Received: 27 November 2022 | Revised: 13 January 2023 | Accepted: 26 January 2023

DOI: 10.1096/fj.202201977R

The FASEB Journal

RESEARCH ARTICLE

Machine-learning methods applied to integrated transcriptomic data from bovine blastocysts and elongating conceptuses to identify genes predictive of embryonic competence

Maria Belen Rabaglino[1] | **Dessie Salilew-Wondim**[2,3] | **Adriana Zolini**[4] | **Dawit Tesfaye**[5] | **Michael Hoelker**[3] | **Pat Lonergan**[1] | **Peter J. Hansen**[4]

[1]School of Agriculture and Food Science, University College Dublin, Dublin 4, Ireland

[2]Institute of Animal Sciences, Animal Breeding, University of Bonn, Bonn, Germany

[3]Department of Animal Science, Biotechnology & Reproduction in Farm Animals, University of Goettingen, Goettingen, Germany

[4]Department of Animal Sciences, D.H. Barron Reproductive and Perinatal Biology Research Program, and Genetics Institute, University of Florida, Gainesville, Florida, USA

[5]Animal Reproduction and Biotechnology Laboratory, Department of Biomedical Sciences, Colorado State University, Fort Collins, Colorado, USA

Correspondence
Maria Belen Rabaglino, School of Agriculture and Food Science, University College Dublin, Belfield, Dublin 4, Ireland.
Email: maria.rabaglino@ucd.ie

Funding information
EC | H2020 | PRIORITY 'Excellent science' | H2020 Marie Skłodowska-Curie Actions (MSCA), Grant/Award Number: 101021311

Abstract

Early pregnancy loss markedly impacts reproductive efficiency in cattle. The objectives were to model a biologically relevant gene signature predicting embryonic competence for survival after integrating transcriptomic data from blastocysts and elongating conceptuses with different developmental capacities and to validate the potential biomarkers with independent embryonic data sets through the application of machine-learning algorithms. First, two data sets from in vivo-produced blastocysts competent or not to sustain a pregnancy were integrated with a data set from long and short day-15 conceptuses. A statistical contrast determined differentially expressed genes (DEG) increasing in expression from a competent blastocyst to a long conceptus and vice versa; these were enriched for KEGG pathways related to glycolysis/gluconeogenesis and RNA processing, respectively. Next, the most discriminative DEG between blastocysts that resulted or did not in pregnancy were selected by linear discriminant analysis. These eight putative biomarker genes were validated by modeling their expression in competent or noncompetent blastocysts through Bayesian logistic regression or neural networks and predicting embryo developmental fate in four external data sets consisting of in vitro-produced blastocysts (i) competent or not, or (ii) exposed or not to detrimental conditions during culture, and elongated conceptuses (iii) of different length, or (iv) developed in the uteri of high- or subfertile heifers. Predictions for each data set were more than 85% accurate, suggesting that these genes play a key role in embryo development and pregnancy establishment. In conclusion, this study integrated transcriptomic data from seven independent

Abbreviations: AAPH, 2,2-azobis (2- amidinopropane) dihydrochloride; Bl, blastocyst dataset; BLR, Bayesian logistic regression; BSO, buthionine sulfoximine; Cc, conceptus dataset; DEG, differentially expressed genes; ET, Embryo transfer; FDR, False discovery rate; HF, High-fertile; IFNT, interferon-tau; IVT, in vitro; IVV, In vitro; NEFA, non-esterified fatty acid; NN, neural network; NP, Non-pregnant; P4, Progesterone; PCA, principal component analysis; PR, pregnant; ROS, reactive oxygen species; SF, sub-fertile; sPLS-DA, sparse Partial Least-Squares Discriminant Analysis.

experiments to identify a small set of genes capable of predicting embryonic competence for survival.

KEYWORDS

bioinformatics, biomarker, embryo transfer, omics, pregnancy

1 | INTRODUCTION

Since the introduction of artificial insemination over 80 years ago, assisted reproductive technologies in cattle have resulted in enormous advances in genetic gain and productivity.[1,2] Among the toolbox of techniques currently available, the use of embryo transfer, particularly of in vitro (IVT)-produced embryos, continues to increase.[3] Implementation of embryo transfer (ET) allows the multiplication of the genetic merit of the donor, and it can represent a potential advantage compared with artificial insemination in terms of pregnancy outcomes as it bypasses potential errors between ovulation and the arrival of the embryo to the uterus, particularly in those scenarios, where the oocyte is more vulnerable than the embryo to suboptimal environmental conditions.[4] For example, in high-producing dairy cows during the postpartum period, which mobilize body fat in response to a negative energy balance, decreasing oocyte quality,[5] or in animals suffering heat stress.[6] Nevertheless, pregnancy loss due to early embryonic mortality is still a major barrier to advancing cattle productivity. Most losses occur early in gestation, often before any detectable interaction between the embryo and the maternal reproductive tract has occurred.[7] Later loss after day 30 of gestation is another feature of IVT-produced embryos (reviewed in[8]).

Embryonic mortality can occur because of intrinsic embryonic defects or inadequate endometrial receptivity. If the embryo is competent to maintain a pregnancy after the blastocyst stage, it will hatch from the zona pellucida and begin to grow. The trophectoderm proliferates, changing the embryonic morphology from a spherical to an oval shape and finally a filamentous form by around days 12 and 14. This layer secretes significant amounts of interferon-tau (IFNT) to induce the maternal recognition of pregnancy and ensure the maintenance of a functional corpus luteum for the production of progesterone (P4).[9,10]

Despite years of research, selecting embryos for transfer still relies on subjective assessment of morphology, usually following well-established guidelines from the International Embryo Technology Society (https://www.iets.org/). Thus, identifying a molecular signature associated with embryonic competence for pregnancy establishment is an urgent challenge. In a series of independent experiments, it has been demonstrated that the transcriptome of blastocysts, produced IVT[11–13] or in vivo (IVV),[12,14–16] differs for those embryos capable of establishing and maintaining a pregnancy to term (or at least until day 60 or 90 depending on the study) compared with those that fail to do so. In these studies, a similar experimental design was employed, consisting of obtaining an embryonic biopsy before ET and retrospectively classifying the embryo as competent (i.e., pregnant, PR) or noncompetent (i.e., nonpregnant, NP) according to the pregnancy outcome. Nonetheless, not a single gene related to embryonic competence was common in all studies.[12]

Another model of differential competence is related to the conceptus length around the period of maternal recognition of pregnancy. Increased conceptus length is associated with greater IFNT production[17] and, therefore, higher chances of pregnancy maintenance.[18,19] Furthermore, short conceptuses fail to upregulate a large number of IFNT-dependent and -independent genes in the endometrium, which are likely important for pregnancy establishment.[20] Barnwell et al.[21] characterized the transcriptome of short and long conceptuses at day 15 of gestation, demonstrating a difference in gene expression potentially associated with developmental capacity. There was no correlation between the top differentially expressed genes (DEG) reported in that study[21] and those described for the blastocyst experiments.[12] These discordances between the results might be attributed to the considerable variation between conditions employed in the studies. Therefore, there is a gap in the knowledge of critical genes involved in embryo survival. The aforementioned studies, while carried out independently and using a variety of transcriptomics platforms, provide an opportunity to combine the data to generate a gene expression signature of competent blastocysts and elongated conceptuses. This approach can identify molecular markers that could be used as predictors of embryonic developmental capacity and further advance knowledge of critical pathways involved in embryonic survival.

The objectives of the present study were to model a biologically relevant gene signature that predicts embryonic competence for survival after integrating transcriptomic data from blastocysts and elongating conceptuses with higher or lower chances of sustaining a pregnancy and to validate the potential biomarkers with independent embryonic data through the application of machine-learning algorithms.

2 | MATERIALS AND METHODS

The workflow employed in this study consists of three sequential steps: (1) integration and analysis of blastocyst and elongated conceptus data, (2) identification of biomarker genes, and (3) validation of the biomarker genes in independent data sets through machine-learning approaches. The data sets employed in these steps are summarized in Table 1.

2.1 | Transcriptomic data integration and analysis

2.1.1 | Data collection

Data were obtained from embryos and elongated conceptuses generated IVV to avoid any potential confounding influences induced by the IVT environment.[22] Blastocyst data (Bl) were downloaded from the data set GSE131178 in the Gene Expression Omnibus[16] or obtained after request.[12] These data sets were renamed Bl_Zo and Bl_SW, respectively. Elongating conceptus data were downloaded from the data set GSE75750.[21]

Data were processed using packages for the R software.[23] Bl_Zo data were generated through RNAseq. The raw counts were transformed through the variance stabilizing transformation method[24] using the vst function from the DESeq2 package.[25] The final data set consisted of samples obtained from blastocysts that resulted in pregnancy at day 60 after the transfer (pregnant, PR, $n = 4$) or not (nonpregnant, NP, $n = 13$). Bl_SW data were obtained after RNA hybridization with the two-channel Agilent-028298 Embryogene Bovine 45K microarray.[12] Data from each channel were processed with the limma package,[26] normalized by the quantile method and log2 transformed. Rows were collapsed to retain the microarray probe with the average value from the group of genes with the same Ensembl ID. Only data obtained from the IVV-produced blastocysts were retained in the final data set, containing PR ($n = 7$) and NP ($n = 5$) samples. Pregnant samples were those obtained from blastocysts that maintained pregnancy until term. Elongated conceptus data (Cc) were generated through hybridization with the Affymetrix Bovine Genome Array.[21] The raw data were processed with the gcRMA package[27] through transformation and normalization by the quantile method. The expression values of genes with the same Ensembl ID were averaged. This data set consisted of samples obtained from long conceptuses (24.7 ± 1.9 mm, $n = 5$) or short age-matched conceptuses (4.2 ± 0.1 mm, $n = 5$). Each blastocyst data set was combined with the conceptus data using the Combat function from the sva package[28] to remove the batch effect and to generate two integrated data sets: Bl_Zo-Cc and Bl_SW-Cc.

2.1.2 | Statistical analysis

For each integrated data set, genes that were differentially expressed between competent Bl-Cc (PR Bl and long Cc) and noncompetent Bl-Cc (NP Bl and short Cc) were determined through the following contrast: (PR_Bl + long Cc) – (NP_Bl + short Cc) after fitting a linear model by generalized least squares to the expression data, with the limma package.[26] This was the first step to later screen the data for biomarkers, so to cast a wide net, a p value of <.1 was employed as the statistical criterion to define a DEG. Next, DEG determined for each integrated data set were compared in a Venn Diagram. The overlapping genes between the up- or downregulated DEG for each integrated data set were employed for the subsequent analyses.

TABLE 1 Overview of the data sets employed for data analysis.

Data set accession no.[a]	Main characteristics	Total *n*	References
GSE131178—Requested	In vivo-produced blastocyst that resulted in pregnancy or not	29	Zolini et al.[16] Salilew-Wondim et al.[12]
GSE75750	Short and long day-15 conceptuses	10	Barnwell et al.[21]
GSE130954—Requested	In vitro-produced blastocyst that resulted in pregnancy or not	26	Zolini et al.[13] Salilew-Wondim et al.[12]
GSE42281—GSE83767	In vitro-produced blastocyst exposed to suboptimal conditions	34	Desmet et al.[33] Cagnone et al.[36]
GSE75750	Short and long day-15 conceptuses	15	Barnwell et al.[21]
GSE107891	Day-17 conceptuses transferred into high-fertile or subfertile heifers	23	Moraes et al.[37]

Note: Data from italicized data sets were integrated and analyzed to identify biomarker genes of pregnancy success. Data from bolded data sets were employed to validate the biomarker genes.

[a]Accession numbers correspond to the Gene Expression Omnibus database.

2.1.3 | Functional analysis

The DAVID software[29] was employed to determine the KEGG pathways enriched (q value or false discovery rate [FDR] < 0.05) by the competent or noncompetent genes.

2.2 | Identification of the biomarker genes

To be meaningful, the biomarker genes should be able to discriminate samples from PR and NP blastocysts. Thus, the first step was to integrate both IVV blastocyst data sets (Bl_Zo and Bl_SW) using the Combat function from the sva package[28] to generate a unique IVV blastocyst data set. Two NP Bl_SW samples identified as outliers were removed at this stage. The final data set consisted of samples from IVV-produced blastocysts (Bl_IVV) classified as PR ($n = 11$) or NP ($n = 16$). Genes involved in the enriched KEGG pathways determined in the previous step were selected as candidate biomarkers genes.

The most discriminative combination of genes representative of competent and noncompetent embryos was selected by sparse Partial Least Squares Discriminant Analysis or sPLS-DA.[30] The change in separation between groups (NP and PR blastocysts) as additional genes from the "competent" and "noncompetent" genes were added to the analysis, were evaluated through linear discriminant analysis, implemented with the lda function of the MASS package.[31] The final group of selected genes represented the minimal signature showing the highest coefficient of linear discriminants in the first component. In other words, the corresponding plot of discriminant function coefficients showed the maximum separation between PR and NP blastocysts.

Sample distribution according to the expression of the selected biomarkers was evaluated through a principal component analysis (PCA) and hierarchical clustering using internal packages of R. ShinyGO[32] was employed to determine enriched terms by the competent and noncompetent biomarker genes (FDR < 0.05).

2.3 | Validation of the biomarker genes

2.3.1 | Data sets for validation

Two data sets from blastocysts and two data sets from elongated conceptuses were used to validate the biomarker genes:

Blastocysts data sets: The first data set, namely an IVT Blastocyst data set (Bl_IVT), was compiled as follows. Data from the IVT blastocysts from the study of Salilew-Wondim et al.[12] were integrated with data from IVT blastocysts from the study of Zolini et al.,[13] with accession number GSE130954, generated through RNAseq. Data integration was performed as detailed for Bl_IVV. However, several samples from the data set GSE130954 had a high proportion of nonexpressed genes. Therefore, only samples with more than 30% of expressed genes were employed. The final integrated data set consisted of 9 PR and 17 NP samples. The second data set corresponded to embryos exposed to suboptimal conditions (Bl_SubOp) and was merged from two data sets: GSE42281 and GSE83767. Data from GSE42281 were generated from blastocysts that were cultured IVT under normal conditions ($n = 4$) or for 6.5 days with a combination of elevated nonesterified fatty acid (NEFA) concentrations ($n = 4$).[33] The concentrations used were equivalent to those found in the serum of cows experiencing negative energy balance, which is detrimental to embryonic development.[34,35] The data set GSE83767 consisted of data obtained from blastocysts that were cultured under normal ($n = 16$) or increased oxidative stress conditions ($n = 16$).[36] Oxidative stress was achieved by adding two pro-oxidant agents: 0.01 mM 2,2′-azobis (2-amidinopropane) dihydrochloride (AAPH), which promotes reactive oxygen species (ROS) production, and 0.4 mM buthionine sulfoximine (BSO), which inhibits glutathione synthesis. These reagents were added separately to the culture medium from day 3 onward. In both studies, the extracted mRNA from control or treated blastocysts was hybridized to the two-channel Agilent-028298 Embryogene Bovine 45 K microarray. Data were processed as detailed for Bl_SW. Data from three arrays identified as outliers were discarded. Samples included in the integrated data set were 17 cultured under normal conditions and 17 cultured under suboptimal conditions.

Elongated conceptus data sets: the data set GSE75750, named CcGr3, consisted of transcriptomic data from long- (24.7 ± 1.9 mm, $n = 5$) and short (4.2 ± 0.1 mm, $n = 5$)-elongated conceptuses (employed in the step 1) plus data from elongated conceptuses that were derived from Grade 3 blastocysts (15.2 ± 2.4 mm, $n = 3$).[21] The other data set was downloaded from GSE107891[37] and consisted of data from elongated conceptuses derived from IVV-produced blastocyst transferred to high-fertile (HF) or subfertile (SF, $n = 10$) heifers (Cc_HFSF). Elongated conceptuses from HF heifers were classified as short (1.2–6.9 cm; $n = 5$) or long (9.8–32.2 cm; $n = 8$). Count data were transformed to stabilize the variance.

2.3.2 | Model construction for machine learning

The model consisted of training and testing data sets. The training data set involved the expression of the biomarker genes in the Bl_IVV data set. The testing data sets were the

four specified above. An addon batch effect adjustment of the testing data with the training data was performed with the bapred package.[38] Two methods were employed for the classification of the testing data, according to the training data: The first method was a Bayesian logistic regression (BLR) model, implemented through a generalized Bayesian algorithm with a binomial function with the caret package.[39] This model does not require tuning parameters. The second method was a neural network (NN) model, implemented with TensorFlow/keras for R (https://tensorflow.rstudio.com/). A ReLU activation function was used for the hidden layer, and a sigmoid output activation function with cross-entropy loss since the task was a binary classification. The error function was the cross-entropy of a binary variable, and the optimization method employed was stochastic gradient descent. The validation data sets were used to tune the number of hidden units employed to fit the final model.

3 | RESULTS

3.1 | Integration of blastocyst and elongating conceptus data identify competent and noncompetent genes

The statistical contrast between PR blastocysts plus long conceptuses versus NP blastocysts plus short conceptuses resulted in 3295 DEG for the integrated Bl_Zo-Cc data set and 1742 for the Bl_SW-Cc data set. For the integrated Bl_Zo-Cc data set, 1302 genes were upregulated after the contrast, that is, genes that were more expressed in PR blastocysts than in NP blastocysts and even more expressed in long conceptuses than short conceptuses (Table S1A). On the other hand, 1993 genes were downregulated after the contrast, that is, more expressed in NP blastocysts than PR blastocysts and even more expressed in short conceptuses than long conceptuses. For the integrated Bl_SW-Cc data sets, the number of DEG was 742 upregulated genes and 1000 downregulated genes (Table S1B). Figure 1 shows the expression pattern for the up- and downregulated genes for each data set. The overlap between up- and downregulated genes for each data set resulted in 341 competent genes: genes increasing in expression from a PR blastocyst to a long conceptus and 699 noncompetent genes: genes increasing in expression from an NP blastocyst to a short conceptus (Figure 2; Table S1C).

Functional analysis using KEGG pathways showed that the 341 competent genes were enriched in Metabolic pathways (bta01100; 70 genes, FDR = 7.2×10^{-10}), Glycolysis/gluconeogenesis (bta00010; 9 genes, FDR = 0.003), and Glycerolipid metabolism (bta00561; 8 genes, FDR = 0.02), while the 669 noncompetent genes were enriched in Spliceosome (bta03040; 29 genes, FDR = 2.4×10^{-10}), Ribosome biogenesis in eukaryotes (bta03008; 20 genes, FDR = 2.4×10^{-08}), Cell cycle (bta04110; 16 genes, FDR = 0.004), RNA polymerase (bta03020, eight genes, FDR = 0.01), and RNA degradation (bta03018, 11 genes, FDR = 0.02). In total, 70 and 79 competent and noncompetent genes, respectively, were involved in these pathways, which were employed to define the biomarker genes (Table S1D).

3.2 | Selected biomarker genes discriminate between PR and NP samples

From the set of all genes in the pathways described above, there were eight genes (three competent and five noncompetent; Table 2) that, after application of sPLS-DA and linear discriminant analysis, best discriminated between PR and NP samples in the Bl_IVV data set. Figure 3 depicts the distribution of PR and NP samples in a PCA plot and hierarchical clustering analysis based on the expression of all the competent and noncompetent genes (1040 genes; panel A), genes in the enriched KEGG pathways (149 genes; panel B) and the eight biomarker genes (panel C). As expected, full discrimination between NP and PR samples was achieved using the expression of the eight biomarker genes.

Of the eight biomarker genes, the three competent genes were involved in functional terms related to carbohydrate metabolism glycolysis, oxidoreductase activity, and chondroitin and aldehyde biosynthesis (Table S2A), and the five noncompetent genes were enriched in terms involved in RNA metabolism, gene expression, spliceosome, and cell cycle (Table S2B).

3.3 | Expression of the biomarker genes predicts embryonic competence in the validation data sets

The rationale was that modeling with the expression signature of the eight discriminating genes between PR and NP IVV-produced blastocysts would predict embryonic competence in independent embryo transcriptomic data. Therefore, it was expected that predictions for each validation data set would indicate increased competence for PR IVT-produced blastocysts (Bl_IVT), embryos cultured under normal conditions (Bl_SubOp), long conceptuses (CcGr3), and long conceptuses developed in an HF heifer (Cc_HFSF), whereas "noncompetence" would be predicted for NP IVT-produced blastocysts (Bl_IVT), embryos cultured under suboptimal conditions (Bl_SubOp), short and grade 3 conceptuses (CcGr3), and conceptuses obtained from SF or HF heifers, but classified as short in the last case (Cc_HFSF).

Figure 4 shows the sample distribution in the PCA and hierarchical clustering plots for the integrated data sets

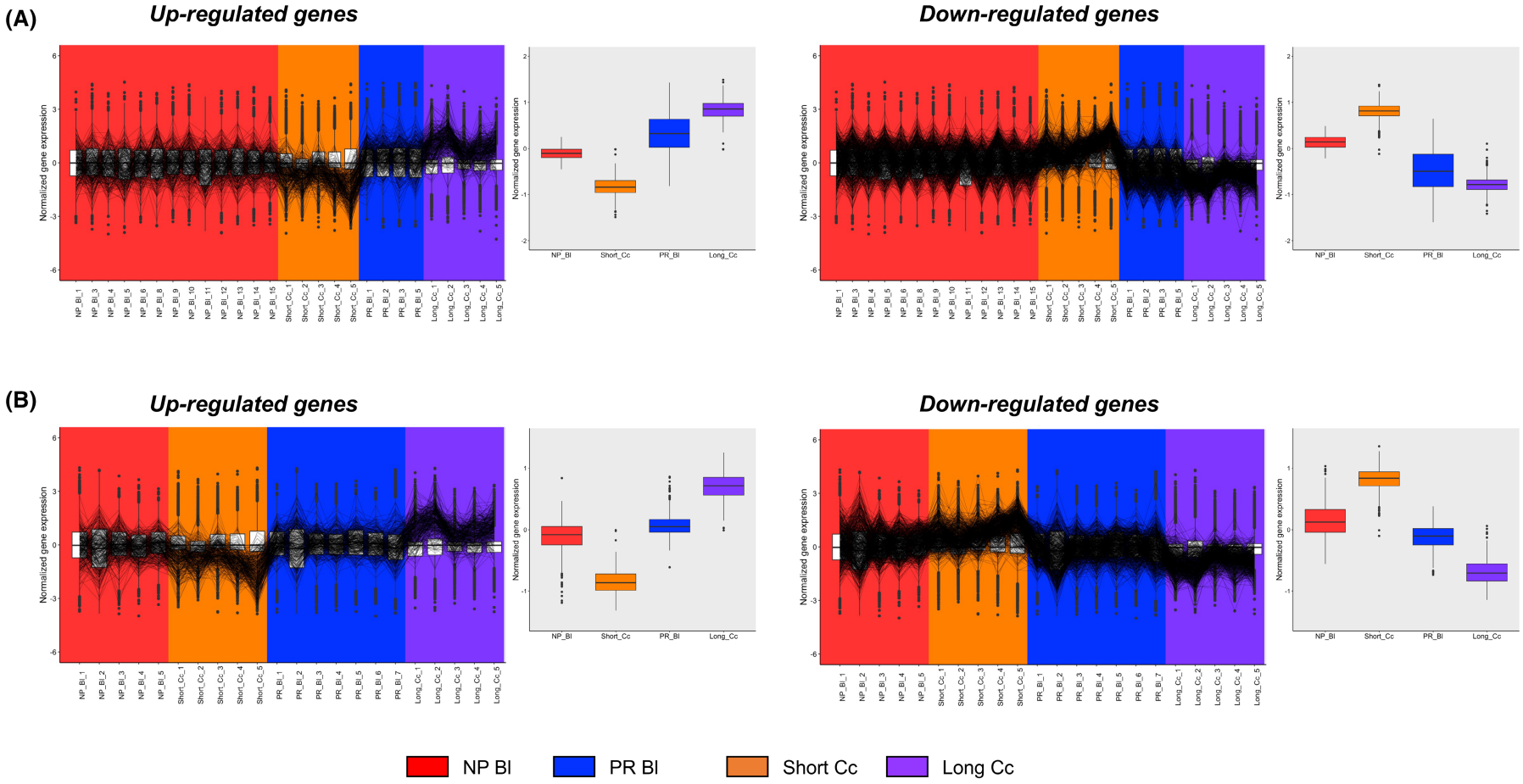

FIGURE 1 Patterns of gene expression related to embryonic fate. Expressions are plotted per sample (line plots) or the average per group (box plots) for upregulated or downregulated genes between long conceptuses (Long Cc) and in vivo-produced blastocysts that resulted in pregnancy (PR Bl) versus short conceptuses (Short Cc) and in vivo-produced blastocysts that did not result in pregnancy (NP Bl). Data from in vivo-produced blastocysts, obtained from two different studies,[12,16] were integrated with elongated conceptus data.[21] The results for the integrated data set between the Zolini study data[16] and the elongated conceptus data, or between the Salilew-Wondim study data[12] and the elongated conceptus data, are shown in Panels A and B, respectively.

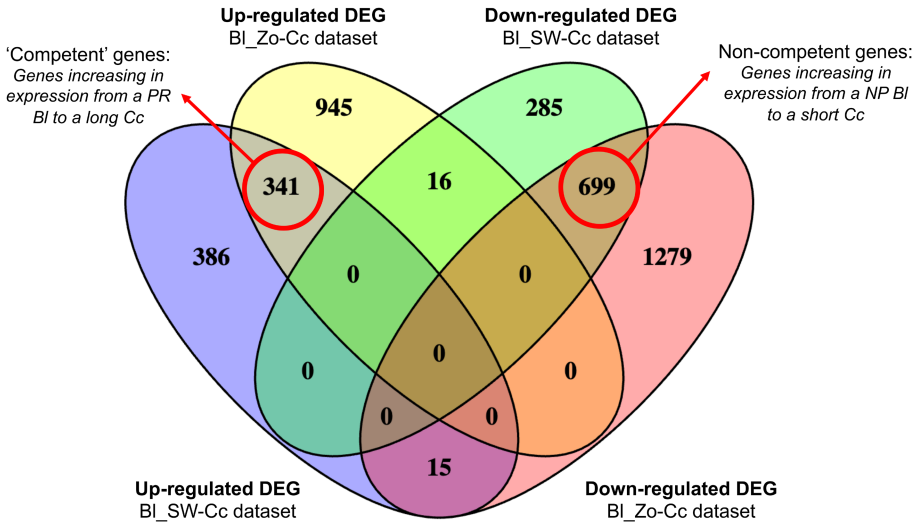

FIGURE 2 Comparison between differentially expressed genes (DEG). The Venn diagram shows the overlap between DEG that were upregulated or downregulated for each integrated data set, as explained in Figure 1.

TABLE 2 List of candidate biomarker genes selected by linear discrimination from the group of genes classified as competent or not according to the differential expression between blastocysts that resulted in pregnancy/long conceptuses and blastocysts that did not result in pregnancy/short conceptuses.

Ensembl ID	Entrez ID	Official symbol	Gene name	Chr	Classification
ENSBTAG00000003989	505642	*GSTO1*	Glutathione S-transferase omega 1	26	Competent
ENSBTAG00000007357	281690	*CHSY1*	Chondroitin sulfate synthase 1	21	Competent
ENSBTAG00000019782	281543	*TPI1*	Triosephosphate isomerase 1	5	Competent
ENSBTAG00000004077	286862	*YWHAG*	Tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein gamma	25	Noncompetent
ENSBTAG00000004943	281667	*CCNA2*	Cyclin A2	6	Noncompetent
ENSBTAG00000008578	613634	*LSM4*	LSM4 homolog, U6 small-nuclear RNA, and mRNA degradation associated	7	Noncompetent
ENSBTAG00000011046	515462	*CDK7*	Cyclin-dependent kinase 7	20	Noncompetent
ENSBTAG00000016023	515145	*EIF4A3*	Eukaryotic translation initiation factor 4A3	19	Noncompetent

between Bl_IVV and each validation data set, according to the expression of all the genes in common among the data sets, or only the biomarker genes, and the corresponding coefficient of the linear discriminant. Competent and noncompetent samples were better separated when the expression of the biomarker genes was considered rather than the expression of all genes. Accordingly, the accuracy of the predictions for each validation data set using a BLR or NN was higher than 85% (Table 3), suggesting that these genes play a key role in embryo survival.

The standardized expression for the competent and noncompetent genes in the competent and noncompetent embryos on each data set is depicted in Figure 5. Competent genes are more expressed in PR IVV-produced blastocysts (Bl_IVV), embryos cultured under normal conditions (CcGr3), and long conceptuses (CcGr3, and Cc_HFSF). However, there are no differences in the expression of these genes for PR and NP IVT-produced blastocysts (Bl_IVT). On the other hand, noncompetent genes are less expressed in the competent (or PR) embryos in all the data sets, including Bl_IVT.

4 | DISCUSSION

Given that most pregnancy losses occur during early gestation, numerous efforts have been made to identify the genes involved in embryonic competence and survival

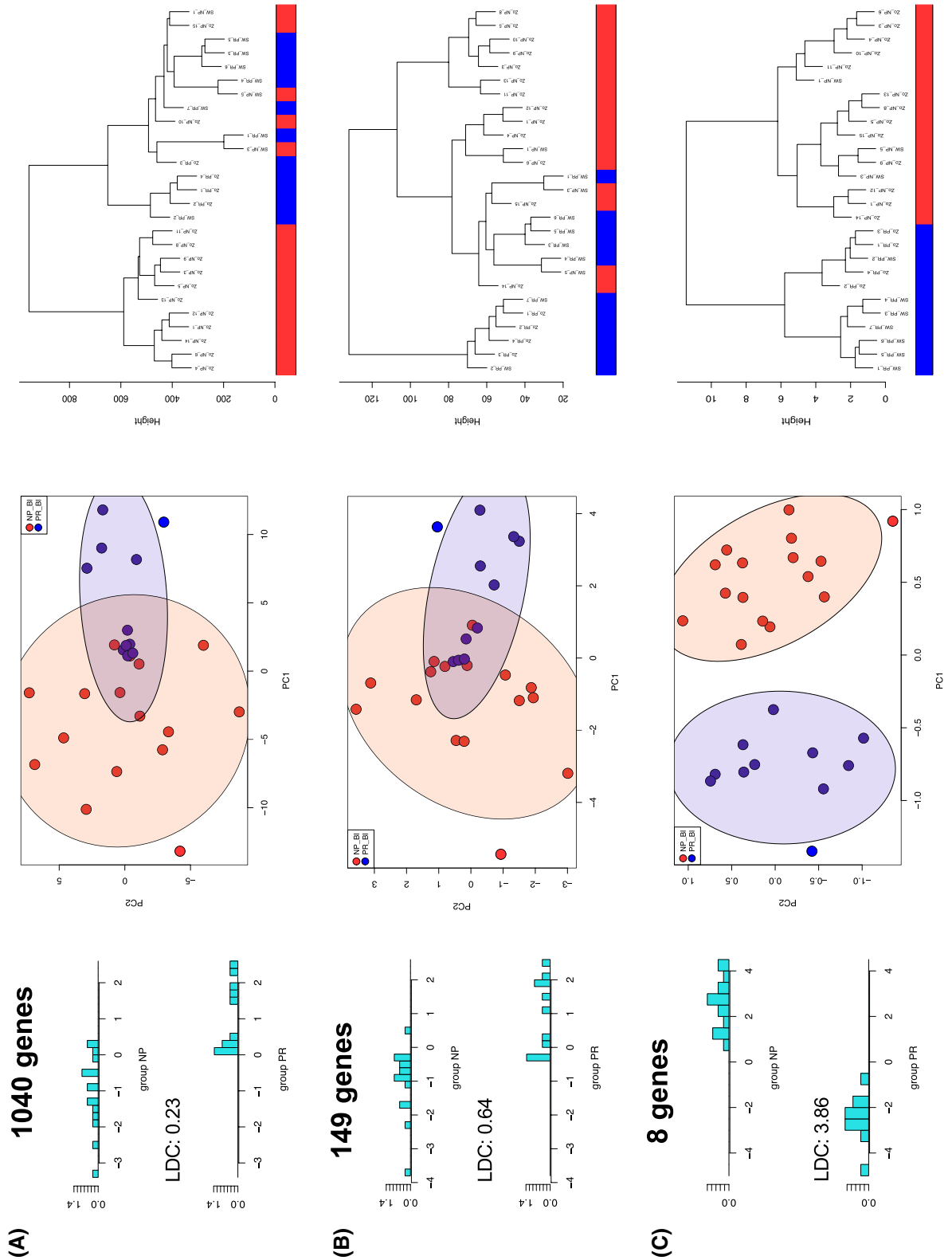

FIGURE 3 Discrimination of samples from blastocysts with divergent developmental capacities. The plots illustrate the results for linear discriminant analysis (left plot), principal component analysis (central plots), and hierarchical clustering (right plot) for the integrated data from in vivo-produced blastocysts. The distribution of the samples was evaluated according to the expression of (A) the 1040 overlapping genes, (B) the 149 genes involved in enriched KEGG pathways, and (C) the eight genes identified as biomarkers of embryonic competence. NP_BI, blastocysts that did not result in pregnancy after transfer; PR_BI, blastocysts that resulted in pregnancy after transfer.

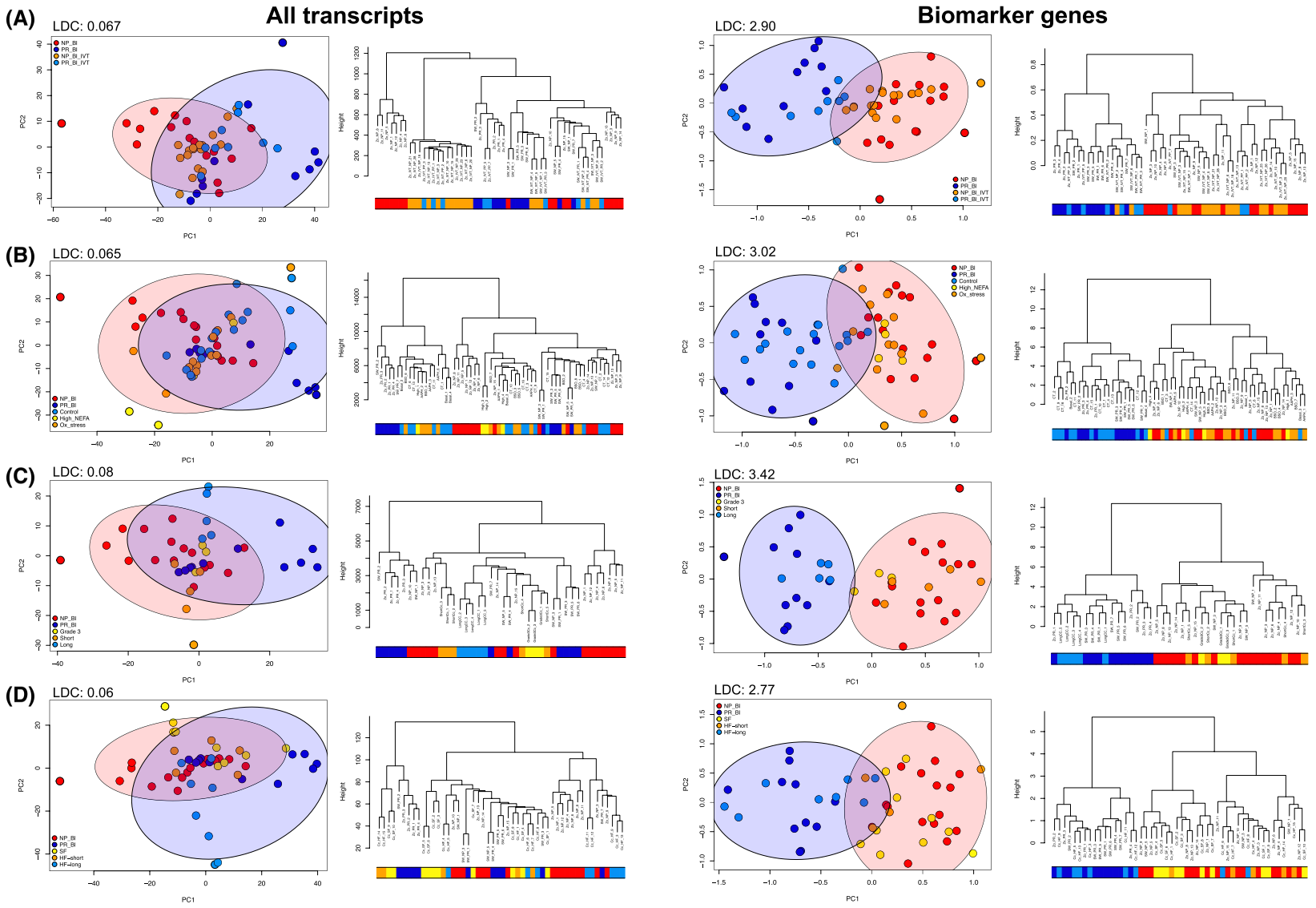

FIGURE 4 Discrimination of samples from blastocysts and elongated conceptuses with divergent developmental capacities. The principal component and hierarchical clustering plots show the distribution of samples according to all the transcripts (left two plots) or the eight biomarker genes (right two plots) for data from (A) in vitro-produced blastocyst that resulted in pregnancy (PR_Bl_IVT) or not (NP_Bl_IVT) integrated with data from in vivo-produced blastocyst that resulted in pregnancy (PR_Bl) or not (NP_Bl); (B) embryos cultured with high levels of nonesterified fatty acids (High_NEFA), pro-oxidative agents (Ox_stress), or normal conditions (Control) integrated with PR_Bl and NP_Bl data; (C) short, long, and grade 3 elongated conceptuses integrated with PR_Bl and NP_Bl data; and (D) elongated conceptus developed in subfertile (SF) or short or long conceptuses in high-fertile (HF) heifers integrated with PR_Bl and NP_Bl data. LDC, linear discriminant coefficient for the first component.

TABLE 3 Evaluation metrics correspond to the classifications of embryonic competence based on the expression of eight biomarker genes.

	Bl_IVT		Bl_SubOp		CcGr3		Cc_HFSF	
Classifier	**BLR**	**NN[33]**	**BLR**	**NN[39]**	**BLR**	**NN[14]**	**BLR**	**NN[30]**
Accuracy	96.15	96.15	85.29	88.24	100	100	86.96	100
Accuracy *p* value	0.0002	0.0002	1.93×10^{-05}	3.0810^{-06}	0.002	0.002	0.019	5.37×10^{-05}
Mcnemar *p* value	1	1	0.074	0.134	—	—	1	—
Sensitivity	100.00	100.00	70.59	76.47	100	100	75.00	100
Specificity	94.12	94.12	100.00	100.00	100	100	93.33	100
Pos Pred value	90.00	90.00	100.00	100.00	100	100	85.71	100
Neg Pred value	100.00	100.00	77.27	80.95	100	100	87.50	100
Misclassified	*One NP sample*		*Five controls*	*Four controls*			*Two long and one short HF*	

Note: The training data sets were data obtained from blastocysts produced in vivo and classified as pregnant (PR) or not (NP). The testing data sets were data from Bl_IVT: blastocysts produced in vitro and classified as PR or NP; Bl_SubOp: blastocysts cultured or not under suboptimal conditions; CcGr3 = long, short, and grade 3 day 15 conceptuses; and Cc_HFSF: long and short conceptuses developed in high-fertile (HF) or subfertile (SF) heifers. The classifiers were Bayesian logistic regression (BLR) and neural network (NN), with the number of hidden units shown in parenthesis.

after deposition in the recipient uterus. Here, we attempted to increase the knowledge in that area by using bioinformatics and machine-learning approaches to integrate data sets from different sources and transcriptomic platforms, identify putative biomarkers genes, and validate these genes using independent data sets. Embryo survival after ET depends not only on its intrinsic properties but also on the maternal reproductive tract environment (reviewed in[40]) and culture conditions if the embryo was produced IVT (reviewed in[41,42]). Therefore, the application of mathematical algorithms to biological data is challenging and subject to errors. The approach followed in the present study to identify potential critical genes for embryo survival attempted to follow a physiological scenario. First, data from IVV-produced blastocyst were integrated with data from IVV-produced day-15 conceptuses to avoid any potential confounding effects of IVT conditions.[22] This first step aimed to determine those genes that were more expressed in the PR than in the NP blastocyst and even more expressed in long conceptuses than short conceptuses and vice versa. The rationale was that genes involved in embryonic competence should continue increasing (or decreasing) in expression from the blastocyst to the elongated conceptus stage. Functional analysis revealed that "competent" genes (those that should increase in expression from blastocyst to elongated conceptus) were involved in energy metabolisms, such as glycolysis and gluconeogenesis. In contrast, "noncompetent" genes (those that should decrease in expression from blastocyst to elongated conceptus) were related to mRNA processing. These results are consistent with the literature: glucose uptake and utilization are critical for embryo survival and development[43,44] and increase from the blastocyst to the expanded blastocyst stage in the bovine.[45] Glycolysis/gluconeogenesis, oxidative phosphorylation, and carbon metabolism, among others, were also identified as enriched pathways by the IVV-produced blastocysts resulting in pregnancy in the study by Salilew-Wondim et al.[12] Downregulation of genes involved in mRNA processing in competent embryos might be explained through the "quiet embryo hypothesis" proposed by Leese[46]: viable embryos favor a metabolism that is "quiet" rather than "active" as they are required to expend less energy rectifying damage to the genome, transcriptome, and proteome. RNA and protein synthesis can account for about 12% and 20%, respectively, of the total energy consumption.[47] Indeed, consistent with this hypothesis, it has been shown that fresh and frozen human embryos with higher capacity for development had a lower rate of amino acid depletion, appearance, and turnover than arresting embryos.[48] Therefore, decreasing the expression of genes involved in certain steps of mRNA processing might impact specific biological functions and favor development beyond the blastocyst stage.

After determining the potential genes related to competence, the next step was to define those that would behave as biomarkers at the blastocyst stage. Narrowing the 70 competent and 79 noncompetent genes through bioinformatics approaches, we could identify eight potential biomarkers. The 3 competent biomarkers were glutathione S-transferase omega 1 (*GST1*), chondroitin sulfate synthase 1 (*CHSY1*), and triosephosphate isomerase 1 (*TPI1*), all of which have essential roles in cellular metabolism. *GST1* is a member of a multigene family of isoenzymes, which have been shown to modulate signal transduction pathways controlling cell proliferation and

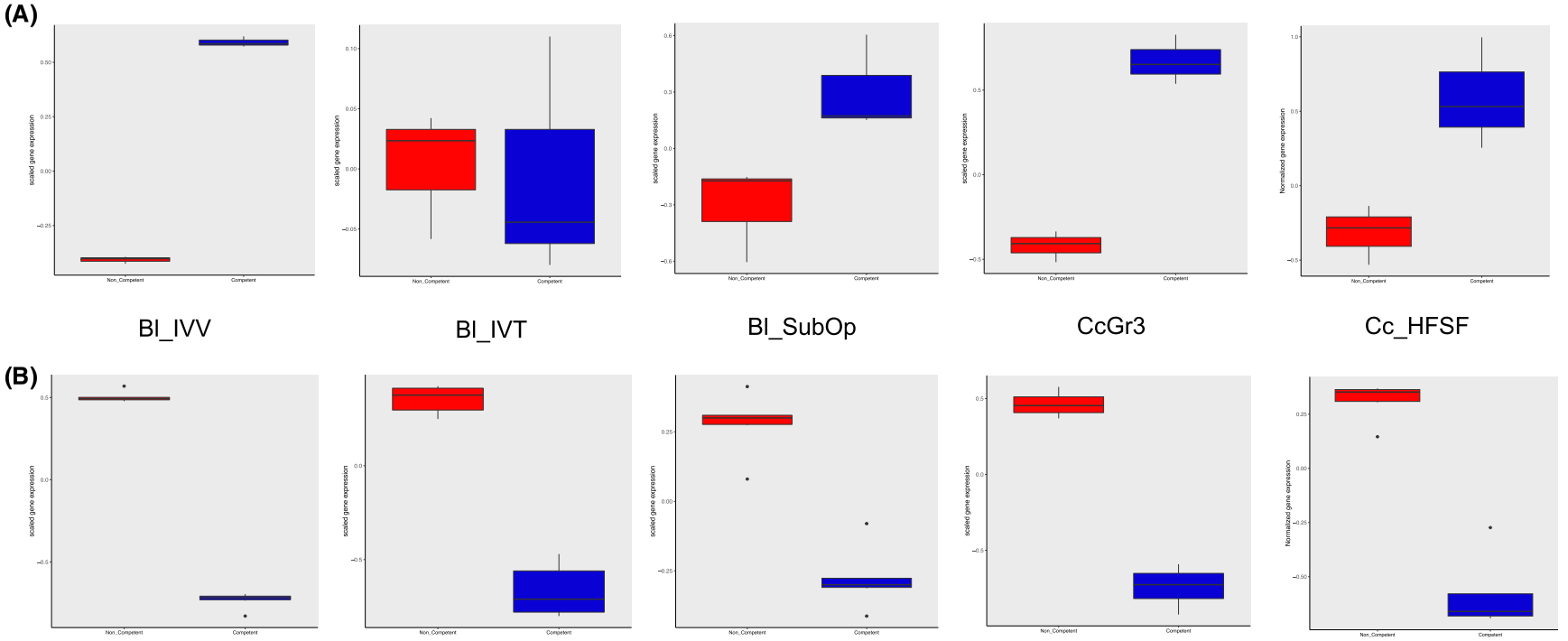

FIGURE 5 Expression of the biomarker genes. The box plots show the normalized expression for (A) the three "competent" biomarkers or (B) the five "noncompetent" biomarkers; in the competent (blue boxes) and noncompetent (red boxed) embryos from each data set. Bl_IVV: in vivo-produced blastocysts that resulted in pregnancy (considered competent) or not (considered noncompetent). Bl_IVT: in vitro-produced blastocysts that resulted in pregnancy (considered competent) or not (considered noncompetent). Bl_SubOp: blastocysts cultured under normal (considered competent) or suboptimal conditions (considered noncompetent). CcGr3: long conceptuses (considered competent), and short and grade 3 elongated conceptuses (considered noncompetent). Cc_HFSF, long conceptuses developed in high-fertile (HF) heifers (considered competent), and short conceptuses developed in high-fertile (HF) heifers or in subfertile (SF) heifers (considered noncompetent).

cell death (reviewed in[49]). The omega is an atypical cytosolic class, and it regulates genes involved in cellular stress response, steroid metabolism, transcription, and cytoskeleton organization.[50] Furthermore, inhibition or knockdown of *GST1* downregulated the Wnt inhibitor dickkopf 1,[50] a progestomedin that promotes conceptus elongation.[51] *CHSY1* synthesizes chondroitin sulfate, one of the major glycosaminoglycans of the extracellular matrix, regulating cell growth, differentiation, morphogenesis, and, thus, embryo development.[52] Finally, *TPI1* catalyzes the interconversion of dihydroxyacetone phosphate and glyceraldehyde 3-phosphate in glycolysis and gluconeogenesis pathways[53] and has been found to be one of the most abundant proteins in the cultured trophectoderm cells of 8-day cattle embryos.[54] Conversely, two of the five noncompetent biomarkers are key cell cycle regulators: cyclin A2 (*CCNA2*) and cyclin-dependent kinase 7 (*CDK7*). Cyclin A2 is required at two critical points during the mitotic cell cycle of somatic adult cells: G1- to S-phase transition and entry into mitosis.[55] However, *Ccna2*-null mutant murine embryos can develop normally from the four-cell to the postimplantation stage, indicating it is not essential for DNA replication in the early mouse embryo.[56] *CDK7* associates with two regulatory subunits: cyclin H, which activates it, and the RING finger protein Mat1 (ménage-à-trois 1) to form the trimeric *CDK7*-cyclin H-Mat1 complex.[57,58] In mouse embryos, cyclin H is important in the normal expansion of the intracellular mass (ICM) from the blastocyst stage, although reduced levels did not impair ICM formation. Thus, it is unknown whether cyclin H is required for development to the blastocyst stage.[59] Another relevant noncompetent gene is eukaryotic translation initiation factor 4A3 (*EIF4A3*), an element of the exon junction complex, which contains proteins playing critical roles in postsplicing events such as mRNA export, cytoplasmic localization, and nonsense-mediated decay.[60] In addition, it has been shown in zebrafish embryos that *EIF4A3* can regulate the Wnt/β-catenin-signaling pathway,[61] which is also involved in blastocyst and ICM formation in the bovine embryo.[62] This gene was also significantly downregulated in the bovine blastocyst compared with the morula stage.[63] Undoubtedly, the product of these noncompetent genes, including *YWHAG* (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma), a highly conserved protein that mediates signal transduction by binding to phosphoserine-containing proteins, and *LSM4* (*LSM4* homolog), U6 small-nuclear RNA, and mRNA degradation associated, an LSM protein that attaches to the 3′-terminal oligo(U) tract of U6 snRNA and plays a role in pre-mRNA splicing are critical components of the cellular machinery for cell cycle progression. Splicing errors frequently occur in the preimplantation embryo and are an important component of attenuated responses of the early embryo to DNA damage.[64] While some studies indicate that expression of the corresponding genes might not be necessary for the blastocyst,[56,59,63] our integrated data analysis suggests that their expression should be downregulated for embryo survival.

We demonstrated here that the expression of the eight biomarker genes strongly discriminated samples from IVV PR and NP blastocysts obtained from two different experiments (Bl_Zo and Bl_SW) and can be employed to construct a predictive model of embryo survival. The expressions of these biomarkers in samples from external independent data sets were used to validate this model. Predictions in the Bl_IVT data set were consistent between both BLR and NN models. All PR samples were correctly classified as such, while the same NP blastocyst was misclassified as PR by both models. Samples obtained from IVT-produced blastocysts that resulted in pregnancy were expected to be classified as such. However, the single NP blastocyst classified as PR could have been a competent embryo that did not result in pregnancy because of the sampling technique or other factors. An interesting finding is that the expression of competent genes between the PR and NP IVT-produced blastocysts did not differ, so the prediction was based on the expression of the noncompetent genes. Searching for candidate biomarker genes was done with data from IVV embryos and may have differed if data from IVT embryos were employed instead. Also, both the intrinsic quality of the IVT-produced embryo and the quality of the recipient animal determine whether an embryo leads to pregnancy after ET. Therefore, even an embryo with strong expression of the competent genes does not necessarily lead to pregnancy if the recipient animal is unsuitable. High expression of noncompetent genes, on the other hand, does not lead to pregnancy in any case, regardless of the quality of the recipient animals. In addition, the lack of upregulation of the competent genes might be one of the reasons for which the transfer of IVT-produced embryos results in lower pregnancy rates than IVV-produced counterparts.[8] Support for this hypothesis comes from the fact that the addition of glycosaminoglycans (synthesized in part by *CHSY1*, see above) to the culture medium improved the development of IVT-produced embryos in cattle[65] and pigs.[66] Of course, further experiments would be needed to confirm this theory. Although embryos in the Bl_SubOp data set were also produced IVT, they were treated with detrimental agents, and thus, expression of these genes would be downregulated in these embryos compared to the controls. Regarding this data set (Bl_SubOp), both BLR and NN models predicted five and four control samples as noncompetent or NP. As mentioned earlier, samples from this data set were obtained from blastocysts exposed to oxidative stress during culture from day 3 to day 7[36] or high NEFA levels during 6.5 days of embryo culture.[33] Gene expression was evaluated in those

embryos that reached the blastocyst stage. Although these embryos were not transferred to a recipient, it is assumed that their ability to establish and maintain a pregnancy would be compromised. Interestingly, both BLR and NN models classified all the NEFA/BSO/AAPH-treated embryos as noncompetent while five (BLR) and four (NN) control embryos (three in common) were classified as such, suggesting that these embryos would not be able to establish a pregnancy. The expression of competent and noncompetent genes in control or treated embryos followed a similar pattern than for Bl_IVV, CcGr3, and Cc_HFSF. Also, the authors of these data sets found that exposure to the pro-oxidant BSO agent downregulated *TPI1* while upregulated *EIF4A2* compared with controls.[36]

Even when the model was constructed with the Bl_IVV data set, that is, data from the IVV-produced blastocysts, predictions in the elongated conceptuses were highly accurate. For the CcGr3 data set, both BLR and NN models predicted, as expected, short and long conceptuses as noncompetent, or NP, and competent, or PR, respectively. In addition, elongated conceptuses generated from grade 3 blastocysts were predicted as noncompetent. The authors of this data set[21] did not find DEG between conceptuses derived from grade 1 or grade 3 day-7 embryos probably because those grade 3 conceptuses that continued until day 15 of gestation were the most developmentally competent, as a much lower proportion of elongated conceptuses derived from grade 3 day-7 embryos were recovered at day 15 (18.3%), compared with those derived from grade 1 embryos (61.3%). Nevertheless, these conceptuses were classified as noncompetent according to the expression of the biomarker genes, suggesting that they might have experienced pregnancy loss subsequently. Finally, concerning the Cc_HFSF data set, all the conceptuses developed in SF heifers were predicted as noncompetent by both BLR and NN models, highlighting the influence of the endometrium on the transcriptomic profile of the conceptus. Indeed, given the sensing ability of the endometrium, dysregulation of the conceptus–endometrial interactions can lead to pregnancy loss after conceptus elongation.[67,68] On the other hand, all the long and short conceptuses collected from HF heifers were correctly predicted as competent or not, respectively, by the NN model, whereas the BLR model misclassified only two long and one short conceptus. The authors of the original study[37] reported 1287 DEG between HF and SF conceptuses but only 18 DEG between short- and long-HF conceptuses, highlighting the effect of the uterine fertility status. Nevertheless, our results suggest that the length of elongated conceptus affects its likelihood of survival, even when it develops in an optimal environment, which underlines the importance of the intrinsic characteristics of the embryo.

In conclusion, this study integrated seven bovine embryonic data sets through bioinformatic and machine-learning approaches to identify a small set of genes capable of predicting embryonic competence. Although data were obtained from seven independent experiments and generated through different sources, the results were consistent and biologically relevant. Furthermore, they can help to understand factors involved in pregnancy establishment after embryo transfer and to identify related noninvasively quantifiable molecules, such as specific microRNAs and metabolites, to develop a robust and practical method for embryo evaluation.

AUTHOR CONTRIBUTIONS

Maria Belen Rabaglino: Analytical design; Data collection, integration, and analysis; Results visualization and interpretation; Writing of the original draft: Dessie Salilew-Wondim, Adriana Zolini, Dawit Tesfaye, Michael Hoelker, and Peter J. Hansen: Data generation; Maria Belen Rabaglino, Peter J. Hansen, and Pat Lonergan: Conceptualization; Peter J. Hansen and Pat Lonergan: Correction of the first draft. All authors: Correction and approval of the final version.

ACKNOWLEDGMENTS

MBL was supported by an H2020-MSCA-Individual Fellowship (Proposal 101021311). Open access funding provided by IReL.

DISCLOSURES

The authors have no competing interests to declare.

DATA AVAILABILITY STATEMENT

Data were derived from the following data sets deposited in the Gene Expression Omnibus repository: GSE131178, GSE75750, GSE13095, GSE42281, GSE83767, and GSE10789.

ORCID

Maria Belen Rabaglino https://orcid.org/0000-0002-0099-045X
Dessie Salilew-Wondim https://orcid.org/0000-0002-8756-0865
Adriana Zolini https://orcid.org/0000-0003-4528-584X
Dawit Tesfaye https://orcid.org/0000-0001-8166-0606
Michael Hoelker https://orcid.org/0000-0002-7878-6491
Pat Lonergan https://orcid.org/0000-0001-5598-5044
Peter J. Hansen https://orcid.org/0000-0003-3061-9333

REFERENCES

1. Foote RH. Development of reproductive biotechnologies in domestic animals from artificial insemination to cloning: a perspective. *Cloning*. 1999;1:133-142.

2. Moore S, Hasler J. A 100-year review: reproductive technologies in dairy science. *J Dairy Sci*. 2017;100:10314-10331.
3. Viana J. 2020 statistics of embryo production and transfer in domestic farm animals. *Embryo Technol Newsl*. 2021;39:24-37.
4. Hansen PJ. The incompletely fulfilled promise of embryo transfer in cattle-why aren't pregnancy rates greater and what can we do about it? *J Anim Sci*. 2020;98:1-20.
5. Leroy JL, Vanholder T, Mateusen B, et al. Non-esterified fatty acids in follicular fluid of dairy cows and their effect on developmental capacity of bovine oocytes in vitro. *Reproduction*. 2005;130:485-495.
6. Hansen PJ. Effects of heat stress on mammalian reproduction. *Philos Trans R Soc Lond B Biol Sci*. 2009;364:3341-3350.
7. Diskin M, Morris D. Embryonic and early foetal losses in cattle and other ruminants. *Reprod Domest Anim*. 2008;43(Suppl 2):260-267.
8. Ealy A, Wooldridge L, McCoski S. BOARD INVITED REVIEW: post-transfer consequences of in vitro-produced embryos in cattle. *J Anim Sci*. 2019;97:2555-2568.
9. Thatcher WW, Guzeloglu A, Mattos R, Binelli M, Hansen TR, Pru JK. Uterine-conceptus interactions and reproductive failure in cattle. *Theriogenology*. 2001;56:1435-1450.
10. Clemente M, Lopez-Vidriero I, O'Gaora P, et al. Transcriptome changes at the initiation of elongation in the bovine conceptus. *Biol Reprod*. 2011;85:285-295.
11. El-Sayed A, Hoelker M, Rings F, et al. Large-scale transcriptional analysis of bovine embryo biopsies in relation to pregnancy success after transfer to recipients. *Physiol Genomics*. 2006;28:84-96.
12. Salilew-Wondim D, Tesfaye D, Rings F, et al. The global gene expression outline of the bovine blastocyst: reflector of environmental conditions and predictor of developmental capacity. *BMC Genomics*. 2021;22:408.
13. Zolini A, Block J, Rabaglino M, et al. Molecular fingerprint of female bovine embryos produced in vitro with high competence to establish and maintain pregnancy. *Biol Reprod*. 2020;102:292-305.
14. Salilew-Wondim D, Hölker M, Rings F, et al. Bovine pretransfer endometrium and embryo transcriptome fingerprints as predictors of pregnancy success after embryo transfer. *Physiol Genomics*. 2010;42:201-218.
15. Ghanem N, Salilew-Wondim D, Gad A, et al. Bovine blastocysts with developmental competence to term share similar expression of developmentally important genes although derived from different culture environments. *Reproduction*. 2011;142:551-564.
16. Zolini AM, Block J, Rabaglino MB, et al. Genes associated with survival of female bovine blastocysts produced in vivo. *Cell Tissue Res*. 2020;382:665-678.
17. Rizos D, Scully S, Kelly AK, et al. Effects of human chorionic gonadotrophin administration on day 5 after oestrus on corpus luteum characteristics, circulating progesterone and conceptus elongation in cattle. *Reprod Fertil Dev*. 2012;24:472-481.
18. Carter F, Forde N, Duffy P, et al. Effect of increasing progesterone concentration from day 3 of pregnancy on subsequent embryo survival and development in beef heifers. *Reprod Fertil Dev*. 2008;20:368-375.
19. Clemente M, de La Fuente J, Fair T, et al. Progesterone and conceptus elongation in cattle: a direct effect on the embryo or an indirect effect via the endometrium? *Reproduction*. 2009;138:507-517.
20. Sánchez J, Mathew D, Behura S, et al. Bovine endometrium responds differentially to age-matched short and long conceptuses. *Biol Reprod*. 2019;101:26-39.
21. Barnwell C, Farin P, Ashwell C, Farmer W, Galphin S, Farin C. Differences in mRNA populations of short and long bovine conceptuses on day 15 of gestation. *Mol Reprod Dev*. 2016;83:424-441.
22. Rabaglino M, O'Doherty A, Bojsen-Møller Secher J, et al. Application of multi-omics data integration and machine learning approaches to identify epigenetic and transcriptomic differences between in vitro and in vivo produced bovine embryos. *PLoS One*. 2021;16:e0252096.
23. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2020.
24. Huber W, von Heydebreck A, Sueltmann H, Poustka A, Vingron M. Parameter estimation for the calibration and variance stabilization of microarray data. *Stat Appl Genet Mol Biol*. 2003;2:1-22.
25. Love M, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
26. Ritchie M, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
27. Wu, J., and Irizarry, R. (2017) gcrma: Background Adjustment Using Sequence Information. R package version 2.52.0.
28. Leek, J., Johnson, W., Parker, H., Fertig, E., Jaffe, A., Zhang, Y., Storey, J. D., and Torres, L. (2021) sva: surrogate variable analysis. R package version 3.42.0.
29. Huang da W, Sherman B, Lempicki R. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
30. Lê Cao K, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinfor*. 2011;12:253.
31. Venables W, Ripley B. *Modern Applied Statistics with S*. Springer; 2002.
32. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 2020;36:2628-2629.
33. Desmet KL, Van Hoeck V, Gagne D, et al. Exposure of bovine oocytes and embryos to elevated non-esterified fatty acid concentrations: integration of epigenetic and transcriptomic signatures in resultant blastocysts. *BMC Genomics*. 2016;17:1004.
34. Leroy JL, Valckx SD, Jordaens L, et al. Nutrition and maternal metabolic health in relation to oocyte and embryo quality: critical views on what we learned from the dairy cow model. *Reprod Fertil Dev*. 2015;27:693-703.
35. Leroy JL, Van Soom A, Opsomer G, Bols PE. The consequences of metabolic changes in high-yielding dairy cows on oocyte and embryo quality. *Animal*. 2008;2:1120-1127.
36. Cagnone GL, Sirard MA. Transcriptomic signature to oxidative stress exposure at the time of embryonic genome activation in bovine blastocysts. *Mol Reprod Dev*. 2013;80:297-314.
37. Moraes JGN, Behura SK, Geary TW, Hansen PJ, Neibergs HL, Spencer TE. Uterine influences on conceptus development in fertility-classified animals. *Proc Natl Acad Sci USA*. 2018;115:E1749-E1758.

38. Hornung R, Causeur D, Bernau C, Boulesteix A. Improving cross-study prediction through addon batch effect adjustment or addon normalization. *Bioinformatics*. 2017;33:397-404.
39. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28:1-26.
40. Lonergan P, Forde N, Spencer T. Role of progesterone in embryo development in cattle. *Reprod Fertil Dev*. 2016;28:66-74.
41. Lonergan P, Fair T. In vitro-produced bovine embryos: dealing with the warts. *Theriogenology*. 2008;69:17-22.
42. Sirard M. The influence of in vitro fertilization and embryo culture on the embryo epigenetic constituents and the possible consequences in the bovine model. *J Dev Orig Health Dis*. 2017;8:411-417.
43. Riley JK, Moley KH. Glucose utilization and the PI3-K pathway: mechanisms for cell survival in preimplantation embryos. *Reproduction*. 2006;131:823-835.
44. Leese HJ. History of oocyte and embryo metabolism. *Reprod Fertil Dev*. 2015;27:567-571.
45. Obeidat Y, Catandi G, Carnevale E, et al. A multi-sensor system for measuring bovine embryo metabolism. *Biosens Bioelectron*. 2019;126:615-623.
46. Leese HJ. Quiet please, do not disturb: a hypothesis of embryo metabolism and viability. *Bioessays*. 2002;24:845-849.
47. Buttgereit F, Brand M. A hierarchy of ATP-consuming processes in mammalian cells. *Biochem J*. 1995;312:163-167.
48. Stokes PJ, Hawkhead JA, Fawthrop RK, et al. Metabolism of human embryos following cryopreservation: implications for the safety and selection of embryos for transfer in clinical IVF. *Hum Reprod*. 2007;22:829-835.
49. Laborde E. Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death. *Cell Death Differ*. 2010;17:1373-1380.
50. Ramkumar K, Samanta S, Kyani A, et al. Mechanistic evaluation and transcriptional signature of a glutathione S-transferase omega 1 inhibitor. *Nat Commun*. 2016;7:13084.
51. Tríbulo P, Rabaglino M, Bo M, et al. Dickkopf-related protein 1 is a progestomedin acting on the bovine embryo during the morula-to-blastocyst transition to program trophoblast elongation. *Sci Rep*. 2019;9:11816.
52. Esko JD, Kimata K, Lindahl U. Proteoglycans and sulfated glycosaminoglycans. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor Laboratory Press; 2009.
53. Daar IO, Artymiuk PJ, Phillips DC, Maquat LE. Human triosephosphate isomerase deficiency: a single amino acid substitution results in a thermolabile enzyme. *Proc Natl Acad Sci USA*. 1986;83:7903-7907.
54. Talbot NC, Powell AM, Caperna TJ, Garrett WM. Proteomic analysis of the major cellular proteins of bovine trophectoderm cell lines derived from IVP, parthenogenetic and nuclear transfer embryos: reduced expression of annexins I and II in nuclear transfer-derived cell lines. *Anim Reprod Sci*. 2010;120:187-202.
55. Pagano M, Pepperkok R, Verde F, Ansorge W, Draetta G. Cyclin A is required at two points in the human cell cycle. *EMBO J*. 1992;11:961-971.
56. Winston N, Bourgain-Guglielmetti F, Ciemerych MA, et al. Early development of mouse embryos null mutant for the cyclin A2 gene occurs in the absence of maternally derived cyclin A2 gene products. *Dev Biol*. 2000;223:139-153.
57. Fisher RP, Morgan DO. A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell*. 1994;78:713-724.
58. Tassan JP, Jaquenoud M, Fry AM, Frutiger S, Hughes GJ, Nigg EA. In vitro assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein. *EMBO J*. 1995;14:5608-5617.
59. Patel SA, Simon MC. Functional analysis of the Cdk7.Cyclin H.Mat1 complex in mouse embryonic stem cells and embryos. *J Biol Chem*. 2010;285:15587-15598.
60. Chan CC, Dostie J, Diem MD, et al. eIF4A3 is a novel component of the exon junction complex. *RNA*. 2004;10:200-209.
61. Wang B, Rong X, Zhou Y, et al. Eukaryotic initiation factor 4A3 inhibits Wnt/beta-catenin signaling and regulates axis formation in zebrafish embryos. *Development*. 2021;148:dev198101.
62. Tribulo P, Leão B, Lehloenya K, Mingoti G, Hansen P. Consequences of endogenous and exogenous WNT signaling for development of the preimplantation bovine embryo. *Biol Reprod*. 2017;96:1129-1141.
63. Yu B, van Tol HTA, Stout TAE, Roelen BAJ. Reverse transcription priming methods affect normalisation choices for gene expression levels in oocytes and early embryos. *Mol Hum Reprod*. 2021;27:1-12.
64. Wyatt CDR, Pernaute B, Gohr A, et al. A developmentally programmed splicing failure contributes to DNA damage response attenuation during mammalian zygotic genome activation. *Sci Adv*. 2022;8:eabn4935.
65. Jang G, Lee BC, Kang SK, Hwang WS. Effect of glycosaminoglycans on the preimplantation development of embryos derived from in vitro fertilization and somatic cell nuclear transfer. *Reprod Fertil Dev*. 2003;15:179-185.
66. Kano K, Miyano T, Kato S. Effects of glycosaminoglycans on the development of in vitro-matured and -fertilized porcine oocytes to the blastocyst stage in vitro. *Biol Reprod*. 1998;58:1226-1232.
67. Forde N, Carter F, Spencer T, et al. Conceptus-induced changes in the endometrial transcriptome: how soon does the cow know she is pregnant? *Biol Reprod*. 2011;85:144-156.
68. Bauersachs S, Ulbrich S, Zakhartchenko V, et al. The endometrium responds differently to cloned versus fertilized embryos. *Proc Natl Acad Sci USA*. 2009;106:5681-5686.

SUPPORTING INFORMATION

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

How to cite this article: Rabaglino MB, Salilew-Wondim D, Zolini A, et al. Machine-learning methods applied to integrated transcriptomic data from bovine blastocysts and elongating conceptuses to identify genes predictive of embryonic competence. *The FASEB Journal*. 2023;37:e22809. doi:10.1096/fj.202201977R