

SCIENTIFIC REPORTS



OPEN

Global misregulation of genes largely uncoupled to DNA methylome epimutations characterizes a congenital overgrowth syndrome

Zhiyuan Chen^{1,3}, Darren E. Hagen^{1,4}, Tieming Ji², Christine G. Elsik¹ & Rocío M. Rivera¹

Assisted reproductive therapies (ART) have become increasingly common worldwide and numerous retrospective studies have indicated that ART-conceived children are more likely to develop the overgrowth syndrome Beckwith-Wiedemann (BWS). In bovine, the use of ART can induce a similar overgrowth condition, which is referred to as large offspring syndrome (LOS). Both BWS and LOS involve misregulation of imprinted genes. However, it remains unknown whether molecular alterations at non-imprinted loci contribute to these syndromes. Here we examined the transcriptome of skeletal muscle, liver, kidney, and brain of control and LOS bovine fetuses and found that different tissues within LOS fetuses have perturbations of distinct gene pathways. Notably, in skeletal muscle, multiple pathways involved in myoblast proliferation and fusion into myotubes are misregulated in LOS fetuses. Further, characterization of the DNA methylome of skeletal muscle demonstrates numerous local methylation differences between LOS and controls; however, only a small percent of differentially expressed genes (DEGs), including the imprinted gene *IGF2R*, could be associated with the neighboring differentially methylated regions. In summary, we not only show that misregulation of non-imprinted genes and loss-of-imprinting characterize the ART-induced overgrowth syndrome but also demonstrate that most of the DEGs is not directly associated with DNA methylome epimutations.

The use of ART has become increasingly common worldwide and each year approximately six percent of infants born in developed countries are conceived employing these technologies¹. The use of ART involves manipulation of gametes and/or embryos in an artificial environment, and it has been reported that ART-conceived children have increased risk for birth defects². For example, a meta-analysis of eight epidemiologic studies indicates that ART-conceived children have a 5.2-fold increased likelihood of developing the congenital overgrowth condition BWS³. However, it is still unresolved whether the use of ART or the infertility per se is the main cause of the loss-of-imprinting disorders observed in ART-conceived humans³. BWS is characterized by complex and variable symptoms including pre- and post-natal overgrowth, enlarged tongue, ear malformation, umbilical hernia, and predisposition to develop childhood tumors⁴. In ruminants, gametes and embryos subjected to *in vitro* manipulations can develop into unusually large offspring that share phenotypes with BWS^{5,6}. LOS, as the congenital overgrowth syndrome is referred to in ruminants, can cause detrimental effects to both the dam and offspring such as delivery difficulty due to the oversized fetus and inability for the newborn to stand and to suckle⁶. It is unclear what triggers the development of these congenital overgrowth conditions and their associated phenotypes and why ART potentiates the syndromes. Serum supplementation of the oocyte maturation medium and the embryo culture medium has been recognized as a mediator of LOS⁶, although the factors in the serum that are responsible for the overgrowth phenotype remain to be identified. Several animal studies have indicated that the

¹Division of Animal Sciences, University of Missouri, Columbia, MO, 65211, USA. ²Department of Statistics, University of Missouri, Columbia, MO, 65211, USA. ³Present address: ZC-159 G Warren Alpert Building, 200 Longwood Avenue, Boston, MA, 02115, USA. ⁴Present address: Department of Animal Science, Oklahoma State University, 311C Noble Research Center, Stillwater, OK, 74078, USA. Correspondence and requests for materials should be addressed to R.M.R. (email: riverarm@missouri.edu) or D.E.H. (email: Darren.Hagen@okstate.edu)

use of ART can alter the epigenome of the gametes and embryos^{7–10} and this can contribute to the etiology of the ART-induced overgrowth conditions^{3,11}.

DNA methylation is an epigenetic modification involving the addition of a methyl group to the 5th carbon of cytosine¹². In mammals, DNA methylation typically occurs in a CpG context with the exception of CpH methylation, which is mostly observed in neural tissues, oocytes and embryonic stem cells^{13,14}. DNA methylation plays a key role in many biological processes such as regulation of tissue-specific gene expression, suppression of parasite DNA in the genome, X-chromosome inactivation, and genomic imprinting¹². Genomic imprinting is an epigenetic phenomenon in which a subset of genes, known as imprinted genes, are transcribed monoallelically in a parental-origin-dependent manner¹⁵. The transcriptional asymmetry of the parental alleles is usually directed by allele-specific DNA methylation (ASM) at imprinted loci established during gametogenesis¹⁵. Proper allele-specific DNA methylation and allele-specific expression of imprinted genes play important roles in the regulation of embryonic and neonatal growth, placental function, postnatal behavior, and metabolism¹⁶. Altered gene expression and DNA methylation at imprinted loci have been associated with congenital overgrowth disorders such as BWS in human¹⁷ and LOS in bovine^{5,18}.

Most BWS cases are sporadic and have been linked to two imprinted loci on chromosome 11p15.5, the *KCNQ1* locus and the *H19/IGF2* locus⁴. Approximately 50% of the BWS cases are associated with the hypomethylation of KvDMR1 at the *KCNQ1* locus and 2–7% are linked to the hypermethylation of the differentially methylated region (DMR) at the *H19/IGF2* locus⁴. In addition, studies have shown that a subset of BWS patients with epimutation at the *KCNQ1* locus also exhibit abnormal DNA methylation at other imprinted loci¹⁹. We have observed loss of methylation of the KvDMR1 on the maternal allele in LOS⁵ and have reported that LOS is a multi-locus loss-of-imprinting condition in which aberrant imprinted gene expression is associated with tissue-specific loss of imprinted DNA methylation¹⁸.

Although it is well accepted that loss-of-imprinting can contribute to these overgrowth syndromes, and as such have been coined “loss-of-imprinting syndromes”, it remains unknown whether aberrant gene expression and DNA methylation occur at non-imprinted loci and to what extent these molecular alterations contribute to the variable phenotypes observed in these conditions. To address this question, we examined the transcriptome of skeletal muscle, liver, kidney, and brain of four control and four LOS day ~105 *Bos taurus indicus* (*B. t. indicus*) × *Bos taurus taurus* (*B. t. taurus*) F₁ fetuses¹⁸. We found that different LOS fetuses exhibit different numbers of DEGs and that each tissue within LOS fetuses have perturbations of distinct gene pathways. Notably, in skeletal muscle, multiple pathways involved in myoblast proliferation and fusion into myotubes are misregulated in LOS fetuses. Further, characterization of the DNA methylome of skeletal muscle revealed numerous local methylation differences between LOS and controls; however, very few DEGs could be linked to neighboring identified DMRs. This study indicates that global misregulation of non-imprinted genes in addition to loss-of-imprinting characterizes the ART-induced overgrowth syndrome. The observation that most identified DMRs could not be directly associated with aberrant gene expression suggests that caution should be exercised when making conclusions about the etiology of such syndromes by interpreting time-point-specific DNA methylation data.

Results

Identification and characterization of DEGs in LOS fetuses. In order to determine to what extent the transcriptome is altered in bovine LOS fetuses, we analyzed RNA sequencing (RNAseq) data of skeletal muscle, liver, kidney, and brain from four control and four LOS day ~105 (d104–106) *B. t. indicus* × *B. t. taurus* F₁ fetuses that we generated in a previous study¹⁸ (Supplementary Table 1). These tissues were selected because they represent the three primary germ layers (*i.e.*, liver for endoderm, kidney and skeletal muscle for mesoderm, and brain for ectoderm). Only females were used in this study to avoid any potential sex-specific gene expression²⁰. The average bodyweight for control and LOS fetuses was 405 g [standard deviation (SD) = 10 g, range = 24 g] and 592 g (SD = 95 g, range = 200 g), respectively ($p = 0.008$) (Supplementary Table 1). The bodyweight differences between the control and LOS fetuses are not expected to be due to paternal effects because all fetuses were sired by one Nelore bull [ABS CSS MR N OB 425/1 677344 29NE0001 97155 (*i.e.*, *B. t. indicus*)]⁵.

As LOS fetuses exhibited dramatic difference in bodyweight, each LOS fetus was considered individually and was compared to the mean of the four controls to identify DEGs. To ensure the validity of grouping four control fetuses for DEG identification, principle component analyses (PCA) were performed using the normalized RNAseq read counts. For the RNAseq the expression libraries of kidney, skeletal muscle, and liver, control fetuses clustered while the libraries from the LOS fetuses did not (Supplementary Fig. S1). It should be noted that the RNAseq libraries of skeletal muscle and liver from Control #2 fetus were sequenced with 50 bp read length; while the RNAseq reads of other fetuses were of 100 bp. The read length differences may explain why Control #2 is segregated from other control fetuses in these two tissue types. In brain, both control and LOS fetuses exhibited a segregated pattern in the PCA plot (Supplementary Fig. S1). In order to minimize the false positive DEGs caused by natural biological variation and technical variation, the DEGs were identified with the consideration of variance among the controls and the RNAseq read length differences using the edgeR package as we previously described^{18,21,22}. The number of DEGs varied between tissues for different LOS fetuses (Supplementary Fig. S2). For example, in the largest LOS fetus analyzed (weight = 714 g), the number of DEGs was 3868, 1581, 441, and 276 for skeletal muscle, liver, kidney, and brain, respectively, while in the smallest LOS fetus (weight = 514 g), the number of DEGs for those tissues were 2270, 3232, 139, and 1106 (Supplementary Fig. S2). The number of DEGs was not associated with fetal weight, but the fetuses with loss-of-imprinting at the *KCNQ1* locus (*i.e.*, LOS #1 and #4) had more DEGs in liver and skeletal muscle than LOS fetuses with correct imprinting at this locus (Supplementary Fig. S2).

Next, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses²³ to gain biological insights of the DEGs identified in the LOS fetuses. We found that several KEGG pathways were shared by DEGs in different tissues while others were unique for a particular tissue (Fig. 1A and Supplementary Fig. S2). For

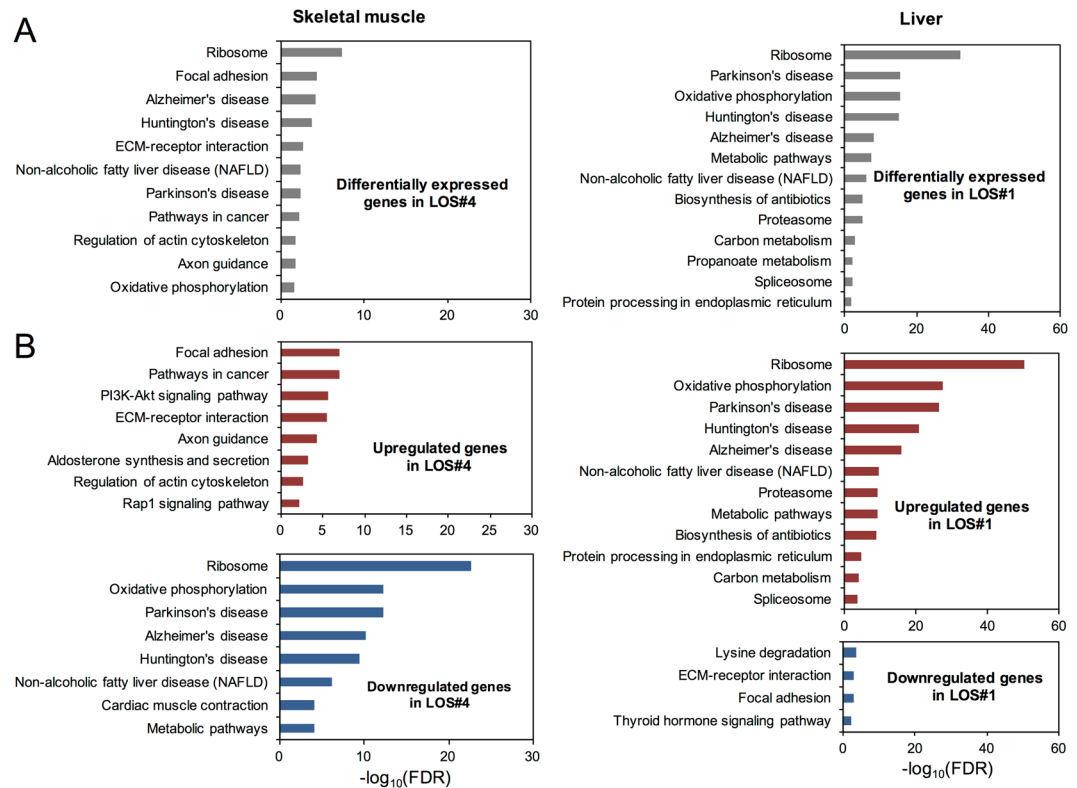


Figure 1. KEGG pathway analyses of DEGs in somatic tissues of LOS fetuses. (A) KEGG pathways enriched for DEGs in skeletal muscle and liver (FDR < 0.05). Shown here are examples of tissues with the most DEGs. (B) KEGG pathways enriched for upregulated and downregulated genes in skeletal muscle and liver (FDR < 0.05). DEG: differentially expressed genes; LOS: large offspring syndrome.

example, “ribosome” and “oxidative phosphorylation” were enriched in DEGs in skeletal muscle, liver, and kidney, while “pathways in cancer” was unique for skeletal muscle (Fig. 1A and Supplementary Fig. S2), suggesting the over-proliferation of the muscle cells in LOS fetuses. Further analyses revealed that even the common pathways could be differentially represented in different tissues. For instance, “ribosome” and “oxidative phosphorylation” were enriched with downregulated genes in skeletal muscle but enriched with upregulated genes in liver and kidney (Fig. 1B and Supplementary Fig. S2). For brain, however, misregulated genes were only found to enrich two KEGG pathways (Supplementary Fig. S2). Overall, we found that numbers of DEGs varied among LOS fetuses and different tissues within LOS fetuses have perturbations of distinct gene pathways.

Multiple pathways involved in myoblast proliferation and fusion into myotubes are misregulated in LOS. To identify DEGs correlated with the overgrowth phenotype of the LOS fetuses, weighted gene correlation network analyses (WGCNA)²⁴ were used to identify clusters of genes that were highly interconnected, namely modules, and to determine which modules were correlated with bodyweight of the fetuses. For all the tissues analyzed, three modules [*i.e.*, turquoise, red, and green (Supplementary Fig. S3)] in skeletal muscle were identified to be significantly correlated (p-value < 0.05) with bodyweight and were enriched with the KEGG pathways associated with overgrowth phenotype (Fig. 2A,B). In both turquoise and red clusters, the enriched pathways were involved in cell proliferation and differentiation (*e.g.*, “pathways in cancer” and “cell cycle”), and cell-cell adhesion and fusion (*e.g.*, “focal adhesion”, “cell adhesion molecules”, and “adherens junction”) (Fig. 2A,B), while the green module was only enriched with genes associated with the “spliceosome” pathway (Supplementary Fig. S3).

Fetal myogenesis involves myoblast proliferation/differentiation and fusion of myoblasts into multinucleated muscle fibers²⁵. Characterization of genes that belong to “pathways in cancer” led us to identify multiple upregulated genes in LOS fetuses that are known to be involved in myoblast proliferation and differentiation (Fig. 2C). For example, IGF1 is a major regulator for skeletal muscle growth and both *IGF1*²⁶ and its downstream signaling component *PIK3CD*²⁶ exhibited increased expression levels in LOS #3 and/or #4, the two largest fetuses. Besides *IGF1*, transcript abundance of several other growth factors and their receptors such as *FGF1*, *FGF18*, *FGFR1*, *FGFR2*, *PDGFB*, *PDGFRA*, *PDGFRB*, *TGFB1*, and *TGFB2* were also increased in LOS #3 and/or #4 (Fig. 2C). These growth factors may work in synergy to promote myoblast proliferation and differentiation²⁵. In addition, numerous upregulated genes in LOS that belong to the “focal adhesion” and the “cell adhesion molecules” pathways have been reported to play a role in myoblast adhesion and fusion. For example, we found *DOCK1*, a prototypical member of the family of Rho GTPase activator that is essential for myoblast fusion²⁷, to be upregulated in both LOS #3 and #4 (Fig. 2C). Further, a number of extra cellular matrix genes, reported to

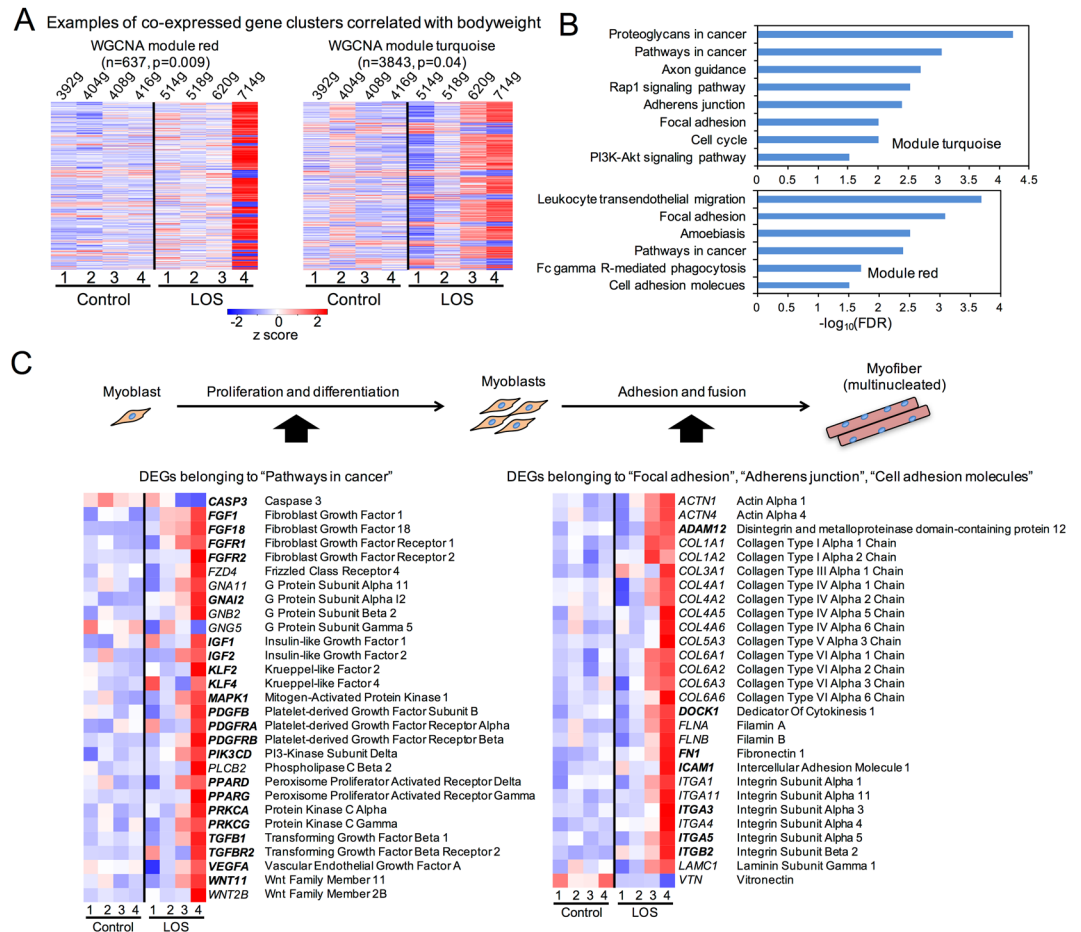


Figure 2. Multiple pathways involved in myogenesis are misregulated in skeletal muscle of LOS fetuses. **(A)** Heat map profiling of the genes of the WGCNA network module red and turquoise. Red and turquoise modules represent clusters of co-expressed genes that are correlated with bodyweight of the fetuses (details may be found in Fig. S2). Shown in the parentheses are the number of genes in each module and p-value of the correlation between the module and bodyweight. The z-score scale represents the mean-subtracted regularized log-transformed read counts. **(B)** Enriched KEGG pathways for module turquoise and red (FDR < 0.05). **(C)** Heat map profiling of the genes that belong to the "pathways in cancer" and the "focal adhesion/adherens junction/cell adhesion molecules" pathways. Bolded genes are those known to be involved in myoblast proliferation and differentiation or fusion of myoblasts into myotubes.

be involved in myoblast migration, adhesion, and fusion (e.g., *ADAM12*, *FNI*, *ICAM1*, *ITGA3*, and *ITGA5*)^{28,29} exhibited increased transcript abundance in LOS #3 and/or #4 (Fig. 2C). Although it is likely that some of the gene expression changes could be the consequential events of the altered fetal myogenesis, these data suggest that the misregulated pathways may be responsible for the increased muscle mass and overgrowth phenotype of the LOS fetuses. In summary, gene co-expression network analyses revealed that multiple pathways involved in fetal skeletal muscle development were disrupted in LOS.

Whole genome bisulfite sequencing (WGBS) and read processing. As differential gene expression can be regulated by differential DNA methylation¹² and DNA methylation can be altered by the use of ART⁷⁻¹⁰, we next sought to address whether altered DNA methylation could be the molecular mechanism responsible for the misregulated gene pathways in skeletal muscle of LOS fetuses. To compare the methylome between control and LOS, we generated WGBS data of skeletal muscle of the fetuses used for RNAseq analyses. For each individual, we sequenced ~400 million 100 bp paired-end reads, which is approximately 26 × coverage of the bovine reference genome (Supplementary Table S1). The average bisulfite conversion rate was 99.31% (Supplementary Table S1). Of note, strong methylation bias was detected at the 5' end of read mate 1 and both ends of read mate 2 (Supplementary Fig. S4). As these biases could be introduced during the library preparation and/or sequencing steps³⁰, the first 3 base pairs (bp) and last 2 bp of each WGBS read were trimmed and not used for determination of CpG methylation levels. Following a series of quality filtering steps, ~80% of the CpGs in the reference genome were covered by ≥ 5 WGBS reads (Supplementary Table S2).

Genome-wide CpG methylation landscape and its relationship with transcriptome in fetal skeletal muscle. Prior to making comparisons between LOS and controls, we sought to establish the DNA

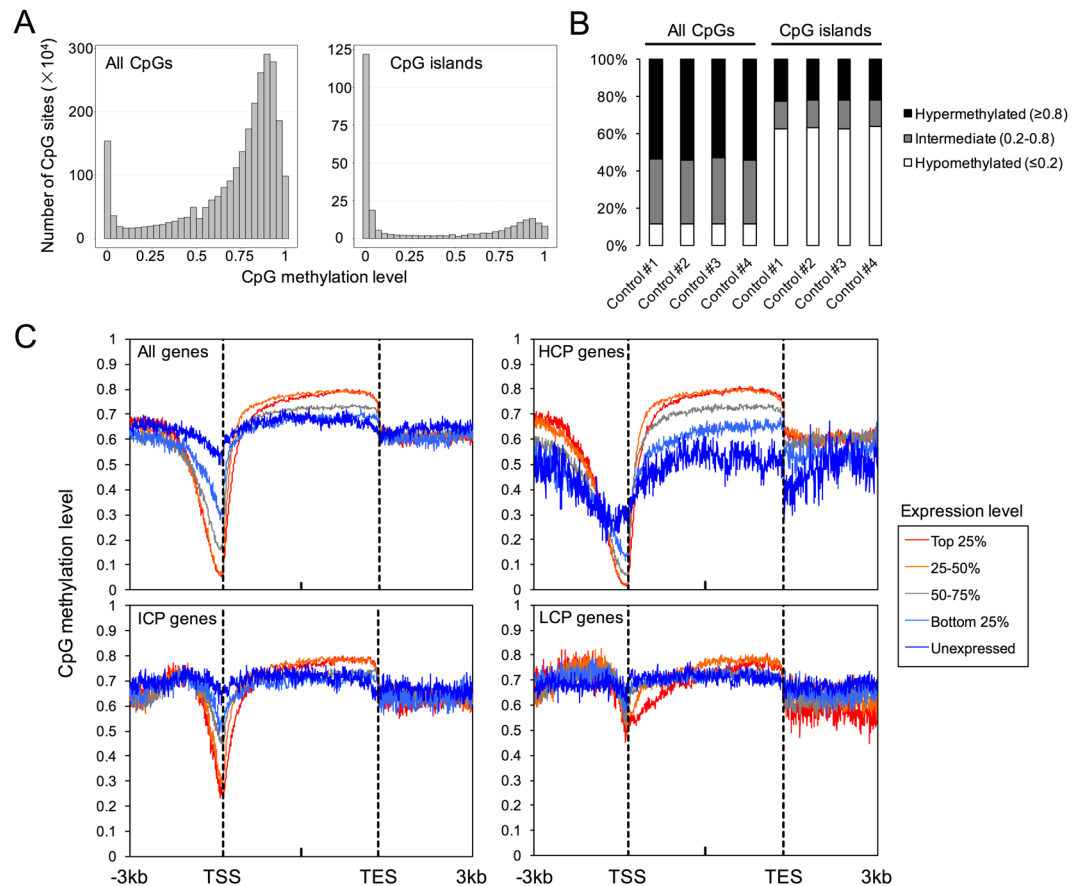


Figure 3. Genome-wide CpG methylation landscape and its relationship to the transcriptome in day ~105 skeletal muscle of control fetuses. **(A)** Distribution of CpGs sites across different methylation levels. The methylation level of each CpG site was calculated as the ratio of the “C” reads to the sum of the “C” and “T” reads. CpGs from all four controls were combined for the generation of this plot. **(B)** Percent of CpGs at each methylation level in control fetuses. **(C)** Averaged DNA methylation of annotated bovine genes along the gene body and 3 kb upstream of the TSS and 3 kb downstream of the TES. Genes were classified into five groups based on their expression levels. HCP: high CpG density promoter; ICP: intermediate CpG density promoter; LCP: low CpG density promoter; TSS: transcription start site; TES: transcription end site.

methylation baseline in bovine fetal skeletal muscle. In controls, the distribution of CpG sites across different methylation levels exhibited a bimodal pattern (Fig. 3A), which is similar to what has been observed in somatic tissues of human and mouse³¹. Further, consistent with the reports that CpG islands are usually unmethylated³¹, the majority of the CpGs within CpG islands in our study were hypomethylated (Fig. 3A,B). In addition, we explored the relationship between the DNA methylome and the transcriptome. We first plotted the average methylation of all the annotated genes along the gene body and 3 kb upstream of the transcription start site (TSS) and 3 kb downstream of the transcription end site (TES). The 3 kb range was chosen because the average DNA methylation beyond this distance was ~0.6–0.7, which was the overall average methylation of bovine fetal skeletal muscle. As shown in Fig. 3C, we found a negative correlation (Spearman rank test, $p < 0.02$) between the gene expression level and the level of DNA methylation at regions near the TSS (–500bp to TSS) (Fig. 3C). Furthermore, we queried the relationship between DNA methylation and transcript abundance for genes with different levels of CpG density at their promoter regions. For genes with high (HCP) and intermediate (ICP) CpG density promoters, there was a negative correlation (Spearman rank test, $p < 0.02$) between the methylation level at promoter regions and the transcript abundance (Fig. 3C). However, low CpG density promoters (LCP) were usually methylated and had no significant correlation with gene expression levels (Fig. 3C). Notably, we observed a positive correlation (Spearman rank test, $p < 0.02$) between DNA methylation at gene bodies and transcript abundance, especially with HCP genes (Fig. 3C). Overall, in our study, the associations between global DNA methylation and global gene expression were consistent with the findings in human and mouse³¹.

Identification and characterization of the ASM regions in bovine fetal skeletal muscle. As only a limited number of imprinted ASM regions have been characterized in bovine^{5,18,32–34}, one objective of our study was to identify ASM regions at base resolution in bovine fetal skeletal muscle using WGBS. To distinguish the parental origins of the alleles in the *B. t. indicus* × *B. t. taurus* F₁ progenies, we obtained heterozygous SNPs of the fetuses from two sources: 1) SNPs identified from the WGBS data using Bis-SNP³⁵ and 2) SNPs identified from the RNAseq data of the F₁ hybrids³³ (Supplementary Fig. S5). We assigned the WGBS reads overlapping

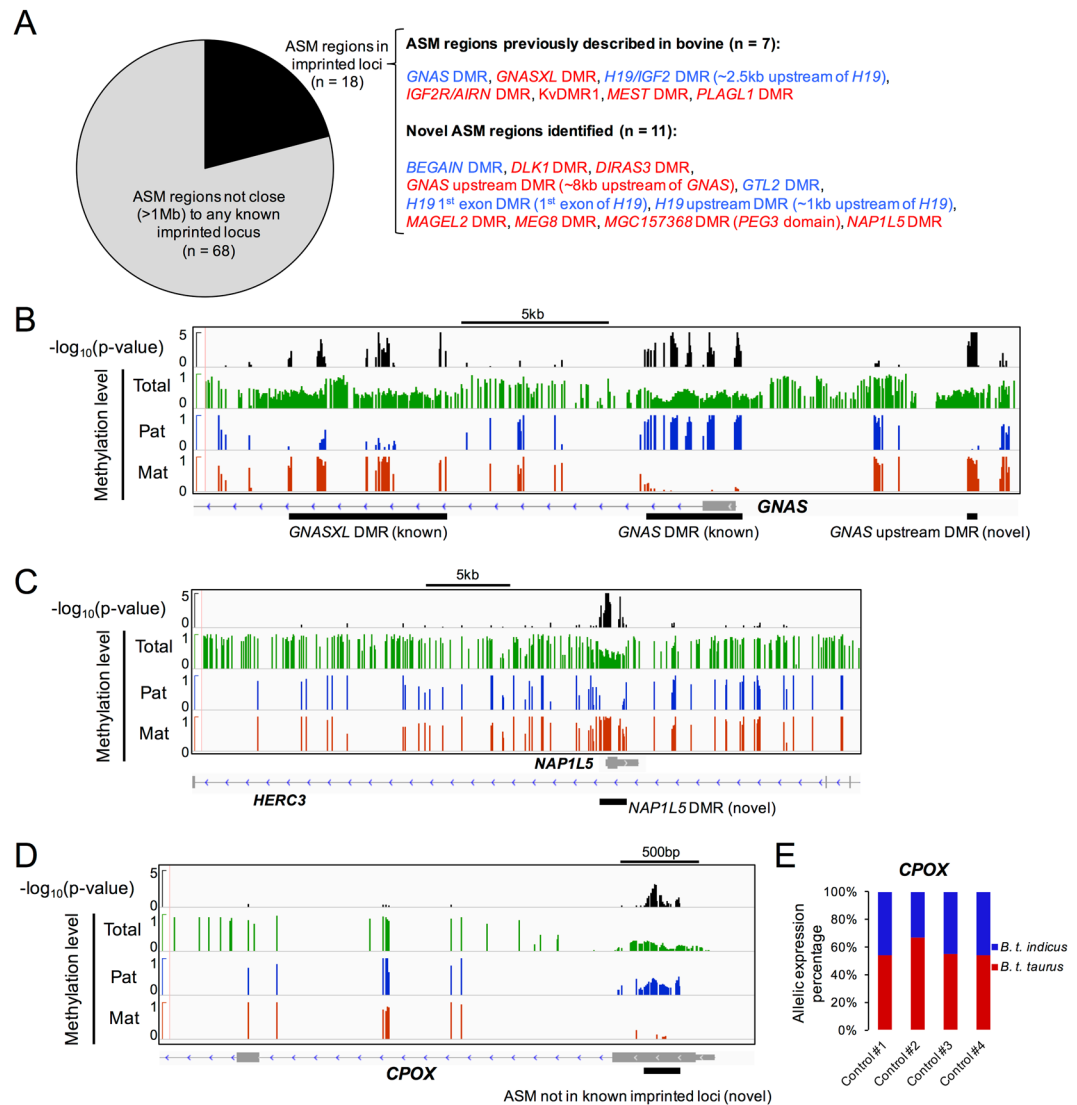


Figure 4. Identification and characterization of ASM regions in skeletal muscle of control fetuses. (A) Summary of the ASM regions identified at imprinted and non-imprinted loci. Red color: DMRs with maternal allele-specific methylation; blue color: DMRs with paternal allele-specific methylation. (B–D) Genome browser views of the imprinted ASM loci: *GNAS/GNASXL* and *NAP1L5/HERC3* and the non-imprinted *CPOX*. The $-\log_{10}(\text{p-value})$ for ASM significance (black) and the CpG methylation levels for total (green), *B. t. indicus* allele (blue, paternal), and *B. t. taurus* allele (red, maternal) are also shown. Each bar represents a single CpG site. (E) Allelic expression percentage of *CPOX* in skeletal muscle of the control fetuses. ASM: allele-specific DNA methylation; DMR: differentially methylated region. More examples may be found in Figs S5–S6.

the SNPs to their parental origins based on the genotype of the *B. t. indicus* sire³³ of the F_1 fetuses. As an initial step to screen CpGs with ASM, allelic WGBS reads were combined from the four controls and CpGs with at least $4 \times$ coverage of each allele ($n = 4,798,414$; 17.4% of CpGs in the bovine genome) were subject to Fisher's exact test to determine whether DNA methylation level was significantly different between alleles. In total, 109,794 ASM candidate CpGs were identified with a p-value cutoff of 0.01 [false discovery rate (FDR) \approx 0.05, Supplementary Fig. S5]. To further minimize the false positive ASM sites, the ASM candidate CpGs were merged into regions using a clustering method as previously described³⁶. In total, 1,070 ASM CpGs were grouped into 86 discrete genomic regions (Fig. 4A and Supplementary Table S3).

We manually annotated the identified 86 ASM regions (Supplementary Table S3) and found that 18 of them were located at known imprinted loci, while others were located >1 Mb away from any known imprinted gene in bovine (Fig. 4A). The ASM regions >1 Mb from known imprinted genes are referred to as the “non-imprinted ASM” in subsequent text. The two types of ASM regions are of comparable size; however, the imprinted ASMs have a higher density of CpGs (Supplementary Fig. S6) located within CpG islands. Both imprinted and non-imprinted ASM loci have overall DNA methylation level around 0.5, but ASM regions at imprinted loci exhibit less variation (range = 0.43 [imprinted] vs 0.8 [non-imprinted] and SD = 0.11 [imprinted] vs. 0.19 [non-imprinted]; Supplementary Fig. S6). Lastly, the methylation differences between the parental alleles for the

imprinted ASM regions are greater when compared to the non-imprinted ASM regions (mean = 0.73 [imprinted] vs. 0.51 [non-imprinted]; Supplementary Fig. S6).

Of the 18 imprinted ASM regions, seven have been previously described in bovine by us and others. These include the paternal DMRs: *GNAS* DMR, *H19/IGF2* DMR^{33,34} and the maternal DMRs: *GNASXL* DMR, *IGF2R/AIRN* DMR, *KvDMR1*, *MEST* DMR, and *PLAGL1* DMR^{18,32–34} (Fig. 4B and Supplementary Fig. S7). Of the novel imprinted ASM regions identified in bovine fetal skeletal muscle, eight of them, including *NAPIL5* DMR, *DIRAS3* DMR, *GTL2* DMR and *DLK1* DMR (Fig. 4C, Supplementary Fig. S7), have been previously reported in human and/or mouse (Supplementary Table S4). The other three novel ASM regions namely, *BEGAIN*, *MEG8*, and *MGC157368* DMRs, were located within the gene bodies of the known imprinted genes in bovine (Supplementary Table S3). All the ASM regions described above could be associated with the nearby imprinted transcriptional activities previously reported by us³³ with the exception of *MEST* DMR, for which no heterozygous SNP was identified to enable allelic gene expression analyses.

We also manually inspected the imprinted ASM regions previously described in human and/or mouse in our bovine WGBS data. We found that five ASM regions (*i.e.*, *GRB10* DMR, *H13* DMR, *NNAT* DMR, *PEG3* DMR, and *PEG13* DMR) were absent in this study due to either poor SNP coverage or low read coverage (Supplementary Table S4). A less stringent criterion (*i.e.*, $3 \times$ coverage per allele instead of $4 \times$) for ASM calling could lead to the identification of another three DMRs (*i.e.*, *INPP5F* DMR, *PEG10* DMR, and *SNRPN* DMR) (Supplementary Table S4). Lastly, we found that some imprinted DMRs in human and/or mouse were simply not subject to ASM in bovine fetal skeletal muscle as these loci were either hypomethylated (*e.g.*, *RASGRF1* DMR, *ZRSR1/IMPACT1* DMR, and *COMMD1* DMR) or partially methylated on both parental alleles (*e.g.*, *MKRN3* DMR and *GPR1/ZDBF2* DMR) (Supplementary Table S4).

For the 68 ASM regions not adjacent to any known imprinted locus, we sought to identify genes subject to allele-specific expression in the nearby regions. As shown in Supplementary Table S3, of the 68 ASM regions, 20 were intergenic (>5 kb away from any annotated gene) while the other 48 regions could be associated with nearby genes. For the 48 genes, parental-allele-specific read counts for 24 genes could be obtained from our previous study³³. We found that all of these genes exhibited biallelic gene expression and had no association with the nearby ASM regions (shown in Fig. 4D,E, and Supplementary Fig. S8 are five examples). The remaining 24 genes were either lowly expressed in skeletal muscle or had no heterozygous SNPs available to ascertain the parental alleles. It is possible that some of these ASM regions are associated with tissue-specific or transcript-isoform-specific monoallelic gene expression.

Lack of direct associations between DMRs and DEGs in LOS. As for the DEG analysis, each LOS fetus was analyzed independently and compared to the average of all four controls to identify DMRs using the Bseq R package³⁰. The hierarchical clustering of CpG methylation within DMRs showed that control fetuses clustered together and were separated from the LOS fetuses (Fig. 5A). The numbers of DMRs identified were similar for the four LOS fetuses with most DMRs identified in LOS #1 (Fig. 5B). Notably, we identified loss of DNA methylation of *KvDMR1* in LOS#1 and #4 (Fig. 5C), indicating the validity of the statistical method used for DMR determination, as we have previously identified these specific epimutations by the bisulfite polymerase chain reaction (PCR), cloning, and sequencing⁵. However, our previous study reported a single DMR with a size of 385 bp (37 CpGs) while the WGBS analyses allowed us to identify methylation differences between control and LOS across the entire CpG island (1957bp, 207 CpGs, UMD3.1 Chr29: 49,552,845–49,554,801). Further characterization of the DMRs revealed that most of them contained less than 20 CpGs and were of methylation differences between 0.2–0.5 (Fig. 5D).

We then characterized the genomic distances between the identified DMRs and the DEGs, in order to determine whether any of the DMRs could be linked to aberrant gene expression in LOS. We first queried the genomic regions within 5 kb range of the DEGs and found that on average only ~1.5% of DEGs in each LOS fetus could be associated with the identified DMRs (Fig. 6A). Increasing the range queried from 5 kb to 20 kb only added an additional ~0.5% of DEGs that associate with the identified DMRs (Fig. 6A). The locations of the identified DMRs were randomly permuted 1000 times to determine whether the numbers of DEGs that are associated with the DMRs were greater than expected by chance. For each dataset, DEGs were considered to be associated with altered DNA methylation if DMRs were within 5 kb or 20 kb range of the DEGs. The expected counts of DEGs were calculated as the average of the 1000 randomized datasets. As shown in Fig. 6A, DEGs are associated with DMRs more often than expected by chance in LOS #1, #3, and #4 but not in LOS #2.

We found that 1407 DEGs are common between LOS #3 and #4 (the two largest LOS fetuses) and we sought to determine if differential DNA methylation can explain misexpression of these genes (Fig. 2C). Only 14 DMRs overlapped between LOS #3 and #4, and only one of these was located within 5 kb of one common DEG (*C5H12orf45*) and another within 20Kb (*RWDD1*). We also noted that not all LOS fetuses with aberrant DNA methylation at the same locus had differential gene expression of the associated gene (an example shown in Fig. 6B,C), suggesting that differential DNA methylation is not the sole cause of differential gene expression. Lastly, minor methylation differences were detected in the TSS regions (–1.5 kb to +500 bp) of the DEGs between control and LOS (Fig. 6F and Supplementary Fig. S9). In summary, only a very small percent DEGs in fetal skeletal muscle could be directly associated with the DMRs between control and LOS.

Association of aberrant DNA methylation and differential gene expression at *IGF2R/AIRN* and *MAGEL2* imprinted loci. Insulin-like growth factor 2 receptor (IGF2R) is a scavenger receptor of the fetal growth factor IGF2, and is maternally expressed in cow^{18,37} (Fig. 7A) and mouse³⁸, but exhibits polymorphic imprinting in human³⁹. In this study, *IGF2R* exhibited decreased transcript abundance in three out of four LOS fetuses (Fig. 7C), and this differential gene expression was associated with the altered DNA methylation at DMR1 and DMR2 of the *IGF2R* locus (Fig. 7A–B, Supplementary Fig. S9). In all four LOS fetuses, we found loss of

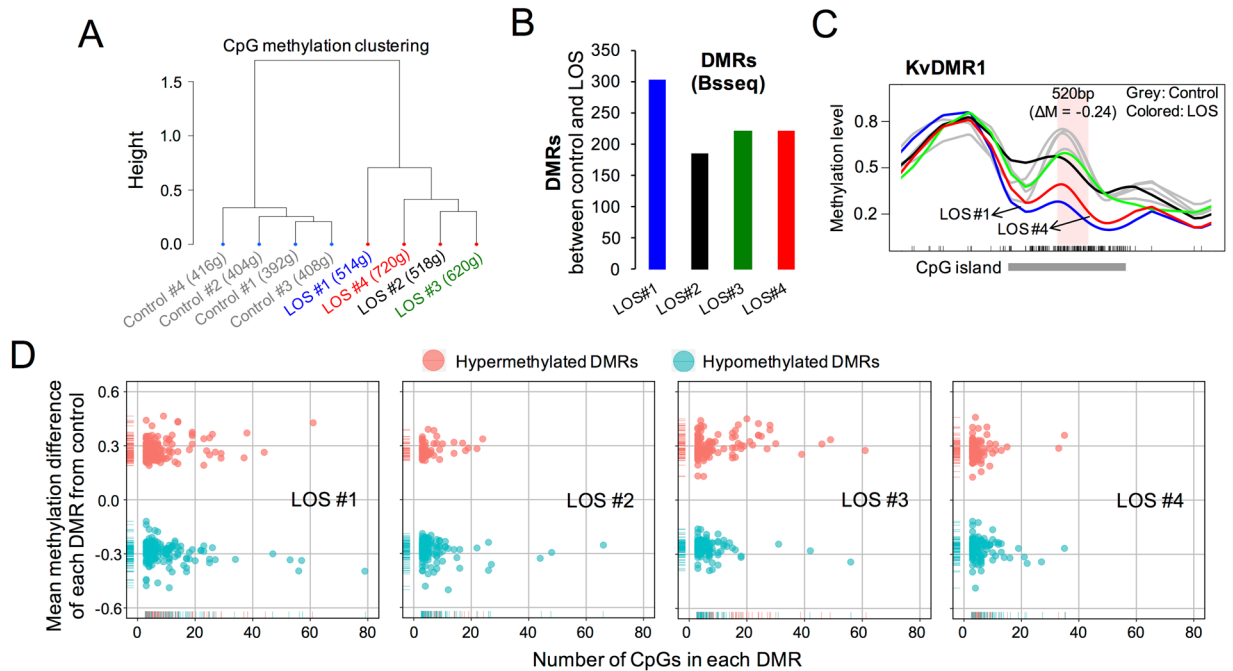


Figure 5. Identification and characterization of DMRs in fetal skeletal muscle between control and LOS. **(A)** Hierarchical clustering of CpG methylation within DMRs between control and LOS using 1-Pearson's correlation distance. Bodyweight for each fetus is also shown. **(B)** Number of DMRs identified by the Bsseq R package between each LOS fetus and all four controls. **(C)** Smoothed methylation profiles of the F₁ fetuses at KvDMR1 with the pink area representing the DMR identified (size = 520 bp; ΔM = mean methylation difference of the DMR between all four LOS and the controls; grey = controls; blue = LOS #1; black = LOS #2; green = LOS #3; red = LOS #4). Ticks at bottom represent CpG sites. **(D)** Number of CpGs and the mean methylation difference of each DMR identified in each LOS fetus.

methylation of DMR2, which is located in the second intron of *IGF2R* with loss of methylation in all four LOS fetuses (Fig. 7B). Normally, DMR2 was methylated on the maternal allele and unmethylated on the paternal allele. Hypomethylation of DMR2 has also been reported in sheep LOS⁴⁰, suggesting the *IGF2R* locus is similarly misregulated in these two species. DMR1, which is close to the promoter region of *IGF2R*, is normally partially methylated on the paternal allele and unmethylated on the maternal allele. In our study, DMR1 was hypermethylated in LOS (Fig. 7A,B).

In mice, paternal silencing of *IGF2R* is mediated by the paternally expressed long non-coding RNA *AIRN*, which is transcribed in the antisense direction to *IGF2R* and is regulated by DMR2 of *IGF2R*⁴¹. Thus, we reasoned that hypomethylation of DMR2 can lead to biallelic expression of *AIRN*, which represses the expression of *IGF2R* on both alleles and results in the decreased level of *IGF2R* in LOS fetuses. However, contrary to our expectation, *AIRN* was biallelically expressed in both control and LOS fetal skeletal muscle (Fig. 7D), suggesting that down-regulation of *IGF2R* is not mediated by biallelic expression of *AIRN*. Nevertheless, decreased transcript abundance of *IGF2R* was associated with a hypomethylated DMR2 and hypermethylated DMR1 in LOS fetuses.

In addition to *IGF2R*, we detected an association between differential gene expression and differential DNA methylation at the *MAGEL2* locus. *MAGEL2* is a paternally expressed gene¹⁸ with its promoter methylated on the maternal allele but unmethylated on the paternal allele (Supplementary Table S3). As shown in Fig. 7E,F, the increased transcript abundance of *MAGEL2* in LOS #4 is associated with the loss of DNA methylation at its promoter region.

Discussion

Since BWS was first described in the early 1960s⁴² and the first cases of ART-induced LOS were reported in 1995⁴³, the main focus in both fields has been to understand how loss-of-imprinting alone can contribute to the complex and variable phenotypic features of these overgrowth conditions. In human, the epimutation and genetic alteration of the two imprinted loci (*i.e.*, *KCNQ1* and *H19/IGF2* loci) affect nearly 70% of BWS patients with the hypomethylation at KvDMR1 accounting for ~50% of all BWS cases¹⁹. In addition, loss-of-imprinting at multiple loci was observed in ~25% of BWS cases with epimutation at the *KCNQ1* locus¹⁹. Similarly in bovine, we have also shown that LOS exhibits loss-of-imprinting at multiple loci including the *KCNQ1* locus^{5,18}. However, to date, there has been no report of molecular features that can consistently and reliably diagnose and predict these syndromes and the associated variable phenotypes. Here we advance the field by describing the altered transcript abundance of genes beyond imprinted loci in LOS fetuses, showing that loss-of-imprinting is not be the only molecular lesion of this condition.

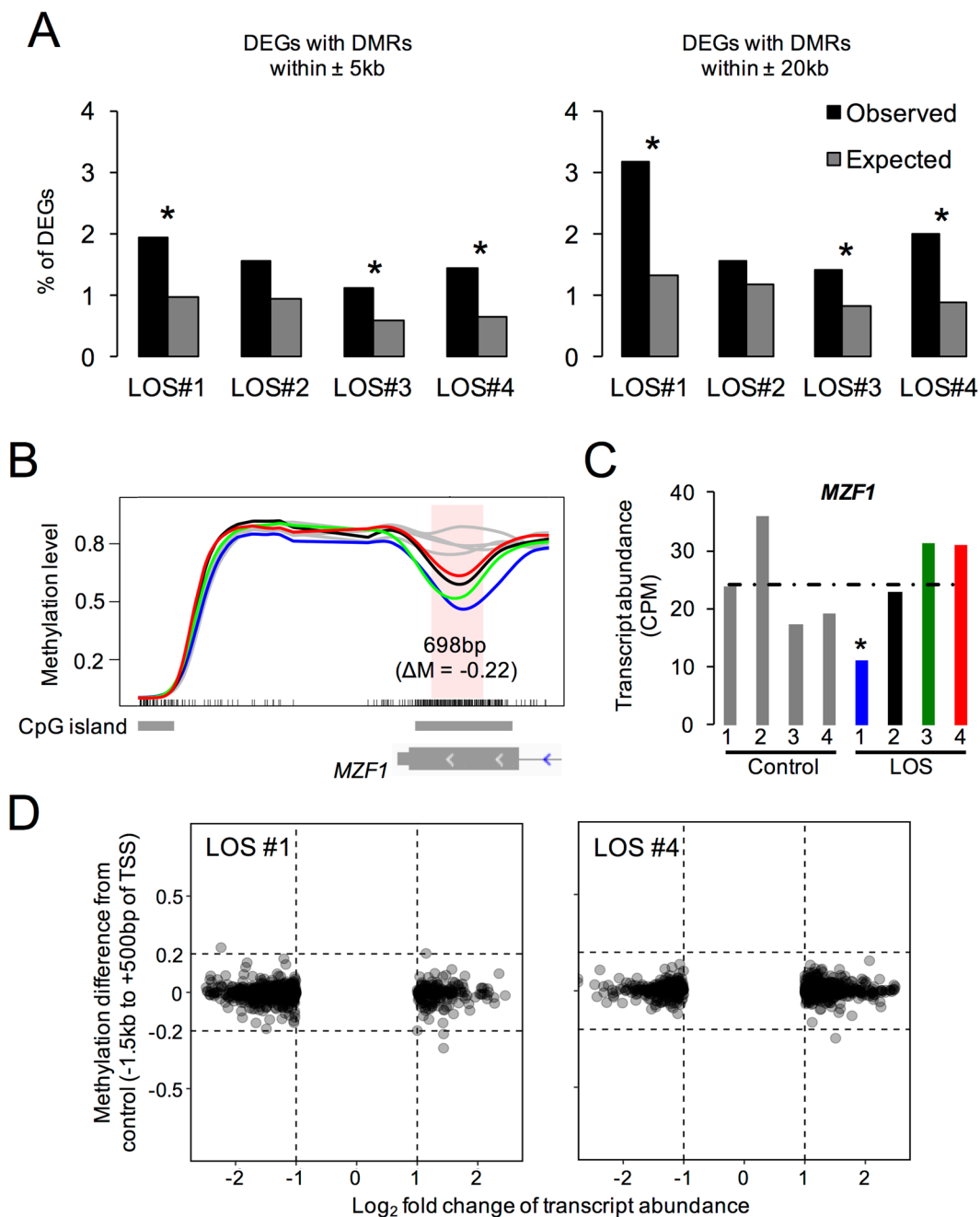


Figure 6. Lack of associations between the DEGs and the DMRs in skeletal muscle of the LOS fetuses. **(A)** Percentage of DEGs that are associated with the DMRs in LOS fetuses. A DMR is considered to be associated with a DEG if it is less than 5 kb or 20 kb away from the DEG. To determine whether the number of DEGs associated with DMRs were greater than expected by chance, DMRs were shuffled within each chromosome 1000 times and the expected numbers of DEGs associated with the DMRs were calculated as the average of 1000 datasets and were used in chi-square tests (* $p < 0.01$). **(B,C)** Examples of lack of association between DEGs and DMRs when the DMRs are within 5 kb range of the DEGs. For **(B)**, the details may be found in Fig. 5. Panel **(C)** shows the transcript abundance of *MZF1* in all fetuses (asterisk: edgeR FDR < 0.05). DEG: differentially expressed genes. **(D)** Mean methylation level differences of the DEGs between control and LOS fetuses at the TSS region (-1.5 kb to +500 bp). Only DEGs that had at least 2-fold change are shown here.

In this study, WGCNA and KEGG pathway analyses identified that multiple pathways involved in myoblast proliferation and fusion of myoblasts into myotubes are misregulated in skeletal muscle of LOS fetuses. This finding is consistent with the previous report that short-term culture of ovine embryos can affect myogenic programming and the altered myogenesis may contribute to the large muscles observed in the oversized fetuses⁴⁴. In addition to skeletal muscle, altered transcript abundance of genes was also detected in liver, brain, and kidney of the LOS fetuses. Interestingly, tissue specific alterations in transcript abundance varied between LOS fetuses.

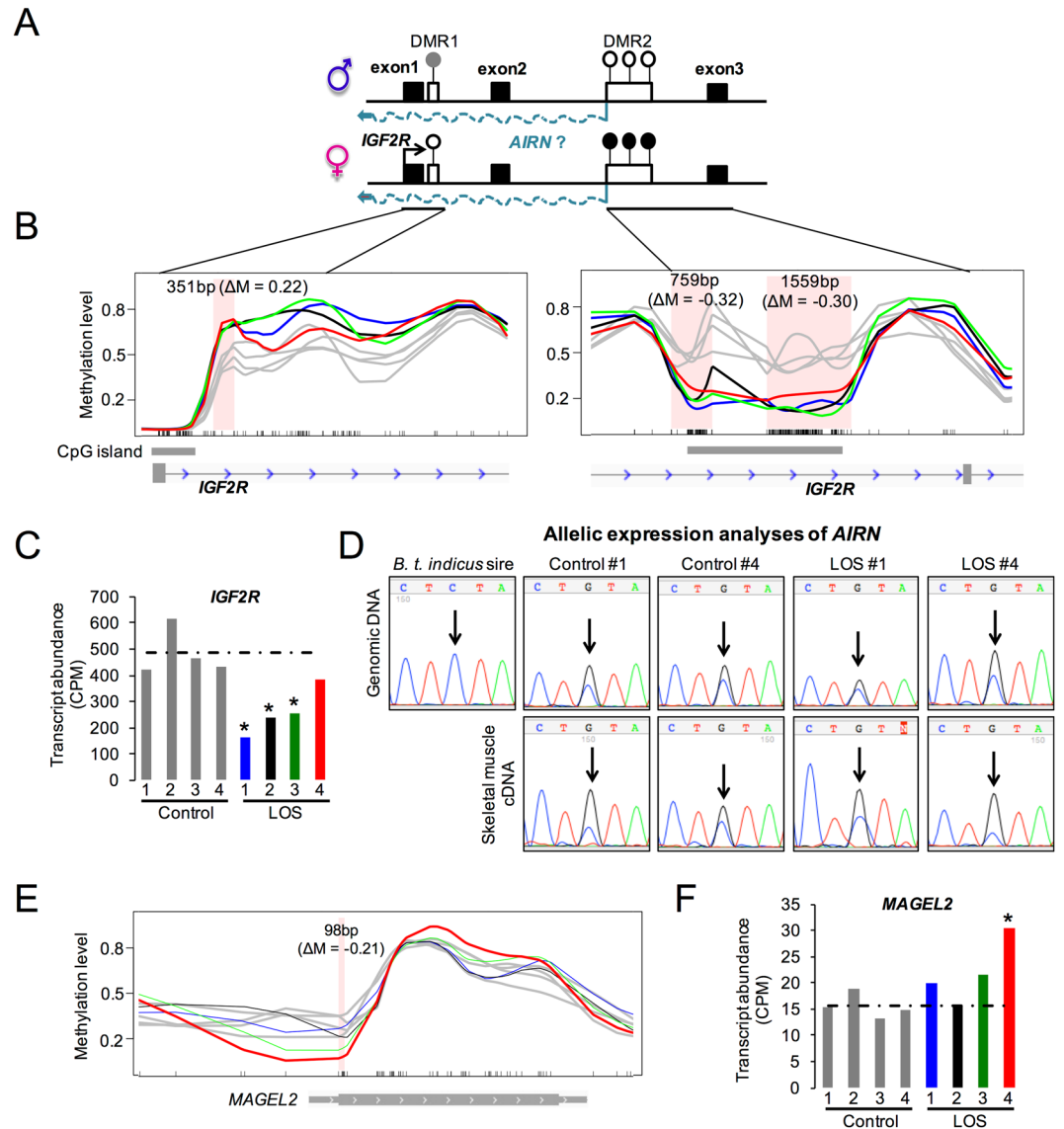


Figure 7. DNA methylation and differential gene expression at *IGF2R/AIRN* and *MAGEL2* imprinted loci. (A) Schematics of the *IGF2R/AIRN* locus. Bent arrows represent the transcription start sites. White boxes represent the ASM regions (white lollipop = unmethylated; grey lollipop = partially methylated; black lollipop = methylated). (B) Smoothed methylation profiles for the two DMRs at *IGF2R* locus between control and LOS fetuses. The pink areas represent the DMR identified (ΔM = mean methylation difference of the DMR between all four LOS and the controls; grey = controls; blue = LOS #1; black = LOS #2; green = LOS #3; red = LOS #4). Ticks at bottom represent CpG sites. (C) Transcript abundance of *IGF2R* in control and LOS fetuses (asterisk: edgeR FDR < 0.05, LOS vs control). (D) Allelic expression analyses for *AIRN* by Sanger sequencing. The SNP site is indicated by an arrow. (E) Smoothed methylation profile of *MAGEL2* locus. The labels are as panel B. (F) Transcript abundance of *MAGEL2* in the F₁ fetuses (asterisk: edgeR FDR < 0.05).

For example, LOS #1 had the most DEGs in liver, LOS #3 in kidney and LOS #4 in muscle. This observation is supported by findings in BWS in which different individuals have different susceptibility for Wilms tumor of the kidney or hepatoblastoma¹⁹. Future studies with in depth characterization of the histology and the transcriptome of the tissues at various developmental stages of LOS will better elucidate relationships between the altered gene pathways and the potential disturbed development of these organs.

An interesting finding of this study was that only a small percent of DEGs could be linked to the identified DMRs between control and LOS. Recently, it has been shown that histone modifications such as H3K27me3 at distal regions in oocytes can escape epigenetic reprogramming and maintain inheritance during the pre-implantation development⁴⁵. It is possible that molecular marks like H3K27me3 in bovine may act as an epigenetic memory of the use of ART procedures and mediate the differential gene expression and the aberrant development in later stages. Alternatively, the DEGs identified in this study could represent the downstream effect of the aberrant expression of developmentally important genes affected by ART during pre-implantation

development. Lastly, it is also possible that some of the DEGs identified in our study can be attributed to germline interaction(s) between *B. t. indicus* and *B. t. taurus* subspecies. In summary, our study demonstrates that altered transcriptome in LOS skeletal muscle is largely uncoupled to DNA methylome epimutations and we submit that caution should be exercised in interpreting results from DNA methylation data alone when studying the etiology of these overgrowth syndromes.

In addition to providing novel insights on the etiology of LOS, our study characterized the cytosine methylation landscape at base resolution in bovine. Consistent to what has been previously reported for other mammals, CpG islands in bovine are hypomethylated and a negative association between DNA methylation and gene expression levels are observed at HCP and ICP, but not LCP promoters³¹. Furthermore, we have identified and characterized novel ASM at both imprinted and non-imprinted loci in bovine. For the imprinted loci, we have expanded the number of imprinted ASM in bovine by describing several novel DMRs such as *DIRAS3* DMR, *GTL2* DMR, and *NAP1L5* DMR. Moreover, we have shown that some known imprinted DMRs in human and/or mouse (e.g., the *L3MBTL* DMR in human and the primary DMRs *RASGRF1* and *IMPACT* in mouse) are not subject to ASM in fetal bovine skeletal muscle, suggesting species-specific imprinting of these genes. Lastly, we have identified ASM regions at non-imprinted loci which others⁴⁶ have proposed may be the result of *cis*-acting regulatory polymorphisms. It should be noted that ASM regions identified in this study are not comprehensive because we could not assign all WGBS reads covering CpGs to either parental alleles due to the lack of informative SNPs. In future studies, analyses of the F₁ hybrids of different crosses may further improve the identification of ASM CpG sites.

In conclusion, we have shown that misregulation of non-imprinted genes in addition to loss-of-imprinting characterizes the ART-induced overgrowth syndrome and have demonstrated that most of the aberrant gene expression in fetal skeletal muscle is not directly associated with aberrant DNA methylation.

Materials and Methods

An expanded and detailed version of the methodology can be found in the Supplementary file.

RNAseq data analyses. The production of the *Bos taurus indicus* × *Bos taurus taurus* (*B. t. indicus* × *B. t. taurus*) F₁ fetuses and collection of the fetal tissues were previously described by us¹⁸. Briefly, for control fetuses, synchronized *B. t. taurus* (Holstein breed) females were artificially inseminated with semen from one *B. t. indicus* bull (Nelore breed; ABS CSS MR N OB 425/1 677344 29NE0001 97155) and the F₁ fetuses were collected via caesarean section at day ~105 of gestation (104–106). For *in vitro* produced fetuses, *B. t. taurus* cumulus-oocyte complexes were matured and fertilized *in vitro* with the same sire as for controls and embryos that developed to the blastocyst stage were transferred into synchronized *B. t. taurus* recipients. On day ~105, conceptuses were collected as described for the control fetuses. All animal procedures were performed at TransOva Genetics by veterinarians, and all procedures were approved by their animal care and use committee (Protocol number – MRP2010–001) and were conducted in a manner conforming to Trans Ova Genetics policies and procedures and the Guide for the Care and Use of Laboratory Animals. RNAseq data of the control and LOS fetuses were obtained from one of our previous studies and were processed as previously described^{18,33}. Briefly, RNAseq reads were subjected to adaptor removal and quality trimming prior to alignment to the genome. To minimize single reference genome alignment bias^{47,48}, HISAT2 (version 2.0.0)⁴⁹ was used to align RNAseq reads to a diploid genome: the *B. t. taurus* reference genome (UMD3.1) and a pseudo *B. t. indicus* genome³³. The alignment of the RNAseq reads to the diploid genome was then merged to identify: i) the RNAseq reads that were aligned to the same single genomic position in both the *B. t. taurus* reference genome and the pseudo *B. t. indicus* genome and ii) reads that were uniquely aligned to either genome but not aligned to the other. DEGs between each LOS fetus and the mean of all four controls were determined with the consideration of variance among the controls using the edgeR package^{21,22}. In addition, RNAseq read length was included as a covariate for edgeR analyses as the RNAseq reads of liver, skeletal muscle, and brain from Control#2 were sequenced with a different read length compared to other samples (50 bp vs. 100 bp). The WGCNA package²⁴ was used to construct a weighted gene co-expression network for the skeletal muscle of the four control and four LOS fetuses. Further, the identified network modules were correlated with the bodyweight to identify genes in modules most likely responsible for the altered bodyweight. In addition, KEGG pathway analyses for the significant network modules and the DEGs in LOS fetuses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID)²³.

WGBS library preparation and sequencing. Genomic DNA isolated from skeletal muscle of four control and four LOS day ~105 *B. t. taurus* × *B. t. taurus* F₁ female fetuses¹⁸ were subject to WGBS analyses. The WGBS libraries were generated using the NEBNext Ultra DNA library Prep Kit for Illumina (NEB E7370) as per the manufacturer's instructions. Prior to sodium bisulfite mutagenesis, 0.5% unmethylated Lambda DNA was added to the bovine DNA to act as an internal control to monitor the bisulfite conversion rate. Each WGBS library was sequenced using Illumina sequencing technology to generate 100 bp paired-end reads. The raw FASTQ files are publicly available at Gene Expression Omnibus (GEO accession no. GSE93775).

WGBS data analyses. Following adaptor removal and quality trimming, WGBS read pairs were aligned to the bovine reference genome assembly UMD3.1 using Bismark (version: 0.15.0)⁵⁰ with the default parameters. Only uniquely aligned reads were used for subsequent analyses. In order to distinguish the C > T SNPs from C > T substitutions that were caused by sodium bisulfite conversion in the F₁ fetuses, Bis-SNP (version: 0.82.2)³⁵ was used to perform the SNP calling. Only CpG sites that showed consensus CpG context for both parental alleles (i.e., do not overlap any SNPs identified by Bis-SNP) were used for the subsequent analyses. The methylation level was determined using the reads from both forward and reverse strands covering the same symmetric CpG site. The methylation level was calculated as: (number of "C" reads)/(number of "C" reads + number of "T" reads).

In order to identify ASM regions, the WGBS read pairs overlapping SNPs were assigned to their parental origin based on the genotype of the *B. t. indicus* sire³³. Further, the allelic WGBS reads were pooled from four controls to determine the sequencing depth and only CpGs that had at least $4 \times$ coverage of each allele were used for Fisher's exact test. To estimate the FDR, WGBS read pairs overlapping SNPs were randomly assigned to be either *B. t. indicus* or *B. t. taurus* in origin in order to generate a randomized dataset of ASM sites³⁶. As an initial screen, ASM candidate CpGs were identified using a p-value of 0.01, resulting in 109,794 ASM sites with a FDR of ~5% (compared to the 5507 CpGs identified as ASM sites in the permuted datasets). To minimize false positives, the identified ASM candidate sites were clustered into regions as previously described³⁶. MethylKit R package⁵¹ was used to perform the hierarchical clustering of the methylation profiles. The DMRs between each LOS fetus and all four controls were identified using the Bseq R package³⁰.

To estimate the expected number of DEGs associated with DMRs, the identified DMRs were randomly shuffled within each chromosome using "shuffleBed" function from BedTools⁵². The UMD3.1 genome assembly gaps from UCSC genome Browser were excluded from the potential shuffled positions. Shuffling was performed 1000 times and for each dataset, a DEG was considered to be associated with a DMR if the DMR falls into the DEG coordinates that had been extended by 5 kb or 20 kb. The averaged DEG counts of the 1000 datasets were used as the expected count for chi-square test.

Statistical analyses and data visualization. Statistical analyses were implemented with R (<http://www.r-project.org/>). Heat maps were generated with the R function "heatmaps". The averaged methylation level of the annotated genes were generated using deepTools⁵³. For ASM regions, bed files were converted to the bigwig format and visualized as custom tracks in the Integrative Genomics Viewer (IGV)⁵⁴.

Allelic expression analyses of *AIRN*. Both DNA and cDNA of the F₁ fetuses and DNA of the sire were amplified by PCR followed by Sanger sequencing to ascribe the parental origin of the transcript. Nucleic acid isolation, cDNA synthesis, and PCR amplifications were as previously reported¹⁸.

References

- European, I. V. F. M. C. *et al.* Assisted reproductive technology in Europe, 2011: results generated from European registers by ESHRE. *Human reproduction* **31**, 233–248, <https://doi.org/10.1093/humrep/dev319> (2016).
- Grafodatskaya, D., Cyttrynbaum, C. & Weksberg, R. The health risks of ART. *EMBO reports* **14**, 129–135, <https://doi.org/10.1038/embor.2012.222> (2013).
- Vermeiden, J. P. & Bernardus, R. E. Are imprinting disorders more prevalent after human *in vitro* fertilization or intracytoplasmic sperm injection? *Fertility and sterility* **99**, 642–651, <https://doi.org/10.1016/j.fertnstert.2013.01.125> (2013).
- Weksberg, R., Shuman, C. & Beckwith, J. B. Beckwith-Wiedemann syndrome. *European journal of human genetics: EJHG* **18**, 8–14, <https://doi.org/10.1038/ejhg.2009.106> (2010).
- Chen, Z., Robbins, K. M., Wells, K. D. & Rivera, R. M. Large offspring syndrome: a bovine model for the human loss-of-imprinting overgrowth syndrome Beckwith-Wiedemann. *Epigenetics: official journal of the DNA Methylation Society* **8**, 591–601, <https://doi.org/10.4161/epi.24655> (2013).
- Young, L. E., Sinclair, K. D. & Wilmut, I. Large offspring syndrome in cattle and sheep. *Rev Reprod* **3**, 155–163 (1998).
- Rivera, R. M. *et al.* Manipulations of mouse embryos prior to implantation result in aberrant expression of imprinted genes on day 9.5 of development. *Hum Mol Genet* **17**, 1–14, <https://doi.org/10.1093/hmg/ddm280> (2008).
- Market-Velker, B. A., Zhang, L., Magri, L. S., Bonvissuto, A. C. & Mann, M. R. Dual effects of superovulation: loss of maternal and paternal imprinted methylation in a dose-dependent manner. *Hum Mol Genet* **19**, 36–51, <https://doi.org/10.1093/hmg/ddp465> (2010).
- de Waal, E. *et al.* Primary epimutations introduced during intracytoplasmic sperm injection (ICSI) are corrected by germline-specific epigenetic reprogramming. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4163–4168, <https://doi.org/10.1073/pnas.1201990109> (2012).
- de Waal, E. *et al.* Gonadotropin stimulation contributes to an increased incidence of epimutations in ICSI-derived mice. *Human molecular genetics* **21**, 4460–4472, <https://doi.org/10.1093/hmg/dds287> (2012).
- Kuentz, P. *et al.* Child with Beckwith-Wiedemann syndrome born after assisted reproductive techniques to a human immunodeficiency virus serodiscordant couple. *Fertil Steril* **96**, e35–38, <https://doi.org/10.1016/j.fertnstert.2011.04.030> (2011).
- Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature reviews. Genetics* **14**, 204–220, <https://doi.org/10.1038/nrg3354> (2013).
- He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu Rev Genomics Hum Genet* **16**, 55–77, <https://doi.org/10.1146/annurev-genom-090413-025437> (2015).
- Shirane, K. *et al.* Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS genetics* **9**, e1003439, <https://doi.org/10.1371/journal.pgen.1003439> (2013).
- Bartolomei, M. S. & Ferguson-Smith, A. C. Mammalian genomic imprinting. *Cold Spring Harbor perspectives in biology* **3**, <https://doi.org/10.1101/cshperspect.a002592> (2011).
- Plasschaert, R. N. & Bartolomei, M. S. Genomic imprinting in development, growth, behavior and stem cells. *Development* **141**, 1805–1813, <https://doi.org/10.1242/dev.101428> (2014).
- Kalish, J. M., Jiang, C. & Bartolomei, M. S. Epigenetics and imprinting in human disease. *Int J Dev Biol* **58**, 291–298, <https://doi.org/10.1387/ijdb.140077mb> (2014).
- Chen, Z. *et al.* Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 4618–4623, <https://doi.org/10.1073/pnas.1422088112> (2015).
- Eggermann, T. *et al.* Prenatal molecular testing for Beckwith-Wiedemann and Silver-Russell syndromes: a challenge for molecular analysis and genetic counseling. *European journal of human genetics: EJHG* **24**, 784–793, <https://doi.org/10.1038/ejhg.2015.224> (2016).
- Ingleby, F. C., Flis, I. & Morrow, E. H. Sex-biased gene expression and sexual conflict throughout development. *Cold Spring Harbor perspectives in biology* **7**, a017632, <https://doi.org/10.1101/cshperspect.a017632> (2014).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
- Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols* **8**, 1765–1786, <https://doi.org/10.1038/nprot.2013.099> (2013).
- Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).

24. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559, <https://doi.org/10.1186/1471-2105-9-559> (2008).
25. Biressi, S., Molinaro, M. & Cossu, G. Cellular heterogeneity during vertebrate skeletal muscle development. *Developmental biology* **308**, 281–293, <https://doi.org/10.1016/j.ydbio.2007.06.006> (2007).
26. Schiaffino, S. & Mammucari, C. Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: insights from genetic models. *Skelet Muscle* **1**, 4, <https://doi.org/10.1186/2044-5040-1-4> (2011).
27. Laurin, M. *et al.* The atypical Rac activator Dock180 (Dock1) regulates myoblast fusion *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15446–15451, <https://doi.org/10.1073/pnas.0805546105> (2008).
28. Thorsteinsdottir, S., Deries, M., Cachaco, A. S. & Bajanca, F. The extracellular matrix dimension of skeletal muscle development. *Developmental biology* **354**, 191–207, <https://doi.org/10.1016/j.ydbio.2011.03.015> (2011).
29. Abmayr, S. M. & Pavlath, G. K. Myoblast fusion: lessons from flies and mice. *Development* **139**, 641–656, <https://doi.org/10.1242/dev.068353> (2012).
30. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome biology* **13**, R83, <https://doi.org/10.1186/gb-2012-13-10-r83> (2012).
31. Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & development* **28**, 812–828, <https://doi.org/10.1101/gad.234294.113> (2014).
32. O'Doherty, A. M., O'Shea, L. C. & Fair, T. Bovine DNA methylation imprints are established in an oocyte size-specific manner, which are coordinated with the expression of the DNMT3 family proteins. *Biology of reproduction* **86**, 67, <https://doi.org/10.1095/biolreprod.111.094946> (2012).
33. Chen, Z. *et al.* Global assessment of imprinted gene expression in the bovine conceptus by next generation sequencing. *Epigenetics: official journal of the DNA Methylation Society* **11**, 501–516, <https://doi.org/10.1080/15592294.2016.1184805> (2016).
34. Robbins, K. M., Chen, Z., Wells, K. D. & Rivera, R. M. Expression of KCNQ1OT1, CDKN1C, H19, and PLAGL1 and the methylation patterns at the KvDMR1 and H19/IGF2 imprinting control regions is conserved between human and bovine. *Journal of biomedical science* **19**, 95, <https://doi.org/10.1186/1423-0127-19-95> (2012).
35. Liu, Y., Siegmund, K. D., Laird, P. W. & Berman, B. P. Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome biology* **13**, R61, <https://doi.org/10.1186/gb-2012-13-7-r61> (2012).
36. Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831, <https://doi.org/10.1016/j.cell.2011.12.035> (2012).
37. Bebbere, D. *et al.* Tissue-specific and minor inter-individual variation in imprinting of IGF2R is a common feature of *Bos taurus* Concepti and not correlated with fetal weight. *PLoS one* **8**, e59564, <https://doi.org/10.1371/journal.pone.0059564> (2013).
38. Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K. & Schweifer, N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* **349**, 84–87, <https://doi.org/10.1038/349084a0> (1991).
39. Xu, Y., Goodyer, C. G., Deal, C. & Polychronakos, C. Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochemical and biophysical research communications* **197**, 747–754 (1993).
40. Young, L. E. *et al.* Epigenetic change in IGF2R is associated with fetal overgrowth after sheep embryo culture. *Nature genetics* **27**, 153–154, <https://doi.org/10.1038/84769> (2001).
41. Wutz, A. *et al.* Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745–749, <https://doi.org/10.1038/39631> (1997).
42. Weksberg, R., Shuman, C. & Smith, A. C. Beckwith-Wiedemann syndrome. *American journal of medical genetics. Part C, Seminars in medical genetics* **137C**, 12–23, <https://doi.org/10.1002/ajmg.c.30058> (2005).
43. Farin, P. W. & Farin, C. E. Transfer of bovine embryos produced *in vivo* or *in vitro*: survival and fetal development. *Biology of reproduction* **52**, 676–682 (1995).
44. Maxfield, E. K. *et al.* Short-term culture of ovine embryos modifies fetal myogenesis. *The American journal of physiology* **274**, E1121–E1123 (1998).
45. Zheng, H. *et al.* Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Mol Cell* **63**, 1066–1079, <https://doi.org/10.1016/j.molcel.2016.08.032> (2016).
46. Tycko, B. & Allele-specific, D. N. A. methylation: beyond imprinting. *Human molecular genetics* **19**, R210–R220, <https://doi.org/10.1093/hmg/ddq376> (2010).
47. Degner, J. F. *et al.* Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* **25**, 3207–3212, <https://doi.org/10.1093/bioinformatics/btp579> (2009).
48. Stevenson, K. R., Coolon, J. D. & Wittkopp, P. J. Sources of bias in measures of allele-specific expression derived from RNA-sequence data aligned to a single reference genome. *BMC genomics* **14**, 536, <https://doi.org/10.1186/1471-2164-14-536> (2013).
49. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357–360, <https://doi.org/10.1038/nmeth.3317> (2015).
50. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572, <https://doi.org/10.1093/bioinformatics/btr167> (2011).
51. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology* **13**, R87, <https://doi.org/10.1186/gb-2012-13-10-r87> (2012).
52. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, <https://doi.org/10.1093/bioinformatics/btq033> (2010).
53. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, W187–191, <https://doi.org/10.1093/nar/gku365> (2014).
54. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24–26, <https://doi.org/10.1038/nbt.1754> (2011).

Acknowledgements

We thank Dr. Susanta Behura for the critical reading of the manuscript. This work was supported by the University of Missouri Research Board (RB 16–07 HAGEN), National Science Foundation (NSF Award No. 1615789), National Institute of Health (Grant 5R21HD062920), and the Food for the 21st Century Program at the University of Missouri.

Author Contributions

Z.C., D.E.H., and R.M.R. designed research; Z.C. performed research and analyzed data; D.E.H. and C.G.E. assisted with data analyses and contributed analytic tools; T.J. assisted with statistical analyses; Z.C. and R.M.R. interpreted the results and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-13012-z>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017