The counterpart congenital overgrowth syndromes Beckwith-Wiedemann Syndrome in human and large offspring syndrome in bovine involve alterations in DNA methylation, transcription, and chromatin configuration.

Yahan Li [1], Ping Xiao [2,a], Frimpong Boadu [3,a], Anna K. Goldkamp [2,a], Snehal Nirgude [4], Jianlin Cheng [3], Darren E. Hagen [2,*], Jennifer M. Kalish [4,5,*], Rocío Melissa Rivera [1,*]

[a] All contributed equally to the work

[1]Division of Animal Sciences, University of Missouri, Columbia, MO, 65211, USA

[2]Department of Animal and Food Sciences, Oklahoma State University, Stillwater, OK 74078, USA

[3]Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA

[4]Division of Human Genetics and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[5]Departments of Pediatrics and Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA

***** Correspondence: riverarm@missouri.edu, kalishj@chop.edu, darren.hagen@okstate.edu

Short Title: Molecular analyses of the counterpart overgrowth syndromes BWS and LOS

Abstract

Beckwith-Wiedemann Syndrome (BWS, OMIM #130650) is a congenital epigenetic disorder in humans which affects approximately 1 in 10,340 children. The incidence is likely an underestimation as the condition is usually recognized based on observable phenotypes at birth. BWS children have up to a 28% risk of developing tumors and currently, only 80% of patients can be corroborated molecularly (epimutations/variants). It is unknown how the subtypes of this condition are molecularly similar/dissimilar globally, therefore there is a need to deeply characterize the syndrome at the molecular level. Here we characterize the methylome, transcriptome and chromatin configuration of 18 BWS individuals together with the animal model of the condition, the bovine large offspring syndrome (LOS). Sex specific comparisons are performed for a subset of the BWS patients and LOS. Given that this epigenetic overgrowth syndrome has been characterized as a loss-of-imprinting condition, parental allele-specific comparisons were performed using the bovine animal model. In general, the differentially methylated regions (DMRs) detected in BWS and LOS showed significant enrichment for CTCF binding sites. Altered chromosome compartments in BWS and LOS were positively correlated with gene expression changes, and the promoters of differentially expressed genes showed significant enrichment for DMRs, differential topologically associating domains, and differential A/B compartments in some comparisons of BWS subtypes and LOS. We show shared regions of dysregulation between BWS and LOS, including several HOX gene clusters, and also demonstrate that altered DNA methylation differs between the clinically epigenetically identified BWS patients and those identified as having DNA variants (i.e. *CDKN1C*

48 microdeletion). Lastly, we highlight additional genes and genomic regions that have the
49 potential to serve as targets for biomarker development to improve current molecular
50 methodologies. In summary, our results suggest that genome-wide alternation of
51 chromosome architecture, which is partially caused by DNA methylation changes, also
52 contribute to the development of BWS and LOS.
53
54
55 Keywords: Beckwith-Wiedemann Syndrome, large offspring syndrome, DNA
56 methylation, chromosome architecture, Hi-C sequencing
57
58

Introduction

Beckwith-Wiedemann Syndrome (BWS, OMIM #130650) in humans and its bovine counterpart large/abnormal offspring syndrome (LOS/AOS) are well-known congenital epigenetic disorders [1–3]. The incidence of spontaneous BWS has been reported to be approximately 1 in 10,340 live births but the incidence of spontaneous LOS is currently unknown [4–6]. BWS and LOS share a broad spectrum of clinical features including macrosomia, macroglossia, abdominal wall defects (umbilical hernia/omphalocele), lateralized overgrowth, ear malformation, organomegaly, and placentomegaly [2,7–19]. In addition, BWS is also associated with increased embryonal tumor incidence, hyperinsulinism, and facial nevus simplex [2], and LOS often results in dystocia and demise of offspring and dam due to offspring oversize [20]. BWS and LOS share epigenetic defects in the form of loss-of-imprinting at *IGF2*, *KCNQ1*, *IGF2R*, *PLAGL1*, *PEG3*, and *DLK1* [2,6–8,21–30]. In addition to primary epimutations, subtypes of BWS involve genetic defects which lead to loss of gene function or secondary epimutations that result in loss-of-imprinting [31,32].

The development of the mammalian embryo is a finely regulated process and adverse external insults during this window can have permanent influence on the health of the individual long after the exposure, a phenomenon known as Developmental Origins of Health and Disease [33,34]. Both BWS and LOS have been reported to occur more frequently in offspring conceived using assisted reproductive technologies (ART). Children conceived using ART have over a 10-fold increased risk (1 in 1,126 live births) for developing BWS [35], while in the bovine genetic improvement industry, the incidence of overgrowth was recently reported to be 10% for males and 24% for

82 females [36]. The causal relationship between ART and the increased incidence of
83 these syndromes is considered to be achieved by inducing errors in the epigenome of
84 oocytes and/or embryos since ART manipulations overlap with critical developmental
85 windows for epigenetic reprogramming [37–41].

86 We and others have shown that alterations in epigenetic mechanisms (*i.e.*
87 genomic imprinting, DNA methylation, and locus-specific chromosome architecture) and
88 the resultant altered expression of small/long non-coding RNAs are detectable during
89 gestation and after birth in these syndromes [42–44]. However, whether genome wide
90 alteration of chromosome architecture exist in these conditions is not known, nor is how
91 altered genome organization is involved in the observed global alterations in methylome
92 and transcriptome and vice versa. Our hypothesis is that genome-wide alteration of
93 chromosome architecture occurs in the counterpart overgrowth syndromes BWS and
94 LOS that lead to regional and global gene expression changes and that some of these
95 topological aberrations are shared between the two species. In this study, we
96 systematically analyzed the alterations of DNA methylation, chromosome architecture,
97 and gene expression in fibroblast primary cells from human BWS patients and bovine
98 LOS fetuses and explored the association among these changes.
99

100 **Results and Discussion**

101 For human, five controls (CON_H) and 18 BWS individuals were used for
102 analyses. The BWS group is divided into three subgroups according to molecular
103 subtypes, namely imprinting center 1 gain of methylation (BWS_IC1_GOM, n=3),
104 imprinting center 2 loss of methylation (BWS_IC2_LOM, n=12), and *CDKN1C* variant

(BWS_CDKN1C_Mut, n=3). For bovine, the experimental groups were control (CON_B, n=5) and LOS (LOS, n=5). It should be noted that one of the LOS samples, LOS_#1, is considered an outlier according to all the sequencing results, therefore it is excluded from the results presented in the main document. Data including LOS_#1 is available in the supplement (LOS_All group, n=6). For each statistical test, overall (non-sex-specific) and sex-specific comparisons between groups were conducted for human and bovine. The sex of human patient BWS_16_Mut shows conflict between all the sequencing results (female) and hospital record (male), so it was excluded from the sex-specific comparisons. Given that BWS and LOS have been characterized as loss-of-imprinting conditions [2,22], parental allele-specific comparisons were performed using the bovine animal model as this type of analysis is not possible with the human samples. A summary of performed comparisons and numbers of statistically significant differences for all the analyses can be found in Table 1 (index 1-36).

DNA methylation profile in BWS and LOS

Whole genome bisulfite sequencing (WGBS) was conducted for human and bovine fibroblast samples to determine CpG methylation status. Read alignment information can be found in Table S1. Differentially methylated regions (DMR) were determined for all the comparisons in Table 1 and the DMR details can be found in Table S1.

Principal component analysis (PCA) clusters the majority of BWS samples while the control samples are more dispersed (Figure S1A), perhaps due to limitations in obtaining age matched healthy human samples. Similarly, hierarchical cluster analysis

groups most of the BWS samples irrespective of subtype (Figure 1A). 5,708 DMRs were detected when comparing all the BWS samples together with the control group, and ~66% of these DMRs were hypermethylated (Figure 1A). A greater number of DMRs were detected between BWS and controls when each subtype was analyzed independently, and hypermethylation was observed in 67% of DMRs in IC2 LOM group, 30% in IC1 GOM group and 96% in the *CDKN1C* variant group (Figure S2A, B, C). We next compared the methylomes of the three BWS subtypes and identified clear and specific differences between them (Figure S2D, E, F). For bovine samples, 10,860 DMRs were identified in LOS with ~75% of those being hypermethylated (Figure 1B), as per the overall BWS analysis.

Figure 1. CpG methylation heatmap and genomic context enrichment for differentially methylated regions (DMR). (A) and **(B)** DMR for human BWS group compared with control group **(A)**, and bovine LOS group compared with control group **(B)**. The left panel is the heatmap with hierarchical clustering and sex annotation which include male (M), female (F), and conflict (C). The right panels are the distribution of genomic contexts at the DMR (centered) and surrounding regions. Genomic contexts include 1kb promoter, exon, intron, CpG island, predicted CTCF binding sites, public CTCF ChIP-seq peaks, and repetitive sequences. The line chart on top of each panel is the average enrichment score for all the DMRs. **(C)** Number of unique and overlapped DMRs between BWS IC2 LOM subtypes and LOS.

The predominant epimutation in BWS is loss of methylation (LOM) on maternal allele at the imprinting control region (ICR) of *KCNQ1OT1*, namely imprinting center 2

(IC2; i.e. KvDMR1), which accounts for ~50% of documented cases [2]. Normal levels of DNA methylation at this ICR should be ~50%, since only the maternal allele is normally methylated. Average DNA methylation at this locus (i.e. chr11:2719755-2722554) was 17.59% in fibroblasts from IC2 patients and 45.18% in control samples, indicating LOM in this BWS subtype. This LOM at IC2 is not complete (0%), because the epimutations in BWS are mosaic which lead to a mixture of affected and not affected cells [2]. Hierarchical clustering based on these DMRs separated most BWS with IC2 LOM samples away from controls (8/12; Figure S2B). No clustering based on sex was observed. To further investigate whether the other four IC2 LOM samples that intermixed with controls could be considered a novel IC2 subgroup, we subdivided the IC2 LOM group into subgroup 1 (BWS_IC2_LOM_S1, n=8) and 2 (BWS_IC2_LOM_S2, n=4) and compared them with the control group and each other. Subgroup-specific comparisons against controls identified a larger number of DMRs (23,582) in subgroup 1 than in subgroup 2 (1,915). 672 DMRs were identified as shared between the two IC2 LOM subgroups (Figure 1C). We then compared DMRs between LOS and the two IC2 LOM subgroups and found that 2.89% (683) DMRs of subgroup 1 and 1.93% (37) DMRs of subgroup 2 are shared with LOS (Figure 1C and Table S1). In total, 17 DMRs were shared between the two IC2 LOM subgroups and LOS, including at the promoters of *KCNQ1OT1* (i.e., KvDMR1) and *HOXA-AS2* and gene bodies of the epigenetic modifiers *HDAC4* and *PRDM8* (Figure 1C and Table S1). Further, one of the shared DMRs is in the gene body of *NR2F1-AS1*, a gene associated with tumors reported in BWS, namely rhabdomyosarcoma and hepatocellular carcinoma [45].

The second most common epimutation in BWS (~5% of cases [2]) is gain of

methylation (GOM) at the *H19/IGF2* imprinted domain (i.e. IC1). Compared to controls, fibroblasts from BWS IC1 samples have 19.09% more DNA methylation at chr11:2015075-2022914, which encompass the IC1 and the entire *H19* gene body. This synchronized DNA methylation change in IC1 and *H19* gene body has also been reported in recurrent spontaneous abortion patients [46]. Cluster analysis of the DMRs grouped the three BWS samples with IC1 epimutation away from controls (Figure S2A). No clustering based on sex was observed. Heritable genetic mutations (including large-scale chromosomal rearrangement but not uniparental disomy) caused loss-of-imprinting accounts for up to 9% of BWS cases [2]. Specifically, genetic mutations at the *CDKN1C* gene occur in ~5% of BWS cases [2]. Figure S2C shows the cluster analysis of DMRs between BWS with *CDKN1C* variant and controls. Of note, no IC1 or IC2 epimutations were detected in this subgroup.

Sex-specific comparisons of the IC2 LOM subtype identified a larger number of DMRs in males than females when compared to controls (53,488 vs. 26,459, respectively). As per BWS, LOS males have twice the number of methylome epimutations than females (75,370 and 34,270 DMRs, respectively). Interestingly, DMRs in male LOS are mainly hypermethylated while DMRs in female LOS are mainly hypomethylated (Figure S3B and S3C).

Allele-specific methylation analysis in the bovine control group clearly separates the maternal and paternal alleles in the PCA plot (Figure S1C) and identified 10,143 DMRs between the parental alleles, with the majority (>99%) being hypomethylated on the paternal allele (Figure S3D). Some of the identified DMRs overlap with known imprinted genes and exhibit the expected allelic patterns. For example, lower paternal

methylation at *IGF2R* ICR (89.37%; chr9:96222606-96223965), lower paternal methylation at IC2 (89.55%; chr29:48908847-48909926), and higher paternal methylation at IC1 (68.16% and 73.02% at chr29:49504687-49505606 and chr29:49507607-49510366, respectively). Interestingly, there seems to be a spread of DNA methylation on the paternal allele from the IC1 to *H19*, creating allelic specific methylation over the gene body (chr29:49511527-49512886) with 9.70% methylation on maternal allele and 55.24% on paternal allele. Allele-specific comparisons between LOS and controls, identified that both alleles are mainly hypermethylated in LOS, which is similar to the finding of the overall comparison (Figure S3E and S3F). However, the LOS-associated allele-specific DMRs are, for the most part (>80%), not shared between the parental alleles. As expected, the normally methylated maternal allele of the *IGF2R* ICR (chr9:96222606-96223885) is hypomethylated in LOS when compared to controls, similar to what has been previously published by us and others [21,24]. We have previously reported maternal LOM at IC2 in some LOS fetuses [7]. In the current study, one LOS fetus (LOS_#3, the heaviest female) shows a 54.36% decrease in DNA methylation at maternal IC2 (chr29:48908887-48909846) than the control group mean (90.98%). However, the group mean difference between LOS and control is 14.7% at maternal IC2 which does not pass the 15% filtering threshold and is not included in the results, although it indeed shows some level of hypomethylation. Similar to our previous findings, maternal hypermethylation at IC1, was not observed in LOS, however, we detected hypermethylation at IC1 (chr29:49504927-49505566, 17.98%) and the *H19* gene body (chr29:49512727-49512926, 21.02%) on the paternal allele, which is a novel finding of unknown significance.

Overall, hypomethylated DMRs in BWS and LOS (including overall, sex-specific, and allelic-specific comparisons) are significantly enriched for gene body, exon, intron, CpG island, CpG shore, CpG shelf, predicted CTCF binding sites, and public CTCF ChIP-seq peaks, and depletion for repetitive sequence according to permutation tests (Figure 1, S2, S3, and Table S1). Hypomethylated DMRs also show significant enrichment for gene promoter (1kb) in BWS related comparisons and LOS allele-specific comparisons. For hypermethylated DMRs in most of the LOS and BWS comparisons, there is also enrichment for CpG island, CpG shore, CpG shelf, and CTCF binding sites. However, the pattern is varied for gene promoter, exon, intron, and repetitive sequence based on sex and BWS subtype. For example, BWS with IC2 LOM and BWS with IC1 GOM show enrichment at these elements but BWS with *CDKN1C* variant shows depletion.

Interestingly, among the top 50 more severely dysregulated regions (both hyper- and hypomethylated) in the BWS with IC2 LOM group, 11 hypomethylated DMRs overlap the gene promoter and/or body of several *HOX* genes, including *HOX* A3, A4, A5, A6, A7, B7, B9, C4, C10, and D4 from four *HOX* clusters located on chromosome 7 (A), 17 (B), 12 (C), and 2 (D) (Figure 2 and Table S1). Most of these genes also had hypomethylated DMRs in the BWS with IC1 GOM group. For example, the gene body of *HOXC4* (chr12:54446778-54448977) had a 41.17% decrease in the BWS IC2 LOM group and 58.58% decrease in the BWS IC1 GOM group. Further, the promoter of *HOXD4* (chr2:177015615-177017974) also had reduced DNA methylation in these groups (30.26% and 36.05%, respectively). In LOS, *HOX* A3, A5, A6, B3, C4, C10, D4, and D11, also have modest alterations (15%-20%) in DNA methylation when compared

to controls. *HOX* genes play roles as both oncogenes and tumor suppressors in regulating cell proliferation and differentiation [47]. Both hyper- and hypomethylated *HOX* genes are frequently observed in cancer, including Wilms tumor, which is the most commonly diagnosed tumor in BWS [47,48]. *HOX* genes are important transcription factors during embryonic development and exhibit spatial and temporal-specific expression patterns [49]. In human, the number in the name of *HOX* genes are based on their order of spatial expression pattern with *HOX* 1 being expressed in the most anterior side (i.e. the head) and *HOX* 13 is the most posterior side (i.e. the trunk) [47]. *HOX* 1 to 4, expressed in the head region, are more frequently associated with DMRs in BWS, especially *HOX* 4. This observation may be partially the result of the source of our BWS fibroblast cells as some samples were collected from the skin behind the ear; further studies with various sources of fibroblasts will be needed to validate this finding. Importantly, hypomethylated DMRs were identified in the top DMR lists in the IC2 LOM and IC1 GOM when compared to BWS with *CDKN1C* variant. Identified DMRs in BWS reflect altered *HOX* expression during embryonic development and have the potential to serve as molecular markers to diagnose and distinguish different subtypes.

Figure 2. DNA methylation alterations at *HOX* gene clusters in BWS and LOS. (A) *HOXA* cluster. **(B)** *HOXB*. **(C)** *HOXC*. **(D)** *HOXD*. Top panel is human genome and bottom panel is bovine genome for each plot. Additional gene annotation from NCBI RefSeq is added since some *HOX* genes are not annotated in the R package for UCSC annotation. All the *HOX* genes from the same cluster are located on the same strand. The order of *HOX* genes may be opposite for human and bovine.

In summary, the abovementioned findings demonstrate methylome distinctions between the epigenetically- and also the genetically-identified syndrome beyond the signature DMRs or deletions. BWS children have up to a 28% risk of developing tumors and currently, only 80% of patients can be corroborated molecularly, future studies will need to be performed to determine their usefulness in diagnosing these children better and more specifically.

Gene expression profile and alternative splicing events in BWS and LOS

Ribosomal RNA depleted total RNA sequencing was conducted to determine the global transcript profile in human and bovine fibroblast samples. In accordance with the cell-type used here, extracellular matrix genes show high expression level in both species, including collagen (*COL1A1/2*, *COL3A1*, *COL5A1/2, COL6A1/2/3*, and *COL12A1*), thrombospondin (*THBS1*), vimentin (*VIM*), filamin (*FLNA*), fibronectin (*FN1*), and tenascin (*TNC*). Differentially expressed genes (DEGs) were determined for all the comparisons in Table 1 and the DEG details can be found in Table S2.

The transcriptomes of the human fibroblasts are not distinct between them and as a result the samples are intermixed in the PCA (Figure S1D). Twelve DEGs were detected when all BWS samples were compared to controls and were all downregulated in BWS. Those genes were *LINGO2*, *MDGA2*, *NGFR*, *MIR92A1*, *FLG*, *BMP2*, *SIGLEC15*, *SLAMF7*, *LIPH*, and three novel genes. Downregulation of *MDGA2*, *NGFR* and *BMP2* have been reported in tumorigenesis [50–52]. DEG based hierarchical clustering separates each of the BWS subtypes from the control samples (Figure 3B);

however, no obvious clustering based on sex was observed in the IC2 LOM group. Subtype-specific analyses revealed that 23 genes were different from controls in IC1 GOM, 8 in IC2 LOM and 10 in *CDKN1C* variant (Figure 3A, 3B, and 3C). When the subgroups of IC2 LOM were taken into consideration, 742 transcripts were identified as different from controls in subgroup 1 and two in subgroup 2. Nine DEGs were shared between IC2 LOM subgroup 1 and IC1 GOM (Table S2), making them potential subtype-specific biomarkers. Those were the upregulated *GRID1*, *KCNA4*, *SESN3*, *GOLGA8CP*, *NTRK3*, *NKD1*, and the downregulated *PTPRE*, *GPRC5A*, and *SERPINB2*. The other DEGs were unique to each BWS subtype. Some of the DEGs are associated with tumorigenesis, including downregulation of *LOXL4* in BWS with IC1 GOM, upregulation of *PECAM1* in BWS with IC2 LOM, and downregulation of *TAP2* in BWS with *CDKN1C* variant [53–55]. In addition, *SESN3* is involved in insulin sensitivity and in tumorigenesis [56,57]. There was no significant group difference, however, in transcript abundance for the signature imprinted genes of BWS (e.g. *H19*, *IGF2*, *CDKN1C*, and *KCNQ1OT1*) in these comparisons. High variations in transcript level were observed for *H19* and *IGF2* in the control group (more than 70-fold difference between the highest and lowest individual). *IGF2* transcript abundance was very high in BWS_#11 (IC2 LOM) and #16 (*CDKN1C* variant), more than four times higher than the highest control sample, and *KCNQ1OT1* level was three-fold higher than controls in BWS_#7 and #11 (IC2 LOM).

Figure 3. Expression heatmap and genomic context enrichment for differentially expressed gene (DEG). DEG for human BWS subtypes compared with control group **(A-C)**, and bovine LOS group compared with control group **(D)**. The left

panel is the heatmap with hierarchical clustering and sex annotation which include male (M), female (F), and conflict (C). Gene expression is shown as log2 transformed difference between sample and row average. The right panels are the distribution of genomic contexts at the DEG (centered at 1kb promoter or gene body) and surrounding regions. Genomic contexts include differentially methylated regions (DMR), differential interacting regions (DIR), differential topologically associating domains (diff_TAD), and differential compartments identified in corresponding group comparisons. The line chart on top of each panel is the average enrichment score for all the DEGs.

For bovine, PCA separates the samples based on treatment (Figure S1E) although not by parental allele (Figure S1F) indicating similar expression profiles from both alleles. Transcriptome comparison between LOS and controls identified 112 DEGs and hierarchical clustering clearly separated the two groups (Figure 3D). As expected, the signature imprinted gene of LOS (*IGF2R*) was found downregulated two-fold in the LOS group. *IGF2R* is normally highly expressed from the maternal allele and its protein can bind to IGF2 protein with high affinity to mediate the degradation of excess IGF2 [58]. Studies in mice have shown that maternal inheritance of *Igf2r* null variant causes increase of IGF2 and IGF binding proteins, ~25%-30% increase of body weight, proportionately increased organ size, and perinatal death [59]. The overgrowth phenotype in LOS could be largely attributed to changes in the IGF2 signaling pathway. In addition, several DEGs in LOS have also been reported in tumor studies in human, such as downregulation of *AOX1*, *CYGB*, and *TMOD2*, and upregulation of *BCL11A*, *AGRN*, and *LEF1* [60–65]. Four DEGs were shared between LOS and BWS IC2 LOM

subgroup 1, namely the upregulated *EMID1*, and the downregulated *SERPINE2*, *LRRC8D*, *GPX1*. When analyzing the transcriptome in a sex-specific manner, male LOS has 109 DEGs when compared to controls while the number of DEGs is doubled in female LOS (213). Of those DEGs, eight were detected in both male and female LOS, namely *FBLN7* (down/down in male/female), *CYGB* (down/down), *LOC112446407* (up/up), *ABLIM3* (down/down), *IGF2R* (down/down), and three novel genes MSTRG.7897 (up/up), MSTRG.13383 (up/down), and MSTRG.22799 (down/up). Downregulation of *ABLIM3* has also been reported in hepatocellular carcinoma [66].

Allele-specific comparison in the bovine control group identified 122 genes with significant biased expression from one allele, including *IGF2R* which has about three-fold higher expression from the maternal allele than the paternal allele. In addition, *GSTK1*, *MAN1B1*, *GSTZ1*, *ARSG*, *SLC25A29*, *TLDC1*, *PDIA4*, *RWDD2A*, and *NARS* have been reported as imprinted genes or to have biased allelic expression in various mammalian species by other researchers[67–75]. The allelic expression pattern of *Gstz1* is associated with expression quantitative trait loci (eQTL) in mice [67]. *SLC25A29*, together with our previously reported *BEGAIN* [76], show biased paternal expression and are located close to the *DLK1-DIO3* imprinted domain, which suggest broader regulatory impacts of this domain in bovine. *TLDC1* is associated with gamete-specific promoter methylation patterns in human [77]. Although not possible to determine the imprinting status of the remaining genes in this study, as we only have F1 samples with one parental cross (*i.e. B.t. indicus X B.t. taurus*), other genes in the list should be further characterized to determine if they are imprinted or rather they behave as eQTL as we have previously shown [78].

Allele-specific DEGs were identified in LOS. *IGF2R* shows a significant decrease (~2.4 fold) from the maternal allele making it similar in expression level to the paternal allele. The expression of this gene from the paternal allele in LOS was similar to control. In addition, among the 122 genes with biased allelic expression in controls, *ECHDC2* has higher paternal expression in controls and is downregulated on the paternal allele in the LOS group, and *IGF1R* has lower maternal expression in controls and is upregulated on the maternal allele in LOS group. *ECHDC2* is involved in branched chain amino acid metabolism and its lower expression increases cell viability, and *IGF1R* encodes a membrane receptor for the IGF signaling and is positively correlated with growth [79,80].

Regarding the association between changes in gene expression and DNA methylation, upregulated DEGs in female BWS with IC2 LOM group showed enrichment for hypomethylated DMRs at promoter and gene body, and downregulated DEGs in this group showed enrichment for both hypo- and hypermethylated DMRs at promoter and gene body according to permutation tests (Table S2). In addition, upregulated DEGs in LOS group showed enrichment for both hypo- and hypermethylated DMRs at promoter and hypomethylated DMRs at gene body, and downregulated DEGs in this group showed enrichment for hypermethylated DMRs at promoter and gene body. Given that DMRs overlapping exons can influence alternative splicing, we analyzed the total RNA sequencing data to identify alternative splicing (AS) events in BWS and LOS [81]. For humans, 1,183, 1,292, and 1,151 AS events were identified in BWS with IC1 GOM, IC2 LOM, and *CDKN1C* variant groups when compared to the control group (Figure 4A, S4A, S4B, and Table S2). For bovine, 746 AS events were detected in the LOS group

when compared to the control group (Figure 4B and Table S2). Regardless of species or subtypes, most AS events were skipped exons with an average of 43.5% across pairwise comparisons. AS events appeared to be evenly distributed across chromosomes in BWS and LOS. Few genes shared alternative splicing events between BWS and LOS, however, we did find instances of shared splicing events in pathways relevant to overgrowth. For example, genes involved in the WNT signaling pathway (*TCF7L2*, *HDAC1*, *PRKC1*) had skipped exons in both BWS with *CDKN1C* variant and LOS groups. WNT signaling pathway regulates cell proliferation and differentiation, and is associated with Hippo pathway to control organ size [82]. YAP1, which is a transcription co-activator in Hippo pathway, has been reported with increased non-phosphorylated (active) protein level in skeletal muscle of LOS fetuses by us [26]. Although none of the exons of the mentioned genes were differentially methylated, these data suggest alternative splicing as a contributor or outcome of the overgrowth phenotype.

Figure 4. Identification of alternative splicing events in Beckwith-Wiedemann Syndrome (BWS) and Large Offspring Syndrome (LOS) and correlation network in BWS. (A) and **(B)** The genomic distribution, numbers, and types of alternative splicing events identified in human BWS IC2 LOM (A) and bovine LOS (B) groups when compared to their control groups. Manhattan plots show alternatively spliced genes aligned along the x-axis by their position on the chromosome and on the y-axis by their log10(false discovery rate). The horizontal dashed black line represents a significance threshold of 0.05. **(C)** Correlation between gene expression profile in 15 modules and BWS diagnosis and scores determined by weighted gene co-expression network analysis (WGCNA). Numbers in the heatmap represent correlation

coefficient. Significance is shown as * (<0.05), ** (<0.01), and ns (≥0.05) in parentheses. (D) Gene interaction network within purple module and hub genes were highlighted.

To understand the correlation between transcriptional regulation and phenotypic characteristics of human BWS features, weighted correlation network analysis (WGCNA) was performed to determine gene clusters with high correlation in BWS. As a result, genes were grouped into 15 modules based on similarity of expression. The grey and purple modules showed a positive correlation (p-value < 0.05) with macroglossia and omphalocele (Figure S4C) suggesting that genes with high expression in these modules may serve as biomarkers of these BWS pathologies. Other suggestive features such as a birthweight greater than two standard deviations (Birthweight_2SDs), facial nevus simplex, and ear creases or pits also showed higher correlation with purple modules (Figure S4D). Based on the standard of BWS scoring from Wang's study [83], a comprehensive BWS score was generated for each individual and expressional trends of purple module showed a significantly positive correlation with BWS score compared to other modules (Figure 4C). The 94 genes in this module showed enriched functions in RNA processing and regulation of immune system process (Table S2). Among the genes within the purple module, 21 hub genes with high BWS score correlation (BWS correlation ≥ 0.5) and module membership score (MM_purple ≥ 0.8) were identified (Figure 4D and Table S2), and one of these genes, Myb-Like, SWIRM And MPN Domain-Containing Protein 1 (*MYSM1*) is a transcription coactivator and participates in chromatin remodeling and repression of innate immunity and autoimmunity [84]. *KCNQ1OT1*, a long-noncoding RNA implicated in BWS, was also enriched [83]. In

addition, *MIR145* which encodes has-miR-145-5p was enriched in this module, which has been reported to be differentially expressed in BWS patients with macroglossia by us [26].

Characterization of chromosome architecture in human and bovine

Hi-C sequencing was performed to determine genome-wide chromosome architecture and its relationship with the methylome and transcriptome in human and bovine fibroblasts. Read alignment information can be found in Table S3.

We first characterized the two fundamental elements of chromosome architecture, namely chromosome compartments and topologically associating domains (TAD), in the human and bovine control groups (Table S3). For chromosome compartment, the A (positive eigenvalue, enriched in euchromatin) and B (negative eigenvalue, heterochromatin) compartments each roughly occupies half of the sequenceable autosomal genome in human (49.4% for A and 50.6% for B) and bovine (47.3% and 52.7%). Sex-specific and allele-specific analyses show a similar ratio (Table S3). 13,912 and 12,699 TADs (including sub-TADs) were found in the autosomes of human and bovine control groups, respectively, with similar TAD length distributions (Figure 5A, 6A, and Table S3). BWS, LOS, and sex-specific groups have similar TAD length distribution as their corresponding control group. For bovine allele-specific groups, the TAD length is decreased likely due to largely reduced read coverage after filtering for the parental alleles. The TAD boundaries in both species and all groups show enrichment for predicted CTCF binding sites, public CTCF ChIP-seq peaks, transcription start sites, and CpG islands, and depletion for repetitive sequences (Figure

459 5B and 6B).

460

461 **Figure 5.** Characterization of human chromosome architecture and alterations in
462 Beckwith-Wiedemann Syndrome (BWS). (A) Length distribution of topologically
463 associating domains (TAD) detected in control group. Level 1 TAD refers to the outmost
464 TADs. The red line marks the median length. (B) Genomic context enrichment at control
465 group TAD and subTAD boundaries (centered) and flanking regions. Genomic contexts
466 include predicted CTCF binding sites, CTCF public ChIP-seq peaks, transcription start
467 sites (TSS), CpG island, and repetitive sequences. The line chart on top of each panel
468 is the average enrichment score for all the boundaries. (C) Genome-wide distribution of
469 differentially methylated regions (DMR), differential compartments (diff_compartment),
470 differential TAD (diff_TAD), differential interacting regions (DIR) frequency, and
471 differentially expressed genes (DEG) identified when comparing BWS group with control
472 group. Red dots indicate an increase in CpG methylation, eigenvalue (towards A
473 compartment), and expression level, and black dots indicate decrease. The vertical
474 location of dots indicates the level of changes. Mb = megabases. (D) Correlation
475 between chromosome compartment changes (measured by eigenvalue) at gene
476 promoters and log2 transformed fold change (logFC) of protein-coding genes and
477 lncRNA. The left panel shows genome-wide correlations, and the right panel shows
478 correlation at regions of differential compartment. (E) Examples of alterations of
479 chromosome architecture associated with gene expression changes in BWS patients.
480 The top matrix plot shows the Hi-C contacts at 10kb resolution. A deeper blue color
481 indicates higher physical contact frequency. Red lines mark the TADs. All the statistical
482 analyses are conducted by comparing the bottom right group with the top left group.
483 Tracks below the matrix show the location of different elements along this genomic
484 region. The compartment track shows the A/B compartment represented by eigenvalue.
485 The blue and red lines indicate the top left and bottom right group, respectively. The
486 grey area marks significantly differential compartments between groups detected at
487 different resolutions which may overlap and show deeper color. The diff_TAD tracks

show the differential TADs in the bottom right group. The DIR tracks show the DIRs that gain or lose contact. The DIRs are also marked in the matrix with purple or black squares. The DMR tracks show DMRs that gain or lose CpG methylation. The DEG tracks show up- or down-regulated DEGs. The DMR_CTCF_pre track shows predicted CTCF binding sites that overlap with DMRs. The CTCF_pub_ChIP track shows the public CTCF ChIP-seq peaks. The UCSC annotation tracks show the UCSC gene annotation. (F) Contact frequency changes between DEG *SESN3* and super-enhancers on Chromosome 11 (position denoted on the X axis) in BWS IC1 GOM group. Two genes (not DEG) surrounding *SESN3* are also plotted for comparisons.

Chromosome architecture disruption in BWS and LOS

Differential chromosome compartments were detected between groups at 1 mb, 500 kb, 250 kb, 100 kb, 50 kb, and 25 kb resolutions for all the comparisons in Table 1 (Table S4). Differential TADs were detected between groups at 25kb resolution for all the comparisons in Table 1 (Table S5). The group mean Hi-C contact matrix was used for the differential TAD analysis instead of all individual samples in each group, due to software limitations. Types of differential TAD between groups are categorized as loss, strength change, merge, split, zoom, and complex (for description refer to Methods). In addition, differentially interacting regions (DIR) were detected for binned genomic regions between groups at 1 mb, 500 kb, 250 kb, 100 kb, and 50 kb resolutions for all the comparisons in Table 1 (Table S6), and changes are shown as gain and loss in contact frequency. Changes in chromosome architecture were associated with DEGs according to permutation tests. For example, upregulated DEGs in BWS with IC1 GOM group showed enrichment for differential compartment with increased eigenvalue

(change towards A compartment) at promoters, downregulated DEGs in female BWS with IC1 GOM group showed enrichment for differential TAD at promoters, and downregulated DEGs in LOS group showed enrichment for differential compartment with decreased eigenvalue at promoters (Table S2). Together with the previously mentioned association between DEGs and DMRs, these observations indicate that the causes of gene misregulation in BWS and LOS are diverse and there is not a single dominant mechanism.

For human BWS subtypes, BWS with IC1 GOM group has higher number of unique differential compartments with longer length (50 and 500 kb for 25th and 75th percentile, same format below) than BWS with IC2 LOM (50 and 300 kb) and BWS with CDKN1C variant (50 and 400 kb) groups when compared to the control group (Table 1 and S4). Increased number and decreased length of unique differential compartments were found in sex-specific comparisons, including male BWS with IC2 LOM (25 and 200 kb), female BWS with IC1 GOM (25 and 250 kb), female BWS with IC2 LOM (25 and 125 kb), and female BWS with CDKN1C variant (25 and 250 kb). For bovine, the lengths of unique differential compartments in the LOS group (250 kb and 1 mb) are longer than any of the human BWS subtypes. Unique differential compartments in male LOS (100 kb and 1 mb) are lower in number but longer in length than female LOS (25 and 250 kb), and in LOS maternal allele (25 and 500 kb) are also lower in number but longer in length than LOS paternal allele (25 and 75 kb).

Given the prevalence of differential compartments in BWS and LOS, Pearson correlation tests were performed between eigenvalue changes at gene promoters, which reflect A/B compartment changes, and expression changes of all protein-coding

genes and lncRNA (not just DEGs). Small RNA genes were not included in the test due to the uncertainty of their promoter locations. For genome-wide test, a very weak positive correlation was observed for both BWS (correlation coefficient equals to 0.078) and LOS (0.13; Figure 5D and 6D). The positive correlation increased in BWS (0.32) and LOS (0.29) when only considering the differential compartment regions, which indicates that the compartment changes have impacts on gene expression regulation in the differential compartment regions (Figure 5D and 6D). The same trend was detected in most other BWS and LOS involved comparisons excluding BWS with IC2 LOM, female LOS, and female LOS paternal allele (figure not shown). These exceptions have similar weak correlation at differential compartment regions as their genome-wide correlation, which indicates the compartment changes have less impact on gene expression in these groups. Interestingly, the differential compartments in these exceptions have a relatively shorter length as mentioned before. The Pearson correlations between eigenvalue changes and methylation changes at DMRs are generally very low (correlation coefficient absolute values are less than 0.05) in BWS and LOS involved comparisons (figure not shown).

BWS with *CDKN1C* variant group had the highest number (286) of differential TAD among BWS subtypes, while in LOS there were 369 differential TADs when compared to the control group (Table 1 and S5). For sex-specific comparisons, both female BWS with IC2 LOM (695) and female LOS (526) showed a higher differential TAD number than male BWS with IC2 LOM (239) and male LOS (231) respectively, when compared to their controls. This pattern could be related to the genome-wide sex-specific DNA methylation changes. As previously mentioned in bovine, DMRs are

mainly hypermethylated in male LOS and occur at regions that have intermediate to high DNA methylation level in the control group. Although these regions are enriched for predicted CTCF binding sites, the normally high DNA methylation level is expected to block the CTCF binding [85] which is seen as depletion of public CTCF ChIP-seq peaks in the heatmap (Figure S3B). And further hypermethylation could have minor impact on CTCF binding change in male LOS. In female LOS, DMRs also mainly occur at regions that have high DNA methylation level in the control group, and have enrichment for predicted CTCF binding sites and depletion for public CTCF ChIP-seq peaks (Figure S3C). However, the global loss of methylation could lead to activation of these potential CTCF binding sites and change the CTCF binding profile and TADs, which is reflected by the higher number of differential TAD detected in female LOS than male LOS. Similarly, male BWS with IC2 LOM also has a higher ratio of hypermethylated DMR than female BWS with IC2 LOM. Another interesting observation is that the boundaries for differential compartments are usually also the boundaries for TADs or sub-TADs, which means TAD is the unit for compartment changes. So, the more frequently disrupted TADs in female LOS may explain the observed higher number but shorter length of differential compartments in female LOS than male LOS, and that the potentially increased active CTCF binding sites in female LOS leads to smaller TADs and further causes fragmentation or length decrease of nearby differential compartments. These shortened differential compartments in female LOS have a much lower impact on gene expression change (correlation coefficient = 0.027) than male LOS (0.27).

Vulnerable chromosomes and chromosomal regions in BWS and LOS

The comparative genomic distribution of DMRs, differential compartments, DIRs, and DEGs are summarized in Figure 5C, 6C, S5, and S6. For humans, when ordering chromosomes by the ratio of total differential compartment length to chromosome length, chr2, 3, 6, 9, 10, 11, 12, 13, 17, 18, and 20 are often in the top ten list in BWS (Figure 5C and S5). It is worth noting that the differential compartments detected on chr9 and 18 are mainly around the unmappable region at the centromere. In sex-specific comparisons, chrX and Y also exhibit frequent compartment changes (figure not shown). For bovine, chr27, 21, 4, 10, 28, 8, 1, 14, 23, and 13 are the top ten in LOS group when ordered by the ratio of total differential compartment length to chromosome length (Figure 6C). Sex-specific comparisons show a somewhat different top ten list for male LOS (chr27, X, 26, 21, 28, 8, 1, 13, 4, 14) and female LOS (chr27, 4, 7, 16, 10, 14, 18, 5, 8, 22; Figure S6B and S6C). When comparing paternal allele to maternal allele in the control group, chr27, 9, 4, and 6 are the most different chromosomes regarding compartment (Figure S6D). Chr27 is largely altered in LOS paternal allele (Figure S6F) and male LOS paternal allele group (figure not shown). Compartment changes on chr9 become frequently detected in LOS involved sex-specific and allele-specific comparisons that female LOS paternal allele shows a similar pattern of change as in Figure S6D, and male LOS maternal allele shows an opposite pattern (figure not shown). Chr4 is largely altered in LOS paternal allele (Figure S6F) and female LOS paternal allele group (figure not shown). Interestingly, the differential compartment pattern on chr4 in LOS paternal allele is almost the opposite of the pattern of allelic comparison in control group (Figure S6D), which suggests the paternal allele of chr4 in LOS has a level of maternalization in chromosome compartment. Chr6 is largely

603 altered in LOS maternal allele (Figure S6E), male LOS maternal allele, female LOS
604 maternal allele, and female LOS paternal allele group (figure not shown).

605

606

607 **Figure 6. Characterization of bovine chromosome architecture and**
608 **alterations in large offspring syndrome (LOS). (A)** Length distribution of topologically
609 associating domains (TAD) detected in control group. **(B)** Genomic context enrichment
610 at control group TAD and subTAD boundaries (centered) and surrounding regions. **(C)**
611 Genome-wide distribution of differentially methylated regions (DMR), differential
612 compartments (diff_compartment), differential TAD (diff_TAD), differential interacting
613 regions (DIR) frequency, and differentially expressed genes (DEG) identified when
614 comparing LOS group with control group. **(D)** Correlation between chromosome
615 compartment changes (measured by eigenvalue) at gene promoters and log2
616 transformed fold change (logFC) of protein-coding genes and lncRNA. **(E)** Examples of
617 alterations of chromosome architecture associated with gene expression changes in
618 LOS fetuses. Detailed descriptions refer to Figure 5.

619

620

621 The maternal and paternal alleles show a similar large scale (chromosome-
622 wide) genomic distribution of allele-specific DMRs in LOS (Figure S6E and S6F). For
623 example, over 61% of DMRs on chr1 are found in the last 30 mb (19%) of this
624 chromosome in LOS maternal allele and LOS paternal allele. This enrichment is not
625 caused by a biased distribution of genomic variants since the methylation differences
626 between control paternal allele and control maternal allele are evenly distributed on this
627 chromosome (Figure S6D). However, as mentioned before, the precise genomic
628 coordinates of the DMRs are largely not shared (>80%) between maternal and paternal

alleles. This observation indicates that the two alleles of certain chromosomes or parts of a chromosome are vulnerable for DNA methylation alterations in LOS, but the alterations are largely not driven by the DNA sequence. The chromosome territory, which refers to the special location of chromosomes in the nucleus, could be involved in this vulnerability. According to human and mouse three-dimensional fluorescence in situ hybridization (3D-FISH) images, many chromosomes have their two alleles separated from each other [86,87]. Evidence have also shown that the two alleles of each chromosome have a symmetrical relative location to other chromosomes in mouse [88]. Assuming bovine has the same features, if changes of microenvironment at symmetrical spaces occur, they could have impacts on the same part of the two alleles. Other large scale chromosome markers may also directly or indirectly contribute to this vulnerability, such as histone modifications, which undergo rapid and dynamic reprogramming during early embryo development, and are known to regulate the *de novo* establishment of DNA methylation [89–91]. Future studies on chromosome territories in bovine will be needed to understand this process.

Given the very low correlation between DMRs and differential compartments and generally similar genomic distribution pattern of DMRs between maternal and paternal alleles in LOS, the distinct distribution pattern of differential compartments between maternal and paternal alleles suggests that other allele-specific alterations exist in LOS and are involved in the regulation of chromosome compartment, such as histone modifications and non-coding RNAs (ncRNAs). Chromosome compartment is considered mainly regulated by phase separation mechanism that chromatin regions with different physicochemical properties form distinct phases; phases with different

properties are spatially separated and phases with similar properties are clustered [92,93]. Histone modifications and ncRNAs have been found to attract binding proteins with intrinsically disordered domains that can lead to phase separation and compartment change [94–96]. In addition, studies in mouse have shown that the maternal and paternal alleles exhibit distinct histone modification profiles in early embryo, which may account for possible allele-specific disruptions [89,91]. Future studies on the histone modification profiles and ncRNA nuclear localization in BWS and LOS will help build the connection and improve the understanding of the underlying mechanisms.

Association among DMRs, altered chromosome architectures, and DEGs in BWS and LOS

Several genomic regions are shown to illustrate the possible impact of DNA methylation and chromosome architecture on gene expression changes in BWS and LOS (Figure 5E, 6E, S7, and S8). For humans, gene sestrin 3 (*SESN3*) is upregulated and located within a differential compartment region that changes toward A compartment in BWS with IC1 GOM group, and no DMR is found within gene body or promoter of *SESN3* (Figure 5E). Downregulated gene platelet derived growth factor subunit A (*PDGFA*) resides in an A compartment region that gains contact with a B compartment region in female BWS with IC2 LOM (Figure S7A). A small change toward B compartment is observed although not significant, and two hypomethylated DMRs overlap the gene body and promoter of *PDGFA*. Gene IQCJ-SCHIP1 readthrough (*IQCJ-SCHIP1*) is downregulated but located within a differential compartment region

that changes toward A compartment in female BWS with CDKN1C variant group, and the promoter and first exon of *IQCJ-SCHIP1* overlap with a hypermethylated DMR (chr3:158785716-158788035, 12.33% increase) which was excluded in Table S1 due to the 15% difference cutoff (Figure S7B). These changes are also shown in female comparison with the other two BWS subtypes (figure not shown), which indicates the alterations around *IQCJ-SCHIP1* are specific to this BWS subtype.

For bovine, gene KIT ligand (*KITLG*) is downregulated and located within a differential compartment region that changes toward B compartment in LOS group, and no DMR is found within gene body or promoter of *KITLG* (Figure 6E). The 5 mb upstream region of the *DLK1-DIO3* imprinted domain shows different patterns of chromosome compartment in male control and female control, with sex-specific changes in male LOS and female LOS (Figure S8). A gigantic TAD (4 mb) is observed mainly belonging to the B compartment in male control but not in female control. The upregulated gene BCL11 transcription factor B (*BCL11B*) resides in this 4 mb TAD which changes toward A compartment in male LOS group, and overlaps with multiple DMRs (mainly hypomethylation) in the gene body (Figure S8A). In female LOS, the 4 mb TAD is not altered, but a small region that changes toward B compartment overlaps with downregulated gene bradykinin receptor B2 (*BDKRB2*), and a hypermethylated DMR is found in its gene body (Figure S8B). We have previously reported dysregulation of multiple microRNAs in the *DLK1-DIO3* imprinted domain in both BWS and LOS [26], and further investigation will be needed to validate whether this is a consequence of compartment changes.

As mentioned before, DNA methylation at promoters and enhancers have

repressive impacts on gene expression through blocking binding sites for transcription factors [97]. Chromosome compartments also have impacts on gene expression. The B compartments are enriched for heterochromatin, mainly localized in the peripheral regions in the nucleus or surrounding the nucleoli, and isolated from areas with enrichment of transcription machineries (such as transcription factories) [98–102]. Oppositely, the A compartments are enriched for euchromatin, mainly localized in the interior of nucleus between the two B compartment regions and associated with transcription factories. The example loci above suggest a model that changes of DNA methylation and chromosome compartment in BWS and LOS can simultaneously affect gene expression in an independent manner. In the DEG examples of human *SESN3* and bovine *KITLG*, *BCL11B*, and *BDKRB2*, no DMR was found at the promoters of these DEGs, and their expression changes follow the alterations of their chromosome compartments that changing towards A compartment leads to upregulation and changing towards B compartment leads to downregulation. In the example of human DEG *PDGFA*, the hypomethylated DMR at its promoter will further open the promoter for transcription factor binding, however, its spatial change towards B compartment will reduce the availability of transcription machineries in its surrounding microenvironment. In the opposite example of human DEG *IQCJ-SCHIP1*, its spatial change towards A compartment will provide more transcription machineries, but the hypermethylated DMR at its promoter will prevent their binding. So, in these two cases, DNA methylation and chromosome compartment form an antagonistic situation where the one with more severe change or stronger effect might overcome the effect of the other on gene expression regulation. This model could explain some of the conflicts between promoter

DNA methylation changes and gene expression changes observed in our BWS and LOS datasets. In addition, alterations in TAD and chromosome compartment in BWS and LOS could change enhancer availability and contacts with gene promoters, and further affect gene expression. To investigate this mechanism, we embarked on the identification of long-range contacts between super-enhancers from dbSUPER database and promoter of the DEG *SESN3* in BWS IC1 GOM group as an example. *SESN3* showed increased promoter contacts with the super-enhancer region chr11:86430802-86457226, chr11:86168452-86236631, and chr11:85855833-85863446 from the closest upstream broad A compartment (Figure 5F). These super-enhancers could also contribute to the upregulation of *SESN3* in BWS.

Conservation between human and bovine

Next, we investigated the conservation of TAD and chromosome compartment between human and bovine genomes. For TAD in the control groups (with both sexes included) a corresponding genomic region of the majority of human TADs can be found in bovine (13,425 out of 13,912). Over these regions, 2,065 (~15.4%) human TADs were found conserved in bovine with greater than 90% overlapping in both species (Table S7). For sex specific analyses that include chrX, 1,960 (~13.7%) and 2,032 (~13.9%) human TADs were conserved in bovine for male and female control group, respectively (Figure 7A and Table S7). For chromosome compartments in the control groups (with both sexes included), continuous 25 kb bins with the same A or B compartment assignment were merged that results in 4,086 regions in human, and a corresponding genomic region can be found in bovine for most of them (3,713). Over

these regions, 1,813 (~48.8%) human compartments were found conserved in bovine with greater than 90% overlapping in any species (Table S7). For sex specific analyses that include chrX, 2,569 (~47.6%) and 1,989 (~46.8%) human compartments were conserved in bovine for male and female control group, respectively (Figure 7C and Table S7). For both TAD and chromosome compartment, the conservation pattern is similar between male and female for autosomes but is different for chromosome X, which is expected for inactivated X to show megadomain (lose TADs) and change towards B compartment in female.

Figure 7. Chromosome architecture conservations between human and bovine. (A) Genome-wide conservation of control group topologically associating domains (TAD) between human and bovine. Each line indicates a conserved TAD. Different colors refer to different human chromosomes. Mb = megabases. **(B)** Overlapped differential TAD identified in some comparisons of Beckwith-Wiedemann Syndrome (BWS) and large offspring syndrome (LOS). Information regarding comparisons used refer to Methods. Each line indicates a shared differential TAD between combined BWS and LOS results with greater than 90% overlapping in both species. Green and black refer to shared differential TAD with same and different type between BWS and LOS, respectively. **(C)** Genome-wide conservation of control group A/B compartment between human and bovine. Each line indicates a conserved compartment. Red and blue refer to A and B compartment, respectively. **(D)** Overlapped differential compartments identified in BWS and LOS. Each line indicates an overlapped differential compartment. Green and black refer to differential compartments with same or opposite direction of change between BWS and LOS, respectively. **(E)** and **(F)** An example region of conserved chromosome architecture in control group between human and bovine, and shared compartment changes between BWS and LOS. Detailed descriptions refer to Figure 5E.

Similar approaches were used to find shared differential compartments between BWS and LOS. For humans, the lists of differential compartments from BWS involved comparisons (index 1, 2, 3, 4, 8, 9, 10, 11, and 12) were combined, which resulted in 2,117 records (allowing overlap and identical records), and a corresponding genomic region can be found in bovine for 1,689 records. The combined bovine list from LOS involved comparison (index 16, 18, and 19) contains 1,118 records. In total, 772 (~45.7%) records of differential compartments from human overlap with bovine records without a cutoff for overlap percentage and regardless of direction of change (Figure 7D and Table S7). If the overlap cutoff in both species is set to 20%, 50% or 80%, 193 (~11.4%), 80 (~4.7%) or 28 (~1.7%) records will remain, respectively. Chromosome X, 6, 10, 9, 7, and 2 in human have the highest number of overlapped differential compartments, and in bovine the most frequent chromosomes are X, 8, 4, 23, 13, and 28 (Figure 7D). Similarly for differential TADs, 2,739 and 1,126 records were combined for BWS and LOS involved comparisons, respectively, and 2,635 human records have a corresponding genomic region found in bovine. And the overlapped record number is 153 (5.8%) with a 90% overlap cutoff in both species. Chromosome 1, 5, 3, 2, and 14 in human have the highest number of overlapped differential TADs, and in bovine the most frequent chromosomes are 7, 16, 1, 10, and 2 (Figure 7B). A 6 mb region is shown as an example to illustrate the TAD and compartment conservation and shared compartment changes between human and bovine (Figure 7E and 7F). In control groups, TADs and compartments are largely conserved in the right half of this region, and a 500 kb differential compartment region that changes toward B compartment is

shared by BWS (chr2:231000000-231500000) and LOS (chr2:118000000-118500000) groups. These shared changes of chromosome compartment and TAD between BWS and LOS will be worthwhile for further detailed investigation and have the potential to serve as molecular biomarkers for this syndrome.

The differential compartments identified in BWS and LOS show preference for conservation on certain chromosomes, and this preference indicates these chromosomes have some similar features that lead to vulnerabilities. As discussed before, the territories of chromosome could be a source of vulnerability for molecular disruptions. Among the human chromosomes that have the highest number of conserved differential compartments in BWS, chr6, 10, 9, and 7 have relatively close chromosome territories according to the human fibroblast 3D-FISH image [87]. Elucidating whether the corresponding chromosomes in bovine have such spatial closeness requires future studies on bovine chromosome territories.

Conclusion

In summary, our results suggest that genome-wide alternation of chromosome architecture, which is partially caused by DNA methylation changes, also contribute to the development of BWS and LOS. We show shared regions of dysregulation between BWS and LOS, including several *HOX* gene clusters, and also demonstrate that altered DNA methylation differs between the clinically epigenetically identified BWS patients and those identified as having DNA variants (i.e. *CDKN1C* microdeletion). Importantly, we demonstrate, for the first time, that subpopulations of IC2 LOM patients exists and can be further separated molecularly into two subpopulations. We also highlight

additional genes and genomic regions that have the potential to serve as targets for biomarker development to improve current molecular methodologies. With the advent of personalized medicine, the bovine model of overgrowth will be invaluable to be able to perform animal and in vitro studies, such as CRISPR/Cas9-based epigenetic editing, of regions identified as altered, in order to develop methodologies to ameliorate the clinical features and treat BWS.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human patients

Patient skin fibroblast samples and clinical information were collected through the BWS Registry, under the oversight of the Children's Hospital of Philadelphia (CHOP) Institutional Review Board protocol (IRB 13-010658) and in accordance with the Declaration of Helsinki. In brief, consent was obtained from all patients and/or legal guardians to collect longitudinal clinical information, in addition to samples that became available through clinical care and publish the findings. Patients were identified from the registry cohort and selected based on the availability of skin fibroblasts and their genetic or epigenetic type of BWS. Non-BWS samples (control) were collected for consideration of other growth difference (i.e. have an omphalocele but do not have BWS).

Patient skin fibroblasts were cultured as previously described [43]. Briefly, skin samples collected from regions either behind the ear or abdomen were split to process one chemically and other mechanically. Chemical disruption was performed using collagenase and the other was mechanically minced using a scalpel blade. Both explants were seeded into a T25 flask and supplemented with RPMI skin media (RPMI

with fetal bovine serum, penicillin–streptomycin antibiotic, and a final concentration of 2 mM L-glutamine). All flasks were incubated at 37°C for up to one month, with periodic media changes. Successful explant cultures were passaged for sustained growth and trypsinized to freeze and stored in liquid nitrogen. Clinical testing for BWS was performed at the University of Pennsylvania Genetic Diagnostic Laboratory as previously described [103].

The identifier and sex of patient included in this study were as follows; 1) control: CON_H_#1 to #3 (female), CON_H_#4 and #5 (male); 2) BWS with IC2 loss of methylation: BWS_#1_IC2 to BWS_#6_IC2 (female), BWS_#7_IC2 to BWS_#12_IC2 (male); 3) BWS with IC1 gain of methylation: BWS_#13_IC1 to BWS_#15_IC1 (female); 4) BWS with *CDKN1C* variant: BWS_#16_Mut (male according to record but female according to sequencing results), BWS_#17_Mut and BWS_#18_Mut (female). The age of fibroblast sample collection (in days with gestational age correction) for each group are control (average 139, range 1-505), BWS with IC2 loss of methylation (327, 1-1283), BWS with IC1 gain of methylation (1274, 116-3459), BWS with *CDKN1C* variant (206, 6-488).

Bovine animals

Day 105 *Bos taurus indicus* (*B. t. indicus*; Brahman breed) x *Bos taurus taurus* (*B. t. taurus*; Angus breed) F1 hybrid fetuses were generated by our laboratory in 2019 and used as tissue donors [19,44]. This breeding strategy introduced genetic variants between the maternal and paternal alleles in the for allele-specific analyses. The control group was generated using artificial insemination (AI) and the ART group was generated

by *in vitro* production procedures. The LOS group was defined as individuals from the ART group with body weight greater than 97th centile of controls (548g). Conceptuses were collected by caesarean section to maintain nucleic acid integrity. Fetal skin fibroblast primary cell lines were established and maintained as described in our previous study [44]. The identifier, original ID, sex, and body weight of fetuses used in this study were as follows; 1) control fetuses: CON_B_#1 (original ID 533, female, 388g), CON_B_#2 (647, female, 396g), CON_B_#3 (640, male, 448g), CON_B_#4 (648, male, 466g), and CON_B_#5 (527, male, 514g); 2) LOS fetuses: LOS_#1 (616, female, 638g), LOS_#2 (656, female, 704g), LOS_#3 (604B, female, 986g), LOS_#4 (512, male, 648g), LOS_#5 (602, male, 752g), and LOS_#6 (664, male, 1080g).

All the animal procedures were approved by the University of Missouri Animal Care and Use Committee under protocol 9455. Trained personnel and Veterinarians performed all animal handling and surgeries.

METHOD DETAILS

In this manuscript, all the chromosomal coordinates for bovine genome refers to assembly ARS-UCD1.2, and for human genome refers to GRCh37 (hg19) [104].

Genomic DNA extraction

Genomic DNA from human skin fibroblasts was isolated using the AllPrep DNA/RNA Micro Kit (QIAGEN) per the manufacturer's instructions, as previously described [43]. Briefly, the cells were lysed in the Buffer RLT Plus. The lysate was then homogenized into a QIAshredder spin column. The homogenized lysate was then

transferred to an AllPrep DNA spin column, washed with washing buffer and then DNA was eluted using elution buffer. The concentration of DNA was measured by using a NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific). Genomic DNA samples were stored at -20°C.

Bovine fibroblast cells were lysed in lysis buffer (0.05 M Tris-HCl (pH 8.0), 0.1 M EDTA, and 0.5% (w/v) SDS) with proteinase K (Fisher BioReagents, BP1700) at 55°C for four hours. Genomic DNA was extracted with Phenol:Chloroform:Isoamyl Alcohol (SIGMA, P3803) following the manufacturer's instructions. The concentration of DNA was measured by using a NanoDrop® ND-1000 Spectrophotometer and DNA integrity was confirmed by electrophoresis on a 0.7% agarose gel. Genomic DNA samples were stored at -20°C.

RNA isolation

Total RNA from human skin fibroblasts was isolated using the AllPrep DNA/RNA Micro Kit (QIAGEN) per the manufacturer's instructions, as previously described [43]. Briefly, the cells were lysed in the Buffer RLT Plus. The lysate was then homogenized into a QIAshredder spin column. The homogenized lysate was then transferred to an AllPrep DNA spin column, and the flowthrough was used for RNA extraction using 70% ethanol. The concentration of RNA was measured by using a NanoDrop® ND-1000 Spectrophotometer and RNA samples were stored at -80°C.

Total RNA was isolated from bovine fibroblast cells using TRIzol™ Reagent (Invitrogen, 15596026) following the manufacturer's instructions. The concentration of RNA was measured by using a NanoDrop® ND-1000 Spectrophotometer. RNA samples

were stored at -80°C.

CTCF binding site prediction

Potential CTCF binding sites were predicted globally for human and bovine genome using TFBSTools 1.26.0 [105] with database JASPAR2020 [106]. CTCF motifs of 'vertebrates' were used for prediction and min.score was set to 81%.

Hi-C sequencing and data analyses

Hi-C sequencing for fibroblast cells was conducted by CD Genomics. Information on library preparation and sequencing obtained from the company is as follows: Cells were cross-linked with 1% formaldehyde at room temperature and quenched with glycine. The cross-linked cells were subsequently lysed. Endogenous nucleases were inactivated with 0.3% SDS, then chromatin DNA were digested by MboI (NEB), marked with biotin-14-dCTP (Invitrogen), and ligated by T4 DNA ligase (NEB). After reversing cross-links, the ligated DNA was extracted through QIAamp DNA Mini Kit (Qiagen) according to manufacturers' instructions. Purified DNA was sheared to 300-500 bp fragments and were further blunt-end repaired, A-tailed and adaptor-added, followed by purification and PCR amplification. Finally, equimolar pooling of libraries was performed based on QC values and sequenced on Novaseq6000 sequencer (Illumina) with a read length configuration of 150 PE for 1,000M PE reads per sample (1,000M in each direction).

The processing of Hi-C raw sequencing reads was based on the 'Hi-C Processing Pipeline' from 4D Nucleome Data Portal

(https://data.4dnucleome.org/resources/data-analysis/hi_c-processing-pipeline). Briefly, reads were aligned to the corresponding reference genome using bwa 0.7.17 with parameters 'mem -SP5M' [107]. Reads with mapping quality (MAPQ) equal or greater than 30 were kept for analyses. Then, pairtools 0.3.0 was used to parse, sort, and deduplicate aligned reads, and pair types 'UU', 'UR', and 'RU' were selected [108]. Cooler and Juicer were used to convert contact matrix to .cool and .hic files [109,110].

Hi-C data was also used to identify genomic variants following the pipeline for 1000 bull genome project [111]. Briefly, Read groups were added using AddOrReplaceReadGroups function of picard 2.25.5 [112] for aligned Hi-C sequencing reads. Duplicated reads were marked using MarkDuplicates function of picard with parameter '--OPTICAL_DUPLICATE_PIXEL_DISTANCE 2500'. For human, the dataset of known variants was acquired from GATK resource bundle (https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle). For bovine, the dataset of known variants was acquired from the 1000 bull genome project, namely ARS1.2PlusY_BQSR_v3.vcf.gz. GATK 4.2.1.0 [113] was used to recalibrate base quality and identify genomic variants in the Hi-C data with the known variant dataset as reference. Parameters used for BaseRecalibrator were '–bqsr-baq-gap-open-penalty 45 --read-filter MappingQualityReadFilter --minimum-mapping-quality 20' for bovine. For human, the same settings were used except with default value for --bqsr-baq-gap-open-penalty. In addition, for bovine, our previously generated genomic sequencing data for semen DNA of the bull used to sire all the fetuses in this study was included for joint genotyping in order to assign alleles for the genomic variants (GEO: GSE197130) [44]. HaplotypeCaller was used with parameter '-ERC GVCF', and

GenomicsDBImport and GenotypeGVCFs were used with default settings. For bovine, raw SNP and INDEL were separately scored using VariantRecalibrator with parameter '--resource:1000G,known=false,training=true,truth=true,prior=10.0 known_variant -an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR' was used. For human, parameter '--resource:hapmap,known=false,training=true,truth=true,prior=15.0 known_variant1 --resource:omni,known=false,training=true,truth=true,prior=12.0 known_variant2 --resource:1000GSNP,known=false,training=true,truth=true,prior=10.0 known_variant3 --resource:1000GINDEL,known=false,training=true,truth=true,prior=10.0 known_variant4 --resource:dbsnp,known=true,training=false,truth=false,prior=2.0 known_variant5 -an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR' was used. Scored variants were filtered using ApplyVQSR with parameter '--truth-sensitivity-filter-level 99.0'.

For bovine, variants used for allele assignment met the following criteria: 1) marked as PASS by ApplyVQSR, 2) GQ>=40 in the bull and the fetus, 3) DP>=10 in the bull, 4) no read supports not assigned alleles in the bull and the fetus, 5) only one allele of the bull is shown in the fetus. In addition, variants on chrX identified in male fetuses were excluded from analyses, since which is likely due to the lack of chrY in the reference genome. Alleles were assigned for bovine Hi-C reads based on qualified variants using custom Perl scripts. For SNP, the nucleotide in the reads at the SNP location was directly used for allele assignment. For INDEL, local maternal and paternal allele pseudo genomes were generated, and the reads were realigned to both pseudo genomes. The reads were assigned to the alleles with more aligned bases, or less

970 mismatches if aligned base numbers were the same. Next, variants with biased allelic
971 ratio of Hi-C reads (one allele is less than 15%) were excluded from analyses. Lastly,
972 reads that cover multiple variants and have conflicting allele assignments were
973 excluded from the analyses.

974 Statistical analyses for detecting differentially interacting regions (DIR),
975 differential TAD, and differential chromosome compartments were performed. The sex
976 chromosomes were only analyzed in sex-specific comparisons to circumvent
977 confounding created by X chromosome inactivation. R package multiHiCcompare
978 1.12.0 was used to identify DIR at 1mb, 500kb, 250kb, 100kb, and 50kb resolutions
979 between groups [114]. The make_hicexp function was used with default parameters
980 except 'remove.regions = NULL', and cyclic_loess and hic_exactTest functions were
981 used with default settings. False discovery rate (FDR) was controlled at 0.05 for
982 significance.

983 For visualization and TAD detection for each group, the size of contact matrix
984 was normalized to the first sample in the Control group for both human (CON_H_#1)
985 and bovine (CON_B_#1), and group mean contact matrix was calculated based on
986 normalized matrix of each sample in the group. R package SpectralTAD 1.10.0 was
987 used with parameter 'qual_filter = TRUE, z_clust = FALSE, levels = level, resolution =
988 25000, window_size = window_size, gap_thresh = 0.2, min_size = 6' to identify TAD
989 structure ranging from 150kb to 4mb at 25kb resolution [115]. Level and window_size
990 was initially set to 10 and 120, respectively, and reduced by 1 and 30 each time if error
991 reported. In addition, diffDomain was used with function and parameter 'dvsd multiple --
992 min_nbin 1 --hicnorm NONE' and 'adjustment fdr_bh' to detect differential TAD

boundaries at 25kb resolution between groups based on TAD identified by SpectralTAD [116]. Since SpectralTAD and diffDomain only support single sample comparison instead of group comparison, to reduce the impacts of sample variations within each group and obtain more consistent differences between groups, filtering for inconsistent TADs in the control groups was performed for detected differential TADs. Inconsistent TADs were obtained by applying diffDomain between each sample and its group mean contact matrix and merging all the differential results from each sample. It is worth noting that TAD comparisons between BWS subtypes were not filtered and resulted in much larger number of differential TAD since there was no control group involved. Types of differential TAD include loss, strength change, merge, split, zoom, and complex. Strength change refers to the situation that the same TAD is detected in the two groups for comparing but the chromosomal contact frequency is different within the TAD. TAD with strength change could reflect forming or losing small sub-TADs and loops within it that requires higher resolution and cannot be properly annotated with the sequencing depth in this study. TAD with strength change could also be a result of chromosome compartment change of the sub-TADs within it. Zoom refers to the extension or shrink of a TAD in another group. Complex refers to the situation that cannot be simply categorized into any other type.

dcHiC was used to assign chromosome compartments and identify differential compartments at 1mb, 500kb, 250kb, 100kb, 50kb, and 25kb resolutions between groups [117]. Default settings for cis, select, analyze, and subcomp functions of dcHiC were used, and FDR was controlled at 0.05 for significance during the analyze step. The quantile normalized principal component (eigenvalue) calculated at 25kb resolution

was used for assigning compartments and plotting.

Whole genome bisulfite sequencing (WGBS) and data analyses

WGBS for fibroblast cells was conducted by CD Genomics. Information on library preparation and sequencing obtained from the company is as follows: Genomic DNA was fragmented by sonication to a mean size of approximately 200-400 bp. Fragmented DNA was end-repaired, 5'-phosphorylated, 3'-dA-tailed and then ligated to methylated adapters. The methylated adapter-ligated DNAs were purified using 0.8× Agencourt AMPure XP magnetic beads and subjected to bisulfite conversion by ZYMO EZ DNA Methylation-Gold Kit (Zymo Research). The converted DNAs then have the addition of 8-bp index primers to create the final cDNA library. Equimolar pooling of libraries was performed based on QC values and sequenced on Novaseq6000 sequencer (Illumina) with a read length configuration of 150 PE for 300M PE reads per sample (300M in each direction). 0.1-1% lambda DNA were added during the library preparation to monitor bisulfite conversion rate.

For WGBS data, raw sequencing reads were deduplicated using the clumpify function of BBMap 38.90 [118] and trimmed for adapter sequences and low quality bases using trimmomatic 0.39 [119] with parameters 'ILLUMINACLIP:adapter_seq:2:30:10:1:true LEADING:20 TRAILING:20 AVGQUAL:20 MAXINFO:0:0.5'. Trimmed reads were aligned to the corresponding reference genome using BSBolt 1.5.0 [120] with default setting. Trimmed reads were also aligned to lambda phage genome to determine bisulfite conversion rates. Samtools 1.13 [121] was used to convert, sort, filter, and index bam files. MarkDuplicates function of picard

2.25.5 [112] was used with parameter '--OPTICAL_DUPLICATE_PIXEL_DISTANCE 2500' to further remove duplicated reads after alignment. CpG methylation information was extracted from the bam files using the function CallMethylation of BSBolt with parameter '-CG -min 1 -BQ 20'. For both human and bovine samples, CpG methylation at C/T SNPs detected in their corresponding Hi-C data were excluded for sense strand reads, and CpG methylation at G/A SNPs were excluded for antisense strand reads. For bovine samples, SNPs identified from Hi-C dataset were also used to assign alleles for WGBS reads using custom Perl scripts. C/T SNPs (based on sense strand sequence) were only applied for antisense strand reads, and G/A SNPs were only applied for sense strand reads. Other types of SNP were applied for both strands. Reads that cover multiple SNPs and have conflicting allele assignments were excluded. For both overall and allele-specific comparisons, statistical analyses were conducted using R package hummingbird [122] with parameter 'minCpGs = 10, minLength = 100, maxGap = 300' to identify differentially methylated regions (DMRs) between groups. DMRs with at least 15% difference in methylation level (both gain and loss of methylation) and at least 4 mean read coverage at CpG sites in all the samples were reported. For allele-specific analyses, the read coverage filtering was not applied since only about 20% of the usable SNPs are shared by all the sample. The sex chromosomes were only analyzed in sex-specific comparisons to circumvent confounding created by X chromosome inactivation.

Total RNA sequencing (total RNA-seq) and data analyses

Total RNA-seq for fibroblast cells was conducted by CD Genomics. Information

on library preparation and sequencing obtained from the company is as follows: The first step involves the removal of ribosomal RNA using RiboZero kit (Illumina). Subsequently, the RNA is fragmented into small pieces using divalent cations under elevated temperatures. The cleaved RNA fragments are copied into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis. These cDNA fragments then have the addition of a single 'A' base and subsequent ligation of the adapter. The products are purified and enriched with PCR to create the final cDNA library. This results in libraries with inserts ranging in size from 120–200 bp with a median size of 150 bp. Equimolar pooling of libraries was performed based on QC values and sequenced on Novaseq6000 sequencer (Illumina) with a read length configuration of 150 PE for 33M PE reads per sample (33M in each direction).

All read data was assessed for quality with FastQC 0.11.7 and low-quality bases (Phred scores ≤ 20) were trimmed using the dynamictrim function of SolexaQA++ 3.1.7.1. Reads less than 60 bases in length were removed using SolexaQA++ LengthSort. Known splice sites were extracted from the gtf file of each reference genome using the Python script hisat2_extract_splice_sites.py. Hisat2 2.1.0 was used to establish a genome index with the hisat2-build and also to align the quality trimmed reads to the reference genome with the following adjusted parameters: --mp 6,6; –score-min L,0,-0.2; known-splicesite-infile. Sam files were converted to bam files and coordinate-sorted using Samtools 1.8. Then, Stringtie 2.1.4 was used with the sorted bam files to assemble transcriptomes for each sample. Stringtie merge was used to combine the assembled transcripts from all samples into a reference assembly. Assembled transcripts were extracted using the gffread module in cufflinks 2.2.1. An

index was built for the merged transcriptome and reads were quantified using Salmon 0.11.1. IsoformSwitchAnalyeR 1.17.04 was used to rescue stringtie annotations and generate a gene count matrix.

For alternative splicing (AS) event analyses, the merged Stringtie annotation file and coordinate-sorted bam files were used, and AS events were called using rMATs 4.0.1 with parameters '-t paired –readlength 150 -cstat 0.0001 -libType fr-firststrand'. FDR was controlled at 0.05 for significance. Manhattan plots of were generated using R package Sushi [123].

To assign allele specificity for total RNA-seq reads in bovine, variants were called from the aligned data according to GATK's Best Practices for Calling Variants in RNA-seq data. Duplicate reads were removed and read groups were assigned using Picard 2.22.4. A dictionary was created for the reference genome using GATK's CreateSequenceDictionary. GATK SplitNCigarReads was used to hard clip sequences extending into intronic regions and GATK BaseRecalibrator to recalibrate base quality scores. Then, HaplotypeCaller was run on calibrated data and called variants were filtered with adjusted parameters (AB < 0.2 & MQ0 > 50). Variants that were shared between genomic and RNA-seq data for each sample were kept. Variants that were heterozygous in the fetus and homozygous in the bull were used to assign alleles. A variant annotation file was created using the annotate function of bcftools 1.8 and gene annotations were added using custom scripts. Custom R scripts were used to retrieve maternal and paternal allele depth read counts from VCF files in order to quantify allele-specific expressions.

Statistical analyses were conducted using R package edgeR 3.36.0 [124] to

identify differentially expressed genes (DEGs) between groups. Parameters used include ‘method = "TMM"’ for calcNormFactors function and ‘robust = TRUE’ for glmFit function. FDR was controlled at 0.05 for significance.

Functional correlation between gene expression and BWS clinical diagnosis

To determine the co-expressional network that gathered by genes showing similar trends across samples, WGCNA was performed in R [125]. Briefly, gene abundance matrix (normalized expression matrix) from IsoformSwitchAnalyeR was log2 transformed and only genes with high expressional variability values (top 60%, 15304 genes retained) were remained for identification of co-expression network and a soft threshold β value was set as 10 according to the criterion of approximate scale-free topology. BWS specific clinical diagnosis was generated through the evaluated features from Wang et al reported study [83]. Clinical BWS score consists of the combination of cardinal and suggestive features, with one more point (plus two points) was given for each diagnosed cardinal feature, while one point was given for each diagnosed suggestive feature. The correlation between eigengene (a representative of the gene expression profiles in a module) in each module and clinical diagnoses was calculated using Pearson’s correlation method, and only relationships with p-value < 0.05 were considered as significantly positive/negative correlation. Hub genes within BWS-related module were identified with both absolute gene-BWS score significance ≥ 0.5 and absolute module membership ≥ 0.8.

Correlation analyses between gene expression and chromosome compartment

Pearson correlation analyses between gene expression changes and eigenvalue changes at gene promoter were performed using the R base function 'cor'.

Genomic content overlap and enrichment analyses

Gene annotation information was obtained from NCBI (GCF_002263795.1_ARS-UCD1.2_genomic.gtf and GCF_000001405.25_GRCh37.p13_genomic.gtf) [126]. Repeated and overlapped exons were merged for each gene, and introns were calculated based on merged exons. Promoters (1kb) were calculated based on transcription start sites annotation and only included protein coding genes and long non-coding RNAs. Annotation of CpG islands and repeated sequences were obtained from UCSC Genome Browser [127]. Locations of CpG shores (flanking 2kb from CpG islands) and shelves (flanking 2-4kb from the CpG island) were calculated based on CpG island annotation. Human public CTCF ChIP-seq data of fibroblast primary cells from different tissues were obtained from the ENCODE portal with the following identifiers: ENCFF148VQH (lung), ENCFF337WIE (foreskin), ENCFF344UBE (cardiac), ENCFF438XHB (mammary gland), ENCFF570FLB (lung), ENCFF638OUB (dermis), ENCFF652MLT (foreskin), ENCFF734DZF (lung), ENCFF738CXX (villous mesenchyme), ENCFF859PRV (aortic adventitia), and ENCFF933VBD (pulmonary artery) [128]. Peaks from these files were merged to use as common CTCF sites. Bovine public CTCF ChIP-seq data of different tissues (no fibroblast primary cell data available) were obtained from the Gene Expression Omnibus (GEO) database with the following identifiers: GSE158430 (replicate for adipose, cerebellum, cortex, hypothalamus, liver, lung, muscle, and

spleen) and GSE129423 (control rumen epithelial primary cells) [129]. GSE129423 was based on genome build UMD3.1.1 and was converted to ARS-UCD1.2 using liftOver from UCSC Genome Browser with parameter '-minMatch=0.1'. Due to lower quality (i.e. occasionally very broad peaks) of these data than the human data, peak regions found in at least four files were kept as common CTCF sites. Bedtools and custom Perl scripts were used to identify overlapped genomic contexts for regions with altered DNA methylation, gene expression, and chromosome architecture, and calculate observed frequencies (Obs) [130]. Permutation tests were performed to determine expected frequencies (Exp) of overlapping by randomly shuffling genomic contexts across the genome for 10,000 times. For each test, the p value was calculated as p = n(|Exp - mean(Exp)| ≥ |Obs - mean(Exp)|)/10,000. EnrichedHeatmap was also used for enrichment visualization in the heatmaps and different flanking distance was chosen based on the length of genomic content examined to best show the background level and enrichment pattern [131].

The genomic coordinates of human super-enhancers were obtained from database dbSUPER and used for finding enhancer-promoter contact for gene *SESN3* at 10kb resolution with Python library bioframe [132,133].

Conservation analyses

For TAD/compartment conservation analyses between human and bovine control groups, the genomic coordinates of human TAD/compartment were converted to bovine coordinates using liftOver from UCSC Genome Browser with parameter '-minMatch=0.1'. Then, custom Perl scripts were used to identify bovine

TAD/compartment overlapping with human TAD/compartment that the overlapped region is greater than 90% in both species for TAD and in any species for compartment.

To identify shared differential compartments between BWS and LOS, for each comparison, bedtools was first used to merge significant results of different resolutions into non-overlapping regions [130]. Merged results of comparisons that compare BWS/LOS to control group were combined for each species. For human, BWS vs Control (index 1 in Table 1), BWS_IC1_GOM vs Control (2), BWS_IC2_LOM vs Control (3), BWS_CDKN1C_Mut vs Control (4), BWS_IC2_LOM_Male vs Control_Male (8), BWS_Female vs Control_Female (9), BWS_IC1_GOM_Female vs Control_Female (10), BWS_IC2_LOM_Female vs Control_Female (11), BWS_CDKN1C_Mut_Female vs Control_Female (12) were used for analyses. For bovine, LOS vs Control (16), LOS_Male vs Control_Male (18), and LOS_Female vs Control_Female (19) were used for analyses. Then human coordinates were converted to bovine using liftOver and regions overlapped with bovine differential compartments were identified using custom Perl scripts.

Plot generation

R package ComplexHeatmap, EnrichedHeatmap, plotgardener, circlize, methylKit, ggpubr, ggfortify, ggplot2, and Python library matplotlib were used for making plots [131,134–141].

Data availability

All the raw sequencing data for human samples, including whole genome

bisulfite sequencing, total RNA sequencing, and Hi-C sequencing, are available in the database of Genotypes and Phenotypes (dbGaP) with accession number (phs001794.v1.p1). The subject ID of each sample used in dbGaP can be found in Table S8. Some of the patients in this study were also included in our previous study and deposited under the same dbGaP subject ID [26]. All the raw sequencing data for bovine samples, including whole genome bisulfite sequencing, total RNA sequencing, and Hi-C sequencing, are available in the Gene Expression Omnibus (GEO) database with accession number (pending).

Acknowledgements

We acknowledge the ENCODE Consortium and the John Stamatoyannopoulos (UW) and Bradley Bernstein (Broad) laboratories that generated the human CTCF ChIP-seq datasets. This research was partially funded by the United States Department of Agriculture-Agriculture and Food Research Initiative (USDA-AFRI), grant 2018-67015-27598. This work was partially supported by National Cancer Institute [K08 CA193915], Alex's Lemonade Stand Foundation, a Damon Runyon Clinical Investigator Award provided by the Damon Runyon Cancer Research Foundation (105–19), the Lorenzo "Turtle" Sartini Jr. Endowed Chair in Beckwith-Wiedemann Syndrome Research, and the Victoria Fertitta Fund through the Lorenzo "Turtle" Sartini Jr. Endowed Chair in Beckwith-Wiedemann Syndrome Research (J.M.K.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Y.L., J.M.K, D.E.H., J.C., and R.M.R. designed research; Y.L. and S.N. performed research; Y.L., P.X., F.B., A.K.G., analyzed data; Y.L., P.X., F.B., A.K.G., S.N., and R.M.R. wrote the paper; all authors edited the paper.

Declaration of interests

The authors declare no competing interests.

Authors' ORCID

Yahan Li, yl5c7@mail.missouri.edu, ORCID 0000-0003-3062-4626
Ping Xiao, ping.xiao@okstate.edu, ORCID 0000-0001-5678-501X
Frimpong Boadu, fbqc9@missouri.edu, ORCID 0000-0002-4464-6191
Anna Goldkamp, anna.goldkamp@okstate.edu, ORCID 0000-0002-1510-465X
Snehal Nirgude, nirgudes@chop.edu, ORCID 0000-0001-8471-1384
Jianlin Cheng, chengji@missouri.edu, ORCID 0000-0003-0305-2853
Darren Hagen, darren.hagen@okstate.edu, ORCID 0000-0001-8295-020X
Jennifer Kalish, kalishj@chop.edu, ORCID 0000-0003-1500-9713
Rocío Melissa Rivera, riverarm@missouri.edu, ORCID 0000-0002-9832-9618

References

1. Rivera RM, Donnelly CG, Patel BN, Li Y, Soto-Moreno EJ. Abnormal Offspring Syndrome. Bovine Reproduction. 2021; 876–895.

2. Brioude F, Kalish JM, Mussa A, Foster AC, Bliek J, Ferrero GB, et al. Clinical and molecular diagnosis, screening and management of Beckwith–Wiedemann syndrome: an international consensus statement. Nature Reviews Endocrinology. 2018;14: 229–249.

3. Farin PW, Piedrahita JA, Farin CE. Errors in development of fetuses and placentas from in vitro-produced bovine embryos. Theriogenology. 2006;65: 178–91. doi:10.1016/j.theriogenology.2005.09.022

4. Li Y, Donnelly CG, Rivera RM. Overgrowth syndrome. Veterinary Clinics: Food Animal Practice. 2019;35: 265–276.

5. Mussa A, Russo S, De Crescenzo A, Chiesa N, Molinatto C, Selicorni A, et al. Prevalence of Beckwith–Wiedemann syndrome in north west of Italy. American journal of medical genetics Part A. 2013;161: 2481–2486.

6. Li Y, Sena Lopes J, Fuster PC, Rivera RM. Spontaneous and ART-induced large offspring syndrome: similarities and differences in DNA methylome. Epigenetics. 2022; 1–20.

7. Chen Z, Robbins KM, Wells KD, Rivera RM. Large offspring syndrome: a bovine model for the human loss-of-imprinting overgrowth syndrome Beckwith-Wiedemann. Epigenetics. 2013;8: 591–601. doi:10.4161/epi.24655

8. Hori N, Nagai M, Hirayama M, Hirai T, Matsuda K, Hayashi M, et al. Aberrant CpG methylation of the imprinting control region KvDMR1 detected in assisted reproductive technology-produced calves and pathogenesis of large offspring syndrome. Animal reproduction science. 2010;122: 303–312.

9. Lazzari G, Wrenzycki C, Herrmann D, Duchi R, Kruip T, Niemann H, et al. Cellular and molecular deviations in bovine in vitro-produced embryos are related to the large offspring syndrome. Biology of reproduction. 2002;67: 767–75.

10. Behboodi E, Anderson GB, BonDurant RH, Cargill SL, Kreuscher BR, Medrano JF, et al. Birth of large calves that developed from in vitro-derived bovine embryos. Theriogenology. 1995;44: 227–32.

11. Farin PW, Farin CE. Transfer of bovine embryos produced in vivo or in vitro: survival and fetal development. Biology of reproduction. 1995;52: 676–682.

12. McEvoy TG, Sinclair KD, Broadbent PJ, Goodhand KL, Robinson JJ. Post-natal growth and development of Simmental calves derived from in vivo or in vitro embryos. Reproduction, Fertility and Development. 1998;10: 459–464.

13. Farin PW, Crosier AE, Farin CE. Influence of in vitro systems on embryo survival and fetal development in cattle. Theriogenology. 2001;55: 151–170.

14. Hasler JF, Henderson WB, Hurtgen PJ, Jin ZQ, McCauley AD, Mower SA, et al. Production, freezing and transfer of bovine IVF embryos and subsequent calving results. Theriogenology. 1995;43: 141–152.

15. van Wagtendonk-de Leeuw AM, Aerts BJ, den Daas JH. Abnormal offspring following in vitro production of bovine preimplantation embryos: a field study. Theriogenology. 1998;49: 883–94.

16. Sinclair KD, Broadbent PJ, Dolman DF. In vitro produced embryos as a means of achieving pregnancy and improving productivity in beef cows. Animal Science. 1995;60: 55–64.

17. Kruip TA, Den Daas JHG. In vitro produced and cloned embryos: effects on pregnancy, parturition and offspring. Theriogenology. 1997;47: 43–52.

18. van Wagtendonk-de Leeuw AM, Mullaart E, de Roos AP, Merton JS, den Daas JH, Kemp B, et al. Effects of different reproduction techniques: AI MOET or IVP, on health and welfare of bovine offspring. Theriogenology. 2000;53: 575–97.

19. Rivera RM, Goldkamp AK, Patel BN, Hagen DE, Soto-Moreno EJ, Li Y, et al. Identification of large offspring syndrome during pregnancy through ultrasonography and maternal blood transcriptome analyses. Scientific Reports. 2022;12: 10540. doi:10.1038/s41598-022-14597-w

20. Nava-Trujillo H, Rivera RM. Large offspring syndrome in ruminants: current status and prediction during pregnancy. animal. 2023;17: 100740.

21. Chen Z, Hagen DE, Ji T, Elsik CG, Rivera RM. Global misregulation of genes largely uncoupled to DNA methylome epimutations characterizes a congenital overgrowth syndrome. Scientific reports. 2017;7: 12667.

22. Chen Z, Hagen DE, Elsik CG, Ji T, Morris CJ, Moon LE, et al. Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. Proc Natl Acad Sci U S A. 2015;112: 4618–23. doi:10.1073/pnas.1422088112

23. Sangalli JR, Chiaratti MR, De Bem THC, de Araujo RR, Bressan FF, Sampaio RV, et al. Development to term of cloned cattle derived from donor cells treated with valproic acid. PLoS One. 2014;9: e101022.

24. Su J, Wang Y, Liu Q, Yang B, Wu Y, Luo Y, et al. Aberrant mRNA expression and DNA methylation levels of imprinted genes in cloned transgenic calves that died of large offspring syndrome. Livestock Science. 2011;141: 24–35.

25. Su H, Li D, Hou X, Tan B, Hu J, Zhang C, et al. Molecular structure of bovine Gtl2 gene and DNA methylation status of Dlk1-Gtl2 imprinted domain in cloned bovines. Animal reproduction science. 2011;127: 23–30.

26. Li Y, Hagen DE, Ji T, Bakhtiarizadeh MR, Frederic WM, Traxler EM, et al. Altered microRNA expression profiles in large offspring syndrome and Beckwith-Wiedemann syndrome. Epigenetics. 2019; 1–27.

27. Krzyzewska IM, Alders M, Maas SM, Bliek J, Venema A, Henneman P, et al. Genome-wide methylation profiling of Beckwith-Wiedemann syndrome patients without molecular confirmation after routine diagnostics. Clinical epigenetics. 2019;11: 53.

28. Tee L, Lim DH, Dias RP, Baudement M-O, Slater AA, Kirby G, et al. Epimutation profiling in Beckwith-Wiedemann syndrome: relationship with assisted reproductive technology. Clinical epigenetics. 2013;5: 1–10.

29. Tenorio J, Romanelli V, Martin-Trujillo A, Fernández G-M, Segovia M, Perandones C, et al. Clinical and molecular analyses of Beckwith–Wiedemann syndrome: Comparison between spontaneous conception and assisted reproduction techniques. American journal of medical genetics Part A. 2016;170: 2740–2749.

30. Rossignol S, Steunou V, Chalas C, Kerjean A, Rigolet M, Viegas-Pequignot E, et al. The epigenetic imprinting defect of patients with Beckwith—Wiedemann syndrome born after assisted reproductive technology is not restricted to the 11p15 region. Journal of medical genetics. 2006;43: 902–907.

31. Cooper WN, Luharia A, Evans GA, Raza H, Haire AC, Grundy R, et al. Molecular subtypes and phenotypic expression of Beckwith–Wiedemann syndrome. European journal of human genetics. 2005;13: 1025.

32. Mussa A, Russo S, De Crescenzo A, Freschi A, Calzari L, Maitz S, et al. (Epi) genotype–phenotype correlations in Beckwith–Wiedemann syndrome. European journal of human genetics. 2016;24: 183.

33. Wadhwa PD, Buss C, Entringer S, Swanson JM. Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. Seminars in reproductive medicine. © Thieme Medical Publishers; 2009. pp. 358–368.

34. Niakan KK, Han J, Pedersen RA, Simon C, Pera RAR. Human pre-implantation embryo development. Development. 2012;139: 829–841.

35. Mussa A, Molinatto C, Cerrato F, Palumbo O, Carella M, Baldassarre G, et al. Assisted Reproductive Techniques and Risk of Beckwith-Wiedemann Syndrome. Pediatrics. 2017;140. doi:10.1542/peds.2016-4311

36. Erik Mullaart, Geart Ludema, Helga Flapper, John Hepburn, Björn Oback. Birthweight data of IVP calves. 39th Annual Meeting of the Association of Embryo Technology in Europe Animal Reproduction. 2023;2. Available: https://animal-reproduction.org/article/64dfa00da953954a9b16f7a5

37. Fauser BC, Devroey P, Diedrich K, Balaban B, Bonduelle M, Delemarre-van de Waal HA, et al. Health outcomes of children born after IVF/ICSI: a review of current expert opinion and literature. Reprod Biomed Online. 2014;28: 162–82. doi:10.1016/j.rbmo.2013.10.013

38. Urrego R, Rodriguez-Osorio N, Niemann H. Epigenetic disorders and altered gene expression after use of assisted reproductive technologies in domestic cattle. Epigenetics. 2014;9: 803–815.

39. Sunderam S, Kissin DM, Zhang Y, Jewett A, Boulet SL, Warner L, et al. Assisted reproductive technology surveillance—United States, 2017. MMWR Surveillance Summaries. 2020;69: 1.

40. Joao Viana. 2020 Statistics of embryo production and transfer in domestic farm animals. Embryo Technology Newsletter. 2021;39.

41. Rivera RM, Stein P, Weaver JR, Mager J, Schultz RM, Bartolomei MS. Manipulations of mouse embryos prior to implantation result in aberrant expression of imprinted genes on day 9.5 of development. Human molecular genetics. 2008;17: 1–14.

42. Rovina D, La Vecchia M, Cortesi A, Fontana L, Pesant M, Maitz S, et al. Profound alterations of the chromatin architecture at chromosome 11p15. 5 in cells from Beckwith-Wiedemann and Silver-Russell syndromes patients. Scientific reports. 2020;10: 1–19.

43. Naveh NSS, Deegan DF, Huhn J, Traxler E, Lan Y, Weksberg R, et al. The role of CTCF in the organization of the centromeric 11p15 imprinted domain interactome. Nucleic Acids Research. 2021;49: 6315–6330.

44. Li Y, Boadu F, Highsmith MR, Hagen DE, Cheng J, Rivera RM. Allele-specific aberration of imprinted domain chromosome architecture associates with large offspring syndrome. iScience. 2022;25: 104269. doi:10.1016/j.isci.2022.104269

45. Li X, Li Y, Bai S, Zhang J, Liu Z, Yang J. NR2F1-AS1/miR-140/HK2 axis regulates hypoxia-induced glycolysis and migration in hepatocellular carcinoma. Cancer Management and Research. 2021; 427–437.

46. Wen X, Zhang Q, Zhou L, Li Z, Wei X, Yang W, et al. Intrachromosomal Looping and Histone K27 Methylation Coordinately Regulates the lncRNA H19-Fetal Mitogen IGF2 Imprinting Cluster in the Decidual Microenvironment of Early Pregnancy. Cells. 2022;11: 3130.

47. Paço A, de Bessa Garcia SA, Freitas R. Methylation in HOX clusters and its applications in cancer therapy. Cells. 2020;9: 1613.

48. Brzezinski J, Choufani S, Romao R, Shuman C, Chen H, Cunanan J, et al. Clinically and biologically relevant subgroups of Wilms tumour defined by genomic and epigenomic analyses. British Journal of Cancer. 2021;124: 437–446.

49. Mallo M, Alonso CR. The regulation of Hox gene expression during animal development. Development. 2013;140: 3951–3963.

50. Wang K, Liang Q, Li X, Tsoi H, Zhang J, Wang H, et al. MDGA2 is a novel tumour suppressor cooperating with DMAP1 in gastric cancer and is associated with disease outcome. Gut. 2016;65: 1619–1631.

51. Yang Z, Chen H, Huo L, Yang Z, Bai Y, Fan X, et al. Epigenetic inactivation and tumor-suppressor behavior of NGFR in human colorectal cancer. Molecular Cancer Research. 2015;13: 107–119.

52. Vishnubalaji R, Yue S, Alfayez M, Kassem M, Liu F-F, Aldahmash A, et al. Bone morphogenetic protein 2 (BMP2) induces growth suppression and enhances chemosensitivity of human colon cancer cells. Cancer cell international. 2016;16: 1–12.

53. Choi SK, Kim HS, Jin T, Moon WK. LOXL4 knockdown enhances tumor growth and lung metastasis through collagen-dependent extracellular matrix changes in triple-negative breast cancer. Oncotarget. 2017;8: 11977.

54. Abraham V, Cao G, Parambath A, Lawal F, Handumrongkul C, Debs R, et al. Involvement of TIMP-1 in PECAM-1-mediated tumor dissemination. International Journal of Oncology. 2018;53: 488–502.

55. Henle AM, Nassar A, Puglisi-Knutson D, Youssef B, Knutson KL. Downregulation of TAP1 and TAP2 in early stage breast cancer. PloS one. 2017;12: e0187323.

56. Tao R, Xiong X, Liangpunsakul S, Dong XC. Sestrin 3 protein enhances hepatic insulin sensitivity by direct activation of the mTORC2-Akt signaling. Diabetes. 2015;64: 1211–1223.

57. Chen Y, Huang T, Yu Z, Yu Q, Wang Y, Hu J, et al. The functions and roles of sestrins in regulating human diseases. Cellular & Molecular Biology Letters. 2022;27: 1–24.

58. Denley A, Cosgrove LJ, Booker GW, Wallace JC, Forbes BE. Molecular interactions of the IGF system. Cytokine & growth factor reviews. 2005;16: 421–439.

59. Lau MM, Stewart CE, Liu Z, Bhatt H, Rotwein P, Stewart CL. Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes & development. 1994;8: 2953–2963.

60. Vantaku V, Putluri V, Bader DA, Maity S, Ma J, Arnold JM, et al. Epigenetic loss of AOX1 expression via EZH2 leads to metabolic deregulations and promotes bladder cancer progression. Oncogene. 2020;39: 6265–6285.

61. Khaled WT, Choon Lee S, Stingl J, Chen X, Raza Ali H, Rueda OM, et al. BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. Nature communications. 2015;6: 5987.

62. Shivapurkar N, Stastny V, Okumura N, Girard L, Xie Y, Prinsen C, et al. Cytoglobin, the newest member of the globin family, functions as a tumor suppressor gene. Cancer research. 2008;68: 7448–7456.

63. Wu X, Li C, Wang Z, Zhang Y, Liu S, Chen S, et al. A bioinformatic analysis study of m7G regulator-mediated methylation modification patterns and tumor microenvironment infiltration in glioblastoma. BMC cancer. 2022;22: 1–17.

64. Zhang H, Liang J, Lu T, Li M, Shan G, Bi G, et al. AGRN promotes lung adenocarcinoma progression by activating Notch signaling pathway and acts as a therapeutic target. Pharmacological Research. 2023; 106819.

65. Yang X, Liu H, Ye T, Ye Z. High expression of LEF1 correlates with poor prognosis in solid tumors, but not blood tumors: a meta-analysis. Bioscience Reports. 2020;40: BSR20202520.

66. Wang T, Dang N, Tang G, Li Z, Li X, Shi B, et al. Integrating bulk and single-cell RNA sequencing reveals cellular heterogeneity and immune infiltration in hepatocellular carcinoma. Molecular Oncology. 2022;16: 2195–2213.

67. Lu L, Pandey AK, Houseal MT, Mulligan MK. The genetic architecture of murine glutathione transferases. PLoS One. 2016;11: e0148230.

68. Lim K-S, Chang S-S, Choi B-H, Lee S-H, Lee K-T, Chai H-H, et al. Genome-Wide analysis of allele-specific expression patterns in seventeen tissues of Korean Cattle (Hanwoo). Animals. 2019;9: 727.

69. Li SM, Valo Z, Wang J, Gao H, Bowers CW, Singer-Sam J. Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. PloS one. 2012;7: e31751.

70. de Souza MM, Zerlotini A, Rocha MIP, Bruscadin JJ, Diniz WJ da S, Cardoso TF, et al. Allele-specific expression is widespread in Bos indicus muscle and affects meat quality candidate genes. Scientific Reports. 2020;10: 10204.

71. Edwards CA, Watkinson WM, Telerman SB, Hulsmann LC, Hamilton RS, Ferguson-Smith AC. Reassessment of weak parent-of-origin expression bias shows it rarely exists outside of known imprinted regions. Elife. 2023;12: e83364.

72. Huo H-N, Zhang C, Wang K, Chen W-N, Zhang Y-J, Yu W-L, et al. A novel imprinted locus on bovine chromosome 18 homologous with human Chromosome 16q24. 1. 2023.

73. Lim K-S, Kim H-C, Choi B-H, Son J-W, Lee K-T, Choi T-J, et al. Identification of Monoallelically Expressed Genes Associated with Economic Traits in Hanwoo (Korean Native Cattle). Animals. 2021;12: 84.

74. Cao W, Douglas KC, Samollow PB, VandeBerg JL, Wang X, Clark AG. Origin and evolution of marsupial-specific imprinting clusters through lineage-specific gene duplications and acquisition of promoter differential methylation. Molecular Biology and Evolution. 2023;40: msad022.

75. Huang W-C, Ferris E, Cheng T, Hörndli CS, Gleason K, Tamminga C, et al. Diverse non-genetic, allele-specific expression effects shape genetic architecture at the cellular level in the mammalian brain. Neuron. 2017;93: 1094-1109. e7.

76. Xie J, Ji T, Ferreira MA, Li Y, Patel BN, Rivera RM. Modeling allele-specific expression at the gene and SNP levels simultaneously by a Bayesian logistic mixed regression model. BMC bioinformatics. 2019;20: 530.

77. Schulze KV, Szafranski P, Lesmana H, Hopkin RJ, Hamvas A, Wambach JA, et al. Novel parent-of-origin-specific differentially methylated loci on chromosome 16. Clinical Epigenetics. 2019;11: 1–10.

78. Chen Z, Hagen DE, Wang J, Elsik CG, Ji T, Siqueira LG, et al. Global assessment of imprinted gene expression in the bovine conceptus by next generation sequencing. Epigenetics. 2016;11: 501–16. doi:10.1080/15592294.2016.1184805

79. Du J, Li Z, Li Q-Z, Guan T, Yang Q, Xu H, et al. Enoyl coenzyme a hydratase domain–containing 2, a potential novel regulator of myocardial ischemia injury. Journal of the American Heart Association. 2013;2: e000233.

80. Riedemann J, Macaulay VM. IGF1R signalling and its inhibition. Endocrine-related cancer. 2006;13: S33–S43.

81. Maunakea AK, Chepelev I, Cui K, Zhao K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell research. 2013;23: 1256–1269.

82. Heallen T, Zhang M, Wang J, Bonilla-Claudio M, Klysik E, Johnson RL, et al. Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. Science. 2011;332: 458–461.

83. Wang KH, Kupa J, Duffy KA, Kalish JM. Diagnosis and management of Beckwith-Wiedemann syndrome. Frontiers in pediatrics. 2020;7: 562.

84. Tian M, Liu W, Zhang Q, Huang Y, Li W, Wang W, et al. MYSM1 represses innate immunity and autoimmunity through suppressing the cGAS-STING pathway. Cell Reports. 2020;33.

85. Wiehle L, Thorn GJ, Raddatz G, Clarkson CT, Rippe K, Lyko F, et al. DNA (de) methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. Genome research. 2019;29: 750–761.

86. Markaki Y, Smeets D, Fiedler S, Schmid VJ, Schermelleh L, Cremer T, et al. The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. Bioessays. 2012;34: 412–426.

87. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol. 2005;3: e157.

88. Han Z, Cui K, Placek K, Hong N, Lin C, Chen W, et al. Diploid genome architecture revealed by multi-omic data of hybrid mice. Genome research. 2020;30: 1097–1106.

89. Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature. 2016;537: 553–557.

90. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nature Reviews Genetics. 2009;10: 295–304.

91. Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nature cell biology. 2018;20: 620–631.

92. Ryu J-K, Hwang D-E, Choi J-M. Current understanding of molecular phase separation in chromosomes. International Journal of Molecular Sciences. 2021;22: 10736.

93. Hildebrand EM, Dekker J. Mechanisms and Functions of Chromosome Compartmentalization. Trends in Biochemical Sciences. 2020.

94. Quinodoz SA, Jachowicz JW, Bhat P, Ollikainen N, Banerjee AK, Goronzy IN, et al. RNA promotes the formation of spatial compartments in the nucleus. Cell. 2021.

95. Wang L, Gao Y, Zheng X, Liu C, Dong S, Li R, et al. Histone modifications regulate chromatin compartmentalization by contributing to a phase separation mechanism. Molecular cell. 2019;76: 646-659. e6.

96. Bian Q, Anderson EC, Yang Q, Meyer BJ. Histone H3K9 methylation promotes formation of genome compartments in Caenorhabditis elegans via chromosome compaction and perinuclear anchoring. Proceedings of the National Academy of Sciences. 2020;117: 11459–11470.

97. Li E, Zhang Y. DNA methylation in mammals. Cold Spring Harbor perspectives in biology. 2014;6: a019133.

98. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY). 2009;326: 289–293.

99. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159: 1665–1680.

100. Karbassi E, Rosa-Garrido M, Chapski DJ, Wu Y, Ren S, Wang Y, et al. Direct visualization of cardiac transcription factories reveals regulatory principles of nuclear architecture during pathological remodeling. Journal of molecular and cellular cardiology. 2019;128: 198–211.

101. Robson MI, Jose I, Czapiewski R, Sivakumar A, Kerr AR, Schirmer EC. Constrained release of lamina-associated enhancers and genes from the nuclear envelope during T-cell activation facilitates their association in chromosome compartments. Genome research. 2017;27: 1126–1138.

102. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature. 2017;544: 59–64.

103. Baker SW, Duffy KA, Richards-Yutz J, Deardorff MA, Kalish JM, Ganguly A. Improved molecular detection of mosaicism in Beckwith-Wiedemann Syndrome. Journal of medical genetics. 2021;58: 178–184.

104. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 2020;9: giaa021.

105. Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinformatics (Oxford, England). 2016;32: 1555–1556.

106. Fornes O, Castro-Mondragon JA, Khan A, Van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic acids research. 2020;48: D87–D92.

107. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics. 2009;25: 1754–1760.

108. Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, et al. Pairtools: from sequencing data to chromosome contacts. bioRxiv. 2023; 2023.02. 13.528389.

109. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell systems. 2016;3: 95–98.

110. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics. 2020;36: 311–316.

111. Hayes BJ, Daetwyler HD. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. Annual review of animal biosciences. 2019;7: 89–102.

112. Broad Institute. Picard toolkit. Broad Institute, GitHub repository. Broad Institute, GitHub Repository; 2021. Available: http://broadinstitute.github.io/picard/

113. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media; 2020.

114. Stansfield JC, Cresswell KG, Dozmorov MG. multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. Bioinformatics. 2019;35: 2916–2923.

115. Cresswell KG, Stansfield JC, Dozmorov MG. SpectralTAD: an R package for defining a hierarchy of topologically associated domains using spectral clustering. BMC bioinformatics. 2020;21: 1–19.

116. Hua D, Gu M, Du Y, Qi L, Du X, Bai Z, et al. DiffDomain enables identification of structurally reorganized topologically associating domains. Research in Computational Molecular Biology: 26th Annual International Conference, RECOMB 2022, San Diego, CA, USA, May 22–25, 2022, Proceedings. Springer; 2022. pp. 302–303.

117. Chakraborty A, Wang JG, Ay F. dcHiC detects differential compartments across multiple Hi-C datasets. Nature Communications. 2022;13: 6827.

118. Bushnell B. BBMap. 2021. Available: sourceforge.net/projects/bbmap/

119. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30: 2114–2120.

120. Farrell C, Thompson M, Tosevska A, Oyetunde A, Pellegrini M. BiSulfite Bolt: A bisulfite sequencing analysis platform. GigaScience. 2021;10: giab033.

121. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.

122. Ji T. A Bayesian hidden Markov model for detecting differentially methylated regions. Biometrics. 2019;75: 663–673.

123. Phanstiel DH, Boyle AP, Araya CL, Snyder MP. Sushi. R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics. 2014;30: 2808–2810.

124. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England). 2010;26: 139–140.

125. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9: 1–13.

126. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research. 2016;44: D733-45. doi:10.1093/nar/gkv1189

127. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome research. 2002;12: 996–1006.

128. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489: 57.

129. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research. 2002;30: 207–210.

130. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842.

131. Gu Z, Eils R, Schlesner M, Ishaque N. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC genomics. 2018;19: 1–7.

132. Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic acids research. 2016;44: D164–D171.

133. Open2C, Abdennur N, Fudenberg G, Flyamer I, Galitsyna AA, Goloborodko A, et al. Bioframe: operations on genomic intervals in pandas dataframes. bioRxiv. 2022; 2022.02. 16.480748.

134. Wickham H. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics. 2011;3: 180–185.

135. Gu Z. Complex heatmap visualization. iMeta. 2022;1: e43.

136. Kramer NE, Davis ES, Wenger CD, Deoudes EM, Parker SM, Love MI, et al. Plotgardener: Cultivating precise multi-panel figures in R. Bioinformatics. 2022;38: 2042–2045.

137. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology. 2012;13: 1–9.

138. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30: 2811–2812.

139. Kassambara A. ggpubr:“ggplot2” based publication ready plots. R package version 04 0. 2020;438.

140. Tang Y, Horikoshi M, Li W. ggfortify: unified interface to visualize statistical results of popular R packages. R J. 2016;8: 474.

141. Hunter JD. Matplotlib: A 2D graphics environment. Computing in science & engineering. 2007;9: 90–95.

Table 1. Summary of statistical comparisons

								Diff_compartment	
Species	Type	Index	Comparison	DMR	DEG	DIR*	Diff_TAD	Raw*	Unique
Human	Overall	1	BWS vs Control	5708	12	1489	191	482	83
		2	BWS_IC1_GOM vs Control	21124	23	702	105	950	157
		3	BWS_IC2_LOM vs Control	7865	8	1399	144	743	125
		4	BWS_IC2_LOM_S1 vs Control	23582	742	54921	256	2050	328
		5	BWS_IC2_LOM_S2 vs Control	1915	2	116	55	1909	532
		6	BWS_CDKN1C_Mut vs Control	37122	10	833	286	551	105
		7	BWS_IC2_LOM vs BWS_IC1_GOM	6535	4	445	2197	471	148
		8	BWS_CDKN1C_Mut vs BWS_IC1_GOM	66575	53	2563	3706	1252	276
		9	BWS_CDKN1C_Mut vs BWS_IC2_LOM	7169	8	1149	1985	560	150
		10	BWS_IC2_LOM_S1 vs BWS_IC2_LOM_S2	20657	484	68396	2508	4592	750
	Sex specific	11	BWS_IC2_LOM_Male vs Control_Male	53488	12	289	239	2623	486
		12	BWS_Female vs Control_Female	12700	7	600	434	926	157
		13	BWS_IC1_GOM_Female vs Control_Female	26034	20	682	105	1529	314
		14	BWS_IC2_LOM_Female vs Control_Female	26459	43	7867	695	1743	448
		15	BWS_CDKN1C_Mut_Female vs Control_Female	12320	19	1719	540	996	242
		16	BWS_IC2_LOM_Female vs BWS_IC1_GOM_Female	13100	3	218	2783	952	354
		17	BWS_CDKN1C_Mut_Female vs BWS_IC1_GOM_Female	69325	31	437	3059	1441	289
		18	BWS_CDKN1C_Mut_Female vs BWS_IC2_LOM_Female	27140	56	598	1055	966	243
Bovine	Overall	19	LOS vs Control	10860	112	20200	369	2328	192
		20	LOS_All vs Control	4071	6	566	216	1010	129
	Sex specific	21	LOS_Male vs Control_Male	75370	109	10299	231	3941	294
		22	LOS_Female vs Control_Female	34270	213	4671	526	2737	632
		23	LOS_All_Female vs Control_Female	26189	23	495	339	1489	471
	Allele specific	24	LOS_Maternal vs Control_Maternal	17611	29	1000	681	716	231
		25	LOS_Paternal vs Control_Paternal	15708	16	1115	651	1550	624
		26	Control_Paternal vs Control_Maternal	10143	122	4940	105	2367	1038
		27	LOS_All_Maternal vs Control_Maternal	15405	2	275	709	307	147
		28	LOS_All_Paternal vs Control_Paternal	14061	0	291	689	1112	511
	Sex and allele specific	29	LOS_Male_Maternal vs Control_Male_Maternal	26690	28	899	339	2146	838
		30	LOS_Male_Paternal vs Control_Male_Paternal	21682	25	1177	261	1566	520
		31	LOS_Female_Maternal vs Control_Female_Maternal	9286	19	1980	96	2609	1491
		32	LOS_Female_Paternal vs Control_Female_Paternal	7364	9	1737	66	4681	2830
		33	Control_Male_Paternal vs Control_Male_Maternal	1322	54	2262	23	1063	542
		34	Control_Female_Paternal vs Control_Female_Maternal	1302	14	1104	16	3551	2171
		35	LOS_All_Female_Maternal vs Control_Female_Maternal	11118	4	1035	138	1674	1139
		36	LOS_All_Female_Paternal vs Control_Female_Paternal	8819	0	956	92	3799	2440

DMR = differentially methylated regions; DEG = differentially expressed genes; DIR = differential interacting regions; diff_TAD = differential topologically associating domains; Diff_compartment = differential compartments; Unique = after merging overlapped and continuous regions with the same direction of change; *Count includes results from all the bin sizes

Table 1

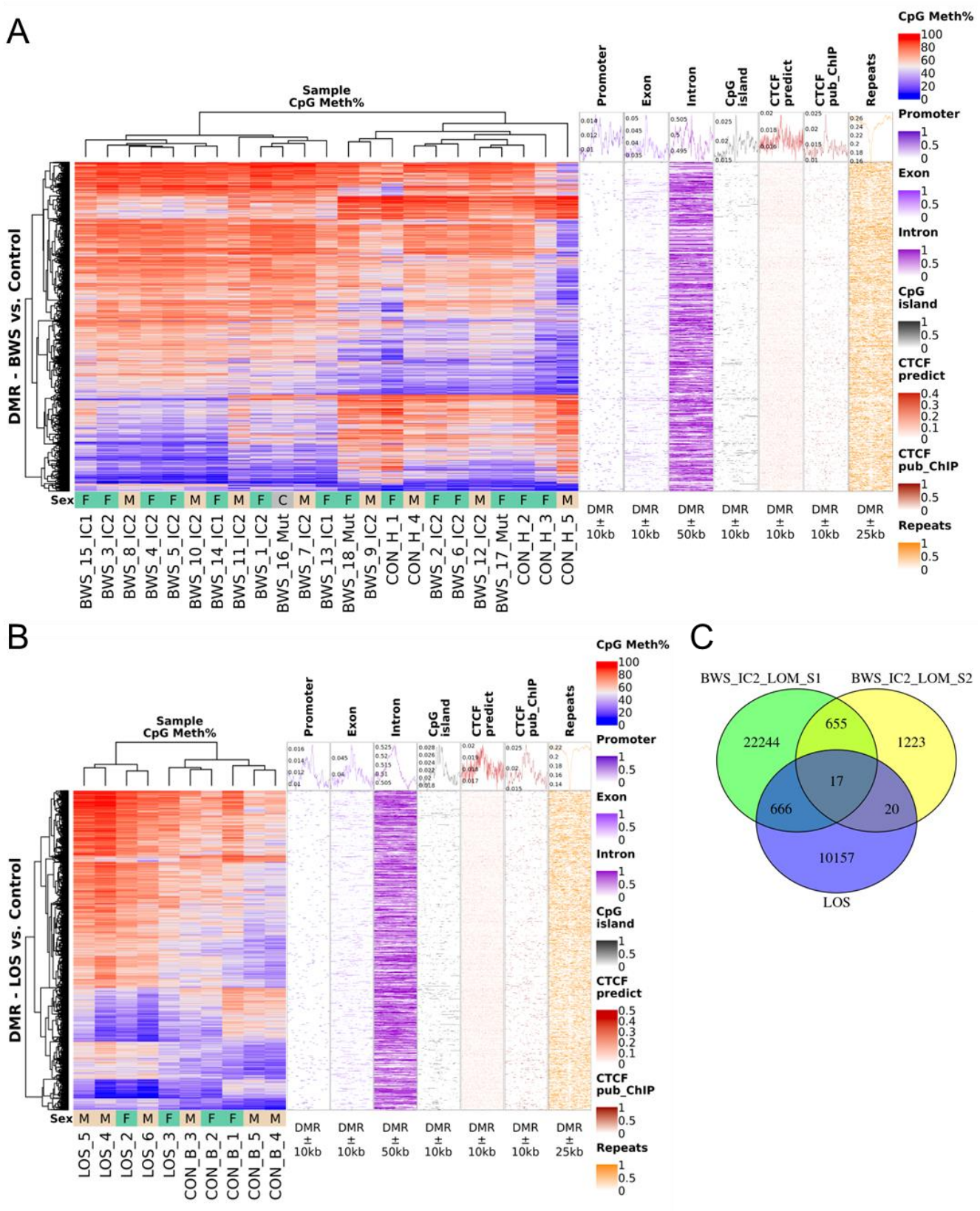

Figure 1

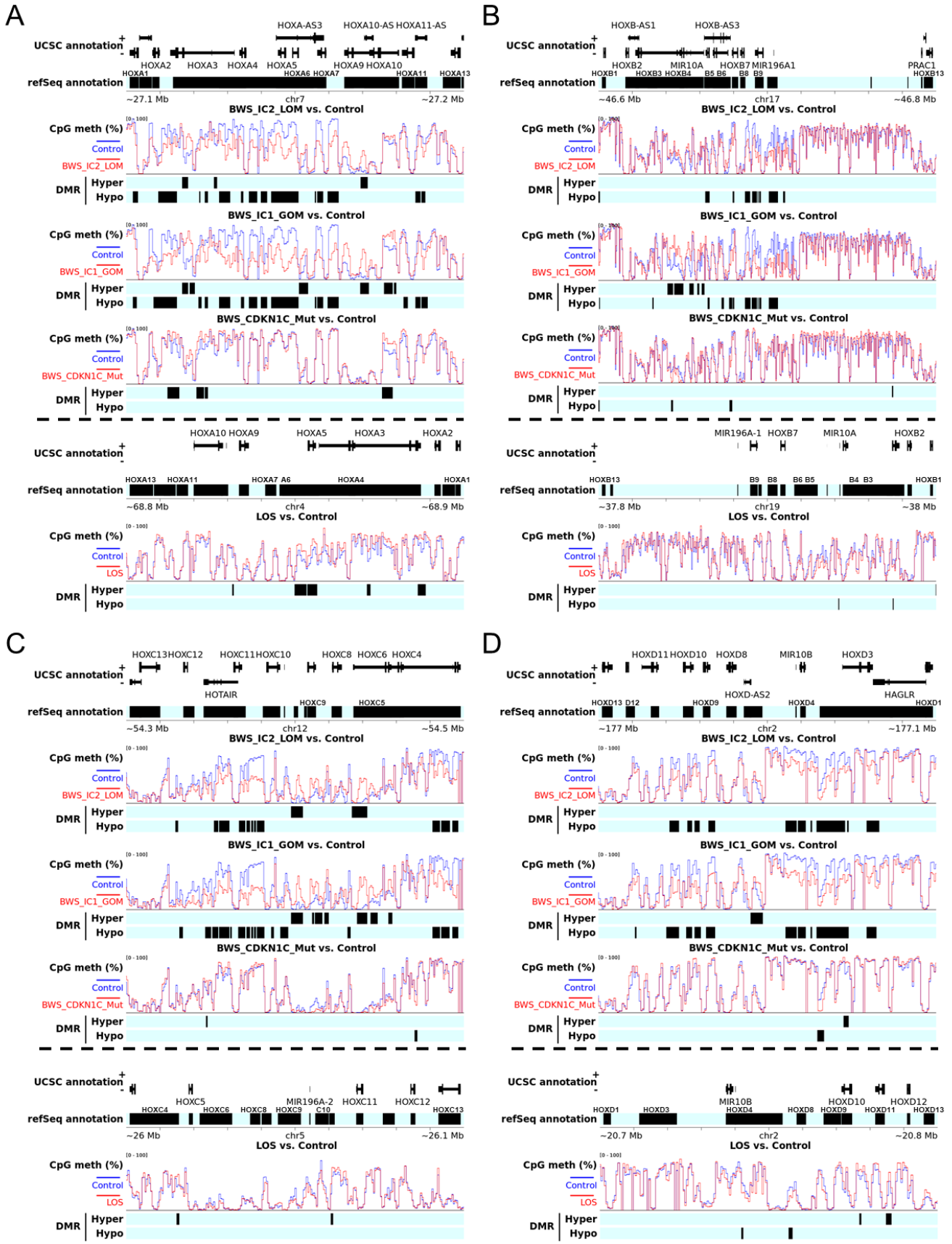

Figure 2

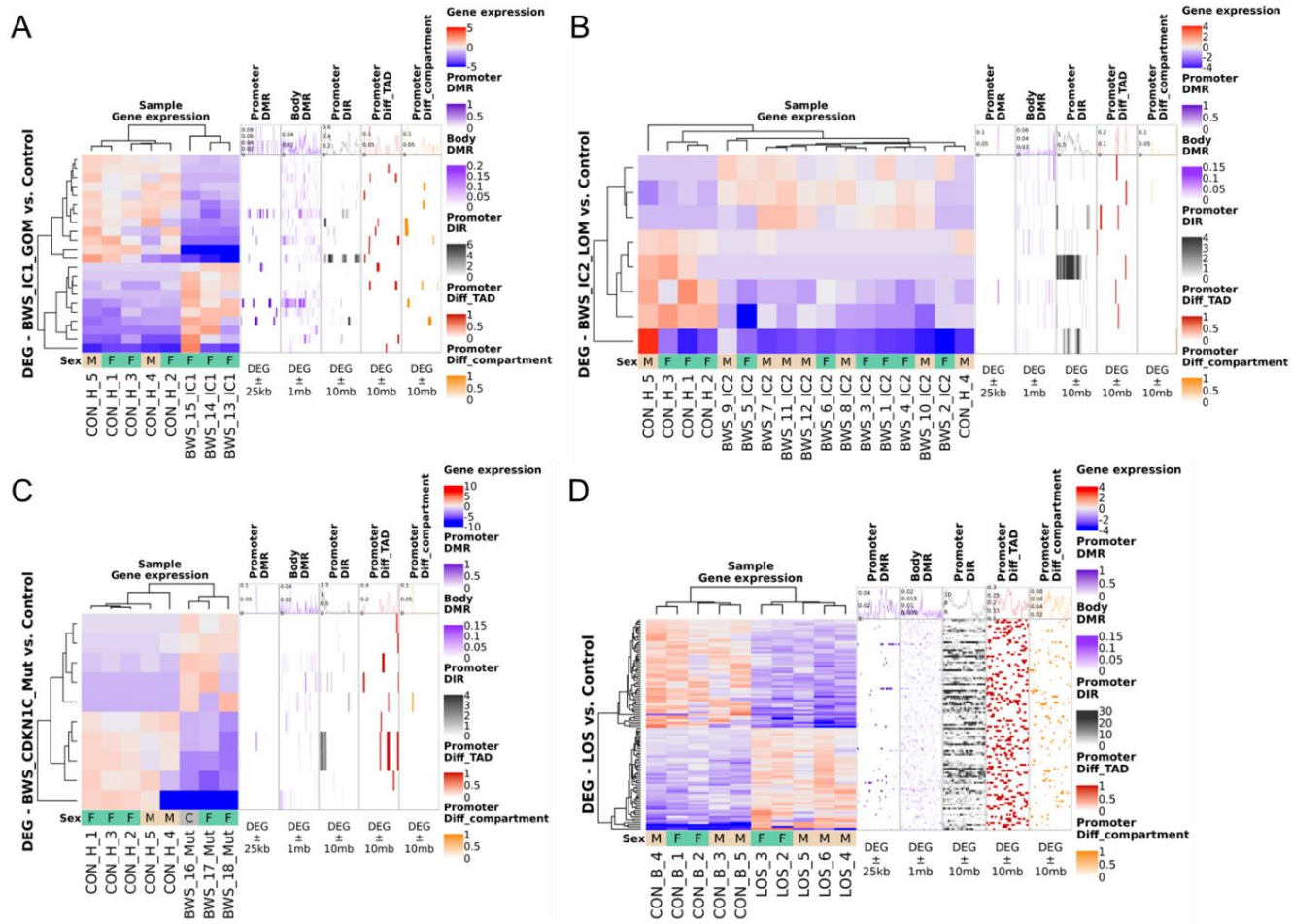

Figure 3

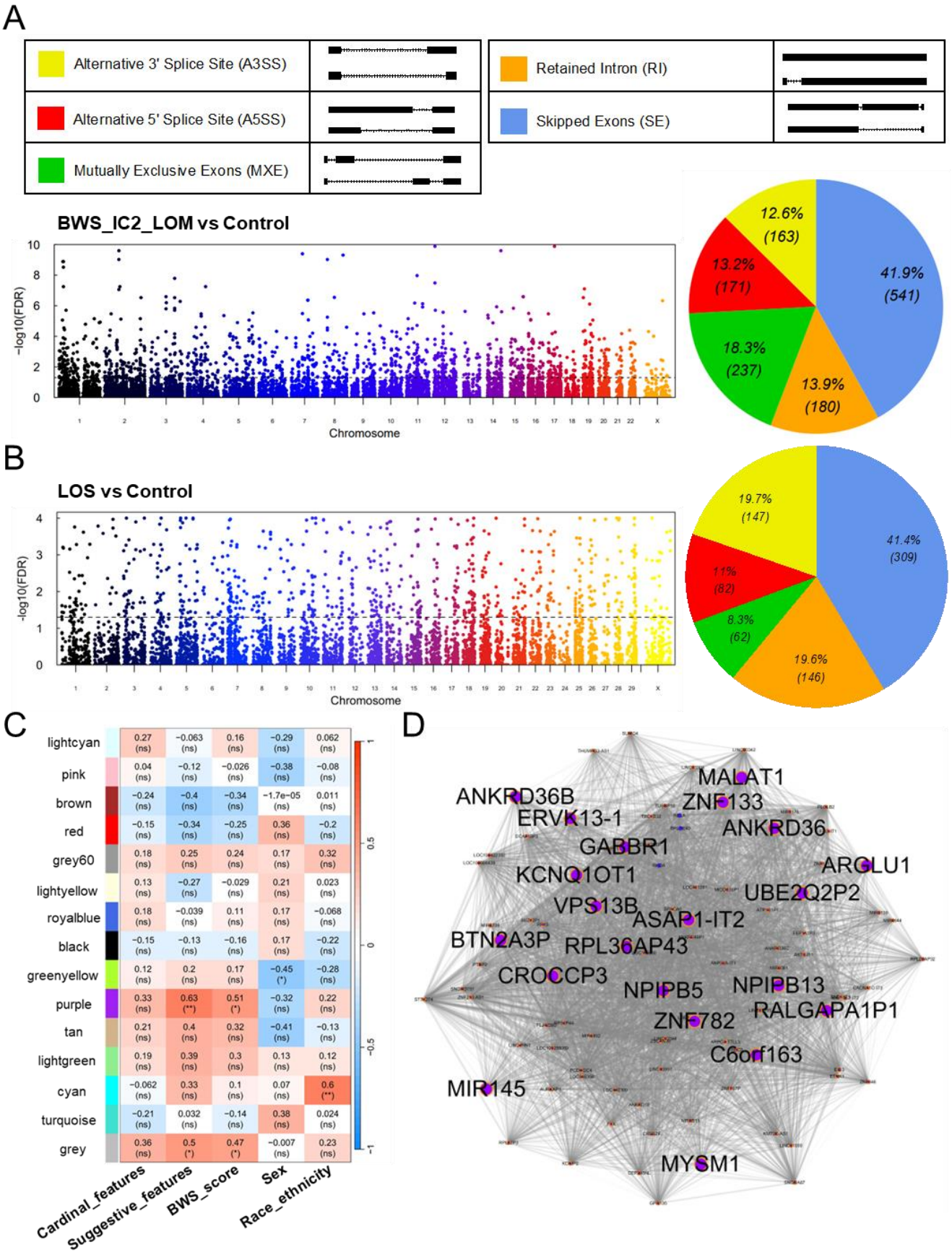

Figure 4

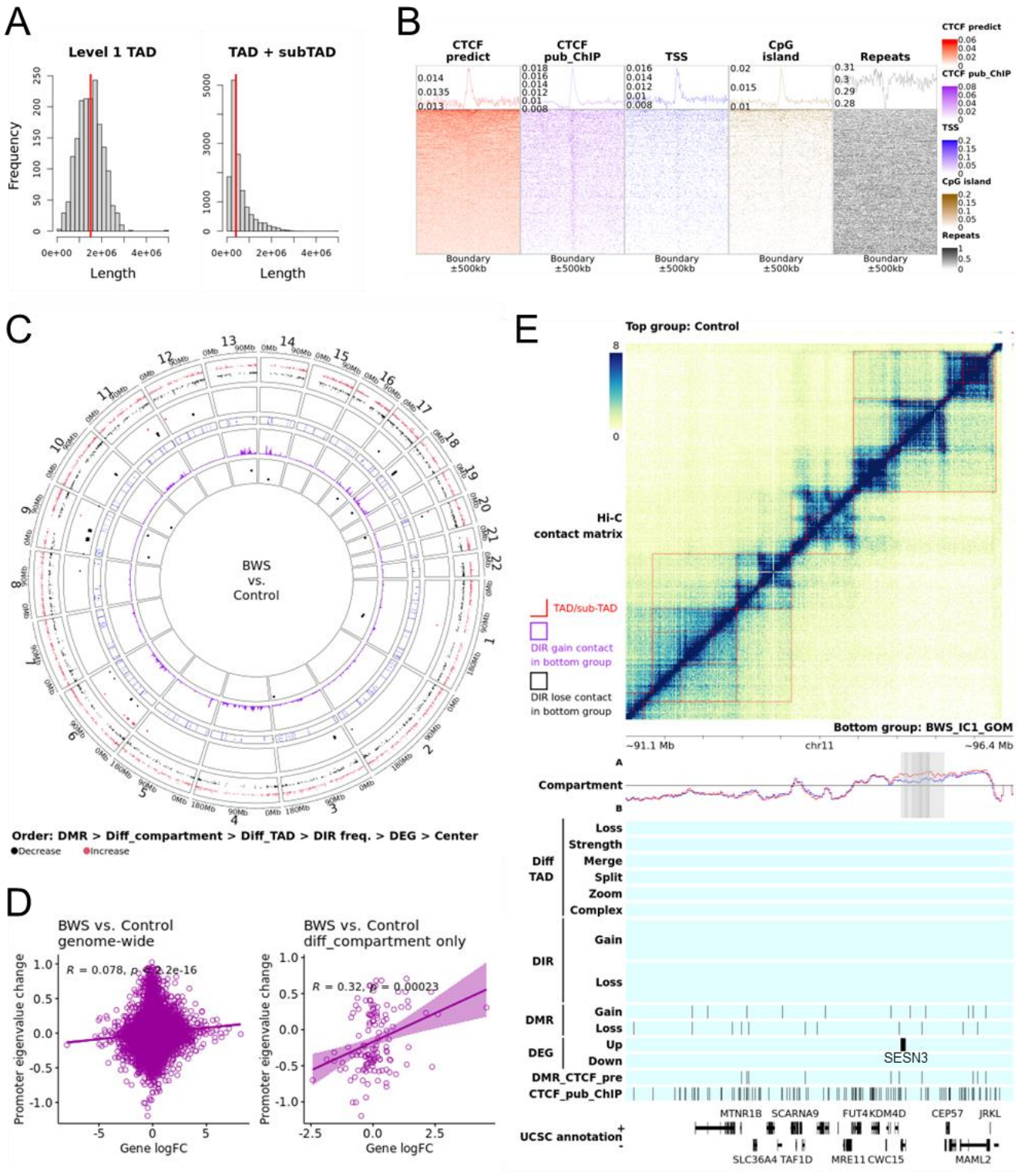

Figure 5

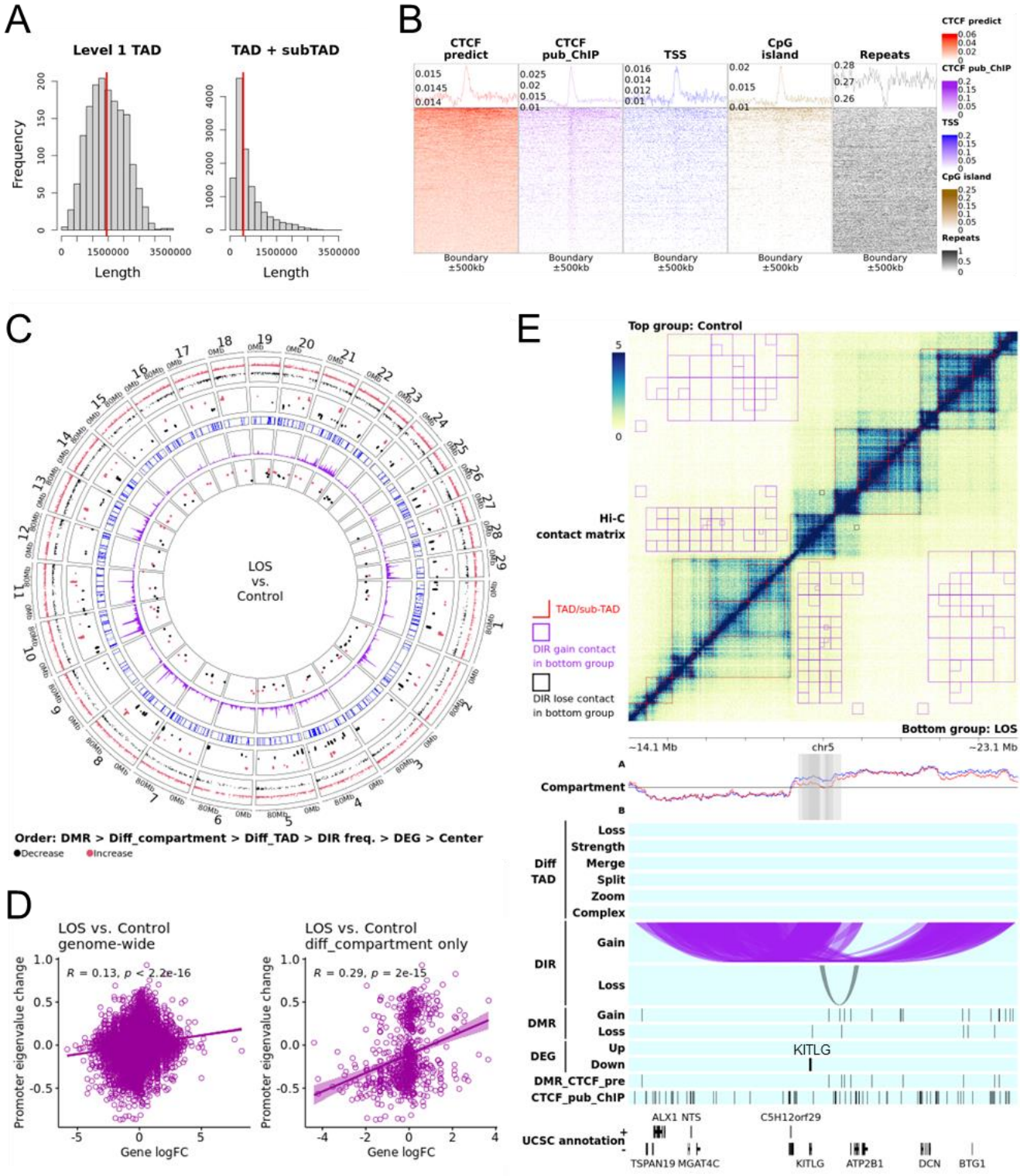

Figure 6

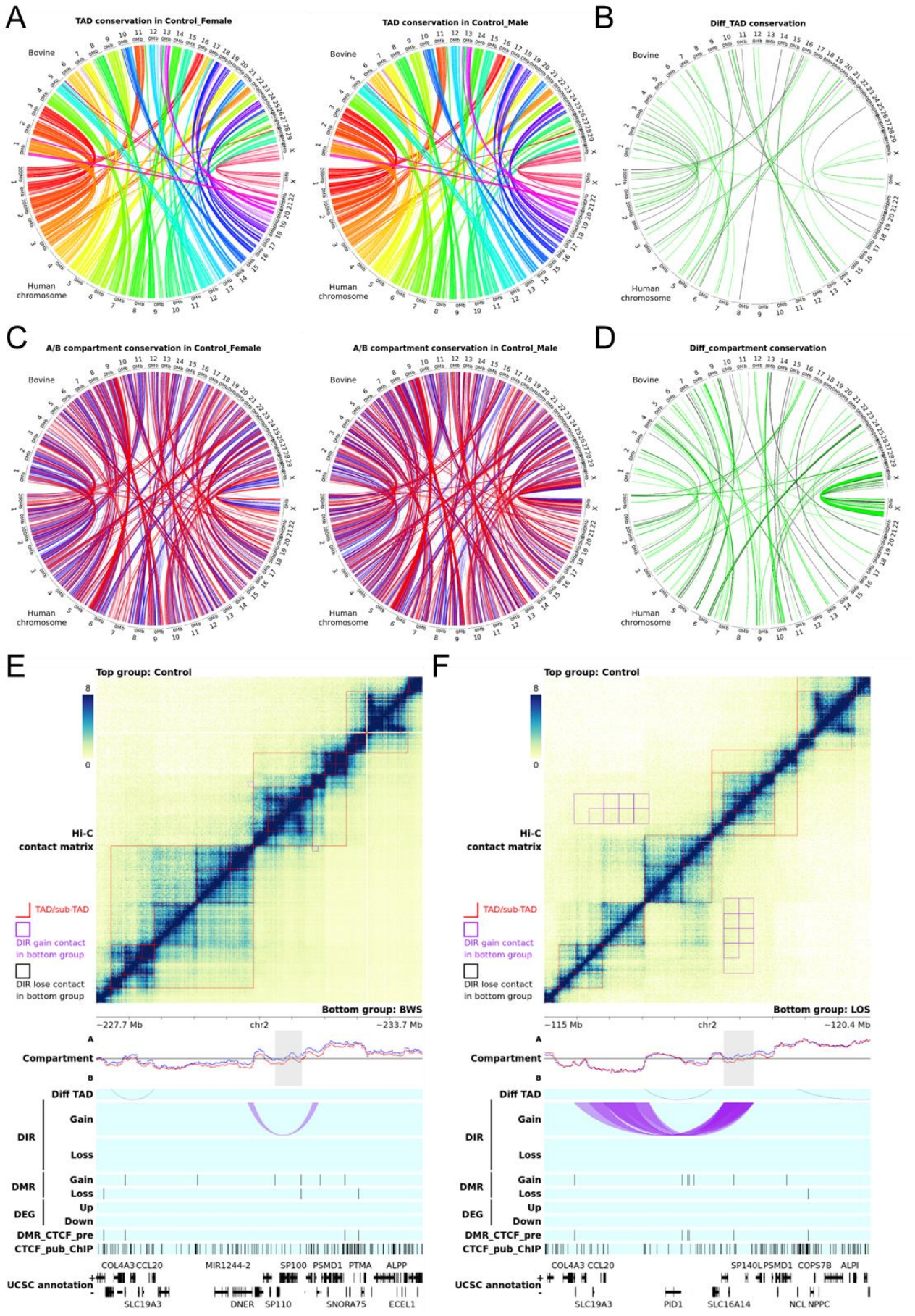

Figure 7