Rewiring of gene expression in PWBC at artificial insemination is predictive of pregnancy outcome in cattle (*Bos taurus*)

Sarah E. Moorey[1], Bailey N. Walker[2], Michelle F. Elmore[3,4], Joshua B. Elmore[4], Soren P. Rodning[3] and Fernando H. Biase[2]

[1] Department of Animal Science, University of Tennessee, Knoxville, TN; [2] Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [3] Department of Animal Sciences, Auburn University, Auburn, AL, USA; [4] Alabama Cooperative Extension System, AL, USA.

Corresponding author:

Fernando Biase

175 West Campus Drive, Blacksburg, VA 24061. Phone 540-231-9520. Email fbiase@vt.edu

ORCID ID https://orcid.org/0000-0001-7895-0223

ABSTRACT

Infertility is a disease that affects humans and cattle in similar ways. The resemblance includes complex genetic architecture, multiple etiology, low heritability of fertility related traits in females, and the frequency in the female population. Here, we used cattle as a biomedical model to test 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 female calves on the day of artificial insemination and analyzed transcript abundance for 10496 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 bloodborne transcript abundance is highly associated with fertility in females.

INTRODUCTION

The Word Health Organization considers Infertility to be a "disease of the reproductive system"[1]. Infertility is a common problem in humans and agriculturally important large mammals. For instance, prevalence of infertility may reach 16%[2] and 15%[3] in humans and beef female calves, respectively. Similar to what has been observed in humans[4], the genetic architecture of fertility in beef heifers is very complex[5,6,7,8,9,10,11,12]. In line with this complexity, female reproductive traits have low heritability. The heritability of traits related to female reproductive fitness (i.e.: birth rate and age at last reproduction) in humans ranges from 0.05 to 0.28[13]. Likewise, reproductive related traits in cattle such as heifer pregnancy and first service conception range from 0.07 to 0.20[5,14,15,16,17,18,19] and 0.03 to 0.18[5,14,17], respectively. Due to genetic and physiological resemblance with human, cattle have been identified as an important biomedical model for dissecting female fertility traits[20,21].

Beyond the importance as a biomedical model, cattle production systems provide approximately 28%[22] of the protein supply globally. Improving cattle production efficiency is essential for farmers to attain sustainable production and support the growing demand for animal protein[22]. Infertility is a major factor that hinders efficiency in cattle production, and it starts with limited success of pregnancy in young female calves. 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[23,24,25,26]. Furthermore, the genetic correlation between yearling pregnancy rate and lifetime pregnancy rate is high (0.92-0.97)[27,28]. 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 examination of the genetic components of fertility in beef heifers have yielded several genes potentially associated with fertility traits[5,6,7,8,9,10,11,12], but the effect of these markers are minimal, and there is no clear redundancy in genetic markers identified across breeds. 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 traits[29]. In line with this rationale, expression-trait association studies have emerged as a means to better understand complex traits[30,31]. Specifically related to fertility, there is evidence that circulating micro RNA (miRNA) profiles are indicative of reproductive function[32]. Furthermore, we have previously identified differences in the transcriptome of peripheral white blood cells (PWBC) among heifers of different fertility potential[33].

In the present study, we set out 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. We also show that transcript abundance in PWBCs can be successfully utilized for the prediction of heifer reproductive outcome.

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 (not-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, 10496 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 not-pregnant samples clustered separately from other subjects based on the expression levels of protein-coding genes. 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 1). 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 2). 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,

97 Supplementary Table 3, Supplementary Fig. 1). Of these 1860 genes, 963 genes formed 716
98 negative, 1053 genes formed 708 positive, four genes lost two positive, and 40 genes formed 23
99 inverted coexpression connections in not-preg heifers relative to both AI-preg and NB-preg
100 heifers. Among the 963 genes that formed new negative coexpression connections in not-preg
101 heifers, there was enrichment of the biological processes "proteolysis" (n=38 genes) and
102 "translation" (n=52 genes, FDR<0.1, see Fig. 2d for selected genes and Supplementary Table 4
103 for list of genes). Five of the 40 genes that formed inverted coexpression connection in not-preg
104 heifers relative to the pregnant counterparts were associated with "negative regulation of
105 transcription by RNA polymerase II" (FDR<0.1, see Fig. 2d for selected genes and Supplementary
106 Table 5 for a list of genes). These results indicate that physiological factors associated with fertility
107 or infertility act on the transcription protein-coding genes in PWBCs.

108 ***Differential miRNA:mRNA coexpression associated with pregnancy outcome***

109 The generation of data from small RNA (from plasma) and mRNA (from PWBCs) from the
110 same subjects (n=17) permitted us to ask whether circulating miRNAs have correlated expression
111 with protein-coding genes expressed in the PWBCs. We identified 141 pairs of miRNA:mRNA
112 with |r|>0.85 (eFDR<0.0001, Supplementary Fig. 2) that were formed by 106 protein-coding genes
113 and 33 miRNAs (Supplementary table 6). Among the 106 protein-coding genes existing in high
114 correlation with miRNA, we identified three significantly enriched biological processes, namely:
115 "peptidyl-threonine phosphorylation", "cilium assembly" and "regulation of gene expression" (Fig.
116 3a, Supplementary table 7). Of notice, 30 of the 141 miRNA:mRNA pairs that presented negative
117 correlation between the expression levels have been previously identified as interacting pairs in
118 the miRWalk database (Fig. 3b). Additionally, bta-mir-744 and bta-mir-320a emerged as the two
119 most connected miRNAs (13 and 6 connections, respectively). It is evident that circulating
120 miRNAs have regulatory roles over transcripts of protein-coding genes expressed in the PWBCs.

121 We also interrogated miRNA:mRNA pairs for differential coexpression associated with
122 pregnancy outcome. We identified 2330 and 2738 miRNA:mRNA pairs with gain of negative and
123 positive correlated expression, respectively, (|r|>0.9, eFDR≤0.01, Supplementary Fig. 3) in not-
124 pregnant heifers that were not identified compared to in pregnant heifers (both AI-preg and NB-
125 preg, Fig. 3c). In addition, there were 51, two and three connections showing inverted, loss of
126 negative and loss of positive correlations in not-pregnant heifers relative to the pregnant (both AI-
127 preg and NB-preg) counterparts (Fig. 3c, see Supplementary table 8 for genes). Among the 1220
128 protein-coding genes associated with gain of negative correlation with miRNAs, there were
129 biological processes significantly enriched, namely: "immunoglobulin mediated immune

130 response", "negative regulation of DNA-binding transcription factor activity", "lymphocyte
131 chemotaxis" and "mitochondrial respiratory chain complex I assembly" (FDR<0.1, Supplementary
132 table 9). Among the 1288 genes associated with gain of positive correlation with miRNAs, the
133 following biological processes were significantly enriched: "regulation of DNA replication" and
134 "cellular response to BMP stimulus" (FDR<0.1, Supplementary table 10). These results indicate
135 that several miRNA and mRNA pairs have their connections altered in relation to the heifer's
136 fertility potential.

137 ***Differential gene expression associated with pregnancy outcome***

138 Differential gene expression analysis revealed 67 (55 up- and 12 down-regulated) and 81
139 (63 up- and 18 down-regulated) protein-coding genes with altered expression profiles between
140 PWBC of AI-preg versus not-pregnant, and NB-preg versus not-pregnant heifers, respectively
141 (eFDR≤0.05, Fig. 4a, Supplementary tables 11 and 12, Supplementary Fig. 4). Notably, 26 genes
142 were common between both pregnant groups versus the not-pregnant group (Fig. 4b). Gene
143 ontology and pathway analyses of differentially expressed genes between AI-preg and Not-
144 pregnant heifers revealed significant enrichment for the terms "signal transduction" (*ARAP3*,
145 *LSP1*, *NFAM1*, *OR12D2*, *RARA*; FDR<0.05). When differentially expressed genes between NB-
146 preg and not-pregnant heifers were analyzed, significant enrichment was determined for the terms
147 "signal transduction" (*ARAP3*, *HRH2*, *LSP1*, *TIAM2*, *TNFRSF17*, *TNFRSF1A*; FDR=0.062) and
148 the KEGG Pathway "cytokine-cytokine receptor interaction" (*IFNGR2*, *IL5RA*, *LTB*, *TNFRSF17*,
149 *TNFRSF1A*; FDR=0.0045). Plasma miRNA profiles were very similar among samples from all
150 pregnancy outcomes with the exception of miR-11995 that was more abundant in AI-preg versus
151 not-pregnant heifers. These results indicate that quantitative differences exist in transcript
152 abundance of protein-coding genes expressed in PWBC relative to the pregnancy outcome.

153 Next, we asked if differential gene expression could be associated with differential
154 miRNA:mRNA coexpression. An intersection of the results identified 12 differentially expressed
155 genes (10 annotated genes) in not-pregnant heifers relative to the AI-preg and NB-preg groups
156 that also gained coexpression with 27 miRNAs (Fig. 4c). Notably, all genes that had higher
157 expression in not-pregnant heifers (*CEACAM4*, *FCGR2*, *ISG20*, *JAML*, *LSP1*, *NCF1*, *NFE2*,
158 *PLEKHO1*) showed a gain in negative coexpression with specific miRNAs. By contrast *MGAT4D*
159 was lower expressed in not-pregnant heifers and gained a positive coexpression with a miRNA in
160 not-pregnant heifers.

161 ***Assessment of transcript levels as predictors of pregnancy outcome***

162 Motivated by the observations that several genes have their abundance altered in
163 relationship to the fertility potential pregnancy (Fig. 4a), we asked whether transcript abundances
164 could serve as predictor features across datasets. We generated gene expression values for a
165 similar dataset we produced previously (GSE103628[34]), which also consisted of transcriptome
166 data from PWBCs collected from heifers at the time of AI and were classified according to their
167 pregnancy outcome as AI-preg (n=6) or not-pregnant (n=6). This dataset was particularly fitting
168 for this interrogation because the sampling occurred one year earlier on the same site. We
169 combined the expression values of the previously produced dataset (year one) with those of the
170 current dataset (year two) and contrasted the transcript abundance between heifers that were
171 categorized as AI-preg versus not pregnant accounting for the different years as a fixed effect in
172 the model. This analysis identified 198 genes whose abundance were significantly different
173 between the two experimental groups (AI-preg, not pregnant, FDR<0.03, Fig. 5a, Supplementary
174 table 13).

175 We used the transcript abundance values for these 198 genes obtained from samples
176 collected in year one as input (training set) for machine learning algorithms to derive models of
177 prediction to discriminate pregnancy outcome (AI-preg or not-pregnant, Fig. 5b). Next, we used
178 the transcript abundance values for these 198 genes obtained from samples collected in year two
179 as input to the algorithms for a blind classification of pregnancy outcome (AI-preg or not-pregnant,
180 Fig. 5c). The accuracy of classifications was assessed based on the known pregnancy outcome,
181 and parallelized random forest was the algorithm that produced the greatest proportion of correct
182 classification of all 11 heifers tested (46.3% out of 2000 randomizations), followed by 53.1%
183 correct classification for 10 out of 11 heifers (Fig. 5d). Notably, the top five genes that showed the
184 largest degree of importance for the classification were *RPL39*, *SMIM26*, *LONRF3*, *GATA3*,
185 *N6AMT1* (Fig. 5e), of which, *RPL39*, *SMIM26*, *N6AMT1* were up-regulated and *LONRF3* and
186 *GATA3* were down-regulated in AI-preg heifers relative to the not- pregnant group (Fig. 5f).

187

188 **DISCUSSION**

189 Here, we report a multilayered transcriptome analysis of bloodborne transcripts sampled
190 on the day of AI. The main objective of our study was to understand gene regulatory networks[35]
191 occurring in PWBCs, and rewiring that can occur based on the heifer's fertility potential. We
192 focused on the day of artificial insemination because it is the last day prior to the fertilization, when
193 the heifer has the first opportunity to become pregnant. We interrogated mRNA and miRNA

194 abundance levels obtained from the same individuals using the coexpression framework, which
195 is a valuable systems biology approach to gain insights into the biology of complex traits[35].
196 Furthermore, using machine learning[36], we detected consistent altered gene transcript abundance
197 associated with infertility across independent datasets. Collectively, our results provide important
198 aspects of regulatory networks in PWBCs that shed light on the physiological status of an animal
199 and its preparedness for becoming pregnant.

200 Similar to work reported in humans[37], our gene network analysis focusing on mRNAs
201 detected several protein-coding genes with strong coexpression in PWBCs. Interestingly, nearly
202 all of the correlated expression inferred as statistically significant across all heifers (|r|>0.98,
203 FDR<0.02) were positive. As expected[38], in silico functional gene co-expression analysis
204 indicated groups of genes that are linked to specific biological processes. A contrast of
205 coexpressing pairs of genes between samples collected from not-pregnant heifers and the two
206 groups of heifers that became pregnant displayed a large number of connections present in not-
207 pregnant heifers that were not identified in both AI-preg and NB-preg heifers. Our strict criteria
208 (|r|>0.99, eFDR<0.002) utilized to identify these altered connections indicate functional regulatory
209 relationships between selected genes.

210 Surprisingly, a great majority of the altered connections between gene pairs were present
211 gained in not-pregnant heifers, while there was minor loss of significant connections in not-
212 pregnant heifers relative to the AI-preg and NB-preg groups. The enrichment for specific biological
213 processes of genes that formed new negative connections was a clear evidence of the biological
214 relevance of these new connections. Notably, among genes annotated with proteolytic activity,
215 *ADAM19*[39], *ASRGL1*[40], *ATG4B*[41], *DPEP2*[42], *IMMP2L*[43], *PSMB9*[44], *SERPINE2*[45], *UCHL3*[46],
216 *USP42*[47] have been implicated in female infertility. A more drastic change in patterns of
217 coexpression was observed with the inversion of gene connections (significant pairwise
218 coexpression in one group with significant coexpression with opposite direction in another group).
219 These gene circuits changing from activation to repression are explained by current models of
220 incoherent feed-forward loops[48]. Cumulatively, the large occurrence of new connections and the
221 inverted connections created in PWBCs exclusively in not pregnant heifers indicate that gene
222 regulation is altered as an adaptative response to a stimulus.

223 Our profiling of small RNAs from plasma identified 263 annotated miRNAs, out of which
224 131 and 122 were previously identified in sera and circulating exosomes, respectively, among
225 other cattle organs[49]. Only 38 annotated miRNAs were profiled in peripheral blood mononucleated
226 cells in humans[50], which demonstrates that our plasma samples were depleted of intact cells.

227 Immuno-related cells composing the PWBC population can uptake circulating small
228 RNAs[51]. Although our data did not allow us to trace the origin of the tissue that produced the
229 miRNAs, we detected 33 circulating miRNAs whose abundance are associated with the
230 abundance of 106 protein-coding genes. As reported in humans[52], we quantified positive and
231 negative correlated abundance between miRNA and protein-coding genes. Negative correlations
232 mostly occur due to the targeting of mRNA by a specific miRNA. Indeed, we found several pairs
233 of miRNA:mRNA negatively correlated that had been identified as interacting pairs that fit this
234 model. By comparison, the positive coexpression between miRNA:mRNA pairs can be explained
235 by a miRNA targeting an intermediate gene, which in turn is negatively associated with a third
236 gene, thereby creating a positive correlation between a miRNA:mRNA pair (see Fig. 3 in
237 reference[53]). In our results, we only identify examples of miRNA:mRNA pairs that follow the
238 intermediate action of miRNAs when relaxing the stringent criteria for inference of significant
239 coexpression. Regardless of the model of interaction, it was notable that many protein-coding
240 genes that were coexpressed with miRNAs were enriched for specific biological processes,
241 including "regulation of gene expression". These results add further evidence to the growing body
242 of literature suggesting that circulating miRNAs act as mediators of cellular function[54,55].

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

251 Critical aspects of female fertility are tied to an appropriate regulation of the immune
252 system. Maternal immune tolerance to pregnancy starts with the exposure to proteins in the
253 semen[56], and is carried throughout placentation and pregnancy establishment[57]. Notably, 24 out
254 of 26 differentially expressed genes between pregnant (AI-preg and NB-preg) and not-pregnant
255 heifers showed greater abundance in not-pregnant heifers relative to the other two groups. (AI-
256 preg and NB-preg). While we did not detect differential abundance of miRNAs relative to
257 pregnancy outcome, the gain of coexpression between miRNAs and mRNAs from differentially
258 expressed genes functionally linked with immunology is in line with regulatory roles that certain
259 miRNAs exert on immune cells[58]. This result lead us to hypothesize that the heifers in the not-

260 pregnant group have a heightened activity of specific cells or functions in the immune system
261 relative to the heifers that became pregnant. It remains unclear, however, if this possible higher
262 activity was stimulated by extrinsic factors or is within the naturally-occurring variation[59].

263 The results of differential gene expression and differential coexpression fostered a
264 hypothesis that transcript abundance in the PWBCs can be used as predictors of pregnancy
265 outcome. We tested this hypothesis by applying machine learning algorithms on two groups of
266 heifers that were bred in 2015 (year one) and 2016 (year two). Parallel random forest emerged
267 as the algorithm with over 90% efficiency of classification nearly all trials executed, which confirms
268 the potential of accurate classification of samples using RNA-seq data under the case-control
269 framework[60,61]. The results show that while not one single gene emerges as a potential biomarker,
270 the accumulated information of transcript abundance from multiple genes can be powerful for the
271 identification of fertility potential in cattle.

272 Although we detected altered coexpression with high confidence, and in some instances,
273 with significant enrichment of genes in specific gene ontology terms, the consequence of the
274 differential coexpression on those biological functions remains undetermined. In addition, it
275 remains unclear whether the altered patterns of gene expression profiles are connected to
276 immune functions that are related to infertility[62]. The data analyzed here does not allow us to infer
277 causality[63], however our results converge to a strong indication that differences in gene
278 expression profiles are associated[63] with fertility in cattle.

279 In summary, analyses of mRNA from PWBCs and circulating extracellular miRNA by
280 multiple approaches (differential gene expression, differential gene coexpression and machine
281 learning classification) provided complementary evidence that gene expression profiles in the
282 PWBCs are associated with the fertility potential in beef heifers. The overall difference detected
283 in transcript abundance for specific genes is reproducible across samples and show promise for
284 the development of predictors of fertility potential in beef cattle.

285

286 **METHODS**

287 **Heifer classification according to pregnancy outcome**

288 All procedures with animals were performed in accordance with protocols approved by
289 Institutional Animal and Care and Use Committee at Auburn University.

290 Crossbred female calves (n=32, Angus-Simmental, *Bos taurus*) were developed as
291 replacement heifers[64,65] at the Black Belt Research and Extension Center (Auburn University).
292 Thirty-four days prior to AI, heifers were evaluated for reproductive tract score (scale of 1-5[66]) and
293 body condition score (BCS; scale of 1-9[67]) for the assertion that morphological and physiological
294 conditions were suitable for breeding. One animal was eliminated from the breeding herd at this
295 time.

296 Thirty one animals, 14 months of age on average, underwent estrous synchronization for
297 fixed-time artificial insemination utilizing the 7-Day Co-Synch + CIDR protocol[68]. An experienced
298 veterinarian inseminated all heifers with semen from sires of proven fertility. Fourteen days after
299 insemination, heifers were exposed to two fertile bulls for natural breeding for 60 days. An
300 experienced veterinarian performed pregnancy evaluation by transrectal palpation on days 69
301 and 120 post artificial insemination. Presence or absence of a conceptus, alongside
302 morphological features indicating fetal age were used to classify heifers as pregnant to AI (AI-
303 pregnant), pregnant to natural breeding (NB-pregnant), or not-pregnant.

304 **Collection, processing and preservation of biological material**

305 Immediately after AI, 10ml of blood was drawn from the jugular vein into vacutainers
306 containing 18mg K2 EDTA (Becton, Dickinson and Company, Franklin, NJ). The tubes were
307 inverted to prevent blood coagulation and immersed in ice for approximately 4 hours. Plasma and
308 buffy coat were separated by centrifugation for 10 minutes at 2000×g at 4°C. Plasma was
309 removed (~1.5ml) and centrifuged 10 minutes at 500×g at 4°C for further removal of cells. The
310 plasma devoid of cells was then transferred to a new microcentrifuge tube and preserved at -
311 80°C. The buffy coat was removed and deposited into 14ml of red blood cell lysis solution (0.15
312 M ammonium chloride, 10mM potassium bicarbonate, 0.1mM EDTA, Cold Spring Harbor
313 Protocols) for 10 minutes at room temperature (24-25°C). After a centrifugation, 800xg for 10
314 minutes, the solution was discarded and the pellet containing PWBCs was re-suspended in 200
315 μl of RNAlater® (Lifetechnologies™, Carlsbad, CA). PWBC samples were stored at -80°C.

316 **RNA extraction, library preparation, and RNA sequencing**

317 Total RNA was isolated from PWBCs of 18 heifers (AI-pregnant (n=6), NB-pregnant (n=6),
318 and not-pregnant (n=6)) using TRIzol™ reagent (Invitrogen, Carlsbad, CA)[69,70] following the
319 manufacturer's protocol. RNA yield was quantified using the Qubit™ RNA Broad Range Assay
320 Kit (Eurogene, OR) on a Qubit® Fluorometer, and integrity was assessed on an Agilent 2100
321 Bioanalyzer (Agilent, Santa Clara, CA) using an Agilent RNA 6000 Nano kit (Agilent, Santa Clara,

322 CA). Aproximatly 500 μg of total RNA was used as input for the TruSeq Stranded mRNA Library
323 Prep kit (Illumina, Inc., San Diego, CA); and library preparation was carried out following
324 manufacturer's instructions. Libraries were quantified with the Qubit™ dsDNA High Sensitivity
325 Assay Kit (Eurogene, OR); and quality was evaluated using the High Sensitivity DNA chip (Agilent,
326 Santa Clara, CA) on an Agilent 2100 Bioanalyzer. Libraries were sequenced in a HiSeq2500
327 system at the Genomic Services Laboratory at HudsonAlpha, Huntsville, AL to generate 125nt
328 long, pair-end reads.

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

336 Total RNA obtained from plasma was desiccated to a volume of 2.5μl for library
337 preparation with the TruSeq Small RNA Library Prep kit (Illumina, Inc., San Diego, CA) following
338 manufacturer's instructions, with the minor alteration that reagent volumes were halved in all
339 reactions until reverse transcription was performed, as described elsewhere[71]. After PCR
340 amplification, libraries were loaded onto a 10% polyacrylamide gel[72] and electrophoresis was
341 carried out for three hours at 100 volts alongside a 25/100 bp mixed DNA ladder (Bioneer,
342 Alameda, CA); and the custom DNA ladder included in the TruSeq Small RNA Library Prep kit.
343 The gel was soaked in GelRed® Nucleic Acid Gel Stain (Biotium, Inc., Fremont, CA) for DNA
344 staining. DNA bands in the range of 145 -160 base pairs were isolated from the gel by the crush-
345 and-soak method[73] with the following modifications. Gel fragments were placed into a 0.5 ml tube
346 containing small holes at the bottom, mounted into a 2ml tube. The tubes were centrifuged at
347 14,000xg for two minutes at room temperature. The crushed gel was soaked in 200μl of water at
348 4°C overnight, followed by removal of the solution containing the library to a new tube.

349 Concentration of each library was determined by Qubit™ dsDNA High Sensitivity Assay
350 Kit, and the volume corresponding to 1.5ng of each library was pooled for further purification with
351 the DNA Clean & Concentrator™-5 (ZymoResearch, Irvine, CA) following the manufacturer's
352 instructions. The pool was quantified with Qubit™ dsDNA High Sensitivity Assay Kit, and the DNA
353 profile was evaluated using the High Sensitivity DNA chip on an Agilent 2100 Bioanalyzer.

354 Libraries were sequenced in a NextSeq system at the Genomic Services Laboratory at
355 HudsonAlpha, Huntsville, AL to generate 75nt long, single end reads.

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

358 **RNA sequencing data processing**

359 Sequences were submitted to a custom build bioinformatics pipeline[74,75,76]. Reads were
360 aligned to the bovine genome (ARS-UCD1.2 obtained from Ensembl[77]) using HISAT2[78].
361 Sequences aligning to multiple places on the genome, or with 5 or more mismatches, were filtered
362 out. The sequences were then marked for duplicates, and non-duplicated pairs of reads (mRNA)
363 or single reads (miRNA) were used for transcript quantification at the gene level. The fragments
364 were counted against the Ensembl gene annotation[79] (version 1.87) using featureCounts[80].
365 RNAseq data from one animal (not-pregnant) were removed from analysis due to low sequencing
366 yield (613,765 reads).

367 **Filtering of lowly expressed genes**

368 All analytical procedures were carried out in R software[81]. Supplementary Code contain
369 all scripts used for analytical procedures described below and the creation of the plots, which can
370 also be obtained on figshare[82].

371 Counts per million (CPM), and fragments per kilobase per million reads (FPKM) were
372 calculated using "edgeR"[83]. In order to remove quantification uncertainty associated to lowly
373 expressed genes and erroneous identification of differentially expressed genes[84,85], we retained
374 genes with at least 2 CPM for PWBC mRNA or at least 1 CPM for plasma miRNA in five or more
375 samples for downstream analyses. Further analytical procedures were carried out with genes
376 annotated as "protein-coding" and "miRNA" for the PWBC RNA-sequencing and plasma small
377 RNA-sequencing, respectively.

378 **Pair-wise gene coexpression and differential coexpression for protein-coding**
379 **genes**

380 First, we calculated pair-wise Pearson's correlation coefficient (r)[86,87] for all 17 animals
381 sampled. Significance of the correlations was assessed using the "TestCor" package based on
382 the empirical statistical test and the large-scale correlation test with normal approximation[88] for
383 the calculation of the false discovery rate (FDR). Further biological interpretation of results was

384 performed on pairs of genes if abs(r)>0.98, which corresponded to a nominal P≤1.2x10^{-16} and
385 FDR<0.02.

386 Next, we calculated within group pair-wise Pearson's correlation coefficient (*r*) for each
387 group: AI-pregnant, NB-pregnant, and not-pregnant. We interpreted a differential correlation as
388 biologically meaningful based on the following scenarios, assuming that groups 1 and 2 can be
389 assigned arbitrarily: inverted correlation, if $|r_{group\ 1} - r_{group\ 2}| > 1.95$; gained correlation in group
390 1 and lost correlation in group 2, if $|r_{group\ 1}| > 0.99$ and $-0.1 < r_{group\ 2} < 0.1$.

391 Significance was estimated by the eFDR approach[75,89,90]. We resampled (5000
392 randomizations) the labels of pregnancy outcome and calculated *r* for each pair of genes in the
393 scrambled data set ($r_{scramble}$). eFDR was obtained based on the proportion of the $r_{scramble}$ greater
394 than a specified *r* (i.e.: for r=0.99, eFDR<0.002, Supplementary Fig. 1), relative to the total number
395 of correlations calculated using the formulas described elsewhere[75,89,90].

396 **mRNA-miRNA coexpression and differential coexpression**

397 We calculated Pearson's correlation (*r*)[86,87] values between miRNA and mRNA genes for
398 all 17 samples. Significance was estimated by the eFDR approach[75,89,90] by shuffling (5000
399 randomizations) the animal identification on mRNA dataset, thus breaking the common source of
400 both RNA types (Supplementary Fig. 2).

401 Pearson's correlation (r) values were calculated for mRNA-miRNA transcript pairs within
402 groups: AI-pregnant, NB-pregnant, and not-pregnant heifers. We interpreted a differential
403 correlation as biologically meaningful based on the following scenarios $|r_{group\ 1}| > 0.99$ and
404 $-0.1 < r_{group\ 2} < 0.1$. Significance was assessed by the eFDR approach (Supplementary Fig. 3).
405 We further interrogated the data for experimentally validated miRNA-mRNA interactions obtained
406 from miRWalk database (version 3)[91,92].

407 **Differential expression of mRNA and miRNA**

408 Differences of transcript levels between groups based on the pregnancy outcome were
409 determined from fragment counts using "edgeR"[83] and "DESeq2"[93]. For each comparison, a gene
410 was initially inferred as differentially expressed if the nominal P value was ≤0.03 in both "edgeR"
411 and "DESeq2" analyses. This nominal P value corresponded to eFDR≤0.05 (Supplementary Fig.
412 4), as calculated according to the procedure and formulas outlined elsewhere[75,89,90] using 10000
413 randomizations of sample reshuffling. Genes inferred as differentially expressed based on P
414 values were further selected through "leave-one-out" cross validation[94,95]. Only genes whose

415 significance were maintained after all rounds of sampling were further considered for biological
416 interpretation.

417 **Testing for enrichment of gene ontology terms and KEGG pathways**

418 Differentially expressed genes and genes that displayed differential coexpression among
419 heifers of AI-pregnant, NB-pregnant, and not-pregnant classifications were analyzed for
420 enrichment of Gene Ontology[96] categories and KEGG Pathways[97]. In all tests, the background
421 gene list consisted of all genes detected as expressed in our dataset[98]. GO annotations and
422 transcript lengths were obtained from BioMart[99] and KEGG annotation was obtained from the
423 "org.Bt.eg.db" package. We used "GoSeq" package[100] to estimate significance of category
424 enrichment using the sampling method[100]. We utilized FDR for multiple testing under
425 dependency[101], and inferred statistical significance of GO terms and KEGG pathways if FDR <
426 0.10.

427 **Assessment of predictive potential of gene expression levels by machine learning**
428 **algorithms.**

429 We focused the mRNA sequencing data from AI-pregnant and not-pregnant heifers from
430 this report and data we collected from the same site on the previous year and published
431 elsewhere[34,76]. The raw sequences from the previous dataset were realigned, and data were
432 treated as described above.

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

439 For the machine learning approach, we used the expression data for the 198 DEGs from
440 year one (2015) as a training set and the data for the same group of genes from year two (2016)
441 as a test set. Initially we processed the training set for the development of models and assessed
442 the prediction accuracy for 86 classification algorithms (Supplementary Fig. 5). Accuracy was
443 determined by the number of samples correctly classified divided by the number of samples tested
444 (n=11). The models that showed prediction accuracy ≥0.9 (classified correctly at least 10
445 samples, accuracy P value ≤0.01), were further subjected to 2000 random predictions with
446 different initial 'seed' numbers.

447 **FIGURE LEGEND**

448 Fig. 1: Overview of the experimental design and data. a Schematics of the experimental design,
449 classification scheme based on fertility, data produced and overview of the analyses. b
450 Dimensionality reduction of the transcriptome data obtained from the mRNA abundance of
451 protein-coding genes in PWBCs and plasma extracellular miRNA.

452

453 Fig. 2: mRNA:mRNA coexpression in PWBCs. a Network of genes with highly correlated mRNA
454 abundance. b Coexpression of 1860 protein-coding genes in pregnant heifers (AI-preg and NB-
455 preg). c Profile of the differential coexpression in heifers classified as not-pregnant relative to the
456 counterparts that became pregnant. d Scatterplot of selected genes associated with "proteolysis"
457 that were coexpressed in heifers classified as not-pregnant relative to the counterparts that
458 became pregnant. e Scatterplot of selected genes associated with "negative regulation of
459 transcription by RNA polymerase II" that showed inverted coexpression in heifers classified as
460 not-pregnant relative to the counterparts that became pregnant.

461

462 Fig. 3: miRNA from plasma coexpressed with mRNA in PWBCs. a miRNA:mRNA pairs
463 significantly coexpressed whose genes are significantly enriched in specific biological processes.
464 b miRNA:mRNA pairs significantly coexpressed with evidence of interaction from miRBase. c
465 Scatterplots of selected pairs representing differential coexpression between miRNA:mRNA in
466 heifers classified as not-pregnant relative to the counterparts that became pregnant

467

468 Fig. 4: Differential coexpression among heifers of different reproductive outcome. a Heatmap
469 depicting the expression levels of genes differentially expressed according to specific contrasts.
470 b Transcript abundance of the genes differentially expressed on both contrasts: AI-preg vs not-
471 pregnant and NB-preg vs not-pregnat. c miRNA:mRNA pairs significantly coexpressed whose
472 protein-coding genes are differentially expressed in PWBCs based on pregnancy outcome.

473

474 Fig. 5: Prediction of pregnancy outcome based on mRNA levels and machine learning algorithms.
475 a Representation of the contrasting of gene expression levels of heifers classified as AI-preg and
476 not-pregnant from two years and the number of differentially expressed genes (DEGs). b
477 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.

DATA AVAIABILITY

All raw sequencing files and raw read counts are deposited in the National Center for Biotechnology Information under the Gene Expression Omnibus database, accession: GSE146041 and GSE103628. All the intermediate data and results produced from raw files can be reproduced based on the codes available on Supplementary Code.

AUTHOR CONTRIBUTIONS

F.B. conceived designed and supervised the experiments. 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, supervised the experiments, contributed to data analysis and wrote the final draft.

499 **REFERENCES**

500

501 1. Zegers-Hochschild F*, et al.* The International Glossary on Infertility and Fertility Care,
502 2017. *Fertil Steril* **108**, 393-406 (2017).
503

504 2. Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility
505 prevalence and treatment-seeking: potential need and demand for infertility medical care.
506 *Human reproduction (Oxford, England)* **22**, 1506-1512 (2007).
507

508 3. Dickinson SE*, et al.* Evaluation of age, weaning weight, body condition score, and
509 reproductive tract score in pre-selected beef heifers relative to reproductive potential. *J*
510 *Anim Sci Biotechnol* **10**, 18 (2019).
511

512 4. Gajbhiye R, Fung JN, Montgomery GW. Complex genetics of female fertility. *NPJ*
513 *Genom Med* **3**, 29 (2018).
514

515 5. Peters SO*, et al.* Heritability and Bayesian genome-wide association study of first service
516 conception and pregnancy in Brangus heifers. *Journal of Animal Science* **2013**, 605-612
517 (2013).
518

519 6. Neupane M*, et al.* Loci and pathways associated with uterine capacity for pregnancy and
520 fertility in beef cattle. *PLoS One* **12**, e0188997 (2017).
521

522 7. McDaneld TG*, et al.* Genomewide association study of reproductive efficiency in female
523 cattle. *Journal of Animal Science* **2014**, 1945-1957 (2014).
524

525 8. McDaneld TG*, et al.* Y are you not pregnant: Identification of Y chromosome segments
526 in female cattle with decreased reproductive efficiency. *Journal of Animal Science* **90**,
527 2142-2151 (2012).
528

529 9. de Camargo GM, Costa RB, de Albuquerque LG, Regitano LC, Baldi F, Tonhati H.
530 Association between JY-1 gene polymorphisms and reproductive traits in beef cattle.
531 *Gene* **533**, 477-480 (2014).
532

533 10. Dias MM*, et al.* Study of lipid metabolism-related genes as candidate genes of sexual
534 precocity in Nellore cattle. *Genet Mol Res* **14**, 234-243 (2015).
535

536 11. Irano N*, et al.* Genome-wide association study for indicator traits of sexual precocity in
537 Nellore cattle. *PLoS One* **11**, e0159502 (2016).
538

539 12. Junior GAO*, et al.* Genomic study and medical subject headings enrichment analysis of
540 early pregnancy rate and antral follicle numbers in Nelore heifers. *Journal of Animal*
541 *Science* **95**, 4796-4812 (2017).
542

543 13. Kosova G, Abney M, Ober C. Colloquium papers: Heritability of reproductive fitness
544 traits in a human population. *P Natl Acad Sci USA* **107 Suppl 1**, 1772-1778 (2010).

14. Fortes MRS*, et al.* Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *Journal of Animal Science* **2012**, 2894-2906 (2012).

15. Doyle SP, Golden BL, Green RD, Brinks JS. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. *Journal of Animal Science* **78**, 2091-2098 (2000).

16. Toghiani S, Hay E, Sumreddee P, Geary TW, Rekaya R, Roberts AJ. Genomic prediction of continuous and binary fertility traits of females in a composite beef cattle breed. *Journal of Animal Science* **95**, 4787-4795 (2017).

17. Bormann JM, Totir LR, Kachman SD, Fernando RL, Wilson DE. Pregnancy rate and first-service conception rate in Angus heifers. *Journal of Animal Science* **84**, 2022-2025 (2006).

18. McAllister CM, Speidel SE, Crews DH, Jr., Enns RM. Genetic parameters for intramuscular fat percentage, marbling score, scrotal circumference, and heifer pregnancy in Red Angus cattle. *Journal of Animal Science* **89**, 2068-2072 (2011).

19. Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, DeNise S. Genomic predictions in Angus cattle: Comparisons of sample size, response variables, and clustering methods for cross-validation. *Journal of Animal Science* **92**, 485-497 (2014).

20. Adams GP, Pierson RA. Bovine Model for Study of Ovarian Follicular Dynamics in Humans. *Theriogenology* **43**, 113-120 (1995).

21. Abedal-Majed MA, Cupp AS. Livestock animals to study infertility in women. *Anim Front* **9**, 28-33 (2019).

22. Henchion M, Hayes M, Mullen AM, Fenelon M, Tiwari B. Future Protein Supply and Demand: Strategies and Factors Influencing a Sustainable Equilibrium. *Foods* **6**, (2017).

23. Cushman RA, Kill LK, Funston RN, Mousel EM, Perry GA. Heifer calving date positively influences calf weaning weights through six parturitions. *Journal of Animal Science* **2013**, 4486-4491 (2013).

24. Marshall DM, Minqiang W, Freking BA. Relative calving date of first-calf heifers as related to production efficiency and subsequent reproductive performance. *Journal of Animal Science* **68**, 1812-1817 (1990).

25. Lesmeister JL, Burfening PJ, Blackwell RL. Date of first calving in beef cows and subsequent calf production. *Journal of Animal Science* **36**, 1-6 (1973).

590 26. Damiran D, Larson KA, Pearce LT, Erickson NE, Lardner BHA. Effect of calving period
591 on beef cow longevity and lifetime productivity in western Canada. *Translational Animal*
592 *Science* **2**, S61-S65 (2018).
593
594 27. Morris CA, Cullen NG. A note on genetic correlations between pubertal traits of males or
595 females and lifetime pregnancy rate in beef cattle. *Livestock Production Science* **39**, 291-
596 297 (1994).
597
598 28. Mwansa PB, Kemp RA, Jr DHC, Kastelic JP, Bailey DRC, Coulter GH. Selection for
599 cow lifetime pregnancy rate using bull and heifer growth and reproductive traits in
600 composite cattle. *Canadian Journal of Animal Science* **80**, 507-510 (2000).
601
602 29. Li-Pook-Than J, Snyder M. iPOP goes the world: integrated personalized Omics profiling
603 and the road toward improved health care. *Chem Biol* **20**, 660-666 (2013).
604
605 30. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating Gene
606 Expression with Summary Association Statistics to Identify Genes Associated with 30
607 Complex Traits. *Am J Hum Genet* **100**, 473-487 (2017).
608
609 31. Gusev A*, et al.* Integrative approaches for large-scale transcriptome-wide association
610 studies. *Nat Genet* **48**, 245-252 (2016).
611
612 32. Ioannidis J, Donadeu FX. Circulating microRNA profiles during the bovine oestrous
613 cycle. *PLoS One* **11**, e0158160 (2016).
614
615 33. Dickinson SE*, et al.* Transcriptome profiles in peripheral white blood cells at the time of
616 artificial insemination discriminate beef heifers with different fertility potential. *BMC*
617 *Genomics* **19**, 129 (2018).
618
619 34. Dickinson SE*, et al.* Transcriptome profiles in peripheral white blood cells at the time of
620 artificial insemination discriminate beef heifers with different fertility potential. *BMC*
621 *Genomics* **19**, (2018).
622
623 35. Emmert-Streib F, Dehmer M, Haibe-Kains B. Gene regulatory networks and their
624 applications: understanding biological and medical problems in terms of networks. *Front*
625 *Cell Dev Biol* **2**, 38 (2014).
626
627 36. Xu C, Jackson SA. Machine learning and complex biological data. *Genome Biol* **20**, 76
628 (2019).
629
630 37. Monaco G*, et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow
631 Absolute Deconvolution of Human Immune Cell Types. *Cell Rep* **26**, 1627-1640 e1627
632 (2019).
633

634 38. van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression
635 analysis for functional classification and gene-disease predictions. *Brief Bioinform* **19**,
636 575-592 (2018).
637
638 39. Sheridan MA*, et al.* Early onset preeclampsia in a model for human placental trophoblast.
639 *P Natl Acad Sci USA* **116**, 4336-4345 (2019).
640
641 40. Fitzgerald HC*, et al.* Idiopathic infertility in women is associated with distinct changes in
642 proliferative phase uterine fluid proteins. *Biol Reprod* **98**, 752-764 (2018).
643
644 41. Yang HL, Mei J, Chang KK, Zhou WJ, Huang LQ, Li MQ. Autophagy in endometriosis.
645 *Am J Transl Res* **9**, 4707-4725 (2017).
646
647 42. Pelch KE, Schroder AL, Kimball PA, Sharpe-Timms KL, Davis JW, Nagel SC. Aberrant
648 gene expression profile in a mouse model of endometriosis mirrors that observed in
649 women. *Fertil Steril* **93**, 1615-1627 e1618 (2010).
650
651 43. Demain LA, Conway GS, Newman WG. Genetics of mitochondrial dysfunction and
652 infertility. *Clin Genet* **91**, 199-207 (2017).
653
654 44. Koks S*, et al.* The differential transcriptome and ontology profiles of floating and
655 cumulus granulosa cells in stimulated human antral follicles. *Mol Hum Reprod* **16**, 229-
656 240 (2010).
657
658 45. Li SH*, et al.* Correlation of cumulus gene expression of GJA1, PRSS35, PTX3, and
659 SERPINE2 with oocyte maturation, fertilization, and embryo development. *Reprod Biol*
660 *Endocrinol* **13**, 93 (2015).
661
662 46. Mtango NR, Sutovsky M, Susor A, Zhong Z, Latham KE, Sutovsky P. Essential role of
663 maternal UCHL1 and UCHL3 in fertilization and preimplantation embryo development.
664 *J Cell Physiol* **227**, 1592-1603 (2012).
665
666 47. Ingham NJ*, et al.* Mouse screen reveals multiple new genes underlying mouse and human
667 hearing loss. *PLoS Biol* **17**, e3000194 (2019).
668
669 48. Reeves GT. The engineering principles of combining a transcriptional incoherent
670 feedforward loop with negative feedback. *J Biol Eng* **13**, 62 (2019).
671
672 49. Sun HZ, Chen Y, Guan LL. MicroRNA expression profiles across blood and different
673 tissues in cattle. *Sci Data* **6**, 190013 (2019).
674
675 50. Vaz C*, et al.* Analysis of microRNA transcriptome by deep sequencing of small RNA
676 libraries of peripheral blood. *BMC Genomics* **11**, 288 (2010).
677
678 51. Fritz JV*, et al.* Sources and Functions of Extracellular Small RNAs in Human
679 Circulation. *Annu Rev Nutr* **36**, 301-336 (2016).

680
681 52. Diaz G, Zamboni F, Tice A, Farci P. Integrated ordination of miRNA and mRNA
682 expression profiles. *BMC Genomics* **16**, 767 (2015).
683
684 53. Ritchie W, Rajasekhar M, Flamant S, Rasko JE. Conserved expression patterns predict
685 microRNA targets. *PLoS Comput Biol* **5**, e1000513 (2009).
686
687 54. Bayraktar R, Van Roosbroeck K, Calin GA. Cell-to-cell communication: microRNAs as
688 hormones. *Mol Oncol* **11**, 1673-1686 (2017).
689
690 55. Rayner KJ, Hennessy EJ. Extracellular communication via microRNA: lipid particles
691 have a new message. *J Lipid Res* **54**, 1174-1181 (2013).
692
693 56. Robertson SA, Sharkey DJ. The role of semen in induction of maternal immune tolerance
694 to pregnancy. *Semin Immunol* **13**, 243-254 (2001).
695
696 57. Moffett A, Loke C. Immunology of placentation in eutherian mammals. *Nat Rev*
697 *Immunol* **6**, 584-594 (2006).
698
699 58. Momen-Heravi F, Bala S. miRNA regulation of innate immunity. *J Leukoc Biol*, (2018).
700
701 59. Hughes HD, Carroll JA, Burdick Sanchez NC, Richeson JT. Natural variations in the
702 stress and acute phase responses of cattle. *Innate Immun* **20**, 888-896 (2014).
703
704 60. Wenric S, Shemirani R. Using Supervised Learning Methods for Gene Selection in RNA-
705 Seq Case-Control Studies. *Front Genet* **9**, 297 (2018).
706
707 61. Zararsiz G*, et al.* A comprehensive simulation study on classification of RNA-Seq data.
708 *PLoS One* **12**, e0182507 (2017).
709
710 62. Brazdova A, Senechal H, Peltre G, Poncet P. Immune Aspects of Female Infertility. *Int J*
711 *Fertil Steril* **10**, 1-10 (2016).
712
713 63. Altman N, Krzywinski M. Association, correlation and causation. *Nat Methods* **12**, 899-
714 900 (2015).
715
716 64. Patterson DJ, Smith MF. Management considerations in beef heifer development and
717 puberty. *Vet Clin North Am Food Anim Pract* **29**, xiii-xiv (2013).
718
719 65. Larson RL, White BJ, Laflin S. Beef Heifer Development. *Vet Clin North Am Food Anim*
720 *Pract* **32**, 285-302 (2016).
721
722 66. Holm DE, Thompson PN, Irons PC. The value of reproductive tract scoring as a predictor
723 of fertility and production outcomes in beef heifers. *J Anim Sci* **87**, 1934-1940 (2009).
724

725 67. Rae DO, Kunkle WE, Chenoweth PJ, Sand RS, Tran T. Relationship of parity and body
726 condition score to pregnancy rate in Florida beef-cattle. *Theriogenology* **1993**, 1143-1152
727 (1993).
728
729 68. Larson JE*, et al.* Synchronization of estrus in suckled beef cows for detected estrus and
730 artificial insemination and timed artificial insemination using gonadotropin-releasing
731 hormone, prostaglandin F2α, and progesterone. *Journal of Animal Science* **2006**, 332–
732 342 (2006).
733
734 69. Rio DC, Ares M, Jr., Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI
735 reagent). *Cold Spring Harb Protoc* **2010**, pdb prot5439 (2010).
736
737 70. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium
738 thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* **1**,
739 581-585 (2006).
740
741 71. Yeri A*, et al.* Total Extracellular Small RNA Profiles from Plasma, Saliva, and Urine of
742 Healthy Subjects. *Scientific Reports* **7**, 44061 (2017).
743
744 72. Chory J, Pollard JD, Jr. Separation of small DNA fragments by conventional gel
745 electrophoresis. *Curr Protoc Mol Biol* **Chapter 2**, Unit2 7 (2001).
746
747 73. Green MR, Sambrook J. Isolation of DNA Fragments from Polyacrylamide Gels by the
748 Crush and Soak Method. *Cold Spring Harb Protoc* **2019**, pdb prot100479 (2019).
749
750 74. Biase FH*, et al.* Massive dysregulation of genes involved in cell signaling and placental
751 development in cloned cattle conceptus and maternal endometrium. *P Natl Acad Sci USA*
752 **113**, 14492-14501 (2016).
753
754 75. Biase FH, Kimble KM. Functional signaling and gene regulatory networks between the
755 oocyte and the surrounding cumulus cells. *BMC Genomics* **19**, 351 (2018).
756
757 76. Dickinson SE, Biase FH. Transcriptome data of peripheral white blood cells from beef
758 heifers collected at the time of artificial insemination. *Data Brief* **18**, 706-709 (2018).
759
760 77. Kinsella RJ*, et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space.
761 *Database* **2011**, bar030 (2011).
762
763 78. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
764 requirements. *Nat Methods* **12**, 357-360 (2015).
765
766 79. Flicek P*, et al.* Ensembl 2014. *Nucleic Acids Res* **42**, D749-755 (2014).
767
768 80. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
769 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
770

81. Ihaka R, Gentleman. R: A Language and Environment for Statistical Computing. *Journal of Computational and Graphical Statistics* **5**, 299-214 (2012).

82. Biase FH.Supplementary code and files to Rewiring of gene expression in PWBC at artificial insemination is predictive of pregnancy outcome in cattle (Bos taurus). https://doi.org/10.6084/m9.figshare.11985666.v2.

83. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

84. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).

85. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

86. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D. How to infer gene networks from expression profiles. *Mol Syst Biol* **3**, (2007).

87. Serin EA, Nijveen H, Hilhorst HW, Ligterink W. Learning from Co-expression Networks: Possibilities and Challenges. *Front Plant Sci* **7**, 444 (2016).

88. Cai TT, Liu W. Large-Scale Multiple Testing of Correlations. *J Am Stat Assoc* **111**, 229-240 (2016).

89. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *P Natl Acad Sci USA* **100**, 9440-9445 (2003).

90. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet* **15**, 335-346 (2014).

91. 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 (2011).

92. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods* **12**, 697 (2015).

93. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

94. Picard RR, Cook RD. Cross-Validation of Regression-Models. *J Am Stat Assoc* **79**, 575-583 (1984).

95. Breiman L, Spector P. Submodel Selection and Evaluation in Regression - the X-Random Case. *Int Stat Rev* **60**, 291-319 (1992).

96. Ashburner M*, et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

97. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Molecular BioSystems* **10**, 2441-2447 (2014).

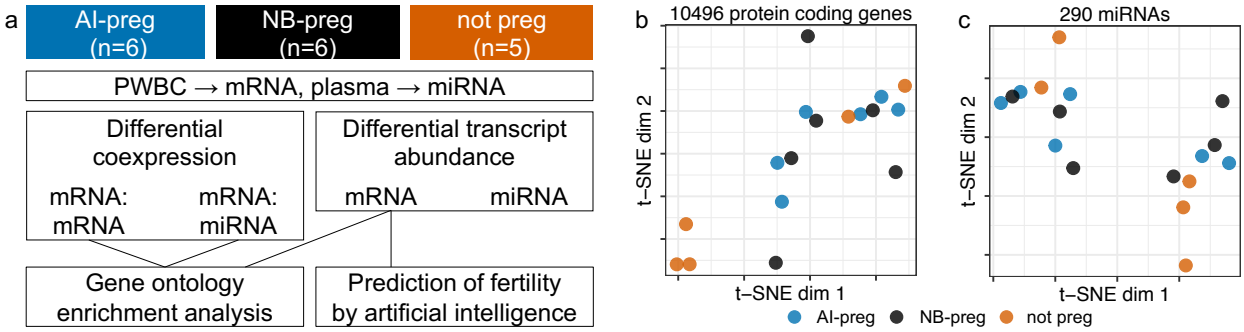
98. Timmons JA, Szkop KJ, Gallagher IJ. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol* **16**, 186 (2015).

99. Durinck S*, et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440 (2005).

100. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, R14 (2010).

101. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Statist* **29**, 1165-1188 (2001).

840 Figure 1



841

842

Figure 2

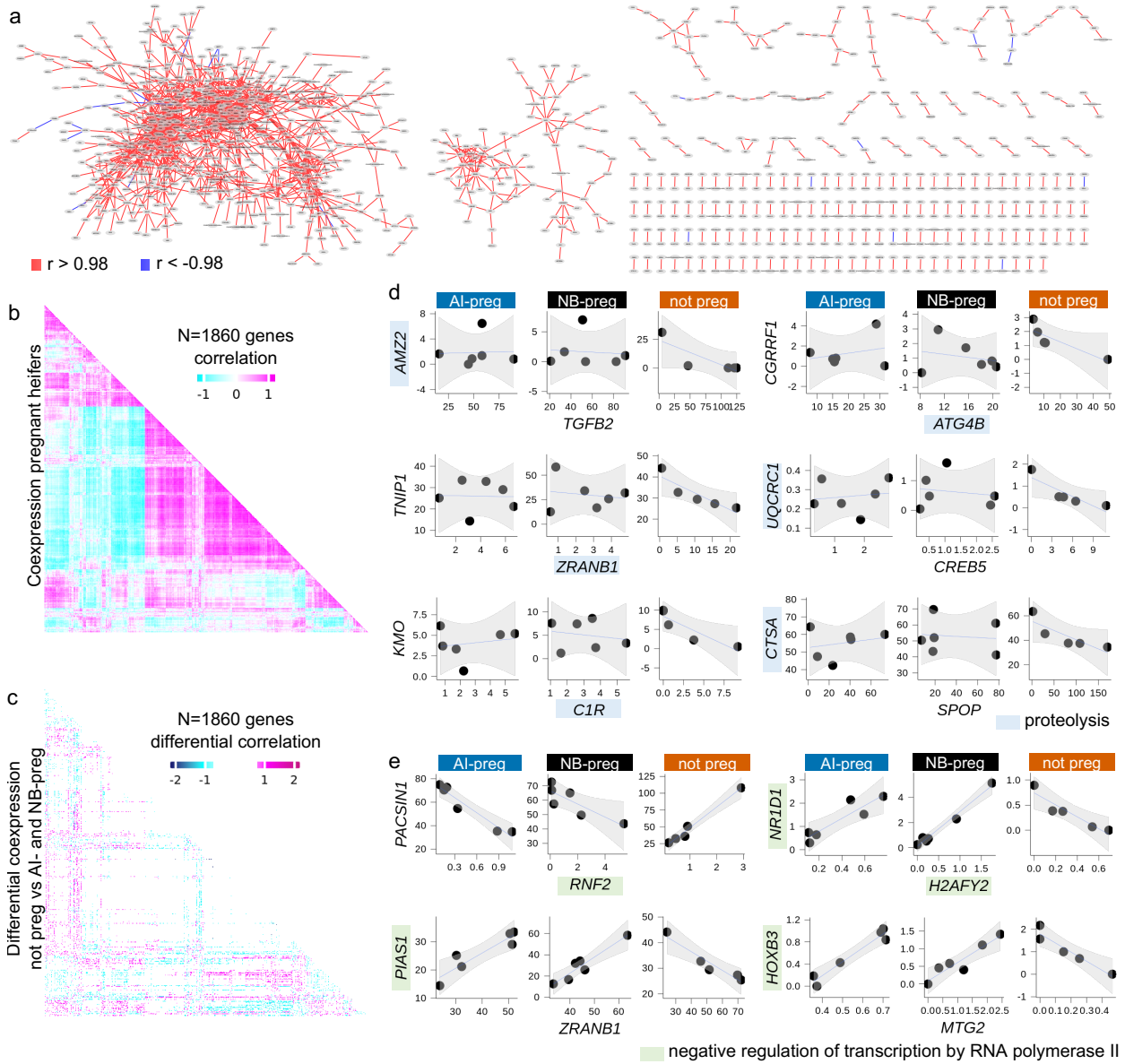

a
r > 0.98
r < -0.98
b
N=1860 genes
correlation
-1 0 1
Coexpression pregnant heifers
c
N=1860 genes
differential correlation
-2 -1 1 2
Differential coexpression
not preg vs AI- and NB-preg
d
AI-preg
NB-preg
not preg
AMZ2
TGFB2
CGRRF1
ATG4B
TNIP1
ZRANB1
UQCRC1
CREB5
KMO
C1R
CTSA
SPOP
proteolysis
e
PACSIN1
RNF2
NR1D1
H2AFY2
PIAS1
ZRANB1
HOXB3
MTG2
negative regulation of transcription by RNA polymerase II

Figure 3

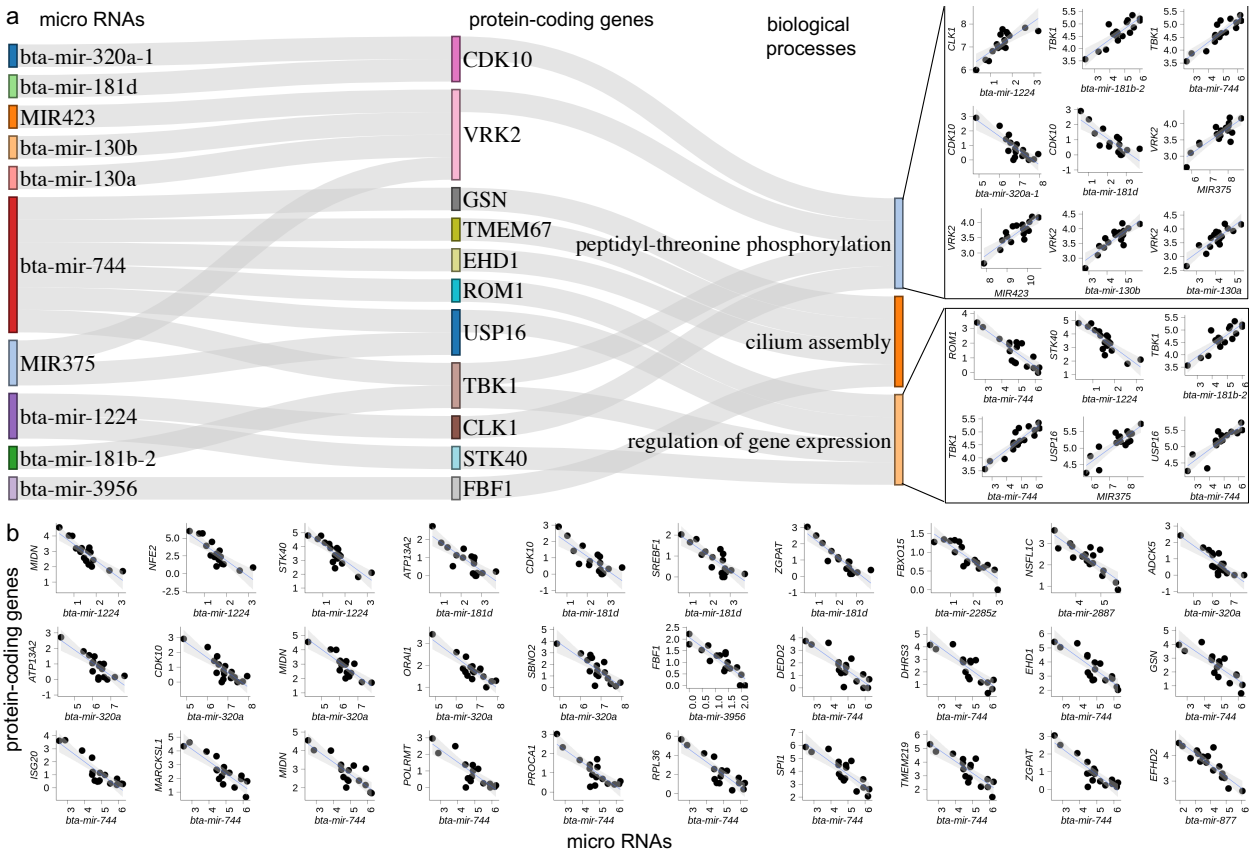

a
micro RNAs
protein-coding genes
biological processes
bta-mir-320a-1
bta-mir-181d
MIR423
bta-mir-130b
bta-mir-130a
bta-mir-744
MIR375
bta-mir-1224
bta-mir-181b-2
bta-mir-3956
CDK10
VRK2
GSN
TMEM67
EHD1
ROM1
USP16
TBK1
CLK1
STK40
FBF1
peptidyl-threonine phosphorylation
cilium assembly
regulation of gene expression
b
protein-coding genes
micro RNAs
c
AI-preg
NB-preg
not preg
ZDHHC3
bta-mir-424
CUEDC2
bta-mir-502b
Pregnant
Not pregnant
NSUN3
bta-mir-138-1
SMG8
MIR185

Figure 4

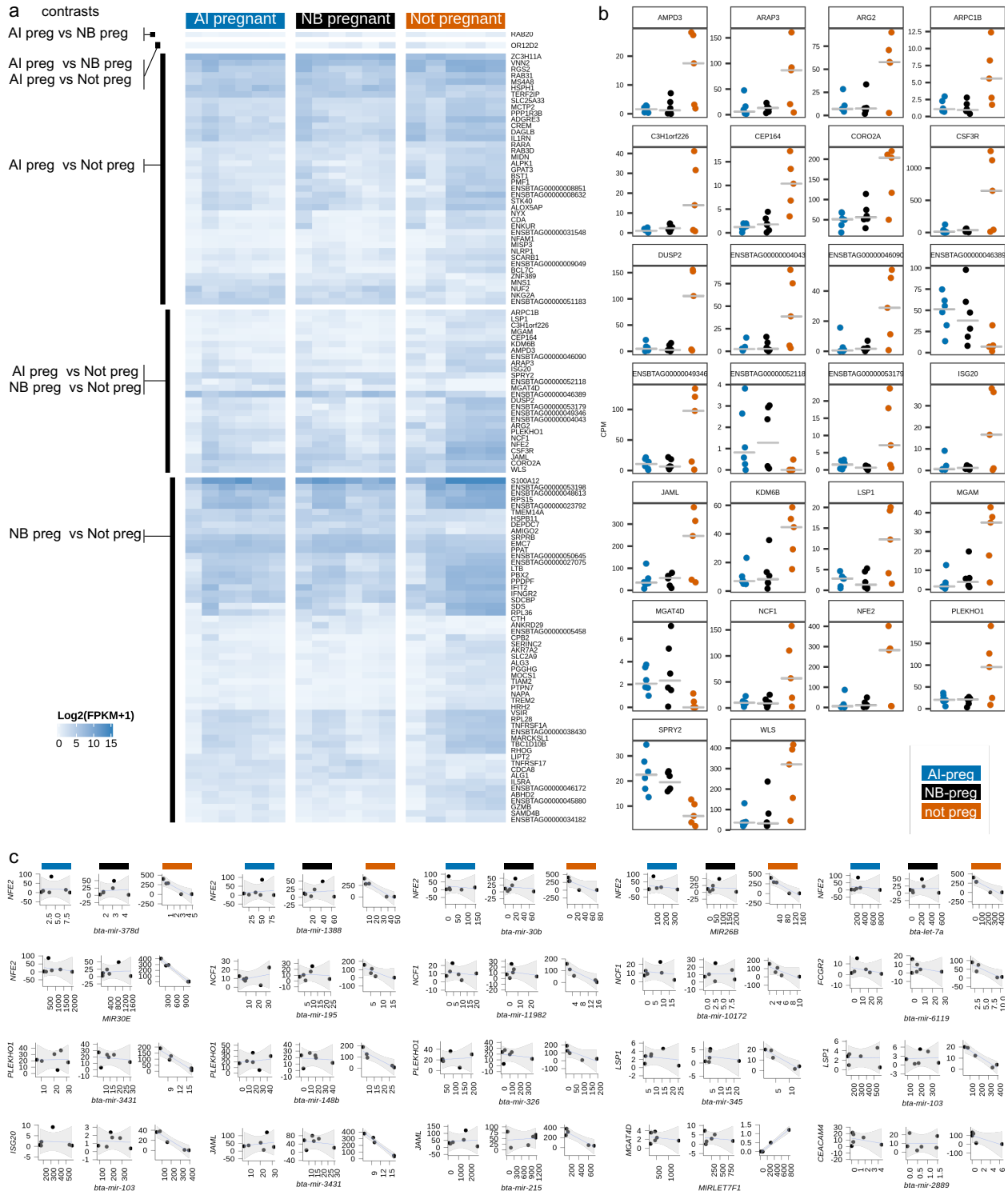
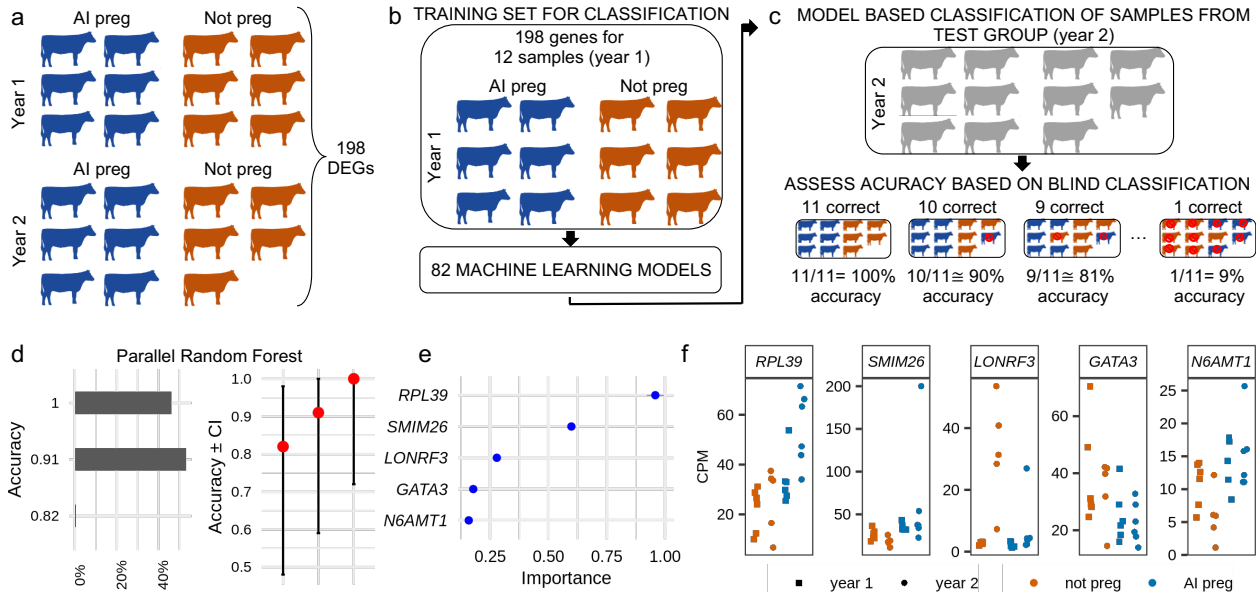

a
contrasts
AI preg vs NB preg
AI preg vs NB preg
AI preg vs Not preg
AI preg vs Not preg
AI preg vs Not preg
NB preg vs Not preg
AI pregnant
NB pregnant
Not pregnant
Log2(FPKM+1)
0 5 10 15
b
CPM
AI-preg
NB-preg
not preg
c
bta-mir-378d
bta-mir-1388
bta-mir-30b
MIR26B
bta-let-7a
MIR30E
bta-mir-195
bta-mir-11982
bta-mir-10172
bta-mir-6119
bta-mir-3431
bta-mir-148b
bta-mir-326
bta-mir-345
bta-mir-103
bta-mir-103
bta-mir-3431
bta-mir-215
MIRLET7F1
bta-mir-2889

Figure 5


a
AI preg
Not preg
Year 1
Year 2
198 DEGs
b
TRAINING SET FOR CLASSIFICATION
198 genes for 12 samples (year 1)
82 MACHINE LEARNING MODELS
c
MODEL BASED CLASSIFICATION OF SAMPLES FROM TEST GROUP (year 2)
ASSESS ACURACY BASED ON BLIND CLASSIFICATION
11 correct
10 correct
9 correct
1 correct
11/11= 100% accuracy
10/11≅ 90% accuracy
9/11≅ 81% accuracy
1/11= 9% accuracy
d
Parallel Random Forest
Accuracy
Accuracy ± CI
0%
20%
40%
e
RPL39
SMIM26
LONRF3
GATA3
N6AMT1
Importance
f
CPM
year 1
year 2
not preg
AI preg