

RESEARCH ARTICLE

# Novel Epigenetic Changes Unveiled by Monozygotic Twins Discordant for Smoking Habits

Alessandra Allione<sup>1,2</sup>, Francesca Marcon<sup>3</sup>, Giovanni Fiorito<sup>1,2</sup>, Simonetta Guarrera<sup>1,2</sup>, Ester Siniscalchi<sup>3</sup>, Andrea Zijno<sup>3</sup>, Riccardo Crebelli<sup>3</sup>, Giuseppe Matullo<sup>1,2\*</sup>

**1** Human Genetics Foundation, HuGeF, Turin, Italy, **2** Department of Medical Sciences, University of Turin, Turin, Italy, **3** Department of Environment and Primary Prevention, Istituto Superiore di Sanità, Rome, Italy

\* [giuseppe.matullo@unito.it](mailto:giuseppe.matullo@unito.it)



OPEN ACCESS

**Citation:** Allione A, Marcon F, Fiorito G, Guarrera S, Siniscalchi E, Zijno A, et al. (2015) Novel Epigenetic Changes Unveiled by Monozygotic Twins Discordant for Smoking Habits. PLoS ONE 10(6): e0128265. doi:10.1371/journal.pone.0128265

**Academic Editor:** Matteo Pellegrini, UCLA-DOE Institute for Genomics and Proteomics, UNITED STATES

**Received:** November 27, 2014

**Accepted:** April 23, 2015

**Published:** June 4, 2015

**Copyright:** © 2015 Allione et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The authors are obliged to protect the study participants' privacy (as outlined in the different informed consent forms and emphasized by Italian privacy law), preventing them from making individual-level data publicly available. Data are available from the HuGeF repository (<http://www.hugef-torino.org/site/index.php?id=52&t=articolo>) for researchers who meet the criteria for access to confidential data. Requests should be addressed to Ms Maria Grazia D'ardes ([mariagrazia.dardes@hugef-torino.org](mailto:mariagrazia.dardes@hugef-torino.org)), secretary of the HuGeF ethics committee.

## Abstract

Exposure to cigarette smoking affects the epigenome and could increase the risk of developing diseases such as cancer and cardiovascular disorders. Changes in DNA methylation associated with smoking may help to identify molecular pathways that contribute to disease etiology. Previous studies are not completely concordant in the identification of differentially methylated regions in the DNA of smokers. We performed an epigenome-wide DNA methylation study in a group of monozygotic (MZ) twins discordant for smoking habits to determine the effect of smoking on DNA methylation. As MZ twins are considered genetically identical, this model allowed us to identify smoking-related DNA methylation changes independent from genetic components. We investigated the whole blood genome-wide DNA methylation profiles in 20 MZ twin pairs discordant for smoking habits by using the Illumina Human-Methylation450 BeadChip. We identified 22 CpG sites that were differentially methylated between smoker and non-smoker MZ twins by intra-pair analysis. We confirmed eight loci already described by other groups, located in *AHRR*, *F2RL3*, *MYOG1* genes, at 2q37.1 and 6p21.33 regions, and also identified several new loci. Moreover, pathway analysis showed an enrichment of genes involved in GTPase regulatory activity. Our study confirmed the evidence of smoking-related DNA methylation changes, emphasizing that well-designed MZ twin models can aid the discovery of novel DNA methylation signals, even in a limited sample population.

## Introduction

The term “epigenetics” refers to heritable changes in gene function, which do not involve changes to the underlying DNA sequence; a change in phenotype without a change in genotype. Epigenetic changes arising during a lifetime are likely to be triggered by individual endogenous factors as well as by environmental factors. Given the central role played by the progressive accumulation of epigenetic changes in the etiopathology of cancer and other degenerative diseases [1], there is growing interest in the effect of modification on disease risk

**Funding:** This study was supported by Compagnia di San Paolo, ([www.compagniadisanpaolo.it](http://www.compagniadisanpaolo.it)) (G.M.); Human Genetics Foundation, ([www.hugef-torino.org](http://www.hugef-torino.org)) (G.M.); Istituto Superiore di Sanita, ([www.iss.it](http://www.iss.it)) (R.C.); Fondazione Umberto Veronesi, ([www.fondazioneveronesi.it](http://www.fondazioneveronesi.it)) (G.M., A.Z.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

exerted by environmental and lifestyle factors, which may ultimately alter the risk of developing cancer. Among lifestyle and behavioral factors, smoking habits deserve major consideration. In fact, epigenetic alterations are likely to play a role in tobacco-induced cancers; an association between smoking and altered DNA methylation of specific genes in lung cancer tissue has been shown [2], as well as the progressive accumulation of epigenetic alterations in the respiratory epithelium of heavy smokers during the carcinogenic process [3, 4]. In addition, recent epigenome-wide association studies identified several genetic loci showing smoking-related DNA methylation changes in DNA from blood [3, 5–10], supporting the use of this surrogate non-tumor tissue to detect altered DNA methylation in smokers.

Monozygotic (MZ) twins provide a powerful model to assess the influence of environmental factors on a particular trait of interest; in subjects who are genetically identical but experience different exposure to environmental or lifestyle factors, any differences can be directly ascribed to this non-shared environment experienced by co-twins [11, 12]. Thus, genetically identical individuals can provide a convenient model to evaluate epigenetic changes where there is no genetic variation.

Taking advantage of the MZ twin-based design, we investigated the effects of smoke exposure on genomic DNA methylation in a population of co-twins with discordant smoking habits [13].

The use of MZ twin pairs who were discordant for smoking habits allowed a perfect match both for the genetic component and the *in utero* and early childhood environmental exposures between smoker and non-smoker co-twins, thus greatly increasing the sensitivity of the analysis. In the present investigation we tested, on a genome-wide level, whether smoking duration and/or intensity could modulate gene-specific methylation in blood DNA.

## Material and Methods

### Sample description

A total of 21 MZ twin pairs who were discordant for smoking habits were enrolled in this cross-sectional study, which was conducted in accordance with the principles of Good Clinical Practice and approved by the Independent Ethics Committee of the University of Rome “Tor Vergata”. Written informed consent was obtained from all subjects according to the Declaration of Helsinki. Information on demographics, medical history, lifestyle, dietary habits, occupational and environmental exposures was collected by means of a questionnaire. Saliva samples were taken from all subjects to verify the zygosity status. Testing for zygosity was performed using the AmpFiSTRs Identifier kit (Applied Biosystems) [13]. All analyses were carried out on coded samples.

### Smoking history

Detailed information on smoking history was collected for all subjects as previously described [13, 14]. Smoking status was assigned as follows: a) smokers: each member of the MZ twin pair that smoked; all smokers had a tar intake of  $\geq 60$  mg tar/day, based on the number of cigarettes smoked per day multiplied by the tar yield of the cigarette brand as declared on the pack (this cut-off was chosen to exclude irregular or occasional smokers); b) non-smokers: each member of the MZ twin pair who has never smoked or had stopped smoking since more than 10 years. To avoid misclassifications based on the self-reported smoker or non-smoker status, urinary cotinine levels were measured in all study subjects using the Accutest NicAlert\_test kit (Jant Pharmaceutical Corp.) according to the manufacturer’s instructions. Overall tobacco exposure during the smoker’s lifetime was quantified by the pack-year parameter which integrates the number of cigarettes smoked per day during the smoking period. In particular, one pack-year is defined as 20

manufactured cigarettes (one pack) smoked per day for one year, and the number of pack-years for each smoker was calculated as (number of cigarettes smoked per day  $\times$  number of years smoked)/20.

### Blood sample collection

Blood samples from all donors were collected within 2 days. Members of the same twin-pair were sampled on the same day, and blood samples were processed in parallel. Differential white blood cells (WBCs) count was determined for each sample, as previously described [13].

### DNA methylation array

Genomic DNA was isolated from whole blood samples using Puregene Core Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions and stored at  $-80^{\circ}\text{C}$ . Bisulphite conversion of 500 ng of each sample was performed using the EZ-96 DNA Methylation Gold kit (Zymo Research, Orange, CA, USA). Bisulphite-converted DNA was hybridized on the Illumina HumanMethylation450 BeadChip (Illumina Inc., San Diego, CA), following the Illumina Infinium HD methylation protocol.

### Statistical analyses

The GenomeStudio Methylation Module v1.0 software (Illumina) was used to convert on-chip fluorescent methylation signals into numerical values ( $\beta$ -values) between 0 and 1, which represent the methylation percentage of each analyzed CpG site.

Raw methylation data were normalized and analyzed according to standard procedures using the open source R statistical environment Bioconductor (methylumi package) [15]. The following criteria were used to filter out samples,  $\beta$ -values, and CpG sites respectively: global call rate  $\leq 95\%$ , detection p-value  $\geq 0.01$ , and detection p-value  $\geq 0.01$  in more than 20% of the assayed samples. To identify differentially methylated CpGs between smokers and non-smokers, we applied the paired Kruskal-Wallis non parametric test. CpG loci with  $p < 0.001$ , and mean methylation differences of at least 0.03 and greater than 3 standard deviations from the mean were selected and further investigated. In order to ascertain whether inter-individual variability in whole blood cell composition could affect the results, for each of the identified differentially methylated sites we compared the estimated differences in methylation by two multivariate linear regression models: the first including cell type composition (from each subject's WBC count) as covariate, and the second without adjusting for cell type.

To investigate the association between time and smoking intensity with DNA-methylation levels of the CpG sites identified in the previous analysis, Spearman correlation coefficients between intra-pair differences (non-smokers vs smokers) in methylation and pack-years were calculated.

### Enrichment analysis

Enrichment analysis was carried out using the Gene Ontology (GO) tool "WEB-based GENE SeT AnaLysis Toolkit" (WebGestalt; <http://bioinfo.vanderbilt.edu/webgestalt/>; [16, 17]). A list of 223 genes in which the top differentially methylated CpGs between smokers and non-smokers were located ( $p < 0.001$ ; S1 Table) was uploaded, 216 of which were recognized by the software and analyzed.

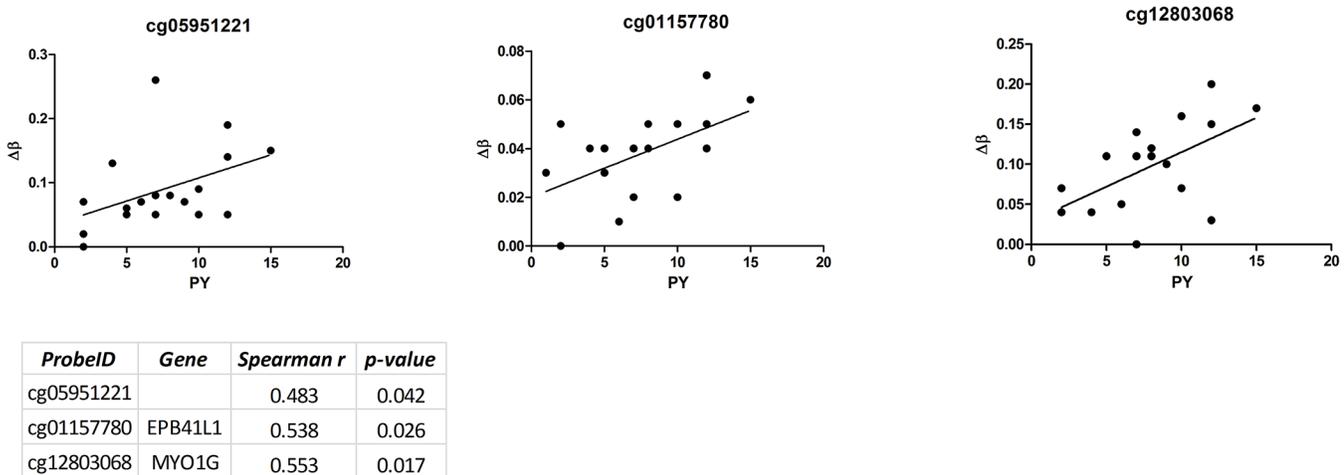
## Results

The final dataset included 20 twin pairs. A total of 436,184 CpG sites passed the quality control, 22 of which (Table 1; CpG list ordered by Mean  $\Delta\beta$ ) had differential methylation levels associated with smoking status (selected as described in the Materials and Methods section). No significant difference was observed comparing DNA methylation between smokers and non-

**Table 1. DNA methylation differences at specific loci between smoking and non-smoking discordant twin pairs.**

ProbeID	Chr	Position	Gene	Mean $\Delta\beta$	SD	p-value	Previously reported
cg05575921	5	373378	AHRR	-0.140	0.018	$1.91 \times 10^{-6}$	Besingi and Johansson 2014; Breitling et al. 2011; Dogan et al. 2014; Elliott et al. 2014; Harlid et al. 2014; Monick et al. 2012; Shenker et al. 2013; Zeilinger et al. 2013
cg21566642	2	233284661		-0.088	0.015	$4.77 \times 10^{-5}$	Monick et al. 2012 (lung macrophages); Shenker et al. 2013; Zeilinger et al. 2013
cg05951221	2	233284402		-0.073	0.013	$5.72 \times 10^{-6}$	Monick et al. 2012 (lung macrophages); Harlid et al. 2014; Shenker et al. 2013
cg06126421	6	30720080		-0.062	0.010	$2.10 \times 10^{-4}$	Besingi and Johansson 2014; Dogan et al. 2014; Elliott et al. 2014; Shenker et al. 2013; Zeilinger et al. 2013
cg01940273	2	233284934		-0.060	0.011	$5.72 \times 10^{-6}$	Shenker et al. 2013; Zeilinger et al. 2013
cg01961092	4	84373320	HELQ	-0.047	0.006	$4.83 \times 10^{-4}$	
cg03636183	19	17000585	F2RL3	-0.046	0.009	$1.91 \times 10^{-5}$	Besingi and Johansson 2014; Harlid et al. 2014; Shenker et al. 2013; Wan et al. 2012; Zeilinger et al. 2013
cg25648203	5	395444	AHRR	-0.042	0.007	$8.20 \times 10^{-5}$	Breitling et al. 2011; Shenker et al. 2013; Zeilinger et al. 2013; Monick et al. 2012 (lung macrophages)
cg13437870	6	52909126	ICK	-0.040	0.007	$4.83 \times 10^{-4}$	
cg04074536	4	798711	CPLX1	-0.038	0.009	$5.86 \times 10^{-4}$	
cg03650233	8	57360226	PENK	-0.038	0.007	$7.08 \times 10^{-4}$	
cg12224879	2	46973883	SOCS5	-0.037	0.005	$2.61 \times 10^{-4}$	
cg09022230	7	5457225	TNRC18	-0.037	0.008	$3.22 \times 10^{-4}$	
cg10655371	7	91749682	CYP51A1	-0.036	0.007	$5.86 \times 10^{-4}$	
cg26979044	16	25045552		-0.036	0.006	$1.68 \times 10^{-4}$	
cg02339888	1	67862336	IL12RB2	-0.034	0.006	$4.83 \times 10^{-4}$	
cg08681409	10	134697042		-0.032	0.007	$2.10 \times 10^{-4}$	
cg23261204	10	133733416		-0.032	0.004	$4.83 \times 10^{-4}$	
cg02567879	1	11761416	C1orf187	-0.030	0.007	$3.95 \times 10^{-4}$	
cg01157780	20	34700376	EPB41L1	0.031	0.006	$1.05 \times 10^{-4}$	
cg06935361	13	32964454	BRCA2	0.049	0.006	$7.08 \times 10^{-4}$	
cg12803068	7	45002919	MYO1G	0.083	0.013	$3.95 \times 10^{-4}$	Zeilinger et al. 2013

doi:10.1371/journal.pone.0128265.t001



**Fig 1. Correlation between methylation levels and smoking intensity (expressed as pack-years) for 2q37.1, EPB41L1 and MYO1G loci.**

doi:10.1371/journal.pone.0128265.g001

smokers with and without adjusting for WBC proportions, thus we described the results obtained without any adjustment for WBC count.

Out of the 22 CpGs with differential methylation levels, 15 were located in known genes. The majority of these CpGs (86%) were found to be less methylated in smokers. The most significant CpG site, cg05575921 ( $\Delta$  smokers vs non-smokers -14%,  $p = 1.91 \times 10^{-6}$ ), was located in the *AHRR* gene. The second region was on chromosome 2q37.1 and comprised three differentially methylated CpG sites ( $\Delta$  smokers vs non-smokers: cg21566642-8.8%,  $p = 4.77 \times 10^{-5}$ ; cg233284402-7.3%,  $p = 5.72 \times 10^{-6}$ ; cg233284934-6%,  $p = 5.72 \times 10^{-6}$ , respectively).

The intergenic region at 6p21.33 was located in a gene desert region and included the CpG locus cg06126421, which is less methylated in smokers than in non-smokers ( $\Delta$  smokers vs non-smokers = -6.2%,  $p = 2.10 \times 10^{-4}$ ).

We found DNA hypomethylation in smokers in the cg01961092 site, located in the *HELQ* gene body ( $\Delta$  smokers vs non-smokers = -4.7%,  $p = 4.83 \times 10^{-4}$ ) and we also report a significant association of tobacco smoking with the CpG site cg03636183, located within the *F2RL3* gene ( $\Delta$  smokers vs non-smokers = -4.6%,  $p = 1.91 \times 10^{-5}$ ).

Other hypomethylated sites in smokers with a difference > 3% were located in the following genes: *ICK*, *CPLX1*, *PENK*, *SOCS5*, *TNRC18*, *CYP51A1* and *IL12RB2* or in regions with no annotated transcripts. Three sites resulted as being hypermethylated in smokers, which were located in the *MYO1G* (cg12803068), *BRCA2* (cg06935361) and *EPB41L1* (cg01157780) genes.

Moreover, a positive correlation between methylation  $\Delta$ s (as described in the Materials and Methods section) and smoking intensity (expressed as pack-years) was found for three of the above mentioned loci: 2q37.1, *EPB41L1* and *MYO1G* (Fig 1).

For the gene-set enrichment analysis, a list of 223 genes in which CpGs differentially methylated between smokers and non-smokers were found to be located (S1 Table) was considered, 216 of which were recognized by WebGestalt software and analyzed for GO enrichment analysis. A total of six molecular function categories (Table 2) were significantly enriched (FDR q-value >0.05). The most enriched molecular function categories were all related to GTPase regulatory activity; in particular the Rho- and Ras-guanyl-nucleotide exchange factor activity categories were the most enriched (FDR q-value = 0.0158 for both). The genes harboring differentially methylated CpGs that were classified as members of the “small GTPase regulator activity”, “Rho guanyl-nucleotide exchange factor activity” and “Ras guanyl-nucleotide exchange factor activity”

**Table 2. Gene enrichment results carried out using the Gene Ontology (GO) enrichment analysis web-based bioinformatics tool WEB-based GENE Set AnaLysis Toolkit (WebGestalt).**

Molecular functions that were enriched for genes differentially methylated in smokers				
GO Term	Description	p-value	FDR q-value	Genes
GO:0005083	small GTPase regulator activity	0.0004	0.0158	<i>PSD4, TBC1D22A, MCF2L, RASGRF2, IQSEC1, KALRN, OBSCN, DENND3, ARHGEF16</i>
GO:0005085	guanyl-nucleotide exchange factor activity	9.85x10 <sup>-05</sup>	0.0158	<i>PSD4, MCF2L, RASGRF2, IQSEC1, KALRN, OBSCN, DENND3, ARHGEF16</i>
GO:0005089	Rho guanyl-nucleotide exchange factor activity	0.0005	0.0158	<i>MCF2L, RASGRF2, KALRN, OBSCN, ARHGEF16</i>
GO:0005088	Ras guanyl-nucleotide exchange factor activity	0.0003	0.0158	<i>MCF2L, RASGRF2, KALRN, OBSCN, DENND3, ARHGEF16</i>
GO:0005543	phospholipid binding	0.0005	0.0158	<i>IRS2, DOC2A, SH3BP2, PSD4, MCF2L, RASGRF2, OSBPL5, IQSEC1, OBSCN, KALRN, APOA1, ARHGEF16</i>
GO:0070644	vitamin D response element binding	0.0004	0.0158	<i>TCF3, RXRA</i>
Cellular components that were enriched for genes differentially methylated in smokers				
GO Term	Description	p-value	FDR q-value	Genes
GO:0017053	transcriptional repressor complex	0.0003	0.0483	<i>TFAP4, NCOR2, ZFPM1, PRDM16, MTA1</i>

doi:10.1371/journal.pone.0128265.t002

categories were: *PSD4, TBC1D22A, MCF2L, RASGRF2, IQSEC1, KALRN, OBSCN, DENND3, ARHGEF16*.

## Discussion

In order to investigate gene-specific DNA methylation changes associated with smoking habits we used a high-throughput whole-genome approach in a sample of MZ twins discordant for smoking habits. Comparing discordant MZ twins represents a powerful improvement over the traditional case-control study to search for disease-associated biological markers and environmental exposure-related phenotypes. The strength of this design to unveil epigenetic effects was demonstrated in recent studies which compared the DNA methylation patterns of MZ twins who were discordant for various diseases [18–22]. Direct comparisons between identical twins constitute an ideal experimental model for testing environmental epigenetics because differences in DNA sequences, including single-nucleotide polymorphisms (SNPs) abundant in singleton-based studies, which could potentially modulate the effects of environmental exposures or the DNA methylation levels, do not interfere in the analysis.

Several of the most significant smoking-associated hypomethylated CpG sites that we have described here have been already reported in the literature (see Table 1).

The underlying relationship of AHRR hypomethylation and smoking is further supported by Monick *et al.* [23], who reported a significant association between cg05575921 DNA methylation and AHRR gene expression and by Philibert *et al.* [24], showing a high correlation of AHRR DNA methylation and serum cotinine levels in smokers. The latter concluded that AHRR DNA methylation status is a sensitive marker of smoking history, which could help with integrating self-reported smoking habits or existing smoking-related biomarkers used in clinical or epidemiological settings (e.g., cotinine). In addition, if properly configured as a clinical assay, AHRR methylation levels could also be used as a screening tool in efforts to target anti-smoking interventions to nascent smokers in the early phases of smoking.

We found a novel CpG hypomethylated in smokers (cg01961092) located in the HELQ gene body. HELQ is a 3'–5' DNA helicase with strand displacement activity that was recently shown

to participate in resistance to the DNA interstrand crosslinking pathway [25], caused also by aldehydes such as acrolein and crotonaldehyde, found in tobacco smoke [26].

The *F2RL3* (coagulation factor II receptor-like 3) gene encodes for a protease-activated receptor-4 (PAR-4), which has been suggested to be involved in the pathophysiology of both cardiovascular and neoplastic diseases. These patterns suggest that the adverse health effects of smoking could be partly mediated by pathways through *F2RL3* methylation. Recently Zhang *et al.*, showed that *F2RL3* methylation intensity is strongly correlated with smoking status [27] and is a predictor of smoking-related mortality [28].

No clear biological relationships between smoking and other identified genes (*CPLX1*, *PENK*, *SOCS5*, *IL12RB2*, *CYP51A1*, *TNRC18*, *ICK*) are evident, however some of these genes have been previously reported as smoking-associated.

We also found CpG loci with higher DNA methylation levels in smokers located in the *MYO1G*, *BRCA2* and *EPB41L1* genes (Table 1).

The *MYO1G* cg12803068 site and other CpGs in the same gene have been already described as hypermethylated in smokers in previous studies [7, 9, 29]. *MYO1G* is a plasma membrane-associated class I myosin that is abundant in T and B lymphocytes and mast cells [30], and contributes to cell elasticity [31]. Altered expression of *MYO1G* could contribute in explaining the increased fibrosis found in several smoking-targeted tissues.

None of the reported differentially methylated loci reached a Bonferroni-corrected p-value threshold ( $p = 1.15 \times 10^{-7}$ ). However, this statistical correction is likely to be too conservative given the non-independence of CpG sites and the identical MZ twin intra-pair genetic background. In view of the above considerations, in order to identify differentially methylated loci, we used a combined analytical approach which took into account both the statistical significance (with  $p < 0.001$  as an arbitrary threshold) and the extent of methylation changes (SD and mean  $\beta$ -value differences). This analytical approach is widely used in genome-wide gene expression studies, and is reported to identify loci with high reproducibility and biological relevance compared with conventional methods that merely rely on single CpG statistical significance. Moreover, the usefulness of this approach is supported by the identification (also in our study) of differentially methylated loci already described in association with smoking, showing the same methylation trends as those previously reported.

MZ twins provide a valuable model to identify environmental effects without the influence of genetic variability. Nevertheless, given the relatively small sample size, replication of our newly identified results in larger samples is needed.

The difference we observed in *AHRR* mean methylation levels (14%, range 1–32%) between smokers and non-smokers was comparable with those reported in larger non-twin independent populations studied by other authors (17% in [3]; 24.4% in [29]; 7.4% in [9]; 15% in [8]; 22% in [7]). However, large-scale epigenetic studies are still ongoing, and a threshold to define the minimum DNA methylation changes able to exert functional significance at the population level has not yet been firmly established.

To examine the possible functional impact of differential methylation at loci reported there, we conducted an exploratory analysis using a gene expression/methylation dataset produced by our group in the context of a previous study [32].

The database comprises paired gene expression and methylation data (whole blood) from the same subjects and allows methylation/gene expression correlation analyses. Unfortunately, most of the genes we identified as differentially methylated in this study were only weakly or not expressed, and only the *HELQ* gene showed evidence of association between gene expression and DNA methylation (Spearman correlation, cg01961092 methylation/gene expression  $r = -0.26$ ,  $p = 0.02$ ). Recently, a large-scale genome-wide expression profiling has been

performed, but none of our differently methylated genes was found to be differentially expressed in smokers' blood cells [33].

We also investigated the relationship between methylation levels and smoking intensity expressed as pack-year. We observed that in three loci, 2q37.1, *EPB41L1*, and *MYO1G*, the methylation differences between smokers and non-smokers increased according to pack-year, indicating that these three loci could be further used as epigenetic markers of smoking intensity.

The most enriched GO molecular functions were all related to GTPase regulatory activity. Interestingly, the same pathways were identified by analyzing differentially methylated genes in alveolar macrophages from smokers [23]. Moreover, molecular functions of these genes are of special interest since a recent study described gene-smoking interactions, identifying several single nucleotide polymorphisms located in genes belonging to the Rho-GTPase pathway associated with an early-onset of coronary artery disease in smokers [34]. Another enriched molecular function was "vitamin D response element binding". Serum vitamin D levels are inversely correlated with the incidence, prevalence and death rates of different tumors [35], and recently Zheng *et al.* suggested a protective effect of vitamin D against colorectal cancer, whose effects are more beneficial in active smokers than in non-smokers [36]. Moreover, vitamin D deficiency have been associated with a diminished lung function and more rapid lung function decline in smokers in a longitudinal cohort of elderly men [37]. These results suggest that vitamin D may have a protective effect against the damaging effects of smoking on lung function. Smoking has a significant effect on calcium and vitamin D metabolism: the repression of the vitamin D-parathyroid hormone system, evident among smokers may represent another potential mechanism for the deleterious effects of smoking on bone metabolism, and may contribute to the reported risk of osteoporosis among smokers [38].

## Conclusions

In this study, we successfully identified 22 DNA methylation sites that were differentially methylated. By using MZ twins who were discordant for smoking habits, we were able to replicate most of the previously published significant sites, with a relatively small number of phenotypically discordant twin pairs. Moreover, we were also able to detect novel smoking-related differentially methylated CpG sites. Considering both our results along with previously published results, it is becoming evident that environmental exposures, such as smoking, may modify the epigenetic profiles of several genes. Given the major burden to individuals and to society of smoking-related diseases, further functional studies are warranted to better explain the pathogenic role exerted by smoking through epigenetic changes.

## Supporting Information

**S1 Table. List of genes where the top differentially methylated CpGs between smokers and non-smokers were located ( $p < 0.0001$ ).**

(DOCX)

## Acknowledgments

The authors would like to thank Dr. Cristina Andreoli for outstanding scientific revision of the manuscript.

## Author Contributions

Conceived and designed the experiments: AZ RC GM. Performed the experiments: FM ES SG AA. Analyzed the data: GF. Contributed reagents/materials/analysis tools: FM. Wrote the paper: AA FM GM.

## References

1. Hattori N, Ushijima T. Compendium of aberrant DNA methylation and histone modifications in cancer. *Biochem Biophys Res Commun*. 2014; 455(1–2):3–9. Epub 2014/09/10. doi: [10.1016/j.bbrc.2014.08.140](https://doi.org/10.1016/j.bbrc.2014.08.140) PMID: [25194808](https://pubmed.ncbi.nlm.nih.gov/25194808/).
2. Tan Q, Wang G, Huang J, Ding Z, Luo Q, Mok T, et al. Epigenomic analysis of lung adenocarcinoma reveals novel DNA methylation patterns associated with smoking. *Onco Targets Ther*. 2013; 6:1471–9. Epub 2013/11/10. doi: [10.2147/OTT.S51041](https://doi.org/10.2147/OTT.S51041) PMID: [24204162](https://pubmed.ncbi.nlm.nih.gov/24204162/); PubMed Central PMCID: PMC3818101.
3. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. 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*. 2013; 22(5):843–51. Epub 2012/11/24. doi: [10.1093/hmg/dds488](https://doi.org/10.1093/hmg/dds488) PMID: [23175441](https://pubmed.ncbi.nlm.nih.gov/23175441/).
4. Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD, et al. Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res*. 2005; 11(18):6505–11. Epub 2005/09/17. doi: [10.1158/1078-0432.CCR-05-0625](https://doi.org/10.1158/1078-0432.CCR-05-0625) PMID: [16166426](https://pubmed.ncbi.nlm.nih.gov/16166426/).
5. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011; 88(4):450–7. Epub 2011/04/05. doi: [10.1016/j.ajhg.2011.03.003](https://doi.org/10.1016/j.ajhg.2011.03.003) PMID: [21457905](https://pubmed.ncbi.nlm.nih.gov/21457905/); PubMed Central PMCID: PMC3071918.
6. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet*. 2012; 21(13):3073–82. Epub 2012/04/12. doi: [10.1093/hmg/dds135](https://doi.org/10.1093/hmg/dds135) PMID: [22492999](https://pubmed.ncbi.nlm.nih.gov/22492999/); PubMed Central PMCID: PMC3373248.
7. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics*. 2014; 6(1):4. Epub 2014/02/04. doi: [10.1186/1868-7083-6-4](https://doi.org/10.1186/1868-7083-6-4) PMID: [24485148](https://pubmed.ncbi.nlm.nih.gov/24485148/); PubMed Central PMCID: PMC3915234.
8. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, et al. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics*. 2014; 15:151. Epub 2014/02/25. doi: [10.1186/1471-2164-15-151](https://doi.org/10.1186/1471-2164-15-151) PMID: [24559495](https://pubmed.ncbi.nlm.nih.gov/24559495/); PubMed Central PMCID: PMC3936875.
9. Besingi W, Johansson A. Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet*. 2014; 23(9):2290–7. Epub 2013/12/18. doi: [10.1093/hmg/ddt621](https://doi.org/10.1093/hmg/ddt621) PMID: [24334605](https://pubmed.ncbi.nlm.nih.gov/24334605/).
10. Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the sister study. *Environ Health Perspect*. 2014; 122(7):673–8. Epub 2014/04/08. doi: [10.1289/ehp.1307480](https://doi.org/10.1289/ehp.1307480) PMID: [24704585](https://pubmed.ncbi.nlm.nih.gov/24704585/); PubMed Central PMCID: PMC4080519.
11. Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet*. 2011; 27(3):116–25. Epub 2011/01/25. doi: [10.1016/j.tig.2010.12.005](https://doi.org/10.1016/j.tig.2010.12.005) PMID: [21257220](https://pubmed.ncbi.nlm.nih.gov/21257220/); PubMed Central PMCID: PMC3063335.
12. Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, et al. A longitudinal study of epigenetic variation in twins. *Epigenetics*. 2010; 5(6):516–26. Epub 2010/05/28. PMID: [20505345](https://pubmed.ncbi.nlm.nih.gov/20505345/); PubMed Central PMCID: PMC3322496.
13. Andreoli C, Gregg EO, Puntoni R, Gobbi V, Nunziata A, Bassi A. Cross-sectional study of biomarkers of exposure and biological effect on monozygotic twins discordant for smoking. *Clin Chem Lab Med*. 2011; 49(1):137–45. Epub 2010/11/19. doi: [10.1515/CCLM.2011.009](https://doi.org/10.1515/CCLM.2011.009) PMID: [21083439](https://pubmed.ncbi.nlm.nih.gov/21083439/).
14. Marcon F, Carotti D, Andreoli C, Siniscalchi E, Leopardi P, Caiola S, et al. DNA damage response in monozygotic twins discordant for smoking habits. *Mutagenesis*. 2013; 28(2):135–44. Epub 2012/11/08. doi: [10.1093/mutage/ges062](https://doi.org/10.1093/mutage/ges062) PMID: [23132286](https://pubmed.ncbi.nlm.nih.gov/23132286/).
15. Davis S, Du P, Bilke S, Triche T, Bootwalla M. Methylumi: Handle Illumina methylation data. R package version 2.6.1. 2013.
16. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT Analysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res*. 2013; 41(Web Server issue):W77–83. Epub 2013/05/25. doi: [10.1093/nar/gkt439](https://doi.org/10.1093/nar/gkt439) PMID: [23703215](https://pubmed.ncbi.nlm.nih.gov/23703215/); PubMed Central PMCID: PMC3692109.

17. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* 2005; 33(Web Server issue):W741–8. Epub 2005/06/28. doi: [10.1093/nar/gki475](https://doi.org/10.1093/nar/gki475) PMID: [15980575](https://pubmed.ncbi.nlm.nih.gov/15980575/); PubMed Central PMCID: PMC1160236.
18. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* 2010; 20(2):170–9. Epub 2009/12/24. doi: [10.1101/gr.100289.109](https://doi.org/10.1101/gr.100289.109) PMID: [20028698](https://pubmed.ncbi.nlm.nih.gov/20028698/); PubMed Central PMCID: PMC2813473.
19. Wong CC, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, et al. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry.* 2014; 19(4):495–503. Epub 2013/04/24. doi: [10.1038/mp.2013.41](https://doi.org/10.1038/mp.2013.41) PMID: [23608919](https://pubmed.ncbi.nlm.nih.gov/23608919/); PubMed Central PMCID: PMC3906213.
20. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* 2011; 7(9):e1002300. Epub 2011/10/08. doi: [10.1371/journal.pgen.1002300](https://doi.org/10.1371/journal.pgen.1002300) PMID: [21980303](https://pubmed.ncbi.nlm.nih.gov/21980303/); PubMed Central PMCID: PMC3183089.
21. Heyn H, Carmona FJ, Gomez A, Ferreira HJ, Bell JT, Sayols S, et al. DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. *Carcinogenesis.* 2013; 34(1):102–8. Epub 2012/10/12. doi: [10.1093/carcin/bgs321](https://doi.org/10.1093/carcin/bgs321) PMID: [23054610](https://pubmed.ncbi.nlm.nih.gov/23054610/); PubMed Central PMCID: PMC3534196.
22. Gervin K, Vigeland MD, Mattingsdal M, Hammero M, Nygard H, Olsen AO, et al. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet.* 2012; 8(1):e1002454. Epub 2012/02/01. doi: [10.1371/journal.pgen.1002454](https://doi.org/10.1371/journal.pgen.1002454) PMID: [22291603](https://pubmed.ncbi.nlm.nih.gov/22291603/); PubMed Central PMCID: PMC3262011.
23. Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet.* 2012; 159B(2):141–51. Epub 2012/01/11. doi: [10.1002/ajmg.b.32021](https://doi.org/10.1002/ajmg.b.32021) PMID: [22232023](https://pubmed.ncbi.nlm.nih.gov/22232023/); PubMed Central PMCID: PMC3318996.
24. Philibert RA, Beach SR, Lei MK, Brody GH. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin Epigenetics.* 2013; 5(1):19. Epub 2013/10/15. doi: [10.1186/1868-7083-5-19](https://doi.org/10.1186/1868-7083-5-19) PMID: [24120260](https://pubmed.ncbi.nlm.nih.gov/24120260/); PubMed Central PMCID: PMC3819644.
25. Takata K, Reh S, Tomida J, Person MD, Wood RD. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat Commun.* 2013; 4:2338. Epub 2013/09/06. doi: [10.1038/ncomms3338](https://doi.org/10.1038/ncomms3338) PMID: [24005565](https://pubmed.ncbi.nlm.nih.gov/24005565/); PubMed Central PMCID: PMC3778836.
26. Kozekov ID, Nechev LV, Moseley MS, Harris CM, Rizzo CJ, Stone MP, et al. DNA interchain crosslinks formed by acrolein and crotonaldehyde. *J Am Chem Soc.* 2003; 125(1):50–61. Epub 2003/01/08. doi: [10.1021/ja020778f](https://doi.org/10.1021/ja020778f) PMID: [12515506](https://pubmed.ncbi.nlm.nih.gov/12515506/).
27. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect.* 2014; 122(2):131–7. Epub 2013/11/26. doi: [10.1289/ehp.1306937](https://doi.org/10.1289/ehp.1306937) PMID: [24273234](https://pubmed.ncbi.nlm.nih.gov/24273234/); PubMed Central PMCID: PMC3915264.
28. Zhang Y, Yang R, Burwinkel B, Breitling LP, Hollecsek B, Schottker B, et al. F2RL3 methylation in blood DNA is a strong predictor of mortality. *Int J Epidemiol.* 2014. Epub 2014/02/11. doi: [10.1093/ije/dyu006](https://doi.org/10.1093/ije/dyu006) PMID: [24510982](https://pubmed.ncbi.nlm.nih.gov/24510982/).
29. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One.* 2013; 8(5):e63812. Epub 2013/05/22. doi: [10.1371/journal.pone.0063812](https://doi.org/10.1371/journal.pone.0063812) PMID: [23691101](https://pubmed.ncbi.nlm.nih.gov/23691101/); PubMed Central PMCID: PMC3656907.
30. Maravillas-Montero JL, Lopez-Ortega O, Patino-Lopez G, Santos-Argumedo L. Myosin 1g regulates cytoskeleton plasticity, cell migration, exocytosis, and endocytosis in B lymphocytes. *Eur J Immunol.* 2014; 44(3):877–86. Epub 2013/12/07. doi: [10.1002/eji.201343873](https://doi.org/10.1002/eji.201343873) PMID: [24310084](https://pubmed.ncbi.nlm.nih.gov/24310084/).
31. Olety B, Walte M, Honnert U, Schillers H, Bahler M. Myosin 1G (Myo1G) is a haematopoietic specific myosin that localises to the plasma membrane and regulates cell elasticity. *FEBS Lett.* 2010; 584(3):493–9. Epub 2009/12/09. doi: [10.1016/j.febslet.2009.11.096](https://doi.org/10.1016/j.febslet.2009.11.096) PMID: [19968988](https://pubmed.ncbi.nlm.nih.gov/19968988/).
32. Matullo G, Guarrera S, Betti M, Fiorito G, Ferrante D, Voglino F, et al. Genetic variants associated with increased risk of malignant pleural mesothelioma: a genome-wide association study. *PLoS One.* 2013; 8(4):e61253. Epub 2013/04/30. doi: [10.1371/journal.pone.0061253](https://doi.org/10.1371/journal.pone.0061253) PMID: [23626673](https://pubmed.ncbi.nlm.nih.gov/23626673/); PubMed Central PMCID: PMC3634031.
33. Charlesworth JC, Curran JE, Johnson MP, Goring HH, Dyer TD, Diego VP, et al. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. *BMC Med Genomics.* 2010; 3:29. Epub 2010/07/17. doi: [10.1186/1755-8794-3-29](https://doi.org/10.1186/1755-8794-3-29) PMID: [20633249](https://pubmed.ncbi.nlm.nih.gov/20633249/); PubMed Central PMCID: PMC2911391.

34. Ward-Caviness C, Haynes C, Blach C, Dowdy E, Gregory SG, Shah SH, et al. Gene-smoking interactions in multiple Rho-GTPase pathway genes in an early-onset coronary artery disease cohort. *Hum Genet.* 2013; 132(12):1371–82. Epub 2013/08/03. doi: [10.1007/s00439-013-1339-7](https://doi.org/10.1007/s00439-013-1339-7) PMID: [23907653](https://pubmed.ncbi.nlm.nih.gov/23907653/); PubMed Central PMCID: PMC3835376.
35. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, et al. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. *Int J Cancer.* 2011; 128(6):1414–24. Epub 2010/05/18. doi: [10.1002/ijc.25439](https://doi.org/10.1002/ijc.25439) PMID: [20473927](https://pubmed.ncbi.nlm.nih.gov/20473927/).
36. Zheng XE, Lipka S, Li T, Shahzad G, Levine E, Vlacancich R, et al. The relationship of vitamin D status, smoking, and colorectal adenoma: a retrospective study in an ethnically diverse community. *J Steroid Biochem Mol Biol.* 2013; 136:280–3. Epub 2012/09/25. doi: [10.1016/j.jsbmb.2012.09.018](https://doi.org/10.1016/j.jsbmb.2012.09.018) PMID: [23000288](https://pubmed.ncbi.nlm.nih.gov/23000288/).
37. Lange NE, Sparrow D, Vokonas P, Litonjua AA. Vitamin D deficiency, smoking, and lung function in the Normative Aging Study. *Am J Respir Crit Care Med.* 2012; 186(7):616–21. Epub 2012/07/24. doi: [10.1164/rccm.201110-1868OC](https://doi.org/10.1164/rccm.201110-1868OC) PMID: [22822023](https://pubmed.ncbi.nlm.nih.gov/22822023/); PubMed Central PMCID: PMC3480523.
38. Brot C, Jorgensen NR, Sorensen OH. The influence of smoking on vitamin D status and calcium metabolism. *Eur J Clin Nutr.* 1999; 53(12):920–6. Epub 1999/12/22. PMID: [10602348](https://pubmed.ncbi.nlm.nih.gov/10602348/).