

Quantification of Unmethylated Alu (QUALu): a tool to assess global hypomethylation in routine clinical samples

Raquel Buj^{1,2}, Izaskun Mallona^{1,2}, Anna Díez-Villanueva^{1,2}, Víctor Barrera¹, Dídac Mauricio^{2,3,4}, Manel Puig-Domingo^{2,3,4}, Jordi L. Reverter^{2,3}, Xavier Matias-Guiu⁵, Daniel Azuara⁶, Jose L. Ramírez^{2,7}, Sergio Alonso^{1,2}, Rafael Rosell^{2,7}, Gabriel Capellà⁶, Manuel Perucho^{1,2,8}, Mercedes Robledo^{9,10}, Miguel A. Peinado^{1,2} and Mireia Jordà^{1,2}

¹ Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Badalona, Barcelona, Spain

² Germans Trias i Pujol Health Sciences Research Institute (IGTP), Badalona, Barcelona, Spain

³ Department of Endocrinology and Nutrition, University Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain

⁴ ISCIII Center for Biomedical Research on Diabetes and Metabolic Associated Diseases (CIBERDEM), Madrid, Spain

⁵ Department of Pathology and Molecular Genetics, University Hospital Arnau de Vilanova and University of Lleida, Biomedical Research Institute of Lleida (IRBLLEIDA), Lleida, Spain

⁶ Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain

⁷ Catalan Institute of Oncology (ICO), Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain

⁸ Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

⁹ Hereditary Endocrine Cancer Group, Spanish National Cancer Research Center (CNIO), Madrid, Spain

¹⁰ ISCIII Center for Biomedical Research on Rare Diseases (CIBERER), Madrid, Spain

Correspondence to: Mireia Jordà, **email:** mjorda@imppc.org

Miguel A. Peinado, **email:** map@imppc.org

Keywords: DNA hypomethylation, Alu repeats, human cancer, biomarker, routine clinical biospecimens

Received: November 23, 2015

Accepted: January 25, 2016

Published: February 07, 2016

ABSTRACT

Hypomethylation of DNA is a hallmark of cancer and its analysis as tumor biomarker has been proposed, but its determination in clinical settings is hampered by lack of standardized methodologies. Here, we present QUALu (Quantification of Unmethylated Alu), a new technique to estimate the Percentage of UnMethylated Alu (PUMA) as a surrogate for global hypomethylation.

QUALu consists in the measurement by qPCR of Alu repeats after digestion of genomic DNA with isoschizomers with differential sensitivity to DNA methylation. QUALu performance has been evaluated for reproducibility, trueness and specificity, and validated by deep sequencing. As a proof of use, QUALu has been applied to a broad variety of pathological examination specimens covering five cancer types.

Major findings of the preliminary application of QUALu to clinical samples include: (1) all normal tissues displayed similar PUMA; (2) tumors showed variable PUMA with the highest levels in lung and colon and the lowest in thyroid cancer; (3) stools from colon cancer patients presented higher PUMA than those from control individuals; (4) lung squamous cell carcinomas showed higher PUMA than lung adenocarcinomas, and an increasing hypomethylation trend associated with smoking habits.

In conclusion, QUALu is a simple and robust method to determine Alu hypomethylation in human biospecimens and may be easily implemented in research and clinical settings.

INTRODUCTION

Extensive evidence describes cancer as a combination of genetic and epigenetic alterations which

cooperate at every step of the tumor progression (reviewed in [1]). DNA methylation is the most well-characterized epigenetic mark in mammals and consists in the covalent addition of a methyl group to the cytosine located

within the CpG dinucleotide. It is frequently associated with silenced chromatin and transcriptional repression (reviewed in [2-3]). Among all the epigenetic alterations that delineate cancer genomes, loss of global DNA methylation has been considered a hallmark. Numerous works have demonstrated that DNA hypomethylation is an early and sustained event in tumorigenesis. Besides, it promotes a permissive landscape for cancer development and progression by encouraging chromosomal instability, imprinting loss, aberrant gene expression and transposon activation (reviewed in [2-3]). More importantly, it has been reported a strong association between the degree of DNA hypomethylation and the tumor grade and stage, which has attracted great interest for its potential clinical value, not only in cancer diagnosis and prognosis [4-9], but also as a marker of cancer risk [9-12].

A wide variety of techniques have been designed to measure global DNA methylation, some of which quantify the overall levels of 5-methylcytosine in the genome compared with unmethylated cytosines (e.g. HPLC, immunochemical assay, etc.), while others assess the methylation levels of specific genome compartments (reviewed in [9, 13]). Among the second group, the most widely used methods are those based on repeat elements, as they exhibit a high copy number and are widespread throughout the human genome. Nevertheless, none of them has been established in the clinical practice due in part to technical, economical and time shortcomings, which preclude a standardized alternative.

Here we present a new method, Quantification of Unmethylated Alu (QUAlu), which uses Alu repeats as surrogate reporter of global DNA methylation. Alu repeats are primate-specific transposable elements that belong to the Short Interspersed Elements (SINEs) family and represent the most abundant class of repetitive sequences in the human genome (1.1 million copies per haploid genome) [14]. Alu elements contain up to 25% of the overall CpG sites in the genome (Table 1) and are highly methylated in somatic tissues. Interestingly, they are located in gene-rich regions [15].

QUAlu is a simple and rapid method based on the digestion of genomic DNA with the methylation-sensitive and insensitive isoschizomers *HpaII/MspI*, the ligation of an adaptor and a qPCR using primers specific for the Alu consensus sequence. We have applied this technique to a broad variety of pathological examination samples including fresh frozen tissues, formalin-fixed paraffin-embedded (FFPE) sections, fine-needle aspiration biopsies (FNAB), stools and liquid biopsies. Our preliminary results underscore the potential clinical utility of the assessment of unmethylated Alu elements by QUAlu.

RESULTS

QUAlu design and technical evaluation

According to the reference human assembly hg19, there are over 28 millions of CpG dinucleotides in the human genome and more than half are located within repeat elements, being Alu elements those containing the highest fraction, namely 25.4% (Table 1). Therefore, we selected Alu elements as the most adequate surrogate reporter of global methylation and developed QUAlu technique, a method to identify unmethylated Alu repeats which shares the quantitative nature of the related technique LUMA [16] and the specificity of QUMA [17]. Fundamentals of QUAlu assay are outlined in Figure 1 and Supplementary Figure S1. QUAlu is based on the different methylation sensitivity of the isoschizomers *HpaII/MspI* (see Materials and Methods and Supplementary Material), whose recognition site is C/CGG, located in the Alu consensus sequence AACCCGG present in 14.4% of Alu elements (Table 1). Noteworthy, analysis of whole genome bisulfite sequencing data [18] showed that the *HpaII/MspI* sites embedded in CpG islands may be used as reporters of the overall methylation of these genomic elements [19]. In this regard, we also confirmed that this postulate may be also applied to Alu repeats: more than 90% of the *HpaII/MspI* sites within the Alu consensus sequence AACCCGG showed concordant methylation levels with the whole Alu sequence (Supplementary Table S4). To determine the virtual representativeness of QUAlu (QUAluome), an electronic qPCR simulation was performed showing a theoretical coverage of 155 878 Alu elements (see Supplementary Material), which corresponded to the 13.65% of the Aluome, with a bias toward amplification of young subfamilies (Supplementary Figure S2).

To assess the linearity of QUAlu, different starting amounts of HCT116 genomic DNA, ranging from 0.3 to 80 ng, were used. As it can be observed in Figure 2A, all quantifications showed excellent linearity ($R^2 > 0.98$ in all cases). Moreover, similar percentages of unmethylated Alu elements were obtained, being the overall average 8.8 ± 2.2 (Supplementary Figure S3A). The same assay was done with clinical samples from normal tissues of lung, colon and thyroid, obtaining an excellent linearity in all cases (Supplementary Table S5).

The inter-assay repeatability was assessed by analyzing the same HCT116 genomic DNA in 39 independent QUAlu assays. The mean of all the analyses was 9.9% and the standard deviation ± 1.6 . Furthermore, replicates of tumor and normal clinical samples were measured in different plates confirming that the technique is reproducible ($R^2 = 0.981$) (Figure 2B). Importantly, the feasibility of QUAlu in samples containing partially degraded DNA was verified (Supplementary Material and

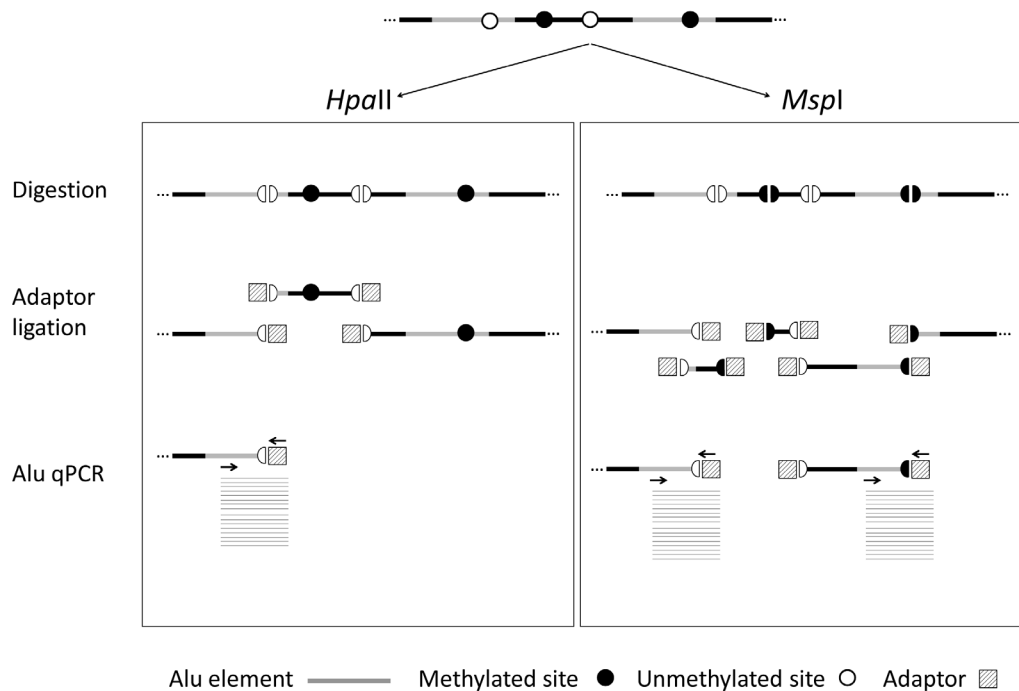


Figure 1: QUAAlu technique diagram. Genomic DNA is digested using *HpaII* and *MspI* isoschizomers (DNA methylation sensitive and insensitive, respectively), ligated to a synthetic adaptor, and Alu elements are specifically amplified by qPCR in two separated reactions. The ratio between the two reactions gives the percentage of unmethylated Alu elements (PUMA). DNA normalization is performed by parallel amplification of L1PA.

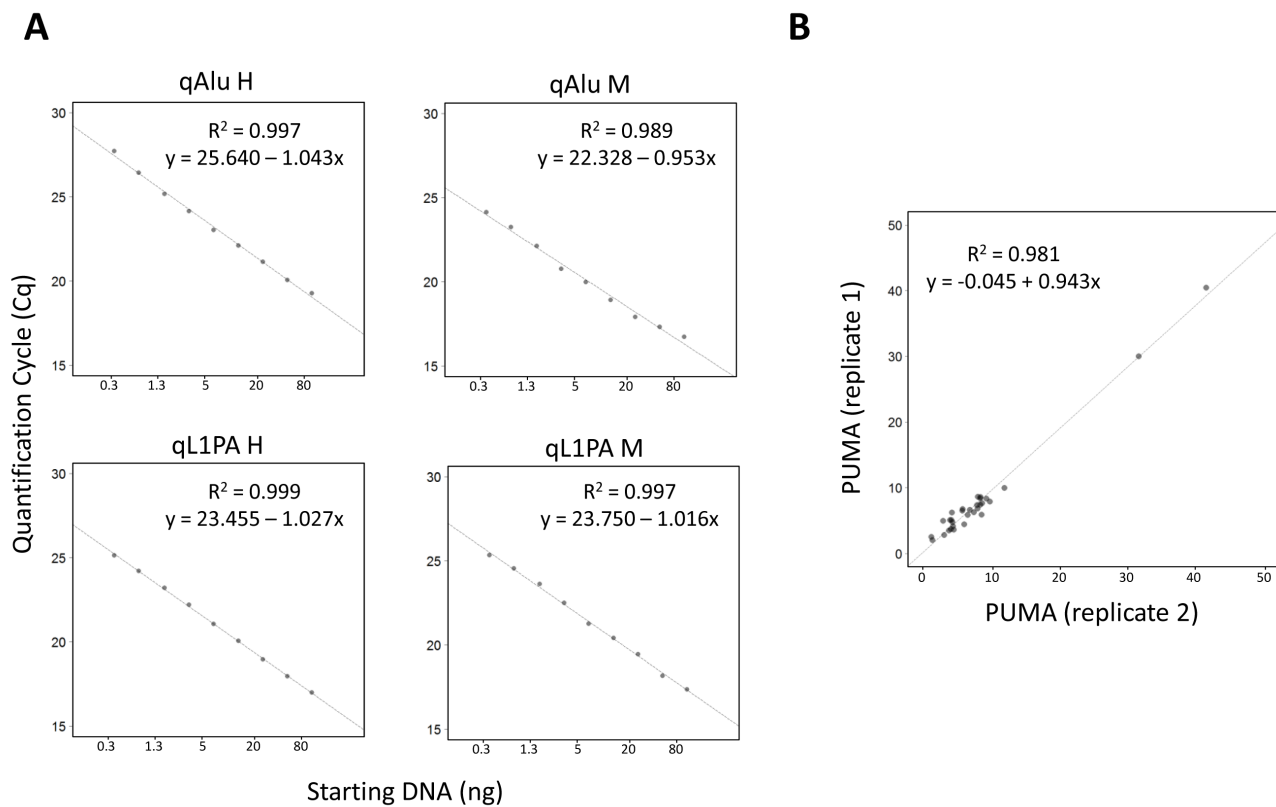


Figure 2: Evaluation of QUAAlu technique. A. Standard curves showing the linear range of the different qPCRs performed in a QUAAlu assay. HCT116 genomic DNA amounts ranging from 0.3 to 80 ng were used. B. Correlation of the percentage of unmethylated Alu elements (PUMA) determined by QUAAlu in two independent experiments.

Sequence	No of elements	Base pairs	No of CpGs	No of HpaII/MspI sites	AACCCGG hits ^d	QALuome ^d
Alu ^b	1,194,734	305,076,148 (9.7%)	7,173,987 (25.4%)	742,725 (32.3%)	172,574 (79.1%)	155,878
LINE ^c	1,498,690	638,481,131 (20.4%)	3,412,416 (12.1%)	155,813 (6.8%)	6,881 (3.2%)	0
CpG islands	28,691	21,842,742 (0.7%)	2,089,537 (7.4%)	270,622 (11.8%)	4,470 (2.1%)	0
Genome	-	~3,200,000,000 (100%)	28,217,009 (100%)	2,297,221 (100%)	218,131 (100%)	0

^aData based on GRCh37/hg19 human genome assembly.

^bRepeatMasker's Alu repFamily members discarding FLAMs and FRAMs.

^cRepeatMasker's LINE repClass members.

^dVirtual QALu amplicons.

	Thyroid	Prostate	Breast	Colon	Lung
Normal tissue	6.2 ± 1.6 (n = 9)	4.8 ± 2.6 (n=7)	5.6 ± 2.66 (n=14)	6.9 ± 1.9 (n=16)	5.9 ± 2.0 (n=37)
Tumor tissue	8.2 ± 3.1 (n=59)	8.5 ± 8.8 (n=18)	10.0 ± 5.9 (n=20)	14.6 ± 5.2 (n=16)	14.6 ± 8.3 (n=39)
p-value ^a	0.032	0.357	0.002	< 0.001	< 0.001

^aNormal vs. Tumor, Mann-Whitney U test

Supplementary Figure S3B-S3C).

It is of note that consistent QALu results may be achieved with DNA amounts well below one haploid genome. In fact, linear range response was reached with as little as 0.005 pg of DNA per PCR tube (equivalent to 0.002 haploid genomes) (Supplementary Figure S4). The low requirements of QALu are due to the multiplex nature of its target (a large pool of more than one hundred thousand Alu repeats), reaching sensitivity detection 4 to 5 orders of magnitude higher than a single copy locus (e.g. promoter region of Digestive organ expansion factor homolog (zebrafish) *-DIEXF-* gene) (Supplementary Figure S4 and Supplementary Material). Noteworthy, the complexity of the QALu product is visualized as a broad melting peak for both the *HpaII* and the *MspI* samples, contrasting with the narrower peak of a single PCR product (Supplementary Figure S5).

Finally, the specificity of QALu to amplify Alu elements was validated by next-generation sequencing of five QALu determinations. The results showed that 97% of the reads (range 96.6-98.1%) obtained from the sequenced samples aligned with Alu repeats (Supplementary Table S6) and confirmed the complex composition of QALu product composed of multiple different Alu elements with similar distributions among all the analyzed samples (Supplementary Figure S6).

QALu application to fresh frozen human cancer samples

QALu technique was applied to analyze the levels of unmethylated Alu elements in different cancer types and their normal counterparts (Table 2 and Supplementary Table S7). Interestingly, the different normal tissue types showed similar values of Percentage of UnMethylated Alu elements (PUMA) (Table 2) (Kruskal Wallis test, p -value = 0.308), being the average PUMA 6.0 ± 2.1 (range 1.8-14.8). However, PUMA showed a broad variation in tumors, ranging from 0.9 to 40.5 (average = 10.7 ± 6.8). Furthermore, important differences were observed between cancer types (Table 2 and Figure 3A) (Kruskal Wallis test, p -value < 0.05), with colon and lung exhibiting 2-3 fold higher levels of unmethylated Alu repeats as compared with thyroid, prostate and breast cancer (Supplementary Tables S7 and S8).

The comparison between normal and tumor samples showed statistically significant differences in most cancer types, except in prostate cancer (Mann-Whitney U test, p -value = 0.357) (Figure 3A and Table 2).

Since there were no differences among normal tissues in the proportion of unmethylated Alu elements, the 99th percentile (PUMA = 12%) was taken as a cut-off value to consider a tumor as hypomethylated. Thus, particular analysis of each tumor type revealed that 62.5% of colon and 64.1% of lung tumors were hypomethylated, while in breast and prostate tumors this figure was 20%

and 16.7%, respectively. Otherwise, only 11.9% of thyroid tumors had percentages of unmethylated Alu elements above the reference value (Supplementary Figure S7). When considering only the matched normal-tumor pairs of all cancer types ($n = 81$), the difference among them was evident in most cases (paired Mann-Whitney U test p -value < 0.001). About one third of colon (5/16) and lung (13/37) tumors displayed a high hypomethylation as compared with the paired normal tissue (fold change greater than 3). Contrarily, this big difference was uncommon in breast (2/14), prostate (1/7) or thyroid (0/7) cancer (Supplementary Tables S7 and S9).

Evaluation of PUMA as biomarker in specific cancers

To estimate the potential value of the percentage of unmethylated Alu elements to discriminate between normal and tumoral tissue we performed ROC analyses (Figure 3B). Area Under the Curve (AUC) values confirmed that PUMA was a good biomarker ($AUC > 0.8$) for breast, colon and lung cancer, with sensitivities and specificities >80%, except for breast cancer, whose sensitivity reached 92.9% but the specificity was lower

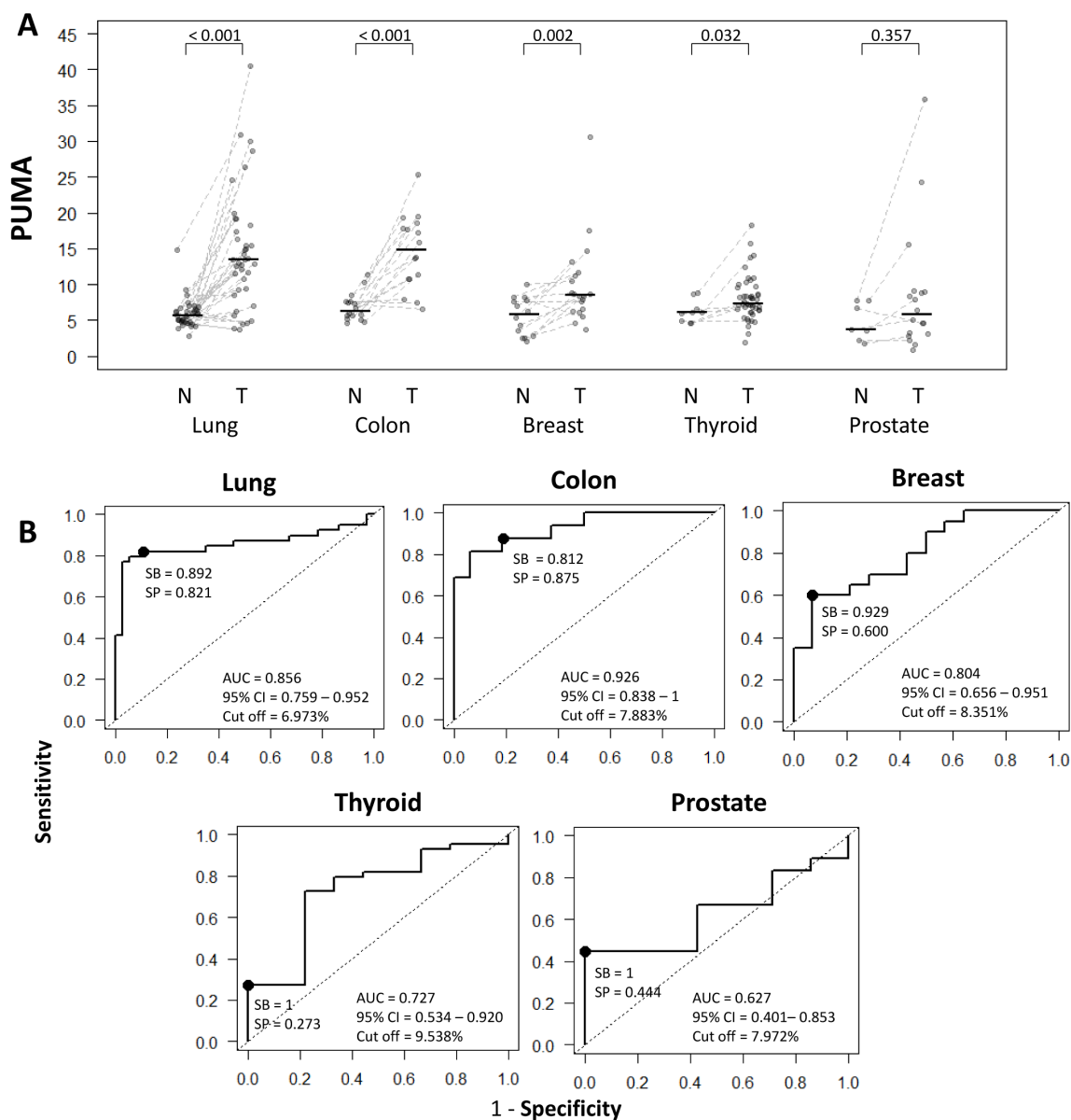


Figure 3: Comparison and diagnostic value of QALu among different cancer types. A. PUMA in different cancer types (normal tissue (N) and tumor (T)); the median of each group is represented by a black line. **B.** Receiver Operating Characteristic curves for the diagnosis of lung, colon, breast, thyroid and prostate cancer according to the percentage of unmethylated Alu (PUMA) elements determined by QALu.

(60%). Otherwise, for both prostate and thyroid cancer, although the sensitivity reached 100%, the specificity was low (44.4% and 27.3%, respectively). The cut-off values varied from tissue to tissue (Figure 3B), with the lowest levels in lung (6.97) and the highest in thyroid (9.54).

Moreover, to evaluate the clinical value of the QALu assay, we performed additional statistical analyses in two cancer types with the highest and the lowest PUMA, namely lung and thyroid cancer.

As described above, lung cancers exhibited high levels of hypomethylated Alu elements, but differences were also detected among lung cancer subtypes. Specifically, lung squamous cell carcinoma showed the highest PUMA (16.8 ± 8.8) compared with lung adenocarcinomas (12.2 ± 6.1) (Mann-Whitney U test

p -value = 0.001) (Figure 4A and Supplementary Table S10). On the other hand, while no significant differences were observed in PUMA among lung normal tissue from never smokers, former smokers (more than 10 years) and current smokers (Kruskal Wallis test p -value >0.05), there was a significant increasing trend (Mann-Whitney U test, p -value = 0.006) of lung tumors to become more unmethylated in current smokers compared to former smokers (Figure 4B).

As regard to thyroid cancer, there were no significant differences in the percentage of unmethylated Alu elements in relation to histology subtype (Kruskal Wallis test, p -value = 0.231) or genetic alteration (RAS or BRAF mutations) (Kruskal Wallis test p -value = 0.147).

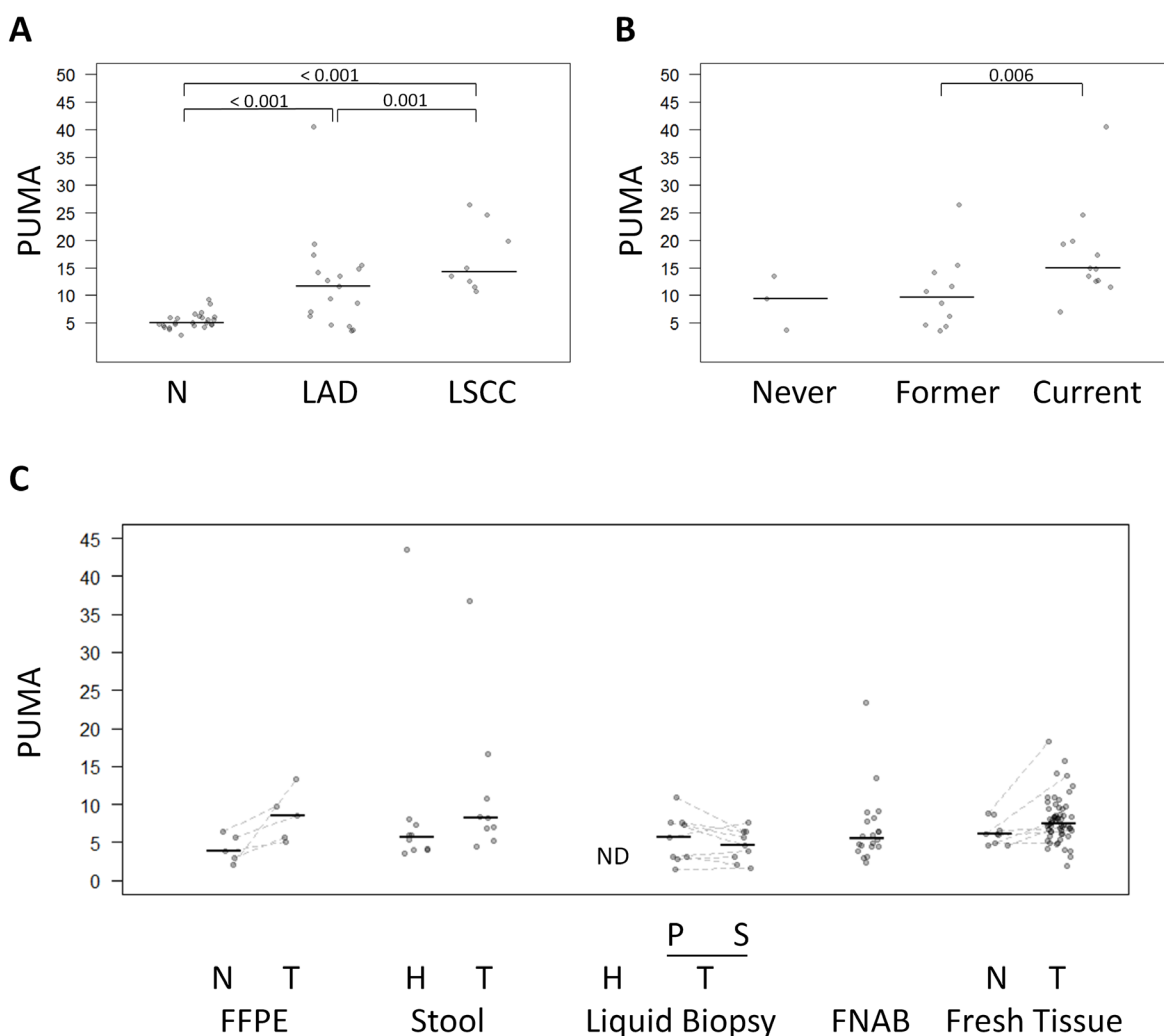


Figure 4: Comparison of QALu among different clinical characteristics and sample types. A. PUMA in normal lung tissue, lung adenocarcinoma (LAD) and lung squamous cell carcinoma (LSCC) and B. in lung cancer patients according to their smoking habits. C. PUMA in different sample types: FFPE (colon cancer patients), stool (healthy donors and colon cancer patients), liquid biopsy (plasma from healthy donors, and plasma and serum from lung cancer patients), FNAB (thyroid goiter patients), fresh tissue (thyroid cancer patients). The median of each group is represented by a black line. Normal (N), tumor (T), healthy donors (H), plasma (P), serum (S), no detectable samples (ND).

QUALu application to diverse pathological examination biospecimens

Next, we evaluated the applicability of QUALu to biospecimens obtained in standard pathological procedures often containing low amount of poor quality DNA (FFPE, FNAB, liquid biopsies and stools). Due to the low amount of starting material, DNA was not quantified and 1 ul of the extracted DNA was used for QUALu analysis (see Supplementary Material). All samples produced detectable levels of amplified Alu elements (qAlu M Cq range: 16 to 30, qAlu H Cq range: 19 to 34) (Supplementary Figure S8), with the exception of plasma samples obtained from healthy individuals and one stool sample from one colon cancer patient.

The low number of samples precluded a robust comparison of the results, but some insightful trends were observed (Figure 4C). FFPE colon tumors showed higher PUMA than the matching normal samples (paired Mann-Whitney U test p -value = 0.062). Moreover, elevated Alu hypomethylation in the stools was more frequent in colon cancer patients than in control individuals (Mann-Whitney U test p -value = 0.079). Interestingly, liquid biopsies produced similar PUMA in plasma and serum from lung cancer patients, but did not amplify in cancer-free controls, which is consistent with the common absence of circulating free DNA in healthy individuals [20].

Finally thyroid goiters FNABs showed a PUMA in the same range than normal tissues with some exceptions: four cases presented a PUMA above 8.8, the highest value in normal thyroid tissue (Figure 4C).

DISCUSSION

Most of human genome is methylated, but in a wide range of pathologies, including cancer, a global DNA hypomethylation occurs affecting in large extent repetitive elements, which constitute ~45% of the genome. Weisenberger et al. [21] demonstrated that the methylation of different repetitive sequences, namely, LINE-1, Alu and satellite 2 (Sat2), significantly correlated with global methylation levels measured by high performance liquid chromatography (HPLC), and proposed the use of these repeats as surrogate reporters of global methylation. LINES have been broadly used to estimate the global levels of hypomethylation [22-24], but Alu elements display some features that make them more suited for this purpose. Namely, Alu repeats constitute the most abundant retrotransposon and contain 25% of all the CpGs in the human genome. Moreover, due to their prevalent localization in gene-rich regions [15] epigenetic variations in Alu repeats may have direct implications in gene regulation, and by extension, in tumor biology.

Here we present QUALu, a new technique to measure the levels of DNA unmethylation in Alu repeats,

with several features that facilitate a direct implementation in clinical and research settings. These include a 100 fold higher sensitivity than other related methods [25], as accurate determinations can be performed with as little as 300 pg of DNA (corresponding to approximately 150 diploid cells). QUALu specificity for Alu elements is extremely high, as demonstrated by deep sequencing of its products, where 97% of the reads mapped in Alu repeats.

Moreover, thanks to the calibration with internal controls (LIPA simultaneously with Alu repeats), QUALu is relatively unaffected by the quantity and quality of the starting material in artificially degraded DNA. Furthermore, we have demonstrated that this technique is amenable to be applied to a broad spectrum of pathological examination specimens routinely collected in clinical settings (frozen tissues, FFPE, liquid biopsies, stools and FNAB). In spite of the low number of samples analyzed, preliminary results are promising, especially for FFPE and stools. Nevertheless, direct applicability of QUALu in different clinical settings requires the analysis of large series of cases to define the threshold, sensitivities and specificities.

With clinical practice in mind, technical benefits of QUALu include the small number of steps (digestion-ligation, real time PCR and analysis), the short time required to complete the determination (less than 5 hours from the DNA to the final result, even without automation of the process), and the low cost (about 6.3 US\$ per sample, including technician labor).

As mentioned above, several techniques have been developed to estimate global methylation, and many of them target repeat elements as surrogate reporters (reviewed in [9, 13]). While most of these methods may constitute good alternatives to compare global methylation levels among a few samples, their implementation as a clinical tool is not a straightforward approach due to either technical complexity or exquisite sample necessities. QUALu simplicity and limited equipment requirements facilitate its implementation in most laboratories.

To assess the clinical potential of QUALu, we determined the extent of Alu hypomethylation in different human cancers by analyzing normal and tumor tissues. In a first analysis we compared the different normal tissues and different individuals, showing that the levels of unmethylated Alu elements were very consistent from tissue to tissue and from individual to individual. This result was in agreement with previous studies analyzing global DNA methylation by MethyLight and HPLC [21] or targeting Alu elements [26-27], but not with other works reporting tissue-associated global methylation differences targeting LINE-1 [28].

Our data suggests that the degree of hypomethylation is variable among different cancer types, with thyroid, prostate and breast cancer exhibiting low levels of hypomethylation, while cancers of colon and lung displayed the highest levels. Chalitchagorn et al.

[28] analyzed LINE-1 methylation using bisulfite based PCR in several cancers, and although they found high levels of hypermethylation in some types (e.g. esophagus cancer), no hypomethylation was observed among the ones analyzed in our study. This might be explained by the low sensitivity of their technique or the low number of samples analyzed.

It is interesting to note that the low-hypomethylation cancer group included hormone-related tissues (thyroid, prostate and breast) with no direct interaction with external factors, while the high-hypomethylation group (colon and lung) was composed by tumors with a high exposure to environmental factors (e.g. diet, air). While we do not know the reason for this association, there are evidences supporting the impact of certain environmental factors (drugs, chemicals, pollutants and other agents) in the deregulation of epigenetic enzymes, which will eventually generate epigenetic changes, including DNA hypomethylation, that may accumulate over the time causing alterations in key cellular processes and promoting cancer [29-30]. Noteworthy, it has been reported that Alu hypomethylation (but not LINE-1 hypomethylation) in esophageal mucosa may reflect an epigenetic field for cancerization in esophageal carcinogenesis [12]. Other alternative explanations may be related with the dynamics of tumor progression in different tumor types and the role of DNA hypomethylation behind specific deregulation of biological pathways, including genomic stability [31-32].

Confirming previous reports [33] we found significant differences in the levels of hypomethylation among the two types of lung cancers considered here. Lung adenocarcinomas, the histological subtype most frequently associated with never-smokers and former smokers, were less hypomethylated than lung squamous cell carcinomas. Many studies have suggested a strong correlation between loss of DNA methylation and smoking habit in cancer patients [34-36], but also in healthy people [37-38]. In this regard, we found a significant increase of hypomethylation in current smokers compared to former smokers, but this trend was not observed in the adjacent normal tissue. This result may indicate different mechanisms of tumor progression in ex-smokers as compared with current smokers.

In summary, we have demonstrated that QALu is a feasible approach to analyze global DNA methylation in almost any type of biospecimen routinely collected in ordinary clinical settings. DNA hypomethylation is a hallmark of cancer but, as we have shown, its degree is highly variable. Its determination with a technique as QALu may have a broad spectrum of applications including diagnostic and prognostic evaluations.

MATERIALS AND METHODS

Samples

This study included a total of 300 pathological examination samples of different sources: 220 fresh frozen tissue samples (Supplementary Tables S1 and S2), 10 FFPE samples, 31 liquid biopsies, 19 stool samples and 20 thyroid goiter FNAB samples. Regarding fresh frozen tissues, 16 colorectal carcinomas and their paired normal adjacent tissues were obtained from Hospital Universitari de Bellvitge (Barcelona, Spain). Forty-four thyroid carcinomas and 9 paired adjacent thyroid tissues were obtained as described in our previous study [39]. DNA of 20 breast carcinomas and 14 paired normal adjacent tissues, 18 prostate carcinomas and 7 paired normal adjacent tissues, and 39 lung carcinomas and 37 normal adjacent tissues were obtained from the Spanish National DNA Bank (BNADN, Salamanca, Spain). Patient characteristics are shown in Supplementary Tables S1 and S2. Five normal and tumoral paired colorectal carcinoma FFPE samples were obtained from Cooperative Human Tissue Network (CHTN). Nine stool samples from colorectal carcinoma patients and 10 from healthy donors were obtained from Hospital Universitari de Bellvitge (Barcelona, Spain). Finally, nine plasma and serum paired lung carcinoma liquid biopsies, 13 plasma liquid biopsies from healthy donors and 20 thyroid goiter FNAB samples were obtained from Hospital Universitari Germans Trias i Pujol (Badalona, Spain). The study was approved by the Hospital Germans Trias i Pujol Ethics Committee. Informed consent was obtained before surgery.

The colorectal carcinoma cell line HCT116 was obtained from the American Type Culture Collection (ATCC) and was authenticated on 3rd March 2014 by using the AmpFLSTR® Identifiler® Plus PCR Amplification Kit (Applied Biosystems). Cells were cultured in D-MEM/F12, supplemented with sodium pyruvate, L-glutamine and 10% fetal bovine serum (Life Technologies, MD, USA) and were maintained at 37°C in a 5% CO₂ atmosphere. Genomic DNA was isolated using different methods as described in the Supplementary Material.

Human genome sequence data sets

We used the GRCh37/hg19 human genome assembly. Genomic positions of Alu elements, LINE sequences, CpG islands and CpG dinucleotides were retrieved from the UCSC MySQL repository (genome-mysql.cse.ucsc.edu). Motif search was performed with EMBOSS (<http://emboss.sourceforge.net/>). To count the number of overlapping features (Table 1) we used the BEDTools package (v2.19.1, [40]). For more technical details see Supplementary Material.

Quantification of Unmethylated Alu (QUAlu) assay

The principle underlying this technique is the selective amplification of Alu repeats containing an unmethylated CpG site within the consensus sequence AACCCGG. Briefly, genomic DNA was digested in parallel in two separated tubes with *HpaII* and *MspI* methylation-sensitive and -insensitive isoschizomers, respectively, which leave identical sticky ends (C/CGG). Next, a synthetic adaptor was ligated to the digested DNA fragments. Quantification was performed by qPCR using a primer complementary to the chimeric sequence of the adaptor plus the consensus Alu sequence after *HpaII* digestion (AACCC + synthetic adaptor) and another one complementary to the Alu consensus sequence and located ~20 nucleotides upstream of the *HpaII* cutting site (Supplementary Figure S1). Thereby, qPCR of *MspI* digestion (qAlu M) allowed the quantification of all the amplifiable Alu elements (irrespective of the methylation status), while qPCR of *HpaII* digestion (qAlu H) only quantified the subset of amplifiable Alu elements containing an unmethylated CpG. Thus, the final result corresponded to the fraction of unmethylated Alu elements respect the total number of amplifiable Alu elements calculated according to the equation described below. Furthermore, two specific qPCRs for L1PA (a Long Interspersed Nuclear Element-1, LINE-1 subfamily) were performed to normalize the DNA input for both *MspI* (qL1PA M) and *HpaII* digestions (qL1PA H). For more technical details see Supplementary Material.

QUAlu data analysis

Statistical analyses were performed using R version 3.1.0. The Percentage of UnMethylated Alu elements (PUMA) for each sample was assigned according to this equation:

$$\text{PUMA} = \frac{\frac{E_{q\text{Alu H}}^{-Cq_{q\text{Alu H}}}}{E_{q\text{L1PA H}}^{-Cq_{q\text{L1PA H}}}}}{\frac{E_{q\text{Alu M}}^{-Cq_{q\text{Alu M}}}}{E_{q\text{L1PA M}}^{-Cq_{q\text{L1PA M}}}}} \times 100$$

The relative amount of unmethylated Alu elements (given by qAlu H normalized by the reference sequence qL1PA (qL1PA H)) and the relative amount of total amplifiable Alu elements (given by qAlu M normalized by qL1PA M) were calculated as a ratio of exponential functions in which the base was the qPCR efficiency (E) and the variable was the quantification cycle (Cq). To tackle the qPCR error propagation, permutation tests were done to obtain a final PUMA and its variation using the qPCR R package (v1.4-0, [41]). Mann-Whitney U test and Kruskal Wallis tests were used, as appropriate, to assess

the significance among the different groups of samples. Correlation analyses were conducted using two-tailed Kendall tests. The significance level was established at $p < 0.05$ for all analyses. Receiver operating characteristic (ROC) curves were generated using the pROC R package (v1.7.3, [42]) to assess the cut-off value that best discriminated between tumor and normal tissue according to the PUMA.

Characterization of QUAlu product by Next Generation Sequencing

To determine the specificity of the technique, the products generated by qAlu H and qAlu M from 5 samples (the colorectal cancer cell line HCT116, a lung squamous carcinoma with its normal matching tissue and a papillary thyroid carcinoma with its normal matching tissue) were sequenced using Ion Torrent technology (Life Technologies). For more technical details see Supplementary Material and Supplementary Table S3.

ACKNOWLEDGMENTS

We thank Mar Muñoz, Esmeralda Castelblanco and Veronica Mancikova for excellent technical support. We also thank all the patients and control donors for making this study possible.

CONFLICTS OF INTEREST

MAP is cofounder and equity holder of Aniling, a biotech company with no interests in this paper. The other authors declare no conflict of interest.

GRANT SUPPORT

RB was supported by a FPI fellowship from Ministerio de Economía y Competitividad. AD-V was supported in part by a contract PTC2011-1091 from Ministerio de Economía y Competitividad. This work was supported by grants from FEDER, the Ministerio de Economía y Competitividad (SAF2011/23638 to MAP), the Instituto de Salud Carlos III (FIS PI11/02421 to JR, FIS PI11/01359 and FIS PI14/00240 to MR, FIS PI14/00308 to MJ, FIS PI12/00511 to MP), and Fundació Olga Torres (to MJ).

Editorial note

This paper has been accepted based in part on peer-review conducted by another journal and the authors' response and revisions as well as expedited peer-review in Oncotarget.

REFERENCES

1. Sandoval J and Esteller M. Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev.* 2012; 22:50-55.
2. Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics.* 2009; 1:239-259.
3. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007; 447:433-440.
4. Li J, Huang Q, Zeng F, Li W, He Z, Chen W, Zhu W and Zhang B. The prognostic value of global DNA hypomethylation in cancer: a meta-analysis. *PLoS One.* 2014; 9:e106290.
5. Fabris S, Bollati V, Agnelli L, Morabito F, Motta V, Cutrona G, Matis S, Grazia Recchia A, Gigliotti V, Gentile M, Delilieri GL, Bertazzi PA, Ferrarini M, Neri A and Baccarelli A. Biological and clinical relevance of quantitative global methylation of repetitive DNA sequences in chronic lymphocytic leukemia. *Epigenetics.* 2011; 6:188-194.
6. Igarashi S, Suzuki H, Niinuma T, Shimizu H, Nojima M, Iwaki H, Nobuoka T, Nishida T, Miyazaki Y, Takamaru H, Yamamoto E, Yamamoto H, Tokino T, Hasegawa T, Hirata K, Imai K, et al. A Novel Correlation between LINE-1 Hypomethylation and the Malignancy of Gastrointestinal Stromal Tumors. *Clin Cancer Res.* 2010; 16:5114-5123.
7. Saito K, Kawakami K, Matsumoto I, Oda M, Watanabe G and Minamoto T. Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. *Clin Cancer Res.* 2010; 16:2418-2426.
8. Frigola J, Sole X, Paz MF, Moreno V, Esteller M, Capella G and Peinado MA. Differential DNA hypermethylation and hypomethylation signatures in colorectal cancer. *Hum Mol Genet.* 2005; 14:319-326.
9. Torano EG, Petrus S, Fernandez AF and Fraga MF. Global DNA hypomethylation in cancer: review of validated methods and clinical significance. *Clin Chem Lab Med.* 2012; 50:1733-1742.
10. Moore LE, Pfeiffer RM, Poscablo C, Real FX, Kogevinas M, Silverman D, Garcia-Closas R, Chanock S, Tardon A, Serra C, Carrato A, Dosemeci M, Garcia-Closas M, Esteller M, Fraga M, Rothman N, 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; 9:359-366.
11. Choi JY, James SR, Link PA, McCann SE, Hong CC, Davis W, Nesline MK, Ambrosone CB and Karpf AR. Association between global DNA hypomethylation in leukocytes and risk of breast cancer. *Carcinogenesis.* 2009.
12. Matsuda Y, Yamashita S, Lee YC, Niwa T, Yoshida T, Gyobu K, Igaki H, Kushima R, Lee S, Wu MS, Osugi H, Suehiro S and Ushijima T. Hypomethylation of Alu repetitive elements in esophageal mucosa, and its potential contribution to the epigenetic field for cancerization. *Cancer Causes Control.* 2012; 23:865-873.
13. Jorda M and Peinado MA. Methods for DNA methylation analysis and applications in colon cancer. *Mutat Res.* 2010; 693:84-93.
14. Deininger P. Alu elements: know the SINEs. *Genome Biol.* 2011; 12:236.
15. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409:860-921.
16. Karimi M, Johansson S and Ekstrom TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation. *Epigenetics.* 2006; 1:45-48.
17. Rodriguez J, Vives L, Jorda M, Morales C, Munoz M, Vendrell E and Peinado MA. Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. *Nucleic Acids Res.* 2008; 36:770-784.
18. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009; 462:315-322.
19. Barrera V and Peinado MA. Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale. *Nucleic Acids Res.* 2012; 40:11490-11498.
20. Schwarzenbach H, Hoon DS and Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer.* 2011; 11:426-437.
21. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, Ehrlich M and Laird PW. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* 2005; 33:6823-6836.
22. Tajuddin SM, Amaral AF, Fernandez AF, Chanock S, Silverman DT, Tardon A, Carrato A, Garcia-Closas M, Jackson BP, Torano EG, Marquez M, Urdinguio RG, Garcia-Closas R, Rothman N, Kogevinas M, Real FX, et al. LINE-1 methylation in leukocyte DNA, interaction with phosphatidylethanolamine N-methyltransferase variants and bladder cancer risk. *Br J Cancer.* 2014; 110:2123-2130.
23. Salas LA, Villanueva CM, Tajuddin SM, Amaral AF, Fernandez AF, Moore LE, Carrato A, Tardon A, Serra C, Garcia-Closas R, Basagana X, Rothman N, Silverman DT, Cantor KP, Kogevinas M, Real FX, et al. LINE-1 methylation in granulocyte DNA and trihalomethane exposure is associated with bladder cancer risk. *Epigenetics.* 2014; 9:1532-1539.
24. Inamura K, Yamauchi M, Nishihara R, Lochhead P, Qian ZR, Kuchiba A, Kim SA, Mima K, Sukawa Y, Jung S, Zhang X, Wu K, Cho E, Chan AT, Meyerhardt JA, Harris CC, et al. Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. *J Natl Cancer Inst.* 2014; 106.

25. Lisanti S, Omar WA, Tomaszewski B, De Prins S, Jacobs G, Koppen G, Mathers JC and Langie SA. Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PLoS One*. 2013; 8:e79044.
26. Choi SH, Worswick S, Byun HM, Shear T, Soussa JC, Wolff EM, Douer D, Garcia-Manero G, Liang G and Yang AS. Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int J Cancer*. 2009; 125:723-729.
27. Wu HC, Delgado-Cruzata L, Flom JD, Kappil M, Ferris JS, Liao Y, Santella RM and Terry MB. Global methylation profiles in DNA from different blood cell types. *Epigenetics*. 2011; 6.
28. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, Voravud N, Sriuranpong V and Mutirangura A. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene*. 2004; 23:8841-8846.
29. Herceg Z and Vaissiere T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics*. 2011; 6:804-819.
30. Mathers JC, Strathdee G and Relton CL. Induction of epigenetic alterations by dietary and other environmental factors. *Adv Genet*. 2010; 71:3-39.
31. Rodriguez J, Frigola J, Vendrell E, Risques RA, Fraga MF, Morales C, Moreno V, Esteller M, Capella G, Ribas M and Peinado MA. Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res*. 2006; 66:8462-9468.
32. Eden A, Gaudet F, Waghmare A and Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science*. 2003; 300:455.
33. Suzuki M, Shiraishi K, Eguchi A, Ikeda K, Mori T, Yoshimoto K, Ohba Y, Yamada T, Ito T, Baba Y and Baba H. Aberrant methylation of LINE-1, SLIT2, MAL and IGFBP7 in non-small cell lung cancer. *Oncol Rep*. 2013; 29:1308-1314.
34. Smith IM, Mydlarz WK, Mithani SK and Califano JA. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. *Int J Cancer*. 2007; 121:1724-1728.
35. Andreotti G, Karami S, Pfeiffer RM, Hurwitz L, Liao LM, Weinstein SJ, Albanes D, Virtamo J, Silverman DT, Rothman N and Moore LE. 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; 9:404-415.
36. Liu F, Killian JK, Yang M, Walker RL, Hong JA, Zhang M, Davis S, Zhang Y, Hussain M, Xi S, Rao M, Meltzer PA and Schrupp DS. Epigenomic alterations and gene expression profiles in respiratory epithelia exposed to cigarette smoke condensate. *Oncogene*. 2010; 29:3650-3664.
37. Shigaki H, Baba Y, Watanabe M, Iwagami S, Miyake K, Ishimoto T, Iwatsuki M and Baba H. LINE-1 hypomethylation in noncancerous esophageal mucosae is associated with smoking history. *Ann Surg Oncol*. 2012; 19:4238-4243.
38. Tajuddin SM, Amaral AF, Fernandez AF, Rodriguez-Rodero S, Rodriguez RM, Moore LE, Tardon A, Carrato A, Garcia-Closas M, Silverman DT, Jackson BP, Garcia-Closas R, Cook AL, Cantor KP, Chanock S, Kogevinas M, et al. Genetic and non-genetic predictors of LINE-1 methylation in leukocyte DNA. *Environ Health Perspect*. 2013; 121:650-656.
39. Mancikova V, Buj R, Castelblanco E, Inglada-Perez L, Diez A, de Cubas AA, Curras-Freixes M, Maravall FX, Mauricio D, Matias-Guiu X, Puig-Domingo M, Capel I, Bella MR, Lerma E, Castella E, Reverter JL, et al. DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *Int J Cancer*. 2014; 135:598-610.
40. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841-842.
41. Ritz C and Spiess AN. qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*. 2008; 24:1549-1551.
42. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC and Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011; 12:77.