



# Loss of Monoallelic Expression of *IGF2* in the Adult Liver *Via* Alternative Promoter Usage and Chromatin Reorganization

Jinsoo Ahn<sup>1</sup>, Joonbum Lee<sup>1,2</sup>, Dong-Hwan Kim<sup>1</sup>, In-Sul Hwang<sup>3,4</sup>, Mi-Ryung Park<sup>3</sup>, In-Cheol Cho<sup>5</sup>, Seongsoo Hwang<sup>3</sup> and Kichoon Lee<sup>1,2\*</sup>

<sup>1</sup>Functional Genomics Laboratory, Department of Animal Sciences, The Ohio State University, Columbus, OH, United States,

<sup>2</sup>The Ohio State University Interdisciplinary Human Nutrition Program, The Ohio State University, Columbus, OH, United States,

<sup>3</sup>Animal Biotechnology Division, National Institute of Animal Science, Rural Development Administration, Jeonbuk, South Korea,

<sup>4</sup>Columbia Center for Translational Immunology, Columbia University Irving Medical Center, Columbia University, New York, NY, United States,

<sup>5</sup>National Institute of Animal Science, Rural Development Administration, Jeju, South Korea

## OPEN ACCESS

### Edited by:

Lucas Lima Verardo,  
Universidade Federal dos Vales do  
Jequitinhonha e Mucuri, Brazil

### Reviewed by:

Takashi Kohda,  
University of Yamaguchi, Japan  
Jiangnan Qu,  
Veracyte, United States

### \*Correspondence:

Kichoon Lee  
lee.2626@osu.edu

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 14 April 2022

**Accepted:** 21 June 2022

**Published:** 22 July 2022

### Citation:

Ahn J, Lee J, Kim D-H, Hwang I-S,  
Park M-R, Cho I-C, Hwang S and  
Lee K (2022) Loss of Monoallelic  
Expression of *IGF2* in the Adult Liver  
Via Alternative Promoter Usage and  
Chromatin Reorganization.  
Front. Genet. 13:920641.  
doi: 10.3389/fgene.2022.920641

In mammals, genomic imprinting operates *via* gene silencing mechanisms. Although conservation of the imprinting mechanism at the *H19/IGF2* locus has been generally described in pigs, tissue-specific imprinting at the transcript level, monoallelic-to-biallelic conversion, and spatio-temporal chromatin reorganization remain largely uninvestigated. Here, we delineate spatially regulated imprinting of *IGF2* transcripts, age-dependent hepatic mono- to biallelic conversion, and reorganization of topologically associating domains at the porcine *H19/IGF2* locus for better translation to human and animal research. Whole-genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq) of normal and parthenogenetic porcine embryos revealed the paternally hypermethylated *H19* differentially methylated region and paternal expression of *IGF2*. Using a polymorphism-based approach and omics datasets from chromatin immunoprecipitation sequencing (ChIP-seq), whole-genome sequencing (WGS), RNA-seq, and Hi-C, regulation of *IGF2* during development was analyzed. Regulatory elements in the liver were distinguished from those in the muscle where the porcine *IGF2* transcript was monoallelically expressed. The *IGF2* transcript from the liver was biallelically expressed at later developmental stages in both pigs and humans. Chromatin interaction was less frequent in the adult liver compared to the fetal liver and skeletal muscle. The duration of genomic imprinting effects within the *H19/IGF2* locus might be reduced in the liver with biallelic conversion through alternative promoter usage and chromatin remodeling. Our integrative omics analyses of genome, epigenome, and transcriptome provided a comprehensive view of imprinting status at the *H19/IGF2* cluster.

**Keywords:** imprinting, *IGF2*, pigs, alternative promoter usage, biallelic conversion, chromatin reorganization

## 1 INTRODUCTION

Genomic imprinting operates in mammals as an epigenetic mechanism that leads to parent-of-origin-specific monoallelic expression of a subset of genes, mostly in a cluster, *via* silencing of either parental allele (Reik and Walter, 2001; Ferguson-Smith, 2011; Barlow and Bartolomei, 2014). Paternal or maternal expression of imprinted alleles is essential for embryonic development, animal growth and behavior, and diseases related to abnormal loss of imprinting (LOI) (Peters, 2014; Tian, 2014; Tucci et al., 2019). The allele-specific silencing is either direct (e.g., DNA hypermethylation on promoters (Ahn et al., 2021a) or indirect (e.g., by antisense non-coding RNAs (Latos et al., 2012) and chromatin insulators (Bell and Felsenfeld, 2000; Hark et al., 2000) and become complex when multiple types of silencing simultaneously occur on transcript isoforms generated by alternative promoter usage (Hayward et al., 1998a; Hayward et al., 1998b; Peters et al., 1999; Plagge et al., 2004; Ahn et al., 2020c). Regarding the insulators, studies have extensively investigated the *H19/Igf2* locus and established the insulator-mediated organization that coordinately regulates them *via* the imprinting control region (ICR) (Bartolomei et al., 1991; Dechiara et al., 1991; Ferguson-Smith et al., 1991; Bartolomei et al., 1993; Thorvaldsen et al., 1998; Bell and Felsenfeld, 2000; Hark et al., 2000). This ICR upstream (5') of *H19* is methylated only in the paternal allele (i.e., paternally imprinted), so that the insulator CCCTC-binding factor (CTCF) is prevented from binding to the paternal allele of ICR and subsequently, the enhancer downstream (3') of *H19* communicates with the far upstream promoter of *Igf2* to drive paternal *Igf2* expression. Insulin-like growth factor 2 (*IGF2*) is a growth factor that plays a central role in fetal and postnatal growth. Transgenic overexpression of the paternally expressed *Igf2* gene increased fetal growth (Sun et al., 1997), and upregulation or downregulation of *IGF2* *via* aberrant imprinting is associated with the overgrowth disorder Beckwith-Wiedemann syndrome and the growth retardation disorder Silver-Russell syndrome, respectively (Jacob et al., 2013). The maternally expressed *H19* gene is a negative regulator of growth and encodes a tumor suppressor (Hao et al., 1993; Yoshimizu et al., 2008). This counteraction between paternally and maternally expressed genes posited in the parental conflict theory regulates balanced growth (Moore and Haig, 1991; Haig, 2004). Although the insulator model for the *H19/IGF2* locus has been established in mice and humans (Barlow and Bartolomei, 2014; Nordin et al., 2014), tissue- and transcript-specific imprinting and changes in chromatin organization during development remain to be identified in detail. These identifications can be achieved in a comparative manner in mammals that serve as biomedical models and are agriculturally important livestock. Pigs are relevant models for translational research due to their anatomical, physiological, as well as genomic similarities with humans (Lunney et al., 2021). Although studies have described DNA methylation and gene expression regarding *IGF2* and *H19* in pigs and a general conservation in the imprinting mechanism (Li et al., 2008; Park et al., 2009; Braunschweig et al., 2011; Criado-Mesas et al., 2019), tissue-specific imprinting at the

transcript level, age-dependent hepatic monoallelic-to-biallelic conversion (loss of monoallelic expression), and spatio-temporal chromatin reorganization remain largely uninvestigated.

The multi-layered epigenetic regulatory machineries responsible for DNA methylation, chromatin accessibility, histone modifications, and gene expression can be investigated using integrative omics approaches. As the gold standard for DNA methylation analysis, whole-genome bisulfite sequencing (WGBS) of animal models including parthenogenetic embryos has been utilized to identify differentially methylated regions (DMRs) (Clark et al., 2006; Ahn et al., 2020a; Ahn et al., 2020c; Ahn et al., 2021b; a; Morrison et al., 2021). To assess chromatin accessibility (Thurman et al., 2012) and capture open chromatin sites, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) has been used (Buenrostro et al., 2013). Transcriptionally active promoters are marked by histone H3 trimethylated at lysine 4 (H3K4me3) (Howe et al., 2017). A subset of genes has an extended H3K4me3 signal, which covers the gene body and overlaps with the active enhancer mark—histone H3 acetylated at lysine 27 (H3K27ac), consisting of the broad epigenetic domain (Beacon et al., 2021). Cell type and tissue-specific enhancers in non-coding regulatory regions serve as key cis-regulatory elements for gene expression (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Andersson et al., 2014; Coppola et al., 2016). During development, epigenetic modifications alter enhancer activities, as shown in H3K27ac enrichment followed by up-regulation of extracellular matrix genes which might reduce myogenic potential in aged skeletal muscle (Zhou et al., 2019). The formation of open chromatin regions and maintenance of enhancer elements are related to activation of tissue-specific genes (Xu et al., 2007; Wiench et al., 2011). Collectively, epigenetic modulations change chromatin structure and thereby alter DNA accessibility, which affects availability of enhancers and promoters to the transcriptional machinery. These epigenetic modifications might affect tissue-specific monoallelic gene expression within the *H19/IGF2* locus, which can be identified by profiling informative single nucleotide polymorphisms (SNPs) in genomic DNA and mRNA of the same individual (Ahn et al., 2021b; a). In addition, chromosomal conformation capture-based methods such as Hi-C have enabled unbiased identification of chromatin interactions across the genome (Lieberman-Aiden et al., 2009). The genome is partitioned into functional domains of different scale including megabase-long and evolutionarily conserved topologically associating domains (TADs) in which intra-domain chromatin interactions are frequent and cis-regulatory elements are coordinately regulated (Shen et al., 2012). The insulation score for genomic intervals along the chromosome is used to detect minima/valleys of insulation profile for areas of reduced chromatin interactions which are classified as TAD boundaries (Crane et al., 2015). As such, investigating the chromatin structure of imprinted domains in terms of TAD organizations (Lleres et al., 2019; Li et al., 2020) improves our understanding on imprinting clusters in the chromosomal context.

Here we aimed to delineate tissue-specific imprinting of major *IGF2* transcripts and hepatic mono- to biallelic conversion during development of pigs in comparison with humans. We found that the monoallelic-to-biallelic switch through liver-specific alternative promoter usage might occur concomitantly with removal of TAD boundaries and lower chromatin interaction frequencies at the porcine *H19/IGF2* locus. Our findings provided a comprehensive view of coordinated action of regulatory elements and chromatin organization and better understanding of tissue-specific and developmentally regulated genomic imprinting at the *H19/IGF2* locus.

## 2 MATERIALS AND METHODS

### 2.1 Ethics Statement

Our experimental protocols for parthenogenetic studies in the pig were approved by the Institutional Animal Care and Use Committee of the National Institute of Animal Science, Rural Development Administration (RDA) of Korea (approval number NIAS 2015-670). Access to the National Bioscience Database Center (NBDC) Human Database for de-identified data was controlled to observe the Ohio State Human Research Protection Program (HRPP) policies on human subjects (study number 2020E1322).

### 2.2 Collection of Parthenogenetic and Control Embryos

Procedures of production of parthenogenetic embryos following *in vitro* maturation (IVM) of pig oocytes have been described in our previous reports (Kwon et al., 2017; Ahn et al., 2020b). In detail, ovaries of Landrace x Yorkshire x Duroc (LYD) pigs were obtained from a local slaughterhouse, transferred to our lab, and maintained in a thermos at 30–35°C. Cumulus-oocyte complexes (COCs) were gathered and washed in Tyrode's lactate-Hepes medium containing 0.1% (w/v) polyvinyl alcohol. Before IVM, 50 COCs were washed three times in TCM-199 (GIBCO, Grand Island, NY, United States) [supplemented with 0.1% polyvinyl alcohol (w/v), 3.05 mM D-glucose, 0.91 mM sodium pyruvate, 0.57 mM cysteine, 0.5 µg/ml luteinizing hormone, 0.5 µg/ml follicle stimulating hormone, 10 ng/ml epidermal growth factor, 10% porcine follicular fluid (pFF), 75 µg/ml penicillin G, and 50 µg/ml streptomycin] and then placed in each well of five 4-well dishes (Nunc, Roskilde, Denmark) containing 500 µL of maturation medium and matured for 40–42 h at 38.5°C in an incubator containing 5% CO<sub>2</sub>. After maturation, cumulus cells were removed and oocytes having the first polar body were selected and activated as follows: oocytes were placed in a fusion chamber with 250 µm diameter wire electrodes (BLS, Budapest, Hungary) covered with 0.3 M mannitol solution containing 0.1 mM MgSO<sub>4</sub>, 1.0 mM CaCl<sub>2</sub>, and 0.5 mM Hepes and two DC pulses (1 s interval) of 1.2 kV/cm were applied for 30 µs using an LF101 Electro Cell Fusion Generator (Nepa Gene Co., Ltd. Chiba, Japan). Then, after 2 h of stabilization period, parthenogenetic embryos were placed into oviducts of two LY (Landrace X Yorkshire) surrogate gilts aged 12 months at onset of

estrus to develop the embryos. Parthenogenetic embryos were recovered at day 21 from the surrogate gilts before they underwent morphological changes around day 30–35 (Bischoff et al., 2009; Hwang et al., 2020). As a control, fertilized embryos were also recovered at day 21 from gilts, after two LY gilts were naturally mated with boars twice with a 6 h interval during the natural heat period at the onset of estrus and confirmed pregnant by ultrasound examination. For the recovery, gilts were euthanized, and their reproductive tracts were sectioned, and the placenta was isolated from the uterus. Embryos were separated from the surrounding placenta and the surface of embryos was dried on cleaning tissues. Morphologically intact embryos with comparable sizes (approx. 2 cm) were selected and stored in liquid N<sub>2</sub> until further use.

### 2.3 Whole-Genome Bisulfite Sequencing

Genomic DNA was isolated from whole collected embryos (triplicates for both parthenogenetic and control embryos) and fragmented. Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences, Inc. Ann Arbor, MI, United States) was used to optimize bisulfite conversion of genomic DNA according to the manufacturer's instructions. PCR was conducted to amplify the bisulfite-treated DNA with adapter primers, Diastar™ EF-Taq DNA polymerase (Solgent, Daejeon, Korea), and the following thermal cycles: 3 m at 95°C followed by 35 cycles of 30 s at 95°C, 30 s at 60°C, and 30 s at 72°C, and a final extension for 5 m at 72°C. After bead-based clean-up, libraries were sequenced on an HiSeqX sequencer by Macrogen (Seoul, Korea) with 151 bp paired-end reads. Data quality was checked using FastQC (v0.11.7). Raw reads in FASTQ format were quality- and adapter-trimmed with the default parameters of Trim Galore (v0.4.5), except for additional trimming of 18 bp off the 3' end of R1 and the 5' end of R2 for removing bases derived from the sequence tag introduced in the library preparation procedure (--three\_prime\_clip\_R1 18 --clip\_R2 18). Trimmed reads (more than 800 million reads for each sample) were aligned to the pig reference genome (Scrofa11.1, GenBank accession: GCF\_000003025.6) using Bismark (v0.22.3) with default parameters including --no\_overlap for paired-end reads (Krueger and Andrews, 2011). After deduplication using the deduplicate\_bismark command, the Bismark methylation extractor was used to calculate methylation percentage of every cytosine in CpG context. Next, the DMR caller, metilene (v0.2-8), was used to identify *de novo* DMRs with default parameters [including maximum distance of 300 bp between CpGs (-M 300), minimum of 10 CpGs (-m 10), and minimum mean methylation difference of 0.2 (-d 0.2)] and a false discovery rate (FDR) option (Juhling et al., 2016). Methylation ratios and DMRs (FDR < 0.05) were visualized on genomic coordinates using the R/Bioconductor package Gviz (v1.28.3) (Hahne and Ivanek, 2016).

### 2.4 RNA Sequencing

Total RNA from whole embryo samples ( $n = 3$  for each of the control and parthenote) was isolated with TRIzol reagent (Sigma-Aldrich, United States) following the manufacturer's instructions. The RNA samples were treated with DNase I to avoid genomic

DNA contamination and electrophoresed in 1.2% agarose gels to evaluate the integrity of RNA, which was confirmed by 28S/18S rRNA ratio > 2 and RNA integrity number (RIN) > 7 using an Agilent 2100 BioAnalyzer. The concentrations of RNA were assessed by the ratios of A260/A280 and A260/A230 (1.8–2.0). One  $\mu$ g of total RNA and the TruSeq RNA Sample Prep Kit v.2 (Illumina, San Diego, CA, United States) were used to construct cDNA libraries, and final libraries were produced using the protocol consisting of polyA-selected RNA extraction, RNA fragmentation, random hexamer primed reverse transcription and amplification. The cDNA libraries were quantified by quantitative Real-Time PCR (qPCR), and qualification of the libraries was assessed using the Agilent 2100 Bioanalyzer. The library products (100 nt paired-end) were sequenced by the Illumina HiSeq2500 RNA-Seq platform. The raw RNA sequencing reads (FASTQ format) were checked for quality by FastQC (v0.11.7) and trimmed and filtered by Trimmomatic v0.38 with default parameters (Bolger et al., 2014). Then, using STAR aligner (v2.7.5) with default parameter settings (Dobin et al., 2013), cleaned sequencing reads were mapped to the pig reference genome sequence (Sscrofa11.1). Duplicated reads were removed using Picard MarkDuplicates and reads were filtered using SAMtools (-q 30) (Li et al., 2009). Read coverages in BAM files were normalized to values equivalent to transcripts per million (TPM) using bamCoverage in deepTools (v3.5.0) with parameters (--binSize 10, --smoothLength 15) (Ramirez et al., 2014) and plotted using the R/Bioconductor package Gviz (v1.28.3) (Hahne and Ivanek, 2016).

## 2.5 Analyses of Differential Gene Expression

Raw RNA-seq reads in the FASTQ format were quantified against indexed pig transcriptome using Salmon (v1.3.0) in the mapping-based mode (Patro et al., 2017). TPM values of each gene were obtained for parthenogenetic embryos (PA) and control embryos (CN) ( $n = 3$  for each) from Salmon output files (quant.sf). The output files were then imported by tximport function to construct a gene-level DESeqDataSet object for the R/Bioconductor package DESeq2 (v1.28.1) (Love et al., 2014). The test for DEGs was conducted by DESeq2. To obtain significant DEGs, combined criteria of  $FDR < 0.05$  and the absolute  $\log_2$ -fold change > 1 were used, where a fold change is defined as read counts in PA divided by read counts in CN.

## 2.6 Profiling Gene Regulatory Elements

Raw FASTQ files deposited with GEO accession number GSE158430 (ATAC-seq, and H3K27ac, H3K4me3, and CTCF ChIP-seq) for 6-month-old pigs (Kern et al., 2021), GSE143288 (ATAC-seq, and H3K27ac and H3K4me3 ChIP-seq) for 2-week-old pigs (Zhao et al., 2021), GSE153452 (CTCF-seq) for pig embryonic fibroblasts (Li et al., 2020), and GSE155324 (CTCF-seq) for human lymphoblasts (Ushiki et al., 2021) were downloaded *via* the European Nucleotide Archive (ENA) Globus GridFTP. The quality of the raw sequencing reads was checked using FastQC (v0.11.8), and raw reads were trimmed and filtered using Trimmomatic

(v0.38) with default settings (Bolger et al., 2014). All trimmed reads were mapped to the pig reference genome (Sscrofa11.1) or the human reference genome (GRCh38.p13, RefSeq assembly accession: GCF\_000001405.39) using BWA-MEM aligner (v 0.7.17-r1198) using default parameters (Li, 2013). For ATAC-seq, mitochondrial genome was removed from the pig genome before alignment to avoid contamination of the mitochondrial genome which is more accessible owing to chromatin packaging deficiencies (Yan et al., 2020). Aligned reads were deduplicated using Picard MarkDuplicates, and filtered for quality using SAMtools (MAPQ > 30) (Li et al., 2009). MACS2 was used with default parameters to call peaks except for broad peaks (--broad) for ATAC-seq and  $FDR < 0.01$  (-q 0.01) (Zhang et al., 2008). Read coverages in BAM files were normalized to 1x depth (reads per genomic content, RPGC) using bamCoverage in deepTools (v3.5.0) with parameters (--binSize 10, --smoothLength 15) (Ramirez et al., 2014). Peaks were visualized on genomic coordinates using the R/Bioconductor package Gviz v1.36.2 (Hahne and Ivanek, 2016).

## 2.7 Analyses of Tissue-Specific and Developmental Stage-Specific Expression

Raw RNA-seq data of normal pigs [PRJEB44486 (under *Sus scrofa* section in the FAANG datasets, <https://data.faaang.org/dataset>), GSE77776 (Li et al., 2017), PRJNA493166 (Zhang et al., 2019), GSE158430 (Kern et al., 2021), PRJNA597972, GSE124484, GSE92433, PRJNA721126, GSE93855 (Tang et al., 2017), and GSE157045 (Yang et al., 2021)] and humans [GSE63634 (Yan et al., 2016), SRP166862 (George et al., 2019), PRJNA395106, and GSE120795 (Suntsova et al., 2019)] were downloaded through ENA's Globus GridFTP, except for raw data files of the human adult liver (accession hum0158.v2 for controlled access) which were downloaded *via* SFTP of the NBDC Human Database. The RNA-seq processing procedures were the same as above.

## 2.8 Analyses of Allele-Specific Expression

Datasets with genomic DNA and mRNA sequencing data from the same individuals were used. In detail, in addition to the above RNA-seq data, raw data for genomic DNA from the same individual were also downloaded from ENA's Globus GridFTP, except for the human adult liver obtained from NBDC's SFTP. For each pig breed, one liver and skeletal muscle sample from a 60-day-old pig was used (PRJNA309108/GSE77776) (Li et al., 2017). The human fetal liver samples were from two fetuses at 12 weeks after gestation (GSE63634) (Yan et al., 2016). The IDs of human liver samples from adults aged 31–74 were RK001, RK003, RK018, RK019, RK024, RK075, RK130, RK141, and RK157 (hum0158.v2). The IDs of human smooth muscle samples from reproductive age adults were MP100N, MP136N, MP169N, NW206N, and GO537N (SRP163897/SRP166862) (George et al., 2019). The IDs of human lung samples of adults aged 68–77 were N1, N3, N5, N12, N19, and N23 (PRJNA395106). Individuals with these IDs had

heterozygous SNPs at the *IGF2* locus in genomic DNA. In addition, an RNA-seq dataset, PRJNA597972, was used to analyze biallelic tendencies.

Whole-genome sequencing (WGS), whole-exome sequencing (WES), and H3K4me1 ChIP-seq data were cleaned and aligned using BWA-MEM as above. The deduplicated BAM files were used to detect SNPs and obtain allele counts in individual samples by generating vcf files through bcftools mpileup piped to bcftools call command. The published pig SNP data (GCA\_000003025.6\_current\_ids.vcf.gz) were obtained from the EBI ftp server (ftp://ftp.ebi.ac.uk/pub/databases/eva/rs\_releases/release\_2/by\_species/sus\_scrofa/Sscrofa11.1/). The human SNPs in the vcf file (GCF\_000001405.39.gz) for the GRCh38.p13 genome were downloaded from the NCBI data repository (https://ftp.ncbi.nih.gov/snp/.redesign/latest\_release/VCF/). For sequencing to generate chromatogram for *IGF2(8)* and *IGF2-AS*, genomic DNA and RNA were isolated from the liver of 6-month-old Berkshire pigs. To amplify a DNA fragment containing a potential SNP on the second exon of the *IGF2(8)* transcript, primers were designed on the first intron (forward: 5'-AGCGTGGA GAGGCTCTCTT-3') and the second intron (reverse: 5'-ACCCAAACTCAATGCAGCTTT-3'). For a potential SNP on the second exon of the *IGF2-AS* transcript, primers were designed on the first intron (forward: 5'-CTGCTC TGGGTTCCCAT-3') and the second intron (reverse: 5'-CTGACAA CCCTGCCCTGTT-3'). After sequencing and confirming heterozygosity of the SNP, primers for cDNA were designed on the first exon (forward: 5'-CCCCATTGGCACCAG TACAG-3') and the third exon (reverse: 5'-GCTGAGCCCGAG GAGATGTG-3') of *IGF2(8)* and the first exon (forward: 5'-GGA CACGCGAGGCGA-3') and second exon (reverse: 5'-CAAGGT CCAGGCGCATGT-3') of *IGF2-AS* to avoid genomic DNA contamination. PCR was conducted using the Taq DNA Polymerase (#M0273S, New England Biolabs, Ipswich, MA, United States) with an initial incubation at 95°C for 30 s, followed by 43 cycles at 95°C for 30 s, 56°C for 35 s, and 68°C for 20 s. The final extension was performed at 68°C for 5 min. After agarose gel electrophoresis, DNA was extracted from separated PCR bands using a QIAquick Gel Extraction Kit (#28104, Qiagen, Venlo, Netherlands) according to the manufacturer's protocol and sent out for Sanger sequencing at The Ohio State University Core Facility.

## 2.9 Hi-C Data Processing

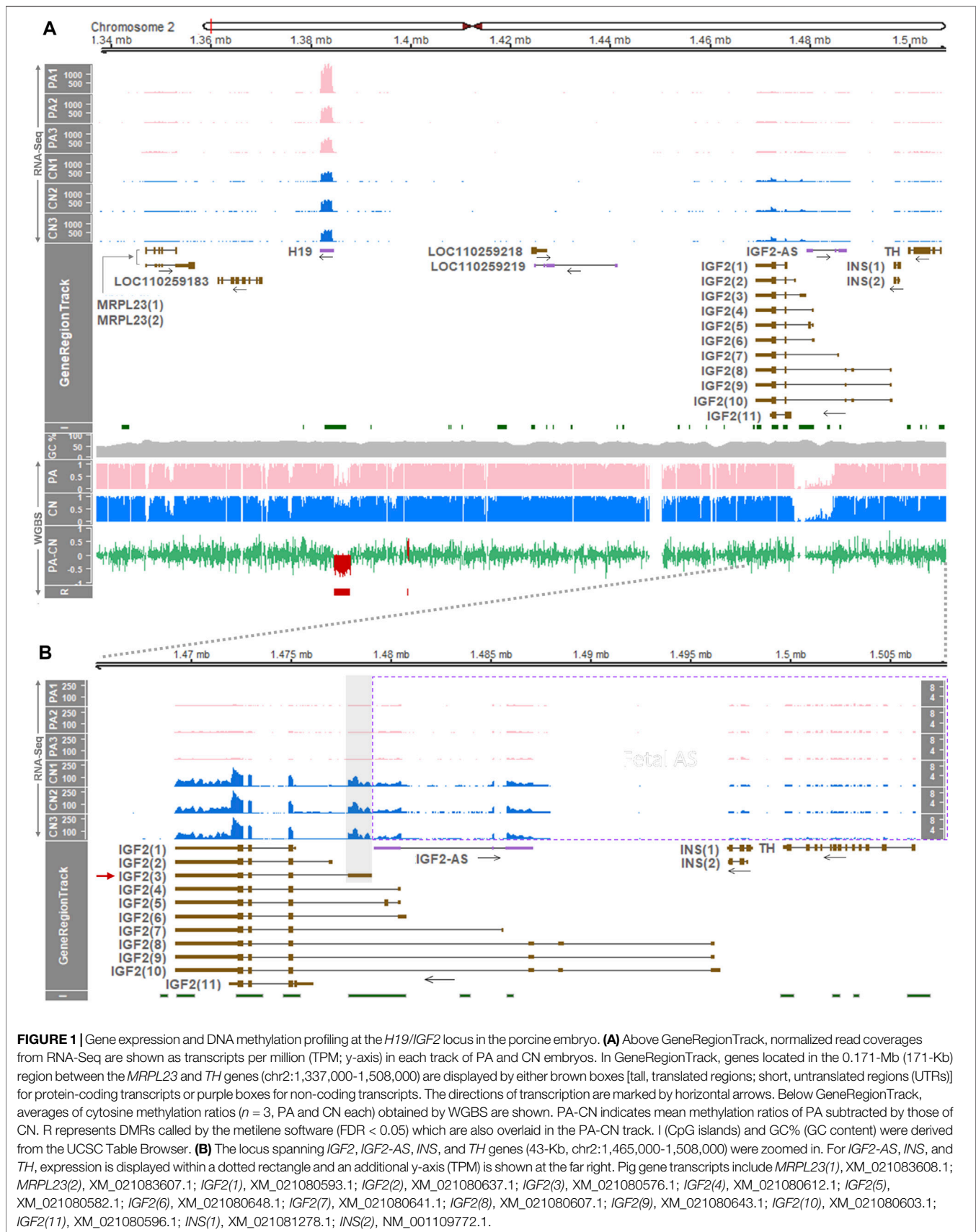
Raw Hi-C data in FASTQ format of the liver of fetal (embryonic day 90) and adult (2-years-old) Bamaxiang pigs (PRJNA482496) (Tian et al., 2020) and skeletal muscle tissues of a Luchuan pig (embryonic day 35) (GSE166346) (Yuan et al., 2021) and Large White pigs (2-week-old) (GSE143288) (Zhao et al., 2021) were retrieved through ENA's Globus GridFTP. After assessing the quality of data using FastQC (v0.11.7), the raw paired-end reads were trimmed and filtered out to remove low quality reads, adapters, and reads shorter than 20 bp by using default settings of Trim Galore (v0.4.5). Cleaned data were processed using HiC-Pro (v3.1.0) with default parameters (Servant et al., 2015) while specifying the index for bowtie2 (v2.4.4) and MboI (or DpnII) restriction fragments according to data submitters'

publications (Tian et al., 2020; Yuan et al., 2021; Zhao et al., 2021). To determine concordance of raw Hi-C matrices, GenomeDISCO was used to produce smoothed matrices and randomly work on the smoothed matrices to obtain concordance scores (Ursu et al., 2018). The validPairs files from matrices with high concordance were merged to increase resolution and normalized by iterative correction and eigenvector decomposition (ICE) using parameters of HiC-Pro (-s merge\_persample -s build\_contact\_maps -s ice\_norm). TADs were identified using insulation scores and ICE normalized matrices were visualized using the GENOVA R-package (Van Der Weide et al., 2021).

## 3 RESULTS

### 3.1 A Differentially Methylated Region Within the Porcine *H19/IGF2* Locus is Paternally Methylated at a CpG Island

Using diploid uni-maternal PA embryos and bi-parental CN embryos, we performed WGBS at an approximately 50X depth (**Supplementary Table S1**). By analyzing the WGBS data, DMRs between the embryos were obtained whose mean methylation difference (i.e., a mean of PA subtracted by a mean of CN) was more than 0.2 (hypermethylation in PA) or less than 0.2 (hypomethylation in PA) with significance (FDR < 0.05) (**Supplementary Figure S1A**). Compared with methylated regions without significance, DMRs tended to be longer in base pairs and greater in CpG numbers, as a result of processing approximately 588 million deduplicated uniquely mapped reads for each replicate on average (**Supplementary Figure S1B** and **Supplementary Table S1**). Using unmethylated lambda phage DNA added to sample DNA prior to fragmentation, bisulfite conversion rates at CpG, CHG, and CHH sites were calculated and estimated to be 99.69%–99.71% across samples representing successful library construction (**Supplementary Table S1**). In order to examine DNA methylation status within the *H19/IGF2* locus in porcine embryos, methylation ratios at single-base resolution were analyzed. Based on the mean methylation ratio at each CpG site, diploid uni-maternal PA carried a broad range of hypomethylation immediately upstream (5') of the non-coding *H19* gene (**Figure 1A** bottom panel, **Supplementary Table S2**). Hemi-methylation occurred in bi-parental CN, and a DMR between PA and CN was identified which mostly overlapped a CpG island (**Figure 1A** bottom panel). It was consistent with a previous study regarding the presence of the porcine *H19*-DMR, where maternal alleles were almost unmethylated and paternal alleles were completely methylated which led to hemi-methylation (Braunschweig et al., 2011). Additionally, a narrow hypermethylated DMR in PA was located farther upstream of *H19* (**Figure 1A** bottom panel). DNA methylation in the upstream and downstream of the *IGF2*, *INS*, *TH* genes did not show differences between PA and CN, and various *IGF2* transcripts did not exhibit a transcript-specific methylation pattern (**Figure 1A** bottom panel). In summary, the hypomethylated region in PA (i.e., paternally methylated *H19*-



DMR) was the only DMR that overlapped a CpG island throughout the *H19/IGF2* locus of the pig embryo, and transcript-specific DMRs were not found in the *IGF2* locus.

### 3.2 Expression of the *IGF2* Gene in Pig Embryos is Paternal Allele-Specific

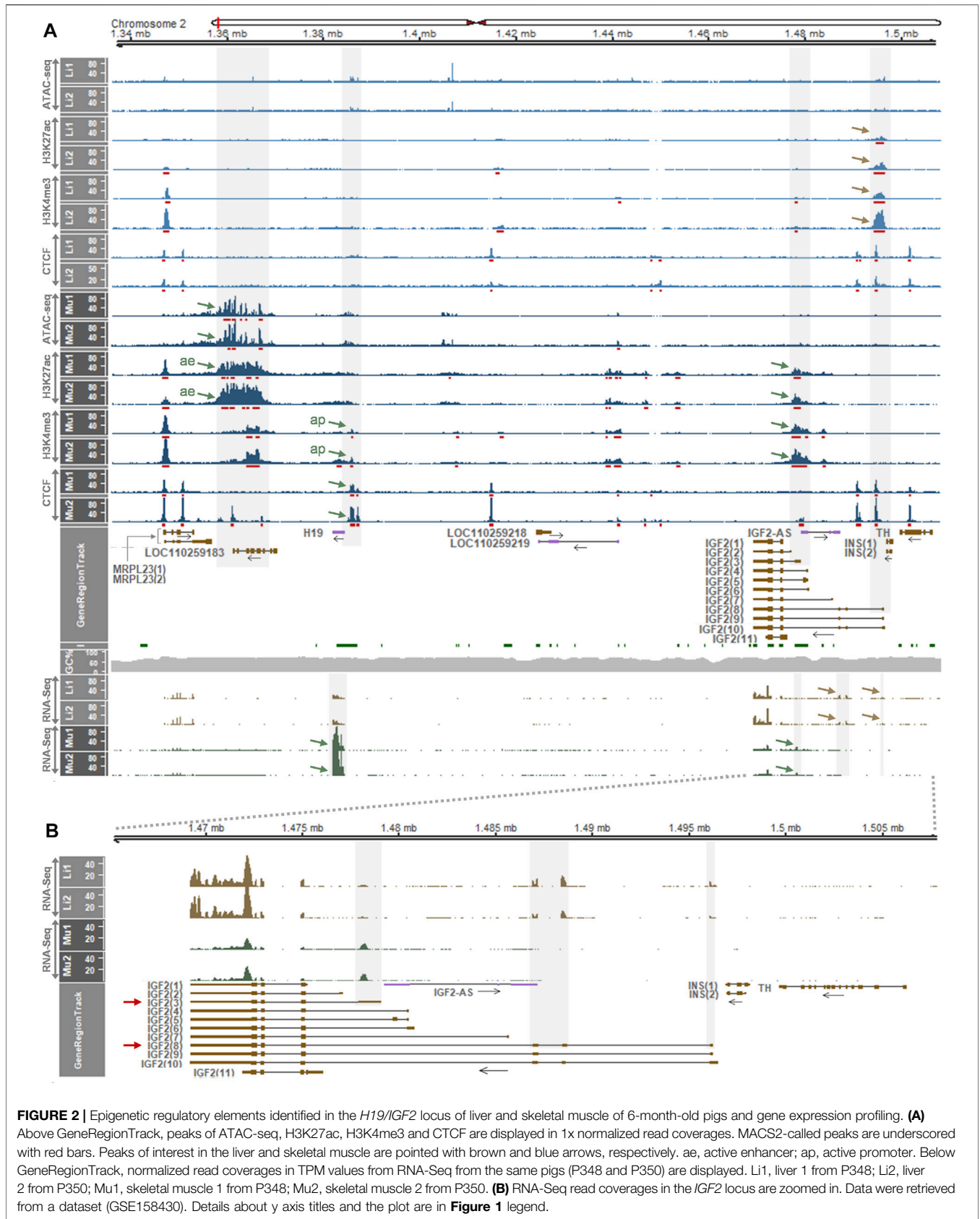
In our model of PA and CN, gene and transcript expressions within the *H19/IGF2* locus were examined to investigate their imprinting status. Analyses of differentially expressed genes (DEGs) on RNA-seq data (**Supplementary Table S1**) revealed that expression of the non-coding *H19* gene tended to increase in PA compared to CN (1.61-fold higher in PA) suggesting its maternal expression, i.e., had a higher expression in two maternal alleles of PA than in one maternal allele of CN (**Figure 1A** top panel, **Supplementary Table S3**, and **Supplementary Figure S1C**). This deviation from a 2-fold increase might be accounted for by gene dosage compensation in diploid uni-parental embryos or loss of imprinting (Shemer et al., 1996; Park et al., 2011). Expression of the *IGF2* genes was almost exclusive in CN embryos (adjusted  $p$ -value < 0.001), indicating expression in the paternal allele of CN while being absent in PA without the paternal allele (**Figure 1B**; **Supplementary Table S3**, and **Supplementary Figure S1C**). Among the *IGF2* transcript isoforms, the major transcript was *IGF2(3)* (short-form; GenBank accession number: XM\_021080576.1) having four exons and its non-overlapping first exon carried predominant read coverages (**Figure 1B**). In addition, paternal expression of antisense of *IGF2* (*IGF2-AS*) was indicated by almost exclusive read coverages in CN (adjusted  $p$ -value < 0.001) (**Figure 1B**; **Supplementary Table S3**, and **Supplementary Figure S1C**). Other genes and gene transcripts including *MRPL23* and *INS* appeared to be expressed biallelically and expression of the *TH* gene tended to increase in PA embryos, while expression of LOC110259183, LOC110259218, and LOC110259219 was almost undetectable (**Figure 1**; **Supplementary Table S3**, and **Supplementary Figure S1C**). Consequently, the imprinted expressions of the *IGF2(3)* and *IGF2-AS* genes were shown to be paternal monoallelic.

### 3.3 Regions Surrounding the Porcine *H19* and *IGF2* Locus Accumulate Distinguishable Gene Regulatory Elements Between the Liver and Skeletal Muscle

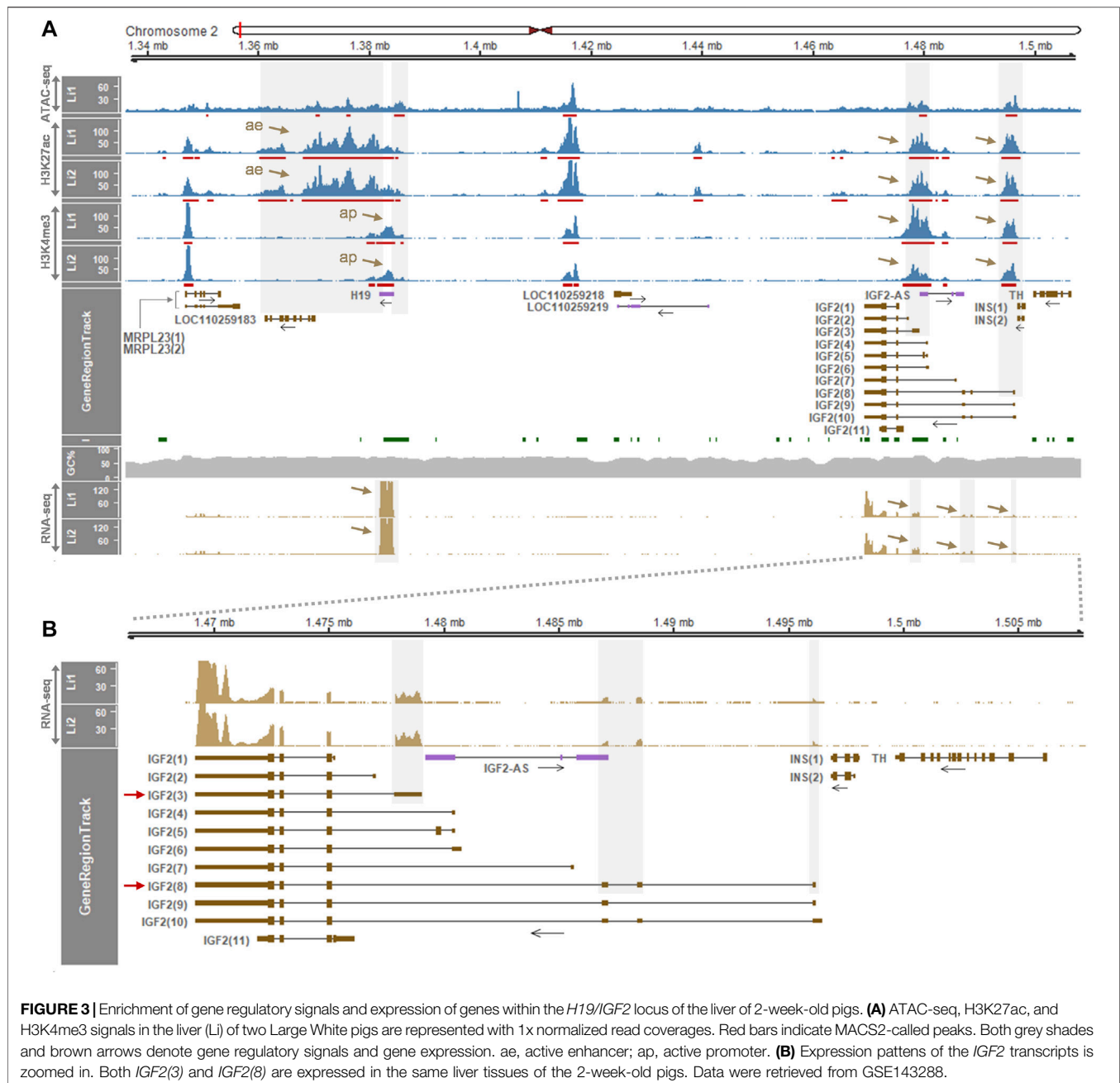
To examine whether imprinted monoallelic expression at the *IGF2* locus is maintained in different developmental stages, we first analyzed expression levels *H19* and *IGF2* in multiple tissues of 6-month-old pigs using a dataset retrieved by the Gene Expression Omnibus (GEO) accession GSE158430. At this stage, expression of *IGF2* was prominent in the liver and expression of *H19* was high in skeletal muscle (**Supplementary Figure S2**). In other pig tissues, including the adipose tissue, brain hypothalamus, lung, and spleen, expression of both *IGF2* and *H19* was substantially low (**Supplementary Figure S2**). Therefore, we investigated gene regulatory elements

in the liver and skeletal muscle of the same pigs within and near the porcine *H19/IGF2* locus that might affect gene expression. ATAC-seq (for open chromatin), H3K27ac (at active enhancers and promoters), H3K4me3 (at active promoters), and CTCF (insulators) were analyzed. Our analysis of these datasets from two biological replicates of skeletal muscle revealed that H3K27ac signals were distributed around ATAC peaks spanning ~20 kb (chr2:1,346,472-1,366,301) approximately 15–35 kb downstream (3') of the *H19* transcription start site (TSS) (ae, **Figure 2A**), suggestive of active enhancers. However, these signals were absent in the liver. On the other hand, H3K27ac peaks upstream (5') of *IGF2(8)* were enriched with H3K4me3 in the liver, and also H3K27ac peaks upstream of *IGF2(3)* were enriched with H3K4me3 in skeletal muscle (approx. 1–2 kb H3K27ac and H3Kme3 signals) (**Figure 2A**), suggesting that they represent regulatory features including active promoters. In addition, in skeletal muscle but not in the liver, H3K4me3 signals located immediate upstream (5') of *H19* where CTCF peaks co-localized (ap, **Figure 2A**). It indicated that the *H19* promoter was activated, and this activation might be attributed to the aforementioned active enhancer downstream of *H19* and monoallelic CTCF binding which leads to monoallelic activation of the *H19* promoter as previously reported regarding the human and mice ICRs (Bell and Felsenfeld, 2000; Hark et al., 2000). Moreover, although monoallelic CTCF binding in the skeletal muscle was not confirmed due to lack of heterozygous sites, the monoallelic CTCF binding was evident in pig embryonic fibroblasts (PEFs) whose CTCF sites were analogous to those of human lymphoblasts (**Supplementary Figure S3**). Noticeably, expression of *H19* was low in the liver and high in skeletal muscle (**Figure 2A** bottom panel, and **Supplementary Figure S2**), and *IGF2* transcripts expressed in this locus were different between the two tissues: *IGF2(8)* in the liver and *IGF2(3)* in skeletal muscle, and total expression of *IGF2* was higher in the liver (**Figure 2A** bottom panel, **Figure 2B**, and **Supplementary Figure S2**).

We further investigated an additional presence of gene regulatory elements in the liver which deviate from the insulator model (Barlow and Bartolomei, 2014; Nordin et al., 2014) using another dataset (GSE143288). In 2-week-old pigs, H3K27ac peaks were distributed in the downstream of *H19* around the open chromatin region indicated by ATAC signals, although the dataset contained ATAC data for the one pig (Li1) (**Figure 3A** top panel). The H3K4me3 signals near *H19* indicated an activated promoter, and all the ATAC, H3K27ac, and H3K4me3 signals around the first exon of *IGF2(3)* suggested the long-range insulator-mediated regulation of *IGF2(3)* expression. However, those ATAC, H3K27ac, and H3K4me3 signals were also detected near the first exon of the *IGF2(8)* transcript (**Figure 3A** top panel), suggesting the presence of an additional transcript-level gene regulation. Consistently, the expression of both *IGF2(3)* and *IGF2(8)* was observed in the same liver tissues while *H19* was expressed in the far downstream of *IGF2* (**Figure 3A** bottom panel and **Figure 3B**). Taken together, it suggested that, compared to the *IGF2(3)* expression in the whole embryo (**Figure 1**) and the *IGF2(8)* expression in the liver from 6-month-old pigs (**Figure 2**), both *IGF2(3)* and *IGF2(8)* are





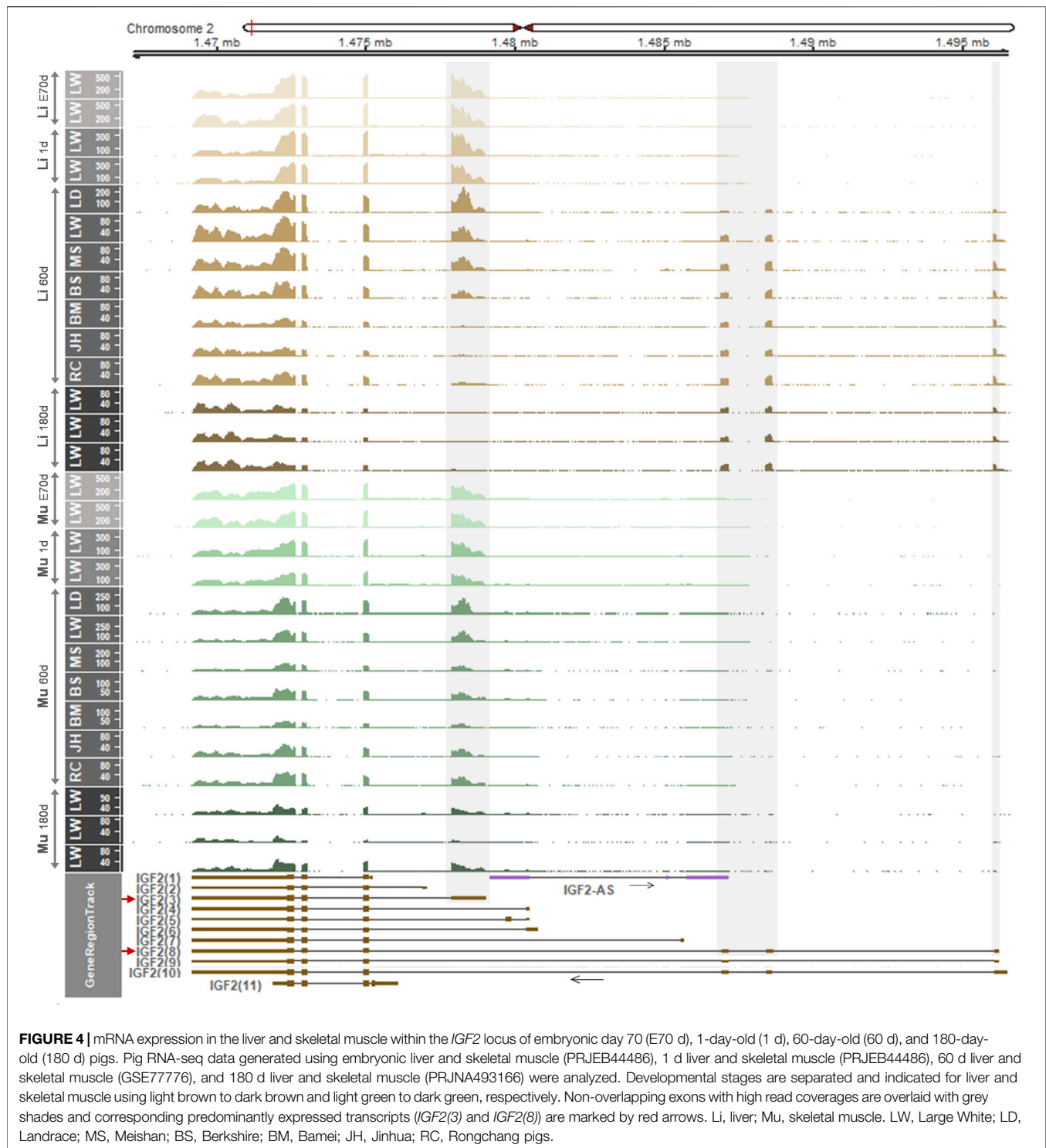


expressed in the liver of two-week-old pigs (Figure 3) under the regulation of two different sets of gene regulatory elements.

### 3.4 Developmental Changes of the Porcine *IGF2* Gene Expression and Bi- or Mono-Allelic Expression are Distinct in the Liver and Skeletal Muscle

Considering the aforementioned distinct gene regulation and expression patterns in the liver, we examined whether the *IGF2* gene expression is regulated developmentally and shows an allele-specific pattern. First, using RNA-seq datasets, read

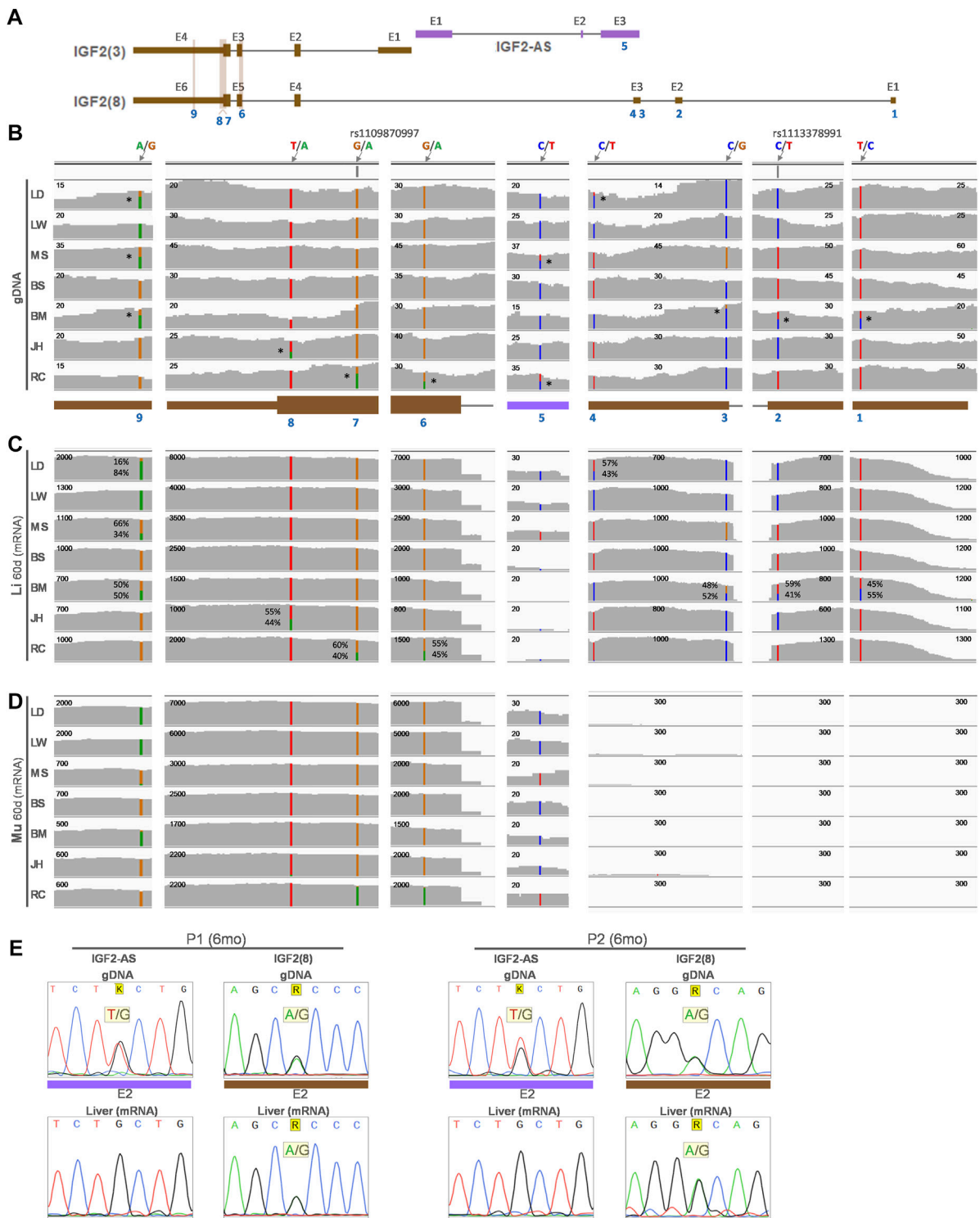
coverages for *IGF2* transcripts were analyzed in the liver and skeletal muscle across developmental stages. In both the liver and skeletal muscle, on embryonic day 70 and postnatal day 1, the *IGF2(3)* transcript was predominantly expressed (Figure 4). On day 60, however, the *IGF2(3)* transcript was not always predominant in the liver of analyzed pigs, while in the skeletal muscle the *IGF2(3)* transcript was predominant (Figure 4). In particular, in the liver of Landrace pigs, the *IGF2(3)* transcript was predominant and the *IGF2(8)* expression was detected at a low level. In the liver of Large White and Meishan pigs both *IGF2(3)* and *IGF2(8)* transcripts were predominant, while Berkshire pigs showed high expression of the *IGF2(8)* transcript and reduced



expression of the *IGF2(3)* transcript. On the other hand, expression of the *IGF2(3)* transcript was almost absent in the liver of Bamei, Jinhua, and Rongchang pigs, whereas the *IGF2(8)* transcript was predominant (Figure 4). On day 180, the *IGF2(8)* transcript was predominant in the liver and the *IGF2(3)* transcript was predominant in skeletal muscle (Figure 4). The

total read coverages of *IGF2* tended to decrease in both the liver and skeletal muscle during development. The transition of *IGF2* expression in the liver, but not in skeletal muscle, was repeatedly shown in additional data (Supplementary Figures S4A and S5).

To examine allelic expression patterns, informative (heterozygous) SNPs in the *IGF2(3)* and *IGF2(8)* transcripts



**FIGURE 5 |** Allelic expression of the *IGF2* transcripts from the liver and skeletal muscle of 60-day-old pigs from 9 breeds. **(A)** Predominant porcine *IGF2* transcripts in liver (Li) and skeletal muscle (Mu) from **Figure 4** are displayed along with an antisense transcript (*IGF2-AS*). **(B)** Location of SNPs on genomic DNA (gDNA) analyzed using WGS (PRJNA309108) are denoted with blue numbers. Reference and alternative alleles are marked in the format of ref/alt (e.g., A/G). If present, the reference SNP ID (rs ID) is shown (e.g., rs1109870997). Heterozygous alleles are denoted with stars (\*) on the right side of SNPs in the genomic DNA. Read coverages of RNA-seq (GSE77776) from the same pigs as WGS are displayed for liver **(C)** and skeletal muscle **(D)**. Numbers on the top-left corner of each read coverage denote the depth of coverage. **(E)** Sequencing of exon 2 (E2) of *IGF2(8)* and *IGF2-AS* for genomic DNA and cDNA from the liver tissues of two 6-month-old (6 months) Berkshire pigs (P1 and 2). LW, Large White; LD, Landrace; MS, Meishan; BS, Berkshire; BM, Bamei; JH, Jinhua; RC, Rongchang pigs.

and *IGF2-AS* were explored (Figures 5A,B). Nine heterozygous SNPs were found in genomic DNA (gDNA) of any of the 60-day-old pigs (Figures 5A,B and Supplementary Table S4). Among those SNPs, four heterozygous SNPs including a previously reported SNP (rs1113378991) were found for non-overlapping exons of the *IGF2(8)* transcript in Bamei (SNP1-3) and Landrace (SNP4) pigs (Figure 5B and Supplementary Table S4) and mRNA expression on those four alleles in the liver was biallelic (Figure 5C and Supplementary Table S4), indicating biallelic expression of the porcine *IGF2(8)* transcript. One heterozygous SNP for *IGF2-AS* (SNP5) in Meishan and Rongchang pigs (Figure 5B and Supplementary Table S4) was expressed monoallelically in both the liver and skeletal muscle, although the mRNA expression level was low (Figures 5C,D and Supplementary Table S4). Four other SNPs (SNP6-9) including a previously reported SNP (rs1109870997) were found in overlapping exons of the *IGF2(3)* and *IGF2(8)* transcripts, and informative SNPs were not found in the non-overlapping exon (E1) of *IGF2(3)* (Figures 5A,B and Supplementary Table S4). In the liver of Rongchang, Jinhua, and Bamei pigs which expressed exclusively the *IGF2(8)* transcript (Figure 4), SNP6 and 7 in Rongchang, SNP8 in Jinhua, and SNP9 in Bamei pigs were biallelically expressed (Figure 5C and Supplementary Table S4). In the liver of Meishan pigs which expressed both the *IGF2(3)* and *IGF2(8)* transcripts (Figure 4), SNP9 showed a decreased biallelic tendency due possibly to monoallelic expression of *IGF2(3)* and biallelic expression of *IGF2(8)* (Figure 5C and Supplementary Table S4). In the liver of Landrace pigs which expressed the *IGF2(3)* transcript predominantly and the *IGF2(8)* transcript at a low level (Figure 4), SNP9 showed a monoallelic tendency due possibly to substantially higher monoallelic expression of the *IGF2(3)* transcript than biallelic expression of the *IGF2(8)* transcript (Figure 5C and Supplementary Table S4). These allelic changes were also found using another dataset which showed a biallelic tendency or a decreased biallelic tendency (Supplementary Figure S4B). In the skeletal muscle, these Rongchang, Jinhua, Bamei, Meishan, and Landrace pigs expressed the *IGF2(3)* transcript predominantly (Figure 4), and corresponding expression of SNP6-9 tended to be monoallelic (Figure 5D and Supplementary Table S4). In the liver of adult Berkshire pigs, biallelic expression of the *IGF2(8)* transcript and monoallelic expression of *IGF2-AS* were confirmed (Figure 5E). Taken together, it suggests that allelic expression patterns were different in between the liver and muscle, in addition to the difference of the expressed form of *IGF2* transcripts.

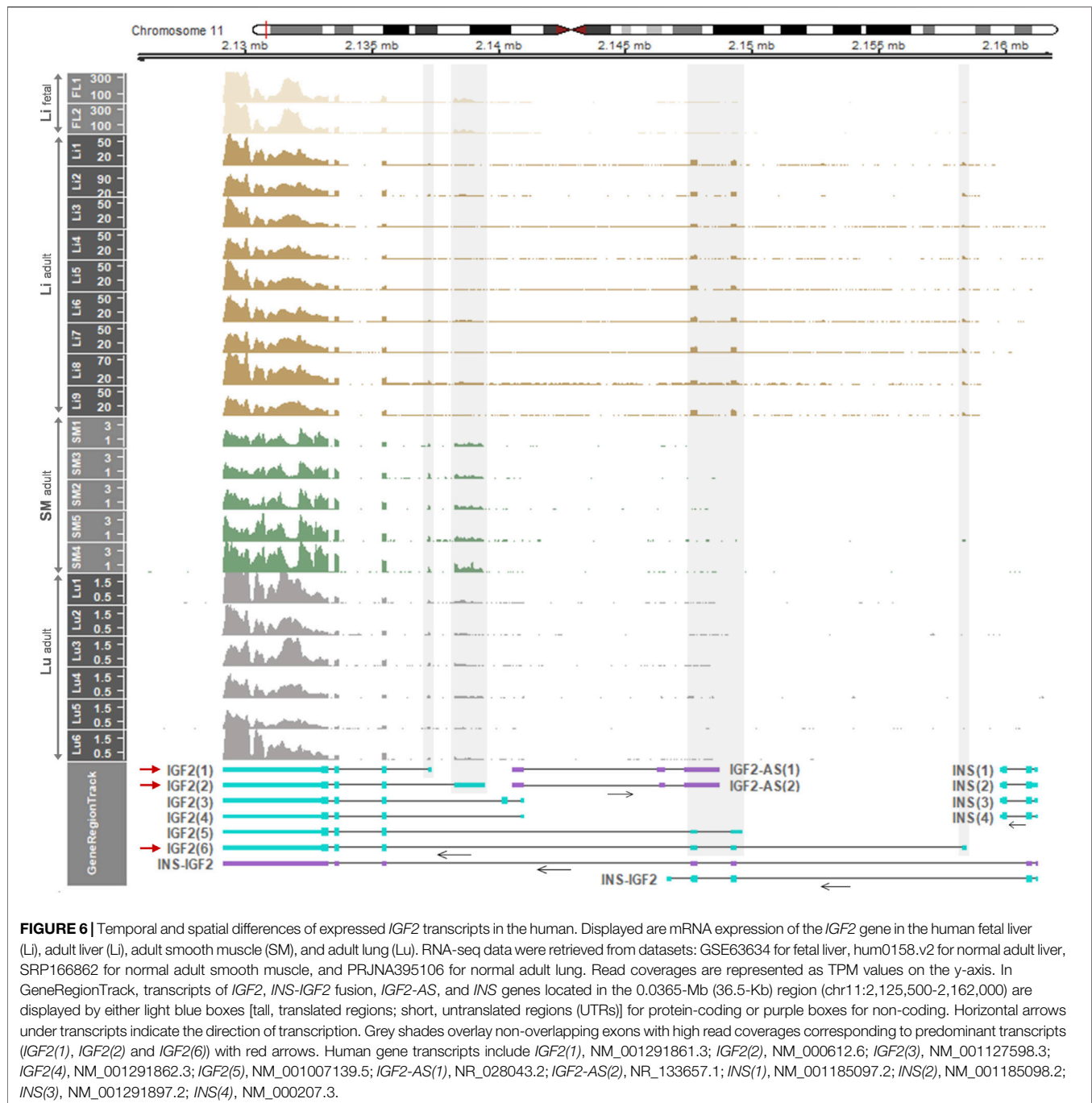
### 3.5 *IGF2* Expression in Humans is Developmentally Regulated and Bi- or Mono-Allelic Expression Patterns Are Tissue-Specific

For comparison with pigs, the *IGF2* gene expression in the fetal and adult liver and other tissues of the human was examined and also allelic expression in those tissues was further investigated. Based on RNA-seq read coverages, when compared with other

tissues, expression of *IGF2* and *H19* was relatively high in the liver and muscle, respectively, while expression of both *IGF2* and *H19* was low in other tissues including the brain, lung, colon, and stomach (Supplementary Figure S6A). The *IGF2(2)* transcript [the orthologous transcript of porcine *IGF2(3)*] was expressed in the fetal liver and the *IGF2(6)* transcript [the orthologous transcript of porcine *IGF2(8)*] was expressed in the adult liver (Figure 6 and Supplementary Figures S6, S7), which are comparable to the findings of the predominant expression of *IGF2(3)* and *IGF2(8)* in fetal and adult livers, respectively, in pigs. Both the *IGF2(1)* and *IGF2(2)* transcripts were expressed in the adult smooth muscle and lung (Figure 6). The overall expression levels tended to be high to low in the order of fetal liver, adult liver, adult smooth muscle, and adult lung. Expression of *INS-IGF2* fusion transcripts and *INS* transcripts was not detectable (Figure 6), while expression of *IGF2-AS* was low but detectable in the fetal liver and decreased in postnatal stages of both pigs and humans (Figure 7B and Supplementary Figure S8). Total expression of *IGF2* was significantly higher in the liver than in skeletal muscle, and total expression of *H19* tended to be higher in skeletal muscle than in the liver (Supplementary Figure S6B), and these patterns in adult humans were similar to those of the adult pigs (Supplementary Figure S2B).

Allelic expression was further explored for the fetal and adult liver of the human, regarding informative SNPs in the expressed *IGF2(2)* and *IGF2(6)* transcripts and *IGF2-AS* (Figure 7A). The *IGF2(2)* and *IGF2-AS* transcripts were expressed in the fetal liver (Figures 7A,B), and six informative SNPs on gDNA were found (Figure 7B and Supplementary Table S5). All of those six SNPs were monoallelically expressed in mRNA of the fetal liver (Figure 7B and Supplementary Table S5). The *IGF2(6)* transcript was expressed in the adult liver and with a very low degree for *IGF2-AS* expression (Figures 7A,C). The same SNP5 between the fetal and adult liver was heterozygous in the adult liver from individual 2, 4, 6, 7, and 9 (Figure 7C top panel and Supplementary Table S5) and tended to be expressed biallelically (Figure 7C bottom panel and Supplementary Table S5), suggesting biallelic expression of *IGF2(6)* unlikely to monoallelic expression of *IGF2(2)* in the fetal liver (Figure 7B and Supplementary Table S5). Also, in the adult liver, SNP8-10 were found to be heterozygous in some individuals (SNP8: 2 and 4-9, SNP9: 2, 6, 7, and 9, SNP10: 2, 4, 7, and 9) and corresponding mRNA expression of *IGF2(6)* tended to be biallelic (Figure 7C and Supplementary Table S5). The SNP7 was found on *IGF2-AS* but the expression was too low to determine its allelic expression (Figures 7A,C and Supplementary Table S5) and therefore, its allelic expression could not be determined.

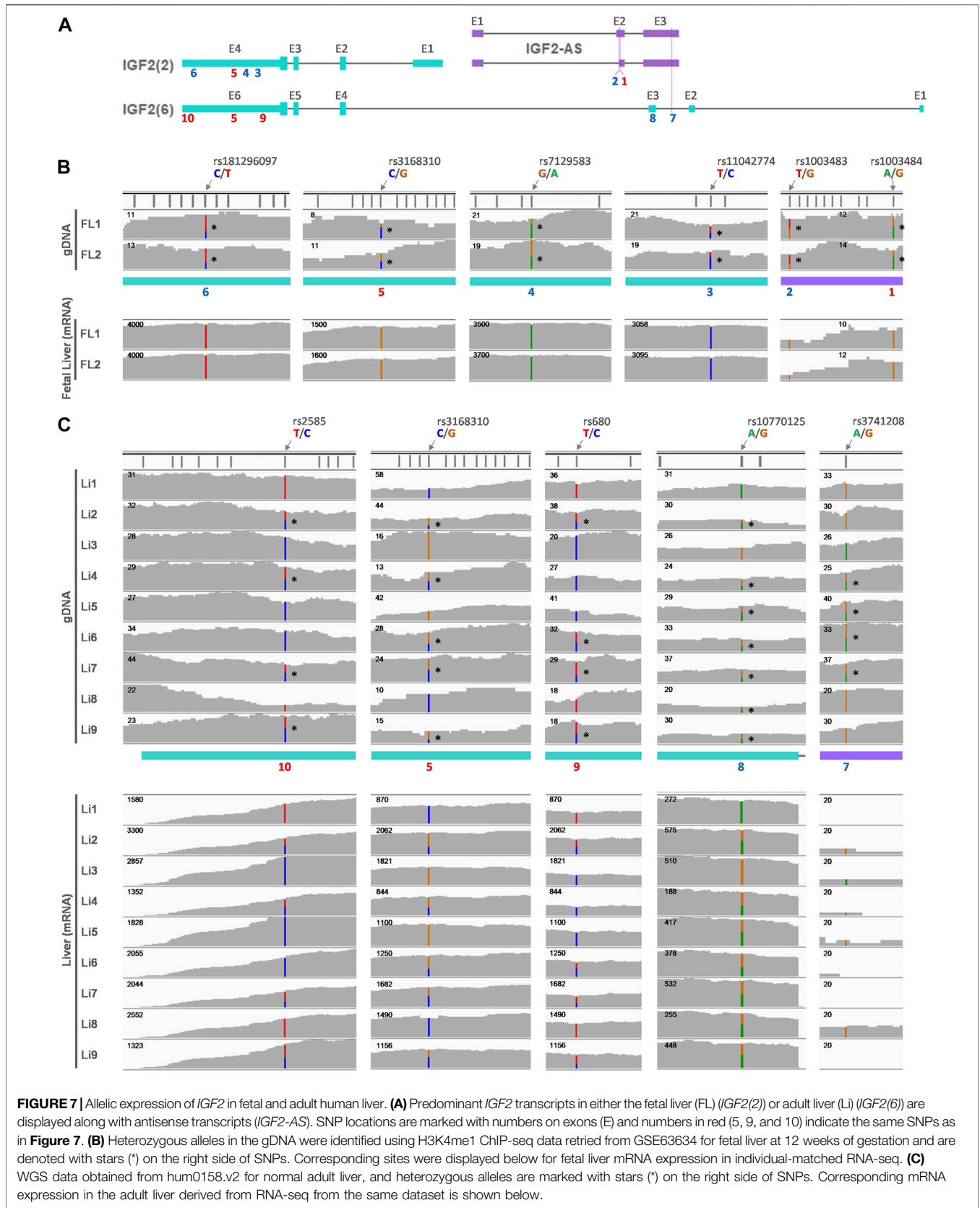
Moreover, in smooth muscle and lungs, informative SNPs on *IGF2(1)*, *IGF2(2)*, and *IGF2-AS* were explored in gDNA (Figure 8A) and their allelic expression was examined. In smooth muscle, both *IGF2(1)* and *IGF2(2)* were expressed, and SNP11-15 were heterozygous in gDNA (Figures 8A,B and Supplementary Table S6). Allelic expression of SNP11-15 tended to be monoallelic (Figure 8B bottom panel and Supplementary Table S6), indicating monoallelic expression of *IGF2(1)* and *IGF2(2)*. The SNP1 was heterozygous in smooth



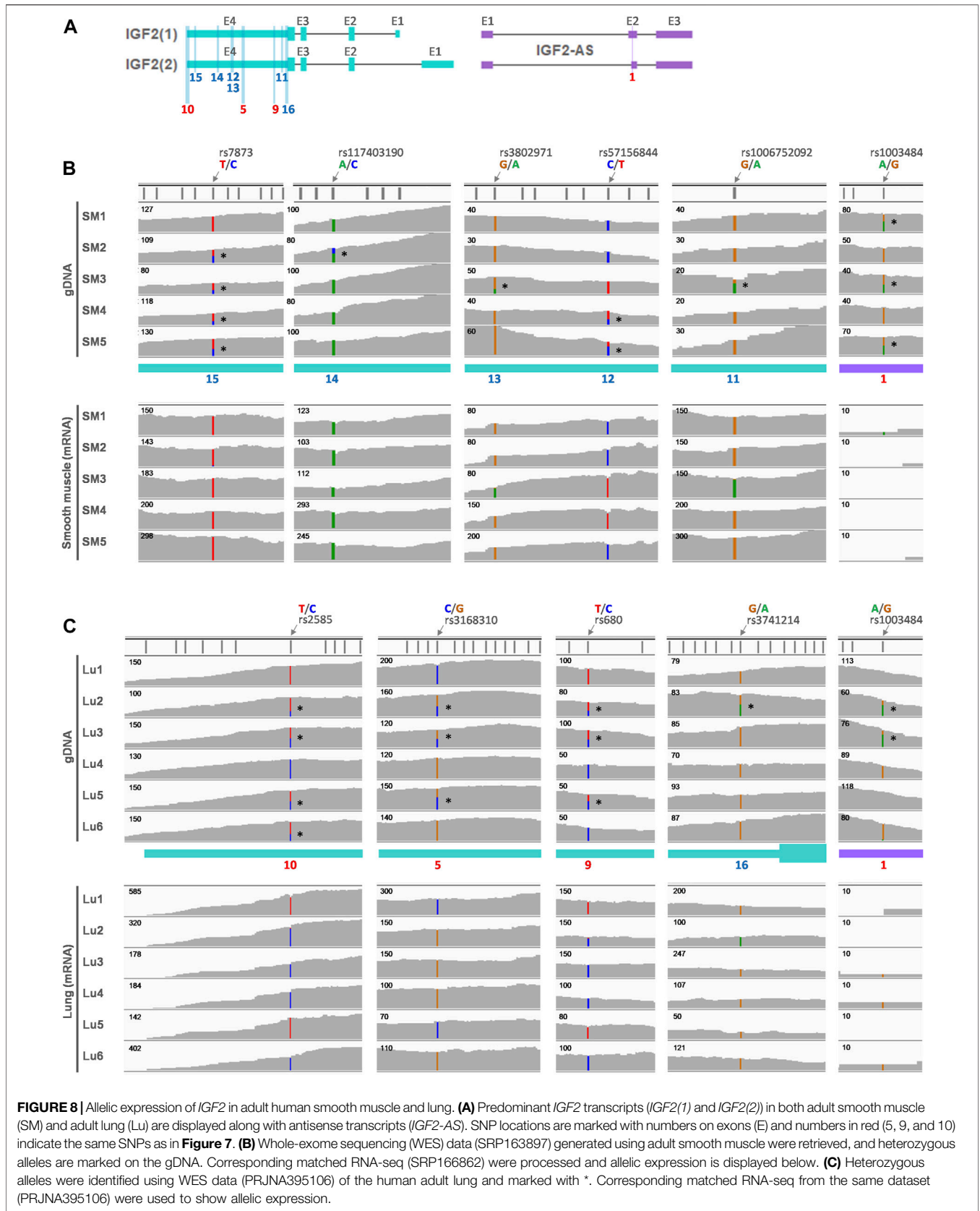
muscle, but the mRNA expression was very low (Figures 8A,B and Supplementary Table S6). In the lungs, both *IGF2(1)* and *IGF2(2)* were expressed, and the same SNP5, 9, and 10 as in the liver were found to be heterozygous (Figures 8A,C and Supplementary Table S6). Unlike biallelic expression in the liver, expression on SNP5, 9, and 10 was monoallelic in the lungs (Figure 8C bottom panel and Supplementary Table S6). In addition, SNP16 was found to be heterozygous in individual 2, and its expression tended to be monoallelic in the lungs (Figures

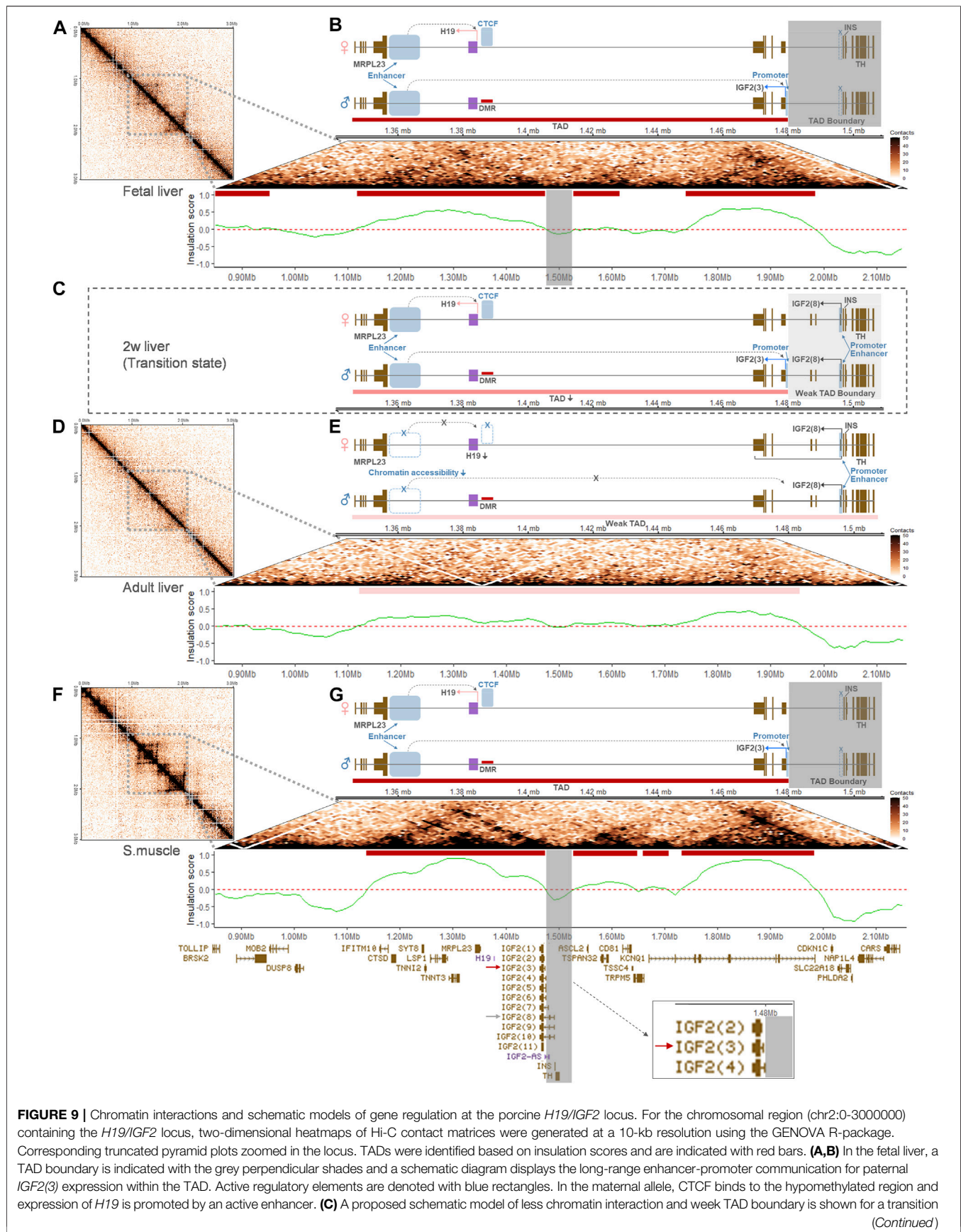
8A,C and Supplementary Table S6). Regarding SNP1, the mRNA expression of *IGF2-AS* was very low in the lungs (Figures 8A,C and Supplementary Table S6).

Overall, in the human, *IGF2(1)* and/or *IGF2(2)* were expressed in the fetal liver, adult smooth muscle, and adult lung, and the expression tended to be monoallelic, whereas *IGF2(6)* was expressed in the adult liver and the expression tended to be biallelic. In addition, the *IGF2-AS* expression was very low, except in the fetal liver where the expression tended to be monoallelic.



**FIGURE 7** | Allelic expression of *IGF2* in fetal and adult human liver. **(A)** Predominant *IGF2* transcripts in either the fetal liver (FL) (*IGF2(2)*) or adult liver (Li) (*IGF2(6)*) are displayed along with antisense transcripts (*IGF2-AS*). SNP locations are marked with numbers on exons (E) and numbers in red (5, 9, and 10) indicate the same SNPs as in Figure 7. **(B)** Heterozygous alleles in the gDNA were identified using H3K4me1 ChIP-seq data retrieved from GSE63634 for fetal liver at 12 weeks of gestation and are denoted with stars (\*) on the right side of SNPs. Corresponding sites were displayed below for fetal liver mRNA expression in individual-matched RNA-seq. **(C)** WGS data obtained from hum0158.v2 for normal adult liver, and heterozygous alleles are marked with stars (\*) on the right side of SNPs. Corresponding mRNA expression in the adult liver derived from RNA-seq from the same dataset is shown below.







**FIGURE 9** | state around 2 weeks (2w) for the porcine liver. Both paternal *IGF2(3)* expression and biallelic *IGF(8)* expression are indicated. **(D,E)** In the adult liver, a week TAD throughout the region is marked with pink bars. Also, low chromatin accessibility in the downstream of *H19* and altered gene regulation are represented in the diagram, along with biallelic *IGF(8)* expression indicated by black bent arrows on both alleles. **(F,G)** In skeletal muscle, the grey perpendicular shades denote a TAD boundary and paternal *IGF2(3)* expression is indicated, similarly to the ones in the fetal liver **(A,B)**. Gene transcripts expressed throughout the region are displayed in the bottom track. A red arrow marks the predominant *IGF2* isoform, *IGF2(3)*, in the fetal liver **(B)** and skeletal muscle **(G)**, and a grey arrow points at the long-form, *IGF2(8)*, which is expressed predominantly in the adult liver **(E)**. Both *IGF2(3)* and *IGF2(8)* are expressed in the proposed transition state of pigs **(C)**.

### 3.6 Topologically Associating Domains in the *H19/IGF2* Imprinted Cluster and Schematic Models Represent Insulation and Imprinting Boundaries

Hi-C datasets were analyzed to identify topologically associating domains (TADs) and TAD boundaries within and adjacent to the *H19/IGF2* locus. To determine whether Hi-C data from replicates of the liver and skeletal muscle can be merged, concordance of the data was estimated using GenomeDISCO (Ursu et al., 2018). For both fetal and adult livers, concordance scores for all pairwise comparisons of Hi-C matrices in multiple resolutions passed a threshold of 0.8 while also passing the recommended threshold of 0.8 at a 50-kb resolution (Ursu et al., 2018), although the scores tended to decrease in higher resolutions (**Supplementary Figure S9**). In addition, smoothed matrices produced from GenomeDISCO procedures displayed similarities within the tissue groups, but not between the groups (**Supplementary Figure S10–S15**). For skeletal muscle of 2-week-old pigs, concordance scores from pairwise comparison also passed the threshold and smoothed matrices were different from those of fetal and adult livers (**Supplementary Figures S16, S17**). Total number of trimmed paired-end reads of each group were similar: ~2.35 billion for fetal liver, ~2.90 billion for the liver, and ~2.24 billion for skeletal muscle (**Supplementary Table S7**). Based on high concordance and comparable amount of the data, matrices of three, three, and two replicates for each group (fetal liver, adult liver, and skeletal muscle), respectively, were merged to increase resolution. After merging, Hi-C matrices at a 10-kb resolution were visualized using two-dimensional heatmaps, and the matrices of the fetal liver and skeletal muscle exhibited stronger contact interactions between approximately 1.0 and 2.0 Mb of pig chromosome 2 than that of the adult liver (**Figure 9**). In the fetal liver and skeletal muscle, boundaries between TADs (i.e., TAD boundaries) encompassing the locus where the first exon of *IGF2(8)* starts, but not the first exon of *IGF2(3)*, was found (**Figures 9A,F**). In addition, muscle from fetal pigs contained a TAD boundary at the locus containing the first exon of *IGF2(8)* which was not overlapped with the first exon of *IGF2(3)* (**Supplementary Figure S18**). In contrast, in the adult liver, a week TAD throughout the region based on insulation scores was revealed, and TAD boundaries were absent at the *H19/IGF2* locus (**Figure 9D**). We expected lowered chromatin interaction in the liver at a transition state based on the enriched regulatory elements and gene expression patterns (**Figure 3**), but Hi-C data were absent (**Figure 9C**). In schematic diagrams, we propose chromatin reorganization in the liver that can weaken TAD and TAD boundaries and alter gene regulation, resulting in conversion of monoallelic expression of

*IGF2(3)* in the fetal liver (**Figure 9B**) via both monoallelic expression of *IGF2(3)* and biallelic expression of *IGF2(8)* in the neonatal liver (**Figure 9C**) to biallelic expression of *IGF2(8)* in the adult liver (**Figure 9E**). In skeletal muscle, consistent *IGF2(3)* expression throughout the development and similar chromatin interaction frequencies between pre- and post-natal stages suggested robust imprinted gene regulation underlying monoallelic expression of the *IGF2* transcript (**Figure 9G**).

## 4 DISCUSSION

In this study, we present comprehensive imprinting status of the conserved paternally imprinted *H19/IGF2* cluster including developmentally regulated and tissue-specific allelic *IGF2* gene expression in the pig and human. By comparing methylome of parthenogenetic (diploid uni-maternal) embryos with bi-parental control embryos, while reducing genetic variability with triplicates of each sample, the porcine *H19* DMR was identified. Previously, the paternal methylation imprint on the *H19* germline DMR, which is fully methylated in sperm and unmethylated in oocytes, was reported in pigs in the form of a group of three DMRs (Park et al., 2009). On the paternal allele of the *H19* DMR, however, demethylation temporarily occurs and then it is remethylated by the morula stage (Park et al., 2009). In addition, differential expression of *IGF2* between androgenetic, parthenogenetic, and *in vitro* fertilized control embryos was previously observed from the blastocyst stage around day 10 (Park et al., 2011). Because, in this study, parthenogenetic and control embryos were recovered later at embryonic day 21 at which the dynamic methylation changes were passed, the detected *H19* DMR could be consistent with the germline DMR between sperm and oocytes. The recovery day 21 was also before morphological degeneration of parthenogenetic embryos occurs at around day 30–35 (Bischoff et al., 2009; Hwang et al., 2020) so that we could prevent confounding effects other than genetic effects. On the other hand, putative *IGF2* DMRs, which were hypermethylated in sperm DNA of Swiss Landrace and Swiss Large White (Giannini and Braunschweig, 2009), were not found in the current study possibly due to breed-specific effects on DNA methylation as described previously (Hwang et al., 2020).

Integrative analyses of ATAC-seq, ChIP-seq, and RNA-seq datasets provide effective strategies to precisely and spatiotemporally elucidate epigenetic regulatory elements and their genetic variations that affect gene expression (Buenrostro et al., 2013; Lara-Astiaso et al., 2014; Floc'hlay et al., 2020). While variations on the regulatory DNA are often buffered and

compensated by other regulatory elements so that redundant regulatory signals might be present in the *H19/IGF2* locus (Floc'hlay et al., 2020), clear distinctions of regulatory layers between the liver and muscle tissues were identified (Figure 2). As the ATAC signals are substantially correlated with H3K27ac (Lara-Astiaso et al., 2014), co-occurrence of ATAC and H3K27ac downstream of *H19* suggested that active enhancers were established from a poised state concomitantly with formation of open chromatin sites upon developmental and signaling cues (Creyghton et al., 2010). Interestingly, this activation of enhancers occurred in skeletal muscle, but not in the liver, of 6-month-old pigs, which leads to recruitment of tissue-specific transcription factors and drives tissue-specific gene expression (Ong and Corces, 2011). In addition, in eukaryotes, H3K4me3 is associated with transcriptional activation on active promoters and typically restricted to narrow regions at the 5' end of the gene body (Santos-Rosa et al., 2002; Schneider et al., 2004; Pena et al., 2006; Wysocka et al., 2006). H3K4me3 marks overlapping the CTCF signal immediately upstream of *H19* might represent the active *H19* promoter regulated by binding of the transcription factor CTCF at close range. Also, H3K4me3 marks overlapping the promoter regions of *IGF2* transcripts [i.e., *IGF2(3)* and *IGF2(8)*] might represent transcriptionally active *IGF2* promoters. At both proximal and distal regions of TSSs the H3K27ac signal can be found (Creyghton et al., 2010), and thus overlaps of H3K27ac with H3K4me3 at close proximity to the promoter regions of *IGF2* transcripts might represent both active enhancers and promoters. We found that these overlaps of H3K27ac and H3K4me3 were present near the first exons of both *IGF2(3)* and *IGF2(8)* transcripts in the liver of 2-week-old pigs, indicating distinct gene regulation during early post-natal liver development.

Although many imprinted genes have been studied in the fetal stage because of their relevance to fetal growth (Peters, 2014; Tian, 2014; Tucci et al., 2019), the current study revealed that the expressed transcript isoform of *IGF2* in muscle tissues might be stably maintained during development and its monoallelic expression was identified in post-natal stages suggesting its role in mature muscle. In the liver of fetal pigs, there was a lack of informative SNPs, but our parthenogenesis studies with whole embryos showed paternal monoallelic expression in the embryonic stage (Figure 1B). In the human, it has been reported that *IGF2* gene transcription is driven by multiple promoters in fetal and non-hepatic adult tissues (Holthuisen et al., 1993; Monk et al., 2006), but we showed that the major form in these tissues of humans was *IGF2(2)* which is orthologous to porcine *IGF2(3)*. On the other hand, the liver-specific promoter (P1) drives *IGF2* gene transcription in the adult liver (Holthuisen et al., 1993; Monk et al., 2006). The corresponding adult liver-specific transcript is not currently annotated in the NCBI Gene database (<https://www.ncbi.nlm.nih.gov/gene/3481>), but we revealed the expression of adult liver-specific *IGF2(6)* (Supplementary Figure S7) which was orthologous to the porcine long-form transcript [*IGF2(8)*]. The *IGF2* transcripts from human fetal tissues including the liver is paternally imprinted and monoallelic, but expression becomes biallelic in the adult liver (Kalscheuer et al., 1993). In addition to this biallelic conversion,

we revealed that a relatively high *IGF2* expression occurred in the adult human liver compared to other analyzed tissues where monoallelic expression remains (Figure 6 and Supplementary Figure S6). Our analyses using pigs support the biallelic conversion and alternative promoter usage that might occur gradually at post-natal ages while ages for initiation of the conversion might vary (Figure 4). These allelic expression patterns were verified based on individual-matched genomic DNA sequence data from WGS and mRNA sequence data from RNA-seq in both pigs and humans (Figures 5, 7, and 8) using informative SNPs found on genomic DNA that served as markers to confirm allelic imbalance of mRNA expression (Castel et al., 2015; Ahn et al., 2021b; a). We primarily examined SNPs in non-overlapping exons to identify and analyze allelic expression at the isoform level while there is a previous study relied on a marker in the last overlapping exon (Braunschweig et al., 2011). In contrast to the *INS-IGF2* read-through script whose expression is spatially regulated in pancreatic islets (Jian and Felsenfeld, 2021), the expression of *IGF2-AS* transcript has been shown to be developmentally regulated as its imprinted paternal expression is relatively high in fetal stages and decreased in adults in both pigs and humans (Okutsu et al., 2000; Braunschweig et al., 2004). Moreover, the notion that the expression of *IGF2-AS* in fetal stages interferes with overlapping *IGF2* (Braunschweig et al., 2004) might be supported by our findings: relatively high expression of *IGF2-AS* in the fetal liver coincided with negligible expression of the long-forms [porcine *IGF2(8)* and human *IGF2(6)*] and low expression of *IGF2-AS* in adults coincided with high expression of the long-forms. However, it is also expected that the antisense role of *IGF2-AS* might be limited in normal tissues due to its low expression compared with high expression of *IGF2*, although increased expression of *IGF2-AS* has been reported in Wilms tumors (Okutsu et al., 2000). Rather, the biallelic conversion in the liver might ensue changes in chromatin structure and regulatory elements as well as antisense expression as discussed below.

The differences in chromatin interaction between fetal and adult livers (and also between skeletal muscle and adult liver) suggested not only transcript conversion, but also chromatin remodeling might occur toward changes in gene regulatory elements and reduce long-range enhancer-promoter communication in the adult liver (Figure 9 and Supplementary Figure S18). In particular, hemimethylation at the *H19* DMR in the adult liver was reported indicating maintenance of the imprint (Braunschweig et al., 2011); however, compared to the fetal liver, chromatin interaction indicated by the self-interacting TADs became weaker in the adult liver of pigs. This lower interaction might be related to less activity of the distal enhancer for the long-range communication, and removal of TAD boundaries might lead to use of the proximal enhancer for the long *IGF2* transcript in the liver. Also, a linkage between less *H19* expression indicating the weak distal enhancer and expression of the biallelic *IGF2* transcript (Ohlsson et al., 1994) was consistently observed in the livers of both pigs and humans (Supplementary Figures S2 and S6). Additionally, in between the fetal and adult stages, there might be a transition state in the liver that is characterized by co-existence of TADs and a weak TAD boundary which is permissive to the proximal enhancer activity

(Gong et al., 2018). In contrast, in skeletal muscle, TADs and TAD boundaries in fetal stages appeared to maintain in post-natal ages. Whether these TADs and TAD boundaries are variable at the single cell level will need to be further investigated (Farabella and Marti-Renom, 2020; Luppino et al., 2020), but existence of TADs and TAD boundaries at the porcine *IGF2* locus was evident. Their significant changes and remodeling in the liver might contribute to liver-specific modifications of *IGF2* allelic expression patterns. Our presentation of the pig Hi-C fills the gap in mammalian genomics, but unfortunately, in the human, Hi-C data from liver (GSE58752) and muscle (GSE87112) tissues that we processed displayed a very low resolution for this relatively narrow range of the *H19/IGF2* locus. Based on our current study that advances our understanding on tissue-specific genomic imprinting in the *H19/IGF2* cluster, studies on other animal species using multi-omics data can further comparatively delineate the *H19/IGF2* locus. Also, gene annotations for porcine *H19* and *IGF2-AS* and human *IGF2(6)* need to be updated due to their lack in the NCBI Gene database ([www.ncbi.nlm.nih.gov/gene](http://www.ncbi.nlm.nih.gov/gene)). In this study, based on previous studies reported *H19* and *IGF2-AS* expression status in pigs (Li et al., 2008; Braunschweig et al., 2011) and *IGF2(6)* in humans (Holthuizen et al., 1993; Monk et al., 2006) as well as our alignment and sequencing results, these genes were analyzed to present the complete landscape of genomic imprinting.

## 5 CONCLUSION

Our integrative omics analyses of genome, epigenome, and transcriptome revealed a comprehensive imprinting status at the *H19/IGF2* locus in pigs in comparison with humans. The porcine *H19/IGF2* imprinting cluster represented a long-term influence of genomic imprinting in muscle tissues but not in the liver which might be similar to that of the orthologous human gene cluster. To the best of our knowledge, this is the first study that describes relatedness between mono- to biallelic conversion of *IGF2* and alternative promoter usage in reorganized chromatin in the liver of adult pigs. The current approaches can be applied in cross-tissue and cross-species analyses to elucidate epigenetic mechanisms that underlie tissue growth and development.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI GEO repository under accession number GSE195528. The publicly available datasets analyzed for this study are

## REFERENCES

- Ahn, J., Hwang, I. S., Park, M. R., Cho, I. C., Hwang, S., and Lee, K. (2020a). *The Landscape of Genomic Imprinting at the Porcine SGCE/PEG10 Locus from Methylome and Transcriptome of Parthenogenetic Embryos*, 10, 4037–4047. doi:10.1534/g3.120.40142G3 *Genes|Genomes|Genetics (Bethesda)*
- Ahn, J., Hwang, I. S., Park, M. R., Hwang, S., and Lee, K. (2021a). *Genomic Imprinting at the Porcine DIRAS3 Locus*, 11, 1315. doi:10.3390/ani11051315 *Anim. (Basel)*
- Ahn, J., Hwang, I. S., Park, M. R., Hwang, S., and Lee, K. (2021b). *Genomic imprinting at the porcine PLAGL1 locus and the orthologous locus in the human*. *Genes (Basel)* 12, 541. doi:10.3390/genes12040541
- Ahn, J., Wu, H., Lee, J., Hwang, I. S., Yu, D., Ahn, J. S., et al. (2020c). *Identification of a novel imprinted transcript in the porcine GNAS complex locus using*

summarized in Materials and Methods and listed in **Supplementary Table S8**.

## ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of the National Institute of Animal Science, Rural Development Administration (RDA) of Korea (approval number NIAS 2015-670).

## AUTHOR CONTRIBUTIONS

JA, SH, and KL conceived this project. JA contributed to methodologies of next-generation sequencing and Hi-C, JL and D-HK conducted polymorphism-based sequencing. I-SH, I-CC, and M-RP contributed to parthenogenesis. JA and KL contributed to the analysis of omics data. JA wrote the manuscript. JA, D-HK, and KL reviewed and edited the manuscript. SH and KL contributed to the funding acquisition. All authors approved the final manuscript.

## FUNDING

This work was partially supported by the United States Department of Agriculture National Institute of Food and Agriculture Hatch Grant (Project No. OHO01304).

## ACKNOWLEDGMENTS

The National Bioscience Database Center (NBDC, <https://humandbs.biosciencedbc.jp/>) allowed our access to WGS and RNA-seq data of normal human liver. High performance computing was conducted using supercomputers at the Ohio Supercomputer Center ([www.osc.edu](http://www.osc.edu)). We are thankful to Michelle Milligan for proofreading this manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.920641/full#supplementary-material>

- methyloome and transcriptome of parthenogenetic fetuses. *Genes (Basel)* 11, E96. doi:10.3390/genes11010096
- Ahn, J., Wu, H., Lee, J., Hwang, I. S., Yu, D., Ahn, J. S., et al. (2020b). Identification of a novel imprinted transcript in the porcine GNAS complex locus using methyloome and transcriptome of parthenogenetic fetuses. *Genes (Basel)* 11, E96. doi:10.3390/genes11010096
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461. doi:10.1038/nature12787
- Barlow, D. P., and Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6, a018382. doi:10.1101/cshperspect.a018382
- Bartolomei, M. S., Webber, A. L., Brunkow, M. E., and Tilghman, S. M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev.* 7, 1663–1673. doi:10.1101/gad.7.9.1663
- Bartolomei, M. S., Zemel, S., and Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature* 351, 153–155. doi:10.1038/351153a0
- Beacon, T. H., Delcuve, G. P., Lopez, C., Nardocci, G., Kovalchuk, I., Van Wijnen, A. J., et al. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clin. Epigenetics* 13, 138. doi:10.1186/s13148-021-01126-1
- Bell, A. C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482–485. doi:10.1038/35013100
- Bischoff, S. R., Tsai, S., Hardison, N., Motsinger-Reif, A. A., Freking, B. A., Nonneman, D., et al. (2009). Characterization of conserved and nonconserved imprinted genes in swine. *Biol. Reprod.* 81, 906–920. doi:10.1095/biolreprod.109.078139
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Braunschweig, M. H., Owczarek-Lipska, M., and Stahlberger-Saitbekova, N. (2011). Relationship of porcine IGF2 imprinting status to DNA methylation at the H19 DMD and the IGF2 DMRs 1 and 2. *BMC Genet.* 12, 47. doi:10.1186/1471-2156-12-47
- Braunschweig, M. H., Van Laere, A. S., Buys, N., Andersson, L., and Andersson, G. (2004). IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3. *Genomics* 84, 1021–1029. doi:10.1016/j.ygeno.2004.09.006
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218. doi:10.1038/nmeth.2688
- Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195. doi:10.1186/s13059-015-0762-6
- Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L., and Frommer, M. (2006). DNA methylation: Bisulphite modification and analysis. *Nat. Protoc.* 1, 2353–2364. doi:10.1038/nprot.2006.324
- Coppola, C. J., Ramaker, C. R., and Mendenhall, E. M. (2016). Identification and function of enhancers in the human genome. *Hum. Mol. Genet.* 25, R190–R197. doi:10.1093/hmg/ddw216
- Crane, E., Bian, Q., Mccord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., et al. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 523, 240–244. doi:10.1038/nature14450
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936. doi:10.1073/pnas.1016071107
- Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castello, A., Benitez, R., Fernandez, A. I., et al. (2019). Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition. *PLoS One* 14, e0220708. doi:10.1371/journal.pone.0220708
- DeChiara, T. M., Robertson, E. J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859. doi:10.1016/0092-8674(91)90513-x
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635
- Farabella, I., and Marti-Renom, M. A. (2020). TADs without borders. *Nat. Genet.* 52, 752–753. doi:10.1038/s41588-020-0670-x
- Ferguson-Smith, A. C., Cattanaach, B. M., Barton, S. C., Beechey, C. V., and Surani, M. A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature* 351, 667–670. doi:10.1038/351667a0
- Ferguson-Smith, A. C. (2011). Genomic imprinting: The emergence of an epigenetic paradigm. *Nat. Rev. Genet.* 12, 565–575. doi:10.1038/nrg3032
- Floc'hlay, S., Wong, E., Zhao, B., Viales, R. R., Thomas-Chollier, M., Thieffry, D., et al. (2020). Cis-acting variation is common across regulatory layers but is often buffered during embryonic development. *Genome Res.* 31, 211–224. doi:10.1101/gr.266338.120
- George, J. W., Fan, H., Johnson, B., Carpenter, T. J., Foy, K. K., Chatterjee, A., et al. (2019). Integrated epigenome, exome, and transcriptome analyses reveal molecular subtypes and homeotic transformation in uterine fibroids. *Cell Rep.* 29, 4069–4085. doi:10.1016/j.celrep.2019.11.077
- Giannini, P., and Braunschweig, M. (2009). DNA methylation patterns at the IGF2-H19 locus in sperm of Swiss Landrace and Swiss Large White boars. *J. Anim. Breed. Genet.* 126, 475–479. doi:10.1111/j.1439-0388.2009.00802.x
- Gong, Y., Lazaris, C., Sakellaropoulos, T., Lozano, A., Kambadur, P., Ntziachristos, P., et al. (2018). Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* 9, 542. doi:10.1038/s41467-018-03017-1
- Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. *Methods Mol. Biol.* 1418, 335–351. doi:10.1007/978-1-4939-3578-9\_16
- Haig, D. (2004). Genomic imprinting and kinship: How good is the evidence? *Annu. Rev. Genet.* 38, 553–585. doi:10.1146/annurev.genet.37.110801.142741
- Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E., and Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 365, 764–767. doi:10.1038/365764a0
- Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., Tilghman, S. M., et al. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489. doi:10.1038/35013106
- Hayward, B. E., Kamiya, M., Strain, L., Moran, V., Campbell, R., Hayashizaki, Y., et al. (1998a). The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10038–10043. doi:10.1073/pnas.95.17.10038
- Hayward, B. E., Moran, V., Strain, L., and Bonthron, D. T. (1998b). Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15475–15480. doi:10.1073/pnas.95.26.15475
- Holthuizen, P., Van Dijk, M. A., Rodenburg, R. J., Koonen-Reemst, A. M., and Sussenbach, J. S. (1993). Transcriptional regulation of the major promoters of the human IGF-II gene. *Mol. Reprod. Dev.* 35, 391–393. doi:10.1002/mrd.1080350412
- Howe, F. S., Fischl, H., Murray, S. C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? *Bioessays* 39, 1–12. doi:10.1002/bies.201600095
- Hwang, I. S., Park, M. R., Lee, H. S., Kwak, T. U., Son, H. Y., Kang, J. K., et al. (2020). *Developmental and Degenerative Characterization of Porcine Parthenogenetic Fetuses during Early Pregnancy*. 10. *Anim. (Basel)*. doi:10.3390/ani10040622
- Jacob, K. J., Robinson, W. P., and Lefebvre, L. (2013). Beckwith-wiedemann and silver-russell syndromes: Opposite developmental imbalances in imprinted regulators of placental function and embryonic growth. *Clin. Genet.* 84, 326–334. doi:10.1111/cge.12143
- Jian, X., and Felsenfeld, G. (2021). Large parental differences in chromatin organization in pancreatic beta cell line explaining diabetes susceptibility effects. *Nat. Commun.* 12, 4338. doi:10.1038/s41467-021-24635-2
- Juhling, F., Kretzmer, H., Bernhart, S. H., Otto, C., Stadler, P. F., Hoffmann, S., et al. (2016). metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* 26, 256–262. doi:10.1101/gr.196394.115
- Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H., and Ropers, H. H. (1993). The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat. Genet.* 5, 74–78. doi:10.1038/ng0993-74
- Kern, C., Wang, Y., Xu, X., Pan, Z., Halstead, M., Chanthavixay, G., et al. (2021). Functional annotations of three domestic animal genomes provide vital

- resources for comparative and agricultural research. *Nat. Commun.* 12, 1821. doi:10.1038/s41467-021-22100-8
- Krueger, F., and Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27, 1571–1572. doi:10.1093/bioinformatics/btr167
- Kwon, D. J., Kim, D. H., Hwang, I. S., Kim, D. E., Kim, H. J., Kim, J. S., et al. (2017). Generation of alpha-1, 3-galactosyltransferase knocked-out transgenic cloned pigs with knocked-in five human genes. *Transgenic Res.* 26, 153–163. doi:10.1007/s11248-016-9979-8
- Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D. A., David, E., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949. doi:10.1126/science.1256271
- Latos, P. A., Pauler, F. M., Koerner, M. V., Senergin, H. B., Hudson, Q. J., Stocsits, R. R., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472. doi:10.1126/science.1228110
- Li, C., Bin, Y., Curchoe, C., Yang, L., Feng, D., Jiang, Q., et al. (2008). Genetic imprinting of H19 and IGF2 in domestic pigs (*Sus scrofa*). *Anim. Biotechnol.* 19, 22–27. doi:10.1080/10495390701758563
- Li, F., Wang, D., Song, R., Cao, C., Zhang, Z., Wang, Y., et al. (2020). The asynchronous establishment of chromatin 3D architecture between *in vitro* fertilized and uniparental preimplantation pig embryos. *Genome Biol.* 21, 203. doi:10.1186/s13059-020-02095-z
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 3, 13033997.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, M., Chen, L., Tian, S., Lin, Y., Tang, Q., Zhou, X., et al. (2017). Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res.* 27, 865–874. doi:10.1101/gr.207456.116
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi:10.1126/science.1181369
- Lleres, D., Moindrot, B., Pathak, R., Piras, V., Matelot, M., Pignard, B., et al. (2019). CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biol.* 20, 272. doi:10.1186/s13059-019-1896-8
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Lunney, J. K., Van Goor, A., Walker, K. E., Hailstock, T., Franklin, J., Dai, C., et al. (2021). Importance of the pig as a human biomedical model. *Sci. Transl. Med.* 13, eabd5758. doi:10.1126/scitranslmed.abd5758
- Luppino, J. M., Park, D. S., Nguyen, S. C., Lan, Y., Xu, Z., Yunker, R., et al. (2020). Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet.* 52, 840–848. doi:10.1038/s41588-020-0647-9
- Monk, D., Sanches, R., Arnaud, P., Apostolidou, S., Hills, F. A., Abu-Amero, S., et al. (2006). Imprinting of IGF2 P0 transcript and novel alternatively spliced INS-IGF2 isoforms show differences between mouse and human. *Hum. Mol. Genet.* 15, 1259–1269. doi:10.1093/hmg/ddl041
- Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: A parental tug-of-war. *Trends Genet.* 7, 45–49. doi:10.1016/0168-9525(91)90230-N
- Morrison, J., Koeman, J. M., Johnson, B. K., Foy, K. K., Beddows, I., Zhou, W., et al. (2021). Evaluation of whole-genome DNA methylation sequencing library preparation protocols. *Epigenetics Chromatin* 14, 28. doi:10.1186/s13072-021-00401-y
- Nordin, M., Bergman, D., Halje, M., Engstrom, W., and Ward, A. (2014). Epigenetic regulation of the Igf2/H19 gene cluster. *Cell Prolif.* 47, 189–199. doi:10.1111/cpr.12106
- Ohlsson, R., Hedborg, F., Holmgren, L., Walsh, C., and Ekstrom, T. J. (1994). Overlapping patterns of IGF2 and H19 expression during human development: Biallelic IGF2 expression correlates with a lack of H19 expression. *Development* 120, 361–368. doi:10.1242/dev.120.2.361
- Okutsu, T., Kuroiwa, Y., Kagitani, F., Kai, M., Aisaka, K., Tsutsumi, O., et al. (2000). Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. *J. Biochem.* 127, 475–483. doi:10.1093/oxfordjournals.jbchem.a022630
- Ong, C. T., and Corces, V. G. (2011). Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283–293. doi:10.1038/nrg2957
- Park, C. H., Kim, H. S., Lee, S. G., and Lee, C. K. (2009). Methylation status of differentially methylated regions at Igf2/H19 locus in porcine gametes and preimplantation embryos. *Genomics* 93, 179–186. doi:10.1016/j.ygeno.2008.10.002
- Park, C. H., Uh, K. J., Mulligan, B. P., Jeung, E. B., Hyun, S. H., Shin, T., et al. (2011). Analysis of imprinted gene expression in normal fertilized and uniparental preimplantation porcine embryos. *PLoS One* 6, e22216. doi:10.1371/journal.pone.0022216
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi:10.1038/nmeth.4197
- Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., et al. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 442, 100–103. doi:10.1038/nature04814
- Peters, J. (2014). The role of genomic imprinting in biology and disease: An expanding view. *Nat. Rev. Genet.* 15, 517–530. doi:10.1038/nrg3766
- Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., et al. (1999). A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3830–3835. doi:10.1073/pnas.96.7.3830
- Plagge, A., Gordon, E., Dean, W., Boiani, R., Cinti, S., Peters, J., et al. (2004). The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding. *Nat. Genet.* 36, 818–826. doi:10.1038/ng1397
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Bruggmann, S. A., Flynn, R. A., Wysocka, J., et al. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283. doi:10.1038/nature09692
- Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. doi:10.1093/nar/gku365
- Reik, W., and Walter, J. (2001). Genomic imprinting: Parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32. doi:10.1038/35047554
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., et al. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411. doi:10.1038/nature01080
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., Kouzarides, T., et al. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.* 6, 73–77. doi:10.1038/ncb1076
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., et al. (2015). HiC-pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259. doi:10.1186/s13059-015-0831-x
- Shemer, R., Birger, Y., Dean, W. L., Reik, W., Riggs, A. D., and Razin, A. (1996). Dynamic methylation adjustment and counting as part of imprinting mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 93, 6371–6376. doi:10.1073/pnas.93.13.6371
- Shen, Y., Yue, F., Mcleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120. doi:10.1038/nature11243
- Sun, F. L., Dean, W. L., Kelsey, G., Allen, N. D., and Reik, W. (1997). Transactivation of Igf2 in a mouse model of Beckwith-Wiedemann syndrome. *Nature* 389, 809–815. doi:10.1038/39797
- Suntsova, M., Gaifullin, N., Allina, D., Reshetun, A., Li, X., Mendeleeva, L., et al. (2019). Atlas of RNA sequencing profiles for normal human tissues. *Sci. Data* 6, 36. doi:10.1038/s41597-019-0043-4
- Tang, Q., Gu, Y., Zhou, X., Jin, L., Guan, J., Liu, R., et al. (2017). Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. *Gigascience* 6, 1–9. doi:10.1093/gigascience/gix105
- Thorvaldsen, J. L., Duran, K. L., and Bartolomei, M. S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev.* 12, 3693–3702. doi:10.1101/gad.12.23.3693

- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82. doi:10.1038/nature11232
- Tian, X. C. (2014). Genomic imprinting in farm animals. *Annu. Rev. Anim. Biosci.* 2, 23–40. doi:10.1146/annurev-animal-022513-114144
- Tian, X., Li, R., Fu, W., Li, Y., Wang, X., Li, M., et al. (2020). Building a sequence map of the pig pan-genome from multiple de novo assemblies and Hi-C data. *Sci. China. Life Sci.* 63, 750–763. doi:10.1007/s11427-019-9551-7
- Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., and Erice Imprinting, G. (2019). Genomic imprinting and physiological processes in mammals. *Cell* 176, 952–965. doi:10.1016/j.cell.2019.01.043
- Ursu, O., Boley, N., Taranova, M., Wang, Y. X. R., Yardimci, G. G., Stafford Noble, W., et al. (2018). GenomeDISCO: A concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics* 34, 2701–2707. doi:10.1093/bioinformatics/bty164
- Ushiki, A., Zhang, Y., Xiong, C., Zhao, J., Georgakopoulos-Soares, I., Kane, L., et al. (2021). Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. *Nat. Commun.* 12, 2282. doi:10.1038/s41467-021-22470-z
- Van Der Weide, R. H., Van Den Brand, T., Haarhuis, J. H. I., Teunissen, H., Rowland, B. D., De Wit, E., et al. (2021). Hi-C analyses with GENOVA: A case study with cohesin variants. *Nar. Genom. Bioinform.* 3, lqab040. doi:10.1093/nargab/lqab040
- Wiench, M., John, S., Baek, S., Johnson, T. A., Sung, M. H., Escobar, T., et al. (2011). DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* 30, 3028–3039. doi:10.1038/emboj.2011.210
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86–90. doi:10.1038/nature04815
- Xu, J., Pope, S. D., Jazirehi, A. R., Attema, J. L., Papathanasiou, P., Watts, J. A., et al. (2007). Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12377–12382. doi:10.1073/pnas.0704579104
- Yan, F., Powell, D. R., Curtis, D. J., and Wong, N. C. (2020). From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* 21, 22. doi:10.1186/s13059-020-1929-3
- Yan, L., Guo, H., Hu, B., Li, R., Yong, J., Zhao, Y., et al. (2016). Epigenomic landscape of human fetal brain, heart, and liver. *J. Biol. Chem.* 291, 4386–4398. doi:10.1074/jbc.M115.672931
- Yang, Y., Fan, X., Yan, J., Chen, M., Zhu, M., Tang, Y., et al. (2021). A comprehensive epigenome atlas reveals DNA methylation regulating skeletal muscle development. *Nucleic Acids Res.* 49, 1313–1329. doi:10.1093/nar/gkaa1203
- Yoshimizu, T., Miroglio, A., Ripoche, M. A., Gabory, A., Vernucci, M., Riccio, A., et al. (2008). The H19 locus acts *in vivo* as a tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* 105, 12417–12422. doi:10.1073/pnas.0801540105
- Yuan, R., Zhang, J., Wang, Y., Zhu, X., Hu, S., Zeng, J., et al. (2021). Reorganization of chromatin architecture during prenatal development of porcine skeletal muscle. *DNA Res.* 28, dsab003. doi:10.1093/dnares/dsab003
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137. doi:10.1186/gb-2008-9-9-r137
- Zhang, Y., Zhang, L., Yue, J., Wei, X., Wang, L., Liu, X., et al. (2019). Genome-wide identification of RNA editing in seven porcine tissues by matched DNA and RNA high-throughput sequencing. *J. Anim. Sci. Biotechnol.* 10, 24. doi:10.1186/s40104-019-0326-9
- Zhao, Y., Hou, Y., Xu, Y., Luan, Y., Zhou, H., Qi, X., et al. (2021). A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat. Commun.* 12, 2217. doi:10.1038/s41467-021-22448-x
- Zhou, J., So, K. K., Li, Y., Li, Y., Yuan, J., Ding, Y., et al. (2019). Elevated H3K27ac in aged skeletal muscle leads to increase in extracellular matrix and fibrogenic conversion of muscle satellite cells. *Aging Cell* 18, e12996. doi:10.1111/acer.12996

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ahn, Lee, Kim, Hwang, Park, Cho, Hwang and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.