RESEARCH

Open Access

CrossMark

Epigenome-wide association study of serum cotinine in current smokers reveals novel genetically driven loci

Richa Gupta[1*] , Jenny van Dongen[2], Yu Fu[1], Abdel Abdellaoui[2], Rachel F. Tyndale[3], Vidya Velagapudi[1], Dorret I. Boomsma[2], Tellervo Korhonen[1,4], Jaakko Kaprio[1,4], Anu Loukola[1,5] and Miina Ollikainen[1,4]

Abstract

Background: DNA methylation alteration extensively associates with smoking and is a plausible link between smoking and adverse health. We examined the association between epigenome-wide DNA methylation and serum cotinine levels as a proxy of nicotine exposure and smoking quantity, assessed the role of SNPs in these associations, and evaluated molecular mediation by methylation in a sample of biochemically verified current smokers ($N = 310$).

Results: DNA methylation at 50 CpG sites was associated (FDR < 0.05) with cotinine levels, 17 of which are novel associations. As cotinine levels are influenced not only by nicotine intake but also by *CYP2A6*-mediated nicotine metabolism rate, we performed secondary analyses adjusting for genetic risk score of nicotine metabolism rate and identified five additional novel associations. We further assessed the potential role of genetic variants in the detected association between methylation and cotinine levels observing 124 *cis* and 3898 *trans* methylation quantitative trait loci (meQTLs). Nineteen of these SNPs were also associated with cotinine levels (FDR < 0.05). Further, at seven CpG sites, we observed a trend ($P < 0.05$) that altered DNA methylation mediates the effect of SNPs on nicotine exposure rather than a direct consequence of smoking. Finally, we performed replication of our findings in two independent cohorts of biochemically verified smokers ($N = 450$ and $N = 79$).

Conclusions: Using cotinine, a biomarker of nicotine exposure, we replicated and extended identification of novel epigenetic associations in smoking-related genes. We also demonstrated that DNA methylation in some of the identified loci is driven by the underlying genotype and may mediate the causal effect of genotype on cotinine levels.

Keywords: Epigenome-wide association study, Smoking, Cotinine, Genetic risk score, Nicotine metabolism, meQTL, Causal inference, Molecular mediation

Background

Smoking remains a major preventable cause of morbidity and mortality worldwide and has been shown to associate extensively with DNA methylation changes across the genome as evidenced by several epigenome-wide association studies (EWAS) [1]. Almost all EWAS so far, including a recent large meta-analysis ($N = 15,907$) [2], assessed the association between DNA methylation and self-reported smoking status or smoking quantity [3–23].

Self-reported smoking status and quantity, however, are prone to inaccuracies due to reporting bias (usually under reporting or recall bias) [24]. Cotinine, the primary metabolite of nicotine, is a reliable measure of nicotine exposure among current smokers [25] and provides higher statistical power compared to self-reported smoking quantity, as shown by Ware et al. [26] in a genome-wide association study (GWAS) meta-analysis of cotinine levels.

The cotinine GWAS meta-analysis [26] identified a locus on chromosome 4, within UDP glucuronosyltransferase family 2 member B10 (*UGT2B10*) gene, a key enzyme in nicotine and cotinine metabolism [27]. However, genetic variants in this gene did not associate with

* Correspondence: richa.gupta@helsinki.fi
[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
Full list of author information is available at the end of the article

self-reported smoking quantity suggesting that cotinine may capture more than mere nicotine exposure [26]. Cotinine level is not only dependent on nicotine intake, but is also affected by the rate of formation from nicotine and the rate of cotinine metabolism to 3-hydroxycotinine, both are mediated by one highly genetically polymorphic enzyme, *CYP2A6* [28]. The rate of nicotine metabolism also influences smoking behavior; for instance, fast metabolizers tend to smoke more cigarettes per day and more intensely than slow metabolizers, likely due to the longer retention of nicotine among slow metabolizers from a given intake [29, 30]. Nicotine metabolite ratio (NMR; 3-hydroxycotinine/cotinine) is a reliable proxy for *CYP2A6*-mediated nicotine and cotinine metabolism [29]. NMR is highly heritable with just three independent SNPs accounting for ~ 30% of variance in NMR in Finnish population [31].

In this study, we performed an EWAS to examine the association of serum cotinine levels with DNA methylation in a sample of current (daily) smokers from the Finnish Twin Cohort (N = 310). To reduce the impact of variation in cotinine levels due to *CYP2A6*-mediated metabolism, we utilized the genetic risk score (GRS) of NMR. As many of the genes identified in our EWAS had genetic variants that were linked to smoking-related traits previously, we further investigated the effects of genetic variants on methylation (methylation quantitative trait loci (meQTLs)) and cotinine levels. Finally, at loci where the genetic variants were associated with both

methylation and cotinine levels, we examined whether altered methylation maybe a molecular mediator in the association observed between the genotype (single nucleotide polymorphism; SNP) and cotinine levels rather than being altered directly as a consequence of smoking.

Results

The overall study design and summary of our main findings is shown in Fig. 1.

Epigenome-wide association study

We examined the association of nicotine exposure, assessed by serum cotinine levels, with DNA methylation in a sample of current smokers (N = 310, serum cotinine > 4.85 ng/ml). In our discovery EWAS, we identified 50 CpG sites (in 35 genes) showing significant association (FDR < 0.05) with serum cotinine levels (Table 1). Among the 50 highlighted CpG sites, 17 (in 17 genes) were novel associations while 33 (in 18 genes) have previously been reported to associate with smoking in EWAS. Consistent with previous studies, cg05575921 in *AHRR* was the most significant CpG site (FDR = 1.1×10^{-12}), with seven additional previously reported CpG associations identified within *AHRR*. Other notable replicated hits include six CpG sites in *ALPPL2* (min FDR = 2.0×10^{-10} for cg05951221) and cg03636183 in *F2RL3* (FDR = 8.7×10^{-07}). For a great majority (84%, 42/50) of the highlighted CpG sites, higher cotinine levels were associated with lower methylation with small effect sizes. For instance, a
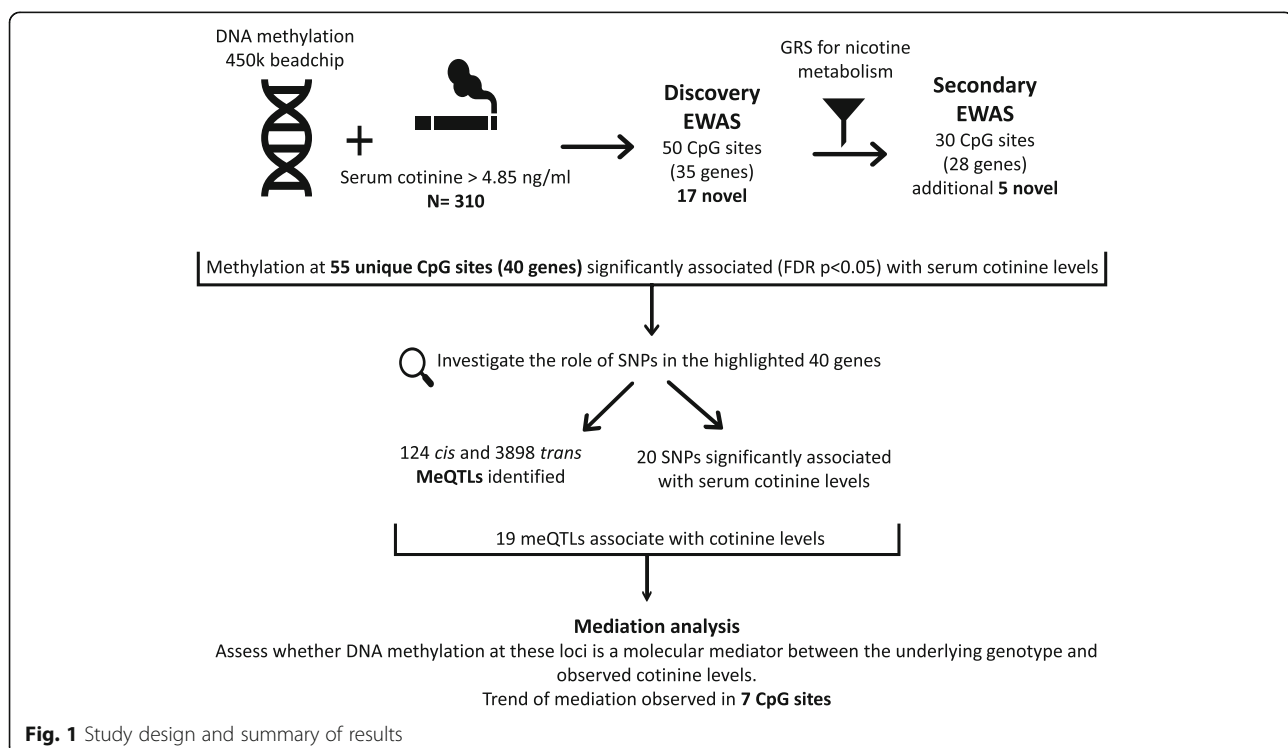

Fig. 1 Study design and summary of results

Table 1 Results from the discovery and secondary epigenome-wide association analyses of serum cotinine levels in current smokers

CpG	Chr	Nearest gene	Discovery analysis				Secondary analysis				References
			Beta	SE	P value	P_{FDR}	Beta	SE	P value	P_{FDR}	
cg25189904	1	GNG12	− 0.000135	0.000026	5.6E−07	9.0E−03	− 0.000134	0.000028	1.8E−06	3.1E−02	[2–13, 52]
cg21033965	1	CLEC20A	− 0.000096	0.000020	4.0E−06	3.7E−02		**			[2]
cg26687670	1	OPN3	− 0.000025	0.000005	6.5E−06	4.4E−02		**			Novel
cg27241845	2	ALPP	− 0.000083	0.000016	3.3E−07	6.2E−03	− 0.000082	0.000016	8.8E−07	1.7E−02	[2–10]
cg03329539	2	ALPPL2	− 0.000070	0.000012	2.7E−08	6.6E−04	− 0.000069	0.000013	1.3E−07	3.0E−03	[2–10, 12, 52]
cg06644428	2	ALPPL2	− 0.000054	0.000012	5.5E−06	4.3E−02		**			[2–11, 19, 52]
cg05951221	2	ALPPL2	− 0.000178	0.000021	1.9E−15	2.0E−10	− 0.000171	0.000022	1.4E−13	1.5E−08	[2–13, 19, 20, 52]
cg21566642	2	ALPPL2	− 0.000210	0.000026	1.1E−14	8.8E−10	− 0.000205	0.000027	4.2E−13	3.4E−08	[2–12, 19, 20, 52]
cg01940273	2	ALPPL2	− 0.000134	0.000017	9.8E−14	6.3E−09	− 0.000129	0.000018	3.4E−12	2.2E−07	[2–13, 20, 52]
cg13193840	2	ALPPL2	− 0.000045	0.000010	3.6E−06	3.4E−02	− 0.000048	0.000010	2.5E−06	3.3E−02	[2–7, 9, 10]
cg02306995	3	SEMA5B	*				0.000047	0.000010	4.4E−06	4.8E−02	Novel ***
cg04776445	4	MSX1	*				0.000085	0.000018	4.3E−06	4.8E−02	Novel ***
cg13740236	4	LSM6	− 0.000053	0.000009	2.9E−08	6.7E−04	− 0.000057	0.000010	1.0E−08	3.0E−04	Novel
cg11902777	5	AHRR	− 0.000071	0.000011	1.1E−09	3.2E−05	− 0.000068	0.000012	2.1E−08	5.6E−04	[3–7, 9]
cg01899089	5	AHRR	− 0.000063	0.000014	6.1E−06	4.4E−02	**				[2–8]
cg05575921	5	AHRR	− 0.000359	0.000039	3.3E−18	1.1E−12	− 0.000351	0.000041	4.4E−16	1.1E−10	[2–13, 16, 18–20, 51, 52]
cg26703534	5	AHRR	− 0.000086	0.000013	5.8E−10	1.9E−05	− 0.000084	0.000014	6.2E−09	2.0E−04	[2–8, 12, 18, 51, 52]
cg14817490	5	AHRR	− 0.000099	0.000021	2.4E−06	2.9E−02	**				[2–10, 12, 13, 18, 52]
cg25648203	5	AHRR	− 0.000091	0.000015	4.3E−09	1.1E−04	− 0.000086	0.000016	1.1E−07	2.8E−03	[2–13, 52]
cg21161138	5	AHRR	− 0.000142	0.000016	1.8E−17	2.9E−12	− 0.000141	0.000017	6.6E−16	1.1E−10	[2–12, 18, 51, 52]
cg24090911	5	AHRR	− 0.000076	0.000015	4.3E−07	7.4E−03	− 0.000074	0.000015	2.5E−06	3.3E−02	[2–7, 10]
cg10961758	5	EDIL3	*				− 0.000069	0.000014	2.5E−06	3.3E−02	Novel***
cg16179182	5	VTRNA1–1	0.000048	0.000010	5.7E−06	4.3E−02	**				Novel
cg14580211	5	SMIM3	− 0.000062	0.000013	1.7E−06	2.2E−02	**				[2–9, 52]
cg06126421	6	IER3	− 0.000132	0.000020	1.1E−10	4.3E−06	− 0.000132	0.000021	6.9E−10	2.8E−05	[2–12, 20, 52]
cg14753356	6	IER3	− 0.000072	0.000015	3.1E−06	3.4E−02	**				[2–9]
cg22856972	6	TBPL1	0.000066	0.000014	3.0E−06	3.4E−02	0.000073	0.000014	9.1E−07	1.7E−02	Novel
cg05469934	7	PDGFA	0.000038	0.000008	3.6E−06	3.4E−02	**				Novel
cg09022230	7	TNRC18	− 0.000079	0.000014	4.1E−08	8.8E−04	− 0.000070	0.000014	2.4E−06	3.3E−02	[2–6, 8]
cg21322436	7	CNTNAP2	− 0.000055	0.000012	5.6E−06	4.3E−02	**				[2–7, 9]
cg25949550	7	CNTNAP2	− 0.000045	0.000009	1.4E−06	1.9E−02	**				[2–9, 11, 52]
cg09267815	8	YTHDF3	− 0.000080	0.000016	7.2E−07	1.1E−02	**				Novel
cg22288066	8	NOV	0.000033	0.000007	6.8E−06	4.4E−02	**				Novel
cg24838345	8	MTSS1	− 0.000088	0.000017	2.6E−07	5.3E−03	− 0.000084	0.000017	2.3E−06	3.3E−02	[2–6, 8, 9]
cg25305703	8	CASC8	− 0.000088	0.000018	1.7E−06	2.2E−02	**				[2–7]
cg01692968	9	SLC44A1	− 0.000075	0.000016	3.3E−06	3.4E−02	− 0.000079	0.000017	2.7E−06	3.3E−02	[2–9]
cg14120703	9	NOTCH1	− 0.000043	0.000009	4.3E−06	3.8E−02	− 0.000045	0.000010	4.3E−06	4.8E−02	[2, 4]
cg05992400	10	CYP2C18	− 0.000050	0.000011	4.9E−06	4.1E−02	**				Novel
cg16421726	10	ZRANB1	0.000056	0.000012	2.7E−06	3.1E−02	**				Novel
cg05335388	11	ABTB2	*				0.000035	0.000007	1.6E−06	2.9E−02	Novel***
cg19254163	11	PTGDR2	− 0.000042	0.000009	6.4E−06	4.4E−02	**				[2–7]
cg21611682	11	LRP5	− 0.000071	0.000010	1.2E−12	6.7E−08	− 0.000067	0.000010	8.2E−11	3.8E−06	[2–11, 52]

Table 1 Results from the discovery and secondary epigenome-wide association analyses of serum cotinine levels in current smokers *(Continued)*

CpG	Chr	Nearest gene	Discovery analysis				Secondary analysis				References
			Beta	SE	P value	P FDR	Beta	SE	P value	P FDR	
cg14624207	11	LRP5	− 0.000063	0.000010	2.6E−10	9.2E−06	− 0.000063	0.000010	1.4E−09	5.1E−05	[2–7, 9]
cg10788371	11	LRRC32	− 0.000064	0.000013	1.3E−06	1.8E−02	− 0.000067	0.000013	8.6E−07	1.7E−02	[2–6]
cg11491407	12	CACNA2D4	− 0.000095	0.000020	5.7E−06	4.3E−02	**				Novel
cg18443392	12	LRP1	0.000027	0.000006	4.2E−06	3.8E−02	**				Novel
cg04524475	13	RNY1P5	− 0.000053	0.000012	6.2E−06	4.4E−02	**				Novel
cg26589665	15	THSD4	− 0.000042	0.000008	3.8E−07	6.9E−03	− 0.000041	0.000008	2.1E−06	3.3E−02	Novel
cg05361530	16	GNG13	0.000044	0.000009	4.8E−06	4.1E−02	**				Novel
cg07991586	16	ADCY7	− 0.000042	0.000009	6.6E−06	4.4E−02	**				Novel
cg05460226	17	PIK3R5	− 0.000072	0.000015	3.4E−06	3.4E−02	**				[2–4]
cg05396243	17	TMEM220-AS1	− 0.000036	0.000007	1.0E−06	1.5E−02	**				Novel
cg07658936	17	ARHGAP44	0.000029	0.000006	6.8E−06	4.4E−02	**				Novel
cg03636183	19	F2RL3	− 0.000151	0.000022	1.9E−11	8.7E−07	− 0.000155	0.000022	2.8E−11	1.5E−06	[2–15, 17, 18, 20–22, 52]
cg12817959	22	SSTR3	*				0.000034	0.000007	4.3E−06	4.8E−02	Novel***

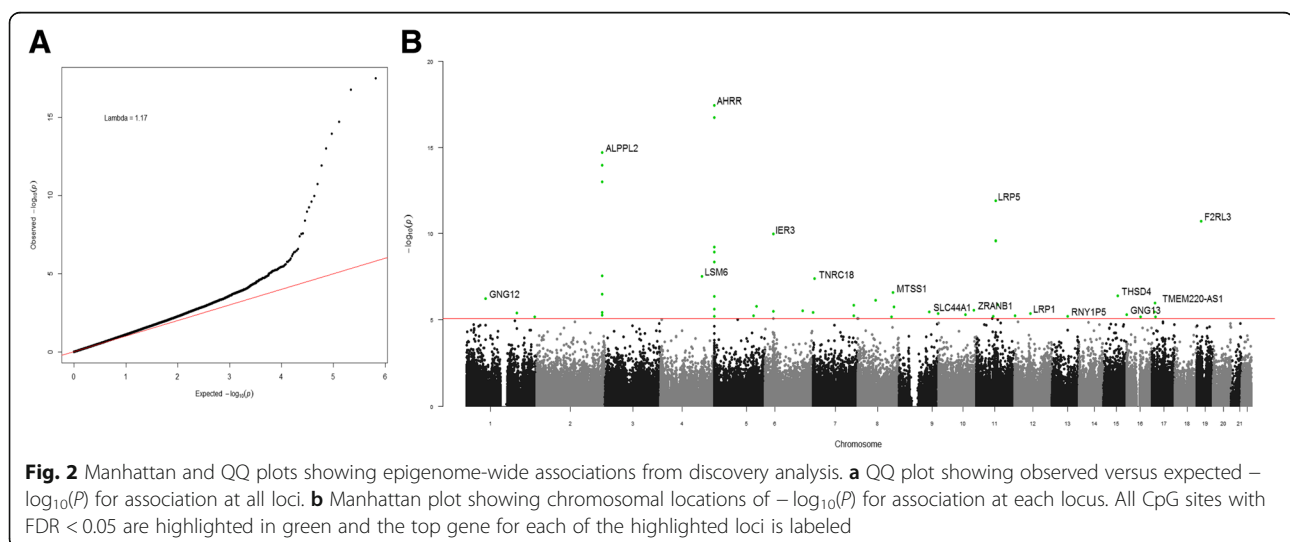*CpG not genome-wide significant in discovery analysis
**CpG not genome-wide significant in secondary analysis
***Novel hits identified in secondary analysis

100 ng/ml increase of cotinine was associated with approximately 3.6% lower methylation at cg05575921 in *AHRR*. Novel CpG associations in smoking-related genes include cg13740236 in *LSM6* (FDR = 6.7×10^{-04}) and cg26589665 in *THSD4* (FDR = 6.9×10^{-03}) (full list in Table 1). Manhattan and quantile-quantile (QQ) plots for the discovery analysis are presented in Fig. 2.

As cotinine levels are influenced not only by nicotine intake but also by nicotine and cotinine clearance rate (both mediated by *CYP2A6*), we performed a secondary analysis, wherein we included the GRS for NMR as an additional covariate in the model. In this secondary EWAS, altogether 30 CpG sites (in 20 genes) were significantly (FDR < 0.05) associated with cotinine levels. Five (in five genes) of these 30 CpG sites were novel and not genome-wide significant in the discovery EWAS (cg04776445 in *MSX1*, cg05 335388 in *ABTB2*, cg10961758 in *EDIL3*, cg12817959 in *SSTR3*, and cg02306995 in *SEMA5B*), and the remaining 25 CpG sites (in 15 genes) overlapped with the discovery results. Half of the discovery EWAS hits (25 CpG sites) were no longer significant after controlling for nicotine and cotinine clearance rate. Manhattan and QQ plots for the secondary analysis are presented in Additional file 1: Figure S1.

Fig. 2 Manhattan and QQ plots showing epigenome-wide associations from discovery analysis. **a** QQ plot showing observed versus expected − log₁₀(P) for association at all loci. **b** Manhattan plot showing chromosomal locations of − log₁₀(P) for association at each locus. All CpG sites with FDR < 0.05 are highlighted in green and the top gene for each of the highlighted loci is labeled

Altogether, 55 unique CpG sites in 40 genes (22 of which were novel in 22 genes) were identified in our discovery and secondary EWAS analyses.

Pathway analyses

We performed gene-level pathway analyses for the 40 genes identified in our EWAS (discovery and secondary) and observed an enrichment of 31 specific gene networks and pathways (FDR < 0.05) (Additional file 2: Table S1). The top ones were colorectal cancer metastasis signaling, thrombin signaling, and purigenic receptor signaling. We also performed CpG-level gene ontology adjusting for non-uniform distribution of probes per gene on the methylation array (450 k) and observed no enrichment (FDR > 0.05) of biological processes (Additional file 2: Table S1).

Genetic association analysis

To assess the potential genetic influence on cotinine levels, we tested the association between cotinine levels and 46,780 polymorphic SNPs in all highlighted genes and observed 20 SNPs (in 9 genes) significantly associated (FDR < 0.05) with cotinine levels (Additional file 3: Table S2; QQ plot in Additional file 4: Figure S2). The strongest association was observed for rs187669467 in *ARHGAP44* (FDR = 2.6×10^{-02}) explaining 7% of variance in cotinine levels. We also tested the association of 375 genome-wide significant SNPs identified in the cotinine GWAS meta-analysis [26] in our discovery sample and observed 152 SNPs to be nominally associated (P < 0.05) with cotinine levels. The 20 SNPs highlighted in the current study were not genome-wide significant in the Ware et al. study [26].

Methylation quantitative trait loci analysis

We further assessed the influence of genotypes on methylation levels of the highlighted CpG sites. We performed meQTL analysis for 46,780 polymorphic SNPs within the 40 genes (with 50 kb flanking regions) identified in our discovery and secondary EWAS and identified 124 *cis* meQTLs (Additional file 5: Table S3) and 3898 *trans* meQTLs (Additional file 6: Table S4) (FDR < 0.05). *Cis* meQTLs were observed for 22 CpG sites (average 5 SNPs per CpG, range 1–23) and 73 SNPs (average 2 CpG sites per SNP, range 1–7) in 15 genes, while *trans* meQTLs were observed for all 55 CpGs (40 genes; average 70 SNPs per CpG, range 2–198) and 321 SNPs (31 genes; average 12 CpG sites per SNP, range 1–51). Figure 3 illustrates the ample *cis* and *trans* meQTLs identified with a circos plot (Circos.ca [32]).

Mediation analysis

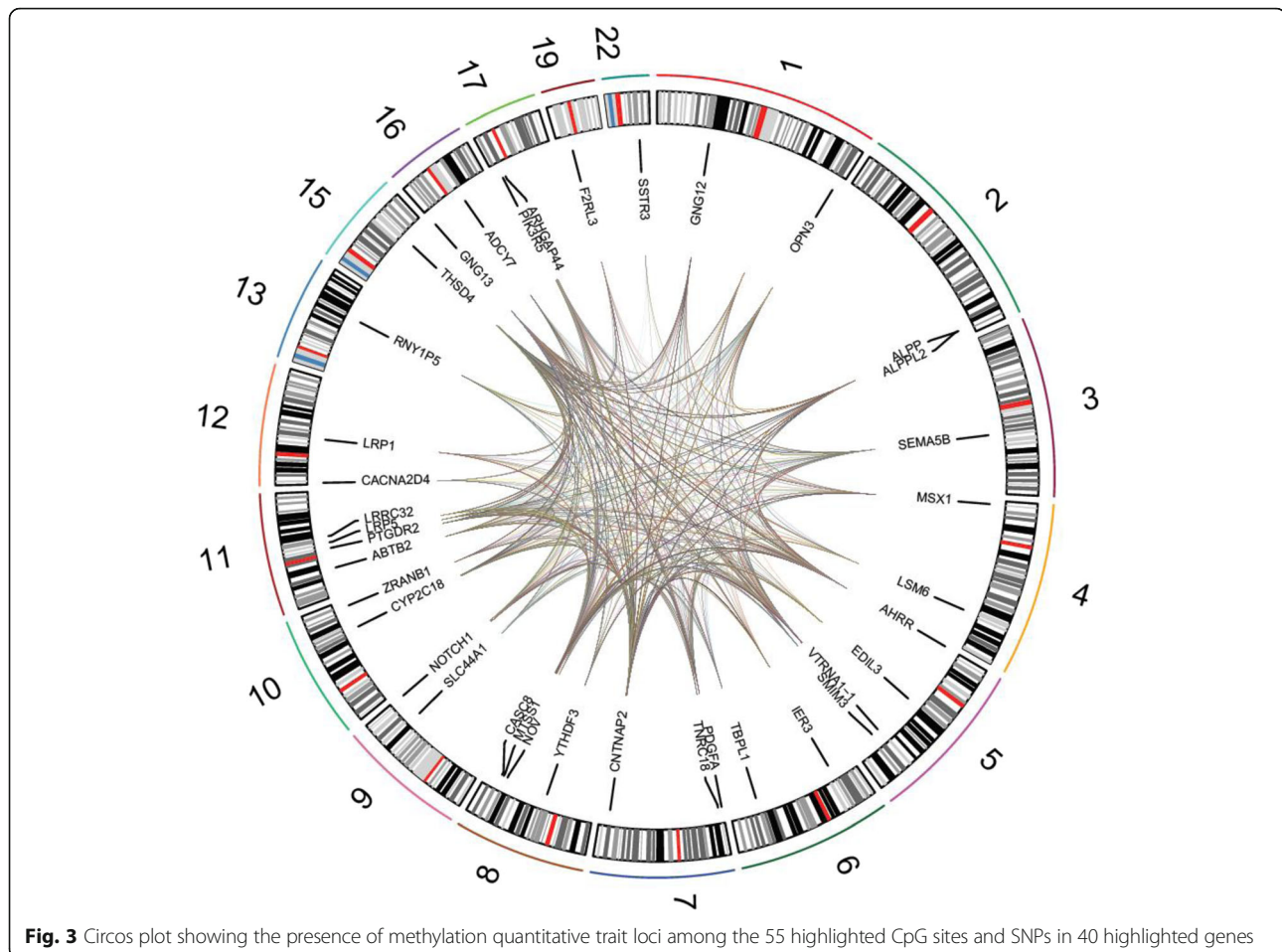At loci where genotype was associated with both methylation (meQTL) as well as cotinine levels (19 SNPs overlap), we performed mediation analysis implemented with causal inference test (CIT) to investigate the role of DNA methylation as a molecular mediator in the observed association between genotype and cotinine levels (Additional file 7: Figure S3). CIT is a widely used approach to infer causal indirect effects of a genetic variant on an outcome using a series of statistical tests (see the "Methods" section and Additional file 7: Figure S3) evaluating conditional independencies between covariates in order to distinguish a mediated effect of the genetic variant (G) on an outcome (cotinine; C) through an intermediate (methylation, M) from a reverse cause and a common cause (pleiotropic) effect [33, 34].

Mediation analysis of 19 SNPs that associated with both cotinine levels and methylation revealed a trend (P < 0.05) for methylation at seven CpG sites (in *LSM6*, *TNRC18*, *MTSS1*, *CASC8*, *PTGDR2*, *THSD4*, and *TMEM220-AS1*) being a molecular mediator between the effects of SNPs (in *GNG12*, *CNTNAP2*, *CYP2C18*, *ABTB2*, and *THSD4*) and cotinine levels (Additional file 8: Table S5). With a *cis* effect observed between rs532712997 and cg26589665 in *THSD4*, the remaining six CpG sites showed mediating effects with multiple SNPs in *trans*. These results indicate that alteration of methylation at these CpG sites may be a mediator in the causal pathway between genotype and observed cotinine levels, rather than a direct consequence of nicotine exposure/smoking. These seven CpG sites (Additional file 8: Table S5) bind multiple transcription factors (TF), include DNase I hypersensitive site (DHS), and overlap with enhancers (Additional file 9: Table S6 and Additional file 10: Table S7).

Replication

To replicate our findings, we assessed the association between serum cotinine levels and methylation at the 55 CpG sites identified in our discovery and secondary EWAS in a Dutch population-based sample from the Netherlands Twin Register (NTR) as well as in an independent Finnish population cohort DIetary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM). Among the 55 CpGs identified in our EWAS, 23 CpG sites were nominally associated (P < 0.05) with cotinine levels in NTR (N = 450) and 11 CpG sites in DILGOM (N = 79) with consistent direction of association, i.e., higher methylation associated with lower cotinine levels (Additional file 10: Table S7).

To replicate the trend observed in mediation analysis, we followed the same procedure as in FTC. For the 11 CpG sites where methylation levels were nominally associated (P < 0.05) with cotinine levels in the DILGOM sample (Additional file 10: Table S7), we performed genetic association and meQTL analysis. Altogether, eight SNPs were observed as meQTLs and were also nominally associated with cotinine levels (P < 0.05) in the

Gupta *et al. Clinical Epigenetics*   (2019) 11:1

Page 6 of 13

Fig. 3 Circos plot showing the presence of methylation quantitative trait loci among the 55 highlighted CpG sites and SNPs in 40 highlighted genes

DILGOM sample. We employed CIT to assess mediation between these eight SNPs and 11 CpG sites and observed a trend ($P < 0.05$) for mediation at seven CpG sites in *AHRR*, *SLC44A1*, *NOTCH1*, *F2RL3*, and *CASC8* (Additional file 11: Table S8) with cg25305703 in *CASC8* consistent with the mediation trend observed in FTC.

Discussion

Smoking has a major influence on methylation changes across the genome, as evidenced by numerous EWAS conducted in last few years and extensively elaborated in previous reviews [1, 35]. Cotinine is a reliable measure of recent nicotine exposure compared to self-reported smoking status or smoking quantity that are prone to reporting bias [24, 25]. In this study, we assessed the association of serum cotinine level with genome-wide DNA methylation in a sample of biochemically verified current smokers (serum cotinine > 4.85 ng/ml) from the Finnish Twin Cohort. With a half-life of 15–20 h, cotinine is a reliable indicator of recent nicotine exposure [25]. In our relatively small sample ($N = 310$), we replicated several previously reported findings (33 CpG sites), many of which were originally identified in studies

utilizing self-reported smoking measures in much larger samples ($N \sim 1000$ or more) [2–4], thus showing that a biomarker with biological proximity to the phenotype provides high statistical power.

Aside from the consistently reported association with CpG sites in *AHRR*, *F2RL3*, *ALPPL2*, *IER3*, and *MTSS1*, which have previously been discussed extensively in the context of smoking [3, 9, 10, 22], we identified novel loci pointing to smoking-related genes such as *THSD4*, *LSM6*, *CACNA2D4*, and *CYP2C18*. An interesting novel candidate identified in our EWAS is *THSD4* (thrombospondin type 1 domain containing 4). Genetic variants in this gene have been associated with several smoking-related phenotypes including nicotine dependence [36], smoking cessation success in clinical trials [37], and lung function affected by smoking [38]. Other genes highlighted in our EWAS with genetic variants associated with smoking-related phenotypes include *LSM6* and *CNTNAP2* associated with nicotine dependence [36] and *CACNA2D4* with pack years (indicator of lifelong accumulated smoking exposure) [39]. In our data, SNPs in *CNTNAP2* and *THSD4* genes were significantly associated with cotinine levels, in line with these

previous findings. Some of the pathways highlighted among the genes identified in our EWAS have been implicated in smoking-associated aberration in cancer pathogenesis [40, 41] and pulmonary disorders [42–44], highlighting the clinical significance of our findings.

The level of cotinine is influenced not only by intake (such as smoking pattern and smoking quantity) but also by the rate of clearance of cotinine [28]. Therefore, we utilized a GRS for NMR (a proxy for nicotine and cotinine clearance rate) to account for the impact of this in our secondary analysis and identified five additional novel loci but also observed that half of the associations from our discovery EWAS were no longer significant. A plausible explanation for this observation could be that adding the GRS adjusts for the genetic variation of nicotine metabolism rate (normal vs faster metabolizers) among the individuals analyzed. Associations which disappear when the NMR GRS is added to the model may be related to the impact of nicotine metabolism rate on nicotine intake or nicotine/cotinine metabolism. Among such associations (no longer genome-wide significant after adjusting for the NMR GRS) are CpG sites in genes *AHRR* and *CYP2C18*, members of the metabolic pathways involved in nicotine and bupropion (prescription medication for smoking cessation [45]) degradation and xenobiotic metabolism [46]. *CYP2C18*, a member of the cytochrome P450 (CYP) family, and *AHRR*, a regulator of CYP genes, are both involved in xenobiotic (which includes components of cigarette smoke) metabolism. Altered expression of *CYP2C18* in the context of smoking has been reported previously [47]; however, the role of methylation as a plausible regulator in such observed alterations warrants further experimental investigation. It should be noted that although just three SNPs used in the NMR GRS account for a relatively high proportion (~ 30%) of variance in NMR, a large portion of variance remains unaccounted for and this is a major limitation of our study. The three independent SNPs constituting the GRS were identified in the Finnish population sample and may not capture the same effects across other populations (as we observed with the NTR replication sample).

As some of the CpGs identified in our EWAS were annotated to genes that have previously been linked to smoking-related phenotypes (36–39), we examined the role of genetic variants in the genes identified in our EWAS. It is noteworthy that a plethora of associations between the methylation levels of highlighted CpG sites and SNPs were observed in both *cis* and *trans.* We can speculate a cross-genome interplay between nicotine exposure-associated DNA methylation and genetic variation in the highlighted genes [48, 49]. Some of the *cis* meQTLs also act in *trans* with other CpG sites. Most of the CpG sites bind multiple TFs, overlapping enhancers or DHS sites hinting at global-directed networks at play

[48, 50]; however, these may be independent effects occurring simultaneously. It should also be noted that we tested meQTLs only among genes highlighted in our analysis potentially missing other *cis* and *trans* meQTLs in the whole genome. Interestingly, for the 19 meQTLs that were also associated with cotinine levels, we observed a trend of molecular mediation by methylation at seven CpG sites. These results suggest that alterations in methylation of these nicotine exposure-associated CpG sites might not always be a consequence of smoking as commonly suggested, but rather methylation at such CpG sites may act as a causal mediator (molecular mechanism) in regulating the effect of genetic variation (in genes that modulate nicotine intake and/or nicotine and cotinine metabolism rate) on cotinine levels or smoking behavior. We observed support for these findings in an independent replication sample (DILGOM). These findings are particularly interesting as these SNPs do not overlap SNPs in the probes. We observed *cis* mediation between CpG sites and SNPs in *THSD4* (FTC) and *AHRR* (DILGOM) both involved in smoking [36–38, 51], while *trans* mediation in CpG sites that bind multiple TFs or overlap DHS sites and enhancers providing plausible action mechanism for molecular mediation. These findings are crucial and provide caution to interpretation of smoking-related EWAS but require further experimental evidence.

Cotinine, even though a reliable measure of nicotine exposure, is not specific to tobacco smoking. Other potential sources of nicotine, such as smokeless tobacco (e.g., snus), e-cigarettes, and/or nicotine replacement therapy, could also contribute to cotinine levels; however, the likelihood of such alternative sources in our sample was low (see the "Methods" section). There have been only a few EWAS so far using cotinine as a continuous phenotype [51, 52], both of which were performed in samples including smokers and non-smokers. When considering replication of prior findings, it should be noted that all the previous EWAS we used to assess the novelty of our findings (see "Methods") were conducted in samples including both smokers and non-smokers. We cannot rule out that our novel findings might be specific to current or heavy smoking as we limited our analyses to self-reported current smokers with elevated cotinine levels only. Having included only current smokers, the range of variability is limited (in contrast to studies including both smokers and non-smokers) resulting in smaller effect sizes observed. Similar results (small effects) were noted by Zhang et al. [52] and Su et al. [22] when restricting analysis to current smokers only. Although we observed a great overlap with previous findings, non-overlapping associations could be due to differences in study design and sample (smokers and non-smokers in previous publications versus current smokers only in our study), population specificity of

methylation levels [12], and precision of cotinine measurement technique (mass spectrometry versus immunoassay) [53]. The lack of replication of the novel associations identified in our EWAS in the NTR and DILGOM samples may be due to several aforementioned factors including population specificity, differences in precision of cotinine measurement (mass spectrometry (FTC and DILGOM) and immunoassay (NTR)) and small sample size (DILGOM N = 79), or small and inconsistent effects of novel associations. Even though we replicated several prior associations between smoking and methylation, the novel associations identified in our study, while biologically meaningful, would require replication in other larger samples.

Conclusions

In conclusion, we examined the epigenetic signature of nicotine exposure in a sample of biochemically verified current smokers and identified several novel loci, including genes involved in nicotine degradation and metabolism. Our study is the first cotinine EWAS in which the rate of nicotine clearance is accounted for, allowing us to discover smoking-pertinent novel associations which may be specific to regular/heavy smoking. We further expose genetic influences on methylation and provide suggestive evidence for the role of methylation as a molecular mediator between genetic variation and cotinine levels, as opposed to a direct consequence of nicotine exposure in some of the genes.

Methods

Study sample

The study sample of cotinine verified current smokers comes from the Finnish Twin Cohort (FTC) [54], a population-based longitudinal study designed to examine genetic and environmental determinants of health-related behaviors. A total of 310 individuals (51 full monozygotic pairs, 44 full dizygotic pairs, and 120 twin individuals without a co-twin) were included in our discovery analysis. Genome-wide DNA methylation was assessed with Infinium HumanMethylation450 BeadChip in peripheral blood using standard protocols [55]. The average age in the sample was 29.5 years (SD 14.2, range 21–69), and 52% were females. Genotype data imputed to 1000Genomes phase 3 (processing and imputation of the genotype data has been described in detail previously [31]) was available for 304 of these individuals, which were consequently included in the secondary analysis.

Phenotype data

Current smokers were selected based on the threshold of serum cotinine above 4.85 ng/ml, as suggested by Benowitz et al. [56]. Selecting current smokers with a threshold of cotinine > 4.85 ng/ml instead of 10 ng/ml

was done to maximize sample size available for discovery analysis, and sensitivity analysis (results not shown) was performed to ensure that using a higher threshold did not affect the results and subsequent inference. Cotinine was measured from frozen serum samples using high-precision LC-MS/MS at University of Toronto, Canada (N = 242) [57] and at the Metabolomics Unit, Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland (N = 68). The protocols and instrument parameters have been described previously [31, 58, 59]. Average cotinine in our discovery sample (N = 310) was 192.7 ng/ml (median = 178.8, SD = 148.4, range = 5.1–820.5). In our discovery sample, two individuals (with cotinine levels above 100 ng/ml) reported using nicotine replacement therapy and none declared use of smokeless tobacco. E-cigarette use was non-existent in Finland at the time of data collection.

Replication samples

DILGOM

To replicate our findings, we utilized a Finnish population cohort DILGOM (DIetary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome) [60, 61]. Briefly, DILGOM originates from the population-based national FINRISK 2007 study and includes a total of 631 unrelated Finnish individuals aged 25–74 years from the Helsinki area. A total of N = 79 individuals with methylation data from the peripheral blood, genome-wide genotype data imputed to 1000Genome phase I, and serum cotinine > 4.85 ng/ml (current smokers) were analyzed. The average age of the sample was 51 years (SD 13.5, range 25–72), and 44% were females. Cotinine was measured with gas chromatograph mass spectrometry at the Laboratory of Analytical Biochemistry at the Institute of Health and Welfare, Helsinki, Finland, as described earlier [57]. All 79 individuals were self-reported current smokers. Average cotinine levels were 134.2 ng/ml (median = 129.6, SD = 98.0, range 5.0–408.8).

Netherlands twin registry

The subjects participated in longitudinal survey studies from the Netherlands Twin Register (NTR) [62] and in the NTR biobank project [63]. Data from 450 samples from 446 individuals (58% females) were analyzed. For four individuals, two longitudinal blood samples were included. The blood sampling procedure [63], Illumina 450 k methylation data [64], and nicotine measurements have been described in detail previously [65]. A standard threshold of plasma cotinine > 50 ng/ml (for immunoassay measurements) was used to select current smokers [66]. Average cotinine levels were 309.6 ng/ml (median = 229.0, SD = 300.5, range = 50.0–2329.0).

DNA methylation data processing and analysis

DNA methylation data was preprocessed and normalized using the pipeline suggested by Lehne et al. [67] implemented using R packages "minfi" [68] and "limma" [69]. We modified the pipeline to accommodate the relatedness in our sample. Altogether the quality control involved exclusion of probes with detection $P > 1 \times 10^{-16}$, bead count < 3 (estimated with "wateRmelon" R package [70]) and call rate < 98%. Samples with sample call rate < 98% and sex mismatch were further excluded. After quality control, 418,302 probes and $N = 310$ samples remained (FTC data). Quantile normalization stratified on probe type, color channel, and probe subtypes was performed to obtain a methylation beta matrix. The first 30 principal components (PCs) from the control probes were calculated to adjust for technical bias. White blood cell subtypes (CD8T, CD4T, Natural Killer cells, B cell, monocyte, and granulocyte) were estimated using "FlowSorted.Blood.450 k" [71] within the "minfi" R package based on a modified version of the Houseman algorithm [72] and included in the regression model. Intermediary residuals were estimated by regressing the methylation beta values on age, sex, body mass index (BMI), 30 control probe PCs, and white blood cell types. PCs from these intermediary residuals were further included in the final regression model to adjust for any unaccounted global covariation. To analyze the association between methylation (dependent variable) at each CpG site and serum cotinine levels (explanatory variable), we used linear mixed effects model with "lmer" function implemented in "lme4" R package [73]. Age, sex, BMI, batch variable (site of cotinine measurement), 30 control probe PCs, and 10 intermediary residual PCs were included as covariates (fixed effects), while family identifier and zygosity were included as random effects to account for the relatedness in our sample. To avoid spurious associations, we further applied ad hoc exclusion of probes reported by Zhou et al. [74] to affect beta methylation at the CpG (probes with non-unique mapping, inconsistent extension base, SNP in the extension base, and overlap with any known SNPs with global minor allele frequency (MAF) and MAF in a Finnish population > 1%).

We applied false discovery rate (FDR) adjustment to the p values for multiple testing correction and considered FDR-adjusted $P < 0.05$ as statistically significant. Manhattan and QQ plots were produced using the R package "qqman" [75]. Lambda value estimates for the QQ plot were calculated using the "estlambda" function in R package "GenABEL" [76]. For replication of our results, analysis pipeline was identical for the replication sample NTR. For the DILGOM sample, analysis was identical except that instead of a mixed effects model, a linear model was used as all individuals were unrelated.

Secondary analysis

To account for the individual rate of nicotine metabolism, we included the GRS for NMR in the regression model (described above) as an additional covariate. We extracted the genotypes for the three independent SNPs (rs56113850, rs113288603, and rs12461964) and calculated the weighted mean of minor allele counts to calculate the GRS, as described previously [31]. One of the top variants (indel esv2663194) identified in the NMR GWAS meta-analysis [31] was not available in our samples (discovery and replication). GRS calculation and analysis pipeline was identical for the replication samples.

Literature search

We used the same literature search strategy as Joehanes et al. [2]. Briefly, PubMed literature database was queried using medical subject heading (MeSH) terms of DNA methylation and smoking. To limit duplication of efforts, we used the supplementary data on literature search used by Joehanes et al. which included articles until 2015 and applied additional filter on year (starting 2015) along with a filter for species (human). We reviewed the abstracts to determine if the studies were as follows: (1) performed in healthy adult human populations (also excluding maternal smoking), (2) sample type analyzed was peripheral blood, (3) studies assessed genome-wide methylation only, and (4) public reporting of P values and gene annotations for the CpGs identified was available. A total of 24 publications met all the inclusion criteria and are listed in the Additional file 12: Table S9. A compiled list of CpGs (19,208 CpG sites) and associated gene names (8582 genes) from all 24 studies were utilized to assess novelty of our findings.

Annotation and pathway analysis

All statistically significant CpG sites were annotated using the data aggregated by Zhou et al. [74]. For the probes that did not have a gene name listed, the name of the nearest gene was fetched from Ensembl [77]. This gene list was analyzed with the Ingenuity Pathway Analysis software (IPA; Ingenuity® Systems, https://www.qiagenbioinformatics.com/) to identify gene networks at play. We applied multiple testing correction and an enrichment for genes functioning in signaling and metabolic pathways [78] was considered significant when FDR < 0.05. A total of 38 of the 40 highlighted genes were recognized by IPA. We also performed pathway analysis (Gene Ontology) to assess enrichment of biological processes among the top 55 CpG sites while taking into account the non-uniform CpG probe distribution per gene [79] on the 450 k array with R package "missMethyl" [80] and corrected for multiple testing and considered FDR < 0.05 significant.

Association of cotinine with genetic variants

As the genes identified in our EWAS harbor genetic variants associated with smoking behavior phenotypes, such as nicotine dependence [36], we examined whether SNPs in the genes identified in our EWAS analysis associate with cotinine levels. We performed association analysis using mixed effects model implemented as "lmer" function in "lme4" R package. Cotinine levels were treated as dependent variable, while genotype dosage (coded as 0, 1 or 2 for copies of effect allele), age, sex, and BMI were included as fixed effects in the model. Family and zygosity were included as random effects to account for the relatedness in our sample. Only polymorphic SNPs ($N = 46,780$) were tested for association with cotinine levels. Analysis was otherwise identical for the DILGOM sample except that linear model was used instead of mixed effects model, as all individuals were unrelated. To estimate proportion of variance explained by individual SNPs, we subtracted the R^2 of the model including only covariates (age, sex, BMI) from the R^2 of the model also including the SNP.

Methylation quantitative trait loci analysis

To examine the association between methylation and SNPs within the 40 genes (with 50 kb flanking regions; gene boundaries are available in Additional file 13: Table S10), we used R package "MatrixEQTL" [81]. Only polymorphic SNPs ($N = 46,780$) were tested using linear model setting with a *cis* distance of 2.5 Mb. The longest gene was approximately 2.4 Mb (Additional file 13: Table S10); thus, a *cis* distance of 2.5 Mb was chosen to ensure that all possible combinations of SNPs, and the CpG within a gene are tested. All association tests between remaining SNP and CpGs were considered *trans*. Genotypes were coded as copies of effect allele (0, 1, or 2) and methylation data was extracted for the 55 probes from the CPACOR normalized data set used in the EWAS. We included, age, sex, BMI, white blood cell counts, control probe PCs, and cotinine as covariates in the model. In addition, owing to the relatedness in our sample, we included a covariance matrix based on CpG sites being tested in the model. Results of meQTL analysis were visualized using R package "RCircos" [82].

Mediation analysis

For the SNPs that were significantly associated with both cotinine levels and methylation, we investigated the potential of methylation (M) being a mediator between the effects of genetic variant (G) on outcome cotinine levels (C) using causal inference test (CIT) [83]. In CIT, four conditions are tested to assess consistent causal mediation as described by Millstein et al. [83]. Conditions for CIT are (1) G is associated with C, (2) G is associated with M conditional on C, (3) M is associated with C conditional on G, and (4) G is independent of C conditional on M. We implemented CIT analysis with R package "cit" and function "cit.cp" with 50 permutations for conditional analysis and included, age, sex, BMI, family, and zygosity as covariates. We report the overall *p* value (omnibus *p* value) based on collective conditional analysis and considered $P < 0.05$ as nominally significant.

Additional files

Additional file 1: Figure S1. Manhattan and QQ plots showing epigenome-wide association results from secondary analysis when accounting for the rate of nicotine metabolism using a GRS. (A) QQ plot showing observed versus expected $-\log_{10}(P)$ for association at all loci. (B) Manhattan plot showing chromosomal locations of $-\log_{10}(P)$ for association at each locus. All CpG sites with FDR < 0.05 are highlighted in green and the top gene for each of the highlighted loci is labeled. (PDF 171 kb)

Additional file 2: Table S1. Gene network (IPA) and pathway analysis (Gene Ontology) of EWAS results. (XLS 3125 kb)

Additional file 3: Table S2. SNPs in the 40 top genes significantly associated (FDR *p* < 0.05) with cotinine levels. (XLS 14 kb)

Additional file 4: Figure S2. QQ plot for genetic association analysis of cotinine levels and 46,780 SNPs in 40 genes. (PDF 17 kb)

Additional file 5: Table S3. *Cis* acting methylation quantitative trait loci in the 40 genes highlighted in the EWAS. (XLS 31 kb)

Additional file 6: Table S4. *Trans* acting methylation quantitative trait loci in the 40 genes highlighted in the EWAS. (XLS 687 kb)

Additional file 7: Figure S3. Mediation analysis to assess whether DNA methylation is a causal mediator to the observed association between genetic variants and cotinine levels. (PDF 588 kb)

Additional file 8: Table S5. Mediation analysis omnibus *P* values for the 19 meQTLs associated with serum cotinine levels. (XLS 33 kb)

Additional file 9: Table S6. Transcription factor binding site information for 55 highlighted CpG sites based on ENCODE data. (XLS 96 kb)

Additional file 10: Table S7. CpG sites showing significant (FDR *p* < 0.05) association with cotinine levels in regular smokers in FTC and replication results for top 55 CpG sites in NTR and DILGOM. (XLS 42 kb)

Additional file 11: Table S8. Results for CIT performed on DILGOM sample replicated CpG sites and SNPs associated with their methylation levels as well as cotinine levels. (XLS 10 kb)

Additional file 12: Table S9. Literature search results for smoking EWAS. (XLS 9 kb)

Additional file 13: Table S10. Gene boundaries, based on Ensembl transcripts for the 40 genes highlighted in the EWAS. (XLS 10 kb)

Availability of data and materials
Due to the consent given by study participants, FTC data cannot be made publicly available. Data are available through the Institute for Molecular Medicine Finland (FIMM) Data Access Committee (DAC) for authorized researchers who have IRB/ethics approval and an institutionally approved study plan. For more details, please contact the FIMM DAC (fimm-dac@helsinki.fi). Data from DILGOM cohort can be accessed by application to the THL Biobank (thl.fi/web/thl-biobank). 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.

Authors' contributions
The study design was conceived and evolved by MO, AL, JK, and RG. RG analyzed and interpreted the data and wrote the paper. JVD, AA, and YF analyzed the replication data. JK, RFT, VV, and DIB generated the phenotype data. JK, AL, and DIB acquired the genotype data. JK, MO, and DIB acquired the DNA methylation data. AL, MO, JK, RFT, TK, and JVD critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Written informed consent was obtained from all subjects who gave DNA samples in accordance to the current edition of the Declaration of Helsinki and the collection of blood samples followed the recommendations given in the Declaration of Helsinki and its amendments. Data collection for the Finnish Twin Cohort has been approved by the Hospital District of Helsinki and Uusimaa, the ethics committee for epidemiology and public health (HUS-113-E3–01, HUS-346-E0–05, HUS 136/E3/01, HUS 270/13/03/01/2008, HUS 154/13/03/00/11). DILGOM participants provided written informed consent and the Ethics committee of Helsinki University Central Hospital has provided the approval (#46 20.2.2007, #90 3.4.2007, #332 13.03.2013). The protocol was designed and performed according to the principles of the Helsinki Declaration and was approved by the coordinating ethics committee of the hospital district of Helsinki and Uusimaa, Finland. Informed consent was obtained from all participants of NTR study. 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).

Consent for publication
Not applicable.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [2]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [3]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, and Departments of Pharmacology & Toxicology and Psychiatry, University of Toronto, Toronto, Canada. [4]Department of Public Health, University of Helsinki, Helsinki, Finland. [5]Department of Pathology, University of Helsinki, Helsinki, Finland.

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. 2015;7:113.
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(5):436–47.
3. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8(5):599–618.
4. Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. Hum Mol Genet. 2015;24(8):2349–59.
5. Dogan MV, Beach SRH, Philibert RA. Genetically contextual effects of smoking on genome wide DNA methylation. Am J Med Genet B Neuropsychiatr Genet. 2017;174(6):595–607.
6. 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.
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. Besingi W, Johansson A. Smoke-related DNA methylation changes in the etiology of human disease. Hum Mol Genet. 2014;23(9):2290–7.
9. Sayols-Baixeras S, Lluis-Ganella C, Subirana I, Salas LA, Vilahur N, Corella D, et al. Identification of a new locus and validation of previously reported loci showing differential methylation associated with smoking. The REGICOR study. Epigenetics. 2015;10(12):1156–65.
10. 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.
11. 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.
12. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics. 2014;6(1):4.
13. Lee MK, Hong Y, Kim SY, London SJ, Kim WJ. DNA methylation and smoking in Korean adults: epigenome-wide association study. Clin Epigenetics. 2016;8:103.
14. 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.
15. 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.
16. Philibert RA, Beach SR, Brody GH. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. Epigenetics. 2012; 7(11):1331–8.
17. 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.
18. Zaghlool SB, Al-Shafai M, Al Muftah WA, Kumar P, Falchi M, Suhre K. Association of DNA methylation with age, gender, and smoking in an Arab population. Clin Epigenetics. 2015;7:6.
19. Flanagan JM, Brook MN, Orr N, Tomczyk K, Coulson P, Fletcher O, et al. Temporal stability and determinants of white blood cell DNA methylation in

Gupta *et al. Clinical Epigenetics*   (2019) 11:1

Page 12 of 13

the breakthrough generations study. Cancer Epidemiol Biomark Prev. 2015;24(1):221–9.

20. Allione A, Marcon F, Fiorito G, Guarrera S, Siniscalchi E, Zijno A, et al. Novel epigenetic changes unveiled by monozygotic twins discordant for smoking habits. PLoS One. 2015;10(6):e0128265.

21. Qiu W, Wan E, Morrow J, Cho MH, Crapo JD, Silverman EK, et al. The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study. Epigenetics. 2015;10(11):1064–73.

22. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, et al. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. PLoS One. 2016;11(12):e0166486.

23. Park SL, Patel YM, Loo LWM, Mullen DJ, Offringa IA, Maunakea A, et al. Association of internal smoking dose with blood DNA methylation in three racial/ethnic populations. Clin Epigenetics. 2018;10(1):110.

24. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res. 2009;11(1):12–24.

25. Benowitz NL. Cotinine as a biomarker of environmental tobacco smoke exposure. Epidemiol Rev. 1996;18(2):188–204.

26. Ware JJ, Chen X, Vink J, Loukola A, Minica C, Pool R, et al. Genome-wide meta-analysis of cotinine levels in cigarette smokers identifies locus at 4q13. 2 Sci Rep. 2016;6:20092.

27. Hukkanen J, Jacob P 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. Pharmacol Rev. 2005;57(1):79–115.

28. Zhu AZ, Renner CC, Hatsukami DK, Swan GE, Lerman C, Benowitz NL, et al. The ability of plasma cotinine to predict nicotine and carcinogen exposure is altered by differences in CYP2A6: the influence of genetics, race, and sex. Cancer Epidemiol Biomark Prev. 2013;22(4):708–18.

29. Dempsey D, Tutka P, Jacob P 3rd, Allen F, Schoedel K, Tyndale RF, et al. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. Clin Pharmacol Ther. 2004;76(1):64–72.

30. Strasser AA, Benowitz NL, Pinto AG, Tang KZ, Hecht SS, Carmella SG, et al. Nicotine metabolite ratio predicts smoking topography and carcinogen biomarker level. Cancer Epidemiol Biomark Prev. 2011;20(2):234–8.

31. Loukola A, Buchwald J, Gupta R, Palviainen T, Hallfors J, Tikkanen E, et al. A genome-wide association study of a biomarker of nicotine metabolism. PLoS Genet. 2015;11(9):e1005498.

32. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.

33. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Hum Mol Genet. 2016;25(R2):R149–R56.

34. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. BMC Genet. 2009;10:23.

35. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. Front Genet. 2013;4:132.

36. Drgon T, Montoya I, Johnson C, Liu QR, Walther D, Hamer D, et al. Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers. Mol Med. 2009;15(1–2):21–7.

37. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, et al. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. Arch Gen Psychiatry. 2008;65(6):683–93.

38. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, et al. Genome-wide association study identifies five loci associated with lung function. Nat Genet. 2010;42(1):36–44.

39. Sung YJ, de Las Fuentes L, Schwander KL, Simino J, Rao DC. Gene-smoking interactions identify several novel blood pressure loci in the Framingham heart study. Am J Hypertens. 2015;28(3):343–54.

40. Ma Y, Li MD. Establishment of a strong link between smoking and cancer pathogenesis through DNA methylation analysis. Sci Rep. 2017;7(1):1811.

41. Aithal MG, Rajeswari N. Role of notch signalling pathway in cancer and its association with DNA methylation. J Genet. 2013;92(3):667–75.

42. Cicko S, Lucattelli M, Muller T, Lommatzsch M, De Cunto G, Cardini S, et al. Purinergic receptor inhibition prevents the development of smoke-induced lung injury and emphysema. J Immunol. 2010;185(1):688–97.

43. Li S, Hu X, Wang Z, Wu M, Zhang J. Different profiles of notch signaling in cigarette smoke-induced pulmonary emphysema and bleomycin-induced pulmonary fibrosis. Inflamm Res. 2015;64(5):363–71.

44. Tilley AE, Harvey BG, Heguy A, Hackett NR, Wang R, O'Connor TP, et al. Down-regulation of the notch pathway in human airway epithelium in association with smoking and chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2009;179(6):457–66.

45. Wilkes S. The use of bupropion SR in cigarette smoking cessation. Int J Chron Obstruct Pulmon Dis. 2008;3(1):45–53.

46. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46(5):2699.

47. Van Dyck E, Nazarov PV, Muller A, Nicot N, Bosseler M, Pierson S, et al. Bronchial airway gene expression in smokers with lung or head and neck cancer. Cancer Med. 2014;3(2):322–36.

48. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017;49(1):131–8.

49. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 2016;17:61.

50. Stueve TR, Li WQ, Shi J, Marconett CN, Zhang T, Yang C, et al. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. Hum Mol Genet. 2017;26(15):3014–27.

51. Philibert RA, Beach SR, Lei MK, Brody GH. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. Clin Epigenetics. 2013;5(1):19.

52. Zhang Y, Florath I, Saum KU, Brenner H. Self-reported smoking, serum cotinine, and blood DNA methylation. Environ Res. 2016;146:395–403.

53. Pesce A, Rosenthal M, West R, West C, Crews B, Mikel C, et al. An evaluation of the diagnostic accuracy of liquid chromatography-tandem mass spectrometry versus immunoassay drug testing in pain patients. Pain Physician. 2010;13(3):273–81.

54. Kaprio J. The Finnish twin cohort study: an update. Twin Res Hum Genet. 2013;16(1):157–62.

55. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288–95.

56. Benowitz NL, Bernert JT, Caraballo RS, Holiday DB, Wang J. Optimal serum cotinine levels for distinguishing cigarette smokers and nonsmokers within different racial/ethnic groups in the United States between 1999 and 2004. Am J Epidemiol. 2009;169(2):236–48.

57. Tanner JA, Novalen M, Jatlow P, Huestis MA, Murphy SE, Kaprio J, et al. Nicotine metabolite ratio (3-hydroxycotinine/cotinine) in plasma and urine by different analytical methods and laboratories: implications for clinical implementation. Cancer Epidemiol Biomark Prev. 2015;24(8):1239–46.

58. Nikkanen J, Forsstrom S, Euro L, Paetau I, Kohnz RA, Wang L, et al. Mitochondrial DNA replication defects disturb cellular dNTP pools and remodel one-carbon metabolism. Cell Metab. 2016;23(4):635–48.

59. Roman-Garcia P, Quiros-Gonzalez I, Mottram L, Lieben L, Sharan K, Wangwiwatsin A, et al. Vitamin B(1)(2)-dependent taurine synthesis regulates growth and bone mass. J Clin Invest. 2014;124(7):2988–3002.

60. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol Syst Biol. 2010;6:441.

61. Inouye M, Silander K, Hamalainen E, Salomaa V, Harald K, Jousilahti P, et al. An immune response network associated with blood lipid levels. PLoS Genet. 2010;6(9):e1001113.

62. Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JH, Draisma HH, et al. The adult Netherlands twin register: twenty-five years of survey and biological data collection. Twin Res Hum Genet. 2013;16(1):271–81.

63. Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks AI, Estourgie-van Burk GF, et al. The Netherlands twin register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet. 2010;13(3):231–45.

64. van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun. 2016;7:11115.

65. Bot M, Vink JM, Willemsen G, Smit JH, Neuteboom J, Kluft C, et al. Exposure to secondhand smoke and depression and anxiety: a report from two studies in the Netherlands. J Psychosom Res. 2013;75(5):431–6.

66. Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. Hum Mol Genet. 2009;18(20):4007–12.

67. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. 2015;16:37.
68. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014; 30(10):1363–9.
69. Smyth GK. Limma: linear models for microarray data, Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer; 2005. p. 397–420.
70. Pidsley R, Wong YCC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013;14:293.
71. 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.
72. 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 Bioinf. 2012;13:86.
73. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. 2015. 2015;67(1):48.
74. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017;45(4):e22.
75. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014;10:005165.
76. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23(10):1294–6.
77. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, García Girón C, Hourlier T, Howe K, Kähäri A, Kokocinski F, Martin FJ, Murphy DN, Nag R, Ruffier M, Schuster M, Tang YA, Vogel JH, White S, Zadissa A, Flicek P, Searle SM. The Ensembl gene annotation system. Database (Oxford). 2016;2016. https://doi.org/10.1093/database/baw093.
78. Kramer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics. 2014;30(4):523–30.
79. Geeleher P, Hartnett L, Egan LJ, Golden A, Raja Ali RA, Seoighe C. Gene-set analysis is severely biased when applied to genome-wide methylation data. Bioinformatics. 2013;29(15):1851–7.
80. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286–8.
81. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353–8.
82. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 2013;14:244.
83. Millstein J, Chen GK, Breton CV. Cit: hypothesis testing software for mediation analysis in genomic applications. Bioinformatics. 2016; 32(15):2364–5.