—Technology Report—

Targeted DNA Methylation Analysis by High Throughput Sequencing in Porcine Peri-attachment Embryos

**Benson H. MORRILL[1), Lindsay COX[1), Anika WARD[1), Sierra HEYWOOD[1),
Randall S. PRATHER[2) and S. Clay ISOM[1)**

[1)*Animal Dairy & Veterinary Sciences Department, Utah State University, Logan, UT 84322, USA*
[2)*Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA*

Abstract. The purpose of this experiment was to implement and evaluate the effectiveness of a next-generation sequencing-based method for DNA methylation analysis in porcine embryonic samples. Fourteen discrete genomic regions were amplified by PCR using bisulfite-converted genomic DNA derived from day 14 *in vivo*-derived (IVV) and parthenogenetic (PA) porcine embryos as template DNA. Resulting PCR products were subjected to high-throughput sequencing using the Illumina Genome Analyzer IIx platform. The average depth of sequencing coverage was 14,611 for IVV and 17,068 for PA. Quantitative analysis of the methylation profiles of both input samples for each genomic locus showed distinct differences in methylation profiles between IVV and PA samples for six of the target loci, and subtle differences in four loci. It was concluded that high throughput sequencing technologies can be effectively applied to provide a powerful, cost-effective approach to targeted DNA methylation analysis of embryonic and other reproductive tissues.
Key words: Bisulfite sequencing, DNA methylation, Illumina, Parthenogenetic embryo

In addition to 'traditional' genetic control, epigenetic phenomena are being recognized as playing important roles in directing cellular function and fate. Epigenetics can be defined as the means whereby heritable patterns of gene expression within a population of cells are regulated without alterations in the actual nucleotide sequence of the genome. Epigenetics, broadly defined, can include such processes as post-translational histone modification, chromatin remodeling, RNA-mediated gene silencing (RNAi), and DNA and mRNA methylation [reviewed in 1, 2]. Of these, DNA methylation is the best characterized, and the most extensively studied [see 1, 3 and references therein]. DNA methylation involves the covalent addition of a methyl group ($-CH_3$) at the C5 position of the cytosine nucleotide base. In mammals, this modification is observed primarily at cytosines in a CpG context – that is, when a cytosine is immediately followed by a guanine in the nucleotide sequence. In mammals, 70–80% of CpG dinucleotides are methylated. Regions of higher-than-expected CpG frequency (CpG islands) are associated with most mammalian genes, and the relative preponderance of methylated cytosines within these CpG islands is generally inversely correlated with transcriptional activity of those genes (i.e. high methylation = low gene expression). DNA methylation is carefully regulated within the cell, and the establishment and maintenance of DNA methylation marks are accomplished by a variety of methyltransferase enzymes (DNMT1, DNMT3A, DNMT3B, e.g.). Precisely coordinated methylation and de-methylation of DNA is essential for proper development [4].

Faulty patterns of DNA methylation are associated with myriad disease conditions, including many cancers [5–9].

In livestock animal production systems, application of some assisted reproductive technologies (ART) such as *in vitro* fertilization (IVF) and somatic cell nuclear transfer (SCNT) is limited because of relatively poor efficiency. In cattle, establishment of pregnancy after transferring embryos derived from IVF into recipient females can be as low as 20% [10, 11]; survival of SCNT embryos after transfer is even lower: 5–10% [reviewed in 12]. Similar – though even less encouraging – data are observed in pigs [13]. Evidence is mounting that embryo and/or gamete manipulations *in vitro* can adversely impact the delicate constitution of the embryonic epigenome, and that even subtle changes in DNA methylation profiles might provide a mechanistic link between *in vitro* culture techniques and poor developmental competence.

More is known about the role of DNA methylation in development and disease than other epigenetic phenomena, at least in part because robust and simple techniques have been developed that allow for accurate assessment of CpG methylation status, even at single base-pair resolution. The most widely-applied approach to DNA methylation analysis is a technique called bisulfite sequencing. When exposed to bisulfite ions under the proper conditions, methylated cytosines of chromosomal DNA remain intact, while unmethylated cytosines undergo de-amination – thus converting cytosine to uracil. In downstream sequencing analyses of bisulfite-converted DNA, uracil residues behave like thymine, instead of the cytosines they were originally derived from. Thus by comparing DNA sequences 'before' and 'after' bisulfite conversion, it is possible to ascertain which cytosines within a particular sequence were methylated and which were unmethylated at the time of analysis. Potential drawbacks to traditional bisulfite sequencing include the inconvenience of cloning of

bisulfite-converted products into *E. coli* for propagation, the relatively low statistical confidence derived from analyses that are typically based on data from only 10–20 sequences per sample, and costs that accumulate quickly as more samples and/or more genetic sequences are targeted for analysis. Alternatives to bisulfite sequencing for evaluating DNA methylation, include Combined Bisulfite Restriction Analysis (CoBRA), Methylated DNA Immunoprecipitation (MeDIP), and Reduced Representation Bisulfite Sequencing (RRBS), among others [reviewed in 14]. These alternative techniques each come with significant limitations, including (but not limited to) high cost and/ or low resolution of sequence-specific methylation data.

The vast majority of the published studies regarding DNA methylation analysis in mammalian samples utilize low-throughput, traditional bisulfite (Sanger) sequencing as outlined above. Recent advances in nucleic acid sequencing technologies (so-called 'next-generation', 'high throughput', or 'deep' sequencing) have made it possible to generate millions of base pairs worth of sequence information in a matter of hours. Until now, most reports of the utilization of high-throughput sequencing technologies for studying DNA methylation detail genome-wide epigenomic analyses [15, 16], which efforts can be extremely costly. In addition, because of the large size of mammalian genomes, the depth of coverage at each CpG site and the associated statistical confidence in the true proportion of methylated vs unmethylated cytosines in the initial cell population can be quite low. Herein, we report our efforts taking a 'hybrid' approach to DNA methylation analysis: using the Illumina GAIIx platform to generate millions of bisulfite sequencing reads from a discrete number of targeted loci, with each read containing quantitative information about the methylation status of our input samples.

As a small part of our larger effort to understand the mechanisms behind early embryonic mortality in swine, we set out to devise simple, straightforward, relatively inexpensive, but powerful techniques for targeted DNA methylation analysis in early embryonic samples. Day 14 embryos were chosen for this particular analysis because the physiology of the developing conceptus during the peri-attachment window (days 11–18 in pigs) is especially critical to the ultimate success or failure of individual embryos growing in utero. It was considered that differences in DNA methylation between porcine embryos derived from artificial insemination (*in vivo*; IVV) and parthenogenetic oocyte activation (PA) could yield insight into the epigenetic determinants of successful embryogenesis and the phenomenon of parent-of-origin genomic imprinting. Genomic imprinting is a poorly understood phenomenon – especially in non-traditional model species – that involves the preferential transcription of one parental allele over the other. Differential DNA methylation of the maternal and paternal alleles is at the heart of genomic imprinting [17, 18]. It has been suggested that imprinted genes are particularly susceptible to insult in response to the *in vitro* manipulations associated with ART [19–22].

From a previous, small-scale pilot study designed to identify epigenetic differences between day 14 IVV and PA embryos, CpG islands were identified that were putatively differentially methylated between embryo types (data not shown). We selected for further study 14 of these CpG islands that were located either within or proximal to characterized gene bodies. For ease of identification, these CpG islands are referred to by the name of their nearest gene neighbor.

Table 1. Summary of the PCR products used and the number of Illumina reads aligned for each CpG island studied

PCR product	Length (bp)	# CpG	IVV reads	PA reads
BCOR	346	20	40,245	57,940
BMP7	454	25	82,961	116,271
BSX	392	28	68,699	42,226
CDC42BPB	340	30	95,711	116,835
CDX2	387	33	84,901	118,953
EBF	476	40	54,772	89,935
FLT1	373	30	61,294	73,924
GCN5	282	21	67,707	124,687
HK2	316	25	85,146	105,566
LHX4	485	34	100,716	99,273
NDUFB11	289	19	49,610	33,268
RRM1	255	21	107,704	87,613
RUNX1T1	249	25	60,929	57,122
WDR27	445	31	105,155	108,185
Total:			1,065,550	1,231,798

PCR Product = The CpG islands queried by bisulfite PCR sequencing are referred to by the name of the gene that is in closest proximity. Length = the predicted length of the PCR product; bp = base pairs. # CpG = the number of potential methylation sites available for analysis within each PCR product. IVV = *in vivo*-produced embryo; PA = embryo derived from parthenogenetic oocyte activation. The values presented in the "IVV Reads" and the "PA Reads" columns are the numbers of short Illumina sequencing reads that were aligned to reference sequences for each CpG island and used for downstream analyses.

Sequence information and chromosomal coordinates for these CpG islands can be found in Suppl Table 1 (on-line only). A high throughput sequencing approach was taken to evaluate the methylation status of these 14 CpG islands in IVV and PA embryos. Genomic DNA from a single intact day 14 embryo from each production method (IVV and PA) was collected and subjected to bisulfite conversion, which provides a mark to differentiate unmethylated from methylated cytosine bases within the nucleotide sequences evaluated. A single PCR amplicon from each of fourteen distinct CpG islands was generated using bisulfite-converted DNA from both embryo types. The average amplicon length was 359 base pairs (bp; range 249–485 bp), covering an average of 27.1 CpG dinucleotides per amplicon (range 19–40 CpG sites). These bisulfite PCR products were subjected to high throughput sequencing using the Genome Analyzer IIx platform from Illumina, and the short (80 bp) reads produced by the instrument were mapped to the reference sequences for each individual amplicon.

A total of 2,297,348 sequencing reads were aligned, with a slight bias overall towards reads from the PA sample. Table 1 provides summary statistics regarding the number of reads aligned and used for downstream methylation analyses for each CpG island queried. In addition to the tendency for more reads from the PA sample, the number of aligned reads per amplicon varied greatly within a sample. Likewise, the distribution of reads across the amplicons was not uniform. Figure 1 shows representative traces for four PCR products, plotting depth of coverage (Y-axis) against nucleotide position (X-axis). Read distribution bias was so extreme for one amplicon (*NDUFB11*; not shown) that large gaps in coverage were
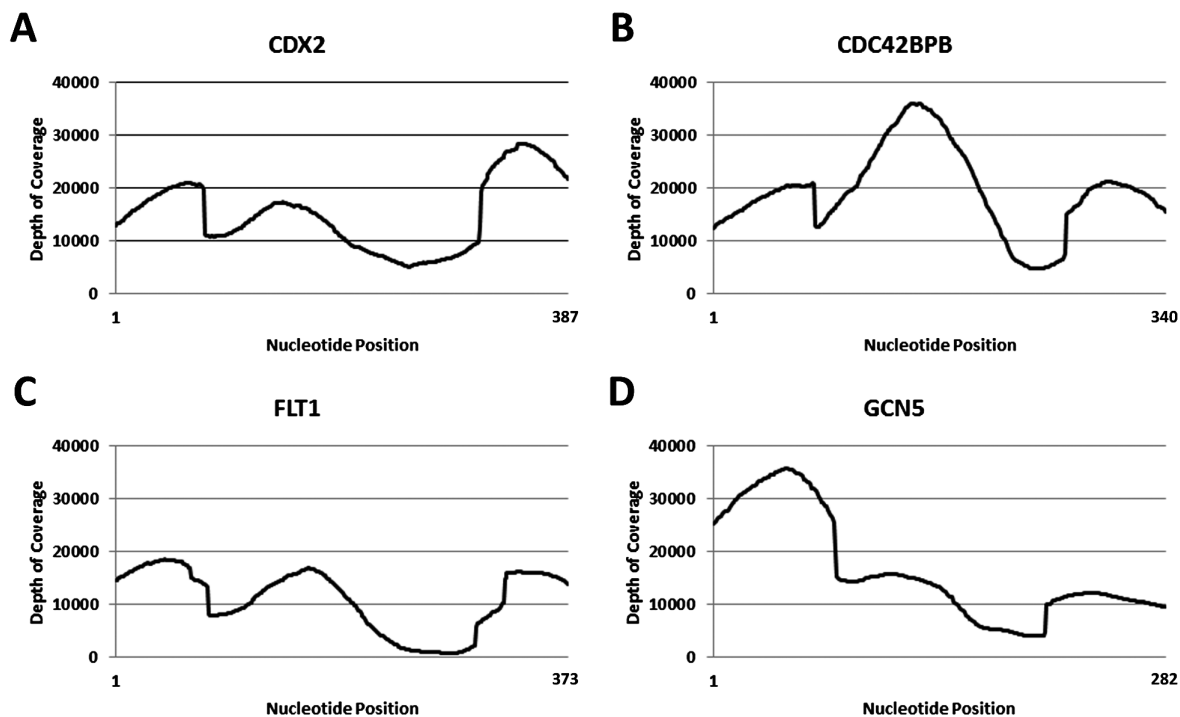
A CDX2

B CDC42BPB

C FLT1

D GCN5

Fig. 1. Depth of coverage obtained using high throughput bisulfite sequencing. Representative traces plotting absolute depth of coverage (Y-axis) against nucleotide position (X-axis) along the entire length of the amplicon for four of the 14 PCR products sequenced and aligned. These four traces are all from IVV samples; the traces from PA samples are virtually identical in shape and differ only slightly regarding the depth of coverage at each position.

observed, and this genomic locus was not evaluated further. While the absolute depth-of-coverage values differed between samples, the shapes of the read distribution traces for a given PCR product were virtually indistinguishable between samples. Notwithstanding these peculiarities associated with read generation and alignment, the average depth of coverage across all nucleotide positions in all amplicons was 14,611 for IVV and 17,068 for PA, with a maximum of 77,944 at nucleotide position 55 of the RRM1-IVV amplicon and a minimum of 718 at nucleotide position 279 of the FLT1-PA amplicon.

Using a sophisticated single nucleotide polymorphism identification algorithm, we were able to precisely determine the proportion of cytosines in CpG context that retained their identity after bisulfite sequencing (i.e. methylated in original template) relative to those that were deaminated by the conversion process (unmethylated in original sample). Data from the comparative methylation analysis are summarized in Fig. 2. We saw clear differences in overall methylation levels (more than 30% difference) between IVV and PA embryos in six of the thirteen amplicons analyzed, while overall percent methylation was not obvious in the remaining seven amplicons. It was interesting to note, however, that subtle differences (greater than 1%, but less than 5% difference) in overall methylation patterns were suggested between sample types for three of the amplicons (BMP7, 4.5% difference; LHX4, 2.9% difference; and RUNX1T1, 4.4% difference).

The physiological significance of these subtle differences is under investigation, and is more appropriately debated in another venue.

The prevailing dogma, though, is that major changes in the overall density of methyl-cytosine bases across large regions of chromosomal DNA alters the conformation of chromatin, which can impact gene expression, and cell physiology. By this definition, these subtle differences in DNA methylation between IVV and PA embryos would not be expected to have any physiological relevance. However, there is no clear threshold of DNA methylation that is required to elicit biological change. As little as 10% change in promoter methylation can cause phenotypic variation in the viable yellow agouti mouse [23], and constantly-emerging evidence suggests that even very minor changes in DNA methylation patterns might have distinct effects on cellular and organismal physiology [see references 24–26, e.g.]. Thus the importance of being able to confidently detect such subtle changes in DNA methylation patterns cannot be overstated.

Traditional bisulfite sequencing experiments typically report methylation statistics for 10–20 DNA strands for each target sequence and for each sample tested. Figure 3 depicts the relationship between the number of DNA strands queried and the resulting statistical power for different magnitudes of discrepancies in percent methylation between two samples. While a traditional bisulfite sequencing experiment utilizing 20 sequences per amplicon would have sufficient statistical power to confidently detect a 50% difference in methylation frequency between sample types (as might be expected in classic instances of genomic imprinting), more subtle differences (≤ 10%) between samples would not be reliably detected with even 100 sequences per amplicon (Fig. 3). Herein lies one major benefit of
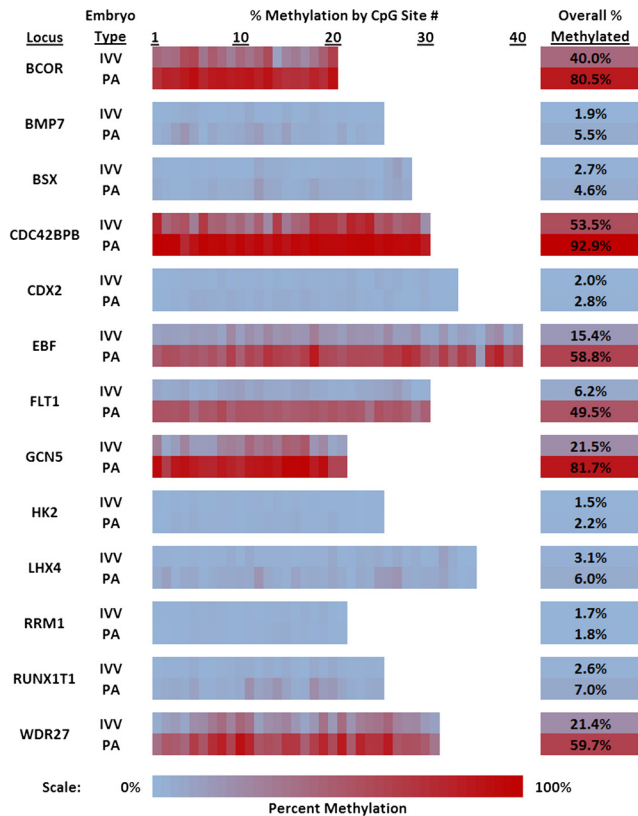
Fig. 2. Comparative analysis of DNA methylation patterns of *in vivo*-produced (IVV) and parthenogenetic (PA) porcine embryos. The percentages of methylated cytosines at each potential methylation site for each of thirteen distinct PCR products are depicted. Deep red colors indicate very high percent methylation; pale blue indicates very low methylation. Locus = gene nearest the CpG island queried.

Fig. 4. Manual methylation analysis of FLT1 sequencing reads from high throughput and traditional bisulfite sequencing techniques. A) The FLT1 locus queried in these experiments corresponds to a CpG island that maps to the 3' end (exon 29 and flanking intron sequences) of the FLT1 gene. The long horizontal line represents the full 864 base-pair CpG island. Each vertical line represents a potential methylation site within the island. The hashed bar marks the approximate location of the full PCR amplicon generated for these experiments, while the green shaded box shows the relative position of exon 29 of the FLT1 coding sequence. The CpG sites marked with red lines are those described in further detail in this figure (CpG dinucleotides A–H in Panels B and C below), which correspond to CpG dinucleotide numbers 15–22 (in reverse order) in Fig. 2. Primer sequences and precise chromosomal coordinates of this and other genomic sequences analyzed in these experiments can be found in Suppl Table S1 (on-line only). B) Short Illumina reads corresponding to the FLT1 locus (n = 54 reads for PA; n = 52 reads for IVV) were aligned to the reference sequence and the methylation status of each DNA fragment was evaluated. Each row in the figure represents a short sequencing read. Each column represents a potential methylation site. Methylated CpG sites are indicated by red boxes, whereas unmethylated sites are colored blue. For this specific segment of the FLT1 locus, 82.1% of the DNA fragments analyzed were either fully methylated or completely unmethylated, while only 17.9% of the clones were partially methylated. These data illustrate the utility of a manual analysis of short reads for evaluating the distribution pattern of methylated *vs.* unmethylated cytosines. C) Methylation analysis by traditional bisulfite sequencing of the same eight CpG dinucleotides within the FLT1 locus that were evaluated in Panel B.
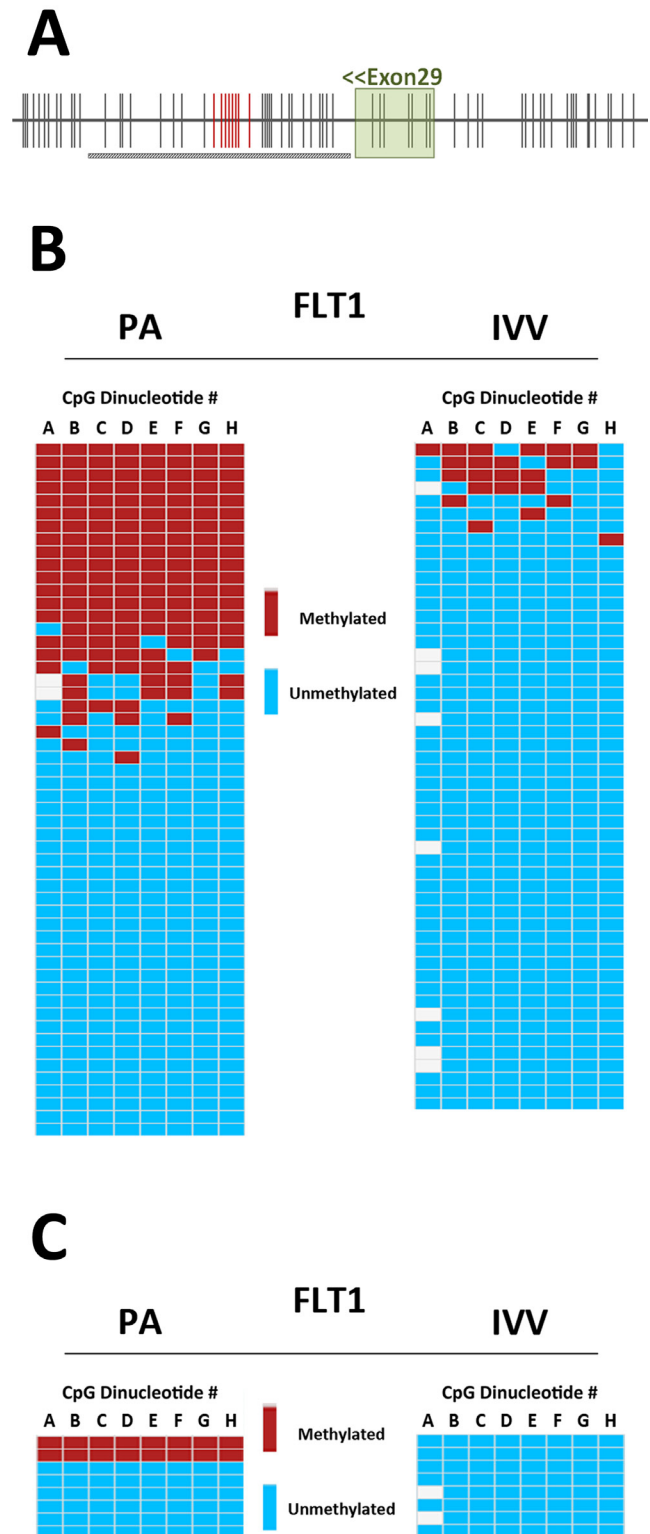
Statistical Power

Difference Between Samples	2	10	20	50	100	500	1000
90%	0.36	1	1	1	1	1	1
80%	0.36	1	1	1	1	1	1
70%	0.22	0.95	1	1	1	1	1
60%	0.18	0.84	0.99	1	1	1	1
50%	0.14	0.67	0.94	1	1	1	1
40%	0.11	0.50	0.81	1	1	1	1
30%	0.09	0.33	0.60	0.95	1	1	1
20%	0.07	0.19	0.35	0.71	0.95	1	1
10%	0.06	0.09	0.14	0.29	0.51	0.99	1
5%	N/A	N/A	0.06	0.12	0.19	0.67	1
1%	N/A	N/A	0.05	0.05	0.06	0.08	0.11

Number of Reads Sequenced

Fig. 3. Statistical power in bisulfite sequencing experiments. The statistical power that is available to detect significant differences in methylation profiles between two samples varies with the number of discrete observations made for each sample and the magnitude of the difference in percent methylation between the two samples. Estimates of statistical power for varying combinations of observations (reads sequenced) and percent methylation differences (difference between samples) are presented in this figure. A power statistic of 0.8 is generally regarded as a minimal value to achieve for appropriate statistical confidence [30]. In the figure above, values shaded in grey suggest that for those hypothetical experimental parameters, there would be insufficient statistical power to confidently reject the null hypothesis (that methylation levels are equivalent between samples). Thus for an observed 40% difference between samples (i.e. 90% methylated vs. 50% methylated), using 20 reads per sample, one might with reasonable confidence reject the hypothesis that the two samples had identical methylation patterns. But for an observed difference of 30%, the statistical power (< 0.8) would be insufficient to make that claim. Note that for relatively subtle differences in methylation ($< 10\%$), even 100 sequences per sample would not provide enough observations for statistically-significant separation of samples. N/A = the software used for power statistic calculations [29] imposes lower limits of samples sizes and proportions used together to ensure numerical stability; below these limits, meaningful statistics are not generated.

this high throughput bisulfite sequencing approach to targeted DNA methylation analysis: given the enormous depth of coverage at each potential methylation site, confidence is extremely high that any differences observed are truly reflective of the overall methylation status of the input material, and not an artifact of statistical sampling.

It should be noted that a comprehensive analysis of the short read sequencing and alignment data generated here provides only a summary of the average percent methylation at each locus for each sample. Thus, it is not immediately apparent whether a particular locus that demonstrates 50% overall methylation is approximately 50% methylated in all DNA fragments coming from that sample, or whether half the fragments are highly methylated (~100%) and the other half very lowly methylated (~0%). These data stand in contrast with those obtained from traditional bisulfite sequencing approaches where the complete methylation profile for individual cloned DNA fragments can be known. In some experimental schemes, being able to distinguish between these scenarios could be very important – as in the study of imprinted genes. Despite the high throughput nature of the approach described here, it is still very possible to evaluate

a small number of randomly selected reads such that the nature of the distribution of methylated cytosines can be assessed. A summary of one such evaluation is presented in Fig. 4, where it was noted that at the FLT1 locus, cytosine methylation was distributed such that the majority of the reads were either completely methylated or completely unmethylated. The utilization of longer sequencing reads (80–100 bp, as described herein) facilitates such analyses, by allowing the evaluation of as many potential methylation sites as possible for each read.

We performed traditional bisulfite sequencing to confirm that the results generated using the Illumina platform would be validated by a more time-tested approach. A summary of the work flow for each of the two sequencing approaches (traditional Sanger sequencing *vs.* high throughput Illumina sequencing) is presented in Fig. 5, and the results from the two methods are summarized in Fig. 6. The correlation coefficient (R) of the methylation statistics between the two techniques was 0.95, indicating very high agreement between the two analysis methods. We surmise that the small differences observed between methods arise because of sampling errors that arise due to the relatively small number of sequences analyzed using traditional bisulfite sequencing techniques.

Importantly, the costs associated with this high throughput bisulfite sequencing approach were not any higher than for a traditional bisulfite sequencing approach. For the current experiment, we sequenced fourteen amplicons from each of two samples at an average depth of coverage of 15,839 for a cost of approximately $1200 (US dollars). Sequencing these same fourteen amplicons from these same two samples at a depth of coverage of 20 using a traditional bisulfite sequencing approach would cost approximately $1800 (US dollars), not including the cost of PCR product subcloning, plasmid minipreps, etc. Depending on the numbers of sample/amplicon combinations queried, a next-generation sequencing approach can be significantly more cost-effective than a traditional approach. The sequencing depth that was attained in the current experiments provided many more observations at each potential methylation site than were needed to accurately ascertain its methylation status. This suggests that even more samples and/or more genomic loci could be queried in a similar experiment without compromising effectiveness or significantly increasing the cost.

In conclusion, the single end, short read Illumina sequencing platform can be effectively implemented to address specific questions regarding the epigenetic constitution of embryonic and other reproductive samples. Ultra-deep bisulfite sequencing can dramatically boost statistical confidence in targeted DNA methylation analyses as compared to traditional bisulfite sequencing. These techniques, as reported here, should be more widely considered as powerful and cost-effective means for evaluating patterns of DNA methylation in reproductive research.

Methods

Collection of embryonic tissue

Embryos derived from parthenogenetic activation of *in vitro* matured porcine oocytes were produced, transferred into recipient females, and harvested on day 14 of gestation as described elsewhere [27]. The IVV embryos were produced by artificial insemination
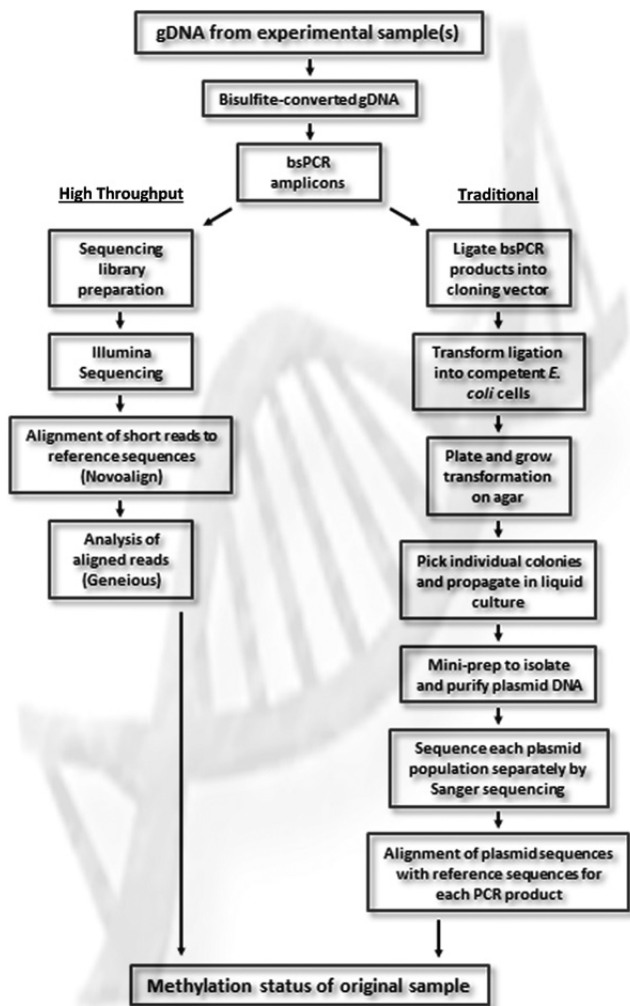
Fig. 5. Workflow summaries for traditional and high throughput bisulfite sequencing. Summarized protocols for traditional and high throughput DNA methylation analysis by bisulfite sequencing are presented. gDNA = genomic DNA; bsPCR = bisulfite PCR.
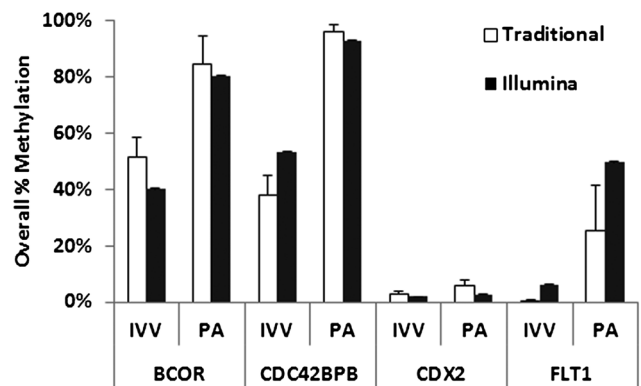

Fig. 6. DNA methylation analysis by traditional versus high throughput bisulfite sequencing. Percent methylation of four target sequences in *in vivo*-derived (IVV) and parthenogenetic (PA) embryos, as determined by traditional bisulfite sequencing (open bars) or Illumina-based deep sequencing (black bars). Genomic sequences and sample types are presented across the X-axis, while overall percent methylation for each sample/amplicon combination is presented on the Y-axis. The correlation coefficient (R) for percent methylation between analysis platforms for all sample/amplicon combinations is 0.95, suggesting a strong positive correlation between datapoints collected using the two distinct methods. Values presented are overall percent methylated (# methyl-C/# total CpG) ± standard error of the mean (SEM). The high throughput analysis approach utilized for the Illumina data does not lend itself to a calculation of a true error statistic based on all individual reads. Therefore, the error terms presented for these data points are estimates based on the standard deviation values calculated for ~300 randomly selected short reads for each sample/locus combination, and using the average depth of sequencing coverage for each amplicon as the value for "n" when calculating the SEM.

according to standard industry practices, and were harvested by uterine flushing on day 14 of gestation as well. Upon collection, embryos were snap frozen in liquid nitrogen and stored at –80 C until use.

Genomic DNA preparation and bisulfite PCR

Genomic DNA was isolated from the IVV and PA embryos using the AllPrep DNA/RNA Micro kit from Qiagen (Valencia, CA, USA), and was stored at –20 C until use. Genomic DNA (500 ng) from both embryo types was subjected to bisulfite conversion using the EZ DNA Methylation-Gold kit from Zymo (Irvine, CA, USA), according to the manufacturer's directions. The EpiDesigner software from Sequenom was used to design primers for bisulfite PCR (Suppl Table S1: on-line only). The HotStarTaq Plus Master Mix kit from Qiagen was utilized to generate PCR amplicons from bisulfite-converted genomic DNA. PCR reaction mixes included 20.0 μl 2× MasterMix, 10.0 μl water, 4.0 μl CoraLoad loading reagent,

2.0 μl bisulfite-converted gDNA (at 20 ng/μl), and 4.0 μl forward/reverse primer mix (final concentration of 1 μM each). Samples were then subjected to thermal cycling for target amplification. Cycling parameters included a 5 min initial enzyme activation cycle followed by 40 cycles of 95 C/15 sec – 55 C/30 sec – 72 C/30 sec and then a single 5 min extension termination step of 72 C. Resulting PCR products were purified using the QIAquick PCR Purification kit from Qiagen and pooled by sample type in equimolar concentrations, then sheared to approximately 100–150 bp fragments using the BioRuptor from Diagenode (Liege, Belgium). Sequencing library preparation proceeded according to standard single-end sequencing protocols provided by Illumina, and both final libraries (IVV and PA) were loaded onto a single lane of the Genome Analyzer IIx flow cell at a final combined concentration of 7 pM. Sequencing consisted of 2×40 bp cycles, resulting in a final read length of 80 bp.

Sequencing read alignment and methylation analysis

Sequencing reads were aligned to reference sequences using the bisulfite sequencing algorithm of the Novoalign software from Novocraft (Selangor, Malaysia). Aligned reads were uploaded into Geneious, a sequence manipulation software product from Biomatters (Auckland, New Zealand), and the SNP detection algorithm was used to determine the frequencies of methylated versus unmethylated cytosines in CpG context in the distinct samples for each PCR

product. To compare with the results of traditional sequencing (see below), the methylation status of the FLT1 locus was determined by selecting short reads that aligned perfectly to the target sequence (n = 54 reads for PA; n = 52 reads for IVV), and evaluating the methylation status of each read using BISMA, a web-based DNA methylation analysis platform [28].

Traditional E. coli-based bisulfite sequencing

Traditional bisulfite sequencing to validate the deep sequencing data was performed by subcloning four of the same PCR products used for Illumina sequencing (*BCOR, CDC42BPB, CDX2, FLT1*) in to the pGEM T-Easy vector from Promega (Madison, WI, USA), transforming subcloned plasmids into E. cloni 10G chemically-competent bacterial cells (Lucigen; Middleton, WI, USA), and sequencing the resulting plasmid preparations. A minimum of seven unique plasmid preparations were sequenced for each sample/amplicon combination. Methylation frequencies of subcloned PCR products were determined using BISMA as described above.

Power analysis calculations

The power analysis calculations were performed using software written and made freely available online by Dr. Russ V. Lenth at the University of Iowa [29]. Repeated power tests to compare two proportions were performed, changing sample sizes and expected differences between samples with each iteration. Alpha (α) at 0.05 and equal sample sizes per treatment group were set for each permutation. The null hypothesis to be tested in each instance was that the two proportions (% methylated) were equal (p1 = p2). For each set of hypothetical experimental parameters, a power statistic was generated and recorded to create the table of statistical power in Fig. 3.

Acknowledgements

References

1. **Bernstein BE, Meissner A, Lander ES.** The mammalian epigenome. *Cell* 2007; **128**: 669–681. [Medline] [CrossRef]

2. **Goldberg AD, Allis CD, Bernstein E.** Epigenetics: a landscape takes shape. *Cell* 2007; **128**: 635–638. [Medline] [CrossRef]

3. **Bird A.** DNA methylation patterns and epigenetic memory. *Genes Dev* 2002; **16**: 6–21. [Medline] [CrossRef]

4. **Feng S, Jacobsen SE, Reik W.** Epigenetic reprogramming in plant and animal development. *Science* 2010; **330**: 622–627. [Medline] [CrossRef]

5. **Esteller M.** Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007; **8**: 286–298. [Medline] [CrossRef]

6. **Baylin SB, Jones PA.** A decade of exploring the cancer epigenome – Biological and translational implications. *Nat Rev Cancer* 2011; **11**: 726–734. [Medline] [CrossRef]

7. **Kwok JB.** Role of epigenetics in Alzheimer's and Parkinson's disease. *Epigenomics* 2010; **2**: 671–682. [Medline] [CrossRef]

8. **Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS.** Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol* 2009; **5**: 401–408. [Medline] [CrossRef]

9. **Hedrich CM, Tsokos GC.** Epigenetic mechanisms in systemic lupus erythematosus and other autoimmune diseases. *Trends Mol Med* 2011; **17**: 714–724. [Medline] [CrossRef]

10. **Farin PW, Slenning BD, Britt JH.** Estimates of pregnancy outcomes based on selection of bovine embryos produced *in vivo* or *in vitro*. *Theriogenology* 1999; **52**: 659–670. [Medline] [CrossRef]

11. **Hasler JF.** The current status of oocyte recovery, *in citro* embryo production, and embryo transfer in domestic animals, with an emphasis on the bovine. *J Anim Sci* 1998; **76**(Suppl 3): 52–74.

12. **Watanabe S, Nagai T.** Survival of embryos and calves derived from somatic cell nuclear transfer in cattle: a nationwide survey in Japan. *Anim Sci J* 2011; **82**: 360–365. [Medline] [CrossRef]

13. **Authority EFS.** Update on the state of play of animal cloning. *EFSA Journal* 2010; **8**: 1784–1804.

14. **Laird PW.** Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 2010; **11**: 191–203. [Medline] [CrossRef]

15. **Boerno ST, Grimm C, Lehrach H, Schweiger MR.** Next-generation sequencing technologies for DNA methylation analyses in cancer genomics. *Epigenomics* 2010; **2**: 199–207. [Medline] [CrossRef]

16. **Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR.** Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009; **462**: 315–322. [Medline] [CrossRef]

17. **Holmes R, Soloway PD.** Regulation of imprinted DNA methylation. *Cytogenet Genome Res* 2006; **113**: 122–129. [Medline] [CrossRef]

18. **Dindot SV, Person R, Strivens M, Garcia R, Beaudet AL.** Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res* 2009; **19**: 1374–1383. [Medline] [CrossRef]

19. **Velker BA, Denomme MM, Mann MR.** Embryo culture and epigenetics. *Methods Mol Biol* 2012; **912**: 399–421. [Medline]

20. **Thompson JR, Williams CJ.** Genomic imprinting and assisted reproductive technology: connections and potential risks. *Semin Reprod Med* 2005; **23**: 285–295. [Medline] [CrossRef]

21. **Manipalviratn S, DeCherney A, Segars J.** Imprinting disorders and assisted reproductive technology. *Fertil Steril* 2009; **91**: 305–315. [Medline] [CrossRef]

22. **Hiura H, Okae H, Miyauchi N, Sato F, Sato A, Van De Pette M, John RM, Kagami M, Nakai K, Soejima H, Ogata T, Arima T.** Characterization of DNA methylation errors in patients with imprinting disorders conceived by assisted reproduction technologies. *Hum Reprod* 2012; **27**: 2541–2548. [Medline] [CrossRef]

23. **Waterland RA, Jirtle RL.** Transposable elements: targets for early nutritional effects on epigenomic gene regulation. *Mol Cell Biol* 2003; **23**: 5293–5300. [Medline] [CrossRef]

24. **Couldrey C, Lee RS.** DNA methylation patterns in tissues from mid-gestation bovine foetuses produced by somatic cell nuclear transfer show subtle abnormalities in nuclear reprogramming. *BMC Dev Biol* 2010; **10**: 27. [Medline] [CrossRef]

25. **Breton CV, Salam MT, Gilliland FD.** Heritability and role for the environment in DNA methylation in AXL receptor tyrosine kinase. *Epigenetics* 2011; **6**: 895–898. [Medline] [CrossRef]

26. **Downing C, Johnson TE, Larson C, Leakey TI, Siegfried RN, Rafferty TM, Cooney CA.** Subtle decreases in DNA methylation and gene expression at the mouse Igf2 locus following prenatal alcohol exposure: effects of a methyl-supplemented diet. *Alcohol* 2011; **45**: 65–71. [Medline] [CrossRef]

27. **Martin L, Besch-Williford C, Lai L, Cheong HT, Im GS, Park KW, Murphy C, Hao Y, Ellersieck MR, Keisler DH, Schatten H, Green JA, Prather RS.** Morphologic and histologic comparisons between *in vivo* and nuclear transfer derived porcine embryos. *Mol Reprod Dev* 2007; **74**: 952–960. E. [Medline] [CrossRef]

28. **Rohde C, Zhang Y, Reinhardt R, Jeltsch A.** BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics* 2010; **11**: 230. [Medline] [CrossRef]

29. **Lenth RV.** Java applets for power and sample size [computer software]. 2006. Retrieved July 15, 2012 from http://www.stat.uiowa.edu/~rlenth/Power.

30. **Cohen J.** Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates; 1988. xxi, 567 p.