OMICS A Journal of Integrative Biology
Volume 27, Number 5, 2023
Mary Ann Liebert, Inc.
DOI: 10.1089/omi.2023.0028

Gene Set Based Integrated Methylome and Transcriptome Analysis Reveals Potential Molecular Mechanisms Linking Cigarette Smoking and Related Diseases

Pashupati P. Mishra,[1–3] Binisha H. Mishra,[1–3] Emma Raitoharju,[4,5] Nina Mononen,[1–3] Jorma Viikari,[6,7] Markus Juonala,[6,7] Nina Hutri-Kähönen,[8] Mika Kähönen,[2,9] Olli T. Raitakari,[10–12] and Terho Lehtimäki[1–3]

Abstract

Advanced integrative analysis of DNA methylation and transcriptomics data may provide deeper insights into smoke-induced epigenetic alterations, their effects on gene expression and related biological processes, linking cigarette smoking and related diseases. We hypothesize that accumulation of DNA methylation changes in CpG sites across genomic locations of different genes might have biological significance. We tested the hypothesis by performing gene set based integrative analysis of blood DNA methylation and transcriptomics data to identify potential transcriptomic consequences of smoking via changes in DNA methylation in the Young Finns Study (YFS) participants ($n = 1114$, aged 34–49 years, women: 54%, men: 46%). First, we performed epigenome-wide association study (EWAS) of smoking. We then defined sets of genes based on DNA methylation status within their genomic regions, for example, sets of genes containing hyper- or hypomethylated CpG sites in their body or promoter regions. Gene set analysis was performed using transcriptomics data from the same participants. Two sets of genes, one containing 49 genes with hypomethylated CpG sites in their body region and the other containing 33 genes with hypomethylated CpG sites in their promoter region, were differentially expressed among the smokers. Genes in the two gene sets are involved in bone formation, metal ion transport, cell death, peptidyl-serine phosphorylation, and cerebral cortex development process, revealing epigenetic–transcriptomic pathways to smoking-related diseases such as osteoporosis, atherosclerosis, and cognitive impairment. These findings contribute to a deeper understanding of the pathophysiology of smoking-related diseases and may provide potential therapeutic targets.

Keywords: cigarette smoking, gene set analysis, DNA methylation, transcriptomics, public health, epigenomics

[1] Department of Clinical Chemistry, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
[2] Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
[3] Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland.
[4] Molecular Epidemiology, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
[5] Tampere University Hospital, Tampere, Finland.
[6] Department of Medicine, University of Turku, Turku, Finland.
[7] Division of Medicine, Turku University Hospital, Turku, Finland.
[8] Department of Paediatrics, Tampere University Hospital, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
[9] Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland.
[10] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
[11] Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
[12] Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland.

Introduction

Cigarette smoking, referred to as smoking hereafter, is the leading preventable risk factor for numerous diseases and affects almost every organ in the body such as heart, blood vessels, lungs, brain, liver, mouth, and bones (Bergen, 1999). Smoking induces diseases through pathogenic mechanisms such as inflammation, oxidative damage, endothelial dysfunction, and alterations in the immune system and epigenetics (Zhou et al., 2016).

However, the molecular cascade underlying these mechanisms is not well understood. Availability of high-throughput multi-omics data and advanced statistical techniques for integrative analysis makes it possible to investigate the interrelationships of the involved molecules and their functions in a biological system across multiple molecular layers.

Understanding the underlying molecular mechanism of smoking-related disease induction is crucial for the development of interventions, treatments, and pharmacological agents aimed at reducing smoking-related health burden. For instance, knowledge of the epigenetic and consequent transcriptomic effects of smoking may aid in the development of drugs targeting specific epigenetic mechanisms to prevent consequences at transcriptomic level.

Such an approach is already in use for cancer treatment (Tzika et al., 2018). The drugs, when used along with the existing evidence-based therapies for treating smoking (Fiore et al., 2000), may play a crucial role in smoking-related health risks management.

Several studies have shown smoking-related alterations in DNA methylation (Joehanes et al., 2016; Kaur et al., 2019; Mishra et al., 2020; Zeilinger et al., 2013). Smoking can affect DNA methylation via several mechanisms. For example, carcinogenic contents of cigarette smoke such as arsenic, chromium, formaldehyde, polycyclic aromatic hydrocarbons, and nitrosamines can damage DNA. The DNA damage leads to recruitment of DNA methyltransferase 1 (DNMT1) at the repair site, consequently affecting DNA methylation (Lee and Pausova, 2013). Also, nicotine in cigarette smoke can affect DNA methylation by directly down-regulating DNMT1 expression.

Cigarette smoke may also alter DNA methylation indirectly through the modulation of expression and activity of DNA-binding factors. The largest epigenome-wide association study (EWAS) of smoking identified 18,760 active smoking-related CpG sites annotated to 7201 genes with false discovery rate (FDR) <0.05 (Joehanes et al., 2016).

This study was based on the Infinium HumanMethylation 450 K BeadChip that measures methylation at about 450,000 CpG sites throughout the genome instead of the >850,000 CpG sites measured by the newer Illumina HumanMethylationEPIC BeadChip (EPIC). A recent EWAS of smoking habit using EPIC chip identified 952 CpG sites in 500 genes differentially methylated between current and never smokers at a genome-wide significance threshold ($p = 6.25 \times 10^{-8}$) (Christiansen et al., 2021).

The impact of smoking on DNA methylation (Joehanes et al., 2016; Kaur et al., 2019; Mishra et al., 2020; Zeilinger et al., 2013) and genome-wide expression (Charlesworth et al., 2010; Huan et al., 2016; Vink et al., 2017) have previously been studied mainly independently. Integrative studies on consequences of the altered DNA methylation on genome-wide expression on the same set of individuals in blood have been based on 450 K chip (Maas et al., 2020; Tsai et al., 2018).

A few multi-omic integrative studies on smoking are based on EPIC chip-based DNA methylation data; however, they are based on specific cell types such as bronchoalveolar lavage cells (Ringh et al., 2019) and monocytes (Wan et al., 2018) but not on whole blood. While the earlier mentioned integrative multi-omic studies on smoking have crucial contributions to understanding of smoking-related disease mechanisms, integrative analysis of DNA methylome and transcriptome from whole blood using EPIC array is an important scientific gap that needs to be addressed.

Further, the previous integrative analyses were based on traditional one-to-one association analysis between genes and methylation sites, which is likely to suffer from lack of sufficient statistical power.

Small changes in individual gene expression levels associated with smoking-related methylation pattern in their genomic regions can be missed by one-to-one association analysis of methylation sites and corresponding genes due to lack of statistical power. Combining genes with smoking specific CpG sites and analyzing them as a set increases the probability of finding transcriptomic consequence of smoking via epigenetic route due to increased statistical power.

Small but coordinated changes in gene expression due to smoking-related DNA methylation changes in their genomic regions can have major biological effects even if the changes are not statistically significant for individual genes. Gene set analysis can capture such results as we have shown in our previous study using 450 K chip-based methylation data analysis (Mishra et al., 2020).

In the present study, we performed advanced gene set based integrative analysis of DNA methylome with 850 K methylation sites and transcriptome from whole blood collected from a cohort of European ancestry to test whether sets of genes containing smoking specific CpG sites within their genomic region (cis-regulation) are differently expressed among smokers, as compared with non-smokers (Fig. 1). In addition to the gene set based integrative analysis, we also conducted individual omics data analysis for the DNA methylation and transcriptome data to identify smoking-related CpG sites and genes respectively (Fig. 1).

Materials and Methods

Cohort description

This study was based on the Young Finns Study (YFS), one of the largest existing prospective multicenter follow-up studies assessing cardiovascular risk factors from childhood to adulthood (Raitakari et al., 2008). The study began in 1980 with 3596 children and adolescents aged 3–18 years randomly selected from 5 university hospital areas in Finland (Turku, Tampere, Helsinki, Kuopio, and Oulu) and they have been followed in regular intervals for over 40 years until 2020.

The study was approved by the Ethics Committee of the Hospital District of Southwest Finland on June 20, 2017 (ETMK: 68/1801/2017). All participants gave their written informed consent, and the studies were conducted in

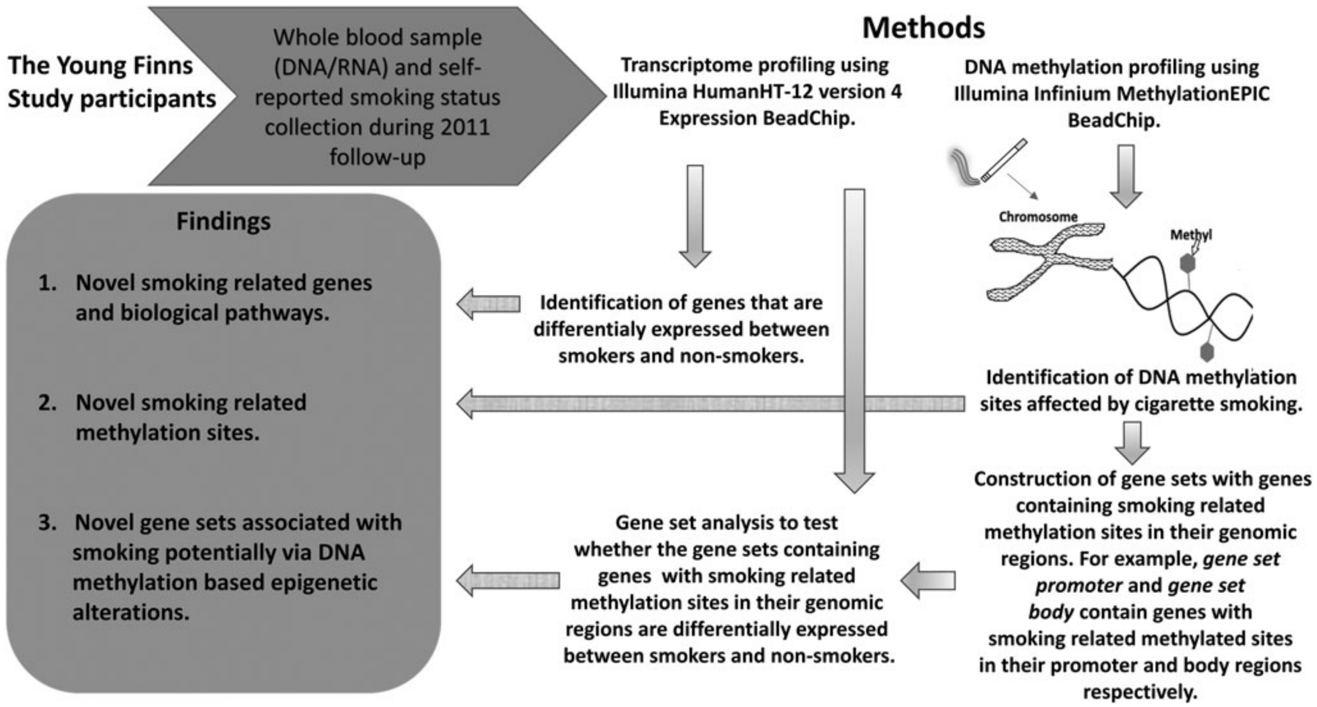
The Young Finns Study participants
Whole blood sample (DNA/RNA) and self-reported smoking status collection during 2011 follow-up
Methods
Transcriptome profiling using Illumina HumanHT-12 version 4 Expression BeadChip.
DNA methylation profiling using Illumina Infinium MethylationEPIC BeadChip.
Chromosome
Methyl
Findings
1. Novel smoking related genes and biological pathways.
2. Novel smoking related methylation sites.
3. Novel gene sets associated with smoking potentially via DNA methylation based epigenetic alterations.
Identification of genes that are differentialy expressed between smokers and non-smokers.
Identification of DNA methylation sites affected by cigarette smoking.
Construction of gene sets with genes containing smoking related methylation sites in their genomic regions. For example, gene set promoter and gene set body contain genes with smoking related methylated sites in their promoter and body regions respectively.
Gene set analysis to test whether the gene sets containing genes with smoking related methylation sites in their genomic regions are differentially expressed between smokers and non-smokers.

FIG. 1. Schematic representation of the overall study design.

accordance with the Declaration of Helsinki. Data protection will be handled according to current regulations as noted next.

This study involved a subpopulation of 1114 participants from the YFS with DNA methylation, transcriptomics, smoking habit, alcohol consumption, and occupation-based socio-economic status data available. Variable for smoking habit was based on self-reported information on whether the participants are daily smokers or never or less than daily smokers. Variable for the participants' alcohol consumption information was generated from their self-reports on their alcohol consumption during the previous week where 1 unit is equivalent to 14 g of alcohol (Juonala et al., 2009).

DNA methylation profiling, pre-processing, and normalization

DNA was obtained from EDTA-blood samples collected during the YFS 2011 follow-up using a Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. DNA integrity was tested by analyzing a subset of the samples with Agilent's Fragment Analyzer. Genome-wide DNA methylation levels were obtained using Illumina Infinium MethylationEPIC BeadChips, following the protocol by Illumina (Marttila et al., 2021). All pre-processing steps were performed using functions implemented in the *minfi* R/Bioconductor package (Fortin et al., 2017).

All analyzed samples have a sum of detection *p*-values <0.01 across all the probes. Only the samples with logged (log2) median of the methylated and unmethylated intensities clustering well based on default threshold of *plotQC* function in *minfi* R package were included. Samples for which the actual sex did not match the predicted sex were excluded.

Background subtraction and dye-bias normalization were performed via the noob method (Triche et al., 2013), followed by stratified quantile normalization using *preprocessQuantile* function in *minfi*. Probes with a detection *p*-value of more than 0.01 in more than 99% of the samples were filtered out.

CpG loci on sex chromosomes were excluded from the analysis to avoid gender-based methylation bias. Also, cross-reactive probes (McCartney et al., 2016; Pidsley et al., 2016) and probes with single nucleotide polymorphisms were excluded from the analysis. After quality control, the total number of autosomal CpGs was 769,683 in 1114 samples (150 active smokers and 964 never or less than daily-smokers).

Transcriptome profiling pre-processing and normalization

RNA isolation was performed from whole-blood samples collected from the YFS participants during the 2011 follow-up. Expression levels were analyzed with Illumina HumanHT-12 version 4 Expression BeadChip (Illumina, Inc.), containing 47,231 expression and 770 control probes. Raw Illumina summary probe-level data were exported from Beadstudio and processed in R (www.r-project.org) using a nonparametric background correction, followed by quantile normalization with control and expression probes, with the *neqc* function in the *limma* package (Smyth et al., 2005) and a log2 transformation. The pre-processing details are also described elsewhere (Mishra et al., 2021).

Biostatistical analysis

All statistical analyses were performed using R statistical software (v.4.1.0) (R Core Team, 2021).

Availability of data and materials

The dataset supporting the conclusions of this article was obtained from the Cardiovascular Risk in YFS, which comprises health-related participant data. The use of data is restricted under the regulations on professional secrecy (Act on the Openness of Government Activities, 612/1999) and on sensitive personal data (Personal Data Act, 523/1999, implementing the EU data protection directive 95/46/EC). Due to these restrictions, the data cannot be stored in public repositories or otherwise made publicly available.

Data access may be permitted on a case-by-case basis on request only. Data sharing outside the group is done in collaboration with the YFS group and requires a data-sharing agreement. Investigators can submit an expression of interest to the chairperson of the publication committee, Prof Mika Kähönen (Tampere University, Tampere, Finland) and Prof Terho Lehtimäki (Tampere University).

DNA methylation data analysis. Beta values, calculated as the ratio of intensities between methylated and unmethylated probe, were used as measures of methylation level. Differentially methylated positions (DMPs) for smoking status were identified using moderated *t*-test implemented in *limma* package in R (Smyth et al., 2005). The analysis was adjusted for age, sex, body mass index (BMI), alcohol consumption, socioeconomic status, and blood cell type proportions by adding them as covariates in the linear model implemented in *limma*.

Blood cell type proportions consisted of proportions of CD8T, CD4T, natural killer cells, B cells, monocytes, and granulocytes in white blood cells estimated through the reference-based Houseman method (Houseman et al., 2012) using the estimateCellCounts function in the minfi Bioconductor package in R (Aryee et al., 2014). Population structure, batch effects, and technical covariates were corrected for by including the first 30 principal components based on both control probes and methylation beta values each as covariates in the multiple linear regression model (Lehne et al., 2015).

Physical activity index, measured as weekly metabolic equivalent task hours (MET-h/week) calculated from information on the frequency, intensity, and duration of physical activity including leisure-time physical activity and commuting to the workplace (Pälve et al., 2018), was significantly associated with smoking habit with odds ratio (OR) of 0.88 and *p*-value of 2.8×10^{-09}. Alcohol consumption was also significantly associated with smoking habit but with less statistical significance ($\mathrm{OR}=1.13$, $p=1.9\times10^{-05}$) as compared with that of the physical activity index.

Therefore, as our main objective was to accurately estimate the individual effect of smoking habit on DNA methylation, we did not adjust the analysis with physical activity index to avoid the collinearity-related problem. Genomic control to reduce the number of false positive results was done by calculating genomic inflation factor (λ) as the ratio of the median of the empirically observed distribution of the test statistic to the expected median.

Novelty of the identified DMPs was assessed through a literature review, including a comparison of the results with the largest (Joehanes et al., 2016) and the latest (Christiansen et al., 2021) studies at FDR <0.05. Differentially methylated CpGs with genome-wide significance of $p<6.5\times10^{-8}$ (0.05/769,683) that are annotated to genes were used to define sets of genes.

A total of 16 gene sets were defined with genes containing hypo- or hypermethylated CpG sites in their: (1) transcription start site (TSS) 1500 region (200–1500 bases upstream of the TSS), (2) TSS200 region (0–200 bases upstream of the transcriptional start site), (3) 5′UTR region (within the 5′ untranslated region, between the TSS and the ATG start site), (4) first exon region (first exon of the gene), (5) body region (between the ATG and stop codon irrespective of the presence of introns, exons, TSS, or promoters), (6) 3′UTR region (between the stop codon and poly A signal), (7) exon boundary region, and (8) promoter region (TSS200, TSS1500, 5′UTR, and first exon regions combined).

Definitions of the gene regions were based on Human Methylation 850 K array and were obtained from *Illumina Human Methylation EPICanno. ilm10b2. hg19* R/Bioconductor package (Hansen et al., 2016). Differentially expressed genes (DEGs) among the smokers identified with transcriptome data analysis in this study (Section transcriptome data analysis) as well as in previous studies were removed from the defined gene sets because the aim of gene set analysis in this study was to test whether the method can identify transcriptomic consequences of smoking-related alterations in DNA methylation that is missed by traditional univariate methods.

Transcriptome data analysis. The DEGs between smokers and non-smokers were identified using moderated *t*-test implemented in *limma* R package. The analysis was adjusted for age, sex, BMI, alcohol usage, socioeconomic status, and the first 10 principal components of the transcriptomic data. Enrichment analysis of the statistically significant DEGs with FDR <0.05 with biological process terms of gene ontology (GO) database (Ashburner et al., 2000; Carbon et al., 2021) was done using *clusterProfiler* R package (Wu et al., 2021; Yu et al., 2012).

The enrichment analysis was done for up- and down-regulated genes separately. Summarization and interpretation of the biological process GO terms was done using REVIGO using default parameters (Supek et al., 2011).

Integrative DNA methylation and transcriptomic data analysis. Integrative analysis of DNA methylation and transcriptomics data concerning the effects of smoking habit was done by conducting self-contained gene set analysis of the differentially methylated gene sets using rotation gene set test (*ROAST*) (Wu et al., 2010) implemented in *limma* R/Bioconductor package (Fig. 2).

Results

Study participants

The characteristics of the YFS cohort participants of this study are shown in Table 1. Only 150 out of the total 1114 participants were daily smokers. The proportion of male participants was higher among daily smokers (53%) as compared with the non-smokers group (45%). Daily smokers were less involved in physical activity and had higher alcohol consumption as compared with never or less than daily smokers.

DMPs between active smokers and never smokers

In EWAS comparing active smokers ($N=150$) and never or less than daily smokers ($N=964$), we identified 272 statistically significant CpG sites or DMPs with methylome-wide significance ($p<6.5\times10^{-8}$) and 1206 DMPs with a more liberal threshold of FDR <0.05 (Fig. 3). The genomic inflation factor (λ) for these results was 1.08, which is lower

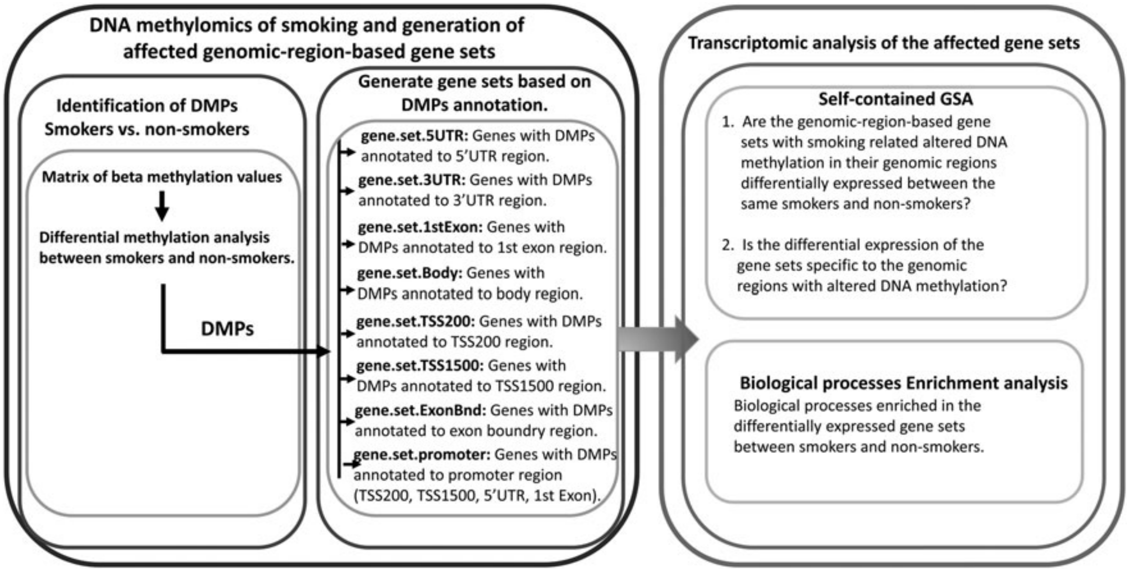
DNA methylomics of smoking and generation of affected genomic-region-based gene sets
Identification of DMPs Smokers vs. non-smokers
Matrix of beta methylation values
Differential methylation analysis between smokers and non-smokers.
DMPs
Generate gene sets based on DMPs annotation.
gene.set.5UTR: Genes with DMPs annotated to 5'UTR region.
gene.set.3UTR: Genes with DMPs annotated to 3'UTR region.
gene.set.1stExon: Genes with DMPs annotated to 1st exon region.
gene.set.Body: Genes with DMPs annotated to body region.
gene.set.TSS200: Genes with DMPs annotated to TSS200 region.
gene.set.TSS1500: Genes with DMPs annotated to TSS1500 region.
gene.set.ExonBnd: Genes with DMPs annotated to exon boundry region.
gene.set.promoter: Genes with DMPs annotated to promoter region (TSS200, TSS1500, 5'UTR, 1st Exon).
Transcriptomic analysis of the affected gene sets
Self-contained GSA
1. Are the genomic-region-based gene sets with smoking related altered DNA methylation in their genomic regions differentially expressed between the same smokers and non-smokers?
2. Is the differential expression of the gene sets specific to the genomic regions with altered DNA methylation?
Biological processes Enrichment analysis
Biological processes enriched in the differentially expressed gene sets between smokers and non-smokers.

FIG. 2. Schematic diagram representing the overall gene set based integrative multi-omics analysis for studying the impact of cigarette smoking on DNA methylation and its transcriptomic consequences. DMPs, differentially methylated positions; GSA, gene set analysis.

than that reported in other studies (Christiansen et al., 2021; Joehanes et al., 2016). The 272 statistically significant DMPs in our results constituted 36 novel smoking-related DMPs (Table 2) and replicated 236 DMPs from previous studies (Christiansen et al., 2021; Joehanes et al., 2016) (Supplementary Table S1). The number of DMPs replicated in our study based on FDR <0.05 was 663.

DEGs between active smokers and never smokers

Differential gene expression (DGE) analysis of transcriptomics data identified 371 genes upregulated and 312 genes downregulated (683 DEGs) among daily smoking participants with FDR <0.05 (Fig. 4) (Supplementary Table S2). While GO based gene set enrichment analysis of the 371 upregulated genes identified 105 biological processes with FDR <0.05 (Supplementary Table S3), only 87 biological processes were identified to be enriched in the 312 downregulated genes with FDR <0.05 (Supplementary Table S4).

Majority of the significantly enriched biological process GO terms in the upregulated genes were related to immune system, metabolism of carbohydrate, metabolism of fat, and signaling cascade. Similarly, GO terms enriched in the downregulated genes included biological processes such as regulation of cell killing, leukocyte-mediated cytotoxicity, lymphocyte proliferation, T cell activation, T cell receptor signaling pathway, immunological memory process, heart trabecula formation, immune response to tumor cell, pyroptosis, and interferon-gamma production.

Among the 683 smoking-related DEGs identified in this study, only 11 genes (6 up- and 5 downregulated) had altered methylation level in their genomic regions. While three of the

Table 1. Population Characteristics of the Young Finns Study Participants

Characteristics	*Daily smokers*	*Never or less than daily smokers*
Number of subjects	150	964
Sex (% women and % men)	47% and 53%	55% and 45%
Age, years	41 ± 5 (Range 34–49)	42 ± 5 (Range 34–49)
Body mass index, kg/m^2	26.1 ± 4.1	26.6 ± 5.1
Total cholesterol (mmol/L)	5.0 ± 0.9	5.0 ± 0.9
Low-density lipoprotein (LDL) cholesterol (mmol/L)	3.0 ± 0.8	3.1 ± 0.8
High-density lipoprotein (HDL) cholesterol (mmol/L)	1.3 ± 0.3	1.3 ± 0.3
Systolic blood pressure (mmHg)	120.4 ± 12.7	120.4 ± 14.4
Diastolic blood pressure (mmHg)	73.9 ± 10.7	75.8 ± 11.4
Alcohol consumption, units/day	1.4 ± 1.6	0.7 ± 1.1
Physical activity index (MET h/week)	12.3 ± 15.5	20.6 ± 22.2
Socioeconomic status	Low: 29% Medium: 43% High: 28%	Low: 18% Medium: 39% High: 43%

Data are mean ± SD or proportions.

MET-h, metabolic equivalent task hour; SD, standard deviation.

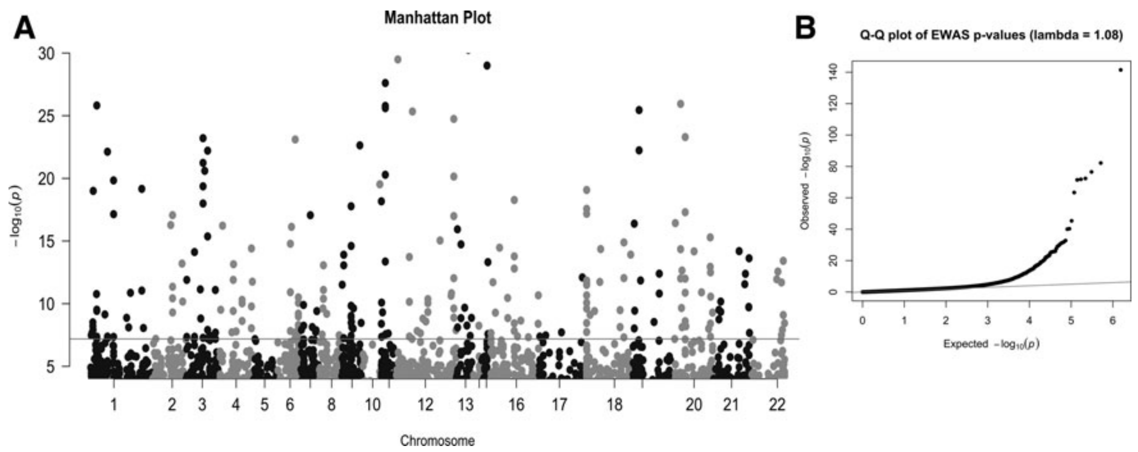

FIG. 3. (**A**) Manhattan plot showing the *p*-values of genome-wide CpG sites. *X*-axis represents the position of the CpG sites on each chromosome. *Y*-axis represents the negative log10 of the *p*-values for the association. The *solid horizontal line* represents the genome-wide significance threshold ($p = 6.5 \times 10^{-8}$). (**B**) Quantile-quantile plot showing genomic inflation factor ($\lambda = 1.08$) of the epigenome-wide association study. The genomic inflation factor (ratio of the median of the empirically observed distribution of the test statistic to the expected median) represents the extent of inflation and false positive rate in the results.

Table 2. List of 36 Novel Cigarette Smoking-Related CpG Sites Identified in This Study with Genome-Wide Significance of *p*-Value $<6.5 \times 10^{-8}$ (0.05/769,683), Their Corresponding Genomic Location, Corresponding Genes, Coefficients, Standard Error, and *p*-Values

ProbeID	*Chromosome*	*Position*	*Genes*	*Coefficients*	*SE*	p
cg12637027	chr5	56690874		–0.026	0.003	4.4×10^{-15}
cg14486033	chr2	54643636		–0.018	0.002	1.9×10^{-14}
cg19136686	chr16	17464401	*XYLT1*	–0.009	0.001	8.6×10^{-14}
cg00592949	chr9	112680911	*PALM2*; *PALM2-AKAP2*	–0.032	0.004	2.7×10^{-13}
cg12739216	chr12	131706350		–0.025	0.004	1.7×10^{-12}
cg15775568	chr2	54643284		–0.013	0.002	7.1×10^{-11}
cg24947681	chr15	39760933		–0.021	0.003	1.3×10^{-10}
cg16485845	chr8	141802466	*PTK2*	–0.014	0.002	1.8×10^{-10}
cg24087280	chr17	48193712	*SAMD14*	0.015	0.002	2.3×10^{-10}
cg15548246	chr6	13121401	*PHACTR1*	0.013	0.002	3.6×10^{-10}
cg02511321	chr5	32098574	*PDZD2*	0.015	0.002	3.9×10^{-10}
cg10682119	chr15	93182494	*FAM174B*	–0.034	0.005	3.9×10^{-10}
cg22311669	chr16	30466567		–0.016	0.003	6.4×10^{-10}
cg06563667	chr1	58016025	*DAB1*	–0.019	0.003	7.2×10^{-10}
cg04481318	chr9	134282053		–0.022	0.004	8×10^{-10}
cg26823705	chr1	145435523	*NBPF20*; *NBPF10*	–0.024	0.004	1.3×10^{-9}
cg15102575	chr3	124510809	*ITGB5*	–0.021	0.003	2.2×10^{-9}
cg11075883	chr7	146658441	*CNTNAP2*	0.013	0.002	3.9×10^{-9}
cg07815896	chr15	40385132	*BMF*	–0.017	0.003	5.7×10^{-9}
cg26894575	chr1	153518054	*S100A4*	–0.006	0.001	7.8×10^{-9}
cg13518852	chr1	212892006		0.026	0.005	8.7×10^{-9}
cg18734657	chr7	139420591	*HIPK2*	–0.023	0.004	1×10^{-8}
cg00442581	chr9	130733834	*FAM102A*	–0.022	0.004	1.2×10^{-8}
cg08151621	chr19	28995456	*LOC100420587*	–0.017	0.003	2.4×10^{-8}
cg01990910	chr16	12207648	*SNX29*	–0.016	0.003	2.9×10^{-8}
cg06467473	chr9	127054510	*NEK6*	–0.011	0.002	3.1×10^{-8}
cg19467605	chr11	94349883	*PIWIL4*	0.030	0.005	3.4×10^{-8}
cg07411532	chr20	56266785	*PMEPA1*	0.013	0.002	3.7×10^{-8}
cg09465516	chr2	54751985	*SPTBN1*	–0.021	0.004	4.2×10^{-8}
cg16762439	chr7	146795245	*LOC101928700*; *CNTNAP2*	0.013	0.002	4.3×10^{-8}
cg20430809	chr19	2089006	*MOB3A*	–0.014	0.003	4.4×10^{-8}
cg05702597	chr3	154829011	*MME*	–0.025	0.004	4.7×10^{-8}
cg11786988	chr17	8826387	*PIK3R5*	–0.012	0.002	5.6×10^{-8}
cg21811986	chr3	5053132		–0.012	0.002	6×10^{-8}
cg17340043	chr10	13202710	*MCM10*	0.021	0.004	6.4×10^{-8}
cg11674355	chr2	65610261	*SPRED2*	–0.024	0.004	6.4×10^{-8}

Full list of smoking-related CpG sites identified in this study can be found in Supplementary Table S1.
SE, standard error.

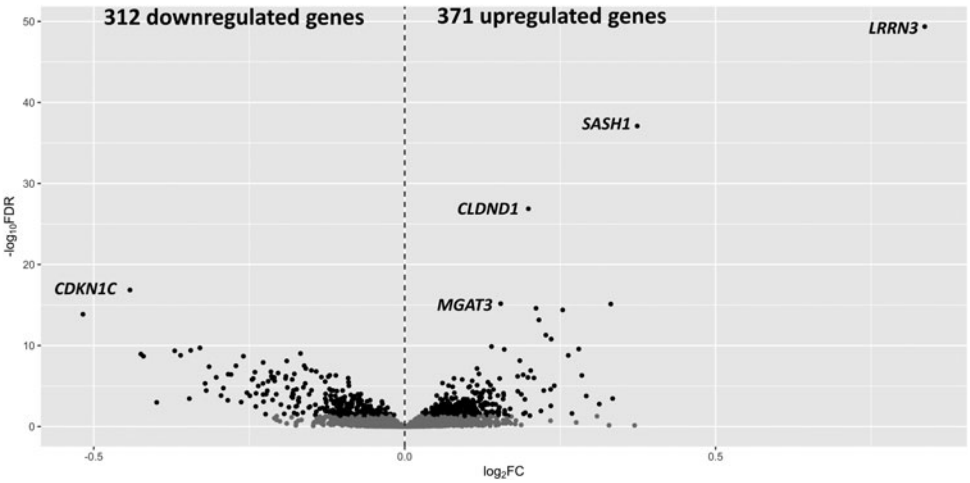

FIG. 4. Volcano plot of differentially expressed genes between active smokers and non-smokers. The dots on the right and left sides of the dashed vertical line represent 371 up and 312 downregulated genes, respectively, among active smokers with FDR <0.05. The *grey dots* represent statistically not significant genes with FDR >0.05. The five most significant differentially expressed genes are labeled. The *y*-axis represents negative log (base10) of FDR and the *x*-axis represents log (base 2) of the fold change between active smokers and non-smokers. FDR, false discovery rate.

six upregulated genes (leucine rich repeat neuronal 3 [*LRRN3*], *MGAT3* and G protein-coupled receptor 15 [*GPR15*]) had hypomethylated CpGs in their promoter region, the other three (claudin domain-containing protein 1 [*CLDND1*], *FAM102A*, and *EPHA4*) had hypomethylation in their body region.

Among the five downregulated genes, two (*PRSS23* and *SLAMF7*) had hypomethylation in their promoter region and the other three (*MTSS1*, *GFI1*, and *CCM2*) had hypomethylation in their body region.

Association between transcriptomic level gene expression and smoking-related alterations in DNA methylation

The gene set containing 49 genes with hypomethylated CpG sites in their body region was differentially expressed among smokers for both mixed hypothesis (the genes are up- or downregulated) and up hypothesis (the genes are upregulated) with FDR.mixed = 0.0005 and FDR.up = 0.03, respectively. The three most statistically significant upregulated genes (based on DGE analysis) in the gene set were *SIN3B* (FDR = 0.06), *BMF* (FDR = 0.08), and *PDE1C* (FDR = 0.19).

Difference in expression levels of all the 32 upregulated genes in this gene set is shown in Figure 5. Similarly, the three most DGE analysis-based downregulated genes in the gene set were *ITPK1* (FDR = 0.49), *CDH23* (FDR = 0.49), and *NBR1* (FDR = 0.59). Difference in expression levels of all the 17 downregulated genes in this gene set is shown in Figure 6. Overall, genes in the gene set were enriched in GO terms related to biological processes such as ossification, metal ion transport, response to oxygen levels, fat cell differentiation, cell death, and peptidyl-serine phosphorylation (Supplementary Table S5).

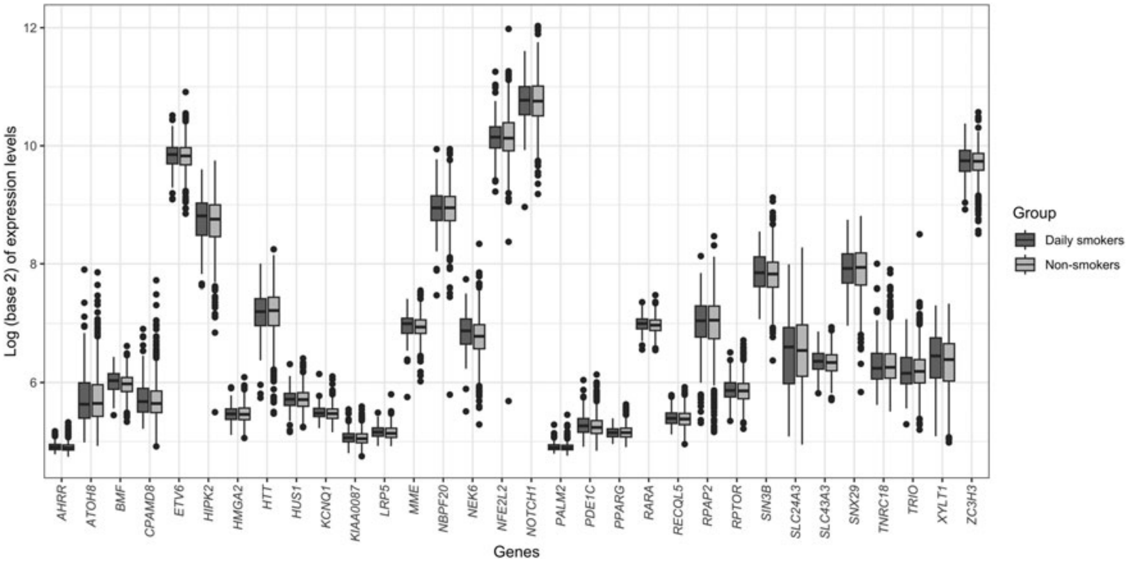

FIG. 5. Box plots for gene expression changes for the 32 upregulated genes with hypomethylated CpG sites in their body region between daily smokers and never or less than daily smokers.

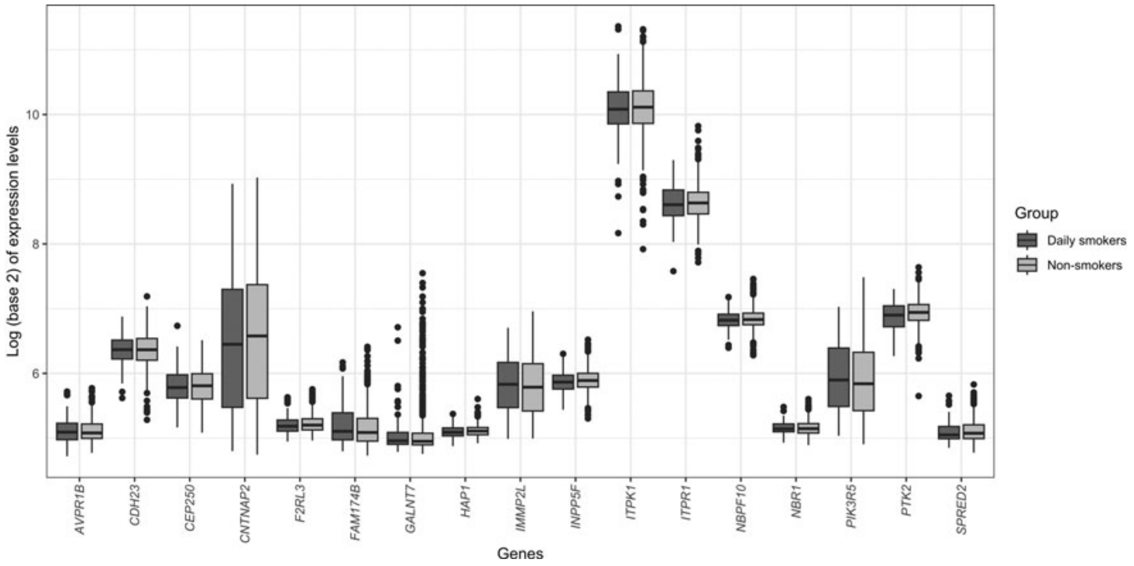

FIG. 6. Box plots for gene expression changes for the 17 downregulated genes with hypomethylated CpG sites in their body region between daily smokers and never or less than daily smokers.

The other gene set containing 33 genes with hypomethylated CpG sites in their promoter region was differentially expressed among smokers only for mixed hypothesis with FDR.mixed = 0.001. The three most DGE analysis-based statistically significant upregulated genes in the gene set were *SLC23A2* (FDR = 0.09), *ANPEP* (FDR = 0.16), and *NEC6* (FDR = 0.28). Difference in expression levels of all the 19 upregulated genes in this gene set is shown in Figure 7.

The three most downregulated genes in the gene set were *NCALD* (FDR = 0.07), *INPP4A* (FDR = 0.21), and *BACH2* (FDR = 0.25). Difference in expression levels of all the 14 downregulated genes in this gene set is shown in Figure 8. Member genes of this gene set were enriched in the biological process GO term related to the cerebral cortex development process.

None of the gene sets containing genes with hypermethylated genomic positions were differentially expressed among smokers with a statistical significance threshold of FDR <0.05.

Discussion

In the present study, we performed a novel advanced gene set based integrative analysis of DNA methylomic and transcriptomic data to identify epigenetic alterations associated with smoking and its transcriptomic consequences. The integrative system-level approach allowed us to identify sets of novel genes that have smoking-related alterations in DNA methylation within their genomic regions and are also differentially expressed among smokers.

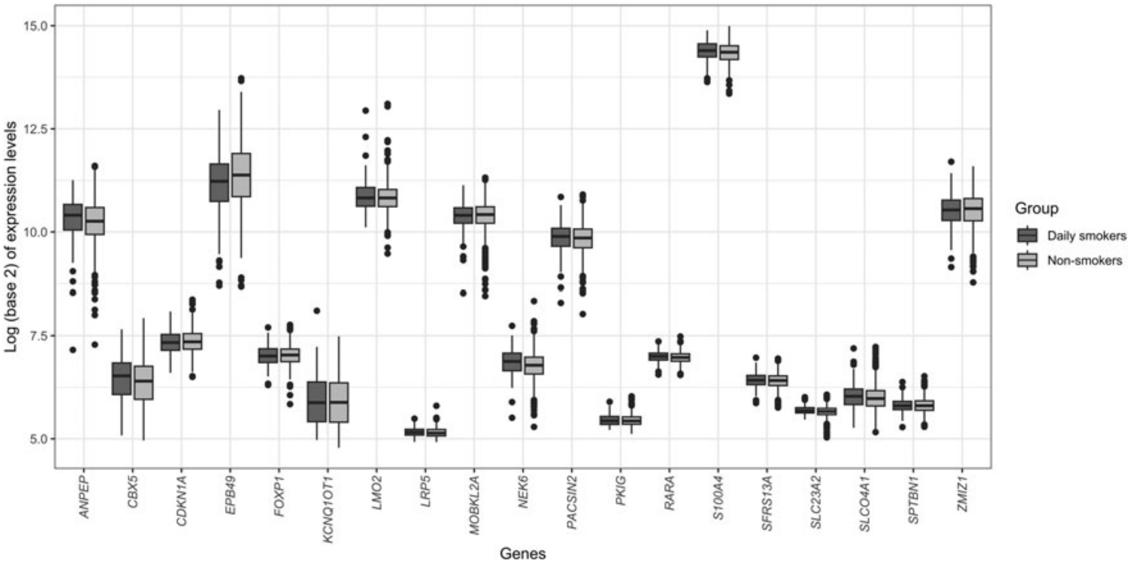

FIG. 7. Box plots for gene expression changes for the 19 upregulated genes with hypomethylated CpG sites in their promoter region between daily smokers and never or less than daily smokers.

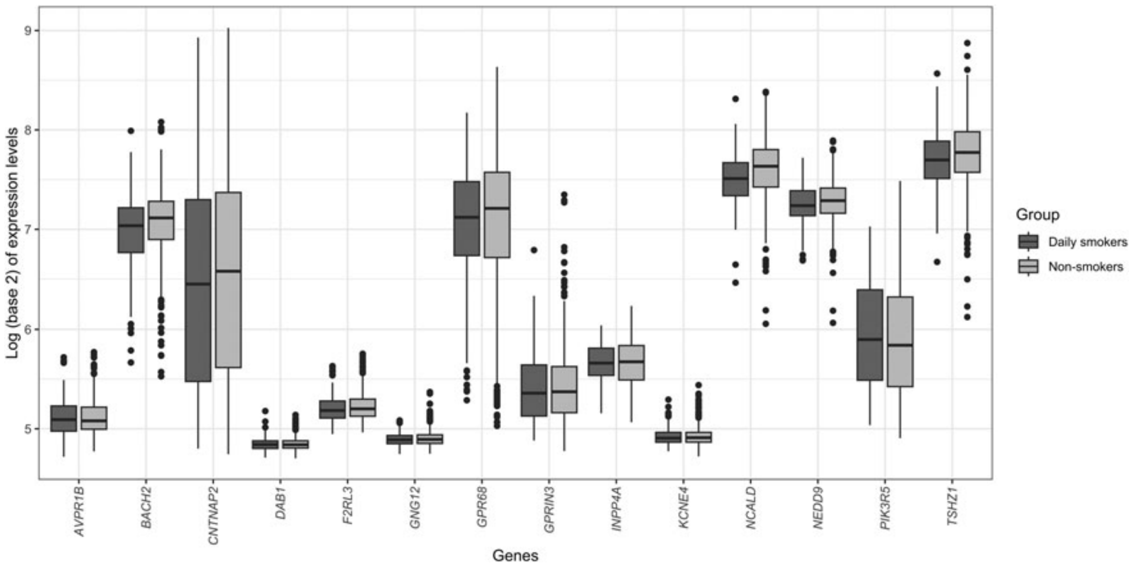

FIG. 8. Box plots for gene expression changes for the 14 downregulated genes with hypomethylated CpG sites in their promoter region between daily smokers and never or less than daily smokers.

The novelty of this study rests in the usage of both a novel data platform and a novel advanced system-level statistical method to study epigenetic and transcriptomic consequences of smoking. To our knowledge, this is the first multi-omics integrative study of smoking that is based on EPIC BeadChip-based DNA methylation and transcriptomic data from whole blood.

The number of measured CpG sites in EPIC BeadChip is nearly twice as many the number in HumanMethylation450 BeadChip (450 K), generating an important scientific gap that needs to be addressed. On the statistical method side, we propose a novel system-level integrative approach that pools and analyzes together all the genes based on their genomic regions where smoking-related DNA methylation alterations have occurred. Our results showed that this approach is statistically more powerful as compared with traditional one-to-one association analysis between genes and methylation sites such as the study by Maas et al. (2020).

Gene sets containing genes with hypomethylated CpG sites in body as well as in promoter regions were significantly differentially expressed among smokers. While the gene set containing genes with hypomethylated CpG sites in their body region was upregulated among smokers, the one with hypomethylated CpG sites in the promoter region contained both up- and downregulated genes among smokers. Interestingly, genes from the gene sets based on both promoter and body regions were not found to be differentially expressed using the traditional gene-wise differential gene expression analysis in this study.

Similarly, to the best of our knowledge, the effects of smoking-related differentially methylated genes on gene expression have not been identified in any of the previous transcriptome-wide studies despite the pronounced association of smoking with DNA methylation in the genomic regions of some of the member genes such as *AHRR* and *F2RL2* reported by several studies such as Joehanes et al. (2016), Kaur et al. (2019), and Zeilinger et al. (2013).

The novel genes within the gene sets identified in this study are enriched in biological processes that are known in literature to have associations with smoking; however, the underlying molecular mechanisms remain mostly unexplained. This study revealed the epigenetic mechanisms through which these biological processes are affected by smoking. For example, smoking is a well-known risk factor for osteoporosis (Yoon et al., 2012); this study identified the smoking-triggered epigenetic mechanism that affects genes involved in the process of bone formation.

Similarly, this study identified that genes involved in metal ion transport are affected by smoking-related alterations in DNA methylation. Imbalance in metal ion concentration in cerebrospinal fluid caused by smoking can lead to cognitive impairment (Li et al., 2021). Similarly, alterations in metal ion transport system can affect copper homeostasis, which, in turn, can lead to dyslipidemia and low-density lipoprotein oxidation (Meyer et al., 2014) and its related diseases such as osteoporosis (Poiana et al., 2013) and atherosclerosis (Linton et al., 2019).

Despite being one of the most important and preventable risk factors for atherosclerosis, the underlying molecular mechanism linking smoking and atherosclerosis remains largely unknown. Our findings support the hypothesis that the atherogenic effect of smoking might involve endothelial cell death (Messner et al., 2012). We identified that genes involved in cerebral cortex development are differentially expressed among smokers through the epigenetic mechanism. Our finding supports a previous study that reported an association between smoking and accelerated cortical thinning that causes cognitive decline in adults (Karama et al., 2015).

Univariate analysis of DNA methylation and transcriptomic data separately concerning smoking also generated novel results in this study. Biological implications of the novel DMPs along with the replicated ones from DNA methylation data analysis were studied with the downstream integrative approach as described earlier.

The largest transcriptomic study of smoking so far by Huan et al. (2016) reported 1270 DEGs between active smokers and never smokers at FDR <0.1, out of which 289 were replicated in this study at the same statistical significance threshold (FDR <0.1). Similarly, we also replicated 69 smoking-related genes reported by another study (Vink et al., 2017).

Considering the smoking-related genes reported in both the studies (Huan et al., 2016; Vink et al., 2017), there were 428 new genes at FDR <0.05 and 694 new genes at FDR <0.1 reported in this study. The most significant smoking-related genes identified in this study such as *LRRN3*, *CLDND1*, *PID1* (phosphotyrosine interaction domain-containing protein 1), *GPR15*, and *S1PR5* (sphingosine-1 phosphate receptor 5) were consistent in all other recent studies reviewed here.

Biological processes linked to the DEGs from transcriptomic data analysis such as immune system, cell death, fat metabolism, and signaling cascade are known to play a central role in the development of atherosclerosis (Hultén and Levin, 2009; Kong et al., 2022; Messner et al., 2012; Wolf and Ley, 2019).

This study has several limitations, one of them being the self-reported smoking status. However, this study replicated the most consistent findings across the literature from both DNA methylation and transcriptomics data indicating the robustness of the data and analysis approach. Another limitation is that the study was based on Infinium MethylationEPIC (EPIC) BeadChips for DNA methylation and Illumina HumanHT-12 version 4 Expression BeadChip for transcriptome, which provide suboptimal coverage of regulatory elements as compared with sequencing-based platforms such as whole-genome bisulfite sequencing and RNA-Seq.

The study was based on cross-sectional data, and, therefore, changes in DNA methylation after cessation of smoking could not be studied. Also, the participants of this study are of European origin and therefore further studies with populations of different ethnicities are needed. However, this study contributed a novel gene-set based integrative DNA methylome and transcriptome analysis approach. The approach is statistically more powerful as compared with the traditional single-molecule analysis methods and can allow identification of small but coordinated changes in gene expression associated with DNA methylation.

Conclusions

Using system-level integrated analysis of DNA methylation and transcriptomics data, we identified novel sets of genes associated with smoking through epigenetic alterations, uncovering the potential molecular cascade underlying the disease induction mechanism by smoking. The genes and their epigenetic regulation explain the underlining mechanism of how smoking affects different biological processes such as those related to the immune system, metal ion transport, osteoblast differentiation, hypoxia, cell death, and the cerebral cortex development process and can lead to related diseases such as atherosclerosis, osteoporosis, and cognitive impairment.

Importantly, this study proposes an alternative system-level integrative multi-omics analysis method that can uncover small but coordinated changes in gene expression potentially due to DNA methylation changes in their genomic regions. Such small and coordinated changes in gene expression can be missed by traditional linear association analysis methods due to lack of sufficient statistical power. The proposed integrative method can be applied to study a wide range of biological problems, for example, to uncover epigenetic alterations associated to a disease or a trait of interest, their effects on gene expression and related biological processes.

Authors' Contributions

P.P.M.: conceptualization, investigation, data analysis, writing—original draft; B.H.M.: data analysis, reviewed and edited the manuscript; E.R.: reviewed and edited the manuscript; N.M.: reviewed and edited the manuscript; J.V.: data acquisition, reviewed and edited the manuscript; M.J.: reviewed and edited the manuscript; N.H.-K.: data acquisition; M.K.: data acquisition, reviewed and edited the manuscript; O.T.R.: data acquisition, reviewed and edited the manuscript; T.L.: data acquisition, reviewed and edited the manuscript.

Author Disclosure Statement

The authors declare they have no conflicting financial interests.

Funding Information

The YFS has been financially supported by the Academy of Finland: grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), 41071 (Skidi), 330809, and 338395; the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere, and Turku University Hospitals (grant X51001, 9AC077, 9X047, 9S054, and 9AB059); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Sigrid Jusélius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Laboratoriolääketieteen Edistämissäätiö Sr; and Diabetes Research Foundation of Finnish Diabetes Association; This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 848146 (To Aition) and No. 755320 (TAXINOMISIS). This project has received funding from the European Research Council (ERC) advanced grants under grant agreement No. 742927 (MULTIEPIGEN project); Tampere University Hospital Supporting Foundation; and Finnish Society of Clinical Chemistry. P.P.M. was supported by the Academy of Finland (Grant No. 349708).

Supplementary Material

Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4
Supplementary Table S5

References

Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30(10):1363–1369; doi: 10.1093/bioinformatics/btu049

Ashburner M, Ball CA, Blake JA, et al. Gene ontology: Tool for the unification of biology. Nat Genet 2000;25(1):2529; doi: 10.1038/75556

Bergen AW, Caporaso N. Cigarette smoking. J Natl Cancer Inst 1999;91(16):1365–1375; doi: 10.1093/jnci/91.16.1365

Carbon S, Douglass E, Good BM, et al. The Gene Ontology resource: Enriching a GOld mine. Nucleic Acids Res 2021; 49(D1):D325–D334; doi: 10.1093/nar/gkaa1113

Charlesworth JC, Curran JE, Johnson MP, et al. Transcriptomic epidemiology of smoking: The effect of smoking on gene expression in lymphocytes. BMC Med Genomics 2010;3(1): 1–11; doi: 10.1186/1755-8794-3-29

Christiansen C, Castillo-Fernandez JE, Domingo-Relloso A, et al. Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. Clin Epigenetics 2021;13(1):1–13; doi: 10.1186/s13148-021-01018-4

Fiore MC. A clinical practice guideline for treating tobacco use and dependence: A US Public Health Service report. JAMA 2000;283(24):3244–3254; doi:10.1001/jama.283.24 .3244

Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics 2017;33(4):558–560; doi: 10.1093/ bioinformatics/btw691

Hansen KD. IlluminaHumanMethylationEPICanno. ilm10b2. hg19: Annotation for Illumina's EPIC Methylation Arrays. R Package Version 0.6.0; 2016.

Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13(1):1–16; doi: 10.1186/1471-2105-13-86

Huan T, Joehanes R, Schurmann C, et al. A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking. Hum Mol Genet 2016;25(21): 4611–4623; doi: 10.1093/hmg/ddw288

Hultén LM, Levin M. The role of hypoxia in atherosclerosis. Curr Opin Lipidol 2009;20(5):409–414; doi: 10.1097/MOL .0b013e3283307be8

Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. Circ Cardiovasc Genet 2016;9(5):436–447; doi: 10.1161/CIRCGENETICS.116.001506

Juonala M, Viikari JS, Kähönen M, et al. Alcohol consumption is directly associated with carotid intima-media thickness in Finnish young adults. The Cardiovascular Risk in Young Finns Study. Atherosclerosis 2009;204(2):e93–e98; doi: 10.1016/j.atherosclerosis.2008.11.021

Karama S, Ducharme S, Corley J, et al. Cigarette smoking and thinning of the brain's cortex. Mol Psychiatry 2015;20(6): 778–785; doi: 10.1038/mp.2014.187

Kaur G, Begum R, Thota S, et al. A systematic review of smoking-related epigenetic alterations. Arch Toxicol 2019; 93:2715–2740; doi: 10.1007/s00204-019-02562-y

Kong P, Cui ZY, Huang XF, et al. Inflammation and atherosclerosis: Signaling pathways and therapeutic intervention. Signal Transduct Target Ther 2022;7(1):131; doi: 10.1038/ s41392-022-00955-7

Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Front Genet 2013;4:132; doi: 10.3389/fgene.2013.00132

Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol 2015;16(1):1–12; doi: 10.1186/s13059-015-0600-x

Li H, Mu Q, Kang Y, et al. Association of cigarette smoking with male cognitive impairment and metal ions in cerebrospinal fluid. Front Psychiatry 2021;12:738358; doi: 10.3389/ fpsyt.2021.738358

Linton MF, Yancey PG, Davies SS, et al. The role of lipids and lipoproteins in atherosclerosis. Endotext [Internet] 2019.

Maas SC, Mens MM, Kühnel B, et al. Smoking-related changes in DNA methylation and gene expression are associated with cardio-metabolic traits. Clin Epigenetics 2020;12:1–16; doi: 10.1186/s13148-020-00951-0

Marttila S, Viiri LE, Mishra PP, et al. Methylation status of nc886 epiallele reflects periconceptional conditions and is associated with glucose metabolism through nc886 RNAs. Clin Epigenetics 2021;13:1–18; doi: 10.1186/s13148-021-01132-3

McCartney DL, Walker RM, Morris SW, et al. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genomics Data 2016;9:22–24; doi: 10.1016/j.gdata.2016.05.012

Messner B, Frotschnig S, Steinacher-Nigisch A, et al. Apoptosis and necrosis: Two different outcomes of cigarette smoke condensate-induced endothelial cell death. Cell Death Dis 2012;3(11):e424; doi: 10.1038/cddis.2012.162

Meyer JM, Ji A, Cai L, et al. Minimally oxidized LDL inhibits macrophage selective cholesteryl ester uptake and native LDL-induced foam cell formation [S]. Journal of Lipid Research 2014;55(8):1648–1656; doi: 10.1194/jlr.M044644

Mishra BH, Mishra PP, Raitoharju E, et al. Modular genome-wide gene expression architecture shared by early traits of osteoporosis and atherosclerosis in the Young Finns Study. Sci Rep 2021;11(1):1–13; doi: 10.1038/s41598-021-86536-0

Mishra PP, Hänninen I, Raitoharju E, et al. Epigenome-450K-wide methylation signatures of active cigarette smoking: The Young Finns Study. Biosci Rep 2020;40(7); doi: 10.1042/ BSR20200596

Pälve KS, Pahkala K, Magnussen CG, et al. Association of physical activity in childhood and early adulthood with carotid artery elasticity 21 years later: The cardiovascular risk in Young Finns Study. J Am Heart Assoc 2018;3(2):e000594; doi: 10.1161/JAHA.113.000594

Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 2016;17(1):1–17; doi: 10.1186/s13059-016-1066-1

Poiana C, Radoi V, Carsote M, et al. New clues that may link osteoporosis to the circulating lipid profile. Bone Research 2013;1(1):260–266; doi: 10.4248/BR201303004

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria; 2021.

Raitakari OT, Juonala M, Rönnemaa T, et al. Cohort profile: The cardiovascular risk in young Finns study. Int J Epidemiol 2008;37(6):1220–1226; doi: 10.1093/ije/dym225

Ringh MV, Hagemann-Jensen M, Needhamsen M, et al. Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells. EbioMedicine 2019;46:290–304; doi: 10.1016/j .ebiom.2019.07.006

Smyth GK. limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using

R and Bioconductor 2005;397–420; doi: 10.1007/0-387-29362-0_23

Supek F, Bošnjak M, Škunca N, et al. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 2011;6(7):e21800; doi: 10.1371/journal.pone.0021800

Triche TJ, Weisenberger DJ, Van Den Berg D, et al. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 2013;41(7):e90; doi: 10.1093/nar/gkt090

Tsai PC, Glastonbury CA, Eliot MN, et al. Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. Clin Epigenetics 2018;10(1):1–21; doi: 10.1186/s13148-018-0558-0

Tzika E, Dreker T, Imhof A. Epigenetics and metabolism in health and disease. Front Genet 2018;11:361; doi: 10.3389/fgene.2020.00428

Vink JM, Jansen R, Brooks A, et al. Differential gene expression patterns between smokers and non-smokers: Cause or consequence? Addict Biol 2017;22(2):550–560; doi: 10.1111/adb.12322

Wan M, Bennett BD, Pittman GS, et al. Identification of smoking-associated differentially methylated regions using reduced representation bisulfite sequencing and cell type-specific enhancer activation and gene expression. Environ Health Perspect 2018;126(4);047015; doi: 10.1289/EHP2395

Wolf D, Ley K. Immunity and inflammation in atherosclerosis. Circ Res 2019;124(2):315–327; doi: 10.1161/CIRCRESAHA.118.313591

Wu D, Lim E, Vaillant F, et al. ROAST: Rotation gene set tests for complex microarray experiments. Bioinformatics 2010; 26(17):2176–2182; doi: 10.1093/bioinformatics/btq401

Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation 2021;2(3):100141; doi: 10.1016/j.xinn.2021.100141

Yoon V, Maalouf NM, Sakhaee K. The effects of smoking on bone metabolism. Osteoporos Int 2012;23:2081–2092; doi: 10.1007/s00198-012-1940-y

Yu G, Wang LG, Han Y, et al. ClusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 2012;16(5):284–287; doi: 10.1089/omi.2011.0118

Zeilinger S, Kühnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 2013;8(5):e63812; doi: 10.1371/journal.pone.0063812

Zhou Z, Chen P, Peng H. Are healthy smokers really healthy? Tob Induc Dis 2016;14(1):1–12; doi: 10.1186/s12971-016-0101-z

Address correspondence to:
Pashupati P. Mishra, PhD
Department of Clinical Chemistry
Faculty of Medicine and Health Technology
Tampere University
Tampere 33520
Finland

E-mail: pashupati.mishra@tuni.fi

Abbreviations Used

BMI	= body mass index
CLDND1	= claudin domain-containing protein 1
DEG	= differentially expressed gene
DMP	= differentially methylated position
DNMT1	= DNA methyltransferase 1
EWAS	= epigenome-wide association study
FDR	= false discovery rate
GO	= gene ontology
GPR15	= G protein-coupled receptor 15
GSA	= gene set analysis
LRRN3	= leukine-rich repeat neuronal 3
MET-h	= metabolic equivalent task hours
OR	= odds ratio
SD	= standard deviation
SE	= standard error
TSS	= transcription start site
UTR	= untranslated region
YFS	= Young Finns Study