,61E-12	0.3923 (0.36-0.42)	0.4043 (0.38-0.43)	0.3709 (0.35-0.39)	0.3813 (0.36-0.40)
cg19572487	17	38476024	RARA	-10,02	9,37E-40	14,02	-7,81	7,56E-15	11,92	9,85E-53	0.4477 (0.40-0.51)	0.3475 (0.30-0.43)	0.4502 (0.40-0.51)	0.3722 (0.31-0.44)
cg04956244	17	38511592	RARA	2,50	3,44E-10	3,64	1,59	4,29E-06	4,40	7,50E-15	0.5979 (0.57-0.63)	0.6229 (0.59-0.65)	0.5706 (0.55-0.59)	0.5865 (0.57-0.61)
cg16255816	17	39889811	HAP1	-0,57	9,19E-11	3,63	-0,90	2,67E-05	3,59	1,16E-14	0.1406 (0.13-0.15)	0.1348 (0.12-0.14)	0.1392 (0.13-0.15)	0.1302 (0.12-0.14)
cg03373393	17	39890193	HAP1	-0,34	1,38E-11	3,82	-0,40	2,94E-05	3,37	2,03E-15	0.0849 (0.08-0.09)	0.0814 (0.07-0.09)	0.0763 (0.07-0.08)	0.0723 (0.07-0.08)
cg25809905	17	42467728	ITGA2B	6,53	2,74E-12	3,58	3,87	1,64E-06	4,64	2,34E-17	0.5129 (0.44-0.58)	0.5782 (0.52-0.64)	0.5170 (0.47-0.57)	0.5557 (0.51-0.61)
cg21280392	17	47304116	PHOSPHO1	2,37	1,56E-08	2,74	1,14	3,35E-06	4,03	3,11E-13	0.2667 (0.24-0.29)	0.2903 (0.27-0.32)	0.2527 (0.24-0.27)	0.2641 (0.25-0.29)
cg07465627	17	53167407	STXBP4	-1,38	3,67E-12	4,57	-1,46	6,45E-06	4,19	1,20E-16	0.1802 (0.16-0.21)	0.1665 (0.15-0.19)	0.1617 (0.15-0.18)	0.1471 (0.13-0.17)
cg02186444	17	73120977	ARMC7	3,82	4,67E-16	5,06	3,01	1,95E-08	6,79	5,49E-23	0.5410 (0.50-0.58)	0.5792 (0.53-0.61)	0.5200 (0.48-0.56)	0.5501 (0.52-0.58)
cg07251887	17	73641809	LOC100130933	-5,13	5,61E-09	3,52	-2,94	1,64E-07	5,30	8,26E-15	0.3329 (0.29-0.38)	0.2816 (0.24-0.32)	0.3216 (0.29-0.35)	0.2922 (0.25-0.32)
cg06459104	18	5456880	EPB41L3	-5,45	1,78E-10	3,68	-4,70	2,11E-07	5,33	2,54E-16	0.2918 (0.21-0.38)	0.2373 (0.17-0.32)	0.2842 (0.22-0.35)	0.2372 (0.17-0.31)
cg00073090	19	1265879	x[a]	-0,26	9,10E-14	3,32	-0,90	6,17E-08	4,84	3,34E-20	0.1727 (0.16-0.18)	0.1701 (0.16-0.18)	0.1440 (0.13-0.15)	0.1350 (0.13-0.14)
cg15187398	19	2093896	MOBKL2A	-2,75	8,51E-09	3,00	-1,77	6,68E-07	4,88	4,15E-14	0.2507 (0.22-0.28)	0.2232 (0.20-0.25)	0.2113 (0.19-0.23)	0.1937 (0.18-0.21)
cg07381806	19	2094327	MOBKL2A	-5,12	3,46E-17	7,12	-4,27	4,74E-06	4,73	1,37E-21	0.4132 (0.36-0.47)	0.3621 (0.30-0.42)	0.3898 (0.36-0.42)	0.3471 (0.30-0.40)
cg00835193	19	2291780	LINGO3	-8,23	2,44E-08	2,26	-11,66	4,40E-07	5,01	9,20E-14	0.8822 (0.75-0.93)	0.7999 (0.61-0.92)	0.8938 (0.77-0.94)	0.7772 (0.59-0.91)
cg03636183	19	17000585	F2RL3	-14,74	2,42E-80	22,45	-17,63	1,65E-39	26,94	5,58E-118	0.4930 (0.45-0.54)	0.3456 (0.28-0.43)	0.5152 (0.47-0.56)	0.3390 (0.28-0.43)
cg15159987	19	17003890	CPAMD8	-3,53	6,39E-15	4,62	-2,27	1,33E-06	4,77	4,90E-20	0.6339 (0.59-0.67)	0.5986 (0.56-0.64)	0.6090 (0.57-0.64)	0.5863 (0.55-0.61)
cg23973524	19	18873222	CRTC1	4,21	1,23E-25	6,59	2,16	1,06E-06	4,65	4,18E-30	0.3062 (0.28-0.34)	0.3483 (0.32-0.38)	0.2886 (0.26-0.32)	0.3102 (0.28-0.35)
cg11902728	19	35786580	MAG	5,76	4,05E-13	1,31	1,39	4,41E-05	1,79	1,08E-16	0.3524 (0.32-0.42)	0.4100 (0.35-0.46)	0.3680 (0.31-0.42)	0.3818 (0.33-0.43)
cg22649124	19	39282077	LGALS7B	-0,71	1,15E-09	2,97	-0,54	4,32E-05	3,37	2,25E-13	0.1161 (0.10-0.13)	0.1090 (0.10-0.12)	0.1012 (0.09-0.11)	0.0957 (0.09-0.11)
cg11701312	19	58897497	RPS5	-1,41	3,40E-10	3,71	-1,16	2,72E-05	3,66	4,30E-14	0.3002 (0.28-0.32)	0.2861 (0.27-0.31)	0.2918 (0.27-0.31)	0.2802 (0.26-0.29)
cg16201146	20	19191526	x[a]	-3,57	7,91E-14	4,32	-3,55	5,75E-07	4,64	2,48E-19	0.6390 (0.57-0.72)	0.6033 (0.55-0.68)	0.5935 (0.54-0.65)	0.5580 (0.49-0.61)
cg07339236	20	50312490	ATP9A	-1,10	1,57E-21	7,43	-0,83	5,04E-16	12,29	1,57E-35	0.0828 (0.07-0.09)	0.0717 (0.06-0.08)	0.0707 (0.07-0.08)	0.0624 (0.06-0.07)
cg00871610	21	37093012	MIR802	-4,33	5,46E-08	2,71	-3,89	1,13E-07	5,50	7,57E-14	0.5267 (0.47-0.57)	0.4834 (0.42-0.54)	0.4496 (0.41-0.48)	0.4107 (0.36-0.46)
cg06595162	21	40141040	NCRNA00114	-2,71	2,65E-08	2,72	-2,29	3,88E-06	3,85	6,17E-13	0.8358 (0.80-0.86)	0.8088 (0.77-0.84)	0.7731 (0.75-0.80)	0.7503 (0.72-0.78)
cg23110422	21	40182073	ETS2	-3,68	1,72E-16	5,91	-3,60	2,54E-07	5,29	2,66E-22	0.9005 (0.87-0.92)	0.8637 (0.82-0.90)	0.8547 (0.82-0.88)	0.8187 (0.78-0.86)
cg22635096	21	46550644	ADARB1	3,36	5,81E-11	3,33	2,64	7,60E-06	3,99	2,15E-15	0.2511 (0.20-0.30)	0.2847 (0.23-0.36)	0.2046 (0.18-0.25)	0.2310 (0.20-0.26)
cg02532700	22	37257404	NCF4	-1,15	1,88E-16	6,28	-0,94	1,07E-07	5,88	1,20E-22	0.0924 (0.08-0.11)	0.0809 (0.07-0.10)	0.1014 (0.09-0.12)	0.0920 (0.08-0.10)
cg01127300	22	38614796	x[a]	-3,83	3,08E-16	6,22	-3,87	7,04E-06	3,75	1,65E-20	0.4576 (0.40-0.51)	0.4193 (0.34-0.48)	0.4485 (0.39-0.51)	0.4098 (0.35-0.47)

Displayed are **a**) the results of the linear model calculated with M-value adjusted for age, sex, BMI, alcohol and white blood count (p-value and explained variance), as well as the median ß-value methylation difference between current and never smokers for the F4 discovery panel with genome-wide significance (p-value ≤ 1E-07) and **b)** the corresponding results of the same CpG sites for the F3 replication panel for comparison (significance level ≤ 5E-05); (c) the corresponding p-value gained by meta-analysis of F4 and F3, sorted by chromosome and mapinfo (Genome build 37)

[a] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites [b] According to UCSC Genome Browser no annotated transcripts are associated with these CpG sites, but SNPs within the same region (shore of a CpG Island) have a predicted function on the *ALPPL2* gene, which is located several kb apart from this CpG island

List of publications and presentations

Publications included in this thesis

Zeilinger, S., Kuhnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A., Strauch, K., Waldenberger, M. & Illig, T (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8: e63812.

Publications not included in this thesis

Zeilinger, S.*, Petersen, A.K.*, Kastenmuller, G., Romisch-Margl, W., Brugger, M., Peters, A., Meisinger, C., Strauch, K., Hengstenberg, C., Pagel, P., Huber, F., Mohney, R.P., Grallert, H., Illig, T., Adamski, J., Waldenberger, M., Gieger, C. & Suhre, K (2014) Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. Hum Mol Genet 23: 534-45. *joint first authors

Albrecht E*, Waldenberger M*, Krumsiek J, Evans AM, Breier M, Adamski J, Koenig W, **Zeilinger S**, Klopp N, Theis FJ, Wichmann HE, Suhre K, Illig T, Strauch K, Peters A, Gieger C, Kastenmüller G, Doering A and Meisinger C (2014) Metabolite Profiling Reveals New Insights into the Regulation of Serum Urate in Humans. Metabolomics 10: 141-151. *joint first authors

Schieck M, Michel S, Suttner K, Illig T, **Zeilinger S**, Franke A, Vogelberg C, v. Berg A, Bufe A, Heinzmann A, Laub O, Rietschel E, Simma B, Frischer T, Genuneit J, Kerzel S, Kabesch M (2013) Genetic variation in Th17 pathway genes, childhood asthma, and total serum IgE levels. J Allergy Clin Immunol.

Potaczek DP, Michel S, Sharma V, **Zeilinger S**, Vogelberg C, v. Berg A, Bufe A, Heinzmann A, Laub O, Rietschel E, Simma B, Frischer T, Genuneit J, Illig T, Kabesch M (2013) Different FCER1A polymorphisms influence IgE levels in asthmatics and non-asthmatics. Pediatr Allergy Immunol 24: 441-9.

Mezger M, Hector M, Eichler L, **Zeilinger S**, Kormann M, Riethmüller J, Stern M, Hartl D. The association of TLR5 polymorphisms with cystic fibrosis lung disease. Submitted.

Lattka E, Koletzko B, **Zeilinger S**, Hibbeln JR, Klopp N, Ring SM, Steer CD (2013) Umbilical cord PUFA are determined by maternal and child fatty acid desaturase (FADS) genetic variants in the Avon Longitudinal Study of Parents and Children (ALSPAC). Br J Nutr 109: 1196-1210.

Kormann MSD, Hector A, Marcos V, Mays LE, Kappler M, Illig T, Klopp N, **Zeilinger S**, Carevic M, Rieber N, Eickmeier O, Zielen S, Gaggar A, Moepps B, Griese M and Hartl D (2012) CXCR1 and CXCR2 haplotypes synergistically modulate cystic fibrosis lung disease. European Respiratory Journal 39: 1385-1390.

Koletzko B, Lattka E, **Zeilinger S**, Illig T, Steer C (2011) Genetic variants of the fatty acid desaturase gene cluster predict amounts of red blood cell docosahexaenoic and other polyunsaturated fatty acids in pregnant women: findings from the Avon Longitudinal Study of Parents and Children. Am J Clin Nutr 93: 211-219.

Poster presentations

Zeilinger S, Klopp N, Kühnel B, Gieger C, Illig T. Tobacco smoking leads to significant genome-wide DNA methylation changes: 450K methylation chip analysis. CSC SYMPOSIUM, Epigenetic Regulation: From Mechanism to Intervention, London, UK, June 20th – 22th 2012.

Zeilinger S, Brockhaus C, Klopp N, Gieger C, Illig T. Tobacco smoking leads to significant DNA methylation changes in several 100 loci: 450K methylation chip analysis. Statustreffen SFB/TR22, Lübeck, Germany, October 26th - 28th 2011.

Zeilinger S, Klopp N, Baurecht H, Rodriguez E, Illig T. The study of DNA methylation patterns in the forkhead transcription factor (FOXP3) gene in asthmatics and controls. Epigenomics of Common Diseases, Hinxton, Cambridge, UK, September 13th-16th 2011

Zeilinger S, Klopp N, Baurecht H, Rodriguez E, Illig T. The study of DNA methylation patterns in the forkhead transcription factor (FOXP3) gene in asthmatics and controls. European Human Genetics Conference, Gothenburg, Sweden, June 12th-15th 2010.

Oral presentations

'Genome-wide DNA methylation analysis with the example of smoking'; Workshop zur Genetischen Epidemiologie: 'New Avenues for Genetic and Molecular Epidemiology', Grainau, Germany, April 11th – 13th 2012.

'Epigenome wide association studies (EWAs) with the Illumina 450K methylation bead chip – Example tobacco smoking'; SFB/TR22 Retreat 'Angerer Alm 2012', St. Johann, Austria, March 5th – 7th 2012.

'Omics platforms of Z3a'; SFB/TR22 Meeting, Frankfurt, Germany, November 29th 2011.

'Epigenetic analysis in molecular epidemiology: Methylation analysis of the Foxp3 (forkhead transcription factor) gene in asthmatic and non-asthmatic control subjects from the KORA study – results and perspectives'; SFB/TR22 Retreat 'Angerer Alm 2011', St. Johann, Austria, February 2nd – 4th 2011.

'Functional Genomics Core Facility'; SFB/TR22 Meeting, Fürstenried, Germany, October 27th-29th 2010.

'Determination of methylation pattern – the user point of view'; Frühjahrsakademie der Gesellschaft für Genetik 2010: 'Genetic Epidemiology – Status quo, Where to go', Wittenberg, Germany, April 28th-30th 2010.

'Epigenetics and Asthma: Do prenatal and/or postnatal environmental exposures change epigenetic pattern leading to asthma later in life?'; SFB/TR22 Retreat 'Angerer Alm 2010', St. Johann, Austria, February 3rd – 5th 2010.

Danksagung (Acknowledgments)

Ich möchte mich gerne bei allen Menschen bedanken die mich während dieser Arbeit direkt oder indirekt unterstützt haben.

Mein besonderer Dank gilt vor allem Herrn Prof. Dr. Jerzy Adamski, der die Betreuung meiner Dissertation an der TU München übernahm, und seiner uneingeschränkten Unterstützung dieser Arbeit. Ebenso möchte ich mich herzlich bei Herrn Prof. Dr. Thomas Illig für das entgegengebrachte Vertrauen in meine Arbeit und für das immer offene Ohr bei allen Fragen bedanken. Die Art der Betreuung und Führung beider Mentoren hat mir viel Freiraum für eigene Ideen und Entwicklung gegeben, wofür ich sehr dankbar bin. Für die Übernahme der Rolle des zweiten Prüfers und somit freundlichen Unterstützung dieser Arbeit, möchte ich mich auch bei Herrn Prof. Dr. Hans-Rudolf Fries bedanken.

Frau Dr. Melanie Waldenberger und Frau Dr. Eva Reischl danke ich für die sorgfältige Durchsicht der Rohfassung der Arbeit, und im Besonderen für ihre wissenschaftliche und persönliche Unterstützung vor allem während der Endphase dieser Dissertation.

Mein Dank gilt ebenfalls Herrn Dr. Norman Klopp für inspirierende Diskussionsbereitschaft und seine stets hilfreichen Ratschläge.

Des Weiteren bedanke ich mich bei Frau Brigitte Kühnel, Frau Simone Wahl, Herrn Hansjörg Baurecht sowie Herrn Michael Brugger für ihre hilfsbereite Unterstützung in allen statistischen Angelegenheiten.

Für ihre freundliche und kollegiale Unterstützung bedanke ich mich bei allen ehemaligen und aktuellen Kollegen meiner Arbeitsgruppe. Insbesondere Frau Dr. Elke Rodríguez und Frau Jennifer Kriebel für diverse wissenschaftliche Diskussionen, Frau Nadine Lindemann, Frau Franziska Scharl und Frau Nicole Spada für ihre wertvolle Unterstützung im Labor, sowie Frau Anja Kleinschmidt für die Durchführung der EMSAs.

Ganz herzlich danke ich auch meinen Eltern und Brüdern sowie meinen Schwiegereltern in Spe für ihre bedingungslose Unterstützung und ihren Glauben an mich.

Lieber Andi, ein besonderer Dank gilt dir. Für das richtige Gespür, mich in die Arme zu nehmen oder anzutreiben. Für deine tatkräftige und seelische Unterstützung, deine Liebe und Geduld. Danke, dass es dich gibt…

Curriculum vitae

Persönliche Daten:

Name:	Sonja Zeilinger

Werdegang:

seit 09/2009	Doktorandin an der LMU-IBE (Gastvertrag am HMGU)
04/2009 – 08/2009	Wissenschaftliche Mitarbeiterin bei der Fa. Schwertner (Gastvertrag am HMGU)
04/2008 – 08/2009	Masterstudium Genetik/ Biotechnologie an der Naturwissenschaftlichen Fakultät der Universität Salzburg (AUT) Schwerpunkt: Immunologie Abschluss: Master rer.nat. in Genetik/ Biotechnologie
03/2006 – 03/2009	Wissenschaftliche Mitarbeiterin der LMU München - Dr. von Haunersches Kinderspital, KUBUS Forschungszentrum (Allergogenetik), München
05/2005 – 02/2006	Praktikum bei Bavaria Film GmbH, München
10/2001 – 04/2005	Bachelorstudium der Biologie an der Naturwissenschaftlichen Fakultät der Universität Salzburg (AUT) Fachrichtung: Genetik und Molekularbiologie Abschluss: Bachelor in Genetik und Molekularbiologie
05/2001 – 09/2001	Angestellte bei CIMA – Stadtmarketing GmbH, Ried in Innkreis (AUT)
03/2000 – 04/2001	Au Pair in Tiburon, San Francisco (USA)
1999	Abitur und Diplom für Kindergartenpädagogik