Check for updates

ARTICLE Genome-wide DNA methylation differences in nucleus accumbens of smokers vs. nonsmokers

Christina A. Markunas¹, Stephen A. Semick², Bryan C. Quach ¹, Ran Tao², Amy Deep-Soboslay², Megan U. Carnes¹, Laura J. Bierut ³, Thomas M. Hyde^{2,4,5}, Joel E. Kleinman^{2,4}, Eric O. Johnson ¹⁶, Andrew E. Jaffe ^{2,4,7,8} and Dana B. Hancock ¹

Numerous DNA methylation (DNAm) biomarkers of cigarette smoking have been identified in peripheral blood studies, but because of tissue specificity, blood-based studies may not detect brain-specific smoking-related DNAm differences that may provide greater insight as neurobiological indicators of smoking and its exposure effects. We report the first epigenome-wide association study (EWAS) of smoking in human postmortem brain, focusing on nucleus accumbens (NAc) as a key brain region in developing and reinforcing addiction. Illumina HumanMethylation EPIC array data from 221 decedents (120 European American [23% current smokers], 101 African American [26% current smokers]) were analyzed. DNAm by smoking (current vs. nonsmoking) was tested within each ancestry group using robust linear regression models adjusted for age, sex, cell-type proportion, DNAm-derived negative control principal components (PCs), and genotype-derived PCs. The resulting ancestry-specific results were combined via meta-analysis. We extended our NAc findings, using published smoking EWAS results in blood, to identify DNAm smoking effects that are unique (tissue-specific) vs. shared between tissues (tissue-shared). We identified seven CpGs (false discovery rate < 0.05), of which three CpGs are located near genes previously indicated with blood-based smoking DNAm biomarkers: ZIC1, ZCCHC24, and PRKDC. The other four CpGs are novel for smoking-related DNAm changes: ABLIM3, APCDD1L, MTMR6, and CTCF. None of the seven smoking-related CpGs in NAc are driven by genetic variants that share association signals with predisposing genetic risk variants for smoking, suggesting that the DNAm changes reflect consequences of smoking. Our results provide the first evidence for smokingrelated DNAm changes in human NAc, highlighting CpGs that were undetected as peripheral biomarkers and may reflect brainspecific responses to smoking exposure.

Neuropsychopharmacology (2021) 46:554-560; https://doi.org/10.1038/s41386-020-0782-0

INTRODUCTION

Cigarette smoking remains the leading cause of preventable death, resulting in more than 7 million deaths annually worldwide [1]. Despite cigarette smoking causing significant morbidity and mortality and ~70% of adult smokers wanting to quit [2], 14% of all U.S. adults remain current smokers [2].

Becoming a regular smoker involves multiple stages, with evidence for heritability at each stage (37%–55% for smoking initiation [3, 4] 46%–59% for smoking persistence [3], and up to 75% for nicotine dependence [4]). Genome-wide association studies (GWAS) have established nicotinic acetylcholine receptor and other loci for varied smoking behaviors [5], and, most recently, implicated >400 loci using Ns up to 1.2 million from the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) [6]. Despite the strong heritability of smoking behaviors and GWAS successes to date, much remains unknown about the neurobiology underlying the trajectory of smoking from predisposing genetic loci [6, 7] to consequences of smoking exposure having downstream neurobiological effects that reinforce smoking behaviors.

To better characterize the biological determinants and consequences of smoking, gene regulation studies (e.g., epigenetic or gene expression changes) in disease-relevant brain tissues are needed. Previously, we reported differential RNA expression (RNAexp) in postmortem dorsolateral prefrontal cortex samples of smokers vs. non-smokers [8]. Here, we focus on nucleus accumbens (NAc), another key tissue in addiction through its role in cognitive processing of motivation, pleasure, and reward/ reinforcement learning [9, 10]. The NAc influences the dopamine reward system in the binge/intoxication stage of the addiction cycle [11]. Its disruption has been described as lying at the core of drug addiction [12].

Numerous studies [13–15] have identified DNA methylation (DNAm) changes related to cigarette smoking in peripheral blood. However, because of the tissue specificity of gene regulation, the neurobiological relevance of these changes is unknown. In addition, these blood-based studies may miss biologically-relevant smoking-related DNAm differences that are only detectable in brain. Here, we report the first epigenome-wide association study

¹GenOmics, Bioinformatics, and Translational Research Center, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC, USA; ²Lieber Institute for Brain Development (LIBD), Baltimore, MD, USA; ³Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA; ⁴Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ⁵Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ⁶Fellow Program, RTI International, Research Triangle Park, NC, USA; ⁷Department of Mental Health and Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA and ⁸Department of Human Genetics and Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA Correspondence: Dana B. Hancock (dhancock@rti.org)

These authors contributed equally: Andrew E. Jaffe, Dana B. Hancock

Received: 29 March 2020 Accepted: 21 July 2020 Published online: 30 July 2020

(EWAS) of smoking in human postmortem brain samples, focusing on NAc. Using RNAexp data from the same brain samples, we evaluated smoking-related CpGs for association with nearby RNAexp levels. We extended our DNAm findings to blood to identify tissue-specific and tissue-shared smoking DNAm effects, and applied a blood-derived smoking DNAm polyepigenetic score [16] to our brain data to predict current vs. nonsmoking status. Lastly, we used DNAm guantitative trait loci (mQTL) mapping to tease apart genetically driven DNAm that underlie predisposing factors for initiating smoking and developing addiction (evidenced by QTL variants associated with smoking phenotypes) vs. gene regulatory responses to smoking exposure that may exert downstream neurobiological effects, reinforce smoking behavior, and contribute to failed smoking cessation. Both types of smoking-related DNAm changes are informative for broadening our understanding of neurobiological pathways leading to smoking and its consequences.

METHODS AND MATERIALS

Current smokers and nonsmokers among human postmortem NAc samples

Postmortem human NAc tissues were obtained at autopsy as part of the Lieber Institute for Brain Development (LIBD) Human Brain Repository. See Supplementary Methods for details on the data collection, cotinine measurement, and sample exclusion criteria.

Smoking status was determined based on both cotinine levels in blood and/or brain samples and next-of-kin reporting to minimize phenotype misclassification. Cases (current smokers) were defined by one of the following: cotinine level >12 ng/mL (threshold indicative of current active smoking rather than passive smoking [17]); or cotinine level detected above the reporting limit, but <12 ng/mL, in blood and/or brain tissues and a next-of-kin report that corroborates current smoking. Controls (nonsmokers) were defined by having both negative cotinine levels (below the reporting limit) and a next-of-kin report of no current smoking.

DNAm data, quality control (QC), and pre-processing

Illumina HumanMethylation EPIC data were generated using DNA extracted from NAc samples of 239 eligible decedents, as described previously [18, 19]. Individual-level data are available in the Gene Expression Omnibus (accession GSE147040) [20]. Data quality assessment and pre-processing were conducted using the R package, minfi [21]. DNAm data processing included stratified quantile normalization, and computing principal components (PCs) of the negative control probe intensities to correct for technical artifacts. Because cell type proportions can confound DNAm associations with outcomes of interest, we included neuronal cell type proportions to control for cellular heterogeneity [22], estimated using the Houseman method [23], as covariates in our statistical models. QC resulted in the exclusion of 18 samples and 76,413 probes. See Supplementary Methods for more details on the final analysis dataset (221 samples and 789.678 probes). Sample characteristics are shown in Table 1.

Smoking EWAS meta-analysis in NAc

EWAS of β -values, representing the percentage of DNAm at each CpG (ratio of methylated intensities relative to the total intensity), on smoking case/control status as the outcome was conducted separately by ancestry (European American [EA] and African American [AA] samples). The ancestry-specific results were combined by fixed-effects inverse variance-weighted metaanalysis using *METAL* [24]. Details regarding the final model selection are provided in Supplementary Methods. Robust linear regression analyses were conducted using the R package, *MASS* [25], to test the association between current vs. nonsmoking and DNAm (β -value), while adjusting for age at death, sex, negative control PCs, genotype PCs, and estimated proportion of Genome-wide DNA methylation differences in nucleus accumbens of smokers... CA Markunas et al.

555

Table 1. Description of LIBD NAc samples $(N = 221)^a$.						
Variable	Case (N = 53)	Control (N = 168)	P value ^b			
Sex, N						
Female	14	42	0.86			
Male	39	126				
Race, N						
African American	26	75	0.57			
European American	27	93				
Age at death (years),						
Mean + SD	47.25 ± 12.5	45.75 ± 13.9	0.46			
Range (minimum–maximum)	18.8–73.9	17.4–71.8				
Proportion of non-neuronal cells ^c , mean $+$ SD	0.76 ± 0.07	0.76 ± 0.07	0.56			
Negative control PCs ^d , mean +	- SD					
PC1	-0.08 ± 8.31	-0.10 ± 11.0	0.99			
PC2	0.06 ± 4.04	-0.002 ± 3.67	0.92			
PC3	-0.06 ± 0.91	0.01 ± 2.02	0.73			
PC4	-0.15 ± 0.86	0.04 ± 1.90	0.30			

categorical and continuous variables, respectively.

PC principal component, SD standard deviation, NA not applicable.

^aFinal dataset after quality control and study-specific exclusions were applied to the starting N = 239.

 ^{b}P values are based on a Fisher's exact test and *t*-test for categorical and continuous variables, respectively.

^cEstimated using NAc DNAm data following the Houseman method (refer to Methods for details).

^dGenerated using the negative control probes on the DNAm array (refer to Methods for details).

non-neuronal cells. We accounted for multiple testing by controlling the false discovery rate (FDR) at 5% [26]. An overview of our study design to extend and characterize our EWAS findings is provided in Fig. 1.

Functional annotation of differential DNAm sites using RNAexp, mQTL, and colocalization analyses

We generated corresponding RNA sequencing (RNA-seq)-based expression profiles for each NAc sample that had DNAm data (Table S1) and integrated these RNAexp data using a two-step approach. First, we directly correlated the levels of DNAm and nearby (<u>+</u>500 Kb) expressed genes across samples, regardless of smoking status, to identify those CpGs associated with RNAexp (and thus more likely to be functional). Next, we tested genes with a significant RNAexp-DNAm association for differential RNAexp by smoking in NAc, with and without adjustment for nearby DNAm levels. Details regarding RNA-seq data generation, processing, QC, RNAexp-DNAm association analyses, and differential RNAexp by smoking analyses are provided in Supplementary Methods.

We performed *cis*-mQTL analyses using *Matrix eQTL* [27] testing the association of DNAm and genotypes of variants \pm 500 Kb, without accounting for smoking, and then expanding the model using a joint 2 degrees-of-freedom (d.f.) test [28] accounting for both variant main effects and variant-by-smoking status interaction on DNAm. To test if any mQTL variants were also associated with predisposing genetic risk of smoking-related traits (i.e., mQTL and GWAS signals were shared), we performed a colocalization analysis, applying the *coloc* [29] function (coloc.abf) to evaluate whether the mQTL association signals were shared with the GWAS signals for smoking initiation (i.e., ever vs. never smoking, most

Genome-wide DNA methylation differences in nucleus accumbens of smokers... CA Markunas et al.

556



Fig. 1 Overview of the study design. Dotted lines indicate previously published results. SmPEGS: Smoking methylation PolyEpigenetic Score. ^aSugden et al. [16]. ^bJoehanes et al. [14]. ^cGenes proximal to differentially methylated CpGs.

closely resembling the current smoker vs. nonsmoker phenotype in LIBD) and cigarettes per day (most indicative of nicotine dependence) from GSCAN [6]. See details in the Supplementary Methods.

Comparison of smoking-related DNAm changes in NAc and blood To investigate tissue-specific vs. -shared smoking-related DNAm changes, we compared our EWAS meta-analysis results in NAc (FDR < 0.05) to published results from the largest bloodbased EWAS meta-analysis of current vs. never smoking to date (N = 15,907) [14]. We assessed the overlap between the NAc- and blood-based smoking DNAm results at both the CpG- and genelevel. Details regarding these comparisons are provided in Supplementary Methods.

Application of a blood-based Smoking methylation PolyEpigenetic Score (SmPEGS) in brain

We applied the blood-based SmPEGS [16] to our brain DNAm data (Supplementary Methods). Logistic regression models were run separately by ancestry and in the pooled sample to test the association between the blood-derived SmPEGS and current vs. nonsmoking in NAc, adjusting for the same covariates from our brain EWAS analysis. To assess the ability of the SmPEGS to classify smoking status in brain, we performed a receiver operating characteristic (ROC) analysis using the R package, *pROC* [30].

RESULTS

Epigenome-wide association results by smoking in NAc

We identified seven CpGs that exceeded FDR < 0.05 in our crossancestry EWAS meta-analysis (Fig. 2 and Table 2; Figs. S1 and S2). Each CpG is located within or proximal to a unique gene locus (Fig. S3). The difference in their mean percentages of DNAm in smokers and nonsmokers ranged from 0.92% to 4.13%. Three of the seven smoking-associated CpGs exhibited decreased DNAm by smoking, with consistent directions of effect between EA and AA ancestries (Table 2). Overall, smoking-related effect sizes were small, and there was no evidence of systematically increased or decreased DNAm by smoking (Fig. S1). More complete EWAS results are provided in Table S2 (meta-analysis P < 0.05). Full results are provided in Table S3.

Linking DNAm associations to RNAexp and mQTL variants in NAc Using a two-step approach, we integrated DNAm data with corresponding RNAexp data in the same NAc samples to identify smoking-related CpGs associated with nearby RNAexp and genes with differential RNAexp by smoking. We first focused on the seven smoking-related CpGs in NAc, and tested for association between DNAm levels and RNAexp of each gene within +500 Kb (Table S4). We identified one long noncoding RNA (IncRNA) with RNAexp levels that were significantly associated with DNAm levels at cg08395748: *RP11-297D21.4* (ENSG00000270049: antisense to *AGRP* [Agouti related neuropeptide]; $P = 1.54 \times 10^{-4}$, gene start site located 0.09 Mb upstream of the CpG). We next evaluated *RP11-297D21.4* for differential RNAexp by smoking and found no evidence of association, regardless of controlling for nearby DNAm levels at cg08395748 (without adjustment for DNAm: log₂ fold change = -0.06, P = 0.16; with adjustment for DNAm: log₂ fold change = -0.03, P = 0.60). Overall, these results provide limited functional support for the smoking-related DNAm changes in NAc.

To tease apart potential genetic effects on DNAm (i.e., variants with colocalized mQTL and smoking phenotype associations indicating predisposing risk factors for smoking) compared to DNAm changes resulting from smoking exposure (i.e., lacking mQTL variants, or having mQTL variants present but them not being associated with smoking phenotypes), we performed cismQTL analyses for each of the seven smoking-related CpGs. We observed statistically significant mQTL variants ($P < 2.2 \times 10^{-6}$ based on correction for testing 22,625 variants) for one CpG: a linkage disequilibrium block with the lead SNP rs10740507 (G allele) being associated with increased DNAm at cg09959332 $(P = 2.83 \times 10^{-19})$; Table S5); rs10740507 is also a significant *cis*eQTL in NAc and other tissues from the Genotype-Tissue Expression (GTEx) project [31]. However, coloc results for testing if both traits are associated and share a causal variant [29], found no overlap between GSCAN (N = 1,232,091 for smoking initiation GWAS, N = 337,334 for cigarettes per day GWAS) and mQTL signals from each smoking-related CpG (posterior probability <0.01). The cis-mQTL results (e.g., Table S5 for cg09959332) and lack of colocalized mQTL and smoking GWAS signals were consistent whether or not smoking was taken into account in the QTL model.

Comparisons to infer tissue-specific and -shared smoking DNAm effects

CpG-level: comparison of differentially methylated CpGs. We assessed our NAc-identified CpGs in the largest reported blood-based EWAS of smoking, conducted on the Illumina 450k array [14]; none of the five available smoking-related CpGs were associated in blood (Table 3). The two other smoking-related CpGs in NAc were captured on the EPIC array, but not the 450 K array, and thus not tested in blood. Effect sizes for all five available smoking-related CpGs in NAc were smaller in blood than in brain, and the directions of effect were consistent for three of the five CpGs. The mean percentages of DNAm in blood and brain were similar.

We performed the reverse look-up of blood-based findings in NAc. None of the 16,706 available smoking-related CpGs in blood were associated in NAc (Table S6), and the effect sizes were not correlated (r = 0.08). Overall, we found no evidence of tissue-shared smoking-related DNAm changes at the CpG-level.

Genome-wide DNA methylation differences in nucleus accumbens of smokers... CA Markunas et al.

557



Fig. 2 EWAS meta-analysis in NAc of smokers (N = 53) vs. nonsmokers (N = 168). CpGs are shown according to their position on chromosomes 1–22 (alternating red/blue) and plotted against their $-\log_{10} p$ values. The dotted horizontal line indicates genome-wide significance based on FDR < 0.05. The genomic inflation factor (λ) was 1.05 for the meta-analysis and 1.01 and 0.98 for European American-and African American-specific analyses, respectively.

CpG N	Nearest gene	EWAS meta-analysis ($N =$ 221)		EA EWAS (<i>N</i> = 120)		AA EWAS (<i>N</i> = 101)	
		Beta	P value ^a	Beta	P value ^a	Beta	P value ^a
cg18384794	ABLIM3	-0.014	6.25 × 10 ⁻⁹	-0.012	1.19 × 10 ⁻³	-0.016	9.58 × 10 ⁻⁷
cg17884843	ZIC1	0.016	6.05×10^{-8}	0.011	1.67×10^{-2}	0.019	4.04×10^{-7}
cg05224975	APCDD1L	0.009	8.63×10^{-8}	0.011	3.11×10^{-6}	0.007	5.13×10^{-3}
cg09959332	ZCCHC24	-0.041	1.95×10^{-7}	-0.034	6.76×10^{-4}	-0.053	4.17 × 10 ⁻⁵
cg11912754	PRKDC	0.016	2.81×10^{-7}	0.007	0.12	0.024	1.17 × 10 ⁻⁶
cg13555116	MTMR6	-0.022	3.90×10^{-7}	-0.026	2.40×10^{-6}	-0.016	2.74×10^{-3}
cq08395748	CTCF	0.010	4.01×10^{-7}	0.013	7.45×10^{-7}	0.006	0.05

Gene-level: comparison of genes proximal to differentially methylated CpGs. We annotated the seven smoking-related CpGs in NAc to the nearest gene. Of these, three genes (ZIC1 [Zic family member 1], ZCCHC24 [Zinc finger CCHC-type containing 24], PRKDC [Protein kinase, DNA-activated, catalytic subunit]) were also annotated as the nearest gene to a CpG showing differential DNAm by smoking in blood (Tables 3 and S7). Among the three overlapping genes (Fig. S3), ZIC1 was the only gene for which the smoking-associated CpG in NAc fell within the gene body. While the specific CpGs were not shared between tissues, their directions of association were consistent between tissues for ZIC1 (higher DNAm levels among smokers) and ZCCHC24 (lower DNAm

levels among smokers), but not for *PRKDC*. Overall, our results support both tissue-specific (*ABLIM3* [Actin binding LIM protein family member 3], *APCDD1L* [APC down-regulated 1 like], *MTMR6* [Myotubularin related protein 6], and *CTCF* [CCCTC-binding factor]) and -shared (*ZIC1*, *ZCCHC24*, *PRKDC*) genes proximal to differentially methylated CpG(s) by smoking.

Blood-based SmPEGS tests for discrimination of smoking status in brain

We applied a blood-based SmPEGS [16] to our brain DNAm data to predict smoking status (Table S8). The SmPEGS was associated with smoking in the combined (OR = 1.78 [95% confidence

Genome-wide DNA methylation differences in nucleus accumbens of smokers... CA Markunas et al.

558

СрG	Chr Position		Nearest gene	NAc EWAS (<i>N</i> = 221)		Blood EWAS ^a (<i>N</i> = 15,907)		
				Beta	P value	Beta	P value	Gene-level ^b
cg18384794	5	148509515	ABLIM3	-0.014	6.25×10^{-9}	0.0009	0.36	No
cg17884843	3	147141132	ZIC1	0.016	6.05×10^{-8}	-0.0006	0.59	Yes
cg05224975 ^c	20	57049253	APCDD1L	0.009	8.63×10^{-8}	NA	NA	No ^d
cg09959332 ^c	10	81224116	ZCCHC24	-0.041	1.95×10^{-7}	NA	NA	Yes
cg11912754	8	48676898	PRKDC	0.016	2.81×10^{-7}	0.0001	0.91	Yes
cg13555116	13	25862332	MTMR6	-0.022	3.90×10^{-7}	-0.0014	0.23	No
cg08395748	16	67605701	CTCF	0.010	4.01 × 10 ^{−7}	0.0004	0.52	No

NA not applicable.

^aBlood-based smoking cross-ancestry EWAS meta-analysis results [14]. Blood-based targeted look-up via personal communication with Dr. Roby Joehanes, first author of smoking EWAS.

^bNearest gene (e.g., *ABLIM3*) associated with NAc-based smoking-related CpG (FDR < 0.05) is also associated with a smoking-related CpG (FDR < 0.05) in blood. ^cEPIC probe not available on the Illumina 450 K DNAm array.

^dAPCDD1L-AS1 was identified in the blood-based EWAS of smoking.

interval: 1.22, 2.59], P = 0.003) and EA (OR = 2.00 [1.12, 3.59], P = 0.02) samples. The association among AAs was non-significant but consistent in direction (OR = 1.45 [0.79, 2.67], P = 0.24). The blood-based SmPEGS provided limited discrimination between smokers and nonsmokers in NAc with an area under the ROC curve (AUC) = 0.62 (95% confidence interval: 0.54, 0.71) in the combined sample and an AUC = 0.67 (0.55, 0.78) among EAs.

DISCUSSION

Our results strongly support tissue-specific smoking DNAm effects at the CpG-level in blood vs. brain, with some support for tissueshared DNAm effects at the gene-level. Our study is the first to report epigenome-wide DNAm changes by smoking in human postmortem brains. Among the seven CpGs exceeding FDR < 0.05 in NAc, only cg08395748 was associated with nearby RNAexp (RP11-297D21), but further evaluation indicated that RP11-297D21 was not differentially expressed by smoking. To investigate both tissue-specific and -shared smoking DNAm effects, we extended our findings to blood. None of the available NAc- or bloodsmoking-associated CpGs were associated with smoking in the other tissue. However, three of the seven NAc-identified CpGs were closest to genes that were also identified as proximal to smoking-associated CpGs in the prior blood-based EWAS. Further, we used cis-mQTL and colocalization analyses with well-powered GWAS of smoking phenotypes to separate predisposing vs. consequential dysregulation of smoking. We concluded that the observed DNAm changes at all seven smoking-related CpGs were likely consequences due to smoking exposure. These changes may in turn influence downstream neurobiological pathways in the NAc that reinforce smoking behaviors, contribute to failed smoking cessation, and influence other health-related effects.

Three genes were proximal to smoking-related DNAm changes in both blood and NAc: *ZIC1*, *ZCCHC24*, and *PRKDC*. While different CpGs were implicated in NAc vs. blood, the directions of effect for DNAm levels by smoking were consistent between tissues for *ZIC1* and *ZCCHC24*. All genes have been previously related to smoking in other contexts. Zic1 is a zinc finger transcription factor that plays a role in early development and formation of the cerebellum [32]. All five smoking-associated *ZIC1* CpGs in blood and NAc (FDR < 0.05) showed increased DNAm in current smokers. However, the opposite direction of effect has been reported for sperm samples (N = 156): decreased DNAm at *ZIC1* in current smokers ($P = 9.6 \times 10^{-8}$) [33]. *ZIC1* was also implicated in the smoking initiation GWAS from GSCAN (gene-based $P = 4.14 \times 10^{-7}$) [6]. In our study, *ZIC1* genetic variation that predisposes to smoking initiation was not colocalized with *cis*-mQTL-driven dysregulation, suggesting that the *ZIC1* locus is also vulnerable to DNAm changes due to smoking exposure. Dysregulation of *ZIC1* in brain has been observed for cocaine, with *ZIC1* downregulated in response to repeated administration in mouse NAc ($P = 3.7 \times 10^{-7}$) [34]. These multiple lines of evidence link the *ZIC1* locus to both predisposing and consequential effects of smoking, and possibly drug use more broadly.

While little is known about the function of *ZCCHC24*, increased DNAm at *ZCCHC24* in newborn blood samples has been associated with sustained maternal smoking during pregnancy ($P = 5.29 \times 10^{-4}$) [15]. *PRKDC* encodes the DNA-dependent protein kinase catalytic subunit and plays a key role in the DNA repair pathway, non-homologous end-joining (NHEJ) [35]. Genetic variants in NHEJ genes [36], including *PRKDC* [37], have been previously associated with lung cancer, of which cigarette smoking is a major risk factor.

We also report 4 genes with evidence of proximal differential DNAm by smoking that is specific to NAc: (1) ABLIM3, (2) APCDD1L, (3) MTMR6, and (4) CTCF. ABLIM3 is involved in axonal guidance signaling and is downregulated in dopaminergic neurons in the ventral tegmental area in relation to perinatal nicotine exposure in rats [38]. Little is known about the function of APCDD1L, but it was selected as part of two blood-based gene expression signatures developed to classify smokers vs. nonsmokers [39]. While APCDD1L was not annotated to any smoking-associated CpGs in blood, the antisense non-coding RNA of APCDD1L (APCDD1L-AS1) was identified (cg14546153: $\beta = 0.003$, P = 0.0004) [14]. *MTMR6* is a member of the myotubularin-related protein family of phosphatidylinositol 3-phosphatases and plays a key role in 3phosphoinositide lipid metabolism [40]. While MTMR6 dysregulation in brain has not been linked to smoking or addiction phenotypes, differential MTMR6 expression in brain has been implicated in schizophrenia [41], which is highly comorbid and genetically correlated with smoking [42]. CTCF, a well-known transcriptional regulator, is thought to play a role in brain development, regulation of neural genes [43], and response to dopamine (a neurotransmitter involved in addiction and reward pathways) [44].

Only tissue-specific effects were observed at the CpG-level. NAcbased smoking-related CpGs were not identified in blood, and blood-based smoking-related CpGs were not identified in NAc. For example, cg05575921 within *AHRR* (Aryl-Hydrocarbon Receptor Repressor) is a robustly replicated blood-based DNAm biomarker for cigarette smoking, but was not associated with smoking in NAc (meta-analysis P = 0.57; 0.19% difference in DNAm between smokers and nonsmokers). For comparison, cg05575921 showed an 18% difference in DNAm between smokers and never smokers in blood samples ($P = 4.5 \times 10^{-26}$) [14]. While unlikely to fully account for the lack of association in brain, the likely presence of former smokers among our nonsmoker controls may reduce power to detect DNAm changes at cg05575921, as cg05575921 is associated with former vs. never smoking, although its effect is attenuated by 77.5% compared with current vs. never smokers [14].

To further extend blood-based findings, we applied a bloodderived smoking DNAm polyepigenetic score to our brain data to predict smoking status. We found that the SmPEGS provided limited discrimination between smokers and nonsmokers in NAc (AUC = 0.62 [0.54, 0.71]), indicating that many effects detected in blood may not directly translate to strong effects in brain. For comparison, AUCs of 0.81–0.93 for never vs. current smoking and 0.77–0.78 for never vs. ever (current/former) smoking were reported in blood-based samples [16].

While this study's data provide a new resource to begin investigating the neurobiological underpinnings of smoking phenotypes, there are several limitations. The data (i.e., multiomic data in human postmortem brain with available smoking data) are highly unique, but our sample size and thus statistical power were limited, possibly impacting our ability to detect smoking-related DNAm changes in brain and link those changes with nearby RNAexp. The case and control definitions for cigarette smoking were based on multiple data sources and corroboration of board-certified psychiatrists, suggesting that phenotype misclassification was unlikely; any misclassification would be expected to have identified cases as controls and biased the results towards the null. The present study was designed as a nested case-control sample selected from decedents with postmortem NAc tissue available in the LIBD Human Brain Repository with exclusion criteria for trauma, cancer, and other brain-damaging conditions. We set additional exclusion criteria to enable discovery of DNAm associations specifically with cigarette smoking in late adolescents and adults and with minimal confounding by co-morbid psychiatric diseases or alcohol/ substance use (i.e., no intoxication at time of death or next-ofkin reporting of lifetime alcohol or substance use disorder). With these criteria, the potential for ascertainment or other selection bias cannot be excluded, given inherent differences associated with smoking (for example, a greater portion of cases having died of natural causes [90.6%], likely due to smoking's widespread detrimental effects, compared with controls [79.2%]). However, applying the same exclusion criteria to cases and controls and achieving similar distributions in age, sex, and ancestry distributions (Table 1) likely helped to mitigate potential selection biases between the case/control groups.

There are currently no external NAc DNAm datasets for replication of our novel findings, so we relied on internal consistency between ancestral groups by focusing on metaanalysis results, rather than ancestry-specific findings. We also relied on a large-scale blood-based EWAS of smoking (N = 15,907) [14] to compare CpG-smoking associations in NAc vs. blood; ideally, these comparisons would be done in the same cohort, but to our knowledge, no cohorts exist with genome-wide DNAm data in NAc and blood tissues from decedents with known smoking histories.

Our comparisons may have been affected by differences between the NAc- and blood-based datasets. Statistical power to detect blood-based findings in brain was more limited, but there should be sufficient power to detect the NAc-based findings from the blood-based EWAS [14]. Although different Illumina DNAm arrays were used (450 K [blood] vs. EPIC [NAc]), the more recently developed EPIC array targets over 90% of the 450 K CpG probes and both arrays use the same probe chemistries and generally exhibit high levels of correlation [45]. For the gene-level comparison between tissues, 'target' 559

genes were defined based on physical proximity due to the lack of available RNAseq and DNAm data in blood to conduct cisexpression quantitative trait DNAm analyses, as we performed in brain. As a result, we may miss true target genes and limit our ability to define shared and unique smoking-related effects between blood and brain. In addition, the reference data used to derive estimates of cell-type proportions were based on Illumina 450 K data generated in dorsolateral prefrontal cortex brain samples [21]. While we used cell type-discriminating CpGs that overlapped with the EPIC array for cell-type proportion estimates in NAc, there could be measurement error in the cell type proportions arising from the different tissues. There could be residual cell type differences that were unaccounted for in our analysis that would result in false-positive associations. Because we analyzed a mixture of cell subtypes concurrently, it is possible that we may miss true smoking-related DNAm changes that occur in only a specific cell subtype. Further, measurement error in the DNAm data, if any, would be expected to be nondifferential, given the randomization of cases and controls across plates, and bias the results towards the null. Nonetheless, our statistical model adjusted for PCs of the DNAm negative control probes to control technical artifacts. Measurement error in the other potential confounders included in our model (age, sex, ancestry) were likely negligible. Finally, we are unable to rule out unmeasured or residual confounding due to other exposures or traits (e.g., former alcohol use [no decedent was intoxicated at time of death]) or reduced power due to smoking-related DNAm changes which persist in former smokers

Our study is the first to demonstrate smoking-related DNAm changes in human brain and lays the foundation for future studies to expand and extend our findings in brain. Our results highlight CpGs, previously undetected as peripheral DNAm biomarkers of smoking that may reflect brain-specific processes, and provide support for tissue-shared genes proximal to smoking-related DNAm changes, suggesting that the susceptibility of certain epigenomic regions to smoking may be conserved between tissues. By incorporating genetic data to differentiate smokingrelated DNAm changes that are genetically-driven (i.e., DNAm as a mediator of genetic risk of smoking) vs. exposure-driven (i.e., DNAm changes as a consequence of exposure), we found that the observed smoking-related DNAm changes at 7 CpGs most likely reflect neurobiological processes related to smoke exposure. Exposure-driven epigenetic associations, as we observed, may have smaller effect sizes in bulk tissue, compared with cell typespecific studies, and may not capture biologically meaningful effects that accumulate over time and with repeated exposures when measured cross-sectionally [46]. We expect that increasing sample size and expanding the number of brain tissues and cell types examined in future studies will identify additional DNAm and RNAexp changes, expanding knowledge of both underlying predisposition to smoking and consequences of smoking exposure on the epigenome, and enable the investigation of ancestryspecific effects.

FUNDING AND DISCLOSURE

This work was supported by the National Institute on Drug Abuse R01 DA042090 (PI: DBH). LBJ is listed as an inventor on U.S. Patent 8,080,371, "Markers for Addiction" covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. There are no other potential conflicts to disclose.

AUTHOR CONTRIBUTIONS

RT, ADS, TMH, JEK, AEJ collected and maintain the human brain tissue repository and its data; CAM, BCQ, LJB, EOJ, AEJ, DBH conceptualized the study design and statistical analyses for cigarette smoking; RT generated the data; CAM, BCQ, CUM analyzed the

560

data; CAM, SAS, BCQ, MUC, EOJ, AEJ, DBH interpreted results; CAM drafted the manuscript; all authors critically revised the manuscript and approve of the final version.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41386-020-0782-0).

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

REFERENCES

- 1. World Health Organization. Report on the Global Tobacco Epidemic. Geneva: World Health Organization; 2017.
- Wang TW, Asman K, Gentzke AS, Cullen KA, Holder-Hayes E, Reyes-Guzman C, et al. Tobacco product use among adults — United States, 2017. MMWR Morb Mortal Wkly Rep. 2018;67:1225–32.
- Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction. 2003;98:23–31.
- Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav Genet. 2005;35:397–406.
- Hancock DB, Markunas CA, Bierut LJ, Johnson EO. Human genetics of addiction: new insights and future directions. Curr Psychiatry Rep. 2018;20:8.
- Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51:237–44.
- Culverhouse RC, Saccone NL, Stitzel JA, Wang JC, Steinbach JH, Goate AM, et al. Uncovering hidden variance: pair-wise SNP analysis accounts for additional variance in nicotine dependence. Hum Genet. 2011;129:177–88.
- Semick SA, Collado-Torres L, Markunas CA, Shin JH, Deep-Soboslay A, Tao R, et al. Developmental effects of maternal smoking during pregnancy on the human frontal cortex transcriptome. Mol Psychiatry. (In press).
- Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. Annu Rev Neurosci. 2006;29:565–98.
- Nestler EJ. Is there a common molecular pathway for addiction? Nat Neurosci. 2005;8:1445–9.
- Koob GF, Volkow ND. Neurocircuitry of addiction. Neuropsychopharmacology. 2010;35:217–38.
- Grace AA, Floresco SB, Goto Y, Lodge DJ. Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. Trends Neurosci. 2007;30:220–7.
- 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:e63812.
- 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.
- Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. Am J Hum Genet. 2016;98:680–96.
- Sugden K, Hannon EJ, Arseneault L, Belsky DW, Broadbent JM, Corcoran DL, et al. Establishing a generalized polyepigenetic biomarker for tobacco smoking. Transl Psychiatry. 2019;9:92.
- Avila-Tang E, Al-Delaimy WK, Ashley DL, Benowitz N, Bernert JT, Kim S, et al. Assessing secondhand smoke using biological markers. Tob Control. 2013;22:164–71.
- Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, et al. DNA methylation signatures in development and aging of the human prefrontal cortex. Am J Hum Genet. 2012;90:260–72.
- Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nat Neurosci. 2016;19:40–7.
- Research dataset. Hancock D. Genome-wide DNA methylation differences in nucleus accumbens of smokers vs. nonsmokers. 2020. https://www.ncbi.nlm.nih. gov/geo/query/acc.cgi?acc=GSE147040.
- 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;30:1363–9.

- Jaffe AE, Shin J, Collado-Torres L, Leek JT, Tao R, Li C, et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nat Neurosci. 2015;18:154–61.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.
- 25. Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New york: Springer; 2002.
- Benjamini Y, Hochberg Y. Controlling for false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57:289–300.
- Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28:1353–8.
- Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Hum Hered. 2007;63:111–9.
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10:e1004383.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an opensource package for R and S+ to analyze and compare ROC curves. BMC Bioinforma. 2011;12:77.
- 31. GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45:580–5.
- 32. Merzdorf CS. Emerging roles for zic genes in early development. Dev Dyn. 2007;236:922–40.
- Jenkins TG, James ER, Alonso DF, Hoidal JR, Murphy PJ, Hotaling JM, et al. Cigarette smoking significantly alters sperm DNA methylation patterns. Andrology. 2017;5:1089–99.
- Feng J, Wilkinson M, Liu X, Purushothaman I, Ferguson D, Vialou V, et al. Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. Genome Biol. 2014;15:R65.
- 35. Chen BP, Chan DW, Kobayashi J, Burma S, Asaithamby A, Morotomi-Yano K, et al. Cell cycle dependence of DNA-dependent protein kinase phosphorylation in response to DNA double strand breaks. J Biol Chem. 2005;280:14709–15.
- Tseng RC, Hsieh FJ, Shih CM, Hsu HS, Chen CY, Wang YC. Lung cancer susceptibility and prognosis associated with polymorphisms in the nonhomologous end-joining pathway genes: a multiple genotype-phenotype study. Cancer. 2009;115:2939–48.
- Hsia TC, Chang WS, Chen WC, Liang SJ, Tu CY, Chen HJ, et al. Genotype of DNA double-strand break repair gene XRCC7 is associated with lung cancer risk in Taiwan males and smokers. Anticancer Res. 2014;34:7001–5.
- Keller RF, Dragomir A, Yantao F, Akay YM, Akay M. Investigating the genetic profile of dopaminergic neurons in the VTA in response to perinatal nicotine exposure using mRNA-miRNA analyses. Sci Rep. 2018;8:13769.
- Giordano M, Tripathi KP, Guarracino MR. Ensemble of rankers for efficient gene signature extraction in smoke exposure classification. BMC Bioinforma. 2018;19:48.
- Mochizuki Y, Ohashi R, Kawamura T, Iwanari H, Kodama T, Naito M, et al. Phosphatidylinositol 3-phosphatase myotubularin-related protein 6 (MTMR6) is regulated by small GTPase Rab1B in the early secretory and autophagic pathways. J Biol Chem. 2013;288:1009–21.
- Bergon A, Belzeaux R, Comte M, Pelletier F, Herve M, Gardiner EJ, et al. CX3CR1 is dysregulated in blood and brain from schizophrenia patients. Schizophr Res. 2015;168:434–43.
- Hartz SM, Horton AC, Hancock DB, Baker TB, Caporaso NE, Chen LS, et al. Genetic correlation between smoking behaviors and schizophrenia. Schizophr Res. 2018;194:86–90.
- 43. Arzate-Mejia RG, Recillas-Targa F, Corces VG. Developing in 3D: the role of CTCF in cell differentiation. Development. 2018;145:dev137729.
- Galloway A, Adeluyi A, O'Donovan B, Fisher ML, Rao CN, Critchfield P, et al. Dopamine triggers CTCF-dependent morphological and genomic remodeling of astrocytes. J Neurosci. 2018;38:4846–58.
- Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the illumina methylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:208.
- 46. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, et al. Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: the children's environmental health and disease prevention research center's epigenetics working group. Environ Health Perspect. 2017;125:511–26.