RNA-seq data of pig placenta and endometrium during late gestation

Sarah Maman-Haddad, Laure Gress, Amandine Suin, Nathalie Vialaneix, Agnès Bonnet

Data in Brief 57 (2024) 111178

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

RNA-seq data of pig placenta and endometrium during late gestation

Sarah Maman-Haddad[a,b,*], Laure Gress[a], Amandine Suin[c], Nathalie Vialaneix[d,e], Agnès Bonnet[a]

[a] GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, F31326 Castanet Tolosan, France
[b] INRAE, Sigenae, GenPhySE, MIAT UR875, F31326 Castanet Tolosan, France
[c] GeT-PlaGe, INRAE, Genotoul, Castanet-Tolosan, France
[d] Université de Toulouse, INRAE, UR MIAT, Castanet-Tolosan 31326, France
[e] Plateforme Biostatistique, Genotoul, Toulouse, France

ARTICLE INFO

Article history:
Received 28 May 2024
Revised 6 November 2024
Accepted 21 November 2024
Available online 28 November 2024

Dataset link: PRJEB75252 (Original data)

Keywords:
Transcriptome
End of gestation
Placenta
Endometrium
Pig
lncRNA

ABSTRACT

Limiting the level of piglet losses before weaning is a growing demand from producers and society to improve the welfare and health of sows and piglets. In particular, perinatal mortality, which can be defined as the complete development allowing survival at birth, is mostly due to reduced piglet maturity that occurs at the end of gestation. Fetal growth and maturation depend on a fine balance between the nutrient requirements for optimal fetal growth and the maternal nutrient requirements. This balance occurs at the feto-maternal interface, defined as the interaction between the mother (uterus/endometrium) and the fetus (placenta). Thus, the CO-LOCATION project (ANR20-CE20-0020-01) studies the feto-maternal system in relation to fetal maturation and piglet survival at birth. To this end, we documented the transcriptome of endometrial and placental tissues in late gestation from pure and reciprocal crossbred fetuses using two breeds with extreme fetal maturity: Large White and Meishan, showing substantial and low neonatal mortality, respectively. 224 endometrial and 224 placental samples were selected from the PORCINET tissue collection (ANR-09-GENM-005) together with sow breed, day of gestation, sex, fetal genotypes and maturity. RNA was processed for RNA-

* Corresponding author at: GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, F31326 Castanet Tolosan, France.
E-mail address: sarah.maman@inrae.fr (S. Maman-Haddad).

https://doi.org/10.1016/j.dib.2024.111178

seq analysis using NovaSeq6000. with an average of 107 and 105 million reads per endometrial and placental sample, respectively. Sequences were processed using the Nextflow nf-core/rnaseq pipeline for transcript and gene quantification. The average mapping rate was 91 % and 86 % for endometrial and placental samples, respectively. Then, the TAGADA pipeline was used to reconstruct RNA-seq de novo, predict lncRNA and quantify them. The data generated from this analysis provides a complete transcriptional profile of the feto-maternal interface during late gestation. These data sets are the starting point for further analyses, including differential expression analysis, enrichment analysis and investigation of the feto-maternal dialog.

Specifications Table

Subject	Biological sciences.
Specific subject area	Omics: RNA-seq analysis of the pig feto-maternal interface in late gestation (90 and 110 days of gestation, birth at 116 days of gestation).
Data format	Raw (fastq), Processed (csv format)
Type of data	Tables (csv format), Figures
Data collection	The RNA-seq data were obtained from 224 samples of porcine endometrial tissues and 224 samples of porcine placental tissues, which are adjacent tissues. RNA from the samples was extracted, quality controlled and sequenced individually using the Illumina NovaSeq6000 platform with a paired-end read length of 2 × 150 bp. The data were preprocessed by Nextflow nf-core/rnaseq pipeline and TAGADA Nextflow pipeline for transcriptome quantification and lncRNA prediction, respectively
Data source location	Institution: INRAE Laboratory: Genotoul genomics facility GeT-PlaGe (http://get.genotoul.fr/en/) and GenPhySE Town: Toulouse Country: FRANCE
Data accessibility	Data can be accessed from Repository name: FAANG Data identification number: PRJEB75252 Direct URL to data: https://data.faang.org/dataset?searchTerm=PRJEB75252&sortTerm=accession&sortDirection=desc
Related research article	

1. Value of the Data

- These raw RNA-seq data sets provide a unique picture that simultaneously interrogates the two adjacent tissues (placenta and endometrium). These tissues are the players of the feto-maternal interface during late gestation.
- The quality control for mRNA and sequencing has already been performed and validated.
- The data sets may provide detailed information on the overall genes expressed in the two tissues and the differentially expressed genes (DEGs) between breeds and during late gestation in pigs.
- These data may be used in integrative analyses to characterize the feto-maternal dialogue.

- By applying different workflows, the raw RNA-seq data and bioinformatics processes provided here can be used for further analyses to investigate the role of coding and non-coding genes in the regulation of maternal-fetal compromise mechanisms.

2. Background

The reduction of the piglet losses before weaning is a key issue. Indeed, in commercial pig herds, mortality from birth to weaning ranges from 10 % to 20 % [1] and is not compatible with an efficient and sustainable production. Perinatal mortality is mostly due to reduced piglet maturity, defined as the complete development allowing survival at birth [2,3]. The maturation process varies with the fetal genome and the uterine environment [4] and depends on the dialog between two tissues, the placenta (fetal genome) and the endometrium (maternal genome), to regulate nutrient resource allocation [5].

To study the complex mechanisms within and between the tissues (endometrium and placenta) during the fetal maturation process, we performed RNA-seq on the samples collected in a previous project (PORCINET project, ANR-09-GENM-005) that included pure and reciprocal fetal genotypes.

These data will allow us to gain new biological insights into the feto-maternal interface at the end of gestation, a period corresponding to fetal growth and maturation.

3. Data Description

This study is the first to describe the transcriptome of the two tissues at the feto-maternal interface from four fetal genotypes. These fetuses were generated using two extreme maternal breeds for fetal maturity and piglet survival: Large White (LW: high neonatal mortality) and Meishan (MS: low neonatal mortality) inseminated with mixed semen from both breeds (LW+MS boars).

The data presented in this article include RNA-seq reads from 224 samples for each of the two tissues, the placenta of fetuses and the adjacent endometrium of sows. Fetuses were collected at two days of gestation (D90-D110, term at 114 days). The feto-maternal interface of each fetus was carefully dissected to separate the two adjacent tissues and stored at −80 °C. Fetuses were selected according to their fetal and maternal genotypes (see Experimental Design Material and Methods section), sex and fetal maturation status (seven replicates per condition). Total RNA from the samples was randomly extracted and quality controlled: the mean RNA Integrative Number (RIN), which determines quality, was found to be equal to 7.9 and 6.6 for endometrial and placental samples, respectively. In this study, the placenta had a lower RIN quality than the endometrium. Fig. 1 shows the range of RNA quality by tissue and genotype. MM placentas were also found to have lower RNA quality than other genotypes. Nevertheless 97 % and 75 % of the endometrial and placental samples, respectively, had a RIN $\geq$ 6. Given the number of conditions to be studied, all the RNA samples were sequenced to ensure high statistical power. RNA samples were distributed on five lanes (50 samples per lane) and sequenced using the Illumina NovaSeq6000 platform with a paired-end read length of 2 $\times$ 150 bp.

Sequencing generated an average of 107 and 105 million of reads per endometrium and placenta FASTQ file, respectively. Bioinformatic pre-processing was performed using the Nextflow nf-core/rnaseq pipeline [6–8] to ensure sequence quality and robust transcriptome assembly and quantification. Fig. 2 A–D shows that all FASTQ sequencing files have an average Phred score per sequence >30, indicating high sequence quality. After sequence alignment, the average mapping rate was 91 % and 86 % for endometrium and placenta, respectively. Gene coverage plots (Fig. 2 E–H) show high genome coverage in the genes and lower values at the 5′ than 3′ ends of genes, with an average 5′-3′ bias of 1.38 for endometrial samples and 1.40 for placental samples. These plots showed that the 5'-3' bias is similar across samples and ensured that the downstream analyses are not to be going affected.

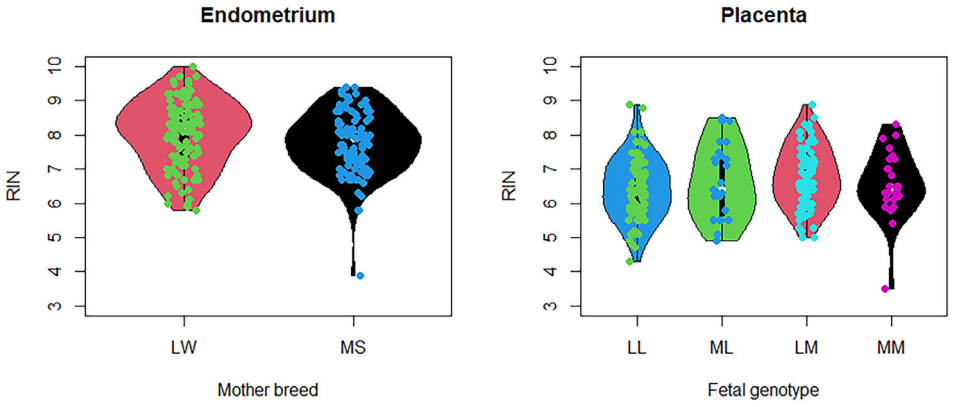

Fig. 1. RNA quality. The RNA quality distribution is displayed with respect to the condition (genotype), which itself depends on the tissue. For the endometrium (maternal tissue), the condition is the sow genotype (Large White (LW) or Meishan (MS)) and for the placenta, the condition is the fetal genotype: (purebred LWxLW (LL), crossbred from LW mother (ML), crossbred from MS mother (LM), purebred MSxMS (MM)).

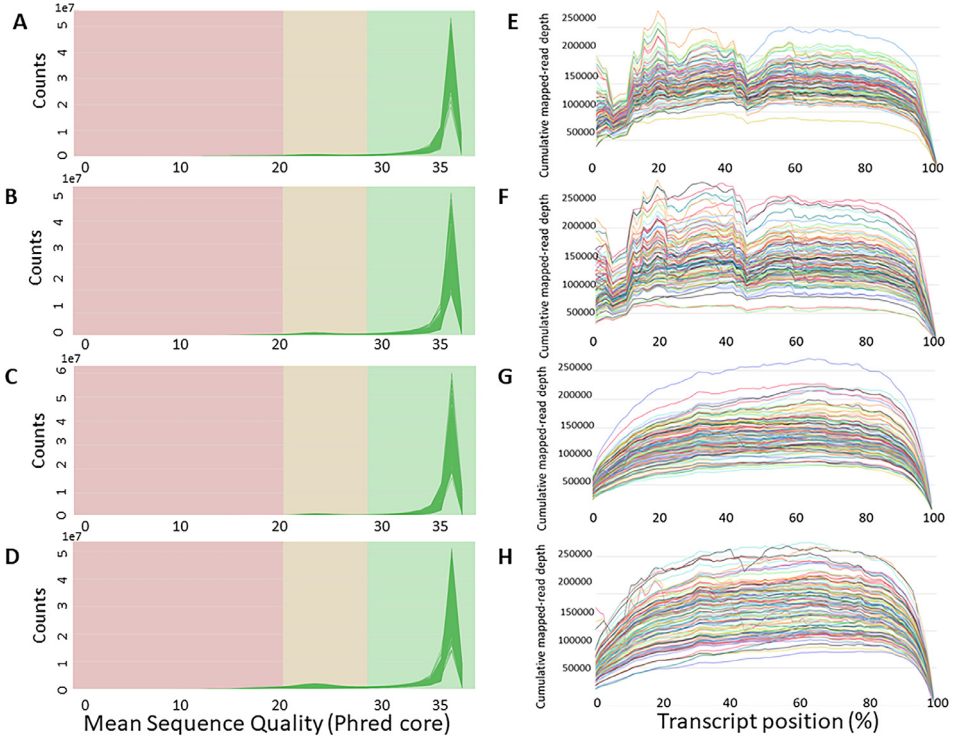

Fig. 2. Bio-informatic quality control. Evaluation of the quality of the sequences, (A), (B): Mean sequence Phred Score for the two batches from the endometrial samples; (C), (D): Mean sequence Phred Score for the two batches from the placental samples. Evaluation of the sequencing alignment, (E), (F): Mean gene coverage profile for each endometrial sample with the count value for the two batches; (G), (H): Mean gene coverage profile for each placental sample for the two batches. Each sample is represented by a different color.

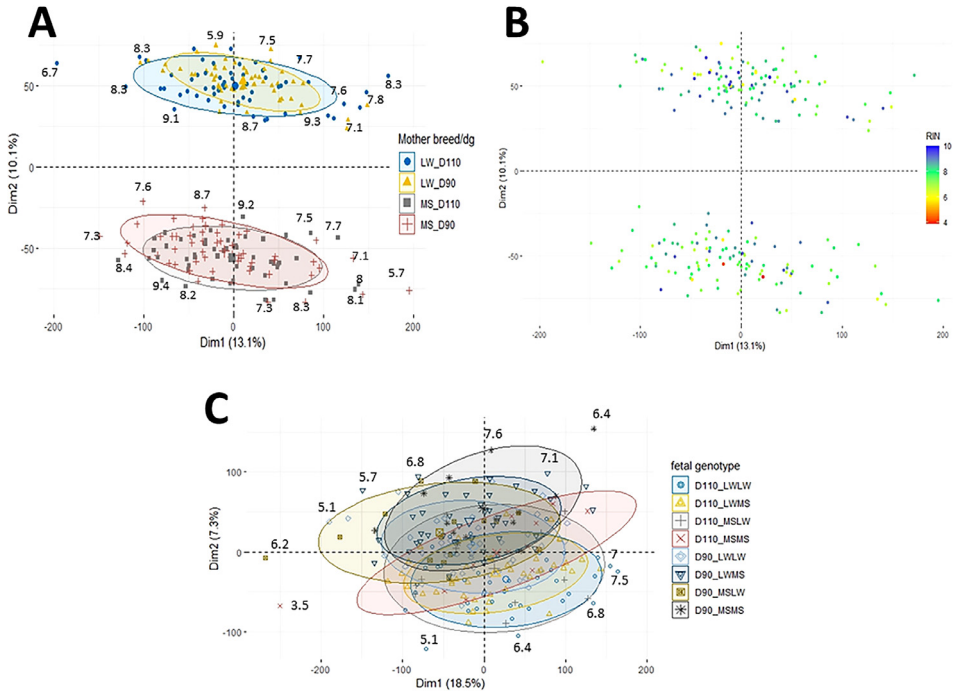

Fig. 3. Principal component analysis. Principal component analysis (PCA) sample plot illustrating the sample variability of in the two datasets. The PCAs were performed on filtrated, RIN adjusted (for endometrium) and transformed data. Sample projection colored by: (A) maternal breed (LW, MS) and day of gestation in the endometrium (D90, D110). (B) RIN in the endometrium. (C) fetal genotype (LL, ML, LM, MM) and day of gestation in the placenta. The ellipses were added to group the samples (ellipse level=0.75, i.e. 75 % of the samples were included in the ellipse). On the plot, the RIN of the sample on the periphery or outside the ellipses was reported.

Raw sequence FASTQ files were deposited in the European Nucleotide Archive (ENA) at EMBL-EBI via the FAANG data portal (https://www.faang.org/).

Supplementary File 1 provides sample information, RNA quality, FASTQ accession numbers, and sequencing information (read count, mapping rate and 5'-3' bias).

The endometrial and placental datasets were then analyzed using Principal Component Analysis (PCA) to detect potential outliers and assess RNA quality effect. The PCAs on data without considering the transcript count threshold identified a RIN effect on the second axis of the PCA (Supplementary file 2) for the endometrial dataset, which needs to be considered in downstream analyses. After transcript count threshold filtration, RIN adjustment (for endometrium using ComBat-seq function of the R package sva) and transformation of the data, PCA plots (Fig. 3A and C) showed the variability of gene expression according to the main conditions of the experimental design, i.e. maternal and fetal genotypes, respectively, and the two days of gestation. Although some outliers were identified for the endometrial tissue (Fig. 3A), they were not associated with the lowest RIN values. In addition, two PCA analyses based on sample group and RIN $\geq$6 or RIN$\geq$7 were performed on endometrium and placenta datasets to further assess how low RIN affects the quality of the analysis. The PCA plots are provided in Supplementary file 3 and showed no improvement in clustering by conditions using samples with RIN $\geq$6 or RIN $\geq$7. These analyses (Fig. 3 and Supplementary file 3) demonstrate that RIN scores are not associated with the conditions of interest in our study and validate the RNA-seq data by showing coherent clustering of samples for each tissue (Fig. 3A and C).

Finally, the RNA-seq data were used to identify novel predicted long non-coding RNA (lncRNA) using the TAGADA Nextflow pipeline [9]. Fig. 4 compares the TAGADA *de novo* recon-

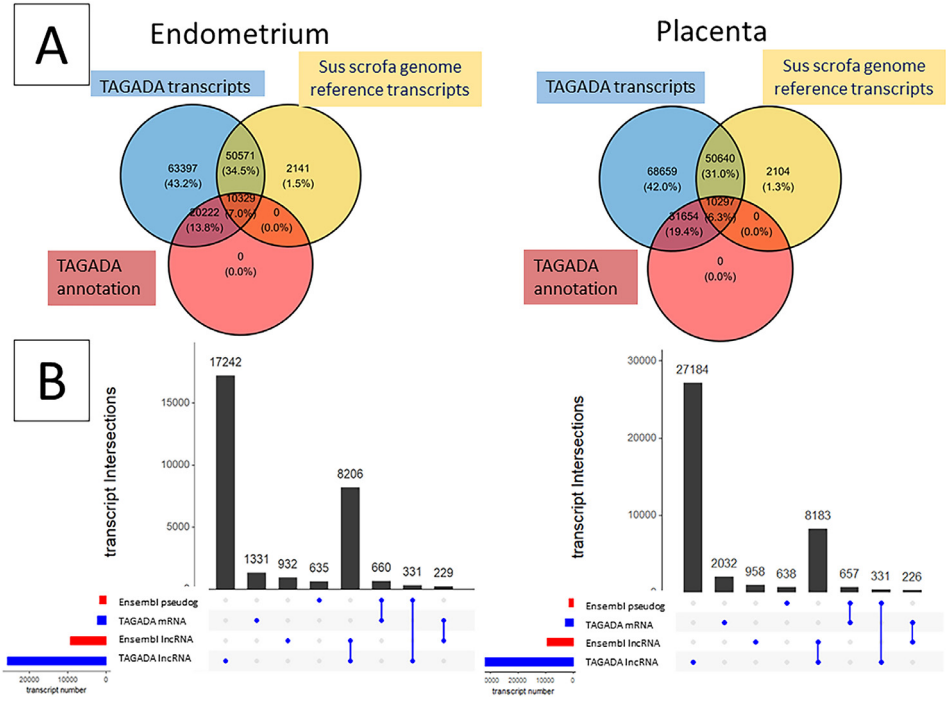

Fig. 4. Comparison of the TAGADA *de novo* reconstruction results with the reference annotation (Sus_scrofa.Sscrofa11.1.104). TAGADA transcripts correspond to the number of transcripts predicted by TAGADA. TAGADA annotation correspond to the transcript biotypes predicted by TAGADA.

struction with the Ensembl annotation (Sus_scrofa.Sscrofa11.1.104) and shows a good overlap between the TAGADA pipeline and the Ensembl lncRNAs but also the identification of many newly annotated lncRNAs.

These data demonstrate the high quality of our sequencing and alignment. The bio-informatic analyses (transcriptome quantification, lncRNA prediction) show that our RNA-seq data may be useful and of relevant to explore the complex mechanisms of feto-maternal dialog during the fetal maturation process via differential gene expression analyses.

4. Experimental Design, Materials and Methods

4.1. Experimental animals

18 Meishan (MS) and 21 Large White (LW) sows were inseminated with mixed semen (LW and MS boars) at the GenESI experimental farm (10.15454/1.5572415481185847E12) using 3 MS boars and 3 LW boars, so that each litter consisted of purebred (LW or MS) and crossbred (LM from MS sows and ML from LW sows) fetuses. The MS and LW breeds are two extreme breeds in terms of fetal maturity and piglet mortality at birth. The LW breed is a highly selected breed with high mortality and reduced maturity at birth, whereas the MS breed produces piglets with extremely low mortality despite being lighter at birth. All fetuses were obtained by cesarean section and euthanized for tissue sampling.

For each of 407 fetuses, a predetermined area of the two adjacent tissues (placenta/endometrium) was dissected, dissociated with forceps, immediately frozen in liquid nitrogen, and stored at −80 °C. Each endometrium sample was paired with its purebred or crossbred

fetus: LWxLW (LL) or MSxLW (ML) from Large White sows (LW) and MSxMS (MM) or LWxMS (LM) from Meishan sows (MS). The detailed sampling protocol has been deposited in FAANG (https://data.faang.org/api/fire_api/samples/INRAE_SOP_COLOcATION-tissues_20210817.pdf).

Finally, 224 fetuses were selected from 28 sows (seven sows per breed and for the two days of gestation (D90/D110)), according to their fetal and maternal genotypes, sex and fetal maturation status (seven replicates per condition). The corresponding tissue sampling is described in the supplementary File 1.

4.2. RNA extraction

Samples were randomized for multiple effects: gestational day, fetal genotype, fetal sex and fetal maturation status. The first level of randomization consisted of 5 series of 50 samples to match the 5 sequencing lanes. A second level of randomization was applied for each sequencing lane, i.e. 2 RNA extraction series of 24 samples each. This randomization was applied to both tissues.

Total RNA was isolated from each of the 448 samples. Briefly, tissue samples were disrupted, homogenized, and ground to a fine powder. An aliquot of 80–100 mg of the fine powder was then processed for total RNA isolation and purification using Trizol (Invitrogen, France) and Nucleospin RNA II kit (Macherey-Nagel, France) according to the manufacturer's instructions. Quantification was performed using a Nanodrop and RNA quality was determined using a 5200 Fragment Analyzer at the Genotoul Genomics Facility GeT-PlaGe (http://get.genotoul.fr/en/). The detailed RNA protocol has been deposited in FAANG (https://api.faang.org/files/protocols/experiments/INRAE_SOP_COLOcATION_RNA_Extraction_20230425.pdf). The samples were extracted in batches of 24. RNA quality was determined by the RNA Integrative Number (RIN) of the 5200 Fragment Analyzer. Fig. 1 reports the RNA quality of the samples according to the maternal breed (endometrium) and fetal genotype (placenta).

4.3. cDNA library and Illumina sequencing

Library preparation and sequencing experiments were performed at the Genotoul Genomics Facility GeT-PlaGe (http://get.genotoul.fr/en/). RNA libraries were prepared according to Illumina's protocols on the Illumina TruSeq RNA Sample. mRNA was first isolated using oligo(dT) beads. mRNA was fragmented and reverse transcribed. Adapter ligation, size selection (150–200 bp) and library amplification were then performed. The detailed library preparation, including quality control (Fragment Analyser profile (Agilent)), was deposited in FAANG (https://api.faang.org/files/protocols/experiments/INRAE_SOP_COLOcATION_Library_Preparation_20230509.pdf). Libraries were sequenced using the Illumina NovaSeq6000 platform with a paired-end read length of 2 $\times$ 150 bp, on 9 different lanes, i.e. 50 samples per lane randomly selected as shown in Supplementary File 1. The generated raw data were trimmed and cleaned by removing low quality reads and the adapters. The samples with low quality or low number of reads were resequenced in lane 10.

4.4. Transcriptome assembly and RNA-seq quantification

For each tissue, the 224 FASTQ files were grouped into 2 batches in order to run the Nextflow nf-core/rnaseq pipeline 6:8] on an accurate number of samples and to optimize the resources and memory allocated to each run. In each tissue, the batch contained two common samples to verify that there was no difference in counts between the batches. Finally, a local configuration file was loaded for each pipeline to optimize resources (option cleanup = true) and to retrieve pipeline traces.

The Nextflow nf-core/rnaRNAseq pipeline, was launched with STAR mapping software and RSEM for transcript abundance calculation.

The pipeline (version nfcore-Nextflow-v20.11.0-edge) was run on the four batches (one batch per folder) using 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). The DESeq2 step was skipped in the pipeline (-skip_deseq2_qc argument).

Finally, the pipeline produced a gene count and a transcript count matrix per batch, which are available for statistical analysis.

4.5. Bio-informatic quality control

Quality control of sequencing reads and sequence alignments was performed using FastQC and QualiMap software, respectively (from Nextflow nf-core/rnaseq pipeline MultiQC tool). We validated that the Phred quality scores of sequencing reads and the sequence alignments are of high quality for downstream analysis (Fig. 2).

4.6. RNA-seq tool versions

- Bedtools, version 2.29.2
- dupradar, version 1.18.0
- FASTQC, version 0.11.9
- nexflow, version 20.11.0
- nf-core/rnaseq, version 3.0
- picard, version 2.23.9
- preseq, version 2.0.3
- QualiMap, version 2.2.2-dev
- rsem, version 1.3.1
- rseqc, version 3.0.1
- samtools, version 1.10
- stringtie, version 2.1.4
- subread, version 2.0.1
- trimgalore, version 377
- MultiQC, version 1.9

4.7. Multivariate exploratory analyses

Transcript count data were RIN adjusted using ComBat-seq function of the R package sva). The transcripts with fewer than five counts in 100 % of replicates for at least one of the conditions analysed were filtered out using Bioconductor's HTSFilter package. Transcript count normalization and transformation was carried out by applying the variance stabilizing transformation (VST) function from DeSeq2. Then, after scaling and centering, PCA was used, to detect potential outliers (Fig. 3) (FactoMineR package, version 2.4).

To evaluate the RIN effect, transcript count data were log 2 transformed ($\log2(x+1)$) before PCA analysis (Supplementary File 2).

4.8. lncRNA prediction

The lncRNA prediction was performed on the 448 RNA-seq samples using the TAGADA Nextflow pipeline [9]. TAGADA pipeline includes a FEELnc step for lncRNA detection [10], which

required the selection of a prediction method (shuffle or intergenic) before running the TAGADA pipeline. The selection of the best FEELnc prediction method was based on the evaluation of the quality of the prediction and of the quantification for the two methods.

4.8.1. Choice of the prediction method

Two lncRNA prediction methods (shuffle/intergenic) were evaluated using the FEELnc pipeline (FlExible Extraction of Long non-coding RNAs) version FEELnc-v.0.1.1 [9] based on reconstructed transcripts from RNA-seq data. GTF files of samples generated by the Nextflow nf-core/rnaseq pipeline were merged with StringTie and passed to the first module of FEELnc. Three modules were started in the FEELnc lncRNA prediction pipeline.

To select the best prediction method, we evaluated the quality of the two methods (shuffle or intergenic) by assessing at the receiver operating characteristic (ROC) curve and at the quality of the lncRNA quantification. Supplementary File 4 Fig. 1 shows the evaluation of the quality of prediction for the two methods (shuffle or intergenic) using the two graph ROC curves (TGROC). In addition, Supplementary File 4 Table 1 compares the quality of the lncRNA quantification for the intergenic and shuffle methods. The results showed fewer ambiguous transcripts with the intergenic method, both for the endometrium and the placenta (Supplementary file 4 Fig. 1), and more lncRNAs predicted for the two tissues with the shuffle method (supplementary File 4 Table 1). Therefore, the shuffle method was selected for lncRNA detection.

4.8.2. RNA-seq de novo reconstruction and lncRNA quantification

Finally, the TAGADA pipeline was launched for *de novo* RNA-seq reconstruction and for lncRNA prediction and quantification. These steps were performed using the BAM files obtained from the Nextflow nf-core/rnaseq pipeline as input data to save processing time and resources.

Limitations

In this study, we observed samples with low RNA quality on measurements of endometrial gene expression levels. However, we were able to show that RNA degradation was not associated with the conditions of interest. Nevertheless, to avoid any effect of RNA degradation, we recommend to explicitly correct the endometrial data for the effect of RIN using the ComBat-seq function of the R package sva [11] by stratifying RIN measures (10 classes). In accordance with the study of Romero et al. [12], we believe that this approach is the most appropriate to recover biologically meaningful signals from samples with RNA of heterogeneous quality.

Ethics Statement

The use of animals and the procedures performed in this study were approved by European Union legislation (Directive 86/609/EEC) and by French legislation in the Midi-Pyrénées region of France (Decree 2001-464 29/05/01; accreditation for animal facilities C-35-275-32). The technical and scientific personnel have received individual accreditation (MP/01/01/01/11) from the Ethics Committee (Région Midi-Pyrénées, France) for experiments involving live animals. Under these conditions, this study follows the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines and is committed to the 3Rs of laboratory animal research, using the minimum number of animals to achieve statistical significance.

Credit Author Statement

Sarah Maman-Haddad: Bioinformatic and analysis management, FAIR management, validation, Writing- Original draft preparation. **Laure Gress:** Data management plan supervision, Bioinformatic analysis, Writing draft. **Amandine Suin:** RNA-seq sequencing. **Nathalie Vialaneix:**

Data management plan, FAANG/ENA RNA-seq submission. **Agnes Bonnet:** Conceptualization, Methodology, Bioinformatic analysis, Data curation, Supervision, Project administration, Writing and reviewing draft and Funding acquisition.

Data Availability

PRJEB75252 (Original data) (FAANG).

Acknowledgments

This work is a part of the CO-LOcATION project that is supported by the French National Agency of Research (grant N° ANR20-CE20-0020-01).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that have influenced, or could be perceived to have influenced, the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.111178.

References

[1] R. Muns, M. Nuntapaitoon, P. ummaruk, Non-infectious causes of pre-weaning mortality in piglets, Livest. Sci. 184 (2016) 46–57, doi:10.1016/j.livsci.2015.11.025.
[2] V. Voillet, M. SanCristobal, Y. Lippi, P.G. Martin, N. Iannuccelli, C. Lascor, F. Vignoles, Y. Billon, L. Canario, L. Liaubet, Muscle transcriptomic investigation of late fetal development identifies candidate genes for piglet maturity, BMC Genom. 15 (1) (2014) 797, doi:10.1186/1471-2164-15-797.
[3] G. Lefort, R. Servien, H. Quesnel, Y. Billon, L. Canario, N. Iannuccelli, C. Canlet, A. Paris, N. Vialaneix, L. Liaubet, The maturity in fetal pigs using a multi-fluid metabolomic approach, Sci. Rep. 10 (2020) 19912, doi:10.1038/s41598-020-76709-8.
[4] N.J. Biensen, M.E. Wilson, S.P. Ford, The impacts of uterine environment and fetal genotype on conceptus size and placental vascularity during late gestation in pigs, J. Anim. Sci. 77 (4) (1999) 954–959, doi:10.2527/1999.774954x.
[5] K.A. Vonnahme, C.O. Lemley, Programming the offspring through altered uteroplacental hemodynamics: how maternal environment impacts uterine and umbilical blood flow in cattle, sheep and pigs, Reprod. Fertility Dev. 24 (1) (2012) 97–104, doi:10.1071/RD11910.
[6] H. Patel, P. Ewels, A. Peltzer, O. Botvinnik, G. Sturm, D. Moreno, P. Vemuri, M. U Garcia, silviamorins, L. Pantano, M. Binzer-Panchal, nf-core bot, R. Syme, M. Zepper, G. Kelly, F. Hanssen, J.A. Fellows Yates, C. Cheshire, rfenouil, J. Espinosa-Carrasco, marchoeppner, E. Miller, A. Talbot, P. Zhou, S. Guinchard, M. Hörtenhuber, G. Gabernet, C. Mertes, D. Straub, P. Di Tommaso, nf-core/rnaseq (2023), doi:10.5281/zenodo.1400710.
[7] P. Di Tommaso, M. Chatzou, E.W. Floden, P.P. Barja, E. Palumbo, C. Notredame, Nextflow enables reproducible computational workflows, Nat. Biotechnol. 35 (4) (2017) 316–319, doi:10.1038/nbt.3820.
[8] P.A. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M.U. Garcia, P. Di Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines, Nat. Biotechnol. 38 (3) (2020) 276–278, doi:10.1038/s41587-020-0439-x.
[9] C. Kurylo, C. Guyomar, S. Foissac, S. Djebali, TAGADA: a scalable pipeline to improve genome annotations with RNA-seq data, NAR Genom. Bioinform. 5 (10) (2023), doi:10.1093/nargab/lqad089.
[10] V. Wucher, F. Legeai, B. Hédan, G. Rizk, L. Lagoutte, T. Leeb, V. Jagannathan, E. Cadieu, A. David, H. Lohi, S. Cirera, M. Fredholm, N. Botherel, P. Leegwater, C. Le Béguec, H. Fieten, C. Johansson, J. Johnsson, LUPA consortium, J. Alifoldi, C. André, K. Lindblad-Toh, C. Hitte, T. Derrien, FEELnc: a tool for Long non-coding RNAs annotation and its application to the dog transcriptome, Nucl. Acids Res. 45 (8) (2017) e57, doi:10.1093/nar/gkw1306.

[11] Y. Zhang, G. Parmigiani, WE. Johnson: batch effect adjustment for RNA-seq count data, NAR Genom. Bioinform. 2 (3) (2020) 284–292, doi:10.1093/nargab/lqaa078.
[12] I. Gallego Romero, AA. Pai, J. Tung, Y. Gilad, RNA-seq: impact of RNA degradation on transcript quantification, BMC Biol. 12 (2014) 42, doi:10.1186/1741-7007-12-42.