VERITAS

D A S H

DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD

Urinary Benzene Biomarkers and DNA Methylation in Bulgarian Petrochemical Workers: Study Findings and Comparison of Linear and Beta Regression Models

Citation	Seow, Wei Jie, Angela Cecilia Pesatori, Emmanuel Dimont, Peter B. Farmer, Benedetta Albetti, Adrienne S. Ettinger, Valentina Bollati, et al. 2012. Urinary benzene biomarkers and DNA methylation in Bulgarian petrochemical workers: study findings and comparison of linear and beta regression models. PLoS ONE 7(12): e50471.
Published Version	doi:10.1371/journal.pone.0050471
Accessed	February 19, 2015 11:54:37 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:10587967
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

PLOS ONE

Urinary Benzene Biomarkers and DNA Methylation in Bulgarian Petrochemical Workers: Study Findings and Comparison of Linear and Beta Regression Models

Wei Jie Seow[1]*, Angela Cecilia Pesatori[2], Emmanuel Dimont[3], Peter B. Farmer[4], Benedetta Albetti[2], Adrienne S. Ettinger[5], Valentina Bollati[2], Claudia Bolognesi[6], Paola Roggieri[6], Teodor I. Panev[7], Tzveta Georgieva[8], Domenico Franco Merlo[9], Pier Alberto Bertazzi[2], Andrea A. Baccarelli[1]

1 Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, United States of America, 2 Department of Clinical Sciences and Community Health, University of Milan & Istituto Di Ricovero e Cura a Carattere Scientifico Fondazione Ca Granda Maggiore Policlinico Hospital, Milan, Italy, 3 Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, 4 Department of Cancer Studies and Molecular Medicine, University of Leicester, Leicester, United Kingdom, 5 Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, United States of America, 6 Environmental Carcinogenesis Unit, Istituto Di Ricovero e Cura a Carattere Scientifico AOU San Martino-IST-National Cancer Research Institute, Genoa, Italy, 7 Department of Chemical Analyses, National Centre of Public Health and Analyses, Sofia, Bulgaria, 8 Department of Public Health and Risk Assessment, National Center of Public Health and Analyses, Sofia, Bulgaria, 9 Epidemiology, Biostatistics and Clinical Trials, Istituto Di Ricovero e Cura a Carattere Scientifico AOU San Martino-IST-National Cancer Research Institute, Genoa, Italy

Abstract

Chronic occupational exposure to benzene is associated with an increased risk of hematological malignancies such as acute myeloid leukemia (AML), but the underlying mechanisms are still unclear. The main objective of this study was to investigate the association between benzene exposure and DNA methylation, both in repeated elements and candidate genes, in a population of 158 Bulgarian petrochemical workers and 50 unexposed office workers. Exposure assessment included personal monitoring of airborne benzene at work and urinary biomarkers of benzene metabolism (S-phenylmercapturic acid [SPMA] and trans,trans-muconic acid [t,t-MA]) at the end of the work-shift. The median levels of airborne benzene, SPMA and t,t-MA in workers were 0.46 ppm, 15.5 µg/L and 711 µg/L respectively, and exposure levels were significantly lower in the controls. Repeated-element DNA methylation was measured in Alu and LINE-1, and gene-specific methylation in MAGE and p15. DNA methylation levels were not significantly different between exposed workers and controls ($P > 0.05$). Both ordinary least squares (OLS) and beta-regression models were used to estimate benzene-methylation associations. Beta-regression showed better model specification, as reflected in improved coefficient of determination (pseudo R^2) and Akaike's information criterion (AIC). In beta-regression, we found statistically significant reductions in LINE-1 (−0.15%, $P < 0.01$) and p15 (−0.096%, $P < 0.01$) mean methylation levels with each interquartile range (IQR) increase in SPMA. This study showed statistically significant but weak associations of LINE-1 and p15 hypomethylation with SPMA in Bulgarian petrochemical workers. We showed that beta-regression is more appropriate than OLS regression for fitting methylation data.

Introduction

Benzene is a known human carcinogen classified by the International Agency for Research on Cancer (IARC) as a Group I carcinogen due to sufficient evidence for causing acute myeloid leukemia (AML) in humans [1]. It has also been associated with other diseases of the blood-forming systems such as myelodysplastic syndromes (MDS), lymphocytic leukemias and other cancers [2–5]. Some of the major current exposures to benzene occur in petrochemical industries, coke and steel plants, and from traffic exhaust and cigarettes smoke. Following findings on the effects of low-dose occupational exposure to benzene, such as hematotoxic effects seen at levels as low as 1 ppm [6], whether there is a safe

level of occupational benzene exposure has been highly debated. Once absorbed into the body, benzene is metabolized to benzene oxide before forming either S-phenylmercapturic acid (SPMA) or trans,trans-muconic acid (t,t-MA) in the liver and lungs [7]. SPMA and t,t-MA are the most commonly used biomarkers of occupational benzene exposure [8–10]. The exact mechanism by which benzene exerts its toxicity still remains unclear.

Aberrant DNA methylation, which may lead to genomic instability and altered gene expression, is frequently observed in AML [11]. Repeated-element DNA hypomethylation, as well as gene-specific hypermethylation or hypomethylation are commonly seen in hematological cancers [12]. Hypermethylation of tumor

suppressor *p15*, which leads to inactivation of the gene and subsequent uncontrolled proliferation of the cell, is commonly observed in AML and other hematological cancers [13–16]. The melanoma antigen family A (*MAGE*) gene, which encodes tumor-rejection antigens, is widely expressed in cells from hematological malignancies and is found to be hypomethylated with benzene exposure [17–20]. Studies have found significant correlation between SPMA and 8-oxo-2′-deoxyguanosine (8-oxodG) – a biomarker of oxidative DNA damage – measured in lymphocytes, which may then induce epigenetic modifications [21,22].

To date, most studies have performed statistical analyses of DNA methylation using ordinary least squares (OLS) regression, in which the Gaussian distribution is assumed for the methylation level. However, methylation is usually measured as percentage or proportion values (%5 mC), which are bounded in the (0,1) interval. In principle, the beta distribution, as opposed to the Gaussian distribution, would be the most natural statistical model for analyzing proportional data bounded in the (0,1) interval [23] and has also been shown in many empirical studies, to be an appropriate distributional model for proportion data where most of the linear assumptions tend to fail [24]. Beta regression has been proven to be more flexible than OLS regression, as it uses two sets of parameters to model trends in both the conditional mean and the precision (hence, conditional variance) of the proportion data distributed over the (0,1) interval with respect to a set of explanatory covariates [25]. However, whether a beta distribution is more appropriate to fit methylation data has yet not been evaluated on empirical data.

The main objective of this study was to investigate possible mechanisms of benzene toxicity via DNA methylation, by examining both repeated-element and gene-specific DNA methylation in a population of Bulgarian petrochemical workers. We also explored using beta-regression models to fit the methylation data as a methodological alternative to more commonly used OLS regression. Since DNA methylation is potentially modifiable through lifestyle or pharmacological intervention, identifying the association between benzene and methylation may be critical to identify biological effects from benzene exposure and design public health interventions among exposed workers.

Materials and Methods

Ethics Statement

Written informed consent was obtained prior to enrollment. The study was approved by the Genoa, Italy National Cancer Research Institute Ethics Committees.

Study Population

The study population consisted of 158 petrochemical workers employed at the Lukoil-Neftochim petrochemical plant in Burgas, Bulgaria and 50 unexposed office workers, used as controls, from the same plant. All subjects were recruited between October 1999 and September 2000 and had worked in the same position at the plant for at least one year. Self-administered questionnaires collected information regarding exposure history, medical history, lifestyle factors, smoking habits, occupational activities and other personal information.

Benzene Exposure Assessment

Individual airborne benzene measurements were based on personal monitoring sampling on the day before phlebotomy. Personal exposure to airborne benzene was measured by active sampling using charcoal sampling tubes as described previously [26]. Briefly, personal full-shift air monitoring was performed

during one working day using active samplers with flow of 20–30 cm^3/min worn near the breathing zone during a typical work-shift (midweek, from 8:00 a.m. to 2:00 p.m.), and analyzed with a Perkin Elmer Model 8300 gas chromatograph with flame-ionization detector and column DB-1–30 m. The limit of detection (LOD) for benzene was 0.023 ppm for the whole sample volume.

In addition to air benzene, urinary biomarkers of benzene exposure, i.e., S-phenylmercapturic acid (SPMA) and trans-trans-muconic acid (t,t-MA), which represent intermediary metabolites of benzene metabolism, were also assessed. Both biomarkers were measured in urine samples collected at the end of the work shift in the same day of air monitoring using previously-described methods [26,27]. Briefly, determination of urinary t,t-MA was performed by pre-purification of urine with solid-phase extraction using a strong-anion exchange column (SAX, 300 mg, Supelco Analytical, St. Louis, MO) followed by high performance liquid chromatography and UV detection as reported by Buratti and colleagues [28]. The LOD for t,t-MA was 50 µg/l, and forty-six measurements below LOD were assigned a value corresponding to half of LOD (i.e. 25 µg/l) with thirty-three (71.7%) of them being controls. Determination of urinary SPMA was based on an immunoassay technique according to a previously published procedure [29]. The immunoassay plate-based kit and reagents were from AB Biomonitoring Ltd. (Cardiff, Wales, United Kingdom). The detection limit of the procedure was 0.2 µg/L and none of the measurements were below LOD.

Methylation Outcomes

Four ml of whole blood was collected in EDTA tubes from each subject. DNA was extracted using the Wizard Genomic DNA purification kit (Promega, Madison, WI, USA) following the manufacturer's instructions. We performed DNA methylation analyses on bisulfite-treated DNA using a highly-quantitative analysis based on PCR-Pyrosequencing. 500 ng DNA (concentration 25 ng/µl) was treated using the EZ DNA Methylation-Gold^{TM} Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's protocol. Final elution was performed with 200 µl M-Elution Buffer. We performed DNA methylation analyses of Alu and LINE-1 repeated sequences, which allow for the amplification of a representa-tive pool of repetitive elements, as previously described [17].

To measure methylation of the *MAGE* and *p15* promoters, a 50 µL PCR was carried out in 25 µL GoTaq Green Master mix (Promega, Madison, WI, USA), 25 ng bisulfite-treated genomic DNA, with primer concentrations and PCR cycling conditions for each assay as shown in Table S1. PCR products were purified and sequenced by pyrosequencing as previously described using 0.3 µM sequencing primer [17]. PCR primers are shown in Table S2. Compared to other common methods of DNA methylation analysis, pyrosequencing-based assays have the advantage of producing individual measures of methylation at more than one CpG dinucleotide, thus more accurately reflecting DNA methylation in the region. In each gene promoter assay, we measured %5 mC at different CpG dinucleotides within the promoter region as shown in Table S3. Every sample was pyrosequenced two times for each assay to confirm reliability and the average of the two runs was used in the statistical analysis.

Statistical Analysis

Demographic, exposure and methylation variables were compared between workers and controls using Pearson's chi-square test for categorical variables, Welch two-sample t-test for normally-distributed continuous variables and Wilcoxon's rank sum test for

non-normally distributed continuous variables. First, simple linear ordinary least squares (OLS) models were fitted to Alu, LINE-1, *MAGE,* and *p15* methylation levels as dependent variables with each of the exposure variables separately. Additional multiple linear regression models were fitted that adjusted for the set of potential confounders of age, sex, current smoking habits, daily hours of environmental tobacco smoke (ETS) exposure and education level. Since DNA methylation was measured in white blood cells, and each white blood cell subtype may have different methylation profiles [30], we also measured proportions of the various leukocyte cell types (basophils, eosinophils, lymphocytes, monocytes and neutrophils) and adjusted for all analyses in order to eliminate influences from differences in proportions of cell types (Table S4). We then fit a second set of models using beta-regression [31]. We hypothesized that the beta-regression approach to be more appropriate and potentially provide more statistical power than OLS regression for methylation data, which are measured as proportions ranging from 0 to 1 (reported here, as customary, as percent values). Both the conditional mean and the precision (hence, conditional variance) of each methylation variable were modeled with the same set of covariates using the logit link for the conditional mean response and the log link for the conditional precision. The estimates from all models are reported as 'increase in percent methylation per IQR increase in exposure' in order to make the associations in the different models comparable for all exposure variables. Fitted mean methylation level trajectories for an individual with average characteristics (40 years old, non-smoking male worker with middle-school education and 5.3 ETS [Environmental Tobacco Smoking] hours) were plotted to provide a graphical representation of the results. Regular adjusted (for OLS) and pseudo adjusted (for beta regression) coefficients of determination (R^2), as well as Akaike's information criterion (AIC) were also computed for each of the adjusted models in order to compare relative goodness of fit between OLS and beta regression models. We also generated Pearson residual plots to compare model fit (Figures S1 and S2). All statistical analyses were conducted using the *betareg* and base packages in R version 2.13.1 (R-project.org. © 1998–2012) All tests were conducted as two-sided and considered significant with a p-value <0.05.

Results

The study participants were predominantly male and most of the individual characteristics were similar in both workers and controls (Table 1). However, education level was significantly lower in workers as compared to controls ($P<0.001$). ETS hours were significantly higher in workers compared to controls ($P<0.001$). Therefore both education and ETS hours were adjusted for in our regression models.

All exposure indices were significantly higher in workers than controls (Table 2). Median air benzene levels for controls were not calculated as almost all of the controls, except for one, were exposed to benzene levels below the limit of detection (LOD). Both urinary SPMA and t,t-MA were significantly higher among the workers (all $P<0.001$). Median SPMA levels were approximately 15.5 µg/L among workers and 10.1 µg/L among controls, and median t,t-MA levels were 711 µg/L among workers and 25 µg/L among controls. The correlation between the three different exposure variables in workers was moderate: 0.23 between airborne log-benzene and SPMA ($P<0.001$); 0.40 between the two biomarkers SPMA and t,t-MA ($P<0.001$); and 0.48 between airborne log-benzene and t,t-MA ($P<0.001$). We found no statistically significant difference between workers and controls in

the four methylation outcome variables as shown in Table 2. Because all of the controls except for one had air benzene below LOD, we did not use air benzene in the regression models reported below.

Table 3 presents the OLS regression results for the association between percent methylation of each methylation variable with each urinary biomarker. The magnitudes of association estimates were expressed as % change in DNA methylation per an IQR increase in the exposure biomarkers. Using urinary SPMA as exposure, we found small but statistically significant decreases in the methylation levels of both LINE-1 (−0.089% per IQR increase, $P=0.035$) and *p15* (−0.090% per IQR increase, $P=0.048$) with increasing SPMA in the unadjusted model. Only LINE-1 was significant when adjusted for all other covariates. LINE-1 methylation level was reduced by 0.14% ($P=0.009$) with each IQR increase in urinary SPMA exposure after adjusting for age, sex, education, smoking and ETS hours. There were no significant associations with t,t-MA. Adjusting for the possible effects that may be due to inherent between-participant differences in relative proportions of the various leukocyte cell types (Table S4) did not alter the results, so heterogeneity in leukocyte cell type was not included in the final model.

Table 4 shows the beta-regression results for the association between percent methylation of each methylation variable with each of the two urinary exposure variables. Using unadjusted beta-regression models, we found statistically significant but modest reductions in LINE-1 (0.085%, $P=0.033$) and in *p15* (0.084%, $P=0.023$) methylation levels with each IQR increase in urinary SPMA. After adjusting for all of the covariates, both LINE-1 and *p15* remained significant; we found a significant 0.15% decrease in LINE-1 ($P=0.005$), and 0.096% decrease in *p15* ($P=0.001$) methylation per an IQR increase in urinary SPMA. Significant associations were not observed with t,t-MA. Scatterplots showing SPMA and each of the methylation markers, as well as regression lines from beta-regression models are shown in Figure 1.

In general, the results from beta-regression were overall consistent with those from OLS regression. However, beta-regression models had higher adjusted pseudo R^2 values and lower AICs as compared to those generated by OLS regression models (Tables 3 and 4), indicating that beta regression performed better, in terms of model fit, than OLS regression models. We note that for the association of *MAGE* methylation with both SPMA and t,t-MA, OLS regression models showed negative adjusted R^2 values, which indicate particularly inadequate fit to the methylation data [32,33]. In addition, Pearson residual plots showed that beta regression generally provides better fit of the data (Figures S1 and S2).

Discussion

In this study, we investigated the association between occupational benzene exposure with both repeated-element and gene-specific DNA methylation in a population of Bulgarian petrochemical workers. Using beta-regression, we found small but statistically significant associations of LINE-1 and *p15* hypomethylation with SPMA after adjusting for covariates. We also showed that beta-regression provided better goodness-of-fit, as reflected in higher adjusted pseudo R^2 and lower AIC values, compared to standard linear OLS regression models. Our results show that exposure to benzene, as reflected in the internal dose measured by urinary SPMA was associated with both repeated-element and gene-specific hypomethylation.

Our finding showing a decrease in LINE-1 methylation with increasing urinary SPMA levels in beta-regression models is

Table 1. Characteristics of 158 Petrochemical Workers and 50 Controls, Bulgaria from 1999– 2000.

Variables[a]		Workers	Controls	P-value[b]
Sex	Male	138 (87.3%)	38 (76%)	0.089
	Female	20 (12.7%)	12 (24%)	
Age, years		39.9 (8.39)	41.3 (10.8)	0.41
Education level	None/Elementary School	13 (8.23%)	2 (4%)	<0.001*
	Middle School	121 (76.6%)	21 (42%)	
	High School and above	24 (15.2%)	27 (54%)	
Job-years		13.4 (8.79)	11.1 (10.9)	0.20
Smoking status	Current	103 (65.2%)	29 (58%)	0.13
	Past	127 (80.4%)	34 (68%)	
	Never	31 (19.6%)	16 (32%)	
Number cigarettes per day		12.7 (11.9)	9.72 (10.3)	0.092
Pack-years		15.3 (15.7)	13.8 (17.5)	0.61
ETS exposure	Yes	140 (89.2%)	42 (84%)	0.54
	No	17 (10.8%)	8 (16%)	
ETS hours		5.32 (4.05)	3.52 (2.41)	<0.001*
% Leukocytes	Basophils	0.78 (1.29)	0.34 (0.82)	0.006*
	Eosinophils	0.84 (1.49)	1.22 (1.40)	0.11
	Monocytes	1.35 (1.47)	1 (1.68)	0.19
	Lymphocytes	42.1 (9.44)	41.2 (9.29)	0.60
	Neutrophils	53.8 (9.46)	55.3 (8.90)	0.32

Abbreviations: ETS, environmental tobacco smoke.
*P<0.05.
[a]Categorical variables are expressed as n (%), and continuous variables are expressed as mean (SD).
[b]P-values were obtained from Pearson's chi-square test for categorical variables and Welch two-sample t-test for continuous variables and Wilcoxon rank-sum test for non-normally distribution variables.
doi:10.1371/journal.pone.0050471.t001

consistent with the global hypomethylation reported in response to various environmental exposures [17,34–38] and has been suggested to represent an early step preceding the severe hypomethylation typical of leukemia and other malignancies [12,39–42]. The observed association was small in size, consistent with previous results by Bollati et al. in a study of gas-station attendants and police officers exposed to low-level benzene [17]. Similar effect sizes on DNA methylation measured in blood leukocytes from healthy participants have been suggested to

Table 2. A Comparison of exposure and methylation variables between workers and controls, Median (IQR).

Variables[a]	Workers (N = 158)						Controls (N = 50)						P-value[b]
	Min	p25	Median (IQR)	p75	Max	% <LOD	Min	p25	Median (IQR)	p75	Max	% <LOD	
Benzene, ppm	0.012[c]	0.19	0.46 (1.27)	1.47	23.9	2/158 (1.3%)	0.012[c]	0.012[c]	0.012 (0)[c]	0.012[c]	0.65	49/50 (98%)	–
SPMA, µg/L	0.24	9.95	15.5 (44.9)	54.9	349.4	0/158 (0%)	0.24	7.73	10.1 (6.6)	14.3	59.7	0/50 (0%)	<0.001*
t,t-MA, µg/L	25.0[d]	311.5	711 (1270)	1581	9961	13/158 (8.2%)	25.0[d]	25.0[d]	25.0 (75.0)[d]	100	754	33/50 (66%)	<0.001*
Alu, %	22.4	25.1	25.6 (1.03)	26.1	27.2	–	21.9	24.6	25.3 (1.6)	26.2	27.9	–	0.15
LINE-1, %	71.6	80.7	81.8 (1.97)	82.7	85.3	–	75.4	80.5	81.9 (2.23)	82.8	84.3	–	0.72
MAGE, %	47.0	90.4	91.6 (2.16)	92.5	96.4	–	86.2	90.2	91.6 (2.05)	92.3	93.7	–	0.60
p15, %	1.12	2.60	3.32 (1.86)	4.46	14.1	–	1.80	2.87	3.45 (1.23)	4.10	7.65	–	0.78

Abbreviations: SPMA, S-phenylmercapturic acid; t,t-MA, Trans-trans-muconic acid.
Data were missing and excluded in the analysis for the following variables: SPMA (10 individuals), t,t-MA (7 individuals), Alu (14 individuals), LINE-1 (15 individuals), *MAGE* (16 individuals) and *p15* (28 individuals).
*P<0.05.
[a]All exposure and outcome variables are expressed as minimum (Min), 25th quantile (p25), median (IQR), 75th quantile (p75), maximum (Max).
[b]P-values were obtained from Wilcoxon's rank sum test.
[c]Almost all controls (49/50) have values less than LOD of 0.023 ppm and were assigned a value corresponding to half of LOD in the analysis (i.e. 0.012 ppm).
[d]A majority of the controls (33/50) have values less than LOD of 50 µg/L and were assigned a value corresponding to half of LOD in the analysis (i.e. 25 µg/l).
doi:10.1371/journal.pone.0050471.t002

Table 3. Association Estimates of Urinary Biomarkers, SPMA and t,t-MA, on Repeated-element and Gene-Specific Percent Methylation using OLS Regression.

Biomarker	Outcome	Unadjusted		Adjusted[a]			
		Coefficient[b]	P-value	Coefficient[b]	P-value	R[2c]	AIC
SPMA	Alu	0.010	0.72	0.0072	0.83	0.032	444.0
	LINE-1	−0.089	0.035*	−0.14	0.009*	0.054	581.1
	MAGE	0.025	0.80	0.011	0.93	−0.034	838.7
	p15	−0.090	0.048*	−0.077	0.20	0.021	562.2
t,t-MA	Alu	0.027	0.51	0.034	0.48	0.017	486.0
	LINE-1	−0.054	0.38	−0.044	0.56	0.012	633.0
	MAGE	0.056	0.71	0.034	0.85	−0.025	901.9
	p15	0.057	0.39	0.081	0.32	0.011	608.7

Abbreviations: SPMA, S-phenylmercapturic acid; t,t-MA, Trans-trans-muconic acid; AIC, Akaike's information criterion.
*$P < 0.05$.
[a]Model adjusted for age, sex, smoking history, education, ETS hours.
[b]Coefficient refers to the change in methylation % per IQR change in exposure variable.
[c]Adjusted R^2.
doi:10.1371/journal.pone.0050471.t003

represent an early deviation from normal methylation patterns induced by benzene exposure [17].

We also found a decrease in *p15* methylation in relation to increasing SPMA levels, contrary to the hypermethylation of *p15* reported in the previous investigation by Bollati et al. [17]. Also, we did not find any significant associations with Alu and *MAGE* methylation, which showed a small but significant negative association with airborne benzene levels in the Bollati et al. study [17]. Several differences in the levels and sources of exposure might account for the discrepancies between the two studies. Bollati et al. studied traffic-exposed subjects in an urban area and investigated associations of DNA methylation with airborne benzene levels. In our study instead, we examined a population of industrial workers with higher workplace exposure levels, and evaluated associations with urinary biomarkers of benzene metabolism. The different settings of the exposures, as well as the different exposure levels may account for the discrepancies in the results.

This is the first study to compare the use of beta-regression and OLS regression methods for analyzing proportion methylation outcomes. We found that beta-regression provided a better goodness-of-fit, as reflected in higher adjusted pseudo R^2 values and lower AIC values. This finding is consistent with the concept that beta-regression is more appropriate for analyzing proportion data, such as DNA methylation, compared to standard linear OLS regression models. We obtained negative adjusted R^2 values for *MAGE* in the OLS regression model, because the adjusted R^2 takes into account the number of parameters in the model by subtracting a function of that number from the value, thus potentially resulting in negative values in cases of particularly poor fit [32,33]. Although results from beta-regression were overall consistent with OLS regression, in beta-regression we found some significant, but weak, associations of benzene exposure and DNA methylation which would have been missed by OLS regression. This finding suggests that beta-regression may be more sensitive for detecting small changes in percent methylation. This is due to the more flexible nature of the beta distribution, which unlike the Gaussian distribution having independent mean and variance parameters, allows for the mean and the variance to be related, where the variance is a function of the mean [43]. In fact, being a member of the exponential family, the beta distribution has a

variance function which directly depends on the mean and, as such, differences in variance within a variable can be just as informative as differences in the mean [31]. In addition, beta-regression allows the use of various link functions linking the outcome measure to the set of explanatory covariates, allowing one to fit models with non-linear mean trajectories. Alternatively, simplex regression can also be used to fit data bounded in (0,1) and has been shown to give estimates concordant with beta regression [44].

In the present study, we found associations of urinary SPMA with DNA methylation markers, whereas no gross methylation differences were observed between exposed workers and controls. In general, biomarkers of exposure, including SPMA, are used to yield more accurate measures of biological dose than occupational group categories, as they reflect the level of absorption, metabolism, as well as of the clearance of the chemical. SPMA in particular, has been previously found to be associated with oxidative stress [21], a benzene-induced cellular stressor and well-established modifier of methylation patterns [45]. On the contrary, we did not observe any significant correlation of t,t-MA levels with DNA methylation. Previous studies have also found the correlation of both SPMA and t,t-MA with airborne benzene to be relatively low and SPMA to be a more reliable biomarker for benzene exposure due to its longer half-life than t,t-MA [46–48].

The present study has a number of limitations, including a relatively small number of participants; and possible unmeasured confounding, such as non-occupational benzene exposures (e.g. from vehicle traffic). It should also be noted that other factors that might affect DNA methylation, such as nutritional factors, were unmeasured and could not be adjusted for in this study. Our study has also several strengths including: accurate outcome assessment with pyrosequencing duplicates completed for each outcome; biomarkers measured to take into account the internal dose of each subject with complete data for urinary SPMA; and the use of a population of workers exposed to high benzene levels which represents a different setting than those investigated in previous epigenetic studies. The workers in our population were exposed to benzene for an average of 13.4 years, and therefore any potential aberrant changes in methylation if any, caused by chronic benzene exposure should have been clearly visible after such prolonged
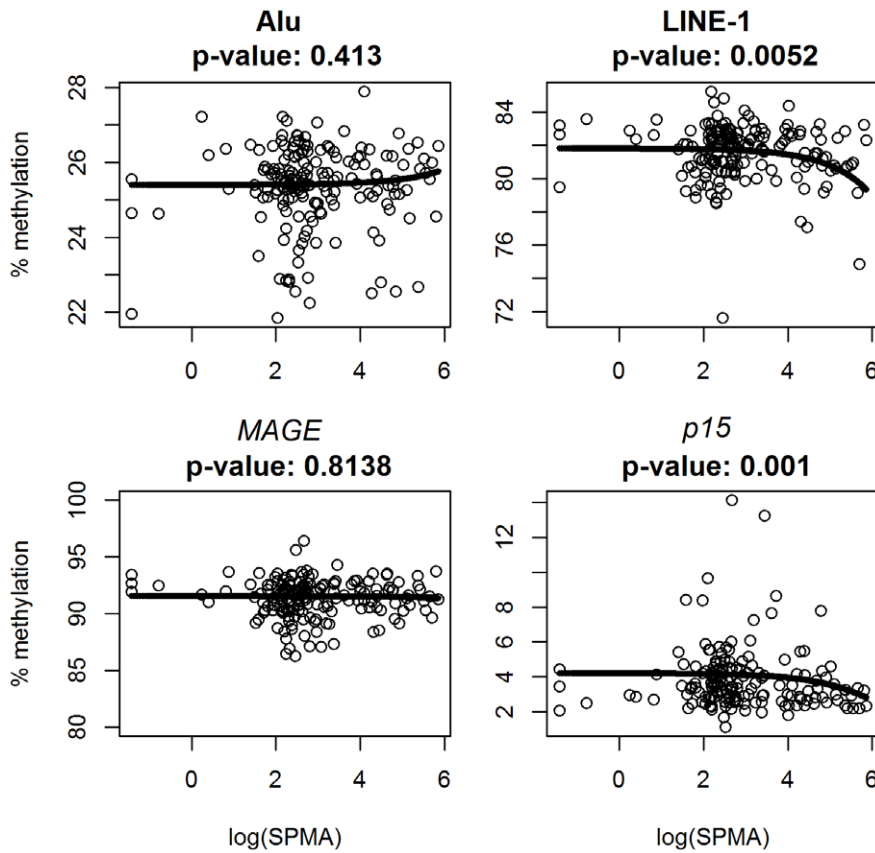
Figure 1. Association of S-phenylmercapturic acid (SPMA) with DNA methylation in Alu, LINE-1, *MAGE* and *p15* methylation. Fitted beta regression models of repeated-element and gene-specific methylation % versus log(SPMA), adjusted for potential confounders as described in the text. Lines correspond to fitted mean trajectories from beta regression models using the logit link, evaluated for hypothetical individuals with sample mean covariate values (petrochemical workers, non-smoking male, age: 40, ETS: 5.3 hours, education: middle-school). P-values shown correspond to main associations of SPMA in each model.
doi:10.1371/journal.pone.0050471.g001

exposure to high levels of benzene. We also considered potential influences of various leukocyte cell types on between-subject heterogeneity in the response.

Our findings showed that occupational benzene exposure was associated with small but statistically significant DNA methylation changes in the workers. Our results lacked consistency across

Table 4. Association Estimates of Urinary Biomarkers, SPMA and t,t-MA, on Repeated-element and Gene-Specific Percent Methylation using Beta-regression.

Biomarker	Outcome	Unadjusted		Adjusted[a]		R² [c]	AIC
		Coefficient[b]	*P*-value	Coefficient[b]	*P*-value		
SPMA	Alu	0.010	0.71	0.023	0.41	0.050	−908.0
	LINE-1	−0.085	0.033[*]	−0.15	0.005[*]	0.10	−768.6
	MAGE	0.010	0.88	−0.010	0.81	0.000018	−652.4
	p15	−0.084	0.023[*]	−0.096	0.001[*]	0.070	−755.5
t,t-MA	Alu	0.028	0.51	0.013	0.70	0.048	−976.0
	LINE-1	−0.050	0.40	−0.0029	0.97	0.056	−834.4
	MAGE	0.018	0.86	−0.026	0.71	0.0021	−723.0
	p15	0.12	0.81	0.11	0.12	0.051	−824.2

Abbreviations: SPMA, S-phenylmercapturic acid; t,t-MA, Trans-trans-muconic acid; AIC, Akaike's information criterion.
[*]$P < 0.05$.
[a]Model adjusted for age, sex, smoking history, education, ETS hours.
[b]Coefficient refers to the change in methylation % per IQR change in exposure variable.
[c]Pseudo-adjusted R^2.
doi:10.1371/journal.pone.0050471.t004

different biomarkers and exposure metrics and should therefore be interpreted with caution. Also, some of our findings, such as the negative associations between SPMA and DNA methylation of p15, are at variance with previous findings on benzene-exposed individuals [17]. Because DNA methylation is potentially modifiable through lifestyle and pharmacological interventions, if confirmed, our findings may open new paths for prevention against benzene-induced carcinogenesis. If benzene-related hypomethylation were to be found to be a mechanistic intermediate along the path linking benzene to its carcinogenic effects, then targeted interventions, such as providing benzene-exposed workers with dietary supplements, including folic acid which may contrast the association of benzene with DNA hypomethylation [49] could be developed and tested among benzene-exposed workers.

Supporting Information

Figure S1 OLS Pearson Residual Plots of the Association of S-phenylmercapturic acid (SPMA) with DNA methylation in Alu, LINE-1, *MAGE* and *p15* methylation. OLS Pearson residuals of the association of S-phenylmercapturic acid (SPMA) with DNA methylation in Alu, LINE-1, *MAGE* and *p15* methylation %, adjusted for potential confounders as described in the text. Lines correspond to zero reference residual values on the y-axis.
(TIFF)

Figure S2 Beta Regression Pearson Residual Plots of the Association of S-phenylmercapturic acid (SPMA) **with DNA methylation in Alu, LINE-1, *MAGE* and *p15* methylation.** Beta Regression Pearson residuals of the association of S-phenylmercapturic acid (SPMA) with DNA methylation in Alu, LINE-1, *MAGE* and *p15* methylation %, adjusted for potential confounders as described in the text. Lines correspond to zero reference residual values on the y-axis.
(TIFF)

Table S1 PCR Primers Concentrations and Cycling Conditions.
(PDF)

Table S2 Primers for PCR assay.
(PDF)

Table S3 Localization of Gene Promoters and Regions Amplified.
(PDF)

Table S4 Model Estimates of OLS and Beta Regression adjusting for Leukocyte Proportions.
(PDF)

Author Contributions

Conceived and designed the experiments: AP ASE VB DFM PAB AB. Performed the experiments: BA. Analyzed the data: WJS ED. Wrote the paper: WJS. Provided critical revisions for the manuscript: AP ED AB. Data collection and exposure assessment: CB PR TP TG PF DFM.

References

1. IARC (1987) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Supplement 7. Lyon, France, Geneva, Switzerland: The Agency; Secretariat of the World Health Organization distributor. 120.
2. Khalade A, Jaakkola MS, Pukkala E, Jaakkola JJ (2010) Exposure to benzene at work and the risk of leukemia: a systematic review and meta-analysis. Environ Health 9: 31.
3. Smith MT (2010) Advances in understanding benzene health effects and susceptibility. Annu Rev Public Health 31: 133–148.
4. Steinmaus C, Smith AH, Jones RM, Smith MT (2008) Meta-analysis of benzene exposure and non-Hodgkin lymphoma: biases could mask an important association. Occup Environ Med 65: 371–378.
5. Schnatter AR, Rosamilia K, Wojcik NC (2005) Review of the literature on benzene exposure and leukemia subtypes. Chem Biol Interact 153–154: 9–21.
6. Lan Q, Zhang L, Li G, Vermeulen R, Weinberg RS, et al. (2004) Hematotoxicity in workers exposed to low levels of benzene. Science 306: 1774–1776.
7. Kim S, Vermeulen R, Waidyanatha S, Johnson BA, Lan Q, et al. (2006) Using urinary biomarkers to elucidate dose-related patterns of human benzene metabolism. Carcinogenesis 27: 772–781.
8. Farmer PB, Kaur B, Roach J, Levy L, Consonni D, et al. (2005) The use of S-phenylmercapturic acid as a biomarker in molecular epidemiology studies of benzene. Chem Biol Interact 153–154: 97–102.
9. Fustinoni S, Consonni D, Campo L, Buratti M, Colombi A, et al. (2005) Monitoring low benzene exposure: comparative evaluation of urinary biomarkers, influence of cigarette smoking, and genetic polymorphisms. Cancer Epidemiol Biomarkers Prev 14: 2237–2244.
10. Qu QS, Melikian AA, Li GL, Shore R, Chen LC, et al. (2000) Validation of biomarkers in humans exposed to benzene: Urine metabolites. Am J Ind Med 37: 522–531.
11. Kuang SQ, Tong WG, Yang H, Lin W, Lee MK, et al. (2008) Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. Leukemia 22: 1529–1538.
12. Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301: 89–92.
13. Herman JG, Civin CI, Issa JP, Collector MI, Sharkis SJ, et al. (1997) Distinct patterns of inactivation of p15INK4B and p16INK4A characterize the major types of hematological malignancies. Cancer Res 57: 837–841.
14. Yin SN, Hayes RB, Linet MS, Li GL, Dosemeci M, et al. (1996) A cohort study of cancer among benzene-exposed workers in China: overall results. Am J Ind Med 29: 227–235.
15. Melki JR, Vincent PC, Clark SJ (1999) Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. Cancer Res 59: 3730–3740.

16. Guo SX, Taki T, Ohnishi H, Piao HY, Tabuchi K, et al. (2000) Hypermethylation of p16 and p15 genes and RB protein expression in acute leukemia. Leuk Res 24: 39–46.
17. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, et al. (2007) Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res 67: 876–880.
18. Traversari C, van der Bruggen P, Luescher IF, Lurquin C, Chomez P, et al. (1992) A nonapeptide encoded by human gene MAGE-1 is recognized on HLA-A1 by cytolytic T lymphocytes directed against tumor antigen MZ2-E. J Exp Med 176: 1453–1457.
19. van Baren N, Brasseur F, Godelaine D, Hames G, Ferrant A, et al. (1999) Genes encoding tumor-specific antigens are expressed in human myeloma cells. Blood 94: 1156–1164.
20. Greiner J, Ringhoffer M, Simikopinko O, Szmaragowska A, Huebsch S, et al. (2000) Simultaneous expression of different immunogenic antigens in acute myeloid leukemia. Exp Hematol 28: 1413–1422.
21. Sorensen M, Skov H, Autrup H, Hertel O, Loft S (2003) Urban benzene exposure and oxidative DNA damage: influence of genetic polymorphisms in metabolism genes. Science of the Total Environment 309: 69–80.
22. Lim SO, Gu JM, Kim MS, Kim HS, Park YN, et al. (2008) Epigenetic Changes Induced by Reactive Oxygen Species in Hepatocellular Carcinoma: Methylation of the E-cadherin Promoter. Gastroenterology 135: 2128–2140.
23. Spanos A (1999) Probability theory and statistical inference : econometric modeling with observational data. Cambridge, UK; New York, NY, USA: Cambridge University Press. xxvii, 815 p.
24. Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions. New York: Wiley.
25. Kieschnick R, McCullough BD (2003) Regression analysis of variates observed on (0,1): percentages, proportions and fractions. Statistical Modelling 3: 193–213.
26. Pesatori AC, Garte S, Popov T, Georgieva T, Panev T, et al. (2009) Early effects of low benzene exposure on blood cell counts in Bulgarian petrochemical workers. Med Lav 100: 83–90.
27. Fustinoni S, Buratti M, Campo L, Colombi A, Consonni D, et al. (2005) Urinary t,t-muconic acid, S-phenylmercapturic acid and benzene as biomarkers of low benzene exposure. Chem Biol Interact 153–154: 253–256.
28. Buratti M, Fustinoni S, Colombi A (1996) Fast liquid chromatographic determination of urinary trans,trans-muconic acid. J Chromatogr B Biomed Appl 677: 257–263.
29. Aston JP, Ball RL, Pople JE, Jones K, Cocker J (2002) Development and validation of a competitive immunoassay for urinary S-phenylmercapturic acid and its application in benzene biological monitoring. Biomarkers 7: 103–112.

30. Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, et al. (2012) Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. Int J Epidemiol 41: 126–139.

31. Ferrari SLP, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. Journal of Applied Statistics 31: 799–815.

32. Brooks C (2008) Introductory econometrics for finance. Cambridge England; New York: Cambridge University Press. xxiv, 648 p.

33. Everitt B, Howell DC (2005) Encyclopedia of statistics in behavioral science. Hoboken, N.J.: John Wiley & Sons.

34. Rusiecki JA, Baccarelli A, Bollati V, Tarantini L, Moore LE, et al. (2008) Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic Inuit. Environ Health Perspect 116: 1547–1552.

35. Zhao CQ, Young MR, Diwan BA, Coogan TP, Waalkes MP (1997) Association of arsenic-induced malignant transformation with DNA hypomethylation and aberrant gene expression. Proc Natl Acad Sci U S A 94: 10907–10912.

36. Takiguchi M, Achanzar WE, Qu W, Li G, Waalkes MP (2003) Effects of cadmium on DNA-(Cytosine-5) methyltransferase activity and DNA methylation status during cadmium-induced cellular transformation. Exp Cell Res 286: 355–365.

37. Chen H, Li S, Liu J, Diwan BA, Barrett JC, et al. (2004) Chronic inorganic arsenic exposure induces hepatic global and individual gene hypomethylation: implications for arsenic hepatocarcinogenesis. Carcinogenesis 25: 1779–1786.

38. Ji Z, Zhang L, Peng V, Ren X, McHale CM, et al. (2010) A comparison of the cytogenetic alterations and global DNA hypomethylation induced by the benzene metabolite, hydroquinone, with those induced by melphalan and etoposide. Leukemia 24: 986–991.

39. Suter CM, Martin DI, Ward RL (2004) Hypomethylation of L1 retro-transposons in colorectal cancer and adjacent normal tissue. Int J Colorectal Dis 19: 95–101.

40. Lin CH, Hsieh SY, Sheen IS, Lee WC, Chen TC, et al. (2001) Genome-wide hypomethylation in hepatocellular carcinogenesis. Cancer Res 61: 4238–4243.

41. Wahlfors J, Hiltunen H, Heinonen K, Hamalainen E, Alhonen L, et al. (1992) Genomic hypomethylation in human chronic lymphocytic leukemia. Blood 80: 2074–2080.

42. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, et al. (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res 22: 246–258.

43. Paolino P (2001) Maximum likelihood estimation of models with beta-distributed dependent variables. Political Analysis 9: 325–346.

44. Wang XF (2012) Joint generalized models for multidimensional outcomes: a case study of neuroscience data from multimodalities. Biom J 54: 264–280.

45. Bollati V, Baccarelli A (2010) Environmental epigenetics. Heredity (Edinb) 105: 105–112.

46. Lovreglio P, Barbieri A, Carrieri M, Sabatini L, Fracasso ME, et al. (2010) Validity of new biomarkers of internal dose for use in the biological monitoring of occupational and environmental exposure to low concentrations of benzene and toluene. Int Arch Occup Environ Health 83: 341–356.

47. Carrieri M, Tranfo G, Pigini D, Paci E, Salamon F, et al. (2010) Correlation between environmental and biological monitoring of exposure to benzene in petrochemical industry operators. Toxicol Lett 192: 17–21.

48. Boogaard PJ, vanSittert NJ (1996) Suitability of S-phenyl mercapturic acid and trans-trans-muconic acid as biomarkers for exposure to low concentrations of benzene. Environ Health Perspect 104: 1151–1157.

49. Pufulete M, Al-Ghnaniem R, Khushal A, Appleby P, Harris N, et al. (2005) Effect of folic acid supplementation on genomic DNA methylation in patients with colorectal adenoma. Gut 54: 648–653.