

RESEARCH

Open Access

# Evolutionary age of repetitive element subfamilies and sensitivity of DNA methylation to airborne pollutants

Hyang-Min Byun<sup>1\*†</sup>, Valeria Motta<sup>1,2†</sup>, Tommaso Panni<sup>3</sup>, Pier Alberto Bertazzi<sup>2</sup>, Pietro Apostoli<sup>4</sup>, Lifang Hou<sup>5</sup> and Andrea A Baccarelli<sup>1</sup>

## Abstract

**Background:** Repetitive elements take up >40% of the human genome and can change distribution through transposition, thus generating subfamilies. Repetitive element DNA methylation has associated with several diseases and environmental exposures, including exposure to airborne pollutants. No systematic analysis has yet been conducted to examine the effects of exposures across different repetitive element subfamilies. The purpose of the study is to evaluate sensitivity of DNA methylation in differentially-evolved LINE, *Alu*, and HERV subfamilies to different types of airborne pollutants.

**Methods:** We sampled a total of 120 male participants from three studies (20 high-, 20 low-exposure in each study) of steel workers exposed to metal-rich particulate matter (measured as PM<sub>10</sub>) (Study 1); gas-station attendants exposed to air benzene (Study 2); and truck drivers exposed to traffic-derived elemental carbon (Study 3). We measured methylation by bisulfite-PCR-pyrosequencing in 10 differentially-evolved repetitive element subfamilies.

**Results:** High-exposure groups exhibited subfamily-specific methylation differences compared to low-exposure groups: L1PA2 showed lower DNA methylation in steel workers ( $P=0.04$ ) and gas station attendants ( $P=0.03$ ); L1Ta showed lower DNA methylation in steel workers ( $P=0.02$ ); *AluYb8* showed higher DNA methylation in truck drivers ( $P=0.05$ ). Within each study, dose-response analyses showed subfamily-specific correlations of methylation with exposure levels. Interaction models showed that the effects of the exposures on DNA methylation were dependent on the subfamily evolutionary age, with stronger effects on older LINEs from PM<sub>10</sub> (p-interaction=0.003) and benzene (p-interaction=0.04), and on younger *Alus* from PM<sub>10</sub> (p-interaction=0.02).

**Conclusions:** The evolutionary age of repetitive element subfamilies determines differential susceptibility of DNA methylation to airborne pollutants.

**Keywords:** Environment, Exposures, DNA methylation, Repetitive elements, Subfamily

## Background

Approximately half of the human genome is made of repetitive elements [1]. Repetitive elements, particularly L1s, are DNA sequences that only encode for proteins instrumental to their replication and insertion into new locations within the genome. Also often referred to as

“junk DNA” [2] or “jumping genes” [3], repetitive elements have been active in mammalian genomes for over 100 million years and are therefore a source of evolution and genome structure organization. Repetitive elements are believed to have reshaped the human genome through continuous jumping to remote genomic locations and through the activation of alternative transcription of nearby genes [4-7]. *De novo* retrotransposition insertions occurring in the germ line are inherited through generations. However, many newly inserted repetitive elements have been observed in

\* Correspondence: hmbyun@hsph.harvard.edu

†Equal contributors

<sup>1</sup>Laboratory of Environmental Epigenetics, Exposure Epidemiology and Risk Program, Harvard School of Public Health, Boston, MA, USA  
Full list of author information is available at the end of the article

somatic tissues and, albeit non heritable, have been linked with human disease and cancers [8,9].

Among repetitive elements, the best studied families are the long interspersed nuclear elements (LINE-1 or L1), short interspersed nuclear element family (SINEs), and human endogenous retrovirus (HERV) family. *Alu* and long-terminal repeats (LTRs) are the most abundant representatives of SINEs and HERVs, respectively. All these families are classified as retroelements because of their ability to retrotranspose, i.e., copy themselves into an RNA intermediate and insert back into the genome as a new cDNA copy [10]. Among repetitive elements, only LINE-1 and *Alu* have been unequivocally shown to be still active and retrotransposition-competent in the current human genome [11]. Since the mobilization of repetitive elements has been linked to genomic instability and consequent genetic disorders, mechanisms such as DNA methylation are believed to have developed in cells to control the proliferation of retrotransposons. CpG-dinucleotide DNA methylation in repetitive elements has been proposed to contribute to repress retrotransposition activity as part of a self-defense system of the host genome [12,13]. On the other hand, decreased DNA methylation in repetitive elements has been linked with increased transcription and higher rates of retrotransposon activity *in vitro* [14,15]. In human studies, differences in DNA methylation of LINE-1, *Alu*, and HERV have been consistently demonstrated in response to stress [16] and infections [17], as well as in autoimmune diseases [18], cardiovascular diseases [19] and cancers [20,21].

Through their “copy and paste” retrotransposition activity, each repetitive element family has gradually accumulated new base substitutions, insertions or deletions referred to as “diagnostic mutations” [22], which allow for distinguishing various subfamilies within each repetitive element family. Through the analysis of these mutations and the divergence of the repetitive element consensus sequences from the original ancestral sequence, it is possible to estimate an “evolutionary age” for each subfamily [1,23]. The mutation rate depends on the CpG density; since CpG di-nucleotides are more sensitive than non-CpG sites, they tend to be eliminated through evolution by substitution to either TpG or CpA dinucleotides. These substitutions are frequent events, as the mutation rate of CpG sites is 9.2 fold faster than non-CpG changes [24,25]. In the mammalian genome, CpG dinucleotides are found to be highly represented in repetitive elements [26]. Through many cycles of substitutions, old subfamilies remain less rich in CpGs and show weakened or no retrotransposon activity, whereas young subfamilies are richer in CpGs and still transcriptionally active in the human genome [25,27].

Recent investigations have repeatedly linked DNA methylation of repetitive elements with environmental exposures. In human studies, methylation of repetitive elements in blood DNA has been correlated with a number of exposures that generate oxidative stress and inflammation [28,29], including airborne pollutants [21,30-32], metals [33-35], and persistent organo-pollutants [36,37]. Several investigators have proposed that these effects may contribute to generate – or at least reflect – the altered retrotransposon methylation states related with aging, autoimmune diseases (e.g., multiple sclerosis, lupus), cardiovascular disease, or cancer development and progression [38]. DNA methylation of repetitive elements has been also widely proposed as an indicator of global genomic methylation level [39]. However, current investigations have only focused on the analysis of a single common sequence for each family, which is not expected to comprehensively reflect DNA methylation patterns within specific subfamilies. Recent reports have shown that methylation of an individual common sequence is not correlated with the global content of methylation in normal tissues [7]. A recent observational study showed that CpG content is a primary predictor of changes over time in DNA methylation at individual CpG sites [40]. Because repetitive element CpG content and DNA methylation levels vary dramatically across subfamilies with different evolutionary age, repetitive element subfamilies may also have differential sensitivity to environmental exposures. However, no systematic analysis has yet been conducted to examine the effects of environmental exposures across different subfamilies of repetitive elements.

In the present work, we conducted a comprehensive investigation of DNA methylation in repetitive element subfamilies of different evolutionary ages. We examined multiple groups of participants exposed to different types of airborne pollutants (Table 1). We sampled groups with high- and low-exposure from three independent studies including steel workers and low-exposed controls to metal-rich particulate matter (Study 1); gas station attendants and low-exposed controls to traffic-derived benzene (Study 2); and truck drivers and low-exposed controls to traffic-derived Elemental Carbon (EC) in Beijing, China. We examined DNA methylation of four LINE-1 subfamilies (L1PA5, L1PA2, L1Hs, and L1Ta) [41], three *Alu* subfamilies (*AluSx*, *AluYb8*, and *AluYd6*) [42], and three HERV subfamilies (MLT1D, ERV1, and ERV9) [43] selected within each family to represent different evolutionary ages ranging from old to young elements. We showed that the effects of the exposures on DNA methylation are dependent on the subfamily evolutionary age. To the best of our knowledge, this is the first report showing that the evolutionary age

**Table 1 Characteristics and exposure levels of the study participants**

			Highly-exposed steel workers (n=20)	Low-exposed controls (n=20)
<b>Study 1<sup>a</sup></b> Exposure to metal-rich particulate matter (PM)	PM <sub>10</sub> [ $\mu\text{g}/\text{m}^3$ ]	Mean $\pm$ SD	203.7 $\pm$ 22.9	100.9 $\pm$ 28.9
		Range	[152.2 ; 227.9]	[73.7 ; 150.0]
	Participants' characteristics	Age [Years], mean $\pm$ SD	42.4 $\pm$ 7.9	37.8 $\pm$ 3.0
		Ex/current smokers, n (%)	12 (60)	14 (70)
<b>Study 2<sup>b</sup></b> Exposure to air benzene	Air benzene [ $\mu\text{g}/\text{m}^3$ ]	Mean $\pm$ SD	78.6 $\pm$ 42.5	7.0 $\pm$ 5.5
		Range	[31.2 ; 180.1]	[4.2 ; 23.0]
	Participants' characteristics	Age [Years], mean $\pm$ SD	39.9 $\pm$ 11.2	39.7 $\pm$ 10.4
		Ex/current smokers, n (%)	7 (35)	5 (25)
<b>Study 3<sup>c</sup></b> Exposure to traffic-derived elemental carbon	Elemental carbon [ $\mu\text{g}/\text{m}^3$ ]	Mean $\pm$ SD	21.3 $\pm$ 4.7	13.4 $\pm$ 2.1
		Range	[16.6 ; 35.6]	[7.8 ; 16.1]
	Participants' characteristics	Age [Years], mean $\pm$ SD	35.2 $\pm$ 5.1	33.4 $\pm$ 6.0
		Ex/current smokers, n (%)	8 (40)	6 (30)

The participants were recruited from different occupations and divided in high- and low-exposed group according to their personal levels of exposure.

<sup>a</sup> Steel workers with Particulate Matter with aerodynamic diameter  $\leq 10 \mu\text{m}$  [PM<sub>10</sub>] >152.2  $\mu\text{g}/\text{m}^3$  in the highly-exposed group; steel workers with PM<sub>10</sub><150.0  $\mu\text{g}/\text{m}^3$  in the low-exposed control group.

<sup>b</sup> Gas station attendants in the highly-exposed group; indoor office workers in the low-exposed control group.

<sup>c</sup> Truck drivers in the highly-exposed group; indoor office workers in the low-exposed control group.

of repetitive elements determines their vulnerability to environmental exposures.

## Results

### Repetitive element evolutionary age, CpG density and methylation levels

As a measure of CpG density we calculated the ratio between observed and expected CpG content (CpG<sub>o/e</sub>) for each subfamily (Table 2). The subfamily evolutionary age and CpG density showed opposite directions, as younger subfamilies showed higher CpG density. The subfamily evolutionary age and DNA methylation levels – estimated among all the low-exposed control groups from Study 1, Study 2, and Study 3 – also showed opposite directions, as younger subfamilies tended to have higher mean methylation, except for HERV (Table 2).

### DNA methylation of repetitive element subfamilies by exposure group

We first examined DNA methylation in repetitive element subfamilies by contrasting the high- vs. low-exposure groups in analysis adjusted for age and smoking. Among LINE-1 subfamilies, L1PA2 showed significantly lower DNA methylation both in highly PM<sub>10</sub>-exposed steel workers in Study 1 (mean differences=-1.2%,  $P=0.04$ ), and in gas station attendants in Study 2 (mean differences=-1.3%,  $P=0.03$ ) relative to the respective low-exposed controls (Figure 1B and see Additional file 1: Table S1). L1Ta,

the youngest of the LINE-1 showed in Study 1 significantly lower DNA methylation in steel workers with high exposure to metal-rich PM<sub>10</sub> compared to the low-exposed controls (mean differences=-1.5%,  $P=0.02$ ) (Figure 1D and see

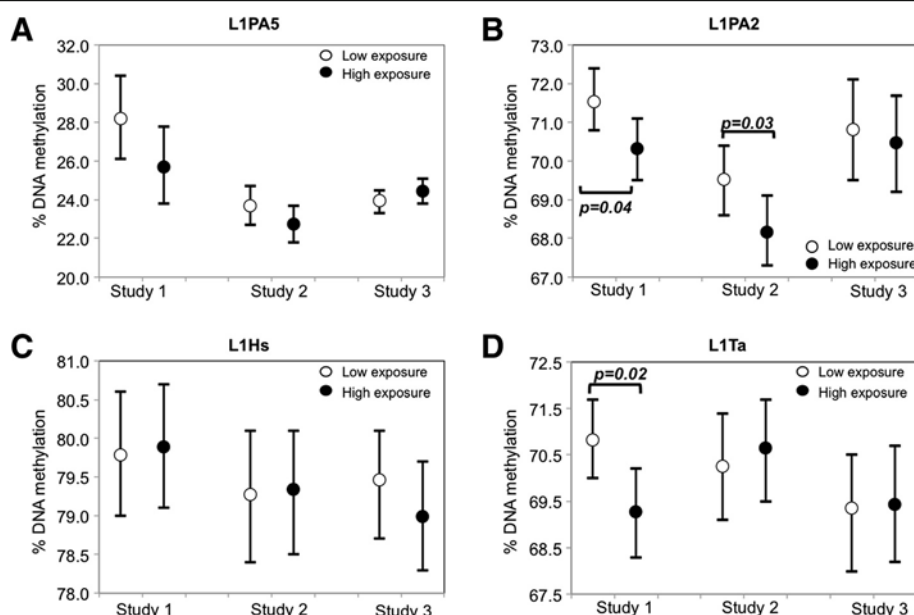
**Table 2 Characteristics of the repetitive element subfamilies examined in the present study**

Family	Subfamily	Evolutionary age (Mya) <sup>a</sup>	CpG density (CpG <sub>o/e</sub> ) <sup>b</sup>	Mean methylation (%) <sup>c</sup>
LINE-1	L1PA5	20.4	0.073	25.3
	L1PA2	7.6	0.104	70.6
	L1Hs	5	0.155	79.5
	L1Ta	1.9	0.282	70.1
Alu	AluSx	40	0.878	24.7
	AluYb8	2.9	0.884	89.9
	AluYd6	2	0.95	89.9
HERV	MLT1D	98.2	0.043	97.9
	ERV1	24.4	0.074	24.5
	ERV9	15	0.439	52

<sup>a</sup> Estimated time in 'Million years ago' (Mya) based on the time when the repetitive element subfamily appeared in the human genome.

<sup>b</sup> Approximate ratio of the observed to the expected CpG content (CpG<sub>o/e</sub>) of the sequence. CpG observed/expected (CpG<sub>o/e</sub>). The ratio is calculated using the formula ((Num of CpG)/(Num of C  $\times$  Num of G))  $\times$  Total number of nucleotides in the sequence).

<sup>c</sup> Mean methylation of all low-exposure control groups from Study 1, Study 2, and Study 3.



**Figure 1** DNA methylation differences in LINE-1 subfamilies between low and high exposure groups. Mean DNA methylation levels and 95% confidence intervals of L1PA5 (panel A), L1PA2 (panel B), L1Hs (panel C), and L1Ta (panel D) in low and high exposure groups are shown for each of the studies. Open circles represent low exposure group (Steelworkers in low exposure job position in Study 1; indoor office workers in Study 2 and Study 3); closed circles represent high exposure group (Steel workers in high exposure job position in Study 1; gas station attendants in Study 2; and truck drivers in Study 3). Significant p-values ( $< 0.05$ ) for DNA methylation differences between low and high exposure groups are shown in the figures.

Additional file 1: Table S1). Neither L1PA5 – the oldest LINE-1 subfamily in this study – nor L1Hs, a relatively young subfamily, showed significant DNA methylation differences between low and high exposure groups in any of the three studies (Figure 1A and 1C; and see Additional file 1: Table S1). *AluYb8* – a relatively younger *Alu* – showed in Study 3 significantly higher DNA methylation in truck drivers with high EC exposure compared to indoor office workers (mean difference=0.4%,  $P=0.039$ ) (Figure 2B and see Additional file 1: Table S1). Neither *AluSx* – the oldest *Alu* subfamily in this study, nor *AluYd6* – the youngest *Alu* subfamily in this study – showed significant DNA methylation differences between high and low-exposure groups in any of the three studies (Figure 2A and 2C; and see Additional file 1: Table S1). None of the HERV subfamilies showed significant DNA methylation differences between high and low-exposure groups in any of the three studies (see Additional file 1: Table S1).

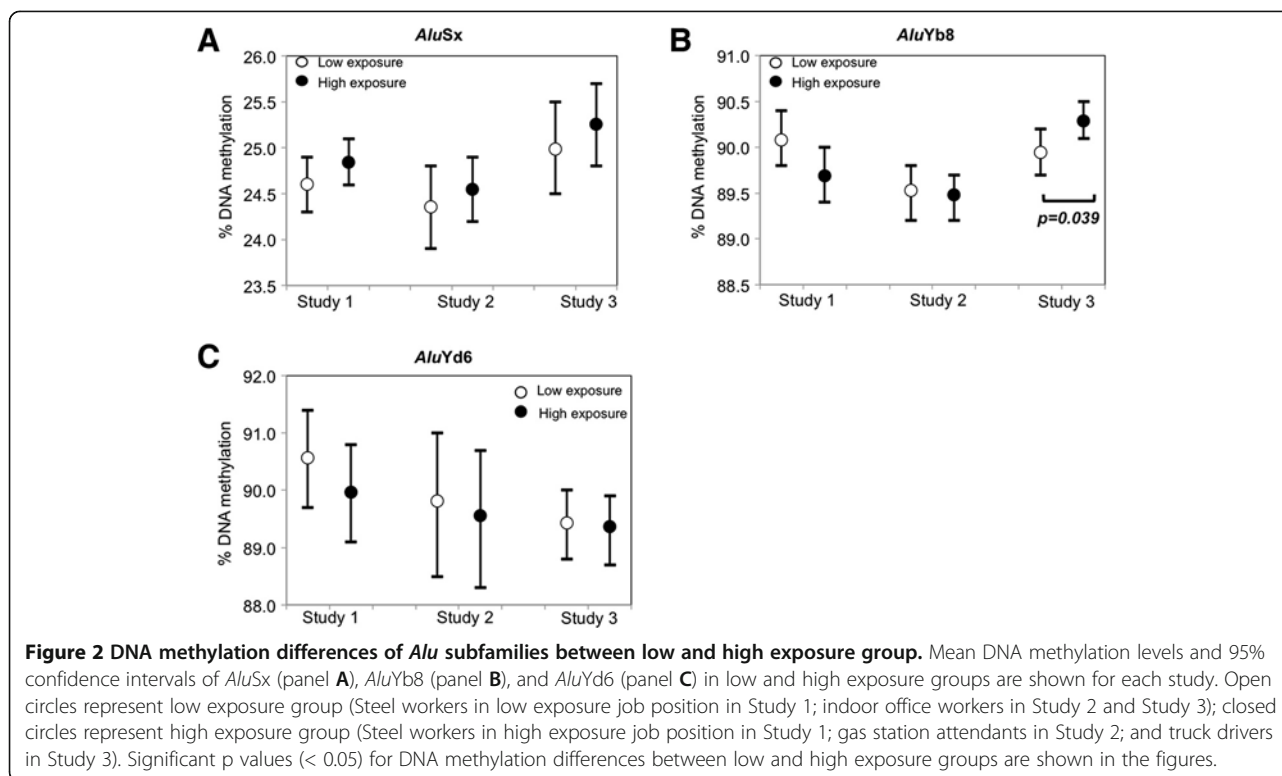
#### Dose–response relationship in the correlation of DNA methylation in repetitive element subfamilies with exposure levels

To evaluate the dose–response relationships between exposure levels and DNA methylation, we fitted regression models estimating the relative change – expressed as percent – of the ratio of methylated/unmethylated DNA

associated with an increase in the exposure from the 25th to the 75th percentile, adjusted for age and smoking (Table 3). Among LINE-1 subfamilies, all the significant correlations of exposure levels with DNA methylation across studies were observed in the two oldest subfamilies, i.e., L1PA5 and L1PA2. In Study 1, DNA methylation of L1PA5 and L1PA2 was negatively correlated with the levels of exposure to metal-rich  $PM_{10}$  ( $\tau=-15.2$ ,  $P=0.02$  and  $\tau=-5.8$ ,  $P=0.03$ , respectively). In Study 2, L1PA2 methylation showed a negative correlation with air benzene exposure ( $\tau=-4.3$ ,  $P=0.01$ ). In Study 3, L1PA5 methylation showed a negative correlation with EC exposure ( $\tau=5.6$ ,  $P=0.01$ ). In all the three studies, there were no or marginal correlations of methylation of *Alus* and HERVs with the levels of exposure (Table 3).

#### Evolutionary age of subfamilies and sensitivity of DNA methylation to airborne pollutants

To determine whether repetitive element sensitivity to environmental exposures was dependent on evolutionary age, we used an interaction analysis that modeled the correlation between environmental exposure levels and DNA methylation as a function of the evolutionary age. Figure 3 shows a graphical representation of the results for the combinations of repetitive element families and exposures that showed significant interactions with arbitrary numbers at regular intervals of 2, 6, 11, 15, and 20



million years ago (Mya), reflecting a plausible range and spaced intervals of evolutionary age, as well as for those that showed significant correlations of methylation of any of the subfamilies with the exposure levels in the section above. Complete results are shown in Additional file 1: Table S2.

In Study 1, the negative effect of PM<sub>10</sub> exposure on LINE-1 methylation was stronger with increasing

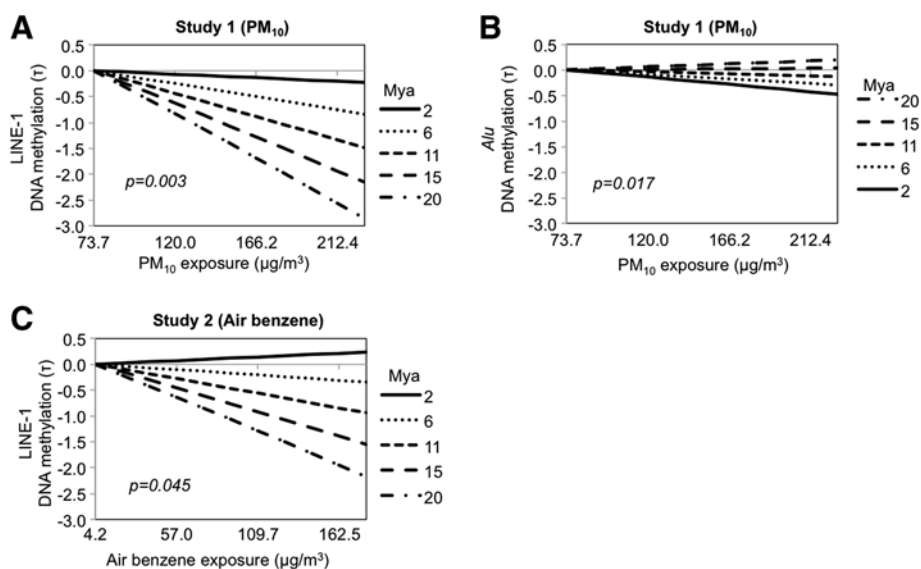
evolutionary age ( $\tau$  for interaction = -0.6;  $P=0.003$ ) (Figure 3A). Also in Study 1, the interaction model showed a positive significant interaction between PM<sub>10</sub> exposure and evolutionary age in predicting *Alu* family methylation ( $\tau=0.2$  and  $P=0.017$ ) (Figure 3B). However, as shown in the slopes in Figure 3, the differences of the effects of PM<sub>10</sub> in Study 1 across different evolutionary ages were more pronounced for LINE-1 methylation (Figure 3A)

**Table 3 Dose-response relationship between levels of personal air pollutants exposure and DNA methylation in the participants**

Repetitive element		Study 1 associations with metal-rich particulate matter (PM <sub>10</sub> )		Study 2 associations with air benzene		Study 3 associations with traffic-derived elemental carbon	
Family	Subfamily	$\tau^a$	p value	$\tau^a$	p value	$\tau^a$	p value
LINE-1	L1PA5	-15.2	0.02*	-2.3	0.28	5.6	0.01*
	L1PA2	-5.8	0.03*	-4.3	0.01*	-2.3	0.51
	L1Hs	1.4	0.68	1.0	0.60	-2.7	0.19
	L1Ta	-4.2	0.17	0.7	0.74	1.6	0.68
<i>Alu</i>	<i>AluSx</i>	0.0	1.00	1.0	0.21	0.0	1.00
	<i>AluYb8</i>	-3.8	0.11	-0.4	0.74	2.4	0.14
	<i>AluYd6</i>	-6.6	0.31	-4.1	0.37	4.0	0.29
HERV	MLT1D	-47.6	0.16	54.6	0.13	-41.8	0.09
	ERV1	0.4	0.78	0.6	0.56	0.0	0.99
	ERV9	1.9	0.45	1.4	0.42	0.5	0.83

<sup>a</sup>  $\tau = (2^{\tau} - 1) * 100$  represents the percent-change of the ratio methylated/unmethylated associated with an increase in the exposure from the 25th to the 75th percentile, adjusted for age and smoking.  
 \* p < 0.05.





**Figure 3 Interaction of evolutionary ages (Mya) with air pollutant exposures in determining repetitive element DNA methylation.**

Effects on DNA methylation estimated by modeling interactions of metal-rich  $PM_{10}$  exposure with ages of LINE-1 subfamilies in Study 1 (panel **A**); interactions of metal-rich  $PM_{10}$  exposure with ages of *Alu* subfamilies in Study 1 (panel **B**); interactions of airborne benzene exposure with ages of LINE-1 subfamilies in Study 2 (panel **C**). The differences in DNA methylation in each subfamily are represented using  $\tau$ . The evolutionary ages shown are arbitrary numbers at regular intervals of 2, 6, 11, 15, and 20 million years ago (Mya), reflecting a plausible range and spaced intervals of evolutionary ages. The interaction  $p$ -values are shown in the figures.

compared to *Alu* methylation (Figure 3B). In Study 2, the negative effect of airborne benzene in LINE-1 DNA methylation was also progressively stronger with increasing evolutionary age ( $\tau$  for interaction =  $-0.2$ ;  $P=0.045$ ) (Figure 3C). No interactions were found between evolutionary ages and exposure levels in determining HERV methylation (see Additional file 1: Table S2).

## Discussion

DNA methylation of repetitive elements has been extensively studied in relation to environmental exposures and human disease. Nonetheless, most if not all of the previous studies have investigated one single sequence in one or at most two repetitive element subfamilies. The present work is based on a comprehensive methylation analysis of 10 repetitive element subfamilies that were examined in three groups of participants with well-characterized exposure, including steel workers with exposure to  $PM_{10}$  in Study 1; gas station attendants exposed to air benzene in Study 2; and truck drivers exposed to EC in Study 3. We found that effects on DNA methylation of individual repetitive element subfamilies were specific to the exposure type. We selected different ages of repetitive element subfamilies – including old, intermediate, and young subfamilies – based on the time when they appeared in the human genome. We showed exposure-related effects that depended on the subfamily evolutionary age.

The different susceptibility of repetitive element methylation to environmental pollutants may be explained by the sequence variation and GC-content differences between the subfamilies. Subfamilies with older evolutionary age have lower CpG content due to higher substitution rates. As a measure of CpG density, we calculated the ratio between observed and expected CpG content ( $CpG_{o/e}$ ) for each subfamily (Table 2). Our data confirmed that the subfamily age was inversely correlated with DNA methylation levels in the CpG sites. Also, DNA methylation of those CpG sites was positively correlated with the ratio of  $CpG_{o/e}$ . These findings show that older subfamilies have lower CpG density and are prone to have lower DNA methylation. This also supports the concept that each repetitive element family has different patterns of DNA methylation, which might reflect varying degrees of regulation and help explaining the different responses to environmental exposures.

An alternative potential explanation for the exposure-related differences in DNA methylation observed in the present study relates to the genomic position of repetitive elements in the genome. Repetitive element families show different insertional preference in the human genome; for instance, LINE-1s are frequently inserted in AT rich regions, as TTTT/A is the site to prime reverse transcription [44]. *Alus* and HERVs are more likely inserted into GC rich regions, i.e., in regions near genes or gene-related features such as CpG islands [45]. In some instances, methylation spreads from upstream *Alus*

into the nearby gene-promoter region [46]. Due to the functional relationships between repetitive elements and their surrounding regions, it is possible that differential sensitivity of the insertion regions to environmental exposures could affect DNA methylation of the inserted repetitive element. The methylation difference in this pool of repetitive elements is small (less than 2% in low versus high exposed groups) when expressed in percentage over the total number of cytosines in the considered position. Since we designed subfamily-specific assays, even a small difference in methylation levels might lead to instability of the genomic structure.

Methylation of individual sequences in the repetitive element families LINE-1, *Alu*, and HERV has been already investigated in relation to environmental exposures such as PM<sub>10</sub>, black carbon, and persistent organic pollutants [21,30-37,44]. However, due to the similarity of the sequences and the difficulty in designing primers for specific subfamilies, most of previous studies have analyzed only one single subfamily (i.e., L1Hs for the LINE-1 family and *AluSx* for the *Alu* family) and – to the best of our knowledge – no study has yet investigated multiple subfamilies. In the present study, we found significant associations of DNA methylation of specific repetitive element subfamilies in both the analysis using exposure groups (high vs. low) and in the dose–response analysis using continuous levels of exposure to metal-rich PM<sub>10</sub>, air benzene, or EC. However, not all the effects on DNA methylation were consistently found in both the group and dose–response analysis. For instance, L1Ta showed a significant difference in highly-exposed steel workers in Study 1, but the group analysis was not confirmed in the dose–response analysis using continuous PM<sub>10</sub> levels. Similarly, in the *Alu* family, DNA methylation of the intermediate-age *AluYb8* sequence showed a significant difference in the highly-exposed group of truck drivers in Study 3, which was not confirmed in the dose–response analysis using continuous EC levels. The small number of individuals in each group might explain at least part of these inconsistencies. It is worth noting that, even in the cases with discordance of statistical significance, group and dose–response analysis were concordant in showing similar directions for the exposure-related methylation differences.

Repetitive element subfamilies were inserted in the host human genome at different evolutionary ages. To provide more stable estimates of the general effects of air pollutants, as well as to elucidate the biological bases of the heterogeneity of effects within repetitive element family, we investigated whether the correlations between subfamily methylation and exposures depended on the subfamily evolutionary age. We applied arbitrary numbers at regular intervals of 2, 6, 11, 15, and 20 million years ago (Mya), reflecting a plausible range and spaced

intervals of evolutionary ages. We observed that the effects of air pollutants on repetitive element methylation – particular in LINE-1 subfamilies – were significantly affected by the age of subfamilies. The interaction analysis of environmental exposure and ages of repetitive element subfamilies suggested that DNA methylation in older LINE-1 subfamilies might be more vulnerable to environmental exposure than in younger subfamilies.

Our findings are consistent with the hypothesis that exposures activate pollutant-specific biological pathways, which may in turn result in signature differences in DNA methylation in specific repetitive element subfamilies. DNA methyltransferases (DNMTs) play a fundamental role in the generation of DNA methylation by transferring methyl groups from S-adenosyl-methionine (SAM) to the C5 position of the pyrimidine ring of cytosines. DNMTs are environmentally sensitive [9] and may represent vulnerable targets in the biological process linking pollutant exposures to DNA methylation. Specific DNMT isoforms show different sensitivity to the environment, as potentially each pollutant might target one or a combination the several DNMT isoforms. DNMT isoforms also have different activity in the methylation of individual repetitive element subfamilies [47]. For instance *in vitro* studies showed that LINE-1 sequences are preferentially methylated by the DNMT3B1, DNMT3B2, and DNMTΔ3B isoforms, which however do not produce any methylation on *AluYb8*. Taken together, these data indicate that different susceptibility of DNMTs to environmental exposures could modify their subfamily-specific activity on DNA methylation.

The present study has a number of limitations. The small sample size of 40 participants from each of the three studies might have limited the power to detect exposure-related differences. Despite designing PCR primers on highly homologous sequence regions between subfamilies, our assays might have missed some copies of each subfamily due to the sequence variations inherent to repetitive element subfamilies. In addition, because of the general characteristics of sodium-bisulfite conversion, we could not distinguish between CpG to TpG mutation and cytosine methylation in CpG sites. Separate genomic sequencing would be necessary to identify *bona-fide* cytosine methylation in CpG sites. The selection of repetitive elements in this study was limited to representative sequences with different evolutionary ages. Future studies are needed including larger numbers of subfamilies. Nonetheless, to the best of our knowledge, the present study includes the most comprehensive selection to date of subfamilies ever examined in relation with environmental exposure.

This study has also a number of strengths that support the validity of the results. We conducted a comprehensive DNA methylation analysis of repetitive element

subfamilies of different evolutionary ages within three distinct families. We used a highly quantitative bisulfite-PCR-pyrosequencing approach for DNA methylation analysis, which is the gold standard for DNA methylation analyses in short (up to 80–100 bp) sequences. Finally, we evaluated three different airborne pollutants, whose exposure was assessed through directly measured or estimated levels at the personal level. Significant differences in methylation level were found in repetitive element subfamilies that had not yet been analyzed in environmental studies. Our results suggest that previous studies that aimed at evaluating global genomic methylation by just considering specific subfamilies of L1HS and *AluSx* might have missed significant associations of repetitive element DNA methylation with environmental exposures. Using an augmented panel of repetitive element subfamilies might help to identify novel effects of environmental exposures.

## Conclusions

In conclusion, the present study on DNA methylation of 10 repetitive element subfamilies showed family- and subfamily-specific effects in three distinct conditions of exposures to different types of air pollutants. We also showed that sensitivity to environmental exposures is dependent on the evolutionary ages of the repetitive element subfamilies. Our results provide better understanding of the effects of the exposures on methylation of repetitive element subfamilies and may help to elucidate the role of repetitive elements in response to environmental risk factors related to human health and disease.

## Methods

### Study participants and exposure levels

We selected a total of 120 healthy individuals who participated in three previous studies of different types of airborne pollutants. From each of the three studies, we selected two groups of 20 highly-exposed individuals and 20 low-exposed controls matched by age and smoking status. In consideration of the predominance of males in the three original studies and to limit potential confounding by gender, we only sampled male participants. Table 1 shows the characteristics and the exposure levels of the three studies.

In Study 1, the participants were recruited from steel workers in Brescia, Italy [48]. In brief, we selected among workers in a steel production plant 20 steel workers with high exposure to metal-rich particles (Particulate Matter with aerodynamic diameter  $>10 \mu\text{m}$  [ $\text{PM}_{10}$ ]  $\geq 152.2 \mu\text{g}/\text{m}^3$ ) and 20 office workers as controls with low exposures ( $\text{PM}_{10} \leq 150.0 \mu\text{g}/\text{m}^3$ ).

In Study 2, we selected 20 gas station attendants exposed to airborne benzene in Milan, Italy as the highly-

exposed group (air benzene  $\geq 31.2 \mu\text{g}/\text{m}^3$ ); and 20 office workers as the low-exposed control group (air benzene  $\leq 23.0 \mu\text{g}/\text{m}^3$ ) [31].

In Study 3, we selected 20 truck drivers with exposure Elemental Carbon (EC), taken as a tracer of traffic particles, in Beijing, China as the highly-exposed group (EC  $\geq 16.6 \mu\text{g}/\text{m}^3$ ); and 20 office workers as the low-exposed control group (EC  $\leq 16.1 \mu\text{g}/\text{m}^3$ ) [49].

Exposure assessment is described in the Additional file 1: Table S3.

### Sample preparation

Sample preparation was conducted in all studies using the same standardized procedures. Buffy coat from whole blood was collected from each participant and immediately stored at  $-80^\circ\text{C}$  until genomic DNA was isolated. Genomic DNA from all participants within each study was isolated with the same batch of reagents in narrow time windows in order to minimize technical and operator variations. The isolated genomic DNA was stored at  $-80^\circ\text{C}$  for future use. Details on sample collection and DNA isolation for each study have been described in previous publications [30,31,49].

### DNA methylation analysis

One  $\mu\text{g}$  of genomic DNA was treated using the EZ DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's protocol. Final elution was performed with 30  $\mu\text{L}$  M-Elution Buffer. Bisulfite-treated DNA was aliquoted and stored at  $-80^\circ\text{C}$  until ready for use. We performed DNA methylation analyses on bisulfite-treated DNA using highly quantitative analysis based on PCR-pyrosequencing. The PCR and pyrosequencing primer sequences for L1HS, *AluSx*, and *AluYb8* have been previously published by Yang *et al.* [39] and by Choi *et al.* [50]. We developed additional assays specific for three LINE-1 subfamilies (L1PA5, L1PA2, and L1Ta), for *AluYd6* subfamily, and for three HERV subfamilies (MLT1D, ERV1, and ERV9) (see Additional file 1 for primer sequences and PCR conditions). A minimum number of 1 CpG and a maximum of 5 CpGs were evaluated in each assay. PCR amplification was performed at standard conditions using the GoTaq<sup>®</sup> Hot Start Polymerase (Promega, Madison, WI).

We used a PSQ Q96 MD pyrosequencing System (QIAGEN, Valencia, CA), as previously described [30]. Each sample was tested two times for each assay to confirm reproducibility. As a quality control check to estimate bisulfite conversion efficiency, we placed duplicate genomic DNA samples on each bisulfite conversion plate to estimate the internal plate variation of bisulfite conversion and the pyrosequencing reaction. We also added universal PCR products amplified from cell-line DNA on each pyrosequencing plate to check run-to-run



and plate-to-plate variation. In addition, the Pyrogram peak pattern from every sample was visually inspected to confirm the quality of the reaction.

### Characteristics of the repetitive element subfamilies

The repetitive element evolutionary age is defined based on the estimated time when each subfamily appeared in the human genome. We identified an evolutionary age (expressed in Mya [million years ago]) for each repetitive element subfamily based on information available in previous literature [41-43,51,52] (Table 2). As a measure of CpG density we calculated the ratio between observed and expected CpG content ( $CpG_{o/e}$ ) for each subfamily. The CpG observed/expected ( $CpG_{o/e}$ ) ratio is calculated by formula;  $(\text{Num of CpG}/(\text{Num of C} \times \text{Num of G})) \times \text{Total number of nucleotides in the sequence}$  [53] (Table 2).

### Statistical analysis

We used mixed-effect regression models to evaluate the effects of the exposures on DNA methylation levels of each subfamily, as previously reported [54]. This approach yields a global estimate of the effect on multiple CpGs within each subfamily sequence by modeling correlated data in adjacent CpG sites within each sequence, as well as the measures from duplicate pyrosequencing runs. Mixed-effect models have the advantage over standard methods of using the entirety of the information in the data, thus maximizing statistical power by distinguishing different sources of variance. See Additional file 1 for the detailed description of the models used.

### Additional file

**Additional file 1: Table S1.** DNA methylation in the three populations investigated by high and low exposure. **Table S2.** Interaction of evolutionary ages (Millions year ago, Mya) with airborne pollutant exposures in determining repetitive element DNA methylation. Selected results from the same analyses are reported in graphical form in Figure 3. **Table S3.** Primer sequences and PCR conditions.

### Abbreviations

$CpG_{o/e}$ : Observed and expected CpG content; DNMT: DNA methyltransferases; EC: Elemental carbon; HERV: Human endogenous retrovirus; LINE-1 or L1: Long interspersed nuclear elements; LTR: Long-terminal repeats; PCR: Polymerase chain reaction; PM: Particulate matter; PSQ: Pyrosequencing; SAM: S-adenosyl-methionine; SINE: Short interspersed nuclear element family.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

HMB, VM and AAB designed the experiment. TP analyzed the data. HMB and VM performed the experiments. PAB, PA, LF, and AAB contributed participant's materials. HMB, VM and AAB wrote the manuscript and AAB oversaw the research. All authors have read and approved the final manuscript.

### Acknowledgments

This work has been supported in part by the HSPH-NIEHS Center for environmental health (ES000002); and by related work in the NIEHS grants ES021733 and ES021895. The Beijing air pollution study has been supported by NIEHS grants ES020010 and ES020984.

### Author details

<sup>1</sup>Laboratory of Environmental Epigenetics, Exposure Epidemiology and Risk Program, Harvard School of Public Health, Boston, MA, USA. <sup>2</sup>Department of Environmental and Occupational Health, Università degli Studi di Milano and IRCCS Ca' Granda Maggiore Policlinico Hospital, Milan, Italy. <sup>3</sup>Department of Statistics, University of Milano-Bicocca, Milan, Italy. <sup>4</sup>Department of Experimental and Applied Medicine, Occupational Medicine and Industrial Hygiene, University of Brescia, Brescia, Italy. <sup>5</sup>Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago 60611, USA.

Received: 3 January 2013 Accepted: 5 July 2013

Published: 15 July 2013

### References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
2. Ohno S: **So much "junk" DNA in our genome.** *Brookhaven Symp Biol* 1972, **23**:366-370.
3. Mc CB: **The origin and behavior of mutable loci in maize.** *Proc Natl Acad Sci U S A* 1950, **36**:344-355.
4. Snyder MP, Kimbrell D, Hunkapiller M, Hill R, Fristrom J, Davidson N: **A transposable element that splits the promoter region inactivates a Drosophila cuticle protein gene.** *Proc Natl Acad Sci U S A* 1982, **79**:7430-7434.
5. Gallus GN, Cardaioli E, Rufa A, Da Pozzo P, Bianchi S, D'Eramo C, Collura M, Tumino M, Pavone L, Federico A: **Alu-element insertion in an OPA1 intron sequence associated with autosomal dominant optic atrophy.** *Mol Vis* 2010, **16**:178-183.
6. Byun HM, Heo K, Mitchell KJ, Yang AS: **Mono-allelic retrotransposon insertion addresses epigenetic transcriptional repression in human genome.** *J Biomed Sci* 2012, **19**:13.
7. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G: **Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer.** *PLoS Genet* 2010, **6**:e1000917.
8. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, et al: **Landscape of somatic retrotransposition in human cancers.** *Science* 2012, **337**(6097):967-971.
9. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al: **Somatic retrotransposition alters the genetic landscape of the human brain.** *Nature* 2011, **479**:534-537.
10. Rogers J: **Retroposons defined.** *Nature* 1983, **301**:460.
11. Cordaux R, Batzer MA: **The impact of retrotransposons on human genome evolution.** *Nat Rev Genet* 2009, **10**:691-703.
12. Slotkin RK, Martienssen R: **Transposable elements and the epigenetic regulation of the genome.** *Nat Rev Genet* 2007, **8**:272-285.
13. Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL: **Impact of transposable elements on the evolution of mammalian gene regulation.** *Cytogenet Genome Res* 2005, **110**:342-352.
14. Liu WM, Schmid CW: **Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation.** *Nucleic Acids Res* 1993, **21**:1351-1359.
15. Liu WM, Maraia RJ, Rubin CM, Schmid CW: **Alu transcripts: cytoplasmic localisation and regulation by DNA methylation.** *Nucleic Acids Res* 1994, **22**:1087-1095.
16. Schmid CW: **Does SINE evolution preclude Alu function?** *Nucleic Acids Res* 1998, **26**:4541-4550.
17. Nellaker C, Yao Y, Jones-Brando L, Mallet F, Yolken RH, Karlsson H: **Transactivation of elements in the human endogenous retrovirus W family by viral infection.** *Retrovirology* 2006, **3**:44.
18. Laska MJ, Brudek T, Nissen KK, Christensen T, Moller-Larsen A, Petersen T, Nexø BA: **Expression of HERV-Fc1, a human endogenous retrovirus, is**

- increased in patients with active multiple sclerosis. *J Virol* 2012, **86**:3713–3722.
19. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J: **Ischemic heart disease and stroke in relation to blood DNA methylation.** *Epidemiology* 2010, **21**:819–828.
  20. Kitkumthorn N, Tuangsintanakul T, Rattanatanayong P, Tiwawech D, Mutirangura A: **LINE-1 methylation in the peripheral blood mononuclear cells of cancer patients.** *Clin Chim Acta* 2012, **413**:869–874.
  21. Zhu ZZ, Sparrow D, Hou L, Tarantini L, Bollati V, Litonjua AA, Zanobetti A, Vokonas P, Wright RO, Baccarelli A, Schwartz J: **Repetitive element hypomethylation in blood leukocyte DNA and cancer incidence, prevalence, and mortality in elderly individuals: the Normative Aging Study.** *Cancer Causes Control* 2011, **22**:437–447.
  22. Jurka J, Smith T: **A fundamental division in the Alu family of repeated sequences.** *Proc Natl Acad Sci U S A* 1988, **85**:4775–4778.
  23. Giordano J, Ge Y, Gelfand Y, Abrusan G, Benson G, Warburton PE: **Evolutionary history of mammalian transposons determined by genome-wide defragmentation.** *PLoS Comput Biol* 2007, **3**:e137.
  24. Hough-Evans BR, Britten RJ, Davidson EH: **Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo.** *Dev Biol* 1988, **129**:198–208.
  25. Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselte TD, Hoppens CL, Deininger PL: **Structure and variability of recently inserted Alu family members.** *Nucleic Acids Res* 1990, **18**:6793–6798.
  26. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C: **Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells.** *Nucleic Acids Res* 1982, **10**:2709–2721.
  27. Matera AG, Hellmann U, Schmid CW: **A transpositionally and transcriptionally competent Alu subfamily.** *Mol Cell Biol* 1990, **10**:5424–5432.
  28. Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC: **Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2).** *Nucleic Acids Res* 2004, **32**:4100–4108.
  29. Baccarelli A, Tarantini L, Wright RO, Bollati V, Litonjua AA, Zanobetti A, Sparrow D, Vokonas P, Schwartz J: **Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study.** *Epigenetics* 2010, **5**:222–228.
  30. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, Zanobetti A, Sparrow D, Vokonas PS, Schwartz J: **Rapid DNA methylation changes after exposure to traffic particles.** *Am J Respir Crit Care Med* 2009, **179**:572–578.
  31. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, Byun HM, Jiang J, Marinelli B, Pesatori AC, et al: **Changes in DNA methylation patterns in subjects exposed to low-dose benzene.** *Cancer Res* 2007, **67**:876–880.
  32. Peluso M, Bollati V, Munnia A, Srivatanakul P, Jedpiyawongse A, Sangrajrang S, Piro S, Ceppi M, Bertazzi PA, Boffetta P, Baccarelli AA: **DNA methylation differences in exposed workers and nearby residents of the Ma Ta Phut industrial estate.** Rayong: Thailand. *Int J Epidemiol*; 2012.
  33. Pilsner JR, Hall MN, Liu X, Ahsan H, Ilievski V, Slavkovich V, Levy D, Factor-Litvak P, Graziano JH, Gamble MV: **Associations of plasma selenium with arsenic and genomic methylation of leukocyte DNA in Bangladesh.** *Environ Health Perspect* 2011, **119**:113–118.
  34. Wilhelm CS, Kelsey KT, Butler R, Plaza S, Gagne L, Zens MS, Andrew AS, Morris S, Nelson HH, Schned AR, et al: **Implications of LINE1 methylation for bladder cancer risk in women.** *Clin Cancer Res* 2010, **16**:1682–1689.
  35. Wright RO, Schwartz J, Wright RJ, Bollati V, Tarantini L, Park SK, Hu H, Sparrow D, Vokonas P, Baccarelli A: **Biomarkers of lead exposure and DNA methylation within retrotransposons.** *Environ Health Perspect* 2010, **118**:790–795.
  36. Kim KY, Kim DS, Lee SK, Lee IK, Kang JH, Chang YS, Jacobs DR, Steffes M, Lee DH: **Association of low-dose exposure to persistent organic pollutants with global DNA hypomethylation in healthy Koreans.** *Environ Health Perspect* 2010, **118**:370–374.
  37. Rusiecki JA, Baccarelli A, Bollati V, Tarantini L, Moore LE, Bonefeld-Jorgensen EC: **Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic Inuit.** *Environ Health Perspect* 2008, **116**:1547–1552.
  38. Wilson AS, Power BE, Molloy PL: **DNA hypomethylation and human diseases.** *Biochim Biophys Acta* 2007, **1775**:138–162.
  39. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP: **A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements.** *Nucleic Acids Res* 2004, **32**:e38.
  40. Byun HM, Nordio F, Coull BA, Tarantini L, Hou L, Bonzini M, Apostoli P, Bertazzi PA, Baccarelli A: **Temporal stability of epigenetic markers: sequence characteristics and predictors of short-term DNA methylation variations.** *PLoS One* 2012, **7**:e39220.
  41. Lee J, Cordaux R, Han K, Wang J, Hedges DJ, Liang P, Batzer MA: **Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons.** *Gene* 2007, **390**:18–27.
  42. Xing J, Hedges DJ, Han K, Wang H, Cordaux R, Batzer MA: **Alu element mutation spectra: molecular clocks and the effect of DNA methylation.** *J Mol Biol* 2004, **344**:675–682.
  43. van de Lagemat LN, Medstrand P, Mager DL: **Multiple effects govern endogenous retrovirus survival patterns in human gene introns.** *Genome Biol* 2006, **7**:R86.
  44. Landry JR, Rouhi A, Medstrand P, Mager DL: **The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter.** *Mol Biol Evol* 2002, **19**:1934–1942.
  45. Gu Z, Wang H, Nekrutenko A, Li WH: **Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence.** *Gene* 2000, **259**:81–88.
  46. Schulz WA, Steinhoff C, Florl AR: **Methylation of endogenous human retroelements in health and disease.** *Curr Top Microbiol Immunol* 2006, **310**:211–250.
  47. Choi S, Park SY, Jeong J, Cho E, Park S, Lee M, Kwak D, Lim JY, Jung WW, Sul D: **Identification of toxicological biomarkers of di(2-ethylhexyl) phthalate in proteins secreted by HepG2 cells using proteomic analysis.** *Proteomics* 2010, **10**:1831–1846.
  48. Dioni L, Hoxha M, Nordio F, Bonzini M, Tarantini L, Albetti B, Savarese A, Schwartz J, Bertazzi PA, Apostoli P, et al: **Effects of short-term exposure to inhalable particulate matter on telomere length, telomerase expression, and telomerase methylation in steel workers.** *Environ Health Perspect* 2011, **119**:622–627.
  49. Baccarelli A, Barretta F, Dou C, Zhang X, McCracken JP, Diaz A, Bertazzi PA, Schwartz J, Wang S, Hou L: **Effects of particulate air pollution on blood pressure in a highly exposed population in Beijing, China: a repeated-measure study.** *Environ Health* 2011, **10**:108.
  50. Choi SH, Worswick S, Byun HM, Shear T, Soussa JC, Wolff EM, Douer D, Garcia-Manero G, Liang G, Yang AS: **Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer.** *Int J Cancer* 2009, **125**:723–729.
  51. Khan H, Smit A, Boissinot S: **Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates.** *Genome Res* 2006, **16**:78–87.
  52. Batzer MA, Deininger PL: **Alu repeats and human genomic diversity.** *Nat Rev Genet* 2002, **3**:370–379.
  53. Gardiner-Garden M, Frommer M: **CpG islands in vertebrate genomes.** *J Mol Biol* 1987, **196**:261–282.
  54. Tarantini L, Bonzini M, Apostoli P, Pegoraro V, Bollati V, Marinelli B, Cantone L, Rizzo G, Hou L, Schwartz J, et al: **Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation.** *Environ Health Perspect* 2009, **117**:217–222.

doi:10.1186/1743-8977-10-28

Cite this article as: Byun et al.: Evolutionary age of repetitive element subfamilies and sensitivity of DNA methylation to airborne pollutants. *Particle and Fibre Toxicology* 2013 **10**:28.