



OPEN

## A multi-omics analysis identifies molecular features associated with fertility in heifers (*Bos taurus*)

Mackenzie A. Marrella &amp; Fernando H. Biase

Infertility or subfertility is a critical barrier to sustainable cattle production, including in heifers. The development of heifers that do not produce a calf within an optimum window of time is a critical factor for the profitability and sustainability of the cattle industry. In parallel, heifers are an excellent biomedical model for understanding the underlying etiology of infertility because well-nourished heifers can still be infertile, mostly because of inherent physiological and genetic causes. Using a high-density single nucleotide polymorphism (SNP) chip, we collected genotypic data, which were analyzed using an association analysis in PLINK with Fisher's exact test. We also produced quantitative transcriptome data and proteome data. Transcriptome data were analyzed using the quasi-likelihood test followed by the Wald's test, and the likelihood test and proteome data were analyzed using a generalized mixed model and Student's t-test. We identified two SNPs significantly associated with heifer fertility (rs110918927, chr12: 85648422,  $P = 6.7 \times 10^{-7}$ ; and rs109366560, chr11:37666527,  $P = 2.6 \times 10^{-5}$ ). We identified two genes with differential transcript abundance (eFDR  $\leq 0.002$ ) between the two groups (Fertile and Sub-Fertile): Adipocyte Plasma Membrane Associated Protein (*APMAP*, 1.16 greater abundance in the Fertile group) and Dynein Axonemal Intermediate Chain 7 (*DNAI7*, 1.23 greater abundance in the Sub-Fertile group). Our analysis revealed that the protein Alpha-ketoglutarate-dependent dioxygenase FTO was more abundant in the plasma collected from Fertile heifers relative to their Sub-Fertile counterparts (FDR < 0.05). Lastly, an integrative analysis of the three datasets identified a series of molecular features (SNPs, gene transcripts, and proteins) that discriminated 21 out of 22 heifers correctly based on their fertility category. Our multi-omics analyses confirm the complex nature of female fertility. Very importantly, our results also highlight differences in the molecular profile of heifers associated with fertility that transcend the constraints of breed-specific genetic background.

### Abbreviations

$\mu\text{g}$	Microgram
AI	Artificial insemination
ALB	Albumin
APMAP	Adipocyte plasma membrane associated protein
APOC2	Apolipoprotein C-II
ASAP3	ArfGAP with SH3 domain, ankyrin repeat and PH domain 3
ATP5MC1	ATP synthase membrane subunit c locus 1
C8B	Complement component C8 beta chain
CCDC34	Coiled-coil domain containing 34
CEP170	Centrosomal protein 170
CIDR	Controlled internal drug release
CPM	Counts per million
DNAI7	Dynein axonemal intermediate chain 7
DTT	Dithiothreitol
DGE	Differential gene expression
DPA	Differential protein abundance
eFDR	Empirical false discovery rate
EML6	Echinoderm microtubule associated protein like 6

School of Animal Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. email: fbiase@vt.edu

FDR	False discovery rate
FPKM	Fragments per kilobase per million
<i>FTO</i>	Alpha-ketoglutarate-dependent dioxygenase FTO
GnRH	Gonadotrophin-releasing hormone
<i>GPLD1</i>	Phosphatidylinositol-glycan-specific phospholipase D
GWAS	Genome-wide association study
IAA	Indole-3-acetic acid
IACUC	Institutional animal care and use committee
IM	Intramuscular
K2 EDTA	Dipotassium ethylenediaminetetraacetic acid
LCP1	Lymphocyte cytosolic protein 1
LOC525947	Serotransferrin-like
M	Molar
m <sup>6</sup> A	N6-methyladenosine
m <sup>6</sup> A <sub>m</sub>	N6mena2'-O-dimethyladenosine
mg	Milligram
Min	Minutes
ml	Milliliter
mRNA	Messenger RNA
ms	Millisecond
<i>MYDGF</i>	Myeloid derived growth factor
ng	Nanogram
nm	Nanometer
nl	Nanoliter
P	Probability
PGF2 $\alpha$	Prostaglandin F2 alpha
<i>PROZ</i>	Vitamin K-dependent protein Z
<i>RAD51AP1</i>	RAD51 associated protein 1
RIN	RNA integrity number
<i>SERPINF1</i>	Pigment epithelium-derived factor
SNP	Single nucleotide polymorphism
TAF5	TAF5 chemokine-like family member 5
<i>UQCRQ</i>	Ubiquinol-cytochrome c reductase complex III subunit VII
VT-MSI	Virginia tech mass spectrometry incubator
Xg	Relative centrifugal force
$\mu$ l	Microliter

The latest data from the Food and Agriculture Organization show that in 2020 more than 46% of the daily protein supply in the world was from animal-based foods (FAO-STATS). Bovine meat and milk accounted for 12.8% of the total protein supply in the world in 2020 (FAO-STATS). These numbers underscore the importance of cattle production to sustain a growing demand for protein globally<sup>1</sup>. Infertility or subfertility is a critical barrier to sustainable cattle production<sup>2</sup>, including in heifers. For example, approximately 15%<sup>3</sup> and 5%<sup>4</sup> of beef and dairy heifers, respectively, do not calve at 24 months of age. Heifers that calve at an optimum age have greater productivity and longevity in the herd<sup>5–10</sup>. Therefore, identifying heifers with optimum fertility is a promising approach to improving sustainability in cattle production.

The heritability of breeding values for heifer fertility is often low for beef<sup>11–17</sup> and dairy<sup>18–23</sup> heifers, which indicate that there are multiple genetic factors impacting this complex trait beyond additive genetic effects. Another potential avenue for the understanding of infertility is the use of molecular phenotyping<sup>24</sup>. The pioneering efforts focused on genome-wide association studies (GWAS) to identify genetic markers associated with heifer fertility<sup>4,12,25–38</sup>, but only a few seem to be reproducible across populations<sup>37</sup>. More recent efforts have also focused on transcriptome<sup>39–41</sup> and metabolome<sup>42</sup> datasets characterizing these molecules in blood samples. Again, limited genes have been identified with differential transcript abundance across datasets<sup>39</sup>. Much research is needed for the identification of molecular features that can help explain fertility fitness.

Altogether, approximately 5% of heifers are infertile<sup>4,43</sup>, and this cohort is a great biological model for studying the genetic bases of infertility for several reasons. First, neither dairy nor beef heifers are under the challenging metabolic demand required for milk production<sup>44–46</sup>. Second, post-partum cows need to undergo a critical period of physiological and anatomical recovery before the next breeding<sup>47–49</sup>. Third, there are several postpartum diseases with negative consequences on reproduction success<sup>50–52</sup>. Reproductive problems in well-managed heifers are inherent to their physiology<sup>20</sup>, most of which are also under genetic control<sup>53</sup>, or directly related to mutations<sup>54</sup> that impair female reproductive functions.

Angus and Holstein heifers have similar frequencies of infertility or subfertility<sup>4,43</sup> despite the selection pressures directed at beef or dairy production, and thus have distant genetic background. Most studies involving the identification of biological features associated with fertility in heifers have used either one group of pure-bred or crossbred animals<sup>4,12,25–42</sup>. Here, we carried out a case-control<sup>55</sup> experiment to test the hypothesis that differences in genetic variants, gene transcript, and protein abundance due to fertility fitness would be shared between heifers of different genetic background. Our objective was to contrast genetic variants, gene transcripts, and protein abundance between Fertile and Sub-Fertile heifers from Angus and Holstein genetic backgrounds. We show that both the independent analysis and multi-omics approach identified molecular signatures that

capable of discriminating heifers of differing fertility potential, and thus with an underlying biology associated with fertility that is shared between both breeds.

## Methods

All analytical procedures are presented in Additional file 1 and accessible at <https://biase-lab.github.io/MultiOmics/>.

**Ethics statement.** Animal handling for this experiment was approved by the Institutional Animal Care and Use Committee (IACUC) at Virginia Polytechnic Institute and State University.

**Experimental design.** We collected blood samples from purebred Angus heifers ( $n = 12$ ), averaging 14 months in age, at the time of their first artificial insemination (AI) service. Heifers were subjected to a 7-Day Co-Synch + CIDR estrus synchronization protocol prior to breeding. Briefly, heifers were administered an intramuscular (IM) injection of gonadotrophin-releasing hormone (GnRH, 100  $\mu$ g; Factrel<sup>®</sup>; Zoetis Inc.) on Day 0, followed by the insertion of a controlled internal drug release (CIDR, 1.38 g Progesterone; Eazi-Breed<sup>™</sup> CIDR<sup>®</sup>; Zoetis Inc.). On Day 7, the CIDR was removed and an injection of prostaglandin F2 alpha (PGF2 $\alpha$ , 25  $\mu$ g; Lutalyse<sup>®</sup>; Zoetis Inc.) was delivered. Fixed-time AI was performed  $54 \pm 2$  h following CIDR removal alongside a second injection of GnRH.

Additionally, we collected blood samples from purebred Holstein ( $n = 10$ ) heifers, averaging 12 months in age, at the time of the first AI service. Heifers were enrolled in a 5-Day CIDR-Synch protocol before insemination. Briefly, an IM injection of GnRH was delivered on Day 0 with the insertion of a CIDR device. The CIDR device was removed on Day 5, followed by an IM injection of PGF2 $\alpha$ . A second injection of PGF2 $\alpha$  was administered 24 hours later. Then, timed AI was performed with a second GnRH injection on Day 8.

Heifers were identified as Fertile (Holstein,  $n = 5$ ; Angus,  $n = 5$ ) or Sub-Fertile (Holstein,  $n = 5$ ; Angus,  $n = 7$ ) based on their pregnancy outcome, following similar criteria used previously<sup>39,40</sup>. Fertile animals were identified as those who became pregnant and subsequently delivered a calf following the first insemination service. Angus heifers were categorized as Sub-Fertile after failing to achieve pregnancy following two insemination services and exposure to a bull for natural breeding. Holstein heifers were identified as Sub-Fertile after needing four or more artificial inseminations.

Heifers were synchronized with protocols that have been identified by prior research to have high success for a heifer to become pregnant to AI<sup>56,57</sup>. Hence the different protocols for beef and dairy heifers. The criteria for classification were different for each group due to differences in management that are inherent to beef and dairy replacement heifers. Most importantly, each heifer had multiple opportunities to become pregnant before being classified as sub-fertile.

The heifers utilized in this study were not part of a nutritional experiment, and thus nutrition was not accounted as a variable nor was it a factor in the selection of heifers. All dairy heifers were raised with equivalent exposure to feed. Similarly, all beef heifers were raised with equivalent exposure to feed.

**Blood sample collection and white blood cell isolation.** Fifty ml of blood were drawn from each animal by venipuncture of the jugular vein using 18 mg K2 EDTA vacutainers (Becton, Dickinson, and Company). The tubes were inverted for proper mixing with the anticoagulant and then immediately placed on ice until further processing.

We processed the blood samples following procedures described elsewhere<sup>39,40,58</sup> within three hours of sampling<sup>58</sup>. Tubes containing whole blood samples were centrifuged for 25 minutes (min) at 4 °C and 2000 $\times$ g to separate the buffy coat. The buffy coat was then aspirated and mixed with 14 ml of red blood cell lysis buffer (1.55 M ammonium chloride, 0.12 M sodium bicarbonate, 1 mM EDTA (Cold Spring Harbor Protocols)). Then, the solution was centrifuged for 10 min at 4 °C at 800 $\times$ g and the supernatant was discarded. The remaining pellet was mixed with 200  $\mu$ l TRIzol<sup>™</sup> Reagent (Invitrogen<sup>™</sup>, Thermo Fisher Scientific, Waltham, MA) in a 2 ml cryotube (Corning Inc., Corning, NY) prior to snap-freezing with liquid nitrogen. Samples were then stored at  $-80$  °C until further processing.

**Total RNA and DNA extraction.** The buffy coat samples were thawed at room temperature in a total volume of 525  $\mu$ l TRIzol<sup>™</sup> Reagent. Then, total RNA was extracted from peripheral white blood cells using the Zymo Research Direct-zol<sup>™</sup> DNA/RNA Miniprep kit (Zymo Research Corporation, Irvine, CA), according to the manufacturer's protocol. Next, we assessed the quality of the RNA by quantifying the RNA integrity number (RIN) for each sample using the Agilent RNA 6000 Pico kit (Agilent, Santa Clara, CA) on the Agilent 2100 Bio-analyzer (Agilent, Santa Clara, CA).

**Genotyping and data processing.** We submitted 400 ng of DNA for each heifer to Neogen (Neogen Corporation, Lincoln, NE) for genotyping. The samples were genotyped using the Illumina BovineHD Beadchip (Illumina Inc., San Diego, CA) genotyping array (777K). We processed the data for quality control<sup>59</sup> using PLINK<sup>60</sup>. First, we removed SNPs that were preferentially called in one of the groups in the case and control. This was followed by the removal of samples with more than 10% of the genotypes missing, and removal of SNPs with a minor allelic frequency less than 1%, a missing rate greater than 10%, or deviation from the Hardy-Weinberg equilibrium ( $P < 0.00001$ ). Next, we carried out variant pruning. We considered a window size of 50 kilobases with five variants in each window at a correlation threshold of 0.2. After pruning, we calculated relatedness and

inbreeding coefficients using the parameter ‘--make-rel’ in PLINK (Additional file 2). All reported SNP coordinates are relative to btau9 assembly converted with the LiftOver tool<sup>61</sup>.

**Library preparation and sequencing.** For sequencing library construction, 900 ng of total RNA was diluted into 25 µl of nuclease-free water, and RNA quantity was confirmed using the Qubit™ RNA High Sensitivity Assay kit (Invitrogen™, Thermo Fisher Scientific, Waltham, MA) on the Qubit™ 4 Fluorometer (Invitrogen™, Thermo Fisher Scientific, Waltham, MA). Libraries were prepared for next-generation sequencing using the Illumina Stranded mRNA Prep kit (Illumina, Inc., San Diego, CA) and the IDT® for Illumina RNA UD indexes (Illumina, Inc., San Diego, CA) according to the manufacturer’s instructions. Sequencing was conducted on the NovaSeq 6000 sequencing system (Illumina, Inc., San Diego, CA) using the NovaSeq 6000 SP Reagent kit v1.5 (Illumina, Inc., San Diego, CA) to produce paired-end reads 150 nucleotides in length. Sequencing was performed by the VANTAGE laboratory at Vanderbilt University Medical Center (Nashville, TN).

**Sequence alignment and filtering.** We aligned the sequences to the cattle reference genome (*Bos taurus*.ARS-UCD1.2.105) in the Ensembl<sup>62</sup> database with hisat2<sup>63–65</sup> using the -very-sensitive parameter. Then, we used Samtools<sup>66,67</sup> to filter sequences and remove secondary alignments, duplicates, and unmapped reads. Next, we used biobambam2<sup>68</sup> to mark and remove duplicates.

**Transcript quantification and gene filtering.** The number of fragments that matched to the Ensembl<sup>62</sup> cow gene annotation (*Bos taurus*.ARS-UCD1.2.105) was quantified using featureCounts<sup>69</sup>, and we preserved genes annotated as protein-coding, pseudogenes, or long non-coding RNA. Genes were then retained for further analysis if counts per million (CPM) and fragments per kilobase per million (FPKM) were >1 in at least five samples.

**Proteomics data and processing.** One hundred µl of plasma per sample was submitted to the Virginia Tech Mass Spectrometry Incubator (VT-MSI) facility at the Fralin Life Sciences Institute, Virginia Tech, for protein extraction and data collection.

Plasma samples (100 µl) were acidified by the addition of 11.1 µl 12% (v/v) o-phosphoric acid (MilliporeSigma, St. Louis, MO), then proteins were precipitated by the addition of 725 µl LC/MS grade methanol and incubated at -80°C overnight. Precipitated protein was collected by centrifugation and solubilized in S-trap lysis buffer (10% (w/v) SDS in 100 mM triethylammonium bicarbonate (MilliporeSigma, St. Louis, MO, pH 8.5)). Protein concentration was determined by measuring the absorbance at 280 nm, then 150 µg of protein for each sample was reduced using DTT (4.5 mM) then alkylated with iodoacetamide (10 mM, MilliporeSigma, St. Louis, MO). Unreacted I iodoacetamide was quenched with DTT (10 mM, MilliporeSigma, St. Louis, MO) and samples were acidified using o-phosphoric acid (MilliporeSigma, St. Louis, MO). Protein was again precipitated using methanol and incubated at -80°C overnight as above. Precipitated protein was loaded onto a micro S-trap and washed with methanol then digested overnight with trypsin. Peptides were recovered and five µg, as determined by measuring the absorbance at 215 nm using a DS-11 FX+ spectrophotometer/fluorometer (DeNovix, Wilmington, DE), of each sample was analyzed twice (duplicates) using ESI-MS/MS Orbitrap Fusion Lumos (Thermo Fisher Scientific (Waltham, MA)).

Samples were first loaded onto a precolumn (Acclaim PepMap 100 (Thermo Scientific, Waltham, MA), 100 µm × 2 cm) after which flow was diverted to an analytical column (50 cm µPAC (PharmaFluidics, Woburn, MA). The UPLC/autosampler utilized was an Easy-nLC 1200 (Thermo Scientific, Waltham, MA). Flow rate was maintained at 150 nl/min and peptides were eluted utilizing a 2 to 45% gradient of solvent B in solvent A over 88 min. Spray voltage on the µPAC compatible Easy-Spray emitter (PharmaFluidics, Woburn, MA) was 1300 volts, the ion transfer tube was maintained at 275 °C, the RF lens was set to 30% and the default charge state was set to 3.

MS data for the m/z range of 400–1500 was collected using the orbitrap at 120,000 resolution in positive profile mode with an AGC target of 4.0e5 and a maximum injection time of 50 ms. Peaks were filtered for MS/MS analysis based on having isotopic peak distribution expected of a peptide with an intensity above 2.0e4 and a charge state of 2–5. Peaks were excluded dynamically for 15 s after 1 scan with the MS/MS set to be collected at 45% of a chromatographic peak width with an expected peak width (FWHM) of 15 s. MS/MS data starting at m/z of 150 was collected using the orbitrap at 15000 resolution in positive centroid mode with an AGC target of 1.0e5 and a maximum injection time of 200 ms. Activation type was HCD stepped from 27 to 33.

Data were analyzed utilizing Proteome Discoverer 2.5 (Thermo Scientific, Waltham, MA) combining a Sequest HT and Mascot 2.7 (Matrix Science, Boston, MA) search into one result summary for each sample. Both searches utilized the UniProt reference *Bos taurus* proteome database and a common protein contaminant database provided with the Proteome Discoverer (PD) software package<sup>70</sup>. Each search assumed trypsin-specific peptides with the possibility of 2 missed cleavages, a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.1 Da. Sequest HT searches also included the PD software precursor detector node to identify MS/MS spectra containing peaks from more than one precursor. Sequest HT searches included a fixed modification of carbamidomethyl at Cys and the variable modifications of oxidation at Met and loss of Met at the N-terminus of a protein (required for using the INFERYS rescoring node). Peptide matches identified by Sequest HT were subjected to INFERYS rescoring to further optimize the number of peptides identified with high confidence.

Mascot searches included the following dynamic modifications in addition to the fixed modification of Cys alkylated by iodoacetamide (carbamidomethylated): oxidation of Met, acetylation of the protein N-terminus, cyclization of a peptide N-terminal Gln to pyro-Glu, and deamidation of Asn/Gln residues.

Protein identifications were reported at a 1% false discovery rate (high confidence) or at 5% false discovery rate (medium confidence) based on searches of decoy databases utilizing the same parameters as above. The

software matched peptide peaks across all runs, and protein quantities are the sum of all peptide intensities associated with the protein.

**Principal component analysis.** We carried out principal component analysis for the genotypes after pruning using the parameter ‘--pca’ in PLINK. The eigenvectors were used for plotting. For the transcriptome data, first we obtained the variant stabilized data using the function ‘vst’ from the R package ‘DESeq2’. Next, we calculated the components using the function ‘plotPCA’ in R. For the protein data, we averaged the values for each technical duplicate and used these values as input for the function ‘prcomp’ in R.

**Statistical analyses.** *SNP association analysis.* After filtering, 575,053 genotypes from 22 animals were used for association analysis conducted in PLINK<sup>60</sup> using Fisher’s exact test. We adjusted the nominal *P* values to correct for multiple hypothesis testing using the adaptive permutation procedure<sup>71</sup> in PLINK<sup>60</sup>. Locus association was inferred at  $\alpha = 1 \times 10^{-5}$ , as reported by The Wellcome Trust Case Control Consortium<sup>72</sup> for case-control studies, as well as by previous GWAS analyses of reproductive traits in cows or heifers<sup>4,25,32</sup>, which corresponded to an adjusted *P* value <0.005.

*Differential transcript abundance.* We compared transcript abundance between samples from each breed and each fertility group. The R packages ‘edgeR’<sup>73,74</sup>, with the quasi-likelihood test, and ‘DESeq2’<sup>75</sup>, using the Wald’s and likelihood test, were utilized to conduct the analyses. We adjusted the raw *P* values for multiple hypothesis testing by calculating the empirical false discovery rate (eFDR<sup>76</sup>), with 10,000 permutations. Differences in transcript abundance were deemed statistically significant when eFDR <0.002 in the results obtained from the three tests.

*Differential protein abundance.* To identify differential protein abundance that is robust to the algorithm utilized, we analyzed the protein data using two different algorithms. First, we transformed the protein data using natural logarithm ( $\text{Log}_e(x)$ ). We analyzed the transformed data using a generalized mixed model<sup>77</sup> using the R package ‘lme4’, which included the fertility group (Fertile or Sub-Fertile), breed (Angus or Holstein), and the random effect of the subject. Random effect was included in this analysis as samples were assayed twice to provide a more robust estimate of differential protein abundance. Then, we used the function ‘emmeans’, which tests the significance of the difference ( $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$ ) with the Student’s *t* test<sup>78</sup>, to calculate the estimated differences in protein abundance between fertility groups within each breed. We also analyzed the log-transformed data using the R package ‘limma’<sup>79</sup>. We accounted for the same independent variables mentioned above (fertility group and breed), in addition to accounting for the correlation between the duplicated data for each individual with the function ‘duplicateCorrelation’. We tested for a differential abundance of the identified proteins using the empirical Bayes Statistics implemented in the function ‘eBayes’<sup>80,81</sup>. In both analyses, we adjusted the nominal *P* values using FDR<sup>82</sup>. Significance was assumed if  $\text{FDR} < 0.05$  in both approaches.

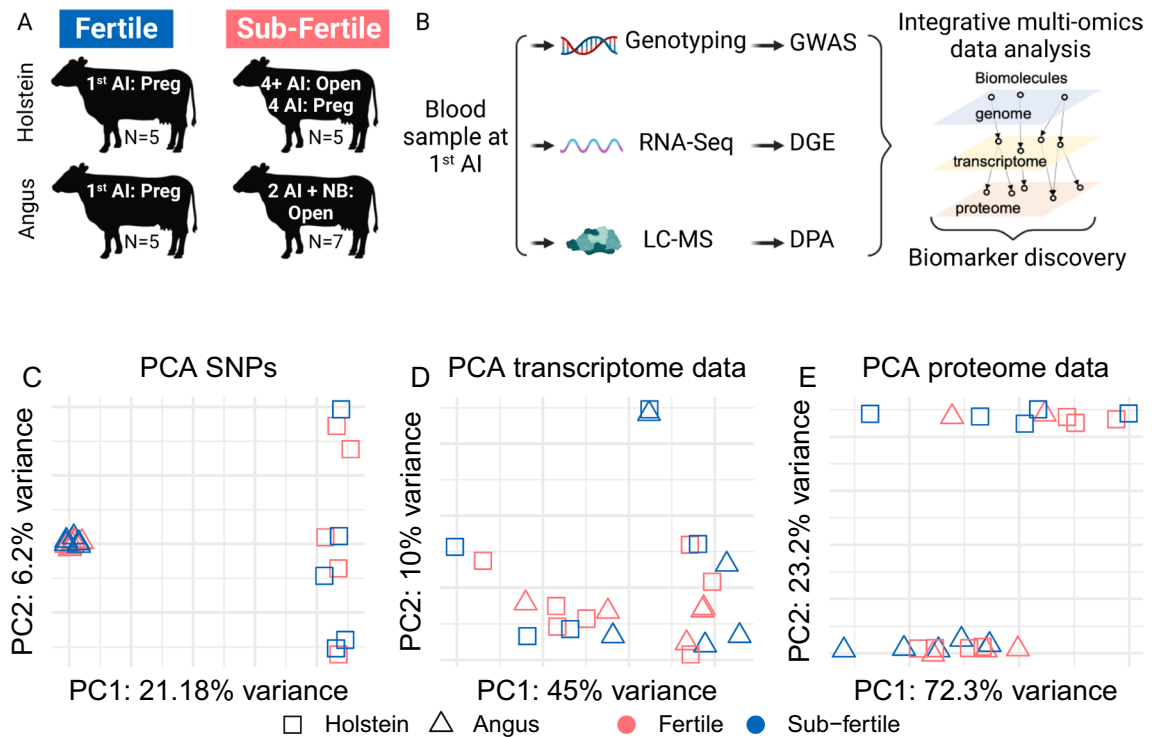
*Multi-omics factor analysis.* We analyzed the multimodal multi-omics datasets (genome, transcriptome, and proteome) interactively using Multi-omics Factor Analysis approach<sup>83,84</sup>. We subset the genotypes, transcriptome, and proteome data to reduce the global profiling. We retained SNPs with a *P* value < 0.001 for the Fisher’s test, genes with a *P* value < 0.01 for all three statistical tests employed, and proteins with a *P* value < 0.05 in both statistical tests used. We conducted the analysis using the R package ‘MOFA2’<sup>83,84</sup>, accounting for the breed as a group.

## Results

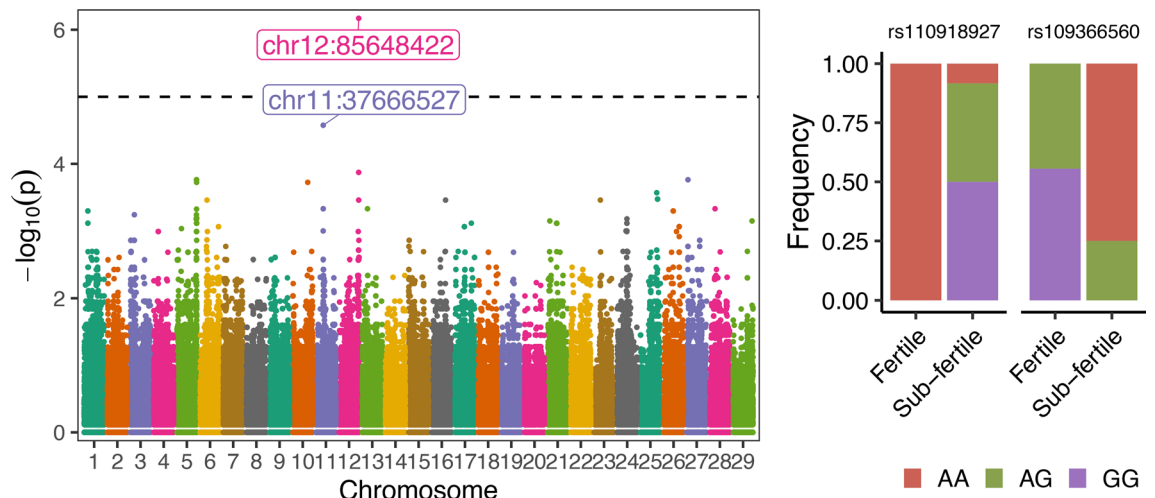
**Overview of the data produced.** We selected 22 *Bos taurus* heifers of Angus ( $n = 12$ ) and Holstein ( $n = 10$ ) breeds based on their fertility fitness (Fig. 1A). We isolated total RNA from circulating white blood cells, averaging  $16.3 \mu\text{g} \pm 4.0$ , and quality, measured by the RIN, averaging  $9.4 \pm 0.4$ . The extraction of genomic DNA yielded  $1.1 \mu\text{g} \pm 0.4$ . We produced RNA-sequencing data (Fig. 1B) and quantified the transcript abundance of 12,445 genes (12,105 protein-coding genes, 228 long non-coding RNAs, and 112 pseudogenes). We also analyzed 575,053 nucleotide positions across the bovine genome (Fig. 1B). Lastly, we produced untargeted proteomics data from plasma that resulted in the relative quantification of 213 proteins. As expected, the genotypic and proteomic data clustered the heifers of different genetic background separately (Fig. 1A). Conversely, there was no clustering of the samples based on the transcriptome data of the peripheral white blood cells (Fig. 1D,E).

**GWAS identifies SNPs associated with fertility in Angus and Holsteins heifers.** Our analysis identified two SNPs significantly associated with heifer fertility (rs110918927, chr12: 85648422,  $P = 6.7 \times 10^{-7}$ ; and rs109366560, chr11:37666527,  $P = 2.6 \times 10^{-5}$ , Fig. 2A, Additional file 3). For the SNP rs110918927, all heifers that delivered a calf after one artificial insemination presented the genotype AA ( $f_{(A)} = 1, f_{(G)} = 0$ ), whereas 11 out of 12 heifers classified as sub-fertile presented at least one copy of the allele G ( $f_{(A)} = 0.29, f_{(G)} = 0.71$ ). This polymorphism sits in an intergenic region of the genome with the closest gene located > 73 kilobases downstream relative to the SNP. For the SNP rs109366560, none of the heifers classified as sub-fertile were homozygous for the allele G ( $f_{(G)} = 0.12, f_{(A)} = 0.88$ ), and five out of the nine fertile heifers genotyped were homozygous GG ( $f_{(G)} = 0.78, f_{(A)} = 0.22$ ). This SNP is located on intron 22 of the gene Echinoderm microtubule associated protein like 6 (*EML6*).



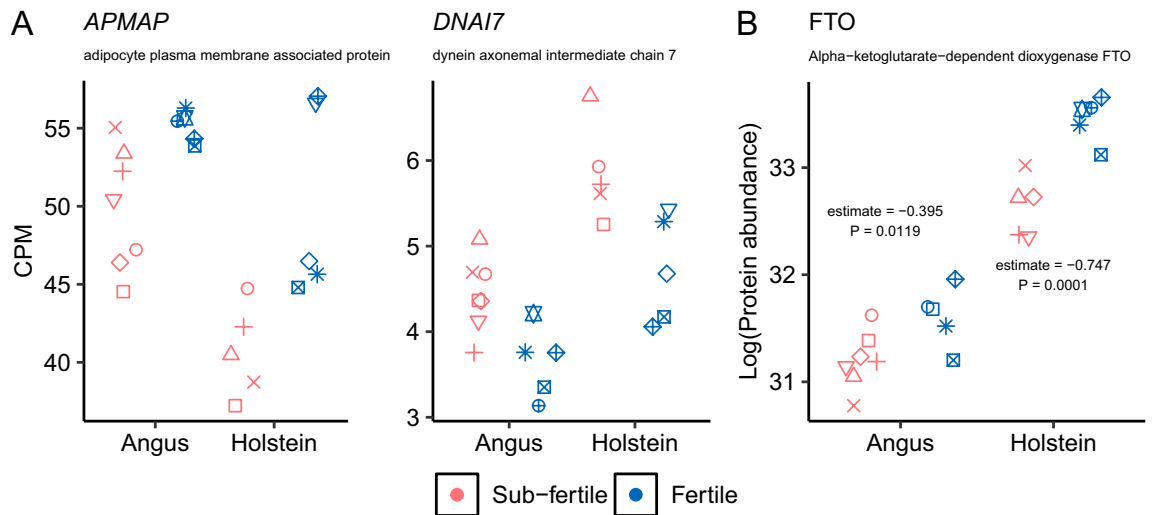


**Figure 1.** Overview of data produced. (A) Breeds and classification used in this study, including sample size. (B) Schematics of the data produced, and analysis undertaken. Principal component analysis of the genome-wide single nucleotide polymorphisms (C), transcriptome (D) and (E) proteome data. GWAS: genome-wide association analysis, DGE: differential gene expression, DPA: differential protein abundance.



**Figure 2.** Genome-wide association analysis of fertility in beef and dairy heifers. (A) Manhattan plot with the distribution of SNPs across their genome and their  $P$  values from Fisher's exact association test. (B) Genetic frequencies of the two SNPs that are putatively linked to fertility in heifers.

**Transcriptome analysis identifies differential transcript abundance between Fertile and Sub-Fertile heifers.** Next, we sought to determine if there were differences in transcript abundance from circulating white blood cells between the Fertile and Sub-Fertile heifer groups, accounting for their genetic background. We identified two genes whose transcript abundance differed (eFDR  $\leq 0.002$ ) between the two groups (Fertile and Sub-Fertile), namely Adipocyte Plasma Membrane Associated Protein (*APMAP*, 1.16 greater abundance in the Fertile group) and Dynein Axonemal Intermediate Chain 7 (*DNAI7*, 1.23 greater abundance in the Sub-Fertile group) (Fig. 3A, Additional file 4).



**Figure 3.** Differential transcript and protein abundance associated with fertility. **(A)** Transcript abundance. **(B)** Protein abundance. In each plot, within each breed, shapes indicate the same animal.

### Proteomic analysis identifies differential protein abundance between Fertile and Sub-Fertile heifers.

We also tested if there were differential abundance in proteins present in the plasma of heifers classified based on their fertility groups in both genetic backgrounds. The protein Alpha-ketoglutarate-dependent dioxygenase FTO was more abundant in the plasma collected from Fertile heifers relative to their Sub-Fertile counterparts (FDR < 0.05, Fig. 3B, Additional file 5).

### Integrative multi-omics analysis identifies molecular features that classify heifers based on their fertility potential.

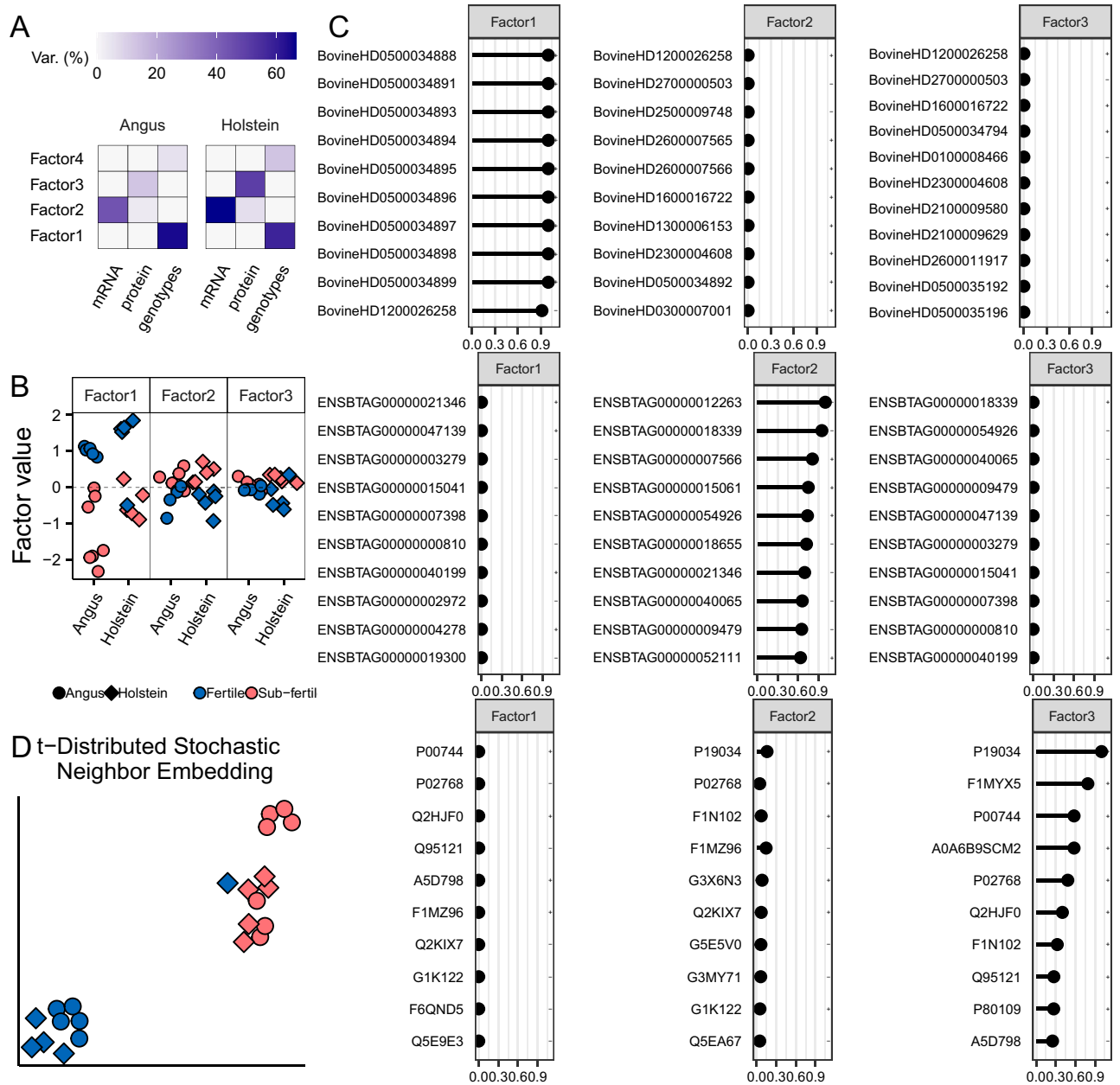
When each data were evaluated independently, the quantification of 22 and 23 gene and protein relative abundances accounted for 44.1% and 16.6% of the variance associated with fertility classification, respectively, and the genotypic information of 59 SNPs explained 70.1% of the variance associated with fertility classification. Overall, there were four factors identified in the analysis with the potential to distinguish the samples based on their fertility status, out of which three were most representative with Factors one, two, and three being mostly dominated by genotype, transcript, and protein data, respectively (Fig. 4A). Factors one, two, and three separated most of the samples based on their fertility classification except two, three, and five samples, respectively (Fig. 4B).

Notably, the top nine SNPs that explained most of the variance related to Factor one are located in a window on chromosome 5 spanning from nucleotide 118332762 to 118345383. The tenth SNP was the top significant polymorphism identified on chromosome 12 nucleotide 85648422 according to our Fisher's exact test contrasting heifers of different fertility potential (Fig. 4C). Among the genes whose transcript abundance explained the variance related to Factor two, we identified the following annotated genes: ArfGAP with SH3 domain, ankyrin repeat and PH domain 3 (*ASAP3*), ATP synthase membrane subunit c locus 1 (*ATP5MC1*), Centrosomal protein 170 (*CEP170*), Myeloid derived growth factor (*MYDGF*), Coiled-coil domain containing 34 (*CCDC34*), RAD51 associated protein 1 (*RAD51AP1*), and Ubiquinol-cytochrome c reductase complex III subunit VII (*UQCRCQ*) (Fig. 4C). Among the proteins whose abundance explained the variance related to Factor three, the following were annotated to known genes: Apolipoprotein C-II (*APOC2*), Lymphocyte cytosolic protein 1 (*LCPI1*), Vitamin K-dependent protein Z (*PROZ*), Albumin (*ALB*), Serotransferrin-like (*LOC525947*), Complement component C8 beta chain (*C8B*), Pigment epithelium-derived factor (*SERPINF1*), Phosphatidylinositol-glycan-specific phospholipase D (*GPLD1*), Alpha-ketoglutarate-dependent dioxygenase FTO (*FTO*) (Fig. 4C). Collectively, data from the genotypes, transcriptome, and proteome clustered 21 out of 22 heifers correctly based on their fertility status, with only one Fertile heifer clustering with the group of Sub-Fertile heifers.

### Discussion

Reproduction is a multidimensional biological function in mammals that can be partitioned into multiple components or traits<sup>85</sup>, and as a consequence, infertility is a complex phenotype with multifactorial origins, including a strong genetic component<sup>53,54</sup>. Our study was not designed to identify molecular markers for future use in selection programs. Rather, our work addressed two critical questions regarding the underlying biology of infertility: (a) whether multiple layers of molecular information, present in the circulatory system, would differ based on female fertility fitness; and (b) whether the integrative analysis of multiple layers of molecular information would be a better predictor of the causes of infertility. Our analysis identified molecular signatures in the genome, transcriptome, and proteome that provide important insights about the root causes of infertility.

Neither one of the significant SNPs were located in a region previously associated with female reproductive traits<sup>86</sup>. These SNPs have also not been previously reported to be associated with fertility traits in previous investigations that focused on sire-centric models<sup>4,33–37</sup>, nor on studies that focused on genotyped heifers only<sup>32,38</sup>.



**Figure 4.** Multi-omics analysis of heifer fertility. **(A)** Percentage of variance explained by each factor within each data modality. **(B)** Relative separation of samples based on their phenotype for each factor plotted. **(C)** The relative weight of importance of the top ten features (SNP identifiers are from the array annotation, gene identifiers are from Ensembl, protein identifiers are from UniProt) in each data modality and factor plotted. **(D)** sample clustering using all three data modes.

However, it is notable that the polymorphism rs110918927 is in the gene *EML6*, which produces a protein that participates in the function of spindle microtubules in oocytes<sup>87</sup>. Knockdown of this protein in mice oocytes at the germinal vesicle stage impairs spindle morphology and increases aneuploidy<sup>87</sup> in oocytes that progress to the metaphase II stage in the absence of *EML6*<sup>88</sup>. The gene *EML6* also produces transcripts in bovine oocytes<sup>89</sup>, and the significant SNP in this gene is a strong indication of a functional connection to reduced oocyte developmental competence in the Sub-Fertile group of heifers.

Genes differentially expressed in the peripheral white blood cells have been associated with fertility in heifers<sup>39–41</sup>. The protein APMAP exhibits arylesterase activity, which is known to protect lipoproteins from oxidation<sup>90</sup>. Importantly, the APMAP protein regulates adipose composition and metabolic health, and the disruption of the *APMAP* gene in mice leads to an increase in visceral adipose tissue expansion<sup>91</sup>. This protein was also shown to be less abundant in the omental tissue of women diagnosed with polycystic ovary syndrome<sup>92</sup>. Therefore, lower expression of *APMAP* in the peripheral white blood cells of Sub-Fertile heifers is possibly connected with a metabolic, hormonal or inflammatory disorder that disrupts fertility in heifers.



The Protein DNAI7 composes the axonemal dynein complex and participates in beta-tubulin binding activity and microtubule binding activity, and thus contributes to ciliary beating<sup>93</sup>. Variants that impair the function of DNAI7 are associated with Primary Ciliary Dyskinesia, with one potential consequence being the abnormal function of cilia and possible impaired transport of the cleaving embryos into the uterus<sup>94</sup>. DNAI7 may also function as a cell cycle regulator, and dysregulated transcript abundance of DNAI7 was associated with nasopharyngeal neoplasm in mice<sup>95</sup> and lung adenocarcinoma in humans<sup>96</sup>. Since Sub-Fertile heifers have greater abundance of *DNAI7* transcripts in their circulating white blood cells, it is possible that dysregulation in the cell cycle has a biological link with subfertility. Further research is required, however, to evaluate whether a dysregulation in the cell cycle linked to upregulation of *DNAI7* is connected with increased inflammation<sup>91</sup> associated with less transcripts from *APMAP*.

The protein Alpha-ketoglutarate-dependent dioxygenase FTO has oxidative demethylation activity of abundant N6-methyladenosine (m<sup>6</sup>A) residues in RNA<sup>97</sup>. The protein FTO preferentially demethylates N6,2'-O-dimethyladenosine (m<sup>6</sup>A<sub>m</sub>) rather than m<sup>6</sup>A and contributes to a reduced stability of m<sup>6</sup>A<sub>m</sub> mRNAs<sup>98</sup>. On a systemic level, genomic variants in FTO were associated with symptoms of metabolic disorders<sup>99</sup>, although the effects observed in humans, such elevated body mass index<sup>100,101</sup>, and mice<sup>102</sup> may be contradictory. Also worth noting, a variant on the FTO gene was associated with polycystic ovary syndrome<sup>103</sup>. Interestingly, in mice, the FTO gene is downregulated due to a deficiency in essential amino-acids<sup>104</sup>, and deficiency in the FTO protein causes postnatal growth retardation and a significant reduction in adipose tissue and lean body mass<sup>105</sup>. Our observation of the FTO abundance in heifers of different fertility potential is an indication that Sub-Fertile heifers could be experiencing a metabolic imbalance, contributing to their lower fertility. We note that the heifers utilized in this experiment were not nutritionally challenged and thus, our observations are a consequence of their intrinsic biological system and how it may utilize nutrients.

The next step was to interrogate the data we produced in a comprehensive manner. Interestingly, the largest source of variability was observed in the genomic data. Nine of the top ten SNPs that were assigned to Factor one were located in an intron of the TAF<sub>A</sub> chemokine-like family member 5 (*TAF<sub>A</sub>5*) gene. These SNPs are within a quantitative trait loci for milk yield<sup>106</sup>, a trait negatively correlated with reproductive traits<sup>107</sup>, however, no relationship between genetic variants in this gene and female fertility has been reported previously. None of the top ten genes with transcript abundance relevant for the modeling of the variance were identified as differentially expressed when analyzed independently. This result is not surprising because the identification of significant features using standard statistical approaches for association analysis is not necessarily the best approach for identifying predictive genes associated with complex traits<sup>108,109</sup>. It was surprising that three out of nine annotated proteins, which composed the top ten proteins that explained most of the variance in factor three, were also identified in our analyses using general linear mixed models. The most interesting result, however, was that all three data modalities were able to separate 21 out of 22 heifers correctly based on their fertility potential. Our results show that molecular differences have strong signals linked to fertility fitness that surpasses their differing genetic background.

## Conclusions

Our interrogation of multiple levels of biological information (genome, transcriptome, and proteome) at a systemic level in heifers highlighted the molecular complexity of female fertility. While the genomic data pointed to a disruption of oocyte developmental competence, the transcriptome and proteomic data point to metabolic dysregulation contributing to subfertility or infertility. Although the differences in molecular profiles identified in our study need to be further validated by mechanistic studies, our results, supported by the current literature, highlight differences in the molecular profile associated with female fertility that transcend the constraints of breed-specific genetic background.

## Data availability

The transcriptome and proteome data generated and analyzed during the current study are available in the Gene Expression Omnibus and ProteomeXchange repositories under the following identifiers: GSE220220 and PXD038756, respectively. The genotypic data are available from the corresponding author upon reasonable request.

Received: 25 April 2023; Accepted: 1 August 2023

Published online: 04 August 2023

## References

- Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M. & Tiwari, B. Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods* **6**(7), 53 (2017).
- Davis, T. C. & White, R. R. Breeding animals to feed people: The many roles of animal reproduction in ensuring global food security. *Theriogenology* **150**, 27–33 (2020).
- Moorey, S. E. & Biase, F. H. Beef heifer fertility: Importance of management practices and technological advancements. *J. Anim. Sci. Biotechnol.* **11**, 97 (2020).
- Galliou, J. M. *et al.* Identification of loci and pathways associated with heifer conception rate in U.S. Holsteins. *Genes (Basel)* **11**(7), 767 (2020).
- 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.* **91**(9), 4486–4491 (2013).
- 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), 1–6 (1973).
- Damiran, D., Larson, K. A., Pearce, L. T., Erickson, N. E. & Lardner, B. H. A. Effect of calving period on beef cow longevity and lifetime productivity in western Canada. *Transl. Anim. Sci.* **2**(suppl\_1), S61–S65 (2018).

8. Boulton, A. C., Rushton, J. & Wathes, D. C. A study of dairy heifer rearing practices from birth to weaning and their associated costs on UK dairy farms. *Open J. Anim. Sci.* **5**, 185–197 (2015).
9. Hoffman, P. C. Optimum body size of Holstein replacement heifers. *J. Anim. Sci.* **75**(3), 836–845 (1997).
10. Heinrichs, A. J. Raising dairy replacements to meet the needs of the twenty-first century. *J. Dairy Sci.* **76**(10), 3179–3187 (1993).
11. Bormann, J. M., Totir, L. R., Kachman, S. D., Fernando, R. L. & Wilson, D. E. Pregnancy rate and first-service conception rate in Angus heifers. *J. Anim. Sci.* **84**, 2022–2025 (2006).
12. Peters, S. O. *et al.* Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. *J. Anim. Sci.* **90**(10), 3398–3409 (2012).
13. Fortes, M. R. S. *et al.* Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *J. Anim. Sci.* **90**, 2894–2906 (2012).
14. Doyle, S. P., Golden, B. L., Green, R. D. & Brinks, J. S. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. *J. Anim. Sci.* **78**(8), 2091–2098 (2000).
15. Toghiani, S. *et al.* Genomic prediction of continuous and binary fertility traits of females in a composite beef cattle breed. *J. Anim. Sci.* **95**(11), 4787–4795 (2017).
16. McAllister, C. M., Speidel, S. E., Crews, D. H. Jr. & Enns, R. M. Genetic parameters for intramuscular fat percentage, marbling score, scrotal circumference, and heifer pregnancy in Red Angus cattle. *J. Anim. Sci.* **89**(7), 2068–2072 (2011).
17. Bodhireddy, P. *et al.* Genomic predictions in Angus cattle: Comparisons of sample size, response variables, and clustering methods for cross-validation. *J. Anim. Sci.* **92**(2), 485–497 (2014).
18. Raheja, K. L., Burnside, E. B. & Schaeffer, L. R. Heifer fertility and its relationship with cow fertility and production traits in Holstein dairy cattle. *J. Dairy Sci.* **72**(10), 2665–2669 (1989).
19. Muir, B. L., Fatehi, J. & Schaeffer, L. R. Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holsteins. *J. Dairy Sci.* **87**(9), 3029–3037 (2004).
20. Walsh, S. W., Williams, E. J. & Evans, A. C. A review of the causes of poor fertility in high milk producing dairy cows. *Anim. Reprod Sci.* **123**(3–4), 127–138 (2011).
21. Kuhn, M. T., Hutchison, J. L. & Wiggans, G. R. Characterization of Holstein heifer fertility in the United States. *J. Dairy Sci.* **89**(12), 4907–4920 (2006).
22. Tiezzi, F., Maltecca, C., Cecchinato, A., Penasa, M. & Bittante, G. Genetic parameters for fertility of dairy heifers and cows at different parities and relationships with production traits in first lactation. *J. Dairy Sci.* **95**(12), 7355–7362 (2012).
23. Jagusiak, W. Fertility measures in Polish Black-and-White cattle. 1. Genetic parameters of heifer fertility traits. *J. Anim. Feed Sci.* **14**(3), 423–433 (2005).
24. Koltes, J. E. *et al.* A vision for development and utilization of high-throughput phenotyping and big data analytics in livestock. *Front. Genet.* **10**, 1197 (2019).
25. Neupane, M. *et al.* Loci and pathways associated with uterine capacity for pregnancy and fertility in beef cattle. *PLoS ONE* **12**(12), e0188997 (2017).
26. McDaneld, T. G. *et al.* Genomewide association study of reproductive efficiency in female cattle. *J. Anim. Sci.* **92**, 1945–1957 (2014).
27. McDaneld, T. G. *et al.* Y are you not pregnant: Identification of Y chromosome segments in female cattle with decreased reproductive efficiency. *J. Anim. Sci.* **90**, 2142–2151 (2012).
28. de Camargo, G. M. *et al.* Association between JY-1 gene polymorphisms and reproductive traits in beef cattle. *Gene* **533**(2), 477–480 (2014).
29. Dias, M. M. *et al.* Study of lipid metabolism-related genes as candidate genes of sexual precocity in Nelore cattle. *Genet. Mol. Res.* **14**(1), 234–243 (2015).
30. Irano, N. *et al.* Genome-wide association study for indicator traits of sexual precocity in Nelore cattle. *PLoS ONE* **11**(8), e0159502 (2016).
31. Junior, G. A. O. *et al.* Genomic study and medical subject headings enrichment analysis of early pregnancy rate and antral follicle numbers in Nelore heifers. *J. Anim. Sci.* **95**(11), 4796–4812 (2017).
32. Stegemiller, M. R. *et al.* Genome-wide association analyses of fertility traits in beef heifers. *Genes (Basel)* **12**(2), 217 (2021).
33. Frischknecht, M. *et al.* Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. *BMC Genomics* **18**(1), 910 (2017).
34. Jiang, J. *et al.* A Large-scale genome-wide association study in U.S. Holstein cattle. *Front. Genet.* **10**, 412 (2019).
35. Nayeri, S. *et al.* Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* **17**(1), 75 (2016).
36. Chen, S. Y. *et al.* Identifying pleiotropic variants and candidate genes for fertility and reproduction traits in Holstein cattle via association studies based on imputed whole-genome sequence genotypes. *BMC Genomics* **23**(1), 331 (2022).
37. Kiser, J. N. *et al.* Validation of 46 loci associated with female fertility traits in cattle. *BMC Genomics* **20**(1), 576 (2019).
38. Kiser, J. N. *et al.* Identification of loci associated with conception rate in primiparous Holstein cows. *BMC Genomics* **20**(1), 840 (2019).
39. Moorey, S. E. *et al.* Rewiring of gene expression in circulating white blood cells is associated with pregnancy outcome in heifers (*Bos taurus*). *Sci. Rep.* **10**(1), 16786 (2020).
40. Dickinson, S. E. *et al.* Transcriptome profiles in peripheral white blood cells at the time of artificial insemination discriminate beef heifers with different fertility potential. *BMC Genomics* <https://doi.org/10.1186/s12864-018-4505-4> (2018).
41. 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 (2018).
42. Phillips, K. M. *et al.* Plasma metabolomic profiles differ at the time of artificial insemination based on pregnancy outcome, in *Bos taurus* beef heifers. *Sci. Rep.* **8**(1), 13196 (2018).
43. Bormann, J. M., Totir, L. R., Kachman, S. D., Fernando, R. L. & Wilson, D. E. Pregnancy rate and first-service conception rate in Angus heifers. *J. Anim. Sci.* **84**(8), 2022–2025 (2006).
44. Pryce, J. E., Royal, M. D., Garnsworthy, P. C. & Mao, I. L. Fertility in the high-producing dairy cow. *Livest. Prod. Sci.* **86**(1–3), 125–135 (2004).
45. Wathes, D. C. *et al.* Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. *Theriogenology* **68**(Suppl 1), S232–S241 (2007).
46. Diskin, M. G. & Kenny, D. A. Managing the reproductive performance of beef cows. *Theriogenology* **86**(1), 379–387 (2016).
47. Randel, R. D. Nutrition and postpartum rebreeding in cattle. *J. Anim. Sci.* **68**(3), 853–862 (1990).
48. Breuel, K. F. *et al.* Factors affecting fertility in the postpartum cow: Role of the oocyte and follicle in conception rate. *Biol. Reprod.* **48**(3), 655–661 (1993).
49. Okano, A. & Tomizuka, T. Ultrasonic observation of postpartum uterine involution in the cow. *Theriogenology* **27**(2), 369–376 (1987).
50. Sheldon, I. M., Williams, E. J., Miller, A. N., Nash, D. M. & Herath, S. Uterine diseases in cattle after parturition. *Vet. J.* **176**(1), 115–121 (2008).
51. Sheldon, I. M., Molinari, P. C. C., Ormsby, T. J. R. & Bromfield, J. J. Preventing postpartum uterine disease in dairy cattle depends on avoiding, tolerating and resisting pathogenic bacteria. *Theriogenology* **150**, 158–165 (2020).

52. Sheldon, I. M., Lewis, G. S., LeBlanc, S. & Gilbert, R. O. Defining postpartum uterine disease in cattle. *Theriogenology* **65**(8), 1516–1530 (2006).
53. Crowley Jr., W.F., Pitteloud, N., & Seminara, S. New genes controlling human reproduction and how you find them. *Trans. Am. Clin. Climatol. Assoc.* **119**, 29–37; discussion -8 (2008).
54. Yatsenko, S. A. & Rajkovic, A. Genetics of human female infertility. *Biol. Reprod.* **101**(3), 549–566 (2019).
55. Mann, C. J. Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emerg. Med. J.* **20**(1), 54–60 (2003).
56. Lima, F. S. *et al.* Hormonal manipulations in the 5-day timed artificial insemination protocol to optimize estrous cycle synchrony and fertility in dairy heifers. *J. Dairy Sci.* **96**(11), 7054–7065 (2013).
57. Patterson, D., Kojima, F. & Smith, M. A review of methods to synchronize estrus in replacement beef heifers and postpartum cows. *J. Anim. Sci.* **81**(14\_suppl\_2), E166–E177 (2003).
58. Wilson, C., Dias, N. W., Pancini, S., Mercadante, V. & Biase, F. H. Delayed processing of blood samples impairs the accuracy of mRNA-based biomarkers. *Sci. Rep.* **12**(1), 8196 (2022).
59. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**(9), 1564–1573 (2010).
60. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**(3), 559–575 (2007).
61. Hinrichs, A. S. *et al.* The UCSC genome browser database: Update 2006. *Nucleic Acids Res.* **34**(90001), D590–D598 (2006).
62. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**(D1), D749–D755 (2014).
63. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**(9), 1650–1667 (2016).
64. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**(4), 357–360 (2015).
65. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**(8), 907–915 (2019).
66. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* <https://doi.org/10.1093/gigascience/giab008> (2021).
67. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).
68. Tischler, G. & Leonard, S. biobambam: Tools for read pair collation based algorithms on BAM files. *Source Code Biol. Med.* <https://doi.org/10.1186/1751-0473-9-13> (2014).
69. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**(7), 923–930 (2014).
70. Orsburn, B. C. Proteome Discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* <https://doi.org/10.3390/proteomes9010015> (2021).
71. Che, R., Jack, J. R., Motsinger-Reif, A. A. & Brown, C. C. An adaptive permutation approach for genome-wide association study: Evaluation and recommendations for use. *BioData Min.* **7**, 9 (2014).
72. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature.* **447**(7145), 661–678 (2007).
73. McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
74. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
75. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
76. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**(16), 9440–9445 (2003).
77. Nakayasu, E. S. *et al.* Tutorial: Best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **16**(8), 3737–3760 (2021).
78. Kalpić, D., Hlupić, N. & Lovrić, M. Student's t-tests. In *International Encyclopedia of Statistical Science* (ed. Lovric, M.) 1559–1563 (Springer, 2011).
79. Smyth, G. K. *Limma: Linear Models for Microarray Data* 397–420 (Springer, 2005).
80. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genetics Mol. Biol.* <https://doi.org/10.2202/1544-6115.1027> (2004).
81. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hyper-variable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**(2), 946–963 (2016).
82. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Met.* **57**(1), 289–300 (1995).
83. Argelaguet, R. *et al.* Multi-omics factor analysis—A framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**(6), e8124 (2018).
84. Argelaguet, R. *et al.* MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**(1), 111 (2020).
85. Orr, T. J. & Garland, T. Jr. Complex reproductive traits and whole-organism performance. *Integr. Comp. Biol.* **57**(2), 407–422 (2017).
86. Hu, Z. L., Park, C. A. & Reecy, J. M. Bringing the Animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. *Nucleic Acids Res.* **50**(D1), D956–D961 (2022).
87. Yin, H., Hou, X., Zhang, T., Shi, L. & Su, Y. Q. Participation of EML6 in the regulation of oocyte meiotic progression in mice. *J. Biomed. Res.* **34**(1), 44–53 (2019).
88. Yin, H. *et al.* Echinoderm microtubule associated protein like 1 is indispensable for oocyte spindle assembly and meiotic progression in mice. *Front. Cell Dev. Biol.* **9**, 687522 (2021).
89. Walker, B. N. & Biase, F. H. The blueprint of RNA storages relative to oocyte developmental competence in cattle (*Bos taurus*). *Biol. Reprod.* **102**(4), 784–794 (2020).
90. de la Iglesia, R., Mansego, M. L., Sanchez-Muniz, F. J., Zulet, M. A. & Martinez, J. A. Arylesterase activity is associated with antioxidant intake and paraoxonase-1 (PON1) gene methylation in metabolic syndrome patients following an energy restricted diet. *EXCLI J.* **13**, 416–426 (2014).
91. Pessenheiner, A. R. *et al.* APMAP interacts with lysyl oxidase-like proteins, and disruption of Apmmap leads to beneficial visceral adipose tissue expansion. *FASEB J.* **31**(9), 4088–4103 (2017).
92. Corton, M. *et al.* Proteomic analysis of human omental adipose tissue in the polycystic ovary syndrome using two-dimensional difference gel electrophoresis and mass spectrometry. *Hum. Reprod.* **23**(3), 651–661 (2008).
93. Braschi, B. *et al.* Consensus nomenclature for dyneins and associated assembly factors. *J. Cell Biol.* <https://doi.org/10.1083/jcb.202109014> (2022).
94. Blyth, M. & Wellesley, D. Ectopic pregnancy in primary ciliary dyskinesia. *J. Obstet Gynaecol.* **28**(3), 358 (2008).
95. Manenti, G. *et al.* Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (Pas1) locus activity in mice. *Oncogene* **23**(25), 4495–4504 (2004).

96. Kabbout, M. *et al.* ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clin. Cancer Res.* **19**(13), 3383–3395 (2013).
97. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**(12), 885–887 (2011).
98. Mauer, J. *et al.* Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. *Nature* **541**(7637), 371–375 (2017).
99. Zhao, X., Yang, Y., Sun, B. F., Zhao, Y. L. & Yang, Y. G. FTO and obesity: Mechanisms of association. *Curr. Diab. Rep.* **14**(5), 486 (2014).
100. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**(1), 25–34 (2009).
101. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**(1), 18–24 (2009).
102. Fischer, J. *et al.* Inactivation of the Fto gene protects from obesity. *Nature* **458**(7240), 894–898 (2009).
103. Liu, A. L. *et al.* Association between fat mass and obesity associated (FTO) gene rs9939609 A/T polymorphism and polycystic ovary syndrome: A systematic review and meta-analysis. *BMC Med. Genet.* **18**(1), 89 (2017).
104. Cheung, M. K., Gulati, P., O'Rahilly, S. & Yeo, G. S. FTO expression is regulated by availability of essential amino acids. *Int. J. Obes. (Lond.)* **37**(5), 744–747 (2013).
105. Boissel, S. *et al.* Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *Am. J. Hum. Genet.* **85**(1), 106–111 (2009).
106. Michenet, A., Saintilan, R., Venot, E. & Phocas, F. Insights into the genetic variation of maternal behavior and suckling performance of continental beef cows. *Genet. Sel. Evol.* **48**(1), 45 (2016).
107. Albarran-Portillo, B. & Pollott, G. E. The relationship between fertility and lactation characteristics in Holstein cows on United Kingdom commercial dairy farms. *J. Dairy Sci.* **96**(1), 635–646 (2013).
108. An, N., Yu, Z. & Yang, X. Expression differentiation is not helpful in identifying prognostic genes based on TCGA datasets. *Mol. Ther. Nucleic Acids.* **11**, 292–299 (2018).
109. Lo, A., Chernoff, H., Zheng, T. & Lo, S. H. Why significant variables aren't automatically good predictors. *Proc. Natl. Acad. Sci. U.S.A.* **112**(45), 13892–13897 (2015).

## Acknowledgements

We express our appreciation for Chad Joines, Director of Beef Cattle Operations, Shane Brannock, Director of Dairy Complex, and their respective crews for supporting our sampling. We also thank Jada Nix from our group for the critical reading of our manuscript.

## Author contributions

FB designed, supervised, acquired funding and wrote the paper. FB collected and processed the samples for data collection. MM contributed to funding acquisition, analysis, and interpretation of the data. All authors have read and approved the submitted version.

## Funding

The authors thank the funding from the Virginia Agriculture Council, Virginia Cattle Industry Board, Virginia State Dairy Association. This project was partially supported by Agriculture and Food Research Initiative Competitive Grant no. 2020-67015-31616 from the USDA National Institute of Food and Agriculture. Funding institutions had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-39858-0>.

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

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023