Appendix S1. Methods

Biological samples

Skeletal muscle samples were obtained from a flock of Dorset/Suffolk/Rambouillet cross-bred sheep raised at Utah State University. Samples were taken of *longissimus lumborum* (LL) skeletal muscles from three fetuses at 100 days of gestation and three lambs at 12 weeks postpartum (Vuocolo *et al.* 2007; Byrne *et al.* 2010). The fetal samples corresponded with tertiary myotube formation (Bonnet *et al.* 2010). This near-term fetal-lamb developmental transition has been associated with a major switch in gene expression in skeletal muscle as well as changes in muscle fibre oxidative status and muscle cell hypertrophy (Byrne *et al.* 2010). The LL muscle was dissected from the animals at equivalent pre-determined sites within 15 minutes of euthanasia, weighed, frozen under liquid nitrogen and stored at -80°C. The fetuses included one male and two females while the lambs were all male. Animals were reared and euthanased in accordance with the animal ethics guidelines of Utah State University (Logan, Utah, USA).

Microarray transcript profiling

Gene expression analyses were performed by using RNA isolated from the fetal and lamb ovine LD muscle samples (n=3/ group). Affymetrix Bovine GeneChip microarrays (Affymetrix, Santa Clara, CA), containing 24,072 probe sets were employed, as previously described (Fleming-Waddell *et al.* 2007; Vuocolo *et al.* 2007; Byrne *et al.* 2010). All microarray images and quality control measurements were within recommended limits. Previous studies indicated that the bovine Affymetrix microarray could be used for analysis of ovine gene expression although there was some minor data loss due to species specific sequence differences in the probe sets (Vuocolo *et al.* 2007). Expression values were determined with the GeneChip Operating Software (GCOS) v1.1.1 and the MAS5.0 processing algorithm (Fleming-Waddell *et al.* 2007; Vuocolo *et al.* 2007; Byrne *et al.* 2010). All data are available at the National Centre for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), as raw and processed (MAS5.0) data (Accession number GSE20112) . There was also good correspondence between changes in gene expression measured by the microarrays and qRT-PCR.

Chromatin immunoprecipitation followed by sequencing

Experimental design: Enrichment for histone 3 lysine 27 trimethylation (H3K27me3) at genomic loci was measured using chromatin immunoprecipitation followed by next generation sequencing (ChIP-Seq) for the fetal (n=3) and lamb (n=3) LL muscle samples. ENCODE minimum standards were used for ChIP-Seq unless otherwise specified (Consortium 2011a). Each sample was initially independently processed. Pooled input nucleosomal DNA from the three animals was used as the input control for each developmental age.

Chromatin immunoprecipitation: Chromatin immunoprecipitation was performed essentially according to an established method (Wagschal *et al.* 2007). Briefly, nuclei were isolated from 1.2 g of thawed LL muscle and the chromatin subsequently treated with micrococcal nuclease (MNase) (10 U/ µl) (New England BioLabs, Ipswich, MA) to generate nucleosomes. Preliminary experiments optimised the concentration of MNase and time of chromatin digestion. Soluble MNase digested chromatin was recovered using mild sonication and centrifugation. To assess the extent of MNase digestion and the production of nucleosomes, the soluble chromatin was digested with proteinase K and DNA extracted by phenol/

chloroform (1:1) and precipitated in the presence of 250 mM NaCl and glycogen (20 µg/ml). The DNA was further purified using the PCR mini-elute purification kit (Qiagen) and subjected to 1.5% agarose gel electrophoresis with DNA visualised using ethidium bromide. A single strong ~150 bp band, corresponding to DNA associated mononucleosomes, was present in all samples. Direct incubation of purified genomic DNA with MNase did not produce a distinct ~150 bp band. Although MNase has a cleavage site nucleotide bias this is not a problem for nucleosome mapping (Allan *et al.* 2012; Gaffney *et al.* 2012).

The soluble chromatin was pre-cleared with Protein A-Sepharose / salmon sperm DNA (Millipore). A proven and highly specific antibody to H3K27me3 (Millipore 07-449; 10 µg per incubation) was used for immunoprecipitation (Bock *et al.* 2010; Egelhofer *et al.* 2011). Immune complexes were precipitated with Protein A-Sepharose, washed sequentially with low salt buffers supplemented with 0.1% NP-40 in 1xTE buffer and then eluted in ChIP Elution Buffer (50 mM NaCl, 50 mM Tris-Cl (pH 7.5), 0.1 mM PMSF, 5 mM EDTA, 1% SDS (w/v)). The H3K27me3 immunoprecipitated DNA and the input nucleosomal DNA were quantified using Quant-IT^{TM} PicoGreen double stranded DNA Reagent (Invitrogen). Approximately 50-100 ng of purified DNA was isolated from each H3K27me3 immunoprecipitation. The antibody to H3K27me3 has been widely validated and in the current investigation its specificity was demonstrated by the strong association of H3K27me3 with (i) homeobox encoding transcription factors (Pauler *et al.* 2009), (ii) X chromosome inactivation in females (King *et al.* 2008; Escamilla-Del-Arenal *et al.* 2011), as well as (iii) strong enrichment of the modified genes in a variety of datasets available in GSEA which were generated by manipulation of the polycomb protein complex responsible for H3K27 methylation , and (iv) substantial agreement with human ENCODE data for H3K27me3 in a variety of tissues (Zhou *et al.* 2011).

ChIP-Seq data generation: DNA isolated from purified H3K27me3 nucleosomes and input nucleosomes was sequenced by using the Illumina Genome Analyser II (GAII) sequencing-by-synthesis technology in accordance with standard operating procedures (Genome Analysis Services, Geneworks, Adelaide, SA). End repaired and adaptor-ligated DNA fragments of ~170 bp were size selected from acrylamide gels, purified and amplified by PCR. Each library was validated using an Agilent 2100 Bioanalyser and then sequenced. Single-end 65 bp sequence reads were generated and base quality scores were produced using the GAII data analysis software pipeline (v 1.3.2). Approximately 10 million sequence reads per sample were generated. A total of eight libraries were produced corresponding to three fetal and three lamb H3K27me3 samples, as well as the two input nucleosomal DNA controls for each developmental stage. The number of reads per sample was consistent with ENCODE recommendations (CONSORTIUM 2011b; Ma & Wong 2011). Mean quality scores were plotted as a function of sequence position for each library and all libraries showed Q_{20} scores extending beyond 60 bp. A total of 95.8 million reads (6.2 Gb) was generated. The ChIP-Seq data has been deposited in the NCBI Gene Expression Omnibus (GEO) under the SuperSeries GSEXXXXX.

Mapping sequence reads to the bovine genome sequence

Sequence reads were mapped to the bovine genome sequence as an annotated ovine genome sequence was unavailable. Orthologous ovine and bovine sequences show considerable identity i.e. ~ 97% in gene coding sequences and ~93% in intergenic sequences (Kijas *et al.* 2006) (our unpublished data). Moreover, both species have substantial similarities in their chromosomal organisation and extensive gene synteny (Goldammer *et al.* 2009). Therefore, it was anticipated that there would be efficient mapping of ovine sequence reads onto the bovine

annotated genome sequence. Sequence reads were initially screened for adaptor and filtered for mammalian repeats using RepeatMasker (RM-3.2.8 and db-20090604) and reads with ≥ 50 bp of unmasked sequence were subsequently used. Studies using the human genome revealed that 79.6% of the genome is uniquely mapable using 30 bp reads (Rozowsky *et al.* 2009). Although a number of sequence alignment algorithms were available, the cross-species nature of the alignment, particularly the potential for indels, was not easily accommodated by most programs, which are typically optimised for identity mapping within a species. Consequently, the ovine sequence reads were mapped to the bovine genome (Btau_4.0 (Elsik *et al.* 2009; Tellam *et al.* 2009), including the unassigned scaffolds, using BLAT (Kent 2002) but optimised with the following parameters: -minIdentity=75; -tileSize=11; maxIntronSize=5; StepSize=5 –mask=lower and -qMask=lower. Ovine sequence reads with > 75% identity over > 50 bp, allowing for an indel size of up to 5 bp, were identified. Of the total number of ChIP-Seq reads, 11.6 and 9.9 million from the fetal and lamb samples, respectively, passed three filters: (i) unique mapping to the bovine genome; (ii) length of the sequence match ≥ 50 bp, and; (iii) sequence identity ≥ 75 %. The uniquely mapped reads for each sample were converted to alignment files detailing the genomic co-ordinates and strand for each read. Btau_4.0 is 89% complete (Elsik *et al.* 2009) and thus some reads were likely lost due to the incomplete nature of the genome sequence. Other sequences were not mapped due to localised species specific sequence divergence, particularly as a consequence of large indels, or the inability to be mapped uniquely to the genome. The Y chromosome sequence was unavailable for mapping.

H3K27me3 peak detection

The software package CisGenome was used to identify genomic regions of H3K27me3 nucleosome enrichment for each sample relative to the corresponding input nucleosome control (Ji *et al.* 2008). CisGenome is an integrated computational tool for ChIP-Seq analysis. The uniquely mapped sequence reads were extended to a read length of 147 bp (direction dependent on strand), representing the length of the input nucleosomal DNA. The signal was the count of the number of overlapping DNA reads at each nucleotide position. H3K27me3 peaks were identified by using the CisGenome software based on the number of reads derived from the H3K27me3 immunoprecipitated DNA in a sliding window of 200 bp, stepped every 50 bp, compared to the number of reads in the same region for the input nucleosomal control i.e. (a *two sample analysis*). A conservative threshold of 10 or more sequence reads in the 200 bp window was applied. CisGenome normalised for library size when the peaks were generated. FDR estimates were based on the level of H3K27me3 nucleosome enrichment in this window compared to the input nucleosome control using a conditional binomial model with a FDR ≤ 0.1 (Ji *et al.* 2008). Windows of enriched H3K27me3 nucleosomes that overlapped were merged into one region, retaining the minimal FDR and maximum fold change enrichment values. This pair-wise analysis was performed for each of the biological replicates. Gaps of H3K27me3 modification within broad regions of strong modification are likely due to local sequence regions that do not uniquely map, or local sequence divergence between the ovine and bovine genome. Reads mapping to genomic sequence unassigned to a bovine chromosome (ChrUN) were not included in subsequent analyses.

The H3K27me3 enriched regions showed strong overlaps in pairwise comparisons between the three biological replicates at each developmental stage. The percentage of H3K27me3 peaks that overlapped in pairwise comparison by more than 1 bp in each group was 62-74% and 67-77% for the fetal and lamb samples, respectively. Moreover, the H3K27me3 read counts in RefSeq gene promoters showed pairwise correlation coefficients of 0.87- 0.94 and

0.86-0.91 for the biological replicates within the fetal and lamb groups, respectively. The top ranked 500 genes showed greater than 96% overlap between biological replicates in each biological state. Moreover, there was greater than 95% overlap between the top ranked 500 genes for the fetal and lamb biological replicates. Thus, there was good agreement between the biological replicates in terms of peak position and overlap of genes ranked according to H3K27me3 promoter read counts in the fetal and lamb samples.

To examine the architecture of the H3K27me3 modification across the genome in each developmental state, the uniquely aligned reads from the three biological replicates within each developmental group were pooled and compared to their age matched input nucleosomal control reads using CisGenome for H3K27me3 peak detection, as described above (*two sample analysis*), but excluding the X chromosome from analysis. Combining sequence reads from the biological replicates provided additional statistical power for detection of H3K27me3 peaks. The X chromosome was excluded due to the largely chromosome-wide enrichment of H3K27me3 nucleosomes on one of the two X chromosomes in females due to X chromosome inactivation. Peak list tables (*.cod files) produced by CisGenome defined the peak parameters (chromosome, peak start, peak end, peak strand, peak length, FDR, $\log_2$ (fold change) and strand specific read counts (forward and reverse)) for the fetal and lamb states. These data were used to identify correlations between H3K27me3 peaks and transcriptional start sites (TSS), CpG Islands (CGIs) and genes (CDS). They were also used to define genes with promoters enriched with H3K27me3 peaks (see below).

Nucleotide profiles of input nucleosomes and H3K27me3 nucleosomes

The G+C contents of the 147 bp H3K27me3 nucleosomes located in peaks by CisGenome and the input nucleosomes were determined for each biological state. For the latter, the input nucleosome 65 bp sequences that uniquely mapped to the bovine genome (excluding ChrUN and chromosome X) were extended to 147 bp as previously described and input nucleosome peaks were independently detected using a CisGenome *one sample analysis*, a sliding window of 200 bp and minimum read count of 10. False discovery rates were estimated by modelling the read count in non-binding windows using a negative binomial distribution. By comparing the observed number of windows that have k reads with the expected number of windows that have k reads, an estimate of the false positive rate was made (Ji *et al.* 2008). A FDR cutoff of 0.1 was used. The nucleosome and H3K27me3 nucleosome peak co-ordinates were used to extract relevant 147 bp sequences. In addition, one million random 147 bp sequences were extracted from the unmasked portion of the bovine genome. The G+C content of each 147 bp nucleosome sequence and frequency of nucleotides at each position in the sequence were then determined for all three data sets i.e. (i) H3K27me3 enriched nucleosomes; (ii) input nucleosomes, and; (iii) the randomised control. This was performed for both the lamb and fetal samples. Differences in G+C content were defined using the Mann-Whitney U test. The ChIP-Seq method involved preparation of nucleosomes using MNase (Dingwall *et al.* 1981). The cleavage site for MNase was associated with a nucleotide bias however this has been shown to not be a problem for nucleosome mapping as similar results can be obtained with other nucleases of differing specificities (Allan *et al.* 2012; Gaffney *et al.* 2012). Exclusion of the first and last five nucleotides from the 147 bp nucleosomal sequences gave essentially the same results for G+C contents.

Visualisation of ChIP-Seq data using Hilbert space filling curves

The H3K27me3 peaks identified in the pooled analyses for each developmental state were visualised using a Hilbert space filling curve for each autosome. A Hilbert curve is a continuous fractal space filling curve that maintains locality information. The Hilbert curve

visualisation tool *HilbertVis* was used, which is adapted for representation of genomic data (Anders 2009). The tool allows the display of very long data vectors (e.g. genomic features of chromosomes) in a space-efficient manner that maintains locality relationships thereby allowing high resolution visualisation of genomic features (i.e. H3K27me3 peaks). The H3K27me3 peak co-ordinates (in *.cod files) for the fetal and lamb developmental time points were converted to gff (General Feature Format) files and uploaded into the HilbertVis GUI tool. A Hilbert curve for each autosome and for each developmental state was generated.

Association of sequence reads within H3K27me3 peaks with genome architecture

Sequence reads within H3K27me3 peaks in both developmental states were examined for their association with structural features of the bovine genome including genes, transcriptional start sites (TSS) and CpG islands (CGI). The *B. taurus* genome coordinates (assembly BTAU_4.0) for RefFlat genes were obtained from the University of Santa Cruz California (UCSC) website (UCSC). (UCSC)(UCSC)(UCSC)The RefFlat gene set contained 10,891 genes annotated for transcription and coding sequence start and stop coordinates. This information was derived from full length cDNA sequences. A further set of 29,027 gene sequences (RefSeq) was obtained by downloading the NCBI *B. taurus* reference gene sequences and their coordinates . RefSeq genes are a collection of curated, non-redundant gene sequences produced by NCBI. 'LOC' RefSeq genes were manually updated for recent annotations. The positions for 37,595 CpG islands (CGI) in BTAU_4.0 were downloaded from the UCSC genome browser. There was no attempt to correct for alternative promoters. For each gene set the distance from the gene feature start site to each H3K27me3 peak was calculated and the number of sequence reads within each 100 bp window derived to a maximum distance of 10 kb in either direction. For the RefFlat dataset, the sequence start was defined as the TSS while for the RefSeq dataset the gene start was defined as the first base of exon 1. The H3K27me3 peak distances from CGI boundaries were similarly calculated. To determine the correlation between H3K27me3 peaks and CGI, the coordinates of each CGI were extended 500 bp either side to capture CpG island shores (Doi *et al.* 2009; Irizarry *et al.* 2009) and all H3K27me3 peaks that overlapped with these extended coordinates, including partial overlaps, were then identified. The number of sequence reads within these windows for each genome feature was then determined. Each analysis was independently performed for the fetal and lamb samples. H3K27me3 peaks overlapping gene promoter regions were calculated using a 2.5 kb window that included 2 kb upstream and 500 bp downstream of the first base of exon 1 of each RefSeq gene (Mikkelsen *et al.* 2007).

Association of sequence reads in H3K27me3 peaks in RefSeq gene promoters with gene expression

In order to correlate the co-localisation of H3K27me3 enriched regions with gene expression, we compared sequence reads in H3K27me3 peaks in promoters of autosomal RefSeq genes with corresponding gene expression data. The bovine Affymetrix microarray was used for measurement of gene expression in the LL skeletal muscle samples (Fleming-Waddell *et al.* 2007; Vuocolo *et al.* 2007; Byrne *et al.* 2010). H3K27me3 peaks were detected in 4,070 and 3,384 promoters from the RefSeq annotated gene list for the fetal and lamb samples, respectively. The H3K27me3 sequence counts within these gene promoter windows were then determined, normalised for library size and corrected for the normalised input control. Of the genes containing H3K27me3 sequence tags in their promoters, gene expression signals were correspondingly measured for 1,953 and 1,466 Affymetrix probe-sets, corresponding to 1,430 and 1,107 annotated unique genes for the fetal and lamb samples, respectively. Probe-sets which reported 'no' expression (called as *Absent* by MAS5.0 in all biological replicates for a developmental state) were also assessed for H3K27me3 levels within the promoter regions of

the corresponding genes. Gene expression levels were divided into four categories: no expression; low expression (log_2 scaled averages <4); medium expression (log_2 scaled averages ≥4 and <8) and high expression (log_2 scaled averages ≥ 8). The independent MAS5 Affymetrix flag calls for the expressed genes were either Present or Marginal for all biological replicates for each developmental state. A *t* test was used to measure statistical differences between the levels of H3K27me3 sequence reads in each gene expression category compared with the 'no' expression gene category.

Differentially modified gene promoters between the fetal and lamb developmental states
The strategy used for detection of RefSeq autosomal genes with significant differences in H3K27me3 promoter density was as follows. The number of sequence reads in the promoters of genes for each of the biological replicates was determined. For each gene these reads were corrected for library size and corresponding normalised input controls were subtracted. A figure of 0.1 reads was added to any zero values and quantile normalisation employed. Significant differences in H3K27me3 promoter reads between the two developmental states were then determined using an adaptation of the statistical package designed to analyse RNA-seq data in the commercial package CLCBIO (CLCBIO). An FDR corrected P value ≤ 0.1 was considered significant. SuppInfo_8 lists the differentially modified genes including fold change relative to the fetal sample, P-value, FDR corrected P-value, fetal normalised mean value, and lamb normalised mean value for each gene. Genes associated with changes in H3K27me3 promoter modification may, in some cases, reflect alternative promoter use (data not shown).

Gene function enrichment analysis
To identify higher level functional themes associated with gene lists, the Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.7) was used (Huang da *et al.* 2009a; Huang da *et al.* 2009b). DAVID provides statistical methods for identification of enriched biological and structural terms within a gene list compared to a relevant background gene list. For the analysis of modified genes in the fetal or lamb samples, statistically over-represented Gene Ontology (GO) terms, keywords and pathways as well as protein structural themes were identified with a Benjamini-Hochberg-adjusted p-value ≤ 0.05 and a minimum number of 10 genes/ term. Functional Annotation Clustering was performed using the DAVID system using the highest stringency option. This higher level analysis displayed related functional annotation terms together based on overlaps of genes associated with each function term. It therefore gives an overview of gene function information associated with large datasets. In this case enrichment scores (- log (geometric mean of the P-values for terms in the cluster)) greater than 1.3 with terms represented by at least 5 genes were considered significant. For the differentially modified genes a cluster enrichment score ≥ 1.0 for Functional Annotation Clustering and a p-value ≤ 0.05 for GO terms were used. The gene backgrounds were either the bovine autosomal RefSeq genes for the intragenome analysis or genes whose promoters showed enrichment of H3K27me3 for the analysis of differential modification between the biological states.

References for Appendix S1

ftp site for NCBI Bos taurus RefSeq genes. URL ftp://ftp.ncbi.nih.gov/refseq/B_taurus/.

Gene Expression Omnibus. URL http://www.ncbi.nlm.nih.gov/geo/.

Geneset enrichment analysis. URL http://www.broadinstitute.org/gsea/index.jsp.

RepeatMasker. URL http://www.repeatmasker.org/.

Allan J., Fraser R.M., Owen-Hughes T. & Keszenman-Pereyra D. (2012) Micrococcal nuclease does not substantially bias nucleosome mapping. *J Mol Biol* **417**, 152-64.

Anders S. (2009) Visualization of genomic data with the Hilbert curve. *Bioinformatics* **25**, 1231-5.

Bock I., Dhayalan A., Kudithipudi S., Brandt O., Rathert P. & Jeltsch A. (2010) Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays. *Epigenetics* **6**.

Bonnet M., Cassar-Malek I., Chilliard Y. & Picard B. (2010) Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal* **4**, 1093-109.

Byrne K., Vuocolo T., Gondro C., White J.D., Cockett N.E., Hadfield T., Bidwell C.A., Waddell J.N. & Tellam R.L. (2010) A gene network switch enhances the oxidative capacity of ovine skeletal muscle during late fetal development. *BMC Genomics* **11**, 378.

CLCBIO CLC Genomics Workbench 6. URL http://www.clcbio.com.

Consortium B.C.B. (2011a) ChIP-Seq Minimal Standards and Guidelines. URL http://www.betacell.org/documents/administered/about/guidelines/ENCODE_BCBC_ChIP-Seq_Standards_V01_20110503.pdf.

CONSORTIUM B.C.B. (2011b) ChIP-Seq Minimal Standards and Guidelines. URL http://genome.ucsc.edu/ENCODE/protocols/dataStandards/ChIP_DNase_FAIRE_DNAme_v2_2011.pdf.

Dingwall C., Lomonossoff G.P. & Laskey R.A. (1981) High sequence specificity of micrococcal nuclease. *Nucleic Acids Res* **9**, 2659-73.

Doi A., Park I.H., Wen B., Murakami P., Aryee M.J., Irizarry R., Herb B., Ladd-Acosta C., Rho J., Loewer S., Miller J., Schlaeger T., Daley G.Q. & Feinberg A.P. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**, 1350-3.

Egelhofer T.A., Minoda A., Klugman S., Lee K., Kolasinska-Zwierz P., Alekseyenko A.A., Cheung M.S., Day D.S., Gadel S., Gorchakov A.A., Gu T., Kharchenko P.V., Kuan S., Latorre I., Linder-Basso D., Luu Y., Ngo Q., Perry M., Rechtsteiner A., Riddle N.C., Schwartz Y.B., Shanower G.A., Vielle A., Ahringer J., Elgin S.C., Kuroda M.I., Pirrotta V., Ren B., Strome S., Park P.J., Karpen G.H., Hawkins R.D. & Lieb J.D. (2011) An assessment of histone-modification antibody quality. *Nat Struct Mol Biol* **18**, 91-3.

Elsik C.G., Tellam R.L., Worley K.C., Gibbs R.A., Muzny D.M., Weinstock G.M., Adelson D.L., Eichler E.E., Elnitski L., Guigo R., Hamernik D.L., Kappes S.M., Lewin H.A., Lynn D.J., Nicholas F.W., Reymond A., Rijnkels M., Skow L.C., Zdobnov E.M., Schook L., Womack J., Alioto T., Antonarakis S.E., Astashyn A., Chapple C.E., Chen H.C., Chrast J., Camara F., Ermolaeva O., Henrichsen C.N., Hlavina W., Kapustin Y., Kiryutin B., Kitts P., Kokocinski F., Landrum M., Maglott D., Pruitt K., Sapojnikov V., Searle S.M., Solovyev V., Souvorov A., Ucla C., Wyss C., Anzola J.M., Gerlach D., Elhaik E., Graur D., Reese J.T., Edgar R.C., McEwan J.C., Payne G.M., Raison J.M., Junier T., Kriventseva E.V., Eyras E., Plass M., Donthu R., Larkin D.M., Reecy

J., Yang M.Q., Chen L., Cheng Z., Chitko-McKown C.G., Liu G.E., Matukumalli L.K., Song J., Zhu B., Bradley D.G., Brinkman F.S., Lau L.P., Whiteside M.D., Walker A., Wheeler T.T., Casey T., German J.B., Lemay D.G., Maqbool N.J., Molenaar A.J., Seo S., Stothard P., Baldwin C.L., Baxter R., Brinkmeyer-Langford C.L., Brown W.C., Childers C.P., Connelley T., Ellis S.A., Fritz K., Glass E.J., Herzig C.T., Iivanainen A., Lahmers K.K., Bennett A.K., Dickens C.M., Gilbert J.G., Hagen D.E., Salih H., Aerts J., Caetano A.R., Dalrymple B., Garcia J.F., Gill C.A., Hiendleder S.G., Memili E., Spurlock D., Williams J.L., Alexander L., Brownstein M.J., Guan L., Holt R.A., Jones S.J., Marra M.A., Moore R., Moore S.S., Roberts A., Taniguchi M., Waterman R.C., Chacko J., Chandrabose M.M., Cree A., Dao M.D., Dinh H.H., Gabisi R.A., Hines S., Hume J., Jhangiani S.N., Joshi V., Kovar C.L., Lewis L.R., Liu Y.S., Lopez J., Morgan M.B., Nguyen N.B., Okwuonu G.O., Ruiz S.J., Santibanez J., Wright R.A., Buhay C., Ding Y., Dugan-Rocha S., Herdandez J., Holder M., Sabo A., Egan A., Goodell J., Wilczek-Boney K., Fowler G.R., Hitchens M.E., Lozado R.J., Moen C., Steffen D., Warren J.T., Zhang J., Chiu R., Schein J.E., Durbin K.J., Havlak P., Jiang H., Liu Y., Qin X., Ren Y., Shen Y., Song H., Bell S.N., Davis C., Johnson A.J., Lee S., Nazareth L.V., Patel B.M., Pu L.L., Vattathil S., Williams R.L., Jr., Curry S., Hamilton C., Sodergren E., Wheeler D.A., Barris W., Bennett G.L., Eggen A., Green R.D., Harhay G.P., Hobbs M., Jann O., Keele J.W., Kent M.P., Lien S., McKay S.D., McWilliam S., Ratnakumar A., Schnabel R.D., Smith T., Snelling W.M., Sonstegard T.S., Stone R.T., Sugimoto Y., Takasuga A., Taylor J.F., Van Tassell C.P., Macneil M.D., Abatepaulo A.R., Abbey C.A., Ahola V., Almeida I.G., Amadio A.F., Anatriello E., Bahadue S.M., Biase F.H., Boldt C.R., Carroll J.A., Carvalho W.A., Cervelatti E.P., Chacko E., Chapin J.E., Cheng Y., Choi J., Colley A.J., de Campos T.A., De Donato M., Santos I.K., de Oliveira C.J., Deobald H., Devinoy E., Donohue K.E., Dovc P., Eberlein A., Fitzsimmons C.J., Franzin A.M., Garcia G.R., Genini S., Gladney C.J., Grant J.R., Greaser M.L., Green J.A., Hadsell D.L., Hakimov H.A., Halgren R., Harrow J.L., Hart E.A., Hastings N., Hernandez M., Hu Z.L., Ingham A., Iso-Touru T., Jamis C., Jensen K., Kapetis D., Kerr T., Khalil S.S., Khatib H., Kolbehdari D., Kumar C.G., Kumar D., Leach R., Lee J.C., Li C., Logan K.M., Malinverni R., Marques E., Martin W.F., Martins N.F., Maruyama S.R., Mazza R., McLean K.L., Medrano J.F., Moreno B.T., More D.D., Muntean C.T., Nandakumar H.P., Nogueira M.F., Olsaker I., Pant S.D., Panzitta F., Pastor R.C., Poli M.A., Poslusny N., Rachagani S., Ranganathan S., Razpet A., Riggs P.K., Rincon G., Rodriguez-Osorio N., Rodriguez-Zas S.L., Romero N.E., Rosenwald A., Sando L., Schmutz S.M., Shen L., Sherman L., Southey B.R., Lutzow Y.S., Sweedler J.V., Tammen I., Telugu B.P., Urbanski J.M., Utsunomiya Y.T., Verschoor C.P., Waardenberg A.J., Wang Z., Ward R., Weikard R., Welsh T.H., Jr., White S.N., Wilming L.G., Wunderlich K.R., Yang J. & Zhao F.Q. (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522-8.

Escamilla-Del-Arenal M., da Rocha S.T. & Heard E. (2011) Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Hum Genet* **130**, 307-27.

Fleming-Waddell J.N., Wilson L.M., Olbricht G.R., Vuocolo T., Byrne K., Craig B.A., Tellam R.L., Cockett N.E. & Bidwell C.A. (2007) Analysis of gene expression during the onset of muscle hypertrophy in callipyge lambs. *Anim Genet* **38**, 28-36.

Gaffney D.J., McVicker G., Pai A.A., Fondufe-Mittendorf Y.N., Lewellen N., Michelini K., Widom J., Gilad Y. & Pritchard J.K. (2012) Controls of nucleosome positioning in the human genome. *PLoS Genet* **8**, e1003036.

Goldammer T., Di Meo G.P., Luhken G., Drogemuller C., Wu C.H., Kijas J., Dalrymple B.P., Nicholas F.W., Maddox J.F., Iannuzzi L. & Cockett N.E. (2009) Molecular

cytogenetics and gene mapping in sheep (Ovis aries, 2n = 54). *Cytogenet Genome Res* **126**, 63-76.
Huang da W., Sherman B.T. & Lempicki R.A. (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.
Huang da W., Sherman B.T., Zheng X., Yang J., Imamichi T., Stephens R. & Lempicki R.A. (2009b) Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 1.
Irizarry R.A., Wu H. & Feinberg A.P. (2009) A species-generalized probabilistic model-based definition of CpG islands. *Mamm Genome* **20**, 674-80.
Ji H., Jiang H., Ma W., Johnson D.S., Myers R.M. & Wong W.H. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**, 1293-300.
Kent W.J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64.
Kijas J.W., Menzies M. & Ingham A. (2006) Sequence diversity and rates of molecular evolution between sheep and cattle genes. *Anim Genet* **37**, 171-4.
King W.A., Coppola G., Pinton A., Joudrey E.M. & Basrur P.K. (2008) Spatial distribution of histone isoforms on the bovine active and inactive x chromosomes. *Sexual Development* **2**, 12-23.
Ma W. & Wong W.H. (2011) The analysis of ChIP-Seq data. *Methods Enzymol* **497**, 51-73.
Mikkelsen T.S., Ku M., Jaffe D.B., Issac B., Lieberman E., Giannoukos G., Alvarez P., Brockman W., Kim T.K., Koche R.P., Lee W., Mendenhall E., O'Donovan A., Presser A., Russ C., Xie X., Meissner A., Wernig M., Jaenisch R., Nusbaum C., Lander E.S. & Bernstein B.E. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60.
Pauler F.M., Sloane M.A., Huang R., Regha K., Koerner M.V., Tamir I., Sommer A., Aszodi A., Jenuwein T. & Barlow D.P. (2009) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res* **19**, 221-33.
Rozowsky J., Euskirchen G., Auerbach R.K., Zhang Z.D., Gibson T., Bjornson R., Carriero N., Snyder M. & Gerstein M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol* **27**, 66-75.
Tellam R.L., Lemay D.G., Van Tassell C.P., Lewin H.A., Worley K.C. & Elsik C.G. (2009) Unlocking the bovine genome. *BMC Genomics* **10**, 193.
UCSC UCSC Genome Bioinfomatics. URL http://genome.ucsc.edu/index.html.
UCSC UCSC Genome Bioinformatics. URL http://genome.ucsc.edu/.
Vuocolo T., Byrne K., White J., McWilliam S., Reverter A., Cockett N.E. & Tellam R.L. (2007) Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. *Physiol Genomics* **28**, 253-72.
Wagschal A., Delaval K., Pannetier M., Arnaud P. & Feil R. (2007) Chromatin Immunoprecipitation (ChIP) on Unfixed Chromatin from Cells and Tissues to Analyze Histone Modifications. *CSH Protoc* **2007**, pdb prot4767.
Zhou X., Maricque B., Xie M., Li D., Sundaram V., Martin E.A., Koebbe B.C., Nielsen C., Hirst M., Farnham P., Kuhn R.M., Zhu J., Smirnov I., Kent W.J., Haussler D., Madden P.A., Costello J.F. & Wang T. (2011) The Human Epigenome Browser at Washington University. *Nat Methods* **8**, 989-90.