

RESEARCH ARTICLE

Open Access



A comprehensive genome-wide scan for parent-of-origin expressed genes in the pig clarifies the conservation landscape of genomic imprinting

Mathilde Perret¹, Nathalie Iannuccelli¹, Sophie Leroux¹, Katia Fève¹, Patrice Dehais², Eva Jacomet¹, Jean-Noël Hubert¹, Carole Iampietro³, Céline Vandecasteele³, Sarah Maman-Haddad², Thomas Faraut¹, Laurence Liaubet¹, Agnès Bonnet¹, Cécile Donnadieu³, Juliette Riquet¹ and Julie Demars^{1*} 

Abstract

Background Genomic imprinting, a mechanism resulting in parent-of-origin expression of genes through epigenetic regulation, intersects with a broad range of biological fields, including evolution, molecular genetics and epigenetics, and determinism of complex traits. Although next generation sequencing technologies nowadays enable imprinted genes to be detected in a genome-wide manner, a wide spectrum of this phenomena is evaluated only in humans and rodents.

Results Here, we propose to map genes showing a parental expression imbalance in hypothalamus, muscle and placenta in piglets around birth using an extensive strategy that minimized biases and relied on reciprocal crosses, reconstruction of parental phases after imputation, and statistical analyses discriminating parent-of-origin from allele-specific expression. We detected 141 genes with strong to exclusive parental expression imbalance (ratio > 25:75). A large proportion (80%) of these genes have never been shown to exhibit parent-of-origin expression and a small proportion (15%) are shared by at least two tissues, suggesting an overall weak conservation landscape of genomic imprinting. Interestingly, we identified novel parent-of-origin expressed genes involved in neurodevelopmental (*PREPL*, Prolyl Endopeptidase Like) and fetal growth (*FAM20B*, Glycosaminoglycan Xylosylkinase, and *POU6F2*, POU Class 6 Homeobox 2) functions. In-depth analyses of specific loci highlighted specific imprinted isoforms of *COPG2* (COP1 Coat Complex Subunit Gamma 2) and confirmed livestock-specific imprinted genes such as the Zinc Finger Protein 300-like gene.

Conclusions Altogether, our results provide an atlas of parent-of-origin expressed genes in the pig, making it the most documented species for genomic imprinting after humans and rodents. Our findings indicate weak conservation of this mechanism across species and tissues, suggesting a small number of core imprinted genes shared across eutherians and another imprinted genes that seem specific to species or tissues. These latter parent-of-

*Correspondence:

Julie Demars
julie.demars@inrae.fr

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

origin expressed genes may have been subjected to evolutionary forces that have determined their imprinting status in either a livestock-specific or a tissue-specific manner.

Background

Genomic imprinting is an original molecular phenomenon that leads to a preferential allele-specific expression dependent on the parental origin. Genomic imprinting is a form of epigenetic regulation found in placental mammals and flowering plants [1]. Parent-of-origin expressed (POE) genes, often called imprinted genes, are found isolated or as clusters across the genome, representing about 1% of the total gene content in the best studied mammals such as mice, rat, and human species [2]. Imprinted genes include different types, with approximately 1/3 being protein-coding genes and the rest being non-coding RNAs genes comprising long (lncRNAs) and small RNAs (microRNAs and snoRNAs) [3]. The greatest variability of imprinting patterns (i.e. parental expressed allele exhibiting a weak to robust parent-of-origin effect) has been observed for protein-coding genes [4].

Imprinted genes play critical roles in foetal and post-natal development and adult tissue function [2], as supported by various human imprinting disorders (as reviewed in [5]), as well as mutations within imprinted genes associated with major agronomic phenotypes (as reviewed in [6]). Initial studies on genomic imprinting, mainly focused on key imprinted genes expressing exclusively one parental allele, suggested a certain level of conservation of imprinted genes between tissues and species [7–9]. More recently, multiple studies have demonstrated differences between both tissues and development stages, with the placenta and brain showing the highest frequency of imprinted genes [10–12]. These results are supported by Bonthuis et al. [13], who highlighted that the number of imprinted genes in the arcuate nucleus of the hypothalamus was 100 to 300% higher than in somatic skeletal tissues.

More and more studies are now focusing on the evolution of genomic imprinting across species. Comparison of the genomic imprinting phenomenon through the phylogeny of eutherians has brought novel insights on molecular mechanisms of the acquisition of imprinting by distinguishing canonical from non-canonical imprinting, the latter being exclusively identified in muridae so far [14]. While canonical imprinting is driven by DNA methylation, non-canonical imprinting, a form of oocyte-specific acquisition of genomic imprinting, is established by the apposition of H3K27me3 on retrovirus sequences, including long terminal repeats (LTR) [15]. Non-canonical imprinting results in maternally DNA methylated imprints that govern paternal-specific gene expression in extra-embryonic cells. To date, humans and rodents are the species that have been most extensively studied and

some specific studies have been dedicated to marsupials [16, 17]. However, genomic imprinting studies on other species, including livestock, are often restricted to a few orthologous genes [8, 9, 18], leading to a gap of knowledge about genomic imprinting across evolutionary space.

Next-generation sequencing technologies focused on RNA have enabled the identification of imprinted genes in a genome-wide manner by allowing the detection of a large spectrum of parental imbalances in transcripts from various biological samples [19]. However, the application of such novel approaches to study genomic imprinting mechanisms requires dedicated experimental designs and specific bioinformatics and statistical methods to avoid false positive results [20]. Many studies have relied on reciprocal crosses between genetically and phenotypically divergent strains of mice [21], rats [14], breeds of pigs [22], and even subspecies (cattle) [23], in order to maximise the heterozygosity in offspring and leverage POE. Although these experimental designs are powerful, they require rigorous and stringent analysis in order to discriminate between the following two classes of mono-allelic expression [19]: allele-specific expression (ASE), which represents unequal expression based on nucleotide identity, which is sometimes referred to as allelotype expression (AGE) [22, 24], and POE, which represents unequal expression based on parental origin. Most, but not all, AGE are breed-specific expression contrasts, since genetically divergent breeds are characterized by having a high number of variants that are fixed for alternate alleles in the breeds that are crossed [14, 22, 24]. Evaluating all genetic variability, rather than focusing only on breed-specific variants, enhances the detection of both AGE and POE. This approach, although rarely performed, can be applied in livestock species since parental breeds are polymorphic and carry heterozygous variants that are shared between breeds.

Another bias when identifying novel imprinted genes stems from the quality of genome annotations, in particular for non-coding RNAs. However, recent data obtained in pigs has highlighted novel POE genes [22, 25–27], including livestock-specific imprinted genes such as *KBTD6* (Kelch Repeat And BTB Domain Containing 6) [25], *ZNF791*-like (Zinc Finger Protein 791) [26], and *ZNF300*-like (Zinc Finger Protein 300) [27]. This underscores the importance of acquiring in-depth knowledge of genomic imprinting in various mammalian species in order to measure its extent and conservation and to study the role of imprinted genes in the determinism of complex traits in domestic species.

Here, we propose to map imprinted genes in hypothalamus, muscle, and placenta in piglets around birth, using an extensive strategy by minimizing biases. This approach relies on reciprocal outbred crosses and reconstruction of the parental phases after imputation in order to discriminate between POE and AGE, taking into account all genetic variability. Genome-wide comparison of both number and distribution of imprinted genes along the pig genome has shown weak conservation across tissues and species (pig vs. human and mouse species). However, when parental expressed genes are shared across eutherians, they exhibit a strong conservation of their imprinting patterns (i.e. the direction of parental expressed allele). In addition, deeper analyses of specific loci confirmed livestock-specific imprinted genes, including the *ZNF300-like* gene (Zinc Finger Protein 300) (*LOC100520903*), as well as specific imprinted isoforms of *COPG2* (COPI Coat Complex Subunit Gamma 2), a gene showing conflicting data in the literature.

Methods

Ethical statement

All procedures and guidelines for animal care were approved by the local ethical committee on animal experimentation (Poitou–Charentes) and the French Ministry of Higher Education and Scientific Research (authorizations no 2018021912005794 and no 11789-2017101117033530). The use of animals and the procedures performed in this study to obtain placental samples (dataset 2) were approved by the European Union legislation (directive 86/609/EEC) and French legislation in the Midi-Pyrénées Region of France (Decree 2001–464 29/05/01; accreditation for animal housing C-35–275-32). The technical and scientific staff obtained individual accreditation for experiments (MP/01/01/01/11) from the Ethics Committee (région Midi-Pyrénées, France). Under these conditions, this study follows the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments) and is committed to the 3Rs of laboratory animal research.

Animals and experimental designs

In order to detect parental expression imbalance, we relied on two datasets that were generated on reciprocal crosses between two genetically distant pig breeds, the European Large White (LW) breed and the Asian Meishan (MS) breed (Fig. 1A). Piglets from the first dataset were produced to highlight parental expression imbalance in muscle and hypothalamus at day 1 after birth by combining transcriptomic data obtained on these tissues from piglets and genomic data obtained from the blood of both the piglets and their parents. Dataset 1 included 5 parents and their 9 offspring distributed as: 2 sows, 1

boar, and 6 piglets from the cross 1 and 1 sow, 1 boar, and 3 piglets from cross 2 (Fig. 1A).

The second dataset was produced to target imprinting genes in the placenta at day 110 of gestation. This developmental stage corresponds to a mature stage in pig, since the gestation length ranges from 111 to 120 days from the first day of mating. Dataset 2 included 24 parents and their 54 offspring, distributed as: 10 sows, 2 boars, and 12 fetuses from cross 1, and 7 sows, 2 boars, and 42 fetuses from cross 2 (Fig. 1A). Details on animal resources and the whole genetic design can also be found in [28] (Fig. 1A).

Animals to produce dataset 1 were generated in the frame of the PIPETTE project (French National Research Agency ANR-18-CE20-0018), as described in [29]. Animals to produce dataset 2 were generated in the frame of the PORCINET project (French National Research Agency ANR-09-GENM005). Animals from both datasets were produced at the GenESI INRAE experimental farm [30].

Biological sample collection and nucleic acid extraction

For the dataset 1, 14 biological samples were used for genomic DNA extraction. Blood samples were collected on EDTA and stored frozen for nine months at -20°C . Biological samples were collected at the adult developmental stage for all parents ($n=5$), while biological samples were collected at day 1 after birth for all offspring ($n=9$). High molecular weight genomic DNA was extracted from blood using the Genomic-tip 100/G kit (Qiagen, Reference 10,243), as recommended by the manufacturer. Genomic DNA concentrations were measured using the Qubit fluorimetry system with the Broad Range kit (Invitrogen, Reference Q32850) Fragment size distributions were assessed using the Femto Pulse Genomic DNA 165 kb Kit (Agilent). Purity was measured using a Nanodrop system (Thermo Fisher). All samples were purified with beads to obtain expected ratios i.e. 260/280: 1.8–2 and 260/230: 2–2.2.

In addition, genomic DNA was extracted from blood, hypothalamus, and muscle ($n=4$) to perform molecular analyses of the *IGF2* locus.

For transcript RNA extraction from *longissimus dorsi* muscle and hypothalamus from the dataset 1, nine piglets were slaughtered at day 1 after birth. A piece of *longissimus dorsi* and hypothalamus was cut and immediately frozen to avoid degradation of transcripts. RNA extractions for muscle were performed using the Quick-RNA Miniprep Plus Kit (Zymo Research, Reference R1057) following the manufacturer's recommendation. RNA extractions for hypothalamus were performed using the NucleoSpin RNA kit (Macherey–Nagel, Reference 740,955) following the manufacturer's recommendation. RNA concentrations were measured using the Qubit

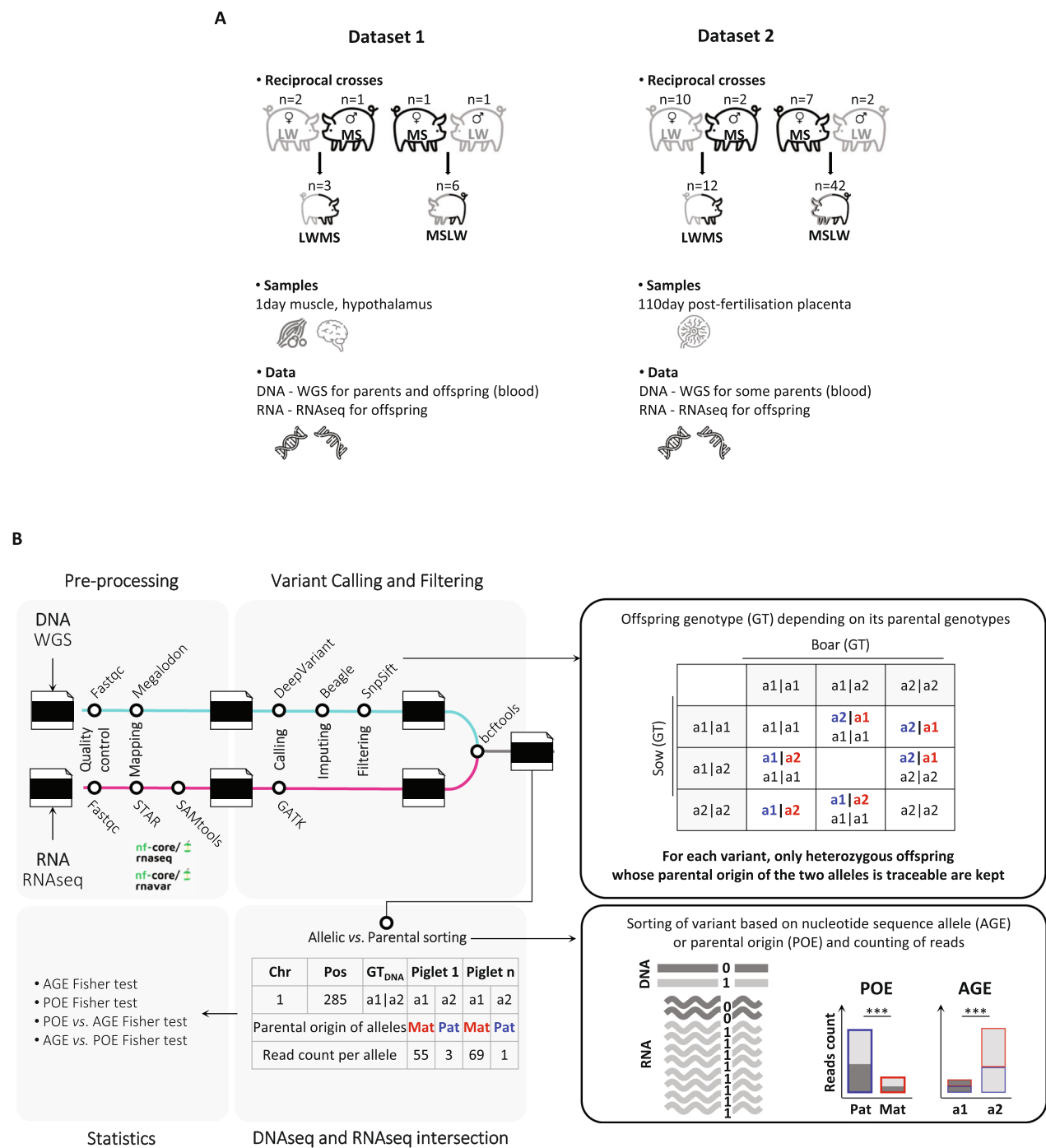


Fig. 1 Experimental designs and approach to detect parental expression. **a** Large White (LW) pigs and Meishan (MS) pigs underwent reciprocal crosses at two different developmental stages (day 1 after birth for dataset 1 and day 110 of gestation for dataset 2). Tissue sample size and sequencing contents across both datasets are mentioned. **b** Roadmap of preprocessing and statistical analyses leading to an extensive strategy to discriminate allele-specific expression (AGE) from parent-of-origin expression (POE). The key step of our strategy relies on the filtering of variant, within trio and duo and for each variant, of heterozygous offspring for which the parental origin of both alleles (a1 and a2) could be identified (in blue for the boar and in red from the sow)

fluorimetry system with the Broad Range kit (Invitrogen, Reference Q10210). Fragment size distributions were assessed using agarose seakem electrophoresis after RNA denaturation. Purity was measured using a Nanodrop 8000 system (Thermo Scientific).

For the dataset 2, the 54 placenta samples and their associated endometrium from crossbred fetuses were collected at day 110 of gestation. RNA extraction was carried out on the frozen powdered placenta samples according to the NucleoSpin RNA, Mini kit for RNA

purification (Macherey–Nagel, Reference 740,955). Preparatory steps, RNA purification, and quality controls are described in the datapaper by Maman et al. [31].

Sequencing technologies

Preparation of libraries and of RNA and DNA sequencing were performed on the GeT-PlaGe core facility in INRAE Toulouse [32]. Long reads Oxford Nanopore Technology (ONT) was used to sequence genomic DNA extracted from blood for dataset 1. ONT libraries were prepared according to the manufacturer's instructions "1D gDNA selecting for long reads (SQK-LSK109)". At each step, DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). DNA purity was tested using the Nanodrop (ThermoFisher) and size distribution and degradation were assessed using the Fragment Analyzer (Agilent) DNF-464 HS Large Fragment Kit. Purification steps were performed using AMPure XP beads (Beckman Coulter). For one Flowcell, 5 µg of DNA was purified and then sheared at 25 kb using the Megaruptor system (diagenode). A size selection step using the Short Read Eliminator M Kit (Circulomics) was performed. A one step DNA damage repair + END-repair + dA tail of double stranded DNA fragments was performed on 2 µg of DNA, after which adapters were ligated to the fragments. The library was loaded onto a FLO-PRO002 R9.4.1 flow cell and sequenced on an PromethION instrument at 20 fmol within 72 h. At 24 h and 48 h, a nuclease flush was performed on the same flow cell, after which the library was re-loaded at 20 fmol. The DNAseq of the dataset 1 (blood) are available from the European Nucleotide Archive (ENA) under accession number PRJEB86771.

For RNA-seq experiments, libraries were prepared by the GeT-PlaGe core facility in INRAE Toulouse [32]. Sequencing was performed on a S4 flow cell of an Illumina NovaSeq6000 machine using a paired-end read length of 2×150 bp. The RNAseq of the first dataset (muscle and hypothalamus) are available from the European Nucleotide Archive (ENA) under the accession number PRJEB86771. For the dataset 2, all protocols and quality controls were submitted to FAANG database and described in Maman et al. [31].

Data preprocessing

Variant calling

All bioinformatic analyses were performed at the genotoul bioinformatics platform Toulouse Occitanie [33] and are presented on Fig. 1B. The ONT (filtered) fastq reads were aligned to the Sscrofa11.1 reference assembly (GCF_000003025.6) using minimap2 with the map_ont flag, resulting in alignments corresponding to a coverage ranging from 30 to 60X of the genome. Variants were called using the PEPPER-Margin-DeepVariant pipeline [34] release 0.4 with the ont flag. This model

was trained for the guppy 4 basecaller, corresponding to the basecaller of the produced ONT reads (MinKNOW-Live-Basecalling version 4.0.11 (flipflop)). Only autosomes were considered. The resulting gvcf files were merged using GLnexus [35], resulting in a single vcf file for the 14 animals containing 27,569,029 variants. In order to fill missing genotypes from genomic ONT variant calling, data were imputed using Beagle 4.0 [36]. The resulting vcf file, called Dataset1_DNAseq_variants_final.vcf.gz, was deposited in the Recherche Data Gouv public database [37].

For RNAseq analyses from *longissimus dorsi* (muscle) and hypothalamus, transcriptome assembly and quantification were performed using the Nextflow nf-core/rna-seq (version nfcore-Nextflow RNA-seq 3.13.2) pipeline. Mapping was performed with STAR software against the Sus scrofa genome reference version 11.1 (Sus_scrofa.Sscrofa11.1.dna.toplevel.fa) and the gene structure annotation version 11.1.109 (Sus_scrofa.Sscrofa11.1.109.gtf). Variant calling from RNAseq was performed following GATK guidelines, leading to a single vcf file for each tissue for the 9 animals, including 2,204,937 and 3,572,855 variants in *longissimus dorsi* and hypothalamus, respectively. These vcf files, called Dataset1_RNAseq_muscle_variants_final.vcf.gz and Dataset1_RNAseq_hypothalamus_variants_final.vcf.gz, have been deposited in the Recherche Data Gouv public database [37].

For the dataset 2, we used data from placenta (fetal tissue) and endometrium (maternal tissue) to reconstruct the parental origin of alleles using imputation and phasing. For RNAseq analyses from placenta and endometrium samples, transcriptome assembly and quantification were performed using the Nextflow nf-core/rnaseq (version nfcore-Nextflow-v20.11.0-edge) pipeline. Mapping was performed with STAR software against the Sus scrofa genome reference version 11.1 (Sus_scrofa.Sscrofa11.1.dna.toplevel.fa) and the gene structure annotation version 11.1.104 (Sus_scrofa.Sscrofa11.1.104.gtf). Associated bioinformatic analyses have been described in a data paper [31]. The RNAseq fastq files have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB75252.

Variant calling from RNAseq dataset 2 was performed using a fork of the Nextflow nf-core RNAVAR code from GitHub (dev version, release 104). Two DSL2 modules were integrated into the Nextflow nf-core/RNAVAR pipeline and the haplotype code was modified to add a gVCF generation step. VCF indexes for dbSNP and indels were built using tabix-0.2.5 and bcftools-1.14. The STAR indexes (v2.7.5a) and the fasta indexes were built using samtools (v1.19) and GATK CreateSequenceDictionary (v4.4.0.0), based on the references Sus_scrofa.Sscrofa11.1.104.gtf, sus_scrofa.vcf (release-104), and Sus_scrofa.Sscrofa11.1.dna.toplevel.fa. Nextflow nf-core/

RNAVAR pipeline was performed using default filtering criteria and generated 54 gVCF files from the 54 placental samples and 28 gVCF files from the 28 endometrium samples. Last, one single vcf file per tissue (placental and endometrium) was generated using GATK GenomicsDBImport (v4.2.6.1). In total, 792,902 and 798,812 variants were called from placenta and endometrium RNAseq data, respectively. The vcf file specific of the placenta called `Dataset2_RNAseq_placenta_variants_final.vcf.gz` has been deposited in the Recherche Data Gouv public database [37].

Imputation and reconstruction of genomes for dataset 2

Since only limited whole genome sequencing was performed for the dataset 2, we reconstructed genomes of both parents and offspring based on endometrium and placenta RNAseq data, which represent mother and offspring, respectively, as well as a dataset of additional genomes ($n=84$) that are not publically available. We used the whole dataset presented in [28, 31], which includes 256 animals comprising RNAseq from the placenta of 224 fetuses and RNAseq from the maternal endometrium of 28 sows. Among the 224 fetuses, 112 were sampled at day 90 of gestation and 112 were sampled at day 110 of gestation. The whole genome sequencing data of 4 sires from dataset 2 were available, resulting in 194 trios and 30 duos for assessing parental origin of alleles in the 224 offspring. First, using bcftools 1.17 [38], only heterozygous biallelic SNPs were kept from both vcf files of the 112 placental fetuses at day 110 of gestation and 28 maternal endometrium. Second, we checked that sibling fetuses had the same maternal genome from endometrium RNAseq variant calling using the identity by state (IBS) parameter of PLINK 1.9 [39]. Since the experimental design involved crosses between the Large White and Meishan breeds, we used a dataset of additional genomes ($n=84$) that are not publically available including Large White ($n=75$) and Meishan ($n=9$) genomes to impute from RNAseq variant calls to whole genome variant calls, using SHAPEIT5 [40]. Only autosomes were considered. Finally, we checked and filtered out Mendelian incompatibilities between mother and fetus using PLINK 1.9 [39]. In total, we reconstructed genomic data for 31,311,840 SNPs after pruning, imputation, and phasing. This vcf file, called `Dataset2_DNAseq_variants_final.vcf.gz`, has been deposited in the Recherche Data Gouv public database [37].

Annotation of variants

Annotation of significant variants was performed using the Variant Effect Predictor (VEP) toolset [41] based on the Ensembl annotation release 114 and the NCBI annotation release 106 databases, since they often

show differences in annotation, especially for 3' and 5' UTR annotation and identification of novel genes.

Detection of parental variants

In contrast to congenic lines, pig breeds are outbred populations, meaning that most variants are polymorphic in both breeds, although some other variants are breed-specific. To take advantage of all genetic variability and exploit parental origin of all variants, we cannot rely on the reference haplotypes from founders or on classical phasing methods [36, 40], which phase chromosome independently, although parents transmit one copy of each chromosome to their offspring. Thus, we used the following dedicated strategy to select only informative parental variants. For each variant, only heterozygous offspring whose parental origin of the two alleles could be traced were retained. To identify such variants, each nuclear family (trios and duos) was evaluated independently and filtering was performed using the SnpSift tool [42]. Thus, for each individual from a trio, three combinations enable us to determine the paternal origin, and three other combinations enable us to determine the maternal origin (Fig. 1B). For duos, only one possible combination per parental origin was used. Using this approach, the combinations were first grouped by parental origin and by individual using bcftools 1.17 [38], resulting in two global parental files for each descendant. In order to obtain a list of positions that could be used to detect specific parental expression in hypothalamus, muscle, and placenta, each genomic parental file, obtained either from ONT (hypothalamus and muscle) or reconstructed (placenta), was then cross-referenced with its associated variant file that was generated from the RNAseq experiments. This resulted in a set of files with common positions between the ONT and RNA-seq data for each parental origin and each individual. These output files were then merged via bcftools 1.17 [38] to generate a global file containing, for each descendant of each cross, all positions that were exploitable for the detection of POE and AGE. The parental alleles of the individuals' genotypes were recoded into paternal origin (Pat) and maternal origin (Mat) alleles using a customized perl script. Consequently, for each variant, two nomenclatures existed, one with alleles coded as a1 and a2 to detect AGE and one coded as Pat and Mat to detect POE.

Statistical analyses

All statistical analyses were performed with R v4.3.1

Relative normalisation of RNAseq raw read counts

Prior to statistical analyses, RNA-seq raw read counts for each SNP and each sample were normalised using relative size factors to account for differences in sequencing

Table 1 Contingency table to test AGE

	a1	a2
Observed	Tot(a1)	Tot(a2)
Expected	$\frac{\text{Tot(a1)} + \text{Tot(a2)}}{2}$	$\frac{\text{Tot(a1)} + \text{Tot(a2)}}{2}$

Table 2 Contingency table to test POE

	Pat	Mat
Observed	Tot(Pat)	Tot(Mat)
Expected	$\frac{\text{Tot(Pat)} + \text{Tot(Mat)}}{2}$	$\frac{\text{Tot(Pat)} + \text{Tot(Mat)}}{2}$

depth. For each sample i , the raw library size, L_i , was calculated as the total number of mapped reads after alignment. A relative size factor, SF_i , was then computed by dividing L_i by the median library size across all samples. Normalised read counts were obtained by dividing the raw read counts (C_{ij} , for SNP j in sample i) by the corresponding size factor, SF_i .

This approach preserves the relative proportions of allele-specific read counts within each sample, while reducing inter-sample variation due to differences in sequencing depth.

Organization of normalised read counts per SNP, depending on both parental and allelic origins

For each variant, only heterozygous offspring for which parental origin was traceable for both alleles were retained. This means that the number of individuals used at each position varied and that the expected read counts for biallelic expression were 50:50 since all analysed individuals were heterozygous. For each SNP with two alleles (a1 and a2), normalised read counts were organized per individual into four categories: Pa1 corresponding to paternal allele 1 reads, Pa2 corresponding to paternal allele 2 reads, Ma1 corresponding to maternal allele 1 reads, and Ma2 corresponding to maternal allele 2 reads.

Total normalised read counts per SNP across all individuals were computed as:

$$\text{Tot(a1)} = \sum_{i=1}^n (\text{Pa1}_i + \text{Ma1}_i) \text{ and}$$

$$\text{Tot(a2)} = \sum_{i=1}^n (\text{Pa2}_i + \text{Ma2}_i) \text{ for allelic origin, and}$$

$$\text{Tot(Pat)} = \sum_{i=1}^n (\text{Pa1}_i + \text{Pa2}_i) \text{ and}$$

$$\text{Tot(Mat)} = \sum_{i=1}^n (\text{Ma1}_i + \text{Ma2}_i) \text{ for parental origin.}$$

Table 3 Contingency table to test POE > AGE

	AGE	POE
Minimum	Min(Tot(a1), Tot(a2))	Min(Tot(Pat), Tot(Mat))
Maximum	Max(Tot(a1), Tot(a2))	Max(Tot(Pat), Tot(Mat))

Table 4 Contingency table to test AGE > POE

	POE	AGE
Minimum	Min(Tot(Pat), Tot(Mat))	Min(Tot(a1), Tot(a2))
Maximum	Max(Tot(Pat), Tot(Mat))	Max(Tot(a1), Tot(a2))

Fisher's exact tests to discriminate allelic imbalance from parental origin imbalance

A two-sided Fisher's exact test, the AGE test, was performed to detect allelic imbalance by comparing observed normalised read counts per allele (a1 and a2) against expected counts under a null hypothesis of the equal (50:50) allelic contribution that is typical of biallelic expression (Table 1).

A two-sided Fisher's exact test, the POE test, was performed to detect parental origin imbalance by comparing observed normalised read counts per parental origin (Pat and Mat) against expected counts under a null hypothesis of the equal (50:50) allelic contribution that is typical of biallelic expression (Table 2).

A one-sided Fisher's exact test using the greater than option, the POE > AGE test, was performed to detect whether the expression imbalance was more important for POE (parental origin imbalance) than AGE (allelic imbalance). This was performed by comparing both minimums and maximums of the distribution of normalised read counts, given their classification by parental origin (POE) or allelic origin (AGE) (Table 3).

A one-sided Fisher's exact test using the greater than option, the AGE > POE test, was performed to detect whether the expression imbalance was more important for AGE (allelic imbalance) than POE (parental origin imbalance). This was performed by comparing both minimums and maximums of the distribution of normalised read counts given their classification by allelic origin (AGE) or parental origin (POE) (Table 4).

A Bonferroni-corrected p-value threshold of 0.01 divided by the number of tests was applied. To determine unique parent-of-origin expressed genes, we used only

significant variants that were annotated. For genes that were tagged by several variants, the mean of all maternal expressed alleles was evaluated.

Fisher’s exact tests to detect a potential sex effect in placenta
To assess whether the parental expression imbalance could be cofounded with a differential expression level between males and females, the two-sided Fisher’s exact tests AGE and POE were performed within groups of males and females. This analysis was possible only in placenta using the dataset 2, which included 27 males and 27 females.

An adaptation of a one-sided Fisher’s exact test using the greater than option was computed as (Tables 5 and 6):

Molecular analyses of the IGF2 locus

Allele-specific PCR for parental informativity

A PCR allele competitive extension (PACE) [43] analysis was performed with 10 ng of purified blood genomic DNA, using the PACE®-IR 2×Genotyping Master mix (3CR Bio-science) and 12 microM of a mix of extended allele specific forward primers and 30 microM of common reverse primer in a final volume of 5 µL. The touch-down PCR amplification condition was 15 min at 94 °C for the hot-start (eurobio SCIENTIFIC, reference PB10.11–05) activation, 10 cycles of 20 s at 94 °C, 58–66 °C for 30 s (dropping 0.8 °C per cycle), then 40 cycles of 20 s at 94 °C and 30 s at 58 °C performed on an ABI9700 thermocycler, followed by a final point read of the fluorescence on an ABI QuantStudio 6 real-time PCR system, and using the QuantStudio software 1.3 (Applied Biosystems). The trio of primers used for genotyping the variant rs3473951016 were GAAGGTGACCAAGTTCA TGCTCCAGGAGCAGAGTGCGCAC, GAAGGTGCGG AGTCAACGGATTATCCAGGAGCAGAGTGCGCAA, and TCCCTGCCCAGCCCCCACT.

Bisulfite conversion of genomic DNA extracted from hypothalamus and muscle

500 ng of hypothalamus and muscle genomic DNA sample collected from 10 piglets were bisulfite converted using an EZ DNA methylation-direct kit (Zymo Research, cat D5020).

PCR conditions before Sanger sequencing

We used Sanger sequencing (i) to validate the accuracy of the PACE genotyped rs3473951016 variant from blood genomic DNA samples and (ii) for epigenotyping rs3473951016 variant on sodium bisulfite-treated genomic DNA obtained from hypothalamus and muscle samples. PCR were performed using the PCR eurobio 5X kit (eurobio SCIENTIFIC, reference PB10.11-05) in the presence of primers at 10 µM, 0.05 U of Taq polymerase

Table 5 Contingency table to test a parental expression imbalance only in females

	Males	Females
Minimum	Min(Tot(Pat), Tot(Mat))	Min(Tot(Pat), Tot(Mat))
Maximum	Max(Tot(Pat), Tot(Mat))	Max(Tot(Pat), Tot(Mat))

Table 6 Contingency table to test a parental expression imbalance only in males

	Females	Males
Minimum	Min(Tot(Pat), Tot(Mat))	Min(Tot(Pat), Tot(Mat))
Maximum	Max(Tot(Pat), Tot(Mat))	Max(Tot(Pat), Tot(Mat))

and the corresponding amount of genomic DNA. The genotyping step required an amount of 10 ng of blood genomic DNA. The epigenotyping step required 25 ng of bisulfite converted hypothalamus and muscle genomic DNA targeting the IGF2 locus. The pair of primers used to genotype the variant rs3473951016 was CCGGGCTT TTTCTAACAGG and CCGTCTGACTAGCTGGTGAA T (genotyping step). The pair of primers used to detect the methylated allele and perform Methylation-sensitive PCR (MS-PCR) was AAAGAGTTTCGTTTTTTTAG GGCGG and AAACCCTAACCCACACCCCTTAC G (epigenotyping step). The PCR cycling conditions on a verity PCR machine (Applied Biosystems) were 5 min at 94 °C for the PCR BIO Taq DNA Polymerase Classic activation, then 30 cycles (genotyping) and 35 cycles (epigenotyping) of 30 s at 94 °C, primer annealing of 30 s at 60 °C, and 30 s of primer extension at 72 °C. The PCR products were then purified using the “Thermosensitive Alkaline Phosphatase” (F0654, Thermo-Scientific) a 0.06 U/microL final, and the “Exonuclease I” (MO293L, NEB) 1.25 U/µM. The reaction was performed at 37 °C for 45 min, followed by an inactivation step at 80 °C during 30 min. The sequencing reaction has been performed by Genewiz company (<https://www.genewiz.com/en-GB/>). The results were analysed with the CLC software.

Results

An extensive strategy to detect parental expression imbalance

Genetic variability between parental genomes is a prerequisite for the characterization of the transcriptome with allele-specific resolution and the systematic identification of putative POE genes. In order to detect POE, we relied on two datasets generated from reciprocal crosses between two genetically distant pig breeds, the European Large White (LW) breed and the Asian Meishan (MS) breed (Fig. 1B). In contrast to congenic lines, the majority of variants are polymorphic in both these parental breeds, although some variants are breed-specific. Piglets from the first dataset were produced to identify POE in muscle and hypothalamus at day 1 and the second dataset targeted detection of POE in the placenta at day 110 of

A

	rs3472903225 7:121663456	rs3472187167 12:24196236	rs700461902 16:47995339																																				
AGE Fisher test Two-sided	<table><tr><td></td><td>a1</td><td>a2</td></tr><tr><td>Obs</td><td>321.71</td><td>439.90</td></tr><tr><td>Exp</td><td>380.80</td><td>380.80</td></tr><tr><td colspan="3">P-value adj = 1</td></tr></table>		a1	a2	Obs	321.71	439.90	Exp	380.80	380.80	P-value adj = 1			<table><tr><td></td><td>a1</td><td>a2</td></tr><tr><td>Obs</td><td>1378</td><td>5.99</td></tr><tr><td>Exp</td><td>691.99</td><td>691.99</td></tr><tr><td colspan="3">P-value adj = 4.62e-241</td></tr></table>		a1	a2	Obs	1378	5.99	Exp	691.99	691.99	P-value adj = 4.62e-241			<table><tr><td></td><td>a1</td><td>a2</td></tr><tr><td>Obs</td><td>405.99</td><td>712.38</td></tr><tr><td>Exp</td><td>559.18</td><td>559.18</td></tr><tr><td colspan="3">P-value adj = 2.58e-05</td></tr></table>		a1	a2	Obs	405.99	712.38	Exp	559.18	559.18	P-value adj = 2.58e-05		
		a1	a2																																				
	Obs	321.71	439.90																																				
	Exp	380.80	380.80																																				
P-value adj = 1																																							
	a1	a2																																					
Obs	1378	5.99																																					
Exp	691.99	691.99																																					
P-value adj = 4.62e-241																																							
	a1	a2																																					
Obs	405.99	712.38																																					
Exp	559.18	559.18																																					
P-value adj = 2.58e-05																																							
POE Fisher test Two-sided	<table><tr><td></td><td>Pat</td><td>Mat</td></tr><tr><td>Obs</td><td>16.34</td><td>745.27</td></tr><tr><td>Exp</td><td>380.80</td><td>380.80</td></tr><tr><td colspan="3">P-value adj = 2.44e-112</td></tr></table>		Pat	Mat	Obs	16.34	745.27	Exp	380.80	380.80	P-value adj = 2.44e-112			<table><tr><td></td><td>Pat</td><td>Mat</td></tr><tr><td>Obs</td><td>563.92</td><td>820.07</td></tr><tr><td>Exp</td><td>691.99</td><td>691.99</td></tr><tr><td colspan="3">P-value adj = 0.39</td></tr></table>		Pat	Mat	Obs	563.92	820.07	Exp	691.99	691.99	P-value adj = 0.39			<table><tr><td></td><td>Pat</td><td>Mat</td></tr><tr><td>Obs</td><td>519.29</td><td>599.07</td></tr><tr><td>Exp</td><td>559.19</td><td>559.18</td></tr><tr><td colspan="3">P-value adj = 1</td></tr></table>		Pat	Mat	Obs	519.29	599.07	Exp	559.19	559.18	P-value adj = 1		
		Pat	Mat																																				
	Obs	16.34	745.27																																				
	Exp	380.80	380.80																																				
P-value adj = 2.44e-112																																							
	Pat	Mat																																					
Obs	563.92	820.07																																					
Exp	691.99	691.99																																					
P-value adj = 0.39																																							
	Pat	Mat																																					
Obs	519.29	599.07																																					
Exp	559.19	559.18																																					
P-value adj = 1																																							
POE > AGE Fisher test One-sided, greater	<table><tr><td></td><td>AGE</td><td>POE</td></tr><tr><td>Min</td><td>321.71</td><td>16.34</td></tr><tr><td>Max</td><td>439.90</td><td>745.27</td></tr><tr><td colspan="3">P-value adj = 3.87e-87</td></tr></table>		AGE	POE	Min	321.71	16.34	Max	439.90	745.27	P-value adj = 3.87e-87			<table><tr><td></td><td>AGE</td><td>POE</td></tr><tr><td>Min</td><td>5.99</td><td>563.92</td></tr><tr><td>Max</td><td>1378</td><td>820.07</td></tr><tr><td colspan="3">P-value adj = 1</td></tr></table>		AGE	POE	Min	5.99	563.92	Max	1378	820.07	P-value adj = 1			<table><tr><td></td><td>AGE</td><td>POE</td></tr><tr><td>Min</td><td>405.99</td><td>519.29</td></tr><tr><td>Max</td><td>712.38</td><td>599.07</td></tr><tr><td colspan="3">P-value adj = 1</td></tr></table>		AGE	POE	Min	405.99	519.29	Max	712.38	599.07	P-value adj = 1		
		AGE	POE																																				
	Min	321.71	16.34																																				
	Max	439.90	745.27																																				
P-value adj = 3.87e-87																																							
	AGE	POE																																					
Min	5.99	563.92																																					
Max	1378	820.07																																					
P-value adj = 1																																							
	AGE	POE																																					
Min	405.99	519.29																																					
Max	712.38	599.07																																					
P-value adj = 1																																							
AGE > POE Fisher test One-sided, greater	<table><tr><td></td><td>POE</td><td>AGE</td></tr><tr><td>Min</td><td>16.34</td><td>321.71</td></tr><tr><td>Max</td><td>745.27</td><td>439.90</td></tr><tr><td colspan="3">P-value adj = 1</td></tr></table>		POE	AGE	Min	16.34	321.71	Max	745.27	439.90	P-value adj = 1			<table><tr><td></td><td>POE</td><td>AGE</td></tr><tr><td>Min</td><td>563.92</td><td>5.99</td></tr><tr><td>Max</td><td>820.07</td><td>1378</td></tr><tr><td colspan="3">P-value adj = 3.59e-184</td></tr></table>		POE	AGE	Min	563.92	5.99	Max	820.07	1378	P-value adj = 3.59e-184			<table><tr><td></td><td>POE</td><td>AGE</td></tr><tr><td>Min</td><td>519.29</td><td>405.99</td></tr><tr><td>Max</td><td>599.07</td><td>712.38</td></tr><tr><td colspan="3">P-value adj = 0.24</td></tr></table>		POE	AGE	Min	519.29	405.99	Max	599.07	712.38	P-value adj = 0.24		
		POE	AGE																																				
	Min	16.34	321.71																																				
	Max	745.27	439.90																																				
P-value adj = 1																																							
	POE	AGE																																					
Min	563.92	5.99																																					
Max	820.07	1378																																					
P-value adj = 3.59e-184																																							
	POE	AGE																																					
Min	519.29	405.99																																					
Max	599.07	712.38																																					
P-value adj = 0.24																																							

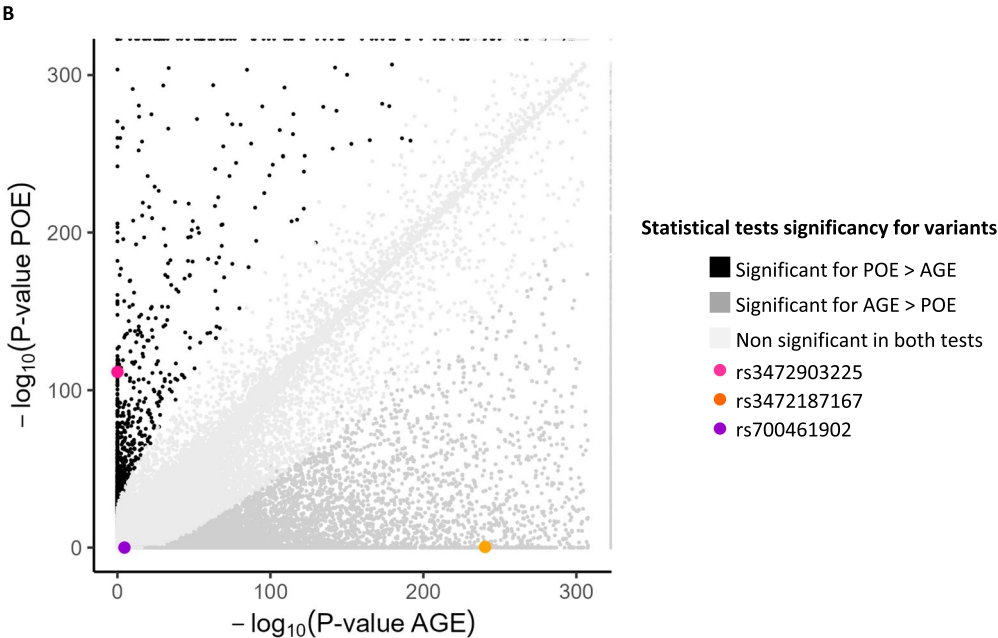


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Discrimination between parental and allelic origin. **a** Examples of variants that are significant either for parental imbalance test (rs3472903225), allelic imbalance (rs3472187167) or non significant (rs700461902). The 4 Fisher's exact tests are shown with real data extracted from normalized read count and structured either on allelic origin (a1 and a2) or on parental origin (Pat and Mat). **b** Scatter plot of statistical analyses for all informative variants in the three tissues (hypothalamus, muscle and placenta). For each variant, the values of $-\log_{10}(\text{Pval AGE})$ and $-\log_{10}(\text{Pval POE})$ from Fisher's exact tests have been plotted and the direction of the effect (AGE or POE) has been colored in black for the POE > AGE significance, in dark grey for the AGE > POE Fisher test significance and in light grey when it was not possible to discriminate between both directions. Only variants in black showing parental expression imbalance have been conserved for further analyses. The 3 variants shown in (a) are plotted on the scatter plot

gestation. We developed a strategy to identify POE genes, exploring all genetic variability, rather than restricting the analysis to breed-specific variants, and considering all variants regardless of their initial annotation in the pig reference genome (Fig. 1B).

Calling of variants followed by imputation allowed the detection of 27,569,028 and 31,311,840 variants in the first and second dataset, respectively. The very large number of variants in both datasets may be explained by (i) the mapping of reads on the Sscrofa11.1 reference genome, which is from a Duroc pig and (ii) the high genetic diversity between the Large White and Meishan parental breeds. A total of 1,490,388 and 566,278 variants in, respectively, datasets 1 and 2 were Duroc-specific and had only alternative alleles in our data and are, therefore, not useful for further analyses. Filtering of variants was performed to keep all informative heterozygous variants in offspring, those that were fixed in each parental breed, as usually done (respectively 3,047,025 and 8,085 datasets 1 and 2), and those for which parental origin was traceable (i.e. one of the two parent was homozygous). The six possible parental genomic combinations (Fig. 1B) enabled us to identify 24,024,152 informative variants in dataset 1 and 13,025,625 in dataset 2 (Additional file 1 Figs. S1 and S2).

Calling of variants from transcriptomic data enabled us to identify 3,572,855, 2,204,937, and 792,903 variants in, respectively, the hypothalamus, muscle, and placenta RNAseq data. The lists of variants from the genomic and transcriptomic data were intersected per tissue to obtain positions for which parental expression imbalance could be exploited (Fig. 1B and Additional file 1 Fig. S2). From the intersection, we only retained bi-allelic genomic positions that consisted of at least one individual per crossing direction, and for which the total number of reads was at least 15 in the hypothalamus and muscle, and at least 90 in the placenta, due to the 6 times greater number of F1s compared with the other two datasets. A total of 539,555, 322,599, and 191,901 variants were considered for further analyses for hypothalamus, muscle, and placenta, respectively (Additional file 1 Fig. S2).

Different statistical analyses were performed to discriminate POE from AGE (Fig. 2A). We first assessed POE and AGE expression using two-sided Fisher's exact tests at each genomic position in the three tissues. We also assessed which of the two mono-allelic expression models (POE: POE > AGE or AGE: AGE > POE) was most

probable, using one-sided Fisher's tests (Fig. 2A). From the 539,555, 322,599, and 191,901 variants analyzed in, respectively, the hypothalamus, muscle, and placenta, bi-allelic expression was concluded for a large proportion, as neither AGE nor POE could be significantly determined for 531,357, 317,735, and 186,847 variants in these respective tissues. We determined AGE to be significantly (0.01 Bonferroni corrected p-value) more probable than POE (AGE > POE, Fisher test) for 8,160, 3,387, and 4,567 variants in these respective tissues. Finally, we identified 510, 348, and 469 variants (0.01 Bonferroni corrected p-value) in these respective tissues for which POE was significantly more likely than AGE (POE > AGE, Fisher test) (Fig. 2B).

Identification of parent-of-origin expressed genes, from weak imbalance to imprinting

Our strategy to identify parental expression imbalance in a genome-wide manner relied on an approach in which all variants were initially considered since no filters were applied on either selection of variants from their original breed or on the annotation of variants, in order to maximize the possibility of identifying novel candidate genes for imprinting. We analyzed both the genomic distribution and the parental origin of the 510, 348 and 469 variants that were detected to show POE in respectively, the hypothalamus, muscle, and placenta. While some variants were grouped in clusters, others were more widespread, suggesting that one gene could be either targeted by several variants or a single variant (Additional file 1 Fig. S3). Annotation of the 510, 348, and 469 variants that show significant POE using the VEP toolset [41] identified 185, 104, and 213 genes in the hypothalamus, muscle, and placenta, respectively (Additional file 2 Table S1), including isolated unannotated variants. In total, this represented at least 450 genes with a significant parental expression imbalance. Several highly significant variants were identified in well-known imprinted genes such as *IGF2* (*Insulin-like Growth Factor 2*), *MEG3* (*Maternally Expressed Gene 3*), and *MEST* (*Mesoderm Specific Transcript*), confirming the validity of our approach (Additional file 2 Table S1).

The power of our statistical analyses of parental imbalance ratios enabled identification of very weak parental imbalance (46:54). Variants showing weak to moderate parental expression imbalance (ratios ranging from 46:54 to 25:75) were not considered further. This resulted in

260, 200 and 91 variants for further analyses for the hypothalamus, muscle, and placenta, respectively (Additional file 3 Table S2). To assess whether potential false-positive calls of parental expression imbalance might be explained by different expression levels between sexes, we performed independent analyses in males ($n=27$) and females ($n=27$) in dataset 2, since dataset 1 includes only nine offspring. Significant differences between males and females were identified for 23 out of 91 variants but the pattern of imprinting was not affected by sex, suggesting that the variants in the three tissues that were significant for POE was not biased (Additional file 1 Fig. S4 and Additional file 3 Table S2). Of the 260, 200 and 91 variants in, respectively, the hypothalamus, muscle and placenta, a total 141 genes were annotated, including 51 with an exclusive parental expression imbalance (ratio greater than 5:95) (Additional file 3 Table S2). Among all genes, 12 did not correspond to a HUGO Gene Nomenclature, established from a committee of the Human Genome Organisation (HUGO), but were annotated to genes with uncertain functions (LOC symbols) from NCBI annotation release 106. A total of 70, 48, and 51 genes showed parental expression imbalance in hypothalamus, muscle, and placenta, respectively. While more genes showed preferential expression of the paternal ($n=37$) versus the maternal ($n=14$) allele in the placenta, paternal versus maternal allele expression was more balanced in the hypothalamus (40 versus 30) and muscle (27 versus 21) (Fig. 3A).

Analyses of the distribution of reads for the known imprinted genes *IGF2*, *MEG3*, and *MEST* clearly highlighted exclusive expression from the paternal allele for *MEST* and from the maternal allele for *MEG3* (Fig. 3B). The conclusion of full paternal expression seemed as sharp for *IGF2* in muscle and placenta but a switch to maternal expression was observed in the hypothalamus. Unfortunately, we could not determine whether a similar switch in parental expression was also present for the *H19* gene, which is regulated by the Imprinting Control Region 1 (ICR1), because there were no informative variants in the *H19* gene. We addressed whether these differences in *IGF2* expression patterns between tissues could be caused by a difference in parental origin of methylation within ICR1. We molecularly characterised the parental origin of ICR1 methylation using the rs3473951016 variant in this region that was informative in dataset 1. Based on chromatograms, we observed the presence of the paternal allele in both tissues for each cross. Thus, the change in the direction of the POE for the *IGF2* gene, which showed maternal expression in the hypothalamus and exclusive paternal expression in muscle tissue, is not associated with a switch in the parental origin of methylation at ICR1, but rather with alternative,

tissue-specific promoters or enhancers (Fig. 3C and Additional file 1 Fig. S5).

Identification of novel imprinted genes shared between tissues and likely livestock-specific

In order to better characterize imprinted genes in piglets and further explore genes showing parental expression imbalance based on statistical tests, we evaluated their distribution between the three tissues. Among the 141 genes identified across tissues, the majority seemed specific to one tissue, with 48, 27 and 44 out of 70, 48, and 51 genes with significant POE for hypothalamus, muscle, and placenta, respectively (Fig. 4A). While 15 genes were shared between hypothalamus and muscle, only 1 (*LOC100520903* also called *ZNF300-like*) was common between hypothalamus and placenta and none was shared between muscle and placenta, suggesting specificity of genomic imprinting mechanisms in hypothalamus versus placenta. We identified 6 genes that showed parental expression imbalance in all three tissues, i.e. *FAM20B*, *IGF2*, *IGF2R*, *MEG3*, *MEST*, and *KBTD6* (Fig. 4A).

Further analyses were conducted on significant POE genes that were shared between at least two tissues ($n=22$). The magnitude of parental expression imbalance and the number of informative offspring were also analysed in greater detail to detect novel imprinted genes (Fig. 4B and Additional file 3 Table S2). Most of these genes were conserved across species, except for *KBTD6*, which is known to be a livestock-specific imprinted gene, as well as *FAM20B* (*Glycosaminoglycan Xylosylkinase*), *HSD17B11* (*Hydroxysteroid 17-beta dehydrogenase 11*), *PDE4DIP* (*Phosphodiesterase 4D Interacting Protein*), *PODXL* (*Podocalyxin Like*), *POU6F2* (*POU Class 6 Homeobox 2*), *PREPL* (*prolyl endopeptidase like*), *RNF34* (*Ring Finger Protein 34*), *SLA-1* (*Swine Leucocyte Antigen 1*), *TK2* (*Thymidine Kinase 2*), *TNKS* (*Tankyrase*) and *LOC100520903* (*ZNF300-like*). In mice, these genes are mainly involved in behavior and growth (Additional file 1 Fig. S6). Focusing on genes for which at least half of offspring were informative for evaluating parental expression imbalance, exclusive paternal expression was observed for *KBTD6* and exclusive maternal expression for *POU6F2*. Although the pattern of imprinting for *FAM20B* and *PDE4DIP* was not restricted to a single parental allele, a high significant imbalance in the direction of the paternal allele was observed (Fig. 4C).

A focused analysis on the unannotated *LOC100520903* gene, a zinc finger protein 300-like gene, was carried out. The *ZNF300-like* gene is located between *RNF216* (*Ring Finger Protein 2016*) and *OR10AH1* (*Olfactory Receptor Family 10 subfamily AH member 1*) and showed a short and a long isoform, as well as features such as a CpG island that spans the first exon of the longest isoform

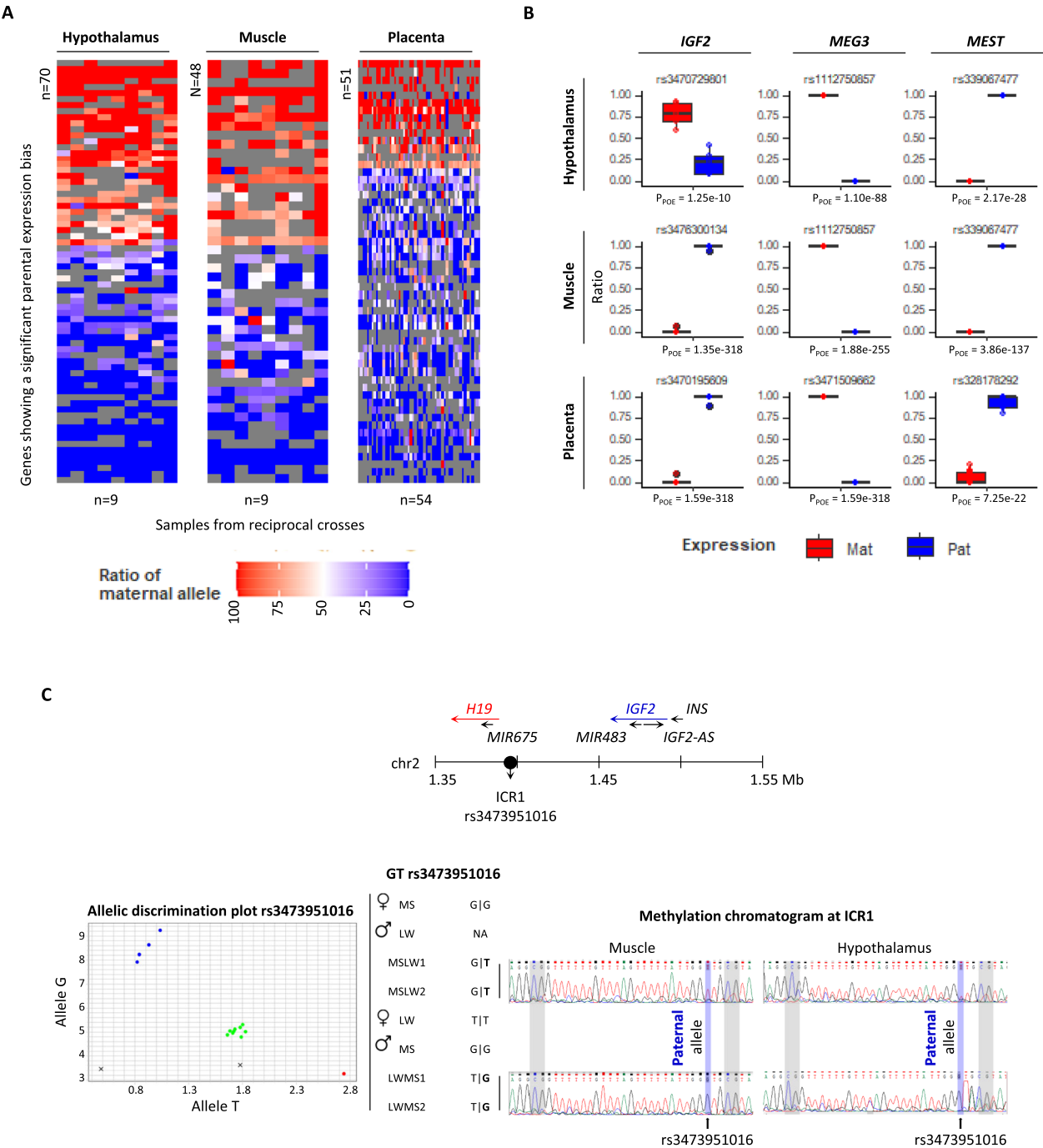


Fig. 3 Genome-wide detection of parental expression imbalance (ratio above 25:75) in piglets. **a** Genes showing parent-of-origin expression (n = 141). Each significant variants for POE > AGE statistical Fisher test has been annotated using both Ensembl annotation release 114 and NCBI annotation release 106. When several variants tagged the same gene, the mean of POE has been considered leading to 70, 48 and 51 genes with parental expression imbalance in hypothalamus, muscle and placenta, respectively. The ratio of maternal allele has been plotted on the heatmap with blue corresponding to paternal expression and red corresponding to maternal expression. **b** Parental ratio for three imprinted genes in each tissue. For each tissue, one significant variant tagging *IGF2*, *MEG3* and *MEST* genes has been considered and number of parental reads for informative individuals has been plotted relative to its parental origin. Reads from paternal and maternal origins have been colored in blue and red, respectively. **c** Parental methylation evaluation at the Imprinting Control Center 1 (ICR1) of the *IGF2* locus. The variant rs3473951016, located within ICR1, was informative in two nuclear families from reciprocal crossed of dataset 1 with the T allele coming from the boar in the first MSLW cross and the G allele coming from the boar in the second LWMS cross. Chromatograms issued from methylation sensitive PCR followed by Sanger sequencing of PCR products in hypothalamus and muscle showed that only paternal allele is methylated in both tissues

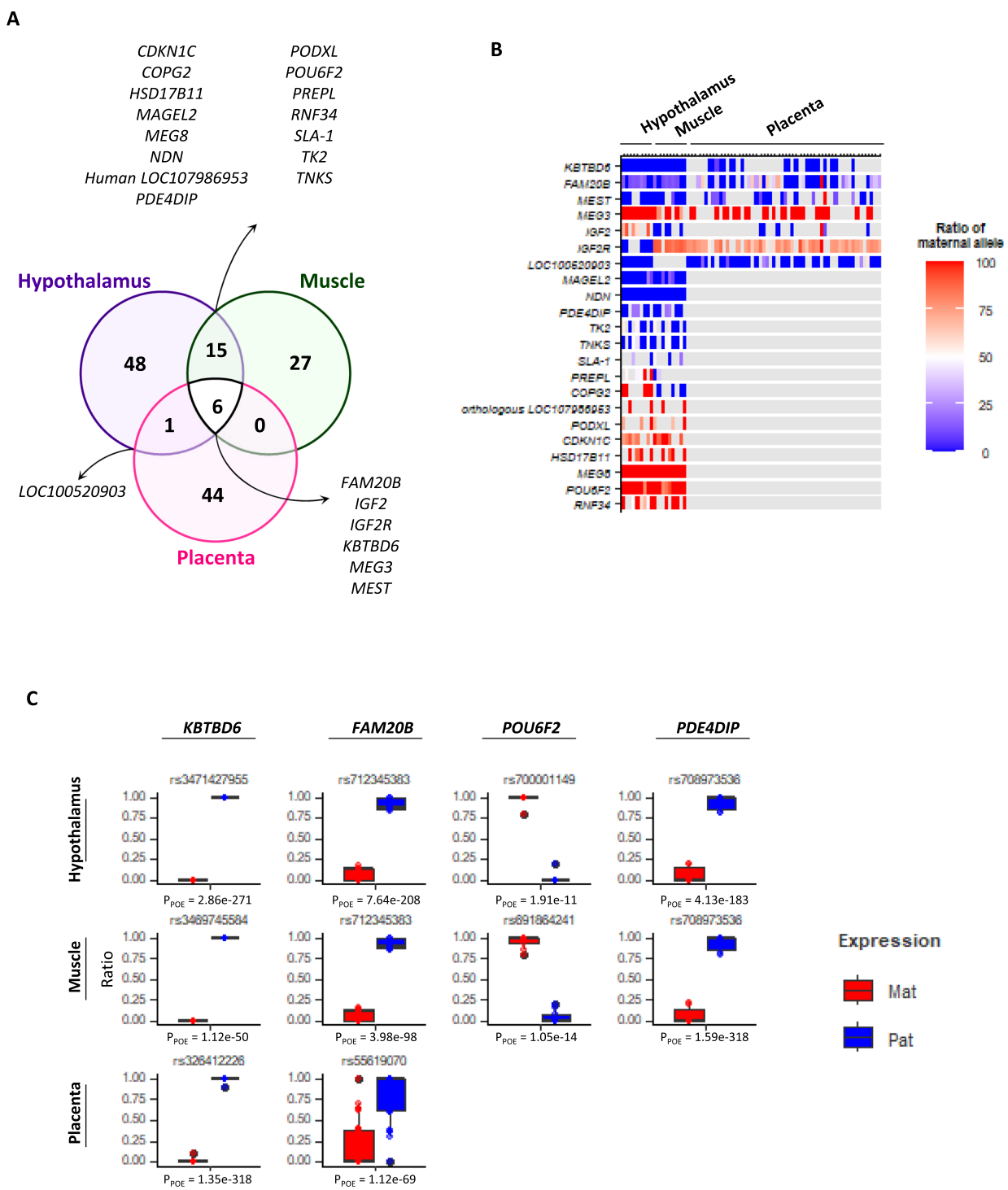


Fig. 4 Conservation of imprinted genes in piglets across tissues. **a** Venn diagram of parent-of origin expressed genes. A venn diagram was build from the 70, 48 and 51 genes with a significant parental expression imbalance in hypothalamus, muscle and placenta, respectively. Most of detected genes seemed tissue-specific. **b** Heatmap of POE focused on genes detected in at least two tissues. The ratio of maternal allele has been plotted on the heatmap with blue corresponding to paternal expression and red corresponding to maternal expression. **c** Parental expression imbalance of novel imprinted genes. Among genes showing POE in at least two tissues, genes with a strong or exclusive parental imbalance (above 25: 75) and unknown so far for imprinting were selected. For each tissue, one significant variant tagging *KBTBD6*, *FAM20B*, *POU6F2* and *PDE4DIP* genes has been considered and number of parental reads for informative individuals has been plotted. Reads from paternal and maternal origins have been colored in blue and red, respectively.

and repeat elements such as LTR (Fig. 5A). This region seems to have undergone different evolutionary forces based on the dot plot between several species in Fig. 5B, which demonstrates the acquisition of complexity in orthologous regions across speciation through an expansion of highly repeated elements of Type I transposons family. In piglets, this novel gene exhibits exclusive paternal expression in both the hypothalamus and the placenta

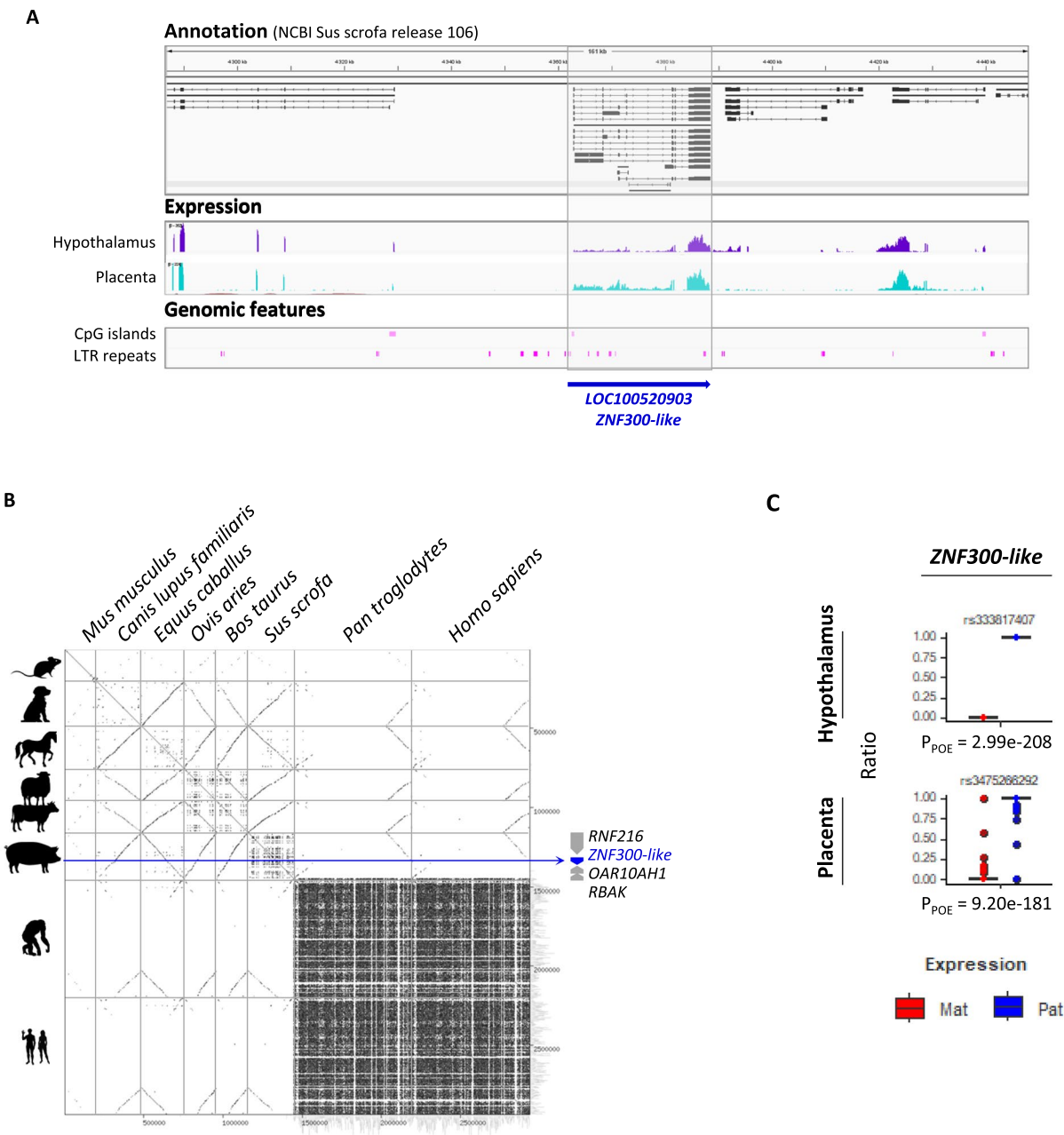
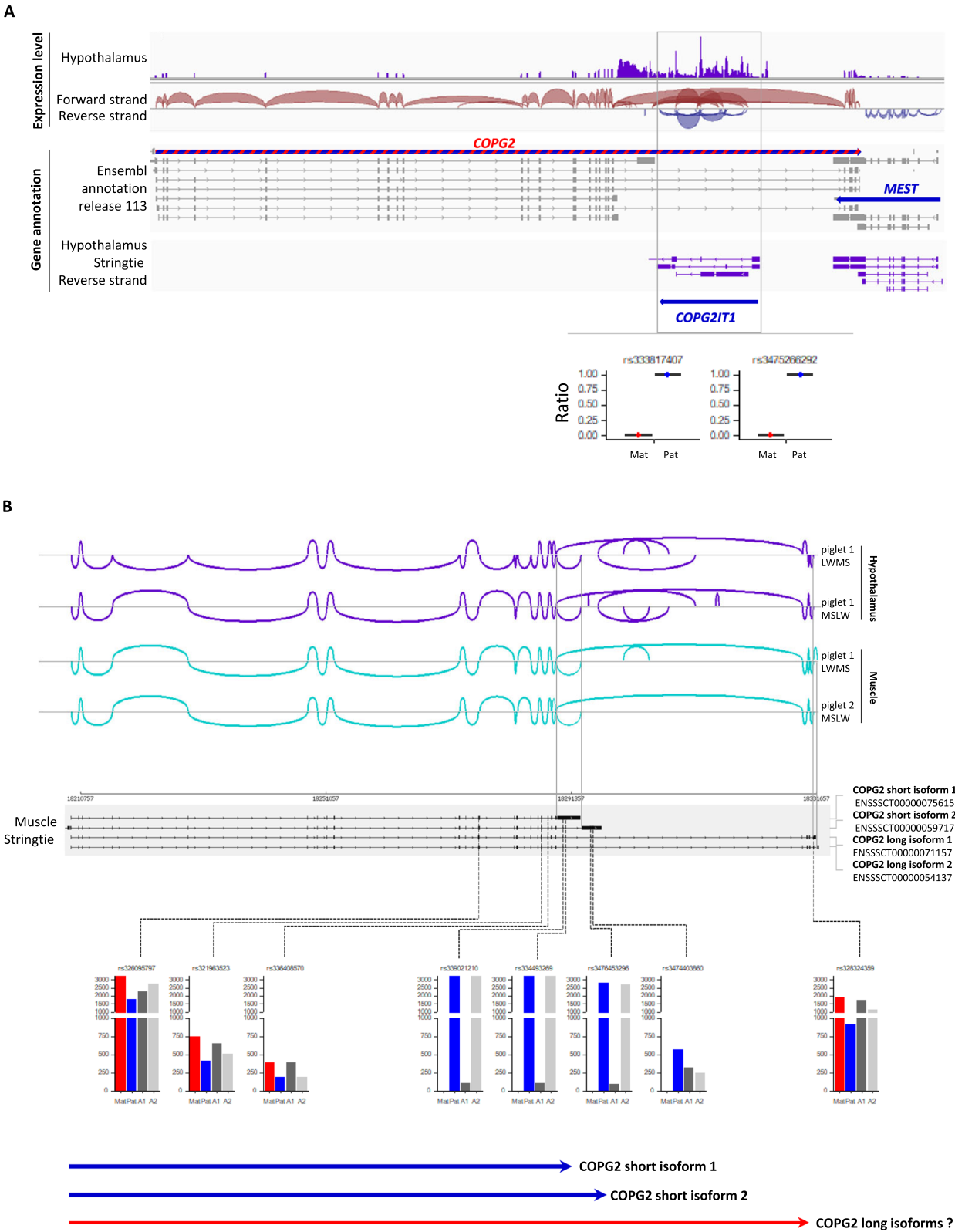


Fig. 5 The paternally expressed *ZNF300-like* gene (*LOC100520903*); a putative lineage-specific imprinted gene. **a** Detection of the *LOC100520903*, a *ZNF300-like* gene, in both hypothalamus and placenta. Screenshot of the *RNF216-RBAK* locus using the Integrative Genome Viewer (IGV [68]). Expression level in hypothalamus in one individual showed coverage from the forward strand of the annotated *LOC100520903* only in the NCBI Sus scrofa release 106. Distribution of CpG island, LTR repeats and chromatin marks regulation along the region using publicly available FAANG dataset. **b** Comparison of sequences of the *RNF216-RBAK* region in several species. Fasta sequences of *RNF216-RBAK* region from mouse (GCA_000001635.9), dog (GCA_011100685.1), horse (GCA_041296265.1), cow (GCA_002263795.4), sheep (GCA_016772045.2), pig (GCA_000003025.6), chimpanzee (GCA_028858775.2) and human (GCA_000001405.29) were aligned and compared across each other. The dot plot showed the evolution of the region through expansion of repeated elements across speciation. **c** Parental expression imbalance of the *ZNF300-like* gene. For hypothalamus and placenta, one significant variant tagging the *LOC100520903* gene has been considered and number of parental reads for informative individuals has been plotted. Reads from paternal and maternal origins have been colored in blue and red, respectively



(See figure on previous page.)

Fig. 6 Detection of *COPG2IT1* and *COPG2* paternally expressed isoforms for the *COPG2* locus. **a** Identification of the paternal *COPG2IT1* gene in hypothalamus. Screenshot of the *COPG2* locus using the Integrative Genome Viewer (IGV [68]). Expression level in hypothalamus in one individual showed high coverage (grey box) from the reverse strand while none gene was located and annotated in the last available Ensembl annotation release 113. Transcript assembly using Stringtie [44] detected an antisense transcript at this location. Two significant variants for POE > ASE statistical Fisher's exact test (rs333817407 and rs3475266292) overlapped this novel transcript and showed an exclusive paternal expressed allele in hypothalamus strongly suggested the identification of *COPG2IT1* gene. **b** Identification of paternal *COPG2* isoforms. The *COPG2* transcripts showed short and long isoforms in both hypothalamus and muscle. Transcript assembly using Stringtie [44] improved 3' UTR of shortest isoforms compared to Ensembl annotation release 113. Analyses of POE of variants spanning *COPG2* gene showed that variants rs339021210, rs3344933269, rs3476453296 and rs3474403860 tagged exclusively last exons of *COPG2* shortest isoforms. Only paternal allele is expressed for all of them, suggesting a paternal expression of shortest isoforms of *COPG2*. None informative and significant variants tagged exclusively longest isoforms of *COPG2*, making their expression patterns unconvincing

(Fig. 5C), although parental expression appeared ambiguous in the placenta for two out of 54 individuals, likely due to their low level of expression.

The *COPG2* locus, detection of the paternal *COPG2IT1* gene and isoform-specific imprinting of *COPG2*

We tried to understand the inconsistencies that we obtained for the *COPG2* gene in piglets since discrepancies were also observed in human and mouse for this gene. Indeed, several variants spanning the *COPG2* gene brought conflicting information on parental expression imbalance, with some variants suggesting bi-allelic expression, some a trend for preferential maternal allele expression, and others exclusive paternal allele expression. A deep visual analysis of reads through the Integrative Genome Viewer (IGV), combined with the identification of isoforms from our transcriptomic data using Stringtie [44], allowed a novel antisense transcript to be detected in the hypothalamus that was not yet annotated in the pig genome (Fig. 6A). This transcript was located between the *COPG2* and *MEST* genes and likely corresponds to the *COPG2IT1* gene, which clearly showed paternal expression based on distribution reads for rs333817407 and rs3475266292 (Fig. 6A).

Besides identification of the *COPG2IT1* non-coding transcript within the *COPG2* locus, visual and manual inspection of the *COPG2* gene itself identified parental expression of specific *COPG2* isoforms (Fig. 6B). Annotation of *COPG2* isoforms based on the Ensembl annotation release 113 and the NCBI annotation release 106 suggested short and long isoforms that share the first exons. Reconstruction of transcripts that were present in hypothalamus and muscle detected at least four distinct isoforms, two short (ENSSSCT00000075615 and ENSSSCT00000059717) and two long isoforms (ENSSSCT00000071157 and ENSSSCT00000054137), as shown on sashimi plots in Fig. 6B. Based on the location of informative variants within the different exons of the *COPG2* gene, we observed a sharp switch between variants that were common to the short and long isoforms (rs326095797, rs321963523, and rs336408570) and variants that were specific to the short isoforms (rs339021210 and rs3344933269 for *COPG2* short isoform 1 and rs3476453296 and rs3474403860 for *COPG2* short

isoform 2) (Fig. 6B). We conclude that both short isoforms were exclusively expressed from the paternal allele, while this was more difficult to conclude for the long isoforms between maternal or bi-allelic expression since some exons were shared with the downstream *MEST* exons.

Discussion

Advantages and limitations of our approach to minimize bias

We provided the results of a large-scale analysis of parental expression imbalance in piglets around birth, with the major objective of minimizing bias, as per latest recommendations [20]. Our strategy was based on reciprocal crosses between two breeds with very high genetic heterogeneity to maximize heterozygosity in offspring and to discriminate between the two classes of mono-allelic expression (AGE and POE). The use of experimentally manipulated embryos, focusing solely on gynogenotes and ignoring androgenotes, does not allow AGE and POE to be distinguished and is less accurate for genome-wide mapping of genes subject to genomic imprinting [25, 26]. The acquisition of genomic data from both parents and offspring allowed us to exploit the all genetic variability by selecting all the heterozygous variants in offspring for which the parental origin could be deduced. Unlike congenic mouse lines, farm animal populations are polymorphic. Although it is possible to select only fixed variants in each breed to mimic the approach used for reciprocal crosses in mice, this is much more restrictive [22, 45–47]. Here, we were able to exploit 24,024,152 parental informative variants in hypothalamus and muscle and 13,025,625 in placenta from genomic data, including 3,047,025 and 8,085 variants that were fixed in Large-White and Meishan breeds for datasets 1 and 2, respectively. Most other studies analyze only breed-specific variants from reciprocal crosses and exploit around five times fewer variants [22].

All previous studies aiming at detecting imprinted genes use variant annotation as an additional filter before statistical analysis in order to consider only annotated genes. This assumes that the genome is perfectly annotated, which is not the case, particularly for non-coding genes, which represent 2/3 of the imprinted genes

identified to date [4]. Their identification in mammalian genomes is difficult and they are often poorly annotated [48]. For example, *KCNQ1OT1* and *NESP-AS*, which are known to regulate imprinting at, respectively, the *CDNIC-KCNQ1* loci and *GNAS* in mammals, do not appear in the porcine genome annotation (Ensembl annotation release 113 and NCBI annotation release 106 databases). Here, we chose to apply an annotation filter a posteriori in order not to restrict the list of significant variants. Several variants did not appear to be located in annotated genes or were poorly annotated, and manual inspection of these variants often revealed a gene. In addition, as non-coding genes are poorly annotated in the current pig genome build, there is a higher risk to incorrectly annotate a protein-coding gene that is adjacent to a not annotated non-coding gene as being imprinted. While we were able to detect the expression of several long non-coding RNAs (*COPG2IT1* and *AIRN*) by visual inspection, *COPG2IT1* and *AIRN* would have been tagged to, respectively, *COPG2* and *IGF2R* instead of as non-coding genes (Additional file 1 Fig. S7). These examples show the importance of genome annotation for studying genomic imprinting in a genome-wide manner. However, our results for the *COPG2* locus demonstrate the complexity and potential for misleading results that can arise from a genome-wide approach.

Although our approach is intended to minimize bias, we still face some limitations to detect an exhaustive list of POE genes. Although we generated whole genome sequencing from ONT for dataset 1, we did not explore the haplotype phasing from such technologies as Li et al. [49] recently performed to take advantage of all genetic variability. The lack of informativity on parental origin still remains the greatest source of incompleteness in the detection of imprinted genes. For example, we were not able to identify the *PEG10* or *H19* genes, which are well-known imprinted genes in pigs, in contrast to Li et al. [49], who used three sets of trio families by crossing divergent pig breeds to generate sufficient heterozygous loci. Furthermore, the threshold that was chosen to identify parental expression imbalance may have impaired the identification of genuine imprinted genes such as *PLAGL1* and *BEGAIN*, which showed parental expression imbalance below the 25:75 ratio. Indeed, some variants that tag *PLAGL1* showed significant parental expression imbalance (p -value = 7.05×10^{-18}) but with weak imbalance in favor of the maternal allele (56:44 ratio) in placenta. In a similar way, some variants that tag *BEGAIN* showed significant parental expression imbalance (p -value = 3.40×10^{-87}) in hypothalamus but with a stronger imbalance in favor of the paternal allele (70:30 ratio). Our stringent criteria using a Bonferroni significance threshold and a minimum imprinting ratio may have resulted in missing the detection of some imprinted genes. In

contrast, false positive calls could also have resulted from the low number of heterozygous animals in both reciprocal crosses, the low number of variants per gene, as well as the level of expression of genes. Thus additional multi-omics experiments using different combinations of breeds [49] will be very useful to obtain a full overview of the genomic imprinting landscape in pigs.

Specificity of genomic imprinting by tissue and developmental stage

The first comparative studies on the presence of imprinted genes across tissues and species focused on a few candidate genes per orthology, particularly in livestock species (as reviewed in [6]), and suggested firstly that these atypical regulatory mechanisms were conserved across species. High-throughput -omics technologies then provided access to a global and more exhaustive overview, leading to more contrasting results, despite the biases inherent in these methods [20]. It was then shown that imprinted expression (i.e. the number of imprinted genes, as well as the strength of the parental imbalance for a given gene) was stronger in embryonic tissues, including hypothalamus and placenta, than in adulthood tissues [4, 45, 50]. More specifically, the hypothalamus has even been characterized to have a higher frequency of imprinted gene expression than other brain regions [10, 51]. Our results on the comparison between tissues were consistent with these conclusions, as, using the most stringent filters, we identified 70, 48, and 51 imprinted genes in the hypothalamus, muscle, and placenta, respectively. In addition, for the 141 genes with strongest POE magnitudes, 86% were genes detected in the hypothalamus and placenta, supporting the notion that genomic imprinting may constitute a conserved mechanism to instruct both neural and placental functions.

Some discrepancies that we observed across tissues concerning patterns of imprinting have already been shown in previous studies, such as the switch from strong maternal allele expression of *IGF2* in the hypothalamus to exclusive paternal expression in the other tissues investigated [4, 51]. The placenta is also the tissue with the most imprinting discrepancies between the mouse and humans [52] although the placentas of these two species are structurally equivalent (discoidal or hemo-chorial placenta). From a functional point of view, the dynamics of gene expression during gestation demonstrates that the mouse placenta looks similar to the human placenta only during the first half of pregnancy [53], suggesting the importance of developmental stage of the placenta. Here, placentas from piglets were sampled around birth. In addition, pigs present an epithelo-chorial placenta [54], unlike humans and mice. We hypothesized that developmental stage and placentation may explain why we observed imprinting patterns that

were not conserved between species, particularly with regard to genes expressed in the placenta such as *CBR1*, *HM13*, and *PLAGL1*. In fact, these genes showed exclusive paternal expression in human and murine placentas [52, 55], while we observed the opposite for pigs, especially for *CBR1*.

Weak conservation of imprinted genes across mammals, in contrast to their imprinting patterns

Our approach revealed a number of genes with strong or even exclusive parental expression imbalance ($n=141$), consistent with studies carried out in humans and mice, for which approximately 200 imprinted have been characterized (as reviewed in [2]). The use of two datasets of very different sizes showed that a larger dataset increases the number of genes with intermediate expression levels *versus* genes with very strong or exclusive parental expression imbalance. Among the 141 genes with strong POE imbalance, 65 and 35% showed paternal and maternal expression, respectively. These results were consistent with the data available for other species for imprinting, such as humans, rats, and mice (<https://www.geneimprint.com/>, accession date 2025/08/20), in contrast to what has been reported in the literature for pigs, where these ratios appear to be balanced between the two parental origins [22, 49].

In order to evaluate the conservation landscape of parental expression imbalance, we compared genes identified in piglets regardless of the tissue analyzed to human and mouse species (as reviewed in [2]). The distribution of the 141 imprinted genes along the pig autosomes and the position of orthologous mouse and human imprinted genes known from the literature is shown in Fig. 7. Only a small proportion of locations overlapped between species, since only 21 genes with a parental expression imbalance were shared between at least two species, while 53 were shared between human and mouse (as reviewed in [2]). As previously mentioned, some known imprinted genes in pigs such as *BEGAIN*, *PLAGL1*, and *HM13* showed parental imbalance below 25:75 and, therefore, were part of our first list of 450 parental expressed genes (Additional file 2 Table S1). Of these, most of the genes that were conserved across piglets, mouse, and human were located in clusters and often showed exclusive parental expression, which is typical of known imprinted genes. For additional common positions between species, some neighbored orthologous genes showed parental expression imbalance (Additional file 1 Fig. S8). In contrast, imprinting patterns (i.e., the direction of parental expressed allele) were very well conserved for 15 out of the 21 imprinted genes that were shared between at least two species. Discrepancies due to conflicting data or tissue dependency were already suggested in the literature

for *COPG2* (*COPI Coat Complex Subunit Gamma 2*) and *IGF2*.

While genes known to be imprinted in pigs were detected, such as *NAP1L5* [56], *MEST* [57], *PEG3* [56], novel imprinted genes were identified, including *POU6F2*, *FAM20B*, *PDE4DIP*, *HSD17B11* (*17-Beta-Hydroxysteroid Dehydrogenase Xi*), and *PREPL*. Among these genes, it is difficult to determine whether they are specific to pigs or shared by a wider phylogenetic clade, since some novel imprinted genes still remain to be discovered [58]. Interestingly, mono-allelic expression was shown for *PDE4DIP* in both *Bos Indicus* [59] and for human patients with acute myeloid leukemia [60]. In addition, these genes appeared to be particularly important for behavioral phenotypes, neural development, growth, and pre- or post-natal mortality, as demonstrated by the effects in knockout mice (Mouse Genome Informatics database). For example, deletions of *FAM20B* in mice led to embryonic lethality or perinatal lethality, with homozygous embryos showing severely stunted growth, with multisystem organ hypoplasia and delayed development [61, 62]. Furthermore, germline mutations in *POU6F2* were significantly associated with Wilms' tumor [63], a child nephroblastoma that is highly prevalent in human imprinting disorders such as Beckwith-Wiedemann syndrome [64]. Although a large majority of imprinted genes that we identified are novel, parental expression imbalance of these genes in humans or mice cannot be ruled out, especially for genes that do not show exclusive paternal or maternal imbalance. Detection of genes with strong parental expression imbalance supports the relevance of our approach [3, 20].

The field of genomic imprinting in pigs is growing, with two recent articles identifying 179 [22] and 121 [49] imprinted genes. The overlap of POE genes we identified with previous studies in pigs is, however, limited, with 15 common genes (Additional file 3 Table S2); differences in developmental stage or breeds used seem to be the most likely explanations [22, 49]. However, some novel parent-of-origin expressed genes are shared with these studies in pigs, including *SLCO1A2* (*Solute Carrier Organic Anion Transporter Family Member 1A2*), *DCAF16* (*DDB1 And CUL4 Associated Factor 16*), *ETNPPL* (*Ethanolamine-Phosphate Phospho-Lyase*), *DOC2B* (*Double C2 Domain Beta*), and *AOX1* (*Aldehyde Oxidase 1*), demonstrating the power of our strategy. While the most robust and extensive approach to date identified 179 genes with POE, only 17 had a 5:95 average expression imbalance in the same tissue for at least one developmental stage, compared to 51 in our study. Moreover, *KBTBD6* [25], *ZNF791* [26], and *ZNF300-like* [27] imprinted genes, identified exclusively in domestic animals, including artiodactyls and carnivores, are among the genes showing total imbalance in our study

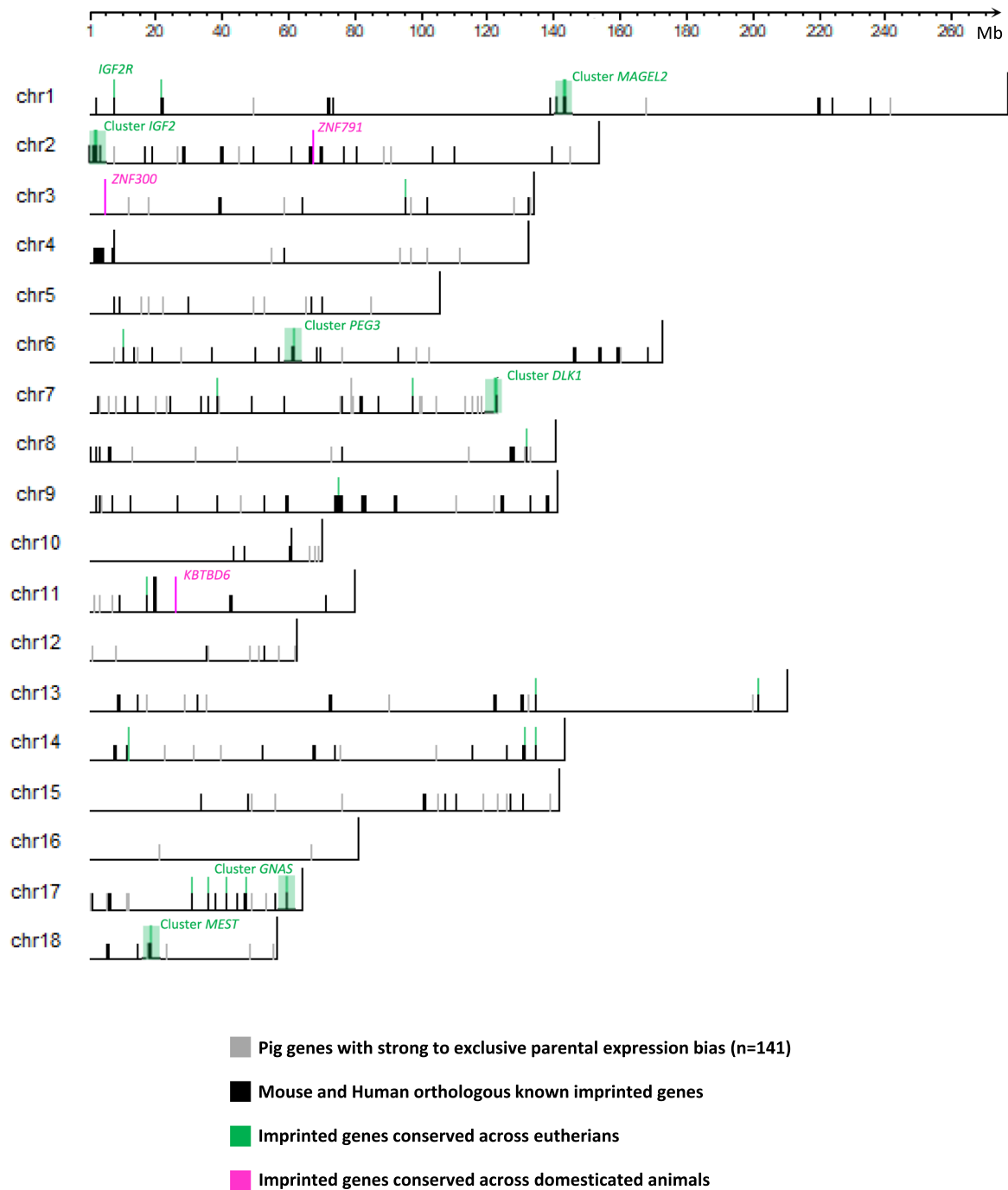


Fig. 7 Conservation of imprinted genes across human and mouse species. Distribution along the pig autosomes of genes showing POE in piglets and orthologous imprinted genes identified in both human and mouse species. The location of the 70, 48 and 51 genes (n= 141) identified in hypothalamus, muscle and placenta, respectively has been plotted on the ideogram in light grey. Positions of mouse and human orthologous imprinted genes [2] have been mentioned in black and overlapping of genes and positions between piglets, mouse and human, suggesting conservation across species, have been highlighted in green. Positions of livestock-specific imprinted genes [25–27] have been mentioned have been highlighted in pink

(Fig. 7). These results increase the spectrum of imprinted genes that may have acquired POE during speciation and genome evolution, likely via the integration of retroviral sequences [65]. Finally, the complexity of data analyses from genome-wide studies is highlighted by our in-depth inspection of specific loci such as the *COPG2* locus, for which we identified the antisense non-coding *COPG2IT1* gene and its paternal expression as well as

an imprinted-specific isoform for the *COPG2* gene. The pattern of expression that we observed for the short isoforms of *COPG2* in hypothalamus may be the result of a tissue-specific alternative promoter [66] or be typical of specific neuronal cells [12, 51]. Although imprinted-isoform-specificity has been highlighted in different species [66, 67], this has not yet been described in pigs.

Conclusions

The ability to detect imprinted genes across a wide range of species provides a valuable opportunity to enhance our understanding of the regulatory mechanisms underlying genomic imprinting. Indeed, the identification in recent years of specific genomic imprinting in domesticated animals emphasizes the importance of acquiring deeper knowledge across eutherians and particularly in livestock species, for which an extensive cartography of imprinted genes could significantly improve the evaluation and contribution of parental gene expression effects on agro-economic complex traits.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-026-01031-2>.

Additional file 1 (PPTX 894 KB)

Additional file 2 (XLSX 16644 KB)

Additional file 3 (XLSX 102 KB)

Acknowledgements

We greatly thank all the people from the INRAE experimental unit [30] for taking care of animals. We greatly thank all the people from the bioinformatics core facility at INRAE Toulouse [33] for their continuous and efficient support. We greatly thank all the people from the genomics core facility at INRAE Toulouse [32] for their skills in various sequencing technologies. We also thank Ingrid David for her help about statistical analyses.

Author contributions

MP performed bioinformatic and statistical analyses, wrote the manuscript. NI performed DNA and RNA extractions for hypothalamus and muscle, managed dataset 1. SL performed sampling and RNA extractions for hypothalamus and muscle. KF supervised Mathilde Perret for molecular experiments. PD performed pearl script for bioinformatic analyses. EJ performed the *IGF2* locus experiments. JNH supervised Eva Jacomet. CI performed ONT experiments. CV performed sequencing quality criteria. SMH performed preprocessing analyses of RNA sequencing and SNP detection from endometrium and placenta. TF performed variant calling from ONT sequencing. LL led the project PORCINET (no ANR-09-GENM005). AB led the project COLOCATION (no ANR-20-CE20-0020). CD led the SeqOcIn project from the Operational program ERDF-FSE MIDI-PYRENEES ET GARONNE 2014–2020. JR is Mathilde Perret supervisor and analysed the data. JD led the project PIPETTE (ANR-18-CE20-0018), supervised Mathilde Perret, analysed the data, wrote the manuscript and draw figures. All authors read, edited and approved the manuscript.

Funding

M.P. has been founded by an INRAE PhD fellowship. J.N.H. has been founded by the French National Agency grant PIPETTE no ANR-18-CE20-0018 and the INRAE Animal Genetics division. Data and analyses from dataset 1 have been founded by both the French National Agency grant PIPETTE no ANR-18-CE20-0018 and the SeqOcIn project from the Operational program ERDF-FSE MIDI-PYRENEES ET GARONNE 2014–2020. Data and analyses from dataset

2 have been founded by both the French National Agency grants PORCINET no ANR-09-GENM005 and COLOCATION no ANR-20-CE20-0020.

Data availability

The DNAseq and RNAseq fastq files of the dataset 1 (blood, muscle and hypothalamus) are available from the European Nucleotide Archive (ENA) under accession number PRJEB86771. For the dataset 2 (placenta), the RNAseq.fastq files have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB75252, a datapaper is also available [31]. Preprocessing variant calling files (.vcf) are deposited in the Recherche Data Gouv public database [37].

Declarations

Ethics approval and consent to participate

All the procedures and guidelines for animal care were approved by the local ethical committee in animal experimentation (Poitou–Charentes) and the French Ministry of Higher Education and Scientific Research (authorizations no 2018021912005794 and no 11789–2017101117033530). Use of animals and the procedures performed in this study were approved by the European Union legislation (directive 86/609/EEC) and French legislation in the Midi-Pyrénées Region of France (Decree 2001–464 29/05/01; accreditation for animal housing C-35–275-32). The technical and scientific staff obtained individual accreditation (MP/01/01/01/11) from the Ethics Committee (région Midi-Pyrénées, France) for experiments involving live. Under these conditions, this study follows the ARRIVE guidelines (Animal Research: Reporting of In Vivo Experiments), and is committed to the 3Rs of laboratory animal research and, thus using the minimum number of animals to achieve statistical significance.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosan, France

²SIGENAE, GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

³GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France

Received: 4 June 2025 / Accepted: 7 January 2026

Published online: 09 March 2026

References

1. Wolf JB, Oakey RJ, Feil R. Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity*. 2014;113:167–75. <https://doi.org/10.1038/hdy.2014.11>.
2. Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC, Tucci V, Bartolomei MS, et al. Genomic imprinting and physiological processes in mammals. *Cell*. 2019;176:952–65. <https://doi.org/10.1016/j.cell.2019.01.043>.
3. Perez JD, Rubinstein ND, Dulac C. New perspectives on genomic imprinting, an essential and multifaceted mode of epigenetic control in the developing and adult brain. *Annu Rev Neurosci*. 2016;39:347–84. <https://doi.org/10.1146/annurev-neuro-061010-113708>.
4. Perez JD, Rubinstein ND, Fernandez DE, Santoro SW, Needleman LA, Ho-Shing O, et al. Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife*. 2015;4:e07860. <https://doi.org/10.7554/eLife.07860>.
5. Monk D, Mackay DJG, Eggermann T, Maher ER, Riccio A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat Rev Genet*. 2019;20:235–48. <https://doi.org/10.1038/s41576-018-0092-0>.
6. Hubert J-N, Perret M, Riquet J, Demars J. Livestock species as emerging models for genomic imprinting. *Front Cell Dev Biol*. 2024;12:1348036. <https://doi.org/10.3389/fcell.2024.1348036>.
7. Bischoff SR, Tsai S, Hardison N, Motsinger-Reif AA, Freking BA, Nonneman D, et al. Characterization of conserved and nonconserved imprinted genes in

- swine. *Biol Reprod.* 2009;81:906–20. <https://doi.org/10.1095/biolreprod.109.078139>.
8. Magee DA, Berkowicz EW, Sikora KM, Berry DP, Park SDE, Kelly AK, et al. A catalogue of validated single nucleotide polymorphisms in bovine orthologs of mammalian imprinted genes and associations with beef production traits. *Animal.* 2010;4:1958–70. <https://doi.org/10.1017/S1751731110001163>.
 9. Park C-H, Uh K-J, Mulligan BP, Jeung E-B, Hyun S-H, Shin T, et al. Analysis of imprinted gene expression in normal fertilized and uniparental preimplantation porcine embryos. *PLoS ONE.* 2011;6:e22216. <https://doi.org/10.1371/journal.pone.0022216>.
 10. Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, et al. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat Genet.* 2015;47:544–9. <https://doi.org/10.1038/ng.3274>.
 11. Daskeviciute D, Chappell-Maor L, Sainty B, Arnaud P, Iglesias-Platas I, Simon C, et al. Non-canonical imprinting, manifesting as post-fertilization placenta-specific parent-of-origin dependent methylation, is not conserved in humans. *Hum Mol Genet.* 2025;34:626–38. <https://doi.org/10.1093/hmg/ddaf009>.
 12. Ho-Shing O, Dulac C. Influences of genomic imprinting on brain function and behavior. *Curr Opin Behav Sci.* 2019;25:66–76. <https://doi.org/10.1016/j.cobeha.2018.08.008>.
 13. Bonthuis PJ, Steinwand S, Stacher Hörndli CN, Emery J, Huang W-C, Kravitz S, et al. Noncanonical genomic imprinting in the monoamine system determines naturalistic foraging and brain-adrenal axis functions. *Cell Rep.* 2022;38:110500. <https://doi.org/10.1016/j.celrep.2022.110500>.
 14. Richard Albert J, Kobayashi T, Inoue A, Monteagudo-Sánchez A, Kumamoto S, Takashima T, et al. Conservation and divergence of canonical and non-canonical imprinting in murids. *Genome Biol.* 2023;24:48. <https://doi.org/10.1186/s13059-023-02869-1>.
 15. Inoue K, Hirose M, Inoue H, Hatanaka Y, Honda A, Hasegawa A, et al. The rodent-specific microRNA cluster within the *Sfmbt2* gene is imprinted and essential for placental development. *Cell Rep.* 2017;19:949–56. <https://doi.org/10.1016/j.celrep.2017.04.018>.
 16. Edwards CA, Takahashi N, Corish JA, Ferguson-Smith AC. The origins of genomic imprinting in mammals. *Reprod Fertil Dev.* 2019;31:1203–18. <https://doi.org/10.1071/RD18176>.
 17. Stringer JM, Pask AJ, Shaw G, Renfree MB. Post-natal imprinting: evidence from marsupials. *Heredity.* 2014;113:145–55. <https://doi.org/10.1038/hdy.2014.10>.
 18. Ishihara T, Suzuki S, Newman TA, Fenelon JC, Griffith OW, Shaw G, et al. Marsupials have monoallelic MEST expression with a conserved antisense lncRNA but MEST is not imprinted. *Heredity.* 2024;132:5–17. <https://doi.org/10.1038/s41437-023-00656-z>.
 19. Wang X, Clark AG. Using next-generation RNA sequencing to identify imprinted genes. *Heredity.* 2014;113:156–66. <https://doi.org/10.1038/hdy.2014.18>.
 20. 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. <https://doi.org/10.7554/eLife.83364>.
 21. Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M, et al. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* 2015;16:112. <https://doi.org/10.1186/s13059-015-0672-7>.
 22. Quan J, Yang M, Wang X, Cai G, Ding R, Zhuang Z, et al. Multi-omic characterization of allele-specific regulatory variation in hybrid pigs. *Nat Commun.* 2024;15:5587. <https://doi.org/10.1038/s41467-024-49923-5>.
 23. Liu R, Tearle R, Low WY, Chen T, Thomsen D, Smith TPL, et al. Distinctive gene expression patterns and imprinting signatures revealed in reciprocal crosses between cattle sub-species. *BMC Genomics.* 2021;22:410. <https://doi.org/10.1186/s12864-021-07667-2>.
 24. St Pierre CL, Macias-Velasco JF, Wayhart JP, Yin L, Semenkovich CF, Lawson HA. Genetic, epigenetic, and environmental mechanisms govern allele-specific gene expression. *Genome Res.* 2022;32:1042–57. <https://doi.org/10.1101/gr.276193.121>.
 25. Ahn J, Hwang I-S, Park M-R, Hwang S, Lee K. Imprinting at the KBTBD6 locus involves species-specific maternal methylation and monoallelic expression in livestock animals. *J Anim Sci Biotechnol.* 2023;14:131. <https://doi.org/10.1186/s40104-023-00931-3>.
 26. Ahn J, Hwang I-S, Park M-R, Rosa-Velazquez M, Cho I-C, Relling AE, et al. Evolutionary lineage-specific genomic imprinting at the ZNF791 locus. *PLOS Genet.* 2025;21:e1011532. <https://doi.org/10.1371/journal.pgen.1011532>.
 27. Ahn J, Hwang I-S, Park M-R, Hwang S, Cho I-C, Lee K. Genomic imprinting at a porcine ZNF locus via a canonical imprinting mechanism. *Anim Genet.* 2025;56:e70034. <https://doi.org/10.1111/age.70034>.
 28. Lefort G, Servien R, Quesnel H, Billon Y, Canario L, Iannuccelli N, et al. The maturity in fetal pigs using a multi-fluid metabolomic approach. *Sci Rep.* 2020;10:19912. <https://doi.org/10.1038/s41598-020-76709-8>.
 29. Hubert J-N, Iannuccelli N, Cabau C, Jacomet E, Billon Y, Serre R-F, et al. Detection of DNA methylation signatures through the lens of genomic imprinting. *Sci Rep.* 2024;14:1694. <https://doi.org/10.1038/s41598-024-52114-3>.
 30. GenESI. Pig innovative breeding experimental facility. 2018. <https://doi.org/10.15454/1.5572415481185847E12>. Accessed 04 Jan 2026.
 31. Maman-Haddad S, Gress L, Suin A, Vialaneix N, Bonnet A. RNA-seq data of pig placenta and endometrium during late gestation. *Data Brief.* 2024;57:111178. <https://doi.org/10.1016/j.dib.2024.111178>.
 32. GeT-PlaGe. GeT-PlaGe genomics and transcriptomics core facility of toulouse. 2022. <https://doi.org/10.17180/NVXJ-5333>. Accessed 04 Jan 2026.
 33. GenoToul Bioinfo. GenoToul bioinformatics facility. 2018. <https://doi.org/10.15454/1.5572369328961167E12>. Accessed 04 Jan 2026.
 34. Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods.* 2021;18:1322–32. <https://doi.org/10.1038/s41592-021-01299-w>.
 35. Lin MF, Rodeh O, Penn J, Bai X, Reid JG, Krashenina O, et al. GLNexus: joint variant calling for large cohort sequencing. *BioRxiv.* 2018. <https://doi.org/10.1101/343970>.
 36. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* 2018;103:338–48. <https://doi.org/10.1016/j.ajhg.2018.07.015>.
 37. Demars J, INRAE. Datasets of variants from DNAseq and RNAseq in pigs; 2025. <https://doi.org/10.57745/RF5PUT>.
 38. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience.* 2021;10:giab008. <https://doi.org/10.1093/gigascience/giab008>.
 39. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
 40. Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet.* 2023;55:1243–9. <https://doi.org/10.1038/s41588-023-01415-w>.
 41. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol.* 2016;17:122. <https://doi.org/10.1186/s13059-016-0974-4>.
 42. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program. *Snpsift Front Genet.* 2012;3:35. <https://doi.org/10.3389/fgenet.2012.00035>.
 43. von Maydell D. PCR allele competitive extension (PACE). *Methods Mol Biol Clifton NJ.* 2023;2638:263–71. https://doi.org/10.1007/978-1-0716-3024-2_18.
 44. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. Stringtie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33:290–5. <https://doi.org/10.1038/nbt.3122>.
 45. Andergassen D, Dotter CP, Wenzel D, Sigl V, Bammer PC, Muckenhuber M, et al. Mapping the mouse allome reveals tissue-specific regulation of allelic expression. *Elife.* 2017;6:e25125. <https://doi.org/10.7554/eLife.25125>.
 46. Oczkowicz M, Szmatola T, Piórkowska K, Ropka-Molik K. Variant calling from RNA-seq data of the brain transcriptome of pigs and its application for allele-specific expression and imprinting analysis. *Gene.* 2018;641:367–75. <https://doi.org/10.1016/j.gene.2017.10.076>.
 47. Wu Y-Q, Zhao H, Li Y-J, Khederzadeh S, Wei H-J, Zhou Z-Y, et al. Genome-wide identification of imprinted genes in pigs and their different imprinting status compared with other mammals. *Zool Res.* 2020;41:721–5. <https://doi.org/10.24272/jissn.2095-8137.2020.072>.
 48. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen L-L, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol.* 2023;24:430–47. <https://doi.org/10.1038/s41580-022-00566-8>.
 49. Li C, Ge M, Long K, Han Z, Li M, Zhang Z, et al. Mechanism of parent-of-origin effects revealed by multi-omic data in euro-chinese hybrid pigs. *Nat Commun.* 2025;16:7542. <https://doi.org/10.1038/s41467-025-62243-6>.
 50. Pinter SF, Colognori D, Beliveau BJ, Sadreyev RI, Payer B, Yildirim E, et al. Allelic imbalance is a prevalent and tissue-specific feature of the mouse

- transcriptome. *Genetics*. 2015;200:537–49. <https://doi.org/10.1534/genetics.115.176263>. 51.
- Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*. 2010;329:643–8. <https://doi.org/10.1126/science.1190830>. 52.
- Monk D. Genomic imprinting in the human placenta. *Am J Obstet Gynecol*. 2015;213:S152–62. <https://doi.org/10.1016/j.ajog.2015.06.032>. 53.
- Soncin F, Khater M, To C, Pizzo D, Farah O, Wakeland A, et al. Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Development*. 2018;145:dev156273. <https://doi.org/10.1242/dev.156273>. 54.
- Stenhouse C, Seo H, Wu G, Johnson GA, Bazer FW. Insights into the regulation of implantation and placentation in humans, rodents, sheep, and pigs. *Adv Exp Med Biol*. 2022;1354:25–48. https://doi.org/10.1007/978-3-030-85686-1_2. 55.
- Monk D, Arnaud P, Apostolidou S, Hills FA, Kelsey G, Stanier P, et al. Limited evolutionary conservation of imprinting in the human placenta. *Proc Natl Acad Sci U S A*. 2006;103:6623–8. <https://doi.org/10.1073/pnas.0511031103>. 56.
- Jiang CD, Li S, Deng CY. Assessment of genomic imprinting of PPP1R9A, NAP1L5 and PEG3 in pigs. *Genetika*. 2011;47:537–42. 57.
- Zhang FW, Han ZB, Deng CY, He HJ, Wu Q. Conservation of genomic imprinting at the NDN, MAGEL2 and MEST loci in pigs. *Genes Genet Syst*. 2012;87:53–8. <https://doi.org/10.1266/ggs.87.53>. 58.
- Hu Y, Yuan S, Du X, Liu J, Zhou W, Wei F. Comparative analysis reveals epigenomic evolution related to species traits and genomic imprinting in mammals. *Innovation*. 2023;4:100434. <https://doi.org/10.1016/j.xinn.2023.100434>. 59.
- de Souza MM, Zerlotini A, Rocha MIP, Bruscadin JJ, da Diniz WJS, Cardoso TF, et al. Allele-specific expression is widespread in *Bos indicus* muscle and affects meat quality candidate genes. *Sci Rep*. 2020;10:10204. <https://doi.org/10.1038/s41598-020-67089-0>. 60.
- Mulet-Lazaro R, van Herk S, Erpelinck C, Bindels E, Sanders MA, Vermeulen C, et al. Allele-specific expression of *GATA2* due to epigenetic dysregulation in *CEBPA* double-mutant AML. *Blood*. 2021;138:160–77. <https://doi.org/10.1182/blood.2020009244>. 61.
- Liu X, Li N, Zhang H, Liu J, Zhou N, Ran C, et al. Inactivation of *Fam20b* in the neural crest-derived mesenchyme of mouse causes multiple craniofacial defects. *Eur J Oral Sci*. 2018;126:433–6. <https://doi.org/10.1111/eos.12563>. 62.
- Vogel P, Hansen GM, Read RW, Vance RB, Thiel M, Liu J, et al. Amelogenesis imperfecta and other biomineralization defects in *Fam20a* and *Fam20c* null mice. *Vet Pathol*. 2012;49:998–1017. <https://doi.org/10.1177/0300985812453177>. 63.
- Perotti D, De Vecchi G, Testi MA, Lualdi E, Modena P, Mondini P, et al. Germline mutations of the *POU6F2* gene in Wilms tumors with loss of heterozygosity on chromosome 7p14. *Hum Mutat*. 2004;24:400–7. <https://doi.org/10.1002/humu.20096>. 64.
- Anvar Z, Acurzio B, Roma J, Cerrato F, Verde G. Origins of DNA methylation defects in Wilms tumors. *Cancer Lett*. 2019;457:119–28. <https://doi.org/10.1016/j.canlet.2019.05.013>. 65.
- Andergassen D, Smith ZD, Kretzmer H, Rinn JL, Meissner A. Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. *Dev Cell*. 2021;56:2995–3005.e4. <https://doi.org/10.1016/j.devcel.2021.10.010>. 66.
- Stelzer Y, Bar S, Bartok O, Afik S, Ronen D, Kadener S, et al. Differentiation of human parthenogenetic pluripotent stem cells reveals multiple tissue- and isoform-specific imprinted transcripts. *Cell Rep*. 2015;11:308–20. <https://doi.org/10.1016/j.celrep.2015.03.023>. 67.
- Newman T, Bond DM, Ishihara T, Rizzoli P, Gouil Q, Hore TA, et al. *PRKACB* is a novel imprinted gene in marsupials. *Epigenetics Chromatin*. 2024;17:29. <https://doi.org/10.1186/s13072-024-00552-8>. 68.
- Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*. 2023;39:btac830. <https://doi.org/10.1093/bioinformatics/btac830>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.