



OPEN

Rewiring of gene expression in circulating white blood cells is associated with pregnancy outcome in heifers (*Bos taurus*)

Sarah E. Moorey¹, Bailey N. Walker², Michelle F. Elmore^{3,4}, Joshua B. Elmore⁴, Soren P. Rodning³ & Fernando H. Biase²✉

Infertility is a challenging phenomenon in cattle that reduces the sustainability of beef production worldwide. Here, we tested the hypothesis that gene expression profiles of protein-coding genes expressed in peripheral white blood cells (PWBCs), and circulating micro RNAs in plasma, are associated with female fertility, measured by pregnancy outcome. We drew blood samples from 17 heifers on the day of artificial insemination and analyzed transcript abundance for 10,496 genes in PWBCs and 290 circulating micro RNAs. The females were later classified as pregnant to artificial insemination, pregnant to natural breeding or not pregnant. We identified 1860 genes producing significant differential coexpression (eFDR < 0.002) based on pregnancy outcome. Additionally, 237 micro RNAs and 2274 genes in PWBCs presented differential coexpression based on pregnancy outcome. Furthermore, using a machine learning prediction algorithm we detected a subset of genes whose abundance could be used for blind categorization of pregnancy outcome. Our results provide strong evidence that transcript abundance in circulating white blood cells is associated with fertility in heifers.

Cattle production systems provide approximately 28%¹ of the global protein supply. Improving cattle production efficiency is essential for farmers to attain sustainable production and support the growing demand for animal protein¹. Infertility is a critical problem in agriculturally important large animals. For instance, the prevalence of infertility ranges from 64 to 95%, averaging 15% in beef heifers^{2–14}. First breeding success greatly influences the lifetime efficiency of beef replacement heifers. Heifers that calve early in their first calving season experience increased productivity and longevity than their later calving herd mates^{15–18}. Furthermore, the genetic correlation between yearling pregnancy rate and lifetime pregnancy rate is high (0.92–0.97)^{19,20}. Therefore, the ability to identify heifers that experience optimal fertility during the first breeding is essential to the sustainability of beef cattle production systems. The genetic architecture of fertility in beef heifers is very complex^{21–28}. In line with this complexity, female reproductive traits have low heritability, with traits related to female reproductive fitness in cattle, such as heifer pregnancy and first service conception, range from 0.07 to 0.20^{2,21,29–33} and 0.03 to 0.18^{2,21,29}, respectively. The examination of the genetic components of fertility in beef heifers have yielded several genes potentially associated with fertility traits^{21–28}, but the effect of these markers are minimal, and there is no clear redundancy in genetic markers identified across breeds or subspecies.

Beyond the genomic profiling, the analysis of multiple layers of an individual's molecular blueprint is likely key for understanding the underlying biology of complex traits³⁴. In line with this rationale, expression-trait association studies have emerged as a means to better understand complex traits^{35,36}. Specifically related to fertility, there is evidence that circulating micro RNA (miRNA) profiles are indicative of reproductive function³⁷. Furthermore, we have previously identified differences in the transcriptome of peripheral white blood cells (PWBC) among heifers of different fertility potential³⁸.

¹Department of Animal Science, University of Tennessee, 2506 River Drive, Knoxville, TN 37996, USA. ²Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, 175 West Campus Drive, Blacksburg, VA 24061, USA. ³Department of Animal Sciences, Auburn University, 107 Comer Hall, Auburn, AL 36849, USA. ⁴Alabama Cooperative Extension System, 107 Comer Hall, Auburn, AL 36849, USA. ✉email: fbias@vt.edu

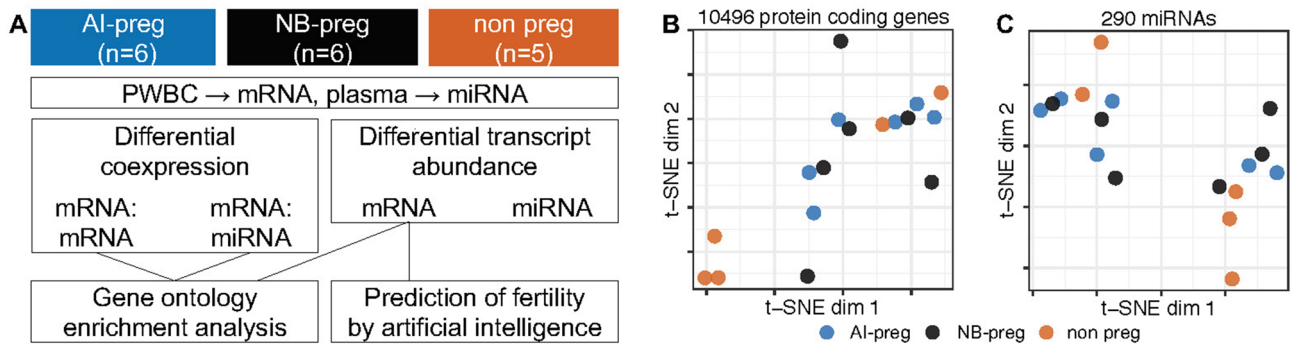


Figure 1. Overview of the experimental design and data. (A) Schematics of the experimental design, classification scheme based on fertility, data produced and overview of the analyses. (B,C) Dimensionality reduction of the transcriptome data obtained from the mRNA abundance of protein-coding genes in PWBCs and plasma extracellular miRNA.

In the present study, our objective was to investigate gene transcripts in PWBCs and circulating micro RNA (miRNA) profiles in heifers enrolled in fixed time artificial insemination (AI). We utilized a multi-dimensional approach to investigate differences in coexpression of protein-coding genes in PWBCs, and coexpression between circulating miRNA and mRNAs in PWBCs.

Results

Overview of the experiment and data generated. We generated transcriptome data for mRNA from PWBC and small RNA from plasma collected at the time of AI from heifers that became pregnant to AI (AI-preg, $n=6$), pregnant later in the breeding season from natural breeding with bulls NB (NB, NB-preg, $n=6$), or failed to become pregnant (non-pregnant, $n=5$, Fig. 1A). We produced, on average, 25.8 and 7.1 million reads from mRNA and small RNA, respectively. Following the filtering for lowly expressed genes, 10,496 protein coding genes in PWBC and 290 miRNAs in plasma samples were used for differential gene expression, coexpression and prediction analyses (Fig. 1A). Notably, three of the five non-pregnant samples clustered separately from other subjects based on the expression levels of protein-coding genes (Fig. 1B). We note that all heifers were phenotypically similar; and no marked distinction was observed among these three heifers relative to the remainder of the group on the day of AI. Contrariwise, there was no separation of samples based on the miRNA expression levels (Fig. 1C).

Differential mRNA:mRNA coexpression associated with pregnancy outcome. First, we asked whether there was a general coexpression profile of the protein-coding genes expressed in PWBCs amongst all 17 samples. We identified that 1633 genes that formed 2017 pairs with highly significant correlated expression levels ($|r| > 0.98$, $FDR < 0.02$, Fig. 2A, Supplementary Table S1). Notably, ~99% of those significant correlations were positive (Fig. 2A). Several genes showing coexpression were associated with ‘translation’, ‘positive regulation of gene expression’, ‘B cell differentiation’, and ‘negative regulation of extrinsic apoptotic signaling pathway’ amongst other biological processes ($FDR < 0.1$, Supplementary Table S2). This result indicates the occurrence of gene regulatory networks in PWBCs that are biologically meaningful for the immune system.

Next, we investigated if the pair-wise gene coexpression differs between animals that became pregnant relative to those that did not become pregnant. We identified 1860 genes with significant differential correlation based on the pregnancy outcome ($eFDR < 0.002$, Fig. 2B,C, Supplementary Table S3, Supplementary Fig. S1). Of these 1860 genes, 963 genes formed 716 negative, 1053 genes formed 708 positive, four genes lost two positive, and 40 genes formed 23 inverted coexpression connections in non-pregnant heifers relative to both AI-preg and NB-preg heifers. Among the 963 genes that formed new negative coexpression connections in non-pregnant heifers, there was enrichment of the biological processes “proteolysis” ($n=38$ genes) and “translation” ($n=52$ genes, $FDR < 0.1$, see Fig. 2D for selected genes and Supplementary Table S4 for list of genes). Five of the 40 genes that formed inverted coexpression connection in not-preg heifers relative to the pregnant counterparts were associated with “negative regulation of transcription by RNA polymerase II” ($FDR < 0.1$, see Fig. 2E for selected genes and Supplementary Table S5 for a list of genes). These results indicate that physiological factors associated with fertility or infertility act on the transcription of protein-coding genes in PWBCs.

Differential miRNA:mRNA coexpression associated with pregnancy outcome. The generation of data from small RNA (from plasma) and mRNA (from PWBCs) from the same subjects ($n=17$) permitted us to ask whether circulating miRNAs have correlated expression with protein-coding genes expressed in the PWBCs. We identified 141 pairs of miRNA:mRNA with $|r| > 0.85$ ($eFDR < 0.0001$, Supplementary Fig. S2) that were formed by 106 protein-coding genes and 33 miRNAs (Supplementary Table S6). Among the 106 protein-coding genes existing in high correlation with miRNA, we identified three significantly enriched biological processes, namely: “peptidyl-threonine phosphorylation”, “cilium assembly” and “regulation of gene expression” (Fig. 3A, Supplementary Table S7). Of notice, 30 of the 141 miRNA:mRNA pairs that presented negative correlation between the expression levels have been previously identified as interacting pairs in the miRWalk data-

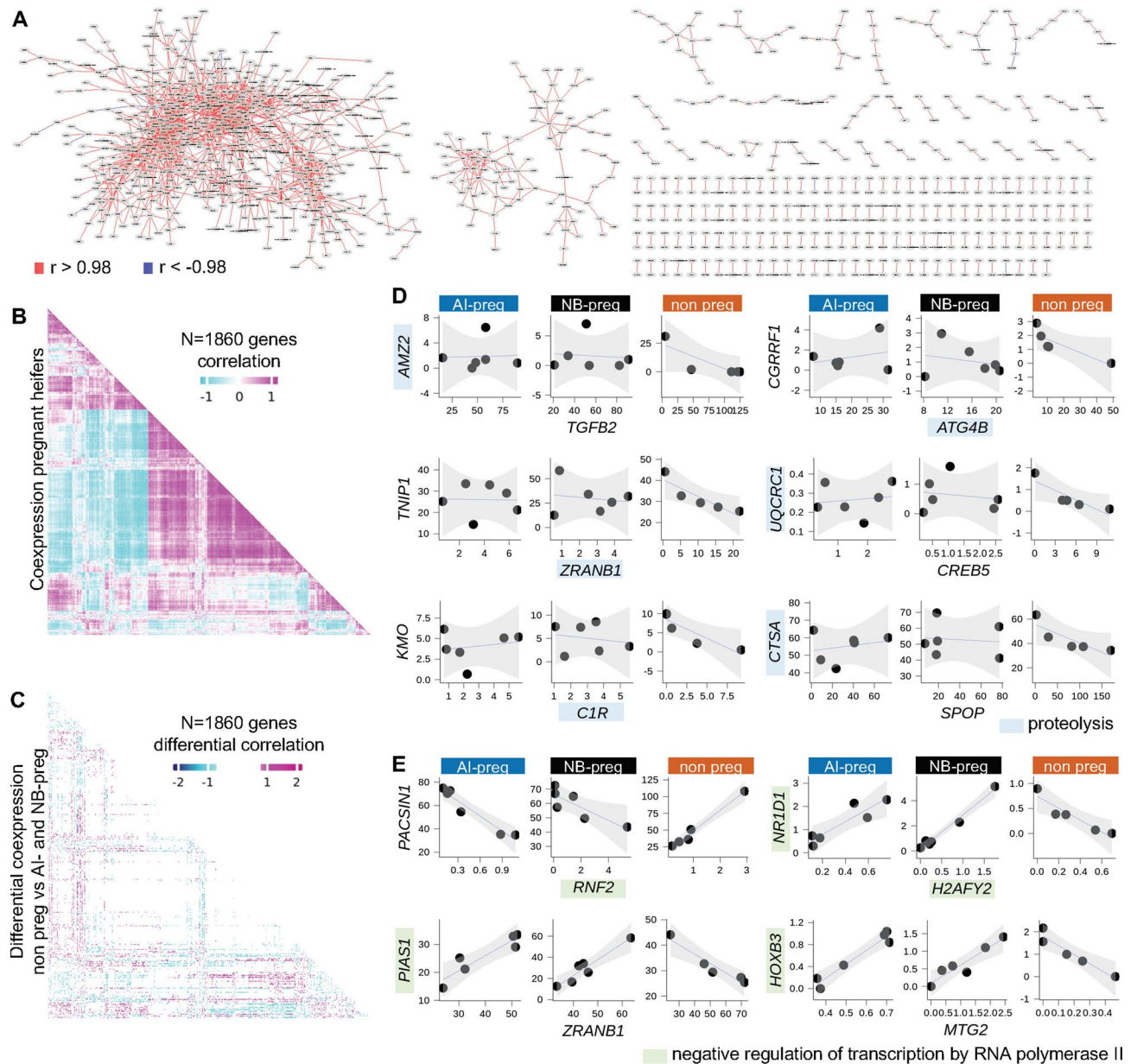


Figure 2. mRNA:mRNA coexpression in PWBCs. (A) Network of genes with highly correlated mRNA abundance. (B) Coexpression of 1860 protein-coding genes in pregnant heifers (AI-preg and NB-preg). (C) Profile of the differential coexpression in heifers classified as non-pregnant relative to the counterparts that became pregnant. (D) Scatterplot of selected genes associated with “proteolysis” that were coexpressed in heifers classified as non-pregnant relative to the counterparts that became pregnant. (E) Scatterplot of selected genes associated with “negative regulation of transcription by RNA polymerase II” that showed inverted coexpression in heifers classified as non-pregnant relative to the counterparts that became pregnant. All scatterplots present transcript abundance in $\text{Log}_2(\text{counts per million})$.

base (Fig. 3B). Additionally, bta-mir-744 and bta-mir-320a emerged as the two most connected miRNAs (13 and 6 connections, respectively). It is evident that circulating miRNAs have regulatory roles over transcripts of protein-coding genes expressed in the PWBCs.

We also interrogated miRNA:mRNA pairs for differential coexpression associated with pregnancy outcome. We identified 2330 and 2738 miRNA:mRNA pairs with gain of negative and positive correlated expression, respectively, ($|r| > 0.9$, $\text{eFDR} \leq 0.01$, Supplementary Fig. S3) in non-pregnant heifers that were not identified compared to pregnant heifers (both AI-preg and NB-preg, Fig. 3C). In addition, there were 51, two and three connections showing inverted, loss of negative and loss of positive correlations in non-pregnant heifers relative to the pregnant (both AI-preg and NB-preg) counterparts (Fig. 3C, see Supplementary Table S8 for genes). Among the 1220 protein-coding genes associated with gain of negative correlation with miRNAs, there were biological processes significantly enriched, namely: “immunoglobulin mediated immune response”, “negative regulation

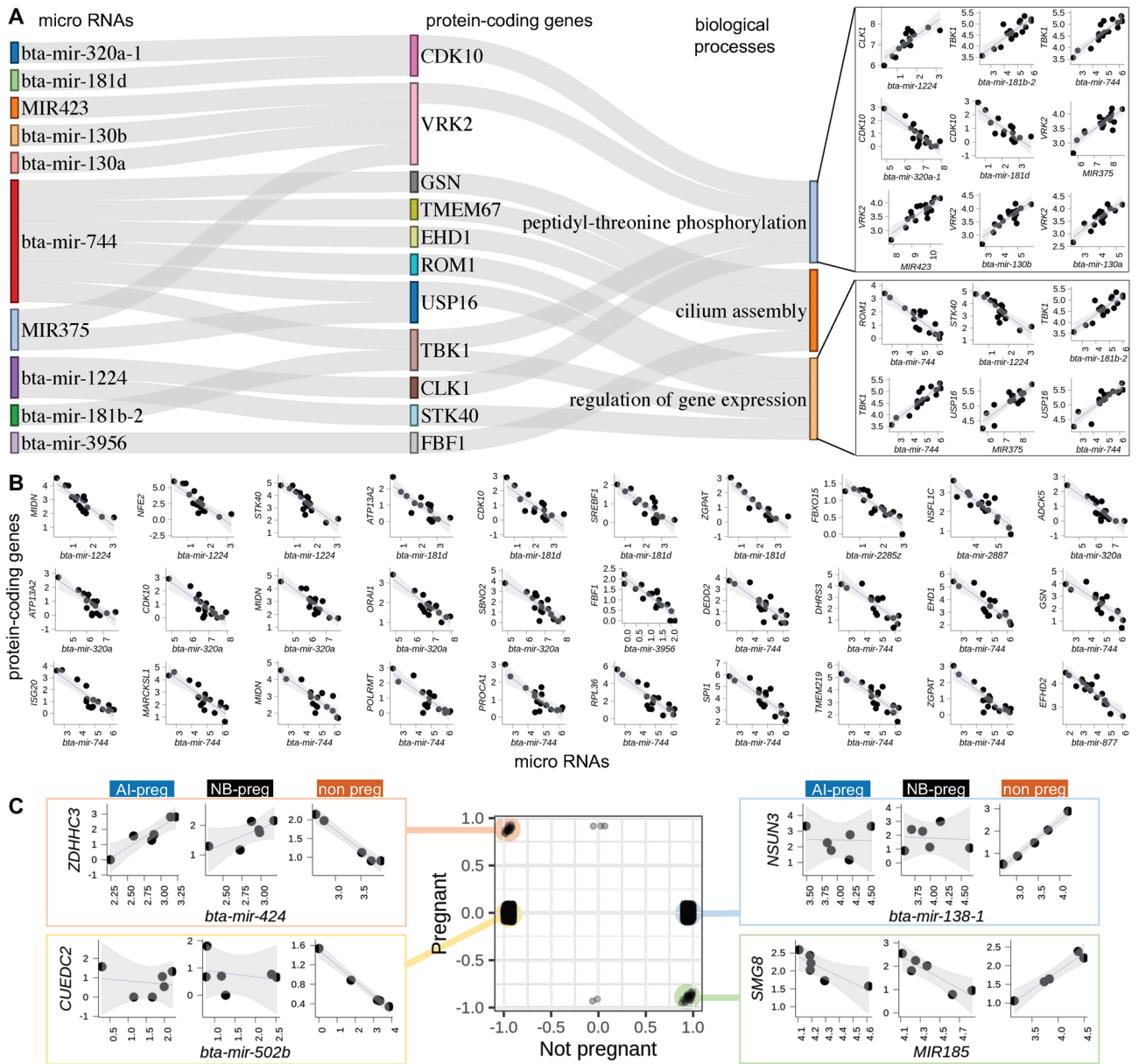


Figure 3. miRNA from plasma coexpressed with mRNA in PWBCs. **(A)** miRNA:mRNA pairs significantly coexpressed whose genes are significantly enriched in specific biological processes. **(B)** miRNA:mRNA pairs significantly coexpressed with evidence of interaction from miRBase. **(C)** The center chart presents pairwise gene Pearson's correlation for the corresponding group. The scatterplots to each side show selected miRNA:mRNA pairs representing differential coexpression between miRNA:mRNA in heifers classified as non-pregnant relative to the counterparts that became pregnant. All scatterplots presenting transcript abundance are in Log₂(counts per million).

of DNA-binding transcription factor activity”, “lymphocyte chemotaxis” and “mitochondrial respiratory chain complex I assembly” (FDR < 0.1, Supplementary Table S9). Among the 1288 genes associated with gain of positive correlation with miRNAs, the following biological processes were significantly enriched: “regulation of DNA replication” and “cellular response to BMP stimulus” (FDR < 0.1, Supplementary Table S10). These results indicate that several miRNA and mRNA pairs have their connections altered in relation to the heifer’s fertility potential.

Differential gene expression associated with pregnancy outcome. Differential gene expression analysis revealed 67 (55 up- and 12 down-regulated) and 81 (63 up- and 18 down-regulated) protein-coding genes with altered expression profiles between PWBC of AI-preg versus non-pregnant, and NB-preg versus non-pregnant heifers, respectively (eFDR ≤ 0.05, Fig. 4A, Supplementary Tables S11 and S12, Supplementary Fig. S4). Notably, 26 genes were common between both pregnant groups versus the non-pregnant group (Fig. 4B). Gene ontology and pathway analyses of differentially expressed genes between AI-preg and Non-pregnant heifers revealed significant enrichment for the terms “signal transduction” (*ARAP3*, *LSP1*, *NFAM1*,

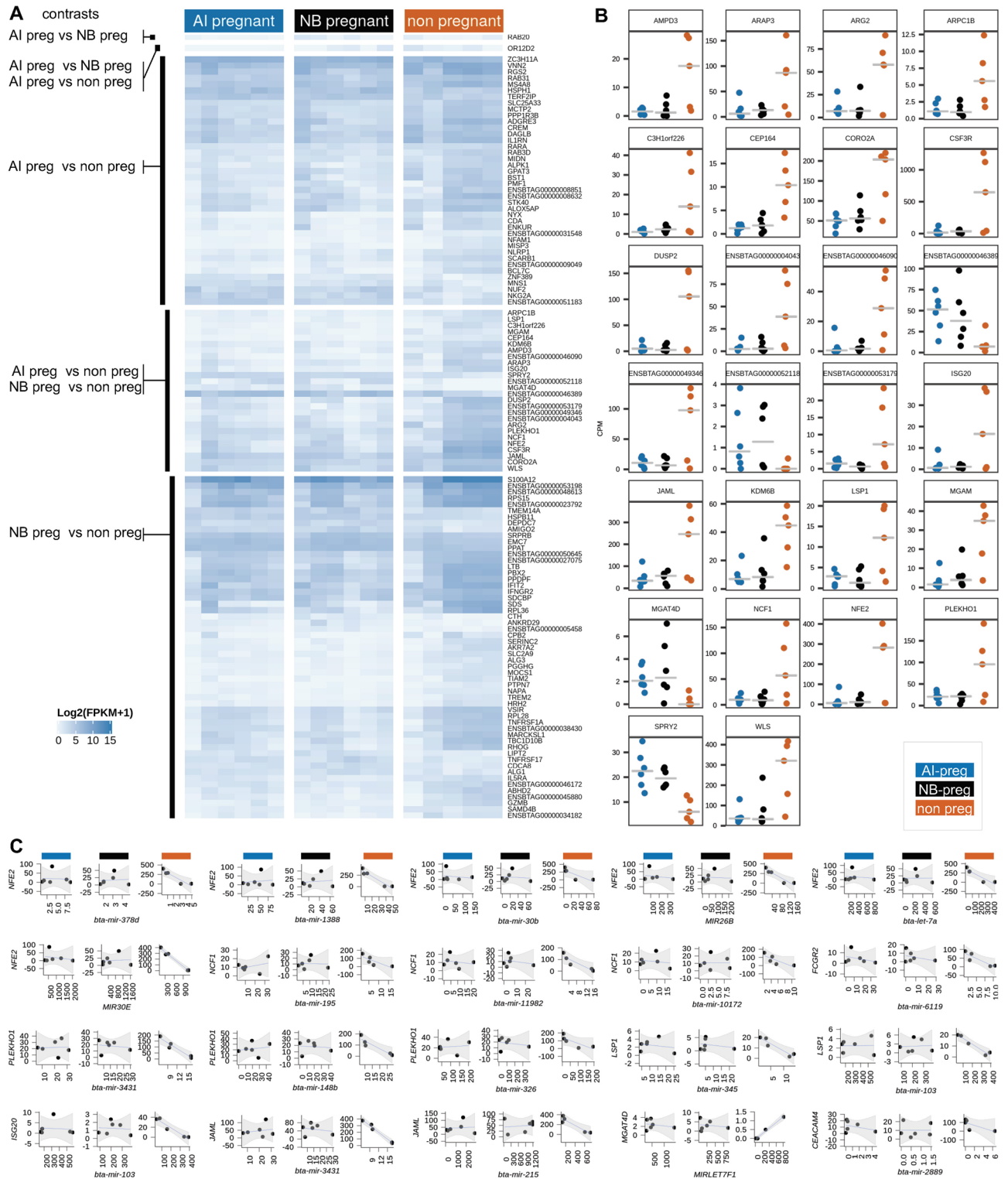


Figure 4. Differential coexpression among heifers of different reproductive outcome. **(A)** Heatmap depicting the expression levels of genes differentially expressed according to specific contrasts. **(B)** Transcript abundance in counts per million of the genes differentially expressed on both contrasts: AI-preg versus non-pregnant and NB-preg versus non-pregnant. **(C)** miRNA:mRNA pairs significantly coexpressed whose protein-coding genes are differentially expressed in PWBCs based on pregnancy outcome. All scatterplots present transcript abundance in Log2(counts per million).

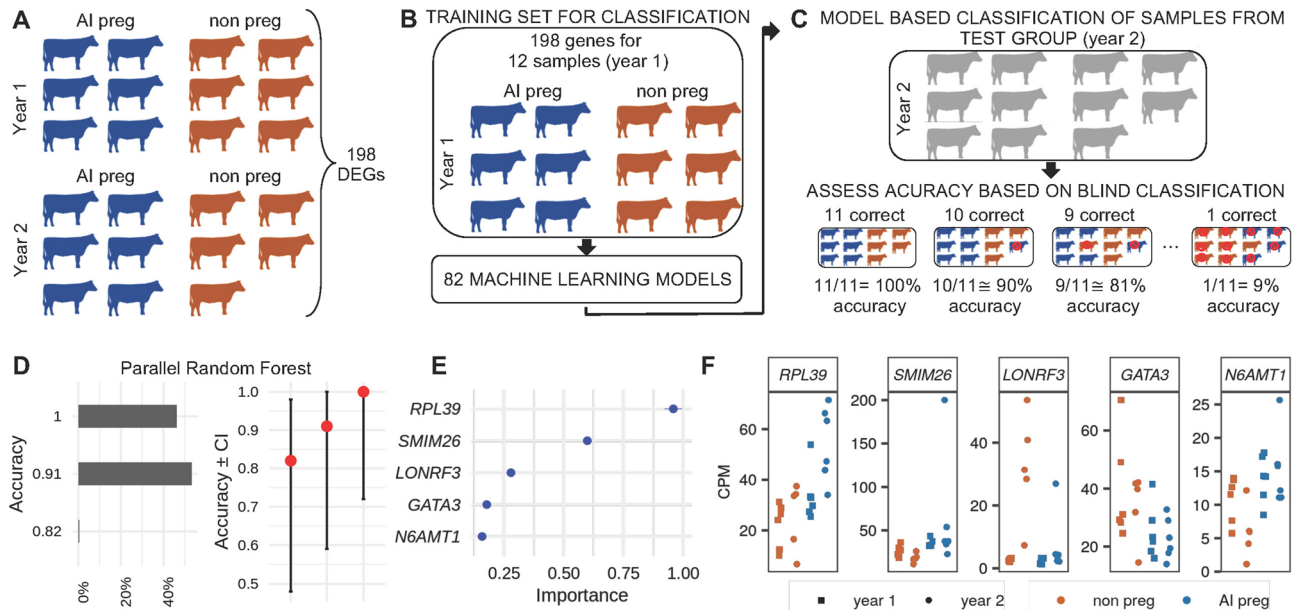


Figure 5. Prediction of pregnancy outcome based on mRNA levels and machine learning algorithms. **(A)** Representation of the contrast between gene expression levels of heifers classified as AI-preg and non-pregnant from two years, and the number of differentially expressed genes (DEGs). **(B)** Depiction of the training set used as input for the modeling of the transcriptome dataset by multiple machine learning algorithms. **(C)** Depiction of the data used as the test set for prediction of pregnancy outcome, and the representation of the calculation of accuracy. **(D)** Accuracy of blind pregnancy prediction based on the algorithm “parallel random forest” in 2000 trials. **(E)** Variable importance for the top five genes. **(F)** Transcript levels for the top five genes most informative for the machine learning algorithm to classify pregnancy outcome.

OR12D2, *RARA*; $FDR < 0.05$). When differentially expressed genes between NB-preg and non-pregnant heifers were analyzed, significant enrichment was determined for the terms “signal transduction” (*ARAP3*, *HRH2*, *LSP1*, *TIAM2*, *TNFRSF17*, *TNFRSF1A*; $FDR = 0.062$) and the KEGG Pathway “cytokine-cytokine receptor interaction” (*IFNGR2*, *IL5RA*, *LTB*, *TNFRSF17*, *TNFRSF1A*; $FDR = 0.0045$). Plasma miRNA profiles were very similar among samples from all pregnancy outcomes with the exception of miR-11995 that was more abundant in AI-preg versus non-pregnant heifers. These results indicate that quantitative differences exist in transcript abundance of protein-coding genes expressed in PWBC relative to the pregnancy outcome.

Next, we asked if differential gene expression could be associated with differential miRNA:mRNA coexpression. An intersection of the results identified 12 differentially expressed genes (10 annotated genes) in non-pregnant heifers relative to the AI-preg and NB-preg groups that also gained coexpression with 27 miRNAs (Fig. 4C). Notably, all genes that had higher expression in non-pregnant heifers (*CEACAM4*, *FCGR2*, *ISG20*, *JAML*, *LSP1*, *NCF1*, *NFE2*, *PLEKHO1*) showed a gain in negative coexpression with specific miRNAs. By contrast *MGAT4D* was more lowly expressed in non-pregnant heifers, and gained a positive coexpression with a miRNA in non-pregnant heifers.

Assessment of transcript levels as predictors of pregnancy outcome. Motivated by the observations that several genes have their abundance altered in relation to the fertility potential pregnancy (Fig. 4A), we asked whether transcript abundances could serve as predictive features across datasets. We previously generated gene expression values for a similar dataset (GSE103628³⁸), which also consisted of transcriptome data from PWBCs collected from heifers at the time of AI, and were classified according to their pregnancy outcome as AI-preg ($n = 6$) or non-pregnant ($n = 6$). This dataset was particularly fitting for this interrogation because the sampling occurred at the same site in the year prior. We combined the expression values of the previously produced dataset (year one) with those of the current dataset (year two) and contrasted the transcript abundance between heifers that were categorized as AI-preg versus non-pregnant accounting for the different years as a fixed effect in the model. This analysis identified 198 genes whose abundance were significantly different between the two experimental groups (AI-preg, non-pregnant, $FDR < 0.03$, Fig. 5A, Supplementary Table S13).

We used the transcript abundance values for these 198 genes obtained from samples collected in year one as input (training set) for machine learning algorithms to derive models of prediction to discriminate pregnancy outcome (AI-preg or non-pregnant, Fig. 5B). Next, we used the transcript abundance values for these 198 genes obtained from samples collected in year two as input to the algorithms for a blind classification of pregnancy outcome (AI-preg or non-pregnant, Fig. 5C). The accuracy of classifications was assessed based on the known pregnancy outcome, and parallelized random forest was the algorithm that produced the greatest proportion of correct classification of all 11 heifers tested (926 out of 2000 predictions, 46.3%), followed by 1062 out of 2000 predictions with correct classification for 10 out of 11 heifers (53.1%, Fig. 5D). Notably, the top five genes that

showed the largest degree of importance for the classification were *RPL39*, *SMIM26*, *LONRF3*, *GATA3*, *N6AMT1* (Fig. 5E), of which, *RPL39*, *SMIM26*, *N6AMT1* were up-regulated and *LONRF3* and *GATA3* were down-regulated in AI-preg heifers relative to the non-pregnant group (Fig. 5F).

Discussion

Here, we report a multilayered transcriptome analysis of transcripts from blood sampled on the day of AI. The main objective of our study was to understand gene regulatory networks³⁹ occurring in PWBCs, and rewiring that can occur based on the heifers' fertility potential. We focused on the day of artificial insemination because it is the last day prior to the fertilization, when the heifer has the first opportunity to become pregnant. We interrogated mRNA and miRNA abundance levels obtained from the same individuals using the coexpression framework, which is a valuable systems biology approach to gain insights into the biology of complex traits³⁹. Furthermore, using machine learning⁴⁰, we detected consistent altered gene transcript abundance associated with infertility across independent datasets. Collectively, our results provide important aspects of regulatory networks in PWBCs that shed light on the physiological status of an animal and its preparedness for becoming pregnant.

We must note that the empirical results reported in this study must be considered in light of some limitations. Here, we evaluated transcripts present in a subset of cells (white blood cells) in one tissue (blood). It is important to acknowledge that the data generated does not capture the totality of the complexity involving infertility, nor the multitude of factors influencing infertility. Furthermore, this is a snapshot of one day in the heifer's life which we are using to make inferences on the first breeding season. Thus, the results presented here cannot be extrapolated to the consecutive breeding seasons in the heifer's life. Finally, the sample size does not capture all the variability existent within cattle. Thus, we do not generalize the results beyond the scope of this research.

Similar to work reported in humans⁴¹, our gene network analysis focusing on mRNAs detected several protein-coding genes with strong coexpression in PWBCs. Interestingly, nearly all of the correlated expression inferred as statistically significant across all heifers ($|r| > 0.98$, $FDR < 0.02$) were positive. As expected⁴², in silico functional gene co-expression analysis indicated groups of genes that are linked to specific biological processes. A contrast of coexpressing pairs of genes between samples collected from non-pregnant heifers and the two groups of heifers that became pregnant displayed a large number of connections present in non-pregnant heifers that were not identified in either AI-preg or NB-preg heifers. Our strict criteria ($|r| > 0.99$, $eFDR < 0.002$) utilized to identify these altered connections indicate functional regulatory relationships between selected genes.

Surprisingly, a great majority of the altered connections between gene pairs present were gained in non-pregnant heifers, while there was minor loss of significant connections in non-pregnant heifers relative to the AI-preg and NB-preg groups. The enrichment for specific biological processes of genes that formed new negative connections was clear evidence of the biological relevance of these new connections. Notably, among genes annotated with proteolytic activity, *ADAM19*⁴³, *ASRGL1*⁴⁴, *ATG4B*⁴⁵, *DPEP2*⁴⁶, *IMMP2L*⁴⁷, *PSMB9*⁴⁸, *SERPINE2*⁴⁹, *UCHL3*⁵⁰, *USP42*⁵¹ have been implicated in female infertility. A more drastic change in patterns of coexpression was observed with the inversion of gene connections (significant pairwise coexpression in one group with significant coexpression with opposite direction in another group). These gene circuits changing from activation to repression are explained by current models of incoherent feed-forward loops⁵². Cumulatively, the large occurrence of new connections and the inverted connections created in PWBCs exclusively in non-pregnant heifers indicate that gene regulation is altered as an adaptive response to a stimulus.

Our profiling of small RNAs from plasma identified 263 annotated miRNAs, out of which 131 and 122 were previously identified in sera and circulating exosomes, respectively, among other cattle organs⁵³. Only 38 annotated miRNAs were profiled in peripheral blood mononucleated cells in humans⁵⁴, which demonstrates that our plasma samples were depleted of circulating white blood cells.

Immuno-related cells composing the PWBC population can uptake circulating small RNAs⁵⁵. Although our data did not allow us to trace the origin of the tissue that produced the miRNAs, we detected 33 circulating miRNAs whose abundance are associated with the abundance of 106 protein-coding genes. As reported in humans⁵⁶, we quantified positively and negatively correlated abundances between miRNA and protein-coding genes. Negative correlations mostly occur due to the targeting of mRNA by a specific miRNA. Indeed, we found several pairs of miRNA:mRNA negatively correlated that had been identified as interacting pairs that fit this model. By comparison, the positive coexpression between miRNA:mRNA pairs can be explained by a miRNA targeting an intermediate gene, which in turn is negatively associated with a third gene, thereby creating a positive correlation between a miRNA:mRNA pair (see Fig. 3 in reference⁵⁷). In our results, we only identify examples of miRNA:mRNA pairs that follow the intermediate action of miRNAs when relaxing the stringent criteria for inference of significant coexpression. Regardless of the model of interaction, it was notable that many protein-coding genes that were coexpressed with miRNAs were enriched for specific biological processes, including "regulation of gene expression". These results add further evidence to the growing body of literature suggesting that circulating miRNAs act as mediators of cellular function^{58,59}.

We also tested whether circulating miRNA and PWBC mRNA pairs have differential coexpression relative to the pregnancy outcome. Following the same trend observed for miRNA:mRNA coexpression in PWBCs, there were 1000-fold more miRNA:mRNA connections that were present in heifers that did not become pregnant compared to the loss in coexpression. Results from gene enrichment analysis support the notion that genes forming new miRNA:mRNA coexpression connections in non-pregnant heifers function in the immune system, although our data do not permit inferences on whether particular immune processes were more or less functional in non-pregnant heifers.

Critical aspects of female fertility are tied to an appropriate regulation of the immune system. Maternal immune tolerance to pregnancy starts with the exposure to proteins in the semen⁶⁰, and is carried throughout placentation and pregnancy establishment⁶¹. Notably, 24 out of 26 differentially expressed genes between pregnant

(AI-preg and NB-preg) and non-pregnant heifers showed greater abundance in non-pregnant heifers relative to the other two groups. (AI-preg and NB-preg). While we did not detect differential abundance of miRNAs relative to pregnancy outcome, the gain of coexpression between miRNAs and mRNAs from differentially expressed genes functionally linked with immunology is in line with regulatory roles that certain miRNAs exert on the immune system⁶². Specifically, bta-miR-30b⁶³, bta-miR-148b, bta-miR-195⁶⁴, bta-miR-326⁶⁵, bta-miR-378d⁶⁶, bta-miR-1388⁶³, bta-miR-2889⁶⁶, bta-miR-11982⁶⁷ have been reported to be differentially expressed in pathogen infected cells or tissues relative to control counterparts. This result lead us to hypothesize that the heifers in the non-pregnant group have a heightened activity of specific cells or functions in the immune system relative to the heifers that became pregnant. It remains unclear, however, if this possible higher activity was stimulated by extrinsic factors or is within the naturally-occurring variation⁶⁸.

The results of differential gene expression and differential coexpression fostered a hypothesis that transcript abundance in the PWBCs can be used as predictors of pregnancy outcome. We tested this hypothesis by applying machine learning algorithms on two groups of heifers that were bred in 2015 (year one) and 2016 (year two). Parallel random forest emerged as the algorithm with over 90% efficiency of classification nearly all trials executed, which confirms the potential of accurate classification of samples using RNA-seq data under the case-control framework^{69,70}. The results show that while not one single gene emerges as a potential biomarker, the accumulated information of transcript abundance from multiple genes can be powerful for the identification of fertility potential in cattle.

Although we detected altered coexpression with high confidence, and in some instances, with significant enrichment of genes in specific gene ontology terms, the consequence of the differential coexpression on those biological functions remains undetermined. In addition, it remains unclear whether the altered patterns of gene expression profiles are connected to immune functions that are related to infertility⁷¹. The data analyzed here does not allow us to infer causality⁷², however our results converge to a strong indication that differences in gene expression profiles are associated⁷² with fertility in cattle.

In conclusion, analyses of mRNA from PWBCs and circulating extracellular miRNA by multiple approaches (differential gene expression, differential gene coexpression and machine learning classification) provided complementary evidence that gene expression profiles in the PWBCs are associated with the fertility potential in beef heifers. The overall difference detected in transcript abundance for specific genes is reproducible across samples and show promise for the development of molecular features indicating fertility potential in beef heifers. The analysis of transcriptome data is critical for the understanding of economically important complex traits⁷³; and our findings show that transcript levels may be relevant for selection of animals more likely to contribute to a sustainable beef production system.

Methods

All procedures with animals were approved by Institutional Animal and Care and Use Committee at Auburn University. All methods related to animal handling were performed in accordance with the guidelines stipulated by the Institutional Animal and Care and Use Committee at Auburn University and the Guide for the Care and Use of Laboratory Animals.

Heifer classification according to pregnancy outcome. Crossbred heifers ($n = 32$, Angus-Simmental, *Bos taurus taurus*) were developed as replacement heifers^{74,75} at the Black Belt Research and Extension Center (Auburn University). All animals were exposed to the same vaccination protocol which included an injection of Ultraback 7 and Bovi-Shield Gold FP 5 L5 at approximately 6 months of age followed by a second injection of each product at approximately 9 months of age. A final injection of Bovi-Shield Gold FP 5 VL5 HB was administered to all heifers at approximately 12 months of age. Thirty-four days prior to AI, heifers were evaluated for reproductive tract score (scale of 1–5⁷⁶) and body condition score (BCS; scale of 1–9⁷⁷) for the assertion that morphological and physiological conditions were suitable for breeding. One animal was eliminated from the breeding herd at this time due to a reproductive tract score of 2. The remaining animals presented body condition score 6 and reproductive tract score 3 ($n = 9$), 4 ($n = 14$) or 5 ($n = 8$).

Thirty-one animals, 14 months of age on average, underwent estrous synchronization for fixed-time artificial insemination utilizing the 7-Day Co-Synch + CIDR protocol⁷⁸. An experienced veterinarian inseminated all heifers with semen from one sire of proven fertility (Deer Valley All In, Select Sires). Fourteen days after insemination, heifers were exposed to two fertile bulls for natural breeding for 60 days. An experienced veterinarian performed pregnancy evaluation by transrectal palpation on days 69 and 120 post artificial insemination. Presence or absence of a conceptus, alongside morphological features indicating fetal age were used to classify heifers as pregnant to AI (AI-pregnant), pregnant to natural breeding (NB-pregnant), or non-pregnant. At the end of the breeding season, 16, nine and six heifers were classified as AI-pregnant, NB-pregnant or non-pregnant, respectively.

Collection, processing and preservation of biological material. Immediately after AI, 10 ml of blood was drawn from the jugular vein into vacutainers containing 18 mg K2 EDTA (Becton, Dickinson and Company, Franklin, NJ). The tubes were inverted to prevent blood coagulation and immersed in ice for approximately 4 h. Plasma and buffy coat were separated by centrifugation for 10 min at $2000 \times g$ at 4 °C. Plasma was removed (~1.5 ml) and centrifuged 10 min at $500 \times g$ at 4 °C for further removal of cells. The plasma devoid of cells was then transferred to a new microcentrifuge tube and preserved at –80 °C. The buffy coat was removed and deposited into 14 ml of red blood cell lysis solution (0.15 M ammonium chloride, 10 mM potassium bicarbonate, 0.1 mM EDTA, Cold Spring Harbor Protocols) for 10 min at room temperature (24–25 °C). After a cen-

trifugation, 800 × g for 10 min, the solution was discarded and the pellet containing PWBCs was re-suspended in 200 µl of RNAlater (Lifetechnologies, Carlsbad, CA). PWBC samples were stored at –80 °C.

RNA extraction, library preparation, and RNA sequencing. Total RNA was isolated from PWBCs of 18 heifers [AI-pregnant (n = 6), NB-pregnant (n = 6), and non-pregnant (n = 6)] using TRIzol reagent (Invitrogen, Carlsbad, CA)^{79,80} following the manufacturer's protocol. RNA yield was quantified using the Qubit RNA Broad Range Assay Kit (Eurogene, OR) on a Qubit Fluorometer, and integrity was assessed on an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA) using an Agilent RNA 6000 Nano kit (Agilent, Santa Clara, CA). RNA was submitted for sequencing if RNA integrity number was >7. Approximately 500 µg of total RNA was used as input for the TruSeq Stranded mRNA Library Prep kit (Illumina, Inc., San Diego, CA); and library preparation was carried out following manufacturer's instructions. Libraries were quantified with the Qubit dsDNA High Sensitivity Assay Kit (Eurogene, OR); and quality was evaluated using the High Sensitivity DNA chip (Agilent, Santa Clara, CA) on an Agilent 2100 Bioanalyzer. Libraries were sequenced in a HiSeq2500 system at the Genomic Services Laboratory at HudsonAlpha, Huntsville, AL to generate 125nt long, pair-end reads.

Total RNA (circulating extracellular RNA) was isolated from 1 ml of plasma by adding 9 ml of TRIzol™ reagent and following manufacturer's protocols through organic phase separation. The aqueous phase containing the RNA was then collected and purified with the RNA Clean and Concentrator-5 kit (ZymoResearch, Irvine, CA), according to manufacturer's instructions. Total RNA yield was quantified in test samples using the Qubit RNA Broad Range Assay Kit averaging 21.4 ng, whereas samples prepared for sequencing were not quantified and all of the eluted RNA was directly used for library preparation.

Total RNA obtained from plasma was desiccated to a volume of 2.5 µl for library preparation with the TruSeq Small RNA Library Prep kit (Illumina, Inc., San Diego, CA) following manufacturer's instructions, with the minor alteration that reagent volumes were halved in all reactions until reverse transcription was performed, as described elsewhere⁸¹. After PCR amplification, libraries were loaded onto a 10% polyacrylamide gel⁸² and electrophoresis was carried out for three hours at 100 V alongside a 25/100 bp mixed DNA ladder (Bioneer, Alameda, CA); and the custom DNA ladder included in the TruSeq Small RNA Library Prep kit. The gel was soaked in GelRed Nucleic Acid Gel Stain (Biotium, Inc., Fremont, CA) for DNA staining. DNA bands in the range of 145–160 base pairs were isolated from the gel by the crush-and-soak method⁸³ with the following modifications. Gel fragments were placed into a 0.5 ml tube containing small holes at the bottom, mounted into a 2 ml tube. The tubes were centrifuged at 14,000xg for two minutes at room temperature. The crushed gel was soaked in 200 µl of water at 4 °C overnight, followed by removal of the solution containing the library to a new tube.

Concentration of each library was determined by Qubit dsDNA High Sensitivity Assay Kit, and the volume corresponding to 1.5 ng of each library was pooled for further purification with the DNA Clean and Concentrator-5 (ZymoResearch, Irvine, CA) following the manufacturer's instructions. The pool was quantified with Qubit dsDNA High Sensitivity Assay Kit, and the DNA profile was evaluated using the High Sensitivity DNA chip on an Agilent 2100 Bioanalyzer. Libraries were sequenced in a NextSeq system at the Genomic Services Laboratory at HudsonAlpha, Huntsville, AL to generate 75nt long, single end reads.

Raw sequences and metadata are deposited in the Gene Expression Omnibus database (GSE146041).

RNA sequencing data processing. Sequences were submitted to a custom build bioinformatics pipeline^{84–86}. Reads were aligned to the bovine genome (ARS-UCD1.2 obtained from Ensembl⁸⁷) using HISAT2⁸⁸. Sequences aligning to multiple places on the genome, or with 5 or more mismatches, were filtered out. The sequences were then marked for duplicates, and non-duplicated pairs of reads (mRNA) or single reads (miRNA) were used for transcript quantification at the gene level. The fragments were counted against the Ensembl gene annotation⁸⁹ (version 1.87) using featureCounts⁹⁰. RNAseq data from one animal (non-pregnant) were removed from analysis due to low sequencing yield (613,765 reads).

Filtering of lowly expressed genes. All analytical procedures were carried out in R software⁹¹. Supplementary code contains all scripts used for analytical procedures described below and the creation of the plots. This code and all original data can be obtained on figshare⁹².

Counts per million (CPM), and fragments per kilobase per million reads (FPKM) were calculated using “edgeR”⁹³. In order to remove quantification uncertainty associated to lowly expressed genes and erroneous identification of differentially expressed genes^{94,95}, we retained genes with at least 2 CPM for PWBC mRNA or at least 1 CPM for plasma miRNA in five or more samples for downstream analyses. Further analytical procedures were carried out with genes annotated as “protein-coding” and “miRNA” for the PWBC RNA-sequencing and plasma small RNA-sequencing, respectively.

Pair-wise gene coexpression and differential coexpression for protein-coding genes. First, we calculated pair-wise Pearson's correlation coefficient (r)^{96,97} for all 17 animals sampled. Significance of the correlations was assessed using the “TestCor” package based on the empirical statistical test and the large-scale correlation test with normal approximation⁹⁸ for the calculation of the false discovery rate (FDR). Further biological interpretation of results was performed on pairs of genes if $\text{abs}(r) > 0.98$, which corresponded to a nominal $P \leq 1.2 \times 10^{-16}$ and $\text{FDR} < 0.02$.

Next, we calculated within group pair-wise Pearson's correlation coefficient (r) for each group: AI-pregnant, NB-pregnant, and non-pregnant. We interpreted a differential correlation as biologically meaningful based on the following scenarios, assuming that groups 1 and 2 can be assigned arbitrarily: inverted correlation, if $|r_{\text{group1}} - r_{\text{group2}}| > 1.95$; gained correlation in group 1 and lost correlation in group 2, if $|r_{\text{group1}}| > 0.99$ and $-0.1 < r_{\text{group2}} < 0.1$.

Significance was estimated by the eFDR approach^{85,99,100}. We resampled (5000 randomizations) the labels of pregnancy outcome and calculated r for each pair of genes in the scrambled data set ($r_{scramble}$). The eFDR was obtained based on the proportion of the $r_{scramble}$ greater than a specified r (i.e.: for $r = 0.99$, eFDR < 0.002, Supplementary Fig. S1), relative to the total number of correlations calculated using the formulas described elsewhere^{85,99,100}.

mRNA–miRNA coexpression and differential coexpression. We calculated Pearson's correlation (r)^{96,97} values between miRNA and mRNA genes for all 17 samples. Significance was estimated by the eFDR approach^{85,99,100} by shuffling (5000 randomizations) the animal identification on mRNA dataset, thus breaking the common source of both RNA types (Supplementary Fig. S2).

Pearson's correlation (r) values were calculated for mRNA–miRNA transcript pairs within groups: AI-pregnant, NB-pregnant, and non-pregnant heifers. We interpreted a differential correlation as biologically meaningful based on the following scenarios $|r_{group1}| > 0.99$ and $-0.1 < r_{group2} < 0.1$. Significance was assessed by the eFDR approach (Supplementary Fig. S3). We further interrogated the data for experimentally validated miRNA–mRNA interactions obtained from miRWalk database (version 3)^{101,102}.

Differential expression of mRNA and miRNA. Differences of transcript levels between groups based on the pregnancy outcome were determined from fragment counts using “edgeR”⁹³ and “DESeq2”¹⁰³. For each comparison, a gene was initially inferred as differentially expressed if the nominal P value was ≤ 0.03 in both “edgeR” and “DESeq2” analyses. This nominal P value corresponded to eFDR ≤ 0.05 (Supplementary Fig. S4), as calculated according to the procedure and formulas outlined elsewhere^{85,99,100} using 10,000 randomizations of sample reshuffling. Genes inferred as differentially expressed based on P values were further selected through “leave-one-out” cross validation^{104,105}. Only genes whose significance were maintained after all rounds of sampling were further considered for biological interpretation.

Testing for enrichment of gene ontology terms and KEGG pathways. Differentially expressed genes and genes that displayed differential coexpression among heifers of AI-pregnant, NB-pregnant, and non-pregnant classifications were analyzed for enrichment of Gene Ontology¹⁰⁶ categories and KEGG Pathways¹⁰⁷. In all tests, the background gene list consisted of all genes detected as expressed in our dataset¹⁰⁸. The GO annotations and transcript lengths were obtained from BioMart¹⁰⁹ and KEGG annotation was obtained from the “org.Bt.eb” package. We used “GoSeq” package¹¹⁰ to estimate significance of category enrichment using the sampling method¹¹⁰. We utilized FDR for multiple testing under dependency¹¹¹, and inferred statistical significance of GO terms and KEGG pathways if FDR < 0.10.

Assessment of predictive potential of gene expression levels by machine learning algorithms. We focused the mRNA sequencing data from AI-pregnant and non-pregnant heifers from this report and data we collected from the same site on the previous year and published elsewhere^{38,86}. The raw sequences from the previous dataset were realigned, and data were treated as described above.

To remove the genes not informative to the machine learning algorithm, we identified differentially expressed genes between the two groups (AI-pregnant and non-pregnant). We used the same procedures and standards aforementioned with the exception that we added year as a fixed effect into the model. There were 198 genes differentially expressed between these two groups (AI-preg and non-pregnant). We performed a variance stabilizing transformation on the counts for these 198 genes, and the transformed data were used as input.

For the machine learning approach, we used the expression data for the 198 DEGs from year one (2015) as a training set and the data for the same group of genes from year two (2016) as a test set. Initially we processed the training set for the development of models and assessed the prediction accuracy for 86 classification algorithms (Supplementary Fig. S5). Accuracy was determined by the number of samples correctly classified divided by the number of samples tested ($n = 11$). Six models that showed prediction accuracy ≥ 0.9 (classified correctly at least 10 samples, accuracy P value ≤ 0.01 , Supplementary Fig. S5), were further subjected to 2000 predictions with different initial ‘seed’ numbers.

Ethics of research with animals

All procedures with animals were approved by Institutional Animal and Care and Use Committee at Auburn University. All methods related to animal handling were performed in accordance with the guidelines stipulated by the Institutional Animal and Care and Use Committee at Auburn University and the Guide for the Care and Use of Laboratory Animals.

Data availability

All raw sequencing files and raw read counts produced in this work are deposited in the National Center for Biotechnology Information under the Gene Expression Omnibus database, accession: GSE146041. All the intermediate data and results produced from raw files can be reproduced based on the codes available on Supplementary Code.

Received: 12 May 2020; Accepted: 18 September 2020

Published online: 08 October 2020

References

- Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M. & Tiwari, B. future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods* <https://doi.org/10.3390/foods6070053> (2017).
- Bormann, J. M., Totir, L. R., Kachman, S. D., Fernando, R. L. & Wilson, D. E. Pregnancy rate and first-service conception rate in Angus heifers. *J. Anim. Sci.* **84**, 2022–2025. <https://doi.org/10.2527/jas.2005-615> (2006).
- Roberts, A. J., Geary, T. W., Grings, E. E., Waterman, R. C. & MacNeil, M. D. Reproductive performance of heifers offered ad libitum or restricted access to feed for a one hundred forty-day period after weaning. *J. Anim. Sci.* **87**, 3043–3052. <https://doi.org/10.2527/jas.2008-1476> (2009).
- Peters, S. O. *et al.* Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. *J. Anim. Sci.* **91**, 605–612. <https://doi.org/10.2527/jas.2012-5580> (2013).
- Grings, E. E., Geary, T. W., Short, R. E. & MacNeil, M. D. Beef heifer development within three calving systems. *J. Anim. Sci.* **85**, 2048–2058. <https://doi.org/10.2527/jas.2006-758> (2007).
- Funston, R. N. & Deutscher, G. H. Comparison of target breeding weight and breeding date for replacement beef heifers and effects on subsequent reproduction and calf performance. *J. Anim. Sci.* **82**, 3094–3099 (2004).
- Funston, R. N. & Larson, D. M. Heifer development systems: Dry-lot feeding compared with grazing dormant winter forage. *J. Anim. Sci.* **89**, 1595–1602. <https://doi.org/10.2527/jas.2010-3095> (2011).
- Gutierrez, K. *et al.* Effect of reproductive tract scoring on reproductive efficiency in beef heifers bred by timed insemination and natural service versus only natural service. *Theriogenology* **81**, 918–924. <https://doi.org/10.1016/j.theriogenology.2014.01.008> (2014).
- Martin, J. L. *et al.* Effect of prebreeding body weight or progesterin exposure before breeding on beef heifer performance through the second breeding season. *J. Anim. Sci.* **86**, 451–459. <https://doi.org/10.2527/jas.2007-0233> (2008).
- Lynch, J. M. *et al.* Influence of timing of gain on growth and reproductive performance of beef replacement heifers. *J. Anim. Sci.* **75**, 1715–1722 (1997).
- Mallory, D. A., Nash, J. M., Ellersieck, M. R., Smith, M. F. & Patterson, D. J. Comparison of long-term progesterin-based protocols to synchronize estrus before fixed-time artificial insemination in beef heifers. *J. Anim. Sci.* **89**, 1358–1365. <https://doi.org/10.2527/jas.2010-3694> (2011).
- Patterson, D. J., Corrah, L. R., Kiracofe, G. H., Stevenson, J. S. & Brethour, J. R. Conception rate in *Bos taurus* and *Bos indicus* crossbred heifers after postweaning energy manipulation and synchronization of estrus with melengestrol acetate and fenprostalene. *J. Anim. Sci.* **67**, 1138–1147 (1989).
- Dickinson, S. E. *et al.* Evaluation of age, weaning weight, body condition score, and reproductive tract score in pre-selected beef heifers relative to reproductive potential. *J. Anim. Sci. Biotechnol.* **10**, 18. <https://doi.org/10.1186/s40104-019-0329-6> (2019).
- Dickinson, S. E. *et al.* Evaluation of age, weaning weight, body condition score, and reproductive tract score in pre-selected beef heifers relative to reproductive potential. *J. Anim. Sci. Biotechnol.* **10**, 18. <https://doi.org/10.1186/s40104-019-0329-6> (2019).
- Cushman, R. A., Kill, L. K., Funston, R. N., Mousel, E. M. & Perry, G. A. Heifer calving date positively influences calf weaning weights through six parturitions. *J. Anim. Sci.* **4486–4491**, 2013. <https://doi.org/10.2527/jas2013-6465> (2013).
- Marshall, D. M., Minqiang, W. & Freking, B. A. Relative calving date of first-calf heifers as related to production efficiency and subsequent reproductive performance. *J. Anim. Sci.* **68**, 1812–1817 (1990).
- Lesmeister, J. L., Burfening, P. J. & Blackwell, R. L. Date of first calving in beef cows and subsequent calf production. *J. Anim. Sci.* **36**, 1–6. <https://doi.org/10.2527/jas1973.3611> (1973).
- Damiran, D., Larson, K. A., Pearce, L. T., Erickson, N. E. & Lardner, B. H. A. Effect of calving period on beef cow longevity and lifetime productivity in western Canada. *Transl. Anim. Sci.* **2**, S61–S65. <https://doi.org/10.1093/tas/txy020> (2018).
- Morris, C. A. & Cullen, N. G. A note on genetic correlations between pubertal traits of males or females and lifetime pregnancy rate in beef cattle. *Livest. Prod. Sci.* **39**, 291–297. [https://doi.org/10.1016/0301-6226\(94\)90291-7](https://doi.org/10.1016/0301-6226(94)90291-7) (1994).
- Mwansa, P. B. *et al.* Selection for cow lifetime pregnancy rate using bull and heifer growth and reproductive traits in composite cattle. *Can. J. Anim. Sci.* **80**, 507–510 (2000).
- Peters, S. O. *et al.* Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. *J. Anim. Sci.* **605–612**, 2013. <https://doi.org/10.2527/jas2012-5580> (2013).
- Neupane, M. *et al.* Loci and pathways associated with uterine capacity for pregnancy and fertility in beef cattle. *PLoS ONE* **12**, e0188997. <https://doi.org/10.1371/journal.pone.0188997> (2017).
- McDanel, T. G. *et al.* Genomewide association study of reproductive efficiency in female cattle. *J. Anim. Sci.* **1945–1957**, 2014. <https://doi.org/10.2527/jas2012-6807> (2014).
- McDanel, T. G. *et al.* Y are you not pregnant: Identification of Y chromosome segments in female cattle with decreased reproductive efficiency. *J. Anim. Sci.* **90**, 2142–2151 (2012).
- de Camargo, G. M. *et al.* Association between JY-1 gene polymorphisms and reproductive traits in beef cattle. *Gene* **533**, 477–480. <https://doi.org/10.1016/j.gene.2013.09.126> (2014).
- Dias, M. M. *et al.* Study of lipid metabolism-related genes as candidate genes of sexual precocity in Nelore cattle. *Genet. Mol. Res.* **14**, 234–243. <https://doi.org/10.4238/2015.January.16.7> (2015).
- Irano, N. *et al.* Genome-wide association study for indicator traits of sexual precocity in Nelore cattle. *PLoS ONE* **11**, e0159502. <https://doi.org/10.1371/journal.pone.0159502> (2016).
- Junior, G. A. O. *et al.* Genomic study and medical subject headings enrichment analysis of early pregnancy rate and antral follicle numbers in Nelore heifers. *J. Anim. Sci.* **95**, 4796–4812. <https://doi.org/10.2527/jas2017.1752> (2017).
- Fortes, M. R. S. *et al.* Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *J. Anim. Sci.* **2894–2906**, 2012. <https://doi.org/10.2527/jas2011-4601> (2012).
- Doyle, S. P., Golden, B. L., Green, R. D. & Brinks, J. S. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. *J. Anim. Sci.* **78**, 2091–2098. <https://doi.org/10.2527/2000.7882091x> (2000).
- Toghiani, S. *et al.* Genomic prediction of continuous and binary fertility traits of females in a composite beef cattle breed. *J. Anim. Sci.* **95**, 4787–4795. <https://doi.org/10.2527/jas2017.1944> (2017).
- McAllister, C. M., Speidel, S. E., Crews, D. H. Jr. & Enns, R. M. Genetic parameters for intramuscular fat percentage, marbling score, scrotal circumference, and heifer pregnancy in Red Angus cattle. *J. Anim. Sci.* **89**, 2068–2072. <https://doi.org/10.2527/jas.2010-3538> (2011).
- Bodhireddy, P. *et al.* Genomic predictions in Angus cattle: Comparisons of sample size, response variables, and clustering methods for cross-validation. *J. Anim. Sci.* **92**, 485–497. <https://doi.org/10.2527/jas.2013-6757> (2014).
- Li-Pook-Than, J. & Snyder, M. iPOP goes the world: Integrated personalized Omics profiling and the road toward improved health care. *Chem. Biol.* **20**, 660–666. <https://doi.org/10.1016/j.chembiol.2013.05.001> (2013).
- Mancuso, N. *et al.* Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. *Am. J. Hum. Genet.* **100**, 473–487. <https://doi.org/10.1016/j.ajhg.2017.01.031> (2017).
- Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252. <https://doi.org/10.1038/ng.3506> (2016).

37. Ioannidis, J. & Donadeu, F. X. Circulating microRNA profiles during the bovine oestrous cycle. *PLoS ONE* **11**, e0158160. <https://doi.org/10.1371/journal.pone.0158160> (2016).
38. Dickinson, S. E. *et al.* Transcriptome profiles in peripheral white blood cells at the time of artificial insemination discriminate beef heifers with different fertility potential. *BMC Genomics* <https://doi.org/10.1186/s12864-018-4505-4> (2018).
39. Emmert-Streib, F., Dehmer, M. & Haibe-Kains, B. Gene regulatory networks and their applications: Understanding biological and medical problems in terms of networks. *Front. Cell. Dev. Biol.* **2**, 38. <https://doi.org/10.3389/fcell.2014.00038> (2014).
40. Xu, C. & Jackson, S. A. Machine learning and complex biological data. *Genome Biol.* **20**, 76. <https://doi.org/10.1186/s13059-019-1689-0> (2019).
41. Monaco, G. *et al.* RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640.e1627. <https://doi.org/10.1016/j.celrep.2019.01.041> (2019).
42. van Dam, S., Vosa, U., van der Graaf, A., Franke, L. & de Magalhaes, J. P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform.* **19**, 575–592. <https://doi.org/10.1093/bib/bbw139> (2018).
43. Sheridan, M. A. *et al.* Early onset preeclampsia in a model for human placental trophoblast. *Proc. Natl. Acad. Sci. USA* **116**, 4336–4345. <https://doi.org/10.1073/pnas.1816150116> (2019).
44. Fitzgerald, H. C. *et al.* Idiopathic infertility in women is associated with distinct changes in proliferative phase uterine fluid proteins. *Biol. Reprod.* **98**, 752–764. <https://doi.org/10.1093/biolre/iyoy063> (2018).
45. Yang, H. L. *et al.* Autophagy in endometriosis. *Am. J. Transl. Res.* **9**, 4707–4725 (2017).
46. Pelch, K. E. *et al.* Aberrant gene expression profile in a mouse model of endometriosis mirrors that observed in women. *Fertil. Steril.* **93**, 1615–1627.e1618. <https://doi.org/10.1016/j.fertnstert.2009.03.086> (2010).
47. Demain, L. A., Conway, G. S. & Newman, W. G. Genetics of mitochondrial dysfunction and infertility. *Clin. Genet.* **91**, 199–207. <https://doi.org/10.1111/cge.12896> (2017).
48. Koks, S. *et al.* The differential transcriptome and ontology profiles of floating and cumulus granulosa cells in stimulated human antral follicles. *Mol. Hum. Reprod.* **16**, 229–240. <https://doi.org/10.1093/molehr/gap103> (2010).
49. Li, S. H. *et al.* Correlation of cumulus gene expression of GJA1, PRSS35, PTX3, and SERPINE2 with oocyte maturation, fertilization, and embryo development. *Reprod. Biol. Endocrinol.* **13**, 93. <https://doi.org/10.1186/s12958-015-0091-3> (2015).
50. Mtango, N. R. *et al.* Essential role of maternal UCHL1 and UCHL3 in fertilization and preimplantation embryo development. *J. Cell Physiol.* **227**, 1592–1603. <https://doi.org/10.1002/jcp.22876> (2012).
51. Ingham, N. J. *et al.* Mouse screen reveals multiple new genes underlying mouse and human hearing loss. *PLoS Biol.* **17**, e3000194. <https://doi.org/10.1371/journal.pbio.3000194> (2019).
52. Reeves, G. T. The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback. *J. Biol. Eng.* **13**, 62. <https://doi.org/10.1186/s13036-019-0190-3> (2019).
53. Sun, H. Z., Chen, Y. & Guan, L. L. MicroRNA expression profiles across blood and different tissues in cattle. *Sci. Data* **6**, 190013. <https://doi.org/10.1038/sdata.2019.13> (2019).
54. Vaz, C. *et al.* Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. *BMC Genomics* **11**, 288. <https://doi.org/10.1186/1471-2164-11-288> (2010).
55. Fritz, J. V. *et al.* Sources and functions of extracellular small RNAs in human circulation. *Annu. Rev. Nutr.* **36**, 301–336. <https://doi.org/10.1146/annurev-nutr-071715-050711> (2016).
56. Diaz, G., Zamboni, F., Tice, A. & Farci, P. Integrated ordination of miRNA and mRNA expression profiles. *BMC Genomics* **16**, 767. <https://doi.org/10.1186/s12864-015-1971-9> (2015).
57. Ritchie, W., Rajasekhar, M., Flamant, S. & Rasko, J. E. Conserved expression patterns predict microRNA targets. *PLoS Comput. Biol.* **5**, e1000513. <https://doi.org/10.1371/journal.pcbi.1000513> (2009).
58. Bayraktar, R., Van Roosbroeck, K. & Calin, G. A. Cell-to-cell communication: MicroRNAs as hormones. *Mol. Oncol.* **11**, 1673–1686. <https://doi.org/10.1002/1878-0261.12144> (2017).
59. Rayner, K. J. & Hennessy, E. J. Extracellular communication via microRNA: Lipid particles have a new message. *J. Lipid Res.* **54**, 1174–1181. <https://doi.org/10.1194/jlr.R034991> (2013).
60. Robertson, S. A. & Sharkey, D. J. The role of semen in induction of maternal immune tolerance to pregnancy. *Semin. Immunol.* **13**, 243–254. <https://doi.org/10.1006/smim.2000.0320> (2001).
61. Moffett, A. & Loke, C. Immunology of placentation in eutherian mammals. *Nat. Rev. Immunol.* **6**, 584–594. <https://doi.org/10.1038/nri1897> (2006).
62. Momen-Heravi, F. & Bala, S. miRNA regulation of innate immunity. *J. Leukoc. Biol.* <https://doi.org/10.1002/JLB.3MIR1117-459R> (2018).
63. Jin, W. *et al.* Transcriptome microRNA profiling of bovine mammary epithelial cells challenged with *Escherichia coli* or *Staphylococcus aureus* bacteria reveals pathogen directed microRNA expression profiles. *BMC Genomics* **15**, 181. <https://doi.org/10.1186/1471-2164-15-181> (2014).
64. Li, R. *et al.* Transcriptome microRNA profiling of bovine mammary glands infected with *Staphylococcus aureus*. *Int. J. Mol. Sci.* **16**, 4997–5013. <https://doi.org/10.3390/ijms16034997> (2015).
65. Vegh, P. *et al.* MicroRNA profiling of the bovine alveolar macrophage response to *Mycobacterium bovis* infection suggests pathogen survival is enhanced by microRNA regulation of endocytosis and lysosome trafficking. *Tuberculosis (Edinb)* **95**, 60–67. <https://doi.org/10.1016/j.tube.2014.10.011> (2015).
66. Ma, S., Tong, C., Ibeagha-Awemu, E. M. & Zhao, X. Identification and characterization of differentially expressed exosomal microRNAs in bovine milk infected with *Staphylococcus aureus*. *BMC Genomics* **20**, 934. <https://doi.org/10.1186/s12864-019-6338-1> (2019).
67. Pang, F. *et al.* Comprehensive analysis of differentially expressed microRNAs and mRNAs in MDBK cells expressing bovine papillomavirus E5 oncogene. *PeerJ* **7**, e8098. <https://doi.org/10.7717/peerj.8098> (2019).
68. Hughes, H. D., Carroll, J. A., Burdick Sanchez, N. C. & Richeson, J. T. Natural variations in the stress and acute phase responses of cattle. *Innate Immun.* **20**, 888–896. <https://doi.org/10.1177/1753425913508993> (2014).
69. Wenric, S. & Shemirani, R. Using supervised learning methods for gene selection in RNA-seq case-control studies. *Front. Genet.* **9**, 297. <https://doi.org/10.3389/fgene.2018.00297> (2018).
70. Zararsiz, G. *et al.* A comprehensive simulation study on classification of RNA-seq data. *PLoS ONE* **12**, e0182507. <https://doi.org/10.1371/journal.pone.0182507> (2017).
71. Brazdova, A., Senechal, H., Peltre, G. & Poncet, P. Immune aspects of female infertility. *Int. J. Fertil. Steril.* **10**, 1–10. <https://doi.org/10.22074/ijfs.2016.4762> (2016).
72. Altman, N. & Krzywinski, M. Association, correlation and causation. *Nat. Methods* **12**, 899–900. <https://doi.org/10.1038/nmeth.3587> (2015).
73. Fang, L. *et al.* Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res.* **30**, 790–801. <https://doi.org/10.1101/gr.250704.119> (2020).
74. Patterson, D. J. & Smith, M. F. Management considerations in beef heifer development and puberty. *Vet. Clin. North Am. Food Anim. Pract.* **29**, 13–14. <https://doi.org/10.1016/j.cvfa.2013.07.014> (2013).
75. Larson, R. L., White, B. J. & Laffin, S. Beef heifer development. *Vet. Clin. North Am. Food Anim. Pract.* **32**, 285–302. <https://doi.org/10.1016/j.cvfa.2016.01.003> (2016).

76. Holm, D. E., Thompson, P. N. & Irons, P. C. The value of reproductive tract scoring as a predictor of fertility and production outcomes in beef heifers. *J. Anim. Sci.* **87**, 1934–1940. <https://doi.org/10.2527/jas.2008-1579> (2009).
77. Rae, D. O., Kunkle, W. E., Chenoweth, P. J., Sand, R. S. & Tran, T. Relationship of parity and body condition score to pregnancy rate in Florida beef-cattle. *Theriogenology* **1143–1152**, 1993. [https://doi.org/10.1016/0093-691x\(93\)90013-U](https://doi.org/10.1016/0093-691x(93)90013-U) (1993).
78. Larson, J. E. *et al.* Synchronization of estrus in suckled beef cows for detected estrus and artificial insemination and timed artificial insemination using gonadotropin-releasing hormone, prostaglandin F_{2α}, and progesterone. *J. Anim. Sci.* **2006**, 332–342 (2006).
79. Rio, D. C., Ares, M. Jr., Hannon, G. J. & Nilsen, T. W. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb. Protoc.* <https://doi.org/10.1101/pdb.prot5439> (2010).
80. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: Twenty-something years on. *Nat. Protoc.* **1**, 581–585. <https://doi.org/10.1038/nprot.2006.83> (2006).
81. Yeri, A. *et al.* Total extracellular small RNA profiles from plasma, saliva, and urine of healthy subjects. *Sci. Rep.* **7**, 44061. <https://doi.org/10.1038/srep44061> (2017).
82. Chory, J. & Pollard, J. D. Jr. Separation of small DNA fragments by conventional gel electrophoresis. *Curr. Protoc. Mol. Biol.* <https://doi.org/10.1002/0471142727.mb0207s47> (2001).
83. Green, M. R. & Sambrook, J. Isolation of DNA fragments from polyacrylamide gels by the crush and soak method. *Cold Spring Harb. Protoc.* <https://doi.org/10.1101/pdb.prot100479> (2019).
84. Biase, F. H. *et al.* Massive dysregulation of genes involved in cell signaling and placental development in cloned cattle conceptus and maternal endometrium. *Proc. Natl. Acad. Sci. USA* **113**, 14492–14501. <https://doi.org/10.1073/pnas.1520945114> (2016).
85. Biase, F. H. & Kimble, K. M. Functional signaling and gene regulatory networks between the oocyte and the surrounding cumulus cells. *BMC Genomics* **19**, 351. <https://doi.org/10.1186/s12864-018-4738-2> (2018).
86. Dickinson, S. E. & Biase, F. H. Transcriptome data of peripheral white blood cells from beef heifers collected at the time of artificial insemination. *Data Brief* **18**, 706–709. <https://doi.org/10.1016/j.dib.2018.03.062> (2018).
87. Kinsella, R. J. *et al.* Ensembl BioMarts: A hub for data retrieval across taxonomic space. *Database* <https://doi.org/10.1093/database/bar030> (2011).
88. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360. <https://doi.org/10.1038/nmeth.3317> (2015).
89. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749–755. <https://doi.org/10.1093/nar/gkt1196> (2014).
90. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2014).
91. Ihaka, R. & Gentleman, R. A language and environment for statistical computing. *J. Comput. Graph. Stat.* **5**, 299–314 (2012).
92. Biase, F. H. Supplementary code and files to rewiring of gene expression in circulating white blood cells is associated with pregnancy outcome in heifers (*Bos taurus*). *Figshare* <https://doi.org/10.6084/m9.figshare.11985666.v3> (2020).
93. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. <https://doi.org/10.1093/bioinformatics/btp616> (2010).
94. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94. <https://doi.org/10.1186/1471-2105-11-94> (2010).
95. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25. <https://doi.org/10.1186/gb-2010-11-3-r25> (2010).
96. Bansal, M., Belcastro, V., Ambesi-Impiombato, A. & di Bernardo, D. How to infer gene networks from expression profiles. *Mol. Syst. Biol.* <https://doi.org/10.1038/msb4100120> (2007).
97. Serin, E. A., Nijveen, H., Hilhorst, H. W. & Ligterink, W. Learning from co-expression networks: Possibilities and challenges. *Front. Plant. Sci.* **7**, 444. <https://doi.org/10.3389/fpls.2016.00444> (2016).
98. Cai, T. T. & Liu, W. Large-scale multiple testing of correlations. *J. Am. Stat. Assoc.* **111**, 229–240. <https://doi.org/10.1080/01621459.2014.999157> (2016).
99. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445. <https://doi.org/10.1073/pnas.1530509100> (2003).
100. Sham, P. C. & Purcell, S. M. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* **15**, 335–346. <https://doi.org/10.1038/nrg3706> (2014).
101. Dweep, H., Sticht, C., Pandey, P. & Gretz, N. miRWalk-database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J. Biomed. Inform.* **44**, 839–847. <https://doi.org/10.1016/j.jbi.2011.05.002> (2011).
102. Dweep, H. & Gretz, N. miRWalk2.0: A comprehensive atlas of microRNA-target interactions. *Nat. Methods* **12**, 697. <https://doi.org/10.1038/nmeth.3485> (2015).
103. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/PREACCEPT-8897612761307401> (2014).
104. Picard, R. R. & Cook, R. D. Cross-validation of regression-models. *J. Am. Stat. Assoc.* **79**, 575–583. <https://doi.org/10.2307/2288403> (1984).
105. Breiman, L. & Spector, P. Submodel selection and evaluation in regression—The X-random case. *Int. Stat. Rev.* **60**, 291–319. <https://doi.org/10.2307/1403680> (1992).
106. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
107. Du, J. *et al.* KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol. Biosyst.* **10**, 2441–2447. <https://doi.org/10.1039/C4MB00287C> (2014).
108. Timmons, J. A., Szkop, K. J. & Gallagher, I. J. Multiple sources of bias confound functional enrichment analysis of global-omics data. *Genome Biol.* **16**, 186. <https://doi.org/10.1186/s13059-015-0761-7> (2015).
109. Durinck, S. *et al.* BioMart and bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525> (2005).
110. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **11**, R14. <https://doi.org/10.1186/gb-2010-11-2-r14> (2010).
111. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188. <https://doi.org/10.1214/aos/1013699998> (2001).

Author contributions

S.E.M. performed sample collection, library preparations, contributed to the data analysis, prepared the first draft and edited the final draft. B.N.W. contributed to sample collection, assisted on the laboratory assays and editing of the final draft. M.F.E. is responsible for the acquisition and curation of the animal database. J.B.E. and S.P.R. were responsible for the health and reproductive management of the animals. F.B. conceived designed and supervised the experiments. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-73694-w>.

Correspondence and requests for materials should be addressed to F.H.B.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 Author(s) 2020