Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions

Zachary A. Kaminsky[1,2], Abbas Assadzadeh[1], James Flanagan[1] and Arturas Petronis[1,2,*]

[1]The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, ON, Canada M5T 1R8 and [2]University of Toronto, Toronto, ON, Canada

ABSTRACT

The development and use of high throughput technologies for detailed mapping of methylated cytosines (metC) is becoming of increasing importance for the expanding field of epigenetics. The single nucleotide primer extension reaction used for genotyping of single nucleotide polymorphisms has been recently adapted to interrogate the bisulfite modification induced 'quantitative' C/T polymorphism that corresponds to metC/C in the native DNA. In this study, we explored the opportunity to investigate C/T (and G/A) ratios using the Applied Biosystems (ABI) SNaPshot technology. The main effort of this study was dedicated to addressing the complexities in the analysis of DNA methylation in GC-rich regions where interrogation of the target cytosine can be confounded by variable degrees of methylation in other cytosines (resulting in variable C/T or G/A ratios after treatment with bisulfite) in the annealing site of the interrogating primer. In our studies, the mismatches of the SNaPshot primer with the target DNA sequence resulted in a biasing effect of up to 70% while these effects decreased as the location of the polymorphic site moved upstream of the target cytosine. We demonstrated that the biasing effect can be corrected with the SNaPshot primers containing degenerative C/T and G/A nucleotides. A series of experiments using various permutations of quantitative C/T and G/A polymorphisms at various positions of the target DNA sequence demonstrated that SNaPshot is able to accurately report cytosine methylation levels with <5% average SD from the true values. Given the relative simplicity of the method and the possibility to
multiplex C/T and G/A interrogations, the SNaPshot approach may become a useful tool for large-scale mapping of metC.

INTRODUCTION

Technological advancement in DNA methylation analysis is an important and ongoing endeavor of epigenetic research. The gold standard technique for the fine mapping of methylated cytosines (metC) utilizes bisulfite modification, which converts unmethylated cytosines into uracils, while the metC remain unchanged. Following PCR, uracils are replaced by thymines (T), allowing differentiation between what was metC and C by examining the proportions of C to T present in the PCR product at positions of interest. However, because DNA from different cells exhibit various degrees of epigenetic heterogeneity, it is necessary to clone the target sequences of the bisulfite-treated DNA and sequence numerous clones [in the literature, the number of clones varies from 3 to >65 (1,2)]. Therefore, although precise, the method is very labor intensive and as a rule is limited to the analysis of relatively short (<1 kb) DNA fragments. In mammals, methylation targets are cytosines in CpG dinucleotides (3) and in some cases CpNpG (4), therefore sequencing the entirety of the clone is not necessarily efficient as only a small fraction of the nucleotides are of interest. Because of this, a method that could efficiently analyze methylatable cytosines only is highly desirable. Since bisulfite modification induces DNA sequence changes ('induced DNA polymorphism'), the objective to identify metC and C, which become C and T (or G and A in the complementary strand) after bisulfite modification, is quite similar to that of genotyping single nucleotide polymorphisms (SNPs). The difference, however, is that SNP genotyping generates three discrete outcomes (a heterozygote and two alternative homozygotes), while C/T (G/A in the complementary strand)

ratios may exhibit a high degree of variation from 0%/100% to 100%/0% and anywhere in between. There also have been successful attempts to develop a protocol for SNP allele frequency estimation in DNA pools where allele frequency may vary between 100 and 0% (5). Among the myriad of techniques for the detection and measurement of SNPs, a frequently used approach is the single nucleotide primer extension reaction (6). Several studies have provided evidence that methylation-sensitive single nucleotide primer extension-based SNP genotyping (Ms-SNuPe) can be performed in a quantitative manner and thus could be useful for evaluation of ^{met}C/C ratios (5,7–14). However, several important issues remained to be addressed. First, can the mapping of ^{met}C/C be performed in an automated high throughput fashion? Second, what are the effects of neighboring cytosines when their methylation status is unknown *a priori*? The latter is a typical situation for CpG islands, regions of primary interest in DNA methylation studies. Until recently, many such regions have been investigated using Ms-SNuPe by avoiding CpG dinucleotides in the annealing site of the interrogating primer (1), which is a marked limitation of the technology. Recent attempts have been made to overcome the problem of Ms-SNuPe primer mismatch effects in order to interrogate CpG sites independent of sequence context, including GC-rich regions, using MALDI mass spectrometry (15). While it is intuitive that primer binding mismatches may affect results, the effects of such mismatches have not been analyzed in previous studies.

As the need for high throughput experimental designs to investigate methylation profiles increases, it seems that the techniques employed to accurately investigate them require increasingly specific equipment platforms that may not be available at all facilities. At present, the platforms that are capable of investigating complex methylation patterns unhindered by CpG-rich regions are mass spectrometers, pyrosequencing machines and a software analysis program of direct sequencing called ESME (15–20), all of which are not very common in standard molecular biology laboratories. It was therefore one of our goals to successfully employ a SNaPshot technique that is both cost-effective and easy to perform on an alternate equipment platform and therefore widen the availability of high throughput Ms-SNuPe reactions to those facilities with access to electrophoresis machines, such as the Applied Biosystems (ABI) Avant 3100 genetic analyzer. The main effort was dedicated to interrogation of methylated status in the GC-rich regions where the density of methylatable cytosines is high. After bisulfite modification such methylated and unmethylated cytosines will result in numerous 'quantitative' C/T and G/A polymorphisms, which may compromise binding of the interrogating primer and distort SNaPshot results. We have examined the effect of mismatches in the primer binding region and found a position-dependent biasing effect. More importantly, however, we have demonstrated that SNaPshot primers containing degenerative C/T or G/A bases at potential mismatches can accurately quantify methylation at a given target and may produce multiple peak patterns that are indicative of methylation differences upstream of that target.

The analyses performed in this study can be stratified into three categories. The first category consisted of experiments to operationalize and test the ability of multiplexed SNaPshot primers to quantitatively measure C/T (G/A) ratios in the target sequences where there were no methylatable cytosines and therefore no possible primer mismatches with the bisulfite-induced variable C/T (G/A) in the SNaPshot primer binding region. In the second category, the more complex scenarios were explored, such as when a SNaPshot primer binding region contains such quantitative C/T (G/A) polymorphisms, and strategies to reliably measure target C/T (G/A) ratios were developed. Finally, the third category experiments were dedicated to verification of the adapted SNaPshot approach on various oligonucleotides and bisulfite-modified DNA targets.

MATERIALS AND METHODS

DNA sequence targets for SNaPshot interrogation

Three types of DNA targets were generated to address various issues surrounding the adaptation of the SNaPshot approach for the quantitative evaluation of C/T (G/A) ratios:

(1) A fragment of the promoter region of the gene encoding human catecholamine *O*-methyltransferase (*COMT*) was amplified as follows. 10× PCR Buffer, 2 mM MgCl$_2$, 2.5 mM dNTP, 1 M Betaine, 0.4 mM primers and 1 U of Taq polymerase (New England Biolabs), primers: comtF1 5′-agaccacaggtgcagtcagcacag-3′ and comtR1 5′-cacc-tatcccagtgttccacccta-3′ at 95°C for 5 min, 30 cycles (94°C for 1 min, 61°C for 1.5 min and 72°C for 1 min), and 72°C for 5 min. CCGG and GCGC sites of the amplicon were subsequently methylated using M-HpaII and M-HhaI, respectively, in two separate fractions. The third fraction of the amplicon was left unmethylated. Both methylated and unmethylated DNA samples were then subjected to bisulfite modification (21). Briefly, DNA was boiled for 5 min, cooled on ice and denatured for 15 min at 50°C after adding 4 μl of fresh 2 M NaOH in a total reaction volume of 25 μl. Two volumes of 2% LMP agarose in distilled water was added and 10 μl aliquots of this solution was pipetted into cold mineral oil and placed immediately back into dry ice to create beads. The mineral oil was removed and a solution of 1.9 g sodium metabisulfite in 2.5 ml H$_2$O, 720 μl of 2 M NaOH and 500 μl of 1 mM hydroquinone was added. Samples were incubated on ice for 30 min followed by incubation at 50°C for 3.5 h. The agarose beads were washed four times for 15 min with 1 ml TE, two times for 15 min with 0.2 M NaOH, three times for 10 min with 1 ml TE and two times for 15 min with H$_2$O. This was followed by semi-nested PCR using identical reaction and cycling conditions as above with semi-nested primers: BisF1, 5′-gaagggg-gttatttgtggttagaa-3′, BisF2 5′-gatttttgagtaagattagattaag-3′ and BisR1 5′-aacaaccctaactaccccaa-3′. C (^{met}C in the amplicon) containing templates were mixed with the T (unmethylated C in the amplicon) containing fraction to create a standard curve from 100 to 0% of C signal in increments of 5%: (100% C: 0% T), (95% C: 5% T), (90% C: 10% T), . . . , (0% C: 100% T). This was done for those templates containing C at M-HpaII sites and M-HhaI sites separately. The M-HpaII sites were interrogated with three forward primers, while three primers for the M-HhaI sites were added as negative controls. In a similar way, M-HhaI

sites were interrogated with all six forward primers (three of which were for the M-HpaII sites as negative controls) in one run and the three reverse primers in a second run (Figure 1). The interrogating primers were designed to have a T_m close to 50°C to allow for similar annealing dynamics in the multiplexed reaction. To vary the length of the primers, non-complementary tails were designed on the 5′ end of each primer by repeating the sequence GACT (shown in brackets): at least two sets of the GACT for oligos with the total length <40 nt and by one set of GACT for those >40 nt. The interrogating SNaPshot primers were

(a) 5′-agtaagattagattaagaggt-3′, 5′-[gact]$_1$gatatttttatgagga- tattt-3′ and 5′-[gact]$_6$ttatggtttgtgtttgttat-3′ for the HpaII sites;

(b) 5′-[gact]$_4$ggatattttggttattgttg-3′, 5′-[gact]$_6$ttttgattttatt- ttatttgttg-3′ and 5′-[gact]$_7$agtgtttttttaattttttgtag-3′ for the HhaI sites (direct primers);

(c) 5′-ccacaataaatatccac-3′, 5′-[gact]$_2$tataacaaacaaaataca- aaac-3′ and 5′-[gact]$_3$acactacaaaaattaaaaaaac-3′ for the remaining three HhaI sites (reverse primers).

(2) In order to investigate the effects of quantitative G/A and C/T polymorphisms in the SNaPshot primer binding region, sets of oligonucleotides containing variable C/T and G/A were synthesized. Five polymorphic positions were investigated: −2, −5, −10, −15 and −18 (A_{-2}/G_{-2}, A_{-5}/G_{-5}, T_{-10}/C_{-10}, A_{-15}/G_{-15} and T_{-18}/C_{-18}) upstream of the nucleotide that was interrogated ('target' nucleotide, N_{target}) (Figure 2). SNaPshot primers in the experiment were named according to the polymorphic site, while the DNA template itself was named by the nucleotide in the polymorphic position and also the target nucleotide. Therefore, the primer T_{-2} is fully complementary to the templates $A_{-2}A_{target}$ and $A_{-2}G_{target}$, but not comple- mentary at the −2 position to the templates $G_{-2}A_{target}$ and $G_{-2}G_{target}$. It is evident that the T_{-2} primer will preferentially bind to (and interrogate) the DNA sequence that contains A_{-2}, in comparison to G_{-2} at the upstream position. The degree of such bias, however, is unknown as is the impact of the location of the mismatch proximal to the target nucleotide. To elucidate the degree of bias, DNA templates containing an upstream polymorphism, e.g. $G_{-2}G_{target}$ and $A_{-2}A_{target}$, were added in equal amounts which resulted in a polymorphic G/A site at the −2 position. This template mix was tested in two different primer sce- narios: first with primer T_{-2}, then with primer C_{-2}. All other polymorphic sites at positions −5, −10, −15 and −18 were analyzed in the same way. For the mismatch bias correction, numerous DNA template combinations with varying percentages of polymorphic nucleotides in the primer binding site were tested using different primer combinations (see Results).

(3) To verify that degenerative SNaPshot primers are able to interrogate numerous polymorphic C/T and A/G containing DNA sequence targets, two types of DNA templates were used:

(a) Six oligonucleotide templates were synthesized with quantitative G/A polymorphisms in different positions of the SNaPshot primer annealing region (Figure 3). The target nucleotide A/G proportions were synthes- ized to be 50%:50% in each template while the upstream A/G ratios were synthesized according to Figure 3. SNaPshot primers contained a 50%:50% proportion of C/T at degenerative positions corresponding to poly- morphic positions in the templates.

(b) Two human genomic DNA samples from brain and placenta were bisulfite modified and subjected to both SNaPshot interrogation and to cloning plus sequencing-based measurement of ^{met}C density. The two selected CpG island regions were identified as exhibiting DNA methylation differences according

Figure 1. A continuous sequence of DNA representing the bisulfite-modified *COMT* promoter region amplified as a template for the SNaPshot multiplexing experiments. Only the top strand of DNA is depicted along with the positions of all SNaPshot primers used. Polymorphic target sites created by bisulfite modifications of M-HpaII and M-HhaI methylated amplicons are marked in bold capital letters. Primers overlapping M-HhaI sites (bottom right) were designed with T at the overlapping CpG positions. Forward SNaPshot primers are boxed above the sequence while reverse SNaPshot primers are boxed in red below the sequence. The non-binding GACT repeat tails (placed on the 3′ end of some primers) are denoted by a number, the purpose of which is to vary the primer length in order to distinguish them in the ABI 3100 Genetic Analyzer.
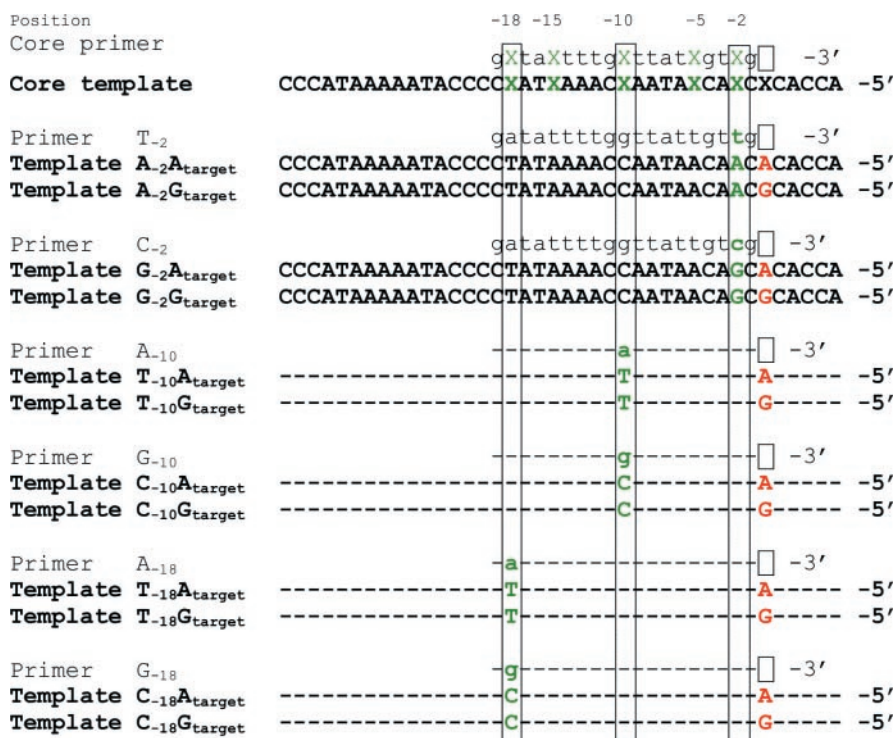
Figure 2. Oligonucleotide templates and complementary primers were synthesized to test the effects of C/T and G/A polymorphisms upstream of the target nucleotide at positions −2, −5, −10, −15 and −18 bp. Sequences are depicted for C/T and G/A polymorphisms at positions −2, −10 and −18; however, primers and templates were also tested for positions −5 and −10. Templates are named for the complementary strand to which the SNaPshot primer binds, the two nucleotides representing the polymorphic position (signified by the number) and the target nucleotide, respectively. SNaPshot primers are named according to the nucleotide complementary to the upstream polymorphism.

Figure 3. A list of oligonucleotide templates synthesized to contain between 1 and 4 polymorphic positions in the SNaPshot primer binding region and respective primers containing degenerative bases at positions corresponding to those polymorphisms. Percentages of nucleotides synthesized into the templates are depicted while SNaPshot primers were designed with a 50%:50% proportion of C/T at all polymorphic positions.

to our microarray-based DNA methylation profiling (A. Schumacher, A. Petronis, *et al*., unpublished data). These regions are located between 28 and 276 bp upstream of known genes coding for *LGALS1* (lectin, galactoside-binding, soluble, 1), otherwise referred to as galectin 1 and humanin, respectively. Three CpG positions were selected for each CpG island and will be referred to as gal1, gal2 and gal3 for galectin 1 and hum1, hum2 and hum3 for humanin (Figure 4).

Bisulfite modification reactions were performed as described above. Target sequences were amplified using fully nested PCR. PCR conditions were as follows: 10× PCR Buffer, 2 mM MgCl$_2$, 2.5 mM dNTP, 0.4 mM primers and 1 U of Taq polymerase. The first PCR was performed using primers, for galectin 1: 22_f1 5′-gtagaatgttaattttgggtagaaataat-3′ plus 22_r1 5′-ctcaaccatcttctctaaacacc-3′; and for humanin: 52_f1 5′-agtttgtattaaggagatttataaggatag-3′ plus 52_r1 5′-aaccaa-caaaacacacaaacc-3′. The second (nested) PCR used primers, for galectin 1: 22_f2: 5′-gttattgaggtttagaaaagagaaggtat-3′ plus

Figure 4. SNaPshot primers for the galectin1 (**A**) and humanin (**B**) genes that were identified as being differentially methylated between placenta and brain tissue.

22_r2 5′-acttataaacctaactcatcatcaaactat-3′ and for humanin: 52_f2: 5′-aatttagattttgagtttttgaaag-3′ plus 52_r2 5′-aacacaaca-taacaacaaacaaaac-3′ site. Two successive rounds of touch down PCR were used with the following cycling conditions: 95°C for 3 min, 10 cycles [94°C for 1 min, 60°C for 30 s (minus 1°C/cycle), 72°C for 40 s], 30 cycles of (94°C for 1 min, 50°C for 30 s, 72°C for 40 s) and 72°C for 5 min. The sequences of the interrogating SNaPshot primers were gal1 5′-gttattgggggyggagtt-3′, gal2 5′-[gact]$_2$gaggatgttttygggtagg-3′ and gal3 5′-[gact]$_4$gatyggatygggtgagttt-3′. Primers for humanin were hum1 5′-acagttyggattttttygaaaggggg-3′, hum2 5′-aactcccaatatcrtacratac-3′ and hum3 5′-ygagggtgatagg-gaag-3′. Amplicons were cloned into the pDrive plasmid (Qiagen) that were used for transformation of DH5-α competent cells. Individual colonies were grown at 37°C for 15 h followed by plasmid purification using the Qiagen Spin Miniprep kit. Sequencing of 12–15 plasmid inserts per template was carried out with M13 reverse primer using ABI Big Dye Terminator kit 3.1. Six CpG positions were investigated using SNaPshot primers individually and five were multiplexed in two groups (gal1, gal2 and gal3, and hum1 and hum3). The differences in length of degenerative primers to be multiplexed were in accordance with the specifications recommended by the manufacturer (ABI). All SNaPshot experiments on bisulfite-modified DNA were repeated in quadruplicate.

SNaPshot. SNaPshot reactions were carried out using 0.01–0.4 pmol of DNA template, 2 μM primers in the final reaction, 5 μl of SNaPshot master mix (ABI), and water to a final reaction volume of 10 μl. Reactions were carried out in the ABI 9700 Thermocycler at 96°C for 10 s, 50°C for 5 s and 60°C for 30 s, and repeated for a total of 25 cycles. Following cycling, samples were treated with 3 U of calf intestine

phosphatase at 37°C for 1 h and heat inactivated at 72°C for 15 min. In an ABI optical plate, 9 μl HI-DI formamide, 0.5 μl of Genescan 120 LIZ size standard (ABI) and 0.5 μl of the reaction were combined, denatured at 95°C for 5 min, and immediately placed on ice for 2 min. Samples were then loaded on the ABI Avante 3100 Genetic Analyzer for analysis. To determine which target produced which signal in a multiplexed reaction, peak position was correlated with the length of the primer designed for a given target. The *x*-axis reports the relative length of the primers compared to the loaded size marker. Therefore, shorter SNaPshot primers will produce peaks closer to the origin on the *x*-axis while longer ones will produce peaks farther from it. It should be noted that a C and T peak incorporated by the same primer should be offset by at least 0.5 nt spacing because primers with different ddNTP compositions travel at different rates through the polymer matrix in the ABI machine. The precise peak height was determined by the Genescan 6.0 analysis software, which provides a specific data output below the graph of peaks that includes the relative size (in bp), peak height and peak area for each primer. The data output corresponding to any given peak can be determined by clicking on a peak, which will in turn highlight the corresponding data cell below. The percentage of C at a given position was determined by the formula: $C\% = [(C_i)/(C_i + T_i)] \times 100$, where C_i and T_i stand for the peak height of C and T signal, respectively. When interrogating the reverse strand, G and A peaks will be produced in place of C and T peaks. The same formula can be used in this case by substituting G_i and A_i for C_i and T_i, respectively. Peak area should not be used as diffusion over large distances in the polymer matrix may affect the consistency of results. In cases where multiple peaks for a single target were observed, cumulative intensities for all C peaks and all T peaks

were substituted into C_i and T_i, respectively, in the above formula.

RESULTS

Simple DNA templates: multiplexed SNaPshot reactions

The maximum number of SNapShot primers that would fit on the *COMT* amplicon in the forward direction without competing for annealing sites was six (Figure 1). The remaining three primers were designed for the reverse strand and run in a separate reaction (Figure 1). To test the multiplexing capability, we created a standard curve of decreasing C signal—increasing T signal from 100 to 0% in increments of 5% (100% C: 0% T), (95% C: 5% T), (90% C: 10% T), ..., (0% C: 100% T) in each of the six sites by mixing methylated (C) and unmethylated (T) templates (Figure 5). In the SNaPshot reaction, the primers bind to their complementary strand and incorporate a single fluorescent ddNTP. Primers are then separated and measured proportionally in the ABI 3100 Genetic Analyzer. Consequently, those primers designed in the forward and reverse directions produce C/T and G/A peaks, respectively. Methylated HpaII or methylated HhaI site-derived templates were separately tested with all six forward SNaPshot primers. Three sets of experiments were performed in triplicate for templates comprised of (i) three HpaII sites and (ii) three HhaI sites interrogated by forward primers; and (iii) three HhaI sites interrogated by reverse primers. The results showed that all primers produced mean C/T and G/A (for the reverse primers) ratios reflecting the expected values (the range of SD: 2.7–4.6%) (Figure 6).

### Complex DNA templates: SNaPshot primer mismatch bias and bias correction

Mismatch bias. One of the key objectives of the study was to investigate the suitability of the primer extension technique for the investigation of DNA methylation patterns when dealing with CpG dinucleotides (that can be methylated to different proportions) in close proximity to the target cytosine. Depending on their *a priori* unknown methylation status, such CpGs may generate a wide variety of C/T or G/A upstream polymorphisms that may lead to one or more mismatches between the DNA template and the primer, and thus result in incorrect measurement of C/T or G/A ratios at the target site. Templates listed in Figure 2 were mixed in equal amounts to produce a 50%:50% ratio of the target and upstream position. For example, the mix of $G_N G_{target}$ and $A_N A_{target}$ templates was tested with each complementary primer, C_N and T_N, individually. The degree to which a base mismatch could bias the reflected proportion of target G/A decreased when it moved farther upstream in the 5′ direction (Figure 7). The G_{-2}/A_{-2} mismatch biased the results between 40 and 65% while the G_{-5}/A_{-5} mismatch resulted in 25–70% bias. The average for the T_{-10}/C_{-10}, G_{-15}/A_{-15} and T_{-18}/C_{-18} mismatches was 35, 23 and 9%, respectively (Figure 7).

Mismatch bias correction. Equal template mixes used to determine the degree of mismatch bias in the above experiment, e.g. $G_N G_{target}$ and $A_N A_{target}$, were now interrogated with a 50% mix of SNaPshot primers complementary to each template. In all cases, the expected 50% proportion of represented targets was observed. An additional set of experiments was performed to test if such an equal representation of primers could
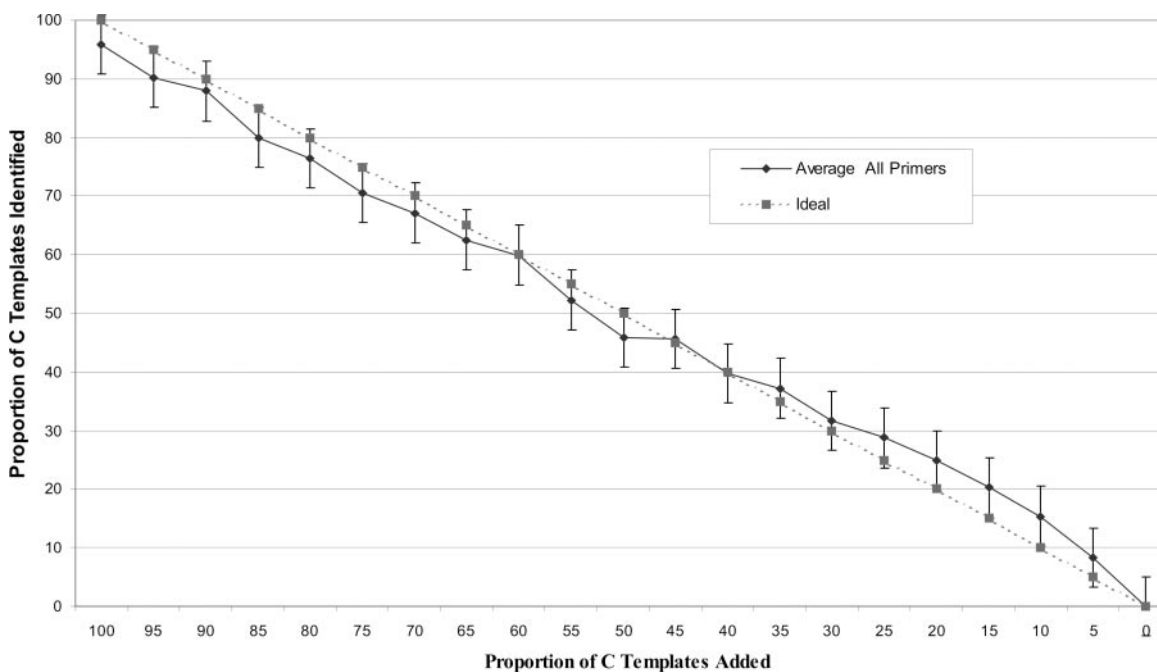
Figure 5. A graph of the average methylation values reported by nine primers interrogating control templates created to contain only C or T at the CpG islands of interest. Each dilution series was tested in triplicate for each primer so that each data point is an average of 27 experiments. Templates with all CpG islands of interest containing C were diluted and mixed in increments of 5% with templates containing T at all CpG being interrogated to test the ability of the primers to accurately measure the amount of methylation.
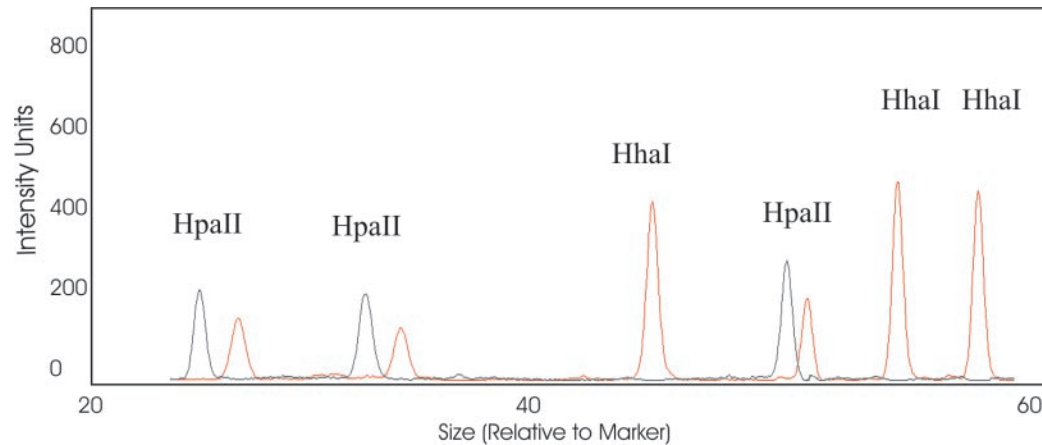
Figure 6. Data output from the ABI Avante 3100 combining 60% bisulfite-treated M-HpaII methylated template and 40% unmethylated template. The peak heights show 60% C to 40% T signal for those peaks methylated with M-HpaII (peak pairs 1, 2 and 4). Peaks 3, 5 and 6, representing M-HhaI sites, show no C signal and hence no methylation. Peak order is determined by primer size.
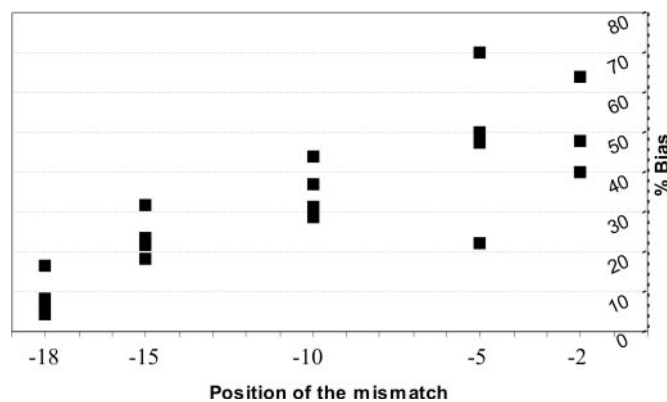
Figure 7. A graphical representation of the percentage that a mismatch in the primer binding region at various positions upstream from the target CG can affect the reading. There is a correlation between the proximity of a primer mismatch to the target and the degree to which the resulting ^{met}C/C reading will be inaccurate. Zero baseline represents 0% biasing effect.

accurately identify known amounts of target A/G nucleotides independent of the upstream polymorphisms that cause bias. A series of dilutions of oligonucleotide templates were prepared to interrogate G_{target}/A_{target} in decrements of 25% from 100% G to 0% G (Figure 8). For each dilution series, the −2 and −5 positions (G_{-2}/A_{-2} and G_{-5}/A_{-5}) polymorphisms also varied. Each dilution series was tested in duplicate with each of the following five scenarios of upstream polymorphism concentration: (i) 100% G_N versus 0% A_N; (ii) 75% G_N versus 25% A_N; (iii) 50% G_N versus 50% A_N; (iv) 25% G_N versus 25% A_N; and (v) 0% G_N versus 100% A_N. A 50%:50% primer mix, primers C_{-2}/T_{-2} (50%:50%) and primers C_{-5}/T_{-5} (50%:50%), was able to accurately quantify the amounts of the target present in each of the above scenarios (SD 2.7–5.6%) (Figure 8A and B). To test the effects of multiple upstream mismatches, two sets of experiments were performed in triplicate using randomly selected 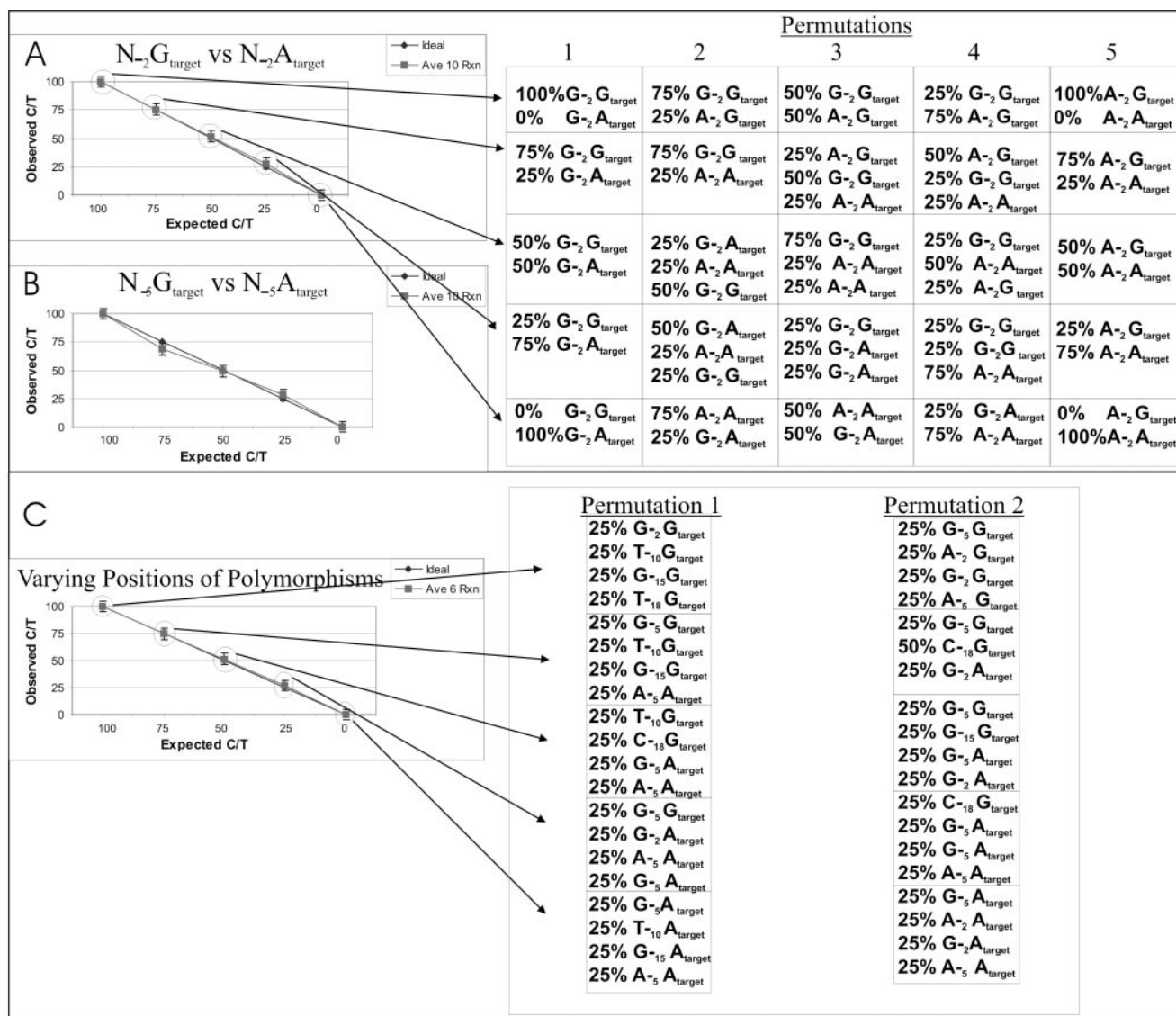combinations of templates with all five upstream polymorphisms (Figure 8C). These dilution series were tested with all 10 primers represented in equal amounts, accounting for each template polymorphism (primers C_{-2}, T_{-2}, C_{-5}, T_{-5}, G_{-10}, A_{-10}, C_{-15}, T_{-15}, A_{-18} and G_{-18} (each at 10%). Data output reflected the

known amounts of target (SD 0.79–5.0%). The presence of primers in the reaction mixture for which there was no complementary template did not bias results. For all experiments run without an equal representation of all primers, curves were drastically biased and results did not reflect the known amounts of target nucleotides (Supplementary Figure 1).

C/T (G/A) interrogations in complex oligonucleotide templates and bisulfite modified DNA

Degenerative primer experiments on oligonucleotide templates. To test the ability of primers designed with C/T and G/A degenerative bases at polymorphic positions to interrogate a given target, six oligonucleotide templates and corresponding primers were designed as shown in Figure 3. All templates were 50%:50% of A and G ($A_{0.5}$ and $G_{0.5}$) at the target position and variable upstream proportions (0.1–0.9; 0.2–0.8; 0.3–0.7; 0.4–0.6; etc.) of C/T and G/A ranging from one to four positions. Primers had a 50%:50% mix of C/T ($C_{0.5}$ and $T_{0.5}$) at positions corresponding to any polymorphic positions in the template. All corresponding SNaPshot primers with degenerate bases at polymorphic positions accurately reported equal intensities of fluorescent C and fluorescent T that correspond the 50%:50% of G and A in templates (SD = 0.22–3.5%) (data not shown).

Degenerative primer experiments on bisulfite-modified DNA. The SNaPshot approach was used to identify DNA methylation differences in several genes that exhibit tissue (brain and placenta) specific differential epigenetic modification (A. Schumacher, A. Petronis, *et al.*, unpublished data). Such target sequences have also been subjected to cloning of bisulfite-modified DNA and direct sequencing of individual clones was performed on at least 12 clones. Five CpG positions were investigated using SNaPshot primers in multiplexed primer groups (gal2, gal3 and gal1, and hum1 and hum3). Primer hum2 was not multiplexed because it was designed as a reverse primer and exhibited a high degree of sequence homology with primers hum1 and hum3. All SNaPshot experiments on bisulfite-modified DNA were repeated in quadruplicate. In all experiments, an identical peak pattern was observed and the proportions of all peaks remained identical

Permutations

	1	2	3	4	5
A $N_{-2}G_{target}$ vs $N_{-2}A_{target}$	100% $G_{-2}G_{target}$ / 0% $G_{-2}A_{target}$	75% $G_{-2}G_{target}$ / 25% $A_{-2}G_{target}$	50% $G_{-2}G_{target}$ / 50% $A_{-2}G_{target}$	25% $G_{-2}G_{target}$ / 75% $A_{-2}G_{target}$	100% $A_{-2}G_{target}$ / 0% $A_{-2}A_{target}$
	75% $G_{-2}G_{target}$ / 25% $G_{-2}A_{target}$	75% $G_{-2}G_{target}$ / 25% $A_{-2}A_{target}$	25% $A_{-2}G_{target}$ / 50% $G_{-2}G_{target}$ / 25% $A_{-2}A_{target}$	50% $A_{-2}G_{target}$ / 25% $G_{-2}G_{target}$ / 25% $A_{-2}A_{target}$	75% $A_{-2}G_{target}$ / 25% $A_{-2}A_{target}$
B $N_{-5}G_{target}$ vs $N_{-5}A_{target}$	50% $G_{-2}G_{target}$ / 50% $G_{-2}A_{target}$	25% $G_{-2}A_{target}$ / 25% $A_{-2}A_{target}$ / 50% $G_{-2}G_{target}$	75% $G_{-2}G_{target}$ / 25% $A_{-2}G_{target}$ / 25% $A_{-2}A_{target}$	25% $G_{-2}G_{target}$ / 50% $A_{-2}A_{target}$ / 25% $A_{-2}G_{target}$	50% $A_{-2}G_{target}$ / 50% $A_{-2}A_{target}$
	25% $G_{-2}G_{target}$ / 75% $G_{-2}A_{target}$	50% $G_{-2}A_{target}$ / 25% $A_{-2}A_{target}$ / 25% $G_{-2}G_{target}$	25% $G_{-2}G_{target}$ / 25% $G_{-2}A_{target}$ / 25% $G_{-2}G_{target}$	25% $G_{-2}G_{target}$ / 25% $G_{-2}G_{target}$ / 75% $A_{-2}A_{target}$	25% $G_{-2}A_{target}$ / 75% $A_{-2}A_{target}$
	0% $G_{-2}G_{target}$ / 100% $G_{-2}A_{target}$	75% $A_{-2}A_{target}$ / 25% $G_{-2}A_{target}$	50% $A_{-2}A_{target}$ / 50% $G_{-2}A_{target}$	25% $G_{-2}A_{target}$ / 75% $A_{-2}A_{target}$	0% $A_{-2}G_{target}$ / 100% $A_{-2}A_{target}$

C Varying Positions of Polymorphisms

Permutation 1
- 25% $G_{-2}G_{target}$
- 25% $T_{-10}G_{target}$
- 25% $G_{-15}G_{target}$
- 25% $T_{-18}G_{target}$
- 25% $G_{-5}G_{target}$
- 25% $T_{-10}G_{target}$
- 25% $G_{-15}G_{target}$
- 25% $A_{-5}A_{target}$

- 25% $T_{-10}G_{target}$
- 25% $C_{-18}G_{target}$
- 25% $G_{-5}A_{target}$
- 25% $A_{-5}A_{target}$

- 25% $G_{-5}G_{target}$
- 25% $G_{-2}A_{target}$
- 25% $A_{-5}A_{target}$
- 25% $G_{-5}A_{target}$

- 25% $G_{-5}A_{target}$
- 25% $T_{-10}A_{target}$
- 25% $G_{-15}A_{target}$
- 25% $A_{-5}A_{target}$

Permutation 2
- 25% $G_{-5}G_{target}$
- 25% $A_{-2}G_{target}$
- 25% $G_{-2}G_{target}$
- 25% $A_{-5}G_{target}$
- 25% $G_{-5}G_{target}$
- 50% $C_{-18}G_{target}$
- 25% $G_{-2}A_{target}$

- 25% $G_{-5}G_{target}$
- 25% $G_{-15}G_{target}$
- 25% $G_{-5}A_{target}$
- 25% $G_{-2}A_{target}$
- 25% $C_{-18}G_{target}$
- 25% $G_{-5}A_{target}$
- 25% $G_{-5}A_{target}$
- 25% $A_{-5}A_{target}$

- 25% $G_{-5}A_{target}$
- 25% $A_{-2}A_{target}$
- 25% $G_{-2}A_{target}$
- 25% $A_{-5}A_{target}$

Figure 8. (**A**) Data points produced by 25% increments of G_{target}/A_{target} templates while varying the percentage of polymorphic G/A 2 bp upstream from the target nucleotide. (**B**) Data points produced by 25% increments G_{target}/A_{target} templates while varying the percentage of polymorphic G/A 5 bp upstream from the target nucleotide. Results for each data point in the −2 (A) and −5 (B) permutations represent an average of 10 experiments. (**C**) Data points produced by diluting the G_{target}/A_{target} 25% increments while varying the percentage of polymorphic nucleotides in the amounts shown to the right.

between runs for each sample. The methylation status of target positions determined by SNaPshot and by sequencing individual clones was similar within 5% (SD 1.0–3.16%) (Figures 9 and 10A).

Some of the experiments revealed multiple C or T peaks. Such patterns were observed only when SNaPshot primers with degenerative positions were used, which indicates that this is due to subtle differences in electrophoretic mobility of primers containing nucleotide differences. If two primer variants, e.g. T_N and C_N, both incorporate a fluorescent ddTTP at the target position, two T peaks may be observed. Results for those primers producing multiple peaks were calculated by incorporating all C signals for a given primer cumulatively and comparing them to the cumulative T signals. Calculating C/T ratios in this way increased the resolution of the method and decreased the SD between replicates on average by 0.7%.

An interesting example of multiple peaks was observed from SNaPshot primer gal1, which generated two T peaks, the ratio of which was consistently 22.5%/77.5% (Figure 10B and C). This finding suggested that the only degenerative nucleotide (Y) in the primer corresponded to a differentially modified cytosine in the genomic DNA. Investigation of 15 clones of bisulfite-treated DNA from this tissue identified that this CpG position was in fact differentially methylated at a frequency of 13%/87% ^{met}C/C. A dedicated SNaPshot primer was synthesized to directly interrogate the CpG position located within the primer gal1 binding region and reported a C/T ratio of 15%/85% (Figure 10D). Additionally, the primer 1/template 1 interrogation (Figure 3) produced a set of two C peaks, template 1 being synthesized with a single degenerative position displaying 20% C and 80% T. In this case, the proportion of the smaller peak of the C peaks to the cumulative C signal ranged from 19 to 23% C between replicates.
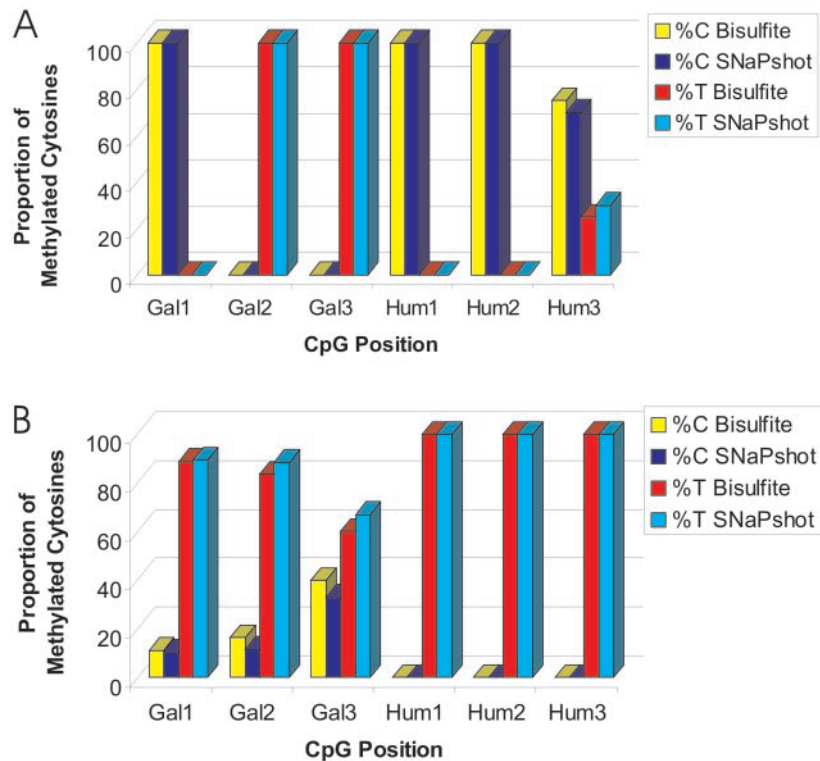
Figure 9. Graphical representation of methylation profiles quantified by sequencing of at least 12 clones of bisulfite-modified genomic DNA and SNaPshot for six CpG dinucleotide positions in bisulfite-modified DNA amplified from brain tissue (**A**) and placenta tissue (**B**).

This suggests that in addition to the target CpG, sometimes a SNaPshot primer containing one degenerative base can quantitatively measure ^{met}C/C proportion at the upstream CpG. Primers with more than one degenerative base, however, have more than two possible variants, making identification of the presence of upstream methylation differences possible, though not quantifiable.

DISCUSSION

In order to expedite mapping of methylated cytosines, we have operationalized the Applied Biosystems method known as SNaPshot to quantitatively discern C/T and G/A proportions in the bisulfite-modified DNA sequence, which reflects ^{met}C/C ratios in the native DNA. In this method, the primer extension reaction is performed using fluorescent 2′,3′-dideoxynucleotide triphosphates that extend the 3′ end of an interrogating primer designed to bind exactly to the target site. The primers, now labeled by specific fluorescent dyes, are measured by an electrophoresis platform and indicate the ratios of the polymorphic C/T and G/A nucleotides at the target position. The mean C/T and G/A ratios generated by all primers displayed an SD range of 2.7–4.6% with the known C/T and G/A values falling within this range. This SD range is similar to other studies in which the SNaPshot method displayed a range between 0 and 3.7% in measuring C/T ratios (14).

Accurate measures of bisulfite-induced C/T (G/A) polymorphisms, however, strongly depend on a perfect match of the interrogating primer and the DNA sequence. When methylatable cytosines are present in the primer annealing region, bisulfite modification will generate quantitative C/T (or G/A) polymorphisms that may lead to partial or full mismatch with the primer. According to our data, a mismatched nucleotide is capable of biasing the results up to 70%, depending on the position of the mismatch. In an unknown system where the upstream polymorphisms could vary by any degree, as in the case of the unknown *a priori* methylation status of bisulfite-treated CpG dinucleotides, it becomes increasingly important to have reliable interrogative capability. A relatively large degree of bias was observed by one group using the SNaPshot approach to interrogate a single CpG that had previously been shown to vary in methylation status between tumor and non-tumor tissue (14,22). Our interpretation is that the observed bias could have been produced by a mismatch in the SNaPshot primer binding region. This might have occurred if primers were designed to bind template DNA containing a methylatable CpG or CpNpG, or in the case of incomplete bisulfite conversion upstream of the target nucleotide. Another possibility is the fact that the authors used peak area instead of peak height for their analysis. While this is less likely to produce the levels of bias observed, ABI technical support warns that peak area readings may vary as a result of primer diffusion in the ABI polymer matrix. Consistently, with other published work (1,8), we detected that primer design in GC-rich areas is of primary importance to address the biasing effects of neighboring polymorphisms. According to our data, primers containing up to five degenerative bases at the positions of the putative mismatches can accurately interrogate a given target. Equal mixture of all possible primers
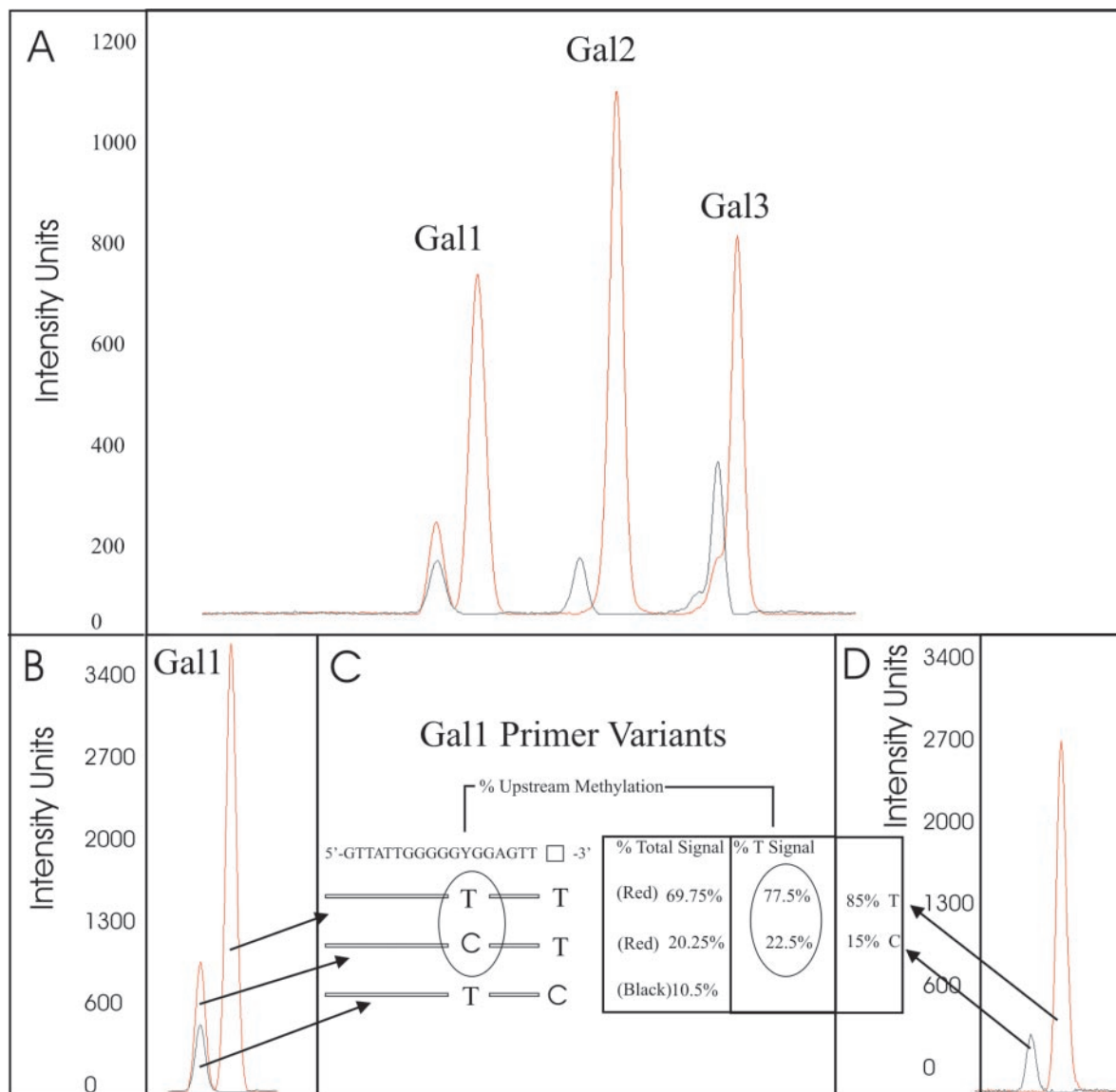
Figure 10. (A) SNaPshot results on bisulfite-modified DNA from brain interrogated with primers gal1, gal2 and gal3. (B) The peak pattern identified with gal1 SnaPshot primer only. (C) A depiction of how the proportion of multiple peaks in the scenario of a single upstream polymorphism were indicative of the methylation profile of the upstream CpG (Y position) measured by gal1 and verified by sequencing of bisulfite-modified genomic DNA. (D) SNaPshot peaks resultant from an interrogation of the gal1 upstream CpG (Y) using primer sequence: 5′-TTGGGGGTTATTGGGGG-3′.

complementary to any of the possible polymorphisms is able to accurately interrogate and reflect the actual percentages of target nucleotides, independent of the quantitative upstream polymorphisms.

It is important to note that degenerative primers may exhibit subtle differential mobility in the electrophoresis platform, which results in multiple SNaPshot peaks. For example, if two primer variants, one with an upstream C and one with an upstream T, both incorporate a fluorescent T at the target position, two T peaks may be observed. This is consistent with the finding that multiple peaks were only observed on templates containing upstream methylation differences (although not all templates containing upstream differences exhibited multiple peaks). We detected that accurate results were obtained from combining together all C intensities belonging to the same target site and comparing them to the cumulative T

intensities. That is, while the migration behavior of primer variants may split a given C or T target signal, the cumulative C and T proportions represented in the data output remain the same relative to each other. Absence of multiple peaks may indicate that there is no sequence variation upstream of the target nucleotide or that the migration dynamics of the moiety of primers is similar.

The presence of multiple peaks may be informative to the method as SNaPshot primer variants may reflect C/T or G/A variation in binding regions of the bisulfite-treated DNA templates. In some cases (e.g. gal1, see Results and Figure 10), the methylation status of a CpG located upstream of the target CpG was identified from a multiple peak pattern. In such cases, only two primer variants are present and a maximum of two peaks can be produced from either or both the C or T signal resultant from upstream methylation. The nucleotide differences between

primer variants causes (i) a preferential binding to the most complementary template and (ii) a spatial segregation of these primer variants in the electrophoretic matrix. Therefore, the percentage of methylation appears to be proportional to the difference of the two multiple peaks. Because multiple peaks may be small, it becomes important to take the signal intensities of the overall reaction into consideration. According to the ABI machine specifications, a peak intensity ranging between 50 and 2000 is optimal for accurate readings. Those peaks with an intensity <50 cannot be differentiated from the background and will not be recognized by the machine. Peaks with very high signal intensities have the danger of pulling up background that could be confused for multiple peaks.

When there is more than one degenerative base in the SNaPshot primer, it is not trivial to assign multiple peaks to a specific CpG in the primer annealing region. In order to avoid possible overlaps of peaks belonging to different primers, one possibility is to increase the size difference of those primers designed. If the result is still in question, re-running a given SNaPshot primer individually on the same template can be recommended. In general, since SNaPshot primers with degenerative bases can identify the presence of ^{met}C/C in the primer binding site, DNA methylation analysis can be optimized by a two-tiered approach. The first tier would use a single degenerate SNaPshot primer to scan high CpG density areas for possible complex peak (i.e. differential DNA methylation) patterns. The second round of SNaPshot primers may then be designed to interrogate specific CpG dinucleotides implicated in the first round of SNaPshot experiments.

In comparison to other methods used to quantitatively estimate C/T (G/A) ratios, the SNaPshot approach exhibits several advantages. Adding non-binding 'tails' to the 5′ ends of multiple SNaPshot primers changes their lengths relative to each other, and results in a spatial segregation during electrophoresis. This multiplexing technique enables multiple CpG sites to be interrogated in a single reaction. In our experiments, six primer extension reactions were multiplexed without any evidence for compromising the quality of interrogations. Given longer amplicons, the SNaPshot method can potentially accommodate 10 primers according to the manufacturer's specifications (Applied Biosystems, Inc., Foster City, CA) (23). It should be noted that another approach has avoided the issues of reading multiple primer variants produced by primers using degenerate bases by incorporating tagged bases in the interrogating primer. This allows degenerative bases to be removed, thereby creating a core sequence of uniform size, which is in turn measured by MALDI mass spectrometry (15). Major differences between the SNaPshot and the MALDI mass spectrometry approaches involve the generation of a 4 nt core sequence to avoid confounding signal and the use of B-cyanoethyl phosphoramidite nucleotide tags as opposed to fluorescent ddNTPs. While the two assays are very similar at the early stages, advantages of the electrophoresis platform lie in a greater multiplexing capability and no restrictions regarding incorporation of a degenerate base within 4 nt of the target position. Additionally, the production of multiple peaks in the electrophoresis platform adds informativeness over a larger DNA area using fewer primers, which translates into a lower cost.

While the multiple peaks highlighted in this method allow for a two-tiered investigation of larger CpG-rich regions of DNA, other technologies such as pyrosequencing (16–18) and the newly developed ESME approach (19) enable identification of methylation profiles of multiple CpG dinucleotides as well as the surrounding sequence in a single reaction. The situation in pyrosequencing, however, would be optimal when interrogating a number of CpG dinucleotides within close proximity, while the SNaPshot method can interrogate any CpG for which a primer has been designed without spatial constraint. The detection limit of methylation differences in the ESME approach is 20% or greater, while pyrosequencing and SNaPshot can provide accurate measures of methylation within ~5%. While the pyrosequencing and the ESME approaches offer attractive features, the added informativeness, ease and affordability of the modified SNaPshot approach makes it a convenient alternative for those laboratories with an electrophoresis platform.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Dahl,C. and Guldberg,P. (2003) DNA methylation analysis techniques. *Biogerontology*, **4**, 233–250.
2. Petronis,A., Gottesman,II, Kan,P., Kennedy,J.L., Basile,V.S., Paterson,A.D. and Popendikyte,V. (2003) Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr. Bull.*, **29**, 169–178.
3. Gruenbaum,Y., Stein,R., Cedar,H. and Razin,A. (1981) Methylation of CpG sequences in eukaryotic DNA. *FEBS Lett.*, **124**, 67–71.
4. Clark,S.J., Harrison,J. and Frommer,M. (1995) CpNpG methylation in mammalian cells. *Nature Genet.*, **10**, 20–27.
5. Norton,N., Williams,N.M., Williams,H.J., Spurlock,G., Kirov,G., Morris,D.W., Hoogendoorn,B., Owen,M.J. and O'Donovan,M.C. (2002) Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum. Genet.*, **110**, 471–478.
6. Sokolov,B.P. (1990) Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res.*, **18**, 3671.
7. Hong,K.M., Yang,S.H., Guo,M., Herman,J.G. and Jen,J. (2005) Semiautomatic detection of DNA methylation at CpG islands. *Biotechniques*, **38**, 354, 356, 358.
8. El-Maarri,O., Herbiniaux,U., Walter,J. and Oldenburg,J. (2002) A rapid, quantitative, non-radioactive bisulfite-SNuPE-IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res.*, **30**, e25.
9. El-Maarri,O. (2004) SIRPH analysis: SNuPE with IP-RP-HPLC for quantitative measurements of DNA methylation at specific CpG sites. *Methods Mol. Biol.*, **287**, 195–205.

10. Eads,C.A., Danenberg,K.D., Kawakami,K., Saltz,L.B., Blake,C., Shibata,D., Danenberg,P.V. and Laird,P.W. (2000) MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.*, **28**, E32.

11. Gonzalgo,M.L. and Jones,P.A. (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.*, **25**, 2529–2531.

12. Gonzalgo,M.L. and Jones,P.A. (2002) Quantitative methylation analysis using methylation-sensitive single-nucleotide primer extension (Ms-SNuPE). *Methods*, **27**, 128–133.

13. Nguyen,T.T., Mohrbacher,A.F., Tsai,Y.C., Groffen,J., Heisterkamp,N., Nichols,P.W., Yu,M.C., Lubbert,M. and Jones,P.A. (2000) Quantitative measure of c-abl and p15 methylation in chronic myelogenous leukemia: biological implications. *Blood*, **95**, 2990–2992.

14. Uhlmann,K., Brinckmann,A., Toliat,M.R., Ritter,H. and Nurnberg,P. (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis*, **23**, 4072–4079.

15. Tost,J., Schatz,P., Schuster,M., Berlin,K. and Gut,I.G. (2003) Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.*, **31**, e50.

16. Tost,J., Dunker,J. and Gut,I.G. (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques*, **35**, 152–156.

17. Fakhrai-Rad,H., Pourmand,N. and Ronaghi,M. (2002) Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.*, **19**, 479–485.

18. Colella,S., Shen,L., Baggerly,K.A., Issa,J.P. and Krahe,R. (2003) Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques*, **35**, 146–150.

19. Lewin,J., Schmitt,A.O., Adorjan,P., Hildmann,T. and Piepenbrock,C. (2004) Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. *Bioinformatics*, **20**, 3005–3012.

20. Rakyan,V.K., Hildmann,T., Novik,K.L., Lewin,J., Tost,J., Cox,A.V., Andrews,T.D., Howe,K.L., Otto,T., Olek,A. *et al.* (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.*, **2**, e405.

21. Hajkova,P., el-Maarri,O., Engemann,S., Oswald,J., Olek,A. and Walter,J. (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol. Biol.*, **200**, 143–154.

22. Uhlmann,K., Marczinek,K., Hampe,J., Thiel,G. and Nurnberg,P. (1999) Changes in methylation patterns identified by two-dimensional DNA fingerprinting. *Electrophoresis*, **20**, 1748–1755.

23. Lindblad-Toh,K., Winchester,E., Daly,M.J., Wang,D.G., Hirschhorn,J.N., Laviolette,J.P., Ardlie,K., Reich,D.E., Robinson,E., Sklar,P. *et al.* (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet.*, **24**, 381–386.