# BRIEF COMMUNICATION: Development of epigenomic analysis in agricultural animals

C Couldrey[a]*, R Brauning[b], EF Retzel[c] and JC McEwan[b]

*[a]AgResearch Ruakura,Private Bag 3123, Hamilton 3240, New Zealand; [b]AgResearch Invermay, Private Bag 50-034, Mosgiel 9053, New Zealand; [c]National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, NM, USA 87505*
*Corresponding author. Email: christine.couldrey@agresearch.co.nz*

**Keywords**: epigenetics; DNA methylation; sheep; Carwell; muscular hypertrophy; epigenomic pipeline

## Introduction

Advances made in DNA sequencing technologies over the past 15 years (Zhao & Grant 2011) have resulted in genome wide selection for production traits becoming a reality. Soon all sequence variations in each individual in closed breeding schemes will be known. What is not currently known is how to rank these variations, especially those that involve changes in gene expression rather than amino acid sequence. In order to continue making genetic progress we will need an understanding of how epigenetics affects gene expression and ultimately phenotype.

Epigenetics is the study of changes in phenotype caused by mechanisms other than changes in the underlying DNA sequence. Epigenetic mechanisms include but are not limited to, DNA methylation, acetylation and methylation of histone proteins that bind and stabilise DNA, and non-coding RNA molecules. These epigenetic processes act in an interrelated manner in order to regulate gene expression (Bird 2002). One of the key determinants in the control of gene expression in mammals is DNA methylation – a mechanism known to play a central role in regulating many aspects of growth and development (Bird 2002). Aberrant DNA methylation is able to preclude normal development and lead to disease (Gopalakrishnan et al. 2008). Despite the clear importance of DNA methylation, it remains poorly understood how methylation patterns are set during development, maintained, change in response to environmental conditions, and regulate gene expression.

Massively parallel DNA sequencing has recently become a vital tool in the analysis of DNA methylation and reduced representation bisulfite sequencing (RRBS) has proven to be effective in understanding DNA methylation landscapes (Cokus et al. 2008; Meissner et al. 2008; Smith et al. 2009). However, to date, mammalian genome-wide epigenetic studies have focused on humans and mice. Here we describe development of a RRBS protocol and bioinformatic analysis in sheep using the Carwell phenotype as a proof of principle. This phenotype is an inherited muscular hypertrophy trait which, in spite of its localisation to a well-defined 350kb region at the distal end of *Ovis aries* (OAR) chromosome 18 (Nicoll et al. 1998), considerable resequencing efforts, determination of imprinted inheritance and gene expression/splice variant patterns, the causative mutation has still not been identified. The commercially desirable phenotype, apparent absence of a causative mutation and known involvement of DNA methylation in a genomic imprinted region, offers a unique opportunity for use as a model system in which to develop capabilities for generating and analysing next generation bisulfite sequencing data in order to find (epi)mutations that correlate with economically important agricultural phenotypes across the whole genome.

## Materials and methods

A sample of the *M. longissimus dorsi* from one wild-type eight month old lamb was collected after slaughter and high quality DNA was extracted using standard phenol chloroform methodology (Sambrook et al. 1989). A RRBS methodology based on previously published studies (Cokus et al. 2008; Meissner et al. 2008; Smith et al. 2009) was developed. MspI restriction enzyme was used to digest 5 µg genomic DNA. Illumina sequencing adapters containing methylated cytosines were ligated onto digested DNA. Agarose gel size selection of DNA fragments was performed to obtain inserts of 150-250 bp. Bisulfite conversion of non-methylated Cs was performed on size selected fragments using EZ-DNA bisulfite conversion kit (Zymo, CA, USA) following the manufacturer's instructions except with a modification to bisulfite conversion conditions as recommended by (Smith et al. 2009); 99°C for 5 min, 60°C 25 min, 99°C 5 min, 60°C 85 min, 99°C 5 min, 60°C 175 min, 6 x (95°C 5 min, 60°C 90 min). PCR amplification was performed on converted DNA and sequenced on two lanes of an Illumina HiSeq sequencer using 50bp paired-end reads (National Center for Genome Resources, Santa Fe NM, USA). RRBS was performed in duplicate.

Quality control of data was undertaken using FastQC software (Babraham Bioinformatics, Cambridge, UK) in order to check for appropriate base composition cytosine (C), guanine (G), thymine (T) and adenine (A), extent of bisulfite conversion, and quality of sequence. Sequences were mapped to sheep genome assembly OARv2 using Bismark software (Babraham Bioinformatics, Cambridge, UK). Sequencing read counts and levels of methylation at individual CpG sites (cytosine

**Figure 1** Base composition of reduced representation bisulfite sequencing (RRBS) sequence from two lanes of Illumina HiSeq next generation sequencing from one technical replicate (219,712,546 individual sequences). For clarity of the first three bases, the spacing of the x-axis has been generated using the square root of the base pair position number.

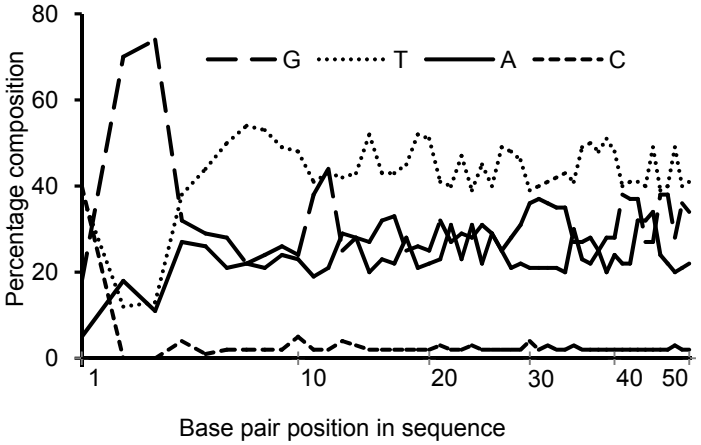
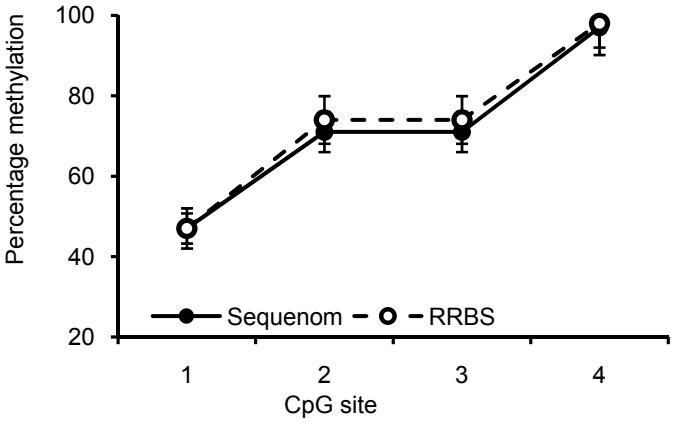



**Figure 2** Comparison of DNA methylation measurement using reduced representation bisulfite sequencing (RRBS) or Sequenom analysis at four representative CpG sites in the Carwell region of sheep muscle. Sequenom analysis (n = 6) with standard error of mean; RRBS with two technical replicates with standard error of mean.




followed by a guanine) were calculated and visualised using Seqmonk software (Babraham Bioinformatics, Cambridge, UK).

Accuracy of RRBS protocol was assessed by using Chi-squared analysis to compare DNA methylation levels at 134 selected CpG sites within the Carwell locus measured using RRBS and Sequenom MassARRAY (Coolen et al. 2007; Couldrey & Lee 2010; Couldrey et al. 2011) in six *M. longissimus dorsi* samples collected after slaughter from eight month old lambs.

## Results and discussion

Two lanes of Illumina HiSeq sequencing generated approximately 220 million sequence reads for each of the technical replicates. Approximately 180 million reads were of sufficiently high quality and suitable for mapping in each replicate.

Quality control analysis using FastQC indicated that on average the 50 base pair sequences displayed the expected nucleotide composition. An example of this is shown in Fig. 1. Because DNA was fragmented with MspI restriction enzyme, sequencing from the 5' end of the inserts should result in either a "C" or "T" (depending on original methylation status) in position 1, "G" in positions 2 and 3 and the remaining sequence being almost devoid of "C" and enriched for "T" because of the conversion of all non-methylated C's to T's. On average 70% of the sequences began with the expected nucleotides TGG or CGG when sequencing from the 5' end of the insert. The remaining base composition was C poor and T rich as expected. Similarly sequencing from the 3' end of fragments resulted in ~70% of sequences beginning with the expected CAA triplet and these sequences enriched for "A" and almost devoid of "G".

Bismark sequence mapping parameters were optimised for the sheep genome and capacity of servers currently available for analysis. Genome wide DNA methylation was able to be determined at a single nucleotide resolution in regions selected by reduced representation. Repeatability of the RRBS protocol and pipeline was determined by comparison of biological replicates. Depending on the stringency of analysis, the repeatability of the RRBS pipeline developed varied between R = 0.82 when the minimum number of reads covering a CpG site was set at 10, to R = 0.95 when a minimum of 100 reads were required for a CpG site to be included in the analysis. Given the complex protocol for generation of RRBS and the massively parallel sequencing, the precision of the epigenomic pipeline presented here is sufficient to allow high resolution analysis for the comparison of different phenotypes.

In addition to RRBS producing precise methylation measurements, it also accurately measured the extent of DNA methylation as illustrated in four representative sites shown in Fig. 2. A comparison of data from the RRBS methylation pipeline and Sequenom analysis of 134 CpG sites within the Carwell region identified that the proportion methylation measured by the two independent methods were indistinguishable at 132/134 CpG sites.

CpG sites are not randomly distributed throughout the mammalian genome (Lander et al. 2001). RRBS therefore aims to enrich CpG sites analysed while minimising the sequencing required. Approximately 1% of the genome was selected for sequencing and, using a lower limit of 10 sequencing

reads for any given CpG to be included in analysis, we have calculated that >20% of all CpG sites in the genome were able to be analysed. This level of enrichment of CpG sequencing represents a dramatic cost saving compared to sequencing the entire genome following bisulfite conversion. There are no published data on the enrichment of RRBS in humans or mice at a single nucleotide level as previous studies in these species have not sequenced sufficiently deep. A comparison of enrichment of CpG islands coverage is also problematic because the CpG island rich regions of many sheep genes are missing in the current genome assembly.

We have generated what is believed to be the first sheep methylome and optimised RRBS for use in agricultural animals. This protocol and bioinformatic analysis will have widespread use in the future through facilitating the identification of superior individuals to enhance productivity through further selective breeding.

## Acknowledgements

The authors would like to thank Ovita for sheep muscle samples, Tim Manley and Jeremy Bracegirdle for the Sequenom MassARRAY sample analysis, and Harold Henderson for statistical assistance. Work was supported by an AgResearch Research and Capability Fund Emerging Leaders grant.

## References

Bird A 2002. DNA methylation patterns and epigenetic memory. Genes and Development 16: 6–21.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215–219.

Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ 2007. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Research 35: E119.

Couldrey C, Lee RS 2010. DNA methylation patterns in tissues from mid-gestation bovine foetuses produced by somatic cell nuclear transfer show subtle abnormalities in nuclear reprogramming. BMC Developmental Biology 10: 27.

Couldrey C, Wells DN, Lee RS 2011. DNA methylation patterns are appropriately established in the sperm of bulls generated by somatic cell nuclear transfer. Cellular Reprogramming 13: 171–177.

Gopalakrishnan S, Van Emburgh BO, Robertson KD 2008. DNA methylation in development and human disease. Mutation Research 647: 30–38.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W and others 2001. Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.

Nicoll GB, Burkin HR, Broad TE, Jopson NB, Greer GJ, Bain WE, Wright CS, Dodds KG, Fennessy PF, McEwan JC 1998. Genetic linkage of microsatellite markers to the Carwell locus for rib-eye muscling in sheep. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, NSW, Australia 26: 529–532.

Sambrook J, Fritsch EF, Maniatus T 1989. Molecular cloning a laboratory manual. 2nd edn. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Smith ZD, Gu H, Bock C, Gnirke A, Meissner A 2009. High-throughput bisulfite sequencing in mammalian genomes. Methods 48: 226–232.

Zhao J, Grant SF 2011. Advances in whole genome sequencing technology. Current Pharmaceutical Biotechnology 12: 293–305.