ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12721

Pre- and post-puberty expression of genes and proteins in the uterus of *Bos indicus* heifers: the luteal phase effect post-puberty

M. R. S. Fortes[*], L. F. Zacchi[*,1], L. T. Nguyen[*,†], F. Raidan[‡], M. M. D. C. A. Weller[§], J. J. Y. Choo[*], A. Reverter[‡], J. P. A. Rego[¶], G. B. Boe-Hansen[**], L. R. Porto-Neto[‡], S. A. Lehnert[‡], A. Cánovas[††], B. L. Schulz[*], A. Islas-Trejo[‡‡], J. F. Medrano[‡‡], M. G. Thomas[§§] and S. S. Moore[¶¶]

[*]School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD 4072, Australia. [†]Faculty of Biotechnology, Vietnam National University of Agriculture, Gialam, Hanoi, Vietnam. [‡]Animal Science Department, Universidade Federal de Viçosa, Vicosa, Minas Gerais, 36570-900, Brazil. [§]CSIRO Agriculture and Food, Queensland Bioscience Precinct, Brisbane, QLD 4072, Australia. [¶]Instituto Federal de Educação, Ciência e Tecnologia do Ceara, Fortaleza, Ceará, 62930-000, Brazil. [**]School of Veterinary Sciences, The University of Queensland, Gatton, QLD 4343, Australia. [††]Department of Animal Biosciences, Centre of Genetic Improvement for Livestock, University of Guelph, Guelph, ON, N1G 2W1, Canada. [‡‡]Department of Animal Science, University of California Davis, Davis, CA, 95616, USA. [§§]Department of Animal Science, Colorado State University, Fort Collins, CO, 80523, USA. [¶¶]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD 4072, Australia.

Summary

Progesterone signaling and uterine function are crucial in terms of pregnancy establishment. To investigate how the uterine tissue and its secretion changes in relation to puberty, we sampled tissue and uterine fluid from six pre- and six post-pubertal Brahman heifers. Post-pubertal heifers were sampled in the luteal phase. Gene expression of the uterine tissue was investigated with RNA-sequencing, whereas the uterine fluid was used for protein profiling with mass spectrometry. A total of 4034 genes were differentially expressed (DE) at a nominal P-value of 0.05, and 26 genes were significantly DE after Bonferroni correction ($P < 3.1 \times 10^{-6}$). We also identified 79 proteins (out of 230 proteins) that were DE ($P < 1 \times 10^{-5}$) in the uterine fluid. When we compared proteomics and transcriptome results, four DE proteins were identified as being encoded by DE genes: *OVGP1*, *GRP*, *CAP1* and *HBA*. Except for *CAP1*, the other three had lower expression post-puberty. The function of these four genes hypothetically related to preparation of the uterus for a potential pregnancy is discussed in the context of puberty. All DE genes and proteins were also used in pathway and ontology enrichment analyses to investigate overall function. The DE genes were enriched for terms related to ribosomal activity. Transcription factors that were deemed key regulators of DE genes are also reported. Transcription factors *ZNF567*, *ZNF775*, *RELA*, *PIAS2*, *LHX4*, *SOX2*, *MEF2C*, *ZNF354C*, *HMG20A*, *TCF7L2*, *ZNF420*, *HIC1*, *GTF3A* and two novel genes had the highest regulatory impact factor scores. These data can help to understand how puberty influences uterine function.

Keywords beef cattle, RNA-sequencing, gene expression, genetics, proteomics

Introduction

The pre-pubertal uterus is considered less able to support embryonic or fetal development in gilts, as it has been reported to lead to pregnancy loss in embryo transfer experiments (Henning & Davis 2003). In the cow, lower circulating progesterone concentrations have been associated with a uterine environment with reduced ability to sustain conceptus elongation and therefore pregnancy (Forde *et al.* 2011a). As embryo implantation is related to uterine receptivity, it is important to characterize the uterine environment at the molecular level.

Previous studies have investigated the endometrial transcriptome of cows in early pregnancy and cyclic conditions. In early pregnancy, endometrial gene expression was found to be independent of conceptus presence until maternal recognition, which occurs on day 15–16 in cattle (Forde

Address for correspondence

M. R. S. Fortes, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia.
E-mail: m.fortes@uq.edu.au
[1]Present address: ARC Training Centre for Biopharmaceutical Innovation, The University of Queensland, St. Lucia, Qld 4072, Australia

Accepted for publication 26 July 2018

et al. 2011b). The endometrial transcriptome of heifers was found to be sensitive to cyclic variation and circulating progesterone concentrations (Forde *et al.* 2011a). Together, this evidence suggests that preparation for pregnancy occurs independently from the presence of the conceptus, as a natural part of the uterine changes during the estrous cycle. Preparation for pregnancy is primarily a response to progesterone signaling. To further understand this mechanism, we posed the question: what is the difference in transcriptome profile and uterine secretion between heifers that have never experienced progesterone signaling (those with pre-pubertal status) and heifers that are experiencing progesterone signaling as they enter reproductive life (post-puberty)?

The uterine fluid composition changes with the estrous cycle, presumably in preparation for pregnancy. The abundance of proteins in uterine fluid changes mostly in the luteal (or progesterone-driven) phase of the cycle (Soleilhavoup *et al.* 2016). The uterine fluid composition also changes before implantation of the conceptus (Forde *et al.* 2014). In this study, uterine fluid composition was explored using mass spectrometry methods. The objective of this study was to investigate the effect of pubertal transition and the related progesterone impact on uterus gene expression and secreted proteins in uterine fluid using transcriptomics and proteomics to compare pre-pubertal and post-pubertal cycling *Bos indicus* heifers.

Materials and methods

Animal management and puberty observation

Management, handling and euthanasia of animals were approved by the Animal Ethics Committee of The University of Queensland, Production and Companion Animal group (certificate number QAAFI/279/12). A total of 20 Brahman heifers, which were not pedigreed animals but had a characteristic *B. indicus* phenotype and were typical beef industry animals, were sourced from two commercial herds in Queensland, Australia, as weaners (<250 kg) and kept at grazing conditions at the University of Queensland Gatton Campus facilities. The heifers were all kept under the same grazing conditions.

Heifers were examined every fortnight from October 2012 to May 2013 for physical indications of puberty. Ovarian activity was observed using ultrasonography [HS-2000(VET), Honda Electronics Inc.]. Pubertal status was defined by presence of a corpus luteum (CL) observed using ultrasound (Johnston *et al.* 2009). Euthanasia was carried out by stunning with a captive bolt followed by exsanguination. Euthanasia plans were based on date of first CL observation. Six heifers were pre-pubertal and six heifers were post-pubertal at euthanasia. When a post-pubertal heifer was identified, it was paired with a pre-pubertal heifer that was randomly drawn from the remaining animals and processed on the same day. Post-pubertal heifers were euthanized while in the luteal phase of their second estrous cycle, confirmed by the presence of a CL on ovarian tissue post euthanasia. Progesterone measurements were also carried as detailed before (Fortes *et al.* 2016), and concentrations of progesterone were 0.4 ± 0.2 ng/ml for pre-pubertal heifers and 2.0 ± 0.7 ng/ml for post-pubertal heifers.

Sample harvesting

Post-euthanasia, the entire uterus and oviduct were identified and dissected. A mixture of oviduct and uterine fluid samples were obtained by micro-perfusion with PBS flushing (at 37 °C) on the initial portion of both (right and left) infundibulum. The fluid was collected by suction in the medial portion of the uterine horn. The oviduct and uterine fluid were combined in clean tubes and subjected to centrifugation at 700 ***g*** for 10 min (at 4 °C) to separate fluid from cells. The supernatant was further centrifuged at 10 000 ***g*** for 60 min at 4 °C, divided into aliquots and kept at −80 °C until proteomics analyses. For transcriptomics, a biopsy of uterine tissue containing the endometrial layer was collected from the tip of the right and left horn in approximation of the junction with the isthmus, preserved by snap freezing in liquid nitrogen and then kept at −80 °C until RNA extraction. Biopsies were taken after micro-perfusion flushing of the uterine fluid, and the vast majority of cells in the sample were from the uterine tissue (not the fluid). Biopsies were taken from the uterine proximal portion, as close as possible to the utero–tubal junction. In post-pubertal heifers, the biopsies were taken from this same proximal portion in the uterine horn that was ipsilateral to the CL observed. RNA expression of the tissue, not the fluid, was captured in the RNA-sequencing (RNA-seq) experiment.

Proteomics mass spectrometry and data analysis

For proteomics, we used all 12 animals described in the experiment—six pre-pubertal heifers and six post-pubertal heifers—as biological replicates. To prepare the samples for mass spectrometry proteomics, oviduct and uterine fluid containing approximately 50 µg of protein was pipetted into Protein LoBind tubes (Eppendorf) containing a solution with a final concentration of 0.5% SDS, 50 mM Tris-HCl buffer pH 8 and 10 mM DTT. Samples were incubated at 95 °C for 20 min to denature the proteins and break disulphide bonds. Samples were cooled to room temperature, alkylated by addition of acrylamide to a final concentration of 25 mM and incubated at room temperature for 1 h to block the reformation of disulphide bridges. Samples were precipitated by addition of 4 volumes of 1:1 methanol:acetone followed by incubation at −20 °C for 16 h. After 10 min of centrifugation at 16 000 ***g***, the protein pellet was dried and resuspended in 100 µl of 50 mM NH_4HCO_3, and proteins were proteolitically digested by addition of 1 μg of trypsin (Sigma) at 37 °C with agitation for 18 h. Insoluble material was pelleted at 16 000 ***g*** for 1 min, and soluble peptides were desalted using C_{18} ZipTips (Millipore) and resuspended in 9.25% acetonitrile and 0.1% formic acid. Desalted peptides were analyzed by liquid chromatography electrospray ionization tandem mass spectrometry with a TripleTOF 5600 instrument (SCIEX), as previously described (Xu *et al.* 2015), except that a 70-min LC gradient was used to separate the peptides. Proteins were identified from Information Dependent Acquisition analysis of one randomly chosen case sample and one randomly chosen control sample with PROTEINPILOT v5.0.1 (SCIEX), searching against a database of proteins (downloaded from Uniprot on 28 March 2016) with a total of 43 813 entries assigned to *B. taurus* and which also included bovine viruses for thoroughness (a total of 6870 reviewed entries from Swiss-prot and 36 948 unreviewed entries from TrEMBL). Uniprot contains only 1310 entries assigned to *B. indicus* (49 reviewed entries from Swiss-prot and 1261 entries from TrEMBL), and no *B. indicus* proteome is available. For the purpose of analysis, a more complete proteome was required, and thus *B. taurus* was chosen, given that genome sequence differences between the two subspecies would not affect the overall protein quantification process. Sequential window acquisition of all theoretical mass spectra (SWATH-MS) relative quantitative proteomics (Gillet *et al.* 2012) data were analyzed with PEAKVIEW v2.1 (SCIEX). Statistical analyses and Gene Ontology (GO) analyses were performed using MSSTATS and GOSTATS in R, as previously described (Zacchi & Schulz 2016), using an in-house-designed Python script to include entries from TrEMBL. MSSTATS fits linear mixed models and reports adjusted and unadjusted *P*-values considering multiple testing (Choi *et al.* 2014). The proteomics dataset is available from the data repository known as Pride (submission reference: 1-20180401-140710).

RNA extraction and sequencing

Total RNA was isolated separately from ~30 mg of the homogenized uterine tissue samples from pre- and post-pubertal heifers. Tissue samples were added to 1 ml of TRIzol reagent (Sigma-Aldrich) with a QIAshredder homogenizer (QIAGEN Pty, Ltd.). Homogenized samples were added to 200 μl of chloroform (Sigma-Aldrich), mixed vigorously by hand and allowed to stand for 5 min at room temperature. This was followed by centrifugation at 14 000 ***g*** for 10 min at 4 °C. The aqueous layer with RNA was transferred to a new tube, and 200 μl of ethanol (100%; Chem-supply) was added. The sample mixture with ethanol was transferred to an RNeasy mini column, and RNA purification followed the kit's protocol (QIAGEN Pty, Ltd.). Sample concentrations were measured with a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific) with an optimal 260/280 ratio between 1.8 and 2.1. Sample quality was confirmed using RNA integrity number (RIN) values verified with a Bioanalyser (Agilent Technologies, Inc.). All samples passed quality control, with RINs ranging from 6.6 to 8.4 and an average A260/A230 ratio of around 1.95, indicating RNA of sufficient quality for our analyses (Gallego Romero *et al.* 2014). cDNA libraries for each sample were made using an Illumina TrueSeq stranded mRNA Sample Preparation Kit (Illumina). This kit uses poly-T oligo attached magnetic beads, which select for RNA with a poly-A tail—mostly mRNA but also a subset of non-coding RNA. Sequencing was performed on a HiSeq 2000 Sequencer Analyzer (Illumina), following standard methods (Fortes *et al.* 2016). The UMD3.1 reference assembly for the bovine genome (release 77; ftp://ftp.ensembl.org/pub/release-77/genbank/bos_taurus/) was used to guide sequence assembly, performed with CLC workbench software (CLC Bio). The same software was used for quality control of sequence data and calculation of RPKM (reads per kilo base per million mapped reads) per gene (Mortazavi *et al.* 2008; Cánovas *et al.* 2013). By normalizing for RNA length and total reads in each sample, the RPKM measure facilitates comparisons of gene expression levels across samples (Mortazavi *et al.* 2008). FASTQC-project protocols were used (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequence-read lengths and base coverage, nucleotide contributions and base ambiguities were considered in the analyses, as described before (Cánovas *et al.* 2014a, 2014b). A threshold of RPKM greater than 0.2 was used to select genes expressed in the samples (Wickramasinghe *et al.* 2012). Samples analyzed passed all the quality control parameters, and we were able to continue with data mining, described as follows.

The RNA sequence data are freely available and can be retrieved from the BioSamples data portal: https://www.ebi.ac.uk/biosamples/samples?text=SAMEA104618936. Biosamples identifiers are: SAMEA104618936, SAMEA104618947, SAMEA104618901, SAMEA104618921, SAMEA104618958, SAMEA104618907, SAMEA104618919, SAMEA104618915, SAMEA104618894, SAMEA104618939, SAMEA104618932 and SAMEA104618923. This dataset is publically available through the Functional Annotation of Animal Genomes (FAANG) initiative (http://data.faang.org/home).

Normalization of RNA-seq data

In the present study, we applied a base-2 log transformation of the RPKM reads. The log transformation helps to stabilize the variance of RPKM values, an issue of critical importance as differential expression of particularly low counts can be easily biased without transformation (Bullard *et al.* 2010). We then adopted methodology initially proposed for the normalization of gene expression microarray intensity signals and based on mixed model linear equations (Byrne *et al.* 2005; Norris *et al.* 2005; Reverter *et al.* 2005). Accordingly, we fitted the following mixed effect model to the log-transformed RPKM values:

$$Y_{ijkp} = \mu + L_i + G_j + GA_{jk} + GP_{jp} + e_{ijkp},$$

where Y_{ijkp} represents the base-2 log-transformed RPKM value from the *i*th library (with 72 levels), the *j*th gene (with 16 973 levels) of the *k*th animal (with 12 levels) in the *p*th physiological state (pre- and post-puberty); μ is the overall mean; L_i represents the fixed effect of the *i*th library; G_j represents the random effect of the *j*th gene; GA and GP represent the random interaction effects of the gene × animal and gene × physiological states respectively; and e_{ijkp} is the residual term associated with the measurement in Y_{ijkp}. Using standard stochastic assumptions, the effects of *G*, *GA*, *GP* and *e* were assumed to follow a normal distribution with a zero mean and between-gene, between-gene within animal, between-gene within-physiological state and within-gene components of variance respectively. Restricted maximum likelihood estimates of variance components and solutions to model effects were obtained using VCE6 software (ftp://ftp.tzv.fal.de/pub/vce6/).The linear combination of solutions $G + GA + GP$ was used to obtain the normalized mean expression of each gene in each of the samples under scrutiny. Following normalization, we used a *t*-test to compare averages of pre-pubertal and post-pubertal heifers and identify differentially expressed (DE) genes ($P < 0.05$).

Identification of key transcription factors

The regulatory impact factor (RIF) metrics described by Reverter *et al.* (2010) were used to identify the transcription factors (TFs) with the highest evidence of contributing to differential expression in the two physiological states: pre- and post-puberty. Herein, the experimental contrast was pre- vs. post-puberty. We used all known TFs in the RIF metrics for the *i*th regulator ($i = 1, 2, \ldots, 1085$) and estimated RIF values using the following formulae:

$$\mathrm{RIF1}_i = \frac{1}{n_{\mathrm{de}}} \sum_{j=1}^{j=n_{\mathrm{de}}} \hat{a}_j \times \hat{d}_j \times \mathrm{DC}_{ij}^2$$

and

$$\mathrm{RI2}_i = \frac{1}{n_{\mathrm{de}}} \sum_{j=1}^{j=n_{\mathrm{de}}} [(e1_j \times r1_{ij})^2 - (e2_j \times r2_{ij})^2],$$

where n_{de} is the number of DE genes; $\hat{a}_j$ is the estimated average expression of the *j*th DE gene, averaged across the two conditions being contrasted (pre- and post-puberty) and computed as $\hat{a}_j = (e1_j - e2_j)/2$, where $e1_j$ and $e2_j$ represent the expression of the *j*th DE genes in conditions 1 (pre-puberty) and 2 (post-puberty) respectively; $\hat{d}_j$ is the estimated differential expression of the *j*th DE gene and computed as $\hat{d}_j = e1_j - e2_j$; and DC is the differential correlation between the *i*th TF and the *j*th DE gene and is computed from the difference between $r1_{ij}$ and $r2_{ij}$, the co-expression correlation between the *i*th TF and the *j*th DE gene in conditions 1 (pre-puberty) and 2 (post-puberty) respectively.

In practical terms, RIF1 captures TFs showing differential connectivity (i.e. differential wiring of the network) to abundant genes showing differential expression between pre- and post-puberty, whereas RIF2 focuses on TFs that showed evidence as predictors of change in abundance of genes with differential expression between pre- and post-puberty. Using a nominal threshold, $P < 0.05$, a TF was deemed as a key TF if either of the two RIF scores was greater than 1.96 SD.

Functional enrichment analysis

Enrichment analyses of GO terms were performed using 4034 DE genes ($P < 0.05$) as target gene lists and the genes expressed in these tissues (16 973 genes) as a background list. The biological terms enriched for DE genes were identified using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) functional annotation chart (Huang da *et al.* 2009). Significant GO terms and pathways after Benjamini-Hochberg correction are reported.

Enrichment analyses for proteomics results were also carried out. GO analyses were performed using MSSTATS and GOSTATS in R as previously described (Zacchi & Schulz 2016).

Results

Differentially expressed proteins

In total, 260 proteins were identified by IDA LC-MS/MS (Table S1) using PROTEINPILOT (SCIEX), of which 258 were identified with a 1% global false discovery rate. The uterine fluid proteome was qualitatively very similar to serum, with highly abundant proteins including serum albumin, complement proteins, serotransferrin and alpha-2-macroglobulin (Wiktorowicz & Soman 2016; Moulder *et al.* 2017; Kuscuoglu *et al.* 2018). One of the most confidently identified proteins, and clearly distinct from the serum proteome, was the oviduct-specific glycoprotein (Q28042), encoded by the *OVGP1* gene. Of the proteins identified, 230 were quantified by SWATH-MS relative quantitative proteomic analysis. Proteins were either higher ($n = 118$) or lower ($n = 112$) in abundance when pre- and post-pubertal animals were compared (Fig. 1). After adjusting the *P*-value for multiple testing, 79 proteins were found to be DE ($P < 1 \times 10^{-5}$). The expression of 35 of these proteins was lower in post-pubertal heifers, whereas for 44 it was higher. From the adjusted *P*-values, four DE proteins were also DE genes in the RNA-seq data: *OVGP1*, *GRP*, *CAP1* and *HBA* (Table 1). Data on the 230 proteins analyzed is provided in Table S2.

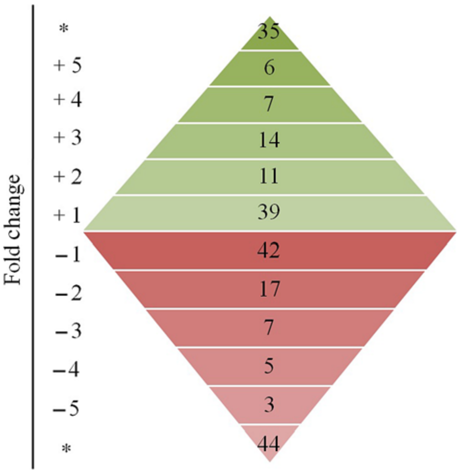

Figure 1 Distribution of uterine fluid proteins with different fold change in the comparison between six pre-pubertal heifers, which had no previous experience of a luteal phase, and six post-pubertal heifers in the luteal phase. Proteins that were higher in pre-pubertal heifers are marked in green; proteins that were lower are marked in pink. Numbers represent the numbers of proteins identified according to fold change. *The number of proteins that were differentially expressed ($P < 1 \times 10^{-5}$).

From the two groups of higher and lower DE proteins, we identified 18 significant enriched GO terms ($P < 0.01$) as shown in Fig. 2. The difference between the pre- and post-pubertal uterine fluid can be summarized as the presence of extracellular proteins (secreted) or intracellular proteins; these were some of the most significant GO terms.

Differentially expressed genes

Data normalization of RPKM values of 0.2 or higher identified 16 973 genes expressed in the uterus. Of these expressed genes, 4034 genes were DE in post-pubertal heifers ($P < 0.05$) when compared to the pre-pubertal heifers (Table S3). A heat map illustrating the hierarchical clustering of the 4034 DE genes between pre- and post-pubertal groups validated the optimality of data normalization in this study (Fig. 3). After Bonferroni correction, 26 genes were found to be significantly DE ($P < 3.1 \times 10^{-6}$). The genes that were DE and had an absolute fold change higher than 3 are highlighted in a volcano plot (Fig. 4). The most significant DE gene was *EDAR*.

Enrichment analyses of GO terms were performed using the 4034 DE genes as a target gene list and the genes expressed in these tissues (16 973 genes) as a background list. The biological terms enriched for DE genes were identified using the DAVID functional annotation chart (Huang da *et al.* 2009). After correction for multiple testing, DE genes in the uterus were enriched for 11 GO terms (Table S4). The GO terms included mostly terms related to ribosome and ribosomal activity. The target DE genes were also enriched for two KEGG pathways: Ribosome and Cell Adhesion Molecules (CAMs).

Key regulatory genes

Using the RIF metrics, we identified 100 significant top-ranking TFs ($P < 0.05$) in the uterus (Table S5). Of these 100 TF, 19 were also DE genes (Table 2). Twenty-five of the expressed genes encoded for TFs of the zinc finger family. Of these 25 zinc finger TFs, two (*ZNF146* and *ZNF84*) were lower in expression at puberty ($P < 0.05$), one (*ZNF367*) was higher at puberty ($P < 0.05$) and the remaining were not DE. Transcription factors *ZNF567*, *ZNF775*, *RELA*, *PIAS2*, *LHX4*, *SOX2* and a novel gene had the highest RIF1 scores (>3 SD units) in the uterus. For RIF2, *MEF2C*, *ZNF354C*, *HMG20A*, *TCF7L2*, *ZNF420*, *HIC1*, *GTF3A* and a novel gene had the highest RIF2 scores (>2 SD units). Here we use the term ‘novel genes’ in reference to features identified as genes in the bovine genome that have yet to receive a name and an official symbol in NCBI. Together, proteomics and transcriptomics data serve as evidence for puberty being a complex, polygenetic process, regulated by multiple TFs.

Discussion

The purpose of this study was to compare gene expression in the uterus and protein expression in uterine fluid of pre- and post-pubertal heifers using transcriptomic and proteomic approaches. The post-pubertal heifers were euthanized in the luteal phase of the cycle, and thus it was expected that the presence of progesterone would impact the expression of genes and abundance of proteins, as previously described in ewes (Soleilhavoup *et al.* 2016).

Four genes—*OVGP1*, *GRP*, *CAP1* and *HBA*—were identified as being significantly DE in the transcriptome and in

Table 1 Genes and proteins that were differentially expressed between 6 pre- and 6 post-pubertal samples according to both proteomics and transcriptomics approaches.

UNIPROT accession	log2FC	Adj. *P*-value	Gene	Gene name	DE RNA-seq	*t*-test *P*-value
Q28042	−0.67	1.0×10^{-16}	*OVGP1*	*Oviductal glycoprotein 1*	−1.11	2.0×10^{-2}
Q863C3	−2.47	1.0×10^{-16}	*GRP*	*Gastrin-releasing peptide*	−1.96	1.5×10^{-4}
Q3SYV4	0.56	1.0×10^{-16}	*CAP1*	*Cyclase associated actin cytoskeleton regulatory protein 1*	0.31	7.6×10^{-3}
P01966	−0.8	1.0×10^{-16}	*HBA*	*Hemoglobin alpha 2*	−2.35	1.7×10^{-2}

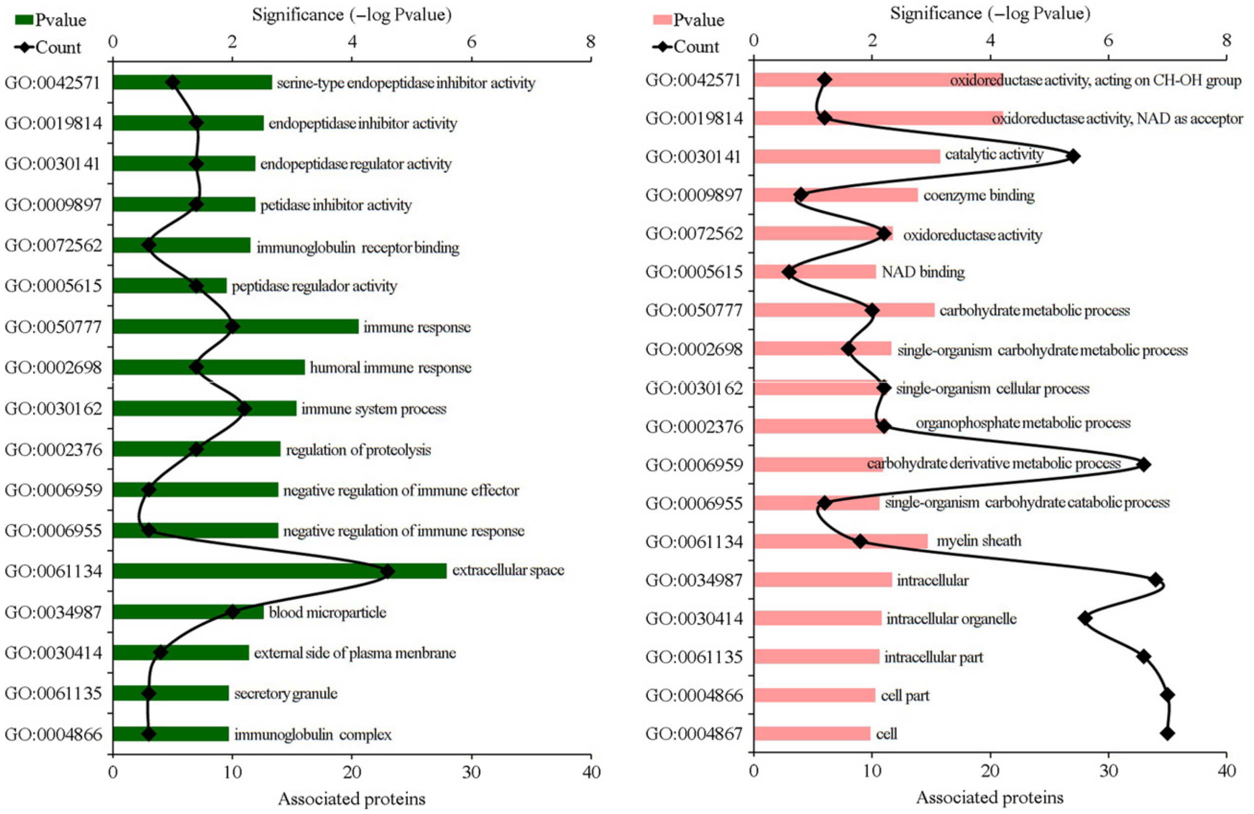

Figure 2 Enriched Gene Ontology (GO) terms (bars, $P > 0.01$) and number of proteins (count) related to each GO term (lines) present in the uterine fluid sampled of pre-pubertal heifers and post pubertal heifers. GO terms associated with proteins that were lower in post-pubertal heifers are marked in green; GO terms associated with proteins that were higher are marked in pink. These comparisons considered six pre-pubertal heifers and six post-pubertal heifers.

the proteome comparisons. OVGP1 is an oviduct specific glycoprotein that has been observed to be less abundant in the luteal phase in sheep compared to the estrus phase (Soleilhavoup *et al.* 2016). Identified as DE by two approaches in our experiment and by mass spectrometry in the ewe experiment, OVGP1 is an interesting target for further investigation of puberty and its role in the uterus of ruminants. In other mammals this glycoprotein has been reported to be critical for sperm–egg interactions and fertility (Lyng & Shur 2009; Yang *et al.* 2015). The abundance of proteins in the uterine fluid is a sum of secretory activity from oviduct and uterine cells and cells of the immune system (Lee *et al.* 2015). OVGP1 is a large, carbohydrate-rich, epithelial glycoprotein that is expressed in uterine tissues from late follicular development through early embryonic development, being downregulated when exposed to lipopolysaccharide challenge in an infection model (Ibrahim *et al.* 2015). Other proteins that were DE in their abundance may be part of the immune system operating in the uterus.

GRP has been reported to be IFNT (interferon-tau)-stimulated and influenced by progesterone signaling in sheep (Song *et al.* 2008). This evidence seems contrary to our results, as we found a lower expression of *GRP* and its protein in post-pubertal animals. It is possible that progesterone alone, without IFNT stimulation, leads to lower levels of *GRP* expression in post-pubertal cattle, a hypothesis that merits further investigation. Changes in *GRP* expression in the cow endometrium during the estrus cycle have also been reported (Mitko *et al.* 2008). It is also important to note that *GRP* produces a number of peptides with varying biological activity (Giraud *et al.* 2010). One of its peptides, oGRP1–46, seems to play a crucial role during embryo implantation in sheep (Giraud *et al.* 1994, 2010).

The expression of the protein encoded by *CAP1* has been observed in the uterine fluid of pregnant cows, and the transcriptomics profile suggests a higher mRNA abundance in the conceptus (Forde *et al.* 2014). In studied heifers, *CAP1* has been detected as a DE gene and the associated protein has been found in higher abundance in the post-pubertal luteal phase. Therefore, *CAP1* may be secreted by uterine cells to prepare the uterus environment for a potential pregnancy in cattle.

HBA (*hemoglobin alpha 2*) is the fourth DE gene that was also found to be significant in the proteomics experiment and is recognized for its function in basic oxygen transport. Recently, it has emerged that epithelial cells, including cells of the reproductive tract, also express hemoglobin genes

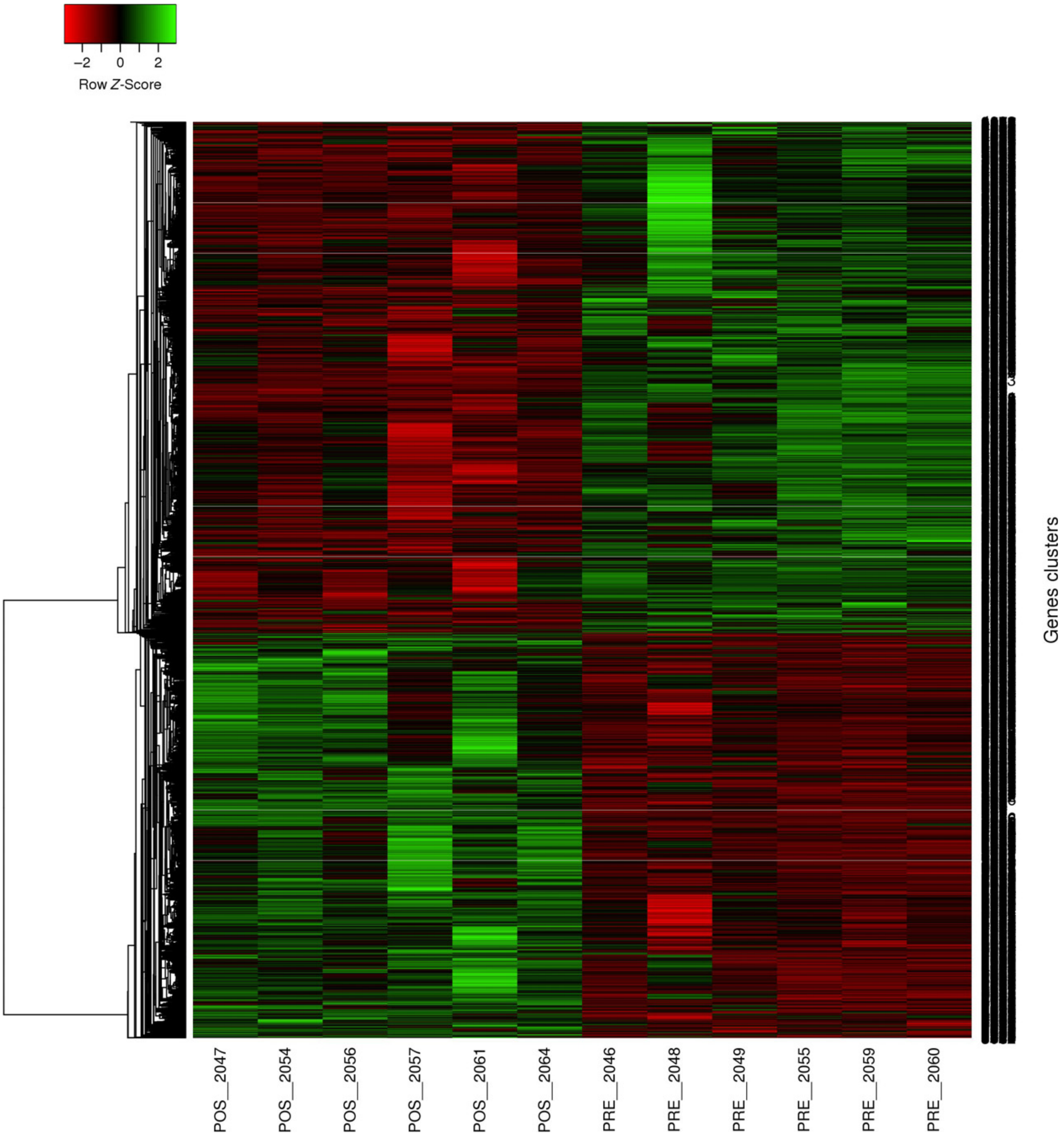

Figure 3 Heat map representing the hierarchical clustering of differentially expressed (DE) genes in the uterus tissue between pre- and post-pubertal Brahman heifers. Unsupervised hierarchical clustering heat map used Pearson correlation and centroid linkage. A total of 4034 DE genes were analysed and visualized. Expression level is color coded: red for under-expressed genes, black for unchanged expression and green for over-expressed genes. The *y*-axis represents gene clusters, and the *x*-axis represents physiological states and samples, POS meaning post-puberty and PRE meaning pre-puberty.

(Saha *et al.* 2017). It has been speculated that its function in epithelia may be to protect tissues from oxidative stress-induced inflammation. In a study of endometritis in Korean cattle, hemoglobin beta gene expression was undetectable in endometritis samples compared to normal controls, further pointing to a potential role in epithelial immune system regulation in the reproductive tract (Choe *et al.* 2010).

These four genes emerged as related to proteins that are differentially abundant, possibly as a direct result of variation in mRNA expression. The DE genes and proteins reported herein may play a role in preparing or maturing

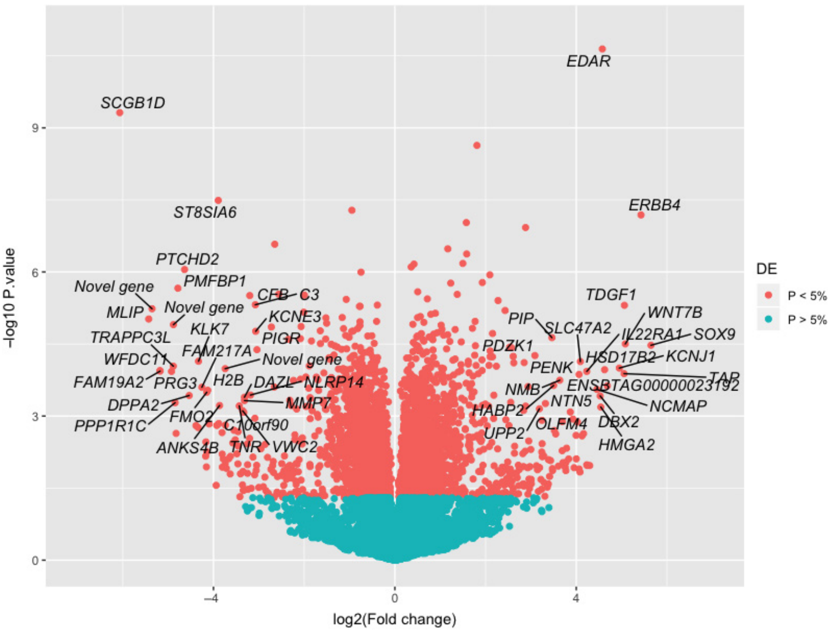

Figure 4 Volcano plot with gene expression results for the uterus samples. Log 2-fold changes are represented on the *x*-axis and −log10 *P*-values are represented on the *y*-axis. Differentially expressed (DE) genes are highlighted in red, and gene symbols are shown for genes with an absolute fold change higher than 3. These comparisons considered six pre-pubertal heifers and six post-pubertal heifers.

Table 2 Transcription factors that were differentially expressed (DE) in the uterus of six pre- vs. six post-pubertal heifers and deemed significant regulators according to the regulatory impact factor metrics (RIF1 and RIF2).

Gene	Symbol	DE	*P*-value	RIF1	RIF2	Gene name
ENSBTAG00000013895	*RELA*	−0.27	1.5×10^{-2}	−3.12	−0.11	*RELA proto-oncogene, NF-kB subunit*
ENSBTAG00000014738	*HOXA11*	0.82	3.0×10^{-4}	−2.8	0.8	Homeobox A11
ENSBTAG00000031687	*ZNF146*	−0.1	3.7×10^{-2}	−2.6	−0.75	Zinc finger protein 146
ENSBTAG00000017763	*NFIL3*	1.64	1.0×10^{-4}	−2.21	1.53	Nuclear factor, interleukin 3 regulated
ENSBTAG00000027442	*NFIB*	0.3	4.4×10^{-2}	−2.2	−0.45	Nuclear factor I B
ENSBTAG00000021869	*THAP5*	0.37	1.2×10^{-2}	−2.2	1.22	THAP domain containing 5
ENSBTAG00000009975	*PBX4*	0.79	5.0×10^{-4}	−2.13	−0.32	PBX homeobox 4
ENSBTAG00000017824	*IRF8*	−0.75	1.0×10^{-6}	−2.06	0.89	Interferon regulatory factor 8
ENSBTAG00000006587	*ZNF367*	0.34	1.0×10^{-2}	−1.97	1.29	Zinc finger protein 367
ENSBTAG00000008063	*PPARA*	0.91	8.2×10^{-5}	−1.95	2.23	Peroxisome proliferator activated receptor alpha
ENSBTAG00000014692	*CREB3L4*	1.11	6.0×10^{-3}	−1.21	2.16	cAMP responsive element binding protein 3 like 4
ENSBTAG00000020441	*HMG20A*	0.19	3.0×10^{-2}	−0.58	−2.19	High mobility group 20A
ENSBTAG00000045877	*TSC22D3*	0.9	5.0×10^{-3}	−0.04	2.45	TSC22 domain family member 3
ENSBTAG00000010978	*PROX1*	−0.89	1.4×10^{-2}	0.52	2.18	Prospero homeobox 1
ENSBTAG00000021699	*RORB*	1.08	9.9×10^{-3}	0.61	2.66	RAR related orphan receptor B
ENSBTAG00000039316	*ZNF84*	−0.24	8.0×10^{-3}	0.85	2.46	Zinc finger protein 84
ENSBTAG00000013412	*NFAT5*	−0.65	3.1×10^{-2}	1.98	−0.93	Nuclear factor of activated T cells 5
ENSBTAG00000009496	*STAT5A*	−0.35	3.0×10^{-3}	2.05	−1.13	Signal transducer and activator of transcription 5A
ENSBTAG00000014949	*SP9*	3.69	1.3×10^{-3}	2.15	−1.68	Sp9 transcription factor

the uterus for the fertilization process and/or a subsequent pregnancy. It has been suggested that many genes and proteins contribute to this system to support early conceptus life (Killian 2004). This systemic view of the uterine environment is supported by our results, where over 4000 genes were found to be DE, some already known to play a role in reproduction and others being novel candidate genes. The most upregulated gene post-puberty was the *EDAR* gene, a member of the tumor necrosis factor receptor family, which has been shown to have relevance to embryonic development in mice and humans (Botchkarev & Fessing 2005; Lindfors *et al.* 2013). Other DE genes and key TFs seem to contribute to a tightly regulated robust system that underpins fertility.

Among the TFs that were also DE genes, two were zinc finger genes (ZNF): *ZNF367* and *ZNF84*. Zinc fingers are known as transcriptional repressors regulating puberty in rats and nonhuman primates (Lomniczi *et al.* 2013). The decrease of ZNF genes' expression before puberty has been reported in rats and primates (Lomniczi *et al.* 2015). In addition, an association between puberty in women and a single nucleotide polymorphism (SNP) situated near *ZNF462* and *ZNF483* has been reported (Perry *et al.* 2009; Elks *et al.* 2010; Chen *et al.* 2012; Demerath *et al.* 2013). The significance of ZNF genes has previously been reported on the transcriptomics of the reproductive axis (hypothalamus, pituitary gland, ovaries and liver) associated with puberty in these same Brahman heifers (Fortes *et al.* 2016; Nguyen *et al.* 2017, 2018). Together, the evidence suggests a major role for ZNF genes inducing puberty in Brahman heifers and other mammals.

The pathway and GO enrichment analyses for the DE genes showed that terms related to the ‘ribosome pathway’ were significant in our target list. It has long been known that the estrous cycle and estrogen signaling influences ribosome quantity and capacity of the uterus tissue (Greenman & Kenney 1964). Post-puberty, the uterus is larger, and especially during the luteal phase it is much larger, than in pre-puberty heifers. Any tissue growth requires ribosomal activity and the syntheses of multiple proteins.

It is also relevant to mention that age at puberty is an economic trait in beef cattle systems (Cushman *et al.* 2013). Enhancing understanding about biomarkers involved in puberty is an attractive possibility for selection of *Bos indicus* heifers.

Acknowledgements

We are grateful to Dr. Toan Phung for his bioinformatics assistance in the analyses of the proteomics results. L.F.Z. was funded by a post-doctoral fellowship from CONICET and a Research Endeavour Fellowship. B.L.S. holds a National Health and Medical Research Council RD Wright Biomedical (CDF Level 2) Fellowship APP1087975. M.R.S.F. was funded by a University of Queensland Postdoctoral Fellowship. L.T.N. was funded by a University of Queensland Postgraduate Scholarship.

References

Botchkarev V.A. & Fessing M.Y. (2005) Edar signaling in the control of hair follicle development. *Journal of Investigative Dermatology Symposium Proceedings* **10**, 247–51.

Bullard J.H., Purdom E., Hansen K.D. & Dudoit S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94.

Byrne K.A., Wang Y.H., Lehnert S.A., Harper G.S., McWilliam S.M., Bruce H.L. & Reverter A. (2005) Gene expression profiling of muscle tissue in Brahman steers during nutritional restriction. *Journal of Animal Science* **83**, 1–12.

Cánovas A., Rincón G., Bevilacqua C. *et al.* (2014b) Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-sequencing. *Scientific Reports* **4**, 5297.

Cánovas A., Rincón G., Islas-Trejo A., Jimenez-Flores R., Laubscher A. & Medrano J.F. (2013) RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. *Journal of Dairy Science* **96**, 2637–48.

Cánovas A., Reverter A., DeAtley K.L. *et al.* (2014a) Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PLoS ONE* **9**, e102551.

Chen C.T., Fernandez-Rhodes L., Brzyski R.G. *et al.* (2012) Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women’s Health Initiative SHARe Study. *Human Molecular Genetics* **21**, 1419–32.

Choe C., Park J.W., Kim E.S., Lee S.G., Park S.Y., Lee J.S., Cho M.J., Kang K.R., Han J. & Kang D. (2010) Proteomic analysis of differentially expressed proteins in bovine endometrium with endometritis. *Korean Journal of Physiology and Pharmacology* **14**, 205–12.

Choi M., Chang C.-Y., Clough T., Broudy D., Killeen T., MacLean B. & Vitek O. (2014) MSSTATS: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–6.

Cushman R.A., Kill L.K., Funston R.N., Mousel E.M. & Perry G.A. (2013) Heifer calving date positively influences calf weaning weights through six parturitions. *Journal of Animal Science* **91**, 4486–91.

Demerath E.W., Liu C.T., Franceschini N. *et al.* (2013) Genome-wide association study of age at menarche in African-American women. *Human Molecular Genetics* **22**, 3329–46.

Elks C.E., Perry J.R., Sulem P. *et al.* (2010) Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nature Genetics* **42**, 1077–85.

Forde N., Beltman M.E., Duffy G.B., Duffy P., Mehta J.P., O’Gaora P., Roche J.F., Lonergan P. & Crowe M.A. (2011a) Changes in the endometrial transcriptome during the bovine estrous cycle: effect of low circulating progesterone and consequences for conceptus elongation. *Biology of Reproduction* **84**, 266–78.

Forde N., Carter F., Spencer T.E. *et al.* (2011b) Conceptus-induced changes in the endometrial transcriptome: how soon does the cow know she is pregnant? *Biology of Reproduction* **85**, 144–56.

Forde N., McGettigan P.A., Mehta J.P., O’Hara L., Mamo S., Bazer F.W., Spencer T.E. & Lonergan P. (2014) Proteomic analysis of uterine fluid during the pre-implantation period of pregnancy in cattle. *Reproduction* **147**, 575–87.

Fortes M.R.S., Nguyen L.T., Weller M.M. *et al.* (2016) Transcriptome analyses identify five transcription factors differentially expressed in the hypothalamus of post-versus pre-pubertal Brahman heifers. *Journal of Animal Science* **94**, 3693–702.

Gallego Romero I., Pai A.A., Tung J. & Gilad Y. (2014) RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology* **12**, 42.

Gillet L.C., Navarro P., Tate S., Rost H., Selevsek N., Reiter L., Bonner R. & Aebersold R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular and Cellular Proteomics* **11**, O111.016717.

Giraud A., Salamonsen L., Whitley J. & Shulkes A. (1994) A peptide related to gastrin-releasing peptide is synthesized and secreted by the ovine endometrium in early-pregnancy. *Endocrinology* **135**, 2806–9.

Giraud A.S., Dumesny C., Whitley J.C., Parker L.M., Jennings I., Kemp B., Moody T.W., Sancho V., Jensen R.T. & Shulkes A. (2010) Isolation, identification and biological activity of gastrin-releasing peptide 1-46 (oGRP 1-46), the primary *GRP* gene-derived peptide product of the pregnant ovine endometrium. *Peptides* **31**, 284–90.

Greenman D.L. & Kenney F.T. (1964) Effects of alterations in hormonal status on ribosomes of rat uterus. *Archives of Biochemistry and Biophysics* **107**, 1–6.

Henning H.J. & Davis D.L. (2003) Evaluation of the uterine environment and embryos of prepubertal gilts. *Journal of Animal Science* **81**, 2575–82.

Huang da W., Sherman B.T. & Lempicki R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57.

Ibrahim S., Salilew-Wondim D., Rings F., Hoelker M., Neuhoff C., Tholen E., Looft C., Schellander K. & Tesfaye D. (2015) Expression pattern of inflammatory response genes and their regulatory micrornas in bovine oviductal cells in response to lipopolysaccharide: implication for early embryonic development. *PLoS ONE* **10**, e0119388.

Johnston D.J., Barwick S.A., Corbet N.J., Fordyce G., Holroyd R.G., Williams P.J. & Burrow H.M. (2009) Genetics of heifer puberty in two tropical beef genotypes in northern Australia and associations with heifer- and steer-production traits. *Animal Production Science* **49**, 399–412.

Killian G.J. (2004) Evidence for the role of oviduct secretions in sperm function, fertilization and embryo development. *Animal Reproduction Science* **82–83**, 141–53.

Kuscuoglu D., Janciauskiene S., Hamesch K., Haybaeck J., Trautwein C. & Strnad P. (2018) Liver–master and servant of serum proteome. *Journal of Hepatology* **69**, 512–24.

Lee S.K., Kim C.J., Kim D.J. & Kang J.H. (2015) Immune cells in the female reproductive tract. *Immune Network* **15**, 16–26.

Lindfors P.H., Voutilainen M. & Mikkola M.L. (2013) Ectodysplasin/NF-kappaB signaling in embryonic mammary gland development. *Journal of Mammary Gland Biology and Neoplasia* **18**, 165–9.

Lomniczi A., Wright H., Castellano J.M., Sonmez K. & Ojeda S.R. (2013) A system biology approach to identify regulatory pathways underlying the neuroendocrine control of female puberty in rats and nonhuman primates. *Hormones and Behavior* **64**, 175–86.

Lomniczi A., Wright H., Castellano J.M., Matagne V., Toro C.A., Ramaswamy S., Plant T.M. & Ojeda S.R. (2015) Epigenetic regulation of puberty via Zinc finger protein-mediated transcriptional repression. *Nature Communications* **6**, 10195.

Lyng R. & Shur B.D. (2009) Mouse oviduct-specific glycoprotein is an egg-associated ZP3-independent sperm-adhesion ligand. *Journal of Cell Science* **122**, 3894–906.

Mitko K., Ulbrich S.E., Wenigerkind H., Sinowatz F., Blum H., Wolf E. & Bauersachs S. (2008) Dynamic changes in messenger RNA profiles of bovine endometrium during the oestrous cycle. *Reproduction* **135**, 225–40.

Mortazavi A., Williams B.A., Mccue K., Schaeffer L. & Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–8.

Moulder R., Bhosale S.D., Goodlett D.R. & Lahesmaa R. (2017) Analysis of the plasma proteome using iTRAQ and TMT-based isobaric labeling. *Mass Spectrometry Reviews* **3**, 7.

Nguyen L.T., Reverter A., Canovas A., Venus B., Islas-Trejo A., Porto-Neto L.R., Lehnert S.A., Medrano J.F., Moore S.S. & Fortes M.R. (2017) Global differential gene expression in the pituitary gland and the ovaries of pre- and postpubertal Brahman heifers. *Journal of Animal Science* **95**, 599–615.

Nguyen L.T., Reverter A., Cánovas A. *et al.* (2018) STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in Brahman heifers. *Frontiers in Genetics* **9**, 87.

Norris B.J., Bower N.I., Smith W.J.M., Cam G.R. & Reverter A. (2005) Gene expression profiling of ovine skin and wool follicle development using a combined ovine-bovine skin cDNA microarray. *Australian Journal of Experimental Agriculture* **45**, 867–77.

Perry J.R.B., Stolk L., Franceschini N. *et al.* (2009) Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nature Genetics* **41**, 648–50.

Reverter A., Barris W., McWilliam S., Byrne K.A., Wang Y.H., Tan S.H., Hudson N. & Dalrymple B.P. (2005) Validation of alternative methods of data normalization in gene co-expression studies. *Bioinformatics* **21**, 1112–20.

Reverter A., Hudson N.J., Nagaraj S.H., Perez-Enciso M. & Dalrymple B.P. (2010) Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* **26**, 896–904.

Saha D., Koli S. & Reddy K.V.R. (2017) Transcriptional regulation of Hb-alpha and Hb-beta through nuclear factor E2-related factor-2 (Nrf2) activation in human vaginal cells: a novel mechanism of cellular adaptability to oxidative stress. *American Journal of Reproductive Immunology* **77**, e12645.

Soleilhavoup C., Riou C., Tsikis G., Labas V., Harichaux G., Kohnke P., Reynaud K., de Graaf S.P., Gerard N. & Druart X. (2016) Proteomes of the female genital tract during the oestrous cycle. *Molecular and Cellular Proteomics* **15**, 93–108.

Song G., Satterfield M.C., Kim J., Bazer F.W. & Spencer T.E. (2008) Gastrin-releasing peptide (GRP) in the ovine uterus: regulation by interferon tau and progesterone. *Biology of Reproduction* **79**, 376–86.

Wickramasinghe S., Rincon G., Islas-Trejo A. & Medrano J.F. (2012) Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* **13**, 45.

Wiktorowicz J.E. & Soman K.V. (2016) Discovery of candidate biomarkers. *Advances in Experimental Medicine and Biology* **919**, 443–62.

Xu Y., Bailey U.M. & Schulz B.L. (2015) Automated measurement of site-specific N-glycosylation occupancy with SWATH-MS. *Proteomics* **15**, 2177–86.

Yang X., Zhao Y., Yang X. & Kan F.W. (2015) Recombinant hamster oviductin is biologically active and exerts positive effects on sperm functions and sperm-oocyte binding. *PLoS ONE* **10**, e0123003.

Zacchi L.F. & Schulz B.L. (2016) SWATH-MS glycoproteomics reveals consequences of defects in the glycosylation machinery. *Molecular and Cellular Proteomics* **15**, 2435–47.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Proteins identified in the uterine fluid: list of 260 proteins identified in the uterine fluid, Uniprot accession identifiers, peptides identified, score and coverage.

Table S2 Proteins in the comparison of pre- and post-pubertal data.

Table S3 Genes differentially expressed ($P < 0.05$) in the uterus tissue.

Table S4 Results for Gene Ontology (GO) and pathway enrichment analyses.

Table S5 Regulatory impact factor metrics (RIF1 and RIF2) identified 100 significant top ranking transcription factors ($P < 0.05$) in the uterus.