Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry

Mathias Ehrich*, Matthew R. Nelson*, Patrick Stanssens†, Marc Zabeau†, Triantafillos Liloglou§, George Xinarianos§, Charles R. Cantor*,§, John K. Field*,§, and Dirk van den Boom*§

*SEQUENOM, Inc., 3595 John Hopkins Court, San Diego, CA 92121; †Methexis Genomics NV, Technologiekpark 4, B-9052 Zwijnaarde, Belgium; ‡Roy Castle Lung Cancer Research Programme, University of Liverpool Lung Cancer Research Centre, 200 London Road, Liverpool L3 9TA, United Kingdom; and §Clinical Dental Sciences, University of Liverpool, Liverpool L69 3BX, United Kingdom

Contributed by Charles R. Cantor, September 10, 2005

Methylation is one of the major epigenetic processes pivotal to our understanding of carcinogenesis. It is now widely accepted that there is a relationship between DNA methylation, chromatin structure, and human malignancies. DNA methylation is potentially an important clinical marker in cancer molecular diagnostics. Understanding epigenetic modifications in their biological context involves several aspects of DNA methylation analysis. These aspects include the de novo discovery of differentially methylated genes, the analysis of methylation patterns, and the determination of differences in the degree of methylation. Here we present a previously uncharacterized method for high-throughput DNA methylation analysis that utilizes MALDI-TOF mass spectrometry (MS) analysis of base-specifically cleaved amplification products. We use the IGF2/H19 region to show that a single base-specific cleavage reaction is sufficient to discover methylation sites and to determine methylation ratios within a selected target region. A combination of cleavage reactions enables the complete evaluation of all relevant aspects of DNA methylation, with most CpGs represented in multiple reactions. We successfully applied this technology under high-throughput conditions to quantitatively assess methylation differences between normal and neoplastic lung cancer tissue samples from 48 patients in 47 genes and demonstrate that the quantitative methylation results allow accurate classification of samples according to their histopathology.

In recent years it has become apparent that genetic information is stored not only in the arrangement of four nucleotide bases, but also in the covalent modification of selected bases. In mammals, the main modification is methylation of cytosine. The covalent addition of methyl groups to cytosine in CpG dinucleotides is catalyzed by the DNA methyltransferase enzyme family (1, 2). In the human genome, CpG dinucleotides are generally underrepresented and are concentrated in distinct areas called CpG islands. A large proportion of these CpG islands are found in the promoter regions of genes. The conversion of cytosine to 5-methylcytosine in promoter-associated CpG islands can cause changes in chromatin structure and usually results in transcriptional silencing of the associated gene. These events have been linked to mammalian development, imprinting, X-chromosome inactivation (3), suppression of parasitic DNA (4), and cancer etiology (5–8). Cancer research, in particular, may benefit from DNA methylation analysis, because changes in the methylation status of nucleosomal DNA promise to be a powerful marker in the early detection of neoplastic events (9–11).

Although DNA methylation analysis is a rapidly developing field, studies demonstrating its usefulness in a clinical environment are still scarce. This situation is partly due to technical limitations in current DNA methylation analysis. Several analysis techniques are available, but each suffers several shortcomings. A cost-effective, high-throughput DNA methylation analysis tool would help tremendously to understand the relationship between the status of the tens of thousands of potential methylation sites and disease and to develop diagnostic and prognostic methylation markers.

An excellent review of current methylation analysis techniques by Laird (11) has recently been published. It concluded that “no one technique or general approach is superior, as the competing goals of quantitative accuracy, sensitive detection, and high local or global informational content, compatibility with formalin fixed tissues and compatibility with automation are not found in a single technique.” In this study, we introduce a previously uncharacterized approach for methylation analysis that utilizes base-specific cleavage of nucleic acids. Samples are analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) (12). This method permits the high-throughput identification of methylation sites and their semiquantitative measurement at single or multiple CpG positions and, therefore, achieves objectives mutually exclusive in many other currently used methods.

Materials and Methods
Bisulfite Treatment. Bisulfite treatment was performed according to a protocol from Paulin et al. (13). Details are provided in Supporting Materials and Methods, which is published as supporting information on the PNAS web site.

PCR and in Vitro Transcription. The IGF2/H19 region (chr11: 1,983,678–1,984,097, assembly July 2003) was PCR amplified from bisulfite-treated human genomic DNA by using primers that incorporate the T7 5′-TAG TAA TAC GAC TCA CTA TAG GGA GA] promoter sequence. Two sets of primers were designed to incorporate the T7 promoter sequence either to the forward 5′-TAG TAA TAC GAC TCA CTA TAG GGA GAA GGCTGT TAG TTT TTA TTT TTT TTT TAA T-3′; 5′-AGG AAG AGA GAA GAA CCA CTA TCT CCC CTC AAA AAA-3′) or to the reverse 5′-AGG AAG AGA GGT TAG TTT TTT TTT TTT TAAA T-3′; 5′-AGG TAA TAC GAC TCA CTA TAG GGA GAA GGC GTA CCA CTA TCT CCC CTC AAA AAA-3′) strand. Alternatively, we cloned the derived PCR product into a pGEM-T vector system (Promega) and reamplified from the cloned DNA.

A description of the regions used for methylation analysis in non-small cell lung cancer can be found as Data Set 1, which is published as supporting information on the PNAS web site.

Details of the PCR conditions and transcription reaction have been described in ref. 12. They are also supplied in Supporting Materials and Methods.
Statistical Methods. Relative methylation was compared between normal and tumor samples by using the Wilcoxon signed-rank test, a nonparametric counterpart of the paired t test. The two-way hierarchical cluster analysis clustered the 96 tissue samples and 76 most variable CpG fragments (variance >0.02) based on pairwise Euclidean distances and the complete linkage clustering algorithm. This clustering was carried out by using a modified version of the heatmap.2 function of the GREGMISC package by using the R statistical environment. The tree-based classifier was found by using the recursive partitioning package RPART in R. A complete six-node tree was pruned to the four-node tree that resulted in the lowest 10-fold cross-validation error.

Results

Assay Concept. Our DNA methylation analysis by MALDI-TOF MS employs base-specific cleavage of single-stranded nucleic acids. This approach has already proved to be a powerful DNA sequence analysis tool (12, 14–16). The idea is to generate a PCR amplification product from bisulfite-treated DNA, which is transcribed in vitro into a single-stranded RNA molecule and, subsequently, cleaved base specifically by an endoribonuclease. The method resembles earlier approaches to sequencing nucleic acids developed by Maxam-Gilbert and Sanger. The conversion of unmethylated cytosine to uracil during bisulfite treatment will generate base-specific cleavage products that reflect underlying methylation patterns and that can be readily analyzed by MALDI-TOF MS. A schematic of the assay concept is shown in Fig. 1.

Because the method was intended for discovery of methylated CpGs and the relative quantitation of methylated versus nonmethylated DNA copies, we avoided preferential amplification of methylated or nonmethylated DNA by excluding any CpG sites from the primer region.

Because earlier experiments revealed that RNase A cleavage is more robust and does not produce side products such as cyclic phosphates (15), we used the cleavage method introduced by Stanssens et al. (12), which uses RNase A and incorporation of either noncleavable dCTP or dTTP to achieve U or C specificity, respectively. The combination of RNase A cleavage with transcription from forward and reverse strands of the PCR amplicon allows base-specific cleavage after each of the four bases. Cleavage at A and G are mimicked by cleavage at C and U of the reverse strand.

Methylation affects the base-specific cleavage mass signal pattern through C → T sequence changes introduced into the bisulfite-treated genomic DNA. The exact change in the mass signal pattern depends on the cleavage scheme and can be classified into three categories: (i) methylation can introduce new cleavage sites resulting in new, shorter products; (ii) methylation can lead to replacement of an existing cleavage site with a noncleavable nucleotide and, therefore, connect two existing fragments together resulting in a new, longer product; and (iii) methylation can generate a sequence change in an existing cleavage product that does not affect cleavage but generates a mass shift (see the example displayed in Fig. 1 and Table 1).

The most effective discrimination between fully methylated and nonmethylated template DNA can be achieved in reactions such as C-specific cleavage on the forward strand, where every methylated CpG introduces a cleavage site. A completely nonmethylated template results in an RNA transcript that does not have any cleavable nucleotides and, hence, does not generate any cleavage products. A methylated template, on the contrary, will result in an RNA transcript that contains several cleavable nucleotides and, hence, generates defined cleavage patterns. This distinct difference allows rapid and sensitive discovery of methylation sites.

Quantitative analysis of DNA mixtures with variable ratios between methylated and nonmethylated template DNA can best be achieved with reactions where methylation introduces a mass shift within the affected cleavage products. Here nonmethylated and methylated CpG sites will generate status-specific mass signals. Comparing their intensities allows the determination of the ratio of methylated and nonmethylated DNA. A combination of cleavage reactions can be used when complete identification of the methylation status and the relative quantitation of each CpG site within a given amplicon is desired.

Assay Applications: Analysis of IGF2/H19 as a Model System. We tested methylation analysis by MALDI-TOF MS by using the IGF2/H19 imprinted region on chromosome 11p15.5. Changes in the methylation pattern of the IGF2/H19 locus are associated with Beckwith–Wiedemann syndrome and Wilms tumor. The methylation pattern in this region has been described and examined by several authors and, therefore, provides an ideal test case (17–20). Vu et al. (19) performed detailed bisulfite-sequencing examinations in several tissues of adult and fetal specimens. It has been shown that only the paternal allele of the IGF2/H19 region is methylated in adult blood samples.

The model amplicon used was a 416-bp PCR product (chr11: 1,975,945–1,976,360) derived from the IGF2/H19 region, which includes 26 CpG sites. We analyzed the top strand of this region according to Vu et al. (19). Twenty-five of the 26 CpG sites were described as frequently methylated in the paternal allele, and one methylated CpG site was also reported. For method development, we initially amplified the bisulfite-treated genomic DNA and
then cloned the amplification products to exclude bisulphite treatment-related process variance. Later, we also applied our method to direct analysis of bisulphite-treated genomic DNA without cloning, in particular, for experiments exploring the quantitative capabilities of our method.

**Discovery of Genomic Methylation Sites.** The most dramatic differences in mass signal patterns are produced by C-specific cleavage of the forward PCR strand. Only template strands in which methylation has protected the cytosine from conversion into uracil will have cleavage sites in the amplification product. These sites yield cleavage product lengths determined by the distances between neighboring methylated CpGs. An unmethylated template generates an amplification product devoid of cytosine and, therefore, yields no cleavage products. The resulting uncleaved, full-length transcription product is too long to be detected by MALDI-TOF MS. To discriminate between an unmethylated template DNA and a failed reaction, we designed primer tags that are transcribed only after successful PCR reactions and result in specific mass signals confirming the success of the reaction.

Fig. 2a shows the mass spectra of unmethylated and fully methylated DNA after the forward transcript is cleaved C specifically. The mass spectrum derived from methylated template shows signals arising from the expected methylation sites. Each mass signal represents two CpG sites (cleavage at the beginning of the fragment and at the end), and most CpG sites are therefore represented by two cleavage products. The nonmethylated template creates a mass spectrum devoid of any methylation-associated signals, except for a mass signal representing a primer-related cleavage product. The marked difference between the mass signal pattern of methylated and unmethylated DNA in the C cleavage reaction is ideal for discovery of methylation and for discrimination between methylated and unmethylated samples.

We used the unique mass signals representing methylation events to test the sensitivity of detecting methylated DNA in mixtures. For this purpose, we cloned PCR products derived from the IGF2/H19 region and selected fully methylated and nonmethylated clones. We mixed the DNA derived from these clones in different ratios (fully methylated template DNA, a mixture of 50:50 methylated/unmethylated DNA, 25% methylated DNA, 5% methylated DNA and no methylated DNA at all) and carried out an analysis of these mixtures. Fig. 2b shows an overlay of spectra (mass window from 3,000 to 5,000 Da) resulting from C-specific cleavage of the IGF2/H19 region. Signals representing methylation events at CpGs 6–8, 10–12, 15, and 16 of the IGF2/H19 region are shown. The results indicate that 5% methylated DNA can be detected in a mixture without performing methylation-specific PCR. The concurrence of multiple mass signals, when several CpGs in the target region are methylated, increases the reliability of analysis.

**Relative Methylation Analysis.** Although the C-specific cleavage provides a suitable means for discovery and sensitive detection of methylation, it cannot be directly used to quantify the relative amount of methylated DNA in an individual sample. The intensity of a mass signal can only be used for quantification if a nearby mass signal exists for comparison. Determination of the relative amount of methylation is best achieved with cleavage reactions producing a pair of mass signals representing methylated and unmethylated DNA. Methylation-induced C → T changes on the forward PCR strand are represented as G → A changes on the reverse strand. When using C- and U-specific cleavage on the reverse strand, we will not cleave at the methylation site (see Fig. 1; see also Fig. 5, which is published as supporting information on the PNAS web site). Methylation events will appear as mass shifts of 16 Da (mass difference between G and A), or multiples of 16 Da, when more than one CpG is contained in a cleavage product. We refer to cleavage products harboring one or more CpG sites as CpG units.

We can estimate the relative methylation by dividing the peak intensity or area of the methylated DNA by the sum of the intensities or areas of the methylated and nonmethylated DNA peaks. The accuracy and precision of this approach have been studied extensively for estimating allele frequencies in DNA pools by primer extension and MALDI-TOF MS. To discriminate between an unmethylated template DNA and a failed reaction, we designed primer tags that are transcribed only after successful PCR reactions and result in specific mass signals confirming the success of the reaction.

**Table 1. Impact of methylation on base-specific cleavage pattern**

<table>
<thead>
<tr>
<th>Cleavage reaction</th>
<th>Nonmethylated DNA</th>
<th>Type of change</th>
<th>Methylated DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TAAATGTAT</td>
<td>Introduction of cleavage nucleotide</td>
<td>TAAACGTAT</td>
</tr>
<tr>
<td>Forward cleavage</td>
<td></td>
<td>Removal of cleavage nucleotide</td>
<td></td>
</tr>
<tr>
<td>C-specific</td>
<td>TAAAT GTAT</td>
<td>Mass shift</td>
<td>ATACG TTTTA</td>
</tr>
<tr>
<td>U-specific</td>
<td>ATAC T T A</td>
<td>Mass shift</td>
<td>ATACGT T T A</td>
</tr>
</tbody>
</table>

Nonmethylated DNA: TAAATGTAT. Methylated DNA: TAAACGTAT. Theoretical cleavage products of sequence TAAACGTAT, which will be converted through bisulfite treatment to the sequence TAAATGTAT if methylated at the CpG and to TAAACGTAT if not methylated. Cleavage products are shown as DNA species. In actual practice, the DNA sequence is converted to a mixture of ribonucleotides and deoxynucleotides by transcription.
mixing experiments we found a direct correlation between the estimated relative methylation based on the signal intensities of the two cleavage products and the relative amount of methylated template DNA. As depicted in Fig. 3b, the linear relationship between the actual and estimated relative methylation spans the entire range in these mixtures. The average standard deviation of relative methylation was ±5% for mixtures between 10% and 90%. Outside of this range, quantitative data can be analyzed, but limitations in the dynamic range of current axial MALDI-TOF mass spectrometers may decrease the precision of relative methylation estimates.

**Complete Methylation Pattern Analysis.** We tested the capability of base-specific cleavage to determine the methylation status of every single CpG in a target region. The C-specific forward reaction contains a cleavage site for each methylated CpG in the amplicon. The occurrence of a cleavage product in a mass spectrum implies the existence of two cleavage sites and, hence, the presence of methylated DNA in a sample, can be detected in mixtures containing as little as 5% methylated DNA without methylation-specific PCR.

The occurrence of a cleavage product in a mass spectrum implies the existence of two cleavage sites and, hence, two methylated cytosines. The practical mass window for current instrumentation ranges from 1,000 to 10,000 Da. In this mass window, cleavage products with lengths between 4 and 30 nucleotides can be detected. We simulated methylation detection rates for 27,757 CpG islands annotated by the UCSC genome browser (status October 2004). In this simulation, we included all CpG islands longer than 100 bp. For those that exceeded 500 bp in length (15,365 of the 27,757 CpG islands) we randomly assigned a 500-bp stretch within the annotated CpG islands. Our simulation covered a total nucleotide count of 11,638,876 bp and 1,145,069 individual CpGs. A single cleavage reaction reveals the methylation status of ~82% of all CpG sites, either as individual CpGs or as CpG units. The use of a second base-specific cleavage reaction increases the detection efficiency to...
89%, and a combination of four reactions assesses the methylation status of 98% of the CpG sites. For every target region analyzed, these results may vary with the distribution of the CpG sites and their sequence context. For example, the IGF2/H19 region described here would have required only two reactions to obtain the methylation status of every CpG site. Fig. 6, which is published as supporting information on the PNAS web site, shows the reconstruction of the methylation pattern in the IGF2/H19 region by using all four base-specific cleavage spectra. Our methylation results for the IGF2/H19-imprinted region in adult blood samples confirmed the methylation patterns reported by Vu et al. (19).

Our results demonstrate the capability of the method to discriminate methylated and unmethylated DNA target regions simultaneously and to reconstruct the exact methylation pattern. For a more realistic test, we analyzed bisulfite-treated genomic DNA directly. The mass spectra showed signal patterns that are representative of the methylated template and those that are characteristic of the unmethylated template. The signal intensities for methylation-specific signals and nonmethylation-specific signals were compared, and the 50/50 ratio expected for hemimethylated DNA, in control blood samples, was confirmed. The results of this experiment are shown in Fig. 7, which is published as supporting information on the PNAS web site.

**Analysis of Cytosine Methylation in Lung Cancer.** To evaluate the performance of this approach in a high-throughput setting with clinical samples, we characterized changes in the cytosine methylation pattern that occur in non-small cell lung cancer. We analyzed the degree of methylation of 47 candidate gene promoter regions in lung tumors and adjacent normal specimens from 48 cases with a history of smoking. The patient collection consists of 22 females and 26 males, ages 45 to 83 years (median = 68 years). DNA was extracted from frozen tissue specimens by using a standard phenol/chloroform protocol. The candidate genes, provided in Data Set 1, were selected from public databases, because they have been shown to change expression levels during cancer development and progression, or because their biological function is related to cell adhesion and cell interaction. We selected a single CpG island within each of the genes and usually focused on those residing in the promoter and 5′ untranslated region. The selected gene regions contain a total of 1,426 CpGs. A quantitative assessment of the relative methylation was carried out for each CpG unit according to the protocol described above. We used only the U-specific cleavage of the reverse transcript, because it provides sufficient semiquantitative information for identification of differentially methylated CpG sites. In this cleavage reaction, the 1,426 CpGs were distributed among 613 unique CpG units.

Of the 47 promoter regions analyzed, three resulted in poor quality spectra that could not be analyzed. The remaining 44 regions provided at least 10 valid methylation measurements in 599 CpG units (range = 3–25, median = 13). The majority of CpG units resulted in quantitative measurements of methylation for >88% of the samples analyzed in a high-throughput, single-pass process. The average degree of methylation in normal compared with tumor tissues for each CpG unit is shown in Fig. 8, which is published as supporting information on the PNAS web site. The majority of CpG units had an average relative methylation <0.1 in both normal and tumor tissues. One hundred thirteen units were completely nonmethylated in every sample (degree of methylation = 0) and 2 units were completely methylated (degree of methylation = 1). Among the units with a relative methylation <0.5 in normal tissue (n = 513), CpG units of tumor tissues were significantly more likely to have higher levels of methylation (hypermethylation) than normal tissues (P = 1.5 × 10^{-10}). Conversely, those units with a relative methylation >0.5 in normal tissue (n = 86) were much more likely to be hypomethylated in tumor tissue (P = 1.3 × 10^{-10}).

![Color Key](image.png)

**Fig. 4.** A two-way hierarchical cluster analysis of the relative methylation of 68 CpG units (columns) measured on 96 tissue samples from 48 lung cancer cases (rows). Tissue samples are identified on the left vertical axis as white boxes (normal tissue) and black boxes (tumor tissue). CpG units are identified at the bottom horizontal axis as the gene and the fragment number within the gene. The relative methylation of each fragment within each sample is presented in the central image plot with values ranging from zero (red) to one (yellow; see color key). Missing values are represented as gray.
A comparison of relative methylation between normal and tumor tissues (paired by patient) within each CpG unit identified 41 highly statistically significant differences \( P < 0.0001 \), yielding a conservative \( 1 \% \) experimentwide significance level spread among 12 of the 47 targeted promoter regions. The direction of change in relative methylation was consistent among all significant CpG units within every gene. In three significant promoter regions, the average relative methylation was elevated in tumor compared with normal tissue. The remaining nine regions were hypomethylated in tumor tissue.

To test the ability of CpG methylation measurements to discriminate between normal and tumor samples, we carried out a two-dimensional hierarchical cluster analysis by using the methylation results on the 41 significant units and 27 other CpG units from the same genes with an intersample standard deviation >0.2 (see Fig. 4). The tissue samples formed two distinct clusters separating the normal from the tumor tissues. Seven tumor tissues clustered among the normal tissues, whereas only one normal tissue demonstrated a methylation pattern similar to the tumor tissues. Multiple major CpG subclusters are evident, some made up of units from single genes and others comprised of multiple promoter regions with very similar methylation patterns. This result supports the hypothesis that the relative amount of methylated DNA can be used to differentiate between normal and neoplastic tissue.

### Discussion

There are several prior methods for assessment of cytosine methylation. Many of these techniques can only analyze a restricted set of CpG sites in their target regions, and one must extrapolate to estimate the degree of methylation in the whole region. Misinterpretation of the methylation status has been reported, caused, for example, by incomplete bisulfite treatment, improper primer design in methods relying on selective amplification, or partial digestion when restriction-based methods are applied. A further complication arises for those methods restricted to selected CpGs when their methylation within the examined genomic region is not representative.

In this study, we present a previously uncharacterized approach for DNA methylation analysis that employs base-specific cleavage and MALDI-TOF MS analysis to overcome those limitations. Using a combination of four base-specific cleavage reactions, each CpG of a target region can be assessed individually and is represented by multiple indicative mass signals. The information acquired about the methylation status of the examined region is based on numerous independent observations. The redundancy of this information can be leveraged to achieve higher confidence in qualitative analysis and to obtain highly precise quantitative averages. The approach can be adapted to various needs in DNA methylation analysis, including discovery of methylation in large stretches of genomic DNA with a single cleavage reaction, where fractions with at least 5% methylated DNA can be detected in mixtures, semiquantitative analysis of the relative methylation, and methylation pattern analysis, where the methylation status of each CpG site within a target region can be determined.

The cost for a set of four base-specific cleavage reactions is comparable to a single fluorescent deoxy sequencing reaction. However, because a single reaction covers ~82% of the targeted CpG sites, most applications could be carried out at a quarter of the cost. A conventional high-throughput MALDI-TOF mass spectrometer is capable of processing 6,000 samples per day. This number of samples can comprise any combination of DNA samples, amplicons, and cleavage reactions. For example, the study described here encompassed 47 promoter regions measured in 96 individuals (1,426 CpG sites × 96 individuals = 136,896 CpGs total) and can be completed in a single day (assay design and ordering of oligonucleotides excluded).

The large-scale analysis of methylation in non-small cell lung cancer is the first high-throughput implementation of this method for quantitative assessment of the degree of methylation. Initial analysis of methylation ratios in a subset of the 12 most significant genes resulted in the clustering of samples according to their histological status. Several CpG units within those 12 significant genes showed small but characteristic changes in DNA methylation. Their identification relies on the precision of the quantitative measurement of our method. Less precise methods, such as fluorescent sequencing or hybridization-based methods, will likely fail to identify them. Although the robustness of the methylation data as a predictor for tissue histology must be verified through independent replication, these results support the suitability and robustness of base-specific cleavage and MALDI-TOF MS for semi-quantitative methylation analysis.

There are further technological features that could be highly valuable in the exploration of the role of methylation. Large-scale studies, for example, can use multiplexing if they focus on discovering methylation in a region and do not need exact reconstruction of methylation sites. Because of the universal nature of the RNA polymerase recognition sequence, it is possible to multiplex several target regions in one reaction from PCR onwards. Such multiplexing of base-specific cleavage reactions has been performed successfully in our laboratory for SNP discovery applications.

We have shown that methylation analysis by base-specific cleavage in combination with MALDI-TOF MS detection is a sensitive, accurate, and reliable technique. The implementation of this technique will provide an excellent tool for cost-effective high-throughput methylation analysis and, therefore, will help enhance our ability to detect disease genes and to increase our understanding of epigenetic modifications.

We thank Christiane Schmidt for the implementation of suitable bisulfite treatment protocols and Johnathan Miller at the Liverpool University for his assistance in tissue sectioning and DNA extraction. T.L. and G.X. are supported by the Roy Castle Lung Foundation, United Kingdom.

### References