DNA Methylation as a Long-term Biomarker of Exposure to Tobacco Smoke

Natalie S. Shenker,[a] Per Magne Ueland,[b] Silvia Polidoro,[c] Karin van Veldhoven,[d] Fulvio Ricceri,[c] Robert Brown,[a] James M. Flanagan,[a] and Paolo Vineis[c,d]

Background: Most biomarkers of exposure tend to have short half-lives. This includes cotinine, a metabolite of nicotine widely used to assess smoke exposure. Cotinine is thus unsuitable as a determinant of past exposure to cigarette smoke.
Methods: We used bisulphite pyrosequencing of a set of four genomic loci (*AHRR*, 6p21, and two at 2q37) that had differential DNA methylation levels in peripheral blood DNA dependent on tobacco exposure to create a predictive model of smoking status.
Results: Combining four gene loci into a single methylation index provided high positive predictive and sensitivity values for predicting former smoking status in both test (n = 81) and validation (n = 180) sample sets.
Conclusions: This study provides a direct molecular measure of prior exposure to tobacco that can be performed using the quantitative approach of bisulphite pyrosequencing. Epigenetic changes that are detectable in blood may more generally act as molecular biomarkers for other exposures that are also difficult to quantify in epidemiological studies.

(*Epidemiology* 2013;24: 712–716)

Epidemiological studies increasingly rely on biomarkers of exposure.[1,2] However, most biomarkers tend to be short-lived, with half-lives of only days to months. This is an important limitation for the investigation of diseases, such as cancer, with long latency periods. A well-validated biomarker for tobacco smoking is cotinine, which has a half-life of only 16 hours.[3,4] As a result, cotinine does not distinguish between former smokers and those who have never smoked and can validate only whether former smokers have actually quit smoking.[5] The identification of a persistent biomarker of tobacco exposure would not only be useful for molecular epidemiology but also would suggest a paradigm for quantification of other exposures that are also difficult to measure.

Epigenetic modifications such as DNA methylation and histone modification are key determinants of chromatin structure and gene expression. These modifications are maintained during cell division and, when perturbed, play a key role in cancer development.[6–8] Epigenetic changes may also represent a biological indicator of lifetime accumulation of environmental exposures related to aging,[9] hormones,[10] ionizing radiation,[11] alcohol,[12] smoking,[1,13] and perhaps many others.

We have previously performed a DNA methylation study of white blood cell DNA within a large prospective cohort of current, former, and never smokers, based on the results of two epigenetic-wide association studies including 374 subjects (half who subsequently developed breast or colon cancer and half who were healthy controls) and a validation cohort of 180 subjects.[14] Decreased methylation levels at eight genomic loci were associated with current smoking using a cutoff of $P < 10^{-7}$ (Bonferroni corrected) and several were also validated by bisulphite pyrosequencing in an independent sample set. Here, we assess the performance of DNA methylation measured by bisulphite pyrosequencing at four selected genomic loci, combined into a methylation index (MI), as a biomarker of former exposure to tobacco smoke.

METHODS

Study Subjects

All study participants were drawn from the Turin component of the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) cohort, a general population cohort that consists of approximately 10,000 persons with standardized lifestyle and personal history questionnaires, anthropometric data, and blood samples collected for DNA extraction.[15,16] Smoking status was ascertained from questionnaire data. For the test sample set, 81 healthy persons were sampled from 1,805 who had been previously measured for

Submitted 11 January 2013; accepted 5 April 2013; posted 17 July 2013.

From the [a]Department of Surgery and Cancer, Epigenetics Unit, Imperial College London, London, United Kingdom; [b]Section of Pharmacology, Institute of Medicine, University of Bergen, Bergen, Norway; [c]HuGeF Foundation, Torino, Italy; and [d]MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom.

J.M.F. is funded by a Breast Cancer Campaign Fellowship. J.M.F. and R.B. acknowledge funding from Cancer Research UK (A13086) and the Imperial Biomedical Research Centre. P.V. is funded by the HuGeF Foundation, Torino, Italy. N.S.S. is funded by a Medical Research Council UK graduate scholarship.
The authors have declared that they have no relationships that could be construed as resulting in an actual, potential or perceived conflict of interest with regard to the article submitted.

Correspondence: James M. Flanagan, Epigenetics Unit, Division of Cancer, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, 4th Floor IRDB, Hammersmith Campus, Du Cane Road, London W12 0NN. E-mail: j.flanagan@imperial.ac.uk.

Copyright © 2013 by Lippincott Williams & Wilkins
ISSN: 1044-3983/13/2405-0712
DOI: 10.1097/EDE.0b013e31829d5cb3

serum cotinine,[17] including 33 nonsmokers, 30 former smokers, and 18 current smokers. These included 66 men and 15 women with similar smoking status. Smoking duration was calculated as age at recruitment minus age at smoking initiation (for current smokers) and age at quitting minus age at smoking initiation (for former smokers). For the validation component of this study, 180 healthy women were randomly sampled from the EPIC-Turin cohort (n = 102 nonsmokers, n = 45 former smokers, n = 33 current smokers).

Laboratory Analysis

DNA samples were extracted from buffy coats using the QIAsymphony DNA Midi Kit (Qiagen, Crawley, UK). Genomic DNA (250 ng) from each subject was bisulphite converted and pyrosequenced as described previously.[18] The loci included one in the aryl hydrocarbon receptor repressor (*AHRR*) gene (Chr5: 373,299), two intergenic loci at 2q37 (Chr2: 233,284,112 and Chr2:233,284,661), and one intergenic locus at 6p21.33 (Chr6: 30,720,080). Individuals (test, n = 11 and validation, n = 42) that were heterozygous for a previously identified G→A single nucleotide polymorphism at the first locus in *AHRR* (Chr5: 373,299) were removed from the analysis to avoid confounding of the *AHRR* methylation results. The pyrosequencing assay for this locus determines both the methylation assay and the genotype in the same assay for this cytosine-guanine dinucleotide (CG) site.[14] Alternative CG sites that could be used in place of this *AHRR* site included cg21161138, for which a pyrosequencing assay was developed, or cg05575921, which requires the further development of a pyrosequencing assay. Illumina 450K DNA methylation microarray processing and analysis methods have been described elsewhere.[14] The pyrosequencing data for the four loci in the validation cohort (n = 180) were generated in a previous study.[14] As these data represent actual methylation values, there was no normalization performed on the pyrosequencing data.

Statistical Analysis

Wilcoxon ranked-sum tests were used to compare nonparametric data. Pearson tests were performed to assess any correlation between methylation values at various genomic loci. The methylation model was constructed using a stepwise iterative generalized linear regression model of the data for never smokers versus former smokers starting with pyrosequencing data from six loci with previously reported assays.[14] We excluded one *AHRR* CG site (cg21161138) and F2RL3 (cg03636183) that were not independent of other loci ($P > 0.05$ in the linear regression model) and were highly correlated with the other four markers ($R > 0.6$). The sensitivity, specificity, positive predictive value (PPV), and negative predictive value were calculated using receiver operative curve (ROC) analyses. Area under the curve (AUC) values were calculated for each genomic locus and the overall MI and compared with those for cotinine in differentiating between never versus former smokers. Binomial tests were used to assess PPVs. All statistical analyses were performed in R, v2.13.1.

RESULTS

The association between methylation levels in four of the genomic loci and smoking status that was observed previously[14] was confirmed in the test set of 81 persons (Table 1 and Figure 1). Cotinine levels above the cutoff of 15 ng/mL[5,19] were associated with current smoking status ($P < 0.0001$, Table 2) with a high predictive value for current

TABLE 1. Mean Methylation Percentages for the Four Cytosine–Guanine Nucleotides Identified from the Microarray Study as Differentially Methylated in Current Smokers vs. Former and Nonsmokers, in Addition to the Mean Cotinine Values (ng/mL) in Each Group of Individuals

Marker	Genomic Locus (Illumina Identification)	Smoking Status			Sensitivity (%)	Specificity (%)	AUC Never vs. Former (95% CI)	PPV Using Optimum Thresholds
		Never (n = 33)	Former (n = 30)	Current (n = 18)				
Test set (n = 33 never smokers, 30 former smokers, 18 current smokers)								
AHRR_p1	Chr5:373,299 (cg23576855)	79.6	73.3	59.6	0.65	0.67	0.71 (0.56–0.86)	0.57
2q37_p1	Chr2.233,284,112 (cg06644428)	19.0	15.9	18.0	0.67	0.77	0.68 (0.56–0.85)	0.72
2q37_p3	Chr2.233,284,661 (cg21566642)	59.6	53.9	43.5	0.61	0.64	0.66 (0.46–0.81)	0.58
6p21.33	Chr6:30,720,080 (cg06126421)	73.0	66.1	58.5	0.63	0.65	0.63 (0.45–0.85)	0.52
Cotinine (ng/mL); mean		5.0	3.8	980.7	0.04	0.90	0.47 (0.32–0.63)	0.25
MI					0.69	0.90	0.82 (0.64–0.99)	0.92
Validation set (n = 102 never smokers, 45 former smokers, 33 current smokers)								
MI					0.71	0.80	0.83 (0.70–0.96)	0.85

AUC values are shown for the individual markers to distinguish former smokers from never smokers in the test set.

© 2013 Lippincott Williams & Wilkins

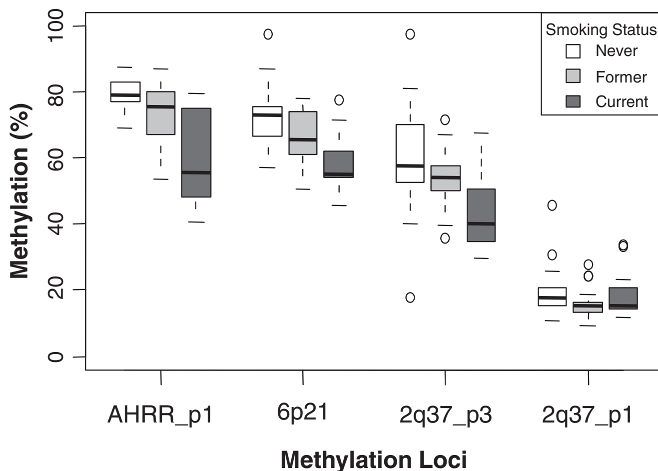

FIGURE 1. Box plots of distribution for methylation at four genomic loci for never, former, and current smokers.

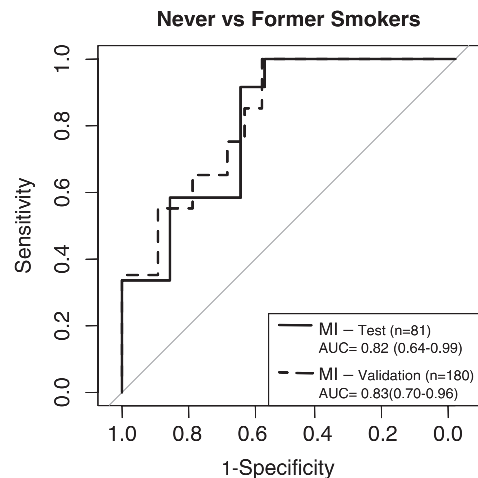

FIGURE 2. ROC for prediction of former smoking status based on MI.

smokers compared with never smokers using ROC analysis (AUC = 0.97) (Table 1). Table 1 indicates the AUC values based on methylation values for each genomic locus in predicting former smoker status for the test set, in addition to that for cotinine: the AUC for AHRR_p1 (AUC = 0.71), 6p21 (AUC = 0.63), 2q37_p1 (AUC = 0.68), and 2q37_p3 (AUC = 0.66) individually had a greater ability to distinguish former from nonsmokers than cotinine levels (AUC = 0.47). We combined the methylation values of all four loci into a single MI using the MI model = ($\beta_1 M_1 \times \beta_2 M_2 \times \beta_3 M_3 \times \beta_4 M_4$), where β represents the β-coefficient for the methylation locus in association with smoking status and M represents the methylation level of each locus as a percentage (equivalent to raw β values from 450K methylation array data). For the purpose of this analysis, we have not transformed methylation values to M values as this did not alter the performance of the model.

The AUC value for the combined MI of these four loci in differentiating never from former smokers was 0.82 (95% confidence interval [CI] = 0.64–0.99) in the test set and 0.83 (95% CI = 0.70–0.96) in the validation set (Figure 2) with a sensitivity of 69% and 71%, respectively. Using previously published 450K methylation array data,[14] we show the MI defined in the present study correlates strongly with duration of smoking in former smokers (Pearson's correlation R = 0.47, Figure 3). Furthermore, using the current pyrosequencing data in the validation set, the MI also correlated strongly with the time since quitting in former smokers (Pearson's correlation R = −0.51, Figure 4), as well as duration of smoking (R = 0.63, data not shown).

TABLE 2. Cotinine Levels (ng/mL) in the Test Sample Set (n = 81)

	N	Median	Minimum	Maximum	IQR
Never smoker	32	2.03	0.83	45.15	1.70–3.15
Former smoker	25	2.04	0.57	24.47	1.59–3.41
Current smoker	18	904.30	3.27	3181.89	583–1,156

IQR, interquartile range.

DISCUSSION

Any biomarker of tobacco exposure should reflect the degree of exposure, including the intensity and duration of smoking.[5] Our study gives strong evidence for long-standing methylation changes as a result of smoking. Methylation is a relatively stable DNA modification.[20] However, methylation changes may be reversible after the cessation of an exposure, although in the case of tobacco the timing of reversion is unknown.

We and others have found, using the Illumina 450K methylation beadchip, that a particular region of DNA methylation in the *AHRR* gene is strongly associated with smoking

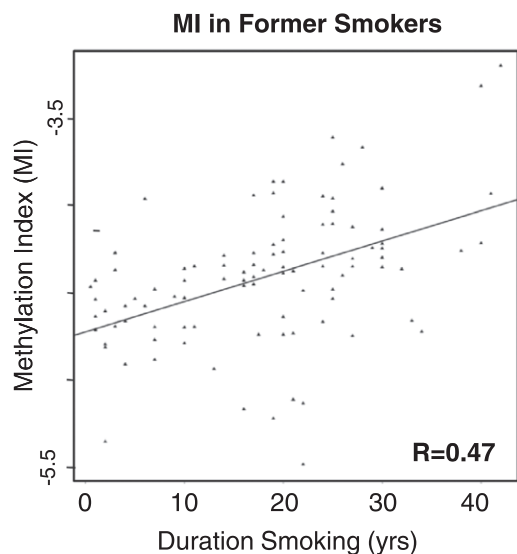

FIGURE 3. MI compared with duration of smoking in former smokers.

© 2013 Lippincott Williams & Wilkins

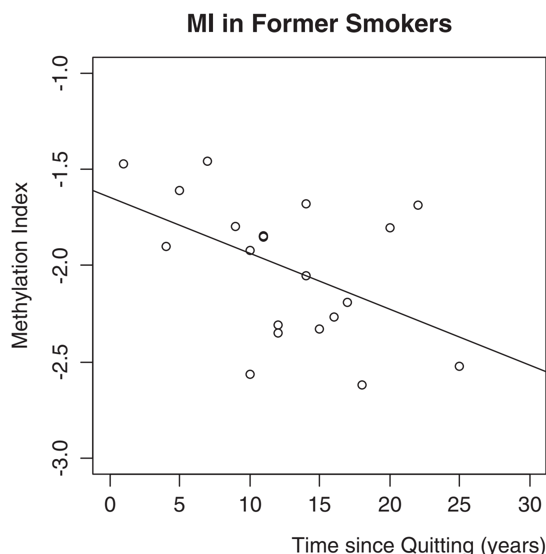

FIGURE 4. MI versus time since quitting in former smokers.

and can be a marker of past smoking exposure.[14,21–23] *AHRR* is part of the aryl hydrocarbon pathway that metabolizes cigarette smoke components, including the carcinogenic dioxins and dioxin-like compounds.[24,25] Monick and colleagues,[22] using the 450K array, found that a single probe in the *AHRR* gene was associated with smoking status (false discovery rate $P < 0.05$) in cultured lymphoblast cell lines; Philibert et al[23] also using the same 450K platform identified the same site in both men and women in peripheral blood lymphocyte DNA (false discovery rate $P < 0.05$). Additionally, in a much larger study (n = 1,062) on the 450K platform, Joubert et al extended this by identifying 26 CG sites associated with smoking in cord blood DNA (Bonferroni $P < 0.05$), which showed that maternal smoking could potentially affect methylation in the newborn baby.

In the majority of loci, we noted that smoking induces hypomethylation (loss of methylation). This has been observed in all studies to date. We validated this hypomethylation using an alternative method of bisulphite pyrosequencing for six loci.[14] Specifically regarding one of the top hits in the *AHRR* gene (eg, cg05575921), Monick et al,[22] Joubert et al,[21] Philibert et al,[23] and our own study have all shown a decrease in methylation, which suggests that the direction of association is consistent. A notable exception is CYP1A1, for which increased methylation in relation to maternal smoking in pregnancy was identified and replicated in an independent population in a study by Joubert et al.[21] The contrasting effects of maternal smoking during pregnancy on methylation at CGs in *AHRR* and CYP1A1 are of interest because of the opposing function of these genes in the aryl-hydrocarbon receptor pathway.[26] Numerous other loci have been associated with methylation changes in smokers, including two that we previously identified (the intergenic loci 2q37 and 6p21); the functions of these are currently unknown.

One of the limitations of this study is that it has been conducted in a single population enrolled in the EPIC cohort in Turin. Further validation in other populations is needed. In particular, investigators who have serum cotinine measurements along with smoking history and 450K methylation data can perform a similar analysis to generate a biomarker of past smoking specific for the 450K platform. Another limitation in this analysis is that there is a sex imbalance in the test and validation sample sets. An analysis of larger cohorts may identify sex-specific differences. Also, we had few persons who were self-described as former smokers with high cotinine levels (>15 ng/mL, n = 1) or current smokers with low cotinine levels (<15 ng/mL, n = 2). It would be useful to assess the performance of the MI versus self-reported smoking status in a larger study.

We have shown that the epigenetic changes associated with smoking are detected in blood DNA in former smokers many years after they have quit smoking[14] (median 13 years [interquartile range 9–18] in the present study, Figures 3 and 4). Given that the majority of white blood cell types have lifespans of ~30 days, this suggests that the exposure must also be affecting the hematopoietic stem and progenitor cells, which perpetuate the epigenetic alterations in the daughter differentiated cells. Further evidence of long-term perpetuation of methylation associated with exposures was observed in the Dutch famine study, with insulin-like growth factor 2 hypomethylation in exposed persons detected 60 years later.[27] If exposures such as smoking can increase a person's risk of cancer, even in former smokers, then we hypothesize that those exposures throughout life must also affect the tissue-specific stem and progenitor cells and potentially the cell of origin for the initial carcinogenic events. Although we have measured this biomarker in blood DNA, which proves a suitable DNA source, other sources of DNA (such as buccal swabs) may be equally useful for measuring this long-term biomarker of cigarette smoke exposure. Further investigation into DNA methylation of other cell types affected by tobacco smoking is warranted.

In sum, we have determined a set of differentially methylated genomic loci dependent on tobacco exposure that can predict former smoking status with high positive predictive and sensitivity values when combined into a single DNA MI. This provides a direct molecular measure of prior exposure to tobacco that can be performed using pyrosequencing. These data suggest that epigenetic patterns detected in blood may provide molecular biomarkers of other exposures that are also difficult to quantify in epidemiological studies.

REFERENCES

1. 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;88:450–457.
2. Sasco AJ, Secretan MB, Straif K. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung Cancer*. 2004;45(suppl 2):S3–S9.

© 2013 Lippincott Williams & Wilkins

3. Benowitz NL. Clinical pharmacology of nicotine: implications for understanding, preventing, and treating tobacco addiction. *Clin Pharmacol Ther*. 2008;83:531–541.
4. Hannan LM, Jacobs EJ, Thun MJ. The association between cigarette smoking and risk of colorectal cancer in a large prospective cohort from the United States. *Cancer Epidemiol Biomarkers Prev*. 2009;18:3362–3367.
5. Florescu A, Ferrence R, Einarson T, Selby P, Soldin O, Koren G. Methods for quantification of exposure to cigarette smoking and environmental tobacco smoke: focus on developmental toxicology. *Ther Drug Monit*. 2009;31:14–30.
6. Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008;358:1148–1159.
7. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004;4:143–153.
8. Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. *Trends Genet*. 2004;20:350–358.
9. Rakyan VK, Down TA, Maslau S, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. 2010;20:434–439.
10. Li S, Hursting SD, Davis BJ, McLachlan JA, Barrett JC. Environmental exposure, DNA methylation, and gene regulation: lessons from diethylstilbesterol-induced cancers. *Ann N Y Acad Sci*. 2003;983:161–169.
11. Ma S, Liu X, Jiao B, Yang Y, Liu X. Low-dose radiation-induced responses: focusing on epigenetic regulation. *Int J Radiat Biol*. 2010;86:517–528.
12. Christensen BC, Kelsey KT, Zheng S, et al. Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet*. 2010;6:e1001043.
13. Vineis P, Chuang SC, Vaissière T, et al.; Genair-EPIC Collaborators. DNA methylation changes associated with cancer risk factors and blood levels of vitamin metabolites in a prospective study. *Epigenetics*. 2011;6:195–201.
14. Shenker NS, Polidoro S, van Veldhoven K, 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;22:843–851.
15. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol*. 1997;26(suppl 1):S6–S14.
16. Riboli E. The European Prospective Investigation into Cancer and Nutrition (EPIC): plans and progress. *J Nutr*. 2001;131:170S–175S.
17. Timofeeva MN, McKay JD, Smith GD, et al. Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. *Cancer Epidemiol Biomarkers Prev*. 2011;20:2250–2261.
18. Brennan K, Garcia-Closas M, Orr N, et al.; KConFab Investigators. Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res*. 2012;72:2304–2313.
19. Soo-Quee Koh D, Choon-Huat Koh G. The use of salivary biomarkers in occupational and environmental medicine. *Occup Environ Med*. 2007;64:202–210.
20. Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Mol Biotechnol*. 2010;44:71–81.
21. Joubert BR, Håberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2012;120:1425–1431.
22. Monick MM, Beach SR, Plume J, et al. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet*. 2012;159B:141–151.
23. Philibert RA, Beach SR, Brody GH. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics*. 2012;7:1331–1338.
24. Evans BR, Karchner SI, Allan LL, et al. Repression of aryl hydrocarbon receptor (AHR) signaling by AHR repressor: role of DNA binding and competition for AHR nuclear translocator. *Mol Pharmacol*. 2008;73:387–398.
25. Chiba T, Uchi H, Yasukawa F, Furue M. Role of the arylhydrocarbon receptor in lung disease. *Int Arch Allergy Immunol*. 2011;155(suppl 1):129–134.
26. Kawajiri K, Fujii-Kuriyama Y. Cytochrome P450 gene regulation and physiological functions mediated by the aryl hydrocarbon receptor. *Arch Biochem Biophys*. 2007;464:207–212.
27. Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*. 2008;105:17046–17049.

© 2013 Lippincott Williams & Wilkins