



# TonguExpert: A Deep Learning-Based Algorithm Platform for Fine-Grained Extraction and Classification of Tongue Phenotypes

Ting Li<sup>1</sup> · Ling Zuo<sup>2</sup> · Pengyu Wang<sup>1</sup> · Liangfu Yang<sup>1</sup> · Zijia Liu<sup>3</sup> · Xu Wang<sup>4</sup> · Jingze Tan<sup>5,6</sup> · Yajun Yang<sup>5,6</sup> · Jiucun Wang<sup>5,7,8</sup> · Yong Zhou<sup>9</sup> · Li Jin<sup>5,7,8</sup> · Guangtao Zhai<sup>3</sup> · Jianxin Chen<sup>2,4</sup> · Qianqian Peng<sup>1</sup> · Guoqing Zhang<sup>1</sup> · Sijia Wang<sup>1,10</sup>

Received: 17 May 2024 / Revised: 3 November 2024 / Accepted: 12 November 2024  
© International Human Phenome Institutes (Shanghai) 2024

## Abstract

Tongue analysis holds promise for disease detection and health monitoring, especially in traditional Chinese medicine. However, its subjectivity hinders clinical applications. Deep learning offers a path for automated tongue diagnosis, yet existing methods struggle to capture subtle details, and the lack of large datasets hampers the development of robust and generalizable models. To address these challenges, we introduce TonguExpert (<https://www.biosino.org/TonguExpert>), a free platform for archiving, analyzing, and extracting phenotypes from tongue images. Our deep learning framework integrates cutting-edge techniques for tongue segmentation and phenotype extraction. TonguExpert analyzes a massive dataset of 5992 tongue images from a Chinese population and extracts 773 phenotypes including five predicted labels and their probabilities, 355 global features (entire tongue, tongue body, and tongue coating) and 408 local features (fissures and tooth marks) in a unified process. Besides, 580 additional features for five tongue subregions are also available for future study. Notably, TonguExpert outperforms manual classification methods, achieving high accuracy (ROC-AUC 0.89–0.99 for color, 0.97 for fissures, 0.88 for tooth marks). Additionally, the model generalizes well to predict new phenotypes (e.g., greasy coating) using external datasets. This allows the model to learn from a broader spectrum of data, potentially improving its overall performance. We also release the largest publicly available dataset of tongue images and phenotypes, which is invaluable for advancing automated analysis and clinical applications of tongue diagnosis. In summary, this research advances automated tongue diagnosis, paving the way for wider clinical adoption and potentially expanding the applications in the future.

**Keywords** TonguExpert · Tongue images · Tongue phenotype extraction · Automated platform · Traditional Chinese medicine · Deep learning

## Abbreviations

ALT	Alanine aminotransferase	mAP	Mean Average Precision
AUC	Area under the curve	mIoU	Mean IoU
CKD	Chronic kidney disease	MSE	Mean squared error
CLARA	Clustering large applications	PCA	Principal component analysis
CNN	Convolutional neural network	RGB	Red, Green and Blue
HOG	Histogram of oriented gradients	ROC	Receiver operating characteristic
HSV	Hue, Saturation and Value	SAM	Segmentation Anything model
IoU	Intersection over union	TCM	Traditional Chinese medicine
LASSO	Least absolute shrinkage and selection operator	VGG16	Visual Geometry Group 16
LBP	Local binary patterns		

Ting Li and Ling Zuo contributed equally to this work.

Extended author information available on the last page of the article

## Introduction

Tongue phenotypes hold great potential in the early detection of diseases and health monitoring, especially in traditional Chinese medicine (TCM) (Kim et al. 2014; Lo et al. 2012; Morita et al. 2021). Tongue diagnosis has long been utilized by TCM practitioners for disease differentiation and therapeutic decision-making. Traditionally, this involves visually examining various attributes of the tongue such as shape, color, coating, fissures, and tooth marks. These observed features are regarded as manifestations indicative of the internal organs and overall health status (Lo et al. 2012). Recent research has demonstrated that changes in the color and thickness of the tongue body or coating can reflect alterations in the microbiota of the tongue coating, which is associated with various conditions like tumors (Chen et al. 2024; Yuan et al. 2023), rheumatoid arthritis (Liu et al. 2024), and hepatitis B (Tang et al. 2022). A tooth-marked tongue, characterized by tooth marks caused by an enlarged tongue pressing against teeth (Yanagisawa et al. 2007), is considered a significant indicator for TCM pathogenic factors such as spleen deficiency. It's closely associated with clinical manifestations related to gastric diseases (Zhou et al. 2022). Clinical evidence suggests a correlation between tooth marks and elevated levels of alanine aminotransferase (ALT) (Hu et al. 2019), chronic kidney disease (CKD) (Chung et al. 2023) and acute ischemic stroke (Huang et al. 2022). While a fissured tongue is often perceived as a benign condition characterized by deep fissures of varying depth, size, and number on the tongue dorsum, limited evidence exists regarding its genetic and pathological mechanisms. Some studies have associated fissured tongue with psoriasis, age, *IL-17RC* mutation, and hormonal changes associated with pregnancy (Lyng Pedersen et al. 2015; Monshi et al. 2021; Picciani et al. 2018; Xie et al. 2023; Yun et al. 2007).

Tongue diagnostic offers great potential in early detection of diseases and long-term health monitoring due to its simplicity, effectiveness, cheapness, and non-invasiveness. However, the subjective nature of traditional tongue diagnosis limits its medical applications, as diagnostic accuracy relies heavily on the experience and knowledge of practitioners, leading to unreliable and inconsistent results. To address this limitation, researchers are increasingly focusing on developing objective and quantitative tongue diagnostic methods (Tania et al. 2019).

In recent years, efforts have been directed to developing tongue image processing techniques, including tongue edge segmentation and color analysis (Cao et al. 2016; Liu et al. 2022; Song et al. 2022; Wang et al. 2013; Xu et al. 2020; Yuan et al. 2021). Tooth marks and fissures, being local symptoms of tongue conditions, present a fine-grained

classification challenge. Previous studies have approached this problem by extracting these features as classified phenotypes. For instance, classification models of tooth-marked tongue have been constructed based on colors, concavity, and changes in brightness of the tooth-marked region (Shao et al. 2014). With the advancements in artificial intelligence and deep learning, convolutional neural network (CNN) algorithms are being increasingly applied to tongue image processing and classification. For example, Sun et al. used a seven-layer CNN model to recognize tooth marks (Sun et al. 2019), while Wang et al. employed a 34-layer CNN model for classifying tooth-marked tongues (Wang et al. 2014). Tang et al. utilized a cascaded CNN to detect the tongue region and tooth marks (Tang et al. 2020).

While existing research on automated tongue diagnosis offers promise, several key shortcomings remain: (1) Fragmented Feature Analysis. Current methods typically analyze tongue features in isolation (color, morphology, local features) neglecting potential synergy for a more comprehensive assessment. (2) Limited Detail Extraction. Prior research often focuses solely on presence/absence of tooth marks or fissures, failing to capture crucial details like number, length, and location. This lack of granularity hinders capturing subtle changes in tongue appearance, impacting clinical interpretation. (3) Data Scarcity. Limited datasets restrict the robustness and generalizability of trained models. (4) Restricted Prediction Scope. Previous models struggle to predict new phenotypes based on existing features, limiting their adaptability. (5) Data Inaccessibility. The absence of publicly available datasets impedes model comparison and advancement.

To address these challenges, this study proposes an automated framework for tongue phenotype recognition and classification using deep learning algorithms (Fig. 1). Leveraging a large-scale dataset of individuals, our framework offers several key advancements: (1) Unified Feature Extraction. Our approach extracts both global (color, morphology) and local features (fissures, tooth marks) in a unified process. (2) Standardized Dataset Creation. We introduce the largest known dataset of tongue images, including tongue region images, corresponding mask images, and extracted tongue phenotype labels with associated information. This standardized dataset facilitates future research. (3) State-of-the-Art Segmentation and Prediction. Our framework incorporates cutting-edge image segmentation algorithms like Segmentation Anything model (SAM) (Kirillov et al. 2023) along with other deep learning techniques, to capture fine-grained details of tongue phenotypes. Additionally, our model demonstrates the ability to predict new phenotypes based on existing features, validated through an independent dataset. (4) User-Friendly Implementation and Broader Impact. The developed algorithms are integrated into a

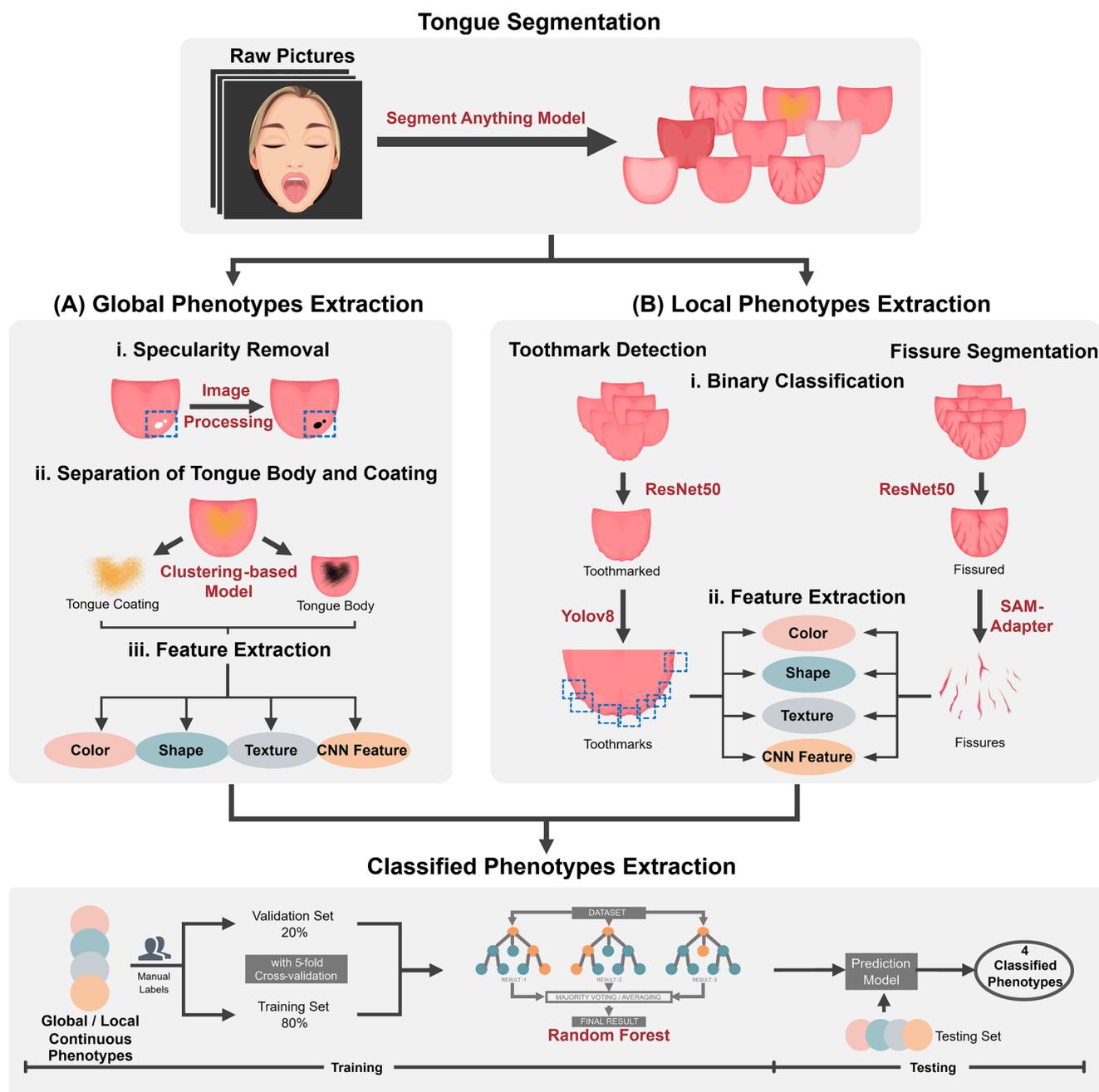


Fig. 1 Study overview

user-friendly website (<https://www.biosino.org/TonguExpert>), allowing users to upload tongue images for analysis and access the standardized dataset for further exploration. This research advances automated tongue diagnosis, paving the way for its wider adoption in clinical settings and potentially expanding the applications of tongue diagnosis in the future.

## Materials and Methods

### Samples and Dataset

Our tongue image dataset was collected from two cohorts of volunteers. The first cohort included 2088 individuals who were recruited in four regional districts of China from 2015 to 2019: Zhengzhou, Taizhou, Miaojiang and Nanning (NSPT, males: females=1:1.65, aged from 18 to 78 years old, mean±SD=48.46±12.91). This cohort is a sub project

of The National Science & Technology Basic Research Project which was approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (No. 14117).

Samples in the second cohort included 3904 Han Chinese individuals who were recruited from 2018 to 2019 as volunteers in the Staff Hospital of Jidong oil-field, Tangshan City, China (JD, males: females=1:1.09, aged from 20 to 80 years old, mean $\pm$ SD=45.54 $\pm$ 13.26). This study was approved by the Institutional Review Board of Shanghai Institute of Life Sciences, Chinese Academy of Sciences (ER-SIBS-261410-A1801). To retain as many samples as possible and ensure that the tongue images are representative and diverse, we excluded images based only on photo quality rather than participants' health status. Further details of participants recruitment can be found in the previous papers (Peng et al. 2024; Xia et al. 2020). All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the Declaration of Helsinki and its later amendments or comparable ethical standards. All participants provided written informed consent.

The greasy coating dataset utilized for assessing the generalization capability of the model originates from one of our previously published articles (Wang et al. 2022), including 1486 tongue images of three tongue image labels: non-greasy ( $N=85$ ), greasy ( $N=759$ ), and thick greasy ( $N=642$ ). Details of image processing are available in the original paper.

### Tongue Images Quality Control

To ensure the stability and reliability of imaging conditions, all images were captured using a specialized tongue diagnostic instrument. This device comprises a dark box equipped with a stable artificial light source and a high-performance digital camera, both set in fixed positions. Such an arrangement ensures consistent irradiation of the tongue, with the positions of the light source and camera being static, thereby mitigating the impact of external lighting on the photographic outcomes.

Quality control was rigorously applied to all tongue images. Any images exhibiting blurriness or showing incomplete tongues were excluded. Notably, images displaying improper tongue extension were also discarded. Improper extension, characterized by the tension or curling of the tongue due to excessive force during extension, can interfere with the tongue's blood circulation. This may lead to alterations in the tongue's color, coating, or moisture levels. Since such changes could obscure the true characteristics of the tongue, images depicting improper tongue

extension were removed to maintain the integrity of the study's findings. Ultimately, we obtained a dataset consisting of 5992 tongue diagnosis photos, with a male-to-female ratio of 1:1.26 and an age range of 46.55 $\pm$ 13.21 years.

### Tongue Segmentation

To mitigate the impact of extraneous elements such as facial features, lips, and the inner walls of the device, the tongue region in the original images was carefully segmented and extracted for detailed analysis. In this project, the segmentation of the tongue body was achieved using a combined approach of Grounding DINO and SAM. Developed by Meta AI, the SAM enables zero-shot transfer, facilitating effortless interactive and automatic segmentation of any image or video without the need for additional training. Grounding DINO, an advanced zero-shot object detection model, integrates the Transformer-based DINO detector with grounding pre-training. By merging Grounding DINO with the SAM, precise segmentation of the tongue region from tongue diagnosis photos is attainable without training dataset.

The tongue region images obtained from SAM exhibit some shadows cast by lips on the tongue surface; to eliminate these shadows, we employed the SAM-Adapter model. This method aims to optimize the SAM's performance in complex tasks such as detecting camouflaged objects, shadow elimination, and medical image segmentation. The SAM-Adapter enhances the SAM's adaptability to specific domain information or visual cues by incorporating efficient adapter modules into the segmentation network, thereby achieving more precise segmentation results in complex image processing scenarios. In this study, 242 shadow-free tongue images were used as a training set to deeply train the SAM-Adapter model. After several rounds of iterative training, the model's performance significantly improved, and it was eventually saved and used for precise segmentation of shadow-free tongue images.

Considering possible future clinical applications, we divided the tongue into five subregions based on previous literature (Xu 2017): tongue bottom, tongue tip, tongue center, and the two tongue margins. Specifically, the bottom 1/5 of the entire tongue image was classified as the tongue tip, then the 1/5 on both the left and right sides as the tongue margins, the top 1/5 as the tongue bottom, and the remaining part as the tongue center. These divided subregions were used for extraction of continuous phenotypes (except for shape phenotypes, which are not applicable). Additionally, we pairwise compared the gray median values of the five subregions. For data meeting normality and homogeneity of variance, we used the t-test; for data meeting normality but not homogeneity of variance, we used Welch's t-test;

for data not meeting normality but meeting homogeneity of variance, we used the Mann-Whitney U test; and for data meeting neither, we used the Kruskal-Wallis test.

### Global Tongue Phenotypes Extraction

To extract more accurate tongue color phenotypes, it is generally necessary to preprocess the tongue images first, with the detection of reflective spots being a crucial aspect of this preprocessing. The variations in lighting conditions during the capture of each photograph lead to inconsistent intervals of grayscale distribution in the highlights of these images. Utilizing a fixed threshold to identify highlights will lead to significant inaccuracies. Consequently, we apply histogram equalization to all photos in order to establish a dependable interval for highlight distribution. Subsequently, we utilize 230 as the threshold value to create a binary mask, which is then employed to eradicate the highlighted area in the original image.

Separation of the tongue coating and body is a prerequisite and fundamental step for the qualitative and quantitative analysis of their colors. Typically, the k-means clustering algorithm is employed for the separation of tongue coating and body. The k-means algorithm, a partition-based clustering method, is simple but can be influenced by outliers and is not suitable for large datasets. Upon comparison, it was found that the Clustering Large Applications (CLARA) algorithm, capable of handling large data sets, was more suitable. Therefore, in this project, the CLARA clustering algorithm was used to achieve the separation of coating and body. After separation, the next step is to automatically determine the tongue coating and tongue body using an algorithm. Since the tongue body has a more vivid red color compared to the coating, we use a method based on the proportion of the red value in determining the color components. Specifically, if the red value occupies a larger proportion in the Red, Green and Blue (RGB) color space, the system automatically identifies the image as tongue body; conversely, if the red value's proportion is smaller, it is identified as tongue coating.

To comprehensively describe the color distribution of the entire tongue, tongue body and tongue coating, histograms of color distributions in RGB, L\*ab, Hue, Saturation and Value (HSV) color spaces, and grayscale values were extracted, with each color space divided into five intervals to obtain the frequency values of colors in each interval. Subsequently, the first moment (representing the mean of the data distribution), the second moment (representing the variance of the data distribution), and the third moment (representing the skewness of the data distribution) of these color spaces were calculated. Additionally, we also calculated the median, mean, variance and skewness values for

each color space. Texture features encompass local binary patterns (LBP), histogram of oriented gradients (HOG), information entropy, contrast, and uniformity. Shape features include the length, width, aspect ratio, and area of the entire tongue, the area of both tongue body and coating, as well as the ratio of tongue coating area to tongue body area. CNN features were extracted using the Visual Geometry Group 16 (VGG16) network. The resulting 4096-dimensional feature vector was reduced using principal component analysis (PCA), ultimately yielding a ten-dimensional CNN network feature.

### Tongue Tooth Mark Detection

Before detecting tooth marks, a binary classification model for tongue tooth marks was developed based on the ResNet50 network to obtain images containing tooth marks. The training set included 291 images without teeth marks and 325 images with teeth marks, with the model utilizing five-fold cross-validation.

Tongue tooth mark detection in this project was conducted using the YOLOv8 model, the latest version in the You Only Look Once (YOLO) series of real-time object detectors, renowned for its state-of-the-art accuracy and speed. YOLOv8 uses advanced backbone networks and neck architectures, enhancing feature extraction and object detection performance. Compared to anchor-based methods, it achieves greater precision and detection efficiency. For the training set, two professionals collaboratively annotated the tongue tooth mark locations in tongue images, with only the consistently annotated results included. The final training set comprised 468 annotated images of tongue tooth marks. Additionally, due to the tendency of object detection networks to annotate the same tongue tooth mark multiple times, further correction of the detection results was necessary. This was achieved by calculating the Intersection over Union (IoU) values between different detection boxes. If the IoU value exceeded 0.35, it indicated a duplicate annotation, and the duplicates were removed, retaining only one.

Various features of tongue tooth marks were extracted from the images, including color, texture, shape, and CNN network features. The method for extracting color features is analogous to that used for tongue color. Additionally, the phenotypes of three color spaces and grayscale in tooth marks were calculated as relative values to their corresponding phenotypes in the entire tongue. Shape features encompass the number of tooth marks and the mean, maximum, and minimum values of their width, height, and diagonal lengths. The methods for extracting texture and CNN features are same as those for tongue coating.

## Tongue Fissure Detection

Before segmenting tongue fissures, a binary classification model for tongue fissures was trained using ResNet50 to obtain images containing tongue fissures. The training set included 473 images without fissures and 473 images with fissures, with the model using five-fold cross-validation.

In this study, the SAM-Adapter model was also used for the segmentation of tongue fissures. We used 412 tongue fissure images as training data to train the SAM-Adapter model. To optimize the model's performance, we utilized a pre-trained model for camouflaged objects. After several iterations of training, the model's performance significantly improved, and it was eventually saved and used to segment the tongue fissures in the entire tongue image dataset.

Similarly, we extracted the color, texture, shape, and CNN network features of tongue fissures from the images, employing the same methods used for extracting tooth mark phenotypes.

## Manual Annotation of Classified Tongue Phenotypes

Two experts manually labeled the severity of fissures and tooth marks on the same set of tongue diagnosis images (how many). They categorized tongue color (light red, red, dark red), coating color (white, light yellow, dark yellow), and classified fissures and tooth marks into two categories: light and severe. Ultimately, we constructed a labeled dataset for a quantitative prediction model containing various tongue phenotype images, which includes 142 white tongue coating images, 198 light yellow tongue coating images, 127 dark yellow tongue coating images; 110 light red tongue body images, 149 red tongue body images, 172 dark red tongue body images; 408 light tongue fissure images, 378 severe tongue fissure images; as well as 539 light tongue teeth mark images and 269 severe tongue teeth mark images.

## Prediction of Classified Tongue Phenotypes

Building upon the extracted global and local phenotypes, we investigated the ability to predict classified phenotypes based on this information. To evaluate the performance of our model, the manual labels of tongue body color, tongue coating color, tooth marks, and fissures, as mentioned above, were employed as benchmarks. Only samples agreed on by both experts were included in the dataset used for predicting classified phenotypes. And the labeled dataset was divided into a training set and a test set at an 8:2 ratio. Considering the potential collinearity in such a large number of phenotypes, we applied the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm for feature selection on the training dataset before modeling. Using the default

alpha value of 1, we employed 10-fold cross-validation to determine the optimal value of the regularization parameter lambda, and selected the lambda value that minimized the mean squared error (MSE). This approach reduces the complexity of the model by shrinking some coefficients to zero, thereby selecting the most relevant features, and mitigates the impact of multicollinearity by penalizing large coefficients.

We developed a tongue phenotype classification model using the Random Forest algorithm. Features of tongue coating color, tongue color, tongue fissures, or tongue teeth marks were used as input vectors for training the Random Forest model. Five-fold cross-validation was employed during the training process. Subsequently, the model was validated on the test set, and the area under the curve (AUC) was calculated. The top 10 most important features for predicting tongue phenotype categories were identified based on the feature importance scores from the model. It should be noted that we used random forest rather than other deep learning algorithms to establish the classification model, thus maintaining the interpretability of the classification model and helping users identify the phenotypes most relevant to specific classifications. Additionally, to provide clinicians with valuable probabilistic insights into the diagnostic predictions, we provide both the classification results and the prediction probabilities of the prediction model.

## Predicting Greasy Coating Labels Based on Quantitative Phenotypes

This study explored a novel approach: predicting qualitative tongue phenotypes based on quantitative phenotype data to extend the clinical diagnostic applications of quantitative tongue phenotypes and apply this prediction model to a broader range of tongue phenotype labels. To validate this approach, the study applied it to a dataset from a published paper (Wang et al. 2022). The paper included a dataset of tongue diagnosis images and corresponding greasy coating labels, classifying the images into non-greasy, greasy, and thick greasy categories. We applied the tongue image phenotype feature extraction methods developed in this study to these diagnostic images, extracting color, texture, and shape features of the tongue coating, tongue body, and entire tongue using a series of algorithms, as well as deep features using CNN. Based on the quantitative phenotype data extracted from these images, we further conducted quality control and feature selection on these phenotypes to select the phenotypes included in predictive model. This study built a greasy coating prediction model using the random forest algorithm. The labeled dataset is divided into three categories including 85 non-greasy tongue images, 759

greasy tongue images, and 642 thick-greasy tongue images, and split into a training set and a test set in an 8:2 ratio.

## Results

### Tongue Segmentation

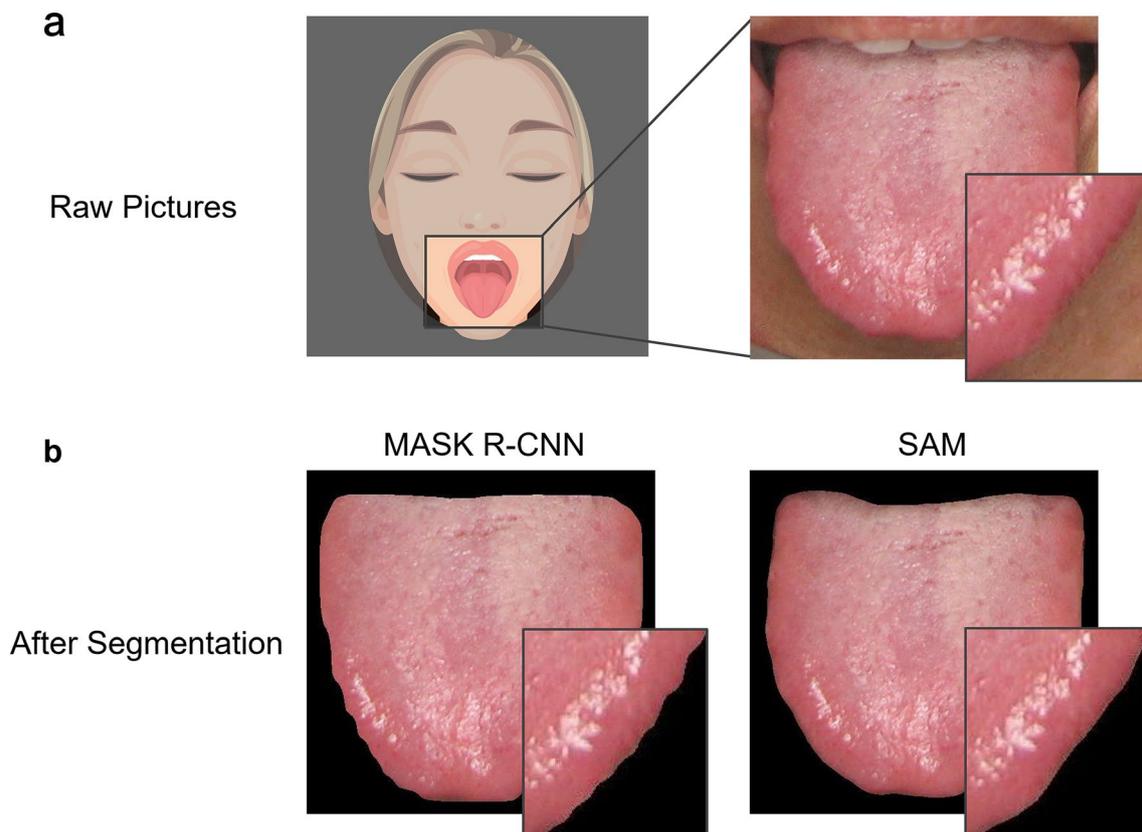
Employing a combined approach of Grounding DINO (object detection) and SAM (segmentation), we achieved precise tongue region segmentation in the original images. Our method offers a clear advantage in edge segmentation compared to the widely used Mask region-based CNN (Mask R-CNN) network. As illustrated in Fig. 2, Mask R-CNN generates wavy segmentation edges around the tongue, which could hamper the subsequent extraction of tongue tooth marks. In contrast, SAM produces smoother edges, leading to more accurate analysis.

Considering that clinicians often focus on specific regions of the tongue, we referred to previous literature and further divided the tongue into five subregions (tongue bottom, tongue tip, tongue center, and the two tongue margins) to enhance clinical applicability. For example, in the case of color phenotypes, we extracted the average color of these five subregions in JD cohort (Fig. S1a). Visually,

the tongue tip subregion appeared the reddest (usually due to less coating), the tongue bottom subregion appeared the darkest (usually due to thicker and darker coating), and the colors of the two tongue margins were similar. Additionally, we pairwise compared the grayscale values of the five subregions. As expected, the results showed statistically significant differences between all subregions except the two tongue margins (Fig. S1b). These results showed that our extracted phenotypes can effectively capture the detailed features of different regions of the tongue.

### Tongue Phenotypes Extraction

After completing the segmentation of the tongue body, we extracted the phenotypes of coating color, tongue color, fissures, and tooth marks. Before extracting the coating and tongue phenotypes, it was necessary to remove reflective spots and segment the entire tongue into tongue body and coating. The tongue coating and tongue body were properly distinguished, as shown in Fig. S2, the correlation between coating color and tongue color was significantly lower than the internal correlation within the coating and tongue color themselves. The correlations of all color phenotypes of tongue coat and tongue substance are shown in Fig. S3.



**Fig. 2** Tongue segmentation comparison of Mask R-CNN (a) and SAM (b)

**Table 1** Five-fold cross validation results of tongue fissure binary classification network

	Loss	Accuracy (%)	Precision (%)	Recall (%)	F1 score
Fold 1	0.005	99.47	98.75	100	99.37
Fold 2	0.007	98.41	97.25	100	98.61
Fold 3	0.003	98.94	98.10	100	99.04
Fold 4	0.003	97.88	96.91	98.95	97.92
Fold 5	0.040	98.41	98.00	98.99	98.49
Average	0.011	98.62	97.80	99.59	98.69

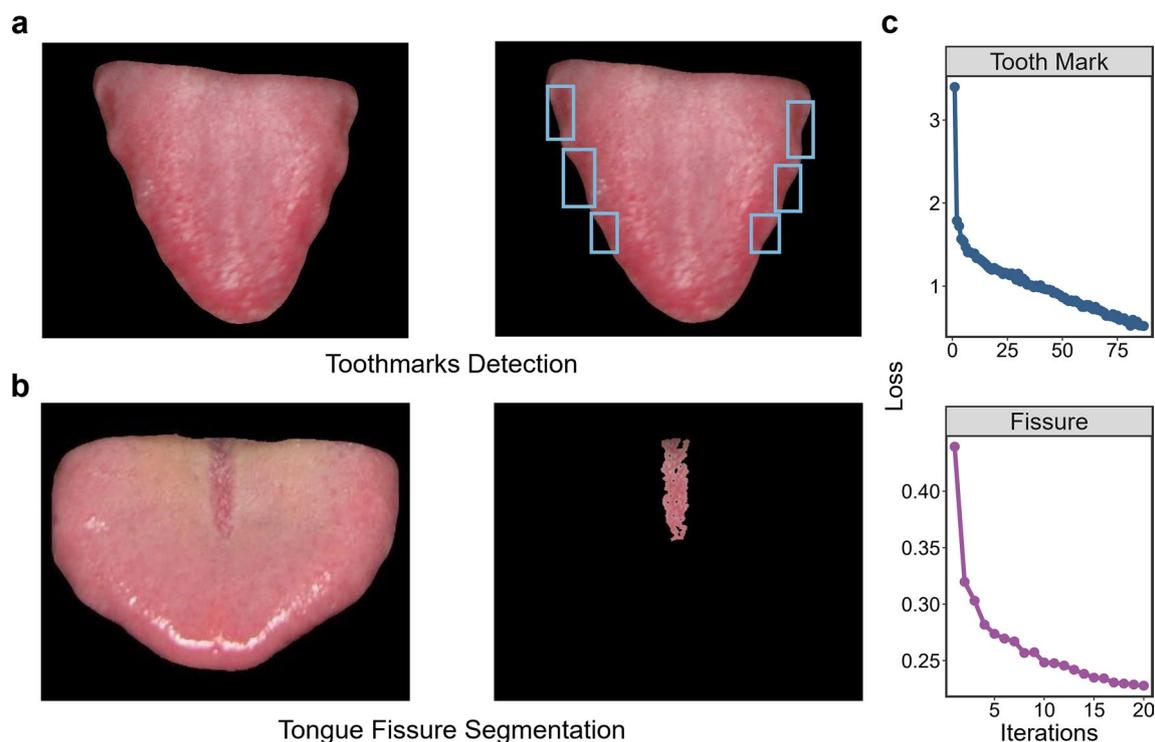
**Table 2** Five-fold cross validation results of tooth mark binary classification network

	Loss	Accuracy (%)	Precision (%)	Recall (%)	F1 score
Fold 1	0.037	98.37	100.00	96.61	98.28
Fold 2	0.013	99.19	98.33	100.00	99.16
Fold 3	0.059	98.37	100.00	96.67	98.31
Fold 4	0.059	97.56	100.00	94.34	97.09
Fold 5	0.010	99.19	98.39	100.00	99.19
Average	0.036	98.54	99.34	97.52	98.41

Our deep learning models achieved impressive results in classifying tongue fissures and tooth marks (Tables 1 and 2). The fissure classification model reached an average accuracy of 98.62%, demonstrating high sensitivity and specificity (ability to correctly identify both positive and negative cases). Precision (accuracy of positive identifications) was 97.80%, recall (ability to detect all positive cases) was

99.59%, and F1-score (balanced metric for precision and recall) was 98.69%. Similarly, the tooth mark classification model achieved an average accuracy of 98.54% with high precision (99.34%), recall (97.52%), and F1-score (98.41%). These results indicate that our neural networks effectively extracted tongue fissures and tooth marks (Fig. 3a and b). Training successfully optimized the models, as evidenced by the steadily decreasing loss values for both YOLOv8 and SAM-Adapter networks (Fig. 3c and d). After training, the SAM-Adapter network achieved a final mean IoU (mIoU) of 0.72 for fissure segmentation, indicating good overlap between predicted and actual segmentation masks. At a threshold of 0.5, the mean Average Precision (mAP) of the tooth mark detection network reached 0.83, signifying high average precision in identifying tooth marks. To describe the position and distribution of the tooth marks, our phenotype includes the carefully extracted number, relative width, and height of the tooth marks. The correlogram in Fig. S4 shows that the detected tooth marks are mainly distributed on the left and right margins and the lower middle part of the entire image, which aligns with the actual distribution of tooth marks primarily on the tongue margins and tip.

To assess the reliability of our phenotypic quantification method, we captured two consecutive images of the same sample within a short interval. As shown in Fig. S5b, all 11 samples exhibited highly significant correlations between the two measurements (Pearson  $r > 0.98$ ,  $p < 0.05$ ). Different

**Fig. 3** Tongue tooth mark and fissure extraction results. (a) An example of tongue tooth mark extraction results. (b) An example of tongue fissure extraction results. (c) Loss curve of the tongue tooth mark detection network (top) and the tongue fissure segmentation network (bottom)

classes of phenotypes were selected as examples (Fig. S5c) to further demonstrate that there were no significant differences between the two measurements ( $p > 0.05$ , paired Wilcoxon signed-rank test), highlighting the robustness of our model.

### Correlation of Extracted Phenotypes and Manual Labels

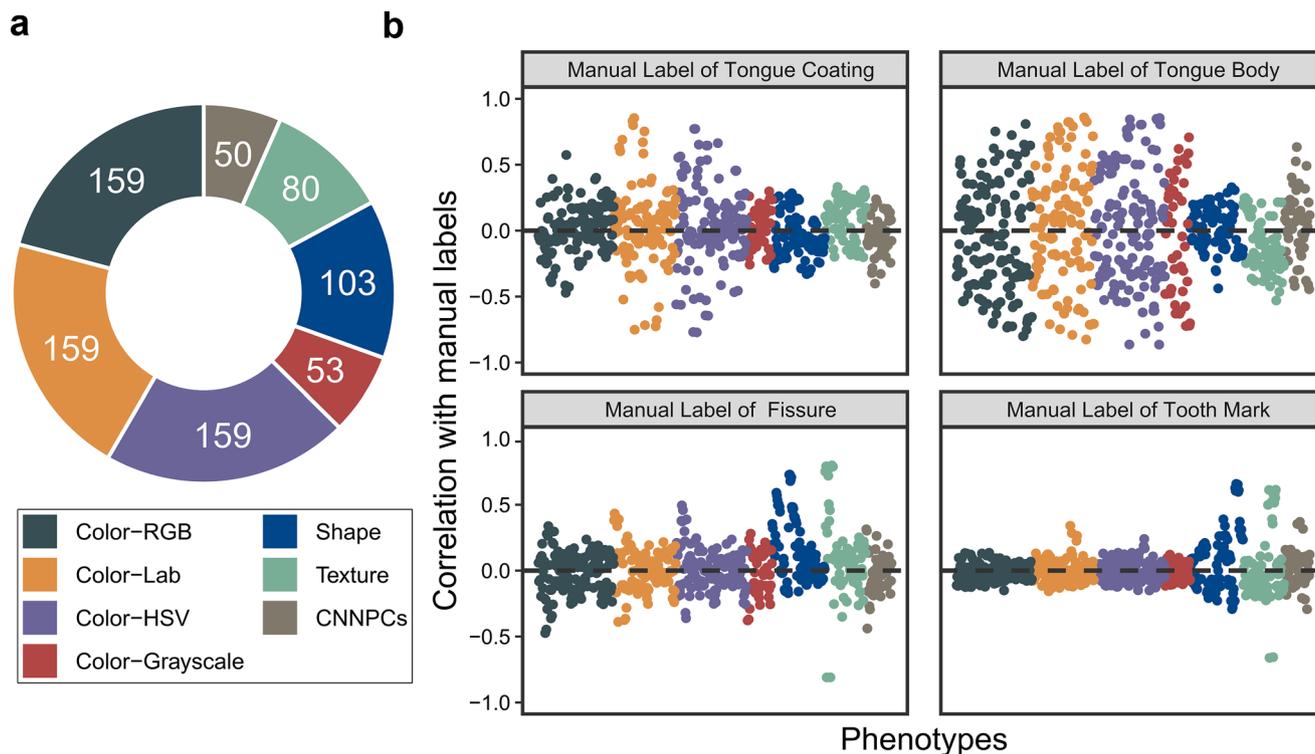
TonguExpert extracts a total of 763 phenotypes including 355 global features (entire tongue, tongue body, and tongue coating) and 408 local features (fissures, tooth marks). The 763 quantitative phenotypes were used later in the classification prediction model, as shown in Fig. 4a. There are 530 color phenotypes extracted, including 90 phenotypes each for the entire tongue, tongue body and coating, and 130 phenotypes each for tooth mark and fissure. There are 103 shape phenotypes extracted, including three phenotypes for tongue body and coating (area of tongue body and coating, the ratio of tongue coating area to tongue body area), four phenotypes for the entire tongue (area, width, length and aspect ratio), and 48 phenotypes each for tooth mark and tongue fissure. There are 80 texture phenotypes extracted, including 16 phenotypes each for the entire tongue, tongue body, tongue coating, tooth mark and tongue fissure. There are 50 CNN feature phenotypes extracted, including 10 each for the entire tongue, tongue body, tongue coating, tooth

mark and tongue fissure. A detailed description of all phenotypes is available in File S2.

We examined the correlation between all quantitative phenotypes and manual assigned labels for traditional classified phenotypes. The results (Fig. 4b) revealed that, as expected, tongue coating and tongue body color labels were strongly correlated with color-based quantitative phenotypes, while tongue fissure and tooth marks labels were primarily associated with shape and texture phenotypes but not color. Furthermore, unlike tongue body label, the correlation patterns of tongue coating label with grayscale and three color space phenotypes were inconsistent. This suggests that extracting more fine-grained color phenotypes may help us differentiate between different tongue phenotypes.

### Prediction of Classified Tongue Phenotypes

To evaluate how well our extracted features predict traditional classified tongue phenotypes, we compared model predictions for the color levels of tongue body and tongue coating, as well as the severity levels of tongue fissures and tooth marks, with manually annotated labels. We fed a series of previously extracted variables (coating color, tongue color, tooth marks, and fissures) into a random forest model to predict their classifications. Initially, two experts manually assessed and classified these phenotypes. The classification between the two raters for the same sample is



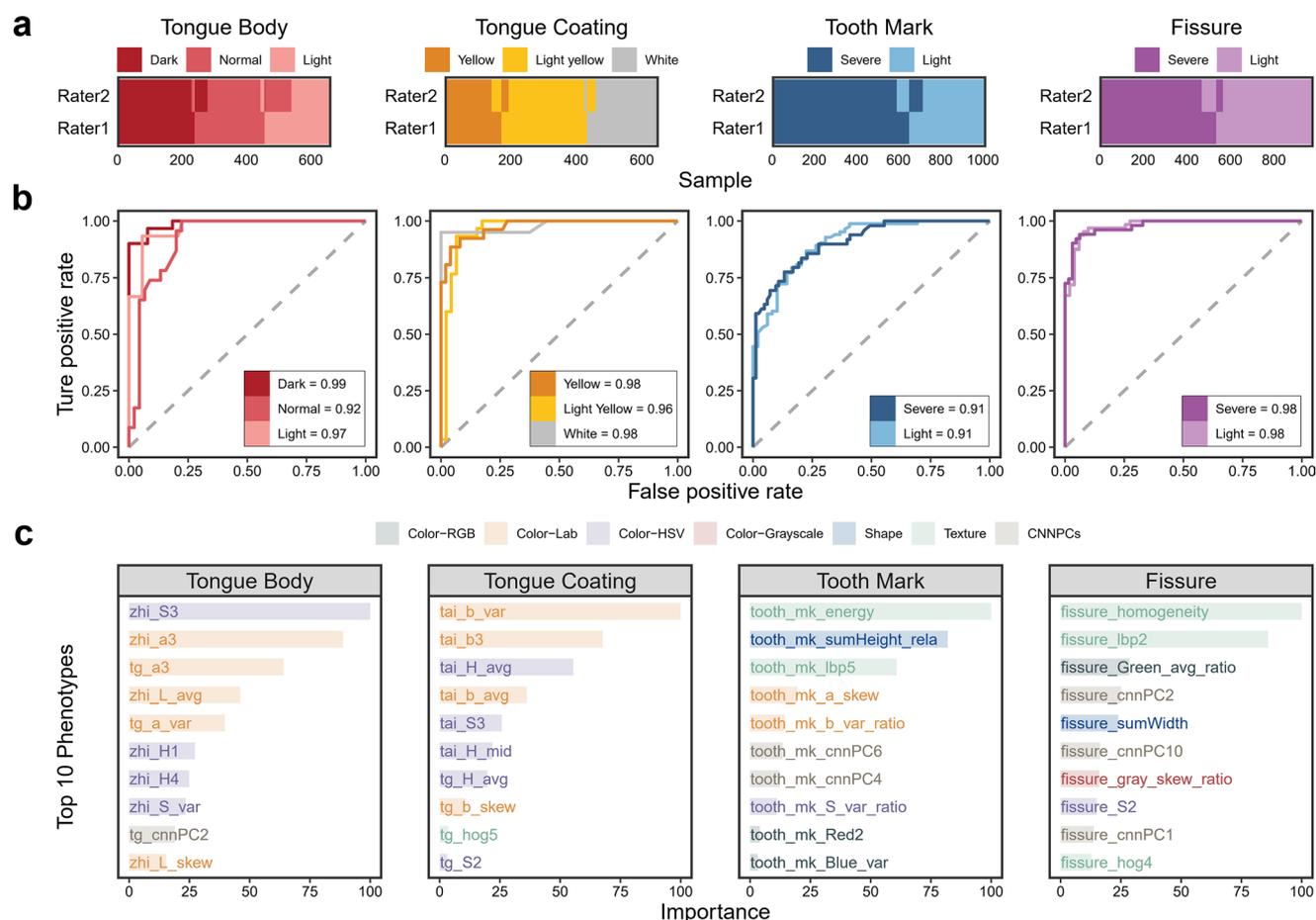
**Fig. 4** Count of extracted quantitative phenotypes (a) and Spearman correlation coefficient ( $r$  value) between these phenotypes and four manual labels (b)

shown in Fig. 5a. Only images with consistent annotations by both raters were included in the random forest training set. The model's performance was evaluated using receiver operating characteristic (ROC) curves and AUC (shown in Fig. 5b). Here, the ROC-AUC values for coating color classification ranged between 0.96 and 0.98, for tongue body color between 0.92 and 0.99, for fissures it was 0.98, and for tooth marks it was 0.91. These high values indicate good discrimination ability.

We identified the top 10 most important features used by the model for each prediction (Fig. 5c). As expected, features related to the tongue body itself dominated the model predicting tongue body color, highlighting the accuracy of our extractions. Consistent with the correlation analysis, tongue body and coating were primarily associated with color phenotypes (H and S value of HSV, L and a value of Lab for body; H and S value of HSV, b value of Lab for coating). Fissure and tooth mark predictions relied more on shape or texture features.

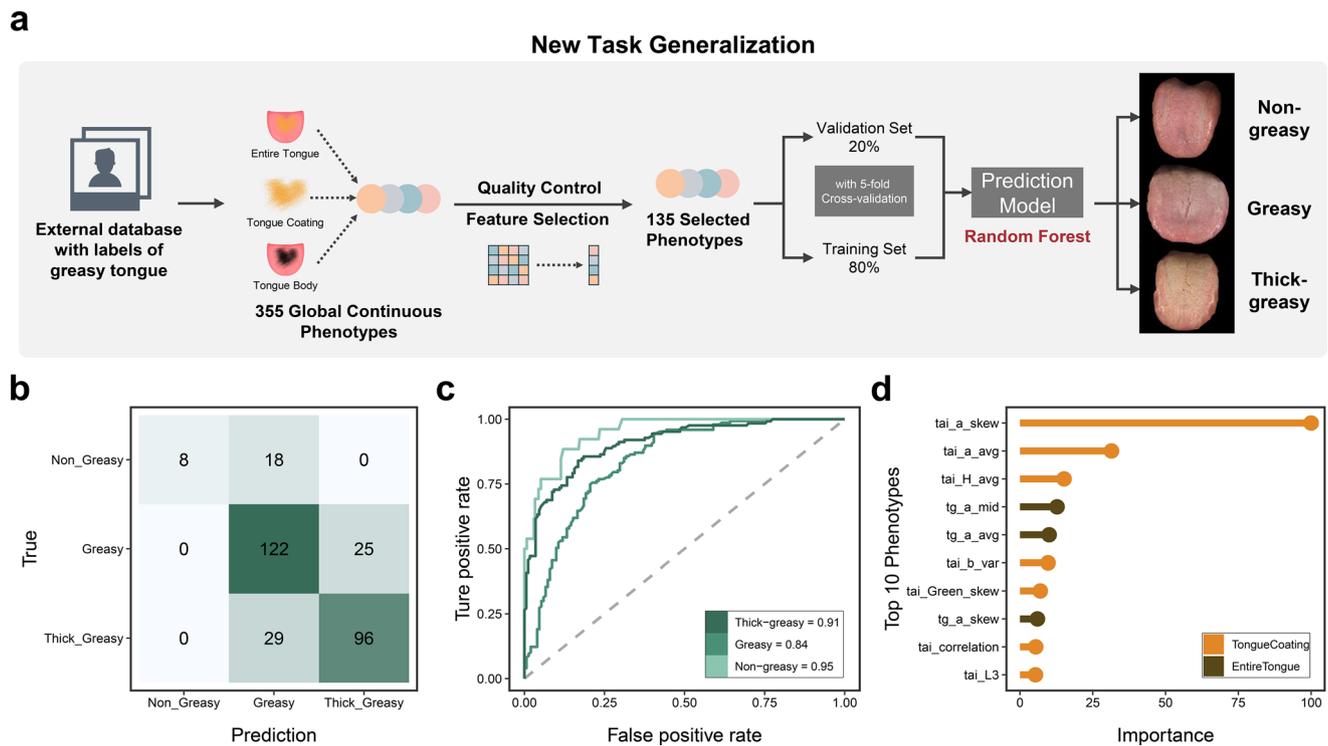
## Prediction of Greasy Coating Phenotypes as a Novel Phenotype in an External Dataset

To evaluate the model's ability to handle unseen data (generalization), we applied our method to an external dataset to extract the global phenotypes (color, shape, texture and CNN feature phenotypes of the entire tongue, tongue body and coating) and predict greasy coating (non-greasy, greasy, and thick greasy) based on these phenotypes (Fig. 6a). We extracted 355 phenotypes, after quality control and feature engineering, 135 phenotypes remained. Interestingly, 29 of these features were related to color, and only one was related to texture. Additionally, the distribution of these features is informative: 15 features originated from the tongue coating itself, 10 from the entire tongue, and just five from the tongue body, highlighting the importance of the tongue coating for greasy coating classification. The model's predictions (horizontal axis) and the actual classifications (vertical axis) are in high correlation (Fig. 6b), and the ROC-AUC values indicate excellent performance:



**Fig. 5** Tongue phenotype classification results. **(a)** Manual classification labels for tongue color, coating color, fissures, and tooth marks phenotypes. **(b)** ROC curves for the model predictions of classified phenotypes of tongue color, coating color, fissures, and tooth marks.

**(c)** Top 10 most important phenotypes in each predictive model. The colors of bars represent the region of the tongue from which the phenotype was extracted



**Fig. 6** Greasy tongue phenotype classification results. **(a)** Process for predicting greasy coating classification using global quantitative phenotypes. **(b)** Confusion matrix for the model predictions of greasy

0.95 for non-greasy, 0.84 for greasy, and a remarkable 0.91 for thick greasy coatings (Fig. 6c). Notably, the top three most influential features all came from the tongue coating, further emphasizing the model’s focus on relevant areas (Fig. 6d). These results showcase the power of combining feature selection with machine learning models. By extracting informative phenotypes from tongue images, the model can effectively predict greasy coating classifications. This approach has significant advantages for applications in, for example, traditional medicine or automated diagnostic tools.

## Discussion

Tongue diagnosis is an effective and non-invasive method that does not rely on instruments and is not limited by location, contributing to the establishment of primary healthcare (Wen et al. 2020). However, the utilization of tongue images is contingent upon the prior knowledge and clinical experience of doctors, resulting in a lack of objectivity and robustness in tongue diagnosis results (Matos et al. 2021; Wang et al. 2020). Previous studies have made significant efforts towards standardizing tongue diagnosis, but fragmented and coarse-grained feature analyses impede the accumulation of knowledge regarding tongue characteristics. While

coating. **(c)** ROC curves for the model predictions of greasy coating. **(d)** Top 10 important phenotypes in model predicting greasy coating

data-driven models exhibit excellent performance, the lack of interpretability in features limits clinical applications and generalization to new phenotypes. Additionally, scarce data impedes model comparison and advancement. To address these challenges, our study introduces TonguExpert (<https://www.biosino.org/TonguExpert>), a free, automated platform for archiving, analyzing and extracting corresponding phenotypes from tongue images.

Current methods typically analyze tongue features in isolation (color, morphology, local features) neglecting potential synergy for a more comprehensive assessment. This study proposes an automated framework for tongue phenotype recognition and classification using deep learning algorithms. The comprehensive framework integrated algorithms such as DINO and SAM for tongue segmentation, and Yolov8, ResNet50 for tongue phenotype extraction. Leveraging a large-scale dataset of natural individuals, including 5992 de-identified tongue images and phenotypes from a Chinese population, our approach extracts 773 phenotypes including 355 global features (entire tongue, tongue body, and tongue coating) and 408 local features (fissures, tooth marks) in a unified process. Besides, 580 additional features for five tongue subregions are also available for future study.

In addition to the features extracted from the primary dataset, our model can also be further enhanced by incorporating

external labels. Notably, TonguExpert achieves robust performance (ROC-AUC 0.92–0.99 for color, 0.98 for fissures, 0.91 for tooth marks) demonstrating its stability and reliability. Interestingly, the correlation analysis showed that different color space phenotypes may vary in importance across different prediction tasks, suggesting that finer-grained phenotypes could improve our ability to describe unseen phenotypes more accurately in the future. Furthermore, the model's generalization ability was validated using an external dataset for predicting greasy coating phenotypes. This yielded promising results (ROC-AUC 0.84–0.95) with good interpretability, highlighting the potential of integrating artificial intelligence into tongue diagnosis for a more objective, reliable, and efficient approach.

Although deep learning models have achieved impressive performance in previous tongue diagnosis studies, the black-box nature of deep learning has also led to some challenges (Liu et al. 2023; Yuan et al. 2023): in the clinical, it may hinder the interpretability of results, limiting the progress of clinical applications; while in the laboratory, it makes it difficult for us to accumulate understanding of tongue features, which is not conducive to further research on tongue diagnosis. Therefore, in this study, we used a variety of deep learning models to extract as many and as fine-grained phenotypes as possible, and used traditional machine learning models (random forest) for prediction, to balance the performance and interpretability of the models. The phenotypes we extracted form a universal phenotype library, which can serve as the basis for interpreting unseen phenotypes in the future, helping us generalize clinical observations, monitor the early stages of diseases, and enabling more targeted research into the biological mechanisms of tongue features.

Limited data hinders the robustness and generalizability of deep learning models, while the lack of publicly available datasets impedes model comparison and advancement. To address these challenges, we introduce the largest known dataset of tongue images. This dataset includes tongue region images, corresponding mask images, and extracted tongue phenotype labels with associated information. The dataset and the developed algorithms are integrated into a user-friendly website, allowing users to upload tongue images for analysis and access the standardized dataset for further exploration. This research advances automated tongue diagnosis, paving the way for its wider adoption in clinical settings and potentially expanding the applications of tongue diagnosis in the future.

## Limitations and Future Directions

Additional high-quality manual labels can further enhance model performance. While satisfactory results were achieved with different devices, mobile phone images may

not perform as well. Future work will focus on improving model compatibility with mobile images.

Current studies on the genetics and multi-omics of tongue phenotypes are scarce. The next step is to analyze these aspects using multi-omics approaches to understand individual differences in tongue phenotypes and their link to diseases.

## Conclusion

This study introduces TonguExpert (<https://www.biosino.org/TonguExpert>), a platform for archiving tongue images and extracting corresponding phenotypes and the large publicly available dataset hold significant promise for advancing research in tongue image analysis and diagnosis. This study demonstrates the effectiveness of combining feature selection with deep learning for extracting informative features from tongue images and accurately predicting tongue phenotypes, paving the way for its wider adoption in clinical settings and potentially expanding the applications of tongue diagnosis in the future.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s43657-024-00210-9>.

**Acknowledgements** We are grateful to X.Chen from Taizhou Institute of Health Sciences of Fudan University, Y. Fan from Human Phenome Institute of Fudan University, and Y. Hu from Shanghai Institute of Nutrition and Health for sample information and tongue images collection. We thank Y. Ling from Shanghai Institute of Nutrition and Health for help in developing the TonguExpert website. We also thank the participants of the NSPT and JD cohorts who consented to participate in research, and the related teams, including interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

**Authors' Contributions** S. Wang, G. Zhang, Q. Peng and J. Chen conceived the project and revised this work. T. Li built the main model framework, G. Zhai and Z. Liu contributed to the constructed spicularity removal algorithm. L. Zuo and T. Li labeled the classification of phenotypes of TonguExpert dataset, performed the data analysis, data visualization, and draft the manuscript. T. Li, L. Zuo, P. Wang, L. Yang, Q. Peng and G. Zhang contributed to the development of the TonguExpert online platform. L. Jin, J. Wang, J. Tan and Y. Yang contributed to the design and acquisition of data in NSPT cohort. Y. Zhou contributed to the design and acquisition of data in JD cohort. X. Wang and J. Chen contributed to the design and acquisition of data in the greasy coating dataset. All authors approved the final version of the manuscript.

**Funding** The authors acknowledge the following funding sources: the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant no. XDB38020400 to S. Wang and XDB38030100 to G. Zhang), CAS Youth Innovation Promotion Association (2020276 to Q. Peng), CAS Young Team Program for Stable Support of Basic 488 Research (YSBR-077 to S. Wang), Hutian Scholar funded by the Beijing University of Chinese Medicine (2022-XJ-KYQD-003 to J.

Chen), CAS Interdisciplinary Innovation Team to S. Wang, Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01 to S. Wang, L. Jin, J. Wang, J. Tan, and Q. Peng; 2023SHZDZX02 to L. Jin and G. Zhang), the National Natural Science Foundation of China (NSFC) (32325013, 92249302 to S. Wang, 32271186 to J. Tan, 82004222 to X. Wang, 82025036 to J. Chen), the National Key Research and Development Project (2018YFC0910403 to S. Wang), Ministry of Science and Technology of the People's Republic of China (2015FY111700 to L. Jin), Shanghai Science and Technology Commission Excellent Academic Leaders Program (22XD1424700 to S. Wang), 111 Project (B13016) to L. Jin, and CAMS Innovation Fund for Medical Science (2019-I2M-5-066 to J. Wang and L. Jin).

**Data Availability** The dataset used for modeling in our study is available at <https://www.biosino.org/TonguExpert>. The external validation dataset of greasy coating can be obtained according to the original publication.

## Declarations

**Ethical Approval** All study protocols were approved by the institutional review boards of the pertinent research institutions. This TonguExpert dataset consists of data from the project which was approved by the Ethics Committee of Human Genetic Resources at the Shanghai Institute of Life Sciences, Chinese Academy of Sciences (ER-SIBS-261410-A1801) and project of The National Science & Technology Basic Research Project which was approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (No. 14117).

**Consent to Participate** Written informed consent was obtained from all individual participants included in the study.

**Consent for Publication** All the participants approved to publish.

**Competing Interests** The authors declare that there is no conflict of interest. L. Jin is the Editor-in-Chief of Phenomics, and he was not involved in reviewing this paper.

## References

- Cao G, Ding J, Duan Y et al (2016) Classification of tongue images based on doublet and color space dictionary. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM):1170–1175. <https://doi.org/10.1109/bibm.2016.7822686>
- Chen Q, Huang X, Zhang H et al (2024) Characterization of tongue coating microbiome from patients with colorectal cancer. *J Oral Microbiol* 1(1):2344278. <https://doi.org/10.1080/20002297.2024.2344278>
- Chung CJ, Wu CH, Hu WL et al (2023) Tongue diagnosis index of chronic kidney disease. *Biomed J* 4(1):170–178. <https://doi.org/10.1016/j.bj.2022.02.001>
- Hu MC, Lan KC, Fang WC et al (2019) Automated tongue diagnosis on the smartphone and its applications. *Comput Methods Programs Biomed* 174:51–64. <https://doi.org/10.1016/j.cmpb.2017.12.029>
- Huang YS, Wu HK, Chang HH et al (2022) Exploring the pivotal variables of tongue diagnosis between patients with acute ischemic stroke and health participants. *J Tradit Complement Med* 1(5):505–510. <https://doi.org/10.1016/j.jtcme.2022.04.001>
- Kim J, Han G, Ko SJ et al (2014) Tongue diagnosis system for quantitative assessment of tongue coating in patients with functional dyspepsia: a clinical trial. *J Ethnopharmacol* 15(1):709–713. <https://doi.org/10.1016/j.jep.2014.06.010>
- Kirillov A, Mintun E, Ravi N et al (2023) Segment anything. *arXiv e-prints arXiv:230402643*. <https://doi.org/10.48550/arXiv.2304.02643>
- Liu H, Feng Y, Xu H et al (2022) MEA-Net: multilayer edge attention network for medical image segmentation. *Sci Rep* 1(1):7868. <https://doi.org/10.1038/s41598-022-11852-y>
- Liu J, Dan W, Liu X et al (2023) Development and validation of predictive model based on deep learning method for classification of dyslipidemia in Chinese medicine. *Health Inf Sci Syst* 1(1):21. <https://doi.org/10.1007/s13755-023-00215-0>
- Liu Q, Shi K, Bai Y et al (2024) Biology of tongue coating in different disease stages of RA and its value in disease progression. *Microb Pathog* 191:106644. <https://doi.org/10.1016/j.micpath.2024.106644>
- Lo LC, Cheng TL, Huang YC et al (2012) Analysis of agreement on traditional Chinese medical diagnostics for many practitioners. *Evid Based Complement Alternat Med* 2012:178081. <https://doi.org/10.1155/2012/178081>
- Lynge Pedersen AM, Nauntofte B, Smidt D et al (2015) Oral mucosal lesions in older people: relation to salivary secretion, systemic diseases and medications. *Oral Dis* 2(6):721–729. <https://doi.org/10.1111/odi.12337>
- Matos LC, Machado JP, Monteiro FJ et al (2021) Can Traditional Chinese Medicine Diagnosis Be Parameterized and Standardized? A Narrative Review. *Healthcare (Basel)* (2). <https://doi.org/10.3390/healthcare9020177>
- Monshi B, Grabovac S, Gulz L et al (2021) Psoriasis is associated with fissured tongue but not geographic tongue: a prospective, cross-sectional, case-control study. *J Dtsch Dermatol Ges* 1(8):1170–1176. <https://doi.org/10.1111/ddg.14451>
- Morita A, Murakami A, Noguchi K et al (2021) Combination image analysis of Tongue Color and Sublingual Vein improves the diagnostic accuracy of Oketsu (blood stasis) in Kampo Medicine. *Front Med (Lausanne)* 8:790542. <https://doi.org/10.3389/fmed.2021.790542>
- Peng Q, Liu X, Li W et al (2024) Analysis of blood methylation quantitative trait loci in East asians reveals ancestry-specific impacts on complex traits. *Nat Genet* 5(5):846–860. <https://doi.org/10.1038/s41588-023-01494-9>
- Picciani BLS, Teixeira-Souza T, Pessoa TM et al (2018) Fissured tongue in patients with psoriasis. *J Am Acad Dermatol* 7(2):413–414. <https://doi.org/10.1016/j.jaad.2017.08.020>
- Shao Q, Li XQ, Fu ZC (2014) Recognition of teeth-marked Tongue based on gradient of Concave Region. 2014 Int Conf Audio Lang Image Process (Icalip) 1–2:968–972. <https://doi.org/10.1109/ICALIP.2014.7009938>
- Song H, Huang Z, Feng L et al (2022) RAFF-Net: an improved tongue segmentation algorithm based on residual attention network and multiscale feature fusion. *Digit Health* 8:20552076221136362. <https://doi.org/10.1177/20552076221136362>
- Sun Y, Dai S, Li J et al (2019) Tooth-marked Tongue Recognition using gradient-weighted class activation maps. *Future Internet* 1(2):45. <https://doi.org/10.3390/fi11020045>
- Tang W, Gao Y, Liu L et al (2020) An automatic recognition of tooth-marked Tongue based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning. *IEEE Access* 8:153470–153478. <https://doi.org/10.1109/ACCESS.2020.3017725>
- Tang YS, Guo JC, Xu L et al (2022) Pathological change of Chronic Hepatitis B patients with different Tongue Coatings by Circular Multi-omics Integrated Analysis. *Chin J Integr Med* 2(1):28–35. <https://doi.org/10.1007/s11655-020-3275-4>

- Tania MH, Lwin K, Hossain MA (2019) Advances in automated tongue diagnosis techniques. *Integr Med Res* 142–56. <https://doi.org/10.1016/j.imr.2018.03.001>
- Wang X, Zhang B, Yang Z et al (2013) Statistical analysis of tongue images for feature extraction and diagnostics. *IEEE Trans Image Process* 2(12):5336–5347. <https://doi.org/10.1109/TIP.2013.2284070>
- Wang H, Zhang XF, Cai YH (2014) Research on Teeth Marks Recognition in Tongue Image. 2014 International Conference on Medical Biometrics (Icmb 2014):80–84. <https://doi.org/10.1109/Icmb.2014.21>
- Wang ZC, Zhang SP, Yuen PC et al (2020) Intra-rater and Inter-rater Reliability of Tongue Coating diagnosis in traditional Chinese Medicine using Smartphones: Quasi-delphi Study. *JMIR Mhealth Uhealth* 7e16018. <https://doi.org/10.2196/16018>
- Wang X, Wang X, Lou Y et al (2022) Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in noninvasive ethnopharmacological evaluation. *J Ethnopharmacol* 285:114905. <https://doi.org/10.1016/j.jep.2021.114905>
- Wen G, Ma J, Hu Y et al (2020) Grouping attributes zero-shot learning for tongue constitution recognition. *Artif Intell Med* 109:101951. <https://doi.org/10.1016/j.artmed.2020.101951>
- Xia X, Chen X, Wu G et al (2020) Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle. *Nat Metab* 9946–957. <https://doi.org/10.1038/s42255-020-00270-x>
- Xie Y, Zhou P, Gao Y et al (2023) Chronic mucocutaneous candidiasis presenting as refractory fissured tongue in a patient with IL-17RC mutation: the first reported case of Chinese ethnicity. *Emerg Microbes Infect* 1(2):2231567. <https://doi.org/10.1080/221751.2023.2231567>
- Xu J (2017) Clinical illustration of Tongue diagnosis of traditional Chinese medicine. Chemical Industry, Beijing
- Xu Q, Zeng Y, Tang W et al (2020) Multi-task Joint Learning Model for Segmenting and Classifying Tongue images using a deep neural network. *IEEE J Biomed Health Inf* 2(9):2481–2489. <https://doi.org/10.1109/JBHI.2020.2986376>
- Yanagisawa K, Takagi I, Sakurai K (2007) Influence of tongue pressure and width on tongue indentation formation. *J Oral Rehabil* 3(11):827–834. <https://doi.org/10.1111/j.1365-2842.2007.01734.x>
- Yuan Y, Liao W (2021) Design and implementation of the traditional Chinese Medicine Constitution System based on the diagnosis of Tongue and Consultation. *Ieee Access* 9:4266–4278. <https://doi.org/10.1109/Access.2020.3047452>
- Yuan L, Yang L, Zhang S et al (2023) Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. *EClinicalMedicine* 57:101834. <https://doi.org/10.1016/j.eclinm.2023.101834>
- Yun SJ, Lee JB, Kim SJ et al (2007) Recurrent geographical tongue and fissured tongue in association with pregnancy. *J Eur Acad Dermatol Venerol* 2(2):287–289. <https://doi.org/10.1111/j.1468-3083.2006.01861.x>
- Zhou J, Li S, Wang X et al (2022) Weakly supervised deep learning for tooth-marked Tongue Recognition. *Front Physiol* 13:847267. <https://doi.org/10.3389/fphys.2022.847267>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Ting Li<sup>1</sup>  · Ling Zuo<sup>2</sup>  · Pengyu Wang<sup>1</sup>  · Liangfu Yang<sup>1</sup>  · Zijia Liu<sup>3</sup> · Xu Wang<sup>4</sup>  · Jingze Tan<sup>5,6</sup>  · Yajun Yang<sup>5,6</sup>  · Jiucun Wang<sup>5,7,8</sup>  · Yong Zhou<sup>9</sup> · Li Jin<sup>5,7,8</sup>  · Guangtao Zhai<sup>3</sup> · Jianxin Chen<sup>2,4</sup> · Qianqian Peng<sup>1</sup>  · Guoqing Zhang<sup>1</sup>  · Sijia Wang<sup>1,10</sup> 

✉ Jianxin Chen  
cjsx@bucm.edu.cn

✉ Qianqian Peng  
pengqianqian@sinh.ac.cn

✉ Guoqing Zhang  
gqzhang@sinh.ac.cn

✉ Sijia Wang  
wangsjia@sinh.ac.cn

<sup>1</sup> CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup> School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China

<sup>3</sup> Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, 800 Dongchuan RD, Shanghai 200240, China

<sup>4</sup> School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China

<sup>5</sup> State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, Human Phenome Institute, Fudan University, Shanghai 200433, China

<sup>6</sup> Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China

<sup>7</sup> Taizhou Institute of Health Sciences, Fudan University, Taizhou, Jiangsu 225316, China

<sup>8</sup> Research Unit of Dissecting the Population Genetics and Developing New Technologies for Treatment and Prevention of Skin Phenotypes and Dermatological Diseases (2019RU058), Chinese Academy of Medical Sciences, Beijing 100730, China

<sup>9</sup> Clinical Research Institute, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

<sup>10</sup> Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650201, China