Running title: Transcriptome profiles of infertility

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

Sarah E. Moorey[a], Bailey N. Walker[b], Michelle F. Elmore[c,d], Joshua B. Elmore[d], Soren P. Rodning[c] and Fernando H. Biase[b*]

[a] Department of Animal Science, University of Tennessee, 2506 River Drive Knoxville, TN, 37996

[b] Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, 175 West Campus Drive, Blacksburg, VA, 24061

[c] Department of Animal Sciences, Auburn University, 107 Comer Hall, Auburn, AL, 36849

[d] Alabama Cooperative Extension System, 107 Comer Hall, Auburn, AL, 36849

*Corresponding author:

Fernando Biase

Email fbiase@vt.edu Phone (540)231-9520

## ABSTRACT

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 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 transcript abundance in circulating white blood cells is associated with fertility in heifers.

Key words: infertility, systems biology, transcriptome

## INTRODUCTION

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 protein [1]. 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

46 increased productivity and longevity than their later calving herd mates [15-18]. Furthermore, the
47 genetic correlation between yearling pregnancy rate and lifetime pregnancy rate is high (0.92-
48 0.97) [19,20]. Therefore, the ability to identify heifers that experience optimal fertility during the first
49 breeding is essential to the sustainability of beef cattle production systems.The genetic
50 architecture of fertility in beef heifers is very complex [21-28]. In line with this complexity, female
51 reproductive traits have low heritability, with traits related to female reproductive fitness in cattle,
52 such as heifer pregnancy and first service conception, range from 0.07 to 0.20 [2,21,29-33] and 0.03
53 to 0.18 [2,21,29], respectively. The examination of the genetic components of fertility in beef heifers
54 have yielded several genes potentially associated with fertility traits [21-28]; but the effect of these
55 markers are minimal, and there is no clear redundancy in genetic markers identified across breeds
56 or subspecies.

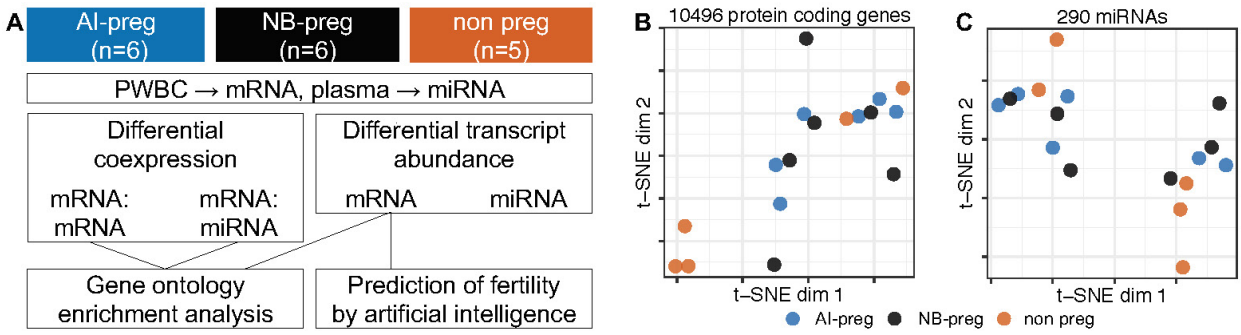
57 Beyond the genomic profiling, the analysis of multiple layers of an individual's molecular
58 blueprint is likely key for understanding the underlying biology of complex traits [34]. In line with this
59 rationale, expression-trait association studies have emerged as a means to better understand
60 complex traits [35,36]. Specifically related to fertility, there is evidence that circulating micro RNA
61 (miRNA) profiles are indicative of reproductive function [37]. Furthermore, we have previously
62 identified differences in the transcriptome of peripheral white blood cells (PWBC) among heifers
63 of different fertility potential [38].

64 In the present study, our objective was to investigate gene transcripts in PWBCs and
65 circulating micro RNA (miRNA) profiles in heifers enrolled in fixed time artificial insemination (AI).
66 We utilized a multi-dimensional approach to investigate differences in coexpression of protein-
67 coding genes in PWBCs, and coexpression between circulating miRNA and mRNAs in PWBCs.
68

69 **RESULTS**

70 ***Overview of the experiment and data generated***

71 We generated transcriptome data for mRNA from PWBC and small RNA from plasma
72 collected at the time of AI from heifers that became pregnant to AI (AI-preg, n=6), pregnant later
73 in the breeding season from natural breeding with bulls NB (NB, NB-preg, n=6), or failed to
74 become pregnant (non-pregnant, n=5, Fig. 1A). We produced, on average, 25.8 and 7.1 million
75 reads from mRNA and small RNA, respectively. Following the filtering for lowly expressed genes,
76 10496 protein coding genes in PWBC and 290 miRNAs in plasma samples were used for
77 differential gene expression, coexpression and prediction analyses (Fig. 1A). Notably, three of
78 the five non-pregnant samples clustered separately from other subjects based on the expression
79 levels of protein-coding genes (Fig. 1B). We note that all heifers were phenotypically similar; and
80 no marked distinction was observed among these three heifers relative to the remainder of the
81 group on the day of AI. Contrariwise, there was no separation of samples based on the miRNA
82 expression levels (Fig. 1C).

83 


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

88

89 ***Differential mRNA:mRNA coexpression associated with pregnancy outcome***

90 First, we asked whether there was a general coexpression profile of the protein-coding
91 genes expressed in PWBCs amongst all 17 samples. We identified that 1633 genes that formed
92 2017 pairs with highly significant correlated expression levels (|r|>0.98, FDR<0.02, Fig. 2A,
93 Supplementary Table S1). Notably, ~99% of those significant correlations were positive (Fig. 2A).
94 Several genes showing coexpression were associated with 'translation', 'positive regulation of
95 gene expression', 'B cell differentiation', and 'negative regulation of extrinsic apoptotic signaling
96 pathway' amongst other biological processes (FDR<0.1, Supplementary Table S2). This result
97 indicates the occurrence of gene regulatory networks in PWBCs that are biologically meaningful
98 for the immune system.

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

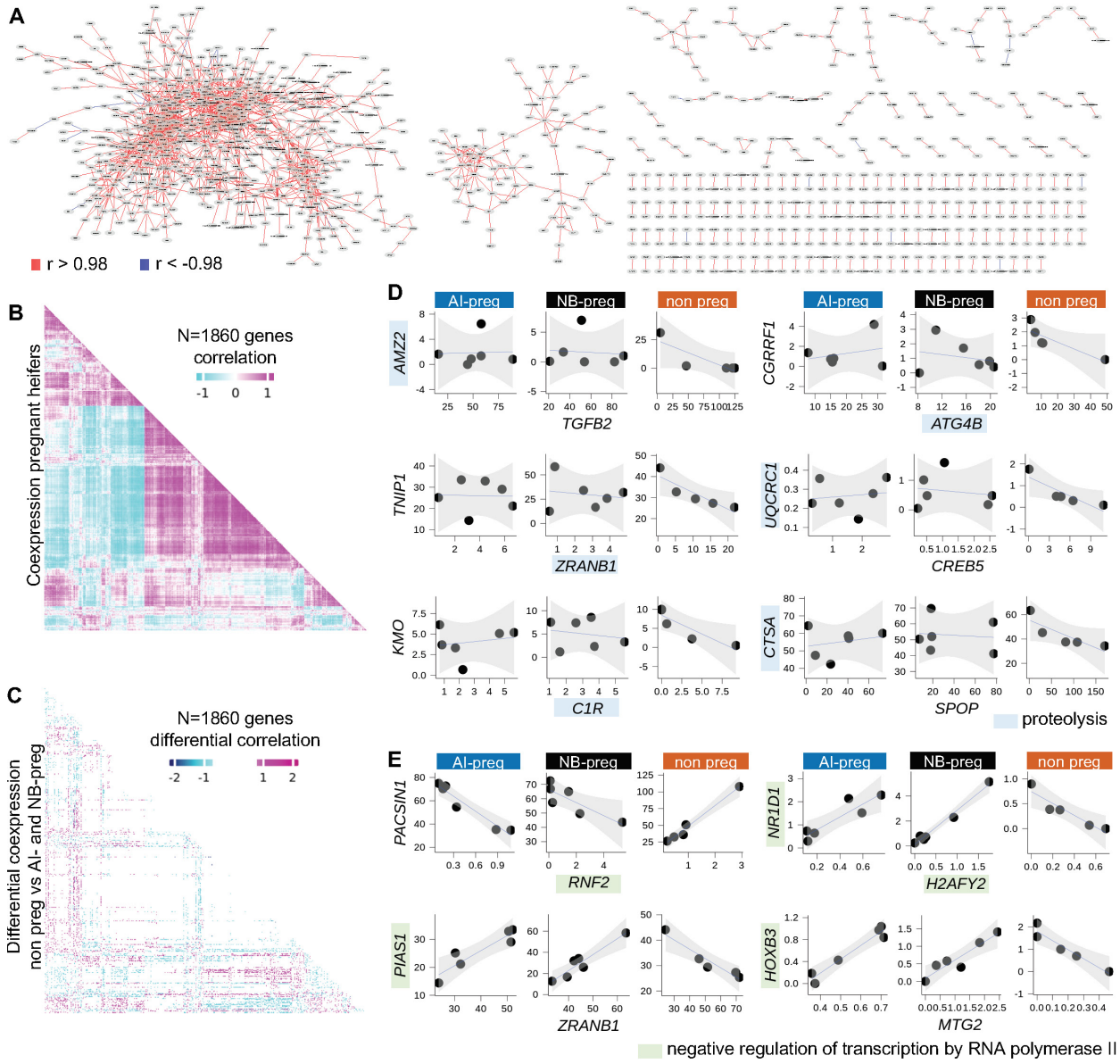



Fig. 2. mRNA:mRNA coexpression in PWBCs. (A) Network of genes with highly correlated mRNA abundance. (B) Coexpression of 1860 protein-coding genes in pregnant heifers (AI-preg and NB-preg). (C) Profile of the differential coexpression in heifers classified as non-pregnant relative to the counterparts 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

121 classified as non-pregnant relative to the counterparts that became pregnant. All scatterplots
122 present transcript abundance in Log2(counts per million).

123

124

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

126 The generation of data from small RNA (from plasma) and mRNA (from PWBCs) from the
127 same subjects (n=17) permitted us to ask whether circulating miRNAs have correlated expression
128 with protein-coding genes expressed in the PWBCs. We identified 141 pairs of miRNA:mRNA
129 with |r|>0.85 (eFDR<0.0001, Supplementary Fig. S2) that were formed by 106 protein-coding
130 genes and 33 miRNAs (Supplementary Table S6). Among the 106 protein-coding genes existing
131 in high correlation with miRNA, we identified three significantly enriched biological processes,
132 namely: “peptidyl-threonine phosphorylation”, “cilium assembly” and “regulation of gene
133 expression” (Fig. 3A, Supplementary Table S7). Of notice, 30 of the 141 miRNA:mRNA pairs that
134 presented negative correlation between the expression levels have been previously identified as
135 interacting pairs in the miRWalk database (Fig. 3B). Additionally, bta-mir-744 and bta-mir-320a
136 emerged as the two most connected miRNAs (13 and 6 connections, respectively). It is evident
137 that circulating miRNAs have regulatory roles over transcripts of protein-coding genes expressed
138 in the PWBCs.

139 We also interrogated miRNA:mRNA pairs for differential coexpression associated with
140 pregnancy outcome. We identified 2330 and 2738 miRNA:mRNA pairs with gain of negative and
141 positive correlated expression, respectively, (|r|>0.9, eFDR≤0.01, Supplementary Fig. S3) in non-
142 pregnant heifers that were not identified compared to pregnant heifers (both AI-preg and NB-preg,
143 Fig. 3C). In addition, there were 51, two and three connections showing inverted, loss of negative
144 and loss of positive correlations in non-pregnant heifers relative to the pregnant (both AI-preg and

145 NB-preg) counterparts (Fig. 3C, see Supplementary Table S8 for genes). Among the 1220
146 protein-coding genes associated with gain of negative correlation with miRNAs, there were
147 biological processes significantly enriched, namely: “immunoglobulin mediated immune
148 response”, “negative regulation of DNA-binding transcription factor activity”, “lymphocyte
149 chemotaxis” and “mitochondrial respiratory chain complex I assembly” (FDR<0.1, Supplementary
150 Table S9). Among the 1288 genes associated with gain of positive correlation with miRNAs, the
151 following biological processes were significantly enriched: “regulation of DNA replication” and
152 “cellular response to BMP stimulus” (FDR<0.1, Supplementary Table S10). These results indicate
153 that several miRNA and mRNA pairs have their connections altered in relation to the heifer’s
154 fertility potential.

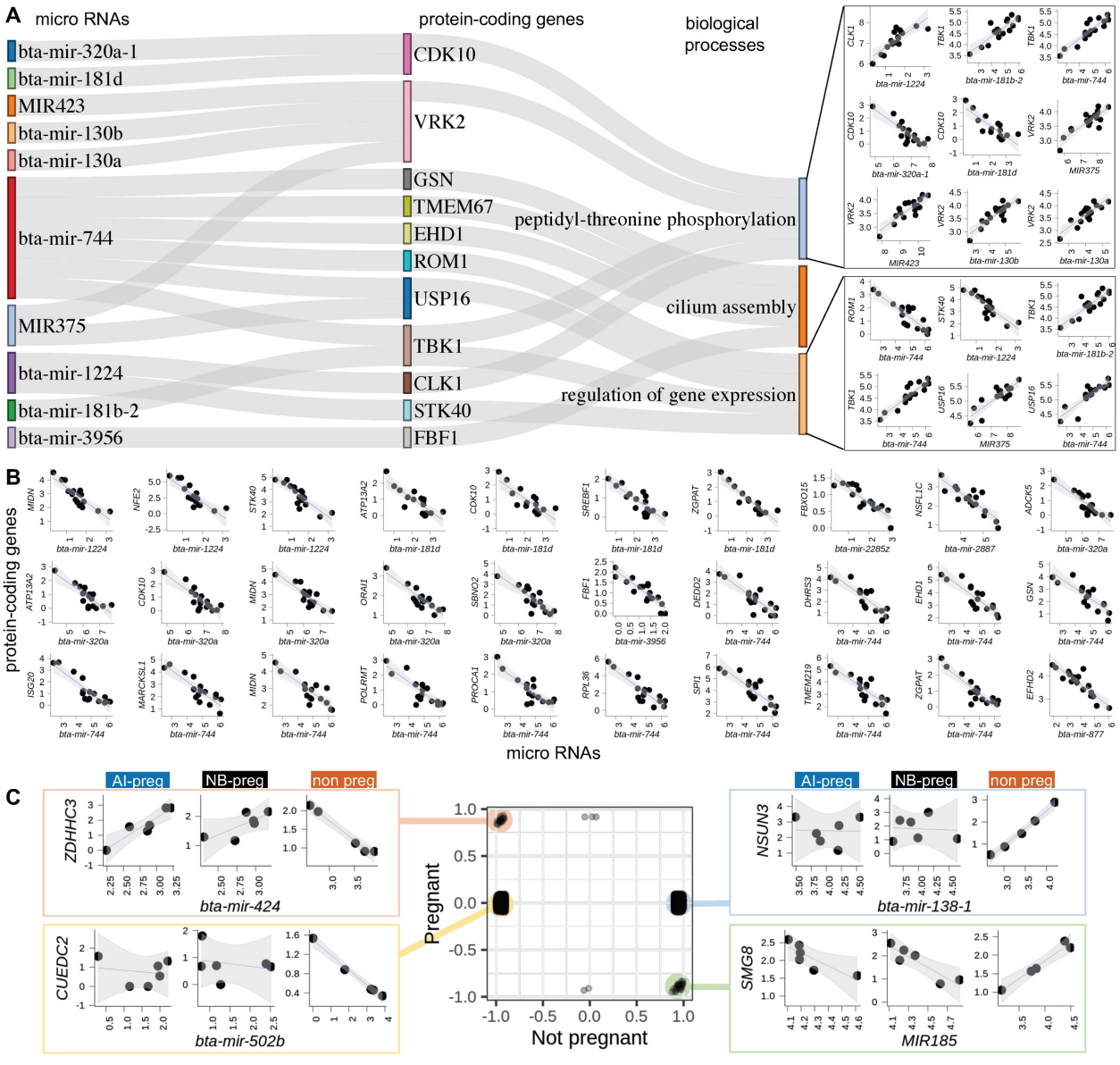



Fig. 3. miRNA from plasma coexpressed with mRNA in PWBCs. (A) miRNA:mRNA pairs significantly coexpressed whose genes are significantly enriched in specific biological processes. (B) miRNA:mRNA pairs significantly coexpressed with evidence of interaction from miRBase. (C) The center chart presents pair-wise 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

162 counterparts that became pregnant. All scatterplots presenting transcript abundance are in
163 Log2(counts per million).

164

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

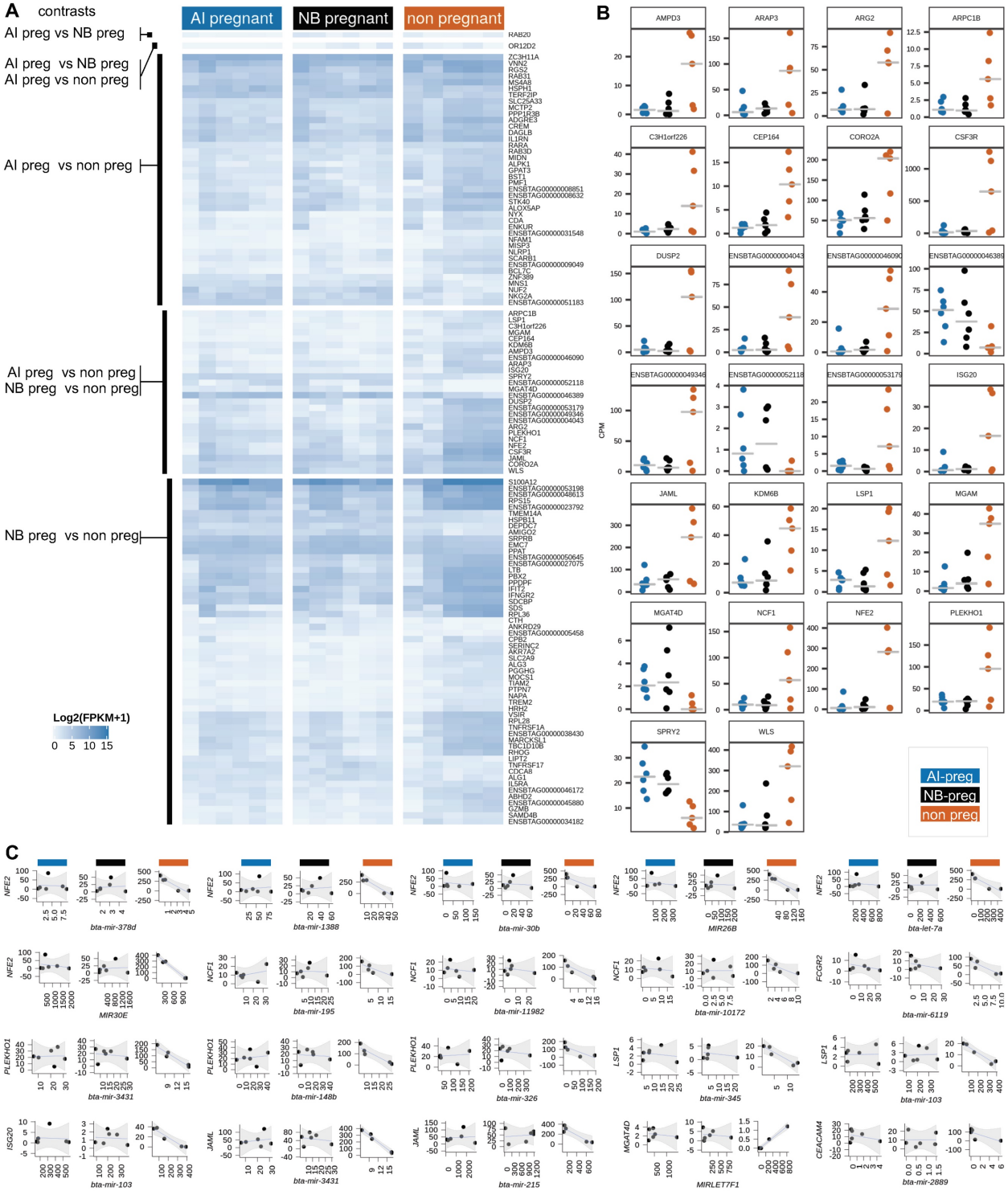
166 Differential gene expression analysis revealed 67 (55 up- and 12 down-regulated) and 81
167 (63 up- and 18 down-regulated) protein-coding genes with altered expression profiles between
168 PWBC of AI-preg versus non-pregnant, and NB-preg versus non-pregnant heifers, respectively
169 (eFDR≤0.05, Fig. 4A, Supplementary Tables S11 and S12, Supplementary Fig. S4). Notably, 26
170 genes were common between both pregnant groups versus the non-pregnant group (Fig. 4B).
171 Gene ontology and pathway analyses of differentially expressed genes between AI-preg and Non-
172 pregnant heifers revealed significant enrichment for the terms “signal transduction” (*ARAP3*,
173 *LSP1*, *NFAM1*, *OR12D2*, *RARA*; FDR<0.05). When differentially expressed genes between NB-
174 preg and non-pregnant heifers were analyzed, significant enrichment was determined for the
175 terms “signal transduction” (*ARAP3*, *HRH2*, *LSP1*, *TIAM2*, *TNFRSF17*, *TNFRSF1A*; FDR=0.062)
176 and the KEGG Pathway “cytokine-cytokine receptor interaction” (*IFNGR2*, *IL5RA*, *LTB*,
177 *TNFRSF17*, *TNFRSF1A*; FDR=0.0045). Plasma miRNA profiles were very similar among
178 samples from all pregnancy outcomes with the exception of miR-11995 that was more abundant
179 in AI-preg versus non-pregnant heifers. These results indicate that quantitative differences exist
180 in transcript abundance of protein-coding genes expressed in PWBC relative to the pregnancy
181 outcome.

182 Next, we asked if differential gene expression could be associated with differential
183 miRNA:mRNA coexpression. An intersection of the results identified 12 differentially expressed
184 genes (10 annotated genes) in non-pregnant heifers relative to the AI-preg and NB-preg groups
185 that also gained coexpression with 27 miRNAs (Fig. 4C). Notably, all genes that had higher
186 expression in non-pregnant heifers (*CEACAM4*, *FCGR2*, *ISG20*, *JAML*, *LSP1*, *NCF1*, *NFE2*,

187 *PLEKHO1*) showed a gain in negative coexpression with specific miRNAs. By contrast *MGAT4D*

188 was more lowly expressed in non-pregnant heifers, and gained a positive coexpression with a

189 miRNA in non-pregnant heifers.




190

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

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

199 Motivated by the observations that several genes have their abundance altered in relation
200 to the fertility potential pregnancy (Fig. 4A), we asked whether transcript abundances could serve
201 as predictive features across datasets. We previously generated gene expression values for a
202 similar dataset (GSE103628 [38]), which also consisted of transcriptome data from PWBCs
203 collected from heifers at the time of AI, and were classified according to their pregnancy outcome
204 as AI-preg (n=6) or non-pregnant (n=6). This dataset was particularly fitting for this interrogation
205 because the sampling occurred at the same site in the year prior. We combined the expression
206 values of the previously produced dataset (year one) with those of the current dataset (year two)
207 and contrasted the transcript abundance between heifers that were categorized as AI-preg versus
208 non-pregnant accounting for the different years as a fixed effect in the model. This analysis
209 identified 198 genes whose abundance were significantly different between the two experimental
210 groups (AI-preg, non-pregnant, FDR<0.03, Fig. 5A, Supplementary Table S13).

211 We used the transcript abundance values for these 198 genes obtained from samples
212 collected in year one as input (training set) for machine learning algorithms to derive models of
213 prediction to discriminate pregnancy outcome (AI-preg or non-pregnant, Fig. 5B). Next, we used
214 the transcript abundance values for these 198 genes obtained from samples collected in year two
215 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).

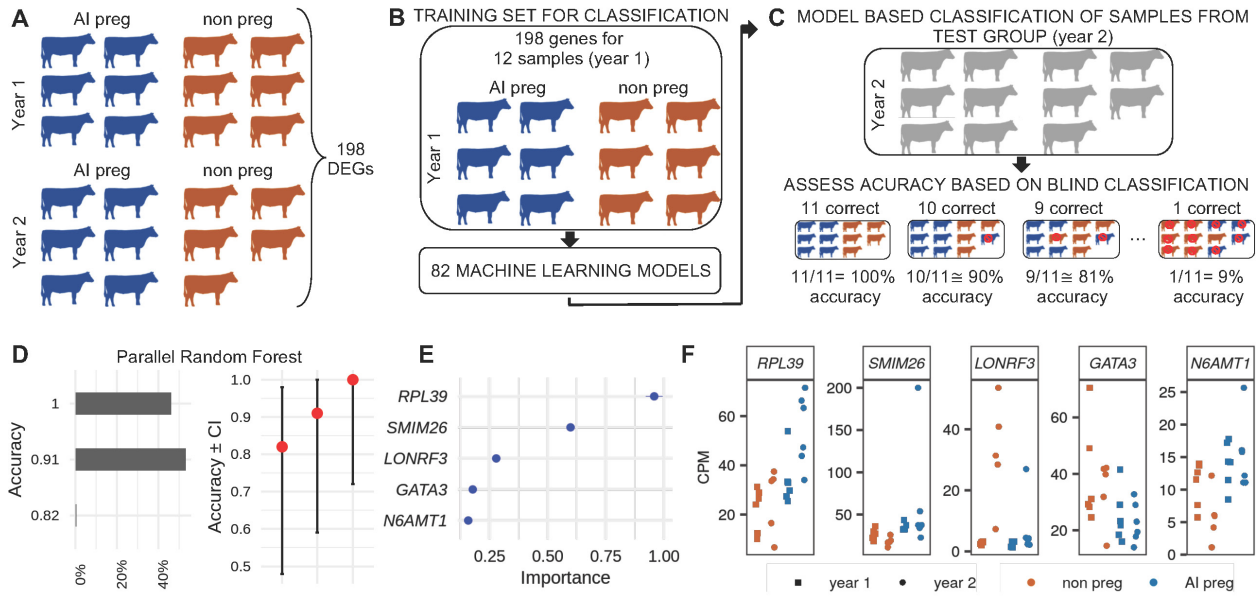



Fig. 5. Prediction of pregnancy outcome based on mRNA levels and machine learning algorithms. (A) Representation of the contrast between gene expression levels of heifers classified as AI-preg and non-pregnant from two years, and the number of differentially expressed genes (DEGs). (B) Depiction of the 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

232 importance for the top five genes. (F) Transcript levels for the top five genes most informative for
233 the machine learning algorithm to classify pregnancy outcome.
234

235 **DISCUSSION**

236 Here, we report a multilayered transcriptome analysis of transcripts from blood sampled
237 on the day of AI. The main objective of our study was to understand gene regulatory networks [39]
238 occurring in PWBCs, and rewiring that can occur based on the heifers' fertility potential. We
239 focused on the day of artificial insemination because it is the last day prior to the fertilization, when
240 the heifer has the first opportunity to become pregnant. We interrogated mRNA and miRNA
241 abundance levels obtained from the same individuals using the coexpression framework, which
242 is a valuable systems biology approach to gain insights into the biology of complex traits [39].
243 Furthermore, using machine learning [40], we detected consistent altered gene transcript
244 abundance associated with infertility across independent datasets. Collectively, our results
245 provide important aspects of regulatory networks in PWBCs that shed light on the physiological
246 status of an animal and its preparedness for becoming pregnant.

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

256 Similar to work reported in humans [41], our gene network analysis focusing on mRNAs
257 detected several protein-coding genes with strong coexpression in PWBCs. Interestingly, nearly
258 all of the correlated expression inferred as statistically significant across all heifers (|r|>0.98,
259 FDR<0.02) were positive. As expected [42], in silico functional gene co-expression analysis
260 indicated groups of genes that are linked to specific biological processes. A contrast of
261 coexpressing pairs of genes between samples collected from non-pregnant heifers and the two
262 groups of heifers that became pregnant displayed a large number of connections present in non-
263 pregnant heifers that were not identified in either AI-preg or NB-preg heifers. Our strict criteria
264 (|r|>0.99, eFDR<0.002) utilized to identify these altered connections indicate functional regulatory
265 relationships between selected genes.

266 Surprisingly, a great majority of the altered connections between gene pairs present were
267 gained in non-pregnant heifers, while there was minor loss of significant connections in non-
268 pregnant heifers relative to the AI-preg and NB-preg groups. The enrichment for specific biological
269 processes of genes that formed new negative connections was clear evidence of the biological
270 relevance of these new connections. Notably, among genes annotated with proteolytic activity,
271 *ADAM19* [43], *ASRGL1* [44], *ATG4B* [45], *DPEP2* [46], *IMMP2L* [47], *PSMB9* [48], *SERPINE2* [49], *UCHL3* [50],
272 *USP42* [51] have been implicated in female infertility. A more drastic change in patterns of
273 coexpression was observed with the inversion of gene connections (significant pairwise
274 coexpression in one group with significant coexpression with opposite direction in another group).
275 These gene circuits changing from activation to repression are explained by current models of
276 incoherent feed-forward loops [52]. Cumulatively, the large occurrence of new connections and the
277 inverted connections created in PWBCs exclusively in non-pregnant heifers indicate that gene
278 regulation is altered as an adaptative response to a stimulus.

279 Our profiling of small RNAs from plasma identified 263 annotated miRNAs, out of which
280 131 and 122 were previously identified in sera and circulating exosomes, respectively, among

281 other cattle organs [53]. Only 38 annotated miRNAs were profiled in peripheral blood
282 mononucleated cells in humans [54], which demonstrates that our plasma samples were depleted
283 of circulating white blood cells.

284 Immuno-related cells composing the PWBC population can uptake circulating small RNAs
285 [55]. Although our data did not allow us to trace the origin of the tissue that produced the miRNAs,
286 we detected 33 circulating miRNAs whose abundance are associated with the abundance of 106
287 protein-coding genes. As reported in humans [56], we quantified positively and negatively correlated
288 abundances between miRNA and protein-coding genes. Negative correlations mostly occur due
289 to the targeting of mRNA by a specific miRNA. Indeed, we found several pairs of miRNA:mRNA
290 negatively correlated that had been identified as interacting pairs that fit this model. By
291 comparison, the positive coexpression between miRNA:mRNA pairs can be explained by a
292 miRNA targeting an intermediate gene, which in turn is negatively associated with a third gene,
293 thereby creating a positive correlation between a miRNA:mRNA pair (see Fig. 3 in reference [57]).
294 In our results, we only identify examples of miRNA:mRNA pairs that follow the intermediate action
295 of miRNAs when relaxing the stringent criteria for inference of significant coexpression.
296 Regardless of the model of interaction, it was notable that many protein-coding genes that were
297 coexpressed with miRNAs were enriched for specific biological processes, including “regulation
298 of gene expression”. These results add further evidence to the growing body of literature
299 suggesting that circulating miRNAs act as mediators of cellular function [58,59].

300 We also tested whether circulating miRNA and PWBC mRNA pairs have differential
301 coexpression relative to the pregnancy outcome. Following the same trend observed for
302 miRNA:mRNA coexpression in PWBCs, there were 1000-fold more miRNA:mRNA connections
303 that were present in heifers that did not become pregnant compared to the loss in coexpression.
304 Results from gene enrichment analysis support the notion that genes forming new miRNA:mRNA
305 coexpression connections in non-pregnant heifers function in the immune system, although our

306 data do not permit inferences on whether particular immune processes were more or less
307 functional in non-pregnant heifers.

308 Critical aspects of female fertility are tied to an appropriate regulation of the immune
309 system. Maternal immune tolerance to pregnancy starts with the exposure to proteins in the
310 semen [60], and is carried throughout placentation and pregnancy establishment [61]. Notably, 24 out
311 of 26 differentially expressed genes between pregnant (AI-preg and NB-preg) and non-pregnant
312 heifers showed greater abundance in non-pregnant heifers relative to the other two groups. (AI-
313 preg and NB-preg). While we did not detect differential abundance of miRNAs relative to
314 pregnancy outcome, the gain of coexpression between miRNAs and mRNAs from differentially
315 expressed genes functionally linked with immunology is in line with regulatory roles that certain
316 miRNAs exert on the immune system [62]. Specifically, bta-miR-30b [63], bta-miR-148b, bta-miR-195
317 [64], bta-miR-326 [65], bta-miR-378d [66], bta-miR-1388 [63], bta-miR-2889 [66], bta-miR-11982 [67] have
318 been reported to be differentially expressed in pathogen infected cells or tissues relative to control
319 counterparts. This result lead us to hypothesize that the heifers in the non-pregnant group have
320 a heightened activity of specific cells or functions in the immune system relative to the heifers that
321 became pregnant. It remains unclear, however, if this possible higher activity was stimulated by
322 extrinsic factors or is within the naturally-occurring variation [68].

323 The results of differential gene expression and differential coexpression fostered a
324 hypothesis that transcript abundance in the PWBCs can be used as predictors of pregnancy
325 outcome. We tested this hypothesis by applying machine learning algorithms on two groups of
326 heifers that were bred in 2015 (year one) and 2016 (year two). Parallel random forest emerged
327 as the algorithm with over 90% efficiency of classification nearly all trials executed, which confirms
328 the potential of accurate classification of samples using RNA-seq data under the case-control
329 framework [69,70]. The results show that while not one single gene emerges as a potential biomarker,

330 the accumulated information of transcript abundance from multiple genes can be powerful for the
331 identification of fertility potential in cattle.

332 Although we detected altered coexpression with high confidence, and in some instances,
333 with significant enrichment of genes in specific gene ontology terms, the consequence of the
334 differential coexpression on those biological functions remains undetermined. In addition, it
335 remains unclear whether the altered patterns of gene expression profiles are connected to
336 immune functions that are related to infertility [71]. The data analyzed here does not allow us to infer
337 causality [72], however our results converge to a strong indication that differences in gene
338 expression profiles are associated [72] with fertility in cattle.

339 In conclusion, analyses of mRNA from PWBCs and circulating extracellular miRNA by
340 multiple approaches (differential gene expression, differential gene coexpression and machine
341 learning classification) provided complementary evidence that gene expression profiles in the
342 PWBCs are associated with the fertility potential in beef heifers. The overall difference detected
343 in transcript abundance for specific genes is reproducible across samples and show promise for
344 the development of molecular features indicating fertility potential in beef heifers. The analysis
345 of transcriptome data is critical for the understanding of economically important complex traits
346 [73]; and our findings show that transcript levels may be relevant for selection of animals more
347 likely to contribute to a sustainable beef production system.

348 **METHODS**

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

353 **Heifer classification according to pregnancy outcome**

354 Crossbred heifers (n=32, Angus-Simmental, *Bos taurus taurus*) were developed as
355 replacement heifers [74,75] at the Black Belt Research and Extension Center (Auburn University).
356 All animals were exposed to the same vaccination protocol which included an injection of
357 ULTRABAC® 7 and BOVI-SHIELD GOLD® FP™ 5 L5 at approximately 6 months of age followed
358 by a second injection of each product at approximately 9 months of age. A final injection of BOVI-
359 SHIELD GOLD® FP® 5 VL5 HB was administered to all heifers at approximately 12 months of
360 age. Thirty-four days prior to AI, heifers were evaluated for reproductive tract score (scale of 1-5
361 [76]) and body condition score (BCS; scale of 1-9 [77]) for the assertion that morphological and
362 physiological conditions were suitable for breeding. One animal was eliminated from the breeding
363 herd at this time due to a reproductive tract score of 2. The remaining animals presented body
364 condition score 6 and reproductive tract score 3 (n=9), 4 (n=14) or 5 (n=8).

365 Thirty one animals, 14 months of age on average, underwent estrous synchronization for
366 fixed-time artificial insemination utilizing the 7-Day Co-Synch + CIDR protocol [78]. An experienced
367 veterinarian inseminated all heifers with semen from one sire of proven fertility (Deer Valley All
368 In, Select Sires). Fourteen days after insemination, heifers were exposed to two fertile bulls for
369 natural breeding for 60 days. An experienced veterinarian performed pregnancy evaluation by
370 transrectal palpation on days 69 and 120 post artificial insemination. Presence or absence of a
371 conceptus, alongside morphological features indicating fetal age were used to classify heifers as
372 pregnant to AI (AI-pregnant), pregnant to natural breeding (NB-pregnant), or non-pregnant. At the
373 end of the breeding season, 16, nine and six heifers were classified as AI-pregnant, NB-pregnant
374 or non-pregnant, respectively.

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

376 Immediately after AI, 10ml of blood was drawn from the jugular vein into vacutainers
377 containing 18mg K2 EDTA (Becton, Dickinson and Company, Franklin, NJ). The tubes were
378 inverted to prevent blood coagulation and immersed in ice for approximately 4 hours. Plasma and

379 buffy coat were separated by centrifugation for 10 minutes at 2000×g at 4°C. Plasma was
380 removed (~1.5ml) and centrifuged 10 minutes at 500×g at 4°C for further removal of cells. The
381 plasma devoid of cells was then transferred to a new microcentrifuge tube and preserved at -
382 80°C. The buffy coat was removed and deposited into 14ml of red blood cell lysis solution (0.15
383 M ammonium chloride, 10mM potassium bicarbonate, 0.1mM EDTA, Cold Spring Harbor
384 Protocols) for 10 minutes at room temperature (24-25°C). After a centrifugation, 800xg for 10
385 minutes, the solution was discarded and the pellet containing PWBCs was re-suspended in 200
386 μl of RNAlater® (Lifetechnologies™, Carlsbad, CA). PWBC samples were stored at -80°C.

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

388 Total RNA was isolated from PWBCs of 18 heifers (AI-pregnant (n=6), NB-pregnant (n=6),
389 and non-pregnant (n=6)) using TRIzol™ reagent (Invitrogen, Carlsbad, CA) [79,80] following the
390 manufacturer's protocol. RNA yield was quantified using the Qubit™ RNA Broad Range Assay
391 Kit (Eurogene, OR) on a Qubit® Fluorometer, and integrity was assessed on an Agilent 2100
392 Bioanalyzer (Agilent, Santa Clara, CA) using an Agilent RNA 6000 Nano kit (Agilent, Santa Clara,
393 CA). RNA was submitted for sequencing if RNA integrity number was >7. Approximately 500 μg
394 of total RNA was used as input for the TruSeq Stranded mRNA Library Prep kit (Illumina, Inc.,
395 San Diego, CA); and library preparation was carried out following manufacturer's instructions.
396 Libraries were quantified with the Qubit™ dsDNA High Sensitivity Assay Kit (Eurogene, OR); and
397 quality was evaluated using the High Sensitivity DNA chip (Agilent, Santa Clara, CA) on an Agilent
398 2100 Bioanalyzer. Libraries were sequenced in a HiSeq2500 system at the Genomic Services
399 Laboratory at HudsonAlpha, Huntsville, AL to generate 125nt long, pair-end reads.

400 Total RNA (circulating extracellular RNA) was isolated from 1ml of plasma by adding 9 ml
401 of TRIzol™ reagent and following manufacturer's protocols through organic phase separation. The
402 aqueous phase containing the RNA was then collected and purified with the RNA Clean &
403 Concentrator-5 kit (ZymoResearch, Irvine, CA), according to manufacturer's instructions. Total

404 RNA yield was quantified in test samples using the Qubit™ RNA Broad Range Assay Kit
405 averaging 21.4ng, whereas samples prepared for sequencing were not quantified and all of the
406 eluted RNA was directly used for library preparation.

407 Total RNA obtained from plasma was desiccated to a volume of 2.5μl for library
408 preparation with the TruSeq Small RNA Library Prep kit (Illumina, Inc., San Diego, CA) following
409 manufacturer's instructions, with the minor alteration that reagent volumes were halved in all
410 reactions until reverse transcription was performed, as described elsewhere [81]. After PCR
411 amplification, libraries were loaded onto a 10% polyacrylamide gel [82] and electrophoresis was
412 carried out for three hours at 100 volts alongside a 25/100 bp mixed DNA ladder (Bioneer,
413 Alameda, CA); and the custom DNA ladder included in the TruSeq Small RNA Library Prep kit.
414 The gel was soaked in GelRed® Nucleic Acid Gel Stain (Biotium, Inc., Fremont, CA) for DNA
415 staining. DNA bands in the range of 145 -160 base pairs were isolated from the gel by the crush-
416 and-soak method [83] with the following modifications. Gel fragments were placed into a 0.5 ml tube
417 containing small holes at the bottom, mounted into a 2ml tube. The tubes were centrifuged at
418 14,000xg for two minutes at room temperature. The crushed gel was soaked in 200μl of water at
419 4°C overnight, followed by removal of the solution containing the library to a new tube.

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

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

429 **RNA sequencing data processing**

430 Sequences were submitted to a custom build bioinformatics pipeline[84-86]. Reads were
431 aligned to the bovine genome (ARS-UCD1.2 obtained from Ensembl [87]) using HISAT2 [88].
432 Sequences aligning to multiple places on the genome, or with 5 or more mismatches, were filtered
433 out. The sequences were then marked for duplicates, and non-duplicated pairs of reads (mRNA)
434 or single reads (miRNA) were used for transcript quantification at the gene level. The fragments
435 were counted against the Ensembl gene annotation [89] (version 1.87) using featureCounts [90].
436 RNAseq data from one animal (non-pregnant) were removed from analysis due to low sequencing
437 yield (613,765 reads).

438 **Filtering of lowly expressed genes**

439 All analytical procedures were carried out in R software [91]. Supplementary code contains
440 all scripts used for analytical procedures described below and the creation of the plots. This code
441 and all original data can be obtained on figshare [92].

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

449 **Pair-wise gene coexpression and differential coexpression for protein-coding**
450 **genes**

451 First, we calculated pair-wise Pearson's correlation coefficient (*r*) [96,97] for all 17 animals
452 sampled. Significance of the correlations was assessed using the "TestCor" package based on

453 the empirical statistical test and the large-scale correlation test with normal approximation[98] for
454 the calculation of the false discovery rate (FDR). Further biological interpretation of results was
455 performed on pairs of genes if abs(r)>0.98, which corresponded to a nominal $P \leq 1.2 \times 10^{-16}$ and
456 FDR<0.02.

457 Next, we calculated within group pair-wise Pearson's correlation coefficient (*r*) for each
458 group: AI-pregnant, NB-pregnant, and non-pregnant. We interpreted a differential correlation as
459 biologically meaningful based on the following scenarios, assuming that groups 1 and 2 can be
460 assigned arbitrarily: inverted correlation, if $|r_{group\ 1} - r_{group\ 2}| > 1.95$; gained correlation in group
461 1 and lost correlation in group 2, if $|r_{group\ 1}| > 0.99$ and $-0.1 < r_{group\ 2} < 0.1$.

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

467 **mRNA-miRNA coexpression and differential coexpression**

468 We calculated Pearson's correlation (*r*) [96,97] values between miRNA and mRNA genes for
469 all 17 samples. Significance was estimated by the eFDR approach [85,99,100] by shuffling (5000
470 randomizations) the animal identification on mRNA dataset, thus breaking the common source of
471 both RNA types (Supplementary Fig. S2).

472 Pearson's correlation (r) values were calculated for mRNA-miRNA transcript pairs within
473 groups: AI-pregnant, NB-pregnant, and non-pregnant heifers. We interpreted a differential
474 correlation as biologically meaningful based on the following scenarios $|r_{group\ 1}| > 0.99$ and
475 $-0.1 < r_{group\ 2} < 0.1$. Significance was assessed by the eFDR approach (Supplementary Fig.

476 S3). We further interrogated the data for experimentally validated miRNA-mRNA interactions
477 obtained from miRWalk database (version 3)[101,102].

478 **Differential expression of mRNA and miRNA**

479 Differences of transcript levels between groups based on the pregnancy outcome were
480 determined from fragment counts using "edgeR" [93] and "DESeq2" [103]. For each comparison, a
481 gene was initially inferred as differentially expressed if the nominal P value was ≤0.03 in both
482 "edgeR" and "DESeq2" analyses. This nominal P value corresponded to eFDR≤0.05
483 (Supplementary Fig. S4), as calculated according to the procedure and formulas outlined
484 elsewhere [85,99,100] using 10000 randomizations of sample reshuffling. Genes inferred as
485 differentially expressed based on P values were further selected through "leave-one-out" cross
486 validation [104,105]. Only genes whose significance were maintained after all rounds of sampling
487 were further considered for biological interpretation.

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

489 Differentially expressed genes and genes that displayed differential coexpression among
490 heifers of AI-pregnant, NB-pregnant, and non-pregnant classifications were analyzed for
491 enrichment of Gene Ontology [106] categories and KEGG Pathways [107]. In all tests, the background
492 gene list consisted of all genes detected as expressed in our dataset [108]. The GO annotations and
493 transcript lengths were obtained from BioMart [109] and KEGG annotation was obtained from the
494 "org.Bt.eg.db" package. We used "GoSeq" package [110] to estimate significance of category
495 enrichment using the sampling method [110]. We utilized FDR for multiple testing under dependency
496 [111], and inferred statistical significance of GO terms and KEGG pathways if FDR < 0.10.

497 **Assessment of predictive potential of gene expression levels by machine learning**
498 **algorithms.**

499 We focused the mRNA sequencing data from AI-pregnant and non-pregnant heifers from
500 this report and data we collected from the same site on the previous year and published elsewhere
501 [38,86]. The raw sequences from the previous dataset were realigned, and data were treated as
502 described above.

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

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

517

518 **REFERENCES**

519


520 1 Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M. & Tiwari, B. Future Protein Supply
521 and Demand: Strategies and Factors Influencing a Sustainable Equilibrium. *Foods* **6**,
522 doi:10.3390/foods6070053 (2017).
523 2 Bormann, J. M., Totir, L. R., Kachman, S. D., Fernando, R. L. & Wilson, D. E. Pregnancy
524 rate and first-service conception rate in Angus heifers. *J. Anim. Sci.* **84**, 2022-2025,
525 doi:10.2527/jas.2005-615 (2006).

3 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, doi:10.2527/jas.2008-1476 (2009).

4 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, doi:10.2527/jas2012-5580 (2013).

5 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, doi:10.2527/jas.2006-758 (2007).

6 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).

7 Funston, R. N. & Larson, D. M. Heifer development systems: dry-lot feeding compared with grazing dormant winter forage. *J. Anim. Sci.* **89**, 1595-1602, doi:10.2527/jas.2010-3095 (2011).

8 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, doi:10.1016/j.theriogenology.2014.01.008 (2014).

9 Martin, J. L. *et al.* Effect of prebreeding body weight or progestin exposure before breeding on beef heifer performance through the second breeding season. *J. Anim. Sci.* **86**, 451-459, doi:10.2527/jas.2007-0233 (2008).

10 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).

11 Mallory, D. A., Nash, J. M., Ellersieck, M. R., Smith, M. F. & Patterson, D. J. Comparison of long-term progestin-based protocols to synchronize estrus before fixed-time artificial insemination in beef heifers. *J. Anim. Sci.* **89**, 1358-1365, doi:10.2527/jas.2010-3694 (2011).

12 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).

13 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. and Biotechnology* **10**, 18, doi:10.1186/s40104-019-0329-6 (2019).

14 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, doi:10.1186/s40104-019-0329-6 (2019).

15 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.* **2013**, 4486-4491, doi:10.2527/jas2013-6465 (2013).

16 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).

17 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, doi:10.2527/jas1973.3611 (1973).

18 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, doi:10.1093/tas/txy020 (2018).

575 19 Morris, C. A. & Cullen, N. G. A note on genetic correlations between pubertal traits of
576 males or females and lifetime pregnancy rate in beef cattle. *Livest. Prod. Sci.* **39**, 291-
577 297, doi:10.1016/0301-6226(94)90291-7 (1994).
578 20 Mwansa, P. B. *et al.* Selection for cow lifetime pregnancy rate using bull and heifer
579 growth and reproductive traits in composite cattle. *Can J. Anim. Sci.* **80**, 507-510 (2000).
580 21 Peters, S. O. *et al.* Heritability and Bayesian genome-wide association study of first
581 service conception and pregnancy in Brangus heifers. *J. Anim. Sci.* **2013**, 605-612,
582 doi:10.2527/jas2012-5580 (2013).
583 22 Neupane, M. *et al.* Loci and pathways associated with uterine capacity for pregnancy
584 and fertility in beef cattle. *PLoS One* **12**, e0188997, doi:10.1371/journal.pone.0188997
585 (2017).
586 23 McDaneld, T. G. *et al.* Genomewide association study of reproductive efficiency in
587 female cattle. *J. Anim. Sci.* **2014**, 1945-1957, doi:10.2527/jas2012-6807 (2014).
588 24 McDaneld, T. G. *et al.* Y are you not pregnant: Identification of Y chromosome segments
589 in female cattle with decreased reproductive efficiency. *J. Anim. Sci.* **90**, 2142-2151
590 (2012).
591 25 de Camargo, G. M. *et al.* Association between JY-1 gene polymorphisms and
592 reproductive traits in beef cattle. *Gene* **533**, 477-480, doi:10.1016/j.gene.2013.09.126
593 (2014).
594 26 Dias, M. M. *et al.* Study of lipid metabolism-related genes as candidate genes of sexual
595 precocity in Nellore cattle. *Genet. Mol. Res.* **14**, 234-243, doi:10.4238/2015.January.16.7
596 (2015).
597 27 Irano, N. *et al.* Genome-wide association study for indicator traits of sexual precocity in
598 Nellore cattle. *PLoS One* **11**, e0159502, doi:10.1371/journal.pone.0159502 (2016).
599 28 Junior, G. A. O. *et al.* Genomic study and medical subject headings enrichment analysis
600 of early pregnancy rate and antral follicle numbers in Nelore heifers. *J. Anim. Sci.* **95**,
601 4796-4812, doi:10.2527/jas2017.1752 (2017).
602 29 Fortes, M. R. S. *et al.* Gene network analyses of first service conception in Brangus
603 heifers: Use of genome and trait associations, hypothalamic-transcriptome information,
604 and transcription factors. *J. Anim. Sci.* **2012**, 2894-2906, doi:10.2527/jas2011-4601
605 (2012).
606 30 Doyle, S. P., Golden, B. L., Green, R. D. & Brinks, J. S. Additive genetic parameter
607 estimates for heifer pregnancy and subsequent reproduction in Angus females. *J. Anim.*
608 *Sci.* **78**, 2091-2098, doi:10.2527/2000.7882091x (2000).
609 31 Toghiani, S. *et al.* Genomic prediction of continuous and binary fertility traits of females
610 in a composite beef cattle breed. *J. Anim. Sci.* **95**, 4787-4795, doi:10.2527/jas2017.1944
611 (2017).
612 32 McAllister, C. M., Speidel, S. E., Crews, D. H., Jr. & Enns, R. M. Genetic parameters for
613 intramuscular fat percentage, marbling score, scrotal circumference, and heifer
614 pregnancy in Red Angus cattle. *J. Anim. Sci.* **89**, 2068-2072, doi:10.2527/jas.2010-3538
615 (2011).
616 33 Boddhireddy, P. *et al.* Genomic predictions in Angus cattle: Comparisons of sample size,
617 response variables, and clustering methods for cross-validation. *J. Anim. Sci.* **92**, 485-
618 497, doi:10.2527/jas.2013-6757 (2014).
619 34 Li-Pook-Than, J. & Snyder, M. iPOP goes the world: integrated personalized Omics
620 profiling and the road toward improved health care. *Chem. Biol.* **20**, 660-666,
621 doi:10.1016/j.chembiol.2013.05.001 (2013).
622 35 Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to
623 Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100**, 473-487,
624 doi:10.1016/j.ajhg.2017.01.031 (2017).

625 36 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association
626 studies. *Nat. Genet.* **48**, 245-252, doi:10.1038/ng.3506 (2016).
627 37 Ioannidis, J. & Donadeu, F. X. Circulating microRNA profiles during the bovine oestrous
628 cycle. *PLoS One* **11**, e0158160, doi:10.1371/journal.pone.0158160 (2016).
629 38 Dickinson, S. E. *et al.* Transcriptome profiles in peripheral white blood cells at the time of
630 artificial insemination discriminate beef heifers with different fertility potential. *BMC*
631 *Genomics* **19**, doi:10.1186/s12864-018-4505-4 (2018).
632 39 Emmert-Streib, F., Dehmer, M. & Haibe-Kains, B. Gene regulatory networks and their
633 applications: understanding biological and medical problems in terms of networks. *Front.*
634 *Cell. Dev. Biol.* **2**, 38, doi:10.3389/fcell.2014.00038 (2014).
635 40 Xu, C. & Jackson, S. A. Machine learning and complex biological data. *Genome Biol.* **20**,
636 76, doi:10.1186/s13059-019-1689-0 (2019).
637 41 Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute
638 Deconvolution of Human Immune Cell Types. *Cell Rep.* **26**, 1627-1640 e1627,
639 doi:10.1016/j.celrep.2019.01.041 (2019).
640 42 van Dam, S., Vosa, U., van der Graaf, A., Franke, L. & de Magalhaes, J. P. Gene co-
641 expression analysis for functional classification and gene-disease predictions. *Brief*
642 *Bioinform.* **19**, 575-592, doi:10.1093/bib/bbw139 (2018).
643 43 Sheridan, M. A. *et al.* Early onset preeclampsia in a model for human placental
644 trophoblast. *P. Natl. Acad. Sci. USA* **116**, 4336-4345, doi:10.1073/pnas.1816150116
645 (2019).
646 44 Fitzgerald, H. C. *et al.* Idiopathic infertility in women is associated with distinct changes
647 in proliferative phase uterine fluid proteins. *Biol. Reprod.* **98**, 752-764,
648 doi:10.1093/biolre/ioy063 (2018).
649 45 Yang, H. L. *et al.* Autophagy in endometriosis. *Am. J. Transl. Res.* **9**, 4707-4725 (2017).
650 46 Pelch, K. E. *et al.* Aberrant gene expression profile in a mouse model of endometriosis
651 mirrors that observed in women. *Fertil. Steril.* **93**, 1615-1627 e1618,
652 doi:10.1016/j.fertnstert.2009.03.086 (2010).
653 47 Demain, L. A., Conway, G. S. & Newman, W. G. Genetics of mitochondrial dysfunction
654 and infertility. *Clin. Genet.* **91**, 199-207, doi:10.1111/cge.12896 (2017).
655 48 Koks, S. *et al.* The differential transcriptome and ontology profiles of floating and
656 cumulus granulosa cells in stimulated human antral follicles. *Mol. Hum. Reprod.* **16**, 229-
657 240, doi:10.1093/molehr/gap103 (2010).
658 49 Li, S. H. *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, doi:10.1186/s12958-015-0091-3 (2015).
661 50 Mtango, N. R. *et al.* Essential role of maternal UCHL1 and UCHL3 in fertilization and
662 preimplantation embryo development. *J. Cell Physiol.* **227**, 1592-1603,
663 doi:10.1002/jcp.22876 (2012).
664 51 Ingham, N. J. *et al.* Mouse screen reveals multiple new genes underlying mouse and
665 human hearing loss. *PLoS Biol* **17**, e3000194, doi:10.1371/journal.pbio.3000194 (2019).
666 52 Reeves, G. T. The engineering principles of combining a transcriptional incoherent
667 feedforward loop with negative feedback. *J. Biol. Eng.* **13**, 62, doi:10.1186/s13036-019-
668 0190-3 (2019).
669 53 Sun, H. Z., Chen, Y. & Guan, L. L. MicroRNA expression profiles across blood and
670 different tissues in cattle. *Sci. Data* **6**, 190013, doi:10.1038/sdata.2019.13 (2019).
671 54 Vaz, C. *et al.* Analysis of microRNA transcriptome by deep sequencing of small RNA
672 libraries of peripheral blood. *BMC Genomics* **11**, 288, doi:10.1186/1471-2164-11-288
673 (2010).

674 55 Fritz, J. V. *et al.* Sources and Functions of Extracellular Small RNAs in Human
675 Circulation. *Annu. Rev. Nutr.* **36**, 301-336, doi:10.1146/annurev-nutr-071715-050711
676 (2016).
677 56 Diaz, G., Zamboni, F., Tice, A. & Farci, P. Integrated ordination of miRNA and mRNA
678 expression profiles. *BMC Genomics* **16**, 767, doi:10.1186/s12864-015-1971-9 (2015).
679 57 Ritchie, W., Rajasekhar, M., Flamant, S. & Rasko, J. E. Conserved expression patterns
680 predict microRNA targets. *PLoS Comput. Biol.* **5**, e1000513,
681 doi:10.1371/journal.pcbi.1000513 (2009).
682 58 Bayraktar, R., Van Roosbroeck, K. & Calin, G. A. Cell-to-cell communication: microRNAs
683 as hormones. *Mol. Oncol.* **11**, 1673-1686, doi:10.1002/1878-0261.12144 (2017).
684 59 Rayner, K. J. & Hennessy, E. J. Extracellular communication via microRNA: lipid
685 particles have a new message. *J. Lipid. Res.* **54**, 1174-1181, doi:10.1194/jlr.R034991
686 (2013).
687 60 Robertson, S. A. & Sharkey, D. J. The role of semen in induction of maternal immune
688 tolerance to pregnancy. *Semin. Immunol.* **13**, 243-254, doi:10.1006/smim.2000.0320
689 (2001).
690 61 Moffett, A. & Loke, C. Immunology of placentation in eutherian mammals. *Nat. Rev.*
691 *Immunol.* **6**, 584-594, doi:10.1038/nri1897 (2006).
692 62 Momen-Heravi, F. & Bala, S. miRNA regulation of innate immunity. *J. Leukoc. Biol.*,
693 doi:10.1002/JLB.3MIR1117-459R (2018).
694 63 Jin, W. *et al.* Transcriptome microRNA profiling of bovine mammary epithelial cells
695 challenged with Escherichia coli or Staphylococcus aureus bacteria reveals pathogen
696 directed microRNA expression profiles. *BMC Genomics* **15**, 181, doi:10.1186/1471-
697 2164-15-181 (2014).
698 64 Li, R. *et al.* Transcriptome microRNA profiling of bovine mammary glands infected with
699 Staphylococcus aureus. *Int. J. Mol. Sci.* **16**, 4997-5013, doi:10.3390/ijms16034997
700 (2015).
701 65 Vegh, P. *et al.* MicroRNA profiling of the bovine alveolar macrophage response to
702 Mycobacterium bovis infection suggests pathogen survival is enhanced by microRNA
703 regulation of endocytosis and lysosome trafficking. *Tuberculosis (Edinb)* **95**, 60-67,
704 doi:10.1016/j.tube.2014.10.011 (2015).
705 66 Ma, S., Tong, C., Ibeagha-Awemu, E. M. & Zhao, X. Identification and characterization
706 of differentially expressed exosomal microRNAs in bovine milk infected with
707 Staphylococcus aureus. *BMC Genomics* **20**, 934, doi:10.1186/s12864-019-6338-1
708 (2019).
709 67 Pang, F. *et al.* Comprehensive analysis of differentially expressed microRNAs and
710 mRNAs in MDBK cells expressing bovine papillomavirus E5 oncogene. *PeerJ* **7**, e8098,
711 doi:10.7717/peerj.8098 (2019).
712 68 Hughes, H. D., Carroll, J. A., Burdick Sanchez, N. C. & Richeson, J. T. Natural variations
713 in the stress and acute phase responses of cattle. *Innate. Immun.* **20**, 888-896,
714 doi:10.1177/1753425913508993 (2014).
715 69 Wenric, S. & Shemirani, R. Using Supervised Learning Methods for Gene Selection in
716 RNA-Seq Case-Control Studies. *Front. Genet.* **9**, 297, doi:10.3389/fgene.2018.00297
717 (2018).
718 70 Zararsiz, G. *et al.* A comprehensive simulation study on classification of RNA-Seq data.
719 *PLoS One* **12**, e0182507, doi:10.1371/journal.pone.0182507 (2017).
720 71 Brazdova, A., Senechal, H., Peltre, G. & Poncet, P. Immune Aspects of Female
721 Infertility. *Int. J. Fertil. Steril.* **10**, 1-10, doi:10.22074/ijfs.2016.4762 (2016).
722 72 Altman, N. & Krzywinski, M. Association, correlation and causation. *Nat. Methods* **12**,
723 899-900, doi: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, doi: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**, xiii-xiv, doi:10.1016/j.cvfa.2013.07.014 (2013).
75 Larson, R. L., White, B. J. & Laflin, S. Beef Heifer Development. *Vet Clin North Am Food. Anim. Pract.* **32**, 285-302, doi: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, doi: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* **1993**, 1143-1152, doi: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 F2α, 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.* **2010**, pdb prot5439, doi: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, doi: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, doi:10.1038/srep44061 (2017).
82 Chory, J. & Pollard, J. D., Jr. Separation of small DNA fragments by conventional gel electrophoresis. *Curr. Protoc. Mol. Biol.* **Chapter 2**, Unit2 7, doi: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.* **2019**, pdb prot100479, doi: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. *P. Natl. Acad. Sci. USA* **113**, 14492-14501, doi: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, doi: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, doi: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* **2011**, bar030, doi: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, doi:10.1038/nmeth.3317 (2015).
89 Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749-755, doi: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, doi:10.1093/bioinformatics/btt656 (2014).
91 Ihaka, R. & Gentleman. R: A Language and Environment for Statistical Computing. *J. Comput. Graph. Stat.* **5**, 299-214 (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. Online resource. 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, doi: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, doi: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, doi: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.* **3**, doi:ARTN 78 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, doi: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, doi:10.1080/01621459.2014.999157 (2016).
99 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *P. Natl. Acad. Sci. USA* **100**, 9440-9445, doi: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, doi: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, doi: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, doi: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, doi:10.1186/PREACCEPT-8897612761307401 (2014).
104 Picard, R. R. & Cook, R. D. Cross-Validation of Regression-Models. *J. Am. Stat. Assoc.* **79**, 575-583, doi:Doi 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, doi:Doi 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, doi: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, doi: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, doi: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, doi: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, doi: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, doi:10.1214/aos/1013699998 (2001).

826

827

828 **Author contributions**

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

835

836 **Competing interests**

837 The author(s) declare no competing interests.

838

839 **Data availability**

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

844

845 **Ethics of research with animals**

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

850