Berry S, Giraldo NA, Green BF, Cottrell TR, Stein JE, Engle EL, et al. Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. Science 2021;372:eaba2609.

Fujimura T, Sato Y, Tanita K, Kambayashi Y, Otsuka A, Fujisawa Y, et al. Serum levels of soluble CD163 and CXCL5 may be predictive markers for immune-related adverse events in patients with advanced melanoma treated with nivolumab: a pilot study. Oncotarget 2018;9:15542–51.

Goodman RS, Lawless A, Woodford R, Fa'ak F, Tipirneni A, Patrinely JR, et al. Extended follow-up of chronic immune-related adverse events following adjuvant anti-PD-1 therapy for high-risk resected melanoma. JAMA Netw Open 2023;6:e2327145.

Jimenez-Sanchez D, Will EM, Lai J, Warrier G, Lipson EJ, Deutsch JS, et al. 1024MO Interpretable AI-identified spatial tumor microenvironment (TME) neighborhoods associate with severe immune-related adverse events (irAE) and progressive disease (PD) in patients with melanoma treated with anti-PD-1-based therapy. Ann Oncol 2023;34:S623.

Knight A, Karapetyan L, Kirkwood JM. Immunotherapy in melanoma: recent advances and future directions. Cancers (Basel) 2023;15:1106.

Lu S, Stein JE, Rimm DL, Wang DW, Bell JM, Johnson DB, et al. Comparison of biomarker modalities for predicting response to PD-1/PD-L1 checkpoint blockade: a systematic review and meta-analysis. JAMA Oncol 2019;5:1195–204.

Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315–6.

Nikas JB, Low WC. ROC-supervised principal component analysis in connection with the diagnosis of diseases. Am J Transl Res 2011;3:180–96.

Thompson LL, Krasnow NA, Chang MS, Yoon J, Li EB, Polyakov NJ, et al. Patterns of cutaneous and noncutaneous immune-related adverse events among patients with advanced cancer. JAMA Dermatol 2021;157:577–82.

GWASs of the Nasolabial Fold Identified Variants Related to Genes that Also Affect Facial Morphology

Journal of Investigative Dermatology (2026) **146,** 1146–1148; doi:10.1016/j.jid.2025.09.382

TO THE EDITOR

The nasolabial fold (NLF) is a prominent dermatological phenotype of the aging midface. Previous anatomical studies have clarified that the NLF is potentially induced by the aging changes in the superficial musculoaponeurotic system architecture, cutaneous ligament, midface musculature and fat compartments, and craniofacial skeleton (Coleman and Grover, 2006; Cotofana et al, 2016). The NLF plays a critical role in facial esthetics and is a key feature in perceived aging (Mayes et al, 2010). It also predicts certain craniofacial abnormalities linked to genetic factors (Cha et al, 2018).

Methods for quantifying the NLF include clinical grading scales, image-based analyses, and device-based technologies (Hamer et al, 2018). These approaches have improved precision but still highlight variability in accuracy and reproducibility. Although the mechanism of the NLF formation have been studied extensively, the underlying genetic factors have not yet been identified, presumably because of the limitations of previous phenotyping methodologies and the limited sample size.

In this study, we developed a deep learning–based method to measure the NLF and performed the largest GWAS of the NLF to date in 10,235 Han Chinese. The discovery sample included a total of 3091 individuals from National Survey of Physical Traits cohort. Studies in this cohort were approved by the Ethics Committees of Fudan University (14117) and the Shanghai Institutes for Biological Sciences (ER-SIBS-261410). The replication sample included a total of 7144 individuals from 2 cohorts: Taizhou Longitudinal cohort of Chinese origin (n = 2911) and Jidong cohort (n = 4233), which were conducted with the approval of the Ethics Committee of Fudan University (Ethics Research Approval 85) (Shanghai, China) and the Ethics Committee of Kailuan General Hospital of Tangshan City, Jidong Oilfield Branch, National Petroleum Corporation in July, 2013 (approval number 2013 YILUNZI1). All participants provided written informed consent. Phenotyping details are provided in Supplementary Materials and Methods. In brief, we generated the NLF score (0–100) from high-resolution 2-dimensional facial images, with a 2-step deep learning network, involving a multitask learning of facial segmentation and a detection solution. Significant differences were observed among the 3 cohort in terms of age ($P = 3.87 \times 10^{-286}$), sex distribution ($P = 1.45 \times 10^{-50}$), and the NLF score ($P < .001$) (Table 1). Among 7324 individuals in the National Survey of Physical Traits and Jidong cohorts, we also collected the 3-dimensional facial data using the 2-pod 3dMDface camera system. Each 3-dimensional facial images contained 21 landmarks for the quantification of facial morphology. The deep learning–based NLF evaluation system demonstrated strong correlation with human assessment in 2 independent cohorts (Pearson correlation coefficient for National Survey of Physical Traits cohort = −0.82, Pearson correlation coefficient for Taizhou Longitudinal cohort = −0.62) (Supplementary Figure S1).

Distributions of the NLF were statistically not normal (Shapiro–Wilk test, $P < .05$). Z-transformed NLF score was used in the subsequent genetic association analysis (Figure 1a). The meta-analysis identified a total of 29 SNPs at 2q31.1, reaching the genome-wide significance ($P < 5 \times 10^{-8}$) (Figure 1b and Supplementary Table S1) and being consistent in 3 cohorts, all at least reaching the nominal significance ($P < .05$) and with betas of the same

Abbreviation: NLF, nasolabial fold

Accepted manuscript published online October 24, 2025; corrected proof published online November 14, 2025

Table 1. Baseline Characteristics of the Individuals

Characteristics	NSPT (n = 3091)	TZL (n = 2911)	JD (n = 4233)	*P*-Value
Age, y, mean (SD)	50.00 (12.87)	56.30 (9.50)	45.19 (13.34)	3.87×10^{-286}
Female, n (%)	1942 (62.83)	1845 (63.38)	2047 (48.36)	1.45×10^{-50}
The NLF, mean (SD)	55.19 (29.09)	36.21 (23.60)	71.75 (22.23)	<.001

Abbreviations: JD, Jidong; NLF, nasolabial fold; NSPT, National Survey of Physical Traits; TZL, Taizhou longitudinal.

Baseline characteristics are presented as means (SD) for quantitative traits and as numbers (percentages) for categorical variables. *P*-values were calculated using the Kruskal–Wallis test for continuous variables and chi-square tests for categorical variables.

direction. No genome-wide significant association signal was observed after conditioning on the lead SNP rs6729788 (Supplementary Figure S2). Fine-mapping analysis further prioritized rs6729788 as the most likely causal variant within this locus (posterior probability = 0.66) (Supplementary Table S2).

Functional annotation analysis of rs6729788 identified 8 interacting genes from antp homeobox family (*HOXD*) (Figure 1c), and these genes are highly enriched in reproductive system tissues, kidney tissues, adipose tissues, and skin tissues (Supplementary Figure S3). Genes near this locus have both been reported to play important roles in facial morphology, particularly among East Asian populations (Huang et al, 2021; Wu et al, 2019). Moreover, *Hox* genes exhibit pronounced age-dependent expression changes in dermal fibroblasts (Ko et al, 2024), potentially influencing NLF severity by promoting matrix metalloproteinase 2–mediated collagen degradation and altering superficial musculoaponeurotic system integrity. Epigenetic markers indicated a regulatory function of rs6729788 (Figure 1d). The derived T allele of the lead SNP rs6729788 increased the severe level of the NLF ($\beta = -2.00$, $P = 1.94 \times 10^{-15}$) (Figure 1e and Supplementary Figure S4). This allele is low frequent in East Asian populations but more common in European populations (Figure 1e), consistent with reports of more severe NLFs in Europeans (Vashi et al, 2016)

To further investigate the relationship between facial morphology and the NLF, we calculated the Euclidean distances between 21 landmark points derived from 3-dimensional facial data to measure facial morphology features. We then analyzed the correlation between these Euclidean distances and the NLF, finding that the morphology of the lower facial region is significantly associated with the NLF (Supplementary Table S3). In addition, SNP-based causal inference test analysis suggested that the SNP influences the NLF through its effect on the Euclidean distances

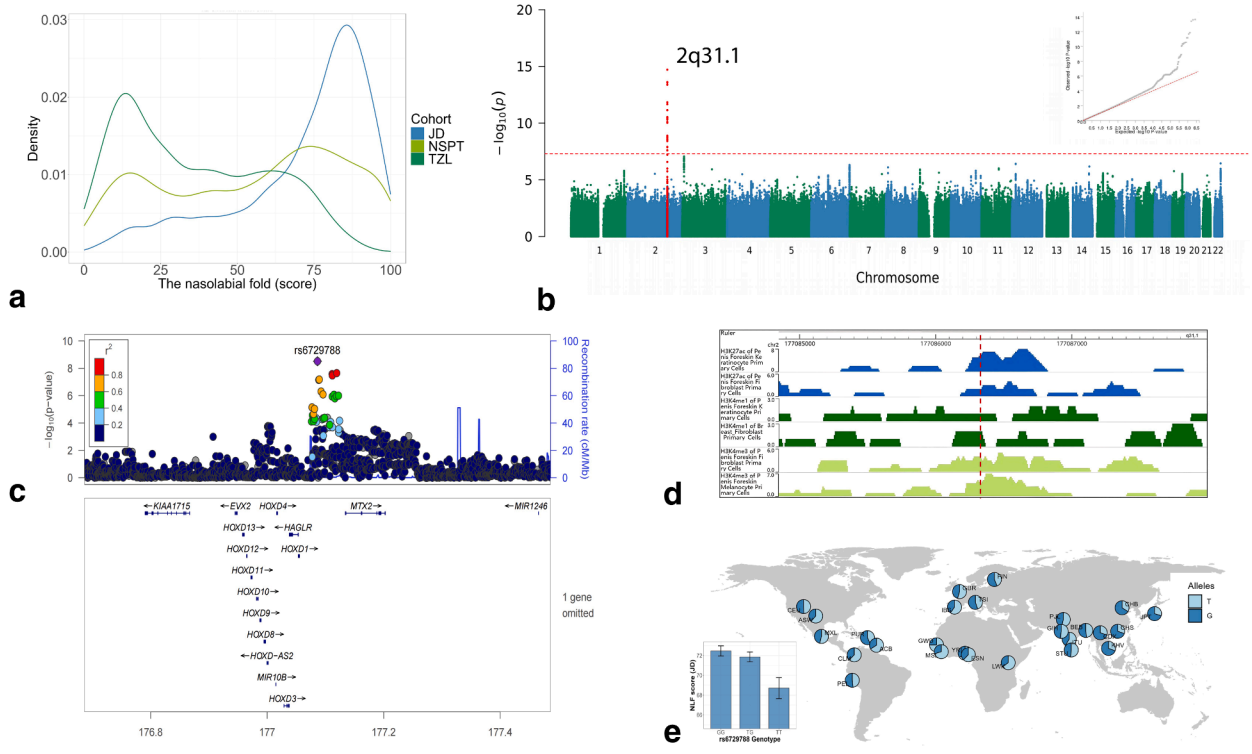


Figure 1. A meta-analysis of GWAS for the NLF identified a signal at 2q31.1. (**a**) Histogram of the NLF in the discovery cohort (NSPT) and replication cohorts (JD and TZL). (**b**) Manhattan plot of the meta-analysis results for the NLF from the GWASs (n = 10,235). The $-\log_{10}$*P*-values for association were plotted for all SNPs according to their physical positions (genome-build GRCh37). The red line was corresponding to the threshold for genome-wide significance ($P = 5 \times 10^{-8}$). (**c**) Regional association plot for the associated region at 2q31.1. Color from blue to red represents increasing linkage disequilibrium (r^2) with rs6729788. (**d**) Epigenetic markers H3k27ac, H3K4me1, and H3Kme3 at 2q31.1 region. The dotted line showed the position of rs6729788. (**e**) Geography of allele frequency on rs6729788. The effect allele T on rs6729788 showed an increasing effect on more severe nasolabial folds. Data are from the 1000 Genomes Project (phase 3) through Ensembl Genome Browser, including 26 populations across 5 continental groups (n = 2504 individuals in total; sample size per population: 60–113). JD, Jidong; NLF, nasolabial fold; NSPT, National Survey of Physical Traits; TZL, Taizhou longitudinal.

representing facial morphology features. This result remained significant after applying false discovery rate correction (adjusted *P* for causal inference test < .05) (Supplementary Table S4), consistent with findings from previous anatomical studies (Sandulescu et al, 2019). To further validate the robustness of the result, we conducted colocalization and 2-sample Mendelian randomization analyses for the facial traits with significant causal inference test results. These analyses revealed that only a subset of traits (ie, AlR_ExR, ChR_ExR, EnR_EnL, and EnR_ExL) showed strong evidence of sharing a causal variant with NLF (Supplementary Tables S4–6) and exerted significant causal effects on NLF severity. Average face showed a significant difference in the lower-face morphology between the highest and the lowest NLF score groups (n = 60). Notably, the highest NLF group showed a longer middle face with protruding mouth (Supplementary Figure S5).

In conclusion, by applying a deep learning–based phenotyping approach, we identified a genome-wide significant association between the *HOXD* locus and NLF severity, highlighting a genetic link between facial morphology and skin aging. One limitation of our study is the lack of direct experimental validation for the identified genetic locus. Although functional analysis suggests regulatory roles for nearby *HOXD* genes, causal mechanisms remain unconfirmed. In addition, because our analysis was conducted exclusively in Han Chinese individuals, the generalizability of the findings to other ancestral populations remains unknown. Future studies incorporating multiethnic cohorts, CRISPR-based perturbation in skin cell models, and single-cell transcriptomic data from human and mouse skin to clarify functional consequences are needed.

ETHICS STATEMENT

This study was approved by the Ethics Committees of Fudan University (approval numbers 14117 and 85); Shanghai Institutes for Biological Sciences (ER-SIBS-261410); and the Ethics Committee of Kailuan General Hospital of Tangshan City, Jidong Oilfield Branch, National Petroleum Corporation (approval number 2013 YILUNZI1). Written informed consent was obtained from all participants.

DATA AVAILABILITY STATEMENT

The GWAS summary statistics are available from the National Omics Data Encyclopedia (http://www.biosino.org/node/analysis/detail/OEZ00021305). Data usage must be in full compliance with the Regulations on Management of Human Genetic Resources in China. Individual genotype and phenotype data cannot be shared owing to Institutional Review Board restrictions on privacy concerns. Other relevant data supporting the key findings of this study are available within the letter and supplementary materials or from the corresponding author on reasonable request. The deep learning–based nasolabial fold scoring model developed in this study has been deployed as a cloud-based API: https://u.ss5.xyz/nlfscore.

KEY WORDS

Facial morphology; GWASs; Nasolabial fold

ORCIDs

Fudi, Wang: https://orcid.org/0000-0002-0208-3343
Zhao, Yuepu: https://orcid.org/0000-0002-4179-7533
Siyuan, Du: https://orcid.org/0000-0003-1602-1669
Jiarui, Li: https://orcid.org/0000-0003-0581-5850
Zhiyang, Li: https://orcid.org/0009-0007-7686-0700
Yonghao, Luo: https://orcid.org/0009-0000-1510-0790
Siying, Fu: https://orcid.org/0009-0009-5341-1333
Wang, Sijia: https://orcid.org/0000-0001-6961-7867

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant number XDB38020400 to SW), National Natural Science Foundation of China (grant number 32325013), CAS Project for Young Scientists in Basic Research (grant number YSBR-077), and Shanghai Science and Technology Commission Excellent Academic Leaders Program (22XD1424700) and was supported by the Human Phenome Data Center of Fudan University.

AUTHOR CONTRIBUTIONS

Conceptualization: SW; Data Curation: FW, YZ, SD, JL, ZL; Formal Analysis: FW, YZ; Funding Acquisition: SW; Investigation: FW, YZ; Methodology: FW, YZ, YL; Project Administration: SW; Resources: SW, ZL, SF; Software: FW, YZ, YL; Supervision: SW; Validation: FW, YZ, SW; Visualization: FW, YZ, SW; Writing - Original Draft Preparation: FW, YZ, SW; Writing - Review and Editing: FW, YZ, SD, JL, ZL, YL, SF, SW

***Fudi Wang*[1,2,3], *Yuepu Zhao*[1,3], *Siyuan Du*[1], *Jiarui Li*[1], *Zhiyang Li*[2], *Yonghao Luo*[2], *Siying Fu*[2] and *Sijia Wang*[1,*]**

[1]*CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China; and* [2]*EveLab Insight (Singapore), Singapore, Singapore*

[3]*These authors contributed equally to this work.*

[*]*Corresponding author e-mail: wangsijia@sinh.ac.cn*

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at 10.1016/j.jid.2025.09.382.

REFERENCES

Cha S, Lim JE, Park AY, Do JH, Lee SW, Shin C, et al. Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genomics 2018;19:481.

Coleman SR, Grover R. The anatomy of the aging face: volume loss and changes in 3-dimensional topography. Aesthet Surg J 2006;26:S4–9.

Cotofana S, Fratila AA, Schenck TL, Redka-Swoboda W, Zilinsky I, Pavicic T. The anatomy of the aging face: a review. Facial Plast Surg 2016;32:253–60.

Hamer MA, Pardo LM, Jacobs LC, Deelen J, Uitterlinden AG, Slagboom E, et al. Facial wrinkles in Europeans: a genome-wide association study. J Invest Dermatol 2018;138:1877–80.

Huang Y, Li D, Qiao L, Liu Y, Peng Q, Wu S, et al. A genome-wide association study of facial morphology identifies novel genetic loci in Han Chinese. J Genet Genomics 2021;48:198–207.

Ko D, Mun S, Kim M, Nho YH, Lee DG, Kang S, et al. A glance into the destiny of transcriptomic activity, embodied by the HOX genes, in neonatal and aging dermal cells. Adv Biol (Weinh) 2024;8:e2300325.

Mayes AE, Murray PG, Gunn DA, Tomlin CC, Catt SD, Wen YB, et al. Ageing appearance in China: biophysical profile of facial skin and its relationship to perceived age. J Eur Acad Dermatol Venereol 2010;24:341–8.

Sandulescu T, Franzmann M, Jast J, Blaurock-Sandulescu T, Spilker L, Klein C, et al. Facial fold and crease development: a new morphological approach and classification. Clin Anat 2019;32:573–84.

Vashi NA, de Castro Maymone MB, Kundu RV. Aging differences in ethnic skin. J Clin Aesthet Dermatol 2016;9:31–8.

Wu W, Zhai G, Xu Z, Hou B, Liu D, Liu T, et al. Whole-exome sequencing identified four loci influencing craniofacial morphology in northern Han Chinese. Hum Genet 2019;138:601–11.

SUPPLEMENTARY MATERIALS AND METHODS

Study populations

The Jidong cohort. The Jidong cohort is a community-based, long-term observational cohort study to evaluate health-related risk factors. The baseline data were collected from 2013 to 2014 in the Staff Hospital, Jidong Oilfield Branch. Approval was obtained from the Ethics Committee of Kailuan General Hospital of Tangshan City and the Medical Ethics Committee, Staff Hospital, Jidong Oilfield Branch, China National Petroleum Corporation in July 2013 (approval number 2013 YILUNZI1). The approval had been renewed in 2018. To date, 9078 individuals aged >18 years have been enrolled after excluding individuals who were unable or unwilling to participate. Written informed consent was obtained from all participants. This study included a total of 4233 individuals (2176 males and 2047 females, aged 20–80 years) who paid the return visit in 2018. The facial photograph and blood samples were collected in the Staff Hospital at the same time.

The National survey of physical traits cohort. The National Survey of Physical Traits cohort is a population-based prospective cohort study of Han Chinese individuals recruited from 3 sites (ie, Taizhou, Nanning, and Zhengzhou) between 2015 and 2018, designed to explore environmental and genetic factors associated with physical traits and diseases. The National Survey of Physical Traits cohort study was conducted with the official approval of the Ethics Committees of Fudan University (14117) and the Shanghai Institutes for Biological Sciences (ER-SIBS-261410). All individuals provided written informed consent. Portrait photos of 3091 individuals (1149 males and 1942 females, aged 17–83 years) were collected in accordance with phenotyping standard operating procedure in this study.

The Taizhou longitudinal cohort. The Taizhou longitudinal cohort study is a long-term observational cohort study to explore the environmental and genetic risk factors for common and noncommunicable diseases. This research program was conducted with the approval of the Ethics Committee of Fudan University (ethics research approval 85). The detailed characteristics have been described before (Wang et al, 2009). Our replication set included 2911 health Han Chinese with portrait photos (1039 males and 1869 females), ranging from ages 31 to 87 years.

Facial data collection

All participants were asked not to take part in vigorous exercise an hour before their study visit, not to wear make-up, and to refrain from alcohol and tobacco use 24 hours before the visit. Facial photographs were taken in a standardized space with a stabilized LED light source to ensure uniform illumination across subjects. Besides, all participants wore a shawl to help give consistent light illumination. A Canon 70D digital camera (lens: Canon EF 40 mm f/2.8) was used for all participants without the flash. The facial photograph for each participant consisted of a frontal facial shot with the eyes closed and no facial expression. The resolution of photographs was 300 dpi.

Phenotyping

Nasolabial folds (NLFs) are bilateral skin creases that extend from the sides of the nose (alae nasi) to the corners of the mouth. Traditional evaluation of the NLFs was performed using the SCINEXA scale (Vierkötter et al, 2009). To better quantify NLFs, this study adopted the following 3 main steps to train the algorithm for NLF phenotyping:

1. Facial landmark detection: Facial keypoint detection algorithms were used to locate the regions of interest for NLFs. The green boxed region in the figure indicates the periocular wrinkle area (Supplementary Figure S6a).
2. Independent models for segmentation and score prediction: Independent segmentation and score prediction models are trained for NLF. These models including the following:
   (i) U-Net–like structure: the overall architecture is based on a U-Net, with modifications to the connection methods between feature layers (Supplementary Figure S6a).
   (ii) IBN module integration in encoder: the convolutional layers in the encoder were combined with IBN modules (Pan et al, 2018), leveraging domain adaptation properties to enhance compatibility with input images across different skin tones and lighting conditions. It consisted of 5 downsampling stages, each constructed using IBN-b convolution blocks as the fundamental unit. The IBN-b blocks were repeated 2, 3, 4, 4, and 3 times sequentially across the stages, respectively, progressively increasing and then stabilizing the feature extraction depth. This design enables hierarchical feature learning while maintaining computational efficiency in deeper layers.
   (iii) Decoder: This employed bilinear interpolation to upsample lower-resolution feature maps, followed by concatenation with corresponding higher-resolution skip connections from the encoder. Each merged feature map was then processed through 2 consecutive Conv-BN-ReLU blocks (convolution → batch normalization → ReLU activation) to refine the upsampled features.
   (v) Multiscale feature fusion for score prediction: features from intermediate layers of the encoder and decoder are fused to improve the accuracy and stability of score predictions (Supplementary Figure S6b).
3. Loss functions: segmentation loss: On the basis of BASNet (Qin et al, 2019), this consisted the following 3 components:
   (i) Weighted dice loss: an image-level loss function derived from dice loss (Sudre et al, 2017), making it particularly effective for imbalanced datasets where target regions occupy small areas. It optimized the score by introducing a weight parameter $w \in [0,1]$ to balance precision and recall, with a logarithmic function applied to amplify gradient values. A constant C prevents division by zero:

$$\mathscr{L}_{\text{weighed\_dice}} = -\log \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} S\left(x_i, y_j\right) G\left(x_i, y_j\right) + C}{w\sum_{i=1}^{W}\sum_{j=1}^{H} S\left(x_i, y_j\right) + (1-w)\sum_{i=1}^{W}\sum_{j=1}^{H} G\left(x_i, y_j\right) + C}$$

(ii) Structural Similarity Index Measure (SSIM) loss: a patch-level loss function that calculates structural similarity for patches of size K × K between segmentation results and ground truth. It is originally designed for image quality evaluation, with particular emphasis on structural feature retention. For a given patch, the SSIM loss is computed as follows:

$$\mathscr{L}_{\text{ssim\_patch}} = -\log \frac{\left(2\mu_s\mu_g + C_1\right)\left(2\sigma_{sg} + C_2\right)}{\left(\mu_s^2 + \mu_g^2 + C_1\right)\left(\sigma_s^2 + \sigma_g^2 + C_2\right)}$$

(iii) Binary Cross-Entropy (BCE) loss: a pixel-level loss function employing cross-entropy computation at each image pixel location. It is the most widely used loss in binary segmentation, which is computed as follows:

$$\mathscr{L}_{bce} = -\frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\left[G\left(x_i, y_j\right)\log S\left(x_i, y_j\right) + \left(1 - G\left(x_i, y_j\right)\right)\log\left(1 - S\left(x_i, y_j\right)\right)\right]$$

Score loss: standard regression loss is used:

$$\mathscr{L}_{score} = \text{Smooth}L_1\left(y_s - y_g\right)$$

4. Evaluation metrics: To assess segmentation performance, we adopted multiple pixel-level validation metrics. Specifically, the F-score (0.75) was used to summarize the harmonic mean of precision and recall, offering a balanced measure of model accuracy in delineating the NLF region. The mean absolute error between predicted NLF severity scores and expert-assigned wrinkle grades was 0.79, indicating strong agreement with clinical evaluation. These metrics collectively demonstrate the robustness and clinical relevance of our deep learning framework.
5. Experimental setup: We conducted training and validation using the Jidong dataset comprising 4233 facial images. The dataset was randomly split into 3809 images for training and 424 for validation (90−10% ratio). For model input, bilateral NLF regions were cropped to 256 × 128 pixels. Following the skin aging atlas protocol (Bazin and Flament, 2010), all images were clinically graded from 100 (mild) to 0 (severe) on the basis of NLF depth, the base of the nose, and the corner of lips. The segmentation masks were annotated to align with the NLF's central axis, defined as the deepest fold trajectory. We used a weight of w = 0.4 and C = 1e-7 in the weighted dice loss and set K = 11 in the SSIM loss. The network was optimized using Adam optimizer (Kingma and Ba, 2014[1]) with hyper parameters as follows: learning rate lr = 1e-3, betas = (0.9, 0.999), eps = 1e-8, and weight decay = 1e-5. Training converged after 500 epochs (batch size = 64) as monitored by validation loss plateau.

[1] Kingma DP, Adam BJ. A method for stochastic optimization. arXiv 2014.

Genotyping

For both Jidong and National Survey of Physical Traits cohorts, genomic DNA was extracted from blood samples using the MagPure Blood DNA KF Kit. All samples were genotyped using the Illumina Infinium Global Screening Array consisting of about 710,000 SNPs. We implemented exclusion criteria for quality control using PLINK (version 1.9) (Chang et al, 2015). In detail, we excluded participants with >5% missing data, the duplicated participants, and participants who failed the X-chromosome sex concordance check or had ethnic information incompatible with their genetic information. We excluded SNPs that had more than 2% missing data, those with a minor allele frequency of less than 1%, and the ones that failed Hardy−Weinberg deviation test ($P < 1 \times 10^{-5}$). Imputation was performed using the 1000 Genomes Project phase 3 as the reference panel. The chip genotype data were first phased using SHAPEIT3 (Delaneau et al, 2019), and IMPUTE2 (Howie et al, 2009) was then used to impute genotypes. SNPs with an imputation quality score (INFO) <0.6, minor allele frequency <0.01, or a missing rate >0.01 were eliminated from further analyses. Finally, 8,039,700 SNPs passed quality control and were used for further analyses.

For Taizhou longitudinal cohort, blood samples were collected, and DNA was extracted. All samples were genotyped using the Illumina HumanOmniZhongHua-8 chip, which

interrogates 894,517 SNPs. After quality control with PLINK (version 1.9), the genotype data were phased using SHAPEIT and were imputed using IMPUTE2 with the 1000 Genomes Project phase 3.

Statistical analysis

GWASs and meta-analysis. GWASs were separately conducted in National Survey of Physical Traits and 2 replication cohorts (Taizhou longitudinal and Jidong cohorts) on the Z-transformed NLF score using software package PLINK. Genetic principal components were calculated from the chip genotype data. Additive allele effects were tested in linear models adjusted for covariates (age, sex, sampling locations, and the top 5 genomic principal components). GWAS outputs were combined using inverse variance fixed-effect meta-analysis as implemented in METAL (Willer et al, 2010). *P*-values $\leq 5 \times 10^{-8}$ from the meta-analysis were considered as genome-wide significant. The inflation factor was estimated close to 1.0 ($\lambda = 0.995$) and not further considered. GWAS results were visualized using Manhattan plots, Q–Q plots. Regional linkage disequilibrium plots and association plots were generated using Haploview and LocusZoom.

Causal inference test analysis. The causal inference test was performed with the R package 'cit' (version 2.3.2) (Millstein et al, 2016). Briefly, the causal inference test requires the following 4 criteria to be met for a valid causal inference: (i) SNP and the NLF are associated, (ii) SNP is associated with facial morphology feature after adjusting for the NLF, (iii) facial morphology feature is associated with the NLF after adjusting for SNP, and (v) SNP is independent of the NLF after adjusting for facial morphology feature. The maximum *P*-value among the 4 tests was reported as the P_{CIT} value. Results with P_{CIT} value < 0.05 were considered statistically significant.

Colocalization analysis. Colocalization analysis was performed using a Bayesian approach in the R package 'coloc' (version 5.1.0) (Giambartolomei et al, 2014). The method evaluates whether 2 association signals share a common causal variant. Genetic association summary statistics for all facial Euclidean distance traits were obtained from a meta-GWAS, conducted in the same cohort as the NLF GWAS described earlier. Analyses were conducted within ±250 kb of the lead SNP, using default prior probabilities. Posterior probabilities were calculated for 5 hypotheses, with PP.H4.abf ≥ 0.9 considered strong evidence for colocalization.

Two-sample Mendelian randomization analysis. The Mendelian randomization analysis was conducted using the 'TwoSampleMR' package (version 0.5.4) (Hemani et al, 2018) to explore the association between the NLF and facial morphology. Given the limited number of SNPs that met the GWAS significance threshold of 5e-8, we relaxed the threshold to 5e-5. Multiple complementary approaches (Mendelian randomization Egger, inverse variance weighted, weighted median, simple mode, and weighted mode) were used to estimate the effects of the exposure and outcome. Heterogeneity was assessed using Cochran's *Q* statistic, and horizontal pleiotropy was evaluated using the Mendelian randomization PRESSO global test (Verbanck et al, 2018), with $P < .05$ indicating significant pleiotropy.

Fine-mapping analysis. To prioritize likely causal variants within the genome-wide significant locus at 2q31.1, we performed statistical fine mapping using the PAINTOR (version 3.0) algorithm. We selected all SNPs located within ±50 kb of the lead SNP rs6729788 and included their Z-scores from GWAS summary statistics. Analyses were performed under models assuming either 1 or 2 causal variants. Posterior probabilities for each SNP were calculated to assess their likelihood of being causal.

Functional analysis. The gene expression heatmap of signal SNPs was visualized using FUMA (Watanabe et al, 2017) on the basis of Genotype-Tissue Expression data. Functional annotation was then visualized using the WashU Epigenome Browser (Li et al, 2022) where we displayed histone modification peaks (H3K27ac, H3K4me1, H3K4me3) for normal human foreskin fibroblasts and keratinocytes, obtained from the ENCODE database. These genomic elements are indicative of regulatory regions, providing insights into the potential functional roles of the SNPs in gene expression and facial morphology.

Facial traits and average face

Three-dimensional facial scans were captured using the dual-lens 3dMDface system. The raw data were preprocessed using MeshMonk (White et al, 2019) following the following steps: (i) the scans were preprocessed by manually marking the nasal tip as the origin, (ii) normalization of all face using an anthropometric mask template with 7906 dense landmarks, (iii) rigid registration based on the Iterative Closest Point algorithm, (iv) nonrigid registration based on the Thin-Plate Spline algorithm, and (v) the samples were aligned to a mean face using generalized Procrustes analysis. Each mapped image was manually checked, and any abnormal images with poor quality were deleted.

We further extracted facial morphological features on the basis of 3-dimensional facial data. In this study, we focused on 119 facial traits as paired distance between 21 previously defined facial landmarks (Xiong et al, 2019) (Supplementary Figure S7). To visualize the relationship between the NLF and facial morphology, principal component analysis clustering was performed on the Euclidean distances in the nasolabial region. On the basis of the first principal component, the top 30 and bottom 30 samples were selected to represent the largest and smallest changes in lower face, respectively. The average face for each group was generated by calculating the mean X, Y, Z coordinates from the 7906 facial landmarks of these selected samples.

SUPPLEMENTARY REFERENCES

Bazin R, Flament F. Skin aging atlas. In: Asian type2. Paris: Med'Com; 2010.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4. s13742-015-0047-8:7.

Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. Nat Commun 2019;10:5436.

Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383.

Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;7: e34408.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the

next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.

Li D, Purushotham D, Harrison JK, Hsu S, Zhuo X, Fan C, et al. WashU epigenome browser update 2022. Nucleic Acids Res 2022;50:W774–81.

Millstein J, Chen GK, Breton CV. Cit: hypothesis testing software for mediation analysis in genomic applications. Bioinformatics 2016;32:2364–5.

Pan X, Luo P, Shi J, Tang X. Two at once: enhancing learning and generalization capacities via ibn-net. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Proceedings of the European conference on computer vision (ECCV). Cham, Switerzland: Springer; 2018. p. 484–500.

Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M. Basnet: boundary-aware salient object detection. Long Beach, CA: Paper presented at: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 15–20 June 2019.

Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2017) 2017;2017:240–8.

Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases [published correction appears in Nat Genet 2018;50:1196]. Nat Genet 2018;50:693–8.

Vierkötter A, Ranft U, Krämer U, Sugiri D, Reimann V, Krutmann J. The SCINEXA: a novel, validated score to simultaneously assess and differentiate between intrinsic and extrinsic skin ageing. J Dermatol Sci 2009;53:207–11.

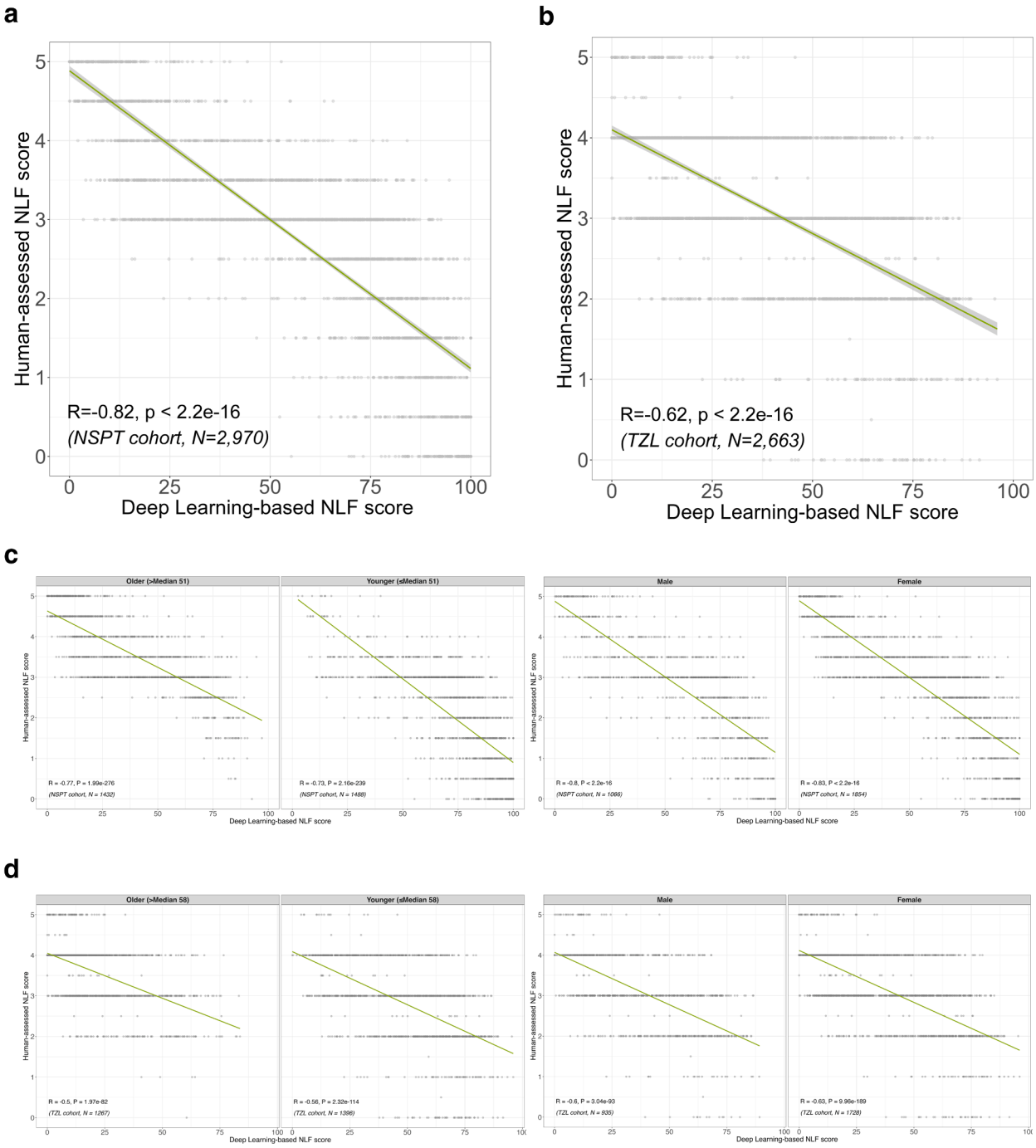
Wang X, Lu M, Qian J, Yang Y, Li S, Lu D, et al. Rationales, design and recruitment of the Taizhou Longitudinal Study. BMC Public Health 2009;9:223.

Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.

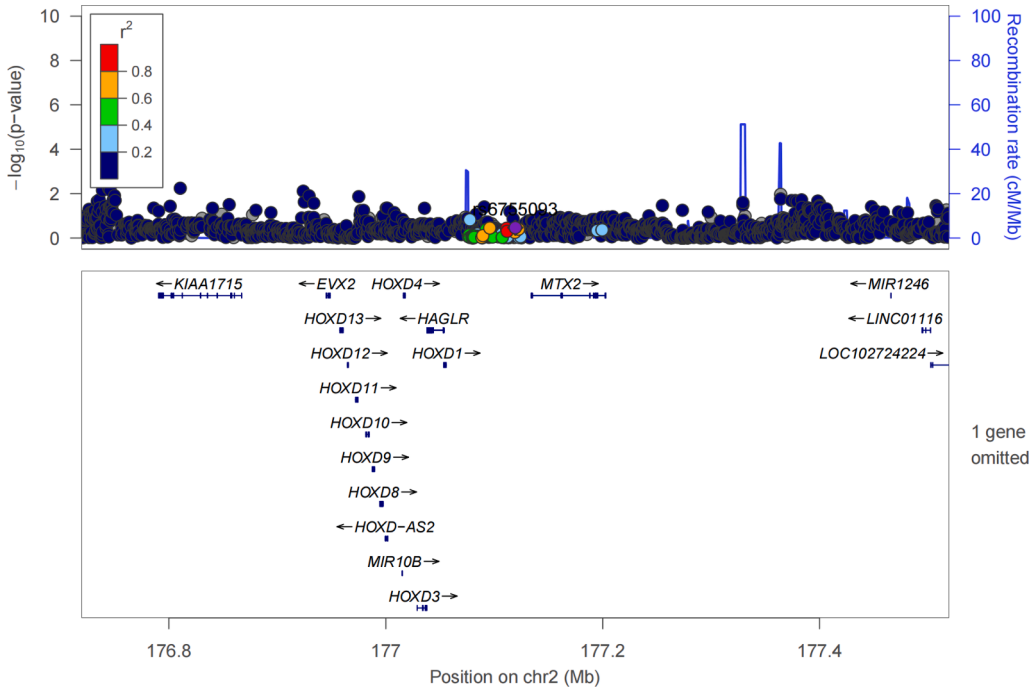
White JD, Ortega-Castrillón A, Matthews H, Zaidi AA, Ekrami O, Snyders J, et al. MeshMonk: open-source large-scale intensive 3D phenotyping. Sci Rep 2019;9:6085.

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.

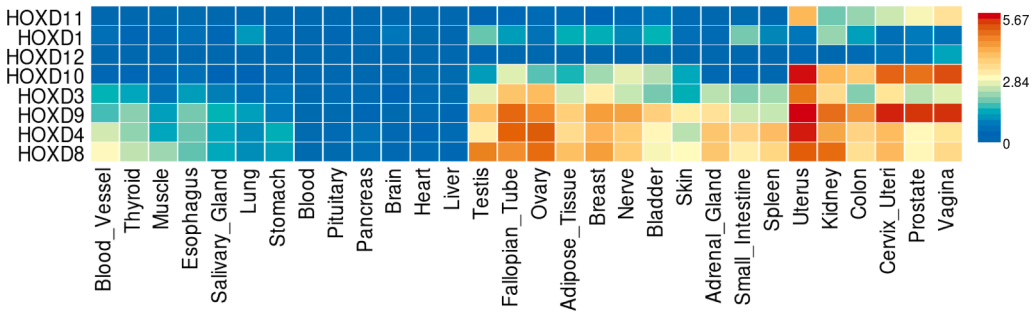
Xiong Z, Dankova G, Howe LJ, Lee MK, Hysi PG, De Jong MA, et al. Novel genetic loci affecting facial shape variation in humans. eLife 2019;8:e49898.



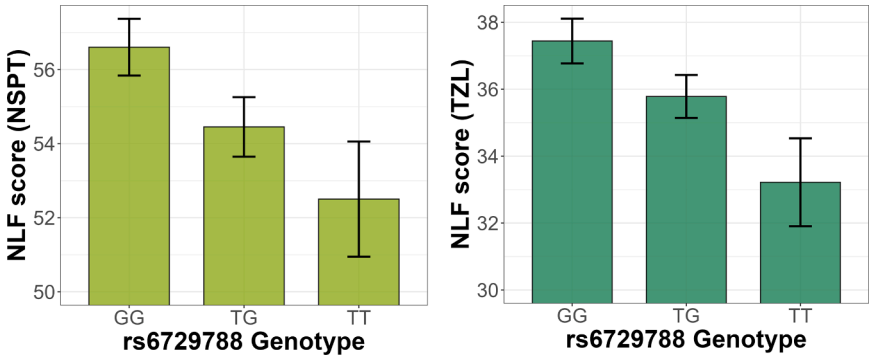
Supplementary Figure S1. Pearson correlation of NLF between human assessment and deep-learning evaluating system in NSPT and TZL. X-axis indicates NLF scores from the deep-learning system, and Y-axis indicates NLF scores from human assessment. The correlation coefficient (R) is provided for each cohort to indicate the strength of the relationship between the 2 methods. (**a, b**) Results for NSPT and TZL. (**c, d**) Results stratified by age and sex in NSPT and TZL, respectively. NLF, nasolabial fold; NSPT, National Survey of Physical Traits; TZL, Taizhou longitudinal.



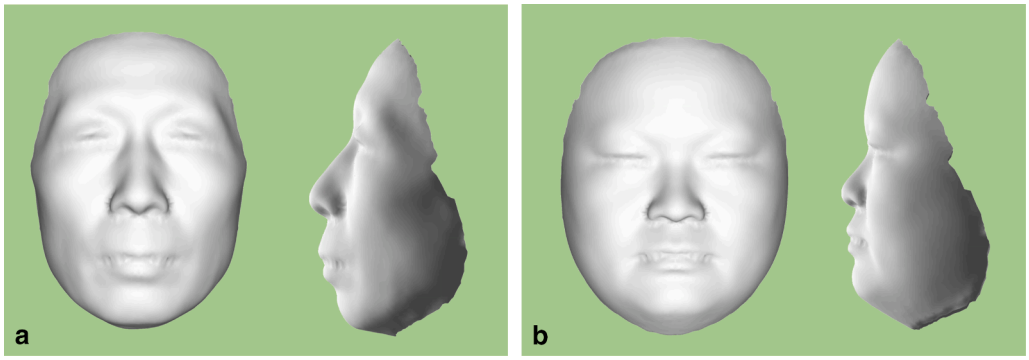
Supplementary Figure S2. Functional annotation analysis at 2q31.1. Regional association plot for the associated region at 2q31.1 after controlling the leading SNP rs6729788.



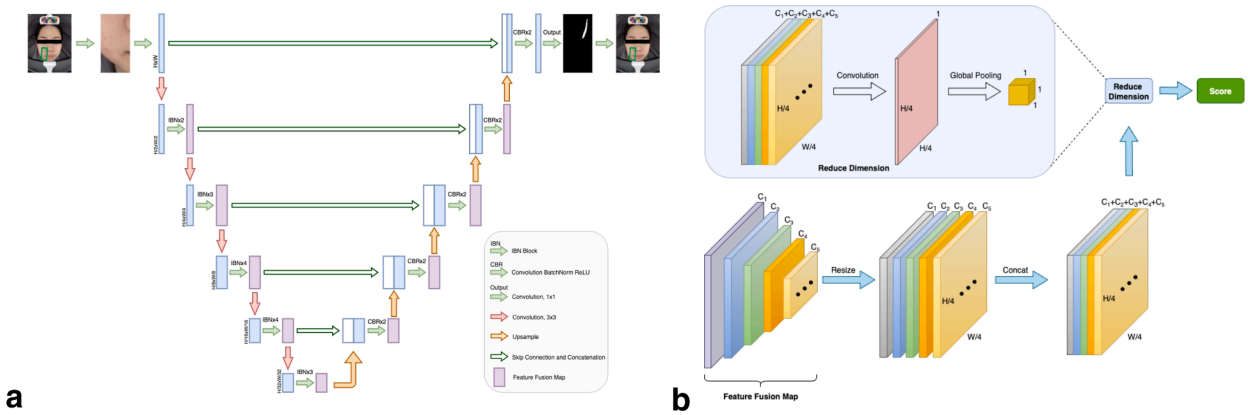
Supplementary Figure S3. Heatmap showing the expression of genes near rs6729788 in different tissues from GTEx data. The rows represent the genes, and the columns represent different tissue types. Gene expression is shown on a blue-to-red scale, with blue indicating low expression and red indicating high expression, highlighting the variation across tissues. GTEx, Genotype-Tissue Expression.



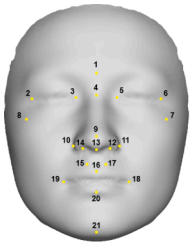
Supplementary Figure S4. Distribution of NLF scores by genotype groups of rs6729788. Bar plots show the NLF scores in the (**a**) NSPT cohort and the (**b**) TZL cohort. Each bar represents the mean NLF score for a genotype group, with error bars indicating the SEM. NLF, nasolabial fold; NSPT, National Survey of Physical Traits; TZL, Taizhou longitudinal.



Supplementary Figure S5. Average facial morphology differences by NLF severity in the NSPT cohort. Panels show the (**a**) highest and (**b**) lowest NLF score groups. NLF, nasolabial fold; NSPT, National Survey of Physical Traits.



Supplementary Figure S6. Diagram of the nasolabial fold detection algorithm. (**a**) Deep learning framework. (**b**) Feature maps schematic diagram. The individual shown is a volunteer participant in our study who provided written informed consent for the deidentified image to be used in publication.



No.	Name	Abbr.	No.	Name	Abbr.	No.	Name	Abbr.
1	Glabella	G	8	Zygion Left	ZyL	15	Philtrum Left	PhL
2	Exocanthion Left	ExL	9	Pronasale	Prn	16	Labiale Superius	Ls
3	Endocanthion Left	EnL	10	Alare Ledt	AIL	17	Philtrum Right	PhR
4	Nasion	N	11	Alare Right	AIR	18	Cheilion Right	ChR
5	Endocanthion Right	EnR	12	Subnasale Right	SnR	19	Cheilion Left	ChL
6	Exocanthion Right	ExR	13	Subnasale	Sn	20	Labiale Inferius	Li
7	Zygion Right	ZyR	14	Subnasale Left	SnL	21	Gnathion	Gn

Supplementary Figure S7. A detailed overview of the 21 facial anatomical landmarks, including their precise locations and definitions. The left panel provides a schematic representation of the human face with the precise locations of the 21 facial anatomical landmarks. The right panel presents a detailed description of each landmark, including its name and abbreviation. Abbr denotes abbreviations, and No denotes number.