



## Data in Brief

# Gene expression profiling by high throughput sequencing to determine signatures for the bovine receptive uterus at early gestation



Veerle Van Hoeck<sup>a</sup>, Saara C. Scolari<sup>a</sup>, Guilherme Pugliesi<sup>a</sup>, Angela M. Gonella-Diaza<sup>a</sup>, Sônia C.S. Andrade<sup>a</sup>, Gustavo R. Gasparin<sup>b</sup>, Luiz L. Coutinho<sup>b</sup>, Mario Binelli<sup>a</sup>

<sup>a</sup> Laboratório de Fisiologia e Endocrinologia Molecular, Department of Animal Reproduction, FMVZ-USP, Pirassununga, Sao Paulo, Brazil

<sup>b</sup> Laboratório de Genética Animal, Departament of Animal Science, ESAIQ-USP, Piracicaba, Sao Paulo, Brazil

## ARTICLE INFO

## Article history:

Received 24 May 2015

Accepted 25 May 2015

Available online 3 June 2015

## Keywords:

Cow

Receptivity

Pregnancy

Uterus

RNA-Seq

## ABSTRACT

The uterus plays a central role among the reproductive tissues in the context of early embryo-maternal communication and a successful pregnancy depends on a complex series of endometrial molecular and cellular events. The factors responsible for the initial interaction between maternal and embryonic tissues, leading to the establishment of pregnancy, remain poorly understood. In this context, Illumina's next-generation sequencing technology has been used to discover the uterine transcriptome signature that is favourable for ongoing pregnancy. More specifically, the present report documents on a retrospective in vivo study in which data on pregnancy outcome were linked to uterine gene expression signatures on day 6 (bovine model). Using the RNA-Seq method, 14,654 reference genes were effectively analysed for differential expression between pregnant and non-pregnant uterine tissue. Transcriptome data revealed that 216 genes were differently expressed when comparing uterine tissue from pregnant and non-pregnant cows. All read sequences were deposited in the Sequence Read Archive (SRA) of the NCBI (<http://www.ncbi.nlm.nih.gov/sra>). An overview of the gene expression data has been deposited in NCBI's Gene Expression Omnibus (GEO) and is accessible through GEO Series accession number GSE65117. This allows the research community to enhance reproducibility and allows for new discoveries by comparing datasets of signatures linked to receptivity and/or pregnancy success. The resulting information can serve as tool to identify valuable and urgently needed biomarkers for scoring maternal receptivity and even for accurate detection of early pregnancy, which is a matter of cross-species interest. Beyond gene expression analysis as a marker tool, the RNA-Seq information on pregnant uterine tissue can be used to gain novel mechanistic insights, such as by identifying alternative splicing events, allele-specific expression, and rare and novel transcripts that might be involved in the onset of maternal receptivity. This concept is unique and provides a new approach towards strategies that are highly needed to improve efficiency of fertility treatments.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Specifications

Organism/cell line/tissue	<i>Bos indicus</i> /uterine cells/uterine tissue
Sex	Female
Sequencer or array type	Illumina HiScanSQ
Data format	Raw data BioProject PRJNA268916 SRA Study SRP05036
Experimental factors	Pregnant versus non-pregnant uterine biopsies
Experimental features	Day 6 endometrial transcriptome signatures from bovine uterine biopsies that were collected in the same cycle of the AI. Data on pregnancy outcome were retrospectively linked with endometrial transcriptome. Samples were collected from contralateral uterine horn in relation to the pregnancy without a negative effect on pregnancy outcome as previously reported.[1]

(continued)

## Specifications

Consent	n/a
Sample source location	The experiment was carried out in Pirassununga, Brazil, at University of São Paulo. It was conducted at the research farm of the University of São Paulo (USP) in Pirassununga, SP, Brazil, using 51 clinically healthy, cycling, multiparous, non-lactating Nellore cows ( <i>Bos indicus</i> ).

## 1. Direct link to deposited genomic data

<http://www.ncbi.nlm.nih.gov/sra/>

E-mail address: [binelli@usp.br](mailto:binelli@usp.br) (M. Binelli).

<http://dx.doi.org/10.1016/j.gdata.2015.05.030>

2213-5960/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 2. Experimental design, materials and methods

### 2.1. Animals, reproductive management and experimental design

The experiment was carried out in accordance with the Ethics and Animal Handling Committee of the School of Veterinary Medicine and Animal Science of the University of São Paulo. It was conducted at the research farm of the University of São Paulo (USP) in Pirassununga, SP, Brazil, using 51 clinically healthy, cycling, multiparous, non-lactating Nellore cows (*Bos indicus*), 4 to 10 years old, with a body condition score between 3 and 4 (scale: 0, emaciated; 5, obese), maintained under the same pasture regimen supplemented with sugar cane and/or corn silage, concentrate and free access to water and mineralized salt.

Estruses were synchronized ( $n = 51$ ) by injecting two im doses of prostaglandin F<sub>2</sub> $\alpha$  (PGF<sub>2</sub> $\alpha$ ) analogue (500  $\mu$ g sodium cloprostenol; Sincrocio, Ouro Fino Animal Health, Cravinhos, Brazil) 14 days apart. At the time of the second PGF<sub>2</sub> $\alpha$  injection, animals were equipped with an Estroject (Rockway, Inc., USA) device, and animals were observed twice daily until the sixth day after the last PGF<sub>2</sub> $\alpha$  injection for standing heat. Ultrasound exams were performed with the aid of a duplex B-mode (grey-scale) and Colour-Doppler instrument (MyLab30 Vet Gold; Esaote Healthcare, São Paulo, SP, Brazil) equipped with a multi-frequency linear transducer 12 h after the animals displayed estrus in order to confirm the presence and location of a pre-ovulatory follicle and AI was performed. Six days post-AI, an endometrial biopsy was collected from the uterine horn contralateral to ovulation, as determined by the presence of a corpus luteum (CL). The biopsy procedure was described in detail elsewhere [1]. Once collected, the biopsy was placed in cryotubes and immediately immersed in liquid nitrogen, then transferred to  $-80$  °C for storage. Blood sampling for determination of progesterone concentrations was performed at the time of biopsy and processed as described by Mesquita et al. [2]. After the biopsy procedure, cows received two 1 g doses of the anti-inflammatory medication flunixin meglumine (Desflan, Ouro Fino Animal Health, Cravinhos, Brazil) in a 24 h interval and one dose of a Penicillin/Streptomycin based antibiotic (Penfort Reforçado, Ouro Fino Animal Health, Cravinhos, Brazil). Plasma P4 concentrations were measured by a solid-phase radioimmunoassay (Coat-A-Count; Siemens Medical Solutions Diagnostics, Los Angeles, USA). The animals were observed for signs of abnormal vaginal secretions and/or discomfort after the biopsy. Pregnancy diagnosis was performed 30 days post-AI via trans-rectal ultrasonography exam.

### 2.2. Animal selection for RNA-Seq

RNA-Seq was performed on samples selected retrospectively based on the day 30 pregnancy diagnosis ( $n = 6$  samples from each pregnant and non-pregnant animals). Animals were selected based on two criteria: plasma P4 concentrations on day 6 must have been within a 1.5 ng/mL interval among all animals and, ovulations must have been confirmed within 12 h after AI as determined by ultrasound exam.

### 2.3. RNA isolation and cDNA synthesis

Frozen uterine samples (30 mg) were macerated in liquid nitrogen using a stainless steel apparatus, immediately mixed with buffer RLT from the RNeasy Mini columns kit (Qiagen, São Paulo, SP, Brazil) and further DNA, RNA and protein extractions performed using a commercial purification kit (Qiagen AllPrep DNA/RNA/PROTEIN Mini Kit, São Paulo, SP, Brazil) according to the manufacturer's recommendations. RNA purity and concentration were determined by spectrophotometry (Thermo Fisher Scientific, Suwanee, GA, USA) and RNA integrity was assessed using the 2100 Bioanalyzer (Agilent Technologies Brasil Ltda, São Paulo, SP, Brazil). RNA integrity numbers (RIN) of individual samples ranged from 7 to 8. The synthesis of deoxyribonucleic acid (cDNA) was performed by reverse transcription using the High Capacity

cDNA Reverse Transcription Kit (Life Technologies, Frederick, Maryland, USA) according to manufacturer's instructions.

### 2.4. RNA-Seq and analysis

For the transcriptome analyses of expression patterns in pregnant and non-pregnant animals, cDNA was generated using a routine RNA library preparation TruSeq protocol developed by Illumina Technologies (San Diego, CA) using 1  $\mu$ g of total RNA as input. Using the kit, mRNA was first isolated from total RNA by performing a polyA selection step, followed by construction of paired-end sequencing libraries with an insert size of about 300 bp. Briefly, polyA selected RNA was cleaved as per Illumina protocol and the cleaved fragments were used to generate first strand cDNA using SuperScript II reverse transcriptase and random hexamers. Subsequently second strand cDNA was synthesized with RNase H and DNA polymerase enzyme. Adapter ligation and end repair steps followed second strand synthesis. Resulting products were amplified via PCR and cDNA libraries were then purified and validated using the Bioanalyzer 2100 (Agilent Technologies). Paired-end sequencing was performed using the Illumina HiScanSQ platform. Samples were multiplexed with unique six-mer barcodes and run on multiple lanes to obtain  $2 \times 100$  bp reads. The paired end (PE) reads were filtered using the Seqclean v1.4.13 package (<https://bitbucket.org/izhbannikov/seqclean>) which removed all reads with a mean quality under 26 and removed the primer and vector contaminants using the UniVec database (<http://www.ncbi.nlm.nih.gov/tools/vecscreen/univect/>). The reads were mapped using the local alignment with Bowtie2 v2.1.0 [3] against the masked *Bos taurus* genome masked assembly (*Bos taurus*UMD3.1) and read counts were obtained using HTSeq-count v0.5.4p2 (<http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>). From the six biopsies collected in the non-pregnant cows and used for RNA-Seq, one sample did not correctly align with the bovine genome and was consequently omitted from the analysis. The differential expression analysis of the RNA-Seq data was performed with package DESeq2 v1.12.1 [4], from R [5]. The normalized counts were then obtained for each annotated gene. After normalization, the log<sub>2</sub> fold change was obtained through a general linear model approach, the significance test was followed by the FDR-Benjamini-Hochberg [6] correction for multiple tests. Genes with the normalized read counts <5 across all samples were filtered out from the analysis. The enrichment analyses of different functional databases was done using the functional annotation tool of the Database for Annotation, Visualization, and Integrated Discovery (DAVID Bioinformatics Resources 6.7, NIAID/NIH) using as background the genes used on the differential expression analysis [7]. The total list of genes derived from pregnant and non-pregnant endometrium was subjected to gene ontology annotation. From the total list, genes involved in different cellular and molecular processes could be identified.

### 2.5. Quantitative Real-time Polymerase Chain Reaction

Nine of the differentially regulated genes as well as exhibiting the highest fold changes between pregnant and non-pregnant cows were selected from the RNA-Seq results for qRT-PCR analysis. Primers for each selected gene were designed using PrimerQuestQM tools (IDT; <http://idtdna.com/Scitools/Applications/Primerquest>). Amplicon sequence identity was confirmed with NCBI Basic Local Alignment Search Tool software (Blast <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Quantitative Real-time PCR was performed using the StepOnePlus Applied Biosystem Real-Time PCR System and amplification was conducted in triplicate. Reactions were performed in a final volume of 20  $\mu$ L using 10.0  $\mu$ L of Power SYBR Green PCR Master Mix (Warrington, UK), 10  $\mu$ M of each primer (forward and reverse), and 4  $\mu$ L of cDNA (diluted 1:40). Temperature profiles comprised an initial denaturation step at 95 °C for 10 min, and 40 cycles with denaturation at 95 °C for 15 s and annealing at 60 °C for 1 min. Cyclophilin was used as the normalizing

reference gene. Relative expression levels of the selected target genes were calculated with the LinRegPCR software method. The cycle threshold (Ct) values determined for the target genes were normalized against the reference gene. PCR product identity was confirmed by sequencing. Statistical analyses of qRT-PCR data were performed using Student's paired t-test. The transcript levels of nine genes were re-evaluated using Quantitative Real-time PCR (qRT-PCR). RNA from the same animals assigned to the RNA-Seq analysis (pregnant,  $n = 6$ ; non-pregnant,  $n = 6$ ) were used for the qRT-PCR validation experiment. Moreover, transcript levels from uterine biopsies of 15 additional animals (pregnant,  $n = 5$ ; non-pregnant,  $n = 10$ ) from animals of the same experiment were quantified. Overall, there was an agreement for the results obtained from RNA-Seq and qRT-PCR [8].

## References

- [1] G. Pugliesi, S.C. Scolari, F.S. Mesquita, M. Maturana Filho, E.R. Araújo, D. Cardoso, et al., Impact of probing the reproductive tract during early pregnancy on fertility of beef cows. *Repr. Domest. Anim.* 49 (2014) 35–39.
- [2] F.S. Mesquita, G. Pugliesi, S.C. Scolari, M.R. França, R.S. Ramos, M. Oliveira, et al., Manipulation of the periovulatory sex steroidal milieu affects endometrial but not luteal gene expression in early diestrus nelore cows. *Therio.* 81 (2014) 861–869.
- [3] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2. *Nat. Methods.* 9 (2012) 357–359.
- [4] M.J. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 15 (2014) 550.
- [5] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5 (2004) 80 PMID: 15461798.
- [6] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JRSS Ser. B.* (1995) 289–300.
- [7] J.R. Dennis, G. Sherman, B.T. Hosack, D.A. Yang, J. Gao, DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4 (2003) 3.
- [8] M. Binelli, S.C. Scolari, G. Pugliesi, V. Van Hoeck, A.M. Gonella-Diaza, S.C.S. Andrade, et al., The transcriptome signature of the receptive bovine uterus determined at early gestation. *PLoS ONE* 10 (2015) e0122874.