

RESEARCH ARTICLE

Editorial Process: Submission:05/05/2019 Acceptance:08/01/2019

## An Internal Control for Evaluating Bisulfite Conversion in the Analysis of *Short Stature Homeobox 2* Methylation in Lung Cancer

Vo Thi Thuong Lan<sup>1\*</sup>, Vu Lan Trang<sup>2</sup>, Nguyen Thuy Ngan<sup>1</sup>, Ho Van Son<sup>3</sup>, Nguyen Linh Toan<sup>4</sup>

### Abstract

**Objective:** The methylation status is considered as powerful diagnostic, prognostic, and predictive biomarkers. However, the limited DNA amount and conversion efficiency after bisulfite treatment are considerable hindrances in quantitative methylation analysis. In this study, we designed an artificial internal control (IC) system that contained the cytosine-free fragment (CFF) following CpG sequences of the *SHOX2* promoter whose methylation status has been described as a valuable biomarker of lung cancer. Its performance in quantifying DNA recovery and bisulfite conversion efficiency as well as in detecting false-positive *SHOX2* methylation was determined on samples from lung cancer patients. **Material and Methods:** The IC system is composed of two pConIC and pUnIC plasmids that both contain a cytosine-free (CF) sequence derived from the CFF and the CpG containing *SHOX2* sequences. They are identical in sequence, except that in the ConIC insert, all cytosines have been converted into thymines. Thus, the ConIC can be used as calibrator of 100% bisulfite conversion efficiency, while the UnIC is the indicator in order to evaluate the DNA recovery, bisulfite conversion efficiency of the *SHOX2* promoter sequence by quantitative real time PCR. **Results:** The copy number of the target sequences impacted on both DNA recovery rates and bisulfite conversion efficiency. An amount of 0.005 ng pUnIC ( $10^6$  copies) showed recovery rate of 18%, similar to that of pConIC, and a bisulfite conversion efficiency of the *SHOX2* reaching 98.7%. On the contrary, higher copy number of pUnIC showed incomplete conversion (<85%) and over recovery (~42%). Using this calibrator/indicator couple, we were able to detect false-positive *SHOX2* methylation (3.77% instead of 0.03%) due to incomplete bisulfite conversion. **Conclusion:** Our results proposed a customizable internal control using the ConIC/UnIC as calibrator/indicator to quantify simultaneously and accurately the DNA recovery and bisulfite conversion efficiencies of individual sequence as well as whole genome in methylation assays, thus promoting the validation of standardized clinical DNA methylation biomarker values to progress toward clinical applications.

**Keywords:** *SHOX2* promoter methylation- DNA recovery- Bisulfite conversion efficiency- Internal Control – IC

*Asian Pac J Cancer Prev*, **20 (8)**, 2435-2443

### Introduction

Wide studies of CpG methylation and its association with diseases such as mental illness (Liu et al., 2018), autoimmune disorders (Sun et al., 2016), infectious diseases and cancer (Silmon de Monerri and Kim, 2014; Dugué et al., 2018) have confirmed that DNA methylation is a powerful diagnostic, prognostic and predictive biomarker. However, only 14 DNA methylation-based biomarker assays are currently commercially available (Koch et al., 2018), which is mainly due to the lack in methods for the evaluation and validation of standardized clinical biomarker values. For instance, an eight-fold difference (ranging from 5% to 40%) in the *BRCA1* methylation in

ovarian cancer was reviewed among different methods (Senturk et al., 2010). In addition, a false positive detection of DNA methylation using pyrosequencing as compared to bisulfite sequencing has been observed (Owa et al., 2018). Using the same bisulfite-based methylation specific PCR (MSP) method, two groups reported an overestimation of DNA methylation due to remaining trace of unconverted DNA and incomplete conversion of unmethylated cytosines to uracil during bisulfite treatment (Lan et al., 2014; Rand et al., 2002). Thus, the critical evaluation of individual targeted region (s) with the precise CpG number and the detailed validation of all the technical steps are recommended in performing DNA methylation-based biomarker assays (Koch et al., 2018; Bock et al., 2016).

<sup>1</sup>Faculty of Biology, VNU University of Science, Hanoi, <sup>3</sup>175 Hospital, Ho Chi Minh City, <sup>4</sup>Department of Pathophysiology, Medical University, Ha Dong, Vietnam, <sup>2</sup>Sorbonne Universités, UPMC Univ. Paris 06, École normale supérieure, PSL Research University, CNRS, INSERM, APHP, Laboratoire des Biomolécules (LBM), Paris, France. \*For Correspondence: [vothithuonglan@hus.edu.vn](mailto:vothithuonglan@hus.edu.vn)

The analysis of DNA methylation of specific genes/regions now exclusively relies on bisulfite-treated DNA followed by PCR amplification or bisulfite sequencing (Kurdyukov and Bullock, 2016). This chemical specifically converts unmethylated cytosine, but not methylated cytosine, to uracil residues (Clark et al., 1994). However, the low pH and high temperature in this reaction can lead to incomplete or inappropriate bisulfite conversion as well as DNA degradation, resulting in over or underestimation of DNA methylation (Olova et al., 2018; Ji et al., 2014; Geneux et al., 2008). Recently, different commercial bisulfite conversion kits, applied to various DNA sources with high and low molecular weight (HMW/LMW), have been comprehensively evaluated for their DNA recovery and bisulfite conversion efficiencies (Holmes et al., 2014; Leontiou et al., 2015; Ørntoft et al., 2017; Tierling et al., 2018). In these performance evaluations, highly sensitive detection methods such as quantitative and digital PCR (qPCR, dPCR) were used to amplify different target regions in genomic DNA (HMW-DNA) or in circulating cell-free DNA (LMW-DNA) (Ørntoft et al., 2017; Tierling et al., 2018; Kint et al., 2018). In addition, custom-designed PCR fragments or cloned target sequences spiked in the HMW/LMW DNAs were used as references (Leontiou et al., 2015; Ørntoft et al., 2017; Munson et al., 2007). These artificially or endogenously targeted templates included (i) cytosine-free fragments (CFF) that allow the simultaneous quantification of total DNA before and after bisulfite treatment thus the determination of bisulfite conversion recovery; (ii) non-CpG cytosine sequences that allow the quantification of converted DNA thus the determination of bisulfite conversion efficiency, and (iii) CpG cytosine sequences that allow the quantification of inappropriate conversion to methylated cytosines thus the determination of bisulfite conversion specificity (Leontiou et al., 2015; Kint et al., 2018). However, in most of those performance evaluations of commercial kits, only one concentration of the references was used and yet the nucleotide sequence of the references used to estimate the conversion efficiency were not related with those of targeted DNA regions, thus limiting the in-depth performance assessment. Indeed, it has been reported that DNA concentration, sequence complexity, GC content, secondary structure elements and even a given cytosine in particular sequence context can interfere with bisulfite conversion (Geneux et al., 2008; Harrison et al., 1998). Recently, Owa et al. (2018) provided evidence that localized incompleteness of bisulfite conversion is a major source of false-positive detection of methylated cytosines. Therefore, the use of an optimal indicator to monitor bisulfite conversion and DNA recovery in methylation assays has been highly recommended (Ørntoft et al., 2017).

In the present study, we constructed an internal control (IC) harboring both cytosine-free (CF) and CpG cytosine sequences; the former was derived from CFF while the latter was derived from the *SHOX2* gene promoter whose methylation status has been described as a valuable biomarker for lung cancer detection (Darwiche et al., 2013; Song et al., 2015). The IC was mixed with genomic DNA extracted from formalin-fixed, paraffin-embedded (FFPE) samples from patients with lung cancer

or with non-cancerous pulmonary diseases. We showed that appropriate concentration of the IC combined with adequate DNA amount allowed the optimal simultaneous monitoring of DNA recovery and bisulfite efficiency using quantitative realtime PCR. Subsequently, based on the bisulfite conversion efficiency assessed on the IC, false positive *SHOX2* methylation level and successful conversion of the *SHOX2* sequence was accurately detected in patients with lung cancer patients.

## Materials and Methods

### Construction of the IC system

Four single-stranded synthetic oligonucleotides were designed, each containing CF sequence derived from the CFF (GRCh37:Chr13,19555120-19555208) (Holmes et al., 2014), and CpG-containing sequence (named MP-*SHOX2*) of the *SHOX2* promoter (GRCh38:Chr3,158103550-158103675) (Song et al., 2015). All the cytosines (C) were converted to thymines (T) in two oligonucleotides (named ConIC-Oligos) but not converted in two others (named UnIC-Oligos). The two oligonucleotides in each couple (ConIC-Oligos or UnIC-Oligos) have been designed to overlap each other so that they annealed together in the PCR reaction for amplifying the duplex DNAs by primer extension. The resulting dsDNAs, named respectively ConIC and UnIC, were then subsequently cloned into pTZ57R/T vector using InsTAclone PCR Cloning Kit (Thermo Fisher Scientific) (Figure 1). The recombinant plasmids, named respectively pConIC and pUnIC, were extracted, and the insert sequences were confirmed by sequencing. Nucleotide sequence of the ConIC was identical to that of the UnIC except that all C were replaced by T, thus the ConIC can be used as a calibrator for 100% conversion efficiency of the UnIC sequence. Both ConIC and UnIC sequences formed the IC system used in this study to evaluate DNA recovery and bisulfite conversion efficiencies. The nucleotide sequences of the ConIC and UnIC inserts and their synthesis by PCR extension were presented in Supplementary Figure 1.

### pConIC and pUnIC spike preparation

Both pConIC and pUnIC recombinant plasmids were linearized with *SmaI* restriction enzyme (New England Biolabs) and purified using the PureLink™ PCR Purification Kit (Invitrogen Life Technologies). Plasmid concentrations were quantified with the NanoDrop2000 device (Invitrogen Life Technologies). The copy number was estimated according to the molecular weight, amount, and length of each plasmid. To evaluate the amplification efficiency, qPCR were performed on serial concentrations ranging from 200 pg to 0.4 pg (equivalent to  $\sim 5 \times 10^7$ - $10^5$  copies) of the linearized pConIC diluted in DNA-free water or a background of 10 ng bisulfite treated genomic DNA. Since in the ConIC all cytosines had been already converted to thymines, this sequence is used to evaluate DNA recovery after bisulfite processing without chemical conversion: to do so 5 ng and 1 ng (equivalent to  $\sim 10^9$ - $3 \times 10^8$  copies) of linearized pConIC was mixed with 500 ng of genomic DNA and converted by bisulfite; then the pConIC copy number was determined by qPCR.

To evaluate DNA recovery after bisulfite conversion along with bisulfite conversion efficiency, serial concentrations ranging from 5 ng to 0.005 ng (equivalent to  $\sim 10^9 - 10^6$  copies) of linearized pUnIC diluted in a background of 500 ng genomic DNA were bisulfite converted and subsequently subjected to 2 qPCR assays. All reaction of bisulfite conversion was performed with the EpiTect Bisulfite Kit (Qiagen).

#### Sample collection, genomic DNA isolation and bisulfite conversion

Lymphocytes were collected from healthy volunteers; formalin-fixed, paraffin-embedded (FFPE) tissue specimens were collected from 97 lung cancer patients and 72 patients suffering from non-cancerous pulmonary diseases (whose classification was examined by pathologists) at the Department of Pathology, 175 Hospital (Ho Chi Minh City) during 2016. Informed consent was obtained from healthy participants and patients in written form and the study was approved by the guidelines of the VNU University of Science ethical committee in Vietnam (9/2016/108/HĐTN/106-YS.06-2015.07). Genomic DNAs were extracted from lymphocytes of healthy volunteers or FFPE lung tissues using the QIAamp DNA Mini Kit (Qiagen) or QIAamp DNA FFPE Tissue Kit (Qiagen). An amount of 500 ng DNAs quantified with the NanoDrop2000 (Invitrogen) was submitted to bisulfite conversion using the EpiTect Bisulfite Kit (Qiagen).

#### Quantitative SYBR Green real time PCR assay

Bisulfite specific primers that are complementary to the sense strand of the converted DNA were designed using the Methyl Primer Express Software v1.0. The primers used for methylated *SHOX2* templates contained CpGs to ensure their specific annealing to methylated sequences. Real time PCR reactions with primer sets specifically designed to ConIC and UnIC were performed on bisulfite treated or nontreated DNAs extracted from healthy volunteers' lymphocytes in order to verify their specificity.

The methylation status of the UnIC was evaluated by real time PCR that was carried out in 20  $\mu$ l per reaction using 10 ng of bisulfite converted DNA as template and qPCRBIOSyGreen Mix Hi-ROX (PCR Biosystem). Real time PCR assays were duplexed in 2 reactions: (1) a reaction to quantify the total input UnIC; (2) a methylation specific reaction to quantify the bisulfite converted UnIC. Water with no DNA template was included in each PCR reaction as a control for contamination. All qPCR

reactions were replicated and performed using a 7500 Real-time PCR instrument (Applied Biosystems, CA). Primer sequences, amplicon lengths and qPCR conditions are shown in Supplementary Table S1.

#### Statistical Analysis

The relative percentage of UnIC conversion by bisulfite was calculated for each sample according to the delta-delta cycle threshold ( $\Delta\Delta CT$ ) method as described previously (Livak and Schmittgen, 2001).

$$\Delta\Delta CT_{\text{Sample}} = \Delta CT_{\text{UnIC}} - \Delta CT_{\text{Calibrator}}; \Delta CT_{\text{UnIC}} = CT_{\text{UnIC/MP-SHOX2}} - CT_{\text{UnIC/CF}}; \Delta CT_{\text{Calibrator}} = CT_{\text{ConIC/MP-SHOX2}} - CT_{\text{ConIC/CF}}$$

Efficiency of bisulfite conversion was calculated using the following formula: Percentage conversion<sub>UnIC</sub> = Percentage conversion<sub>ConIC</sub>  $\times 2^{-\Delta\Delta CT_{\text{Sample}}} = 100\% \times 2^{-\Delta\Delta CT_{\text{Sample}}}$ .

For all statistical analyses, a p-value  $\leq 0.05$  was considered as significant. All analyses were performed with the STATA program version 12 (<https://www.stata.com/>) and Graphpad Prism program version 7 (<https://www.graphpad.com/scientific-software/prism/>).

## Results

#### Validation of the specificity of primer sets

The CF and *SHOX2* promoter sequences contained in the IC system (both ConIC and UnIC) differ in positions from genomic DNA (Figure 1). Thus, specific primer sets (named CF-F/R and MP-*SHOX2*-F/R, see Table S1), which specifically recognised CF and *SHOX2* sequences, respectively, amplified only the targets in the IC but not those in genomic DNA. Actually, no PCR product was obtained when genomic DNA from healthy volunteers' lymphocytes was used as templates, with or without previous bisulfite conversion. On the contrary, the primer sets specifically amplified the CF and *SHOX2* sequences in ConIC and UnIC (Table 1). Moreover, as false positive result in DNA methylation analysis could be due to mismatches of primers to unspecific targets (Dohmet al., 2008; Lan et al., 2014), the specificity of Me-*SHOX2*-F/R primers, designed to specifically recognize methylated *SHOX2* promoter in genomic DNA, was tested using bisulfite treated and untreated DNAs as templates for qPCR. No PCR product was amplified from untreated DNAs. Note that these primers neither amplified any PCR product from bisulfite treated or untreated IC sequences. The specificity of the primer sets used in the study were

Table 1. Specificity of Primer Sets Used for the Amplification Targets in the IC Sequences and in Genomic DNA (gDNA)

Primer pair	PCR products amplified from the templates					
	Non converted gDNA	Bisulfite converted gDNA	Bisulfite converted gDNA+pConIC	Bisulfite converted (pUnIC+gDNA)*	pConIC without bisulfite treatment	pUnIC without bisulfite treatment
CF-F and CF-R	neg	neg	pos	pos	pos	pos
MP- <i>SHOX2</i> -F and MP- <i>SHOX2</i> -R	neg	neg	pos	pos	pos	neg
Me- <i>SHOX2</i> -F and Me- <i>SHOX2</i> -R	neg	pos	pos	pos	neg	neg

neg, negative; pos, positive; \*, genomic DNA was mixed with linearized pUnIC and converted by bisulfite.

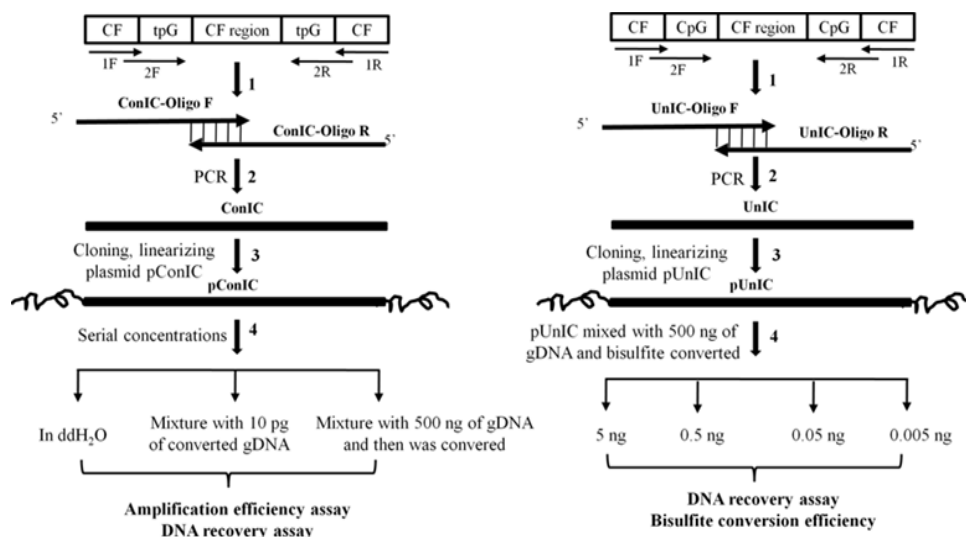


Figure 1. Design of the ConIC and UnIC. Four oligonucleotides containing CF sequence and CpG sequence of the *SHOX2* promoter were extended and amplified by PCR resulting in the ConIC and UnIC sequences (1) that were cloned into pTZ57R/T vector (2); the resulting pConIC and pUnIC plasmids were then linearized (3). The ConIC and UnIC fragments were identical, excepted that all C in the UnIC were replaced by t in the ConIC. A serial concentrations of pConIC and pUnIC were mixed with genomic DNA (gDNA) for evaluating the amplification and bisulfite conversion efficiencies (4). 1F/1R and 2F/2R: CF-F/R and MP-*SHOX2*-F/R primer sets used for evaluating DNA recovery and conversion efficiency. CF region: cytosine-free region.

presented in Table 1.

*Calculation of qPCR amplification efficiency using the ConIC sequence*

The UnIC sequence contain cytosines that were converted into uracils by bisulfite treatment and subsequently to thymines after the first run in PCR reaction. The ConIC sequence, in which all cytosines had

been already converted into thymines, would be identical to the PCR amplified UnIC sequence if this latter were fully converted (100% conversion efficiency). Thus, the ConIC sequence can be used as a calibrator in real time PCR assay to calculate the DNA recovery and bisulfite conversion efficiencies of the UnIC. The two primer sets used here, named CF-F/R and MP-*SHOX2*-F/R, are directed against the CF (reference) and *SHOX2* (target)

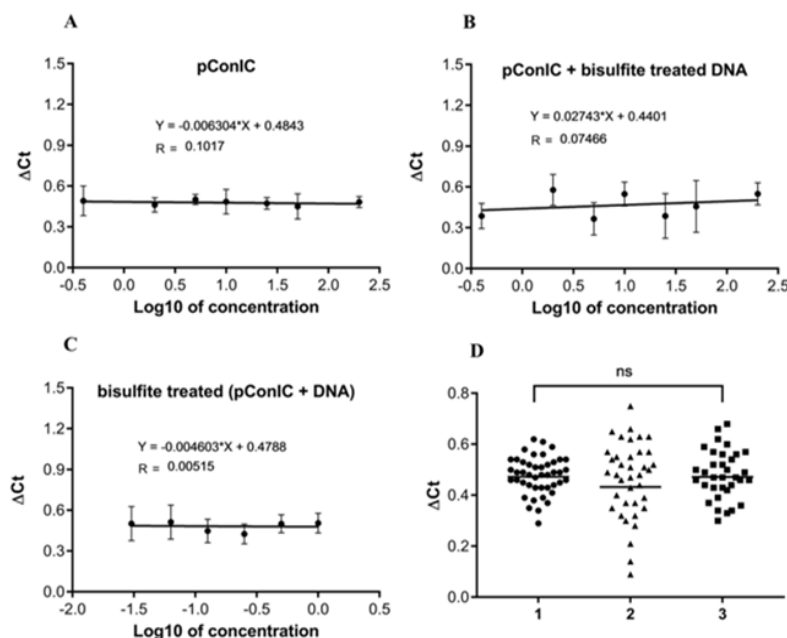


Figure 2. Validation of the  $\Delta\Delta C_t$  Method to Quantify pConIC Amplification Efficiency. (A-C) The amplification efficiency of the CF and *SHOX2* targets on the same pConIC template was examined using qPCR. Serial dilution of the pConIC were amplified by 2 primer sets named CF-F/R and MP-*SHOX2*-F/R, specific to CF and *SHOX2* sequences, respectively. The  $\Delta C_t$  ( $C_{T_{MP-SHOX2}} - C_{T_{CF}}$ ) was calculated for each pConIC dilution. (D) Difference in  $\Delta C_t$  value analysed by one-way ANOVA statistical method when using qPCR templates as following: pConIC alone (1), pConIC mixed with 10 ng of bisulfite-treated genomic DNA (2) and 1 ng pConIC mixed with 500 ng of genomic DNA then subsequently treated by bisulfite (3). ns: non significant.

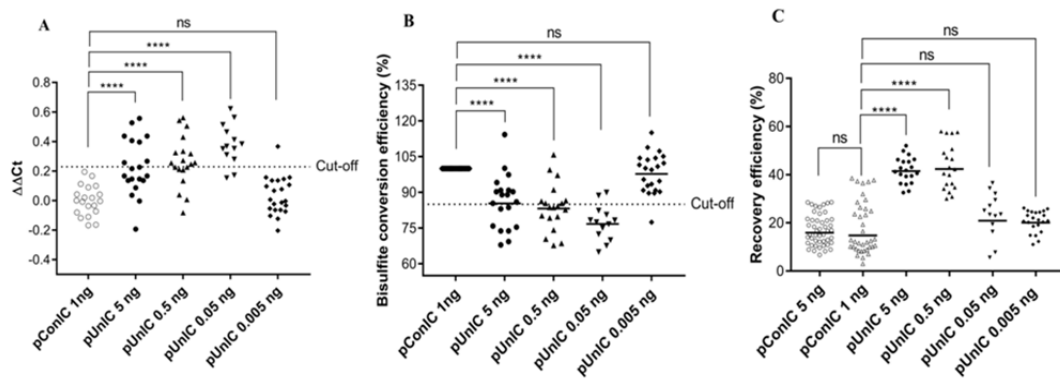


Figure 3. Monitoring of DNA Recovery and Bisulfite Conversion Efficiencies in *SHOX2* Methylation Assay. (A) Serial concentrations of the pUnIC was mixed with 500 ng of genomic DNA and subsequently bisulfite converted in order to determination of the optimal pUnIC concentration. The conversion efficiency where determined by the  $\Delta\Delta CT$  method using pConIC as calibrator for 100% conversion. A cut-off value was determined in order to dichotomize the quantitative  $\Delta\Delta CT$  value into a qualitative result (successful or unsuccessful conversion). (B) Bisulfite conversion efficiency using serial pUnIC concentrations expressed as % as compared to the calibrator pConIC set at 100% conversion. (C) DNA recovery determined by the  $2^{-\Delta CT}$  method and expressed as % compared to input DNA amount. All statistical analyses were performed using one-way ANOVA. ns: non significant; \*\*\*\*  $p < 0.0001$ . The horizontal bars represent the mean.

sequences, respectively, located on the same IC templates (either ConIC or bisulfite-treated UnIC). In order to use the ConIC as a calibrator, real time PCR reactions should have the same amplification efficiency, meaning the  $\Delta CT$  ( $CT_{MP-SHOX2} - CT_{CF}$ ) should stay unchanged with different pConIC concentrations (Livaket al., 2001). In this case, the  $\Delta\Delta CT$  method can be applied to calculate the relative DNA quantification.

In order to quantify the amplification efficiency using pConIC, serial dilutions of the linearized pConIC plasmid, ranging from 200 pg and 0.4 pg (equivalent to  $5 \times 10^7$  to  $10^5$  copies), were prepared either in DNA-free water (group 1) or in a solution containing 10 ng of bisulfite-treated genomic DNA (group 2). In a third group, 1 ng of linearized pConIC plasmid was mixed with 500 ng of genomic DNAs, then the mix was subsequently treated with bisulfite, and serial dilutions were prepared for qPCR templates. Subsequently, real time PCR was performed on all samples of the three groups using the CF-F/R and MP-*SHOX2*-F/R specific primer sets. In Figure 2, the  $\Delta CT$  is plotted according to the serial pConIC concentrations, showing slopes of 0.006, 0.027 and 0.005 for groups 1, 2,

3, respectively, thus meaning that  $\Delta CT$  stayed unchanged along with pConIC concentrations in the three groups (Livaket al., 2001). In group 3 where pConIC was bisulfite treated, we observed a slight and constant decrease in amplification efficiency of both CF and *SHOX2* sequences probably due to complexes of single-stranded AT rich DNAs created during the conversion process (Pughet al., 2008; Underhill et al., 2016), but since this decrease concerned both CF and *SHOX2* sequences, the  $\Delta CT$  stayed unchanged. Therefore, the  $\Delta\Delta CT$  calculation was suitable to analyze the data. Moreover, no significant difference in  $\Delta CT$  values was observed among the three experimental groups, as assessed by one-way ANOVA (Figure 2D). This result indicated that ConIC alone can be used as a calibrator for relative quantification of UnIC recovery and bisulfite conversion efficiencies in further analysis.

*Calculation of bisulfite conversion and DNA recovery efficiencies using ConIC as calibrator and UnIC as indicator*

After having validated the use of ConIC as a calibrator for 100% bisulfite conversion, we proceeded to determine

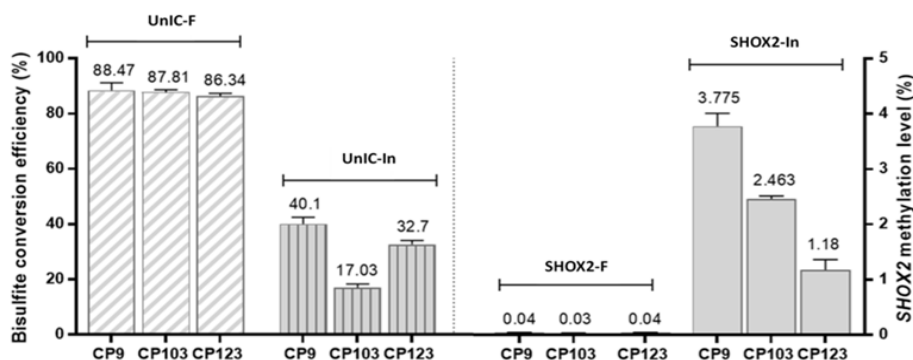


Figure 4. Overestimation of *SHOX2* Methylation Level due to Its Incomplete Bisulfite Conversion. On the left: Bisulfite conversion efficiency was determined using the UnIC indicator in methylation assay performed on 3 patient samples (CP9, CP103 and CP123) previously treated by bisulfite for the recommended time (UnIC-F) or for half-time (UnIC-In). On the right: *SHOX2* methylation level in the respective samples sufficiently (*SHOX2*-F) or insufficiently (*SHOX2*-In) bisulfite treated.

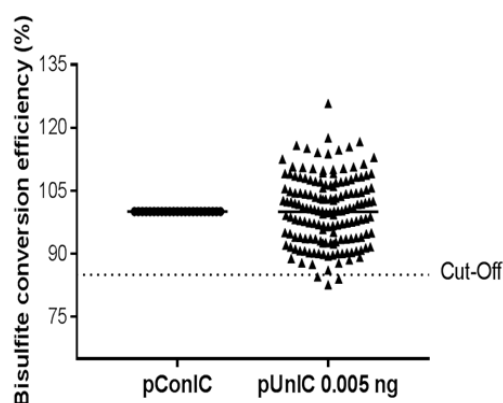


Figure 5. Determination of Bisulfite Conversion Efficiency. The bisulfite conversion efficiency of the DNA extracted from the samples was determined using the UnIC indicator. The horizontal bars represent the mean.

the optimal pUnIC concentration to use in methylation assays for the highest bisulfite conversion efficiency. To do so, we prepared 4 pools of PCR templates by mixing the linearized pUnIC at concentrations of 0.005 ng, 0.05 ng, 0.5 ng and 5 ng (equivalent to  $10^6$ - $10^9$  copies) with a constant amount of genomic DNA equal to 500 ng. This concentration of genomic DNA was chosen according to the manufacturer's instruction for input DNA. Replicates of these mixed DNA templates were then treated with bisulfite (the smallest group had  $n=13$ ). UnIC bisulfite conversion efficiency was calculated by the  $\Delta\Delta CT$  method using the ConIC as calibrator having 100% conversion. Significant difference was observed between pConIC and the groups using from 5 ng to 0.05 ng pUnIC, but not with the group using 0.005 ng, as assessed by one-way ANOVA (Figure 3A); thus, we chose 0.005 ng pUnIC as the optimal concentration. Since the  $\Delta\Delta CT$  value using pConIC template ranged from -0.2 to 0.2 (around 0 value) as observed with  $n=20$  replicates (probably caused by complex of single stranded AT rich DNAs as mentioned above), in order to use pConIC as the calibrator for 100% conversion, it was necessary to determine a cut-off value to dichotomize quantitative  $\Delta\Delta CT$  values into qualitative result (successful or unsuccessful conversion). Such a cut-off  $\Delta\Delta CT$  value was chosen so that 5% of  $\Delta\Delta CT$  values obtained with pUnIC are considered as false positive (unsuccessful) conversion. As showed in Figure 3A, the cut-off  $\Delta\Delta CT$  value of 0.23 was chosen so that all pConIC replicates and 21/22 (>95%) pUnIC replicates were classified as successful conversion. According to this cut-off in  $\Delta\Delta CT$ , samples with conversion efficiency of pUnIC more than 85% calculated on the basis of the  $\Delta\Delta CT$  method with assumption that pConIC having 100% conversion were classified as successful conversion (Figure 3B). The proportion of unmethylated cytosines fully converted and unconverted using 0.005 ng to 5 ng of pUnIC were quantified by direct sequencing (Supplementary Figure 2). In addition, we also assessed DNA recovery efficiency (compared to input DNA) via the relative quantification using  $2^{-\Delta CT}$  method. Between 2 concentrations of pConIC, 5 ng and 1 ng, no significant difference in DNA recovery was observed, both showing

around 17% recovery. There was neither significant recovery difference between 1 ng pConIC and 0.05 ng or 0.005 ng pUnIC, the latter showed around 20% -24% recovery after bisulfite conversion (Figure 3C). On the contrary, higher amount of pUnIC input (5 ng and 0.5 ng) showed higher recovery rates of around 41% -44%, but since they have been shown to be incompletely bisulfite converted, these pUnIC input quantities were not chosen for further analysis. Thus, based on these results, 0.005 ng pUnIC can be used as a reliable internal control to quantitatively monitor both bisulfite conversion efficiency and DNA recovery rate in *SHOX2* methylation assay.

After bisulfite treatment, all unmethylated cytosines in the pUnIC plasmid were converted into uracils. In order to access the PCR amplification efficiency using uracils-containing pUnIC versus thymines-containing pConIC as initial template, 0.005 ng of linearized pUnIC or pConIC where separately mixed with 500 ng of genomic DNA, then the mix was subsequently treated with bisulfite. Subsequently, real time PCR was performed on both samples using the CF-F/R and MP-*SHOX2*-F/R specific primer sets. We found that the CT values are not significantly different ( $p>0.05$ ) (data not shown), indicating that pConIC can be used as representative of the 100% bisulfite converted pUnIC. This result was consistent with previous report describing that DNA polymerases recognise uracil as efficiently as thymine during extension (Heyn et al., 2010).

In order to confirm the performance of pUnIC as an indicator of bisulfite conversion efficiency, aliquots of 0.005 ng of linearized pUnIC were mixed with 500 ng genomic DNA extracted from 3 FFPE samples of patients with non-cancerous pulmonary diseases and then treated with bisulfite following (1) manufacturer's instruction using the EpiTect Bisulfite Kit (Qiagen) or (2) with a 50% reduced time for bisulfite treatment compared to recommended time. qPCR were performed in triplex using MP-*SHOX2*-F/R and Me-*SHOX2*-F/R primer sets, which recognized methylated *SHOX2* targets in pUnIC and in genomic DNA, respectively. All samples sufficiently treated with bisulfite showed bisulfite conversion rates above 86% indicating successful conversion, and 0.03% methylation level of *SHOX2* promoter. In contrast, a disastrous decrease in conversion rate to 17% along with an increase in *SHOX2* methylation level (3.77%) were detected in samples with half-time bisulfite treatment (Figure 4). PCR products amplified by primer sets specific to CF and methylated *SHOX2* promoter were directly sequenced, revealing that the proportion of unconverted cytosines belonging to the pUnIC and *SHOX2* promoter was strictly related to that in DNAs previously bisulfite treated for insufficient time (Supplementary Figure 3). Therefore, using the pUnIC indicator, we were able to detect a false positive *SHOX2* methylation (3.77% instead of 0.03%). Based on these results, 0.005 ng of linearized pUnIC mixed with 500 ng genomic DNA extracted from patient samples was used for further validation of bisulfite conversion of the *SHOX2* promoter from patients with pulmonary disease.

Bisulfite conversion efficiency was performed on 169 FFPE samples: 97 from patients with lung cancer and 72

from patients with non-cancerous lung diseases, using 500 ng genomic DNA mixed with 0.005 ng pUnIC. However, in some samples, DNAs were seriously fragmented and impacted on DNA recovery after bisulfite conversion; thus, out of the 169 samples, only 159 (94 cancer and 65 non-cancerous) yielding sufficiently bisulfite treated DNAs were retained for analysis. Bisulfite conversion efficiency in those 159 DNA samples was evaluated via the pUnIC indicator. Direct sequencing of PCR products amplified by the primer set specific to methylated *SHOX2* showed that all cytosines in the CpG sites remained to be cytosines and the cytosines alone were converted to thymines (data not shown). Taking into account that 2 samples from the control group showed conversion rate (82% and 84%) below the 85% cut-off value (Figure 5), these 2 samples were excluded from further specificity-sensitivity analysis of *SHOX2* methylation. The mean conversion efficiency was estimated to 98.7% (Figure 5). In addition, the pUnIC recovery rate was around 18% compared to 0.005ng input pUnIC (data not shown), which is in agreement with the results previously presented in Figure 3C.

## Discussion

This study aimed at designing an IC system for bisulfite-based methylation qPCR assay in order to simultaneously and quantitatively monitor the DNA recovery and bisulfite conversion efficiencies, 2 parameters that are crucial for the accurate quantification of methylated targets. The IC designed here was validated in the context of quantifying the methylation status of *SHOX2*, known as a valuable biomarker in lung cancer detection (Darwiche et al., 2013; Song et al., 2015). The whole method was then validated on samples from patients with lung cancer and non-cancerous pulmonary diseases.

Both elements of the IC system, ConIC and UnIC, were designed to contain two CpG sequences derived from the *SHOX2* promoter and inserted into a CFF (Figure 1). It is worth noting that double stranded DNA processed through bisulfite conversion became two single strands that are no longer complementary. Here, the UnIC and ConIC sequences correspond actually to the same sequence before and after bisulfite full conversion, respectively. In the IC, nucleotide sites specifically recognized by the CF-F/R and Me-*SHOX2*-F/R primer sets were translocated so that these primers annealed only to targets on the IC but not to those in the genomic DNA (Figure 1 and Supplementary Figure 1). Indeed, using bisulfite treated and non treated genomic DNA as template, no PCR products was amplified by these primer sets (Table 1), indicating that the presence of the IC in DNA samples did not influence on the specific amplification of both endogenous and exogenous targets.

The UnIC insert has a length of 122 bp, which was compatible in length with LMW DNA such as cell-free DNA (cfDNA) of short size DNA (~ 135 bp) in the plasma (Underhill et al., 2016; Walter et al., 2018). The linearized pUnIC plasmid has a length of 3000 bp, which makes it suitable to be used as a control in bisulfite treatment of HMW DNA (Clark et al., 2006).

Therefore, the design of UnIC provides a flexible choice for an IC to be used to evaluate bisulfite conversion of target sequences of different sizes in HMW/LMW DNA samples. Indeed, a 131 bp PCR fragment has been custom-designed and used as LMW-DNA for comparative analysis of different commercial kits for bisulfite conversion efficiency (Ørntoft et al., 2017), whereas linearized plasmids containing target sequences have been used for the evaluation of DNA recovery of bisulfite-converted sequences (Munson et al., 2007). Artificial DNA with different length derived from Lambda DNA was also spiked with genomic DNAs in order to perform comparison of different bisulfite conversion commercial kits (Leontiou et al., 2015). However, all these artificial DNAs published so far did not contain in the same template both CF sequence and the specific CpG sequence of interest, which is the case of the IC designed in our study.

Since in the ConIC sequence all the cytosines had been already converted to thymines, it can be used to evaluate DNA recovery after purification without bisulfite conversion. This recovery rate was around 17% both using 5ng or 1ng input pConIC, which is consistent with post-bisulfite recovery of the endogenous CFF described previously (Holmes et al., 2014; Ørntoft et al., 2017). Using a PCR product mimicking cytosine-free L1RE1 sequence as reference, Bryzgunova et al. (2013) also indicated a recovery rate around 20% of endogenous L1RE1 in the input genomic DNA. Regarding the effect of bisulfite conversion on DNA recovery, it has been mentioned that low DNA recovery rate is rather due to DNA lost by fragmentation, depyrimidination or inefficient purification, than to chemical conversion (Munson et al., 2007; Tanaka and Okamoto, 2007). Uniform recovery of fully bisulfite converted DNA across a wide range of DNA input using different kits has been reported (Ørntoft et al., 2017). In our experimental system, 0.005ng of pUnIC input that had been successfully bisulfite converted (98.7% efficiency as assessed in Figure 5) showed a similar recovery rate to pConIC (Figure 3C). Using higher pUnIC amount (0.5 ng- 5 ng), we obtained higher recovery rate around 41% - 44%, but since the remaining DNA also included incompletely converted DNA (as assessed in Figure 3B and Supplementary Figure 2), these pUnIC concentrations are clearly not optimal. Indeed, the incomplete conversion of unmethylated cytosines may lead to false positive result thus compromising the specificity and sensitivity of methylation assay (Lan et al., 2014; Owa et al., 2018).

Moreover, our results also revealed that the copy number of particular target impacted on bisulfite conversion efficiency, with high amount of templates ( $10^7$  -  $10^9$  copies corresponding to 0.05 ng - 5 ng pUnIC) leading to incomplete conversion (Figure 3B). So far, most qPCR and dPCR assays for evaluating bisulfite conversion efficiency have been performed on single copy sequence such as endogenous *ACTB*, *MYOD1* genes whose maximal number is around  $10^6$  copies corresponding to maximal amount (2 µg) of input DNA recommended by different kits for bisulfite conversion (Holmes et al., 2014; Ørntoft et al., 2017). In the context where the methylation profile of repetitive elements can be used to estimate global DNA

methylation for biomarker determination (Walter et al., 2018; Bryzgonova et al., 2013; Yanget al., 2004), our result highlighted a considerable attention when performing qPCR assay on such repetitive methylated sequences presented in high copy number in the human genome.

The pUnIC at 0.005 ng (106 copies) was used as an indicator of bisulfite conversion efficiency of the *SHOX2* sequence, with a mean value reaching 98.7% conversion. The minute remnant of non-converted/incompletely converted DNA could be partly accounted for the small discrepancy in all the values of pUnIC conversion efficiency that are considered as successful in our system (85% - 100%). Nonetheless, bisulfite conversion rate above the 85% cut-off value have been observed across numerous analytical as well as clinical samples: 88% - 100% conversion has been described for DNA from plasma samples (Ørntoft et al., 2017), as well as 90–100% conversion efficiency of endogenous or artificial sequences reported by numerous evaluations of commercial kits (Holmes et al., 2014; Leontiou et al., 2015; Ørntoft et al., 2017; Tierling et al., 2018). Thus, our IC system demonstrated the reliability and confidence of this calibrator-indicator couple in monitoring bisulfite conversion efficiency. Using this indicator, a false positive *SHOX2* methylation (3.77% methylation instead of 0.03%) was detected in samples that are bisulfite treated for an insufficient time (Figure 4), thus again highlighting the full bisulfite conversion as a crucial requirement in methylation quantification. Particularly, the UnIC used as reference contained sequences of the *SHOX2* promoter; avoiding the interference on bisulfite conversion due to sequence complexity, GC content, secondary structure elements and even a single cytosine in a specific nucleotide context (Harrison et al., 1998; Genereux et al., 2008), thus ensuring more accurate evaluation of bisulfite conversion of the *SHOX2* target. Recently, a panel of DNA methylation markers has shown great increase in sensitivity and specificity in the diagnosis and prognosis analyses of different types of cancers (Mojtabanezhad et al., 2018; Zhanget al., 2017; Ren et al., 2017; Yiet al., 2013). In this context, the ConIC/UnIC system could be a technical asset, offering the customizable insertion of different sequences specific to different methylated targets, becoming a multi-indicator for quantitatively monitoring the bisulfite conversion of multiple target genes in one experiment.

There are some limitations inherent to our study design that should be noted. First, log<sub>10</sub> fold serial dilution of the UnIC was performed in this study (5 ng to 0.005 ng), while a more accurate investigation of the impact of pUnIC copy number on bisulfite conversion efficiency may require a log<sub>2</sub> fold serial dilution. Moreover, evaluation of bisulfite conversion efficiency was limited to small FFPE samples while most of the analyses so far were carried out on blood plasma or specimens from bronchial aspirate or washing during bronchoscopy (Schmidt et al., 2010; Kneip et al., 2011; Dietrich et al., 2012). In our future studies, we intend to increase the samples number as well as to collect samples from different specimen types, from confirmed versus suspicious cancer patients, in order to validate the clinical diagnostic value of *SHOX2* methylation.

In conclusion, to summarize, this study has proposed a customizable internal control using the ConIC/UnIC as calibrator/indicator to quantify simultaneously and accurately the DNA recovery and bisulfite conversion efficiencies of individual sequence as well as whole genome in methylation assays. We have shown that the ConIC/UnIC designed for *SHOX2* target gene demonstrated a bisulfite conversion rate reaching 98.7% for this sequence; thus, ensuring sufficient specificity and sensitivity in subsequent analyses for its methylation status. These encouraging results prompt us to quantitatively assess *SHOX2* methylation in large specimens types collected from different non-cellular and cellular samples. In long term, we are headed at the optimization of the ConIC/UnIC system so that it can be applied in methylation assays specific to multi-targets or repetitive sequences for biomarker development and clinical applications.

## Acknowledgements

The authors would like to thank Nguyen T Trang, Pham AT Duong, Nguyen T Nga, Tran TC Van at the Biomedical Lab, Faculty of Biology, VNU University of Science, Hanoi, Vietnam, and the Company BCE Vietnam Co, Ltd for supporting 7500 Real-time PCR instrument. This study was financially supported by Grant 106-YS.06-2015.07 from the Ministry of Science and Technology, Viet Nam.

## Conflict of interest

The authors Vo Thi Thuong Lan, Vu Lan Trang, Nguyen Thuy Ngan, Ho Van Son and Nguyen Linh Toan declare that they have no conflict of interest.

## References

- Bock C, Halbritter F, Carmona FJ, et al (2016). Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat Biotechnol*, **34**, 726-37.
- Bryzgonova O, Laktionov P, Skvortsova T, et al (2013). Efficacy of bisulfite modification and recovery of human genomic and circulating DNA using commercial kits. *Eur J Mol Biol*, **1**, 1-8.
- Clark SJ, Harrison J, Paul CL, Frommer M (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*, **22**, 2990-7.
- Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M (2006). DNA methylation: Bisulphite modification and analysis. *Nat Protoc*, **1**, 2353-64.
- Darwiche K, Zarogoulidis P, Baehner K, et al (2013). Assessment of *SHOX2* methylation in EBUS-TBNA specimen improves accuracy in lung cancer staging. *Ann Oncol*, **24**, 2866-70.
- Dietrich D, Kneip C, Raji O, et al (2012). Performance evaluation of the DNA methylation biomarker *SHOX2* for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. *Int J Oncol*, **40**, 825-32.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res*, **36**, e105.
- Dugué PA, Bassett JK, Joo JE, et al (2018). DNA methylation-based biological aging and cancer risk and survival: Pooled



- analysis of seven prospective studies. *Int J Cancer*, **142**, 1611-9.
- Genereux DP, Johnson WC, Burden AF, Stöger R, Laird CD (2008). Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Res*, **36**, e150.
- Harrison J, Stirzaker C, Clark, SJ (1998). Cytosines adjacent to methylated CpG sites can be partially resistant to conversion in genomic bisulfite sequencing leading to methylation artifacts. *Anal Biochem*, **264**, 129-32.
- Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, Meyer M (2010). Road blocks on paleogenomes - polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Res*, **38**, e161.
- Holmes EE, Jung M, Meller S, et al (2014). Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One*, **9**, e93933.
- Ji L, Sasaki T, Sun X, et al (2014). Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Front Genet*, **5**, 341.
- Kint S, Spiegelare WD, Kesel JD, Vandekerckhove L, Crieckinge WV (2018). Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PLoS One*, **13**, e0199091.
- Kneip C, Schmidt B, Seegebarth A, et al (2011). *SHOX2* DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *J Thorac Oncol*, **6**, 1632-8.
- Koch A, Joosten SC, Feng Z, et al (2018). Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncol*, **15**, 459-66.
- Kurdyukov A, Bullock M (2016). DNA methylation analysis: choosing the right method. *Biology*, **5**, pii: E3.
- Lan TTV, Ha TN, Uyen QN, et al (2014). Standardization of the methylation-specific PCR method for analysing BRCA1 and ER methylation. *Mol Med Reports*, **9**, 1844-50.
- Leontiou CA, Hadjidaniel MD, Mina P, et al (2015). Bisulfite Conversion of DNA: Performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing. *PLoS One*, **10**, e0135058.
- Liu C, Jiao C, Wang K, Yuan N (2018). DNA methylation and psychiatric disorders. *Prog Mol Biol Transl Sci*, **157**, 175-232.
- Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, **25**, 402-8.
- Mojtabanezhad SA, Yassi M, Nouraei M, et al (2018). The importance of stool DNA methylation in colorectal cancer diagnosis: A meta-analysis. *PLoS One*, **13**, e0200735.
- Munson K, Clack J, Kupsik KL, Smith SS (2007). Recovery of bisulfite-converted genomic sequences in the methylation-sensitive qPCR. *Nucleic Acids Res*, **35**, 2893-2903.
- Olova N, Krueger F, Andrews S, et al (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol*, **19**, 33.
- Ørntoft WM-B, Jensen SØ, Hansen TB, Bramsen JB, Andersen CL (2017). Comparative analysis of 12 different kits for bisulfite conversion of circulating cell-free DNA. *Epigenetics*, **12**, 626-36.
- Owa C, Poulin M, Yan L, Shioda T (2018). Technical adequacy of bisulfite sequencing and pyrosequencing for detection of mitochondrial DNA methylation: Sources and avoidance of false-positive detection. *PLoS One*, **13**, e0192722.
- Pugh TJ, Delaney D, Farnoud N, et al (2008). Impact of whole genome amplification on analysis of copy number variants. *Nucleic Acids Res*, **36**, e80.
- Rand K, Qu W, Ho T, Clark SJ, Molloy P (2002). Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods*, **27**, 114-20.
- Ren M, Wang C, Sheng D, et al (2017). Methylation analysis of *SHOX2* and *RASSF1A* in bronchoalveolar lavage fluid for early lung cancer diagnosis. *Ann Diagn Pathol*, **27**, 57-61.
- Schmidt B, Liebenberg V, Dietrich D, et al (2010). *SHOX2* DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer*, **10**, 600.
- Senturk E, Cohen S, Dottino PR, Martignetti JA (2010). A critical re-appraisal of BRCA1 methylation studies in ovarian cancer. *Gynecol Oncol*, **119**, 376-83.
- Silmon de Monerri NC, Kim K (2014). Pathogens hijack the epigenome: a new twist on host-pathogen interactions. *Am J Pathol*, **184**, 897-911.
- Song L, Yu H, Li Y (2015). Diagnosis of lung cancer by *SHOX2* gene methylation assay. *Mol Diagn Ther*, **19**, 159-67.
- Sun B, Hu L, Luo ZY, et al (2016). DNA methylation perspectives in the pathogenesis of autoimmune diseases. *Clin Immunol*, **164**, 21-7.
- Tanaka K, Okamoto A (2007). Degradation of DNA by bisulfite treatment. *Bioorg Med Chem Lett*, **17**, 1912-5.
- Tierling S, Schmitt B, Walter J (2018). Comprehensive evaluation of commercial bisulfite-based DNA methylation kits and development of an alternative protocol with improved conversion performance. *Genet Epigenet*, **10**, 1-7.
- Underhill HR, Kitzman JO, Hellwig S, et al (2016). Fragment length of circulating tumor DNA. *PLoS Genetics*, **12**, e1006162.
- Walter RFH, Rozynek P, Casjens S, et al (2018). Methylation of LIRE1, RARB, and RASSF1 function as possible biomarkers for the differential diagnosis of lung cancer. *PLoS One*, **13**, e0195716.
- Yang AS, Estecio M, Doshi K, et al (2004). A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res*, **32**, e38.
- Yi JM, Guzzetta AA, Bailey VJ, et al (2013). Novel methylation biomarker panel for the early detection of pancreatic cancer. *Clin Cancer Res*, **19**, 6544-55.
- Zhang C, Yu W, Wang L, et al (2017). DNA methylation analysis of the *SHOX2* and *RASSF1A* panel in bronchoalveolar lavage fluid for lung cancer diagnosis. *J Cancer*, **8**, 3585-91.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.