The American Journal of Human Genetics, Volume *111*

Supplemental information

The association of cigarette smoking with DNA methylation and gene expression in human tissue samples

James L. Li, Niyati Jain, Lizeth I. Tamayo, Lin Tong, Farzana Jasmine, Muhammad G. Kibriya, Kathryn Demanelis, Meritxell Oliva, Lin S. Chen, and Brandon L. Pierce

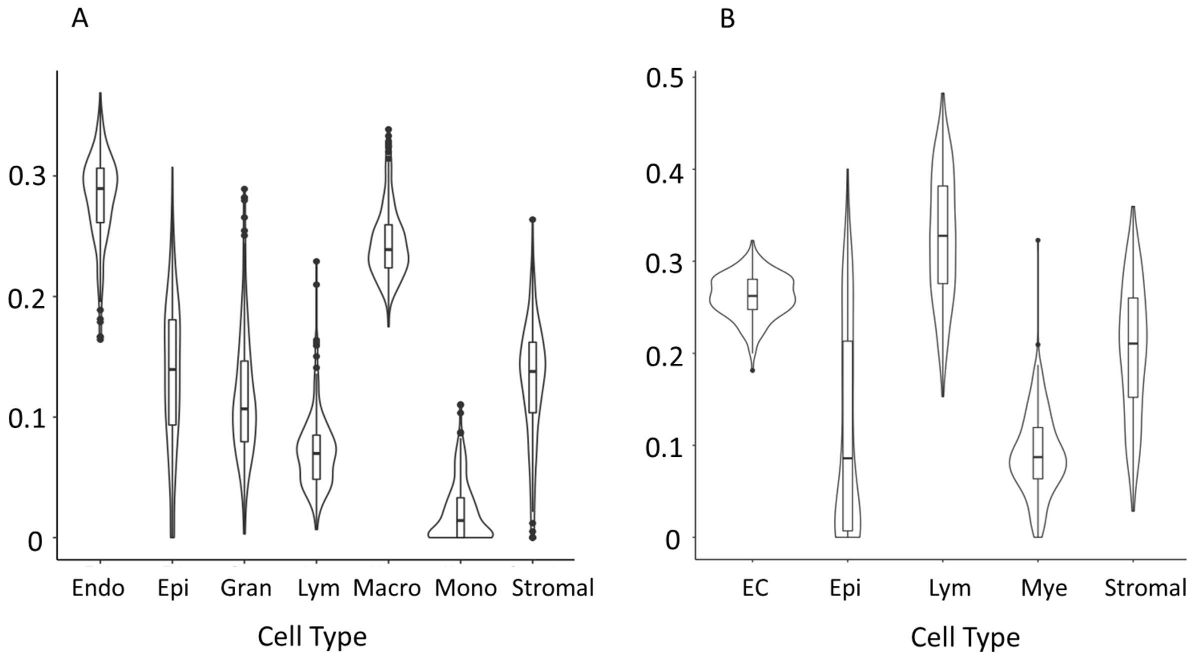


Figure S1. Distribution of cell type percentages

(A) Estimated cell type percentages for lung using the EPISCORE method with the pan-tissue DNAm atlas as a reference dataset (B) Estimated cell type percentages for colon using the EPISCORE method with the pan-tissue DNAm atlas as a reference dataset.

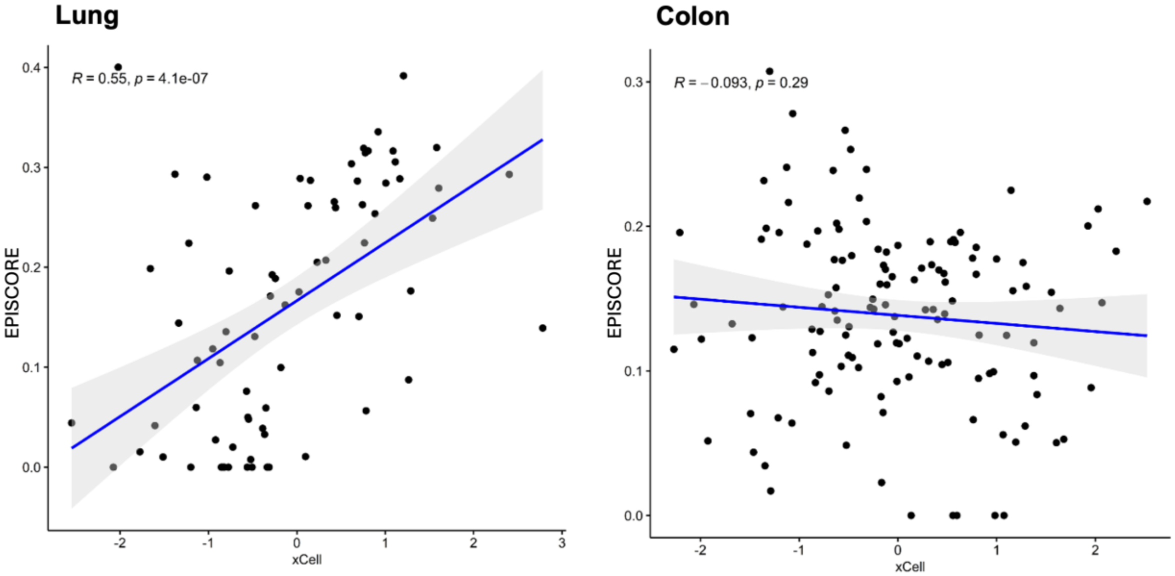
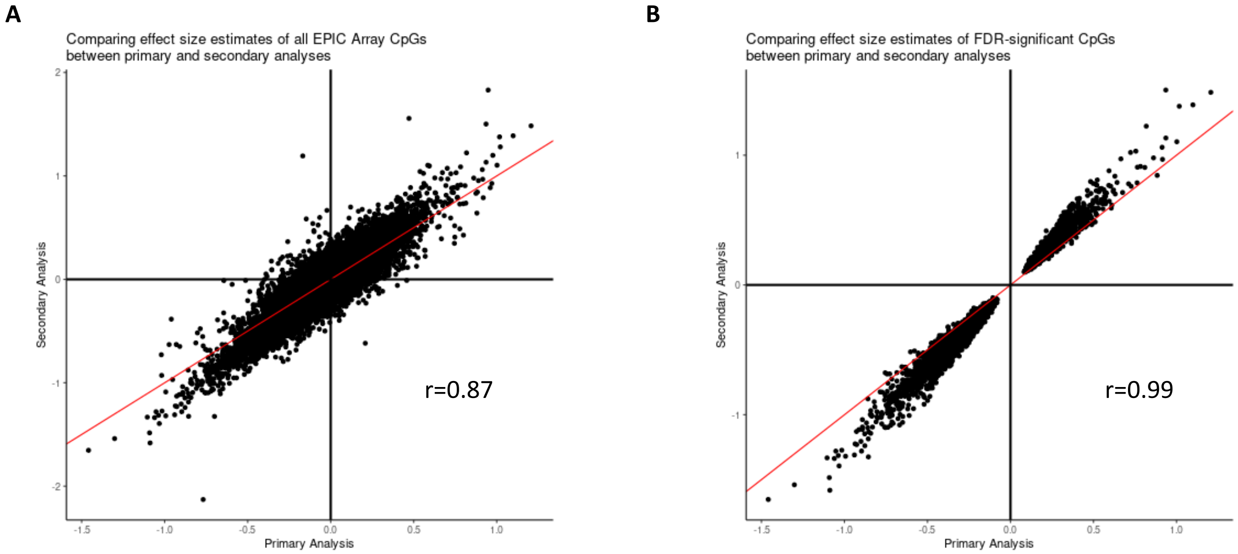


Figure S2. Correlation between EPISCORE and xCell cell type proportions
(A) Spearman correlation between EPISCORE and xCell computed epithelial cell proportions for lung. (B) Spearman correlation between EPISCORE and xCell computed epithelial cell proportions for colon.



A

Comparing effect size estimates of all EPIC Array CpGs between primary and secondary analyses

B

Comparing effect size estimates of FDR-significant CpGs between primary and secondary analyses

Figure S3. Comparison of effect size estimates between the primary EWAS and secondary EWAS analysis

Scatterplot comparing effect size estimates of each CpG between the EWAS of ever vs. never smokers (primary analysis) and current vs. never smokers (secondary analysis) in lung tissue for A) All CpGs on the EPIC Array and B) CpGs that were significantly associated with smoking at an FDR of 0.05.

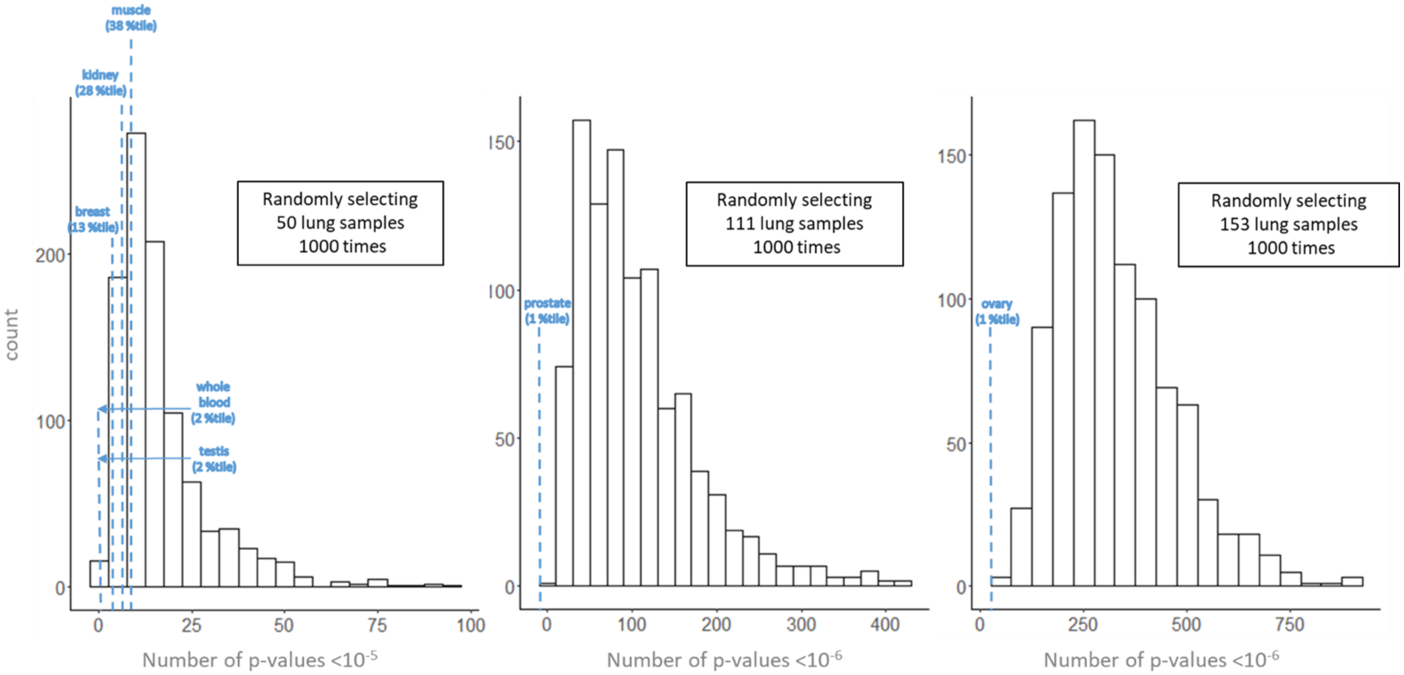


Figure S4. Lung shows more prominent effects of smoking than other tissue types. Number of smoking-associated CpGs in lung compared to the number in other tissues that pass p-value thresholds after down sampling to sample sizes of n=50, n=111, and n=153. Each distribution shown is generated by randomly selecting samples from the 212 lung samples (and conducting EWAS analyses) 1000 times.

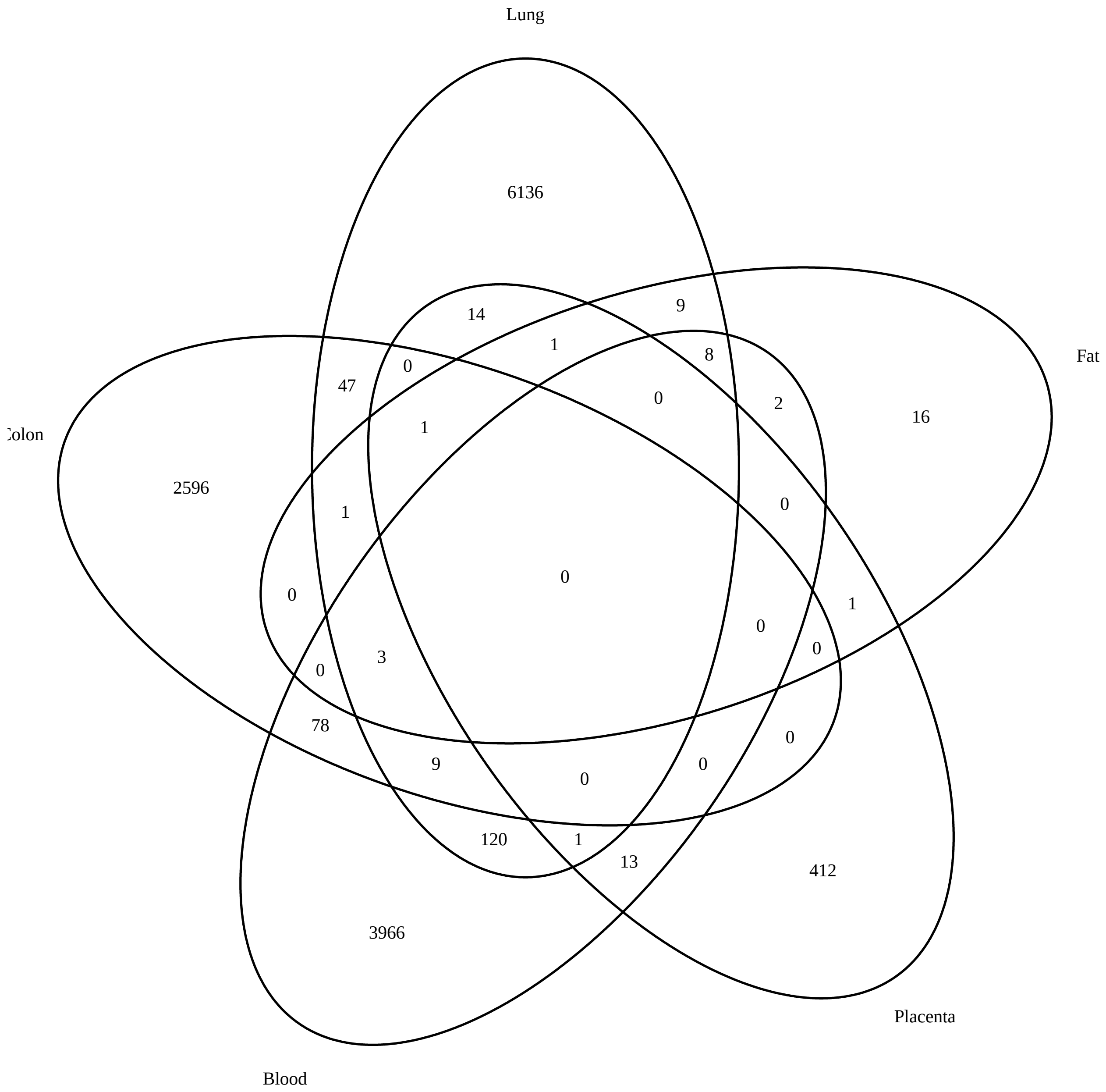


Figure S5. Venn diagram of the number of CpGs previously identified in EWAS conducted in different tissue types

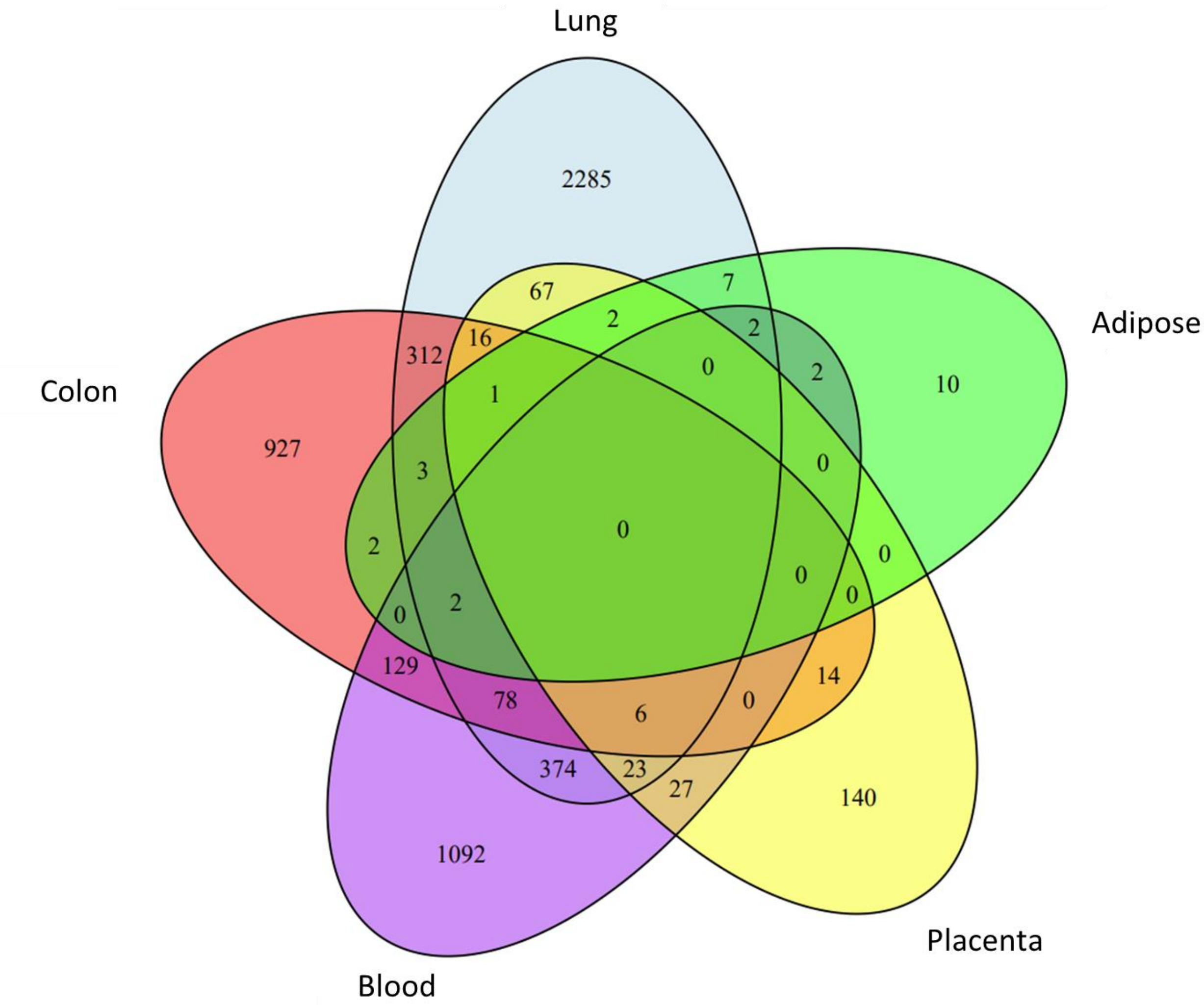


Figure S6. Venn diagram of the genes annotated to smoking-associated CpGs previously identified in EWAS conducted in different tissue types

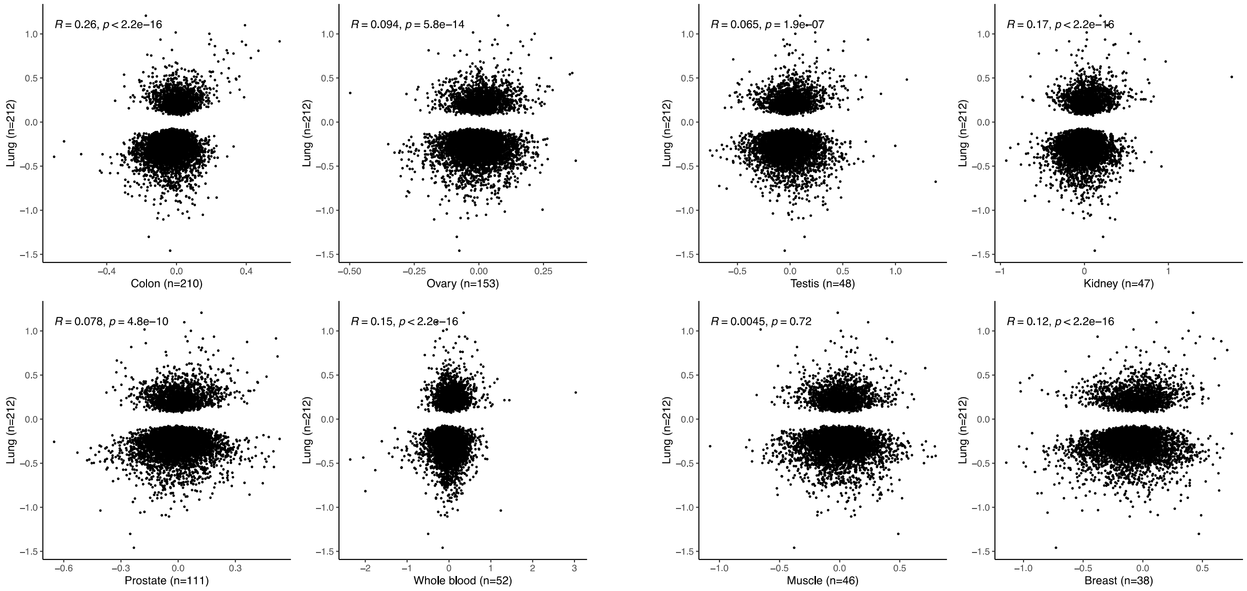


Figure S7. Comparison of effect size estimates between pairs of tissue types

Scatterplots showing the correlations between association estimates observed across pairs of tissue types. Smoking-associated CpGs that pass FDR 0.05 in lung are shown on the vertical axes, with all other tissues shown on the horizontal axes.

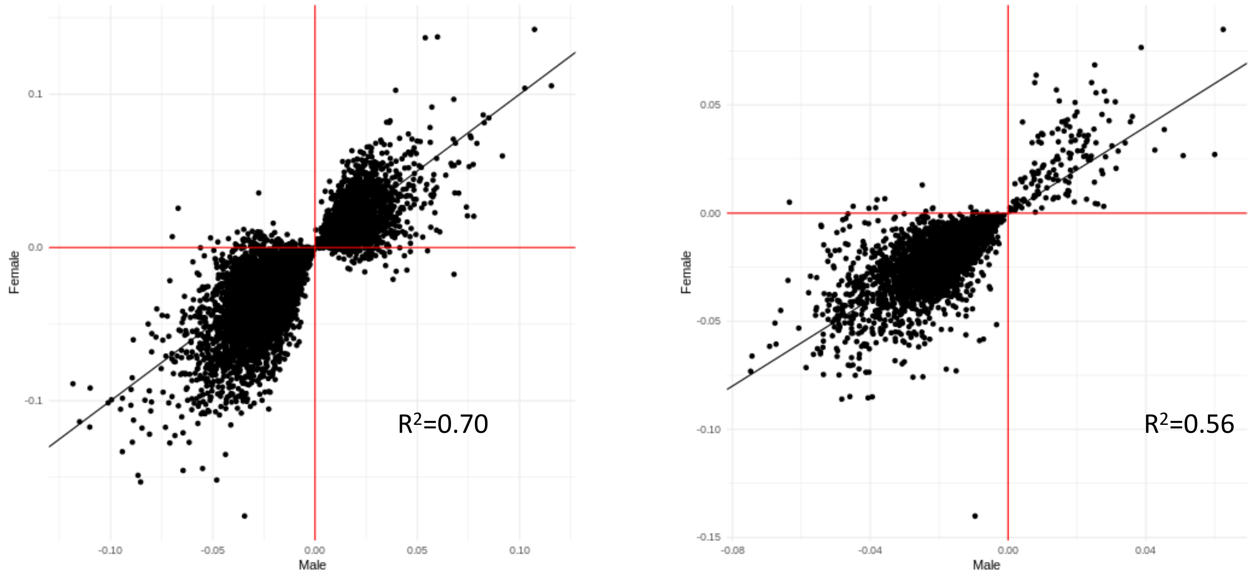


Figure S8. Comparison of effect size estimates between females and males

Scatterplot comparing effect size estimates of each CpG when stratifying by sex for all CpGs that were significantly associated with smoking at an FDR of 0.05 in lung (left) and in colon (right). Black line represents the identity line.

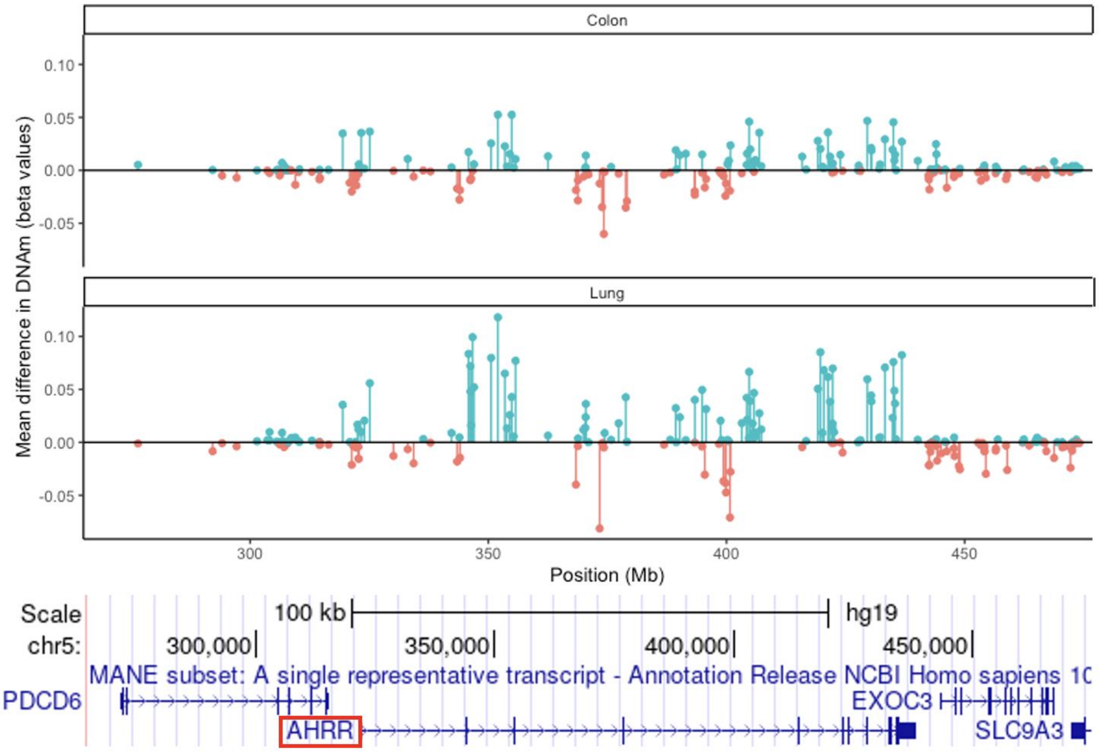


Figure S9. Difference in DNAm beta values between smokers and non-smokers across CpGs measured around the AHRR gene.

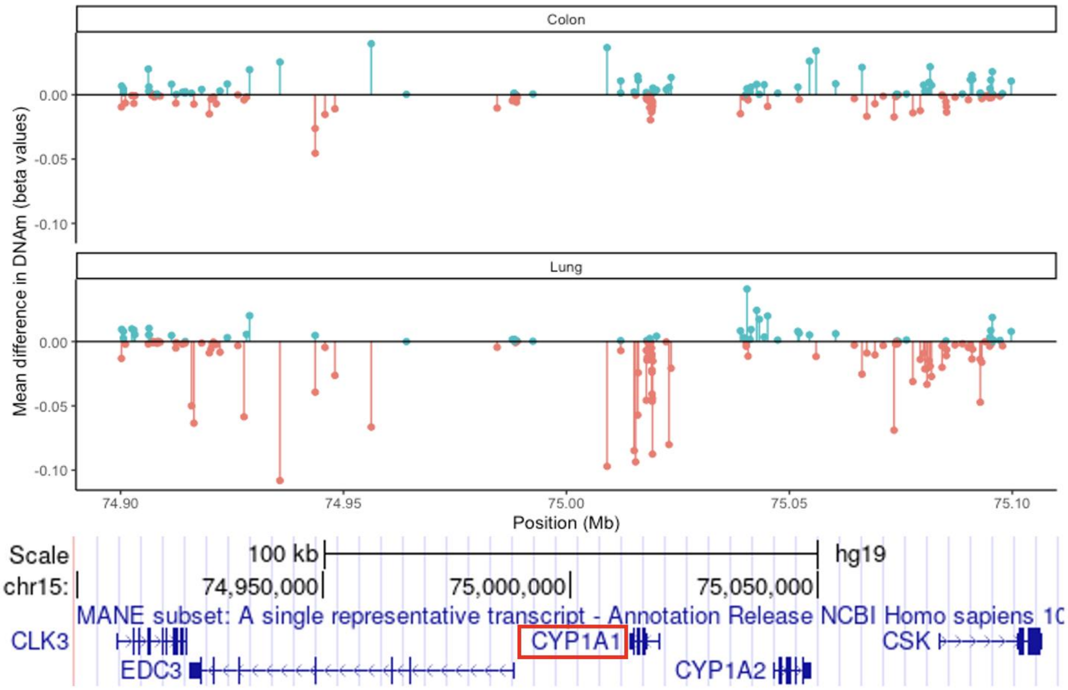


Figure S10. Difference in DNAm beta values between smokers and non-smokers across CpGs measured around the CYP1A1 gene.

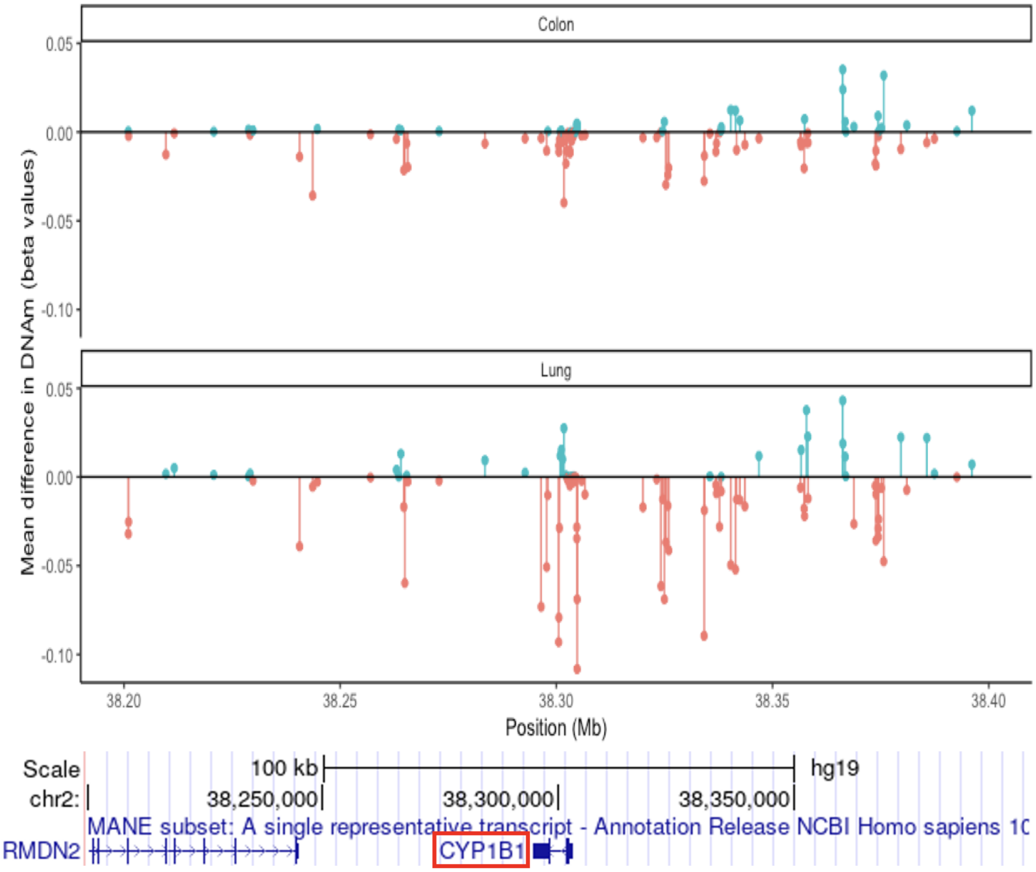


Figure S11. Difference in DNAm beta values between smokers and non-smokers across CpGs measured around the CYP1B1 gene.

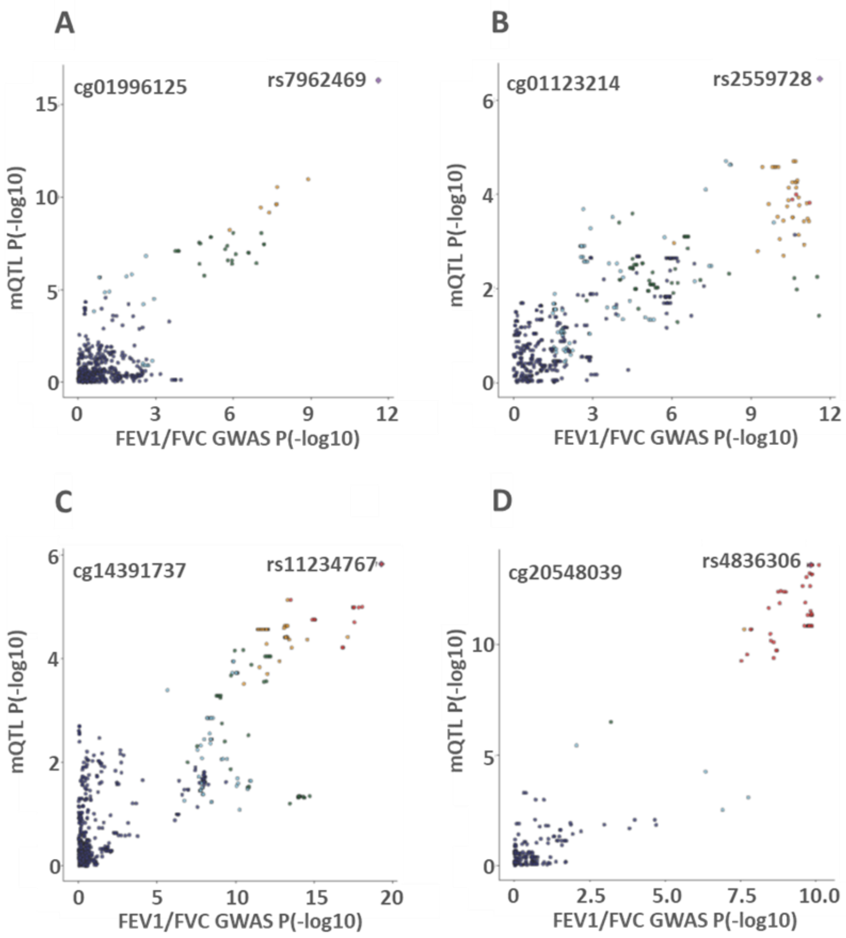


Figure S12. Comparison of P-values for mQTLs vs FEV1/FVC GWAS signals
Scatter plot of P-values for mQTL signals vs. FEV1/FVC GWAS signals for four loci with mQTL/GWAS co-localization including (A) ACVR1B, (B) SFTPA1, (C) PRSS23, and (D) MARCHF3/MARCH3

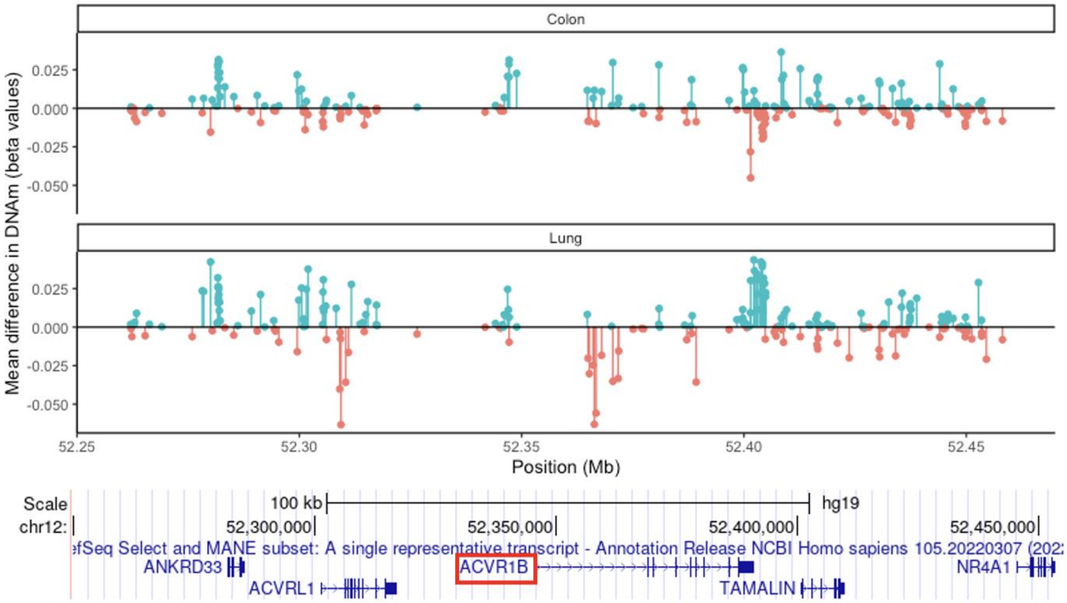


Figure S13. Difference in DNAm beta values between smokers and non-smokers across CpGs measured around the ACVR1B gene.

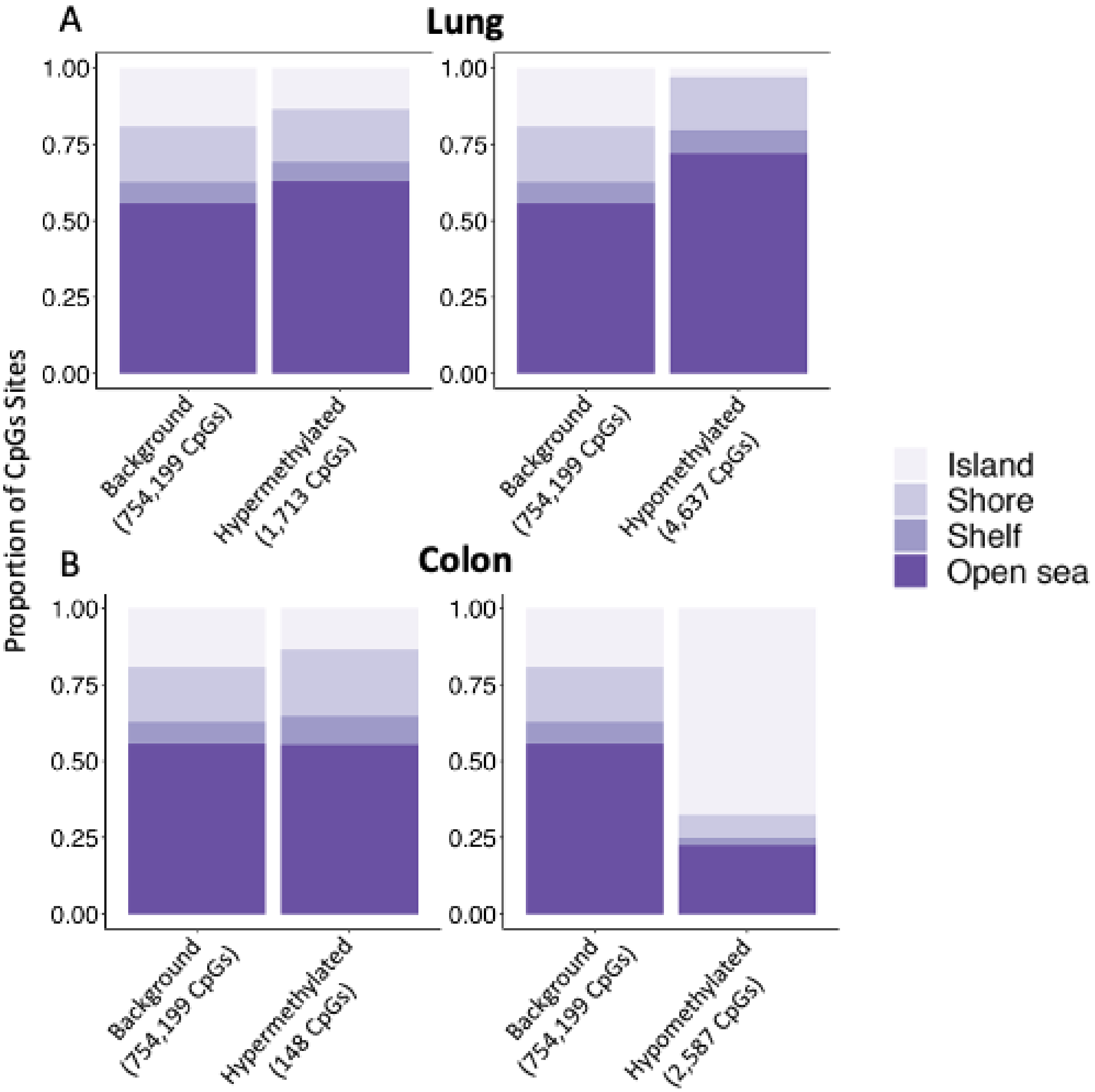


Figure S14. Enrichment of smoking-related CpGs in relation to CpG Island status

Locational distribution of significant smoking-related CpGs sites (FDR < 0.05) in relation to CpG islands in (A) lung tissue (B) colon tissue. Colors represent location of CpG site. Background: All CpGs assayed in the Infinium MethylationEPIC array included in our analyses.

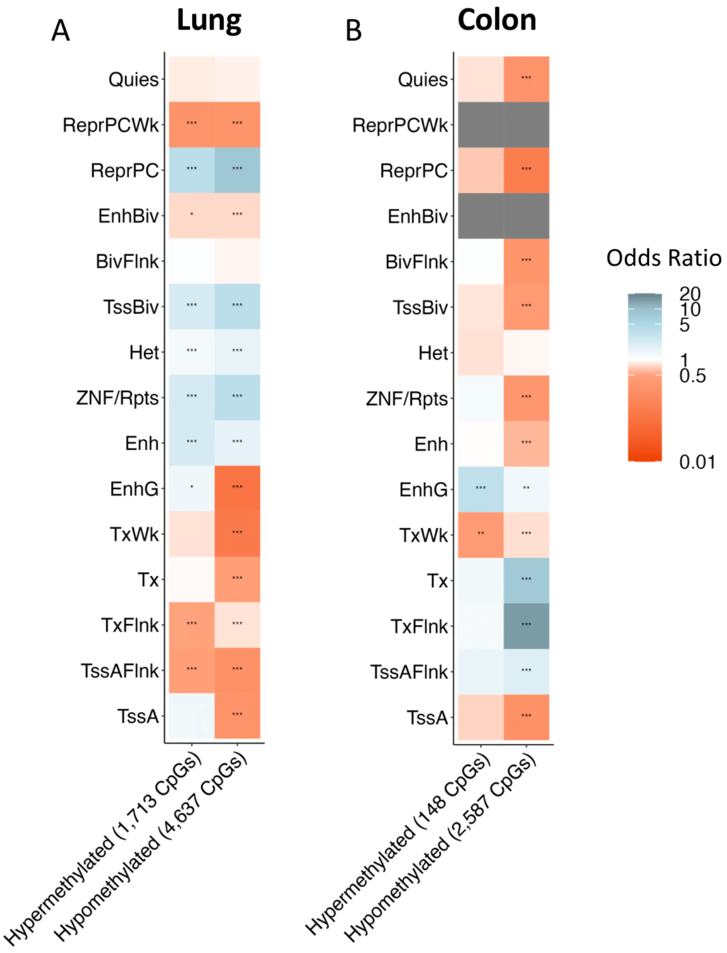
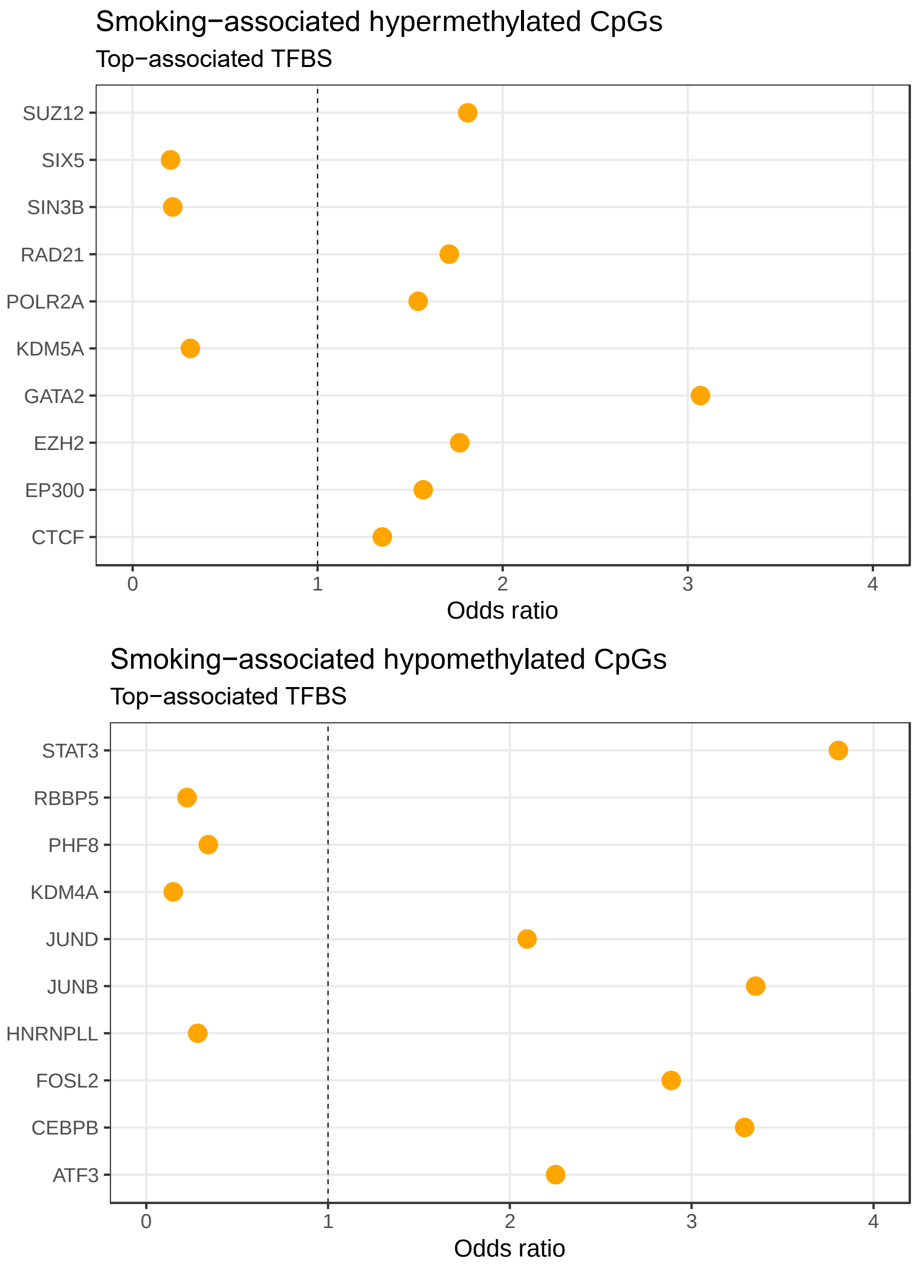


Figure S15. Enrichment of smoking-related CpG sites among chromatin segmentation features

Enrichment of smoking-related CpG sites (FDR <0.05) expressed as odd ratios in (A) lung tissue (B) colon tissue. Fisher's exact P value * < 0.05, **<0.01, ***<0.001. Active chromatin states: active transcription start site (*TssA*), flanking active TSS (*TssAFlnk*), transcription at gene 5′ and 3′ showing both promoter and enhancer (*TxFlnk*), strong transcription (Tx), weak transcription (*TxWk*), genic enhancers (*EnhG*), enhancers (*Enh*) , zinc finger protein genes and repeats (*ZNF/Rpts* ZNF*).* Inactive chromatin states: heterochromatin (*Het*), bivalent/poised TSS (*TssBiv*), flanking bivalent TSS/Enh (*BivFlnk*), bivalent enhancer (*EnhBiv*), repressed polycomb (*ReprPC*), weak repressed polycomb (*ReprPCWk*), quiescent/low (*Quies*).



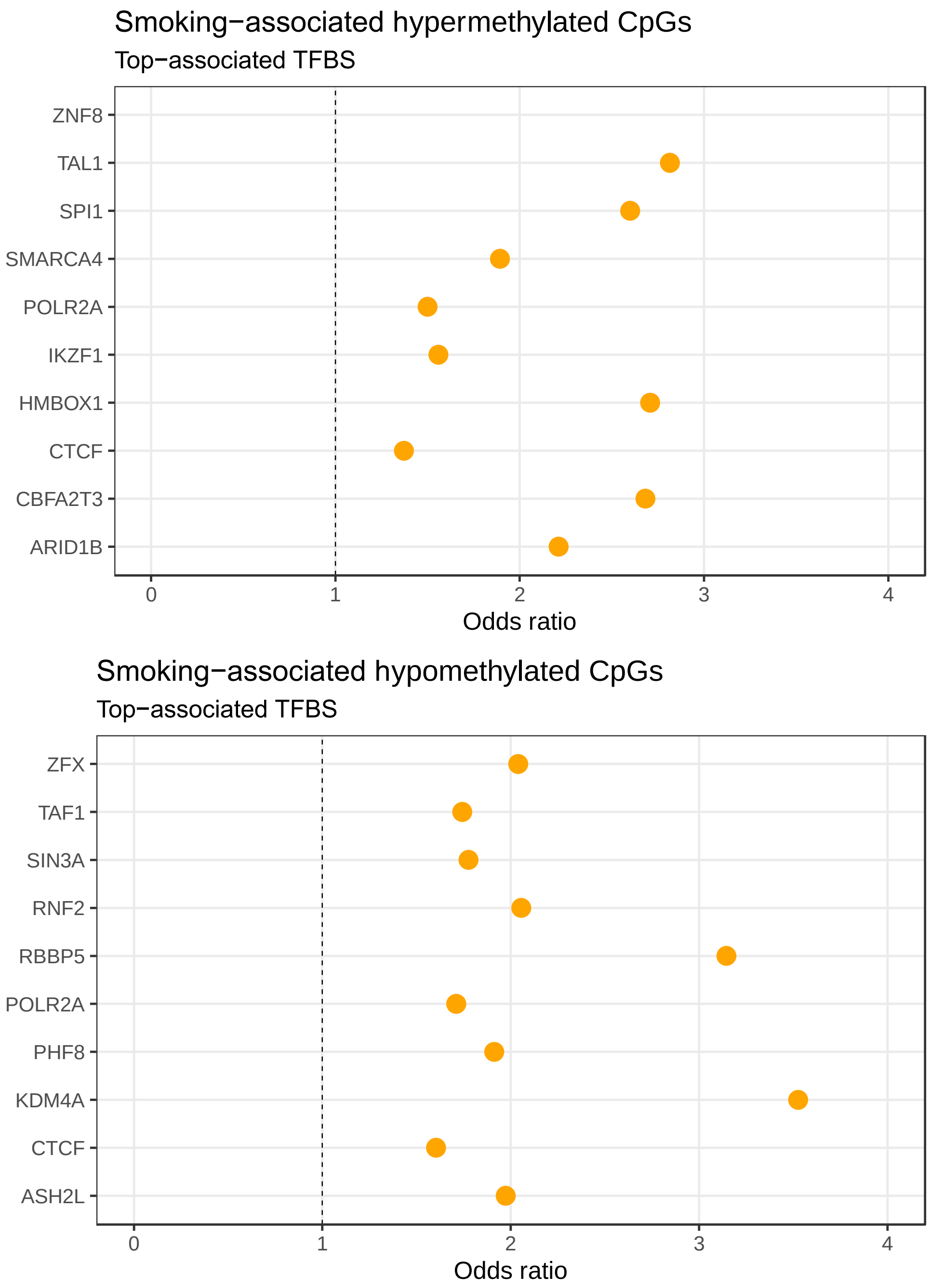
Smoking−associated hypermethylated CpGs

Top−associated TFBS

Smoking−associated hypomethylated CpGs

Top−associated TFBS

Figure S16. Enrichment of lung smoking-associated CpGs in transcription factor binding sites.


Smoking−associated hypermethylated CpGs

Top−associated TFBS

Smoking−associated hypomethylated CpGs

Top−associated TFBS

Figure S17. Enrichment of colon smoking-associated CpGs in transcription factor binding sites.

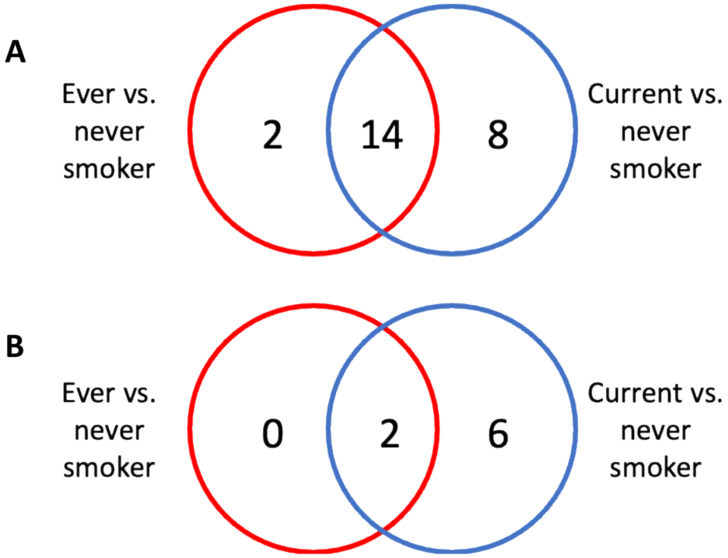


Figure S18. Comparison of the number of Hallmark gene sets and KEGG pathways

Venn diagrams comparing the number of A) Hallmark gene sets and B) KEGG pathways identified in the primary vs. secondary EWAS analysis in lung tissue.

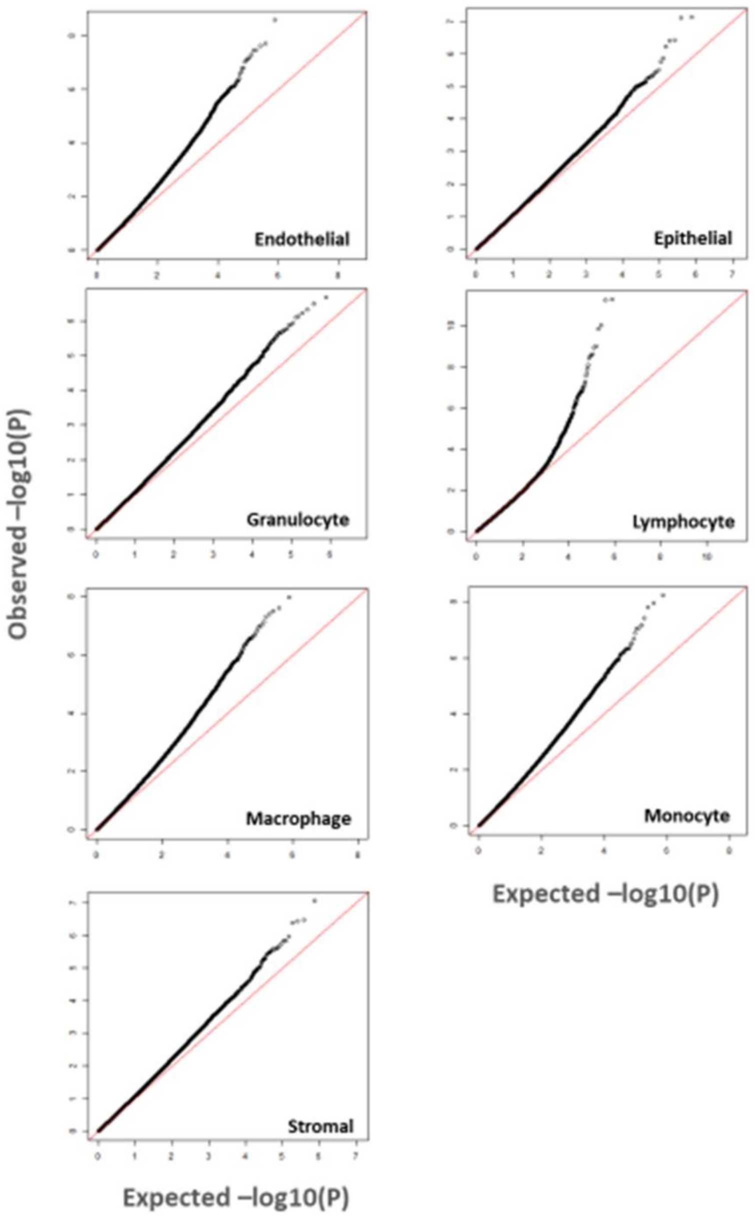


Figure S19. Q-Q plots for smoking by cell-type interactions in lung

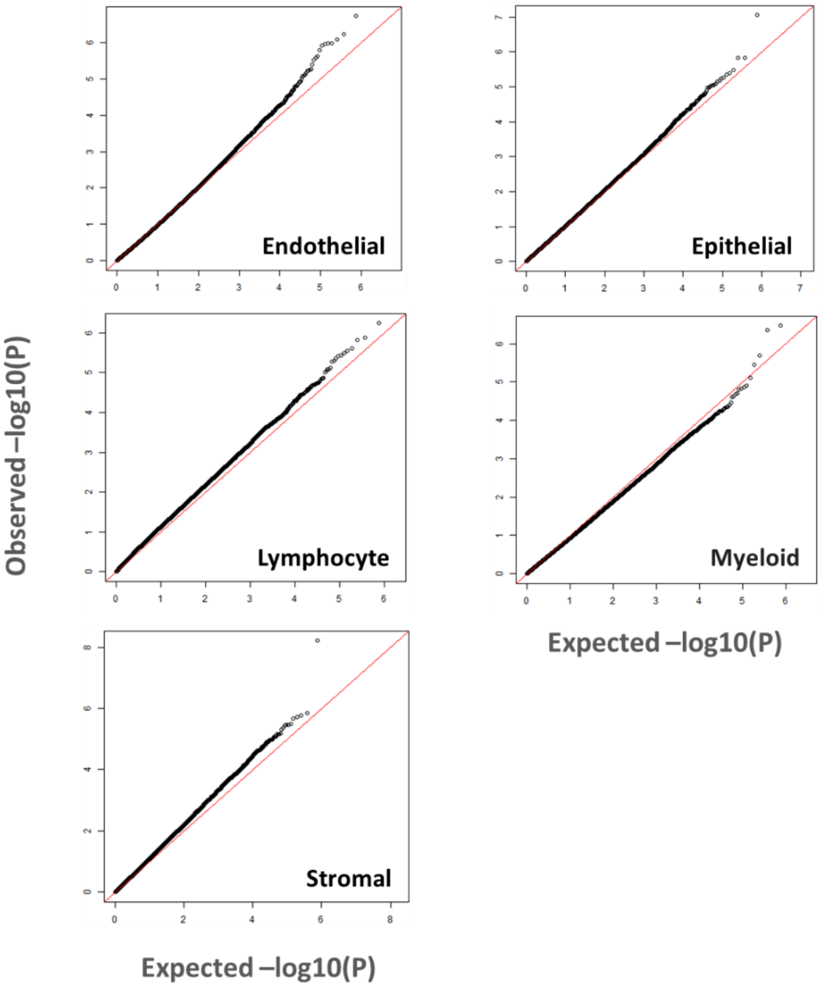


Figure S20. Q-Q plots for smoking by cell-type interactions in colon

Table S1. Correlations between EPISCORE estimates of cell type proportions with DNAm-derived SVs in lung and colon. Endo: endothelial, Epi: epithelial, Gran: granulocytes, Lym: lymphocytes, Macro: macrophages, Mono: monocytes, Stromal: stromal cells, EC: enteroendocrine cells, Mye: myeloid cells

Lung

Pearson Correlation Coefficients and P-values, n = 212										
Cell Types	**SV1**	**SV2**	**SV3**	**SV4**	**SV5**	**SV6**	**SV7**	**SV8**	**SV9**	**SV10**
Endo	-0.28	-0.4	0.47	-0.48	-0.15	-0.19	0.31	-0.14	0.003	0.12
	<.0001	<.0001	<.0001	<.0001	0.03	0.006	<.0001	0.05	0.97	0.09
Epi	-0.49	-0.65	-0.37	0.15	0.28	0.02	0.02	-0.06	0.05	-0.08
	<.0001	<.0001	<.0001	0.03	<.0001	0.73	0.73	0.41	0.44	0.25
Gran	0.59	0.65	0.27	0.02	-0.05	0.003	0.03	0.001	-0.11	-0.03
	<.0001	<.0001	<.0001	0.72	0.46	0.96	0.65	0.99	0.1	0.68
Lym	0.17	0.58	-0.34	0.28	-0.42	-0.14	0.11	0.04	-0.09	0.22
	0.01	<.0001	<.0001	<.0001	<.0001	0.05	0.11	0.59	0.21	0.001
Macro	0.74	0.48	-0.04	0.01	0.2	0.06	0.02	0.11	0.01	0.03
	<.0001	<.0001	0.59	0.94	0	0.39	0.75	0.12	0.88	0.66
Mono	0.62	0.62	0.01	0.07	0.1	0.02	-0.13	-0.04	-0.02	0.06
	<.0001	<.0001	0.9	0.29	0.14	0.75	0.06	0.56	0.78	0.37
Stromal	-0.67	-0.55	0.05	-0.07	-0.1	0.14	-0.31	0.1	0.11	-0.14
	<.0001	<.0001	0.48	0.3	0.17	0.04	<.0001	0.13	0.12	0.04

Colon

Pearson Correlation Coefficients and P-values, n = 209										
Cell Types	**SV1**	**SV2**	**SV3**	**SV4**	**SV5**	**SV6**	**SV7**	**SV8**	**SV9**	**SV10**
EC	-0.12	-0.59	0.13	-0.04	-0.09	-0.08	-0.23	0.19	-0.14	0.12
	0.08	<.0001	0.06	0.52	0.21	0.23	0.0009	0.01	0.05	0.08
Epi	<.0001	-0.6	0.14	-0.04	-0.02	-0.01	0.05	-0.005	-0.07	-0.06
	<.0001	<.0001	0.05	0.61	0.73	0.93	0.47	0.95	0.34	0.42
Lym	-0.35	0.69	-0.18	0.15	0.11	0.16	0.19	0.03	0.07	-0.15
	<.0001	<.0001	0.01	0.03	0.1	0.02	0.01	0.7	0.34	0.04
Mye	0.17	0.76	-0.18	0.14	0.23	0.26	-0.09	-0.1	0.15	0.13
	0.01	<.0001	0.01	0.05	0.0009	0.0001	0.21	0.16	0.03	0.06
Stromal	-0.86	-0.02	0.04	-0.16	-0.18	-0.28	-0.13	-0.02	-0.01	0.11
	<.0001	0.75	0.61	0.02	0.01	<.0001	0.07	0.79	0.89	0.12

Table S2a. Number of smoking-associated CpG Sites in the primary analysis of ever vs. never smokers detected in each tissue type (and the tissue-specific P-value threshold) based on false-discovery rates (FDR) of 0.01 and 0.05.

Tissue	**FDR adjusted P-value threshold**	
	0.01	**0.05**
lung (n=212)	2,478 (3.3e-5)	6,350 (0.0004)
colon (n=209)	662 (8.8e-6)	2,735 (0.0001)
ovary (n=153)	0 (N/A)	0 (N/A)
prostate (n=111)	0 (N/A)	0 (N/A)
whole blood (n=52)	0 (N/A)	0 (N/A)
breast (n=38)	0 (N/A)	0 (N/A)
testis (n=48)	0 (N/A)	0 (N/A)
kidney (n=47)	0 (N/A)	0 (N/A)
muscle (n=46)	0 (N/A)	0 (N/A)

Table S2b. Number of smoking-associated CpG Sites in the secondary analysis of current vs. never smokers.

Tissue	**Smoking-associated CpG sites (Tissue-specific P-value threshold)**	
	FDR 0.01	**FDR 0.05**
lung (n=151)	4,589 (6.11e-5)	10,495 (0.0007)
colon (n=148)	923 (1.22e-5)	4,797 (0.0003)

Table S3. Power to detect the effect sizes of smoking-associated CpGs observed in lung at different sample sizes. *Footnote*: cg01584760, cg20291548, cg09138315 are the smoking-associated CpGs with the maximum, median, and minimum effect sizes, respectively.

CpG	randomly selected n of samples		
	n=150	n=100	n=50
cg01584760	982	349	0
cg20291548	54	4	0
cg09138315	50	1	0

Table S4. Enrichment of smoking-associated CpGs based on prior studies of blood samples,[1] adipose samples,[2] placenta samples,[3] or reported in lung tissue within the current study (p-values computed through a one-sided two-proportion z-test). Red highlighted cells indicate p-values less than 0.05.

	Blood CpGs (p-value threshold: 1E-3)	**Lung CpGs (p-value threshold: 1E-3)**	**Adipose CpGs (p-value threshold: 1E-3)**	**Placenta CpGs (p-value threshold: 1E-3)**	**Blood CpGs (p-value threshold: 1E-5)**	**Lung CpGs (p-value threshold: 1E-5)**	**Adipose CpGs (p-value threshold: 1E-5)**	**Placenta CpGs (p-value threshold: 1E-5)**
Breast	3.30E-02	3.49E-08	1.38E-02	8.08E-02	5.00E-01	5.00E-01	5.00E-01	5.00E-01
Colon	3.09E-47	5.79E-43	4.35E-23	5.00E-01	1.23E-30	1.34E-06	9.10E-17	4.17E-01
Kidney	2.83E-02	1.92E-05	2.29E-33	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01
Muscle	5.00E-01	3.03E-01	1.06E-12	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01
Ovary	7.22E-05	4.60E-09	5.00E-01	4.26E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01
Prostate	4.65E-01	2.57E-02	1.43E-09	5.00E-01	5.00E-01	1.83E-01	5.00E-01	5.00E-01
Testis	5.00E-01	5.82E-02	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01
Whole Blood	8.16E-46	2.22E-08	3.00E-196	5.00E-01	5.00E-01	5.00E-01	5.00E-01	5.00E-01
Lung	7.46E-116	0.00E+00	4.44E-265	1.48E-10	1.43E-91	0.00E+00	0.00E+00	8.14E-07

Table S5. Genes with associations between smoking and both DNAm and gene expression data at an FDR<0.05 in lung. (table in separate file)

Table S6. Genes with associations between smoking and both DNAm and gene expression data at an FDR<0.05 in colon. (table in separate file)

Table S7. Colocalization between smoking-associated CpGs in lung (FDR<0.01), mQTLs, and the 10 SNPs reaching genome-wide significance in the UK Biobank FEV1/FVC GWAS. (table in separate file)

Table S8. Colocalization between smoking-associated CpGs in colon (FDR<0.05), mQTLs, and genome-wide significant SNPs identified in genome-wide association studies of colon-related diseases. (table in separate file)

Table S9. Top hallmark gene sets detected in the primary analysis of hypomethylated smoking-associated CpGs in lung.

Description	Genes in gene set	Genes with smoking-associated CpGs*	Enrichment P	FDR-adjusted p-value
Hallmark Gene Sets				
TNF-alpha signaling via NFKb	199	49	5.01E-07	2.50E-05
P53 pathway	196	46	5.01E-05	1.25E-03
Apoptosis	155	38	1.58E-04	2.64E-03
Hypoxia	190	45	4.68E-04	5.85E-03
IL6-JAK-STAT3 signaling	81	20	9.44E-04	9.44E-03
Early response to estrogen	194	49	1.75E-03	1.35E-02
IL2-STAT5 signaling	194	44	1.90E-03	1.35E-02
MTORC1 signaling	194	38	3.02E-03	1.89E-02
Cholesterol homeostasis	71	18	3.59E-03	1.99E-02
TGF-beta signaling	53	17	6.18E-03	2.87E-02
PI3K-AKT-MTOR signaling	103	26	6.31E-03	2.87E-02
Androgen response	97	25	8.80E-03	3.41E-02
Xenobiotic metabolism	197	36	8.86E-03	3.41E-02
Genes down-regulated in response to ultraviolet (UV) radiation	142	40	1.00E-02	3.58E-02

*Genes with CpGs (as assigned by Illumina) that are associated with smoking

Table S10. Top hallmark gene sets detected in the pathway analysis of smoking-associated CpGs in colon tissue.

Hallmark gene sets	Genes in gene set	Genes with smoking-associated CpGs [1]	Enrichment P	FDR
Epithelial to mesenchymal transition	192	43	9.59e-8	4.80e-6
UV response DN	142	30	2.12e-3	0.05

[1] Genes with CpGs (as assigned by Illumina) that are associated with smoking

Table S11. Top hallmark gene sets and KEGG pathways detected in the secondary analysis of smoking-associated CpGs in lung.

Description	Genes in gene set	Genes with smoking-associated CpGs*	Enrichment P	FDR-adjusted p-value
Hallmark Gene Sets				
TNF-alpha signaling via NFKb	199	76	5.91E-07	2.96E-05
P53 pathway	196	75	1.42E-05	3.54E-04
IL2-STAT5 signaling	194	77	9.89E-05	1.65E-03
Early response to estrogen	194	82	3.07E-04	3.84E-03
Apoptosis	155	58	5.66E-04	5.57E-03
Complement	194	69	6.68E-04	5.57E-03
Allograft rejection	191	63	8.56E-04	5.98E-03
IL6-JAK-STAT3 signaling	81	31	1.06E-03	5.98E-03
Xenobiotic metabolism	197	65	1.08E-03	5.98E-03
Inflammatory response	196	62	1.30E-03	6.50E-03
Late response to estrogen	194	72	1.52E-03	6.89E-03
Adipogenesis	196	65	1.74E-03	7.04E-03
TGF-beta signaling	53	27	1.92E-03	7.04E-03
Cholesterol homeostasis	71	29	1.97E-03	7.04E-03
Genes up-regulated by KRAS activation	192	66	3.01E-03	1.00E-02
MTORC1 signaling	194	63	3.85E-03	1.20E-02
Genes down-regulated in response to ultraviolet (UV) radiation	142	65	4.41E-03	1.30E-02
Fatty acid metabolism	149	43	9.35E-03	2.60E-02
Hypoxia	190	65	1.53E-02	4.02E-02
Interferon gamma response	200	58	1.67E-02	4.16E-02
Apical junction	193	70	1.82E-02	4.34E-02
Bile acid metabolism	110	34	2.04E-02	4.63E-02
KEGG Pathways				
Lipid and atherosclerosis	205	79	6.41E-06	2.28E-03
PPAR signaling pathway	72	31	1.20E-04	2.13E-02
Pathways in cancer	510	181	2.48E-04	2.80E-02

Parathyroid hormone synthesis, secretion and action	105	51	3.15E-04	2.80E-02
Circadian entrainment	96	47	5.64E-04	4.00E-02
PI3K-Akt signaling pathway	342	124	1.00E-03	4.57E-02
Insulin resistance	105	45	1.02E-03	4.57E-02
Non-small cell lung cancer	71	36	1.03E-03	4.57E-02

*Genes with CpGs (as assigned by Illumina) that are associated with smoking

Table S12. CpGs showing strongest evidence of being impacted by the interaction between smoking and estimated cell types.

					Interaction Effect		Main Effect	
Cell-type	**Name**	**Chromosome**	**Position**	**Gene**	**log2[Fold Change]**	**P-Value**	**log2[Fold Change]**	**P-Value**
Lymphocyte	cg26075905	14	102391388	*PPP2R5C*	0.22	5.10E-12	0.06	0.08
Lymphocyte	cg17616967	2	234234525	*SAG*	0.94	5.40E-12	0.34	0.01
Lymphocyte	cg26848446	11	3382329	*ZNF195*	0.48	9.20E-11	0.23	0.002
Lymphocyte	cg12529083	12	13197360	*KIAA1467*	-0.28	1.38E-10	-0.15	0.0006
Lymphocyte	cg12722058	15	23086394	*NIPA1*	-0.28	9.10E-10	-0.22	2.01E-06
Lymphocyte	cg12901038	11	2815078	*KCNQ1*	0.39	1.15E-09	0.13	0.05
Lymphocyte	cg09197895	14	105512268	N/A	-0.22	2.36E-09	-0.14	0.0001
Lymphocyte	cg04180093	7	99686359	*COPS6*	-0.52	2.70E-09	-0.24	0.007
Lymphocyte	cg12285709	14	102975664	*ANKRD9*	-0.32	2.74E-09	-0.07	0.17
Lymphocyte	cg06509613	18	53588274	*LINC01416*	0.38	3.78E-09	0.04	0.5
Mono	cg24096902	1	27754893	*WASF2*	0.23	5.90E-09	0.16	8.50E-05
Mono	cg23606722	12	133319848	*ANKLE2*	0.41	1.10E-08	0.18	0.01
Macro	cg11538937	2	241559124	*GPR35*	-0.20	1.10E-08	-0.11	0.003
Macro	cg24096902	1	27754893	*WASF2*	0.23	2.50E-08	0.18	6.40E-05
Endo	cg06232680	21	39608414	*KCNJ15*	0.346	2.60E-09	-0.049	0.367
Endo	cg14800014	5	125800764	*GRAMD3*	0.212	1.90E-08	-0.131	0.0003

Footnote: *Abbreviations*: Mono, Monocyte; Macro, Macrophage; Endo, Endothelial cell.

Table S13. Smoking-by-cell type interaction results for the CpGs in lung showing the strongest evidence of association with smoking in the primary EWAS analysis. (table in separate file)

References:

1. Silva CP, Kamens HM. Cigarette smoke-induced alterations in blood: A review of research on DNA methylation and gene expression. *Exp Clin Psychopharmacol*. 2021;29(1):116-135. doi:10.1037/pha0000382

2. Tsai PC, Glastonbury CA, Eliot MN, et al. Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenetics*. 2018;10(1):126. doi:10.1186/s13148-018-0558-0

3. Everson TM, Vives-Usano M, Seyve E, et al. Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth. *Nat Commun*. 2021;12(1):5095. doi:10.1038/s41467-021-24558-y