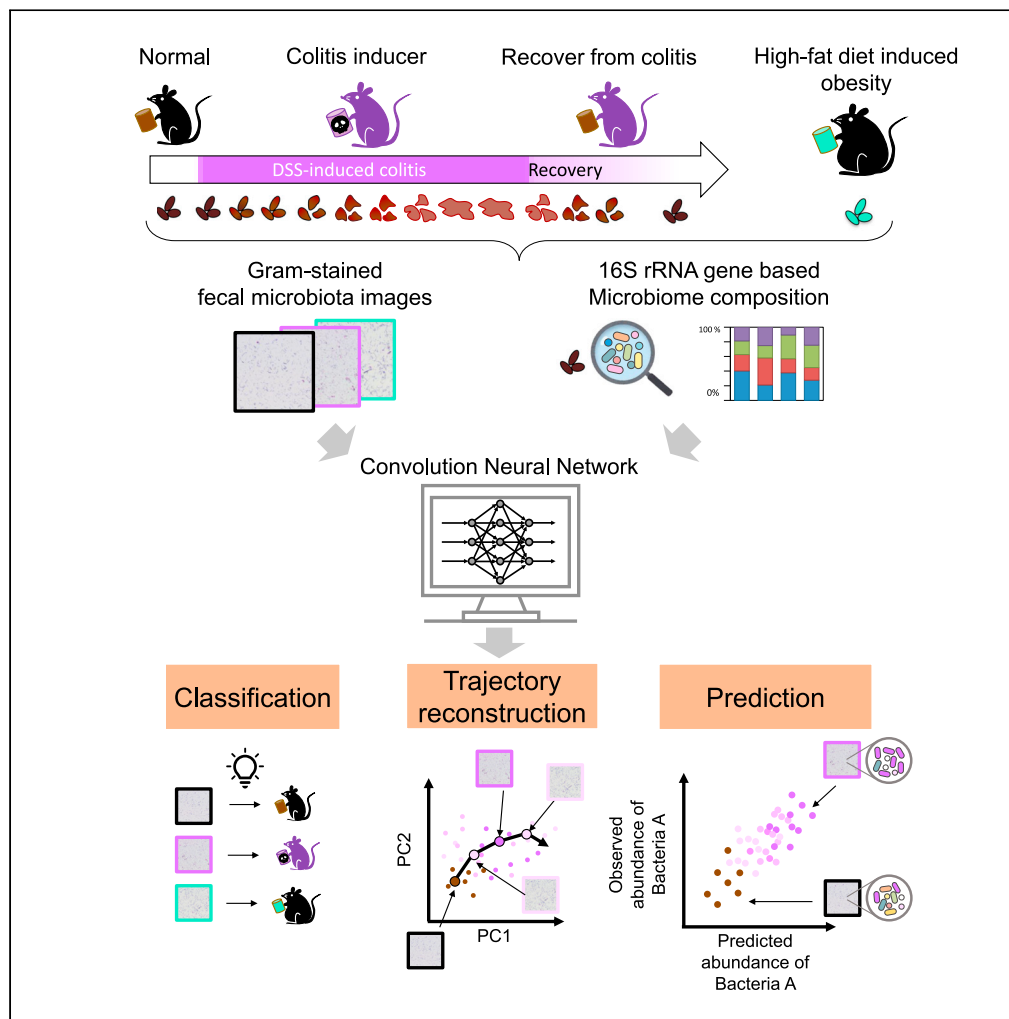


Article

Decoding gut microbiota by imaging analysis of fecal samples



Chikara Furusawa,
Kumi Tanabe,
Chiharu Ishii,
Noriko Kagata,
Masaru Tomita,
Shinji Fukuda

chikara.furusawa@riken.jp
(C.F.)
sfukuda@sfc.keio.ac.jp (S.F.)

Highlights

A deep convolution network classifies gut microbiota based on fecal sample images

Image-based quantitative prediction of gut microbiota composition is demonstrated

This result provides a simple and inexpensive method of gut microbiota analysis



Article

Decoding gut microbiota by imaging analysis of fecal samples

Chikara Furusawa,^{1,2,7,8,*} Kumi Tanabe,^{1,7} Chiharu Ishii,^{3,7} Noriko Kagata,³ Masaru Tomita,³ and Shinji Fukuda^{3,4,5,6,*}

SUMMARY

The gut microbiota plays a crucial role in maintaining health. Monitoring the complex dynamics of its microbial population is, therefore, important. Here, we present a deep convolution network that can characterize the dynamic changes in the gut microbiota using low-resolution images of fecal samples. Further, we demonstrate that the microbial relative abundances, quantified via 16S rRNA amplicon sequencing, can be quantitatively predicted by the neural network. Our approach provides a simple and inexpensive method of gut microbiota analysis.

INTRODUCTION

The gut microbiota is an essential component of the human body, and its activities are important for maintaining human health (Clemente et al., 2012; Lynch and Pedersen, 2016; Sommer and Bäckhed, 2013; Tropini et al., 2017). It plays a critical role in the immune response (Fulde and Hornef, 2014), providing protection against pathogen growth (Kamada et al., 2013), energy biogenesis (Canfora et al., 2015), drug metabolism (Javdan et al., 2020; Jia et al., 2008), and so on. Gut microbiotic abnormalities can promote diseases including cardiovascular disease (Koeth et al., 2013; Tang et al., 2013), central nervous system disorder (Wang and Kasper, 2014), metabolic syndrome (Crovesy et al., 2020; Torres-Fuentes et al., 2017), and cancer (Vivarelli et al., 2019). Thus, analysis of the gut microbiota contributes significantly to human health and medicine.

The gut microbiota is conventionally analyzed using 16S rRNA sequencing data (Hamady and Knight, 2009). The 16S rRNA gene has conserved and variable regions, allowing the design of universal primers with unique amplicons for each bacterial species. 16S rRNA amplicon high-throughput sequencing data are usually processed using bioinformatic tools, and clustered into groups of closely related sequences, known as operational taxonomic units (OTUs). The states of the gut microbiota and of the hosts can be characterized using OTU data in terms of, for instance, disease (Duvall et al., 2017; Forbes et al., 2018) and age-related changes (de la Cuesta-Zuluaga et al., 2019). In such analyses, machine learning algorithms, including random forest and support vector machine algorithms, are often used to classify and extract low-dimensional features (Duvall et al., 2017; Pasolli et al., 2016; Roguet et al., 2018). For large datasets, deep learning methods are also applied, and show high classification performance (Asgari et al., 2018; Fiannaca et al., 2018). Although advances in high-throughput sequencing technologies enable us to acquire huge amounts of 16S amplicon data, these analyses require complex experimental procedures, limiting the number of samples and measurement time points. Simple experimental methods for the systematic analysis of gut microbiota are needed in diagnosis and medical treatment.

We therefore propose a method to characterize the state of gut microbiota based on fecal-sample imaging analysis, using neural network analysis. Deep neural networks, such as convolutional neural networks (CNN), are effective for recognizing biological images, and are now widely applied for image classification and analysis in the biological and medical fields. For example, they have been used to classify images of biological targets such as skin (Esteve et al., 2017) and breast cancer tissues (Cireşan et al., 2013; Han et al., 2017), to classify echocardiograms (Madani et al., 2018), to differentiate cells in hematopoietic lineages (Buggenthin et al., 2017), and for protein localization (Kraus et al., 2017).

We demonstrate that a deep neural network using low-resolution microscope Gram-staining fecal-sample images can characterize dynamic changes in the murine gut microbiota. We obtained images of fecal samples from mice treated with dextran sulfate sodium (DSS), which induces acute colitis via toxicity to gut

¹Center for Biosystem Dynamics Research, RIKEN, Suita, Japan

²Universal Biology Institute, The University of Tokyo, Tokyo, Japan

³Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan

⁴Gut Environmental Design Group, Kanagawa Institute of Industrial Science and Technology, Kawasaki, Japan

⁵Transborder Medical Research Center, University of Tsukuba, Tsukuba, Japan

⁶Metabologenomics, Inc., Tsuruoka, Japan

⁷These authors contributed equally

⁸Lead contact

*Correspondence: chikara.furusawa@riken.jp (C.F.), sfukuda@sfc.keio.ac.jp (S.F.)
<https://doi.org/10.1016/j.isci.2021.103481>



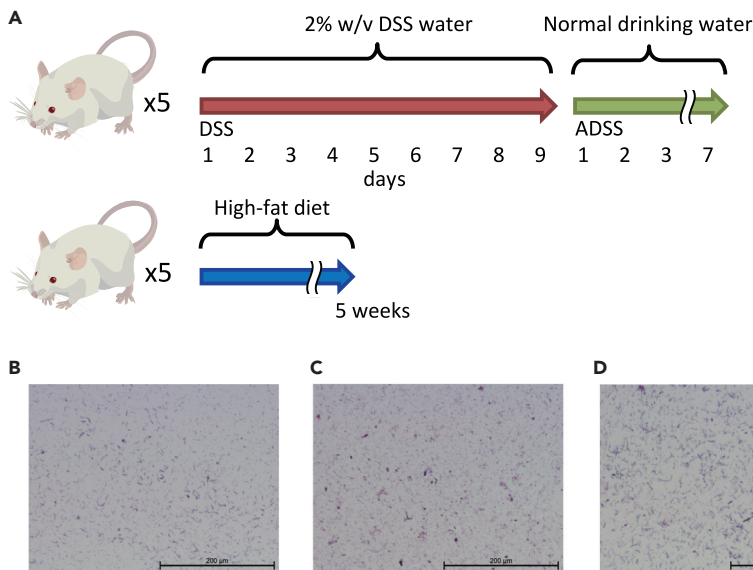


Figure 1. Experimental set-up

(A) Acquisition of fecal samples. Five mice were administered DSS for 9 d to induce colitis, followed by a 7 d recovery process. Fecal samples were obtained daily during DSS induction (DSS0–DSS9) and on days 1, 2, 3, and 7 after stopping DSS administration (ADSS1, ADSS2, ADSS3, and ADSS7). Fecal samples were also obtained from five mice fed a high-fat diet for 5 weeks (HF5). The mouse illustration was obtained from the Togo picture gallery (<http://g86.dbcls.jp/~togoriv/>). (B–D) Examples of fecal-sample images. Fecal-sample images from (B) DSS0, (C) HF5, and (D) DSS9. The scale bar is 200 μm for all images.

epithelial cells (Cooper et al., 1993), and from high-fat-diet-induced obese mice. These fecal-sample images were analyzed using CNN, and microbiotic states were successfully characterized. Following CNN training using a 16S amplicon dataset, the CNN could quantitatively predict fecal-sample relative OTU abundances from the fecal-sample images. This image-based approach provides a simple and inexpensive method for analyzing gut microbiota.

RESULTS

Sampling of murine feces for imaging analysis and 16S amplicon sequencing

To characterize temporal changes in gut microbiota, we obtained fecal samples daily from five mice administered DSS in their drinking water for 9 days (Figure 1A). The samples are named DSS0 to DSS9, with the digit indicating days since starting DSS addition; DSS0 refers to the sample before DSS addition. Recovery from colitis induction was measured on days 1, 2, 3, and 7 after stopping DSS administration (samples ADSS1, ADSS2, ADSS3, and ADSS7). As a different gut microbiota perturbation, we analyzed fecal samples from five mice following five weeks of high-fat (HF)-diet feeding. The samples before and after high-fat feeding are named HF0 and HF5, respectively.

Each fecal sample was homogenized in phosphate-buffered saline (PBS) (1:50, w/v) and Gram-stained before imaging using an ordinary optical microscope. Typical images are shown in Figures 1B–1D. For each sample, we obtained ten images of 1,360 × 1,024 pixels. After background subtraction, each image was divided into non-overlapping sub-images of three different sizes (64 × 64, 128 × 128, or 256 × 256 pixels). To avoid bias in the learning process due to differences in cell density among the images, we selected images in which the ratio of pixels exceeding specific threshold signal intensity was between 0.01 and 0.5.

The same fecal samples were subjected to 16S amplicon sequencing analysis, in which the V1–V2 regions of 16S rRNA-encoding genes were sequenced. Quality filtered sequence reads were clustered into operational taxonomic units (OTUs) to quantify their relative abundance (Figure 2A). OTU relative abundance was drastically altered by adding DSS to the drinking water (comparing DSS0 and DSS9), and recovered

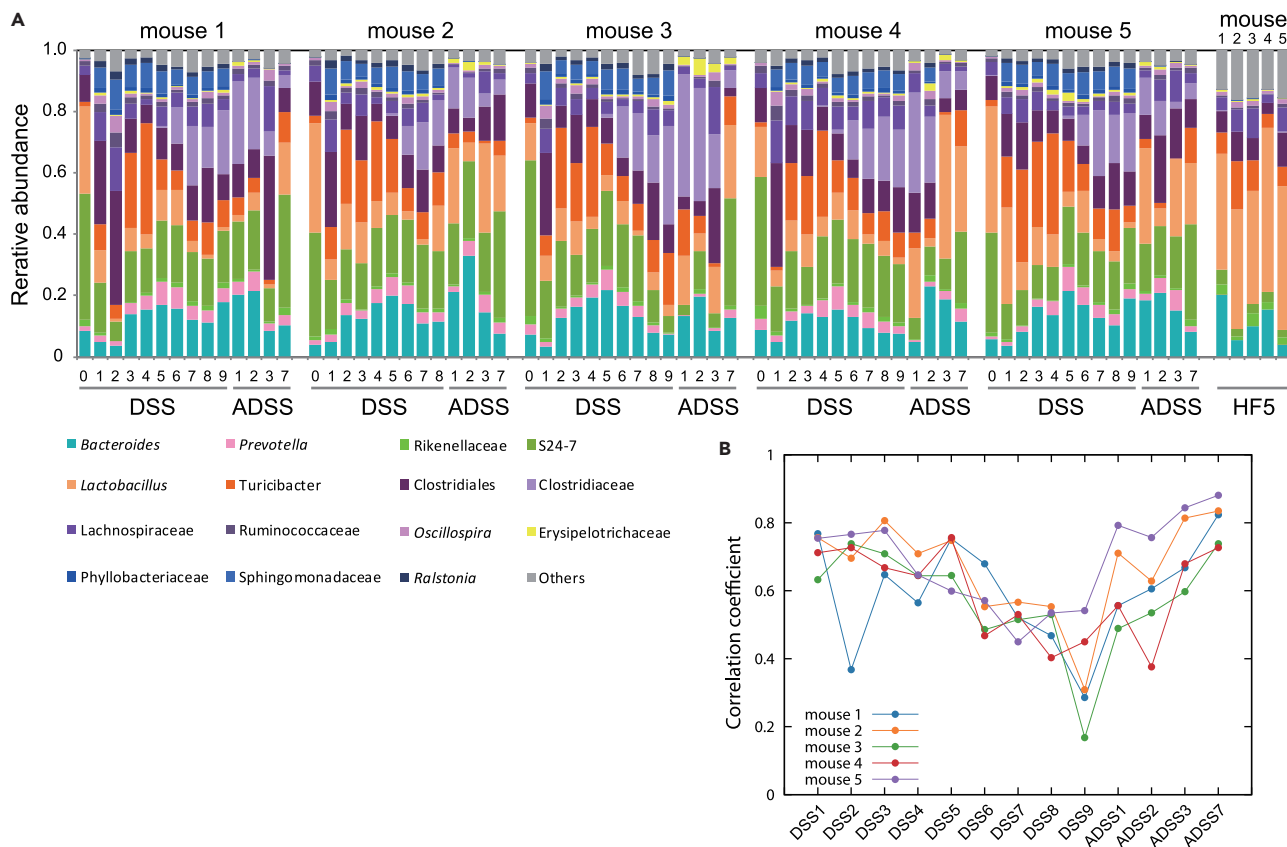


Figure 2. Fecal-sample 16S amplicon sequencing analysis

(A) Relative abundance of operational taxonomic units (OTUs), quantified via 16S rRNA gene sequencing.

(B) For each mouse, the plot presents the change over time in the Spearman's correlation coefficients obtained by comparing the relative OTU abundances before and after the addition of DSS.

to close to its original state 7 days after stopping DSS treatment. This change and recovery were quantified by calculating Spearman's correlation coefficients between relative OTU abundance before and after DSS addition (Figure 2B).

Selection of CNN architecture

We compared four CNN models, LeNet (LeCun et al., 1998), VGG-16 (Simonyan and Zisserman, 2014), ResNet-50 (He et al., 2016), and Xception (Chollet, 2017), to find a suitable model for fecal-sample image feature extraction (see Figure S1 for the details of the CNN models). We selected CNN models using cross-validation (Figure 3A): images from four mice were used to train and validate CNNs, while images from the remaining mouse were used to test classification accuracy. The former images were randomly sampled and partitioned into five sets (folds). At each 5-fold cross-validation iteration, one of the folds was selected to validate the trained model, and the other folds were used for training. Because the number of training images was insufficient to train the CNN from scratch, we used the pretrained ImageNet weights (Deng et al., 2009) on the entire CNN architectures without freezing any layers. For each CNN model and image pixel size, we evaluated the 5-fold cross-validation accuracy. Xception, with a 256 × 256 pixel images, achieved the best validation accuracy (Table S1). We therefore decided to use Xception (Figure 3B) in the following analysis, using 850 images for each category (2,720 images for each mouse; 13,600 images in total).

Fecal-sample image classification using CNN

The CNN was trained using fecal-sample images from four mice, and prediction accuracy was evaluated using the images from the one remaining mouse. This procedure was repeated five times, obtaining the

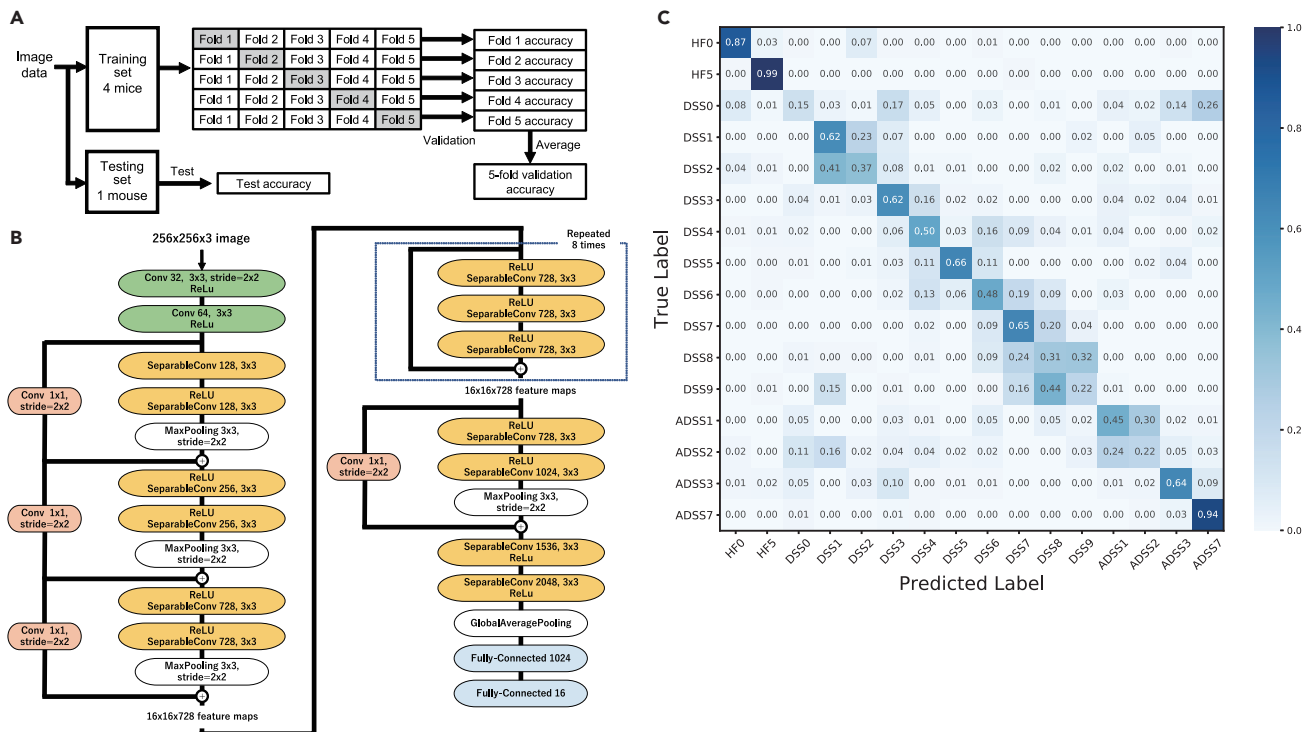


Figure 3. Classification of fecal-sample images using a convolutional neural network

(A) Graphical representation of 5-fold cross-validation.

(B) Architecture of Xception, which was used in this work.

(C) Confusion matrix for evaluating prediction accuracy. The CNN was trained using fecal-sample images from four mice, while the classification results were obtained from the one remaining mouse.

test data from a different mouse each time. The confusion matrix of prediction accuracy averaged over the five test-data mice (Figure 3C), revealed that the CNN achieved 54.4% overall accuracy (averaging the classification accuracy in all categories). This value reflects substantial inaccuracy in the CNN predictions. Although there is a significant amount of data in the off-diagonal elements, such incorrectly predicted data tend to be concentrated close to the diagonal elements (Figure 3C). These incorrect predictions reflect misclassification of samples from adjacent time points (days) in the dataset of DSS treatment; this misclassification was difficult to avoid because such samples share similar microbiotic states. Allowing a one-day difference between the true and predicted categories (for instance, predictions based on DSS2, DSS3, and DSS4 are considered valid predictions for DSS3) generated 76.5% accuracy.

In addition, several misclassifications occurred between specific distant time points. For example, 26% of DSS0 images were misclassified as ADSS7. This misclassification might be because, after stopping DSS for 7 days, the state of the gut microbiota had recovered to almost the original state, making these states difficult to distinguish. This recovery in ADSS7 was supported by the relative OTU abundance determined via 16S amplicon sequencing (Figure 2B). In the classification, we used data from both the DSS and high-fat diet treatments. Using two types of treatment might bias the predictions. To evaluate this possibility, we also classified the DSS and ADSS datasets without the HF dataset, obtaining similar prediction accuracy (Figure S2A).

Further, the image classifications in Figure 3C might reflect differences in the distribution of pixel intensity among samples. This could be because the image background intensity and contrast fluctuated slightly between sampling days, even after normalization. To test this, we trained the CNN using training-data images that had been subjected to random pixel shuffling without changing their intensity distribution. This created difficulty in classification (Figure S2B), indicating that spatial structure in the images contributed to classification accuracy.

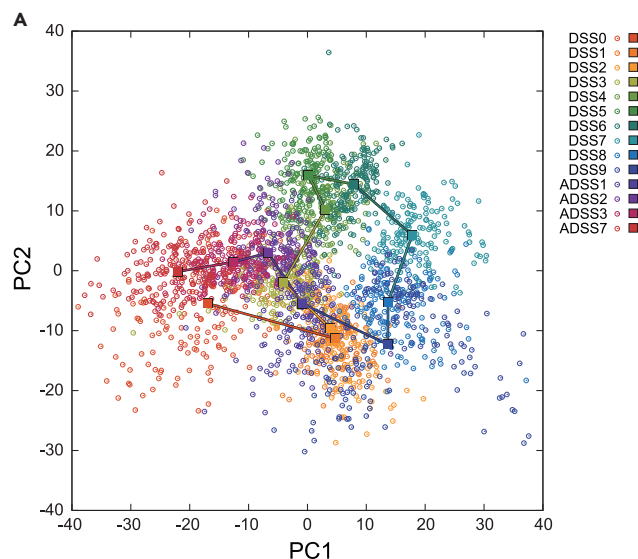


Figure 4. PCA visualization of the activity in the fully connected last layer of the CNN

Each circle represents the activity in the last layer of an input image on the 2-dimensional PCA plane. The activities of the test data images from one mouse are plotted. The PCA was performed using data in 14 categories (DSS0–DSS9, and ADSS1, ADSS2, ADSS3, and ADSS7). The squares represent the median PCA score for each category, while the lines connect the medians of categories sampled on adjacent days.

To evaluate the performance of the CNN for image classification, we analyzed an identical set of sample images using two other widely used classifiers, the random forest and support vector machine (SVM) algorithms. Their classification accuracies were 15.5% and 18.1%, respectively (Figures S2C and S2D), significantly lower than that of CNN.

Reconstructing dynamic changes in gut microbiota

To quantify similarity among the states estimated using the fecal-sample images, we focused on the activations in the (fully connected) last layer, which is a vector with 1,024 components. The CNN converts the input data into linearly separable activities in the last layer, followed by a softmax classification layer to produce the classification probabilities for each category. Thus, the distance values in the high-dimensional activity space in the last layer can be interpreted as similarities between images in terms of the features extracted by the CNN (Eulenberg et al., 2017). This means that images with similar feature representations have more similar microbiotic states than those with dissimilar feature representations.

In the principal component analysis (PCA) visualization of the last-layer activity (Figure 4), the activities in adjacent categories occupy closer regions in the PCA plane. The similarity of the PC scores of ADSS7 and DSS0 suggests that the gut microbiota had recovered by 7 days after removal of DSS. Further, this suggests that CNN feature learning successfully reconstructed the changes in microbiotic state following DSS addition. In contrast, although the t-SNE and UMAP methods clustered images with closer time points (in their visualizations of last-layer activity; Figure S3), they did not reproduce the global structure of the changes in gut microbiotic state over time.

Prediction of 16S amplicon sequencing data using fecal-sample images

To further determine how effectively the imaging-based method distinguishes the gut microbiotic state, we assessed whether fecal Gram-staining images could be used to predict changes in the relative OTU abundances estimated via 16S amplicon sequencing. The same CNN architecture and cross-validation procedures shown in Figures 3A and 3B were used in this analysis. That is, images from four mice were used for CNN training and validation, and images from the one remaining mouse were used to evaluate the prediction accuracy of the log-transformed relative OTU abundances (Figure 2A). Training and prediction were performed for 15 OTUs independently, starting from pretrained ImageNet weights (Deng et al., 2009). In Figure 5A, which relates the observed and predicted relative abundances, the predictions for the 15

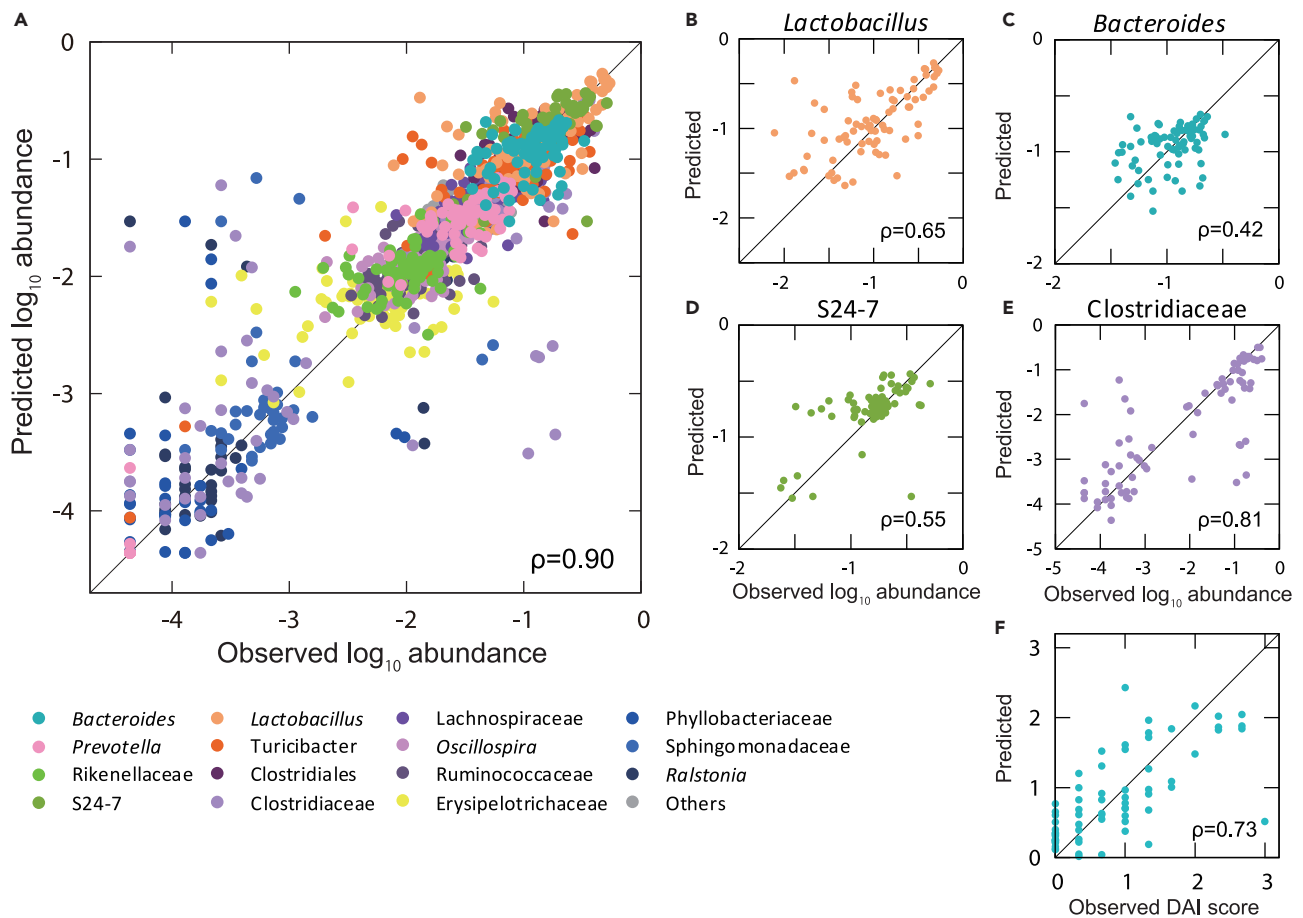


Figure 5. Prediction of relative OTU abundance from fecal-sample images

(A) Comparison of the observed and predicted relative abundances for 15 OTUs. The Spearman's correlation coefficient (ρ) between the observed and predicted abundances is presented. (B–E) Observed and predicted relative abundances for *Lactobacillus*, *Bacteroides*, S24-7, and Clostridiaceae, as examples. Predictions for the other OTUs are presented in Figure S4. (F) Prediction of DAI score based on fecal Gram-staining images.

OTUs are overlaid. Each dot represents the median-predicted relative OTU abundance for the test images (170 images per sample), and panels b–e present the correlations for each of the four OTUs with relatively high abundances (plots for all 15 OTUs are presented in Figure S4). These results strongly suggested that CNN feature extraction of fecal-sample images can quantitatively predict changes in microbial ratios over time. Furthermore, using the same prediction procedure as was used for the 16S amplicon data, we confirmed that the disease activity index (DAI) score (Cooper et al., 1993) (representing colitis severity) can also be quantitatively predicted using fecal-sample images (Figure 5F). This provides further evidence that image-based analysis can represent gut microbiotic changes during and after the development of DSS-induced colitis.

DISCUSSION

This study demonstrates that a CNN can be used to predict the gut microbiotic state based on low-resolution fecal Gram-staining images (Figure 3C). The analysis of last-layer activities suggested that the CNN represents the low-dimensional changes in the gut microbiota over time (Figure 4). Furthermore, using fecal-sample images, the CNN quantitatively predicted the changes in the relative microbial abundances in the gut, quantified via 16S amplicon sequencing (Figure 5).

Gram-staining is a widely used and relatively simple way to identify microorganisms in microscope images. Nonetheless, other staining methods, or combinations of methods, might be more informative for feature

extraction by neural networks, and provide more accurate predictions. On the other hand, if fecal-sample images without staining are available, this can significantly reduce the time and effort required for image acquisition. Improving prediction accuracy depends on improving the methods for microorganism staining and image acquisition and optimizing image resolution.

This study presents low-resolution Gram-staining fecal-image analysis as an inexpensive and simple method to analyze the gut microbiota. It can estimate the gut microbiotic physiological state in terms of its relative microorganism abundances, without requiring complex procedures such as PCR amplification and high-throughput sequencing. Although we studied this in mice, we expect it to be equally applicable to human fecal samples. If this is the case, collecting large quantities of fecal Gram-staining image data, and linking these to medical information, will facilitate diagnosis and treatment planning for hospitalized patients.

Limitations of the study

The prediction accuracy of this CNN-based method was clearly low, at 54.4% (Figure 3C). One reason for this is that the microbiotic states on adjacent days were similar and often difficult to classify accurately: when prediction accuracy was recalculated, allowing for the misclassification of adjacent time points, it increased to 76.5%. Furthermore, to avoid data leakage, we trained the CNN using fecal images from four mice, and used the images from the other mouse to evaluate prediction accuracy. Although the five mice were fed the same diet and housed under the same conditions, they would inevitably differ in gut microbiota (Hoy et al., 2015), reducing prediction accuracy. These individual differences may be reflected in the relatively low prediction accuracy (67.5%) obtained by classifying the 16S amplicon sequencing data into 16 categories using the random forest algorithm (Figure S5). The effects of individual differences might be avoided by obtaining fecal images from individual mice at multiple time points and determining their individual microbiotic states. The development of an experimental design for recording the gut microbiota remains an important problem.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Animals
- METHOD DETAILS
 - Imaging acquisition
 - Image preprocessing
 - DNA extraction
 - 16S rRNA gene sequencing
 - Clinical evaluation of DSS-induced colitis
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Analysis of 16S rRNA gene sequence data
 - Image classification
 - Prediction of 16S rRNA gene amplicon sequencing data and DAI from fecal-sample images

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103481>.

ACKNOWLEDGMENTS

We thank Dr. Yasushi Okada for fruitful discussion; Ms. Yuka Ohara, Mrs. Mitsuko Komatsu, and Noriko Fukuda for technical support; and Dr. Natsumi Saito for conducting the animal experiment. This work was supported in part by the RIKEN Aging Project and the interdisciplinary research program Integrated Symbiology (iSYM), JSPS KAKENHI (17H06389 and 19H05626 to C. F.; 18H04805 to S.F.), JST PRESTO (JPMJPR1537

to S.F.), JST ERATO (JPMJER1902 to S.F. and C.F.), AMED-CREST (JP20gm1010009 to S.F.), the Takeda Science Foundation (to S.F.), the Food Science Institute Foundation (to S.F.), the Yamagata Prefectural Government, and the City of Tsuruoka.

AUTHOR CONTRIBUTIONS

C.F. and S.F. conceived and designed the study. C.I., N.K., and S.F. performed the experiment. C.F., K.T., C.I., M.T., and S.F. analyzed the data. C.F., K.T., C.I., and S.F. wrote the paper.

DECLARATION OF INTERESTS

S.F. is founder and CEO of Metabologenomics, Inc., which is focused on the design and control of the gut environment for human health.

Received: November 23, 2019

Revised: September 21, 2021

Accepted: November 19, 2021

Published: December 17, 2021

REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). TensorFlow: a system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI16), pp. 265–283.
- Asgari, E., Garakani, K., McHardy, A.C., and Mofrad, M.R.K. (2018). MicroPheno: predicting environments and host phenotypes from 16S rRNA gene sequencing using a k-mer based representation of shallow sub-samples. *Bioinformatics* 34, i32–i42. <https://doi.org/10.1093/bioinformatics/bty296>.
- Buggenthin, F., Buettner, F., Hoppe, P.S., Ende, M., Kroiss, M., Strasser, M., Schwarzfischer, M., Loeffler, D., Kokkalis, K.D., Hilsenbeck, O., et al. (2017). Prospective identification of hematopoietic lineage choice by deep learning. *Nat. Methods*. <https://doi.org/10.1038/nmeth.4182>.
- Canfora, E.E., Jocken, J.W., and Blaak, E.E. (2015). Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* 11, 577–591. <https://doi.org/10.1038/nrendo.2015.128>.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*. <https://doi.org/10.1038/nmeth.f.303>.
- Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251–1258. <https://doi.org/10.1109/CVPR.2017.195>.
- Chollet, F. (2018). Keras: The Python Deep Learning Library (Astrophysics Source Code Library). <https://keras.io/>.
- Cireşan, D.C., Giusti, A., Gambardella, L.M., and Schmidhuber, J. (2013). Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, Volume 8150, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, eds. (Springer: Lecture Notes in Computer Science), pp. 411–418. https://doi.org/10.1007/978-3-642-40763-5_51.
- Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell*. <https://doi.org/10.1016/j.cell.2012.01.035>.
- Cooper, H.S., Murthy, S.N.S., Shah, R.S., and Sedergran, D.J. (1993). Clinicopathological study of dextran sulfate sodium experimental murine colitis. *Lab. Invest.* 69, 238–249. <https://europepmc.org/article/med/8350599>.
- Crovesy, L., Masterson, D., and Rosado, E.L. (2020). Profile of the gut microbiota of adults with obesity: a systematic review. *Eur. J. Clin. Nutr.* 74, 1251–1262. <https://doi.org/10.1038/s41430-020-0607-6>.
- de la Cuesta-Zuluaga, J., Kelley, S.T., Chen, Y., Escobar, J.S., Mueller, N.T., Ley, R.E., McDonald, D., Huang, S., Swafford, A.D., Knight, R., and Thackray, V.G. (2019). Age- and sex-dependent patterns of gut microbial diversity in human adults. *mSystems* 4. <https://doi.org/10.1128/mSystems.00261-19>.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: a large-scale hierarchical image database. *IEEE Comput. Vis. Pattern Recognit.* 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* <https://doi.org/10.1128/AEM.03006-05>.
- Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* 8. <https://doi.org/10.1038/s41467-017-01973-8>.
- Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. <https://doi.org/10.1038/nature21056>.
- Eulenberg, P., Köhler, N., Blasi, T., Filby, A., Carpenter, A.E., Rees, P., Theis, F.J., and Wolf, F.A. (2017). Reconstructing cell cycle and disease progression using deep learning. *Nat. Commun.* <https://doi.org/10.1038/s41467-017-00623-3>.
- Fiannaca, A., La Paglia, L., La Rosa, M., Lo Bosco, G., Renda, G., Rizzo, R., Gaglio, S., and Urso, A. (2018). Deep learning models for bacteria taxonomic classification of metagenomic data. *BMC Bioinformatics* 19. <https://doi.org/10.1186/s12859-018-2182-6>.
- Forbes, J.D., Chen, C.Y., Knox, N.C., Marrie, R.A., El-Gabalawy, H., De Kievit, T., Alfa, M., Bernstein, C.N., and Van Domselaar, G. (2018). A comparative study of the gut microbiota in immune-mediated inflammatory diseases - does a common dysbiosis exist? *Microbiome* 6, 1–15. <https://doi.org/10.1186/s40168-018-0603-4>.
- Fulde, M., and Hornef, M.W. (2014). Maturation of the enteric mucosal innate immune system during the postnatal period. *Immunol. Rev.* 260, 21–34. <https://doi.org/10.1111/imr.12190>.
- Hamady, M., and Knight, R. (2009). Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.* 19, 1141–1152. <https://doi.org/10.1101/gr.085464.108>.
- Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., and Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Sci. Rep.* 7, 1–10. <https://doi.org/10.1038/s41598-017-04075-z>.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Hoy, Y.E., Bik, E.M., Lawley, T.D., Holmes, S.P., Monack, D.M., Theriot, J.A., and Relman, D.A. (2015). Variation in taxonomic composition of the fecal microbiota in an inbred mouse strain across

- individuals and time. *PLoS ONE* 10, 1–17. <https://doi.org/10.1371/journal.pone.0142825>.
- Ishii, C., Nakanishi, Y., Murakami, S., Nozu, R., Ueno, M., Hioki, K., Aw, W., Hirayama, A., Soga, T., Ito, M., et al. (2018). A metabolomic approach reveals changes in the intestinal environment of mice fed on american diet. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19124079>.
- Javdan, B., Lopez, J.G., Chankhamjon, P., Lee, Y.C.J., Hull, R., Wu, Q., Wang, X., Chatterjee, S., and Donia, M.S. (2020). Personalized mapping of drug metabolism by the human gut microbiome. *Cell* 181, 1661–1679.e22. <https://doi.org/10.1016/j.cell.2020.05.001>.
- Jia, W., Li, H., Zhao, L., and Nicholson, J.K. (2008). Gut microbiota: a potential new territory for drug targeting. *Nat. Rev. Drug Discov.* 7, 123–129.
- Kamada, N., Chen, G.Y., Inohara, N., and Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. *Nat. Immunol.* 14, 685–690. <https://doi.org/10.1038/ni.2608>.
- Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T., Britt, E.B., Fu, X., Wu, Y., Li, L., et al. (2013). Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* 19, 576–585. <https://doi.org/10.1038/nm.3145>.
- Kraus, O.Z., Grys, B.T., Ba, J., Chong, Y., Frey, B.J., Boone, C., and Andrews, B.J. (2017). Automated analysis of high-content microscopy data with deep learning. *Mol. Syst. Biol.* <https://doi.org/10.15252/msb.20177551>.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., et al. (2016). Jupyter Notebooks – A Publishing Format for Reproducible Computational Workflows (Positioning and Power in Academic Publishing: Players, Agents and Agendas). <https://doi.org/10.3233/978-1-61499-649-1-87>.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE.* <https://doi.org/10.1109/5.726791>.
- Lynch, S.V., and Pedersen, O. (2016). The human intestinal microbiome in health and disease. *N. Engl. J. Med.* 375, 2369–2379. <https://doi.org/10.1056/nejmra1600266>.
- Madani, A., Ong, J.R., Tibrewal, A., and Mofrad, M.R.K. (2018). Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *Npj Digit. Med.* 1, 1–11. <https://doi.org/10.1038/s41746-018-0065-x>.
- Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btr507>.
- Ono, K., Nimura, S., Nishinakagawa, T., Hideshima, Y., Enjyoji, M., Nabeshima, K., and Nakashima, M. (2014). Sodium 4-phenylbutyrate suppresses the development of dextran sulfate sodium-induced colitis in mice. *Exp. Ther. Med.* <https://doi.org/10.3892/etm.2013.1456>.
- Pasolli, E., Truong, D.T., Malik, F., Waldron, L., and Segata, N. (2016). Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput. Biol.* 12, 1–26. <https://doi.org/10.1371/journal.pcbi.1004977>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Roguet, A., Eren, A.M., Newton, R.J., and McLellan, S.L. (2018). Fecal source identification using random forest. *Microbiome* 6, 1–15. <https://doi.org/10.1186/s40168-018-0568-3>.
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>.
- Sommer, F., and Bäckhed, F. (2013). The gut microbiota-masters of host development and physiology. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/nrmicro2974>.
- Tang, W.H.W., Wang, Z., Levison, B.S., Koeth, R.A., Britt, E.B., Fu, X., Wu, Y., and Hazen, S.L. (2013). Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* 368, 1575–1584. <https://doi.org/10.1056/nejmoa1109400>.
- Torres-Fuentes, C., Schellekens, H., Dinan, T.G., and Cryan, J.F. (2017). The microbiota–gut–brain axis in obesity. *Lancet Gastroenterol. Hepatol.* 2, 747–756. [https://doi.org/10.1016/S2468-1253\(17\)30147-4](https://doi.org/10.1016/S2468-1253(17)30147-4).
- Tropini, C., Earle, K.A., Huang, K.C., and Sonnenburg, J.L. (2017). The gut microbiome: connecting spatial organization to function. *Cell Host Microbe.* <https://doi.org/10.1016/j.chom.2017.03.010>.
- Vivarelli, S., Salemi, R., Candido, S., Falzone, L., Santagati, M., Stefani, S., Torino, F., Banna, G.L., Tonini, G., and Libra, M. (2019). Gut microbiota and cancer: from pathogenesis to therapy. *Cancers (Basel)* 11, 1–26. <https://doi.org/10.3390/cancers11010038>.
- Wang, Y., and Kasper, L.H. (2014). The role of microbiome in central nervous system disorders. *Brain Behav. Immun.* 38, 1–12. <https://doi.org/10.1016/j.bbi.2013.12.015>.
- Zeiler, M.D. (2012). ADADELTA: an adaptive learning rate method. <https://arxiv.org/abs/1212.5701>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Sodium Dextran Sulfate MW36,000 - 50,000	MP Biomedicals	Cat# 160110
Sodium Dodecyl Sulfate	Wako	Cat# 192-13981
2-Amino-2-hydroxymethyl-1,3-propanediol	Wako	Cat# 204-07885
2NA(EDTA.2NA)	Dojindo	Cat# 345-01865
Phenol/Chloroform/Isoamyl alcohol 25:24:1 Mixed, pH 7.9	Nacalai tesque	Cat# 25970-56
Ethanol (99.5)	Wako	Cat# 057-00456
RNase A	Roche	Cat#10109142001
Critical commercial assays		
Crystal Violet Solution	Wako	Cat# 405-51065
Pfeiffer's Solution	Wako	Cat# 416-53095
Iodine Solution	Wako	Cat# 402-51075
Acetone.Ethanol Solution	Wako	Cat# 409-51085
Agencourt AMPure XP	Beckman coulter	Cat# A63881
Tks Gflex DNA Polymerase	Takara Bio	Cat# R060A
MiSeq Reagent Kit v3(600 Cycles)	Illumina	Cat# MS-102-3003
Deposited data		
16S rRNA gene sequence of murine microbiota	DDBJ	DRA008811
Experimental models: Organisms/strains		
C57BL/6J mice	CLEA Japan	C57BL/6J
Oligonucleotides		
27Fmod with an overhang adapter (5'-ACACT CTTCCCTACACGACGCTCTCCGATCTAG RGTTCGATYMTGGCTCAG-3')	Ishii et al. (2018)	NA
338R with an overhang adapter (5'-GTGACTGG AGTTCAGACGTGTGCTCTCCGATCTTGCTG CCTCCGTAGGAGT-3')	Ishii et al. (2018)	NA
Software and algorithms		
FLASH 1.2.11	Magoč and Salzberg (2011)	https://ccb.jhu.edu/software/FLASH/
QIIME 1.9.1	Caporaso et al. (2010)	
Keras 2.3.1	Chollet (2018)	https://keras.io
Tensorflow-gpu 1.14.0	Abadi et al. (2016)	https://www.tensorflow.org
Scikit-learn 0.24.2	Pedregosa et al. (2011)	https://scikit-learn.org/
Jupyter lab 1.0.2	Kluyver et al. (2016)	https://jupyter.org
Other		
CLEA Rodent Diet CE-2	CLEA Japan	CE-2
High Fat Diet 32	CLEA Japan	HFD32
Leica DM2500	Leica	DM2500
Zirconia beads 0.1φ	TOMY	Cat# ZSB-01
Zirconia beads 3.0φ	TOMY	Cat# ZB-30

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Chikara Furusawa (chikara.furusawa@riken.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data and code for discriminating microbiotic states are available via <https://github.com/kumi-tanabe/GutMicrobiota>. The microbiome analysis data have been deposited in the DDBJ Sequence Read Archive: <https://dbj.nig.ac.jp/DRAsearch/> under accession number DRA008811.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

The animal experiment was performed using protocols approved by the Animal Studies Committee of Keio University School of Medicine (Tokyo, Japan) and the National Institute of Technology, Tsuruoka College (Tsuruoka, Yamagata, Japan). All mice were housed at the National Institute of Technology, Tsuruoka College, under a 12 h light–dark cycle, with *ad libitum* access to food and water. For the high-fat-diet group, 5-week-old male C57BL/6J mice (N = 5, purchased from CLEA Japan, Inc., Tokyo, Japan) were fed a normal diet (CE-2, CLEA Japan, Inc.) and tap water for 1 week to acclimatize to the new environment. After acclimatization, the diet was replaced with a high-fat diet (HFD32, CLEA Japan, Inc.) for 5 weeks. Fecal samples were obtained from each animal at just before diet replacement (HF0) and after 5 weeks of high-fat diet (HF5). For the DSS-treated group, 8-week-old male C57BL/6J mice (N = 5; CLEA Japan, Inc.) were fed CE-2 and tap water for 1 week to acclimatize to the new environment. Thereafter, their drinking water was replaced with tap water containing 2.0% (w/v) DSS (MP Biomedicals, Santa Ana, CA) for 9 d, then replaced with tap water for 7 d to allow recovery from the DSS-induced colitis. Fecal samples were obtained daily from each animal, from immediately before the day of DSS treatment (DSS0) to 3 d after stopping DSS treatment (ADSS3), and again 7 d after stopping DSS treatment (ADSS7). The fecal samples were freeze-dried and stored at -80°C .

METHOD DETAILS

Imaging acquisition

Each freeze-dried fecal sample was suspended in PBS (1:50, w/v) and homogenized using a pellet pestle homogenizer (Sigma-Aldrich, St. Louis, MO). The concentration of each solution was measured by determining the optical density at 600 nm (OD₆₀₀); the solutions were then diluted with PBS to OD₆₀₀ = 1.0. The 500-fold dilutions were made in PBS; these were centrifuged for 3 s using a tabletop centrifuge. Ten microliters of supernatant was spread on a slide, and after rapid heat fixation of the smears, the slides were Gram-stained with stabilized iodine (Wako, Japan). Ten microscopic fields of Gram-staining images were obtained from each slide using a 20 \times objective (DM2500 optical microscope; Leica, Wetzlar, Germany).

Image preprocessing

The original RGB-color images were 1,360 \times 1,024 pixels in size. First, we normalized the background intensity level by subtracting the 95% quantile intensity from each image. Then, after grayscale conversion by averaging the red, blue, and green pixel intensities, we obtained sub-images of 64 \times 64, 128 \times 128, and 256 \times 256 pixels, respectively, at 256 non-overlapping intensity levels.

DNA extraction

Fecal DNA isolation was performed as described previously (Ishii et al., 2018), with some modifications. In short, approximately 10 mg of each freeze-dried fecal sample was combined with four 3.0 mm zirconia beads and subjected to vigorous shaking (1,500 rpm for 10 min) using a Shake Master (Biomedical Science, Tokyo, Japan). Thereafter, each fecal sample was combined with approximately 100 mg of 0.1 mm zirconia/silica beads, 400 μL DNA extraction buffer (TE containing 1% (w/v) sodium dodecyl sulfate), and 400 μL phenol/chloroform/isoamyl alcohol (25:24:1), and subjected to vigorous shaking (1,500 rpm for 5 min) using a Shake Master. The resulting emulsion was subjected to centrifugation at 17,800 $\times g$ for 10 min at room

temperature, and the bacterial genomic DNA was purified from the aqueous phase via a standard phenol/chloroform/isoamyl alcohol protocol. RNA was removed from the sample via RNase A treatment. The resulting DNA sample was then purified again using phenol/chloroform/isoamyl alcohol treatment.

16S rRNA gene sequencing

16S rRNA genes in the fecal DNA samples were analyzed using a MiSeq sequencer (Illumina, San Diego, CA) as described previously (Ishii et al., 2018). The V1-V2 region of the 16S rRNA genes was amplified from the DNA (approximately 10 ng per reaction) isolated from feces using a bacterial universal primer set consisting of the 27Fmod primer with an overhang adapter (5'-ACACTCTTCCCTACACGACGCTCTCCGATCTAGRGTGATYMTGGCTCAG-3') and 338R primer with an overhang adapter (5'-GTGACTG GAGTTCAGACGTGTGCTCTTCCGATCTTGCTGCCTCCGTTAGGAGT-3')⁴¹. PCR was performed using Tks Gflex DNA Polymerase (Takara Bio, Inc., Shiga, Japan) with the following amplification program: one cycle of denaturation at 98°C for 1 min; 20 cycles of amplification at 98°C for 10 s, 55°C for 15 s, and 68°C for 30 s; and a final extension at 68°C for 3 min. The amplified products were purified using Agencourt AMPure XP kits (Beckman Coulter, Atlanta, GA). The purified products then were further amplified using the following primer pair: a forward primer (5'-AATGATACGGCGACCACCGAGATCTACAC-NNNNNNN NN-ACACTCTTCCCTACACGACGC-3') containing the P5 sequence, a unique 8 bp barcode sequence for each sample (indicated by the string of Ns), and an overhang adapter; and a reverse primer (5'-CAAG-CAGAAGACGGCATAACGAGAT-NNNNNNNNN-GTGACTGGAGTTCAGACGTGTG-3') containing the P7 sequence, a unique 8 bp barcode sequence for each sample (indicated by the string of Ns), and an overhang adapter. After purification, the purified products were mixed in approximately equal molar concentrations to generate a 4 nM library pool, after which the final library pool was diluted to 6 pM, including a 10% PhiX Control v3 (Illumina) spike-in for sequencing. Finally, MiSeq sequencing was performed according to the manufacturer's instructions. In this study, 2 × 300 bp paired-end sequencing was conducted.

Clinical evaluation of DSS-induced colitis

DSS-induced colitis was evaluated daily as previously described (Ono et al., 2014), with some modifications. DAI was calculated by summing the scores for three variables: Bodyweight loss (0, none; 1, 1–5%; 2, 5–10%; 3, 10–15%; and 4, >15%), stool consistency (0, normal; 2, loose stools; and 4, diarrhea), and stool blood (0, negative; 2, half of the feces contain blood; and 4, bloody stool). Bodyweight loss was calculated as the percentage difference between the bodyweight on the day immediately before starting DSS treatment and that on the day the animal was weighed.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of 16S rRNA gene sequence data

Analysis of 16S rRNA gene sequences was performed as described previously (Ishii et al., 2018), with some modifications. Initially, to assemble the paired-end reads, Fast Length Adjustment of SHort reads (FLASH) (v1.2.11) (Magoč and Salzberg, 2011) was used. Assembled reads with an average Q-value < 25 were filtered out using an in-house script. A total of 23,000 filter-passed reads were randomly selected from each sample and used for further analysis. Reads then were processed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (ver. 1.9.1) (Caporaso et al., 2010). Sequences were clustered into OTUs based on 97% sequence similarity, and OTUs were analyzed using the Greengenes Database (ver. 13.8) (DeSantis et al., 2006).

Image classification

For CNN analysis, we evaluated LeNet (LeCun et al., 1998), VGG-16 (Simonyan and Zisserman, 2014), ResNet-50 (He et al., 2016), and Xception (Chollet, 2017), to find a suitable architecture to analyze the fecal-sample images (Figure S1). For each CNN model and image size, we calculated the 5-fold validation accuracy (Table S1): Xception, with 256 × 256 pixel images, achieved the best performance, hence we used it in the present study. We used the pretrained ImageNet weights (Deng et al., 2009) on the entire CNN architectures without freezing any layers. We used ADADELTA (Zeiler, 2012), a variant of the stochastic gradient descent algorithm, as the optimization method, with cross-entropy as an objective function. Training was performed using a batch size of 32, for 300 epochs. For implementation, we used the Keras platform with a TensorFlow backend.

To avoid classification bias due to differences in cell density, we selected images in which the ratio of pixels exceeding a specific threshold signal intensity was within a given range. To determine this range, using the test data, we used the Xception model with 256×256 pixel images, and examined how varying the signal intensity threshold affected the prediction accuracy. We tested nine sets of upper and lower intensity thresholds (three lower bounds: 0.005, 0.01, 0.02; and three upper bounds: 0.4, 0.5, 0.75; [Table S2](#)), achieving comparable prediction accuracy using each set. We therefore decided to use the intensity threshold range 0.01–0.5 for image selection.

The SVM and random forest algorithms used for image classification ([Figures S2C and S2D](#)) were implemented using the scikit-learn library ([Pedregosa et al., 2011](#)). SVMs were created with a radial basis function (RBF) kernel. A grid search was performed to choose the values of γ , the coefficient in the RBF kernel, and C, the penalty term of the error. For the random forest algorithm, the following hyperparameters were obtained by a grid search: `n_estimators`, `max_features`, `random_state`, `min_samples_split`, and `max_depth`.

Prediction of 16S rRNA gene amplicon sequencing data and DAI from fecal-sample images

For the prediction of log-transformed relative OTU abundances based on fecal-sample images ([Figures 5 and S4](#)), we used the Xception architecture ([Figure 3B](#)), in which a fully connected layer with 1,024 nodes is followed by the last fully connected regression layer. Again, we used the pretrained ImageNet weights ([Deng et al., 2009](#)) without freezing any layers. The Xception network was trained using 256×256 pixel fecal-sample images, and relative OTU abundance prediction accuracy was evaluated against the relative OTU abundances of the one remaining mouse. The DAI score was predicted in the same way as OTU abundance, using the Xception network.