Genome-wide DNA methylation profiles in smoking discordant and concordant monozygotic twin pairs

Jenny van Dongen[1,2,3], Gonneke Willemsen[1,2], BIOS Consortium[4], Eco J.C. de Geus[1,2], Dorret I. Boomsma[1,2,3], Michael C. Neale[5]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

[3]Amsterdam Reproduction and Development (AR&D) Research Institute, Amsterdam, The Netherlands

[4]Biobank-based Integrative Omics Study Consortium

[5]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, USA

Corresponding author: Jenny van Dongen, Ph.D., Department of Biological Psychology, Vrije Universiteit Amsterdam, van der Boechorststraat 7, 1081 BT Amsterdam, Amsterdam, Netherlands, tel: +31 (20) 5983570, j.van.dongen@vu.nl .

Abstract

Smoking-associated DNA methylation levels identified through epigenome-wide association studies (EWAS) are generally ascribed to smoking-reactive mechanisms, but the contribution of a shared genetic predisposition to smoking and DNA methylation levels is typically not accounted for. We exploited a strong within-family design, i.e., the discordant monozygotic twin design, to study reactiveness of DNA methylation in blood cells to smoking and reversibility of methylation patterns upon quitting smoking. Illumina HumanMethylation450 BeadChip data were available for 769 monozygotic twin pairs (mean age=36 years,range=18-78, 70% female), including pairs discordant or concordant for current or former smoking. In pairs discordant for current smoking, 13 differentially methylated CpGs were found between current smoking twins and their genetically identical co-twin who never smoked. Top sites include multiple CpGs in *CACNA1D and GNG12,* which encode subunits of a calcium voltage-gated channel and G protein, respectively. These proteins interact with the nicotinic acetylcholine receptor, suggesting that methylation levels at these CpGs might be reactive to nicotine exposure. All 13 CpGs have been previously associated with smoking in unrelated individuals and data from monozygotic pairs discordant for former smoking indicated that methylation patterns are to a large extent reversible upon smoking cessation. We further showed that differences in smoking level exposure for monozygotic twins who are both current smokers but differ in the number of cigarettes they smoke are reflected in their DNA methylation profiles. In conclusion, by analysing data from monozygotic twins, we robustly demonstrate that DNA methylation level in human blood cells is reactive to cigarette smoking.

Key words: Cigarettes, EWAS, twin study, addiction, causality, epigenetic, smoke, Netherlands Twin Register

Background

Epigenome-wide association studies (EWAS) have identified robust differences in DNA methylation between smokers and non-smokers[1]. In a meta-analysis of blood-based DNA methylation studies (N=15,907 individuals; the largest EWAS of smoking to date), 2,623 CpG sites passed the Bonferroni threshold for genome-wide significance in a comparison of current and never smokers[2]. Based on comparison with loci identified in large genome-wide association studies (GWAS), differentially methylated sites were significantly enriched in genes implicated in well-established smoking-associated diseases, such as cancer, cardiovascular disease, inflammatory disease and lung disease, as well as in genes associated with schizophrenia and educational attainment[2]. It has been hypothesized that smoking-induced methylation changes might also contribute to the addictive effect of smoking[3].

Importantly, smoking-associated DNA methylation levels, as established in human EWA studies, may reflect different mechanisms. They may reflect causal effects of smoking on methylation, causal effects of methylation on smoking behaviour, methylation differences associated with epiphenomena of other exposures that correlate with smoking (e.g. alcohol use[4]), or they may reflect a shared genetic predisposition to smoking and methylation level. To distinguish these different mechanisms requires incisive study designs[5]. Establishing whether methylation levels in smokers revert to levels of never smokers upon smoking cessation is a first step. A previous study of 2,648 former smokers with cross-sectional methylation data from the Framingham Heart Study suggested that methylation levels at most CpGs return to the level of never smokers within five years after quitting smoking, but 36 CpGs were still differentially methylated in former smokers, who had quit smoking for 30 years[2]. In the large EWAS meta-analysis of smoking[2], 185 CpGs were differentially methylated between former and never smokers (compared to 2623 between current and never smokers). In addition, differences between former and never smokers were smaller than between current and never smokers. Reversible DNA methylation patterns may suggest

that DNA methylation is reactive to smoking. However, it is also possible that the different methylation level in current smokers reflects a higher genetic liability to smoking behavior (that makes them more likely to initiate and keep smoking). Similarly, differences between former smokers and never smokers could reflect that smoking has caused a persistent methylation change but can also be driven by genetic factors.

In population-based studies, smoking cases and non-smoking individuals may differ on many aspects, including their genetic predisposition to smoking. On the other hand, monozygotic twins are genetically identical (except for *de novo* mutations, but these are rare[6,7]), and are matched on sex, age and early environment. They have been exposed to similar prenatal conditions, which may include second hand smoke from smoking mothers and others. Therefore, smoking discordant monozygotic twin pairs offer a unique opportunity to assess smoking-reactive DNA methylation patterns[5,8]. Despite the large number of previous population-based smoking EWASs, only one previous study compared genome-wide DNA methylation in smoking discordant monozygotic twin pairs[9]. This study analysed whole blood Illumina 450k array methylation data from 20 discordant pairs, and reported 22 top loci, many of which had been previously associated with cigarette smoking in previous studies. However, following the correction for multiple testing, none of the differentially methylated loci were statistically significant, and this previous twin study did not examine reversibility of smoking effects, i.e., where methylation status changes again following smoking cessation.

Here, we analyse unique data from a large cohort of monozygotic twin pairs. This cohort is sufficiently large to include current smoking discordant and concordant pairs, as well as pairs discordant for former smoking. These groups allow identification of loci that are reactive to smoking, and examination of the extent to which effects are reversible upon quitting smoking. Monozygotic pairs in which both twins are current smokers, but who differ in quantity smoked, enable examination of the effects of smoking intensity. Finally, concordant pairs who never smoked allow assessment of the amount of DNA methylation variation at smoking-reactive loci that is due to non-genetic sources of variation other than smoking. In

secondary enrichment analyses, we examined whether smoking-reactive methylation patterns are enriched 1) at loci detected in previous epigenome-wide association studies of other traits and exposures, 2) at loci detected in a previous large Genome-wide Association Study meta-analysis of smoking initiation[10] – these loci are presumed to have a causal effect on smoking behaviour, and 3) within Gene Ontology and Kegg Pathways. Finally, we examined the relationship between DNA methylation and RNA transcript levels in blood for smoking-reactive loci.

Methods

Participants

In the Netherlands Twin Register (NTR), DNA methylation data are available for 3087 whole blood samples from 3055 individuals in twin families, as described in detail previously[11]. The samples were obtained from twins and family members, who participated in NTR longitudinal survey studies[12] and the NTR biobank project[13]. In the current study, methylation data from monozygotic twin pairs were analysed. Among 768 monozygotic twin pairs with genome-wide methylation data and information on smoking and covariates, we identified the following discordant pairs: 53 discordant pairs, in which one twin was a current smoker at blood draw and the other never smoked, 72 discordant pairs, in which one twin was a former smoker at blood draw and the other never smoked, 66 discordant pairs of which one twin was a former smoker and the other a current smoker at blood draw. In addition, we identified the following concordant pairs: 83 twin pairs concordant for current smoking, 88 twin pairs concordant for former smoking, and 406 concordant twin pairs who never smoked. Informed consent was obtained from all participants. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute codes, NTR 03-180).

Peripheral blood DNA methylation and cell counts

Genome-wide DNA methylation in whole blood was measured by the Human Genomics facility (HugeF) of ErasmusMC, the Netherlands (http://www.glimdna.org/). DNA methylation was assessed with the Infinium HumanMethylation450 BeadChip Kit (Illumina, San Diego, CA, USA). Genomic DNA (500 ng) from whole blood was bisulfite treated using the Zymo EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA), and 4 µl of bisulfite-converted DNA was measured on the Illumina 450k array[14] following the manufacturer's

protocol. A custom pipeline for quality control and normalization of the methylation data was developed by the BIOS consortium. First, sample quality control was performed using MethylAid[15]. Next, probe filtering was applied with DNAmArray[16] to remove: ambiguously mapped probes[17], probes with a detection *P* value > 0.01, or bead number < 3, or raw signal intensity of zero. After these probe filtering steps, probes and samples with a success rate < 95% were removed. Next, the DNA methylation data were normalized using functional normalization[18], as implemented in DNAmArray[16] using the cohort-specific optimum number of control probe-based principal components. Probes containing a SNP, identified in a DNA sequencing project in the Dutch population[19], within the CpG site (at the C or G position) were excluded irrespective of minor allele frequency, and only autosomal probes were analysed, leading to a total number of 411,169 methylation sites. The following subtypes of white blood cells were counted in blood samples: neutrophils, lymphocytes, monocytes, eosinophils, and basophils[13].

Smoking

Information on smoking behavior was obtained by interview during the home visit for blood collection as part of the NTR biobank project (2004–2008 and 2010-2011). Participants were asked: “Did you ever smoke?”, with answer categories: (1) no, I never smoked (2) I’m a former smoker (3) yes. Current smokers were asked how many years they smoked and how many cigarettes per day they smoked, while ex-smokers were asked how many years ago they quit, for how many years they smoked and how many cigarettes per day they smoked. Data were checked for consistencies and missing data were completed by linking this information to data from surveys filled out close to the time of biobanking within the longitudinal survey study of the NTR. Pack-years were calculated as the (number of cigarettes smoked per day)/20 × number of years smoked.

Statistical analyses

Overview and hypotheses

All analyses were performed in R[20]. Analyses were performed in six groups of monozygotic twin pairs (**Fig. 1**). To identify DNA methylation differences in smoking discordant monozygotic twin pairs, we first compared the twin pairs, in which one twin had never smoked, and the other was a current smoker at the time of blood sampling. Second, to identify which of these DNA methylation differences might be reversible, we analysed data from 1) monozygotic pairs in which one twin had never smoked, and the other was a former smoker at the time of blood sampling, 2) from monozygotic pairs in which one twin was a current smoker, and the other was a former smoker at the time of blood sampling, and 3) from monozygotic pairs who were both former smokers. Third, to quantify within-pair methylation differences that occur by chance alone, we compared the within-pair differences monozygotic twins concordant for never having smoked. Forth, data monozygotic twins concordant for current smoking were analysed to examine the effects of smoking intensity. Our hypotheses were as follows: 1) if DNA methylation level is reactive to cigarette smoking, methylation differences will be present between smokers and non-smokers after ruling out genetic differences, i.e. in smoking discordant monozygotic twins, and these differences will be larger than in monozygotic pairs concordant for never smoking, 2) if DNA methylation patterns are reversible upon quitting smoking, methylation differences (ΔM) in monozygotic pairs will show the following pattern: ΔM discordant current-never > ΔM discordant current-former and ΔM discordant former-never > ΔM concordant never, 3) a correlation between time since quitting smoking and ΔM in pairs discordant for former smoking is consistent with a gradual reversibility of methylation levels upon quitting smoking, 4) a correlation between ΔM and the difference in number of cigarettes smoked per day in smoking concordant pairs is consistent with smoking-reactive methylation patterns that show a dose-response relationship with amount of cigarettes smoked.

Epigenome-wide association study

In the entire data set of 3087 blood samples, we used linear regression analysis to correct the DNA methylation levels (β -values) for HM450k array row, bisulphite plate (dummy-coding) and white blood cell percentages (% neutrophils, % monocytes, and % eosinophils). White blood cell percentages were included to account for variation in cellular composition between whole blood samples. Lymphocyte percentage was not included in models because it was strongly correlated with neutrophil percentage (r = -0.93), and basophil percentage was not included because it showed little variation between subjects, with a large number of subjects having 0% of basophils. We did not adjust for sex and age, because monozygotic twins have the same sex and age. The residuals from this regression analysis were used in the within-pair EWAS analyses. Specifically, the residuals were used as input for paired *t*-tests to compare the methylation of the smoking twins with that of their non-smoking co-twins. Similarly, paired *t*-tests were applied to data from smoking concordant pairs, and to test for differences in white blood cell profiles and smoking frequency. Statistical significance was assessed following Bonferroni correction for the number of methylation sites tested ($\alpha = 0.05/411{,}169 = 1.2 \times 10^{-7}$). A previous power analysis for DNA methylation studies in discordant monozygotic twins indicated that with 50 discordant pairs, there's 80% power to detect methylation differences of 15% (at epigenome-wide significance; i.e. following multiple testing correction)[21]. Power quickly drops for smaller effect sizes; e.g. with 50 discordant pairs, the power to detect a 10% methylation difference is 10% and the power to detect a methylation difference of 5% approaches alpha[21].

Dose-response relationships

For significant CpGs from the EWAS of discordant monozygotic twin pairs, we examined dose-response relationships in smoking concordant pairs (both twins were current smokers) by correlating within-pair differences in DNA methylation with within-pair differences in smoking packyears and cigarettes per day. Secondly, in twin pairs discordant for former smoking (one twin never smoked and the other one is a former smoker), we correlated and plotted within-pair differences in DNA methylation with the time since quitting smoking to

assess the relationship between time since quitting smoking and reversal of methylation differences within monozygotic twin pairs.

Enrichment analyses

We used the EWAS Toolkit from the EWAS atlas[22] to perform enrichment analyses of Gene Ontology Terms, Kegg Pathways, and previously associated traits among top-sites from the EWAS in discordant monozygotic twin pairs (current versus never). With the trait enrichment tool of the EWAS analysis, we tested for enrichment of all traits (680) that were present in the atlas on April 26, 2022. Because the software requires a minimum of 20 input CpGs, we selected the top 20 CpGs from the EWAS in discordant monozygotic pairs for the enrichment analyses using the EWAS toolkit.

To study overlap of EWAS signal with genetic findings for smoking, we considered the largest GWAS meta-analysis of smoking phenotypes, which is the meta-analysis of smoking initiation by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN)[10]. We obtained leave-one out meta-analysis results with NTR excluded. From the GWAS, we selected all SNPs with a p-value $<5.0\text{x}10^{-8}$ and determined the distance of each Illumina 450k methylation site to each SNP. We then tested whether methylation sites within 1 Mb of genome-wide significant SNPs from the GWAS showed a stronger signal in the within-pair EWAS of smoking discordant monozygotic pairs compared to other genome-wide methylation sites, by regressing the EWAS test statistics on a variable ("GWAS locus") indicating if the CpG is located within a 1 Mb window from SNPs associated with smoking initiation (1 = yes, 0 = no):

$$|t| = \text{Intercept} + \beta_{\text{GWAS locus}} * \text{GWAS locus}$$

Where $|t|$ represents the absolute t-statistic from the paired t-test comparing within-pair methylation differences in smoking discordant pairs and $\beta_{\text{GWAS locus}}$ represents the estimate for GWASlocus, i.e. the change in the t-test statistic associated with a one-unit change in the

variable GWAS locus (e.g. being within 1 Mb of SNPs associated with smoking initiation). For each enrichment test, bootstrap standard erors were computed with 2000 bootstraps with the R-package 'simpleboot'.

Gene Expression

For significant CpGs from the EWAS of discordant monozygotic twin pairs (current versus never), we examined whether the DNA methylation was associated with gene expression levels in *cis*. To this end, we used an independent whole blood RNA-sequencing dataset from the Biobank-based Integrative Omics Study (BIOS) consortium that did not include NTR, and tested associations between genome-wide CpGs and transcripts in cis (<250 kb), as described in detail previously[23]. In short, methylation and expression levels in whole-blood samples (n=2,101) were quantified with Illumina Infinium HumanMethylation450 BeadChip arrays and with RNA-seq (2x50bp paired-end, Hiseq2000, >15M read pairs per sample). For each target CpG (epigenome-wide significant DMPs), we identified transcripts in cis (<250 kb), for which methylation levels were significantly associated with gene expression levels at the experiment-wide threshold applied by this study (FDR<5.0%), after regressing out mQTL and eQTL effects. We also examined whether significant CpGs from the EWAS of discordant monozygotic twin pairs mapped to genes that were previously reported to be differentially expressed in monozygotic pairs of which one twin never smoked, and the other was a current smoker at the time of blood sampling (based on Affymetrix U219 array data; n = 56 pairs; note: the 53 discordant pairs included in the current study of DNA methylation are a sub-set of the 56 discordant pairs included in the study of gene expression)[5].

Results

Descriptives of the smoking discordant and concordant monozygotic twin pairs are given in **table 1**. In twin pairs discordant for current smoking status (N=53 pairs, mean age=33

289 years), the smoking twin on average smoked 8.9 cigarettes per day at the time of blood
290 sampling, and had an average smoking history equivalent to 6.8 pack years. The
291 epigenome-wide association study (EWAS) analysis in pairs discordant for current smoking
292 status identified 13 epigenome-wide significant ($P<1.20\times10^{-7}$) differentially methylated
293 positions (DMPs; **Fig. 2a**). Genome-wide test statistics were not inflated (**Additional file 1**).
294 Absolute differences in methylation ranged from 2.5-13% (0.025-0.13 on the methylation β-
295 value scale), with a mean of 5.4% (**Table 2**). Eight of the 13 CpGs (61.5%) showed lower
296 methylation in the current smoking twins compared to their non-smoking twins.

297 In twin pairs discordant for former smoking (N=72 pairs, mean age=41 years), the twins, who
298 used to smoke, had quit smoking on average 14 years ago, while the other twins had never
299 initiated regular smoking. In this group, no epigenome-wide significant DMPs were identified,
300 and within-pair differences at the 13 significant DMPs identified in the previous analysis were
301 diminished (average reduction: 81%, range=61-96%; **Fig. 2b, Table 2**). By contrast, in twin
302 pairs of which one twin was a current smoker at blood draw and the co-twin had quit smoking
303 (on average 9 years, ago, N=66 pairs), the reduction of within-pair differences at the 13 top
304 CpGs was much smaller (on average, 31%, rage 15-52%; **Fig. 2b, Additional file 2**), and 5
305 of the 13 DMPs identified by comparing current and never smoking twins were also
306 epigenome-wide significant in this group. Furthermore, 5 additional epigenome-wide CpGs
307 were identified in current/former smoking discordant pairs (**Additional file 3**). **Fig. 2b**
308 illustrates the pattern of within-pair differences at the 13 top DMPs identified in current/never
309 discordant monozygotic pairs: largest differences in current/never smoking discordant pairs,
310 smaller differences in former/never discordant pair, and current/former discordant pairs are
311 intermediate. Differences are smallest within smoking concordant pairs. This pattern is in line
312 with smoking-associated methylation patterns in blood cells being to a large extent reversible
313 upon quitting smoking.

314 Distributions of within-pair differences in smoking-discordant and concordant pairs for the top
315 1000 CpGs of the EWAS in discordant pairs are shown in **Fig. 3a**. The distributions illustrate

that differences are largest, as expected, within monozygotic twin pairs discordant for current smoking (current/never smoking pairs), followed by discordant current/former smoking discordant pairs, followed by former/never smoking discordant monozygotic twin pairs. Monozygotic pairs concordant for current smoking also show notable within-pair differences at these CpGs that are substantially larger compared to monozygotic pairs concordant for never smoking (**Fig. 3a**). This could be explained by within-pair differences in the number of cigarettes smoked by monozygotic twins who were concordant for current smoking. The twin correlations in current smoking monozygotic twin pairs were r=0.50, p= 2.2×10^{-6} for cigarettes per day (**Fig. 3b)** and r=0.43, p=3.2×10^{-4} for packyears, respectively. Within-pair differences in DNA methylation at the 13 top-CpGs correlated with within-pair differences in the number of cigarettes smoked per day (mean absolute r= 0.38, range (for different CpGs): -0.56-0.41; **Table 3**, **Fig. 3c**) and with within-pair differences in packyears (mean absolute r= 0.46; range: -0.65-0.42; **Table 3**). In twin pairs discordant for former smoking, within-pair differences in DNA methylation at the 13 top-CpGs were weakly correlated with time since quitting smoking (mean r= -0.11, range=-0.28- 0.05, **Additional file 4).** Based on scatterplots of the within-pair methylation differences against time since quitting smoking (**Fig. 3d**), we hypothesized that the lack of a strong correlation with time since quitting smoking might be explained by most of the reversal taking place within the first years after quitting smoking. We therefore repeated the analysis restricting to those pairs of which the smoking twin had quit smoking less than 5 years ago (N=15 pairs). In this group, within-pair differences in DNA methylation at the 13 top-CpGs were on average more strongly correlated with time since quitting smoking (mean r= -0.16, range=-0.48- 0.23) but the sample size was greatly reduced and correlations were non-significant.

All 13 differentially methylated CpGs identified in current smoking-discordant pairs have been previously associated with smoking. To the study the overlap of methylation differences between smoking discordant twins with loci that have a causal effect on smoking, we considered the largest GWAS meta-analysis of smoking phenotypes, the meta-analysis of

smoking initiation by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN)[10]. Three of the 13 epigenome-wide significant DMPs detected in smoking-discordant monozygotic pairs (cg13411554, cg00336149, and cg21188533 in *CACNA1D*) are located within 1Mb of a GWAS locus associated with smoking initiation. The methylation sites within 1 Mb of genome-wide significant SNPs from the GWAS overall did not show a stronger signal in the within-pair EWAS of smoking discordant monozygotic pairs compared to other genome-wide methylation sites (beta=-0.002, se=0.004, p=0. 0.56, **Fig. 2c**).

We tested for enrichment of methylation sites previously associated with 680 traits reported in the EWAS atlas[22], among the top differentially methylated loci in smoking discordant pairs, which showed strong enrichment of smoking-related traits (**Additional file 5**). Enrichment analysis based on Kegg Pathways showed one significantly enriched pathway; Dopaminergic Synapse (hsa04728; **Additional file 6**), with 3 of the top differentially methylated loci in smoking discordant monozygotic pairs mapping to this pathway; *CACNA1D, GNG12,* and *ARRB1*. No significant enrichment was seen in GO pathways after multiple testing correction (**Additional file 7**).

To examine potential functional consequences of top DMPs, we used previously published data on whole-blood DNA methylation and RNA sequencing (n = 2,101 samples). At four of the 13 CpGs, DNA methylation level in blood was associated with the expression level of nearby genes (**Table 4**). At three CpGs, a higher methylation level correlated with lower expression level. None of the 13 CpGs overlapped with six genes that were differentially expressed in monozygotic pairs discordant for current smoking[5].

Discussion

Previous epigenome-wide association studies (EWAS) have identified robust differences in DNA methylation between smokers and non-smokers at a number of loci. These differences may reflect true smoking-reactive DNA methylation patterns, but can also be driven by (genetic) confounding or reverse causation. We exploited a strong within-family design, i.e., the discordant monozygotic twin design[24], to identify smoking-reactive loci. By analysing

whole blood genome-wide DNA methylation patterns in 53 monozygotic pairs discordant for current smoking, we found 13 CpGs with a difference in methylation level between the current smoking twin and the twin who never smoked. All 13 CpGs have been previously associated with smoking in unrelated individuals and in line with previous studies that compared unrelated smokers and controls[2], our data from monozygotic pairs discordant for former smoking also indicate that methylation patterns are to a large extent reversible upon smoking cessation. We further showed that differences in smoking level exposure for monozygotic twins who are both current smokers but differ in the number of cigarettes they smoke are reflected in their DNA methylation profiles.

The strongest smoking-associated loci typically detected in human blood EWAS are genes involved in detoxification pathways of aromatic hydrocarbons, such as *AHRR* and *CYP1A1*[1], of which *AHRR* was also present among the top differentially methylated loci in our analysis of discordant twins. Mainstream tobacco smoke is a mixture of thousands of chemicals[25]. Although the effects of many of the compounds present in cigarette smoke are unknown, several mechanisms have been described through which cigarette smoking may affect global or gene-specific DNA methylation levels. These include DNA damage induced by certain compounds such as arsenic, chromium, formaldehyde, polycyclic aromatic hydrocarbons, and nitrosamines that all cause double-stranded breaks[26] (which causes increased methylation near repaired DNA) [27],[28], hypoxia induced by carbon monoxide[29] (causing global CpG island demethylation by disrupting methyl donor availability), and modulation of the expression level or activity of DNA-binding proteins, such as transcription factors[30]. Nicotine, presumed to be the major addictive compound in cigarette smoke (although other putative addictive compounds have also been described[31]), has gene regulatory effects. Binding of nicotine to nicotinic acetylcholine receptors causes downstream activation of cAMP response element binding protein, which is a key transcription factor for many genes[32]. In mouse brain, nicotine downregulates the DNA methyl transferase gene *Dnmt*[33].

Importantly, effects of smoking on DNA methylation in brain cells have been hypothesized to contribute to addiction[3], but it is largely unknown to what extent addiction-related DNA methylation dynamics are captured in other tissues such as blood. Nicotinic receptors are especially abundant in the central and peripheral nervous system, but are also present in other tissues. In peripheral blood, nicotinic receptors are present on lymphocytes and polymorphonuclear cells[34], suggesting that EWA studies performed on blood cells might capture nicotine-reactive methylation patterns. Interesting in this regard is our finding that among the top differentially methylated CpGs in smoking discordant pairs are multiple CpGs in *CACNA1D and GNG12,* which encode subunits of a calcium voltage-gated channel and G protein, respectively; proteins that interact with the nicotinic acetylcholine receptor, and the related enrichment of Kegg Pathway dopaminergic neuron. Methylation levels at these CpGs might be reactive to nicotine exposure. Furthermore, the CpGs in *CACNA1D* are in proximity of a GWAS locus for smoking initiation, suggesting that this might be a locus that is not only reactive to smoking exposure, but may also contribute to smoking behaviour. Although it remains to be established if the epigenetic and genetic variation at this locus are functionally connected (i.e. have the same downstream consequences on gene expression), these results suggest that these CpGs can be interesting candidates for further studies into peripheral biomarkers of smoking addiction. Since we applied a discordant monozygotic twin design, the methylation differences identified at this locus in our study cannot be driven by mQTL effects of the SNPs associated with smoking.

The main strength of our study is the use of the discordant monozygotic twin design to examine the effects of smoking, because it rules out genetic confounding, as well as many other confounding factors. The value of studying smoking effects against an identical genetic background is clear if one considers that one of the most strongly associated genetic variants for nicotine dependence is located in the DNA methyltransferase gene *DNMT3B*[35]. This strongly implies a role for DNA methylation in nicotine addiction, but it also suggests that horizontal genetic pleiotropy might contribute to associations between DNA methylation and

smoking in ordinary case-control EWASs, where differences in DNA methylation between unrelated smokers and non-smokers may reflect differences in genotype. Our analysis had adequate power to detect large effects (i.e., the top hits identified in typical smoking EWAS)[21]. Larger sample sizes are required to achieve adequate power to detect smaller effects. While the pattern of within-pair differences in current/never, current/former and former/never discordant monozygotic twin pairs was clearly in line with reversal of methylation patterns following smoking cessation, we did not find a strong correlation between within-pair differences in DNA methylation and time since quitting smoking in former smoking discordant pairs. If most reversal takes place gradually in the first view years after smoking cessation, it might require larger sample sizes of twin pairs discordant for recently quitting smoking to detect such a correlation. Larger samples sizes may be achieved by combining data from multiple twin cohorts in a meta-analysis. Common limitations that apply to many EWA studies including ours are that we only analysed DNA methylation data from blood and that the technique used to measure DNA methylation only covers a small sub-set of all CpG sites in the genome.

Conclusion

In conclusion, we studied reactiveness of DNA methylation in blood cells to smoking and reversibility of methylation patterns upon quitting smoking in monozygotic twins. Analyses in special groups such as monozygotic twins are valuable to validate results from large population-based EWAS meta-analyses, or to train more accurate methylation scores for environmental exposures that are not confounded by genetic effects. Our results illustrate the potential to utilize DNA methylation profiles of monozygotic twins as a read out of discordant exposures at present and in the past.

Availability of data and materials

The HumanMethylation450 BeadChip data from the NTR are available as part of the Biobank-based Integrative Omics Studies (BIOS) Consortium in the European Genome-

phenome Archive (EGA), under the accession code EGAD00010000887. The pipeline for DNA methylation-array analysis developed by the Biobank-based Integrative Omics Study (BIOS) consortium is available here: https://molepi.github.io/DNAmArray_workflow/ (https://doi.org/10.5281/zenodo.3355292). All other analysis code is available upon request from the corresponding author.

Acknowledgments

NTR warmly thanks all participants. We thank Conor Dolan for providing feedback on the manuscript. We acknowledge the contributions of the investigators of the BIOS consortium (Additional file 8): Bastiaan T. Heijmans, Peter A.C. 't Hoen, Joyce van Meurs, Aaron Isaacs, Rick Jansen, Lude Franke, René Pool, Jouke J. Hottenga, Marleen MJ van Greevenbroek, Coen D.A. Stehouwer, Carla J.H. van der Kallen, Casper G. Schalkwijk, Cisca Wijmenga, Sasha Zhernakova, Ettje F. Tigchelaar, P. Eline Slagboom, Marian Beekman, Joris Deelen, Diana van Heemst, Jan H. Veldink, Leonard H. van den Berg, Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs, André G. Uitterlinden, P. Mila Jhamai, Michael Verbiest, H. Eka D. Suchiman, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij, Nico Lakenberg, Hailiang Mei, Maarten van Iterson, Michiel van Galen, Jan Bot, Dasha V. Zhernakova, Rick Jansen, Peter van 't Hof, Patrick Deelen, Irene Nooren, Matthijs Moed, Martijn Vermaat, René Luijk, Marc Jan Bonder, Freerk van Dijk, Michiel van Galen, Wibowo Arindrarto, Szymon M. Kielbasa, Morris A. Swertz, Erik. W van Zwet.

Funding

We acknowledge funding from the National Institute on Drug Abuse grant DA049867, the Netherlands Organization for Scientific Research (NWO): Biobanking and Biomolecular Research Infrastructure (BBMRI–NL, NWO 184.033.111) and the BBRMI-NL-financed BIOS Consortium (NWO 184.021.007), NWO Large Scale infrastructures X-Omics (184.034.019), Genotype/phenotype database for behavior genetic and genetic epidemiological studies (ZonMw Middelgroot 911-09-032); Netherlands Twin Registry Repository: researching the

interplay between genome and environment (NWO-Groot 480-15-001/674); the Avera Institute, Sioux Falls (USA) and the National Institutes of Health (NIH R01 HD042157-01A1, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995); epigenetic data were generated at the Human Genomics Facility (HuGe-F) at ErasmusMC Rotterdam. DIB acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635).

Competing financial interests

None.

Author contributions

Phenotype acquisition or harmonization (including design/funding of contributing studies): GW, ECDG, DIB, MCN. Methylation data acquisition and/or QC: JVD, GW, DIB. Statistical Analysis: JVD. Interpretation of results: JVD, MCN, Draft manuscript preparation: JVD. All authors critically read and approved the final version of the manuscript.

References

1. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics [Internet]. Clinical Epigenetics; 2015;7:113. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4609112&tool=pmcentrez&rendertype=abstract

2. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9:436–47.

3. Zillich L, Poisel E, Streit F, Frank J, Fries GR, Foo JC, et al. Epigenetic Signatures of Smoking in Five Brain Regions. J. Pers. Med. . 2022.

4. Liu C, Marioni RE, Hedman KK, Pfeiffer L, Tsai PC, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2016;

5. Vink JM, Jansen R, Brooks A, Willemsen G, van Grootheest G, de Geus E, et al. Differential gene expression patterns between smokers and non-smokers: cause or consequence? Addict Biol. 2017;22:550–60.

6. Ouwens KG, Jansen R, Tolhuis B, Slagboom PE, Penninx BWJH, Boomsma DI. A characterization of postzygotic mutations identified in monozygotic twins. Hum Mutat. 2018;

7. Jonsson H, Magnusdottir E, Eggertsson HP, Stefansson OA, Arnadottir GA, Eiriksson O, et al. Differences between germline genomes of monozygotic twins. Nat Genet. 2021;53:27–34.

8. Leeuwen DM va., Agen E van, Gottschalk RWH, Vlietinck R, Gielen M, Herwijnen MHM va., et al. Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs. Carcinogenesis [Internet]. 2007;28:691–7. Available from: https://doi.org/10.1093/carcin/bgl199

528 9. Allione A, Marcon F, Fiorito G, Guarrera S, Siniscalchi E, Zijno A, et al. Novel epigenetic
529 changes unveiled by monozygotic twins discordant for smoking habits. PLoS One [Internet].
530 2015;10:e0128265. Available from:
531 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4456379&tool=pmcentrez&rendert
532 ype=abstract

533 10. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2
534 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat
535 Genet. 2019;1.

536 11. van Dongen J, Nivard MG, Willemsen G, Hottenga J-J, Helmer Q, Dolan C V, et al.
537 Genetic and environmental influences interact with age and sex in shaping the human
538 methylome. Nat Commun [Internet]. 2016;7:11115. Available from:
539 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4820961&tool=pmcentrez&rendert
540 ype=abstract

541 12. Ligthart L, van Beijsterveldt CEM, Kevenaar ST, de Zeeuw E, van Bergen E, Bruins S, et
542 al. The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family
543 Designs. Twin Res Hum Genet. 2019;

544 13. Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks a I, Estourgie-van
545 Burk GF, et al. The Netherlands Twin Register biobank: a resource for genetic
546 epidemiological studies. Twin Res Hum Genet [Internet]. 2010;13:231–45. Available from:
547 http://www.ncbi.nlm.nih.gov/pubmed/20477721

548 14. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA
549 methylation array with single CpG site resolution. Genomics. 2011;98:288–95.

550 15. Iterson M Van, Tobi EW, Slieker RC, Hollander W Den, Slagboom PE, Heijmans BT.
551 MethylAid□: visual and interactive quality control of large Illumina 450k datasets.
552 Bioinformatics. 2014;30:3435–7.

16. Maarten van Iterson, Elmar Tobi, Roderick Slieker, Wouter den Hollander, Koen Dekkers, Rene Luijk ES, Heijmans and B. Streamlined workflow for the quality control, normalization and bias-free analysis of Illumina methylation array data - The Leiden approach [Internet]. [cited 2018 Jun 22]. Available from: https://molepi.github.io/DNAmArray_workflow/

17. Chen Y, Lemire M, Choufani S, Butcher DT, Zanke BW, Gallinger S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;2294.

18. Fortin J, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;1–17.

19. Francioli LC, Menelaou A, Pulit SL, Van Dijk F, Palamara PF, Elbers CC, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet. 2014;

20. Team RC. R: A language and environment for statistical computing. 2013;

21. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol. 2015;44:1429–41.

22. Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: A curated knowledgebase of epigenome-wide association studies. Nucleic Acids Res. 2019;

23. Bonder MJ, Luijk R, Zhernakova D V, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. United States; 2017;49:131–8.

24. Bell JT, Spector TD. DNA methylation studies using twins: what are they telling us? Genome Biol [Internet]. 2012;13:172. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23078798%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3491399

578 25. Rodgman A, Perfetti TA. The Chemical Components of Tobacco and Tobacco Smoke.
579 Chem. Components Tob. Tob. Smoke. 2008.

580 26. Smith CJ, Hansch C. The relative toxicity of compounds in mainstream cigarette smoke
581 condensate. Food Chem Toxicol. 2000;38:637–46.

582 27. Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H. Recruitment of DNA
583 methyltransferase I to DNA repair sites. Proc Natl Acad Sci U S A. 2005;102:8905–9.

584 28. Cuozzo C, Porcellini A, Angrisano T, Morano A, Lee B, Di Pardo A, et al. DNA damage,
585 homology-directed repair, and DNA methylation. PLoS Genet. 2007;3:1144–62.

586 29. Olson KR. Carbon monoxide poisoning: Mechanisms, presentation, and controversies in
587 management. J Emerg Med. 1984;1:233–43.

588 30. Lee KWK, Pausova Z. Cigarette smoking and DNA methylation. Front. Genet. 2013.

589 31. Talhout R, Schulz T, Florek E, Van Benthem J, Wester P, Opperhuizen A. Hazardous
590 Compounds in Tobacco Smoke. Int. J. Environ. Res. Public Heal. . 2011.

591 32. Shen JX, Yakel JL. Nicotinic acetylcholine receptor-mediated calcium signaling in the
592 nervous system. Acta Pharmacol. Sin. 2009. p. 673–80.

593 33. R. S, E. M, A. Z, F. P, M. H, E. C, et al. Nicotine decreases DNA methyltransferase 1
594 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic
595 interneurons. Proc Natl Acad Sci [Internet]. Proceedings of the National Academy of
596 Sciences; 2008;105:16356–61. Available from: https://doi.org/10.1073/pnas.0808699105

597 34. Benhammou K, Lee M, Strook MA, Sullivan B, Logel J, Raschen K, et al. [3H]Nicotine
598 binding in peripheral blood cells of smokers is correlated with the number of cigarettes
599 smoked per day. Neuropharmacology. 2000;39:2818–29.

600 35. Hancock DB, Guo Y, Reginsson GW, Gaddis NC, Lutz SM, Sherva R, et al. Genome-
601 wide association study across European and African American ancestries identifies a SNP in
602 DNMT3B contributing to nicotine dependence. Mol Psychiatry. 2017;

603 ***Figure Legends***

604

605 **Figure 1. DNA methylation analysis in smoking discordant and smoking concordant**
606 **monozygotic twin pairs.** Blood DNA methylation profiles (Illumina 450k array) from six
607 groups of monozygotic twin pairs were analysed.

608

609 **Figure 2. Top differentially methylated loci identified in monozygotic twin pairs**
610 **discordant for current smoking** a) Manhattan plot of the EWAS in 53 smoking discordant
611 monozygotic twin pairs (current versus never). The red horizontal line denotes the
612 epigenome-wide significance threshold (Bonferroni correction) and 13 CpGs with significant
613 differences are highlighted. b) Mean within-pair differences in monozygotic twin pairs at the
614 13 CpGs that were epigenome-wide significant in smoking discordant monozygotic pairs.
615 Mean within-pair differences of the residuals obtained after correction of methylation beta-
616 values for covariates are shown for 53 monozygotic pairs discordant for current/never
617 smoking, 66 monozygotic pairs discordant for current/former smoking, 72 monozygotic pairs
618 discordant for former/never smoking, 83 concordant current smoking monozygotic pairs, 88
619 concordant former smoking monozygotic pairs, and 406 concordant never smoking
620 monozygotic pairs. c) QQ-plot showing p-values from the EWAS in 53 smoking discordant
621 monozygotic twin pairs (current versus never). P-values for CpGs located nearby significant
622 SNPs from the GWAS of smoking initiation are plotted in blue and all other genome-wide
623 CpGs are plotted in orange.

624

625 **Figure 3. DNA methylation differences in smoking discordant and smoking**
626 **concordant pairs.** a) Distributions of the mean absolute within-pair differences in discordant
627 and concordant pairs at the top 1000 CpGs with the lowest p-value from the EWAS in
628 discordant monozygotic pairs (current versus never smokers). b) Scatterplot of cigarettes
629 smoked per day in 80 concordant current smoking monozygotic pairs with complete data c)
630 Scatterplot of within-pair differences in cigarettes smoked per day versus DNA methylation

631 at cg05575921 (*AHRR*) in 80 concordant current smoking monozygotic pairs with complete
632 data. d) Scatterplot of within-pair differences in DNA methylation at cg05575921 (*AHRR*)
633 versus time since quitting smoking (years) in 63 pairs discordant for former smoking.
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651

Table 1 Descriptive statistics for smoking discordant and concordant monozygotic twin pairs

	Discordant Current/never (53 pairs)				Discordant Former/never (72 pairs)				Discordant Current/former (66 pairs)				Concordant Current (83 pairs)				Concordant never (406 pairs)				Concordant former (88 pairs)			
	Current smoker	Never-smoker	Mean diff	P-value	Former smoker	Never-smoker	Mean diff	P-value	Current smoker	Former smoker	Mean diff	P-value	Twin 1	Twin 2	Mean diff	P-value	Twin 1	Twin 2	Mean diff	P-value	Twin 1	Twin 2	Mean diff	P-value
% Female pairs	60.4%	60.4%	n.a.	n.a.	77.80%	77.80%	n.a.	n.a.	69.7%	69.7%	n.a.	n.a.	61.4%	61.4%	n.a.	n.a.	73.6%	73.6%	n.a.	n.a.	64.8%	64.8%	n.a.	n.a.
age at blood sampling, mean (SD)	33.1 (8.0)	33.0 (7.9)	0.10	0.34	41.4 (13.2)	41.4 (13.1)	0.02	0.83	42.2 (12.6)	42.2 (12.5)	-0.06	0.45	33.8 (10.3)	33.9 (10.5)	-0.12	0.10	33.1 (11.3)	33.0 (11.2)	0.06	0.08	45.2 (13.4)	45.2 (13.4)	0.09	0.29
Cigarettes per day at blood sampling, mean (SD), N missings	8.9 (6.4), 6	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	11.9 (7.2),9	n.a.	n.a.	n.a.	11.1 (7.0),2	10.9 (6.9), 1	0.00	1.00	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Packyears, mean (SD), N missings	6.8 (7.0), 13	n.a.	n.a.	n.a.	5.9 (11.1), 15	n.a.	n.a.	n.a.	13.6 (13.2),9	9.3 (8.7),10	4.2	0.02	9.7 (9.3),10	8.3 (7.6), 9	0.22	0.82	n.a.	n.a.	n.a.	n.a.	10.6 (11.5),7	9.8 (10.4),11	0.78	0.55
Years since quitting smoking, mean (SD), N missings	n.a.	n.a.	n.a.	n.a.	13.5 (11.4), 9	n.a.	n.a.	n.a.	n.a.	9.0 (10.2),2	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	11.9 (9.1),8	13.6 (11.8),7	-1.62	0.20
Percentage monocytes, mean (SD), N missings	8.0 (2.3), 0	8.5 (2.4), 0	-0.44	0.19	8.6 (2.0), 0	8.3 (1.8), 0	0.29	0.19	8.6 (1.9),0	9.2 (3.1),0	-0.57	0.14	8.3 (2.1), 0	8.1 (1.9), 0	0.20	0.36	8.5 (2.0), 0	8.5 (2.2), 0	0.03	0.75	8.4 (2.4),0	8.5 (2.4),0	-0.06	0.75
Percentage lymphocytes, mean (SD), N missings	35.0 (8.9), 0	35.9 (10.0)	-0.94	0.50	33.6 (8.5), 0	34.0 (8.7)	-0.37	0.77	35.8 (8.2),0	35.6 (8.5),0	0.25	0.84	33.7 (8.3), 0	34.1 (8.3), 0	-0.44	0.67	36.3 (8.4), 0	36.2 (8.4), 0	0.04	0.92	35.0 (7.7),0	34.1 (8.4),0	0.95	0.26
Percentage neutrophils, mean (SD), N missings	53.4 (9.5), 0	52.1 (9.8), 0	1.34	0.38	54.4 (9.1), 0	54.5 (9.1), 0	-0.08	0.95	51.6 (9.0),0	51.7 (8.9),0	0-0.06	0.96	53.7 (8.9), 0	53.7 (9.3), 0	0.03	0.98	51.8, (8.7), 0	51.9 (9.3), 0	-0.08	0.86	52.8 (8.2),0	53.7 (8.4),0	-0.84	0.35
Percentage eosinophils, mean (SD), N missings	3.1 (2.5), 0	3.1 (2.1), 0	.05	0.91	3.1 (1.9), 0	2.9 (2.0), 0	0.15	0.53	3.3 (1.9),0	3.1 (1.7),0	0.21	0.33	3.4 (2.2), 0	3.4 (1.8), 0	-0.03	0.91	2.9 (1.8), 0	2.9 (1.9), 0	-0.04	0.66	3.1 (1.9),0	3.2 (2.4),0	-0.06	0.77
Percentage basophils, mean (SD), N missings	0.5 (0.7), 0	0.5 (0.7), 0	-0.01	0.96	0.3 (0.3), 0	0.4 (0.5), 0	-0.02	0.76	0.6 (0.9),0	0.4 (0.4),0	0.18	0.12	0.9 (3.1), 0	0.6 (1.1), 0	0.25	0.49	0.5 (0.7), 0	0.4 (0.7), 0	0.06	0.21	0.6 (1.1),0	0.6 (0.9),0	-0.01	0.97

Table 2 Epigenome-wide significant differentially methylated CpGs in monozygotic pairs discordant for current smoking status

					Current smoking discordant pairs					Former smoking discordant pairs (former/never)				
IlmnID	CHR	MAPINFO	Gene*	*Nearest gene*	Mean Diff	p-value	95conf_L	95conf_H	T-statistic	Mean Diff	P-value	95conf_L	95conf_H	T-Statistic
cg05575921	5	373378	*AHRR*	*AHRR*	0.132	$4.9x10^{-11}$	0.100	0.165	8.265	0.027	$3.3x10^{-4}$	0.013	0.041	3.778
cg21566642	2	233284661		*ALPPL2*	0.092	$1.5x10^{-10}$	0.069	0.115	7.960	0.033	$3.2x10^{-6}$	0.020	0.046	5.067
cg05951221	2	233284402		*ALPPL2*	0.066	$1.8x10^{-9}$	0.048	0.084	7.270	0.026	$4.6x10^{-6}$	0.016	0.037	4.964
cg01940273	2	233284934		*ALPPL2*	0.060	$2.1x10^{-9}$	0.044	0.077	7.240	0.018	$1.3x10^{-4}$	0.009	0.027	4.052
cg13411554	3	53700276	*CACNA1D*	*CACNA1D*	-0.038	$6.0x10^{-9}$	-0.049	-0.027	-6.947	-0.007	0.10	-0.016	0.002	-1.655
cg01901332	11	75031054	*ARRB1*	*ARRB1*	0.025	$8.0x10^{-9}$	0.018	0.033	6.868	0.006	0.16	-0.002	0.013	1.425
cg21161138	5	399360	*AHRR*	*AHRR*	0.046	$1.9x10^{-8}$	0.032	0.059	6.642	0.002	0.64	-0.006	0.009	0.466
cg00336149	3	53700195	*CACNA1D*	*CACNA1D*	-0.027	$2.0x10^{-8}$	-0.035	-0.019	-6.615	-0.002	0.60	-0.008	0.005	-0.524
cg22132788	7	45002486	*MYO1G*	*MYO1G*	-0.056	$2.4x10^{-8}$	-0.073	-0.039	-6.596	-0.011	$4.5x10^{-3}$	-0.019	-0.004	-2.930
cg21188533	3	53700263	*CACNA1D*	*CACNA1D*	-0.036	$3.9x10^{-8}$	-0.047	-0.025	-6.437	-0.006	0.24	-0.015	0.004	-1.196
cg09935388	1	92947588	*GFI1*	*GFI1*	0.052	$4.1x10^{-8}$	0.035	0.068	6.423	0.002	0.61	-0.006	0.010	0.519
cg25648203	5	395444	*AHRR*	*AHRR*	0.035	$5.3x10^{-8}$	0.024	0.046	6.353	0.002	0.48	-0.004	0.008	0.710
cg19089201	7	45002287	*MYO1G*	*MYO1G*	-0.040	$7.5x10^{-8}$	-0.053	-0.028	-6.260	-0.007	0.13	-0.017	0.002	-1.529

Coordinates are given based on genome build 37. Mean differences represent non-smoking twin minus smoking-twin (hence positive values indicate a higher methylation level in non-smoking twins). The table shows the 13 epigenome-wide significant CpGs from the within-pair EWAS in 53 discordant monozygotic twin pairs (current versus never smokers). Results from the comparison within 72 monozygotic pairs discordant for former smoking are also shown. * CpGs without a gene name are located in intergenic regions. 95conf_L=95% confidence interval lower bound, 95conf_H=95% confidence interval upper bound.

Table 3 Correlations of within-pair differences in DNA methylation with within-pair differences in cigarettes per day and packyears in 83 concordant current smoking monozygotic pairs

					cigarettes per day		Packyears	
cgid	CHR	Position	Gene	Nearest gene	*r*	p-value	*r*	p-value
cg05575921	5	373378	*AHRR*	*AHRR*	-0.52	5.9×10^{-7}	-0.55	1.2×10^{-6}
cg21566642	2	233284661		*ALPPL2*	-0.49	4.7×10^{-6}	-0.56	8.1×10^{-7}
cg05951221	2	233284402		*ALPPL2*	-0.44	4.7×10^{-5}	-0.56	9.0×10^{-7}
cg01940273	2	233284934		*ALPPL2*	-0.56	7.0×10^{-8}	-0.65	2.0×10^{-9}
cg13411554	3	53700276	*CACNA1D*	*CACNA1D*	0.27	1.4×10^{-2}	0.32	7.4×10^{-3}
cg01901332	11	75031054	*ARRB1*	*ARRB1*	-0.23	3.8×10^{-2}	-0.34	5.5×10^{-3}
cg21161138	5	399360	*AHRR*	*AHRR*	-0.52	8.4×10^{-7}	-0.52	7.4×10^{-6}
cg00336149	3	53700195	*CACNA1D*	*CACNA1D*	0.36	1.0×10^{-3}	0.42	3.5×10^{-4}
cg22132788	7	45002486	*MYO1G*	*MYO1G*	0.41	2.1×10^{-4}	0.41	8.4×10^{-4}
cg21188533	3	53700263	*CACNA1D*	*CACNA1D*	0.32	4.1×10^{-3}	0.38	1.4×10^{-3}
cg09935388	1	92947588	*GFI1*	*GFI1*	-0.42	9.1×10^{-5}	-0.59	1.4×10^{-7}
cg25648203	5	395444	*AHRR*	*AHRR*	-0.28	1.2×10^{-2}	-0.42	4.3×10^{-4}
cg19089201	7	45002287	*MYO1G*	*MYO1G*	0.15	1.8×10^{-1}	0.27	2.6×10^{-2}

675 **Table 4** Significantly associated transcripts in cis for CpGs that are differentially methylated in smoking discordant monozygotic twin pairs

CpG	gene	Z score	p-value	FDR
cg25648203	*EXOC3*	-7.34	2.11e-13	0
cg19089201	*RP4-647J21.1*	5.55	2.84e-8	0
cg05575921	*EXOC3*	-4.86	0.00000119	0.00039
cg21161138	*EXOC3*	-3.82	0.000133	0.0254

676

677

678

679

680

681

682

683

684

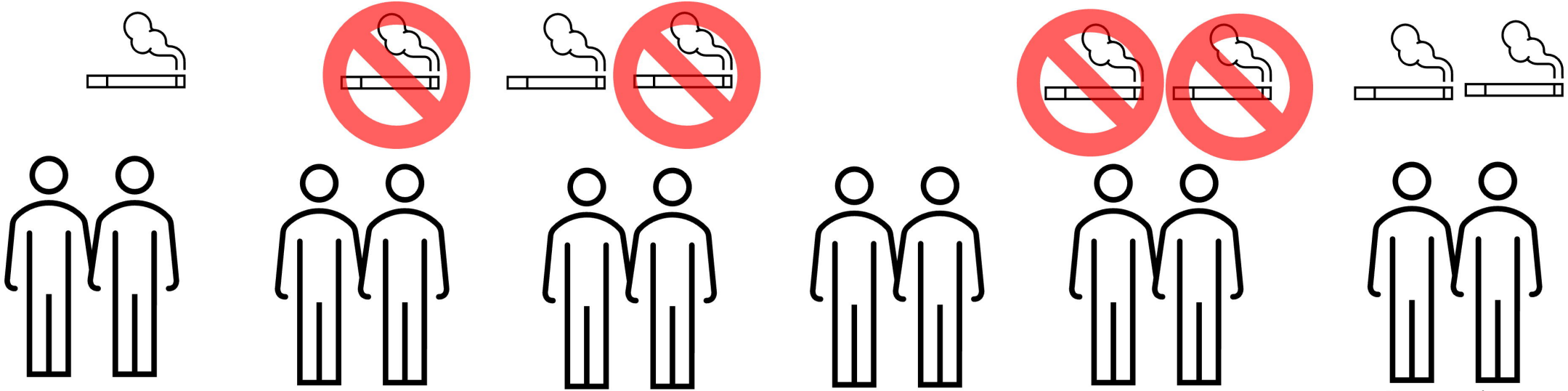
685

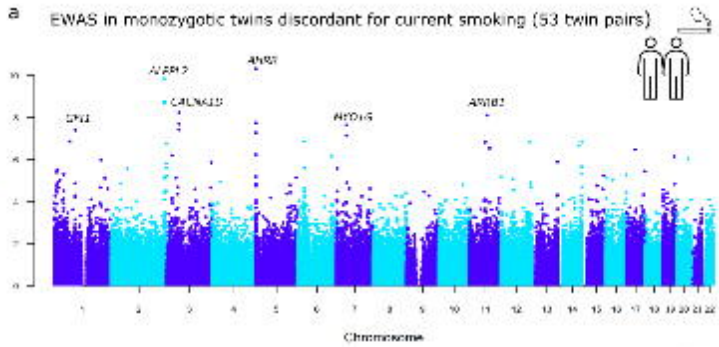
686

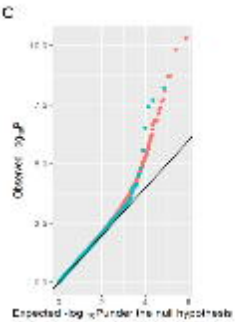
687

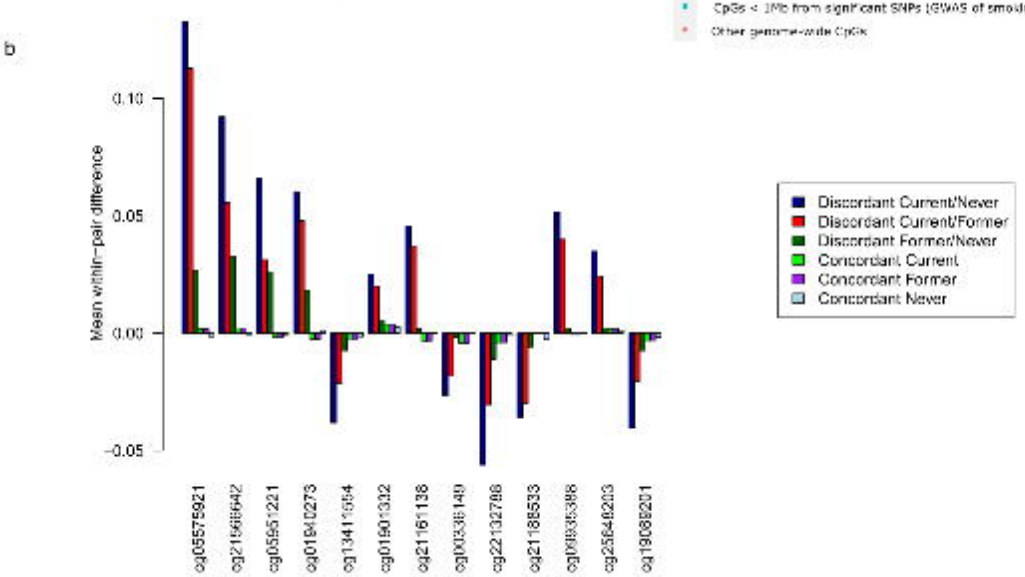
688

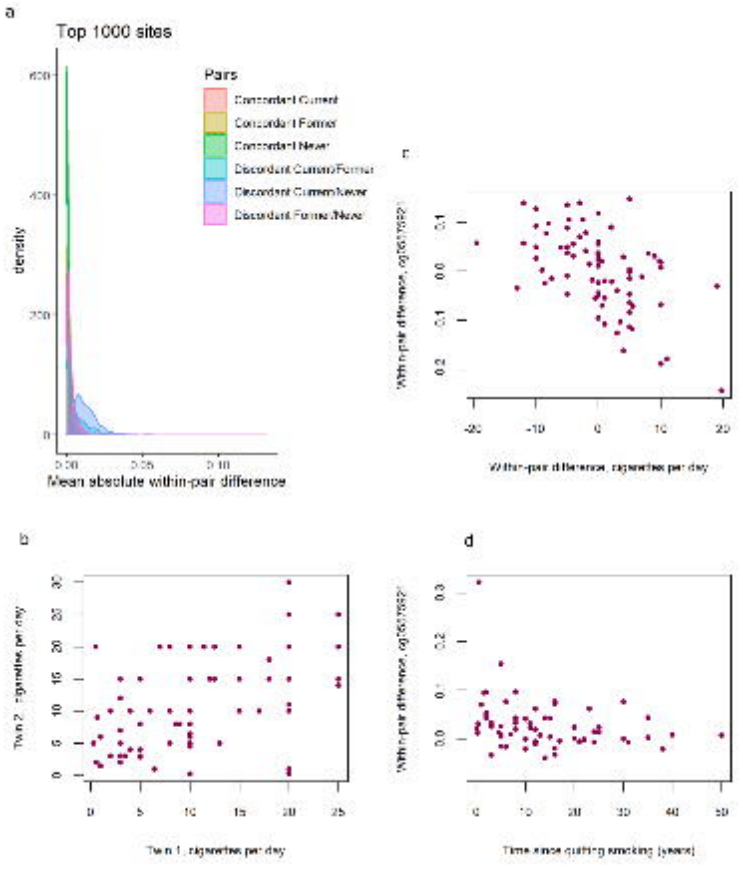
689


Discordant current smoking 53 twin pairs
Discordant former-never smoking 72 pairs
Discordant current-former smoking 66 pairs
Concordant Never smoked 406 pairs
Concordant former smoking 88 pairs
Concordant current smoking 83 pairs


a
EWAS in monozygotic twins discordant for current smoking (53 twin pairs)
AHRR
CYP1A1
CACNA1D
AHRR
AHRR
AHRR
Chromosome


c
Observed -log10P
Expected -log10P under the null hypothesis


CpGs < 1Mb from significant SNPs (GWAS of smoking initiation)
Other genome-wide CpGs
b
Mean within−pair difference
0.10
0.05
0.00
−0.05
cg05575921
cg21566642
cg05951221
cg01940273
cg13411554
cg01901332
cg21161138
cg00338149
cg22132786
cg21188533
cg00835398
cg25648203
cg19089201
Discordant Current/Never
Discordant Current/Former
Discordant Former/Never
Concordant Current
Concordant Former
Concordant Never


a
Top 1000 sites
Pairs
Concordant Current
Concordant Former
Concordant Never
Discordant Current/Former
Discordant Current/Never
Discordant Former/Never
density
Mean absolute within-pair difference
c
Within-pair difference, cg16173932
Within-pair difference, cigarettes per day
b
Twin 2, cigarettes per day
Twin 1, cigarettes per day
d
Within-pair difference, cg16173932
Time since quitting smoking (years)