Supplementary Data

Materials & Methods

Animal model of skin inflammation

Phorbol 12-myristate 13-acetate (PMA) was dissolved in dimethyl sulfoxide (DMSO) at a concentration of 0.3 μg/μL, and 10 μL (3 μg) of dissolved PMA was topically applied to the surface of the dorsal side of mouse ears using a micropipette. The ears treated with PMA were collected after 6, 12, and 24 hours and subjected to the experiments as PMA-6, PMA-12, and PMA-24 groups, respectively. Ears not treated with PMA were also used as the control group. Ear samples were placed on a 0.5-mm thick quartz-covered glass slide and set on the stage of the Raman microscope.

Raman spectroscopy and spectrum analysis

A schematic diagram of the Raman microscope is shown in Figure S1. The Raman apparatus MicroRam-TE-US (Lambda Vision Inc., Sagamihara, Japan) containing a 785-nm continuous wave laser (power: 300 mW, linewidth: <0.2 nm), a dichroic mirror (LPD02-785U-25, Semrock, Illinois, the US), and a polychromator (spectral resolution: 0.4 cm^{-1}/ch) with a cooled CCD detector (2,048 ch, 300,000 pixels), was attached to an inverted microscope Eclipse TE300 with a Plan Fluor ELWD 40x (NA 0.60, WD 3.7-2.7)

Nikon objective lens. The Raman apparatus was controlled using RamanScope ver.2.62 (Lambda Vision Inc., Sagamihara, Japan). To exclude spatial heterogeneity of the Raman spectrum, spectroscopic analysis was performed at 20 points in each ear sample, five points at 1-mm intervals in the craniocaudal direction, and four points at 1-mm intervals in the dorsoventral direction (Fig. S2A *Left*). At each measurement point, after focusing the laser on the glass surface, Raman spectra were recorded at different depths by moving the objective lens to shift focus: 15- and 50-μm deep in the case of the control group, and 15-, 50-, and 100-μm deep in the case of PMA-6, PMA-12, and PMA-24 groups, respectively. To improve the signal-to-noise ratio, the spectral data obtained by one-second laser irradiation were integrated 10 times.

Spectrum data were exported in csv file format. To remove the spectrum due to the mechanical noise of the instrument, background data were subtracted from the raw sample data. The subtracted data were subjected to principal component analysis (PCA) and cluster analysis based on the extracted principal components (PCs), which was performed using a dendrogram and heatmap. To evaluate the accuracy in predicting skin inflammation based on several spectra, the control group was regarded as the normal group, while the PMA groups were regarded as the inflammation group. ROC curves were created to evaluate the sensitivity and specificity in these groups.

Histological analysis

After Raman spectroscopy, ear samples were fixed with 4% paraformaldehyde to prepare the pathological specimens. The fixed ear samples were divided into five pieces along four measurement points aligned in the dorsoventral direction (dashed lines in Fig. S2A *Left*). Tissue sections were prepared from the cut surface of each divided tissue sample, and haematoxylin and eosin (HE) staining was performed for histological assessment (Fig. S2A *Right*). Histological evaluation was carried out at four points in each section, corresponding to the laser-irradiated sites (Fig. S2A, *right*, red dotted circles). The thicknesses of the epidermis and dermis were measured. Infiltration of inflammatory cells, vascularization or vasodilation, and interstitial oedema were rated on a 4-point scale for histological scoring (Fig. S2B & S2C).

Artificial intelligence analysis of Raman spectrum

Raman spectroscopic data measured for 63 mice (120 ear samples) in preliminary experiments were added to the AI analysis. The additional data were not included in the PCA of the main experiment because of differences in measurement methods, such as measurements at 20 points over the entire ear (Figure S2) and at two depths (15, 50 μm)

in the main experiment, whereas the preliminary experiment was conducted at two points (the base and apex of the ear) in 10-μm increments from 0 μm to 250 μm. However, since there were many similarities in the spectral changes, we used the data to increase the efficiency of machine learning.

The original AI analysis program was constructed in the following three steps: pre-processing, creation of learning models, and improvement of prediction accuracy. The flowchart of the AI analysis is shown in Figure S3. For pre-processing, all data were divided into two groups per mouse: learning data for machine learning and test data for prediction. All additional data from preliminary experiments were included in the learning dataset. The main data, which were also used for manual PCA as described in the main text, were divided into learning and test data in three different ways, and the final number of datapoints is shown in Table S3. This process is not necessary for machine learning performed with a sufficient amount of data. However, due to the small amount of data in this study, the training model and the verification of prediction accuracy using test data were performed independently using three different combinations of data (Table S3), and finally their prediction accuracies were integrated. Three different learning models for AI analysis were prepared for each data set: (a) One was based on three wavenumber regions (1,500–1700 cm^{-1}, 2,800–3,000 cm^{-1}, and 3,000-3,250 cm^{-1})

extracted from the raw data, (b) one included the 64 PCs with the highest contribution obtained from the PCA over the entire wavenumber region, and (c) and one included the 64 PCs with the highest contribution in the PCA of the above three wavenumber regions (i.e., 64 PCs x 3 regions = 192 PCs in total). These processes reduced the dimensions in each analysis and allowed us to analyse data trends that were effective in discriminating normal or inflammatory skin from a large number of explanatory variables.

Additionally, in the final step of the analysis, a more accurate determination of normality or inflammation was made by comparing the percentages derived from the three models. The learning data were further randomly divided into 80% training data and 20% validation data, and the learning model was designed to repeat machine learning using the training data and to evaluate prediction accuracy using the validation data 100 times. To compare predictive accuracy among the models, we first assessed various AI algorithms (e.g., support vector machine, multi-layer perceptron, convolutional neural network, and recurrent neural network). The multilayer perceptron method was determined to have the highest learning efficiency and was therefore adopted for the current study. Finally, batch normalisation, dropout, and ensemble learning were used to improve the prediction accuracy for normal and inflammatory conditions.

The above algorithm was independently adapted to the three patterns of data

combinations shown in Table S3 to predict normal skin and inflammation for the test data. These three results were integrated to produce the final ROC curve, following which the prediction accuracy was calculated (Figure 6).

Results

Principal component analysis of Raman spectra by depth

The data set for another PCA was arranged according to the thickness of the epidermis and dermis (Fig. 1C). The 15-μm data for the control and PMA-6 groups and the 50-μm data for the PMA-12 and PMA-24 groups were arranged as epidermis data. In addition, the 50-μm data of the control and PMA-6 groups and 100-μm data of the PMA-12 and PMA-24 groups were integrated as dermis data. Using these rearranged datasets, we conducted the same PCA shown in Figure 3 (Fig. S3). Although we observed slight differences, the results were similar: PC1 to PC4 mainly contributed to the spectral waveform (Fig. S3A and S3B), and the control group and the three groups treated with PMA could be roughly classified by the cluster analysis based on the coefficients of PC1 to PC4 (Fig. S3C and S3D). The control group could be separated from the PMA-12 and PMA-24 groups based on the distribution in the two-dimensional scatter plot for PC2 and PC4, while the PMA-6 group was located between them (Fig. S3E & S3F). The results

were relatively consistent with the results in the analysis according to depth, as shown in Figure 3.

Spatial distribution of Raman score

The inflammation score was estimated based on the heatmap shown in Figure 3. Each spectrum was scored on a four-point scale based on cluster analysis using a dendrogram for PC1 to PC4. The scores were then assigned to a light green to magenta colour along a gradient from low to high and displayed on the left edge of each heatmap (Fig. 3). The colours of each score were assigned to 4 × 5-measurement points to create a colour map (Fig. S4). Although there was spatial heterogeneity, lower scores (light green colouring) tended to be dominant in the control group, while higher scores (light or dark magenta colouring) that increased over time tended to be dominant in the PMA groups.

Prediction of inflammation based on Raman spectra and their principal components

ROC curves were created to evaluate the ability to predict skin inflammation based on the Raman spectra. Here, the control group was regarded as the normal group, and the three groups treated with PMA were regarded as the inflammation group. An intensity ratio of 1,550 cm^{-1}/1,450 cm^{-1} (Fig. S5 *left* panels), an intensity ratio of 2,930 cm^{-1}/2,860

cm^{-1} (Fig. S5 *middle* panels), and an intensity at 3,190 cm^{-1} (Fig. S5 *right* panels) were used for evaluation of Raman spectra. Among these indicators, the area under the ROC curve (AUC) for the intensity ratio of 1,550 cm^{-1}/1,450 cm^{-1} at 50 μm was high at 0.815. When the cut-off value was set to 0.446, the sensitivity and specificity were 75.0% and 85.6%, respectively. The ROC curve for the intensity ratio of 1,550 cm^{-1}/1,450 cm^{-1} at 15 μm also exhibited a high AUC value of 0.760 and a sensitivity of 89.8%, while the specificity was as low as 47.8%. On the other hand, predictions based on the intensity ratio of 2,930 cm^{-1}/2,860 cm^{-1} and intensity at 3,190 cm^{-1} were not satisfactory.

Depth limit for acquiring Raman-scattered light using a near-infrared laser

A mouse ear sample was set close to the top of a quartz glass field, and the laser was focused on the top surface of the quartz glass. The objective lens was then raised using a micrometre to focus the laser at determined depths (15, 50, 100, 150, 200, and 250 μm) in the tissue, following which the Raman-scattered light was recorded. Although the intensity of Raman-scattered light became weaker as the depth increased, and it became difficult to confirm the spectra at 1,600 cm^{-1} and 2,800–3,000 cm^{-1}, it was possible to record Raman-scattered light to some extent down to a depth of 250 μm (Fig. S6).

Table S1. Statistical information for the principal component analysis of Raman spectroscopic data at a depth of 15 µm

	PC1	PC2	PC3	PC4
Standard deviation	**24.171**	**2.785**	**1.740**	**0.951**
Variance (%)	**97.374**	**1.292**	**0.504**	**0.151**
Cumulative (%)	**97.374**	**98.666**	**99.170**	**99.321**

Table S2. Statistical information for the principal component analysis of Raman spectroscopic data at a depth of 50 µm

	PC1	PC2	PC3	PC4
Standard deviation	**24.455**	**3.577**	**1.675**	**1.107**
Variance (%)	**96.459**	**2.064**	**0.452**	**0.198**
Cumulative (%)	**96.459**	**98.523**	**98.975**	**99.173**

Table S3. Number of datapoints used for AI analysis

	1st Session		2nd Session		3rd Session	
	Normal	Inflammation	Normal	Inflammation	Normal	Inflammation
Machine learning	1,399	1,967	1,420	1,928	1,293	2,036
prediction	183	540	162	579	289	471
Total	1,582	2,507	1,582	2,507	1,582	2,507

Figures & Legends

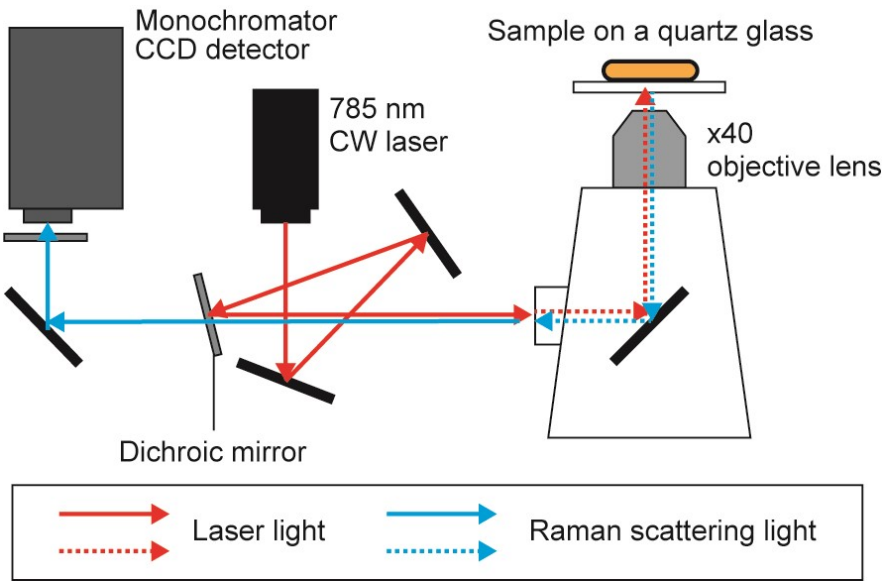

Figure S1. ***Schematic Diagram of the Raman microscope.*** A Raman apparatus equipped with a 785-nm continuous wave laser, dichroic mirror, and monochromator was connected to an inverted microscope with a 40 × objective lens. The sample was mounted on quartz-covered glass and set on the microscopic stage. The laser was used to irradiate the sample (red arrow), and the generated Raman-scattered light was guided to a monochromator through a dichroic mirror and subjected to spectroscopic analysis (light blue arrow). CCD, charge coupled device; CW, continuous wave.

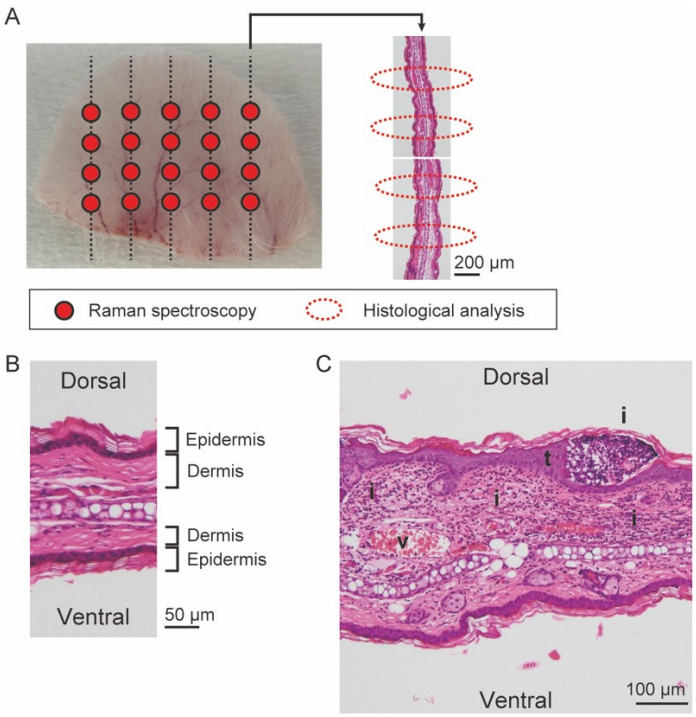

Figure S2. ***Preparation of mouse ear samples and structure of normal or inflamed skin in mouse ears. A:*** Raman spectroscopy was performed at 20 points (red dots), five points in a craniocaudal direction, and four points in a dorsoventral direction for each ear sample (*Left*). The dotted lines indicate the line dividing the ear sample when making the tissue section. A tissue specimen was prepared from each section, and a haematoxylin-and-eosin-stained pathological tissue sample was prepared (*Right*). Histological assessment was performed at four points on each section (red dotted circle) according to the laser-irradiated sites. ***B*****:** In the normal mouse ear, the epidermis and dermis form a layered structure with cartilage in between. ***C:*** In the inflamed mouse ear, vasodilation (v), infiltration of inflammatory cells into the epidermis and dermis (i), and thickening of the epidermis (t) were observed.

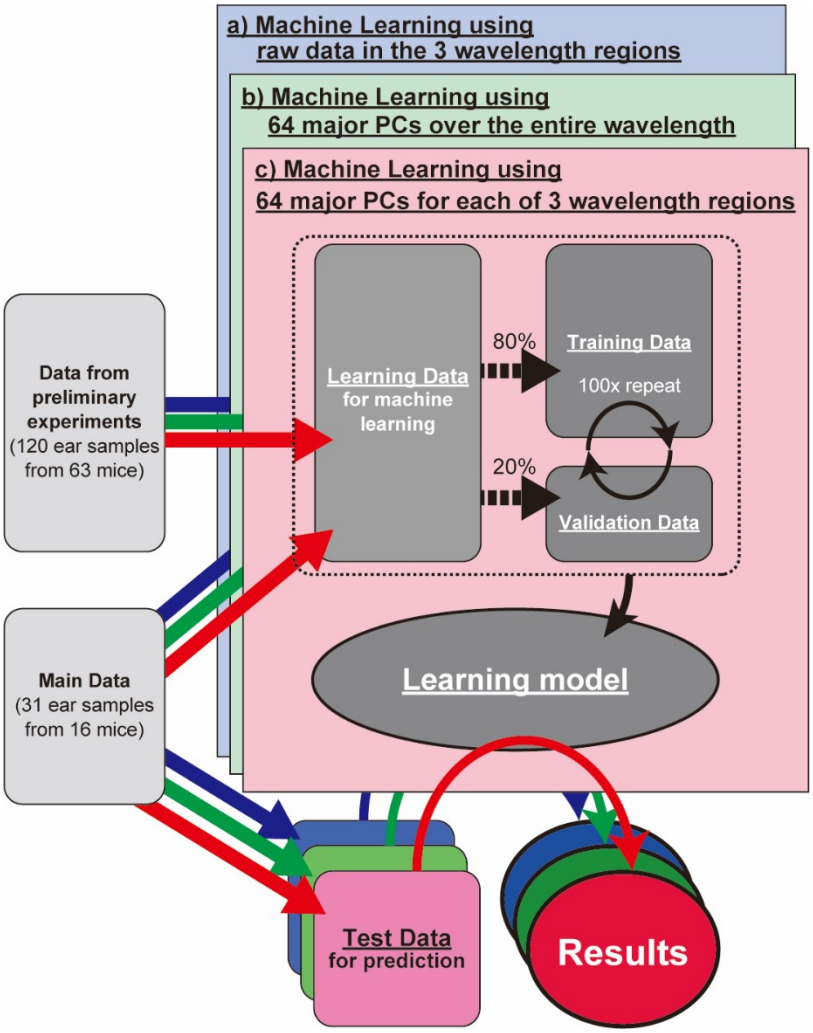

Figure S3. ***Scheme of data processing and AI analysis.*** The main data from 16 mice were divided into training and test data in three ways, and all data from preliminary experiments were included in the learning data (final data numbers are shown in Table S3). All data were pre-processed in three different ways: (a) raw data (blue arrows) in three wavenumber regions (1,500–1,700 cm^{-1}, 2,800-3,000 cm^{-1}, and 3,000-3,250 cm^{-1}), (b) 64 PCs (green arrows) obtained from PCA in all wavenumber regions, and (c) 64 PCs obtained from PCA in three wavenumber regions (i.e., 64 PCs x 3 regions = 192 PCs in

total, red arrow). Then, different learning models (blue box, green box, and red box) were created using each pre-processed dataset. The learning model was further constructed by randomly dividing the learning data into 80% training data and 20% validation data, machine learning was repeated with training data using a multilayer perceptron algorithm and validating using the validation dataset 100 times. Finally, the test data were assessed using the completed training model (blue, green, and red curved arrows), and ROC curves were produced to determine the prediction accuracy. PCA, principal component analysis; PC, principal components.

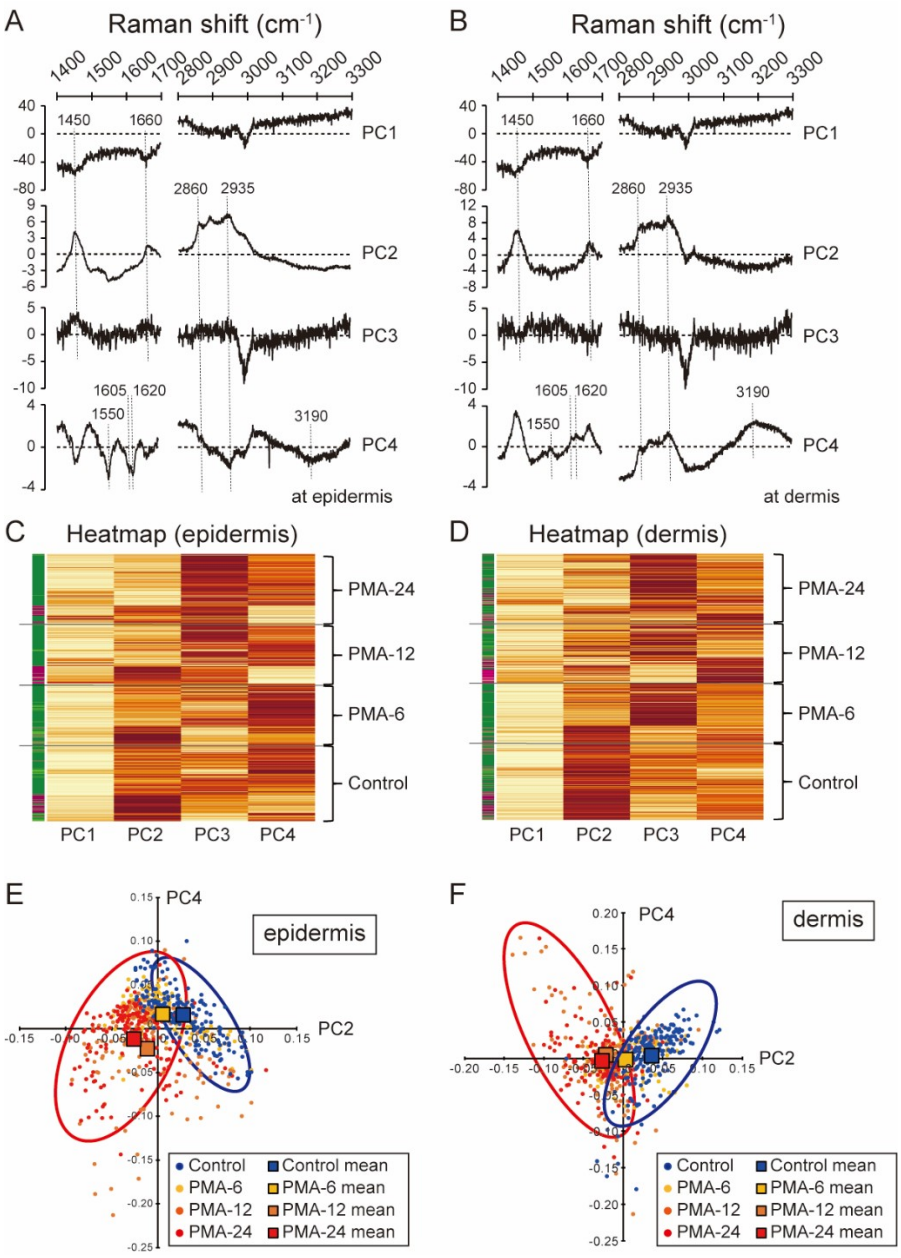

Figure S4. ***Results of principal component analysis and cluster analysis according to tissue structure.*** PCs were extracted from Raman spectra obtained from the ear samples at epidermal (***A***) and dermal (***B***) depths. PC1 to PC4 are shown as major PCs. Vertical dotted lines indicate spectral peaks that could be assigned to molecular bindings. A cluster analysis using a heatmap was then applied to PC1 to PC4 at the epidermis (***C***) and dermis

(***D***). Each spectrum was scored along a four-point scale based on a cluster analysis using a dendrogram of PC1 to PC4, and the scores were assigned to a light green to magenta colour, representing the gradient from low to high, and displayed on the left edge of the heatmap. The contributions of PC2 and PC4 to the spectrum at the level of the epidermis (***E***) and dermis (***F***) are shown in the two-dimensional scatter plot. Blue, yellow, orange, and red dots indicate the plot of each spectral data from control, PMA-6, PMA-12, and PMA-24 groups, respectively. Blue, yellow, orange, and red boxes indicate the average of the contribution value for control, PMA-6, PMA-12, and PMA-24 data, respectively. The blue and red circles indicate the areas in which data for the control and PMA-treated groups are mainly distributed, respectively. PMA, phorbol 12-myristate 13-acetate; ROC, receiver operating characteristic; PC, principal components.

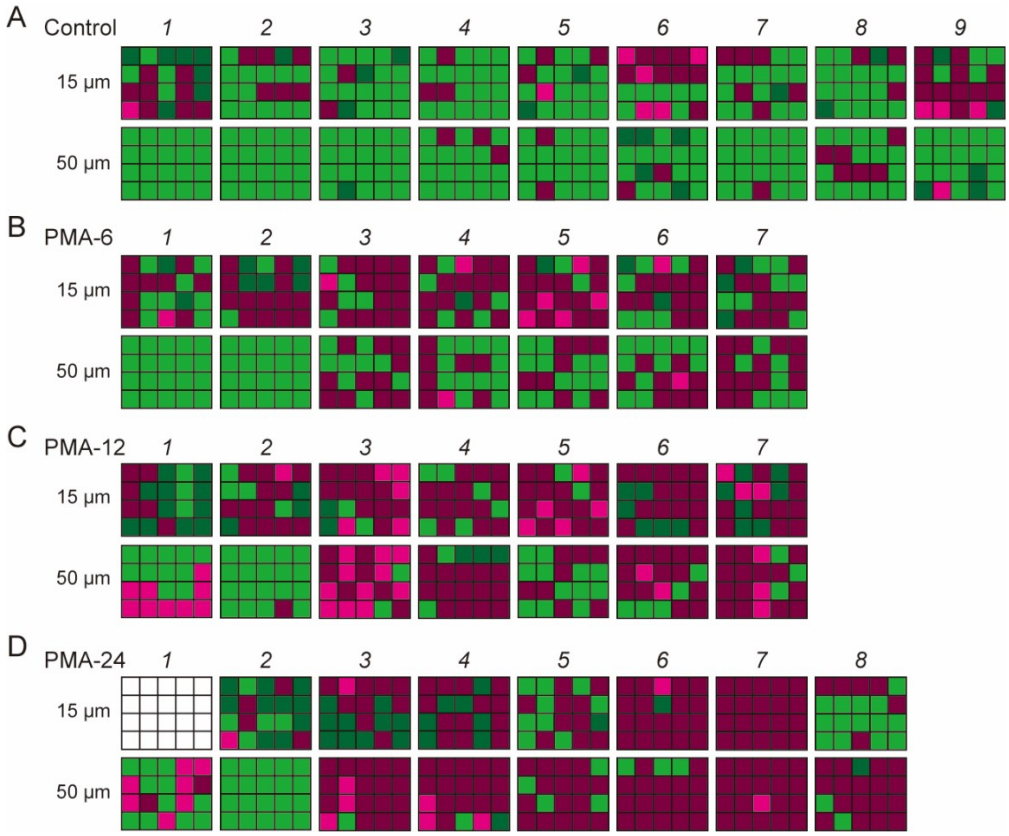

Figure S5. ***Colour mapping of 20 points at each depth in each ear based on Raman score.*** A four-level colour gradient from green to magenta was assigned to the score based on the principal component analysis of the Raman spectrum, and the score was mapped to 20 measurement points. Colour maps of Raman scores obtained from all samples of control group (***A***), PMA-6 group (***B***), PMA-12 group (***C***), and PMA-24 group (***D***) at 15 µm (top row) and 50 µm (bottom row) are shown. The top number indicates the ear number assigned to each group. PMA, phorbol 12-myristate 13-acetate.

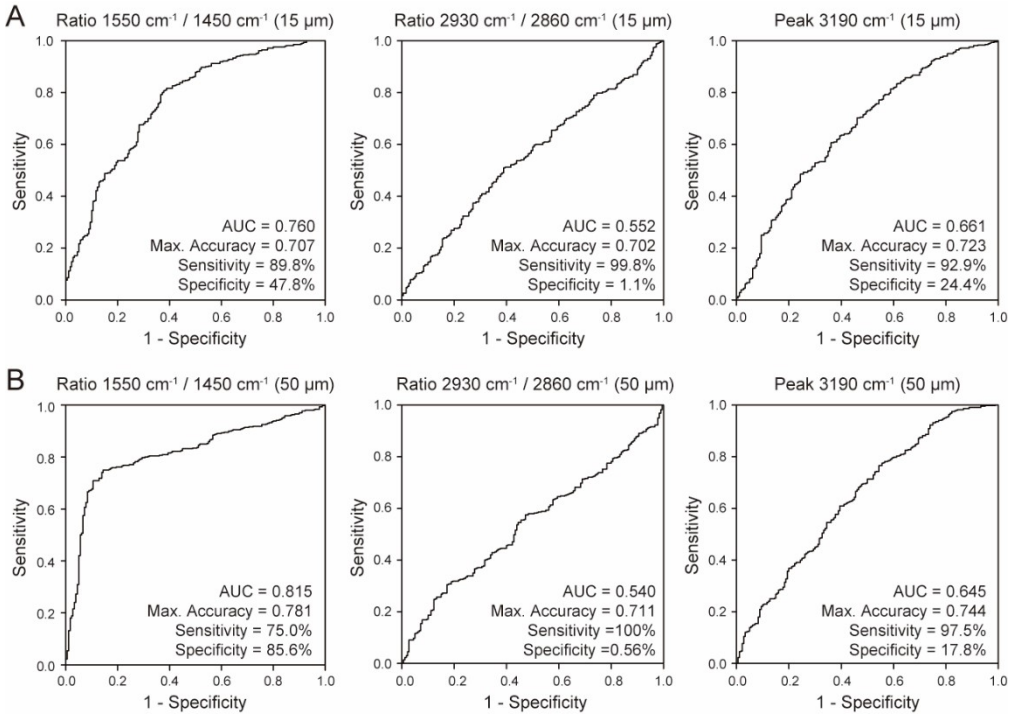

Figure S6. ***ROC curves for the prediction of inflammation based on principal components.*** The control group was regarded as the normal group, and the group treated with PMA for 24 hours was regarded as the inflammation group. ROC curves were created to assess the sensitivity and specificity in predicting the normal and inflammation groups according to representative Raman spectra obtained at depths of 15 μm (***A***) and 50 μm (***B***). PC1 (*Left panels*), PC2 (*Left Middle panels*), PC3 (*Right Middle panels*), and PC4 (*Right panels*) were used to generate ROC curves. The area under the curve (AUC), maximum accuracy, and sensitivity and specificity at the maximum accuracy are shown in the lower right inset of each graph. PMA, phorbol 12-myristate 13-acetate; ROC, receiver operating characteristic; PC, principal components.

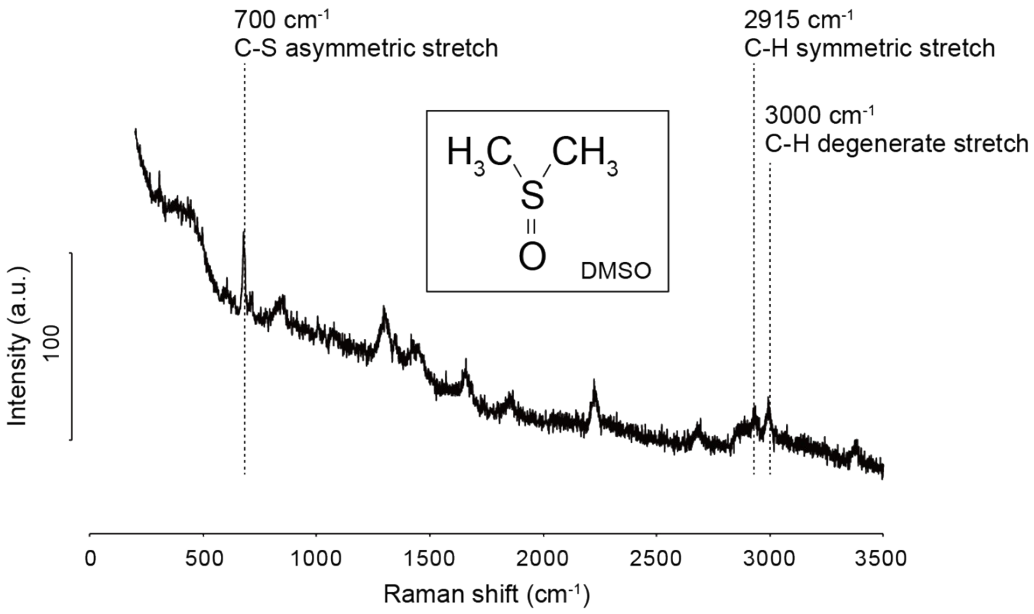

Figure S7. ***Representative Raman spectra including peaks derived from DMSO, obtained from a PMA-6 group mouse.*** The typical peak derived from the C-S and C-H bonds of DMSO was observed in this mouse. DMSO, dimethyl sulfoxide; PMA, phorbol 12-myristate 13-acetate.

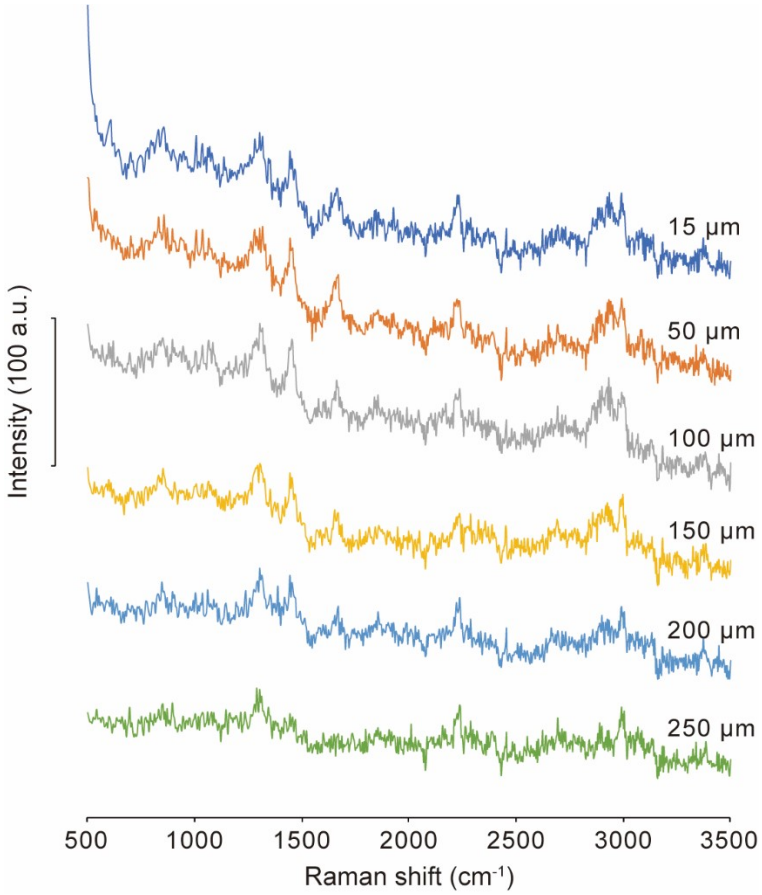

Figure S8. ***Raman spectra recorded at the same location at different depths.*** Raman spectra were recorded at the same location while moving the objective lens to change the measurement depth to 15, 50, 100, 150, 200, and 250 µm (blue, brown, grey, orange, sky blue, and green lines, respectively).