Genomics 102 (2013) 38–46

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Exploring the utility of human DNA methylation arrays for profiling mouse genomic DNA

Nicholas C. Wong [a,g,*], Jane Ng [a,1], Nathan E. Hall [b,c,1], Sebastian Lunke [d], Marika Salmanidis [e], Gabriela Brumatti [e], Paul G. Ekert [e], Jeffrey M. Craig [f], Richard Saffery [a]

[a] Cancer and Disease Epigenetics Group, Murdoch Childrens Research Institute, Department of Paediatrics, The University of Melbourne, Royal Children's Hospital, Parkville, Victoria 3052, Australia
[b] Department of Genetics, La Trobe University, Melbourne, Victoria, Australia
[c] Life Sciences Computation Centre, Victorian Life Sciences Computation Initiative, VIC 3010, Australia
[d] Translational Genomics Laboratory, Centre for Translational Pathology, Department of Pathology, The University of Melbourne, Parkville, Victoria 3010, Australia
[e] Cell Signalling and Cell Death Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia
[f] Early Life Epigenetics, Murdoch Childrens Research Institute, Department of Paediatrics, The University of Melbourne, Royal Children's Hospital, Parkville, Victoria 3052, Australia
[g] Ludwig Institute for Cancer Research, Olivia Newton John Cancer and Wellness Centre, Austin Hospital, Heidelberg, Victoria, 3084, Australia

ARTICLE INFO

Article history:
Received 6 June 2012
Accepted 21 April 2013
Available online 29 April 2013

Keywords:
DNA methylation
Infinium BeadChip
Xeno-array
Human
Mouse
Epigenetics

ABSTRACT

Illumina Infinium Human Methylation (HM) BeadChips are widely used for measuring genome-scale DNA methylation, particularly in relation to epigenome-wide association studies (EWAS) studies. The methylation profile of human samples can be assessed accurately and reproducibly using the HM27 BeadChip (27,578 CpG sites) or its successor, the HM450 BeadChip (482,421 CpG sites). To date no mouse equivalent has been developed, greatly hindering the application of this methodology to the wide range of valuable murine models of disease and development currently in existence. We found 1308 and 13,715 probes from HM27 and HM450 BeadChip respectively, uniquely matched the bisulfite converted reference mouse genome (mm9). We demonstrate reproducible measurements of DNA methylation at these probes in a range of mouse tissue samples and in a murine cell line model of acute myeloid leukaemia. In the absence of a mouse counterpart, the Infinium Human Methylation BeadChip arrays have utility for methylation profiling in non-human species.

1. Introduction

The role of DNA methylation in various physiological processes and related diseases is becoming increasingly apparent as the tools to study this epigenetic mark at the genome-wide level become more developed. These can largely be categorized as those that directly measure methylation at specific CpG sites within a biological sample and those that utilize less specific affinity-based approaches (such as antibody enrichment) to generate a profile of regional methylation for multiple CpG sites [1]. The recent development of the sensitive and highly reproducible Infinium Human Methylation BeadChip arrays [2] has greatly expanded the potential of studies exploring genome-scale methylation in development and disease. However, for model organisms such as the mouse, equivalent tools to study DNA methylation at a genome-scale are not currently available. More labour intensive and less sensitive methodologies are still required.

Here we investigated the feasibility of interrogating the mouse genome for DNA methylation using the Illumina Infinium HM27 and HM450 platforms. Such cross species array experiments have been performed in the past for gene expression analysis of species where a microarray has not been available [3,4]. Probes from the array of a closely related species that perfectly match the test species genome have been used to measure gene expression reliably [5].

The Illumina Infinium chemistry relies on the hybridization of bisulfite converted genomic DNA to 50-mer probes on the BeadChip. Base extension chemistry is then employed to incorporate a fluorescent nucleotide onto each hybridized probe [6]. The Infinium I chemistry used by the HM27 BeadChip uses two probes (one for the methylated and another for the unmethylated allele) per probe sequence and interrogated within the same colour channel. While the Infinium II chemistry uses one probe employing two colour chemistries to measure DNA methylation [2], the HM450 BeadChip uses a combination of both Infinium Type I and Type II chemistries which needs to be accounted for during downstream analysis [7,8].

Here, we determined the utility of the Infinium HM27 and HM450 BeadChip for mouse genomic DNA. We first performed *in-silico* mapping analysis using four popular short read aligners to the mouse reference genome in base space to determine the number of uniquely

* Corresponding author at: Ludwig Institute for Cancer Research, Olivia Newton John Cancer and Wellness Centre, Austin Hospital, Department of Pathology, The University of Melbourne, Australia.
E-mail address: nwon@unimelb.edu.au (N.C. Wong).
[1] Equal contribution.

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.04.014

aligned probe sequences that could directly interrogate DNA methylation in mouse genomic DNA. We also performed mapping of the probes to the bisulfite converted reference genome employing four bisulfite sequence aligners. Mouse samples were then analysed using HM27 and HM450 BeadChips in parallel to human samples to empirically determine if the probes identified by in-silico analysis are informative. Several probes demonstrating differential DNA methylation in the mouse samples were replicated using SEQUENOM EpiTYPER. We present a list of candidate probes that could be used to investigate mouse DNA methylation using the Human Methylation BeadChip in the absence of a mouse-specific counterpart .

2. Results

2.1. HM27 BeadChip in-silico analysis

We define unique mapping 100% identity across all 50 bp of the query sequence to a single region of the reference genome. Therefore the unique annotation in Table 1 is directly comparable between the short read aligners used.

SHRiMP, BWA and Novoalign identified comparable numbers of uniquely mapped HM27 probes to mm9 with 219, 219 and 236 probes respectively. 219 probes were identified using Blat (Table 1 and Fig. 1A). These 219 probes were called by all four aligners (Fig. 1A, Supplementary Table 1) and are likely to represent highly conserved sequences between the human and mouse reference genomes. To determine the mappability of the probes, we performed probe sequence alignment on the human reference genome, both with (hg19) and without reference haplotype sequences (hg19 nohap).

Of the 27,578 probes on the HM27 BeadChip, SHRiMP identified 26,905 uniquely mapped probes while BWA, Novoalign and Blat identified 26,951, 26,946 and 26,938 uniquely aligned probes respectively (Table 1). 26,740 probes were found to uniquely align to hg19 by all four aligners (Supplementary Table 2). No alignment was found for two probes (cg03886110 and cg22467071) to the reference hg19 genome. The remaining probe sequences appear to align to multiple non-unique locations in the hg19 reference genome. We found that these probes mapped to the haplotype sequences of the hg19 reference genome. We then repeated our alignment analysis to a reference genome where the haplotype sequences were excluded, the number of unique mapped probes increased when aligned to SHRiMP, BWA, Novoalign and Blat (Table 1).

2.2. HM450 BeadChip in-silico analysis

We extended our analysis to the current generation HM450 BeadChip and performed probe sequence alignment to both the mouse (mm9) and human (hg19) reference genomes. Of the 485,577 probes, SHRiMP, BWA and Novoalign identified 1988, 1989 and 2076 probes with unique alignment to mm9 respectively (Table 1). Blat identified 1988 probes uniquely aligning to mm9. Using all four aligners, 1989 probes aligned uniquely to mm9 (Fig. 1B, Supplementary Table 3). We found a substantial number of probe sequences that appear to not align uniquely to our reference genome hg19 with the number of perfectly aligned probes ranging from 467,862 to 485,277 across the four base space aligners used (Table 1). As with the HM27 BeadChip, these probes aligned to the haplotype sequences we included in our analysis. Accounting for haplotype, the number of unique mapping probes increased ranging from 482,785 to 485,293 (Table 1).

2.3. Accounting for bisulfite conversion of genomic DNA

Sodium bisulfite reduces the sequence complexity of genomic DNA by converting all unmethylated cytosine bases to uracil that present as a thymidine after PCR amplification [9]. In addition, the DNA strands become non-complementary. As the Infinium BeadChip chemistry relies on the hybridization of probes to bisulfite converted genomic DNA, there is an increased likelihood that probes thought to uniquely align in normal genome space can align to multiple locations across the genome after bisulfite conversion. The increased degeneracy induced by bisulfite conversion could also increase the cross-reactivity of array probes to other species such as mouse.

To investigate this, we took the probe sequences from the HM27 and HM450 BeadChips, and bisulfite converted these sequences in-silico and aligned them to the bisulfite converted mouse genome using four short read alignment algorithms specifically designed for bisulfite massively parallel sequencing. We used Novoalign Bisulfite Mode (http://www.novocraft.com), Brat [10], Bsmap [11] and Bismark [12].

Given the variation in output formats from these bisulfite short read aligners, it was a challenge to extract probes that mapped uniquely according to our definition above. Therefore, the mapped probes from each aligner are presented in Table 1 and Figs. 1C and D.

Table 1
Infinium probe alignment summary.

Genome space	Aligner	Alignments	HM27			HM450		
			mm9	hg19	hg19 nohap	mm9	hg19	hg19 nohap
Base space	SHRiMP	Mapped	1867	27,545	27,545	21,893	484,492	484,487
		Mapped 50M	224	27,532	27,532	2017	483,256	483,251
		Unique	219	26,905	27,148	1988	470,509	482,785
		Multiple	5	627	384	29	12,747	466
	BWA	Mapped	1266	27,574	27,574	13,196	485,764	485,764
		Unique	219	26,951	27,192	1989	472,898	485,185
	Novoalign	Mapped	2515	27,575	27,575	30,635	485,764	485,764
		Unique	236	26,946	27,189	2076	472,999	485,293
	Blat	Mapped[a]	224	27,565	27,568	2017	485,277	485,261
		Unique	219	26,938	27,183	1988	472,513	484,790
		Multiple	5	627	385	29	12,764	471
Bisulfite space	Novoalign BM	Mapped	3739	27,576	ND	51,259	485,763	ND
		Unique	3516	26,847	ND	47,632	470,417	ND
		Not mapped	23,839	2	ND	435,427	923	ND
	Bismark	Mapped	1507	27,574	27,574	17,329	485,764	485,764
		Unique	1451	26,796	27,037	16,604	468,983	481,260
		Not mapped	26,071	4	4	469,358	923	923
	Bsmap	Mapped	1352	27,574	27,574	14,295	485,765	485,765
	Brat	Mapped	333	26,846	ND	2986	470,315	ND

Unique requires alignment of 50M and no mismatches.
[a] Blat does not output in SAM format, we required 100% identity mapping alignment and filtered for multiple hits.

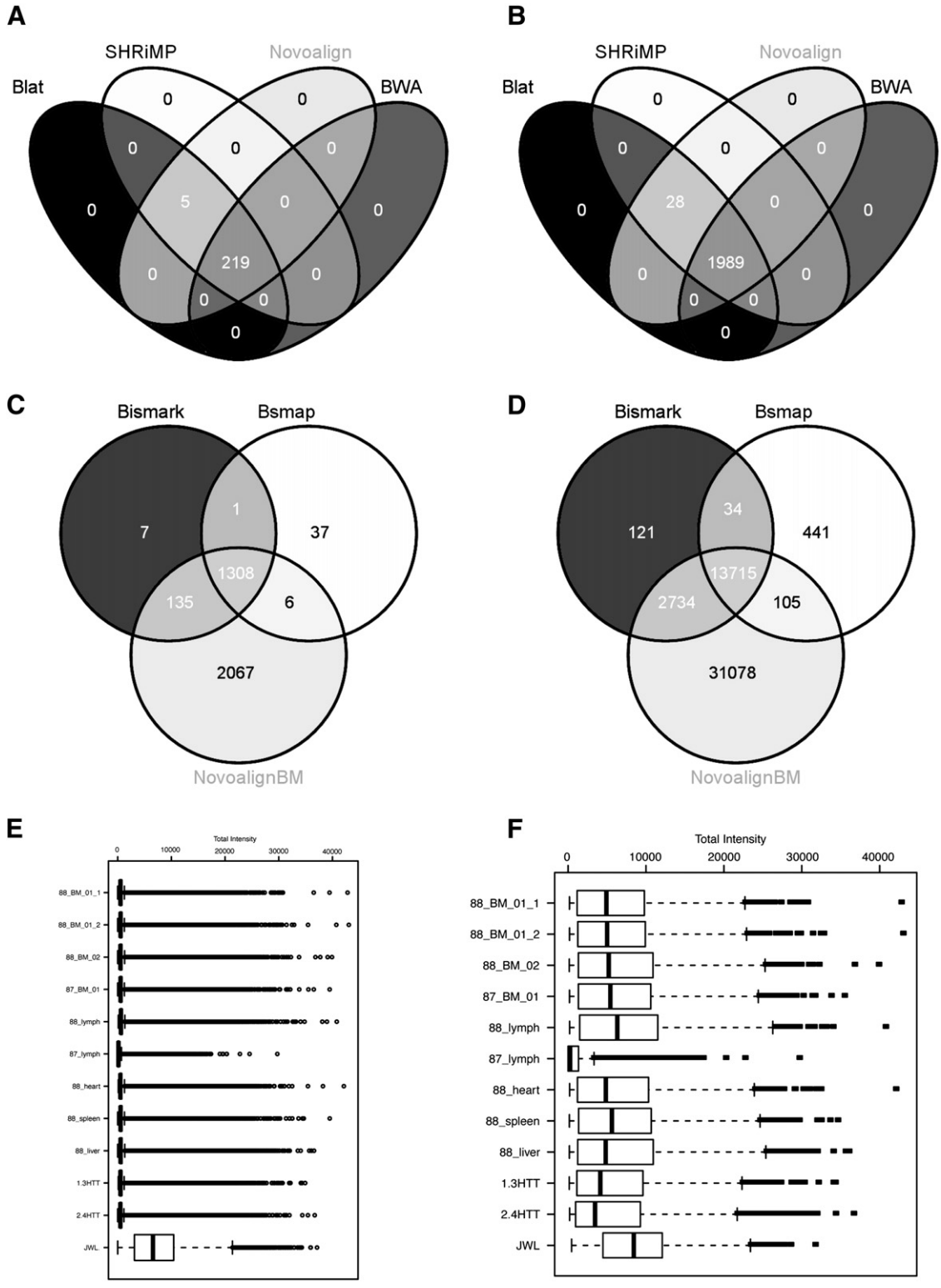


Fig. 1. A. Venn diagram of uniquely mapping HM27 probes aligned using Blat, Novoalign, BWA and SHRiMP using default parameters. 219 probes were commonly identified as uniquely aligning to the mm9 reference genome. Unique was defined as 100% identity to the reference genome with only one alignment. B. Venn diagram of uniquely mapping HM450 probes aligned using Blat, Novoalign, BWA and SHRiMP using default parameters. 1975 probes were commonly identified as uniquely aligning to the mm9 reference genome. C. Venn diagram of uniquely mapping HM27 probes aligned using three bisulfite short read aligners to the bisulfite converted mm9 reference genome. The aligners used included Novoalign Bisulfite Mode (http://www.novocraft.com), Bsmap [11] and Bismark [12]. D. Venn diagram of uniquely mapping HM450 probes to the bisulfite converted mm9 reference genome. E. The signal intensity distribution of all 485,577 probes found on the HM450 BeadArray. Mouse samples demonstrate low probe intensity distributions indicative of a failed array. Tissue samples were taken from two C57BL/6 wild type mice (87 and 88) from bone marrow (BM), lymph nodes (lymph), heart, spleen and liver. The suffix numbers denote the replicates performed. 1.3HTT and 2.4HTT represent the two *Hoxb8* mouse cell lines. JWL is a human lymphoblastoid cell line that was used as a control. F. The signal intensity distribution of 13,715 uniquely aligning probes in bisulfite space. With the exception of 87_lymph which failed overall quality control, the intensity distribution of these probes is comparable to the human sample, JWL.

After accounting for bisulfite conversion, the number of probe sequence alignments from the HM27 array to the mouse mm9 reference genome increased to 333 using Brat and up to 3516 probes using Novoalign in bisulfite mode (Table 1). In total, 123 probes uniquely aligned to the bisulfite converted mouse reference genome using all four aligners. However, 1185 probes were in consensus with 3 out of 4 bisulfite short read aligners used (Novoalign Bisulfite Mode, Bsmap and Bismark (data not shown)). Given the variation of alignment output from these aligners, it was difficult to identify unique alignments. Further, given the paucity of information on alignments made by Brat and that Brat listed the least number of alignments, we believe Brat may be overly conservative in calling unique alignments. We observed this further by performing alignment to the probes found on the HM450 BeadChip and subsequently confirmed by assessing empirical signal intensity from mouse samples analysed by this array (see below). With all four aligners, 182 probes were identified to commonly align while 13,533 probes were found using Novoalign, Bsmap and Bismark (Supplementary Fig. 1A). We therefore used these three bisulfite short read aligner for downstream analysis (Supplementary Table).

Using Novoalign Bisulfite Mode, Bsmap and Bismark we found 1308 and 13,715 probes to uniquely align to the mm9 reference genome after bisulfite conversion (Figs. 1C and D) on the HM27 and HM450 Infinium BeadChips respectively (Supplementary Tables 5 and 6 respectively). We also performed the same in-silico alignment of probes to the bisulfite converted hg19 reference human genome. The majority of probes from both the HM27 and HM450 BeadChips were found to align to the bisulfite-converted human genome sequence (Table 1). However, we observed notable differences between the number of probes that mapped to the hg19 reference genome and the number of unique alignments suggesting that a subset of probes align to multiple locations across our human reference genome. As we included the reference haplotype sequences within our reference genome, the majority of these probes aligned to these sequences. After removing haplotype sequences from our reference genome we found that the number of mapped unique probes increased (Table 1). These are presented in in Supplementary Tables 7 and 8.

2.4. Infinium HM450 BeadChip DNA methylation analysis

To determine the extent of the utility of the probes we have identified in-silico for mouse DNA methylation analysis, we first looked at the distribution of probe signal intensity across the HM450 array after analysis with a panel of genomic DNA derived from various mouse tissues and cell lines. We surmised that the intensity distribution of Infinium probes we predict to interrogate mouse genomic DNA samples would have a similar distribution to an analysis of a human genomic DNA sample using the same probes. The intensity distribution of all 485,577 probes on the HM450 BeadChip in all 11 mouse samples tested was very low compared to that of a human sample used in this study, JWL (Fig. 1E).

Taking only the 13,715 probes we have identified as uniquely mapping to the bisulfite converted mm9 genome revealed an intensity distribution for 10/11 mouse samples similar to the human sample JWL (Fig. 1F). One sample with a very low intensity distribution, 87_lymph, was indicative of a failed Infinium analysis and this was confirmed with the on-board bisulfite conversion control probes on the HM450 BeadChip (Supplementary Fig. 3).

The performance of these unique probes could also be assessed by plotting the distribution of Beta scores called for all 485,577 probes. This revealed a distinct distribution curve for all mouse samples analysed when compared to the human sample, JWL (Supplementary Fig. 1B). With the distribution peaking at a Beta score of 0.3 such a curve in human sample analysis is indicative of a failed analysis. However, plotting the beta distribution of the unique subset of 13,715 probes, the distribution curves from the mouse samples were a similar shape to the human sample analysed in the same experiment (Fig. 2A).

Unsupervised hierarchal clustering of samples using this subset of probes suggests that distinct DNA methylation profiles can be measured using the HM450 array on mouse tissue samples (Fig. 2B). We found a similar dendrogram structure taking gene expression data from the Mouse Gene Atlas V3 (http://biogps.org, Table 4) of the 11,585 genes associated with this unique subset of Infinium HM450 probes (Supplementary Fig. 1C).

To assess the reproducibility of the HM450 platform for analysing mouse genomic DNA, we performed several replicate analyses. We assessed array replication, bisulfite treatment replication, and biological replication. High correlation was observed across all replicate analyses performed. A Pearson's correlation coefficient of 0.988 was observed for array replication (Fig. 2C) where two HM450 BeadChips were used to analyse the same mouse bone marrow sample (88_BM_01). The same correlation was observed performing parallel bisulfite conversions on genomic DNA derived from the same mouse tissue sample (88_BM_01_1 and 88_BM_02 Fig. 2D). We assessed biological replication by comparing DNA methylation from bone marrow tissue of two mice and found a very high correlation coefficient of 0.989 (Supplementary Fig. 1D).

Table 2
Infinium versus SEQUENOM DNA methylation summary.

Gene	Infinium probe ID	Difference hg19 and mm9	SEQUENOM CpG analysed	Infinium HM27			Infinium HM450			SEQUENOM		
				1.3HTT	2.4HTT	Difference	1.3HTT	2.4HTT	Difference	1.3HTT	2.4HTT	Difference
HOXC12	cg19928450	2MM	CpG 6	0.13	0.62	0.49	0.12	0.7	0.58	0.02	0.68	0.66
ZNF312 (FEZF2)	cg19629292	PM	CpG 6	0.06	0.73	0.67	0.08	0.56	0.48	0.06	0.49	0.43
DLX5	cg00096922	PM	CpG 18	0.07	0.55	0.48	0.09	0.51	0.43	0	0.45	0.45
HOXA5	cg02248486	PM	CpG 1	0.93	0.03	0.9	0.81	0.39	0.42	0.86	0.08	0.78
ARC	cg24981018	2MM	CpG 12	0.53	0.07	0.46	0.53	0.14	0.39	0.69	0.48	0.21
PAQR6	cg10046892	2MM	CpG 3	0.25	0.73	0.48	0.53	0.9	0.37	0.84	0.94	0.1
HR	cg12748258	2MM	CpG 2.3.4.5	0.07	0.81	0.74	0.56	0.35	0.21	0.71	0.04	0.67
GDF10	cg07773116	2MM	CpG 4	0.83	0.09	0.74	0.56	0.35	0.21	0.71	0.04	0.67
PRKCDBP	cg18392783	7MM	CpG 2.3.4	0.1	0.7	0.6	0.19	0.38	0.19	0.1	0.27	0.17
JARID2	cg01769037	1MM	CpG 3	0.71	0.23	0.48	0.09	0.16	0.07	0.08	0.06	0.02
WIPI2	cg20592700	2MM	CpG 32.33.34	0.02	0.02	0	0.25	0.2	0.05	0.03	0.05	0.02
KCNC3	cg06572160	2MM	CpG 17	0.03	0.03	0	0.05	0.1	0.05	0	0.01	0.01
PEX5	cg15754084	2MM	CpG 2	0.22	0.22	0	0.82	0.78	0.05	0.08	0.01	0.07
RAB20	cg20371891	3MM	CpG 14.15	0.03	0.03	0	0.18	0.17	0.02	0.01	0.04	0.03
CREBL2	cg00261552	7MM	CpG 12	0.04	0.04	0	0.07	0.07	0	0.03	0.06	0.03
GNA12	cg21685565	1MM	CpG 22.23	0.02	0.02	0	0.01	0.02	0	0.09	0.07	0.02

PM = perfect match, 1MM = 1 mismatch, 2MM = 2 mismatches, 3MM = 3 mismatches, 7MM = 7 mismatches.

Table 3
SEQUENOM primers used in this study.

Infinium ID	Chr	Start	Symbol	Forward primer	Reverse primer
cg02248486	chr6	52,154,271	HOXA5	aggaagagagGTAAATTTTGTTTGATGATTTTTAGAG	cagtaatacgactcactatagggagaaggctTAATTTTACCATAATAAACTATAACCTCAA
cg07773116	chr14	34,737,279	GDF10	aggaagagagTGTAGGGGGATAGGGATTTTTAGTA	cagtaatacgactcactatagggagaaggctCAAAAAATCTACCACCAAATACCAC
cg12748258	chr14	70,954,710	HR	aggaagagagTATAGTTTTGGTTTAGAAGTTGGTG	cagtaatacgactcactatagggagaaggctTTAAAATCAACTCATAAACTCCCCC
cg19629292	chr14	13,178,127	ZNF312	aggaagagagGGGAATTGGGTATTGAAAGAGTAAA	cagtaatacgactcactatagggagaaggctATCATTCCACTTCAAATAACTTCATA
cg19928450	chr15	102,767,199	HOXC12	aggaagagagTTTTTATTTAGGATTGGGGAAGG	cagtaatacgactcactatagggagaaggctCACAAACCCAAAATTCAAAAAATTA
cg00096922	chr6	6,832,079	DLX5	aggaagagagTTGGTTTATGGAGGGAAAGATGTAT	cagtaatacgactcactatagggagaaggctCCATTATCTACCAAAAAACTCTAACC
cg10046892	chr3	88,167,813	PAQR6	aggaagagagGTAGGGTTTGGAGAGTTTTGTTGTA	cagtaatacgactcactatagggagaaggctCCAACTTAAACCCCTAACCTAATCT
cg01769037	chr13	44,827,083	JARID2	aggaagagagTTGATTGTAAAAGGGGATAATTGTT	cagtaatacgactcactatagggagaaggctACAAATCAAATACTAAAAACCTTCAC
cg24981018	chr15	74,502,498	ARC	aggaagagagTGGTAGTAGGGTTTTGGATTTTTTA	cagtaatacgactcactatagggagaaggctCCAAACCCAATATAATCCTACAAATTA
cg20371891	chr8	11,478,337	RAB20	aggaagagagGTAAGGTAAGTGGGAATAGAGGTGT	cagtaatacgactcactatagggagaaggctTCTCAAACCAAAACCACTAAAATTC
cg06572160	chr7	51,846,699	KCNC3	aggaagagagGTATTTTGTTAGGGATTAGATTGGT	cagtaatacgactcactatagggagaaggctTAAATCATCCAACAACAAACCTCC
cg15754084	chr6	124,364,404	PEX5	aggaagagagAGGGGTTAGGTTTTAGTTTTTTTTG	cagtaatacgactcactatagggagaaggctAAAACACAACCTCCTCTTAAACTCC
cg20592700	chr5	143,105,745	WIPI2	aggaagagagTTTTGGAGTATAAATAAGAGTGGGGA	cagtaatacgactcactatagggagaaggctCTCAACCTAACCAACCCTAATCTCT
cg21685565	chr5	141,305,942	GNA12	aggaagagagTTTTTGAGAGTGGTTTAGTTTGTATTTT	cagtaatacgactcactatagggagaaggctACAAATCCACCTTCCTCAAACAAAT
cg00261552	chr6	134,780,245	CREBL2	aggaagagagGAGGGGGTTAATTAGTTTGTAGAGT	cagtaatacgactcactatagggagaaggctTCAAAACTTACCTTACTATCATCCA
cg18392783	chr7	112,630,274	PRKCDBP	aggaagagagGAGAGGGATGTTTGAGGTTATTGAT	cagtaatacgactcactatagggagaaggctCCTTACTAAAAAAATTAACCACCAT

Table 4
Publically available mouse expression datasets used in our study.

GEO identifier	Tissue type	Sample ID
GSM258627	Bone marrow	4MJW06120841
GSM258628	Bone marrow	4MJW06120842
GSM258669	Heart	4MJW061208131
GSM258670	Heart	4MJW061208132
GSM258691	Lymph nodes	4MJW06120839
GSM258692	Lymph nodes	4MJW06120840
GSM258767	Spleen	4MJW061208161
GSM258768	Spleen	4MJW061208162

2.5. Annotation and conservation between mm9 and hg19

Before embarking on this study we originally surmised that highly conserved genomic regions exist and the HM450 BeadChip is agnostic to which organism the genomic DNA analysed originated from. Our in-silico analysis identified 13,715 probes that uniquely mapped to the mm9 reference genome. The proportion of valid Infinium Type I and Infinium Type II probes for mouse interrogation is comparable to the proportion found on the entire HM450 BeadChip (31% and 69% respectively (Fig. 2F i)). The total proportion of Type I and Type II probes on the HM450 BeadChip is 28% and 82% respectively. To determine the extent of sequence conservation, the co-ordinates of the uniquely mapped probes were used to create an annotation table to the mm9 reference genome using the *annotatePeaks.pl* tool from Homer (http://biowhat.ucsd.edu/homer/ngs/index.html) [13], a ChIP-Seq peak finding and annotation tool. We compared the gene assigned to the probe using mm9 annotations with the gene assigned to the probe annotated to the hg19 reference genome. By comparing the gene symbol annotation and the RefSeq ID annotations, if a match is found between the mouse and human annotations of these probes, we would accept that the sequence is conserved between these two species. Of the 13,715 probes annotated, 75% (10,331 probes) were conserved between mouse and human representing 4542 conserved genes (Supplementary Table 11) the HM450 BeadChip can interrogate on mouse samples.

2.6. Using the HM450 BeadChip to investigate mouse biology

Of the 13,715 probes we have taken to measure DNA methylation in mouse samples, we calculated the standard deviation of Beta scores for each probe across the 11 samples we have analysed using the HM450 BeadChip. Setting a variation cut-off of 0.2 or greater standard deviations, 237 probes demonstrated high variation of Beta score across the samples we analysed (Fig. 2E). We performed DAVID [14] and GOrilla [15] ontology analysis on the genes associated with these probes. Ontology analysis revealed strong enrichment in developmental processes and cell signalling (Supplementary Table 9) indicative of tissue specific DNA methylation changes in the mouse tissue samples analysed.

We also investigated the extent of DNA methylation changes in an inducible *Hoxb8* mouse cell line model of acute myeloid leukaemia. This model cell line contains an inducible *Hoxb8* expression system under the control of tamoxifen.

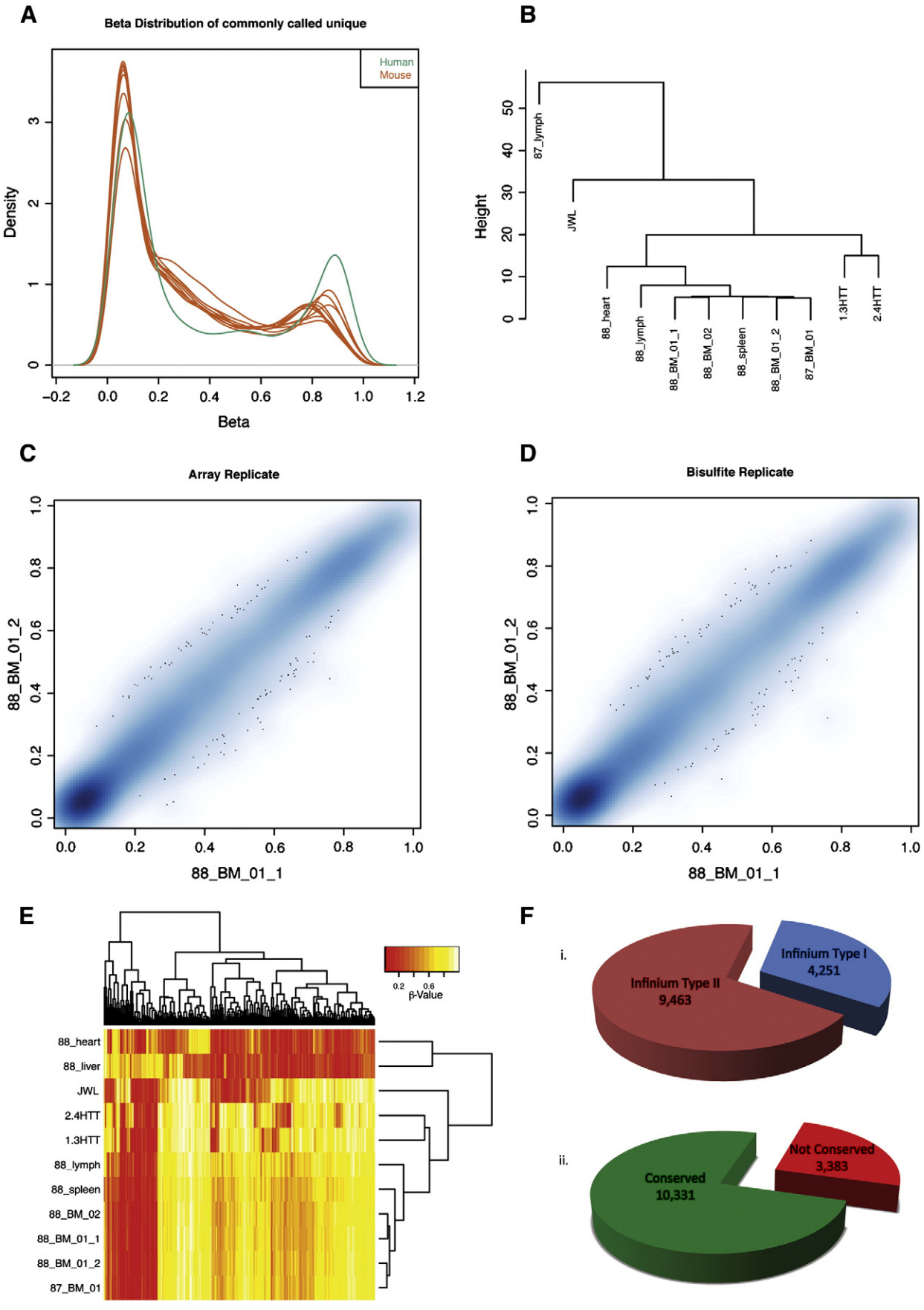
Continuous expression of *Hoxb8* is absolutely required to maintain the leukemic phenotype of this cell line and we can investigate changes in DNA methylation as a result of *Hoxb8* expression [16].

We measured DNA methylation before (1.4HTT) and after Hoxb8 (2.4HTT) induction to identify changes in DNA methylation as a consequence of *Hoxb8* expression using the HM450 BeadChip. We identified 738 probes having greater than 0.2 standard deviations of variation before and after *Hoxb8* induction. These probes were associated with 568 genes and after DAVID ontology analysis, found to be enriched with developmental, differentiation and homeobox processes known to be affected by the induction of *Hoxb8* [16] (Supplementary Table 10).

2.7. SEQUENOM EpiTYPER validation of HM450 mouse DNA methylation calls

We selected 16 Infinium probes demonstrating a range of Beta score differences before and after *Hoxb8* induction in the model mouse cell line for SEQUENOM EpiTYPER replication. The probes selected for validation had Beta score differences as high as 0.58 and contained up to 7 mismatches to the reference bisulfite mouse genome (Table 2). An overall correlation between SEQUENOM and HM450 methylation measurements of 0.48 was observed (Supplementary Fig. 2A). We observed good correlation in DNA methylation between SEQUENOM and HM450 selected with a large or no beta difference (Supplementary Fig. 2A). Poor correlation between SEQUENOM and HM450 was observed for *HoxA5* and *GDF10* which all demonstrated intermediate beta differences by HM450 BeadChip analysis (Supplementary Fig. 2B). From our results, it would seem

Fig. 2. A. The Beta score distribution of 13,715 uniquely mapping probes in bisulfite space of mm9 found on the HM450 BeadArray. The shape of the distribution curve from all the mouse samples analysed, was comparable to the human sample, JWL. B. Dendrogram depicting the level of similarity in DNA methylation profiles across the samples analysed in this study. The tree layout is comparable to a tree derived from the publicly available Mouse Gene Atlas dataset (http://biogps.org). C. Scatter plot depicting highly reproducible results from array replication (same sample, two distinct arrays). D. Scatter plot depicting highly reproducible results from bisulfite conversion replication. E. Heatmap of most variable DNA methylation probes in mouse samples. F. i. Distribution of Infinium Type I and Infinium Type II probes that could be used to interrogate mouse genomic DNA was comparable to the distribution of all probes on the HM450 BeadChip. F. ii. The proportion of highly conserved probes was also identified.


A
Beta Distribution of commonly called unique
Human
Mouse
Density
Beta
B
Height
87_lymph
JWL
88_heart
88_lymph
88_BM_01_1
88_BM_02
88_spleen
88_BM_01_2
87_BM_01
1.3HTT
2.4HTT
C
Array Replicate
88_BM_01_2
88_BM_01_1
D
Bisulfite Replicate
88_BM_01_2
88_BM_01_1
E
β-Value
88_heart
88_liver
JWL
2.4HTT
1.3HTT
88_lymph
88_spleen
88_BM_02
88_BM_01_1
88_BM_01_2
87_BM_01
F
i.
Infinium Type II
9,463
Infinium Type I
4,251
ii.
Conserved
10,331
Not Conserved
3,383

the Beta score measurements of HM450 probes containing up to 3 mismatches are tolerable (Table 2).

3. Discussion

With the increasing popularity of genome-wide DNA methylation BeadChips in human studies, interest in an equivalent mouse platform has been mounting. In its absence, we investigated the utility of Illumina Infinium Human Methylation BeadChips for the analysis of DNA methylation in mouse samples. Recent xeno-array studies have investigated genome wide gene expression and found that it was possible to estimate within-species differences using a microarray from a closely related species [4,17]. It has also been demonstrated that sequences with greater than 75% similarity to a 50-mer target probe are able to cross hybridize giving rise to unreliable intensity measurements [18]. Unlike gene expression microarrays that solely rely on hybridization, the chemistry used by this Infinium BeadChip requires both probe hybridization and base extension to fluorescently label the probe for a positive read out [6]. Nevertheless, to minimize the possibility of cross hybridization, we decided to investigate Infinium probes that mapped uniquely to the mouse reference genome using a panel of short read aligners.

Both the reference genome sequence (base space) and bisulfite converted reference genome sequence (bisulfite space) were used as reference genomes from hg19 and mm9 for in-silico alignment. We found that Blat, Novoalign, BWA and SHRiMP identified similar uniquely mapping probes to the mouse genome even though their underlying alignment algorithms are different (Figs. 1A and B).

The Illumina Infinium Human Methylation Bead Chip interrogates genomic DNA in bisulfite genome space. Therefore it is more appropriate to perform alignment analysis taking into account bisulfite conversion of genomic DNA. Unlike the genome space short read aligners above, bisulfite short read aligners have been developed more recently. It is unclear which bisulfite aligner performs the best. Here we tested four aligners that account for bisulfite conversion, Bsmap [11], Bismark [12], Brat [10] and Novoalign in bisulfite mode (http://www.novocraft.com). These aligners take into account nucleotide degeneracy afforded by bisulfite treatment, unmethylated cytosine converts to thymine after bisulfite conversion and PCR amplification while methylated cytosines remain as cytosine. Furthermore, the Watson and Crick strands of genomic DNA become non-complimentary after bisulfite conversion effectively doubling the size of the genome and increasing the chances of alignment of a 50-mer probe. Bsmap, Bismark, Brat and Novoalign in bisulfite mode handle this in their own way so we set out to identify probes that mapped and were in common with the aligners used. We found early in our analysis that Brat seemed overconservative and it was unclear how mapping was scored by this aligner. We therefore used Bsmap, Bismark and Novoalign in bisulfite mode for analysis.

After mapping to the bisulfite converted reference genome, we found 1308 and 13,715 probes uniquely mapped to bisulfite space from the HM27 and HM450 BeadChips respectively. The performance of these probes was assessed and we found Beta score measurements from them to be reproducible with good correlation between technical and biological replicates and, reliable with good correlation of DNA methylation measured using an independent assay, SEQUENOM EpiTYPER.

These probes could be used to identify tissue specific DNA methylation changes in mouse tissues and offer a potentially robust and viable option of array-based DNA methylation screening in the absence of a purpose built microarray for measuring DNA methylation in mouse samples. Using a model mouse cell line of acute myeloid leukaemia [16] we have been able to detect large changes of DNA methylation as a result of *Hoxb8* activation. The DNA methylation differences of these probes are associated with genes important in cell differentiation and homeobox processes confirming that the HM450 platform can measure biologically relevant changes in mouse samples.

Finally, in-silico alignment and annotation of the Infinium HM27 and HM450 probes to the reference mouse genome (mm9) revealed a significant proportion of usable probes that are highly conserved between human and mouse. Of note, performing in-silico alignment using Blat to hg19 we found 627 HM27 probes and 12,764 HM450 probes that appeared to map to multiple locations within our reference genome (Table 1). Our reference hg19 genome included haplotype reference sequences and the majority of these probes mapped to the haplotype reference sequences. Nevertheless these probes should be noted in future epigenome-wide association studies (EWAS) where the genetic effect of haplotype could confound epigenetic effects.

When using short read aligners that take into account bisulfite conversion of the genome, we found that up to 3% of probes on both the HM27 (729 and 778 probes from Novoalign Bisulfite Mode and Bismark respectively) and HM450 (15,346 and 16,781 probes from Novoalign Bisulfite Model and Bismark respectively) BeadChips mapped to haplotype sequences (derived from Table 1). This can been seen from Bismark alignments where we found 468,983 uniquely aligning probes to the hg19 reference genome with the haplotype sequences that increased to 481,260 unique probes with the haplotype sequences removed from our reference sequence (Table 1). It has been highlighted previously that sex-specific methylation differences observed on the Infinium HM27 array arose from probe sequence cross-reactivity [19]. We speculate that haplotype in a population-based study of DNA methylation by EWAS could confound methylation observations from probes that map to the reference haplotype sequences. We have identified probes that map to the haplotype reference sequences from the current Infinium HM27 and HM450 BeadChips and list them in Supplementary Tables 7 and 8.

In the absence of an Infinium Methylation Mouse BeadChip, we present the utility of the Illumina Infinium HM27 and HM450 BeadChip for analysing mouse samples for DNA methylation. We demonstrate that a subset of probes found on this array is able to interrogate the mouse genome for DNA methylation. Through our analysis, we have also identified potentially unreliable probes on both 27k and 450k platforms that map to multiple locations across the reference human genome (hg19) that should also be taken into account along with sex-specific probes [19] and single nucleotide polymorphisms [20]. From our study, we suggest that the currently available HM450 BeadChip could be a viable and affordable option to gain biological insight into the role of DNA methylation in model mouse systems .

4. Methods

4.1. In-silico analysis

Probe target sequences interrogated by the HM27 and HM450 array were extracted from the annotation library IlluminaHumanMethylation27k.db and IlluminaHumanMethylation450k.db respectively in Bioconductor (http://www.bioconductor.org). Probe target sequences were aligned using a range of short read aligners against the mouse (mm9) or human (hg19) reference genomes. For the human reference genome we constructed two reference genomes, one that contained all known reference haplotype sequences and another without these reference haplotype sequences (nohap). Genome indices required for each aligner were constructed as outlined in the respective user manuals. Aligners included, BWA 0.5.9 [21], SHRiMP 2.1.1 [22], Novoalign 2.05.32 (http://www.novocraft.com) and Blat v34 [23]. Default parameters were used for all aligners.

Using Samtools, probes were filtered for alignment, no mismatches (using SAM tags including H0:i:1 (number of perfect hits), NM:i:0 (edit distance from reference) and NH:i:1 (number of reported

alignments) depending on the output of the aligner used) and in Genome Space, 50 base alignments. Non-unique alignments were filtered using the uniq command. Specific tags generated by BWA after alignment including X0:i:1 (number of best hits) and XT:A:U (unique alignment) were used to extract unique alignments. To obtain perfect alignments from Blat, a minimum identity of 100% was set. The script files and resultant alignment files can be found in Supplementary data.

Novoalign had the added feature of alignment in bisulfite sequence space that aligns bisulfite converted short reads to both the Watson and Crick strands after bisulfite conversion. Infinium probe sequences were mock-converted using the find and replace function of the Unix command sed. Additional aligners that take into account bisulfite conversion of the genome were also used. These were Bsmap [11], Bismark [12], and Brat [10] with default parameters.

All alignments were performed at the Peak Computing Facility, Victorian Life Sciences Computation Initiative (http://www.vlsci.org.au). Scripts and result files are supplied as Supplementary information.

4.2. Mouse cell lines, tissues and growth conditions

Cell lines were generated as previously described [24]. Briefly, haematopoietic progenitor cells were harvested from E14.5 C57BL/6 embryos and infected with lentivirus encoding murine Hoxb8 under the control of a 4-hydroxy tamoxifen (4-OHT) inducible promoter. Cells were cultured in 0.25 ng/ml interleukin-3 (IL-3) and 0.1 μM 4-OHT. Clones were then selected after replating in semi-solid culture media (0.3% soft agar, 20% fetal calf serum (FCS), 0.3 ng/ml IL-3, 0.1 μM 4-OHT in DMEM). All clones were tested to ensure that viability and proliferation remained strictly dependent on IL-3 and Hoxb8. Cells are maintained in DMEM with 10% FCS, IL-3 and 4-OHT. Cells no longer expressing Hoxb8 were prepared by removing 4-OHT from culture for a period of four days. Mouse tissue were harvested from two C57BL/6 mice and stored in DirectPCR (Tail) lysis reagent (Viagen) at 4 °C. Tissue was macerated prior to lysis, proteinase K digestion and phenol chloroform extraction of genomic DNA. After spectrophometry, genomic DNA was quantified then processed as outlined below.

4.3. Infinium HM27 and HM450 BeadChip analysis

Genomic DNA extracted from a model Hoxb8 transformed myeloid cells and from various tissues from C57BL/6 mice. Hoxb8 is under the control of a 4-hydroxy-tamoxifen inducible promoter [16]. We compared the DNA methylation profile from cells before (−HOXB8) and after (+HOXB8) Hoxb8 induction using the Illumina HumanMethylation27 and HumanMethylation450 BeadChips [6]. DNA was bisulfite converted with MethylEasy Xceed Kit (Genetic Signatures, Sydney, Australia) according to manufacturer's instructions. Samples were then processed and analysed by the Australian Genomics Research Facility (AGRF, Melbourne, Australia) according to manufacturer's instructions (Illumina, San Diego). Raw data after scanning and hybridization was processed using lumi, a Bioconductor library within the R environment [25]. Raw data was initially used to determine intensity distributions. Infinium methylation probes were ranked by beta value difference and selected for SEQUENOM verification. Raw data has been submitted to GEO (GSE43226).

4.4. SEQUENOM EpiTYPER DNA methylation analysis

Bisulfite primers were designed to amplify the region of the mm9 genome that flanked the Infinium probe sequences of interest. The mm9 genomic location of the Infinium probes of interest were searched using Blat [23]. 500 bp flanking the probe sequence was extracted for SEQUENOM EpiTYPER primer design using EpiDesigner (http://www.epidesigner.com). The primer sequences used are listed in Table 3. SEQUENOM EpiTYPER chemistry was performed as outlined in [26].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.04.014.

Acknowledgments

The SEQUENOM Platform Facility at Murdoch Childrens Research Institute is managed and maintained by Benjamin Ong, Kirby Simmering, Stephen Wilcox and Melinda Ziino from the Australia Genome Research Facility.

This research was supported by a Victorian Life Sciences Computation Initiative (VLSCI) grant number [VR0002] on the Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government, Australia.

The Leukaemia Foundation of Australia supported this work through a Philip Desbrow Postdoctoral Fellowship (NW) and a Grant-in-Aid (NW, RS, and JC). NW and JC are currently supported by an NHMRC Project Grant and a Victorian Cancer Agency Grant. NW is also supported by the My Room and the Children's Cancer Centre Foundation. RS is supported by an NHMRC Senior Research Fellowship. MCRI is supported by the Victorian Government's Operational Infrastructure Support Program.

References

[1] P.W. Laird, Principles and challenges of genome-wide DNA methylation analysis, Nat. Rev. Genet. 11 (2010) 191–203.
[2] M. Bibikova, B. Barnes, C. Tsan, V. Ho, B. Klotzle, J.M. Le, D. Delano, L. Zhang, G.P. Schroth, K.L. Gunderson, J.-B. Fan, R. Shen, High density DNA methylation array with single CpG site resolution, Genomics (2011) 1–32.
[3] C. Bar-Or, H. Czosnek, H. Koltai, Cross-species microarray hybridizations: a developing tool for studying species diversity, Trends Genet. 23 (2007) 200–207.
[4] A. Oshlack, A.E. Chabot, G.K. Smyth, Y. Gilad, Using DNA microarrays to study gene expression in closely related species, Bioinformatics 23 (2007) 1235–1242.
[5] L. Lin, S. Liu, H. Brockway, J. Seok, P. Jiang, W.H. Wong, Y. Xing, Using high-density exon arrays to profile gene expression in closely related species, Nucleic Acids Res. 37 (2009) e90.
[6] M. Bibikova, J. Le, B. Barnes, S. Saedinia-Melnyk, L. Zhou, R. Shen, K.L. Gunderson, Genome-wide DNA methylation profiling using Infinium® assay, Epigenomics 1 (2009) 177–200.
[7] S. Dedeurwaerder, M. Defrance, E. Calonne, H. Denis, C. Sotiriou, F. Fuks, Evaluation of the Infinium Methylation 450K technology, Epigenomics 3 (2011) 771–784.
[8] J. Maksimovic, L. Gordon, A. Oshlack, SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips, Genome Biol. 13 (2012) R44.
[9] S.J. Clark, A. Statham, C. Stirzaker, P.L. Molloy, M. Frommer, DNA methylation: bisulphite modification and analysis, Nat. Protoc. 1 (2006) 2353–2364.
[10] E.Y. Harris, N. Ponts, A. Levchuk, K.L. Roch, S. Lonardi, BRAT: bisulfite-treated reads analysis tool, Bioinformatics 26 (2010) 572–573.
[11] Y. Xi, W. Li, BSMAP: whole genome bisulfite sequence MAPping program, BMC Bioinforma. 10 (2009) 232.
[12] F. Krueger, S.R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications, Bioinformatics 27 (2011) 1571–1572.
[13] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, C.K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Mol. Cell 38 (2010) 576–589.
[14] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2008) 44–57.
[15] E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists, BMC Bioinforma. 10 (2009) 48.
[16] P.G. Ekert, Apaf-1 and caspase-9 accelerate apoptosis, but do not determine whether factor-deprived or drug-treated cells die, J. Cell Biol. 165 (2004) 835–842.
[17] Y. Lu, P. Huggins, Z. Bar-Joseph, Cross species analysis of microarray expression data, Bioinformatics 25 (2009) 1476–1483.
[18] M. Kane, T. Jatkoe, C. Stumpf, J. Lu, Assessment of the sensitivity and specificity of oligonucleotide (50 mer) microarrays, Nucleic Acids Res. 28 (22) (2000) 4552–4557.
[19] Y.-a. Chen, S. Choufani, J.C. Ferreira, D. Grafodatskaya, D.T. Butcher, R. Weksberg, Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray, Genomics 97 (2011) 214–222.
[20] D. Wang, L. Yan, Q. Hu, L.E. Sucheston, M.J. Higgins, C.B. Ambrosone, C.S. Johnson, D.J. Smiraglia, S. Liu, IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data, Bioinformatics 28 (2012) 729–730.
[21] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics 25 (2009) 1754–1760.
[22] M. David, M. Dzamba, D. Lister, L. Ilie, M. Brudno, SHRiMP2: sensitive yet practical short read mapping, Bioinformatics 27 (2011) 1011–1012.

[23] W.J. Kent, BLAT—the BLAST-like alignment tool, Genome Res. 12 (2002) 656–664.
[24] A.M. Jabbour, C.P. Daunt, B.D. Green, S. Vogel, L. Gordon, R.S. Lee, N. Silke, R.B. Pearson, C.J. Vandenberg, P.N. Kelly, S.L. Nutt, A. Strasser, C. Borner, P.G. Ekert, Myeloid progenitor cells lacking p53 exhibit delayed up-regulation of puma and prolonged survival after cytokine deprivation, Blood 115 (2010) 344–352.
[25] P. Du, W.A. Kibbe, S.M. Lin, Lumi: a pipeline for processing illumina microarray, Bioinformatics 24 (2008) 1547–1548.
[26] N.C. Wong, D. Ashley, Z. Chatterton, M. Parkinson-Bates, H.-K. Ng, M.S. Halemba, A. Kowalczyk, J. Bedo, Q. Wang, K. Bell, E. Algar, J.M. Craig, R. Saffery, A distinct DNA methylation signature defines pediatric pre-B cell acute lymphoblastic leukemia, Epigenetics 7 (2012).