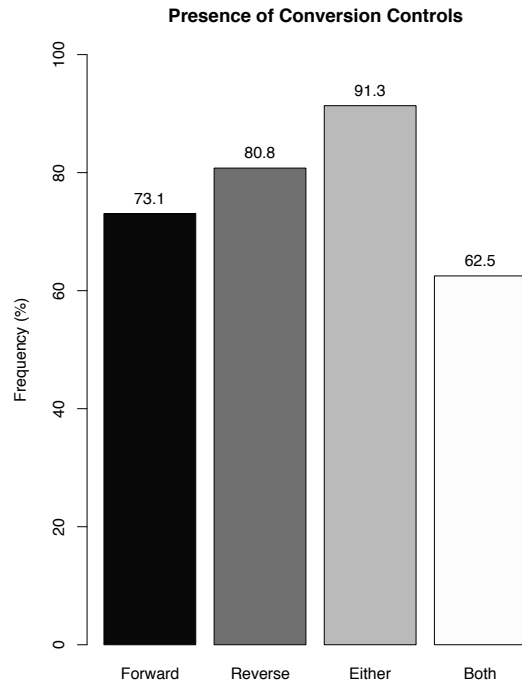


SUPPLEMENT

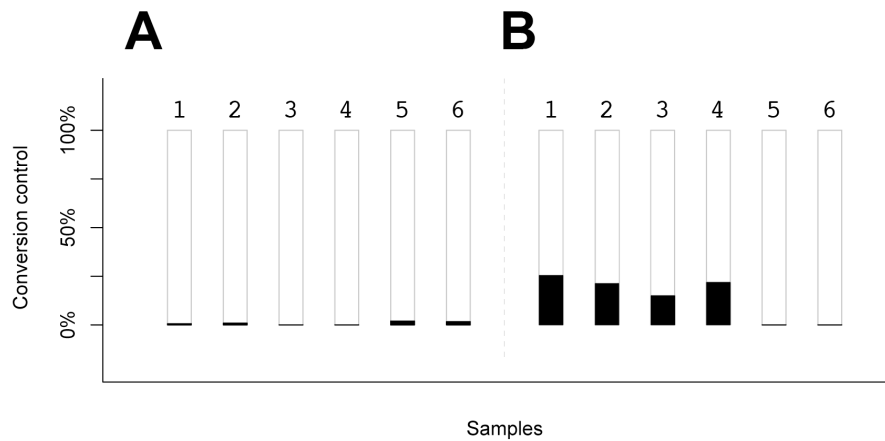
A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
Thompson *et al.*



Supp. Fig. 1. Performance of conversion control identification using amplicons from Sequenom’s Standard EpiPanel (rev. December 2007). One-hundred and four amplicons for thirteen genes (*AKT1*, *ARF1*, *BRAF*, *BRCA1*, *C.MYC*, *CDH1*, *CDH13*, *CDKN2A*, *CEBPA*, *DAPK1*, *DUSP1*, *EGFR*, and *ESR1*) from Sequenom’s Standard EpiPanel were analyzed for the presence or absence of usable conversion controls. The bar plot shows the frequency of amplicons with usable conversion controls on each individual strand (“Forward” and “Reverse”, relative to the amplicon design), on either one or both strands (“Either”), or only on both strands (“Both”).

SUPPLEMENT

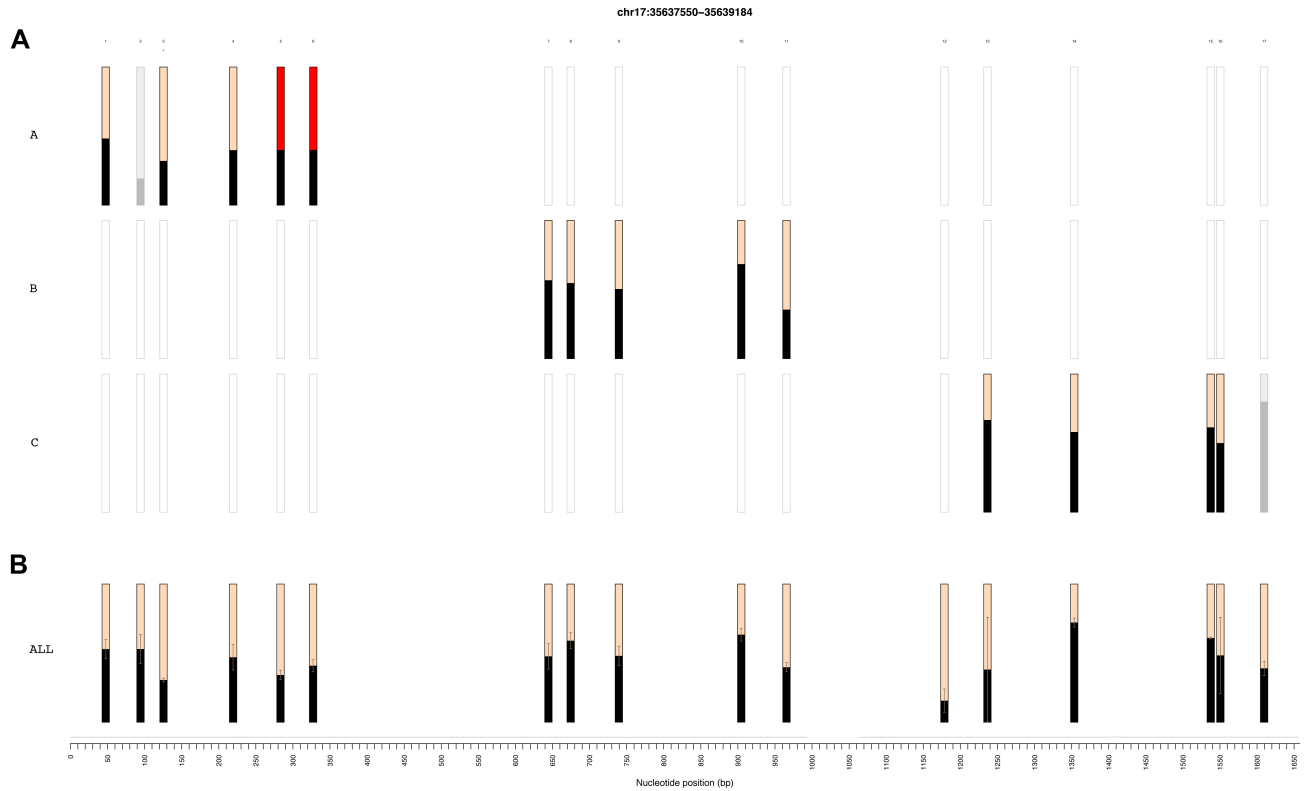
A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
Thompson *et al.*



Supp. Fig. 2. Measurement of conversion controls for two amplicons shows variable extent of bisulphite conversion. (A) Conversion controls were measured for an amplicon of the rat genome (chr1:27718536-27718918, rn4 Nov. 2004 assembly, UCSC Genome Browser) using six biological replicates (numbered 1-6). Bar height (black) depicts the percentage of unconverted cytosines detected, with 0% indicating complete conversion of measured cytosines and 100% indicating a complete failure of conversion. (B) Conversion controls were measured for a second amplicon of the rat genome (chr17:48916975-48917295, rn4 Nov. 2004 assembly, UCSC Genome Browser) using six biological replicates (numbered 1-6) from a separate bisulphite conversion reaction. As before, bar height (black) depicts the percentage of unconverted cytosines detected, with 0% indicating complete conversion of measured cytosines and 100% indicating a complete failure of conversion. Four of six samples show significant retention of unconverted cytosines (numbered 1-4), as measured by conversion controls.

SUPPLEMENT

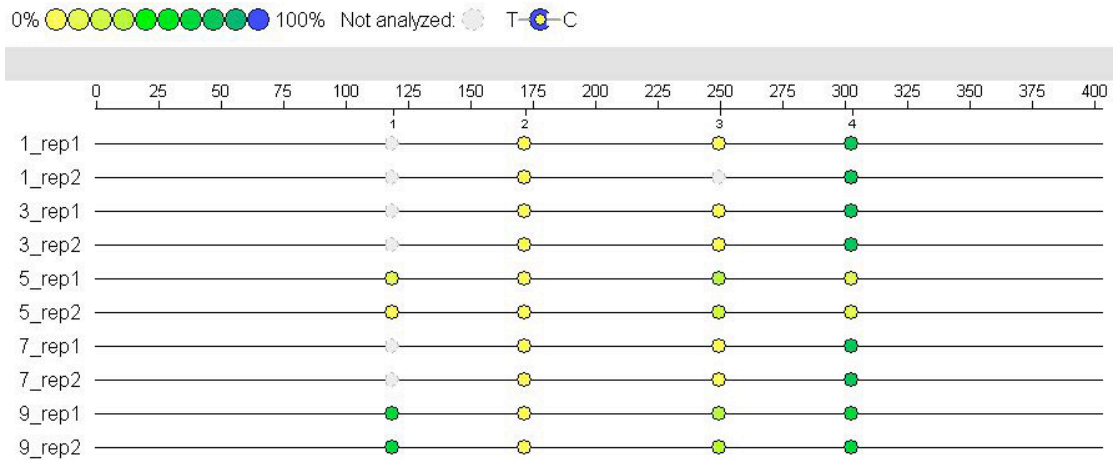
A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
 Thompson *et al.*



Supp. Fig. 3. Integration of MassArray data for multiple samples from three independent, adjacent amplicons upstream of the mouse *Oct4* gene. (A) Methylation measurements shown are calculated from T-cleavage reaction data for individual samples “A,” “B,” and “C” for the amplicons representing chr17:35637550-35637980, chr17:35637956-35638532, and chr17:35638604-35639184, respectively (mouse mm9 Jul. 2007 assembly, UCSC Genome Browser). The bar height (black) denotes percent methylation on a scale from 0 (low) to 100% (high), red stars indicate user-specified “required” sites. CG sites having a molecular weight overlap with other CG-containing fragments are colored in red. CG sites that are not assayable are shown in gray outline, with any data unexpectedly recovered from such loci displayed in shades of gray. (B) MassArray data for both the T and C-cleavage reactions for four samples, combined across all three amplicons in (A). As before, black bar height denotes percent methylation on a scale from 0 (low) to 100% (high), error bars indicate median absolute deviation.

SUPPLEMENT

A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
Thompson *et al.*



Supp. Fig. 4. An example of the epigram visualization generated by Sequenom's EpiTYPER software. The T-cleavage reaction was performed for an amplicon of the rat genome (chr1:27718536-27718918, m4 Nov. 2004 assembly, UCSC Genome Browser) on five different biological samples, each with a pair of technical replicates (note that the same underlying data was used to generate Figure 2). Methylation data is displayed as color-filled circles with yellow indicating 0% methylation and darker colors of green indicating increasing levels of methylation, with blue as 100%. Empty gray circles correspond to CG sites that failed analysis (note that in this case, the missing data correlates directly with SNP#1 identified in Figure 2).

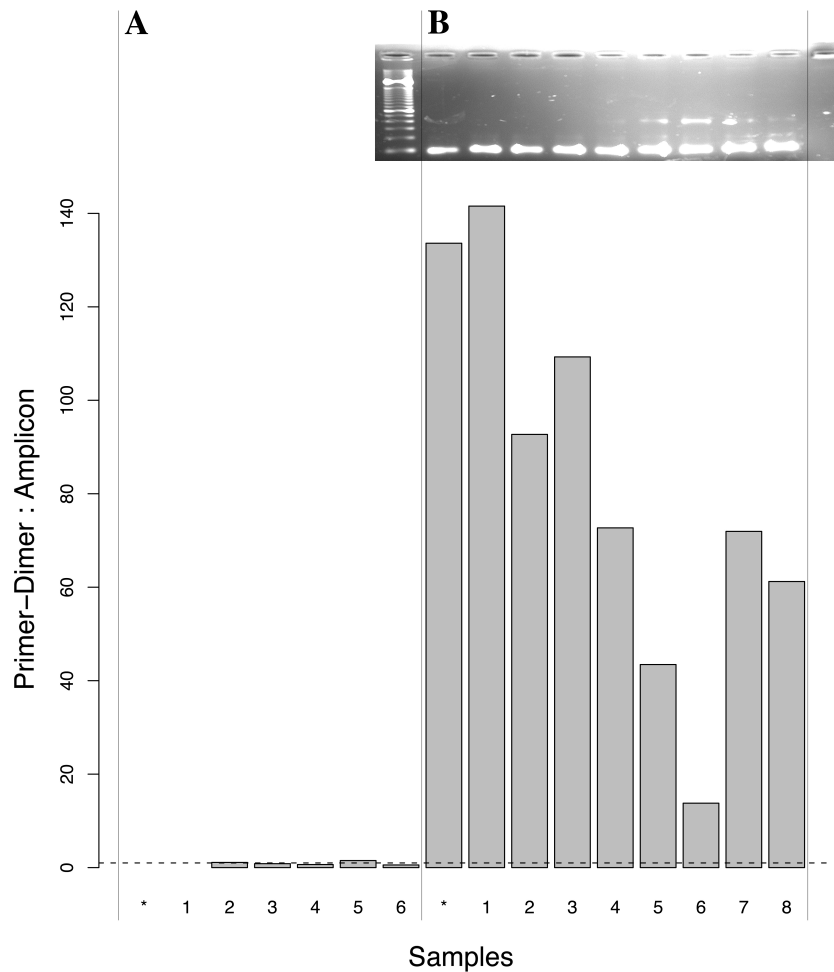
SUPPLEMENT

A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
Thompson *et al.*

Supp. Fig. 5. *In silico* assay prediction. (A) Putative fragmentation patterns are shown for T and C-cleavage reactions on both the plus and minus strands of an amplicon of the rat genome (chr1:70508994-70509366, rat m4 Nov. 2004 assembly, UCSC Genome Browser). CG dinucleotides (filled circles) are numbered and color-coded according to their ability to be assayed, where gray indicates that the CG is located on a fragment whose molecular weight is outside the usable mass window, red indicates a molecular weight overlap with another fragment, and blue indicates a uniquely assayable site. Linked arrowheads denote molecular weight overlaps between multiple CG-containing fragments. Fragmentation patterns are shown in corresponding colors, with the addition of green fragments indicating usable conversion controls. Yellow highlights represent tagged or primer sequences, while lavender highlights denote user-specified “required” sites.

SUPPLEMENT

A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
 Thompson *et al.*



Supp. Fig. 6. Primer dimer estimation using MassArray data successfully detects primer dimer contamination of samples. (A) Primer dimer estimation was applied to an amplicon of the rat genome (chr15:23184831-23185302, rat rn4 Nov. 2004 assembly, UCSC Genome Browser) showing no detectable primer dimers by gel electrophoresis. The average ratio of primer-unique peak intensities to amplicon-unique peak-intensities was calculated by the `estimatePrimerDimer()` function from MassArray data for one water control (*) and six samples (labeled 1-6). Bars are shown for each sample along the x-axis, with heights indicating estimated primer to amplicon ratios. (B) Primer dimer estimation was applied to an additional amplicon of the rat genome (chr15:23185922-23186278, rat rn4 Nov. 2004 assembly, UCSC Genome Browser) which showed significant levels of primer dimer by gel electrophoresis (image included above bar plot, with a 100bp ladder and each following lane matched to its appropriate sample). Water control is indicated by an asterisk (*) and eight samples are labeled (1-8). The height of each bar corresponds, as before, to the ratio of primer-derived signal to amplicon-derived signal. The horizontal dashed line across both panels indicates a ratio of 1:1, where levels of primer-derived signal are equivalent to amplicon-derived signal indicating no presence of primer dimer.

SUPPLEMENT

A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
Thompson *et al.*

Supp. Table 1. Primer pairs and cycling conditions for amplicons used in this study.

		PRIMERS		T _m (°C)
AMPLICON		Forward*	Reverse ^{T7}	
mm9	chr17:35637550-35637980	GTTTTTGAGTGTTTATTTGGGATTG	CCAAAAAATATCACTACCTTCCCT	63
	chr17:35637956-35638532	AGGGAAGGTAGTGATATTTTTTTGG	AAACTACCCCTAATAACCAAAACCC	63.5
	chr17:35638604-35639184	TAAAAAGATTTGTAGGGGGTAGGAG	CAAAAAATCAAAAATTCAAAACC	62
m4	chr1:27718536-27718918	GTTTGAATTTGTTTTTTTTTGTAGAGTTG	CCCAACCAAACTATACTTAACATATTTAC	60
	chr17:48916975-48917295	TATTTGGTATGGTATGAATGGGTATG	CTCAACTCCAATAAACAAAAATATC	59
	chr1:221405426-221405799	TTAGGTTTAAAGTTTAGATTAGTTTGAA	TACACCAACCACTAAACAATAAAAATC	58
	chr11:70508994-70509366	TGGAATATTTAGTAAAGATTAAGTAGGA	TCTCAAAAACCTAAAATTACAAAAATC	58
	chr15:23184831-23185302	GGTGTAAAGATTTTTTGATTAATTTTTATGA	CAACAACAACAACCCCAAAATTAC	57.5
	chr15:23185922-23186278	GATAGATGTTTTGAAGTTAATGGAAGA	AAAAAATCAAAACCAATTCAACTTCC	58.5

T_m, annealing temperature in degrees Celsius used for PCR amplification; mm9, mouse genome version mm9 Jul. 2007, UCSC Genome Browser; m4, rat genome version m4 Nov. 2004, UCSC Genome Browser.

* Forward primer sequences listed all contain a 10bp tag at their 5' ends (AGGAAGAGAG)

^{T7} Reverse primer sequences listed all contain a 31bp tag at their 5' ends (CAGTAATACGACTCACTATAGGGAGAAGGCT)

SUPPLEMENT

A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry
Thompson *et al.*

Supp. Table 2. Signal-to-noise ratio (SNR) data for all expected and novel peaks corresponding to two SNP-containing fragments of an amplicon of the rat genome (chr1:27718536-27718918).

SAMPLE	avg(SNR)	SNP = chr1:27718644 (G->A)				SNP = chr1:27718776 (GTGAG->TG)		
		<u>WT PEAKS</u>		<u>SNP PEAKS</u>		<u>WT PEAKS</u>		<u>SNP PEAKS</u>
		2798.772 Da ¹	2814.771 Da ¹	1272.797 Da ²	1561.982 Da ²	2509.587 Da ¹	2525.586 Da ¹	2236.831 Da*
1	8.528393	1.5615	0.0133	2.8763	12.994	5.6134	0	21.2878
1	7.789843	1.0108	0.0394	4.9997	14.4343	3.5464	0	18.3859
3	14.27818	1.1674	0	13.4333	25.8314	6.6402	0	22.9658
3	13.8562	1.964	0.1283	6.7829	18.4272	7.8867	0	23.1186
5	11.22155	60.1022	6.9064	5.7959	11.164	38.6727	17.5765	0
5	7.254444	35.0049	3.5652	4.4536	8.9225	23.535	8.7874	0
7	12.74913	0.6665	0	11.9235	22.8064	6.0861	0	16.6564
7	13.158	1.8646	0.097	10.2238	13.3508	6.5859	0	20.6557
9	13.86768	18.436	29.9326	13.668	14.8575	26.1533	16.4928	8.9787
9	14.546	20.6775	36.5588	13.879	20.1773	27.0928	16.8212	8.6664

avg(SNR), sum of all SNR data for each sample divided by the expected number of fragments from the amplicon sequence; WT, set of peaks that are expected to arise based on amplicon sequence; SNP, sequence polymorphism affecting fragmentation pattern, with corresponding set of peaks of alternate molecular mass; Da, molecular mass in Daltons of a given peak.

² A peak of indicated mass is expected to arise from two independent fragments within the amplicon

¹ A peak of indicated mass is expected to arise from a single unique fragment within the amplicon

* A peak of indicated mass is not expected to arise from the amplicon