CLINICAL EPIGENETICS

RESEARCH

CrossMark

A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood

Mario Bauer[1], Gunter Linsel[2], Beate Fink[1], Kirsten Offenberg[1], Anne Maria Hahn[5], Ulrich Sack[4], Heike Knaack[4], Markus Eszlinger[3] and Gunda Herberth[1*]

Abstract

Background: Many recent epigenetic studies report that cigarette smoking reduces DNA methylation in whole blood at the single CpG site cg19859270 within the *GPR15* gene.

Results: Within two independent cohorts, we confirmed the differentially expression of the *GPR15* gene when smokers and non-smokers subjects are compared. By validating the GPR15 protein expression at the cellular level, we found that the observed decreased methylation at this site in white blood cells (WBC) of smokers is mainly caused by the high proportion of CD3+GPR15+ expressing T cells in peripheral blood. In current smokers, the percentage of GPR15+ cells among CD3+ T cells in peripheral blood is significantly higher (15.5 ± 7.2 %, mean ± standard deviation) compared to non-smokers (3.7 ± 1.6 %). Treatment of peripheral blood mononuclear cell (PBMC) cultures with aqueous cigarette smoke extract did not induce a higher proportion of this T cell subtype.

Conclusions: Our results underline that DNA hypomethylation at cg19859270 site, observed in WBCs of smokers, did not arise by direct effect of tobacco smoking compounds on methylation of DNA but rather by the enrichment of a tobacco-smoking-induced lymphocyte population in the peripheral blood.

Keywords: DNA methylation, CpG, GPR15, Tobacco smoking, Blood, EWAS cell composition

Background

A surge in recent publications shows a potential link between epigenetic variation and environmental exposure or the etiology of human diseases. For example, epigenetic studies have identified strong associations between tobacco smoking and altered DNA methylation at single sites (CpG) in peripheral blood cells [1–7] (Table 1). To date, eight loci (including *GPR15*, *AHRR*, *F2RL3*) have been confirmed at the genome-wide level to show differential methylation when current smokers and non-smokers are compared [5]. Epigenome-wide association studies (EWAS) of DNA methylation using whole-blood-derived DNA is complicated by the heterogeneity of cell types within blood. For instance, DNA from peripheral blood is a mixture of genetic substrate from various

leukocyte subtypes, and variation in leukocyte proportions may confound true epigenetic associations between methylation and a dependent variable of interest. There are some approaches which aim to deconvolute subject-specific blood composition by using DNA methylation signatures of cell lineage markers [8–14]. Here, in most cases, differentially methylated regions (DMRs) for the lineage markers for B cells (CD19+), granulocytes (CD15+), monocytes (CD14+), NK (CD56+), CD3+CD4+ T cells, CD3+CD8+ T cells, NKT (CD3+CD56+), other T cells (CD3+) are used to determine the frequency of these cells in whole blood. However, variations in minor immune cell fractions or subtypes not covered by the yet analyzed lineage markers may also skew DNA methylation results relating to environmental exposure or disease outcomes. From an immunological point of view, it is obvious that most diseases, when T cell dependent, are linked to the frequency or activation of certain T cell subtypes rather than to CD3+ T cells in general. Asthma and inflammatory

* Correspondence: gunda.herberth@ufz.de
[1]Department of Environmental Immunology, Helmholtz Centre for Environmental Research-UFZ, Leipzig 04318, Germany
Full list of author information is available at the end of the article

BioMed Central

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 2 of 11

Table 1 Summary of reports indicating that tobacco smoking evokes hypomethylation at single CpG sites in white blood cells

Report	Hypomethylation[a]	Methylation change		
		Max[b]	cg19859270[c]	cg03636183[c]
	% of CpG sites (n)	(%)		
Whole blood				
Breitling et al., 2011 [1]	84 (13)	−14	−4.3	−12.6
Wan et al., 2012 [6]	60 (15)	−8	−2; −2	−8; −6
Zeilinger et al., 2013 [7]	62 (29)	−25	−1.4; −2	−15; −17
Shenker et.al., 2013 [3]	83 (36)	−16	n/a	−7; −8
Sun et al., 2013 [4]	93 (15)	−10	−2.1; −2.2	−9.6; −3
Tsaprouni et al., 2014 [5]	96 (30)	−28	−3.5; −2.7	−13.4; −10.6
PBMC				
Dogan et al., 2014 [2]	60 (30)	−19	−10	n/a

n/a not available
[a]Describes the methylation change in smokers vs. non-smokers of the prominent highlighted single CpGs
[b]Indicates the maximum of methylation change at single CpG
[c]Highlighted are values for the main and replicated cohorts of appropriate studies

bowel diseases, for example, are characterized by low frequencies of mucosal associated invariant T cells, a T cell subtype in peripheral blood expressing CD3/CD8/CD161/TCRa7.2 [15, 16]. A difference at CD3 level between controls and cases is not detectable. This example highlights the importance of accounting for the T cell subtype in EWAS, independently of whether the variable of interest is environmental exposure or disease. Therefore, in our present study, we aimed to extend the already published epigenetic observations regarding tobacco smoking and single CpG hypomethylation by examining the biological significance at cellular level. We focused on the CpG site cg19859270 which is located in the gene body of GPR15, an orphan chemoattractant receptor and co-receptor for HIV. Our data show that the observed decreased methylation at this site in white blood cells (WBC) of smokers is directly associated to the proportion of CD3+GPR15+ expressing T cells. In current smoker, the percentage of GPR15+ cells among CD3+ T cells in peripheral blood is significantly higher (15.5 ± 7.2 %, mean ± standard deviation) compared to non-smoker (3.7 ± 1.6 %). However, treatment of WBC cultures with aqueous cigarette smoke extract (CSE) did not induce a higher proportion of this T cell subtype, suggesting that an immunological cascade is evoking GPR15 expressing lymphocytes rather than a direct and specific impact of tobacco smoke on methylation.

Results

GPR15—most differentially expressed gene in smoker vs. non-smoker

To identify a phenotypic consequence of prominent reported hypomethylation at single CpG sites by tobacco smoking, gene expression of single CpG linked genes (*GPR15, F2RL3, AHRR, LIM2, LRRN3, MYLK, PTPRO*) was analyzed in a Working Place Cohort comprising 42 active smokers and 31 non-smokers as well as 20 formerly smokers (Table 1). Out of the 159 implemented genes of interest, differential expression of the two CpG linked genes *GPR15* and *AHRR* next to other non-CpG linked genes (*CCR5, CXCL7, FOXP3, ALOX12,* and *CCL2*) between smoker and non-smoker remained significant after correction for multiple testing (see Additional file 1). *GPR15* expression in whole blood differed mostly with respect to smoking status and independent of the daily time of blood donation (see Additional file 2). In active smokers (< or >10 cigarettes/day) the *GPR15* expression was 5.3-fold higher than in non-smokers ($p = 4.7 \times 10^{-19}$) (Fig. 1b). Former smokers included in the Working Place Cohort showed a significant intermediate value with a 2.0-fold higher ($p = 2.9 \times 10^{-8}$) and 2.8-fold lower ($p = 5.5 \times 10^{-3}$) extent toward non-smoker and smoker (<10 cigarettes/day), respectively (Fig. 1a).
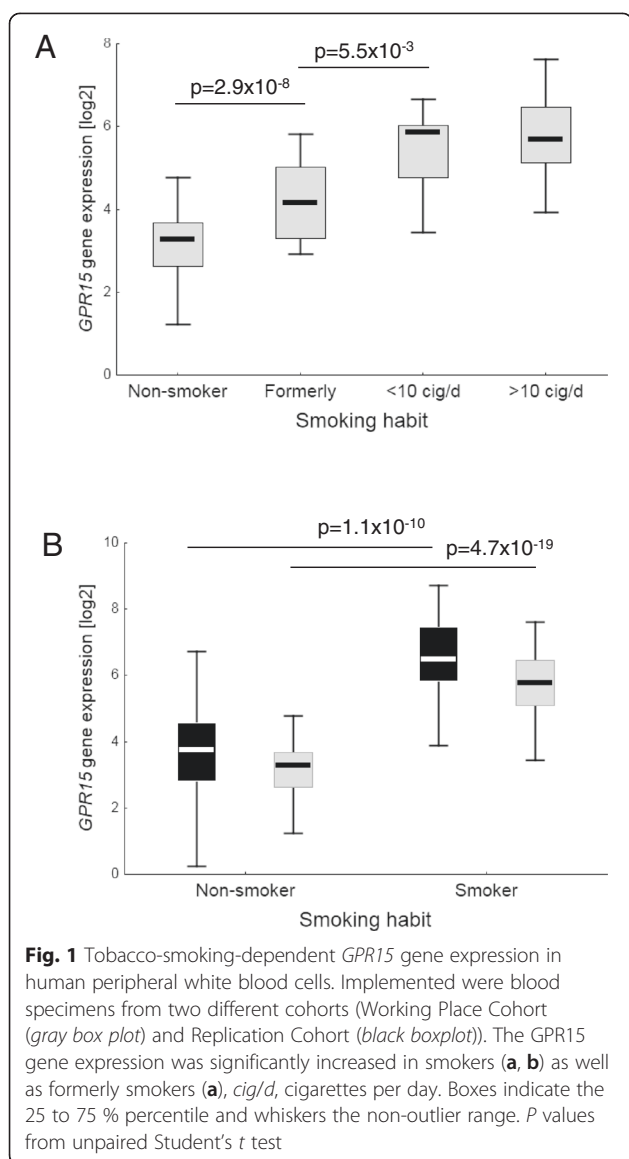
Further validation studies were performed only for *GPR15*.

Replication of *GPR15* gene expression

In order to confirm the relationship between differential expression of *GPR15* and smoking behavior, a replication cohort with 100 randomly selected volunteers was generated. The Replication Cohort comprised 18 active smokers and 82 non-smokers (Table 1). In this cohort, and in a very similar range to the Working Place Cohort, we found that *GPR15* was significantly higher expressed in blood samples of active smokers compared to non-smoker (Fig. 1b).

Validation of GPR15 protein expression at the cellular level

In order to identify the blood cell type expressing GPR15, flow cytometric analyses were performed using

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 3 of 11

Fig. 1 Tobacco-smoking-dependent *GPR15* gene expression in human peripheral white blood cells. Implemented were blood specimens from two different cohorts (Working Place Cohort (*gray box plot*) and Replication Cohort (*black boxplot*)). The GPR15 gene expression was significantly increased in smokers (**a**, **b**) as well as formerly smokers (**a**), cig/d, cigarettes per day. Boxes indicate the 25 to 75 % percentile and whiskers the non-outlier range. *P* values from unpaired Student's *t* test

whole blood samples from the Replication Cohort as well as peripheral blood mononuclear cells (PBMCs) from buffy coats. The main population expressing GPR15 was CD3+ T cells (Fig. 2a, see Additional file 3). Smokers had a significantly ($p = 1.8 \times 10^{-10}$) increased proportion of GPR15+ cells among CD3+ T cells (15.5 ± 7.2 %, mean ± standard deviation) compared to non-smoker (3.7 ± 1.6 %, Fig. 2b). By setting a cutoff of 9 % GPR15 expressing cells among CD3+ T cells in blood, the flow cytometric analysis could distinguish smoker from non-smoker with high sensitivity (0.88 %) and high specificity (0.99 %). In addition to T cells, a low proportion of B cells expressed GPR15 (Fig. 2a). The proportion of CD19+GPR15+ B cells was significantly ($p = 1.5 \times 10^{-5}$) higher in smoker (7.98 ± 7.54 %) compared to non-smoker (3.81 ± 2.89 %, Fig. 2b).

Correlation between *GPR15* gene expression and frequency of GPR15+ cells in whole blood

In the Replication Cohort, *GPR15* gene expression as well as the frequency of GPR15+ lymphocytes among CD3+ T cells and CD19+ B cells were measured in white blood cells. The strongest correlation was found between *GPR15* gene expression and the proportion of CD3+GPR15+ T cells ($R^2 = 0.72$, $p = 8.4^{-28}$). The amount of CD19+GPR15+ B cells was only marginally correlated with the *GPR15* gene expression in WBC ($R^2 = 0.12$, $p = 5.0^{-4}$), Fig. 3.

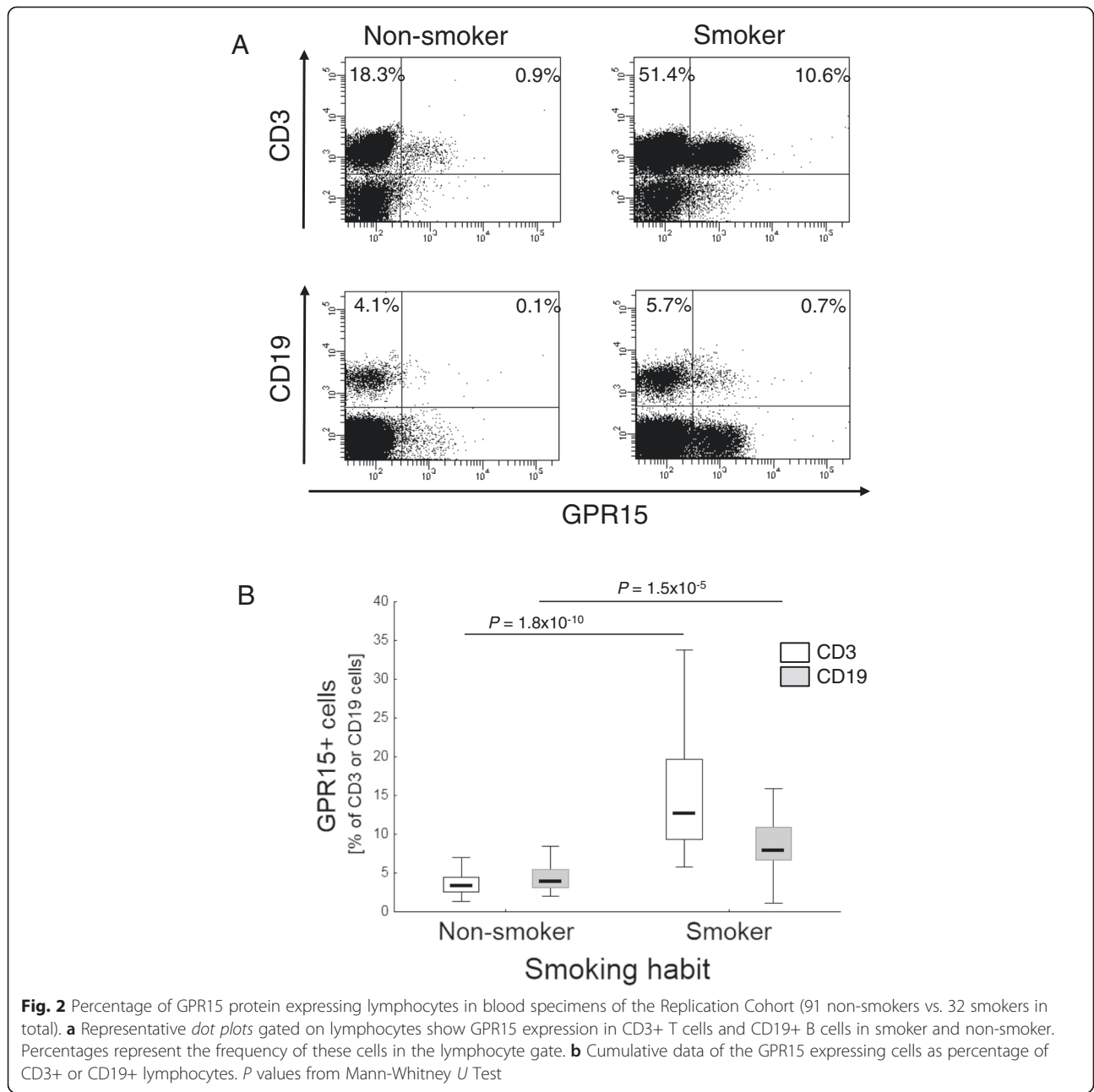## Replication of *GPR15* methylation differences in flow cytometric-sorted cells

Analysis of methylation of CpG site cg19859270 located within the *GPR15* gene was performed in isolated PBMCs, in flow cytometric-sorted CD3+GPR15+ as well as CD3+GPR15- T cells of smokers ($n = 6$) and non-smokers ($n = 6$). At the PBMC level, a methylation difference of 3.0 % ($p = 0.009$) between smoker and non-smoker was observed as expected (Fig. 4). A hypomethylation at cg19859270 was specific for GPR15 expressing cells independent on smoking habit. The methylation difference between GPR15– and GPR15+ T cells was 49.5 % in smoker ($p = 0.005$) and similar in non-smoker (38 %, $p = 0.005$). Within the CD3+GPR15+ population, the hypomethylation of cg19859270 site was slightly more pronounced in smokers than non-smokers (delta methylation = –15.0 %, $p = 0.029$). Due to the low frequency of GPR15+ B cells, sorting for this cell type was not performed.

Allele frequencies of SNPs within the *GPR15* gene

Although the protein expression of GPR15 was only detectable when the CpG site cg19859270 was markedly hypomethylated, the hypomethylation in GPR15 expressing cells reached only values of about 50 %. To exclude an underlying imprinting effect at this CpG site, we analyzed the occurrence of alleles at two single nucleotide polymorphisms (SNP, rs3749260, rs2230344), located within the gene body of *GPR15*, in genomic DNA as well as *GPR15* transcript in WBCs of randomly selected participants of the Replication Cohort (non-smoker and smoker, each $n = 18$). Since both gDNA and *GPR15* transcripts have shown identical proportions of alleles in each subject, the imprinting at cg19859270 may be excluded, and thus, a random monoallelic expression of GPR15 can be supposed (Table 2).

Association between methylation, smoking behavior, and lymphocyte cell subpopulations

In our study, as well as in published data, a significant hypomethylation at cg19859270 within the *GPR15* gene has been observed in smoker at WBC or PBMC level. Thus, an apparent smoking-induced hypomethylation is

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 4 of 11

Fig. 2 Percentage of GPR15 protein expressing lymphocytes in blood specimens of the Replication Cohort (91 non-smokers vs. 32 smokers in total). **a** Representative *dot plots* gated on lymphocytes show GPR15 expression in CD3+ T cells and CD19+ B cells in smoker and non-smoker. Percentages represent the frequency of these cells in the lymphocyte gate. **b** Cumulative data of the GPR15 expressing cells as percentage of CD3+ or CD19+ lymphocytes. *P* values from Mann-Whitney *U* Test

expectable. However, we clearly show that in smokers, the proportion of GPR15 expressing cells is many times higher than in non-smoker, implying that the observed hypomethylation at WBC or PBMC level is determined by the proportions of these cells in peripheral blood. By using a linear regression model we confirmed this observation, assessing the impact of smoking, age, gender, the proportion of CD3+ and frequency of GPR15+ cells among CD3+ T cells on methylation at cg19859270 in PBMCs (Table 3). In univariate analysis smoking ($p = 0.011$) as well as the frequency of CD3+GPR15+ T cells ($p = 0.007$) show significant association to methylation at

cg19859270 (Methyl-PBMC), whereas age, gender and the proportion of CD3+ T cells did not influence this association. However, in the multiple regression model, the smoking-dependent association with methylation at cg19859270 was lost by adjustment to the proportion of CD3+GPR15+ T cells ($p = 0.904$).

Stimulation of PBMCs with CSE in vitro

To identify any causative role of active cigarette smoking to the excess of GPR15+ cells in blood, we investigated the impact of different CSE concentrations on PBMCs isolated from smokers and non-
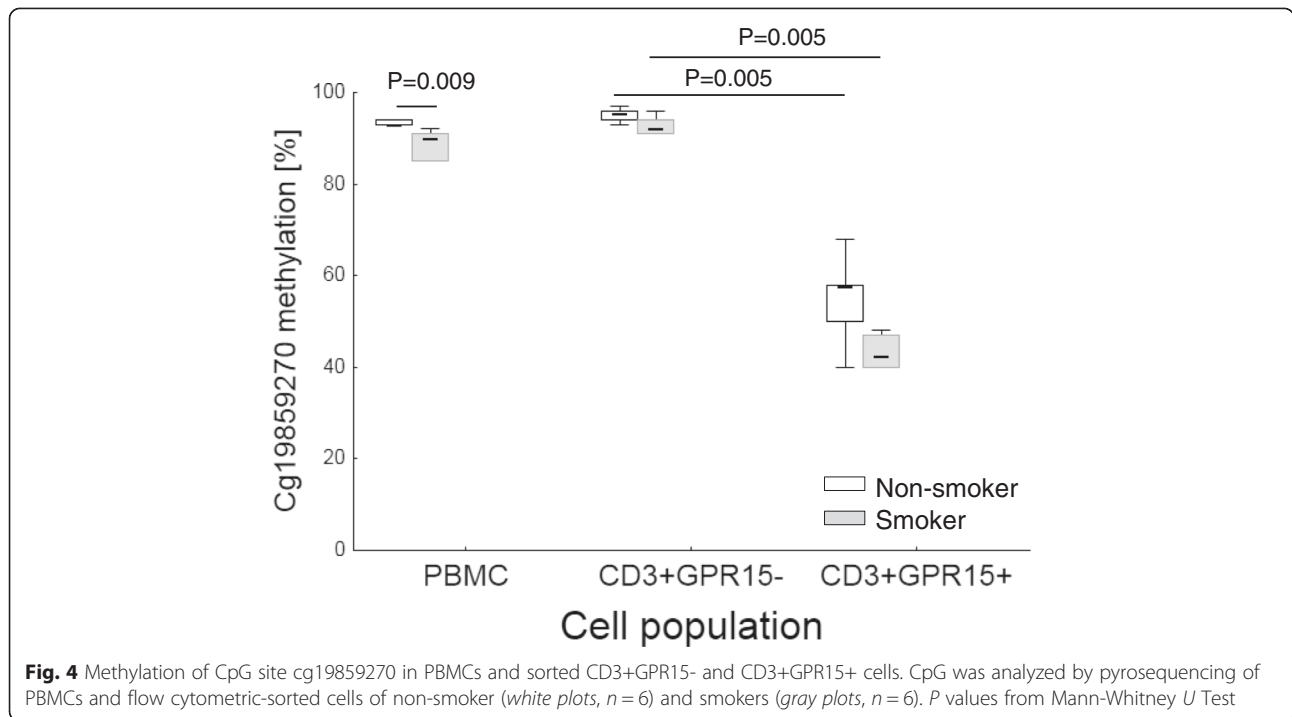
Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 5 of 11
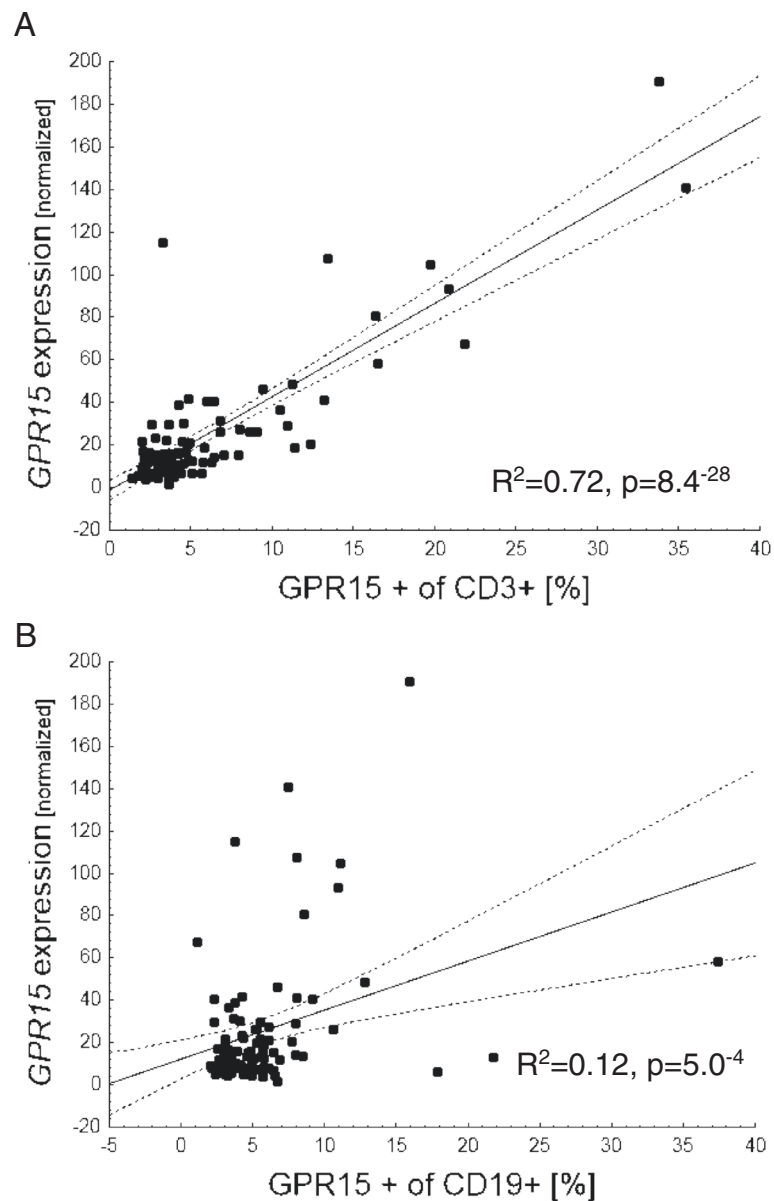
Fig. 3 *GPR15* gene expression in whole blood versus frequency of GPR15 expressing T lymphocytes. Percentage of GPR15+ of CD3+ T cells (**a**) and of CD19+ B cells (**b**). Correlation coefficient from linear regression analysis

smokers (Fig. 5). The CSE did not increase the proportion of GPR15-expressing cells.

Discussion

In this study, we provide the first evidence that reported tobacco smoke induced methylation changes at single CpG site in DNA of WBC is mainly seen due to an increased proportion of specialized cell subtypes in blood rather than by direct impact of tobacco smoke on DNA methylation. Here, we show that only the GPR15-expressing cells are hypomethylated at cg19859270, located within the *GPR15* gene body, and thus, the observed hypomethylation at this CpG site [1–7] in smokers is the consequence of a higher proportion of GPR15 expressing cells being present in the peripheral blood. In two independent cohorts, we confirmed the differential expression of *GPR15* at the RNA level in WBC between smoker and non-smoker and continued the validation at the cellular protein level. By analyzing the GPR15 protein expression in WBC subtypes, it was evident that the proportion of CD3+GPR15+ T cells and to a lesser extent the CD19+GPR15+ B cells were responsible for the apparent smoking-induced hypomethylation in the *GPR15* gene since cg19859270 hypomethylation was

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 6 of 11

Fig. 4 Methylation of CpG site cg19859270 in PBMCs and sorted CD3+GPR15- and CD3+GPR15+ cells. CpG was analyzed by pyrosequencing of PBMCs and flow cytometric-sorted cells of non-smoker (*white plots*, n = 6) and smokers (*gray plots*, n = 6). *P* values from Mann-Whitney *U* Test

specifically found in GPR15 expressing cells. Our results highlight the importance of accounting not only for main cell composition but also for cell subtypes in EWAS performed in whole blood or white blood cells.

Recent EWAS have indicated a potential role of epigenetic changes in the etiology of human diseases or in relation to environmental exposure. However, when performed on whole blood or in general on tissue samples, one major challenge is to distinguish true epigenetic variation from epigenetic changes caused by differences in cellular composition between cases and controls [17]. For example, during aging, the cellular composition of blood is often altered [18, 19], and thus, an apparent epigenetic variation may be the result of age specific

profiles for different blood cell subsets. On the other hand, the level of methylation *per se* varies across the life course, declining through aging [20, 21]. Thus, adjustment for age and cell composition is crucial in EWAS. Information regarding subject age is easily accessible; however, in contrast determination of whole blood cell composition at a cellular level is more challenging.

Recently, a set of statistical methods have been published aimed at overcoming cell composition bias in EWAS. Beside reference-free methods [22, 23], the majority of methods for whole blood are based on existing

Table 2 Allele frequencies of two single nucleotide polymorphisms (rs2230344, rs3749260) located within *GPR15* gene at genomic and transcript level

Allele	gDNA	*GPR15* transcript
	(n)	
rs2230344		
CC	25	25
CT	11	11
TT	0	0
rs3749260		
CC	27	27
CA	7	7
AA	2	2

Table 3 Correlation of DNA methylation at cg19859270 in PBMCs and frequency of GPR15 expressing CD3 lymphocytes in non-smoker versus smoker

	Methyl-PBMC		
	Beta	SE	*P* value
Univariate analysis			
Smoking habit	−4.17	1.32	0.011
Gender	−1.51	1.83	0.428
Age	0.03	0.05	0.558
CD3+	0.11	0.07	0.171
GPR15+ of CD3+	−0.35	0.10	0.007
Multiple regression			
~Smoking habit[a]	−0.71	0.33	0.008
~Smoking habit[b]	0.53	4.19	0.904
~GPR15+ of CD3+[a]	−0.38	0.09	0.003

[a]Adjusted to gender, age, and CD3+
[b]Adjusted to gender, age, CD3+, and GPR15+ of CD3+

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 7 of 11
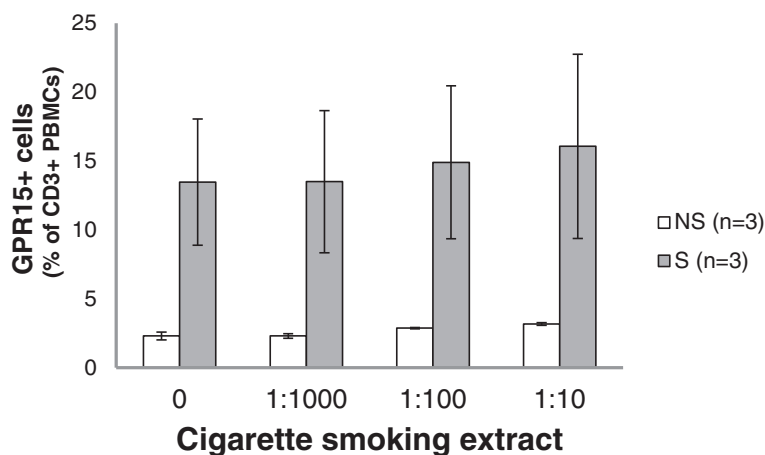
Fig. 5 Influence of cigarette smoke extract (CSE) on GPR15 expression in PBMCs of randomly selected non-smokers (NS, *n* = 3) and smokers (S, *n* = 3). PBMCs were exposed in vitro for 5 days

reference database of sorted blood cells by using differentially methylated loci across major leukocyte types [8, 14, 18]. For example, the reference set provided by Houseman et al. [14] distinguish granulocytes from other cell types and CD4+ and CD8+ T cells from other cells that are not lymphocytes. However, it does not differentiate T cell subtypes like Th1, Th2, regulatory T cells, or memory T cells. Thus, an adjustment for the proportion of blood cells in whole blood is possible only at the level of cell lineage markers. Unmeasured subpopulations or activated versions of measured cell types with impact on the variable of interest are not accounted for in this case. In our study we identified a T cell subtype expressing the orphan chemoattractant receptor GPR15 which is specifically enriched in active smokers. Starting at the level of white blood cells, we identified GPR15 as the most differentially expressed gene in smokers vs. non-smokers in two independent cohorts. Former smokers showed an intermediate level of expression of GPR15 (Fig. 1). These data are in line with results in recently published studies [24, 5]. Tsaprouni et al. observed a correlation between DNA hypomethylation in cg19859270 and GPR15 gene expression in white blood cells of smokers indicating that cigarette smoking reduces DNA methylation [5]. However, GPR15 protein expression at the cellular level has not been analyzed yet in relation to smoking status. Therefore, in order to further validate our findings regarding methylation and gene expression levels, we performed flow cytometric analyses for GPR15 protein expression in WBCs and PBMCs in our replication cohort. We found that CD3+ T cells were the main population expressing GPR15. The proportion of these cells was nearly 5-fold increased in smokers. To a less extent, CD19+ B cells were responsible for the GPR15 expression in whole blood (Fig. 2).

Without further characterization of cell subtypes, it has long been known that smoking changes leukocyte count in whole blood [25, 26]. Here, we demonstrate that within leukocytes, only a particular immune subset was associated with smoking, this subset being enriched in the peripheral blood of smokers. To our knowledge, to date, GPR15-expressing T cells seem to be the strongest cellular determinant in whole blood in discriminating individually active smokers from non-smokers. Based on the frequency of GPR15-bearing CD3+ T cells, a specificity of 0.99 % was reached by flow cytometric analysis of GRP15 expression.

The frequency of these cells was strongly correlated with *GPR15* gene expression at the WBC level, indicating that *GPR15* gene expression detectable in blood is mainly a result of the expression in CD3+GPR15+ T cells. The lower correlation of *GPR15* gene expression in blood with the frequency of CD19+GPR15+ B cells may be explained by the fact that only a small proportion of cells in whole blood are B cells (1.5 %).

To find out the origin of methylation changes at cg19859270 due to smoking habit, we investigated the methylation at this CpG site in PBMCs as well as sorted CD3+GPR15- and CD3+GPR15+ cells from smokers and non-smokers. Here, for the first time, we show that the hypomethylation at cg19859270 within the *GPR15* gene is strongly associated with the GPR15 protein expression and is independent of smoking habit. Irrespective of whether the isolated CD3+GPR15+ cells originated from smokers or non-smokers, these cells carry about 50 % methylation at cg19859270 compared to a greater than 90 % methylation in the CD3+GPR15- population in each WBC sample of blood donor, leading to the assumption of a monoallelic expression of GPR15. Besides this GPR15-associated methylation pattern at cg19859270, the hypomethylation in CD3+GPR15+ cells was slightly more

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 8 of 11

pronounced in smokers (–15 %). The reason for this remains speculative and warrants further investigations which are in focus of future studies.

The apparent smoking-dependent observed difference in cg19859270 methylation at the PBMC or WBC level to an extent seen by ourselves and others of 2–4 % [1–7] seems to be determined by the enrichment of specific cg19859270 hypomethylated cell population in smokers. To statistically evaluate this observation, we analyzed the impact of smoking habit on cg19859270 methylation at the PBMC level by using a multiple linear regression model. In univariate analyses, smoking habit and the proportion of CD3+GPR15+ T cells are both independently associated with PBMC methylation. Neither gender, age, nor the proportion of CD3+ T cells had an influence on PBMC methylation. However, when the model with the PBMC methylation as the outcome and smoking habit as predictor was adjusted, besides age and gender, to the proportion of CD3+GPR15+ expressing cells, the positive association between smoking and cg19859270 methylation lost significance.

Conclusion

With these findings, we want to point out, firstly, the eminent importance of cell subtype adjustment in EWAS. As we show here, an adjustment only at the level of known cell lineage markers [4, 5, 7] does not overcome the cell composition bias in general. Thus, for EWAS, it is strongly recommended to exclude a cell type specificity for all significant findings.

Secondly, we want to point out, that the smoking-induced hypomethylation at cg19859270 in WBC is not caused by direct action of smoking-related soluble compounds. In vitro experiments with cigarette smoking extract (CSE) did not increase the proportion of these GPR15-bearing cells, indicating that CSE is not responsible for the hypomethylation at CpG site cg19859270 located in the gene body of GPR15. To reiterate, these result underline that there is no causative effect of cigarette smoking on DNA methylation observed at the WBC or PBMC level at least for the cg19859270 locus. How tobacco smoke may increase the proportion of GPR15-expressing T cells in the peripheral blood of active smokers remain to be elucidated and requires in vivo studies which were not performed in the present investigation, representing a limitation of our study. However, excluding a direct action of CSE on the proportion of GPR15-expressing T cell in vitro, we favor the hypothesis of a complex immunological cascade toward tobacco-smoking-induced disturbance of tissue homeostasis including the interaction of antigen-presenting cells. One other limitation of our study is that we validated only one, GPR15, of the genes whose expression was differentially expressed in smokers compared to non-smokers.

Thus, although we cannot generalize our findings to other CpGs or other environmental exposures, it might serve as an example for the misleading situation when differences at the methylation level, even after replication at the gene expression level, lead to false assumptions especially when further validation at the protein and cellular level is not performed.

We suggest that even though in many EWAS performed in whole blood an adjustment for cellular composition has been made, these corrected descriptive results should be interpreted with caution when the cell subtype expressing the differentially methylated gene is not identified.

Methods

Subjects

For the present study, data from two independent human cohorts were used (Table 4). A workplace selected cohort comprising 107 volunteers employing in industrial duck production (Working Place Cohort) located in the federal state Brandenburg (Germany). This study was designed to assess the impact of workplace dust on human health. Blood samples were collected in PAXgene Blood RNA Tube (Qiagen, Hilden, Germany) at one working day before and after work. All participants gave written informed consent. The study was approved by the ethics commission of the medical association of Berlin (eth-013/07). The second cohort, used as replication cohort, comprised 123 randomly selected pseudonymous blood samples from healthy volunteers. These samples consisted of 100 venous whole blood samples as well as 23 buffy coats, obtained from the Institute of Transfusion Medicine at the University of Leipzig, Germany. All volunteers were HIV-tested negative and gave written informed consent. The study was approved by Ethics Committees of the University of Leipzig (079-15-09032015). In both cohorts, smoking behavior, age, and gender were recorded via questionnaires.

Sample preparation

Total RNA was prepared from blood samples by PAXgene Blood RNA Kit (Qiagen, Hilden, Germany) and from isolated PBMCs from buffy coat by using peqGold RNA Pure (peqlab, Erlangen, Germany), according to manufacturer's instructions. The cDNA synthesis was carried out with 1 μg of RNA by using ImProm-II™ Reverse Transcription System (Promega, Mannheim, Germany).

Gene expression analysis

Working Place Cohort

In the *Working Place Cohort*, multiple genes (159 genes, see Additional file 1) were analyzed by 96.96 Dynamic Array (Fluidigm, San Francisco, CA, USA). Intron-spanning primers were designed, and UPL probes were selected by the Universal Probe Library Assay Design Center (http://qpcr.probefinder.com/organism.jsp). A

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 9 of 11

Table 4 Description of human specimens

	Male	Female	Cell population	Expression
Working Place Cohort			WBC[b]	Gene
Gender	52	41		
Age[a]	44 (24–72)	43.5 (21–70)		
Tobacco smoking				
Non-smoker	20	11		
Formerly smoker	9	11		
Smoker				
<10 cigarettes/day	4	5		
>10 cigarettes/day	19	14		
Replication cohort			WBC[b]	Gene, protein
WBC				
Gender	53	47		
Age[a]	44 (21–47)	46 (19–66)		
Tobacco smoking				
No	41	41		
Yes	12	6		
Buffy coat			PBMC[c]	gDNA-methylation, protein
Gender	14	9		
Age[a]	36 (24–64)	39 (21–71)		
Tobacco smoking				
No	6	3		
Yes	8	6		

[a]Median (min-max)
[b]White blood cells
[c]Peripheral blood mononuclear cells

preamplification reaction was performed by pooling all primers (final concentration, 50 nM), 5 μl of cDNA, and 2X PreAmp Master Mix (Applied Biosystems/Life Technologies GmbH, Darmstadt, Germany). The cycling program consisted of 95 °C for 10 min, followed by 14 cycles of 95 °C for 15 s, and 60 °C for 4 min on a LightCycler 480 (Roche Applied Science, Mannheim, Germany). The qPCRs of 1:5 diluted with TE buffer preamplified templates were performed following manufacturer's instruction for UPL (Roche, Mannheim, Germany) assays. Briefly, for each individual assay, a 10X Assay Mix that contained 2 μM of each forward and reverse primers, 1 μM UPL probe and 0.025 % Tween-20 were prepared, and 5 μl of the mix was loaded into the assay inlets of the array. Into the sample inlets, 5 μl of the following solution was dispensed: 2.5 μl of PreAmp sample in 1.1X of FastStart Universal Probe Master Mix (Roche, Mannheim, Germany). The cycling program consisted of 2 min at 50 °C, 10 min at 95 °C, followed by 35 cycles of 95 °C for 15 s, 1 min at 60 °C, and 70 °C for 5 s. All reactions were performed in triplicates. Gene expression values were determined by using the $2^{-\Delta\Delta CT}$

method [27] and GAPD and GUSB as reference genes and normalized to the lowest measured value.

Replication Cohort
Specific single gene expressions were performed for GPR15 (primer-for 5′-tggctgcccttcaatacttt, -rev 5′-tag tgttcttgccgcaacc, UPL 72) and reference genes GAPD and GUSB (primer-for 5′-gctctctgctcctcctgttc, -rev 5′-acgaccaaatccgttgactc, UPL 60; -for 5′-cgccctgcctatctg-tattc, -rev 5′-tccccacagggagtgtgtag, UPL 57) with identical PCR-mix as described above in 384-well format on a LightCycler 480 cycler (Roche, Mannheim, Germany).

Analysis of protein expression at cellular level
The translation of differentially expressed gene *GPR15* in smoker vs. non-smoker was analyzed at cellular level via flow cytometry. In the *Replication Cohort*, 100 μl of whole blood was incubated with mouse-anti-human-GPR15 antibody (1:500; R&D Systems, Wiesbaden-Nordenstadt, Germany) supplemented with 5 % goat serum for 1 h. After washing in PBS/1% fetal calf serum (FCS), the GPR15 was stained with R-phycoerythrin-labeled goat-anti-mouse IgG2b (1:500, 1 h; Biozol, Eching,

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 10 of 11

Germany) following an additional wash step. Thereafter, cells were incubated with 5 % mouse serum for 30 min following double-staining step for leucocyte differentiation marker (1 h; anti-CD3-FITC (Beckman Coulter, Krefeld, Germany)), -CD4-BV510, -CD8-PerCP, and -CD19-APC-H7 (BD Biosciences, Heidelberg, Germany). At the end, erythrocytes were lysed in FACS Lysing solution (BD Bioscience, Heidelberg, Germany) according to manufacturer's instruction immediately before measurement.

Isolated PBMCs from buffy coats were stained for GPR15 and thereafter for the other surface receptors according to the above mentioned method but without erythrocyte lysing step. All measurements were performed on a FACS Canto II and analyzed with the BD FACSDIVA software (version 8.0.1, BD Biosciences, Heidelberg, Germany).

Analysis of methylation differences in flow-sorted cells
For GPR15-specific cell sorting, PBMCs were isolated from buffy coats and stained as mentioned above. CD3+GPR15+ cell populations were sorted with purity greater than 99 %. Flow cytometric cell sorting was performed at the laboratory of cytometry of the Core Facility at the University of Leipzig.

Pyrosequencing
Genomic DNA was extracted from PBMC by using the Blood DNA extraction kit according to the manufacturer's protocol (Qiagen, Hilden, Germany). DNA bisulfite treatment was performed using the Epitect kit (Qiagen) according to manufacturer's instruction. Samples were immediately stored at –20 °C and thereafter simultaneously analyzed by pyrosequencing. Methylation assays were designed using the PyroMark Assay Design Software 2.0 (www.qiagen.com). Primer sequences for pyrosequencing are indicated in Additional file 4. Methylation levels for the CpG site were assessed using Pyromark Q24 pyrosequencer (Qiagen).

Validation of smoking behavior—Cotinine ELISA
Smoking behavior in the *Replication Cohort* was validated by cotinine measurements. The cotinine concentration in plasma of blood or blood buffy coat samples was measured using the Cotinine direct ELISA Kit according to manufacturer's instruction (DRG Instruments GmbH, Marburg, Germany). Cotinine level exceeding the sensitivity level of the assay (1 ng/ml) was considered as smoker.

Preparation of aqueous cigarette smoke extract (CSE)
CSE was prepared according to the protocol described by Adenuga and co-workers [28]. Briefly, research-grade reference cigarettes (3R4F) from the University of Kentucky (Tobacco Health Research, Lexington, KY, USA) were used to prepare cigarette smoke extract (CSE) by slowly bubbling smoke from one cigarette into 10 ml of RPMI 1640 without supplements at a rate of 1 cigarette/min. Afterwards, CSE was sterile-filtered through a 0.22-μm filter (Sartorius, Göttingen, Germany).

In vitro exposure of PBMCs
Peripheral blood mononuclear cells (PBMCs) from blood or from blood buffy coat were obtained by density-gradient centrifugation using Ficoll-Paque (GE Healthcare, Berlin, Germany). After separation cells were immediately cultured or cryopreserved and stored in liquid nitrogen. PBMCs (5×10^5 cells/well) were held in vitro culture in 96-well round-bottom plates with RPMI 1640 medium supplemented with 10 % FCS, glutamine, and 25 mM HEPES (Life Technologies, Darmstadt, Germany) without antibiotics. CSE was added in concentrations 1:1000, 1:100, and 1:10. After 5 days, cells were harvested and analyzed by flow cytometry as mentioned before.

Statistical analysis
Statistical significance of parametric distributed values was calculated with unpaired Student's t test. Otherwise, the nonparametric Mann-Whitney U test was applied for comparison of dependent and independent variables, respectively (Statistica for Windows version 10, StatSoft Inc., Europe). Boxes in figures indicate the 25 and 75 % percentiles, whiskers the non-outlier range and dots the outlier. All p values <0.05 were considered to be significant. In analyzing gene expression data (159 genes) from high-throughput approach, the Bonferroni correction for multiple testing ($p < 0.00031$) was considered to be significant (Additional file 1). The sensitivity and specificity as statistical measures of performance of GPR15-dependent discrimination between smokers and non-smokers were calculated.

To test the relationship between smoking and DNA methylation, we used a multiple linear regression model with methylation at PBMC level as outcome (Methyl-PBMC) and smoking habit as predictor. The model was further adjusted to age, gender, and the proportion of CD3+ T cells or CD3+GPR15+ T cells. These possible confounders for Methyl-PBMC were tested also in univariate analyses.

Additional files

Additional file 1: A table listing "Comparison of gene expression (*n*=159) in smokers (s) versus non-smokers (ns) of the Work Place Cohort. Differences in gene expression were estimated by t-test. Significant values after Bonferroni correction for multiple testing (p<0.00031) are held in bold".

Additional file 2: A figure showing "GPR15 gene expression versus circadian rhythm. For the Working Place Cohort, peripheral blood was collected before (morning) and after (afternoon) work".

Bauer *et al. Clinical Epigenetics* (2015) 7:81

Page 11 of 11

Additional file 3: A figure showing "Representative dot plots for the gating strategy for isotype control and GPR15 staining. Gated lymphocytes were further characterized for the expression of CD3 and CD19.

Additional file 4: A table listing "Primer sequences used for pyrosequencing".

Abbreviations
ELISA: enzyme-linked immunosorbent assay; WBC: whole blood cells; PBMC: peripheral blood mononuclear cells; CSE: cigarette smoke extract.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MB and GH wrote the manuscript. GH edited the manuscript. MB, AMH, ME, KO, HK, BF, and US designed and/or performed the experiments. GL designed the Working Place Cohort. All authors read and approved the final manuscript.

Author details
[1]Department of Environmental Immunology, Helmholtz Centre for Environmental Research-UFZ, Leipzig 04318, Germany. [2]Federal Institute for Occupational Safety and Health, Berlin 10317, Germany. [3]Division of Endocrinology and Nephrology, University of Leipzig, Leipzig 04103, Germany. [4]Institute of Clinical Immunology, Medical Faculty, University of Leipzig, Leipzig 04103, Germany. [5]Nikolaus-Fiebiger-Zentrum for Molecular Medicine, Institute of Genetics, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen 91054, Germany.

References
1. 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.
2. 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.
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.
4. Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, et al. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. Hum Genet. 2013;132(9):1027–37.
5. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics. 2014;9(10):1382–96.
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.
7. 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.
8. Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. Genome Biol. 2014;15(3):R50.
9. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, et al. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS One. 2012;7(10):e46705.
10. Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet. 2012;13(10):705–19.
11. Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, et al. Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. Epigenetics. 2013;8(8):816–26.
12. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, et al. Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci U S A. 2012;109 Suppl 2:17253–60.
13. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7(7):e41361.
14. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
15. Hinks TS, Zhou X, Staples KJ, Dimitrov BD, Manta A, Petrossian T et al. Innate and adaptive T cells in asthmatic patients: relationship to severity and disease mechanisms. J Allergy Clin Immunol. 2015. doi:10.1016/j.jaci.2015.01.014.
16. Serriari NE, Eoche M, Lamotte L, Lion J, Fumery M, Marcelo P, et al. Innate mucosal-associated invariant T (MAIT) cells are activated in inflammatory bowel diseases. Clin Exp Immunol. 2014;176(2):266–74.
17. Lowe R, Rakyan VK. Correcting for cell-type composition bias in epigenome-wide association studies. Genome Med. 2014;6(3):23.
18. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15(2):R31.
19. Simone R, Zicca A, Saverino D. The frequency of regulatory CD3 + CD8 + CD28- CD25+ T lymphocytes in human peripheral blood increases with age. J Leukoc Biol. 2008;84(6):1454–61.
20. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. 2012;8(4):e1002629.
21. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, et al. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. Mech Ageing Dev. 2009;130(4):234–9.
22. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics. 2014;30(10):1431–9.
23. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. Nat Methods. 2014;11(3):309–11.
24. Paul S, Amundson SA. Differential effect of active smoking on gene expression in male and female smokers. J Carcinog Mutagen. 2014;5. doi:10.4172/2157-2518.1000198.
25. Petitti DB, Kipp H. The leukocyte count: associations with intensity of smoking and persistence of effect after quitting. Am J Epidemiol. 1986;123(1):89–95.
26. Friedman GD, Siegelaub AB, Seltzer CC, Feldman R, Collen MF. Smoking habits and the leukocyte count. Arch Environ Health. 1973;26(3):137–43.
27. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402–8.
28. Adenuga D, Yao H, March TH, Seagrave J, Rahman I. Histone deacetylase 2 is phosphorylated, ubiquitinated, and degraded by cigarette smoke. Am J Respir Cell Mol Biol. 2009;40(4):464–73.