

網羅的エピゲノム解析を用いた化学物質による次世代影響の解明：新しい  
試験スキームへの基礎的検討

研究代表者	岸 玲子	北海道大学環境健康科学研究教育センター	特別招へい教授
研究分担者	三宅 邦夫	山梨大学大学院総合研究部医学域	准教授
研究分担者	三浦 りゅう	北海道大学環境健康科学研究教育センター	特任助教
研究分担者	荒木 敦子	北海道大学環境健康科学研究教育センター	特任准教授

### 研究要旨

**背景：**胎児期や乳幼児期の化学物質曝露が子供の様々な疾患リスクに影響することが報告されている。その分子メカニズムとして DNA メチル化形成異常が関与している可能性が考えられているが、これまで疫学研究、特に日本人集団での検討はあまりなされていない。

**目的：**本研究は、胎児期の喫煙曝露により影響を受ける DNA メチル化領域を同定し、次世代シーケンサーを用いたメチル化解析による検証法を確立することを目的とした。

**方法：**北海道スタディ札幌コーホート参加者 514 名のうち、初期調査票・出産時カルテ情報、臍帯血 DNA のそろそろ母児ペア 291 名を対象とし、喫煙経験のない非喫煙群 (n = 124)、妊娠中も喫煙を継続した喫煙群 (n = 46)、妊娠がわかって喫煙を止めた喫煙中止群 (n = 77) の 3 群を解析対象とした。Infinium HumanMethylation450 BeadChip による網羅的解析から喫煙曝露と DNA メチル化変化を解析し、再現性の検証は次世代シーケンサーを用いたバイサルファイトシーケンス法でメチル化率を算出した。

**結果：**非喫煙群と喫煙群の比較で 46 CpGs 領域 (27 遺伝子) が同定された。さらに喫煙を中止することにより DNA メチル化が非喫煙者と同等レベルになる 9 CpGs、喫煙を中止しても変化しない 1 CpG を見出すことができた。再現性検証を行なった結果、8 CpGs のうち 7 CpGs で網羅的 DNA メチル化解析結果と同様のメチル化変化が確認でき、スピアマンの相関係数は  $> 0.6$  で高い相関を示した。

**考察：**以上の結果から、別の大規模集団において次世代シーケンサーを用いた DNA メチル化解析の利用が可能であると考えられる。

### A．研究目的

胎児期の化学物質曝露が子供の様々な疾患リスクに影響することが報告されている (Tran et al., 2017)。将来的ながんや喘息のリスクだけでなく、近年、自閉症スペクトラム障害 (ASD) や注意欠陥多動性障害 (ADHD) などの発達障害のリスクを増加させる可能性も指摘されている (Melchior et al., 2015, Joelsson et al., 2016)。胎児期曝露の影響が児の健康リスクに影響を及ぼすメカニズムとしてエピジェネティクスの中心的メカニズムであ

る DNA メチル化変化の関与が考えられている。DNA メチル化の形成は胎児期や乳幼児期で特に重要 (臨界期) (Lister et al., 2015) であることから、この時期の化学物質曝露と疾患発症リスクをつなぐ分子メカニズムとして DNA メチル化形成異常が関与している可能性が考えられるが、これまで疫学研究、特に日本人集団での検討はあまりなされていない。

網羅的エピゲノム解析は全遺伝子領域を含む約 45 万 CpG を解析できる一方で、偽陽性を検出する可能性も考えられる。ま

た疫学研究のような大規模集団において全て網羅的解析を行うことは費用面から考えても困難であり、網羅的解析により同定された特定の遺伝子領域について別手法による検証法の確立が必要である。

そこで本研究では、胎児期の喫煙曝露に着目し、網羅的 DNA メチル化解析により喫煙曝露により変化する DNA メチル化領域を同定し、次世代シーケンサーを用いたメチル化解析による検証法を確立することを目的とし、胎児期喫煙曝露と子供の疾患リスクをつなぐ DNA メチル化領域を明らかにする。

## B．研究方法

### 1．対象

北海道スタディ札幌コーホート参加者 514 名のうち、初期調査票・出産時カルテ情報、臍帯血 DNA のそろう母児ペア 291 名を対象とし、喫煙経験のない非喫煙群 (n = 124)、妊娠中も喫煙を継続した喫煙群 (n = 46)、妊娠がわかって喫煙を止めた喫煙中止群 (n = 77) の 3 群を解析対象とした。なお妊娠前喫煙中止群 (中止群) (n = 44) については、禁煙時期が妊娠時を基準とした時期から対象者毎に異なるため解析から除外した。

### 2．方法

網羅的な DNA メチル化は Infinium HumanMethylation450 BeadChip (イリミナ社) により約 45 万 CpGs のデータを取得し、先行文献を参考にデータの前処理 (標準化、バッチ効果補正など) (Aryee et al., 2014, Fortin et al., 2014, Chen et al., 2013, Touleimat et al., 2012, Leek et al., 2012), 統計モデル (細胞組成、母年齢、児性別、母教育歴で調整) を構築 (Fox et al., 2012, Smyth et al., 2004, Houseman et al., 2014), 喫煙曝露によりメチル化変

化する領域 (CpG) の同定を行った。同定された CpG 領域の再現性を検証するため、同一サンプルの DNA をバイサルファイト PCR, エンドリペア, アダプター・バーコード付加によりライブラリー調整を行い、IonChef によるテンプレート調整後、Ion PGM 次世代シーケンサー (サーモフィッシャー社) を用いてシーケンスデータの取得し、各 CpG サイトのメチル化率を算出した。

### (倫理面への配慮)

本研究は、北海道大学環境健康科学研究教育センター、北海道大学大学院医学研究科及び山梨大学医学部の倫理委員会の承認を得た。本研究によって得られた個人名及び個人データの漏洩については、データの管理保管に適切な保管場所を確保するなどの方法により行うとともに、研究者の道義的責任に基づいて個人データをいかなる形でも本研究の研究者以外の外部の者に触れられないように厳重に保管し、取り扱った。

## C．研究結果

網羅的エピゲノム解析から、ゲノムワイド有意水準 (FDR < 0.05) かつメチル化変化の大きさ ( $|$ 偏回帰係数 $| > 0.02$ ) の条件で、非喫煙群と喫煙群の比較で 46CpGs 領域 (27 遺伝子) が同定された。また同様の条件で非喫煙群と中止群の比較で 15CpGs 領域 (5 遺伝子)、中止群と喫煙群の比較で 64CpGs 領域 (38 遺伝子) をそれぞれ同定した。これらの 3 つの解析から共通する CpGs を抽出することにより、喫煙を中止することにより DNA メチル化が非喫煙者と同等レベルになる 9CpGs (Figure 1), 喫煙を中止しても変化しない

1CpG (Figure 2)を見出すことができた。

さらに同一のDNAサンプルを用いて次世代シーケンサー(NGS)による再現性検証を行った結果、8CpGsのうちAHRH (cg21161138)を除く7CpGsで網羅的DNAメチル化解析(450K)結果と同様のメチル化変化が確認でき、450KとNGSのスピアマンの順位相関係数は $> 0.6$ で高い相関を示した(Figure 3)。

網羅的メチル化解析では設計されていない周辺CpGについてのメチル化状態を次世代シーケンサー解析により検討した結果、多くのCpGで同方向にメチル化変化していることが確認できた(Table 1)。

#### D．考察

本研究により、胎児期喫煙曝露により影響を受ける遺伝子として従来の報告と同様にAHRH, CYP1A1, MYO1Gなどが含まれる一方で、シナプス関連分子であるSHANK2やタンパク質の分解に関連するTRIM36などこれまで報告されていない新たな遺伝子領域を同定することができた。特にSHANK2遺伝子は自閉症の原因とされる遺伝子の1つとして報告されていることから、喫煙曝露による神経発達障害発症メカニズムに関与しているかもしれない。また妊娠が分かって喫煙を中止することによりDNAメチル化が非喫煙者と同等レベルになる遺伝子を多数見いだせたことから、妊娠初期に喫煙を止めればDNAメチル化レベルで影響を及ぼさない可能性が示唆された。

大規模コーホートにおける多サンプルの解析では網羅的メチル化解析はコスト面からも現実的ではない。そこで特異的な領域に対して多サンプルの定量的解析が求められる。本研究において同一DNAサ

ンプルを用いて次世代シーケンサーによるメチル化解析を行いデータの再現性を検証した。その結果、8CpGsのうち7CpGsで同様のメチル化変化が確認できたことから結果の妥当性ならびに次世代シーケンサーによるメチル化解析法が確立できた。しかしながら、1CpGについては網羅的解析の偽陽性である可能性と次世代シーケンサー解析の影響も考えられることから更なる検証を行う必要がある。今後、大規模コーホートを用いて、化学物質の影響を受ける特定のDNAメチル化領域について様々な疾患との関与について解析していく予定である。

#### E．結論

網羅的エピゲノム解析から、胎児期の喫煙曝露により影響を受ける遺伝子を同定することができた。次世代シーケンサーを用いて特異的な遺伝子領域のメチル化解析法を確立し、同定されたほとんどのCpGで再現性が確認できた。

#### F．研究発表

##### 1. 論文発表

なし

##### 2. 学会発表

三宅邦夫, 三浦りゅう, 小林祥子, 小林澄貴, 宮下ちひろ, 荒木敦子, 山縣然太郎, 岸玲子. 妊娠中の喫煙曝露における臍帯血を用いたDNAメチル化変化領域の同定. 第46回日本環境変異原学会. 東京都千代田区. 2017.11.6.-7.

#### G．知的財産権の出願・登録状況

該当なし

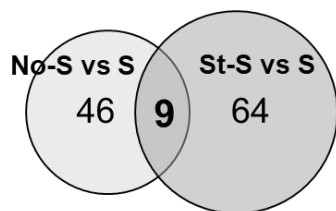
## 参考文献

1. Tran NQV, Miyake K. 2017. Neurodevelopmental Disorders and Environmental Toxicants: Epigenetics as an Underlying Mechanism. *Int J Genomics*. 2017:7526592.
2. Melchior M, Hersi R, van der Waerden J, Larroque B, Saurel-Cubizolles MJ, Chollet A, Galéra C; EDEN Mother-Child Cohort Study Group. 2015. Maternal tobacco smoking in pregnancy and children's socio-emotional development at age 5: The EDEN mother-child birth cohort study. *Eur Psychiatry*. 30(5):562-8.
3. Joelsson P, Chudal R, Talati A, Suominen A, Brown AS, Sourander A. 2016. Prenatal smoking exposure and neuropsychiatric comorbidity of ADHD: a finnish nationwide population-based cohort study. *BMC Psychiatry*. 16:306.
4. Lister R, Mukamel EA. 2015. Turning over DNA methylation in the mind. *Front Neurosci*. 9:252.
5. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 30:1363-1369.
6. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD. 2014. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*. 15: 503..
7. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 8: 203-209.
8. Touleimat, N. & Tost, J. 2012. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 4: 325-341.
9. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 28: 882-883.
10. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R. 2011. High density DNA methylation array with single CpG site resolution. *Genomics*. 98: 288-295.
11. Fox, J. & Weisberg, S. 2011. Robust Regression in R: *An Appendix to An R Companion to Applied Regression*.

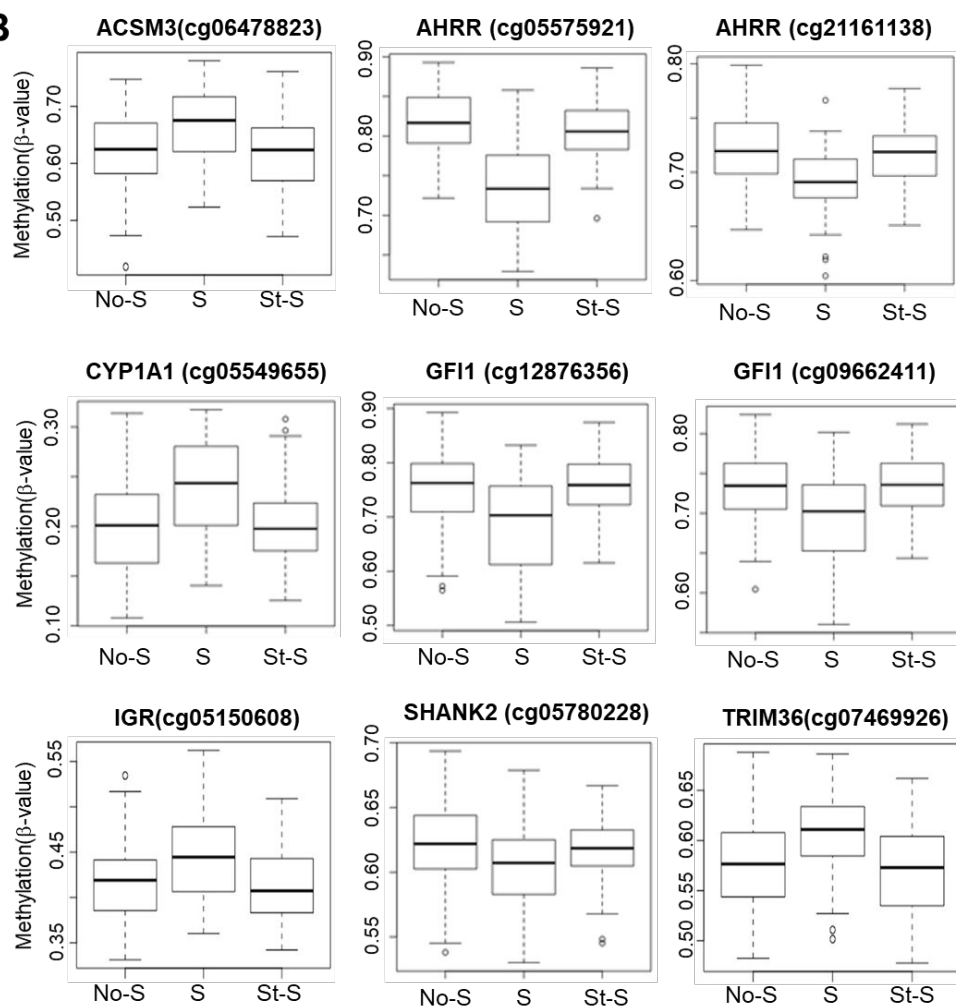
2nd edn.

12. Smyth, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3.
- C.J. 2014. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics*. 30:1431-9.
13. Houseman, E.A., Molitor, J., Marsit,

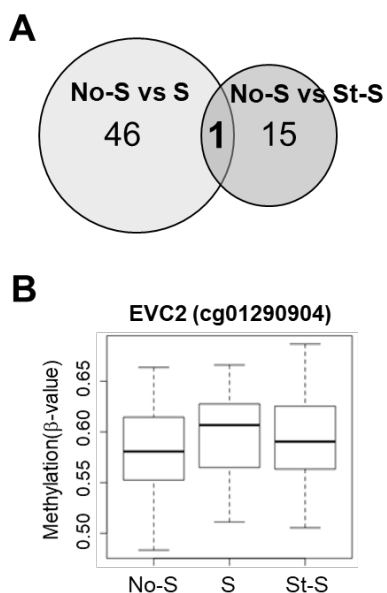
**A**



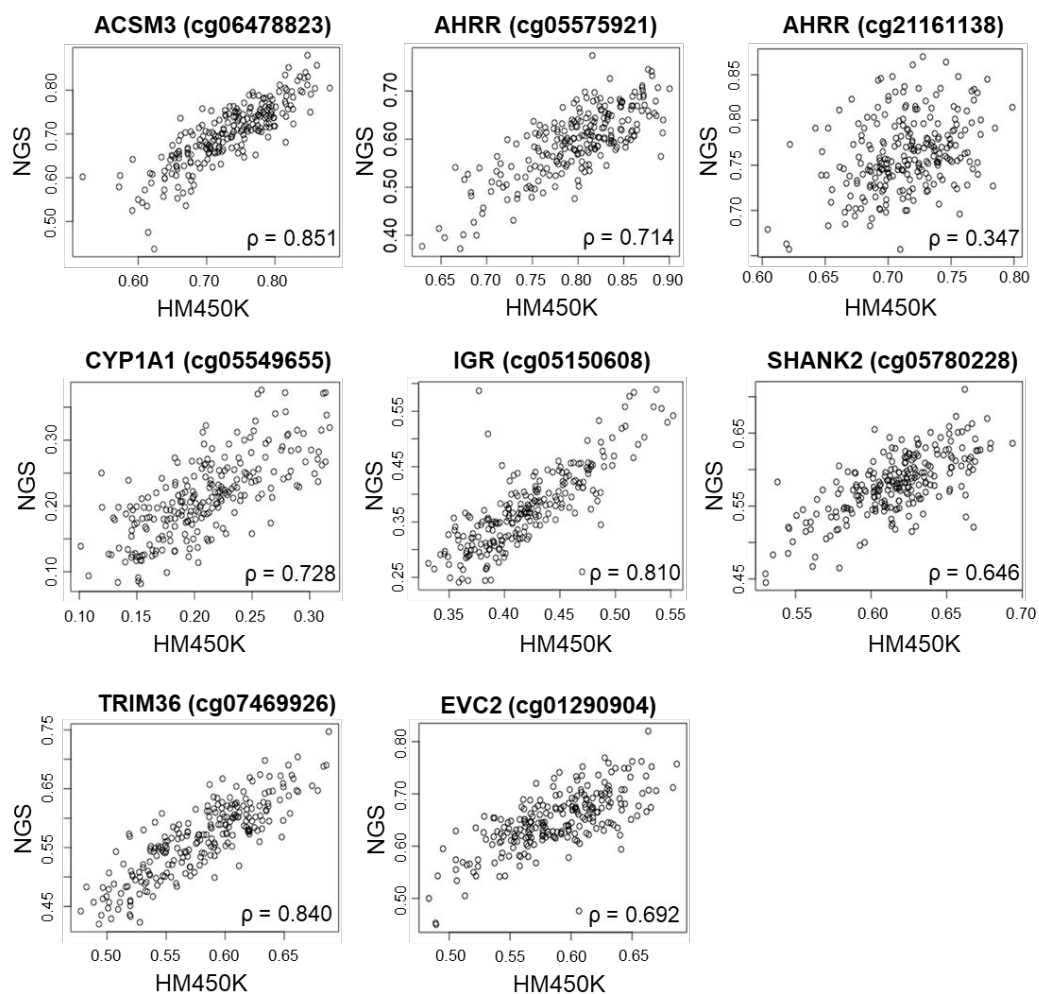
**B**



**Figure 1.** Secondary comparison of differentially methylated CpG sites identified in the Ne-S vs. Su-S comparison relative to those identified in the St-S vs. Su-S comparison. (A) Venn diagram showing 9 CpG sites that were common between the Ne-S vs. Su-S comparison (46 sites) and the St-S vs. Su-S comparison (64 sites). (B) Box plots showing methylation rates ( $\beta$ -value) at 9 common CpG sites.



**Figure 2.** Secondary comparison of differentially methylated CpG sites identified in the Ne-S vs. Su-S comparison relative to those identified in the Ne-S vs. St-S comparison. (A) Venn diagram showing one CpG site that was common between the Ne-S vs. Su-S comparison (46 sites) and the Ne-S vs. St-S comparison (15 sites). (B) Box plots showing methylation rates ( $\beta$ -value) at one common CpG site.



**Figure 3.** Correlation between HumanMethylation450K array (HM450K) and next generation sequencing (NGS) data analysis for the DNA methylation status of eight CpG sites after removing outliers. Values of Spearman correlation coefficient ( $\rho$ ) are indicated.



**Table 1.** Verification result for selected CpG sites using Next Generation Sequencing

Gene region	CpG site <sup>a</sup>	P-value			Mean (SD <sup>e</sup> ) methylation (ratio)		
		Ne-S <sup>b</sup> vs Su-S <sup>c</sup>	Ne-S vs St-S <sup>d</sup>	St-S vs Su-S	Ne-S	St-S	Su-S
ACSM3	cg59 (cg06478823)	0.032 *	0.728	0.033 *	0.700 (0.062)	0.690 (0.078)	0.722 (0.080)
AHRR	cg26	1.17E-10 **	0.258	1.55E-08 **	0.644 (0.052)	0.636 (0.049)	0.564 (0.068)
	cg28	6.64E-12 **	0.205	6.24E-09 **	0.640 (0.054)	0.630 (0.050)	0.549 (0.067)
	cg51 (cg05575921)	1.05E-11 **	0.625	9.76E-11 **	0.615 (0.060)	0.609 (0.050)	0.519 (0.072)
	cg71	3.73E-12 **	0.587	1.48E-10**	0.738 (0.064)	0.733 (0.057)	0.611 (0.097)
	cg96	1.33E-10**	0.388	6.18E-08 **	0.679 (0.058)	0.670 (0.057)	0.579 (0.082)
AHRR	cg26	0.964 (0.984)	0.768 (0.826)	0.980 (0.980)	0.854 (0.034)	0.851 (0.037)	0.854 (0.035)
	cg56	0.007 ** (0.656)	0.827 (0.892)	0.042* (0.429)	0.785 (0.150)	0.814 (0.111)	0.768 (0.150)
	cg79 (cg21161138)	0.064 (0.064)	0.669 (0.576)	0.280 (0.350)	0.767 (0.040)	0.755 (0.063)	0.749 (0.046)
CYP1A1	cg26 (cg05549655)	8.47E-04 **	0.941	7.06E-04 **	0.210 (0.061)	0.202 (0.053)	0.245 (0.061)
	cg31	0.008 **	0.986	0.005 **	0.242 (0.066)	0.236 (0.054)	0.274 (0.060)
	cg42	0.009 **	0.312	3.67E-04 **	0.480 (0.081)	0.461 (0.073)	0.518 (0.071)
	cg49	0.166	0.288	0.007 **	0.470 (0.070)	0.453 (0.065)	0.491 (0.070)
	cg57	8.93E-4 ** (0.002 **)	0.781 (0.781)	7.07E-5 ** (2.19E-04 **)	0.566 (0.077)	0.558 (0.064)	0.604 (0.069)
IGR	cg26	0.72 (0.849)	0.600 (0.600)	0.303 (0.422)	0.371 (0.076)	0.367 (0.070)	0.374 (0.087)
	cg75 (cg05150608)	0.024 *	0.736	0.010 *	0.368 (0.066)	0.365 (0.071)	0.403 (0.080)
	cg77	0.044 *	0.707	0.041 *	0.376 (0.078)	0.372 (0.085)	0.415 (0.110)
SHANK2	cg39 (cg05780228)	2.50E-04 ** (1.98E-04 **)	0.287 (0.246)	0.034 * (0.034 *)	0.590 (0.042)	0.579 (0.040)	0.558 (0.046)
	cg67	0.09 (0.120)	0.986 (0.964)	0.163 (0.163)	0.937 (0.018)	0.938 (0.013)	0.932 (0.017)
	cg74	0.025 *	0.793	0.022 *	0.848 (0.027)	0.849 (0.028)	0.837 (0.027)
TRIM36	cg26	0.003 **	0.777	0.003 **	0.549 (0.085)	0.538 (0.088)	0.602 (0.098)
	cg38	0.031 *	0.562	0.009 **	0.569 (0.082)	0.556 (0.087)	0.606 (0.088)
	cg40 (cg07469926)	0.005 **	0.678	0.002 **	0.563 (0.057)	0.551 (0.065)	0.595 (0.063)
EVC2	cg34 (cg01290904)	0.007 ** (0.007) **	0.0526 (0.042 *)	0.429 (0.527)	0.637 (0.056)	0.658 (0.088)	0.667 (0.054)

<sup>a</sup>CpG sites are numbered according to their position on the sequencing products, overlapped CpG sites with Human Methylation 450K array are indicated in parentheses.

<sup>b</sup>Never-smoker. <sup>c</sup>Sustained-smoker. <sup>d</sup>Stopped-smoker. <sup>e</sup>Standard deviation.

Steel-Dwass tests. \*, P < 0.05; \*\*, P < 0.01, P value from analysis before removal of outliers are indicated in parentheses.