Loop-mediated isothermal amplification (LAMP) and machine learning application for early pregnancy detection using bovine vaginal mucosal membrane

Hiroki Kunii[1], Tomoaki Kubo[2], Natsuki Asaoka[1], Ahmed Z. Balboula[3], Yu Hamaguchi[4], Tomoya Shimasaki[1], Hanako Bai[1], Manabu Kawahara[1], Hisato Kobayashi[4], Hidehiko Ogawa[5], Masashi Takahashi[6]

1. Graduate School of Agriculture, Hokkaido University, Sapporo, Hokkaido, 060-8589,
2. Dairy Research Center, Hokkaido Research Organization, Nakashibetsu, Hokkaido 086-1135, Japan
3. Division of Animal Sciences, University of Missouri Columbia, MO 65211, USA
4. NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo 156-8502, Japan
5. Department of Bioscience, Tokyo University of Agriculture, Tokyo 156-8502, Japan
6. Graduate School of Global Food Resources/Global Center for Food, Land and Water Resources, Hokkaido University, Sapporo, Hokkaido, 060-8589, Japan

Corresponding author:
Masashi Takahashi, e-mail address: mmasashi@anim.agr.hokudai.ac.jp

Present address
Yu Hamaguchi: Department of Electrical Engineering and Bioscience, Waseda University, Tokyo 169-8555, Japan
Hisato Kobayashi: Department of Embryology, Nara Medical University, Kashihara, Nara, 634-8521, Japan

Abstract

An early and accurate pregnancy diagnosis method is required to improve the reproductive performance of cows. Here we developed an easy pregnancy detection method using vaginal mucosal membrane (VMM) with application of Reverse Transcription-Loop-mediated Isothermal Amplification (RT-LAMP) and machine learning. Cows underwent artificial insemination (AI) on day 0, followed by VMM-collection on day 17-18, and pregnancy diagnosis by ultrasonography on day 30. By RNA sequencing of VMM samples, three candidate genes for pregnancy markers (*ISG15* and *IFIT1*: up-regulated, *MUC16*: down-regulated) were selected. Using these genes, we performed RT-LAMP and calculated the rise-up time (RUT), the first-time absorbance exceeded 0.05 in the reaction. We next determined the cutoff value and calculated accuracy, sensitivity, specificity, positive prediction value (PPV), and negative prediction value (NPV) for each marker evaluation. The *IFIT1* scored the best performance at 92.5% sensitivity, but specificity was 77.5%, suggesting that it is difficult to eliminate false positives. We then developed a machine learning model trained with RUT of each marker combination to predict pregnancy. The model created with the RUT of *IFIT1* and *MUC16* combination showed high specificity (86.7%) and sensitivity (93.3%), which were higher compared to *IFIT1* alone. In conclusion, using VMM with RT-LAMP and machine learning algorithm can be used for early pregnancy detection before the return of first estrus.

Keywords: Cow; Early pregnancy detection; Loop-mediated Isothermal Amplification(LAMP); Machine learning; Vaginal mucosa.

Introduction

Early pregnancy-detection is required to improve reproductive performance in cows. One of the most widely used methods for pregnancy diagnosis is a transrectal diagnosis that projects the embryo or fetus. Despite the high accuracy of this method, it can not only be performed until 30 days after artificial insemination (AI). Early detection and re-insemination of non-pregnant cows will increase the chances of pregnancy resulting in improved reproductive performance of the dairy herd.

Several studies have been conducted to detect pregnancy in the earlier stages of gestation. Pregnancy-associated glycoproteins were used as pregnancy markers in milk or plasma of cows around 30 days after AI [1], as well as progesterone concentration in milk [2]. The ELISA kit using these markers is commercially available and optimized to be a simple method for pregnancy diagnosis. These methods could be useful, but still not in time for pregnancy detection earlier than first return to estrus after AI.

In ruminants, the elongating conceptus secretes interferon-tau (IFNT) during pre-implantation period [3, 4] and induces high expression of Interferon stimulated genes (ISGs) in uterine endometrium [5]. Thus, the endometrial biopsy can be one of the reliable methods for early pregnancy detection; however, it is highly invasive to the embryo and/or uterine tissues.

On the other hand, using ISGs in peripheral blood leukocytes has been expected as an early pregnancy detection method. The expression of ISGs was reported as candidates for pregnancy detection on d 18 to 21 in dairy cows [6] and Japanese-Black cattle [7]. However, a reliable judgment can be done only in heifers; cows have low sensitivity [8]. Also, for a reliable non-pregnancy detection in dairy cows, serial blood sampling from d 17 to 25 was required [9]. Thus, pregnancy determination using ISGs could still be improved, especially to achieve high sensitivity and specificity at d 18, regardless of parity.

We hypothesized that pregnancy also affects extra-uterine tissues by direct connection to the uterus, and first reported high expression of representative ISGs, such as *ISG15* in mucosal membrane from non-invasively collected cervical lumen and the deep vaginal wall on day (d) 17-18 of pregnancy [10]. Thus, using samples collected from the vaginal wall could be a novel approach to establish non-invasive and easy pregnancy detection. However, early pregnancy detection from vaginal mucosa has never been reported.

To establish quick and early pregnancy detection from vaginal mucosa, we focused on the minimalization of working procedure, especially the process of gene amplification. The Loop-mediated isothermal amplification (LAMP) amplifies DNA rapidly under the condition of constant temperature using Bst DNA polymerase [11], is recently used as an alternative way for gene expression analysis. LAMP technology can be applied to the direct amplification of RNA samples by adding Avian Myeloblastosis Virus reverse transcriptase, called reverse transcription-LAMP (RT-LAMP).

To predict pregnancy based on RT-LAMP reactions, it is necessary to perform RT-LAMP reaction on a large scale and establish a model that can accurately determine pregnancy. In this perspective, applying machine learning with the output of LAMP assay can be considered a suitable approach. The machine learning approach has been used to address stockbreeding industry issues, such as estrus detection [12] and breeding management [13]. However, no machine learning models for early pregnancy prediction has been created.

The aim of this study is to develop a simple and prompt method of pregnancy detection with high sensitivity and specificity, before first return to estrus. We first explored pregnancy marker genes in the deep vaginal wall on d 17-18 After AI, using comprehensive gene expression data analysis. Then, we analyzed the selected marker genes expression by the LAMP method. Finally, we created a machine learning model

with high accuracy based on the LAMP reaction, proposing a fast and simple way to detect early pregnancy.

Materials and methods

Animals and sampling.

The animals used in this study were multiparous Holstein Friesian cows, which belong to Research experimental station of Hokkaido University (Sapporo, Hokkaido, Japan) and Dairy Research Center, Hokkaido Research Organization (Nakashibetsu, Hokkaido, Japan). All procedures were conducted in accordance with Hokkaido University guidelines regarding the care and use of animals (Approval No. 16-0019) and the Dairy Research Center, Hokkaido Research Organization (Approval No. 2019314-3, 2020304-3). Cows were subjected to AI, and sampling was conducted on d 17 or d 18 after AI (d 0: day of AI). On d 30, pregnancy was assessed by ultrasonography. Based on the diagnosis, the collected samples were classified into non-pregnant and pregnant (n = 40, each group).

Vaginal mucosal membrane collection for RNA sequencing.

On d18 after AI, vaginal mucosal membrane (VMM) sampling was conducted from non-pregnant and pregnant cows (n = 3, each) as described previously [10]. Collected samples were stored at –80 °C and used for RNA sequencing.

VMM collection for early pregnancy evaluation.

To establish simple and quick sampling process for a practical pregnancy evaluation, cotton swab (Handle length: 75mm, head width: 5 mm, P3S-100, Japan Cotton Buds. Industry Ltd, Tokyo, Japan) was used for sampling. After cutting the shaft

to 5 cm, the swab was set to a pick-up tool attached with an endoscope (Fig. 1A). The pick-up tool with a cotton swab was inserted deep into the vagina by using a vaginal speculum. The VMM of deep vaginal wall close to the external os was gently scraped up by monitoring the sampling position while illuminated with LED scope light (Fig. 1B). The collected VMM samples were stored at –80°C until further use.

RNA extraction.

For RT-LAMP, total RNA was extracted using Easy RNA Extraction Kit (Kaneka, Tokyo, Japan). For effective RNA extraction from VMM clinging to swab, the manufacture's procedure was slightly modified. In brief, a cotton swab head with VMM was put to a 1.5 mL tube filled with 100 μL of CelLytic™M Cell Lysis Reagent (Sigma-Aldrich, St. Louis, MO) and incubated for 20 min at 4°C. The swab head was squeezed with a crystal chip in the 1.5 mL tube, and 2 μL of the incubated solution was mixed with 20 μL of solution A (Kaneka) and incubated at 75°C for 5 min. After cooling down to room temperature, 1 μL of DNAse I solution included in the kit was added and then incubated at 42°C for 10 min followed by 75°C for 5 min. The concentration of each RNA sample was measured by a spectrophotometry (NanoDrop ND-2000, Thermo Scientific, Wilmington, DE, USA). RNA samples were stored at –80°C until the RT-LAMP reaction.

RNA sequencing.

Total RNA extraction was conducted using ISOGEN II (Nippon Gene, Tokyo, Japan) as described previously [10]. The concentration of each sample was measured by NanoDrop ND-2000 and provided for sequencing. Construction of sequencing libraries of VMM-derived RNA was conducted by TruSeq RNA Library Prep kit v2 (Illumina, San Diego, CA, USA) according to the manufacturer’s instruction. After checking quality

of the sequence library using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), sequence libraries were read on the Illumina HiSeq 2500 platform. The sequencing data used in this experiment have been deposited in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) with the accession number PRJDB6759. The raw reads were subjected to adapter sequence trimming and quality control using fastp (version 0.20.1) [14]. The trimmed sequences underwent mapping to the bovine reference genome Bos_taurus_UMD_3.1.1 and then count quantification using STAR (version 2.7.8a) and RSEM (version 1.3.3) software, respectively, with default parameters. Differential expression genes (DEGs) between non-pregnant and pregnant samples were assessed using edgeR in R package (version 3.30.3). In this study, genes with adjusted *P*-value < 0.001 and the value of log2 fold-change ≥ 2 or ≤ –2 were defined as DEGs, as strict filtering thought to allow for efficient marker selection.

RT-LAMP.

For the RT-LAMP assay, the Loop-amp RNA Amplification Kit (Eiken Chemical Co. Ltd., Tokyo, Japan) was used. Primers were designed using Primer Explorer v5 (Eiken Chemical Co., Ltd., https://primerexplorer.jp/lampv5/index.html) (Supplementary Table). The reaction mixture consisted of 12.5 µL of 2× Reaction Mix, 1 µL Enzyme mix, 1 µL F3 and B3 primers (5 µM) each, 0.8 µL FIP and BIP primers (50 µM) each, and 1 µL of the total RNA with nuclease-free water up to 25 µL for a final volume. The LAMP assay was performed with MyAbscope® (Kaneka), absorption spectrophotometer with a temperature control function, under isothermal condition of 65°C for 70 min. To provide input data for later analysis, the rise-up time (RUT) in the LAMP reaction was used. The RUT was defined as the first time when the absorbance exceeded 0.05. If the absorbance did not exceed 0.05 throughout the reaction, RUT was determined as the total reaction

time, 4200 seconds. The RUT of each sample was calculated from the output of the MyAbscope® (Kaneka) application for further analysis.

Development of machine learning model.

To develop a binary classification algorithm for determining pregnancy or non-pregnancy based on RUT, Light Gradient Boosting Machine (LightGBM), a decision tree-based algorithm [15], was adopted. The 80-VMM samples (non-pregnant and pregnant: n = 40, each sample) were randomly divided into 50 samples (for training data set) and 30 samples (for test data set). In developing the LightGBM model using the training data set, the hyperparameters were optimized by leave-one-out cross-validation. Using the determined parameters, 50 training data were retrained to create a discrimination model for pregnancy. The input values of 30 test data were applied to the developed models, and the predicted results were compared with the actual pregnancy diagnosis results on d 30. The whole model development and evaluation process is shown as a schematic workflow in supplementary figure. All programing was conducted based on Python 3.8 code with scikit-learn library (version 0.24.1).

Assessment of the prediction performance.

To visualize and compare the performance of each pregnancy predictor, the Receiver Operating Characteristic (ROC) curve was plotted, and the Area Under the Curve (AUC) scores were calculated. Based on each ROC curve, Youden index was used to determine a proper cutoff value. To evaluate the discrimination performance of each predictor, evaluation metrics were calculated using number of True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN) as follows: Accuracy = (TP + TN) / (TP + TN + FP + FN), Sensitivity (probability that a positive test result will be obtained when a cow is pregnant) = TP / (TP + FN), Specificity (probability that a

negative test result will be obtained when a cow is non-pregnant) = TN / (TN + FP), PPV (probability that a cow is pregnant when the test result is positive) = TP / (TP + FP), NPV (probability that a cow is non-pregnant when the test result is negative) = TN / (TN + FN). All calculation was conducted using the scikit-learn library based on python 3.8.

Statistical analysis.

Welch's *t*-test was conducted to determine the significant difference between the mean values of TPM (transcripts per million) or RUT in pregnant and non-pregnant cows by R (version 4.0.2).

Results

Screening of candidate genes for early pregnancy detection.

To select pregnancy marker genes, we screened a comprehensive gene expression analysis dataset of pregnant and non-pregnant VMM. We defined the DEGs with strict filtering (adjusted-$P < 0.001$, log2 fold-change ≥ 2 or ≤ -2) to ensure that the selected genes can be applied to the high accuracy-determination of pregnancy and extracted 55 genes that fit the criteria (Supplementary data). Of the 55 DEGs, 42 were upregulated, whereas 13 were downregulated by pregnancy. Among these, we selected three possible pregnancy marker genes, *ISG15* and *IFIT1* as up-regulated, and *MUC16* as a down-regulated marker, based on the absolute amount of TPM. The TPM values of both *ISG15* and *IFIT1* were significantly higher in the pregnant sample compared to non-pregnant (Fig. 2A, B), whereas the TPM values of *MUC16* significantly decreased in the pregnant sample compared to that in non-pregnant samples (Fig. 2C). Collectively, we decided to use *ISG15*, *IFIT1*, and *MUC16* as suitable marker genes for LAMP reaction.

Predicting performance using RUT of each gene.

We conducted the RT-LAMP reaction for *ISG15*, *IFIT1*, and *MUC16* with RNA from pregnant or non-pregnant VMM. For *ISG15*, the plots were scattered between 2,000 and 3,000 seconds, regardless of the sample concentration (Fig. 3A). Although significant difference ($P = 0.003$) was observed in RUT between non-pregnant and pregnant samples, it showed low prediction performance (AUC = 0.69) (Fig. 3D). The *IFIT1*, almost all pregnant samples raised up between 1,500 s and 3,000 s, while 62.5% of non-pregnant cows (n = 25) never exceeded the threshold throughout LAMP assay (Fig. 3B) with significant difference ($P = 4.253\times$e-11) between non-pregnant and pregnant RUT. The ROC curve of *IFIT1* showed the highest performance (AUC = 0.88) of the three genes (Fig. 3D). The *MUC16*, a candidate gene for reduced expression on d 17-18, no significant difference was observed ($P = 0.241$) in RUT between non-pregnant and pregnant samples (Fig. 3C), leading to the lowest performance (AUC = 0.58) (Fig. 3D). The determined cutoff values for each predictor showed that *IFIT1* was the most prospective marker of the three genes with the highest accuracy value (85.0%) when the threshold set to 2,740 s (Table 1). The value of sensitivity and NPV was sufficiently reliable, 92.5% and 91.2%, respectively. However, the score of specificity was 77.5%, showing the difficulty of reducing false positive samples with a single marker.

Pregnancy prediction using the models trained with RUT of multiple markers.

Plotting RUT for all three candidate genes in 3 dimensions, there can be seen a rough separation between clusters consisting mainly of non-pregnant and pregnant samples (Fig. 4), suggesting the possibility of increasing prediction performance with multiple use of genes. To enhance each score of metrics, we developed multiple gene predictors trained with RUT of each combination of the markers. Using these models, we

predicted the practical pregnancy outcome of 30 cows. Table 2 shows the performance evaluation of the four models developed with each of the three marker gene combinations. Overall, the model created with *IFIT1* and *MUC16* showed the best performance with 90.0% accuracy. On the other hand, the triple marker model overcame the low specificity and PPV, but resulted in a decrease in sensitivity and PPV compared to the single use of *IFIT1*. The two sets of *IFIT1* and *ISG15*, *ISG15* and *MUC16*, had the same lowest performance.

Discussion

In the present study we demonstrated that three candidate marker genes, identified by RNA sequencing, enable early pregnancy detection with the LAMP assay and machine learning. We also highlight the entire process from sample collection to the judgments can be shortened with the perspective of performing the screening for entire AI-received herd.

In our comprehensive gene expression analysis to select candidate marker genes, 55 DEGs were identified. The result revealed that almost all upregulated DEGs belong to known ISGs. This corresponds with the results of several RNA-sequencing analyses conducted in bovine endometrium during pre-implantation [16], suggesting that conceptus secreted IFNT may alter the vaginal gene expression profile. The *MUC16* expression was significantly low in the VMM of pregnant cows compared to non-pregnant cows. The expression level of *MUC16* was significantly lower during luteal phase than at estrus in cervicovaginal mucus of ewes [17] and in cervical tissue of heifers [18] correlated with increasing estradiol levels. Thus, the high expression of *MUC16* on d 17-18 in non-pregnant cows compared to pregnant cows can be considered an increasing expression toward estrus.

The expression of *IFIT1* showed a clear difference in RUT between non-pregnant and pregnant samples. In contrast, the RUT of *ISG15*, the most popular ISGs, seemed not suitable for the pregnancy detection, although TPM comparison between non-pregnant and pregnant showed a significant difference. Focusing on non-pregnant samples, almost all RUT of *ISG15* exceeded the threshold before the end of the LAMP reaction, whereas for *IFIT1*, 62.5% samples did not reach the threshold until the reaction (RUT = 4200) (Fig. 3A, B). This suggests that lower expression level in non-pregnant sample is of importance for the successful judgment. Correspondingly, the TPM value of *ISG15* is about 10-fold higher compared to that of *IFIT1* in non-pregnant samples (Fig. 2A, B). Thus, to select more appropriate markers from ISGs, further study should be focus on sufficiently low expression genes in non-pregnant cows rather than high expression in pregnant cows. On the other hand, the combination use of *MUC16* and ISGs allowed for a more robust determination of pregnancy than ISGs alone. Although the RUT of *MUC16* did not clearly identify non-pregnant and pregnant cows, the approach of using downregulated genes will have a positive effect on early pregnancy predictions. Further study is needed to discover more proper down-regulated genes instead of *MUC16* for improvement of judgment performance.

Peripheral blood leukocyte-based early pregnancy detection has been conducted on d 20 or 21 with approximately 80% accuracy in Holstein heifers or cows [6, 19]. We demonstrated that the accuracy was high enough to practical use at d 17-18 in judging with *IFIT1* (= 85.0%) using VMM samples. Moreover, the combination-use of *IFIT1* and *MUC16* scored the high accuracy at 90.0% for 30 untrained data. A closer comparison of each metric in the Table 1 shows that sensitivity scores were better than specificity scores due to more false-positive samples compared with false-negative. Since most early embryonic death occurs before implantation [20, 21], it is possible that some of the cows that were determined to be non-pregnant by ultrasound in this study got embryonic

mortality between d 17 and d 30, resulting in false positives and the relatively low specificity score (77.5%) in judging with *IFIT1*. The machine learning algorithm can successfully utilize the advantages of *IFIT1* and *MUC16* and was able to increase the specificity to 86.7%. Considering these metrics were calculated with 30 cows that were not used for model training, we expect this model would work well for other untrained data. Since the generalized performance highly depends on the number of training data, a larger scale study will provide greater stability and reliability to the prediction model.

In the present study, we chose three marker genes; however, determining other genes or combination analysis may increase reliability as well as modifying sample collection, preparation and amplification condition.

In conclusion, we propose a new approach for easy pregnancy detection method using VMM with RT-LAMP assay. We developed machine learning models and found the combination use of *IFIT1* and *MUC16* gave the highest performance. Although more large-scale tests are warranted for commercial use, this study is expected to provide promising avenues to enhance the reproductive performance in dairy industry.

Acknowledgements

Funding: This study was supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (KAKENHI, 19H03099, 20J11776) and Japan Racing Association (JRA) and Cooperative Research Grant of the Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture. The authors thank Experimental farm, Hokkaido University for managing and assistant sampling. The authors thank Ms. Kathy Craighead, the University of Missouri, USA, for English language editing the manuscript.

References

[1] L. Commun, K. Velek, J.-B. Barbry, S. Pun, A. Rice, A. Mestek, C. Egli, S. Leterme, Detection of pregnancy-associated glycoproteins in milk and blood as a test for early pregnancy in dairy cows, J. Vet. Diagn. Invest. 28 (2016) 207–213. https://doi.org/10.1177/1040638716632815.

[2] W. Holtz, H. Niggemeyer, Reliable identification of pregnant dairy cows by double milk progesterone analysis, Livest. Sci. 228 (2019) 38–41. https://doi.org/10.1016/j.livsci.2019.07.014.

[3] J.D. Godkin, F.W. Bazer, J. Moffatt, F. Sessions, R.M. Roberts, Purification and properties of a major, low molecular weight protein released by the trophoblast of sheep blastocysts at day 13-21, J. Reprod. Fertil. 65 (1982) 141–150. https://doi.org/10.1530/jrf.0.0650141.

[4] K. Imakawa, R.V. Anthony, M. Kazemi, K.R. Marotti, H.G. Polites, R.M. Roberts, Interferon-like sequence of ovine trophoblast protein secreted by embryonic trophectoderm, Nature. 330 (1987) 377–379. https://doi.org/10.1038/330377a0.

[5] T.R. Hansen, K.J. Austin, D.J. Perry, J.K. Pru, M.G. Teixeira, G.A. Johnson, Mechanism of action of interferon-tau in the uterus during early pregnancy, J. Reprod. Fertil. Suppl. 54 (1999) 329–339.

[6] H. Yoshino, N. Toji, K. Sasaki, K. Koshi, N. Yamagishi, T. Takahashi, T. Ishiguro-Oonuma, H. Matsuda, T. Yamanouchi, Y. Hashiyada, K. Imai, Y. Izaike, K. Kizaki, K. Hashizume, A predictive threshold value for the diagnosis of early pregnancy in cows using interferon-stimulated genes in granulocytes, Theriogenology. 107 (2018) 188–193. https://doi.org/10.1016/j.theriogenology.2017.11.014.

[7] H. Yoshino, K. Kizaki, K. Iga, T. Hirata, H. Matsuda, T. Yamanouchi, Y. Hashiyada, N. Toji, T. Ishiguro-Oonuma, T. Takahashi, K. Hashizume, Use of a prediction method for early pregnancy status utilizing receiver operating characteristic curve analysis of

peripheral blood leukocyte interferon-stimulated genes in Japanese-Black cattle, Anim. Reprod. Sci. 214 (2020) 106283. https://doi.org/10.1016/j.anireprosci.2020.106283.

[8] J.C. Green, C.S. Okamura, S.E. Poock, M.C. Lucy, Measurement of interferon-tau (IFN-τ) stimulated gene expression in blood leukocytes for pregnancy diagnosis within 18–20d after insemination in dairy cattle, Anim. Reprod. Sci. 121 (2010) 24–33. https://doi.org/10.1016/j.anireprosci.2010.05.010.

[9] H. Han, K.J. Austin, L.A. Rempel, T.R. Hansen, Low blood ISG15 mRNA and progesterone levels are predictive of non-pregnant dairy cows, J. Endocrinol. 191 (2006) 505–512. https://doi.org/10.1677/joe.1.07015.

[10] H. Kunii, K. Koyama, T. Ito, T. Suzuki, A.Z. Balboula, T. Shirozu, H. Bai, M. Nagano, M. Kawahara, M. Takahashi, Hot topic: Pregnancy-induced expression of interferon-stimulated genes in the cervical and vaginal mucosal membranes, J. Dairy Sci. 101 (2018) 8396–8400. https://doi.org/10.3168/jds.2017-14251.

[11] T. Notomi, H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, T. Hase, Loop-mediated isothermal amplification of DNA, Nucleic Acids Res. 28 (2000) e63.

[12] S. Higaki, R. Miura, T. Suda, L.M. Andersson, H. Okada, Y. Zhang, T. Itoh, F. Miwakeichi, K. Yoshioka, Estrous detection by continuous measurements of vaginal temperature and conductivity with supervised machine learning in cattle, Theriogenology. 123 (2019) 90–99. https://doi.org/10.1016/j.theriogenology.2018.09.038.

[13] S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke, K. Weigel, Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms, J. Dairy Sci. 97 (2014) 731–742. https://doi.org/10.3168/jds.2013-6693.

[14] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics. 34 (2018) i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

[15] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Adv. Neural Inf. Process. Syst. 30 (2017) 3146–3154.

[16] N. Forde, F. Carter, T.E. Spencer, F.W. Bazer, O. Sandra, N. Mansouri-Attia, L.A. Okumu, P.A. McGettigan, J.P. Mehta, R. McBride, P. O'Gaora, J.F. Roche, P. Lonergan, Conceptus-Induced Changes in the Endometrial Transcriptome: How Soon Does the Cow Know She Is Pregnant?, Biol. Reprod. 85 (2011) 144–156. https://doi.org/10.1095/biolreprod.110.090019.

[17] J.W. Maddison, J.P. Rickard, N.C. Bernecic, G. Tsikis, C. Soleilhavoup, V. Labas, L. Combes-Soia, G. Harichaux, X. Druart, T. Leahy, S.P. de Graaf, Oestrus synchronisation and superovulation alter the cervicovaginal mucus proteome of the ewe, J. Proteomics. 155 (2017) 1–10. https://doi.org/10.1016/j.jprot.2017.01.007.

[18] K. Pluta, P.A. McGettigan, C.J. Reid, J.A. Browne, J.A. Irwin, T. Tharmalingam, A. Corfield, A. Baird, B.J. Loftus, A.C.O. Evans, S.D. Carrington, Molecular aspects of mucin biosynthesis and mucus formation in the bovine cervix during the periestrous period, Physiol. Genomics. 44 (2012) 1165–1178. https://doi.org/10.1152/physiolgenomics.00088.2012.

[19] G. Dalmaso de Melo, B.P. Mello, C.A. Ferreira, C.A. Souto Godoy Filho, C.C. Rocha, A.G. Silva, S.T. Reese, E.H. Madureira, K.G. Pohler, G. Pugliesi, Applied use of interferon-tau stimulated genes expression in polymorphonuclear cells to detect pregnancy compared to other early predictors in beef cattle, Theriogenology. 152 (2020) 94–105. https://doi.org/10.1016/j.theriogenology.2020.04.001.

[20] J.F. Roche, M.P. Bolandl, T.A. McGeady, Reproductive wastage following artificial insemination of heifers, Vet. Rec. 109 (1981) 401–404. https://doi.org/10.1136/vr.109.18.401.

[21] M.G. Diskin, J.J. Murphy, J.M. Sreenan, Embryo survival in dairy cows managed under pastoral conditions, Anim. Reprod. Sci. 96 (2006) 297–311. https://doi.org/10.1016/j.anireprosci.2006.08.008.

Figure/Table legends

Figure 1
The sampling equipment used in this study consisted of a cotton swab, a pick-up tool, and fiber scope (A). A representative figure of sampling location (B).

Figure 2
TPM (transcripts per million) comparison of *ISG15* (A), *IFIT1* (B), and *MUC16* (C) between non-pregnant and pregnant VMM. NP: non-pregnant, P: pregnant (n = 3, each). The bar plots are shown as the mean ± SEM; the significant differences between pregnancy status are determined by Welch's *t*-test. *: $P < 0.05$.

Figure 3
Plots of the RUT (rise-up time, the first time when absorbance exceeded 0.05 in RT-LAMP reaction) and RNA concentration of *ISG15* (A), *IFIT1* (B), and *MUC16* (C). Dotted lines represent the mean value of RUT in non-pregnant (black) and pregnant (red) samples. (D) Receiver Operating Characteristic (ROC) curves drawn with the RUT of *ISG15* (blue line), *IFIT1* (orange line), and *MUC16* (green line). NP: non-pregnant, P: pregnant. AUC: Area under the curve.

Figure 4
A Three-dimensional plot of RUT (rise-up time) for each marker gene. NP: non-pregnant, P: pregnant.

Table 1
Discrimination ability of each single gene for pregnancy detection on d 17-18 after AI.

Table 2
Performance of the pregnancy prediction models assessed by untrained 30 samples.

Supplementary Figure
A schematic flow of machine learning model development and assessment. RUT: rise-up time (the first time when the absorbance exceeded 0.05 in RT-LAMP reaction).

Supplementary Table
Primer sequences for RT-LAMP.

Highlights

Quick and high-sensitive pregnancy detection in cows by LAMP-machine learning

Quick and non-invasive collection of vaginal mucosa for pregnancy detection

Identification of up and downregulated genes in pregnant bovine vaginal mucosa

A combined use of up and downregulated genes can increase prediction accuracy

Table 1

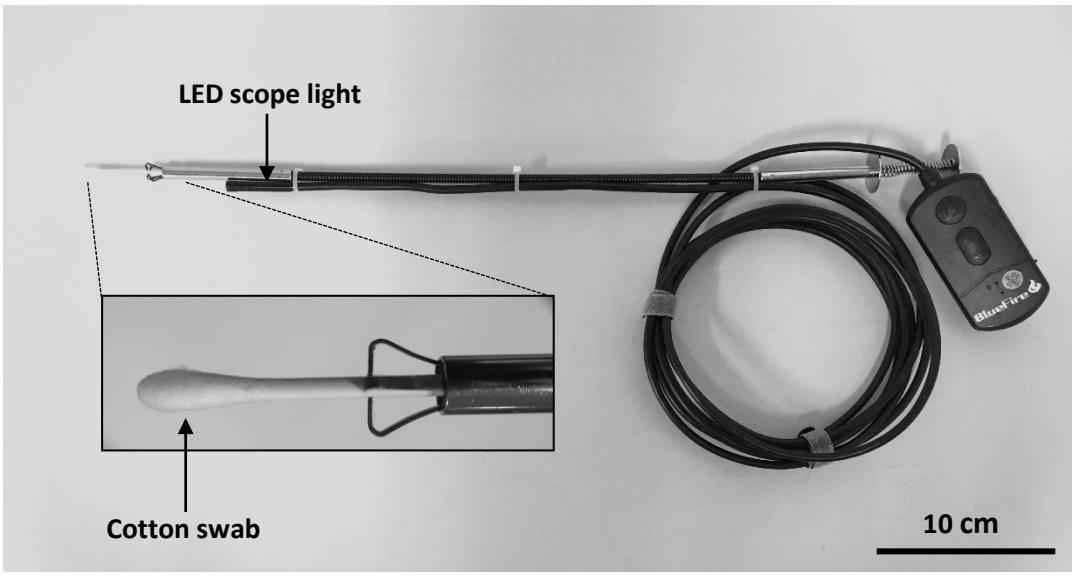
Gene	Threshold time (s)	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
ISG15	2,440	66.3	77.5	55.0	63.3	71.0
IFIT1	2,740	85.0	92.5	77.5	80.4	91.2
MUC16	2,880	60.0	70.0	50.0	58.3	62.5

Table 2

Gene combination	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
ISG15, *IFIT1*, *MUC16*	86.7	86.7	86.7	86.7	86.7
IFIT1, MUC16	90.0	93.3	86.7	87.5	92.9
IFIT1, *ISG15*	53.3	66.7	40.0	52.6	54.5
ISG15, *MUC16*	53.3	66.7	40.0	52.6	54.5

Fig. 1

A


LED scope light
Cotton swab
10 cm

B

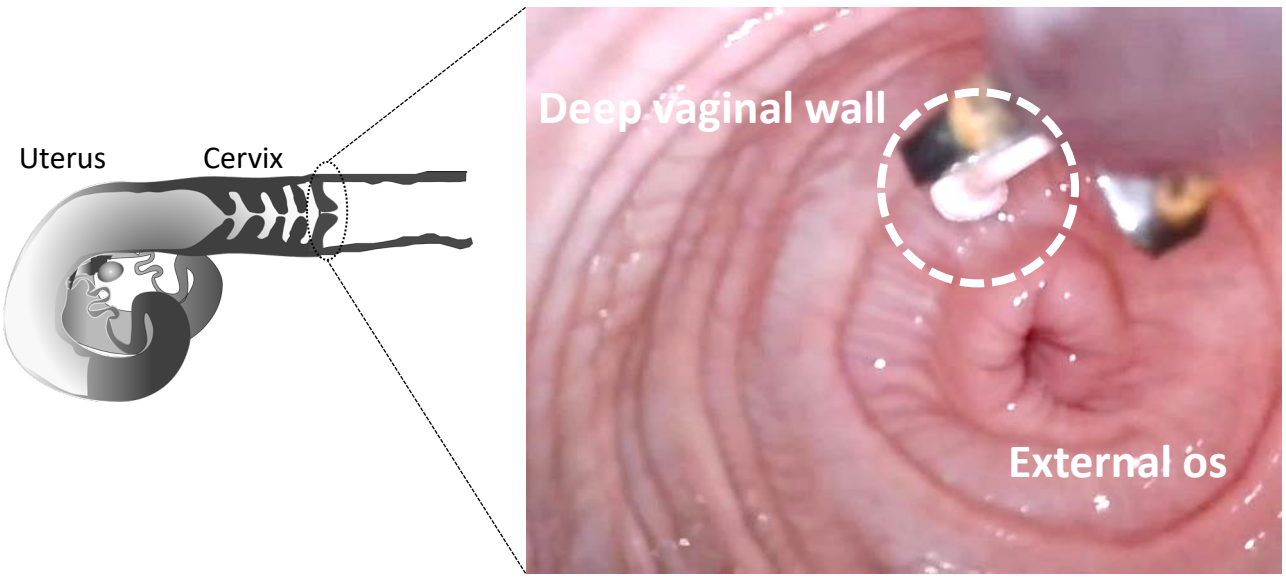

Uterus
Cervix
Deep vaginal wall
External os

Figure. 2

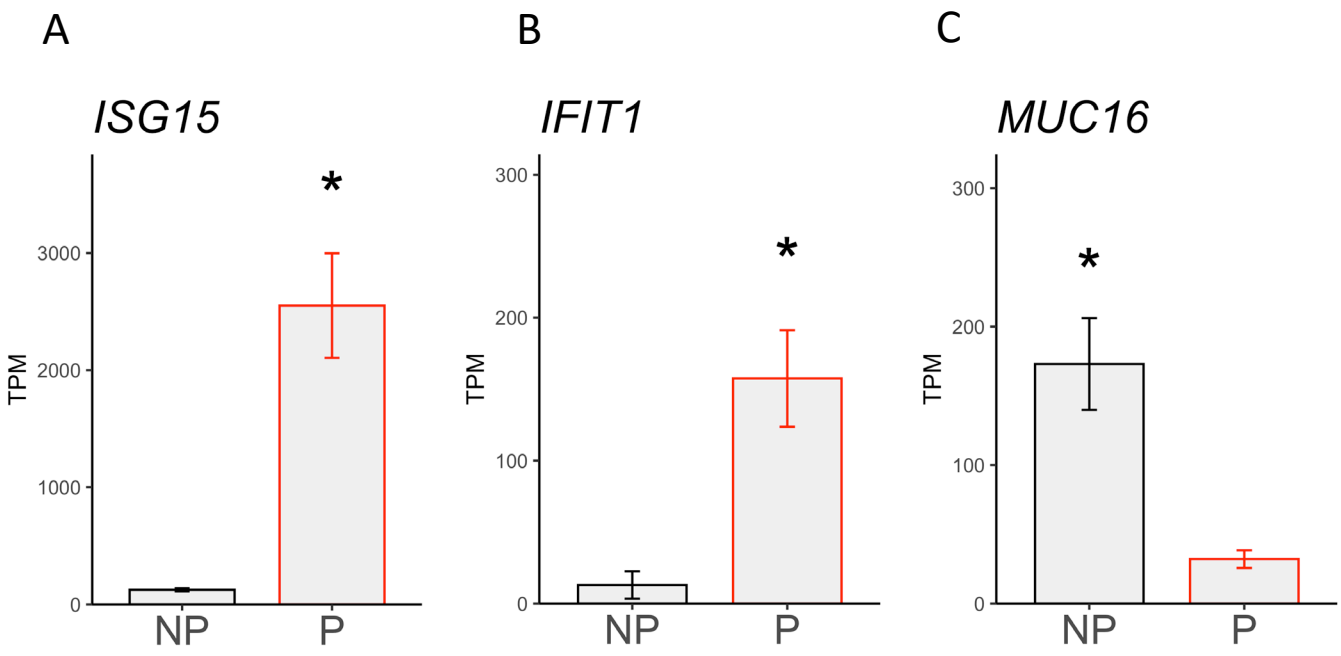
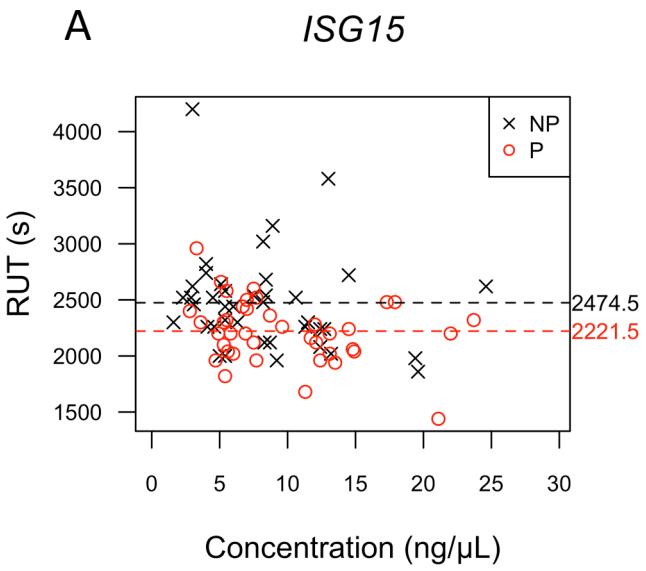
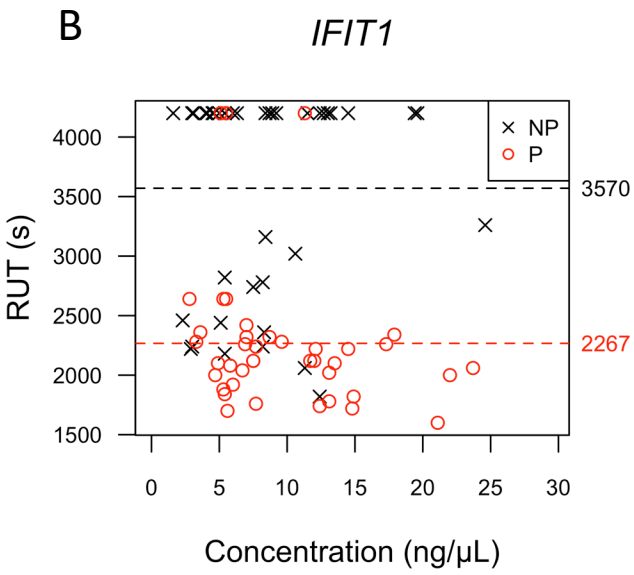
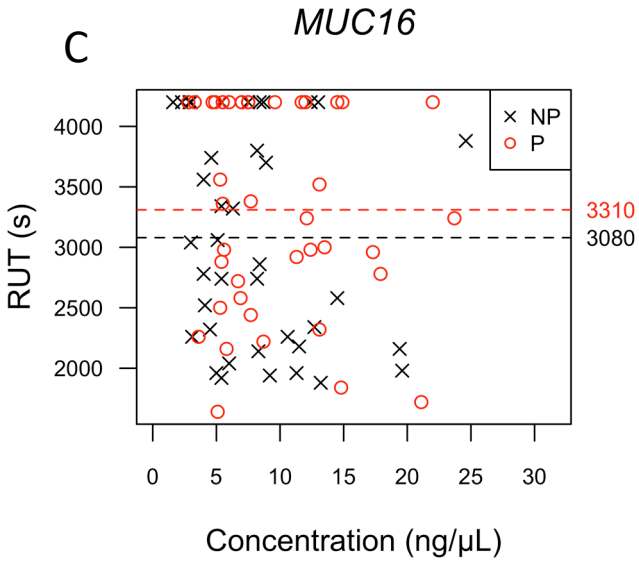

A
ISG15
TPM
3000
2000
1000
0
*
NP
P
B
IFIT1
TPM
300
200
100
0
*
NP
P
C
MUC16
TPM
300
200
100
0
*
NP
P

Figure. 3


A
ISG15
NP
P
RUT (s)
4000
3500
3000
2500
2000
1500
2474.5
2221.5
0
5
10
15
20
25
30
Concentration (ng/μL)


B
IFIT1
NP
P
RUT (s)
4000
3500
3000
2500
2000
1500
3570
2267
0
5
10
15
20
25
30
Concentration (ng/μL)


C
MUC16
NP
P
RUT (s)
4000
3500
3000
2500
2000
3310
3080
0
5
10
15
20
25
30
Concentration (ng/μL)

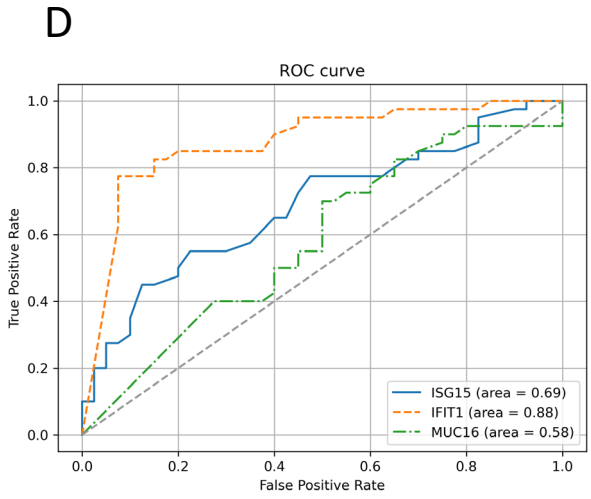
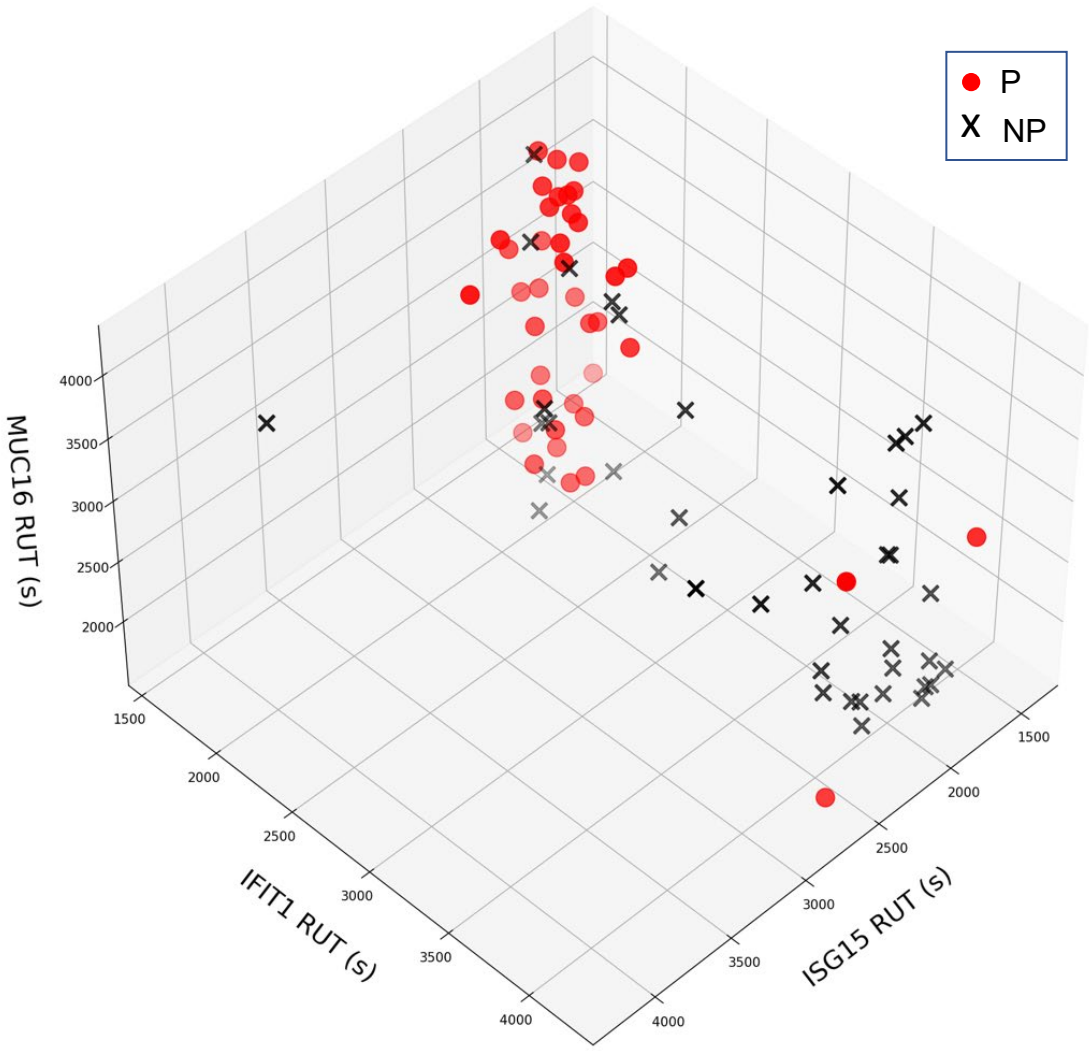

D
ROC curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ISG15 (area = 0.69)
IFIT1 (area = 0.88)
MUC16 (area = 0.58)

Figure. 4


P
NP
MUC16 RUT (s)
4000
3500
3000
2500
2000
IFIT1 RUT (s)
1500
2000
2500
3000
3500
4000
ISG15 RUT (s)
1500
2000
2500
3000
3500
4000