

<https://doi.org/10.1038/s41538-025-00421-y>

# Utilizing drip metabolites and predictive modeling for non-destructive freshness assessment in pork loin



Hyun-Jun Kim<sup>1</sup>, Hye-Jin Kim<sup>2</sup>, Heesang Hong<sup>2</sup>, Minwoo Choi<sup>2</sup>, Azfar Ismail<sup>2</sup>, Daye Mun<sup>2</sup>, Younghoon Kim<sup>2</sup>, Gap-Don Kim<sup>1,3</sup> & Cheorun Jo<sup>1,2</sup>✉

This study validated the use of pork drip metabolites for non-destructive freshness prediction. The pork loin was vacuum-packaged and stored for 27 days at 4 °C. The pH, drip loss, total aerobic bacterial counts (TAB), microbial composition and drip metabolites were examined. LASSO and Random Forest (RF) were selected and used for variable selection, while Ridge regression and Support Vector Regression were utilized to develop predictive models. Validation was performed using leave-one-out cross-validation. LASSO and RF selected 13 and 10 metabolites, respectively. The metabolites selected by each method were trained using Ridge regression and SVR. Each of the four trained models achieved  $R^2$  values of over 0.9. In the validation step, the model trained by Ridge regression using drip metabolites selected through LASSO showed the lowest RMSE value of 0.283 log CFU/g. Therefore, selected drip metabolites can be used to predict TAB and microbial composition of pork loin through mathematical modeling.

The deterioration in the quality of refrigerated pork is primarily due to the proliferation of microorganism<sup>1</sup>. In particular, microbial contamination that occurs during slaughtering processes can lead to deterioration of freshness. If microbes in meat grow to high levels and cause odors, exudates, and slime, the meat is considered unsuitable for human consumption (EU Regulation (EC) No 178/2002, Article 14). In relation to the EU's regulation, if contamination at the production site exceeds  $5 \times 10^6$  CFU/g (total bacterial count), improvements in hygiene practices during production are required (European Commission Regulation (EC) No 2073/2005). Typically, the inspection of microbial count has been used as a plate counting method<sup>2</sup>, which is a method to destroy food materials and cause economic losses<sup>3</sup>. Additionally, the economic losses make it challenging to test many samples<sup>4</sup>. Thus, it is crucial to develop a non-destructive method to predict pork freshness, ensuring consumer safety and maintaining trust in the meat industry.

The traditional methods for assessing meat freshness include sensory evaluation, total microbial content analysis, and the measurement of volatile basic nitrogen (VBN) levels<sup>5</sup>. However, these conventional approaches often face limitations, such as slow processing times and the requirement for sample destruction, which limit their practicality<sup>6,7</sup>. Recently, innovative non-destructive technologies have been developed for monitoring meat spoilage<sup>6-9</sup>. Spectroscopic methods utilize information from overtones and

O-H, C-H, and N-H bonds to measure VBN and TBC<sup>6,8,9</sup>. Smart films detect spoilage through color changes in response to volatile compounds such as CO<sub>2</sub>, H<sub>2</sub>S, and volatile nitrogen emitted during the process of meat spoilage<sup>7</sup>. These innovations enhance food safety and quality assurance.

Pork contains about 70% water, which is found within muscle fibers, between muscle fibers, cell membranes, muscle cells, and muscle bundles<sup>10</sup>. In meat, water is classified into bound water, immobilized water, and free water<sup>5</sup>. Both immobilized and free water can exude to the surface during storage due to structural changes in the meat<sup>5</sup>. These structural changes are influenced by the contamination and proliferation of spoilage microbes<sup>11</sup>. The exuded water is referred to as drip<sup>5</sup>. It was reported that drip can represent the metabolites of meat<sup>5</sup>, and that meat quality can be predicted through drip analysis<sup>12</sup>. Thus, the drip can be used to non-destructively assess the condition of pork, as they are a result of the activity of spoilage microbial.

Metabolomics, the study of metabolites produced by intracellular biochemical reactions, has been used in food science related to food quality and safety<sup>4</sup>. The metabolites in meat can be altered by the growth of spoilage microbes<sup>13</sup>. Kim et al.<sup>4</sup> applied metabolites to a classification model to categorize meat spoilage. Additionally, metabolites have been used to distinguish between fresh and frozen meat using selected biomarkers through multivariate models<sup>14</sup>. Therefore, by utilizing an appropriate model, the

<sup>1</sup>Institutes of Green Bio Science and Technology, Seoul National University, Pyeongchang, 25354, Republic of Korea. <sup>2</sup>Department of Agricultural Biotechnology, Center for Food and Bioconvergence, and Research Institute of Agriculture and Life Science, Seoul National University, Seoul, 08826, Republic of Korea. <sup>3</sup>Graduate School of International Agricultural Technology, Seoul National University, Pyeongchang, 25354, Republic of Korea. ✉e-mail: [cheorun@snu.ac.kr](mailto:cheorun@snu.ac.kr)

metabolites in drip can be analyzed to effectively predict meat spoilage and microorganisms.

When modeling with metabolomics data, it is essential to select and reduce the number of variables<sup>15</sup>. Because a high correlation among variables can lead to multicollinearity, which may cause errors in the estimation of model coefficients and lead to overfitting issues<sup>15</sup>. Additionally, reducing the number of variables makes model training more efficient and the resulting model easier to interpret<sup>16</sup>. To reduce the number of variables, various variable selection models are used<sup>16,17</sup>. For instance, the least absolute shrinkage and selection operator (LASSO) applies L1 regularization to drive the coefficients of less important variables to zero, automatically performing variable selection and dimension reduction, thus preventing model overfitting and reducing complexity<sup>18</sup>. On the other hand, Random Forest (RF) is an ensemble learning model where each decision tree independently learns from bootstrap samples and randomly selects variables during tree construction<sup>17</sup>. This randomness increases the diversity of the model overall, preventing overfitting and reducing variance<sup>17</sup>. Modeling based on the selected variables is also very important<sup>16</sup>. Ridge regression uses L2 regularization to adjust the coefficients of all variables, ensuring model stability and helping prevent overfitting<sup>16</sup>. Meanwhile, Support Vector Regression (SVR) utilizes the kernel trick to effectively capture patterns in the data and solve the optimization problem posed, modeling complex relationships<sup>19</sup>. In this process, SVR aims to include as many data points as possible within the set margin, contributing to the model's generalization performance<sup>19</sup>. Thus, by using variables chosen by LASSO and RF, Ridge regression and SVR effectively leverage the strengths of each technique to predict complex biological information from metabolomics data.

Therefore, this study aimed to predict the freshness of pork using metabolites from the drip and predictive models. Additionally, it intended to estimate the proportion of microorganisms contaminating the pork.

## Results and discussion

### Changes in pork quality characteristics during storage

During the storage period, there was no significant difference in the pH of the pork. The pH was observed to be 5.91 on the first day of storage and decreased to 5.62 by day 27 (Fig. 1a). The pork producer's council has reported that the ideal pH range for pork is between 5.6 and 5.9<sup>20</sup>. The pH of pork is significantly influenced by microbial composition<sup>21,22</sup>. An increase in pH during storage is linked to the predominance of spoilage microorganisms like *Brochothrix thermosphacta*, *Shewanella putrefaciens*, *Pseudomonas* spp<sup>22</sup>, while lactic acid bacteria are associated with pH decreases<sup>22</sup>. Thus, pH changes in pork during storage are likely driven by microbial composition.

The drip loss in pork significantly increased from a 2.12% initially to 11.41% by 27 days of storage (Fig. 1b). The increase in drip is influenced by

the breakdown of muscle tissue<sup>11</sup>, and Chen et al.<sup>23</sup> reported that the destruction of muscle cell structure in pork affects drip loss. Thus, the increased drip loss might be affected by changes in the muscle tissues over the storage period.

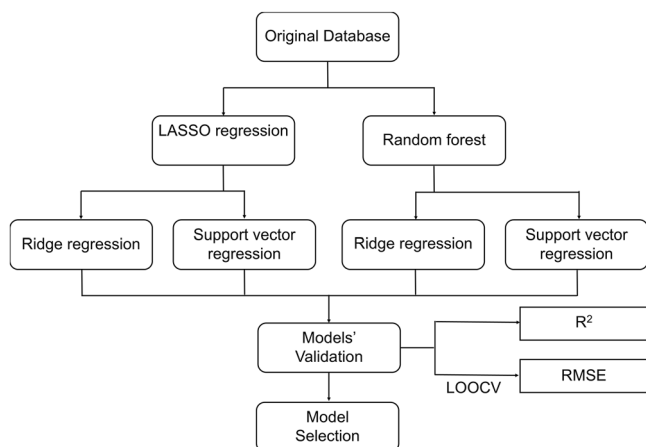
TAB significantly increased from 3.82 log CFU/g on the first day of storage and increased to 7.53 log CFU/g by day 20 (Fig. 2a). It was reported that enzymes secreted by microbes influence changes in the structural tissues of meat<sup>24</sup>, which might have affected the drip loss in pork. To identify the contaminated microbes in pork, microbial compositions were analyzed using 16S rRNA sequencing on days 4, 8, and 20 (Fig. 2b). It was observed that the proportion of *Enterobacteriaceae*, *Lactobacillaceae*, and *Leuconostoc* increased in all three carcasses over the storage period. Microbes contaminate the pork at the time of slaughter<sup>25</sup>, and their composition can change depending on storage conditions<sup>1</sup>. The increase in microbial composition seems influenced by facultative anaerobic conditions in vacuum storage, likely affecting quality factors like pH (Fig. 3).

During the storage period, a total of 63 metabolites were qualitatively analyzed in the drip using <sup>1</sup>H-<sup>13</sup>C HSQC 2D NMR (Table S1). In overall, 30 types of carboxylic acids and derivatives, 12 types of organooxygen compounds, 7 types of purine nucleotides, 4 types of organonitrogen compounds, and 10 types of other compounds were identified. The metabolites were quantitatively analyzed using <sup>1</sup>H zg30 NMR (Table S2). The metabolites that significantly increased over the storage period were acetate, agmatine, alanine, asparagine, aspartate, cadaverine, ethanol, fucose, glutamate, glutamine, inosine, isoleucine, leucine, lysine, malate, methionine, phenylalanine, proline, serine, threonine, tyramine, tyrosine, uridine, valine, and phosphocholine. The increase in microbes during the storage period can enhance the protein degradation process, which increases the amino acids<sup>3</sup>. In particular, biogenic amines such as cadaverine and tyramine can be produced by *Lactobacillaceae* like *Leuconostoc*<sup>26</sup>. Additionally, amino, carboxyl, and hydroxyl groups, which are polar, can easily mix into the drip and be exuded<sup>27</sup>. Thus, the metabolite changes in the drip could be influenced by the TAB. Therefore, a model of the TAB was developed using the drip metabolites to predict meat freshness.

### Selection of metabolites for modeling using mathematical methods and machine learning

Metabolites for TAB modeling were selected using LASSO regression and RF. LASSO regression was applied to the entire set of metabolites for variable selection optimization through cross-validation, and the model was then rerun with the selected metabolites<sup>18</sup>. The metabolites selected by LASSO include ADP, carnosine, creatine, ethanol, fumarate, glutamine, glycerol, hypoxanthine, IMP, malate, O-acetylcarnitine, tyrosine, and uracil. RF selected variables that showed an importance score of 10 or higher after running the seed 10 times<sup>28</sup>. The metabolites chosen by RF are acetate, adenosine, alanine, asparagine, aspartate, cadaverine, hypoxanthine, IMP, isoleucine, and tyramine. Each of selected metabolites by different method was used for modeling.

To assess the effectiveness of feature selection in reducing multicollinearity, the variance inflation factor (VIF) was calculated for all metabolites before and after selection (Table S3). Prior to selection, all metabolites exhibited extreme multicollinearity (VIF = ∞). However, after LASSO selection, the VIF values for the selected metabolites were significantly reduced, with ADP (9.960), ethanol (3.411), fumarate (4.693), glycerol (4.048), hypoxanthine (9.413), IMP (6.154), malate (7.505), O-acetylcarnitine (2.626), tyrosine (1.833), uracil (5.400) showing improved collinearity. Similarly, RF selection effectively minimized multicollinearity, reducing VIF values for acetate (3.352), adenosine (1.310), cadaverine (2.528), hypoxanthine (3.267), IMP (2.674), tyramine (4.617). These results indicate that both LASSO and RF contributed to eliminating redundant variables while preserving essential predictors for microbial composition modeling. Each set of selected metabolites was subsequently used for modeling to predict TAB.



**Fig. 1 | The illustration of workflow for variable selection and modeling to predict total microbial and microbial composition in pork using drip.** LOOCV leave-one-out cross-validation, RMSE root mean square error.

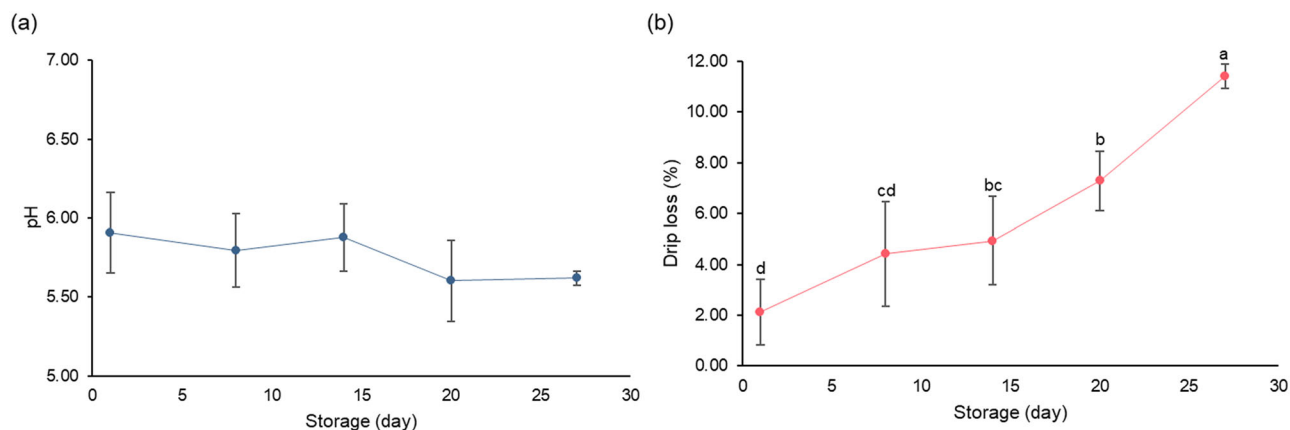


Fig. 2 | Quality changes in pork during storage period. a pH. b Drip loss. Different letters (a–d) indicate significant differences ( $p < 0.05$ ).

### Modeling using mathematical method and machine learning for TAB

Metabolites selected by each method were modeled using the mathematical method of Ridge regression and the machine learning method of SVR. When the metabolites, selected by LASSO regression, were applied to Ridge regression (Eq. (1)) and SVR (Eq. (2)), the  $R^2$  values were 0.968 and 0.970, respectively. Using the metabolites selected by RF, Ridge regression (Eq. (3)) and SVR (Eq. (4)) models were constructed, and the  $R^2$  values were 0.847 and 0.876, respectively. The developed models were validated using RMSE values obtained through LOOCV (Fig. 4). The RMSE values were 0.283, 0.387, 0.598, and 0.519 for Eq. (1) through Eq. (4), respectively. Previous studies have demonstrated that using LASSO and Ridge regression offers significant advantages in predictive modeling. For instance, Wang et al.<sup>29</sup> showed that the LASSO-logistic regression model provided superior predictive accuracy for identifying neurotoxicity biomarkers. Similarly, Zhang et al.<sup>30</sup> found that LASSO regression outperformed other models in predicting neonatal sepsis, achieving the highest accuracy. Furthermore, Krstic et al.<sup>31</sup> reported that LASSO excelled in variable selection, while Ridge regression stabilized models by effectively handling multicollinearity. Therefore, the mathematical model (LASSO regression) that selected the metabolites and modeled them using Ridge regression demonstrated the best performance (Eq. (1)), as it had the high  $R^2$  and the lowest RMSE values.

In the early post-mortem period, metabolite changes were primarily influenced by intrinsic enzymatic processes rather than microbial activity<sup>32</sup>, as indicated by minimal bacterial proliferation (Fig. 2a). Metabolites such as IMP and hypoxanthine (Table S2) are typical byproducts of post-mortem ATP degradation, reflecting endogenous biochemical processes<sup>32</sup>. As storage progressed, microbial proliferation increased significantly (Fig. 2a). For example, *Lactobacillaceae* species contributed to the production of ethanol and biogenic amines such as cadaverine and tyramine<sup>13,33</sup>. In general, microbial nucleic acid metabolism and energy production processes are reflected in metabolites such as ADP, IMP, hypoxanthine, and uracil<sup>34,35</sup>. ADP, IMP, hypoxanthine, and uracil are involved in microbial nucleic acid metabolism and energy production processes<sup>34</sup>. ADP is produced when energy is released from ATP, while IMP and hypoxanthine are produced through the breakdown of nucleic acids<sup>34</sup>. Uracil, a component of RNA, is directly involved in the metabolism of microbial genetic material<sup>35</sup>. Amino acid-related metabolites such as glutamine, tyrosine, carnosine, and creatine have crucial roles in the synthesis and breakdown of proteins by microbial<sup>36</sup>. O-acetylcarnitine and glycerol can be affected by microbial fatty acid metabolism<sup>37</sup>.

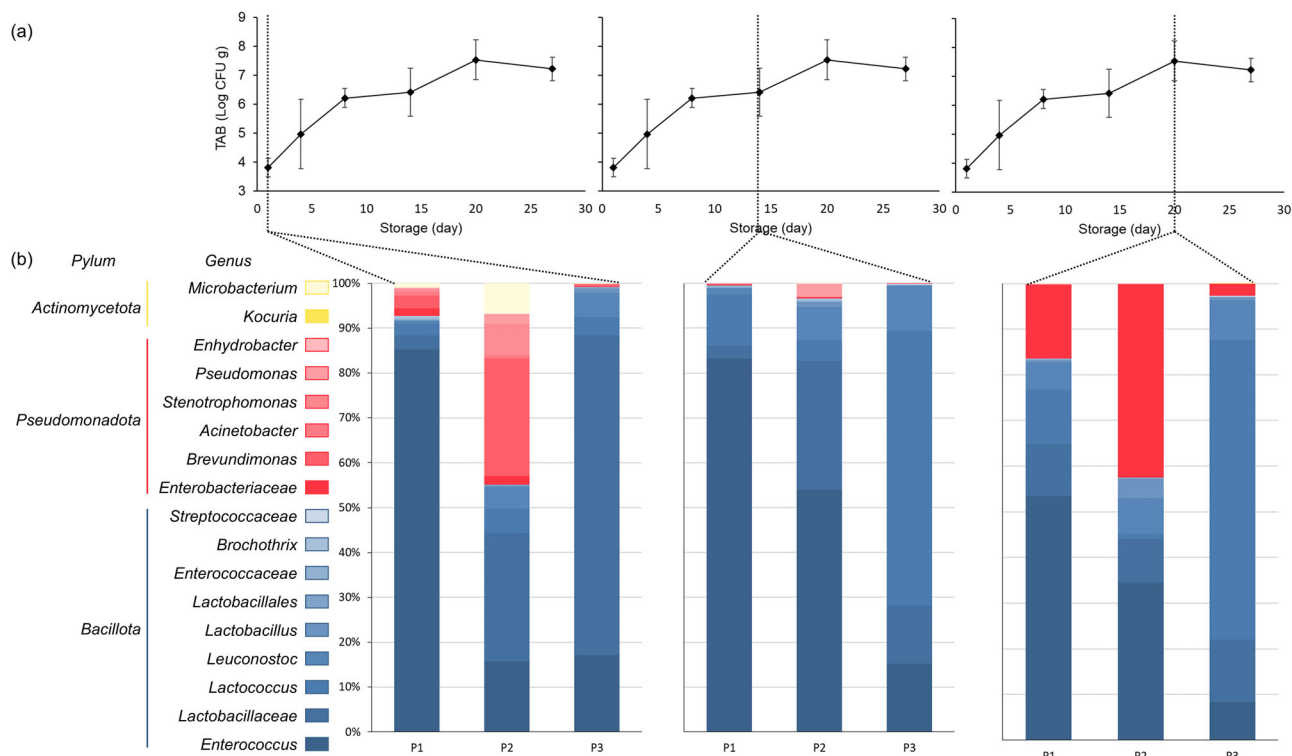
Therefore, each metabolite was influenced not only by the total bacterial count but also by the composition of specific microbes. Consequently, we also verified whether the metabolites in the drip could be used to predict specific microbial compositions.

$$y = 6.2400 + 0.0462(\text{ADP}) - 0.2203(\text{Carnosine}) - 0.2249(\text{Creatine}) + 0.0916(\text{Ethanol}) + 0.1236(\text{Fumarate}) + 0.1817(\text{Glutamine}) + 0.1891(\text{Glycerol}) + 0.1692(\text{Hypoxanthine}) - 0.2165(\text{IMP}) + 0.1741(\text{Malate}) - 0.1527(\text{O - acetylcarnitine}) + 0.1954(\text{Tyrosine}) + 0.1206(\text{Uracil}) \quad (1)$$

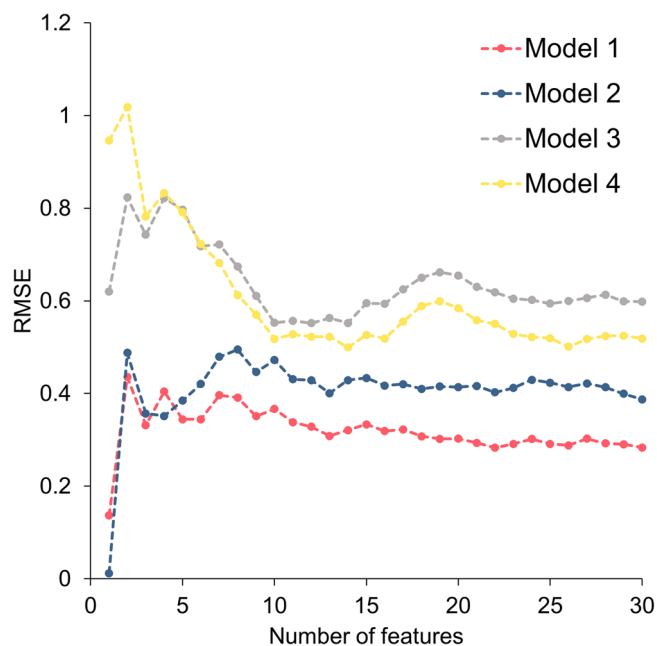
$$y = -0.0006 + 0.0800(\text{ADP}) - 0.2090(\text{Carnosine}) - 0.1729(\text{Creatine}) + 0.0150(\text{Ethanol}) + 0.1105(\text{Fumarate}) + 0.1270(\text{Glutamine}) + 0.1898(\text{Glycerol}) + 0.1692(\text{Hypoxanthine}) - 0.2190(\text{IMP}) + 0.0372(\text{Malate}) - 0.1589(\text{O - acetylcarnitine}) + 0.1605(\text{Tyrosine}) + 0.1014(\text{Uracil}) \quad (2)$$

$$y = 6.2400 + 0.1261(\text{Acetate}) - 0.2644(\text{Adenosine}) + 0.1813(\text{Alanine}) + 0.0920(\text{Asparagine}) + 0.2162(\text{Aspartate}) + 0.0434(\text{Cadaverine}) + 0.3270(\text{Hypoxanthine}) - 0.4480(\text{IMP}) + 0.0287(\text{Isoleucine}) - 0.3851(\text{Tyramine}) \quad (3)$$

$$y = -0.0008 + 0.1612(\text{Acetate}) + 0.0206(\text{Adenosine}) + 0.3128(\text{Alanine}) + 0.0366(\text{Asparagine}) + 0.4623(\text{Aspartate}) + 0.1616(\text{Cadaverine}) + 0.2302(\text{Hypoxanthine}) - 0.6273(\text{IMP}) - 0.5543(\text{Isoleucine}) - 0.5472(\text{Tyramine}) \quad (4)$$



**Fig. 3 | Microbial changes in pork during storage. a** Total aerobic bacterial (TAB) counts. **b** Microbial composition changes across different carcass repetitions (P1, P2, and P3).



**Fig. 4 | Results of RMSE value using LOOCV. The RMS values were calculated cumulatively as the number of features increased.** RMSE root mean square error, LOOCV leave-one-out cross-validation. Model 1: Ridge regression using metabolites selected by LASSO regression. Model 2: Support vector regression (SVR) using metabolites selected by LASSO regression. Model 3: Ridge regression using metabolites selected by Random Forest. Model 4: SVR using metabolites selected by Random Forest.

### Mathematical method for predicting microbial composition model

The model used for prediction employed elastic net regression, which combines LASSO and Ridge models<sup>15</sup>. Prior to modeling, metabolites were input as independent variables and microbial composition values as

**Table 1 | Microbial composition prediction modeling using net-elastic regression (N = 1800, n = 600)**

Phylum	Genus	Spearman ( $r_s$ )	RMSE
Actinomycetota	<i>Microbacterium</i>	0.962	0.1606
Actinomycetota	<i>Kocuria</i>	0.810	0.1403
Pseudomonadota	<i>Enhydrobacter</i>	0.906	0.1552
Pseudomonadota	<i>Pseudomonas</i>	0.959	0.1741
Pseudomonadota	<i>Stenotrophomonas</i>	0.962	0.1606
Pseudomonadota	<i>Acinetobacter</i>	0.959	0.1658
Pseudomonadota	<i>Brevundimonas</i>	0.978	0.1650
Pseudomonadota	<i>Enterobacteriaceae</i>	0.943	0.1797
Bacillota	<i>Streptococcaceae</i>	0.892	0.1824
Bacillota	<i>Brochothrix</i>	0.930	0.1703
Bacillota	<i>Enterococcaceae</i>	0.952	0.1838
Bacillota	<i>Lactobacillaceae</i>	0.940	0.1739
Bacillota	<i>Lactobacillus</i>	0.937	0.1786
Bacillota	<i>Leuconostoc</i>	0.968	0.1890
Bacillota	<i>Lactococcus</i>	0.920	0.1837
Bacillota	<i>Lactobacillales</i>	0.926	0.1735
Bacillota	<i>Enterococcus</i>	0.960	0.1741

RMSE Root Mean Squared Error.

N the total number of samples.

n the number of samples by storage date.

dependent variables in the elastic net regression, and the results are presented in Table 1. The model's fit was assessed using Spearman correlation ( $r_s$ ), with results showing values above 0.95 for the *Microbacterium*, *Pseudomonas*, *Stenotrophomonas*, *Acinetobacter*, *Brevundimonas*, *Enterococcaceae*, *Leuconostoc*, and *Enterococcus*. For *Enhydrobacter*, *Enterobacteriaceae*, *Brochothrix*, *Lactobacillaceae*, *Lactobacillus*,

*Lactococcus*, and *Lactobacillales*, the  $r_s$  values were confirmed to be above 0.9. Additionally,  $r_s$  values above 0.8 were observed for the remaining microbes. A Spearman correlation coefficient ( $r_s$ ) is interpreted as follows: absolute values between 0.3 and 0.5 suggest a weak to moderate correlation; values between 0.5 and 0.7 indicate a moderate to strong correlation; and values above 0.7 are typically considered to reflect a strong correlation<sup>38</sup>. Thus, the elastic net regression model is considered to have a high reliable results based on the high Spearman correlation observed. Consequently, to accurately assess the model's performance, we conducted model validation using the RMSE, which evaluates the similarity between the actual values and the predictions effectively (Table 1). The RMSE value is the square root of the mean squared error between predicted values and actual values<sup>39</sup>. Therefore, RMSE values distributed in the range of 0.15 to 0.19 indicate that the model, utilizing elastic net regression, has a high prediction accuracy. Consequently, the metabolites in the drip can also be used to predict the specific composition of microbes through mathematical modeling.

We conducted variables selection and modeling to predict TAB and microbial composition in pork using drip. For variables selection, we used LASSO regression and RF, while Ridge regression and SVR were utilized for the modeling phase. The combination of LASSO regression for variable selection followed by Ridge regression proved to be the most appropriate method for predicting TAB, as evidenced by an  $R^2$  value of 0.968 and an RMSE of 0.283. Additionally, integrated elastic net regression, which combines LASSO and Ridge regression, was effective for predicting microbial compositions. Therefore, the combination of LASSO and Ridge regression provides high accuracy, making it a valuable tool for prediction of meat freshness and microbial composition. Traditional methods, such as total bacterial count (TBC), typically require over 48 h and involve destructive sampling, leading to material loss and delayed results. In contrast, our method using drip metabolites can reduce the analysis time to just a few hours while maintaining non-destructive sampling, offering a more efficient and practical solution for meat freshness assessment. However, this study has limitations due to the small sample size, which may affect the statistical robustness and generalizability of the findings. While the model showed high accuracy through cross-validation, its performance in real-world industrial environments has not been fully validated. Factors such as sample variability, environmental conditions, and measurement inconsistencies could influence the model's robustness in practical applications. Further research is needed under various conditions, including packaging, storage, drip collection methods, and additional batch experiments, to assess the model's generalizability. Moreover, studies should distinguish between metabolite changes from microbial activity and those from natural aging. Thus, this study serves as a preliminary investigation into methods of measuring spoilage using drip metabolomic analysis and underscores the need for future research based on these preliminary results.

## Methods

### Sample preparation

Pork loin (*Longissimus thoracis*) from both sides of three crossbred ((Landrace × Yorkshire) × Duroc) pigs was acquired from three individual butcher shops in different locations. Each pig was slaughtered 2 days prior to purchase. After trimming both ends, each loin was sliced into 7 pieces, each 11 cm thick. The sliced samples were vacuum-packaged (HFV-600L; Hankook Fudge Machinery Co., Ltd., Hwaseong, Korea) in a polyethylene/nylon bag (oxygen permeability, 22.5 mL/m<sup>2</sup> per 24 h at an atmospheric pressure of 60% relative humidity and 25 °C; moisture vapor transmission rate, 4.7 g/m<sup>2</sup> per 24 h at 100% relative humidity and 25 °C). In total, 30 loin slices (5 days × 3 pigs × 2 sides) were incubated at 4 ± 2 °C and sampled on days 1, 8, 14, 20, and 27 for subsequent analyses.

### pH

This procedure was slightly modified based on the methodology described by Jung et al.<sup>40</sup>. Each 3 g sample was homogenated with 27 mL of distilled water using an Ultra-Turrax T25 homogenizer (Ika-Werke) at a speed of 9600 rpm for 30 s. Following this, the mixtures were centrifuged in a

Continent 512R centrifuge (Hanil) at 2265 × *g* for 10 min and filtered using Whatman No. 4 paper (Whatman plc, Maidstone, UK). The pH of the samples was measured using a Seven2Go pH meter and an InLab Expert Go-ISM pH probe (Mettler-Toledo International Inc., Schwerzenbach, Switzerland). The pH meter was first calibrated at room temperature using buffer solutions with pH values of 4.0, 7.0, and 9.21. The pH of sample was measured twice, with the average of these measurements taken as the final value for each sample.

### Drip loss

Assessing drip loss in pork was based on the method described by Choi et al.<sup>41</sup>. To calculate the drip loss, each piece of pork was initially weighed prior to packaging. Following a period of storage, the pieces were reweighed. The drip loss was then determined by calculating the difference in weight before and after storage using Eq. (4).

$$\text{Drip loss} = \frac{\text{original sample weight} - \text{sample weight after incubation}}{\text{original sample weight}} \times 100 \quad (5)$$

### Total aerobic bacterial counts

The total aerobic bacterial count (TAB) was conducted following the methods described by Ismail et al.<sup>42</sup>. Initially, 10 g of the sample were placed into a sterile bag containing 90 mL of a 0.85% saline solution. The mixture was then agitated using a BagMixer 400 P (Interscience Ind., St. Nom, France). This was followed by creating serial dilutions to secure measurable concentrations. Afterward, aliquots of 100 µL from these dilutions were spread on plate count agar (Difco Laboratories, Detroit, MI, USA) and were then incubated at 37 °C for 48 h. Post-incubation, the number of bacterial colonies was counted. The final results were recorded in log CFU/g.

### Metabolites

The NMR analysis of metabolites was carried out using the method described by Kim et al.<sup>4</sup>. Initially, 1 mL of pork drip collected from pork loin samples was mixed with 4 mL of 0.6 M perchloric acid and homogenized at 16,000 rpm for 1 min with a T25 Ultra homogenizer (Ika-Werke). This homogenate was then centrifuged at 3000 × *g* for 20 min at 4 °C in a Continent 512 R centrifuge (Hanil Co., Ltd., Incheon, Korea). The supernatant was poured into a new tube, and its pH was adjusted to 7.0 using KOH, measured by a SevenGo pH meter (Mettler-Toledo, Schwerzenbach, Switzerland), followed by another centrifugation under the same conditions. The supernatant was then filtered through Whatman No. 1 filter paper (Whatman plc) and freeze-dried for 4 h (Freezer Dryer 18, Labco Corp., Kansas City, USA). The dried sample was rehydrated with 1 mM 3-(trimethylsilyl)propionic acid-2,2,3,3-d<sub>4</sub> (TSP) in D<sub>2</sub>O, adjusted to pH 7.4 with a 20 mM phosphate buffer, then vortexed and incubated at 37 °C for 10 min. The samples centrifugation at 17,800 × *g* for 20 min in an HM-150IV centrifuge (Hanil Co., Ltd., Incheon, Korea), and the prepared samples were placed into 5 mL NMR tubes for analysis using a Bruker 850 MHz NMR spectrometer (Bruker Biospin GmbH, Baden-Wuttemberg, Germany). The analysis employed the zg30 pulse sequence in Bruker's Topspin 4.2.0 software, collecting data at 298 K. The entire process from metabolite extraction to NMR measurement took approximately 5 h.

The NMR spectra used 64 K data points, a sweep width of 17,007.803 Hz, and 128 scans with an acquisition time of 4.20 s. TSP resonance was used as a reference for chemical shift calibration in both qualitative and quantitative analyses. Manual baseline corrections and peak identifications were conducted using 2D NMR spectra techniques such as correlation spectroscopy (COSY) and heteronuclear single quantum coherence (HSQC). COSY was performed with 2 K data points in the  $t_2$  domain and 128 increments in the  $t_1$  domain at 16 scans, covering a spectral width of 11 ppm. HSQC utilized 2 K data points in the  $t_2$  domain and 256 increments in the  $t_1$  domain at 16 scans, with spectral widths of 223 ppm and 11 ppm for the  $f_1$  and  $f_2$  axes, respectively, with a coupling constant of 145 Hz to establish delay durations for short-range correlations. Peak identifications were referenced against the Human Metabolome Database

(HMDB). Quantification of the metabolites from  $^1\text{H}$  NMR spectra was processed using the Chenomx NMR Suite 10.0 (Chenomx, Inc., Edmonton, Alberta, Canada), with TSP as the internal standard and metabolite concentrations expressed in mg/dL, using five replicates for each sample.

### DNA extraction and sequencing of microorganisms in pork

The method for extracting DNA from microorganisms in pork was modified on the techniques described by Kim et al.<sup>4</sup>. To assess the changes in microbial composition during storage, 16S rRNA analysis was performed on samples collected on days 1, 14, and 20, representing the initial, mid-point, and late stages of the storage period, respectively. Five grams of pork mixed with 10 mL sterile saline (0.85%) solution was shaken on a gyratory shaker for 15 min. The mixture was centrifuged at  $200\times g$  for 5 min. The supernatant was collected in a new test tube, then are centrifuged at  $12,000 \times g$  at  $4^\circ\text{C}$  for 10 min. After the centrifugation, the supernatant was discarded, then the pellet was stored at  $70^\circ\text{C}$ .

DNA extraction, polymerase chain reaction (PCR) amplification, and amplicon preparation for sequencing were analyzed as described by Mun et al.<sup>43</sup>. Genomic DNA was extracted from the cell pellet using DNeasy PowerSoil Kit (Qiagen, Hilden, Germany). Primer was using the V3-4 region of the 16S rRNA gene: forward, TCGTCGGCAGCGTCAGATGTGATAAGAGACAGGTGCCAGCMGCCGCGGTAA-3'; reverse, 5' -GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT-3'. The DNA libraries were sequenced using the Illumina iSeq 100 platform (Illumina, San Diego, CA, USA). The fastq files obtained from the iSeq100 sequencing data were analyzed using Mothur (v.1.41). In Mothur, contigs were read combined using the make.contig instruction and filtered for quality using the screen.seqs command. The sequences were aligned using the SILVA database v.138 and removed chimeric sequences using the VSEARCH program v.2.11.1. Taxonomic classification was assessed using the Greengenes format database 14 and removed from the dataset after chloroplast, archaeal, mitochondrial, and eukaryotic sequences were removed from the dataset. The low abundance taxonomic units and singletons were removed using the Mothur subroutine "split. Abund", and a distance of 0.03 calculation (97% sequence similarity) was used to classify operational taxonomic units (OTUs). Sequence data used in this study were deposited to the Short Read Archives with the project number PRJNA1144167.

### Statistical analyses and Modeling

Pork quality values and quantified metabolites were statistically analyzed using analysis of variance (ANOVA) in SAS software (Version 9.4, SAS Institute Inc., Cary, NC, USA). Differences among the means for each storage date were assessed using Tukey's multiple comparison test. Statistical significance was set at  $p < 0.05$ .

The analysis of metabolomics data in this study was conducted using the R programming language and several packages. The modeling process is illustrated in Fig. 1. During the variable selection process, the glmnet package was utilized to implement LASSO regression. LASSO applies L1 regularization to drive the coefficients of less important variables to zero, effectively selecting variables<sup>18</sup>. RF analysis was performed using the randomForest package, where each decision tree independently learned from bootstrap samples and evaluated the importance of variables to select significant ones. RF selected metabolites with an importance score of 10 or higher after 10 iterations. Each set of selected metabolites from LASSO and RF was subsequently used for modeling. For modeling with the selected variables, the Ridge regression was implemented using the ridge package, and the SVR was implemented using the e1071 package. The Ridge regression applied L2 regularization to reduce the coefficients of all variables, and a linear kernel was used for the SVR. The performance of the model was evaluated using the coefficient of determination ( $R^2$ ) and the root mean square error (RMSE).  $R^2$  indicates how well a model explains the variability in the data, with values ranging from 0 to 1; values closer to 1 signify a better model fit<sup>44</sup>. RMSE is calculated by taking the square

root of the average of the squared differences between the predicted and actual values, where lower values indicate higher predictive accuracy<sup>39</sup>. Leave-one-out cross-validation (LOOCV) was used to calculate the RMSE for assessing the model's prediction accuracy.

Elastic net regression was applied to analyze the interactions between metabolomics data and various microbial compositions using the glmnet package<sup>45</sup>. The data used in the model underwent noise data augmentation using the dplyr and readr packages<sup>46</sup>. A total of 600 data points were generated for each storage period (day 1, day 14 and day 20) using Monet Carlo simulation<sup>5</sup>, resulting in a total of 1800 data points ( $n = 600$ ,  $N = 1800$ ) from the original data ( $n = 6$ ,  $N = 18$ ). This method combines LASSO and Ridge regularization to perform variable selection and coefficient shrinkage simultaneously, thereby enhancing the predictive power and interpretability of the model. The performance of the implemented model was evaluated using the Spearman's correlation coefficient, which indicates how well the model explains the data. Additionally, to validate the model's efficacy, RMSE was employed.

### Data availability

All relevant data supporting this study is included within the manuscript and supplementary data. Additional raw data will be made available from the corresponding author upon reasonable request.

Received: 6 November 2024; Accepted: 10 April 2025;

Published online: 23 April 2025

### References

- Bassey, A. P. et al. Assessment of quality characteristics and bacterial community of modified atmosphere packaged chilled pork loins using 16S rRNA amplicon sequencing analysis. *Food Res. Int.* **145**, 110412 (2021).
- Paulsen, P., Schopf, E. & Smulders, F. J. M. Enumeration of total aerobic bacteria and *Escherichia coli* in minced meat and on carcass surface samples with an automated most-probable-number method compared with colony count protocols. *J. Food Prot.* **69**, 2500–2503 (2006).
- Gu, M. et al. Novel insights from protein degradation: deciphering the dynamic evolution of biogenic amines as a quality indicator in pork during storage. *Food Res. Int.* **167**, 112684 (2023).
- Kim, H. J. et al. Mathematical modeling for freshness/spoilage of chicken breast using chemometric analysis. *Curr. Res. Food Sci.* **7**, 100590 (2023).
- Kim, H. J. et al. A non-destructive predictive model for estimating the freshness/spoilage of packaged chicken meat using changes in drip metabolites. *Int. J. Food Microbiol.* **419**, 110738 (2024).
- Zuo, J. et al. Integrating transfer learning and spectroscopy for enhanced pork spoilage assessment using correlation analysis. *Food Chem.* **465**, 142117 (2025).
- Tahir, H. E. et al. Smart films fabricated from natural pigments for measurement of total volatile basic nitrogen (TVB-N) content of meat for freshness evaluation: a systematic review. *Food Chem.* **396**, 113674 (2022).
- Peyvasteh, M. et al. Meat freshness revealed by visible to near-infrared spectroscopy and principal component analysis. *J. Phys. Commun.* **4**, 095011 (2020).
- Zuo, J. et al. Nondestructive intelligent detection of total viable count in pork based on miniaturized spectral sensor. *Food Res. Int.* **197**, 115184 (2024).
- Huff-Lonergan, E. & Lonergan, S. M. Mechanisms of water-holding capacity of meat: The role of postmortem biochemical and structural changes. *Meat Sci.* **71**, 194–204 (2005).
- Im, C. et al. Assessing /uality. *Food Sci. Anim. Resour.* **44**, 758 (2024).
- Setyabrata, D. et al. Proteomics and metabolomics profiling of meat exudate to determine the impact of postmortem aging on oxidative stability of beef muscles. *Food Chem. X* **18**, 100660 (2023).

13. Gänzle, M. G. Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage. *Curr. Opin. Food Sci.* **2**, 106–117 (2015).
14. Kim, H. C. et al. Using 2D qNMR analysis to distinguish between frozen and frozen/thawed chicken meat and evaluate freshness. *npj Sci. Food*. <https://doi.org/10.1038/s41538-022-00159-x> (2022).
15. Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc.* <https://doi.org/10.1186/1753-6561-6-S2-S10> (2012).
16. Luo, X., Chang, X. & Ban, X. Regression and classification using extreme learning machine based on L1-norm and L2-norm. *Neurocomputing* **174**, 179–186 (2016).
17. Kocev, D., Vens, C., Struyf, J. & Džeroski, S. Tree ensembles for predicting structured outputs. *Pattern Recognit.* **46**, 817–833 (2013).
18. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Method.* **58**, 267–288 (1996).
19. Salcedo-Sanz, S., Rojo-Álvarez, J. L., Martínez-Ramón, M. & Camps-Valls, G. Support vector machines in engineering: an overview. *Data Min. Knowl. Discov.* **4**, 234–267 (2014).
20. NPPC Pork Quality Solutions Team. *Pork Quality Targets*. *Pork Information Gateway* (2006).
21. Knox, B. L., Van Laack, R. L. J. M. & Davidson, P. M. Relationships between ultimate pH and microbial, chemical, and physical characteristics of vacuum-packaged pork loins. *J. Food Sci.* **73**, M104–M110 (2008).
22. Chai, X. et al. Impact of packaging methods coupled with high barrier packaging loaded with TiO<sub>2</sub> on the preservation of chilled pork. *J. Food Sci.* **44**, 1142 (2024).
23. Chen, L., Zhou, G. H. & Zhang, W. G. Effects of high oxygen packaging on tenderness and water holding capacity of pork through protein oxidation. *Food Bioprocess. Technol.* **8**, 2287–2297 (2015).
24. Herrero, A. M. et al. Raman spectroscopy study of the structural effect of microbial transglutaminase on meat systems and its relationship with textural characteristics. *Food Chem.* **109**, 25–32 (2008).
25. Zhao, F. et al. Microbial changes in vacuum-packed chilled pork during storage. *Meat Sci.* **100**, 145–149 (2015).
26. Hemme, D. & Foucaud-Scheunemann, C. *Leuconostoc*, characteristics, use in dairy technology and prospects in functional foods. *Int. Dairy J.* **14**, 467–494 (2004).
27. Rivas, B. L., Pereira, E. D., Palencia, M. & Sánchez, J. Water-soluble functional polymers in conjunction with membranes to remove pollutant ions from aqueous solutions. *Prog. Polym. Sci.* **36**, 294–322 (2011).
28. Degenhardt, F., Seifert, S. & Szymczak, S. Evaluation of variable selection methods for random forests and omics data sets. *Brief. Bioinform.* **20**, 492–503 (2019).
29. Wang, Z. et al. Lasso-Logistic regression model for the identification of serum biomarkers of neurotoxicity induced by strychnos alkaloids. *Toxicol. Mech. Methods* **33**, 65–72 (2023).
30. Zhang, P. et al. Machine learning applied to serum and cerebrospinal fluid metabolomes revealed altered arginine metabolism in neonatal sepsis with meningoencephalitis. *Comp. Struct. Biotechnol. J.* **19**, 3284–3292 (2021).
31. Krstic, N. et al. The impact of methodological choices when developing predictive models using urinary metabolite data. *Stat. Med.* **41**, 3511–3526 (2022).
32. Xu, L., Liu, S., Cheng, Y. & Qian, H. The effect of aging on beef taste, aroma and texture, and the role of microorganisms: a review. *Crit. Rev. Food Sci. Nutr.* **63**, 2129–2140 (2021).
33. Swart, R. M. et al. Fumarate production with *Rhizopus oryzae*: utilising the Crabtree effect to minimise ethanol by-product formation. *Biotechnol. Biofuels* **13**, 22 (2020).
34. Jiang, Y. et al. Calcium (Ca<sup>2+</sup>)-regulated exopolysaccharide biosynthesis in probiotic *Lactobacillus plantarum* K25 as analyzed by an omics approach. *J. Dairy Sci.* **104**, 2693–2708 (2021).
35. Xiao, K., Ghalei, H. & Khoshnevis, S. RNA structural probing of guanine and uracil nucleotides in yeast. *PLoS ONE* **18**, e0288070 (2023).
36. Xiao, K. et al. Anaerobic digestion of sludge by different pretreatments: changes of amino acids and microbial community. *Front. Env. Sci. Eng.* <https://doi.org/10.1007/s11783-021-1458-7> (2022).
37. Laus, F. et al. Donkey colostrum and milk: how dietary probiotics can affect metabolomic profile, alkaline sphingomyelinase and alkaline phosphatase activity. *Metabolites* **13**, 622 (2023).
38. Akoglu, H. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **18**, 91–93 (2018).
39. Sun, H. et al. Comparative analysis of pork tenderness prediction using different optical scattering parameters. *J. Food Eng.* <https://doi.org/10.1016/j.jfoodeng.2018.12.006> (2019).
40. Jung, H. Y. et al. Exploring effects of organic selenium supplementation on pork loin: Se content, meat quality, antioxidant capacity, and metabolomic profiling during storage. *J. Anim. Sci. Technol.* **66**, 587 (2024).
41. Choi, M. et al. The impact of overnight lairage on meat quality and storage stability of pork loin. *J. Anim. Sci. Technol.* **66**, 412 (2024).
42. Ismail, A. et al. Quality evaluation of mackerel fillets stored under different conditions by hyperspectral imaging analysis. *Food Sci. Anim. Resour.* **43**, 840 (2023).
43. Mun, D. et al. Alleviation of DSS-induced colitis via bovine colostrum-derived extracellular vesicles with microRNA let-7a-5p is mediated by regulating Akkermansia and β-hydroxybutyrate in gut environments. *Microbiol. Spectr.* **11**, e00121–e00123 (2023).
44. Saunders, L. J., Russell, R. A. & Crabb, D. P. The coefficient of determination: what determines a useful R<sup>2</sup> statistic?. *Invest. Ophthalmol. Vis. Sci.* **53**, 6830–6832 (2012).
45. Chin, N. et al. Cytomegalovirus infection disrupts the influence of short-chain fatty acid producers on Treg/Th17 balance. *Microbiome* **10**, 168 (2022).
46. Moreno-Barea, F. J., Franco, L., Elizondo, D. & Grootveld, M. Application of data augmentation techniques towards metabolomics. *Comput. Biol. Med.* **148**, 105916 (2022).

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant-funded by the Korea government (MIST) under grant numbers 2022R1A6A3A01085938 and RS-2024-00453016.

## Author contributions

H.J.K. (Hyun-Jun Kim) Conceptualization, methodology, formal analysis, software, data curation, visualization, writing-original draft, writing—review & editing, funding acquisition; H.J.K. (Hye-Jin Kim) validation, data curation, writing-original draft, writing—review & editing, funding acquisition; H.H. data curation, methodology; M.C. data curation, methodology, formal analysis, investigation; A.I. data curation, methodology, formal analysis, investigation; D.M. methodology, formal analysis, investigation; Y.K. methodology, investigation; G.D.K. methodology, investigation; C.J. conceptualization, writing—review & editing, supervision, project administration, funding acquisition.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41538-025-00421-y>.

**Correspondence** and requests for materials should be addressed to Cheorun Jo.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2025