

# LINE-1 methylation in granulocyte DNA and trihalomethane exposure is associated with bladder cancer risk

Lucas A Salas<sup>1,2,3</sup>, Cristina M Villanueva<sup>1,2,3,4,\*</sup>, Salman M Tajuddin<sup>5</sup>, André F S Amaral<sup>5</sup>, Agustín F Fernandez<sup>6</sup>, Lee E Moore<sup>7</sup>, Alfredo Carrato<sup>8,9</sup>, Adonina Tardón<sup>3,10</sup>, Consol Serra<sup>11,12</sup>, Reina García-Closas<sup>13</sup>, Xavier Basagaña<sup>1,2,3</sup>, Nathaniel Rothman<sup>7</sup>, Debra T Silverman<sup>7</sup>, Kenneth P Cantor<sup>7</sup>, Manolis Kogevinas<sup>1,2,3,4,14</sup>, Francisco X Real<sup>2,15</sup>, Mario F Fraga<sup>6</sup>, and Núria Malats<sup>5,\*</sup>

<sup>1</sup>Centre for Research in Environmental Epidemiology (CREAL); Barcelona, Spain; <sup>2</sup>Universitat Pompeu Fabra (UPF); Barcelona, Spain; <sup>3</sup>CIBER Epidemiología y Salud Pública (CIBERESP); Barcelona, Spain; <sup>4</sup>IMIM (Hospital del Mar Medical Research Institute); Barcelona, Spain; <sup>5</sup>Genetic and Molecular Epidemiology Group; Spanish National Cancer Research Center (CNIO), Madrid, Spain; <sup>6</sup>Cancer Epigenetics Laboratory; Instituto Universitario de Oncología del Principado de Asturias (IUOPA-HUCA); University of Oviedo; Oviedo, Spain; <sup>7</sup>National Cancer Institute; Bethesda, MD USA; <sup>8</sup>University Hospital of Elche; Elche, Spain; <sup>9</sup>Hospital Ramon y Cajal; Madrid, Spain; <sup>10</sup>University of Oviedo; Oviedo, Spain; <sup>11</sup>Universitat Pompeu Fabra (UPF); Center of Research in Occupational Health; Spain; <sup>12</sup>Consorci Hospitalari Parc Taulí; Sabadell, Spain; <sup>13</sup>Hospital Universitario de Canarias; La Laguna, Spain; <sup>14</sup>National School of Public Health; Athens, Greece; <sup>15</sup>Epithelial Carcinogenesis Group; Spanish National Cancer Research Center (CNIO); Spain

**Keywords:** DNA methylation, epigenetic repression, long interspersed nucleotide elements, LINE-1, trihalomethanes, urinary bladder cancer

**Abbreviations:** DBP, Disinfection by-products; LINE-1, Long Interspersed Element 1; OR, Odds ratio; SBC/EPICURO, Spanish Bladder Cancer/EPICURO Study; THM, Trihalomethanes; UBC, Urothelial bladder cancer; y, years; %5mC; Percentage of 5-methylcytosine, 95% CI; 95% confidence intervals.

DNA methylation changes contribute to bladder carcinogenesis. Trihalomethanes (THM), a class of disinfection by-products, are associated with increased urothelial bladder cancer (UBC) risk. THM exposure in animal models produces DNA hypomethylation. We evaluated the relationship of LINE-1 5-methylcytosine levels (LINE-1%5mC) as outcome of long-term THM exposure among controls and as an effect modifier in the association between THM exposure and UBC risk. We used a case-control study of UBC conducted in Spain. We obtained personal lifetime residential THM levels and measured LINE-1%5mC by pyrosequencing in granulocyte DNA from blood samples in 548 incident cases and 559 hospital controls. Two LINE-1%5mC clusters (above and below 64%) were identified through unsupervised hierarchical cluster analysis. The association between THM levels and LINE-1%5mC was evaluated with  $\beta$  regression analyses and logistic regression was used to estimate odds ratios (OR) adjusting for covariables. LINE-1%5mC change between percentiles 75<sup>th</sup> and 25<sup>th</sup> of THM levels was 1.8% (95% confidence interval (CI): 0.1, 3.4%) among controls. THM levels above vs. below the median (26  $\mu$ g/L) were associated with increased UBC risk, OR = 1.86 (95% CI: 1.25, 2.75), overall and among subjects with low levels of LINE-1%5mC (n = 975), OR = 2.14 (95% CI: 1.39, 3.30), but not associated with UBC risk among subjects' high levels of LINE-1%5mC (n = 162), interaction  $P = 0.03$ . Results suggest a positive association between LINE-1%5mC and THM levels among controls, and LINE-1%5mC status may modify the association between UBC risk and THM exposure. Because reverse causation and chance cannot be ruled out, confirmation studies are warranted.

## Introduction

Disinfection by-products (DBP) constitute complex mixtures of undesired toxic chemicals, including carcinogens formed as side-products when disinfectants (i.e., chlorine) react with organic matter. Trihalomethanes (THM), one of the most prevalent groups of chlorination DBP, have been used as surrogates of

total DBP in epidemiological studies. Levels of THM in Spain have been among the highest in Europe.<sup>1</sup> Lifetime exposure to THM in drinking water has been related to urothelial bladder cancer (UBC) risk,<sup>2,3</sup> but the mechanism of action is poorly understood. UBC is the fourth most prevalent cancer among Spanish males,<sup>4</sup> and is consistently associated with smoking habits.<sup>5</sup>

© Lucas A Salas, Cristina M Villanueva, Salman M Tajuddin, André F S Amaral, Agustín F Fernandez, Lee E Moore, Alfredo Carrato, Adonina Tardón, Consol Serra, Reina García-Closas, Xavier Basagaña, Nathaniel Rothman, Debra T Silverman, Kenneth P Cantor, Manolis Kogevinas, Francisco X Real, Mario F Fraga, and Núria Malats

\*Correspondence to: Núria Malats; Email: nmalats@cnio.es; Cristina M Villanueva; Email: cvillanueva@creal.cat

Submitted: 09/03/2014; Revised: 10/01/2014; Accepted: 10/03/2014

<http://dx.doi.org/10.4161/15592294.2014.983377>

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

The carcinogenicity of DBP is attributed to both genotoxic and non-genotoxic mechanisms. The proposed genotoxic mechanism requires intracellular conjugation of the brominated compounds through glutathione S-transferase theta 1, GSTT1.<sup>6-10</sup> Given that this mechanism does not explain completely the carcinogenic effect observed in animals or humans, non-genotoxic mechanisms remain a plausible explanation. Experiments in rodents show that exposure to THM and haloacetic acids induces alterations in global and specific genes DNA methylation, increases mRNA expression of protooncogenes *c-myc* and *c-jun*, and induces kidney and liver tumors.<sup>11-14</sup> Although experimental evidence suggests the plausibility of this mechanism of action, the relationship between chronic THM exposure and epigenetic changes has not been explored in human populations.

Long Interspersed Element-1 (LINE-1) is a retrotransposon highly repeated in the genome, and its CpG site methylation can be used as a surrogate measure of global DNA methylation.<sup>15</sup> LINE-1 is composed of highly repeated sequences, each 6,000 to 8,000 base pairs in length.<sup>16,17</sup> The human genome contains more than 500,000 copies of LINE-1, of which 3000 are potentially transposable.<sup>18</sup> LINE-1 retrotransposition has been related to several diseases including various types of cancer.<sup>19,20</sup> DNA methylation is the main mechanism that inhibits retrotransposon expression preventing genomic damage.<sup>20,21</sup> Several studies have used LINE-1 methylation as a biomarker of susceptibility to cancer.<sup>22</sup> Both low and high levels of 5-methylcytosine (5mC) within LINE-1 sequences in leukocytes have been linked to an increased risk of UBC.<sup>23-26</sup> We evaluated the relationship of LINE-1 5-methylcytosine levels (LINE-1%5mC) as

outcome of long-term THM exposure and as an effect modifier in the association between long-term THM exposure and UBC risk.

## Results

Overall, 2,090 study subjects provided a blood sample for DNA extraction (1,083 cases and 1,007 controls).<sup>27</sup> Of these, 1,849 [88.5%; 957/1,083 (88.4%) cases and 892/1,007 (88.6%) controls] had available samples for LINE-1%5mC measurements. A total of 1,107 (53.0%) subjects had reliable or high quality interview data, as reported by the interviewer, and more than 70% of y in the exposure window (see methods below) with valid THM data, representing 548 (50.6%) cases and 559 (55.5%) controls in the final analysis. We compared sex, age, area and smoking status between subjects included and excluded from the analysis. Statistically significant differences were found for age (excluded subjects were, on average, 2.7 y older than those included,  $P < 0.001$ ) and area (fewer subjects excluded from Alicante, and more subjects excluded from Manresa and Barcelona). No differences were found for case-control status, sex ratio, and smoking status.

Median age of study participants was 66 y (range = 20–80 y) and 88.1% were men. Smoking status showed statistically significant differences between cases and controls (Table 1), which were maintained in logistic regression models adjusted for age, sex, and area (data not shown). Average LINE-1%5mC level showed a bimodal distribution, overall, by case-control status (Supplemental Material, Figure S1, panels A and B), and within

**Table 1.** Distribution of cases and controls with LINE-1% 5-methylcytosine (LINE-1%5mC) and trihalomethane (THM) data in the Spanish Bladder Cancer/EPICURO (SBC/EPICURO) study

Variable	Cases n = 548 n(%)	Controls n = 559 n(%)	P <sup>1</sup>
<b>Sex</b>			
Male	477 (87.0)	498 (89.1)	0.3
Female	71 (13.0)	61 (10.9)	
<b>Age (y)</b>			
Mean ± SE	64.5 ± 0.4	63.3 ± 0.4	0.07
<b>Area of residence</b>			
Alicante	47 (8.6)	60 (10.7)	0.6
Asturias	238 (43.4)	241 (43.1)	
Barcelona	86 (15.7)	97 (17.4)	
Manresa	24 (4.4)	17 (3.0)	
Sabadell	51 (9.3)	49 (8.8)	
Tenerife	102 (18.6)	95 (17.0)	
<b>Smoking</b>			
Never smokers	101 (18.4)	195 (34.9)	<0.001
Moderate smokers	55 (10.0)	89 (15.9)	
Heavy smokers	392 (71.5)	275 (49.2)	
<b>Average residential THM levels (µg/L)</b>			
Mean ± SE	33.2 ± 1.2	34.2 ± 1.2	0.6
≤26 µg/L	251 (45.8)	272 (48.7)	0.3
>26 µg/L	297 (54.2)	287 (51.3)	
<b>LINE-1%5mC</b>			
Mean ± SE	59.6 ± 2.1	58.8 ± 2.1	0.8
<64 %	456 (83.2)	489 (87.5)	0.05
≥64 %	92 (16.8)	70 (12.5)	

<sup>1</sup>χ<sup>2</sup> test for proportions, t-test for continuous variables comparing cases vs. controls.

the different study areas, both sexes, all the hospitals, and over the 4 y of recruitment (data not shown). By CpG, the 1st and 3rd sites showed higher methylation levels than the 2nd and 4th sites (Supplemental Material, Figure S1, panels C and D). The dendrogram from the unsupervised hierarchical cluster showed 2 clusters. The *k*-means cluster analysis selected a cut-off at 64% between clusters. LINE-1%5mC was classified as low-intermediate (<64%) and high (≥64% methylation). We visually inspected the scatterplots of the first vs. the second components from a principal component analysis of the 4 CpGs, according to different variables to observe potential batch effects (technical and not attributable to covariables). A suggestion of a batch effect was found by the y of interview, which corresponds to time of inclusion and blood collection and was not related to case/control status (Supplemental Material, Figure S2).

LINE-1%5mC levels dichotomized at 64% showed statistically significant differences between cases and controls (Table 1), which remained in logistic regression models adjusted for age, sex, and area (data not shown). Average LINE-1%5mC levels as a continuous variable were slightly different between cases and controls but multivariable analysis showed no statistically significant differences (Table 2). LINE-1%5mC levels were lower among females (both among cases and controls), and were not significantly associated with age and smoking status (Table 2). A statistically significant interaction was found for case-control

status and age (interaction *P* = 0.03). This interaction disappeared after including THM in the models and was not further explored. Methylation levels by area were similar, and adjustment for area instead of hospital as a cluster showed similar results (not shown in tables). Change between the percentiles 25th and 75th of average residential THM levels were associated with increased LINE-1%5mC levels among controls ( $\beta$  = 1.8%) and a negative, not significant association was observed among cases ( $\beta$  = -2.1%) (interaction *P* = 0.02) (Table 2). The negative association among cases was driven by a small group with highest LINE-1%5mC levels (≥64%, *n* = 162) (results not shown). The association with THM levels categorized by the median showed similar trends but there were no statistically significant differences for controls ( $\beta$  = 1.2%), while the association among cases was statistically significant ( $\beta$  = -6.6%) (Table 2). The association between THM levels and LINE-1%5mC is shown graphically in Figure 1. Methylation levels were similar among controls (59%), non-muscle invasive cases (59%) and invasive cases (60%), heterogeneity *P* = 0.3. No differences were observed between muscle invasive (>T2, *n* = 111) and non-invasive (Ta, and T1 tumors, *n* = 437) bladder cancer risk vs. controls.

Odds ratios of UBC were 2.03 (95% CI: 1.27, 3.82) among moderate smokers and 4.48 (95% CI: 3.11, 6.44) among heavy smokers, compared to never smokers, *P*-trend<0.001 (results not shown in tables). Average lifetime THM levels were

**Table 2.** Levels of LINE-1% 5-methylcytosine (LINE-1%5mC) in cases and controls, and  $\beta$  coefficient from  $\beta$  regression of LINE-1%5mC, according to individual characteristics in the Spanish Bladder Cancer/EPICURO (SBC/EPICURO) study

Variable	LINE-1%5mC Mean $\pm$ SE		LINE-1%5mC change <sup>1</sup> (95% CI)		
	Cases	Controls	Cases	Controls	All
<b>Case-control status</b>					
Control					Reference
Case					0.4 (-2.9, 3.6)
<b>Age (y)</b>					
Mean (64 yrs) centered			0.2 (-0.01, 0.3)	-0.1 (-0.2, 0.1)	0.04 (-0.1, 0.1)
interaction <i>P</i>				0.03	
<b>Sex</b>					
Male	59.6 $\pm$ 2.2	59.0 $\pm$ 2.2	Reference	Reference	Reference
Female	59.4 $\pm$ 5.8	57.6 $\pm$ 6.3	-2.2 (-8.7, 4.2)	-7.2 (-12.7, -1.7)	-4.5 (-9.6, 0.5)
interaction <i>P</i>				0.14	
<b>Smoking</b>					
Never	59.6 $\pm$ 4.9	58.5 $\pm$ 3.5	Reference	Reference	Reference
Moderate	59.7 $\pm$ 6.6	59.1 $\pm$ 5.2	-3.4 (-11.3, 4.5)	3.1 (-0.2, 6.5)	1.1 (-2.7, 4.8)
Heavy	59.6 $\pm$ 4.9	58.5 $\pm$ 3.5	-1.6 (-7.7, 4.4)	2.1 (-3.0, 7.1)	0.9 (-3.4, 5.2)
interaction <i>P</i>				0.30	
<b>Average residential THM levels (<math>\mu</math>g/L)</b>					
All ( <i>n</i> = 1107)					
As continuous Change between 25th (10 $\mu$ g/L) and 75th (50 $\mu$ g/L) percentile <sup>2</sup>			-2.1 (-4.4, 0.2)	1.8 (0.1, 3.4)	-0.03 (-1.2, 1.2)
interaction <i>P</i>				0.02	
In categories					
≤26 $\mu$ g/L	59.9 $\pm$ 2.5	59.0 $\pm$ 2.5	Reference	Reference	Reference
>26 $\mu$ g/L	59.1 $\pm$ 2.2	58.8 $\pm$ 2.3	-6.6 (-11.5, -1.7)	1.2 (-3.1, 5.5)	-2.6 (-5.9, 0.7)
interaction <i>P</i>				0.02	

<sup>1</sup>Beta coefficient (in percentages) from robust  $\beta$  regression adjusted for sex, age, y of blood drawing and hospital (as cluster). The All column is additionally adjusted for case-control status.

<sup>2</sup>Based on the fitted curve between LINE-1%5mC and squared root THMs.

**Table 3.** Odds ratio (OR) and 95% confidence interval (CI) of bladder cancer associated with trihalomethane levels and effect modification by LINE-1% 5-methylcytosine levels (LINE-1%5mC) status in the Spanish Bladder Cancer/EPICURO (SBC/EPICURO) study

Average residential THM level ( $\mu\text{g/L}$ )	LINE-1%5mC <64 % (n = 945)		LINE-1%5mC $\geq$ 64 % (n = 162)		All (n = 1107)	
	Cases/ Controls	OR <sup>1</sup> (95% CI)	Cases/ Controls	OR <sup>1</sup> (95% CI)	Cases/ Controls	OR <sup>1</sup> (95% CI)
As continuous						
Change between 25th (10 $\mu\text{g/L}$ ) and 75th (50 $\mu\text{g/L}$ ) percentile <sup>2</sup>	456/489	1.47 (1.12, 1.94)**	92/70	0.62 (0.32, 1.20)	548/559	1.29 (1.00, 1.65)*
interaction <i>P</i> (continuous)						0.03
In categories						
$\leq$ 26 $\mu\text{g/L}$	198/239	Reference	53/33	Reference	251/272	Reference
$>$ 26 $\mu\text{g/L}$	258/250	2.14 (1.39, 3.30)***	39/37	0.80 (0.28, 2.26)	297/287	1.86 (1.25, 2.75)**
interaction <i>P</i> (categories)						0.08

*P* compared with controls in the reference \* $\leq$ 0.05, \*\* $\leq$ 0.01, and \*\*\* $\leq$ 0.001.

<sup>1</sup>OR adjusted for sex, age, area of residence, smoking (never smokers, moderate smokers, heavy smokers).

<sup>2</sup>Based on the fitted curve between LINE-1%5mC and squared root THMs.

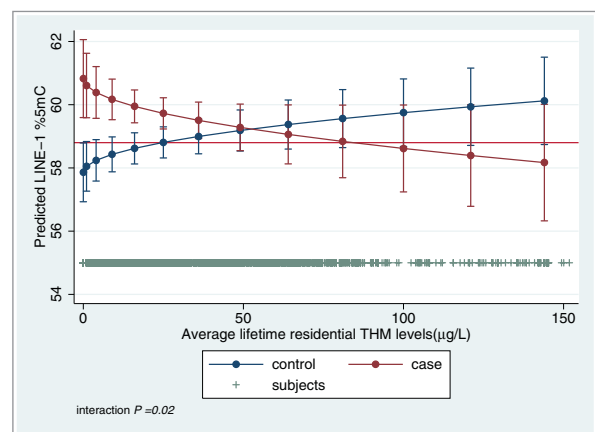
associated with an increased UBC risk in the overall population, with an OR of 1.86 (95% CI: 1.25–2.75) when comparing subjects exposed to THM levels above vs. below the population median (26  $\mu\text{g/L}$ ), *P*-trend = 0.002 (Table 3). The OR of bladder cancer for P75th compared to the P25th of THM levels was 1.29 (95% CI: 1.004, 1.65). Among the interactions tested (THM\*LINE-1%5mC low-intermediate/high methylated subjects, smoking\*THM, and smoking\*LINE-1%5mC low-intermediate/high methylated subjects), only the THM\*LINE-1%5mC was statistically significant (interaction *P* = 0.03). In the less methylated group (<64% LINE-1%5-mC), OR of UBC increased with THM levels whereas THM levels were not associated with risk of UBC in the most methylated subpopulation ( $\geq$ 64% LINE-1%5mC) (Table 3). Alternative analyses using generalized additive models and conditional logistic regression models did not provide a better model fit.

## Discussion

Our results show a slight positive association between THM exposure and LINE-1%5mC levels among controls. A non-statistically significant negative trend was found among cases, driven by a small group with highest LINE-1%5mC levels ( $\geq$ 64%, *n* = 162). The LINE-1 methylation status modified the association between THM levels and UBC risk. The OR of UBC increased with THM levels in the low-intermediate methylated subjects (<64%), whereas the association between THM exposure and UBC risk was not observed among those with the highest LINE-1%5mC levels ( $\geq$ 64%).

DBPs, and specifically some THM and haloacetic acids, had been related to a decrease in DNA methylation levels in mice. Our results differ from the experimental evidence showing lower methylation levels among rats exposed to THM.<sup>11</sup> In our data, THM exposure did not show an association with LINE-1%5mC for all the subjects. The difference with experimental evidence may be partly explained by interspecies differences and non-comparable exposures. In rodent experiments, THMs were

administered at high doses and usually per gavage and methylation sites selected were specific oncogenes.<sup>11</sup> On the contrary, in human populations the drinking water exposure is chronic, at very low doses, and through multiple pathways (ingestion, inhalation, skin absorption) resulting in different pharmacokinetics, and may be altered due to water handling (filtering, fluid mixtures, boiling, cooling, and freezing).<sup>28</sup> Our results are plausible given that one of the mechanisms proposed for THM carcinogenicity is related to cytotoxicity and cell regeneration.<sup>29</sup> This mechanism may affect DNA methylation landmarks because of rapid cell proliferation generating hypermethylated and hypomethylated adapted cell clones,<sup>30</sup> the former being clusters of cells more resilient to environmental toxicants, as seen in controls, and the latter being more susceptible to DNA dysregulation and/or to enter into apoptosis, as seen in cases. However, the observed change of magnitude in LINE-1%5mC levels cannot conclusively associate this observation to THM exposure and/or either could not be biologically relevant in terms of LINE-1 activation.



**Figure 1.** Marginal predictive values of LINE-1%5mC in cases and controls at different lifetime exposures to trihalometananes with 95% confidence intervals.

Changes in LINE-1%5mC levels have been associated with an increased bladder cancer risk in several observational studies in humans. In a larger subset of the current study, a U-shape relationship was observed between LINE-1 and bladder cancer risk.<sup>25</sup> In a bladder cancer case control study conducted in China, non-smokers with LINE-1%5mC levels in the lowest tertile (<81.2%) had an odds ratio of bladder cancer of 1.91 compared to those in the highest tertile ( $\geq 82.52\%$ ).<sup>23</sup> Similarly, in the New Hampshire US bladder cancer study only subjects in the lowest decile of LINE-1%5mC (<74%) compared to the highest decile (>84%) had an odds ratio of bladder cancer of 1.80. This relationship was confounded by gender (lower levels in females), cumulative arsenic exposure and non-invasive cancer versus invasive cancer, but no interaction was found for smoking.<sup>24</sup> In a pooled nested case/control studies of US and Finland<sup>26</sup> an odds ratio of bladder cancer of 1.82 in male smokers was found comparing the lowest quartile ( $\leq 82.9\%$  in US and  $\leq 78.1\%$  in Finnish study) to the quartiles 2 to 4. These results are comparable to our findings, although our population levels are lower than those reported in these studies. LINE-1%5mC levels reported can be summarized as Chinese > US > Finnish-Northern European > Spain-Southern European populations, and differences can be attributed to ethnic differences and specific lifestyle variables such as diet.<sup>23</sup> Our data corroborated that LINE-1%5mC levels were lower in females, but, as in all the previous studies, age and smoking were not consistently related to methylation levels. However, we did not find differences in LINE-1%5mC levels between invasive and non-invasive bladder cancer subjects as observed in other studies.<sup>24</sup> Finally, the less methylated subgroup showed an association between UBC risk and THMs. This is consistent with previous reports in which less methylated subjects, either at LINE-1 or global DNA methylation, were at higher risk of UBC.<sup>23,24,30</sup>

In our study, methylation status was assessed using LINE-1%5mC levels in granulocyte DNA while other studies used lymphocyte, buffy coat and whole blood cells DNA. In our data, granulocyte and lymphocyte LINE-1%5mC levels were highly correlated.<sup>31</sup> Nonetheless, the mean LINE-1%5mC observed (59.2%) was lower compared to previous reports in other ethnic groups and DNA cell sources.<sup>23,24,26</sup> However, even if a proportion of the differences among studies may be attributed to different cell epigenetic landmarks or the ethnic origin of subjects, we cannot disregard other unmeasured technical aspects such as different performance of the sequencing primers in our population.<sup>15</sup> It is important to acknowledge that LINE-1, and among them LIHs, only provide a limited view of the methylation status and do not measure whole DNA methylation, which may be differently altered under the same exposure conditions.

The sample size, the laboratory techniques, and the statistical analyses constitute important strengths in our study. Data were derived from a large hospital based case-control study with a rigorous exposure assessment of THM and information of potential confounders. High quality gold-standard quantifiable laboratory techniques, pyrosequencing, were used to estimate methylation levels in granulocyte DNA LINE-1%5mC. Robust statistical strategies including  $\beta$

regression for proportions and cluster analysis were used to reduce bias in the analyses. The  $\beta$  regression deals with outcomes bounded between 0 and 1, which may present heteroskedasticity and asymmetries in the distribution.<sup>32</sup> Methylation status of a CpG is dichotomous (methylated vs. unmethylated), but it is quantified in the laboratory as a proportion (% methylated CpGs) making this regression suitable for analyzing LINE-1%5mC levels as the outcome. When LINE-1%5mC was used as a confounder, the variable was categorized to avoid non-linear trends of the distribution or complicated transformations, which obscure the interpretation of the data. In literature, methylation is either categorized using arbitrary cut-offs such as <20%, 20–80%, and >80%,<sup>33,34</sup> or using population specific cut-off based on percentiles of controls.<sup>23,24,26</sup> In our study, only few subjects ( $n = 9$ ) were above 80%. Instead of selecting a percentile, the selected approach was using a data-driven unsupervised cluster approach. This strategy separated naturally clustered subpopulations for the analysis, reducing the chance to bias the analysis because of an arbitrary selected cut-off.

A limitation of this study is the inability to establish temporality of the outcomes evaluated in relation with UBC development. When evaluating LINE-1%5mC as an outcome of THM exposure, we assume that methylation levels are the consequence of cumulative lifetime exposure to THM. However, given that we only have one spot sample, we cannot exclude reverse causality (methylation levels due to cancer or other disease, instead of being a consequence of THM exposure). This limitation is inherent to the case-control design in long latency diseases such as cancer. Nevertheless, as no differences were found between different cancer stages and most of the study cases were at non-invasive stages; changes in methylation levels due to a cancer systemic response are very unlikely. Given that most of the historical exposure data is based on questionnaires, we cannot exclude recall bias, expected to be non-differential. A categorical variable that assessed the reliability of the data provided though the questionnaire, as evaluated by the interviewer, was included. Those subjects with unreliable interviews were excluded from analyses. In addition, only those subjects with more than 70% of yearly historical THM estimates in the exposure time window were analyzed to reduce misclassification. However, measurement error in THM level estimates may still be present.

Other limitations are technical due to the laboratory methods used or potential batch effects. Pyrosequencing had some important limitations: first, as we used predesigned primers and the measured point is methylated vs unmethylated, we cannot rule out polymorphisms in the DNA strands (i.e., presence of SNP sites) that may affect performance of the technique.<sup>35</sup> Second, pyrosequencing detection limit per CpG site is 5%, thus measured lower differences may be unreliable.<sup>36</sup> In addition, the most methylated and the least methylated extremes are not easily detected and even 100% methylated blanks may only produce signals around  $\sim 80\%$  methylation.<sup>26</sup> On the other hand, we observed variation that was not explained by any covariable tested. Year of interview (blood collection) and hospital were used as proxies of

technical variability (time of storage and/or differences in blood sample handling). These factors may affect biomarker performance due to DNA degradation by sample handling, specially freezing delays and blood conservation. The protocols, storage conditions and sample management were not changed during the collection, so this only can be explained due to unmeasured individual sample conditions, presence of SNPs or actual LINE-1%5mC levels differences because of other unmeasured confounders. We included the proxies as confounders in adjusted models when modeling LINE-1%5mC.<sup>37</sup> However, if this was actually a batch effect, we cannot assess the extent to which it may alter our results.

In summary, results suggest a positive association between LINE-1%5mC and THM levels among controls, and LINE-1%5mC status may modify the association between UBC risk and THM exposure. Because samples were collected post-diagnosis and the small group in the highest methylation levels, reverse causality and chance cannot be ruled out and future studies are warranted to confirm these results.

## Materials and Methods

### Study design and participants

Data were obtained from participants in the Spanish Bladder Cancer (SBC)/EPICURO study, a multicenter, hospital-based, case-control study conducted from June 1998 to December 2001 in Spain. Subjects were recruited at 18 hospitals in Alicante, Asturias, Barcelona, Manresa, Sabadell and Tenerife. Cases were patients aged 20–80 y living in the geographic catchment areas of the participating hospitals, with histologically confirmed primary bladder cancer based on the 1999 system of the WHO and the International Society of Urological Pathology.<sup>38</sup> Controls were inpatients of the participating hospitals selected to have diagnoses unrelated with the main UBC risk factor (smoking). Controls were individually matched to cases by sex, age ( $\pm 5$  y), and geographic area of residence. The study was approved by the Institutional Review Boards of the participating centers and participants provided informed consent before participation. A total of 1457 eligible cases and 1465 eligible controls were identified. Participation rates were 84% for cases ( $n = 1219$ ) and 87% for controls ( $n = 1271$ ).

### Personal interview

Trained interviewers administered a comprehensive computer-assisted personal questionnaire to participants during their hospitalization. Collected data included socio-demographic characteristics, smoking habits, family history of cancer, medical, occupational, and residential histories from birth to the time of interview (all residences of at least 1 y), average daily consumption of water and water-based fluids (e.g., coffee and tea), and average frequency and duration of showering and bathing. For all dwellings, the y lived in a residence, full address, city, province, region, and country were requested. Zip codes in Barcelona were obtained through the address to estimate the water purveyors' areas.

### THM exposure assessment

Exposure assessment has been described elsewhere.<sup>39,40</sup> In brief, historical water source, treatment and quality data were requested from water companies and local authorities. In the absence of historical THM measurements, water source and current levels were used to estimate the levels in the past. Annual THM levels were modeled back to 1920 assuming unchanged level for a constant water source. Levels changed proportionally to surface water percentage if water source changed. A zero THM level was assumed for those y before chlorination started. Residential THM levels per subject were calculated by merging individual and municipal databases by y and municipality of residence. A time-weighted average level in all residences since 15 y old to the time of interview was calculated for all subjects as previously described.<sup>39</sup> We refer to this time period as the study 'exposure window'.

### Quantification of LINE-1%5mC levels in granulocytes

A blood sample or a saliva sample was requested from all the subjects in the main case-control study for genetic analyses.<sup>27</sup> Buffy coat, whole blood and purified lymphocytes have been used in previous studies. For this study, the granulocyte fraction was used and results were compared against purified lymphocytes.<sup>31</sup> Granulocyte DNA was extracted using standard methods. DNA was treated with bisulfite using EZ-96 DNA METHYLATION-GOLD™ KIT (Zymo Research, Irvine, CA, USA) to transform unmethylated cytosines into uracil, leaving the methylated cytosine unchanged. Polymerase chain reaction amplified the bisulfite-treated DNA sequence using a group of modified 3 DNA primers, as previously described.<sup>41</sup> This approach measures%5mC levels of the 4 CpG sites located immediately after the sequencing primer at LINE-1 promoter region, between nucleotides 318–331, within the human LINE-1 transposon (L1Hs) DNA 5' UTR, GenBank accession number X58075.<sup>31,42</sup> Methylation levels were quantified by pyrosequencing using the PyroMark™ Q24 System (QIAGEN, Valencia, CA, USA). The subject LINE-1%5mC is the percentage of methylated cytosines over the sum of methylated and unmethylated cytosines. Levels of methylation of the 4 CpG sites were averaged to a single LINE-1%5mC level. A total of 302 randomly selected samples (174 cases and 128 controls) were run in duplicate. The intra-assay coefficient of variation was 4.53%, technical causes of variation has been discussed elsewhere.<sup>31</sup> The intraclass correlation coefficient (ICC) of LINE-1 duplicates among samples with less than 5% coefficient of variation ( $n = 216$ , 72% of duplicates) and adjusting by technical duplicates (using a 2 way random effects ANOVA test) was 0.25 (95% CI: 0.12, 0.37) within subjects and 0.40 (95% CI: 0.22, 0.54) between subjects. We used the mean of the duplicates in the analysis. As we were unable to assess the full set of measurements by duplicate, we imputed the coefficient of variation for all the subjects included in the analyses. We compared a model including this variable and without it and estimates were consistent. We thus decided to pool all the available samples for the analyses.

## Statistical methods

Bivariate analyses included  $\chi^2$  to compare proportions, *t*-test to compare mean differences, and proportion test to compare proportion differences. Analyses of average methylation levels (%) of several CpG sites in the LINE-1 repetitive sequences usually show a high peaked distribution which obscures underlying modes due to differential methylation between CpG sites.<sup>33</sup> To observe differentially methylated subpopulations, LINE-1%5mC distribution was classified using an unsupervised hierarchical cluster dendrogram to observe potential clustering and a *k*-means cluster approach was used to select a cut-off. Principal component analysis using the 4 CpGs was used to test potential batch effects (defined as technical sources of variation that have been added to the samples during handling) related to blood sample manipulation (i.e., *y* of blood drawing or hospital).

We used  $\beta$  regression to estimate differences in levels of LINE-1%5mC by different covariates in multivariable analyses, to deal with a variable bounded between 0 and 1.<sup>32</sup> The relationship between THM and LINE-1 levels was assessed separately for cases and controls, and for all combined by including an interaction term in the model. Beta coefficients were multiplied by 100 to interpret results as percentages. Models were adjusted by sex, age, *y* of blood drawing and hospital as cluster. Standard errors were robust adjusted (Huber-White Sandwich method). To obtain a linear relationship between average lifetime residential THM levels and LINE-1%5mC, THM concentration ( $\mu\text{g/L}$ ) was square root transformed. Based on this model, we derived the change in LINE-1%5mC between percentiles P25th to P75th, in order to interpret results in the original (untransformed) scale. We used generalized additive models to observe graphically the dose-response relationship between THM levels and LINE-1%5mC and identify eventual non-linear associations.

We applied logistic regression to assess the UBC risk associated with THM exposure adjusted for age, sex, geographic area of residence and smoking. Smoking was classified in three groups: never smokers, moderate smokers (i.e., former or current smokers of less than 20 pack-y), and heavy smokers (i.e., more or equal to 20 pack-y). Stratified analyses and interaction terms were used to test the modification of the UBC risk associated with THM exposure by LINE-1%5mC status. Different ordered and multinomial logistic regression approaches were considered for sensitivity analyses. Stereotype logistic regression, a class of ordered logistic regression,<sup>43</sup> was used to evaluate risk of non-muscle-invasive and muscle-invasive bladder cancer risk vs. controls for different covariates. In order to interpret results in the original (untransformed) scale, we derived the change in UBC

risk between percentiles 25th to 75th of THMs in the model using the square root of THMs.

Interaction was evaluated by introducing the product of the variables. Likelihood ratio tests comparing models with and without multiplicative interaction terms were used to calculate the interaction *P*. In robust models, the Wald test was used to approximate the interaction *P*. Marginal effects were calculated using the predicted probabilities of the events at different cut-offs for continuous or categorical by continuous interactions holding constant the other equation variables.<sup>44,45</sup>

To rule out bias due to averaging the 4 tested CpGs, 3 different variables were used in alternative analyses instead of average LINE-1%5mC: 1) the first CpG; 2) the first and second component from a principal component analysis of the 4 CpG sites; and 3) a mixed effects model using the 4 CpG as the outcome, using the nucleotide order as a random effect (data not shown). Results were considered statistically significant if the nominal *P* was <0.05. All tests were 2-tailed. Statistical analyses were performed using Stata Statistical Software; release 12.1 (StataCorp LP, College Station, TX), and the Befit module.<sup>46</sup> Results were similar between models so we used the average of CpGs and a  $\beta$  model for our analyses.

## Disclosure of Potential Conflicts of Interest

KPC owns KP Cantor Environmental LLC, Silver Spring, MD, USA.

## Funding

This study was partially supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics (Contract NCI NO2-CP-11015); the Spanish Health Ministry (Fondo de Investigaciones Sanitarias-FIS, Instituto de Salud Carlos III, Spain 00/0745, ISIII-GO3/174, PI080533, PI051436, PI061614, PI09-02102, and PI11/00226) and the European Union (BMH4-98-3243); Red Temática de Investigación Cooperativa en Cáncer- RD12/0036/0050-RTICC; USA-NIH-RO1-CA089715; a postdoctoral fellowship awarded to AFSA from the Fundación Científica de la AECC; Fundació Marató TV3. The work was partially supported by the Association for International Cancer Research (AICR, #09-0780, including a PhD scholarship awarded to S.M.T.). The current analyses were supported by a Colciencias PhD Scholarship, Colombia (Grant: 529/2011 to L. A.S.). This work was also supported by grants from the Instituto de Salud Carlos III FEDER, (PI11/00226).

## References

1. Garcia-Villanova RJ, Mera BB, Gonzalez Paramas AM, Hernandez Hierro JM, Albajar RA, Toruno Fonseca IM, Blanca Mera B, González Paramás AM, Hernández Hierro JM, Ardanuy Albajar R, et al. A multi-year survey of organic disinfection by-products in drinking waters of Castilla y Leon, Spain. The need and difficulty to comply with the legal limit of 2009. *J Env Monit* 2010; 12:200-7. PMID: 20082014; <http://dx.doi.org/10.1039/B911269C>
2. Villanueva CM, Cantor KP, Grimalt JO, Malats N, Silverman D, Tardon A, Garcia-Closas R, Serra C, Carrato A, Castaño-Vinyals G, et al. Bladder cancer and exposure to water disinfection by-products through ingestion, bathing, showering, and swimming in pools. *Am J Epidemiol* 2007; 165:148-56; PMID:17079692; <http://dx.doi.org/10.1093/aje/kwj364>
3. Villanueva CM, Cantor KP, Cordier S, Jaakkola JJK, King WD, Lynch CF, Porru S, Kogevinas M. Disinfection Byproducts and Bladder Cancer: A Pooled Analysis. *Epidemiology* 2004; 15:357-67; PMID:15097021; <http://dx.doi.org/10.1097/01.ede.0000121380.02594.fc>
4. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: 2013.
5. Samanic C, Kogevinas M, Dosemeci M, Malats N, Real FX, Garcia-Closas M, Serra C, Carrato A, Garcia-Closas R, Sala M, et al. Smoking and Bladder Cancer

- in Spain: Effects of Tobacco Type, Timing, Environmental Tobacco Smoke, and Gender. *Cancer Epidemiol Biomarkers Prev* 2006; 15:1348-54; PMID:16835335; <http://dx.doi.org/10.1158/1055-9965.EPI-06-0021>
6. Landi S, Hanley NM, Warren SH, Pegram Ra, DeMarini DM. Induction of genetic damage in human lymphocytes and mutations in Salmonella by trihalomethanes: role of red blood cells and GSTT1-1 polymorphism. *Mutagenesis* 1999; 14:479-82; PMID:10473651; <http://dx.doi.org/10.1093/mutage/14.5.479>
  7. Landi S, Naccarati A, Ross MK, Hanley NM, Dailey L, Devlin RB, Vasquez M, Pegram Ra, DeMarini DM. Induction of DNA strand breaks by trihalomethanes in primary human lung epithelial cells. *Mutat Res* 2003; 538:41-50; PMID:12834753; [http://dx.doi.org/10.1016/S1383-5718\(03\)00086-X](http://dx.doi.org/10.1016/S1383-5718(03)00086-X)
  8. Ross MK, Pegram Ra. Glutathione transferase theta 1-1-dependent metabolism of the water disinfection byproduct bromodichloromethane. *Chem Res Toxicol* 2003; 16:216-26; PMID:12588193; <http://dx.doi.org/10.1021/tx0200820>
  9. Ross MK, Pegram Ra. In vitro biotransformation and genotoxicity of the drinking water disinfection byproduct bromodichloromethane: DNA binding mediated by glutathione transferase theta 1-1. *Toxicol Appl Pharmacol* 2004; 195:166-81; PMID:14998683; <http://dx.doi.org/10.1016/j.taap.2003.11.019>
  10. Pegram Ra, Andersen ME, Warren SH, Ross TM, Claxton LD. Glutathione S-transferase-mediated mutagenicity of trihalomethanes in Salmonella typhimurium: contrasting results with bromodichloromethane off chloroform. *Toxicol Appl Pharmacol* 1997; 144:183-8; PMID:9169083; <http://dx.doi.org/10.1006/taap.1997.8123>
  11. Coffin JC, Ge R, Yang S, Kramer PM, Tao L, Pereira MA. Effect of Trihalomethanes on Cell Proliferation and DNA Methylation in Female B6C3F1 Mouse Liver. *Toxicol Sci* 2000; 58:243-52; PMID:11099637; <http://dx.doi.org/10.1093/toxsci/58.2.243>
  12. Pereira MA, Kramer PM, Conran PB, Tao L. Effect of chloroform on dichloroacetic acid and trichloroacetic acid-induced hypomethylation and expression of the c-myc gene and on their promotion of liver and kidney tumors in mice. *Carcinogenesis* 2001; 22:1511-9; PMID:11532874; <http://dx.doi.org/10.1093/carcin/22.9.1511>
  13. Tao L, Wang W, Li L, Kramer PK, Pereira MA. DNA Hypomethylation Induced by Drinking Water Disinfection By-Products in Mouse and Rat Kidney. *Toxicol Sci* 2005; 87:344-52; PMID:16014735; <http://dx.doi.org/10.1093/toxsci/kfi257>
  14. Tao L, Wang W, Li L, Kramer PM, Pereira Ma. Effect of Dibromoacetic Acid on DNA Methylation, Glycogen Accumulation, and Peroxisome Proliferation in Mouse and Rat Liver. *Toxicol Sci* 2004; 82:62-9; PMID:15342954; <http://dx.doi.org/10.1093/toxsci/kfh266>
  15. Yang AS, Estéicio MRH, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucl Acids Res* 2004; 32:e38; PMID:14973332; <http://dx.doi.org/10.1093/nar/gnh032>
  16. Beck CR, Garcia-Perez JL, Badge RM, Moran J V. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet* 2011; 12:187-215; PMID:21801021; <http://dx.doi.org/10.1146/annurev-genom-082509-141802>
  17. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 2007; 41:331-68; PMID:18076328; <http://dx.doi.org/10.1146/annurev.genet.40.110405.090448>
  18. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 2009; 10:691-703; PMID:19763152; <http://dx.doi.org/10.1038/nrg2640>
  19. Hancks DC, Kazazian HH. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 2012; 22:191-203; PMID:22406018; <http://dx.doi.org/10.1016/j.gde.2012.02.006>
  20. Kitkumthorn N, Mutirangura A. Long interspersed nuclear element-1 hypomethylation in cancer: biology and clinical applications. *Clin Epigenet* 2011; 2:315-30; <http://dx.doi.org/10.1007/s13148-011-0032-8>
  21. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997; 13:335-40; PMID:9260521; [http://dx.doi.org/10.1016/S0168-9525\(97\)01181-5](http://dx.doi.org/10.1016/S0168-9525(97)01181-5)
  22. Terry MB, Delgado-Cruzata L, Vin-Raviv N, Wu HC, Santella RM. DNA methylation in white blood cells: Association with risk factors in epidemiologic studies. *Epigenetics* 2011; 6:828-37; PMID:21636973; <http://dx.doi.org/10.4161/epi.6.7.16500>
  23. Cash HL, Tao L, Yuan J-M, Marsit CJ, Houseman EA, Xiang Y-B, Gao Y, Nelson HH, Kelsey KT. LINE-1 hypomethylation is associated with bladder cancer risk among nonsmoking Chinese. *Int J Cancer* 2012; 130:1151-9; PMID:21445976; <http://dx.doi.org/10.1002/ijc.26098>
  24. 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; 16:1682-9; PMID:20179218; <http://dx.doi.org/10.1158/1078-0432.CCR-09-2983>
  25. Tajuddin SM, Amaral AFS, Fernández AF, Chanock S, Silverman DT, Tardón A, Carrato A, García-Closas M, Jackson BP, Toráño EG, et al. LINE-1 methylation in leukocyte DNA, interaction with phosphatidylethanolamine N-methyltransferase variants and bladder cancer risk. *Br J Cancer* 2014; 110:2123-30; PMID:24595004; <http://dx.doi.org/10.1038/bjc.2014.67>
  26. Andreotti G, Karami S, Pfeiffer RM, Hurwitz L, Liao LM, Weinstein SJ, Albanes D, Virtamo J, Silverman DT, Rothman N, et al. LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. *Epigenetics* 2014; 9:404-15; PMID:24316677; <http://dx.doi.org/10.4161/epi.27386>
  27. García-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, Tardón A, Serra C, Carrato A, García-Closas R, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* 2005; 366:649-59; [http://dx.doi.org/10.1016/S0140-6736\(05\)67137-1](http://dx.doi.org/10.1016/S0140-6736(05)67137-1)
  28. Chowdhury S, Rodríguez MJ, Serodes J. Model development for predicting changes in DBP exposure concentrations during indoor handling of tap water. *Sci Total Env* 2010; 408:4733-43; <http://dx.doi.org/10.1016/j.scitotenv.2010.07.006>
  29. DeAngelo AB, Geter DR, Rosenberg DW, Crary CK, George MH. The induction of aberrant crypt foci (ACF) in the colons of rats by trihalomethanes administered in the drinking water. *Cancer Lett* 2002; 187:25-31; PMID:12359347; [http://dx.doi.org/10.1016/S0304-3835\(02\)00356-7](http://dx.doi.org/10.1016/S0304-3835(02)00356-7)
  30. Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, Silverman D, García-Closas R, Chanock S, Tardón A, Serra C, et al. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol* 2008; 9:359-66; PMID:18339581; [http://dx.doi.org/10.1016/S1470-2045\(08\)70038-X](http://dx.doi.org/10.1016/S1470-2045(08)70038-X)
  31. Tajuddin SM, Amaral AFS, Fernández AF, Rodríguez-Rodero S, Rodríguez RM, Moore LE, Tardón A, Carrato A, García-Closas M, Silverman DT, et al. Genetic and non-genetic predictors of LINE-1 methylation in leukocyte DNA. *Env Heal Perspect* 2013; 121:650-6; PMID: 23552396; <http://dx.doi.org/10.1289/ehp.1206068>
  32. Ferrari SLP, Cribari-Neto F. Beta regression for modeling rates and proportions. *J Appl Stat* 2004; 31:799-815; <http://dx.doi.org/10.1080/0266476042000214501>
  33. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010; 11:587; PMID:21118553; <http://dx.doi.org/10.1186/1471-2105-11-587>
  34. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox T V, Davies R, Down TA, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006; 38:1378-85; PMID:17072317; <http://dx.doi.org/10.1038/ng1909>
  35. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protoc* 2007; 2:2265-75; PMID:17853883; <http://dx.doi.org/10.1038/nprot.2007.314>
  36. Mikeska T, Felsberg J, Hewitt CA, Dobrovic A. Analyzing DNA methylation using bisulphite pyrosequencing. *Methods Mol Biol* 2011; 791:33-53; PMID:21913070; [http://dx.doi.org/10.1007/978-1-61779-316-5\\_4](http://dx.doi.org/10.1007/978-1-61779-316-5_4)
  37. Goldinger A, Henders AK, McRae AF, Martin NG, Gibson G, Montgomery GW, Visscher PM, Powell JE. Genetic and nongenetic variation revealed for the principal components of human gene expression. *Genetics* 2013; 195:1117-28; PMID:24026092; <http://dx.doi.org/10.1534/genetics.113.153221>
  38. Epstein JI, Amin MB, Reuter VR, Mostofi FK. The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder. Bladder Consensus Conference Committee. *Am J Surg Pathol* 1998; 22:1435-48; PMID:9850170; <http://dx.doi.org/10.1097/00000478-199812000-00001>
  39. Villanueva CM, Cantor KP, Grimalt JO, Castañón-Vinyals G, Malats N, Silverman DT, Tardón A, Garcia-Closas R, Serra C, Carrato A, et al. Assessment of lifetime exposure to trihalomethanes through different routes. *Occup Env Med* 2006; 63:273-7; PMID:16556748; <http://dx.doi.org/10.1136/oem.2005.023069>
  40. Villanueva CM, Kogevinas M, Grimalt JO. Haloacetic acids and trihalomethanes in finished drinking waters from heterogeneous sources. *Water Res* 2003; 37:953-8; PMID:12531279; [http://dx.doi.org/10.1016/S0043-1354\(02\)00411-6](http://dx.doi.org/10.1016/S0043-1354(02)00411-6)
  41. Estéicio MRH, Gharibyan V, Shen L, Ibrahim AEK, Doshi K, He R, Jelinek J, Yang AS, Yan PS, Huang TH-M, et al. LINE-1 hypomethylation in cancer is highly variable and inversely correlated with microsatellite instability. *PLoS One* 2007; 2:e399; PMID:17476321; <http://dx.doi.org/10.1371/journal.pone.0000399>
  42. National Library of Medicine. GenBank [Internet]. 2013 ; [cited 2013 Mar 13]; Available from: <http://www.ncbi.nlm.nih.gov/genbank/>
  43. Anderson JA. Regression and Ordered Categorical Variables. *J R Stat Soc Ser B Stat Methodol* 1984; 46:1-30.
  44. Han SS, Rosenberg PS, Garcia-Closas M, Figueroa JD, Silverman D, Chanock SJ, Rothman N, Chatterjee N. Likelihood ratio test for detecting gene (G)-environment (E) interactions under an additive risk model exploiting G-E independence for case-control data. *Am J Epidemiol* 2012; 176:1060-7; PMID:23118105; <http://dx.doi.org/10.1093/aje/kws166>
  45. Buis ML. Stata tip 87: Interpretation of interactions in non-linear models. *Stata J* 2010; 10:305-8.
  46. Buis ML, Cox N, Jenkins SP. BETAFIT: Stata module to fit a two-parameter beta distribution. 2012.